



Examining the Influence of Socioeconomic Factors on Premature Death Rates Across Racial Groups

Tasnim Rida, Lissandro Alvarado, Oreoluwa Williams, Naomi James
TA Advisor: Princess Allotey



Introduction

- Social Determinants of Health: environmental conditions that affect health outcomes and quality of life
- Socioeconomic disparities and health
- Premature Deaths: deaths occurring before average age of 75 in the U.S, often preventable
 - **Examining premature death rates can highlight areas for healthcare improvement.**

**Do income inequality,
unemployment, high school
completion rates affect the number
of premature deaths of certain
racial groups at the county level?**

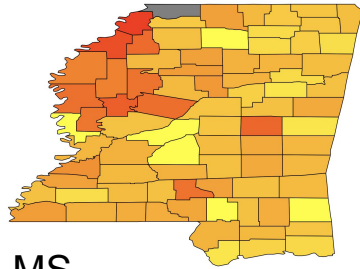
Data Overview

- Dataset ranks counties within each state based on various health outcomes and health factors
 - Source: County Health Rankings Data by the University of Wisconsin Population Health Institute
- Variables of Interest:
 - **Income Inequality:** Ratio of household income at the 80th percentile to income at the 20th percentile
 - **Unemployment:** Percentage of population ages 16 and older unemployed but seeking work
 - **High School Completion:** Percentage of adults ages 25 and over with a high school diploma or equivalent
 - **Percent Population:** Percentages of the population by county.
 - American Indian or Alaska Native (AIAN), Asian and Pacific Islander (AAPI), Non-Hispanic Black, Hispanic, Non-Hispanic White, Native Hawaiian or Other Pacific Islander (NHOPI), Non-Hispanic White
 - **Premature Death:** Years of potential life lost before age 75 per 100,000 population (age-adjusted)

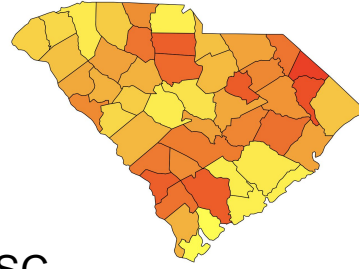
EXPLORATORY DATA ANALYSIS

States with higher level of poverty tend to have more years of premature death.

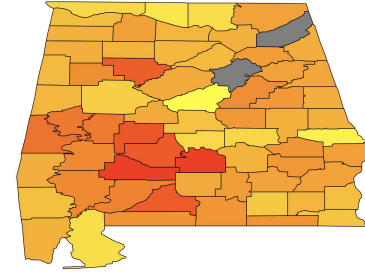
States with the highest premature deaths



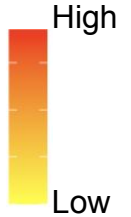
MS



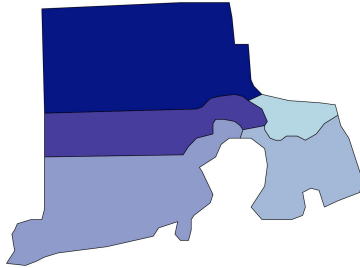
SC



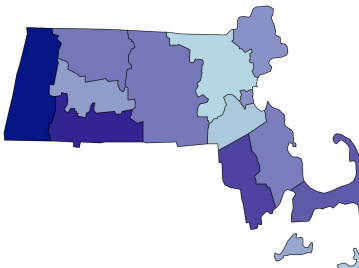
AL



States with the lowest premature deaths



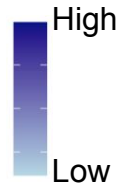
RI



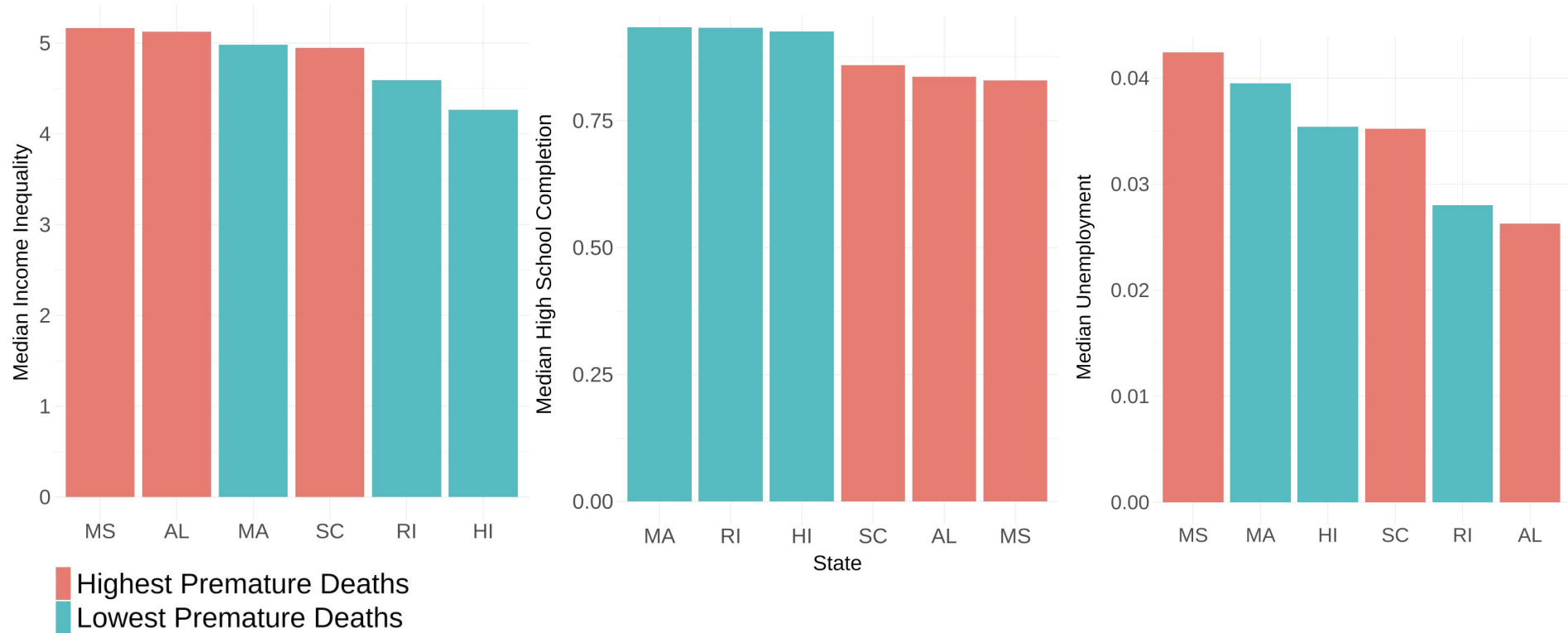
MA



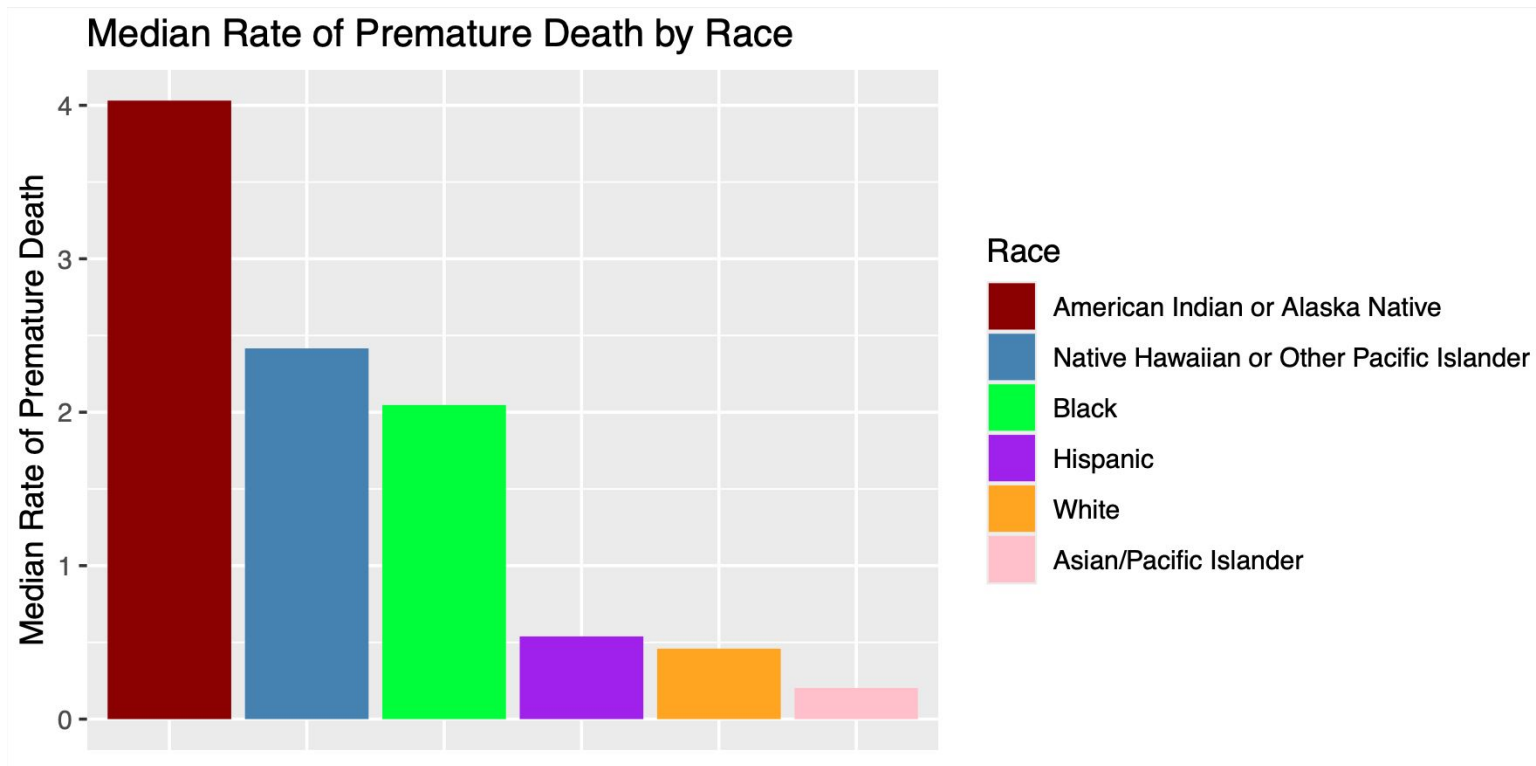
HI



States with higher premature deaths tend to have higher income inequality and high school completion percentages.

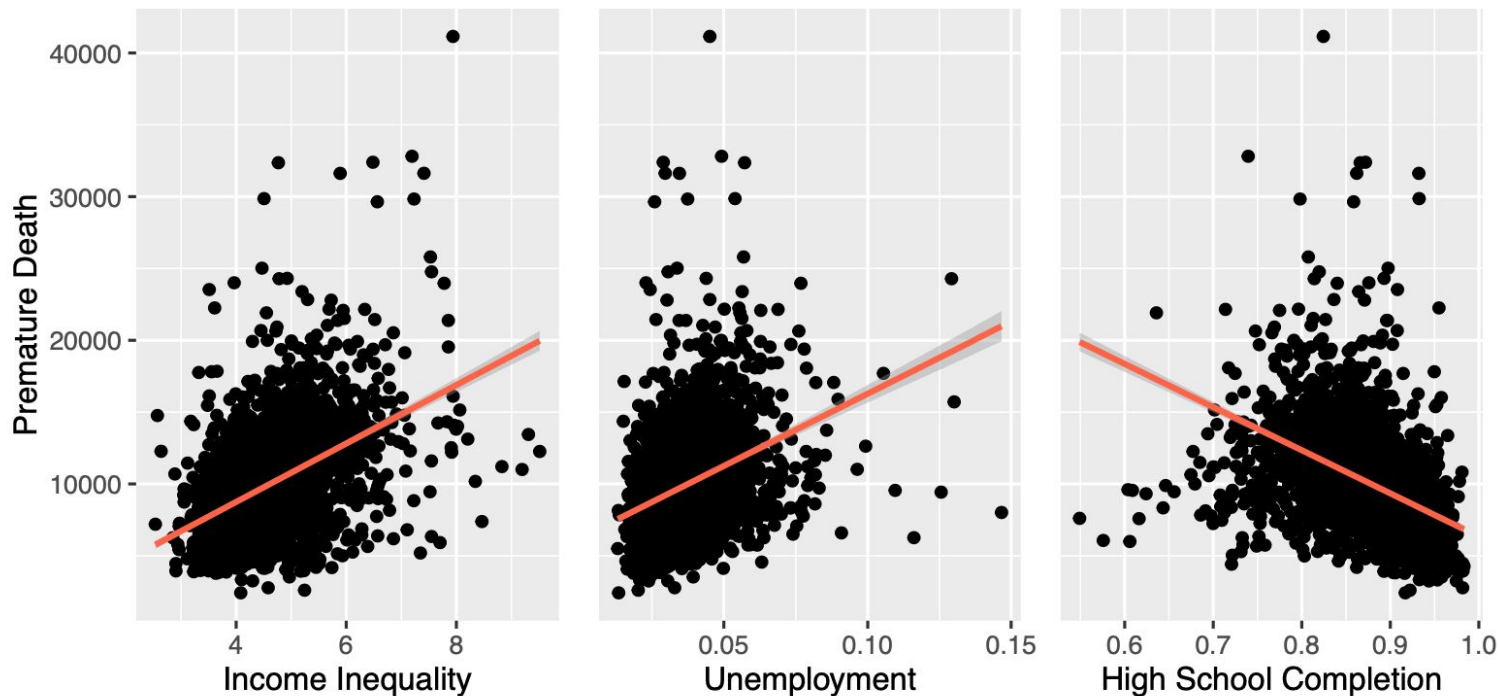


Minority groups are more likely to experience premature deaths.



There are reasonably linear relationships between each of the predictor variables and premature death.

Relationship Between Socioeconomic Factors and Premature Deaths

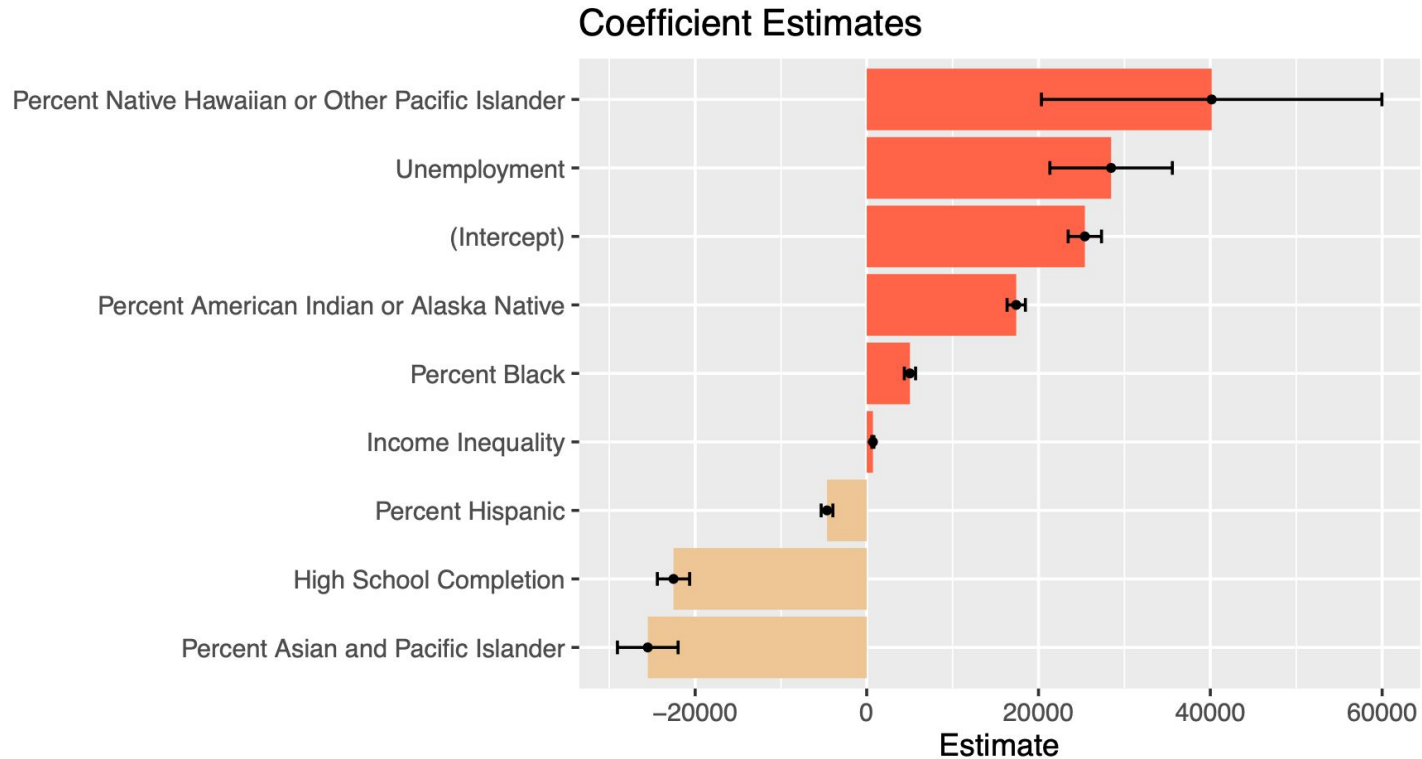


Methods

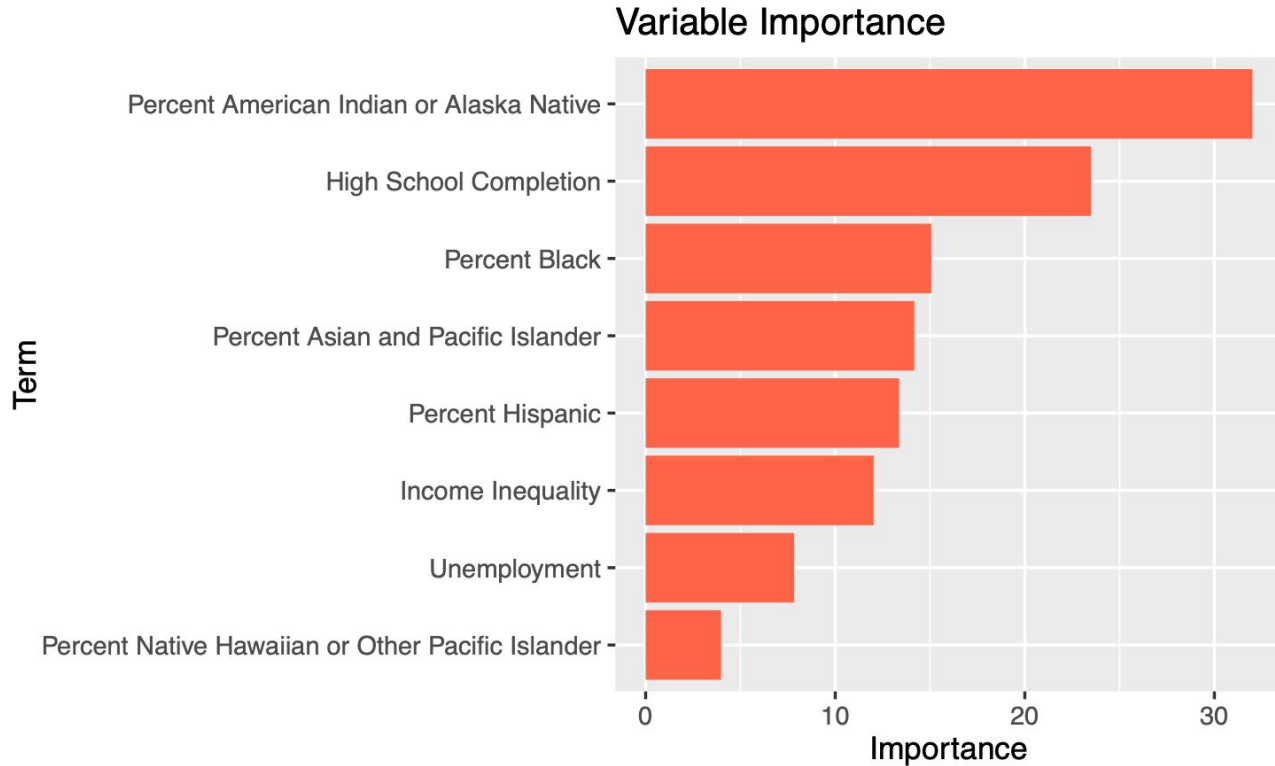
- Build models to predict premature deaths
 - Multiple linear regression
 - Gradient boosted tree
- Evaluate model fits using RMSE

RESULTS

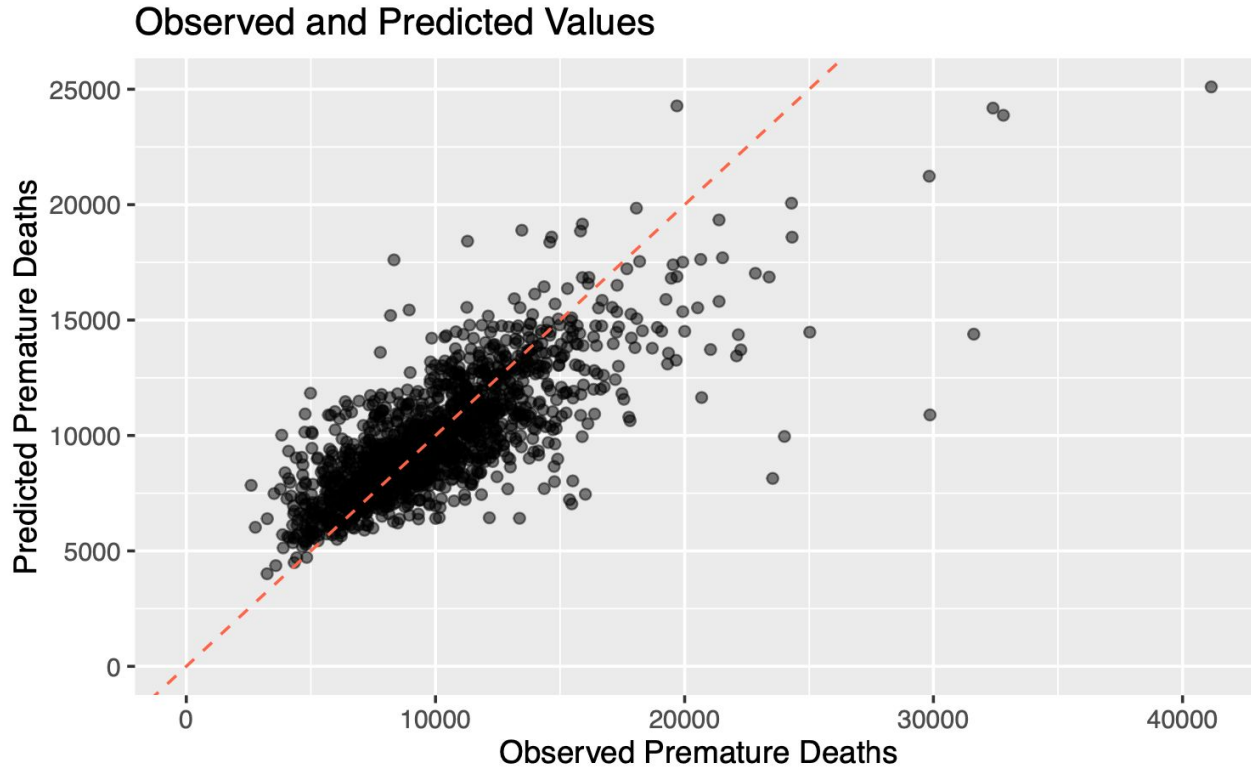
The linear model yielded significant coefficient estimates for each of the socioeconomic variables of interest.



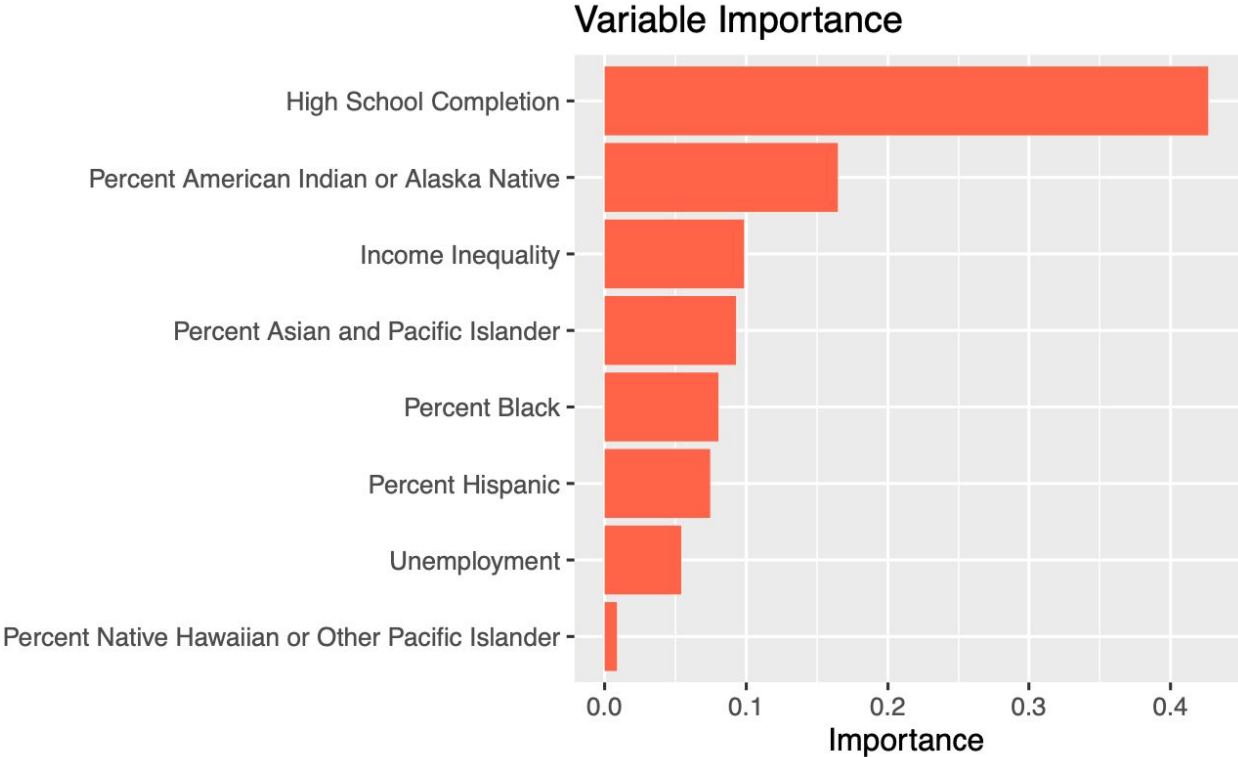
The variable with the highest importance is the percent population of AIAN and percent population of NHOPI has the lowest importance.



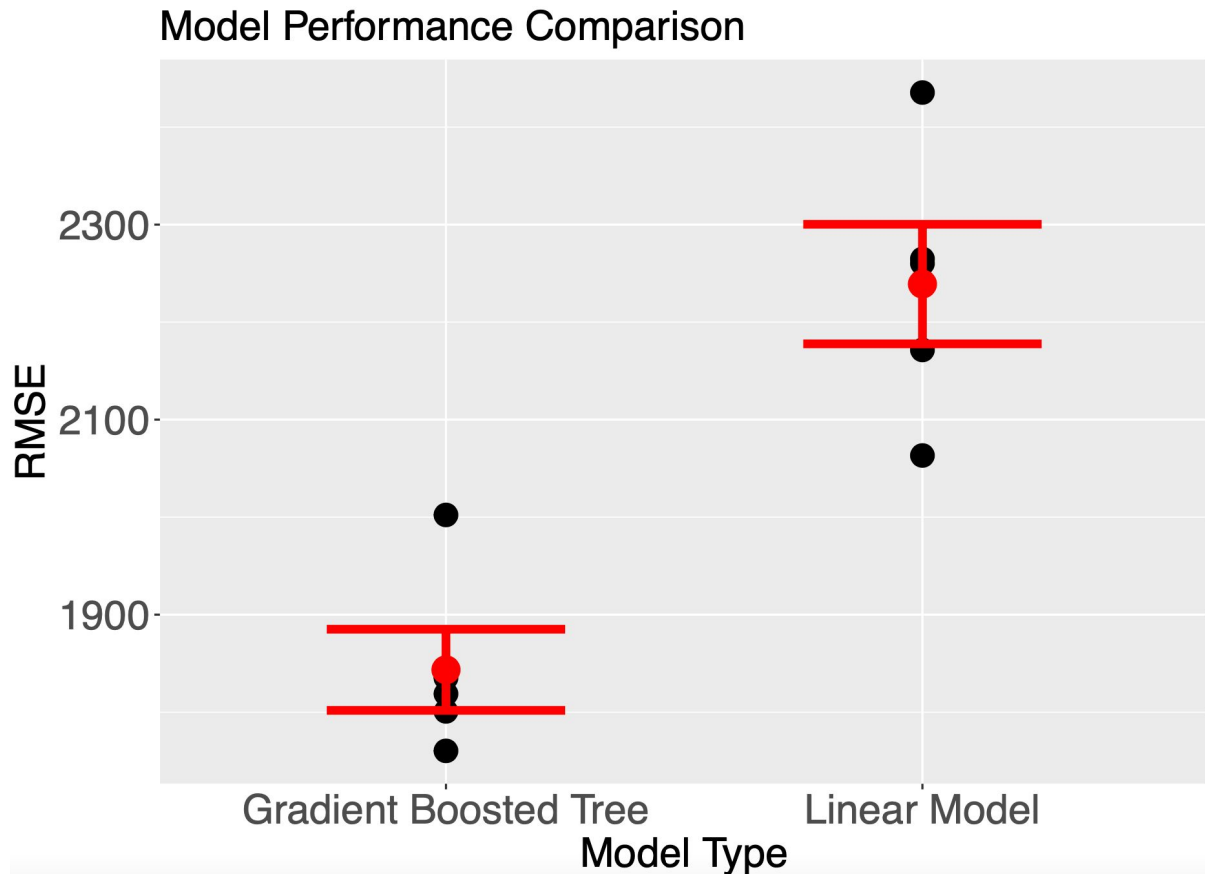
The predicted values of premature death from the gradient boosted tree closely resemble the actual values of the response variable.



In contrast with the previous model, the gradient boosted tree identifies the high school completion rate as the variable with the highest importance.



The gradient boosted tree has a lower RMSE and lower variance than the linear model.



Discussion

- The gradient boosted tree model has the greatest predictive accuracy
 - High school completion is the most significant predictor
 - Native Hawaiian and other Pacific islander is the least significant predictor
- Insights from the gradient boosted tree model contrast with the results of the multiple linear regression model

Limitations

- County-level analysis does not provide insights for individual data
 - Unable to understand individual disparities within a county
- Incomplete racial breakdown data for socioeconomic factors
 - Unable to explore interactions with race
- Variable importance scores only highlight individual variables
 - Unable to explore interactions between predictor variables

Future Work

- Obtain racial breakdowns for income inequality rates, unemployment rates, and high school completion rates
 - Fit interaction models
- Include age and gender as additional predictors
- Increase the sample size

Conclusions

- The boosted tree model is our best predictor of premature deaths
- High school completion percentages was the most important variable in predicting premature deaths
 - Improving education could reduce premature deaths and health disparities
- Targeting access to healthcare for low economic populations may reduce premature deaths