International Journal of Statistics and Probability; Vol. 10, No. 1; January 2021
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Some Multiple Regression Models for the Number of COVID-19 Cases and Deaths in the United States

John Tenenholtz[1], Florence George[1], Sneh Gulati[1]

[1] Department of Mathematics and Statistics, Florida International University, Miami, FL, USA

Correspondence: Florence George, Department of Mathematics and Statistics, Florida International University, Miami, FL, USA

Received: September 28, 2020 Accepted: November 12, 2020 Online Published: November 30, 2020

doi:10.5539/ijsp.v10n1p28 URL: https://doi.org/10.5539/ijsp.v10n1p28

Abstract

The whole world has been affected by the COVID-19 pandemic. It has changed life drastically, affecting both social and business behavior and causing major economic distress throughout the world. The disease is often denominated a “novel coronavirus,” meaning that it is a new strain, that none of us carry antibodies to it and that there is much to be learned about its pathology. This obviously makes it hard to control. While several countries seem to have grasped ways to contain the virus, the United States (the “U.S.”) has seen steady growth in the number of cases and deaths. This paper uses multiple regression models to examine the differences among the several U.S. states in the numbers of cases and deaths and investigates several possible contributing factors to these totals.

Keywords: COVID-19, novel coronavirus, multiple regression

1. Introduction

The first confirmed case of COVID-19 was reported in the US on January 22, 2020. As of October 30, 2020, the U.S. had recorded over nine million confirmed cases and more than 230,000 deaths. The initial cases in the US were travel related and for some time there appeared to be no indications of community spread. On February 26, 2020, the CDC reported its first known case of community spread when a man in California became infected with no travel history or known contact with an infected person. At that time, the U.S. had only 15 cases in total, 12 of them travel related (Hauck, Gelles, Bravo & Thorson, 2020).

Initially, it was believed that the virus could be controlled through testing and contact tracing and that, like the flu, it would dissipate in the summer heat. Unfortunately, the virus is resilient and has defied efforts to contain it in the U.S. The U.S. surpassed 10,000 cases on March 19, 100,000 cases on March 26 and 1 million on April 28. To put the numbers in perspective, the New York Times provided the following comparison: By summer, the total number of infections in the U.S. was more than the combined populations of Nebraska, Vermont and Montana and the national death toll by summer exceeded the population of Syracuse, N.Y. (Almukhtar, Bloch, Aufrichtig, Calderone, Collins, Conlen et al., 2020).

The overall federal response consisted of a series of non-mandatory guidelines combined with inconsistent messaging from various federal agencies. Americans were first urged to stay at home under Presidential guidelines issued on March 16, 2020. Because these guidelines were not mandatory, the states were forced to respond independently, with various actions taken and results obtained. Many (but not all) states issued stay at home orders; however, the level of restriction, the length of the orders and the enforcement of the orders were far from uniform across states. The overall result was that the economy was shut down in most states for the better part of April. Millions lost their jobs and unemployment reached depression-era levels. The virus has led to some drastic changes in American lifestyles. A visit to the theater (should one be open) or dining out at a restaurant or even simple nights out with friends are events that appear to be fraught with danger. Large gatherings are forbidden in most states, weddings have been postponed or scaled down to name a few changes. Many wonder if life will ever return to what it was pre-pandemic.

As with the state responses, the spread of cases within the various states has not been uniform. Northeastern states like New York and New Jersey were hit hard early and saw wide-spread pain but clamped down early so that the disease became manageable. On the other hand, states like Florida, Texas and Arizona were spared early in the pandemic but later saw cases soar, and appeared to be on a more downward trend at the time this analysis was undertaken. The difference among the states in the number of cases could have multiple causes. The states differ in their demographics, education level, population density, economic prosperity, political control of their emergency responses and the actual initial response (such as the extent, length, and enforcement of stay at home orders). All these factors can potentially influence the initial outbreak of the disease as well as its trajectory.

In addition to attempts to reduce the virus' spread, attempts were also made to expand testing. Initial problems with testing included 1) defective tests, (2) insufficient numbers of testing kits, and (3) delays in getting test results from medical lab facilities (Shear, Goodnough, Kaplan, Fink, Thomas & Weiland 2020). Despite these issues, testing did expand, though at less than ideal rates.

The number of COVID-19 cases per state and the statewide trajectory of these cases is widely available and published almost daily. See for example: ("The New York Times, The Coronavirus Outbreak", 2020), ("Centers for Disease Control and Prevention, Coronavirus (COVID-19)", 2020), ("Johns Hopkins Coronavirus Resource Center", 2020). There have also been many epidemiological and clinical studies on Covid-19. For example, Yang and Wang et. al. (2020) examined 150 patients in Wuhan, China early in the epidemic to investigate the clinical predictors of mortality due to Covid. Zhang et al. (2020) studied looked at estimation of the reproductive number and outbreak size of the disease on the Diamond Princess Cruise Ship. Williamson et al. (2020) looked at factors associated with COVID-19 deaths in England. However, as far as we know, to date no statistical analysis has been done to investigate the factors that affect the number of state-wide COVID cases and deaths in USA. We believe that the results would be of social interest and a study like this could perhaps lead to better nationwide planning with more resources directed towards states that are at higher risk. Towards this goal, we investigate multiple regression models using the factors that might influence our variables of interest. For this study, we have considered data at two critical dates in the path of each state's outbreak: (1) The date 5 weeks (35 days) after the 100th confirmed case, and (2) the date 13 weeks (91 days) after the 100th confirmed case. We chose the 5-week date because that is roughly a period of time after lockdown procedures had been instituted and results were being seen. We chose the 13-week period because that is roughly a period after the lifting of lockdown procedures in which effects were being seen.

As of October 30, the US had a per 100,000 population confirmed case rate of about 2,720. A quick comparison with other highly populated countries like China (approximately 6), Pakistan (approximately 153), Indonesia (approximately 149) and India (approximately 590) shows the per capita number of cases is much lower compared to the U.S. Europe, which was hit very early and very hard, was able to control the spread for a time, but has experienced a resurgence of the virus, though levels are still somewhat lower than the US, Spain's case rate is approximately 2,482 per 100,000, while Sweden's is 1,207 per 100,000, the UK's is 1,434 per 100,000, France's is 2,038 per 100,000 and Italy's is 1,108 per 100,000.

The rest of the paper is organized as follows. Section 2 describes the data sources, the data and presents descriptive summaries of the data. In Section 3 we develop regression models to describe the number of cases and deaths on Day 35 as a function of factors like population density, GDP, mobility index etc. In Section 4 we develop similar regression models for the number of cases and deaths on Day 91. Section 5 discusses model assumptions. Section 6 contains some concluding remarks. A table of the data is in the Appendix.

2. Data Description

The initial runs consist of 9 independent variables which we regressed onto 4 response variables. The data are contained in the Appendix 1 and 2. The predictor variables (the "Predictor Variables") are:

- Proportion of the population that is African-American (af_am). It has been widely noted that minority communities have been very hard hit by the virus. We use the proportion of the population that is African-American as a proxy for minority composition.
- Population density (popdens). Epidemiological models predict that the rate of interaction among the population is a major contributing factor to disease spread. We will use population density as a proxy for the rate of interaction.
- Per capita GDP (GDPpercap). We use this as a proxy for the overall wealth of the population.
- Proportion of population with a college degree (coll_deg). Because higher education leads to higher paying jobs and less poverty, this is also a proxy for the wealth of the population.
- Proportion of population that is over 65 (over65). It has been noted that older members of the population are more susceptible to the disease and are more likely to succumb to it.
- Monthly flights into the state before travel bans (flights_to). We use this measure as a proxy for the likelihood that the disease will travel into the state by such flights.
- Party in control of state governor's office. (party_ctrl, 0=GOP, 1=Dem). This variable test whether results have differed depending on the party in control of the governor's office.

- Proportion of the initial 35 days where distancing restrictions in place (prop_dist). This variable is a proxy for the extent of the stay at home restrictions, i.e., the extent of the statewide economic shutdown.
- Change in mobility index from the day of the 100th case (calculated for 35 and 91 days)(dmob). The mobility index is maintained by Descartes Labs. ("Descartes Labs", 2020)

The four response variables we considered in the four different models are the following:

- Confirmed cases per 100,000 population 5 weeks (i.e., on the 35th day) after 100th case (cases_100k35).
- Confirmed cases per 100,000 population 13 weeks (i.e., on the 91st day) after 100th case (cases_100k91).
- Deaths per 100,000 population 5 weeks (i.e., on the 35th day) after 100th case (deaths_100k35).
- Deaths per 100,000 population 13 weeks (i.e., on the 91st day) after 100th case (deaths_100k91).

We can see differing trajectories of the epidemic in the time series plots for various states. Figures 1 and 2 contain time series plots of confirmed cases (Figure 1) and deaths (Figure 2) for each of the six most populous states: California, Texas, New York, Florida, Pennsylvania and Illinois. It is clear that New York was hit very hard very early and had high numbers of cases and deaths, and then got things under control. Pennsylvania had similar, but less severe timing. Illinois has had a double peak in number of cases; the first peak occurred at the same time as New York and Pennsylvania while the second occurred later on, however fatalities were much less pronounced the second time around. Florida, Texas and California had later outbreaks which were more severe in terms of cases, but less so in terms of fatalities.

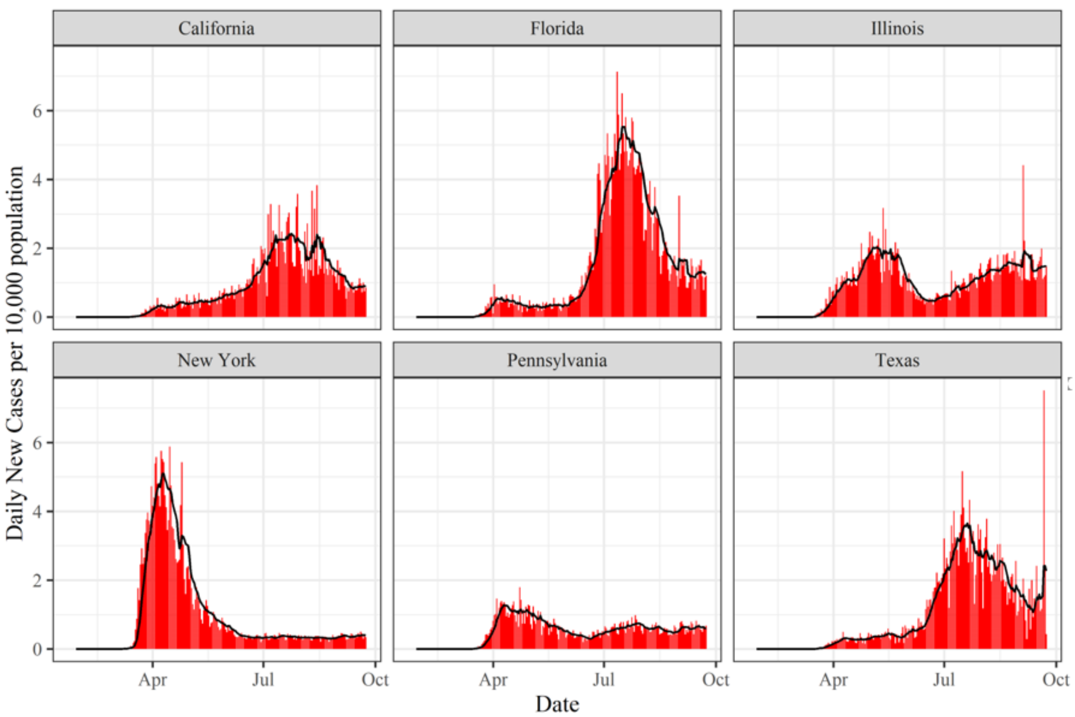

Figure 1. Daily deaths new cases per 100,000 for most the six most populous states

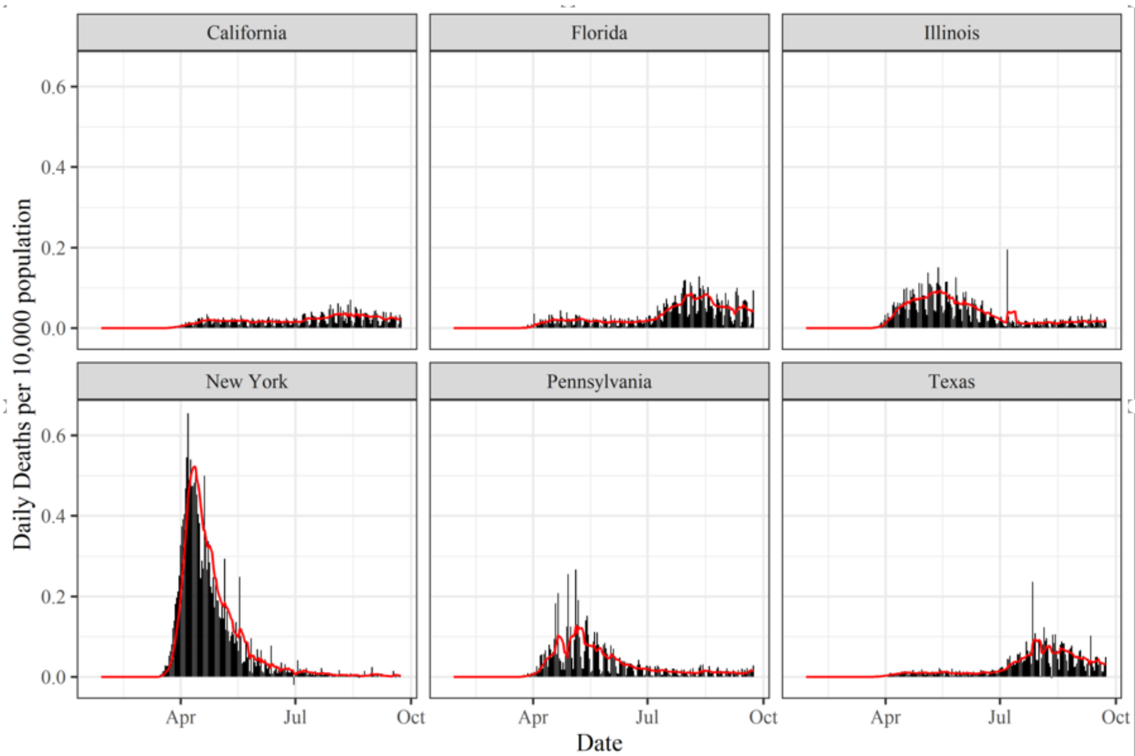

Figure 2. Daily deaths new cases per 100,000 for most the six most populous states

3. Method and Results

We fit a regression models to (1) the number of cases per 100,000 population on the 35th day (5th week) after the 100th confirmed case in the state (2) the number of deaths per 100,000 population on the 35th day after the 100th confirmed case in the state (3) the number of cases per 100,000 population on the 91st day (13th week) after the 100th confirmed case in the state (4) the number of deaths per 100,000 population on the 91st day after the 100th confirmed case in the state. The results are in the following paragraphs.

3.1 Model 1: Dependent Variable: Confirmed Cases as of 35th Day (5th Week)

In the first regression model, we fit a regression model to the number of cases per 100,000 population for the several states 5 weeks (the 35th day) after the 100th confirmed case within each state. The response variable is the number of cases per 100,000 residents on that day. The initial predictor variables are the Predictor Variables enumerated in Section 2.

The initial model run shows that the model is significant with an R-squared of 0.743 and an adjusted R-squared of 0.685, however, the only variable that is significant at $\alpha = 0.05$ (in the presence of all the other independent variables) is population density. The party in control of the governorship and the number of flights into the city are marginally significant (p-value < 0.06) in the presence of the other variables.

We used the backward selection method to choose the best subset of regressors resulting in retention of the following variables in our final model:

- Population Density,
- Per-capita GDP,
- Proportion of the population with a college degree,
- Flights into the state (pre-travel bans),
- Party in control of governorship, and
- Proportion of the initial 35-day period with distancing restrictions in place.

The final model is significant with an R-squared of 0.733 (adjusted R-squared of 0.696). Of the predictor variables, population density (popdens) is highly significant (p-value near 0), and party in control (party_ctrl) of the governorship is significant (p-value=0.0381). All other variables are marginally significant (0.05 < p-value < 0.10). The scatter plot matrix for the variables contained in the final regression model for confirmed cases per 100,000 on Day 35 is shown in Figure 3 while the summary of the model is in Table 1.

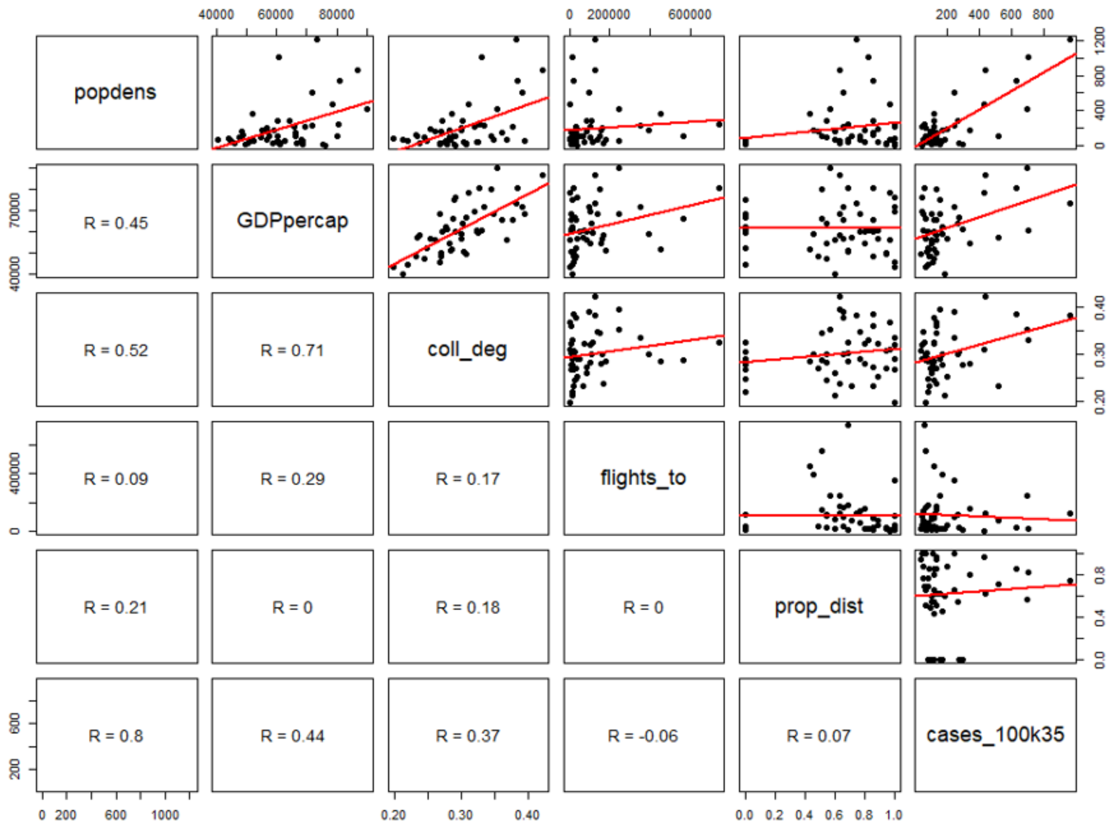

Figure 3. Scatter plot matrix and correlation coefficients of variables in Final Regression Model for confirmed cases per 100,000 on Day 35

Table 1. Final Regression Model summary of confirmed cases per 100,000 on Day 35

(R^2: 0.7332, adj.R^2 : 0.696)

Term	*Coefficients*	*Standard Error*	*t Stat*	*P value*
(Intercept)	**161.1**	**115.35**	**1.397**	**1.397**
popdens	**0.6199**	**0.0704**	**8.801**	**<0.00001**
GDPpercap	**0.0036**	**0.0020**	**1.768**	**0.0841**
coll_deg	**-855.3**	**468.2**	**-1.827**	**0.0747**
flights_to	**-0.00021**	**0.000106**	**-1.985**	**0.0535**
party_ctrl	**80.15**	**37.46**	**2.140**	**0.0381**
prop_dist	**-109.1**	**59.24**	**-1.842**	**0.0724**

It was of special interest to examine the effect of the party in control on the dependent variable. To that end, we decided to look at scatter plots between the dependent variable and some of the independent variables using different markers for the party in control (Red is Republican and blue is Democratic.) These plots are shown in Figures 4 and 5. Note that that the number of cases (as well as the deaths) seem to be higher in Democratic States rather than the Republican States. This may be because the states that were hit hardest at the beginning of the pandemic were in the primarily

Democratic northeast U.S.

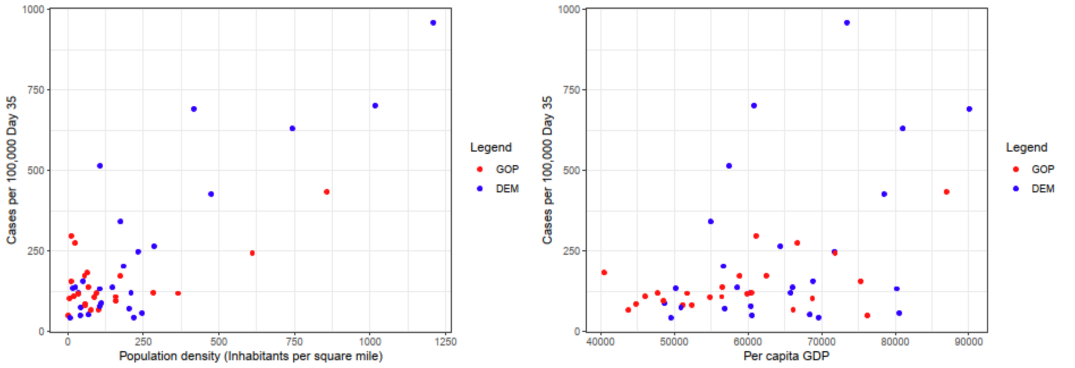

Figure 4. (Left) population density and (Right) per capita GDP vs cases per 100,000 cases, Day 35

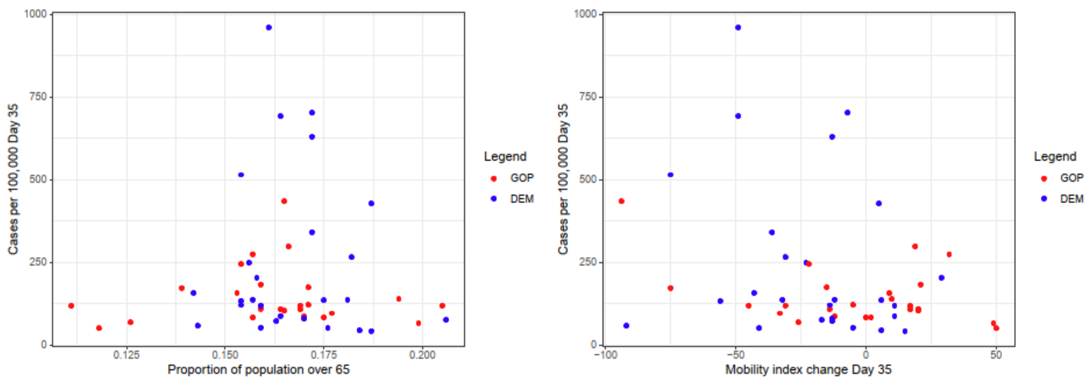

Figure 5. (Left) proportion of population over 65 and (Right) mobility index vs cases per 100,000 cases, Day 35

3.2 Model 2: Dependent Variable: Deaths as of 35th Day (Week 5)

In this model we looked at a regression model fitting the number of deaths per 100,000 population for the states for the time period of 5 weeks (the 35th day) after the 100th case. The dependent variable is the number of deaths per 100,000 residents on that day. As before, the initial predictor variables are the Predictor Variables enumerated in Section 2. The initial regression analysis shows that the model is significant with an R-squared of 0.609 and an adjusted R-squared of 0.521. The analysis shows that population density is extremely significant (in the presence of other variables) and party in control is marginally significant.

Just as in the first model, we used the backward selection procedure to parse the model which results in the retention of the following variables:

- Population Density,
- Per-capita GDP,
- Proportion of the population that is African-American,
- Proportion of the population with a college degree,
- Flights into the state (pre-travel bans), and
- Party in control of governorship.

The final model has an R-squared of 0.601 and adjusted R-squared of 0.545. The significant variables are population density and party in control of the governor’s office. Population density (popdens) is once again highly significant (p-value near 0). Party in control (party_ctrl) of the governorship is significant (p-value < 0.05). No other variables are significant. The scatterplot matrix for the variables in the final model for Deaths as of 35th Day is shown in Figure 6 while the summary of the model is given in Table 2.

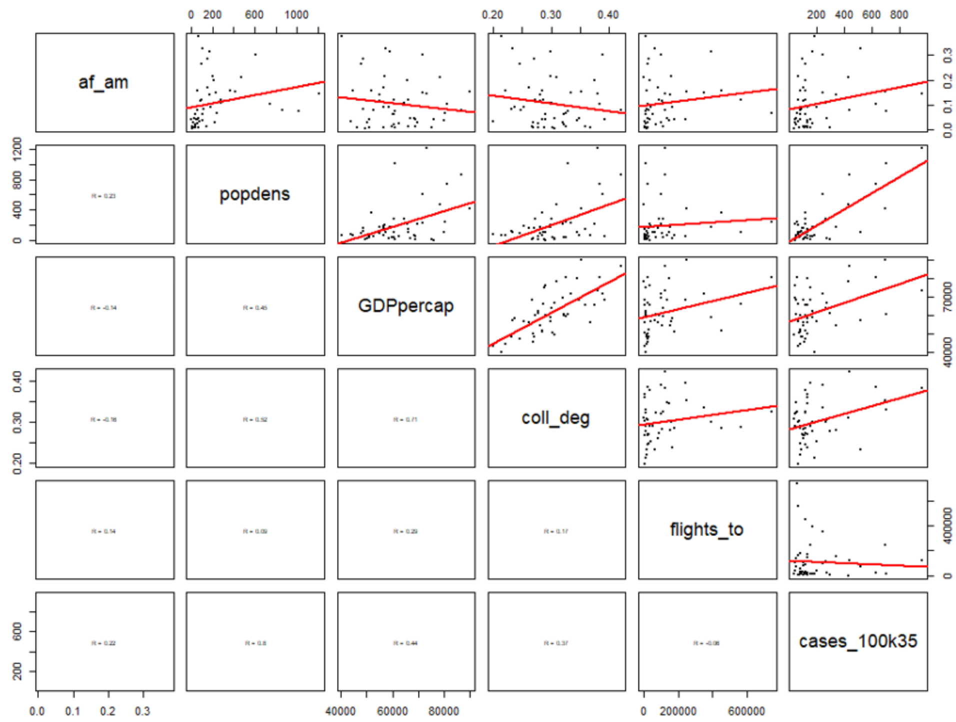

Figure 6. Scatter plot matrix and correlation coefficients of variables in Final Regression Model for Deaths per 100,000 on Day 35

Table 2. Final Regression Model summary of Deaths per 100,000 on Day 35

(R^2: 0.601, adj.R^2: 0.545)

Term	Coefficients	Standard Error	t Stat	P value
(Intercept)	**-3.8029**	**9.627**	**-0.395**	**0.6948**
popdens	**0.03030**	**0.0059**	**5.129**	<0.00001
GDPpercap	**0.00025**	**0.0002**	**1.619**	**0.1127**
coll_deg	**-41.50**	**37.02**	**-1.121**	**0.2685**
flights_to	**-0.000009**	**0.000008**	**-1.079**	**0.2864**
party_ctrl	**5.556**	**2.527**	**2.198**	**0.0333**
af_am	**-16.932**	**14.199**	**1.193**	**0.2396**

3.3 Model 3: Dependent Variable: Confirmed Cases on Day 91 (13th Week)

In this model we considered as the dependent variable the number of confirmed cases as of 13 weeks (or 91 days) after the 100th case. The predictor variables were the same as the ones used for day 35 except that the change in mobility index is computed for the 91st day.

The model is significant with an R-squared of 0.745 and an adjusted R-squared of 0.694. The significant variables (in the presence of all the other independent variables) are population density, the proportion of the first 35 days that were under stay at home orders, and percentage of the population that is African-American.

As in Models 1 and 2, we have used backward selection method resulting in the retention of the following predictors:

- Proportion of the initial 35-day period with distancing restrictions in place.
- Proportion of the population that is African-American,

- Flights into the state (pre-travel bans),
- Per-capita GDP, and
- Party in control of governorship.

The final model is significant with an R-squared of 0.743 and adjusted R-squared of 0.707. Population density (popdens) is once again highly significant (p-value near 0). The proportion of the initial 35 days that were under distancing restrictions (prop_dist) and the proportion of the population that are African American (af_am) are both significant (p-value < 0.05). Per-capita GDP (GDPpercap) and the number of pre-travel ban flights into the jurisdiction were both marginally significant (p-value < 0.10). The scatter plot matrix and the output for the final regression model for Confirmed Cases on Day 91 are given below in Figure 7 and Table 3.

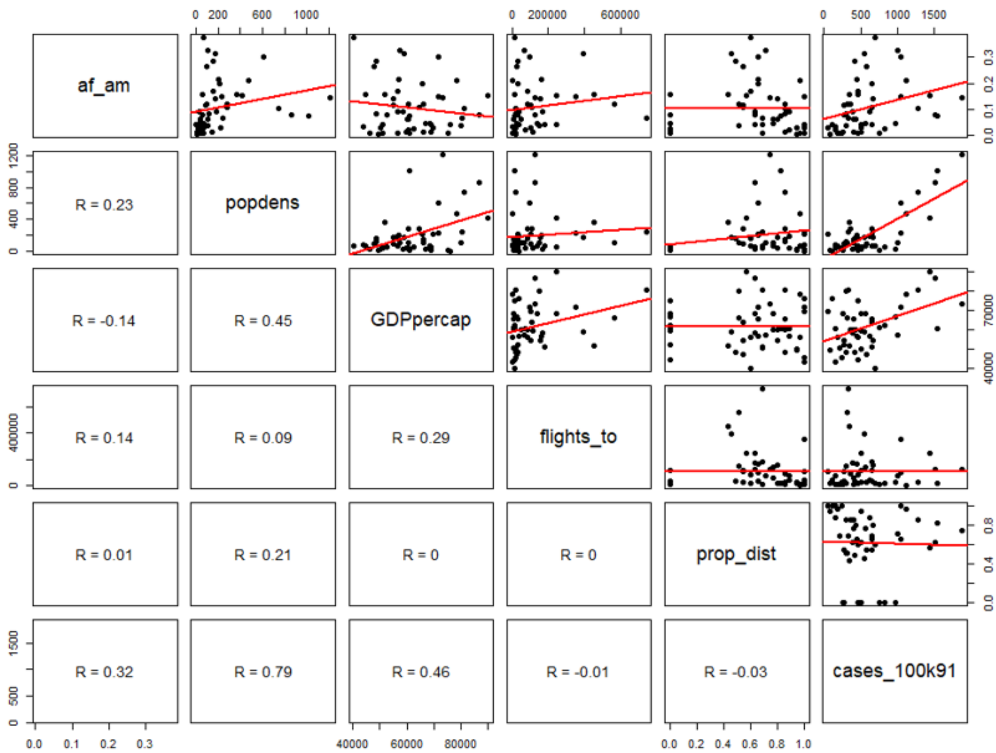

Figure 7. Scatter plot matrix and correlation coefficients of variables in Final Regression Model for confirmed cases per 100,000 on Day 91

Table 3. Final Regression Model summary of Confirmed cases per 100,000 on Day 91

(R^2: 0.7433, adj.R^2 : 0.7074)

Term	*Coefficients*	*Standard Error*	*t Stat*	*P value*
(Intercept)	**43.744**	**228.25**	**0.192**	**0.8489**
popdens	**1.086**	**0.14835**	**7.323**	<0.00001
GDPpercap	**0.0068**	**0.0035**	**1.941**	**0.0588**
flights_to	**-0.00044**	**0.00022**	**-2.003**	**0.0515**
party_ctrl	**110.033**	**77.083**	**1.427**	**0.1607**
af_am	**944.96**	**366.025**	**2.582**	**0.0133**
prop_dist	**-323.286**	**120.727**	**-2.678**	**0.0105**

3.4 Model 4: Dependent Variable: Deaths per 100,000 on Day 91(13th Week)

In this model we considered as the dependent variable the number of deaths as of 13 weeks (or 91 days) after the 100th case. The predictor variables were the same as the ones used for day 35 except that the change in mobility index is computed for the 91st day. The final model, as with the other models, is significant with an R-squared of 0.777 and an adjusted R-squared of 0.727. Here the only significant variable (in the presence of all the other independent variables) is population density. As before, we use backward selection to choose the best regressors resulting in the following independent variables to be retained in the model:

- Population Density,
- Per-capita GDP,
- Proportion of the population that is African-American,
- Flights into the state (pre-travel bans), and
- Party in control of governorship.

The final model is significant with an R-squared of 0.700 and adjusted R-squared of 0.666. Population density (popdens) is extremely significant (p-value near 0). Per-capita GDP (GDPpercapita) and the proportion of the population that are African-American (af_am) are both significant (p-value < 0.05). The number of pre-travel ban flights into the jurisdiction is marginally significant (p-value < 0.10). The scatter plot matrix and the output for the final model for deaths per 100,000 on Day 91 are given below in Figure 8 and Table 4.

Table 4. Final Regression Model summary of Deaths per 100,000 on Day 91

(R^2: 0.7004, adj.R^2 : 0.6664)

Term	*Coefficients*	*Standard Error*	*t Stat*	*P value*
(Intercept)	**-219.303**	**220.005**	**-0.997**	**0.3243**
popdens	**1.0009**	**0.1547**	**6.470**	<0.00001
GDPpercap	**0.00897**	**0.00365**	**2.459**	**0.0179**
flights_to	**-0.00047**	**0.00024**	**-1.960**	**0.0564**
party_ctrl	**3.449**	**70.489**	**0.049**	**0.9612**
af_am	**969.307**	**390.73**	**2.481**	**0.0170**

3.5 Model Assumptions

In checking model assumptions, we found that multicollinearity was not a problem in any of the models. The normality assumption was violated in Models 1 and 2 while Models 3 and 4 did not present any serious violation. The homogeneity of variance assumption was violated in Model 2, but none of the other models posed serious violations. We did not attempt any transformations of the variables because the main purpose of the models was to describe the relationship between the independent variables and the response variables, and transformations would have made interpretation of the results much less straightforward.

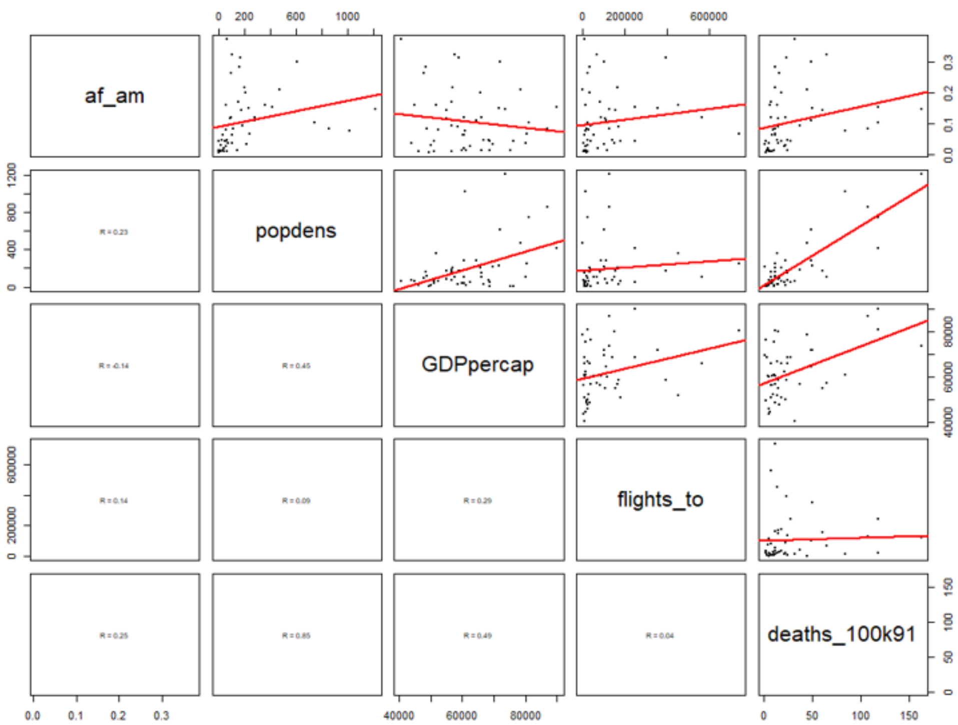

Figure 8. Scatter plot matrix and correlation coefficients of variables in Final Regression Model for Deaths per 100,000 on Day 91

4. Discussion and Conclusion

In this paper, we used multiple linear regression models to identify the significant factors affecting the number of confirmed COVID-19 cases and the number of deaths per 100,000 in the states of the US. We identified population density as the most influential factor in all the models. Interestingly, the model suggests Party control is a significant factor for number of deaths and number of cases at the fifth week, but not at the thirteenth week. The percentage of African American citizens is another interesting factor that was not significant at the fifth week but became significant at the thirteenth week. Per-capita GDP was significant or marginally significant in all the models except the number of deaths at the fifth week. We believe that this paper is an important first step towards a deeper understanding of the factors which influence the number of Covid-19 cases and deaths. Understanding the factors will help us get a better understanding of how to control the virus. However, the trajectory of this disease continues to evolve. In the United States, the first states to feel the wrath of this virus were in the Northeast back in March. Later in the summer we saw the Sun Belt states get hit and as of the time of this analyis it appears to be the Midwest feeling the worst effects of the virus. Another factor to consider is that early in the onset of the virus, the death rate from the disease was extremely high but now seems to have decreased. It is clearly of interest then to see why the waves of infection hit different parts of the country at different times and why the death rate seems to be more stable now. Towards that end, for future research, we would like to see if the addition of a predictor variable that distinguishes between states that were hit by the virus early and those that were hit later and see if that changes the effects of the independent variable. Another factor of interest would be to investigate what if any are the differences in testing protocols and data collection methods in different states and how that affects the models. One important factor in data collection methods is the difference between officially reported deaths due to COVID-19 and excess deaths (the difference between total statewide deaths and the average number of deaths in the past several years). If there are significant differences in this count between states, then that warrants a deeper investigation. Finally, this paper looked at two discrete points in time to model the number of cases and deaths for the different US states. A time series analysis of our two dependent variables would definitely add to the understanding of this new disease.

References

Almukhtar, S., Bloch, M., Aufrichtig, A., Calderone, J., Collins, K., Conlen, M., … Yourish, K. (2020, August 12). Covid in the U.S.: Latest Map and Case Count. *New York Times.* Retrieved from https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html/

Centers for Disease Control and Prevention, Coronavirus (COVID-19). Retrieved from www.cdc.gov/coronavirus/2019-ncov/index.html

Descartes Labs. https://github.com/descarteslabs/DL-COVID-19

Hauck, G., Gelles, K., Bravo, V., & Thorson, M. (2020, June 23). Five months in: A timeline of how COVID-19 has unfolded in the US. *USA Today*. Retrieved from https://www.usatoday.com/in-depth/news/nation/2020/04/21/coronavirus-updates-how-covid-19-unfolded-u-s-timeline/2990956001/

Johns Hopkins Coronavirus Resource Center. Retrieved from www.coronavirus.jhu.edu

Ruan, Q., Yang, K., & Wang, W. *et al.* (2020). Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *Intensive Care Med, 46*, 846–848 (2020). https://doi.org/10.1007/s00134-020-05991-x

Shear, M. D., Goodnough, A., Kaplan, S., Fink, S., Thomas, K., & Weiland, N. (2020, April 1). The Lost Month: How a Failure to Test Blinded the U.S. to Covid-19. *New York Times.* Retrieved from nytimes.com/2020/03/28/us/testing-coronavirus-pandemic.html

The New York Times, The Coronavirus Outbreak. Retrieved from www.nytimes.com/news-event/coronavirus

Williamson, E. J., Walker, A. J., & Bhaskaran, K. *et al.* (2020). Factors associated with COVID-19-related death using Open safely. *Nature, 584*, 430–436. https://doi.org/10.1038/s41586-020-2521-4

Zhang, S., Diao, M., Yu, W., Pei, L., Lin, Z., & Chen, D. (2020). Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *Int J Infect Dis., 93*, 201-204. https://doi.org/10.1016/j.ijid.2020.02.033

Appendix 1. State Statistics

State	af_am	popdens	GDPpercap	coll_deg	over65	flights_to	party_ctrl	prop_dist
Alabama	0.264	95.4	47735	0.245	0.169	23843	0	0.5428571
Alaska	0.043	1.3	76220	0.290	0.118	35897	0	1.0000000
Arizona	0.042	58.3	51179	0.284	0.175	176531	0	0.6857143
Arkansas	0.158	56.9	44808	0.220	0.170	14974	0	0.0000000
California	0.067	246.1	80563	0.326	0.143	740782	1	0.6857143
Colorado	0.043	50.8	68828	0.394	0.142	245548	1	0.6285714
Connecticut	0.103	742.6	81055	0.384	0.172	21501	1	0.8571429
Delaware	0.210	475.1	78468	0.310	0.187	0	1	0.9714286
Florida	0.159	364.6	51745	0.285	0.205	452180	0	0.4285714
Georgia	0.314	173.7	58896	0.299	0.139	391889	0	0.4571429
Hawaii	0.031	218.6	69593	0.320	0.184	102602	1	1.0000000
Idaho	0.010	19.5	46043	0.268	0.159	21500	0	1.0000000
Illinois	0.149	232.0	71727	0.334	0.156	352071	1	1.0000000
Indiana	0.091	183.4	56702	0.253	0.158	39282	1	0.8857143
Iowa	0.027	55.3	62493	0.277	0.171	12817	0	0.0000000
Kansas	0.062	35.4	60310	0.323	0.159	9703	1	0.8000000
Kentucky	0.082	111.3	48697	0.232	0.164	16178	1	0.8571429
Louisiana	0.324	107.1	57445	0.234	0.154	67165	1	0.7142857
Maine	0.010	43.1	50915	0.303	0.206	7299	1	0.6857143
Maryland	0.301	610.8	71838	0.390	0.154	99031	0	0.6571429
Massachusetts	0.081	858.0	86942	0.421	0.165	125110	0	0.6285714
Michigan	0.142	175.0	54928	0.281	0.172	155896	1	0.8000000
Minnesota	0.046	68.1	68427	0.348	0.159	138461	1	0.7714286
Mississippi	0.373	63.7	40464	0.213	0.159	10068	0	0.6000000
Missouri	0.115	87.9	54879	0.282	0.169	104404	0	0.5428571
Montana	0.007	7.0	49540	0.307	0.187	17004	1	0.9428571
Nebraska	0.045	24.3	66737	0.306	0.157	22970	0	0.0000000
Nevada	0.090	25.4	58570	0.237	0.157	167198	1	0.6285714
New Hampshire	0.012	147.8	66069	0.360	0.181	6364	1	0.8571429
New Jersey	0.145	1210.1	73451	0.381	0.161	122974	1	0.7428571
New Mexico	0.030	17.2	50201	0.269	0.175	21399	1	0.9428571
New York	0.152	417.0	90043	0.353	0.164	246409	1	0.5714286
North Carolina	0.216	202.6	56862	0.299	0.163	162329	1	0.6571429
North Dakota	0.011	10.5	75321	0.289	0.153	10726	0	0.0000000
Ohio	0.120	283.2	60464	0.272	0.171	85730	0	0.8571429
Oklahoma	0.080	56.1	52409	0.248	0.157	30020	0	0.0000000
Oregon	0.020	40.9	60558	0.323	0.176	71099	1	0.8857143
Pennsylvania	0.108	285.5	64412	0.301	0.182	110556	1	0.5428571
Rhode Island	0.075	1017.1	60830	0.330	0.172	13151	1	0.8285714
South Carolina	0.285	158.8	48547	0.270	0.177	30263	0	0.4857143
South Dakota	0.011	11.1	61104	0.278	0.166	8052	0	0.0000000
Tennessee	0.168	157.5	56451	0.261	0.164	81729	0	0.6000000
Texas	0.119	101.2	66149	0.287	0.126	565292	0	0.5142857
Utah	0.013	35.3	59892	0.325	0.111	114239	0	0.0000000
Vermont	0.009	68.0	56525	0.368	0.194	3708	0	0.9714286
Virginia	0.199	209.2	65824	0.376	0.154	32205	1	0.6571429
Washington	0.037	104.9	80170	0.345	0.154	148410	1	0.5142857
West Virginia	0.036	77.1	43806	0.199	0.199	2387	0	1.0000000
Wisconsin	0.061	106.0	60425	0.290	0.170	51257	1	0.7714286
Wyoming	0.013	6.0	68757	0.267	0.165	7917	0	0.0000000

Appendix 2: COVID-19 Related State Statistics

State	cases_100k35	deaths_100k35	dmob35	cases_100k91	deaths_100k91	dmob91
Alabama	118.94310	4.119771	-31	575.25873	16.519874	23
Alaska	49.89440	1.230273	50	116.32914	1.913758	86
Arizona	83.05035	3.654490	0	644.48171	18.148803	40
Arkansas	86.12223	1.491151	-12	461.52765	6.892429	22
California	57.70366	1.619752	-92	330.67489	11.722955	-69
Colorado	157.10045	6.754955	-43	500.23130	27.471306	12
Connecticut	630.21574	43.306472	-13	1274.20317	118.335494	35
Delaware	427.41362	12.836786	5	1111.15219	44.672015	48
Florida	118.69034	3.482676	-45	342.45694	13.618753	16
Georgia	172.36763	6.470497	-75	543.26742	23.084698	-26
Hawaii	43.64801	1.130046	6	60.10430	1.200673	33
Idaho	109.22938	3.357460	20	246.32568	4.980233	68
Illinois	248.68565	10.645668	-23	1049.69917	49.921791	6
Indiana	203.21680	11.006772	29	620.09272	37.372522	81
Iowa	173.56192	3.740012	-15	822.89775	21.742782	41
Kansas	119.21132	4.324972	11	426.59322	8.855894	42
Kentucky	87.62951	4.588519	11	305.08052	11.728701	70
Louisiana	514.71414	27.878198	-75	1002.81923	64.834019	-26
Maine	75.50892	3.719651	-17	219.98018	7.588089	60
Maryland	244.38938	11.545434	-22	1041.55364	49.556047	12
Massachusetts	434.08033	16.075437	-94	1511.14914	108.146489	-48
Michigan	340.92808	27.496138	-36	663.56212	60.419409	32
Minnesota	52.16652	3.546331	-5	561.65012	24.398755	39
Mississippi	182.58494	7.022498	21	693.54727	31.517239	55
Missouri	107.26837	5.387564	17	272.50210	14.683754	65
Montana	42.38485	1.497037	15	75.13253	1.964861	80
Nebraska	274.86445	3.773764	32	970.58118	13.802672	81
Nevada	136.61646	6.330848	-32	395.07739	15.421297	23
New Hampshire	137.08795	4.412702	-12	407.73370	24.931769	42
New Jersey	960.36000	61.977958	-49	1878.82718	163.416905	-32
New Mexico	134.72725	4.721415	6	510.00821	22.367108	38
New York	691.95119	48.838622	-49	1442.24727	118.394574	-21
North Carolina	71.46205	2.526677	-13	447.49832	11.260398	10
North Dakota	156.28650	3.280573	9	458.62410	10.366611	54
Ohio	120.77063	5.218537	-5	362.91930	22.337049	48
Oklahoma	82.91696	4.978556	2	271.26810	9.325315	24
Oregon	50.42989	1.967880	-41	150.93402	4.433657	18
Pennsylvania	264.91196	10.529614	-31	652.93760	48.765860	15
Rhode Island	702.21577	21.333615	-7	1542.15607	84.390496	45
South Carolina	95.49958	2.913349	-33	418.22094	12.061264	10
South Dakota	297.40273	2.373796	19	755.20624	10.286449	85
Tennessee	108.27078	2.298960	-14	464.93178	7.219028	19
Texas	68.11657	1.941655	-26	311.11660	7.459680	-4
Utah	117.96786	1.216485	17	512.32736	4.834748	64
Vermont	138.14346	7.532184	10	186.54175	8.974517	69
Virginia	120.27388	4.088796	-14	653.44591	18.546031	11
Washington	132.88434	6.342834	-56	301.94778	15.088853	-12
West Virginia	66.12181	2.678352	49	155.23280	5.189306	78
Wisconsin	79.43413	4.190687	-13	398.42444	12.073988	43
Wyoming	102.97896	1.209484	20	250.53606	3.455670	74