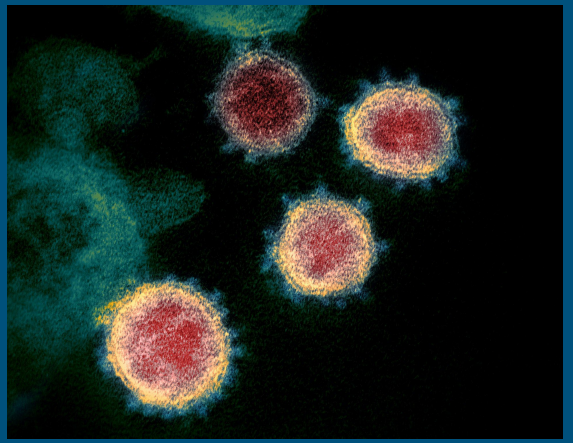
Analysis of COVID-19 and County Demographics

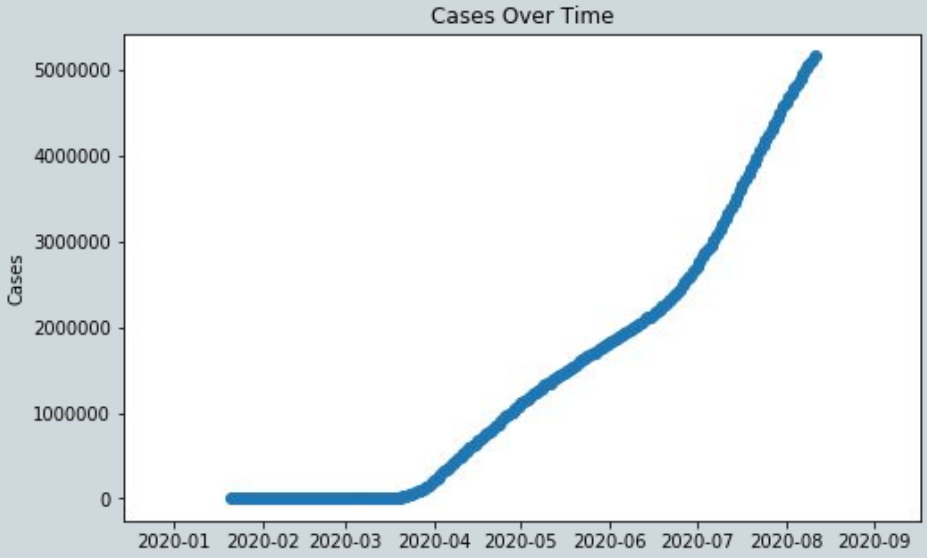
By Brendan Chan

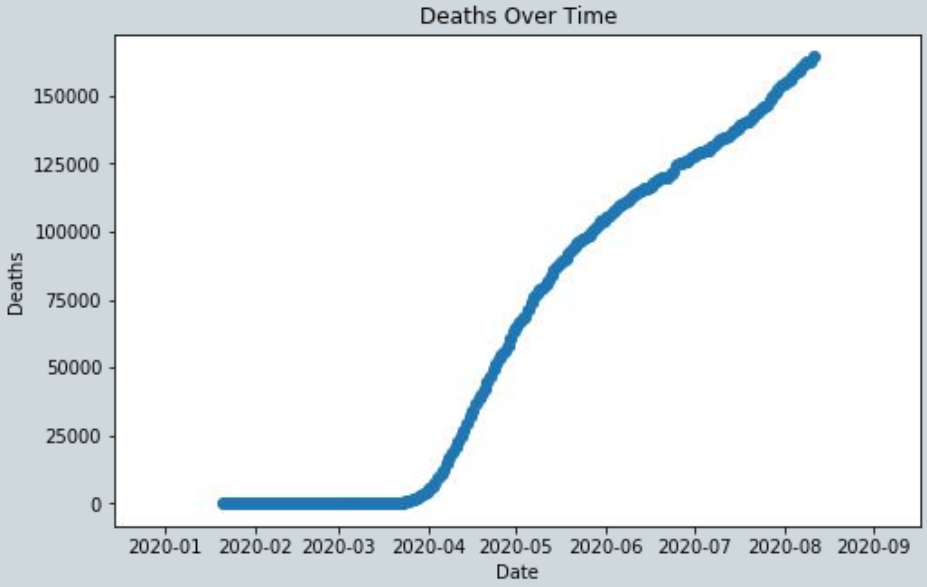
COVID -19

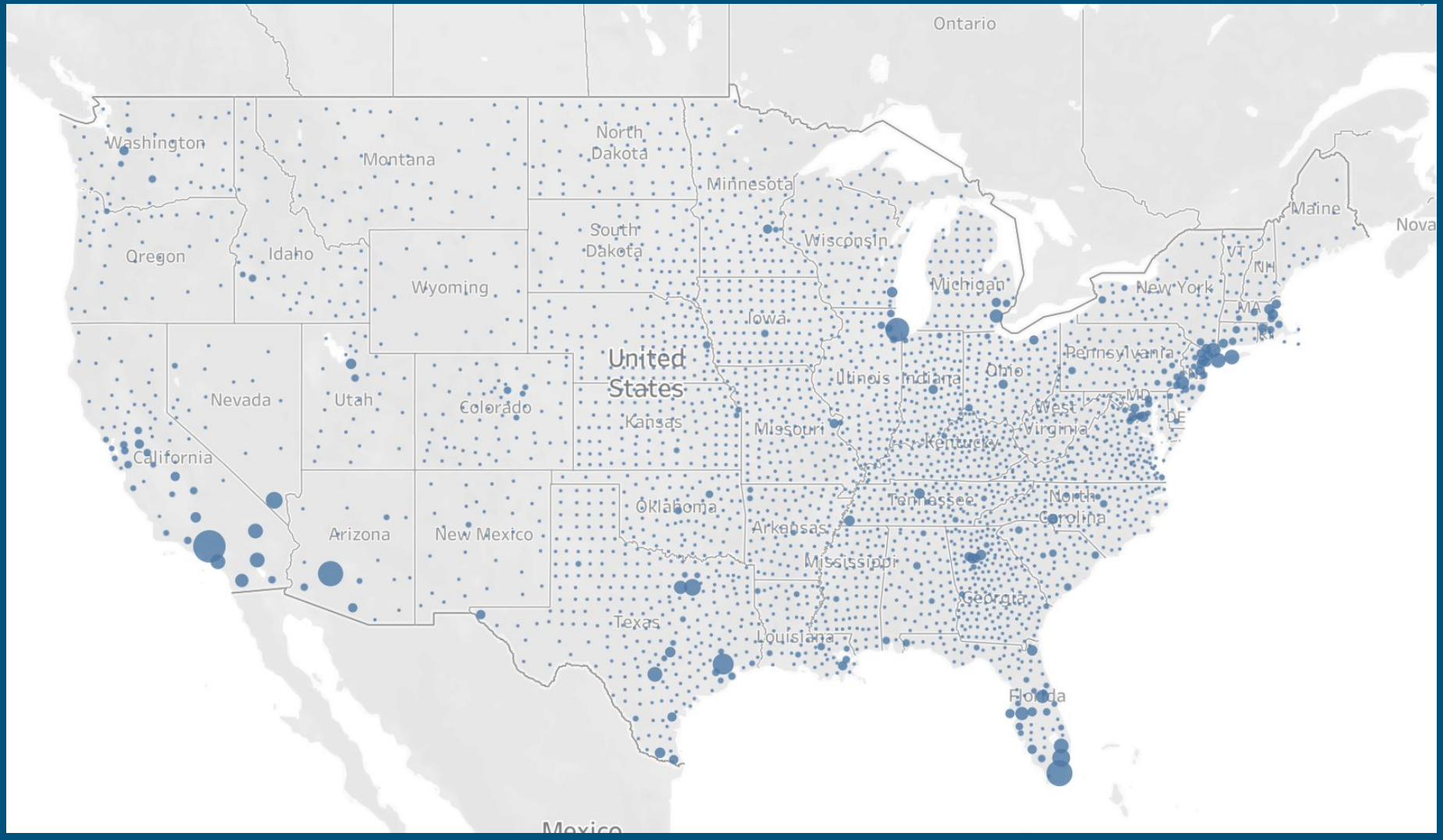
- Began in Wuhan, China
- Confirmed cases in the U.S. starting in late January
- Spread via droplets from talking, coughing, sneezing
- Prevention: social distancing, wearing masks
- Symptoms vary widely with many asymptomatic



With Respect to Time

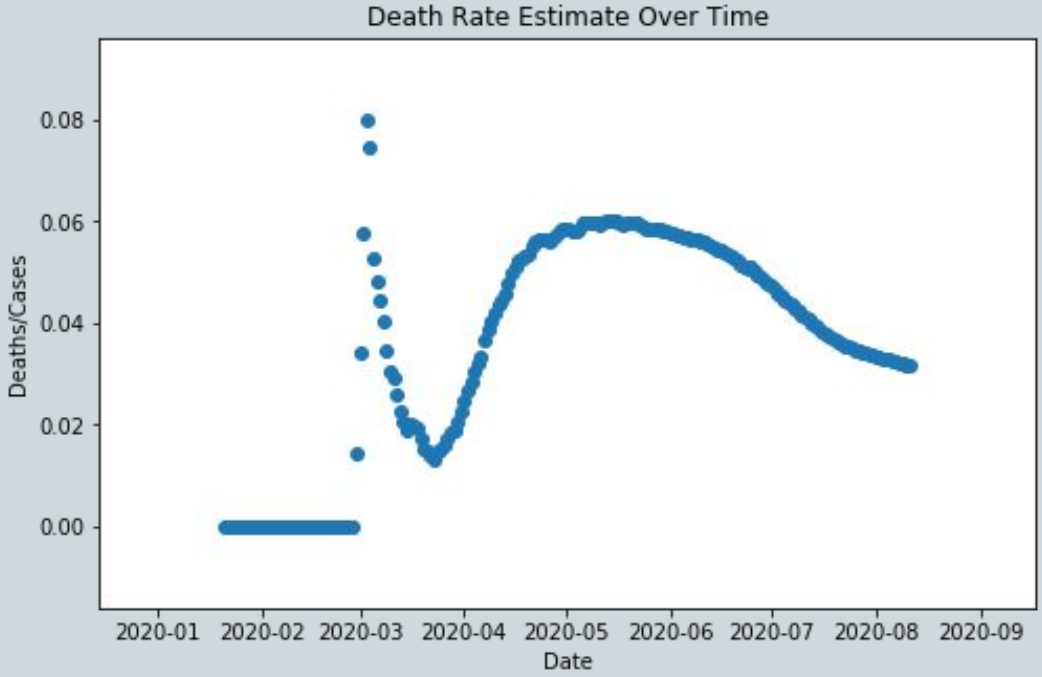




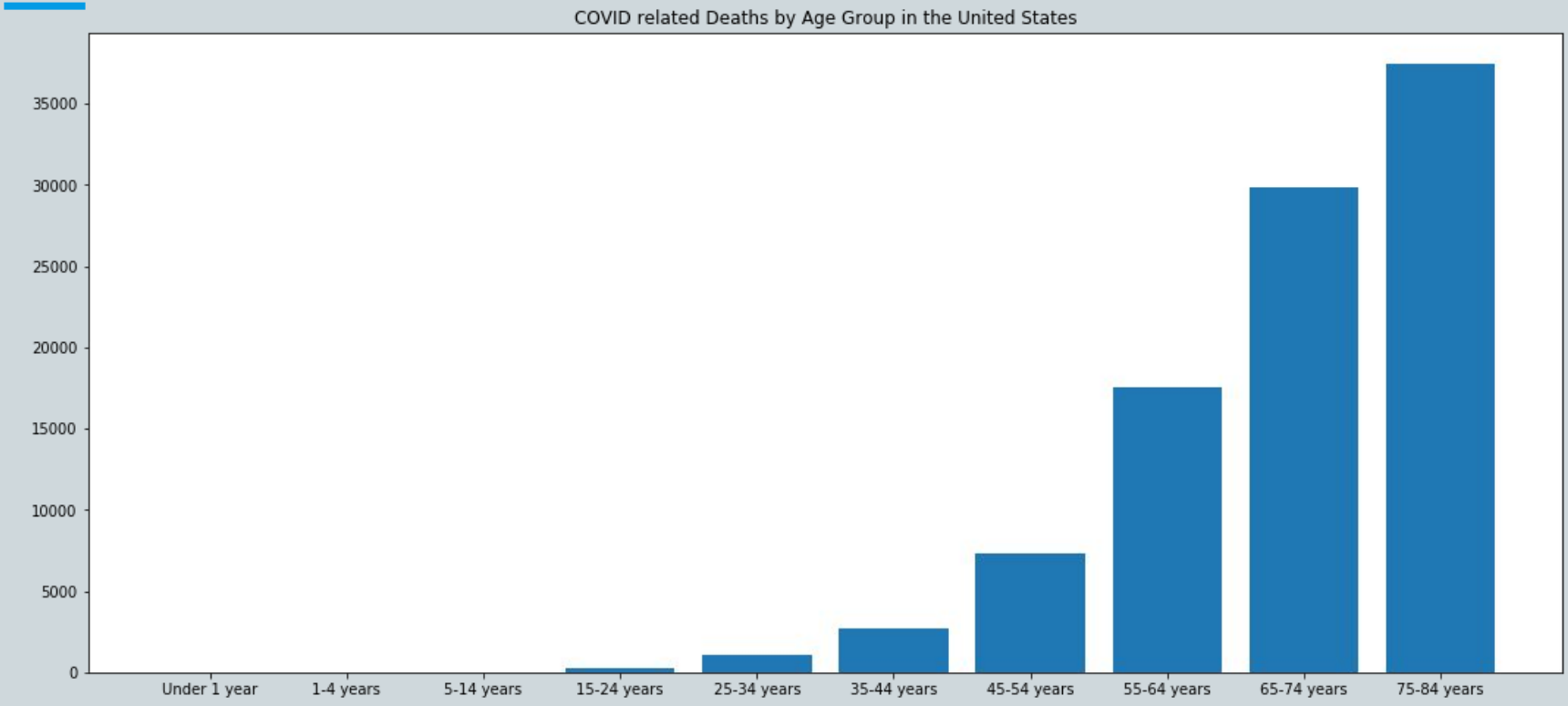

Ontario
Washington
Montana
North Dakota
Minnesota
Maine
Nova
Oregon
Idaho
South Dakota
Wisconsin
VT
NH
Wyoming
Michigan
New York
MA
Iowa
United States
Pennsylvania
Illinois
Indiana
Ohio
Nevada
Utah
Colorado
MD
DE
West Virginia
Kansas
Missouri
Kentucky
California
Tennessee
North Carolina
Oklahoma
Arkansas
Arizona
New Mexico
Mississippi
Georgia
Texas
Louisiana
Florida
Mexico

Mortality Rates

- Around 2-3% on average.
- Possibly < 2% from underreporting
- Demographic Differences?

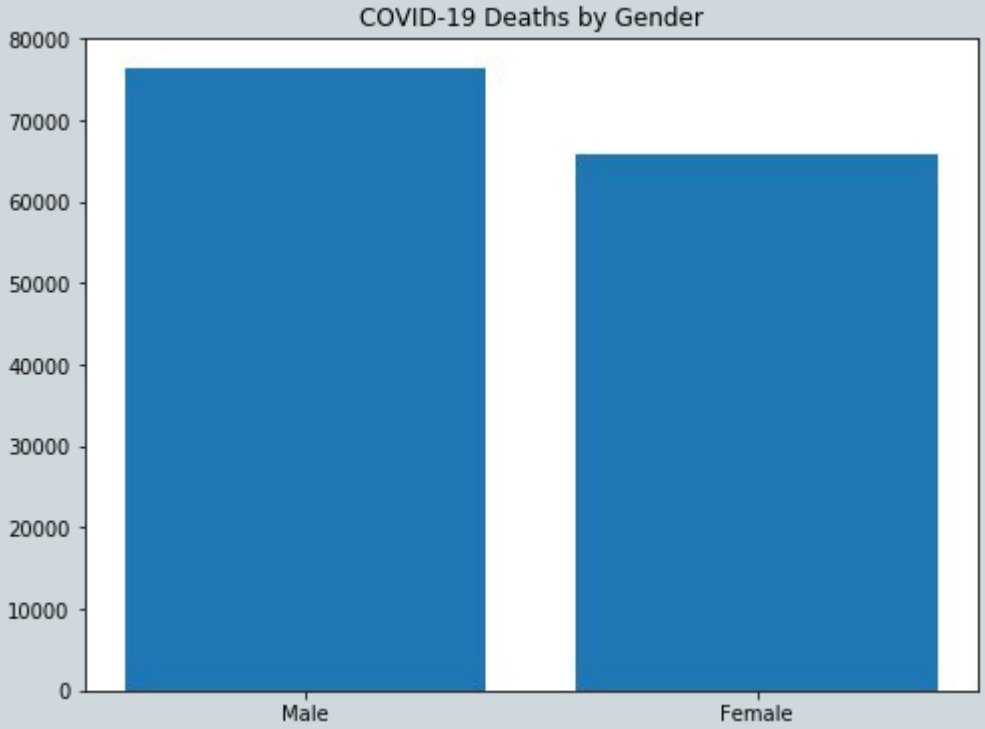


Older People are more vulnerable

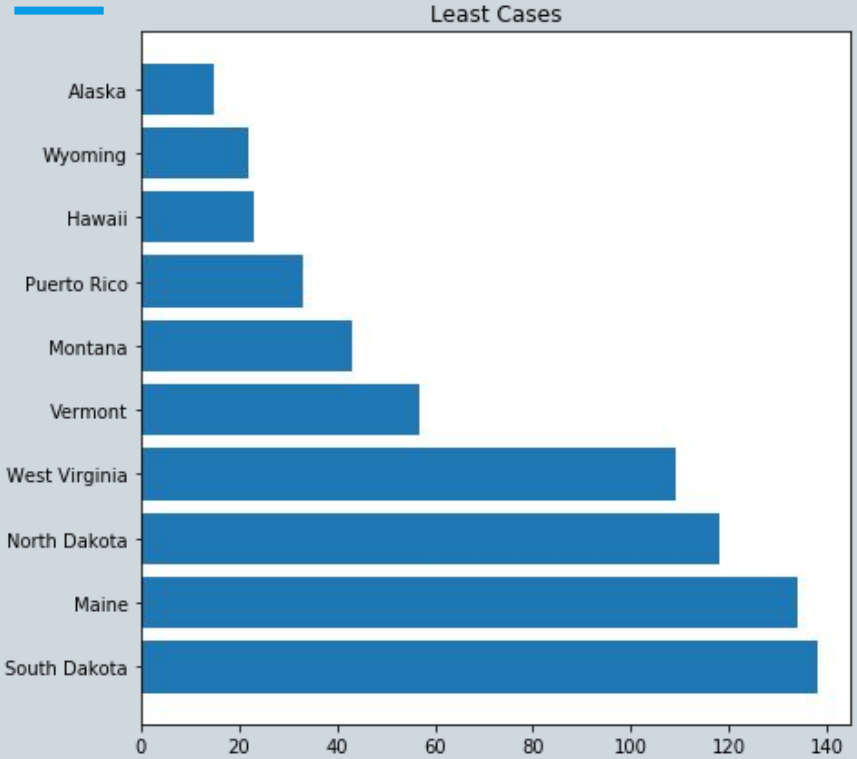


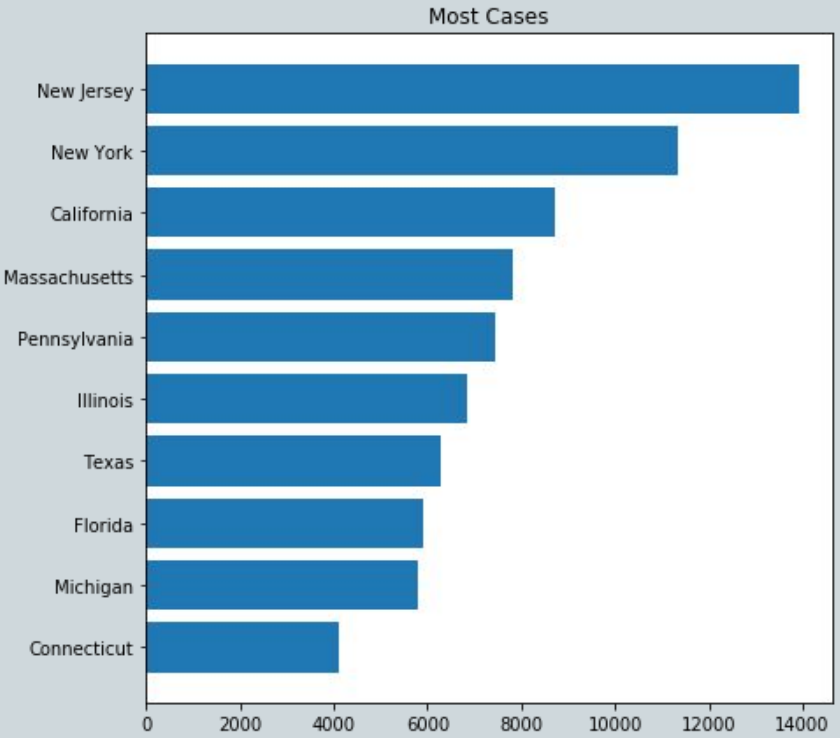
Men are at higher risk

- 54% of deaths are men
- Men are 16% more likely to die than women
- Women generally have a stronger immune system

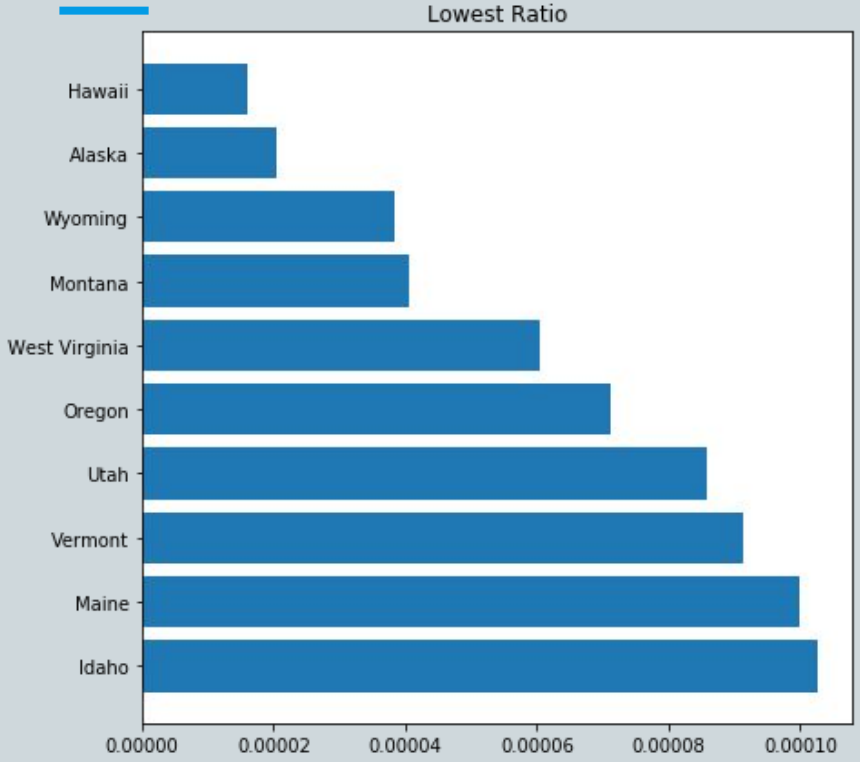


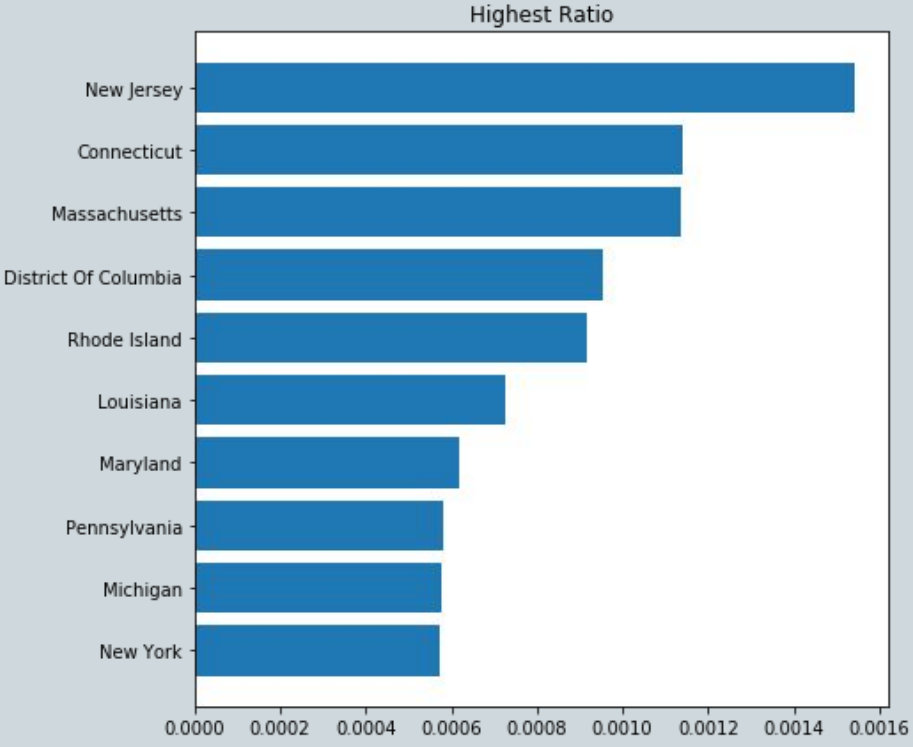
States


Least Cases
Alaska
Wyoming
Hawaii
Puerto Rico
Montana
Vermont
West Virginia
North Dakota
Maine
South Dakota
0
20
40
60
80
100
120
140


Most Cases
New Jersey
New York
California
Massachusetts
Pennsylvania
Illinois
Texas
Florida
Michigan
Connecticut
0
2000
4000
6000
8000
10000
12000
14000

Ratio of Deaths to State Population



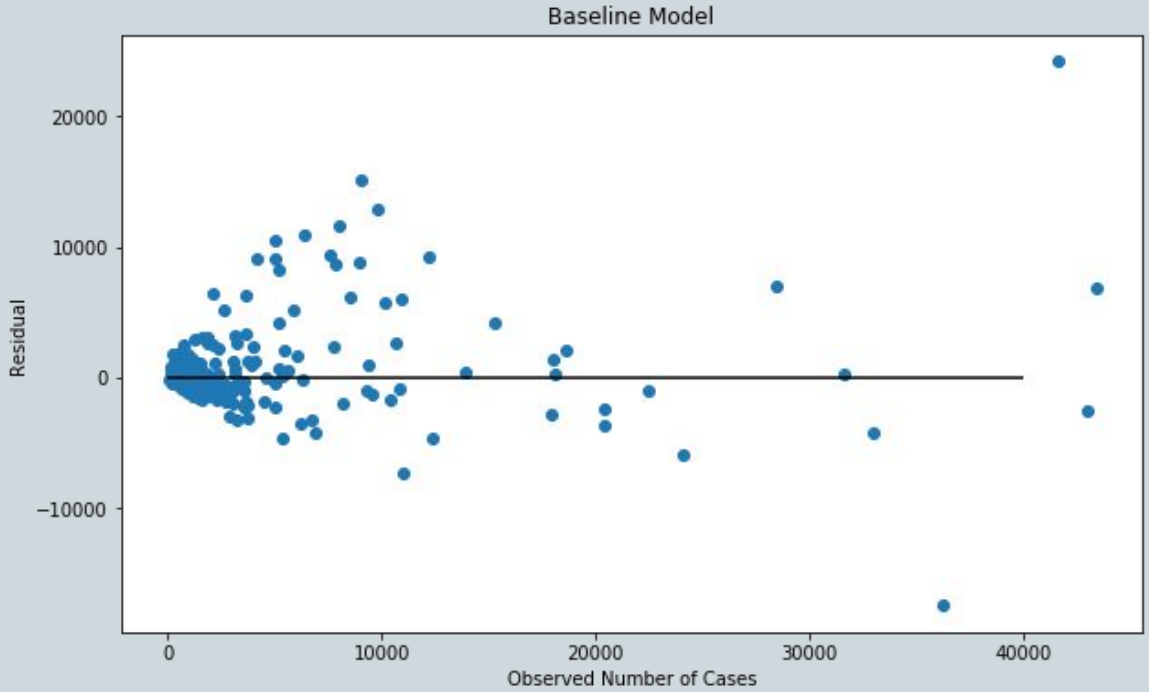


Census Data at the County Level

- 2018 county demographics data from government census database
- Demographic Groups: Age, sex, race
- Problem Statement: Build a model and make inferences on how covid affects different populations based on age, sex, race, and state they live in.

Baseline Model: Predicting Cases

- OLS model
- Use only population
- 3:1 split
- Train R2: 0.87
- Test R2: 0.73
- RMSE: 3979



Throw Everything at the Wall Model

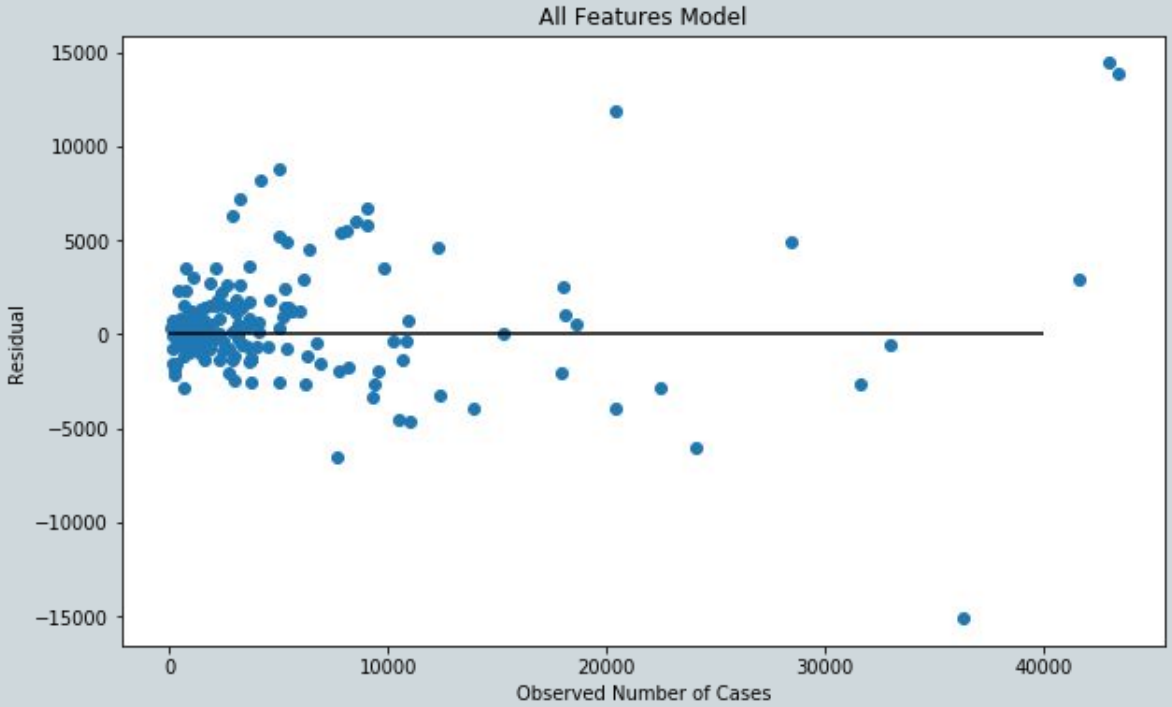
- Train R2: 0.97
- Test R2: 0.85
- RMSE: 2940
- Feature correlations
- Inferences:
  - New Mexico, lower cases
  - Louisiana, higher cases

Deep Dive: New Mexico

- New Mexico shut down its economy sooner than most states
- Slower reopening of economy in June
- Contact tracing efforts
- Significantly Ramped up testing
- Close to testing facilities, reducing turnaround time
- Republicans critical, irreparable damage to economy

LASSO Model + OLS Model

- Features chosen using LASSO
- Train R2: 0.94
- Test R2: 0.85
- RMSE: 3020

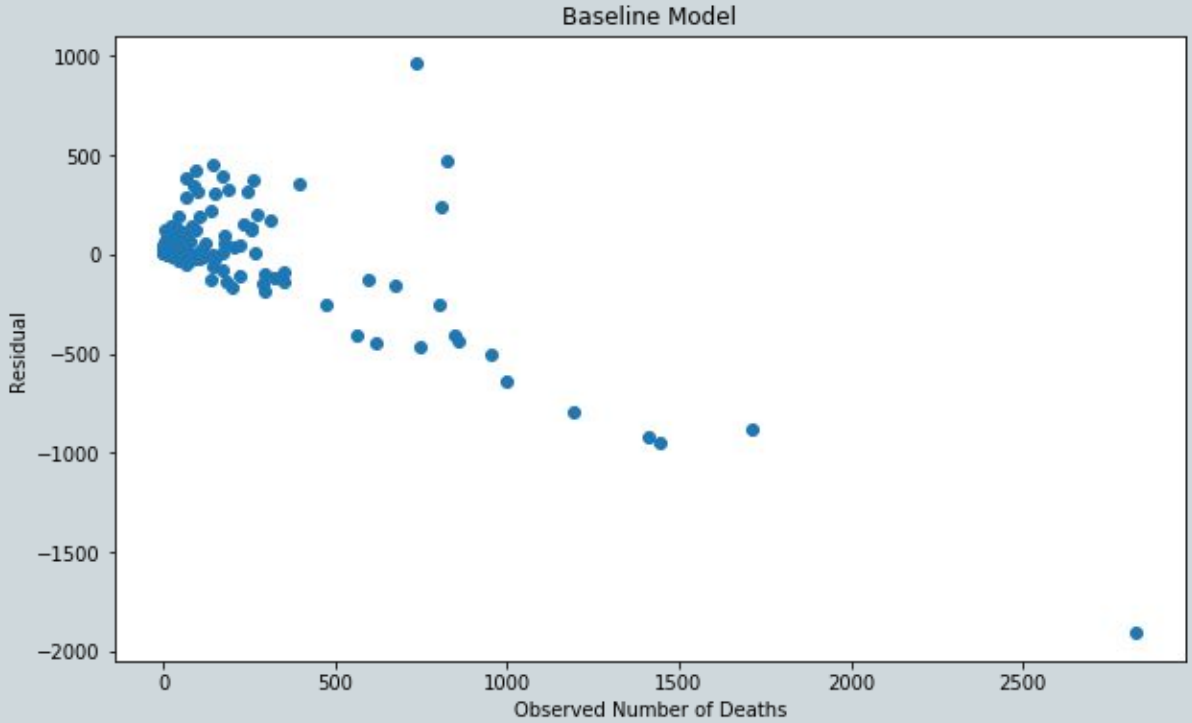


Inferences

- Statistically Significant Coefficients:
  - White Americans: 0.0138
  - African Americans: 0.0229
  - Hispanic or Latino American: 0.0284
  - American Indians & Alaska Natives: 0.1574
  - Asian Americans: -0.0226
- Suggests that hispanic, black, and American indian populations have been hit hardest
- Overrepresentation in essential workforce, lower income communities

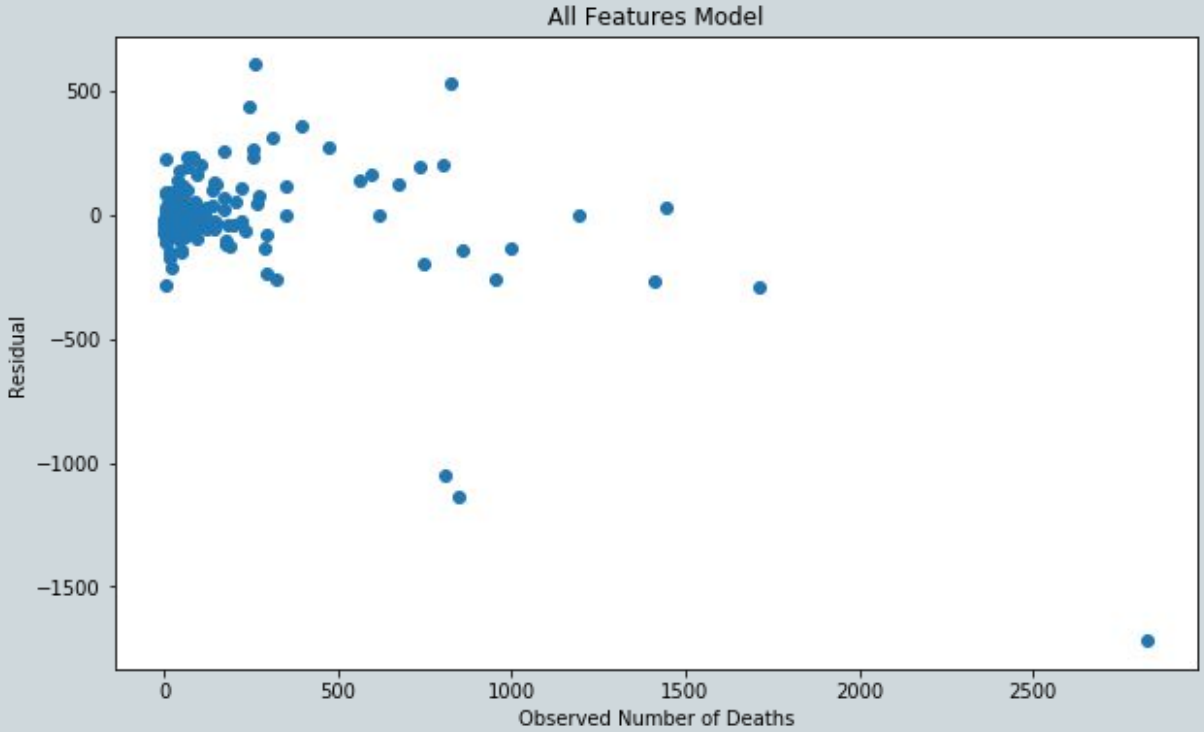
Baseline Model: Predicting Deaths

- Train R2: 0.67
- Test R2: 0.44
- RMSE: 249
- Not very good!



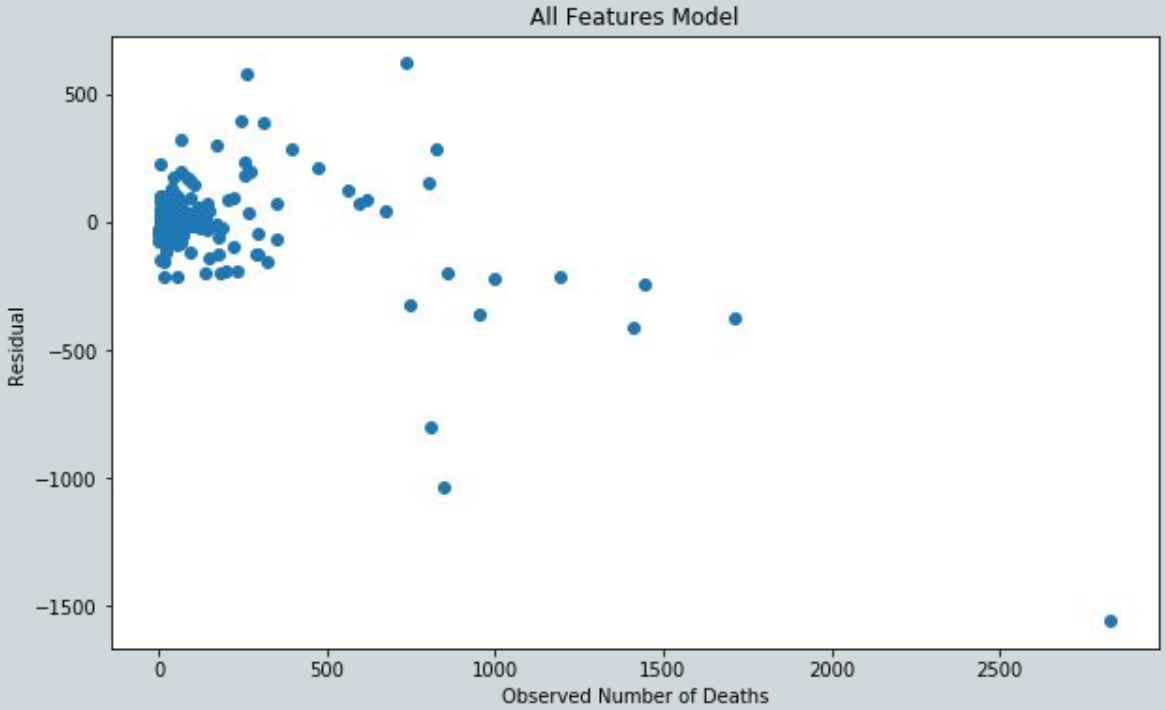
Throw Everything at the Wall Model

- Train R2: 0.93
- Test R2: 0.62
- RMSE: 206



LASSO + OLS Model

- Train R2: 0.90
- Test R2: 0.66
- RMSE: 196
- High Number of features
- Difficult to make inference



Recommendations

- Be cautious about reopening schools. Kids could be spreading COVID.
- Allow for flexibility: households with both children and elderly should consider homeschool or online school.
- Frequent testing and faster results
- Contact Tracing
- Hazard pay for essential workers
- Balance between economy and public health. This is a marathon.

Thanks For Listening!