Identifying South Carolina's Affluent and Deprived Counties: Computing with Standard Scores and Visualizing with Tableau Choropleth Maps

Dr. Coddy Carter

Research and Policy Initiatives

South Carolina Commission for Minority Affairs

October 2018

 Please direct correspondence to Coddy Carter via e-mail at ccarter@cfma.sc.gov.

Abstract

The research report sought to develop a methodology to expressly diagnose problem areas of socioeconomic deprivation in South Carolina. Standardized scores were used to develop ordinal categorizations in regard to poverty, unemployment, per capita income, and median household income. Ordinal categories included below average, average, and above average. The categories were based on raw values, means, and standard deviations of the selected variables. Poverty and unemployment were defined as negative socioeconomic variables because higher values indicated lower socioeconomic status. Above average poverty and unemployment indicated deprivation, while below average measures on these variables indicated affluence. Per capita and median household income were defined as positive socioeconomic variables such that higher values indicated higher socioeconomic status and lower values indicated deprivation. Above average ordinal categorization in per capita and median household income signaled affluence, while below average figures pinpointed deprivation. Anderson, Beaufort, Berkeley, Charleston, Dorchester, Greenville, Lexington, Richland, and York counties were found to be affluent. Allendale, Bamberg, Barnwell, Chester, Clarendon, Dillon, Fairfield, Lee, Marion, Marlboro, Orangeburg, and Williamsburg counties were found to be deprived. Choropleth maps indicated that deprived counties were concentrated in rural areas while affluent counties were clustered in metropolitan statistical areas. The findings of the spatial analysis suggested a need for future studies to examine the metropolitan nature of affluence and the rural pattern of deprivation.

Identifying South Carolina's Affluent and Deprived Counties: Computing with Standard Scores and Visualizing with Tableau Choropleth Maps

In the *FY 2017-18 Agency Accountability Report* (AAR) submitted by the South Carolina Commission for Minority Affairs, it was reported that the responsibility of the Research and Policy Initiatives division would be responsible for the diagnosis of socioeconomic deprivation. The present paper seeks to put forth a statistically-based diagnostic criterion for identifying the presence of either affluence or deprivation and create a conceptual foundation for future research from the Research and Policy Initiatives division of the agency.

Standardized Scores as Comparison Points

The statistical foundation for the diagnostic criterion is a standard score or z-score. A z-score is based on a calculation that incorporates a raw score, the arithmetic mean, and the standard deviation. Position measures such as z-scores allow one to observe where scores fall in comparison to a sample of other scores. A z-score can specifically be categorized as a norm-referenced, linear standard score (Mertler, 2007). The Scholastic Assessment Test (SAT) is an example of a norm-referenced test in that it utilizes positions measures to assess how students perform in comparison to their peers whether at the national, state, county, or any desired level. The position of the metrics in the present paper will determined by how each compares to sampled state averages with the sample consisting of South Carolina's 46 counties.

Z-scores are calculated by using the following equation:

$$z = \frac{x - \overline{x}}{s}$$

where: z = obtained z-score,

x = raw individual value for a county,

$\overline{x}$ = sample mean, and

s = sample standard deviation.

Bluman (2015) defined a z-score as the number of standard deviations that an individual value is above or below the mean. Scores within one standard deviation of the mean were considered to be average, with below average scores being equal to or less than -1 standard deviations from the mean ($z \leq -1$) and above average scores being equal or greater than +1 standard deviations from the mean (Mertler, 2007). Typically, 99% of z-scores ranged between -3.00 and +3.00 (Mertler, 2007). Negative z-scores indicated that a raw value was below the mean, while positive values indicated the opposite.

Z-scores, by definition and by having a range that includes both negative and positive number might present interpretation issues. First, standard deviation is not an intuitive statistic for a lay reader even when given the criterion for determining when scores are below average, average, and above average. Second, a negative z-score can be interpreted negatively and a positive score could be interpreted as being above average. For example, the respective z-scores of -0.4 and +0.4 appear to be different, but they would not actually differ by the below average/average/above average criterion described earlier in the report. Despite having different valences, both would be within one standard deviation of the mean, and therefore, average.

To address this issue and develop the diagnostic method for deprivation and affluence, z-scores will be transformed into t-scores. T-scores are more intuitive as the mean or middle of the distribution is 50 ($T = 50$). Rather, than 99% of scores ranging from -3 to 3 as observed with z-scores, 99% of t-scores should range from 20 to 80 (Mertler, 2007).

T-scores are calculated through the following equation:

$$T = [z \cdot 10] + 50$$

where: z = obtained z-score.

Finally, a diagnostic criterion is determined by observed t-scores. The proposed diagnostic criterion is similar to the below average/average/above average criterion recommended by Mertler (2007), but it is more stringent in extracting scores that would be considered average. Normally, 68.26% of scores would fall within the average category by virtue of being within one standard deviation of the mean. The present criterion is designed such that only 52% of scores would be considered average. This method allows fewer scores to be considered average and more scores to be either below or above average, which allows one to more readily determine which scores differ from the mean.

T-scores equal to and less than 42.6 ($T \leq 42.6$) are considered below average; these scores correspond to the first 23% of the area under a normal distribution. This can be interpreted as the 1st through 23rd percentiles. T-scores equal to or greater than 57.4 ($T \geq 57.4$) are considered above average; these scores correspond approximately to the area from 77 through 100 % under the normal distribution. This area can be interpreted as approximately the 77th through 99th percentiles. T-scores between 42.6 and 57.4 ($42.6 \geq T \geq 57.4$) are considered average.

Critical Mapping and Variables of Interest

Choropleth maps will be used to visually establish if there are any spatial patterns in either deprivation or affluence. Choropleth maps use arbitrary class boundaries to determine how data is visualized (Andrienko, Andrienko, & Savinov, 2001). Creating arbitrary class boundaries based on some statistical parameter facilitates the interpretation of maps by users (Kumar, 2003). In the case of the present report, these arbitrary classes are the ordinal categories (i.e., below average, average, above average) determined by t-scores, and these ordinal classes allow the interpretation of South Carolina's counties that are significant high, low, or average in the variables of interest.

Table 1 shows the choropleth map coding that will be used in diagnostic maps for negative socioeconomic variables. The negative socioeconomic variables that will be examined include percent below poverty and unemployment rate. These are considered negative socioeconomic variables because higher figures indicate lower socioeconomic status. Specifically, above average figures in the context of the diagnostic criterion indicate concern or problem areas. Below average figures signal affluence.

Table 1
Mapping Categories and Color Coding for Negative Socioeconomic Variables

Category	Color Code
Above Average	
Average	
Below Average	

Note: Negative socioeconomic variables may include poverty, unemployment, and high school dropout rates.

Table 2 shows the choropleth map coding for positive socioeconomic variables. Income is a positive socioeconomic variable; higher income indicates higher socioeconomic. Thus, below average income figures will indicate deprivation in regard to this variable, while higher incomes will indicate affluence. What should be noted between both Tables 1 and 2 is that the

color codes are reversed. This is done to ensure that blue coloring indicates affluence and red coloring indicates a problem or deprivation area across all maps.

Table 2

Mapping Categories and Color Coding for Positive Socioeconomic Variables

Category	Color Code
Below Average	
Average	
Above Average	

Note: Positive socioeconomic variables may include per capita income and median household income.

Results

Negative Socioeconomic Variables

Percent below Poverty

Figure 1 indicates that there are 12 counties that could be diagnosed as having significantly high poverty. Spatial patterns were observed with two clusters. One circular cluster included Lee, Chesterfield, Marlboro, Dillon, Marion, Williamsburg, and Clarendon counties. A contiguous cluster of poverty included Allendale, Bamberg, and Barnwell counties. Fairfield and Greenwood counties were neither part of the circular nor contiguous poverty clusters. Figure 1 also indicates that 10 counties were possibly affluence as measured by significantly low poverty rates. These counties could be described as consisting of three spatial clusters in addition to York County. The first spatial cluster included Anderson and Greenville counties. The second cluster included Lexington, Richland, and Kershaw counties. Finally, the third cluster appeared included Berkeley, Charleston, Dorchester, and Beaufort counties.

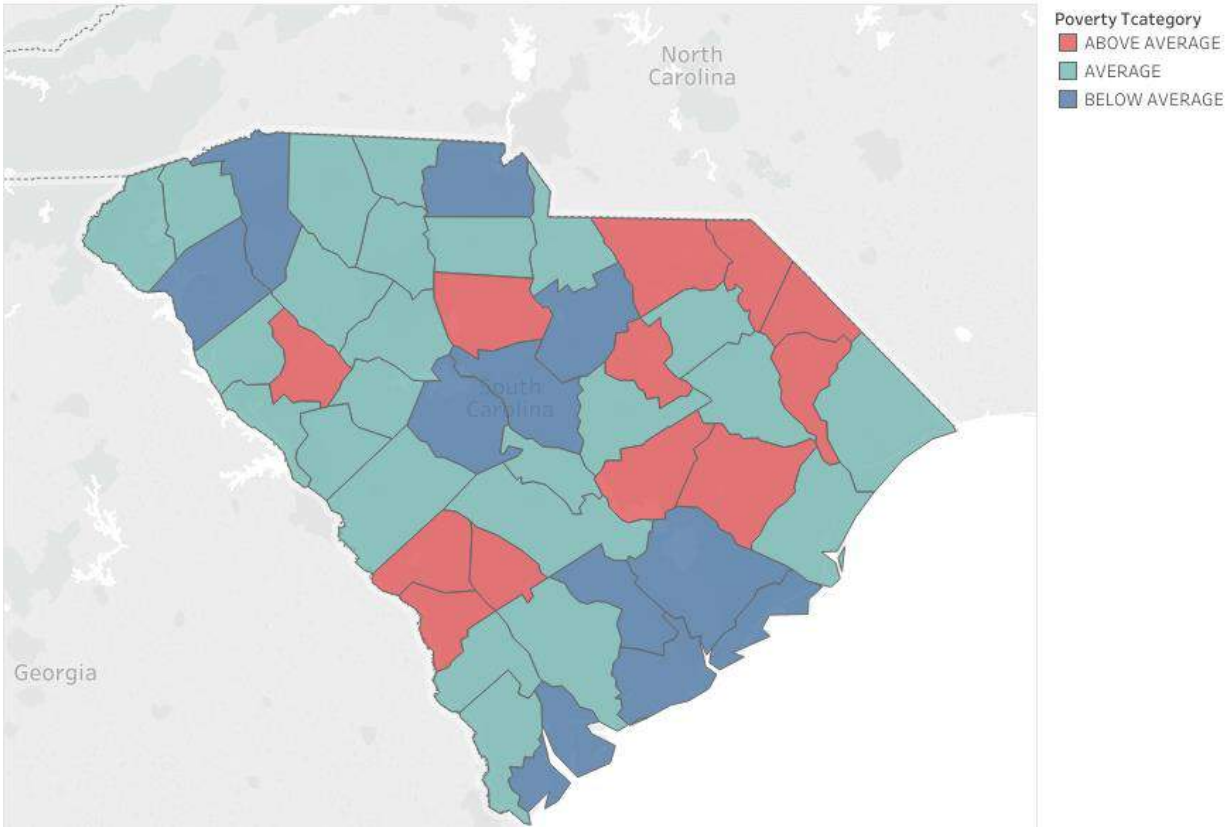

Figure 1. *Spatial Mapping of Percent below Poverty.*

Table 3 shows the counties with the highest and lowest percent below poverty, respectively. Allendale, Bamberg, Dillon, Lee, and Williamsburg counties had the highest percent below poverty. The lowest poverty counties included Beaufort, Berkeley, Dorchester, Lexington, and York.

Table 3
Highest and Lowest Percent below Poverty by County

	Top 5			Bottom 5	
Rank	County	Value	Rank	County	Value
1	Dillon	30.6%	42	Lexington	13.7%
2	Allendale	29.7%	43	Berkeley	13.2%
3	Williamsburg	29.3%	44	York	12.5%
4	Lee	28.2%	45	Beaufort	12.4%
5	Bamberg	27.1%	46	Dorchester	12.1%

Note: Sample mean percent below poverty rate across counties was 20.46% with a standard deviation of 4.89%. Source: American Community Survey, 2016.

Unemployment Rate

Figure 2 indicates that 10 counties were of concern in regard to unemployment rates. There were two contiguous spatial clusters and a circular cluster. The first contiguous cluster included Chester and Fairfield counties. The second contiguous cluster included Allendale, Bamberg, Barnwell, and Orangeburg counties. The final, circular cluster was similar to that observed in regard to poverty in its inclusion of Marlboro, Marion, Williamsburg, and Lee counties. The figure also indicated that 13 counties had below average unemployment. There appeared to be 4 clusters. The first cluster included Anderson, Greenville, and Spartanburg counties. The second cluster included Saluda, Newberry, Lexington, and Richland counties. The third cluster included Jasper and Beaufort counties. Finally, the fourth cluster included Charleston, Dorchester, and Berkeley counties. York County was not part of a cluster.

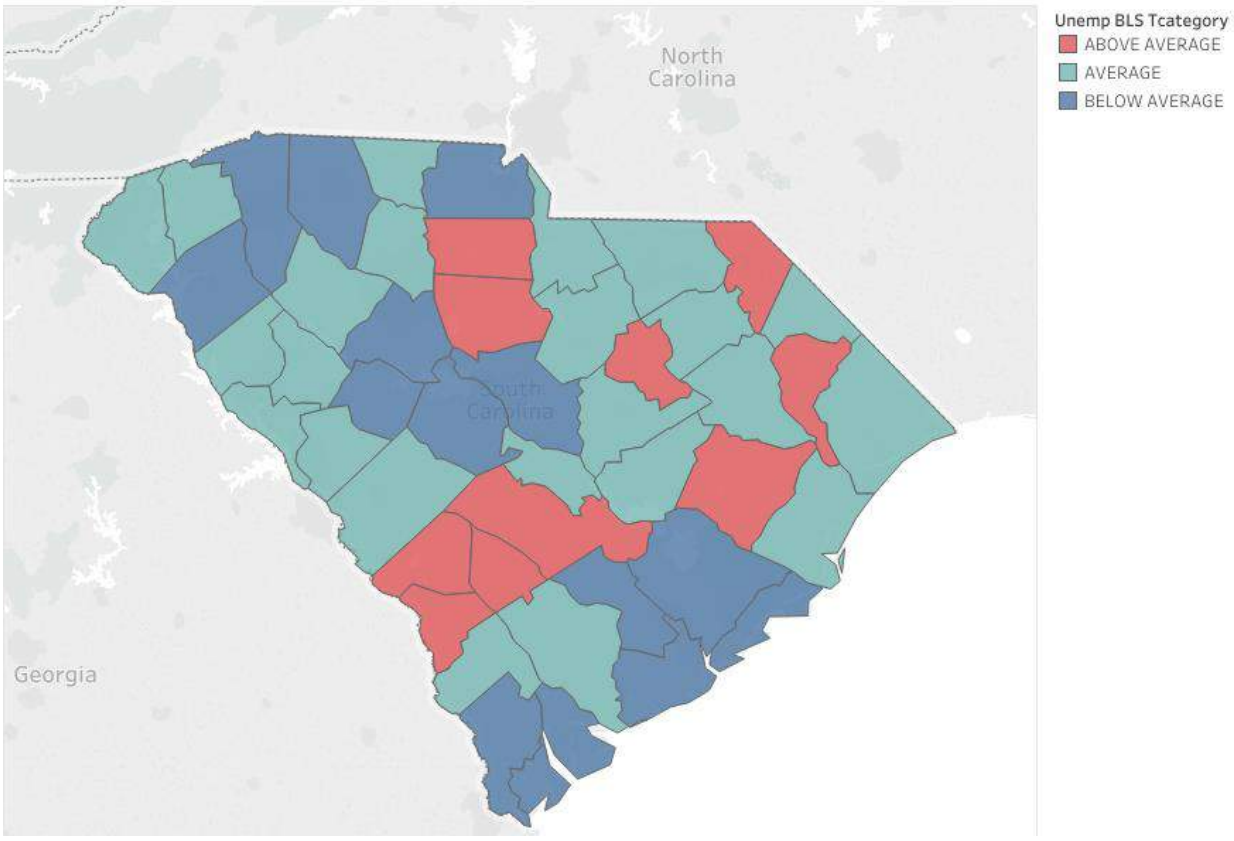

Figure 2. *Spatial Mapping of Unemployment Rate.*

Table 4 shows the counties with the highest and lowest respective unemployment rates. Allendale, Bamberg, Marion, Marlboro, and Orangeburg counties had the highest unemployment rates. Counties with the lowest rates of unemployment included Charleston, Greenville, Jasper, Lexington, and Newberry.

Table 4
Highest and Lowest Annual Unemployment Rates by County

	Top 5			Bottom 5	
Rank	County	Value	Rank	County	Value
1	Bamberg	9.0%	42	Newberry	4.4%
2	Orangeburg	8.8%	43	Jasper	4.3%
3	Marion	8.7%	44	Greenville	4.2%
4	Allendale	8.5%	45	Charleston	4.1%
5	Marlboro	8.4%	46	Lexington	4.1%

Note: Sample mean unemployment rate across counties was 5.84% with a standard deviation of 1.36%. Source: Bureau of Labor Statistics, 2016.

Positive Socioeconomic Variables

Per Capita Income

Figure 3 indicates that 10 counties were significantly low in average per capita income, and 11 counties were very high. Regarding counties particularly low in per capita income, there was one spatial pattern in addition to the standalone counties of Lee and Abbeville. The spatial cluster of per capita income deprivation included Allendale, Bamberg, Orangeburg, Clarendon, Williamsburg, Marion, Dillon, and Marlboro counties. Beaufort, York, Greenville, and Oconee counties were above average in per capita income but not part of any contiguous cluster. There were two contiguous clusters for above average per capita income. The first cluster included Aiken, Lexington, and Richland counties. The second cluster included Georgetown, Berkeley, Charleston, and Dorchester counties.

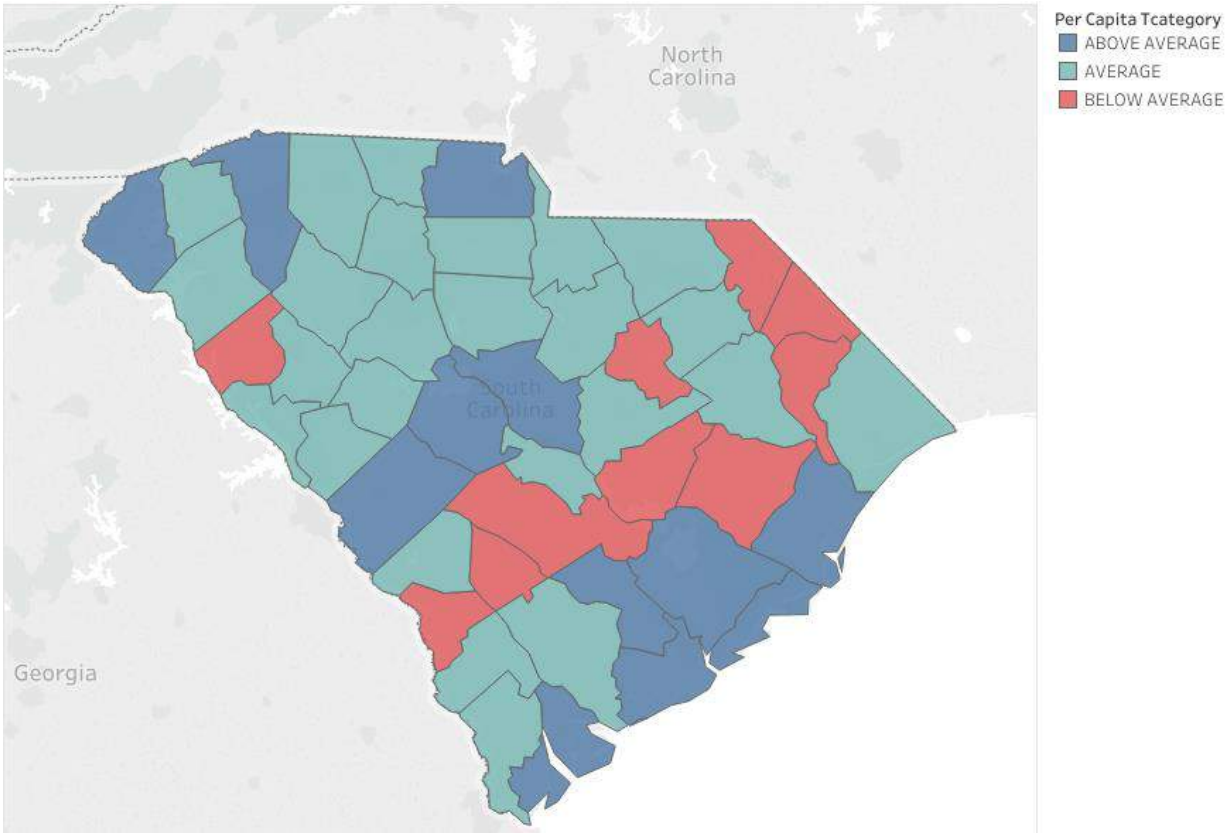

Figure 3. *Spatial Mapping of Per Capita Income.*

Table 5 shows the counties with the highest and lowest per capita incomes. Beaufort, Charleston, Greenville, Lexington, and York counties had the highest per capita incomes. Allendale, Dillon, Lee, Marlboro, and Williamsburg counties had the lowest per capita incomes.

Table 5
Highest and Lowest Per Capita Incomes by County

	Top 5			Bottom 5	
Rank	County	Value	Rank	County	Value
1	Beaufort	$33,877	42	Williamsburg	$16,650
2	Charleston	$33,700	43	Marlboro	$15,849
3	York	$28,830	44	Dillon	$15,729
4	Lexington	$28,141	45	Lee	$15,545
5	Greenville	$28,065	46	Allendale	$12,649

Note: Sample mean per capita income across counties was $22,105.91 with a standard deviation of $4,422.27. Source: American Community Survey, 2016.

Median Household Income

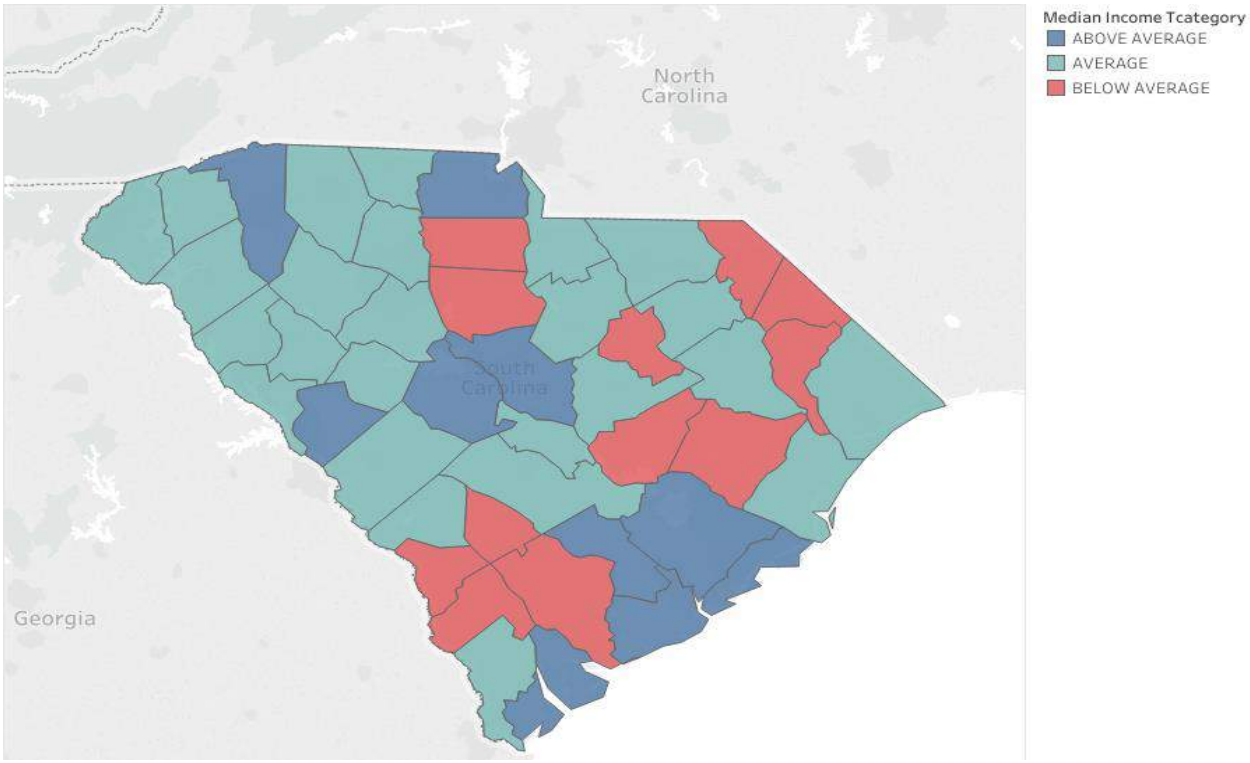

Figure 3. *Spatial Mapping of Median Household Income.*

Table 6 shows the counties with the highest and lowest median household incomes. Beaufort, Charleston, Dorchester, Lexington, and York counties had the highest median household incomes. Allendale, Dillon, Lee, Marion, and Williamsburg counties had the lowest median household incomes.

Table 6
Highest and Lowest Median Household Incomes by County

	Top 5			Bottom 5	
Rank	County	Value	Rank	County	Value
1	Beaufort	$59,227	42	Lee	$31,169
2	York	$56,482	43	Dillon	$30,955
3	Dorchester	$56,345	44	Marion	$30,562
4	Lexington	$55,412	45	Williamsburg	$28,494
5	Charleston	$54,931	46	Allendale	$24,817

Note: Sample mean-median household income across counties was $40,876.20 with a standard deviation of $8,472.85. Source: American Community Survey, 2016.

Discussion and Conclusions

Based on being coded in the choropleth maps as red across all four variables, Allendale, Bamberg, Lee, Marion, Marlboro, Williamsburg counties were unanimously deprived counties. Beaufort, Berkeley, Charleston, Dorchester, Greenville, Lexington, Richland, and York counties were coded as blue across all four variables, which means that they were unanimously affluent. Counties that were red or blue on at least two of the four variables could also be identified as deprived or affluent, respectively. Barnwell, Chester, Clarendon, Fairfield, and Orangeburg counties were red on at least two indicators, indicating deprivation. Anderson County, with blue coding in percent below poverty and unemployment rate, could be identified as affluent. In all there were a total of 11 deprived counties and nine affluent counties.

A few counties were identified as either deprived or affluent in particular variables but not unanimously or on at least two factors. Chesterfield, Greenwood, and Abbeville counties were coded red in either poverty or per capita income but not in more than one variable. These findings showed that there was possible socioeconomic concern in regard to those variables, but the three counties were not necessarily deprived. Jasper and Spartanburg counties were found to be below average in unemployment rates, but they were blue in that variable alone. Thus, they could not be considered affluent. Similarly, Aiken and Oconee were above average in per capita income. Taken together, these findings indicate a need for future studies. These counties might not have been deprived or affluent by the statistical parameters of the present report, but future studies could pinpoint the why scores fell where they did on particular variables.

There appeared to be affluent clusters centric to the metropolitan areas and deprivation clusters centric to rural areas. Berkeley, Charleston, and Dorchester counties, which were affluent, are all located in the Charleston-North Charleston Metropolitan Statistical Area (U.S. Census

Bureau, 2003). Greenville and Anderson counties are located in the Greenville-Anderson-Mauldin Metropolitan Statistical Area. Lexington and Richland counties are located in the Columbia Metropolitan Statistical Area. York County is located in the Charlotte-Concord-Gastonia Metropolitan Statistical Area centric to Charlotte, North Carolina. Beaufort County is located in the Hilton Head Island-Bluffton-Beaufort Metropolitan Statistical Area. The bulk of deprived counties, including Allendale, Bamberg, Barnwell, Clarendon, Dillon, Lee, Marion, and Williamsburg, were not located in core-based statistical areas. This indicated remoteness or distance from populated areas. The rural-metro nature of deprivation and affluence may be of interest for future studies that couple the findings of the present research report with more socioeconomic factors such as education. Education level could be a determinant of employment opportunities which affect the variables examined in the present paper. Thus, studying socioeconomic factors could provide a more complete picture of affluence and deprivation.

References

Andrienko, G., Andrienko, N., & Savinov, A. (2001, August). Choropleth maps: Classification revisited. In *Proceedings ICA* (pp. 1109-1219).

Bluman, A. G. (2015). *Elementary statistics: A step by step approach, a brief version*. (7th Ed.). New York: McGraw-Hill Education.

Kumar, N. (2003, October-December). Mapping spatial and statistical distributions in a choropleth map. *ArcUser*, 48-49.

Mertler, C. A. (2007). *Interpreting standardized test scores: Strategies for data-driven instructional decision making*. Thousand Oaks, CA: Sage Publications.

South Carolina Commission for Minority Affairs (2018, September). *Agency Accountability Report*. Submitted to South Carolina Department of Administration.

U.S. Census Bureau (2013, February). South Carolina Core Based Statistical Areas (CBSAs) and Counties. Retrieved October 18, 2016 at https://www2.census.gov/geo/maps/metroarea/stcbsa_pg/Feb2013/cbsa2013_SC.pdf.

U.S. Census Bureau. *2012-2016 American Community Survey (5-Year Estimates).* B19013: Median Household Income in the Past 12 Months (In 2016 Inflation-Adjusted Dollars).

U.S. Census Bureau. *2012-2016 American Community Survey (5-Year Estimates).* B19301: Per Capita Income in the Past 12 Months (In 2016 Inflation-Adjusted Dollars).

U.S. Census Bureau. *2012-2016 American Community Survey (5-Year Estimates).* S1701: Poverty Status in the Past 12 Months.

US Department of Labor, Bureau of Labor Statistics. (2016). Unemployment rates by county, not seasonally adjusted, annual.