Remediation of Computer Science Education

in the State of Massachusetts

By

Kelvin Lin, Junior, kelvin@bu.edu

Changxuan Fan, Senior, fanchx@bu.edu

Sai Tejaswini Junnuri, Graduate, jteja@bu.edu

1. Introduction - Project Goals and Impacts

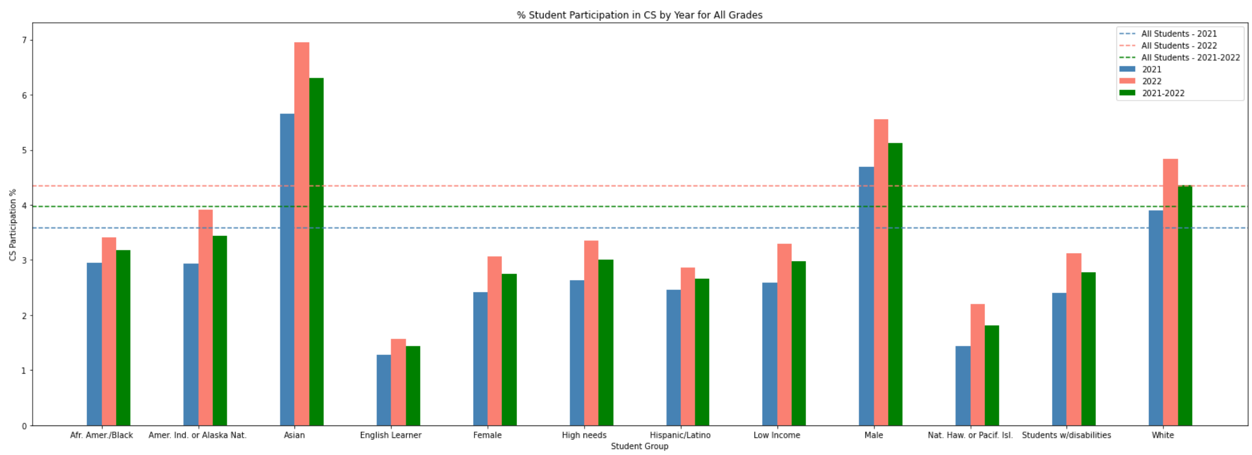
Studies have shown that there exists an educational divide across school districts in Massachusetts with a particular focus on the quantity of quality computer science instruction and AP test taking for many historically underrepresented peoples. In today's modern age, computer science education is becoming increasingly crucial for navigating competitive job markets, and adapting to rapidly changing technologies.

Our focus, in particular, is to develop strategies that can help undo the systemic biases in computer science education that have developed due to years of not considering the perspectives of marginalized communities. Through the analysis of data, the creation of models, and insightful visualizations, we seek to fill the gap between racial and social justice within the world of computer science. The ultimate goal of this semester-long partnership between our team, BU Spark! and the New Commonwealth Fund (NCF) is to develop an effective approach for the use of limited financial resources to implement an equitable ecosystem within technology.

2. Base Analysis - CS Participation Analysis

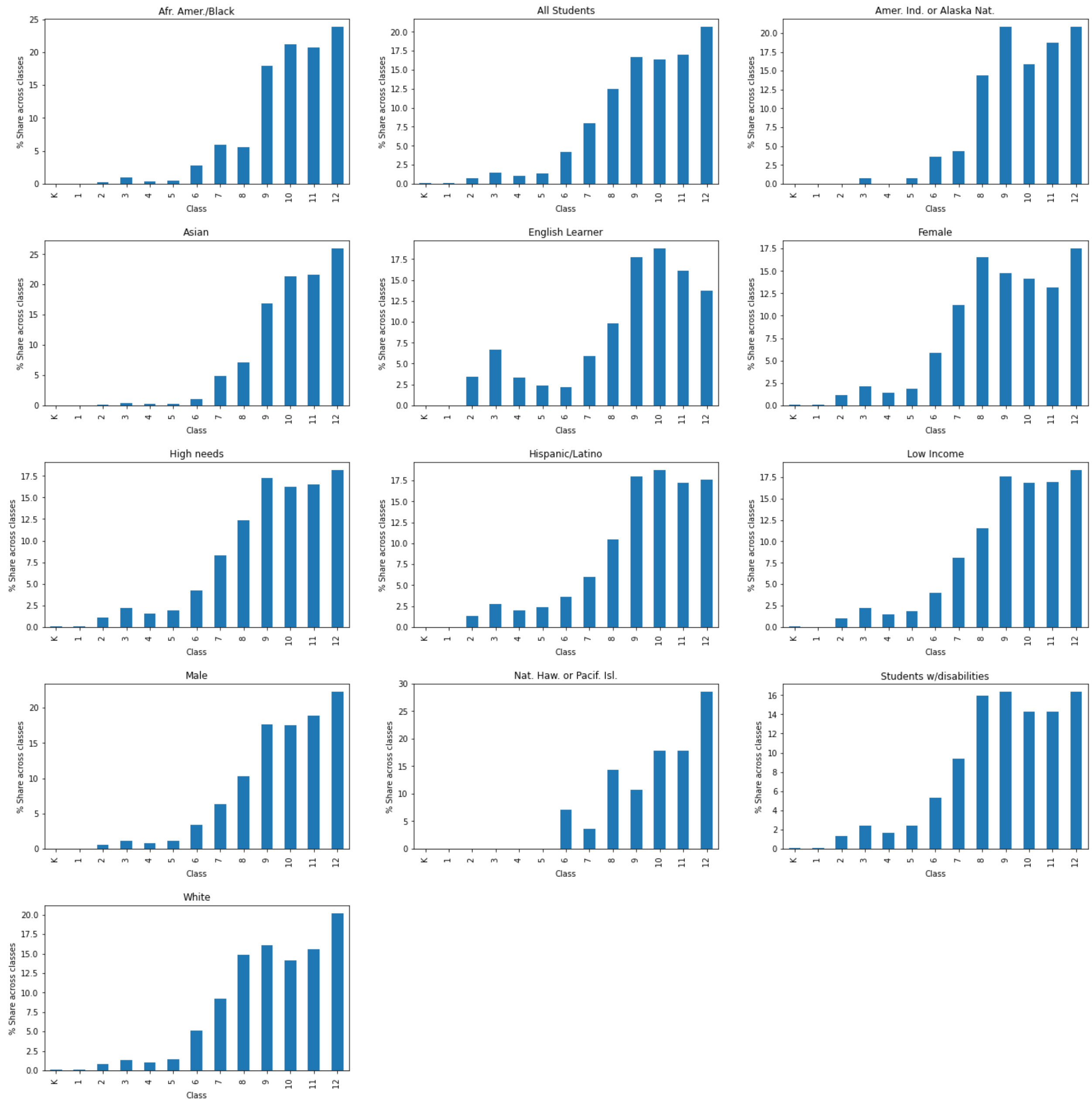
1 - By Student Group and Year:

a. Asian and Male student groups are above average
b. Increase in CS participation for all student groups
c. English Learners have the least participation followed by Nat.Haw or Pacf.Isl



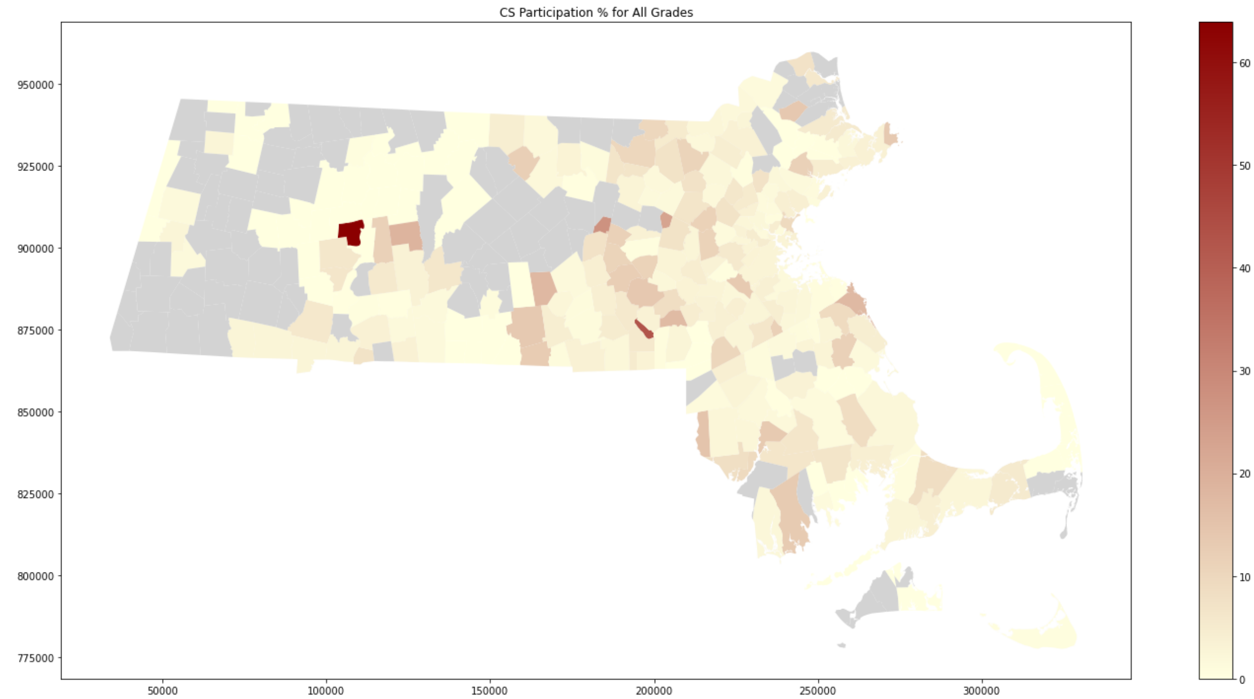
By School Year:

a. Asian and Male student groups, the share of classes keeps increasing with the class
b. English Learners have the opposite trend
c. Nat. Haw or Pacf.Isl have later participation than the rest



2 - By Region:

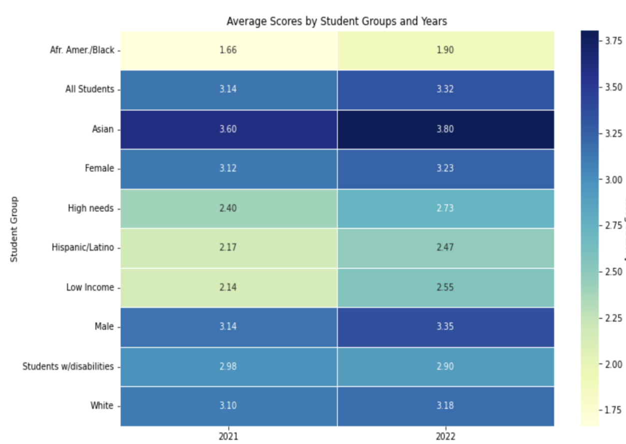
The Participation percentage is consistent in the Greater Boston area and except for a few districts in western Massachusetts, the participation is low.



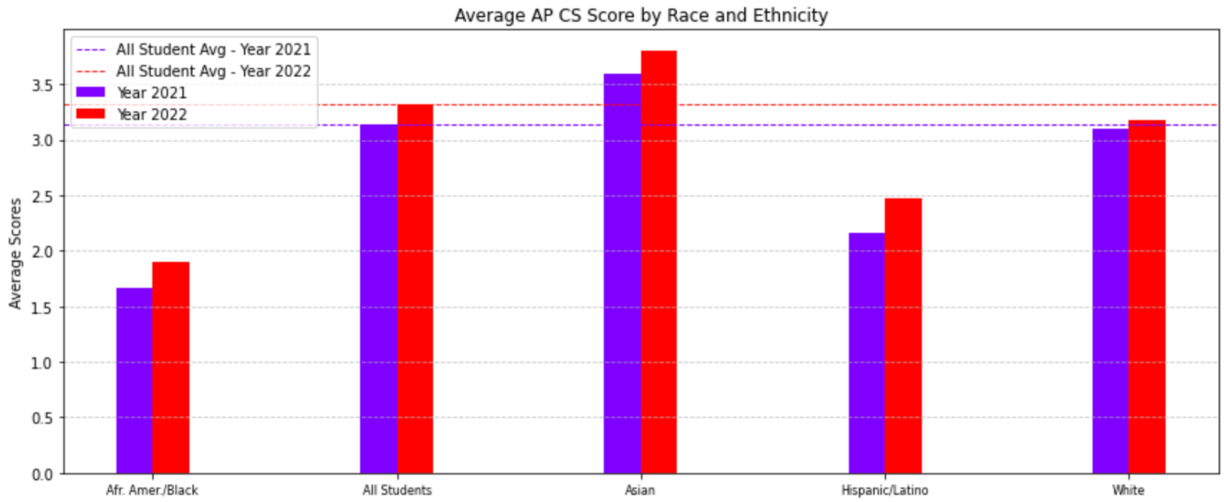
CS Performance by Student Groups

a. Overall View of Performance by All Student Groups

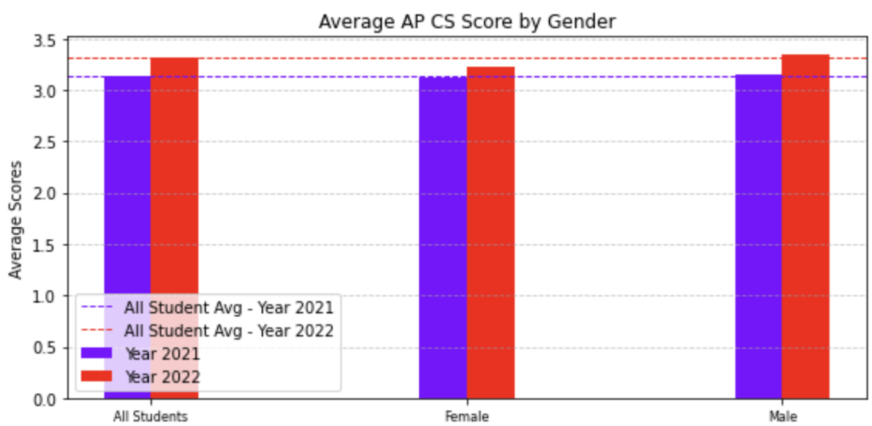
Here is a heat graph illustrating score differences among various student groups. The intensity of the color corresponds to the magnitude of the scores, with deeper colors indicating higher scores.



b. Performance by Race and Ethnicity



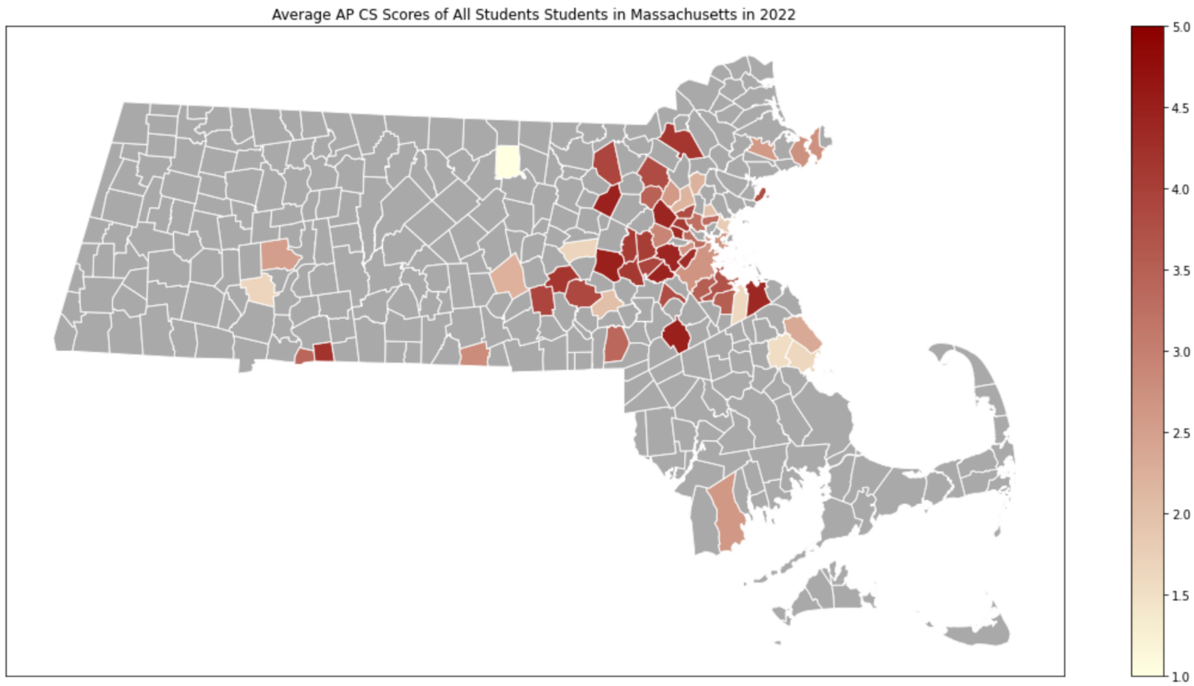
c. Performance by Gender



d. Analysis

The graph above depicts the average AP CS scores across various student groups. Observing the graph, it is evident that the overall average surpasses 3.0. Notably, Asians exhibit the highest scores, while African Americans record the lowest scores. Also, males and females have relatively the same performance. Additionally, the graph indicates an improvement in performance in 2022 compared to 2021.

CS Performance by Regions



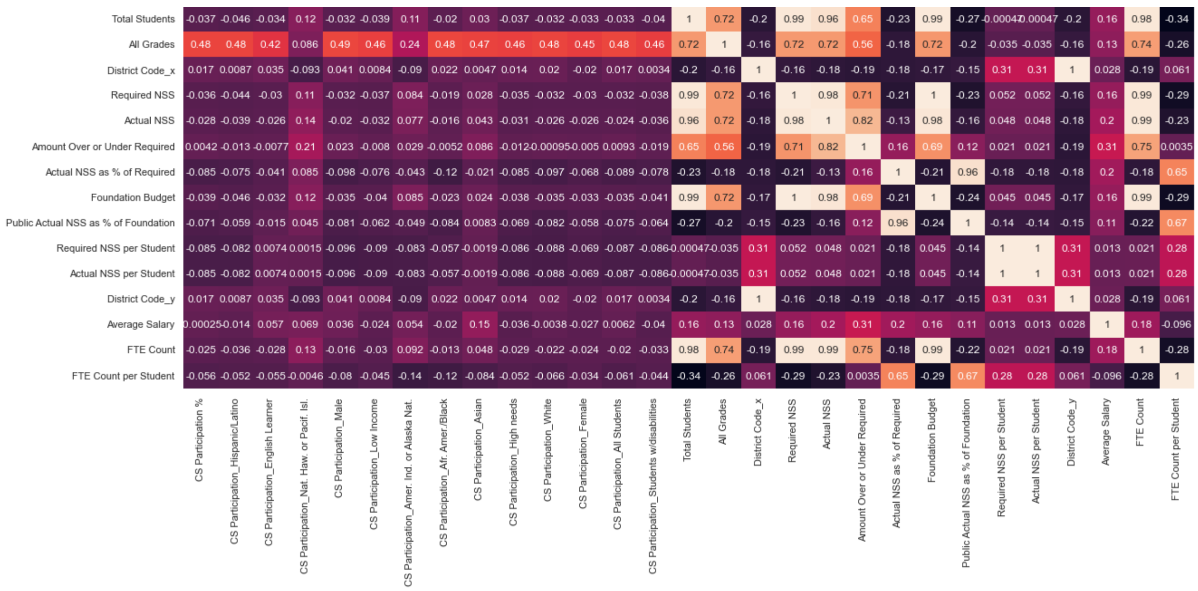
Analysis

When analyzing the average AP CS scores of all students in Massachusetts, we observe that the available data is primarily concentrated in Greater Boston. When examining the color scale, it becomes evident that students in these regions achieve notably high average CS scores, with most of them being represented by darker shades of red. This implies that a significant majority of students in these areas attain scores of at least 4.0.

Furthermore, even though there is missing data, we can still observe a clear trend: areas further from Boston exhibit relatively lower scores, often around 2.0 or 2.50. From this analysis, it can be inferred that regions outside of Boston, apart from the city itself, may not prioritize computer science education to the same extent.

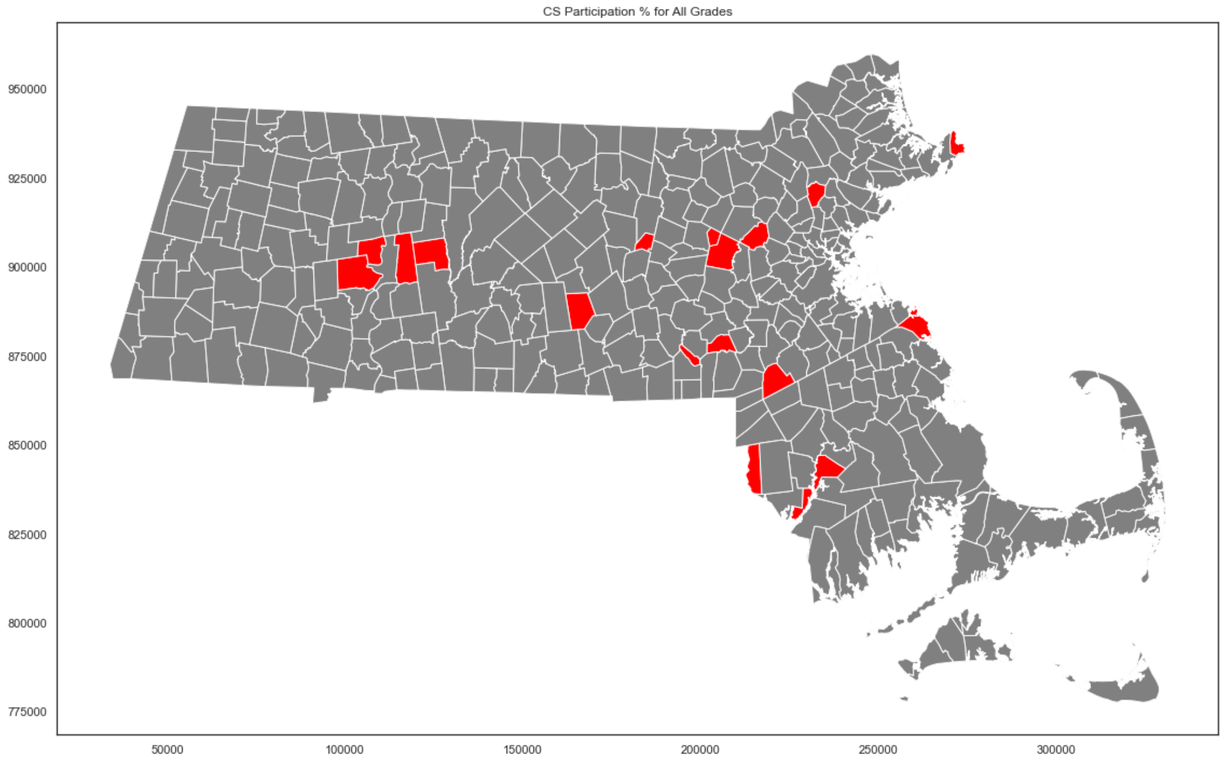
3. Extended Analysis - Specific School Districts in Need

Studies have also shown that although states like Massachusetts have generally a strong educational system there exists less gap in education than decades ago. Findings in the base analysis guided our focus onto the topic of providing support for individual school districts that have been lagging rather than funding generalized programs that benefit all school districts in Massachusetts. This sentiment is further supported by our recent analysis of disparities between groups and students as of recently, and the following is our correlation analysis in the following confusion matrix.



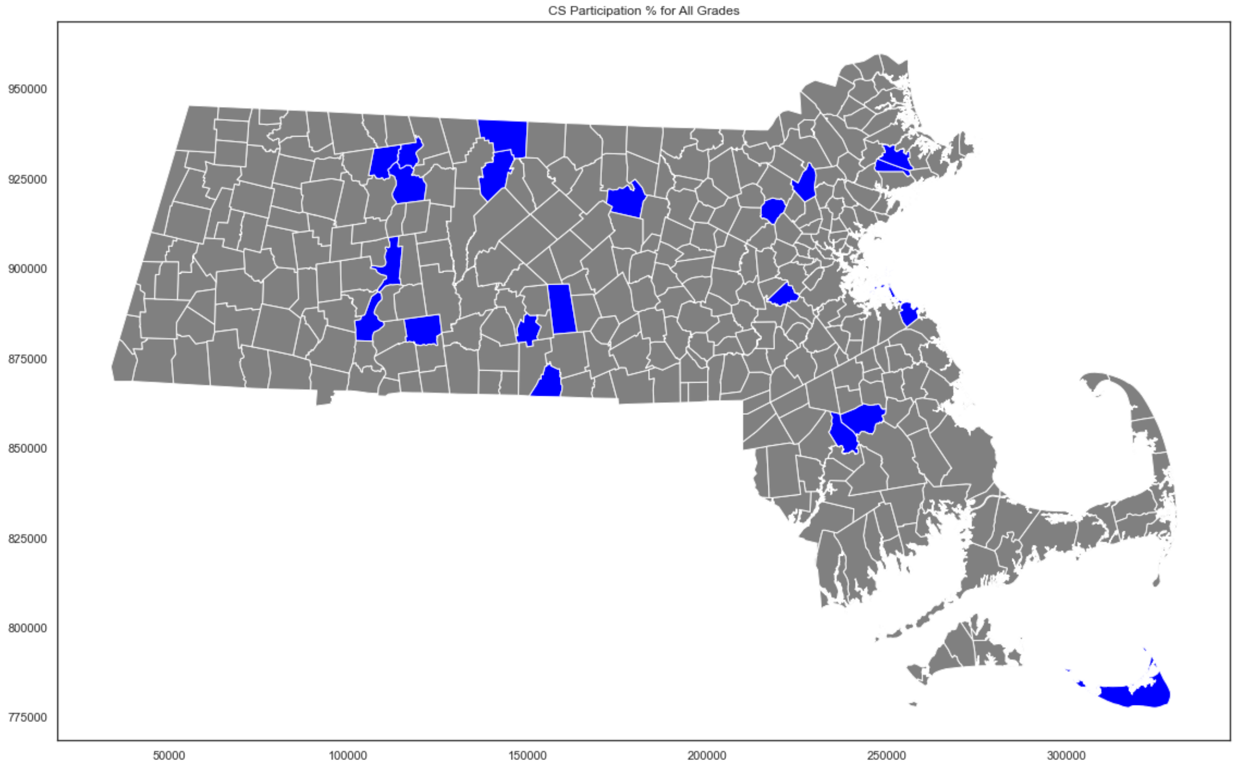
From the confusion matrix, we can note the All Grades row. This row indicates that the values are similar, which means that there is less of a discrepancy between groups, funding, and CS education. However, this does not mean that groups are on equal footing. There may be other factors that contribute to differences in CS outcomes.

What are the School Districts That Have Historically Performed Well?



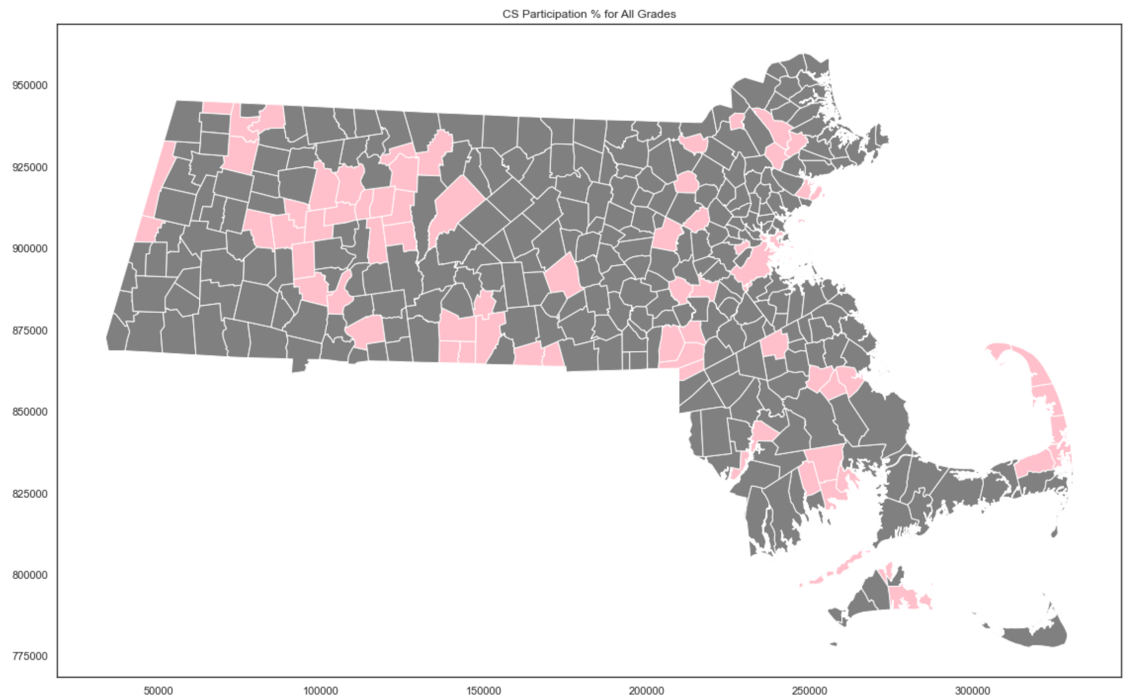
Visualizing the 20 best AP CS participating schools will serve as a useful comparison to the findings of our next questions. Analysis of past historical AP CS participation reveals that some of the best-performing districts in this aspect are the cities of Hatfield, Hopedale, Clinton, Northeast Metropolitan Regional Vocational Technical, Somerset Berkley Regional School District, Maynard, North Middlesex. By aggregating and averaging the data, it was found that the average school district size for these areas is ~15,000 people.

On the Contrary, What are the School Districts That Have Historically Performed Badly?



The graph above depicts some of the districts and areas that require a lot of aid and assistance. These districts include Spencer-E Brookfield, Greenfield, Greater New Bedford Regional Vocational Technical, Southbridge, Athol-Royalston, Gill-Montague, Gateway, and South Hadley. It is also interesting to note that these districts have an average school size of ~26,000 people. These are larger school districts than the districts with the best participation. It may be harder to implement specific policies with a larger student body.

Finally, What are the School Districts That Have Had Close To No CS Participation?



Finally, this is the graph that depicts school districts with very limited or no participation in CS education. These include the school districts of Acushnet, Alma del Mar Charter School (District), Baystate Academy Charter Public School, Benjamin Banneker Charter Public (District), Benjamin Franklin Classical Charter Public (District), Boston Green Academy Horace Mann Charter School (District), Boston Renaissance Charter Public (District). Note that most of these schools are charter schools which means that although they are government-funded, these schools operate independently of established state school systems. This means that many of these schools do not adhere to a standardized education model, which means that their focus may not be on computer science education. It is also of import to note that these are school districts with a very small student population with an average of ~4000 students per school district. This is in stark contrast to the districts above where there is much more of a student population.

4. Implementation of Strategy - Focusing on Specific School Districts

The following are the strategies that are recommended based on the analysis that we have performed. Keep in mind that these are targeting districts where there is poor CS performance or no CS participation at all.

- Creation of computer literacy classes in districts where there are CS education initiatives
- Creation of CS teaching positions at the high school level
- Specifically, provide more resources for people who belong to a group that has a high need for quality computer science education
- Providing financial support for students who cannot pay for CS IB/AP exams to encourage more students to take and qualify for computer science college credit
- Providing computers and necessary tools for coding to eliminate financial barriers to CS

The creation of CS classes in districts where there are few CS educational initiatives is an important step in introducing CS education into schools that have no previous exposure to a well-rounded education in computer science. These classes can range from general technological competency classes to college-level courses that can count for college credit. If students have previous exposure to computer science, even if there are few computer science classes in high school, they may be inspired to pursue computer science in higher education.

In a similar vein, the creation of CS teaching positions to teach the aforementioned classes is of utmost importance. It is important to hire and onboard qualified teaching staff that can inspire students to pursue CS as a possible career in the future. Therefore, funds must also be directed towards hiring qualified and passionate staff to inspire the students of the future.

Another step that should be taken is funding initiatives to provide resources to student groups that have demonstrated need. Due to historical biases and trends, they may be less likely to pursue computer science as a career. That means that it is important to provide incentives to these students, such as CS group-specific scholarships which can incentivize students to pursue CS in higher education.

Providing the necessary tools to be able to access this career is also of utmost importance. Although there are programs in place to provide students with laptops and other tools, it may be important to fund programs that provide these tools under the caveat that they are used for CS purposes while also functioning as tools for general schoolwork. This places a larger emphasis on CS rather than a generalized education which can have a positive impact on CS participation.

5. Limitations - Challenges Faced

The challenges that we faced during this project are as follows…

1. **Missing Data Points**
During analysis, we realized that there were a lot of missing data points within the Massachusetts Department of Education's datasets. While there were many categories, groups of people and subject topics that they covered, these datasets were often incomplete. This was especially prevalent in smaller schools. For an accurate and skew-free analysis, we adjusted the data to account for this issue, and mitigated this issue in accordance with best practices.
2. **Merging Datasets**
Another barrier to analysis and visualization of datasets was the problem of merging datasets. During the data preparation process, we encountered many issues with file names, and how to best organize the data. To mitigate the confusion and issues faced, we found it very helpful to adopt a standardized file naming convention. We also faced the issue of merging datasets from multiple years. To make this an easier and simpler process, we developed a web crawler to aggregate and combine data for easier analysis.

6. Conclusions Drawn - Key Points

There are many key points that one can draw from analysis and visualization of these datasets. It is important to distinguish between the CS educational outcomes between different groups. Most notably, we notice a stark contrast in results between different student groups. Analysis also sees relationships between regions and school years where there seems to be the most CS participation in the Boston area; this is not desirable as a state because this can lead to mass inequity just based on geographic factors.

Regional analysis was something that we wanted to emphasize as well. Particularly, the analysis and visualization of the specific school districts highlighted the disparities in CS educational outcomes between these areas. When we find the school districts that must be focused on, the best approach to implementing an effective solution are to have targeted and specific initiatives, rather than broad programs that do not strictly apply to CS. This would be the most effective allocation of the limited resources that Massachusetts schools have, and would lead to the most effective outcomes in terms of bringing computer science to all individuals.

7. Individual Contributions

Kelvin Lin: Created *BU Spark! Demo-Day* poster, visualizations, and analysis wrote the introduction and conclusions, challenges, and extension analysis.
Changxuan Fan: Team Lead. Responsible for group coordination, leading the editing of the report, and leading the direction of the project. Created visualization of heatmaps.

Sai Tejaswini Junnuri: Performed the CS participation analysis w.r.t to demographics, race, gender, and age. Extracted the district-level school spending information and analyzed it with CS Participation.