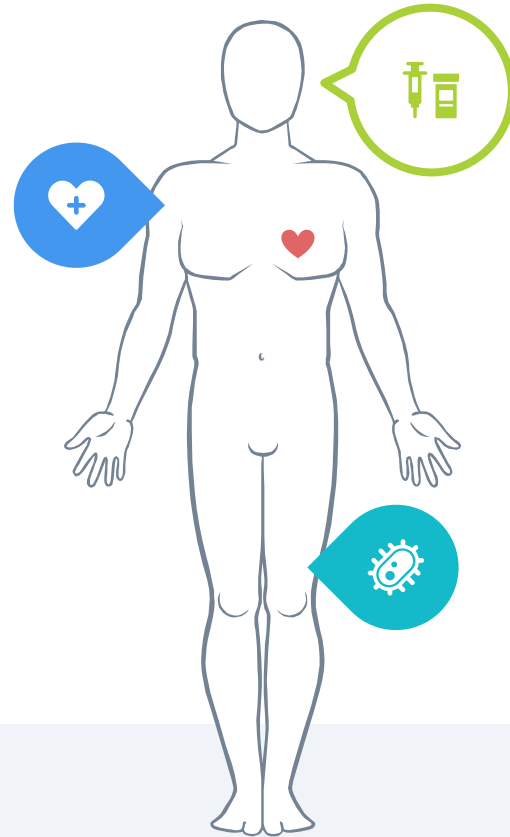


# COVID-19 Vaccination Rate Analysis



# Big question

Do race and income play a factor in whether people get vaccinated?



# Agenda



01

Background

02

Data &  
Methodologies

03

Conclusion & Impact

04

Q & A



**Background**

**1**

# Vaccine Distribution



- ▶ When California started distributing vaccines they wanted to make sure it was equitably distributed
  - ▷ Implemented a system which gave minority groups access codes to get priority for being vaccinated
- ▶ We were curious if this system would actually combat the issue
- ▶ Inequity in vaccine distribution is a global scale issue

# Data



- ▶ Took data from the California government organized by county
- ▶ Underrepresented minorities: American Indian, Black, Latino
- ▶ Represented groups: Asian, White, Multiracial
- ▶ Removed Native Hawaiian/Pacific Islander, Other Race, and Unknown demographics

# Hypotheses

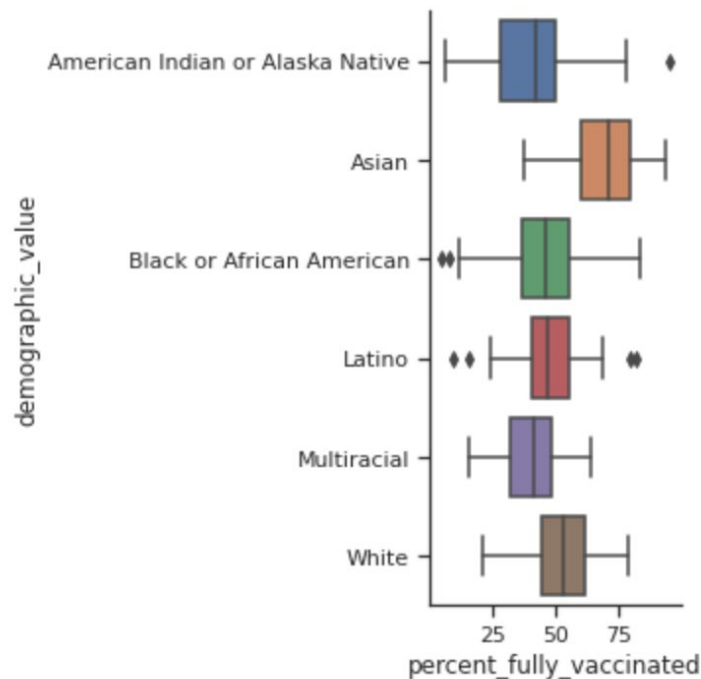
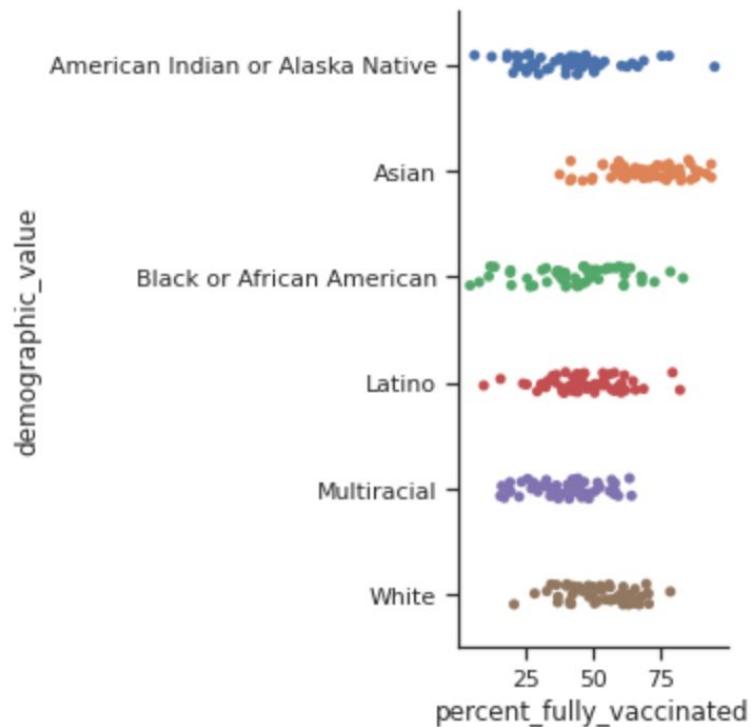
- ▶ Represented racial groups are more likely to be vaccinated than underrepresented racial groups
- ▶ Counties that have lower median incomes are more likely to have lower vaccination rates



**Data & Methodologies**

**2**





Catplot and boxplot of race/ethnicity vs. percent fully vaccinated

# A/B Testing



**Null:** In the total population, the distribution of percent vaccinated is the same for underrepresented racial groups as it is for other racial groups

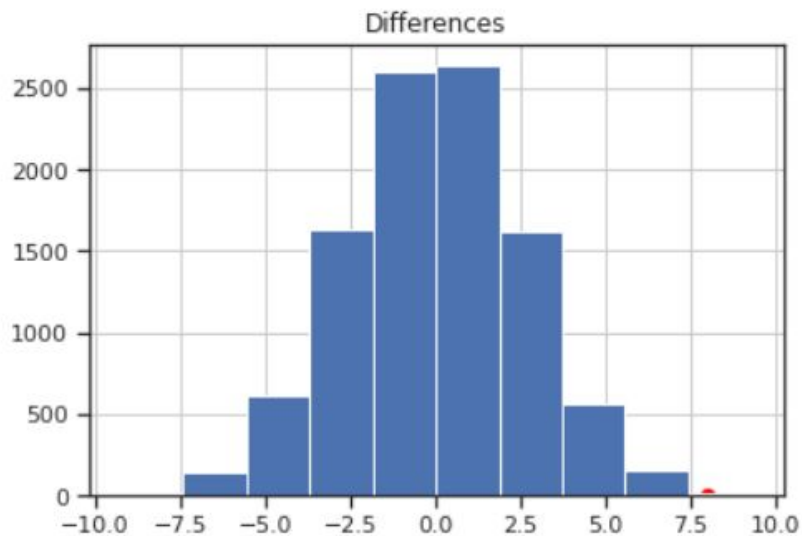
**Alternative:** In the total population, the distribution of percent vaccinated is lower for underrepresented racial groups than for other racial groups

**Test Statistic:** (average percent of represented people who are fully vaccinated)  
– (average percent of underrepresented people who are vaccinated)





```
# plotting the results of the simulations  
# the red dot represents the observed test statistic  
  
differences_df.hist()  
plt.scatter(observed_difference, -0.002, color='red', s=40);
```



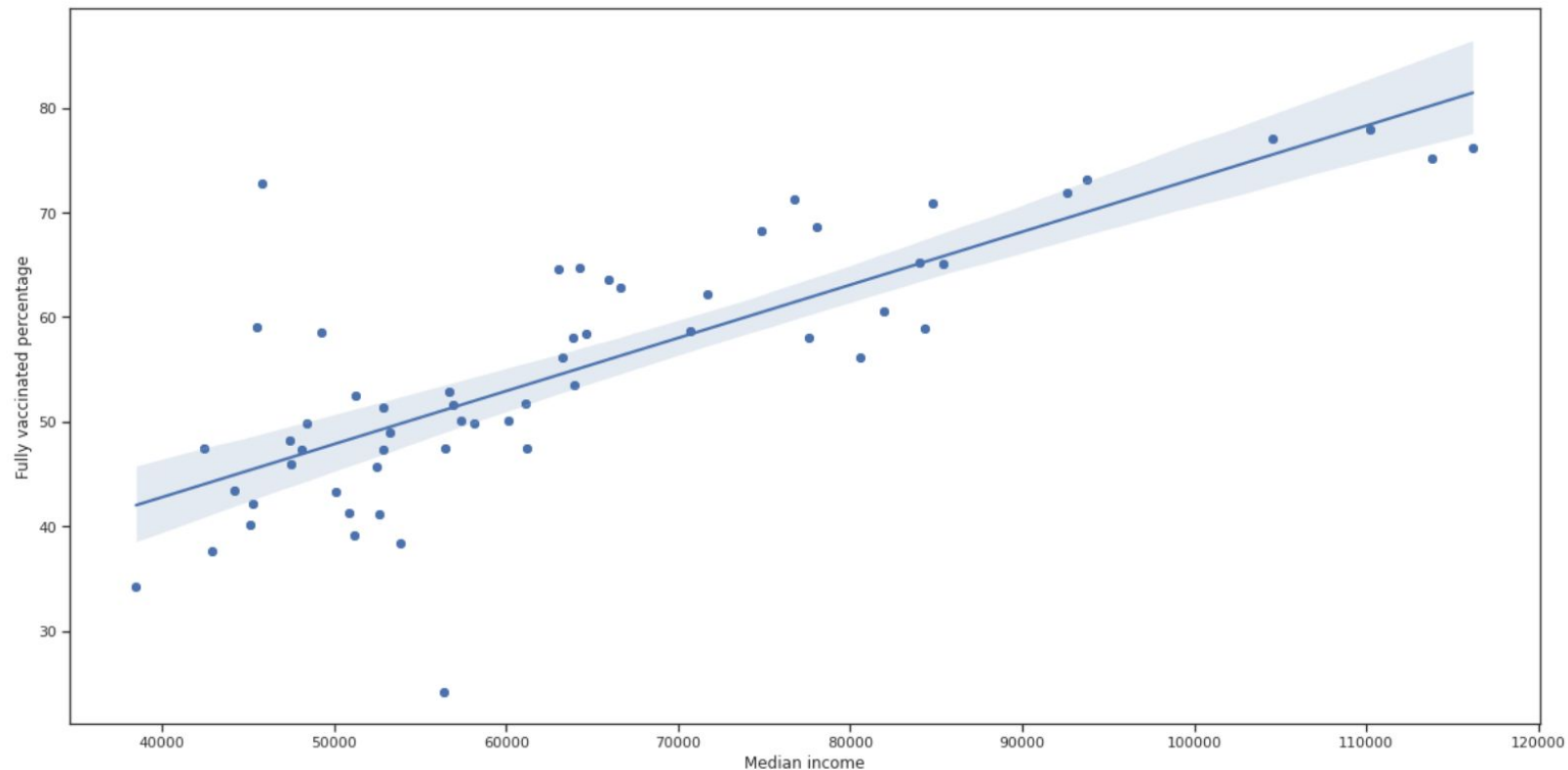
# Linear Regression



- Analyzed association between median income and percentage of fully vaccinated and at least one dose per county
- Dataset containing:
  - County
  - Median income
  - Fully vaccinated percentage
  - At least one dose percentage

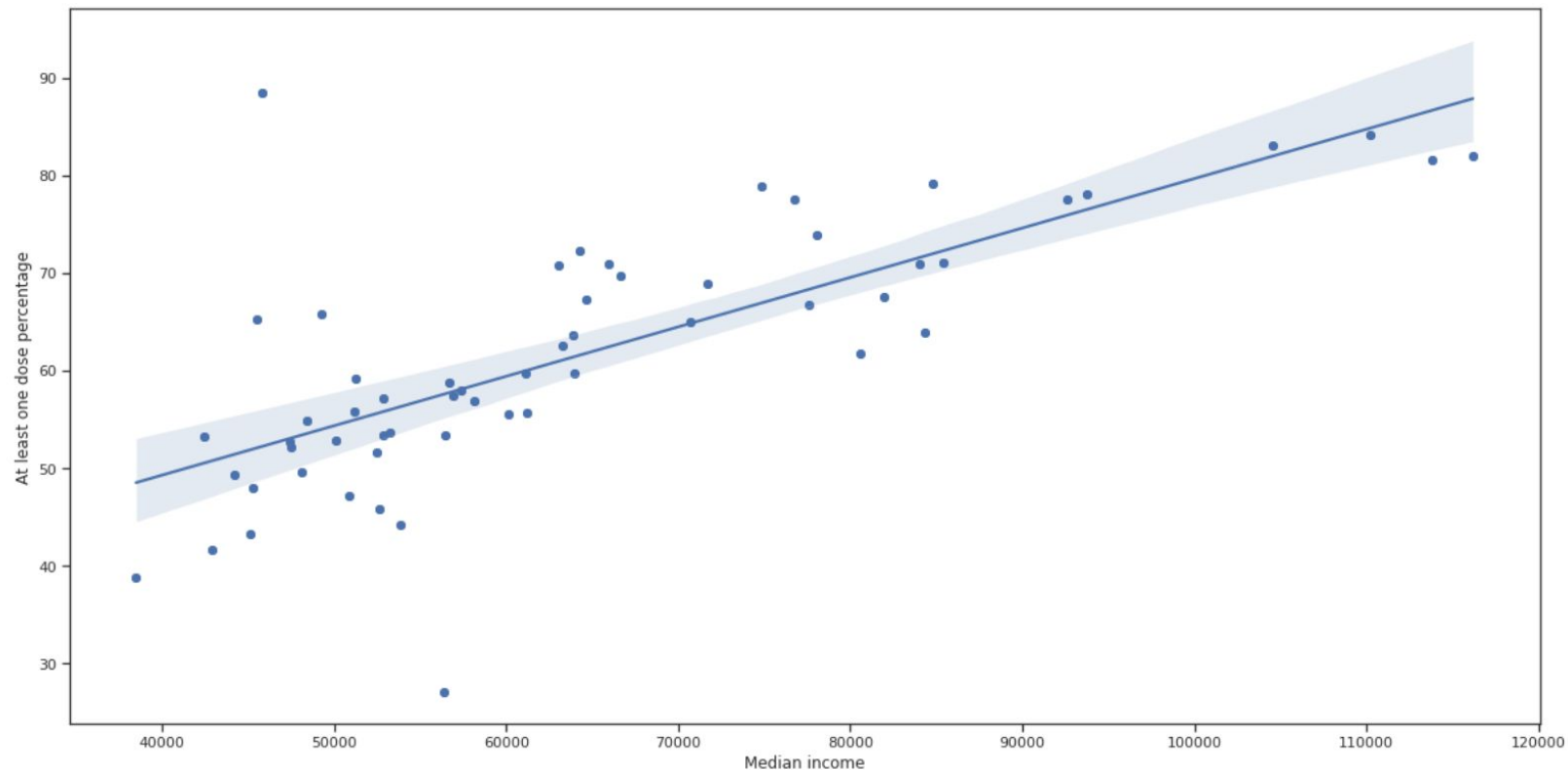


The correlation coefficient is: 0.7889963375011697



Fully vaccinated percentage vs. median income

The correlation coefficient is: 0.7431907440294258



At least one dose percentage vs. median income

# Modelling



- Using sklearn API to create a model using 80/20 split of training/test data
- Coefficient of determination:
  - 0.74 (predicting fully vaccinated from median income)
  - 0.69 (predicting at least one dose from median income)
- Reducing mean squared error by standardizing units
  - From 25.5 to 0.18 (fully vaccinated)
  - From 28.3 to 0.17 (one dose)

**Conclusion and Impact**

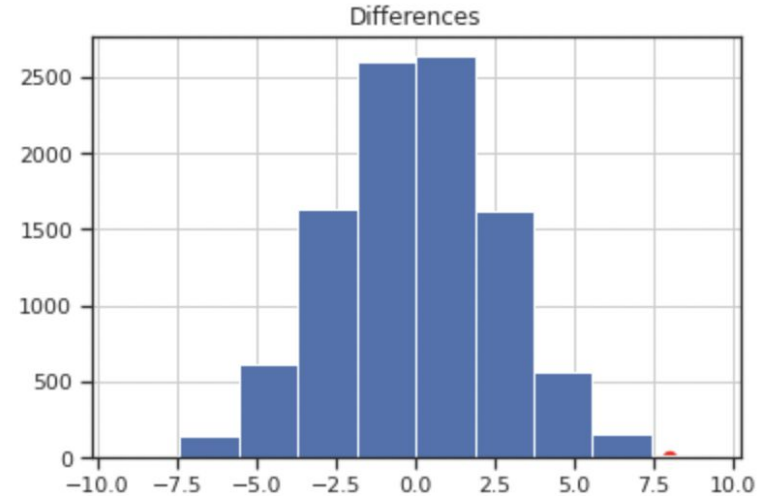
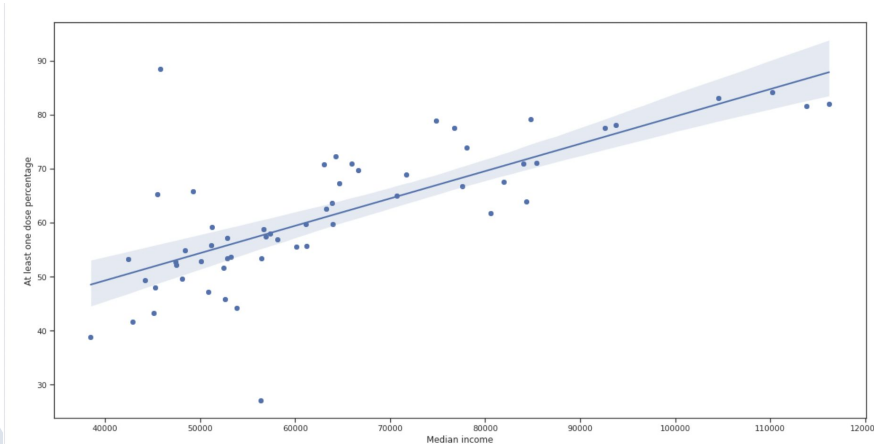
**3**



# Our Original Hypotheses



- P-value and correlation coefficients demonstrate statistical significance
- Race and income do play factors in vaccination rates



# Limitations

## Confounding Variables

Education level

Number of hospitals/clinics

Political affiliation/personal beliefs

## Ecological Correlation

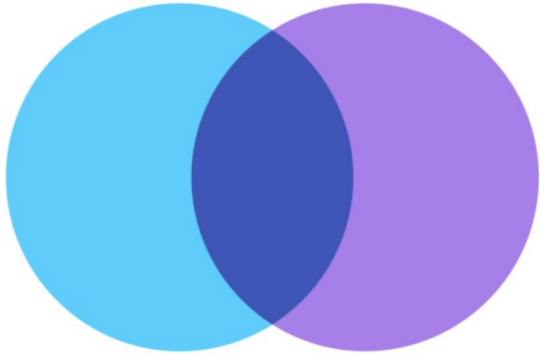
Each point on our graphs represents a county rather than individuals

Can't look at distributions within individual counties

# Further Analysis and Research



Combine Race and  
Income



Include Native  
Hawaiian Group



Increase Scale



## Applications + Impact

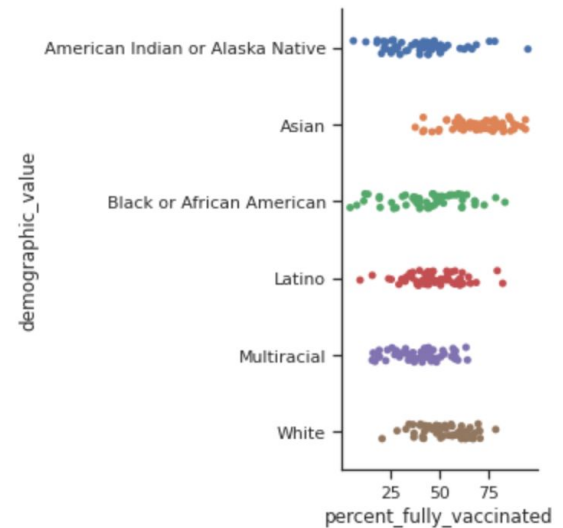
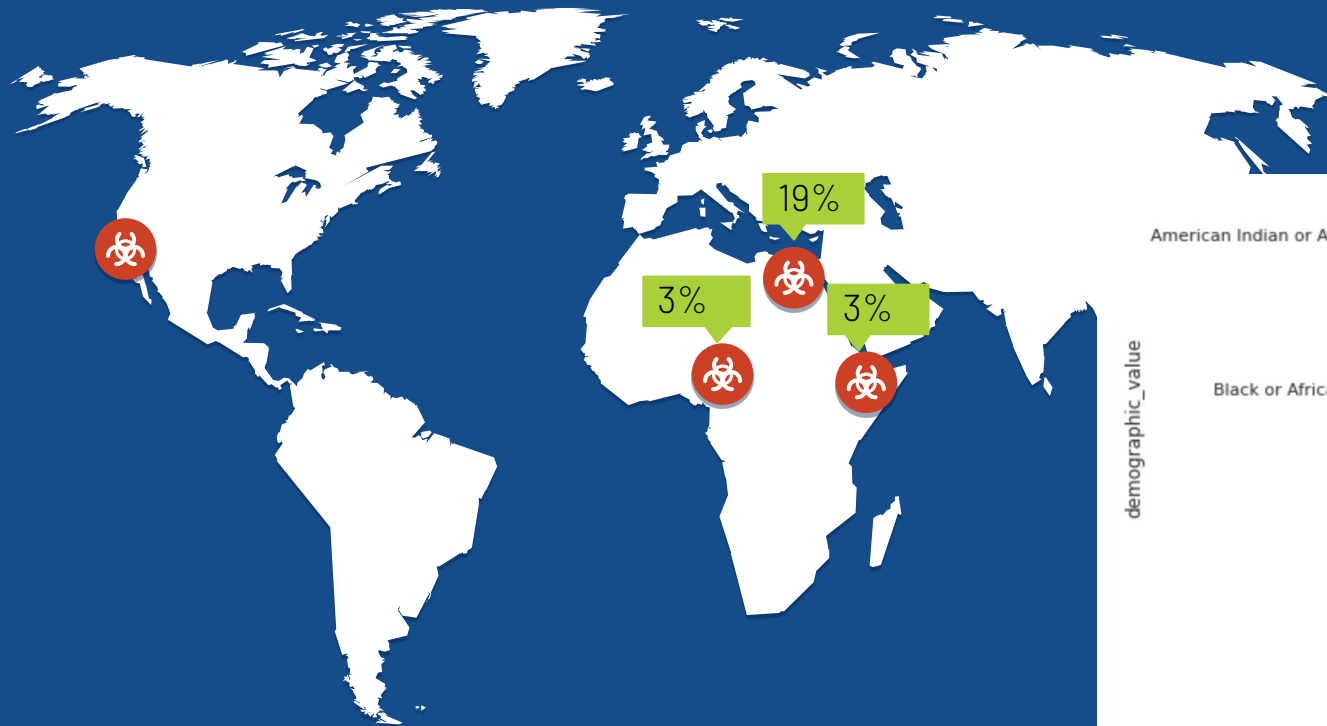
More work needs to be done besides vaccine equity codes

Target specific counties and demographics





# Comparing to Third-World Countries



// Ultimately, we can be only as healthy as the least healthy people among us.

- **Andy Slavitt**

**Interim Administrator of the Centers for Medicare and Medicaid Services**



Q&A

4