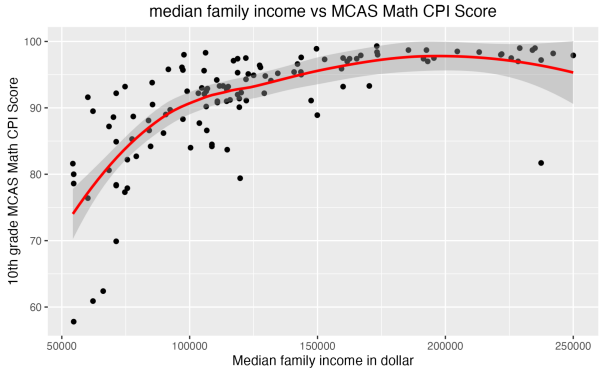
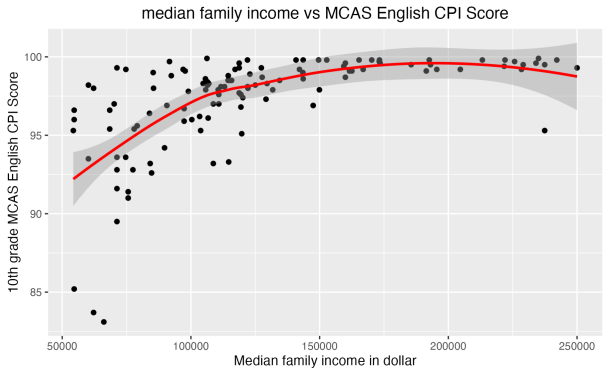
Title: Socioeconomic Demographics and Standardized Test Performance

In modern society, people with higher socioeconomic status often have access to high quality educational resources and thus, benefit more from education than their counterparts with lower socioeconomic status . This dataset investigates the relationship between median family income and students' MCAS CPI Scores. By definition, the median family income is the income level at midpoint that divides a population into two equal-sized groups. The MCAS CPI score, standing for Composite Performance Index, is a mathematically rigorous way to measure the proficiency level of students in a high school in the Massachusetts Comprehensive Assessment System. Below are two scatter plots illustrating the relationship between median family income and 10th-grade MCAS CPI scores in Math and English.

Each black point on the graph represents a high school in Massachusetts. There are a total of 128 high schools in this dataset. The red regression curved line indicates that as median family income increases, CPI scores grow logarithmically. The gray shaped area above and below the red regression line illustrates the confidence level and captures most high schools. This is an indication that underlying patterns are being accurately captured. I performed a t-test regression analysis on these two distinct variables, and the p-values are $1.321 * 10^{-10}$ and $4.087 * 10^{-14}$ for the subject in English and Math, respectively. This indicates that the results are statistically significant. In other words, family income significantly affects how well a student performs in a standardized exam. In conclusion, students from families with higher median incomes tend to achieve higher scores on the MCAS.



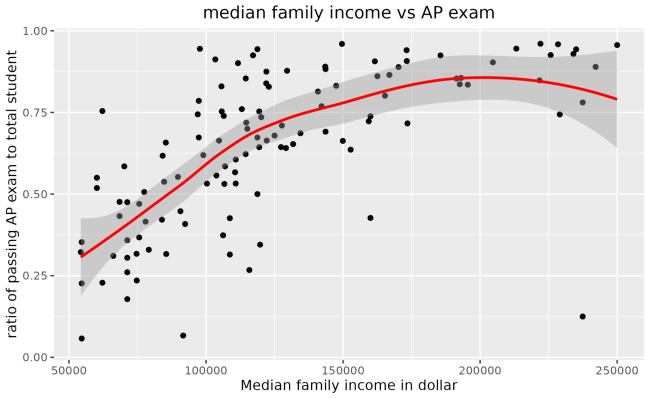


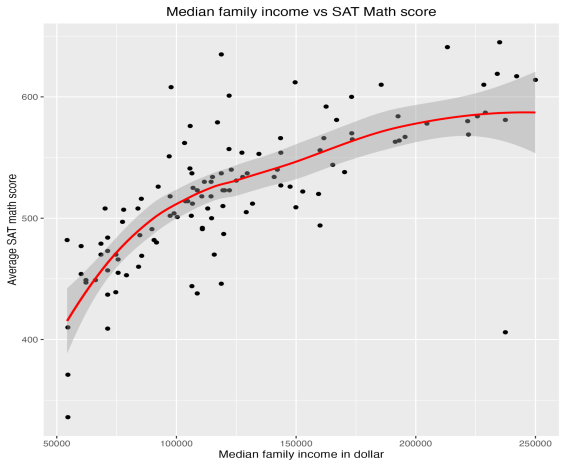
Knowing this shocking relationship between income and CPI Score, we would like to investigate if the underlying pattern remains true in other standardized exams such as the Advanced Placement (AP) and Scholastic Assessment Test (SAT). AP and SAT exam assessment evaluates ability of high school students in English and problem solving skills. These exams provide an indicator of academic success.

AP exam is one of the direct ways for high school students to earn college credit in high school by taking college courses if they pass the exam with 3 above. The SAT exam is another important component of college application when high schoolers apply to colleges. Having some background with what these exams are is important to understand the investigation of this research. As you may see, when income increases, the ratio of number of students passing the

AP exam to total number of students who take the AP exams increases logarithmically shown by the red curved regression line. A p-test was performed on this interesting relationship, which then returns a p-value of $2.218 * 10^{-14}$. This indicates a statistically significant correlation between family income and AP exam success rates. This finding is important as it demonstrates the influence of socioeconomic status on educational achievement and highlights the unfairness of standardized tests towards students from low-income families.

The second graph on the right supports the hypothesis that lower mean family income is correlated with lower MCAS scores and SAT scores. A similar analysis was performed on family income and SAT scores. A statistical analysis with a p-test returns a p-value of $2.2 * 10^{-16}$. In short, having lower family income leads to unfair educational journeys, especially in standardized exams that determine a student's college admissions.





Title: Socioeconomic Demographics and Standardized Test Performance

Group Members: Elizabeth Mackenzie, Yonge Tan, Maekala Turner

Objectives & Interests: Our objective is to explore the relationship between household income and standardized test performance. Our interest is in investigating how demographic predictors correspond to test results. The achievement gap is a prominent and ongoing socio-economic issue within education, so we expect our analysis to support the hypothesis that lower mean family income is correlated with lower MCAS scores and SAT scores.

Group Member responsibilities:

Yongye:

- Find the dataset on Kaggle
- Preprocess and clean the dataset
    - Remove unnecessary columns
    - Remove missing/ NA values
    - Filter original dataset with 1800+ rows to 120 rows
        - filter out high school
        - Filter out number of enrollment greater than 800
- Web Scraping
    - Scrape median household income, population, and household on a HTML table in this wikipedia page
    - Add the scraped data to the preprocessed dataset
- Analyze relationship between median family income in dollar and MCAS Math and English CPI Score
- Analyze relationship between median family income in dollar and number of students who pass the AP exam
- Regression analysis, p-value, correlation test

Elizabeth:

-Bootstrap analysis to confirm p-value

-Monte Carlo power test

Maekala:

- Visualizations delving deeper into school-based predictors of test performance such as average class size, expenditure per pupil, and enrollment numbers.
- Set up Google Doc
- Refine abstract
- Presentation visuals/details