

Lab 2: Impact of Government Interventions on COVID-19 deaths

Carolina Arriaga, Sunit Carpenter, Heather Rancic

Contents

1	Introduction	2
1.1	Research Question	2
2	Model Building Process	3
2.1	Data	3
2.2	Data Loading	4
2.3	Data Wrangling	4
2.4	Study Design	6
2.5	Model goals	7
2.6	Statistical Methods	9
3	Regression Table	13
4	Model Limitations	15
4.1	IID	15
4.2	Linearity	15
4.3	No perfect collinearity	16
4.4	Homoskedastic errors	16
4.5	Normality of errors	17
4.6	zero conditional mean	18
5	Discussion of Omitted Variables	19
6	Conclusion	20
7	References	21

1 Introduction

The year 2020 is known as the worst year in memorable history. When the first wave of Covid-19 hit the U.S. in full force in March of 2020, the U.S. government did not have a plan on how to handle a pandemic.

State governors were left to fend largely for themselves with little Federal government guidance or relief. They individually made decisions to provide protection and guidance to their constituents in order to keep them safe amidst the covid-19 pandemic whilst trying to keep businesses afloat and manage the economy to the best of their abilities. Individual states have opened and closed businesses, required masks, and taken other measures with varying degrees of success. Because of this, it's difficult to ascertain what works, what doesn't, and what the impact of those well intentioned measures are on covid-19 cases.

It is critical that we reach a deeper understanding of these differences in order to respond with appropriate public policies. Our team will clean and analyse data to see what patterns we can find that can help inform U.S. state governors as they seek guidance on what policies are effective and to help them to make the best decisions possible for their constituents. We recognize that the data will not be as perfect as we would like. We plan to clean the data in order to enable a conformed dataset to ensure our research is performed with scientific rigor and integrity to the best of our ability. The plan is to take into account differences between states but also see what correlations we can find overall across the U.S by leveraging data across 4/12/2020 through 3/25/2021.

1.1 Research Question

We want to understand if state level government interventions are correlated with Covid-19-deaths. We defined government interventions as mask required in public, shelter in place, quarantine for travellers as it relates to covid-19 deaths.

Our research question is: Is there an impact from government interactions on covid-19 death toll?

A Structural Equation Model was created to understand the relationship between the Covid-19 death toll and other factors, including the mandates (Fig 1).

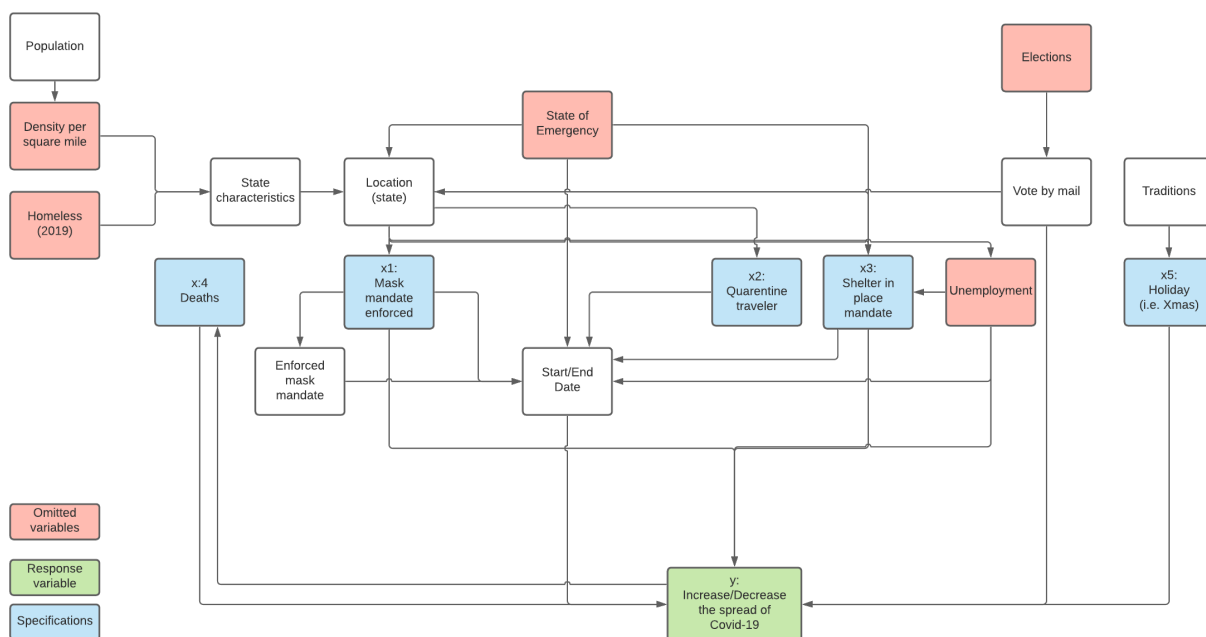


Fig 1: Structural Equation Model for Covid-19 deaths and government interventions.

After completing the SEM, we considered also two subquestions:

- Is the period of +/- fifteen days prior to and after a holiday increasing the Covid-19 deaths?
- Is the population density influencing Covid-19 deaths?

From all the US holidays, we focused this study on three of them: Mother's day, Election day and Thanksgiving. We selected Mother's day because it's an important US holiday that occurred early in the pandemic, we chose election day because we expected many people might ignore recommended mandates in order to vote in person which may have an impact in Covid-19 deaths, and we picked Thanksgiving because it's the most popular travel holiday in the US.

Mandates were not enforced in multiple states consistently during the holidays. We calculated the death toll during a holiday period by subtracting the death toll at the end and start dates. Then we evaluated the effect of the mandates. For each covariate, we aggregated the total days when the mandate was in effect.

For example, the case when the mandate is enforced from the holiday to the next three days was considered as a total of four days.

2 Model Building Process

2.1 Data

In order to answer the research question, our team has gathered information from the following sources which are described below:

- The raw data was sourced from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [1]. Data columns we are considering for our study:

\caption{(#tab:table__raw)Table 1. Filtered raw data from source}

spec	description
US State	US State name or district
Date	Date in format MM/DD/YYYY
Covid-19 Cases	Counts include confirmed and probable (where reported)
Mask mandate enforced - start	Date in format MM/DD/YYYY
Mask mandate enforced - end	Date in format MM/DD/YYYY
Shelter in place - start	Date in format MM/DD/YYYY
Shelter in place - end	Date in format MM/DD/YYYY
Quarantine for travelers - start	Date in format MM/DD/YYYY
Quarantine for travelers - end	Date in format MM/DD/YYYY
Population	Population from 2018
Density	Population per square mile
Elected party	State results in 2020 presidential elections

\end{table}

- Holidays were taken from the python-holidays module.
- State results from presidential election were taken from wikipedia.

2.2 Data Loading

We wrangled the data to achieve the following structure:

```
\begin{table}
\caption{(\#tab:table_processed)Table 2. Processed data from raw data}
```

spec	description
US State	US State name or district
Covid-19 Cases	Case count for the state per 100k people
Deaths	Death toll for the state per 100k people
Mask in public	Aggregated days when mask in public mandate was required (end date - start date)
Shelter in place	Aggregated days when shelter in place was enforced (end date - start date)
Quarantine for travellers	Aggregated days when Quarantine for travellers was enforced (end date - start date)
Density	Population per square mile
Party	Democrat or Republican

```
\end{table}
```

2.3 Data Wrangling

The raw data included a full year across US States. We sliced the data specifically specifically for the three aforementioned holidays. We considered population density as it relates the three holidays we chose to study.

The population density might affect how fast the virus spreads. We divided the states in three groups: High, Med, Low. High density includes all the states with more than 200 people per square mile, then med density is the interval between a 100 and 199, and low density is any value less than 100.

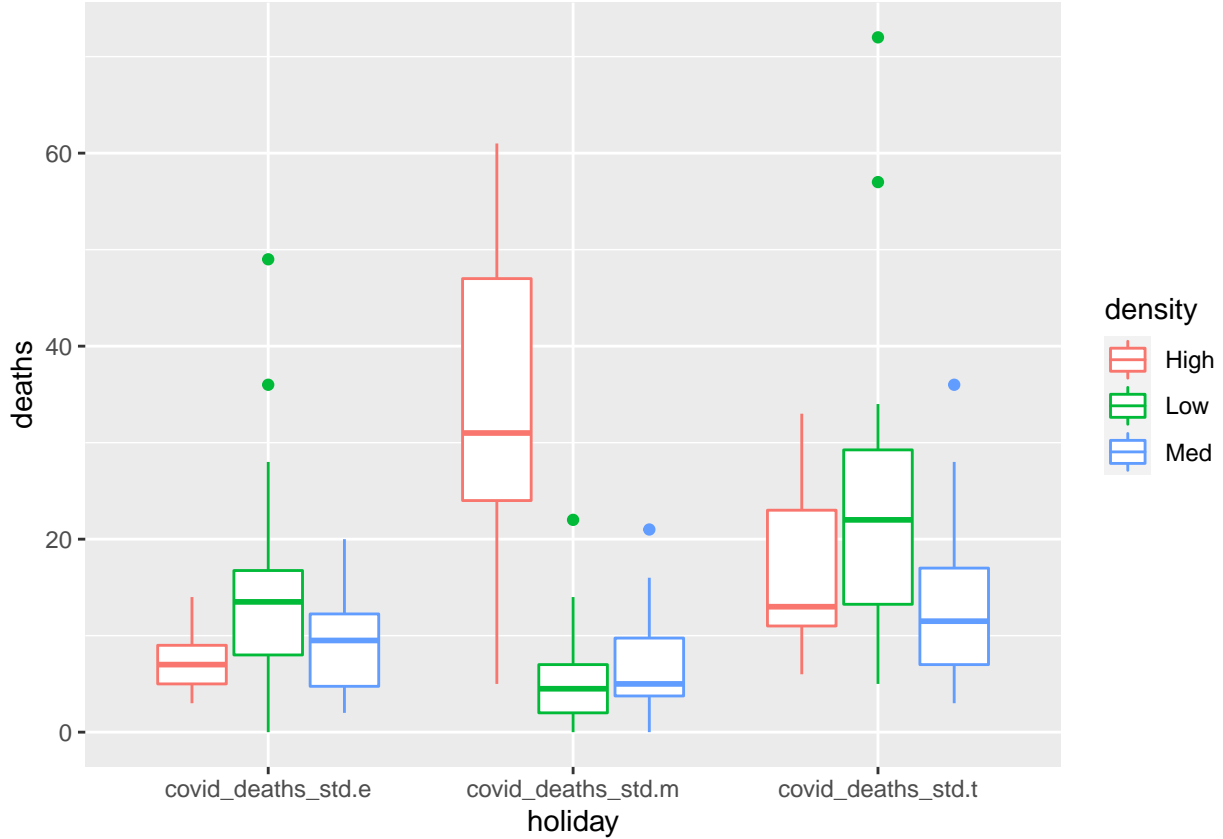


Figure 2. Death toll during Election day, Mother's day and Thanksgiving.

The plot above shows how in low density states, the spread during multiple holidays didn't change significantly. On the other hand, medium and high density populations seem to have a lot of variation during thanksgiving. Finally, we notice how high density populations are more susceptible to change during holidays. The two states with very high death toll during the Thanksgiving holiday are Texas and Indiana. These two are far away from the average value of approximately 35 amongst the three holidays.

The next plot shows the impact of interventions as they relate to the selected holidays. Red observations represent Republicans and blue represent Democrats. We noticed that the shelter in place mandate seem to be more effective for the holiday earlier in the year (Mother's day). It could be that in the holidays later in the year, the mandate was lifted due to Covid-19 fatigue.

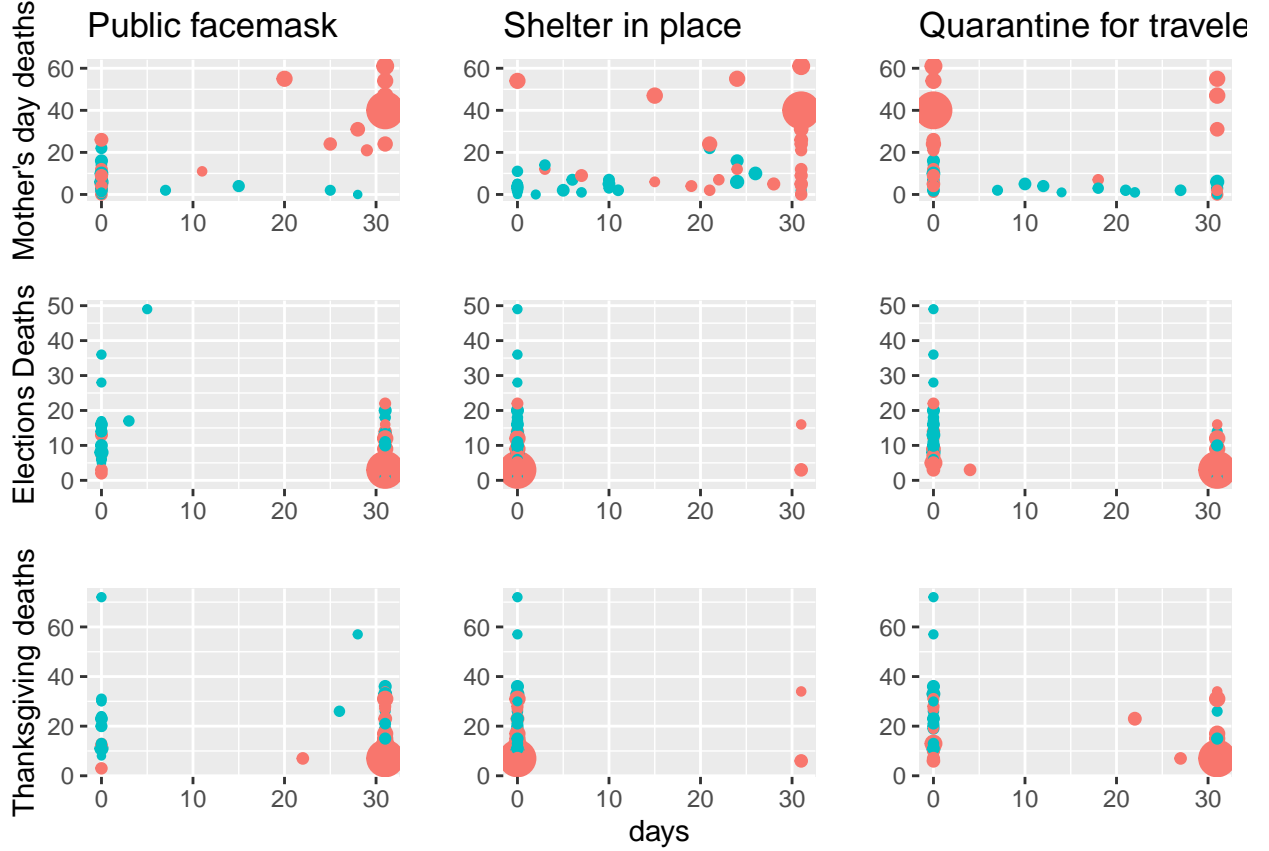


Figure 3. Death toll during multiple holidays and relationship to Public face mask, shelter in place, and quarantine for travelers mandates.

We focused on a specific holiday, such as mother's day. The reason for focusing on this particular period of time is because we can see how multiple states were implementing interventions during the 30 days (holiday +/- 15 days). As we can see in the other holidays, the interventions become more binary (zero or the max value), sometimes with very few observations in each of them.

For example, the shelter in place with only two states (California and New Mexico) with the mandate effective. We don't believe that would be a representative option to measure the impact of that intervention. This study will cover only the Mother's day period. California is a state with the highest population. New Mexico is the poorest state in the USA and with less hospitals and beds for Covid-19 patients [9].

2.4 Study Design

We want to measure the impact of government interactions in the Covid-19 death toll during the Mother's day period. Each observation will be recorded at a state level. The model will be created by 51 observations, each of them composed of the following features:

\begin{table}

\caption{(#tab:processed_data)Table 2. Processed data from Mother's day}

code	description	domain
D	Death toll per 100k people	0 - 61
M	Aggregated days when mask in public required	0 - 30
S	Aggregated days when shelter in place was enforced	0 - 30
Q	Aggregated days when Quarantine for travelers was enforced	0 - 30
P	Elected party affiliation	Democrat (0) or Republican (1)

\end{table}

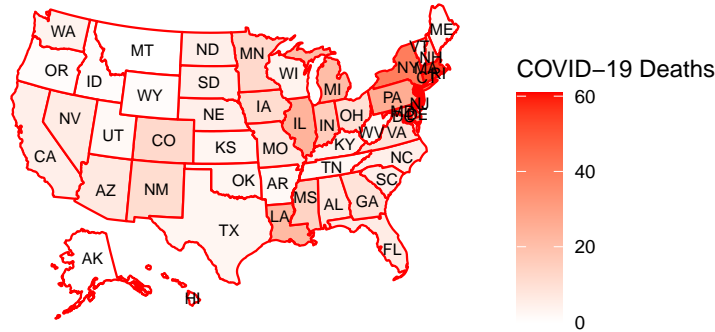
2.5 Model goals

Three models seek to explain the impact of government interventions considering the specifications stated in the SEM in Figure 1, which showed the theoretical relationships between the increase or decrease of Covid-19 deaths and multiple factors. The heatmap below shows similar patterns by Covid-19 deaths and cases. We've chose deaths as our response variable because we believe there's a smaller variation embedded in how the deaths were recorded compared to Covid-19 cases across all the USA.

It must be flagged here that Oregon changed the way they reported the total cases of confirmed cases and switched from number of tests to number of persons tested. There is a lot of variation from state to state that is linked to different approaches when measuring total cases. For this reason, we decided to use the variable of death toll instead of confirmed cases.

U.S. States

COVID-19 Deaths by State.



U.S. States

COVID-19 Cases by State.

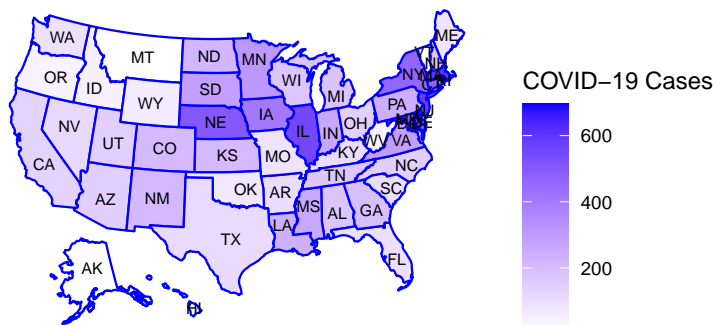


Figure 4. (Left) Death toll by state. (Right) Covid-19 cases by state.

2.5.1 Model 1: Limited model

The mask mandate has been subject to controversy since the pandemic started, having multiple countries and the WHO providing multiple standards on face mask use in community settings [2]. We believe that there is a negative correlation between Covid-19 deaths and the total days the mask in public mandate was required.

$$y = D = \beta_0 + \beta_1 M$$

2.5.2 Model 2: Extended model

The extended model includes the rest of the government interventions which are shelter in place and quarantine for travelers. We believe that during the holiday season, the mobility was increased in all the states. Is not always the case that the interventions coincided with the holidays. If our original intuition about mask mandate has a large effect, we expect to see that in the holidays when people had a higher interaction and consequentially, the Covid-19 deaths would be reduced if more interventions were in place.

In all of the variables proposed in the extended model we expect a negative correlation to the response variable.

$$y = D = \beta_0 + \beta_1 M + \beta_2 S + \beta_3 Q$$

2.5.3 Model 3: Full model

The full model will take into account the political party the state voted during 2020 elections. We chose this variable because democratic states supported the introduction of government interventions in a larger degree compared to republican states:

$$y = D = \beta_0 + \beta_1 M + \beta_2 S + \beta_3 Q + \beta_4 P$$

2.6 Statistical Methods

2.6.1 EDA

Before running the proposed models, we looked the correlation matrix of the variables of interest.

The correlation matrix (Figure 5) provides the big picture of the relationships between all the variables of interest. It seems that Covid-19 deaths is not normally distributed, showing a skew to the left. Days of shelter in place has a large skew to the left. The remaining variables, days of face mask enforced, days of quarantine for travelers, and days of mask enforced during holiday seemed to be part of two populations.

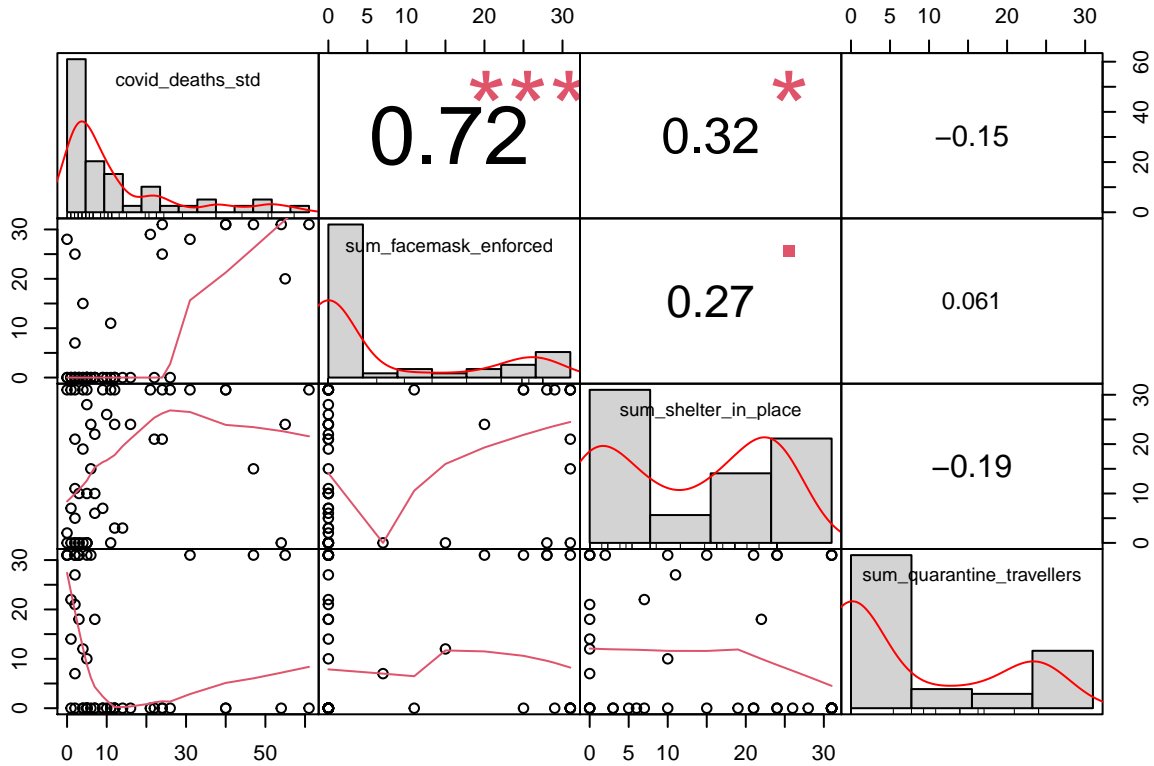


Figure 5. Correlation matrix for death toll, public face mask, shelter in place, and quarantine for travelers during Mother's day.

2.6.1.1 Facemask required in public spaces

Deaths and public facemask – Mother's day

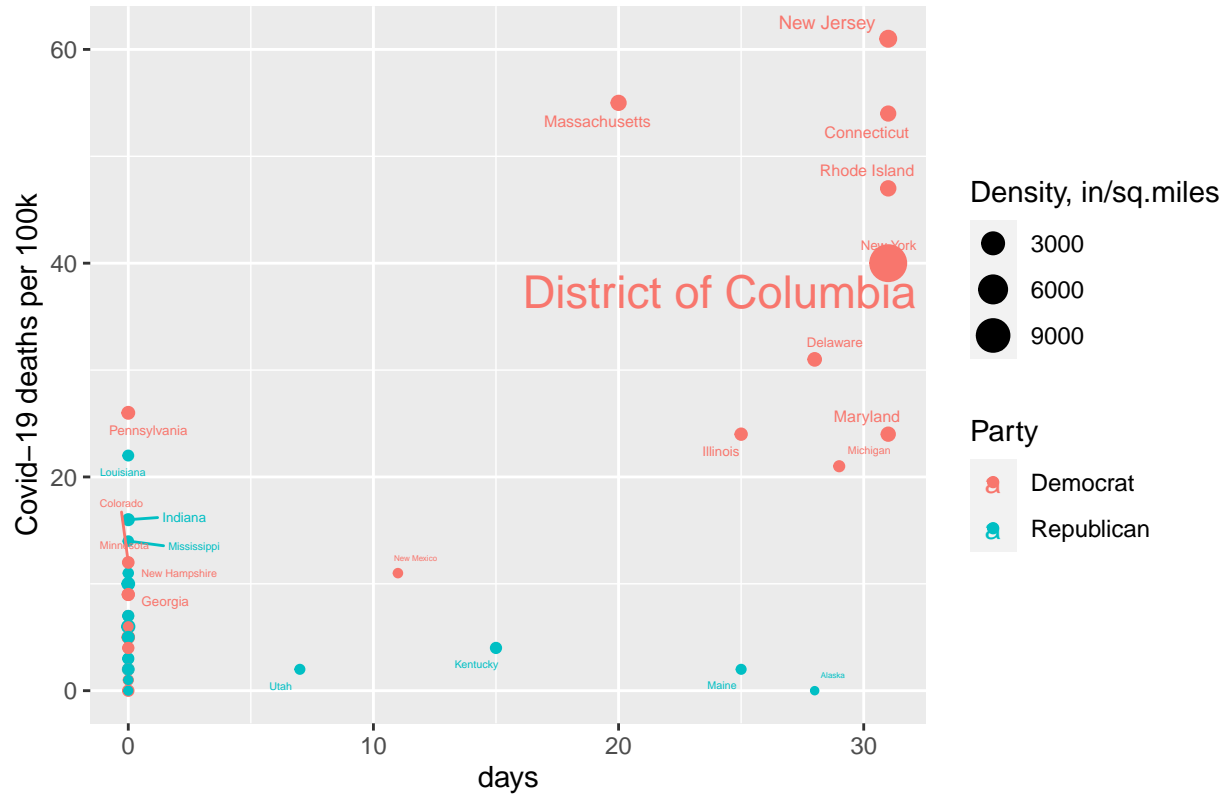


Figure 6. Relationship between death's and effective days of public face masks.

The states with lowest Covid-19 deaths are Hawaii, Alaska, Maine, Vermont and Oregon. They all seem to have a similar population density except for Alaska. Hawaii is naturally isolated to the rest of the US States. It might even be considered an outlier in terms of connectivity with the rest of the states [3].

It is evident that states with the highest density seem to land towards the top right of the plot (Figure 6). We can also notice that most of the democratic states had a longer public face mask mandate requirement period while the majority of republican states didn't required it.

The second state with lowest deaths is Vermont, which has been set as an example to the US, following a strict interventions from the beginning of the pandemic [4]. Oregon seems to have follow a strict path similar to Vermont, in both states, there has been 'relatively good adherence with guidelines' [5].

2.6.1.2 Shelter in place

Deaths and Shelter in place – Mother's day

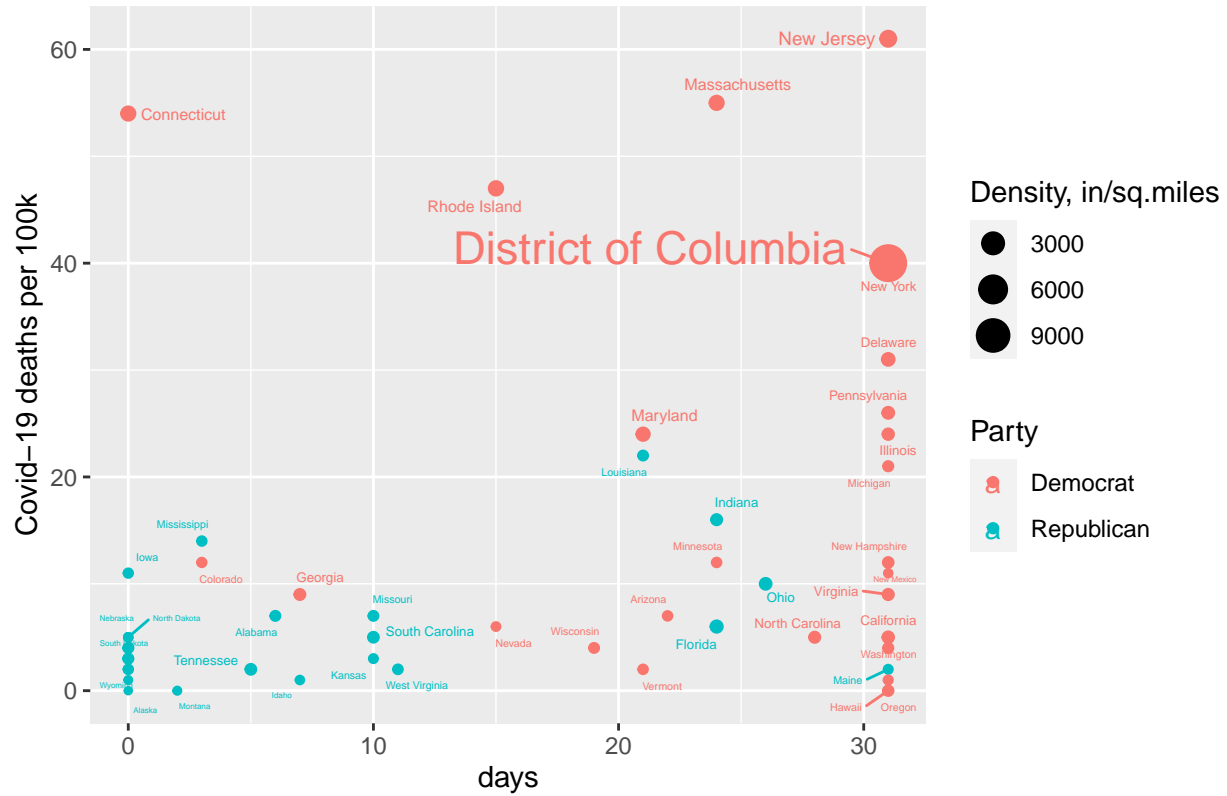


Figure 7. Relationship between death's and effective days of shelter in place.

In this plot we see only two points with a longer shelter in place mandate compared to the rest of the states. In general, days of shelter in place doesn't seem to have an influence on our response variable. We do see that most Democrat states seem to had the mandate for longer compared to Republican states.

2.6.1.3 Quarantine for travellers

Deaths and Quarantine for travelers – Mother's day

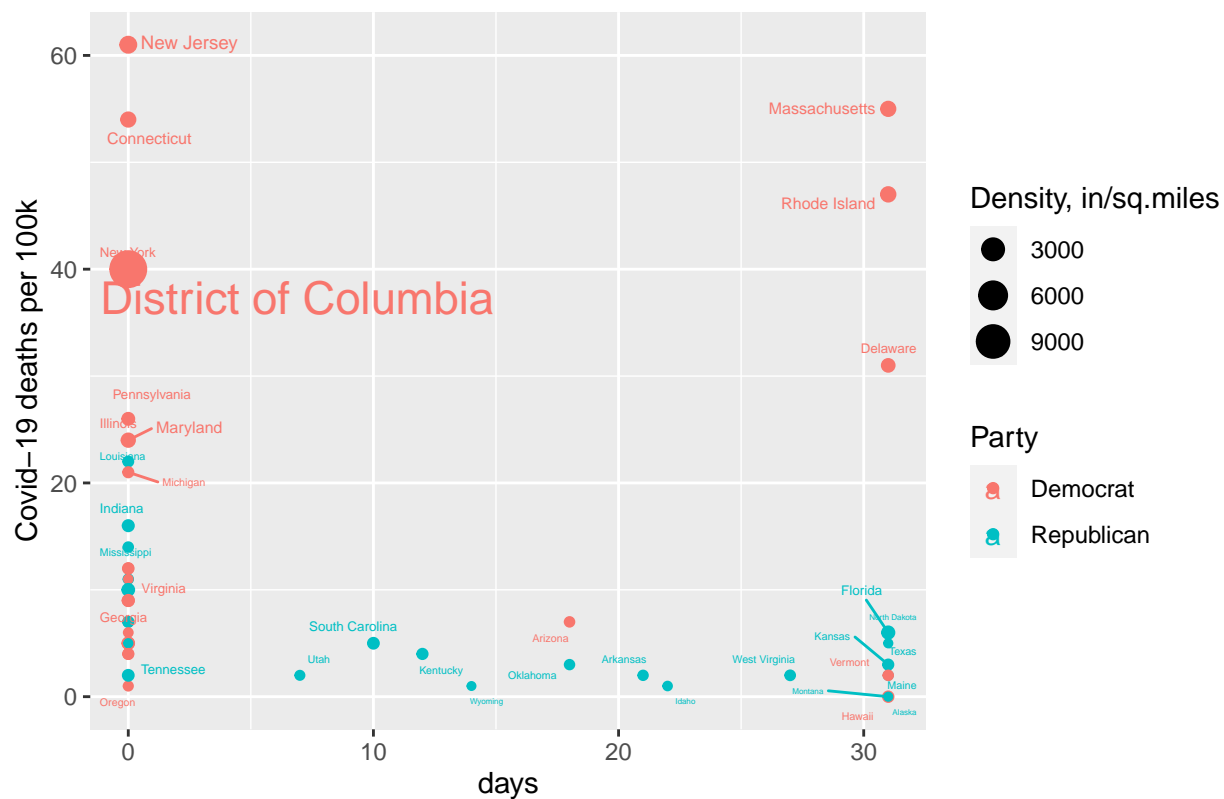


Figure 8. Relationship between death's and effective days of quarantine for travelers.

The histogram of quarantine for travelers (Figure 6) shows the distribution of quarantine for travelers with a large skew to the left. It implies that most of the states didn't have the mandate for a long time.

Possibly a transformation of the variable quarantine for travelers might be used. We will evaluate the necessity in the following scatter plot.

2.6.1.4 Party Elected in 2020 President Election

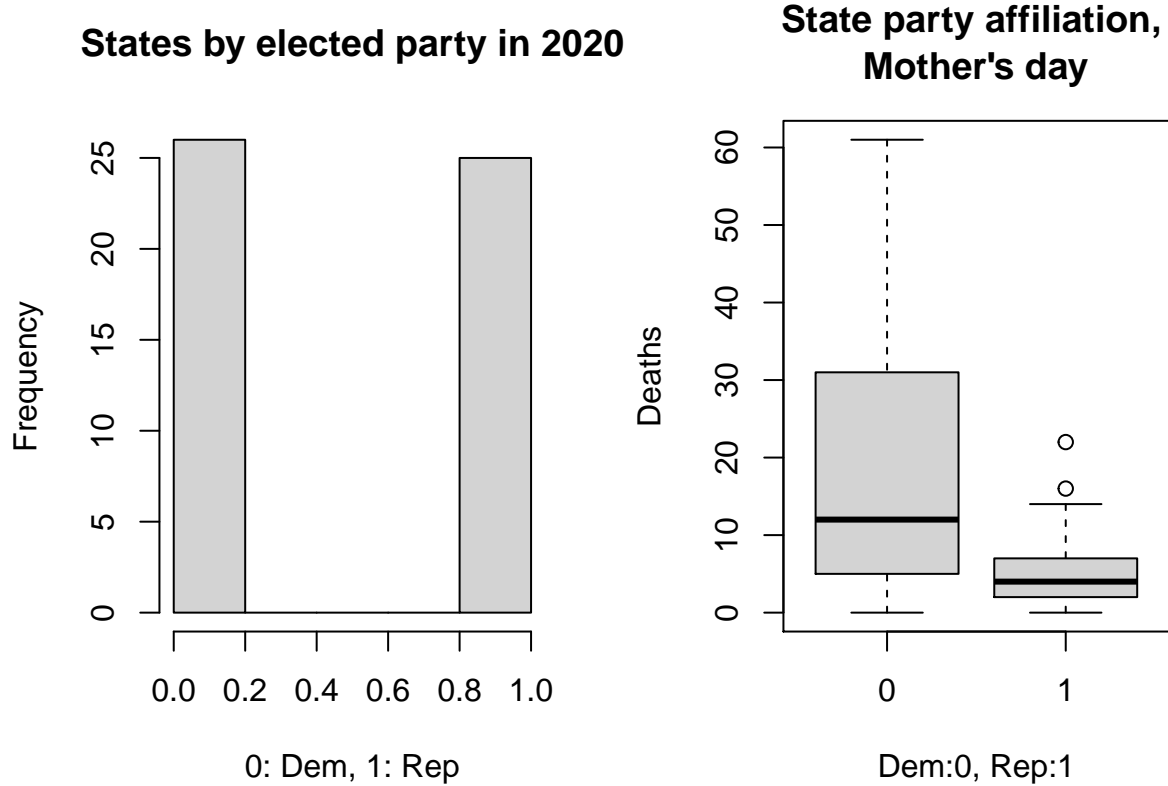


Figure 9. (Left) Histogram of state affiliation. Figure 10. (Right) Distribution of death's by party affiliation in 2020 presidential elections.

The histogram shows that the states are almost evenly divided by Republican versus Democrat, where the latter has a minor lead (+1). Democrats have a larger average death toll and a wider distribution compared to Republicans. We also want to point at the impact of high population density may be the reason for such variation (Figure 7) because all the highest population states are democratic.

3 Regression Table

3.0.1 Model 1: Optimized Limited model

To reduce standard errors we modeled the public face mask required variable as a second degree polynomial. We observed that the residuals didn't have a significant deviation from zero after the optimization (Figure 11).

$$y = D = \beta_0 + \beta_1 M + \beta_2 M^2$$

3.0.2 Model 2: Optimized Extended model

The extended model was optimized by modeling the shelter in place variable as a second degree polynomial, residuals showed a smaller deviation from zero.

$$y = D = \beta_0 + \beta_1 M + \beta_2 M^2 + \beta_3 S + \beta_4 S^2 + \beta_5 Q$$

3.0.3 Model 3: Full model

We kept the model two and added the party affiliation variable as planned.

$$y = D = \beta_0 + \beta_1 M + \beta_2 M^2 + \beta_3 S + \beta_4 S^2 + \beta_5 Q + \beta_6 P$$

After analyzing the variance inflation, we decided to drop the variable shelter in place creating an optimized model three.

$$y = D = \beta_0 + \beta_1 M + \beta_2 M^2 + \beta_5 Q + \beta_6 P$$

We present the results of the three models studied.

=====				
Dependent variable:				

Death toll per 100k inhabitants				
	M	M+S+Q	M+Q+P	
	(1)	(2)	(3)	

mask in public	-0.451	-0.082	-0.216	
	(1.550)	(1.695)	(1.516)	
mask in public sqrd	0.047	0.035	0.036	
	(0.051)	(0.056)	(0.052)	
shelter in place		0.753		
		(0.780)		
shelter in place sqrd		-0.021		
		(0.024)		
quarantine travelers		-0.194	-0.134	
		(0.136)	(0.154)	
party			-5.926	
			(4.339)	
Constant	6.802***	4.992	11.515***	
	(1.018)	(4.287)	(2.476)	

optimized model	Yes	Yes	Yes	
Observations	51	51	51	
R2	0.548	0.599	0.605	
Adjusted R2	0.529	0.554	0.570	

```
## Residual Std. Error      10.681 (df = 48)      10.388 (df = 45)      10.198 (df = 46)
## F Statistic             29.043*** (df = 2; 48) 13.431*** (df = 5; 45) 17.596*** (df = 4; 46)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

4 Model Limitations

4.1 IID

Each state is geographically dependent of the neighbors and political affiliations. Even when interventions were in place by the government, there is no way to know if the people followed the guidelines en mass. The model hardly complies with this requirement, however we consider for this study that the sample is representative of the USA population at the time period of data collection used in the study and the sample size can be considered not small.

4.2 Linearity

For the model one, we have only one specification: mask required in public. We found the residuals had a parabolic trend. For model two we shelter in place residuals were also parabolic. The linearity assumptions was threaten and we optimized the model modeling public face mask and shelter in place as polynomials of second degree. See Figure 11 for model three optimized model plots.

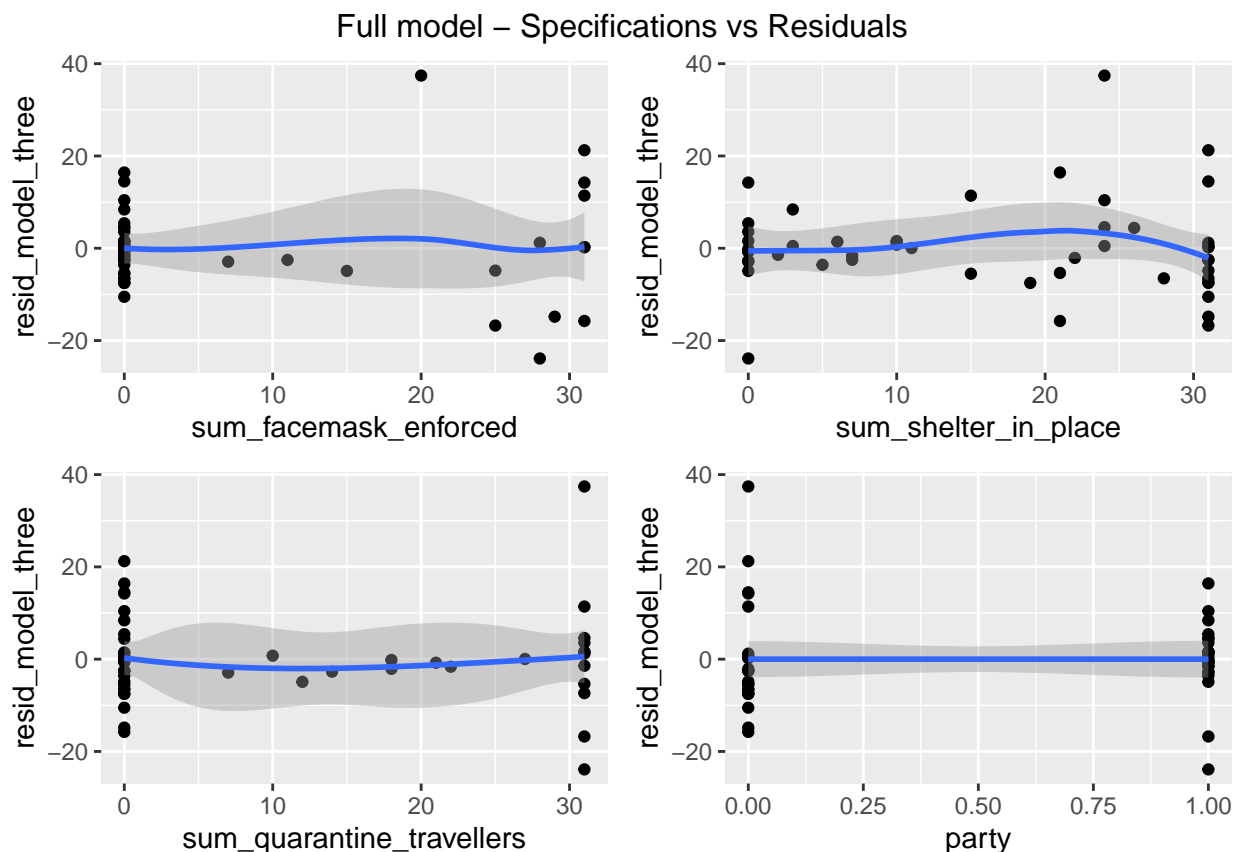


Figure 11. Specifications vs residuals for optimized model three.

4.3 No perfect collinearity

We know the variables chosen didn't have perfect collinearity because we were able to run the model without singularity problems. Also, we analyzed in the full model which variable is bringing the most inflation to our predictions. Public face mask and shelter in place are both increasing the model variation significantly. We decided to drop shelter in place from model three to optimize the error in the model. We

kept public face mask because it's the variable of interest in this study.

```
##      sum_facemask_enforced I(sum_facemask_enforced^2)
##                37.388374                37.627015
##  sum_quarantine_travellers                party
##                1.127255                1.278619

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.514870    2.476034  4.6505 2.811e-05 ***
## sum_facemask_enforced -0.216492    1.515808 -0.1428  0.8871
## I(sum_facemask_enforced^2) 0.036378    0.051887  0.7011  0.4868
## sum_quarantine_travellers -0.134189    0.154077 -0.8709  0.3883
## party          -5.926448    4.339074 -1.3658  0.1786
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After dropping the shelter in place specification from model three a slight improvement in the R square adjusted was obtained.

4.4 Homoskedastic errors

It is clear that the model three has a clear deviation from zero when values are small. A similar behavior is seen in the Scale-location plot where the slop dramatically decreases as the values increase, instead of a flat smooth curve. For this reason, the validity of the t-tests and p-values are threatened.

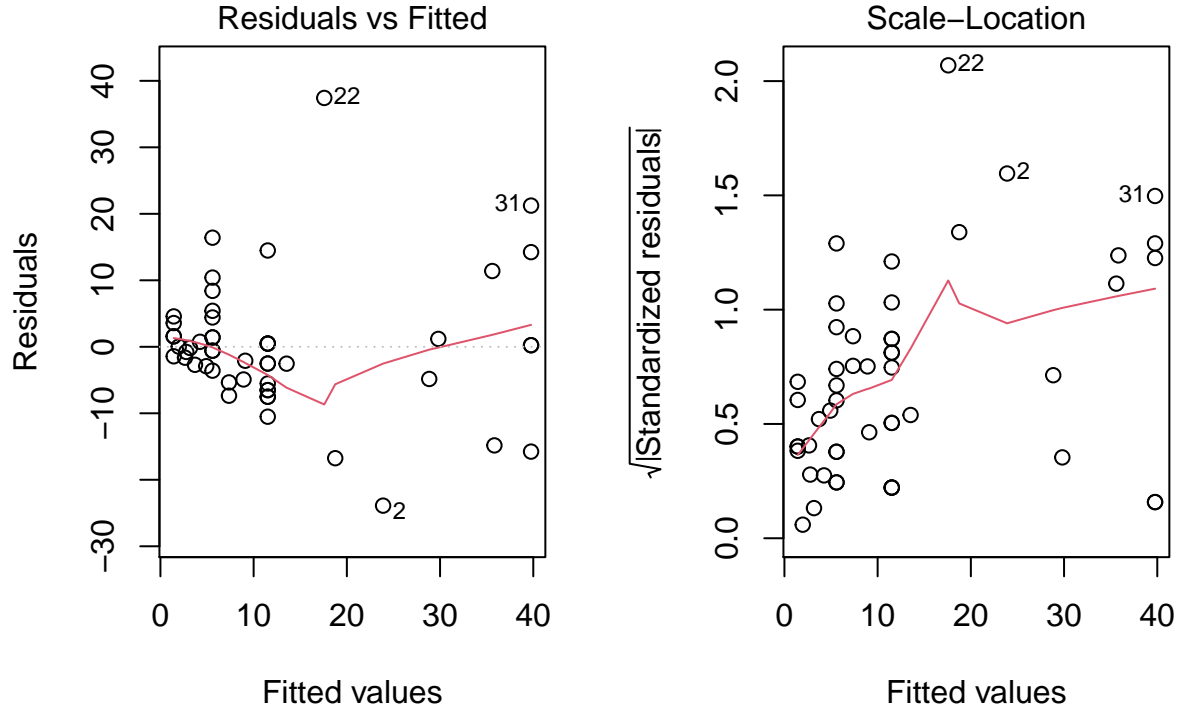


Figure 12. (Left) Residuals vs Fitted for optimized model three show deviation in the middle. Figure 13. (Right) Scale-Log shows a constant increase in the standardized residual values as the fitted value increases.

4.5 Normality of errors

The qq-plot shows that the errors are not normally distributed in both ends of the tail for values less than -1 and 1. The validity of the standard errors reported in this study is threatened.

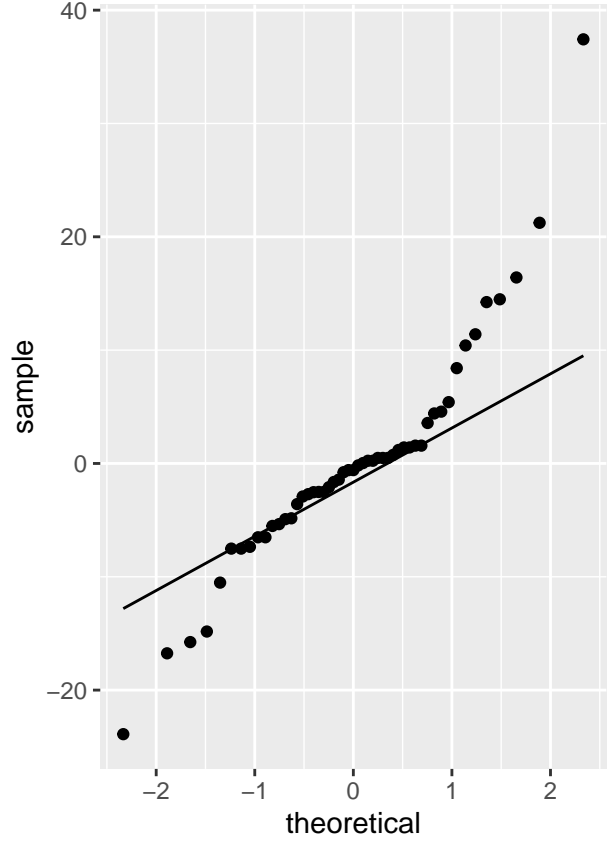
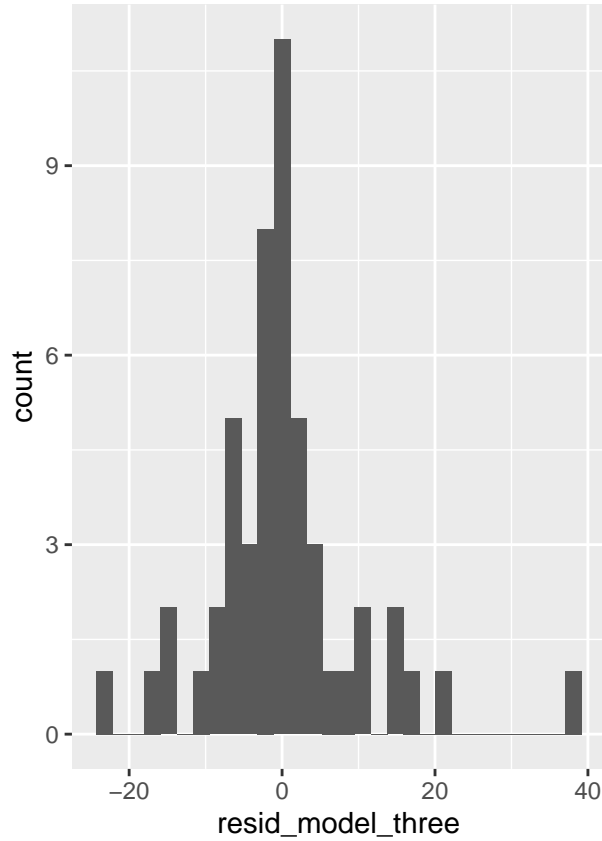
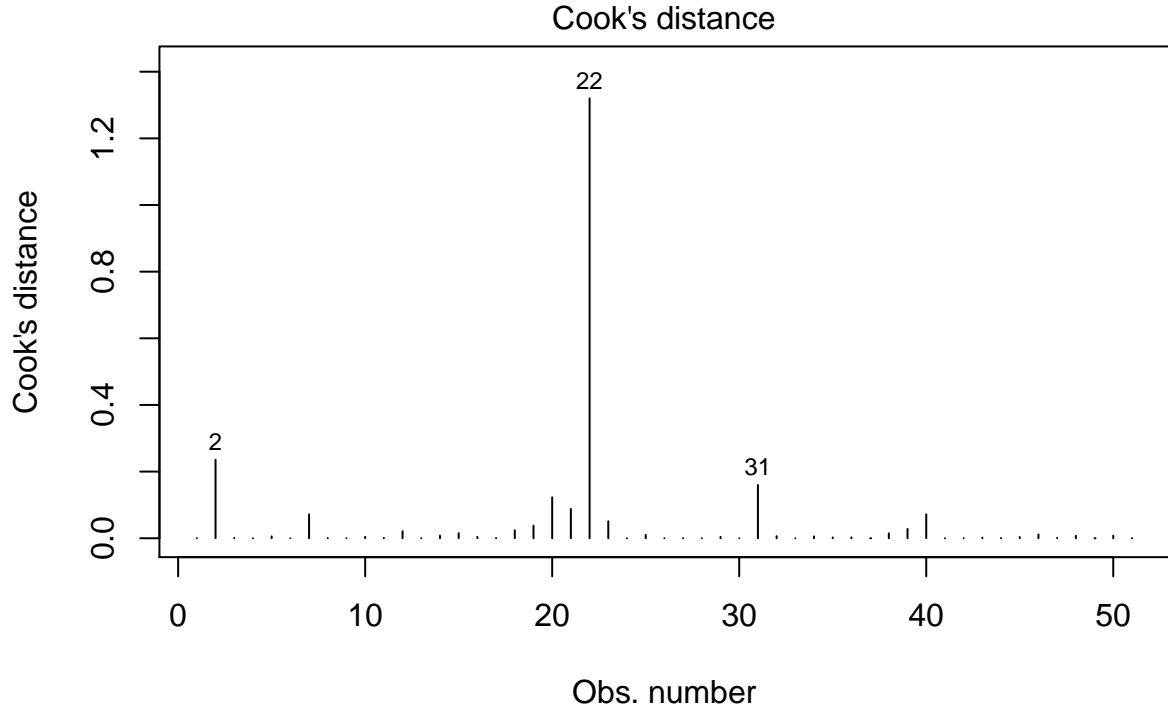


Figure 14. (Left) Residuals histogram for model three. Figure 15. (Right) QQ-Plot of residuals of optimized model three.

4.6 zero conditional mean

The average residual deviates from zero in a parabolic share (Figure 14). Also, we believe that Massachusetts might be an outlier in the data set because of the Cook's D value. We theorize this is due to Massachusetts proximity to New York, which was the epicenter of the pandemic earlier year.



$\text{lm}(\text{covid_deaths_std} \sim \text{sum_facemask_enforced} + \text{l}(\text{sum_facemask_enforced}^2) + .$

Figure 16. Cook's D for optimized model three.

5 Discussion of Omitted Variables

We believe that there are multiple factors at play when it comes to considerations around COVID-19 deaths. It is likely that our models are biased due to omitted variables. We listed out those variables as possible limitations to our models.

Density per square mile (d)

We were not able to get accurate current data on the population density for all of the states. Also, getting the population density of a given state may not give us a true picture of the population density. For example, majority of the people in Nevada live in the Las Vegas area. Nevada has a low population density at the state level, but in reality has a much higher population density as the population is concentrated around the city of Las Vegas. Covid-19 infection is spread by close contact with infected people and higher the population density leads to higher rate of infection and death. [6].

$D = M$, $OV=d$ $D \& M$ (-), $D \& d$ (+), $M \& d$ (+) Away from zero, large

Restaurant occupancy (r)

Restaurants were closed by mandate inconsistently across states. Even for those restaurants that were ordered to close, some businesses chose to defy the guidance and opened anyway. The volume of patrons visiting restaurants during Mother's day, may have influenced the spread and deaths related to Covid-19.

$D = M$, $OV=r$ $D \& M$ (-), $D \& e$ (+), $M \& r$ (-) Towards zero, small

Unemployment (u)

Table 1: Table 3. Omitted variable bias for optimized model three.

OV	Bias	Effect_size
Density per square mile	Away from zero	Large
Restaurant occupancy	Toward from zero	Small
Unemployment	Toward from zero	Small
Homelessness	Away from zero	Small
Covid-19 Fatigue	Away from zero	Large

The US unemployment rate peaked at 14.8% in April, 2020. [8] Millions of people lost their jobs and were not able to work. We do not have data on the impact of people not being able to go to work on the Covid-19 infections. If the unemployed stayed at home, they had the potential of infecting others in the household. On the other hand, by staying at home, the unemployed had less exposure to Covid-19 compared to physically going to work in an office/restaurant type environment.

D = M, OV=u D&M (-), D&u (-), M&u (-) Towards zero, small

Homelessness (h)

We do not have data on the homeless population for each state and how they potentially contributed to the spread of Covid-19 infections. Also, homeless people are less likely to have the resources to purchase masks and other means of protection from Covid-19 infections. We also do not have the data on location of the homeless population - urban or rural areas.

D = M, OV=h D&M (-), D&h (+), M&h (+) Away zero, small

Covid-19 Fatigue (f)

While it is difficult to measure Covid-19 fatigue, as the pandemic wore on, people had increased difficulty with social distancing. People began to weigh mental health impact on isolation vs. the risk of following safety guidelines in favor of ignoring safety guidelines. We saw evidence of this from people gathering for spring break and end of year holidays such as Thanksgiving and Christmas.

D = M, OV=f D&M (-), D&f (+), M&f (+) Away zero, large

6 Conclusion

After evaluating three models and optimized them to comply with the classical linear model assumptions, we found the model three was the best. The optimized model three used only data from mother's day (+/- 15 days), a time relatively early in the pandemic when guidelines in multiple states were in place and people were heeding the guidelines.

The optimized model three considered public face mask mandate, quarantine for travelers and the affiliated party during the presidential election as covariates. We encountered threats to validity of the standard error, F-test, t-value and p-value due to homoskedastic error and fail to comply with normality of errors. The results of this study will focus only in the coefficients of the input variables of the optimized model three.

$$y = D = 11.51 - 0.22 \cdot M_d + 0.04 \cdot M_d^2 - 0.13 \cdot Q_d - 5.92 \cdot P_d \quad (1)$$

We found a negative correlation between deaths and public face mask, quarantine for travelers and the state's political affiliation. This relationship matched our initial theory.

Based on equation one, we can see that over a thirty day period, a maximum of six lives were saved per hundred thousand people in the states that had a public mask mandate in effect during Mother's day.

Quarantine for travelers saves approximately three lives per hundred thousand. The political affiliation of the state based on 2020 presidential elections had a negative correlation with deaths. If the state was Republican, an estimated of 6 lives were saved per thousand people. We found this effect may be influenced due to low population density in republican states. Highest population density states were largely democratic (see Figure 6).

Having said that, we must emphasize that our optimized model three is not a good estimator for deaths. The R square adjusted is low, with a value of 0.57, meaning that there's a large amount of variation not explained by our model.

7 References

- [1] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis.* 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1
- [2] Feng et al. (2020, March). Rational use of face masks in the COVID-19 pandemic. Retrieved April 05, 2021, from [https://www.thelancet.com/journals/lanres/article/PIIS2213-2600\(20\)30134-X/fulltext](https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(20)30134-X/fulltext)
- [3] <https://time.com/5915084/hawaii-covid-coronavirus/>
- [4] <https://www.bloomberg.com/news/articles/2021-01-06/n-y-s-neighbor-keeps-cases-lowest-in-u-s-throughout-pandemic>
- [5] <https://www.wweek.com/news/2020/12/23/oregon-fares-better-than-nearly-all-states-by-two-key-covid-19-measures/>
- [6] <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html> [7]
<http://electionlab.mit.edu/sites/default/files/2020-12/How-we-voted-in-2020-v01.pdf> [8]
<https://fas.org/sgp/crs/misc/R46554.pdf>
- [9] <https://www.nytimes.com/2020/04/24/us/coronavirus-new-mexico.html>
- [10] <https://www.npr.org/2020/10/24/927384374/rural-midwestern-communities-hit-hard-as-coronavirus-surges-in-u-s-again>