Investigating the Impact of School Funding on Academic Performance

Background & Motivation

School funding has emerged as a critical national concern, with debates intensifying in states like **Idaho**, where policymakers are actively reevaluating education budgets. Claims about the impact on student outcomes often lack empirical support, prompting this data-driven analysis to answer:

Does increased school spending directly correlate with improved academic performance?

This study aims to provide actionable insights for policymakers, including:

- Evidence-based recommendations for funding allocation.
- Identification of high-impact investment areas to maximize student success.

Research Objectives

This study aims to provide actionable insights for policymakers and voters, including:

1. Evidence-based recommendations for funding allocation.
2. Identification of high-impact investment areas to maximize student success.

Primary Question:

Analyze whether higher per-student funding correlates with higher graduation rates

Secondary Investigations:

Identify which state performs best despite funding opportunities and restrictions.

Identify which state performs word despite funding opportunities and restrictions.

Additional investigation and comparison between these two would be warranted to improve practices for schools overall.

Methodology

This information is available from the National Center for Education Statistics (NCES), a government office tasked with unbiased tracking of school performance nationwide.

- **Data Sources**:
  - **Performance Metrics**: State-level graduation rates.
  - **Funding Data**: Per-student expenditure totals and categorical breakdowns (where available).
- **Analysis Approach**:
  - Correlation analysis between funding levels and academic outcomes.
  - Regression modeling to isolate impactful spending categories.

Approach and Results

I gathered data gathered from the **NCES**, starting with the **Adjusted Cohort Graduation Rate (ACGR)**. Though the data ranged up through 2022, I opted to **ignore any data after 2019**, owing to changing variables caused by the global pandemic.

-States responded differently to challenges caused by the pandemic, causing too many variables to be accounted for in this study.

There were a few **null** values; these were replaced with the **mean value** for the state.

I tested the usability of the data by generating four predictive models before deciding which two to employ for the remainder of the study:

ARIMA: Handles individual state data best, and can look into the future

ETS: Also good for future estimation, well-suited for data over time

Linear Regression: Great for trendlines for multiple states simultaneously

Random Forest Regression: Similar to Linear Regression, but also able to handle complex information

The **Linear Regression** and **Random Forest** models most closely matched Actual and Predicted outcomes. And, as they can also handle information for multiple states simultaneously, I opted to use these for analysis once all the data had been gathered.

I gathered the spending information for the same timeframe, using the spending-per-student data. Dollar amounts were based on the **Consumer Price Index (CPI)**, prepared by the **Bureau of Labor Statistics, U.S. Department of Labor,** adjusted to a school-year basis to ensure consistency. The dollar mounts have been **adjusted to compensate for inflation.**

-It must be noted that these expenses, by ending in the year 2019, do not include expenditures from funds authorized by programs such as the CARES, CRRSA, ARP and similar supportive acts, which would otherwise dramatically alter the average spending amounts for the years.

Plotting the year-by-year spending amounts showed **entirely right-skewed distributions** and indications that more states spend below the median than above.

With all of the data prepped and ready, I began by comparing the average graduation rate with the average spending, providing confidence **shading for variability**. At first look, there **appears to be a strong correlation** between the two variables, both showing a positive relationship, increasing together over time. But the numbers at 2011 seem to show an inverse relationship; unfortunately there is not enough data to investigate further at this level.

I plotted a **comparison of the percent change** between the two data sets, and saw that there is a significant amount of **shift in spending amounts** from one year to the next, but **not nearly as much of a change in the graduation rate.** This belies the suggestion of a causational relationship, so I investigated the correlation directly:

Pearson correlation coefficient: 0.02823
Pearson r: 0.0282320519487733
p-value: (rounded to 5 decimal places)s (0.00005)

A Pearson correlation that low suggests there is **almost no confirmed correlation** between the two, and the p-value suggests that, while there is a **real relationship between the two, it appears to be weak.**

Wanting to see this for myself, I plotted each value of spending compared to the graduation rate for each state. It became evident that there were **some instances of a positive relationship** as well as **instances of a negative relationship.** It does not look as though one causes the other.

I plotted a list of all states comparing their per-student-spending to one another, and saw a **wide difference between the highest- and lowest-spending states**. Idaho, for example, spends consistently less than half of what is spent per student in New York.

Even though there won't be as much of a difference in terms of absolute percentages, I also plotted the compared graduation rates between the states. While there is definitely some distribution, the **states at the top and bottom differ from those regularly seen in the per-student-spending comparison.**

I plotted the graduation rates of all states of all years just to see how much variability there might be within one state and saw a definite **upward shift for most states** overall, but the **spending total does not seem to be related** to other states directly. The graduation rates did not show the same level of variability.

Looking at the state spending over time, there was too much for anything to be seen clearly. Instead, I looked at the change in spending and, aside from a handful of outliers, **spending rates are fairly constant from one year to the next.**

I plotted a color-coded representation of how much of a change in spending there was between 2011 and 2018. A deeper shade of green represented the most significant change in spending, scaling to red, which represented an overall decrease in spending. Seeing that there were **some states that decreased their spending, but there is still an overall upward trend in graduation rates**, seriously raises the question about the relationship between spending and graduation rates.

With this information, I locked in the coloration for each state, then plotted the overall change in the rate of graduation between 2011 and 2018 for each state. The resulting plot showed the surprising result: the colorization appears almost randomly distributed; **states that more greatly increased spending did not experience a corresponding increase in their rate of graduation**, nor did those states that more greatly reduced spending over this time frame. One state, Vermont, even spent **much more** and experienced a **decrease** in their graduation rate!

I reproduced this using the absolute average dollar amount spent over the time frame for each state and saw a similar result: **states that spend more per student do not experience a higher rate of graduation**. D.C. on average spent the most between 2011 and 2018, but has the lowest average graduation rate compared to other states.

In order to determine which school was able to perform the best regardless of spending, I plotted a forecast for each state using both **Linear Regression** and **Random Forest Regression**, as both are well-suited to this type of data and worked well in testing earlier.

I extrapolated data to predict graduation rates over the next **three years.** The Random Forest Regression only seemed to show a rough continuation where each set was finalized, but the Linear Regression plot showed a measure of trend, mostly in a positive direction. Overall, it is predicted that **almost all schools will maintain or improve graduation rates.**

West Virginia has the **highest** estimated graduation rate at 97.20.
New Mexico has the **lowest** estimated graduation rate at 77.87.

Conclusions

The Pearson numbers and p-values demonstrate there is no correlation between spending and graduation rate.

Examining the data state-by-state confirms this assessment, as there are sufficient examples of increased spending and increased dollar amount not resulting in a corresponding increase in graduation rate. The opposite example is also absent, in that states that have decreased spending and lower absolute spending per year have not seen a corresponding decrease in graduation rate.

From this data, it is therefore safe to conclude there is no correlation between the amount spent and the graduation rate. There is no recommendation to adjust spending for the purpose of causing an impact on graduation rates.

West Virginia boasts the most success at achieving a graduation rate regardless of spending, and New Mexico has the lowest success with graduation rates. It can be reasonably concluded that an analysis of policy and procedure differences between the states will prove beneficial.

Because the model predicts a future based on data that does not reflect the impact of the global pandemic, it is interesting to note that the actual result as to which states would have the highest and lowest graduation rates matched reality according to the actual data set; West Virginia did, in fact, have the highest graduation rate nationwide and New Mexico had the lowest graduation rate nationwide.

Proposals for Future Research

Something this data does not demonstrate is the quality of the education received. It is not known whether states follow the same standard to qualify for graduation; are graduates equally qualified between states?

Additionally, the data does not make clear if it includes private and homeschooling students in their calculations. If more affluent locations choose to provide extra education, is the expense part of this calculation, and is the student counted in the “graduation rate” calculations?

More rural locations may have a lower graduation rate because students leave school in order to begin professional employment in a trade or family farm. This would cause a negative impact on the numbers that is not reflective of the education provided to the students.