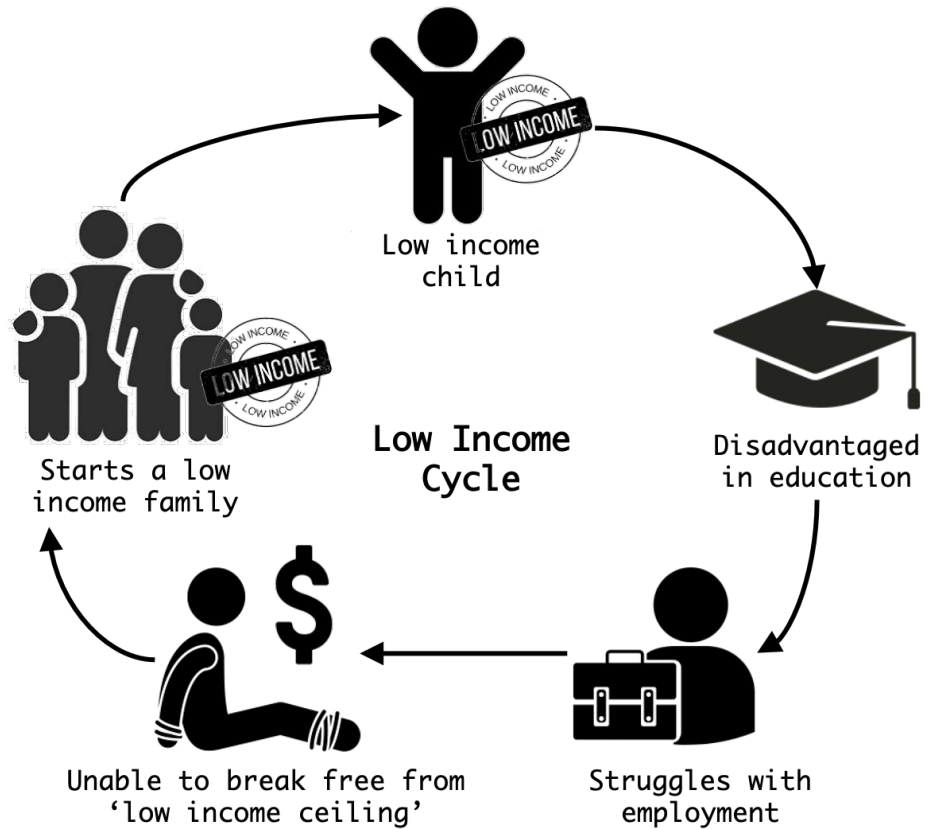
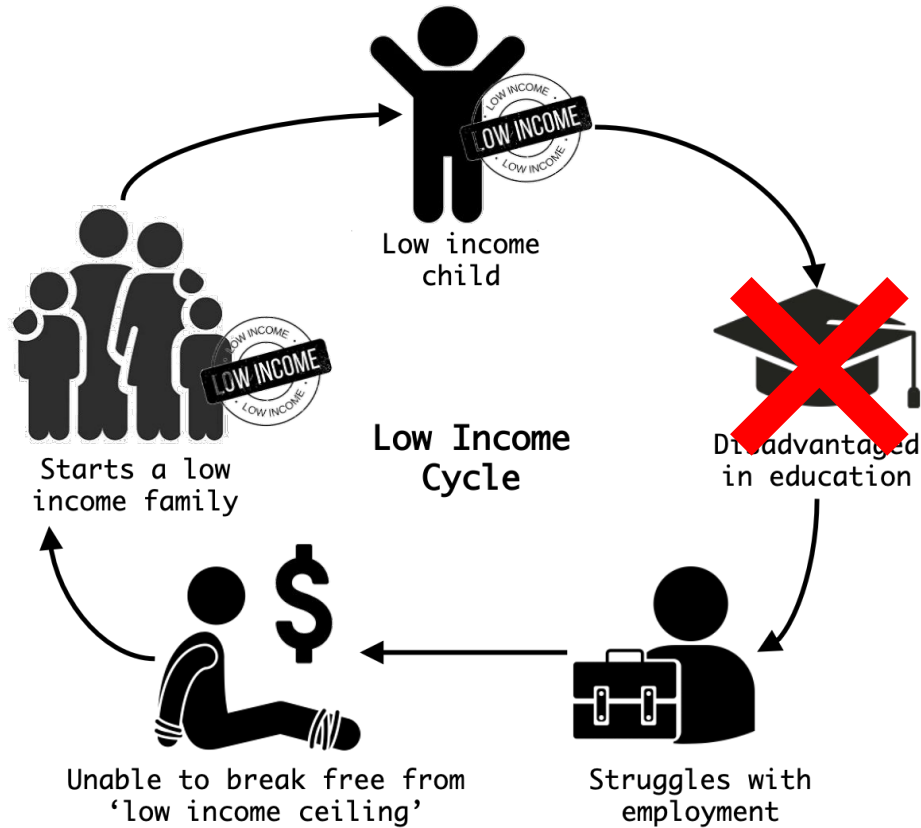




{‘DSIF3’ : ‘Group5’:  
[‘Project\_1’]}



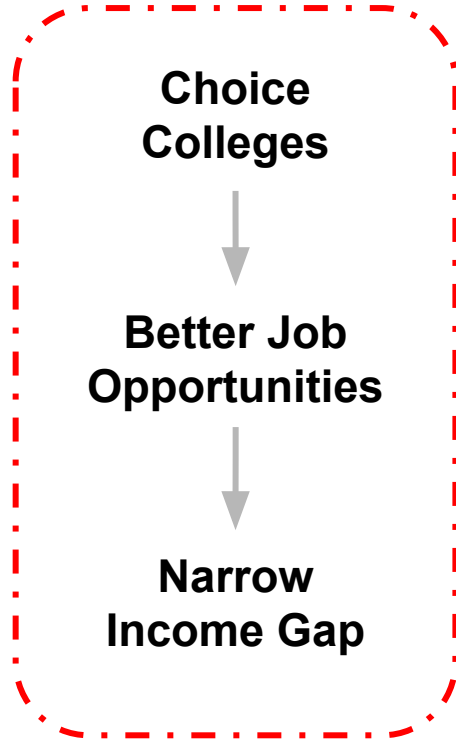
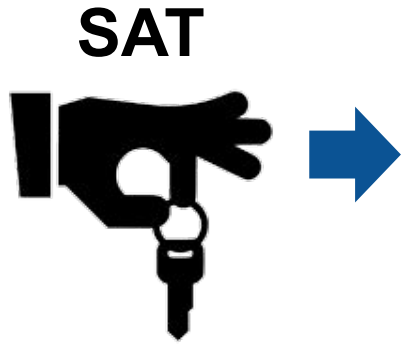




**Break the cycle here!**

*“**Education**, then, beyond all other devices of human origin, is **the great equalizer** of the conditions of men, the balance wheel of the social machinery.”*

— Horace Mann, 1848.



As a team of data scientists from the Board of Education of California, we want to...



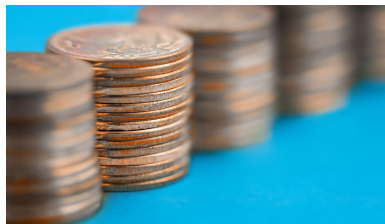
Identify the relationship between **County Income** and **SAT Participation Rate**.

Should there be a **Positive Correlation**, we want to propose qualitative solutions to **improve SAT participation rates of students from lower-income counties**.

# Executive Summary

1. Background Problem Statement
2. EDA (Procedures and methodologies)
3. Visualisation
4. Scholarly Findings
5. Conclusion and Recommendation

## Procedures & Methodologies



Income?



Location?



Culture?



Why Low  
SAT participation?

## Procedures & Methodologies (contd): Importing libraries



CSV files used:

- 2019 Participation rate and scores of counties for SATs in the various counties in California, US
- 2019 Household median income by county

# Procedures & Methodologies (contd): Cleaning Data

## Checklist for Cleaning Data

Data Quality (validity, accuracy, completeness, consistency, uniformity)

- ☐ Data drawn from U.S. Census and U.S. College Board
- ☐ Constraints considered (eg data type, mandatory, unique)
- ☐ Cross field validation (e.g. Total score of SAT cannot be less than Math score)

## The workflow

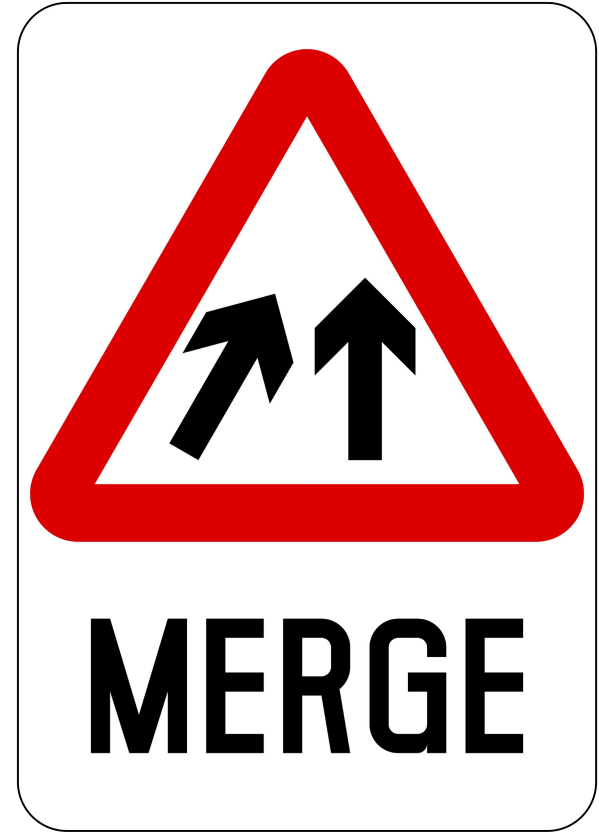
- ☐ Inspection (data profiling, visualizations. etc)
  - ☐ Detect unexpected, incorrect, and inconsistent data.
  - ☐ Check for missing values (NaN values) and management of it.
- ☐ Cleaning
  - ☐ Drop unwanted/irrelevant/duplicate columns
  - ☐ Type conversion, Fix syntax errors (e.g. white spaces), typos
  - ☐ Column names changed to lowercase
  - ☐ Replace weird/undesirable values
  - ☐ Ensure column names are unique and informative
  - ☐ Made values more readable by removing unnecessary words
  - ☐ Standardized!





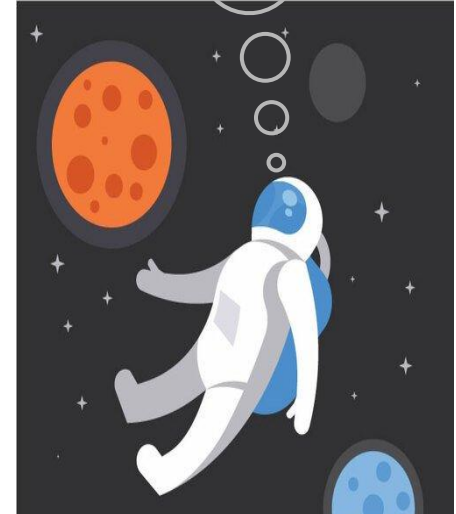
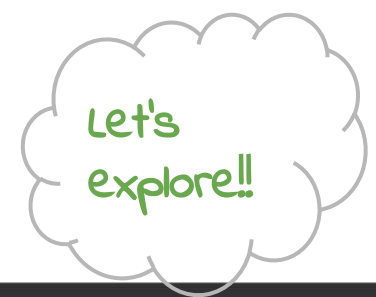
## Procedures & Methodologies (contd): Merging Data

- **Merged 2 files:**
  - 2019 Participation rate and scores of counties for SATs in the various counties in California, US
  - 2019 Household median income by county
- New merged dataframe (e.g. sat2019calinew) saved as CSV in code folder
- Dataframes merged so statistical analysis can be done



## Procedures & Methodologies (contd): Data exploration using the new dataframes

- Inspection part 2 (data profiling, visualizations, etc)
- Data profiling again:
  - Measures of central tendency: e.g. mean, median mode.
  - Measures of variability: e.g. Standard deviation, minimum and maximum values.
    - E.g. `df.describe()` to find mean value of participation rates across counties in grade 11 and 12 respectively
  - Shape of a dataset's distribution
  - Sorted values functions often used E.g. to find the top 10 counties with highest SAT participation rates in grade 11 and 12, and bottom 10 counties with lowest SAT participation rates in grade 11 and 12
- Visualization
  - Generated Plot box, heatmap and Bar charts to discover data trends



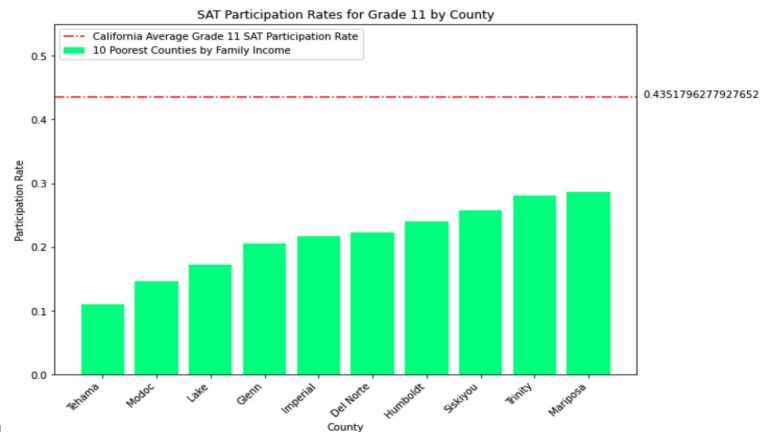
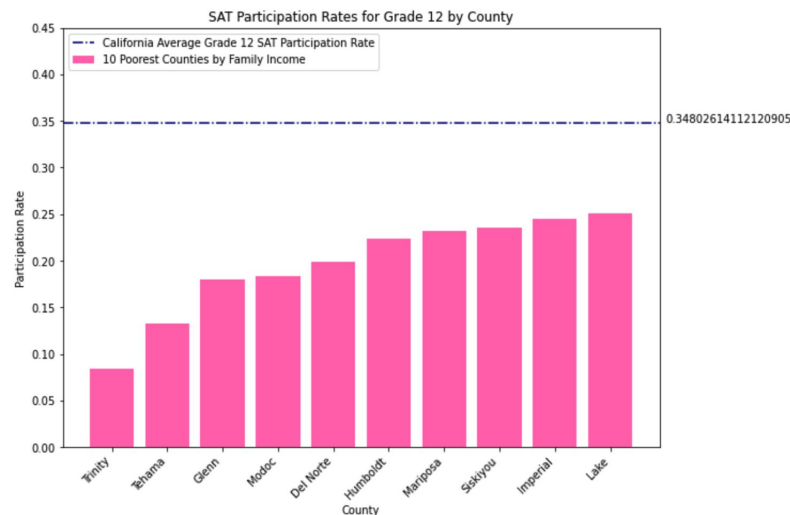
# Visualization (Overview)

- 1) Charts of Low Income Household vs Participation Rate
- 2) Heatmap - 0.5 Correlations
- 3) Linear Regression Model - Positive Correlations

# Participation rate vs Low income

Charts shows low income counties has much lower participation rate in taking SAT for Grade 11 & 12

- For example, Trinity with the lowest participation rate for SAT G12.
- Followed by Tehama which have low participation rates for both SAT G11 & 12.

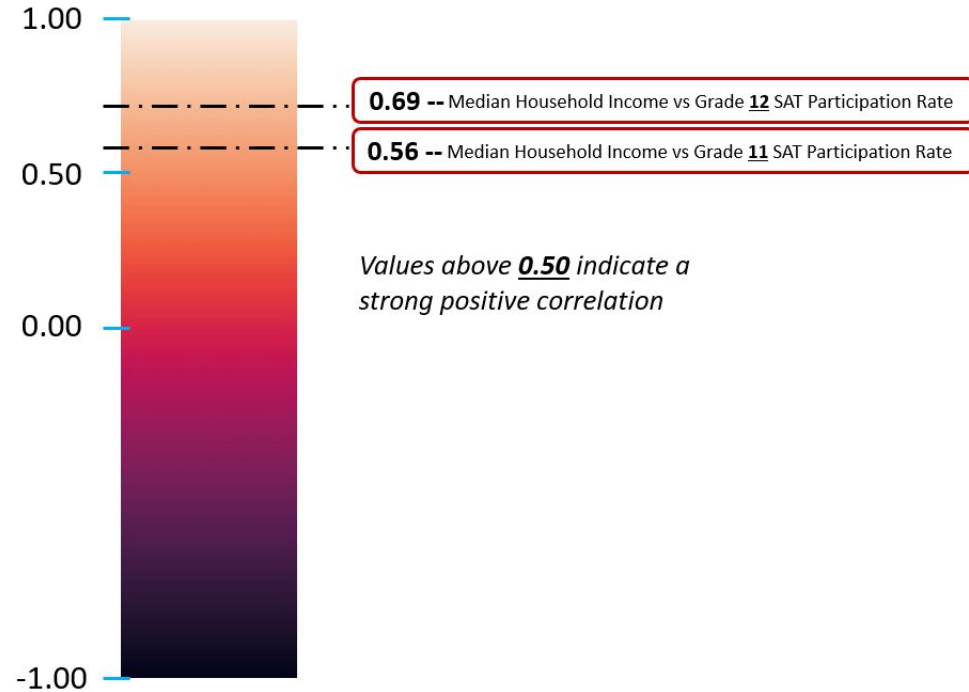


## HEATMAP- Participation rate vs Low income

Figure on right shows, Heatmap of Participation Rate vs low income Household

Grade 12 SAT = 0.69

Grade 11 SAT = 0.56

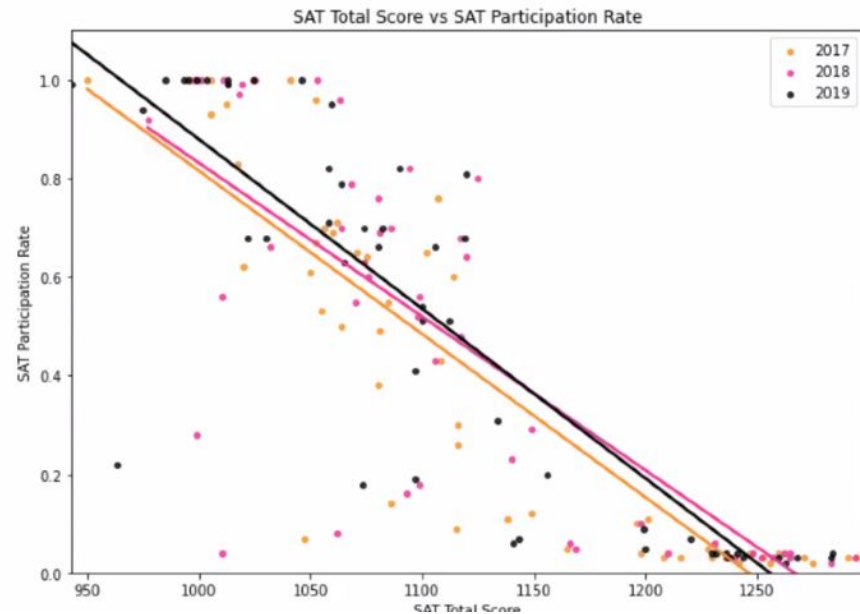


# linear regression model

Fig shows participation rate being affected by low income househ

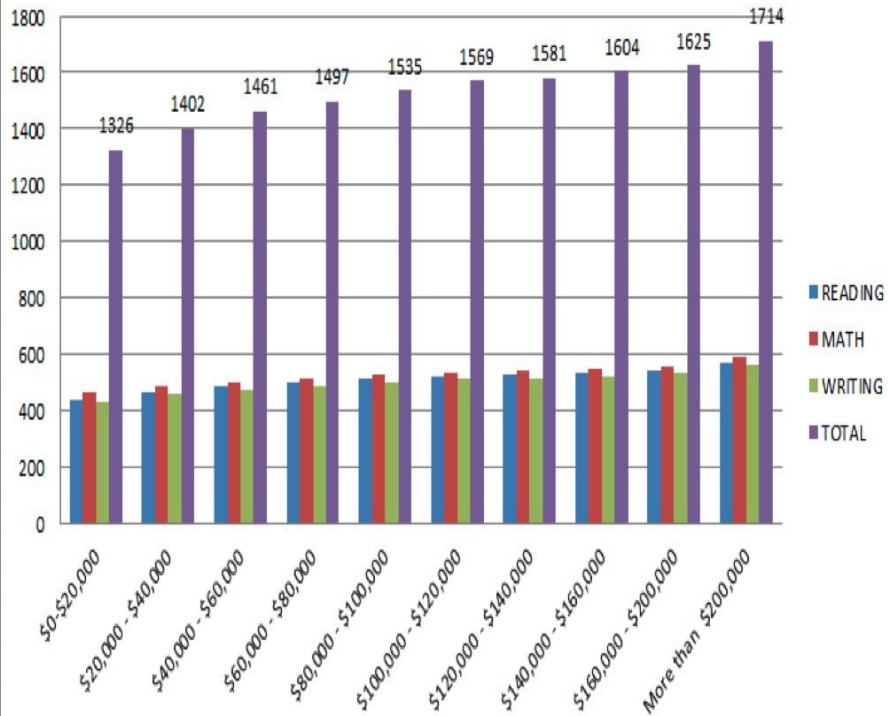
There is a positive correlation.

- Low income counties increases = decrease in participation



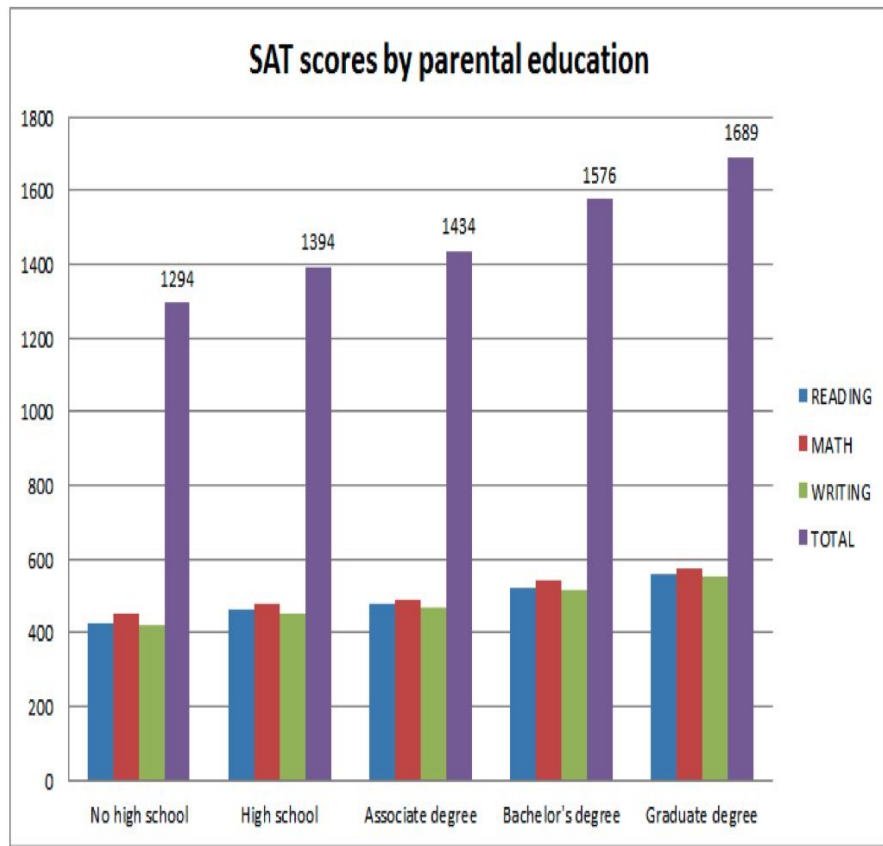
$$y = \beta_0 + \beta_1 x_1$$

**SAT scores by family income**



Parental income and socioeconomic status are correlated to academic achievement. (Dixon-Román, 2007; Rothstein, 2004)

Research shows that parental income has a substantial effect on academic achievement and accounts for a meaningful proportion of the score gap between high income families and the low income families.



Parental education is correlated to academic achievement. (Mattern, Shaw, & Williams, 2008; Zwick, Brown, & Sklar, 2004)

Research shows that parental education has a significant effect on academic achievement and accounts for a meaningful proportion of the score gap between high income families and the low income families.



# Strategies that work to increase SAT participation

```
1 sat_2019[sat_2019['State'].isin(['Delaware', 'Idaho'])]
```

	State	Participation Rate	EBRW	Math	Total
7	Delaware	100%	499	486	985
12	Idaho	100%	505	488	993

1. Offering SAT at no cost
2. Free access to the College Board's official SAT online course
3. Conducting SAT exam during regular school hours with transportation provided.
4. Engagement with the school counselor

# Conclusion and Recommendations

## Conclusion:

The most direct methods in increasing the SAT participation rate could be making the SAT test mandatory statewide and offering fee waiver for SAT exam. However, it is quite common for fee waiver recipient to be absent during the test day, mainly due to lack of confidence and lack of transport.

Hence, our recommendation to increase SAT participation rate are listed below:

1. Offering fee waiver
2. Offering SAT exam during regular school hours with transport provided
3. Regular engagement between school counselor and test takers

# Reference:

Camara, W. J. & Schmidt, A. E. (1999). Group differences in standardized testing and social stratification. College Board Report (99)5. New York, NY: College Board.

Zwick, R. (2004). Is the SAT a "Wealth Test"? The link between educational achievement and socioeconomic status (pp. 203–216). In R. Zwick, (Ed.). Rethinking the SAT in university admissions. New York, NY: Routledge Falmer.

"Free SAT registration now open for public school juniors." Delaware PTA, February 10, 2014.

<http://delawarepta.org/free-sat-registration-now-open-for-public-school-juniors/>

Adams, C. "Delaware Gives All Juniors SAT During School Day." *Education Week*, January 25, 2011.

[http://blogs.edweek.org/edweek/college\\_bound/2011/01/all\\_high\\_school\\_juniors\\_in\\_delaware\\_take\\_sat\\_starting\\_in\\_april.html](http://blogs.edweek.org/edweek/college_bound/2011/01/all_high_school_juniors_in_delaware_take_sat_starting_in_april.html)

"SAT." Idaho State Department of Education. <http://www.sde.idaho.gov/site/assessment/SATstudentParent.htm>