Investment in Education in the United States of America

| Martim Norte  Instituto Superior Técnico  Lisbon, Portugal  martim.norte@tecnico.ulisboa.pt | Miguel Neves  Instituto Superior Técnico  Lisbon, Portugal  miguelscfneves@tecnico.ulisboa.pt |
| --- | --- |

# ABSTRACT

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

* ACM copyright: ACM holds the copyright on the work. This is the historical approach.
* License: The author(s) retain copyright, but ACM receives an exclusive publication license.
* Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Times New Roman 8-point font. Please do not change or modify the size of this text box.

Each submission will be assigned a DOI string to be included here.

Education is one of the most important aspects of a persons’ life, it has as strong role in shaping their future and their personality. The education of a population is essential for the development of nations and society. However, when governments try to improve their school system, investing their funds in the right places is key, even if resources may not be the most important aspect for granting students a better education.

In this project we aim to analyze the impacts of previous investments in education in the United States of America while we also try to find some patterns and curiosities.

This project was developed for the Information Visualization course from Instituto Superior Técnico, in Lisbon, Portugal during the school year 2021/22.

## Author Keywords

Education, Investment

## ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See<http://acm.org/about/class/1998> for the full list of ACM classifiers. This section is required.

# INTRODUCTION

The United States of America (USA) is the wealthiest nation of the world [Davies et al. 2021] but the results from the 2018 Programme for International Student Assessment (PISA) [2] show us the nation falling behind several less-wealthy countries.

Even though the USA are represented in the assessment as one nation, the states have different realities: they all have different levels of wealth and may invest their funds in different ways - namely in education - and different ruling parties. Therefore, we may find different levels of academic success among the states.

In this project we showcase an interactive visualization of data from the U.S. Census Bureau and the National center for Education Statistics that allows users to explore the relation between state-level investment in education and students’ academic success and enrolment rates and provides a comparison between states and how it all evolved through the years between 1992 and 2015.

We also present our own analysis, which focuses mainly on answering the following set of questions:

* How did changes in investment in education impact grades and student enrolment?
* Does state revenue impact academic success in a better way than state investment in education?
* Which type of investment (instruction, support services, capital outlay) has the greatest impact in enrolment and academic success?
* Is there any correlation between the party that governs a state and its investment in education?
* Is there a nationwide trend regarding student ethnicity and their academic success?

# Related work

When it comes to education, there are several articles addressing its importance and its impact in society and economy. [Woessmann 2006] found that the school’s resources are not as important as its environment for academic success. [Hanushek 1996] had already come to similar conclusions and suggested that a “serious restructuring of schools” in the USA was necessary to improve student performance.

# the data

The first dataset we found was the U.S. Education Dataset: Unification Project [5] which aggregated data from the U.S Census Bureau and the National Center for Education Statistics (NCES). This dataset contained information aggregated by state and year from 1992 to 2016 about revenue, expenditure on instruction, support services, capital outlay and NAEP grades for mathematics and reading for fourth and eighth grade students.

We also used data from the National Governors association that was aggregated by [Kaplan, J. 2021] which included information regarding the governing parties in each state throughout the history of the USA.

Finally, to obtain values per capita we used a dataset from the U.S. Census Bureau [7] that contains every state’s population in each decade.

## Data Processing

We merged all the data from the three datasets in one using Python and the Pandas library. We only kept the information we needed and had to handle some missing values. For the U.S Census dataset we had to fill the missing population values by interpolating between the existing ones, those were missing because the censuses only happen every 10 years. While going through the U.S. Governors dataset we noticed some values were missing, more specifically, for the years between the last election and 2020, so we filled those with the last known value for each state. Some regions had very little information so those were removed, fortunately every mainland state had the information we needed so it didn’t impact the visualization a lot.

The dataset ended up as table keyed by state and year.

From that dataset we later had to calculate for each value the percentage of change from the previous year, the values per capita for the attributes that involved money using the gross value and the population for the corresponding year and state. Finally, after having visualized some of the data, we recognized we also had to interpolate the missing values for the students’ grades because the original dataset only had data for every two years – the only years the NAEP exams were held.

# Visualization

## Layout

Our visualization consists of a dashboard with a bottom section with five different idioms with different purposes and a top section with a slider to select the year, two drop-down menus to select the type of analysis and the attribute to be visualized, and two lists, one for the attributes selected and another for the states.

The bottom section is divided in two subsections, the left one shows data aggregated by year and features a choropleth map and a dot plot whereas the right section is aggregated by state and features a line chart, a Gantt chart, and a parallel coordinates plot.

The selection and filtering of data is done both by selecting the desired values in the top section and by interacting with the idioms. The next paragraphs explain how that interaction works for each idiom.

### Choropleth Map

On the top left side of the bottom section, we have the choropleth map of the USA. In this map, each state’s shade of blue represents the state’s value for the attribute selected and for the year on the slider, according to the scale below it and when the user hovers the mouse over a state, a tooltip shows up with its name and the corresponding value. This map is also where the state selection occurs. By clicking on a state, the user can select and deselect it. The selected ones become highlighted with a colored outline. The effects of this selection on the visualization are addressed in the other idioms’ subsections.

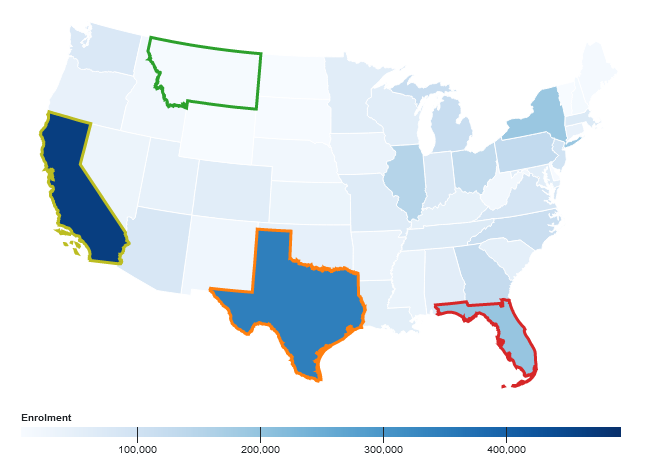


Figure 1. Choropleth Map representing gross values for Enrolment in 2007 with four states selected

### Dot Plot

The dot plot is on the bottom left of the bottom section and, like the choropleth map above it, displays data only from the year selected. But, in this case, it shows us the average results from students of different ethnicities, with its values represented by the dot’s position on the y-axis. Each line corresponds to a different state and each color a different ethnicity. By hovering the mouse over a dot, a tooltip shows up, displaying the name of the state and the item’s value. Whenever the user clicks on a dot, the corresponding state becomes selected or deselected for all the idioms (with a maximum limit of 5 states and minimum of 1). The line’s color also changes to match the state’s color in the other idioms.

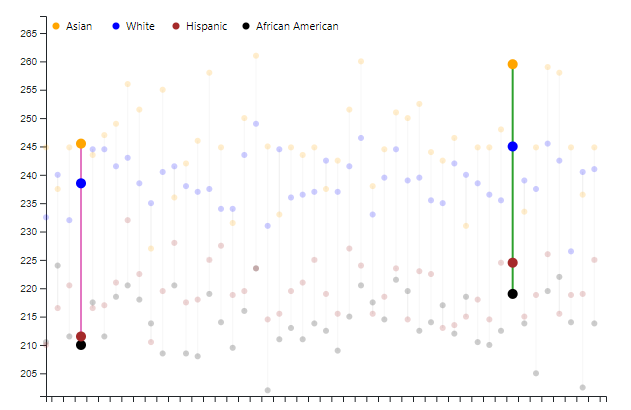


Figure 2. Dot Plot representing the average NAEP results for different ethnicities in the year 2015 with two states highlighted

### Line Chart

The line chart on the top right of the bottom section represents the evolution through time of the selected attributes for the selected states. If only one state is selected, the chart will display the lines for all the selected attributes with their color being defined by the attribute but if multiple states are selected the lines displayed will correspond to the first attribute on the list in the top section and their color will encode the state. This is easy to tell apart because the colors are consistent throughout the visualization i.e., the outer line on the choropleth map has the same color as the line in in the line chart for that same state. If the user wants to highlight one specific line, they just need to hover the mouse over the line for a couple seconds and the others will fade out, this also affects the choropleth map, in which the highlighted line will reflect the highlighted state by lowering the others’ opacity.

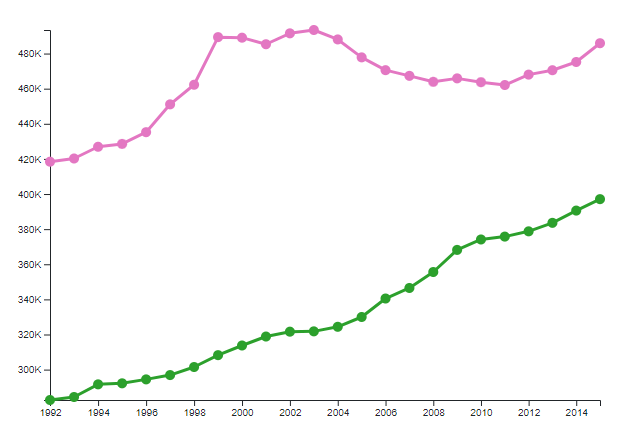


Figure 3. Line Chart showing the evolution of students enrolled from 1992 until 2016 in California (top, pink) and in Texas (bottom, green)

### Gantt Chart

[TODO]

### Parallel Coordinates

On the bottom right of the bottom section we find the parallel coordinates plot with all the attributes – a vertical axis for each of them. Like the idioms mentioned before, the lines in this plot are also color coded according to the state they represent, and each of them corresponds to a different year, with every year being represented. To highlight a specific state on this plot, the user can hover the mouse over any of the state’s representation in the other idioms for a couple seconds, the other lines will then fade out. To facilitate the interpretation of the lines, we also allow the reordering of the attributes by dragging and dropping their axis.

## Rationale

When first addressing the ways we could visualize the data, the easiest and most obvious idiom to pick was the choropleth map, it is able to represent data for every state in the simplest way, fits the problem well and would allow us to tell if there was any correlation between the values and the geographical position (North/South, East/West, interior/littoral, etc.).

When picking the rest, we took in consideration the questions we wanted to answer with them. We first considered using a scatter plot to figure out whether there were any correlations between attributes but because we wanted to compare multiple attributes at the same time, we opted by changing it to a parallel coordinates plot.

We also needed to find a way to visualize the student enrolment and performance data for different ethnicities at the same time. We wanted to have a nation-wide perspective of that and because the year didn’t feel very relevant for this issue, we found the dot plot to be a good way to do it. We can represent every state at the same time in the horizontal axis and the values on the vertical axis, each dot corresponds to a different ethnic group and that is encoded by the dots’ color.

Finally, we figured the line chart would be a good way to visualize the evolution of each attribute through time and would allow us to make some comparisons both between attributes and between states.

## Potential

When it comes to analyzing trends regarding student ethnicity, if we select ‘Gross’ from the analysis menu and any year from 2011 onwards, we can then look at the dot plot and tell if some dots’ colors show up in higher positions more frequently than others. And the answer for the 5th question we proposed comes up clearly, even after using the slider to change the year we are looking into we conclude that students from Asian and White background, on average, have higher grades than those from Hispanic and African American background. We can also tell that between White and Asian, it is usually Asian students who have the best scores and between Hispanic and African American, it is usually the latter with the lowest grades.

But if we change the type of analysis to ‘Percentage change’, the perspective we get is different and we can visualize whether some ethnicities seem to be improving more than others and the answer is not clear but there appears to be a bigger improvement in students with lower grades.

When exploring the data using the choropleth map and the slider, we noticed that the enrolment values per capita kept lowering through the years even though for most states the gross values increased. At first, we found that odd, but a possible justification would be that the adult population is increasing more rapidly than the younger one, maybe because people live longer or maybe due to immigration factors.

# IMPLEMENTATION DETAILS

[TODO]

# conclusion and future work

In this project we created an interactive dashboard to visualize and analyze data related to education in the USA with focus on investment and results.

[what we found]

In the early stages of development, we proposed to answer five different questions. While at first it seemed straightforward to find a way to answer them using visualization, drawing conclusions from what we were seeing in the plots we created proved to be tougher than expected.

If we were to start anew, we would have tried to focus on more specific aspects of the theme to provide a more in depth and more worthy analysis.

[future work]

# REFERENCES

1. Davies, J., R. Lluberas and A. Shorrocks, Credit

Suisse Global Wealth Databook 2021, Credit

Suisse Research Institute, Zurich.

[https://www.credit-suisse.com/media/](https://www.credit-suisse.com/media/assets/corporate/docs/about-us/research/publications/global-wealth-databook-2021.pdf )

[assets/corporate/docs/about-us/research/publications/global-wealth-databook-2021.pdf](https://www.credit-suisse.com/media/assets/corporate/docs/about-us/research/publications/global-wealth-databook-2021.pdf )

1. OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/5f07c754-en>.
2. Woessmann, L. (2006). *Why Students in Some Countries Do Better*. Education Next. Retrieved November 14, 2021, from <https://www.educationnext.org/whystudentsinsomecountriesdobetter/>
3. Hanushek, E. A. (1996). Measuring investment in education. *Journal of economic perspectives*, *10*(4), 9-30.
4. Garrard, R. (2018). *U.S. Education Datasets: Unification Project* [Dataset]. <https://www.kaggle.com/noriuk/us-education-datasets-unification-project>
5. Kaplan, Jacob. United States Governors 1775-2020. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-01-16. <https://doi.org/10.3886/E102000V3>
6. US Census Bureau. (2021, October 8). *Historical Population Change Data (1910–2020)* [Dataset]. <https://www.census.gov/data/tables/time-series/dec/popchange-data-text.html>