SPATIAL ATTAINMENT TRENDS OF RACIAL AND ETHNIC GROUPS IN HOUSTON, TEXAS, 1970 TO 2000

A Dissertation

by

WARREN WAREN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2008

Major Subject: Sociology

SPATIAL ATTAINMENT TRENDS OF RACIAL AND ETHNIC GROUPS IN

HOUSTON, TEXAS, 1970 TO 2000

A Dissertation

by

WARREN WAREN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Mark Fossett
Committee Members,	Dudley Poston
	Rogelio Saenz
	Don Albrecht
Head of Department,	Mark Fossett

December 2008

Major Subject: Sociology

ABSTRACT

Spatial Attainment Trends of Racial and Ethnic Groups in
Houston, Texas, 1970 to 2000. (December 2008)

Warren Waren, B.A., Northeastern State University;
M.A., University of Arkansas
Chair of Advisory Committee: Dr. Mark Fossett

Previous research in the spatial assimilation of racial and ethnic groups has not assessed trends over time due to methodological difficulties and data limitations. I use an innovative method to assess the intercensal changes in neighborhood spatial attainment for African Americans, Hispanics, and non-Hispanic whites in Houston, Texas, between 1970 and 2000. I extend the current literature by showing that an accepted and commonly used method for assessing longitudinal change in spatial attainment is flawed and yields incorrect results. I highlight an alternative approach which makes use of data readily available in Census Summary Files to estimate individual-level spatial attainment regressions. I also show that the choice of neighborhood size affects estimates of spatial attainment effects. Although the influence of spatial scale has been demonstrated in the segregation literature, its consequences for spatial attainment research have not. I investigate and report findings from four geographic scales useful to and commonly used by spatial attainment researchers: the block group, the Census tract, the Zip Code Tabulated Area, and the Public Use Micro Data Area. I compare the benefits and drawbacks of estimating spatial attainment at each level of geography.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

1. INTRODUCTION

My dissertation project assesses patterns and changes in spatial attainment for racial and ethnic groups in Houston, Texas over four time points spanning three decades. The project makes methodological and substantive contributions to the literature on spatial attainment and spatial assimilation. Methodologically, I critique an accepted method for assessing spatial attainment, aggregate regression, and identify a viable alternative method. Substantively, I investigate something not previously reported in the literature—change in spatial attainment over time for multiple racial and ethnic groups. For African Americans in Houston, spatial assimilation into white neighborhoods based on education was not evident in 1970. But by 2000, clear patterns of spatial assimilation emerge for African Americans at the highest levels of education. I assess these patterns at the macro-level using P* contact scores, and at the micro-level using individual-level models of spatial attainment. In my final section I explore the impact of the decision to use small or large areas when assessing spatial attainment.

In Section 2, I review the literature touching on current research in spatial attainment and antecedent works in urban ecology, assimilation, and status attainment. Assimilation is the process through which distinct groups become less distinguishable (Fossett and Cready 1998). Drawing on the work of Gordon (1964), Yinger (1981; 1994), and Alba and Lee (2003) I consider ethnic group assimilation conceptually as a

This dissertation follows the style of *American Sociological Review*.

multi-dimensional process. For my empirical analysis, I specify residential spatial assimilation as an individual spatial attainment process analogous to familiar models of status attainment (Blau and Duncan 1967; Hauser and Featherman 1977; Massey and Denton 1985). This is considered as just one of many possible dimensions of spatial assimilation.[1] I review the development of spatial assimilation research in an urban context from an urban ecological perspective (Lieberson 1963; Park, Burgess, McKenzie, and Wirth 1925). Also, I note recent attempts to refine the conceptualization of assimilation (Alba and Lee 2003; Massey and Fischer 1999; Portes and Rumbaut 2001; Waters 1999; Wright, Ellis, and Parks 2005; Yinger 1981). In an appendix, I review the variety of methodologies used to assess spatial attainment. I compare ideal data and measurement strategies to existing data and methods to highlight the strengths and weaknesses of different approaches to assessing spatial attainment.

Section 3 reviews and critiques an accepted approach for investigating spatial attainment. I show that this approach, which relies on aggregate regression, is flawed. In a later section, I introduce a viable strategy for conducting longitudinal research which can adequately assess spatial attainment.

Sections 4 and 5 present the key set of analyses that investigate my main substantive question—Do spatial attainment effects vary by group and over time? In Section 3, I examine trends in spatial attainment for racial and ethnic groups between 1970 and 2000 in Houston at the macro level. I present macro-level contact scores for

---

[1] Spatial assimilation may be assessed by variables with group differences on area outcomes. Other widely studied aspects of spatial assimilation include income, housing tenure, homeownership, crime, etc.

race and ethnic groups using decade-specific data for the city of Houston between 1970 and 2000. In Section 5 I estimate a simple spatial attainment model estimated using data from Census Summary File. I interpret my results in light of demographic changes in the city over the study period.

In Section 6 I explore the impact of using different areal units in empirical studies assessing spatial attainment. Because of data constraints, spatial attainment researchers have relied on very large areas to approximate neighborhoods. In this analysis, I hypothesize that this miscalculates spatial attainment effects for minorities which are stronger and more easily detected in analyses that draw on smaller spatial areas. Conversely, I anticipate that research using large spatial areas may underestimate the magnitude of spatial attainment effects and/or their statistical significance. I test my hypothesis by estimating spatial attainment models at four distinct levels of spatial analysis: block group, tract, and Zip Code Tabulated Area (ZCTA), and Public Use Microdata Area (PUMA). To perform these analyses, I use data from the Census and from the Houston Area Survey—a survey which provides extensive geographic identification codes for individual respondents.

In the final section of my dissertation I review and discuss the conclusions from each of the previous analytic sections. In my discussion, I discourage further research which relies on aggregate regressions to estimate spatial attainment effects; I encourage researchers to be aware of neighborhood size when assessing spatial attainment; I highlight the utility of using the simple spatial attainment model; I demonstrate the applicability of the simple method to assess trends in spatial attainment over time; and finally, I show that, over the study period, spatial attainment trends are emerging for African Americans in Houston.

2. LITERATURE REVIEW AND CRITICAL ISSUES

2.1. Overview of Perspectives

My dissertation contributes to the literature on spatial assimilation and spatial attainment. As their names imply, both terms apply a spatial dimension to their core concepts. Below, I outline the development of spatial assimilation theory and the use of spatial attainment outcomes to estimate assimilation. I follow this overview with a discussion of some of the more important critiques to spatial assimilation theory. I close with a review and rebuttal of the main theoretical challenge to spatial assimilation, *place stratification theory*.

Spatial assimilation theory draws on the concepts of social distance, assimilation, and status attainment. Early in the history of sociology Goerg Simmel coined the phrase "social distance," using this useful construct to discuss the social construction of space. Simmel posits that sociological differences are often expressed spatially. Lechner (1991: p. 197) notes that Simmel sees, "boundaries themselves … [as] 'sociological,' not spatial facts." Simmel's student Robert Park adapted and applied the concept of social distance to geographic space within metropolitan areas. Groups of different social standing are found to be separated in space. The Chicago School famously mapped the spatial distribution of many sociological variables such as race, ethnicity, income, and language use.

Park also considered social distance a crucial variable in the process of assimilation. Park's views of assimilation were refined by Gordon (1964) and Yinger

(1981; 1994) who describe assimilation as a contingent process, which allows for multiple paths of assimilation over many variables, including language fluency, cultural norms, education, occupation, income, co-residence, friendship, and intermarriage. Different groups may or may not assimilate along different social domains. Contemporary statements on assimilation stress that assimilation is not considered a one-way process, inevitable, or irreversible (Alba and Lee 2003; Fossett and Cready 1998; Lieberson 1980; Yinger 1981). Assimilation may proceed in either direction. For example, it is possible for the majority group to adopt minority culture; such as food, music, or language usage. Assimilation may not proceed at all if the groups in question do not seek to assimilate on certain dimensions. For example, minority groups may wish to protect and preserve established minority culture or social structure. Alternatively, assimilation may not occur because groups may be barred from entering into an assimilative process through discrimination (Gordon 1964; Massey and Denton 1993b; Yinger 1981).

Lieberson (1980) and Fossett and Cready (1998) discuss assimilation theory, as it applies in ecological studies of group competition. They note that the timing, pace, and extent of assimilation all can vary. Its onset may or may not occur. Once initiated, it can proceed slowly or rapidly. It may proceed to the maximum point of dissolving group differences, or it may stop short. Finally, movement toward assimilation can be reversed.

Spatial assimilation is a process through which social distance and associated spatial differences that distinguish group membership are dissolved. However, how do we assess a group’s level of spatial attainment? To answer this question, researchers turned to the literature on status attainment. The status attainment literature founded by Blau and Duncan (1967) investigates the occupational attainment of respondents based on individual-level characteristics such as the respondent's education, the education of the respondent's father, and the status of father's occupation. This literature offers an appropriate methodology to measure the independent and interactive effects of different individual variables on status attainment. Blau and Duncan focused on occupational attainment, but the approach can be extended to a wide range of outcomes such as education, employment, income, or homeownership. Just as status attainment models are used to assess group assimilation on status outcomes, spatial attainment models can be used to assess group assimilation on residential outcomes.

Drawing on the conceptual frameworks outlined in the spatial assimilation and the status attainment literatures, Massey and Mullan (1984) and Massey and Denton (1985) initiated current work in spatial attainment. Their work posits that residential outcomes, including location in urban space, are an attainment outcome analogous to socioeconomic status. Therefore, the status attainment method used by Blau and Duncan (1967) is applicable to residential outcomes. As Park and colleagues argue, the boundaries and distance between groups in a city reflect social distance. Yet, as groups assimilate culturally and structurally, they are also incorporated more proportionally into the area of a city. Space is viewed as a status variable, like education, in which all groups seek to improve. The main assumption of the spatial attainment model is that (Massey and Mullan 1984, p. 94)—"as SES rises . . . minorities attempt to convert their socioeconomic achievements into an improved spatial position, which usually implies assimilation with majority groups."

Critics of the spatial attainment model have suggested possible weaknesses in some of its underlying assumptions. Specifically, critics have raised concerns about the practice of taking percent white as a proxy for neighborhood status; they caution against invoking a normative assumption that percent white is the standard for assimilation; they note that neighborhoods with high percent white are not necessarily the goal of all minority groups; and most significantly, they point out that spatial attainment models do not account for the extreme disadvantage of some groups, especially African Americans. I now review these concerns and discuss their implications for my project.

Wright, Ellis, and Parks (2005) question whether percent white is an adequate proxy for neighborhood status—especially over the last half of the 20th century as US metropolitan cities have become less white. They argue that neighborhoods with higher status yet lower percent white have emerged as middle class minority groups increased in size. Therefore, their position is that percent white is no longer a valid proxy for status.

I note two responses to this position. First, percent white reflects contact with whites and need not be viewed as a proxy for status. Spatial attainment models therefore provide a means for assessing whether assimilation in the form of co-residence has occurred or not. The critical issue then is not the level of percent white attained per se—that may change with changing ethnic demography—the key issue is group *differences* in contact with whites and their differences in how percent white varies with individual characteristics such as education.

Second, the relevance of percent white as a proxy for other residential outcomes—for example, school quality, neighborhood safety, or other amenities—is not easily dismissed. But its relevance should be defended empirically, not simply asserted. My review of this issue for Houston, Texas indicates that percent white *does* correlate strongly with other measures of status. If percent white were not a valid proxy for status, then the correlation coefficients would be quite small. Correlation coefficients from the Houston Area Survey and the Census Summary File for Houston show that percent white correlates strongly with other measures of status: median income in tract (0.75); percent poverty in tract (-0.72); and percent with Bachelor's degree or higher (0.70). Therefore, even though new neighborhood patterns may be emerging, the continuing high correlation of percent white and other measures of status supports its use in spatial attainment models. However, Wright et al.'s (2005) position is well taken. When possible, neighborhood status should be measured directly; not simply assumed to be a correlate of percent white.

Wright et al. (2005) also caution against adopting a normative view that percent white is the desired yardstick of spatial assimilation for minority groups. They warn that using the white middle-class suburbanite as the standard of spatial assimilation reinforces the dominant group (i.e., non-Hispanic white) by measuring every other group's assimilation against something the dominant group has almost by definition (i.e., high percent white). The authors caution,

> Unless spatial assimilation research explicitly decouples neighborhood attainment from proximity to whites in suburbs, rhetorically and empirically, it risks supporting this hegemonic impulse (113).

In my view, their concerns are misplaced. Spatial attainment theory and research need not endorse the goal of achieving proximity to whites. Researchers can readily estimate spatial attainment models without invoking normative assumptions. First, there is the simple descriptive question of whether spatial assimilation is or is not occurring. Second, spatial assimilation theories offer predictions regarding particular patterns of assimilation. These can be tested to see if the patterns are supported or disconfirmed by data. Assimilation is merely predicted to be likely under certain conditions. The idea that high percent white neighborhoods are desirable and prescribed is not a sociological assumption or conclusion. Percent white is only one among many characteristics used by researchers to gauge social interaction between groups. Other characteristics such as neighborhood median income, percent of neighborhood with a college degree, and percent of neighborhood in poverty are frequently employed by spatial attainment researchers in the literature.

Portes and Rumbaut (2001) point out that not all minority groups desire high percent white neighborhoods. Some immigrant groups may wish to protect and preserve their enclaves resulting in segmented assimilation. Spatial assimilation theory anticipates this eventuality. Groups may seek to assimilate specifically to obtain higher contact with whites. Or they may seek socioeconomic outcomes that indirectly promote or impede contact with whites and other groups. Alternatively, groups may have other goals entirely. For example, the Amish have residential goals that are based on neither race nor socioeconomic status. In such a case, assimilation theory predicts little spatial assimilation.

Spatial attainment models provide a means for assessing what in fact is the case. Spatial attainment models which take the group percentage as the dependent variable reveal *whether* groups differ in terms of co-residence and how this is patterned based on individual social characteristics (e.g., education). When a difference is documented, the finding calls for an explanation. Some potential explanations, such as discrimination, are not easily included in the models. But indirect evidence of their impact may be revealed in the residual differences between groups—based on the strong assumption that those residual differences reflect only the impact of discrimination.

The most significant critique current in the literature challenges both spatial assimilation as a theory and spatial attainment as an outcome on empirical grounds. Place stratification theorists emphasize the consistent finding that spatial attainment models do not predict the spatial attainment of African Americans well. They contend that structural forces external to the individual, most notably discrimination, may be overlooked because of spatial attainment's focus on individual-level characteristics. These critics argue that structural considerations such as discrimination are more salient for African Americans than for other groups. Discriminatory practices such as redlining, restrictive zoning, and outright violence and intimidation directed against pioneering families enforce a hierarchy of space (Alba and Logan 1993). To the extent that this view is correct, studies of spatial attainment serve to document the lack of efficacy of key resources such as education and income in minority spatial attainment. This raises the possibility that the differences reflect the structural constraints imposed on the group by discriminatory practices.

Massey and Denton (1985) provide an example. They report that spatial attainment models do not predict much spatial assimilation for African Americans. Accordingly, they argued that spatial attainment as an outcome and spatial assimilation as a process, are fundamentally different for African Americans than for other minorities. The authors explain:

> phenotype and white prejudice alters the process of spatial assimilation to the degree that segregation of blacks is distinctly different, not only from segregation of white immigrant groups, but also from segregation of other nonwhite groups such as Hispanics and Asians (Massey and Denton 1985)

Alba and Logan (1991) introduced the phrase *place stratification* to describe the condition when African Americans are unable to convert their individual-level characteristics (e.g., education) into neighborhood-level outcomes (e.g., higher percent white neighborhoods). This condition exists, for example, when high SES blacks are unable to move into whiter neighborhoods. Later, this type of place stratification was relabeled as the “strong” version of place stratification. it was contrasted with a “weak” version of place stratification, wherein African Americans *can* convert their individual-level characteristics into neighborhood-level outcomes, but at much lower rates than other groups (Adleman 2005; Alba and Logan 1993; Crowder 2001).

Place stratification theory is offered as a “supplement” to spatial assimilation theory in an effort to assess group-level, structural differences (Alba and Logan 1993).

Place stratification theory claims to add the cost of group membership to the individual-level independent variables normally included in spatial attainment models.

Strictly speaking, this is a refinement in interpreting spatial attainment models, not a change in the methodology. Spatial attainment regressions directly compare the "cost" of group membership by modeling group differences in the attainment process. Group differences in education and income are considered when assessing effects on residential outcome. Since race is included as an independent variable in the model, the effect of race is directly assessed. The difference in attainment outcome by group is equivalent to what place stratification terminology labels the cost of group membership. Significantly, the notion of hierarchy of place is compatible with the tenets of spatial assimilation theory. The foundation of spatial assimilation theory is that social distance is expressed geographically. Therefore, great social distance exists between advantaged and disadvantaged groups. That social distance is then expressed geographically in the spatial ecology of the city.

While spatial assimilation theory can address the spatial attainment of minority groups and social distance, available data and methods cannot determine the mechanisms that maintain the hierarchy of place. It is true that discrimination, zoning restrictions, and violence against pioneer households are not variables in the individual-level model of spatial attainment. But it is equally true that place stratification models do not include direct measures of these variables either. Place stratification claims to "subsume" discriminatory practices (Alba and Logan 1993: p. 1391), but it never explicitly includes them in the model. The residual difference between groups is merely reinterpreted and attributed to the impact of discrimination. Thus, place stratification predictions regarding discrimination are not assessed directly. The impact of discrimination is assessed indirectly based on the inability of group differences in social characteristics to explain group differences in residential outcomes. Place stratification theory thus stresses a particular interpretation of group differences in spatial attainments.

A more direct test of the hierarchy of place might include the development of multi-level models which specify an individual-level model of spatial attainment that varies across space, time, and group. Thus, for example, one might investigate whether relative minority size, zoning, percent in poverty, or other ecological factors influence spatial attainment effects.

Place stratification itself is the object of much criticism. Tolnay (2003) objects to the expansion of the concept into “strong” and “weak” versions. He points out that, as it is currently presented, there is no way to falsify a place stratification model. If the African American group has no spatial attainment, then it is classified as strong place stratification. If, however, blacks translate higher SES into better residential location—but still not as good as non-Hispanic whites—it is classified as weak place stratification. The only condition where place stratification is not a factor is when blacks translate their individual-level characteristics into residential location at the same rate as whites.

Another criticism of place stratification theory is that it relies on indirect evidence. It first observes a difference between the groups, and then it attributes this difference to the impact of a mechanism—discrimination—that is not included among the model variables. As a result, place stratification theory is limited to offering an interpretation of a statistical residual difference.

I hold that the lack of reliable trend data in patterns of spatial attainment has led place stratification researchers to misidentify emerging spatial attainment for African Americans. All available evidence indicates that blacks had little spatial assimilation and negligible spatial attainment in US cities before the Fair Housing Act of 1968. This condition, of no spatial attainment, equates to strong place stratification. After passage of the Act, there was at least a nominal decline in the level of housing discrimination. With this, spatial assimilation and spatial attainment became a possibility of African Americans. However, many structural barriers remained in the form of direct and indirect institutional discrimination (Massey and Denton 1993b). This allowed for the highest SES blacks to achieve a small degree of spatial attainment, but still less than other groups not hampered by such institutional barriers. I argue that this condition accounts for the finding of “weak” place stratification in many studies.

Spatial assimilation as a theory and spatial attainment as an outcome continue to provide a viable basis for understanding group residential processes in urban areas. The challenges to their assumptions should be heeded, but they do not undercut the potential value of the perspective. Place stratification does not supplant spatial assimilation; it offers no new direct evidence regarding spatial attainment as an outcome. Research will continue in these areas with access to new datasets incorporating micro- and macro-level data. However, there are certain issues which must be addressed for development to continue apace.

2.2. Critical Issues Facing Spatial Attainment Research

The literature regarding spatial attainment has accomplished much. It has identified group differences in various spatial attainment outcomes; cataloged many ecological factors which influence spatial attainment; incorporated the use of a range of data sources; and sought to overcome weaknesses found in available data. However, I note here that there are two related issues which are critical to the continued development of the area: namely, the need for accurate assessment of trends in spatial attainment; and the consideration of spatial scale when assessing spatial attainment.

One critical issue facing spatial attainment research is the need for accurate assessment of trends in spatial attainment. Lack of trend data has led to poor theoretical understanding of spatial attainment process. Early research attempting to address this question relied on the method of aggregate regression. Below I show that this method is flawed and yields incorrect estimates of spatial attainment effects. Accordingly, it should be discontinued.

The main obstacle preventing analysis of trends in spatial attainment is the availability of useable data at relevant (i.e., small) spatial scales. This type of research relies on both micro- and macro-level data. Large-scale datasets like the Census usually release macro-level data aggregated to the area (block group, tract, etc.), but they do not release micro-level data for small areas. Alternatively, small-scale surveys release micro-level information but rarely release residential location information to researchers in order to protect the confidentiality of respondents. I suggest an approach to overcome this problem by adapting readily available Census tables to create datasets that can be used to estimate simple spatial attainment models for residential outcomes measured at small spatial scales. I apply this method to Houston, Texas using Census data from 1970, 1980, 1990, and 2000. With this approach, I chart changes in spatial attainment patterns over time for non-Hispanic white, black, and Hispanic groups.

Another important issue which merits attention is the role of scale in spatial attainment. Scale, as in demographic and geographic size, must be considered when measuring spatial outcomes. This issue is well-documented in macro-level analyses of residential segregation (Cowgill and Cowgil 1958; Roof and Van Valey 1972; Taeuber and Taueber 1965; Van Valey and Roof 1976b). The consensus in the segregation literature for fifty years has been, the smaller the spatial scale, the greater the macro-level segregation scores. However, a disjunction appears between the literature assessing levels and trends in segregation and the literature assessing spatial attainment.
In the literature on spatial attainment, scant attention is given to spatial scale. This situation is unfortunate, because the spatial attainment literature often relies on residential outcomes measured at very large scales. Residential outcomes have been measure based on urban/suburban distinction in New York City (Alba and Logan 1993) or large sub-borough areas of New York City (Freeman 2002).

The reliance on large spatial scales is due to constraints on data, *not for conceptual reasons*. Spatial attainment research needs data which geographically locates individuals. But Census datasets and other large surveys suppress detailed geographic information in order to protect the confidentiality of their respondents. Since there have been few options in terms of scale, the weaknesses of relying on large scales have not been explored adequately in the spatial attainment literature. I directly compare spatial attainment models across various scales in the analysis section of my dissertation. I find that scale matters in spatial attainment, as anticipated by research in residential segregation: using smaller spatial scales reveals stronger patterns of spatial attainment.

In the sections that follow I present analyses which directly address the critical issues outlined above. The remainder of my dissertation is split into four sections: 1) a detailed discussion of the inappropriateness of aggregate regression in spatial attainment research; 2) an assessment of aggregate spatial attainment trends over time; 3) an assessment of micro-level spatial attainment trends over time; and 4) a comparison of spatial attainment across different geographic units. I finish my discussion of issues in spatial attainment research with a critical evaluation of aggregate regression models as used in spatial attainment research.

3. AGGREGATE ANALYSES OF EXPOSURE AND CONTACT

In this section I hope to demonstrate previously unrecognized problems with using aggregate regression to estimate spatial attainment models. I begin by reviewing a previous article in the literature which uses the method. I show that the method leads to incorrect estimates of spatial attainment effects and flawed substantive conclusions. Next, I will prepare analyses comparing the aggregate regression method to an individual-level regression model. To accomplish this, I replicate the aggregate regression research of Massey and Denton (1985); followed by a parallel analysis that disaggregates their data to estimate “true” individual-level attainment models. I then compare results obtained using the two methods.

Aggregate regression, a methodological practice sometimes termed “ecological regression,” involves a regression analysis wherein data for aggregate units are used to estimate effects for individuals. For example, Robinson (1950) reports the effect of percent black in the national region on illiteracy. By using just nine national subregions, the correlation coefficient between percent black and illiteracy is .95. While this informs us about variation in illiteracy rates across different regions, it does not tell us about the illiteracy of African Americans. When Robinson uses individual-level data comparing illiteracy between race groups, the correlation drops to .20. The fallacy in this example is the use of a group characteristic as an independent variable (percent black in the subregion) to predict an individual-level variable (the illiteracy of persons).

Surprisingly, aggregate regression is frequently used to estimate spatial attainment models (Massey and Denton 1985; Massey and Mullan 1984; Massey and Gross 1991; Mullan and Massey 1984). Spatial attainment models hope to answer the question, “How do individual characteristics translate into residential outcomes?” Therefore, the appropriate level of analysis is at the individual-level. For example, we are interested in the effect of individual educational attainment, *not* the effect of mean education of tract. We are correct to assume that an individual’s educational attainment will be correlated with percent white neighborhood. But we run afoul of the fallacy if we rely on a tract’s mean education to infer the individual education of a resident within the tract. Unfortunately for those seeking to understand trends in spatial attainment, the most often cited works rely on aggregate regressions (Massey and Denton 1985; Massey and Gross 1991).

Researchers have turned to aggregate regression to estimate spatial attainment models for two reasons: lack of appropriate individual-level data and the mistaken assumption that independent variables—predictors—can be measured at the aggregate level in the same manner as the dependent variable. Aggregate-level independent variables are used because independent variables are not reported at the individual level. For example, since Massey and Denton (1985) do not have access to the educational attainment of African American individuals in Census tracts, they employ the variable of average African American education in the tract.

Perhaps the use of aggregate regressions arises from confusion regarding the appropriate level of measurement for dependent and independent variables common in spatial attainment models. The dependent variables in these models are usually measured at the aggregate level (e.g., percent white, median income, percent poverty), for the Census tract. However, *the outcome applies to individuals and the process of spatial attainment is an individual process*. Therefore, the predictors should be measured at the individual level.

Although the problems associated with aggregate regression are well-known, Massey and Denton (1985) claim that the problems do not apply to spatial attainment models. They review results of spatial attainment analyses obtained using both individual and aggregate data and conclude that the results are similar between the two methods. They further conclude that aggregate regressions are a viable method for estimating spatial attainment models.

Figure 1 graphs predictions from spatial attainment models estimated from aggregate regressions reported by Massey and Mullan (1984: 855 Figure 4, Panel 1). The figure depicts the relationship of education and the probability of Anglo contact in Los Angeles in 1970. The figure indicates a strong effect of education for both Hispanics and African Americans. For each group, the predicted proportion Anglo (non-Hispanic white) of the group's tract increases as education increases. The curve for Hispanics is higher than that for blacks, indicating Hispanics translate education into contact with whites at a higher rate than blacks. The gap between Hispanics and African Americans is especially significant at the middle range of the educational variable. The dotted lines in the figure highlight the difference. Massey and Mullan (1984) offer the interpretation:

> While a group of black high school graduates could expect to reside in a tract that was 27% Anglo, a similarly educated Hispanic population could expect to reside in a tract that was 91% Anglo (854).

The results they present also indicate that Hispanics with high school diplomas and African Americans with some college education both live in 90% white neighborhoods.

Unfortunately, that conclusion is not supported by any of the literature on segregation in US cities. The conclusion is particularly untenable for Los Angeles in 1970. Massey and Denton (1993a: 48) report that the isolation index of the city at that time was 74—interpreted as the probability that a black resident’s randomly selected neighbor would also be black. Los Angeles also had the third highest dissimilarity score (91) of the thirty major metropolitan areas included in their study (1993: 64).

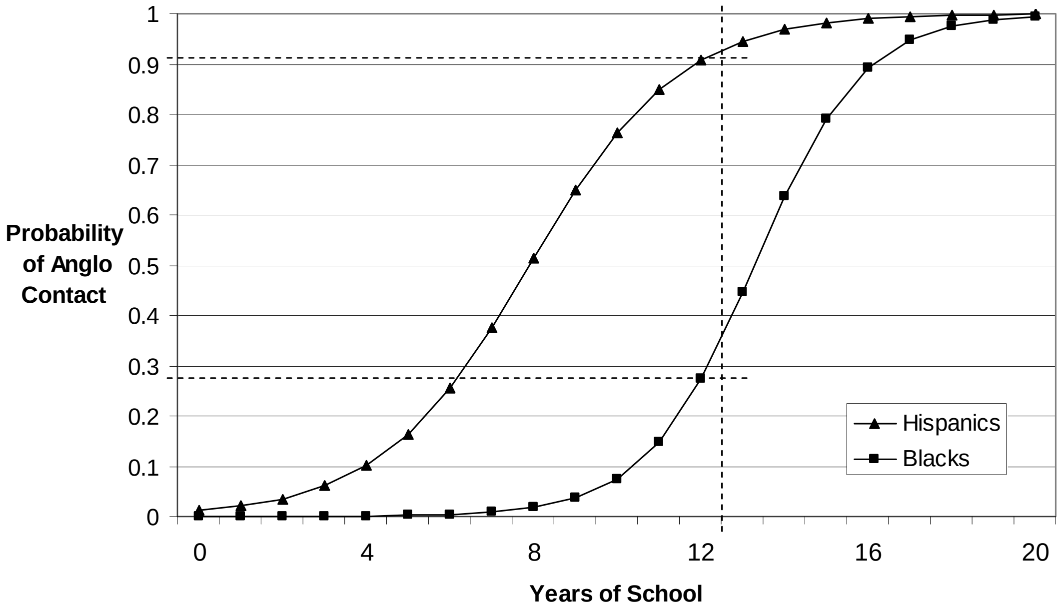

Figure 1. Estimated Relationship between Education and Probability of Anglo Contact in Los Angeles, 1970 (from Massey and Mullan, 1984:855).

Dissimilarity indicates the percentage of African Americans that would need to move to integrated neighborhoods for each neighborhood in the city to have an equal racial composition as the entire city. Both indices indicate very little chance of blacks living in 90% white neighborhoods in 1970's Los Angeles, no matter the level of their education.

In fact, Massey and Mullan's (1984) reported results are severely biased due to their inappropriate use of aggregate data when estimating spatial attainment effects. Specifically, the individual-level spatial attainment effect of education on contact with whites is greatly exaggerated because it is estimated from aggregate data. Langbein and

Lichtman (1978: 50-60) demonstrate that individual-level effects estimated from aggregate data will be biased whenever a factor that affects the dependent variable also covaries with the independent variable. In this case, minority status is the key factor that affects both the dependent variable of probability of Anglo contact and independent variable of mean years of schooling completed, causing them to vary together.

Technically, it is possible, although not likely, that aggregate regressions can yield unbiased estimates of spatial attainment effects. To test this possibility, a direct comparison is needed between individual and aggregate models. If the results are comparable, then the relevant individual-level effects can be reliably estimated from the regressions based on readily available aggregate data.

Massey and Denton (1985) claim to document such a finding. They compare results from aggregate models based on 1970 tract-level data with results from individual-level models based on the 1970 Neighborhood Characteristics File. One reason this research was potentially important was because the Neighborhood Characteristics File was not available after 1970, ostensibly leaving researchers to work only with the aggregate approach. Massey and Denton conducted their analysis by assembling comparable variables from both individual- and aggregate-level data.

Independent variables included race, education, occupation, and income. Massey and Denton (1985) reported that, generally, effects reported at the macro level were found at the micro level, except for the effect of African American income on contact with Hispanics. Aside from this exception, coefficients of neighborhood outcomes are significant and in the expected directions at both micro and macro levels. Massey and

Denton interpreted this as support for the use of the aggregate method in future research. Their argument is that assimilation is a group process, and therefore, there is no fallacy to estimating ecological (i.e., aggregate) models of spatial assimilation.

However, a close inspection reveals that the aggregate results they report grossly exaggerate spatial attainment effects. Massey and Denton (1985: 100) report results for the effects of individual characteristics on the probability of African American contact with Anglos. The effects (i.e., slopes) estimated from aggregate regressions are *an order of magnitude larger* than the effects estimated from individual-level regressions. For example their work, reproduced below in Table 1, reports that the unstandardized regression coefficient of income at the individual-level is 0.017, yet the aggregate regression yields a coefficient of 0.174. All of the effects in the lower panel of this table (the panel highlighting the probability of African American contact with Anglos) are similarly exaggerated. The authors attribute this difference between models to the greater “explanatory power of the macro equations.”

Table 1. Summary of Effects in Path Models of Hispanic and Black Spatial Assimilation: Individuals in U.S. and Census Tracts in Los Angeles SMSA, 1970.

	Individual Level			Aggregate Level		
	Direct	Indirect	Total	Direct	Indirect	Total
Black probability of contact with Anglos and:						
Probability of contact with Hispanic	-.025	.000	-.025	-.065	.000	-.065
Education	.017	.016	.033	.174	.221	.395
Occupation	.024	.004	.028	.201	.036	.237
Income	.022	.000	.022	.201	.008	.209

This table is excerpted from a larger table published on page 100 of Massey and Denton (1985).

I illustrate the inappropriateness of using aggregate regression in spatial attainment by showing the unexamined implications of the previously published results of Massey and Denton (1985). Their table of estimated coefficients reports the coefficient for education on contact with whites for both aggregate- and individual-level regressions. I use that table to illustrate the predicted effect of education on probability of contact with whites implied by the coefficients. I perform a logit transformation on the predictions (p) where:

$$LOGIT(P) = \log(p/(1-p))$$

Then, I plot the predictions implied by both the aggregate- and the individual-level regressions for comparison.

In fact, the greater "explanatory power" alluded to by the authors is a statistical artifact which comes from estimating individual-level spatial attainment using aggregate data. I illustrate this using the results from Massey and Denton's (1985) own analysis. The line graph of Figure 2 compares predictions from the aggregate and individual

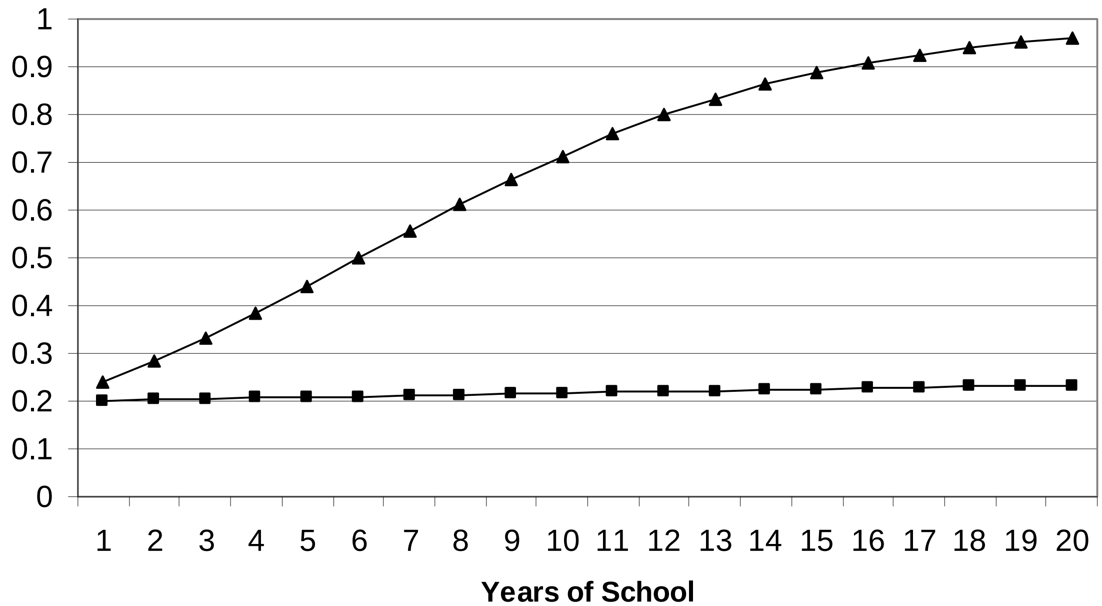

Figure 2. Predictions of African American Contact with Whites Based on Aggregate and Individual-Level Spatial Attainment Models: US, 1970 (from Massey and Denton 1985: 98)*.

*Descriptive statistics are not presented in the published article. However, a point of reference is needed for the intercept. I chose .25 as the average probability of percent white contact for those in the lowest education category. This choice is based on Census data for Houston in 1970.

regressions for Blacks using US data from 1970 presented by Massey and Denton (1985: 98). The aggregate-level prediction has a steep upward slope across the years of school completed implying large differences in spatial attainment outcomes by education. The individual-level prediction is flat with only a very slight shift upward toward the high end of the educational range. The two models predict very different neighborhood outcomes for African Americans with high levels of education. Indeed, it is fair to say that their predictions are almost completely different. The differences in predicted contact are huge at most levels of education. The relevance of this difference for substantive conclusions about spatial assimilation are dramatic.

The line graph of Figure 3 compares predictions from aggregate and individual regressions for Hispanics using US data from 1970 presented by Massey and Denton (1985: 98). The same pattern is apparent: the slope is steep for the aggregate regression predictions; and the relatively flat for the individual-level regression predictions. In this case, the aggregate predictions come close to the upward bound of proximity to whites by the 14th year of school completed.

Figures 2 and 3 indicate that results from aggregate-level analyses suggest very strong spatial assimilation effects, i.e., large differences in proximity with whites between those with high and low levels of education. The individual-level models, on the other hand, reveal much weaker spatial assimilation effects. The spatial attainment effect of education on Hispanic proximity to whites is more pronounced than for African Americans. But still it is much smaller than that suggested by the predictions from the aggregate regressions.

The explicit goal of the Massey and Denton (1985) article is to compare individual and aggregate-level regression models. Aggregation bias is present in the data and the authors acknowledge its existence with regard to standardized regression coefficients. Yet, the authors conclude that the bias is negligible and does not distort substantive conclusions. A closer examination of the results presented in their own tables reveal that aggregate regression leads to a very misleading view of predicted spatial attainment outcomes for African American and Hispanic individuals. As shown in the

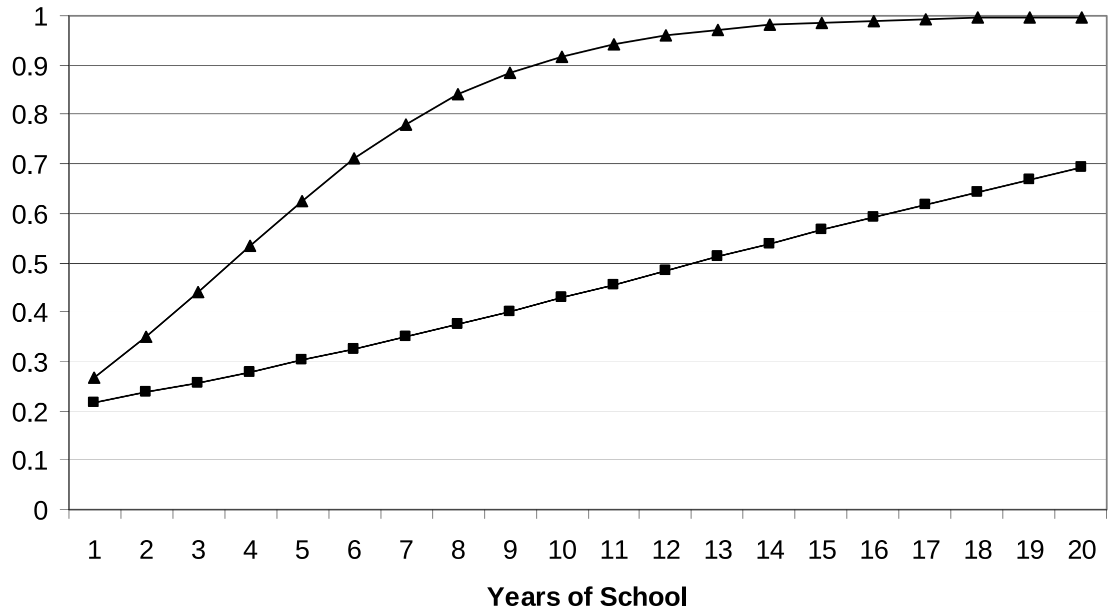

Figure 3. Predictions of Hispanic Contact with Whites Based on Aggregate and Individual-Level Spatial Attainment Models: US, 1970 (from Massey and Denton 1985: 98).

*Descriptive statistics are not presented in the published article. However, a point of reference is needed for the intercept. I chose .25 as the average probability of percent white contact for those in the lowest education category. This choice is based on Census data for Houston in 1970.

figures above, the gap between the predictions implied by individual and aggregate-level models is enormous. Contrary to the author's conclusion, the implied predictions based on aggregate regressions are completely misleading.

In addition to the exaggerated effects given by aggregation bias, the authors mischaracterize spatial attainment theory in arguing for the continued use of aggregate regression in assessing spatial attainment. Spatial attainment is an individual-level theory (i.e., as individual-level characteristics change, residential status changes). Massey and Denton (1985) argue that perhaps, in addition to being an individual-level process, spatial attainment takes place on a group level. This theoretical argument may be plausible. However, it does not correct for the bias found in the aggregate regressions. This idea should be investigated using a multi-level modeling framework in which the micro-level spatial attainment model (at level 1) is itself taken as varying with structural conditions included at level 2. I conclude that, for the purposes of spatial attainment, aggregate regressions such as those used by Massey and Denton (1985) are inappropriate for testing the theory they offer.

4. AGGREGATE TRENDS OF EXPOSURE AND CONTACT

In this section I investigate how the patterns of contact, or exposure, between race/ethnic groups have changed in Houston, Texas between 1970 and 2000. My research here extends the literature in three areas: 1) I report changes in minority exposure to whites in Houston at four points in time spanning three decades; 2) I investigate variation in exposure to whites by minority socioeconomic status classification, and; 3) I examine exposure to the highest status whites by minority status classification. The findings I obtain using P* measures reveal that differences in socioeconomic status within the African American group had little impact on black exposure to whites in Houston in 1970. But by 1980, and continuing through 2000, African Americans with higher socioeconomic status had greater exposure to whites.

Segregation measurement theory identifies several dimensions of segregation (Massey and Denton 1988). The two most widely studied dimensions are uneven distribution—typically assessed using the index of dissimilarity (D), and contact/exposure—typically assessed using P* measures. St. John and Clymer (2000) show that even substantial variation between subgroups (such as socioeconomic status groups by race/ethnic group) can lead to small or insignificant changes in dissimilarity (St. John and Clymer 2000). The dissimilarity index serves well in indexing uneven distribution of residential segregation. But it does not serve well in assessing contact. Accordingly, some researchers have turned to another facet of segregation termed *exposure* to examine spatial attainment (Lieberson 1980; St. John and Clymer 2000).

Exposure (P*) is a more appropriate aggregate measure of spatial attainment. The exposure index reports the probability of interracial contact between two groups based on the group proportions in the city. It shares the advantages of aggregate analysis with dissimilarity—the necessary data are readily available; and like dissimilarity, its numeric calculation allows for direct comparison between cities or for the same city at different points in time. However unlike dissimilarity, exposure is suitable for assessing spatial attainment because of its meaningful individual-level application as a probability of contact between group members.

Although previous research has studied minority exposure to whites over time (Farley and Frey 1994; Lieberson 1980; Massey and Denton 1987; Massey and Denton 1993b), no studies to my knowledge have investigated trends in exposure by socioeconomic status groups in a single city over an extended period of time. Lieberson (1980) charts trends in black exposure to whites in American cities through 1970. Massey and Denton (1993b) compare minority isolation scores in 30 U.S. cities in 1930 and 1970. St. John and Clymer (2000) note increases in spatial attainment for higher educated African Americans in 1990. They do not, however, report changes in spatial attainment over time.

Spatial attainment research often chooses minority group contact with white as an outcome variable. However, there is variation within the white neighborhoods themselves. The goal of spatial attainment may not be to live in white neighborhoods; it may be to live in high status white neighborhoods. If that is the case, then it is informative to assess minority group contact with the highest status whites. St. John and

Clymer (St. John and Clymer 2000) compare contact between whites and blacks of equal status, but not between statuses, and not focusing on contact with highest status whites. Again, I know of no other study that directly assesses minority group contact with high status whites.

Based on segregation trends noted in previous research, I expect to find similar patterns in Houston. Farley and Frey (1994) review trends in segregation between 1980 and 1990 in U.S. cities. They report that segregation for African Americans is slowly declining. However, segregation among Hispanics and Asians increased over the same time period due to increased immigration into American cities. Glaeser and Vigdor (2003) and Logan (2003) confirm this trend using 2000 data—black/white segregation is slowly decreasing, Hispanic and Asian segregation is increasing.

Drawing on these conclusions from previous research, I anticipate four outcomes. First, in racially segregated 1970 Houston, African Americans will have little exposure to whites and no variation in exposure by black status. Second, as *de jure* segregation is formally proscribed, I expect a spatial attainment pattern to emerge for the African American group by 1980—indicated by variation in exposure to whites by black status. Third, black exposure to the highest status whites, the most isolated group, should increase over time. And fourth, patterns of Hispanic exposure to whites will be present from 1970 onward, but scores will decline as Hispanic proportion increases and white proportion decreases in the city.

4.1. Data and Methods for Trends in Exposure and Contact

I investigate aggregate-level trends in racial/ethnic contact in Houston, Texas at four points in time between 1970 and 2000. The data for my study come from four U.S. Censuses: 1970 data are from the Fourth Count Summary Tapes, File A; 1980 data are from *Census Tracts*; 1990 data are from Summary Tape File 3; and 2000 data come from Summary File 3. These datasets are widely used in segregation studies and aggregate-level spatial attainment research.

When investigating the same city of multiple points in time, researchers have developed two approaches to defining the areas which constitute the city: the use of constant boundaries based on the last point in time; and the use of the city definition at the time studied. These approaches attempt to account for the changing area of a metropolitan area. For example, the Houston, Texas metropolitan area was composed of 344 Census tracts in 5 counties in 1970, but by 2000 the city had 886 tracts in 8 counties. One approach takes the 886 tracts of 2000 and includes them into the periphery of 1970 Houston. The other approach compares the two as they were defined at each point in time.

The potential advantage of using constant boundaries is that changes in segregation cannot be attributed to the simple addition of areas over time (Logan, Stults, and Farley 2004). One disadvantage of using constant boundaries is that it includes areas at earlier points in time which were not well populated. This may introduce a downward bias in segregation scores because the large geographic size of sparsely populated peripheral areas can create the appearance of greater integration than would be evident if smaller tracts were used. In addition to population, another significant disadvantage is that the peripheral area's residential patterns may be little related to the patterns of the metropolitan center. A final shortcoming is that, at the time of the study, these peripheral areas may not be socially or economically connected to the metropolitan center.

I do not use constant boundaries; instead I use the approach which relies on the metropolitan area at the time studied. This approach allows me to investigate interracial contact within an area as it is socially defined at the time. It also protects me from the bias of including peripheral areas around Houston which were very lightly populated early in the study. Also, the segregation pattern of metropolitan Houston, Texas is quite different from the segregation patterns of the largely homogenous surrounding areas.

I focus my study on contact between three race/ethnic groups: non-Hispanic white; African American; and Hispanic. Although the category 'Hispanic' is treated as an ethnicity by the Census, I include it in my study as a distinct group comparable to white and African American. The groups are exclusive—no individual is counted in both categories.

I also report contact between groups by education category. Drawing on Census summary file tabulations for education, I create 5 education categories based on completed years of school: 1) 0-8 years; 2) 9-11 years; 3) 12 years, or high school level of education; 4) 13-15 years, or some college; and 5) 16+ years, or college degree. These groupings capture major divisions in educational attainment and can be maintained over all decades.

The main statistical technique I use in this section is the P* exposure index. The formula for the index is:

$$ {}_{x}P_{y} = \sum \left[\left(\frac{x_i}{X} \right) \bullet \left(\frac{y_i}{t_i} \right) \right] $$

Where *X* is the total number of members of group X in the whole city, x_i and y_i are the number of members of group x and y respectively in the i^{th} tract, and t_i is the total population of the i^{th} tract.

The index P* indicates the probability that a randomly selected resident (x_i) will be of the a different race as another randomly selected person (y_i) from the same tract (t_i) (Jaret 1995). The exposure index is computed as a probability and has a range from 0.00 to 1.00. In this study I convert it to a percentage with a range from 0 to 100 for ease of interpretation.

Massey & Denton (1987) give the familiar interpretation of P*:

> Exposure indices measure the extent to which minority and majority members must physically confront one another by virtue of sharing a common tract of residence. The degree of minority exposure may be conceptualized as the likelihood that minority and majority members share a common neighborhood. (p.806)

It is also possible to use P* as a measure of a group's isolation by computing the extent to which the members of a group have contact with their own group.

The formula for the P* isolation index is:

$$ {}_{x}P_{x} = \sum \left[\left(\frac{x_i}{X} \right) \bullet \left(\frac{x_i}{t_i} \right) \right] $$

Where X is the total number of members of group X in the whole city, x_i is the number of members of group x in the i^{th} tract, and t_i is the population of the i^{th} tract.

Interpretations of P* indexes should be compared to the proportion of each group in the total population (Jaret 1995, p.344). This reveals a limitation in the use of the P* indices: it is a function of total group proportions in the city. This means that, all else equal, as the underlying demography of the city changes, P* changes. Charles (2003) explains:

> Isolation is generally low for small groups but is expected to rise with increasing group size even if the group's level of segregation remains constant. Moreover, the larger the relative size of an out-group's population, the greater exposure to that group is likely to be. Both exposure and isolation are influenced by group settlement patterns (Charles 2003, p.172).

To account for demographic affects on P*, I calculate a percent of expected exposure score. This score is simply the ratio of the observed P* score to the score expected under even distribution. For example, if whites make up 58% of the population and African Americans have a P* score of 29, then their percent of expected exposure score is 0.50. The P* score of 29 is quite low, indicating that within the neighborhood of the average African American there is less than a 30% chance of randomly selecting someone who is white. But relative to the fact that only 58% of the population is white, an exposure score of 29 is about half of what would be expected.

This methodological approach is consistent with the aim of the study—which is to examine patterns of spatial attainment among blacks in Houston over the thirty year period of the study. By investigating the exposure and isolation indices of blacks and whites, I can determine patterns of spatial assimilation.

4.2. Results for Contact and Exposure

Table 2 shows the descriptive statistics for Houston, Texas between 1970 and 2000. Population and group counts are necessary to calculate P*; for ease of interpretation across time periods I also report race/ethnic group scores as a percent of total population. Since I investigate exposure by education level, this table shows the education averages by group over time. Finally, this table shows the changing number of Census tracts which constitute the statistical area of Houston at each point in time.

The top row of Table 2 shows that the population of persons over age 25 in Houston tripled over the study period from one million to 3.1 million. The next three rows of the table show changes in the racial composition of Houston during the four time points. Houston underwent significant changes in its racial makeup during this time. Percent white in the city decreased between each census at about the same rate that percent Hispanic increased. The African American percentage of the city remained relatively constant at 17%. The average education of Houstonians increased in 1970 and 1980, and leveled off between 1990 and 2000 at an average level above high school. When broken out by race/ethnic group, all groups increased their education at each point in the study period. The white group consistently has the highest

Table 2. Descriptive Statistics for Population, Race/Ethnicity, Education, and Number of Tracts in Houston, Texas 1970 to 2000.

	1970[a]	**1980**[b]	**1990**[c]	**2000**[d]
Population Over 25	1,022,693	1,728,180	2,273,043	3,181,079
Percent White[e]	74.8	69.6	60.9	47.7
Percent Black	17.0	17.4	19.1	16.6
Percent Hispanic[f]	8.2	11.6	17.2	20.6
Average Education[g]				
Total	1.9	2.2	2.5	2.5
White	2.0	2.3	2.7	2.8
Black	1.6	1.7	2.1	2.3
Hispanic	1.7	1.7	2.0	2.2
Tracts[h]	344	693	819	886

[a] Census of Population and Housing 1970, Fourth Count Summary Tapes, File A (non-Hispanic White, Black, and Hispanic only).

[b] Census of Population and Housing 1980, *Census Tracts* . (non-Hispanic White, Black, Asian, Native Amer., Other, Hispanic)

[c] Census of Population and Housing 1990, Summary Tape File 3. (non-Hispanic White, Black, Asian, Native Amer., Other, Hispanic)

[d] Census of Population and Housing 2000, Summary File 3. (non-Hispanic White, Black, Asian, Hawaiian/Pac. Islander, Native Am., Other, Two or More, and Hispanic)

[e] White includes only non-Hispanic White.

[f] Hispanic includes Hispanic of any race.

[g] Average education is computed across five categories based on years of educational attainment: 0 = less than 9 years; 1 = 9 to 11 years; 2 = 12 years (HS diploma); 3 = 13 to 15 years; 4 = 16+ years.

[h] Tracts are the number of tracts in the counties which constitute the statistical area of Houston.

average education, the Hispanic group the lowest, with the African American group in between.

Table 3 shows the change in exposure between race/ethnic groups in Houston between 1970 and 2000. Because of the significant changes in the demography of the city over the study period, I include percent of expected contact between groups. The expected contact is simply the group's percent of the total racial composition of the city. The column labeled "percent of expected contact" is the ratio of observed to expected contact.

Table 3 shows that African American exposure to whites changed little over the study period, varying from 27.11 in 1970 to 32.15 in 1990, but falling to 29.17 in 2000. Using the common interpretation of exposure, these scores indicate that within an average African American individual's neighborhood in 1970, there was a 27.11 percent chance that a randomly selected person would be African American; in 2000, an 29.17 percent chance. These scores reveal minimal change in African American exposure to whites over the 30 year study period.

Table 3. Probability of Contact (P*) and Percent of Expected Contact Between White, Black, and Hispanic Groups in Houston, Texas 1970 to 2000.

	Group's Probability of Contact with:			Percent of Expected Contact with:		
Year and Group	**White**	**Black**	**Hispanic**	**White**	**Black**	**Hispanic**
1970						
Expected [a]	*74.79*	*16.98*	*8.23*	*100%*	*100%*	*100%*
White	86.67	6.15	7.18	116%	36%	87%
Black	27.11	66.45	6.45	36%	391%	78%
Hispanic	65.26	13.30	21.43	87%	78%	260%
1980						
Expected [a]	*72.10*	*16.38*	*11.52*	*100%*	*100%*	*100%*
White	84.16	6.33	9.51	117%	39%	83%
Black	27.86	63.70	8.43	39%	389%	73%
Hispanic	59.54	11.99	28.46	83%	73%	247%
1990						
Expected [a]	*65.81*	*17.07*	*17.12*	*100%*	*100%*	*100%*
White	78.01	8.34	13.65	119%	49%	80%
Black	32.15	53.79	14.06	49%	315%	82%
Hispanic	52.50	14.02	33.47	80%	82%	196%
2000						
Expected [a]	*57.56*	*17.12*	*25.32*	*100%*	*100%*	*100%*
White	73.03	8.68	18.29	127%	51%	72%
Black	29.17	47.94	22.88	51%	280%	90%
Hispanic	41.59	15.48	42.93	72%	90%	170%

[a] Expected value is equal to the average racial composition of the entire city.
Assuming even distribution (i.e., no segregation) all groups would have this score

However, there is a considerable increase in the percent of expected column for African American exposure to white—from 36% to 51%. This does not change the interpretation of exposure given above. But it does give insight into the role of demographic change in more careful interpretations of the exposure indices. As discussed above, the exposure index of African American to white doesn't change much between 1970 and 2000; neither does the percent African American in the city—which stays constant around 17 percent. It is the percent white in the city that drops from 74.79 to 57.56. The effect of this demographic change is that in 1970, an exposure index of 27.11 was only about one third of the exposure expected (36%). But in 2000, an exposure index of 29.17 is more than half (51%) of the expected exposure of black to white. The raw exposure indices, thus, do not reveal an important movement towards integration.

The exposure of Hispanic to whites decreased each year from 65.26 in 1970 to 41.59 in 2000. The interpretation of contact here is that within an average Hispanic group member's neighborhood, the percent chance that a randomly selected person would be white was 65.26 in 1970 and 41.59 in 2000. This is mirrored in the percent of expected contact score which shows a steady decline from 87% to 72%.

Table 3 also presents contact between minority groups in Houston over the study period. The percent chance of randomly selecting a Hispanic resident within the average African American's neighborhood increased from 6.45 to 22.88, an increase greater than the percentage point increase of Hispanics in the city. This is reflected in the percent of expected contact between African American and Hispanic groups, which approaches unity at 90 by the year 2000.

Moving to isolation, or a group's exposure to itself, white isolation apparently decreased between 1970 and 2000, falling from 86.67 to 73.03. However, the decrease in percent white (from 74.79 to 57.56) outpaced the decrease in isolation—yielding an increase in the percent of expected isolation score of 116 to 127 over the study period. Again, a percent of expected score above one indicates more contact than would be expected based on the racial composition of the city. In this case, whites have more contact with other whites in their neighborhoods than would be expected based on the percent white of the city in 2000, even though the isolation index has decreased somewhat since 1970.

African American isolation decreases each year over the study period from 66.45 in 1970 to 47.94 in 2000. That is, the chance that a randomly selected person in the average African American's neighborhood being another African American was 67% in 1970, but 48% in 2000. The rate of decrease in isolation among African American Houstonians is paralleled in the decrease in percent of expected isolation for the group. Just as African American isolation decreased by 28% (from 66.45 to 47.45), so did scores for percent of expected isolation (from 391 to 280).

Hispanic isolation is more complicated. The isolation of Hispanics increases steadily over the study period from 21.43 to 42.93—interpreted as the percentage chance that a randomly selected person from a Hispanic neighborhood will also be Hispanic. However, even though Hispanics experience greater isolation in their Houston neighborhoods, their level of expected isolation has decreased. Hispanics live in significantly more Hispanic neighborhoods in 2000 than in 1970; but in relation to the race/ethnic composition of the city, Hispanic residents live in neighborhoods which are closer to the Hispanic percentage found in the city. This is shown by the percent of expected isolation, which is much closer to one hundred in 2000 (170) than it was in 1970 (260).

My next analysis focuses on group contact with non-Hispanic whites by socioeconomic status, as indicated by education. Table 4 shows changes in exposure between race/ethnic groups in Houston between 1970 and 2000 by educational category. Similar to the previous table, Table 4 has two columnar panels: one for exposure to white scores; a second for percent of expected contact based on the race/ethnic composition of the city. Also as before, the left-hand side of the table is grouped by decade. But here, each education group is broken out by five categories. To interpret these scores in terms of spatial attainment, I am interested in the range of scores from lowest education to highest. A pattern suggestive of spatial attainment by socioeconomic status is revealed when there is an increase in exposure to whites as education increases.

Table 4. Probability of Residential Contact (P*) and Percent of Expected Contact Between Race/Ethnic Groups and White Group.

	Contact with White Group:			Percent of Expected Contact with White Group:		
	White	**Black**	**Hispanic**	**White**	**Black**	**Hispanic**
1970						
Expected [a]	*74.79*	*74.79*	*74.79*	*100%*	*100%*	*100%*
0-8 years	81.22	28.15	57.42	109%	38%	77%
9-11 years	84.86	25.75	65.20	113%	34%	87%
12 years	87.68	26.70	74.59	117%	36%	100%
13-15 years	89.37	26.78	78.84	119%	36%	105%
16+ years	91.30	28.69	83.18	122%	38%	111%
1980						
Expected [a]	*72.10*	*72.10*	*72.10*	*100%*	*100%*	*100%*
0-8 years	77.49	21.76	50.46	107%	30%	70%
9-11 years	80.72	22.02	58.24	112%	31%	81%
12 years	83.92	27.54	65.80	116%	38%	91%
13-15 years	85.59	35.33	72.27	119%	49%	100%
16+ years	87.67	39.91	78.86	122%	55%	109%
1990						
Expected [a]	*65.81*	*65.81*	*65.81*	*100%*	*100%*	*100%*
0-8 years	70.44	22.99	44.31	107%	35%	67%
9-11 years	73.32	24.67	49.14	111%	37%	75%
12 years	76.40	30.16	56.11	116%	46%	85%
13-15 years	78.50	37.75	63.29	119%	57%	96%
16+ years	81.62	43.59	69.76	124%	66%	106%
2000						
Expected [a]	*57.56*	*57.56*	*57.56*	*100%*	*100%*	*100%*
0-8 years	61.04	21.47	33.32	106%	37%	58%
9-11 years	65.86	22.09	37.12	114%	38%	64%
12 years	69.59	25.81	44.02	121%	45%	76%
13-15 years	72.80	31.51	51.80	126%	55%	90%
16+ years	78.07	39.54	60.79	136%	69%	106%

[a] Expected value is equal to percent of whites in the highest education category.
Assuming even distribution (i.e., no segregation) all groups would have this score

Evidence of no spatial attainment would be no variation in a group's exposure to whites by education.

White and Hispanic exposure/isolation to whites in Houston is clearly tied to education. For these groups, lower education categories always have less exposure to whites than do higher education categories. For example, in 1970, among whites with 0-8 years of education, their white-group exposure was 81.22. Whites with 16+ years of education in 1970 had a white-group exposure score of 91.30—a range of 10.08. Spatial attainment for Hispanics in Houston is even more pronounced. In 1970, Hispanics at the lowest education category had a white-group exposure of 57.42. But the highest group had a score of 83.18—a range of 25.76 points based on educational attainment.

The African American group in 1970 stands in stark contrast to the other groups. Variation in education for African Americans in 1970 had no affect whatsoever on exposure to whites. The exposure of African Americans with 0-8 years of education to the white group was 28.15. The exposure for those with 16+ years of education was 28.69—a range of 0.54 points based on education attainment.

Over the course of the study period, three patterns emerge: 1) a pattern of spatial attainment emerges for the African American group; 2) black exposure to white declines across all education categories between 1990 and 2000; and 3) white and Hispanic exposure scores decline each year (1970-2000) within each education category.

The spatial attainment effect of education on exposure to whites for the African American group is established by 1980 and is little changed through 2000. Compared to 1970 where no spatial attainment was present based on education, in 1980 the African

American group experiences variation in exposure to whites by education. The exposure of blacks in the lowest education category drops from its 1970 level of 28.15 to 21.76. At the same time, the exposure of the African American group in the highest education category jumps from 28.69 to 39.91. These changes create a range of 18.15 points—quite different from the range of zero points the decade before. In 1990, African Americans in all education categories experience an increase in white exposure, but in 2000 exposure returns to levels comparable to 1980.

For the African American group, there is also an expansion and subsequent contraction in 1990 and 2000 in the range of exposure between higher and lower education categories. In 1990 the range between the extremes goes to 20.60; but in 2000 it falls back to 18.07. The contraction is due to demographic shifts in the city, indicated by the percent of expected contact with white score. Although the highest educated African American group members experience a drop in exposure from 43.59 to 39.54 between 1990 and 2000, their exposure relative to the number of whites in the city increases from 66% to 69%. Also, the range of exposure between highest and lowest education categories contracts; but the range of the percent of expected contact actually increases very slightly between 1990 and 2000.

White and Hispanic exposure to whites dropped across all education categories over the study period. However, this decline is primarily the result of the demographic changes in Houston between 1970 and 2000. The percent of expected exposure column for the white group reveals that there was little change in the scores or the range of scores between 1970 and 1990. In 2000, the range of percent of expected exposure (in this case, isolation) increased for all but the lowest education category. In terms of percent of expected isolation, whites in all but the lowest education category are more isolated in 2000 than at any other time in the study period. The lowest education categories have generally held steady over between 1970 and 2000.

Percent of expected exposure to whites by the Hispanic group consistently declines for all education categories over the study time. Highest educated Hispanics have decreased their exposure to whites on average, but have maintained their higher than expected level of exposure to whites throughout the study period, falling only slightly from 111% to 106%—still above expected levels of Hispanic exposure to whites. High school educated Hispanics, on the other hand, have seen their percent of expected exposure to whites drop from 100% in 1970 (observed exposure equaled expected) to 76% in 2000.

Finally, it is noteworthy to compare the two minority groups on their percent of expected contact with whites. The African American group consistently has a much lower percent of expected contact score than the Hispanic group. In fact, the highest education category of African Americans evinces a lower percent of expected score than the lowest education category of Hispanics until 2000. And then, African Americans in the highest category of education scored less than Hispanics with a high school degree (69% for African Americans, compared to 76% for Hispanics).

Table 5 reports exposure scores for race/ethnic groups across all levels of education with the highest education category of whites. Similar to the previous table, Table 3.4 has two columnar panels: one for exposure scores; a second for percent of

Table 5. Probability of Residential Contact (P*) and Percent of Expected Contact Between Race/Ethnic Groups and Highest Education Category of Whites.

	Contact with Highest SES White Group:			Percent of Expected Contact with Highest SES White Group:		
	White	**Black**	**Hispanic**	**White**	**Black**	**Hispanic**
1970						
Expected [a]	*12.38*	*12.38*	*12.38*	*100%*	*100%*	*100%*
0-8 years	7.73	2.46	4.69	62%	20%	38%
9-11 years	10.34	2.27	7.03	83%	18%	57%
12 years	14.55	2.36	10.64	118%	19%	86%
13-15 years	20.15	2.63	13.92	163%	21%	112%
16+ years	26.59	3.52	20.82	215%	28%	168%
1980						
Expected [a]	*18.35*	*18.35*	*18.35*	*100%*	*100%*	*100%*
0-8 years	11.64	3.53	7.52	63%	19%	41%
9-11 years	14.09	3.76	9.52	77%	20%	52%
12 years	19.10	5.31	13.00	104%	29%	71%
13-15 years	24.91	8.10	17.91	136%	44%	98%
16+ years	32.64	10.95	25.85	178%	60%	141%
1990						
Expected [a]	*19.68*	*19.68*	*19.68*	*100%*	*100%*	*100%*
0-8 years	12.49	4.47	9.19	63%	23%	47%
9-11 years	14.18	5.28	10.29	72%	27%	52%
12 years	18.59	6.91	12.95	94%	35%	66%
13-15 years	24.26	10.38	17.44	123%	53%	89%
16+ years	34.79	14.66	27.12	177%	75%	138%
2000						
Expected [a]	*20.04*	*20.04*	*20.04*	*100%*	*100%*	*100%*
0-8 years	13.21	4.78	7.60	66%	24%	38%
9-11 years	14.78	5.22	8.59	74%	26%	43%
12 years	18.89	6.55	11.21	94%	33%	56%
13-15 years	25.12	9.46	16.05	125%	47%	80%
16+ years	38.45	15.15	26.47	192%	76%	132%

[a] Expected value is equal to percent of whites in the highest education category.
Assuming even distribution (i.e., no segregation) all groups would have this score

expected contact. The left-hand side of the table is grouped by decade and each education group is broken out by five categories. The column titled “Percent of Expected Contact with Highest SES White” is the ratio observed score to expected score (number of whites in the highest education category) multiplied by 100 to obtain percentages.

For the whites, the salience of education is apparent at all points in time. This is noted by the range of scores across educational categories. In 1970, the lowest education category whites have an average exposure to the highest education category whites of 7.73. The highest educated whites have an exposure score of 26.59—yielding a range of 18.86. Although these scores may seem low compared to the previous table, when the low expected value is taken into account the exposure of the highest educated whites to other high educated whites (isolation) is more than twice of the expected value (215%).

The Hispanic group likewise has a pattern of variation in contact with high education whites by education in 1970. The range of scores from a lowest of 4.69 to a highest of 20.82 is similar to that of whites, at 16.13. Although the scores are never as high as those for the white group, the two highest education categories of Hispanics have higher than expected exposure to the highest education category whites: those with some college score 1.12; and those with a college degree score 1.68.

For the African American group in 1970, there is only a very slight indication of contact varying by education. The highest educated African American group has slightly more probability of exposure to the highest educated white group than did less educated African Americans. But the range for African Americans from lowest to highest education category in 1970 is 1.06; compared to 16.13 for Hispanics and 18.86 for whites. Also, the proportion of expected exposure is never more than 0.28.

Over time, variation in contact by education effect emerges for African Americans. Clear, monotonic increases in exposure by education are apparent for this group. By 1980, the range of exposure from highest to lowest education category blacks jumps to 7.42. Like exposure to all whites, the range of African American exposure to the highest educated whites grows in 1990 to 10.19, and then stays at that level in 2000 at 10.37.

For whites, exposure (isolation) scores increase from 1970 to 1980, and then hold steady through 2000. For percent of expected exposure, the scores decrease between 1970 and 1980, but then hold steady through 2000. The only exception is that in 2000, the highest educated whites have a greater percent of expected isolation with other highest educated whites (1.92) than at any time since 1970, when it was 2.15.

Once again for the Hispanic group, the demographic changes of the city must be taken into account when assessing exposure. Similar to the white group, Hispanic exposure to the highest educated whites increased between 1970 and 1980, and held constant at that higher level through 2000. However, the percent of expected contact scores reveal a continuous decline for every education category of Hispanics over the entire study period. For example, the highest educated Hispanics have a percent of expected contact with the highest educated whites in 1970 of 1.68. In 1980 the percent of expected drops to 1.41, then to 1.38, then to 1.32. The Hispanic groups with high school education and some college also have large declines in exposure to highly educated whites. Only the lowest education category temporarily breaks this trend in 1990 with a percent of expected score of 0.47; but then the score returns to 0.38 in 2000.

4.3. Discussion of Aggregate Trends in Spatial Attainment in Houston

I find evidence of the emergence of a pattern spatial attainment by education for the African American group in Houston, Texas between 1980 and 2000. By contrast, in 1970 this group has no pattern of spatial attainment by educational category. Although a discernible pattern of spatial attainment is apparent by 1980, the African American group, regardless of education, always has less exposure to whites than do Hispanic group members with a high school education. In 2000 in Houston, African Americans have about half of the exposure to whites as would be expected taking into account the racial composition of the city.

My research reveals several interesting results for all three race/ethnic groups over the study period. I find a pattern of spatial attainment is always present for non-Hispanic whites and Hispanics. By 2000, whites are slightly more isolated than in 1990. Hispanics have seen a consistent decline in exposure to whites relative to what would be expected by the demography of the city. Exposure to the highest educated whites follows the same pattern: white exposure to the highest educated whites is slightly higher in 2000 than in 1990; Hispanic percent of expected exposure decreases each year; and African Americans see no change in exposure between 1990 and 2000.

Importantly, I empirically show that relying on exposure scores alone may misstate segregation. The exposure index offers additional meaning when related to the mathematically expected score based on the group composition in the city—a percent of expected score. For example, using the exposure scores alone, black exposure to white is little changed between 1970 and 2000. But to conclude that there is no spatial attainment occurring for African Americans between those years would be wrong. The percent of expected exposure scores reveal that the African American group has increased its relative exposure to whites across each of the four points in time. The individual experience of contact with whites is not much changed. The difference is that the city has become less white. White exposure scores did not drop as precipitously as percent white I the city did. I feel that this technique is especially important in city case studies, like this one, which evaluates a city over time when substantial changes in race/ethnic composition are taking place. Also, percent of expected scores should be used in three group cities such as Houston where shifts of racial composition are volatile.

When grouped by educational category, spatial attainment emerges on the aggregate level for African Americans in 1980 and continues through 2000. My next section uses a micro-level model to estimate just how much of the spatial attainment effect is from race and how much is from education.

5. MICRO-LEVEL TRENDS IN SPATIAL ATTAINMENT

In this section I use individual-level data to estimate simple regression models which reveal trends in the spatial attainment of race/ethnic groups in Houston, Texas between 1970 and 2000. Although the contact scores used in the previous section reveal spatial attainment patterns among the race/ethnic groups by education and offer a insight into individual-level outcomes, they do not provide significance tests or quantify the explained variance of education on percent white.

The main obstacle preventing analysis of trends in spatial attainment is the paucity of useable data at small spatial scales. This type of research relies on both micro- and macro-level data. Large-scale datasets like the Census usually release macro-level data aggregated to the area (block group, tract, etc.), but they do not release micro-level data for small areas. Alternatively, small-scale surveys will release micro-level information but do not release residential location information to researchers to protect the confidentiality of respondents.

Alba and Logan (1992) suggest an approach to overcome this problem by adapting readily available Census tables to create datasets that can be used to estimate simple spatial attainment models. Using this method opens the way for new research in spatial attainment. I apply this method to Houston, Texas using Census data from 1970, 1980, 1990, and 2000. With this approach, I chart changes in spatial attainment patterns over time for non-Hispanic white, black, and Hispanic groups.

Massey and Fischer (1999) use such an approach to highlight the gap of locational attainment between blacks, Hispanics and Asians. Their research note makes use of a cross-tabulation from Census data that combines race and family income by tract. This treatment allows for the authors to calculate minority segregation from and minority contact with all non-Hispanic whites. The method limits the independent variables to race and income category. They report that, although segregation from whites decreases as income rises for all groups, the gap between blacks and other groups increases as income rises. In suburban tracts, blacks in the highest income category are more segregated from whites than all Hispanics and all but the poorest Asians. Blacks also have less contact with whites than Hispanics and Asians in metro areas and central cities, but black and Hispanic contact are similar in suburban tracts. A limitation to their approach is that all whites are grouped together. Class variation in segregation and contact is to be expected—with higher income whites experiencing higher levels of segregation from minorities. Also, the discussion focuses on the differences between the racial and ethnic groups. But the African American experience of segregation is distinct from the other groups. With that in mind, a comparison with other groups is not as enlightening as the historical experience of segregation by African Americans between 1970 and 1990. Unfortunately, a historical perspective is not presented in their study.

The main limitation to this approach is that individual-level control variables cannot normally be included. Thus, complex multivariate models are possible only if the individual characteristics in question (e.g., education and age) are cross-tabulated by each other at the tract level. In addition, it relies entirely on tract- or block-group-level tabulations to estimate models of spatial attainment. This constrains research to relationships between variables identified by the Census. Another limitation to this methodology is the inability to control the sample. For example, prison and military populations cannot be separated out of the sample universe. A final drawback for researchers using this method is the limitation of the measures reported in Census tables. Large, pan-ethnic groupings are available (e.g., Hispanic and Asian) but more nuanced racial and ethnic groups (e.g., Cuban and South Korean) are not presented.

There are other techniques that have been used to estimate spatial attainment models. I summarize the benefits and shortcomings of those other methods in an extended appendix to my dissertation. I conclude that the Alba and Logan method is appropriate and useful for spatial attainment modeling; can be adapted to chart trends over relatively long periods of time.

Guided by my findings using contact scores, I anticipate that results using individual-level data will reveal that whites and Hispanics translate their education into greater contact with whites (i.e., higher percent white neighborhoods) at a steeper rate than will African Americans. Also, as the percent white of the overall city decreases during the study period, white and Hispanic scores on percent white drop over each decade, while African American scores rise through 1990, then level off or drop in 2000. At the outset of the study, white and Hispanic respondents will exhibit a clear pattern of spatial attainment with whites by education. For African Americans however, this pattern will not be apparent in 1970, but in subsequent years, increases in education for this group will result in increases in percent white neighborhoods—among African

Americans, the highest educated will live in the highest percent white neighborhoods. Additionally, using the individual-level data afforded by my method, I estimate predicted values for percent white neighborhood by group over the decades of the study.

5.1. Data and Methods for Micro-Level Trends

To estimate my individual-level spatial attainment models I use data from the 1970, 1980, 1990, and 2000 long form Censuses. Long forms are sent to a sample of the entire population. Twenty percent of the population was sampled in 1970; one-sixth of the population was sampled in subsequent years. These data provide information on race/ethnicity, education, and percent white of the respondent's tract.

In order to assess spatial attainment trends using Census tables, I draw on the summary file tables of education by race for persons aged 25 and above. For this analysis, race and education are the sole independent variables; percent white in the neighborhood is the dependent variable. I use the counts of race by education in this table to create an individual-level dataset which is suitable for simple spatial attainment modeling. This allows education to be used as an independent variable in my spatial attainment regressions for each race/ethnic group.

I use count data of white non-Hispanic persons to create the dependent variable—percent white within each tract. Percent white is not an appropriate dependent variable for regression analysis because percent white is bounded by 0 and 100. So, for my regressions I perform a logit transformation on the percent white variable (pwpop) where:

$$LOGIT(Pwpop) = \ln(pwpop/(100 - pwpop))$$

Extremely low and extremely high logit scores from percent white scores which are less than 0.25 or greater than 99.75, are bottom- and top-coded at those values. Logit scores for neighborhood percent white are well-suited for statistical modeling purposes, but are less intuitive for interpretation than neighborhood percent white. Therefore, after the models are estimated, I perform an inverse logit transformation to yield the implied values of neighborhood percent white.

I focus my study on three race/ethnic groups: non-Hispanic white; African American; and Hispanic. Although the category 'Hispanic' is treated as an ethnicity by the Census, I include it in my study as a distinct group comparable to white and African American. The groups are exclusive—no individual is counted in both categories.

My independent variable is educational attainment by race for those aged 25 and over. I create 5 education categories based on completed years of school: 1) 0-8 years; 2) 9-11 years; 3) 12 years, or high school level of education; 4) 13-15 years, or some college; and 5) 16+ years, or college degree. These groupings capture major divisions in educational attainment and can be maintained over all decades.

The Houston, Texas metropolitan area was composed of 344 Census tracts in 5 counties in 1970; by 2000, the city had 886 tracts in 8 counties. I use the approach which relies on the metropolitan area at the time studied. This approach allows me to investigate interracial contact within an area as it is socially defined at the time. It also protects me from the bias of including peripheral areas around Houston which were very lightly populated early in the study.

Since my data is limited to one predictor variable, education, I use one-way analysis of variation (ANOVA) to obtain F-test probabilities and coefficients. With a univariate ANOVA, effects are the category means for the categories of the independent variable (in this case, the five categories of education). I create box plot graphs to visually inspect medians and distributions at each point in study period for each group.

I also report the additive effects of education and decade, as well as the interaction of education and year. To do this, I pool all the decades together and create dummy variables for year and education. I then regress the logit of percent white on: 1) education; 2) decade; and 3) education x decade. I report incremental R^2 F-test to assess fit of each model. To account for possible non-linearities in the effect of education, I include a squared education term.

I assess changes in the effects of the interaction of education and year on percent white by education category by reporting graphs of predicted values for each minority group. And finally, to highlight minority changes over time in relation to the majority group, I show graphs of the difference between majority and minority predicted values.

Figure 4 shows an example of the construction of a simple spatial attainment dataset drawn from Summary File tabulations. The first four columns are provided in the race by educational attainment table. The first column shows the geographic identifier which delineates the area of analysis, the Census tract. The second, third, and fourth columns present the race, education, and the count of race by education data. Another

Summary File table gives racial composition by tract. The last column is merged into the dataset by the common geographic identifier in the first column, here labeled “Tract ID”. When the data are in this form, spatial attainment models can be estimated by running weighted regression in which the records are weighted by the number of cases. With weighted regression, each row represents an individual record and the counts (recorded in column 4 of the example) indicate the number of cases (i.e., individuals) with this combination of characteristics. For example, 78 identical cases have the attributes of row 1 (i.e., living in tract 1, Black, education category 1, where 8% of neighborhood is white). Forty-four cases have the attributes of row 2, etc. Each case represents a single individual. So the weighted regressions yield individual-level results from summary-level tables.

Table 6 reviews the sources of the race by education tables used to create my spatial attainment models.

Race by Education Table

Tract ID	Race	Education Category	Count of Race w/ Educ Level
Tract1	2(Black)	1	78
Tract1	2(Black)	2	44
…	…	…	…
Tract2	2(Black)	1	93
Tract2	2(Black)	2	88

Race Count Table

Tract ID	Neighborhood Percent White
Tract1	8%
Tract1	8%
…	…
Tract2	14%
Tract2	14%

Figure 4. Construction of Spatial Attainment Dataset from Summary File Tabulations.

Table 6. Race by Education Tables from 1970-2000 Censuses.

	Sample File	**Table**
1970	Summary File 4(A)	Table 199
1980	Summary File 3(A)	Table P9
1990	Summary Tape File 3(A)	Table P058 (Race) Table P059 (Ethnicity)
2000	Summary File 3	P 148 series

5.2. Micro-Level Attainment Results

The first rows of Table 7 present the population and racial composition for Houstonians aged 25 and older between 1970 and 2000. The table shows that Houston experienced rapid growth over this time. Also, the pattern of composition between the three race/ethnic groups is presented. Whites experienced a steady decline in relative presence from 75% of the population in 1970 to 58% in 2000. The African American percentage stayed constant at 17% over the four censuses. And the Hispanic proportion tripled over the same time going from 8.2 percent in 1970 to 25.3 percent in 2000.

The remaining rows of Table 7 show the education distributions by group in Houston over the four decades of the study period. The total for each group by decade is the count of the sample of persons that completed the long form questionnaire. Generally, each group experienced increases in educational attainment over the study period. Every group has a smaller percentage in the lower education categories in 2000 than in 1970. Conversely, each group has a larger percentage in the higher education categories in 2000 than in 1970. This pattern of increasing educational attainment by group is most evident for African Americans, who more than tripled the percentage point distribution in the highest education categories. The change for the whites over time is less dramatic. Even so, whites experienced a doubling of their percentage distribution in the highest education categories. Hispanics had the least amount of change in their educational distribution over the study period.

Table 7. Population, Racial Composition, and Education Distributions by Race for Houston, Texas 1970 to 2000.

	1970	1980	1990	2000
Population Over 25	1,022,693	1,695,857	2,188,569	2,671,124
Percent White	74.8	72.1	65.8	57.6
Percent Afr. Amer.	17.0	16.4	17.1	17.1
Percent Hispanic	8.2	11.5	17.1	25.3
Non-Hispanic White				
0-8 Years	18.8	9.34	4.4	2.6
9-11 Years	23.0	14.3	10.5	7.9
High School	26.9	30.5	25.3	23.3
13-15 Years	14.7	20.4	29.9	31.4
16+ Years	16.6	25.5	29.9	34.8
Total	152,977	203,783	240,062	256,255
African American				
0-8 Years	38.9	21.2	10.4	5.9
9-11 Years	28.3	21.3	21.5	17.0
High School	20.3	29.5	27.7	28.0
13-15 Years	7.0	16.4	26.0	31.2
16+ Years	5.4	11.7	14.5	18.0
Total	34,729	46,300	62,268	76,220
Hispanic				
0-8 Years	51.2	43.8	37.1	33.4
9-11 Years	16.1	15.6	20.3	22.1
High School	17.3	22.8	19.0	20.0
13-15 Years	8.3	10.6	15.7	16.0
16+ Years	7.3	7.3	7.8	8.5
Total	16,833	32,559	62,430	112,713

Table 8 shows the percent white of the neighborhoods of group members, broken out by educational category, across four decades. Since I have one predictor from my data—education—I use one-way ANOVA to estimate category effects, significance, and explained variance. In this case, the category means are equivalent to category effects, again, because I have a single predictor variable. All effects are statistically significant, but this is not surprising given the large number of individual-level cases for each group. Explained variance in percent white neighborhood based on educational category is generally low—between 2 and 4 percent for whites; between 10 and 12 percent for Hispanics. The story for African Americans is more complicated. For African Americans in 1970 education category explained only $2/10^{ths}$ of one percent of the score of percent white of their neighborhood. By the next decade, however, the explained variance for blacks jumped to 4% and continued at levels comparable to that of whites throughout the remainder of the study period.

The ANOVA results for non-Hispanic whites indicate that they translate higher education into higher percent white neighborhoods at each decade. For example in 1970, the highest educated whites lived in the whitest neighborhoods (87.44% white) while those with the lowest education lived in the least white neighborhoods (75.12% white). This pattern is consistent between all education categories in all decades. From decade to decade another important pattern is revealed: consistent decline in the percent white neighborhood of whites in each education category. Whites in all education categories live in less white neighborhoods than they did in the previous decade over the course of the study period. There may be two reasons for this: 1) the city's overall percent white

Table 8. Means for Percent White by Education and Results of One-Way ANOVA Estimated Separately by Race and Decade for Houston, Texas 1970 to 2000.

	1970	1980	1990	2000
Non-Hispanic White				
0-8 Years	75.12	71.37	64.65	54.06
9-11 Years	79.52	75.03	67.32	58.74
High School	83.16	78.74	70.25	62.03
13-15 Years	85.27	80.56	72.25	64.78
16+ Years	87.44	82.86	75.46	69.42
Prob. of F-test	0.000	0.000	0.000	0.000
Adjusted R^2	0.044	0.038	0.024	0.033
African American				
0-8 Years	22.62	18.49	20.10	17.55
9-11 Years	20.15	18.35	21.57	18.37
High School	21.19	23.39	26.41	21.48
13-15 Years	21.02	30.56	32.99	26.10
16+ Years	22.37	34.74	38.19	32.59
Prob. of F-test	0.000	0.000	0.000	0.000
Adjusted R^2	0.002	0.044	0.055	0.044
Hispanic				
0-8 Years	46.59	42.57	37.26	26.79
9-11 Years	55.52	50.23	41.78	30.09
High School	67.41	58.63	48.72	36.54
13-15 Years	73.09	65.98	56.01	43.65
16+ Years	78.73	73.03	62.65	52.00
Prob. of F-test	0.000	0.000	0.000	0.000
Adjusted R^2	0.129	0.119	0.104	0.107

dropped from 75% to 58% from 1970 to 2000; and 2) the white group lived in extremelywhite neighborhoods at the beginning of the study period—almost any demographic change in the city would cause a decrease in the percent white of their neighborhoods.

For African Americans the ANOVA results indicate dramatic change in the role of education in neighborhood racial composition. In 1970, higher education did not translate into higher percent white neighborhoods for African Americans. As a matter of fact, in 1970 blacks in the lowest education category lived in neighborhoods that were nominally whiter (22.62% white) than any other education category. However, in 1980 those in the higher three education categories saw substantial increases in the percent white of their neighborhoods, while those in the lower two education categories experience a decrease in the percent white. For example in 2000, African Americans in the lowest education category live in neighborhoods which are 17.55% white; while those in the highest education category live in 32.59% white neighborhoods.

For Hispanics, increasing education results in increased white contact at each decade in the study period. For instance, in 1970 the lowest education category of Hispanics lived in 46.59% white neighborhoods; while those in the highest category lived in 78.73% white neighborhoods. Similar to whites, however, from decade to decade Hispanics experience a decrease in percent white neighborhood within each education category. For example, Hispanics with a high school education in 1970 lived in neighborhoods that averaged 67.41% white. But in 2000, high school educated Hispanics live in 36.54% white neighborhoods.

A useful way to visualize group changes in neighborhood percent white by education category is to graph the distributions of each category using a box plot. These box plots are not predictions; they depict the empirical distribution of scores of percent white by race/ethnic group by education category over each decade of the time of the study. Each box represents the interquartile range of values (from the 25th percentile to the 75th percentile) with a horizontal hash mark representing the median value (the 50th percentile score). The x-axis of each of the graphs is comprised of the five education categories. The y-axis is percent white. A bold horizontal bar in each box plot represents the average percent white in the Houston metropolitan area at that time to assist in tracking the changing value of percent white over the study period.

Figure 5 shows the box plot graphs for non-Hispanic whites in Houston over the four decades of the study period. As reported in the ANOVA results for whites, higher education yields higher percent white neighborhoods. Within each decade, the box plots reveal a clear pattern of ascent for increasing education. Also whites in all education categories live in very white neighborhoods. The median scores for each education category (represented by the horizontal line in the middle of the boxes) are higher than the percent white of the city at every year in the study (represented by the bold horizontal line across each graph). In fact, the interquartile range of the highest education category is consistently higher than the percent white of the city—indicating that, by 1990 and 2000, only 25% of the highest educated whites live in neighborhoods that are lower than would be expected under even distribution. Finally, I note that the distributions have increased in size by decade. This is indicated by progressively larger

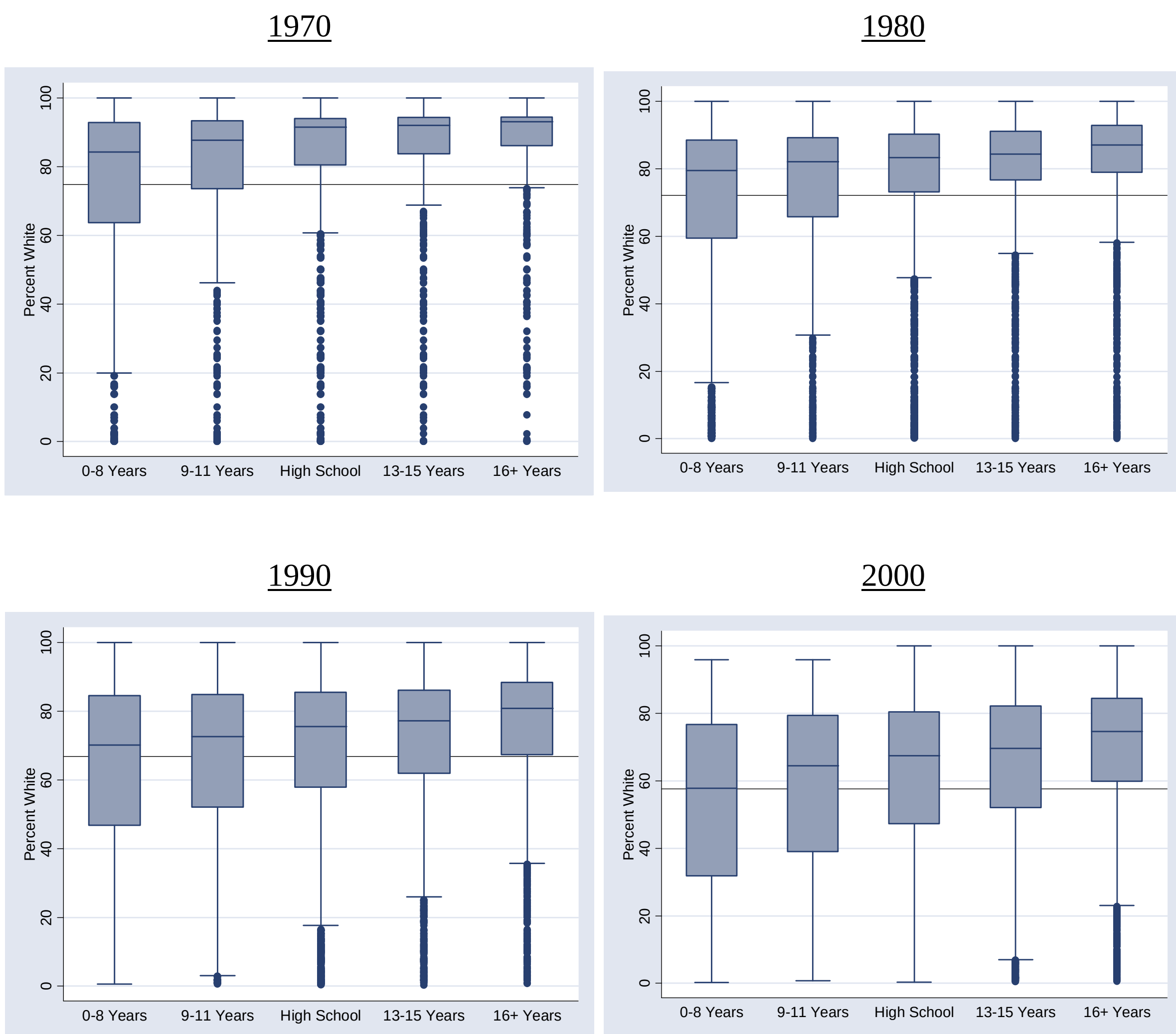

Figure 5. Box Plots Depicting the Distribution of Percent White by Education for Whites in Houston, Texas 1970 to 2000.

boxes over the decades of the study period. It is worth noting that the lowest education categories also have the more variable distributions.

Figure 6 has box plots depicting the distribution of percent white by education category for African American Houstonians between 1970 and 2000. In 1970, the medians for the higher three education categories are nominally higher than the lower two, but all the scores are quite similar. In 1980 and beyond, however, higher education yields higher percent white neighborhoods. Even so, the medians are never close to the percent white of the city which is indicated by the bold horizontal reference line. Even the interquartile range of percent white neighborhood is always below the percent white of the city. Furthermore, in 1970 the distributions (represented by the size of the boxes) are all about the same size. But in 1980 the higher education categories experience a widening distribution, indicated by larger boxes for higher education categories. This general pattern persists but retreats somewhat so that by 2000 the distributions for the higher education categories are not much larger than for the lower categories.

Figure 7 shows box plots depicting the distribution of percent white by education category for Hispanic Houstonians between 1970 and 2000. For Hispanics, higher education always yields higher percent white neighborhoods. However, the pattern for Hispanics is one of consistent decade-to-decade decline. Every median for every education category is lower in a subsequent decade. All categories appear to decline in percent white neighborhood at about the same rate across decades. Until 2000, the median for the highest education group was higher than the percent white of the city. Another interesting pattern for Hispanics is a change in the distribution. In the early

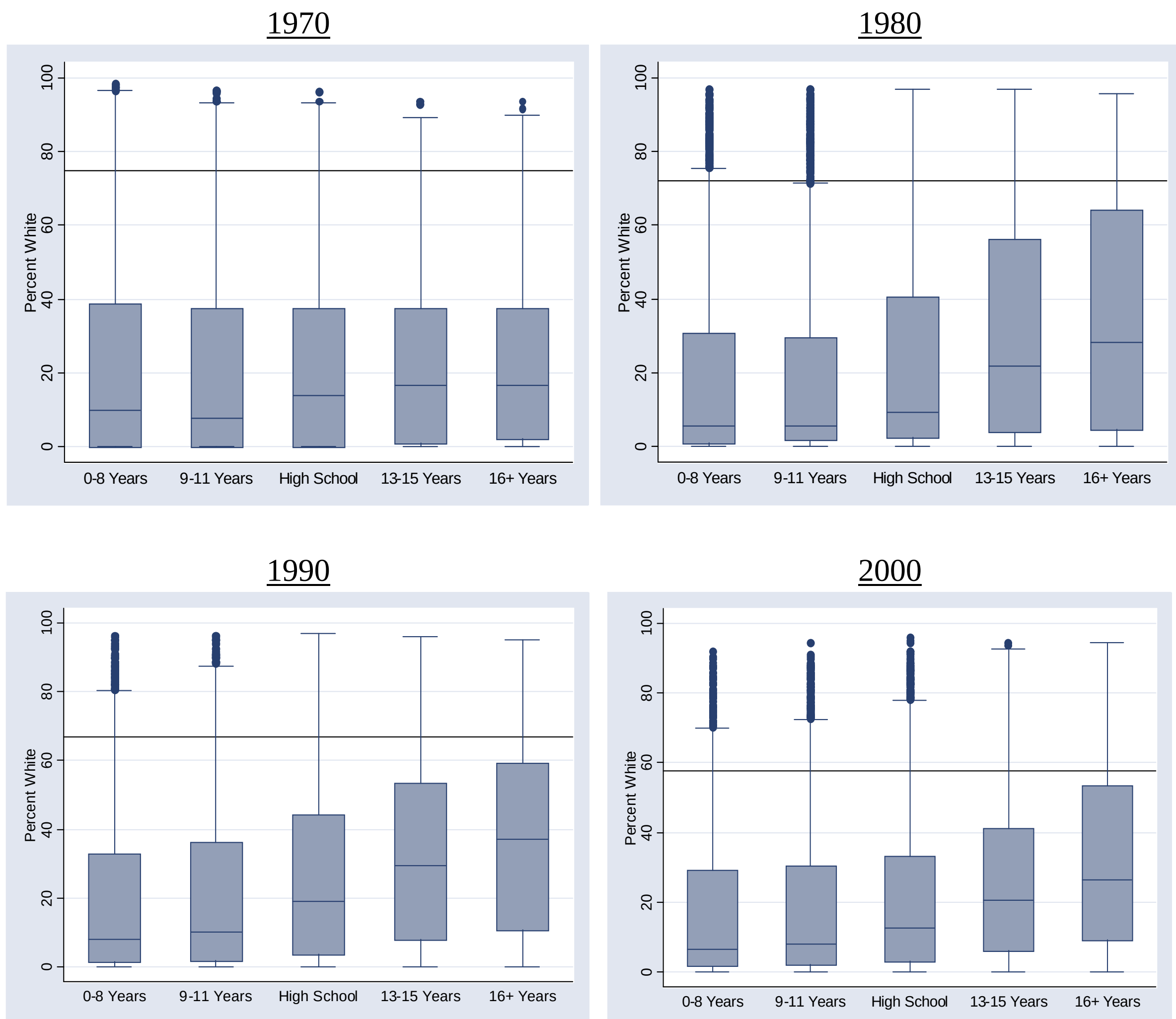

Figure 6. Box Plots Depicting the Distribution of Percent White by Education for African Americans in Houston, Texas 1970 to 2000.

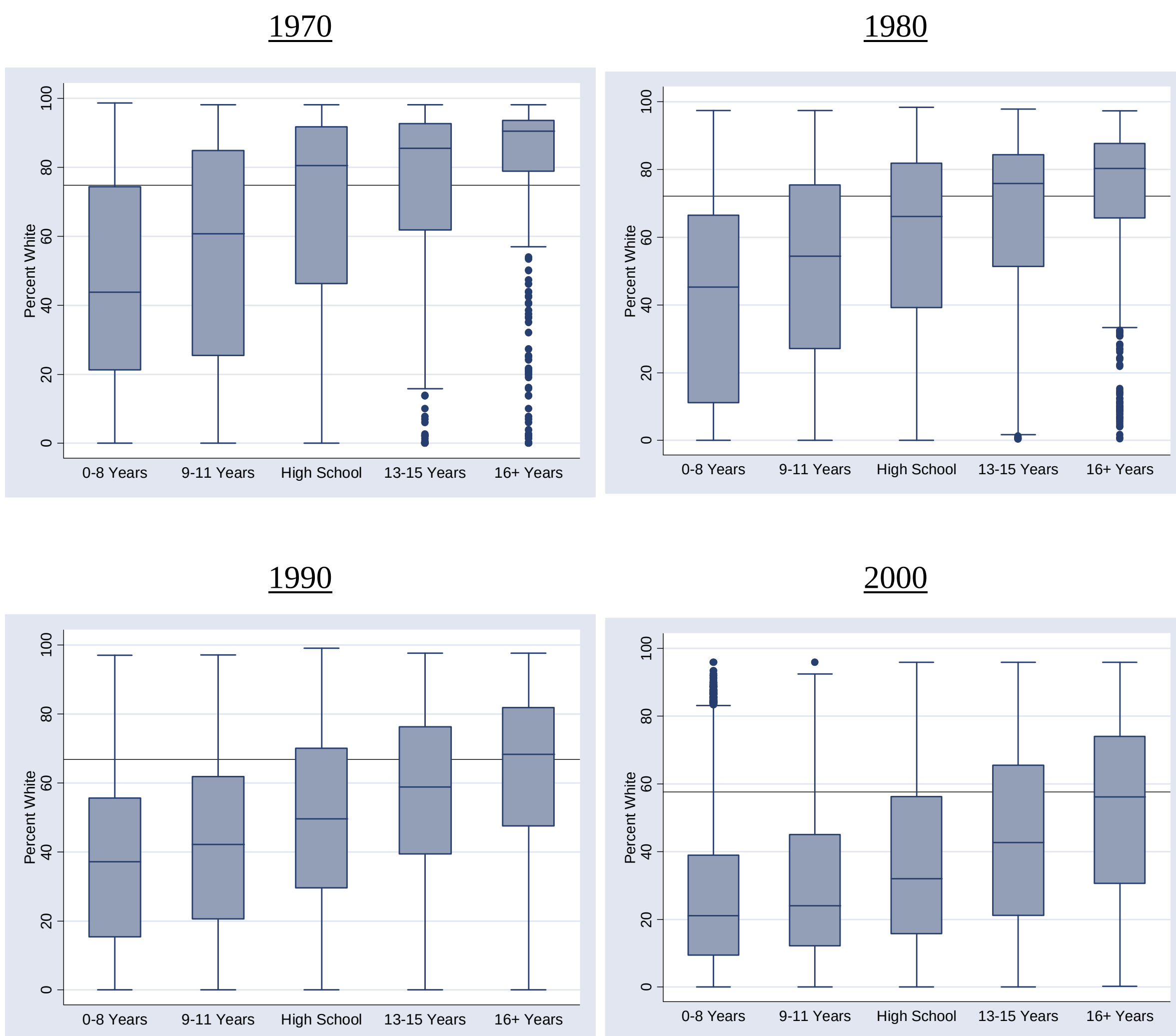

Figure 7. Box Plots Depicting the Distribution of Percent White by Education for Hispanics in Houston, Texas 1970 to 2000.

years of the study, the lower education categories had the widest distributions across percent white neighborhood. But by the later years of the study, the higher education education categories had the wider distributions, while the lower education categories became much narrower (indicated by smaller boxes).

Tables 9, 10, and 11 show the additive effects of education and decade, as well as the interaction of education and decade for each group in successive regression models by race/ethnic group. To accomplish this, I pool the data from all four censuses together and create dummy variables for each decade. I specify 1970 as my omitted dummy category against which the other decades are compared. To account for nonlinearities in the education variable, I square the education term. Since I am using ordinary least squares regression, my dependent variable needs to be unbounded. Therefore, I use a logit-transformed percent white variable as my dependent variable. The resulting coefficients predict changes in the logit of neighborhood percent white for each increase in education category or each change in decade. All results are statistically significant. Also, the increment to R^2 F-test is significant and shows the addition of subsequent variables increases the model fit significantly.

Table 9 shows the regression of the logit of percent on education and decade for non-Hispanic whites. The additive effect of education in Model 1 is very small. By itself, education explains 0.4% of the variation in the logit of percent white of white respondent’s neighborhoods. But when the additive effect of decade is entered into Model 2, the effect of education increases and the explained variation increases to 14.1%. Decade effects for whites are negative and large relative to the constant. From

Table 9. Regression of Logit of Percent White in Neighborhood on Education and Decade for Non-Hispanic White Group in Houston, Texas.

	1	2	3
Education[a]	0.014	0.033	0.041
Year[b]			
1980		-0.344	-0.286
1990		-0.854	-0.762
2000		-1.285	-1.257
Educ x Year[b]			
1980			-0.010
1990			-0.014
2000			-0.007
Constant	1.109	1.699	1.651
Prob. of F	0.000	0.000	0.000
Adjusted R^2	0.004	0.141	0.141
Sample n	4,914,225	4,914,225	4,914,225

All coefficients are statistically significant.
[a]Education term is squared to account for nonlinearity.
[b]Increment to R^2 F-test is statistically significant for all models.

the discussion above it is clear that white neighborhoods were very white in 1970; therefore subsequent decades will yield negative coefficients on the logit of percent white. The coefficients for the interaction of education and year are very small and do not shift the scores of the additive effects much. They are however, statistically significant, due to the large sample size from the pooled data.

Table 10 shows the regression of the logit of percent on education and decade for African American Houstonians. For this group, the effect of education on logit of percent white neighborhood is much larger than for whites, 0.091 compared to 0.014. Also, the amount of variation explained by education is more substantial, 5.1% compared to 0.4%. However, the constant for the education model for African Americans is much lower than it is for whites, -2.43 compared to 1.109. The additive effect of including decade-level changes to the model increases the amount of explained variation to 6.0%. All of the decades have a positive effect on the logit of percent white when compared to the very low 1970 omitted variable. The effect of 1990 is the strongest effect. The interaction terms are small, but their inclusion in Model 3 decreases the additive effects of education and decade substantially.

Table 11 shows the regression of the logit of percent on education and decade for Hispanics in Houston. The additive effect for education is similar to that of African Americans, at 0.09. However, the constant for Hispanics is much higher at -0.86 compared to -2.43. The additive effects for decade are all negative, relative to the high percent white neighborhood composition of Hispanic respondents in 1970. Education

Table 10. Regression of Logit of Percent White in Neighborhood on Education and Decade for African Americans in Houston, Texas.

	1	2	3
Education[a]	0.091	0.083	0.027
Year[b]			
1980		0.328	0.129
1990		0.655	0.429
2000		0.466	0.361
Educ x Year[b]			
1980			0.067
1990			0.069
2000			0.050
Constant	-2.430	-2.813	-2.670
Prob. of F	0.000	0.000	0.000
Adjusted R^2	0.051	0.060	0.063
Sample n	1,281,481	1,218,481	1,218,481

All coefficients are statistically significant.
[a]Education term is squared to account for nonlinearity.
[b]Increment to R^2 F-test is statistically significant for all models.

Table 11. Regression of Logit of Percent White in Neighborhood on Education and Decade for Hispanic Group in Houston, Texas.

	1	2	3
Education[a]	0.090	0.095	0.132
Year[b]			
1980		-0.268	-0.224
1990		-0.661	-0.548
2000		-1.265	-1.142
Educ x Year[b]			
1980			-0.019
1990			-0.040
2000			-0.043
Constant	-0.856	-0.005	-0.105
Prob. of F	0.000	0.000	0.000
Adjusted R^2	0.074	0.150	0.152
Sample n	1,326,583	1,326,583	1,326,583

All coefficients are statistically significant.
[a]Education term is squared to account for nonlinearity.
[b]Increment to R^2 F-test is statistically significant for all models.

and decade explain 15% of the variance in the logit of percent white for Hispanic Houstonians over the study period. And the interaction effects of education and decade are very small, but their inclusion boosts the additive effect of education substantially to 0.132.

Table 12 shows the implied values of percent white for race and education categories by decade in Houston. I predict the values from the regression equations reported in the previous tables titled “Model 3,” the model which includes education and decade, as well as the interaction of the two. These values have been transformed back from logits to percent white to facilitate interpretation. The bottom panel of Table 12 shows the majority-minority gap—that is, the difference between white and minority predicted values. To assist in visualizing the changes within groups and across education and decade, I have plotted the predicted values as line graphs for ease of interpretation. Each line in the graph represents the predicted values of percent white at a particular decade over the categories of education.

Predicted values for non-Hispanic whites are listed across the top rows of Table 12 and are plotted in Figure 8. The predicted values for whites reveal two general patterns: consistent decadal decline in percent white neighborhood for each education category; and a slight increase in the slope of spatial attainment with whites through education over the study period. The decline by decade is evident in Figure 8. The line for 1970 predicted values is on the top, the line for each subsequent decade is progressively lower. The slight increase in the pattern of spatial attainment appears in Figure 8 as an increased slope for 1990 and 2000 values. The drop in percent white

Table 12. Implied Values of Percent White for Race and Education Categories by Decade in Houston, Texas

	1970	1980	1990	2000
Predicted Value[a]				
White				
0-8 Years	83.9%	79.7%	70.9%	59.7%
9-11 Years	84.5%	80.2%	71.4%	60.6%
High School	86.0%	81.6%	73.1%	63.0%
13-15 Years	88.3%	83.8%	75.7%	66.9%
16+ Years	91.0%	86.6%	79.0%	72.1%
African American				
0-8 Years	6.5%	7.3%	9.6%	9.0%
9-11 Years	6.6%	8.0%	10.5%	9.7%
High School	7.2%	10.3%	13.5%	11.9%
13-15 Years	8.1%	15.6%	20.2%	16.6%
16+ Years	9.7%	26.3%	33.2%	25.5%
Hispanic				
0-8 Years	47.4%	41.9%	34.2%	22.3%
9-11 Years	50.7%	44.6%	36.3%	23.9%
High School	60.4%	53.0%	42.9%	29.1%
13-15 Years	74.7%	66.5%	54.3%	39.1%
16+ Years	88.1%	81.4%	69.3%	54.5%
Majority - Minority Gap				
White - Afr. Amer.				
0-8 Years	77.4%	72.4%	61.3%	50.7%
9-11 Years	77.8%	72.2%	60.9%	50.9%
High School	78.8%	71.3%	59.5%	51.1%
13-15 Years	80.2%	68.3%	55.5%	50.4%
16+ Years	81.3%	60.2%	45.8%	46.6%
White - Hispanic				
0-8 Years	36.5%	37.8%	36.6%	37.4%
9-11 Years	33.8%	35.5%	35.1%	36.6%
High School	25.6%	28.6%	30.2%	33.9%
13-15 Years	13.6%	17.3%	21.4%	27.9%
16+ Years	2.9%	5.2%	9.7%	17.6%

[a]Based on Predictions from Regression of Logit of Percent White on Education, Decade, and Education by Decade Interaction.

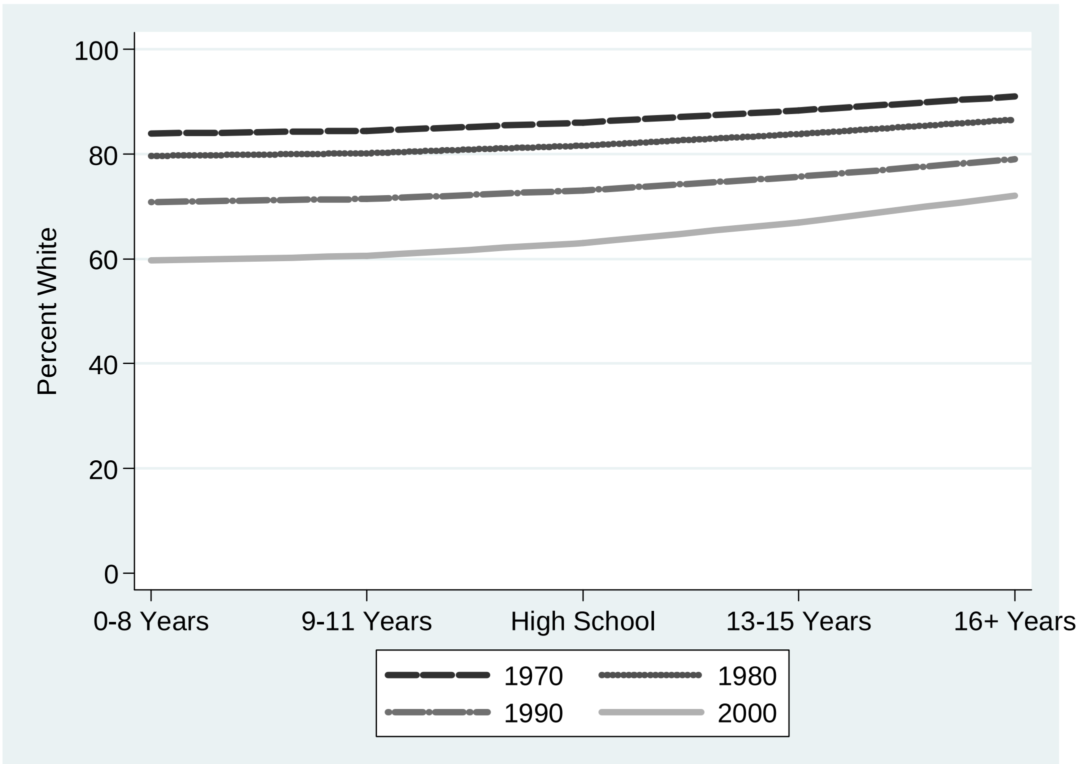

Figure 8. Implied Values of Percent White Based on Predictions from Regression of Logit of Percent White on Education, Decade, and Education by Decade Interaction for Non-Hispanic Whites in Houston, Texas.

neighborhoods for whites appears to be happening at the lower end of the educational categories.

Figure 9 plots the implied values of percent white for African Americans by education over the study period. These same values are listed in Table 12. In 1970, the predicted values are very small, but indicate a very slight slope for increased percent white neighborhood by increasing educational category. However, subsequent years reveal a substantial increase in the percent white neighborhood scores for the top three education categories. The change from 1970 to 1980 is the largest single decade increase. The slope for 1990 has the highest scores. The 2000 predicted values return to levels seen for African Americans in 1980. This drop is due to the contracting percent white of the overall city.

Figure 10 plots the implied values of percent white for Hispanics by education over the study period. These same values are listed in Table 12. For Hispanics, their slope of spatial attainment with whites stays the same over the decades. However, as with the predicted values for non-Hispanic whites, each decade’s line is lower than its predecessor. Again, the slope and shape of the predicted values does not change, but is uniformly lowered each decade due to a steadily falling percent white in the city.

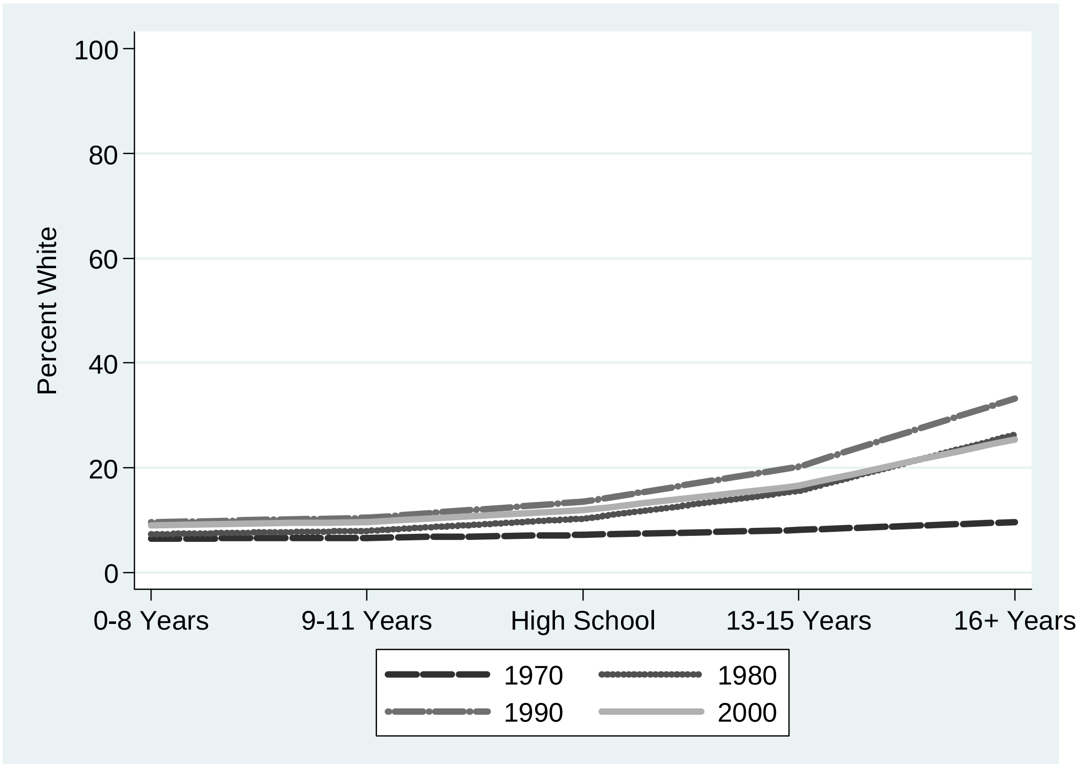

Figure 9. Implied Values of Percent White Based on Predictions from Regression of Logit of Percent White on Education, Decade, and Education by Decade Interaction for African Americans in Houston, Texas.

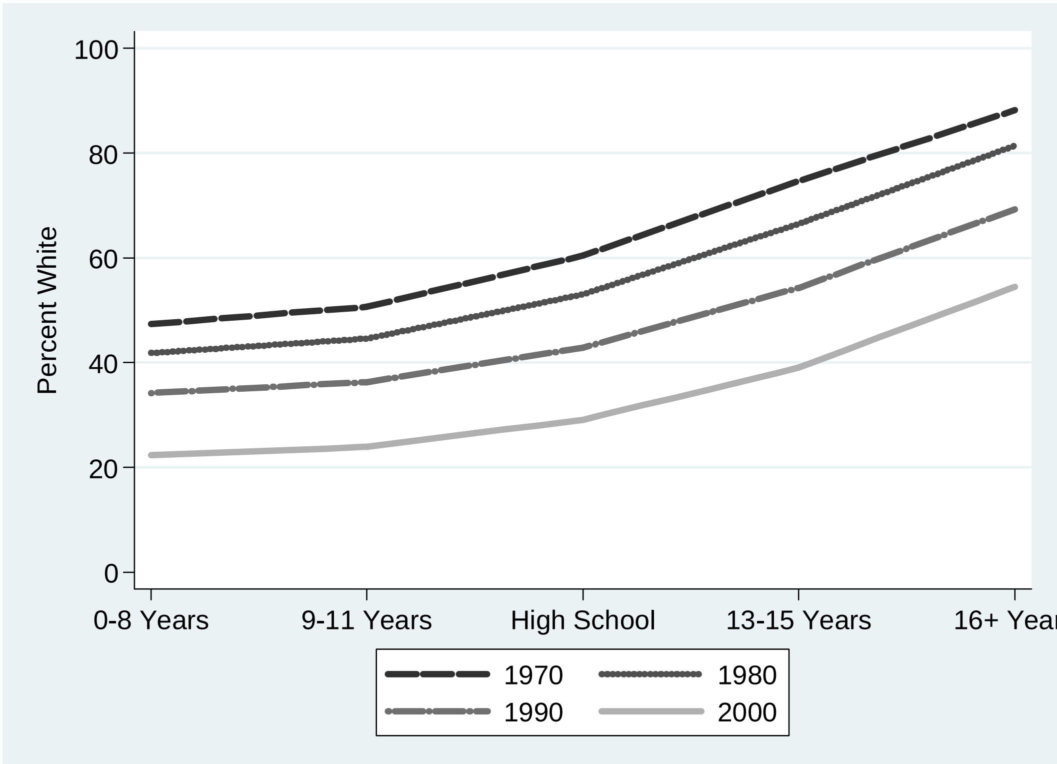

Figure 10. Implied Values of Percent White Based on Predictions from Regression of Logit of Percent White on Education, Decade, and Education by Decade Interaction for Hispanics in Houston, Texas.

Figures 11 and 12 show the difference in white minus black and white minus Hispanic implied values of percent white. These values are also reported in the bottom panel of Table 12. In a theoretical condition of white-black integration, ignoring other groups, each of these scores would be zero. That is, the city average might change over time, but each group's predicted value of percent white neighborhood by education category would be the same—yielding no difference. In a theoretical condition of pure segregation, ignoring other groups, these scores would be 100. If the lines are horizontal, then both the minority and the majority groups experience the same consequences of increased education. If the lines exhibit a negative slope, then increased education for the minority group has more effect on percent white neighborhood than increased education for the white majority.

Figure 11 shows the difference in white minus black implied values of percent white. Figure 11 reveals several interesting findings. In 1970, differences in percent white neighborhood are slightly higher for highly educated blacks than for less educated blacks—but the slope is very slight. In 1980 and 1990, the lines show a monotonic descent across education categories—indicating that education increases yield higher consequences in percent white neighborhood for African Americans than for whites. This effect is especially pronounced for high education category African Americans. However, in 2000, the difference between whites and blacks stays the same across the four lowest education categories, only dropping slightly for the highest category. This horizontal line indicates that whites and blacks had roughly the same educational effect

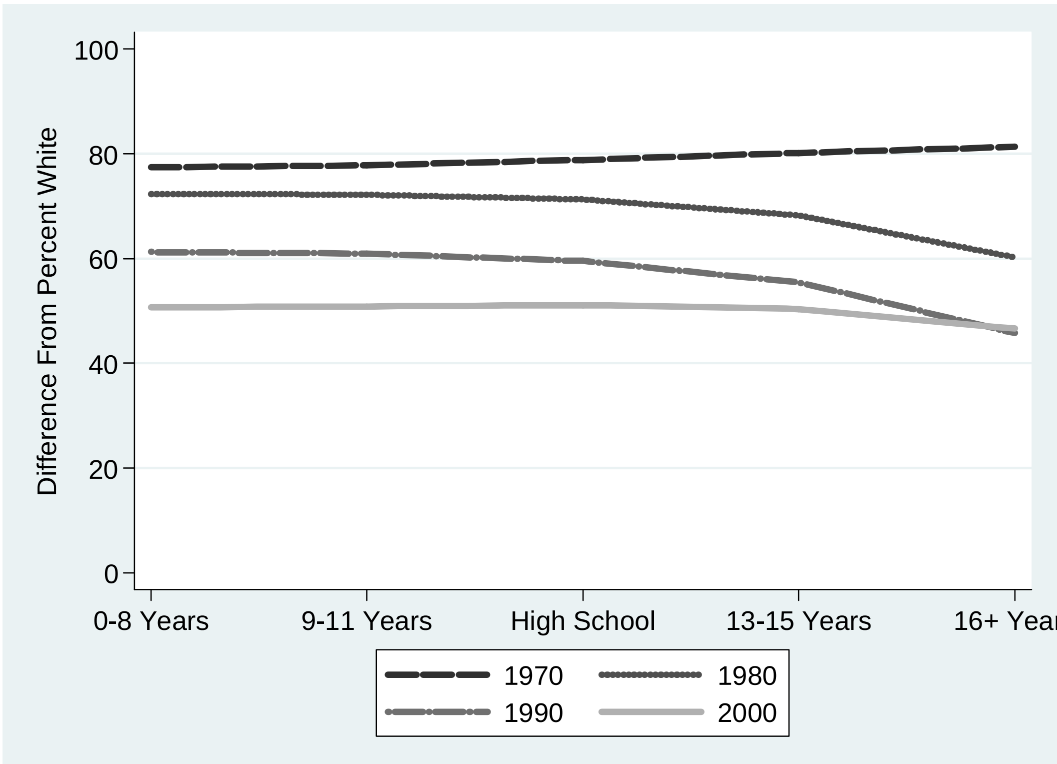

Figure 11. White Minus Black Implied Values of Percent White Based on Predictions from Regression of Logit of Percent White on Education, Decade, and Education by Decade Interaction.

on percent white neighborhood. The 2000 line is lower because lower education category *whites* live in less white neighborhoods.

Figure 12 plots white minus Hispanic implied values of percent white. These values can also be found in the bottom panel of Table 12. The left-hand side of Figure 12 shows that the difference between whites and Hispanics is quite consistent at the two lowest levels of education. The least amount of difference between the groups occurred in 1970, when the predicted value of percent white for the highest education category of Hispanics was 88.1% and for non-Hispanic whites was 91.0%--yielding a difference of 2.9%. Each subsequent decade has an increased difference between white and Hispanic implied values, although the slopes are quite similar. Segregation is slowly increasing between whites and Hispanics, even though a clear pattern of spatial attainment by education exists for Hispanics in all decades of the study period.

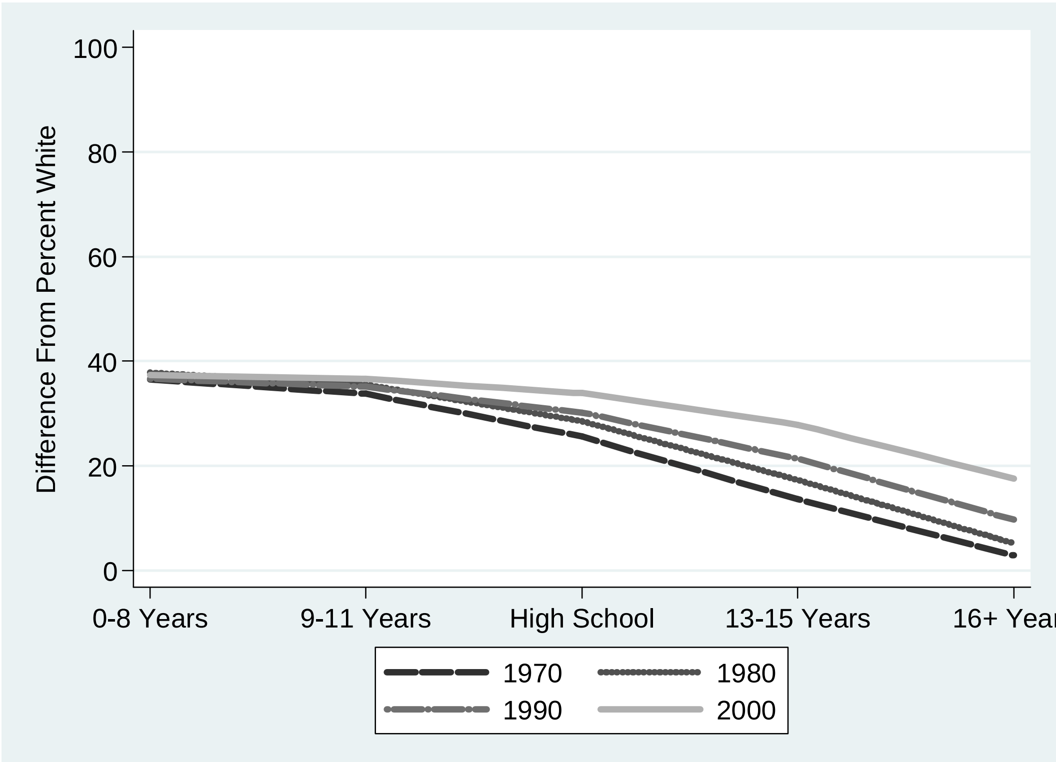

Figure 12. White Minus Hispanic Implied Values of Percent White Based on Predictions from Regression of Logit of Percent White on Education, Decade, and Education by Decade Interaction.

5.3. Discussion of Spatial Attainment Trends in Houston

This section extends the current literature by showing that trends in spatial attainment can be assessed by adapting Census cross-tabulations. The technique introduced by Alba and Logan (1992), had been previously applied to multi-group comparisons at a single point in time. I find this method appropriate and useful in assessing trends in spatial attainment. I demonstrate the method's usefulness by applying it to three race/ethnic groups in Houston, Texas between 1970 and 2000.

Substantively, my models reveal two interesting and simultaneous dynamics affecting spatial attainment: diminishing white presence in the city; and increasing integration. The significant drop in percent white of the city over the study period affects the neighborhood percent white for each group. The integration effects are quite different between the groups: whites and Hispanics see declines each decade in white contact across all education categories; African Americans witness increases in white contact for the higher education categories.

For whites in 1970, whites of any SES lived with the same amount of contact with whites. By 2000, lower SES whites experience less contact with whites than higher status whites—who are simultaneously excluding minorities and lower SES whites. This trend reveals an increasing role of status in contact, even within the majority group.

A more complicated story unfolds in the trends for African. In 1970, African Americans experienced minimal returns in terms of spatial attainment. However, the very next decade after the legal changes affecting housing markets were enacted, strong spatial attainment patterns emerged for this group. In 1980 and 1990, spatial attainment for African Americans had greater returns on education than did whites. The gains in percent white neighborhood peaked in 1990. By 2000, the effect of education had returned to levels found in 1980.

Hispanics exhibit a clear pattern of spatial attainment over the study period. The differences in percent white neighborhood between low and high educated Hispanics remains roughly equivalent between 1970 and 2000. However, segregation is steadily increasing between Hispanic and white Houstonians, as revealed in the overall decline in spatial attainment outcomes for Hispanics at each level of education.

Future research should expand this method to assess the trends of spatial attainment in other cities of interest. A national study could also be performed which estimates spatial attainment for various groups over a long period of time. I am also interested in changes in the spatial attainment of these same groups in Houston in 2010, when the next Census is conducted. Using more recent Census tabulations, similar methods may be useful for Asian Americans and/or biracial households. Also, it is possible to expand the method to other Census cross-tabulations. This would allow for new dependent variables such as family income, poverty status, or employment status. Finally, the spatial attainment literature to date has focused on inter-group comparisons (e.g., between whites, African Americans, and Hispanics); but I find within-group variation to be quite interesting—especially for whites, as their intra-group variation has increased between 1970 and 2000.

6. COMPARISON OF SPATIAL ATTAINMENT OVER VARIOUS SPATIAL SCALES

Spatial attainment research investigates how individual-level characteristics determine neighborhood-level outcomes. Neighborhood-level dependent variables commonly used in the spatial attainment literature include percent white (Massey and Denton 1985; Massey and Mullan 1984); median household income (Logan, Alba, Mcnulty and Fisher (1996); Alba and Logan 1992); and suburban residence (Hwang and Murdock 1998; Logan and Alba (1991). Other useful dependent variables include percent poverty (Jargowsky, 1997) and exposure/isolation to or from other groups (Bayer, McMillan, and Rueben 2004; Massey and Denton 1987; Portes and Rumbaut 2001; St. John and Clymer 2000). Whatever the choice, the dependent variable must be measured at some level of geography.

Since the dependent variable of spatial attainment research is always bounded in geography, it is critical for the "spatial" aspect of spatial attainment to be valid and reliable. For example, arbitrarily assigned areas useful to researchers may not be distinguishable neighborhoods to residents. Or, using areas that are too large might dilute spatial attainment outcomes through homogenization (i.e., there is not enough between-area variation). Incorrectly specified spatial scale, conceptualized here as neighborhood size, leads to invalid and unreliable assessment of neighborhood outcomes.

Researchers have not had much choice in the selection of geographic area. The needed combination of individual-level data and neighborhood-level outcomes runs into the problem of confidentiality. The Census provides one of the better sources of data for spatial attainment models—extensive individual or household level characteristics connected with a very specific geographic boundary. However, the Census Bureau has the responsibility of protecting the confidentiality of its respondents. Therefore, it only publishes microdata through the Public Use Microdata Set (PUMS) which limits geographic information to areas of over 100,000 people.

As a result of this limitation, most spatial attainment research has often had to rely on very large areas. Logan and Alba (1992) rely on suburban places instead of tracts to capture neighborhood outcomes. Logan and Alba (1993) are limited to suburban/urban place distinctions in lieu of more refined areal statistics. Fong and Shibuya (2000) use the Integrated Public Use Microdata Series must use a Public Use Microdata Area that corresponds to relatively large area of 100,000 people. Freeman (2002) likewise, uses a sub-borough area limited to New York City which contains areas averaging 100,000 people. Locational attainment is not possible with the Multi-City Study of Urban Inequality, a widely used dataset for investigations of race relations in four U.S. cities. A website associated with the study notes, “The geographic identifiers (tracts and block group) for the household files are not available to anyone. There are no exceptions to this policy.”

Few researchers have had access to specially prepared microdata suitable for spatial attainment estimation. Such access is usually irregular and expensive. Massey and Denton (1985) use the Neighborhood Characteristics File for the 1970 Census. However, that dataset was suspended in 1980. Gross and Massey (1991) and White, Biddlecom and Guo (1993) purchased a special tabulation of the 1980 Public Use Microdata Set (PUMS-F) to get access to tract-level data with statistical “noise” added to protect confidentiality. They report that a similar tabulation for 1990 data was suspended. Bayer, McMillan, and Reuben (2003) use access to restricted micro-Census data for the San Francisco area to estimate spatial attainment models using the full Census population at the smallest possible geography, the Census block. However, as denoted by the term “restricted,” access to this data is expensive and subject to strict rules of the Census Bureau to protect confidentiality.

The role of spatial scale in the dependent variable of spatial attainment deserves systematic attention. Yet, little research has assessment of spatial attainment effects vary with neighborhood scale. The broader literature on residential segregation suggests that scale matters (Roof and Van Valey 1972; Taeuber and Taeuber 1965; Van Valey and Roof 1976a). The larger each neighborhood area is, the lower the segregation score. For example, if we calculate the segregation score for a city, but instead of tracts or boroughs we use the entire city as the only area, the segregation score is equal to zero—obviously, the city is perfectly proportional to itself. Invariably, the segregation index will increase artifactually when smaller areas are used; even though no changes are being made to the group proportions within the city. Therefore, segregation scores based on PUMAs or Zip Code Tabulated Areas will be lower than segregation scores computed from block groups or blocks.

Bayer, McMillan, and Reuben (2003) offer an important analysis of the impact of spatial scale. In their study using restricted Census data at the block level, they present an appendix which compares ethnic group exposure/isolation findings to illustrate the value of using smaller geographic areas. Their study investigates exposure at the block, block group, tract, PUMA, and county levels. They do not include ZCTAs. As expected, they report that smaller neighborhood scale reveals greater in-group exposure for race/ethnic groups (non-Hispanic Whites, African Americans, Hispanics, and Asians) in San Francisco. Minority in-group exposure is much larger at the tract-level than at the PUMA-level. For African Americans, in-group exposure at the tract-level is 38.3%, but at the PUMA-level it is 25.6%.They also find that block-, block group-, and tract-level exposure rates are always closely correlated. However, the PUMA- and the county-level exposure rates are substantially higher than rates reported at the smaller levels.

Another notable point from Bayer, McMillan, and Reuben (2003) is that the amount of variation explained by the model variables (reported as adjusted R^2) may not fall monotonically as the neighborhood size increases from block through county. They calculate an exposure index using a regression model which yields standard errors and adjusted R^2 values. Since their data come from the restricted Census data, they have very large n's—over 250,000 cases—so their standard errors are very small. We might expect models at the block-level, the smallest unit of aggregation possible, to explain the greatest amount of variation among the geographies. However, they report that the adjusted R^2 is smaller at the block-level than at the block group- or the tract-level. They observe this for all groups, even the white majority of San Francisco.

The variation between findings obtained at differing spatial scales is attributable to an aspect of what is termed the *modifiable areal unit problem*: the scale effect (Openshaw 1983). The scale effect is a statistical variation in the same set of data grouped at a different spatial scale. This effect is applicable to the variation in my results between block groups, tracts, ZCTAs and PUMAs.

Because of the scale effect, I expect to see differences between spatial attainment outcomes at various spatial scales. Generally, spatial attainment models based on smaller geographies will reveal stronger effects and higher levels of explained variance (i.e., higher R^2) on my dependent variable logit of percent white. Models from smaller spatial scales should also generally yield higher coefficients for minority contact with whites. Conversely, models based on larger spatial scales will explain less variance and will yield lower slope coefficients for minority contact with whites.

Also, some slope coefficients from models based on larger geographies may lose statistical significance due to less variation in percent white at larger geographies of the city. Therefore, when using larger spatial scales, researchers should be wary of Type II errors or “false negatives.”

Figure 13 shows the percent white for units at four levels of spatial scale. The block group map in the top left quadrant shows the most geographic detail for percent white and the most variation between neighborhoods. Generally, detail is lost and there are fewer areas of extreme “whiteness” as the geographic extent is increased. For example, the PUMA map shows a smaller area of less than 20% white, and no areas of

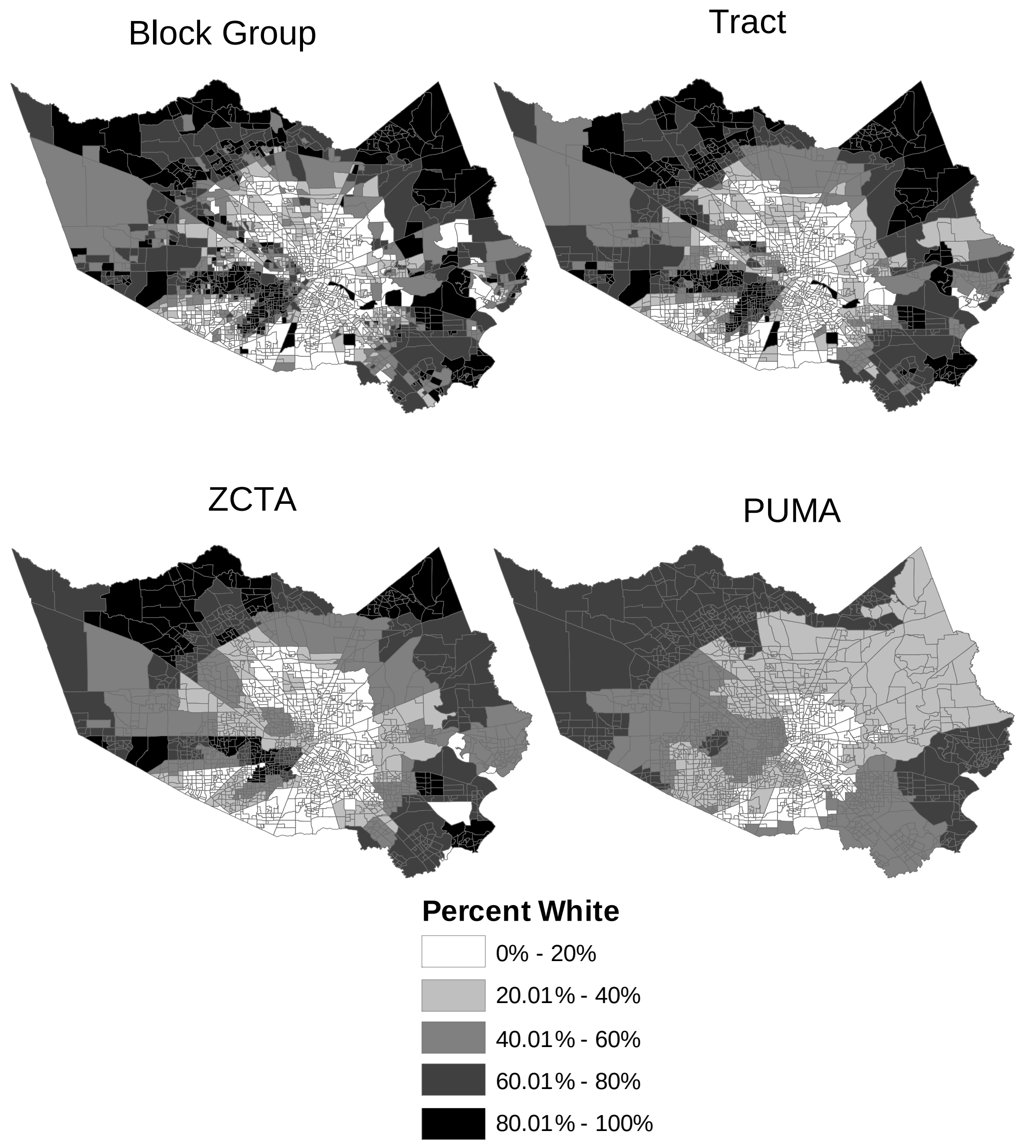

Figure 13. Percent White of Harris County, Texas, by Block Group, Tract, Zip Code Tabulated Area, and Public Use Microdata Area, 2000.

Note: Block group boundaries are shown for each map. Clusters of increasing size appear at each higher level of neighborhood size as laid over the block groups. The PUMA map reveals a smaller area of less than 20% white, and no areas of greater than 80% white.

more than 80% white. Also, the clusters of similarly proportioned white areas are larger at each larger scale.

Public Use Microdata Areas (PUMAs) are frequently used as the geographic scale of spatial attainment models (Fong and Shibuya 2000). They are attractive to spatial attainment researchers because of the associated Public Use Microdata Sample dataset. This dataset, provided by the Census, connects rich individual-level data with PUMA geographies. However, stretching the geographic scale of a PUMA to fit the social context of a neighborhood is problematic. PUMAs are not socially meaningful units for individual residential outcomes. Most people don't change jobs to "move into a better PUMA." Clearly, the logic of spatial attainment analysis requires that the neighborhood areas we use to measure our dependent variable roughly correspond with social neighborhoods.

PUMAs are also problematic for spatial attainment researchers. In order to protect the confidentiality of Census respondents, PUMAs are intentionally created too large to be able to identify an individual in his or her neighborhood. Therefore, each PUMA is required contain more than 100,000 respondents. In Harris County, Texas, the average PUMA has 136,810 people (*s.d.* = 15,327) and the smallest PUMA has 107,656 people. Another potential difficulty for those relying on PUMAs in spatial attainment research is that, because they are so large, there are few PUMAs to work with. For example, in Harris County there are 24 PUMAs. When there are so few areas and their construction artificially limits the minimum quantity of cases the amount of variation to be estimated between the areas will be very low. For example, the standard deviation of my set of 24 PUMAs in Harris County is 11% of the population mean for all PUMAs. For ZCTAs, tracts, and block groups the standard deviations are around 45% of the mean.

Another concern for the use of PUMAs in spatial attainment research is that they contain islands. Because the rules of their construction insist that each one must contain more than 100,000 people, they may include discontiguous areas in order to reach the minimum population mark. However, social neighborhoods are rarely discontiguous. Islands and enclaves are common in all metropolitan areas, but they are not often arbitrarily grouped together. Normally, areas that are very similar demographically but are disconnected geographically are kept quite distinct in the minds of residents. Figure 14 shows an example of a PUMA island. The PUMA numbered 04619 includes multiple, separate areas in southwest Harris County.

A final concern is that PUMAs often cross other Census boundaries. For example, a single Census tract may be in two PUMAs. Conceivably, this could lead to measurement error the neighborhood-level dependent variable. A tract with a white/black composition of 70/30 may be split by a PUMA that includes only one group. Ultimately, this division of Census tracts may not be substantial. The average tract in Harris County has 6,266 people (*s.d.* = 3,096) and the largest tract has a population of 18,550. As mentioned before, the average PUMA in Harris County numbers 136,810 people. On average, a single tract contributes around 5% to the total population of a PUMA. So, a split tract would contribute less than that. Also, the most numerous tracts

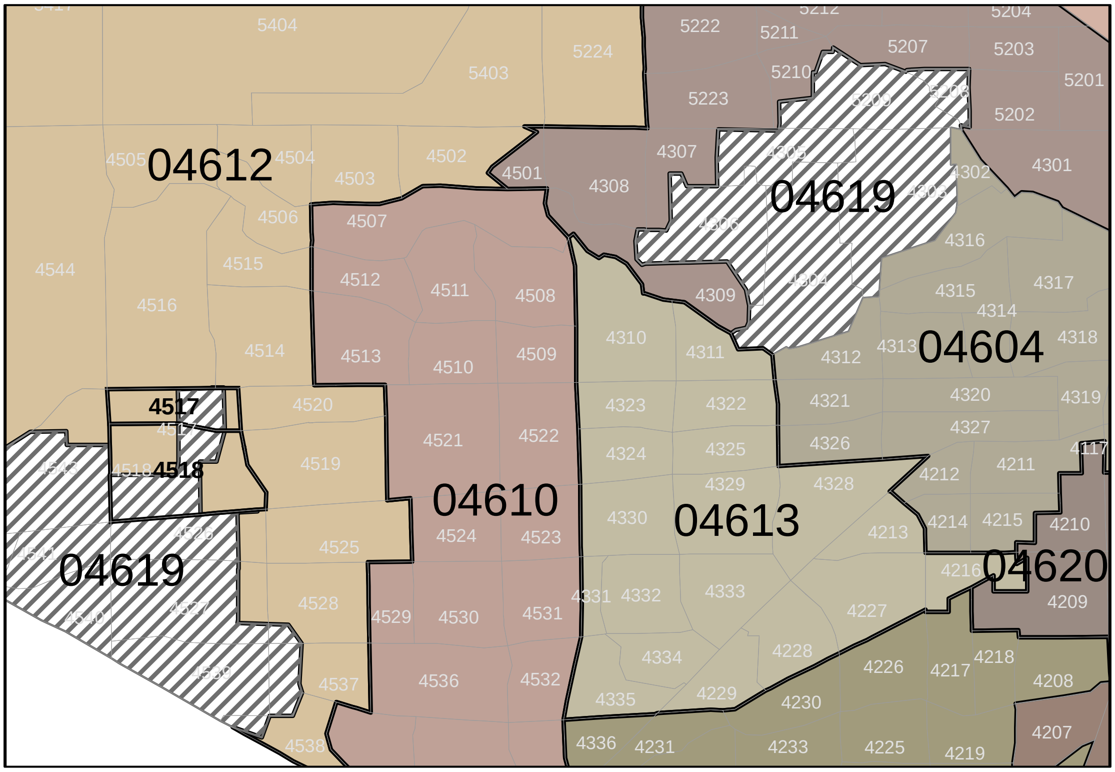

Figure 14 Public Use Microdata Area (PUMA) Islands and Division of Census Tracts in Harris County, Texas, 2000.

Note: This map detail focuses on the southwest corner of Harris County. PUMAs are distinguished here by color and numbered identifier in a large font. Tract boundaries are much smaller. The PUMA numbered 04619 is composed of multiple, separate areas. Also, the section of that PUMA in the lower left divides two tracts—#4517 and #4518.

are in the most numerous PUMAs. Figure 2.3.2 shows that in PUMAs by color, laid over Census tracts. The PUMA numbered 04619, in the bottom left of the image, includes part, but not all of tract number 4517. The rest of that tract is in the PUMA numbered 04612.

A previously unexplored geographic scale is now available to spatial attainment researchers. Census 2000 introduced a new geographic boundary termed “ZIP Code Tabulated Area” (ZCTA). This geographic scale allows for computation of statistical analysis at the ZIP code level provided by the United States Postal Service. ZCTAs are aggregates of Census blocks bounded by ZIP code postal delivery areas. While this new geographic area is much larger than a traditional neighborhood (which is normally associated with the Census tract-level of geography), it is often used in surveys because respondents can readily self-identify their own ZIP code, and are willing to provide this information. In contrast, respondents cannot readily identify their census tract and are often reluctant to offer detailed address information needed for geo-coding.

ZCTAs have not yet been used directly in spatial attainment research. Borjas (1998) uses subsets of respondents of the National Longitudinal Survey of Youth (NLSY) who lived in the same ZIP code. However, to protect the confidentiality of their respondents, the NLSY does not release the actual ZIP codes. Instead, respondents who live in the same ZIP code are grouped together. In effect, Borjas’ work estimates a spatial attainment model backwards—estimating the characteristics of individuals who share the same ZIP code (see *aggregate regression* above). Friedman, Singer, Price, and Cheung (2005) use the ZIP code of intended residence for legal immigrants to

Washington D.C.. However, their data, from visa forms, have no information on race, education, income, occupational status, or household composition. Therefore, little can be concluded from their research about spatial attainment into ZIP codes in US cities. But several benefits of ZCTAs may encourage spatial attainment researchers to adopt their use.

As discussed above, Bayer, McMillan, and Reuben (2003) report large differences between spatial attainment effects assessed at the tract-level and attainment measured at the PUMA-level. ZCTAs fall between tracts and PUMAs and are more attractive to spatial attainment researchers than PUMAs. The average ZCTA in Harris County has 31,589 people. Tracts and PUMAs average 6,266 and 136,810 people, respectively. And ZCTAs do not have an artificial population minimum as do PUMAs (Census Bureau 2001).

There are many more ZCTAs than PUMAs in every metropolitan area. This is due to the smaller geographic extent of ZCTAs and their smaller populations. For example, in Harris County there are 142 valid ZCTAs and only 24 PUMAs.

Frequently, residents distinguish different areas within a metropolis by the area's ZIP code. Unlike PUMAs, residents know their own ZIP code. Also, survey respondents are more likely to share their ZIP code than their full home address. For example, in the Houston Area Survey Data around 95% of respondents offered their ZIP code (which was then translated into a ZCTA); but only 58% revealed their home address.

This may create new opportunities for spatial attainment research. Since ZIP codes are easily self-identified and readily offered to survey researchers, non-Census surveys now have a readily available spatial geography for use in spatial attainment research. Non-Census surveys often provide rich individual-level detail. Now, those investigating spatial attainment can cheaply and easily create useful aggregate level dependent variables for their estimation models. For example, now researchers can use a respondent's individual-level characteristics revealed in a non-Census survey to predict the percent white of the ZCTA in which they live.

However, there are concerns about using ZCTAs in spatial attainment models. Like PUMAs, ZCTAs are much larger than what is usually considered a neighborhood, viewing Census tracts as ostensible neighborhoods. The average ZCTA in Harris County is almost five times as populous as the average tract. Another concern is that ZCTAs do not correspond perfectly with ZIP codes (Census website). Krieger, Waterman, Chen, Soobader, Subramanian, and Carson (2002) note the institutional disconnect between ZIP codes assigned by the Postal Service and ZCTAs assigned by the Census Bureau. It is possible for new ZIP codes to be assigned or old ones retired by the US Postal Service between Censuses. That could result in respondents giving their correct ZIP code, but being placed in an incorrect ZCTA. Also, the allotment of new ZIPs in growing cities could create measurement error in the ZCTAs of growing cities but not in the ZCTAs of cities with a more stable geographic area between Censuses. Finally, like PUMAs, ZCTAs do have discontiguous areas—islands. Figure 15 demonstrates that the ZCTA numbered 77040 in northwestern Harris County is composed of separate areas.

The question arises; do these problems substantively bias spatial attainment models? Although the average ZCTA is five times larger than the average tract in Harris

County, the ZCTA is still five times smaller than the next level of geography commonly used in the literature, the PUMA. The correspondence between ZIP codes and ZCTAs is not perfect, and merits further investigation in future research. However, ZCTAs do cover every residential area in the United States. ZIP codes and ZCTAs change independently. However, each is updated frequently between Censuses, and the timing probably does not effect residential location decisions. The next ZCTA update is scheduled for 2009. New ZIP codes are added before new ZCTAs. But between 1990 and 2000 only 390 new ZIP codes were added (Krieger, et al., 2002). The current number of ZCTAs is around 32,000. Also, new areas are not usually the most populous residential areas. Finally, although ZCTA do produce occasional islands, ZCTAs rarely cross tract boundaries and are normally composed of Census blocks. Figure 2.3.3 is a map detail of the northwestern area of Harris County. It shows that the ZCTA numbered 77040 is composed of separate areas. A larger area appears at 3 o’clock; a smaller island is at 10 o’clock.

All of these concerns are legitimate. All in all, however, I view the case for the validity of the ZCTA as a potentially useful neighborhood geography for spatial attainment research to be strong. Although the area is much larger than Census tracts, it is a much more valid representation of a neighborhood than the PUMA which is mathematically created (based on population) and artificially inflated (with a minimum population threshold).

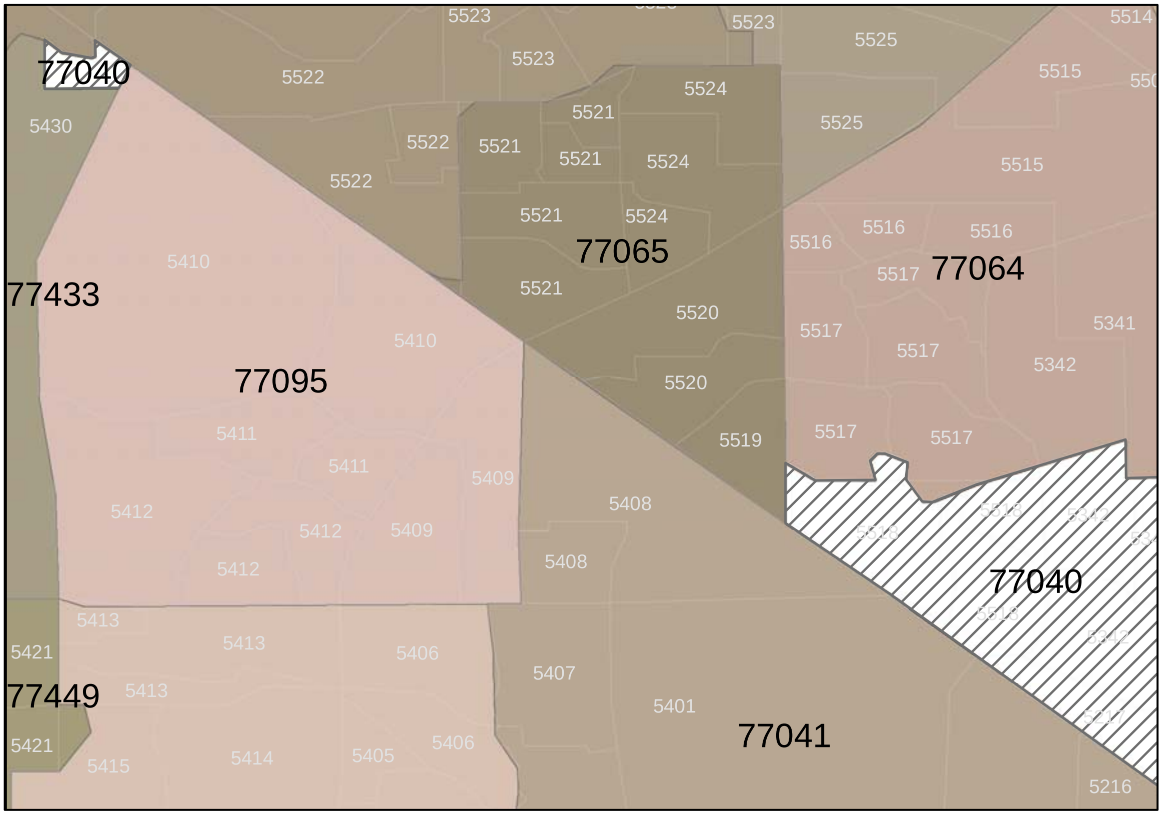

Figure 15. Example of Zip Code Tabulated Area (ZCTA) Island and Division of Census Tracts in Harris County, Texas, 2000.

Note: This map detail focuses on the northwestern area of Harris County. ZCTAs are distinguished here by color and numbered identifier in a large font. Tract boundaries are much smaller. The ZCTA numbered 77040 is composed of separate areas. ZCTAs rarely cross tract boundaries and are normally composed of Census blocks.

In the analytical portion of this section I have two objectives: 1) I investigate the role of spatial scale in spatial attainment research across four geographies—the block group, the tract, the ZCTA, and the PUMA (highlighting comparisons between ZCTA and PUMA models); and 2) I demonstrate the usefulness of incorporating ZCTA dependent variables into non-Census survey research to create spatial attainment models.

6.1. Data and Methods

I obtain my area-level dependent variables used in this section from the US Census Summary File 3 (SF3). The SF3 is based on the 1-in-6 sample of all Census respondents, also referred to as the "long form." I chose this data because of the rich detail it offers. Although I only use percent white as a dependent variable in this section, the SF3 would allow for further research using median income, percent in poverty, or housing tenure as possible spatial attainment outcomes common in the literature. The use of a sample instead of the entire population introduces a small amount of sampling error into the measurement of percent white. But the amount of error should be small and is acknowledged and well-documented in Census literature[2].

I use data for Harris County, the county which contains most of the urban population of the city of Houston, Texas. I limit the geographic extent of my data in this section to the county in order to incorporate non-Census survey data which draws

---

[2] The formula for estimating the large sample standard error of a proportion, used in Census SF3 is:

$$SE_p = \sqrt{\frac{pq}{N}}$$

where p is proportion; q = 1-p; and N is sample size.

samples exclusively from the county. These data are from the Houston Area Survey, detailed below. Future research would benefit from investigating the entire Metropolitan Statistical Area of Houston, which includes the suburban areas surrounding this growing city.

Data for calculating percent white are downloaded at the block group-, the tract-, and the ZCTA-level. As commonly found in the literature, I calculate the proportion white by using the total population count of the geography as the denominator, and the count of non-Hispanic whites as the numerator.

PUMA-level data for counts of non-Hispanic whites and total population within the area are not available for download from the Census. I use PUMA boundaries to aggregate data from block groups to create dependent variables at the PUMA-level. I employ the centroid method to assign every block group to a single PUMA. Using mapping software[3], I overlay a block group map with PUMA boundaries to select the block groups whose centroids fall within the PUMA. Once the block groups are assigned to PUMAs, the block group data are summed, or collapsed, by their PUMA identifier. This creates PUMA-level counts of non-Hispanic whites and the total population within the area.

For this research, I only make use of the PUMA boundaries. The Public Use Microdata Sample (PUMS) data are not used in this research. Although the PUMS is both useful and common for spatial attainment research, my aim in this section is to compare the area of the PUMS—i.e., the PUMA—to the ZCTA.

[3] I use the ArcView© software ArcGIS for this spatial analysis.

As part of my investigation of spatial attainment models using individual-level independent variables from survey research, I use data from the 2003 and 2005 waves of the Houston Area Survey (HAS). The HAS is a telephone survey of respondents in Harris County—Houston, Texas. Respondents are selected in a two-stage random-digit-dialing procedure (Klineberg 2002). Each household is reached by a computer-generated telephone number, and then a respondent is randomly chosen from all household residents 18 years of age or older. The survey is conducted annually. Response rates are normally around 60 percent or higher. Each year ethnic oversamples yield approximately equal numbers of white, African American, and Hispanic respondents. The survey consists of a number of socio-demographic questions as well as many attitudinal instruments.

The HAS contains geographic identifiers for each respondent. In recent waves of the survey, geographic data are collected for a respondent's block group either through answering the survey question concerning street address, or through a reverse-lookup reference based on the telephone number. Alternatively, respondents are asked to give their ZIP code. These geocodes can then be used to append block group-, tract-, ZCTA-, and PUMA-level data from the Census to the records for the individual respondents in the HAS.

6.1.1. Operationalization of Variables

Since the dependent variable, percent white in area, is a proportion bounded between 0 and 1, it is not an appropriate variable for ordinary least squares regression. I adopt a standard approach for dealing with this problem by performing a logit-transformation of neighborhood proportion white. Regression analysis of the logit-transformed variable is appropriate because it meets the linearity and distributional assumptions of linear regression. The calculation of the logit-transformation is:

$$logit\ z = \ln\left(\frac{P}{1-P}\right)$$

where P is the proportion white in the neighborhood.

I guard against allowing the denominator to approach zero or one by coding the proportion white in the neighborhood before performing the logit transformation. If the proportion white is very near zero (less than 0.005) or very near one (higher than 0.995) I recode the variables as 0.005 or 0.995, respectively. This modification of data is common when using the logit transformation. It allows me to include all white or all non-white areas for which logits would be undefined.

For independent variables, I consider the effects of ethnicity or race group, education, income, homeownership, the presence of children in the home, and two attitudinal variables: an index of the respondent’s feeling towards the local schools; and the respondent’s perception of increasing or decreasing home values.

Following spatial assimilation theory, socioeconomic status is expected to have positive effects on contact with whites and neighborhood income level across all groups. I include two measures of socioeconomic status – income and education. Both are presumed to have positive effects on neighborhood proportion white and neighborhood income level. Income is measured by assigning scores of 0-5 for six categories ranging from less than $15,000 to more than $75,000. Education is measured by assigning scores of 0-3 for four categories of less than high school; high school; some college; and professional degree. Because both variables measure the same general concept, multicollinearity is a danger. To minimize multicollinearity, but keep the individual contribution of each factor, both variables are centered (the mean is subtracted from each score).

Homeownership can also be considered a gauge of socioeconomic status – it is a major mechanism of wealth accumulation for American families (Oliver & Shapiro 1995; Shapiro 2004). But expectations regarding its possible effect are complex. On the one hand, it might be expected that minority respondents who own their own homes are of higher socioeconomic standing and would have higher levels of contact with whites and higher income neighborhoods. Alternatively, homeownership is a long-term financial commitment and the possibility exists that a neighborhood could change composition around an individual's residence (Quillian 1999). In this case, percent white might be lower for minority respondents who own instead of rent. Homeownership is operationalized as a dummy variable in the analyses.

Previous research indicates that white respondents in households with children were much more likely to prefer whiter neighborhoods than respondents without children in the home (Emerson, Yancey, and Chai 2001) and may be more likely to make this a priority in location decisions (Ellen 2000). Therefore, I include a count of the number of children in the home.

I employ measures of the respondents' perception of local schools and housing values in the neighborhood. I hypothesize that respondents who perceive better schools and increasing housing values live in whiter areas. Respondents who have negative views of schools and housing values are hypothesized to live in less white areas.

Before moving on to test the model, I would like to note that the model I have created is not complete. I selected these variables because they are commonly used in previous literature and support theoretical arguments. However, the model I present here need not be accepted as valid. Instead, this model should be viewed as a tool to test the effect of spatial scale on any spatial attainment model. This is a purely methodological exercise. I hope that future research will include many more variables and take account of new theoretical advances. Be that as it may, I argue that future research must consider the role of spatial scale.

6.1.2. Statistical Procedures

I present four analyses in this section. First, I present zero-order correlations to describe how percent white varies by geographic extent.

Second, I report two kinds of segregation scores each based on data from the four levels of geography. Segregation is measured with the index of dissimilarity (D). D ranges from 0 (complete integration) to 1.0 (complete segregation). I convert it to a percentage with a range from 0 to 100 for ease of interpretation. The interpretation of D is: *the percentage of minority households that would have to move to effect complete integration—even distribution.*

The computing formula for D between whites and African Americans would be:

$$D = \sum \left| \frac{w_i}{W} - \frac{b_i}{B} \right| \times \frac{1}{2}$$

Where W and B are the total population of whites and African American in the area; i is the count of neighborhoods in the area; and w_i and b_i are the counts of whites and African Americans in each neighborhood (Duncan and Duncan 1955; Massey and Denton 1988).

Also, I report analyses assessing the exposure/isolation or contact score using the four options for geographic scale. Contact is operationalized by computing the Lieberson's Asymmetric Exposure Index (P*). The familiar interpretation of P* is presented by Massey & Denton (1988):

> Exposure indices measure the extent to which minority and majority members must physically confront one another by virtue of sharing a common tract of residence. The degree of minority exposure may be conceptualized as the likelihood that minority and majority members share a common neighborhood (Massey and Denton 1988, p.806).

The formula for the P* exposure index is:

$${}_xP_y = \sum \left[\left(\frac{x_i}{X} \right) \bullet \left(\frac{y_i}{t_i} \right) \right]$$

Where X is the total number of members of group X in the whole city; x_i and y_i are the number of members of group x and y respectively in the ith neighborhood; and t_i is the total population of the ith neighborhood.

The index P* is the probability that a randomly selected resident (x_i) will be of the same race as another randomly selected person (y_i) from the same neighborhood (t_i) (Jaret 1995). The exposure index is computed as a probability and has a range from 0.00 to 1.00. I convert it to a percentage with a range from 0 to 100 for ease of interpretation.

It is also possible to use P* as a measure of a group's *isolation* by computing the exposure of a group with itself in a given neighborhood.

As the underlying demography of the city changes, P* changes. Therefore, interpretations of P* indexes should be compared to the total proportion of each group in the area (Jaret 1995 p. 344)For example, larger groups will exhibit larger isolation and exposure scores, irrespective of dissimilarity. Zubrinsky-Charles (2003) explains:

> Isolation is generally low for small groups but is expected to rise with increasing group size even if the group's level of segregation remains constant. Moreover, the larger the relative size of an out-group's population, the greater exposure to that group is likely to be. Both exposure and isolation are influenced by group settlement patterns (Zubrinsky-Charles 2003, p.172).

I report total group proportions as "expected P*" values with their respective P* scores.

The third analysis I report is a spatial attainment regression of percent white in the neighborhood (at four geographic levels) onto a spatial attainment model composed of independent variables from survey research provided by the HAS. Logit scores for neighborhood percent white are well-suited for statistical modeling purposes, but are less intuitive for interpretation than neighborhood proportion white. To aid in interpretation, I compute partial derivatives to obtain a more intuitive representation of each variable's effect on neighborhood proportion white. These are obtained from:

$$\textit{Partial Derivative} = bP(1-P)$$

Where *b* is the logit coefficient; and *P* is the proportion white in the neighborhood.

The interpretation of the partial derivative is straightforward[4]. It represents the change in neighborhood proportion white associated with a one unit change in the independent variable (Hanushek and Jackson 1977, p. 189). I convert this to a percentage for ease of interpretation.

I also calculate an "additive race effect" calculates the percent white of a respondent' neighborhood assuming group-specific average scores on all the independent variables. This is equivalent to undoing the logit transformation. I sum the products of the coefficients by the group means; then the sum is used as the exponent to the base constant *e* (2.71828); finally, the resultant power is divided by itself plus one. I convert the additive race effect to a percentage for ease of interpretation:

$$\textit{Additive Race Effect} \quad = \alpha + \sum \frac{e^{(b_i * \bar{x}_i)}}{1 + e^{(b_i * \bar{x}_i)}}$$

Where e is the constant 2.71828; b_i is the slope coefficient for the independent variable; $\bar{x}_i$ is the group mean; and α is the constant for the regression model.

I end this section by reporting diagnostic statistics which check the appropriateness of my regression models. I provide details of common diagnostics for multicollinearity, model specificity, overly influential observations, and heteroskedasticity.

[4] I evaluate the partial derivative for all variables at the point on the relationship where the proportion white is equal to 0.4, as found empirically and noted above.

6.2. Results

Table 13 presents the correlations and means of percent white at four levels of neighborhood geography. I report the correlation of raw percent white measured at block group-, tract-, ZCTA-, and PUMA-levels. The left-hand side of Table 13 shows the zero-order correlation matrix of percent white at various neighborhood scales across 1,906 block groups in Harris County. As expected, geographic extents that are close in scale correlate well. For example, percent white at the block group-level correlates positively and strongly with tract-level percent white (r = 0.93). The geographies that are least similar, the block group and the PUMA, have the lowest correlation (r = .60). Note here that the ZCTA-based percent white correlations are much stronger than the PUMA-based correlations for both block groups and tracts.

The rightmost column gives the means of percent white in Harris County by measuring at four geographic extents. The means generally decrease slightly as geographic extent increases. Substantively, the means are very close, only differing by 3 percentage points (ranging from 42% using block groups to 39% white using PUMAs). The mean percent white of all block groups in Harris County is expected to be very close to the mean percent white of all the PUMAs in the same county—the county is around 40% white. Even though I introduce measurement error in the mean of percent white, it is substantively still quite comparable using any of the four geographic extents. The standard deviations do change, substantively, though. Like the means, the standard deviations decrease as the geographic extent increases in size. Block group, tract, and

Table 13. Zero-Order Correlation Matrix of Percent White and Mean Scores at Various Neighborhood Scales of Aggregated Block Groups in Harris County, Texas, 2000.

	Percent White				
	Block Group	Tract	ZCTA	PUMA	Mean (Std Dev.)
Block Group	1.00				0.42 (0.32)
Tract	0.93	1.00			0.41 (0.30)
ZCTA	0.83	0.88	1.00		0.40 (0.27)
PUMA	0.61	0.65	0.73	1.00	0.40 (0.20)

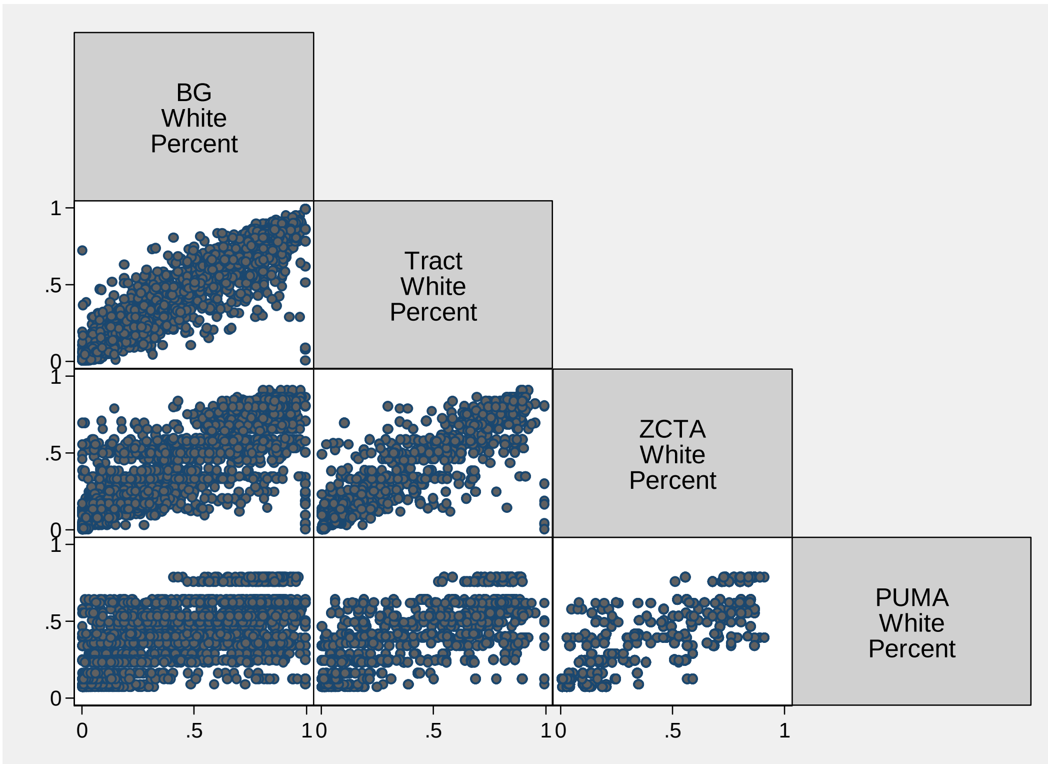

Figure 16 Zero-Order Scatterplot Matrix of Percent White and Means Scores at Various Neighborhood Scales of Aggregated Block Groups in Harris County, Texas, 2000.

ZCTA means share similar scores (ranging from 32% to 27%). The largest change in standard deviation occurs between the ZCTA and the PUMA geographies. Here, the standard deviation drops from 27% to 19% in one step between two geographies.

In addition to the means, I present a matrix scatterplot in Figure 16. Using this matrix, it is easy to visually assess the correlations between the spatial scales. Scales that are close to each other in size show steeper slopes and tighter patterns down the diagonal, reflecting the stronger correlations discussed above.

Just as correlations differ as geographic extent varies, so does segregation. To help see the declining trend, I present D scores of race/ethnic groups in a bar chart. Figure 17 discloses D scores for minority groups with non-Hispanic whites across four levels of geography. The prevailing trend is that D scores decline as geographic extent increases. For example, D scores for the African American group drops from 71.5 at the block group-level to 49.8 at the PUMA-level. The range between the three smallest levels of geography (i.e., block group, tract, and ZCTA) is around 10 points for both minority groups. The difference between PUMA-level scores and its nearest geographic level the ZCTA is larger than the range of the smaller three. Segregation measured at the PUMA-level is more than 10 points lower than at the ZCTA-level.

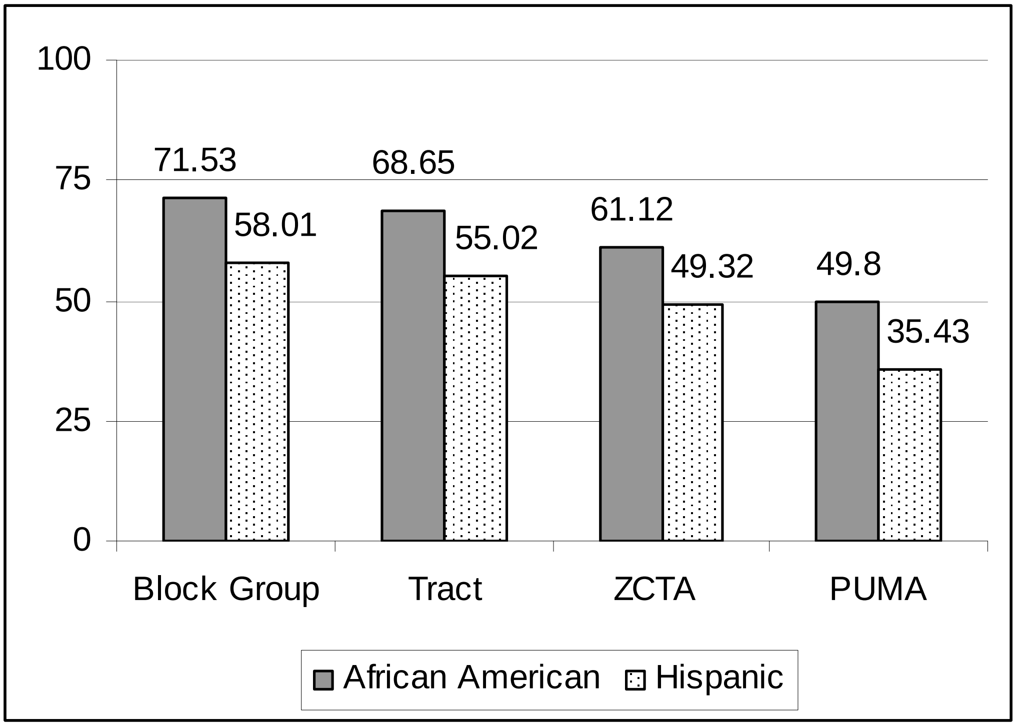

Figure 17. African American and Hispanic Dissimilarity from Non-Hispanic Whites Measured at Four Levels of Geography, in Houston, Texas, 2000.

Table 14 displays P* indexes for all of the pair-wise combinations of race/ethnic groups in the study. As expected, exposure, or out-group contact, increases as geographic extent increases. For example, African American exposure to non-Hispanic whites at the block group-level is 18; at the PUMA-level, their exposure is 29. Using the standard probabilistic interpretation of P*: an African American has an 18% chance that a randomly selected neighbor from the same tract is white; and a 29% that a randomly selected neighbor from the same PUMA is white. The same trend of increasing exposure with increasing levels of geography holds true for all other out-group combinations.

Contact scores are not symmetrical. For instance, African American exposure to non-Hispanic white is not equal to white exposure to African American. Using the score discussed above, the block group exposure for African American to non-Hispanic white is 18. But the non-Hispanic white exposure to African American is 8. An African American has an 18% chance that a randomly selected block group neighbor is white. But a non-Hispanic white only has an 8% chance that his or her neighbor would be African American. The trend of increasing exposure scores with increasing geographies applies to either side of the asymmetric diagonal.

Table 14. Contact Scores (P*) Between Non-Hispanic White, African American and Hispanic Groups across Four Geographic Scales, Harris County, Texas, 2000.

	Block Group			Tract			ZCTA**			PUMA***		
	White	African American	Hispanic	White	African American	Hispanic	White	African American	Hispanic	White	African American	Hispanic
White	64	4	21	62	9	23	57	10	26	50	14	29
African American	18	51	25	19	48	27	22	42	30	30	32	33
Hispanic	27	14	54	28	15	52	29	16	50	34	19	42
Expected Contact****	42	18	33	41	18	34	40	19	36	40	20	34

* Data for Harris County from Census 2000.

**ZCTA is the acronym for "Zip Code Tabulated Area;" a Census geography introduced in Census 2000.

***PUMA is the acronym for "Public Use Microdata Area;" a Census geography commonly used in spatial attainment research.

****Expected Contact is the P* calculation assuming even distribution based on group proportions in the area.

Isolation, on the other hand, decreases as geographic extent increases. In-group contact is greatest at the lowest levels of geography. For example, white exposure to white is 64 at the block group-level; but it is only 50 at the PUMA-level. For a non-Hispanic white in a block group, the odds are almost 2 to 1 (64%) that a randomly selected neighbor will also be white. At the PUMA-level, the odds are even (50%) that a randomly selected neighbor will be white.

In sum, larger levels of geographic aggregation provide a distorted view of inter-group association. The simple measure of percent white in the neighborhood is distorted by the size of the neighborhood used. I find that levels of percent white in PUMAs are only moderately correlated with percent white at the block group extent. Just as percent white varies by level of spatial scale, so do measures of segregation such as distribution (D) and exposure (P*). An index of dissimilarity based on block groups for Houston, Texas in 2000 indicates that more than 70 percent of African American residents would have to relocate to achieve even distribution. However, a computation of the same index, in the same city, in the same year, but based on a PUMA-level geography indicates that only 50 percent of African American residents would need to move to achieve even distribution. In the following section, I demonstrate that spatial scale also affects micro-level spatial attainment models.

Table 15 presents the descriptive statistics for a spatial attainment model based on HAS respondents in Harris County. Note that the unit of analysis switches here from block groups to individuals. The top panel introduces raw percent white scores for HAS respondents. As reported previously in the analysis of block groups, the percent white for all race/ethnic groups varies only slightly by measurement at different scales. The average HAS respondent lives in a neighborhood that is about 39% white. When partitioned by race/ethnic group, though, variation in percent white appears. Whites live in the whitest block groups (64% white); African Americans live in the least white block groups (20% white). Percent white decreases monotonically for whites as geographic scale increases from block group through PUMA-levels. Alternatively, percent white progressively increases for minorities as geographic scale increases.

The bottom panel of Table 15 presents the descriptive statistics for the independent variables used in the spatial attainment regression. I include only those observations which are not missing any variables in the list of independent variables. These are the respondents who will be included in the regression.

Interpreting Table 15 in terms of trends across race/ethnic groups, the non-Hispanic white group scores highest on measures of education and income; followed by the African American group, then the Hispanic group. Non-Hispanic whites also score highest on average homeownership rates and feelings of school quality; followed by Hispanics, then African Americans. White and African American respondents perceive their neighborhood's home values equally as generally increasing; Hispanic respondents are more likely to perceive their home values as increasing. Finally, the Hispanic group averages more children currently living in the home (1.4) than the African American group (0.9) or non-Hispanic whites (0.7).

Table 15. Variation of Percent White at Four Levels of Geography and Descriptive Statistics for Regression Factors by Race and Ethnic Group, Houston Area Survey, 2002-2005.

	White		**Black**		**Hispanic**	
	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev
Variation of Percent White						
Block Group	63.7	24.7	19.6	24.4	29.8	25.3
Tract	62.7	23.6	20.1	23.4	30.5	24.4
ZCTA	58.8	23.6	22.9	23.1	31.0	23.2
PUMA	50.7	18.4	31.4	17.7	35.0	19.9
Individual-Level Independent Variables						
Education	3.0	0.7	2.6	0.7	2.1	0.9
Income	4.7	1.4	3.5	1.7	3.3	1.5
Children at Home	0.7	1.1	0.9	1.2	1.4	1.3
Homeownership	82.5	38.0	55.2	49.7	56.9	49.6
Attitude Towards Schools	2.3	0.9	2.2	0.9	2.5	0.8
Perception of Home Value	0.5	0.5	0.5	0.5	0.5	0.5
N	504		449		376	

Table 16 shows a spatial attainment model based on individual-level data from respondents of the 2003 and 2005 waves of the HAS at four levels of geography. Spatial attainment models estimated at all four levels of geography show similar significance and direction. Education, income, and rating of schools are each significant and positively associated with percent white neighborhoods. However, the coefficients of the significant variables decline as neighborhood geography is increased. For example, at the block group-level a one category increase in education level yields an average increase of .26 in the logit of percent white holding constant the other effects of the other variables on percent white. At the PUMA-level, the same change in education yields only a .12 increase in the logit of percent white. Applying the partial derivative, these logit values amount to 6.2 and 2.9 percentage point increases, respectively, in block group percent white for each increase in educational category.

Table 16 also reveals support for the claim that models will be better fit at smaller spatial scales. I report F-test and R^2 values with spatial attainment regressions at each level of geography. With this particular model, these values are similar ar the block group and tract levels, but they drop slightly at the ZCTA level, and they drop more significantly at the PUMA level. At the block group and tract levels, the regression model explains 20% of the variance in the logit of percent white; around 18% of the variance at the ZCTA level; and only 12% at the PUMA level. This amounts to a 40% reduction in R^2 between the block group and the PUMA levels.

Table 16. Regression of Logit of Percent White on Spatial Attainment Model for All Race/Ethnic Groups in Houston, Texas by Block Group, Tract, ZCTA, and PUMA, Houston Area Survey, 2002 and 2005.

	Block Group		Tract		ZCTA		PUMA	
	Coefficient	Partial Derivative	Coefficient	Partial Derivative	Coefficient	Partial Derivative	Coefficient	Partial Derivative
Education	.26**	6.24	.24**	5.76	.19**	4.56	.12**	2.88
	(.07)		(.22)		(.05)		(.04)	
Income	.51**	12.24	.45**	10.80	.38**	9.12	.19**	4.56
	(.04)		(.03)		(.03)		(.02)	
Children at Home	-.06	-1.44	-.04	-0.96	-.03	-0.72	.004	0.1
	(.04)		(.04)		(.03)		(.02)	
Homeowner	.01	0.24	-.002	-0.05	-.12	-2.88	-.11	-2.64
	(.12)		(.11)		(.09)		(.06)	
School	.34**	8.16	.30**	7.20	.26**	6.24	.17**	4.08
	(.06)		(.05)		(.05)		(.03)	
Home Value	.04	0.96	.08	1.92	.02	0.48	.03	0.72
	(.11)		(.09)		(.08)		(.05)	
Constant	- 1.7 **		- 1.6 **		- 1.3 **		-.88	
	(.18)		(.16)		(.14)		(.09)	
Adjusted R2	.20		.20		.18		.12	
F-test	56.53		55.21		48.15		29.82	
N	1329		1329		1329		1329	

Numbers in parentheses are standard errors
* Significant at .05
** Significant at .01

Tables 17 through 19 allow comparison of spatial attainment by race/ethnic group across varying geographies. These tables apply the same spatial attainment models used above, except each model is now employed separately by race/ethnic group.

Table 17 is the spatial attainment regression for non-Hispanic white respondents to the HAS. For whites, the only variable in the model that is significant at all levels of geography is the attitude toward school. Income is significant at all levels except the PUMA level. Both factors have positive slopes with the logit of percent white. Using the partial derivative to interpret the coefficients, a one category change in perception of the school yields an 8 percentage point change in percent white of the block group, holding constant the other independent variables. A one category increase in income produces a 6 percentage point increase in percent white of neighborhood. At the ZCTA level, school rating and income generate percentage point changes of 6 and 4, respectively.

The presence of children in the home is only significant at the PUMA level. This is unexpected. It indicates that non-Hispanic whites with school-age children live in whiter PUMAs, but not necessarily whiter block groups, tracts, or ZCTAs. Education, homeownership, and perception of housing values have no significant effect on the neighborhood percent white of non-Hispanic whites.

The row titled "additive race" effect shows that a non-Hispanic white respondent, with the average characteristics of that group across all independent variables, would live in a block group that is 66% white, and a PUMA that is 50% white. This estimate is quite close to what is revealed empirically in the descriptive statistics in Table 2.3.3.

The R^2s for this spatial attainment model demonstrate that the model, though relatively weak, is quite consistent across geographies. The R^2s drop consistently as expected as the geographic extent increases. However, the range of difference is small. The amount of variance explained ranges from 10.6% at the block group-level to 9.1% at the PUMA-level.

Table 18 presents the spatial attainment model for African American respondents of the HAS. For this group, income and homeownership are significant at every level of geography. However, whereas income is positively associated with neighborhood percent white, homeownership is negatively associated. The partial derivative columns indicate that, for each increase in income category at the block group-level, there is an average 10 percentage point increase in the percent white of a group-member's neighborhood, holding constant the other independent variables. For the binary category of homeownership at the block group-level, there is an average *decrease* of 20 percentage points in the percent white for those who own their homes compared to those who rent; again, holding constant the effects of the other variables.

Table 17. Regression of Logit of Percent White on Spatial Attainment Model for Non-Hispanic White Group in Houston, Texas, by Block Group, Tract, ZCTA, and PUMA, Houston Area Survey, 2002 and 2005.

	Block Group		Tract		ZCTA		PUMA	
	Coefficient	Partial Derivative	Coefficient	Partial Derivative	Coefficient	Partial Derivative	Coefficient	Partial Derivative
Education	.06	1.44	.03	0.72	-.03	-.72	-.10	-2.4
	(.08)		(.08)		(.07)		(.05)	
Income	.23**	5.52	.20**	4.8	.19**	4.56	.05	1.2
	(.04)		(.04)		(.04)		(.03)	
Children at Home	-.01	-.24	.02	0.48	.08	1.92	.09*	2.16
	(.06)		(.05)		(.05)		(.04)	
Homeowner	.06	1.44	.08	1.92	-.06	-1.44	.02	0.48
	(.16)		(.14)		(.14)		(.10)	
School	.32**	7.68	.28**	6.72	.27**	6.48	.23**	5.52
	(.07)		(.06)		(.06)		(.04)	
Home Value	.06	1.44	.09	2.16	.07	1.68	.10	2.4
	(.12)		(.11)		(.10)		(.07)	
Constant	-.38	-9.12	- .37*	-8.88	- .41*	-9.84	- .63**	-15.12
	(.21)		(.19)		(.18)		(.13)	
Additive Race Effect	.66		.64		.59		.50	
Adjusted R2	.10		.10		.09		.08	
F-test	9.89		9.85		9.39		8.25	
N	504		504		504		504	

Numbers in parentheses are standard errors
* Significant at .05
** Significant at .01

Similar to non-Hispanic whites, attitude towards schools is significant at the block group, tract, and ZCTA levels. Each increase in category of school attitude yields around a 7 percentage point increase in block group and tract percent white for those in the African American group. However, the school factor is not a significant predictor of neighborhood percent white at the PUMA level.

Education is significant for the African American group at all levels except the ZCTA. At other geographic levels, education is positively associated with neighborhood percent white. The partial derivatives at the block group, tract, and PUMA levels are 7, 6, and 5, respectively. That is, for each increase in educational category, the average African American block group resident receives a 7 percentage point increase in percent white; holding constant the effects of the other variables on percent white.

As expected for a numerical minority, the additive race effect increases as the geographic extent increases. An African American respondent, with the average characteristics of that group across all independent variables, would live in a block group that is 8% white, and a PUMA that is 28% white. These scores are consistently and considerably lower than their respective neighborhood proportions reported in Table 15.

Finally, Table 19 shows that the R^2s drop consistently as expected as the geographic extent increases. However, similar to the R^2s reported for non-Hispanic whites, the range of the difference is small. The amount of variance explained ranges from 15% at the block group-level to 12% at the PUMA-level.

Table 18. Regression of Logit of Percent White on Spatial Attainment Model for African Americans in Houston, Texas, by Block Group, Tract, ZCTA, and PUMA, Houston Area Survey, 2002 and 2005.

	Block Group		Tract		ZCTA		PUMA	
	Coefficient	Partial Derivative	Coefficient	Partial Derivative	Coefficient	Partial Derivative	Coefficient	Partial Derivative
Education	.28**	6.72	.24*	5.76	.17	4.08	.20**	4.80
(4 categories)	(.13)		(.12)		(.10)		(.06)	
Income	.43**	10.32	.38**	9.12	.31**	7.44	.15**	3.60
(5 categories)	(.06)		(.05)		(.05)		(.03)	
Children at Home	.01	.24	.03	0.72	.00	0.01	-.01	-.24
(count 0-8)	(.08)		(.07)		(.06)		(.04)	
Homeowner	-.84**	-20.16	-.71**	-17.04	-.63**	-15.12	-.41**	-9.84
(1=yes)	(.19)		(.17)		(.15)		(.09)	
School	.29**	6.96	.29**	6.96	.23**	5.52	.07	1.68
(4 categories)	(.10)		(.09)		(.08)		(.05)	
Home Value	-.01	-.24	.004	0.1	-.1	-2.4	-.11	-2.64
(1=increasing)	(.18)		(.16)		(.14)		(.09)	
Constant	- 2.46**	-59.04	- 2.35**	-56.40	- 1.8 **	-43.2	-.78	-18.72
	(.30)		(.27)		(.24)		(.14)	
Additive Race Effect	.08		.10		.15		.28	
Adjusted R2	.14		.14		.12		.11	
F-test	13.38		12.97		10.99		10.29	
N	449		449		449		449	

Numbers in parentheses are standard errors

* Significant at .05

** Significant at .01

Table 19 is the corresponding spatial attainment regression for the Hispanic group at four levels of geography. Income is significant and positive across all four geographies. Using the partial derivative to interpret the slope coefficient, for each increase in income category at the block group-level Hispanic group members experience on average an 8 percentage point increase in percent white; holding constant the effects of the other variables. The effect decreases with each increase of geography. At the PUMA-level, this effect is around 4 percentage points.

Education is also significant and positive, at all levels except the PUMA-level. Each increase in education category results in a 6 percentage point increase in percent white at the block group-level, all else equal. The slope effects diminish as geography increases. At the ZCTA-level, again using the partial derivative, a one category increase in education yields a 5 percentage point increase in percent white.

Homeownership becomes significant at the larger geographies of ZCTA and PUMA. However it is negatively related to percent white in the neighborhood. For example, Hispanic homeowners live, on average, in ZCTAs that are 9 percent *less white* than do Hispanic renters, all else equal. At the PUMA level homeowners live in 6 percent less white PUMAs than do renters.

Unlike results reported for the non-Hispanic white and African American groups, school attitudes are never significant in the models for Hispanic spatial attainment.

As with the other minority group, the additive race effects increase as geography increases. For a member of the Hispanic group with average scores on all the independent variables, this model predicts residence in a 22 percent white block group. The same characteristics predict a 32 percent white PUMA. These predictions are consistently lower than those empirically observed in Table 15.

The R^2s decrease consistently as expected as the geographic scale increases. As noted with the previous groups, the range of difference is small. The amount of variance explained ranges from 13.5% at the block group-level to 8% at the PUMA-level.

Table 19. Regression of Logit of Percent White on Spatial Attainment Model for Hispanic Group in Houston, Texas by Block Group, Tract, ZCTA, and PUMA, Houston Area Survey, 2002 and 2005.

	Block Group		Tract		ZCTA		PUMA	
	Coefficient	Partial Derivative	Coefficient	Partial Derivative	Coefficient	Partial Derivative	Coefficient	Partial Derivative
Education	.23*	5.52	.24**	5.76	.21**	5.04	.07	1.68
	(.10)		(.09)		(.08)		(.07)	
Income	.33**	7.92	.28**	6.72	.25**	6.00	.18**	4.32
	(.06)		(.05)		(.05)		(.04)	
Children at Home	.04	0.96	.03	0.72	.02	0.48	.04	0.96
	(.06)		(.06)		(.05)		(.04)	
Homeowner	-.11	- 2.64	-.22	-5.28	- .38**	-9.12	- .25*	-6.00
	(.17)		(.15)		(.14)		(.11)	
School	.07	1.68	.04	0.96	.02	0.48	.11	2.64
	(.11)		(.09)		(.08)		(.06)	
Home Value	.15	3.6	.24	5.76	.18	4.32	.18	4.32
	(.16)		(.14)		(.13)		(.10)	
Constant	- 1.22**	-29.28	- 1.01**	-24.24	- .80**	-19.2	- .94**	-22.56
	(.29)		(.26)		(.24)		(.20)	
Additive Race Effect	.22		.24		.26		.32	
Adjusted R2	.12		.12		.11		.06	
F-test	9.61		9.50		8.45		5.32	
N	376		376		376		376	

Numbers in parentheses are standard errors
* Significant at .05
** Significant at .01

6.3. Regression Diagnostics

The preceding analyses relied on sixteen OLS regressions—composed of four groupings (three race/ethnic groups and a regression for all groups) at four levels of geography (block group, tract, ZCTA, and PUMA). Below, I report diagnostics that ensure that these regression models produce efficient, unbiased estimates. Specifically, I test for model specification, multicollinearity, skewness/kurtosis, heteroskedasticity, influential observations, and normality of residuals.

Model specification is scrutinized with the regression specification error test. The test checks for omitted variables in the regression model. Most of my models passed the specification error test. Only the regressions for the Hispanic group at the tract- and the ZCTA-levels were statistically misspecified. A poorly specified model inflates the error term, as reflected in the low R^2s for those models. Other spatial attainment models for Hispanics include variables such as English language proficiency and time in the United States (Bayer, McMillan, and Rueben 2004). Perhaps expanded models appropriately focused on Hispanic assimilation would be better specified. However, for my argument on the role of spatial scale in spatial attainment models, I am comfortable including these models with those that are correctly specified.

Multicollinearity is the result of highly correlated independent variables. When two or more variables are in reality measuring the same concept, standard errors for those variables increase. This results in not finding significance where significance actually exists (Type II error). In my model, education and income measure, conceptually, the same thing: socio-economic status. To guard against multicollinearity, I execute a common transformation on these variables by centering each on its own mean. Centering variables allows for each to remain in the model, and it does not affect the estimate of the regression coefficient. As a rough test for multicollinearity, I compute Variance Inflation Factors[5] (VIF) for all sixteen regressions. VIF scores which are above 10 are considered problematic. In my models, all VIF scores are below 10 except for two variables in the Hispanic model: the VIF for school attitudes is 10.85; and the VIF for perception of home values is 10.52. Most other VIF scores, including those for the centered education and income variables hover around 1.5.

I test for influential observations in my data with an effect diagnostic termed *DFBETA*. I calculate this statistic for each predictor in each equation. Scores of more than |1.0| are problematic. None of the factors in any of my models score greater than |1.0|.

The regressions, as presented above, do, however, violate two important assumptions of OLS regarding the residuals: heteroskedasticity and normality. I find heteroskedasticity in my error term. Seven of the 16 regressions, reveal statistically significant heteroskedastic variance. In addition, a test of the distribution of the residuals reveals that none of my regression equations present normal distributions. The non-normal distribution of residuals is confirmed with studentized residuals. Many observations, more than the 5% or so that might be expected, have studentized residuals

---

[5] The reciprocal of a VIF score is the *tolerance* score. Therefore, a VIF of 10 is equivalent to a tolerance score of 0.10. Tolerances of less than 0.10 are considered problematic (Hamilton 1992 p. 134).

greater than |2|. Violating these assumptions may bring my significance tests into question.

As an exercise to test if the assumptions of OLS are met in my model, I use robust regression to check my estimates. Robust regression does not have the assumption of normality of residuals (skew or heteroskedastic), and is not influenced by outliers (i.e., observations with studentized residuals of greater than |2|). I choose to report my standard OLS regression equations, but to compare them to robust regressions. I find that, in all sixteen regression equations, coefficients which are significant using OLS are also significant with the robust regression. Also, the slopes of OLS and robust regressions are very similar. Of the 96 coefficients (six variables in sixteen regressions), all but two OLS coefficients are less than one robust standard error from the robust coefficient. The two that do not pass this test are the school attitude variable for the non-Hispanic white group at the block group- and the tract-levels.

6.4. Overview of the Role of Spatial Scale

Although the role of scale in spatial attainment outcomes has not been addressed in previous research, my findings support the hypothesis that scale matters. My results are consistent with previously reported findings in the segregation literature (Cowgill and Cowgil 1958; Taeuber and Taueber 1965) and in broader research on spatial analysis (Openshaw 1983).

Larger areas give spatial attainment researchers less variation, lower segregation scores, lower contact scores, and less explained variation for their regression models.

When all race/ethnic groups are pooled, BG, TR and ZCTA models explain 17-20% of the variation in percent white neighborhood. PUMAs explain much less of the variation in percent white with an R^2 of 12%. This finding supports the claim that PUMAs are inappropriate for spatial attainment modeling. It also lends support to the argument that ZCTAs may be appropriate for modeling spatial attainment, since they seem to align well with outcomes of models based on tracts.

I find that the same independent variables are consistently significant regardless of scale. In addition to consistent significance, I find that the direction and magnitude of the slope coefficients are consistent across scale. These results lend support to the continued use of PUMAs in spatial attainment models. Since the risk of using a large area is a false negative (Type II error), the consistency of significance is an important finding for researchers who must rely on the PUMS and PUMA data.

This section of my dissertation has demonstrated the role of spatial scale in spatial attainment models. Prior studies of spatial attainment have relied on larger areas which poorly fit the concept of 'neighborhood'. Other studies have suggested the use of restricted Census tabulations which offer the smallest of neighborhoods (Bayer, McMillan, and Rueben 2004). But these datasets are expensive and difficult to access. I compare spatial attainment models at four levels of geography common in the literature. I find that scale matters, all else equal. Stronger coefficient slopes and significance tests are found consistently at smaller geographies. The amount of variance explained by the spatial attainment models presented in this section act as expected: smaller geographies produce larger R^2s. Future spatial attainment research, which has up to this time used various geographies inconsistently, should be mindful of the size of the neighborhood used as dependent variable.

This section also introduces the use of ZCTAs as a geography in spatial attainment research. Since ZCTAs are smaller in size than PUMAs, ZCTA-based models are consistently more robust than models estimated at the PUMA-level. I find that ZCTA models yield results comparable to those obtained with Census tracts. However, an important advantage of ZCTAs over Census tracts or block groups is that, in non-Census survey research, respondents are much more likely to give their ZIP code than their full street address. The availability of ZCTA-level data opens the door for further spatial attainment research using survey instruments to estimate ZIP code level outcomes.

7. SUMMARY AND CONCLUSIONS

7.1. Summary

In my dissertation I extend the current literature in spatial attainment in four important ways: 1) I show that aggregate regression is an inappropriate method for assessing spatial attainment; 2) I investigate trends in inter-group contact by educational level; 3) I apply existing methods in a new ways to uncover trends in spatial attainment over time; and 4) I show that geographic scale must be considered in spatial attainment models.

7.2. Conclusions

My dissertation yields numerous conclusions for spatial attainment research. First, I suggest that previous research on spatial attainment using aggregate regression be reevaluated in light of the weaknesses I demonstrate regarding that method. Second, is show the index of exposure/contact (P*) is useful for describing spatial attainment patterns because of its intuitive and easy to understand individual-level interpretation. Previously it has been applied to exposure between race/ethnic groups. I recommend its use on more nuanced groupings—such as exposure between education or income categories. Third, I show that the method of utilizing Census cross-tabulations to create individual-level datasets appropriate for spatial attainment modeling can be used to good effect to chart changes in spatial attainment for groups over time. To my knowledge, use of this approach to investigate trends has not been published previously. And fourth, I conclude that spatial attainment models must take into account the size of the neighborhood used in measuring the dependent variable. Previous research has too often relied on very large areas of over 100,000 people. I show that this “stacks the deck” against finding spatial attainment effects that are substantively important and statistically significant. Drawing on data for neighborhoods defined at large spatial scales attenuates spatial attainment effects. In contrast, drawing on data for neighborhoods defined at smaller spatial scales yields spatial attainment effects that are much larger in magnitude, carry greater substantive importance, and more readily attain statistical significance.

REFERENCES

Adleman, Robert. 2005. "The Roles of Race, Class, and Residential Preferences in the Neighborhood Racial Composition of Middle-Class Blacks and Whites." *Social Science Quarterly* 86:209-228.

Alba, Richard and Victor Lee. 2003. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge, MA: Harvard University Press.

Alba, Richard and John Logan. 1992. "Analyzing Locational Attainments: Constructing Individual-Level Regression Models Using Aggregate Data." *Sociological Methods and Research* 20:367-397.

—. 1993. "Minority Proximity to Whites in Suburbs: An Individual Level of Analysis of Segregation." *American Journal of Sociology*, 98:1388-1427.

Alba, Richard and John R. Logan. 1991. "Variations on Two Themes: Racial and Ethnic Patterns in the Attainment of Suburban Residence." *Demography* 28:431-53.

Alba, Richard, John Logan, and B. Stults. 2000. "The Changing Neighborhood Contexts of the Immigrant Metropolis." *Social Forces* 79:587-621.

Bayer, Patrick, Robert McMillan, and Kim Rueben. 2004. "What Drives Racial Segregation? New Evidence Using Census Microdata." *Journal of Urban Economics* 56:514–535.

Blau, Peter Michael and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York,: Wiley.

Borjas, George J. 1998. "To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation." *Journal of Urban Economics* 44:228-253.

Bureau, Census. 2001. "ZIP Code Tabulation Areas (ZCTAs) Frequently Asked Questions." vol. 2006.

Cowgill, D. and Mary Cowgil. 1958. "An Index of Segregation Based on Block Statistics." *American Sociological Review* 16:825-831.

Crowder, Kyle. 2001. "Racial Stratification in the Actuation of Mobility Expectations: Microlevel Impacts of Racially Restrictive Housing Markets." *Social Forces* 79:1377-1396.

Duncan, Otis Dudley and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indices." *American Sociological Review* 20:210-217.

Ellen, I. G. 2000. *Sharing America's Neighborhoods: The Prospects for Stable Racial Integration*. Cambridge, MA: Harvard University Press.

Emerson, M, G. Yancey, and K. J. Chai. 2001. "Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans." *American Sociological Review* 66:922-935.

Farley, Reynolds and William H. Frey. 1994. "Changes in the Segregation of Whites from Blacks during the 1980s: Small Steps towards a More Racially Integrated Society." *American Sociological Review* 59:23-45.

Fossett, Mark A. and Cynthia M. Cready. 1998. "Racial and Ethnic Differentiation and Inequality." 315-355 in *Continuities in Human Ecology*, edited by M. Micklin and D. Poston. New York: Plenum Press.

Freeman, Lance. 2000. "Minority Housing Segregation A Test of Three Perspectives." *Journal of Urban Affairs* 22:15-35.

—. 2002. "Does Spatial Assimilation Work for Black Immigrants in the US?" *Urban Studies* 39:1983-2003.

Glaeser, Edward L. and Jacob L. Vigdor. 2003. "Racial Segregation: Promising News." Pp. 211-234 in *Redefining Urban & Suburban America*, vol. 1, edited by B. Katz and R. E. Lang. Washington D.C.: The Brookings Institution.

Gordon, Milton M. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origin*. New York: Oxford University Press.

Gross, A. and D.S. Massey. 1991. "Spatial Assimilation Models: A Micro-Macro Comparison." *Social Science Quarterly* 72:347-360.

Hanushek, Eric Alan and John Edgar Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic Press.

Harsman, Bjorn and John M. Quigley. 1995. "The Spatial Segregation of Ethnic and Demographic Groups: Comparative Evidence from Stockholm and San Francisco." *Journal of Urban Economics* 37:1-16.

Hauser, Robert M. and David L. Featherman. 1977. *The Process of Stratification: Trends and Analyses*. New York: Academic Press.

Jaret, Charles. 1995. *Contemporary Racial and Ethnic Relations*. New York: Harper-Collins.

Langbein, Laura and Allan Lichtman. 1978. *Ecological Inference*, vol. 10. Thousand Oaks: Sage Publications.

Lechner, Frank J. 1991. "Simmel on Social Space." *Theory, Culture & Society* 8:195-201.

Lieberson, Stanley. 1963. *Ethnic Patterns in American Cities*. New York: Free Press of Glencoe.

—. 1980. *A Piece of the Pie: Blacks and White Immigrants Since 1880*. Berkeley: University of California Press.

Logan, John R. 2003. "Ethnic Diversity Grows, Neighborhood Interaction Lags." Pp. 235-256 in *Redefining Urban & Suburban America*, vol. 1, edited by B. Katz and R. E. Lang. Washington D.C.: The Brookings Institution.

Logan, John R. , Richard D. Alba, Tom McNulty, and Brian Fisher. 1996. "Making a Place in the Metropolis: Locational Attainment in Cities and Suburbs." *Demography* 33:443-53.

Logan, John R., Brian J. Stults, and Reynolds Farley. 2004. "Segregation of Minorities in the Metropolis: Two Decades of Change." *Demography* 41:1-22.

Massey, D.S. and Nancy Denton. 1993a. *American Apartheid: Segregation and the Making of the American Underclass*. Cambridge: Harvard University Press.

Massey, Douglas, Gretchen Condran, and Nancy Denton. 1987. "The Effect of Residential Segregation on Black Social and Economic Well-Being." *Social Forces* 66:29-56.

Massey, Douglas and Nancy Denton. 1985. "Spatial Assimilation as a Socioeconomic Outcome." *American Sociological Review* 50:94-106.

Massey, Douglas and M.J. Fischer. 1999. "Does Rising Income Bring Integration? New Results for Blacks, Hispanics, and Asians in 1990." *Social Science Research,* 28 316-26.

Massey, Douglas and Eric Fong. 1990. "Segregation and Neighborhood Quality: Blacks, Hispanics, and Asians in the San Francisco Metropolitan Area." *Social Forces* 69:15-32.

Massey, Douglas and B.P. Mullan. 1984. "Process of Hispanic and Black Spatial Assimilation." *American Journal of Sociology,* 89: 836-73.

Massey, Douglas S. and A.B. Gross. 1991. "Explaining Trends in Racial Segregation,1970-1980." *Urban Affairs Quarterly* 27:13-35.

Massey, Douglas S. and Nancy A. Denton. 1987. "Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970–1980." *American Sociological Review* 52:802-25.

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

—. 1993b. *American Apartheid : Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Miller, V. and John M. Quigley. 1990. "Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area." *Urban Studies* 27:3-21.

Mullan, Brendan P. and D.S. Massey. 1984. "Processes of Hispanic and Black Spatial Assimilation." *American Journal of Sociology* 89:836-873.

Openshaw, Stan. 1983. *The Modifiable Areal Unit Problem*. Norwick, England: Geo Books.

Park, Robert Ezra, Ernest Watson Burgess, Roderick Duncan McKenzie, and Louis Wirth. 1925. *The City*. Chicago, IL: Univ. of Chicago Press.

Portes, Alejandro and Rubâen G. Rumbaut. 2001. *Legacies: The Story of the Immigrant Second Generation*. Berkeley: University of California Press.

Quillian, Lincoln. 1999. "Migration Patterns and the Growth of High-Poverty Neighborhoods, 1970-1990." *American Journal of Sociology*, 105:1-37.

Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15:351-357.

Roof, Wade and Thomas Van Valey. 1972. "Residential Segregation and Social Differentiation in American Urban Areas." *Social Forces* 51:87-91.

St. John, Craig and Robert Clymer. 2000. "Racial Residential Segregation by Level of Socioeconomic Status." *Social Science Quarterly* 81:701-716.

Taeuber, Karl and Alma Taeuber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change*: Aldine.

Taeuber, Karl and A. F. Taueber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change*. Chicago: Aldine Publishing Company.

Tolnay, Stewart. 2003. "The African American "Great Migration" and Beyond." *Annual Review of Sociology* 29:209-232.

Van Valey, Thomas and Wade Roof. 1976a. "Measuring Residential Segregation in American Cities: Problems of Intercity Comparison." *Urban Affairs Quarterly* 11:453-468.

—. 1976b. "Measuring Residential Segregation in American Cities: Problems of Intercity Comparison." *Urban Affairs Quarterly* 11:453-468.

Waters, Mary C. 1999. *Black Identities : West Indian Immigrant Dreams and American Realities*. Cambridge, MA: Harvard University Press.

Wright, Richard, Mark Ellis, and Virginia Parks. 2005. "Re-Placing Whiteness in Spatial Assimilation Research." *City and Community* 4:111-135.

Yinger, Milton J. 1981. "Toward a Theory of Assimilation and Dissimilation." *Ethnic and Racial Studies* 4:249-264.

—. 1994. *Ethnicity: Source of Strength? Source of Conflict?* Albany, NY: State University of New York Press.

Zubrinsky-Charles, Camille. 2003. "The Dynamics of Racial Residential Segregation." *Annual Review of Sociology*, 29 167-207.

APPENDIX A. COMPARISON OF METHODS FOR ESTIMATING SPATIAL ATTAINMENT EFFECTS

Spatial attainment theory is assessed via a particular set of methods wherein neighborhood outcomes are modeled as individual status attainments. The idea is simple and straight-forward; but in practice, the measurement of spatial attainment is complicated by data limitations and other considerations. In this appendix I review methodologies commonly employed to assess spatial attainment in the research literature. I describe and evaluate each method in terms of its strengths and limitations. I note the datasets that fulfill the needs of each method. And finally, I select a research exemplar that pairs methodology and data from the literature to illustrate research questions addressed, findings reported, and methodological limitations confronted.

As an aid in evaluating each data set and appropriate methodology, I describe the "ideal" set of data and methods to establish a benchmark against which I will evaluate the strengths and weaknesses of available data and prevailing methods of estimating spatial attainment models. Ideally, a spatial assimilation dataset would have the following characteristics:

- An appropriate sample universe—individuals, families, etc.
- Standard socio-demographic data for each individual; including, but not limited to Census items
- Non-Census items such as attitudinal measures towards race found in the Multi-City Study of Urban Inequality)

- Large sample size or full census
- Geographic identifiers to permit location of individual cases with high spatial resolution
- City identifier to allow single city and inter-city comparison

Data sets with these characteristics would permit investigations to estimate very sophisticated individual-level spatial attainment analyses. This ideal has yet to be achieved. But several approaches have been developed utilizing differing combinations of datasets and methods have partially approximated the ideal. I now discuss the strengths and limitations of these approaches.

The most direct method of assessing dynamics of spatial attainment is to regress dependent variables for neighborhood outcomes (e.g., percent white, or median income) on the independent variables (e.g., race, education, marital status) using individual-level data. Studies using restricted-access micro-Census data (e.g., Bayer, McMillan, and Rueben 2004) adopts just such a methodology. Using this data, the sample universe can be specified as desired (e.g., individuals, households, etc.). The dataset assures large sample sizes because all long-form sample records are available. Also, the dataset makes neighborhood information available at small spatial scale for every individual case—allowing for complex model specifications from any set of individual-level Census questions. The major methodological limitation to this approach is that researchers are constrained to items measured by the Census. Non-Census items such as neighborhood racial preferences or an individual’s wealth or net worth are not available.

Although the methodology is attractive—indeed, of all the methods reviewed here it comes closest to approximating the ideal—however, it is difficult to execute. In order to secure the confidentiality of its respondents, the Census Bureau places severe restrictions on access to these data. Access is strictly limited to on-site analysis at special, restricted access Census facilities. Additionally, gaining access to these facilities is prohibitively expensive and requires large-scale funding that precludes routine use of this resource.

Bayer, et al. (2004) provides a research exemplar that utilizes this dataset for assessing spatial attainment. Bayer and associates assess the effects of education, income, language, and immigration status on the average exposure within and between racial groups in San Francisco. The authors report that individual characteristics reported in the micro-Census data explain a substantial portion of racial “segregation” (which they define as in-group contact)—95% for Hispanics, 50% for Asians, and 30% for both Blacks and Whites. These findings contrast sharply with those of previous analyses not based on restricted-access micro-Census data (Borjas 1998; Harsman 1995; Miller 1990). These other studies reported much weaker spatial attainment effects and suggested that the impact of spatial attainment on segregation was much more limited than the impact found by Bayer and colleagues. The difference results primarily from smaller spatial scale. Bayer and colleagues draw upon smaller neighborhood geography to better capture differences in neighborhood outcomes across individuals and groups.

One limitation of this study is that it involves a single metropolitan area for the study—namely, San Francisco. San Francisco has a unique role in the development of the Asian American community on the West coast and it has one of the largest Asian populations of metropolitan city in the U.S.. Accordingly, it is an important case, nevertheless, segregation patterns in San Francisco are not necessarily representative of other cities in California or the nation—particularly if contrasted with traditional black-white cities such as Detroit or Birmingham.

Overall, the prohibitive expense and demands of using the restricted access data account for the reason this approach is not used more often.

Other Census datasets, such as the 1970 Neighborhood Characteristics File, provide individual-level data with neighborhood-level variables in a much more limited way. These data are more readily available to researchers because they are distributed as public access datasets. Like the restricted-access data, special Census tabulations include many individual-level variables for the construction of spatial attainment regressions. The dataset offers large samples which permit investigation of the attainment of numerically small ethnic and racial groups.

Unfortunately, the data have many limitations. First, the Neighborhood Characteristics File of 1970 and the special tabulation of the Public Use Microdata Sample (PUMS-F) of 1980 were both one-of-a-kind tabulations offered by the Census. This precludes the investigation of spatial attainment trends. Second, neighborhood outcomes are limited to less than 10 variables—thus limiting researchers to broad operationalizations of spatial attainment. Third, there is no metropolitan identifier associated with the individual’s case. This limitation prohibits researchers from estimating analyses by city, thus ruling out the possibility of making inter-city comparisons of spatial attainment.

Consider the PUMS-F release for the 1980 Census. This dataset was created in 1980 after the discontinuation of the Neighborhood Characteristics File. The PUMS-F appends selected neighborhood characteristics to individual cases from the Census. To protect confidentiality, a small amount of random error is added to the neighborhood data. Since it is constructed from Census data, large samples of racial and ethnic groups are available.

A research exemplar in the field of spatial attainment that draws on this type of data source is found in Gross and Massey (1991) Their work compares results from aggregate regressions with results from micro-level analyses. They also investigate trends in spatial attainment between 1970 and 1980 by using both the Neighborhood Characteristics file and the PUMS-F to estimate micro models of spatial attainment in 1970 and 1980, respectively. The authors report strong support for spatial assimilation and note evidence of residential convergence in the assimilation of African Americans between 1970 and 1980. The problem, however, is that the results that they apply at a national level. Obviously, this is far from ideal for spatial attainment models, especially since no variables about metropolitan area are included in the dataset. This, plus the fact that comparable files were not released in 1990 and 2000, accounts for why this method has not been used more often.

Spatial attainment models also have been estimated using large scale surveys which include measures of neighborhood outcomes for respondents. This approach allows for non-Census items, such as racial attitudes and residential preferences, to be included in the model in addition to basic socio-demographic characteristics. As a result, even more nuanced models may be estimated than those based solely on Census data. Consistent survey implementation across multiple cities allows for inter-city comparisons. Large samples ensure significance of indicators in the models and the reliability of coefficients.

There are limitations to this approach. Surveys are never as comprehensive as the Census and thus have smaller sample sizes. This makes it difficult to estimate spatial attainment analyses for smaller racial and ethnic groups such as Asians and Hispanics. Also, in contrast to the collection of Census data, large scale surveys are not regularly scheduled. Survey preparation and implementation normally take years to complete. Consequently, large scale representative neighborhood surveys with multiple cities are rarely repeated by the original researchers or replicated by others.

The Multi-City Study of Urban Inequality (MCSUI) is a prime example of this type of large scale neighborhood survey. The survey was conducted in four major US cities (Los Angeles, Atlanta, Boston, and Detroit) from 1992-1994. More than 8,500 interviews were recorded providing large, weighted samples from each locality. Although a geographic identifier for each respondent is not normally available, researchers may sometimes gain access to the identifiers needed to connect neighborhood-level data with individual-level cases—after submitting a confidentiality agreement. This dataset is notable for spatial attainment researchers because it includes a set of questions on racial attitudes and residential preferences not found in Census datasets.

Freeman (2000) and Adleman (2005) use data from the MCSUI in estimating spatial attainment models for racial and ethnic groups. Both works link the individual-level data files with block-group Census data. Taking advantage of the preference instrument in the MCSUI, these researchers address the role of racial preferences in spatial attainment dynamics. The main dependent variable in both studies is percent non-Hispanic white in the block-group. Adleman also investigates percent non-Hispanic black. In addition to the variables normally included in micro-level spatial attainment models, these researchers are able to include wealth, English language ability, and foreign born status. Both studies find support for the inclusion of preferences in spatial attainment models. Adelman reports that African Americans who express the preference for white neighborhoods are able to translate that desire into whiter neighborhoods.

Two major reasons account for why the data are not used more widely. One is that they are limited to only four U.S. cities. The other is that access is no longer granted and the geographic identifiers are no longer in the public domain.

Another method of estimating spatial attainment models relies solely on aggregate-level data. The aggregate method was instrumental in the early development of the literature on spatial attainment and was adopted because of a paucity of datasets that consider ecological outcomes for individuals. The discontinuation of the Neighborhood Characteristics File after the 1970 release constrained researchers to use aggregate data in both their dependent and independent variables. Although this method has the advantage of being feasible because the data are readily available, there are many serious drawbacks to the approach, including ecological inference, aggregation bias, and spatial autocorrelation.

Spatial attainment theory and research methodology is an extension of status attainment theory and methods which are framed at the individual-level (Blau and Duncan 1967; Massey and Denton 1985). However, with the lack of viable data sources that connected individual-level characteristics with neighborhood-level outcomes, researchers were forced to consider the option of using aggregate regressions (see Gross and Massey 1991; Massey, Condran, and Denton 1987; Massey and Fong 1990). For example, Massey and Mullan (1984) used the average SES of minorities within a tract to predict the proportion white within that same tract. Unfortunately, this raises concerns about the aggregate fallacy—estimating individual outcomes from aggregate-level data. One expected negative consequence—which I document in Section 3 of this dissertation—is that of aggregation bias which has the result of grossly exaggerating spatial attainment effects, making them invalid for assessing spatial attainment processes.

The only advantage of this approach is its feasibility based on ready availability of data. The disadvantages of this method, however, strongly suggest that the method is inappropriate for estimating spatial attainment and that results obtained should be considered suspect until verified by other means

Alba and Logan (1992) introduced a methodological technique that uses multiple datasets to build covariance matrices which can be used to estimate multivariate spatial attainment regressions. The technique draws on two kinds of data sets—individual-level datasets with no neighborhood-level information and summary file datasets with neighborhood information and limited individual-level information. The “blended” method they developed yields a true individual-level covariance matrix which can be used to estimate individual-level multivariate spatial attainment regressions which could not otherwise be estimated using only PUMS or Summary File data. As with other spatial attainment models, regressions assessing spatial attainment effects are estimated from a covariance matrix for both individual- and neighborhood-level variables.

The key here, however, is that the elements of the covariance matrix are obtained from multiple datasets rather than a single one. Covariances for neighborhood characteristics and covariances between neighborhood characteristics and individual characteristics are obtained from summary file data. Covariances among individual characteristics are obtained from micro-file data. The two sets of covariance elements are combined to form the full covariance matrix needed to estimate the regression of neighborhood-level dependent variables on individual characteristics. The key insight of the approach is that, although neither the micro data nor the aggregate data are sufficient by themselves, to build the needed covariance matrix, they can be combined to obtain the information necessary to estimate the regressions.

To assist in discussing this complicated strategy, Figures 18, 19, and 20 visually represent the combination of separate data sources into a blended dataset suitable for modeling spatial attainment. Each table presents a hypothetical neighborhood outcome variable Y (e.g., percent white in neighborhood) and three individual-level independent variables labeled X_1, X_2, and X_3 (e.g., education, income, age, etc.) All covariance combinations between variables are necessary to construct a regression equation estimating spatial attainment. Figure 18 shows a checkmark for covariances available directly from summary file data. The weakness in relying solely on summary file data is that the covariances between the independent variables is unavailable. Figure 19 shows checkmarks for covariances found in PUMS data. The disadvantage to only using PUMS data is that the covariances for the neighborhood level data are unavailable. Figure 20 shows the combination of the two datasets to produce an appropriate matrix for regression estimation. Covariances from the two data sources are complimentary with summary file data providing the covariances at the neighborhood-level and PUMS data providing individual-level covariances. Covariances from equivalent variables (e.g., X_1 to X_1) can be obtained from either data source.

The crucial insight of this method is that spatial attainment regressions can be estimated from covariance matrices which are constructed from multiple data sources.

	Y	X_1	X_2	X_3
Y	✓	✓	✓	✓
X1		✓	No	No
X2			✓	No
X3				✓

Figure 18. Incomplete Covariance Matrix from Summary File Data Only.

	Y	X_1	X_2	X_3
Y	No	No	No	No
X1		✓	✓	✓
X2			✓	✓
X3				✓

Figure 19. Incomplete Covariance Matrix from PUMS Data Only.

	Y	X_1	X_2	X_3
Y	SF	SF	SF	SF
X1		SF/ PUMS	PUMS	PUMS
X2			SF/ PUMS	PUMS
X3				SF/ PUMS

Figure 20. Complete Covariance Matrix from Combination of SF and PUMS.

Alba and Logan note four necessary conditions for the method to be used (1992: p. 372)

1. The aggregate characteristics are known for each tract;
2. Individual-level characteristics are known for a sample of individuals;
3. The key individual-level characteristics are tabulated by tract. And, if spatial attainment equations are estimated separately by race, the individual characteristics must be tabulated by race at the tract level;
4. The sources of both individual data and the tract-level tabulations involve the same population.

There are many advantages to this method. It allows ordinary least squares (OLS) regression to be performed using two datasets that separately do not fully connect individual data and neighborhood data. The data are easily accessible to researchers. The method assures large samples sizes, and thus, the opportunity for research focusing on smaller racial and ethnic groups. Also, neighborhood-level data can be identified for individual cities, which permits inter-city comparison. Finally, the technique can be used on data available from 1980 and 1990 to investigate trends in spatial attainment.

Unfortunately, there are important limitations of this technique. Researchers are constrained by the sample universe and individual characteristics tabulated at the tract level. Aggregated tables are sometimes defined by the individual and sometimes by the household.

> A disadvantage is that our approach does not allow us the control over the precise definition of micro-level variables that can be attained with cross-

> level data…; our variables are limited by the tabulations available in Census summary tape files (Alba, Logan, and Stults 2000: p.596).

In addition to the inability to select the sample universe, the individual- and the aggregate-level data draw on different samples (e.g., the PUMS one-percent sample and the summary file one-in-six sample). Consequently, all covariances are not estimated from a common database of individual records. Instead, the different covariances are estimated from different samples and are pieced together to form the full covariance matrix needed to estimate the model. Other limitations include the fact that group differences in spatial attainment effects (i.e., interracial contact) can only be evaluated for variables which are cross-tabulated by race (point 3 above). So in the example matrix above, income could not be added as a control unless a table existed in the micro data that was specified as race by education status by income category.

An example of a study that utilizes this spatial attainment technique is Alba, Logan and Stults (2000).This work investigates the locational attainment of whites, blacks, Hispanics, and Asians in five US cities that receive a disproportionate share of immigration. Also, this study of spatial attainment in 1990 is directly comparable to Logan, Alba, Mcnulty and Fischer’s (1996) work that assesses attainment in 1980 with the same method and data sources. Both studies report strong support for the spatial assimilation model, using percent white in the tract and median tract income as the neighborhood-level outcomes. Both investigations revealed significant gaps between white and black intercepts in locational attainment regressions, thus documenting race differences in spatial attainment.

The final method I consider in this review is the method I ultimately used in the analytical sections of this dissertation. The simple spatial attainment model draws on Census Summary File data that cross-tabulate individual characteristics by geographic areas and reorganizes it in a format that permits the estimation of individual-level spatial attainment analyses. This is accomplished by treating the cells of the cross-tabulation as data points, coding them for the appropriate values of neighborhood outcomes and individual-level independent variables, and performing weighted regressions in which the cell frequencies are applied as weights to estimate the correct individual-level parameters. For example, the tabulation of education by block group performed separately by race provides an individual-level tabulation.

The main limitation to this approach is that individual-level control variables cannot normally be included. Thus, complicated multivariate models are possible only if the individual characteristics in question (e.g., education and age) are cross-tabulated by each other at the tract level. In addition, it relies entirely on tract- or block-group-level tabulations to estimate models of spatial attainment. This constrains research to relationships between variables identified by the Census. Another limitation to this methodology is the inability to control the sample. Prison and military populations cannot be separated out of the sample universe and may attenuate patterns of segregation observed in other parts of a city. A final drawback for researchers using this method is the limitation of the Census tables themselves. Large, pan-ethnic groupings are available (e.g., Hispanic and Asian) but more nuanced racial and ethnic groups (e.g., Cuban and South Korean) are not presented. Once again, the main limitation of the "simple" spatial attainment model is the inability to include multiple control variables commonly used by researchers.

The key advantage is that it does permit the estimation of true, individual-level spatial attainment models across many time points (e.g., 1970, 1980, 1990, and 2000) and, potentially, many metropolitan areas (although I do not explore this possibility in my present study). This method stands as an alternative to the Alba and Logan “blended” method. The trade-off is the ability to do over-time analyses that are not possible with the “blended” method.

Because the simple spatial attainment model is an extension of, yet distinct from, the blended method, I present Figure 21 to highlight the contrasts and comparisons.

Blended	Simple Spatial Attainment Model
Estimates of Individual Covariances	
Covariances may be estimated from samples of different size (PUMS = 1/100 or SF = 1/6)	Covariances are estimated from a consistent sample size (SF=1/6).
Covariances may be estimated from different sample universes	Covariances are estimated from a consistent sample universe.
Sample size for some covariances may be small (PUMS = 1/100)	Sample size for all covariances is always large (SF = 1/6)
Number of Individual Characteristics in Model	
Based on the number of neighborhood-level two-way tabulations e.g., education by race, age by race, income by race, etc.	Based on number of individual variables in a single neighborhood-level crosstabulation e.g., education by race, education by race by gender

Figure 21. Comparison of Blended and Simple Spatial Attainment Models

VITA

WARREN WAREN
Department of Sociology
Loyola University New Orleans
6363 St. Charles Avenue 70115
Monroe Hall, Room 537, Campus Box 30

EDUCATION

PhD	December 2008	Sociology	Texas A&M University, College Station, Texas
MA	May 2003	Sociology	University of Arkansas, Fayetteville, Arkansas
	Master's Thesis: "Social Capital and Hispanic Integration in Springdale, Arkansas: An Exploratory Study." University of Arkansas, Fayetteville.		
BA	May 1993	Sociology	Northeastern State University, Tahlequah, Oklahoma