Analysis of Academic Achievement Fluctuation Due to Revenue in School Districts within the United States

College of Information Science and Technology
Drexel University

March 17, 2019

Abstract - *This project was conducted for the visualization and discovery of how large amounts of money to schools can possibly affect the performance of scores (Academic Achievement) and improvement of (Enrollment) of students from all states in the United States.*

---

1 Introduction

In all of the United States, there is an equal opportunity to attend school, become well-educated, and discover the open resources available for further enhancement of education. Districts in the U.S. receive revenue from either Federal, State, or Local funds. This revenue income is to support the educational system in each district. It is appointed accordingly and dispersed based on a needs basis. According to Data-First.org, who summarized, from the National Center for Education Statistics (NCES), the national total revenues in 2012 were broken down as.

- Local Government: 44.8%
- State Government: 45.1%
- Federal Government: 10.1%

The total revenues include all money that a district receives over a year. Not all districts receive equal amounts from local, state, and federal sources. It varies from state to state and even from district to district.

Stated in Data-First.org, “Across the nation, local revenue sources provide about 40 percent of all education dollars. This ranges from a low of 5 percent in Vermont to a high of 58 percent in Illinois. While cities and counties typically raise these funds through property taxes, they also rely on other measures, like school fees and sales tax. States typically provide about 50 percent of local education funds. Each state has its own system for distributing these funds to school districts, and the formulas can be very complex. The federal government is usually of about 10 percent of local education dollars. The most well-known source of federal education dollars is Title I of the No Child Left Behind/Elementary and Secondary Education Act, which is slated for schools and districts with high percentages of poor children.”

1.1 Gathering of dataset

The importance of education in the United States is a topic that will continue to be relevant and important to many data analysts and statisticians.

We gathered a dataset from Kaggle, titled

"U.S. Education Datasets: Unification Project."

This dataset contains K-12 financial, enrollment, and achievement data in one place.

1.2 Data Processing

This dataset contains important parts from other vital sources of information including.

Enrollment - https://nces.ed.gov/ccd/stnfis.asp

Financials-

https://www.census.gov/programs-surveys/school-finances/data/tables.html

Academic Achievement-

https://www.nationsreportcard.gov/ndecore/xplore/NDE

2 Acquisition and Questions

2.1 Thoughts and Considerations

Interested in the quality and enhancement of revenue money improving academic scores, our group pondered at the idea. We asked several questions.

- Do states that receive more money, necessarily have better scores?
- Do states that receive less money, have worse scores?
- Over the years, what fluctuations have been the most drastic? Which have possibly stayed the same?
- Did enrollment improve when states received more money, did it decrease with less money appointed?
- Or did scores and enrollment keep at a constant?

In this project report, we realized that with this massive data set there are many questions that can be asked and many that can be answered.

These questions are what we are going to answer with our visualizations.

2.2 Hypothesis

Our group hypothesis is that the revenue money given to the schools may have created an improvement on the math and reading scores of all grades in the states. We came to this hypothesis as a positive outlook on the improvement of the education system in the U.S.

For the summary section there will be a calculation and measure of how the states did for the scores and the enrollment of students in the schools. Enrollment in school is important for a thriving district to grow and flourish, without enrollment there is no room for continuation of improvement. Grades are of a great importance to society and to the community in the districts. Math and reading grades will be shown in the data for comparison through out all the states.

There could possibly be no change in the improvement of both scores in exams and enrollment of students due to unknown variables, but for the most part the money appointed by the government should promote a positive aspect to both mentioned.

3 Data Variables

The variables of data that are in our data set is as follows:

Identification:

- ❏ PRIMARY_KEY : A combination of the year and state name.
- ❏ YEAR
- ❏ STATE

Enrollment:

A breakdown of students enrolled in schools by school year.

- ❏ GRADES_PK: Number of students in Pre-Kindergarten education.

- GRADES_4: Number of students in fourth grade.
- GRADES_8: Number of students in eighth grade.
- GRADES_12: Number of students in twelfth grade.
- GRADES_1_8: Number of students in the first through eighth grades.
- GRADES 9_12: Number of students in the ninth through twelfth grades.
- GRADES_KG_12: Number of students in Kindergarten through twelfth grade.
- GRADES_ALL: The count of all students in the state. Comparable to ENROLL in the financial data (which is the U.S. Census Bureau's estimate for students in the state.)

The extended version of states_all contains additional columns that breakdown enrollment by race and gender. For example.

- Grades_ALL_AS: Number of students whose ethnicity was classified as "Asian".
- Grades_ALL_ASM: Number of male students whose ethnicity was classified as "Asian".
- Grades_ALL_ASF: Number of female students whose ethnicity was classified as "Asian".

The represented races include AM (American Indian or Alaska Native), AS (Asian), HI (Hispanic/Latino), BL (Black or African American), WH (White), HP (Hawaiian Native/Pacific Islander), and TR (Two or More Races). The represented genders include M (Male) and F (Female).

Financials:

A breakdown of states by revenue and expenditure.

- ENROLL: The U.S. Census Bureau's count for students in the state. Should be comparable to GRADES_ALL (Which is the NCES's estimate for students in the state).
- TOTAL REVENUE: The total amount of revenue for the states.
  - FEDERAL_REVENUE
  - STATE_REVENUE
  - LOCAL_REVENUE
- TOTAL_EXPENDITURE: The total expenditure for the state.
  - INSTRUCTION_EXPENDITURE
  - SUPPORT_SERVICES_EXPENDITURE
  - CAPITAL_OUTLAY_EXPENDITURE
  - OTHER_EXPENDITURE

Academic Achievement:

A breakdown of student performance as assessed by the corresponding exams (math, reading, grades 4 and 8)

- AVG_MATH_4_SCORE: The state's average score for fourth graders taking the NAEP math exam.
- AVG_MATH_8_SCORE: The state's average score for eighth graders taking the NAEP math exam.
- AVG_READING_4_SCORE: The state's average score for fourth graders taking the NAEP reading exam.
- AVG_READING_8_SCORE: The state's average for eighth graders taking the NAEP reading exam.

3.1 Data Explanation

In-depth explanation of the dataset variables is as follows:

States: For the states, this dataset contains all 50 states. So a level playing field is appropriate when comparing states and their scores.

Year: For the years, this dataset contains from the year 1992 to 2016. 24 years of data.

3.2 Data Discrepancies

There are some missing values. The values are within the column.

Enroll (Column D) from D2 to D52, values are missing.

Other_Expenditure (Column L) from L2 to L52, values are missing.

4 Tools

4.1 Software Used

We used a variety of software tools including.

Python - numerical processing for statistical analysis of mean, differences, median, and range, to name a few. Matplotlib is a visual library that is used for analysis. We use Matplotlib in one of the visualizations towards the end.

Tableau - visual graphs that represented features we could not identify solely by looking at the mass amount of data

5 Python

5.1 Interpretation

Using Python in this report, was important so we could denote possible calculations that could be of importance to the datasets interpretation. First what we did was to convey the dataset into Pandas, which converts the csv file into of tabular data form, also known as a DataFrame. Pandas offers powerful, expressive, and flexible data structures that make data manipulation and analysis simple to calculate. Another module that we decided to implement as well is a popular module named Numpy. Numpy is used for numerical processing within Python. It is a library, adding support for large, multi-dimensional arrays and matrices, included with a large collection of high-level mathematical functions to operate on these arrays.

5.2 Calculations (Overall)

Calculating the arithmetic mean we found out several averages:

Total Revenue Mean: 9092081.85

Federal Revenue Mean: 766372.33

State Revenue Mean: 4216552.94

Local Revenue Mean: 4109156.56

Total Expenditure Mean: 9196680.86

The information provided by Python consists of the averages from Revenues and Expenditure from the years 1992 to 2016.

The average of Enroll from 1993 to 2016 was:

Enroll Mean: 915930.820

The column enroll did not have data from 1992 so that is why its information started from 1993.

Calculating the medians for all the years (1992-2016):

Median of Total Revenue: 5079546.0

Median of Total Expenditure: 5234505.5

5.3 Percentiles

As it turns out, medians are just an instance of a more general kind of analysis called *percentiles.* For the median (the 50th percentile), 50% of the falls on either side. This is why it is thought of as a kind of middle. However, the concept generalizes to provide an idea of where say, the middle 50%, a.k.a the *interquartile range* (IQR), of your data are located.

Calculating the percentiles for all the years (1992-2016):

Total Revenue is:

25th Percentile: 2186305.25

50th Percentile: 5079546.0

75th Percentile: 10859847.75

Total Expenditure is:

25th Percentile: 2165404.5

50th Percentile: 5234505.5

75th Percentile: 10745191.0

With these calculations we can see that Total Revenue and Total Expenditure is closely similar to one another in the millions. This is appropriate since they produce with what they are given. The money that is given to the school districts from the Federal, State, and Local level has to be used for producing better scores and better enrollment for the students and the community. This calculation section focuses on the general means and medians of the overall data, meaning for all 24 years of data. Further centralized information will be discussed in the next section.

6 Tableau / Visualization

6.1 Visualization #1:

6.1.1 Tools

This visualization from tableau is called chart lines (continuous) , it represents several columns from the data in the dataset.

6.1.2 Methods (Overview)

This visualization shows an overview of the selected columns which are named from top to bottom (avg math score 4th grade, avg math score 8th grade, avg reading score 4th grade, avg reading score 8th grade, avg expenditure per student, and avg revenue per student) over the course of the years which are from 1992 to 2016

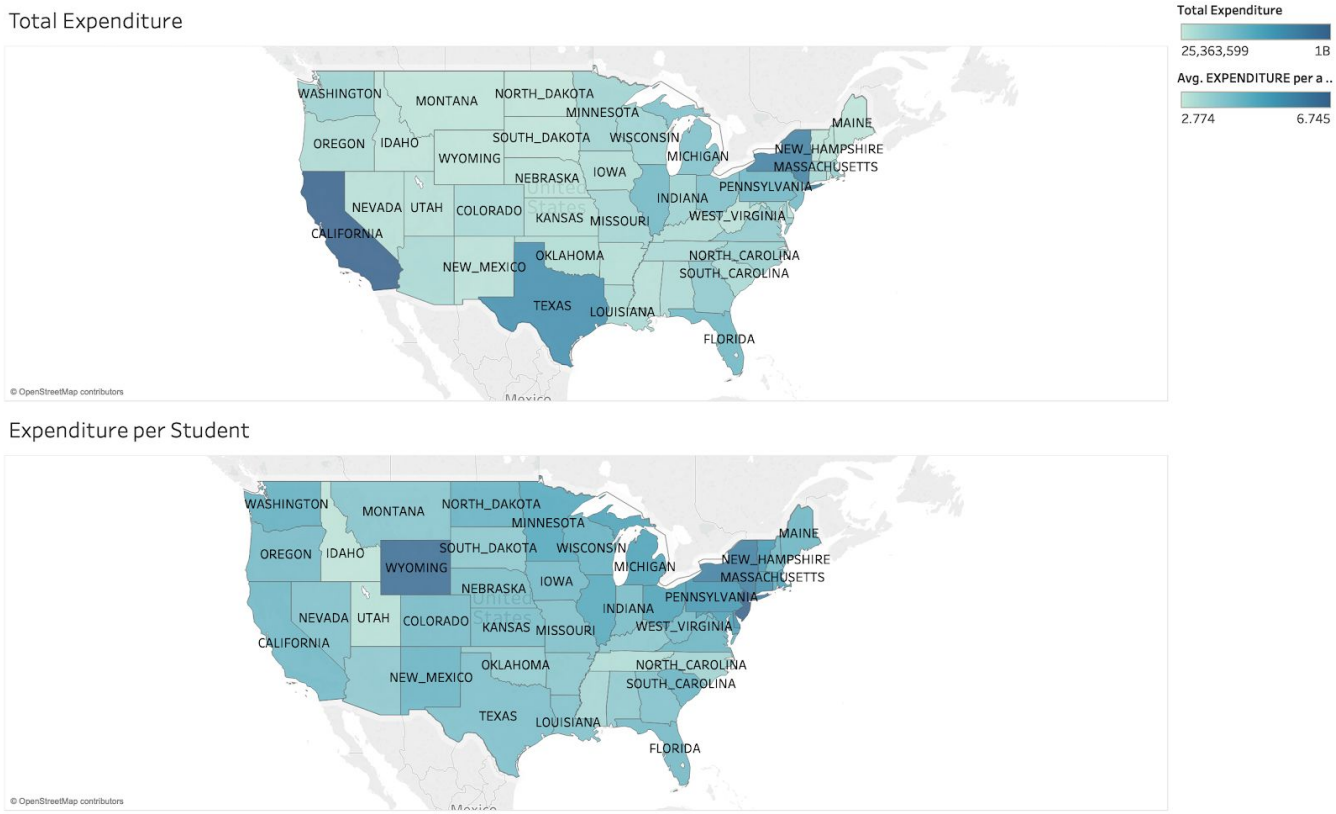
6.1.3 Results

The pattern shown by visualization number 1 reveals the correlation among average math scores between 4th and 8th graders.

6.1.4 Interpretation

With this graph we can interpret that with an increase in expenditure and revenue per student there seems to be an overall increase in the math scores for both 4th and 8th graders. the average reading scores for the 8th graders is the only column of data that does not seem to follow an upward flow.

6.2 Visualization #2:



6.2.1 Tools

This visualization from tableau is called maps, it describes total expenditure and expenditure per student of each state.

6.2.2 Methods (Overview)

This visualization shows an overview of the sum of total expenditure and average expenditure per student of each state throughout the years which are from 1992 to 2016.

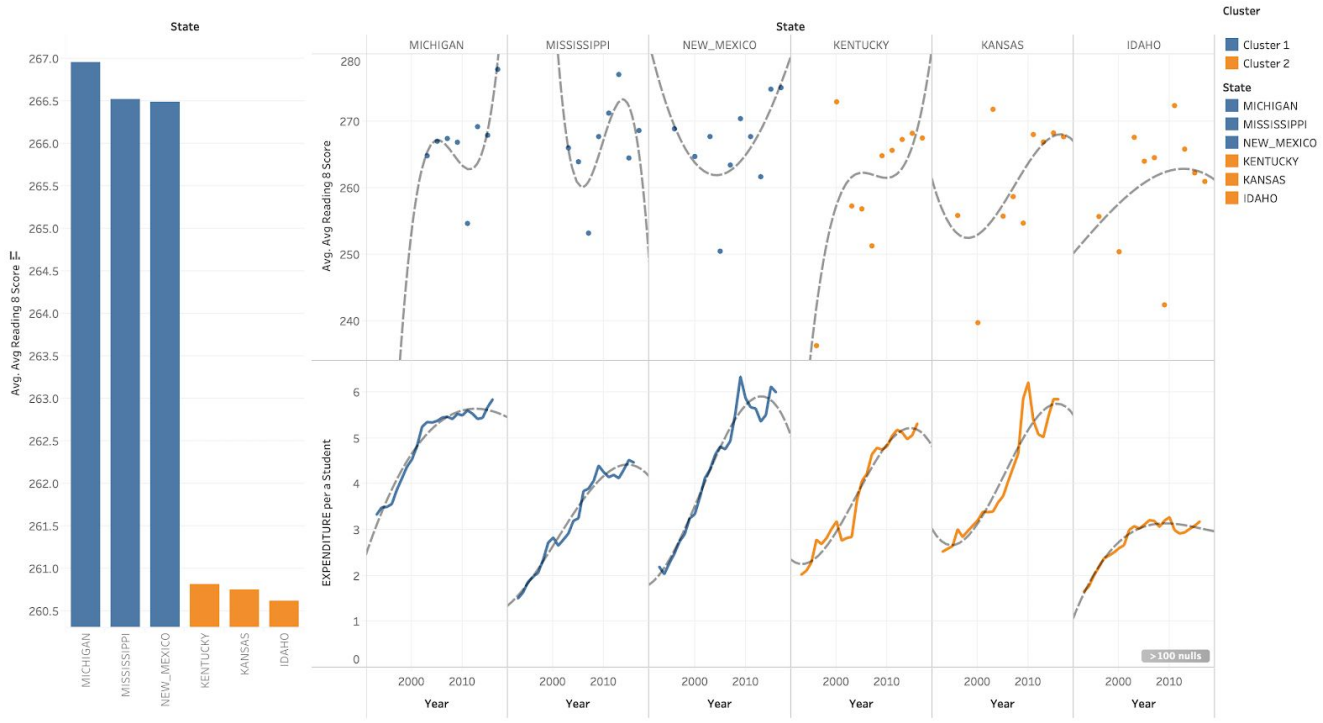
6.2.3 Results

Darker colors represent a relatively higher expenditure. In the first map, California, Texas and New York stand out with darker colors. In the second map, Only Wyoming has dark colors, but other states have similar pale colors.

6.2.4 Interpretation

With this visualization, we can interpret that the first map indicates California, Texas and New York had higher expenditure than any other state from 1992 to 2016. However, the second map presents per-student spending is standardized in all states except Wyoming.

6.3 Visualization #3:



6.3.1 Tools

This visualization from tableau is called Bar chart and chart lines (continuous), it represents the top 3 states and bottom 3 states of average reading score 8th Graders and expenditure per student.

6.3.2 Methods (zoom)

This visualization focuses on the top 3 states and the bottom 3 states of average reading score 8th Graders throughout the years which are from 1992 to 2016.

6.3.3 Results

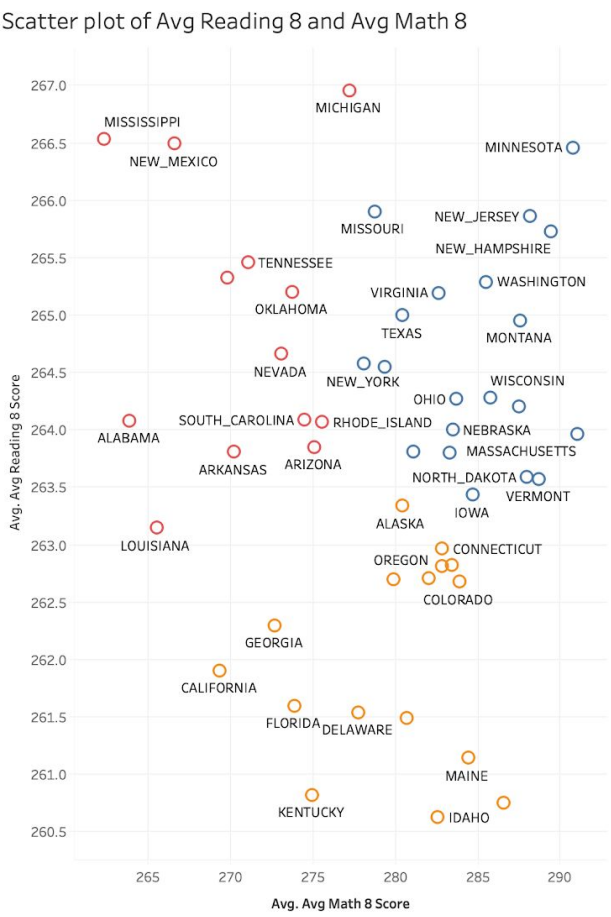
In the bar chart, the average score difference between the top 3 states and the bottom 3 states is about 7 points. The chart lines (continuous) of expenditure per student displays the pattern of rising, but the chart lines (continuous) of average reading score 8th graders presents no pattern.

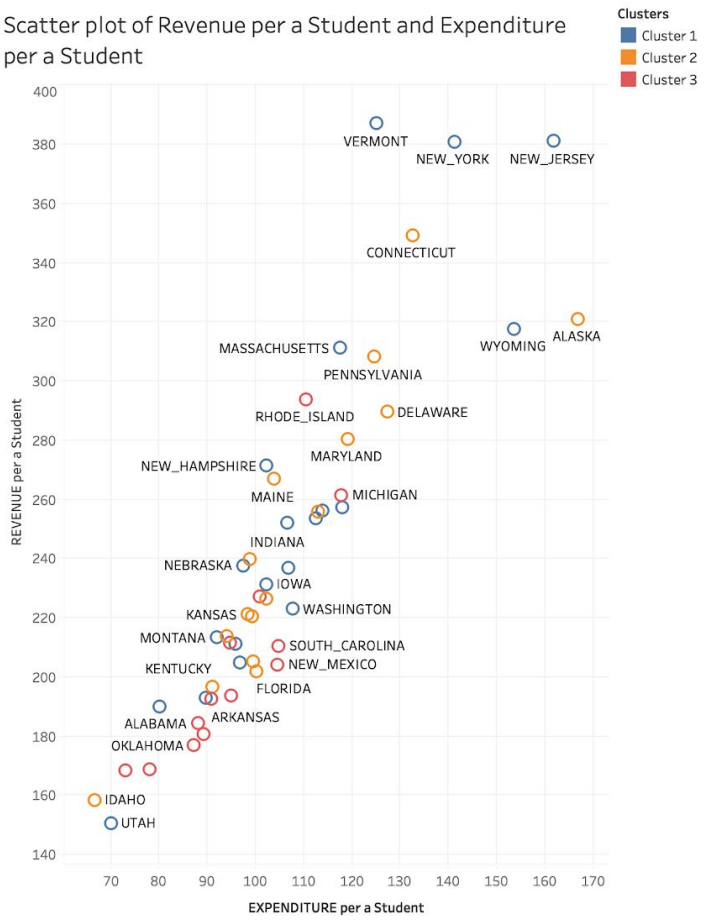
No pattern was found in the reading scores of 8th grade. It delivers it hard to believe that expenditure per-student affects the reading score 8th graders.

Correlations between the reading scores and the expenditure per student are not found in the top three states, but it is observed in the bottom three states, especially in Kansas and Idaho. We can interpret the increase in expenditure per student in the bottom three states as affecting scores.

6.3.4 Interpretation

6.4 VISUALIZATION #4:





6.4.1 TOOLS

This visualization from tableau is called scatter plot chart; it represents the positions of all state on a graph of average math and reading scores 8th graders and a graph of revenue and expenditure per student.

6.4.2 METHODS (ZOOM)

In the first scatter plot, all states are separated into three clusters. The clusters from the first scatter plot are utilized as an indicator in the second scatter plot.

6.4.3 RESULTS

Details of the cluster from the first scatter plot chart are given below.

Cluster 1: good scores in reading and math. (Blue)

Cluster 2: good scores in reading but not in math. (Red)

Cluster 3: Good scores in math but not in reading. (Orange)

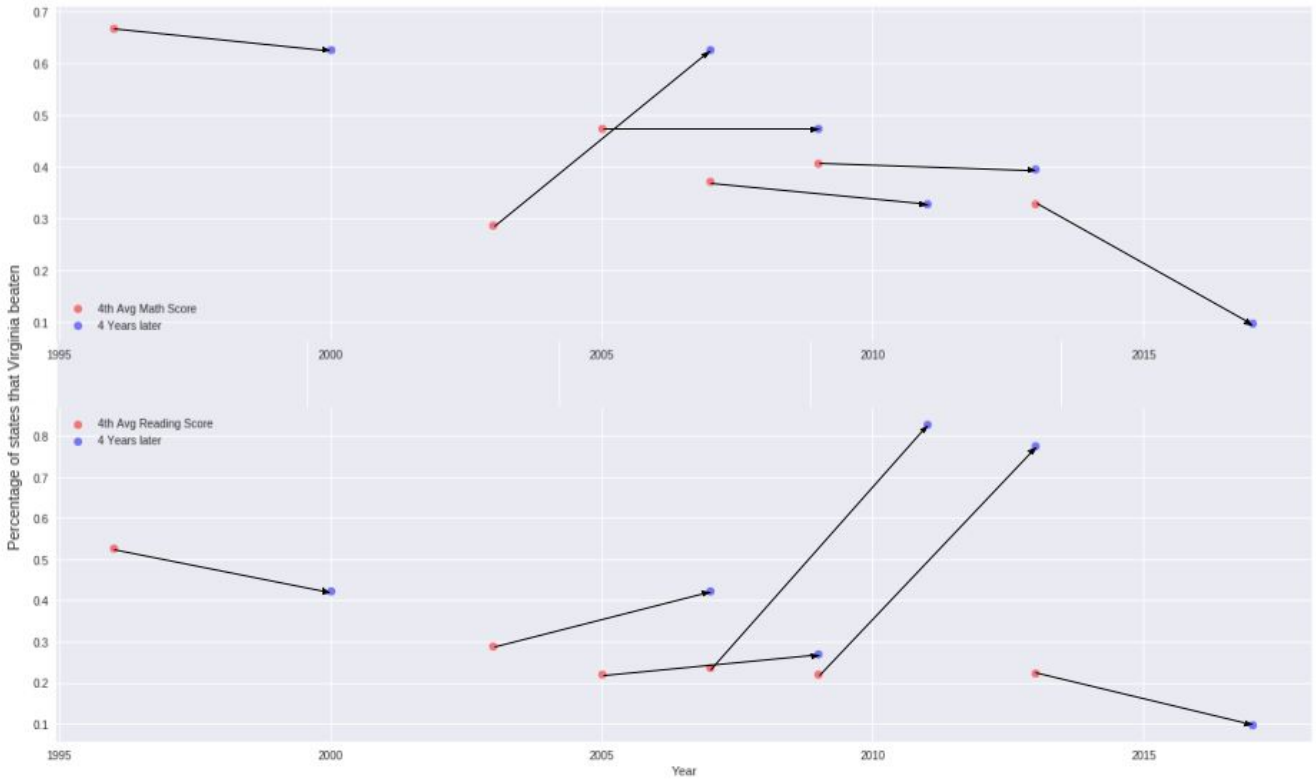
The second scatter plot displays a tendency to increase expenditure per student as revenue per student increases.

6.4.4 INTERPRETATION

UTAH is in cluster 1(Blue) which means good scores in math and reading, has relatively less expenditure per student and revenue per student. It indicates Utah led good scores with spending less budget.

Alaska and Connecticut in cluster 3(Orange) spent comparatively more money but didn't receive good scores.

6.5 VISUALIZATION #5:



6.5.1 TOOLS

This visualization from python matplotlib is called scatter plot chart; it represents the percentage that virginia state beaten in math scores and reading scores cores in 4th and 8th grade. to help connect data pair, i use drawing tools to add arrows between data pairs.

6.5.2 METHODS (FILTER, DETAIL)

States other than virginia are filtered out. To compare the same group of students, we select 4th grade scores at year 1996, 2003, 2005, 2007, 2009 and 2013. correspondingly, 8th grade scores at year 2000, 2007, 2009, 2011, 2013 and 2017 are selected as well. to make arrow more intuitive, we change rank to percentage of states beaten so an arrow pointing to upper means an increasing trend, vise versa.

6.5.3 RESULTS

the plot shows that scores of 4th grade at year 1996, 2013 are decreasing while the others are increasing.

6.5.4 INTERPRETATION

while we found virginia got a much larger sig fund at 2008 rather than other years, we want to check it this data is reasonable. a fund at 2008 could be used on 4th grade students in year 2008, 2007, 2006, 2005, 2004. in the visualization, scores in this interval are all increasing. therefore, this data could be correct and not mistyped.

7 Pitfalls/Limitations

In Visualization #3 there is some pitfall of some sort in the states Kansas and Idaho. There seems to be no pattern involved in those states at all. This pitfall creates a shroud of misinformation that is hard to decipher, since we cannot identify what it means, it is hard to interpret.

Another limitation we have with the data, is a good question that was pointed out in class. Are the test scores standardized throughout all of the states? This question raised some good points, not all tests take the same exact exams.In an online article titled, "What Tests Does Each State Require?" by Catherine Gewertz. This article highlights the exams that states take and exit exams for graduation. According to the article, only one-third of the states use the PARCC or Smarter Balanced tests. Fifteen states and the District of Columbia will administer PARCC or Smarter Balanced tests in the spring of 2019. That's five fewer than in 2016 and 2017. Thirty-two states use tests they designed or bought. Three states give hybrid tests. Two mix their own questions with items from the PARCC/New Meridian item bank, and one adds its own questions onto the full Smarter Balanced test. Twenty-five states require students to take the SAT or ACT, the same number as in 2016 and 2017. That number had been climbing steadily - from seven states a decade ago - as states looked for ways to encourage students to go to college.Thirteen states require students to pass a test to get a high school diploma, one more than in 2017. In some states, students can use projects or portfolios to meet the requirement. Exit exams were more popular before. In 2002 more than half of the US required. The test scores that we acquired from this dataset where from the U.S. Census Bureau and from National Center for Education Statistics (NCES). In hindsight, these test scores were standardized in a certain manner, that is difficult to decipher but these grades are standardized for equal comparison.

8 Conclusion

In conclusion, throughout our analysis we found out very insightful meanings from revenue and expenditure per student. As well as reading and math scores in 4th and 8th grade. For the most part, Visualization #1 shows there is a positive correlation (overall) of improvement of grades as the revenue money is awarded to the districts. This is a good indication of a healthy school environment here in the United States. The few states that saw a slight decrease, need to re-evaluate their models and maybe hire better teachers or principals that can implement better test scores. There are some states that indicate low average expenditure per student will result in a low 8th grade reading score, Visualization #4 proves this in the state of Idaho.

9 References

[1] Center for Public Education. (2012) "How much money does our school district receive from federal, state, and local sources?" Retrieved from http://www.data-first.org/data/how-much-money-does-our-school-district-receive-from-federal-state-and-local-sources/

[2] Garrard, R. (2019) "U.S. Education Datasets: Unification Project" Retrieved from https://www.kaggle.com/noriuk/us-education-datasets-unification-project

[3] Gewertz, Catherine. (2017) "What Tests Does Each State Require?" Retrieved from https://www.edweek.org/ew/section/multimedia/what-tests-does-each-state-require.html