Social Factors and COVID-19 Outcomes: A State-Level Inquiry

Rahul, Lisa

Introduction

Background:

- The COVID-19 response has differed greatly between states
- Large datasets offer a good exercise for BigQuery and Google Cloud Console

Research Questions:

- Are certain states at higher risk than others for adverse COVID-19 outcomes?
- Does the outcomes data suggest that these risks are playing a significant role?

Datasets

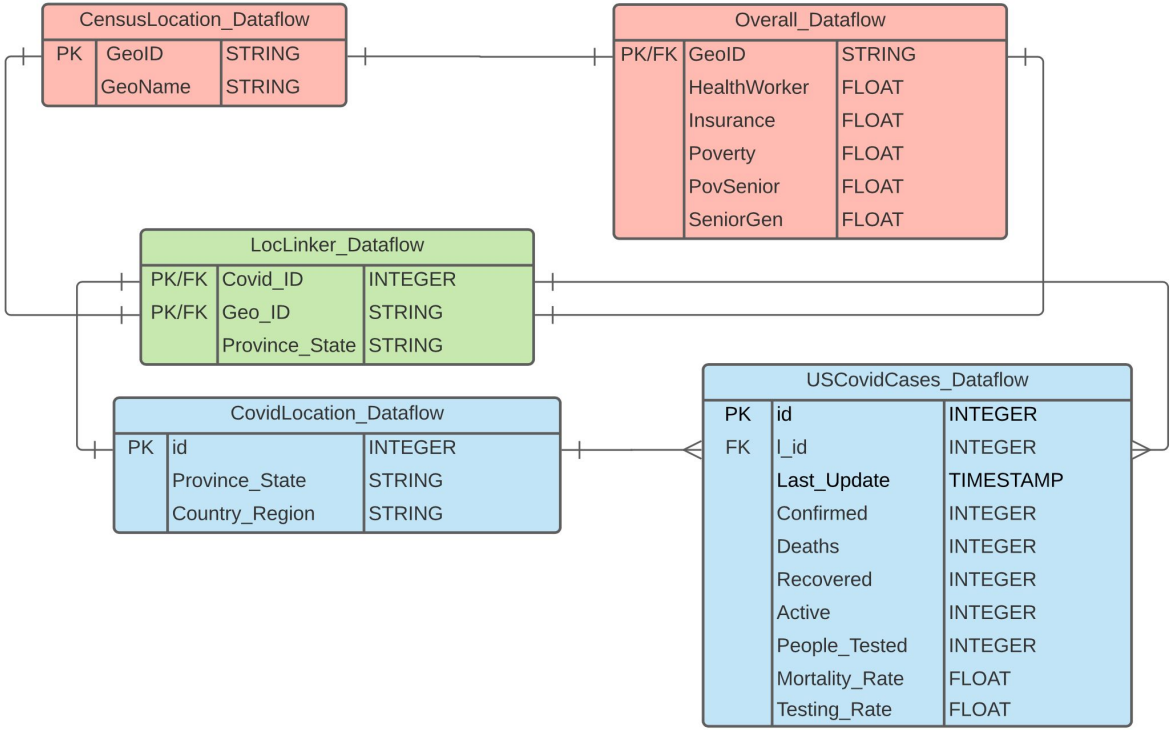
Census (Census): https://www.census.gov/acs/www/data/data-tables-and-tools/geographic-comparison-tables/

- US_65andOver.csv, US_HealthInsurance.csv, US_HealthcareForce.cvs, US_Poverty.csv, US_Poverty65andOver.csv

COVID (JHU): https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports_us

- Csse_covid_19_daily_reports_us.csv (by dates)

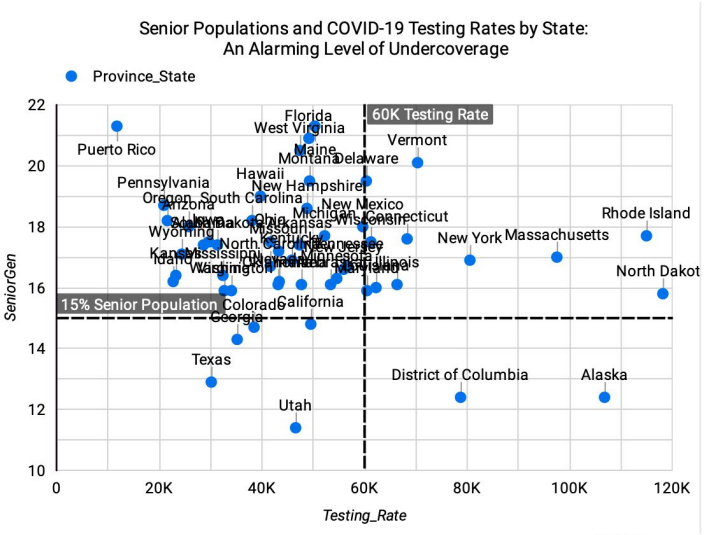
Modeling

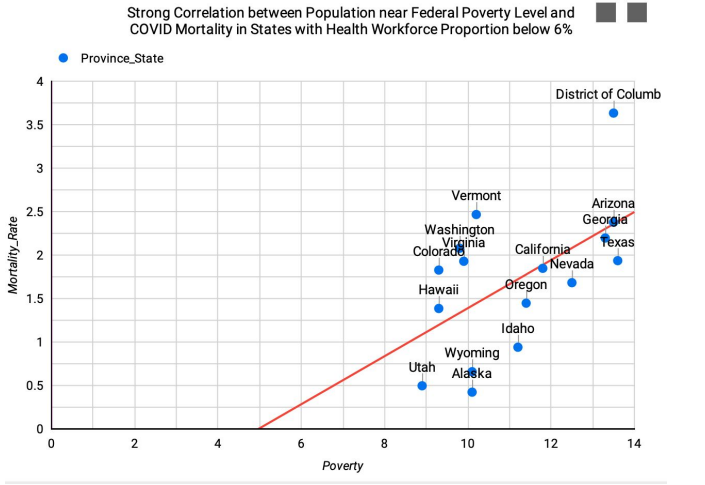

CensusLocation_Dataflow
PK GeoID STRING
GeoName STRING
Overall_Dataflow
PK/FK GeoID STRING
HealthWorker FLOAT
Insurance FLOAT
Poverty FLOAT
PovSenior FLOAT
SeniorGen FLOAT
LocLinker_Dataflow
PK/FK Covid_ID INTEGER
PK/FK Geo_ID STRING
Province_State STRING
CovidLocation_Dataflow
PK id INTEGER
Province_State STRING
Country_Region STRING
USCovidCases_Dataflow
PK id INTEGER
FK l_id INTEGER
Last_Update TIMESTAMP
Confirmed INTEGER
Deaths INTEGER
Recovered INTEGER
Active INTEGER
People_Tested INTEGER
Mortality_Rate FLOAT
Testing_Rate FLOAT

Outcomes

COVID Strategy is out of sync with state demographics

- States with high senior populations are running fewer tests, which increases undetected outbreak risk.
- Mortality rate is strongly correlated with poverty in states with smaller health workforces

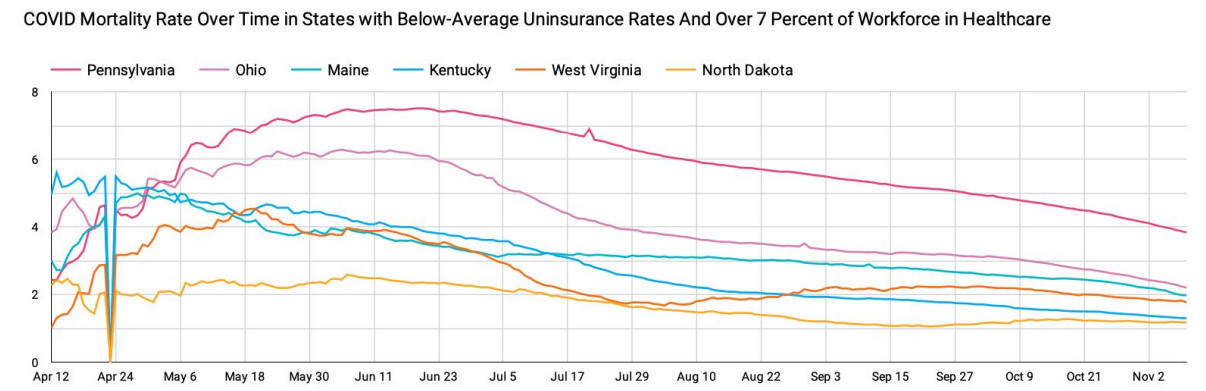


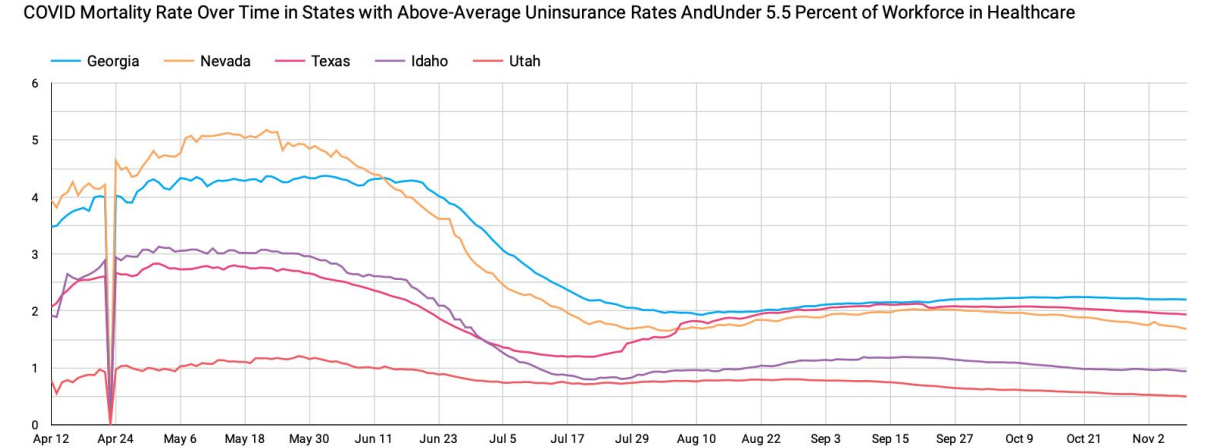


Outcomes

Early waves still present a key data confound

- Well-protected states are still dealing with mortality from early pandemic
- Undercovered states feature lower but stabilized mortality





Conclusions

Results paint a complicated picture

- Certain effects only appear below key thresholds
- Growth in knowledge between outbreaks obscures mortality rate

Key Accomplishments:

- Found a potential critical threshold of healthcare worker proportion and identified states that may need more workers
- Identified states to track for the rest of the pandemic period to understand value of health investment

Challenges and Future Directions

- Census dataset started off with unformatted locations
- Large COVID dataset

In the future, we would like to explore more about the correlation between the percent of poverty and senior citizens to COVID rates in a smaller scale. We would also like to see if other factors affect like population concentration and number of nearby hospitals affect COVID statistics.

Takeaways

- How to use Big Query, Google Cloud Console, Beam Pipeline, Dataflow
- Got familiar with sql : select, join, group by, having, order by
- Sub-queries: aggregate and non-aggregate
- Created views from Big Query
- Data visualization: Certain plots told stories more effectively than others
- Reassessing assumptions: A lot of the connections didn’t pan out