United States Education Study

An Inquiry into Spending and Enrollment Demographics

Katie Tooher
Marquette Computer Science
Marquette University
Milwaukee Wisconsin USA
katie.tooher@marquette.edu

ABSTRACT

This project serves as an analysis of United States K-12 through 12th grade education statistics and data from 1992 – 2016 organized by Roy Garrard. The data is sourced from the U.S. Census Bureau and the National Center for Education Statistics (NCES). The contents of this dataset are not meant for the identification of individual institutions but for a broad view of United States primary to secondary education (K12 - High School). The purposes of this inquiry are to model and test associations that state and federal spending on US education have on enrollment demographics as well as enrichment value of education. For this project, educational enrichment value is quantified by looking at the test score data provided. As a disclaimer, test scores do not always provide the only insight into the value of education but is a single way to quantify this.

BACKGROUND

Education around the world is not always a right or required by law. This discrepancy means that across the globe there is an epidemic of uneducated or undereducated people who struggle to survive and have lower qualities of life. In the United States compulsory education laws are not applicable to all states, but most require that students start their education by age 6 and attend till they are at least 16. There are exemptions to these laws such as home-schooling or religion. A research study by the Economic and Social Research Council, *Identity, Socioeconomic Status and Wellbeing,* reported that "level of education is the strongest predictor of outcomes (compared to age, gender, income, employment status, and marital status) in all models, except for the outcomes of wellbeing and health." This implies that spending on education should be one of the highest priorities in federal and state budgets to promote the improvement and wellbeing of their people.

For educational spending, the United States federal government spends roughly 55% towards total educational expenditure, while state and local governments make up the other 45% according to Jonathan Gruber in his text "*Public Finance and Public Policy"* (2016). Based on data found in the United States Department of Education's, "Education Department Budget History Table: 1980 – 2018," it is apparent (and not surprising) that United States funding for elementary and secondary education has seen an overall increase since the 1980s. Most of the increase is largely due to fluctuations in cost of living and the overall economic health of the United States. The dataset does show decreases for elementary and secondary education between a few different periods, however the reasons for these decreases are not detailed by the dataset and could be the cause of many different factors. In a 2018 breakdown of the entire United States budget, federal spending on education was roughly 15%, as reported by the United States Office of Management and Budget. Thus, it is clear that the United States is making education a

K.Tooher '19, Dec. 2019, Milwaukee, Wisconsin USA

priority for the country and the USA can expect to see educational spending increases with time.

In order to discuss the demographic implications that this study has for increases or decreases in levels of enrollment, it is important to consider the current state of United States population demographics. According to an article written by Bill Chappell for NPR, the U.S. birthrate in 2018 was at its lowest since 1986 and was the fourth consecutive year of birth declines according to the Center for Disease Control and Prevention. "Births: Provisional Data for 2018," a study conducted by the Division of Vital Statistics also reports that "births for United States in 2018 was 3,788,235, down 2% from 2017 and the lowest number of births in 32 years." Although these numbers show decreases in the amount of births in the years since 1986, this means that there is a potential for increase for quality of life and wellbeing as more people are waiting until they are financially stable to start families and resources for family planning and health are showing improvements. For the purposes of this study this means that distributions of students and student enrollment over time could decrease, but this is not a bad thing.

The National Center for Education Statistics reports that the percentage of U.S. public high school students who graduate on time has "increased over the first 7 years … from 79 percent in 2010 – 2011 to 85 percent in 2016 – 2017" as measured by the adjusted cohort graduation rate (ACGR). In 2016 to 2017 for American Indian/Alaska Native 72 percent, Black 78 percent, and Hispanic 80 percent of public high school students were blow the U.S. average of 85 percent of graduating students. White students (89 percent), and Asian/Pacific Islander (91 percent) high school graduates were above the U.S. average. On some state levels such as Wisconsin, the rates for non-Hispanic White students, American Indian students, Asian/Pacific Islander students and Hispanic students showed at least some increase in percentage of high school graduates; whereas graduation rates for Black students decreased, as reported by the Applied Population Laboratory (APL) in a study on the State of Wisconsin high school education statistics projected graduations.

METHODS

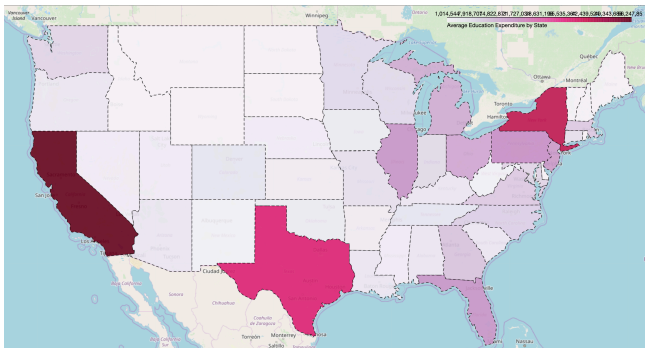
The dataset "U.S. Education Datasets: Unification Project" by Roy Garrard derived from Kaggle that was used for this project was the "states_all_extended.csv". Before using this data to model different associations between spending and enrollment and spending with math test scores it needed to be cleaned.

To start, there were sections of the "States" column that were unique to the rest of the dataset, meaning there was only one instance for that state or territory. Most of the singular pieces of data under the state column that matched this description were either territories that had been renamed, such as Northern Marianas vs Northern Marianas Islands vs Commonwealth of Marinas, Virgin Islands vs U.S. Virgin Islands vs VI, PR vs Puerto Rico. Others had the state acronym listed instead of the full name. The U.S. territory of Guam was removed because of the lack of data in its columns/rows. Other groups listed in the "State" column that were removed were different naming variations of the Department of Defense that did not make sense to belong in this dataset such as DOD_-_FOREIGN, DOD_-_DOMESTIC, DOD_-_OVERSEAS, DOD_(OVERSEAS_AND_DOMESTIC_COMBINED), and DOD_DOMESTIC. The Pandas Python library was used to clean the data. Rows that contained NaN values were also removed. The final cleaned dataset contained 1229 rows and 193 columns.

Before modeling the data, I chose to create generalized data visualizations to get a

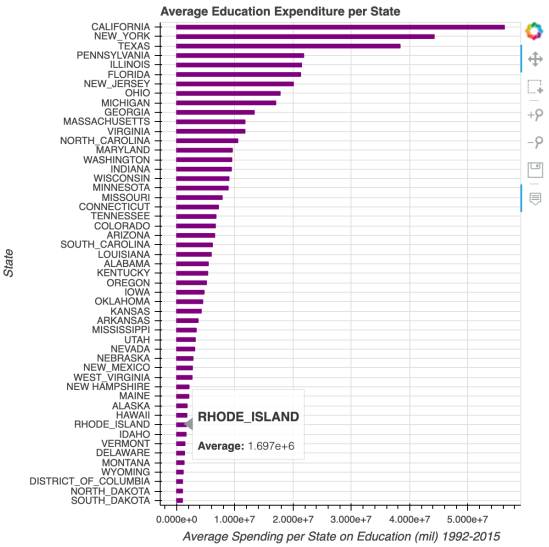
perspective on which key performance indicators would be best to look at, and to find highest, mid-range, and lowest spending states. Finding the different levels of spending for all the states over the period between 1992 and 2016 would help in generalization of findings for this study.

The first visualization created looked at the average expenditure per state between 1992 and 2016. A Choropleth map visualization was chosen because it offers an easy way to view the different levels of expenditure through the gradient shading (dark pink is higher and lighter pink is lower).



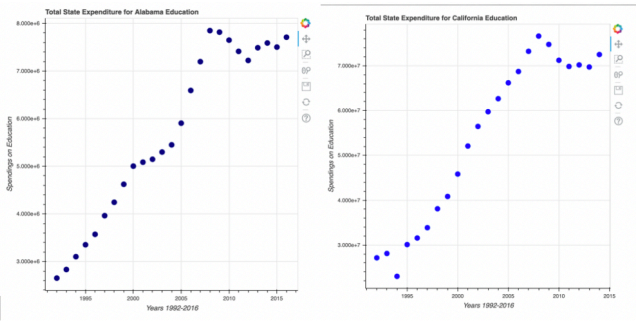
This visualization suggests that the states that spend the most are California, New York, and Texas. This visualization was created using the Python folium library which uses html leaflet to create this gradient map.

Now to see the numbers in a more quantitative manner I created a simple horizontal bar chart using Python Bokeh.



This visualization shows each state's spending and displays the exact amount (in millions) when hovering over a specific state.

To see the different possible KPI's for specific enrollment per demographic as it compares to spending, I created a series of simple scatter plot graphs to see which type of modeling would best fit this study.



The two scatter plots show average spending for Alabama and California. As depicted in the graphs the data exhibits mostly linear data until later years in which spending begins to fluctuate (discussed in the background).

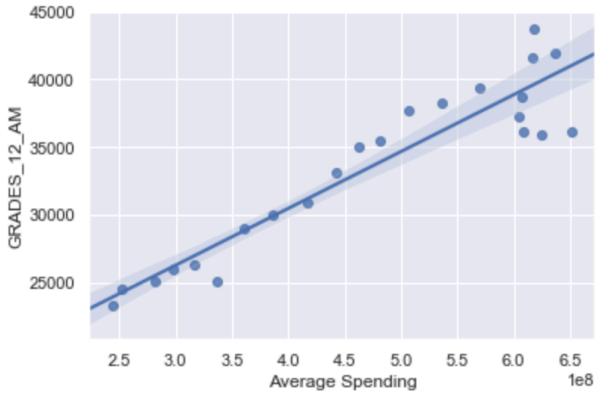
ANALYSIS

Two questions this study set out to answer are:

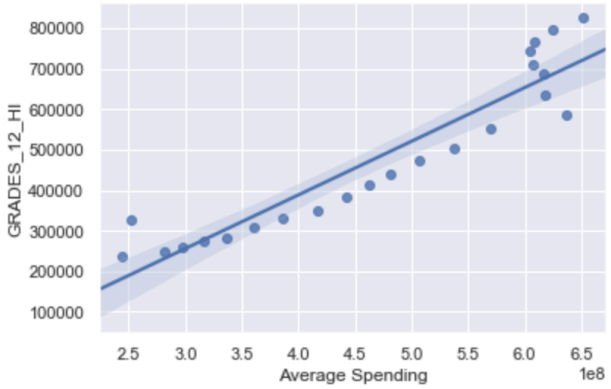
1. Does total spending on education on the federal, state, and local levels of government effected enrollment of different ethnicities of students over time?
2. Does total spending on education on the federal, state, and local levels of government effected test scores (or quantified enrichment value)?

1.1 Linear Regression for Enrollment Demographics and Spending

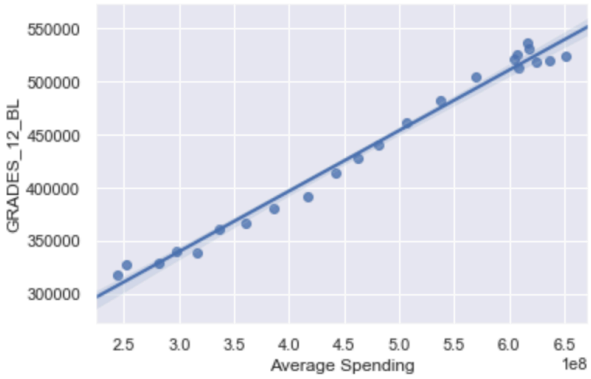
Using Python, the libraries used to do the linear regression were pandas, seaborn (to visualize confidence intervals) sklearn, and statsmodels.api. Linear regression modeling was chosen because the scatter plots generate with the spending data previously showed linear trends. To make the model more specific This regression focuses on grade 12 students of three different races. This decision was made because students that had already made it to grade 12 would have a higher chance of graduating than other grades.



The above linear regression model looks at average spending (millions) over 1992 – 2016 for American Indian students in grade 12.



The above linear regression model looks at average spending (millions) over 1992 – 2016 for Hispanic students in grade 12.



The above linear regression model looks at average spending (millions) over 1992 – 2016 for Black students in grade 12.

Theses visualizations show strong data linearity as well as small confidence intervals, especially for Black students.

Group	P – Value	R^2	F-Statistic
Native American	0.000	0.876	148.5
African American	0.000	0.986	147.6
Hispanic American	0.000	0.876	148.8

Results when each group was modeled separately.

Group	P – Value	R^2	F-Statistic
Native American	0.881	0.986	452.1
African American	0.005	0.986	452.1
Hispanic American	0.698	0.986	452.1

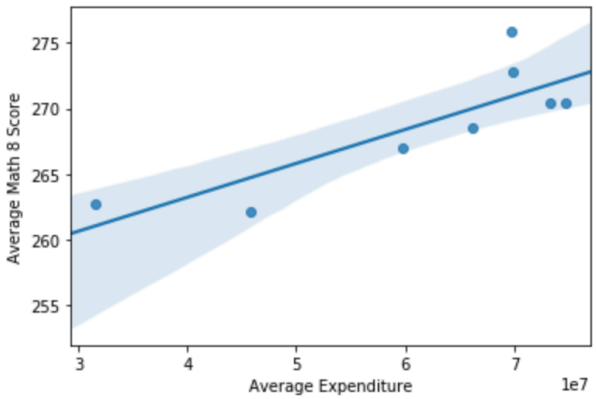
Results when all groups were modeled together.

1.1.1 Results

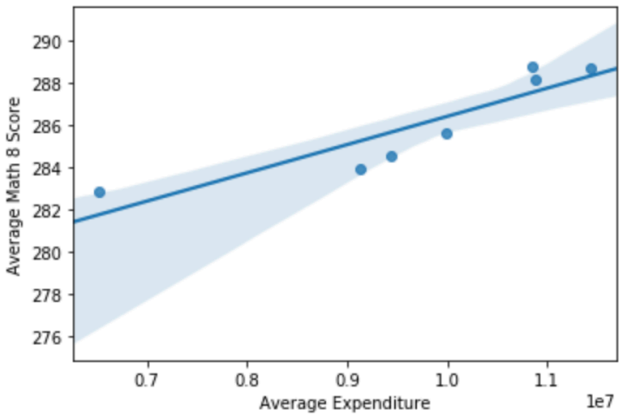
The results show strong correlations between spending and enrollment of different ethnicities. When each grade 12 demographic was compared to spending each had a strong correlation to spending because each p-value is very low (numbers in python truncated, so it looks like 0, but they are not exactly 0). Then looking at the R^2 value, it shows roughly 87% and 98% of enrollment is closely related to rates of spending. The numbers change when the three different races were modeled together. The regression shows that only African American or Black students of grade 12 are closely associated to spending based on the p-value < 0.05. Although the other two groups show good F-statistic and R^2 values for strong correlation, their p-values say that they are not statistically significant.

1.2 Linear Regression for Test Scores and Spending

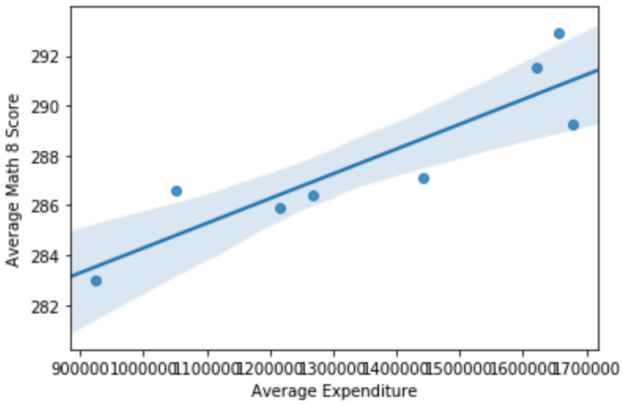
Three linear regression models were created for the test score data to see whether total spending effects test scores (or enrichment value). Data was only given for average reading and math test scores for grades 4 and 8 from odd years beginning after 1996 through 2013. Three states were used for this linear regression, one from higher spending, median spending, and lower range spending: California, Wisconsin, and Montana. Grade 8 average math scores were used.



This model shows average math grade 8 test scores for California and Average Expenditure in millions.



This model shows average math grade 8 scores for Wisconsin and Average Expenditure in millions.



This model shows average math grad 8 scores for Montana and Average Expenditure in thousands.

For this linear regression the models show high confidence intervals and less linear data

State	P-Value	R^2	F-Statistic
California	0.01	0.698	13.9
Wisconsin	0.005	0.814	21.83
Montana	0.003	0.787	22.21

Linear Regression data for State and Average Expenditure only.

State	P-Value	R^2	F-Statistic
California	0.01	0.698	13.90
Wisconsin	0.01	0.698	13.90
Montana	0.01	0.698	13.90

Linear Regression data for all States modeled together with Average Expenditure.

1.2.1 Results

The results for each state compared solely to Average Expenditure show that the test scores are highly correlated to educational spending. Low p-values so we reject the null hypothesis, and high R^2 values that suggest there is on average the data is closely fitted to the linear regression line. When the data is modeled together, it produced a slight correlation, but when compared to the previous regression for enrollment and spending, 69% correlation based on R^2 is not as strong.

DISCUSSION

1.1 Linear Regression for Enrollment Demographics and Spending

The strong associations found by the linear regression for Enrollment of American Indian, Hispanic, and Black grade 12 students suggests that the more spending put into education, then students of all diversities will benefit. More specifically, since Black grade 12 students showed a higher correlation than that of American Indian and Hispanic grade 12 students, the linear regression model suggests that black students would benefit more than the other two groups studied.

1.2 Linear Regression for Test Scores and Spending

The associations found for the linear regression of test score data of grade 8 math tests and spending show a slight association between the two sections of data. Since the Linear regressions showed about a 70% variable variation, we can assume that an increase in spending on education will potentially increase test scores.

CONCLUSION

Increases in educational spending will improve not only the quality and value of education for students (as quantified through test scores), but it will help improve diversity. From this study, decisions can be made to improve education by providing more educational funding for different programs to empower and lift up students. As this data is also locational, this study could assist in the empowerment of different locations that would benefit from increased spending.

Further considerations for this project include comparing this data to data that is scaled to reflect how spending is allocated per student, and how the size of each state makes the spending comparisons different. This is because of the reasoning that: since the state of California has a larger square footage than Colorado, it makes sense that spending is higher in that area because of square footage. Another factor to consider further would be to factor cost of living into the different spending figures as well as how the value of the dollar fluctuated. These changes would improve the models because they scale costs to validate the comparisons.

REFERENCES

[1] Gruber,Jonathan. *Public Finance and Public Policy*., 2016. Print.

[2] The Condition of Education - Preprimary, Elementary, and Secondary Education - High School Completion - Public High School Graduation Rates - Indicator May (2019), https://nces.ed.gov/programs/coe/indicator_coi.asp.

[3] State Compulsory Education Laws." Findlaw, https://education.findlaw.com/education-options/current-state-developments-in-compulsory-education.html.

[4] Univeristy of Wisconsin Madison, et al. Projections for Wisconsin High School Graduates. Applied Population Laboratory (APL), Dec. 2017, https://www.wisconsin.edu/education-reports-statistics/download/educational_statistics/high_school_graduate_projections/Grads_Project_Report_Dec_2017.pdf.

[5] Chappell, Bill. "U.S. Births Fell To A 32-Year Low In 2018; CDC Says Birthrate Is In Record Slump." NPR, NPR, 15 May 2019, https://www.npr.org/2019/05/15/723518379/u-s-births-fell-to-a-32-year-low-in-2018-cdc-says-birthrate-is-at-record-level.

[6] Hamilton, Brady E, et al. Births: Provisional Data for 2018. Division of Vital Statistics, National Center for Health Statistics, May 2019, https://www.cdc.gov/nchs/data/vsrr/vsrr-007-508.pdf.

[7] Manstead, Antony. The Wellbeing Effect of Education. Economic and Social Research Council, July 2014, https://esrc.ukri.org/files/news-events-and-publications/evidence-briefings/the-wellbeing-effect-of-education/.

[8] "U.S. Department of Education Budget History." Home, US Department of Education (ED), 18 Mar. 2019, https://www2.ed.gov/about/overview/budget/history/edhistory.pdf.