School District Secession in East Baton Rouge

Caitlin Moroney

Abstract

In this project, I examine the spatial relationships between school performance and socioeconomic variables, such as race and income. I take as a case study for my analysis the region of East Baton Rouge Parish in southern Louisiana. I conduct two sets of analyses, one taking census block groups as the unit of analysis. These analyses consist of variations on the Local Moran's I which help to identify statistically significant spatial autocorrelation. The second set of analyses are regression models which use schools as the unit of analysis. I conduct both OLS regression and spatial regression, including the spatial error model and spatial lag model. From the Local Moran's analyses, I find statistically significant clusters of the white percentage of the population, median household income, and the collocation of these variables across census block groups. My regression analyses diverge in the selection of dependent variable. The models which seek to predict the percentage of economically disadvantaged students suggest that there is a statistically significant negative relationship between the percentage of white students and the percentage of economically disadvantaged students. The models which take school performance scores as the dependent variable produce evidence of a statistically significant postiive relationship between school performance and the percentage of white students.

Proposal

I am interested in examining the relationships between school performance, race, and socioeconomic indicators (including income and percentage of "economically disadvantaged"[1] students). Furthermore, I am interested in examining this data within the context of school district "secession."[2] I have chosen to focus on schools within East Baton Rouge Parish, Louisiana for two reasons: (1) there are stark differences among school districts with regard to school and school district performance,[3] and (2) residents in the southeast corner of the Parish recently voted to incorporate as the City of St. George with the intention of forming a new school district.[4] If successful in creating a new school district, St. George will be the fourth new district to "secede" from the original East Baton Rouge Parish School District in two decades.

I propose to investigate the following research questions:

- Is there evidence that suggests the clustering of racial groups in East Baton Rouge (EBR) Parish?
- Is there evidence of the clustering of income levels in EBR Parish?
- Is there evidence of a spatial relationship between race and income?
- Is there evidence to suggest that school performance is linked to the racial makeup of the student body, the percentage of economically disadvantaged students, and/or the percentage of students fully proficient in English?

[1] I use the Louisiana Department of Education's data on percentages of economically disadvantaged students. I have yet to find sufficient metadata explaining how they define "economically disadvantaged," but comparing current school and school district performance reports with past reports suggests that this may be the percentage of students on free or reduced lunch plans.

[2] See Lockhart (2019) for a thorough explanation of the school district "secession" phenomenon.

[3] Lussier (2019c)

[4] Vincent and Foster (2019)

Introduction

School district secession in the U.S. is propelling the re-segregation of public schools and school districts across the nation.[5] Two metrics commonly used to assess segregation include (1) "exposure" or "isolation," which measure the relative proportion of racial groups in distinct geographic areas (schools, school districts, census block groups, counties, etc.), and (2) "unevenness," which is concerned with the distribution of racial groups across these geographic regions.[6]

Another piece of this puzzle is the racial segregation between private versus public schools. This could be a significant factor fueling the difference in racial compositions between the total populations of residential areas and their public student populations. For example, if private schools tend to have significantly higher shares of white students than public schools in the same geographic region, this may help explain why public student populations tend to be significantly less white than the total residential populations of school districts. Unfortunately, an examination including private schools is beyond the scope of this analysis.

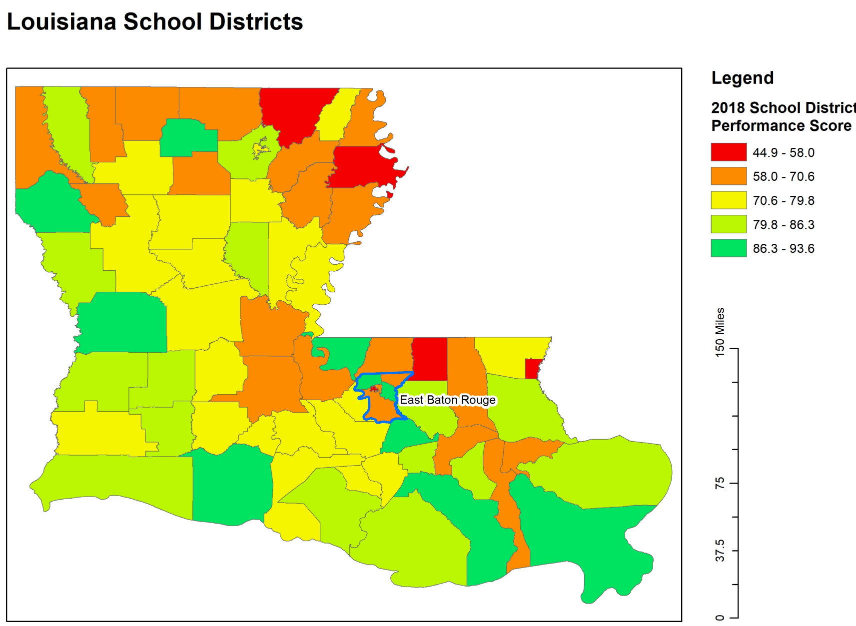

Figure 1: Louisiana school district performance for the 2017-2018 academic year. East Baton Rouge Parish is outlined in blue and labeled.

I have chosen to use East Baton Rouge Parish in Louisiana as a case study because this Parish is split into multiple school districts, which is not the case for many parishes in Louisiana (and counties across the South, generally speaking). Furthermore, there has been a marked difference in school performance among these districts, with Zachary Community School District consistently performing as the top-ranked school district in the state, Central Community School District regularly performing in the top five school districts in the state, and both East Baton Rouge Parish School District and City of Baker School District generally scoring in the bottom fifteen school districts in the state.[7] There is therefore a clear disparity among school districts in this region with respect to school quality. I am interested in investigating whether there are similar disparities among East Baton Rouge Parish school districts with regard to race and socioeconomic indicators.

[5] Lockhart (2019)
[6] Chang (2018)
[7] ("2018 District Performance Scores and Letter Grades" 2018)

East Baton Rouge Parish is no stranger to this school district “splintering” phenomenon. Originally, the entire Parish comprised a single school district, the East Baton Rouge Parish School System. Then, in 2003, both Zachary Community School District and the City of Baker School District formed, largely based on the incorporated regions of the two pre-existing cities. In 2007, another section of the East Baton Rouge Parish School District split off to form Central Community School District in the eastern part of the Parish. The Central Community School District was allowed to form only after the region incorporated into a city.[8] In 2012 and 2013, residents of the southeast section of East Baton Rouge Parish began petitioning to form a new school district.[9] Their initial efforts ultimately fell short, but after waiting the mandated time period, the group began petitioning again in March 2018 with the intention of following Central’s example: residents planned to incorporate as the City of St. George in order to subsequently form a new school district based on the incorporated region.[10] The movement to incorporate recently achieved success; the new City of St. George is now focused on creating a new school district.[11] According to U.S. Census survey estimates, there were approximately 16,300 students from kindergarten through 12th grade in census tracts fully or partly included within the City of St. George boundaries, 7,700 of whom were enrolled in private schools.[12]

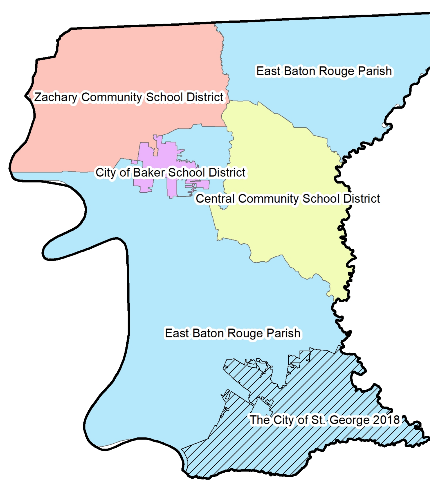

Figure 2: East Baton Rouge Parish school districts and the area of the City of St. George as proposed in 2018.

The rest of the paper is organized as follows: in the Methods section I cover the methodology used in this analysis, in Results section I report the results, and in the Discussion section I discuss the findings.

Methods

I gathered data from the Louisiana Department of Education, the U.S. Department of Education’s National Center for Education Statistics, East Baton Rouge GIS Open Data, the U.S. Census Bureau, and the City of St. George’s website.

Feature classes used in the final project include:

- East Baton Rouge Parish boundary (polygon)[13]
- U.S. Census Bureau Unified School Districts for Louisiana (polygon)

[8] Lussier (2019a)
[9] Lussier (2019a)
[10] Lussier (2019a)
[11] Vincent and Foster (2019)
[12] Lussier (2019b)
[13] City of Baton Rouge and Parish of East Baton Rouge (2020c)

- East Baton Rouge Parish schools (point)[14]
- St. George area (polygon)[15]
- East Baton Rouge Parish Census Block Groups (polygon)[16]

Tabular data joined to spatial feature classes include:

- East Baton Rouge public schools' enrollment by race, school performance scores, percentage of economically disadvantaged students, and percentage of fully English proficient students data for 2017-2018 academic year[17]
- School district performance scores for 2017-2018 academic year[18]

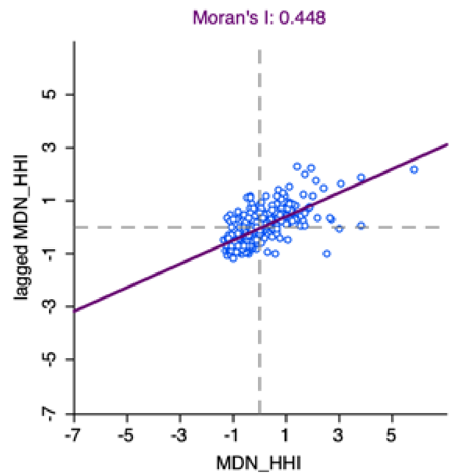

Figure 3: Lagged median household income versus median household income.

After downloading these various data sources, I then performed the following operations: data cleaning, file format conversion, new field calculation, attribute table editing, table joining, clipping, selecting by location, and selecting by attribute. I cleaned the tabular data extensively in R and converted the .xls files to .csv files so as to prevent significant issues in importing the data into my ArcMap project. For the City of St. George polygon layer, I downloaded the KML file from Google Maps linked on the St. George website; I then used ArcCatalog to convert the KML file to a layer package I could import into ArcMap. Once I added the layers in ArcMap and created a geodatabase for the project in which I saved the various files, I began calculating new fields based on existing variables in the attribute tables. Namely, I calculated the percentage of white students for each school in the Louisiana Department of Education data tables. I joined the Louisiana Department of Education school enrollment and performance data table to the East Baton Rouge Parish schools point feature class, keeping only the records with corresponding matches. (Colleges and universities, for example, were included in the East Baton Rouge Parish schools layer, but they were not relevant to my analysis). I joined the Louisiana Department of Education school district performance data table to the school districts feature class. Then, I clipped the U.S. Census Bureau Unified School Districts for Louisiana to only select those within the East Baton Rouge Parish boundary.

I conduct two analyses, one taking the census block groups as the units for analysis and one taking the schools as the units for analysis.

[14] City of Baton Rouge and Parish of East Baton Rouge (2020b)
[15] ("The Map. The City of St. George, Louisiana," n.d.)
[16] City of Baton Rouge and Parish of East Baton Rouge (2020a)
[17] ("2018 School Performance Scores and Letter Grades" 2018)
[18] ("2018 District Performance Scores and Letter Grades" 2018)

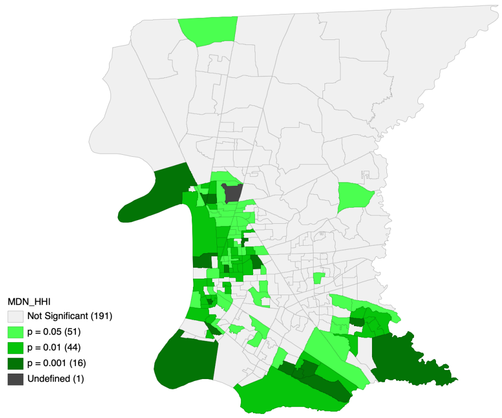

Figure 4: Significance map: median household income.

Census Block Groups

For the census block groups, I focus on median household income and the percentage of the population that identifies solely as white. I construct a spatial weights matrix using Queen's case first-order such that each census block group's neighbors are its immediately adjacent census block groups. I subsequently perform cluster analysis using Local Moran's. I conduct three variations: the first is a standard Local Moran's I which seeks to identify spatial autocorrelation within the study area. The second analysis is Local Moran's I with EB rates, which is an appropriate method when the variable of interest is a rate or proportion. The third analysis is Bivariate Local Moran's I, which identifies statistically significant spatial autocorrelation with respect to two variables. Further details are provided below in the Results section.

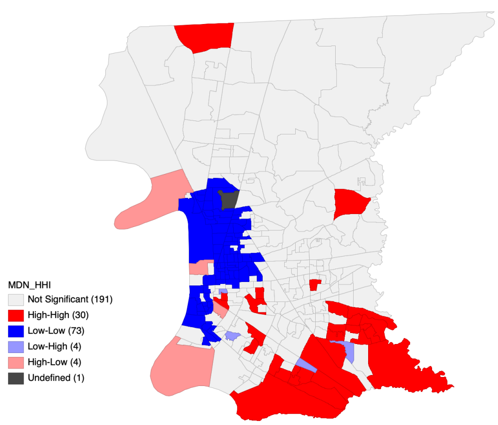

Figure 5: Cluster map: median household income.

Schools

To address my research questions focused on schools, I investigate the relationships between school performance scores, the percentage of the student body that identifies as white, the percentage of students who are economically disadvantaged, the percentage of students who are fully proficient in English, and the size of the school (as measured by the total number of non-pre-kindergarten students). I construct a spatial weights matrix using the distance band approach, where the threshold is set such that every point has at least one neighbor; I also apply inverse distance weighting.

I conduct two regressions, one with the percentage of economically disadvantaged students as the response variable, and the second with school performance scores as the response variable. Using GeoDa, I begin with ordinary least squares (OLS) regression and select the spatial weights matrix in order to view the spatial diagnostics. Based on the regression output, I then determine whether spatial regression is appropriate. If so, I apply either the spatial lag model or the spatial error model.

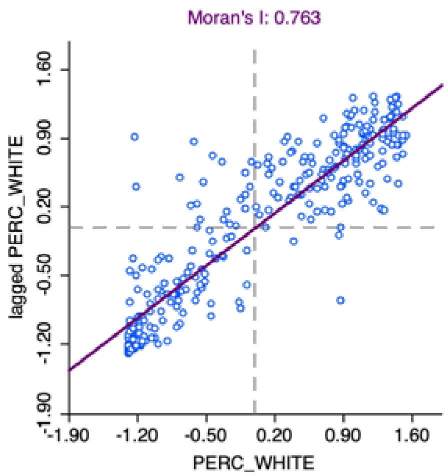

Figure 6: Lagged population percentage white versus population percentage white.

Results

Local Moran's I

Median Household Income

In Figure 4, we see statistically significant evidence of clustering in the southwestern and southeastern parts of the Parish. The general trend indicates spatial autocorrelation, as is evident in the LISA scatter plot (see Figure 3) and the cluster map (see Figure 5). In the southeastern foot of the Parish (where the City of St. George lies), we see significant clustering of census block groups with a higher-than-average median household income. In the central western area, we see mostly low-low clusters, indicating census block groups with lower-than-average median household income surrounded by neighboring census block groups with lower-than-average median household income. While there are a handful of high-low and low-high outliers, most of the significant census block groups show evidence of spatial autocorrelation. As mentioned above, the large pocket of high-high clustering overlaps largely with the City of St. George, whereas the low-low clusters are in the more densely populated part of the Parish which falls within East Baton Rouge Parish School District.

Population Percentage White

In Figure 7, we see statistically significant evidence of clustering in the southwestern and southeastern parts of the Parish. The general trend indicates spatial autocorrelation, as is evident in the LISA scatter plot (see Figure 6) and the cluster map (see Figure 8). In the southern and northeastern parts of the Parish, we see significant clustering of census block groups with a higher-than-average percentage of the population that

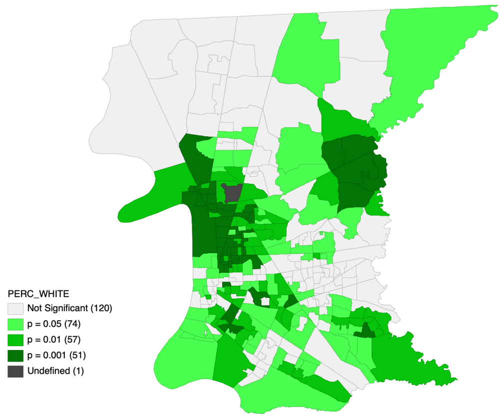

Figure 7: Significance map: population percentage white.

is white. In the central western area, we see mostly low-low clusters, indicating census block groups with a lower-than-average percentage of the population that is white surrounded by neighboring census block groups with a lower-than-average percentage of the population that is white. While there are a handful of statistically significant low-high outliers, most of the significant census block groups show strong evidence of spatial autocorrelation. The high-high clusters overlap to some extent with the City of St. George and Central Community School District. The low-low cluster is in the more densely populated part of the Parish which falls within East Baton Rouge Parish School District.

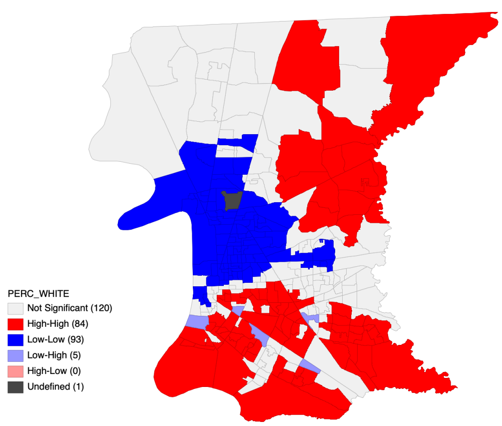

Figure 8: Cluster map: population percentage white.

Local Moran's I with EB rates

Because the percentage of the population that is white is a proportion, it is possible that my first analysis suffered from the identification of spurious outliers. This is due to the fact that variables that are rates or proportions have inherent variance instability. When features have different population sizes, features with smaller populations have more variable rates than features with larger populations. With EB standardization, the rates are adjusted so that rates for features with small populations are pulled closer to the overall average more so than the rates for features with large populations. Note that when the base populations across all features are relatively similar (meaning there is little variance), using local Moran's with EB rates may not be necessary. Although census block groups are statistically designed to contain roughly the same number of

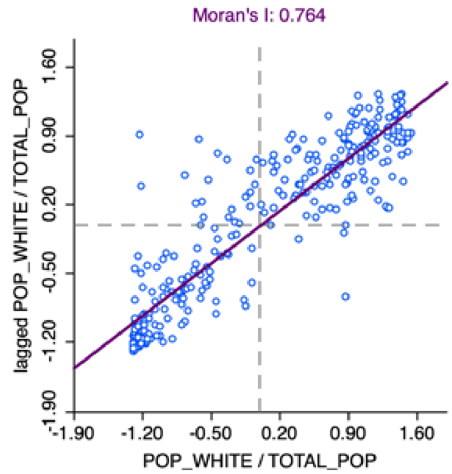

Figure 9: Lagged population percentage white versus population percentage white (with EB rates).

people, I redo the Local Moran's I analysis for the percentage of the population that identifies solely as white, applying the Empirical Bayes rate correction.

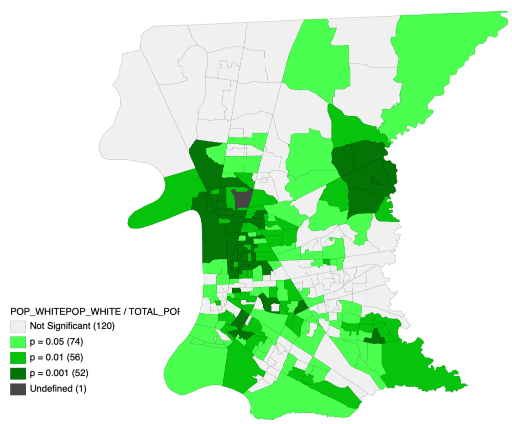

Figure 10: Significance map: population percentage white (with EB rates).

In Figure 10, we see statistically significant evidence of clustering in most parts of the Parish. The general trend indicates spatial autocorrelation, as is evident in the LISA scatter plot (see Figure 9) and the cluster map (see Figure 11). In the southern and northeastern parts of the Parish, we again see significant clustering of census block groups with a higher-than-average percentage of the population that is white. In the central western area, we see mostly low-low clusters, indicating census block groups with a lower-than-average percentage of the population that is white surrounded by neighboring census block groups with a lower-than-average percentage of the population that is white.

In other words, the Local Moran's I with EB rates produces identical to the regular Local Moran's I analysis of this variable. As mentioned above, this is not surprising given that there is unlikely to be significant variance in the census block group populations by design.

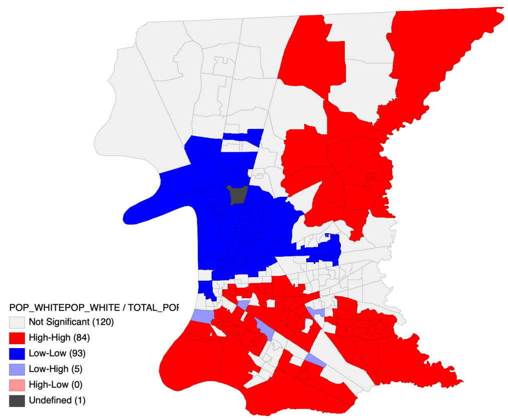

Figure 11: Cluster map: population percentage white (with EB rates).

Bivariate Local Moran's I

The bivariate Local Moran's I allows us to compare the spatial distribution of two variables, X and Y (in this case, X is median household income, and Y is population percentage white). For a contiguous feature set of polygons we can compare the value of X at the i^{th} feature to the values of Y at the neighboring features, $N(i)$. This analysis allows us to determine whether there is a statistically significant relationship between X and Y for each feature in the analysis.

In Figure 13, we see statistically significant evidence of clustering in most of the Parish, aside from the northwest corner. The general trend indicates spatial autocorrelation, as is evident in the LISA scatter plot (see Figure 12) and the cluster map (see Figure 14). In the southern and northeastern parts of the Parish, we see significant high-high clusters. These census block groups have a higher-than-average median household income and are neighbored by census block groups with a higher-than-average percentage of the population that is white. In the central western area, we see mostly low-low clusters, indicating census block groups with a lower-than-average median household income surrounded by neighboring census block groups with a lower-than-average percentage of the population that is white. While there are a handful of statistically significant low-high and high-low outliers, most of the significant census block groups show strong evidence of spatial autocorrelation. The high-high clusters overlap to some extent with the City of St. George and Central Community School District. The low-low cluster is in the more densely populated part of the Parish which falls within East Baton Rouge Parish School District.

Regression: Percentage Economically Disadvantaged Students

In order to answer my research questions about the relationships between school performance, race, and socioeconomic status in schools, I conduct two sets of regressions. The first seeks to determine whether the percentage of white students, the size of the school (as measured by the total number of non-pre-kindergarten students), and/or the percentage of students fully proficient in English can predict the percentage of economically disadvantaged students. The second analysis focuses on how the same explanatory variables can explain school performance scores. Note that for all of my hypothesis tests I am using a confidence level of 0.95; in other words, $\alpha = 0.05$.

OLS

I begin with OLS regression (see Figure 15), focusing first on the standard OLS output to determine whether the predictors I have chosen actually hold explanatory power for the percentage of economically disadvantaged

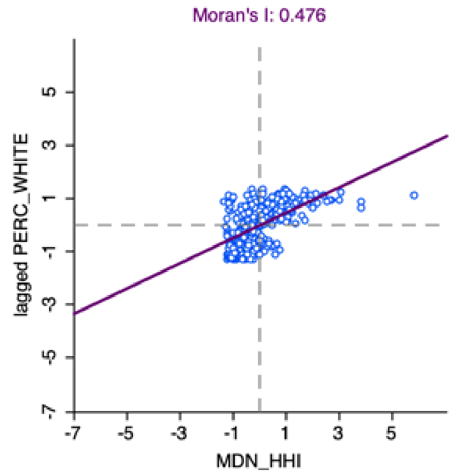

Figure 12: Lagged population percentage white versus median household income.

students. With my first regression, I can see that the percentage of students who are fully proficient in English is not a statistically significant predictor, with a p-value of 0.41656. The multicollinearity number, at 29.79, is quite high which is a further indication that at least one of the predictors is redundant. The other variables appear to be statistically significant, so I will run another OLS regression removing the percentage of students who are proficient in English as a predictor. For comparison, the adjusted R-squared for the first model is 0.767863, the AIC is -128.353, and the SBC is -119.417.

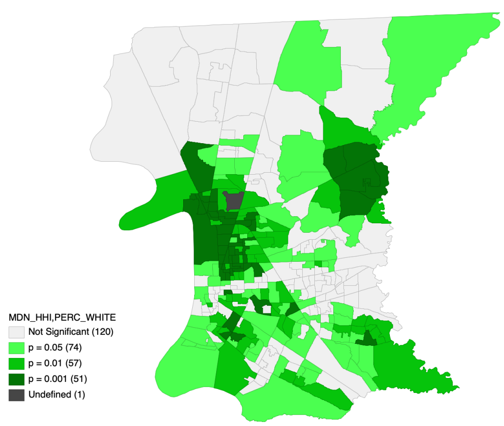

Figure 13: Significance map: lagged population percentage white versus median household income.

For the second OLS regression (see Figure 16), we see that the percentage of white students and the total number of non-pre-kindergarten students are both still statistically significant. The multicollinearity number has decreased significantly to just 4.51, which suggests that we do not have any redundant variables. We can also confirm that this model is better than our first model because the adjusted R-squared has increased from approximately 0.768 to 0.769; also, the AIC and SBC have both decreased to about -129.6 and -122.9, respectively. We can see from the coefficient estimates that the percentage of white students has a negative relationship with the percentage of economically disadvantaged students; in other words, as the share of white students increases, the share of economically disadvantaged students decreases, indicating a significant relationship between race and socioeconomic status in schools. This suggests that schools with a larger share of minority students tend to have a larger share of economically disadvantaged students than whiter schools. Next, we examine the diagnostics.

To determine whether the assumption of normality holds, we can look at the Jarque-Bera statistic, which has a p-value of 0.00000. This suggests that our residuals do not follow a normal distribution, which violates one of the basic assumptions for OLS regression. Although this is a serious issue, it is outside the scope of this paper to address non-normality by applying a different model.

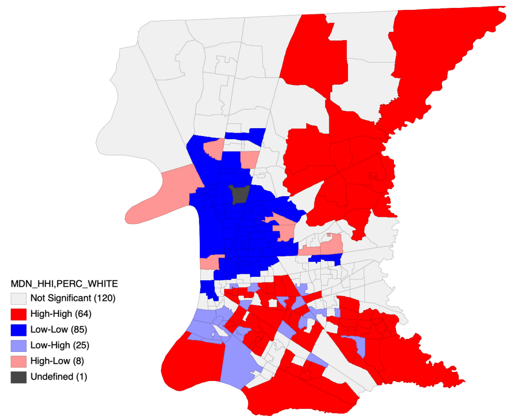

Figure 14: Cluster map: lagged population percentage white versus median household income.

Next, we consider the diagnostics for heteroskedasticity by assessing the results of the Breusch-Pagan, Koenker-Bassett, and White tests. With $\alpha = 0.05$, we fail to reject the null hypothesis (of homoskedasticity) for each of the tests. In other words, we do not have evidence to suggest a violation of the homoskedasticity assumption.

Finally, we examine the spatial dependence diagnostics. The Moran's I analysis produces a p-value of 0.00373, which suggests spatial autocorrelation of the residuals. Similarly, the Lagrange Multiplier (SARMA) has a p-value of 0.00253, which suggests that either the spatial error model or the spatial lag model may be appropriate here. However, neither the Lagrange Multiplier (error) nor the Lagrange Multiplier (lag) produces statistically significant results for $\alpha = 0.05$. In order to investigate further, I apply first the spatial error model and then the spatial lag model in order to determine whether this is an appropriate case for the use of either model.

Spatial Regression

The spatial error model (see Figure 17) adds a term to the OLS regression model whose coefficient, lambda, is the coefficient for the spatial error term. The spatial error model can be represented as follows:

$$y_i = x_i\beta + \lambda w_i \epsilon_i + u_i.$$

With $\alpha = 0.05$, the estimated lambda is just barely statistically significant, with a p-value of 0.04289. To determine whether this model really is an improvement over the original model, I will also note that the R-squared is now 0.786312 (it was 0.775823 for the first model), and the AIC and SBC have both decreased. These results suggest that the spatial error model is an improvement over the regular OLS model. However, the results of the likelihood ratio test, with a p-value of 0.15678, are not statistically significant; the results of this test are inconclusive on whether the spatial error model improves upon the OLS model. Next, I consider the spatial lag model.

The spatial lag model (see Figure 18) also adds a term to the OLS regression model; ρ is the coefficient for the spatial lag term. It captures any spatial dependence in the data. The spatial lag model can be represented as follows:

$$y_i = x_i\beta + \rho w_i y_j + u_i.$$

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh
Dependent Variable  :         ED_  Number of Observations:   69
Mean dependent var  :    0.758299  Number of Variables   :    4
S.D. dependent var  :    0.191239  Degrees of Freedom    :   65

R-squared           :    0.778105  F-statistic           :      75.977
Adjusted R-squared  :    0.767863  Prob(F-statistic)     : 3.23985e-21
Sum squared residual:    0.559953  Log likelihood        :     68.1766
Sigma-square        :  0.00861466  Akaike info criterion :    -128.353
S.E. of regression  :   0.0928152  Schwarz criterion     :    -119.417
Sigma-square ML     :  0.00811526
S.E of regression ML:   0.0900847

-----------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT       1.03967       0.130102        7.99114     0.00000
        PERC_WHITE     -0.720611      0.0589114       -12.2321     0.00000
          TotNonPK  -0.000104197    3.81877e-05       -2.72853     0.00817
        F_Fully_EP      -0.11139       0.136235       -0.81763     0.41656
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   29.787251
TEST ON NORMALITY OF ERRORS
TEST                  DF          VALUE            PROB
Jarque-Bera            2           62.3143         0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF          VALUE            PROB
Breusch-Pagan test     3           11.3293         0.01007
Koenker-Bassett test   3            4.2188         0.23879
SPECIFICATION ROBUST TEST
TEST                  DF          VALUE            PROB
White                  9            6.9514         0.64218

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
   (row-standardized weights)
TEST                          MI/DF       VALUE          PROB
Moran's I (error)             0.1394        3.4493       0.00056
Lagrange Multiplier (lag)       1           1.5542       0.21252
Robust LM (lag)                 1           9.2656       0.00234
Lagrange Multiplier (error)     1           3.9918       0.04572
Robust LM (error)               1          11.7032       0.00062
Lagrange Multiplier (SARMA)     2          13.2574       0.00132
============================== END OF REPORT ================================
```

Figure 15: OLS: percentage of economically disadvantaged students versus percentage of white students, total non-pre-kindergarten students, and percentage of students fully proficient in English.

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh
Dependent Variable  :         ED_  Number of Observations:   69
Mean dependent var  :    0.758299  Number of Variables   :    3
S.D. dependent var  :    0.191239  Degrees of Freedom    :   66

R-squared           :    0.775823  F-statistic           :     114.205
Adjusted R-squared  :    0.769029  Prob(F-statistic)     : 3.71136e-22
Sum squared residual:    0.565712  Log likelihood        :     67.8235
Sigma-square        :  0.00857139  Akaike info criterion :    -129.647
S.E. of regression  :   0.0925818  Schwarz criterion     :    -122.945
Sigma-square ML     :  0.00819872
S.E of regression ML:   0.0905468

-----------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT       0.93493      0.0226996        41.187      0.00000
        PERC_WHITE     -0.727065      0.0582334      -12.4854      0.00000
          TotNonPK  -0.000101488    3.79481e-05      -2.67438      0.00943
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   4.508061
TEST ON NORMALITY OF ERRORS
TEST                  DF          VALUE             PROB
Jarque-Bera            2           70.8995          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF          VALUE             PROB
Breusch-Pagan test     2            4.9923          0.08240
Koenker-Bassett test   2            1.7775          0.41118
SPECIFICATION ROBUST TEST
TEST                  DF          VALUE             PROB
White                  5            3.2372          0.66347

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
   (row-standardized weights)
TEST                          MI/DF       VALUE          PROB
Moran's I (error)             0.1243        2.9005       0.00373
Lagrange Multiplier (lag)       1           1.6857       0.19417
Robust LM (lag)                 1           8.7852       0.00304
Lagrange Multiplier (error)     1           3.1763       0.07471
Robust LM (error)               1          10.2758       0.00135
Lagrange Multiplier (SARMA)     2          11.9615       0.00253
============================== END OF REPORT ================================
```

Figure 16: OLS: percentage of economically disadvantaged students versus percentage of white students and total non-pre-kindergarten students.

With $\alpha = 0.05$, the estimated ρ is statistically significant, with a p-value of 0.00224, and it has a negative effect. Interestingly, the magnitude of the coefficient for percentage of white students increased by a lot, from 0.727 to 0.882 for the OLS model and the spatial lag model, respectively. To determine whether this model is an improvement over the original model, I will note that the R-squared is now 0.800964 (compared to 0.775823 for the OLS model), and the AIC and SBC have both decreased. The AIC is -134.808, and the SBC is -125.871, the lowest values for AIC and SBC across all of the models thus far. These results suggest that the spatial lag model is an improvement over the regular OLS model (and is possibly a better fit than the spatial error model). The results of the likelihood ratio test, with a p-value of 0.00745, further confirm this conclusion. Based on these tests, I conclude that the spatial lag model improves upon the OLS model.

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh
Spatial Weight      : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
Dependent Variable  :         ED_  Number of Observations:   69
Mean dependent var  :    0.758299  Number of Variables   :    3
S.D. dependent var  :    0.191239  Degrees of Freedom    :   66
Lag coeff. (Lambda) :    0.387223

R-squared           :    0.786312  R-squared (BUSE)      : -
Sq. Correlation     : -            Log likelihood        :   68.826033
Sigma-square        :  0.00781511  Akaike info criterion :    -131.652
S.E of regression   :   0.0884031  Schwarz criterion     :     -124.95

-----------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
-----------------------------------------------------------------------------
          CONSTANT       0.94935      0.0265007       35.8236     0.00000
        PERC_WHITE     -0.806172      0.0676507      -11.9167     0.00000
          TotNonPK  -0.000102088     3.4819e-05      -2.93196     0.00337
            LAMBDA      0.387223       0.191243       2.02476     0.04289
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                     DF      VALUE        PROB
Breusch-Pagan test                       2         4.0529     0.13180

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX :
EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
TEST                                     DF      VALUE        PROB
Likelihood Ratio Test                    1         2.0050     0.15678
============================== END OF REPORT ================================
```

Figure 17: Spatial error model: percentage of economically disadvantaged students versus percentage of white students and total non-pre-kindergarten students.

Regression: School Performance Scores

The next series of regressions takes school performance scores as the dependent variable to see whether racial population shares, school size, and/or the percentage of students fully proficient in English can explain differences in school performance.

OLS

With this first regression (see Figure 19), I can see that the percentage of students who are fully proficient in English is again not a statistically significant predictor, with a p-value of 0.79535. The multicollinearity number, as before, suggests that at least one of the predictors is redundant. The other variables appear to be statistically significant, so I will run another OLS regression removing the percentage of students who are proficient in English as a predictor. For comparison, the adjusted R-squared for the first model is 0.433922, the AIC is 572.169, and the SBC is 581.106.

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh
Spatial Weight      : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
Dependent Variable  :         ED_  Number of Observations:   69
Mean dependent var  :    0.758299  Number of Variables   :    4
S.D. dependent var  :    0.191239  Degrees of Freedom    :   65
Lag coeff.   (Rho)  :   -0.388018

R-squared           :    0.800964  Log likelihood        :     71.4038
Sq. Correlation     : -            Akaike info criterion :    -134.808
Sigma-square        :  0.00727924  Schwarz criterion     :    -125.871
S.E of regression   :   0.0853185

-----------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
-----------------------------------------------------------------------------
              W_ED_     -0.388018       0.126941      -3.05667     0.00224
           CONSTANT       1.25997       0.108862        11.574     0.00000
         PERC_WHITE     -0.882099      0.0766171      -11.5131     0.00000
           TotNonPK  -0.000109794    3.50596e-05      -3.13165     0.00174
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                     DF      VALUE        PROB
Breusch-Pagan test                       2         5.7009     0.05782

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX :
EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
TEST                                     DF      VALUE        PROB
Likelihood Ratio Test                    1         7.1605     0.00745
============================== END OF REPORT ================================
```

Figure 18: Spatial lag model: percentage of economically disadvantaged students versus percentage of white students and total non-pre-kindergarten students.

See Figure 20 for the second OLS model output. For $\alpha = 0.05$, the remaining variables remain statistically significant. The adjusted R-squared has increased to 0.441918, the first indication that this model outperforms the first OLS regression. Secondly, the AIC and SBC have both decreased: the AIC is now 570.241, and the SBC is 576.944, which is another indication that this model is a better fit. The percentage of white students and the total number of non-pre-kindergarten students both have a positive relationship with school performance; however, the coefficient magnitude is much larger for the former, suggesting that the percentage of white students has a larger impact on the school performance score. Next, I consider the regression diagnostics.

The results of the non-normality test suggest that the assumption of error normality is not violated. Furthermore, the results of the Breusch-Pagan and Koenker-Bassett tests are not statistically significant, which suggests that our assumption of homoskedasticity is valid. However, I will note that the White robust test results provide some evidence for the presence of heteroskedasticity.

The diagnostics for spatial dependence are similarly conclusive. Neither the Moran's I (error) nor the Lagrange Multiplier (SARMA) is statistically significant. Similarly, the Lagrange Multipliers for spatial lag and spatial error are not statistically significant. Based on these results, I conclude that my OLS model for school performance scores is sufficient; there is no need to perform spatial regression in this case. However, just to cover my bases, I apply a spatial error model and a spatial lag model to confirm whether either can improve upon the OLS model.

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh
Dependent Variable  :    S2018new  Number of Observations:   69
Mean dependent var  :     63.9565  Number of Variables   :    4
S.D. dependent var  :     19.6155  Degrees of Freedom    :   65

R-squared           :    0.458896  F-statistic           :     18.3749
Adjusted R-squared  :    0.433922  Prob(F-statistic)     : 9.62245e-09
Sum squared residual:     14365.7  Log likelihood        :    -282.085
Sigma-square        :     221.011  Akaike info criterion :     572.169
S.E. of regression  :     14.8664  Schwarz criterion     :     581.106
Sigma-square ML     :     208.198
S.E of regression ML:     14.4291

-----------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT       52.9814        20.8388        2.54244     0.01340
        PERC_WHITE       51.9689        9.43598        5.50753     0.00000
          TotNonPK     0.0136832     0.00611662        2.23706     0.02872
        F_Fully_EP      -5.68283        21.8211      -0.260429     0.79535
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   29.787251
TEST ON NORMALITY OF ERRORS
TEST                 DF          VALUE            PROB
Jarque-Bera           2            3.7897         0.15034

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                 DF          VALUE            PROB
Breusch-Pagan test    3            5.2043         0.15743
Koenker-Bassett test  3            3.6561         0.30108
SPECIFICATION ROBUST TEST
TEST                 DF          VALUE            PROB
White                 9           14.5164         0.10511

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
   (row-standardized weights)
TEST                         MI/DF       VALUE         PROB
Moran's I (error)            0.0343        1.3072        0.19114
Lagrange Multiplier (lag)      1           0.4761        0.49021
Robust LM (lag)                1           4.0190        0.04499
Lagrange Multiplier (error)    1           0.2413        0.62328
Robust LM (error)              1           3.7842        0.05174
Lagrange Multiplier (SARMA)    2           4.2603        0.11882
============================== END OF REPORT ================================
```

Figure 19: OLS: school performance score versus percentage of white students, total non-pre-kindergarten students, and percentage of students fully proficient in English.

Spatial Regression

The spatial error model results (see Figure 21) show that the coefficient lambda is statistically significant, with a p-value of 0.03995. The R-squared is approximately 0.486, whereas it was about 0.458 for the OLS model; this suggests that the spatial error term helps to explain more of the observed variance in school performance scores. Similarly, the AIC and SBC have both decreased from about 570.24 to 567.98 and about 576.94 to 574.68, respectively. This is further evidence that the spatial error model improves upon the OLS model. Finally, I consider the results of the likelihood ratio test, which are statistically insignificant. Overall, the results are therefore inconclusive on whether the spatial error model greatly improves upon the OLS model; however, because the R-squared increased, and both the AIC and SBC decreased, I would say that this model is a better fit than the OLS model.

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh
Dependent Variable  :    S2018new  Number of Observations:   69
Mean dependent var  :     63.9565  Number of Variables   :    3
S.D. dependent var  :     19.6155  Degrees of Freedom    :   66

R-squared           :    0.458332  F-statistic           :     27.9229
Adjusted R-squared  :    0.441918  Prob(F-statistic)     : 1.63381e-09
Sum squared residual:     14380.7  Log likelihood        :    -282.121
Sigma-square        :     217.889  Akaike info criterion :     570.241
S.E. of regression  :     14.7611  Schwarz criterion     :     576.944
Sigma-square ML     :     208.416
S.E of regression ML:     14.4366

-----------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT        47.638        3.61919        13.1626     0.00000
        PERC_WHITE       51.6396        9.28462        5.56185     0.00000
          TotNonPK     0.0138214     0.00605037        2.28439     0.02557
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   4.508061
TEST ON NORMALITY OF ERRORS
TEST                  DF          VALUE            PROB
Jarque-Bera            2            3.4561         0.17763

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF          VALUE            PROB
Breusch-Pagan test     2            3.2684         0.19511
Koenker-Bassett test   2            2.3219         0.31319
SPECIFICATION ROBUST TEST
TEST                  DF          VALUE            PROB
White                  5           13.4026         0.01988

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)             0.0341        1.1240         0.26103
Lagrange Multiplier (lag)       1           0.4625         0.49647
Robust LM (lag)                 1           3.9296         0.04744
Lagrange Multiplier (error)     1           0.2388         0.62511
Robust LM (error)               1           3.7059         0.05422
Lagrange Multiplier (SARMA)     2           4.1684         0.12441
============================== END OF REPORT ==============================
```

Figure 20: OLS: school performance score versus percentage of white students and total non-pre-kindergarten students.

The results from the spatial lag model hypothesis tests (see Figure 22) are more conclusive: this model does not appear to be an appropriate fit for the data. While the R-squared (0.460999) is slightly larger than that of the OLS model (R-squared = 0.458332), the AIC increased from 570.24 to 572.00, and the SBC increased from 576.94 to 580.93, which both suggest a worse model fit. Furthermore, the ρ coefficient is relatively small in magnitude and not statistically significant ($p = 0.57263$). The likelihood ratio test is also not statistically significant, with a p-value of 0.6194. Therefore, I conclude that the spatial lag model does not provide a better fit than OLS.

Discussion

For each of the Local Moran's I analyses, we can see statistically significant evidence of spatial autocorrelation. There is significant clustering of median household income, such that the City of St. George region is comprised mainly of census block groups with higher than average median household income, whereas the denser urban area in the central western part of the Parish features clustering of census block groups with lower than average median household income. Refer back to Figure 5. Similarly, both the regular Local Moran's and Local Moran's with EB rates analyses of the percentage of the population that identifies only as white find significant clustering. See Figures 8 and 11, respectively. From these maps, it is evident that census block groups with a higher than average white percentage of the population tend to cluster in the eastern and southern parts of the Parish. Census block groups with a lower than average white percentage of the population tend to group in the central western part of the Parish. Based on these findings, it is unsurprising that my Bivariate Local Moran's I results (see Figure 14) show that census block groups surrounded by census block groups with a higher than average white percentage of the population tend to have a higher than average median household income. Likewise, census block groups neighbored by census block groups with a lower than average white percentage of the population tend to have lower than average median household income.

Taken together, the results of my census block analyses show that there is a statistically significant clustering of racial groups in East Baton Rouge Parish; there is a statistically significant clustering of income groups in the Parish; and that these two clustering trends are collocated, such that relatively richer census block groups tend to be relatively whiter and relatively poorer census block groups tend to have a larger share of minorities. While the relationship between race and income (and the spatial patterns of these variables) is well-documented and considered common knowledge, what I find interesting about these results is that the whiter, more affluent census block groups tend to overlap with what is currently Central School District (a fragmented district) and the City of St. George (soon to form a new school district). These findings lend evidence to the theory that splintering school districts tend to be whiter and wealthier than the original intact school district.

From my regression models, I can draw conclusions about the relationships between school performance and socioeconomic variables. The best model for the regressions taking the percentage of economically disadvantaged students as the dependent variable, according to AIC and SBC values, is the spatial lag model. This model explains about 80.1% of the observed variance in the percentage of economically disadvantaged students, which is quite high. Based on this model, the coefficient for the percentage of white students is approximately -0.882, after controlling for school size and spatial autocorrelation. For each one percentage point increase in the percentage of white students, the percentage of economically disadvantaged students decreases by about 0.882 percentage points, ceteris paribus. This result suggests that the positive relationship between the white share of the population and income exists in schools as it does in the full population, which is perhaps not surprising.

The second set of regressions focus on school performance as the dependent variable. Here, the best performing model was the spatial error model, with the lowest AIC and SBC values as well as the highest R-squared ($R^2 = 0.486$). Controlling for spatial dependence in the error term as well as school size, we see that a one percentage point increase in the percentage of white students results in a 59.9 point increase in the school performance score, on average. Considering that the average school performance score is 63.96 with a standard deviation of approximately 19.62, we can say that the share of white students may have a large

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh
Spatial Weight      : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
Dependent Variable  :    S2018new  Number of Observations:   69
Mean dependent var  :   63.956522  Number of Variables   :    3
S.D. dependent var  :   19.615459  Degrees of Freedom    :   66
Lag coeff. (Lambda) :    0.391138

R-squared           :    0.485807  R-squared (BUSE)      : -
Sq. Correlation     : -            Log likelihood        : -280.990227
Sigma-square        :     197.844  Akaike info criterion :      567.98
S.E of regression   :     14.0657  Schwarz criterion     :     574.683

-----------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
-----------------------------------------------------------------------------
          CONSTANT       45.6869       4.23019       10.8002     0.00000
        PERC_WHITE       59.9077       10.7865       5.55395     0.00000
          TotNonPK     0.0145464    0.00553762       2.62683     0.00862
            LAMBDA      0.391138      0.190398       2.05431     0.03995
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                    DF      VALUE        PROB
Breusch-Pagan test                      2         2.1259     0.34543

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX :
EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
TEST                                    DF      VALUE        PROB
Likelihood Ratio Test                   1         2.2609     0.13267
============================== END OF REPORT ================================
```

Figure 21: Spatial error model: school performance score versus percentage of white students and total non-pre-kindergarten students.

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh
Spatial Weight      : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
Dependent Variable  :    S2018new  Number of Observations:   69
Mean dependent var  :     63.9565  Number of Variables   :    4
S.D. dependent var  :     19.6155  Degrees of Freedom    :   65
Lag coeff.   (Rho)  :   -0.114147

R-squared           :    0.460999  Log likelihood        :    -281.997
Sq. Correlation     : -            Akaike info criterion :     571.995
Sigma-square        :     207.389  Schwarz criterion     :     580.931
S.E of regression   :      14.401

-----------------------------------------------------------------------------
       Variable      Coefficient     Std.Error      z-value    Probability
-----------------------------------------------------------------------------
        W_S2018new     -0.114147      0.202322     -0.564187     0.57263
          CONSTANT       54.3903       12.1341       4.48245     0.00001
        PERC_WHITE       54.9639       11.6139       4.73259     0.00000
          TotNonPK     0.0138227    0.00590525       2.34074     0.01925
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                     DF      VALUE        PROB
Breusch-Pagan test                       2         3.0641     0.21609

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : EBR_selected_public_elem_mid_high_schools_onlyEBRRSAhigh_SWM
TEST                                     DF      VALUE        PROB
Likelihood Ratio Test                    1         0.2467     0.61940
============================= END OF REPORT ================================
```

Figure 22: Spatial lag model: school performance score versus percentage of white students and total non-pre-kindergarten students.

effect on school performance. However, it is important to note that this model explains less than 50% of the observed variation in school performance scores. Overall, these findings suggest that whiter schools tend to perform better, on average, than schools with larger shares of minority students. Though this trend is also well-documented in news reporting, I find it noteworthy that the data and statistical tests do, in the case of East Baton Rouge Parish, provide evidence in support of this claim. I cannot say definitively based on my analysis whether the formation of a St. George school district will result in another well-performing splinter district and a simultaneous dip in the East Baton Rouge Parish School District average school performance. However, I believe these findings, which provide evidence of racial and income segregation as well as a significant linkage between race and school performance, are relevant to the current debate on the incorporation of the City of St. George and the potential formation of a new school district.

Citations

"2018 District Performance Scores and Letter Grades." 2018. Louisiana Department of Education. https://www.louisianabelieves.com/resources/library/performance-scores.

"2018 School Performance Scores and Letter Grades." 2018. Louisiana Department of Education. https://www.louisianabelieves.com/resources/library/performance-scores.

Chang, Alvin. 2018. "The Data Proves That School Segregation Is Getting Worse." *Vox Media*, March. https://www.vox.com/2018/3/5/17080218/school-segregation-getting-worse-data.

City of Baton Rouge, and Parish of East Baton Rouge. 2020a. "Census Block Group. EBRGIS Open Data." December 2020. https://data-ebrgis.opendata.arcgis.com/datasets/census-block-group.

———. 2020b. "School. EBRGIS Open Data." December 2020. https://data-ebrgis.opendata.arcgis.com/datasets/school.

———. 2020c. "Parish Boundary. EBRGIS Open Data." September 24, 2020. https://data-ebrgis.opendata.arcgis.com/datasets/parish-boundary.

Lockhart, P. R. 2019. "Smaller Communities Are 'Seceding' from Larger School Districts. It's Accelerating School Segregation." *Vox Media*, September. https://www.vox.com/platform/amp/2019/9/6/20853091/school-secession-racial-segregation-louisiana-alabama.

Lussier, Charles. 2019a. "In Second Round of St. George Fight, Schools Downplayed but Nevertheless a Driving Force." *The Advocate*, March. https://www.theadvocate.com/baton_rouge/news/education/article_36570154-5163-11e9-8526-4bad11495733.html.

———. 2019b. "A St. George School District Could Start Off with Ample Funding, but Costs Will Grow If Popular." *The Advocate*, September. https://www.theadvocate.com/baton_rouge/news/education/article_c16b3896-dbab-11e9-9798-5f6d6b869334.html.

———. 2019c. "Baton Rouge Area Has 4 of the Top 10 Public School Districts in Louisiana." *The Advocate*, November. https://www.theadvocate.com/baton_rouge/news/education/article_754eed00-009e-11ea-83b1-771344e9c27d.html#tncms-source=infinity-scroll-summary-sticky-siderail-latest.

"The Map. The City of St. George, Louisiana." n.d. http://stgeorgelouisiana.com/map/.

Vincent, Mykal, and Kevin Foster. 2019. "City of St. George Incorporates." *WAFB*, October. https://www.wafb.com/2019/10/13/city-st-george-incoporates/.