Predictors of Death Rate During the COVID-19 Pandemic

Ian Feinhandler [1], Benjamin Cilento [2], Brad Beauvais[3], Jordan Harrop[4]+, Lawrence Fulton [5,]*

[1]   Department of Geography and Political Sciences, Front Range Community College, Longmont, CO, ian.feinhandler@frontrange.edu
[2]   Memorial Hermann Hospital, Houston, TX, ben.cilento@me.com
[3]   School of Health Administration, Texas State University, San Marcos, TX, bmb230@txstate.edu
[4]   Acushnet Corporation, Boston, MA, jordan_harrop@acushnet.golf
[5]   School of Health Administration, Texas State University, San Marcos, TX, lf25@txstate.edu
*   Correspondence: lf25@txstate.edu

Abstract: COVID-19 is a potentially fatal viral infection. This study investigates geography, demography, socioeconomics, health conditions, hospital characteristics, and politics as potential explanatory variables for death rates at the state and county levels. Data from the Centers for Disease Control and Prevention, the Census Bureau, Centers for Medicare and Medicaid, Definitive Healthcare, and USAfacts.org were used to evaluate regression models. Yearly pneumonia and flu death rates (state level, 2014-2018) were evaluated as a function of the governors' political party using repeated measures analysis. At the state and county level, spatial regression models were evaluated. At the county level, we discovered a statistically significant model that included geography, population density, racial and ethnic status, three health status variables along with a political factor. State level analysis identified health status, minority status, and the interaction between governors' parties and health status as important variables.   The political factor, however, did not appear in a subsequent analysis of 2014-2018 pneumonia and flu death rates. The pathogenesis of COVID-19 has greater and disproportionate effect within racial and ethnic minority groups, and the political influence on the reporting of COVID-19 mortality was statistically relevant at the county level and as an interaction term only at the state level.

Keywords: COVID-19, Geospatial Regression, Health Disparities, Public Health

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiologic agent of the coronavirus (COVID-19) pandemic. As of August 31, 2020, the associated death toll in the United States is reported to have surpassed 180,000 [1], the highest of any country in raw numbers but equivalent to many other developed countries when adjusted for population [2]. The proper recognition and remediation of the disease are pressing concerns and each will likely be subject to debate in the months prior to the 2020 presidential election [3,4]. However, there is some concern surrounding the veracity of the data and factors contributing to COVID-19 deaths. Media outlets provide daily updates on the number of cases and deaths but draw this information from data collection and reporting agencies that have adjusted their methods over time [5]. The resulting inconsistencies have led to charges of under-reporting [6,7] and over-reporting [8,9], and have contributed to the politicization of the pandemic.

COVID-19 data inconsistencies and potential political bias in data reporting can have significant implications. If the data that politicians rely on is faulty, subsequent policies may harm public health,

the economy, and other aspects of society. Testing differences, false positives, false negatives, and other factors that likely differ from state-to-state and county-to-county make the underlying official deaths with COVID-19 reports somewhat suspect; however, this study leverages the official data used by the Centers for Disease Control and Prevention (CDC). There are several county-level studies about COVID-19 available from recent research. Badr et al. (2020) evaluated mobility patterns and COVID-19 transmission [10]. This study provided county-level spread data but did not focus on deaths. Scannell et al. (2020) demonstrated racial disparities at the county level for COVID-19 cases and deaths [11]. Cases, unfortunately, suffer from severe measurement problems, as will be discussed. Ives and Bozzuto (2020) analyzed county-level estimates of R0, the basic reproduction number for COVID-19 [12]. Altiera et al. (2020) estimated county-level deaths used in estimating required medical supplies [13]. Two articles consider political factors; Flanders et al. (2020) assessed voter turnout as related to COVID-19 [14], and Makridis and Rothwell (2020) evaluated the effects of political polarization but not in terms of death rates [15]. We found no other paper that addresses death rate disparities by including a political variable. Thus, given the novel nature of the virus and its progression and the known inconsistencies in the reported data, we sought to gain a deeper understanding of the factors that contribute to COVID-19 deaths reported.

1.1. Research Questions

We investigated three research questions. First, what attributes of geography, demography, population density, economy, population health, hospital characteristics, and politics might explain deaths per 100,000 (death rate) at the county-level as of August 31, 2020? Second, did COVID-19 death rates at the state level differ based upon governor party affiliation after accounting for other relevant variables? As a control for our second line of inquiry, we also examined whether variation existed in previous flu/pneumonia death rates (2014-2018) based upon the governor's party affiliation.

1.2. Significance and Motivation

To our knowledge, this research is the first to evaluate COVID-19 using combined data from multiple areas covering demographic, socio-economic, health system, population health, and political factors using spatial regression. It is also the first study to evaluate the effects of state and county political affiliation on COVID-19 death rates. The motivation behind this study is to address the media promulgation of explanatory factors that may or may not be scientifically verifiable (e.g., population density and political factors), particularly when placed in the context of other known factors established at the individual unit of analysis (e.g., race).

2. Methods

2.1. Sample Sizes and Data Sets

Sample sizes for the research questions were 3,116 (county), 51 (states plus Washington D.C.), and 250 (50 states by 5 years). The dependent variable was the death rate per 100,000 population. Cumulative COVID-19 deaths were obtained from USAfacts.org [1] for August 31, 2020. Flu data were from the Centers for Disease Control and Prevention, CDC, from 2014-2018 [16]. Definitive Healthcare data provided descriptive hospital-related information [17]. Population and demographic data were from the Census Bureau [18]. The Centers for Medicare and Medicaid Services (CMS)

provided the source for relevant patient morbidity proportions by state and county [19]. Geographic variables in the analysis included the shapefiles from the Census Bureau's state and county Tiger Files [20].

2.2. Variables

Race and ethnicity variables included the proportion of African American, Native American, Asian, and Hispanic. (The proportion of Caucasians was omitted due to collinearity considerations.) Population density (population per square kilometer), and the proportion 65 and older served as additional control variables, although we anticipated (correctly) that the former might not enter the model, particularly when geospatial effects are considered. Economic variables included median household income and unemployment. Population health status variables included the population proportions with chronic obstructive pulmonary disorder (COPD), heart failure, diabetes, obesity, and cancer; all of which have been identified as risk factors at the individual level [21], as well as other health-related variables including smoking, obesity, alcohol abuse, Alzheimer's Disease, asthma, atrial fibrillation, depression, drug abuse, HIV, hepatitis B, and stroke. Health system capability variables included the number of acute beds in the county or state and the average case-mix index in the county or state.   Case-mix index or CMI adjusts inpatients based on severity, with 1.0 being the "typical" visit and higher average numbers meaning more acute visits than would be expected.

2.3. Reasons for Variable Inclusion and Expected Effects

Geography is included as a known predictor of COVID-19 [22]. Similarly, demographics [23,24], population density [25], economic considerations [26], population health status (comorbidities) [27], and political considerations [28] are also known or hypothetical factors that affect infection and death rates, although the reasons for associations between individual variables and death rates are not fully understood [24].   We include hospital system characteristics to account for the possibility that lack of resources increase death rates [29].

Based on these research studies, we surmise that higher population densities might initially be associated with higher death rates, but that the effects of including spatial models will remove those effects. Increases in population density may place individuals at increased risk of exposure. Better economic status (e.g., lower poverty rates) should result in better access to healthcare systems and thus lower death rates.   (Poverty, for example, results in reduced compliance with COVID-19 protocols [30].) Higher rates of comorbidities (e.g., health status) are likely to be associated with higher death rates [31].   More hospital capability and lower patient severity might reduce death rates [29].   Finally, there is much speculation that political considerations are influencing both death rates and reporting of death rates, where Democratically affiliated geographies are anticipated to have higher death rates [32].

2.4. Transformations

Quantitative variables were standardized. At the state-level of analysis, the small number of observations (51) necessitated data reduction.   We used the first three principal components of all health status variables to proxy the effects of population health. These three components accounted for 75% of the variability of the original 19 variables.

2.5. Models

We evaluated least absolute shrinkage and selection (lasso) models [33] to generate a subset of variables associated with deaths per 100,000 using adaptive p-values as presented by Lockhart et al.

[34] and implemented in the *covTest* package [35] in R [36]. The adaptive p-values address Lindley's paradox, which often requires that significance level changes as sample size increases [37]. We also use 10-fold cross-validation to evaluate R^2 and root mean squared error (RMSE) along with associated standard deviations (SDs). Table 2 is a list of independent variables evaluated.

After fitting the OLS and constrained models, we repeat the same process to fit GIS models. Specifically, we use residual analysis to fit appropriate geospatial models with all of the variables and the subset suggested by lasso. Moran's I and Lagrangian multiplier diagnostics were used to recommend the appropriate geospatial model to be fitted (none, spatial lag, or spatial error).

We also investigated reporting differences that might exist for flu and pneumonia deaths at the state level. Using repeated measures analysis, we modeled the logarithm of flu and pneumonia deaths as a function of year and governor party. All analyses were performed in R Statistical Software [36].

3. Results

All code is available for replication. County-level R code (updated through August 31) is available here: https://rpubs.com/R-Minator/COVIDCounty. State level code (also updated through August 31) and influenza analyses are available here: https://rpubs.com/R-Minator/COVIDState.

3.1. Descriptive Statistics

Table 1 summarizes the descriptive statistics at the county level of analysis. At the county level (as of August 31, 2020), the mean COVID-19 death rate is 33.84, respectively. The mean county population was 9% African American, 2% Native American, 9% Hispanic, 1% Asian, and 20% aged 65 and over. Population density, income, and unemployment averages were 106.45 per square kilometer; $53,000 per county-person; and 4% per county, respectively. The largest comorbidity proportion average was adult obesity (32.85%), and the mean number of acute beds was 215 with a median of 35. Average CMI was 1.06 with a median of 1.17. Sixteen percent of counties voted for the Democratic candidate in 2016.

Table 1. County level descriptive statistics

Variable (n = 3116 counties)	Mean	SD	Median	Minimum	Maximum
Population in 2020	105237	334733.38	26163.00	169	10039107
Population Density (# / km²)	106.45	696.94	17.50	0	27755
Native American %	1.57%	6.48%	0.30%	0.00%	89.60%
Hispanic %	9.30%	13.84%	4.10%	0.00%	99.10%
African American %	8.99%	14.51%	2.20%	0.00%	87.40%
Asian %	1.31%	2.59%	0.60%	0.00%	43.08%
% 65 or older	19.79%	4.76%	19.40%	4.90%	58.20%
Unemployment % (2019)	3.96%	1.39%	3.70%	0.70%	18.30%
Household Income $ (2018) +	$ 52,714.43	$ 13,851.63	$ 50,531.00	$ 25,385.00	$140,382.00
Poverty %	15.17%	6.11%	14.10%	2.60%	54.00%
Smoke %	17.44%	3.56%	16.95%	5.91%	41.49%
Adult Obesity %	32.85%	5.43%	33.10%	12.40%	57.70%
Alcohol Abuse %	2.24%	1.01%	2.21%	0.00%	10.36%
Alzheimer's %	10.17%	2.18%	10.15%	0.00%	25.02%
Asthma %	4.31%	1.34%	4.35%	0.00%	11.64%
Atrial Fibrillation %	8.03%	1.61%	8.12%	0.00%	17.50%
Cancer %	7.41%	1.40%	7.43%	0.00%	12.10%
Kidney % *	22.85%	4.51%	22.94%	0.00%	51.45%
COPD %	12.81%	3.77%	12.44%	0.00%	32.15%

Depression %	17.44%	3.57%	17.48%	0.00%	35.87%
Diabetes %	26.93%	5.09%	27.11%	0.00%	49.62%
Drug Abuse %	3.14%	1.83%	2.93%	0.00%	16.70%
HIV %	0.11%	0.25%	0.00%	0.00%	4.51%
Heart Failure %	14.39%	3.28%	14.15%	0.00%	33.75%
Hepatitis B %	0.47%	0.42%	0.49%	0.00%	4.10%
Hyperlipidemia % **	38.04%	8.80%	39.35%	0.00%	67.55%
Hypertension %	56.51%	8.77%	58.30%	0.00%	74.95%
Ischemia % ***	26.84%	5.44%	26.68%	0.00%	46.91%
Stroke %	3.32%	1.09%	3.35%	0.00%	9.46%
# Acute Beds	215	720.47	35	0	19274
Case Mix Index	1.061	0.587	1.170	0.000	2.710
2016 Winning Party (1 = Democrat)	0.158	0.364	0.000	0.000	1.000
Deaths / 100K	34.030	46.753	17.753	0.000	461.156

[+] collinear with poverty, r = -0.771, *collinear with diabetes, r=0.78, **collinear with hypertension, r=0.80, ***
collinear with heart failure and hypertension, r=0.71 for both

Figure 1 is a notched boxplot of the death rate of Democratic counties versus Republican counties.   The notch indicates statistical significance (median test) at the α=.05 level. There appears to be a statistically significant difference between the two group's death rates per 100,000 people.
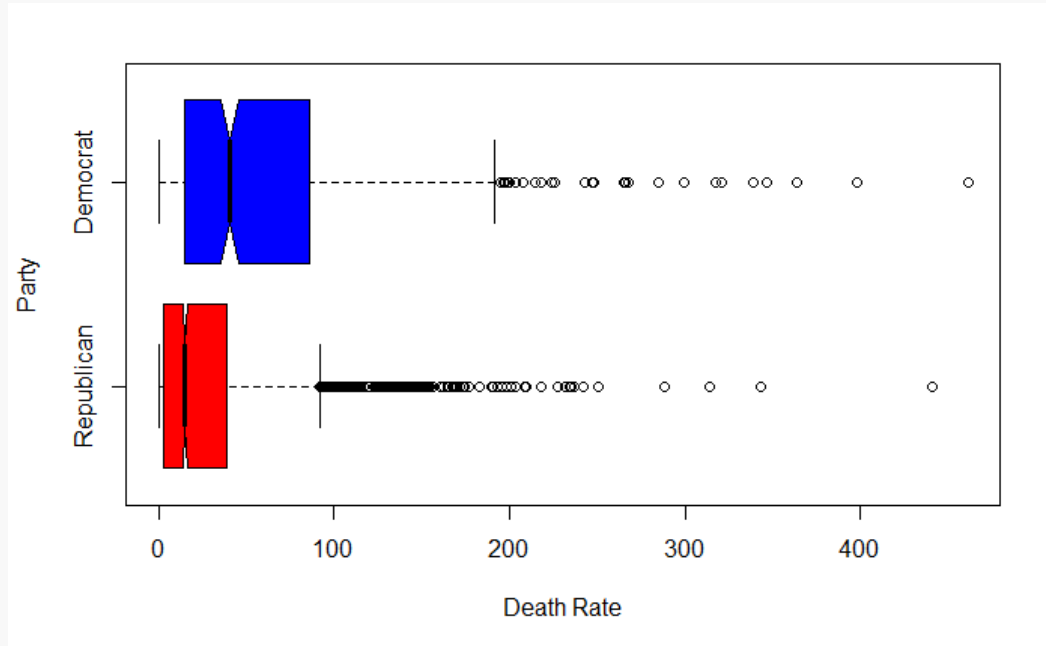
Figure 1.   Boxplots of Republican versus Democratic county death rates per 100,000.

Figure 2 provides a scatterplot of the proportion voting Democrat in a state versus the deaths per 100,000 with symbols showing which states voted for Clinton versus Trump. Seven states have at least 75 deaths per 100,000. Of those states, 6 voted for Clinton. The red and blue dots indicate the current party of the state governor.
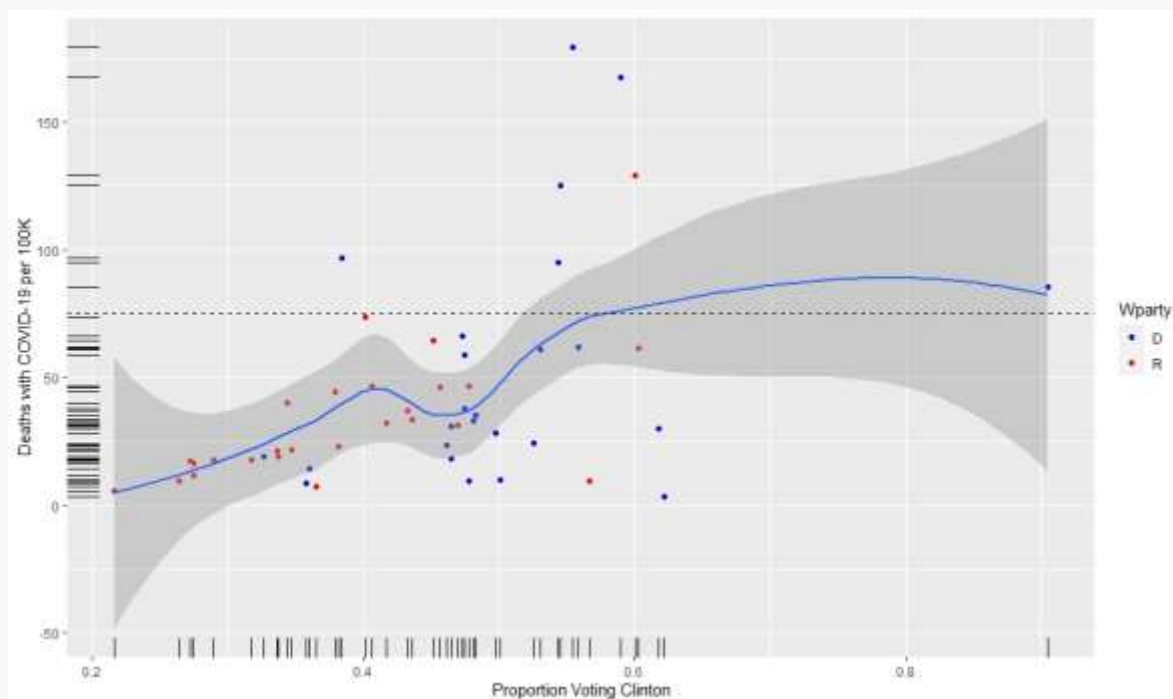
Figure 2. COVID-19 death rates per 100,000 (y axis) as a function of proportion voting for Clinton in 2016 (x axis) and the current party of the governor as a red or blue dot.

Table 2 presents a county-level summary of the association between 2016 presidential election results, population density, and deaths from COVID-19. Population density is higher for counties that voted Democratic (116.2 versus 23.5), as are death rates (71.0 versus 36.8).

Table 2. Population density and COVID-19 deaths by 2016 electoral outcome (August 31, 2020)

Candidate	Counties Won	Avg. Density	Deaths	Death Rate
Clinton	491	116.2	126,554	71.0
Trump	2,625	23.5	55,157	36.8
Total	3,116	41.5	181,711	55.4

At the state level (Table 3), descriptive statistics are provided for variables considered for the final model.   The deaths per 100,000 for COVID-19 were 45.74 versus flu deaths of 15.10 per 100K. The proportions of African American, Native American, Hispanic, and 65 years of age (and older) were 11.27%, 1.62%, 12.01%, and 16.39%, respectively.   Unemployment in 2019 averaged 3.62%, and about 49% of the states had Democratic governors.

Table 3. State level descriptive statistics

Variables (n = 51)	Mean	SD	Median	Minimum	Maximum
% African American	11.27%	10.72%	7.50%	0.40%	46.90%
% Native American	1.62%	2.87%	0.50%	0.20%	14.40%
% Hispanic	12.01%	10.31%	9.52%	1.43%	49.09%
% 65+	16.39%	1.99%	16.40%	11.10%	20.60%
% Unemployment	3.62%	0.82%	3.50%	2.40%	6.10%

% Democratic Governor	49.02%	50.49%	0.00%	0.00%	100.00%
COVID-19 Deaths / 100K	45.74	39.58	32.95	5.01	179.53
Flu Deaths / 100K	15.10	3.76	14.65	7.00	29.60

3.2. COVID-19 Death Analysis, County

The four models estimated for the county analysis are depicted in Table 4. Column 1 shows the estimates for the full OLS model. The lasso model is column 2. Geospatial models (full and reduced based on residual analysis) are columns 3 and 4.

Table 4. Model results (scaled variables)

Variable	OLS Full	p	Lasso	p*	GIS Full	p	GIS Reduced	p
R^2 (Predicted R^2 for Lasso)	*0.377*		*0.352 +/- 0.800*		*0.507*		*0.500*	
Rho					0.634	<0.001	0.589	<.001
Intercept	0.000	0.014	0.000	NA	-0.004	0.732	-0.004	<.001
Pop. Density	0.163	0.017	0.138	0.038	0.066	<0.001	0.051	<.001
% Native American	0.090	0.018	0.057	0.038	0.070	<0.001	0.059	<.001
% Hispanic	0.133	0.022	0.132	<.001	0.082	<0.001	0.071	<.001
% Black	0.408	0.029	0.369	<.001	0.178	<0.001	0.169	<.001
% Asian	0.008	0.019			-0.009	0.581		
% 65+	0.022	0.019			0.022	0.182		
% Unemployed	0.079	0.018	0.075	0.007	0.052	0.001	0.062	<.001
Poverty	0.018	0.027			0.012	0.621		
% Smoke	-0.061	0.026			-0.006	0.815		
% Adult Obesity	-0.045	0.019			0.006	0.721		
% Alcohol	0.041	0.020			0.024	0.170		
% Alzheimers	0.112	0.021	0.149	<.001	0.073	<0.001	0.097	<.001
% Asthma	-0.049	0.020			-0.022	0.217		
% Atrial Fib.	0.017	0.021			0.011	0.563		
% Cancer	-0.010	0.020			-0.016	0.379		
% COPD	-0.074	0.027	-0.104	<.001	-0.047	0.048	-0.053	0.006
% Depression	0.036	0.023			0.043	0.034		
% Diabetes	0.183	0.027	0.162	<.001	0.078	0.001	0.079	<.001
% Drug Abuse	-0.027	0.022			-0.033	0.096		
% HIV	-0.074	0.021			-0.047	0.011		
% Heart Failure	-0.027	0.021			-0.009	0.636		
% Hepatitis B	-0.048	0.021			-0.031	0.095		
% Stroke	0.092	0.022			0.026	0.182		
# Acute Beds	-0.006	0.018			0.009	0.565		
Case Mix Index	0.038	0.017			0.045	0.004		
Winning Party	0.029	0.019	0.024	0.089	0.046	0.007	0.032	0.033

*Adaptive

3.2.1. Ordinary Least Squares (OLS) Full Model

The full OLS model ("OLS Full") is depicted in the first columns of Table 4. The highest variance inflation factor (VIF) was 3.706 (poverty). The model accounted for 37.9% of the variability (R^2). No statistically significant effect for the county's winning party was apparent in the first model evaluation (p=0.242). Figure 3 shows the map of the residuals for the full OLS model, indicating that some spatial autocorrelation exists in the northeast and the southwest areas of the country. Moran's I analysis suggested geospatial correlation as well (I = 0.253, p < .001).
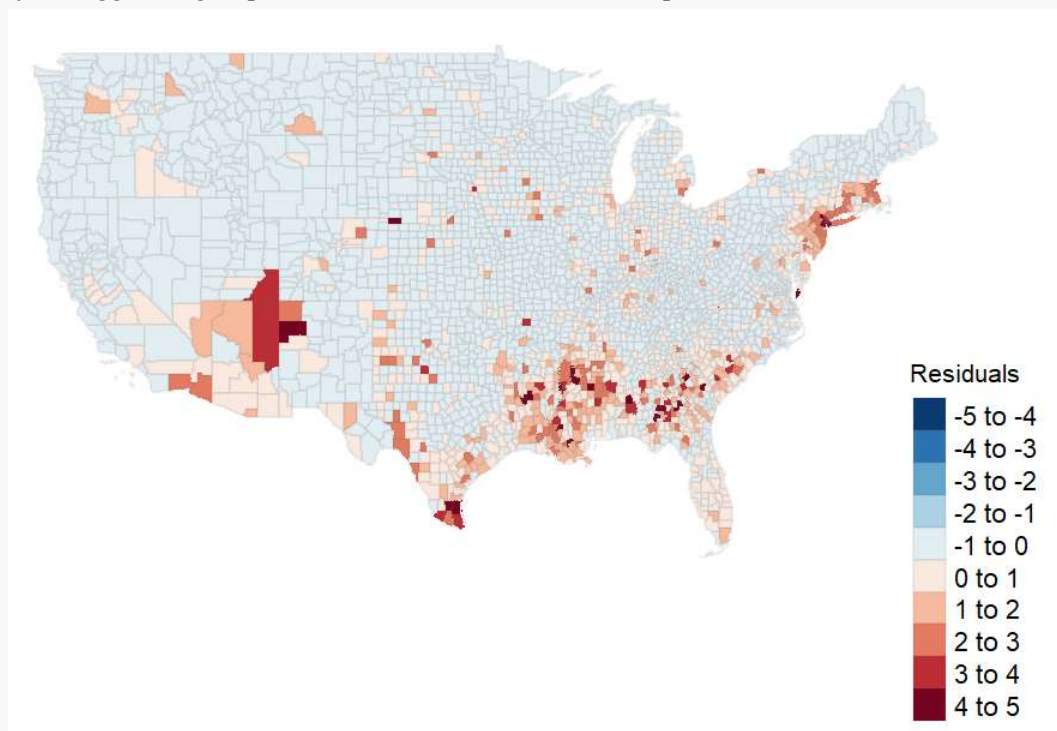
Figure 3. Residual plot from the OLS model shows clusters in the Northeast and Southwest

3.2.2. Lasso Model

The best-tuned lasso model RMSE was 0.800 with a standard deviation (SD) of 0.045. The predicted R^2 was 0.352 with a standard deviation of 0.028. The lasso model ("Lasso", Table 4) using adaptive p-values identified likely predictors as race, ethnicity, and three health status variables (Alzheimer's Disease, COPD, and diabetes). The model produced a similar R^2 as the unconstrained model (R^2 = 0.374). This constrained regression model also suggested that the political factor (winning party) should be considered as a potential explanatory variable (p=0.089). Residual patterns were similar to Figure 2, and Moran's I was statistically significant, indicative of spatial correlation (I = 0.265, p < .001). Lagrangian multiplier diagnostics again recommended a lag model.

3.2.3. Generalized Spatial Two-Stage Least Squares Model, All Variables

A generalized spatial two-stage least squares model (GS2SLS) [38] was used on the full set of independent variables. This model ("GIS Full", Table 4) identified that geospatial location was important for explaining the death rate (ρ = 0.634). Variables in the model again included the political factor (winning party). The residuals from the geospatial model no longer exhibited autocorrelation (Moran's I = -0.098, p = 0.980).

3.2.4.   Generalized Spatial Two-Stage Least Squares Model, Lasso Variables

A final reduced model included the variables identified by the lasso as part of a geospatial lag model. This final model (Table 4, "GIS Reduced") also included the political factor, and again, the residuals were stable based on Monte Carlo simulation of Moran's I (I = -0.070, p = 0.980). For interpretability, the unscaled geospatial model is shown in Table 5.

Table 5.   Unscaled geospatial model

Variable	Estimate	p
Rho	0.598	<.001
(Intercept)	-35.350	<.001
Population Density	0.003	0.001
% Native American	42.728	<.001
% Hispanic	23.226	<.001
% African American / Black	52.703	<.001
Unemployment Rate	2.112	<.001
Alzheimer's Disease	2.077	<.001
Chronic Obstructive Pulmonary Disease (COPD)	-0.664	0.005
Diabetes	0.716	<.001
Winning Party, 2016 Election (1 = Democrat)	4.503	0.021

In Table 5, the reduced geospatial analysis with unscaled variables suggests that geospatial effects, population density, ethnicity and race, unemployment, three health status variables, and winning party are important in explaining death rates per 100,000. Native American, Hispanic, and or African American proportions are associated with a 42.728, 23.226, and 52.703 increase in deaths per 100,000 individuals, respectively. County political leaning based on the 2016 presidential election is associated with an increase of 4.503 deaths per 100,000 individuals (dichotomously-coded variable). Moran's I was not significant (I = -0.070, p=0.9804).

An important result is that while we evaluated population density, its standardized effect size is almost zero (0.003) when other factors are considered. This county-level analysis is congruent with Pew Research findings that death rates are higher in Democratic-led counties [32]. This study suggests that racial/ethnic composition and geographic relationships with the outbreak are important considerations along with political considerations. Further, we note that the results of the spatial analysis are similar to those of the non-spatial analysis.   The implication may be that our county-level models are robust.

3.3.   COVID-19 Death Analysis, State

Given the results of the political analysis at the county level, we further evaluated political leadership at the state level, examining a subset of variables found from the county-level analysis. Since only 51 observations were available, the analysis was restricted to the minority proportion in the state (1-proportion Caucasian only), the first three principal components of health status variables (accounting for 75% of the variability), population density, unemployment, the governor's party, and plurality [20]. Plurality was dichotomously coded with 0 = plurality (the 2016 voting consensus matching the governor's party) and 1= no plurality (voting block different from the governor's party).

We also surmised that there might exist an interaction effect between the governors' party and health status and modeled the interaction terms accordingly. Death rates were mapped, and states in the Northeast (New Jersey, New York, Massachusetts, and Connecticut) had higher death rates than other areas of the country. These states were omitted in secondary analysis to ensure that the results found were not due strictly to outliers.

An OLS model using the aforementioned variables captured 66% of the variability with the highest VIF of 3.24. Statistically significant variables included the minority population, all three health status principal components, and the interaction term between the governor's party and the first principal component (the linear combination representing the primary comorbidities of the population). Moran's I did not suggest that a spatial model was required at the state level (I = .060, p = 0.162). A map of the residuals is shown in Figure 4.   When removing the outliers of New Jersey, New York, Massachusetts, and Connecticut, minority status was the remaining statistically significant variable.   Health status and governor's party interaction with health status fell out of the model (Table 6).

Table 6.   Results of the regression analyses for the state models

Variable	OLS Full	p	OLS without State Outliers	p
R^2	*0.655*		*0.304*	
(Intercept)	-0.007	0.940	0.007	0.961
% Minority	-0.231	0.083	0.421	0.070
Plurality	0.049	0.627	0.078	0.609
Governor's Party	-0.056	0.609	-0.260	0.137
Unemployment	0.188	0.174	0.159	0.437
% in Poverty	0.198	0.243	-0.270	0.273
Population Density	-0.258	0.116	-0.013	0.959
Health PC1	0.201	0.000	-0.074	0.272
Health PC2	0.388	0.000	0.005	0.977
Health PC3	-0.213	0.029	0.145	0.263
Governor's Party x Health PC1	0.084	0.027	0.053	0.332
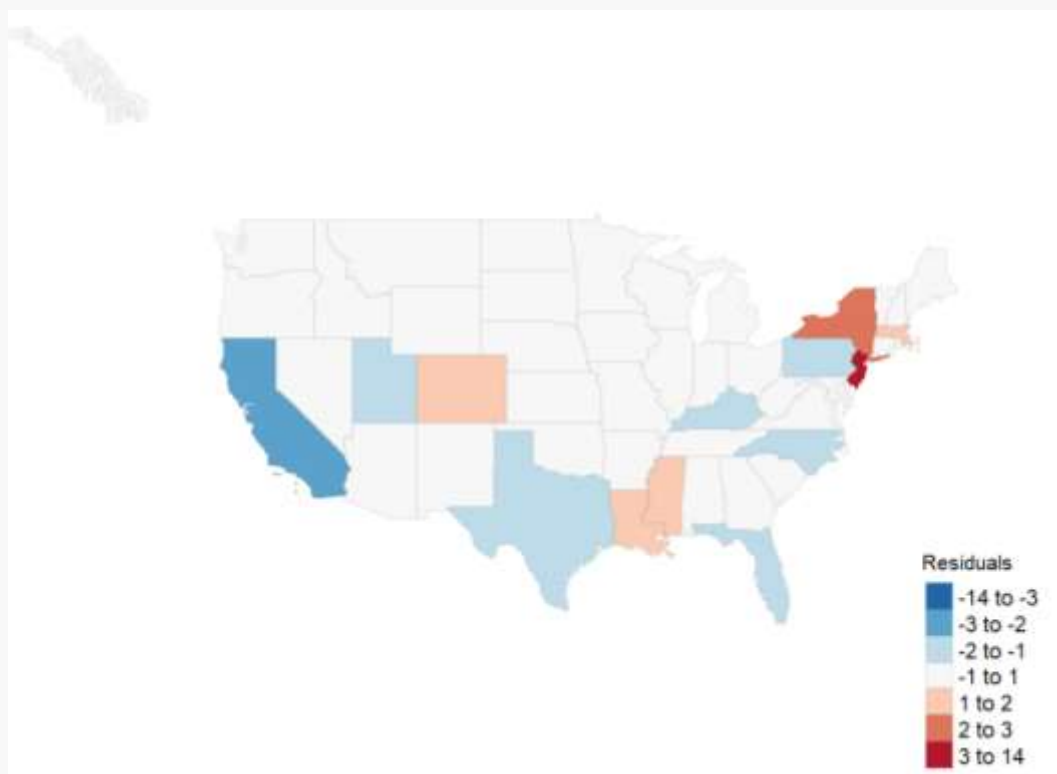
Figure 4.   Residuals, state-level initial analysis

3.4.   Flu Death Analysis, State

As a final analysis, we investigated death rates from past influenza outbreaks and governors' parties, a proxy for party politics.   Since we found an effect at the county level and an interaction effect at the state level, we wanted to see if this was constant over time based on another respiratory disease. To investigate, we ran a repeated measures (by state) analysis of variance on the log-transformed death rate for 2014-2018. The model identified no effects associated with the governor party affiliation ($F_{(1,244)}$=1.531, p=0.217), only reporting year ($F_{(4, 244)}$=2.382, p=0.040).

4. Discussion

4.1.   Summary of Results

In this study, we first ran county-level analysis for death rates based on geographical, socio-economic, health status, health capability, and political groupings.   Our investigations are reduced to two full OLS models and two geospatial models.   From our analysis, it was clear that geospatial models with lags were preferred to the OLS models.   Further, the reduced GIS model using only variables identified from lasso produced nearly the same R^2 as the full GIS model (0.500 versus 0.507, respectively).   Thus, the reduced model performs nearly as well as the full model in estimating county death rates.   In that model, we see significant geospatial effects (ρ), as well as those associated with population density, race, and the winning party in the 2016 election.   The estimate for Democratic counties (untransformed) was 4.503 deaths per 100,000

For the state-level analysis, we found effects associated with the proportion minority, three principal components associated with health status variables, and the interaction between the governor's party and the first health status variable.   However, when removing the four states with the highest death rates (New Jersey, New York, Massachusetts, and Connecticut), we found

that the only predictive variable was the proportion minority in the state. Further, an analysis of influenza death rates showed no effect associated with political party.

4.2. Population Density Effects

Population density has been identified as a predictive factor in disease progression [39,40]. A superficial examination of county-level data indicates that a relationship might exist between population density and death rate from COVID-19 (see Table 2). Consistent with prior analysis [41,42], Table 2 also shows urban areas tended to vote Democrat in the 2016 presidential election. Due to these associations, media outlets have presented the urban-rural divide as a viable explanation for the difference in death rates between counties that voted Democrat in 2016, and those that voted Republican [43,44]. This divide has also provided an explanation for the divergent response to the disease based on party affiliation. For example, Democrats are more concerned about COVID-19 than Republicans, and are more likely to wear a facemask and practice other forms of social distancing [28,45,46]. However, the effect size of population density at the county level is negligible when other factors are considered. (For example, in the reduced GIS model for counties, the standardized coefficient is only 0.051.) Population density does not appear as a significant variable in the state-level models.   The failure of population density to provide a more significant explanation for deaths from COVID-19 has been one of the surprising results from our analysis.

4.3. Race and Ethnicity / Minority Effects

At the county level, our study confirms the findings of numerous researchers pertaining to health care disparities in the United States, particularly with respect to Native American, Hispanic, and African American populations [47-49]. We found an increase in the percentage of these populations to be associated with an increase in mortality from COVID-19 at the county and state levels of analysis. McLaren (2020) attributes this difference to disparities in education, occupation, and commuting patterns [49]. The causes of disparity, however, are not explained by the covariates in this study (see Carl, 2020 [50]). Although we did not include these factors in our analysis, we did find the mortality disparities do not appear to be attributable to differences in unemployment rates or household income. Our county findings suggest that there are healthcare disparities in the United States, but may also be indicative of a pathogenesis of COVID-19 that has greater and disproportionate effect within these three racial groups [51,52]. At the state level, increases in minority population proportions were also associated with increases in death rates per 100,000.

4.4. Health Status Effects

At the state level, health status (measured by three principal components and the interaction between the governor's party and the first principal component) was a predictor for the n=51 state observations.   These health status effects disappeared after removing the four outlier states from the model.   Thus, it would appear that minority status is the predominant predictor such that increases in the proportion of minorities are associated with increases in deaths per 100,000.

4.5. Unemployment Effects

At the county level (and consistent with prior researchers), unemployment characteristics were identified as having a significant association with COVID-19 related deaths [44,45]. Periods of

economic downturn have long been found to associated with declines in health status and higher suicide rates compared with periods of relative prosperity [46-48]. Specifically, we found the degree to which unemployment exists at the county level appears to have an exacerbating influence on deaths related to the disease. Initially, we conjectured this is due to sub-optimal living, transportation, and dietary conditions endured by those without employment as well as the length of time an individual remains unemployed. However, we believe this tells only part of the story. Some are indicating a prolonged and overly restrictive COVID response is deepening an already deleterious economic cycle, the result of which is increased unemployment [49]. And as unemployment increases, so does the mortality rate either directly or indirectly from the disease. In short, extended efforts to eradicate the disease may cause additional harmful secondary and tertiary effects that may be worse than the disease itself.

4.6. Political Party Effect

The influence of politics on the reporting of COVID-19 mortality was a significant finding in our analysis. County level Democratic affiliation was significantly associated with increased COVID-19 deaths, even after controlling for factors such as population density. To the best of our knowledge, this is the first time that population density and urbanization are used as controls when evaluating death rates between Democratic and Republican states.

In past years, the CDC retrospectively tabulated the number of flu-associated illnesses, hospitalizations, and deaths – a process that takes up to two years to generate an estimate. The process relies on estimation modeling in and out of hospitals based on behavioral algorithms [53]. The CDC never relies solely on death certificate data because it recognizes that there is never large-scale testing and that the clinicians do not routinely list influenza data on death certificates if the patient died of pneumonia, heart failure, or deteriorating lung disease. According to the CDC, this leads to significant underreporting of deaths due to flu every year [53].

On February 20, 2020, the CDC published guidelines for the diagnosis and mandatory reporting of COVID-19 for any patients evaluated with "COVID related" illnesses. This applied to all health care practitioners and included a comprehensive set of instructions and codes to document any relationship to COVID-19 on the death certificates [54]. This represents a significant change in reporting of the disease and consequently the inclusion on the death certificate. Three separate additional guidelines put out in March and April affirmed these measures. In addition, the new CDC guidance stated that: "In cases where a definite diagnosis of COVID–19 cannot be made, but it is suspected or likely, it is acceptable to report COVID–19 on a death certificate as 'probable' or 'presumed'"[54]. This change introduced significant potential variation in the tabulation of COVID-19 death tolls.

At approximately the same time, the Centers for Medicare and Medicaid Services (CMS) authorized an additional 20% reimbursement for patients carrying a diagnosis of COVID-19 pursuant to Sections 3710 and 3711 of the CARES Act [55]. These changes created a financial incentive for hospitals to classify patients as positive for COVID-19. Importantly, at the time these measures were introduced, the dominant model used by policy-makers – based on Ferguson et al. [56] – predicted an exceptionally high mortality rate [57]. By late March more accurate estimates predicted a mortality rate well below original expectations [58]. This should have triggered a policy reversal from the CDC and CMS, but no changes were noted. In short, in the politically charged landscape of 2020, the CDC's new way of collecting data, combined with CMS' monetary incentives, may have resulted in the over-

reporting of COVID-19 deaths. The introduction of these two new sources of reporting bias makes historical comparisons unreliable at best. Without reliable data, it is difficult to effectively fight a pandemic. This conundrum associated with the reliability of data on COVID-related deaths highlights the need for objective and uniform standards for case-identification and data-collection.

5. Conclusions

During our analysis, we evaluated the data that pointed toward political interference in the reporting of COVID-related deaths. As of August 21, 2020, it is clear that the national death rate from COVID-19 is higher than from other flu pandemics, but the increase in the reported death rate in states with Democratic governors has been greater than the increase in states with Republican governors. Much more research in the area of politicization of medical reporting is needed, particularly given the political climate of the United States.

One of the major limitations of this study is that the associated methods are unable to estimate causality.   Any variable found to be unimportant in this analysis might have its effects mediated out by others.   The coefficient estimates are associated with the model built, and the associated p-values suggest importance to that model.   A second important limitation is that this analysis is current only as of 1 September 2020.   The analysis will continue to change as the pandemic peaks and subsides.

Future research should supplement this analysis by investigating whether states with contested gubernatorial elections (e.g., those with ballot purges, an issue that is becoming more commonplace [59]) perform worse than those with normal elections. Additional research should focus on time series models as well as simulation to generate forecasts with the external regressors identified by this research.

Supplementary Materials: Variables and sources posted at mdpi.com

Author Contributions: "Conceptualization, I.F & B.C..; methodology,L.F..; software, L.F..; validation, B.B.& J.H.. formal analysis, L.F..; writing—original draft preparation, I.F., B.C., J.H., B.B., L.F>

Funding: This research received no external funding.

Acknowledgments: None.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Description	Source	Description	Source
State FIPS Code	USA Facts	% Hypertension	CMS
State Name	USA Facts	% Ischemic Heart Disease	CMS
County Name	USA Facts	% Stroke	CMS
County FIPS Code	USA Facts	Civilian Labor Force , Feb 2020	BLS
Population 2020	USA Facts	Employed, Feb 2020	BLS
Land Area, square kilometers	CB	Unemployed , Feb 2020	BLS
People per sq. kilometer	Calculated	Percent Unemployed, Feb 2020	BLS
Urban-Rural Classification	NCHS	Civilian Labor Force , 2019	USDA ERS
% < Poverty Line	USDA ERS	Employed , 2019	USDA ERS
% for Clinton in 2016	MIT	Unemployed , 2019	USDA ERS
Winning Party in 2016	MIT	% Unemployed, 2019	USDA ERS
% Below Poverty Line, 2018	USDA ERS	MHI, 2018	USDA ERS

% Smokers	RWF	Population >=65, 2019	CB
% Adult Obesity	RWF	% age 65 and over, 2019	CB
% Abusing Alcohol	CMS	Median Age, 2019	CB
% Alzheimers	CMS	Total Population, 2018	IPUMS
% Asthma	CMS	Racial Data	IPUMS
% Atrial Fibrillation	CMS	# Hospital Physicians	DHC
% Cancer	CMS	# Acute Care Beds	DHC
% Chronic Kidney Disease	CMS	# Intensive Care Beds	DHC
% COPD	CMS	# Staffed Beds	DHC
% Depression	CMS	# Discharges	DHC
% Diabetes	CMS	Sum Average Daily Census	DHC
% Drug Abuse	CMS	Hospital average length of stay	DHC
% HIV	CMS	Average market concentration index	DHC
% Heart Failure	CMS	Average hospital case mix index	DHC
% Hepatitis B or C	CMS	Geographic shape files	CB
% Hyperlipidemia	CMS		

CB = Census Bureau [18], NCHS = National Center for Health Statistics [60], USDA ERS = United States Department of Agriculture Economic Research Service [61], MIT = MIT Election Lab [62], RWF = Robert Woods Foundation County Health Rankings and Roadmaps [63], CMS = Centers for Medicare & Medicaid Services [19], BLS = Bureau of Labor Statistics [64], IPUMS = Integrated Public Use Microdata Series [65], DHC = Definitive Healthcare [17]

References

1. USAFacts. USAFacts | Nonpartisan Government Data. Available online: https://usafacts.org/ (accessed on August 7, 2020).
2. JohnsHopkins. COVID-19 Map. Available online: https://coronavirus.jhu.edu/map.html (accessed on August 4, 2020).
3. Weller, M.W.; Wheat, A.D.; Fleischmann, J.F. COVID-19 Will Reshape the 2020 Presidential Campaign. *The National Law Review* **2020**, *X*.
4. Newall, M. Most Americans support single, national strategy to combat COVID-19. Available online: https://www.ipsos.com/en-us/news-polls/npr-covid-19-strategy-and-immigration (accessed on August 7, 2020).
5. Patino, M. Coronavirus Data in the U.S. Is Terrible, and Here's Why. Available online: https://www.bloomberg.com/news/articles/2020-05-01/why-coronavirus-reporting-data-is-so-bad (accessed on August 4, 2020).
6. Brown, E.; Reinhard, B.; Davis, A.C. Coronavirus death toll: Americans are almost certainly dying of covid-19 but being left out of the official count. *The Washington Post* April 5, 2020, 2020.
7. Kliff, S.; Bosman, J. Official counts understate the U.S. coronavirus death toll. *New York Times* April 5, 2020, 2020.
8. COVID doctors challenge CDC's rules on cause of death, concerned about inflated numbers. Available online: https://justthenews.com/politics-policy/coronavirus/doctors-and-clinical-researchers-challenge-uns-who-cdc-and-dr-faucis (accessed on August 7, 2020).
9. Ingold, J.; Paul, J. Nearly a quarter of the people Colorado said died from coronavirus don't have COVID-19 on their death certificate. *The Colorado Sun* May 15, 2020, 2020.
10. Badr, H.S.; Du, H.; Marshall, M.; Dong, E.; Squire, M.M.; Gardner, L.M. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases* **2020**, https://doi.org/10.1016/S1473-3099(20)30553-3, doi:https://doi.org/10.1016/S1473-3099(20)30553-3.
11. Scannell, C.A.; Oronce, C.I.A.; Tsugawa, Y. Association Between County-Level Racial and Ethnic Characteristics and COVID-19 Cases and Deaths in the USA. *J. Gen. Intern. Med.* **2020**, 10.1007/s11606-020-06083-8, doi:10.1007/s11606-020-06083-8.

12. Ives, A.R.; Bozzuto, C. Predictable county-level estimates of R0 for COVID-19 needed for public health planning in the USA. *medRxiv* **2020**, 10.1101/2020.06.18.20134700, 2020.2006.2018.20134700, doi:10.1101/2020.06.18.20134700.

13. Altieri, N.B., R.; Duncan, J.; Dwivedi, R.; Kumbier, K.; Li, X.; Netzorg, R.; Park, B.; Singh, C.; Tan, Y.S.; Tang, T.; Wang, Y.; Zhang, C.; Yu, B. Curating a COVID-19 data repository and forecasting county-level death counts in the United States. *arXiv e-print* **2020**.

14. Flanders, W.D.; Flanders, W.D.; Goodman, M. The association of voter turnout with county-level coronavirus disease 2019 occurrence early in the pandemic. *Ann. Epidemiol.* **2020**, *49*, 42-49, doi:10.1016/j.annepidem.2020.06.011.

15. Makridis, C.; Rothwell, J.T. The real cost of political polarization: Evidence from the COVID-19 pandemic. *SSRN* **2020**.

16. CDC. Stats of the States - Influenza/Pneumonia Mortality. Available online: https://www.cdc.gov/nchs/pressroom/sosmap/flu_pneumonia_mortality/flu_pneumonia.htm (accessed on August 7, 2020).

17. Healthcare, D. Healthcare Analytics & Provider Data | Definitive Healthcare. Available online: https://www.definitivehc.com (accessed on 7/11/2020).

18. U.S. Census Bureau. Available online: https://www.census.gov/data.html (accessed on 9/7/2020).

19. Research, Statistics, Data & Systems | CMS. Available online: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems (accessed on August 7, 2020).

20. Bureau, U.S.C. TIGER/Line Shapefiles. Available online: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html (accessed on 7/11/2020).

21. CDC. People with Certain Medical Conditions. Available online: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fneed-extra-precautions%2Fgroups-at-higher-risk.html (accessed on August 4, 2020).

22. Bialek, S.; Bowen, V.; Chow, N.; Curns, A.; Gierke, R.; Hall, A.; Hughes, M.; Pilishvili, T.; Ritchey, M.; Roguski, K., et al. Geographic differences in COVID-19 cases, deaths, and incidence — United States, February 12–April 7, 2020. **2020**, *68*.

23. Dowd, J.B.; Andriano, L.; Brazel, D.M.; Rotondi, V.; Block, P.; Ding, X.; Liu, Y.; Mills, M.C. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences* **2020**, *117*, 9696, doi:10.1073/pnas.2004911117.

24. Lynn, R.; Meisenberg, G. Race Differences in Deaths from Coronavirus in England and Wales: Demographics, Poverty, Pre-existing Conditions, or Intelligence? *Mankind Q* **2020**, *60*, 511–524.

25. Rocklöv, J.; Sjödin, H. High population densities catalyse the spread of COVID-19. *J. Travel Med.* **2020**, *27*, doi:10.1093/jtm/taaa038.

26. Dragano, N.; Rupprecht, C.J.; Dortmann, O.; Scheider, M.; Wahrendorf, M. Higher risk of COVID-19 hospitalization for unemployed: an analysis of 1,298,416 health insured individuals in Germany. *medRxiv* **2020**, 10.1101/2020.06.17.20133918, 2020.2006.2017.20133918, doi:10.1101/2020.06.17.20133918.

27. Richardson, S.; Hirsch, J.S.; Narasimhan, M.; Crawford, J.M.; McGinn, T.; Davidson, K.W.; Consortium, a.t.N.C.-R. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* **2020**, *323*, 2052-2059, doi:10.1001/jama.2020.6775.

28. Grossman, G.; Kim, S.; Rexer, J.; Thirumurthy, H. Political Partisanship Influences Behavioral Responses to Governors' Recommendations for COVID-19 Prevention in the United States. *SSRN* **2020**, 10.2139/ssrn.3578695, doi:10.2139/ssrn.3578695.

29. Fang, D.; Pan, S.; Li, Z.; Yuan, T.; Jiang, B.; Gan, D.; Sheng, B.; Han, J.; Wang, T.; Liu, Z. Large-scale public venues as medical emergency sites in disasters: lessons from COVID-19 and the use of Fangcang shelter hospitals in Wuhan, China. *BMJ Global Health* **2020**, *5*, e002815, doi:10.1136/bmjgh-2020-002815.

30. Driscoll, J.; Sonin, K.; Wilson, J.; Wright, A.L. Poverty and Economic Dislocation Reduce Compliance with COVID-19 Shelter-in-Place Protocols *CEPR Discussion Paper No. DP14618* **2020**, NA.

31. CDC. People Who Are at Higher Risk for Severe Illness | Coronavirus | COVID-19 | CDC. Available online: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html (accessed on 7/11/2020).

32. Jones, B. Coronavirus death toll is heavily concentrated in Democratic congressional districts. Available online: https://www.pewresearch.org/fact-tank/2020/05/26/coronavirus-death-toll-is-heavily-concentrated-in-democratic-congressional-districts/ (accessed on August 7, 2020).

33. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **1996**, *58*, 267-268, doi:10.2307/2346178.

34. Lockhart, R.; Taylor, J.; Tibshirani, R.J.; Tibshirani, R. A significance test for the lasso. *Ann. Statist.* **2014**, *42*, 413-468, doi:10.1214/13-AOS1175.

35. Lockhart, R.; Taylor, J.; Tibshirani, R.; Tibshirani, R. covTest: Computes covariance test for adaptive linear modelling. R package version 1.02.

. Available online: https://CRAN.R-project.org/package=covTest (accessed on August 29, 2020).

36. Team, R.C. *R: A language and environment for statistical computing*, 2018.

37. Lindley, D.V. A STATISTICAL PARADOX. *Biometrika* **1957**, *44*, 187-192, doi:10.1093/biomet/44.1-2.187.

38. Drukker, D.M.; Prucha, I.R.; Raciborski, R. Maximum Likelihood and Generalized Spatial Two-Stage Least-Squares Estimators for a Spatial-Autoregressive Model with Spatial-Autoregressive Disturbances. *The Stata Journal* **2013**, *13*, 221-241, doi:10.1177/1536867X1301300201.

39. Reyes, R.; Ahn, R.; Thurber, K.; Burke, T.F. Urbanization and Infectious Diseases: General Principles, Historical Perspectives, and Contemporary Challenges. In *Challenges in Infectious Diseases*, Fong, I.W., Ed. Springer New York: New York, NY, 2013; 10.1007/978-1-4614-4496-1_4pp. 123-146.

40. Neiderud, C.-J. How urbanization affects the epidemiology of emerging infectious diseases. *Infection Ecology & Epidemiology* **2015**, *5*, 27060, doi:10.3402/iee.v5.27060.

41. Badger, E.; Bui, Q.; Pearce, A. The election highlighted a growing rural-urban split. *The New York Times* November 11, 2016, 2016.

42. Evich, H.B. Revenge of the rural voter. Available online: https://politi.co/2BAKpR6 (accessed on August 7, 2020).

43. Davis, E. Study: COVID-19 Deaths Concentrated in Urban Areas Represented by Democrats. *U.S.News* 2020.

44. Elving, R. NPR: What Coronavirus Exposes About America's Political Divide. Available online: https://www.npr.org/2020/04/12/832455226/what-coronavirus-exposes-about-americas-political-divide (accessed on August 7, 2020).

45. Civiqs. Coronavirus: Outbreak concern. Available online: https://civiqs.com/results/coronavirus_concern?annotations=true&uncertainty=true&zoomIn=true (accessed on August 7, 2020).

46. Pew. Health Concerns From COVID-19 Much Higher Among Hispanics and Blacks Than Whites. Available online: https://www.pewresearch.org/politics/2020/04/14/health-concerns-from-covid-19-much-higher-among-hispanics-and-blacks-than-whites/ (accessed on August 7, 2020).

47. Wilder, J.M. The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States. *Clin. Infect. Dis.* **2020**, 10.1093/cid/ciaa959, doi:10.1093/cid/ciaa959.

48. Dorn, A.v.; Cooney, R.E.; Sabin, M.L. COVID-19 exacerbating inequalities in the US. *The Lancet* **2020**, *395*, 1243-1244, doi:10.1016/S0140-6736(20)30893-X.

49. McLaren, J. Racial disparity in COVID-19 deaths: Seeking economic roots with census data. *National Bureau of Economic Research* **2020**, *w27407*.

50. Carl, N. An analysis of COVID-19 mortality at the local authority level in Englan. *SocArXiv* **2020**, *Online*, doi:10.31235/osf.io/dmc2v.

51. Nédélec, Y.; Sanz, J.; Baharian, G.; Szpiech, Z.A.; Pacis, A.; Dumaine, A.; Grenier, J.-C.; Freiman, A.; Sams, A.J.; Hebert, S., et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **2016**, *167*, 657-669.e621, doi:10.1016/j.cell.2016.09.025.

52. Price-Haywood, E.G.; Burton, J.; Fort, D.; Seoane, L. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *New England Journal of Medicine* **2020**, *382*, 2534-2543, doi:10.1056/NEJMsa2011686.

53. CDC. How CDC Estimates the Burden of Seasonal Influenza in the U.S. | CDC. Available online: https://www.cdc.gov/flu/about/burden/how-cdc-estimates.htm (accessed on August 7, 2020).

54. CDC. Guidance for certifying deaths due to coronavirus disease 2019 (COVID-19). Available online: https://www.cdc.gov/nchs/data/nvss/vsrg/vsrg03-508.pdf (accessed on August 7, 2020).

55. CMS. New Waivers for Inpatient Prospective Payment System (IPPS) Hospitals, Long-Term Care Hospitals (LTCHs), and Inpatient Rehabilitation Facilities (IRFs) due to Provisions of the CARES Act. Available online: https://www.cms.gov/files/document/se20015.pdf (accessed on August 7, 2020).

56. Ferguson, N.; Laydon, D.; Nedjati Gilani, G.; Imai, N.; Ainslie, K.; Baguelin, M.; Bhatia, S.; Boonyasiri, A.; Cucunuba Perez, Z.; Cuomo-Dannenburg, G., et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Available online: https://spiral.imperial.ac.uk:8443/handle/10044/1/77482 (accessed on August 7, 2020).

57. Adam, D. Special report: The simulations driving the world's response to COVID-19. *Nature* **2020**, *580*, 316-318, doi:doi:10.1038/d41586-020-01003-6.

58. Holmdahl, I.; Buckee, C. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. *New England Journal of Medicine* **2020**, *383*, 303-305, doi:10.1056/NEJMp2016822.

59. Purges: A growing threat to the right to vote. Available online: https://www.brennancenter.org/sites/default/files/2019-08/Report_Purges_Growing_Threat.pdf (accessed on August 8, 2020).

60. NCHS. National Center for Health Statistics. Available online: https://www.cdc.gov/nchs/index.htm (accessed on July 7, 2020).

61. ERS. United States Department of Agriculture Economic Research Service. Available online: https://www.ers.usda.gov/ (accessed on July 7, 2020).

62. MIT. Data. Available online: https://electionlab.mit.edu/data (accessed on July 7, 2020).

63. RWF. 2020 county health rankings report. Available online: https://www.countyhealthrankings.org/ (accessed on July 7, 2020).

64. BLS. Available online: https://www.bls.gov/ (accessed on July 7, 2020).

65. IPUMS. Download U.S. Census data tables and mapping files. Available online: https://www.nhgis.org/ (accessed on July 7, 2020).