Elizaveta Gonchar
CP 6521
May 3, 2021

Predicting Educational Attainment in St. Louis

OVERVIEW

My research set out to answer the following question: can spatial disparities in educational attainment in St. Louis be predicted and what factors contribute to the outcomes? I am interested in determining whether the data may be used to predict the forces of the Delmar Divide[1] on educational attainment. The motivation and basis of my analysis is from Ansong et al. (2015). In their paper, the authors implemented a geographically weighted regression (GWR) to estimate the district-level factors that contributed to the spatial inequality in academic achievement in Ghana. Identifying factors that correlate with educational attainment is important as they would guide policymakers in developing intervention programs that would promote higher educational attainment in those portions of the population which may not have the resources available to attain the education without support.

For my research, I set out to identify factors which may be related to educational attainment. While there are well-documented socioeconomic and demographic factors that are known to be linked to educational attainment, I set out with the intent of incorporating a spatial approach as well as incorporating variables which may proxy for school quality. My prior going into this analysis was that incorporating school quality would improve model performance but, through my exploratory regressions, I identified that there was too much collinearity to successfully implement such a measure via a GWR. I was successful in identifying factors that are correlated with educational attainment and I present the results of my analysis below.

DATA

I provide a table summarizing the data used for this project at the end of this document. The study area considered in my analysis consists of St. Louis City and St. Louis County. Much of the data used, particularly census and quality measures, pertain to 2019. For my analysis, I had 1,052 census block groups and 23 school districts. Below I provide descriptive statistics of the variables used in the final estimation approach.

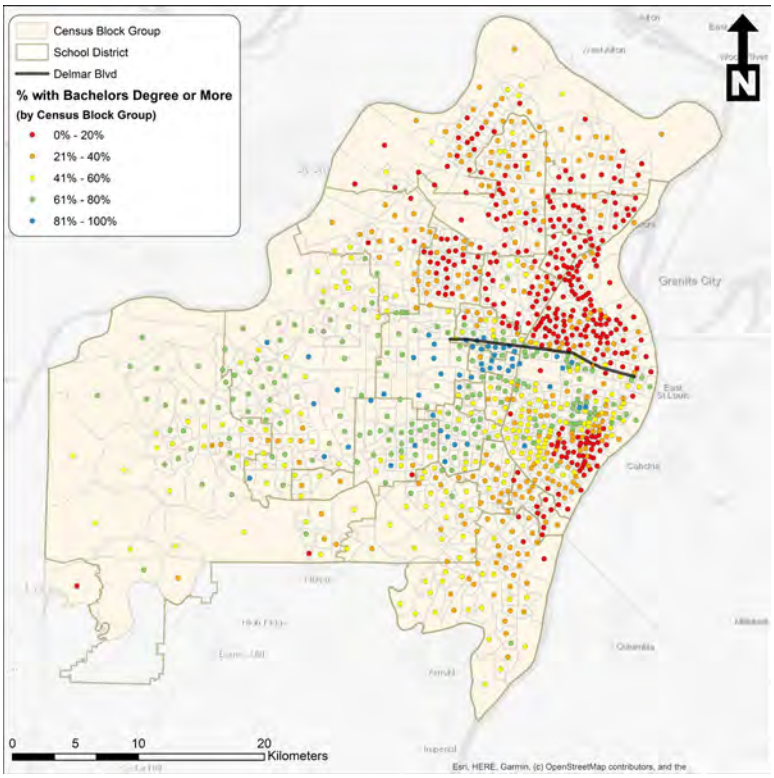
Variable		Min.	Max.	Mean	Standard Dev.
Dependent	% Population with at least Bachelor's degree	0	96.5	38.21	24.84
Independent	Average household income (in US$)	0	558,077	85,948.35	58,323.62
	% Population moved in 2017 or later	0	100	49.16	18.36
	% Households with children	0	100	28.11	13.57
	% White population	0	100	57.70	35.00
	% No internet	0	78.5	10.10	12.17

METHODS

For completeness, I will note that I did use standard methods we discussed in class: data subsetting, table & spatial joins, table pivot, and summary statistics. After preparing, joining (including feature to point tool), and cleaning my data, I followed the following method flow for my analysis. First, to motivate my discussion, I wanted to see how educational outcomes differed in the study area. I present the resulting figure below. From the figure, we see that individuals located above Delmar Blvd in the study area have

[1] Delmar Divide is the term given "to Delmar Boulevard as a socioeconomic and racial dividing line in St. Louis, Missouri." (Wikipedia)

lower educational attainments, as measured by bachelor's degree completion or greater, than the population that resides below Delmar Blvd. I will note that the population located in the southern region of St. Louis City neighboring Illinois does have low educational attainment; I do not have a prior to explain this characteristic.



Second, to motivate my analysis, I conducted an average nearest neighbor (ANN) analysis on the spatial distribution of schools in the study area[2]. I found that the distribution of schools in St. Louis are random for St. Louis City and slightly clustered once I included St. Louis County in the analysis. Since there are rural areas in St. Louis County this result makes sense as the populations are less dense in those area.

Third, to mitigate concerns of multicollinearity in the model, I implemented an exploratory regression on a subset of the variables in my data. It was from this estimation that I found my measures of school quality to not be appropriate predictors of educational attainment outcomes. The exploratory regression did identify one school quality variable as a potential factor (student-teacher ratio), but when implementing the GWR, multicollinearity errors prevented me from incorporating it in the analysis.

Finally, I implemented a GWR. The results of the estimation are presented in the next section.

RESULTS

I present the coefficient estimates from the GWR in the figures presented below on page 4. To reiterate, school quality did not correlate (at least in my specification) with educational attainment. I will briefly summarize the results of the GWR estimation. Before doing so, I will conclude that the model identified demographic and socioeconomic factors that spatially correlated with educational attainment.

[2] A map of the schools in the study area is provided in the Appendix for reference. These did not contribute directly to my final estimation method, so they are not included in the main text.

In Figure A, we see that having children in the households is correlated with lower educational attainment in the urban areas but higher educational attainment in the rural areas. In Figure B, we see that overall household income is positively correlated with educational attainment in the study area. The effects of the income are strongest in the Southern and Northern regions of the study area. In Figure C, we see that increased migration to the block groups in the study area is associated with higher levels of educational attainment. This effect is strongest around Delmar Blvd, particularly South of Delmar in the St. Louis City.

In Figure D, we see that overall having populations with no internet at home is negatively correlated with educational attainment; this effect is strongest in the rural areas of the study area. In Figure E, we see that the white population percentage is positively correlated with educational outcomes; this effect is strongest North of Delmar Blvd, particularly in St. Louis City. This is consistent with the Delmar Divide phenomenon that North of Delmar Blvd has a larger African American population as well as a lower educated population than those living South of Delmar Blvd. Figure F presents the local R^2 values (displayed as quantiles) from the GWR. From the figure, we see that the model explains the data for the study area relatively consistently and has the strongest explanatory power around Delmar Blvd indicating that the model is capturing the phenomenon of the Delmar Divide fairly well. I will note that I found it interesting how similar Figures E and F are as they appear to indicate that race is highly correlated with educational attainment for populations located near Delmar Blvd.

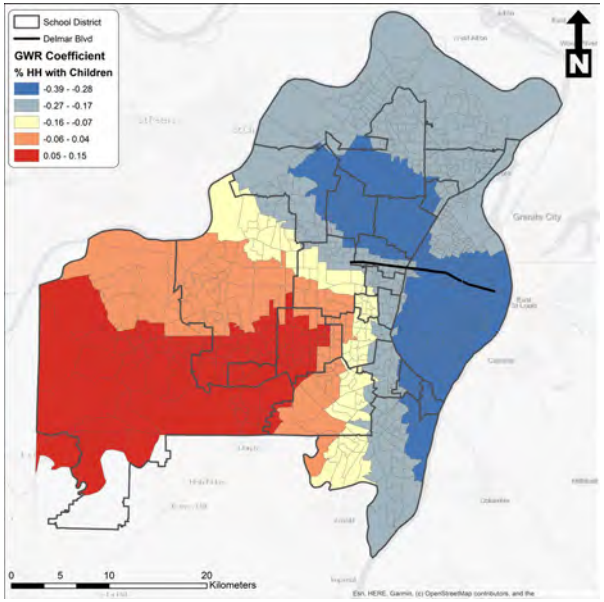
COMMENTS

I will briefly discuss the limitations I faced as well as the limitations of the results presented. First, the variables chosen in this analysis were determined by the data made available to the public. Having more data, particularly intergenerational data is something I strived to incorporate but was unable to do in the time given. Second, with the data used, I did face computational constraints given the size of the data sets. Third, at present, identifying the schools attended by block groups or distances traversed by individuals or households attending schools is challenging and it did limit the empirical approach when I attempted to incorporate school quality.

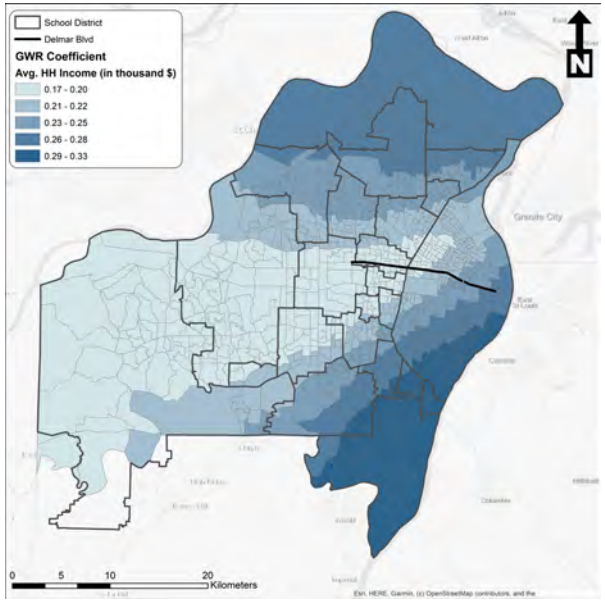
Finally, multicollinearity is everywhere; it was present within the measure of the quality of schools, and it was likely present in the estimation of educational attainment. The level of schooling an individual completes is related to the factors I discussed in my analysis but some of those factors such as income and mobility are likely to be affected by the educational attainment of individuals.

It is clear from the results of this analysis that, from a policy perspective, the influence and damage of the systemic segregation that took place in St. Louis is still present and likely requires intervention to address the effect on the populations of St. Louis.

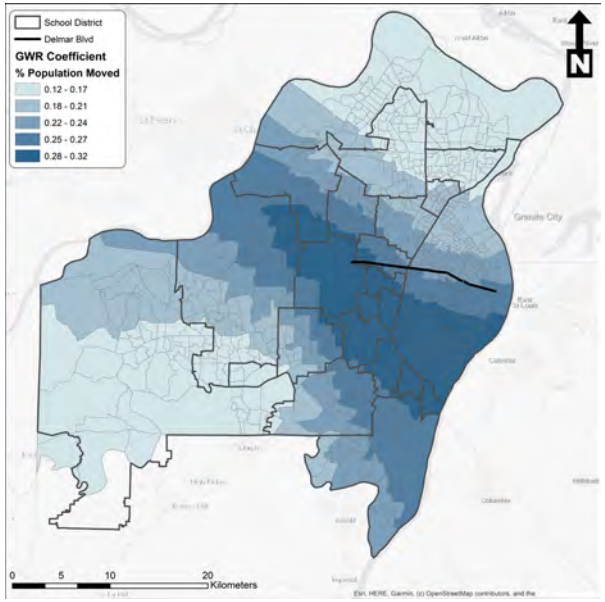
(A)


School District
Delmar Blvd
GWR Coefficient
% HH with Children
-0.39 - -0.28
-0.27 - -0.17
-0.16 - -0.07
-0.06 - 0.04
0.05 - 0.15
N
0 5 10 20
Kilometers

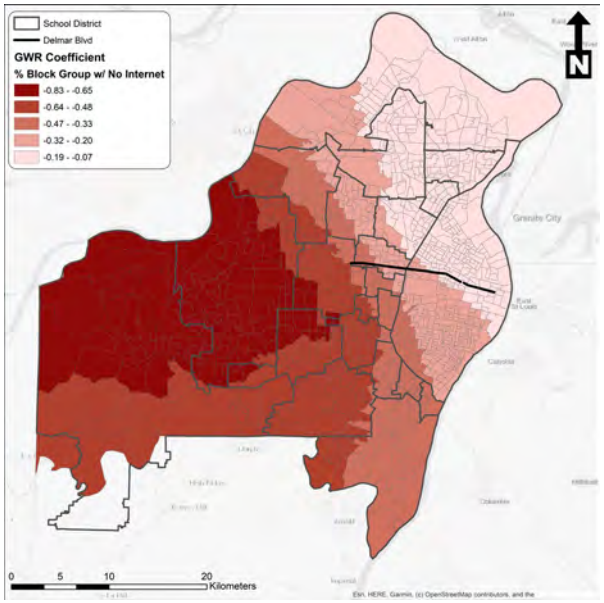
(B)


School District
Delmar Blvd
GWR Coefficient
Avg. HH Income (in thousand $)
0.17 - 0.20
0.21 - 0.22
0.23 - 0.25
0.26 - 0.28
0.29 - 0.33
N
0 5 10 20
Kilometers

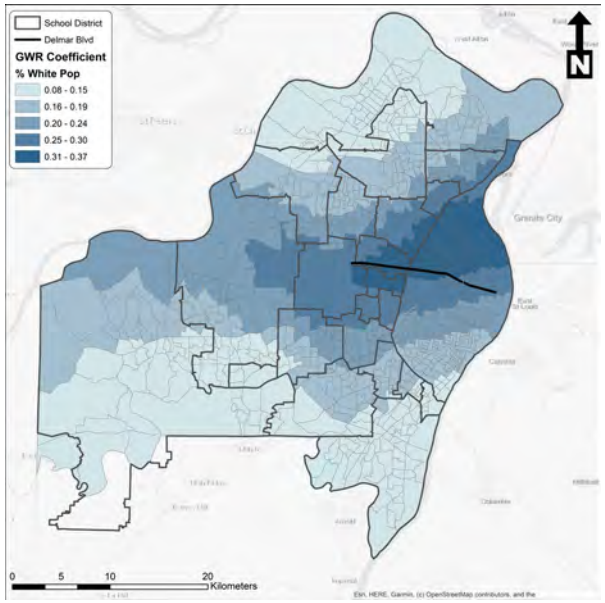
(C)


School District
Delmar Blvd
GWR Coefficient
% Population Moved
0.12 - 0.17
0.18 - 0.21
0.22 - 0.24
0.25 - 0.27
0.28 - 0.32
N
0 5 10 20
Kilometers

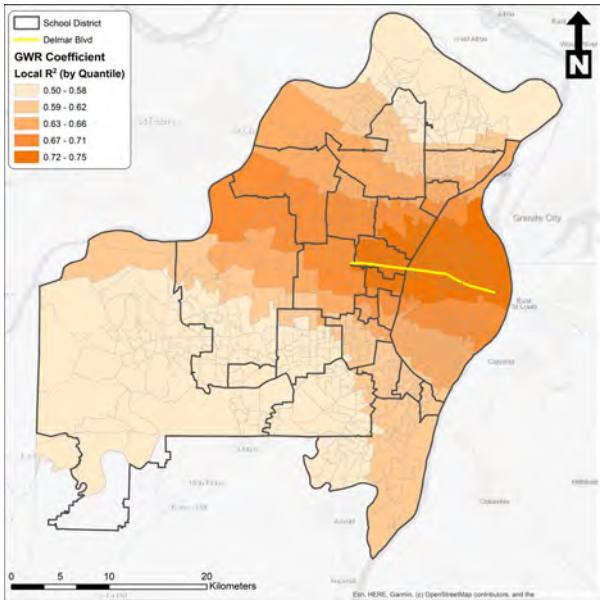
(D)


School District
Delmar Blvd
GWR Coefficient
% Block Group w/ No Internet
-0.83 - -0.65
-0.64 - -0.48
-0.47 - -0.33
-0.32 - -0.20
-0.19 - -0.07
N
0 5 10 20
Kilometers

(E)


School District
Delmar Blvd
GWR Coefficient
% White Pop
0.08 - 0.15
0.16 - 0.19
0.20 - 0.24
0.25 - 0.30
0.31 - 0.37
N
0 5 10 20
Kilometers

(F)


School District
Delmar Blvd
GWR Coefficient
Local R² (by Quantile)
0.50 - 0.58
0.59 - 0.62
0.63 - 0.66
0.67 - 0.71
0.72 - 0.75
N
0 5 10 20
Kilometers

DATA CITED

Below I provide a table of my data sources:

Variables	Description	Data Source (links included)	Date
Block Group Boundary	Administrative boundaries for St. Louis census block groups	TIGER/Line® Shapefiles	2019
Delmar Blvd Street	Line obtained to identify Delmar Blvd on map.	OpenStreetMap	2021
Census Data	Block group census data for 1,052 block groups – variables used from this data are presented in the data section above.	ACS 2019: Period estimates (2015-2019 5-year) for all US block groups.	2015-2019
School District Boundaries	Boundaries for the 23 relevant school districts	Missouri Spatial Data Information Service	2021
Private School Locations	Locations of private schools in St. Louis.		2016
Public School Locations	Locations of public schools in St. Louis.		2020
School Quality Measures	Consisted of MAP test scores (English, math, science), student-teacher ratios, and teacher experience (years teaching & % with master's degrees) by school.	Missouri Comprehensive Data System (MCDS) from Missouri Department of Elementary & Secondary Education	2019

WORK CITED

Ansong, D., Ansong, E. K., Ampomah, A. O., & Adjabeng, B. K. (2015). Factors contributing to spatial inequality in academic achievement in Ghana: Analysis of district-level factors using geographically weighted regession. *Applied Geography*, *62*, 136-146.

APPENDIX

Schools in St. Louis