An Analysis of Socioeconomic Trends in Academic Achievement in the U.S.

Abstract:

In this project, I used two datasets to figure out if there are any interesting correlations between economic variables and educational variables at the county level nationwide. How do students' broader social environments impact their academic achievement? I attempt to answer this question by getting the county-level covariates dataset from the Opportunity Insights and nation-wide test score data for 3rd-8th graders from the Stanford Education Data Archive. The Opportunity Insights dataset describes socioeconomic of each county. The Stanford dataset describes academic variables for 3rd-8th graders across the nation by county. It contains variables for academic achievement and student demographics. In addition to running models on test scores using racial and economic demographics, I also descriptively analyzed each dataset to find trends within the datasets themselves, such as trends in counties with the largest gender gaps in academic achievement.

Background and Significance:

The American education system is a highly fragmented system in which school governance is conducted at the local level. State and local school districts have the power and authority to determine the curriculum and standardized testing standards that schools need to adhere to. While supporters of this system advocate for the potential to customize curriculum according to the community's educational values and students' needs, this system has a myriad of negative consequences on facilitating school improvement. First, the lack of coherence among different school districts acts as a barrier for communication about effective teaching methods and curricula.[i] Second, the fragmented system can lead to an unequal distribution of funding resources between school districts in the same state.[ii] Finally, and most important to the topic at hand, the decentralized system leads to many different methods of storing educational data, often in siloed data storage systems.[iii] The dataset from the Stanford Education Data Archive is innovative in education research because it standardizes over 300 million test scores for 3rd-8th graders in every public school in the U.S. from 2008-2016.[iv] Combined with the dataset from Opportunity Insights on neighborhood-level socioeconomic covariates,[v] important insights on the impact of environmental factors, such as average household income or racial composition, on test scores can be discovered.

The topic of diversity in education is especially relevant today with many discussions of the short- and long-term benefits of diversity, academically and beyond. It may have a major impact on students' academic achievement, which would have large implications for the importance of affirmative action in schools' admissions policies.[vi] Besides having an impact on students' academic performance, it also affects their perceptions of others from different backgrounds.[vii] I will use these two datasets together to gain insights on how environmental economic factors and racial diversity may influence students' academic performance. I also want to descriptively analyze the areas with the largest inequalities in academic achievement between genders since that is also another important dimension of diversity.

Methods:

a. Data Collection

In this project, I will be merging together two nation-wide datasets from the Stanford Education Data Archive and the Opportunity Insights lab by their county FIPS code. The Stanford Education Data Archive dataset specifically looks at test score data for 3rd-8th graders from 2008-2016 in public schools across the U.S. For each county, there are standardized mean test scores and their respective standard deviations. Test score results are further filtered by racial, economic, or gender indicators. Measures of test score gaps between various racial groups and between genders are calculated.[viii] The Opportunity Insights dataset contains many economic and social covariates at the neighborhood and county levels.

b. Variable Creation

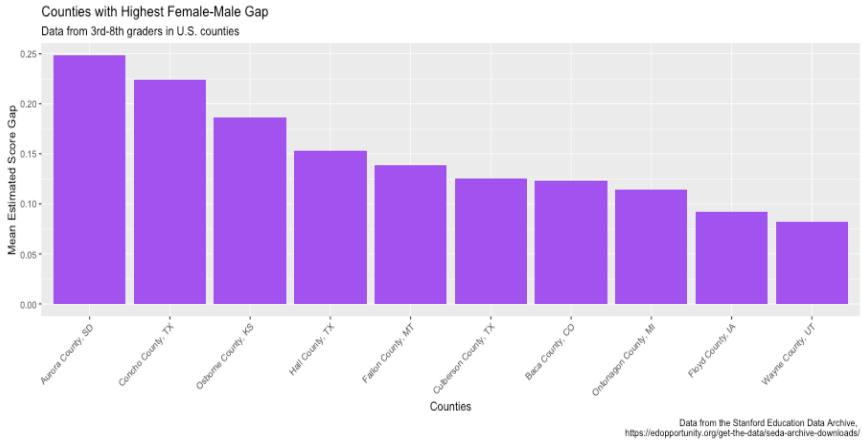
In terms of the variables I used in my analysis, I mostly used existing variables. I mutated the household income variable to units of 10,000 dollars instead of 1 dollar because otherwise, the regression output would not show a significant effect of income on test scores. All existing variables in the dataset are explained in the Stanford Education Data Archive codebook[ix] and the Opportunity Insights website.[x]

c. Analytic Methods

For my modeling analyses, I will be running linear regressions to model test scores using a myriad of socioeconomic covariates found in the Opportunity Insights dataset. I then present the results in a summary regression table and a scatterplot with a fitted regression line showing the correlation between the covariate and test scores. For the descriptive analyses, I will be using bar graphs to show the states with the highest scores for each gender, as well as the counties with the highest female-male test score gaps.

Results:

The first question I tackled was the descriptive question about gender inequalities in test scores. I started with descriptive analyses of the gender gap across states and counties. I first found the top 10 states with the best mean scores for females and males respectively (fig. 1, 2). Some differences between the two plots are that Maryland and Pennsylvania appear in the top 10 states for females, and Kansas and Indiana appear in the top 10 states for males. I then found the top 10 counties with the largest gender gaps (depicted).



The second question that I tackled is the question of how socioeconomic covariates of the county in which students are situated correlate with students' academic performance. First, I created a model of all test scores using a linear regression of average household income (in units of 10,000 dollars) and the proportion of single-parent households. There is a coefficient of 0.06 for this variable, meaning that on average, for each 10,000 dollar increase in household income, there is a .06 increase in scores. For the share of single- parent households, there is a coefficient for -1.88, meaning that for every .1 increase in the proportion of single-parent households, there is a -.188 decrease in the mean of mean scores in the county on average. Both variables have a very small standard error as well as a high absolute T-statistic value, which says that these variables model the scores very well. The p-values are also smaller than .05, which means that at the 95% significance level, the effects of both variables on mean scores are statistically significant and not due to random chance.

Model of Test Scores by Mean Household Income and Share of Single-Parent Households

Data from 3rd-8th graders in 2010

term	Coefficient	Standard Error	Statistic	P-value	5th Percentile	95th Percentile
(Intercept)	0.096	0.002	41.654	0.000	0.091	0.100
hhinc_k	0.055	0.000	234.263	0.000	0.055	0.056
singleparent_share2000	−1.881	0.005	−374.076	0.000	−1.891	−1.871

Data from Opportunity Insights and Stanford Education Data Archive.

Then, I ran another linear regression on test scores using the proportions of non-white residents and the share of poor residents. The coefficient for nonwhite_share2010 says that on average, for every .1 increase in the share of non-white residents in a county, there is a .0431 decrease in the mean test scores for students in the county. I decided to interpret this at .1 increments because the nonwhite_share2010 variable is a percentage share of the total population, not a binary variable. The p-value is 0, which means that at the 95% significance level, it is statistically significant. The coefficient on poor_share2010 says that for every .1 increase in the share

Share of Non-White Minorities and Poor Residents in County vs. All Students' Test Scores

Test Score Data from 8th graders in 2010

term	Coefficient	Standard Error	Statistic	P-value	5th Percentile	95th Percentile
(Intercept)	0.291	0.007	40.465	0.000	0.277	0.305
nonwhite_share2010	−0.431	0.015	−28.209	0.000	−0.461	−0.401
poor_share2010	−1.715	0.048	−35.795	0.000	−1.809	−1.621

Data from Opportunity Insights and Stanford Education Data Archive.

of poor people in the county, there is a decrease of .1715 in test scores on average.

I also ran models of how the proportions of each racial group in the county might affect the test scores of students within that racial group. For Asians, there was a large positive effect on test scores, where on average, a .1 increase in the share of Asians corresponds to a .1 increase in the mean scores of Asian students. For Black and Hispanics, there is a negative correlation for each group respectively, with a stronger effect for Black students. The coefficient for share_black2010 says that on average, for every .1 increase in the share of Black residents in a county, there is a .05 decrease in the mean test scores for Black students in the county. The coefficient for share_hisp2010 says that on average, for every .1 increase in the share of Hispanic residents in a county, there is a .015 decrease in the mean test scores for Hispanic students in the county. Here, I show a sample regression table and scatter plot for the Asian group. The scatter plot's x-axis is scaled by log10 because of the multitude of data points near zero, since Asians typically have small shares of the county's population. The tables and plots for the Black and Hispanic groups can be found in figures 3 and 4 in the Appendix, respectively.

Relationship Between Share of Asians in County vs. Asian Students' Test Scores

Data from 8th graders in 2010

term	Coefficient	Standard Error	Statistic	P-value	5th Percentile	95th Percentile
(Intercept)	0.390	0.015	25.876	0.000	0.361	0.420
share_asian2010	1.021	0.299	3.417	0.001	0.435	1.606

Data from Opportunity Insights and Stanford Education Data Archive.

Discussion/Conclusions

The results substantially answered my research questions about how different socioeconomic covariates correlate with academic performance. Going into the research, I did not have strong hypotheses about what the results may be because I wanted to do more inductive research, making conclusions based on what I can find on the data. In terms of the descriptive analyses on gender, the interpretations were limited by the lack of theoretical understanding about what might be the root causes of gender inequalities in education among different states, but the results do provide guidance for places that education researchers can start from, particularly if they wish to investigate places with the highest gender gaps.

The most surprising finding is the relatively large impact of the share of single-parent households on academic performance, dwarfing the impact of traditional measure of economic status such as household income. This result is surprising because this measure of single-parent households is drawn from the entire county, rather than at the individual-level, so the fact that the status of other households may impact children in the county in general is very surprising. Future research can look at the specific mechanisms of how the proportion of single-parent households may affect students in that community, even those who do not live in single-parent households themselves.

The analyses on racial groups places nuance on the previous literature on the benefits of diversity on students' performance, as different groups seem to experience different effects in both direction and scale from their community's racial composition. Further research can investigate why Black and Hispanic students may see a negative correlation in their test scores with the proportion of Black and Hispanic residents in their county, whereas there is the opposite trend for Asian students. For a more interactive look at the data, all of the results are available online in a Shiny application.[xi]

Appendix:

Figure 1.

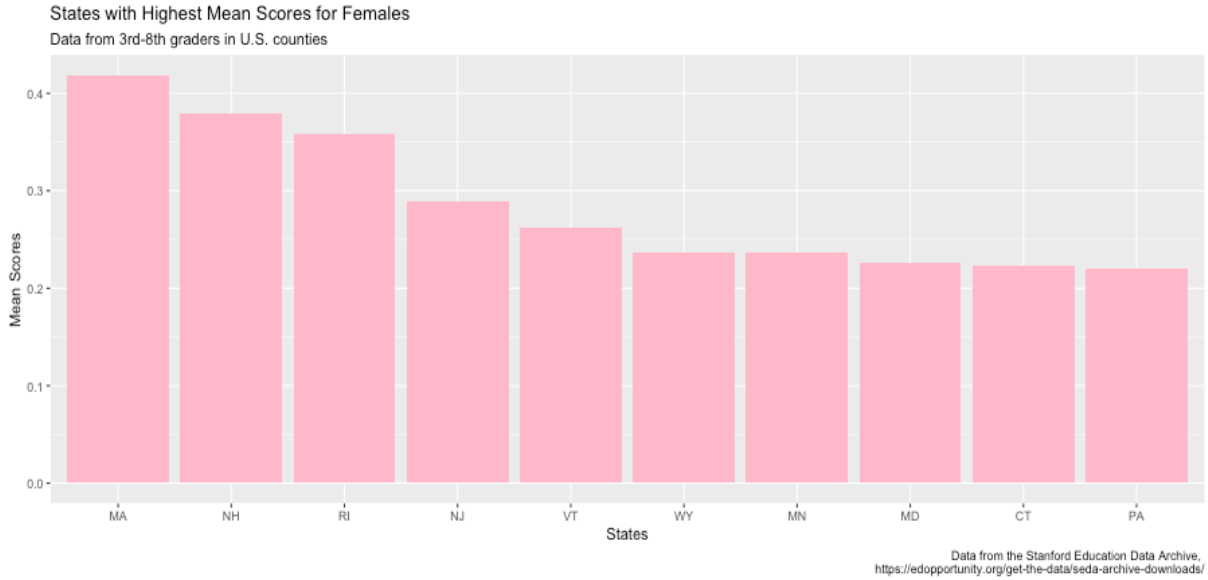


Figure 2.

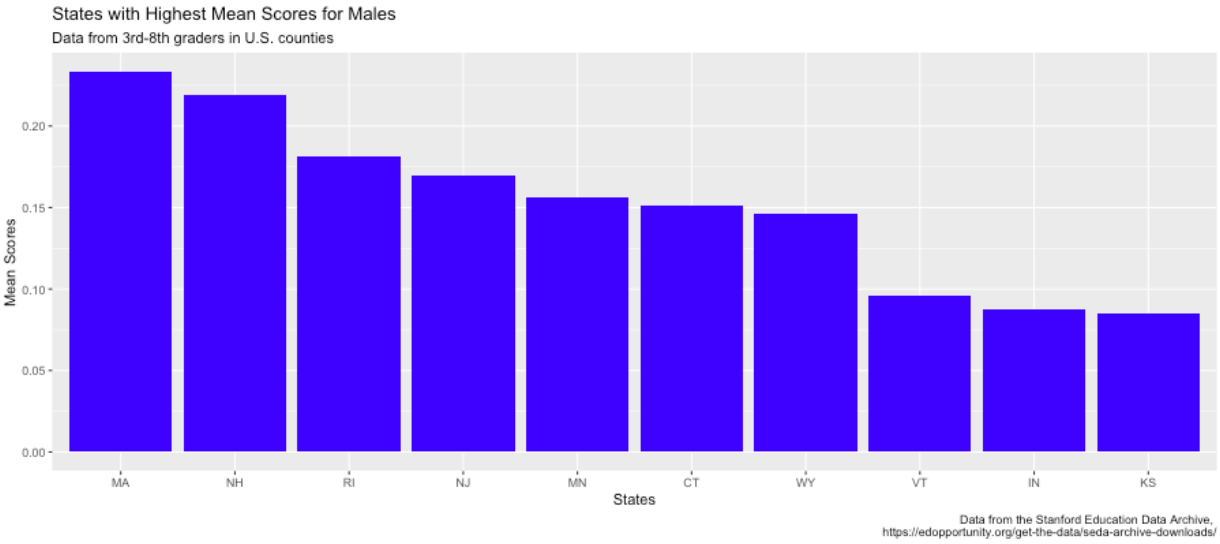


Figure 3a.

Relationship Between Share of Blacks in County vs. Black Students' Test Scores

Test score data from 8th graders in 2010

term	Coefficient	Standard Error	Statistic	P-value	5th Percentile	95th Percentile
(Intercept)	−0.407	0.006	−65.754	0.000	−0.419	−0.395
share_black2010	−0.500	0.024	−20.490	0.000	−0.548	−0.452

Data from Opportunity Insights and Stanford Education Data Archive.

Figure 3b.

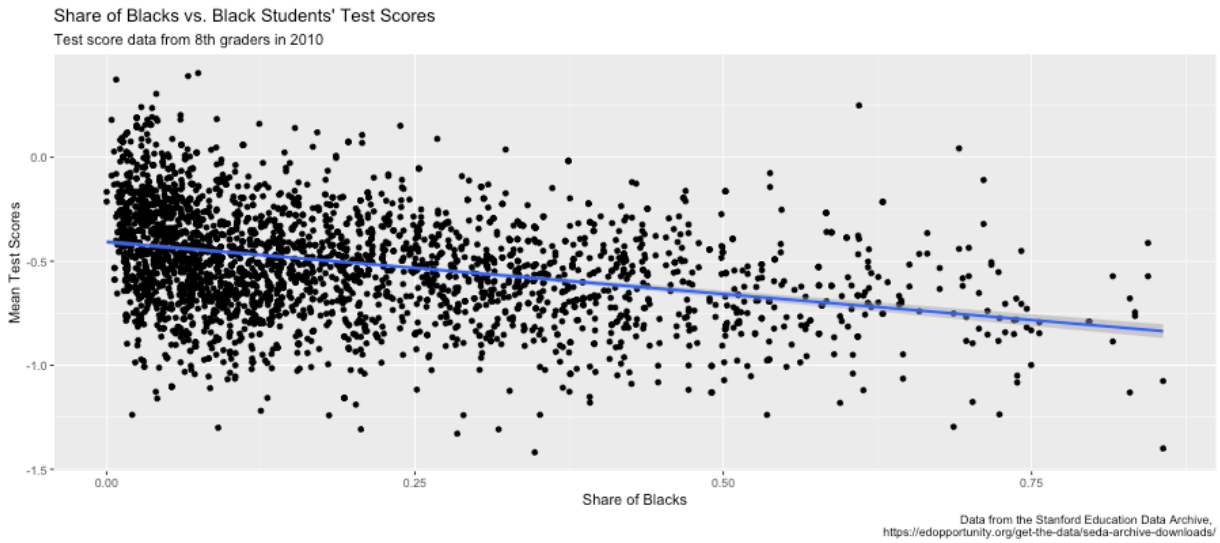


Figure 4a.

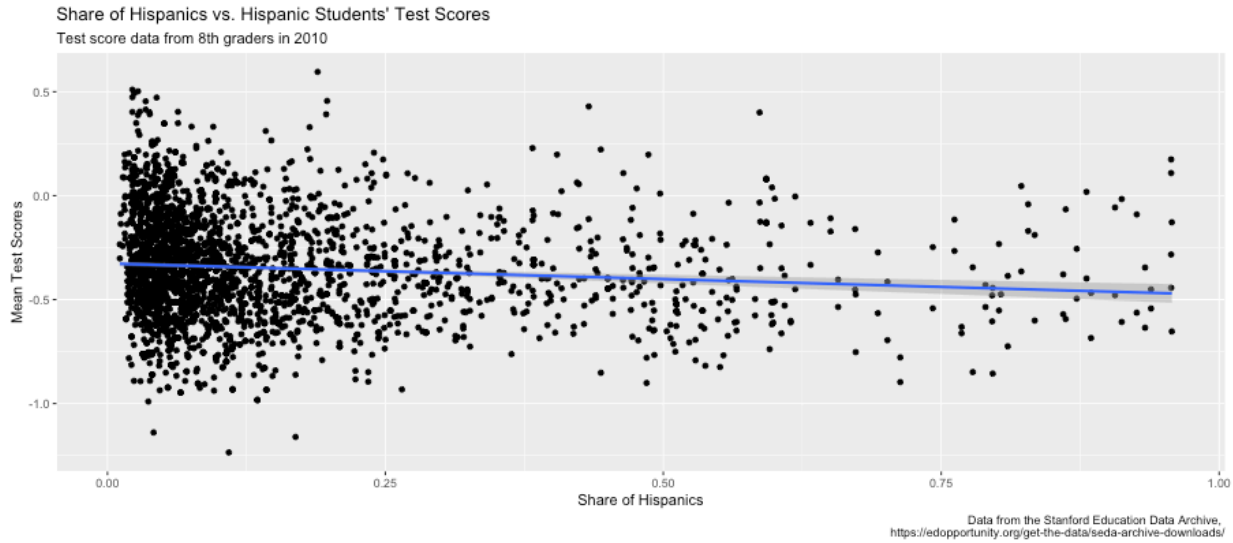
Share of Hispanics in County vs. Hispanic Students' Test Scores

Data from 8th graders in 2010

term	Coefficient	Standard Error	Statistic	P-value	5th Percentile	95th Percentile
(Intercept)	−0.326	0.006	−54.561	0.000	−0.338	−0.314
share_hisp2010	−0.151	0.028	−5.452	0.000	−0.205	−0.096

Data from Opportunity Insights and Stanford Education Data Archive.

Figure 4b.



[i] Stringer, K. (n.d.). Schools Have Plenty of Good Ideas, but They Don't Always Know How to Implement Them. New Report Aims to Improve 'Fragmented' Education Systems. Retrieved December 13, 2019, from https://www.the74million.org/article/schools-have-plenty-of-good-ideas-but-they-dont-always-know-how-to-implement-them-new-report-aims-to-improve-fragmented-education-systems/

[ii] Gunn, D. (n.d.). America's Fragmented Education Funding System. Retrieved December 13, 2019, from Pacific Standard website: https://psmag.com/education/americas-fragmented-education-funding-system

[iii] Ash, K. (2013, March 14). Fragmented Data Systems a Barrier to Better Schools, Experts Say - Education Week. *Education Week*. Retrieved from https://www.edweek.org/ew/articles/2013/03/14/25datadelivery.h32.html

[iv] Badger, E., & Quealy, K. (2017, December 5). How Effective Is Your School District? A New Measure Shows Where Students Learn the Most. *The New York Times*. Retrieved from https://www.nytimes.com/interactive/2017/12/05/upshot/a-better-way-to-compare-public-schools.html, https://www.nytimes.com/interactive/2017/12/05/upshot/a-better-way-to-compare-public-schools.html

[v] Opportunity Insights. (n.d.). Retrieved December 14, 2019, from https://opportunityinsights.org/

[vi] Hallinan, M. T. (n.d.). Diversity Effects on Student Outcomes: Social Science Evidence. *OHIO STATE LAW JOURNAL*, *59*, 22.

[vii] Merlino, Luca and Steinhardt, Max Friedrich and Wren-Lewis, Liam. Forthcoming. "More than Just Friends? School Peers and Adult Interracial Relationships." *Journal of Labor Economics.*

[viii] Matre, J. C. V., Shores, K., Kalogrides, D., & Reardon, S. F. (n.d.). *STANFORD EDUCATION DATA ARCHIVE*. 9.

[ix] https://cepa.stanford.edu/sites/default/files/SEDA%20Technical%20Documentation%20Version1_1.pdf

[x] https://opportunityinsights.org/data/

[xi] https://amytan9.shinyapps.io/test_score_explorer/