Socioeconomic Demographics & Standardized Test Performance

Elizabeth Mackenzie, Yongye Tan, Maekala Turner

Inquiry Development & Data Collection

Continued social and political campaigns against equity practices (Affirmative Action) have only widened the **Achievement Gap** between low-income students and their more affluent peers.

Web Scraping

Scrape median household income, population, and household on a HTML table in this wikipedia page

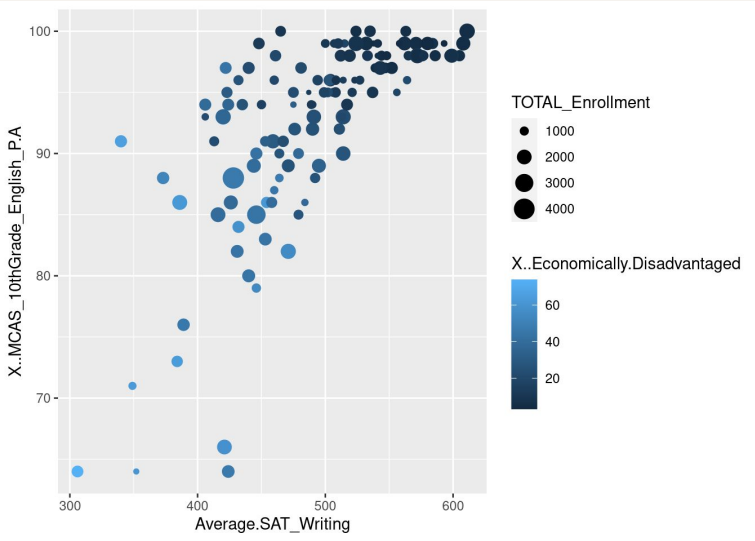
More income -> Higher Score in exam (Is this true?)

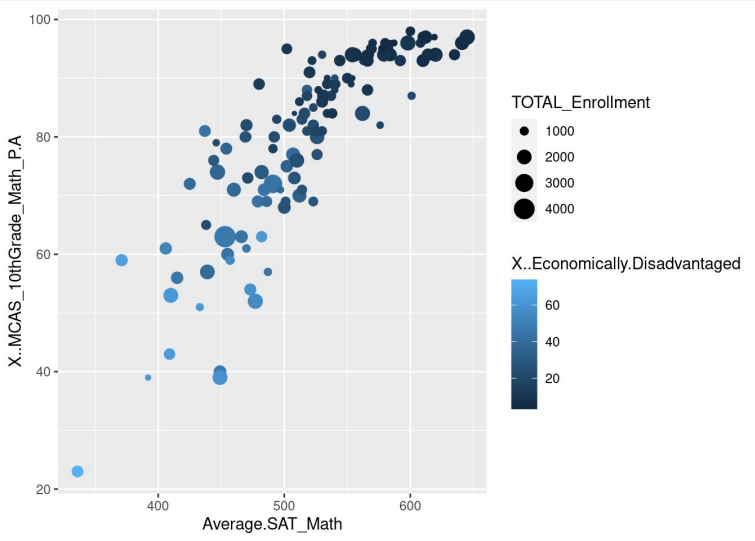
Rank	Municipality	Type	County	Per capita income	Median household income	Median family income	Households	Population
1	Dover	Town	Norfolk	$133,500.00	$250,000+	$250,000+	1,916	5,894
2	Weston	Town	Middlesex	$107,793.00	$220,815.00	$250,000+	3,731	11,806
3	Wellesley	Town	Norfolk	$97,262.00	$226,250.00	$250,000+	8,668	29,365
4	Lexington	Town	Middlesex	$96,170.00	$202,852.00	$235,039.00	12,301	34,235
5	Sherborn	Town	Middlesex	$96,081.00	$218,906.00	$242,443.00	1,515	4,406

Objectives & Interests & Expectations

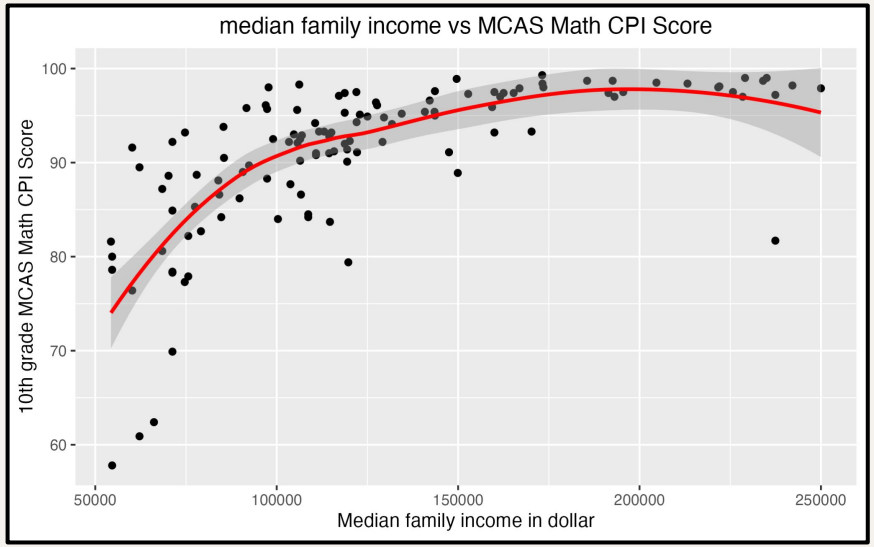
-10th grade English and MCAS CPI
-Average SAT Math and Writing scores
-Ratio of passed AP exams to student body size
-Median family income
-Proportion of high needs students
-Graduation rate

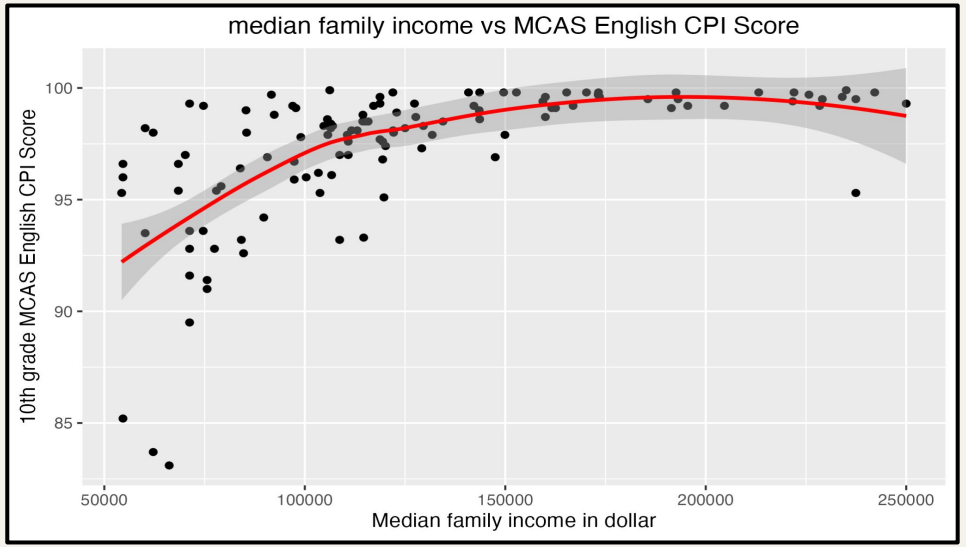
Objectives & Interests & Expectations





Regression Visualization

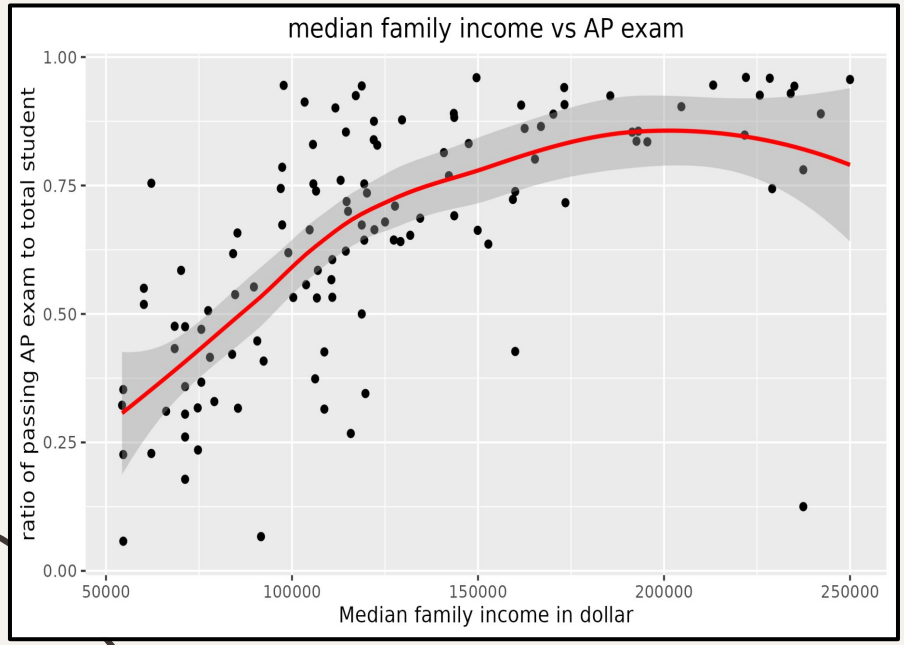


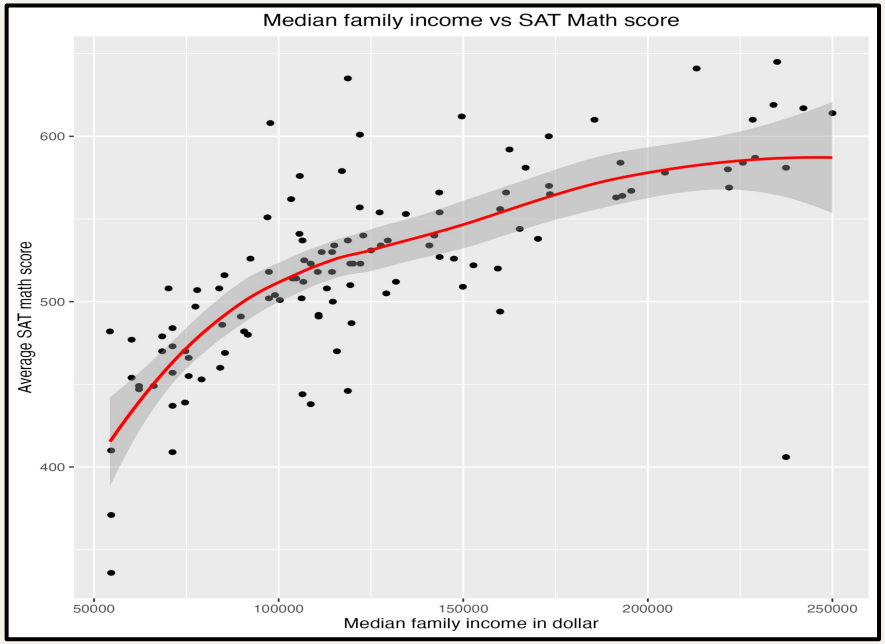


As median family income increases, CPI scores grow logarithmically.

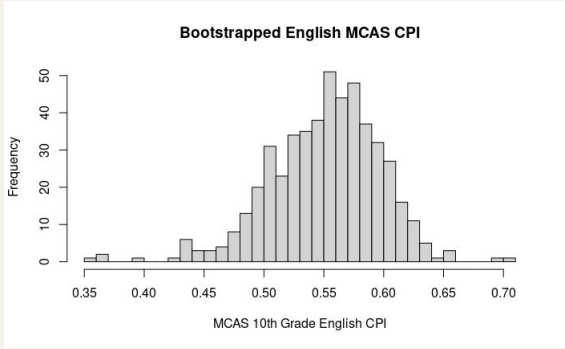
A t-test regression analysis on these two distinct variables, and the p-values are $1.321 * 10^{-10}$ and $4.087 * 10^{-14}$ for the subject in English and Math, respectively.

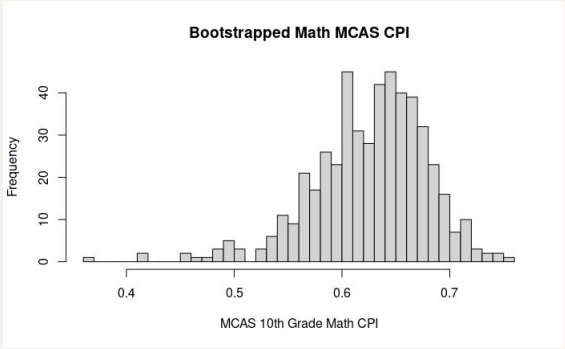
Income ↑ → Exam Scores ↑

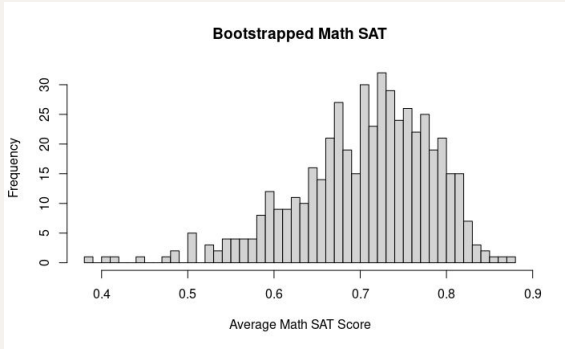


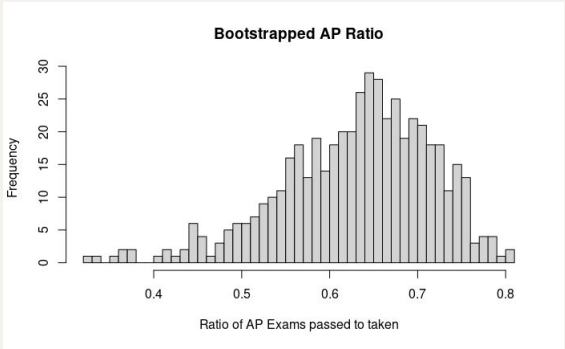


Bootstrap Testing





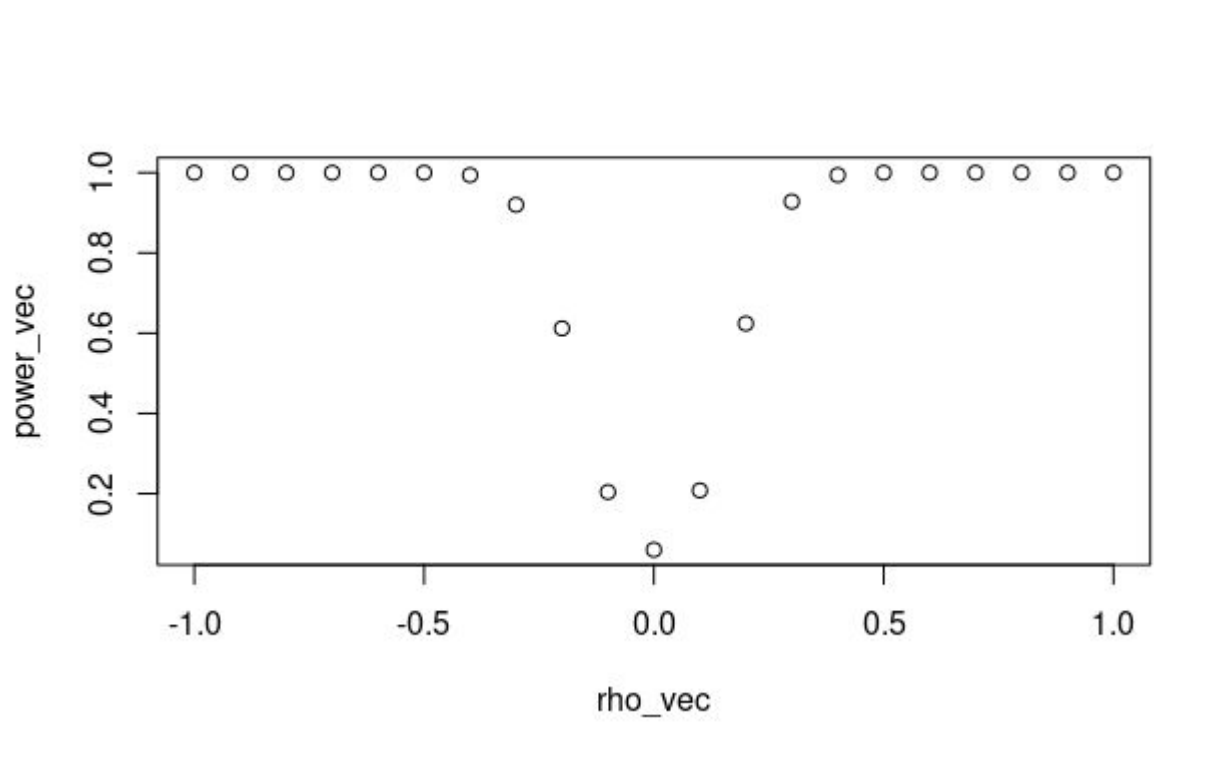


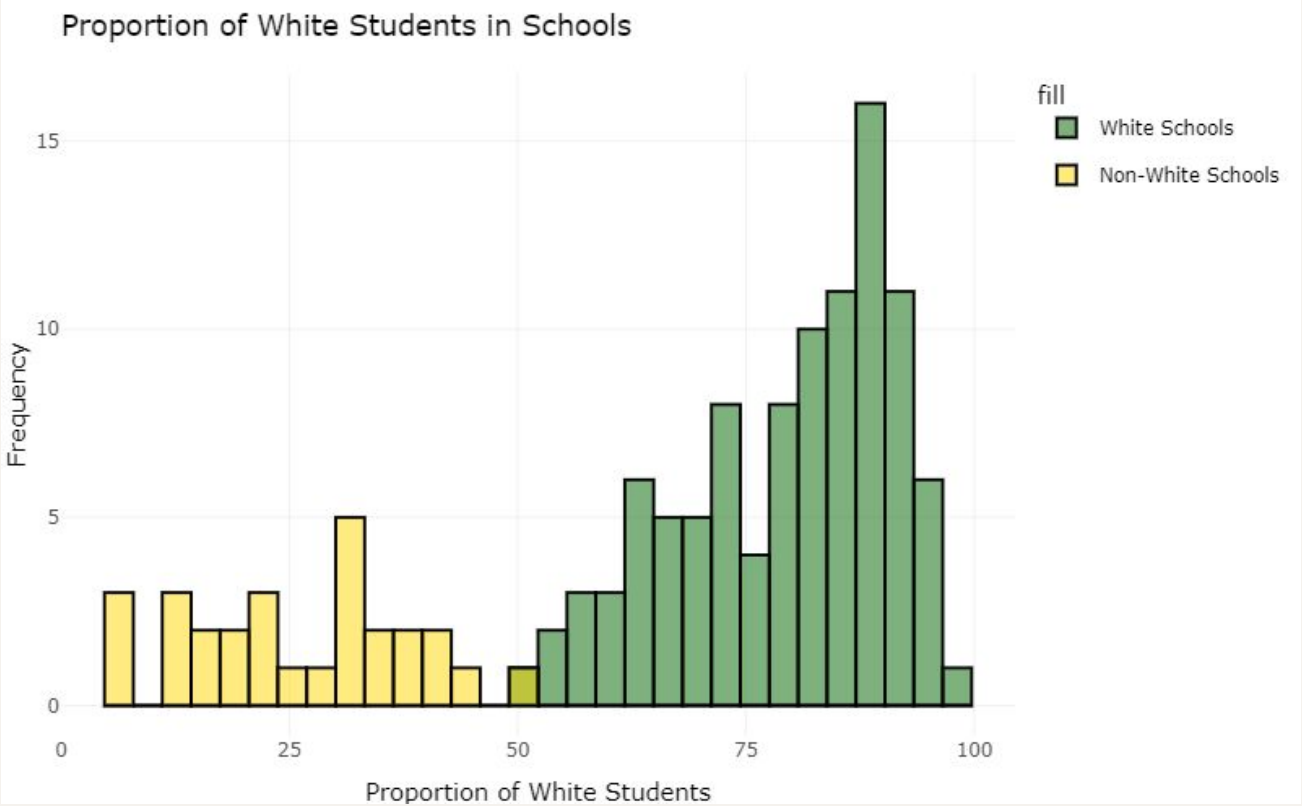


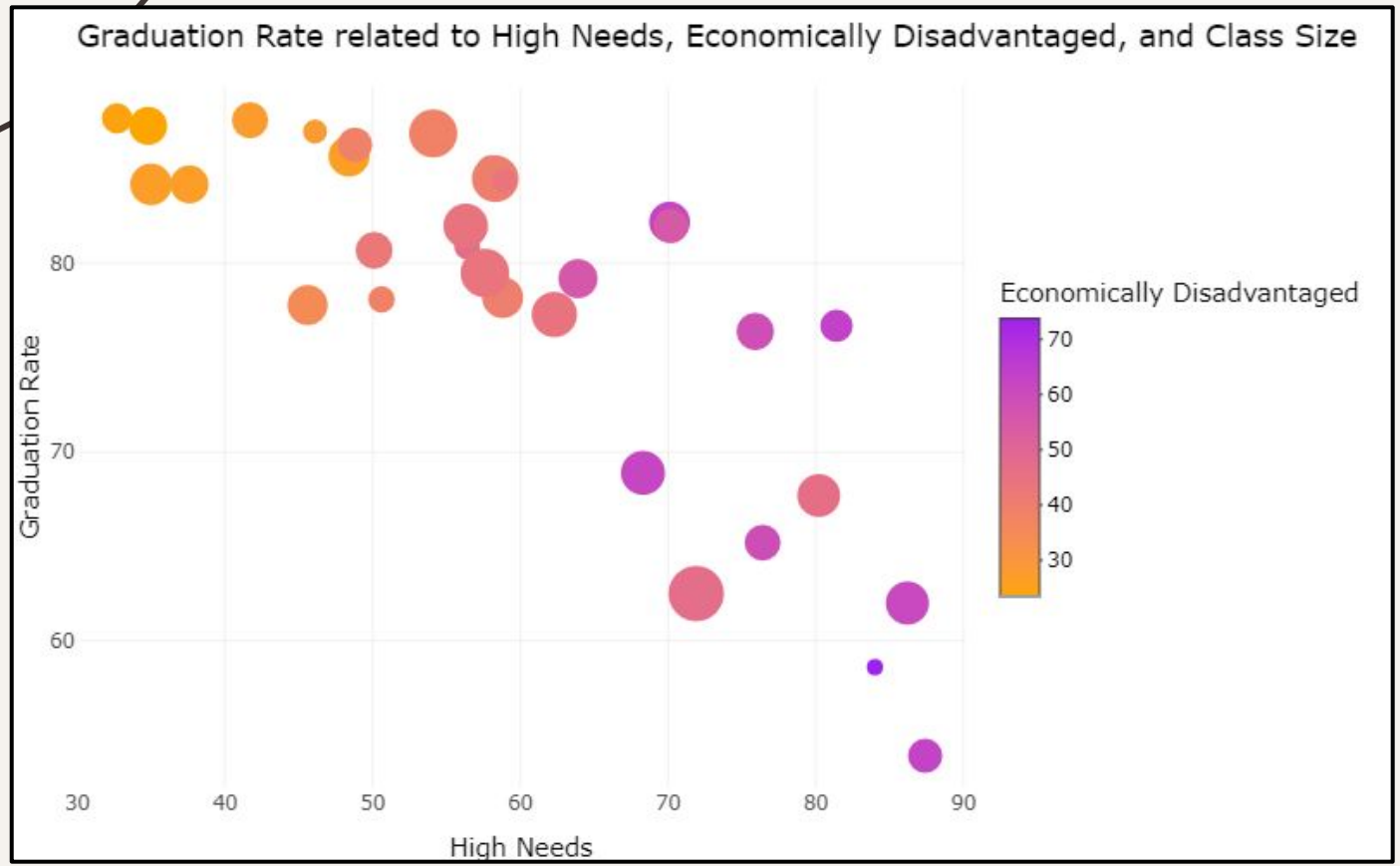
Monte Carlo Power Study

```r
fake_data_fn <- function(N = 125, rho = 0) {
  fake_data = mvrnorm(N, mu = c(125.7, 505), Sigma = matrix(c(50^2, rho*50*58, rho*50*58, 58^2), nrow = 2))
  return(fake_data)
}

fake_data_fn()

cor_test_fn <- function(df, alpha = 0.05) {
  cor_test_result <- cor.test(df[,1],df[,2], conf.level = 1 - alpha)
  return(cor_test_result$p.value < alpha)
}

cor_test_fn(fake_data_fn())

power_test_fn <- function(S = 500, N = 125, rho = 0, alpha = 0.05) {
  dfs <- replicate(S, cor_test_fn(fake_data_fn(N, rho), alpha))
  return(mean(dfs))
}

power_test_fn()


#Power Study

rho_vec <- seq(-1,1,0.1)

power_vec <- lapply(rho_vec, power_test_fn, S = 500, N = 125, alpha = 0.05)

plot(x = rho_vec, y = power_vec)
```

Monte Carlo Power Study

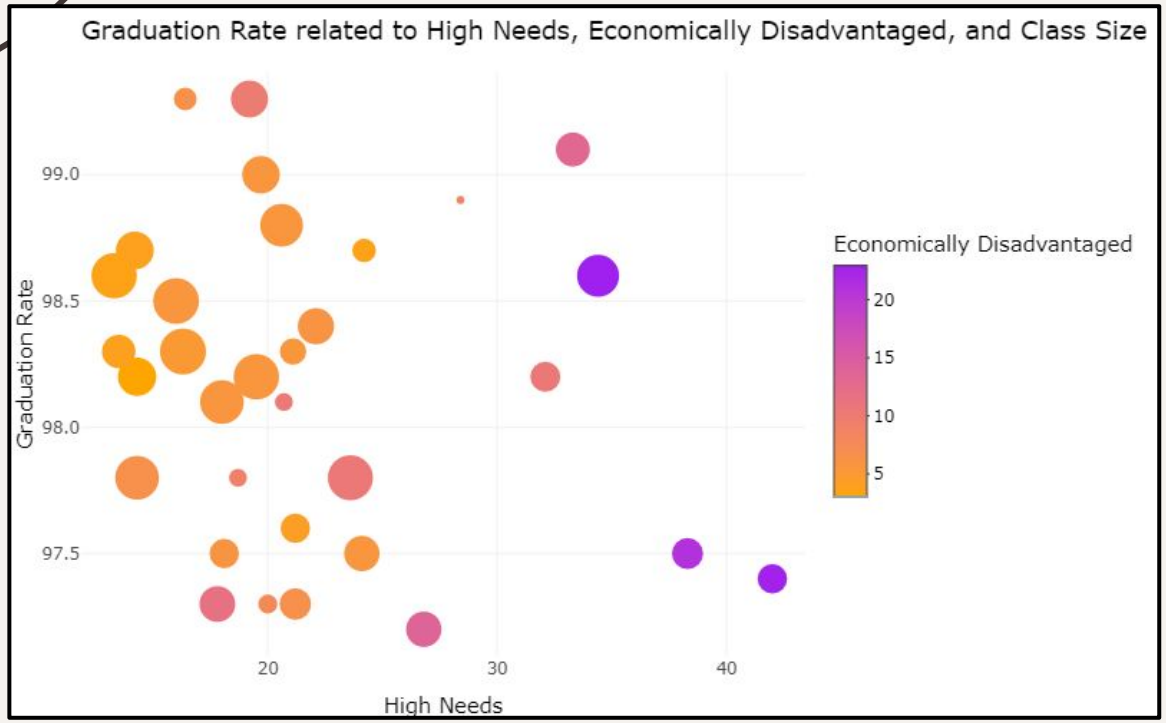



Proportion of White Students in Schools
fill
White Schools
Non-White Schools
Frequency
15
10
5
0
0
25
50
75
100
Proportion of White Students



Bottom Quartile

Regardless of class size, higher rates of high needs students tend to result in lower graduation rates. Schools with the highest graduation rates have the smallest proportions of economically disadvantaged students.



Top Quartile

No matter the class size, schools with above average graduation rates have much smaller proportions of high needs students. Further, we can see that the color gradient is not as smooth and the correlation is not as consistent as the bottom quartile of schools, implying that within the top quartile, high needs students are not necessarily the same students considered economically disadvantaged..

Overall Conclusions

Our analysis supports our hypothesis that lower mean family income is correlated with lower MCAS scores and SAT scores.

- A close relationship between a variety of other socioeconomic factors and academic performance.
- Massachusetts is not immune to the pervasive inequities within education that we most often see on a national scale

Questions?