



# Examining Voting Patterns in Black Belt States

Ryan A. Robert with Dr. James B. Foust  
University of Wisconsin - Eau Claire  
Department of Geography and Anthropology



## Abstract

The purpose of this research is to analyze the relationship between key socioeconomic variables and voting tendencies within the states of Mississippi, Alabama, and Georgia. The Black Belt, a region originally named for its rich black soils but also distinguishable by demographics, runs horizontally through these three states. An overwhelming majority of counties within the "Black Belt" voted majority Democrat in the 2008 Presidential Election.

The dependent variable in this research will be the percentage of Democrat vote by county. The independent variables include measures of education, income, population, age, race, and poverty.

The general hypothesis suggests there is a direct, positive relationship between the percentage of votes cast for Barack Obama and education level, age, and percentage under poverty. There should be an inverse relationship between percentage of votes cast for Democrat and level of income and percentage of black population. The hypothesis in its linear equation form is:

$$Y = a + b_1(\text{BLACK}) - b_2(\text{WHITE}) - b_3(\text{INCOME}) + b_4(\text{POVERTY}) - b_5(\text{EDUCATION}) + b_6(\text{AGE})$$

Where:

Y = percent of votes for Obama per county

A = Intercept

BLACK = percent black population per county

WHITE = percent white population per county

INCOME = median household income per county

EDUCATION = percent of population per county with at least a bachelors degree

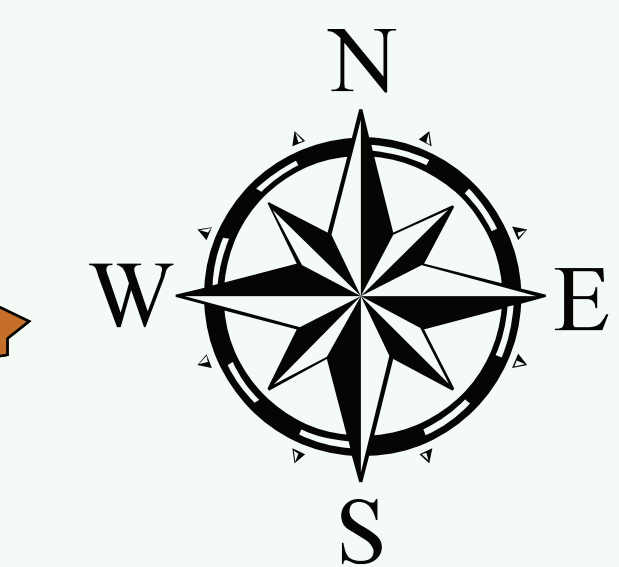
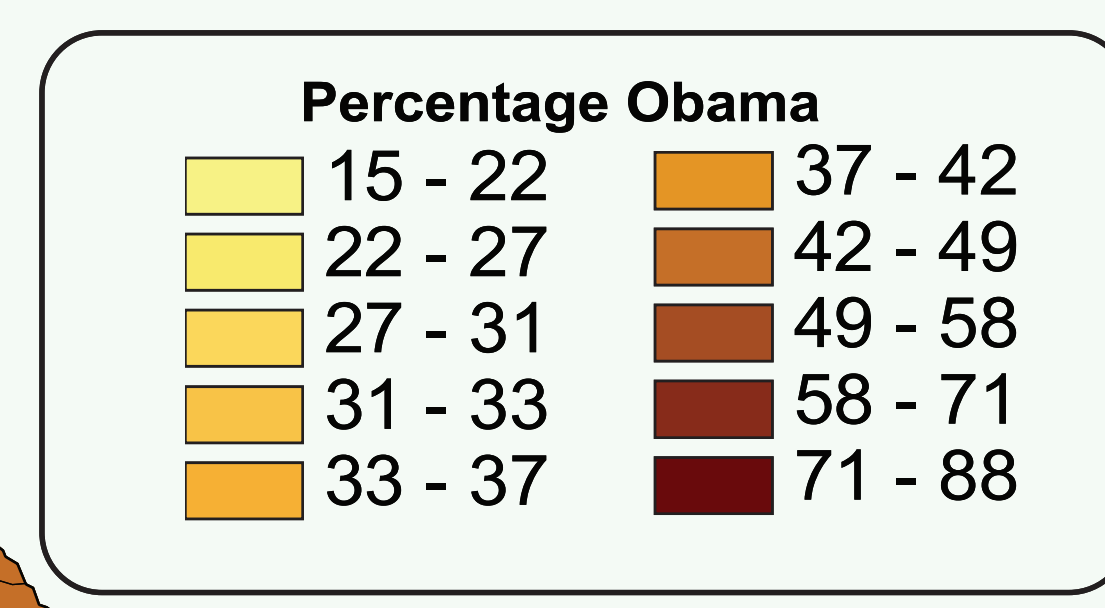
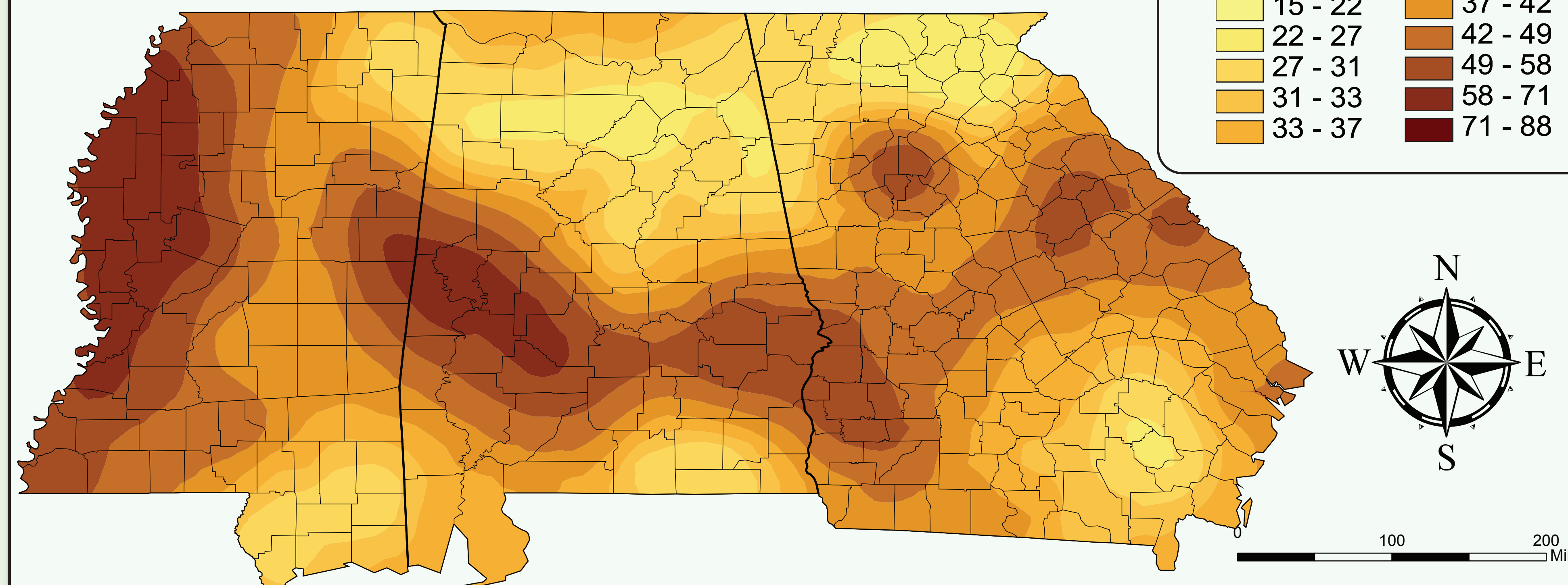
POVERTY = percentage of population under the poverty level per county

AGE = percent of county population in specified age range

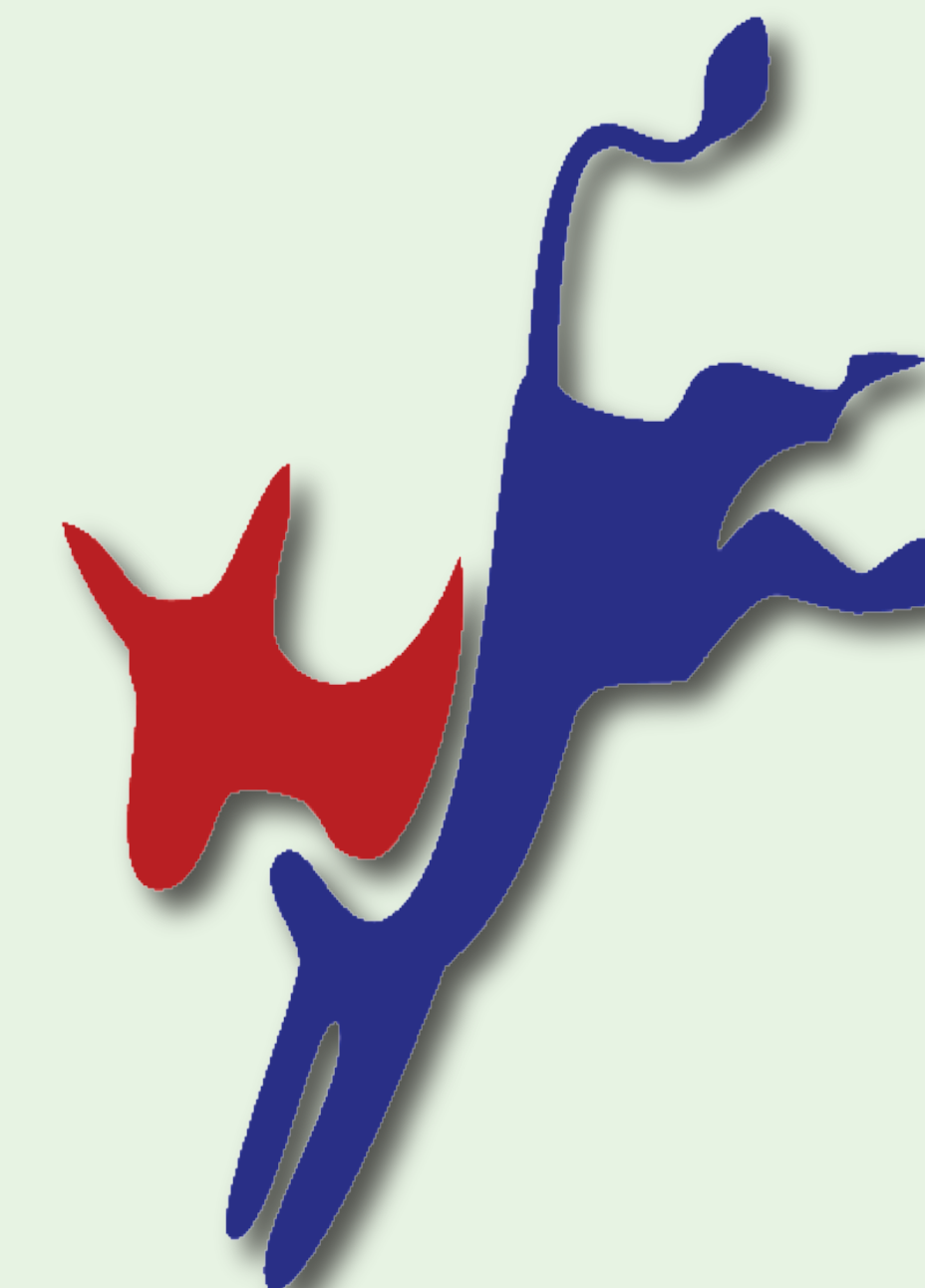


<http://www.britannica.com/blogs/wp-content/uploads/2009/02/obama.jpg>

## Votes Cast for Obama



0 100 200 Miles



<http://images.stuckon-stupid.com/democrats/Democrats%20logo.png>

## Data Availability

Attributes for all independent variables were obtained from the online 2000 Census at the county level. The data was then copied into an Excel spreadsheet where it was further analyzed. Election data was collected at the county level from the New York Times. Voting results from all 308 counties making up the study area were also transferred to Excel.

## Methods

The hypothesis was tested with multiple correlation/regression by using Excel and SPSS. The results of this analysis are given in Table 1 below. The coefficient of multiple correlation (R) is significant at the .05 level. The coefficient of determination (R<sup>2</sup>) indicates that approximately 73 percent of the variation in Obama votes can be 'explained' by variation in the six independent variables. Beta values were also determined for each independent variable. Since the Beta values represent the amount the dependent variable changes as a result of each independent variable, we can examine them to determine the importance of each independent variable. The table results indicate that the percentage of black population is the most important variable as its Beta value (.73) is the highest of the studied variables. Independent variables which had little or no effect on the dependent variable were eliminated and a new hypothesis equation was written. For example, because the age range 18-24 variable has a very low Beta value(.073), it was discarded from the data and a new hypothesis equation was written.

$$Y = a + b_1(\text{BLACK}) - b_2(\text{WHITE}) - b_3(\text{INCOME}) + b_4(\text{POVERTY})$$

Where:

Y = percent of votes for Obama per county

A = Intercept

BLACK = percent black population per county

WHITE = percent white population per county

INCOME = median household income per county

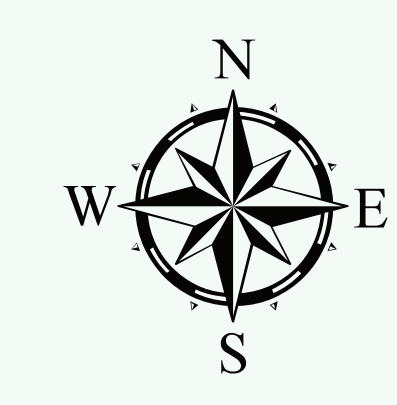
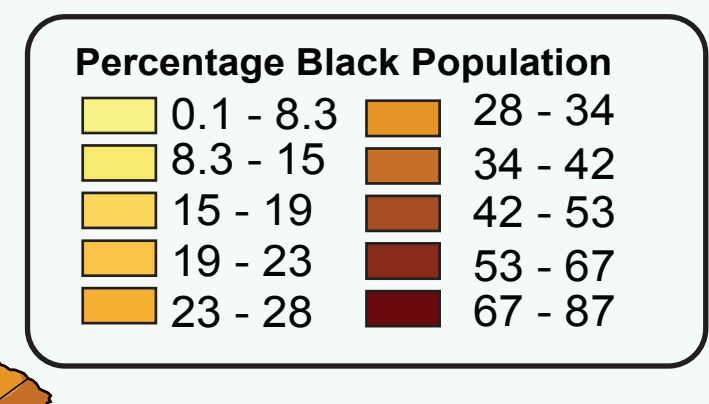
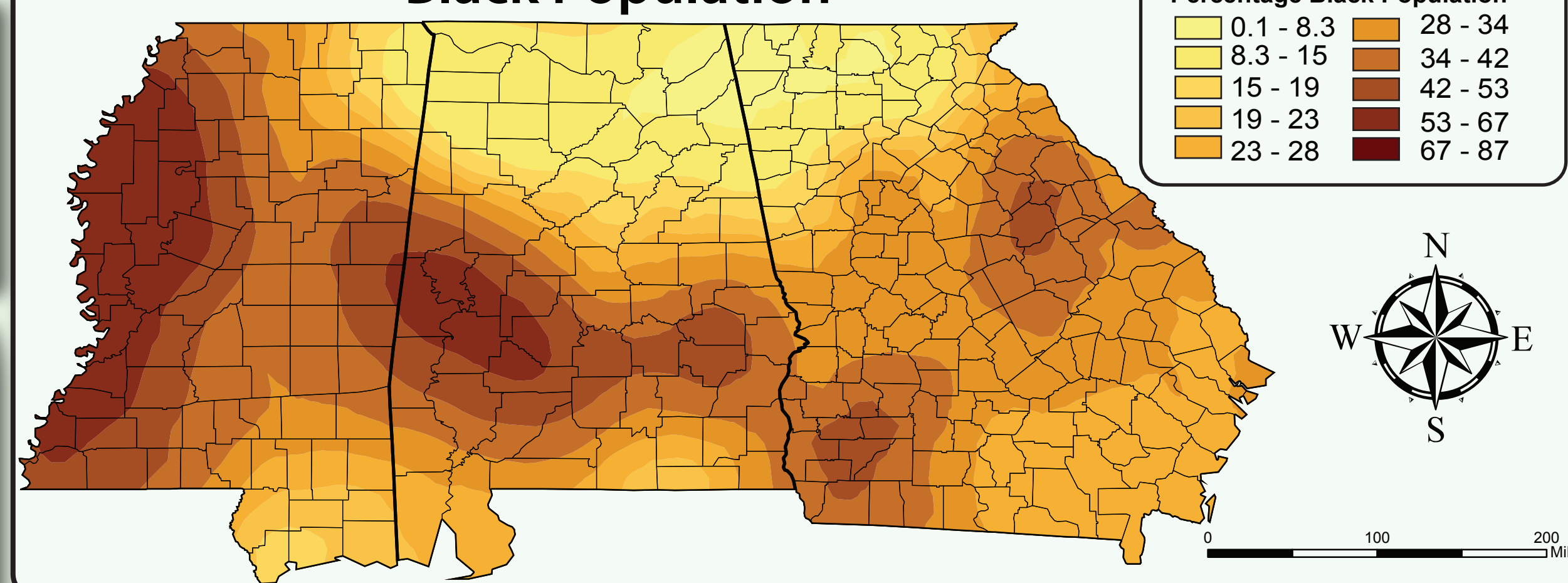
POVERTY = percentage of population under the poverty level per county

Correlation/regression results from the effective variables are also displayed in Table 1.

## Map Creation

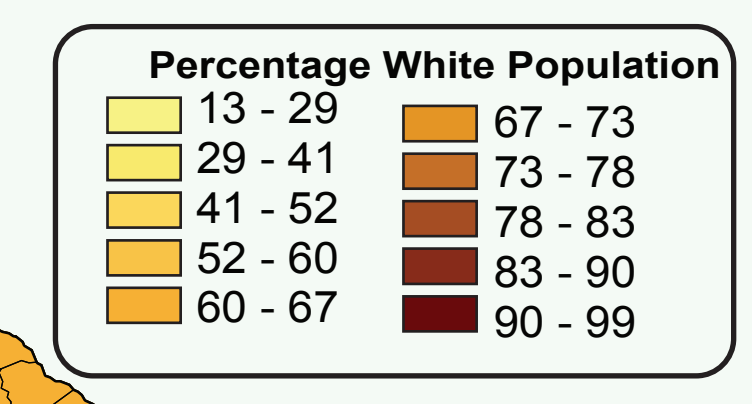
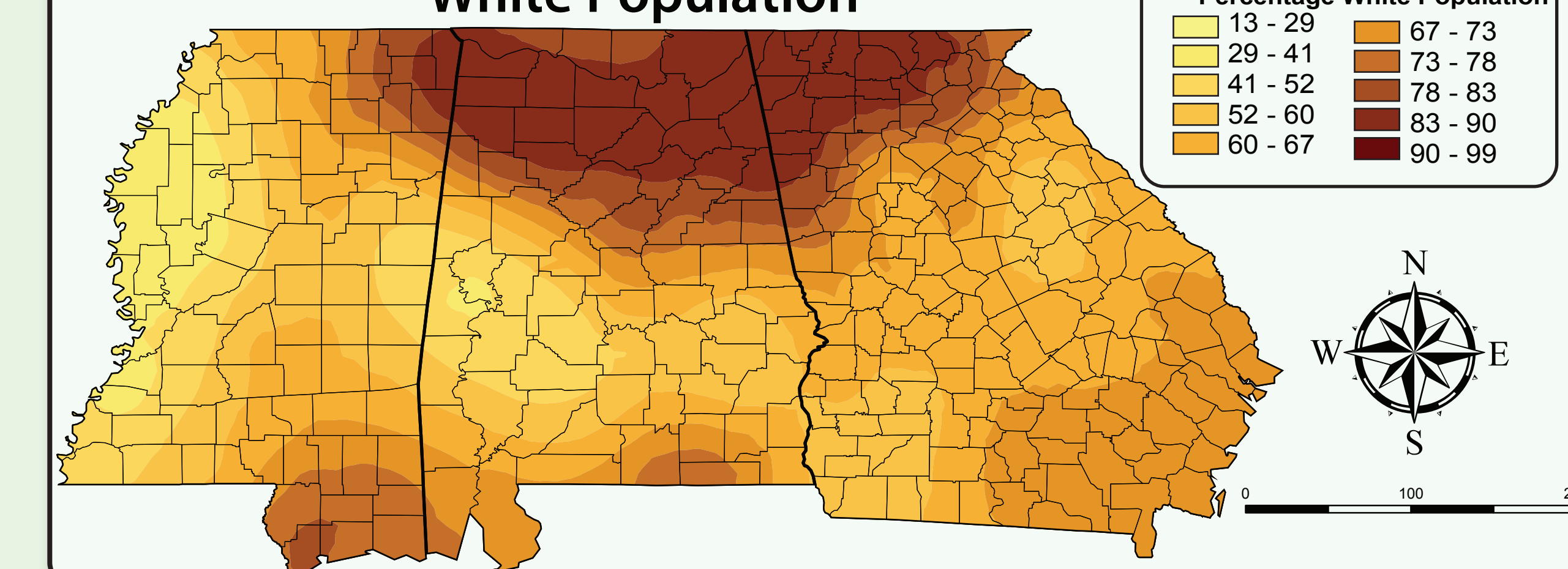
Once attribute data in the form of socioeconomic data for study area counties was developed, the mapping process could begin. Next, ESRI shape files for counties within the study area were acquired. After converting an Excel spreadsheet into dbf format by using Microsoft Access, the dbf table could be spatially joined to the county shape files by a key attribute field containing each county's FIPS code. Kriging maps were interpolated for each variable for visual representation.

## Black Population



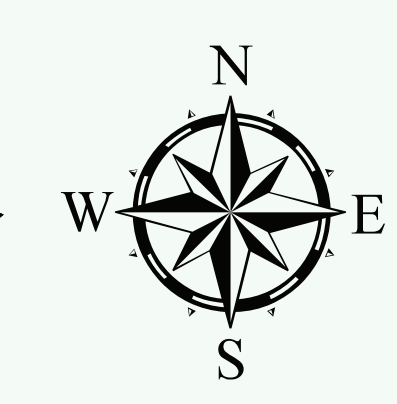
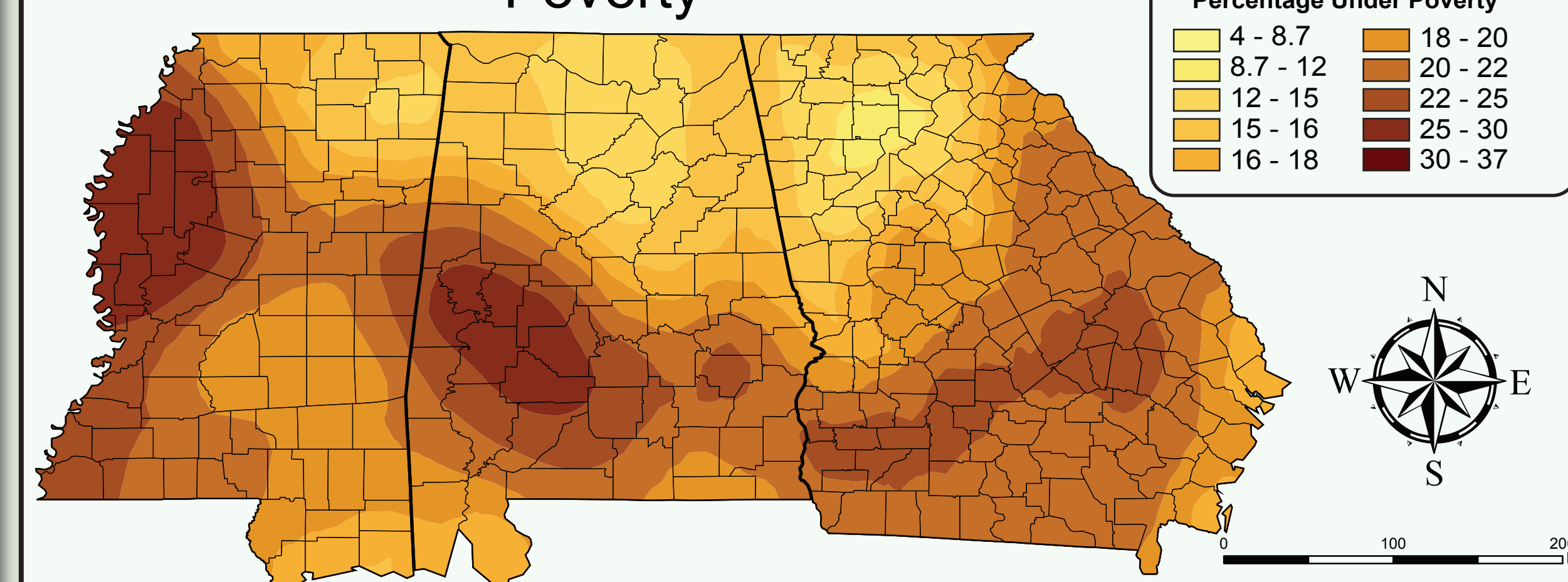
0 100 200 Miles

## White Population



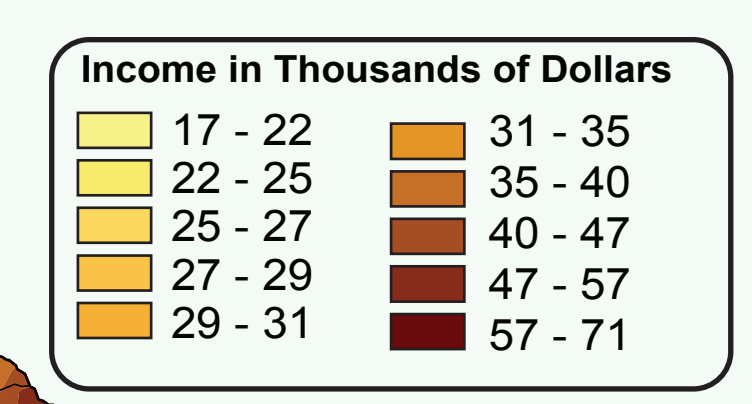
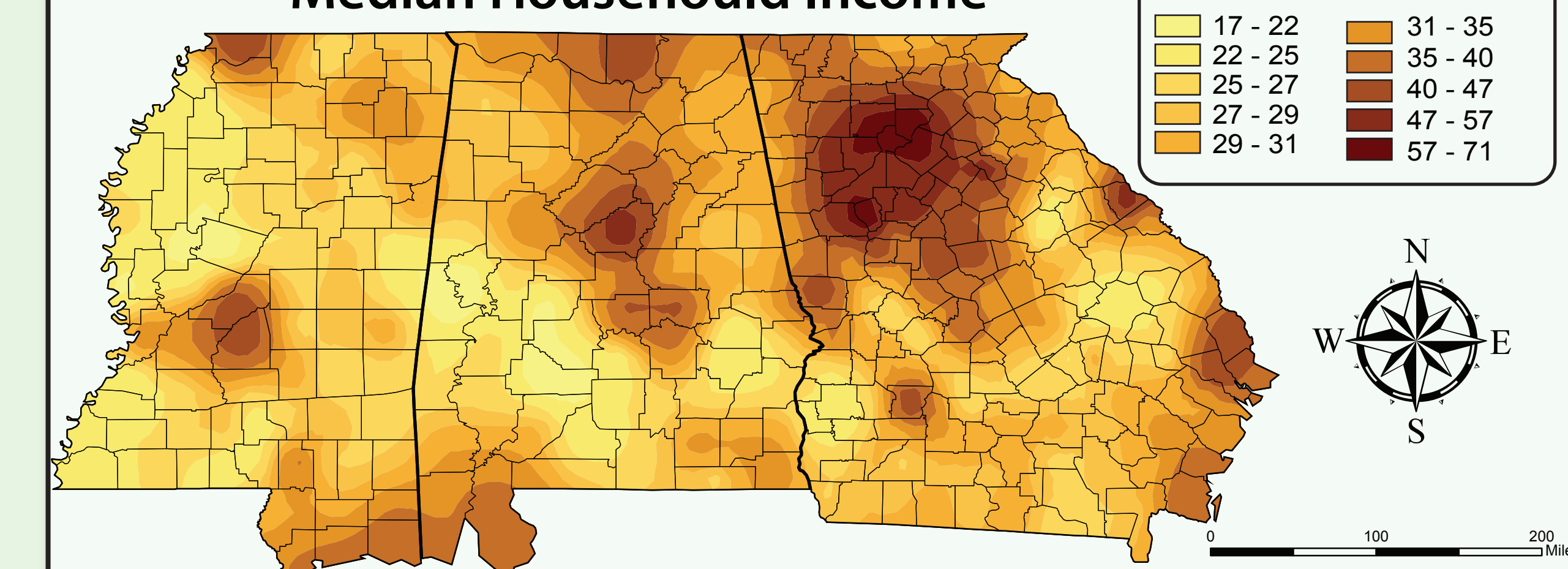
0 100 200 Miles

## Poverty



0 100 200 Miles

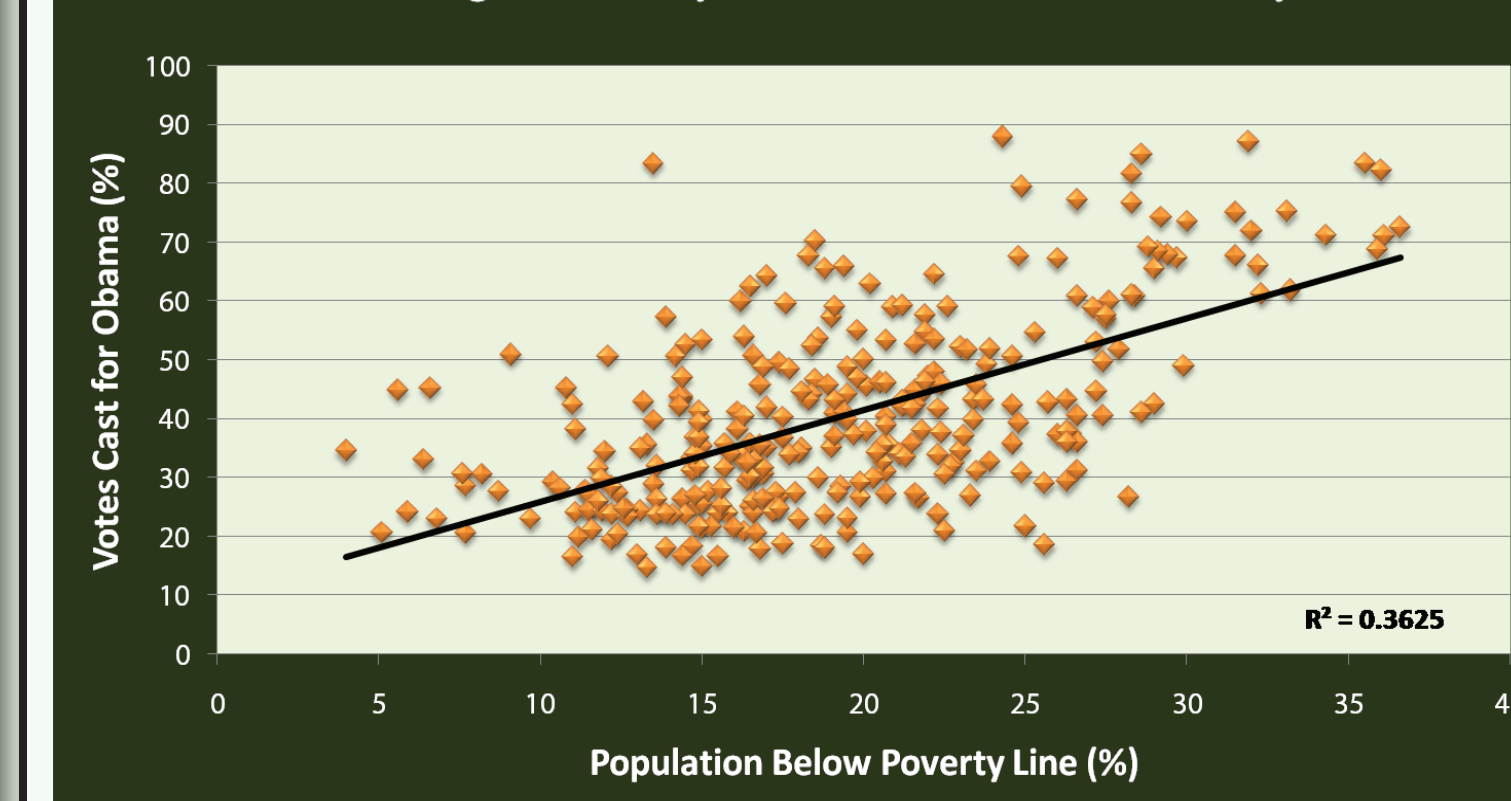
## Median Household Income



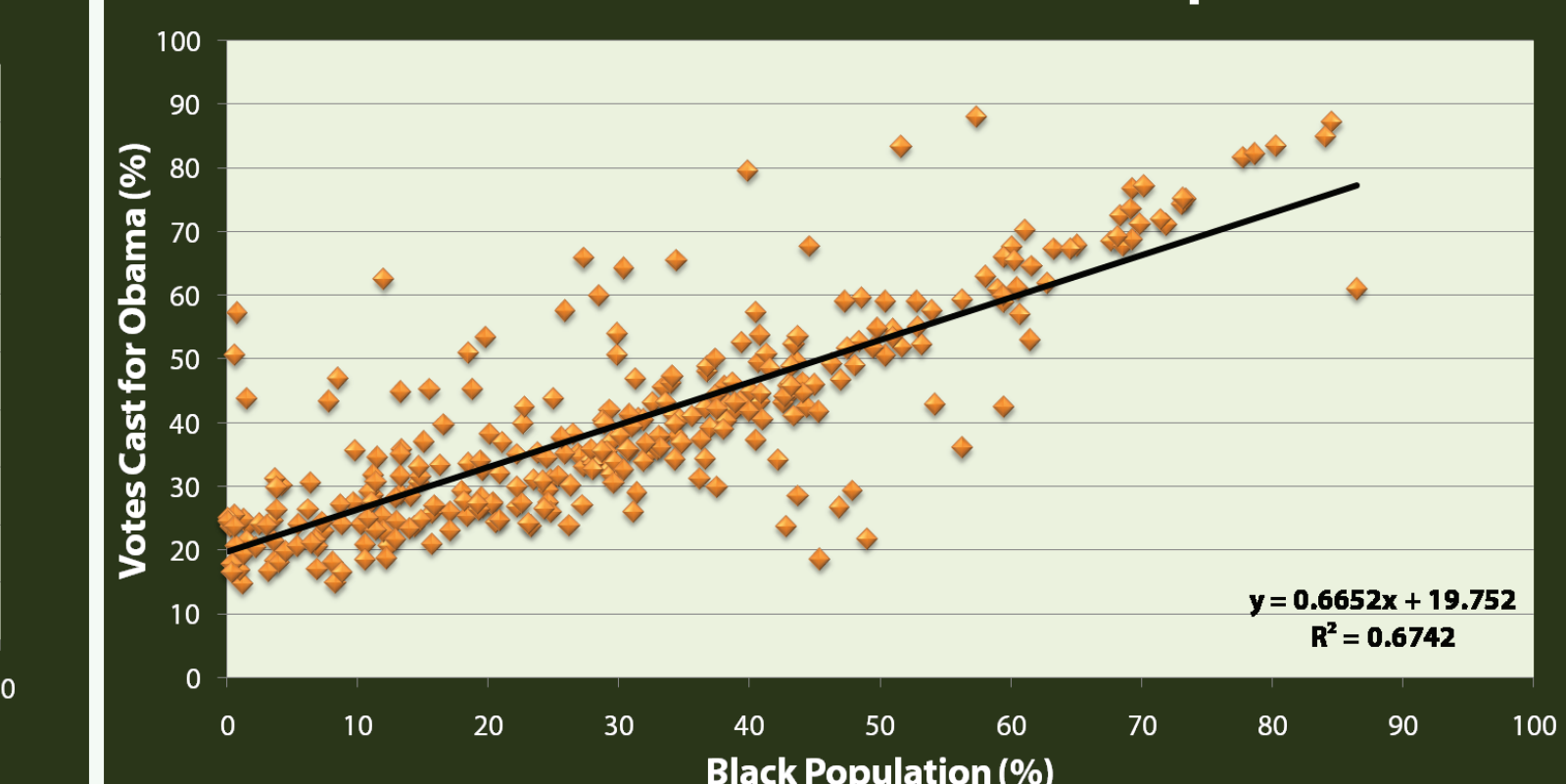
0 100 200 Miles

By visualizing the data in map form, general patterns of the relationship between dependent and independent variables can be seen. Maps of the Obama Vote, Black Population, and Poverty show striking similarities to each other. Each show highest concentrations of the variable running horizontally through the center of the study area. Maps of White population and Median Household Income show an inverse relationship to the percentage of Obama Votes. These maps show strengthen the results derived from Table 1.

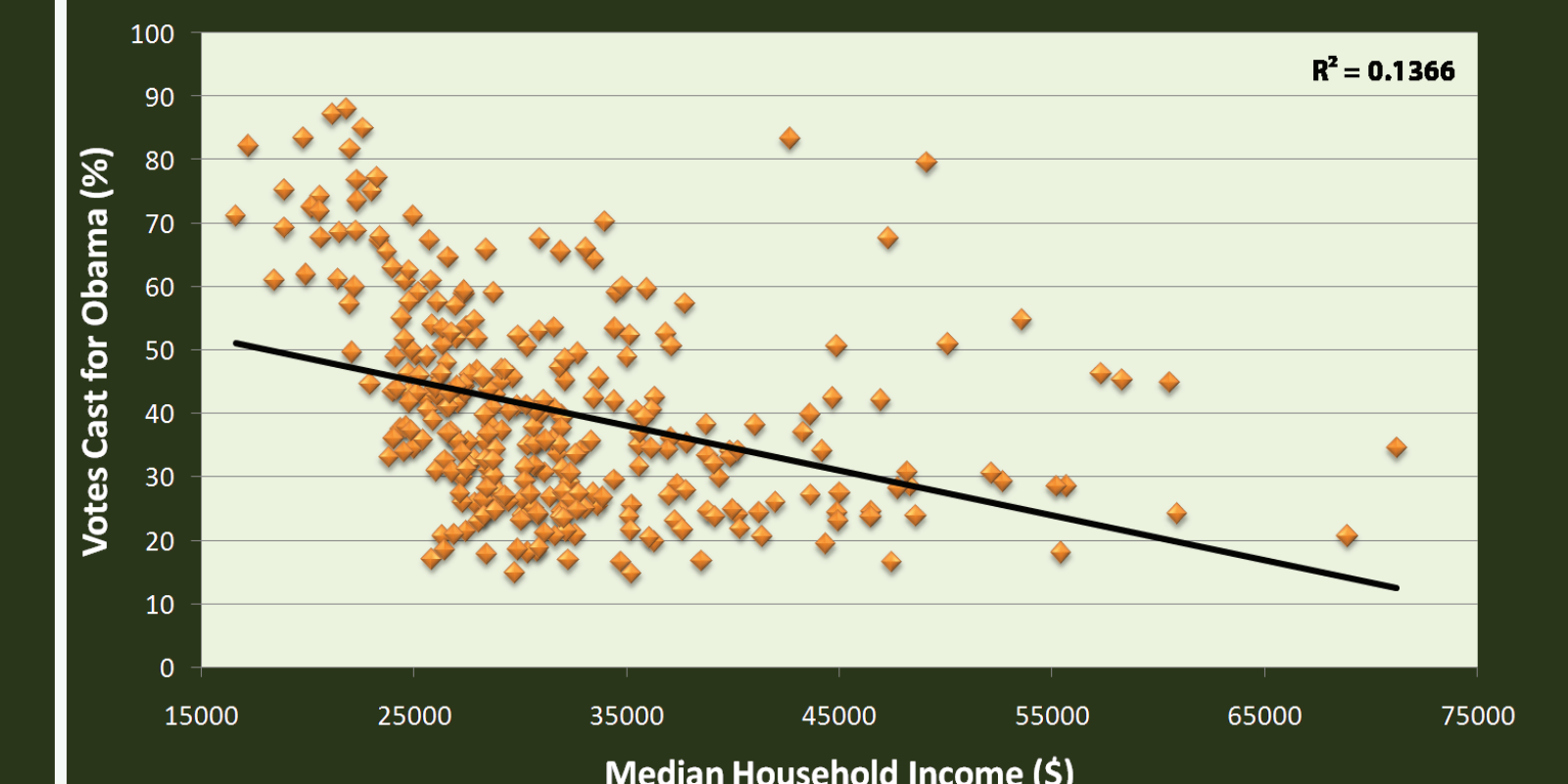
Percentage of County Votes Cast for Obama vs Percentage of County Population Below the Poverty Line



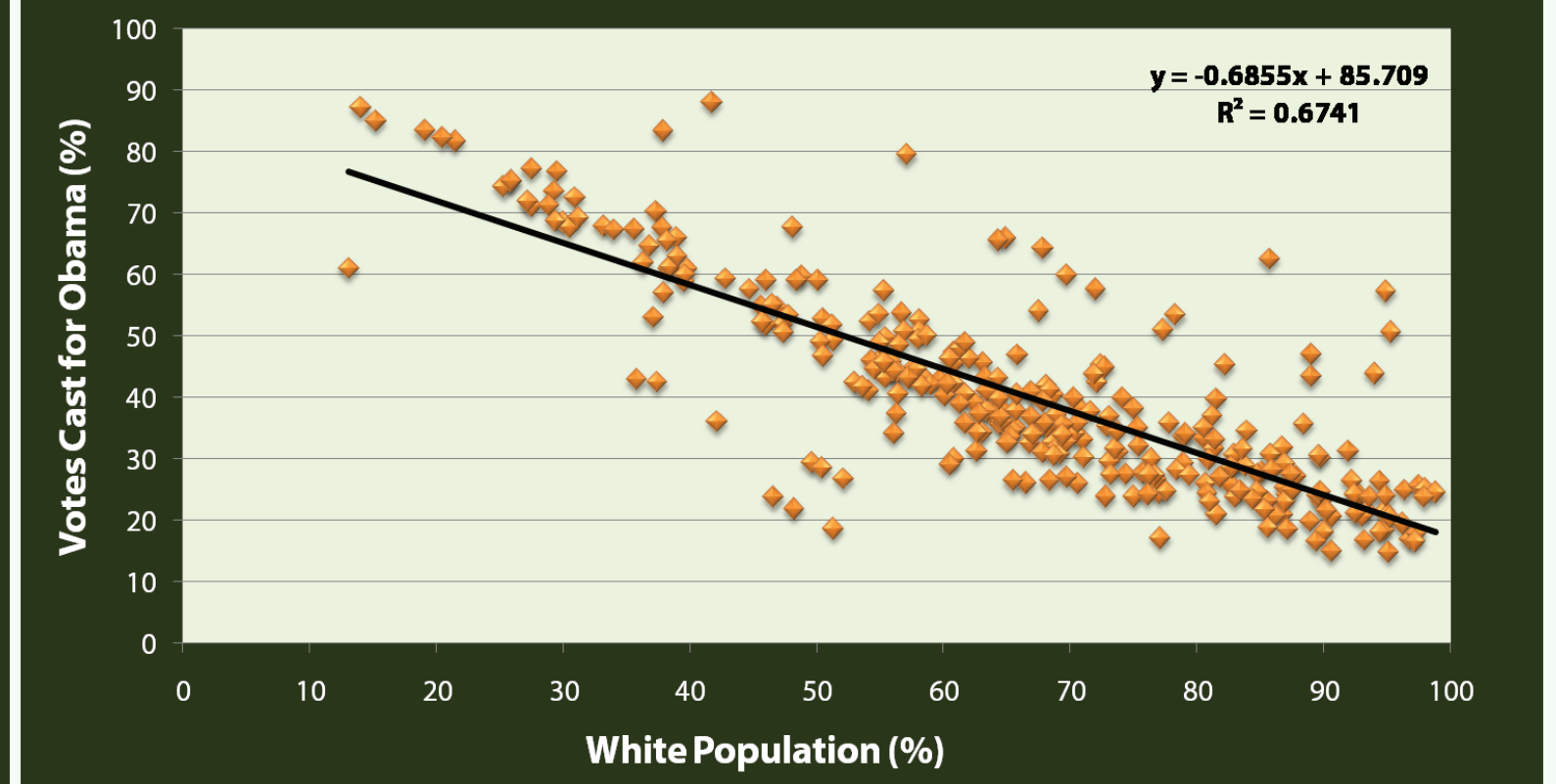
Votes Cast for Obama vs Black Population



Votes Cast for Obama vs Median Household Income



Votes Cast for Obama vs White Population

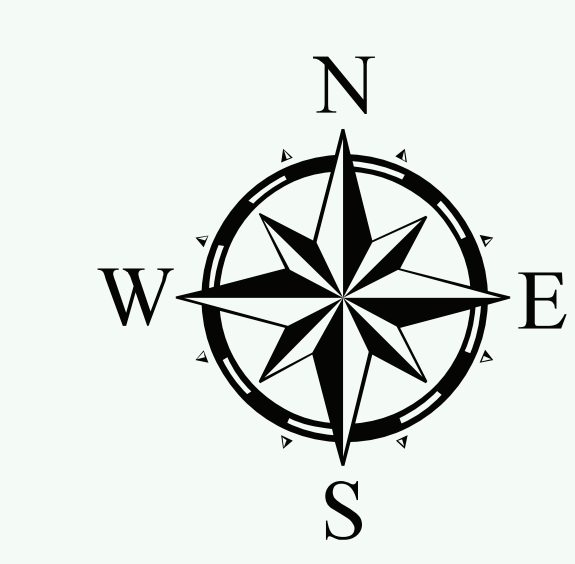
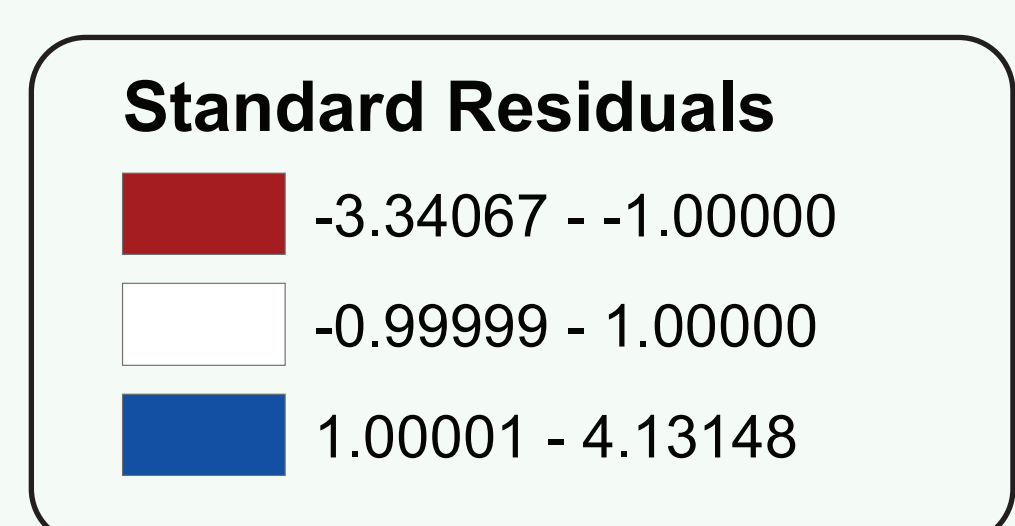
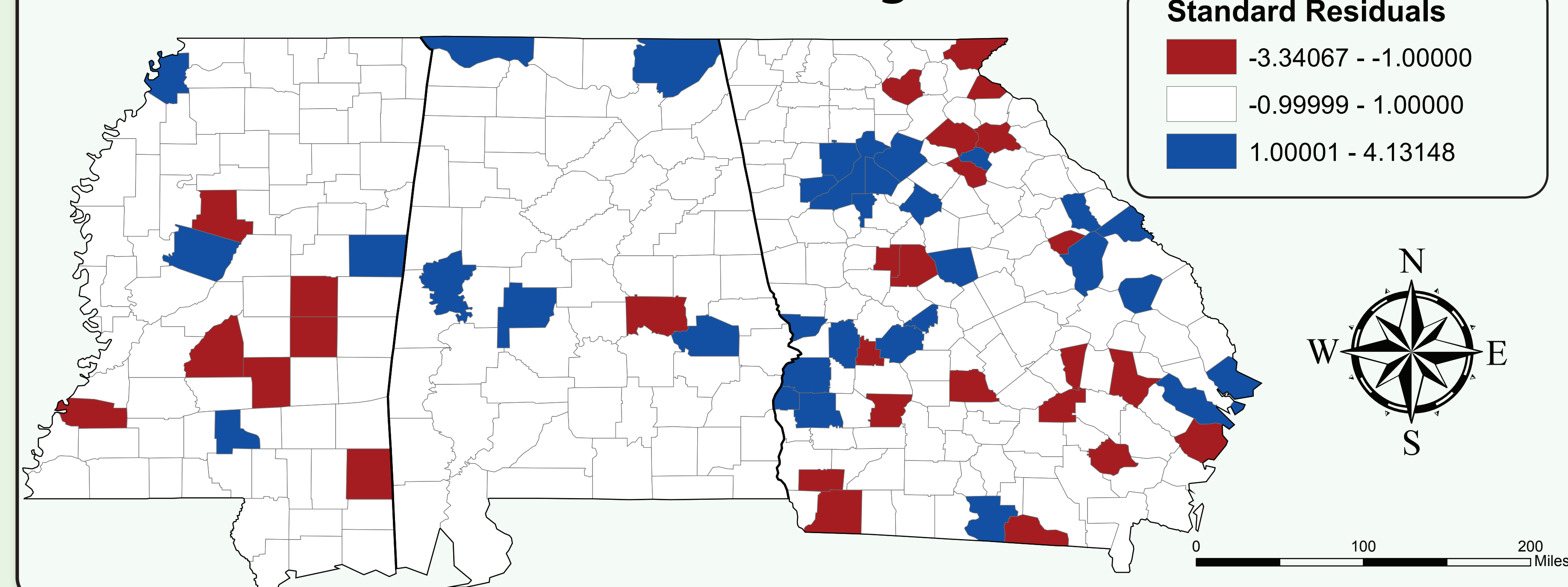


Plotting scatter grams of the dependent and independent variables (above) also shows the relationship between them. An upward sloping line indicates a direct relationship and a downward sloping line indicates an inverse relationship. The scatter grams above show direct relationships between the votes for Obama, black population, and population living in poverty. An inverse relationship exists between votes for Obama, white population, and median household income.

Table 1

All Variables			
Independent Variables (County Level)	b	Beta	Simple r
Black Population (%)	0.5925	0.7314	0.821
White Population (%)	-0.1266	-0.1517	-0.821
Population age 18-24 (%)	0.3691	0.0728	0.248
Population age 25-44 (%)	-0.5916	-0.0997	-0.278
Population age 45-65 (%)	0.8256	0.1204	-0.303
Population older than 65 (%)	-0.8819	-0.1567	0.012
Median Household Income (\$)	-0.0003	-0.1768	-0.370
Population Below Poverty (%)	-0.4341	-0.1673	0.602
Population with at Least Bachelors Degree (%)	0.0002	0.2200	0.016
a = 54.24	R = 0.855	R <sup>2</sup> = 0.731	
Effective Variables			
Independent Variables (County Level)	b	Beta	Simple r
Black Population (%)	0.4202	0.5187	0.821
White Population (%)	-0.3563	-0.4267	-0.821
Median Household Income (\$)	-0.0001	-0.0716	-0.370
Population Below Poverty (%)	-0.5006	-0.1929	0.602
a = 64.91	R = 0.8286	R <sup>2</sup> = 0.6866	

## Standardized Residuals from Regression



0 100 200 Miles

(left) When comparing the R and R<sup>2</sup> values for the data, the removal of insignificant variables is visible to the extent of the beta values of those independent variable removed. For example, the R values and the R<sup>2</sup> values decrease as the variables are removed. This makes sense because although I have considered the relationship insignificant, there is still a weak relationship that if included would account for slightly higher R and therefore R<sup>2</sup> values.

The residuals map (left) shows the residuals from regression for only those independent variables thought to have a significant effect on the outcome of Y. There is stronger emphasis placed on those residuals with values greater than 1 and less than -1. Residuals greater than 1 show areas that overestimate values for the dependent variable. Areas with residual values less than -1 are underestimated. The residuals were standardized by dividing each residual value by the standard error of estimate (SEE). Notice that most counties fall within one SEE.