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

In Massachusetts, research highlights a significant educational gap where students in various districts face disparities in access to quality computer science (CS) education, including preparation for and completion of AP exams. This divide is pronounced among historically marginalized groups. Quality CS education is essential in the current era where technological proficiency is critical for employment and advancement in an increasingly digital world. Addressing this gap is not merely an educational imperative but also pivotal for ensuring equitable participation in a technology-centric job market, necessitating strategies to foster equal opportunities in CS learning.

The collaborative initiative spearheaded by BU Spark! and the New Commonwealth Fund (NCF) is centered on addressing the systemic barriers in computer science education. These barriers have largely emerged from historical oversight and the lack of inclusion of marginalized groups' perspectives. To reverse these entrenched biases, the partnership is utilizing a comprehensive data-driven approach.

By meticulously analyzing educational data, and constructing illuminating visualizations, the collaboration intends to diagnose the issues at hand and present transparent, informed insights into the racial and socio-economic inequities within computer science disciplines. The objective is to align this analytical approach with a commitment to social justice, thereby creating an effective bridge between data and diversity, equity, and inclusion efforts.

The endeavor reaches beyond mere identification of the educational divide; its scope includes crafting and executing actionable strategies that can result in sustainable change. By understanding the distinct needs and barriers faced by underrepresented groups in accessing computer science education, the partnership is keen on developing tailored solutions that address these unique challenges.

2. Data Collection and Cleaning

As part of data collection, we have chosen 8 reports from the Massachusetts Department of Elementary and Secondary Education (DESE) that provide a comprehensive framework for understanding various aspects of the state's educational landscape and are instrumental for analyzing trends and discrepancies:

1. **Enrollment by Grade**: This report provides insights into the distribution of the student population across different grade levels, which is essential for gauging the pipeline of students potentially entering computer science courses.
2. **Computer Science Courses**: This specific report sheds light on the availability and enrollment in computer science courses, offering a direct measure of student engagement and access to computer science education within districts.

3. **Per Pupil Expenditure**: By examining expenditure per student, one can assess the resource allocation across districts, which may influence the quality of education and the availability of courses such as computer science.
4. **AP Participation**: The Advanced Placement (AP) participation report indicates the level at which students engage with college-level curriculum, which can signal academic rigor and preparation across districts.
5. **AP Performance**: This report goes hand-in-hand with participation statistics to provide a fuller picture of not just who is taking AP courses, but also how well students are performing on these college-level assessments.
6. **SAT Performance**: The SAT report provides data on college readiness. It is a useful indicator of academic performance at the district level and can be correlated with access to advanced coursework, including computer science.
7. **Net School Spending Trend**: This report reveals trends in education funding over time, which can highlight whether increasing resources have been dedicated to expanding and enhancing computer science education.
8. **Teacher Salaries**: Teacher compensation can affect the ability to attract and retain qualified educators, which is crucial for maintaining high-quality instruction in all subjects, including computer science.

The amalgamation of these reports was chosen to offer a multi-faceted understanding of educational vigor, resources, and outcomes. Together, they allow for a detailed analysis of the equity and accessibility of computer science education in Massachusetts, helping identify any disparities that exist between different student groups or districts. These disparities become focal points for policy recommendations aimed at leveling the educational playing field.

Post Data Collection, as part of data cleaning we have

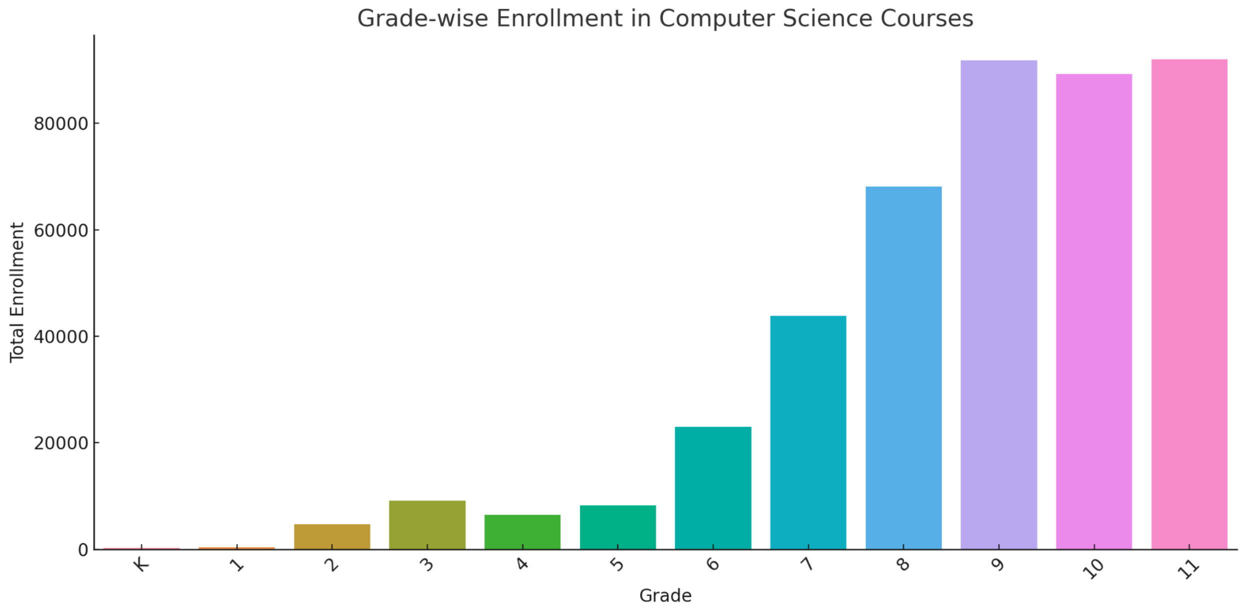
1. Conducted an exploratory review to grasp data structure and identify inconsistencies or missing information.
2. Merged the 2021 and 2022 datasets into one, standardizing categories and data types for uniformity, and facilitating future analyses.
3. Ensured consistency in categorical data such as race, ethnicity, and gender to eliminate variations in labeling.
4. Performed accuracy checks by comparing processed data against sources to validate the data-cleaning process.

3. Exploratory Data Analysis

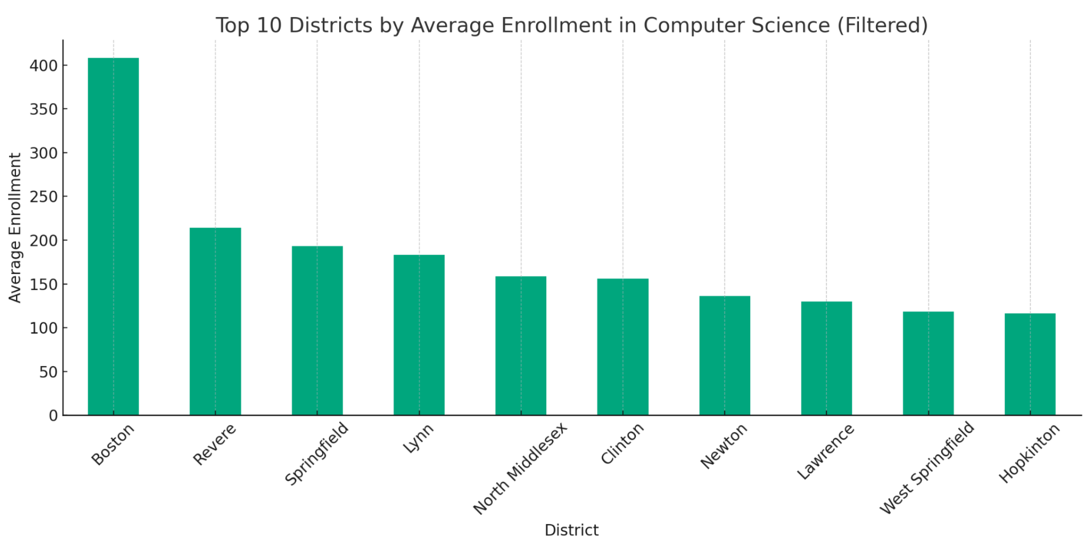
a. CS Course Participation:

Grade-Wise Enrollment:

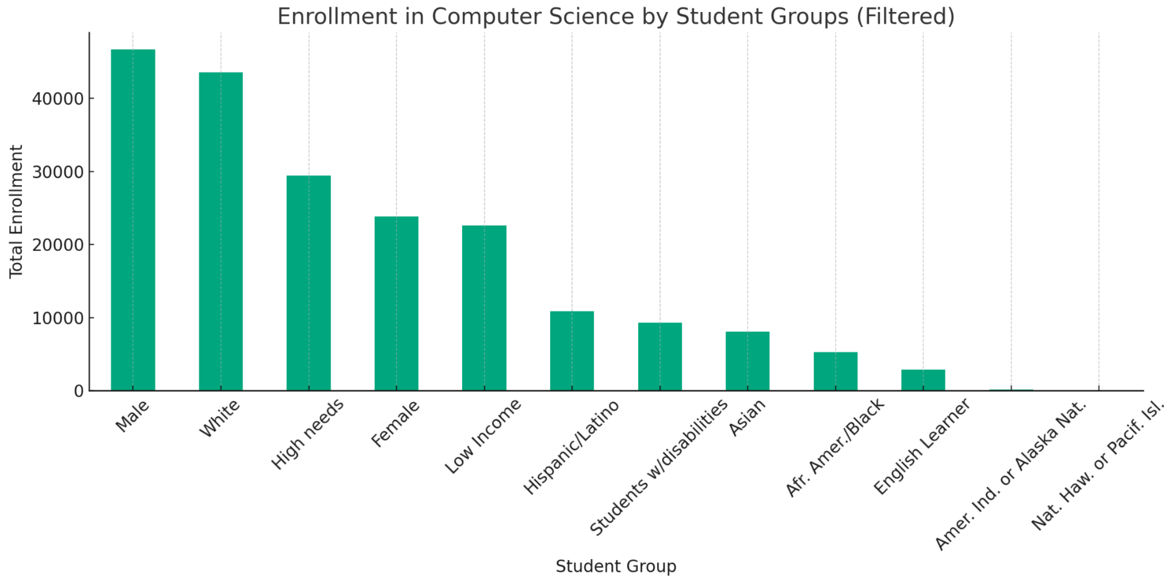
- Enrollment in computer science courses increases significantly in higher grades.
- The lower grades (K to 5) have relatively fewer enrollments in computer science courses.
- The peak enrollment is observed in grade 12, followed by grades 11, 10, and 9, suggesting a growing interest or availability of computer science education as students progress to higher grades.



Top 10 Districts with CS Participation:



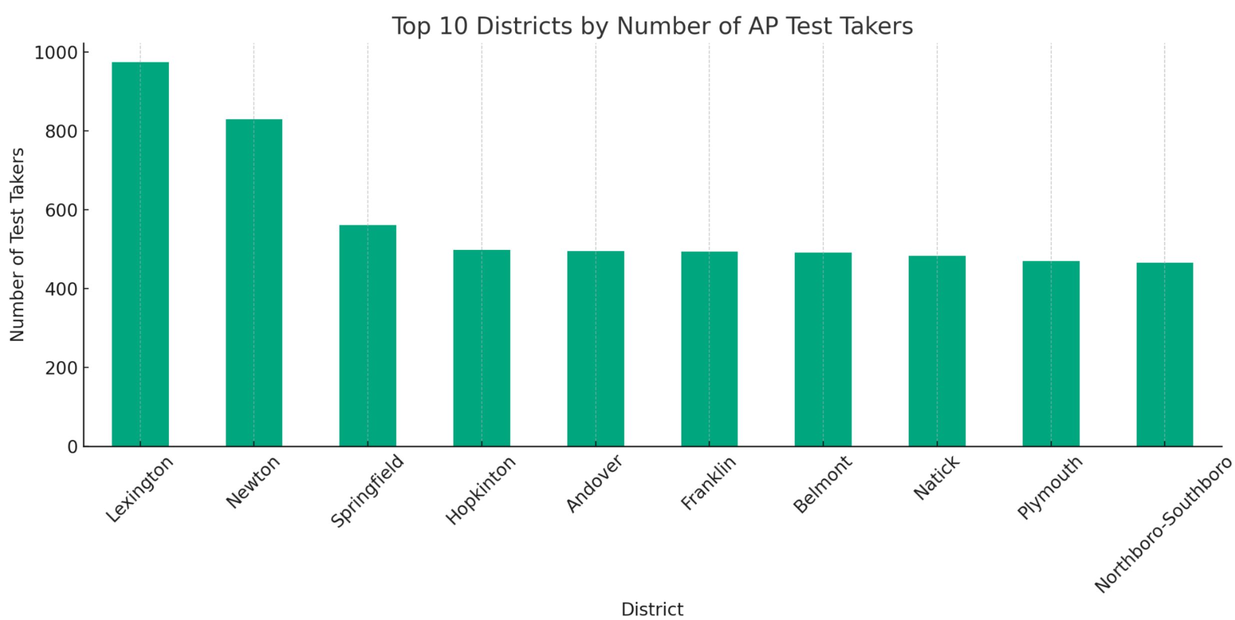
<u>Student Groups by CS Participation:</u>



b. AP Test Taking:

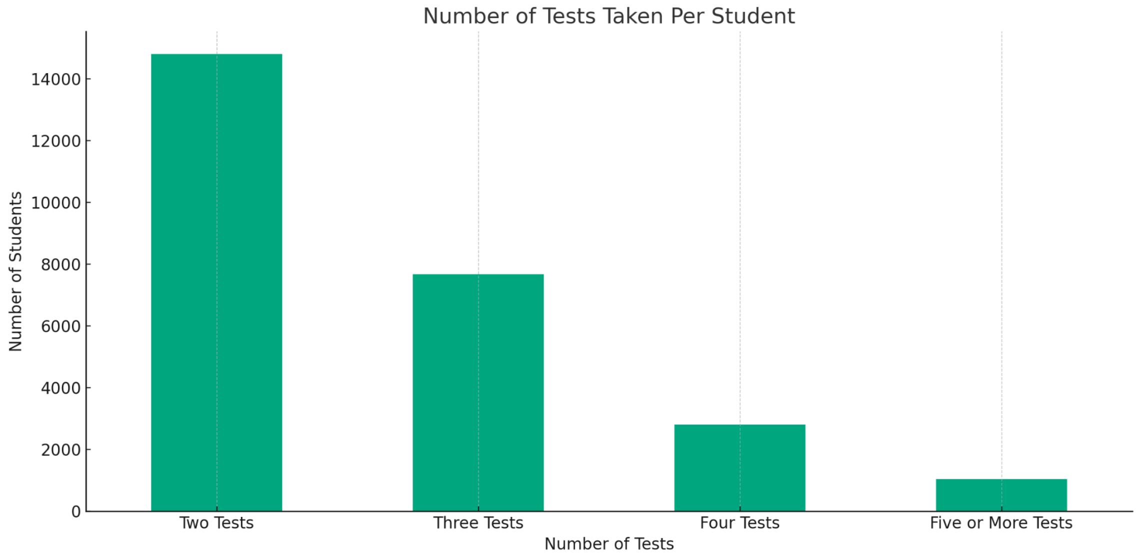
<u>Top 10 Districts by AP Test Taking:</u>

- The top districts by the number of AP test takers show varying levels of participation.
- Some districts have a notably high number of AP test takers, indicating a strong emphasis on AP courses and college-level curriculum in these areas.



Number of Tests per Student:

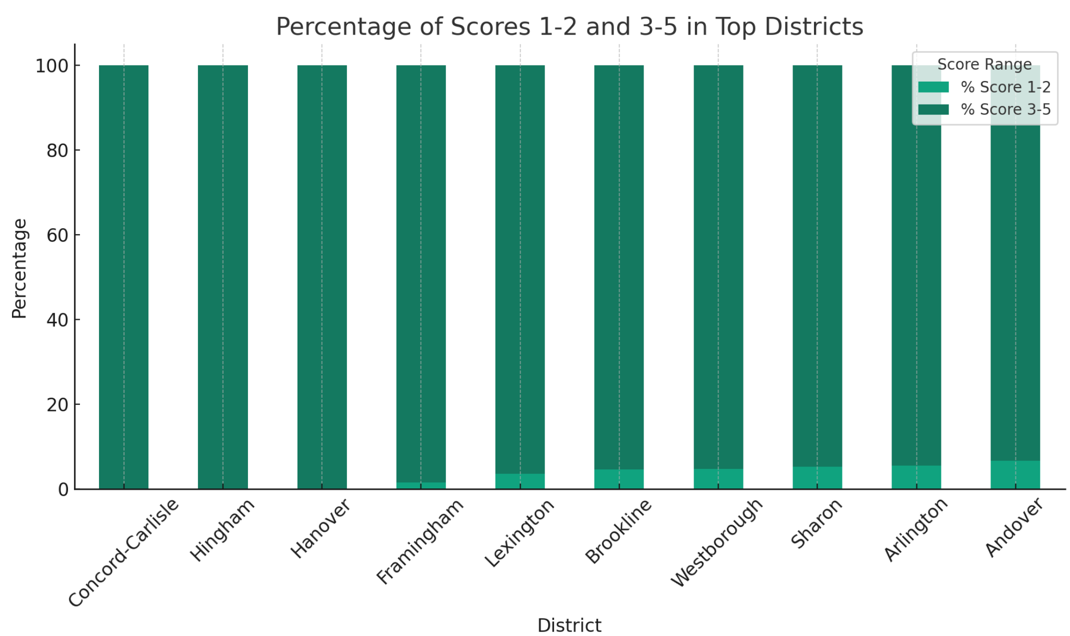
- A significant number of students take only one AP test, but there is also a considerable number of students taking two or more tests.
- The number of students taking three, four, or five or more tests is smaller but still notable, indicating a subset of students who are highly engaged with the AP curriculum.



c. AP Performance:

Top Performing Districts:

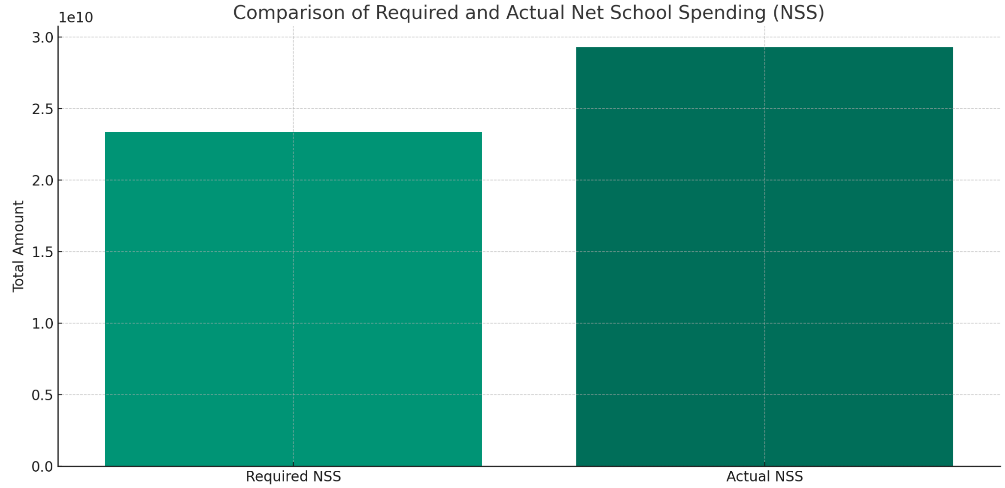
- The top districts by percentage of higher AP scores (3-5) demonstrate strong performance, indicating a high level of academic achievement in these areas.
- The percentage of lower scores (1-2) is significantly lower in these top-performing districts.



d. Net School Spending:

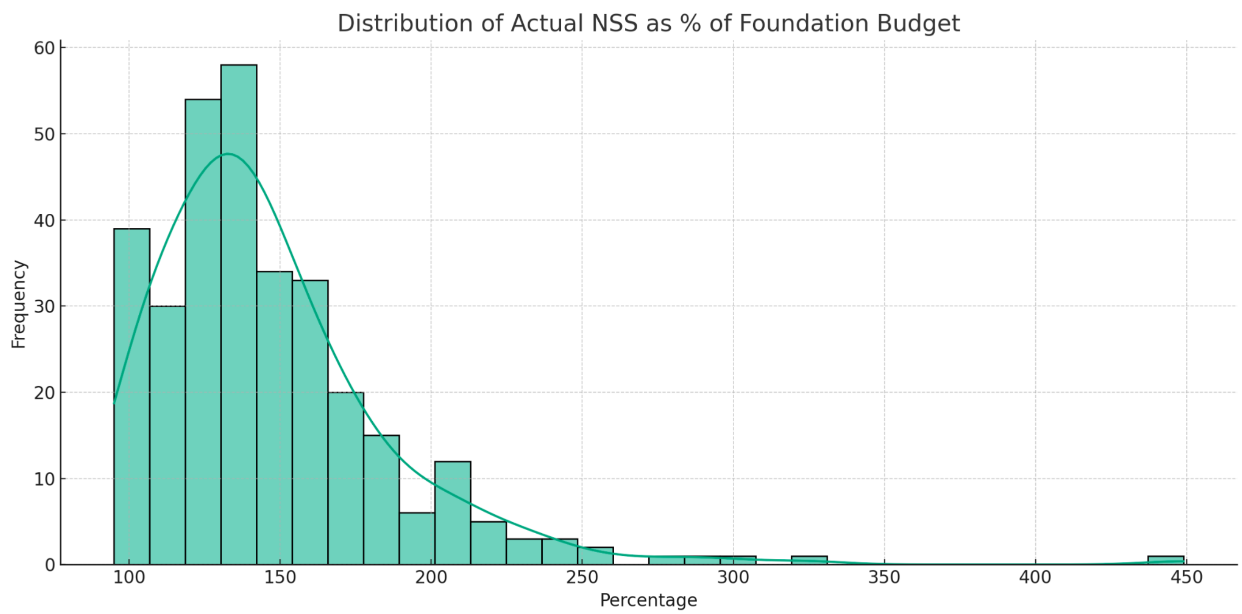
Required NSS vs Actual NSS:

- The total actual NSS across all districts is higher than the total required NSS, suggesting that districts, on average, allocate more resources to schooling than the minimum requirement.
- This trend indicates a commitment to investing in education beyond mandated levels.



Actual NSS as % of Foundation Budget:

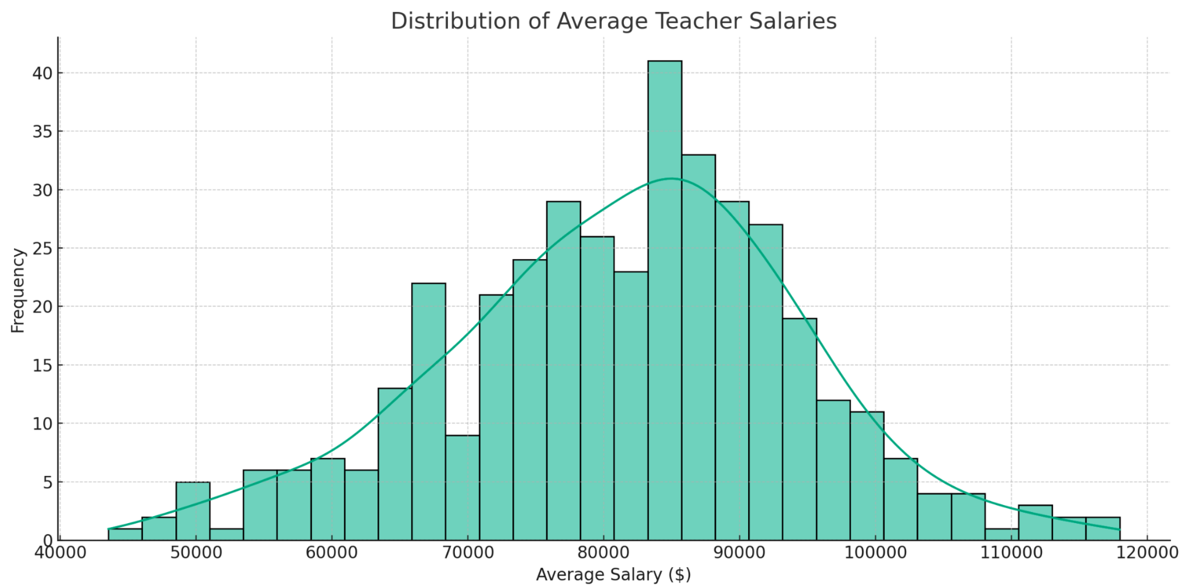
- The distribution of actual NSS as a percentage of the foundation budget varies widely among districts.
- A significant number of districts spend a higher percentage of their foundation budget on education, which might reflect efforts to enhance the quality of education, including computer science programs.



e. Teacher Salaries:

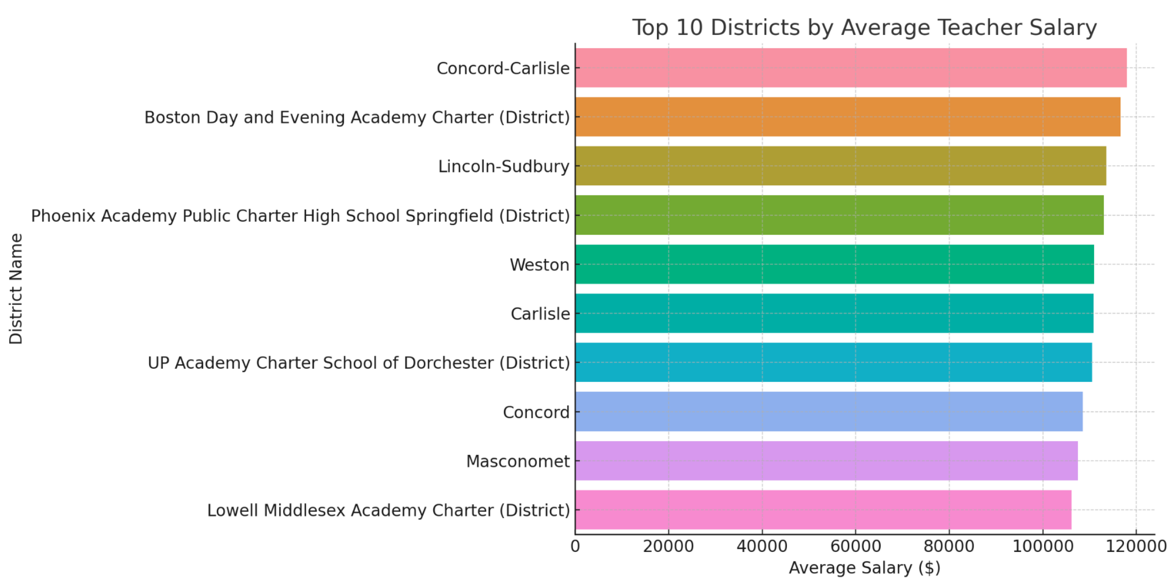
Average Teacher salaries:

- The histogram shows a wide distribution of average salaries, with most salaries clustering around the median value.
- This suggests variability in teacher compensation across districts, which could impact the ability to attract and retain qualified educators.



Top Districts by Average Teacher Salary:

- The bar plot of the top 10 districts by average teacher salary highlights the districts that offer the highest average compensation.
- These districts could potentially be more successful in attracting and retaining qualified educators, including those in computer science.

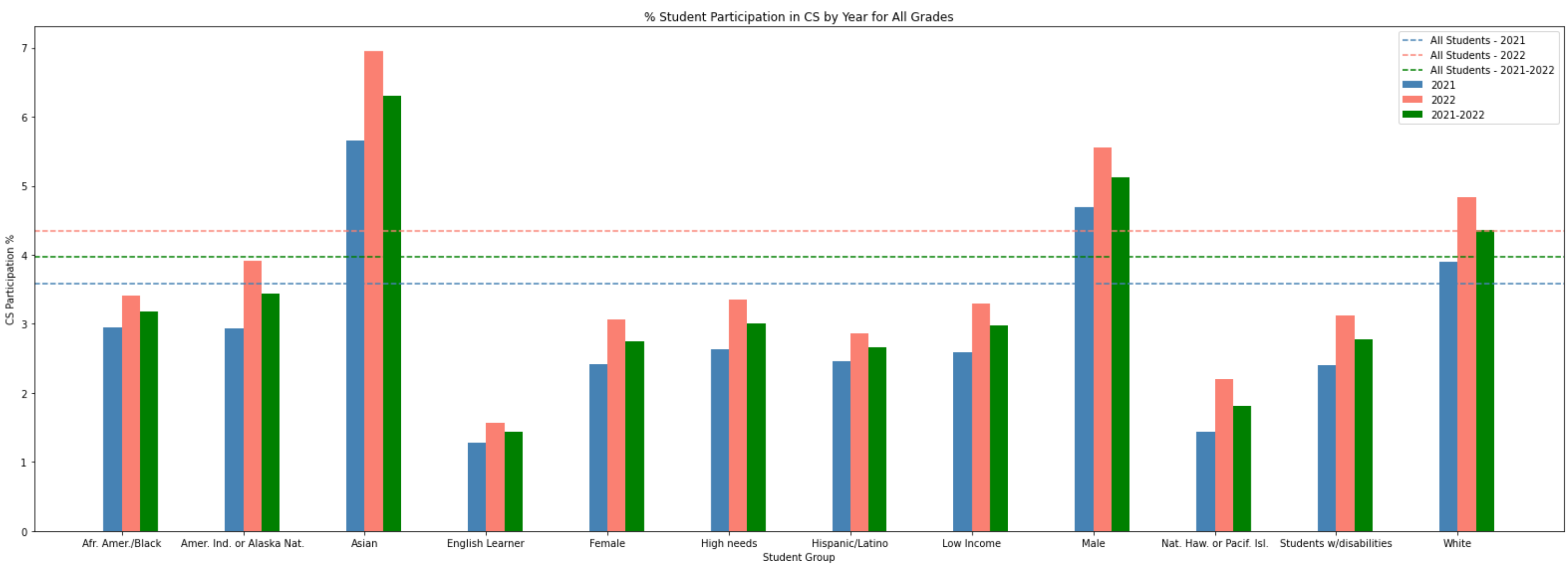


4. Visualizations for Key Questions

CS Enrollment Analysis

1. By Student Group and Year

- **Asian and Male students** show consistently higher engagement in CS courses, exceeding the average participation rates for all students.
- **English Learners** exhibit the lowest participation rates, suggesting significant barriers to access or interest in CS education.
- **Native Hawaiian or Pacific Islander** students demonstrate delayed or lower engagement compared to other groups, indicating a potential need for early intervention.
- Both **High Needs and Low Income** student groups show an upward trend in CS participation, signaling progress that may result from targeted educational initiatives.
- A gender gap is evident, with **Male students participating more than Female students**, highlighting the need for strategies to foster female involvement in CS.

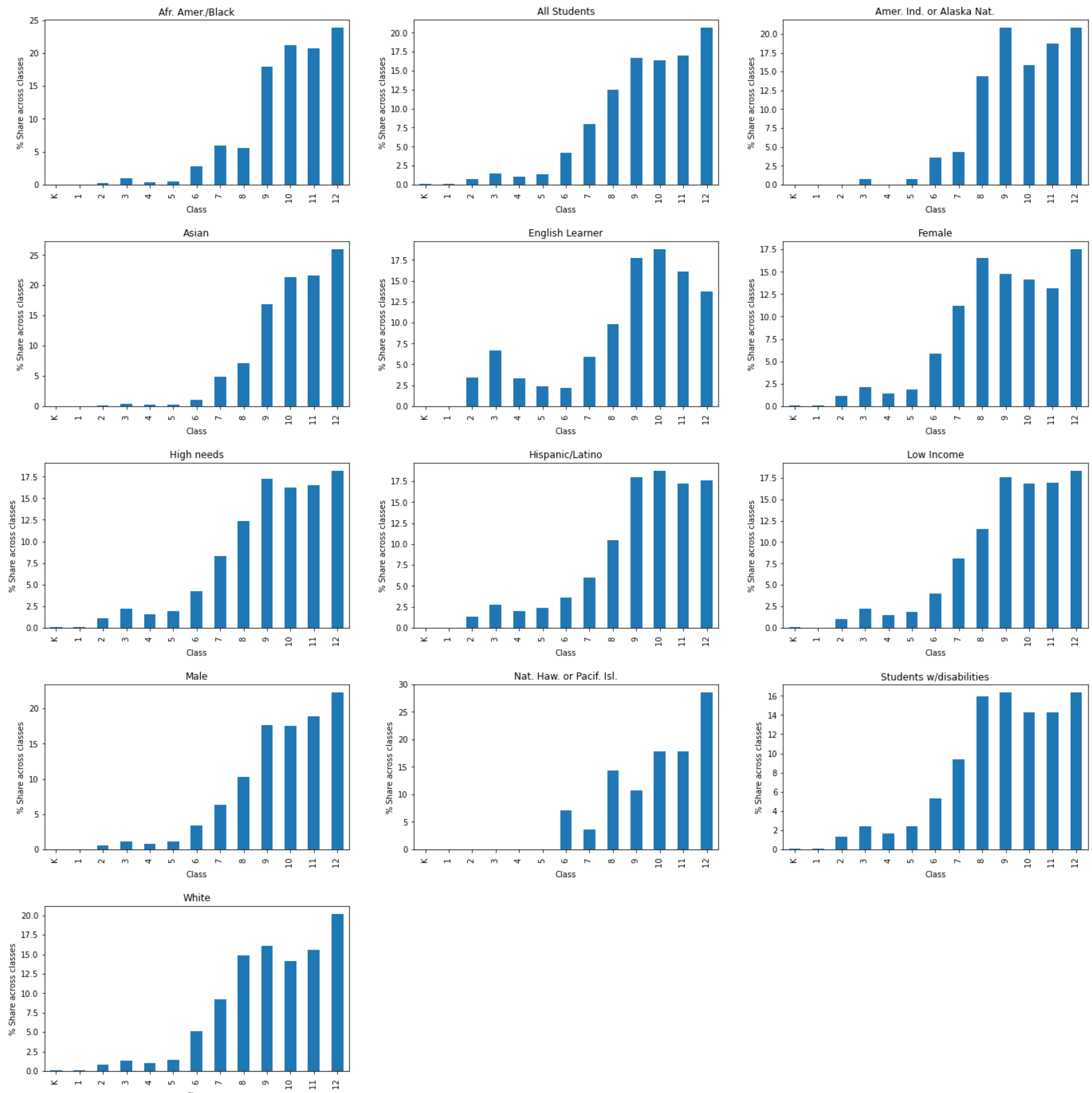


2. By Grade

- Participation peaks at grade 11 for most student groups, with a general trend of increasing engagement from earlier grades.
- African American/Black and Asian students demonstrate significant participation in upper grades, especially grade 11.
- English Learners and Students with Disabilities show consistently lower participation across all grades.
- Participation for both Male and Female students rises until grade 10, then notably dips in grade 12.

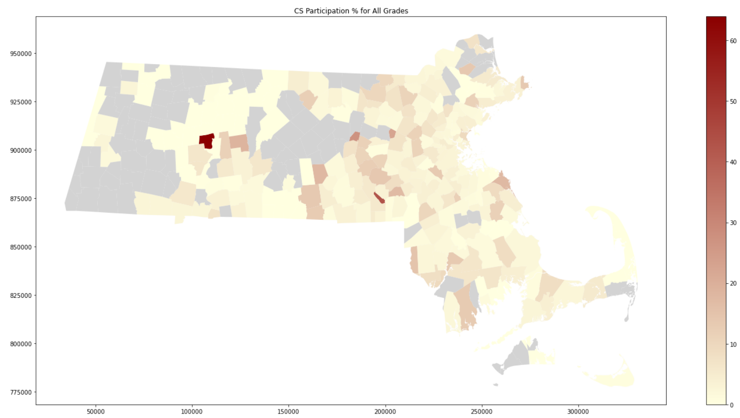
- Hispanic/Latino, High Needs, and Low Income students exhibit a steady rise in participation, reaching the highest levels in grade 11.
- Native Hawaiian or Pacific Islander students show markedly increased engagement in later grades.
- A decline in participation across most groups in grade 12 suggests potential disengagement in the final high school year.

Strategic interventions might be most effective if implemented during the transition to high school and targeted at maintaining interest through grade 12, with a particular focus on supporting English Learners and Students with Disabilities.



3. By Region

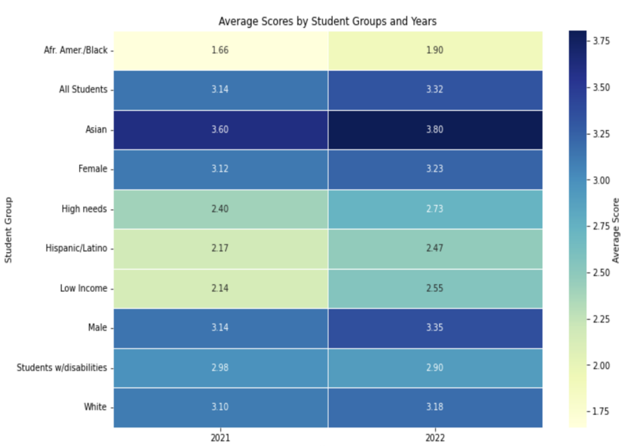
The participation rate in computer science courses exhibits remarkable consistency across the Greater Boston area. However, apart from a handful of districts in Western Massachusetts, there is a noticeable dip in engagement, indicating a regional disparity in participation.

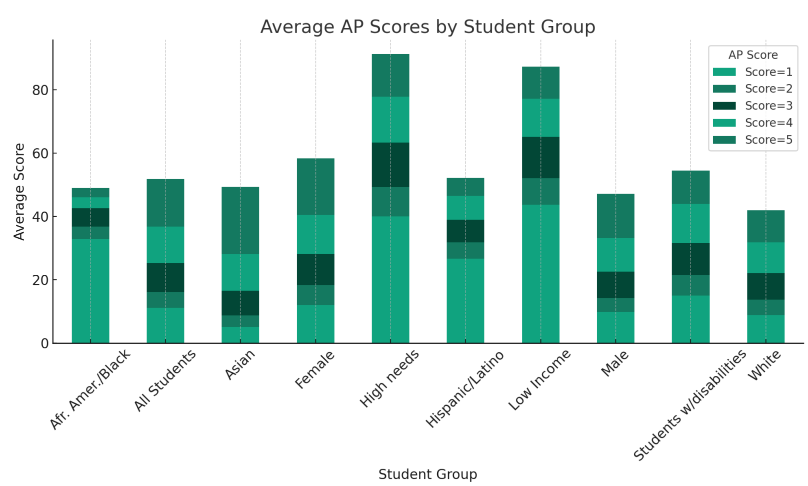


CS Performance Analysis

4. By Student Group

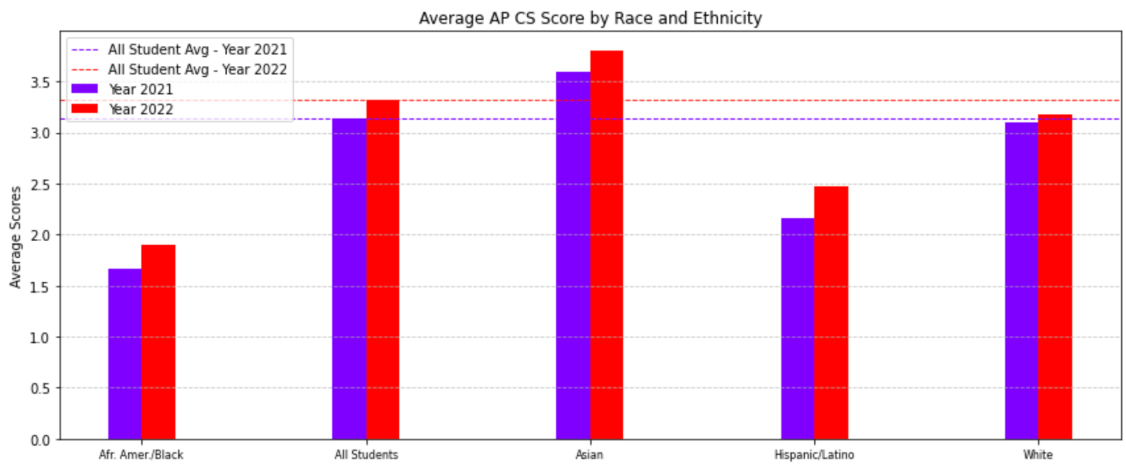
- *<u>All Students Overall:</u>* Reflect a slight increase in the average score from 2021 to 2022, pointing to general improvement in performance.
- *<u>Socioeconomic Factors:</u>* Students identified as low-income or high-needs demonstrate lower average scores compared to the overall population, but like other groups, they show improvement in the following year.
- *<u>Year-Over-Year Trends:</u>* Most student groups display an upward trend in average scores from 2021 to 2022, signaling potential improvements in teaching quality, curriculum, or student engagement strategies.





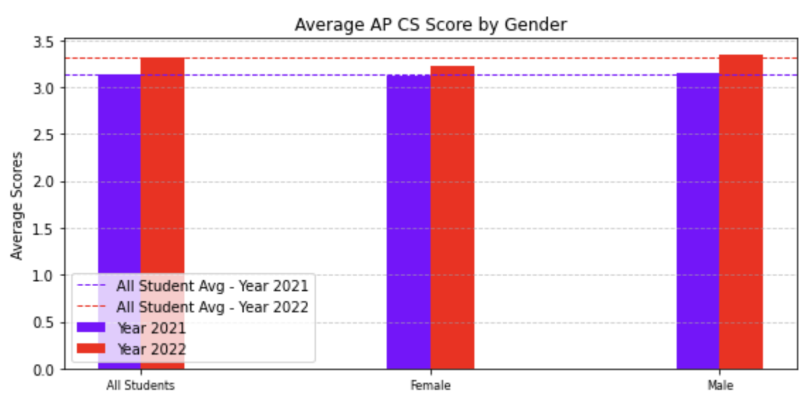
5. By Race and Ethnicity

- *Asian Students:* Maintain the highest average scores across both years, showcasing consistent academic achievement in AP Computer Science.
- *African American Students:* Show an increase in average scores from 2021 to 2022, indicating positive progress, yet remain below other groups for both years.



6. By Gender

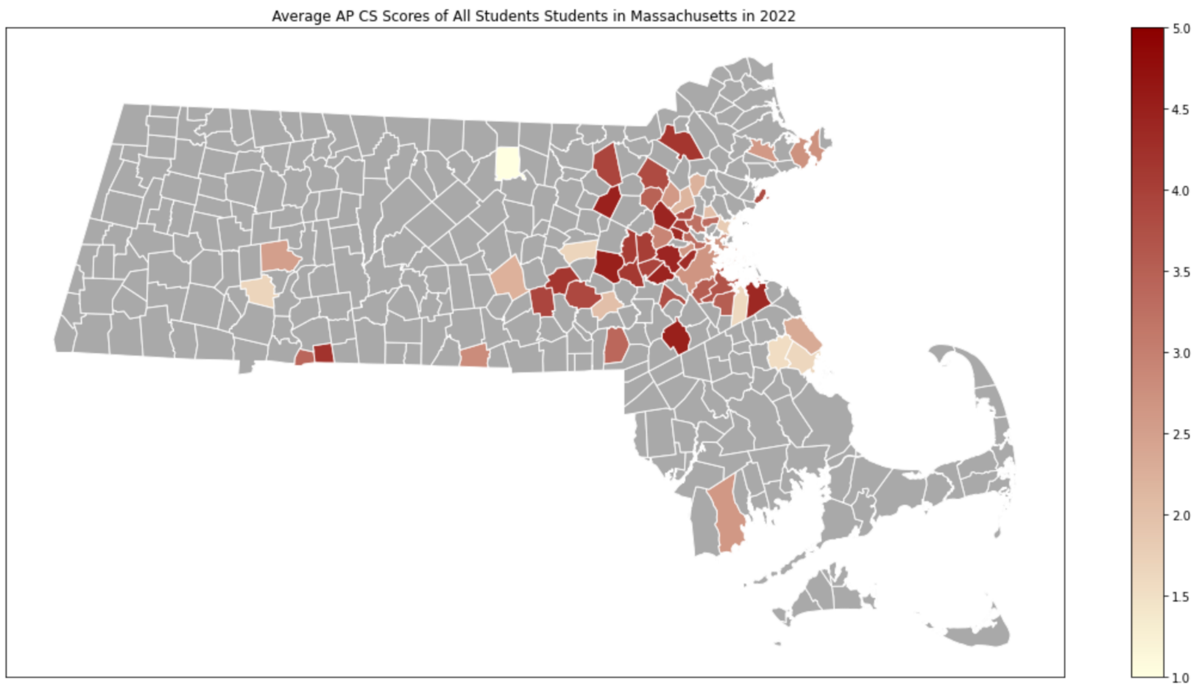
Gender Performance: Both male and female students present very similar average scores each year, suggesting an equitable level of understanding of the AP Computer Science material.



7. By Region

The analysis of average AP Computer Science scores across Massachusetts reveals a geographic concentration of higher academic achievement within the Greater Boston area. The visualization predominantly features darker shades of red in this region, indicating impressive attainment of scores at or above 4.0 by a substantial proportion of students.

Despite some data limitations, a discernible pattern emerges: areas located further from the Boston epicenter tend to register lower average scores, typically around the 2.0 to 2.5 range. This suggests a disparity in educational outcomes, with outlying regions possibly lacking the same level of emphasis on computer science education as found within the Boston vicinity. Such insights point to a potential need for strategic educational investments and initiatives to elevate computer science proficiency in the broader Massachusetts area.



5. Extension Project: Specific School Districts in Need

The extension proposal focuses on how the New Commonwealth Fund can manage its funds effectively, enabling most schools or districts to derive maximum benefits. This, in turn, will lead to a more balanced and improved computer science (CS) education.
Team A has identified three key areas of focus:

- Identify towns that have been lacking and emphasizing CS education in recent years.
- Determine the impact of school spending on CS participation.
- Develop strategies to address disparities in race and gender within CS education.

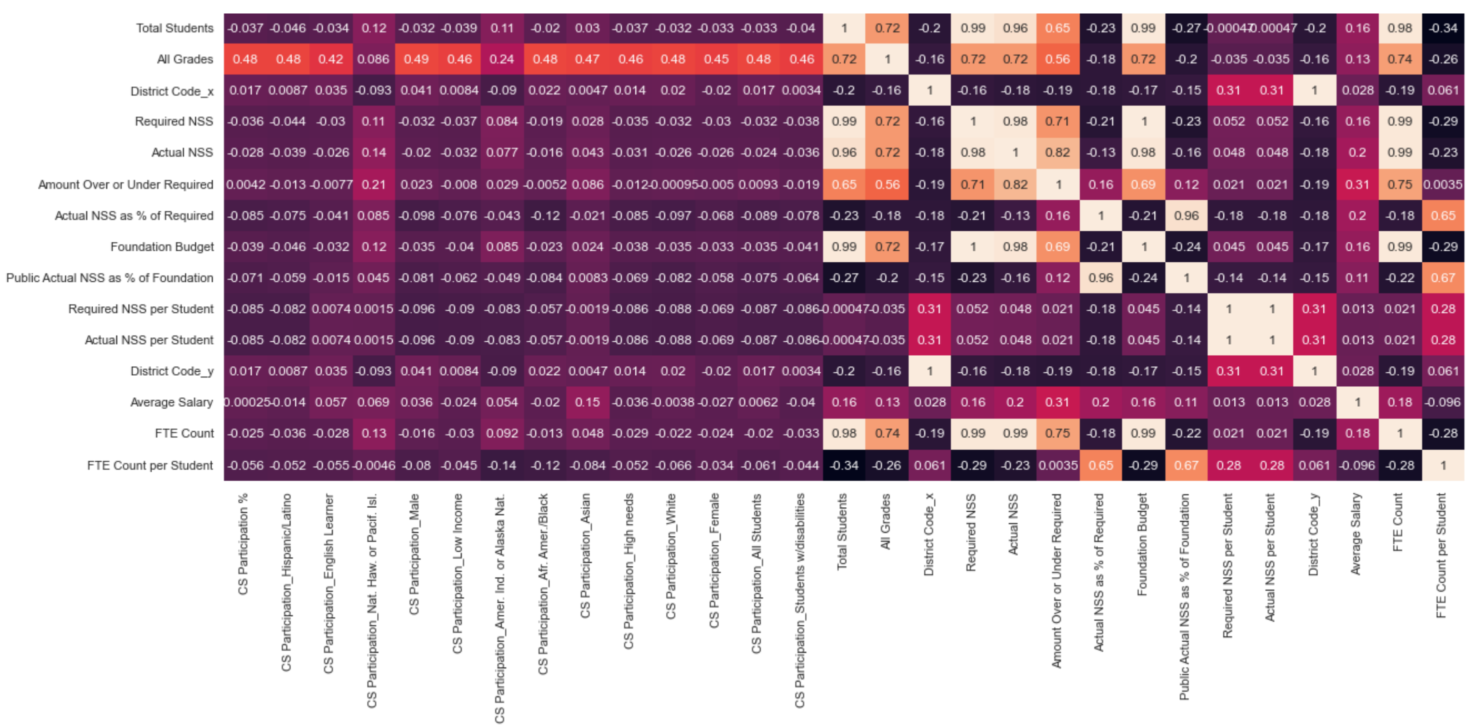
Identifying towns and school districts that have intensified their focus on Computer Science education is crucial to our analysis. Early findings from Team A suggest a positive correlation between funding participation and performance in computer science-related fields. However, it appears that these school districts are concentrated around the Greater Boston area, while the rest of the state experiences lower participation, leading to a decline in AP CS performance.

Determining the towns that need funding, specifically for CS education initiatives, will be pivotal in addressing this issue and disparity. Similar to the positive correlation between funding and AP participation, an equally opposite effect exists for school districts lacking funding for CS initiatives. Identifying these school districts and discerning patterns between high-performing and lower-performing CS schools will aid in remediating existing problem areas and predicting future problem areas accurately.

Extension Project Analysis

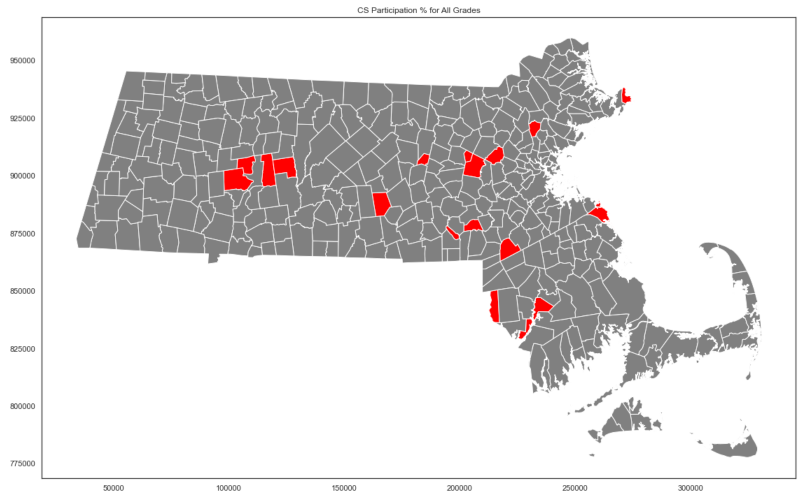
Analysis with School Spending:

We can see below that there is no strong correlation between school spending, teacher salaries with CS course participation for any of the student groups



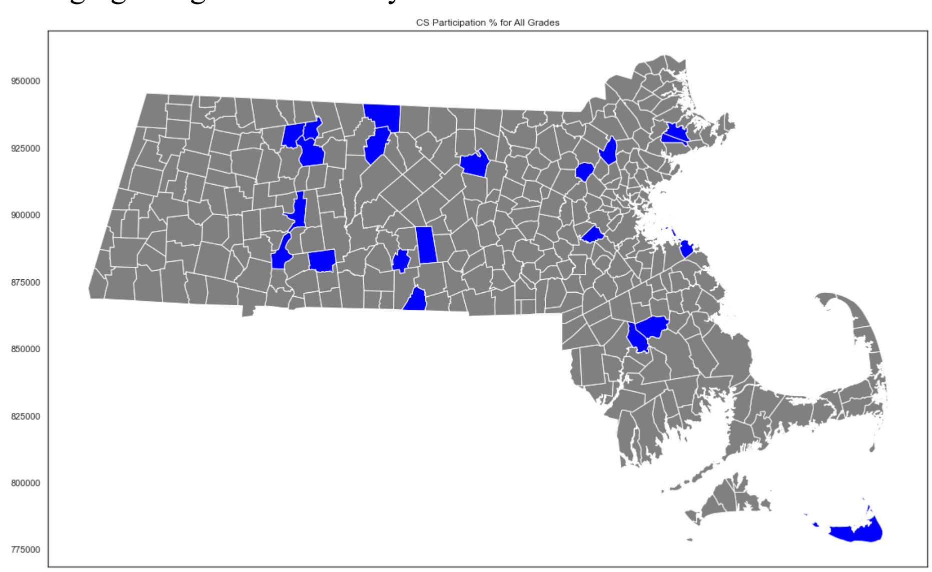
What are the School Districts That Have Historically Performed Well?

Visualizing the top 20 schools in AP Computer Science participation will provide a meaningful benchmark for subsequent inquiries. Historical data analysis uncovers that among the highest performers in AP CS engagement are districts such as Hatfield, Hopedale, Clinton, and Northeast Metropolitan Regional Vocational Technical, as well as the Somerset Berkley Regional School District, Maynard, and North Middlesex. Upon aggregating and averaging district demographic data, it emerges that these areas have an average population size of approximately 15,000, offering a potential demographic context for their success in AP CS participation.



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

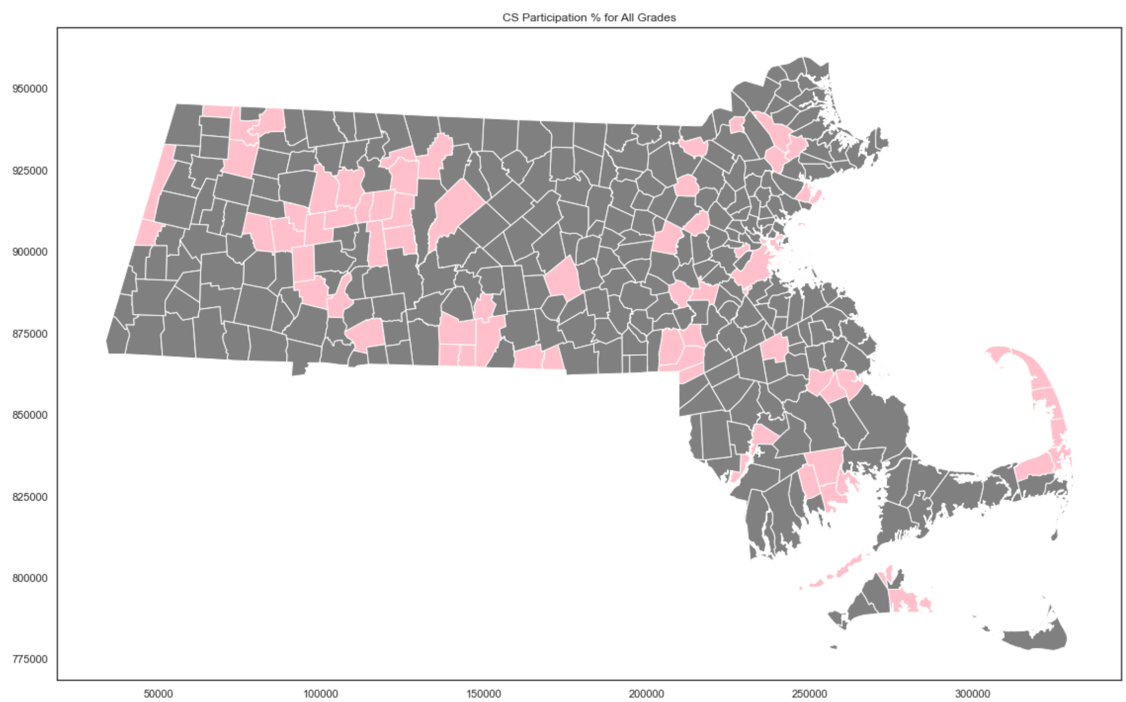
The graph illustrates a selection of districts that are in pressing need of support and resources. These districts are Spencer-East Brookfield, Greenfield, Greater New Bedford Regional Vocational Technical, Southbridge, Athol-Royalston, Gill-Montague, Gateway, and South Hadley. Notably, these districts have an average population size of approximately 26,000, which is notably larger than the populations of the districts exhibiting the highest AP Computer Science participation. The increased size of these school districts may present additional challenges in enacting and enforcing targeted educational policies, due to the complexity that comes with managing a larger student body.



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

The graph in question highlights school districts that exhibit minimal or no engagement in computer science education. The districts identified include 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), and Boston Renaissance Charter Public (District). A common characteristic among these districts is their charter school status, which allows them to operate independently from the conventional state school system. This autonomy often results in varied educational focuses, with computer science education not always being a priority.

Furthermore, it's notable that these districts tend to have considerably smaller student populations, averaging around 4,000 students. This figure is substantially lower compared to the larger districts discussed previously, suggesting that the scale of a school district may have a significant correlation with the presence and quality of computer science education programs.



6. Implementation of Strategy

Based on our comprehensive analysis, we propose the following strategic interventions, specifically aimed at districts with low computer science (CS) performance or no CS participation:

1. **Develop Computer Literacy Initiatives**: Implement computer literacy programs in districts lacking CS education. These should not only establish foundational tech skills but also bridge to more advanced, college-credit-worthy CS coursework.

2. **Establish Dedicated CS Teaching Roles**: Prioritize the creation of CS teaching positions at the high school level to ensure that students receive instruction from educators who are both skilled and passionate about the subject.

3. **Resource Allocation for High-need Groups**: Direct resources towards students from groups that have historically been underrepresented in CS. This includes providing scholarships and financial incentives to encourage the pursuit of CS education and careers.

4. **Financial Aid for CS Examination Fees**: Offer financial support for students unable to afford fees for CS-related IB/AP examinations, thereby promoting broader participation and enabling students to earn college credits.

5. **Equip Students with Necessary Technology**: Allocate funds for programs that supply computers and coding tools required for CS learning, with an emphasis on the utilization of these tools for CS studies alongside general education.

By introducing CS education early in schools with limited exposure, we can ignite interest in technology fields. Hiring dedicated and inspirational CS educators is critical to nurturing future talents. Addressing the needs of historically disadvantaged groups and removing financial barriers are pivotal steps in cultivating a diverse and inclusive environment for CS learning. Emphasizing CS through targeted resource provision can significantly enhance participation and proficiency in this vital field.

7. Limitations - Challenges Faced

Throughout this project, we encountered several obstacles that presented challenges to our analysis:

1. **Incomplete Data**: A significant hurdle was the prevalence of missing data within the Massachusetts Department of Education's datasets. Despite the datasets' breadth in categories and demographics, the incompleteness was particularly pronounced in the case of smaller schools. To ensure a robust and unbiased analysis, we applied best practice data adjustment techniques to address these gaps.

2. **Data Consolidation Difficulties**: The process of synthesizing multiple datasets was met with complexities, primarily due to inconsistencies in file naming and organizational schemes. We overcame these challenges by implementing a uniform file naming system. Additionally, we encountered obstacles when attempting to merge multi-year datasets, a task we streamlined through the development of a web crawler, which facilitated the aggregation and unification of data for more efficient analysis.

8. Conclusions Drawn - Key Points

The analysis and visualization of the datasets yield several critical insights, underscoring the need for a nuanced understanding of computer science (CS) education outcomes across various demographics. Key observations include:

1. **Disparities Among Student Groups**: A pronounced variance in CS educational outcomes is evident among different student groups. This discrepancy underscores the urgency for targeted educational strategies to bridge the gap.

2. **Geographical Disparities**: A concentration of CS participation in the Boston area raises concerns about regional inequities. Such disparities, rooted in geographic location, can perpetuate a state-wide imbalance in educational opportunities.

3. **Focus on Regional Analysis**: Our deliberate focus on regional analysis casts a spotlight on the divergent CS educational outcomes across districts. Identifying districts that lag in CS education is pivotal in tailoring interventions.

4. **Strategic Implementation**: The optimal course of action for deploying effective interventions lies in bespoke, precise initiatives rather than undifferentiated programs with a broader scope. Such targeted efforts promise a judicious use of Massachusetts' educational resources, aiming to democratize access to CS education for all students, irrespective of their location.

In essence, the data calls for a strategic, targeted approach that prioritizes equitable access and quality of CS education, aligning resources with areas of greatest need to ensure comprehensive, statewide educational advancement.

9. 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: Led the data extraction and collection. Performed the EDA for all datasets. 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.