THE COVID 19 PANDEMIC: A WAKE-UP CALL FOR SOCIAL CHANGE?
A QUANTITATIVE ANALYSIS OF THE US.

A Research Paper

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Regional Planning

by

Rishabh Singh

December 2022

CORNELL UNIVERSITY GRADUATE SCHOOL

APPROVAL OF RESEARCH PAPER

Name of candidate: RISHABH SINGH

First Name | Middle Name | Family Name

Graduate Field: CITY AND REGIONAL PLANNING

Degree: M.R.P.

Title of Research Paper:

THE COVID 19 PANDEMIC: A WAKE-UP CALL FOR SOCIAL CHANGE?
A QUANTITATIVE ANALYSIS OF THE US.

COMMITTEE SIGNATURES

Chairperson: ______________________ Date: Dec 19, 2022

Member: ______________________ Date: Dec 19, 2022

Member: ______________________ Date: ____________

Member: ______________________ Date: ____________

Member: ______________________ Date: ____________

LICENSE TO USE COPYRIGHTED MATERIAL

Rishabh Singh

(Student Signature)

ABSTRACT

The COVID-19 pandemic has posed a unique challenge to public health systems and societies around the world. In the United States, COVID -19 has exposed and highlighted existing social vulnerabilities, making what were once hidden issues more visible and pressing. This research paper explores how COVID-19 has exposed and further exacerbated social inequalities in the US, with a focus on how racial disparities and economic disparities have disproportionately affected the people of the US. The paper, through a county-level quantitative spatial analysis, examines the specific incidence of COVID-19 on vulnerable populations highlighting the need for social change. Employing global and local regression models, the paper conducts a spatial analysis to explore three questions in the context of COVID-19– the ‘why’, ‘where’, and ‘how much’. Why do we see the damage to life occur more in one place than the other? Where exactly do we see a more severe impact of the disease? And how much is the difference between these disproportionate impacts?

BIOGRAPHICAL SKETCH

Rishabh Singh is a sustainability and urban planning analyst as well as an environmental engineer, based in Ithaca, New York. He attained his Bachelor of Technology in environmental engineering in 2018 from the Indian Institute of Technology (ISM), Dhanbad after which he worked as an environmental manager in a metal manufacturing and power production industry for a year and a half before pursuing a master's degree in City and Regional Planning at Cornell University. Being interested in sustainability and climate change issues, his past internships in the US include working with the climate resilience and sustainability team of WSP USA, a professional consulting firm for the built and natural environment, and ITCTC (Ithaca Tompkins County Transportation Council), a Metropolitan Planning Organization (MPO). Having contributed to diverse projects involving reducing the scope 1,2 and 3 emissions, sustainable procurement, and transportation planning, he is interested in working on solutions for urban issues related to sustainability and has a deep understanding of methods involving statistical and data analytics tools such as ArcGIS, Python, R programming, Database management, MS Excel, and STATA.

Dedicated to my aging parents, because of who I am what I am.

ACKNOWLEDGMENTS

This project would have not been completed without the guidance and feedback from my academic advisors, Prof Linda Shi, and Prof Stephan Schmidt, whom I thank deeply for having patience with me throughout the project. I would also like to thank Prof Linda K. Nozick for getting this project started and teaching me the methods involved in this project.

I would especially like to thank my family members – my mother, Mrs. Darshan Singh, my father – Dr. Yogendra Singh, and my sister – Dr. Ritika Singh, for their constant support and belief in my abilities to get through the numerous difficult challenges that came along my way throughout the journey. Lastly, I would like to thank all my friends who always cheered me on to finish the task.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF MAPS

LIST OF ABBREVIATIONS

The U.S. - The United States of America

COVID-19 - Coronavirus Disease 2019

SARS-CoV-2 - Severe Acute Respiratory Syndrome Coronavirus 2

MLR - Multiple Linear Regression

SER - Spatial Error Regression

GWR - Geographically Weighted Regression

OLS - Ordinary Least Squares

CDC - Centers for Disease Control and Prevention

SVI - Social Vulnerability Index

CESR - Center for Economic and Social Research

USC - University of Southern California

ATSDR - Agency for Toxic Substances and Disease Registry

GRASP - Geospatial Research, Analysis & Services Program

BRIC - Baseline Resilience Indicators for communities

HVRI - Hazards & Vulnerability Research Institute

CHRR - County Health Rankings & Roadmaps

FIPS - Federal Information Processing System

BRFSS - The Behavioral Risk Factor Surveillance System

ACS 2020 - American Community Survey 2020

MIT - Massachusetts Institute of Technology

GIS - Geographic Information System

BLUE - Best Linear Unbiased Estimator

VIF - Variance Inflation Factor

AIC - Akaike information criterion

BIC - Bayesian information criterion

PRESS - Prediction Residual Error Sum of Squares

p-value - Probability value

INTRODUCTION

Crises expose the social rifts in society. The still ongoing COVID-19 pandemic has been one of the most devastating health events for the world. In the United States, unjust actions in the past have generated disparities in the fundamental social fabric of life, which have ramifications for catastrophic events that damage the whole nation, but disproportionately (“Systematic Inequality and American Democracy,” n.d.). As witnessed in previous pandemics, vulnerable communities were hit harder by disease and had less access to testing and treatment.1 Similarly, current socioeconomic injustices raised the likelihood that historically underprivileged and vulnerable populations experience the current pandemic's most severe effects (Bavel et al., 2020). COVID-19 has highlighted and exacerbated social vulnerabilities in the United States (Perry et al., 2021). The virus has disproportionately affected marginalized populations and has revealed and deepened existing disparities in access to health care, education, and economic opportunity (Patel et al., 2020a).

In order to analyze the reasons behind the disproportionate incidence of COVID-19, three questions take importance – the ‘why’, the ‘where’, and the ‘how much’. Why do we see the damage to life occur more in one place than the other? Where exactly do we see a more severe impact of the disease? And how much is the difference between these disproportionate impacts? The answers to these questions lead us to understand the underlying issues that result in more devastating impacts of a health risk event on some communities and locations than others. For this reason, it is important for such research to be conducted on a national scale that integrates the local impacts on different locations and highlights the difference, in terms of the impact of the disease, between different communities and locations along with being able to quantify the same.

[1] *(From Black Death to Fatal Flu, Past Pandemics Show Why People on the Margins Suffer Most, n.d.)*

This research has utilized publicly available data highlighting the major social vulnerabilities that people face along with controlling for factors that constitute the efforts to reduce the impact of the disease like healthcare availability, vaccinations, and statewide lockdowns imposed. The research employed different regression models on a nationwide county-level dataset (n = 2685), to quantify the effects of different social vulnerabilities on the incidence of COVID-19. The multiple linear regression model (MLR) creates a global model to highlight the average effects of the variables chosen on the overall nation. This model is then improved, to account for spatial dependence that results due to spatial autocorrelation in the error terms (unexplained variation of the dependent variable COVID-19 death rate), in the spatial error regression (SER) model. Lastly, in order to show the variation of effects of each variable with location, the research employs the geographically weighted regression (GWR) model to create a localized model that conducts regression analysis on each feature (county) and highlights the spatial distribution of the coefficients for each variable as they vary across the whole nation. Other methods and tools have been utilized throughout to select the least correlating variables, to create an unbiased model (used in all 3 regressions), that performs well overall, and to show the performance of the same using model diagnostics.

This study confirms the aggravating effects of social vulnerabilities on the incidence of COVID-19 upon the nation overall, along with proving the disproportionate nature of that incidence on not just different communities, but also at different locations. The global regression models while explaining 50 percent of the variation (R2 = 0.5) in the cumulative COVID-19 deaths per 100,000 people per county in the US from Jan. 21, 2020, to Oct. 24, 2021, gave positive coefficients (showing positive correlation between social vulnerabilities and COVID-19 death rate) for each social vulnerability variable with high significance, thus proving on the nation level that these vulnerabilities affected the death rate associated with the disease to worsen the damage caused. The results from the GWR analysis proves that while social vulnerabilities affect a majority of the counties in terms of aggravating the incidence of COVID-19, there are some counties that do not show positive coefficients with some social vulnerabilities. However, the median (most frequent) values for the coefficients give all expected signs for the social vulnerability variables, showing that most of the counties in the US are affected adversely by social vulnerabilities. The maps from the GWR analysis highlight the spatial distribution of coefficients from minimum to maximum as well as where spatially it achieves significance (p-value) and where it does not.

This research answers the questions of why we see more damage to life in one location than another and how social vulnerabilities associated with poverty, racial minority, low education, housing cost burden, unemployment, lack of insurance, limited access to healthy foods and healthcare affected the incidence of COVID-19 across different counties in the United States along with answering where we see this happening more frequently and where we see this happening rarely. The quantitative analysis done via regression models also allows us to quantify the findings in terms of how much, globally, and locally, is the increase in COVID-19 death rate with each increasing social vulnerability, thus answering how disproportionate have the devastating impacts of the pandemic been on different communities.

LITERATURE REVIEW

According to the Centers for Disease Control and Prevention (CDC), social vulnerability, which is defined as the community's overall ability to handle adversity and is measured by indicators of poverty (such as income and wealth) and social stability (such as marital status and employment), may have an impact on how health-related outcomes are affected by disease outbreaks (*CDC/ATSDR Social Vulnerability Index (SVI)*, 2022). For instance, during the most recent COVID-19 pandemic, people from low-resource communities with higher rates of social vulnerability showed higher rates of disease incidence and death than those from more affluent places (Karmakar et al., 2021).

The socially vulnerable are more susceptible before, during, and after a catastrophic event, according to the evidence (*Greater Impact: How Disasters Affect People of Low Socioeconomic Status*, n.d.). The results hold true for members of racial and ethnic minorities, young children, the elderly, or individuals with diseases and disabilities, as well as occupants of particular dwelling types, particularly mobile homes, or high-rise flats. Additionally, these risk elements frequently coexist in combination (Morrow, 1999).The problem becomes worse as groups who are socially fragile are more at danger not just during disasters, but they also have more trouble recovering after them (Juntunen, 2004). This highlights that there are lapses in health equity. Plans for disaster and emergency response do not adequately address the needs of the most vulnerable, especially when organizations fail to take into account where the most vulnerable populations are located in the areas they oversee (Flanagan et al., 2011). Similar to previous pandemics, current socioeconomic injustices raised the likelihood that historically underprivileged and vulnerable populations will experience COVID-19's most severe effects (Bavel et al., 2020). Part of the problem might be people in the system who look at these people as a problem. It is so because, social injustices and such neighborhoods foster the environment for the disease to spread, further complicating efforts to contain the pandemic (Ahmed et al., 2020).

The research studies on the incidence of COVID-19 that followed after the 1st case was detected resulting in the COVID-19 pandemic being perceived as yet another example of how disasters regularly and disproportionately affect socially vulnerable people. Some early studies (Khazanchi et al., 2020) attempted to understand the relationship between social vulnerability and COVID-19 cases and deaths throughout the US, while others suggested that social vulnerability has much stronger local effects rather than equally contributing across states and counties (Mollalo et al., 2020), but higher death rates have been seen to have correlated with higher social vulnerabilities in most of these studies. Past research has shown that disasters, economic shocks, and other disturbances impact different segments of society in very different ways (Cutter et al., 2003; Cutter & Finch, 2008). Because they frequently have to dwell in high-risk places, the poor are more frequently impacted by hazards because when at-risk locations provide chances for employment, public services, or direct amenities, as well as increased productivity and salaries, these areas may be more desirable to settle in (Hallegatte, 2013).

People with low socioeconomic level seek medical attention at a later stage of sickness, which has worse effects on their health. For those who are economically disadvantaged, this results in COVID-19 health outcomes that are worse. Additionally, a person's ability to use health services "with ease, and having trust that you will be treated with respect" determines their access to health care (Patel et al., 2020b).

People with low educational attainment were also adversely impacted by COVID-19. Between February and May 2020, the labor force participation rate for those with a high school education or less fell by 4.0 percentage points, to 51.8 percent (Parkinson,Cody, n.d.). More than three out of four (76%) U.S. adults with at least a bachelor's degree have already received a vaccination or plan to, according to researchers with the Center for Economic and Social Research (CESR) at the USC Dornsife College of Letters, Arts, and Sciences, compared to just over half (53%) of those without a college degree (Medzerian, 2021).

It is a fact that health services in the US are expensive. At the time of a health risk event, the huge population that does not have health insurance are likely to be more adversely hit by the event. Therefore, the COVID-19 pandemic's harm as well as health disparities may be exacerbated by insurance coverage gaps, state decisions to refuse Medicaid expansion and postpone preventative actions. The rising unemployment would probably result in further losses in health coverage, which could expand these discrepancies even more (Gaffney et al., 2020).

Non-White communities in the US are frequently faced with higher rates of chronic disease, lower levels of infrastructure investment, lower access to healthcare and healthy food options, higher crime, and shorter life spans as a result of historical and ongoing patterns of racial and ethnic segregation, environmental degradation, and socioeconomic deprivation (Braveman et al., 2010). Moreover, during COVID-19, people of color still had to go to work and stay at frontlines in healthcare. Experts say, that people of color were a majority of the workforce during the pandemic (Carrazana, 2022).

The most vulnerable individuals are those whose needs are not adequately taken into account in local response and relief organization planning. For instance, during emergencies, those with poor English proficiency, the hearing and visually impaired, and other special needs groups are typically not given proper real-time evacuation information. (U.S. Department of Transportation 2006). People with other health issues were at higher risk from COVID-19. Prevalence of diseases as well as lack of a healthy diet affected the ease of contracting the virus and the severe affects that followed. There is evidence that diabetes and hypertension have been related to higher COVID-19 mortality (Patel et al., 2020b). Researchers also discovered that those who reported eating the most fruits, vegetables, and legumes during the study period had a 41% reduced chance of having severe COVID effects and a 9% lower risk of obtaining COVID compared to those who reported eating the least (Godman, 2021).

Moreover, COVID-19 deaths were unevenly distributed among Democrats and Republicans[2]. Through October 2021, counties with a Republican majority had a higher proportion of COVID deaths than counties with a Democratic majority (Sehgal et al., 2022). Republicans had 153% higher excess death rates than Democrats from April to December 2021, after all adults became eligible for vaccination in Florida and Ohio (Wallace et al., 2022). This also indicates that apart from community factors in question in this research, behavior of people and their willingness to follow government's advice and mandates like lockdowns significantly affected the impact of COVID-19 across the nation. While vaccination has shown indications of reduced deaths as a result of the virus[3], people still chose whether or not to get vaccinated. Hence, it is crucial to understand all these factors and employ them in an analysis of the COVID-19 incidence.

Researchers have constantly advised the employment of county-level and census tracts-level data for demographic information as well as for the COVID -19 incidence that is becoming more and more available with the passing time (Karaye & Horney, 2020a). Kevin Credit, in his paper 'Neighborhood inequity', explored the effects of factors underlying racial and ethnic disparities in COVID -19 testing and infection rates in Chicago and New York analyzing the incidence of COVID -19 on the basis of neighborhoods (Credit, 2020).To understand and highlight the disproportionate impacts of COVID-19 for different communities and different locations, spatial analysis has been critical. While studies have employed spatial analysis tools, they have mostly been utilized for localized purposes. Aggie J. Yellow Horse, Tse-Chuan Yang & Kimberly R. Huyser analyzed how structural inequalities established the architecture for COVID-19 pandemic among native Americans in Arizona using a geographically weighted regression[4]. Another study explored social

---

[2] (*Studies Consistently Find Higher Covid Death Rates among Republicans than Democrats*, n.d.)
[3] *The impact of vaccination on COVID-19 outbreaks in the United States(Moghadas et al., 2021)*
[4] *Structural Inequalities Established the Architecture for COVID-19 Pandemic Among Native Americans in Arizona: a Geographically Weighted Regression Perspective(Yellow Horse et al., 2022)*

vulnerability variables to analyze their spatial-temporal varying impacts on COVID-19 cumulative cases in Texas using the same analysis - geographically weighted regression[5].

Therefore, while it has been noted that localized models perform better in analyzing the effect of social vulnerabilities on COVID-19's incidence in the US, performing the same for the entire nation has been a challenge. However as specified earlier, for highlighting the disproportionate impacts of COVID-19 across space, it is crucial to conduct a nationwide analysis of the US, that is able to integrate the findings across different regions and compare. Among the few nationwide spatial analysis conducted on the topic, there was one conducted by Lauren Andersen which analyzed the spatial determinants of local COVID-19 transmission in the US[6], utilizing cluster analysis, to showcase higher vulnerability of individual variables across space, and the spatial lag regression that created a global regression model ($R^2 = 0.40$) that accounted for spatial dependence. Another study created a localized model of the entire US that showed the distribution of different variables across space[7]. This study utilized the four themes of the social vulnerability index of the CDC that explained 28 percent of the variation ($R^2 = 0.28$) in the COVID-19 death rate across all counties of the US. My research builds on the existing work by firstly creating a global model (n = 2685), that accounts for spatial dependence, to show the average effects of multiple factors on the COVID-19 death rate using the spatial error regression analysis that explains COVID-19 death rate with a higher fit ($R^2 = 0.52$), and secondly by creating localized models using the geographically weighted regression analysis (GWR) that shows the distribution of the coefficients associated with each factor in each county, across the entire nation, thus taking the analogy one step further by employing the localized effects and changes in coefficients over space. Finally, this research employs not only the social vulnerability variables, but it also controls for other factors that affected COVID-19's

[5] *Exploration of spatial-temporal varying impacts on COVID-19 cumulative case in Texas using geographically weighted regression (GWR)*(X. Wu & Zhang, 2021)
[6] *Analyzing the spatial determinants of local Covid-19 transmission in the United States(Andersen et al., 2021)*
[7] *The Impact of Social Vulnerability on COVID-19 in the U.S.: An Analysis of Spatially Varying Relationships* (Karaye & Horney, 2020b)

incidence. Hence, employing variables for all counties while controlling for other factors that affected the incidence of COVID-19, the model explains a higher variation of COVID-19 death rate (SER $R^2 = 0.52$; GWR $R^2 = 0.56$) ensuring that the regression analysis employs better unbiased estimators that show the significant correlation between social vulnerability variables and the COVID-19 death rate. The models in this research control for the effects of healthcare access, lack of community resilience, past health status, and government actions specific to COVID-19 like statewide lockdowns and vaccinations, to try to show the true effect (correlation) of social vulnerabilities on the incidence of the worst pandemic in American history.

DATA

Various variables that capture different factors affecting the COVID-19's impact on people were chosen from a wide range of sources from 6 different themes which are as follows –

1. Socioeconomic status,
2. Racial minority,
3. Lack of community resilience,
4. Conservatism,
5. Health determinants, and
6. Government action.

The Center for Disease Control and Prevention Social Vulnerability Index (CDC SVI or simply SVI), developed by ATSDR's Geospatial Research, Analysis & Services Program (GRASP) uses county-level data to analyze 16 U.S. census variables (data taken from census data as well as ACS 2020 5-year estimates) for 16 social factors to produce county rankings of vulnerability for each factor. These factors include poverty, unemployment, racial minority status, educational attainment, language proficiency, disability, and others. This research utilizes variables from this index to capture – **Socioeconomic status vulnerability** (Theme 1 of CDC SVI), which is a *combination of 5 correlating variables* (created by employing Z scores) that are namely - poverty, unemployment, housing cost burden, lack of higher education, and no health insurance, and **racial minorities** comprising of three variables for three different minority races – African Americans, Hispanics, and American Indian and Alaska Natives. Other important variables like people with disabilities, those living in crowded homes or group quarters along with percentage of elderlies or young ones in household were not selected as they highly correlated with socioeconomic vulnerability, health, and other variables.

Community resilience levels were captured to estimate the factors associated with the community that led to aggravating the impact of COVID-19. The community activity variable, taken from BRIC (Baseline Resilience Indicators for communities) prepared by University of South Carolina - Hazards & Vulnerability Research Institute (HVRI), captures the social networks and connectivity among individuals and groups accounting for interactivity amongst members, volunteerism, and participation in political engagements and civic organizations during the pandemic. Other variables chosen in this theme were the Gini coefficient of income inequality that captures the income gap between the rich and poor, taken from American Community Survey 2020 – 5-year estimates. Limited access to healthy foods was also chosen to be in this theme to account for COVID-19 aggravating factors related to people's diet, which is calculated by using the distance to grocery stores.

Conservatism in people, leading to less vaccinations, and lack of belief in the severity of the virus, has been captured by two variables. All religions adherence rate of people is taken from 2010 religion census conducted by an association of statisticians of American Religious Bodies. Political affiliation for republicans showed evidence of low vaccinations. The variable (pTrump) captures the percent of Trump voters (Republicans) in each county in election 2020. The variable also captures the factors relating to the impact of people participating in election rallies in 2020 on the COVID-19's incidence.

Health determinants were captured using - The County Health Rankings & Roadmaps (CHRR) program, which provides vital health factors, including obesity, smoking, access to healthy foods, access to medical services, access to exercise opportunities, the quality of air and water, etc. for different counties in the US. With most of these variables correlating with each other and socioeconomic status, the two variables finally chosen were percent of physicians per 100,000 people that accounts for health service availability and access, and percent of people diagnosed with diabetes, to account for vulnerable people in terms of their health. Percent of physicians were chosen over the number of nursing homes or hospitals in the area because of lack of proper data for all counties.

Further, several other variables were chosen to account for positive **government actions** during the pandemic from Ballotpedia, like imposed statewide travel restrictions and lockdown durations. To account for not just the duration of the imposed restrictions, but also their timing (earlier is better), these variables have been scored on a scale of 1 to 10 with 1 being the best score for the longest and comparatively early restrictions. Out of both variables, lockdown durations were chosen because of correlation between the two and showing better performance in the model. The variable 'percent of people with vaccination series complete' (2 doses for a 2-dose vaccine and 1 dose for a 1-dose vaccine) was chosen from data provided on Social explorer by IISinfo, to account for the effect of the government provided vaccinations against COVID-19. While a lot of untrusting behaviors (due to conservatism mainly) of people effect this variable, the presence of conservatism in the model accounts for that negative correlation.

Apart from these aforementioned themes, another variable was added which is a regional dummy for locations that were in the 7 states associated with most vulnerability variables' hotspots (highlighted in the hotspot analysis results). These 7 states are namely – South Carolina, Georgia, Alabama, Mississippi, Louisiana, Arkansas, and Texas. The positive coefficient of this variable shows how much the COVID-19 death rate increased for the counties belonging to these states. The table below summarizes all the variables in the dataset, and their sources.

Table 1: Final Dataset with 14 Variables, their Short Description, Year of Data, and the Data Sources

	COMPREHENSIVE MODEL DATASET			
S.No	**VARIABLE**	**MODEL USAGE EXPLANATION**	**YEAR**	**DATA SOURCE**
	n_FIPS	Federal Information Processing System (FIPS) Codes for Counties	-	-
Theme 1: Socioeconomic Status				
X1	(Socioecon_Vul) Socioeconomic status vulnerability	County percentile rankings of a combination of **5 measures** of poverty, unemployment, housing cost burden, lack of higher education, and no health insurance.	2020	CDC/ATSDR Social Vulnerability Index 2020 Database, US
Theme 2: Racial Minority				
X2	(EP_AFAM) Percentage of African American Population	A measure of racial minority vulnerability: African Americans	2020	CDC/ATSDR Social Vulnerability Index 2020 Database, US
X3	(EP_HISP) Percentage of Hispanic Population	A measure of racial minority vulnerability: Hispanics	2020	CDC/ATSDR Social Vulnerability Index 2020 Database, US
X4	(EP_AIAN) Percentage of American Indian or Alaska Native Population	A measure of racial minority vulnerability: American Indians or Alaska Natives	2020	CDC/ATSDR Social Vulnerability Index 2020 Database, US
Theme 3: Lack of Community Resilience				
X5	(LimAc_Food) Percent of Persons with Limited Access to Healthy Foods	A measure of community's access to healthy food by distance to grocery stores. Used as a measure for unhealthy diet.	2020	The Behavioral Risk Factor Surveillance System (BRFSS)
X6	(Comm_Activ) Community Activity	Community score for social networks and connectivity among individuals and groups. A measure of interactivity, volunteering, and political engagement for community members.	2015	University of South Carolina's Hazards and Vulnerability Research Institute (HVRI)'s Baseline Resilience Indicators for Communities (HVRI BRIC)
X7	(Gini_Coeff) Gini Coefficient	A measure of income inequality in the community and the overall population.	2020	Social Explorer: ACS 2020 (5-Year Estimates) American Community Survey 2020
Theme 4: Conservatism				
X8	(pTrump) Percentage of Trump voters in Election 2020	Used as a proxy variable for a measure of political engagement in Election 2020 rallies and people who didn't take COVID mask mandates, and the requirement for getting vaccinated seriously.	2020	MIT Election Data and Science Lab. "County Presidential Election Returns 2000-2020
X9	(All_rel_adh) All Religious	A measure of religious conservatism.	2010	2010 U.S. Religion Census: Religious Congregations &

	Adherence Rate of Total Population			Membership Study. Association of Statisticians of American Religious Bodies.
Theme 5: Health determinants				
X10	(Physicians) Percent of Physicians per 100,000 people	A measure of health service availability and access.	2020	The Behavioral Risk Factor Surveillance System (BRFSS)
X11	(Diabetics) Percent of people diagnosed with diabetes	A measure of diseases capturing the most vulnerable people in terms of their health.	2020	The Behavioral Risk Factor Surveillance System (BRFSS)
Theme 6: Government Action				
X12	(Lockdown_d) Lockdown duration scores	Manual state-level scores assigned on the basis of how early and longer state-level lockdowns were imposed.	2020	Ballotpedia. United States, 2021
X13	(Perc_Vacc) Percentage of people with Vaccination Series Complete	Percent of people who have completed a primary vaccination series' all doses.	2020	COVID-19 Vaccinations in the United States, County. Data Provided by IISInfo
Dummy Variable				
X14	(In_hotspot) Regional dummy	An accounting of locations that were in the 7 states associated with most vulnerability variables' Hotspots.	**-**	Self-created in R studio.
Dependent Variable				
Y	COVID-19 Deaths per 100,000 people	Cumulative COVID-19 Deaths per 100,000 people as of Oct 24,2021. (Dependent Variable)	2021	Coronavirus Case Data 2021, Social Explorer; The New York Times, The COVID Tracking Project.

<u>Expected Regression Equation:</u>

Y(Cumulative COVID-19 Deaths per 100,000 people, or COVID-19 death rate) =

β_1(Socioeconomic status vulnerability) + β_2(Percent of African Americans) + β_3(Percent of Hispanics) + β_4(Percent of American Indians or Alaska natives) + β_5(People with limited access to healthy foods) + β_6(Community activity) + β_7(Gini coefficient of income inequality) + β_8(Republican/Trump voters in election 2020) + β_9(All religions adherence rate) - β_{10}(Percent of Physicians per 100,000 people) + β_{11}(Percent of people diagnosed with diabetes) - β_{12}(State imposed lockdowns) - β_{13}(Percentage of people with vaccination series complete) + β_{14}(County in the southeastern hotspot, i.e. in 7 states – South Carolina, Arkansas, Louisiana, Mississippi, Alabama, Georgia, and Texas) + constant.

METHODOLOGY

This study estimates the COVID-19's cumulative death rate among the counties of the contiguous United States, by analyzing variables that affect our daily lives and increase health risks for one and all at different locations, thus highlighting the effect of social vulnerabilities on the disproportionate incidence of COVID-19 across the US. The research utilizes a variety of tools to analyze quantitative data and draw conclusions from the outcomes. It makes use of multiple data analytics techniques such as variable selection algorithms, spatial visualization maps utilizing hotspot analysis, regression analysis (including Global and Local models of Multiple Linear Regression, Spatial Error Regression, and the Geographically Weighted Regression), to gain a comprehensive understanding of the different factors that cause variation in the prevalence of the disease, along with highlighting the disproportionate impacts of social vulnerabilities and COVID-19 across space.

Spatial analysis is a critical part of this research. It plays an important role to identify areas of health inequity and underserved populations. It is also used to identify the locations of - the most vulnerable populations in the country, the prevalence of COVID-19, the spatial distribution of the vaccination rate and the availability of basic healthcare resources. Moreover, it is employed in this analysis to show how the effects of social vulnerabilities vary across space, especially on the incidence of COVID-19.

The overall analysis is divided into 3 parts –

1. Analysis of variables - This analysis consists of two sub parts conducted on all major variables. A) Spatial visualizations using maps of the variables across the country were created using ArcGIS by employing bivariate symbology on each independent variable and the dependent variable, and B) Hotspot analysis was conducted to show the clusters of high and low values for each variable.

2. Global models - Multiple linear regression or MLR (General global model) and Spatial error model or SER (Global model that accounts for spatial autocorrelation)

using selected 14 variables to identify significant factors that affected the incidence of COVID-19.

3. Localized Model – Geographically weighted regression or GWR (creates 2685 localized regressions for 2685 observations) showing the distribution of coefficients for each county, over the entire nation, on maps.

The following diagram illustrates this analysis as a flowchart.

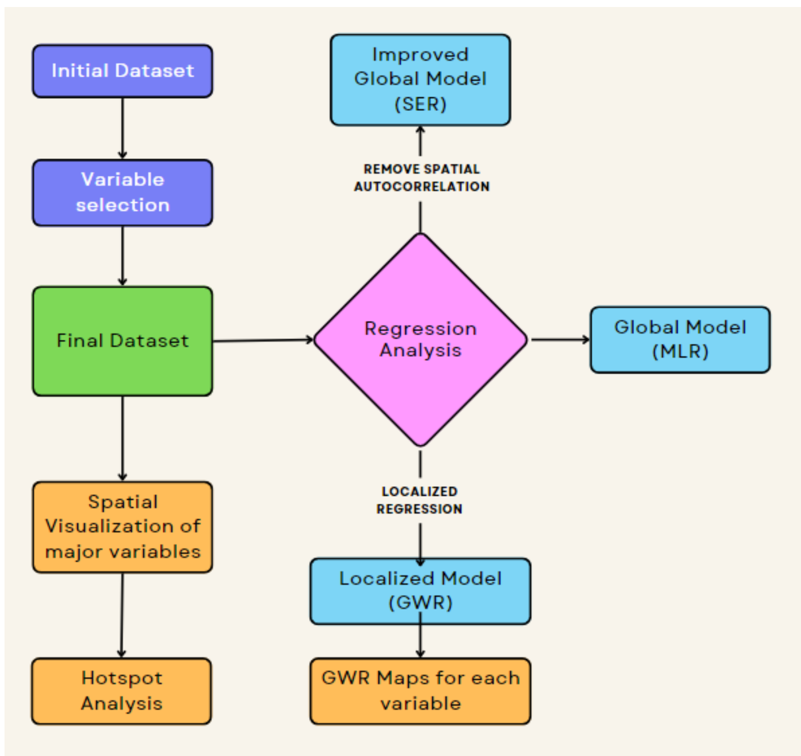


Figure 1: Illustration of the Methodology of the Research

This research analyzes the effect of social vulnerabilities on COVID-19 death rate. In doing so, the main idea behind the research has been to estimate COVID-19 death rate using all factors that affected the spread of the virus. While capturing all factors was not possible for the nationwide study attempted in this paper due to the availability of data, the starting point of the research was to create an initial dataset that comprises of all the factors that could be obtained. However, in order to make statistical inferences about the regression coefficients, it was imperative to obtain a BLUE (Best Linear Unbiased Estimator), for which it was crucial to choose the variables that A) do not correlate with each other[8], and B) have low VIF (Variance Inflation Factor) values[9] in the regression model. This research employed variable selection algorithms for choosing variables closely capturing a theme or parameter. The algorithm takes input of all the correlating variables and chooses the variables, based on the regression model, that give the best model diagnostics in terms of adjusted R-square (shows how much of the variation in the dependent variable is being explained by the explanatory variables), Akaike information criterion (AIC), Bayesian information criterion (BIC), Mallows's Cp, and PRESS statistic.[10] Thus, I obtained the final dataset with which I moved ahead, to explore the variables individually.

Spatial visualization of socially vulnerable populations across the country was conducted using ArcGIS PRO, employing bivariate visualizations (on each independent variable and the dependent variable) and the geoprocessing tool Optimized Hotspot analysis. The bivariate maps (located in the appendix) tell us the locations associated with varying levels (low: white to high: pink) of COVID-19 death rates with varying levels (low: white to high: blue) of major independent variables associated with social vulnerabilities. Their correlation over space is marked by different shades of pink and blue, with shades of pink associated with high values for death rates and low values for the independent variables, and shades of blue associated with low values for death rates and high values for the independent variables. Dark blue color signifies the locations where both the death rates and values of independent variables are high. The Hotspot analysis uses the parameters derived from the characteristics of input data (variable) to reflect the distribution (clusters) of hot spots (high values) and cold spots (low values) for the distribution of the variable over space.

The analysis gave three results – 1) It gave us indicating inferences for why social vulnerabilities associate with higher COVID-19 death rates. 2) It showed us the presence of

[8] (*12.3 - Highly Correlated Predictors | STAT 501*, n.d.)
[9] *Multicollinearity in Regression (S. Wu, 2021)*
[10] *Discovering statistics using IBM SPSS Statistics (Field, 2013)*

spatial autocorrelation in each variable, hence indicating that spatial models that account for spatial dependence by removing the autocorrelation should perform better. This is why I employed the spatial error regression to improve the global model created by the MLR. 3) It indicated a clear pattern that showed us that most social vulnerabilities have a hotspot in the southeastern region of the nation, with health service and vaccinations showing a cold spot in the region, along with high death rate for COVID-19. To analyze this further, a regional dummy variable was added that was given a value of 1 for a county located in the hotspot region, i.e., in the 7 chosen southeastern states and 0 otherwise. Overall, the hotspot analysis shows us spatially where there is clustering of counties that have high and low values for social vulnerabilities and other variables in our model to show us the spatial correlation of different social vulnerability variables with higher rates of COVID-19 deaths.

Before proceeding with the regression analysis, it was important to bring all the different independent variables to the same scale. This was done using a normalization function that reduces the range of each variable from 0 to 1. The final values were given by the function – f(x) = ((x-min(x))/(max(x)-min(x))). It is important to note that since the range of the independent variables was now 0 to 1, the interpretation of the coefficients changed as well. Now, each regression coefficient shows the effect on COVID-19 death rate on moving from 0 (minimum) to 1 (maximum) for that respective variable. After getting each independent variable in the same scale, multiple regression analysis was conducted. However, multiple regression model resulted in spatial autocorrelation among the variables which was identified by the Moran's I statistic of spatial autocorrelation[11](Moran's I: 0.286, p-value $< 2.2e-16$). This happens as regression analysis assumes that the variables being tested are randomly distributed and are not clustered in one location which does not happen for the variables chosen. As seen in hotspot maps (Maps 3-16, located in appendix) our variables tend to be spatially clustered.

[11] Spatial Autocorrelation and Moran's I in GIS(GISGeography, 2014)

The issue of spatial autocorrelation was addressed by employing the spatial error model that accounts for autocorrelation in the error terms of the regression model. Out of the two spatial regression models – Spatial error and spatial lag, that account for spatial dependence, the choice of spatial error regression was made on the basis of the Lagrange Multiplier test and the Robust Lagrange Multiplier index[12]. Its value indicated that the spatial error model will perform better on our dataset and hence I created the improved global model using the spatial error regression. This model reduced the spatial autocorrelation to a negligible value (Moran's I: -0.031, p-value = 0.993), thus improving the fit and giving better coefficients for the global model. However, in order to show the disproportionate impacts of COVID-19 associated with each variable, a local model of each variable or predictor was created using the geographically weighted regression (GWR).

GWR analysis takes place by fitting a regression equation to each feature in the final dataset. These distinct equations are created by GWR by integrating the dependent and independent variables of the features that are located within the neighborhood of each target feature[13]. The GWR model creates two sets of output – 1) On a scale ranging from minimum coefficients to maximum, it creates 5 sets of coefficients related to each variable to show the – minimum, 1st quantile (25th percentile), median (most frequent), 3rd quantile (75th percentile) and maximum values obtained for each coefficient among all the 2685 regressions done on each county. This shows the varying impacts of all the variables over all counties in the US. 2) The results of all 2685 localized regressions are shown on maps to highlight the spatial distribution of coefficients as well as where the model achieve significance (p-value) and where it does not.

[12] (*Lagrange Multiplier - an Overview | ScienceDirect Topics*, n.d.)
[13] Spatial Regression in R (College, n.d.)

RESULTS

1. Analyzing Correlation between variables –

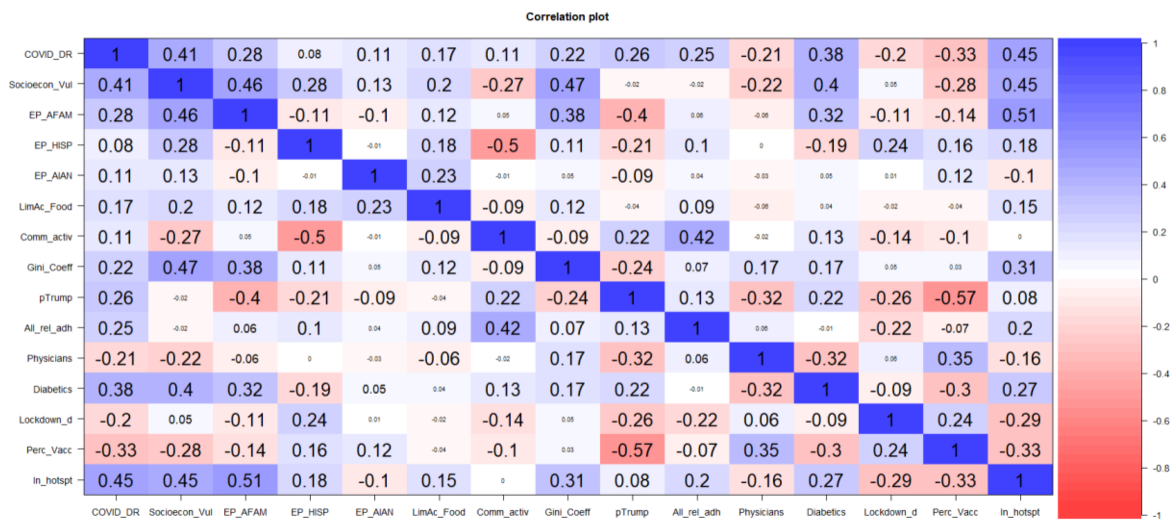


Figure 2: Correlation Matrix of the Final Dataset

While it is crucial to remove correlating variables from the dataset, as one can see from the matrix above, there is still some correlation between some explanatory variables in the final dataset. While most such variables were removed from the final dataset, some crucial variables were kept in the dataset owing to their low VIF values (<5) that shows us that they do not affect the regression on a concerning level.

Socioecon_Vul	EP_AFAM	EP_HISP	EP_AIAN	LimAc_Food
2.601681	3.454154	2.446658	1.257698	1.146675
Comm_activ	Gini_Coeff	pTrump	All_rel_adh	Physicians
2.074256	1.563691	3.296110	1.685746	1.386357
Diabetics	Lockdown_d	Perc_Vacc	In_hotspt	
1.554781	1.384411	2.044551	2.101976	

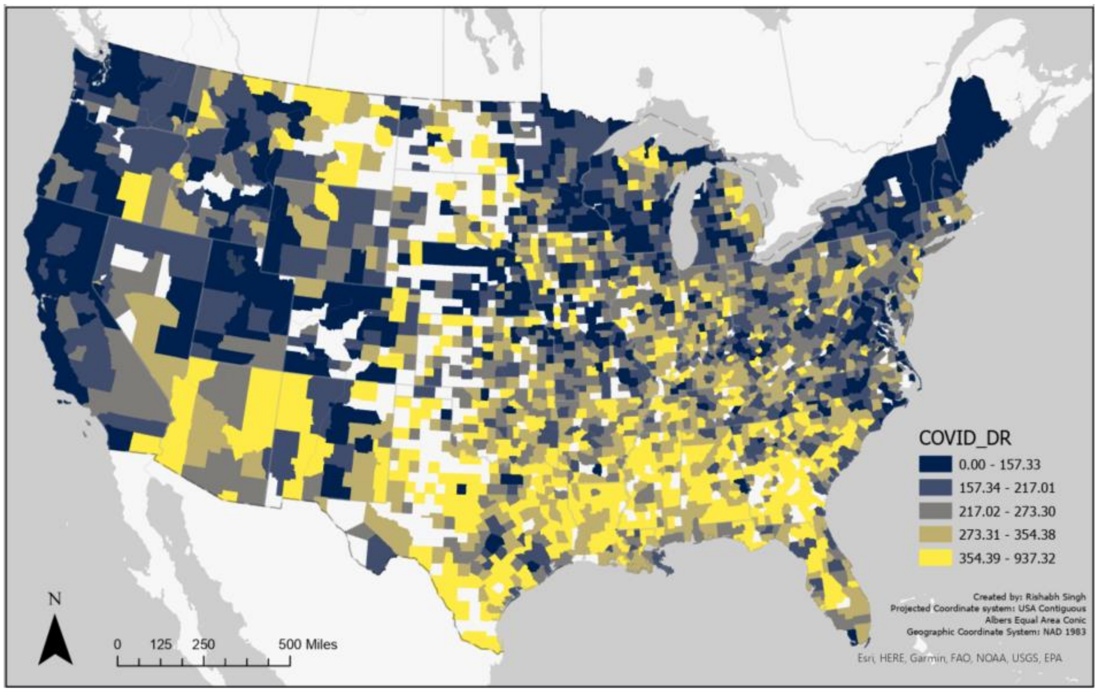
Figure 3: Variance Inflation Factors (VIF) Values of Variables in Final Dataset

Analyzing the correlation between these variables give us further meaningful insights some interesting results are as following – ‘pTrump’ and ‘Perc_Vacc’ having a correlation of -0.57 (negative), indicates strongly that republicans are associated with less vaccinations. The ‘COVID_DR’ row indicates how each variable affected the COVID-19 death rate individually. ‘In_hotspot’, the regional dummy and ‘Socioecon_Vul’ or Socioeconomic vulnerability shows highest correlation of 0.45 and 0.41 (positive) respectively, with the only negative correlation obtained for ‘lockdown_d’ or lockdown duration scores, ‘Physicians’ or Physicians available per 100,000 people and ‘Perc_Vacc’ or percent of people with vaccination series complete.

Another interesting observation is obtained when looking at the ‘In_hotspot’(last) row. As mentioned before this variable is a regional dummy for locations that were in the 7 states associated with most vulnerability variables’ hotspots (highlighted in the hotspot analysis results below), with the 7 states being – South Carolina, Georgia, Alabama, Mississippi, Louisiana, Arkansas, and Texas. The variables correlating with this variable most are ‘COVID_DR’, ‘Socioecon_Vul’ and ‘EP_AFAM’ showing higher COVID-19 death rate, higher socioeconomic status vulnerability and higher proportion of African Americans in the hotspot region. Further, negative correlation for variables like ‘Physicians’, ‘Lockdown_d’ and ‘Perc_Vacc’ indicate that in this region, there is lesser health service available along with lesser vaccination rates. Also, the lockdowns imposed in this region were shorter and imposed later than the rest of the nation. These results will be further discussed later with the results of the spatial analysis.

Other interesting results include - negative correlation obtained between community activity and Hispanics showing a lesser level of community interactivity and political/civic organization engagement among the Hispanics, positive correlation between the proportion of African American population and Socioeconomic status vulnerability showing that among all communities, African Americans are more associated with lower socioeconomic standards, i.e., higher poverty, unemployment, lack of health insurance, and housing cost burden, and lower levels of education attainment. The African American population is also associated with a negative correlation with the Republican population, showing that counties with higher number of African Americans, have lower number of Republicans.

2. Analysis of variables (Bivariate and Hotspot Analysis) –

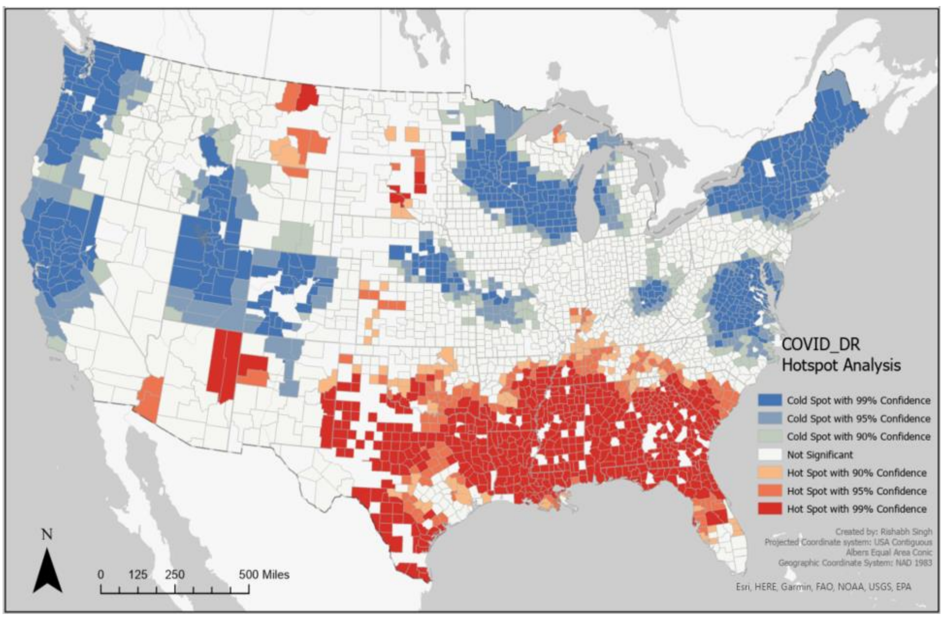


Map 1: Spatial Distribution of COVID-19 Death Rate per 100,000 People (Dependent Variable)

The spatial distribution of COVID-19 deaths per 100,000 population is the dependent variable that the models in this research try to estimate. It is an interesting variable as, if one sees the spatial distribution of COVID-19 deaths instead of deaths per 100,000 it looks quite different from the Map 1 above. While COVID-19 deaths show some states like California, New York, Texas, Florida and Illinois to have high values, that distribution does not necessarily point out a pattern. However, if we analyze the COVID-19 distribution using the deaths per 100,000 people instead, as in Map 1 above, we can see that a general pattern emerges. There are in general, higher values in the southern states, especially the southeastern states. The death rate ranging from 0 to as high as 900s, show maximum levels in the southeastern region along with a clustering pattern. I explored this further along with evaluating the reasons behind the same that gives some interesting results.

The interesting output from the bivariate maps (located in the appendix) include the identification of locations associated with high COVID-19 death rates and high values for social vulnerability variables, shown in dark blue. For the variables expected to show negative signs like health service access (Percent of physicians) and Vaccinations, the locations in dark pink show locations with high COVID-19 death rate and low values for percent of physicians and vaccinations. All of the bivariate maps among (Maps 3 to 16: located in the appendix) show a general pattern of aggravating results for the death rate and their clustering in the southeastern region.

Socioeconomic vulnerability is highly correlated (dark blue) in the southeastern region with only a few counties in other regions showing the same. Also, the maps created from measures captured by the socioeconomic vulnerability like poverty, lack of higher education, and lack of health insurance, or other variables like the Gini coefficient of income inequality and percent diabetics - all show a clustering pattern in the southeastern region. This becomes even more interesting on analyzing the variables that help control the death rates (Vaccinations and Physicians availability). On analyzing these maps (Maps 11 and 13: located in the appendix) we can see a clear pattern of pink in the southeastern region in question, showing lower levels of vaccinations and health service access and availability in the region associated with higher levels of death rates. Thus, these patterns indicate that for the southeastern states of the country, there is high correlation of higher social vulnerabilities, higher death rates due to COVID-19 and lower levels of healthcare and vaccinations. The hotspot analysis confirms these patterns by telling us that while we cannot say with a degree of confidence about the other regions, for the southeastern region of the country, there is a clustering pattern of high vulnerabilities and low healthcare in the area. The hotspot analysis maps (Maps 3 -16: located in the appendix) show the prevalence of clusters of hot spots (high values) and cold spots (low values) for each variable across the entire nation. While it is hard to discern a pattern in the other regions, the southeastern region indicates how areas with high social vulnerability are less provided for, highlighting the gap in the social support provided. It is no surprise that COVID-19's worst damages are also clustered in the same region, where many social vulnerabilities are high. While one expects the major cities to be hotspots, the choice of the dependent variable being COVID-19 deaths per 100,000 people makes it so that those cities with high populations get lesser death rates. Also, for cities that had high death rates in a region of comparatively lower death rates, they do not come up as hot spots since they were actually outliers in a cold spot. The hot spot analysis gives only those regions where many counties, neighbors of each other, had high COVID-19 death rates.



Map 2: Hot Spots and Cold Spots of COVID-19 Death Rates (Dependent Variable)

The regression analysis further investigates this correlation between social vulnerabilities and higher COVID-19 death rates in detail on all counties and specially the southeastern region in the hotspot (variable used: In_hotspot) using the global and the localized models discussed below.

3. Regression Analysis (Global models)-

Global Regression Results

	Mul_LR_Model	SP.ERR_Model
	Dependent variable:	
	COVID_DR	COVID_DR
Socioecon_Vul	46.297*** (10.222)	57.044*** (10.503)
EP_AFAM	180.140*** (20.470)	186.051*** (24.490)
EP_HISP	165.122*** (21.126)	147.203*** (25.467)
EP_AIAN	266.891*** (29.358)	252.142*** (29.909)
LimAc_Food	47.381** (23.115)	56.497** (21.752)
Comm_activ	106.251*** (24.762)	130.330*** (27.212)
Gini_Coeff	71.634*** (21.647)	48.621** (20.589)
pTrump	222.716*** (17.804)	226.586*** (19.791)
All_rel_adh	113.050*** (25.705)	86.573*** (26.871)
Physicians	-4.487 (32.255)	-22.618 (29.196)
Diabetics	143.368*** (18.192)	108.512*** (16.710)
Lockdown_d	-44.367*** (14.150)	-45.654** (20.003)
Perc_Vacc	1.627 (19.454)	16.362 (20.246)
In_hotspt	39.431*** (6.110)	46.045*** (8.452)
Constant	-107.600*** (22.886)	-108.36*** (25.751)
Lambda		0.474*** (0.022)
Observations	2,685	2,685
R2	0.412	0.521
Adjusted R2	0.409	-
Residual Std. Error	92.005	83.134
F Statistic (df = 14; 2665)	133.413***	-
Wald Statistic	-	444.5***
Log likelihood	-15949.2	-15748.1
AIC	31928.4	31526.2
Schwarz Criterion	32016.8	31614.6
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 4: Global Regression Models - MLR and SER Results

The Multiple linear regression (MLR) model revealed that social vulnerability is indeed associated with higher COVID-19 death rate. All the variables related to social vulnerabilities gave positive coefficients showing that for counties in the US, higher vulnerability values correlate with higher COVID-19 death rates for the nation as a whole. The MLR model explained 41% of the variation ($R^2 = 0.41$) in the dependent variable indicating that there were quite a number of factors that affected the mortality related to COVID-19 in the US that were unexplained by our model. The highest coefficients were shown for racial minorities, republicans, and diabetics.

Americans ($\beta = 180.14$), Hispanics ($\beta = 165.12$), and American Indians and Alaska Natives ($\beta = 266.89$) were very positively associated with COVID-19 death rate. Additionally, high proportions of republicans that voted for Donald Trump in election 2020 ($\beta = 222.716$), religious people ($\beta = 113.05$), and people diagnosed with diabetes ($\beta = 143.37$) were all positively associated with higher death rates. As expected, the model also showed that people and counties belonging to the southeastern hotspot, were also associated with higher COVID-19 death rates ($\beta = 39.43$). Socioeconomic vulnerability also showed positive association ($\beta = 46.30$) with higher death rate. Lastly, the model also showed that lack of community resilience including limited access to healthy foods ($\beta = 47.38$), community activity ($\beta = 106.25$), and Gini coefficient ($\beta = 71.63$), showed significant positive association with higher death rates. Among the factors that controlled the death rate, only lockdown duration stayed significant with a negative coefficient as expected ($\beta = -44.37$). Health service access and availability, including percent of physicians per 100,000, and percent of vaccinated people did not significantly impact the results.

Spatial error regression (SER) model results –

While the multiple regression model performed decently, it could not manage the effects of spatial autocorrelation in the dataset (Moran's I: 0.286, p-value < 2.2e-16). Hence, the model results were biased with the possibility of incorrect standard errors. To address this issue, I checked the Spatial error model results. The model diagnostics, given at the bottom, told us if the spatial error model performed better than the Multiple linear regression model. As expected, the residual standard error of the spatial error model comes out to be lower (83.13 compared to 92 for MLR). The log likelihood (higher value for SER), AIC (lower value for SER), and the Schwarz Criterion (lower value for SER) values were all improved for the SER model indicating better performance. Checking for spatial autocorrelation, I obtained the following graphs (left is OLS and right is SER)-

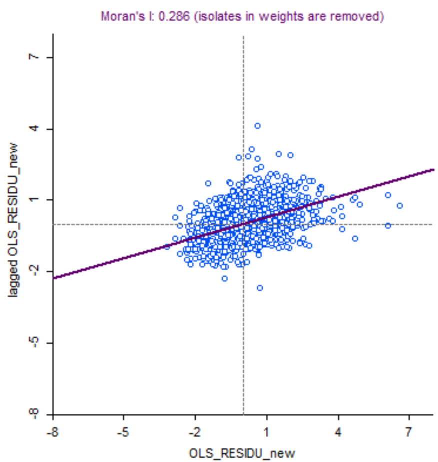


Figure 5: Spatial Autocorrelation (Moran's I) for the MLR Model

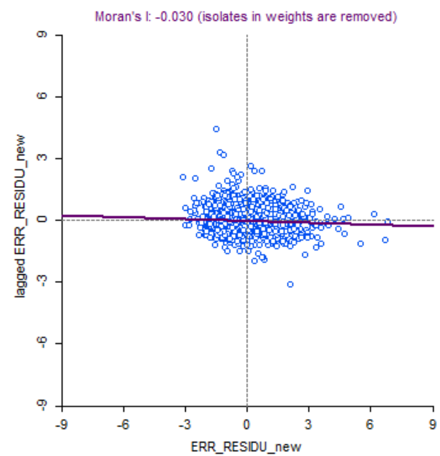


Figure 6: Spatial Autocorrelation (Moran's I) for the SER Model

As can be seen above, spatial error model as expected minimized the spatial autocorrelation in the model to a negligible amount (Moran's I: -0.031, p-value = 0.993), allowing for fairer interpretation of the regression analysis. The spatial error model does so by including a variable LAMBDA to account for the spatial dependence of the error terms. We can see that it is significant but does not have a high coefficient value.

On removing spatial autocorrelation, the SER model explained more of the variation (R^2 = 0.52) in the COVID-19 death rate. Findings demonstrated that counties in the US with greater vulnerability values also have higher COVID-19 death rates for the US as a whole. This correlation was seen for all variables related to social vulnerabilities, which all had positive coefficients. The SER model revealed that race-based minorities, Trump voters

(Republicans) in the 2020 election, involvement in the community, and diabetics had the highest coefficients.

Trump voters (Republicans) in the 2020 election displayed the second highest coefficient value (β = 226.586), after the one associated with percent of American Indians and Alaska Natives (β = 252.14), while all racial minorities, including African Americans (β = 186.05), Hispanics (β = 147.2), and American Indians and Alaska Natives (β = 252.14), gained comparably higher coefficients than most other variables. This indicates that, although racial minorities experienced lots of disproportionate effects from COVID-19, not receiving the vaccine also played a significant role in the same. The percentage of vaccinations variable in the model, however, continued to be inconsequential. The cause of the same could be traced to vaccine reactions that varied between counties with high and low COVID incidence. For instance, despite having a high pandemic death toll, New York had a high pandemic vaccination rate. As a result, even though the variable for vaccinations remained insignificant, the fact that Republicans continued to abstain from vaccinations and actively engaged in 2020 election rallies suggests that the high coefficient linked with them is due to this reason. Moreover, greater death rates were strongly correlated with higher levels of community activity (β = 130.33), limited access to healthy foods (β = 56.497), higher levels of Gini coefficient of income inequality (β = 48.62), higher percentages of religious persons (β = 86.573), diabetics (β = 108.51), and a higher level of socioeconomic vulnerability (β = 57.04). As predicted, the model also revealed that residents of counties in the southeast hotspot had higher COVID-19 death rates (β = 46.05) proving the results of the bivariate and hotspot analysis. Only lockdown duration remained significant among the variables that were expected to reduce the death rate. It obtained a negative coefficient as anticipated (β = -45.65). Access to and availability of healthcare services, particularly the number of physicians per 100,000 people, and the percentage of vaccinated individuals, had no bearing on the findings.

The spatial error model performed better than the MLR model as proven by the model diagnostic values. The major changes between the two models were some slight changes in coefficients (E.g., Socioeconomic vulnerability coefficient changed from 46.3 to 57) but not any major ones. Other changes included some variables becoming less significant (getting a lower p-value) than what they got in the MLR model (E.g., Lockdown duration). This was because the standard errors for the variables changed with more variables showing an increase in the same and a few showing a decrease. Overall, the spatial error could be concluded to have created a better model than the MLR, which was now unbiased and hence allowed for fair interpretation of the coefficients. The 4 variables that showed highest coefficients tell us that in the nation overall, COVID-19 death rate was majorly affected by race-based minorities, conservatism, and political engagements as well as involvement in the community, and also health of individuals, i.e., prevalence of diabetics.

4. Localized Model using GWR –

For this model, the regional dummy had been removed since geographically weighted regression (GWR) by itself shows the regional variations of the coefficients. The model uses weights calculated by row standardization using all links with the neighbors. (sums over all links to n). The model allows neighbors to have zero neighbors by setting the zero policy to be TRUE. (With zero policy set to TRUE, weights vectors of zero length are inserted for regions without a neighbor in the neighbors list). For estimating the bandwidths, the approach was selected to be 'AICc' that obtains the model with the lowest AICc value. A 'bisquare' kernel was chosen with the adaptive criteria set to 'TRUE'.

The model ran on all 2685 features (counties) in the dataset, showing a better fit than the global models explaining 56% of the variation (R2 = 0.56) in the COVID-19 death rate. As seen by the fit and the model diagnostics, (lower AIC value of 30975.8, compared to 31526.2 of SER and 31928.4 of MLR) the GWR model performed much better than both the global models thus proving that localized analysis performs better to estimate COVID-19's impacts.

```
***********************************************************************
*          Results of Geographically Weighted Regression              *
***********************************************************************

*********************Model calibration information*********************
Kernel function: bisquare
Adaptive bandwidth: 269 (number of nearest neighbours)
Regression points: the same locations as observations are used.
Distance metric: Euclidean distance metric is used.

****************Summary of GWR coefficient estimates:******************
                    Min.     1st Qu.      Median     3rd Qu.    Max.
Intercept      -954.27258  -210.87026  -113.90712   -22.44009  403.29
Socioecon_Vul   -79.23255    28.83058    66.60397   101.48765  258.66
EP_AFAM        -651.89025    49.90221   179.18295   335.10787 1000.65
EP_HISP       -1307.54283  -159.23224    16.78444   221.67827  632.34
EP_AIAN       -6057.84522  -314.16677    63.46148   248.09939 2366.68
LimAc_Food     -238.94072   -28.20749    45.47252   132.46871  397.83
Comm_activ     -485.06021   -95.07786    66.55521   226.29879  908.16
Gini_Coeff     -229.26557   -22.48826    39.73222   109.33170  314.65
pTrump            0.30114   215.13383   272.80155   349.29606  808.40
All_rel_adh    -227.29564     6.20149   118.06556   243.07542  658.35
Physicians     -383.17659  -108.98368    43.78090   134.47202  624.26
Diabetics       -98.27404    61.34352   112.36099   184.67271  374.81
Lockdown_d     -505.12627  -204.83075   -53.24236    68.98310 3374.88
Perc_Vacc      -351.27140  -111.25666    -1.65587    92.05199  582.69
************************Diagnostic information*************************
Number of data points: 2685
Effective number of parameters (2trace(S) - trace(S'S)): 420.9728
Effective degrees of freedom (n-2trace(S) + trace(S'S)): 2264.027
AICc (GWR book, Fotheringham, et al. 2002, p. 61, eq 2.33): 31383.6
AIC (GWR book, Fotheringham, et al. 2002,GWR p. 96, eq. 4.22): 30975.79
BIC (GWR book, Fotheringham, et al. 2002,GWR p. 61, eq. 2.34): 30490.03
Residual sum of squares: 14294407
R-square value:  0.6306552
Adjusted R-square value:  0.561949

***********************************************************************
```

Figure 7: Geographically Weighted Regression Model Result

Since I expected disproportionate impacts of the social vulnerabilities and other variables chosen based on the level of prevalence of these vulnerability factors in each county, I expected our localized models to show some counties that give opposite signs. The regression table shows that while the median or most frequent values of coefficients for each variable gave almost all the expected signs (except Physicians availability), all the variables, with the exception of just 1 (pTrump) range from negative values to positive values showing huge variation in the impact of social vulnerabilities on the death rate of COVID-19. As mentioned before, this was expected to occur on the basis of how prevalent the social vulnerabilities in that county were.

Analyzing the median values for the GWR, we can see that the variables with highest coefficients varied from the previous global models highlighting where my global models did not perform well. The variable among the racial minorities that got the highest coefficient in the global model 'EP_AIAN' became low for the median values ($\beta = 63.46$) with 'EP_AFAM' getting the highest coefficient among them ($\beta = 179.18$). Trump voters in the election 2020 still got a high coefficient (in fact, the highest) ($\beta = 272.8$). Other major changes from the global models included EP_HISP getting a low coefficient value ($\beta = 16.8$), availability of physicians attaining a comparatively high positive value ($\beta = 43.78$) against expectation. The overall result shows that while in general most counties are affected adversely by social vulnerabilities and positively by lockdown duration and vaccinations, the effect and the association between variables and the COVID-19 death rate changes across the nation. This can be analyzed better using the SWR output maps (maps 17 to 38: located in the appendix)

The second result of the GWR highlighted these variations across space (counties over the entire nation) by creating maps for coefficients along with their p-values. While some maps have been added in the main document, find the full-size maps in the appendix (Maps 17 to 38). The important results lie in the areas where yellow shaded areas (high positive values) in the coefficient maps correlate with blue values (low p-value shows significance) in the probability value maps. Socioeconomic vulnerability showed highest coefficients ranging up to 250 in the southern and southeastern regions. These areas of the maps also show significance. It is important to note that that the blue areas in the coefficient maps for socioeconomic vulnerability (capturing negative signs for coefficients) do not have significance, thus showing that the counties associated with negative coefficient values did not have significance.

Hence, these maps constitute the localized model that shows the variation of each variable's coefficients across the entire nation with the smallest unit being counties. We can see that while the southern region is associated with high coefficient values with significance, the western region although significant, shows coefficient values ranging from 50 to 100. This also indicates how and where global models indicating a maximum of 57 coefficient value for socioeconomic vulnerability lack.

The following maps can each be assessed in similar manner. Interesting results include high African American vulnerability values in the southern regions that show significance too. However, the percent of diabetics loses significance in the southeastern hotspots, showing overall high values for more northern regions than southern ones indicating the presence of a major northern vulnerability area. The western region is associated with low to medium coefficient values for all variables. Looking at the GWR maps, some other patterns start to emerge apart from the southeastern region as identified earlier. Locations associated with all variables' high vulnerabilities can be seen as yellow and pink areas on the maps. The northern region of the nation shows high coefficient values with multiple variables with negligible p-value showing significance. These regions are states namely – North and South Dakota, Montana as well as Michigan. Other areas with medium to high coefficients with different variables can be seen to be in Central America. States like Missouri, Oklahoma, Iowa, and Ohio show high coefficients for conservatism and limited access to healthy foods. Among the southern region, which is generally associated with higher coefficients, states like Texas, New Mexico, Alabama and Georgia stand out.

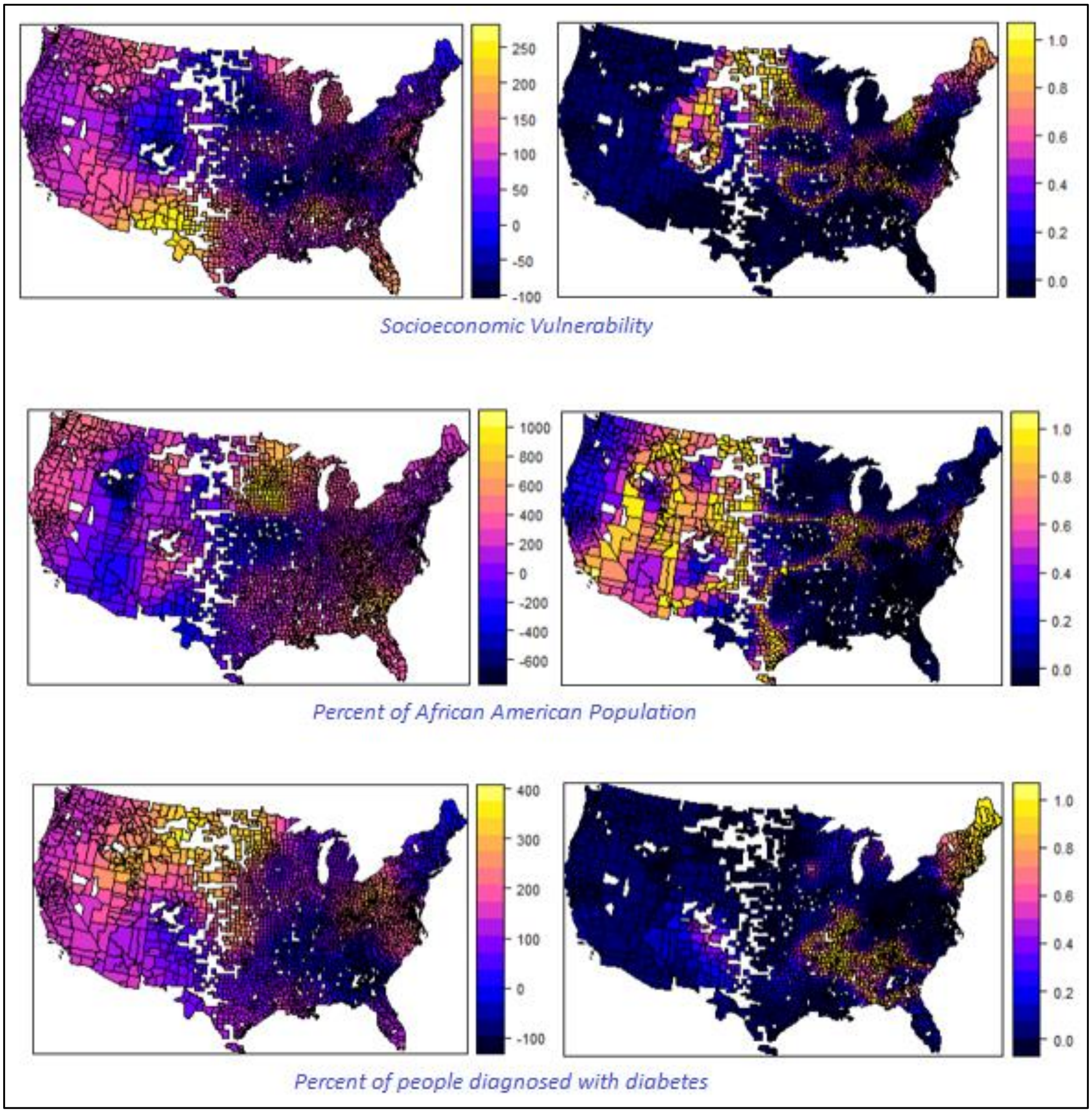


Figure 8: GWR Output Maps: Coefficient Map (left), P-Value Map (Right) for each Variable

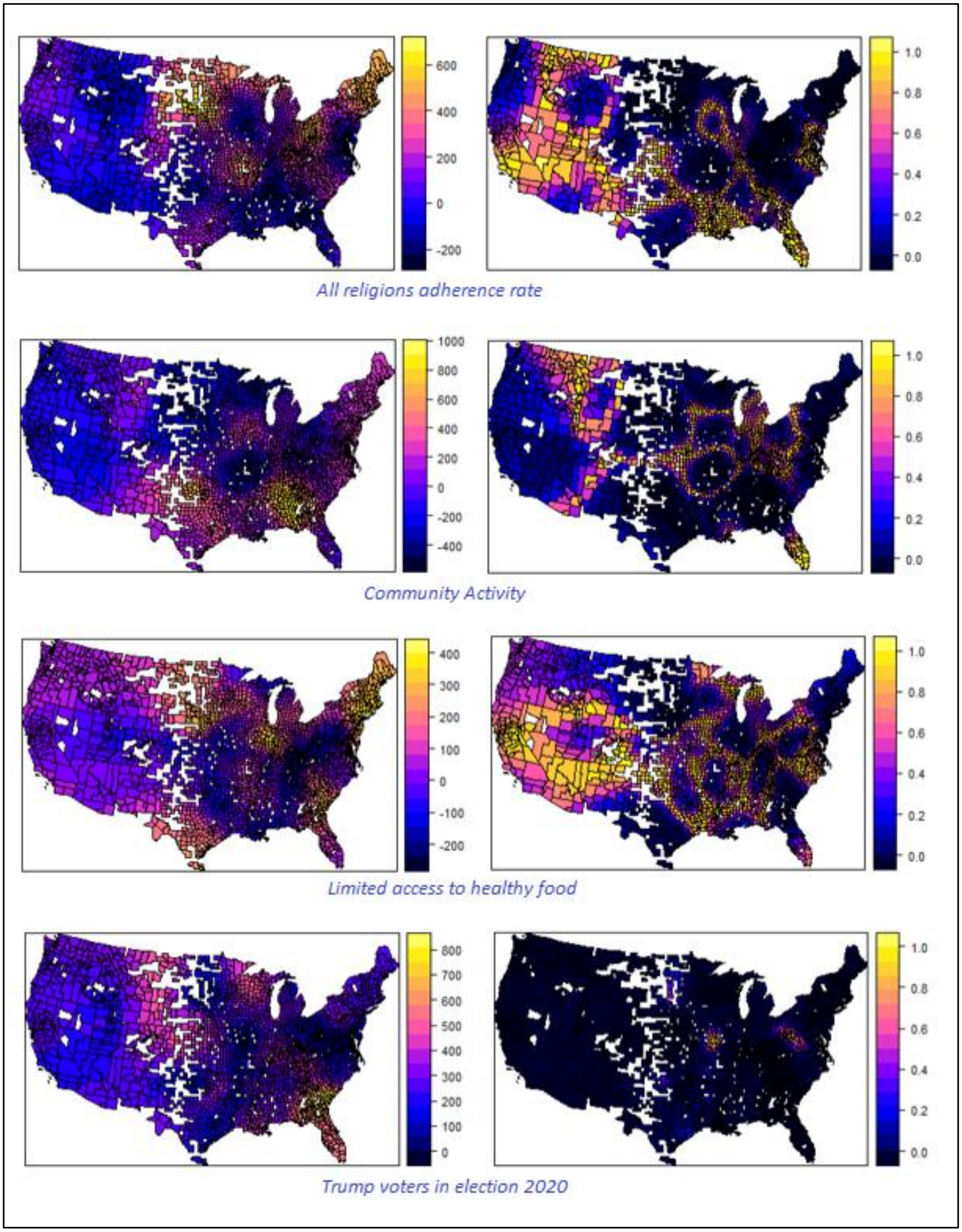


Figure 9: GWR Output Maps: Coefficient Map (left), P-Value Map (Right) for each Variable

High coefficient values for religions adherence rate are situated mostly in the central and North-east region. Here also it is interesting to see that negative values for coefficients exist where the variable loses significance indicating no significant negative coefficient values for all religions adherence rate.

Community activity coefficient values go as high as 900 and that too in the southeastern region and Texas in the south, indicating higher levels of political engagements, community-based organization participation in the area resulting in the high coefficients for COVID-19 death rate. Another interesting observation is the western region associated with negative coefficient values with significance for community activity. Also, the North-east region is associated with medium positive values of coefficients.

Maps related to limited access to healthy foods show that this variable is not very significant overall. However, this is one advantage of the GWR analysis that we can still identify the few areas where it is. It stays significant and associated with high coefficient values in the central north and south as well as the extreme northeast and southeast.

Overall, the GWR output maps[14] gave a lot of opportunity for making meaningful inferences by analyzing each variable's varying coefficient values across the counties in the nation. Showing a better fit they prove how much more advantageous they are for this analysis of social vulnerabilities that tend to vary across the nation.

[14] These maps are used here for reference. Find the actual maps in the appendix (Pg 50).

DISCUSSION

This research paper via the different analysis conducted, explores the effect of social vulnerabilities on the incidence of COVID-19 across all counties of the US. It employs spatial analysis using bivariate visualizations and hotspot analysis to discern distinguishable patterns that emerge for the prevalence of social vulnerabilities, health service lapse and COVID-19 deaths rate. We see that one clear pattern emerges from this analysis, where the southeastern region of the nation comes up as a hotspot showing a clustering pattern of high values of vulnerabilities from different variables in the region. While the region clearly has high social vulnerability across multiple dimensions, what comes out as more surprising is the fact, that our variables accounting for healthcare and vaccinations, show this region as a cold spot, showing that there is actually a lack of both in this region. This finding raises questions of whether or not the social support provided, is actually working to remove the weakness of people, and if it is capable of doing so. Upon exploring this issue further via the regression analysis, we see that the regional dummy variable gives positive association with higher COVID-19 deaths thus enforcing the finding of the hotspot analysis. The GWR results also enforce this finding by showing higher coefficients for counties in the southern and southeastern region.

The results of the regression analysis confirm for the nation average as well as for localized regions that higher social vulnerabilities were associated with higher death rate due to the pandemic. While Vaccinations and Health service access do not show significance in the global models, the GWR model shows why that happens. While for some states, they show positive correlation, for others that show a negative correlation, with the death rates. However, all the other variables show expected signs of coefficients with high significance. The Global model revealed the highest impacting variables to aggravate the incidence of COVID-19 to be racial minorities, Trump voters (Republicans) in the 2020 election, involvement in the community, and diabetics. This tells us that nationwide, vulnerability to

COVID-19 is majorly attributed to conservatism and lack of vaccinations along with excessive political engagement and participation in community-based organizations. Moreover, poor health status of an individual puts him or her in the harm's way. Sadly, but truly, a person's race also determines how much he or she is at risk from a health event like the COVID-19.

While my research has in more than one way proven the disproportionate nature of the pandemic on different communities and locations, analyzing the GWR results gave me insights into how disproportionate the impacts of COVID-19 on the population of the US were. While southern states were highly associated with worse impacts of the pandemic, other patterns also can be seen when looking at the GWR maps, in addition to the southeast, which was already observed by the global model. Other risk zones associated with higher COVID death rates were the northern part of the country that had high coefficient for variables like socioeconomic vulnerability, percent of African Americans, and limited access to healthy foods. These areas belonged to North and South Dakota, Montana, and Michigan. Central America was another risk zone that was identified. This region is mainly associated with high conservatism coefficients and restricted access to healthy foods. They included Missouri, Oklahoma, and Ohio. Other regions that were medium to high-risk zones for certain variables included the northeast and the west. These areas did not show the worst coefficients but were significant with most variables for a medium level impact of the variables on COVID-19. Interaction between multiple vulnerabilities, those not of high scale, can still lead to surprisingly negative health risks. California, for this reason has been known to have taken the brunt of the COVID impact in the west.

While the regression models used in this analysis thoroughly explore the underlying factors behind why and how COVID-19's impacts have been disproportionate on certain socially vulnerable communities, it has some **limitations**. Firstly, as the fit of all 3 models suggests, the models employed in this research still only explain a maximum of 56% of the variation in COVID death rates. Certain critical variables have not been included from themes such as environment, emergency services, hospitalizations, and inter-connecting hazards. Moreover, the model does not account for frontline workers such as nurses, and other health determinants of people such as people with Asthma. Transportation means employed, and other government interventions also affected COVID-19's transmission. Including these variables in the model would surely improve the model. However, adding more variables led to another problem of correlating variables that this research had a problem dealing with. Some crucial variables were let go because of high correlation with one or more of the variables. For example, racial minorities and socioeconomic status correlated with most variables associated with social vulnerabilities related to health service, transportation, etc. Also, as the model employs county level data to answer about populations using demographic data of the counties, it can be improved by using more local data with the best analysis being one conducted on individuals to understand direct outcomes on individuals.

CONCLUSION

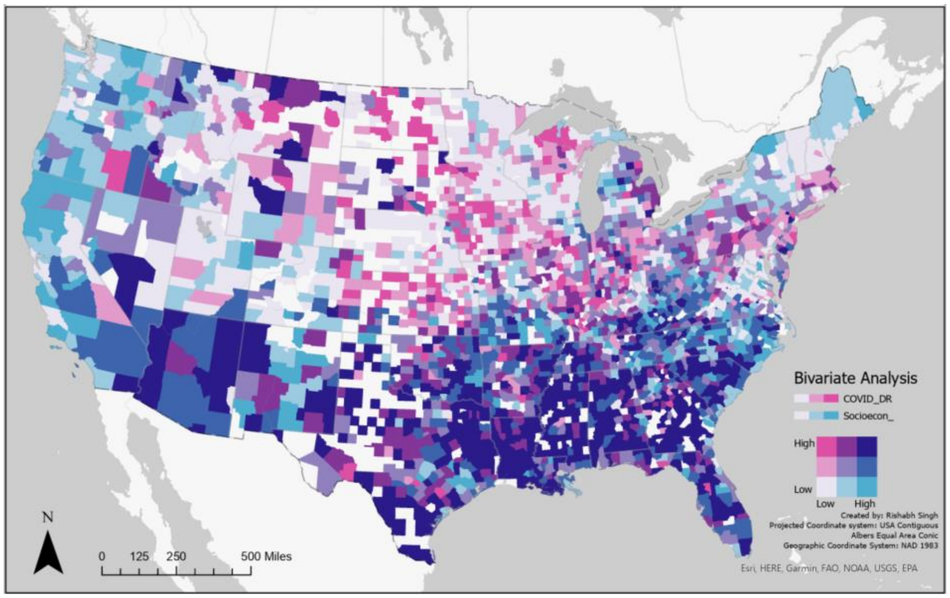
This study accounted for social, racial, community, political, health and government action related determinants by global regression models (MLR and SER) and localized regression model (geographically weighted regression), to explain variation in COVID-19 deaths per 100,000 people across the U.S. at a county level from the start of the pandemic in January 2020 to October 2021using cumulative deaths. Major outputs of the analysis include the identification of regions that will be more likely to be at risk due to future health disasters, as social vulnerabilities do not change very quickly with time. With threats like climate change, global recession, or even war looming large over the world right now, these areas will – A) Need a lot of help from the government, and B) are risky for living, as location also increases one's vulnerability, as established in this research. Major risk zones identified included the southeastern region of the U.S. comprising of 7 states, namely - South Carolina, Georgia, Alabama, Mississippi, Louisiana, Arkansas, and Texas, which was clearly indicated by the spatial analysis done earlier in the research and enforced later with the regression models. Additionally, central America as well as Northern America have high chances of becoming risk zones in the future.

The spatial model gave us a slightly better overall result than the multiple linear regression model by accounting for spatial autocorrelation. Results showed that in the nation overall, COVID-19 death rate was majorly affected by race-based minorities, conservatism, political engagements, and involvement in the community. Further, health of individuals, i.e., history of diabetes played a significant role in worsening the impacts of the pandemic. Another point to be noted in this research is the importance of understanding the severity of such pandemics. The population that did not follow the guidelines issued and did not come to terms with the importance of remedies necessary to mitigate the spread influenced the COVID-19 deaths significantly not only for themselves but also for others, proving that lack of awareness played a big role in the widespread outbreak like COVID-19.

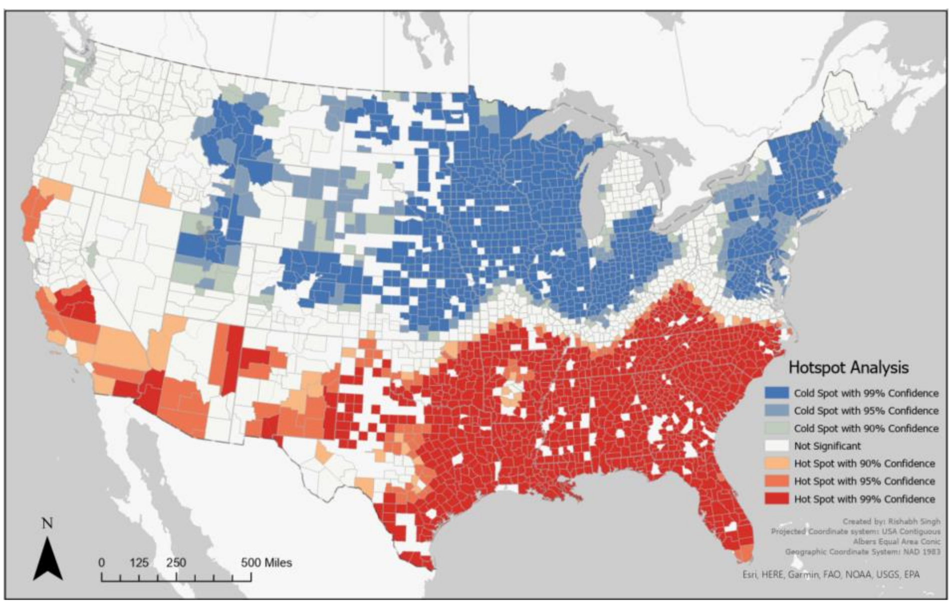
Overall, the methods employed in this project, have been able to explore the underlying themes related to social vulnerabilities in depth therefore answering the three questions it sought to answer - the "why," "where," and "how much." For future disasters and health risk events also, these questions and their answers will play a major role in saving oneself from them. A) Why does life harm seem to happen more frequently in one location than another? The model while not answering this question absolutely, gives significant reasons backed by evidence and statistical confidence as to why this happens and how they are connected with simple things associated with a person apart from the location he is in, like - his race, socioeconomic status, education level, and his health. B) Where specifically do we observe the disease's more severe effects? The spatial analysis pointed out the main clusters of COVID-19's most severely affected areas. These areas will also be risk zones for all health shock events in the future. C) How much do these disparate affects differ from one another? This question has been answered in the terms of COVID-19 death rates, but the severity can be assessed for all risk events. Exploring these questions and the given answers to save the vulnerable can be very crucial to how the US strives ahead in these challenging times.

APPENDIX

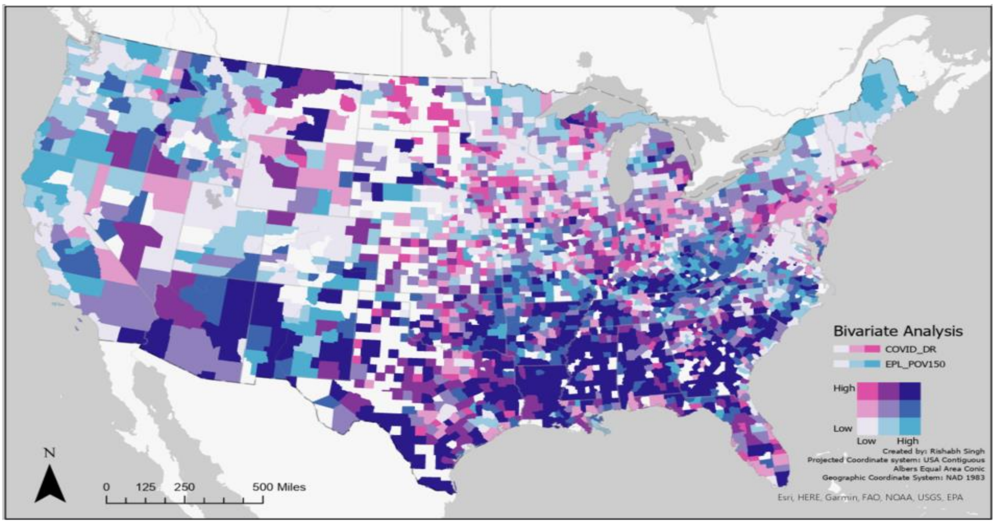
Bivariate and Hotspot Analysis Maps (3 to 16) -



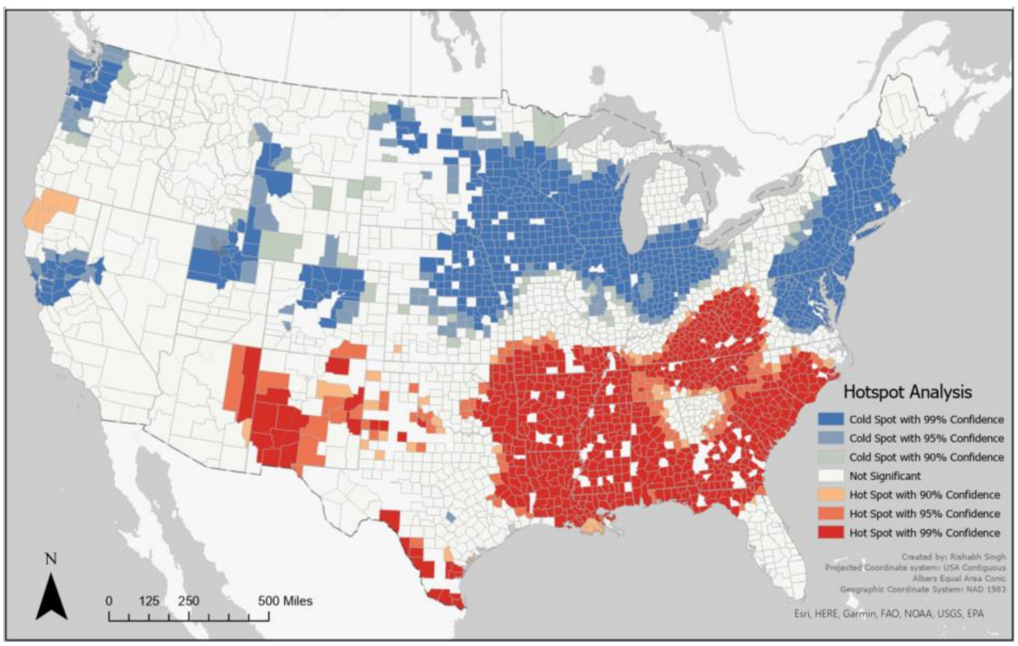
Map 3: Bivariate Visualization of 2 Variables - COVID-19 Death Rate and Socioeconomic Vulnerability



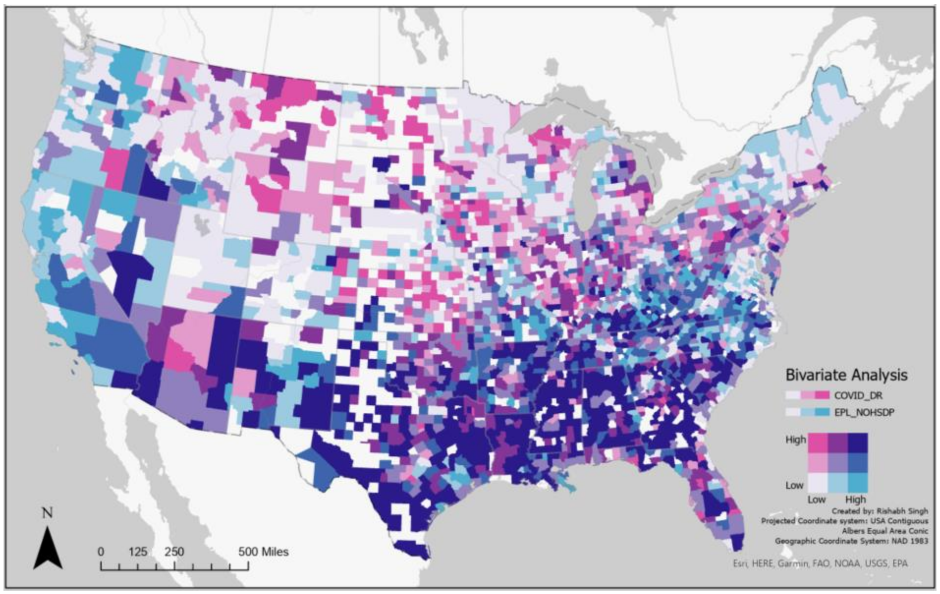
Map 4: Hotspot Analysis Showing Hot Spots (in red) and Cold Spots (in blue) for Variable Socioeconomic Vulnerability



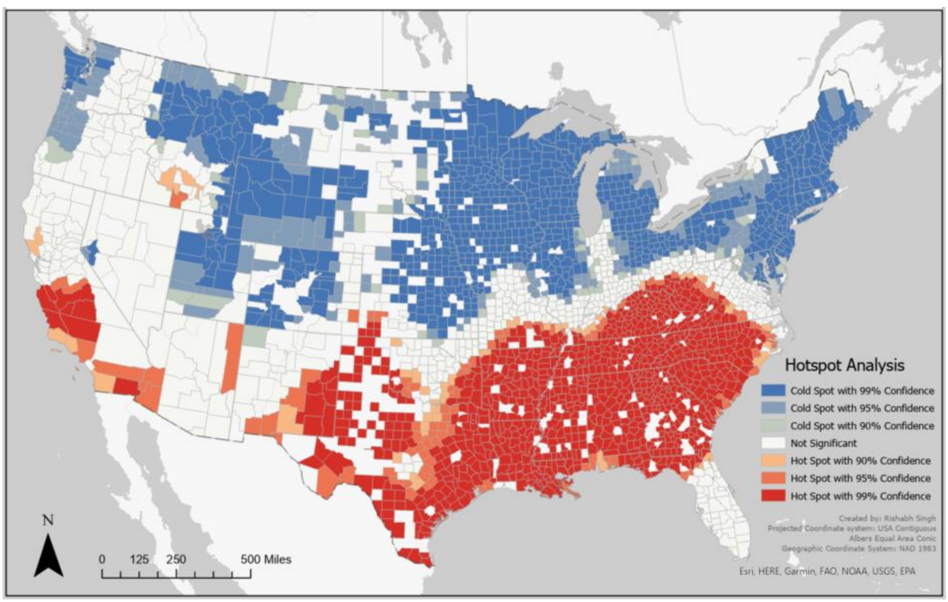
Map 5: Bivariate Visualization of 2 Variables - COVID-19 Death Rate and Poverty Rate (1st measure of Socioeconomic Vulnerability)



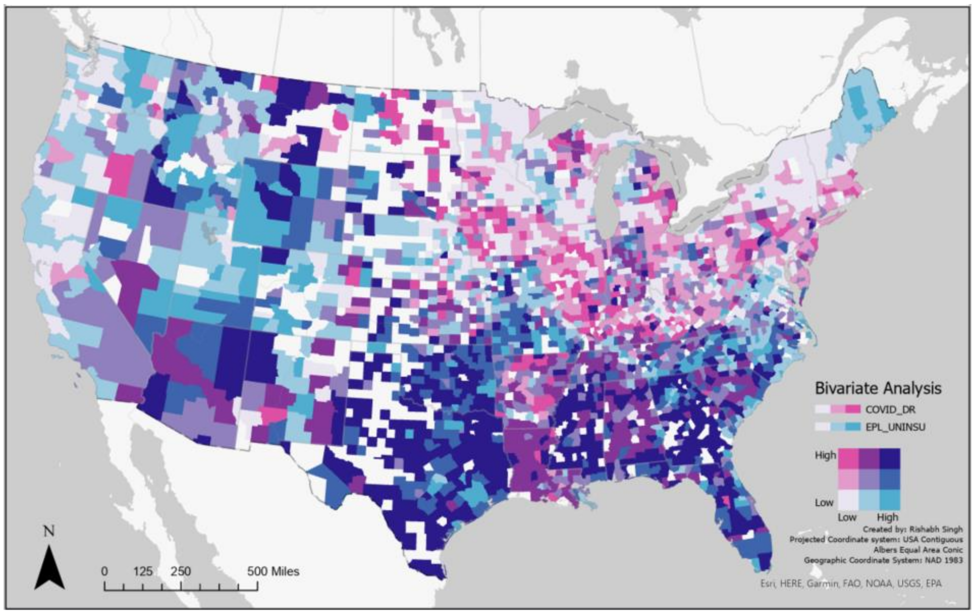
Map 6: Hotspot Analysis Showing Hot Spots (in red) and Cold Spots (in blue) for Poverty Rate (1st Measure in Socioeconomic Vulnerability)



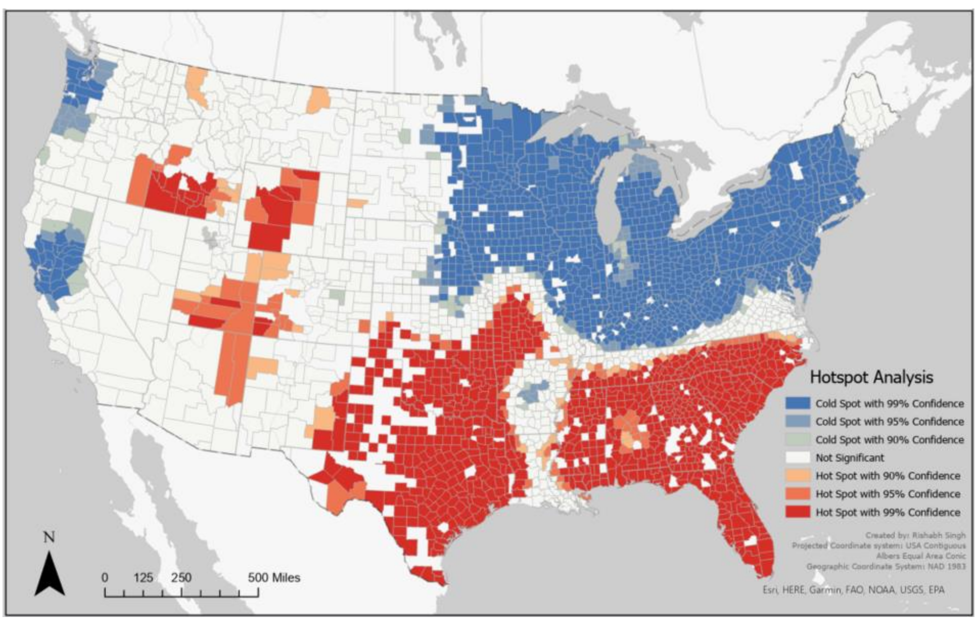
Map 7: Bivariate Visualization of 2 Variables - COVID-19 Death Rate and No High School Education (2nd Measure in Socioeconomic Vulnerability)



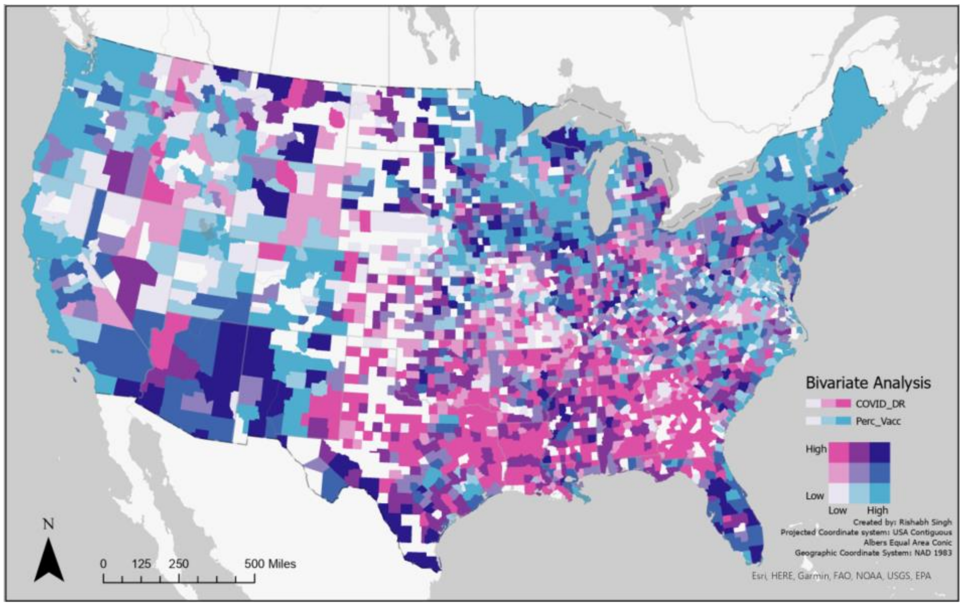
Map 8: Hotspot Analysis Showing Hot Spots (in red) and Cold Spots (in blue) for No High School Education (2nd Measure in Socioeconomic Vulnerability)



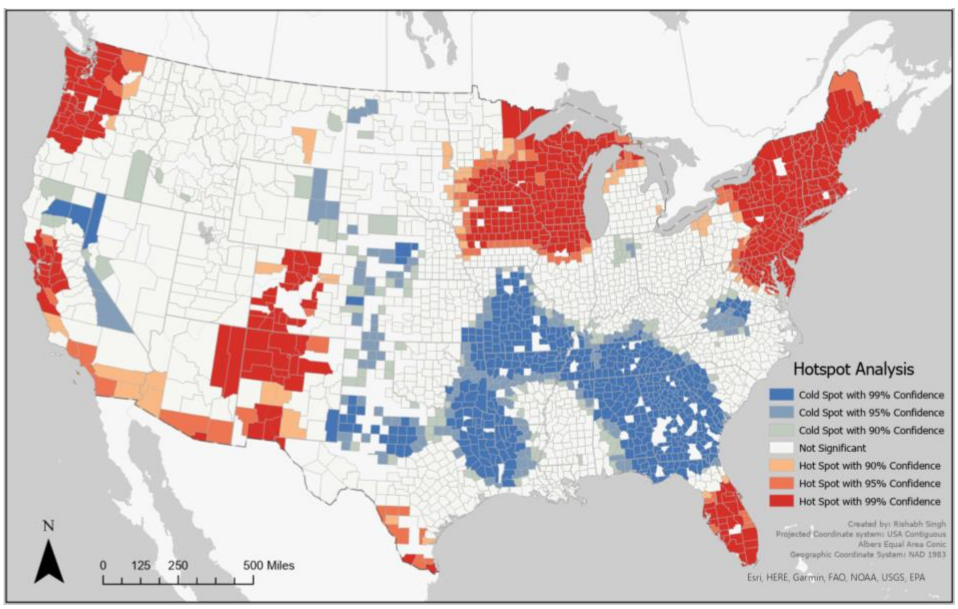
Map 9: Bivariate Visualization of 2 Variables - COVID-19 Death Rate and Uninsured Population (3rd Measure in Socioeconomic Vulnerability)



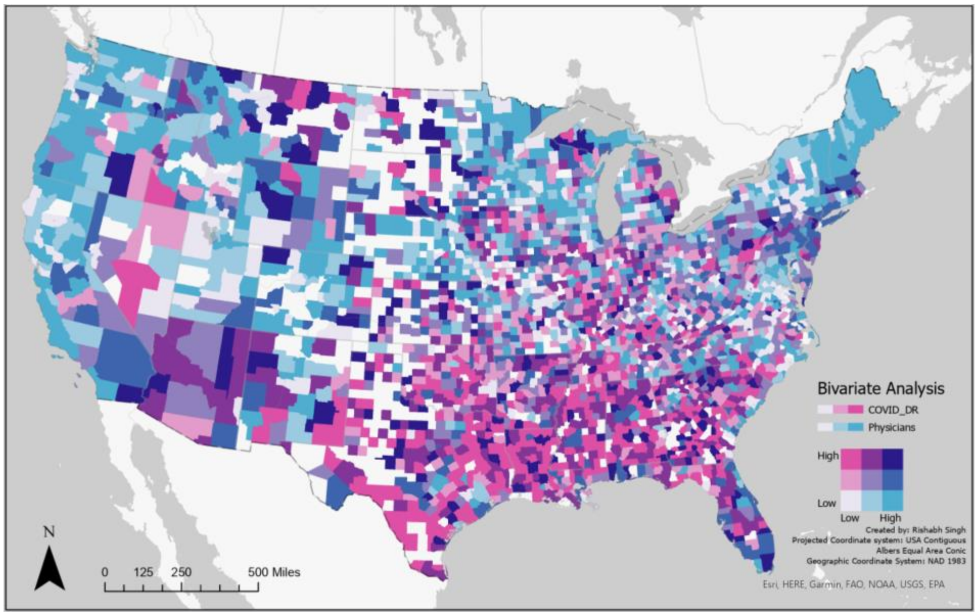
Map 10: Hotspot Analysis Showing Hot Spots (in Red) and Cold Spots (in Blue) for the Uninsured population (3rd Measure in Socioeconomic Vulnerability)



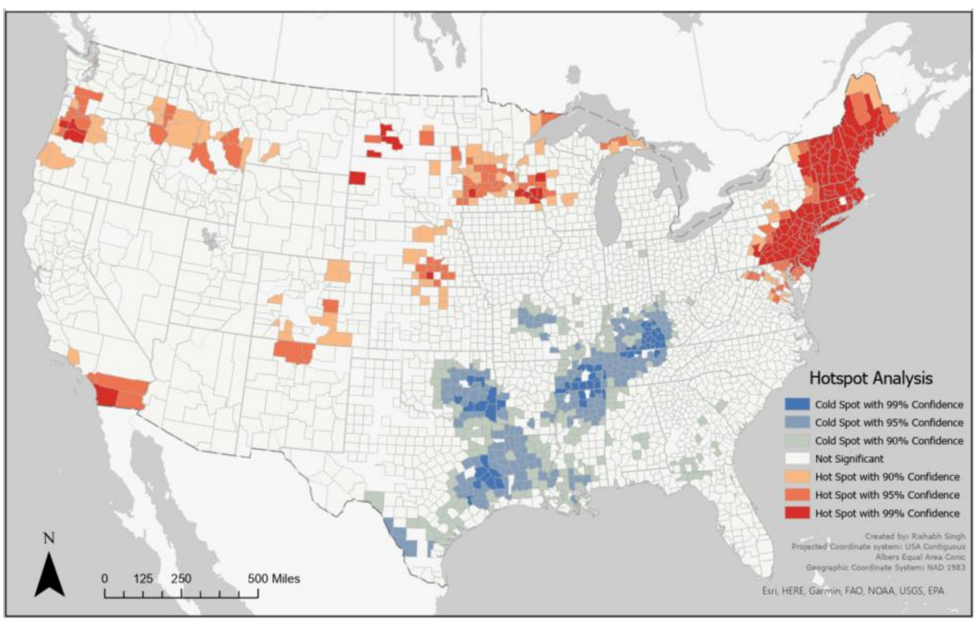
Map 11: Bivariate Visualization of 2 Variables - COVID-19 Death Rate and Percentage People with Vaccination Series Complete



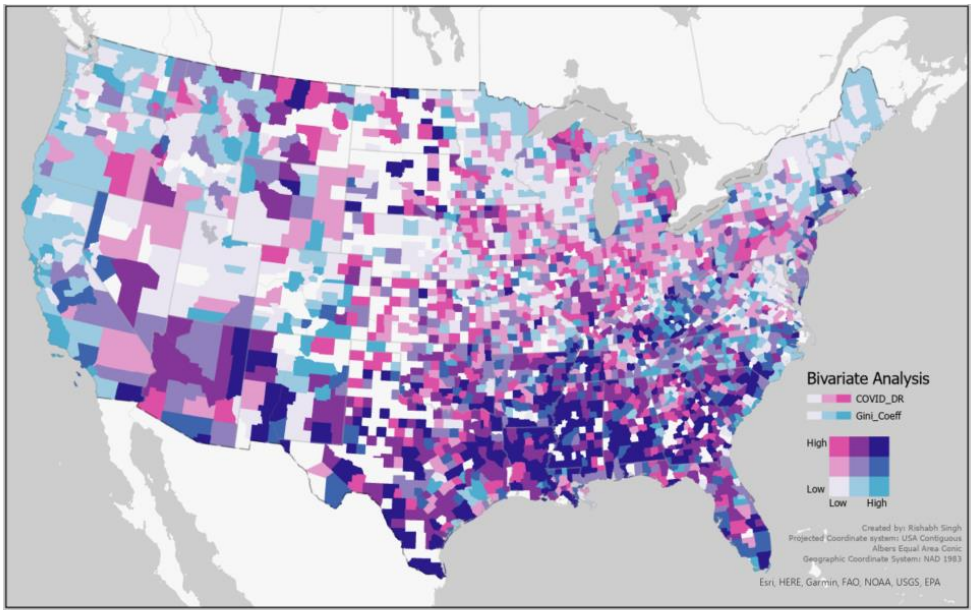
Map 12: Hotspot Analysis Showing Hot Spots (in Red) and Cold Spots (in Blue) for Variable Percent of People with Vaccination Series Complete



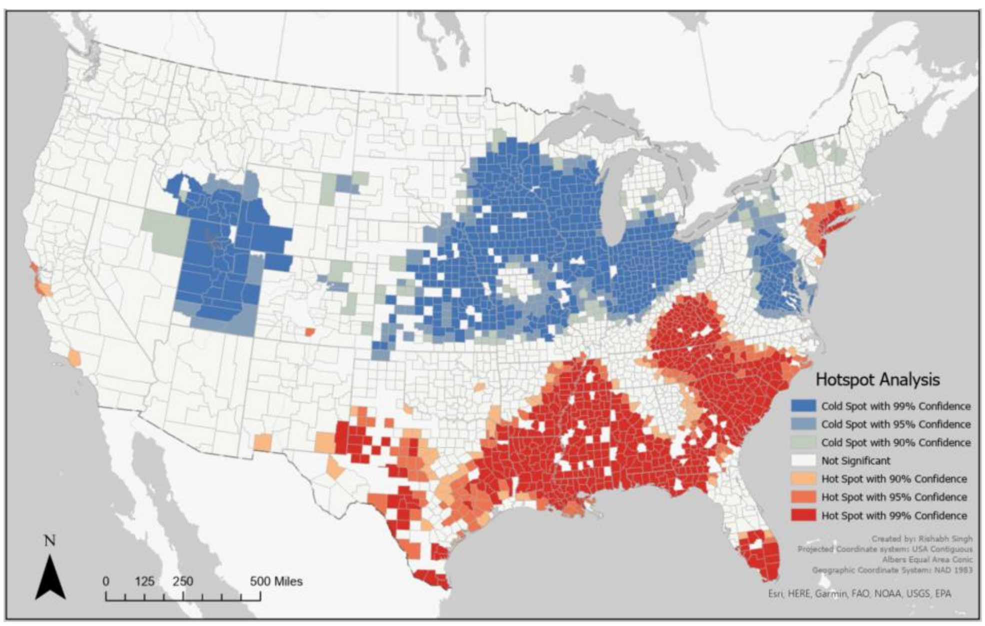
Map 13: Bivariate Visualization of 2 Variables - COVID-19 Death Rate and Percent of Physicians Per 100,000 People



Map 14: Hotspot Analysis Showing Hot Spots (in Red) and Cold Spots (in Blue) for Variable Percent of Physicians Per 100,000 People

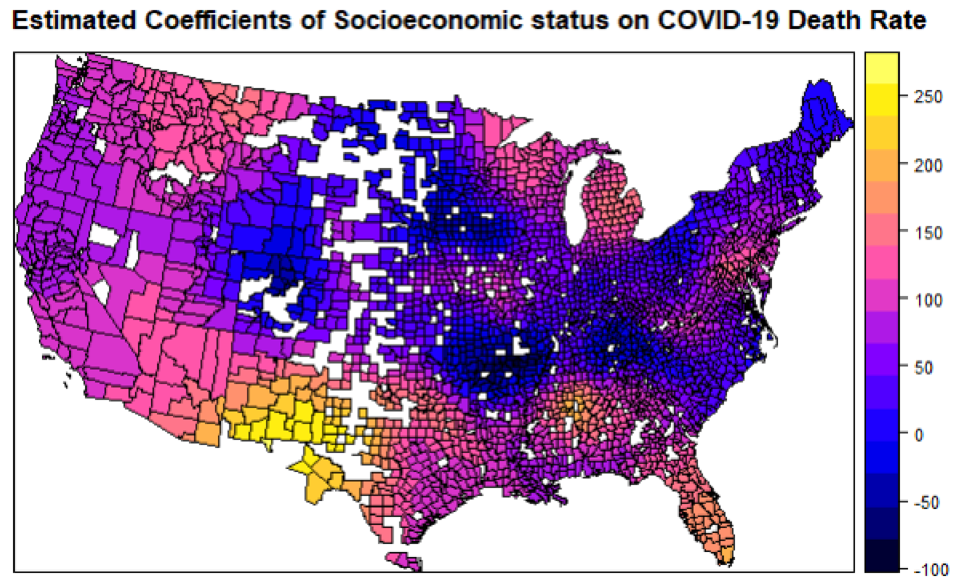


Map 15: Bivariate Visualization of 2 Variables - COVID-19 Death Rate and Gini Coefficient of Income Inequality

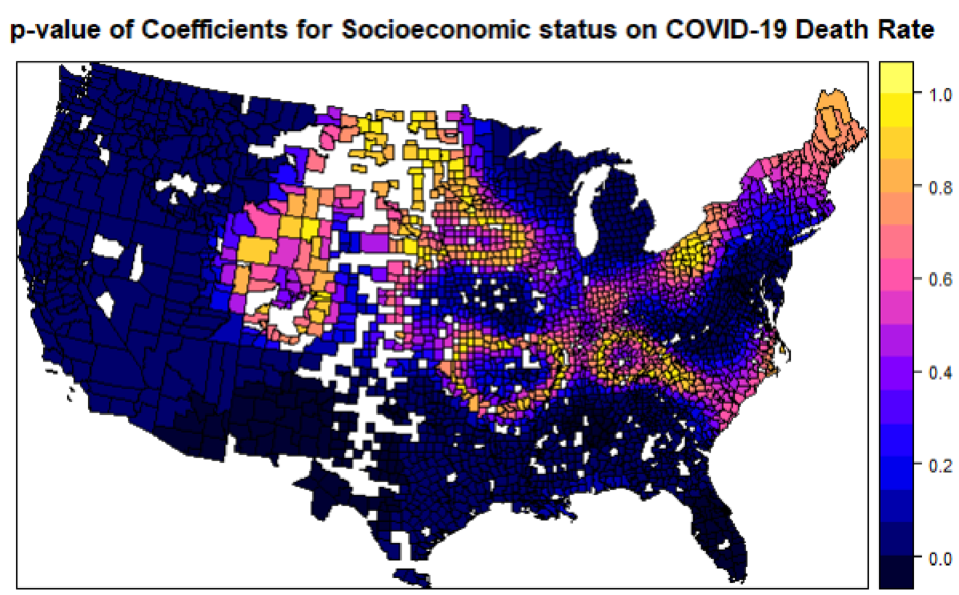


Map 16: Hotspot Analysis Showing Hot Spots (in Red) and Cold Spots (in Blue) for Variable Gini Coefficient of Income Inequality

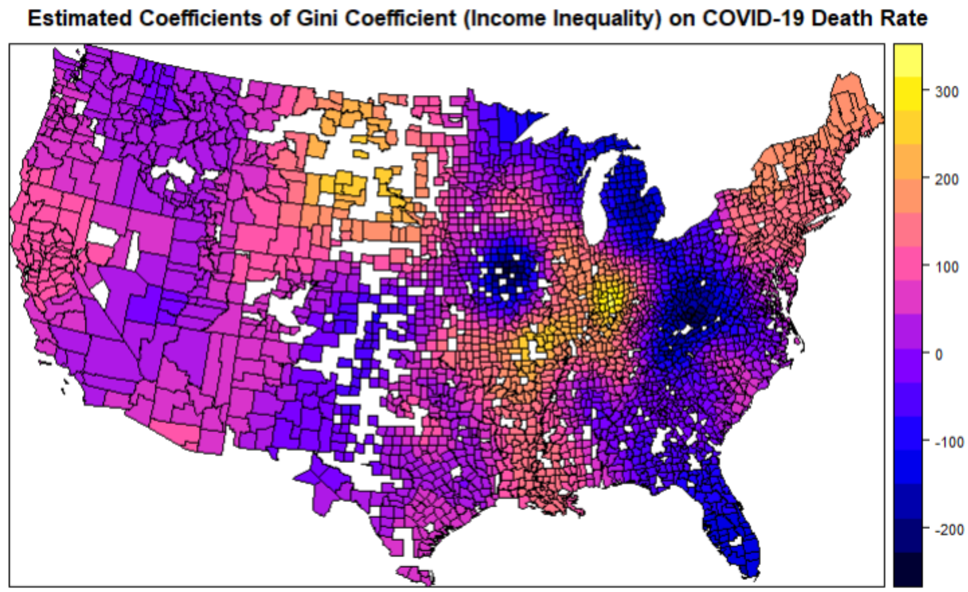
GWR Maps (17 to 38)



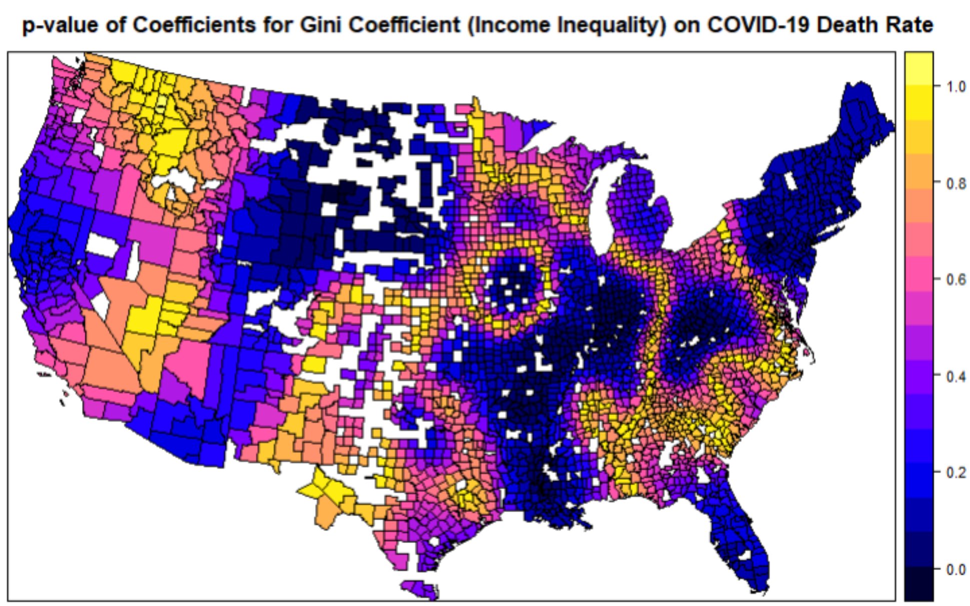
Map 17: GWR Output Map: Socioeconomic Vulnerability Coefficients



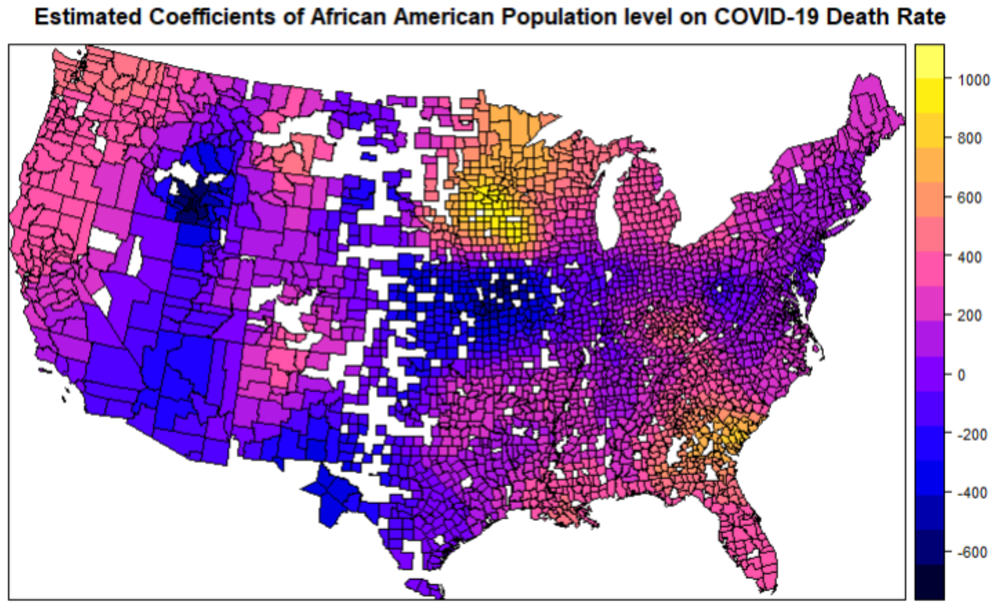
Map 18: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Socioeconomic Vulnerability



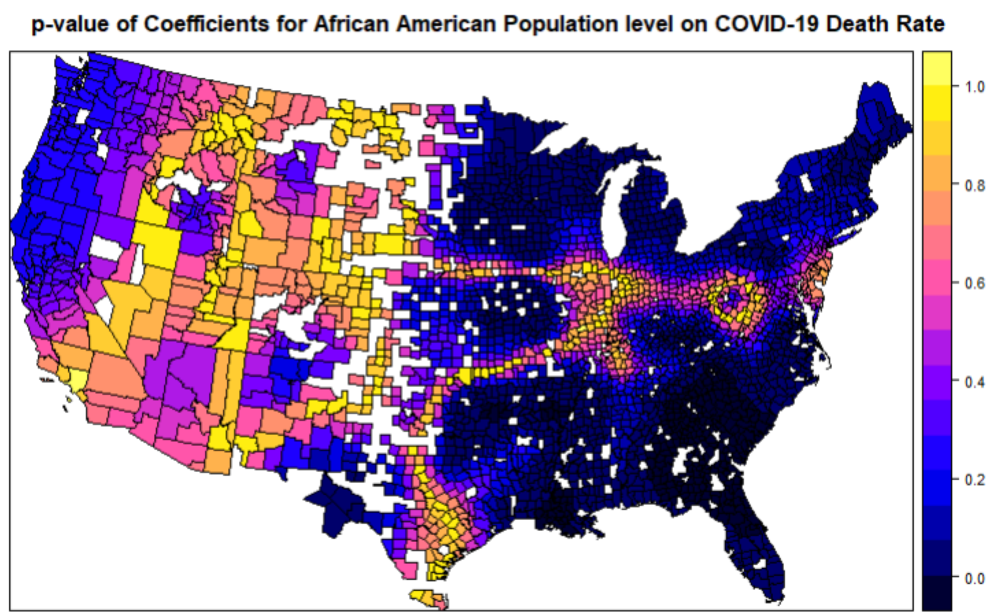
Map 19: GWR Output Map: Spatial Distribution of Coefficients of Gini Coefficient of Income Inequality



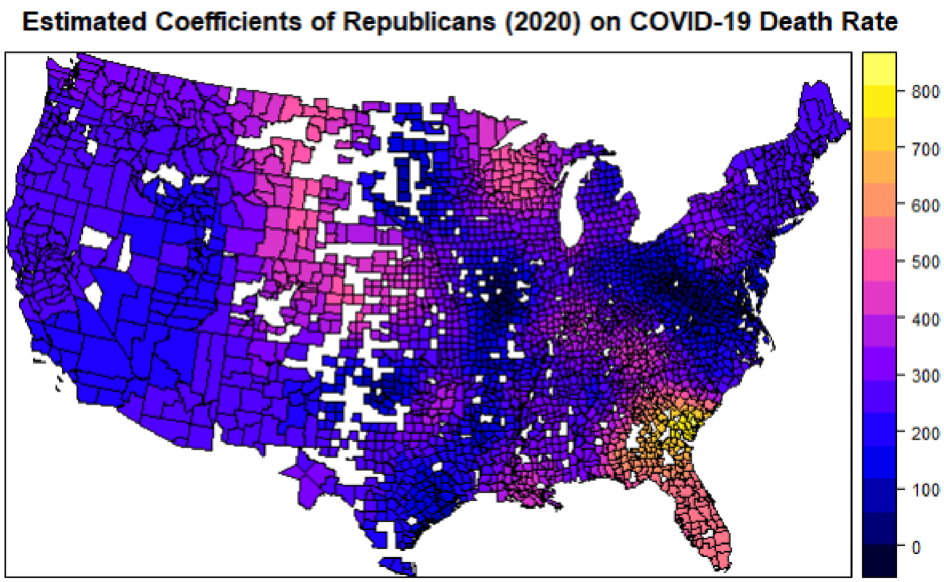
Map 20: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Gini Coefficient of Income Inequality



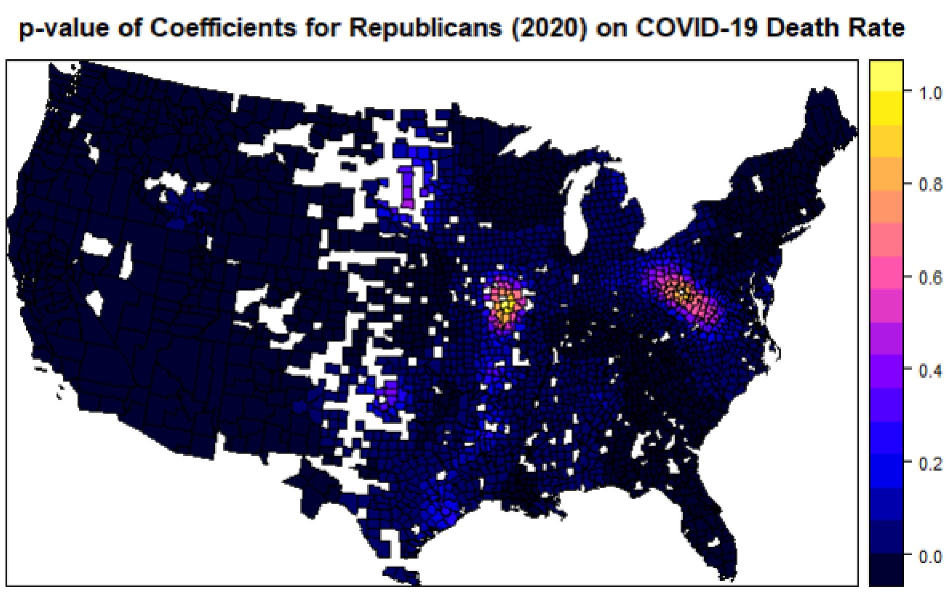
Map 21: GWR Output Map: Spatial Distribution of coefficients of Percent of African Americans



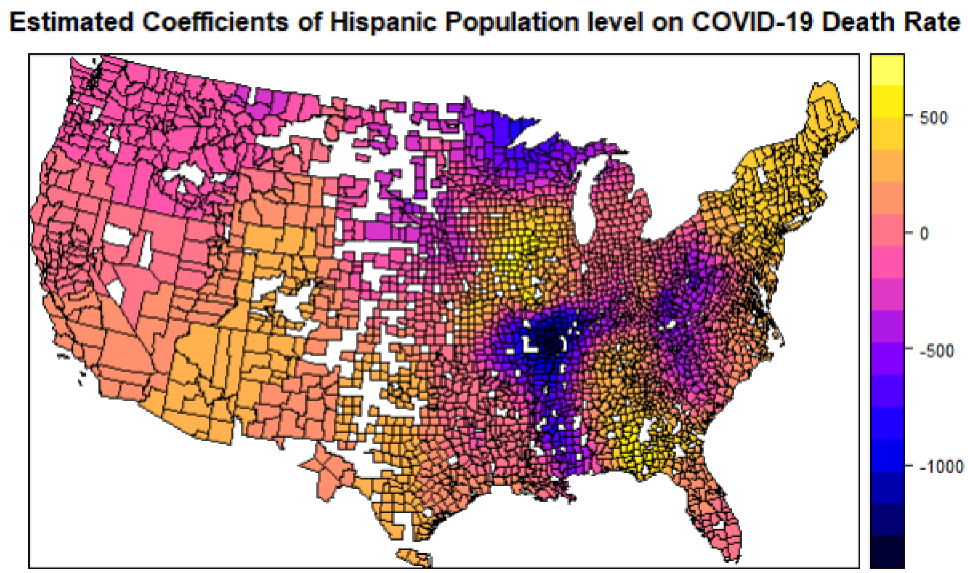
Map 22: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Percent of African Americans



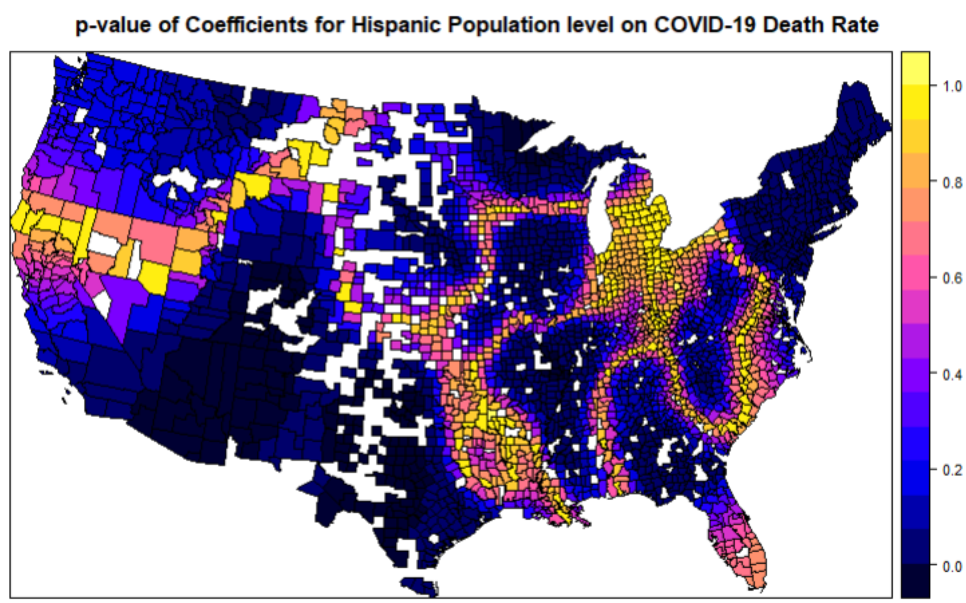
Map 23: GWR Output Map: Spatial Distribution of Coefficients of Republicans Voting in Election 2020



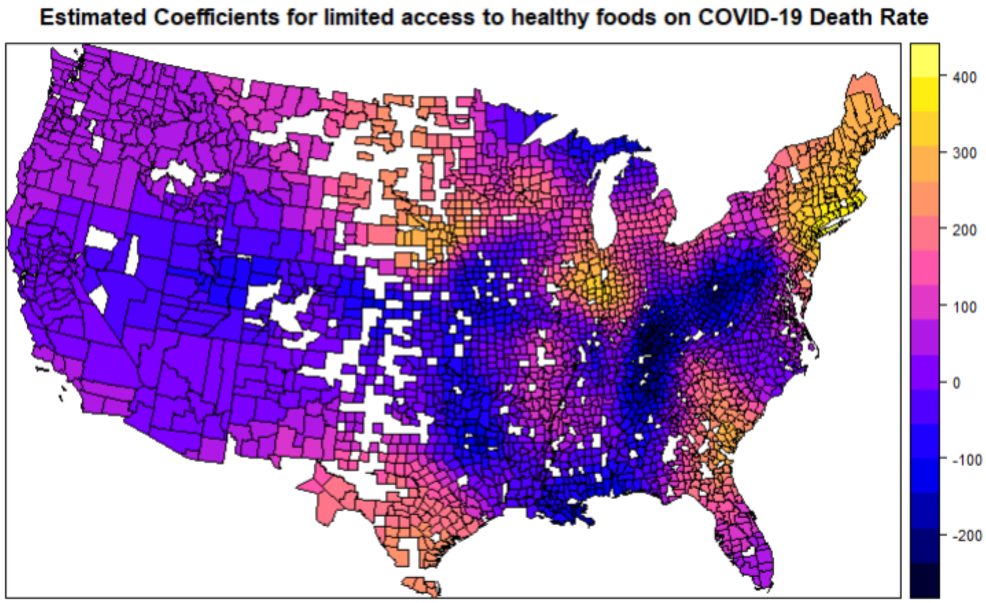
Map 24: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Republicans Voting in Election 2020



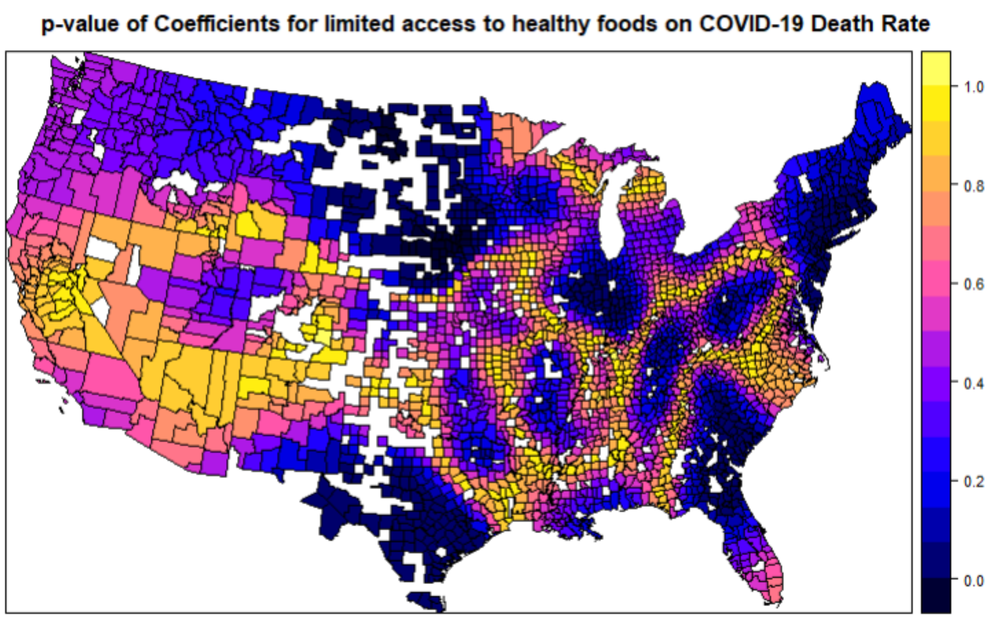
Map 25: GWR Output Map: Spatial Distribution of Coefficients of Percent of Hispanics



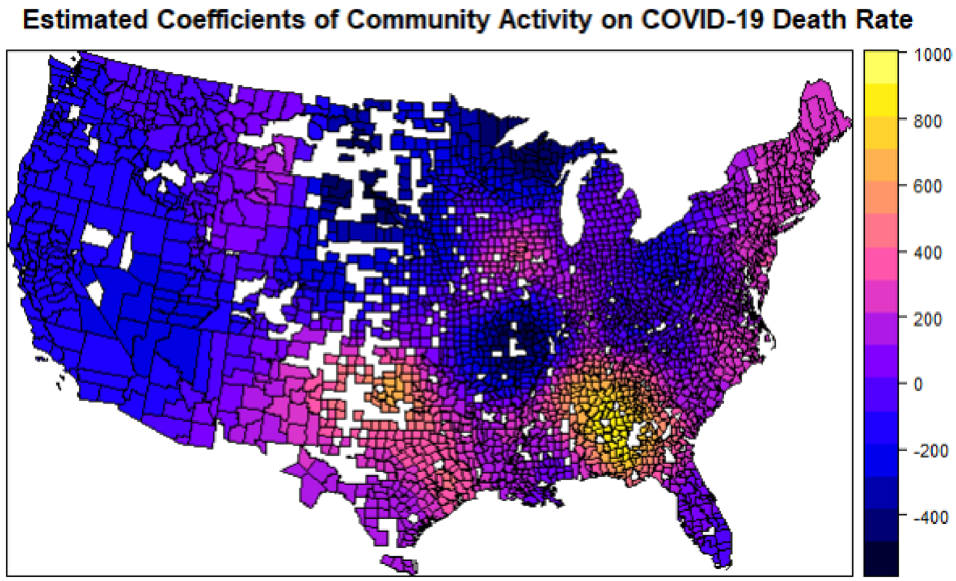
Map 26: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Percent of Hispanics



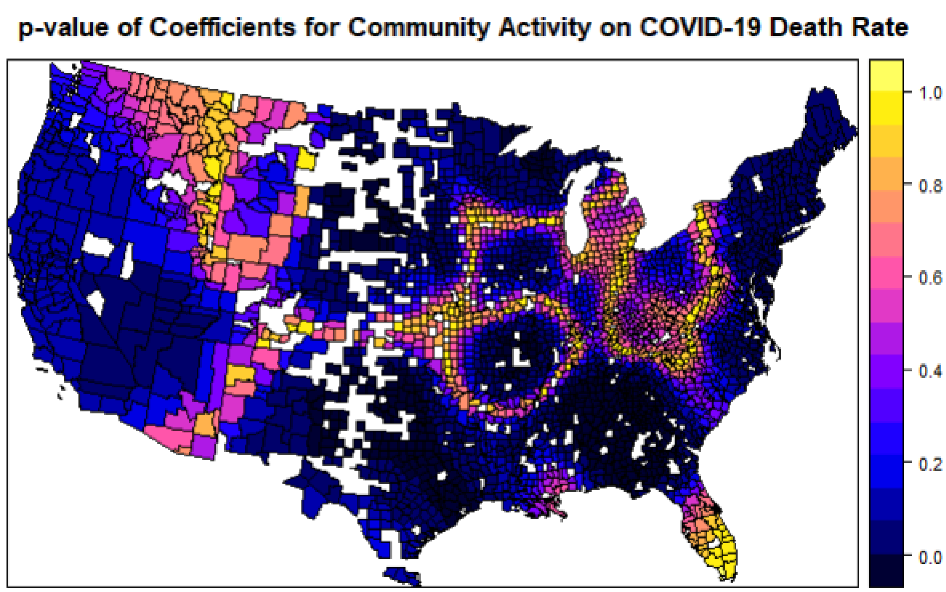
Map 27: GWR Output Map: Spatial Distribution of Coefficients of People with Limited Access to Healthy Foods



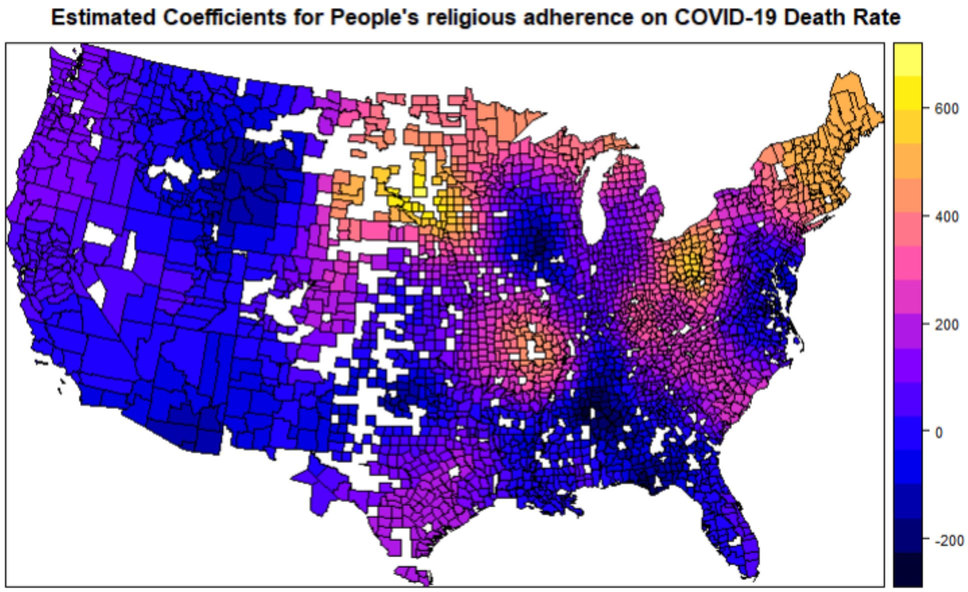
Map 28: GWR Output Map: Spatial Distribution of P-Value of Coefficients of People with Limited Access to Healthy Foods



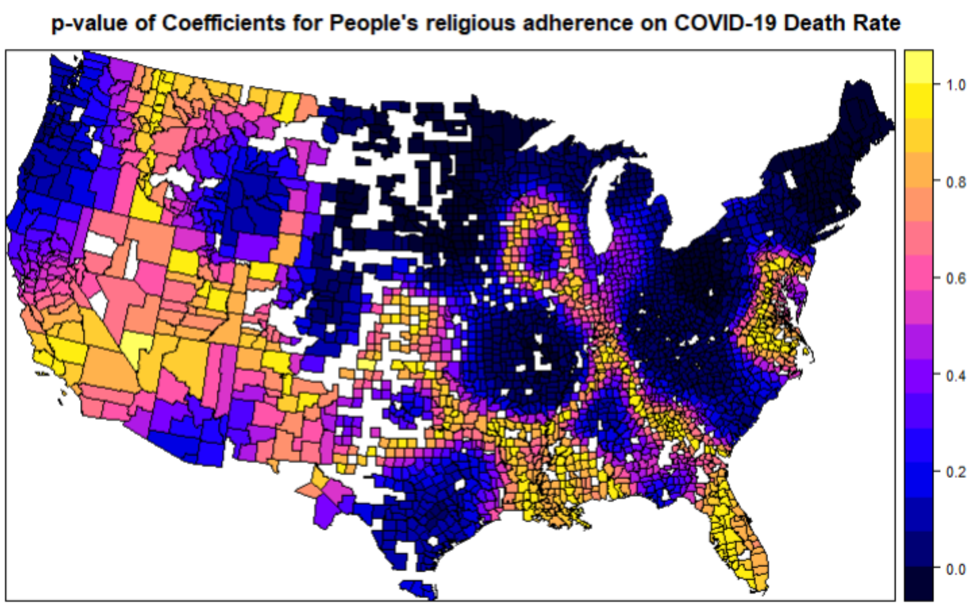
Map 29: GWR Output Map: Spatial Distribution of Coefficients of Community Activity



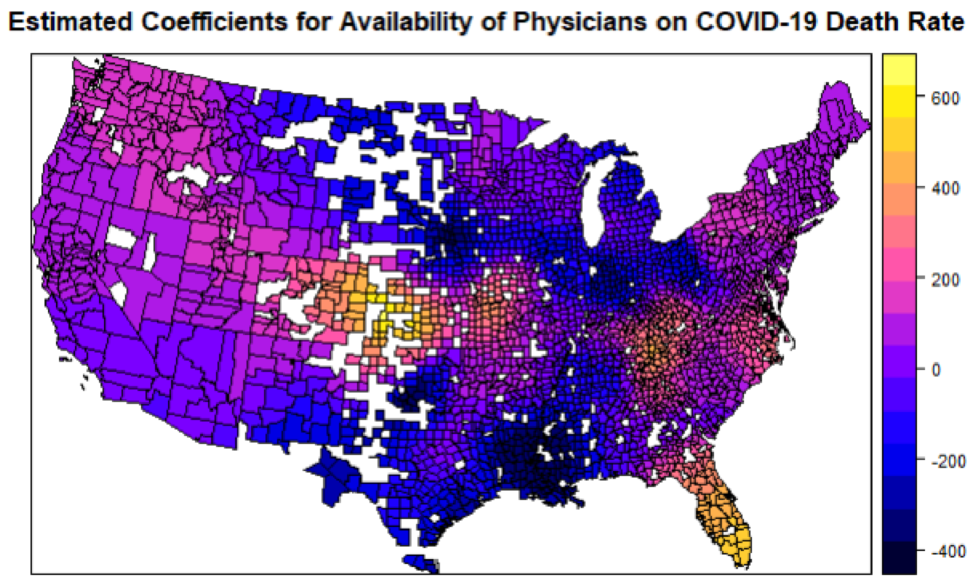
Map 30: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Community Activity



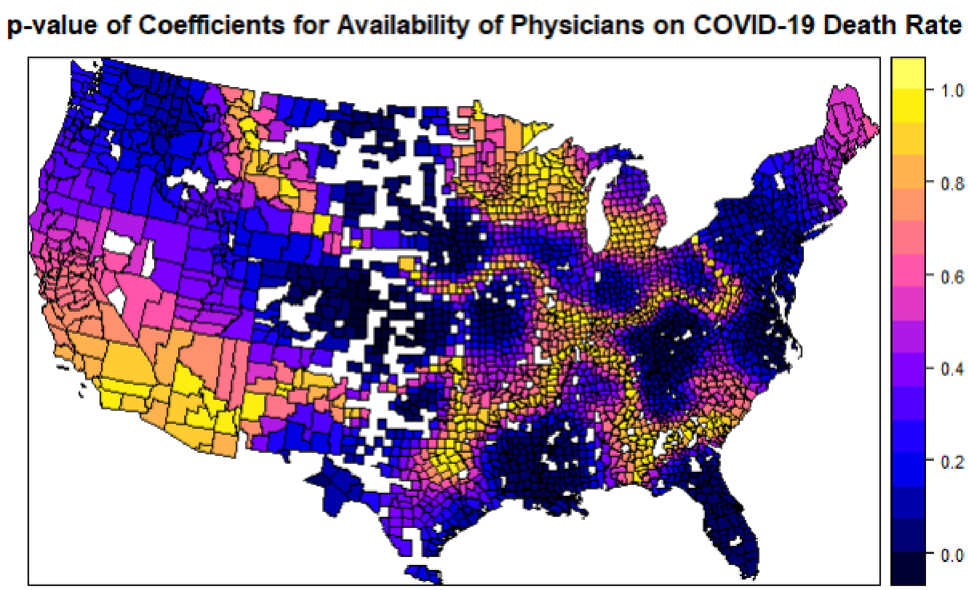
Map 31: GWR Output Map: Spatial Distribution of Coefficients of All Religions Adherence Rate



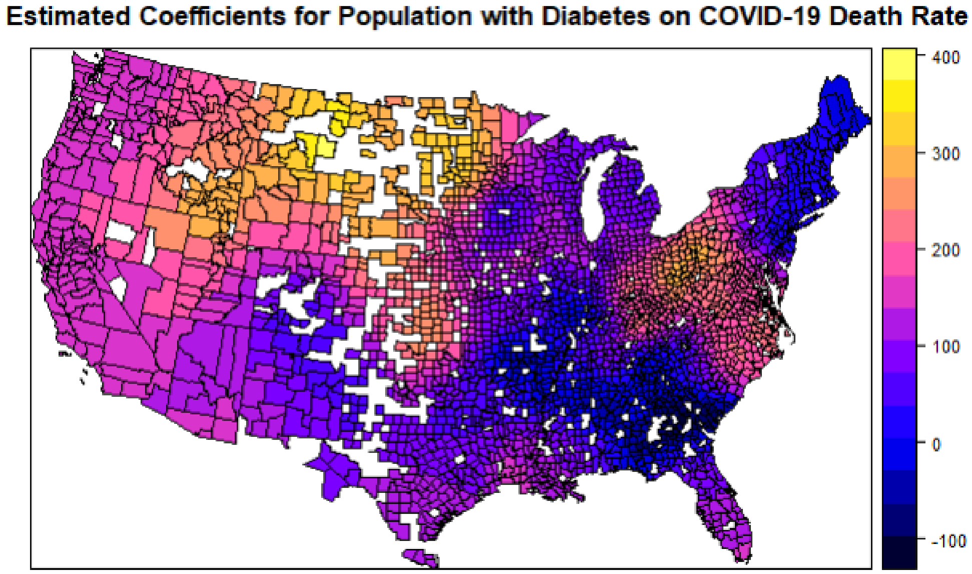
Map 32: GWR Output Map: Spatial Distribution of P-Value of Coefficients of All Religions Adherence Rate



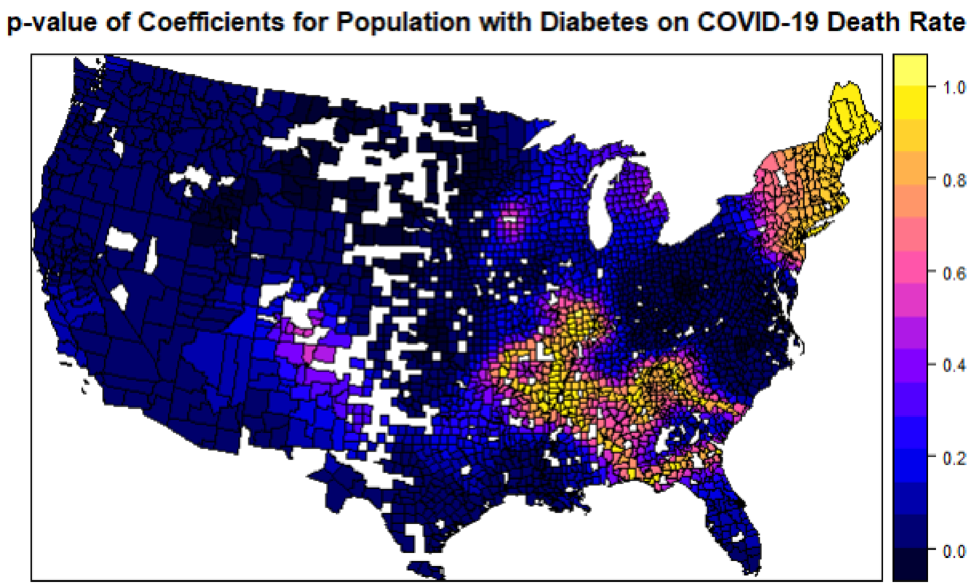
Map 33: GWR Output Map: Spatial Distribution of Coefficients of Availability of Physicians Per 100,000 People



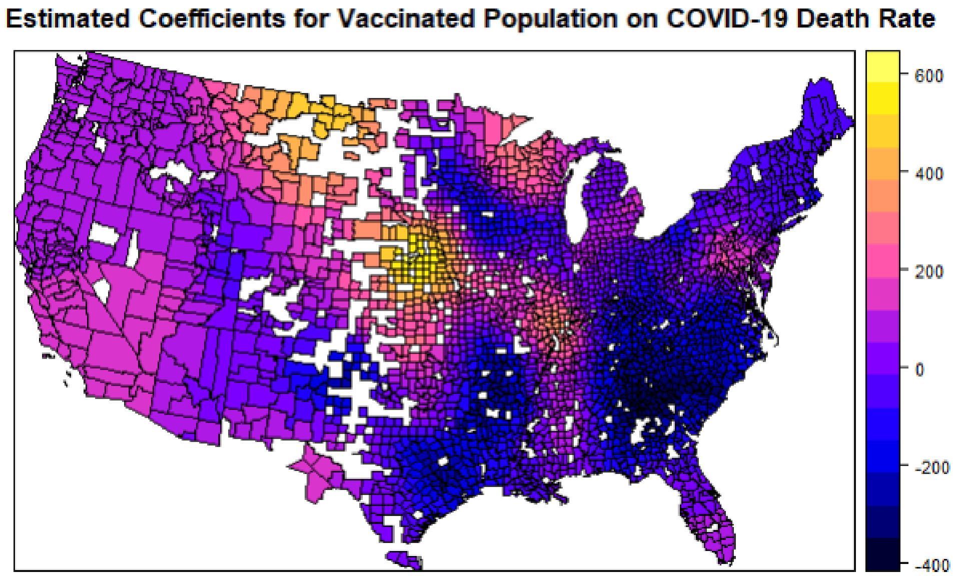
Map 34: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Availability of Physicians Per 100,000 People



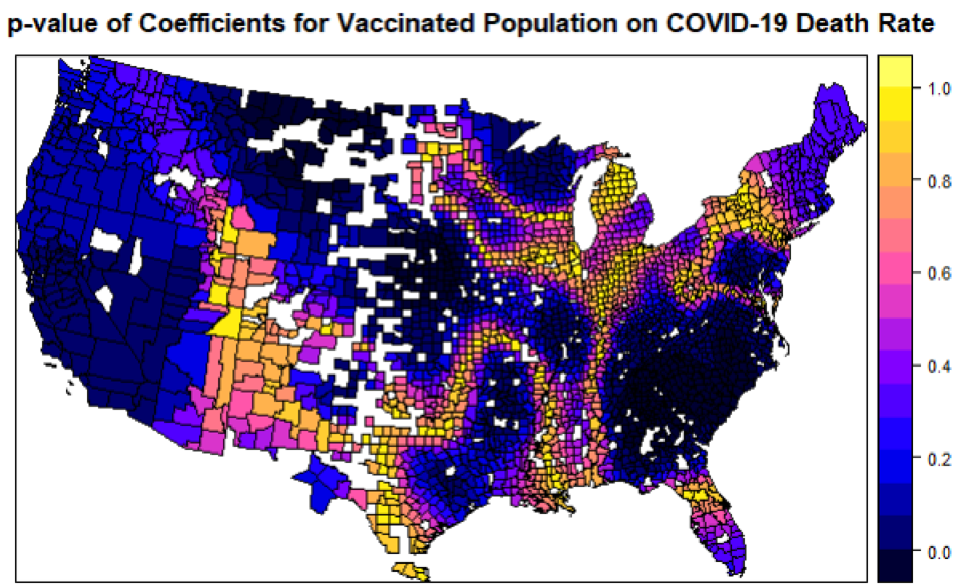
Map 35: GWR Output Map: Spatial Distribution of Coefficients of Percent Diabetics



Map 36: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Percent Diabetics



Map 37: GWR Output Map: Spatial Distribution of Coefficients of Percent of People with Vaccination Series Complete



Map 38: GWR Output Map: Spatial Distribution of P-Value of Coefficients of Percent of People with Vaccination Series Complete

REFERENCES

12.3—Highly Correlated Predictors | STAT 501. (n.d.). PennState: Statistics Online Courses. Retrieved December 13, 2022, from https://online.stat.psu.edu/stat501/lesson/12/12.3

Ahmed, F., Ahmed, N., Pissarides, C., & Stiglitz, J. (2020). Why inequality could spread COVID-19. *The Lancet Public Health*, *5*(5), e240. https://doi.org/10.1016/S2468-2667(20)30085-2

Andersen, L. M., Harden, S. R., Sugg, M. M., Runkle, J. D., & Lundquist, T. E. (2021). Analyzing the spatial determinants of local Covid-19 transmission in the United States. *Science of The Total Environment*, *754*, 142396. https://doi.org/10.1016/j.scitotenv.2020.142396

Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., … Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*(5), Article 5. https://doi.org/10.1038/s41562-020-0884-z

Braveman, P. A., Cubbin, C., Egerter, S., Williams, D. R., & Pamuk, E. (2010). Socioeconomic disparities in health in the United States: What the patterns tell us. *American Journal of Public Health*, *100 Suppl 1*(Suppl 1), S186-196. https://doi.org/10.2105/AJPH.2009.166082

Carrazana, C. (2022, June 2). *Most of the COVID-19 workforce were women of color. What happens now as those jobs end?* The 19th. https://19thnews.org/2022/06/covid-19-workforce-women-of-color-jobs/

CDC/ATSDR Social Vulnerability Index (SVI). (2022, October 27). https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

College, G. and E. S., Hunter. (n.d.). *Chapter 4 Spatial Regression in R | Data Analysis and Visualization with R: Spatial*. Retrieved December 14, 2022, from http://www.geo.hunter.cuny.edu/~ssun/R-Spatial/spregression.html

Credit, K. (2020). Neighbourhood inequity: Exploring the factors underlying racial and ethnic disparities in COVID-19 testing and infection rates using ZIP code data in Chicago and New York. *Regional Science Policy & Practice*, *12*(6), 1249–1271. https://doi.org/10.1111/rsp3.12321

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social Vulnerability to Environmental Hazards*. *Social Science Quarterly*, *84*(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

Cutter, S. L., & Finch, C. (2008). Temporal and spatial changes in social vulnerability to natural hazards. *Proceedings of the National Academy of Sciences*, *105*(7), 2301–2306. https://doi.org/10.1073/pnas.0710375105

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. SAGE.

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*, *8*(1). https://doi.org/10.2202/1547-7355.1792

Gaffney, A. W., Hawks, L., Bor, D. H., Woolhandler, S., Himmelstein, D. U., & McCormick, D. (2020). 18.2 Million Individuals at Increased Risk of Severe COVID-19 Illness Are Un- or Underinsured. *Journal of General Internal Medicine*, *35*(8), 2487–2489. https://doi.org/10.1007/s11606-020-05899-8

GISGeography. (2014, July 15). *Spatial Autocorrelation and Moran's I in GIS*. GIS Geography. https://gisgeography.com/spatial-autocorrelation-moran-i-gis/

Godman, H. (2021, December 1). *Harvard study: Healthy diet associated with lower COVID-19 risk and severity*. Harvard Health. https://www.health.harvard.edu/staying-healthy/harvard-study-healthy-diet-associated-with-lower-covid-19-risk-and-severity

Greater Impact: How Disasters Affect People of Low Socioeconomic Status. (n.d.).

Hallegatte, S. (2013). *An Exploration of the Link between Development, Economic Growth, and Natural Risk* (SSRN Scholarly Paper No. 2251156). https://doi.org/10.2139/ssrn.2251156

Juntunen, L. (2004). *Addressing social vulnerability to hazards* [PhD Thesis]. University of Oregon Eugene.

Karaye, I. M., & Horney, J. A. (2020a). The Impact of Social Vulnerability on COVID-19 in the U.S.: An Analysis of Spatially Varying Relationships. *American Journal of Preventive Medicine*, *59*(3), 317–325. https://doi.org/10.1016/j.amepre.2020.06.006

Karaye, I. M., & Horney, J. A. (2020b). The Impact of Social Vulnerability on COVID-19 in the U.S.: An Analysis of Spatially Varying Relationships. *American Journal of Preventive Medicine*, *59*(3), 317–325. https://doi.org/10.1016/j.amepre.2020.06.006

Karmakar, M., Lantz, P. M., & Tipirneni, R. (2021). Association of Social and Demographic Factors With COVID-19 Incidence and Death Rates in the US. *JAMA Network Open*, *4*(1), e2036462. https://doi.org/10.1001/jamanetworkopen.2020.36462

Khazanchi, R., Beiter, E. R., Gondi, S., Beckman, A. L., Bilinski, A., & Ganguli, I. (2020). County-Level Association of Social Vulnerability with COVID-19 Cases and Deaths in the USA. *Journal of General Internal Medicine*, *35*(9), 2784–2787. https://doi.org/10.1007/s11606-020-05882-3

Lagrange Multiplier—An overview | ScienceDirect Topics. (n.d.). Retrieved December 14, 2022, from https://www.sciencedirect.com/topics/earth-and-planetary-sciences/lagrange-multiplier

Medzerian, D. (2021, February 25). *Education is now a bigger factor than race in desire for COVID-19 vaccine*. USC News. https://news.usc.edu/182848/education-covid-19-vaccine-safety-risks-usc-study/

Moghadas, S. M., Vilches, T. N., Zhang, K., Wells, C. R., Shoukat, A., Singer, B. H., Meyers, L. A., Neuzil, K. M., Langley, J. M., Fitzpatrick, M. C., & Galvani, A. P. (2021). The impact of vaccination on COVID-19 outbreaks in the United States. *MedRxiv*, 2020.11.27.20240051. https://doi.org/10.1101/2020.11.27.20240051

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, *728*, 138884. https://doi.org/10.1016/j.scitotenv.2020.138884

Morrow, B. H. (1999). Identifying and Mapping Community Vulnerability. *Disasters*, *23*(1), 1–18. https://doi.org/10.1111/1467-7717.00102

Parkinson,Cody. (n.d.). *COVID-19, educational attainment, and the impact on American workers: Monthly Labor Review: U.S. Bureau of Labor Statistics*. Retrieved December 13, 2022, from https://www.bls.gov/opub/mlr/2020/beyond-bls/covid-19-educational-attainment-and-the-impact-on-american-workers.htm

Patel, J. A., Nielsen, F. B. H., Badiani, A. A., Assi, S., Unadkat, V. A., Patel, B., Ravindrane, R., & Wardle, H. (2020a). Poverty, inequality and COVID-19: The forgotten vulnerable. *Public Health*, *183*, 110–111. https://doi.org/10.1016/j.puhe.2020.05.006

Patel, J. A., Nielsen, F. B. H., Badiani, A. A., Assi, S., Unadkat, V. A., Patel, B., Ravindrane, R., & Wardle, H. (2020b). Poverty, inequality and COVID-19: The forgotten vulnerable. *Public Health*, *183*, 110–111. https://doi.org/10.1016/j.puhe.2020.05.006

Perry, B. L., Aronson, B., & Pescosolido, B. A. (2021). Pandemic precarity: COVID-19 is exposing and exacerbating inequalities in the American heartland. *Proceedings of the National Academy of Sciences*, *118*(8), e2020685118. https://doi.org/10.1073/pnas.2020685118

Sehgal, N. J., Yue, D., Pope, E., Wang, R. H., & Roby, D. H. (2022). The Association Between COVID-19 Mortality And The County-Level Partisan Divide In The United States. *Health Affairs*, *41*(6), 853–863. https://doi.org/10.1377/hlthaff.2022.00085

Studies consistently find higher Covid death rates among Republicans than Democrats. (n.d.). NBC News. Retrieved December 13, 2022, from https://www.nbcnews.com/health/health-news/covid-death-rates-higher-republicans-democrats-why-rcna50883

Systematic Inequality and American Democracy. (n.d.). *Center for American Progress*. Retrieved December 14, 2022, from https://www.americanprogress.org/article/systematic-inequality-american-democracy/

Wallace, J., Goldsmith-Pinkham, P., & Schwartz, J. (2022). *Excess Death Rates for Republicans and Democrats During the COVID-19 Pandemic* (No. w30512; p. w30512). National Bureau of Economic Research. https://doi.org/10.3386/w30512

Wu, S. (2021, June 5). *Multi-Collinearity in Regression*. Medium. https://towardsdatascience.com/multi-collinearity-in-regression-fe7a2c1467ea

Wu, X., & Zhang, J. (2021). Exploration of spatial-temporal varying impacts on COVID-19 cumulative case in Texas using geographically weighted regression (GWR). *Environmental Science and Pollution Research*, *28*(32), 43732–43746. https://doi.org/10.1007/s11356-021-13653-8

Yellow Horse, A. J., Yang, T.-C., & Huyser, K. R. (2022). Structural Inequalities Established the Architecture for COVID-19 Pandemic Among Native Americans in Arizona: A Geographically Weighted Regression Perspective. *Journal of Racial and Ethnic Health Disparities*, *9*(1), 165–175. https://doi.org/10.1007/s40615-020-00940-2