Deliverable 2 - Team A

Problem Statement, Data Cleaning and Collection:

Problem Statement:

In an era marked by rapid technological advancements and digital integration, computer science education stands as a critical pillar for economic opportunity and social equity. The objective is to identify the gaps and barriers faced by communities in computer science education and AP test-taking.

The insights derived from this analysis will be instrumental for the NCF to pinpoint areas for investment and potential strategic partnerships with nonprofits or corporations. These collaborations can be cultivated to promote racial equity and social justice in education.

The culmination of this effort will result in a proposal that not only illuminates the current state of educational disparity but also recommends actionable steps for the NCF. The aim is to bring about tangible change, ensuring that computer science education serves as a tool for empowerment rather than a gatekeeper of opportunity.

### *Data Collection and Cleaning:*

- **Data Retrieval**: Downloaded datasets for the years 2021 and 2022, focusing on various student groups at the district level from the DESE website.

- **Initial Inspection:** Conducted an initial review of both datasets to comprehend the structure, content, and identify any apparent inconsistencies or missing data.

- **Combining Datasets:** Merged the 2021 and 2022 data files into a unified dataset, ensuring alignment of similar categories and consistency in data types across both years.

- **Categorical Data Consistency:** Ensured that categorical data (e.g., race, ethnicity, gender) were consistent throughout the dataset, with no variations in spelling or categorization.

- **Verification Against Source:** Verified a random sample of the cleaned data against the source to validate the accuracy of the cleaning process.

Exploratory Data Analysis:

CS Participation:

CS Participation % by Student Group and Year:

a. Asian and male student groups are above average.

b. There is an increase in CS participation for all student groups over the years.

c. English Learners had the least participation, followed by Nat. Haw or Pacf.Isl.
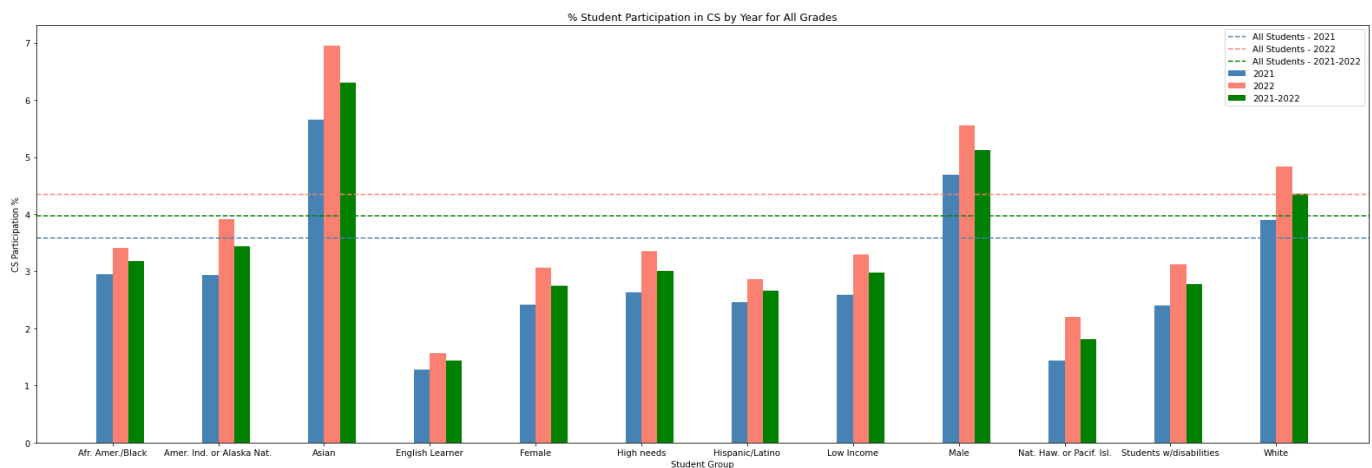
AP Participation:

When examining the heat density maps, we observe that most tests are taken in the Greater Boston area. It is evident that areas outside of the Greater Boston Area either have no tests or are missing data.


Average AP Tests Taken for All Students in Massachusetts in 2022
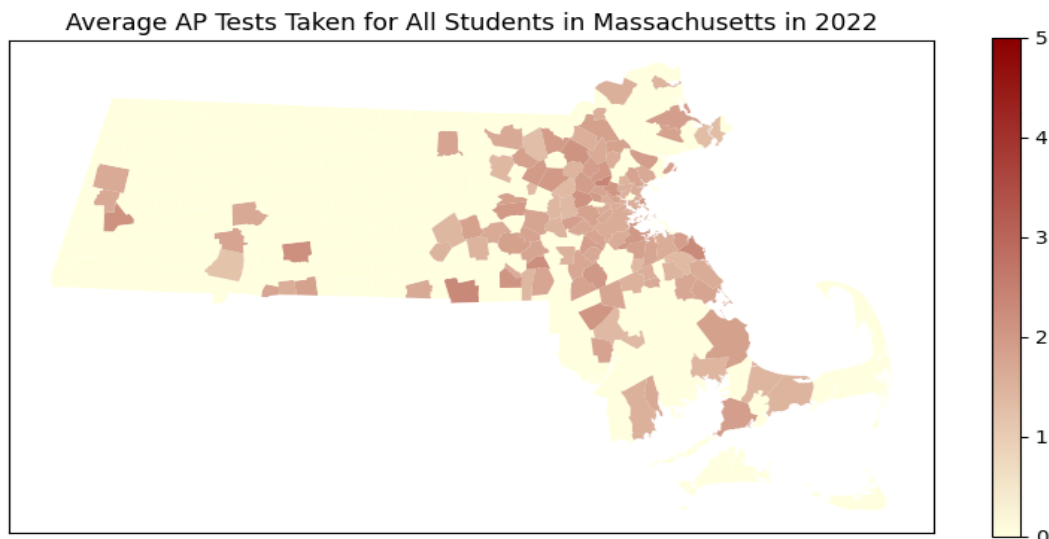
# Methodology, Visualizations, and Insights:

AP CS Participation:

We have observed obvious differences in CS course participation across different student groups. To analyze this further, we look at:

1. **Distribution in CS Participation across grades for all Student Groups:**

   A. *For the highest participating student groups, Asian and Male, the % share across classes keeps increasing.*

   B. *English Learners have the opposite trend which indicates the interest but highlights the challenges.*

   C. *Nat. Haw or Pacf.Isl seems to have started the participation late and can be encouraged or provided adequate resources and information to start early.*
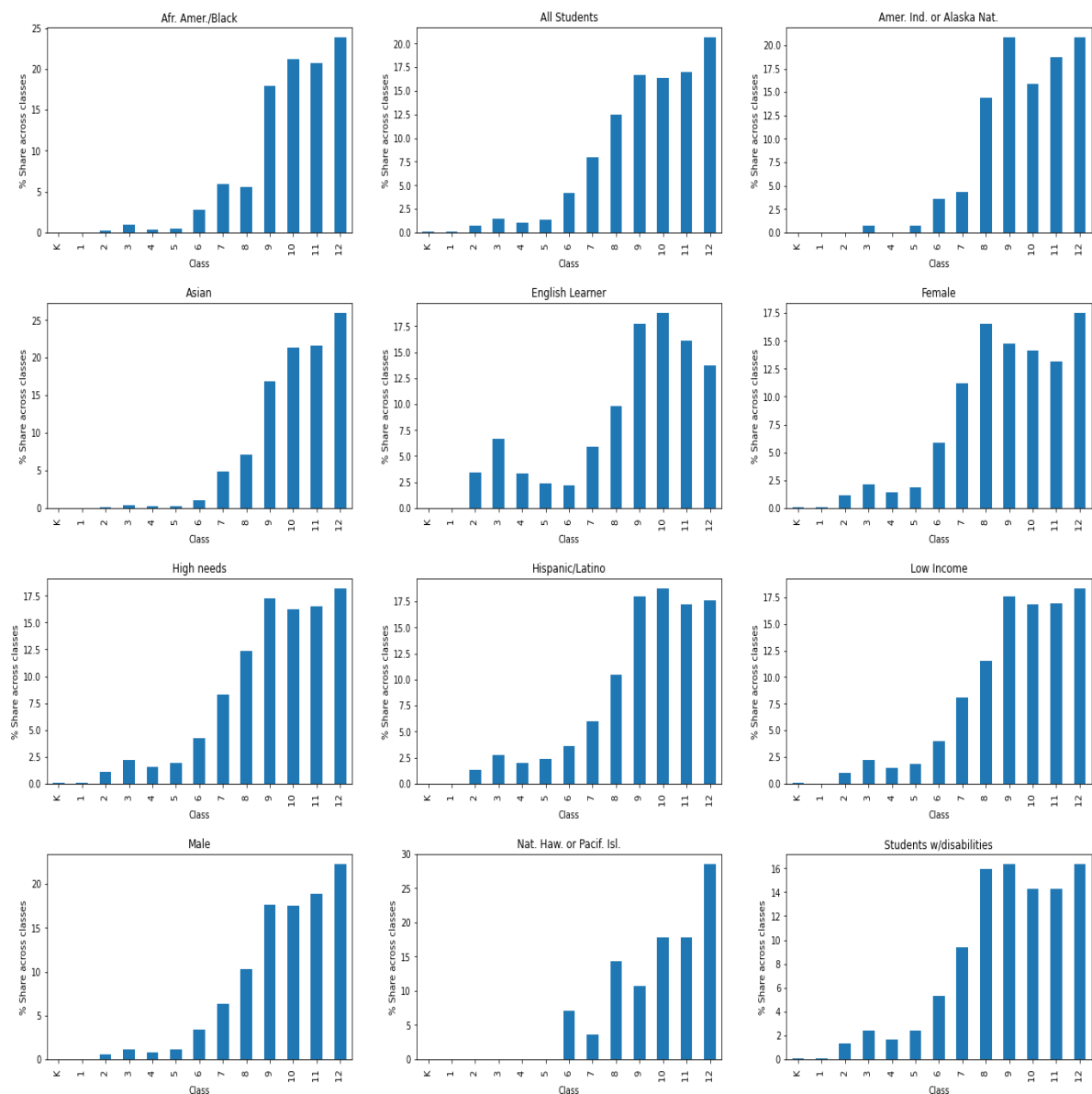
2. Geographic Distribution in CS Participation:

CS Participation % for All Grades
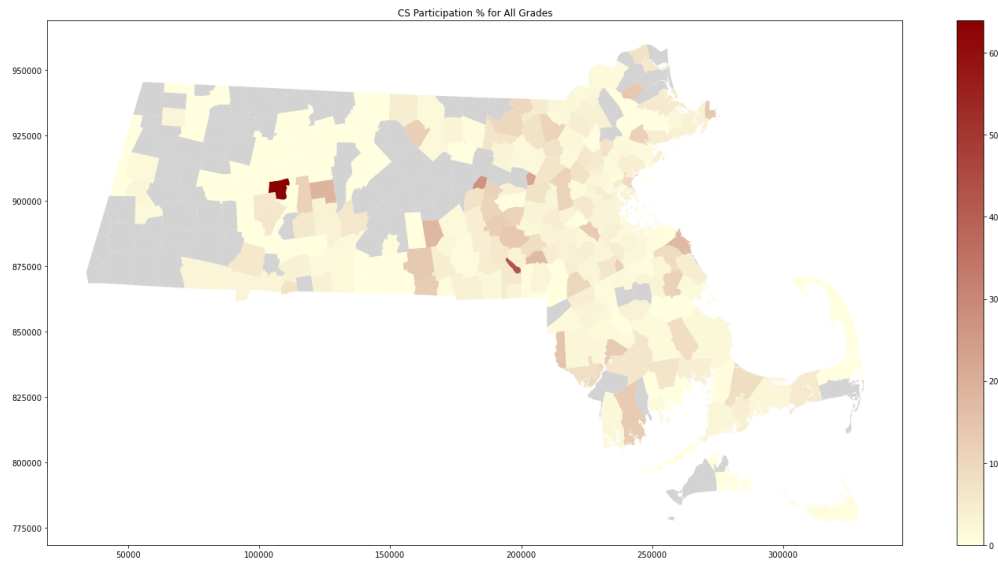
D. Geographically, the performance across the districts with a high urban population seems to be consistent.

E. Some districts have exceptional performance and need to be analyzed to understand if they are outliers or can be used to inform other districts.

AP CS Performance:

The performance across various student groups has less variation compared to CS course taking. Still, we similarly observe that Asian and Male student groups

Average Score by Student Groups for Year 2021

Average Score by Student Groups for Year 2022
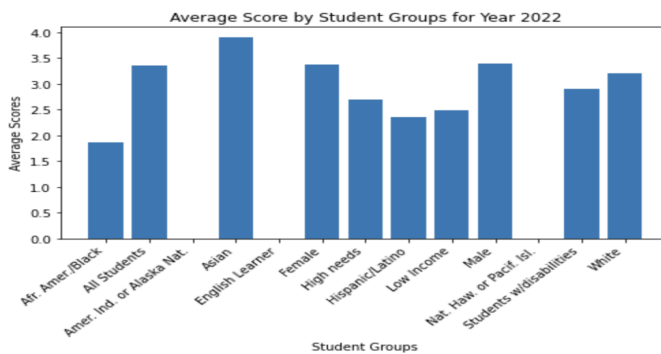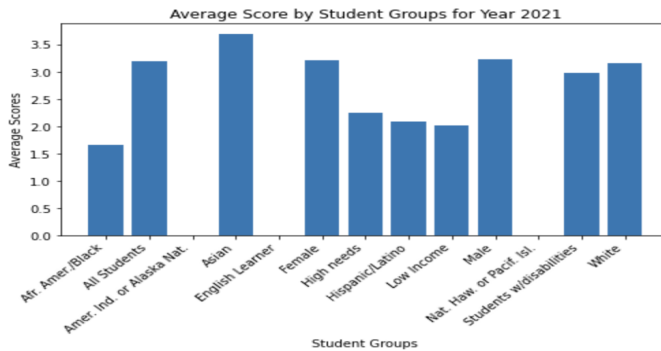
All AP Participation (# Tests Taken):

As students continue to take multiple AP tests, a noticeable contrast becomes apparent between the white male student groups and the rest of the student groups.
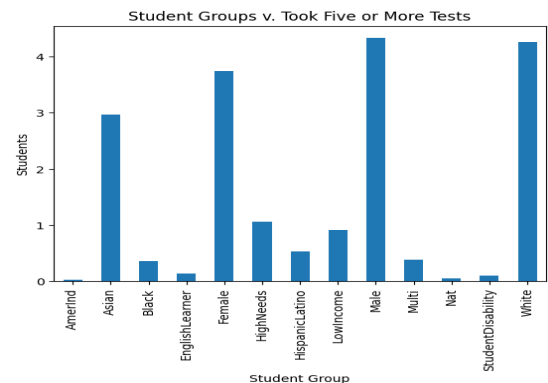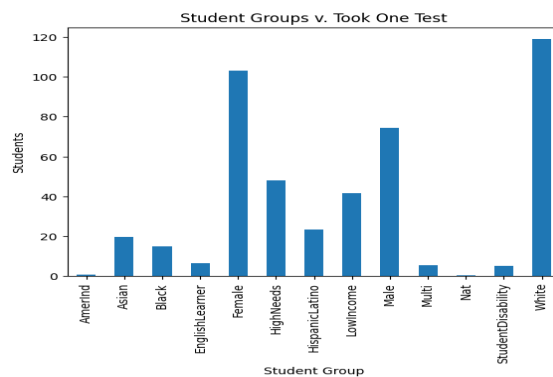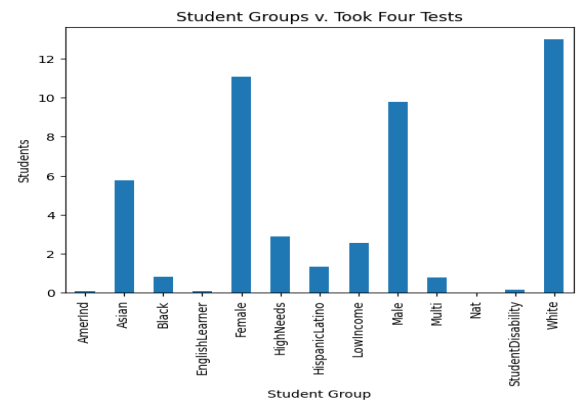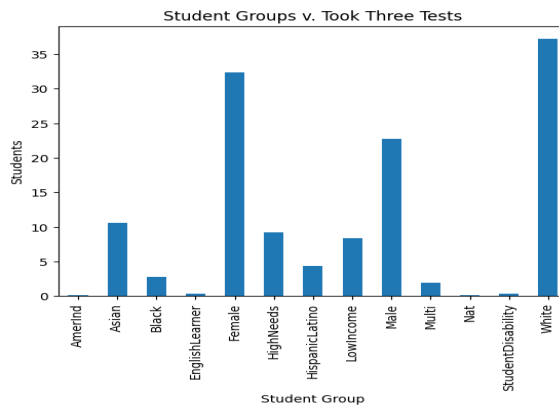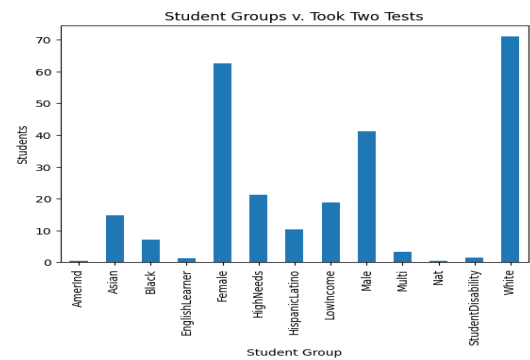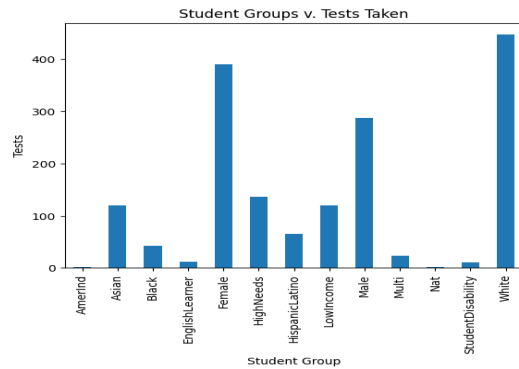
Extension Proposal:

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:

1. Identify towns that have been emphasizing CS education in recent years.

2. Determine towns lacking funding that require increased financial support for CS education.

3. 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 and 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.

Geography is not the sole factor influencing CS education in Massachusetts. It is essential to examine traditional disparities between race and gender within CS education. Minorities in Massachusetts, as with many areas, have been found to have poor CS participation and performance. Research also indicates lower CS participation among females compared to males. Team A is analyzing trends and general patterns among different groups to formulate effective solutions addressing these issues.

The next step involves researching precise strategies to rectify disparities in AP CS participation and performance, both geographically and across group boundaries. Team A plans to utilize heat maps incorporating AP scores data, SAT data, and budget data from the Massachusetts Department of Education. The goal is to identify initiatives that can be implemented to effectively address the issues uncovered in the project.

# Extension Proposal - Methodology, Visualizations, and Insights:

AP CS Average Scores by Student Groups:

The graph below 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. However, certain student groups, such as American Indian and English Learners, lack available data. Additionally, the graph indicates an improvement in performance in 2022 compared to 2021.
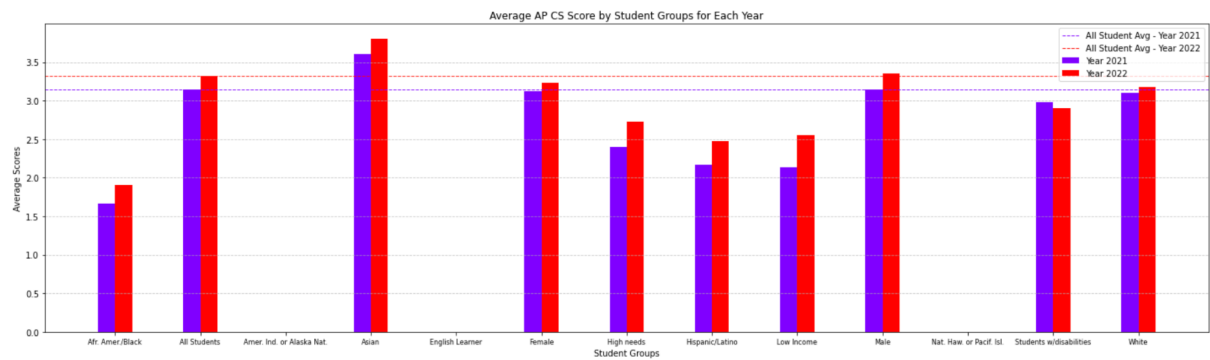
Heat Map of Average Scores:

Here is another 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.
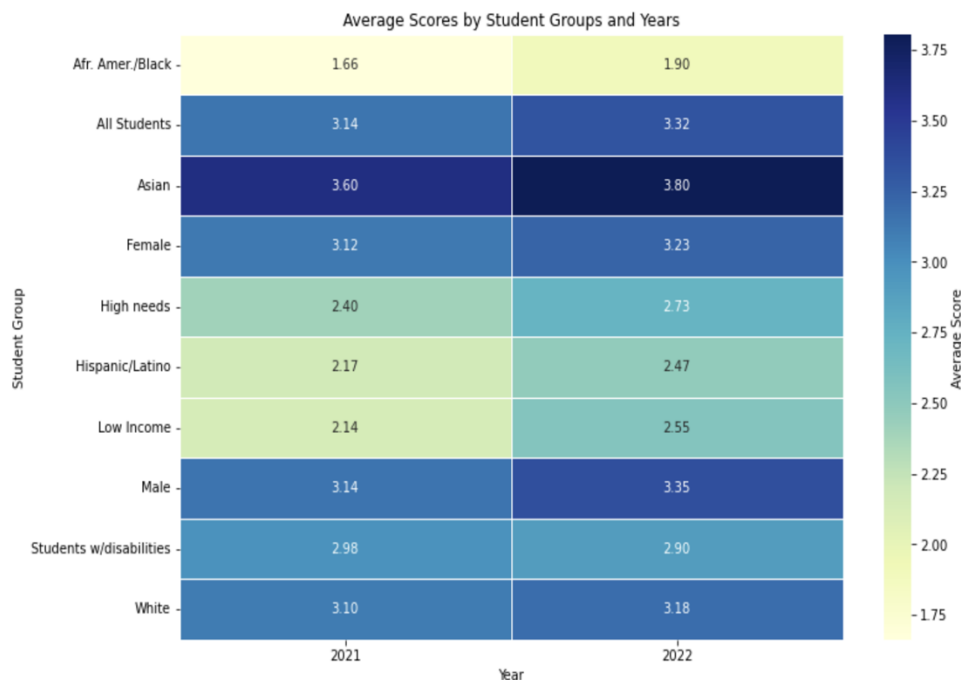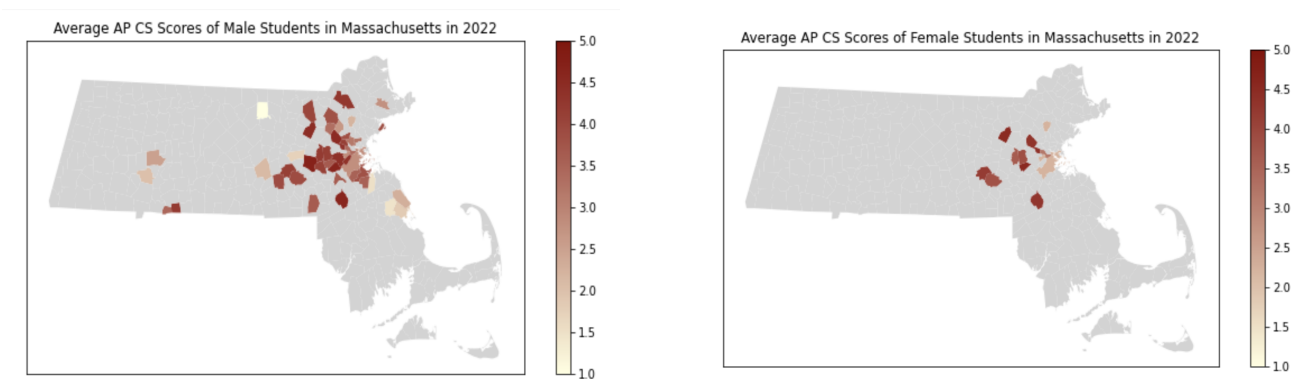
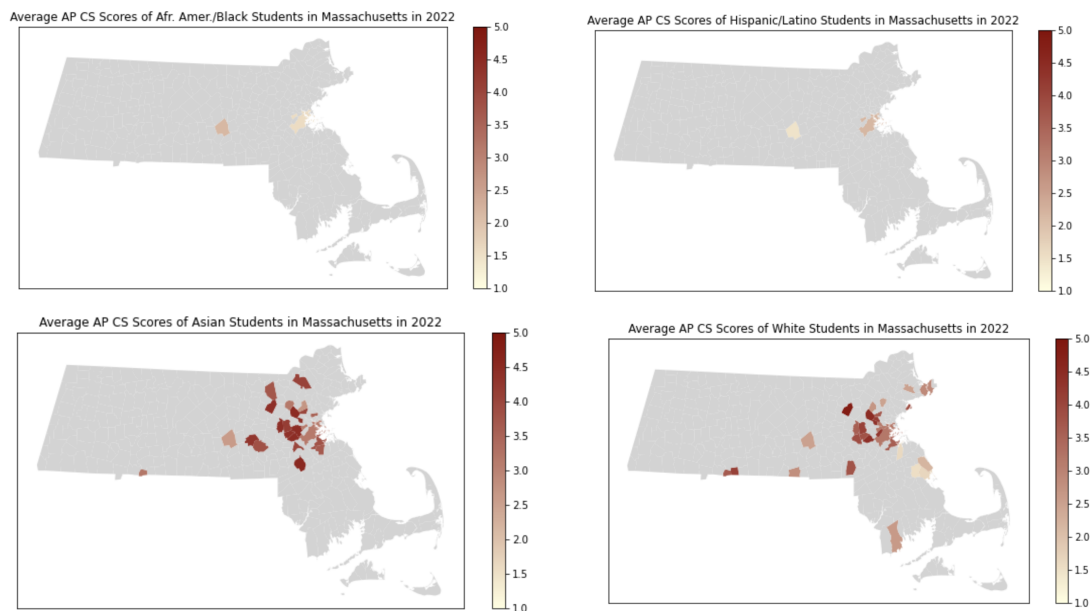# HeatMap of AP CS Performance by Student groups

Analysis:

In analyzing heatmaps across different student groups, it is evident that males and females achieve relatively similar scores. However, there is a notable discrepancy in participation, with males participating more than females. Additionally, a significant concentration of students taking AP CS exams is observed around the Boston area.

Furthermore, there is a marked contrast in participation levels among different ethnic groups. Asian and White students show significantly higher participation compared to African American/Black and Hispanic/Latino students.
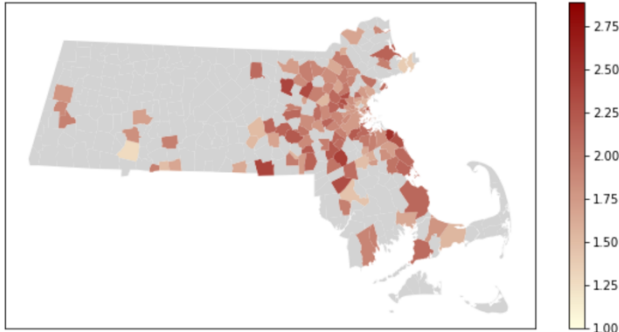
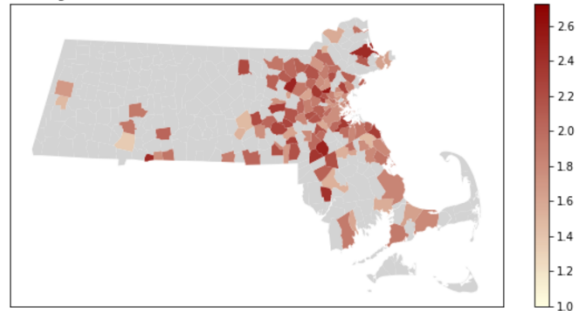## Disparity by Genders:

Disparity by ethnicity:

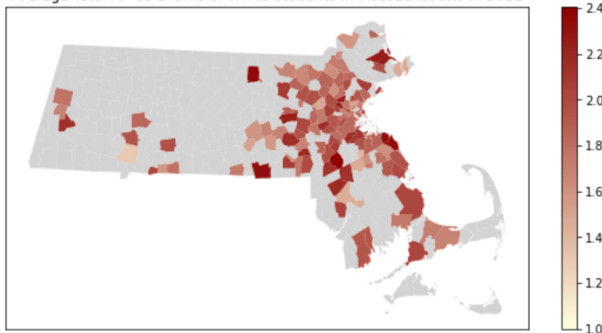# HeatMap of AP CS Participation by Student groups

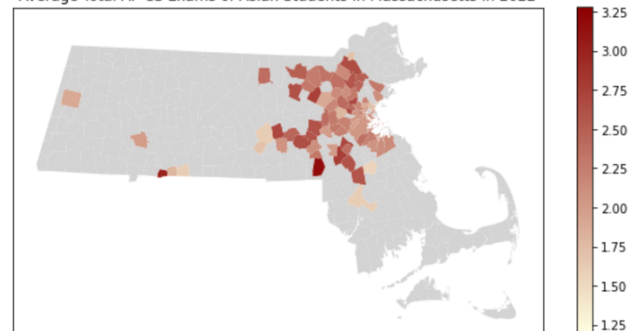Average Total AP CS Exams of Female Students in Massachusetts in 2022

Average Total AP CS Exams of Male Students in Massachusetts in 2022

Average Total AP CS Exams of White Students in Massachusetts in 2022

Average Total AP CS Exams of Asian Students in Massachusetts in 2022

Conclusion:

It is apparent that substantial disparities exist among various student groups and locations. Students in the Boston areas demonstrate higher levels of both participation and performance. Participation and performance are notably higher for Asian and White students, while African American/Black and Hispanic/Latino students exhibit lower levels.

The recommendation is for the New Commonwealth Fund to allocate more funds to areas outside of Boston and encourage increased participation in computer science among African American/Black and Hispanic/Latino students. This could potentially address issues related to lack of interest and financial constraints.

Individual Contribution

Changxuan Fan:
Team Lead. Responsible for group coordination, leading the editing of the report, and leading the direction of the project. Also, responsible for visualization of the Extension proposal and creating heat maps of average AP CS exam performance. Record the video.

Kelvin Lin:
Responsible for the Extension proposal and creating heat maps of average AP CS exam participation. Record the video.

Sai Tejaswini Junnuri:
Extracted the CS Participation dataset and conducted its analysis. Examined the differences in CS participation across student groups, grades, and geography. Compiled the Exploratory Data Analysis (EDA) and methodology sections. Record the video.