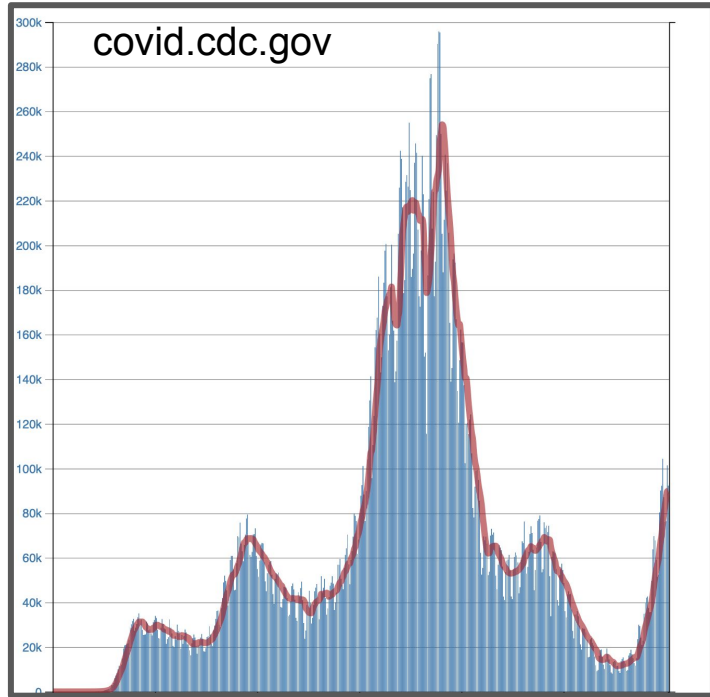# Turning Down the Jab:

Using regression to explore factors associated with low COVID vaccination rates in the US

Maya Remington
August 6, 2021

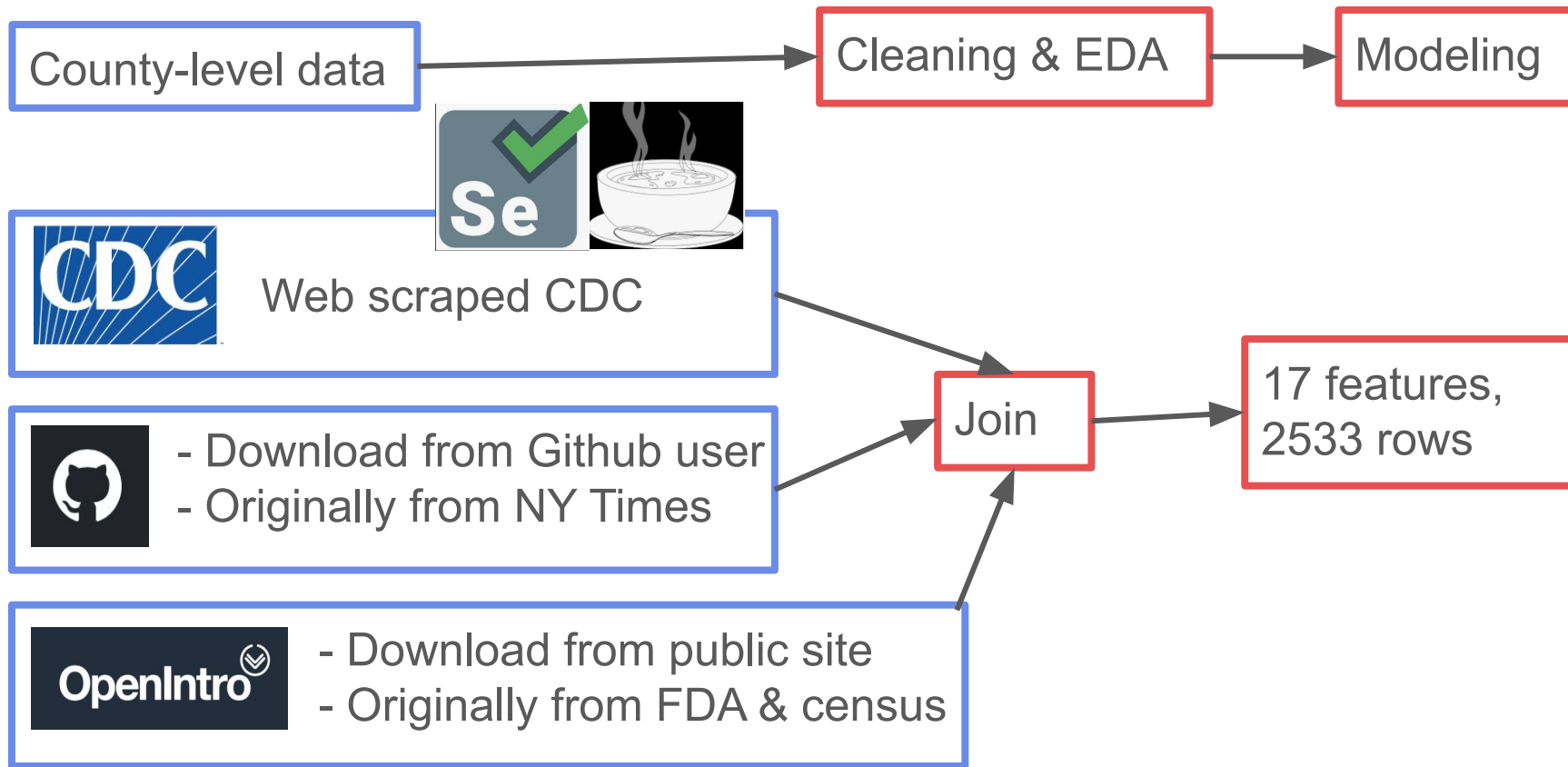# The Problem

covid.cdc.gov

Daily Cases in US

- COVID-19 cases are rising again
  - Fueled by the delta variant
  - Primarily affecting the unvaccinated (~ 30% of US adults)

Q: What political and socioeconomic factors best explain the variation in vaccination rates at the county level?

Data & Methods

County-level data → Cleaning & EDA → Modeling

Web scraped CDC

- Download from Github user
- Originally from NY Times

- Download from public site
- Originally from FDA & census

Join → 17 features, 2533 rows

Key Calculations

Target: **% unvaccinated adults**

 = 1 - % adults with at least 1 shot

% diff (aka 'vote split') in 2020
Presidential election

      = % GOP vote - % Dem vote

The remaining 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: age 65+, population density, household size

The remaining 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: age 65+, population density, ~~household size~~

Improving the model

Feature modeling and regularization applied to 5-fold cross validation:

Model	OLS (baseline)	Interaction terms	Polynomial terms	Lasso	Ridge
R-squared	0.392	0.399	0.398	0.394	0.392

Final model: OLS scaled

Train
- R^2: 0.399
- MAE: 8.70

Test
- R^2: 0.425
- MAE: 8.14

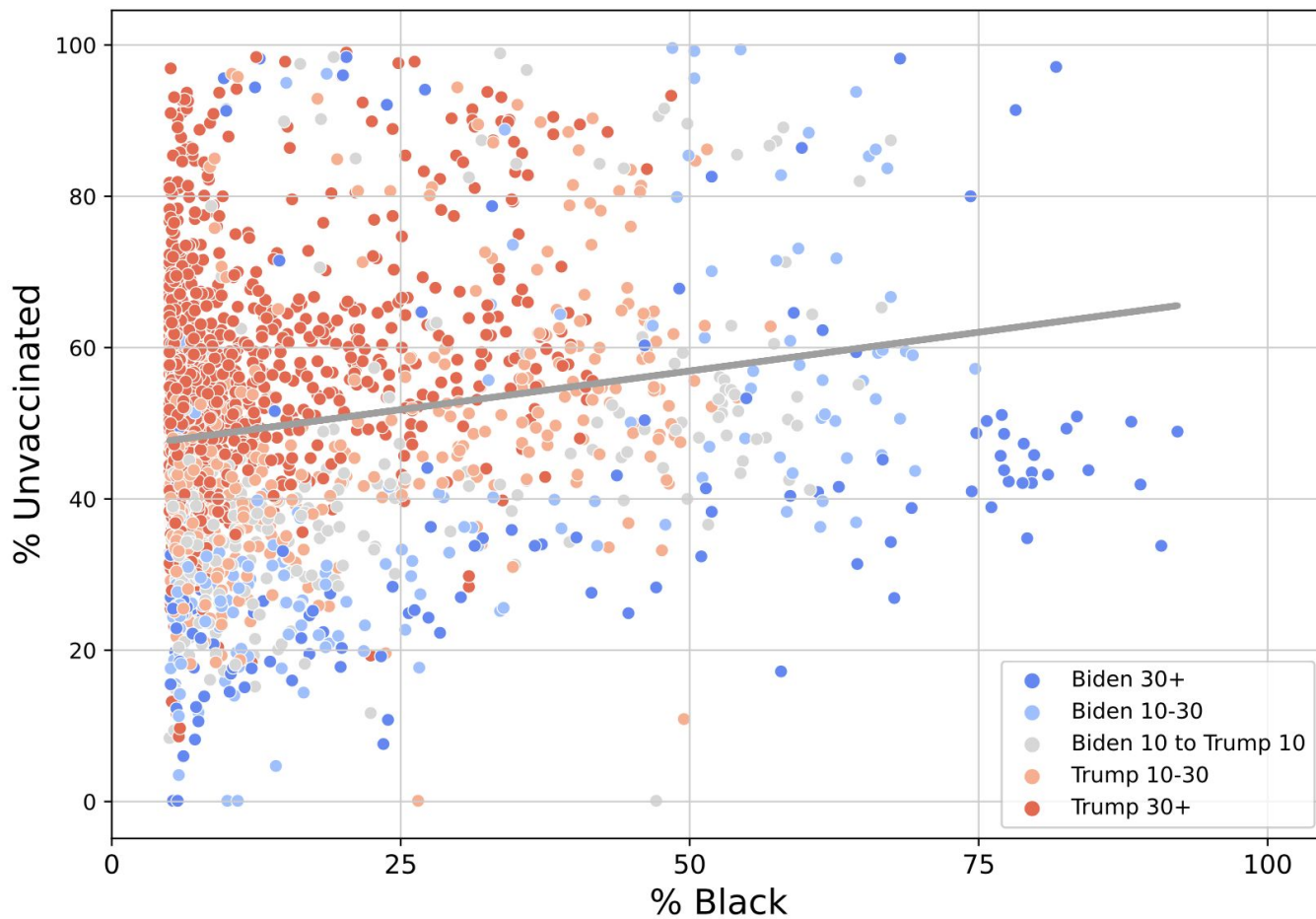Scaled OLS: absolute value of coefficients

| Coefficient |

Positive values
Negative values

Largest coefficient
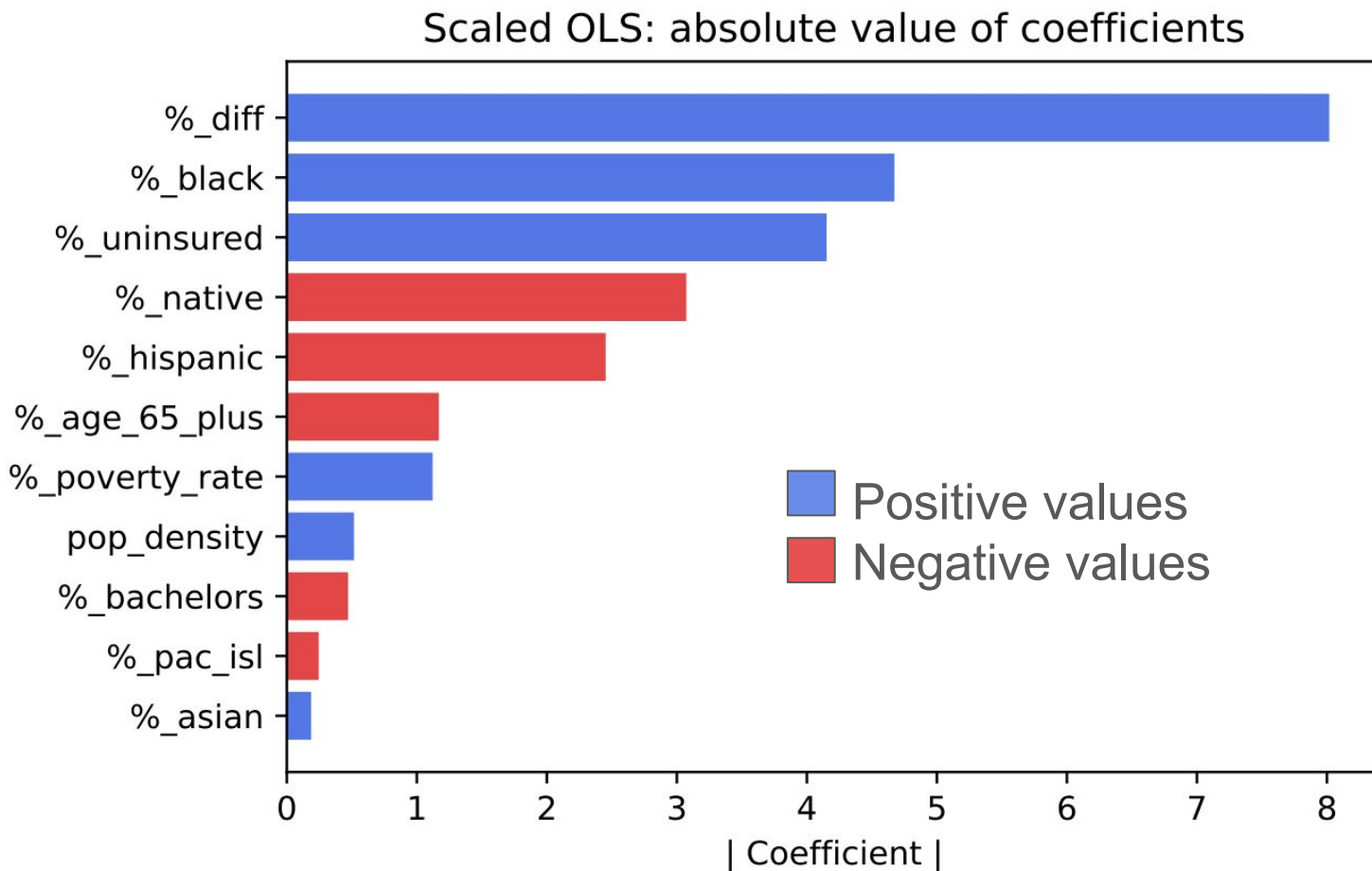
% Unvaccinated vs 2020 Vote Split (% diff) by county

% Unvaccinated

Biden   Trump

- Biden 30+
- Biden 10-30
- Biden 10 to Trump 10
- Trump 10-30
- Trump 30+

2nd largest coefficient

% Unvaccinated

% Black

Biden 30+
Biden 10-30
Biden 10 to Trump 10
Trump 10-30
Trump 30+

3rd largest coefficient

% Unvaccinated (y-axis)

% uninsured (x-axis)

Legend:
- Biden 30+
- Biden 10-30
- Biden 10 to Trump 10
- Trump 10-30
- Trump 30+
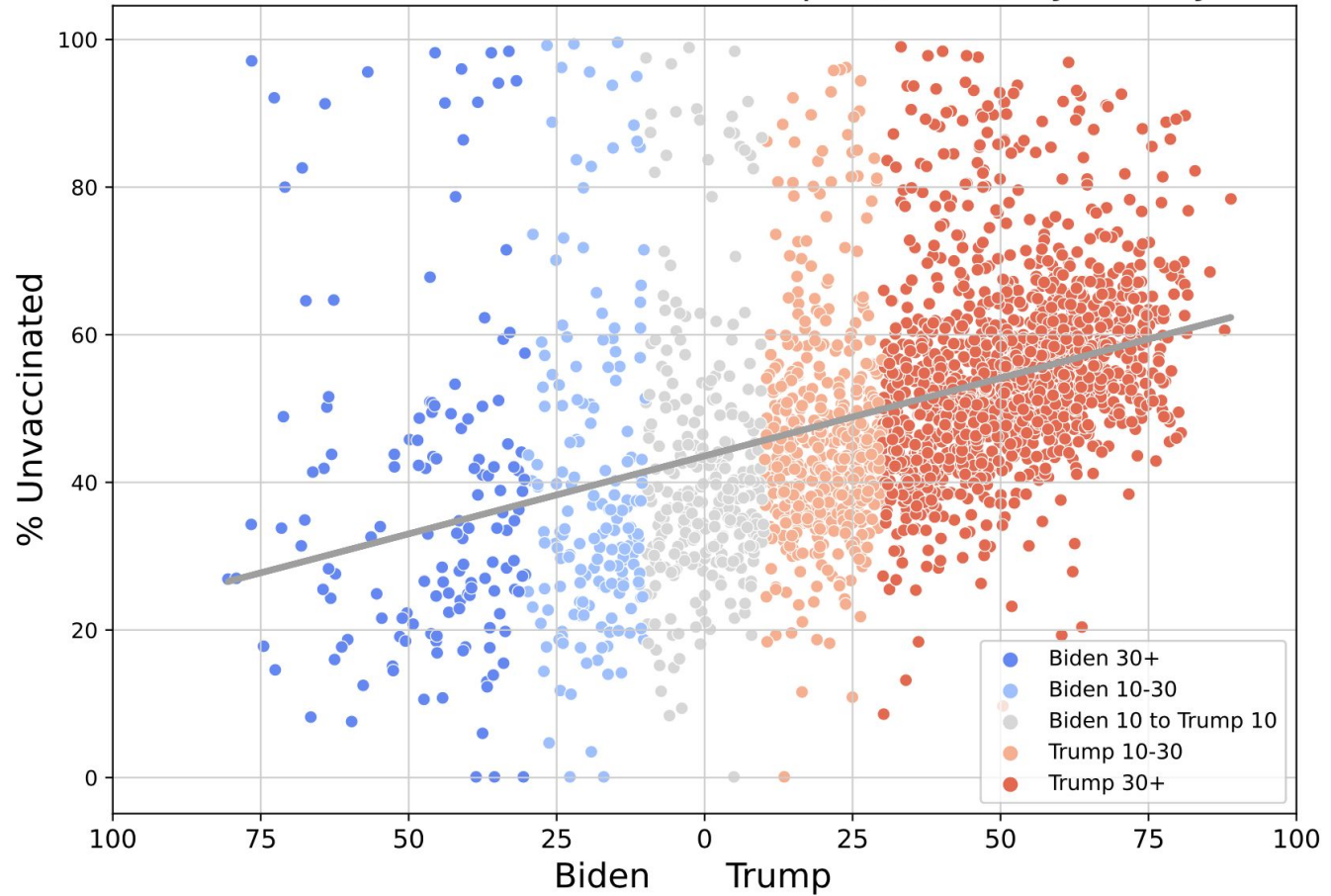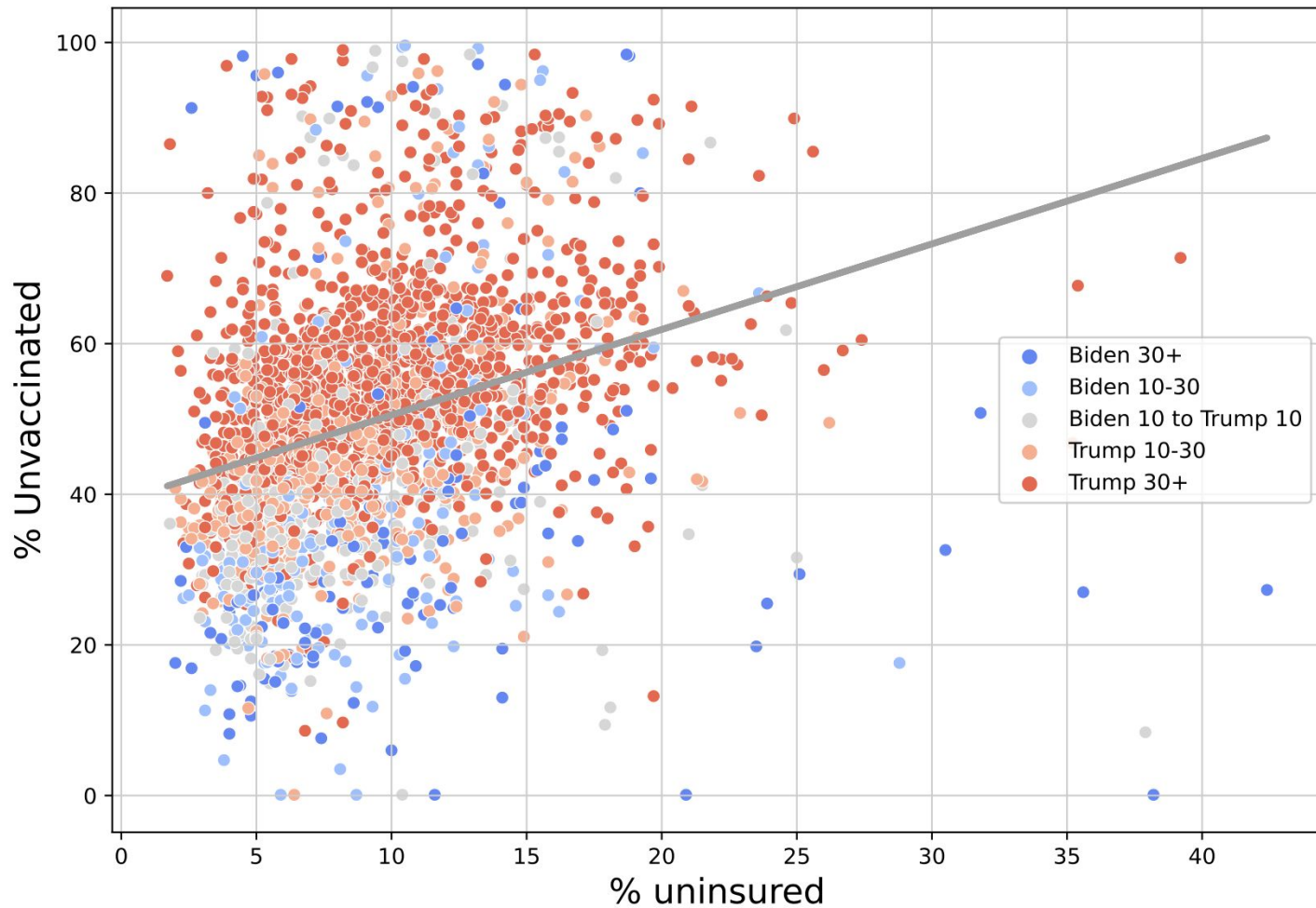
Conclusions

What political and socioeconomic factors best explain the variation in vaccination rates at the county level?

- vote split (% diff)

- % black population

- % uninsured

Future Work

- Try removing some of the features

- Try using more metrics

Thanks!

Urls:

https://covid.cdc.gov/covid-data-tracker/#vaccinations-county-view

https://github.com/tonmcg/US_County_Level_Election_Results_08-2

https://www.openintro.org/data/?data=county_complete