Exploratory Analysis of Racial Disparities in Academic Performance Throughout Missouri

Abstract

This report utilizes academic and economic data reported by the state of Missouri regarding public schools to map performance trends throughout the state. By utilizing K-means clustering, school data was grouped according to differences in academic achievement, per pupil expenditures, teacher experience, teacher/administrator salaries, and student demographics. Though further statistical analysis is needed to establish correlational conclusions of the data, a visual analysis of mapping the schools after clustering shows that St. Louis houses many schools with racial performance disparities, while rural schools have a more even spread of academic performance within their student populations.

Introduction

The socio-spatial perspective of urban sociology (as cited by Gottdiener, Hohle, King, 2019) is a perspective of understanding urban and suburban development that focuses on how the allocation of resources plays an essential role in how communities are formed and function. This perspective does not only try to understand where resources are going in a community, but how access to those resources effect the residents of a community, which is then used to understand the adverse impacts of segregation and other efforts used to ensure prosperity of certain groups of a local population.

The goal of this report was to explore how the socio-spatial perspective can be used to understand academic disparities in Missouri public schools. In doing so, the aim of this report is not only to show the economic correlations with academic disparity, but to map out public schools throughout the state to demonstrate a spatial pattern throughout the state. The results of this report are then used to demonstrate that economic disparities between schools may have a correlation with both social and spatial factors, hence the use of socio-spatial perspective.

By completing this analysis, the report has a specific audience of both academic administrators and local taxpayers of Missouri public schools. The economic model of Missouri public schools, in which the average school is nearly 75% funded by local property taxes (Missouri Comprehensive Data System, 2021), establishes two essential powers in the funding of schools. The taxpayers vote on bonds that fund school districts, as well as elect the board members of those districts that then allocate the use of those funds. By developing this report, these stakeholders can have a new insight in the breadth of power to which their economic decisions can have on their communities.

Method

Data Sources

All data analyzed in this report was retrieved from the Missouri Department of Elementary and Secondary Education (MO DESE) using their Missouri Comprehensive Data System (MCDS) located at https://apps.dese.mo.gov/MCDS/Home.aspx with the Reports and Resources tab activated. This tool allows users to browse relevant school data collected by the state with generated graphs or raw csv files. It is essential to note that only data from 2019 was analyzed in the report, such that it includes data reported before the COVID-19 pandemic. This decision was made to prevent detecting deviations in data that could have been caused by variations in instruction and school management that were a result of the pandemic. Refer to Table 1 for a comprehensive list of locations to which all data was retrieved.

Economic data in this report was defined as any relevant data that reported any general resources of the school, both in monetary and human capital, as well socioeconomic data of the school. The first dataset of interest in the report includes building level per pupil expenditures of each school. This set provided information on the amount of local, state, and federal expenditures of school building per student in that building. Demographic data was retrieved to account for the percent of students in each school of various races, as well as the percent of the student enrollment that had free and reduced lunch status. The latter of the demographic data was used to get a sense of the socioeconomic status of the student population within the school, since most free and reduced lunch program enrollment is dependent on the economic resources of the student. Insight into the human capital of each school was achieved by analyzing the teacher and administrator data, specifically the average years of experience of teachers in each school along with the percent of teachers with master degrees teaching in each school. This information was considered in the analysis to give a sense of quality of instructional resources available at each school. Further, the average salaries of teachers and administrators in each school was accounted for to understand the sort of monetary resources available in each school, such that a school with high salaries and large per pupil expenditures can be assumed to have a high source of revenue.

Performance data was exclusively derived from results reported on the Missouri Assessment Program (MAP) test. This is an end-of-year assessment administered by the state to grades one through eight. Math and English language Arts assessments are administered to all grades, but Science assessments are only administered in grade five, eight, and eleven. Students are then given one of four scores after the test: “below basic” to denoted deficiencies in the subjects, “basic,” “proficient,” and advanced” to denote exceptional performance in the subject. The results of the 2019 assessment were retrieved, and the report was interested in the percent of student of reported racial categories were analyzed. This was because the report was not interested in performance in general, but specifically interested in performance disparities between white and non-white students.

Location data was retrieved for each school using a directory provided by the MCDS. This directory included the school’s name and its street address. In data preparation, the address data was used to look up the geodata of the school for location analysis and mapping.

Data Analysis

The first step of the analysis was to clean and prepare the performance data. This was done by first removing all blank cells in the original dataset. MO DESE has a policy in which all cells that accounts for 10 or less students to be marked with “*”, which needed to be removed for

analysis in the python environment. After removing all object values and setting the data to type "integer," the data was ready to be transformed for the purpose of the analysis. As stated, the analysis was more interested in the performance disparity between white students and non-white students. This was done by transforming the data to produce performance index for each of the four score levels. This index was produced from the ratio between the percent of white students scoring at one level divided by the average percent of all non-white student categories scoring at that same level. In doing so, a score less than one would suggest that more non-white students performed at that level than white students, whereas a score greater than one would suggest that more white students performed at that level than non-white students. An ideal score of one would suggest that a similar number of white students and non-white students scored at the same level. The resulting dataset included eight to twelve new columns of score performance, such that each school at a performance index for each score level in each subject tested.

To gain a greater sense of the spatial environment to which Missouri schools were situated in, data on community venues surrounding each school was included in the analysis. Before retrieving information on surrounding venues, geodata was collected on each school by using the addresses in the directory retrieved from the MCDS. This was done by calling a Nominatim API from the Python package "GeoCoder," which returns coordinates for a searched address. These coordinates were then plotted on a map from the Folium package, to ensure all the retrieved coordinates lied within the state of Missouri. When the location dataset was cleaned, each school location was sent to a Foursquare API that returned a list of closest 100 venues within a 500-meter radius of each set of coordinates.

The Foursquare API provides a category for each venue as a standardized way to understand what each venue is, rather than relying solely on the name of the venue to understand its function. This category data was utilized in the analysis. To consolidate it from the format produced by Foursquare, each venue was one-hot coded. This was done by producing a data frame that had a column for all possible venue categories with an additional column for the id of each school. Each row then identified the school to which each venue was found near and a "1" in the column of the venue category. A pandas function was then used to group all school identifiers and found the average of each column indexed with that school. These averages for each category were then used in the following cluster analysis.

After all performance and venue data was prepared, a new data frame was created to consolidate all venue, performance, and additional economic data for each school. This data frame was then used for a k-means cluster analysis to categorize each school with similar data. Seven clusters were defined for this analysis. Each school was then assigned a label for the cluster the algorithm placed them in. These labels were then used to produce a map of the state with a point for each cluster, such that each cluster was color coded to assist in visualizing spatial patterns throughout the state.

Results

A K-means cluster analysis was run on the remaining data, and each school was assigned one of seven clusters. Please refer to Table 2 for a summary of cluster characteristics.

Table 1: Summary of Cluster Characteristics, Sorted by "Avg Under Performance Indexes"

Cluster Labels	Avg Under Performance Indexes	Avg Over Performance Indexes	Avg Expenditures Per Student	Avg Total Expenditures from State and Local Tax Revenue	Avg % Of Student White	Avg % Of Student with Free or Reduced Lunch	Avg Years Experience of Teachers	Avg % Of Teachers with Masters Degrees	Avg Teacher Salary	Avg Administrator Salary
5	0.62	1.49	12537.85	9392.15	62.17	24.53	14.83	82.22	67583.69	120753.79
0	0.73	1.35	10883.56	8096.82	69.42	27.49	12.84	74.29	59296.39	102287.78
3	0.78	1.57	9481.50	6421.49	73.20	54.38	11.02	46.74	45994.97	81212.68
1	0.79	1.42	10110.15	7117.88	74.71	40.47	12.96	63.23	51552.62	87799.36
4	0.79	1.56	10623.16	7223.83	59.67	56.54	12.78	64.08	54651.20	103284.19
2	0.86	1.59	9309.61	6168.01	70.94	57.89	11.53	51.39	46130.57	80229.10
6	1.36	1.10	9822.01	6260.95	64.57	63.45	10.30	47.45	45678.50	0.00

Note: "Avg Under Performance Indexes" is the average performance index between all below basic and all basic performance indexes for each subject in each cluster. "Avg Over Performance Indexes" is the average performance index between all proficient and advance performance indexes for each subject in each cluster.

Figure 1 is a map of the entire state of Missouri that has each school marked and labeled with their assigned cluster. Figure 2 shows the same data, but zoomed in towards St. Louis, Missouri. Figure 3 shows the same data as Figure 1, but zoomed in towards Kansas City, Missouri.

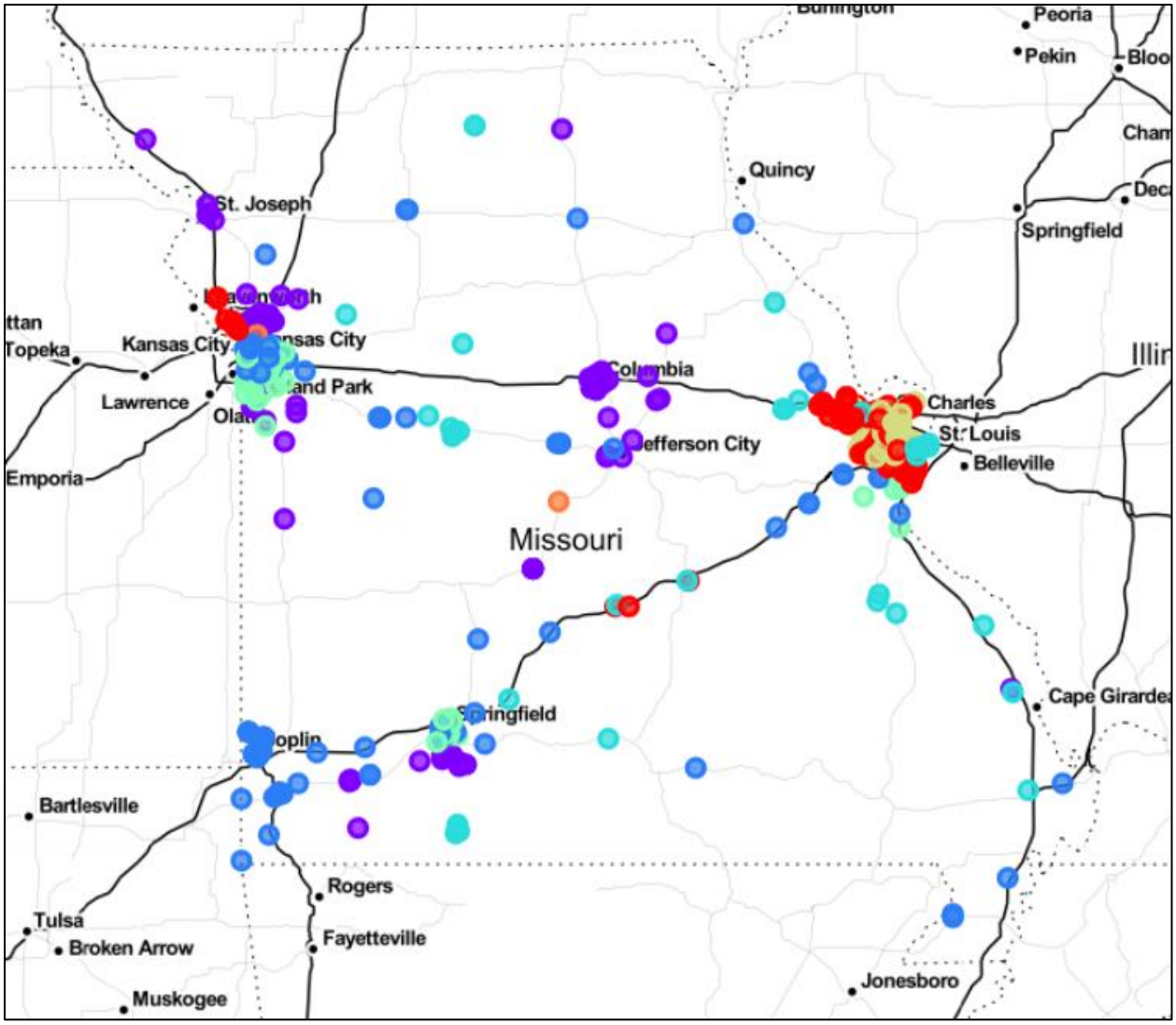

Figure 1. Map of Missouri with points that represent each school. Red is cluster 0, Purple is cluster 1, dark blue is cluster 2, light blue is cluster 3, light green is cluster 4, yellow is cluster 5, and orange is cluster 6.

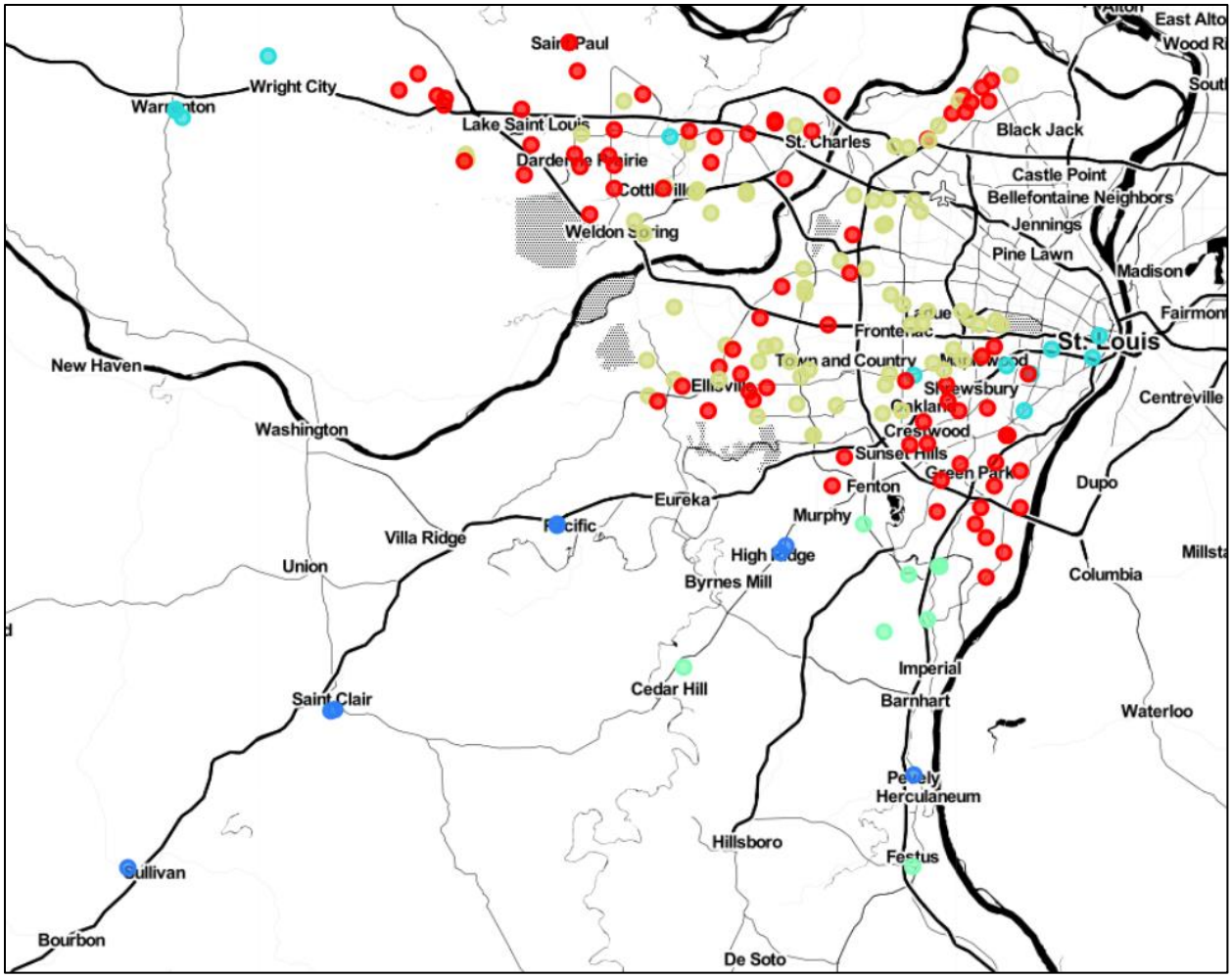

Figure 2. Map of St. Louis with points that represent each school. Red is cluster 0, Purple is cluster 1, dark blue is cluster 2, light blue is cluster 3, light green is cluster 4, yellow is cluster 5, and orange is cluster 6.

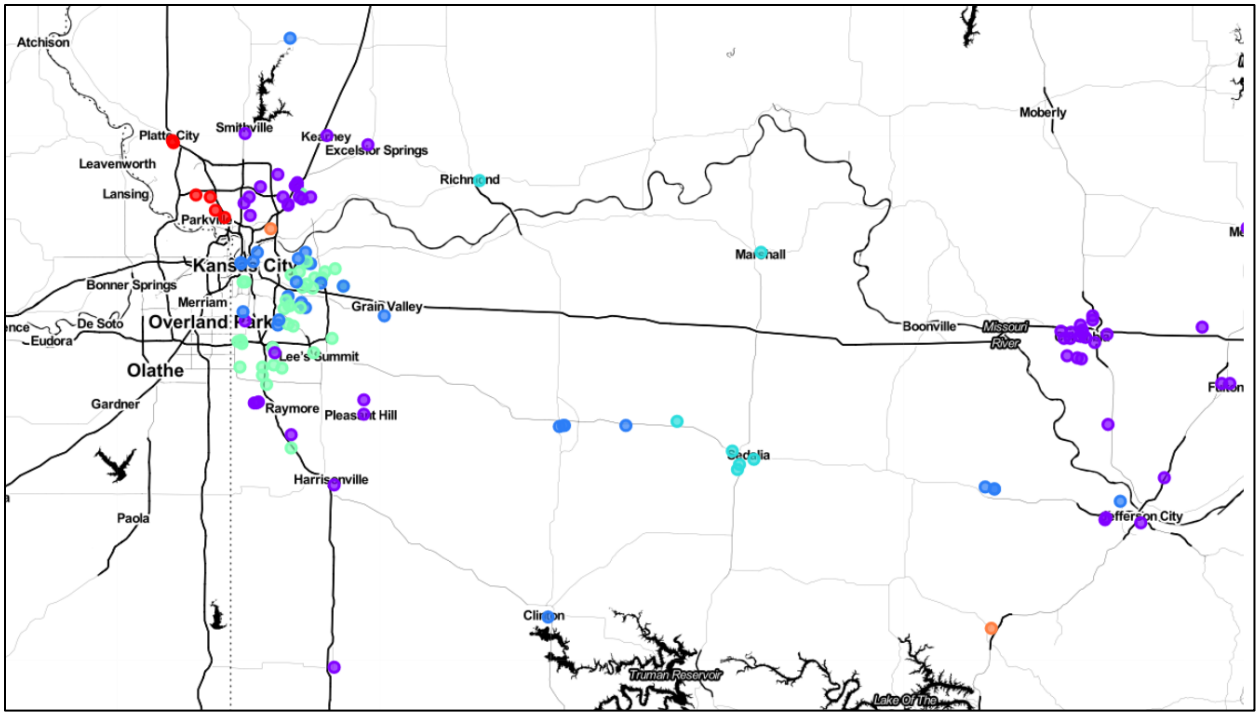

Figure 3. Map of Kansas City with points that represent each school. Red is cluster 0, Purple is cluster 1, dark blue is cluster 2, light blue is cluster 3, light green is cluster 4, yellow is cluster 5, and orange is cluster 6.

Discussion

The first step in understanding the analysis for this repot was to get a sense of what the clusters mean. In this report, the main goal was to understand which clusters represent schools with a high racial disparity, particularly those with a large gap in under performance. This means that clusters with small under performance indexes are of particular interest. Once all the clusters a were sorted on the under-performance index, clusters with outstanding gaps can be easily located.

In this analysis, clusters 5 and 3 have some of the most outstanding under performance gaps. Meaning that these clusters accounted for schools in which more non-white students were performing at below basic and basic score levels than white students in that school. Conversely, cluster 2 had an underperformance index closer to 1, indicating that these schools have a more even share of under performance scores between white students and non-white students, though it still seems far from an ideal 1:1 index.

Once it was established that cluster 5 represent schools with a greater racial disparity than cluster 2, the next step of the analysis was to understand what other factors differ between the three. The first observation is that cluster 5 accounted for schools with less white students, as compared to schools accounted for in cluster 2. The average percent of the student population being white was greater for cluster 2 than for cluster 5. When comparing the economic resources of the two clusters, it comes clear that schools accounted for in cluster 5 enjoy greater expenditures per pupil, hire more qualified teachers, and pays higher salaries for both teachers

and administrators than cluster 2. Additionally, more students in cluster 2 schools are classified with free and reduced lunch than compared to cluster 5 schools.

The next part of this analysis was to understand where each cluster of schools tended to lie within the state of Missouri. After reviewing the maps generated in the analysis, it was observed that most cluster 0 and cluster 5 schools were located in the St. Louis region. Many cluster 2 schools were much for scattered in more rural regions of the state, as well as near smaller cities like Joplin and Kansas City. Cluster 1 schools were relatively well disbursed throughout the state, except in the St. Louis region.

Conclusion

Based on observation, it would seem that the data suggests two main themes of Missouri public schools. The first observation is that there seems to be a correlation between racial disparity in under performance and economic resources, such that a negative correlation exists between these factors. The second observation is that schools with the greatest racial disparity and largest economic resources are centered in the St. Louis region. Additionally, schools with less resources and a greater white student population exists in the rural regions of the state, such as the Joplin area.

These observations could be explained by the rapid development present in the St. Louis area (Gottdiener, Hohle, King, 2019), such that suburban development has diverted much of the available economic resources to regions that are ignoring the needs of non-white students. What is still curious about this conclusion is that suburbanization tends to tract money and greater white populations, but the schools in cluster 5 had both greater economic resources and a greater non-white student population.

The limitations of this conclusion are centered mainly on it being based on observation, such that no statistical analysis was utilized. Further analysis should use a multivariate analysis of variance to ensure that clusters are statistically different from one another. A regression analysis should also be used to test the statistical correlation between economic and performance variables. Additional data cleaning could also be helpful to ensure that more schools are retained for a final analysis.

References

Gottdiener, M., Hohle, R, & King, C., 2019. *The new urban sociology*. Routledge: New York. https://doi.org/10/4324/9780429244452