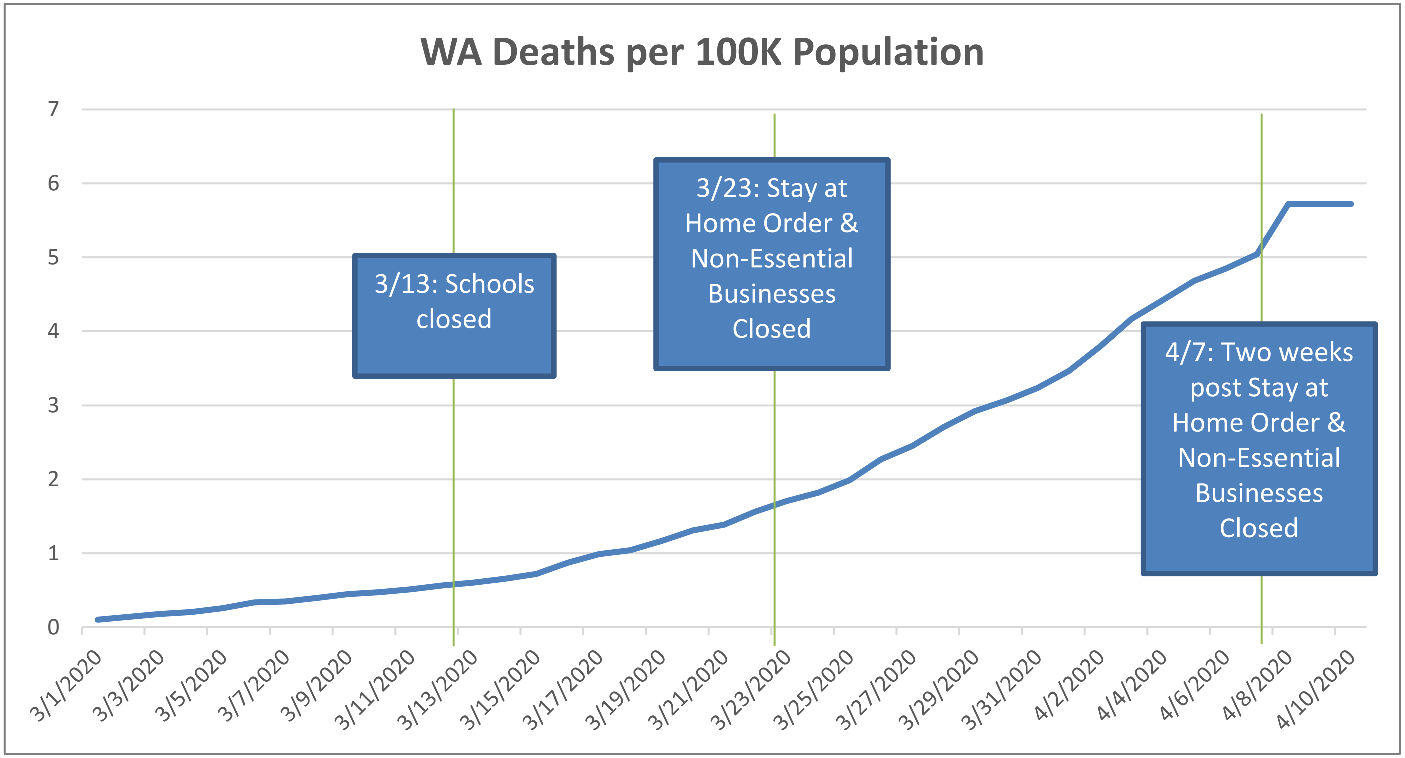
Applying Bayesian Analysis to COVID-19:
What is helping to slow death rates?

ST 540 Final Project

Taylor Krebsbach, Jonathan McMahon, and Shantel Ward

Motivation

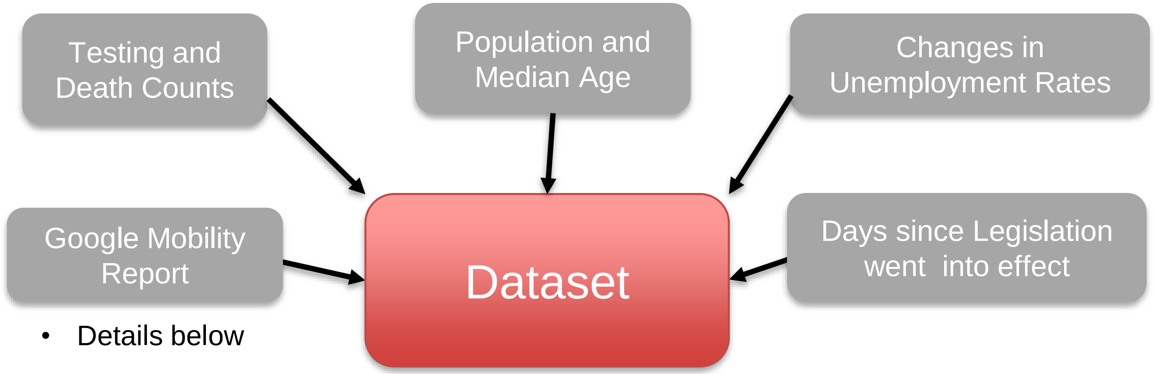
- The COVID-19 pandemic began in Wuhan, China in December 2019. The virus has traveled the globe and drastically changed our lives. It is no surprise for us to wonder if staying at home is slowing the spread. We seek to understand if the actions taken by the U.S. states has significantly impacted the number of deaths caused by COVID-19.
- For example, in Washington state, we see below that the number of deaths per 100,000 population appears to taper off around two weeks after the stay at home order and non-essential business closures.



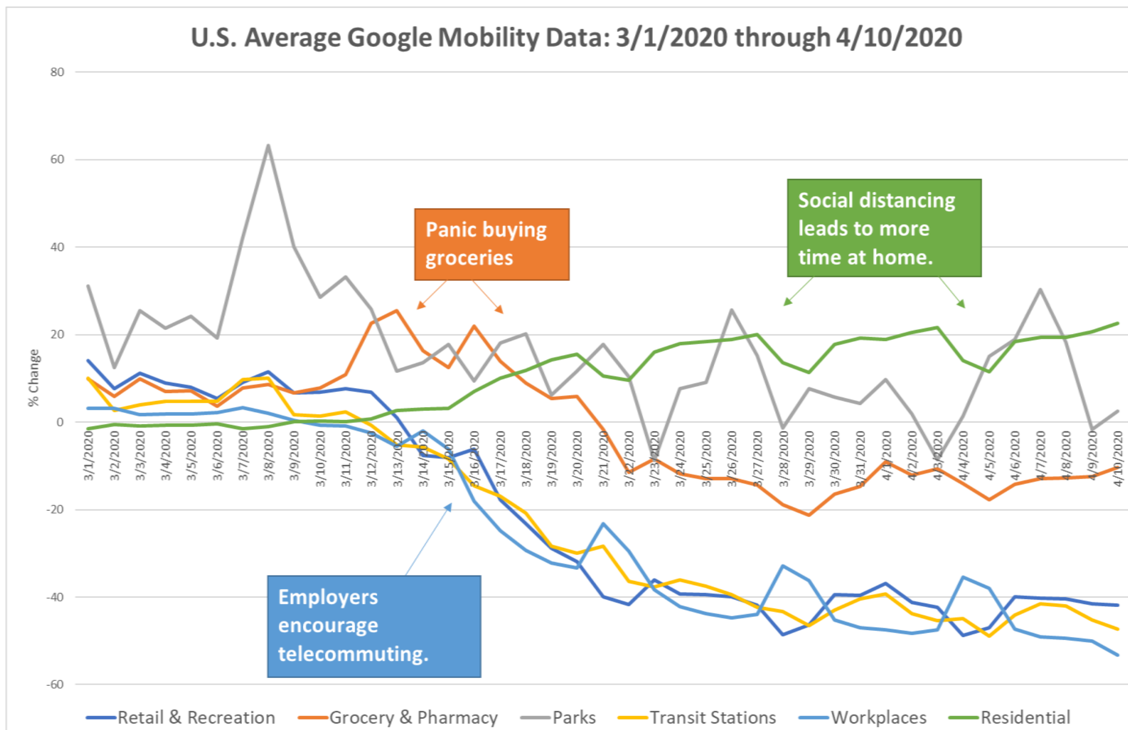
Data Analysis

- We compiled data from multiple sources & aggregated them by state from 3/1-4/10/20. Our model inputs consist of daily snapshots by state of these factors combined with state-specific information (e.g. median age).
  - The dates for state actions e.g. shoool closures) were transformed into "days since effective date of action"
  - Deaths per 100K were normalized by the population of each state
- Google collects mobility data to provide insights into how activity patterns have changed in response to policies aimed at combatting COVID-19.

Data Sources:
covidtracking.com
google.com/covid19/mobility
worldpopulationreview.com
bls.gov
kff.org

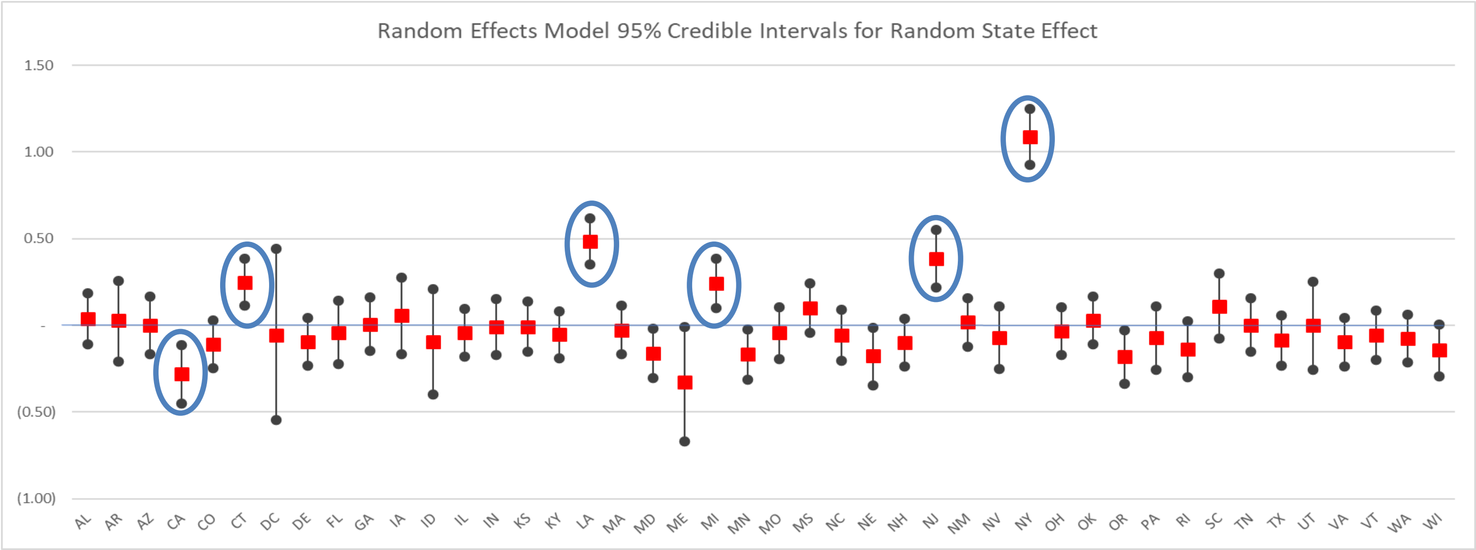


- Details below



Variable Selection & Testing

- Two weeks is the commonly-cited incubation period during which COVID-19 symptoms will manifest. Based on this fact, in order to explore the death rates two weeks into the future, we select variables using data from two weeks prior as our covariates.
- We also removed states that had less than 10 deaths. We considered that these states are “lagging behind” the other states.
- Throughout our modeling selection, we consider the fact that some of the variables in our model may mask the signal of the others.
  - *StayAtHome* variable has statistical significance in the absence of the Mobility data.
- In one model we reviewed the random state effects. As seen below, the states with random effects significantly different from zero are those that were hit the hardest with COVID-19 earlier on: NY, NJ, CT, LA, CA, and MI.



Models & Results

Y_i = new COVID-19 deaths in state two weeks later (per 100k population)

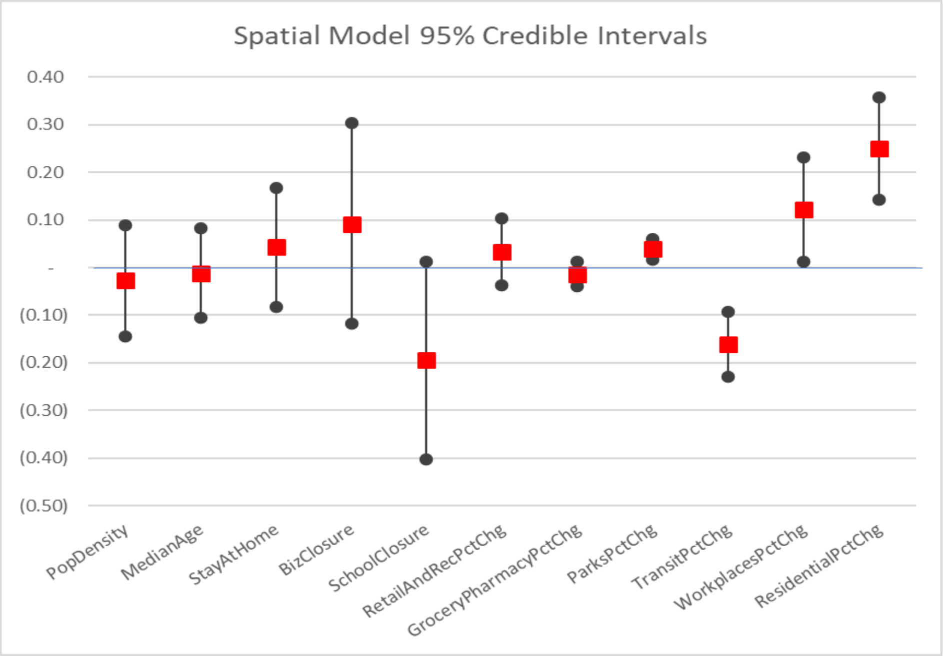
Spatial Model	Stochastic Search Variable Selection Spatial Model	State Random Effects Model
$Y_i = S_{state(i)} + \beta_1 PopDensity + \beta_2 MedianAge$ $+ \beta_3 StayAtHome + \beta_4 BizClosure + \beta_5 SchoolClosure$ $+ \beta_6 RetailAndRec + \beta_7 GroceryPharmacy + \beta_8 Parks$ $+ \beta_9 Transit + \beta_{10} Workplaces + \beta_{11} Residential$ $\underset{s\times s}{S} \sim Normal(0, \Sigma = (\underset{s\times s}{M} - \rho \underset{s\times s}{A})\tau_s)$ A = adjacency matrix for states where $A_{ij} = 1$ if state i is adjacent to j M = diagonal matrix with ith diagonal equal to neighbor count for state i*	$Y_i = S_{state(i)} + \beta_1 PopDensity + \beta_2 MedianAge$ $+ \beta_3 StayAtHome + \beta_4 BizClosure + \beta_5 SchoolClosure$ $+ \beta_6 RetailAndRec + \beta_7 GroceryPharmacy + \beta_8 Parks$ $+ \beta_9 Transit + \beta_{10} Workplaces + \beta_{11} Residential$ $\underset{s\times s}{S} \sim Normal(0, \Sigma = (\underset{s\times s}{M} - \rho \underset{s\times s}{A})\tau_s)$ $\beta_j = \gamma_j \delta_j \quad \gamma_j \sim Bernoulli(0.5) \quad \delta_j \sim Normal(0, \tau)$	$Y_i = R_{state(i)} + \beta_1 PopDensity + \beta_2 MedianAge$ $+ \beta_3 StayAtHome + \beta_4 BizClosure + \beta_5 SchoolClosure$ $+ \beta_6 RetailAndRec + \beta_7 GroceryPharmacy + \beta_8 Parks$ $+ \beta_9 Transit + \beta_{10} Workplaces + \beta_{11} Residential$ $\underset{s\times 1}{R} \sim Normal(0, \tau_s)$
WAIC: 442.67 **DIC: 429.1**	WAIC: 468.74 DIC: 454.6	**WAIC: 442.74** **DIC: 429.1**

- Our Spatial Model yields a spatial dependence parameter ρ=0.61. Therefore we conclude residual spatial dependence is present in these data. We hypothesize that for smaller states this spatial dependence would be higher.
- Among our final three models, the *TransitPctChg* and *ResidentialPctChange* covariates were consistently significant.
- Stochastic Search Variable Selection Model exhibited that covariates *SchoolClosure*, *TransitPctChg*, and *ResidentialPctChange*, should be included in the model. These covariates have mean posterior inclusion probabilities of 0.17, 0.46, and 0.93, respectively.
- WAIC and DIC for the Spatial Model and State Random Effects Model are the lowest. Perhaps this validates that states have a significant impact on the model.
- Model burn-in was 50k iterations then 2 chains x 300k iterations (thin=3). ESS counts of >1,000 and Gelman Diagnostics of 1 revealed that the models converged for all parameters.

*See Page 155 of *Bayesian Statistical Methods* (Reich & Ghosh)

Conclusions & Further Questions

- Conclusions:
  - There is a significant state effect for those states exposed earlier and hit the hardest.
  - Our Spatial Model points to a spatial dependence between states.
  - The Mobility Data has the most consistently significant predictive power in all our models. The *ResidentialPctChange* is the strongest predictor followed by *TransitPctChg*.
  - In the Spatial Model, *ResidentialPctChange* is positively correlated with the death rates, where *TransitPctChg* is negatively correlated.
  - The estimated mean for *SchoolClosure* suggests a strong negative correlation with the death rate, but due to the high level of uncertainty we cannot assign statistical significance to this covariate.
    - Our SSVS Spatial Model variable selection also identified this variable for inclusion.



- Further Exploration:
  - Perform analysis by county level/separate major cities (NYC).
  - Configure random effects model with statistically significant state effects and “all other states”.
  - Our analysis was truncated at 3/27/2020 due to predicting the 4/10/2020 number of deaths. How would our models change with more data?
  - Explore correlations further between covariates to remove any serial correlation or autocorrelation.
  - Incorporate the impacts of hospital bed capacity, hospital utilization and ventilator availability/shortages as covariates to see the significance on death rates.