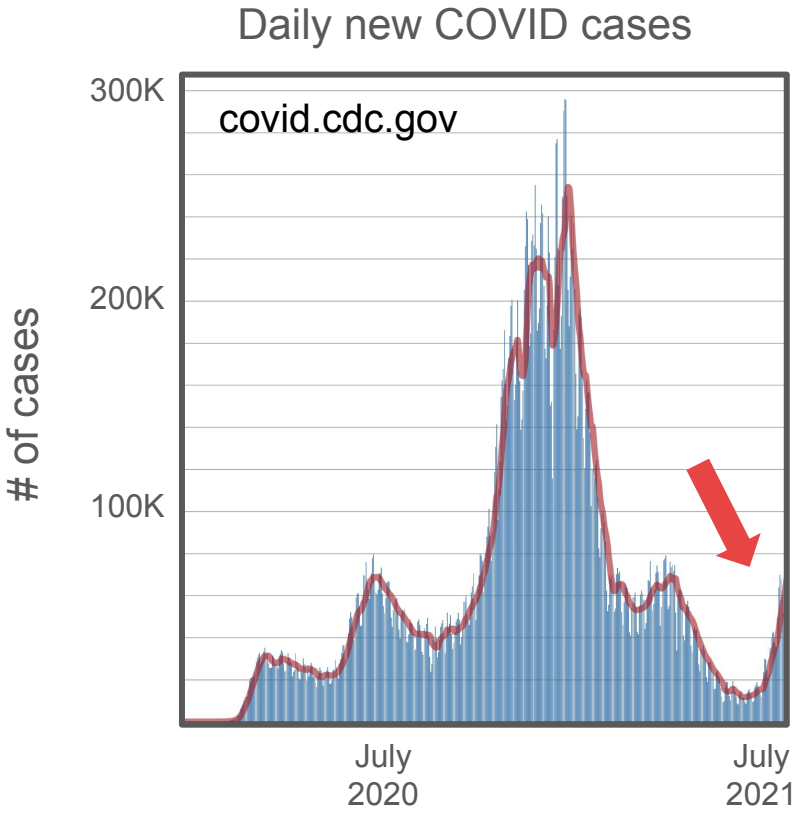
The Great Divide:

Factors associated with low COVID vaccination rates in the US

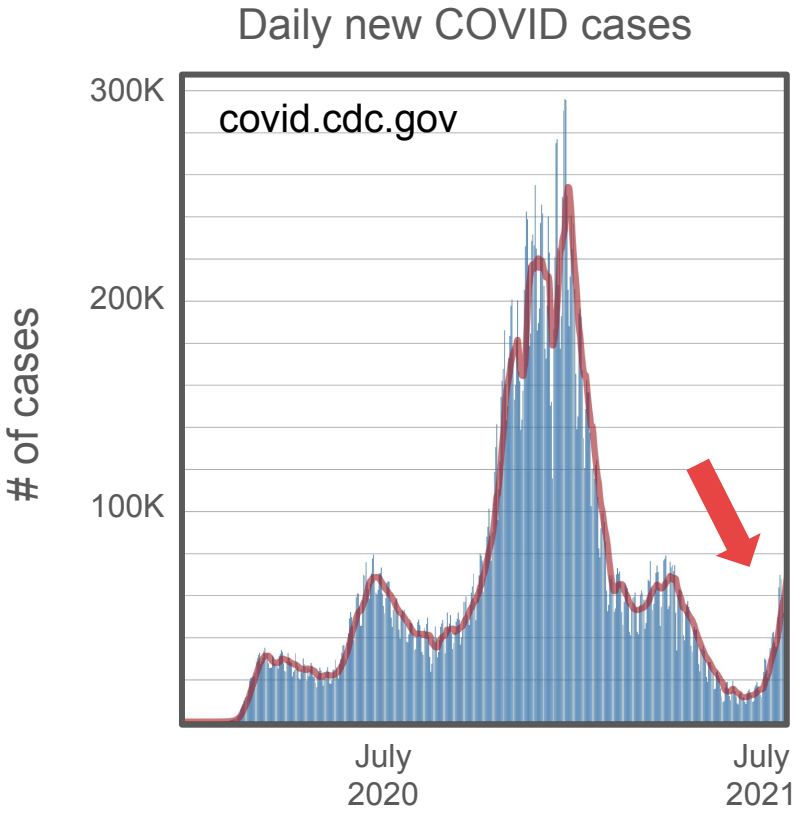
Maya Remington
Summer/Fall 2021

The Problem



- COVID-19 cases are rising again
  - Primarily affecting the unvaccinated
- Approx 30 % of US adults are unvaccinated
  - Varies county-to-county

The Problem



- COVID-19 cases are rising again
  - Primarily affecting the unvaccinated
- Approx 30 % of US adults are unvaccinated
  - Varies county-to-county

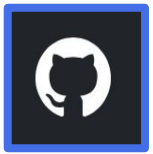
Aim: Use linear regression to answer the question…

At the county level, which political and socio-economic factors best explain high rates of UNvaccinated adults?

Data & Methods

County-level data

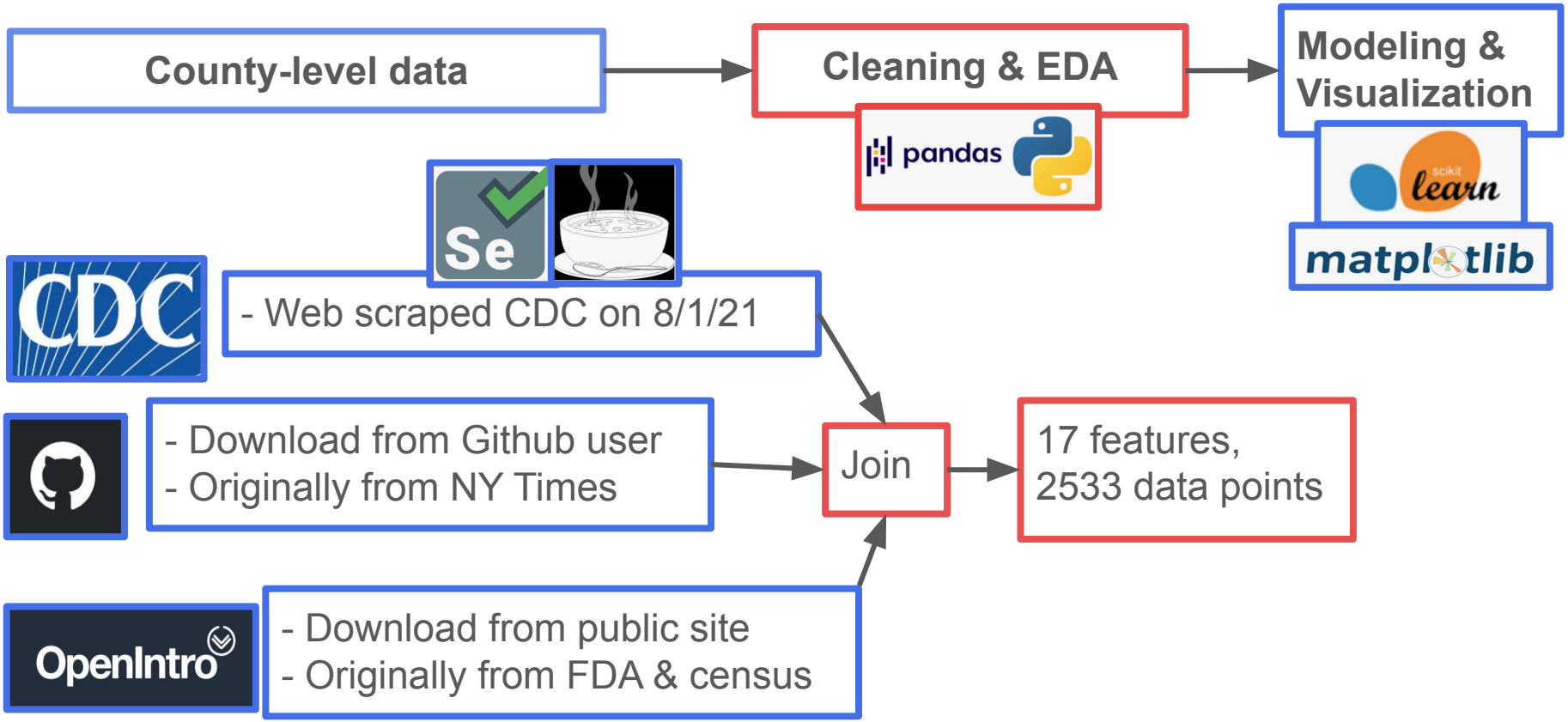
- Web scraped CDC on 8/1/21



- Download from Github user
- Originally from NY Times

- Download from public site
- Originally from FDA & census

Data & Methods



Key Calculations

Target:

% unvaccinated adults

= 1 - % adults with at least 1 shot

Key Calculations

Target:

% unvaccinated adults

= 1 - % adults with at least 1 shot

Important feature:

% Vote split in 2020 Presidential election

= % GOP vote - % Dem vote

17 Features

% Race/ethnicity: asian, black, hispanic, native, pac islander, white

Education level: % HS grad, % bachelors

Economic factors: % poverty, % uninsured, median household income

Access to technology: % have computer, % have smartphone

Other: % vote split, % age 65+, population density, household size

11

~~1~~7 Features

% Race/ethnicity: asian, black, hispanic, native, pac islander, ~~white~~

Education level: ~~% HS grad~~, % bachelors

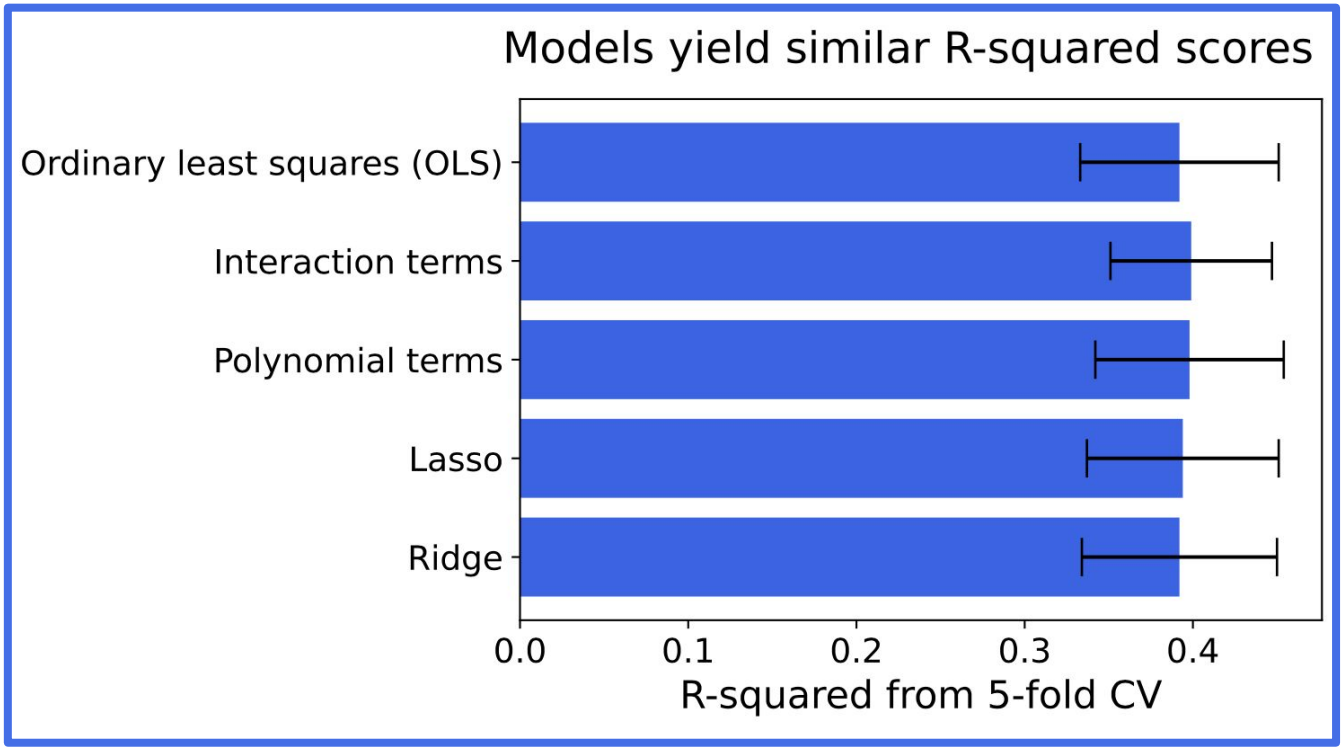
Economic factors: % poverty, % uninsured, ~~median household income~~

Access to technology: ~~% have computer, % have smartphone~~

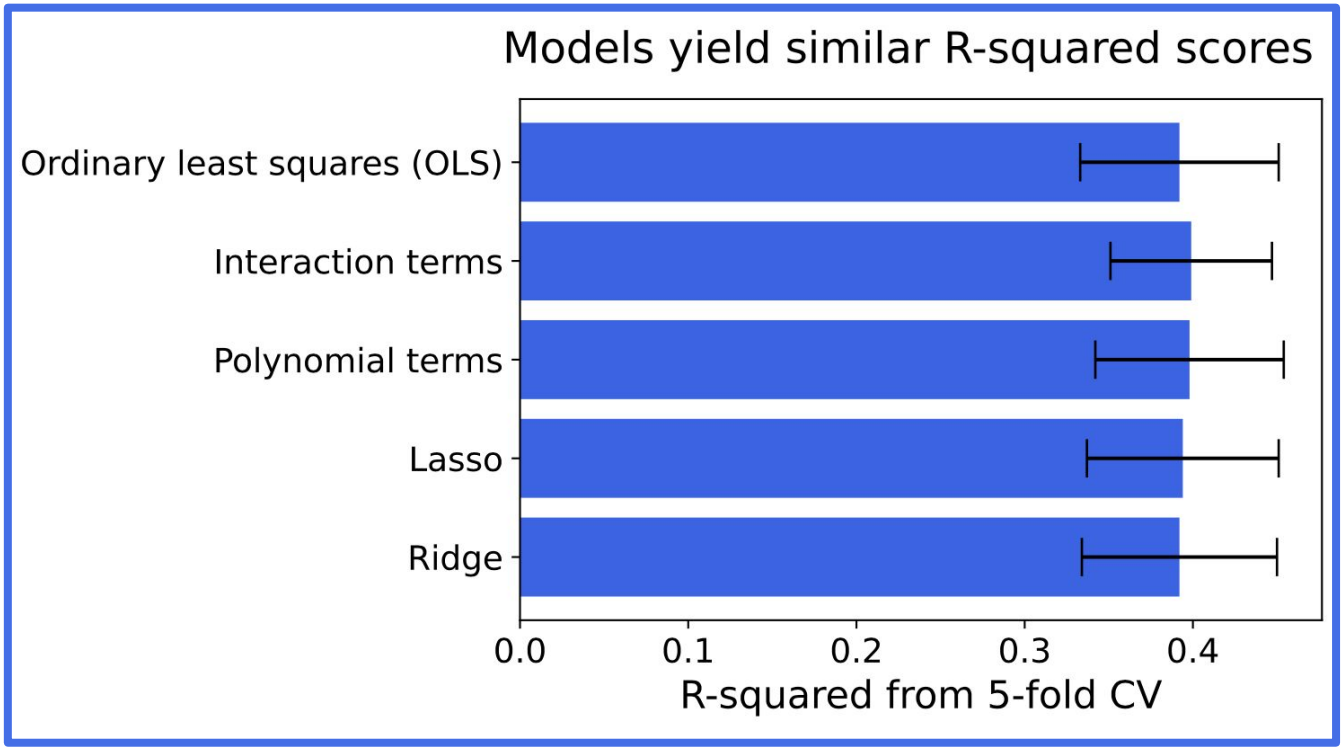
Other: % vote split, % age 65+, population density, ~~household size~~

Removed 6 features due to high collinearity (based on variation inflation factor [VIF])

Linear regression models



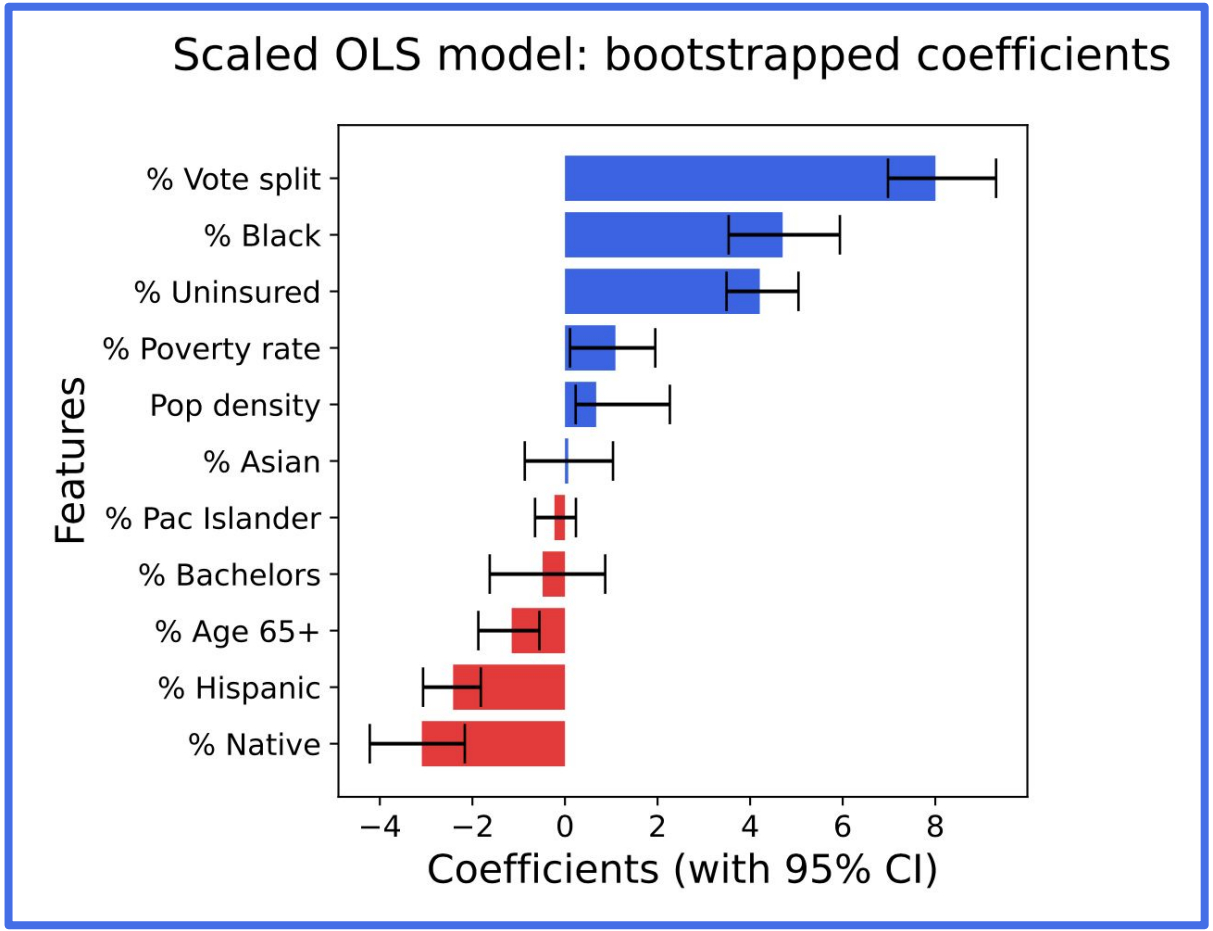
Linear regression models

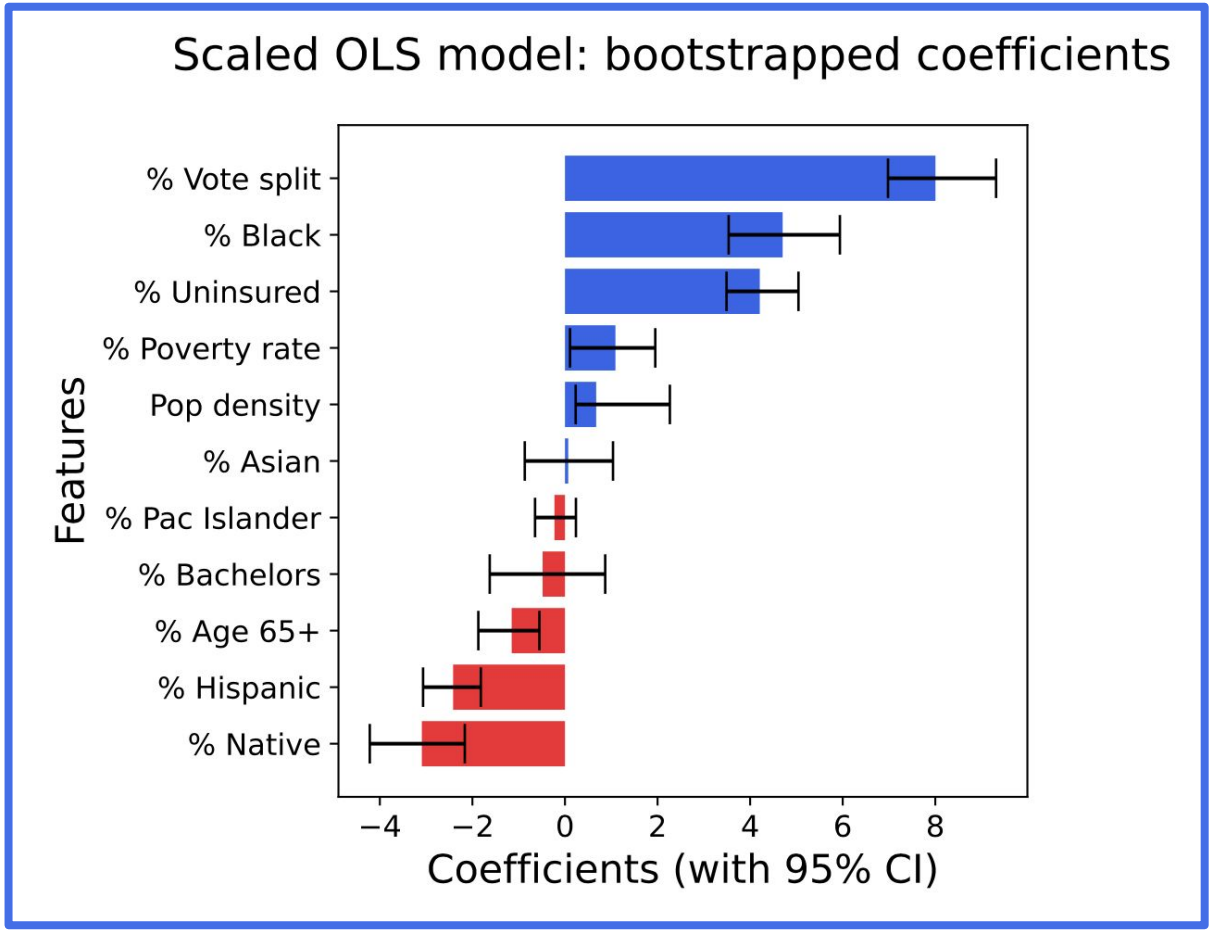


Final model: OLS

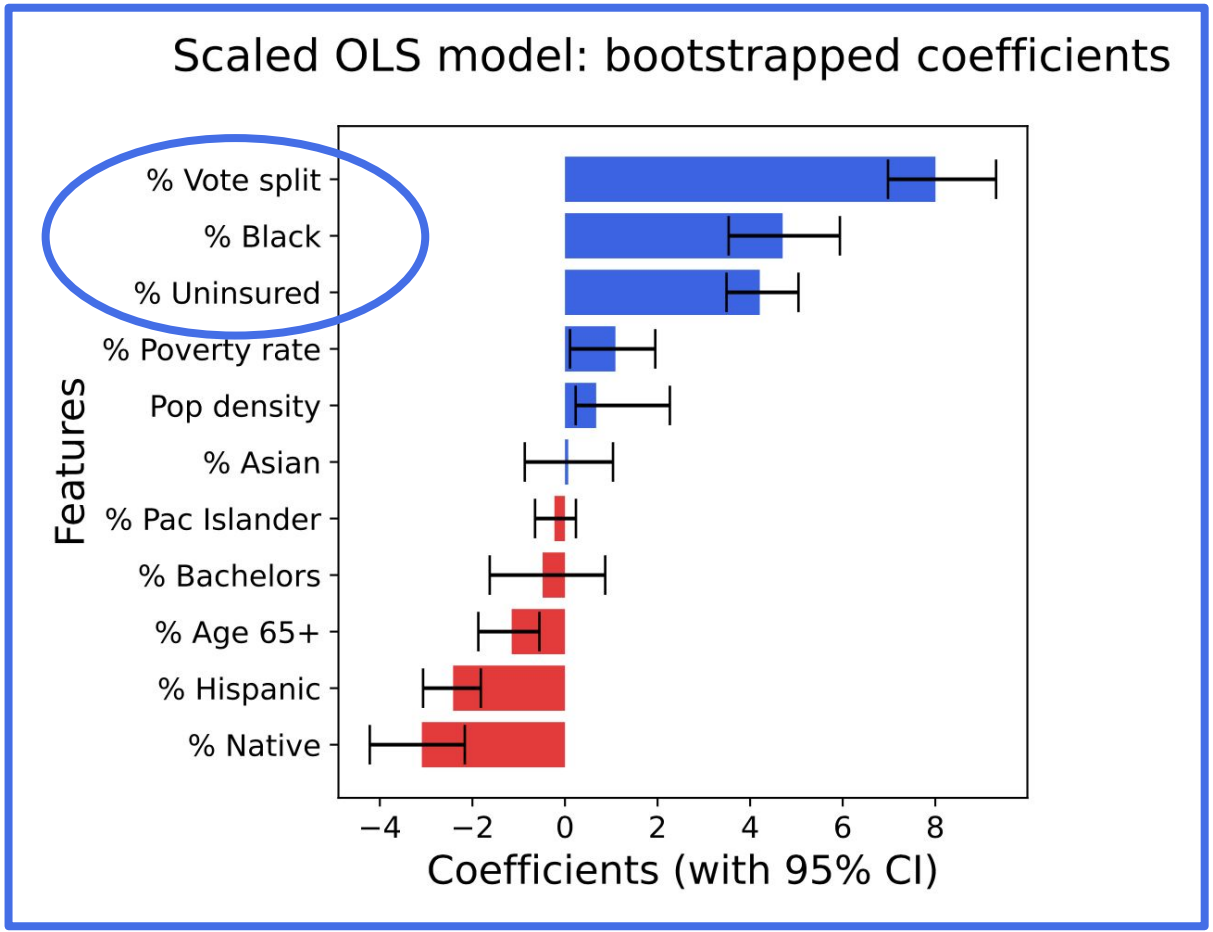
R-squared: 0.425

Mean absolute error (MAE): 8.14


Scaled OLS model: bootstrapped coefficients
% Vote split
% Black
% Uninsured
% Poverty rate
Pop density
% Asian
% Pac Islander
% Bachelors
% Age 65+
% Hispanic
% Native
Features
−4
−2
0
2
4
6
8
Coefficients (with 95% CI)

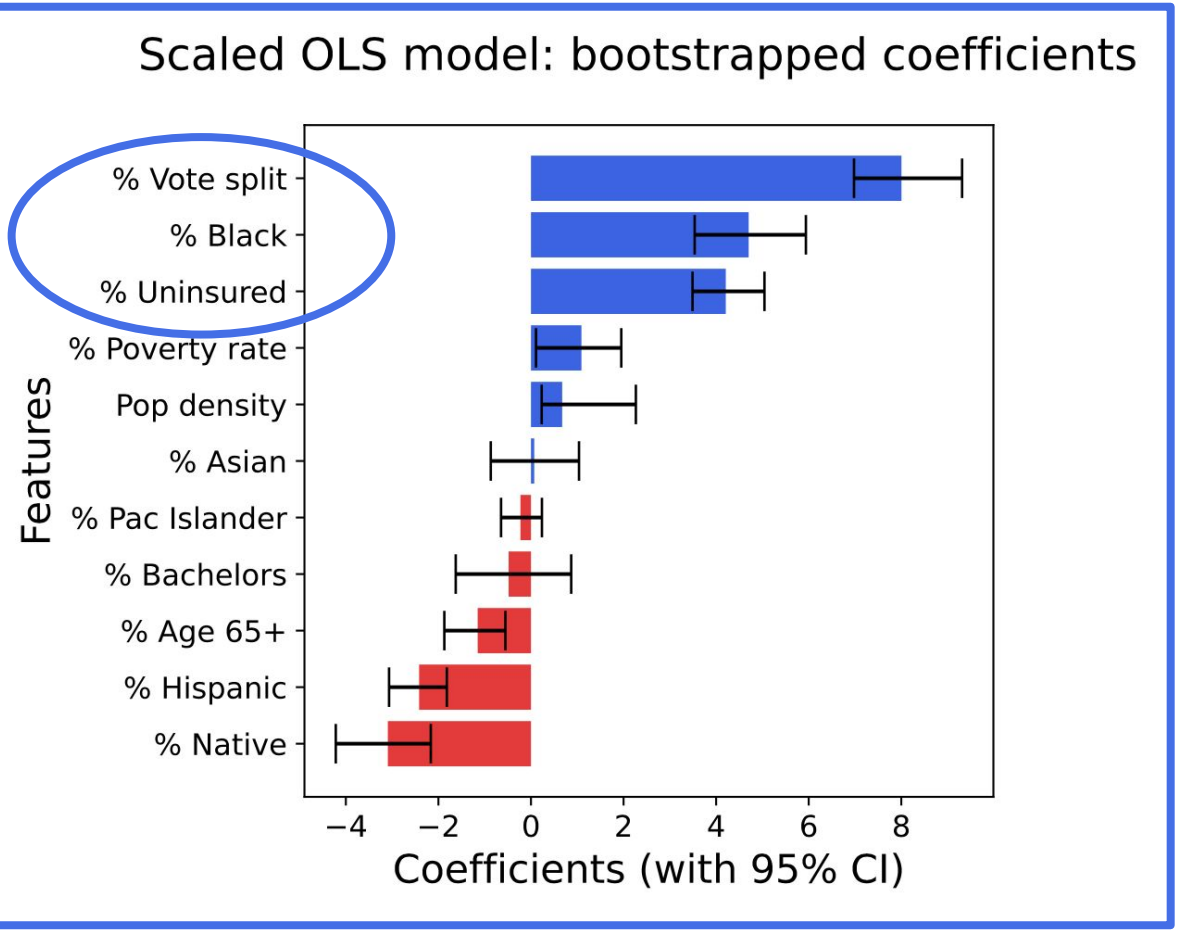


At the county level, which factors best explain high rates of **UNvaccinated adults?**



At the county level, which factors best explain high rates of **UNvaccinated adults?**

1) % Vote split in 2020
= % GOP vote
- % Dem vote



At the county level, which factors best explain high rates of **UNvaccinated adults?**

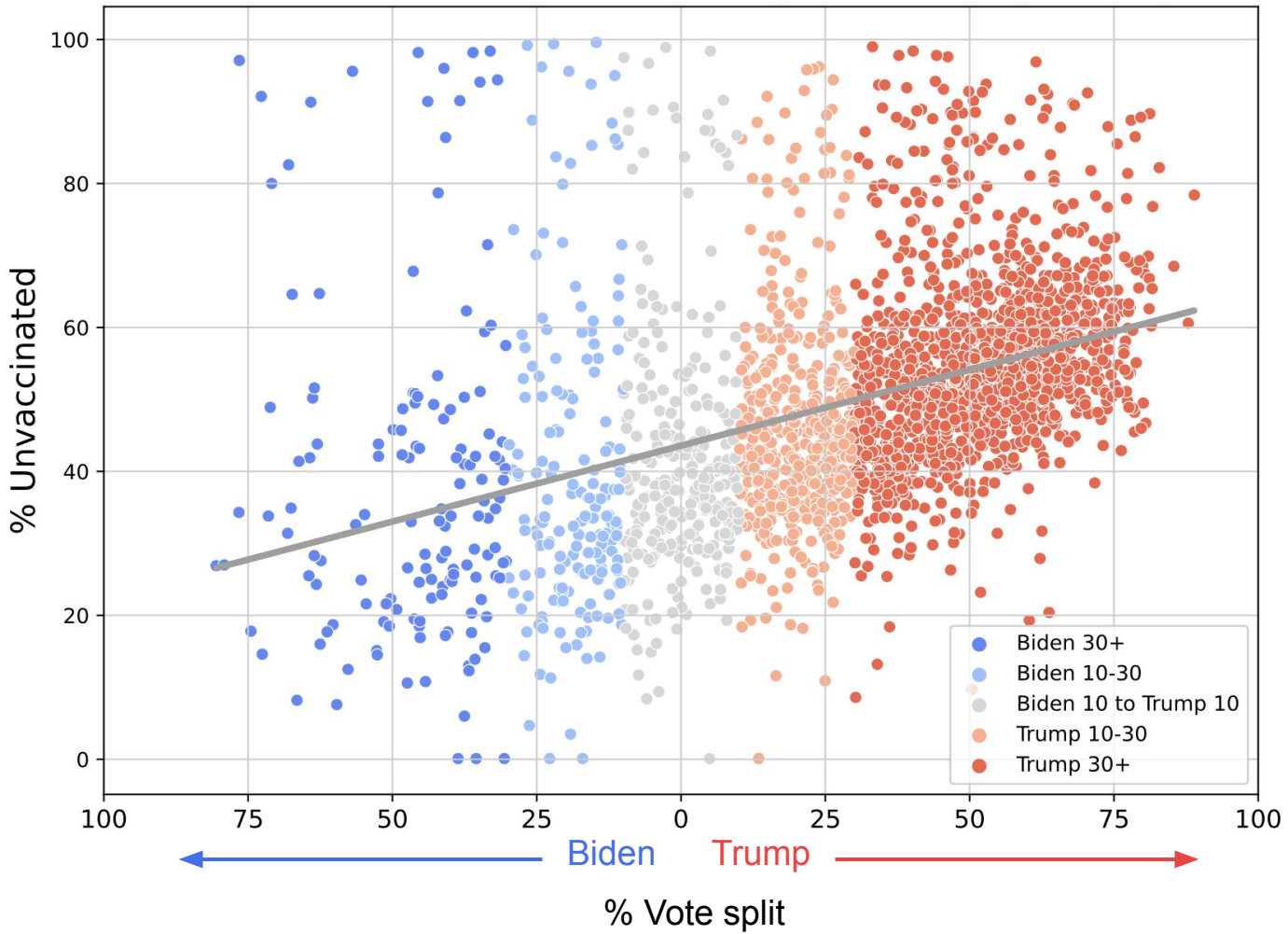
1) % Vote split in 2020
= % GOP vote
- % Dem vote

2) % Black

3) % Uninsured

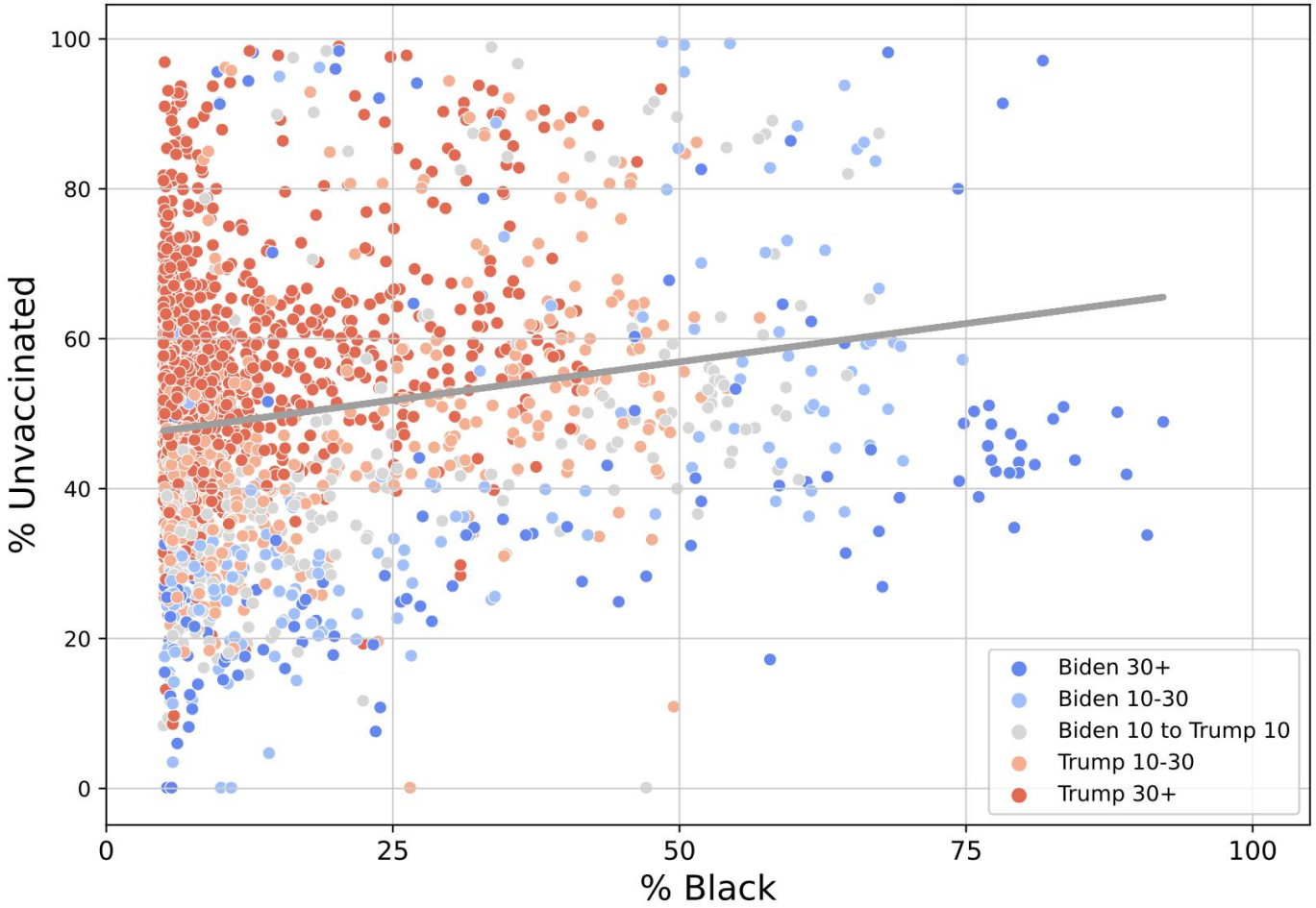
% Vote Split

Red counties have higher rates of unvaccinated adults



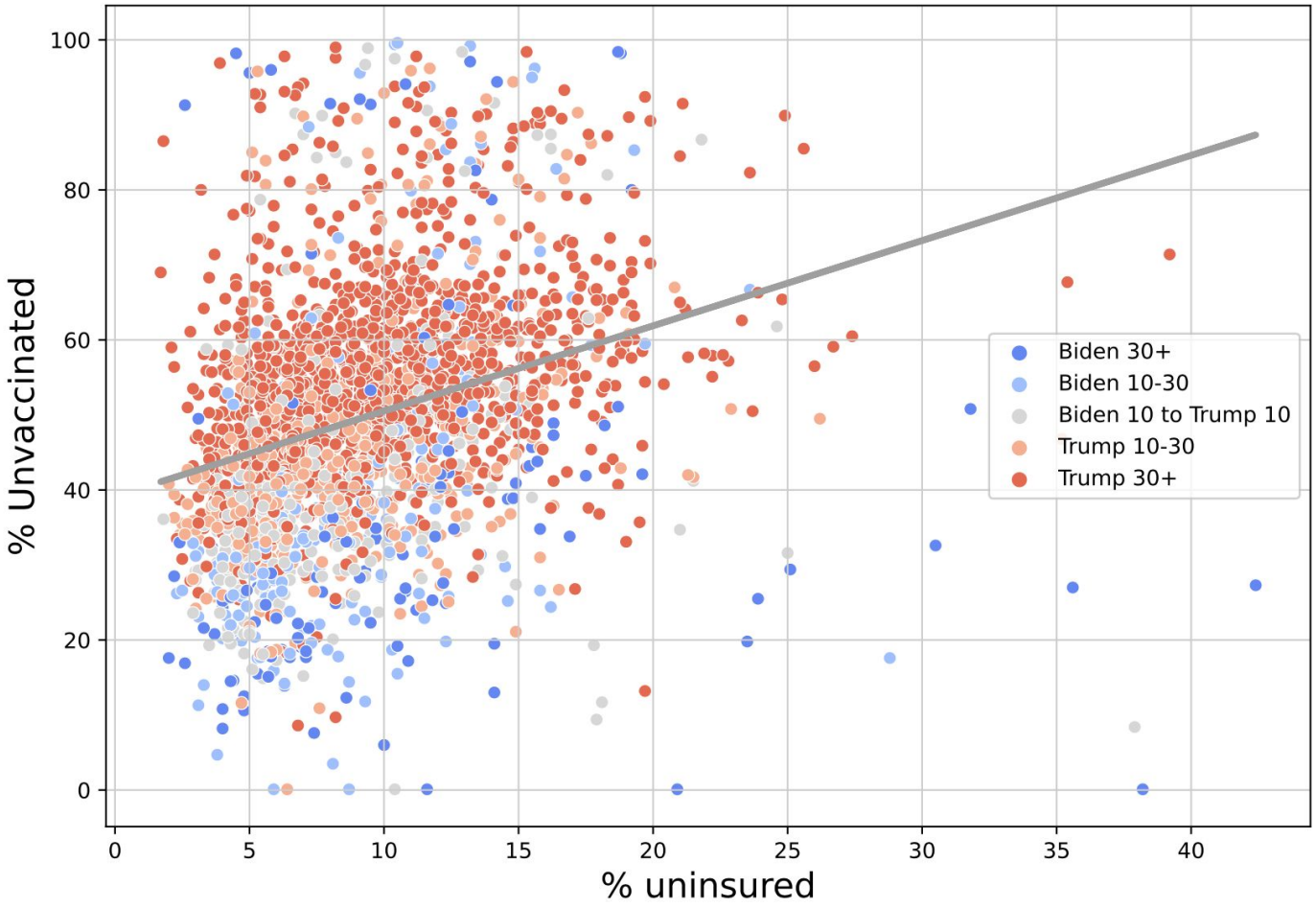
% Black

Counties with larger % black have higher rates of unvaccinated adults



% Uninsured

Counties with larger % uninsured have higher rates of unvaccinated adults



Conclusion

Linear regression models

- Final model: OLS
- Most associated features:
  % Vote split, % Black, % Uninsured

Conclusion

Linear regression models

- Final model: OLS
- Most associated features:
  % Vote split, % Black, % Uninsured

Step toward addressing vaccine hesitancy

- Tailor the vaccination strategy

Thank you!