

Separating the Bulls from the Herd:
Race Distribution and Median Household Income in Durham County

Ryan Mattison¹
Duke University

¹ Ryan Mattison is a junior at Duke University pursuing a Bachelor of Science in Economics degree. His email is mm31@duke.edu. He would like to thank Dr. Charles Becker, whose guidance was instrumental to the production of this paper.

ABSTRACT

This report analyzes U.S. Census data from 1990 and 2000 to find a correlative relationship between race and income in Durham County, North Carolina. This analysis is then extended onto three cities similar to Durham, the results of which are used as a comparison for the base city. Lastly, changes in the correlative relationships from 1990 to 2000 are assessed, providing a dynamic view of Durham in relation to the three other cities.

TABLE OF CONTENTS

1. Introduction.....	pg. 3
2. Background and Data Composition.....	pg. 3
3. Data Issues.....	pg. 4
4. Calculations and Figures.....	pg. 5
5. Results.....	pg. 8
6. Conclusion.....	pg. 11
Work Cited.....	pg. 12
Figure 1, 2.....	pg. 13
Figure 3, 4.....	pg. 14
Figure 5.....	pg. 15
Figure 6.....	pg. 16
Figure 7, 8.....	pg. 17

Figure 9, 10.....
pg. 18

Figure 11, 12.....
pg. 19

Figure 13, 14.....
pg. 20

Figure 15, 16.....
pg. 21

1. Introduction

In this nation's mind, there exists an assumed correlation between average income and race distribution for the standard American neighborhood. The thought of a rich neighborhood consisting predominately of whites and a poor neighborhood consisting predominately of blacks feels as natural as the association between a day at a baseball game and a hot-dog. Yet not all anecdotal deductions pass as law, even if they appear true. Like Jackie Robinson defying the assumed definition of a professional baseball player, anecdotal assumptions on an urban level must be likewise reevaluated on an item by item basis. The best place to look for an urban Jackie Robinson would be a city that breaks other popular assumptions. Durham, North Carolina is one such city.

To Durham, a growing, industrial-aged city in the Southeast, common urban patterns hold less weight than its counterparts. For instance, the high levels of "white flight" seen around the nation in crime-filled cities are not obviously seen in Durham, and unlike the familiar, constricted range of house prices in most neighborhoods, Durham appears to maintain an unusually wide variance of values (Becker). Yet, is this anomalous behavior universal? To what degree does Durham follow the residential axiom of "rich whites, poor blacks" seen throughout the U.S., and does it appear to be moving closer to or farther away from the norm?

2. Background and Data Composition

In science as in politics, race proves to be a sensitive subject. The U.S. Census Bureau, from whose decennial censuses in 1990 and 2000 this study obtained its data, emphasizes that race is a subjective classification. A citizen's race is defined by him or herself subjectively.² The other variable in this study, income, is factually measurable. Each household was asked to report particulars of its income at different points in the census survey. The sum of these, "wage or salary income; net self-employment income; interest, dividends, or net rental or royalty income or income from estates and trusts; social security or railroad retirement income; Supplemental Security Income (SSI); public assistance or welfare payments; retirement, survivor, or disability pensions; and all other income," is defined as total income (USCB, 2005). The median income of a sample space is the number in which half of all incomes are below and half of all incomes are above.

² Although the statistical use of race is hazy from a technical standpoint—the Bureau states that it "should not be interpreted as being scientific or anthropological in nature" (USCB, 2005)—it still suffices as a realistic organizing scheme. Readers should, however, be conscious that one of the two variables in this study is a subjective judgment.

While collecting statistics such as race and median household income, the Bureau divides up counties into Census land tracts. There are 49 in Durham County. The high density of tracts in a small area makes data comparisons more fruitful, as greater specificity of observation lends itself to more specific results.³ For each tract, the Census data provide median household income along with the individual populations of five main races (Black or African American, White, American Indian or Native Alaskan, Asian or Pacific Islander, and Other).⁴ Therefore, each of the 49 tracts has 5 race populations, along with 1 median income per tract. These sets are reported for both years 1990 and 2000, and can also be retrieved for counties and states instead of Census land tracts (USCB, 2005).

3. Data Issues

Although the U.S. Census Bureau works to maintain consistency through time, the evolution of the American housing landscape and the iterative improvement of the Census lead to small discrepancies between decades. The following inconsistencies were found in Durham County land tracts:

- five tracts in 2000 are not reported in 1990
- one tract is listed in 1990 with a median income but without populations
- another tract in 1990 is listed with populations but without a median income

These seven insufficient tracts were excluded from the data sets in this study. Some 42 of the 49 tracts remained usable, which is still a robust data set.

The divisions of race proved likewise inconsistent between the decades. While the 1990 set pairs Asians and Pacific Islanders together, the 2000 set reports them separately. This dilemma was easily remedied by summing the two into one datum per tract. Another disparity that has more drastic implications is the addition in 2000 of the race option “Two or more races,” a classification that was not included in the 1990 set. Had the 1990 citizens been given this option, some may have chosen it. Instead, these people likely chose another option in 1990, and then may have switched to “Two or more races” in 2000. Such switches out of 1990 race options could affect the results of this study, since the raw data do not explain if population decline was caused by exiting or by classification changes. Therefore, population differences between 1990 and 2000 are most likely off by a small amount when assumed to be caused purely by exiting.

³ See section 5, paragraph 2 for an in-depth discussion of dense measurements.

⁴ See section 3, paragraph 2 for data issues relating to race classifications.

Since the “Two or more races” population in the 2000 set have no comparable category in the 1990 set, they were excluded in this study, resulting in the “total population” of each 2000 tract being artificially low by a small degree. As a whole, Durham County’s population was reduced by 2% due to this exclusion.

In terms of overall data, the limitations of the Bureau’s information retrieval program on its website curbed the scope of data comparison. The largest number of data sets that can be returned at a time by the Bureau is 7,000, a cap that hindered the analysis of large sample spaces by Census land tract. The largest area of land tracts that could be retrieved by the website was the Southern states in the region of North Carolina (North Carolina, South Carolina, Virginia, Georgia, West Virginia, and a portion of Tennessee). The entire South and the United States could only be returned and analyzed by county.⁵

From a broad standpoint, another data problem in this study is causality. One cannot tell if race distribution changes because of median income changes or *vice-versa* when comparing the correlation between the two. Yet, although this study cannot inspect at this juncture the cause-and-effect relationship between the two, it can take the first step and confirm that there is or is not a strong correlation in Durham between the two.

4. Calculations and Figures

The first step toward making an accurate comparison of race distribution by median income was computing by what proportions the five races compose each tract. Percentages are necessary, since this study is meant to investigate the relationship of each race to median income per tract. Raw population data are of little use in determining how much influence a race may have on median income, or *vice-versa*, considering 100 whites could be 100% of a population if the total is 100 persons or 1% if the total is 10,000 persons. While both scenarios have the same raw populations, the implications are drastically different. Consequently, proportions were utilized.

Given r_{ij} , the reported population of individual races with i being the classification [1, 5] of {1: Black or African American, 2: White, 3: American Indian or Alaskan Native, 4: Asian or Pacific Islander, 5: Other} and j denoting the land tract, the total population of tract j^* is given as

⁵ See section 5, paragraph 2 for analysis and implications of this dilemma.

$$P_j = \sum_{(i=1,5)} r_{ij} \quad j \in [1,42] \quad (1)$$

Each race population r_{ij} was then divided by P_j , the total population of tract j , to provide R_{ij} , the percentage of race i in the total population of tract j .

$$R_{ij} = r_{ij} / P_j \quad i \in [1,5], j \in [1,42] \quad (2)$$

Equations 1 and 2 supply half of the necessary information, given that this study compares race distribution to median household income of an area. The second half was provided directly from the Census data sets: raw median household income per tract. However, inflation distorts the real value of a dollar over time, thus the 2000 set median incomes were adjusted to have values directly comparable to the 1990 set. Using a Consumer Price Index inflation calculator (CPI...), which stated that 1.00 dollar in 1999 was worth 0.744 dollars in 1989, the adjusted median household incomes of the 2000 set was found by multiplying each income by 0.744.⁶ In summary, for unadjusted income z , the adjusted income Z_j for tract j is:⁷

$$Z_j = z_j * .744 \quad j \in [1,42] \quad (3)$$

The compilation of results can be seen in Figure 1 for 1990 and Figure 2 for 2000. These bar graphs display the composition of race in terms of percentages for each tract. In order to show the relation to median income, the tracts are ordered from least income to greatest (although not all income labels could fit on the x -axis).

The second utilization of this data is more revealing. In Figures 3 and 4, the race proportions of total population for each tract are plotted against the median household incomes of those tracts (for both the 1990 set and the 2000 set). These graphs demonstrate for whites and blacks any pattern of race distribution in relation to median income. To better highlight the patterns, linear regressions are displayed for both data sets (the strength of the relationship is given by the R^2 value). Similar to R^2 , Figure 5 displays the correlation values between median incomes and racial shares in 1990 and 2000. Correlations show how strongly two arrays of data

⁶ The 2000 median household incomes are actually recordings from 1999. In this study, they are referred to as being part of the year 2000 for the sake of simplicity. The same is true for 1989 incomes, which are applied to the 1990 set.

⁷ The 1990 median household income will also be denoted with a Z , even though it is not multiplied by any constant. It can be considered “adjusted” from the outset.

are related, with -1.0 being the strongest possible correlation in opposite directions and 1.0 being the strongest possible correlation in the same direction.⁸

In order to compare the relationship of median income to race distribution in Durham County to other sample spaces, the same process used to create Figures 3 and 4 (equations 1 through 3) was utilized for land tracts and counties of North Carolina in Figures 7 and 8. This process was employed again to find the data used in Figures 6, 9, 10, 15, and 16.⁹ Figure 6 displays the R^2 value of linear regressions for four comparative sample spaces (North Carolina measured by Census land tract, states near North Carolina measured by Census land tract,¹⁰ states in the South measured by county,¹¹ and the U.S. measured by county). Figures 9, 10, 15, and 16 display the linear regressions of Durham (Durham County), North Carolina, of Birmingham, Alabama (Jefferson County), of Greensboro, North Carolina (Guilford County), and of Nashville, Tennessee (Davidson County), all measured by land tract.

The method used to describe how the relationship of racial shares to income distribution changes over time originates from the Figures 3 and 4. The difference in linear regressions of the 1990 and 2000 data sets in these graphs indicate how the linear pattern of race percentage changed during those ten years. To visualize this, the linear regression function of 2000 was subtracted from that of 1990. The y-intercept of this linear regression difference was then set to 0, since vertical shifts indicate pure population increase. See Figures 11 and 12 for these linear regression differences. Figures 13 and 14 make use of this method with the regressions from Figures 9, 10, 15, and 16.

5. Results

Along with Duke University, the Bulls minor league baseball team, and countless tobacco relics, is Durham County home to the “rich whites, poor blacks” maxim seen throughout the

⁸ For example, if person A is throwing baseballs to person B, the correlation value between the number of baseballs that A has and B has would equal -1.0. When A throws to B, B’s stock of baseballs strictly increases while A’s strictly decreases.

⁹ For the sake of relevance and brevity, the original graphs that lent the data are not included.

¹⁰ North Carolina, South Carolina, Virginia, Georgia, West Virginia, and approximately three quarters of Tennessee.

¹¹ The U.S. Census Bureau defines the South as “Maryland, Delaware, West Virginia, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, Texas” (USCB, 2005).

U.S.? A quick glance at Figures 1 and 2 answers that question: yes. Figure 5 lends support to this claim, signifying that median income has a strong positive correlation with whites and a strong negative correlation with blacks in both 1990 and 2000. Figures 3 and 4 solidify this idea and add a dimension of magnitude. In Figure 3, a strongly correlated (high R^2 value), steep, downward-sloping regression line is displayed for blacks—the negative slope indicating that as income increases in an area, the percentage of blacks decrease. The steepness represents how drastically the race percentage decreases for a change in median income. Figure 4 displays a strongly correlated, steep, upward-sloping regression line for whites, which intimates the opposite: higher income areas correlate to significantly higher white percentages.

The data indicate that Durham is like the rest of the U.S, but to what degree? The first step in assessing Durham is to find a suitable comparison. Figure 6 displays the R^2 values of linear regressions found in Figures 3 and 4, but also the R^2 values of the same method of measurement for different sample spaces in America. One might notice that the area of the sample space increases and the density of measurement decreases as the graph is read left to right. Accompanying this movement is decreasing R^2 values. This graph illustrates the importance of high-density measurements and similar-space comparisons. As the sample space area increases past the city edge, it encompasses rural areas that behave in different ways than the city. A wider and wider area encompasses more areas that are unlike the city, and therefore strays more from being an accurate, correlated comparison. Likewise, the density of measurement (land tract versus county) has a large impact. The greater divisions an area is split into, the more precise assessment of similarities and differences. Proof of this are Figures 7 and 8: although both encompass the same sample space, the lower-density, county-measured graph yields R^2 values that are only one-third and one-fourth that of the high-density, tract-measured graph.

If results are skewed by a space that encompasses types of land unlike that of the base, or by a measurement standard less dense than that of the base, then the logical conclusion is to compare Durham to another city (measured by land tract). This study does so with three: one thriving city (Nashville, Tennessee), one typical city (Greensboro, North Carolina), and one stagnant city (Birmingham, Alabama) (as characterized by Becker). As one can see from Figures 9 and 10, Durham in 1990 rests roughly in the middle of the three in terms of racial inequality by income. Given each city's 1990 slope of blacks, if there were two hypothetical tracts in each city, one with a median income of 20,000 dollars and the other 60,000 dollars, then the second, higher income tract would display 77.2% less black population than the first tract in Birmingham, 60.8% less black population than the first tract in Durham, 58.1% less black population than the

first tract in Nashville, and 53.4% less black population than the first tract in Greensboro.¹² Likewise, in the same representative tracts from each city, the 60,000 dollar median income tract would display 77.1% more white population than the 20,000 dollar median income tract in Birmingham, 61.2% more white population in Durham, 58.4% more white population in Nashville, and 54.5% more white population in Greensboro (see Table 1). As this data implies, Durham was roughly in the middle for race versus income inequality.¹³

Table 1: Linear Regression Slopes and Percentage Change for Hypothetical 40,000 Dollar Increase in Median Income (1990)

Race	City	Slope	% Change
Black	Birmingham	-1.93E-05	-77.2%
	<i>Durham</i>	<i>-1.52E-05</i>	<i>-60.8%</i>
	Nashville	-1.45E-05	-58.1%
	Greensboro	-1.33E-05	-53.4%
White	Birmingham	1.93E-05	77.1%
	<i>Durham</i>	<i>1.53E-05</i>	<i>61.2%</i>
	Nashville	1.46E-05	58.4%
	Greensboro	1.36E-05	54.5%

Measured against a number of similar counterparts, Durham appeared to be about average in terms of “rich whites, poor blacks.” Yet, is this changing? Since the R^2 of the regression slopes are high in Figures 9 and 10, the differences between the regression slopes of 2000 and 1990 for both races are solid indicators of the direction that trends are heading. The red linear regression difference lines on Figures 11 and 12 intimate that, although Durham displays “rich whites, poor blacks”, the difference between the two is thinning. Figure 11 shows a difference line that is positively sloped, indicating that higher black percentages were in high income areas in 2000 than in 1999. Similarly, Figure 12 shows the opposite: lower white percentages were in high income areas in 2000. The scales appear to be moving toward a more equitable balance.

How does Durham’s change in race versus income compare with the other cities? Have they lessened inequality at the same pace as Durham? Figure 13, which displays the linear differences for blacks of all four cities, indicates no, the other cities have not reduced inequality as much as Durham. While Durham’s slopes for whites and blacks moved toward equality (*i.e.* a horizontal slope of zero) by decreasing an average of 15.4% from 1990 to 2000, Birmingham’s slopes only decreased an average of 7.7%. Greensboro’s and Nashville’s slopes actually

¹² Given a slope of X percentage/median income, $X * (50,000 - 10,000 \text{ median income}) = Y \text{ percentage change}$.

¹³ Pure equality can be considered a 0% change in population with any change in income.

increased away from equality by an average of 7.5% and 8.0% respectively from 1990 to 2000. The final results of these changes are displayed in Figures 15 and 16, which could allay any person’s fears of inequality in Durham: from 1990 to 2000, Durham has moved from an average city of “rich whites, poor blacks” to having the least inequality of four comparable cities. As before, given the hypothetical second tract with a median income 40,000 dollars higher than the first tract, black population in 2000 would be 71.2% less in the higher income tract in Birmingham, 61.7% less in Nashville, 55.5% less in Greensboro, and 51.6% less in Durham. Similarly, the higher income tract would display a white population 71.2% higher than the first in Birmingham, 64.3% higher in Nashville, 60.5% higher in Greensboro, and 57.9% higher in Durham (see Table 2).

Table 2: Linear Regression Slopes and Percentage Change for Hypothetical 40,000 Dollar Increase in Median Income (2000)

Race	City	Slope	% Change
Black	Birmingham	-1.78E-05	-71.2%
	Nashville	-1.54E-05	-61.7%
	Greensboro	-1.39E-05	-55.5%
	<i>Durham</i>	<i>-1.29E-05</i>	<i>-51.6%</i>
White	Birmingham	1.78E-05	71.2%
	Nashville	1.61E-05	64.3%
	Greensboro	1.51E-05	60.5%
	<i>Durham</i>	<i>1.45E-05</i>	<i>57.9%</i>

6. Conclusion

Although Durham might resist the plagues of white flight and house value conformity, the generality of “rich whites, poor blacks” can call Durham “home.” However, this is not to say that Durham is a welcome abode, as changes between data sets indicate that Durham’s inequality is lessening at a greater rate than all other tested cities. Consequently, although Durham may not be a trend-breaking Jackie Robinson of inequality now, if patterns continue, it may be in the future.

Works Cited

Becker, Charles. Lecture. ECON 195S: Urban Economics. Fall 2005.

CPI Inflation Calculator. *National Aeronautics and Space Administration*. Updated 21 January 2005. <<http://www1.jsc.nasa.gov/bu2/inflateCPI.html>>. 9 November 2005.

U.S. Census Bureau. <<http://factfinder.census.gov/>>. 4 November 2005.

Figure 1

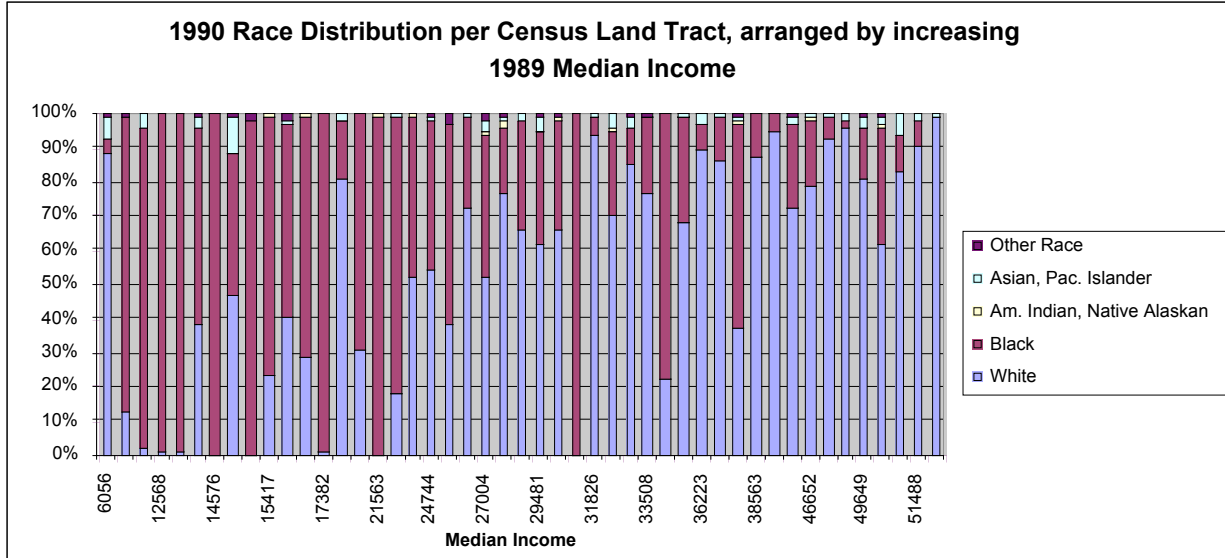


Figure 2

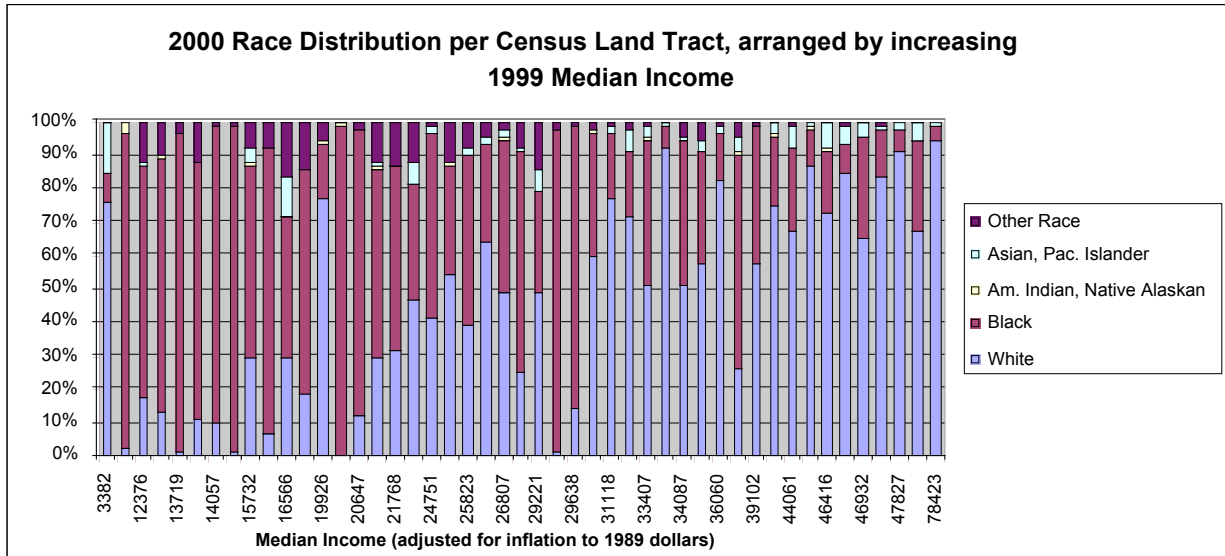


Figure 3

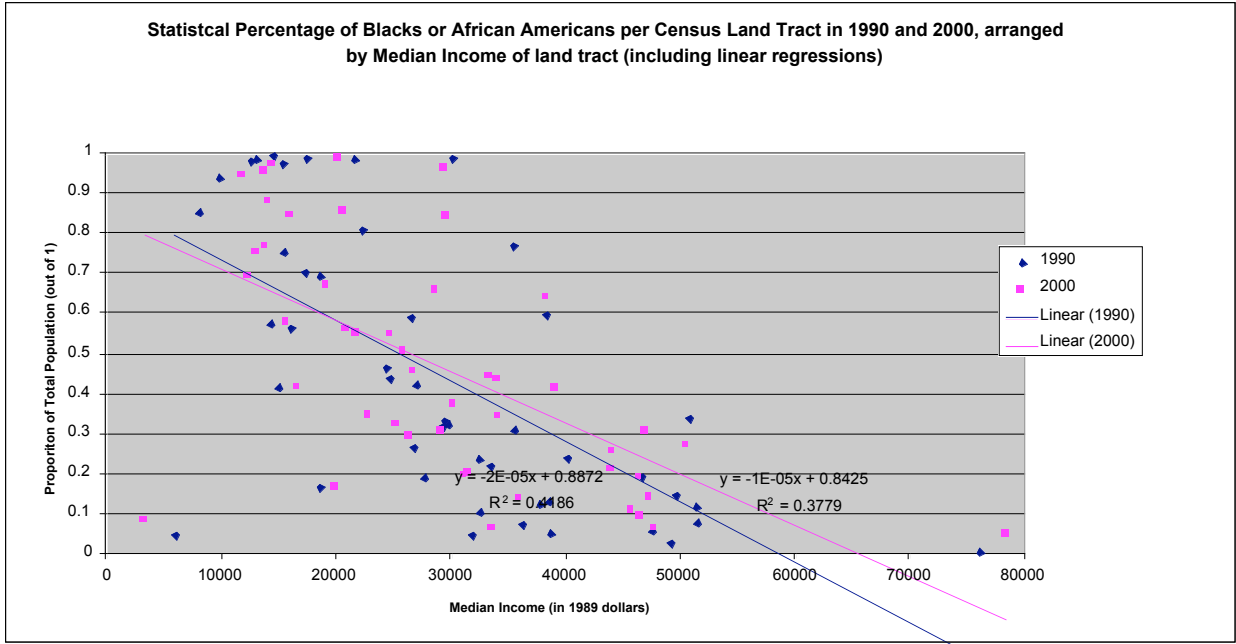


Figure 4

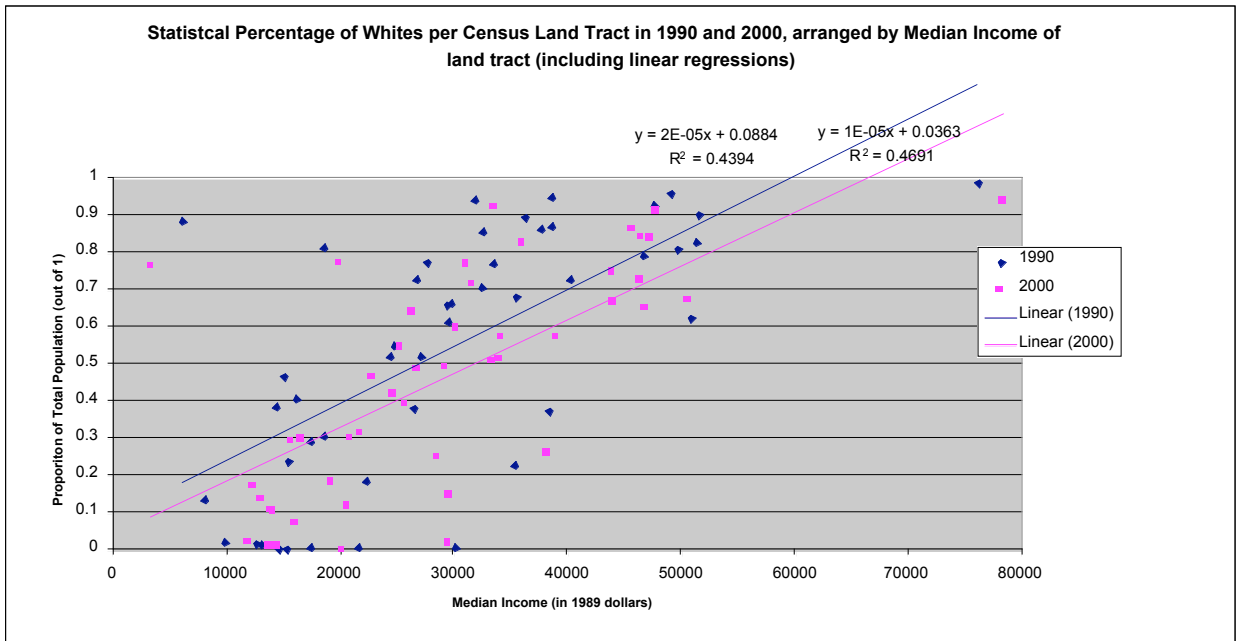


Figure 5

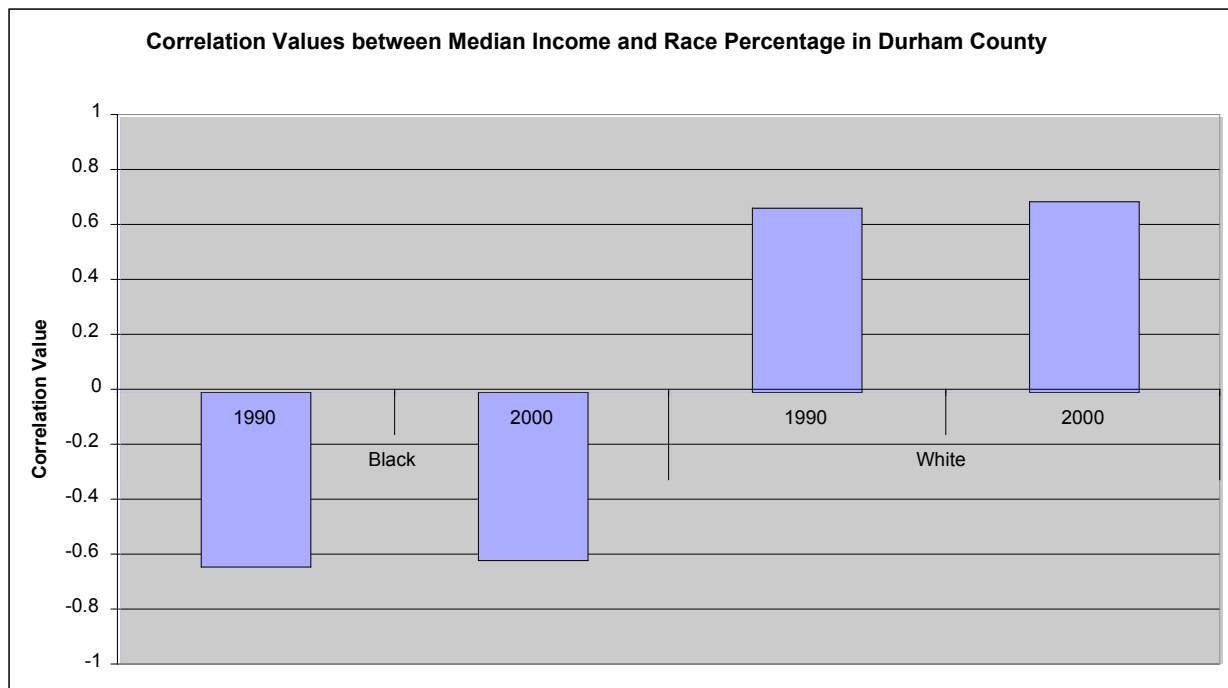


Figure 6

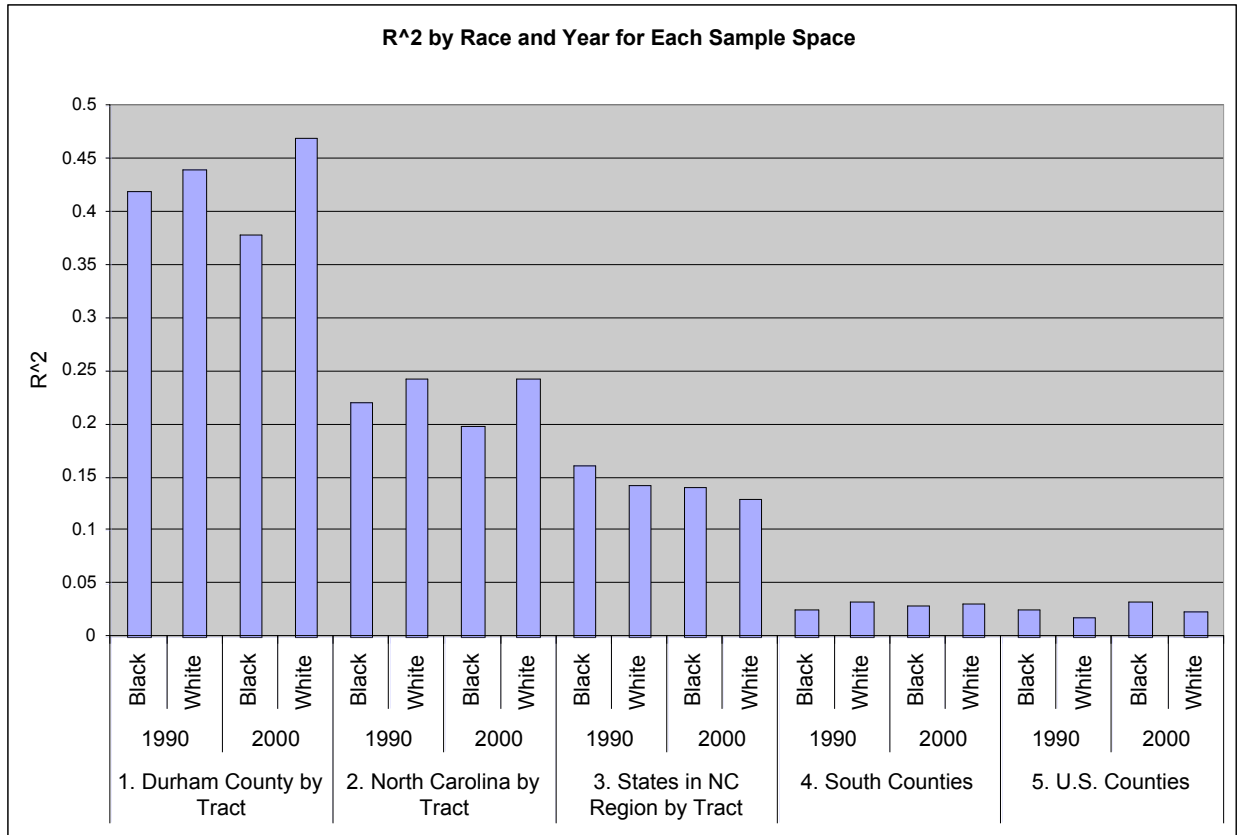


Figure 7

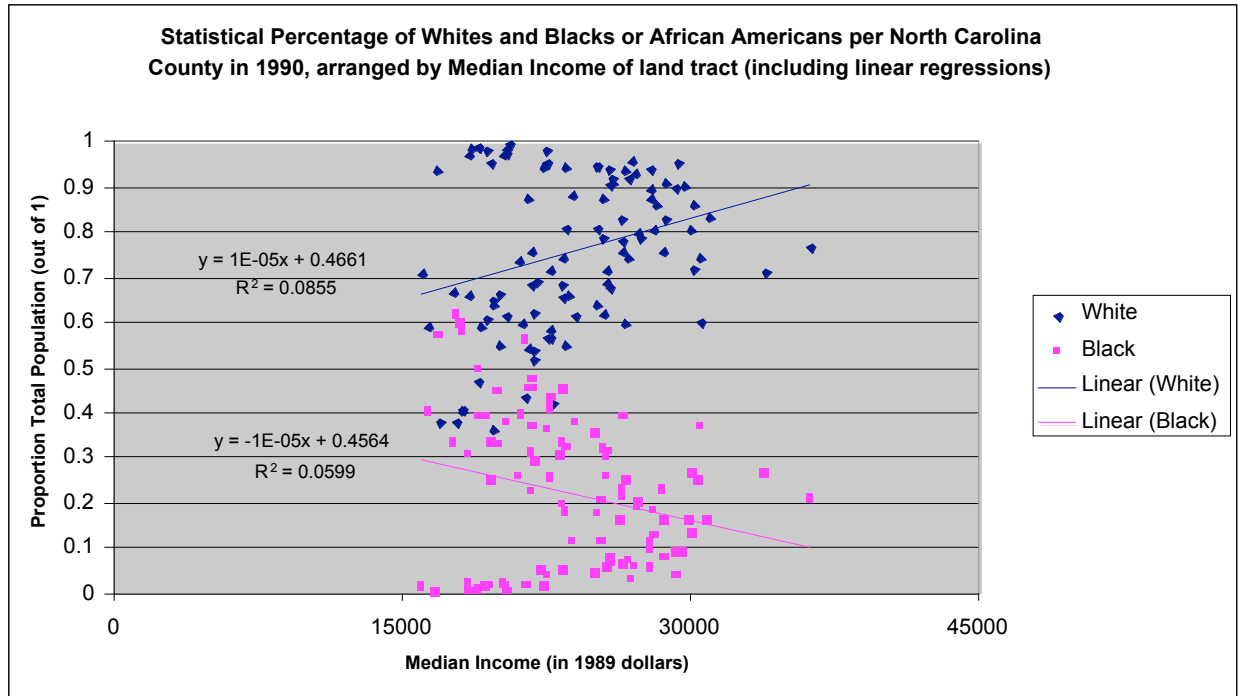


Figure 8

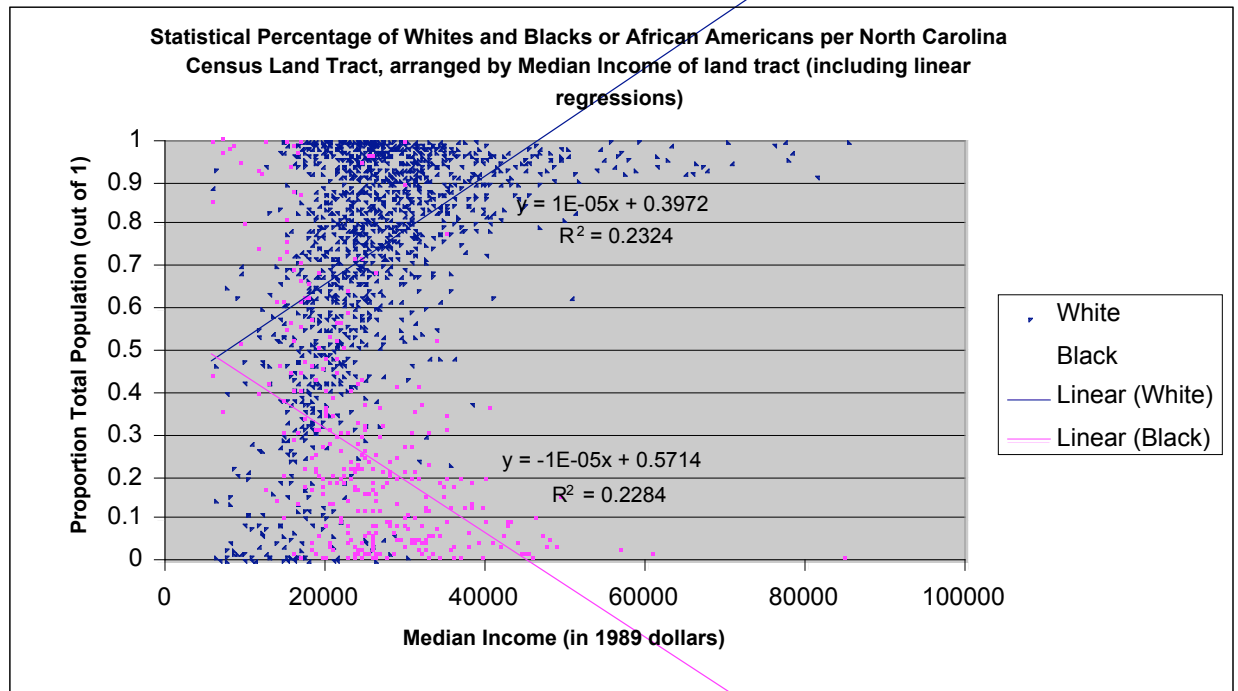


Figure 9

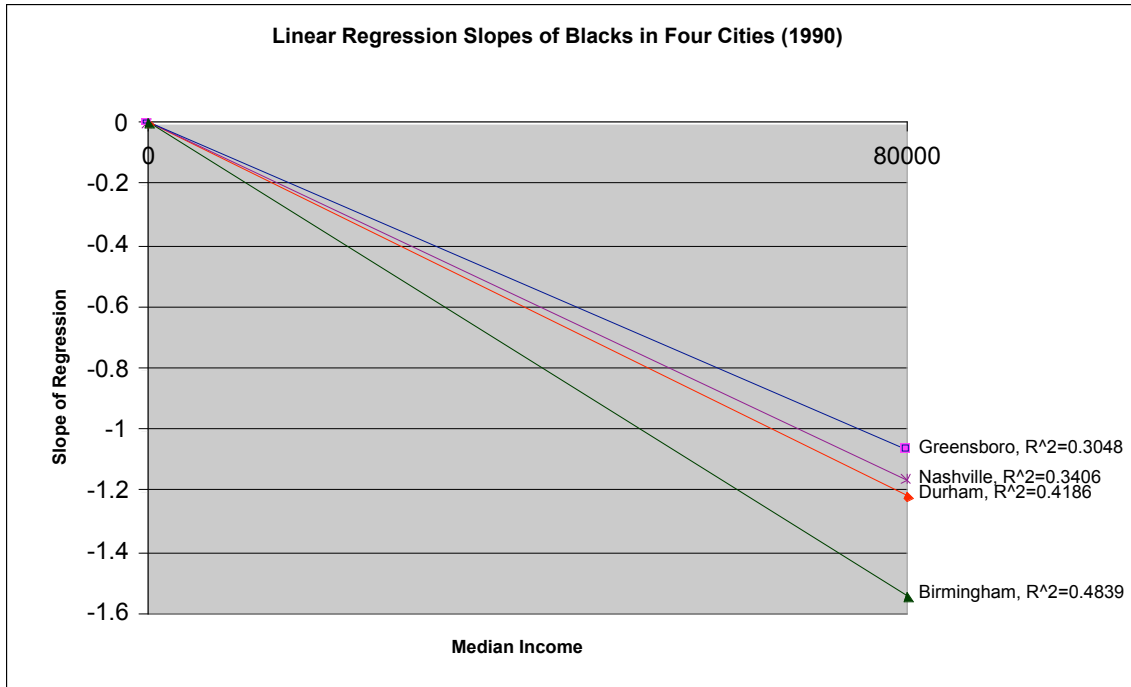


Figure 10

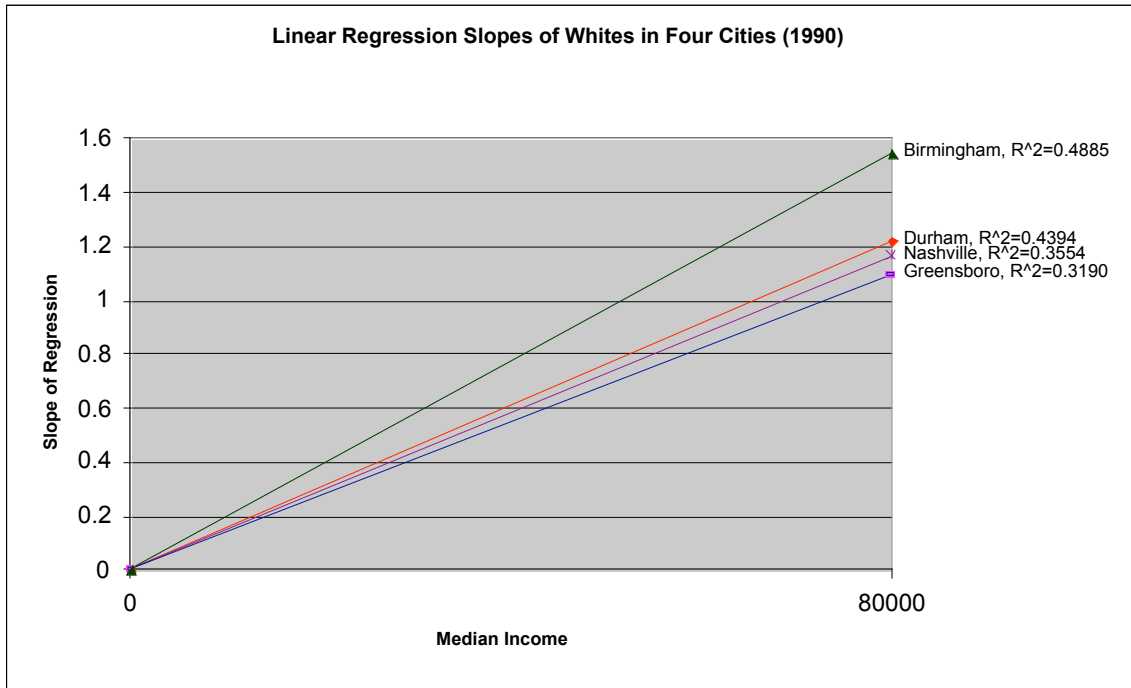


Figure 11

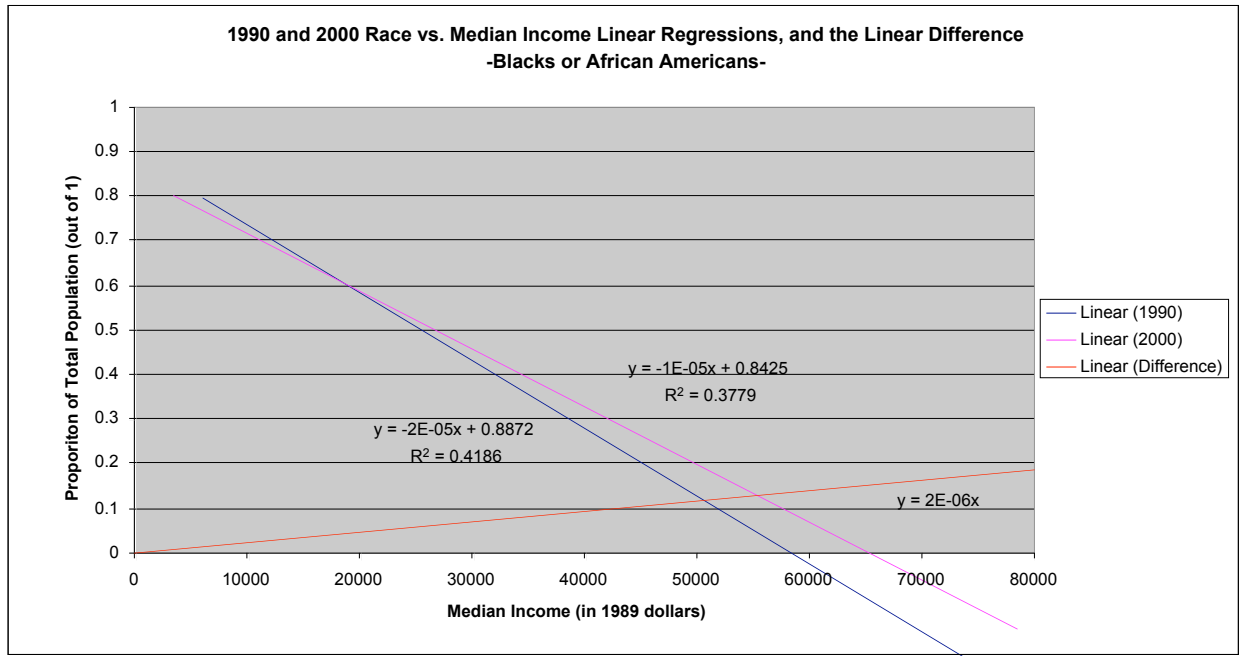


Figure 12

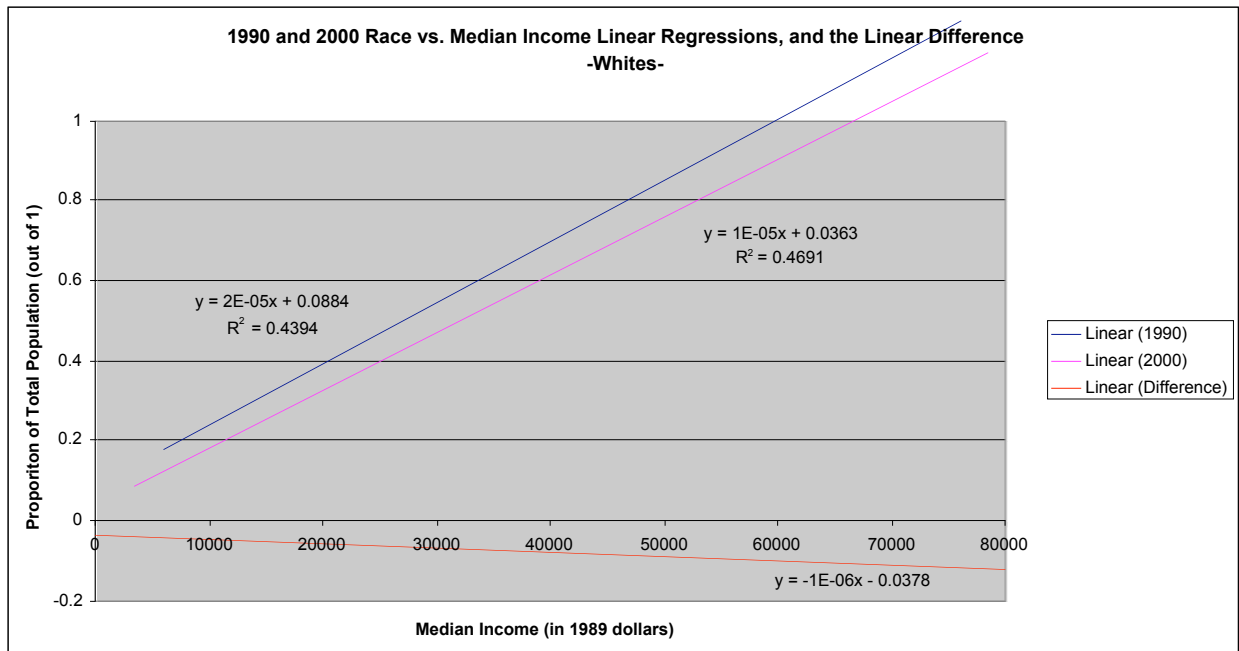


Figure 13

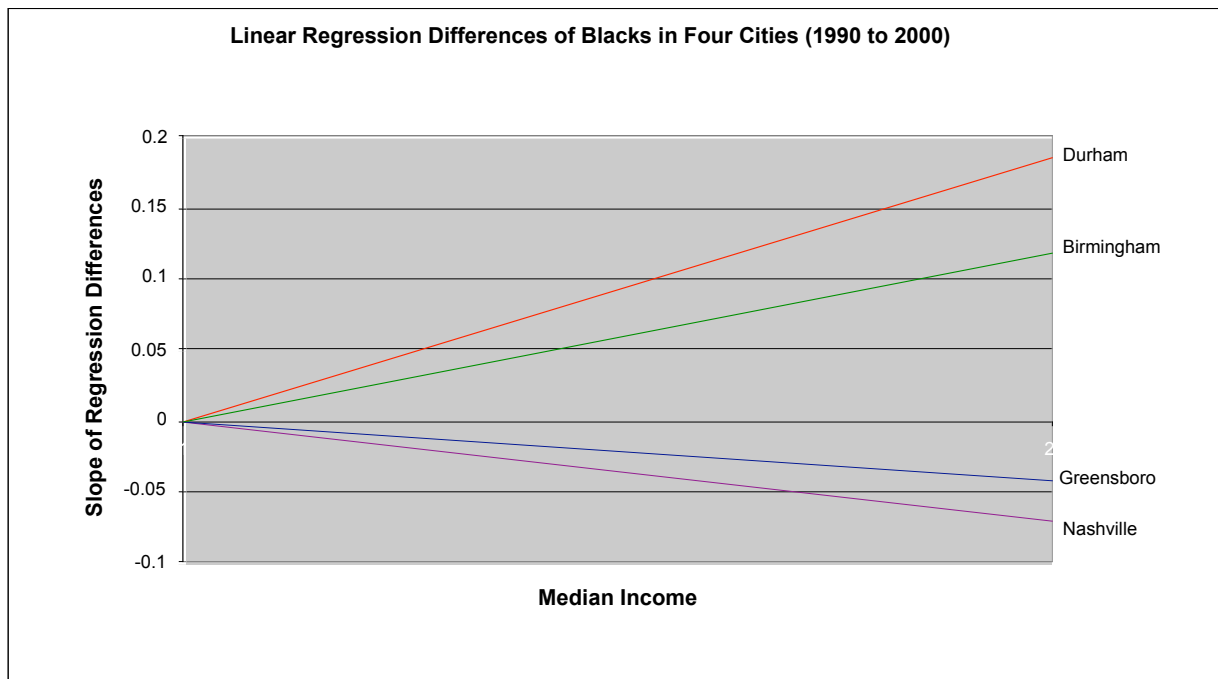


Figure 14

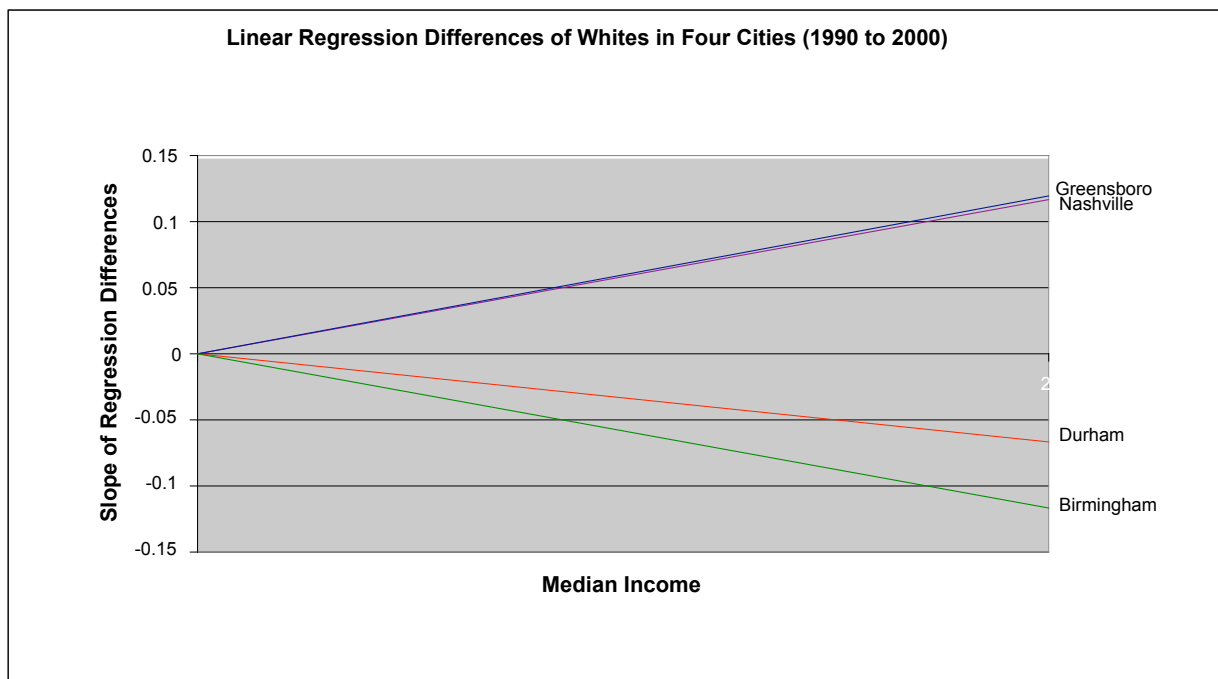


Figure 15

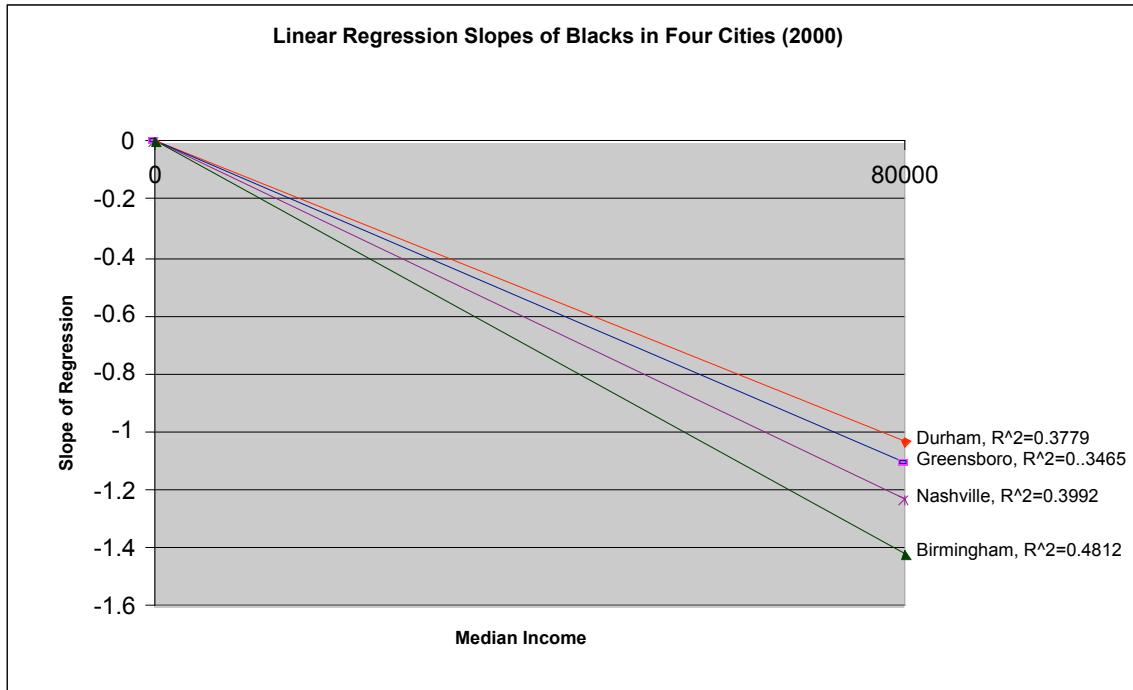


Figure 16

