



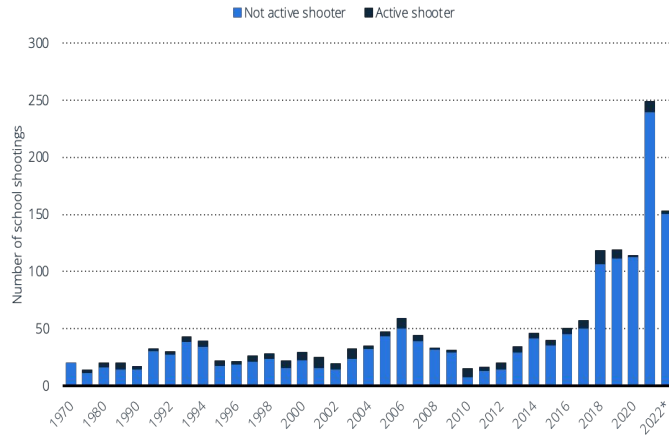
# On School Shootings & Academic Performance & Racial Diversity

**Computational Social Science  
Project Milestone 3**

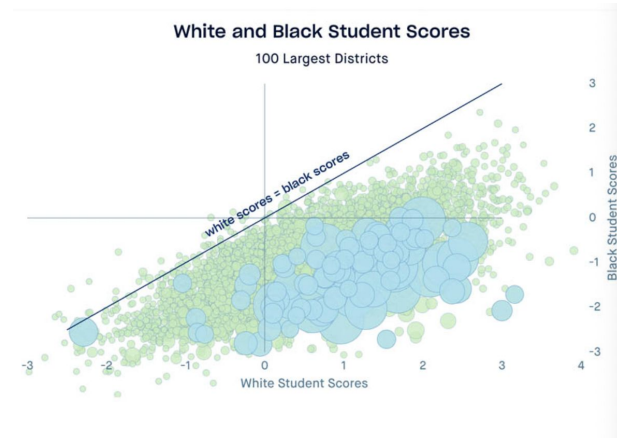
Alex Chae  
Kevishen Valeyatheepillay  
Reem Hazim  
Zachary Raye

# Research Question

“How is the occurrence of **school shootings** in the **US** related to **student’s academic performance**, and to what extent do these relations vary based on **racial composition** where the shootings occur?”



Number of K-12 school shootings U.S. 1970-2022, by active shooter status; Source: Statista



There's a large racial achievement gap in almost all of the US's 100 largest school districts. Image: Stanford University

# Motivation behind research question



## Research gap & urgency

- Current studies show detrimental impact, but relies on limited datasets
- Need for deeper understanding on bigger scale




## Scope of our research

- Determine if negative impact of these events is consistent across all educational institutions, specifically with regard to race



## Addressing educational inequalities

- Existing disparities in access to equal education
- Do these occurrences exacerbate these disparities



“School shootings have a significant **negative correlation** with **students’ academic performance** across national institutions, with a **more pronounced** negative association in schools where a higher proportion of non-white students are enrolled compared to predominantly white institutions.”

## **HYPOTHESIS**

Introduction

Matching

Experiment

Regression

Conclusion

# Stages of Study

Academic  
Performance  
Matching

Enrollment Matching:  
Are scores of students  
enrolled in school after  
shooting affected?

Regression

Introduction

Matching

Experiment

Regression

Conclusion

# Merging and Filtering the data

- We had to **merge** three datasets:
  - Washington Post school shootings dataset
  - U.S. Public School Math and Reading Test Score dataset from US DoE
    - Data about proficiency rates from 2009 to 2021
  - Poverty rate data from U.S. Census Bureau
- We had to **filter** our school shootings dataset:
  - Our shootings dataset originally had 387 schools
  - Our test score data only covered the years 2009 - 2021, so after filtering for shootings in those years, we were left with only 208 schools
  - Many of the shooting schools did not have test score data for the year of the shooting. After filtering those out, we were left with 105 schools

# What Does Our Data Look Like?



School  
ID



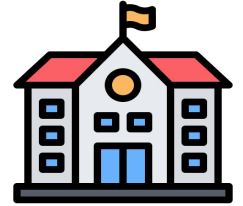
Grade  
Level



Math and  
Reading  
Proficiency  
Rates



Dates



Shooting  
(T/F)

# Data Cleaning

- In small schools where very few students took the state exam, the percentage proficiency rate was reported as an interval rather than a precise number to protect the privacy of the students

Number of Students Reported in the Cell	Ranges Used for Reporting the Percent Proficient and Percent Participation for that Group
6-15	<50%, ≥50%
16-30	≤20%, 21-39%, 40-59%, 60-79% ≥80%
31-60	≤10%, 11-19%, 20-29%, 30-39%, 40-49%, 50-59%, 60-69%, 70-79%, 80-89%, ≥90%
61-300	≤5%, 6-9%, 10-14%, 15-19%, 20-24%, 24-29%, 30-34%, 35-39%, 40-44%, 45-49%, 50-54%, 55-59%, 60-64%, 65-69%, 70-74%, 75-79%, 80-84%, 85-89%, 90-94%, ≥95%
More than 300	≤1%, 2%, 3%, . . . , 98%, ≥99%

- We used the average of the interval in our calculations

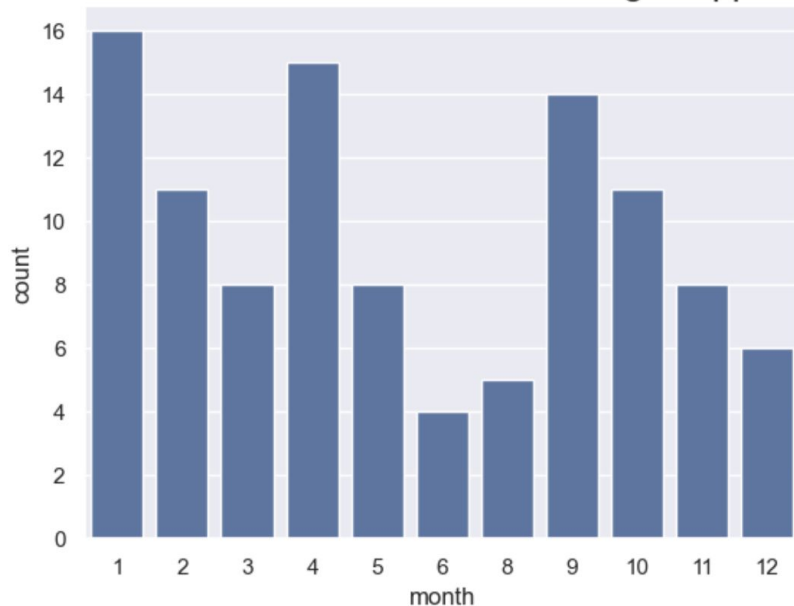


# **Descriptive Statistics**

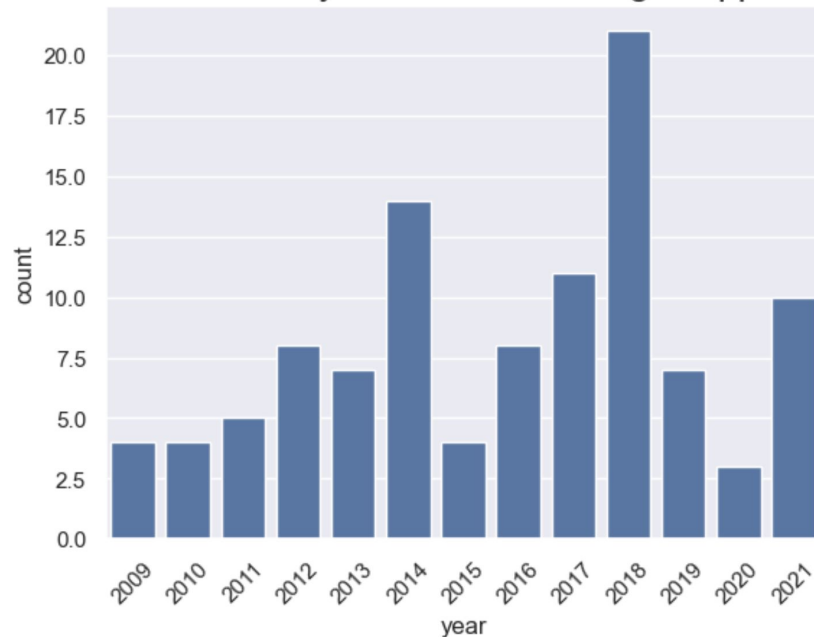


# Data Fast Facts

Distribution of months when shootings happened



Distribution of years when shootings happened



Introduction

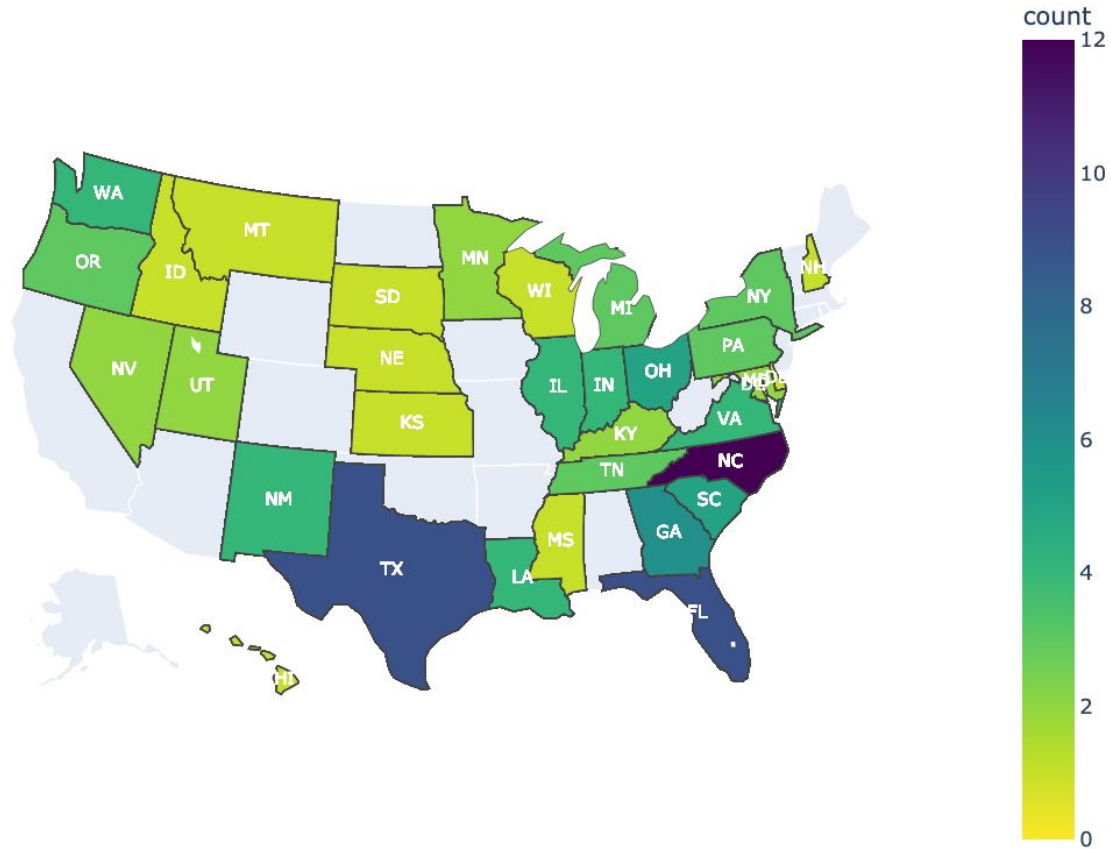
Matching

Experiment

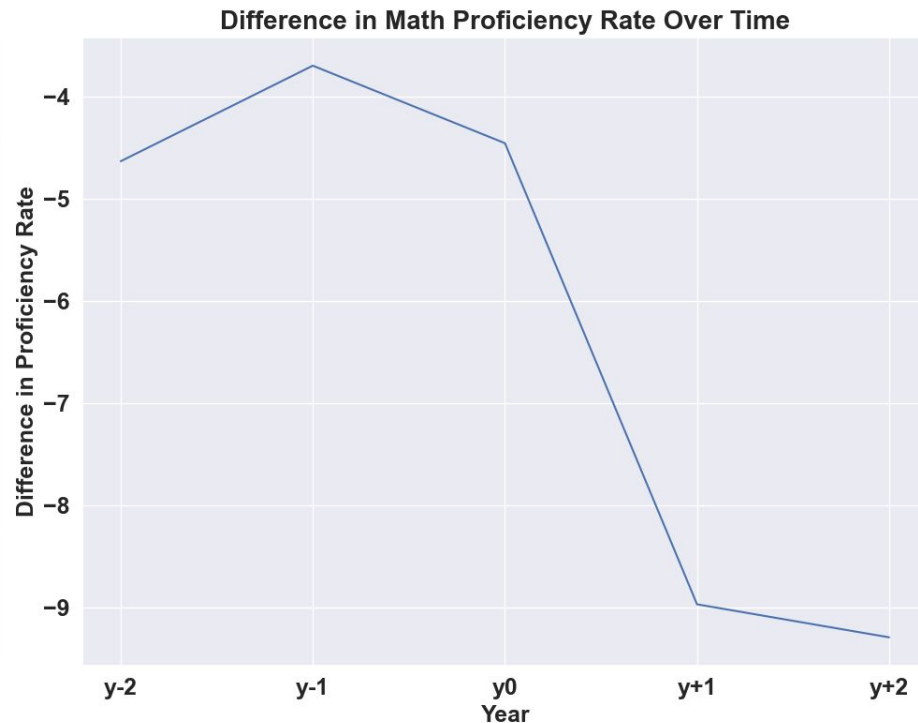
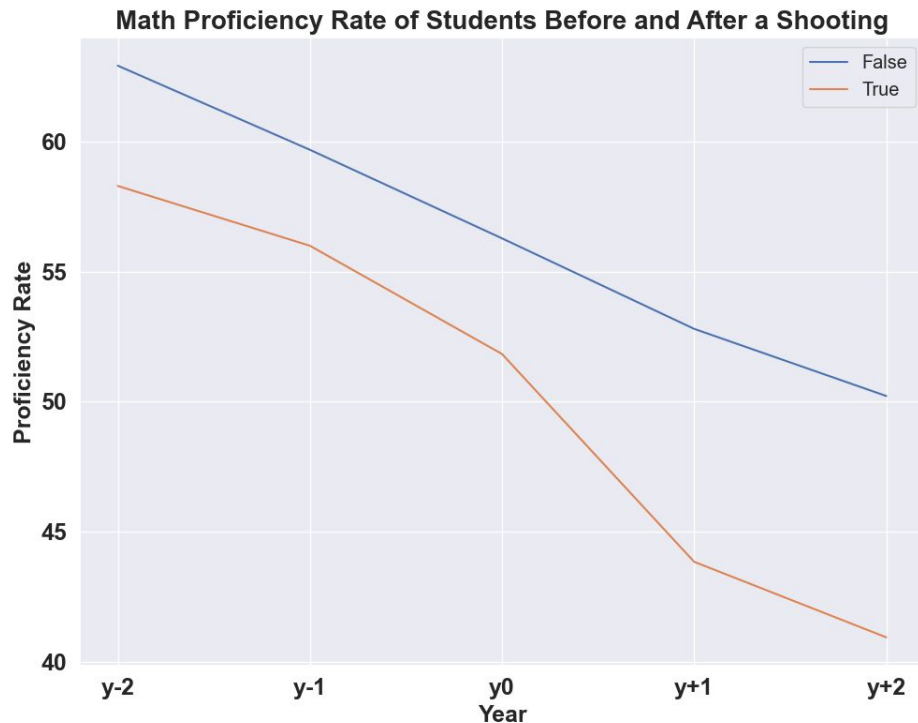
Regression

Conclusion

## Distribution of Shootings over States



# Math Percentage Proficiency Over Time



Introduction

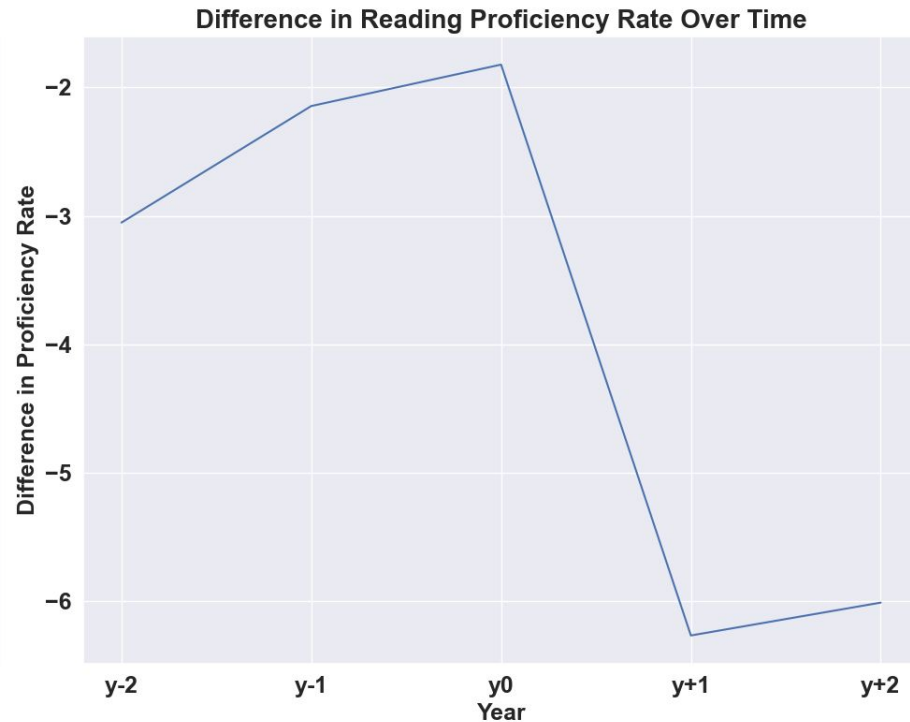
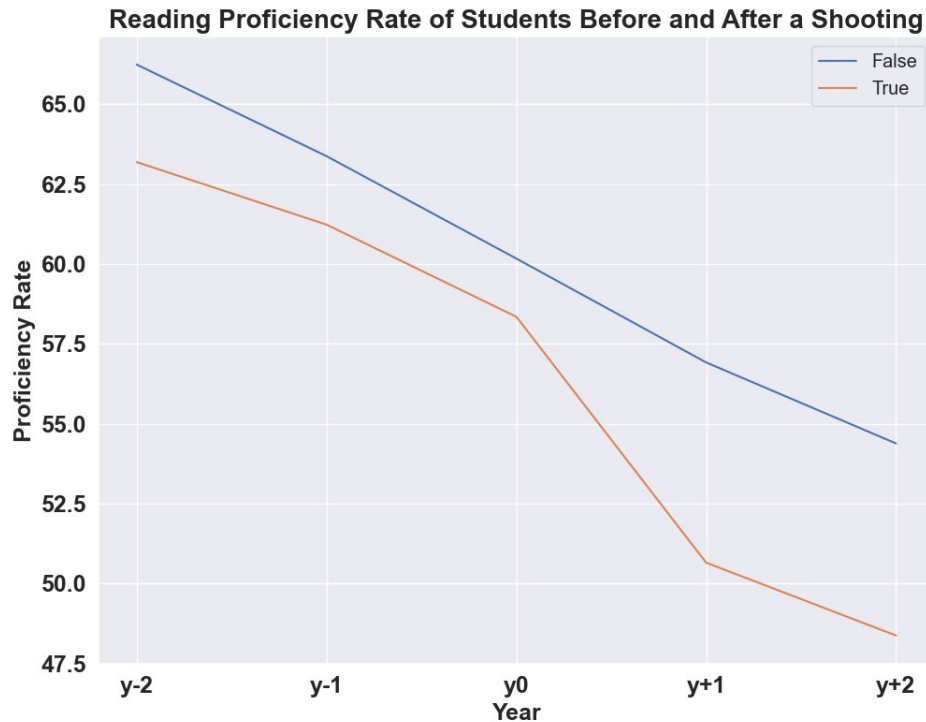
Matching

Experiment

Regression

Conclusion

# RLA Percentage Proficiency Over Time



Introduction

Matching

Experiment

Regression

Conclusion

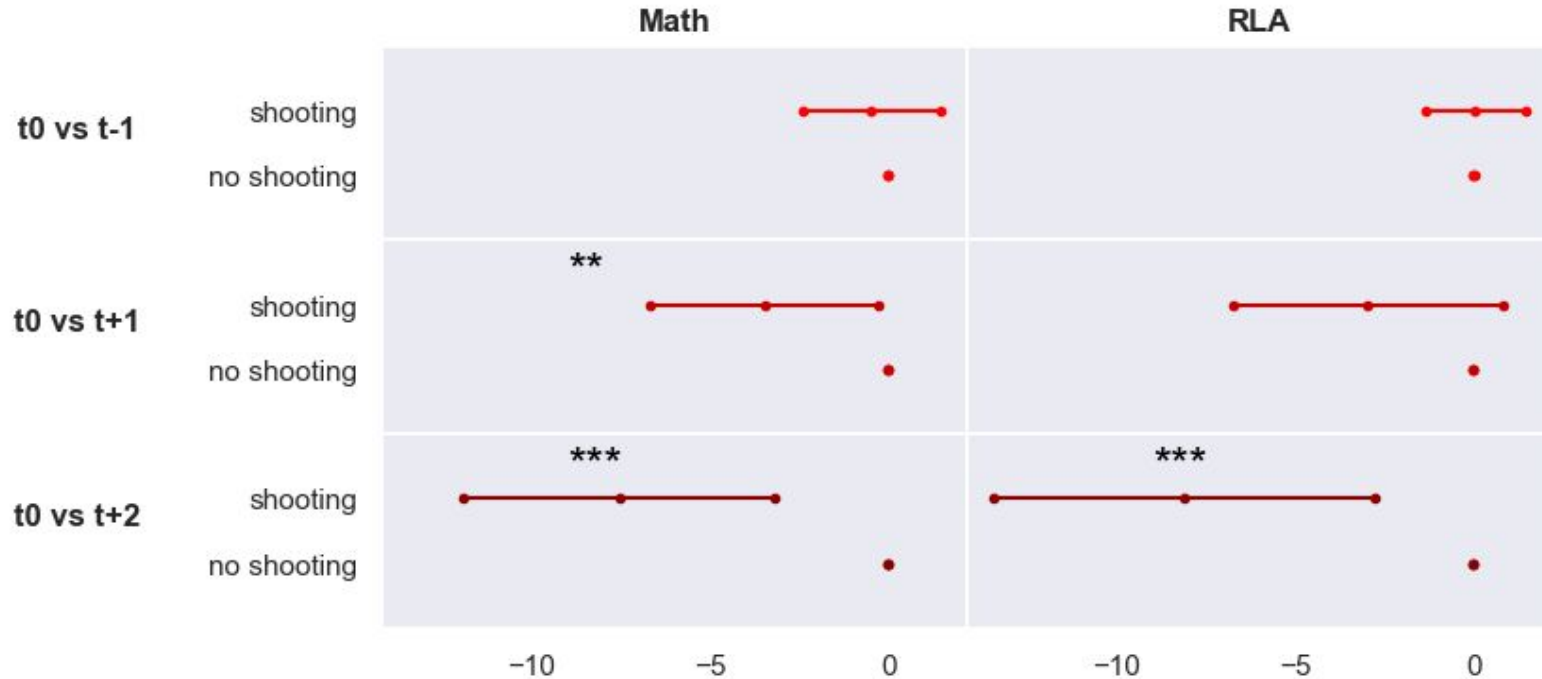
# **The Correlation Between School Shooting and Academic Performance**



# Matching using CEM

- Filter for all the schools that had test score data for
  - The year before the shooting ( $t - 1$ )
  - The year of the shooting ( $t_0$ )
  - The year after the shooting ( $t + 1$ )
  - Two years after the shooting ( $t+2$ )
- We were left with 43 schools where shooting happened
- We binned the schools using the following variables:
  - Test scores from the previous year (similar quality of students)
  - Grade
  - School year
  - District ID
  - Number of test-takers
- Randomly sample some of the control schools in each bin
  - 1 sample (1-to-1 matching)
  - No sample (use entire control population in that bin)

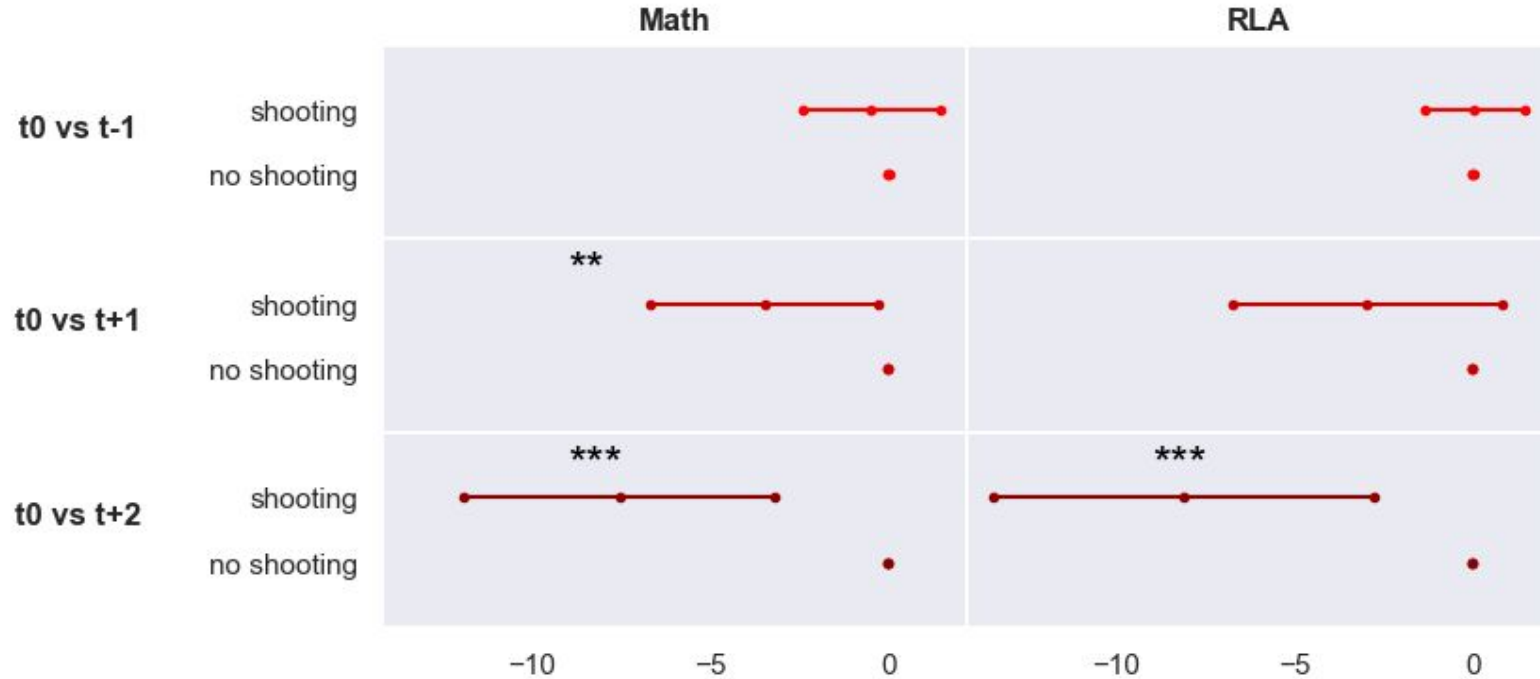
# Results (no sampling)



(\*)  $p < 0.1$ , (\*\*)  $p < 0.05$ , (\*\*\*)  $p < 0.01$



# Results (1 sample)



(\*)  $p < 0.1$ , (\*\*)  $p < 0.05$ , (\*\*\*)  $p < 0.01$

# The Correlation Between School Shootings and School Quality



# Possible explanations for the negative influence of shooting:

## Decrease in school quality

Shooting is associated with a decrease in the school educational quality

## Student mental trauma

Shooting influences mental state of students who experienced shooting

Introduction

Matching

Experiment

Regression

Conclusion

# Are the next batches of students also affected?

## YES!

Shooting changed the trend in student performance in the school even after affected students graduated



School quality ↓

## NO!

Shooting only affected students who experienced shooting & New batches were not affected



Student trauma

Introduction

Matching

Experiment

Regression

Conclusion

# Matching Experiment:

→ Compare academic performance of students right before shooting to that of students that enrolled after the shooting incident

## Elementary Schools

3rd graders right before shooting  
&  
3rd graders 3 years after shooting

## High Schools

Students right before shooting  
&  
Students 4 years after shooting

## Student's t-test

Introduction

Matching

Experiment

Regression

Conclusion

# High Schools

Filters applied:

1. High school
2. Shooting between 2009 - 2017
3. Matched with proficiency rate dataset

Sample size:  $n = 41$  (RLA),  $n = 39$  (Math)

Introduction

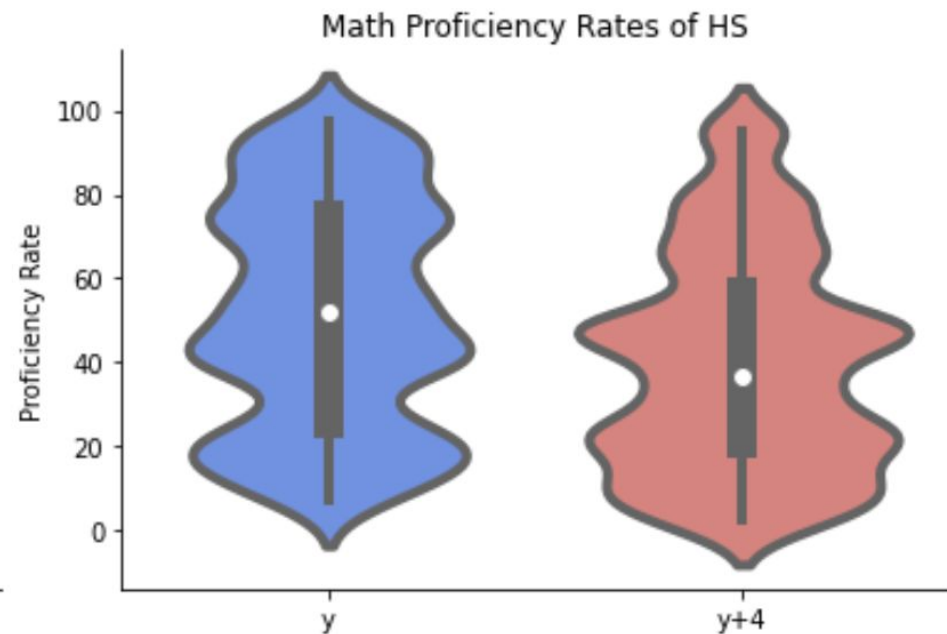
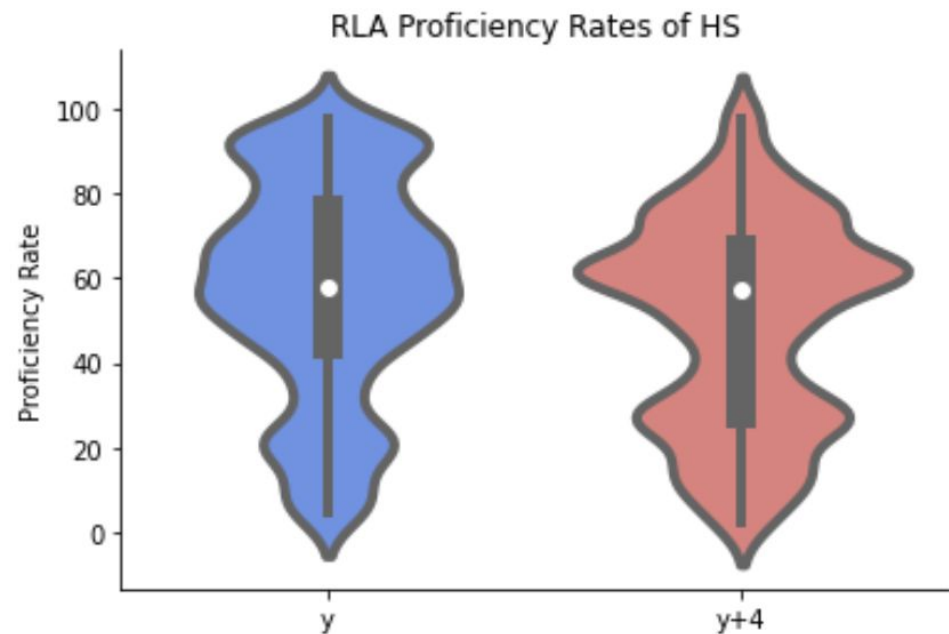
Matching

Experiment

Regression

Conclusion

# High Schools



Introduction

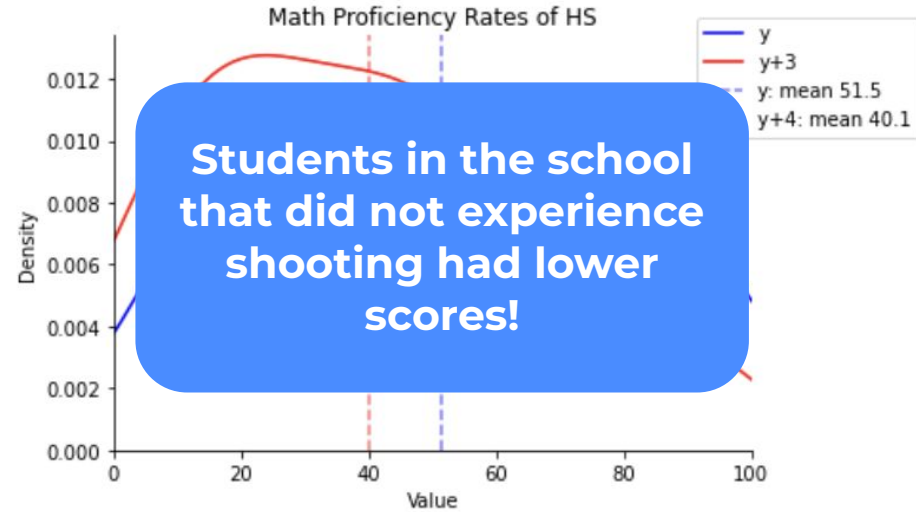
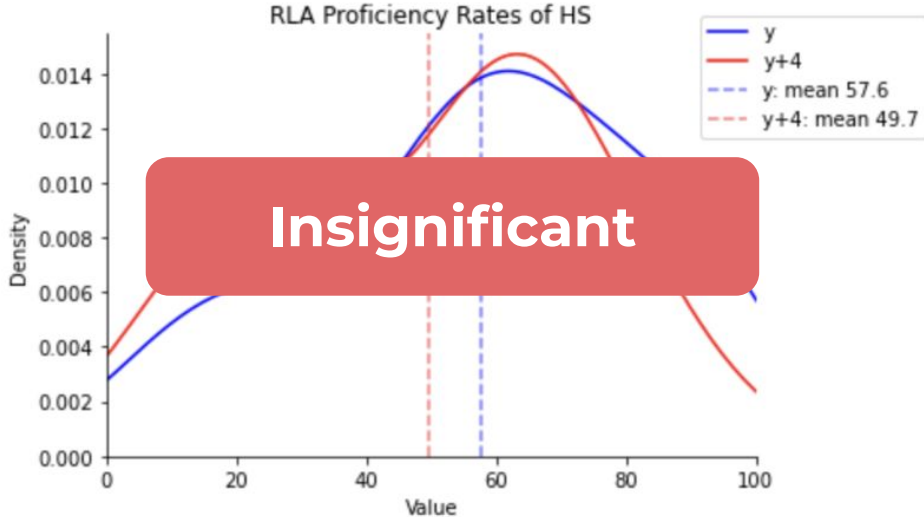
Matching

Experiment

Regression

Conclusion

# High Schools



Introduction

Matching

Experiment

Regression

Conclusion



# High Schools - Confounders?

## Poverty rate

**“Poverty directly affects academic achievement** due to the lack of resources available for student success” (Lacour & Tissington, 2011)

**“Poverty influences the academic performance** of students in Enugu East LGA in terms of classroom teaching and learning” (Chioke, 2021)

## Crime rate

**“Greater numbers of crimes** proximal to school buildings relate to lower levels of academic performance” (Boxer et al, 2020)

**“Exposure to crime has a negative and significant impact** in academic achievement” (Gimenez & Barrado, 2020)

Introduction

Matching

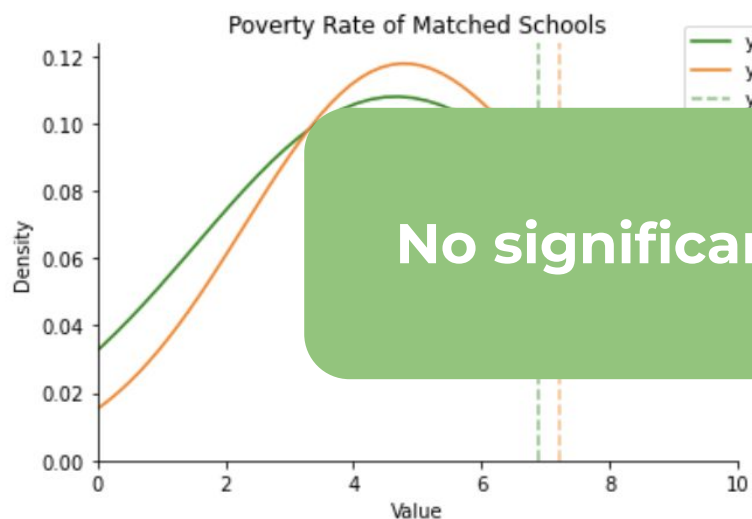
Experiment

Regression

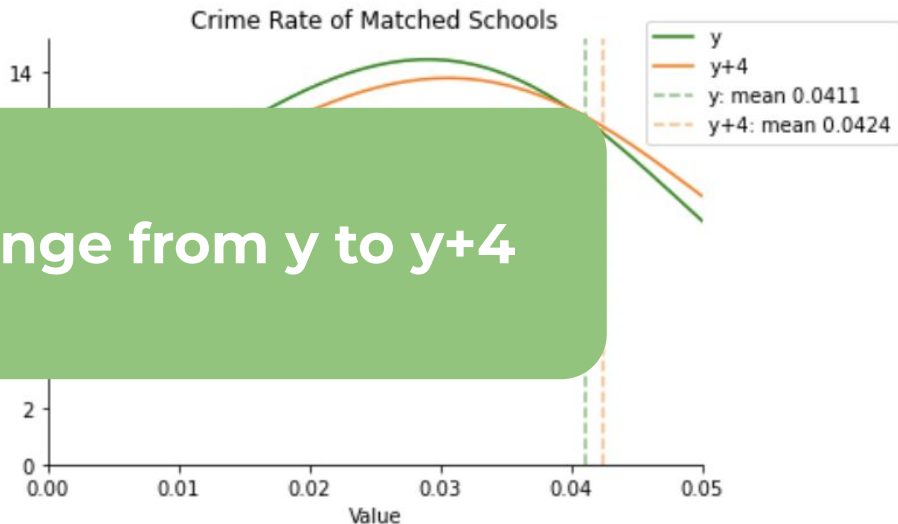
Conclusion

# High Schools - Confounders?

Paired t-test



t-statistic:  $-0.42373454191963994$   
p-value:  $0.6741484247722433$



t-statistic:  $-1.6799349942610122$   
p-value:  $0.10117448338261942$

Introduction

Matching

Experiment

Regression

Conclusion

# Elementary Schools

Filters applied:

1. Elementary school
2. Shooting between 2009 - 2018
3. Matched with proficiency rate dataset

Sample size: **n= 9 (RLA), n= 8 (Math)**

Introduction

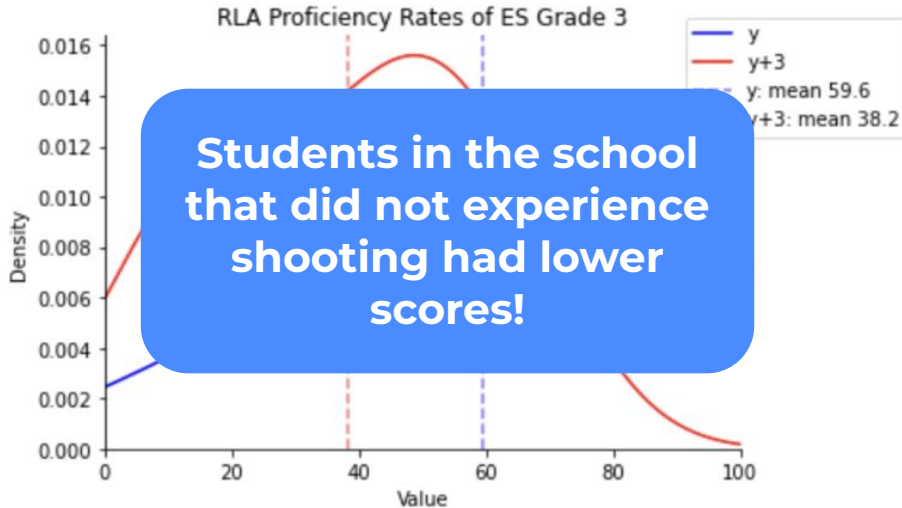
Matching

Experiment

Regression

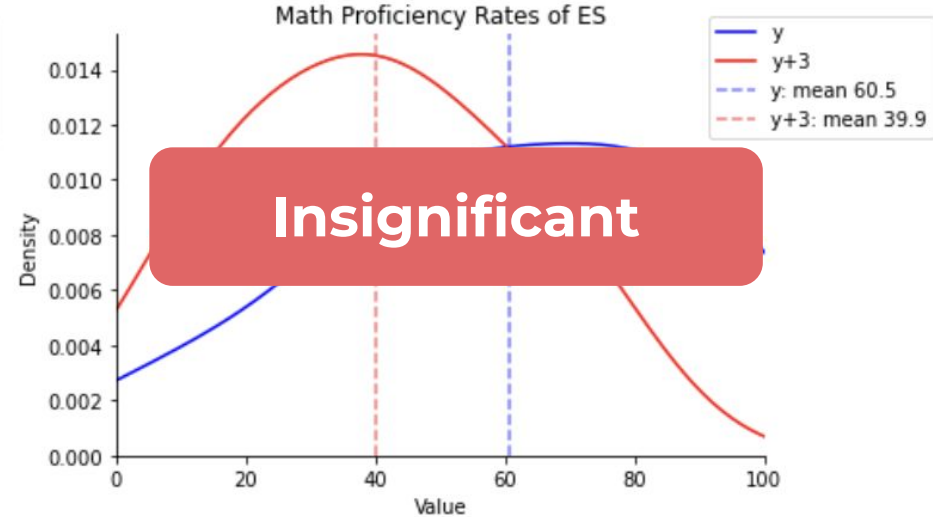
Conclusion

# Elementary Schools



t-statistic: 1.907448768545175

p-value: 0.07457747027217694



t-statistic: 1.6360412047446136

p-value: 0.12410575832438761

Introduction

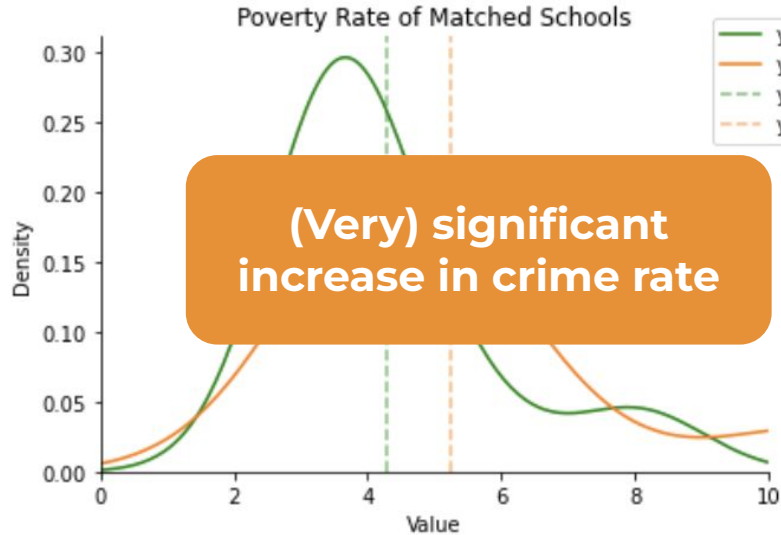
Matching

Experiment

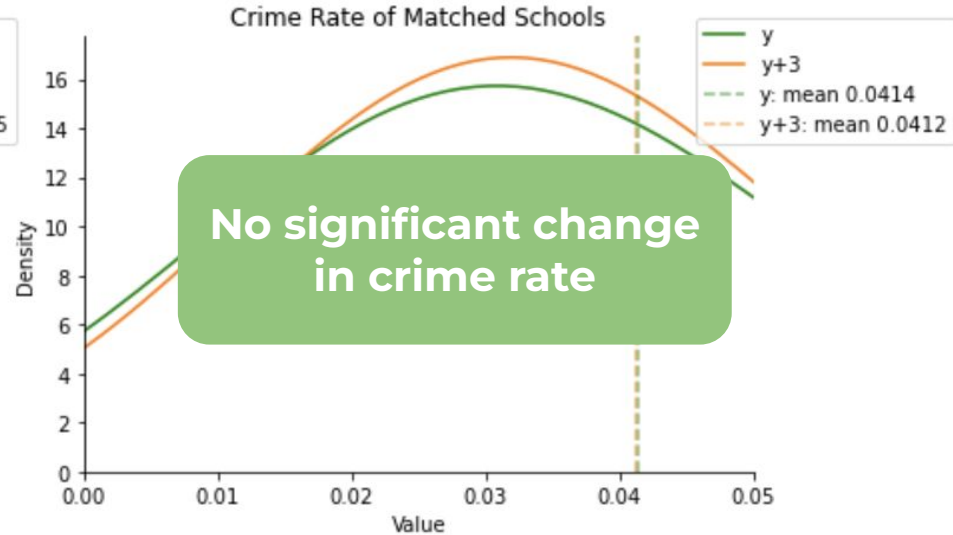
Regression

Conclusion

# Elementary Schools - Confounders?



t-statistic: -3.9379035257397716  
p-value: 0.004308564993170008



t-statistic: 0.21530729785398622  
p-value: 0.8349154153109491

Introduction

Matching

Experiment

Regression

Conclusion

# Matching Experiment:

Elementary schools:

- ❖ **Inconclusive** due to small  $n$  & confounding factors

High schools:

- ❖ New batches of students after shooting incident have lower Math performance than previous students
- ❖ Suggests that shootings might have **decreased the education quality in affected schools**

Introduction

Matching

Experiment

Regression

Conclusion

# Regression



Does the racial distribution of students determine how much schools are affected by school shootings?

# Regression

$$\begin{aligned} \text{Test Scores} = & \alpha + \beta_1 \cdot (\text{After}) + \beta_2 \cdot (\%NonWhite) + \beta_3 (\text{After}) \cdot (\%NonWhite) \\ & + \beta_4 \cdot (\text{SchoolDistrict}) + \xi \end{aligned}$$

- Test Scores: Mean RLA & Math test scores for every school per year
- Population: Schools with *Shooting* = 1 only
- *After* = 1 for period after shootings for schools that had shootings (dummy)
- *%NonWhite*: Proportion of non-white students in schools per year
- *SchoolDistrict* FE: Dummy for school district → Control for state-inherent characteristics



# Regression

$$\begin{aligned} \text{Test Scores} = & \alpha + \beta_1 \cdot (\text{After}) + \beta_2 \cdot (\%NonWhite) + \beta_3 (\text{After}) \cdot (\%NonWhite) \\ & + \beta_4 \cdot (\text{SchoolDistrict}) + \xi \end{aligned}$$

- $\beta_2$ : Effect of higher racial diversity on test scores prior to school shooting
- $\beta_3$ : Effect of higher racial diversity on test scores POST school shooting

# Regression

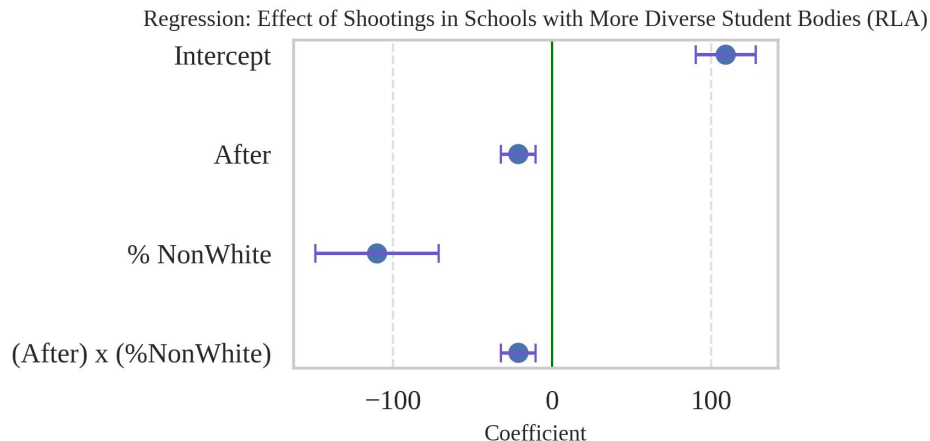
$$\text{Mean RLA Scores} = \alpha + \beta_1 \cdot (\text{After}) + \beta_2 \cdot (\% \text{NonWhite}) + \beta_3 (\text{After}) \cdot (\% \text{NonWhite}) + \beta_4 \cdot (\text{SchoolDistrict}) + \xi$$

Results:

R-squared: 0.914  
Adj. R-squared: 0.833  
F-statistic: 11.36  
Prob (F-statistic): 3.27e-21

**Significant to 1% level**

	coef	std err	t	P> t
const	109.2681	9.478	11.529	0.000
after	-21.1896	5.511	-3.845	0.000
per nonwh	-110.0575	19.496	-5.645	0.000
after_per_nonwh	-21.1349	5.497	-3.845	0.000



- Post shootings, 1% increase in the % of non-white students lead to an incremental decrease in test scores of -21.14 units

Introduction

Matching

Experiment

Regression

Conclusion

# Regression

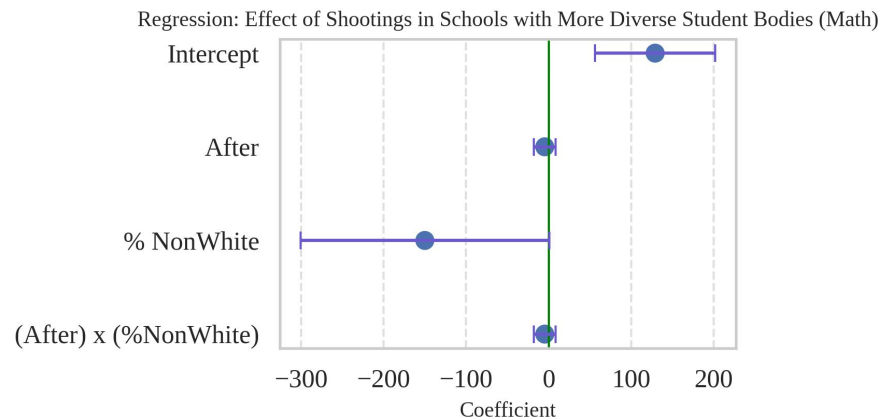
$$\text{Mean Math Scores} = \alpha + \beta_1 \cdot (\text{After}) + \beta_2 \cdot (\% \text{NonWhite}) + \beta_3 (\text{After}) \cdot (\% \text{NonWhite}) + \beta_4 \cdot (\text{SchoolDistrict}) + \xi$$

Results:

R-squared: 0.844  
Adj. R-squared: 0.733  
F-statistic: 7.591  
Prob (F-statistic): 3.33e-19

**Not Significant**

	coef	std err	t	P> t
const	129.0891	36.681	3.519	0.001
after	-4.7406	6.587	-0.720	0.473
per_nonwh	-150.1292	75.930	-1.977	0.051
after_per_nonwh	-4.7284	6.570	-0.720	0.473



- Quite surprisingly, racial Distribution does **not** affect Math scores after the treatment for schools that had shootings
- Caution: No significant coefficients

Introduction

Matching

Experiment

Regression

Conclusion

# Main Findings & Conclusions

## **Coarsened Exact Matching:**

Schools with shooting had a larger decrease in test scores than schools in the same district without shooting

## **Matching Experiment:**

Math performance of high school students who joined the school after shooting are lower than the performance of students in the school right before shooting

## **Regression:**

More racially diverse schools tend to be more influenced by school shootings for academic performance in RLA.

Introduction

Matching

Experiment

Regression

Conclusion

# Implications & Limitations

- Implications:
  - First study to look at school shootings across the US
  - Inform policy about how patterns arise after school shootings. Example: How grades are affected through time and how more diverse schools are the most impacted
- Limitations:
  - We only had **105 schools left after merging** datasets & **43 schools** used for CEM
  - In our matching experiment, we had very **small sample sizes** (n=8 for elementary schools and n=39 for high schools) which might lead cause **biased samples**.
  - We assume that in academic years during which a shooting happened, students took the **assessment test after the shooting** happened
  - Used a very simple regression model where controls were limited due to data availability. Might cause **endogeneity**

Introduction

Matching

Experiment

Regression

Conclusion

# Thank you :)

Alex Chae  
Kevishen Valeyatheepillay  
Reem Hazim  
Zachary Raye