

Clusters of Counties in COVID-19

• • •

Ankita Patil, Jacqueline Simeone, Andy Orfalea & Annie Bishai

Questions

- Can clustering US counties based on COVID vaccine data and demographic makeup give meaningful insights into the political divide in the country?
- Do these clusters mimic this divide?
- Can we predict which way a county voted in 2020 by using COVID vaccine data?

Data

- Sources: CDC (vaccine data), Kaggle (election data), Healthdata.org (additional vaccine data)
- Vaccine rates 3 points in time
- Vaccine hesitancy
- Political affiliation (2020 election)
- Uninsured percentage
- Census Data
- Metro/Non-metro
- Social Vulnerability Index (1-5)

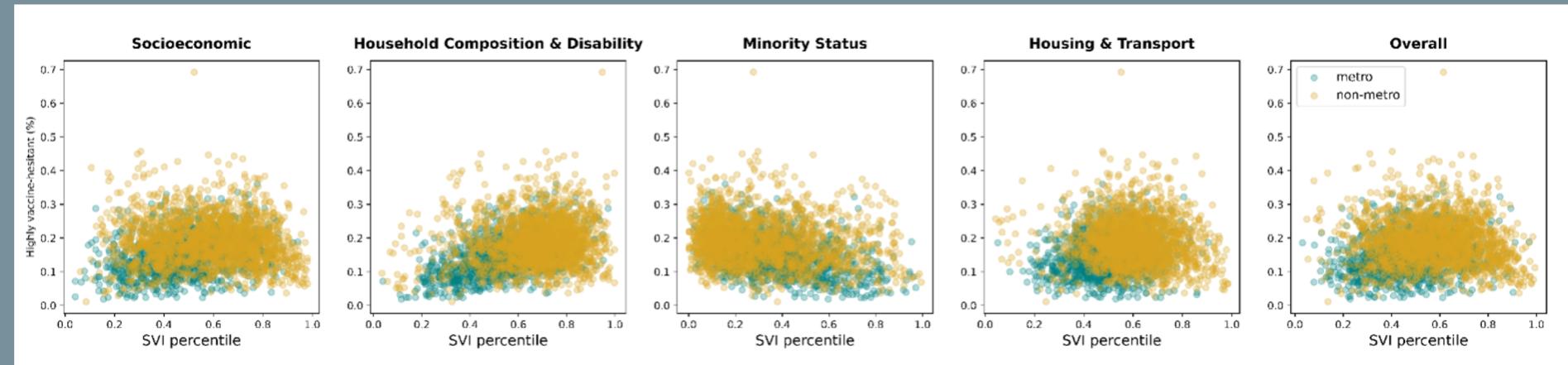
Missing Data

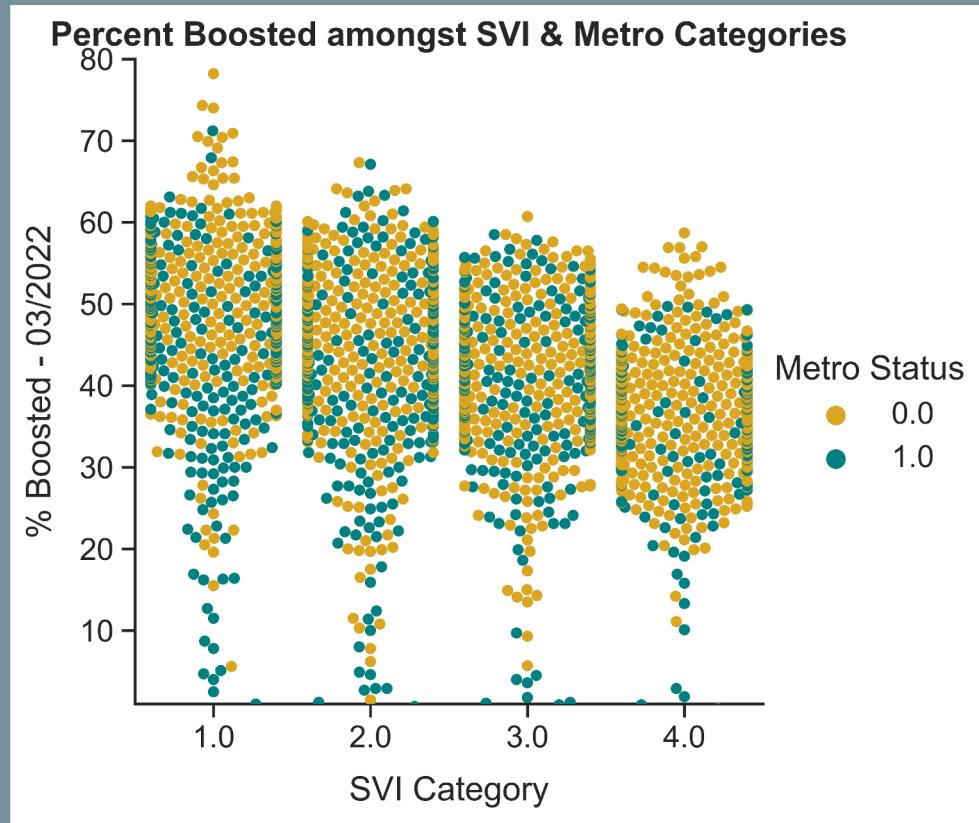
State-Level:

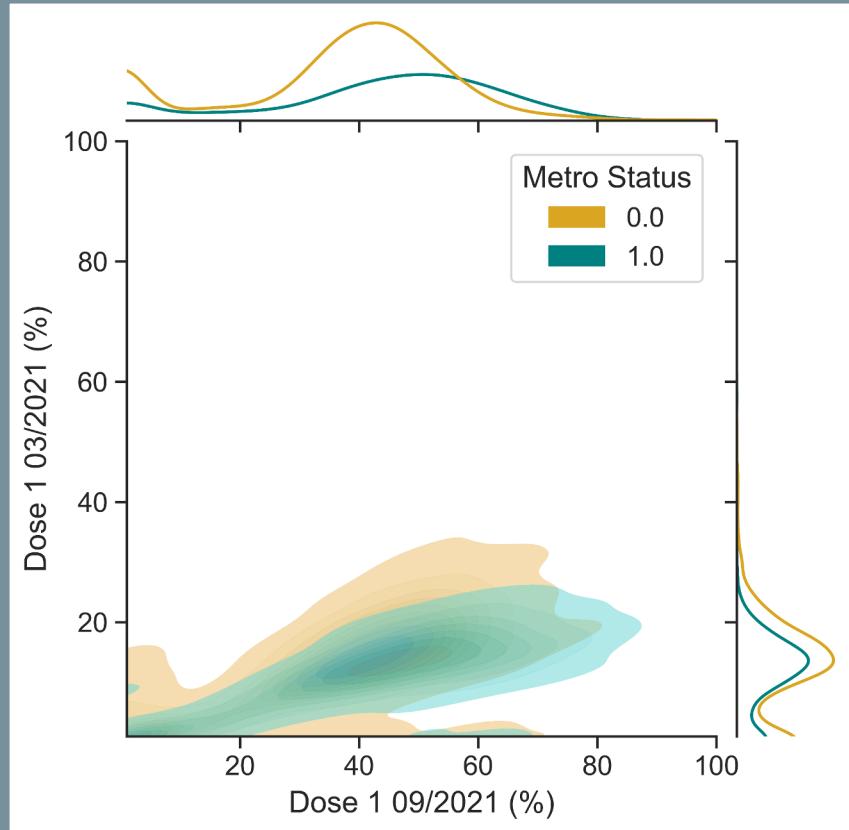
- Systematic missing data in some states
- Vaccine administered outside the county of residency

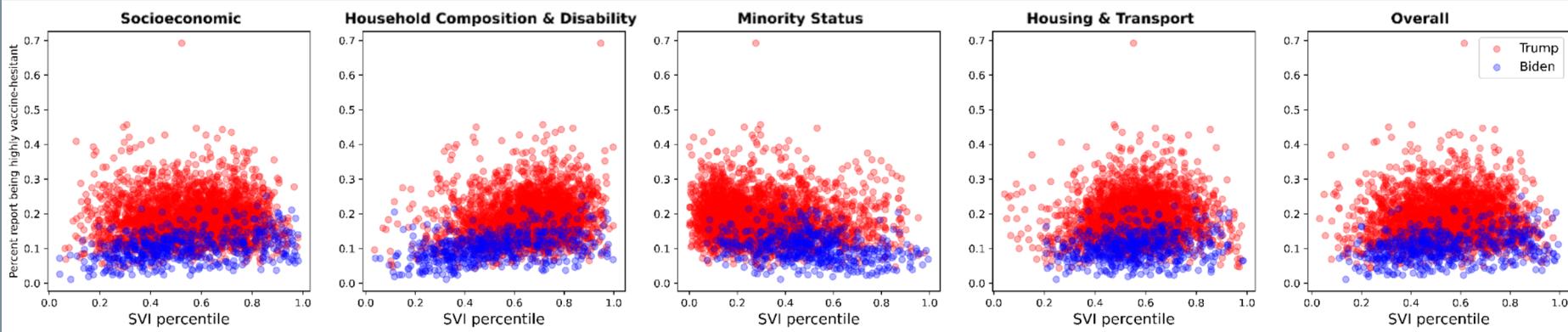
County-Level:

- California
- Hawaii
- Massachusetts
- New Hampshire

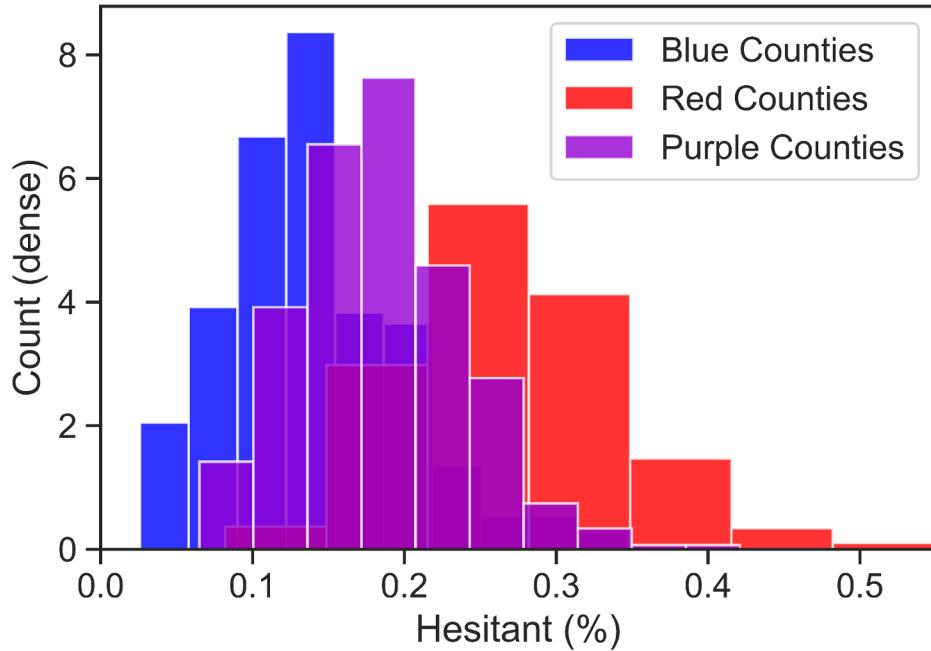




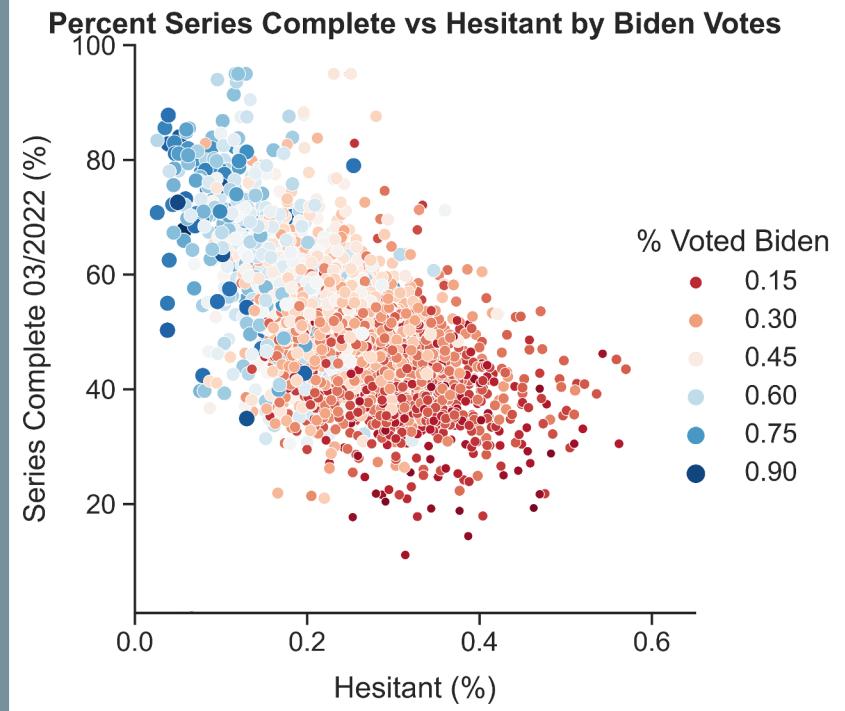




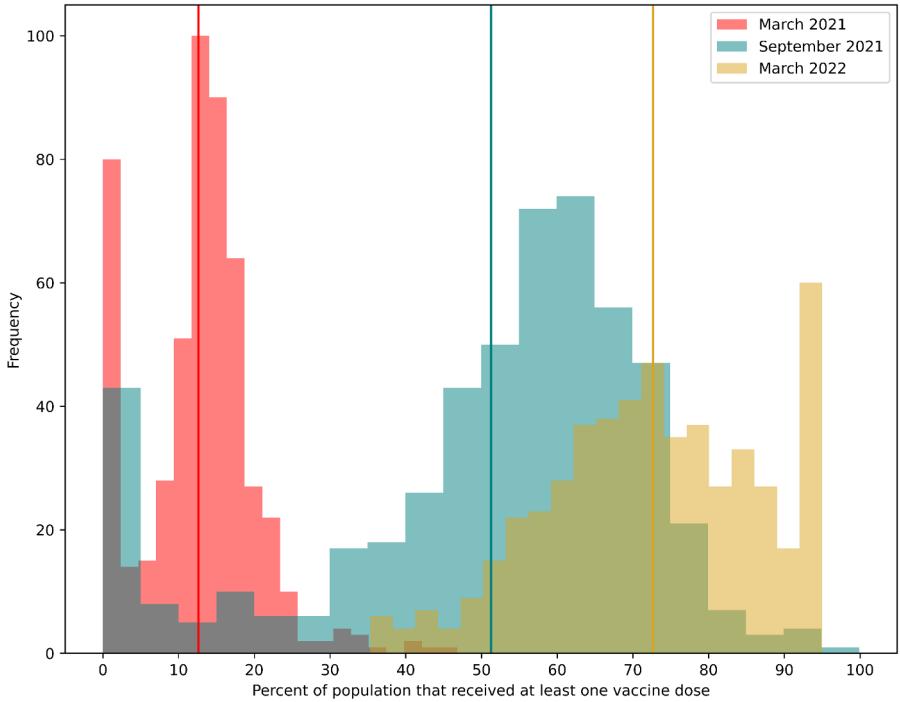
Red, Blue, & Purple County Hesitancy Percent



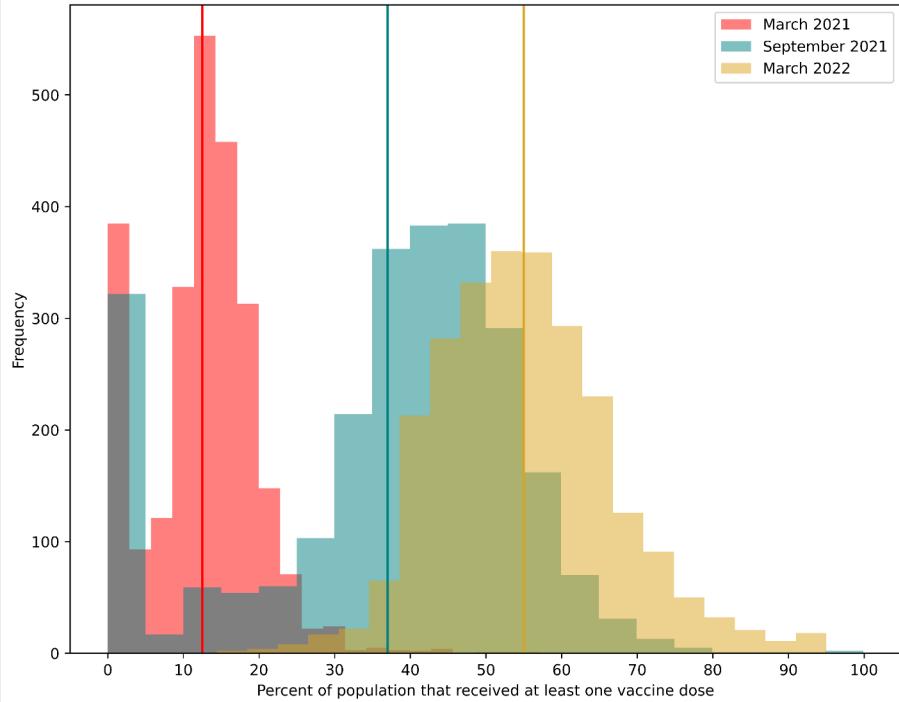
Percent Series Complete vs Hesitant by Biden Votes



Histogram over Time of Vaccination Rates in Counties Won by Biden



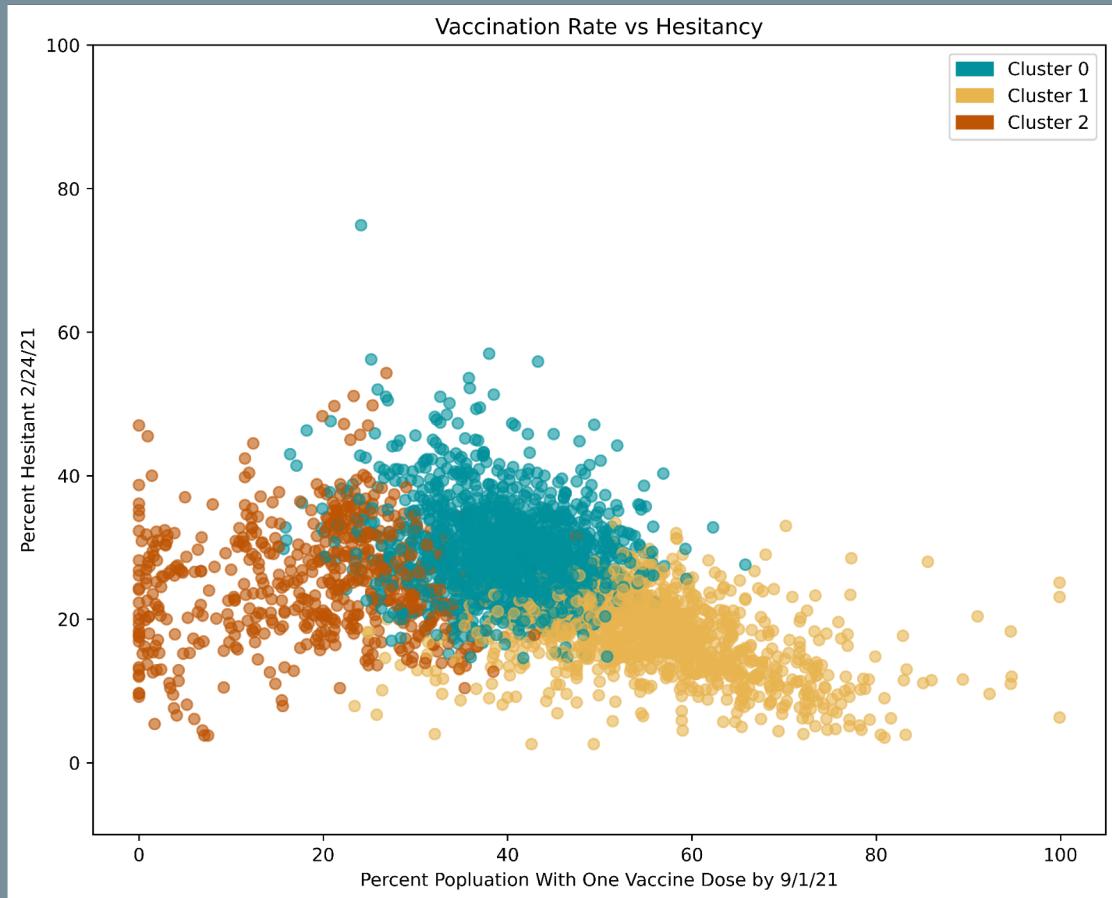
Histogram over Time of Vaccination Rates in Counties Won by Trump



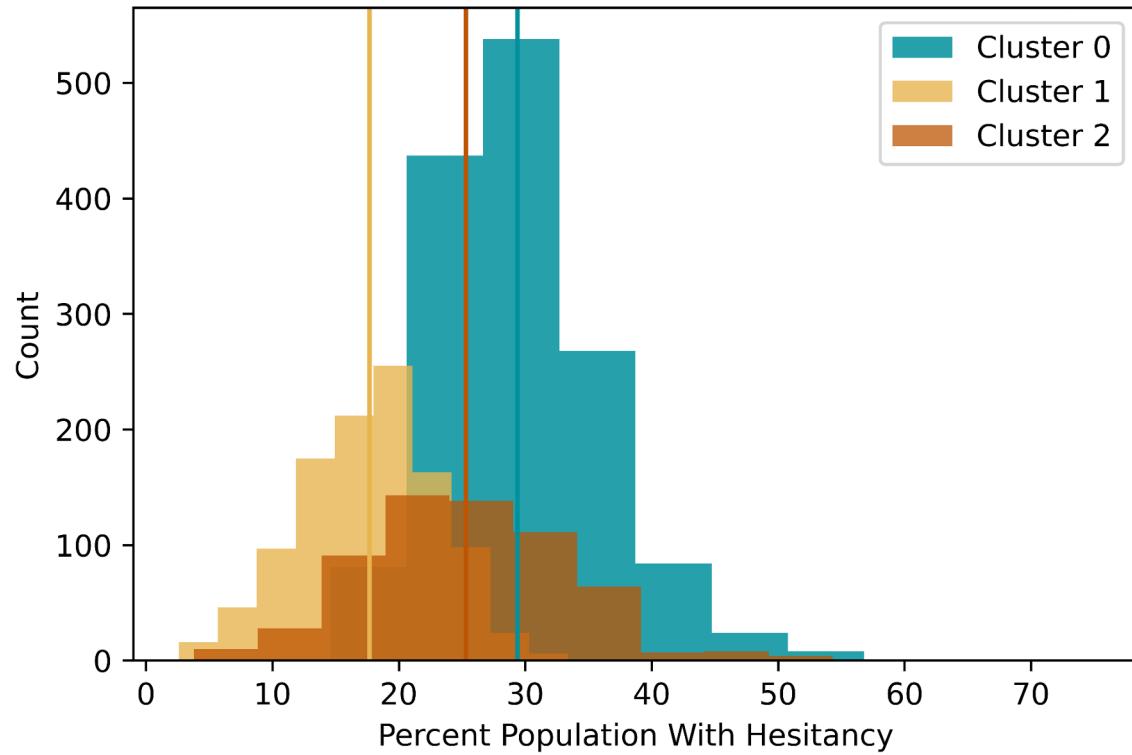
KMeans Clustering Model

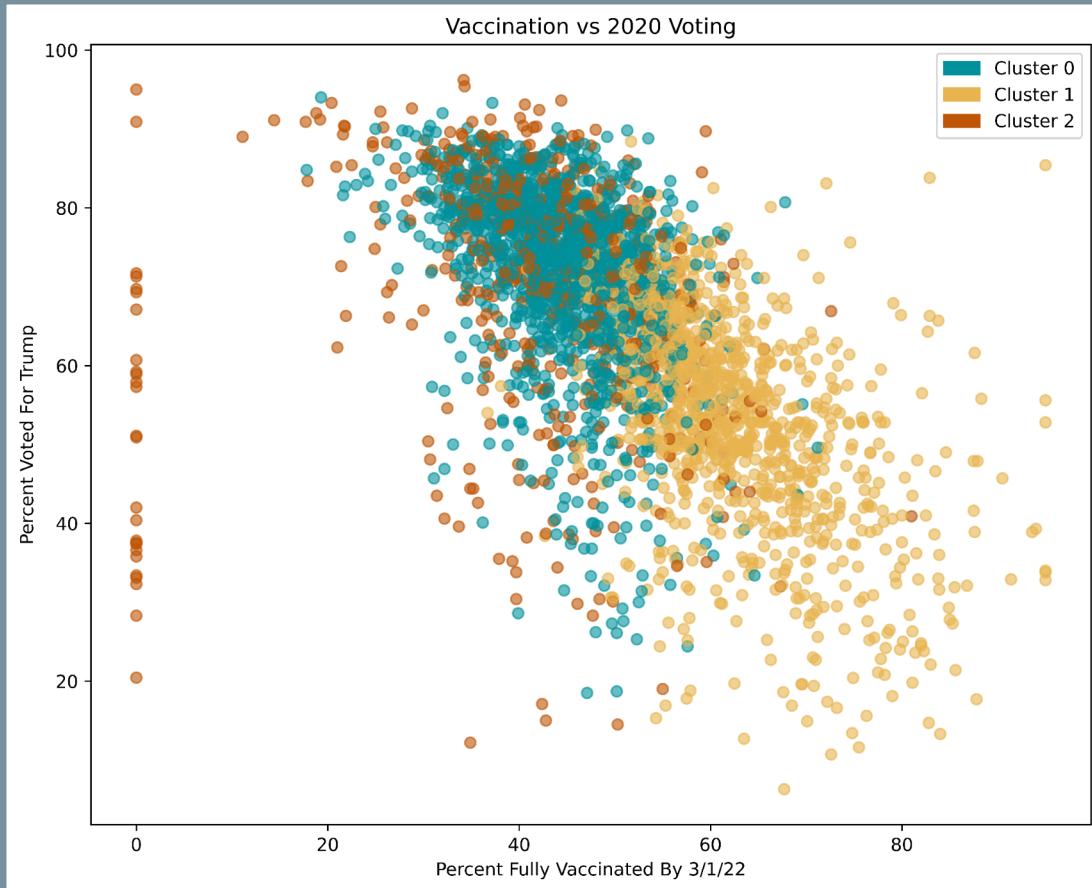
FEATURES

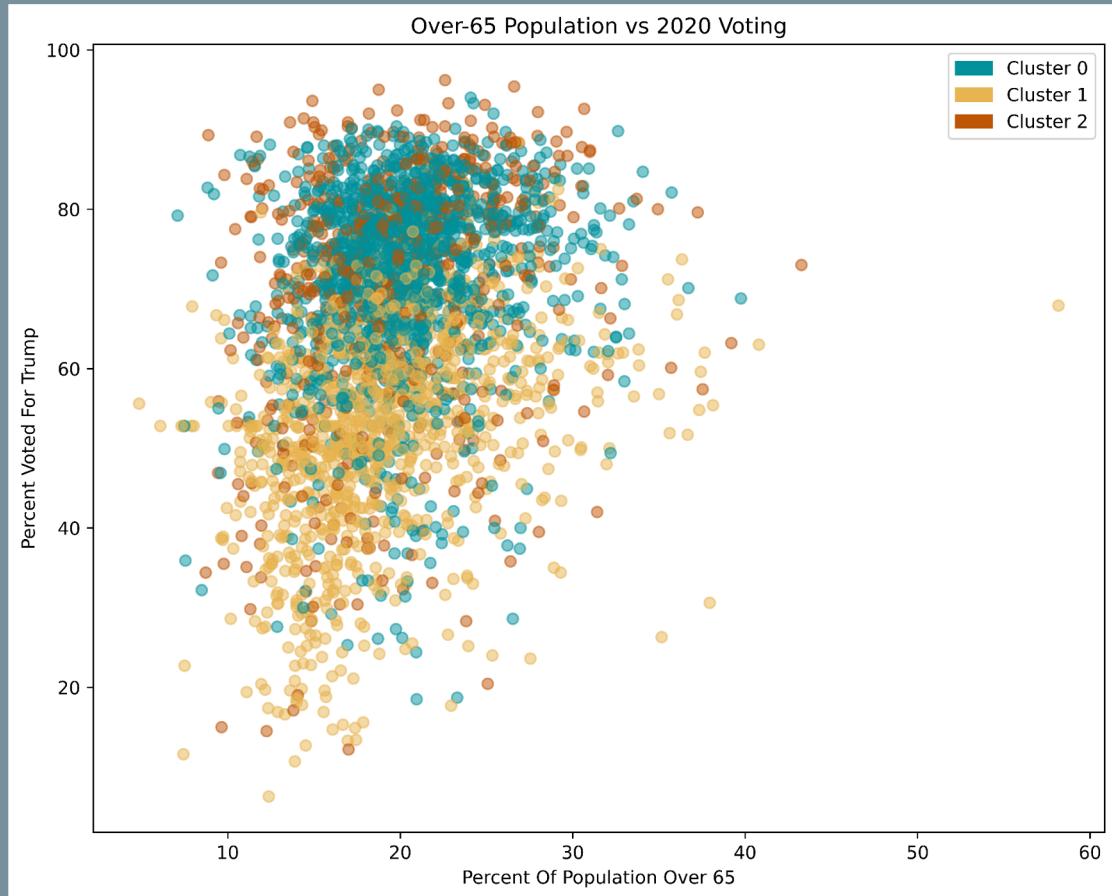
Vaccine hesitancy %	% of population that's taken 1 dose of Covid vaccine
Metro Status	% of population that's completed a vaccine series
% of population 65 years old +	% of population that's taken a booster

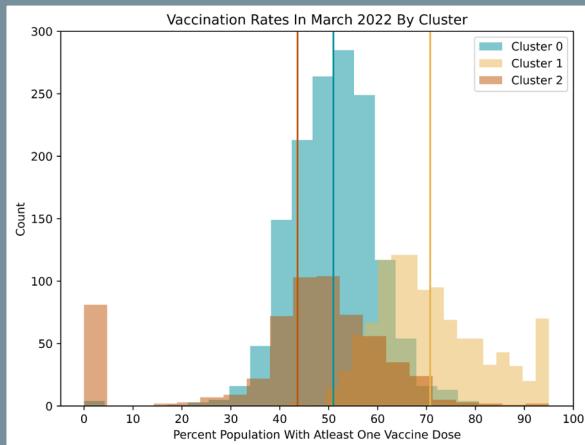
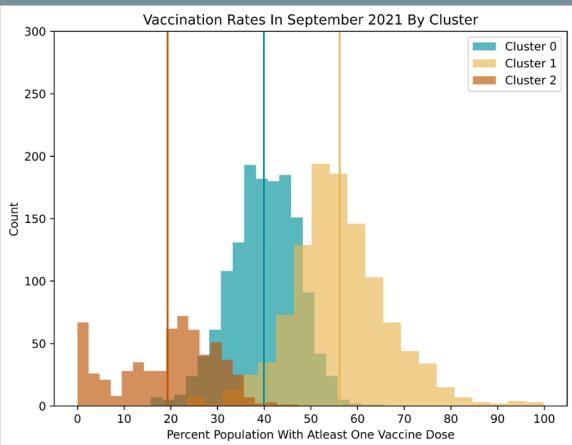
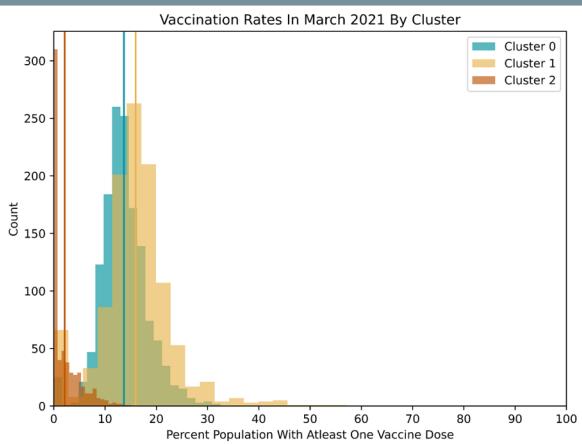


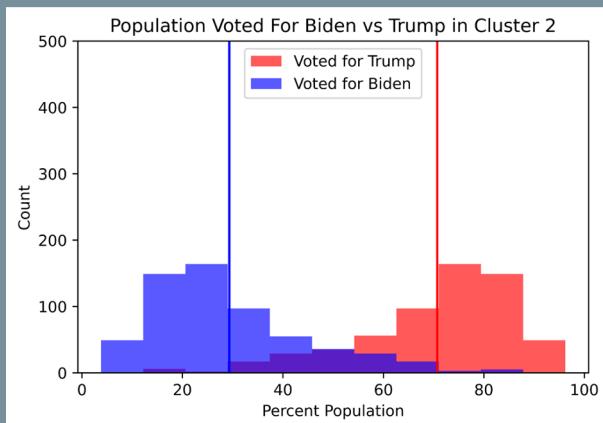
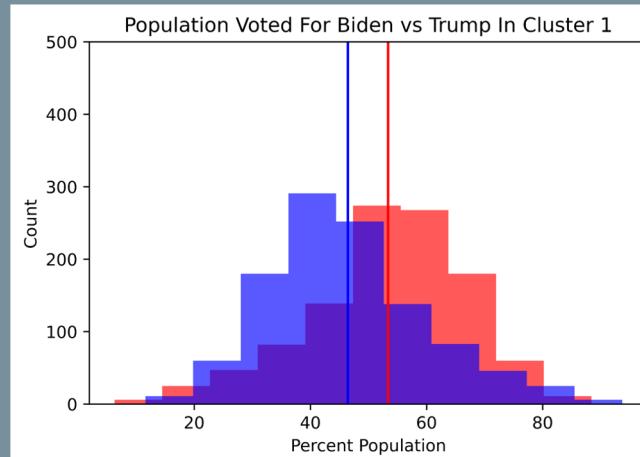
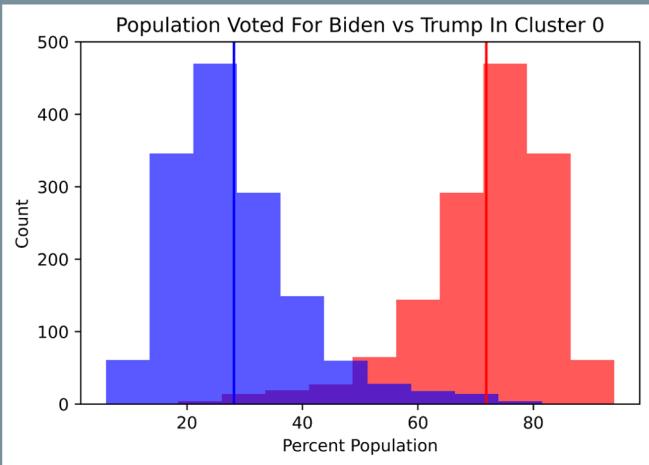
Population With Covid Vaccine Hesitancy

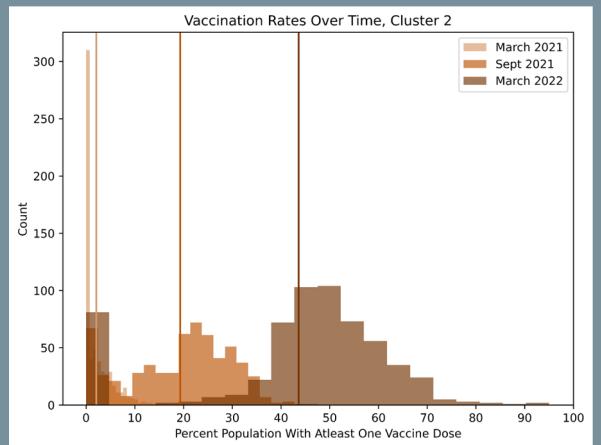
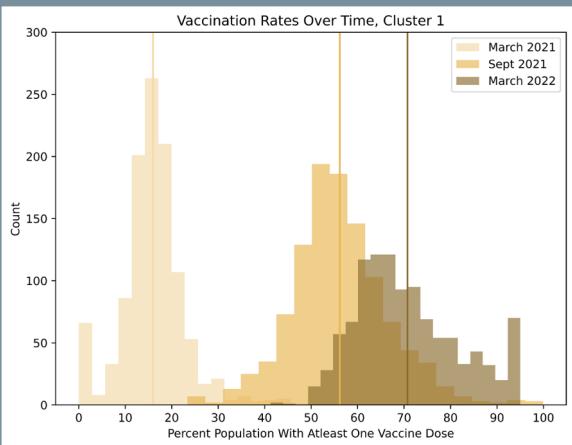
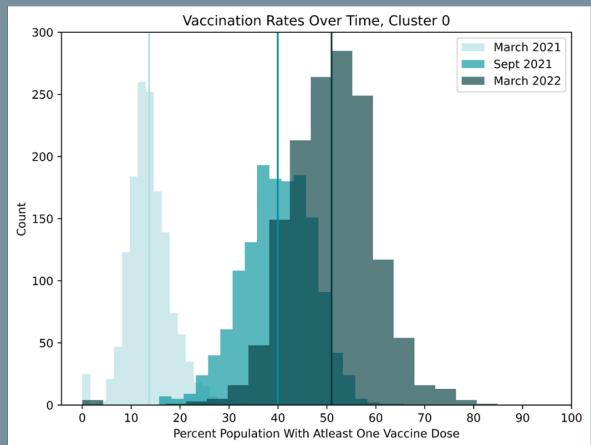


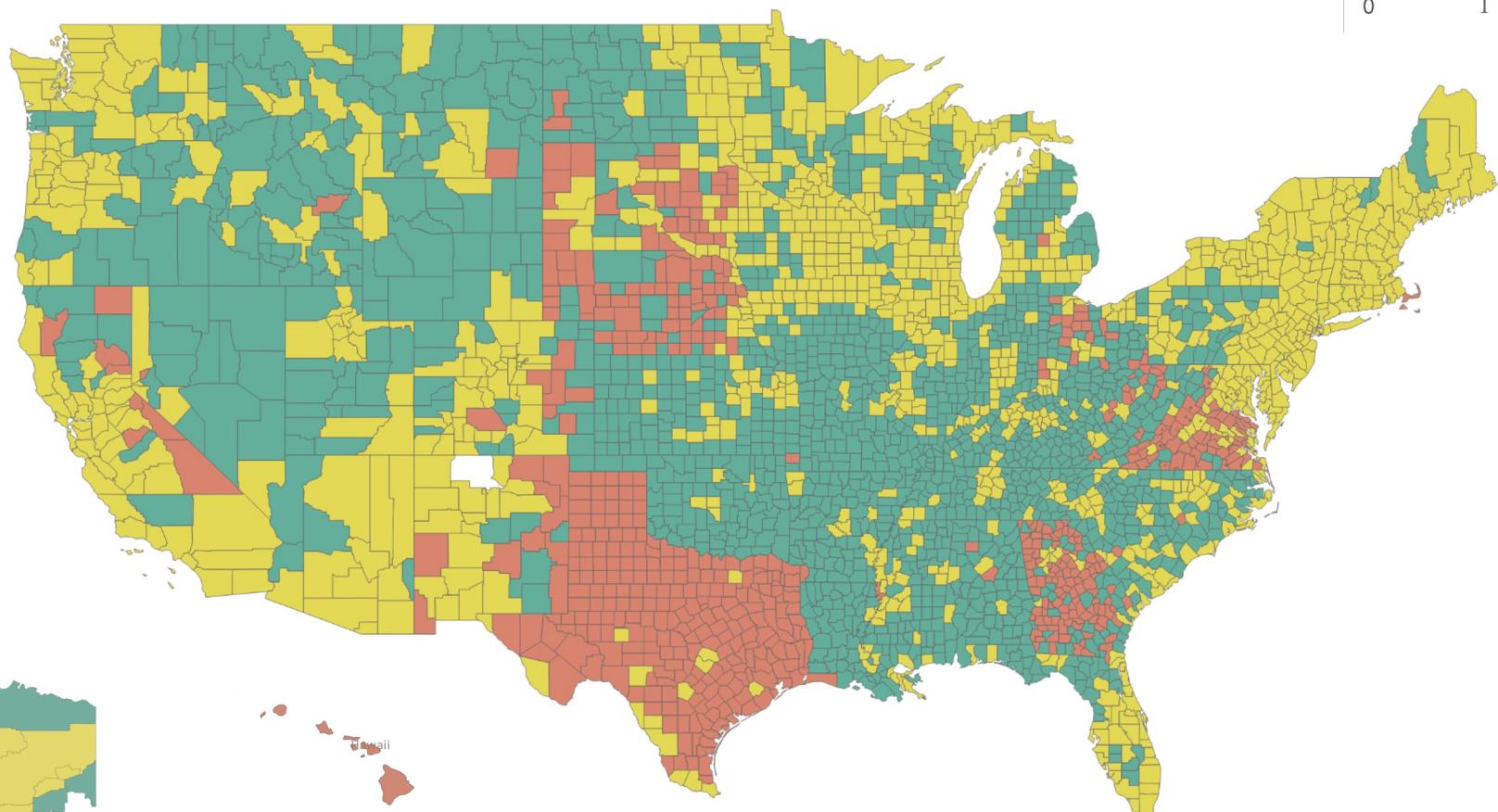
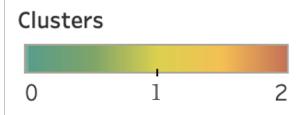


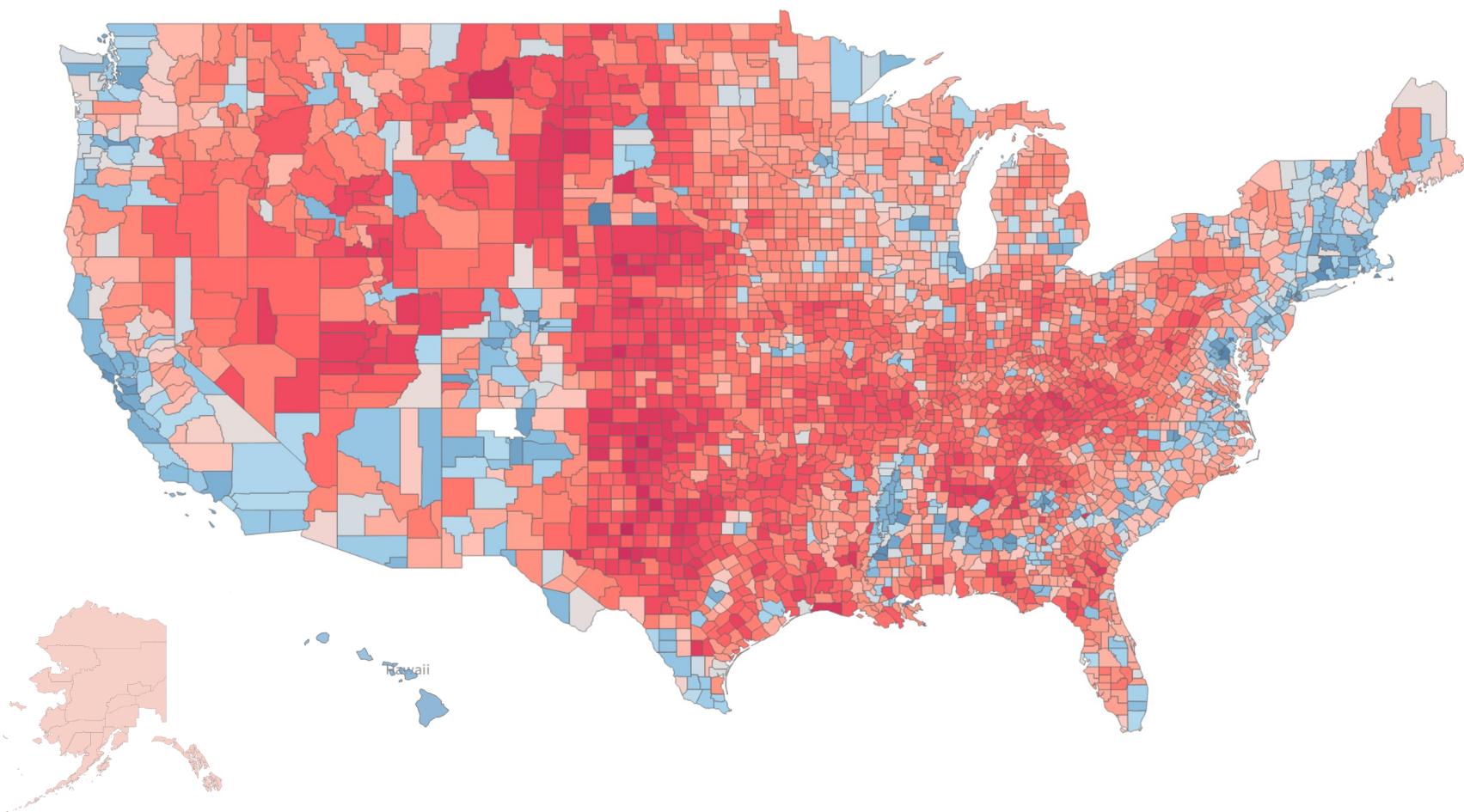
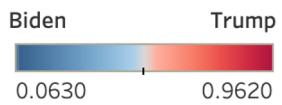






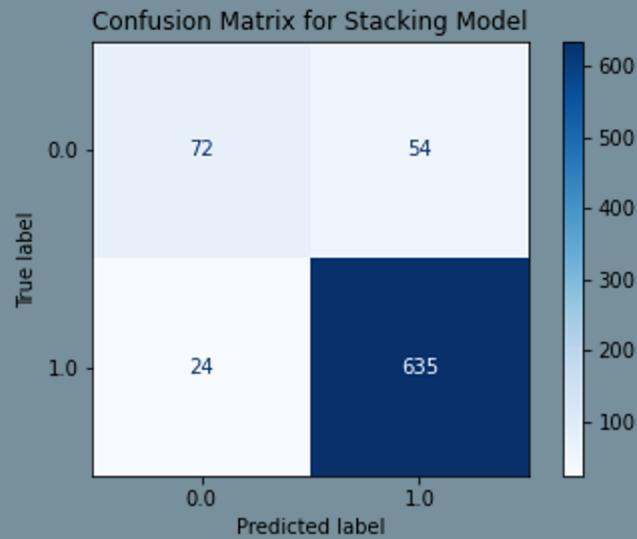






Classification Model

- Covid Vaccine Data
- Baseline Model = 83.1% Accuracy
- Bagging, Random Forest = 88.7% Accuracy
- Stacking Model = 90.3% Accuracy



Conclusions

- Clear and distinct themes within the clusters show how stratified county responses to the covid vaccine were.
- However, it's not a straight 50/50 split, there is some nuance to county reaction as our 3 clusters show.
- Our hypothesis that you could predict which way a county voted in 2020 based on covid vaccine data was correct.

Next Steps

- Strive for improvement in data collection
- Compile more data (mask hesitancy)
- Investigate if covid vaccine discourse has affected non-covid vaccination rates
- Perform a similar clustering analysis for other countries and see how the cluster makeup compares to our clusters

Thank You