Renzo Callejas
Hannah Schurman
Data Mechanics

# School Funding Project

## Introduction

The School Funding Project began as a way to investigate how effective and impactful a school's funding is in regards to that school's educational outcomes. The project was also created to understand the overall impact that school funding has on a region's educational outcomes. Understanding this relationship between school funding and educational success is important for local governments, because it allows them to evaluate their funding decisions and change them in a way that benefits students.

In order to test the relationship between school funding and educational outcomes, we focused on Boston Public High Schools. We defined "educational outcomes" for each high school to be the school's average SAT score and graduation rate. The goal of the project was to answer two questions. One, is there a correlation between a school's educational outcomes and that school's funding? Two, is there a correlation between a school's educational outcomes and the sum total funding of that school's closest three neighboring schools?

In this paper, we will go through the data sets we used and created, and display some visualizations created from this data. Additionally, we will individually look at question one and question two, and explain how we transformed our data in order to put it into the format required by the correlation tool. Then, we will describe how we found

the correlations. Lastly, we will conclude and offer ideas for future improvements to our project. Due to a lack of data received from BPS, we do not currently have results.

**Data Sets and Visualizations**

From the City of Boston Data Portal we retrieved a data set which contained a list of BPS schools and the school's location. From the BPS Offices of Data and Accountability we gathered, for each BPS High School and for a varying amount of years between 2008 and 2016, data on SAT scores, graduation rates, and general funding allocations. For the years between 2008 and 2016 that we did not have data on, we created dummy data in order to allow for testing. This collection of data sets was combined and filtered using different transformations to create the following data sets: Fund_SAT, Fund_GradRates, Fund_Location. As can be inferred from the titles of these data sets, in each data set there is SAT, graduation rates, or location information about our list of BPS High Schools.

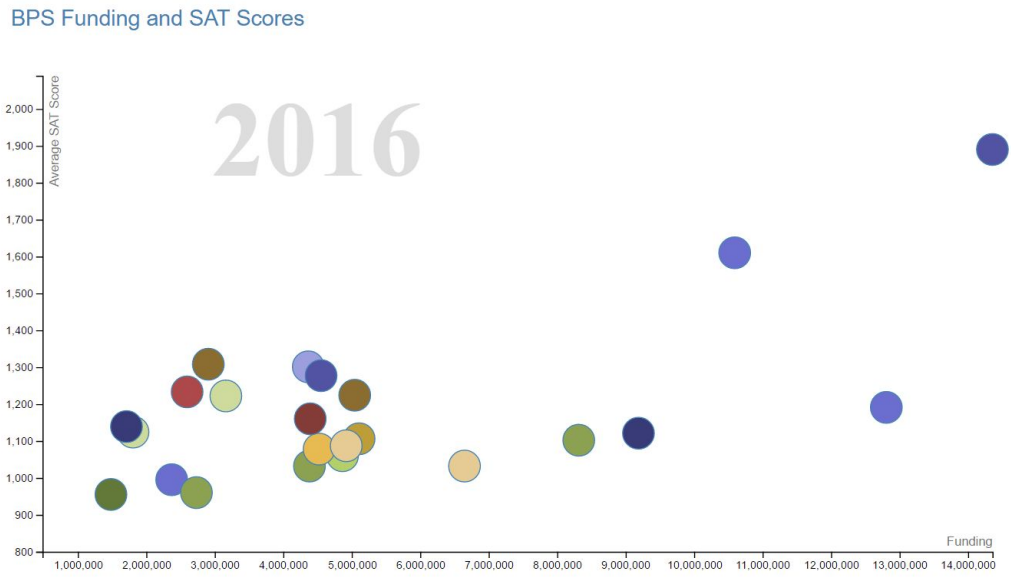


Image #1

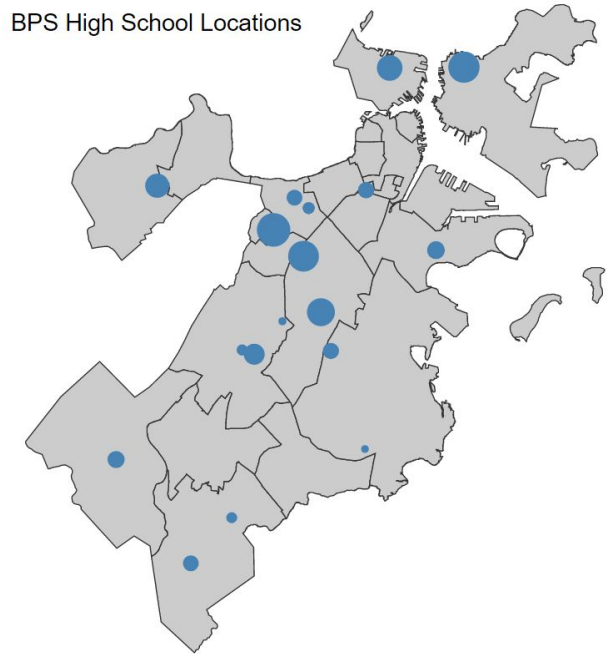


Image #2

**Question 1: Is there a correlation between a school's educational outcomes and that school's funding?**

**Transformations**

In order to find a correlation between funding and SAT scores and funding and graduation rates, we created one data set that listed the funding and SAT data for each school, and another data set that listed the funding and graduation rate data for each school. We also created another data set that listed the funding and location data for each school, which we needed to answer Question 2. Because the initial funding, SAT, and location data were all in the same format, the new Fund_SAT, Fund_GradRates, and Fund_Location datasets were all created using the same transformations. Below is an example and an illustration of how the Fund_SAT dataset was created.

Initial Format of SAT Data: {'Name': ~Name of HS~, 'SAT': {'2008': ~score~, … , '2016': ~score~}}

Initial Format of Funding Data: {'Name': ~Name of HS~, 'Funding': {'2008': ~$funding~, … , '2016': ~$funding~}}

Step 1: Take the product of the funding and SAT data and put it into a new dataset X

Step 2: Select only the tuples in X that have matching school names and put them in dataset Y

Step 3: Project the dataset Y into the following format: {'Name': ~Name of HS~, 'SAT':

{‘2008’: ~score~, … , ‘2016’: ~score~}, ‘Funding’: {‘2008’: ~$funding~, … , ‘2016’: ~$funding~}}

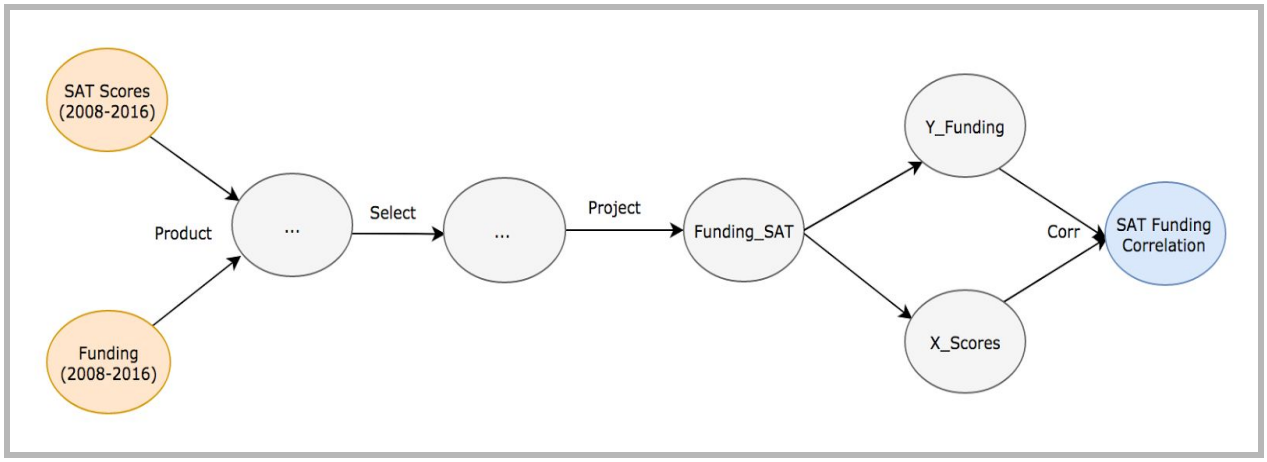


Image #3

**Correlation**

Next we will describe how we found the correlation between funding and SAT scores. We used an identical method to find the correlation between funding and graduation rates.

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

**Question 2: Is there a correlation between a school’s educational outcomes and the sum total funding of that school’s closest three neighboring schools?**

**Transformations**

Initial Format of SAT_GradRates: {‘Name’: ~Name of HS~, ‘SAT’: {‘2008’: ~score~, … , ‘2016’: ~score~}, ‘GradRates’: {‘2008’: ~rate~, … , ‘2016’: ~rate~}}

Initial Format of Closest_Total_Fund: {‘Name’: ~Name of HS~, ‘Closest’: {‘2008’: ~total~, …, ‘2016’: ~total~}}

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 and put them into a new dataset X

Step 4: Select only the tuples in X that have matching school names and put them in dataset Y

Step 5: Project each point in Y to the following format: {‘Name’: ~Name of HS~, ‘SAT’: {‘2008’: ~score~, … , ‘2016’: ~score~}, ‘GradRates’: {‘2008’: ~rate~, … , ‘2016’: ~rate~}, ‘Closest’: {‘2008’: ~total~, …, ‘2016’: ~total~}}

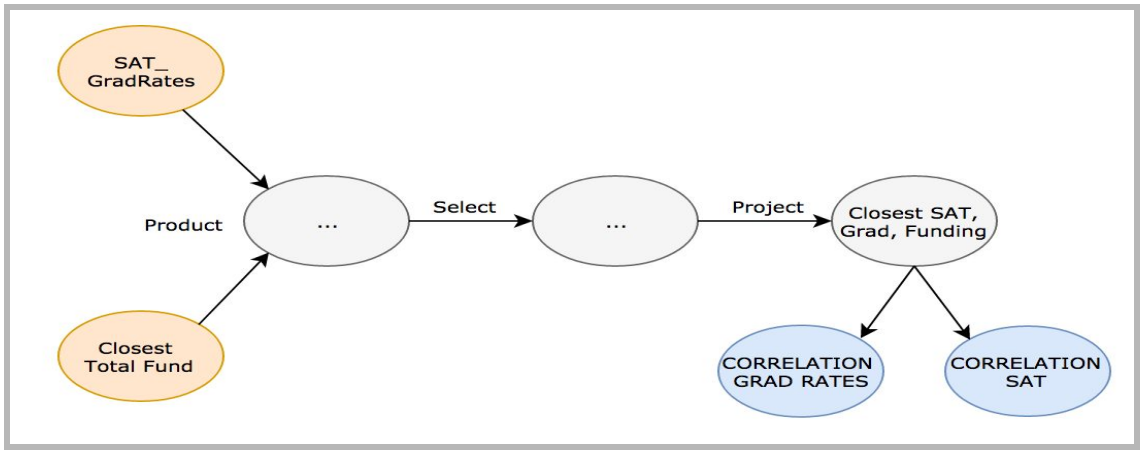


Image #4

**Correlation**

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.

**Conclusions**

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 each school's funding and educational achievement. Then, they can look at each school individually or take an average of each of the school's correlations, to get an average correlation. 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.

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. This will give more accurate results because the average funding over the four years is more responsible for whether a student will do well on their SAT or graduate. This is especially true for SAT scores, because many students take their SATs in Fall of their senior year, and so much of the information tested is based on education they received in their first three years of high school. Further studies should also include additional independent variables into the model, such as neighboring crime and average household income. Adding additional variables can help increase the confidence of the results and paint a larger picture of what truly predicts student success.