BIOS 611 Final Project

Jeffrey Ayers

December 5, 2022

Education Trends for the United States

(Note: if you are reading the PDF of this file from the git repository, it may be out of date. It is included in the repo for convenience but the canonical writeup is in the org file in the repo. See the README instructions for how to build this pdf with make).

Introduction

As a Ph.D. student in the mathematics department instruction is a big part of my job, and one that I've become increasing passionate about. Having grown up with a parent who is a teacher, and now becoming one I have developed an growing interest in seeing what factors lead to student success, especially from a monetary standpoint. I was fortunate to grow up in a state that places high value on education (Massachusetts) and have directly seen the benefits of a large instructional budget, but what about nationwide? Can we see any direct impact on student performance based on this? Moreover, what other hidden trends can we see from data? In this writeup I analyze data from the National Center for Education Statistics to see what trends occur in the US, and the Northeast.

Preliminary Figures

The data was retrieved from Kaggle: US Education Dataset, and minor cleaning was done to it; namely converting column names to lowercase, and removing the underling strings corresponding to state names. These data contained a wide data set of columns corresponding to states, years and a multitude of statistical information: Total enrollment, financial information, test scores from the NAEP exam for grades 4 and 8, as well as expenditure. For example, in the year 2011 we can see what the total enrollment was for each state.

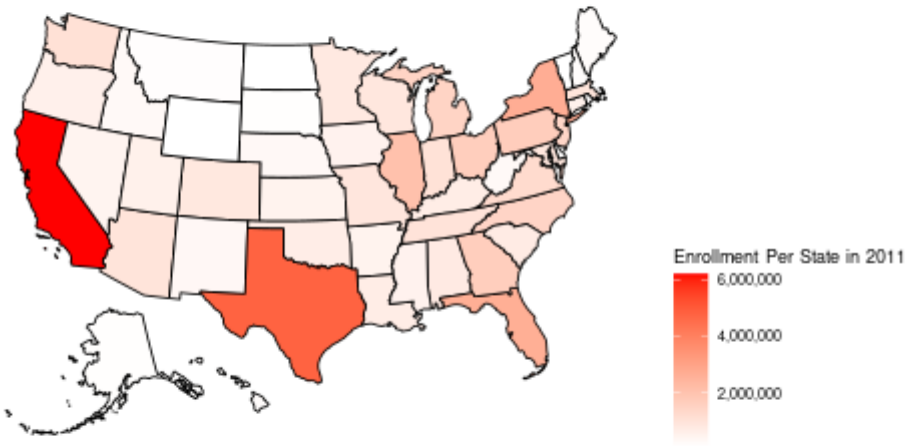

Figure 1: Total student enrollment per state in the year 2011

Correspondingly we can also view the total expenditure in dollars of each state, the total instructional expenditure, and the average NAEP math scores of each student in the states for grades 4 and 8:

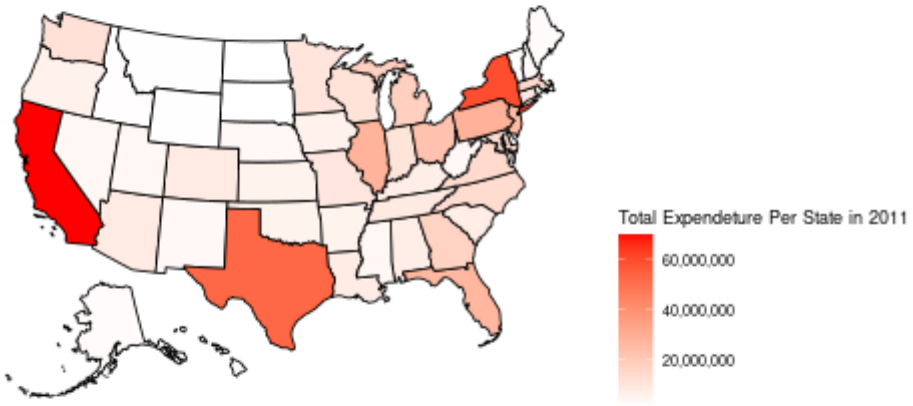

Figure 2: Total expenditure per state in the year 2011

It's no surprise that enrollment seems correlationed with total expenditure: The more students the more money needs to be put in.

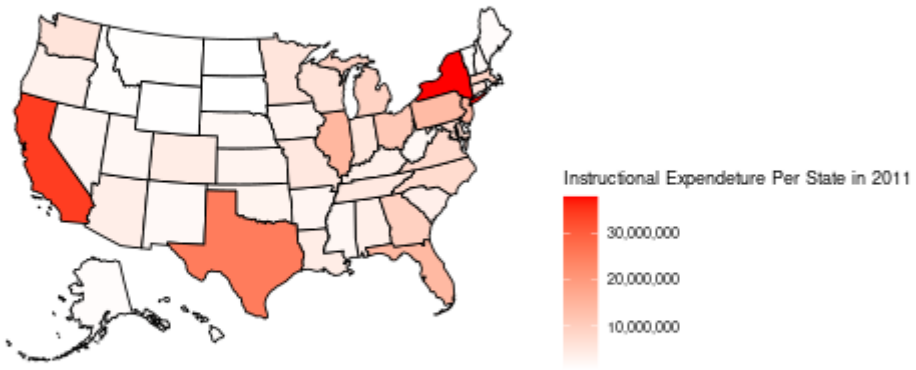

Figure 3: Total instructional expenditure per state in the year 2011

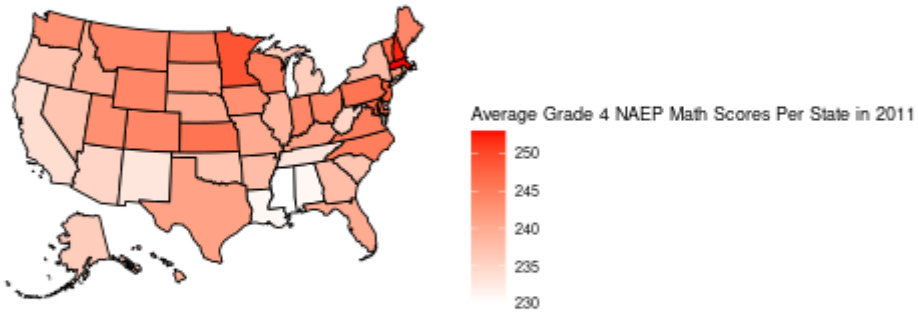

Figure 4: Average math score for grade 4 per state in the year 2011

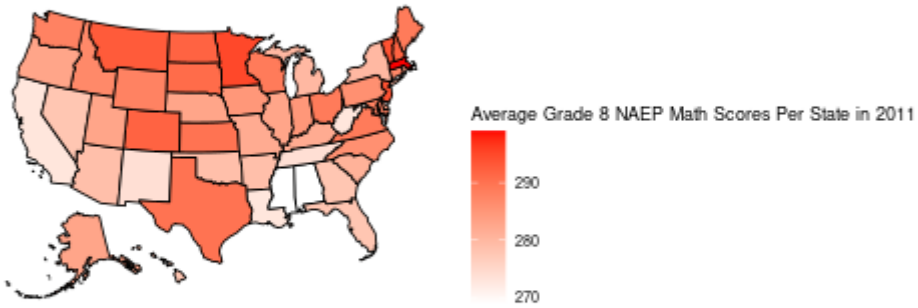

Figure 5: Average math score for grade 8 per state in the year 2011

Interestingly we see that average math scores for both grades 4 and 8 are not always correlated, it seems, with finances. For example in 2011 the mean grade 8 math score for California, the state with the highest total expenditure at $69,847,705, was 273, whereas for Massachusetts, the state with the highest average test score at 299, only spent $15,150,898. To understand I decided to look at how much money was being spent per student on instruction by taking a ratio of the columns:

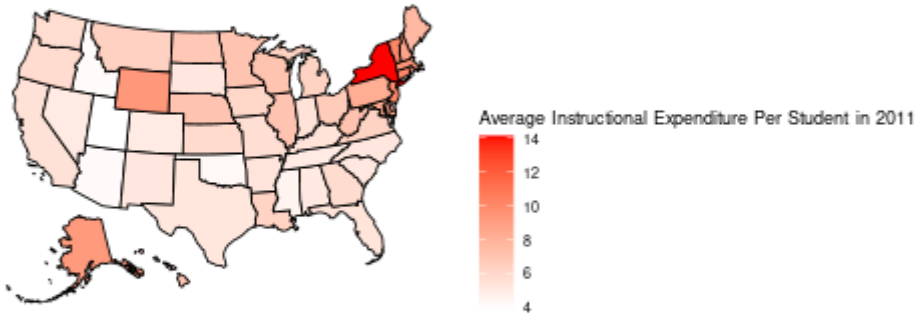

Figure 6: Average math score for grade 8 per state in the year 2011

There is some slightly better information. New York has the highest ratio at 14.13 dollars per student, for an average grade 8 score of 280. California only spent 5.5 dollars per student in 2011, which could be just a factor of having the most students. Mississippi and Alabama both tied for the lowest test scores for grade 8 in 269, and both had a low dollar per student amount at 4.5 and 5 respectively. Massachusetts spent 9.3 dollars per student in instruction, and had the highest test scores. What is not included in this data set is the average salary of teachers in each state, which I suspect could be more likely to cause higher test scores, it would make sense that a higher teacher salary attracts better candidates which in turn leads to higher test scores.

Analysis for New England

Turning to New England specifically was my next step. We first graph the instruction expenditure per year:

We can clearly see that Massachusetts and Connecticut far outpace the rest of the states.

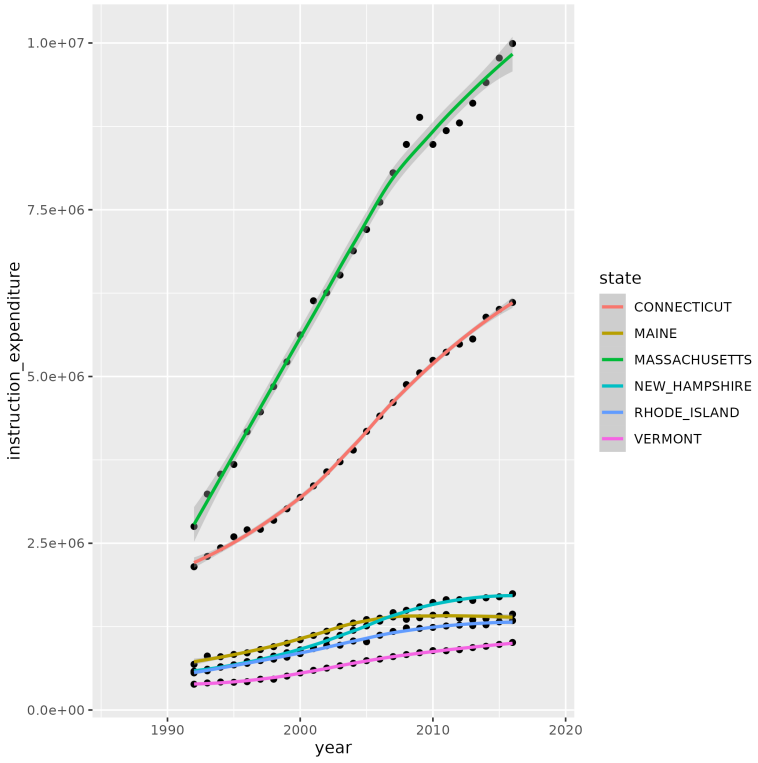

Figure 7: Instruction expenditure per year for New England

Looking at some more information we have a histogram of the average math scores for grade 4 students in New England:

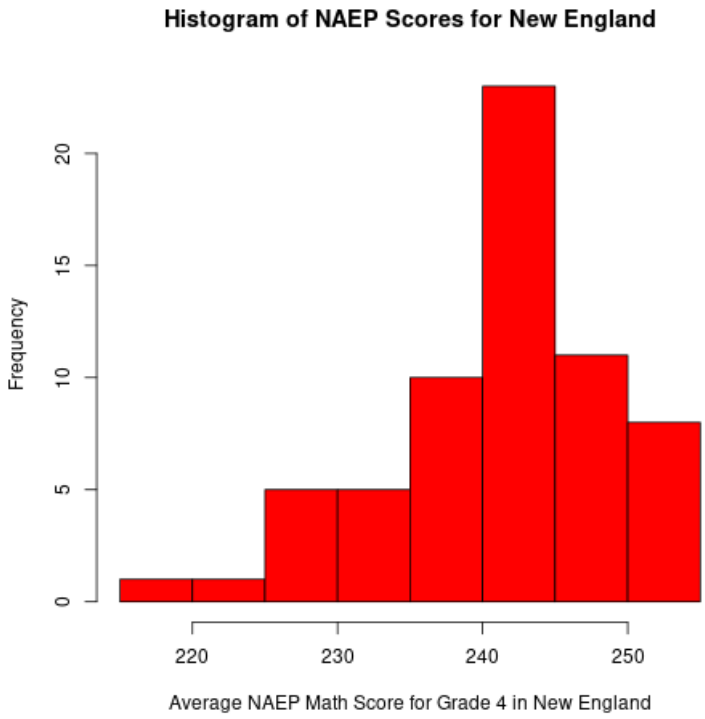

Figure 8: Histogram for New England scores in 2011

Now let's see how this compares with the rest of the US:

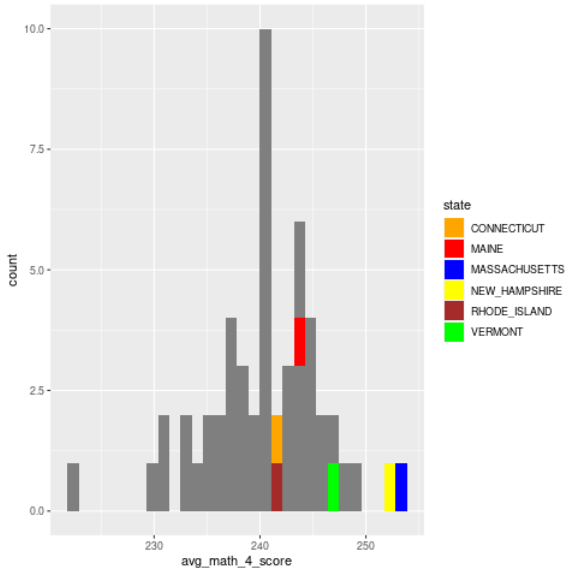

Figure 9: Histogram for US, with New England states colored in 2011

Here we get a good idea how much New England states out perform others, at least for 2011. We can perform a similar histogram with all years and each test to truly see some interesting data:

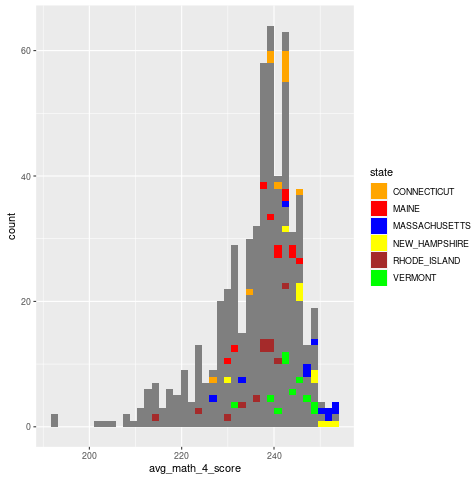

Figure 10: Histogram for US, with New England states colored for average grade 4 math score

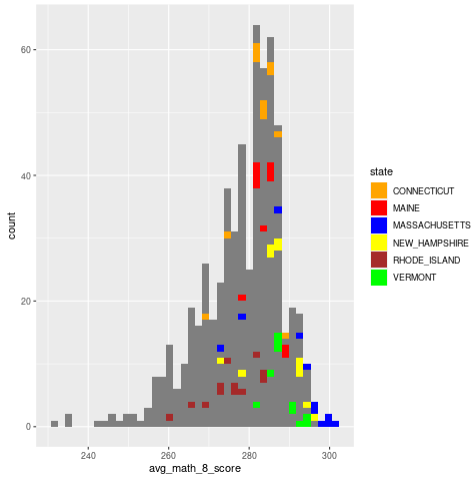

Figure 11: Histogram for US, with New England states colored for average grade 8 math score

While there is nothing definitive certainly we can see that apart from Rhode Island, New England students perform quite well on NAEP tests when they're offered.

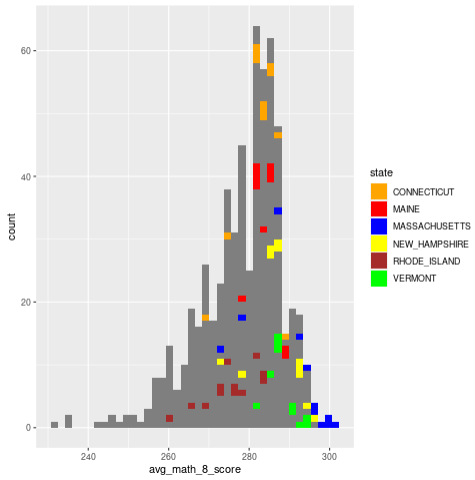

Figure 12: Histogram for US, with New England states colored for average grade 8 reading score

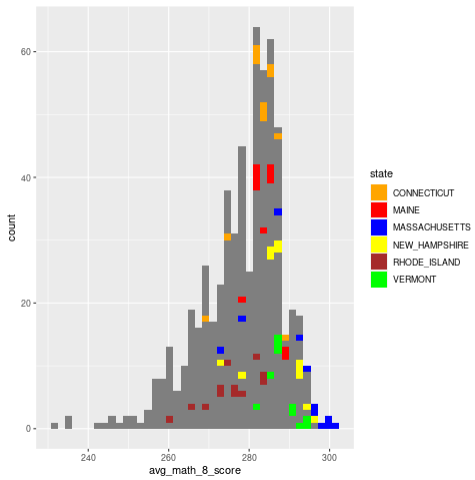

Figure 13: Histogram for US, with New England states colored for average grade 4 reading score

In an effort to see how much these states differ from each other, we perform a PCA. Given that these data are incomplete: various years are missing enrollment, test data, etc. we need to impute the data to ensure a valid PCA. My first attempt was to fill in all NA values with 0. The result is a quite bad association of states:

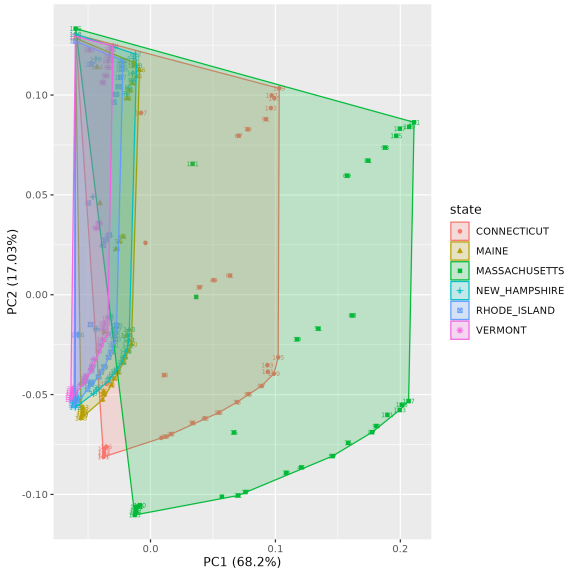

Figure 14: PC1 and PC2 for New England with 0's for NA values

However using the missMDA library, which imputes the missing values for us before we preform a PCA, we can see some nice association of states:

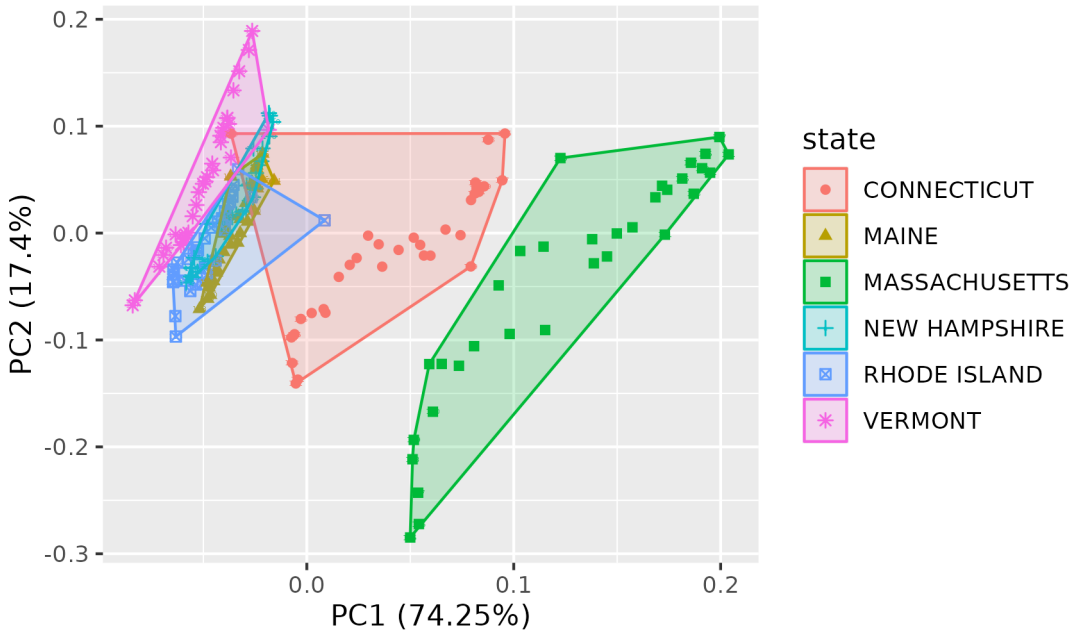

Figure 15: Imputed PC1 and PC2 for New England

With the imputed values we can see that not only do the proportional of variance increase, but we have more well defined clusters for Massachusetts and Connecticut, as well as Vermont. Moreover, the PCA tells us that the difference between Massachusetts is much more different than any other New England state.