Correlation Between School Funding & Educational Achievement

CS 591 L1 Data Mechanics

Renzo Callejas
Hannah Schurman

Questions

We tried to answer the following two questions by using data for **Boston Public High Schools**:
Is there a correlation between a school's average **SAT score** and **graduation rate** and

a. That school's funding?
b. The sum total funding of that school's closest three neighboring schools?

Our study focuses on the algorithms required to find these correlations, as we did not receive the full BPS (Boston Public Schools) data required to find real correlations.

Implications

The answer to the questions above will help the Massachusetts government understand the effectiveness and impactfulness of their school funding. This can provide them the means to investigate why their funding is or is not effective and impactful, and make appropriate changes.

Data

Original: Funding *(from BPS),* SAT *(from BPS),* Grad Rates *(from BPS),* Location *(from City of Boston Data Portal)*
New: Fund_SAT, Fund_GradRates, Fund_Location, corr_SAT, corr_gradrates, corr_location

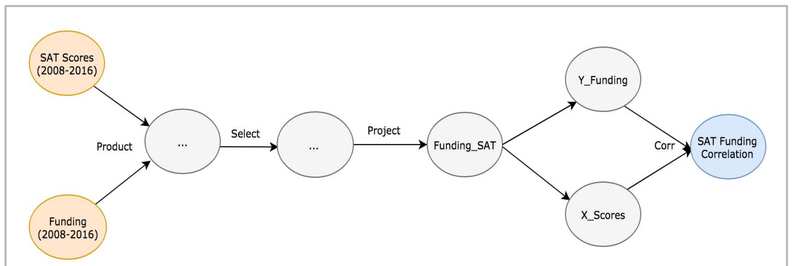
Further Work

More accurate results would come from taking data following an entire high school class and looking at the average funding for the four years they were in high school instead of just the funding their school received during their senior year. Further studies should also include additional independent variables in the model, such as neighboring crime and average household income

School Funding Correlation

Creating Data Sets Using Transformations

Step 1: Take the **product** of the funding and SAT, GradRates, or Location data for 2008 - 2016 and put it into a new dataset X
Step 2: Select only the tuples in X that have funding data and SAT, GradRates, or Location data for the same school and put them in dataset Y
Step 3: Project the dataset Y into the following format: {School Name, [SAT, GradRate, or Location data for 2008-2016], [Funding for 2008-2016]}



Finding Correlations For Each School

To find the correlation between funding and SAT scores & graduation rates for each school, we created a vector F that listed funding for each school for the years 2008 to 2016. We also created vectors S and G that listed SAT scores and graduation rates for those years, respectively. For each school, a correlation coefficient was found between F and S, and F and G.

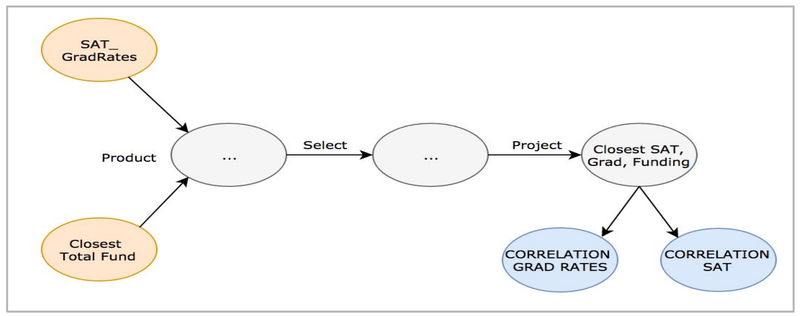
Conclusion

Using the algorithms we have developed, further researchers who are able to collect the required education data will have the ability to find accurate correlations between school funding and educational achievement. The question of whether funding causes change in educational achievement and not the other way around is also important, and will be determined by looking at how states determine funding.

School Neighbor Funding Correlation

Creating Data Sets Using Transformations

Step 1: Using product, select, and project, create a list **SAT_GradRates** that lists the SAT scores and graduation rates for each school.
Step 2: Create the list **Closest_Total_Fund** that lists each school along with the total sum funding of that school's closest three neighbors for each year in 2008-2016.
Step 3: Take the **product** of the two lists.
Step 4: Select the data points that contain information relating to the same school.
Step 5: Project each point to the following format: {School Name, [SAT or GradRate data for 2008-2016], [Total neighbor funding for 2008-2016]}



Finding Correlations For Neighboring School's Funding, SAT & GradRates

To find the correlation between a school A's SAT scores and graduation rates, and the total sum funding of that school's neighbors B, C, and D, for each school A, we created a vector A that listed the total sum funding of schools B, C, and D for the years 2008 to 2016. We then found the correlation between vector A and the vectors S and G that listed school A's SAT scores and graduation rates for 2008-2016.

corr = cov(x,y)/(stddev(x)*stddev(y))