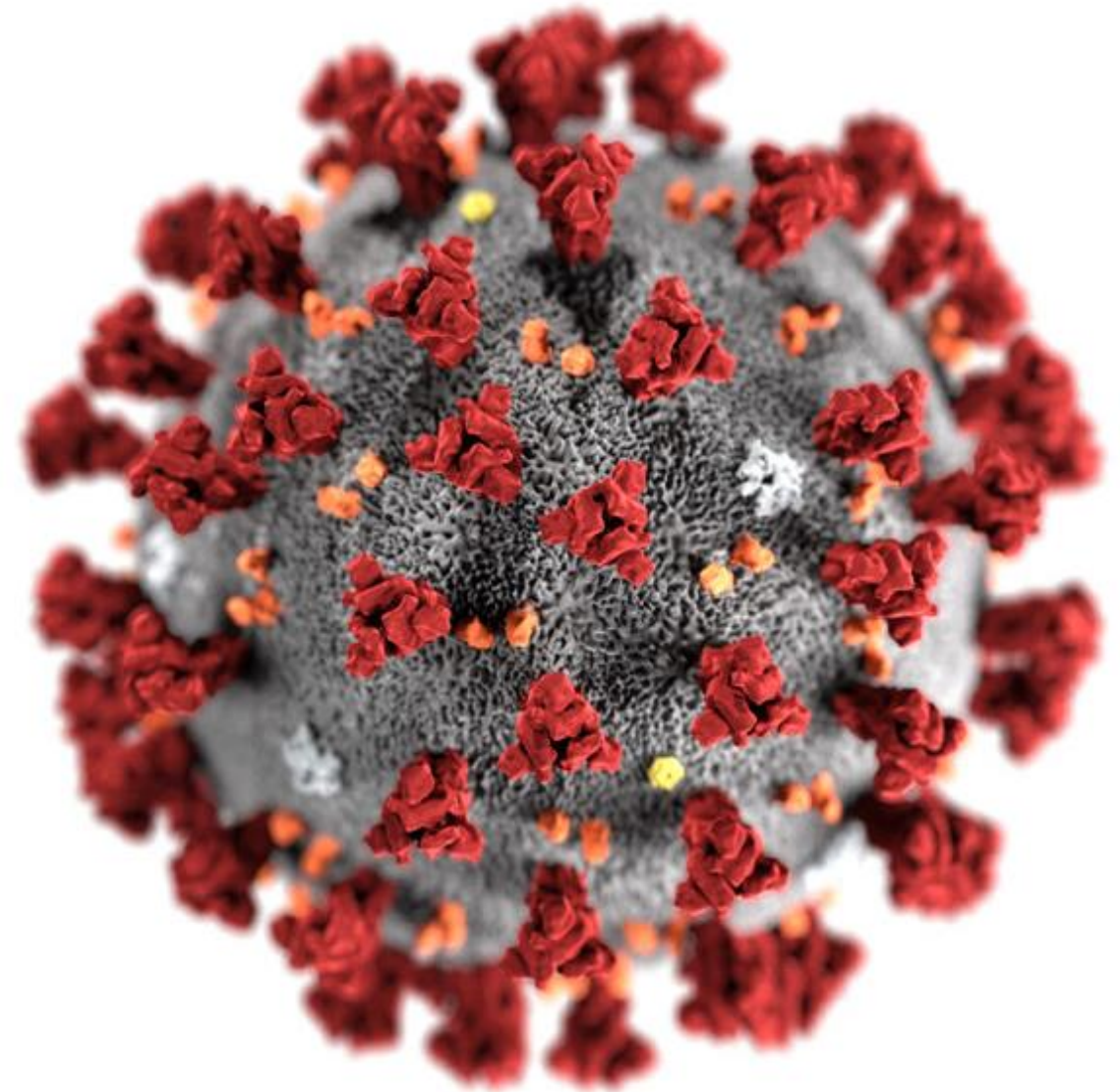
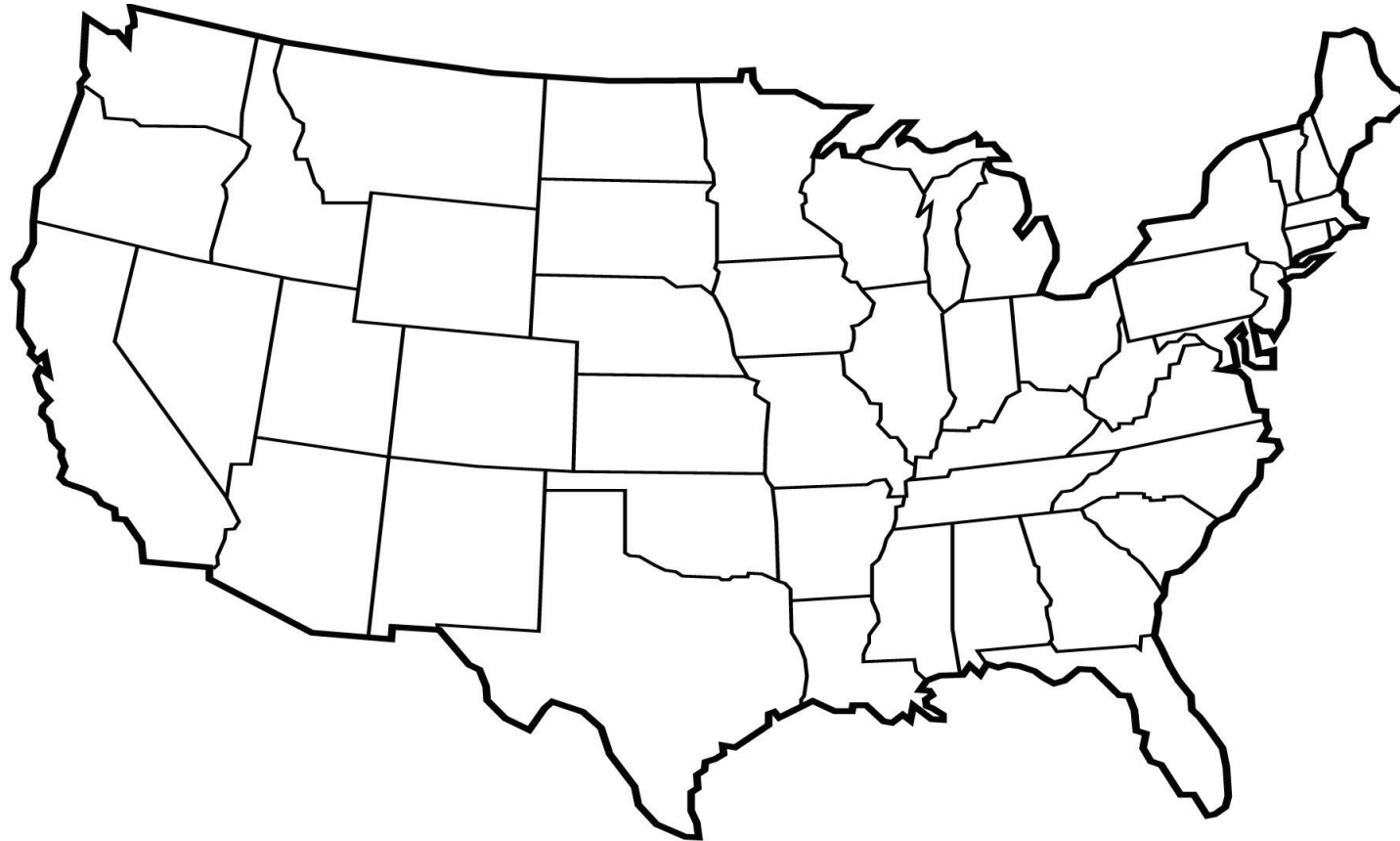


Predicting COVID-19 Using Demographic Data

Caroline Clark, Feras Atwal, James Lee
October 30th, 2020



Can we predict COVID-19 severity using demographic data?



Project Pipeline



DATA
COLLECTION



DATA PRE-
PROCESSING



DATA
VISUALIZATION



MODELING



MAKING THE DATA
INTERACTIVE

Data Collection

County-Level

Area

Population Density

Demographics

Age

Gender

Race

Economic Indicators

Income Per Capita

Health Insurance

Household Income

COVID-19

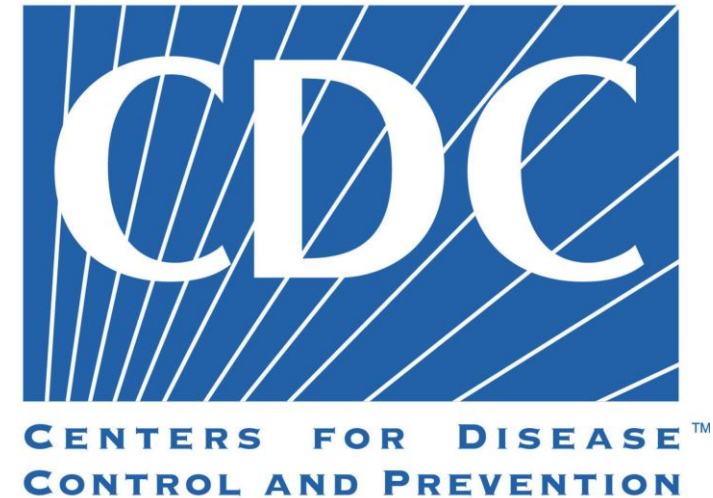
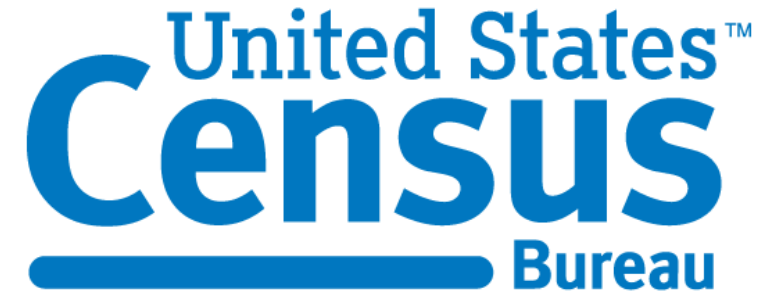
Total Tests

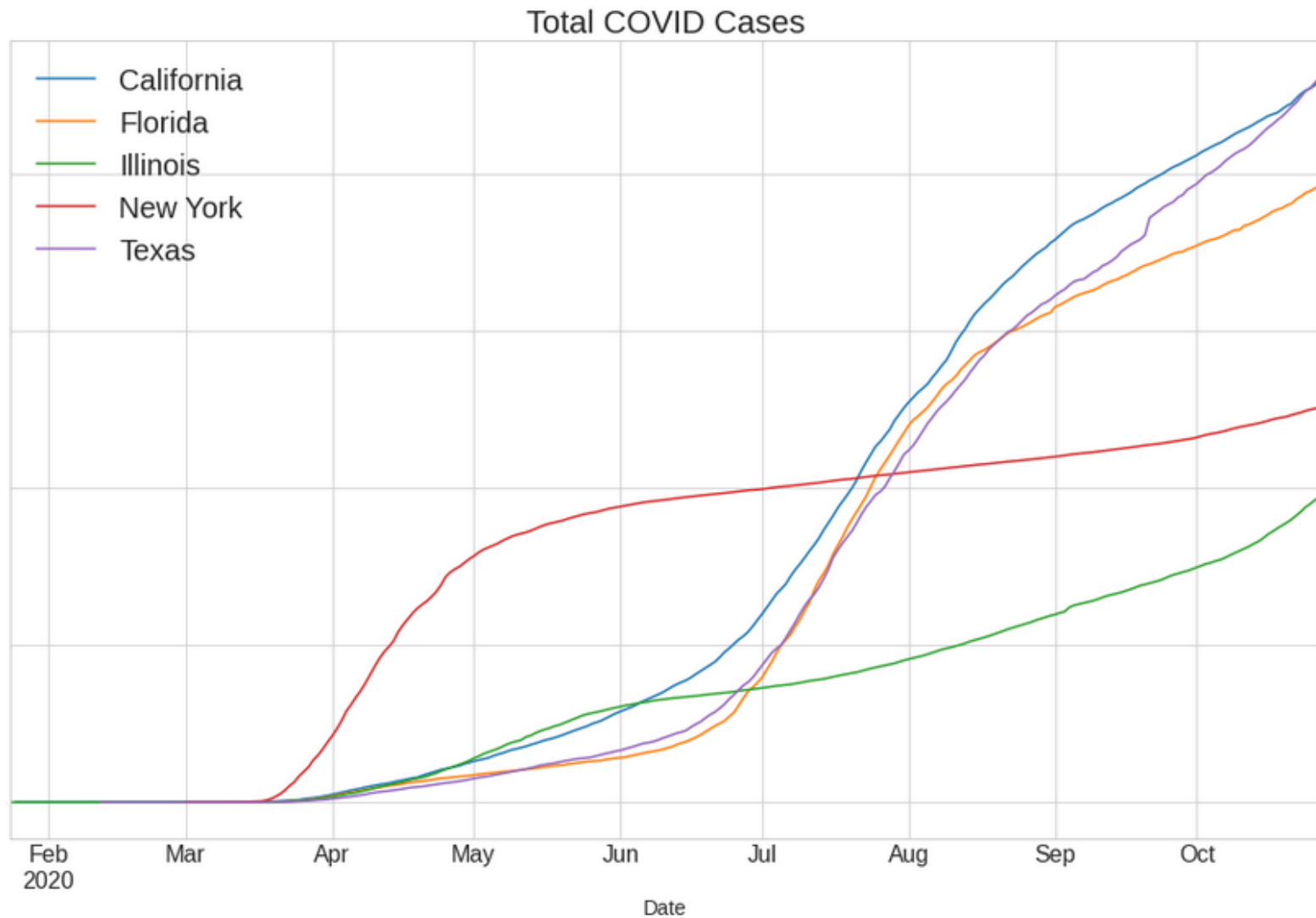
Total Cases

Total Deaths

Health Indicators

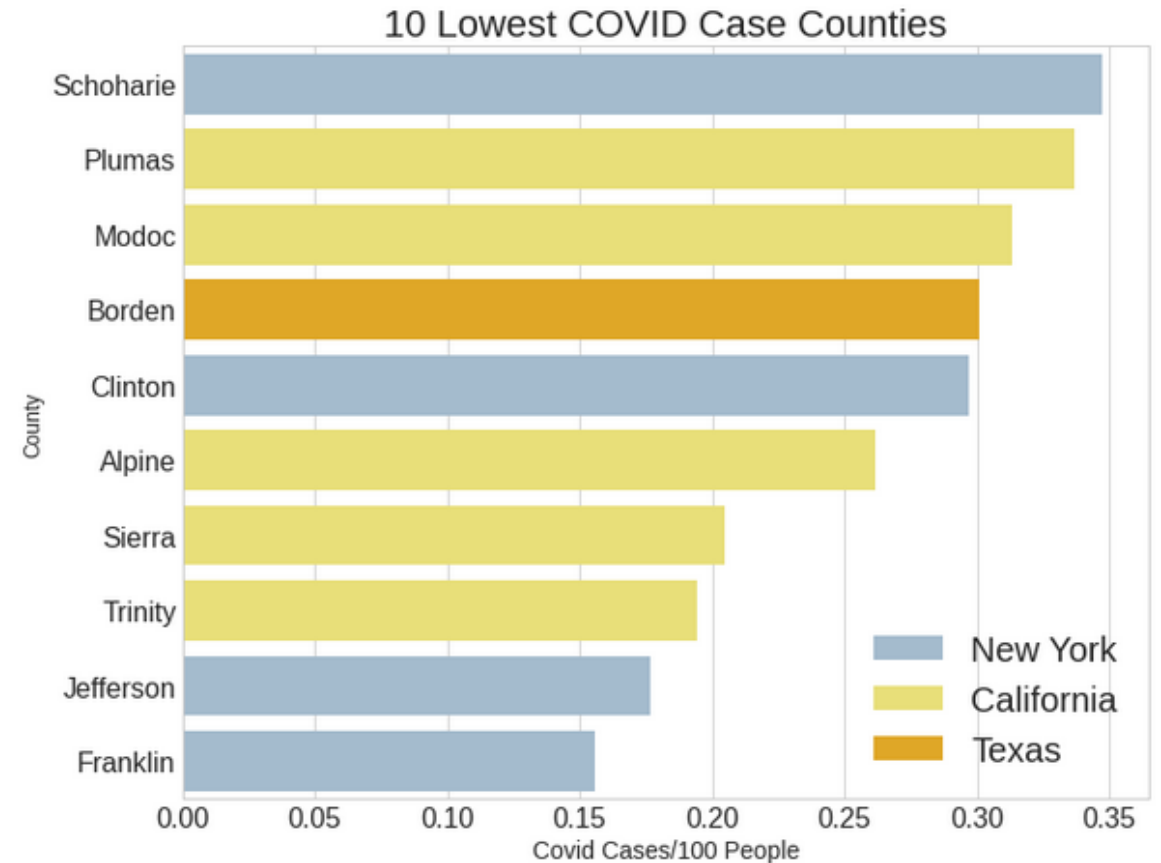
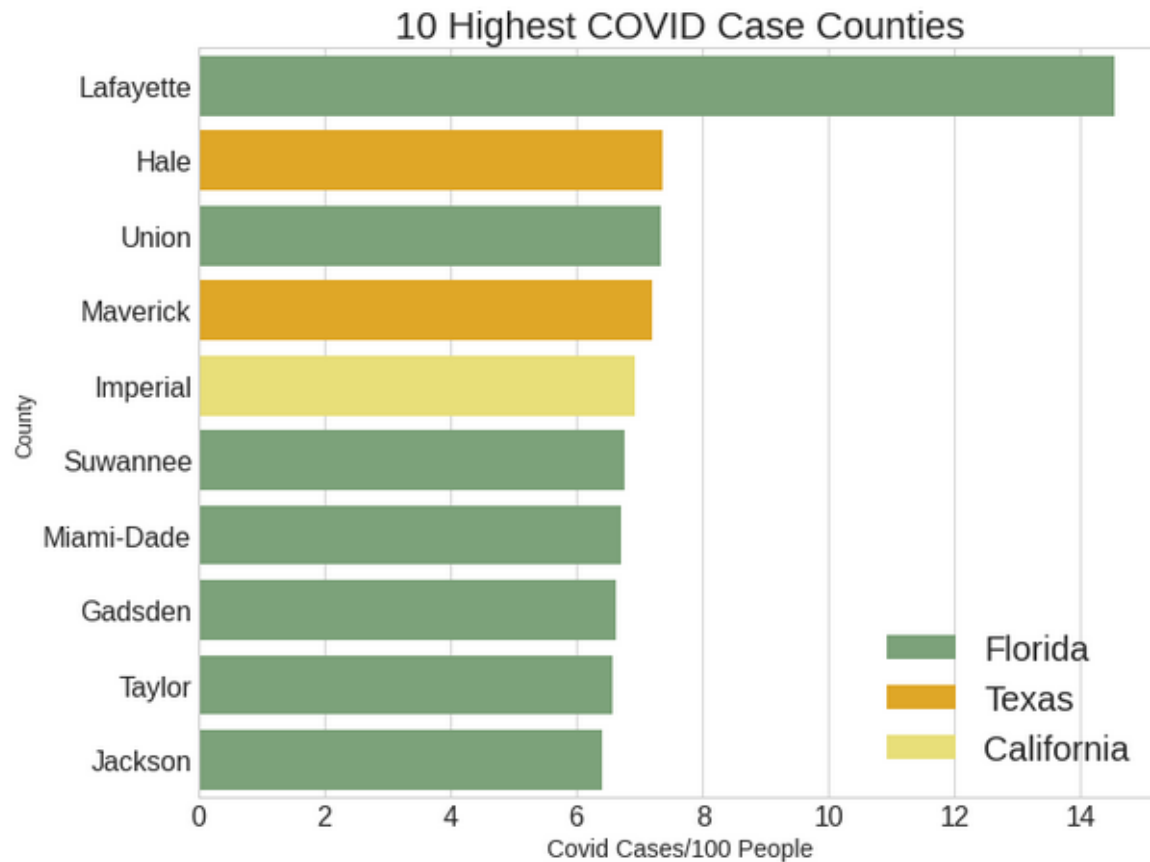
Obesity Rates



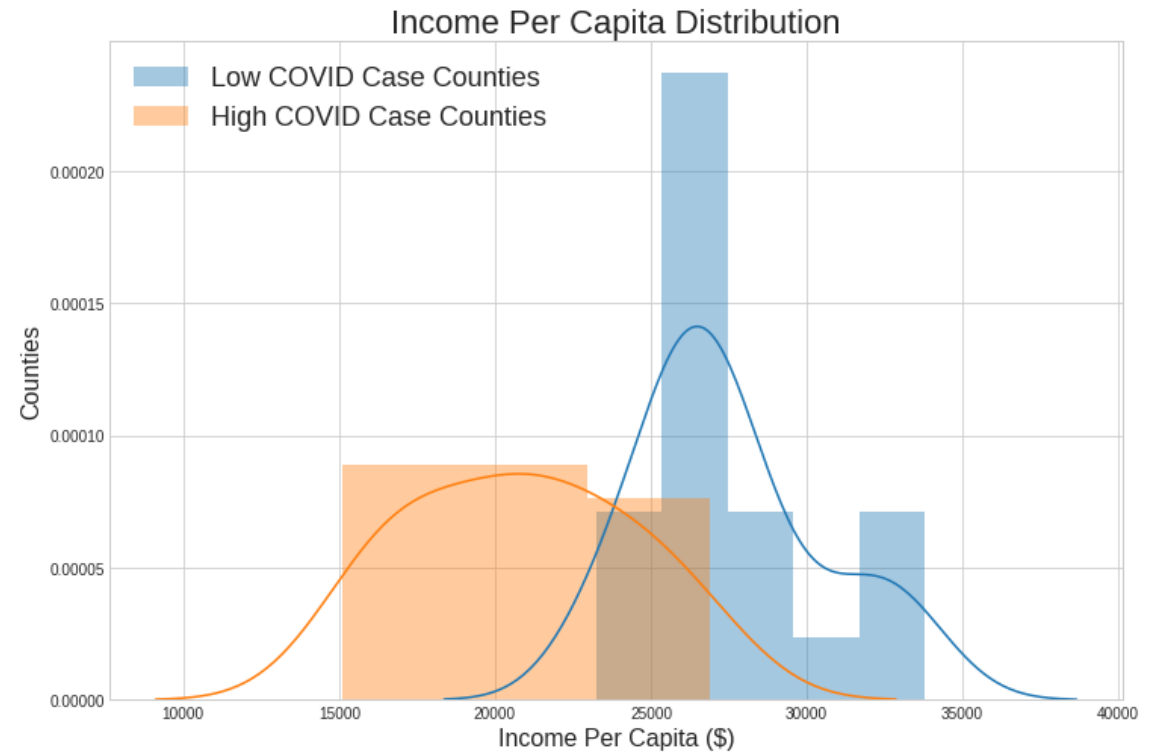
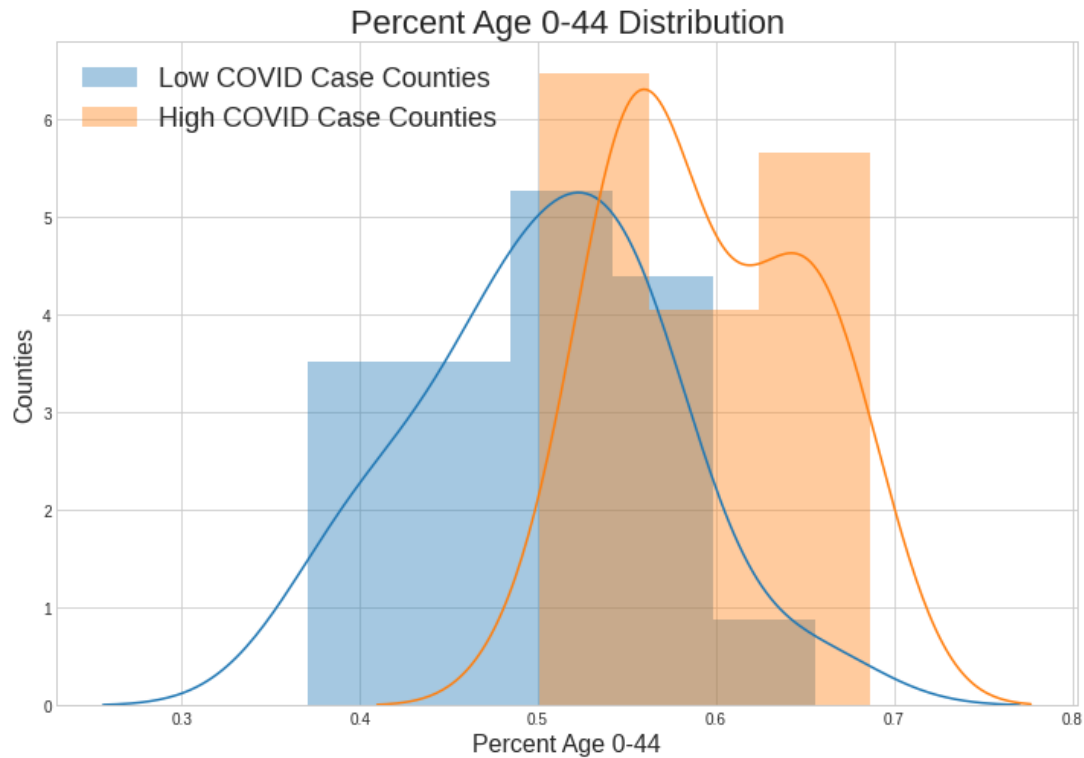


Five States
with the
Most
COVID-19
Data

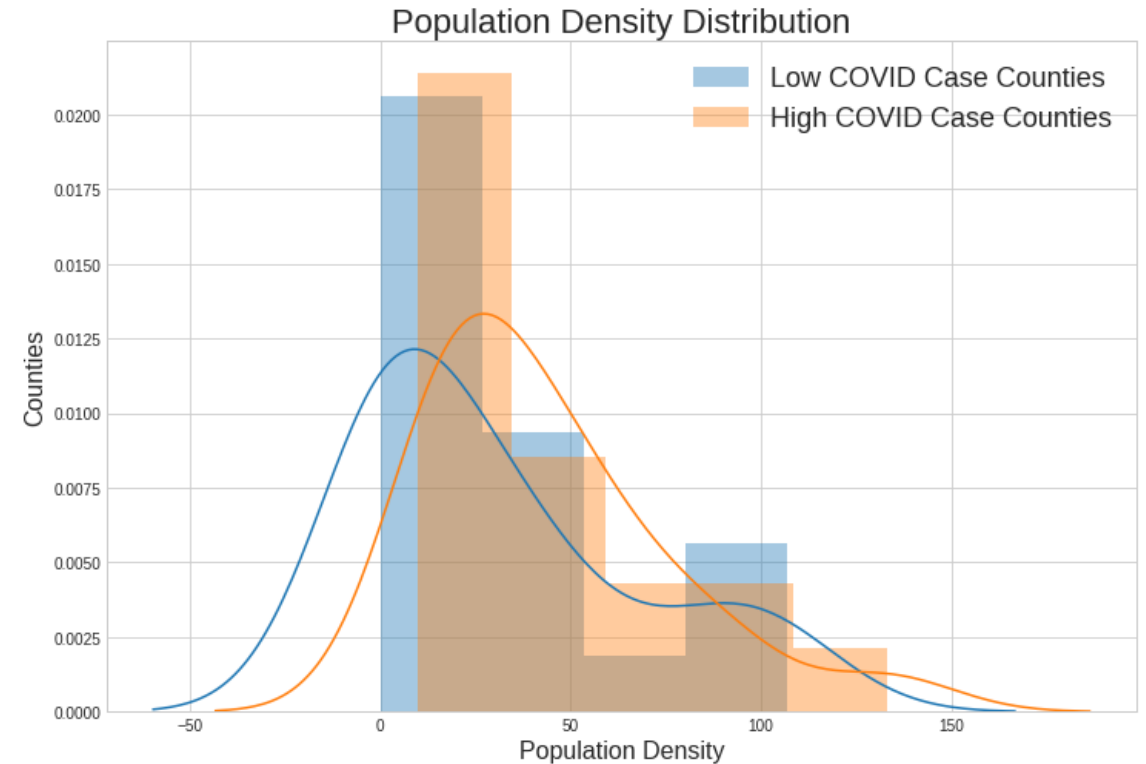
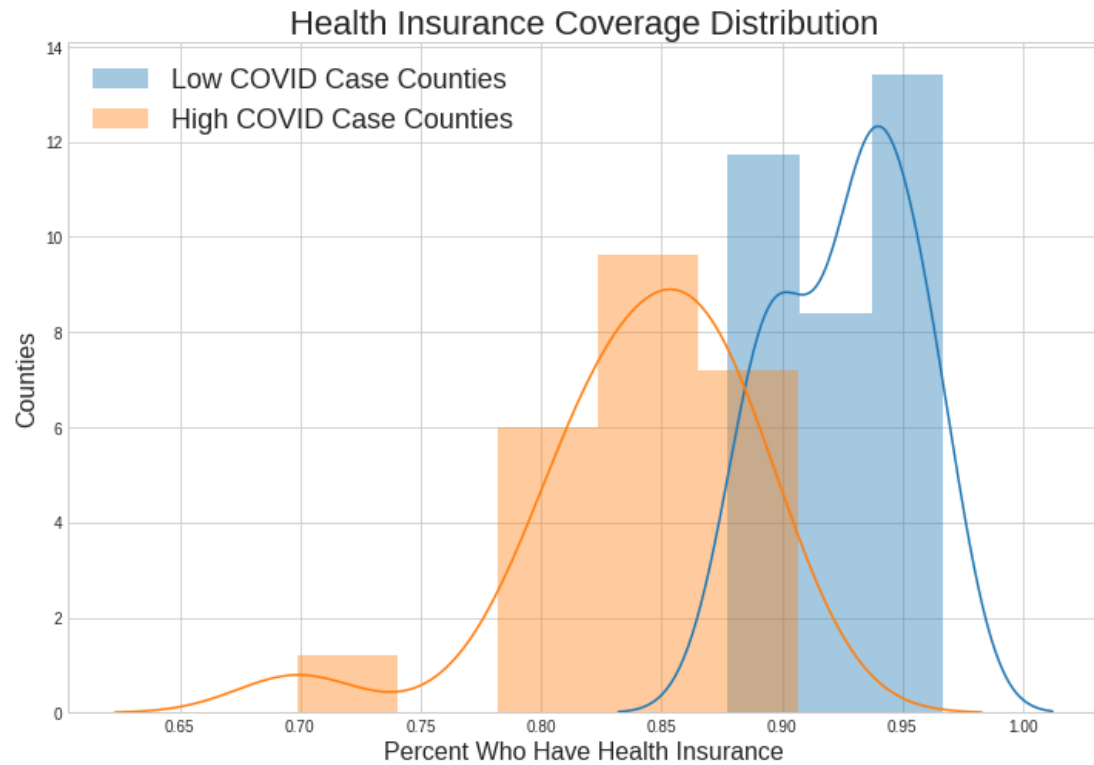
COVID-19 Statistics Vary Widely Among Counties



High COVID Counties Likely to be Younger, Have Lower Income Per Capita



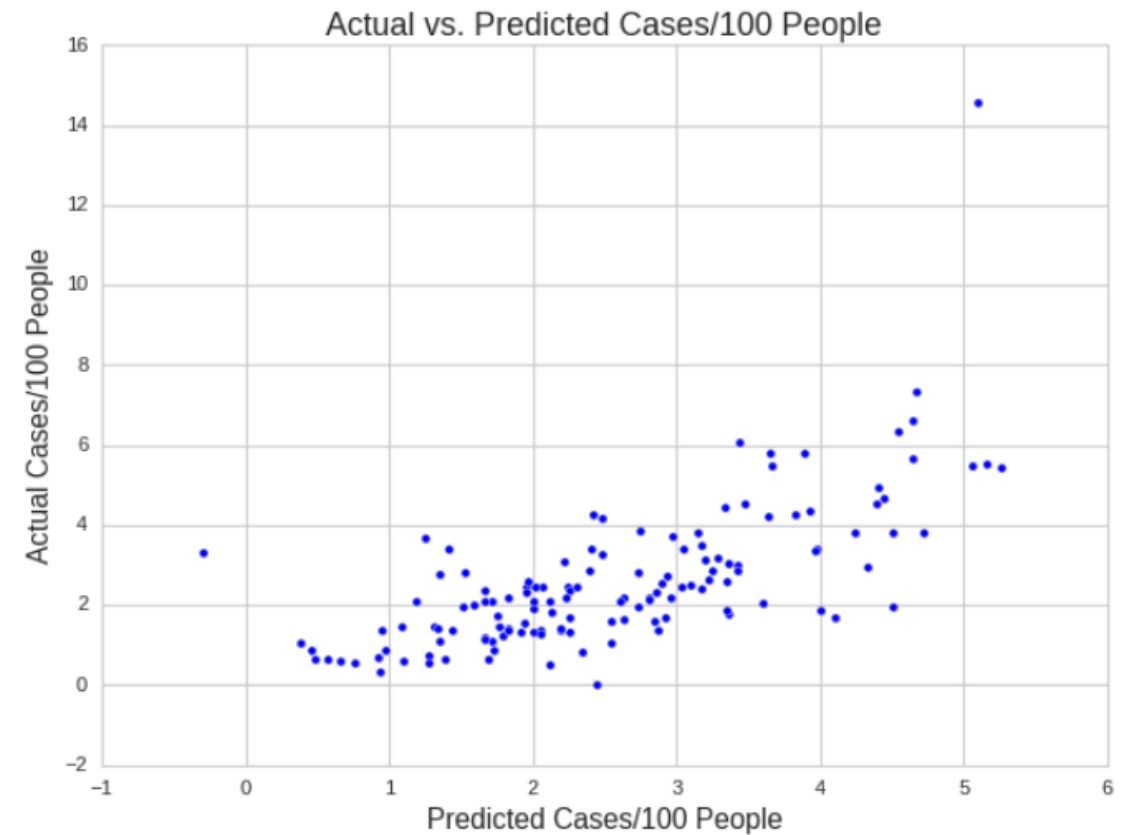
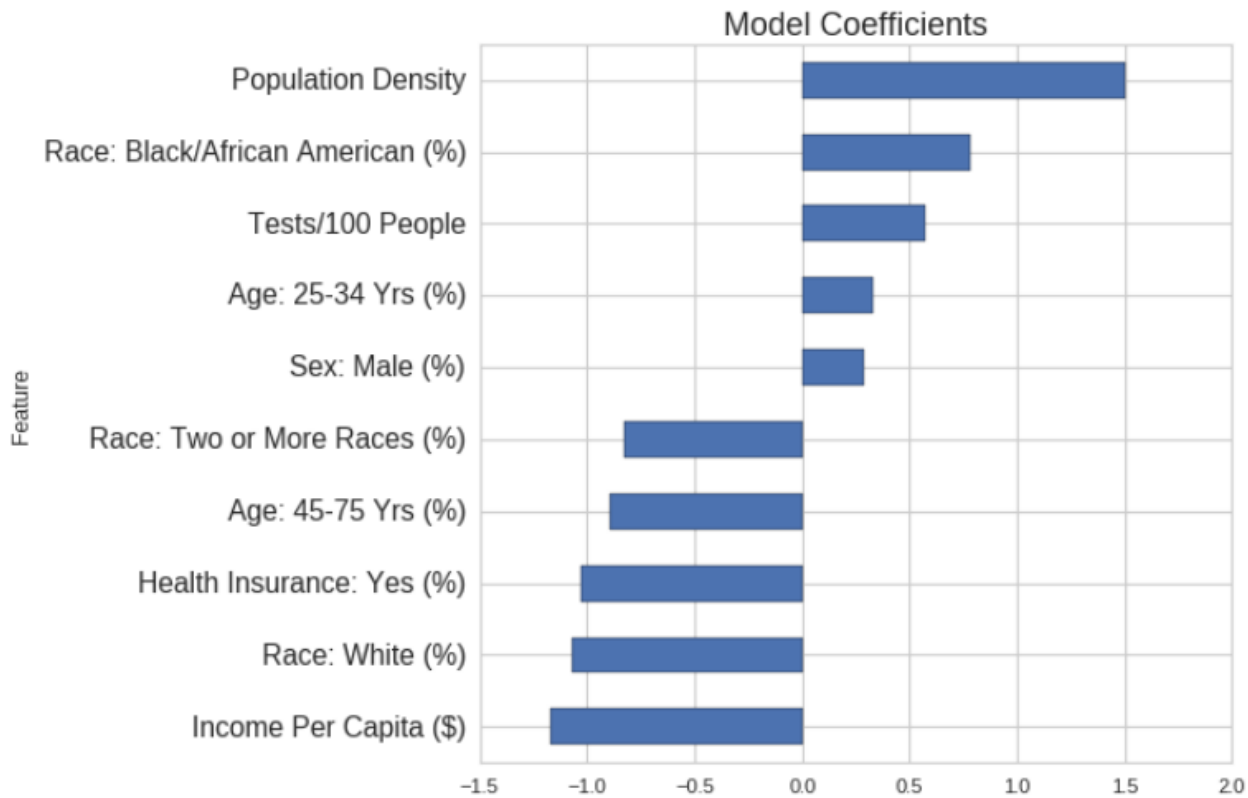
Low COVID Counties Likely to have Insurance Coverage, Lower Population Density



Modeling Successes and Challenges

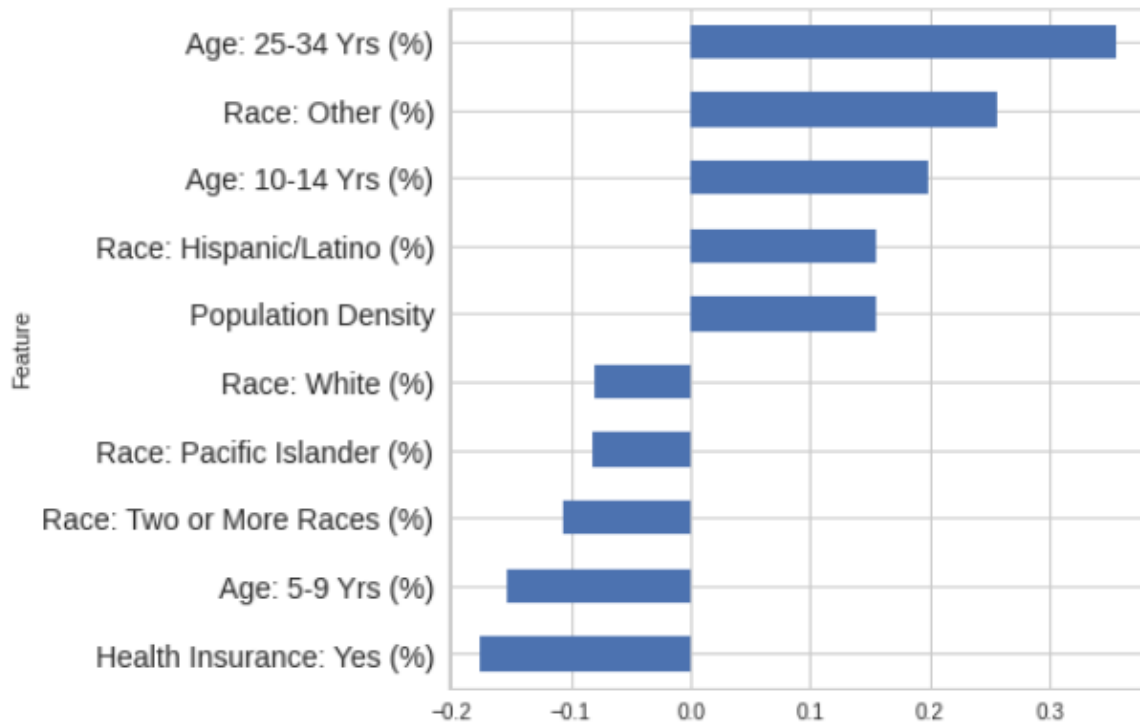
Region	Best Regression R2 Score	Best Classification Accuracy Score	Classification Baseline
All Five States	47%	63%	42%
California	75%	93%	66%
Florida	76%	71%	71%
Illinois	32%	73%	54%
New York	81%	94%	81%
Texas	49%	59%	40%

Population Density and Income Strongest Factors when Modeling All Five States

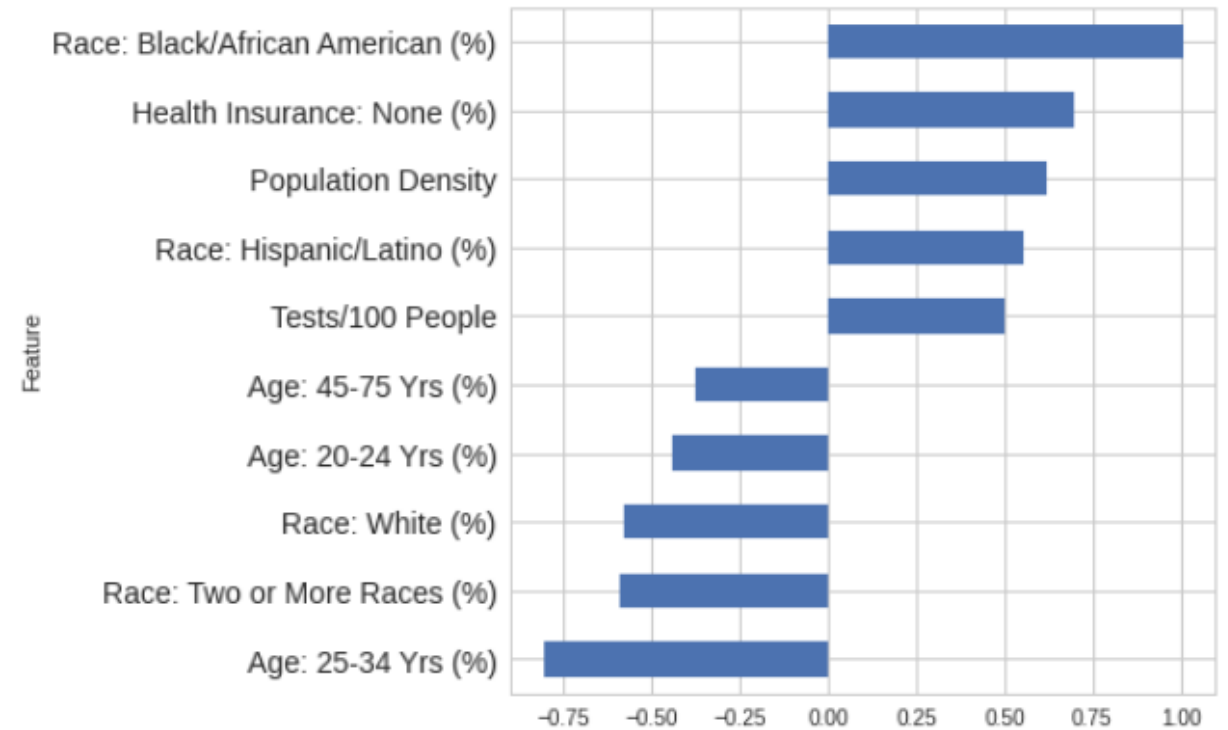


Predictors Varied in State-Level Models

Illinois Model Coefficients



New York Model Coefficients



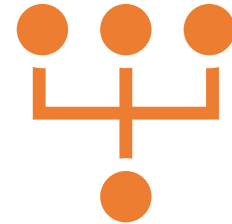
Conclusions and Key Challenges



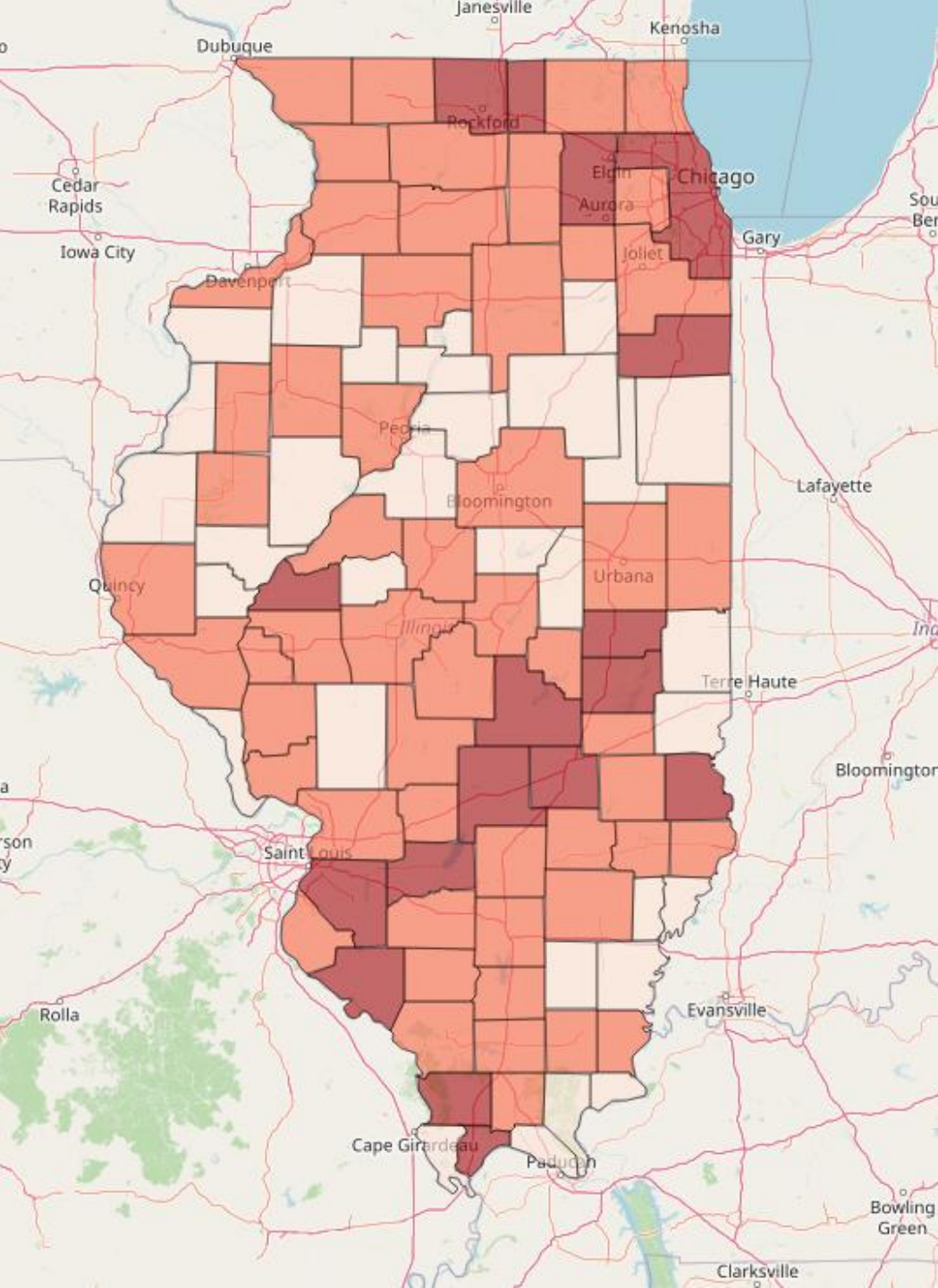
Ongoing event



Widely varying data



More features



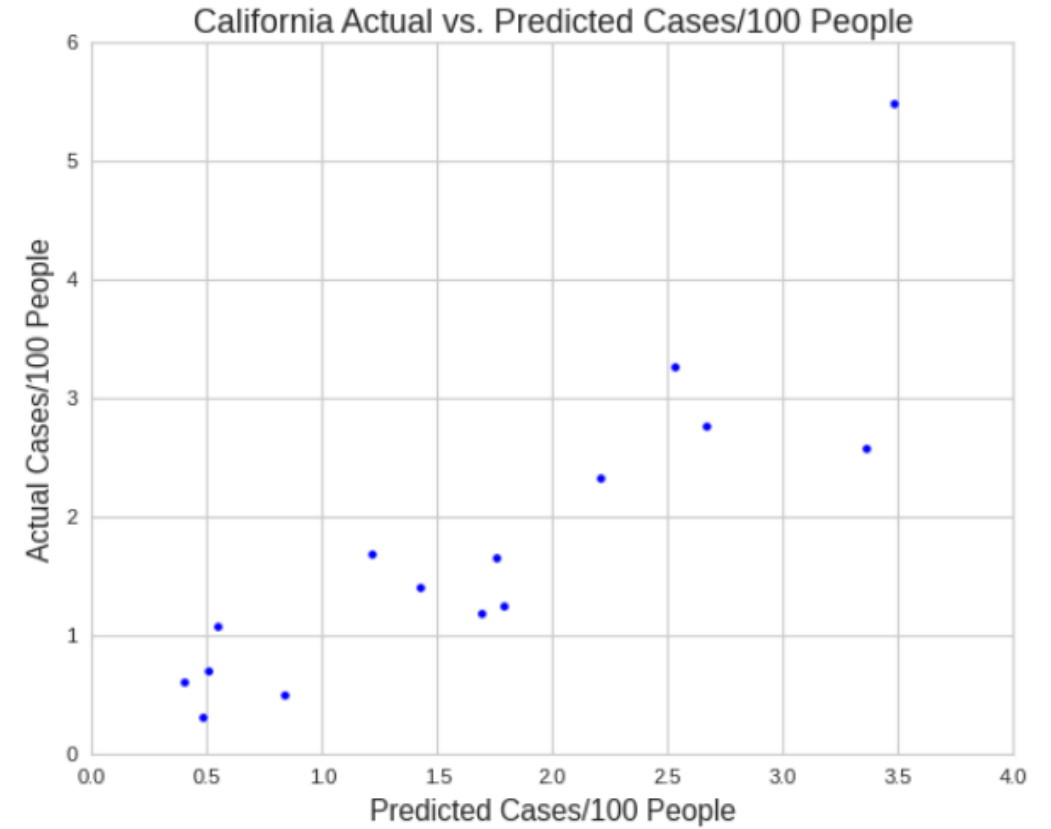
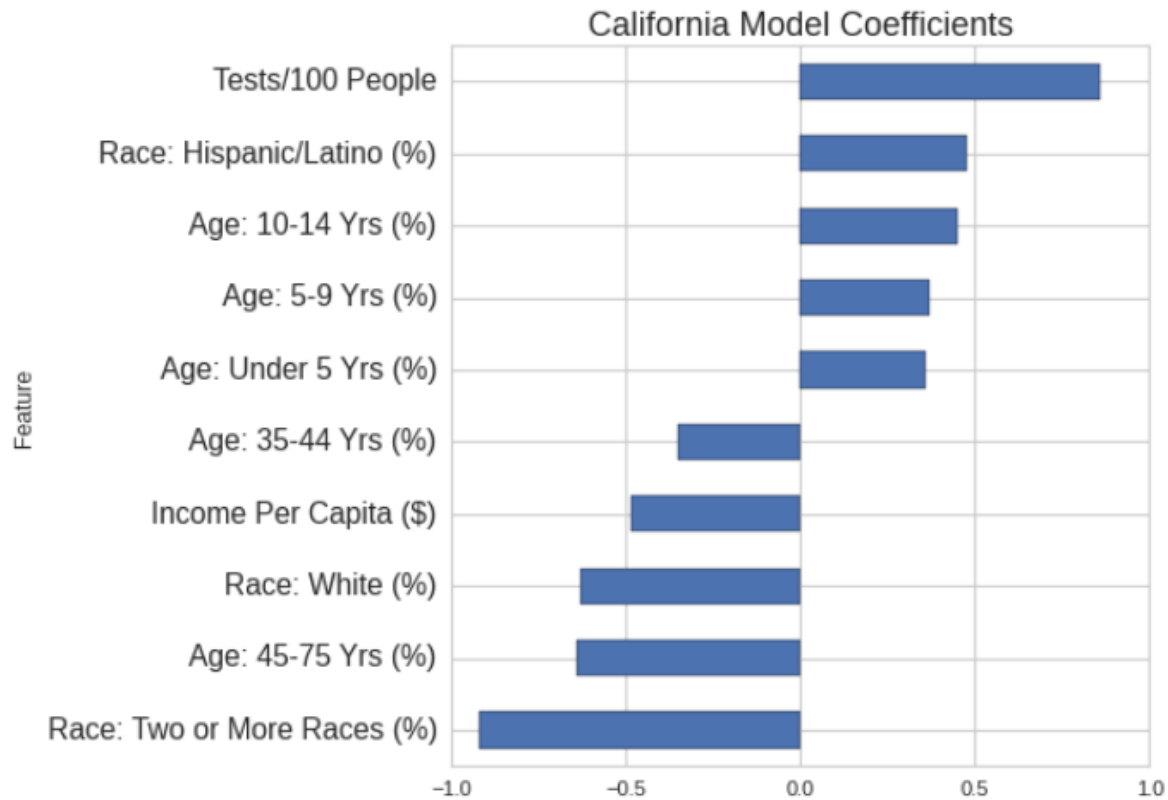
Demo: Interacting with Demographic Data and Classification Model

Thank you

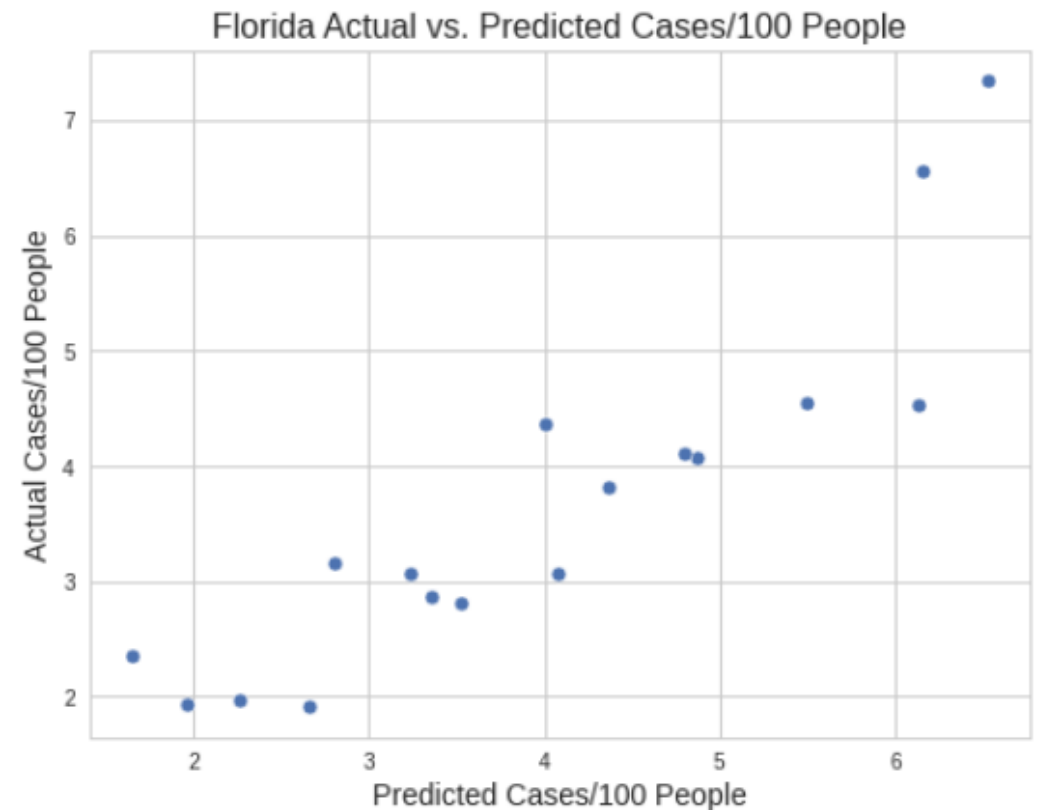
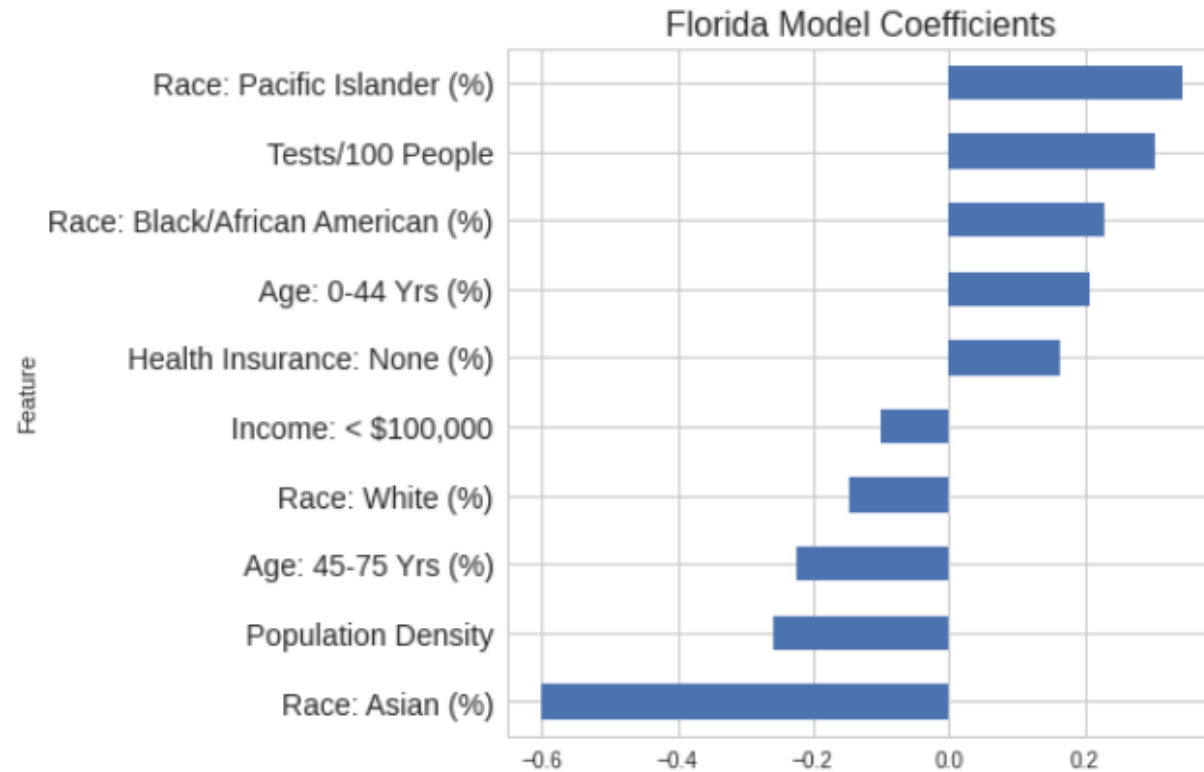


Appendix

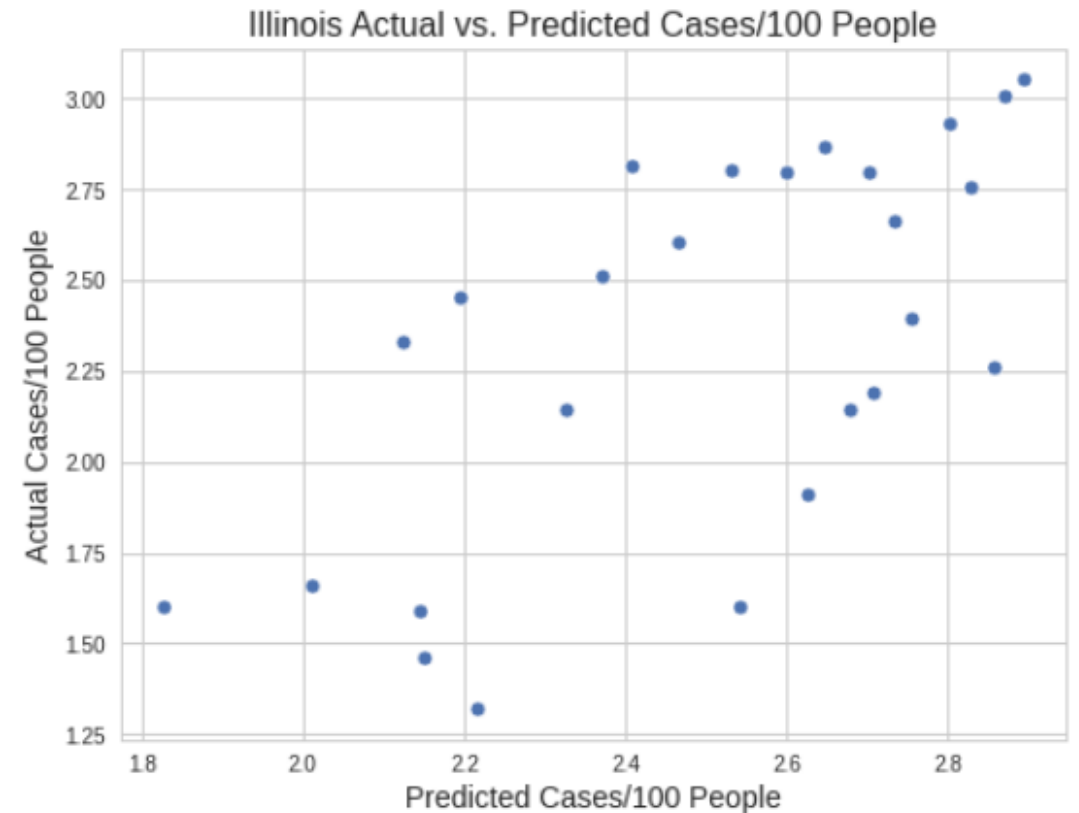
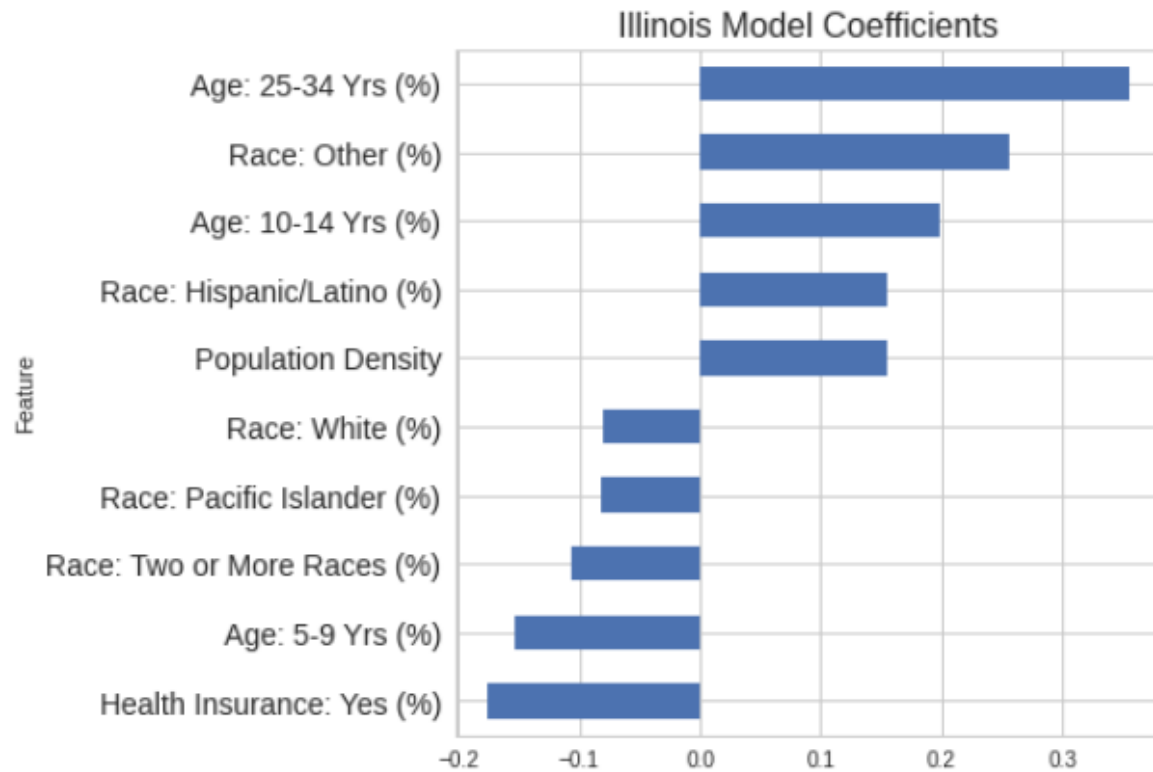
In California, Testing and Race Emerged as Strongest Predictors



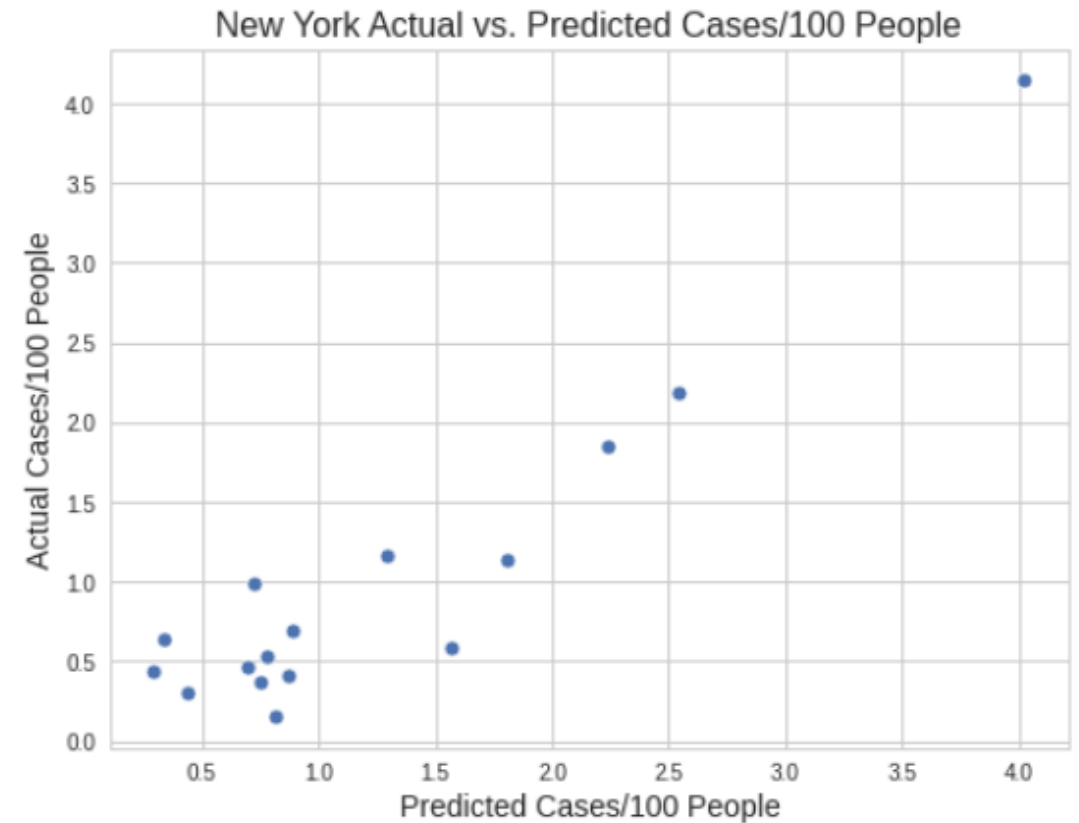
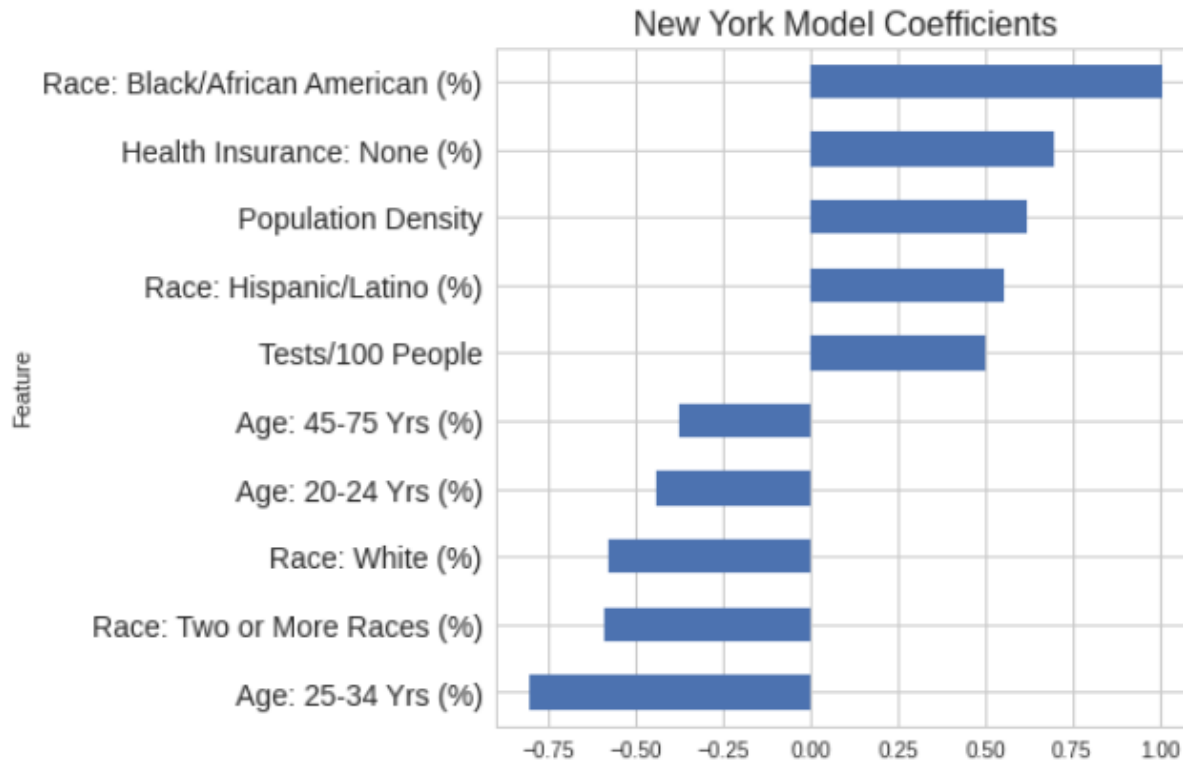
In Florida, Race Emerged as Strongest Predictors



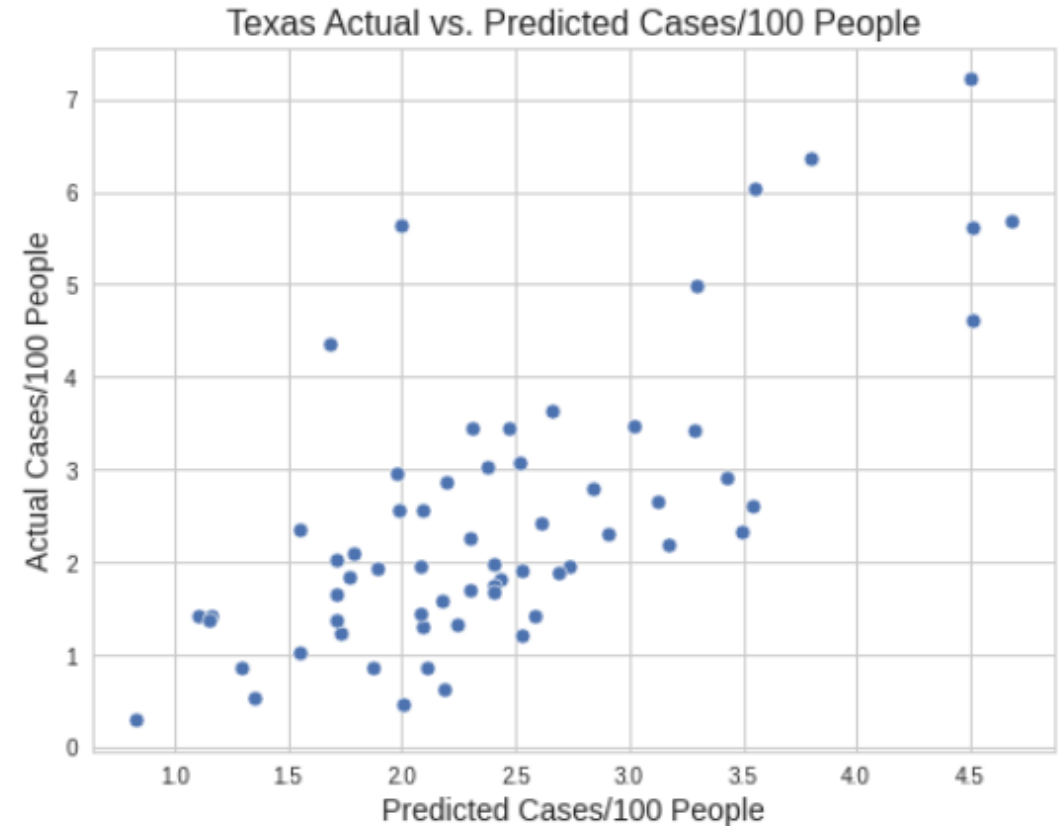
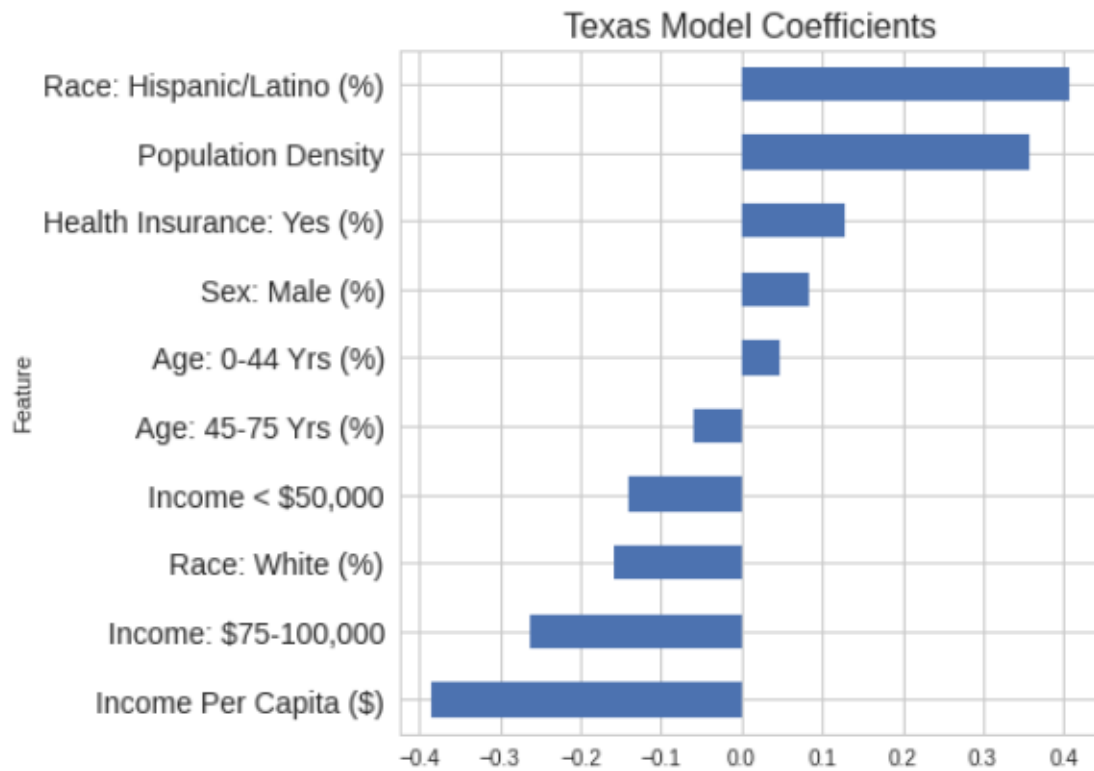
In Illinois, Age and Being Insured Emerged as Strongest Predictors



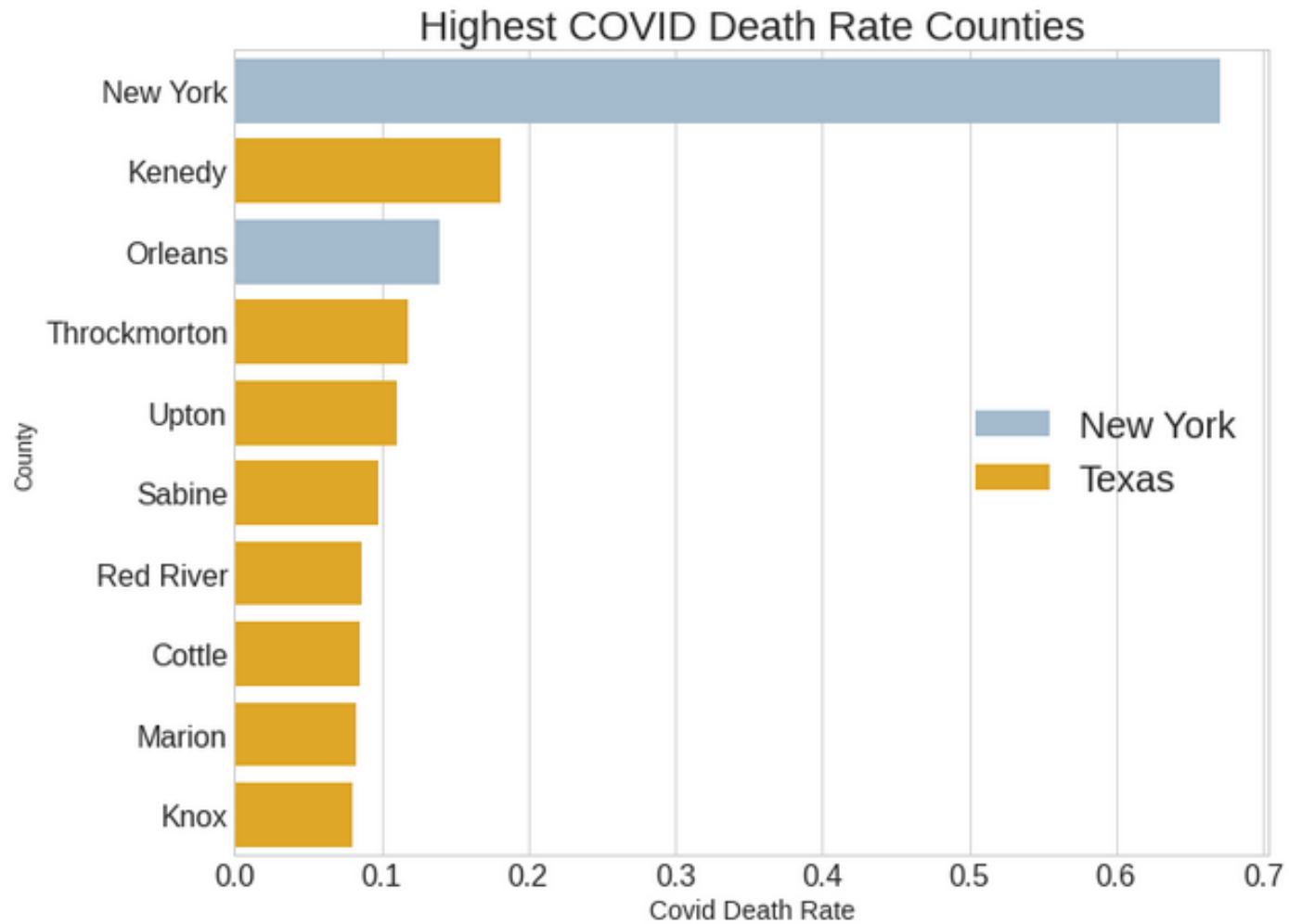
In New York, Race and Age Emerged as Strongest Predictors



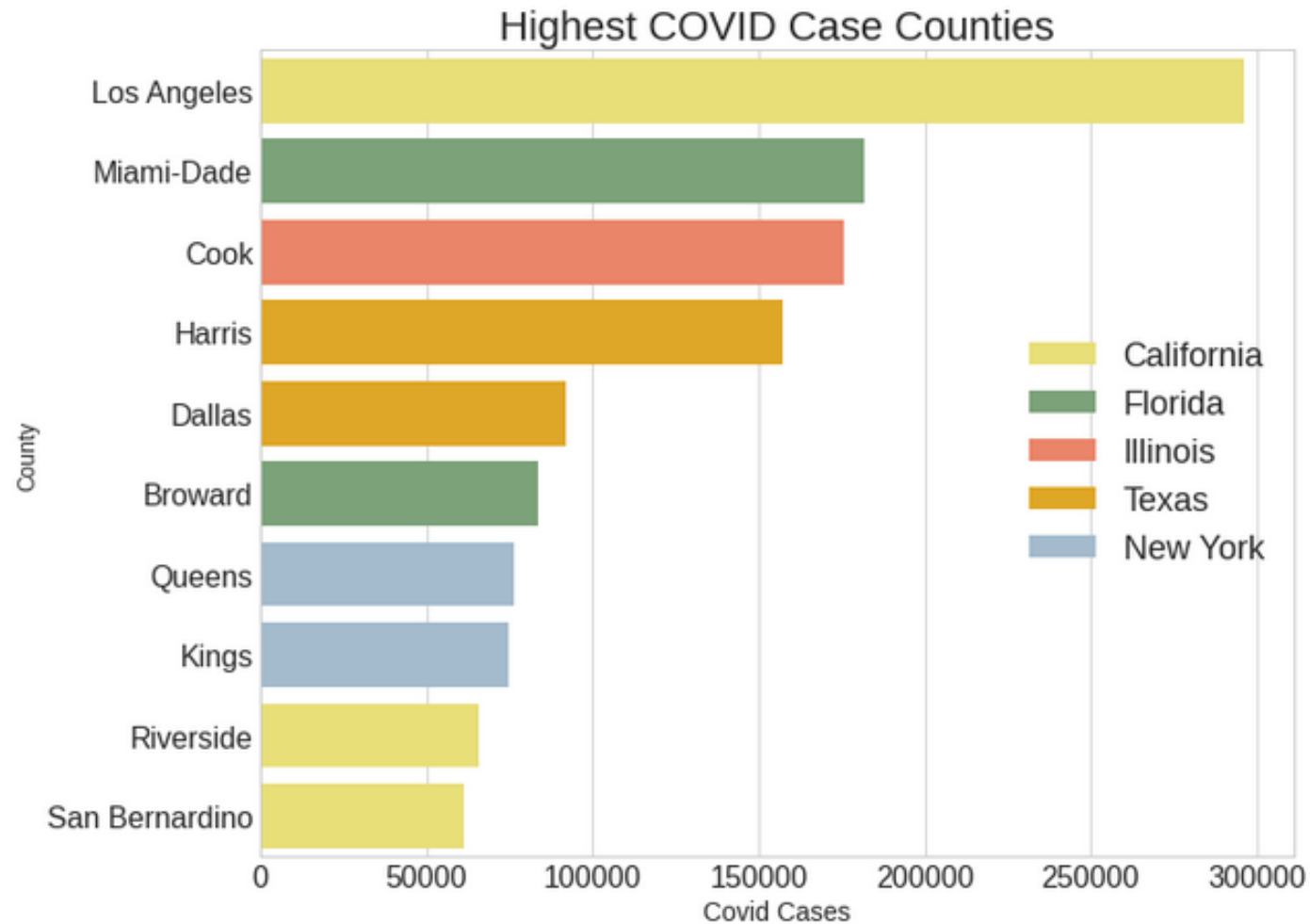
In Texas, Race and Income Emerged as Strongest Predictors



Highest Death Rate Counties



Highest COVID Case Counties Overall



Lowest COVID Case Counties Overall

