BIG DATA APPLICATIONS AND CHALLENGES IN GISCIENCE (CASE STUDIES: NATURAL DISASTER AND PUBLIC HEALTH CRISIS MANAGEMENT)

Amir Masoud Forati

University of Wisconsin-Milwaukee

Follow this and additional works at: https://dc.uwm.edu/etd

Part of the Databases and Information Systems Commons, Geographic Information Sciences Commons, and the Public Health Commons

Recommended Citation


https://dc.uwm.edu/etd/3403

This Dissertation is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact scholarlycommunicationteam-group@uwm.edu.
BIG DATA APPLICATIONS AND CHALLENGES IN GISCIENCE (CASE STUDIES: NATURAL DISASTER AND PUBLIC HEALTH CRISIS MANAGEMENT)

by

Amir Masoud Forati

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Geography at The University of Wisconsin-Milwaukee December 2023
ABSTRACT

BIG DATA APPLICATIONS AND CHALLENGES IN GISCIENCE (CASE STUDIES: NATURAL DISASTER AND PUBLIC HEALTH CRISIS MANAGEMENT)

by

Amir Masoud Forati

The University of Wisconsin-Milwaukee, 2023
Under the Supervision of Professor Rina Ghose

This dissertation examines the application and significance of user-generated big data in Geographic Information Science (GiScience), with a focus on managing natural disasters and public health crises. It explores the role of social media data in understanding human-environment interactions and in informing disaster management and public health strategies. A scalable computational framework will be developed to model extensive unstructured geotagged data from social media, facilitating systematic spatiotemporal data analysis. The research investigates how individuals and communities respond to high-impact events like natural disasters and public health emergencies, employing both qualitative and quantitative methods. In particular, it assesses the impact of socio-economic-demographic characteristics and the digital divide on social media engagement during such crises.

In addressing the opioid crisis, the dissertation delves into the spatial dynamics of opioid overdose deaths, utilizing Multiscale Geographically Weighted Regression to discern local versus broader-scale determinants. This analysis foregrounds the necessity for targeted public health responses and the importance of localized data in crafting effective interventions, especially within communities that are ethnically diverse and economically disparate.
Using Hurricane Irma as a case study, this dissertation analyzes social media activity in Florida in September 2017, leveraging Multiscale Geographically Weighted Regression to explore spatial variations in social media discourse, its correlation with damage severity, and the disproportionate impact on racialized communities. It integrates social media data analysis with political-ecological perspectives and spatial analytical techniques to reveal structural inequalities and political power differentials.

The dissertation also tackles the dissemination of false information during the COVID-19 pandemic, examining Twitter activity in the United States from April to July 2020. It identifies misinformation patterns, their origins, and their association with the pandemic's incidence rates. Discourse analysis pinpoints tweets that downplay the pandemic's severity or spread disinformation, while spatial modeling investigates the relationship between social media discourse and disease spread.

By concentrating on the experiences of racialized communities, this research aims to highlight and address the environmental and social injustices they face. It contributes empirical and methodological insights into effective policy formulation, with an emphasis on equitable responses to public health emergencies and natural disasters. This dissertation not only provides a nuanced understanding of crisis responses but also advances GIScience research by incorporating social media data into both traditional and critical analytical frameworks.
This dissertation is dedicated to those who have always believed in me, even when I found it difficult to believe in myself.

To my parents, Zahra and Ali Mohammad for their unwavering love and support throughout my life. Your sacrifices and encouragement have been the bedrock of my perseverance.

To my mentor, Dr. Rina Ghose, for the invaluable guidance and insights that have shaped both this research and my growth as a scholar.

To my partner, Fahimeh, for the endless patience, understanding, and love that has been a constant source of strength.

And to all the individuals and communities who face and overcome adversity every day, your resilience inspires every page of this work.
# TABLE OF CONTENTS

**Chapter 1 Introduction** ........................................................................................................................................... 1
  Research Significance.................................................................................................................................................. 3

**Chapter 2 Precision Epidemiology and Opioid Crisis: Using Geospatial Analyses to Guide Community-Level Responses in Diverse and Segregated Metropolitan Regions** ............. 5
  Literature review......................................................................................................................................................... 12
  Study Site ................................................................................................................................................................. 19
  Data Collection......................................................................................................................................................... 21
  Methodology............................................................................................................................................................... 22
  Results.......................................................................................................................................................................... 39
  Discussion................................................................................................................................................................... 74
  Conclusions ............................................................................................................................................................... 89

**Chapter 3: Analyzing Twitter Generated Covid-19 Misinformation through A Mixed-Method Approach** ........................................................................................................................................... 91
  Literature Review......................................................................................................................................................... 94
  Methodology............................................................................................................................................................... 102
  Results.......................................................................................................................................................................... 104
  Discussion................................................................................................................................................................... 138
  Conclusions ............................................................................................................................................................... 140

**Chapter 4: A Multiscale Geospatial Analysis of Social Media Activity During Hurricane Irma** ........................................................................................................................................... 142
  Literature Review......................................................................................................................................................... 147
  Study Area................................................................................................................................................................. 156
  Data Sources and Methodology.................................................................................................................................. 157
  Results and Discussion.............................................................................................................................................. 161
  Conclusion................................................................................................................................................................. 176

**Chapter 5: Conclusion** ............................................................................................................................................... 178

**Bibliography/Works Cited/References** .................................................................................................................. 181
<table>
<thead>
<tr>
<th>Figure #</th>
<th>Figure title</th>
<th>Page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Maps of MGWR parameter estimate surfaces for the households with kids at preschool age, educational attainment, percentage of persons with disability, and access to healthcare which tend to show little-to-no spatial heterogeneity.</td>
<td>45</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Maps of MGWR parameter estimate surfaces for crime rate and inequality of household income which tend to show regional patterns of spatial heterogeneity.</td>
<td>46</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Map of MGWR parameter estimate surfaces for Naloxone availability which tends to show local patterns of spatial heterogeneity.</td>
<td>47</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Racial segregation map displaying majority racial group in each census tracts</td>
<td>50</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Monthly overall, fentanyl-related, heroin-related, and other opioid-related OODs</td>
<td>59</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Monthly opioid overdose deaths, vertical line indicates the stay at-home order date (March 23, 2020). Blue shaded region indicates 95% confidence interval for BSTS forecast.</td>
<td>60</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Spatial distribution of the underlying OOD risks at the census tracts level from 2017 to 2020 (OOD is expressed in rate of Overdose per Census Tracts Population)</td>
<td>62</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Increment (yes/no) and increment percentage in the numbers of opioid overdose deaths in Milwaukee County, March 23 to December 31, 2020</td>
<td>66</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Spatial patterns of fentanyl-, heroin- and other opioid-related deaths, March 23 to December 31, 2020 (increments relative to pre-intervention numbers)</td>
<td>67</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Overdose deaths and overdose death centrality in Milwaukee County, WI. Map in 10A shows overdose death locations and census tract hotspots between 2017 and 2020. Each point represents an overdose death location. Warmer colors reflect higher numbers of overdose deaths in the census tracts during the study period. Map in 10B shows overdose journeys between 2017 and 2020 and the degree centrality of census tracts defined using directed spatial social network analysis according to the number of edges for each node/census tract. Each line represents an overdose journey; arrows indicate directionality. Census tracts are colored based on network-weighted degree centrality. Warmer colors reflect higher numbers of edges (journey to or from the census tract).</td>
<td>68</td>
</tr>
</tbody>
</table>
Figure 11  Distances traveled in the journey to overdose. Histogram represents the distribution of distances (km) between residential addresses and overdose death locations.

Figure 12  Hub and authority communities for overdose journeys in Milwaukee, WI. Spatial social network analysis (HITS algorithm) was used to identify hubs (census tracts that are focal points of geographically discordant overdoses) and authorities (communities of residence from which journeys to overdose commonly begin) for overdose journeys during the study periods. Census tracts assigned scores based on hub centrality are depicted in the map in 3A. Warmer colors indicate census tracts with higher hub scores. Census tracts assigned scores based on authority centrality are depicted in the map in 3B. Warmer colors indicate census tracts with higher authority scores.

Figure 13  Hotspots for geographically discordant overdose deaths in Milwaukee, WI. Edge-scan analysis was used to define network interaction hotspots for discordant overdose deaths based on aggregate data over the study timeframe (4A). Warmer colors indicate census tracts that contributed most to overdose journeys. Space-time cube analysis was used to define emerging hotspots categorized as new, consecutive, and sporadic according to the temporal profiles of overdose deaths (4B). Note that “no pattern detected” does not imply the absence of overdoses but rather the absence of a temporal trend in overdoses.

Figure 14  Distribution of Covid-19 related tweets in May and June 2020, throughout conterminous United States

Figure 15  Distribution of Misinformation Tweets in May and June 2020, throughout conterminous United States

Figure 16  Map of MGWR parameter estimate surfaces for Tweet rate.

Figure 17  Map of MGWR parameter estimate surfaces for INFA score.

Figure 18  MGWR R2

Figure 19  Proportion of False Information by Category (a) and Topical Theme (b)

Figure 20  Distribution of Covid-19 related Tweets with False information about Covid-19 throughout United states

Figure 21  The GWPCA CPTV and the covid-19 pandemic myths in social media with the largest loading for GWPC1 and GWPC2.

Figure 22  Daily average wind speed vs. number of tweets

Figure 23  Tweets’ topics categories during hurricane Irma
Figure 24  Geographic distributions of tweets in Pinellas County in conjunction with calculated flood rate 169
Figure 25  Multiscale Geographically Weighted Regression Local R squared. 172
Figure 26  Map of Multiscale Geographically Weighted Regression parameter estimate surface for Tweet rate which tends to show regional patterns of spatial heterogeneity. 174
Figure 27  Maps of Multiscale Geographically Weighted Regression parameter estimate surfaces for a) INFA score which tend to show regional patterns of spatial heterogeneity b) SE score which tend to show local patterns of spatial heterogeneity c) Intercept score which tend to show local patterns of spatial heterogeneity. 176
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table #</th>
<th>Table title</th>
<th>Page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Description of the explanatory variables included in the study.</td>
<td>40</td>
</tr>
<tr>
<td>Table 2</td>
<td>Results from the ordinary least squares regression.</td>
<td>41</td>
</tr>
<tr>
<td>Table 3</td>
<td>MGWR explanatory variable bandwidth and significance (The bandwidth is expressed as the number of neighboring census tracts)</td>
<td>42</td>
</tr>
<tr>
<td>Table 4</td>
<td>Selected explanatory variables included in the study for each region.</td>
<td>51</td>
</tr>
<tr>
<td>Table 5</td>
<td>Model fit metrics global regression, and multiscale geographically weighted regression (MGWR)</td>
<td>52</td>
</tr>
<tr>
<td>Table 6</td>
<td>Regional determinant of opioid overdose death among different races/ethnicities (The bandwidth is expressed as the number of neighboring census tracts)</td>
<td>53</td>
</tr>
<tr>
<td>Table 7</td>
<td>Opioid Overdose Deaths in Milwaukee County disaggregated by mode, drug evident in the toxicology report, cause, gender, race, geographic discordance, and age.</td>
<td>56</td>
</tr>
<tr>
<td>Table 8</td>
<td>Socioeconomic characteristics of focal points and hubs of the journey to overdoses</td>
<td>71</td>
</tr>
<tr>
<td>Table 9</td>
<td>Journey to overdose in Milwaukee County disaggregated by mode, drug evident in the toxicology report, cause, gender, race, geographic discordance, and age (significance: 99%<em><strong>, 95%</strong>, 90%</em>)</td>
<td>73</td>
</tr>
<tr>
<td>Table 10</td>
<td>Top ten counties with the highest number of COVID-19 related geotagged tweets in May and June</td>
<td>104</td>
</tr>
<tr>
<td>Table 11</td>
<td>Most popular hashtags among tweets sharing misinformation</td>
<td>106</td>
</tr>
<tr>
<td>Table 12</td>
<td>Top 10 States Leading in Misinformation Tweets</td>
<td>109</td>
</tr>
<tr>
<td>Table 13</td>
<td>Top 10 Counties with the highest number of tweets sharing misinformation per capita.</td>
<td>111</td>
</tr>
<tr>
<td>Table 14</td>
<td>Partial Distance Correlation Coefficients</td>
<td>114</td>
</tr>
<tr>
<td>Table 15</td>
<td>Global Regression Results</td>
<td>116</td>
</tr>
<tr>
<td>Table 16</td>
<td>MGWR results (The bandwidth is expressed as the number of neighboring census tracts)</td>
<td>117</td>
</tr>
<tr>
<td>Table 17</td>
<td>False Information Categories and Definitions.</td>
<td>122</td>
</tr>
<tr>
<td>Table 18</td>
<td>Themes and Myths within topical content analysis framework Information Categories and Definitions.</td>
<td>123</td>
</tr>
<tr>
<td>Table 19</td>
<td>Definition of Myth categories.</td>
<td>124</td>
</tr>
<tr>
<td>Table 20</td>
<td>Descriptive statistics of Covid-19 pandemic myths</td>
<td>132</td>
</tr>
<tr>
<td>Table 21</td>
<td>The results of global PCA</td>
<td>134</td>
</tr>
<tr>
<td>Table 22</td>
<td>Descriptive statistics for collected tweets in the state of Florida and Pinellas County</td>
<td>159</td>
</tr>
</tbody>
</table>
Table 22  Tweets’ topics in different phases of hurricane and example of tweets  166
Table 23  Comparison of demographic and socioeconomic characteristics between all census tracts in Pinellas County versus the census tracts with tweets (weighted based on number of tweets)  170
Table 24  Ordinary Least Square results  171
Table 25  Multiscale Geographically Weighted Regression results (The bandwidth is expressed as the number of neighboring census tracts)  173
Chapter 1 Introduction

The increasing frequency and intensity of public health crises and natural disasters have highlighted the need for more effective emergency response planning. To improve response efforts, researchers have turned to big data analytics, particularly in the realm of location-based data (Raento et al., 2009; Lane et al., 2010). The use of big location-based data offers many advantages in the context of emergency response planning (Cao et al., 2015; Shaw et al., 2016). It can provide real-time information on the location and magnitude of crises, as well as help to identify vulnerable populations and critical infrastructure (Sui and Goodchild 2011). By leveraging these insights, policymakers and emergency responders can make more informed decisions and allocate resources more efficiently. However, there are also significant challenges associated with the use of big data in emergency response planning.

This dissertation examines the benefits and challenges of using big location-based data to analyze high-impact public crises and natural disasters. Three specific case studies are examined: the opioid epidemic in Milwaukee County, the COVID-19 pandemic in United States, and Hurricane Irma in Pinellas County. By doing so, it aims to contribute to the broader discourse on how data analytics can be used to improve public safety and promote societal benefits. The analysis of these cases can help to advance emergency response planning efforts at the local and national levels, as well as provide broader societal benefits.

The dissertation follows the Three Paper/Publication model at UWM department of geography, and portions of this dissertation have been published in high impact research journals: Drug and Alcohol Dependence, Journal of Urban Health, International Journal of Disaster Risk Reduction, and Applied Geography.
In Chapter 2, we delve into the potential of utilizing extensive location-based datasets to discern community traits that have a bearing on opioid overdoses and recovery processes. The primary objective is to develop a nuanced, outcome-sensitive model that transcends the limitations of generic "one-size-fits-all" community strategies and leverages precision epidemiology coupled with data-driven community involvement. This approach is designed to uncover the complex dynamics affecting overdose fatalities and to tailor the opioid crisis response more effectively at multiple scales. The chapter delineates the methods employed to assess the effectiveness of big data in capturing these community characteristics and discusses the extent to which this analysis can inform support organizations, health agencies, and policy decision-makers. Ultimately, the research presented in this chapter is aimed at enhancing the precision of community interventions in order to reduce the incidence of opioid-related deaths.

Chapter 3 applies a mixed-method framework to study social media misinformation during the COVID-19 pandemic in the USA. The research investigates COVID-19 related Twitter activity in May, June, and July 2020 to examine the origin and nature of disinformation and its relationship with the COVID-19 incidence rate at the state and county level. The study provides significant empirical and methodological insights into COVID-19 misinformation.

In Chapter 4, the research examines Twitter data in conjunction with other public datasets to examine the effects of Hurricane Irma upon affected communities in Pinellas County. The study focuses on the impacts of Hurricane Irma among racial minorities in impoverished neighborhoods to help address inequities in policy responses. The research findings can assist in improved decision-making required for disaster preparedness and emergency response by providing valuable and updated information for supporting critical tasks and activities.
The overall objective of this study is to create greater empirical and methodological insights to address the environmental injustices and public health crises experienced by racialized communities, so that effective policy responses can be formulated. The research findings will provide insights on ongoing research on natural disasters and public health crisis mitigation efforts and can be utilized for effective public policy formulation. Methodologically, the study will advance the research on incorporating location-based data into GIScience research.

Research Significance

This dissertation continues ongoing efforts to take advantage of valuable and updated social media information to handle the societal issues and their applications in decisions required for disaster management (de Albuquerque et al. 2019, Taylor et al. 2012) and public health. By leveraging the Twitter dataset for geospatial processes, we will be able to augment traditional data sources and gain a better understanding of disaster damages and impacted communities. We will also address the dearth of spatial knowledge on how, when, and where people post misinformation on health-related issues. The detailed spatially concentrated Social Media Analysis that proceeds from this study will lay the foundation for future work that incorporates social media into traditional data sources. A thorough investigation of the digital divide and approaches to tackle it will allow future research to understand where user-generated data may provide only selective representations of any issue and that there will always be people and communities that are missing and should be addressed in any social media-based research.

Building on previous work, such as the analysis by Forati and Ghose (2021), which demonstrated how Twitter discussions can aid in predicting COVID-19 trends in the absence of
other indicators, this dissertation will examine the current discourse on COVID-19 on Twitter. Although the immediacy of the pandemic's onset has passed, discussions on Twitter remain a valuable barometer of public sentiment and misinformation. This research will analyze the discourse related to resistance to containment measures, public awareness, and access to reliable information. It will contribute an integrated assessment of misinformation through a novel critical theoretical lens, linking race and ethnicity, socio-political and economic factors, and the propagation of falsehoods across various media platforms. The findings will shed light on the underlying factors contributing to health disparities, particularly in minority populations during the pandemic.
Over the past 16 years, the United States has experienced a dramatic increase in opioid overdose deaths (OOD) and other opioid use disorders. Such disorders have been fueled by the widespread prescription of opioids, but also by the availability of illicit opioid drugs. Consequently, the annual number of deaths from prescription opioids quintupled over the period from 2000 to 2017 (Hedegaard et al., 2020). While the annual number of deaths from prescription opioids remained relatively stable between 2011 and 2017, overdose deaths from illicit opioids (including heroin and synthetic opioids such as fentanyl) increased nine-fold, driven in part by a growing number of people whose use started with prescription opioids (Scholl et al., 2019). A slight decline in OOD was noted in 2018, only to be followed by a resurgence in OOD, likely attributable to factors surrounding the COVID-19 pandemic (Manchikanti et al., 2021; Ghose et al., 2022). This resurgence has continued with a projection of more than 110,000 overdose deaths in the U.S. in 2022, an increase of 11% from 2021 (CDC, 2023). The epidemic is being fueled by synthetic opioids, primarily fentanyl. Disadvantaged and marginalized communities have been hit particularly hard (Forati et al., 2021). High overdose death rates are associated with community-level indicators of socioeconomic duress, such as

---

1 The content of this chapter has been published in the following articles, coauthored with Prof. Ghose and Prof. Mantsch:

neighborhood instability, crime rate, and inequality of household income (Altekruse et al., 2020; Van Draanen et al., 2020; Forati et al., 2021). As we struggle to respond to the overdose crisis, it has become clear that new strategies are needed to better define the problem and its underlying causes.

The opioid crisis has affected both rural and urban areas but is most acute in regions experiencing high levels of poverty and marginalization, where residents have low access to health care. Cities that are sites of great socio-spatial inequalities have experienced high rates of OODs (Persmark et al., 2020). However, studies show that OODs are not uniformly distributed within cities, rather they tend to be spatially clustered in neighborhoods experiencing concentrated poverty, marginalization and deprivation (Paulozzi et al., 2008; Havens et al., 201; Monnat et al., 2016; Cicero et al., 2014; Rigg et al., 2015; Lenardson et al., 2016). Further, studies report that the current overdose crisis is fueled by numerous demographic and socioeconomic factors (Volkow et al., 2019; Blanco et al., 2019), where socio-spatial inequities influence substance use disorder. As Sadler and Furr Holden (2019) note, built environmental characteristics cause OODs to cluster by way of a deprivation amplification effect. An effective response will require an integrated, data-guided approach that involves health care professionals, policymakers, the justice system, and a diverse array of community partners and organizations (Chandler et al., 2020). The spatial analysis, modeling, and mapping capabilities of Geographic Information Systems (GIS) offer a powerful approach to study the opioid crisis (Mennis et al., 2012; Nesoff et al., 2020; Donnelly et al., 2020). This approach has enabled researchers to examine the effectiveness of local policy responses (Sadler et al., 2019; Schneider et al., 2020).
As the opioid crisis worsens, a comprehensive and granular framework to guide initiatives at the community level is desperately needed. Since the influence of key factors varies with geographic scales and across communities, relationships that are evident at more global scales (e.g., state) may not appear at the local scale (e.g., neighborhood) (Pear et al., 2019). As a result, resources and services may be misdirected and opportunities to identify effective community-level solutions may be missed. Thus, the selection of an appropriate scale of analysis is fundamental to meaningful geographic inquiries into public health crises. Policy responses must also be sensitive to such variations across scales. This is especially true in ethnically diverse and racially segregated cities that show pronounced racial disparities in socioeconomic and health conditions. While the general relationship between socioeconomic status and opioid overdose has been well-documented (Pear et al., 2019; Galea et al., 2003; Galea et al., 2005; Boardman et al., 2001; Laditka et al., 2009; Pullen et al., 2014; Tamayo-Sarver et al., 2003; Goyal et al., 2015; Groenewald et al., 2018; Harrison et al., 2018; Smith et al., 2017; Joynt et al., 2013; Hollingsworth et al., 2017; Cordes et al., 2018), there is a clear need for multiscale geospatial analysis and modeling of relationships and their spatial variations across communities.

This dissertation makes significant contributions to the study of the opioid crisis. First, I am attentive to the significance of geographic scale and recognize that a complex, diverse urban environment must be examined through a fine-grained multi-scalar approach (Fotheringham et al., 2017). Therefore, to explicate Milwaukee’s opioid crisis I examine it at the county-wide scale, municipal scale, census tract scale and neighborhood scale. I employ Multiscale Geographically Weighted Regression (MGWR) for modeling, an innovative
methodological approach that explains how geospatial patterns vary across scales, enabling us to examine the differential influence of factors globally, regionally, and locally (Oshan et al, 2019). Such a multi-scalar approach is relatively new to opioid studies, which tend to focus on a single geographical scale (state level or county level) as the unit of analysis. Second, to further examine known and discover new OOD determinants, I compiled a comprehensive dataset of 225 candidate variables, taking into account the various community level factors that shape OODs. I was particularly attentive to the influence of racial segregation and structural inequalities on OODs. Third, to analyze the opioid crisis in the context of racial segregation and associated structural inequalities, I divide Milwaukee County into three regions based on the racial composition of census tracts. Consequently, I examine spatial variations in OODs and policy responses across communities specified by racial/ethnic identities. Our findings show that policy interventions had differential impacts on communities based on race/ethnicity. The study therefore provides new empirical and methodological insights that can guide future research and community-targeted practices, policies, and interventions.

On the other hand, the increase in opioid overdose deaths (OODs) during the pandemic has been attributed to various factors: reduced access to interventions, increased levels of stress due to isolation and loss of mental health support, and changes in the types/combinations and purity of drugs and patterns of drug use. The impacts have severely affected racial/ethnic minority communities, lowering life expectancies in Black and Hispanic Americans. The National Center for Health Statistics notes that the drop in life expectancy among Hispanic Americans is 3 years, Black Americans is 2.9 years, while that of White population is 1.2 years. Simultaneously, the rise in OODs among Black Americans (2013-2018) is
outpacing that of White Americans (Chandler et al., 2020). To understand the effects of the pandemic on the opioid crisis, studies have primarily examined changes in OODs over time. However, the locational or spatial variations in OODs have not been addressed.

This dissertation offers an innovative methodological approach to examining the opioid crisis during the COVID-19 pandemic. It is attentive to place based differences and racialized health disparities and offers new empirical insights for effective policy responses. To assess the impacts of pandemic, I examined OOD variations over time (January 1, 2017, to December 31, 2020) and space (census tracts within Milwaukee County). First, I conducted interrupted time-series analysis, which uses autocorrelational analysis to address seasonality. Next, I examined the spatial changes over time in OODs through spatial modeling in Geographic Information Systems (GIS), as it offers a powerful analytical approach to the study of the opioid crisis (Nesoff et al., 2020; Donnelly et al., 2020; Schneider et al., 2020). Our georeferenced OOD data, used in combination with Census data in GIS, enabled us to examine the demographic shifts in OODs over time and place.

Many drug overdose deaths are geographically discordant (the individuals who die from drug overdoses do not reside in the locale where the overdose occurs). For example, in Milwaukee County, Wisconsin, 26.72% of drug overdose deaths have involved individuals from neighborhoods/communities that differ from those in which the fatal overdose occurs (Ghose et al., 2022). Understanding factors associated with discordant overdoses should provide insight into determinants of overdose risk and therefore support and guide prevention and harm reduction efforts. Indeed, prior work has found that geographic discordance between residence and overdose location increases relapse to prescription opioid misuse, anxiety, and
incarceration in those with opioid use disorders (Oser and Harp, 2017). However, geographic discordance in overdose deaths has not been investigated.

The recognition that there is geographic discordance in many overdose deaths illustrates that for every overdose, there is a journey. While each journey is unique, much can be learned about factors that contribute to overdose risk by understanding common characteristics of those whose overdoses are geographically discordant, their community of origin/export, the overdose destination/import community, as well as the distance traveled from a person’s residence to the overdose location.

In the field of criminology, the distance traveled from an offender’s residence to the crime scene is called the “journey to crime” (Rengert, 2002; Townsley & Sidebottom, 2010). Johnson et al. (2013) studied the journey to crime for illegal drug buyers, highlighting the impact of race and neighborhood characteristics on the distance traveled to drug purchase arrests. Donnelly et al. (2021) studied racial differences in journeys to crime in opioid possession-related arrests, demonstrating how racial and spatial differences can reveal persistent disparities in drug law enforcement. Inspired by criminology studies, this study examines the "journey to overdose," defined as the difference between the drug overdose decedent’s residence and overdose location. Each journey to overdose is complex and may involve multiple stops, backtracking, or loops around areas before the incident (Ackerman & Rossmo, 2015; Johnson et al., 2013). Although a journey can be most easily conceptualized as a distance traveled (see, e.g., Rengert, 2002), journeys can be defined according to geographic discordance in the demographics as well as socioeconomic and cultural contexts between the community of origin/residence and that in which overdose occurs.
Overdose journeys can be studied using network analysis. Substance use occurs in the context of social networks intertwined with the geographic landscape (Fischer, 1982; Andris, 2016). A spatial social network is a collection of nodes connected by edges or agent-based connections between people or institutions embedded in geographic space (Wasserman and Faust, 1994; Ye & Andris, 2021). In geographical network research, nodes are used to represent spatial objects (e.g., neighborhoods), and the edges are used to represent connections. Such analyses have allowed researchers to identify relationships and study them individually or in aggregate within complex networks at varying scales (Uitermark & Van Meeteren 2021).

In this dissertation I propose a novel framework to understand the journey to overdose through spatial social network analysis. The framework will allow us to (i) pinpoint focal points of geographically discordant overdose deaths (ii) unveil hotspots of discordant (e.g., imported) drug overdoses, and (iii) investigate variables that differentiate discordant from non-discordant overdose deaths. I apply this framework to study Milwaukee County, a diverse and segregated metropolitan area in southeastern Wisconsin that has experienced a steady increase in overdose deaths. Prior research has found that overdose death rates vary across diverse communities in Milwaukee County and that many overdose deaths in the county are geographically discordant (Forati et al., 2021; Ghose et al., 2022). Using Milwaukee as a case study, this dissertation is the first, to my knowledge, to apply spatial social network analysis to study the journey to drug overdose, thus providing proof of the principle that this approach can be used to understand and guide community responses to the ongoing drug overdose epidemic.
Literature review

North America is in the middle of an epidemic of opioid abuse. Opioid overdose mortality rates in the United States have increased more than 500 percent since 2000, leading to more than 48,000 deaths in 2017 alone (Centers for Disease Control and Prevention, 2020). There is substantial variation in rates of prescription opioid overdose (POD) between and within states over time (Ruhm, 2017). In urbanized areas, POD rates rose 62% between 1999 and 2004. Metropolitan county rates rose 51%, while non-metropolitan county rates rose 159%, so by 2004, metropolitan and non-metropolitan POD rates were roughly equalized. Therefore, fatal drug overdoses are not a predominantly urban phenomenon since 2004 (Paulozzi and Xi, 2008), and predominantly rural states such as West Virginia, Pennsylvania, Ohio, and New Hampshire have some of the country's highest POD mortality rates (Rudd et al., 2016).

In terms of rural/urban POD rates, national-scale studies have found contradictory results. Some researchers contend that POD rates are higher in rural areas. Paulozzi and Xi (2008) contend that fatal drug overdoses are no longer a predominantly urban phenomenon, and they suggested that deaths involving opioid analgesics was increasing more rapidly in rural areas. Several studies on youth found that rural adolescents were more likely than urban adolescents to have used prescription drugs nonmedically (Havens et al., 2011; Monnat and Rigg, 2016). Cicero et al. (2014) found similar results and argue that the use of heroin has shifted from an inner-city, minority-centered issue to one with a more widespread geographical distribution, affecting predominantly white men and women living outside major urban centers in their late 20s. In contrast, other studies find higher rates in urban areas or no significant
difference. Rigg and Monnat (2015) found that urban adults were more likely to participate in prescription opioid abuse relative to rural adults due to their increased use of other substances, including alcohol, cannabis, and other illegal and prescription drugs. Lenardon et al. (2016) suggest that the prevalence of past-year opioid usage remains much lower in rural counties than in urban counties. However, rural users exhibit several socio-economic disadvantages, which can adversely affect their ability to obtain care and recover. Literature contradictions on rural/urban differences in PODs and mortality may exist because national patterns obscure significant regional and inter-state differences (Pear et al., 2019).

Socioeconomic factors (e.g., high unemployment, poverty, and low education) have been disproportionately affected by drug use over time. Nevertheless, the relationship between Socioeconomic status and opioid overdose has not been extensively studied. Boardman et al. (2001) investigated the neighborhood disadvantage, tension, and the probability of drug use in Detroit City, and their findings indicate that the association between neighborhood disadvantage and drug use is the most among lower-income individuals. Galea et al. (2003) provided a framework to study how structural factors, social norms and attitudes, disadvantages, and physical environment features affect drug use behavior. Galea et al. (2005) proposes a framework to study the characteristics of urban areas that may be associated with drug use and misuse. They list area-level disadvantage, collective efficacy and its relationship with homicide and violence, the quality of the built environment, residential segregation, population density, social norms and attitudes, public and non-public transportation, and physical availability of health and social services as characteristics of the urban environment that may influence drug use and misuse. As a result, communities with a higher concentration
of economic stress factors such as poverty, unemployment, and low educational attainment may experience higher rates of opioid overdose as residents abuse opioids to treat chronic stress resulting from direct or indirect exposure to economic deprivation and the consequent symptoms of depression and anxiety (Boardman et al., 2001; Pear et al., 2019). The effects of exposure to economic stressors, such as complications of mental wellbeing or drug misuse, may be compounded in areas where at-risk communities lack affordable care and social services, overdose treatment programs, and public transit (Laditka et al., 2009; Pullen and Oser, 2014).

Racial residential segregation is a central cause of racial health disparities. The physical division of races by enforced residency in certain places is a systemic method of segregation intended to shield one race from social contact with other ones. Despite the lack of enabling legal legislation, the degree of residential segregation in the United States remains extremely high (Williams and Collins, 2016; Galea et al., 2005; Mayberry et al., 2000). Residential segregation plays a significant role in racial inequalities by determining access to educational and employment opportunities and health-related resources (Williams and Collins, 2016). People living in segregated neighborhoods may have disproportionate exposure, susceptibility, and vulnerability to economic and social inequality, toxic substances, and unsafe conditions. Racial segregation may also influence wellbeing by affecting individual health habits, access to health care services. (Williams and Collins, 2016; Flint and Novotny, 1997).

Similar to rural/urban disparities in PODs and mortality, findings from the several studies of prescription drug use may not be representative of overall national trends. Regional studies show that minorities are less likely to receive opioids than whites on a regular basis (Tamayo-Sarver et al., 2003; Goyal et al., 2015; Groenewald et al., 2018), while results from
national data sets do not clearly demonstrate such disparities (Harrison et al., 2018; Smith et al., 2017), these inconsistencies may occur as national trends obscure major regional and inter-state variations.

Joynt et al. (2013) examined the prescribing of opioids to patients with moderate to severe pain using logistic regression to investigate the association between prescribing opioids, socio-economic status, and race. Socio-economic status was determined based on income, percent poverty, and educational level within a patient's zip code and age, sex, pain level, injury-status, frequency of emergency visits, hospital type, and region used to adjust the model. They found that patients from lower socio-economic status regions were less likely to receive opioids for comparable pain levels than those from more affluent areas. Black and Hispanic patients, regardless of the social background, were also less likely to receive prescriptions for comparable pain levels than whites. The generalizability of these findings is, however, limited due to the measure of socio-economic status used. Since they only applied selected dimensions, such as income and education, to reflect socio-economic status, and therefore may lead to biased conclusions. Hollingsworth et al. (2017) assessed how the economic conditions of local communities relate to opioid overdose incidents, considering data from multiple states over several years. Their analysis indicated that a one-percentage point rise in the unemployment rate of a county corresponds to an increase of 0.19 in the opioid death rate per 100,000 people, which represents a 3.6% increase relative to the average death rate. Similarly, they found that the rate of emergency department (E.D.) visits for opioid overdoses per 100,000 people increases by 0.95, a 7.0% rise compared to the average visit rate. It is important
to note, however, that their study did not factor in the potential for spatial autocorrelation, which could influence opioid overdose rates among geographically adjacent communities.

Cordes (2018) investigated the opioid epidemic in North Carolina along three axes: space, time, and drug type; nevertheless, they incorporated a small number of demographic and socio-economic variables in their Ordinary Least Square (OLS) regression model, failing to include some important factors such as marital status, alcohol consumption, drug availability, unhealthy behavior prevalence, crime rate, digital divide, and neighborhood stability to name a few. On the other hand, OLS assumes that the observations are independent and constant across the study area. The error terms are not correlated and do not consider spatial dependence (Anselin and Arribas-Bel, 2013). However, in reality, we know that variables are spatially correlated. Pear et al. (2019) studied zip codes in 17 states and analyzed the association between zip code-level socio-economic features and counts of POD hospital discharges through Bayesian Poisson space-time models. They showed higher rates of POD could be found in more economically disadvantaged zip codes. As ZIP Codes are based on the location of delivery post offices, so zip codes are not appropriate building blocks for community boundaries (Grubesic, 2008). Even though their model accounts for spatial autocorrelation among neighboring zip codes, it still assumes that the scale of all of the involved relationships are constant over space and thus does not allow for analyzing these relationships at different scales, Whereas, in many cases, including public health crises, this assumption is not valid.

Geographically weighted regression (GWR) is a method that is recently employed to understand how spatial processes associated with opioid overdose vary across space (Nesoff et al., 2020; Donnelly et al., 2020). GWR adjusts for non-stationarity in relationships by the use of
a data-borrowing procedure in order to perform a series of local regressions for each area, which enables the estimation of model parameters at any number of locations in a study area in contrary to a traditional "global" ordinary least squares (OLS) regression model that estimates a single set of parameters, each of which is assumed to be constant across the entire study area. Comparison of local parameter estimates across space is beneficial because it shows whether and how the determinants of opioid overdose vary across geographic space, issues that are overlooked in a global model. Therefore, GWR offers a mechanism not only to explore whether a model is an appropriate representation, but also to define the variables that lead to such a representation for specific locations (Fotheringham et al., 2017; Oshan et al., 2019).

Nesoff et al. (2020) conducted a spatial analysis of fentanyl-involved and non–fentanyl-involved fatal overdoses between 2014 and 2018 in Cook County, Illinois. Using GWR to investigate the spatial variation of covariates at overdose locations, they conclude fentanyl overdoses geographically clustered more than non-fentanyl overdoses, and the odds of a fentanyl-involved overdose were significantly increased for men, Blacks, Latinos/as, and younger individuals. Donnelly et al. (2020) conducted a study to analyze ecological conditions associated with opioid-related possession arrests among Blacks and Whites in Delaware. They noted that community economic deprivation and ethnic inequality more strongly raise rates of possession arrest among Blacks compared to Whites, and how racial discrimination is perpetrated through policing, community composition, economic disadvantage, and the criminal justice system. Their findings demonstrate police responsiveness to drug concerns and discrepancies between Black and White in how drug users engage with the criminal justice system. They identify the ecological conditions associated with opioid-related arrests using geographically weighted
regression (GWR). Existing studies that utilize GWR to analyze local opioid use disorder determinants have several limitations that make it difficult to interpret their results, gain collective insight about opioid crisis processes, and suggest practical policy implementations.

GWR assumes that each determinant operates at the same spatial scale (i.e., the same kernel bandwidth for each variable). However, it is much more likely that the complex social, economic, and demographic factors associated with opioid use disorder and opioid overdose may each vary at different scales (i.e., unique kernel bandwidths for each variable) (Oshan et al., 2019). For example, crime rate can vary sharply across metropolitan areas, and opioid overdose death hot spots spatially overlap with areas of concentrated violence (Carter et al., 2019), so certain citizens are more susceptible to higher opioid overdose rates. In contrast, people are all subject to the effects of aging, and it could be that the relationship between age and opioid overdose rates is independent of geographic context when other factors are taken into consideration. When it is presumed that the same spatial scale applies to all of these relationships, it is likely that the true trends across space are distorted, since the model is mis-specified. As a consequence, when modeling complex spatial processes, it is important to use a multiscale approach, such as multiscale geographically weighted regression (MGWR). MGWR is an extension of GWR that allows studying the relationships at varying spatial scales and achieves that by deriving an optimal bandwidth vector in which each element indicates the spatial scale at which a particular process takes place as opposed to a single, constant bandwidth for the entire study area (Fotheringham et al., 2017).

No research on predictors of prescription opioid overdose has used data with multiple, high-quality socio-economic status measures, despite evidence that different facets of socio-
economic status may vary in their importance for health in different contexts in the United States.

Study Site

In 1999 the first wave of the opioid epidemic started in Wisconsin. That is when opioid deaths started to grow following a spike in opioid prescribing for pain relief. Then the second wave in Wisconsin started in 2010 when deaths involving heroin began to increase. This time more people were using heroin because it was cheaper and more available than prescription opioids. The third wave had begun in 2014. That is when deaths started to increase involving synthetic opioids such as fentanyl. This rise was related to the illicit manufacture of fentanyl and its combination with other narcotics such as heroin (Wisconsin Department of Health Services, 2020).

In Wisconsin, opioids were involved in an estimated 78% of drug overdose deaths in 2018, totaling over 846 (a rate of 15.3). Those involving synthetic opioids other than methadone (mainly fentanyl) continued to grow from 466 in 2017 to 506 in 2018, among opioid-involved deaths. Heroin or prescribed drug deaths decreased to 327 (a rate of 6.0) and 301 (a rate of 5.3) respectively in 2018 (Centers for Disease Control and Prevention, 2020)

Wisconsin has experienced a dramatic rise in opioid abuse, with opioid overdose deaths increasing by 750% between 2000 and 2018. Milwaukee County has been hit particularly hard by the opioid epidemic. Although the number is down slightly from its peak in 2017, there were 383 opioid-related deaths in Milwaukee County in 2018 and at least 1402 overdoses, identified based on the number of incidents requiring naloxone (Narcan) administration by emergency
medical responders (COPE, 2019). 1080 individuals seeking treatment for substance use disorder in Milwaukee County reported opioid use. However, the total number of Milwaukee County residents dealing with Opioid Use Disorder (OUD) is much larger. While opioid use is a state-wide epidemic, the incidence of opioid-related deaths in Milwaukee County (30.1 per 100,000) in 2018 was approximately twice that of the rest of Wisconsin.

Milwaukee's reputation as one of the most racially segregated metropolitan areas in the country is a defining factor influencing the region's socio-economic, political, and cultural environment. By 1970, after the first small wave of black migration to Milwaukee, the metro area had recorded the fifth-highest degree of segregation among the thirty U.S. metropolises with significant black populations (Massey and Denton, 1993; Levine, 2013). Not only has Milwaukee persistently ranked among the most racially segregated metropolitan areas since 1970, but African Americans' residential segregation has also barely declined in Milwaukee over the past thirty years, and studies have found Milwaukee to be the most racially segregated metropolitan area in the U.S. (Frey, 2012). Based on the 2010 census, Milwaukee ranked ninth highest in Hispanic-white segregation rates. Although the segregation of the Hispanic population of Milwaukee is less severe than that of blacks, Hispanic segregation generates "linguistic isolation" (households in which no person aged 14 or over speaks English at least "very well"), which is also part of the demographic and socio-economic environment of Milwaukee (Frey, 2012; Levine, 2013).

A high degree of segregation on any single dimension is problematic as it isolates a minority community from the facilities, opportunities, and resources that influence socio-economic well-being. However, the deleterious effects of segregation are increasing as high
levels of segregation accumulate across dimensions (Massey, 2001). therefore, in the hyper-segregated county of Milwaukee, analyzing the opioid crisis with respect to segregation is needed in order to understand the different layers of the epidemic. Fatal overdose rates do not necessarily follow socioeconomic lines but instead may depend on racial composition, urbanicity, and family organization within a community (Wagner et al., 2019) making Milwaukee an appropriate case study for examining the opioid crisis.

Data Collection

Our study drew data from a wide range of community resources and relied on three basic data categories. The first source of data captured the precise location of fatal opioid overdose deaths. The 2017-2021 dataset was provided by the Medical County Medical Examiner office, which investigates all suspicious deaths occurring in Milwaukee County. The second category of data was health-related, with several subtypes. Public health measures were provided by the Wisconsin Department of Health Services. Emergency medical call for service (EMS) data were provided by the Milwaukee Fire Department. Opioid availability data were derived from The Automation of Reports and Consolidated Orders System (ARCOS) dataset from the Drug Enforcement Administration (DEA) (Washington Post, 2020). The third category of data was demographic- and socioeconomic-related. I collected the data from U.S. Census Bureau’s 2020 American Community Survey 5-Years Estimates of census-tract level data (Data.census.gov). Crime data were provided electronically by the Milwaukee County Police Department for 2017-2021. To consider the effect of digital divide, I calculated the "digital divide index" for all census tracts in Wisconsin, using Gallardo's (2017) proposed formula.
Methodology

MGWR

The term "spatial context" refers to the generally unquantifiable influence of an individual's location on their behavior, preferences, and actions in space. Phenomena are located and happen in the context of time; for example, describing temperature across the country by a single average value would be misleading, as different locales have significantly different temperatures. Likewise, the temperature is dependent on a myriad of variables that vary themselves locally; therefore, if we assume the strength of the effect of each determinant of temperatures is constant across space, it would be as misleading as a single average value.

Spatial heterogeneity and dependence of processes, or in other words, processes variation in the spatial context, basically is Tobler's first law of geography operationalization as "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). Therefore, modeling approaches should account for the possibility that the strength or even direction of processes might vary spatially. Technically, failure to consider the spatial context in modeling will lead to spatial dependency amongst residuals, which invalidates the model and its inferences.

Furthermore, the modifiable areal unit problem (MAUP), which states that the inferences we draw from the study of geographically aggregated data might vary depending on the level to which the data are aggregated, is a well-known problem in dealing with aggregated spatial data. In its most extreme form, it is possible to make radically different inferences from the same underlying data that has been aggregated to different levels. However, MAUP can be recontextualized as it is the result of process spatial dependence and heterogeneity; if the
processes being studied are fundamentally constant across space, our conclusions should be identical regardless of how we aggregate the basic data. However, if the mechanisms affecting the processes differ across space, different data aggregations will yield different outcomes and in extreme circumstances, different inferences about the processes. Since inferences will not depend on a single set of averaged outcomes, local modeling may alleviate the MAUP problem (Fotheringham, 2020).

The most commonly used technique in quantitative analysis, simple linear regression, assumes process changes across geographic space are universal, which is not necessarily the case in every context. When utilizing simple global fitting methods like ordinary least squares (OLS), variations across geographical space, known as spatial heterogeneity and dependence, may be missed (Brunsdon et al., 1996).

Contextual changes in spatial behavior might exist as local common traditions, customs, lifestyles, and daily routines, environmental influences, economic conditions, social imitation and peers influence, and personal differences across space influence individual conduct (Golledge 1997, Altman and Wohlwill 2012, Clark 2003, Garz 2018, Braha and de Aguiar 2017, Rentfrow et al., 2015). Therefore, the impacts of location on human behavior or spatial context should be recognized. Specifically, Gomez et al. (2007) showed how bad weather influences voter turnout and voter behaviors at the local level. Ayllón (2013) highlighted the impacts of economic conditions and their respective stigma effects and discouragement as sources of contextual changes in behavior. Garz (2018) showed how local news coverage could affect people's perceptions. Braha and De Aguiar (2017) emphasized the instrumental impact of social contagion by peers such as family and friends in shaping collective behaviors in communities.
Rentfrow et al. (2015) conducted a study on the geographical distribution of personality traits. Their results revealed how psychological characteristics vary locally and exhibit significant spatial dependency and heterogeneity due to social, ecological, and genetic (selective migration) influences, highlighting the role of collective personality traits in contextual changes in spatial behavior. Coffee et al. (2020), through a study on determinants of residential property due to relative location factor, showed how spatial context provides a more informed measure of socioeconomic status, and inclusion of spatial context enhances research and policy formation.

Moreover, studies showed that local models outperform conventional global models in many situations, even when model complexity is considered (Wang et al., 2018; Zhu et al., 2020; Mollalo et al., 2002). Local models such as Geographically Weighted Regression (GWR) and Multi-Scale Geographically Weighted Regression (MGWR) are recently employed to understand how spatial processes associated with public health and disaster crises vary across space (Nesoff et al., 2020; Donnelly et al., 2021, Chun et al., 2017; Purwaningsih et al., 2018; Rifat and Liu, 2020).

Specifically, regarding disaster management, Wang et al. (2017) employed GWR to examine the relationships between inundation frequency and selected explanatory spatial factors. Johnson et al. (2019) implemented a GWR model to investigate the association between financial services accessibility and natural disaster resiliency. Mardianto et al. (2021) used GWR to estimate the number of flood disasters based on the influence of settlements along riverbanks. Finally, Rifat and Liu (2020) examined disaster resilience's impact on disaster losses among United States' coastal communities.
On the other hand, Nesoff et al. (2020) employed GWR to investigate the spatial variation of covariates at overdose locations. Donnelly et al. (2020) conducted a study to analyze ecological conditions associated with opioid-related possession arrests among Blacks and Whites in Delaware. Dutta et al. (2021) used GWR and MGWR to assess the relationship between adaptive, sensitivity, and exposure determinants of COVID-19 incidence. Similar studies were conducted in different scales and locations (Mollalo et al. 2020; Mansour et al. 2021; Maiti et al. 2021). for example, using MGWR, Iyanda et al. (2020) examined the health and social determinants of the COVID-19 pandemic for 179 counties. All these studies' results highlighted the importance of considering spatial variations of components, especially in larger study areas where dimensions of components differ significantly.

Existing studies that utilize GWR to analyze disaster damage determinants have several limitations that make it difficult to interpret results or to gain collective insight about processes, thereby hampering practical policy implementations. GWR assumes that each determinant operates at the same spatial scale (i.e., the same kernel bandwidth for each variable). However, it is much more likely that the complex social, economic, and demographic factors associated with natural disasters and public health crises may each vary at different scales (i.e., unique kernel bandwidths for each variable) (Oshan et al., 2019). When it is presumed that the same spatial scale applies to all of these relationships, it is likely that the actual trends across space are distorted since the model is misspecified. Consequently, when modeling complex spatial processes, it is essential to use a multiscale approach, such as multiscale geographically weighted regression (MGWR). MGWR is an extension of GWR that allows studying the relationships at varying spatial scales and achieves that by deriving an optimal bandwidth
vector in which each element indicates the spatial scale at which a particular process takes place as opposed to a single, constant bandwidth for the entire study area (Fotheringham et al., 2017).

Ordinary Least Squares Regression (OLS) assumes that the conditioned association between the dependent variable and every independent variable is constant over space; therefore, data from every location in a study area has an equal weighting in the calibration of the model, and a single parameter estimate will be computed for each relationship in the model (Brunsdon et al., 1998). However, many relationships vary across a study area, shaped by socio-spatial differences. Therefore, Brunsdon et al. (1996) introduced GWR as an extension to general regression models. GWR relaxed the spatial stationarity hypothesis associated with global models, allowing relationships to differ from location to location. GWR is denoted by (Fotheringham et al., 2017):

\[
y_i = \beta_{i0} + \sum_{j=1}^{m} \beta_{ij} x_{ij} + \epsilon_i, i = 1, 2, ..., n
\]

where at observation i, \( y_i \) is the dependent variable, \( \beta_{i0} \) is the intercept, \( \beta_{ij} \) is the jth regression parameter, \( x_{ij} \) is the value of the jth explanatory parameter, and \( \epsilon_i \) is a random error term.

The best bandwidth and kernels selection is the most critical criterion for executing any Geographically weighted technique. By considering the choice of the kernel (fixed or adaptive) and how surrounding points are weighted, we may compute an optimum bandwidth for all the
independent variables. The fixed spatial kernel employs distance as a parameter and is commonly used when the data distribution is uniform. As a result, each local regression is calculated using a certain bandwidth distance. However, this type of kernel is insufficient when data is sparse, such as in areas with few regression points, where a fixed spatial kernel gives insufficient fluctuations, resulting in a significant standard deviation of errors (Fotheringham et al., 2003).

An adaptive kernel may be preferable when dealing with nonuniform geographical distributions (Fotheringham et al., 2003; Oshan et al., 2018). However, Murakami et al. (2019) showed that when the bandwidth is too small, most weight matrix elements take near-zero values, leading to the issue of singularity. In other words, small bandwidths introduce overfitting. Furthermore, if subsamples are sparsely dispersed throughout the space, the situation becomes serious. As a result, adaptive bandwidth, which adjusts the kernel window size based on sample density, might be an excellent method to address this issue (Fotheringham et al., 1998).

Another factor to examine is the kernels' weighting functions (Fotheringham et al., 2003). Gaussian weighting takes into account all data points where the weight steadily falls away from the kernel's center. Weights are never allocated zero values in this circumstance. On the other hand, the bi-square weighting has a defined range for which the weights are non-zero. Furthermore, all data points outside of the ideal bandwidth are set to zero, so they have no effect on our local regression. Following the specification of these elements, we may use iterative optimization processes to discover an optimum bandwidth that minimizes either the corrected Akaike Information Criterion (AICc) or the cross-validation (CV) statistic. The AICc
optimization captures the divergence between expected and observed values to calculate the information distance (Charlton et al., 2009). When estimating the expected dependent variable at each regression point, CV minimization investigates the sum of the squared errors (Brunsdon et al., 1996).

In essence, the GWR explicitly integrates geographic context by allowing parameter estimates to be determined for each location of interest. Parameter estimates are obtained at each location by calibrating a locally weighted regression; the spatial weights matrixes of these locally weighted regression encode a data-borrowing scheme designed to allow data points closer to the place of interest to have a more significant effect on the local regression. Parameter estimates are obtained at each census tract by calibrating a locally weighted regression using the following estimator in matrix form (Fotheringham and Oshan, 2016):

$$\hat{\beta}(i) = (X'W(i)X)^{-1}X'W(i)y$$

where $\hat{\beta}$ is the $k \times 1$ vector of parameter estimates, $X$ is the $n \times k$ matrix of the selected explanatory variables, $y$ is the $k \times 1$ vector of observations of the dependent variable, and $W(i)$ is the matrix of spatial weights that encodes a data-borrowing scheme designed to allow data points closer to location $i$ to have a more significant effect on the local regression (Oshan et al., 2019).

The GWR framework assumes the same bandwidth to apply for each relationship in the model, which means the data are weighted at the same scale. While, in many cases, including the public health crises or natural disasters, this assumption is not valid because different processes are involved with varying spatial scales. Fotheringham et al. (2017) proposed an
extension to the GWR framework overcoming this drawback by reformulating equation GWR as a generalized additive model (GAM):

\[ y_i = \sum_{j=0}^{m} f_{ij} + \epsilon \]

where \( f_{ij} \) is the \( j \)th additive term and is a smoothing function applied to \( j \)th explanatory variable at observation \( i \) (Oshan et al., 2019).

GWR is recast as a generalized additive model (GAM) in MGWR, which then employs a back fitting procedure to calibrate a sequence of GWR models based on their partial residuals until the MGWR model converges to a solution (Fotheringham, Yang, and Kang, 2017). In other words, the algorithm allows us to refine the partial residuals and update each iteration process using an appropriate smoothing function. For speedier model calibration, Oshan et al. (2019) argue that an MGWR should be initialized with a beginning value (usually the optimum parameter values based on the GWR model). Therefore, GWR parameter estimates are used to start the backfitting algorithm, and the calibration procedure iterates based on these initial values, and during each iteration, all local parameter estimations and optimal bandwidths are examined. When the difference between the parameter estimations from subsequent iterations converges to a defined threshold, the iteration ends. Thus, due to applying the GWR, the parameter estimates generated are localized to that region rather than being universal across space.

MGWR backfitting algorithm generates a set of bandwidth parameters for each modeled process, indicating a different scale for each process. Bandwidth variations reflect differences in spatial scales, and by capturing the effect of scale in spatial processes, "MGWR can more accurately capture spatial heterogeneity, minimize overfitting, mitigate nonlinear
dependencies among the predictor variables, and reduce bias in the parameter estimates" (Fotheringham et al., 2017, Wolf et al. 2018; Oshan et al., 2019; Yu et al., 2020; Oshan et al., 2020).

The local estimates of the intercept and the covariate-specific optimum bandwidths are possibly the most interesting output from MGWR. These indicate the intrinsic levels of the dependent variable when all other variables in the model are held constant. In essence, the local intercept estimates are a measure of spatial context (Fotheringham et al., 2021).

Local spatial modeling has opened up many doors to the geospatial analysis of critical societal issues; as Fotheringham (2020) calls it, conventional global "one-size-fits-all" approaches are not the most appropriate way to understand complex phenomena. Where spatial dependency and heterogeneity are present when events are located and dependent on their spatiotemporal context, when processes vary at different scales across time and space, local spatial modeling approaches which compute localized parameter estimates yield a more informed conclusion.

Inference in local models is challenging due to the multiple testing issue (different tests of the significance of local parameter estimates in GWR to avoid excessive false discoveries) (da Silva & Fotheringham, 2016), bandwidth selection uncertainty (Li et al., 2020). For example, the parameter estimates acquired in the calibration of a local model will be a function of the dependent variable if the association between the independent and dependent variable is nonlinear but is approximated by a linear function. Meaning, if a covariate has a positive spatial dependency, then the local parameter estimates will also have a positive spatial dependency. As a result, the interpretation of local parameter estimations as indicators of process
nonstationarity or nonlinearity would be muddled. Consequently, Sachdeva et al. (2021) argued that regional variations in parameter estimates given in the calibration of models such as MGWR fundamentally might not reflect intrinsically spatially varying relationships but rather nonlinear conditioned correlations between \( y \) and \( x \). Plotting the local estimates against the respective value of the covariate is a technique to determine the source of spatially variability. If the spatially varying parameter estimates are due to nonlinearity between a covariate \( x \) and the response variable \( y \), this plot would have an apparent structure. If, on the other hand, the resulting plot has no discernible structure, it's safe to presume that the processes being described are spatially dependent.

Overall, modeling such complex issues through a complicated modeling approach would produce massive amounts of output that would need careful examination for robust interpretation (Yu et al., 2020). This dissertation provides the first application of the MGWR framework for modeling opioid overdose death determinants to the knowledge of the author.

I compiled 225 demographic, health-related, and socioeconomic factors, which I identified as explanatory variables (Table 1). All variables were collected, processed, and joined to the administrative boundary shapefile of census tracts collected from the TIGER / Line database (www.census.gov), using ArcGIS Desktop 10.7. A stepwise forward methodology was employed to pick a subset of variables by removing non-significant explanatory variables. Pearson's correlation analysis was subsequently applied to investigate the correlations of all pairs of selected variables. Variance inflation factor (VIF) was used to detect multicollinearity, and thus the most uncorrelated factors were selected as the input of the models (O'Brien, 2007). Then, to investigate opioid overdose death determinants in Milwaukee County,
Wisconsin. I calibrated a global model using regression, which assumes processes to be constant across the study area. Subsequently, an MGWR model was calibrated using a golden section search bandwidth selection routine to obtain optimal bandwidths. Following Oshan et al. (2019), MGWR maps were prepared to visualize the parameter estimates and their statistical significance, with insignificant estimates displayed in grey. The MGWR maps are presented for each variable in order to investigate the parameter estimate spatial heterogeneity. Then the implications of results were discussed.

Interrupted time series

To examine the impact of Covid-19 on opioid overdoses in Milwaukee County, I collected data on opioid-related overdose deaths between January 1, 2017, and December 31, 2020 (I exclude overdose deaths directly attributed to non-opioids). Wisconsin issued its state-wide Stay at Home Order on March 23rd, 2020. Thus, for the current study I established this as the “intervention” date and defined the preceding time period from Jan 1st, 2017, to March 23rd, 2020, as the pre-intervention/pre-pandemic stage and the time period between March 24, 2020, and December 31, 2020, as the post-intervention/post-pandemic stage.

Interrupted time series analysis has been a well-established method used by epidemiologists to evaluate the effectiveness of interventions aimed at mitigating opioid overdose deaths (Puenpatom et al., 2012; Walley et al., 2013; Martin et al., 2018; Ranapurwala et al., 2019; Feder et al., 2020). Moreover, this analytical approach has also been useful for studying the impact of the COVID-19 pandemic and the corresponding responses on opioid overdose deaths (Glober et al., 2020; Slavova et al., 2020).
Social phenomena often exhibit autocorrelation and seasonal effects, which can lead to temporary changes in event rates. Interrupted time series analysis can help to capture these effects and enable researchers to assess the impact of specific policies. This is achieved by analyzing changes in the level and slope of the time series before and after an intervention, comparing the structure of the temporal dynamic before and after the intervention. In recent years, scientists have developed a framework for evaluating the causal influence of interventions using Bayesian statistics. In this study, the effect of COVID-19 containment measures, particularly the "stay at home" order, on the Opioid Overdose Death (OOD) trends in Milwaukee was investigated using Bayesian structural time-series (BSTS) models developed by Brodersen et al. (2015). A diffusion-regression state-space approach was utilized to forecast counterfactual patterns in trends that would have existed if there had been no intervention in a hypothetical scenario. This approach helps to measure the effect and statistical significance of a specific incident on the variable of interest, the OOD rate. The "bsts" (Scott, 2020) and "Causal Impact" (CausalImpact BK, 2015) packages in R were used to implement Bayesian structural time-series (BSTS) models.

*Spatial empirical Bayesian smoothing*

To account for population size, Opioid Overdose Death (OOD) rates were computed by census tract in Milwaukee County. However, these rates were found to exhibit intrinsic variance instability, indicating that the precision of a rate as a measure of underlying risk is inversely proportional to the size of the at-risk population. This can result in significant standard error in rates calculated from small populations, leading to the misclassification of high and low-risk areas. For instance, an area with a smaller population but a high number of OODs may be
misclassified as high-risk, while a larger area with a lower number of OODs may be misclassified as low-risk, based on a fixed boundary such as a 10% rate. To address this issue, spatial empirical Bayes smoothing techniques were applied to reduce variation caused by population size. These techniques improve the precision of the crude rate by borrowing strength from other observations. In Spatial Empirical Bayes, the reference rate is computed for a spatial window surrounding each observation, rather than using the same reference rate for all observations. The Spatial Empirical Bayes smoothed rate is then calculated as a weighted average of the crude rate and the prior or reference rate estimated from the spatial window consisting of the observation and its neighbors. This approach was developed by Anselin et al. (2006).

*Spatial Social Network*

A set of nodes connected by edges representing 'ties' or interactions between individuals or organizations can be defined as a social network (SN) (Wasserman and Faust 1994). SNs have been used to study the association between entities in a myriad of applications including supply chain management (Shaharudin et al., 2019; Han et al., 2020), education research (Dahesh et al., 2020), crime (Bright et al., 2021), etc. However, social phenomena exhibit spatial autocorrelation and spatial heterogeneity that can also produce biases in the social network analysis, as network dynamics and agents' decisions are pretty much driven by geographic context (Andris, 2016). Therefore, over the past few years, there has been a shift toward adding environmental or geographic context to social networks (Sarkar et al., 2019; Albery et al., 2021; Ye & Andris, 2021; Uitermark & Van Meeteren, 2021).
The concept of social flow was introduced by Andris (2016) and defined as a connection created by an agent-based relationship between two places. In other words, social flow is an edge between two geolocated nodes in a social network graph. On the other hand, the geolocation of nodes can be defined as the node's anthroospace. In fact, an anthroospace is a locale where an agent (node) operates. Thus, it can be formalized to associate nodes with specified constrained geography (e.g., points, lines, or polygons). When an edge links two nodes, their anthrospaces get connected. So, in spatial SNs, edges not only link individuals but places. The sum of associations between groups of places creates a place-to-place flows network, called an aggregate social network (Andris, 2016). The dyadic (pairwise) relationship of the journey to an overdose between victims' residences and overdose incident locations is the building block of our aggregate social network. Each node in a journey to overdose dyad was spatially joined to the underlying U.S. Census tract to create an aggregate social network at the census tracts level.

Constructing an aggregate social network as the basis of the study, I propose a novel framework to answer the research question:

Focal points of geographically discordant overdoses

There is a conceptual difference between the definition of distance in geography and social networks. In social network analysis, distance is measured in hops incurred from moving along edges between nodes. Thus, standard network-level properties are non-spatial and can be used to analyze the population dynamics of the social network in its entirety (Sarkar et al., 2019).
Network-level properties such as Average Path Length and Network Diameter provide insights about the network's general connectivity. For example, the mean inter-node distance can be measured by average path length, and the maximum distance between two of the farthest nodes can be calculated by network diameter. On the other hand, betweenness and closeness centrality metrics have been designed to characterize the importance of nodes in the network (Sarkar et al., 2019). Betweenness centrality is the number of shortest paths between all pairs of nodes that pass through the focal node for transitivity. Closeness centrality captures the average distance with which a node can reach all other nodes in the network (Borgatti 2005). However, since we are modeling the journey to overdose, which is a social flow transmission and diffusion embedded in spatial space (nodes represent census tracts), nodes' second-degree links are not meaningful. For example, in a journey to overdose network, flow from neighborhood A to neighborhood B, and flow from neighborhood B to neighborhood C do not necessarily imply that overdose might spread between neighborhoods A and C. Therefore, network-level properties such as Average Path Length and Network Diameter and entity-level metrics such as betweenness and closeness centrality cannot be used to measure journey to overdose aggregate social network structure.

In the aggregate social network, the degree centrality is the most fundamental node importance metric, which counts the number of edges connected to a node. In the context of directed graphs, within a specific domain of application, it is necessary to consider the neighbors of a given node differently, based on the edge's directions. For example, in an academic publications network, a review article cites other important original articles that can be considered authoritative sources of information; in this network, review articles have edges
directed toward many important original articles. On the other hand, important papers are referenced by well-structured review articles; therefore, in the network, they are pointing toward many review papers. Consequently, it can be perceived that there are two important node types in the academic publications network, authorities of important information and hubs that can lead us to authorities.

Jon Kleinberg (1999), developed the Hyperlink-Induced Topic Search (HITS), which is a significant and famous ranking algorithm. In fact, HITS is a node importance metric that calculates both an authority centrality and a hub centrality for all nodes in the network. The authority centrality of a given node is proportional to the sum of hub centralities of all the nodes to which it points to, and the hub centrality is proportional to the sum of authority centralities of all the nodes which point to the given node. The algorithm has been constantly improved and widely used across a wide range of applications, including bioinformatics (Nickerson et al., 2018; Liu et al., 2020), education (Yang & Sun, 2013), and economics (Zhang et al., 2017; Deguchi et al. 2014).

In this study, I adopted the HITS algorithm to study the journey to the overdose network; hub and authority centralities were used to pinpoint the focal point of geographically discordant overdoses and original residences in the directed and weighted journey to the overdose network. Here all census tracts are regarded as nodes with hub scores and authority scores, respectively. A census tract may point to one or more than one census tract through the overdoses link (the link between residence and incidents). To illustrate, suppose a census tract has widely imported overdose incidents from census tracts that are home to many victims with higher hub scores. In that case, it can be regarded as the focal point of geographically
discordant overdoses and ranked with a higher authority score. Conversely, suppose a census tract has been home to many victims and linking them to the focal point of geographically discordant overdoses with higher authority scores. In that case, it can be regarded as the focal point of residence of geographically discordant overdoses and ranked with a higher hub score. An edge weight indicates the degree of involvement or the number of times the journey to overdose happened between two specific census tracts.

To the best of our knowledge, HITS has never been applied in this domain of application. Adopting the HITS algorithm is a unique and promising methodology to study the journey to overdose, which makes the analysis of geographically discordant overdoses of census tracts more reasonable and the recommendation of policymaking for it more reliable since hub and authority values are constantly mutually reinforcing in the calculation process.

Hotspots of imported and domestic drug overdose

Traditionally, most of global and local indexes of spatial autocorrelation, such as Moran's I, local indicators of spatial association (LISA), and Getis–Ord statistics, have been developed to detect spatial clusters of individual points without considering the connectivity between the points and clusters. However, Baker et al. (2020) developed a spatial social network hot spot detection methodology and argued that using traditional hotspot analyses are reliable when dealing with independent incidences of unconnected events; when it comes to connected events such as the journey to overdose, scan methods can provide a better conceptual framework for hot spots detection of dependent/connected events.

In this study, I use Baker's (2020) proposed methodology for hotspot detection in our aggregate spatial social network to detect network interactions hotspots where the number of
edges and network density within the network of nearest neighbors of each node are the highest. In other words, the algorithm detects hotspots where geographically discordant overdoses are not only densely located but also linked, revealing which census tracts host overdoses ties, not just a cluster of overdoses.

Defining characteristics of geographically discordant and non-discordant overdose deaths

I examined overdose death data from January 1, 2017, through December 31, 2020, to define the characteristics of geographically discordant (i.e., imported) vs. domestic (i.e., non-discordant) overdose deaths. Discordant and non-discordant overdose deaths were disaggregated by mode of death, toxicology/cause of death, gender, race, and age. I performed 2-sample t-tests to detect differences in each mode of death, cause of death, gender, race, and age group between discordant and non-discordant deaths. The rationale for using t-tests is that, under the null hypothesis, drug overdose deaths have the same mean for imported and domestic overdoses. All statistical analyses were performed in R.

Results

A dataset of 225 candidate variables was compiled. Nonetheless, these variables exhibited high collinearity based on their Pearson's correlation coefficients and global variance inflation factors (VIFs) when evaluated against one another (O'Brien, 2007). Thus, the most uncorrelated factors were selected as the input of the stepwise mixed methodology to pick a subset of variables by removing non-significant explanatory variables. A variety of different explanatory variables was selectively chosen to develop a robust model to investigate the
variability of the dependent variable “opioid drug overdose death rate” in Milwaukee County (Table 1).

Table 1 Description of the explanatory variables included in the study.

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td>Population</td>
<td>Female (Percentage)</td>
</tr>
<tr>
<td>Demographic</td>
<td>Marital Status</td>
<td>Divorced (Percentage)</td>
</tr>
<tr>
<td>Demographic</td>
<td>Marital Status</td>
<td>Widowed (Percentage)</td>
</tr>
<tr>
<td>Health</td>
<td>Opioid Availability</td>
<td>Total dispensed OXYCODONE 2006-2014</td>
</tr>
<tr>
<td>Health</td>
<td>Prevention</td>
<td>Narcan Direct Program Availability (number of agencies per census tract)</td>
</tr>
<tr>
<td>Health</td>
<td>Prevention</td>
<td>Naloxone Dispensers Availability (number of dispensaries per census tract)</td>
</tr>
<tr>
<td>Health</td>
<td>Prevention</td>
<td>Hospitals and clinics Availability (number of Hospitals and clinics with ER per census tracts)</td>
</tr>
<tr>
<td>Health</td>
<td>Public Health</td>
<td>Persons with a Disability</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>Neighborhood Stability</td>
<td>Households who resided in a housing unit for more than 30 years</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>Neighborhood Stability</td>
<td>Households who resided in a housing unit for 4 to 8 years</td>
</tr>
</tbody>
</table>
Results from the global Ordinary Least Square model for 296 Observations (census tracts) with an R-Squared of 0.371 are summarized and presented in Table 2, in order to provide context for the MGWR results.

Table 2 Results from the ordinary least squares regression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-Statistic</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.001417</td>
<td>-3.469621</td>
<td>------</td>
</tr>
<tr>
<td>Naloxone Availability</td>
<td>0.845867</td>
<td>4.815833</td>
<td>1.196023</td>
</tr>
<tr>
<td>Crime Rate</td>
<td>0.003947</td>
<td>4.054279</td>
<td>1.363406</td>
</tr>
<tr>
<td>Inequality of household income</td>
<td>0.002171</td>
<td>3.687269</td>
<td>1.372546</td>
</tr>
<tr>
<td>Prevalence of disability</td>
<td>0.000038</td>
<td>4.493521</td>
<td>1.4189</td>
</tr>
<tr>
<td>Resided in the same unit for 20 to 30 years</td>
<td>0.000027</td>
<td>3.103784</td>
<td>2.016146</td>
</tr>
<tr>
<td>Total dispensed Opioids</td>
<td>0.000017</td>
<td>2.651645</td>
<td>1.143462</td>
</tr>
<tr>
<td>Resided in the same unit for 4 to 8 years</td>
<td>0.000013</td>
<td>2.294834</td>
<td>1.83644</td>
</tr>
<tr>
<td>Houses built before 1950</td>
<td>0.000003</td>
<td>2.147099</td>
<td>1.222565</td>
</tr>
<tr>
<td>Healthcare resources accessibility</td>
<td>0</td>
<td>-2.266216</td>
<td>1.110823</td>
</tr>
<tr>
<td>Renters spending more than 30% of their income on rent</td>
<td>-0.000009</td>
<td>-2.877497</td>
<td>1.818318</td>
</tr>
<tr>
<td>Households with child</td>
<td>-0.000028</td>
<td>-2.582162</td>
<td>1.156827</td>
</tr>
<tr>
<td>Educational attainment</td>
<td>-0.000532</td>
<td>-1.715223</td>
<td>2.080905</td>
</tr>
</tbody>
</table>
The global model produces a relatively Moderate R2, indicating about 40% of the variation across opioid overdose death rates can be accounted for by the selected variables in this study. Multicollinearity does not seem to be an issue since the VIFs for each explanatory variable are all under 10 (O’brien, 2007). Based on a standard t-value threshold of 1.96 for a 95% confidence level, all of the variables except of education attainment (population over 25 years old with college degree) are statistically significant. From Table 2, it can be seen that the most influential variable is Naloxone Availability in the census tracts, which has a relatively strong positive relationship with opioid overdose death rates, followed by crime rate, inequality of household income, percentage of persons with disability, neighborhood stability, and opioid drug availability. In contrast, the most influential negative association is with educational attainment and households with kids (enrolled in Nursery schools or preschools).

Table 3 MGWR explanatory variable bandwidth and significance (The bandwidth is expressed as the number of neighboring census tracts)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bandwidth</th>
<th>Effective # parameters</th>
<th>Critical t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>277</td>
<td>1.522</td>
<td>2.144</td>
</tr>
<tr>
<td>Resided in the same unit for 4 to 8 years</td>
<td>295</td>
<td>1.41</td>
<td>2.113</td>
</tr>
<tr>
<td>Resided in the same unit for 20 to 30 years</td>
<td>295</td>
<td>1.41</td>
<td>2.113</td>
</tr>
<tr>
<td>Educational attainment</td>
<td>295</td>
<td>1.223</td>
<td>2.054</td>
</tr>
<tr>
<td>Households With child</td>
<td>293</td>
<td>1.496</td>
<td>2.137</td>
</tr>
</tbody>
</table>
The above results assume that relationships are constant across the study area. In order to relax this assumption, deter unexplainable high level of spatial heterogeneity, local multicollinearity and concurvity in the local subsets of the data (Oshan and Fotheringham, 2018; Oshan, et al. 2020), and allow the processes to vary at different scales, it is necessary to employ MGWR. Calibrating an MGWR model produces a vector of optimal bandwidth that describes the spatial scale at which each process in the model varies (Oshan, et al. 2020). MGWR was applied to the same set of explanatory variables used in the global model, The R2 increased to 0.56 in the MGWR model from 0.371 in the global model and the AIC decreased to 675.786 in the MGWR model from 730.974 in the global model.

In Table 3, the bandwidths related to each explanatory variable are listed (theoretically the global model assumes bandwidth of infinity). Eight relationships affect opioid overdose rate at global scale: Neighborhood instability, educational attainment, households with kid at

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>Bandwidth</th>
<th>LMSE 1</th>
<th>LMSE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naloxone Availability</td>
<td>45</td>
<td>16.687</td>
<td>2.993</td>
</tr>
<tr>
<td>Crime Rate</td>
<td>199</td>
<td>3.022</td>
<td>2.41</td>
</tr>
<tr>
<td>Healthcare resources accessibility</td>
<td>294</td>
<td>1.49</td>
<td>2.135</td>
</tr>
<tr>
<td>Total Dispensed opioid</td>
<td>295</td>
<td>1.525</td>
<td>2.145</td>
</tr>
<tr>
<td>Renters spending more than 30% of their income on rent</td>
<td>189</td>
<td>3.445</td>
<td>2.459</td>
</tr>
<tr>
<td>Prevalence of Disability</td>
<td>295</td>
<td>1.286</td>
<td>2.075</td>
</tr>
<tr>
<td>Houses built before 1950</td>
<td>295</td>
<td>1.407</td>
<td>2.112</td>
</tr>
<tr>
<td>Income Inequality</td>
<td>293</td>
<td>1.346</td>
<td>2.094</td>
</tr>
</tbody>
</table>
preschool age, total dispensed opioid drug (2006-2014), access to healthcare, percentage of residents with disability, houses built prior to 1950, and index of household income inequality, with bandwidths indicating almost all the data is included in each local subset. Two associations seem to occur at a regional scale crime rate and percentage of households spending more than 30% of income on rent with bandwidths implying several hundred nearest neighbors. The process of naloxone availability varies locally, having relatively small bandwidth. Several further patterns are apparent upon inspection of the MGWR parameter estimate surfaces along with their uncertainty.

Investigating visual pattern, it can be inferred that Neighborhood instability, educational attainment, households with kid at preschool age, total dispensed opioid drug (2006-2014), access to healthcare, percentage of residents with disability, houses built prior to 1950, and percentage of households spending more than 30% of income on rent, surfaces are effectively global and statistically non-zero, these surfaces display little-to-no spatial heterogeneity (Figure 1).
In concordance with the global model results, Disability, inequality of household income, neighborhood instability, and total dispensed opioid drugs (2006-2014) have positive associations with opioid overdose death rates across the study area while educational attainment, access to healthcare, households with kids, and welfare (percentage of renters spend 30% of their income on rent) have negative associations with opioid overdose death rates across the study area.

Figure 1 Maps of MGWR parameter estimate surfaces for the households with kids at preschool age, educational attainment, percentage of persons with disability, and access to healthcare which tend to show little-to-no spatial heterogeneity.
Inequality of household income and crime rate have a number of positive non-zero parameter estimates and display regional spatial variation (Figure 2). The crime rate surface is clustered in western and southern part of Milwaukee County, where residents are predominantly European Americans, inspection of crime rate surface unveil that crime rate does not play and important role in opioid overdose death rate northeastern part of Milwaukee county. The characterization of this cluster requires further investigation but is in agreement with the global model. Inequality of household income exhibits regional spatial variation too, its surface is clustered in southern part of Milwaukee County, where mostly Hispanic Americans and European Americans reside. Inequality of household income examination shows that it is not an influential determinant of opioid overdose death rate among African Americans.

Figure 2 Maps of MGWR parameter estimate surfaces for crime rate and inequality of household income which tend to show regional patterns of spatial heterogeneity.
Timely administration of naloxone can prevent opioid overdose deaths. Naloxone has been used by emergency medical personnel to pharmacologically reverse overdoses for more than 3 decades (Kim et al. 2009). Since the 1996, community-based programs have provided naloxone to persons who use drugs, their community, and service providers to reverse the potentially fatal overdose of opioids (Wheeler et al., 2012).

Naloxone availability parameter estimates display local spatial variation (Figure 3). As it is expected, naloxone availability has a strong negative association with opioid overdose death rate in affluent white census tracts and rural census tracts. In contrary, in census tracts with high population density, or predominantly African American or Hispanic Americans population, Naloxone availability parameter estimates is either not significantly different from zero or in
specifically Hispanic neighborhood has a positive association with opioid overdose death rate, which can be an indication of failure of governmental programs to control/address opioid overdose death in these neighborhoods.

Naloxone’s efficacy is entirely time dependent. After an overdose, death usually occurs within 1 to 3 hours (Kim et al. 2009). Thus, naloxone is only successful in reversing an overdose if administered before overdose symptoms cause death. Medical first responders and emergency departments are equipped with naloxone. It is always the case, however, that these service providers arrive too late to rescue victims of overdose (Giglio and DiMaggio, 2015). Bystanders might, on the other hand, be hesitant to call 911 due to fear of police interference (Kim et al. 2009). Therefore, overdose victims’ friends or family members are most often the real first responders and are ideally equipped best positioned to intervene within an hour of the onset of overdose symptoms. Giglio and DiMaggio (2015) note that overdose education and lay administration of naloxone is a safe and effective community-based approach to control the opioid overdose epidemic. Wisconsin department of health services’ NARCAN Direct Program agencies have resolved to equip people at risk for an opioid overdose and people who may witness an opioid overdose and their community with the antidote. But apparently, failure to engage suffering communities, lack of education in marginalized communities of color, and failure to address African Americans and Hispanic Americans’ concerns can be the main reasons behind differential patterns of Naloxone availability efficacy in Milwaukee County.

Milwaukee’s reputation as one of the most racially segregated metropolitan areas in the country is a defining factor influencing the region's socio-economic, political, and cultural environment. A high degree of segregation on any single dimension is problematic as it
marginalizes minority communities from the facilities, opportunities, and resources that influence socioeconomic well-being (Massey, 2001). The effects of segregation in Milwaukee County is at the level profound that it seems people are living in three different worlds. For example, Milwaukee County crime rate in 2019 was 42.9 per 1000 residents, however investigation of crime rate in segregated regions of Milwaukee County clearly shows the profound difference between regions, crime rate in African American majority region is 83.9 per 1000 residents, while crime rate in Hispanic majority region is 52.4 per 1000 residents and in European American majority region is 18.1 per 1000 residents. Thus, in a hyper-segregated county of Milwaukee, analyzing the opioid crisis with respect to segregation is needed in order to understand different layers of the epidemic. I divided the Milwaukee County into three regions: census tracts in which European American are majority, census tracts in which African American are majority, and census tracts in which Hispanic American are majority (Figure 4).
Three different Pearson's correlation coefficients and VIFs evaluations were conducted to detect multicollinearity in each region then using the stepwise forward regression analyses were conducted to select a subset of variables by removing non-significant explanatory variables. A variety of different explanatory variables was selectively chosen for each region to develop a robust model to investigate the variability of dependent variable “opioid drug overdose death rate” in three different regions of Milwaukee County. In table 3, different selected impactful variables for each region is presented.
Division of Milwaukee County into three different regions and conducting separate models yield significant improvement judging from comparison of R2s and AICs (Burnham and Anderson, 2002). The R2 for Hispanic increased to 0.995 in the MGWR model from 0.969 in the
global model and the AIC decreased to -28.707 in the MGWR model from 2.959 in the global model; The R2 for blacks increased to 0.718 in the MGWR model from 0.437 in the global model and the AIC decreased to 219.757 in the MGWR model from 252.068 in the global model; and The R2 for whites increased to 0.746 in the MGWR model from 0.48 in the global model and the AIC decreased to 335.372 in the MGWR model from 397.827 in the global model (Table 5).

Table 5 Model fit metrics global regression, and multiscale geographically weighted regression (MGWR)

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>Global</th>
<th>MGWR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R2</td>
<td>AIC</td>
</tr>
<tr>
<td>Hispanics</td>
<td>0.969</td>
<td>2.959</td>
</tr>
<tr>
<td>Blacks</td>
<td>0.437</td>
<td>252.068</td>
</tr>
<tr>
<td>Whites</td>
<td>0.48</td>
<td>397.827</td>
</tr>
</tbody>
</table>

The MGWR results (table 6) support the trend that neighborhoods with higher crime rate and middle-aged population are associated with higher rates of opioid overdose death. These determinants may, therefore, be ideal to focus on if the goal of a policy is to have a broad impact across a study area. Inequality of Household Income is an impactful factor on opioid overdose death among Hispanic and white Americans. Opioid availability and prevalence of Disability have positive association with opioid overdose death among white and black Americans, in contrast, neighborhoods with high percentage of European American or African American families with young kids were more successful at managing the opioid crisis. Living in a stable neighborhood with higher rates of access to healthcare is linked with lower opioid
overdose death among people of color. Marital status, population density, and Alcohol availability are impactful factor on opioid overdose death among Hispanic Americans; college enrollment, and history of physical pain have significant impact on opioid overdose death among African Americans; and Naloxone availability and healthcare services accessibility have influence on white opioid overdose death rate. Although Black Americans are no more likely than Whites to use illicit drugs, they are 6–10 times more likely to be incarcerated for drug offenses (Netherland and Hansen, 2017). Incarceration rate is an important determinant of opioid overdose death among European Americans. The characterization of this determinant is not clear and requires further investigation. These factors might, thus, be ideal to focus on if the policy’s aim is to have an impact across specific neighborhoods in the study area.

Table 6 regional determinant of opioid overdose death among different races/ethnicities (The bandwidth is expressed as the number of neighboring census tracts)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bandwidth</th>
<th>Effective # parameters</th>
<th>t-value (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>172</td>
<td>1.376</td>
<td>2.11</td>
</tr>
<tr>
<td>age</td>
<td>172</td>
<td>1.364</td>
<td>2.106</td>
</tr>
<tr>
<td>households with child</td>
<td>172</td>
<td>1.364</td>
<td>2.106</td>
</tr>
<tr>
<td>Naloxone availability</td>
<td>43</td>
<td>8.753</td>
<td>2.799</td>
</tr>
<tr>
<td>crime rate</td>
<td>138</td>
<td>2.278</td>
<td>2.312</td>
</tr>
<tr>
<td>opioid availability</td>
<td>164</td>
<td>1.629</td>
<td>2.179</td>
</tr>
<tr>
<td>Disability</td>
<td>172</td>
<td>1.347</td>
<td>2.101</td>
</tr>
<tr>
<td>Variable</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Inequality of household income</td>
<td>45</td>
<td>8.898</td>
<td>2.804</td>
</tr>
<tr>
<td>Medically Underserved Area</td>
<td>172</td>
<td>1.153</td>
<td>2.035</td>
</tr>
<tr>
<td>Incarceration rate</td>
<td>75</td>
<td>5.199</td>
<td>2.619</td>
</tr>
<tr>
<td>Blacks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.27</td>
<td>1.157</td>
<td>2.14</td>
</tr>
<tr>
<td>Neighborhood instability</td>
<td>5.75</td>
<td>1.007</td>
<td>2.072</td>
</tr>
<tr>
<td>Age</td>
<td>5.75</td>
<td>1.007</td>
<td>2.072</td>
</tr>
<tr>
<td>Divorce rate</td>
<td>5.75</td>
<td>1.008</td>
<td>2.073</td>
</tr>
<tr>
<td>Crime rate</td>
<td>5.75</td>
<td>1.008</td>
<td>2.073</td>
</tr>
<tr>
<td>Population density</td>
<td>0.36</td>
<td>4.782</td>
<td>2.788</td>
</tr>
<tr>
<td>Alcohol availability</td>
<td>5.75</td>
<td>1.005</td>
<td>2.071</td>
</tr>
<tr>
<td>Healthcare accessibility</td>
<td>5.75</td>
<td>1.011</td>
<td>2.074</td>
</tr>
<tr>
<td>Inequality of household income</td>
<td>0.52</td>
<td>2.869</td>
<td>2.562</td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>98</td>
<td>1.258</td>
<td>2.084</td>
</tr>
<tr>
<td>Age</td>
<td>77</td>
<td>2.424</td>
<td>2.353</td>
</tr>
<tr>
<td>Households with a kid under 7</td>
<td>98</td>
<td>1.523</td>
<td>2.165</td>
</tr>
<tr>
<td>College enrollment rate</td>
<td>98</td>
<td>1.504</td>
<td>2.159</td>
</tr>
<tr>
<td>Crime rate</td>
<td>98</td>
<td>1.272</td>
<td>2.089</td>
</tr>
<tr>
<td>healthcare accessibility</td>
<td>77</td>
<td>2.6</td>
<td>2.38</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>opioid availability</td>
<td>98</td>
<td>1.315</td>
<td>2.103</td>
</tr>
<tr>
<td>disability</td>
<td>98</td>
<td>1.519</td>
<td>2.164</td>
</tr>
<tr>
<td>Houses built prior to 1950</td>
<td>98</td>
<td>1.501</td>
<td>2.159</td>
</tr>
<tr>
<td>public physical health</td>
<td>98</td>
<td>1.438</td>
<td>2.141</td>
</tr>
</tbody>
</table>

These findings are generally consistent with the results of the global model, but spatial variance in the parameter estimations of regional MGWR supports the possibility of interventions targeting particular communities. For instance, if funding or time limitations require resources to be allocated only to a small number of neighborhoods, then it is perhaps wise to concentrate on determinants in areas where there is a verified relationship. Similarly, in the future, a policy campaign that is directed at an area without a verified relationship may be tested for effectiveness to see whether the relationship grows over time or not.

To investigate the impact of the pandemic on OODs, I examined monthly OODs from January 1, 2017 - December 31, 2020. The stay-at-home order issue date, March 23, 2020, was used to divide the study into pre-intervention (pre-pandemic) and post-intervention (pandemic) periods. The average monthly OODs increased by approximately 12. The peak monthly overdose death total during the pre-intervention stage was 37. However, in the post-intervention period, Milwaukee County experienced a peak of 57 OODs per month and a minimum of 23 monthly OODs. These differences suggest that factors related to the pandemic increased OODs in Milwaukee.
In Table 7, the numbers and percentages of OODs prior to and after the March 23rd, 2020 “stay-at-home” directive, disaggregated by mode of death, toxicology/cause of death, gender, race, geographic discordances (when overdose incident location differs from the residential address), and age, are presented. In line with OOD patterns in other U.S. cities (Glober et al., 2020), I found that OODs increased significantly in the post-intervention stage in Milwaukee County. Simultaneously, I find that the principal factors (race, gender, age, mode of death) shaping OODs remained consistent in the pre- and post-intervention stages. For example, the primary mode of death in both periods is accidental, but the percentage shifted slightly from the pre (96%) to the post (95%) stage. Males are the primary victims in both periods, but the female OODs slightly declined from 32% to 29%. In terms of race/ethnicity, the OOD rate was highest among White community members (66%) followed by Black (21%) and Hispanic (9%) community members during the pre-intervention stage. These numbers slightly increased for Black community members (25%) in the post-intervention period but slightly declined among the White (63%) and Hispanic community members (8%). Age groups between 30-40 (26%) and 50-60 (23%) were most affected by OODs in the pre-intervention period, followed by the 40-50 (21%) age group. These rates remained steady in the post-intervention stage, but OODs in the 0-30 age group showed a slight decline from 20% to 13%. I also calculated geographic discordance to see if the county’s OODs primarily affected Milwaukee County residents or outsiders who travel into the county. In both stages, OODs primarily affected county residents.

Table 7 Opioid Overdose Deaths in Milwaukee County disaggregated by mode, drug evident in the toxicology report, cause, gender, race, geographic discordance, and age.
<table>
<thead>
<tr>
<th></th>
<th>Pre-intervention</th>
<th></th>
<th>Post-intervention</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Percentage</td>
<td>Count</td>
<td>Percentage</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undetermined</td>
<td>15</td>
<td>1.49%</td>
<td>7</td>
<td>1.98%</td>
</tr>
<tr>
<td>Accident</td>
<td>978</td>
<td>96.83%</td>
<td>336</td>
<td>95.18%</td>
</tr>
<tr>
<td>Suicide</td>
<td>15</td>
<td>4.49%</td>
<td>10</td>
<td>2.83%</td>
</tr>
<tr>
<td>Homicide</td>
<td>2</td>
<td>0.20%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>Drug Evident in Toxicology Report</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fentanyl</td>
<td>644</td>
<td>63.76%</td>
<td>311</td>
<td>88.10%</td>
</tr>
<tr>
<td>Heroin</td>
<td>431</td>
<td>42.67%</td>
<td>70</td>
<td>19.83%</td>
</tr>
<tr>
<td>Other Opioids</td>
<td>304</td>
<td>30.10%</td>
<td>136</td>
<td>38.53%</td>
</tr>
<tr>
<td>Cocaine</td>
<td>373</td>
<td>36.93%</td>
<td>121</td>
<td>34.28%</td>
</tr>
<tr>
<td>Ethanol</td>
<td>160</td>
<td>15.8%</td>
<td>70</td>
<td>19.83%</td>
</tr>
<tr>
<td>Amphetamines</td>
<td>54</td>
<td>5.35%</td>
<td>27</td>
<td>7.65%</td>
</tr>
<tr>
<td>Sedatives</td>
<td>235</td>
<td>23.27%</td>
<td>90</td>
<td>25.50%</td>
</tr>
<tr>
<td>Gabapentin</td>
<td>56</td>
<td>5.54%</td>
<td>46</td>
<td>13.03%</td>
</tr>
<tr>
<td>Synthetic Cannabinoid</td>
<td>13</td>
<td>1.29%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>690</td>
<td>68.31%</td>
<td>252</td>
<td>71.39%</td>
</tr>
<tr>
<td>Female</td>
<td>320</td>
<td>31.68%</td>
<td>101</td>
<td>28.61%</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Number</td>
<td>Percentage</td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
<td>------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>Hispanic</td>
<td>91</td>
<td>9.01%</td>
<td>30</td>
<td>8.50%</td>
</tr>
<tr>
<td>Black</td>
<td>219</td>
<td>21.68%</td>
<td>89</td>
<td>25.21%</td>
</tr>
<tr>
<td>White</td>
<td>676</td>
<td>66.93%</td>
<td>223</td>
<td>63.17%</td>
</tr>
<tr>
<td>Multiracial</td>
<td>6</td>
<td>0.59%</td>
<td>4</td>
<td>1.13%</td>
</tr>
<tr>
<td>Native American</td>
<td>12</td>
<td>1.19%</td>
<td>5</td>
<td>1.42%</td>
</tr>
<tr>
<td>Asian Pacific Islander</td>
<td>5</td>
<td>0.50%</td>
<td>2</td>
<td>0.57%</td>
</tr>
<tr>
<td>Eastern Indian</td>
<td>1</td>
<td>0.10%</td>
<td>0</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geographic Discordance</th>
<th>Number</th>
<th>Percentage</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>280</td>
<td>27.72%</td>
<td>88</td>
<td>24.92%</td>
</tr>
<tr>
<td>NO</td>
<td>730</td>
<td>72.27%</td>
<td>265</td>
<td>75.07%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Number</th>
<th>Percentage</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-30</td>
<td>201</td>
<td>19.90%</td>
<td>45</td>
<td>12.75%</td>
</tr>
<tr>
<td>30-40</td>
<td>265</td>
<td>26.24%</td>
<td>105</td>
<td>29.75%</td>
</tr>
<tr>
<td>40-50</td>
<td>222</td>
<td>21.98%</td>
<td>72</td>
<td>20.40%</td>
</tr>
<tr>
<td>50-60</td>
<td>229</td>
<td>22.67%</td>
<td>87</td>
<td>24.65%</td>
</tr>
<tr>
<td>60 and more</td>
<td>93</td>
<td>9.21%</td>
<td>44</td>
<td>12.46%</td>
</tr>
</tbody>
</table>

In terms of drug involvement in overdose cases, fentanyl-related OODs increased by 25 and heroin-related OODs decreased by 23. As depicted in Figure 5, fentanyl-related deaths comprised a greater proportion of overall monthly OODs, and heroin-related deaths declined.
Next, I tested for dependence among monthly opioid overdose death counts using the Box-Ljung test. Using a standard value of risk of $\alpha$ (confidence interval) = 0.05 (95% confidence interval), the Box-Ljung test was conducted to investigate whether autocorrelations of OOD time series are different from zero (null hypothesis: “the monthly OODs are independently distributed”; alternative hypothesis: “the monthly OODs exhibit serial correlation). Our results (X-squared = 13.614, df = 1, p-value = 0.0002245) indicate that monthly opioid OODs were dependent on each other. Because of this dependency, I next applied the BSTS model and forecasted a monthly OOD rate for the post-intervention period if pandemic had not taken place and compared the model-estimated mortality rate trend with the actual observed OOD rate in the postintervention period.
As seen in Figure 6, during the post-intervention period, the average number of monthly OODs was 38.00. In the absence of the pandemic and “lockdown”, I would have expected average monthly OODs of 26.29 (95% confidence interval of [22.13, 30.28]). Subtracting this prediction from the observed monthly OOD yields an estimate of the causal effect of various factors arising from the pandemic had on the monthly OOD. This effect is 11.71 with a 95% interval of [7.72, 15.87]. Summing up the individual data points during the post-intervention period (March 23, 2020, to December 31, 2020), there were 342 OODs. Had the pandemic not occurred, the predicted number would be 236.59 with a 95% confidence interval of 199.21 to 272.52. Thus, the monthly opioid overdose death showed an increase of +45% (the 95% confidence interval is +29% and +60%). These results show that the positive trend observed
after March 23, 2020, is statistically significant. The probability of obtaining this effect by
chance is very small (Bayesian one-sided tail-area probability $p = 0.001$).

On May 13, 2020, the Wisconsin Supreme Court struck down 'Stay-At-Home' orders as
unconstitutional (Hageman et al., 2020). Between the stay order and the supreme court ruling,
there was an increase in monthly OODs. However, this increase persisted well after the court
ruling, indicating the high impact of the pandemic upon the increase in substance abuse. In fact,
when we examine the time period between the stay-at-home order and the end of 2020, I find
that for several months, monthly OODs remained above the 99% confidence bounds of the
BSTS forecast (Figure 6), highlighting the importance and impacts of unobserved confounding
factors (e.g., Pandemic induced socio-economic impacts and socio-political changes).

Next, I examined spatial variation in OODs in Milwaukee County from 2017-2020.
Milwaukee's reputation as one of the most racially segregated metropolitan areas in the
country is a defining factor influencing the region's socioeconomic, political, and cultural
environments (Frey, 2012).

To obtain OOD estimate rates and map the underlying risk for OODs within each census
tract, I used a spatial empirical Bayes smoothing technique. The parameters of the prior
distribution were obtained empirically from the Milwaukee County Medical Examiner Office
dataset. The posterior distribution was obtained by combining the Poisson likelihood with the
prior distribution. This resulted in a distribution of possible values of the underlying risks of the
census tracts that were conditional upon the observed number of OODs. The empirical Bayes estimate of the true risk was taken as the mean of the posterior distribution.

Figure 7 Spatial distribution of the underlying OOD risks at the census tracts level from 2017 to 2020 (OOD is expressed in rate of Overdose per Census Tracts Population)
Figure 7 displays the distribution of the underlying OOD risks at the census tract level in Milwaukee County from 2017 to 2020. Of the 296 census tracts examined, a persistently high OOD rate was observed in 6% of them; these are urban census tracts with predominantly young (31 years), Black (39%) and Hispanic (24%) populations, with low educational attainment (10.63% hold a bachelor’s degree or higher) and low median household income ($36,351). The incarceration rate on average is higher (5%) and 10% percent of residents do not have health insurance (Policymap.com). Only 69% of households in these census tracts have internet subscriptions, limiting their access to information. About two-third (63%) of the census tracts with persistently high OOD rates are designated as Medically Underserved Areas (US Department of Health & Human Services, 2021), as characterized by high infant mortality, high poverty, and few primary care providers. A comparison with Figure 4 shows that the OOD distribution aligns with racial segregation patterns. About half (53%) of these communities have experienced concentrated persistent poverty for decades; the health and well-being of residents have been persistently low.

In contrast, 3% of census tracts exhibited persistently low OOD rates. These census tracts are suburban and more affluent, occupied by a predominantly White (83%) and middle-aged (41 median age) population with a higher median household income ($87,079), higher educational attainment (34% hold bachelor’s degree or higher), and relatively high access to health-related resources (99% of the residents have access to health insurance and health care). Other measures of social well-being are evident: 85% of households have an internet subscription, and the average incarceration rate is low (3%).
Overall, pre-pandemic (2017-2019) OOD patterns in Milwaukee County indicate that the opioid crisis has disproportionately impacted historically marginalized Black and Hispanic communities living in central Milwaukee neighborhoods (Figure 7) suggesting that the policies, resources, and interventions that are being designed and implemented in Milwaukee County are primarily benefiting White communities, not communities of color. The low impact of policy interventions in Black and Hispanic communities has further widened their pre-existing health inequalities.

The 2020 map in figure 8 indicates that the pandemic’s effect on OODs has been highest among Milwaukee’s Black and Hispanic communities. But the map also shows a spike in OODs in suburban tracts, which were previously minimally affected by OODs.

In line with prior studies (Linas et al., 2021; Rodda et al., 2020), our interrupted time series analysis showed that OODs have significantly increased during the pandemic. To examine the geographic impact of the pandemic on the overdose crisis in Milwaukee County, differences between the 2020 OOD rate and the average of OOD rate from 2017 to 2019 were calculated for each census tract in Milwaukee County and mapped. Figure 8 shows that 199 census tracts (2/3 of all tracts) in Milwaukee County experienced an increase in the number of OODs relative to pre-intervention levels (2017 to 2019). Geographically, these census tracts are distributed throughout the county, indicating that the pandemic impacted a variety of communities in Milwaukee County.

For all census tracts, OOD increase percentages were calculated to examine the incremental changes in OODs after the pandemic in comparison to the 5-year average number of cases. Results show that OOD rates in two areas were particularly impacted: predominantly
poor, Black neighborhoods in the inner city and predominantly affluent White suburban census tracts.

It is notable that there are areas where the death rates remained steady, showing little percentage change over time (this includes tracts with previously high OOD rates). Importantly, due to relatively low pre-intervention OOD frequency, many suburban tracts displayed sizeable percent increases that do not reflect large increases in the total number of OODs.

The urban tracts that experienced a 10280% increase in OODs suffer from racial and economic segregation and experience concentrated, persistent poverty. The population in these tracts is 72.46% Black (11.52% White, 16.02% other minorities) with a median age of 25.77, a 34% unemployment rate, and only 5% holding a bachelor’s degree or higher with a low median household income ($31,192), 69% internet subscription rate and relatively high (7%) incarceration rate. The effect of the pandemic upon the residents of the urban tracts has further widened pre-existing socio-economic disparities.

In contrast, the suburban census tracts that experienced an 11600% increase are affluent and educated. The population is 83.63% White (6.29% Black, 10.08% other minorities) with a median age of 33.53, 11% unemployed, 45% holding a bachelor’s degree or higher and a median household income is $75,959. Other social well-being indicators include a 90% internet subscription rate and low (2%) incarceration rate. Despite high economic and social well-being, suburban tracts have nonetheless experienced pandemic-related duress, likely contributing to OODs.
Figure 9 shows the spatial patterns of increase changes in fentanyl-, heroin- and other opioid-related OODs. In Milwaukee County, there was a disproportionate increase in OODs involving fentanyl among urban Black and Hispanic community members. However, deaths involving heroin declined significantly, perhaps due to supply chain disruptions (Tyndall, 2020).

These shifts have occurred nationwide; Black Americans are experiencing the highest rate of increase in fentanyl related OODs (Spencer et al., 2019; Althoff et al., 2020; Wu et al., 2021).
To investigate where opioid overdose victims overdose and what journey they go through, I employed the journey to overdose framework. The journey to overdose is the distance between the victims' residence and the location of the overdose incident. Scholars have argued that when it comes to the journey to crime, offenders travel relatively short distances to commit a crime, or in other words, the distribution of journey distances follows a distance decay distribution (Block & Bernasco, 2009; Townsley & Sidebottom, 2010; Johnson et al., 2013; Levine & Lee, 2013). I used Open Source Routing Machine (OSRM) to calculate the distance and travel time between decedent residences and overdose incident locations. The algorithm finds the optimal route by car, bicycle, or foot on the OpenStreetMap road network (Huber & Rust, 2016). In Milwaukee County, I found that 26.72% of all overdose cases are geographically discordant, and among those, on average, opioid crisis victims traveled 24.93 Km.

Figure 9 Spatial patterns of fentanyl-, heroin- and other opioid-related deaths, March 23 to December 31, 2020 (increments relative to pre-intervention numbers)
(22 mins) to their overdose incident location. On a median, 8.91 Km (11 Mins). Only 10.56% traveled greater than 50 miles, and the farthest anyone traveled was 310 miles. Figure 10 shows the distribution of distance traveled using Jenks natural breaks optimization (Jenks, 1963). The histogram of distance distributions indicates that the number of victims only decreases as distance increases. Therefore, the distance decay function appears to hold true for the journey to overdose.

Next, I constructed a U.S. Census tract centered aggregate social network based on the dyadic (pairwise) relationship of the journey to overdose between victims' residences and

![Figure 10 Overdose deaths and overdose death centrality in Milwaukee County, WI. Map in Figure 10A shows overdose death locations and census tract hotspots between 2017 and 2020. Each point represents an overdose death location. Warmer colors reflect higher numbers of overdose deaths in the census tracts during the study period. Map in Figure 10B shows overdose journeys between 2017 and 2020 and the degree centrality of census tracts defined using directed spatial social network analysis according to the number of edges for each node/census tract. Each line represents an overdose journey; arrows indicate directionality. Census tracts are colored based on network-weighted degree centrality. Warmer colors reflect higher numbers of edges (journey to or from the census tract).]
overdose incident locations. The aggregate social network derived from the opioid overdose death database included 375 nodes; connected by 426 weighted, directed edges. Census tracts' degree ranges from 1 to 16, with the average weighted degree of 1.139 (Figure 11).

Next, I implemented The HITS algorithm on the journey to overdose aggregate social network. The HITS algorithm's certain assumption is that an important authority (geographically discordant overdose focal point) is pointed by significant hubs (residence of geographically discordant overdose victims) and that an important hub points to many important authorities. Therefore, the importance of both an authority census tract and a hub census tract can be computed iteratively till results converge. Each census tract's authority value and hub value represent the rated geographically discordant overdose importance value and the rating importance value of each peer, respectively. Adopting the HITS algorithm, I use the hub and
authority centralities to pinpoint hubs and focal points of geographically discordant overdoses (Figure 12).

As shown in Figure 12, the focal points of geographically discordant overdoses (authorities) are concentrated around the central region of Milwaukee County, and original residences (hubs) of the journey to overdoses are mostly in the southern region of Milwaukee County. Demographically, predominantly Hispanic (33%) census tracts have widely imported overdose incidents from census tracts with a predominantly white population (71%), with about 8 years younger population, higher educational attainment, and higher median

Figure 12 Hub and authority communities for overdose journeys in Milwaukee, WI. Spatial social network analysis (HITS algorithm) was used to identify hubs (census tracts that are focal points of geographically discordant overdoses) and authorities (communities of residence from which journeys to overdose commonly begin) for overdose journeys during the study periods. Census tracts assigned scores based on hub centrality are depicted in the map in Figure 3A. Warmer colors indicate census tracts with higher hub scores. Census tracts assigned scores based on authority centrality are depicted in the map in Figure 3B. Warmer colors indicate census tracts with higher authority scores.
household income. In table 8, socioeconomic characteristics of the focal point of geographically discordant overdoses or top authorities and also hubs of original residences of the journey to overdoses are presented.

<table>
<thead>
<tr>
<th></th>
<th>Renter-occupied</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Median age</th>
<th>Single households</th>
<th>Educational attainment</th>
<th>Below Poverty Level</th>
<th>Median Household Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authorities</td>
<td>61.46%</td>
<td>50.68%</td>
<td>11.55%</td>
<td>33.02%</td>
<td>32.45</td>
<td>48.41%</td>
<td>11.60%</td>
<td>25.99%</td>
<td>39636.8</td>
</tr>
<tr>
<td>Hubs</td>
<td>46.66%</td>
<td>71.18%</td>
<td>12.13%</td>
<td>13.11%</td>
<td>40.09</td>
<td>39.01%</td>
<td>16.01%</td>
<td>17.72%</td>
<td>48802.3</td>
</tr>
</tbody>
</table>

Next, to unveil hotspots of imported drug overdoses, we implemented the edge scan methodology developed by Baker et al. (2020) to find hotspots where census tracts are not only host clusters of overdose cases but also connected to other census tracts through the journey to overdose. On the other hand, in order to identify temporal trends in the journey to overdoses, and to investigate in what census tracts the number of geographically discordant overdoses is intensifying or diminishing; we created space-time cubes for each census tract and evaluated the temporal trends using the Mann-Kendall trend test (McLeod, 2005).

As depicted in Figure 13, the hotspots of imported drug overdoses are concentrated around the central and southern regions of Milwaukee County. However, in terms of temporal trends, consecutive hotspots of imported drug overdoses are located in central and northern regions of Milwaukee County, while sporadic and emerging hotspots are also being developed in the vicinity of consecutive hotspots of the journey to overdoses.
Eventually, to investigate differentiating demographic and involved drug characteristics of geographically discordant and non-discordant overdoses, in Table 9, the percentages of domestic and imported overdoses, disaggregated by mode of death, toxicology/cause of death, gender, race, and age, are presented. Then, we performed 2 sample t-tests to detect differentiating factors of geographically discordant and non-discordant overdoses.

Table 9 journey to overdose in Milwaukee County disaggregated by mode, drug evident in the toxicology report, cause, gender, race, and age (significance: 99%***, 95%**, 90%*)
<table>
<thead>
<tr>
<th></th>
<th>Geographic Discordance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td></td>
</tr>
<tr>
<td>Undetermined</td>
<td>2.05%</td>
</tr>
<tr>
<td>Accident</td>
<td>95.90%</td>
</tr>
<tr>
<td>Suicide</td>
<td>1.82%</td>
</tr>
<tr>
<td>Homicide</td>
<td>2.05%</td>
</tr>
<tr>
<td><strong>Drug Evident in Toxicology Report</strong></td>
<td></td>
</tr>
<tr>
<td>Fentanyl</td>
<td>64.46%</td>
</tr>
<tr>
<td>Heroin</td>
<td>32.57%</td>
</tr>
<tr>
<td>Cocaine</td>
<td>44.42%</td>
</tr>
<tr>
<td>Ethanol</td>
<td>17.08%</td>
</tr>
<tr>
<td>Amphetamines</td>
<td>8.43%</td>
</tr>
<tr>
<td>Sedatives</td>
<td>19.82%</td>
</tr>
<tr>
<td>Gabapentin</td>
<td>7.31%</td>
</tr>
<tr>
<td>Synthetic Cannabinoid</td>
<td>1.37%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>69.48%</td>
</tr>
<tr>
<td>Female</td>
<td>30.52%</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>8.20%</td>
</tr>
</tbody>
</table>
Based on our statistical analysis, I found that there is no significant difference in racial and gender composition between geographically discordant and domestic overdoses. Nonetheless, individuals who travel to a new place and commit fatal overdose are significantly younger and less suicidal. On average, victims of domestic overdoses are 44.57 years old; however, the average age of victims of imported overdoses is 40.92 years old. In terms of drugs involved in the overdose cases, fentanyl, cocaine, and amphetamines involved overdoses are significantly higher in imported cases, highlighting the fact that individuals are traveling to procure and use fentanyl, cocaine, and amphetamines in other census tracts and eventually committing overdose.

Discussion

As metropolitan counties struggle to respond to the opioid crisis, it has become evident that new strategies are needed to better define the problem and its underlying causes (Ruhm,
A significant challenge to address the crisis is that influential factors can vary significantly across neighborhoods/communities. As a result, many interventions and policies lack universal effectiveness. This is particularly problematic in cities such as Milwaukee. Indeed, Milwaukee is one of the most ethnically diverse cities in the United States and, at the same time, it is one of the most racially segregated (Frey, 2012, Florida and Mellander, 2015). Milwaukee is also among the U.S. metropolitan areas experiencing the greatest increase in the number of overdose deaths (COPE, 2019). Thus, Milwaukee County provides a unique opportunity to study the intersection among socioeconomic factors, factors related to race and discrimination, and the opioid epidemic.

GIS-based mapping of data is a powerful approach that permits inference of complex interactions among variables based on their temporal-spatial relationships. Geospatial approaches have been previously used to examine factors that influence opioid crisis (Sadler and Furr-Holden, 2019; Schneider et al., 2020; Nesoff et al., 2020; Donnelly et al., 2020). However, prior approaches have been limited as they are based on assumption that all modeled processes operate at the same spatial scale. This creates interpretational challenges and limits the ability to make accurate inferences about relationships that could guide community targeted policies and interventions (Fotheringham et al., 2017; Oshan et al., 2019). MGWR allows a more flexible exploration of the relationships at varying spatial scales. To our knowledge, this is the first published multi-scalar investigation of the opioid crisis.

Our MGWR model identified eight factors that influenced opioid overdoses in Milwaukee County in 2019 at a global (i.e., county-wide) scale. "Disability," "neighborhood instability," and "total dispensed opioid drugs (2006-2014)" had positive associations with
opioid overdose deaths across the whole study area. Overall, these findings are consistent with others suggesting that higher rates of overdose are associated with more impoverished socioeconomic conditions (Cerda et al., 2013; Sadler and Furr-Holden, 2019; Cobert et al., 2020) and over-prescription of opioid medications (Stopka et al., 2019; Cerda et al., 2020). The factors "educational attainment," "access to healthcare," "households with kids," and "percentage of renters spending 30% of their income on rent" all had negative associations with overdose deaths. As the first three factors are all indicators of community stability, their negative relationships with overdose risk are not surprising. By contrast, the negative association between the "percentage of renters spending 30% of their income on rent" and overdose rates was unexpected and will require further investigation.

Two factors displayed regional spatial variation (Figure 2). The “Crime Rate” influenced overdose deaths in southern and western Milwaukee County but had little influence in northeastern Milwaukee County. "Inequality of household income" was associated with overdose deaths in southern Milwaukee County, but not in other regions. Surprisingly, "naloxone availability" parameter estimates displayed local spatial variation (Figure 3). Timely administration of naloxone can prevent opioid overdose deaths, and naloxone has been used by emergency medical personnel to pharmacologically reverse overdoses for decades (Kim et al., 2009). Since 1996, community-based programs have provided naloxone to opioid users, their communities, and service/care providers to reverse potentially fatal opioid overdoses (Wheeler et al., 2012). As expected, "naloxone availability" had a strong negative association with the opioid overdose death rate in affluent, predominantly White and suburban census tracts. In contrast, in census tracts with high population density or predominantly African
American or Hispanic census tracks, "naloxone availability" parameter estimates are either not significantly different from zero or, in Hispanic neighborhoods, have a positive association with "opioid overdose death rate." The reasons for the observed spatial variations are unclear and understanding them will be critical for guiding neighborhood-level policies and practices.

Potential contributing factors include the types of opioids or drug combinations used, hesitancy to engage agencies due to fear of incarceration, lack of adequate educational programs, drug use in isolation versus with peers, the presence of overdose follow-up programs, and barriers to naloxone access (versus availability).

Racial segregation has such a profound effect in Milwaukee County that it is perceivable that residents are living in three different worlds. For example, the overall Milwaukee County crime rate in 2019 was 42.9 per 1000 residents; in predominantly White communities, it was 18.1 per 1000 residents. However, the crime rate in predominantly African American communities was 83.9 per 1000 residents, while in predominantly Hispanic communities it was 52.4 per 1000 residents. Thus, in a hyper-segregated metropolitan area such as Milwaukee, analyzing the opioid crisis with respect to segregation and structural inequalities is needed to understand different layers of the epidemic. Accordingly, many associations with overdose deaths appeared to vary with the racial/ethnic composition of communities. To more closely investigate this, I divided Milwaukee County into three regions: census tracts in which White (non-Hispanic) residents are the majority, tracts in which African American residents are the majority, and tracts in which Hispanic residents are the majority (Figure 4). Indeed, I found that the accuracy of geospatial modeling improved markedly when I established independent models for each of these communities.
The relationships of some factors with overdose deaths were similar across tracts. For example, communities with higher crime rates or those comprised predominantly of middle-aged populations had higher rates of opioid overdose deaths, regardless of racial/ethnic composition. In other cases, relationships varied across census tracts. Positive associations between "inequality of household income" and overdose deaths were evident in the predominantly Hispanic and White communities, but not in the African American community. Prescription opioid availability (historically dispensed oxycodone) and the prevalence of disability had strong positive associations in both African American and White communities, but not within the Hispanic community. In contrast, the percentage of families with young children had strong negative associations with overdose death rates in the African American and White communities, but not in the Hispanic community. Living in a stable neighborhood with more access to health care was associated with lower overdose rates in the African American and Hispanic communities, but not in the White communities. Marital status (divorce or death of a spouse), population density, and alcohol availability were positively associated with overdose deaths in the Hispanic majority, but not among the African American or White communities. College enrollment was positively associated with overdose deaths, and a history of physical pain was negatively associated with opioid overdose deaths in the African American community but not in the predominantly Hispanic or White communities. Health care service accessibility influenced overdose rates in the predominantly White community, but not in the African American or Hispanic majority communities. The precise reasons for geographic variation in these relationships are unclear and require further investigation.
Perhaps the most surprising observation was that naloxone availability negatively influenced overdose in the White majority, but not within the African American or Hispanic majority community. In fact, in areas of Milwaukee that are predominantly Hispanic, there was a positive association between naloxone availability and overdose deaths, potentially reflecting community efforts to increase distribution, despite ineffectiveness in reducing overdose risk. The NARCAN Direct Program, supported by the Wisconsin Department of Health Services, is an integral component of the state's response to the opioid crisis and provides NARCAN at no cost to community agencies to distribute with the goal of reducing overdose deaths. While it is possible that community members are acquiring naloxone from other sources, these findings suggest that, despite naloxone availability, there are obstacles that are preventing its effective use in African American and Hispanic communities. These may include the willingness or ability of community members to engage the agencies through which naloxone is being provided and/or lack of awareness of the availability of naloxone and the benefits of its use. Alternatively, other factors, such as the types/potency of opioids or drug combinations involving non-opioids being used, could vary across communities.

Another surprising observation was related to the association between incarceration rates and overdose deaths. Unexpectedly, incarceration rates influenced overdose deaths in the majority White community but not within the African American or Hispanic communities. African Americans are no more likely than White Americans to use illicit drugs but are 6–10 times more likely to be incarcerated for drug offenses (Netherland and Hansen, 2017). While the threat of incarceration may be a deterrent to drug use and the risk for overdose death is negligible during incarceration, overall overdose rates are higher in former inmates
(Binswanger et al., 2007). Our findings raise questions about whether or not more aggressive policing and criminalization of drug possession in communities of color are effective measures for preventing overdose.

Overdose deaths are declining in White Americans but are on the rise in communities of color, particularly among African Americans (Furr-Holden et al., 2021). This trend is also evident in Milwaukee County (COPE, 2019). It raises concerns that the policies, resources, and interventions that are being designed and implemented are disproportionately benefiting White communities, thereby creating inequity and widening health disparities. Understanding how factors differentially influence overdose deaths across communities should guide targeted policies and approaches aimed at prevention, intervention, and harm reduction. While some of the observations may specifically apply to communities in Milwaukee, many of the relationships are likely evident in communities in other metropolitan areas.

The opioid crisis in the U.S. has profoundly worsened during the pandemic. Over 2/3 of census tracts in Milwaukee County experienced significant increases in OODs. The maximum number of monthly overdose deaths increased from 37 prior to the March 23rd, 2020, state-wide “stay-at-home” order in Wisconsin to 57 afterwards. This increase was consistent across spatially defined communities, across demographic variables (sex, age, race) and across mode of death, indicating that OODs have not only intensified in areas where they were previously evident but have spread to communities that were previously immune to the opioid crisis. Fentanyl has emerged as the prevalent drug involved in OODs, while heroin related OODs have declined. Overall, the involvement of other drugs (e.g., cocaine, ethanol, amphetamines,
sedatives, other opioids) has remained steady, although some variation is evident at the neighborhood scale.

From a spatial perspective, the number of deaths is higher in inner city tracts, where the opioid crisis has been prevalent for many years. The worst-affected areas are primarily Black/Hispanic communities shaped by decades of racial and economic segregation, with concentrated poverty and health inequalities. These vulnerable neighborhoods, with little access to health care, have limited resources to address opioid abuse, leading to a rise in OODs. However, the pandemic has also led to a surge in OODs in more affluent and educated White suburbs with historically high access to health care. I speculate that many of the resources that these communities previously enjoyed were diminished during the pandemic, reducing support for those with opioid use disorder and increasing risk for OODs. Some spatial variation in drug involvement in OODs is evident. Fentanyl is dominant in OODs in poor urban tracts, where there has been a decline in heroin related OODs. In contrast, suburban OODs more commonly involve fentanyl, heroin, and a mix of drugs. Further community-based research is needed to investigate the impact of pandemic on different communities in Milwaukee County so that appropriate policy interventions can be implemented.

Converging factors likely contributed to the increase in OODs during the pandemic. Increased exposure to opioid maintenance treatment reduces the risk of death in opioid-dependent people (Gibson et al., 2008). COVID-19 containment strategies are disrupting existing opioid use disorder (OUD) treatment paradigms, which have focused on in-person examinations, medication distribution, counseling sessions, and group therapy. Complying with social distancing guidelines, OUD treatment centers maximized telehealth, offered larger
amounts of take-home methadone, and limited in-person and inpatient treatments. This abrupt change in treatment practices introduced new risks to patients. Several studies have suggested that telehealth is beneficial to the patient, physician, and healthcare system in many ways (Brown, 2013, O’Gorman and Hogenbirk, 2015). Although telehealth may be an effective alternative to delivering in-person opioid agonist therapy, and it has the potential to expand healthcare accessibility (Eibl et al., 2017), it can further marginalize patients without access to computing devices or internet subscriptions. On the other hand, the unexpected determination to distribute larger amounts of methadone essentially treats all patients as though they were in a stable condition. Consequently, many patients are faced with the responsibility of a possibly fatal privilege that they did not earn and may not be prepared for (Leppla & Gross, 2020).

Increased public and political will has led to an exponential increase in the widespread implementation of harm reduction initiatives to minimize fatal opioid overdose (Hawk et al., 2015). Multiple harm reduction initiatives, such as targeted overdose education, naloxone distribution, and policies to improve bystander assistance during a witnessed overdose, have been employed by public health departments to reduce the morbidity and mortality of opioid use disorder. Although harm reduction initiatives have been facing several challenges, such as preferential focus on substance use treatment and primary prevention instead of harm reduction practice and community-level stigma against people who use drugs as well as the agencies supporting them (Childs et al., 2021); Covid-19 pandemic imposed tremendous pressure of public health departments which provide harm reduction interventions such as syringe service programs, or naloxone distribution. Consequently, they offer less support as they focus resources on responding to the COVID-19 pandemic.
Another possible reason for the significant increase in the number of opioid overdose death might be disrupted drug supplies due to the pandemic. Drug precursors supply chains interruption, probable shut down of abroad synthesis labs, and border closures and their possible effect on drug shipping are potential factors contributing to uncertainties in supply and the quality of the drugs being sold on the street (Tyndall, 2020). Therefore, as recommended by Sun et al. (2020) during COVID-19, authorities should make every effort to ensure accessibility and availability of opioid agonist treatment.

Research supports that the presence of a bystander or witness to the overdose event has been linked to improved outcomes (Ornato et al., 2020; McCann et al., 2021). The presence of a proximal bystander during an overdose event can lead to dispatch codes indicative of an overdose and shorter times to naloxone administration compared with those with distal bystanders, suggesting public education and engagement of overdose harm reduction strategies can reduce the number of fatal opioid overdoses. On the other hand, enforced shelter in home orders can isolate people when they use drugs and make it more likely that they will be alone while doing so, leading to limited emergency medical response access, restricting naloxone administration, and increased risk of fatal overdose.

All preexisting mental health conditions and treatments were strongly associated with higher rates of long-term opioid therapy (Quinn et al., 2018). In fact, Adolescents with anxiety, mood, neurodevelopmental, sleep, post traumatic stress disorder (PTSD), and nonopioid substance use disorders and most mental health treatments are significantly more likely to receive any opioid (Seal et al., 2012; Jerant et al., 2020). Studies showed that common psychiatric disorders like depression and anxiety could predict initiation and ongoing regular
use of opioids in patients with chronic pain leading to adverse outcomes such as declining functional status or opioid misuse (Sullivan et al., 2006). Isolation can have a severe influence on mental health, creating anxiety and depression, which can lead to drug abuse as a coping mechanism, and also increasing the chance of relapse in abstinent individuals; as a result, leading to an increase in the number of opioid overdose death during the covid-19 pandemic.

Based on our observation that, in Milwaukee County, 26.72% of overdose deaths are geographically discordant; I examined characteristics that define the journey to overdose. First, I applied spatial social network analysis using the HITS algorithm to identify hubs (i.e., census tracts that are focal points of geographically discordant overdoses) and authorities (i.e., the communities of residence from which journeys to overdose commonly begin) for overdose deaths and characterized them according to key demographics. Second, I used temporal trend analysis to identify communities in Milwaukee that were consistent, sporadic, and emergent hotspots for overdose deaths. Third, I identified characteristics that differentiated and, therefore, may serve as risk factors for discordant versus non-discordant overdose deaths. To our knowledge, this study is the first to examine the journey to overdose and provides proof of principle that this type of analysis can be applied in metropolitan areas to better understand and guide community responses to the opioid crisis.

On average, victims traveled 51.74 km (8.71 km on the median) to the site of overdose death. Only 13.07% traveled more than 50 km, and the farthest that anyone traveled was 521 km. Consistent with observations in other fields (e.g., Bernasco et al., 2013; Gill et al., 2017), as the distance of the journey increased, the number of overdose deaths generally decreased. Based on aggregated data (2017-2020), focal point communities involved in discordant
overdose deaths were identified using social network analysis. Overall, hubs (communities of residence) were more widely distributed than authorities (communities of overdose); eight census tracts served as the major authorities, while about two times as many census tracts were major hubs. Moreover, hub and authority communities were largely non-overlapping. Authority communities for overdose deaths had lower housing stability and were younger, more impoverished, and less educated relative to either hubs or the county-wide indicators. Overall, these observations are consistent with prior findings (Forati et al., 2021) demonstrating relationships between socioeconomic conditions and overdose deaths in Milwaukee. By contrast, hub values were comparable to those county-wide.

Hub communities had a much higher percentage of White residents relative to authorities or the county overall. Authority communities had a much higher percentage of Hispanic residents relative to hubs or the county overall. These findings suggest that a common journey to overdose is from predominantly White communities to predominantly Hispanic communities. Accordingly, in demographically defined Hispanic communities, the ratio of White overdose deaths victims to White community members is much higher than the ratio of Hispanic overdose death to Hispanic community members. This is consistent with 2021 data showing that fatal and nonfatal overdose rates in the Hispanic population were lower than in the White population in Milwaukee (Guadamuz et al., 2021; Wisconsin DHS, 2022). There may be several explanations for this observation. First, Mexican cartels have long been responsible for supplying cocaine and methamphetamine to Milwaukee via Chicago and have an increased role in trafficking fentanyl (Peterson et al., 2019). Thus, many Hispanic communities may be distribution sites for these drugs in Milwaukee. Second, there are barriers in Hispanic
communities that limit availability of, access to, and utilization of resources for harm reduction, intervention/treatment, and support (Cano, 2020). Further investigation is needed to identify factors that contribute to non-discordant overdoses in these communities. Interestingly, the percentage of Black residents in either hub or authority communities was similar and markedly lower than county-wide numbers. This is surprising, considering that, in line with national trends, both fatal and non-fatal overdose rates have increased precipitously for Black community members in Milwaukee and are now higher than for White community members. Overall, these data suggest that there is a more localized pattern of drug use in Black communities.

Next, I defined hotspots for imported and domestic overdoses in Milwaukee County and applied temporal trend analysis to determine which hotspots emerged during the period of study. Consecutive, sporadic, and new hotspots were identified based on temporal trends (for pattern definitions, refer to ArcGIS pro (2023)). Notably, while many of these emerging hotspots overlapped with communities categorized as authorities for discordant deaths using social network analysis, this analysis revealed communities that displayed temporal patterns of discordant overdoses and identified some communities as new hotspots that did not appear with our time aggregated network analysis. Overall, these findings provide proof of principle that temporal trend analysis can be used to identify hotspots, both existing and new, for imported overdose deaths. Assuming that such analyses can be conducted in a timely manner, the resulting data could guide harm reduction, resource allocation, education, and law enforcement efforts.
Finally, I defined differences in demographics and other characteristics between geographically discordant and non-discordant overdose deaths. Surprisingly, despite the differences in hub and authority community demographics, differences in racial composition were not observed. Similarly, while, overall, men were more likely to be overdose victims than women, the gender breakdown did not differ. Geographically discordant deaths more commonly involved fentanyl, cocaine, and amphetamines than non-discordant deaths, likely reflecting differences in drug supply/availability across census tracts and the need for many to travel to other communities to procure these drugs. By contrast, non-discordant overdoses more commonly involved opioids other than fentanyl or heroin. It is notable that many of these opioids (morphine, hydrocodone, oxycodone) are often procured through diversion and therefore are more likely to be available in communities of residence.

Geographically discordant overdose deaths were more likely to be accidental. This may be attributable to less familiar sources of drugs, the types of drugs being used (e.g., fentanyl), riskier modes and patterns of use, and the lack of availability of and/or access to overdose prevention resources. Geographically non-discordant overdose deaths were more likely to be the result of suicide. This is consistent with observations that most suicides occur at home (Kposowa & McElvain, 2006).

There are a number of research gaps and opportunities for future work that build on the present findings. First, while the list of analyzed factors was extensive, many factors of likely importance (e.g., types of opioids or other drugs that contributed to relapse, types of community programs and treatment approaches, mental health indicators) were not included. Third, I georeferenced overdoses based on the location of the overdose but not the victims'
addresses of residence. Third, there is a need to extend our analyses to regions beyond Milwaukee County so that we can identify and address factors that are differentially influential in suburban and rural communities. In addition to expanding our MGWR analysis, future work will also involve community engagement and qualitative assessments so that we can further understand these relationships prior to sharing data with policymakers and community organizations to support the targeted implementation of policies and initiatives. Fourth, our data collection did not include non-fatal OODs; studies have shown that the number of survived opioid overdoses also increased (Khatri et al., 2021; Ochalek et al., 2020) after the onset of the pandemic. Future research must examine the impact of the pandemic on the nonfatal overdoses across various communities and identify any disparities. Fifth, interrupted time series analyses can be susceptible to history bias due to other interventions or events occurring around the period of analysis (Ewusie et al., 2017; Lopez Bernal et al., 2018). Sixth, there is a need to examine the differential impacts of the COVID-19 pandemic on OODs in suburban and rural communities. Seventh, using distance traveled to define journeys is not always optimal (Gabor & Gottheil, 1984), especially in the context of drug purchasing; availability of illicit substances (Smart, 1980), and enforcement of drug laws (Beckett et al., 2005; Johnson et al., 2013) vary geospatially. Eighth, the demarcation of space is an inherently political process; borders between places are socially constructed, and social phenomena are rarely contained in such locales (Smith & Varzi 2000). Ninth, housing instability, including homelessness and "couch surfing," may result in inaccuracies regarding individuals' official residences and where they actually reside, potentially impacting the computation of authority and hub communities. Future analysis requires careful reflection on how to properly demarcate spatial nodes
(Uitermark & Van Meeteren, 2021; Gibadullina et al., 2021; Ye & Andris, 2021; Chen & Poquet, 2022). Last, I defined the characteristics of communities and individuals involved in overdose journeys according to a small subset of indicators. Further research is needed to analyze the factors that define the journey to overdose.

Conclusions

In conclusion, I report that the factors that influence opioid overdose deaths in Milwaukee County vary across the diverse communities in this highly segregated metropolitan area. The observed geographic variation in relationships includes the impact of naloxone availability and incarceration rates on overdose deaths with pronounced differences between White communities and communities of color. Understanding community-level factors that contribute to overdose risk should guide targeted community-level solutions. Overall, our findings demonstrate the value of precision epidemiology using MGWR analysis for defining and guiding responses to public health challenges. On the other hand, the impact of the COVID-19 pandemic on OODs in Milwaukee has been tremendous and has crossed demographic, socioeconomic, and neighborhood boundaries in this diverse and highly segregated metropolitan area. A better understanding of contributing factors is needed to guide interventions at the local, regional, and national scales. Eventually, understanding the features that define common journeys to overdose is important for guiding community responses. GIS-based analytical approaches can reveal important insight into the community and individual factors that contribute to the risk for geographically discordant overdoses and may provide
actionable data for community agencies and organizations as they formulate strategies for addressing the drug overdose crisis.

Coronavirus (COVID-19) related illness has been identified as a global pandemic by the World Health Organization (WHO). Known in the scientific communities as Severe Acute Respiratory Syndrome Coronavirus 2 or SARS-CoV-2, it was first identified in Wuhan, China in winter, 2019 (Linton et al., 2020; Xie & Chen, 2020). The virus is highly transmissible during the incubation period. Susceptibility to the virus is more common among those who have underlying conditions such as hypertension, diabetes, and heart disease (Xie & Chen, 2020). Combating the spread of virus through containment strategies was advocated by WHO and other public health agencies since the early 2020s. These containment strategies (the practice of using face masks, handwashing, social distancing, self-isolation, quarantine, enforced lockdowns) have been found to be effective in reducing transmissions, while vaccine research and deployment have been underway. As of February 23, 2021, COVID-19 has claimed 2.48 million lives globally, with USA ranking first in both the number of cases (28.2 million) and in the number of deaths (over 550,000) (Centers for Disease Control and Prevention -CDC, 2020).

Since the onset of the pandemic, a 'deluge' of misinformation and conspiracy theories have been disseminated through social media platforms, downplaying the severity of the virus and dismissing its high death toll. Polling from the early stages of the pandemic have shown that many Americans are misinformed about COVID-19. Results from a poll conducted by YouGov and The Economist in early March 2020 revealed that 49% of Americans believed the

---

2 Portions of this chapter have been published in the Applied Geography, coauthored with Prof. Ghose:
Corona virus to be manmade, 44% felt that the dangers of the virus was being exaggerated for political reasons, and 13% believed it was a hoax (Economist, 2020).

The effect of social media on public perception and awareness of COVID-19 is significant as 68% of American adults have reported to receiving news from social media. Further, 59% of Twitter users have found it to be an effective and reliable source for obtaining health information (Singh et al. 2020). Yet, health misinformation is a significant problem in all social media platforms, and the deluge of COVID-19 related misinformation through social media has been identified as ‘infodemic’ (Kulkarni, Prabhu, and Ramraj 2020). When asked about the effect of social media on public's response to COVID-19, Dr. Fauci (chief medical advisor to the President and the director of the National Institute of Allergy and Infectious Diseases) responded "It [social media] has impacted it more negatively than positively. One of the problems is when disinformation gets in there, it has a way of self-propagating itself to the point where you don't know what's true and what's not true" (Gander, 2020). Fueled by such misinformation, political and social resistance to containment measures and vaccinations remains high in the USA. A strong spatiotemporal relationship exists between Twitter misinformation flow and the occurrence of Covid-19 incidents (Singh 2020, Forati and Ghose 2021), highlighting the need for a nuanced investigation of misinformation.

This dissertation investigates the association between social media misinformation and spatial variability in the spread of the virus in USA. I focus on Twitter activity on COVID-19, as 69.3 million Americans use Twitter as their preferred social media platform. Twitter is one of the most popular social media platforms for sharing microblogs coined 'tweets,' which contain a short message and metadata. Geolocation information is available for some tweets showing
where a message is posted. Cao et al. (2015) highlighted the roles of social media as a proxy to understand human behaviors and complex social dynamics in geographic spaces. They developed a scalable computational framework to model extensive unstructured geotagged social media data for systematic spatiotemporal data analysis. This line of research enables us to ask questions of individual and collective behavior that would be very hard to answer without the ubiquity and scope of narrative data produced on Twitter.

This dissertation examines pandemic-related misinformation on Twitter using a mixed-method approach. I use geotagged tweets as this location-based data enables us to analyze the dynamic interactions between human behavior and the environment (Sui and Goodchild, 2011; Shaw et al., 2016). This dissertation (i) examines and categorizes false information narratives through manual, qualitative discourse analysis grounded in social theory, (ii) uses GIS to map and analyze the geographic distribution of false information (iii) uses spatial statistics to examine the spatial variations in Twitter misinformation categories.

Using the Tweepy API (Roesslein, 2009), I collected 84,864 Covid-19-related geotagged tweets from May 1 to July 31, 2020, using trending hashtags such as “corona,” “coronavirus,” “COVID,” “pandemic,” “lockdown,” “quarantine,” “hand sanitizer,” “ppe,” “n95”, “sarscov2”, “nCov,” “COVID-19”, “ncov2019”, “2019ncov”, “flattening the curve,” “social distancing,” “work from home” and the respective hashtag of all these keywords using filter: language “English.” Through manual checks and discourse analysis, multi-level categories are identified. First, I examined tweets as False Information Tactics and categorized them as (i) downplaying, (ii) disinformation, (iii) misinformation (iv) suspected bot activity. Tweets within the last category were eliminated, and the final dataset of 979 tweets was mapped. Next, I categorized the
tweets into four main topical themes: (i) politically charged, (ii) health-related, (iii) re-ligious (iv) freedom rhetoric. These four themes were further subcategorized into nine myths. Finally, through Geographically Weighted Principal Component Analysis (GWPCA) and MGWR, the spatial variations in themes and myths were examined. My research provides significant empirical and methodological insights into covid-19 misinformation. Empirically, these insights can be utilized for effective public policy formulation. Methodologically, this study integrates social theory with spatial analysis to advance mixed-method GIScience research.

Literature Review

Social media platforms have attracted increasing interest regarding human-environment interactions as refinement in spatially identified posts through advances in technology (Tsou 2015) have increased the availability of location-based data (Sui and Goodchild 2011) and, consequently, improved our abilities to show the dynamic interactions between behavior and the environment (Shaw et al. 2016). Twitter is one of the most popular social media platforms for sharing microblogs coined ‘tweets,’ which contain a short message and metadata. Georeferenced tweets contain not only the tweet content but also the tweet’s geolocation or point of origin.

Cao et al. (2015) highlight the roles of social media as a proxy to understand human behaviors and complex social dynamics in geographic spaces and provide a scalable computational framework to model extensive unstructured geotagged social media data for systematic spatiotemporal data analysis. Informed by such insights, I argue that social media data can be used to study high-impact events such as pandemics or hurricanes.
While the COVID-19 virus had been identified in China in late 2019, its arrival in the USA went undetected till January 19th, 2020, when the first COVID-19 case was confirmed in the state of Washington (Holshue et al., 2020). Thereafter, the virus spread rapidly across the USA, and on March 26th, 2020, the USA became the leading country in the number of cases worldwide. A year later, the USA continues to lead the world in terms of both cases and deaths.

Despite the high number of deaths, political and social resistance to containment measures have been profound in the USA, fueled by misinformation spread through social media. The effect of social media on public perception and awareness of COVID-19 is significant as 68% of American adults have reported to receiving news from social media. Further, 59% of Twitter users have found it to be an effective and reliable source for obtaining health information (Singh et al. 2020).

Yet, health misinformation is a significant problem in all social media platforms, and the deluge of COVID-19 related misinformation through social media has been identified as ‘infodemic’ (Kulkarni, Prabhu, and Ramraj 2020). When asked about the effect of social media on the public’s response to COVID-19, Dr. Fauci (chief medical advisor to the President and the director of the National Institute of Allergy and Infectious Diseases) responded, "It [social media] has impacted it more negatively than positively. One of the problems is when disinformation gets in there, and it has a way of self-propagating itself to the point where you don't know what's true and what's not true" (Gander, 2020). Experts have identified several factors that have led to “an unnecessarily brutal pandemic, including a lack of clear messaging from the country's leadership, state and local leaders loosening restrictions too quickly, large
holiday celebrations and continued resistance to wearing face masks or social distancing” (Maxouris, Yan and Vera, 2021).

A primary concern in the face of Covid-19 mitigation efforts is social unrest caused by the perceived loss of liberties arising from quarantine and shelter in place orders (Timms & Brüssow, 2020). Relational theorists Kasapoglu and Akbal (2020) framed COVID-19 in terms of 'uncertainties' in social relations. Such uncertainties are fueled by fear caused by political, social, and economic anxieties, creating a 'moral panic' (ibid). Misinformation circulated through social media contributes significantly to such 'moral panic.' Existential fears of illness and death catastrophize thinking, further destabilizing existing social-relational structures. Therefore, it is crucial for us to examine misinformation in social media and its effects on public perception. Otherwise, misinformation will undermine global efforts to control the COVID-19 virus.

Social media's potential adverse effects on public health (Lewandowsky, Ecker, and Cook, 2017) have prompted scholars to evaluate public belief in and susceptibility to COVID-19 misinformation. Bastani and Bahrami (2020) conducted a qualitative study on COVID-19 related misinformation via social media in Iran. They concluded that cultural factors, high information demand, social media prevalence, and inadequate legal supervision of online content are crucial to misinformation dissemination. Brennen et al. (2020) note that COVID-19 misinformation is presented in different forms gathered from various sources and makes different claims; about 59% of the misinformation in their sample dataset involved various forms of reconfiguration of true information, while 38% was completely fabricated.
In their analysis of Twitter misinformation on Covid-19, Singh et al. (2020, 15) identified five major misinformation myths: These include “Origin of COVID-19, Vaccine Development, Flu Comparison, Heat Kills Disease, Home Remedies”. Their findings show that the myth regarding the origins of the virus were highly dominant in Twitter in January and February of 2020. By the end of February, the dominant myths also included the comparison of COVID-19 to flu and the promotion of home remedies as the cure. Other myths such as the effectiveness of heat-killing COVID-19 and vaccine development theories arose over time as well. Misinformation promoting the use of Chloroquine as an effective preventive measure led to the death of an Arizona resident who consumed a form of Chloroquine used for treating aquariums (Waldrop, Alsup, and McLaughlin 2020). In Nigeria, healthcare workers identified numerous cases of Chloroquine overdose after alleged news from the media claimed its efficacy for the treatment of COVID-19 (Busari and Adebayo, 2020).

Conspiracy theories promoting the pandemic as a hoax or as a bioweapon designed by sinister forces are associated with reduced containment-related behavior. By framing containment strategies as a tyrannical act of the state that violates an individual’s right to freedom, conspiracy theories have championed anti containment behavior (anti-mask, anti-social distancing, anti-isolation/quarantine) as a victory of individual freedom over state’s regulations, leading to faster spread of the virus (Imhoff and Lamberty, 2020). Stanley et al. (2020) concluded that people who are less likely to engage in effortful, deliberative, and reflective cognitive processes are more likely to believe that the pandemic was a hoax and thus less likely to engage in social distancing and handwashing, thus accelerating the spread of the virus (Stanley et al., 2020). Uscinski et al. (2020) studied the psychological foundations of
conspiracy beliefs. They noted that beliefs in conspiracies about the virus are associated with a propensity to reject information from expert authorities, consequently reducing people's willingness to comply with public health guidance. Further, racial prejudice has been associated with COVID-19 in social media discourses. Budhwani and Sun (2020) conducted research on Twitter activities to assess the prevalence and frequency of the phrases "Chinese virus" and "China virus." They found a rise in such tweets over time, suggesting that knowledge translation is likely occurring online, and COVID-19 related racist stigma is likely being perpetuated on social media.

Other studies have been conducted to investigate the association between COVID-19 misinformation and public health guidance compliance. Bertin, Nera, and Delouvée (2020) conducted two cross-sectional studies exploring the relationship between COVID-19 misinformation and attitudes towards vaccines as well as support towards controversial medical treatment such as Chloroquine. They noted that COVID-19 conspiracy beliefs, as well as a conspiracy mentality, are associated with a distrust of vaccines and negatively related to participants' intentions to be vaccinated against COVID-19 in the future. Roozenbeek et al. (2020) investigated how susceptibility to misinformation about Covid-19 affects key self-reported health behaviors. They demonstrate a clear association between susceptibility to misinformation and vaccine hesitancy and, therefore, a reduced likelihood of complying with public health guidance, causing a rapid spread of the virus (Krause et al., 2020). In South Korea, the result of a cross-sectional online survey shows that higher exposure to COVID-19 misinformation led to a decline in individual preventive behaviors; the study highlights the
potential of misinformation to undermine global efforts in COVID-19 disease control and faster and deeper disease spread (Lee et al., 2020).

Our research focuses on the impact of misinformation spread through Twitter, as "59% of Twitter users have reported it as good or extremely good in sharing preventive health information" (Singh et al. 2020, 2). While examining the spread of true and false news stories through Twitter, Vosoughi et al. (2018) found that falsehood diffuses significantly farther, faster, deeper, and more broadly than the truth. The novelty element of false news generates greater public interest over truthful news accounts; consequently, false news is rapidly spread as people are more likely to share novel information (Vosoughi et al., 2018).

Yang et al. (2020) studied the extent of links to low credibility information on Twitter during the pandemic and contend that the combined volume of tweets linking to low-credibility information is comparable to the volume of New York Times articles and CDC links, raising concern about the volume and extent of low credibility information related to COVID-19 on Twitter. Pennycook et al. (2020) suggest that social bots are more likely to be involved in posting and amplifying low-credibility information and called for future research to investigate the impacts of misinformation spread by social bots upon COVID-19 incidence rates at the state and county level.

Brennen et al. (2020) note that prominent public figures play a significant role in spreading misinformation about COVID-19 and that there are significant motivations behind such activities. It is imperative that trusted fact-checkers, social media activists, and media organizations continue to hold prominent figures to account for claims they make on social media. Kouzy et al. (2020) examined a sample of 673 tweets to understand the impacts of
misinformation on public health. Their findings indicate that medical misinformation about the COVID-19 epidemic is being propagated at an alarming rate on social media. Singh et al. (2020) note that a strong spatiotemporal relationship exists between Twitter information flow and new cases of COVID-19. They note that Twitter conversations around COVID-19 myths led to an increase in COVID-19 cases by 2-3 days. Such misinformation is drowning official public health advice on COVID-19, making it extremely problematic for healthcare professionals' voices to be heard, the implications of which may be enormous as the virus spreads faster and deeper (Oxford Analytica, 2020). Singh et al. (2020) conducted multi-level classification in their study of 16,000 Tweets to find five dominant myths: "Origin of COVID-19, Vaccine Development, Flu Comparison, Heat Kills Disease, Home Remedies". Song et al. (2020) use disinformation to denote false information and identify ten different Covid-19 disinformation categories: (i) Public authority, (ii) Community spread and impact, (iii) Medical advice, self-treatments, and virus effects, (iv) Prominent actors (v) Conspiracies (vi) Virus transmission (vii) Virus origins and properties (viii) Public Reaction (ix) Vaccines, medical treatments, and tests; and (x) Other. Kouzy et al. (2020) examine 673 tweets dividing them into two main categories: misinformation and unverifiable information, as well as three main topics: Medical/Public health, Financial and Sociopolitical. Abd-Alrazaq et al. (2020) examine the topical content of tweets; they identify 12 main topics which are grouped under four main themes: (i) the origin of the virus, (ii) its sources, (iii) its impact on people, countries, and the economy (iv) ways of mitigating the risk of infection.

Analysis of false information has primarily been undertaken through machine learning algorithms using keywords. Relying on topic modeling and sentiment analysis, themes of false
information are identified and spatially analyzed (Chen et al., 2020, Qazi, Imran, and Ofli, 2020). These studies are effective both in analyzing high volumes of 'big data' at multiple scales and in examining spatial relationships. Others have used a mix of automated and manual analysis, combining human inductive coding with machine learning (Forati and Ghose 2020, Kouzy et al. 2020, Singh et al. 2020, Zheng et al. 2021). These studies provide a more nuanced understanding of false information and engage in spatial modeling. Others have pursued qualitative manual analysis of tweets to examine hidden narratives, using frameworks grounded in social theory (Thelwall and Thelwall, 2020, Havey, 2020; Jimenez-Sotomayer et al., 2020). This approach helps to provide meaning and context to the data analyzed and is attuned to socio-economic inequalities, racialization, and political power structures. However, such studies tend not to use quantitative methods or spatial analysis.

What is lost in many studies, though, is the reference to critical social knowledge about the underlying narratives behind topics, particularly in regard to political power structures. In this research, employing thorough and extensive discourse analysis, I aim to identify those tweets that (i) downplayed COVID-19, (ii) showed resistance to-ward safety measures, (iii) disseminated COVID-19 conspiracy theories, and (iv) propagated disinformation. In order to examine Covid-19 related Twitter misinformation, analyze the nature and origin of false tweets and assess the associations among covid-19 pandemic myths in social media. This dissertation aims to bridge these theoretical and methodological divides.
Methodology

I collected 84,864 Covid-19 related geotagged tweets from May 1 to July 31, 2020, using the Tweepy API (Roesslein, 2009). Through manual checks and topical content analysis, I classified the tweets into four major types of Covid-19 misinformation: those that (i) downplayed COVID-19, (ii) showed resistance toward safety measures, (iii) disseminated COVID-19 conspiracy theories, and (iv) propagated disinformation (Table 1). Each category included at least ten tweets. Tweets that did not fall within these four categories were discarded as their disinformation was only obliquely pandemic related. Our final dataset contained 979 tweets. Next, I used the discourse analysis technique to reexamine the false information circulated in tweets. To understand their nature and intention, I viewed each tweet as a story where I examined both its textual and visual content (in the form of pictures and memes). Rooted in social theory, our qualitative techniques assist in categorizing the tweets into four major themes and nine major myths (Tables 2 and 3). I conduct statistical and spatial analysis of this data.

Spatial statistics and visualization helped us to statistically investigate the geographic relationship/correlation between several explanatory variables and disease outbreak (Watkins et al., 2007; Wang et al., 2010, Dom et al., 2013). Therefore, I conducted multiscale spatial modeling to (i) measure the spatial scale at which Twitter activity operates based on the covid-19 pandemic incidence rate, (ii) examine the ways that individuals react to and shape the built environment, (iii) capture the spatiotemporal patterns of tweets, (iv) examine its spatial heterogeneity. I conducted partial distance correlation to examine the correlation between social media activity and the COVID-19 incidence rate. Distance correlation is a measure of
dependence between random vectors. The population distance correlation coefficient is zero if and only if the random vectors are independent. Therefore, distance correlation measures both linear and nonlinear correlation between two random variables. This contrasts with Pearson's correlation, which can only detect the linear association between two random variables (Székely et al., 2007). Partial distance correlation measures the correlation between two random variables, while the effect of a set of controlling random variables is eliminated (Székely and Rizzo, 2014). I calculated partial distance correlation coefficient between two variables: the number of tweets and the number of confirmed cases, while the effect of a controlling random variable population is removed to investigate both linear and nonlinear correlation between these variables. I used Poisson Multiscale Geographically Weighted Regression (MGWR) as our spatial modeling tool, as it can explain COVID-19 variation at multiple scales (Mollalo et al., 2020), to investigate the relationship between the potential explanatory variables (counties' number of tweets per 1000 residents and digital divide scores) and the dependent variable (COVID-19 incidence rate per county). An (adaptive) bi-square kernel, which removes the effect of observations outside the neighborhood specified with the bandwidth and (minimize) correct Akaike Information Criterion (AICc), was used to select optimal bandwidth (Oshan et al., 2019). I used MGWR V 2.2.1 as our modeling tool (https://sgsup.asu.edu/sparc/multiscale-gwr).
Results

*Examining the distribution and content of COVID-19 tweets*

As the first step in our analysis, I examined the geographic distribution of the extracted 37,587 COVID-19 related tweets. The Figure 14 shows the distribution of geotagged tweets throughout the conterminous USA.

![Map of Covid-19 related tweets in May and June 2020, throughout conterminous United States](image)

*Figure 14: Distribution of Covid-19 related tweets in May and June 2020, throughout conterminous United States*

The following table (Table 10) shows the top ten counties in terms of the number of COVID-19 related geotagged tweets in May and June.
Table 10 top ten counties with the highest number of COVID-19 related geotagged tweets in May and June

<table>
<thead>
<tr>
<th>County</th>
<th>State</th>
<th>No. of geotagged tweets in May and June</th>
<th>No. of confirmed cases as of July 1st</th>
<th>No. of confirmed deaths as of July 1st</th>
</tr>
</thead>
<tbody>
<tr>
<td>Los Angeles</td>
<td>California</td>
<td>3290</td>
<td>105507</td>
<td>3402</td>
</tr>
<tr>
<td>New York</td>
<td>New York</td>
<td>3075</td>
<td>28518</td>
<td>3088</td>
</tr>
<tr>
<td>Hudson</td>
<td>New Jersey</td>
<td>1541</td>
<td>18842</td>
<td>1457</td>
</tr>
<tr>
<td>Fulton</td>
<td>Georgia</td>
<td>724</td>
<td>7444</td>
<td>314</td>
</tr>
<tr>
<td>Queens</td>
<td>New York</td>
<td>719</td>
<td>65455</td>
<td>7059</td>
</tr>
<tr>
<td>Cook</td>
<td>Illinois</td>
<td>664</td>
<td>90911</td>
<td>4581</td>
</tr>
<tr>
<td>Harris</td>
<td>Texas</td>
<td>658</td>
<td>31422</td>
<td>378</td>
</tr>
<tr>
<td>Travis</td>
<td>Texas</td>
<td>623</td>
<td>9527</td>
<td>124</td>
</tr>
<tr>
<td>Kings</td>
<td>New York</td>
<td>615</td>
<td>59507</td>
<td>7104</td>
</tr>
<tr>
<td>Riverside</td>
<td>California</td>
<td>590</td>
<td>18041</td>
<td>463</td>
</tr>
</tbody>
</table>

To understand the nature of false information and their geographic distribution, I examined the content of each tweet and assigned it an index. Out of 37,587 tweets, 821 tweets contained sentiments that I identified as negative – tweets that (i) downplayed the severity of COVID-19, (ii) propagated conspiracy theories, (iii) disseminated false news/facts about COVID-19. I also examined hashtags used in these tweets. Besides the #covidiot, which has been used by both indexed and unindexed tweets, the most popular hashtags are used by indexed tweets against containment measures, downplaying the virus, or disseminate misinformation is presented in the following table (Table 11).

Table 11 Most popular hashtags among tweets sharing misinformation

<table>
<thead>
<tr>
<th>Plandemic</th>
<th>CovidPropaganda</th>
<th>NoMask</th>
</tr>
</thead>
</table>

105
Below I provide some examples of tweets using these hashtags:

"Welp, add another casualty to the Plandemic, sweettomatoes / soup plantation is officially done for. They are closing all locations permanently. They are closing all 97 locations permanently in wake of the Corona"

"REVOLUTION -Careful- this dangerous - true and proven research will GET you banned and censored from every social media platform: After studying global data from the novel Coronavirus (COVID-19) pandemic, researchers have discovered a strong correlation between severe vitamin D deficiency and mortality rates. #healthmediastar #immunity #healthmedia #healthygut #healthygirl #healthyvegan #veganlifestyle #veganinspiration #naturopath #naturopathicrevolution #vitamnind #lockdownlife #lockdown2020 #quantum”

“Lies they tell??? #lies #covid19 #governmentcontrol #riotworldwide #raisethefire #conjuredhoodoo #nlhealntempal”.

I found that tweets containing such hashtags primarily originated from Orlando, FL, Dallas, TX, Palm Beach, FL, Houston, TX, Los Angeles, CA, and Watchung, NJ. The cities with the highest use of these hashtags in May and June are located in Texas, Florida, and California, which experienced massive surges in COVID-19 cases in July and are emerging as the epicenters of pandemic after reporting record numbers of new confirmed cases for weeks in a row (Hawkins et al., 2020).
I next aimed to categorize tweets containing false information based on content analysis. Based on our content analysis, I categorized all indexed tweets into two groups: those that provide misinformation and those that downplay the severity of the virus. The first category includes 45.33% of tweets which propagated false information about the effectiveness of containment measures. Further, these tweets claimed that religious faith is sufficient protection as God will protect His devout followers from the virus. Alcohol consumption was also promoted as an effective preventive measure from COVID-19 prevention. Tweets also championed various conspiracy theories. One theme framed the COVID-19 pandemic as a hoax created by government to control people and limit their freedom. Anti-state and anti-regulatory in nature, these tweets actively promote violation of containment strategies as an act of individual liberty. Another theme framed COVID-19 as a bio-engineered virus or biological weapon created by sinister political powers. Examples below illustrate the nature of these tweets:

"This came from the CDC website. The 65k deaths claimed to b from Coronavirus is wrong. The actual death number is 11k. So, the lockdown everyone is suffering from is based on lies"

"More tyranny from Dr Fauci. #Covid #EndTheShutdown Where are the lawyers when you need them?"

"Not letting the Covid Communists stop me from seeing my friends and having a good time. Don't be afraid to have a good time. :) @ Los Angeles, California"
The second category includes 55.64% of tweets that downplayed the severity of Coronavirus, the process of transmission and lifetime of the virus. Examples are provided below:

"I come back the covid didn't kill me... à West Hollywood, California"

"I had COVID without symptoms. I have IgG antibodies, I have no virus. Immunoglobulin G provided immunity that can help others. #sars #coronavirus #covidart #covidfitness #recovered #sarscov2 @ Williamsburg, Brooklyn”

This category of tweets promoted anti containment behavior by minimizing the severity of COVID-19. Consequently, both categories of tweets influenced individuals to reject containment measures, leading to further transmission and infections.

Next, I computed the numbers of misinformation tweets per county and state and mapped the results. Figure 15 shows the distribution of misinformation tweets throughout the conterminous USA.
Table 12 shows the list of states with the highest number of indexed tweets per 1000 people.

**Table 12 Top 10 States Leading in Misinformation Tweets**

<table>
<thead>
<tr>
<th>Rank</th>
<th>State</th>
<th>Flagged tweets per 1000 residents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>District of Columbia</td>
<td>0.0632</td>
</tr>
<tr>
<td>2</td>
<td>New Jersey</td>
<td>0.0075</td>
</tr>
<tr>
<td>3</td>
<td>Kansas</td>
<td>0.0049</td>
</tr>
<tr>
<td>4</td>
<td>New York</td>
<td>0.0047</td>
</tr>
<tr>
<td>5</td>
<td>California</td>
<td>0.0045</td>
</tr>
</tbody>
</table>
Notably, five of these top 10 states: Kansas, Oregon, California, Washington, Hawaii, and Nevada, have experienced a massive surge in their known COVID-19 cases as of July 1st (Hawkins et al., 2020), right after the two-month case study period of this study on the extent of misinformation on Covid-19 in Twitter. Our Twitter discourse analysis findings indicate that resistance to containment measures and lack of public awareness bear a significant impact on the patterns of illness and death during the pandemic. The relatively strong association between the number of indexed tweets and the number of cases per capita supports the notion that the sites of misinformation are now experiencing a higher surge of COVID 19 cases. Dowd et al. (2020) emphasized the importance of considering population dynamics and demographic data to mitigate the approaches to combat the pandemic. A strong relationship between population, social media activity, and COVID-19 incidence rates, as our results suggest, highlights the importance of social media monitoring during the COVID-19 pandemic.
Next, I examined the data at the county level to identify the top ten counties with the highest number of indexed tweets per capita. These are Norton (Kansas), Montgomery (Kansas), Grant (New Mexico), Box Butte (Nebraska), Poquoson (Virginia), Martin (Kentucky), Hudson (New Jersey), Okmulgee (Oklahoma), Bremer (Iowa), and Tillamook (Oregon). The number of COVID-19 cases from May and June to early July has almost doubled (1.88 times) in these counties as well.

To examine the demographic context behind such tweets, I used census data in conjunction with our geolocated Twitter data set. Table 13 summarizes the key demographic characteristics of the residents in these counties (www.data.census.gov).

<table>
<thead>
<tr>
<th>Rank</th>
<th>County</th>
<th>State</th>
<th>Population per SQMI</th>
<th>White (%)</th>
<th>Median age</th>
<th>Bachelor’s degree or higher</th>
<th>Below poverty level (%)</th>
<th>Unemployment rate</th>
<th>Median income</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Norton</td>
<td>Kansas</td>
<td>6.50</td>
<td>92.97%</td>
<td>43.30</td>
<td>11.04%</td>
<td>12.70</td>
<td>1.20</td>
<td>49891</td>
</tr>
<tr>
<td>2</td>
<td>Montgomery</td>
<td>Kansas</td>
<td>53.70</td>
<td>84.45%</td>
<td>39.90</td>
<td>11.57%</td>
<td>18.30</td>
<td>5.50</td>
<td>45173</td>
</tr>
<tr>
<td>3</td>
<td>Grant</td>
<td>New Mexico</td>
<td>7.6</td>
<td>83.20%</td>
<td>45.8</td>
<td>28.81%</td>
<td>21.8</td>
<td>7.6</td>
<td>56094</td>
</tr>
<tr>
<td>4</td>
<td>Box Butte</td>
<td>Nebraska</td>
<td>10.60</td>
<td>88.75%</td>
<td>41.30</td>
<td>11.92%</td>
<td>11.90</td>
<td>5.20</td>
<td>56412</td>
</tr>
<tr>
<td>5</td>
<td>Poquoson</td>
<td>Virginia</td>
<td>788.90</td>
<td>94.09%</td>
<td>43.30</td>
<td>29.37%</td>
<td>4.50</td>
<td>5.30</td>
<td>96831</td>
</tr>
<tr>
<td>6</td>
<td>Martin</td>
<td>Kentucky</td>
<td>55.80</td>
<td>92.28%</td>
<td>37.20</td>
<td>5.54%</td>
<td>26.30</td>
<td>13.70</td>
<td>35125</td>
</tr>
<tr>
<td>7</td>
<td>Hudson</td>
<td>New Jersey</td>
<td>13808.8</td>
<td>52.70%</td>
<td>34.3</td>
<td>42.27%</td>
<td>16.3</td>
<td>6.1</td>
<td>97596</td>
</tr>
</tbody>
</table>
The demographic data associated with the geotagged tweets indicate the following characteristics: The Twitter users are predominantly White (84% on average) and relatively middle-aged (average median age is 41 years old), middle-income (average median income is $61,086), and just about 19% of the population 25 years old and over have obtained a Bachelor's degree or higher. Most of these counties are categorized as rural with extremely low population density (Office of Rural Health Policy, 2015). Voting patterns from 2016 indicate a conservative mindset among the Twitter users, as, on average, 62.13% voted for the GOP party (Presidential Election Results: Donald J. Trump Wins, 2016). Lower educational attainment may be a contributing factor in the spread of misinformation. A lack of understanding of epidemiology and scientific medical research can lead to a rejection of scientific knowledge. Low population density can lead to feelings of isolation, which can be mitigated through social media participation. The desire to gain followings by spreading novel and sensationalistic information is also likely a contributing factor behind such Twitter activities. This is particularly damaging as many rural counties either have no intensive care units in their hospitals or no hospitals at all (Ajilore, 2020). Finally, partisan politics may have shaped the attitude of Twitter users. The relationship between Twitter activities that provide misinformation/downplay the significance of the virus has, in turn, affected the spread of COVID-19 (Singh et al. 2020). Therefore, it is imperative to raise public awareness of scientific knowledge and block fake remedies, myths, and false news about COVID-19. The impact of the digital divide on the lack of
public awareness must also be considered. During a pandemic, health officials rely on the Internet and social media sites (and other digital platforms) to communicate vital information to the public. However, the effectiveness of these digital channels depends on whether individuals have access to it. Thus, concerns have been raised about the digital divide, information quality, and biases (Oh et al., 2010; Goodchild and Li, 2012), as well as source credibility (Ostermann and Spinsanti 2011). Recent literature suggests that demographic groups (i.e., low income, low education, and elderly populations) may lack the resources, skills, and motivations to access social media, and therefore, they may be less likely to post relevant information through social media (Xiao & Huang 2015). Sui, Goodchild and Elwood (2013) report that two-thirds of humanity does not have access to the rapidly expanding digital world. At least 10 percent of the US population does not use the internet (Anderson et al. 2019). Therefore, I must recognize that user-generated data will provide only selective representations of any issue and that there will always be people and communities that are missing from the map (Zook et al. 2010; Burns 2015; Meier 2012; Ziemke 2012). Thus, this study's result should be seen not as reflections of on-the-ground conditions but instead as a representational negotiation rooted in spatial inequalities. The impacts of the digital divide upon social media usage are beyond the scope of this dissertation but should be considered in future research.

Examining Correlations between Twitter activity and COVID-19 incidence rate

To determine correlation between twitter activity and COVID-19 incidence rate, I undertook the following steps. First, I calculated the partial distance correlation between COVID-19 Twitter activity and the number of confirmed cases in May and June 2020, while controlling the effect of the population as a third variable. Partial distance correlation measures
association between two random variables with respect to a third random variable, analogous to, but more general than (linear) partial correlation (Székely and Rizzo, 2016). In the next table (Table 14), estimated distance correlations are presented. As expected, there is a relatively weak positive correlation (0.27) between the number of tweets per county and the number of cases, indicating that the higher the number of confirmed COVID-19 cases in a county, the more geotagged COVID-19 tweets originate in that county. Therefore, people who are worse affected by the pandemic are relatively more active on Twitter.

Table 14 Partial Distance Correlation Coefficients

<table>
<thead>
<tr>
<th>Partial distance correlation</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>between the number of tweets and the number of confirmed cases, controlling the effect of population</td>
<td>0.267322</td>
</tr>
<tr>
<td>distance correlation between the number of tweets and population</td>
<td>0.8156</td>
</tr>
<tr>
<td>distance correlation between the number of tweets and the number of confirmed cases</td>
<td>0.7378</td>
</tr>
<tr>
<td>distance correlation between the number of confirmed cases and population</td>
<td>0.7878</td>
</tr>
</tbody>
</table>

Next, I examined the influence of digital divide. The Digital Divide Index (DDI) was proposed by Gallardo (2017) to quantify internet physical access/adoptions and socio-economic characteristics that affect user motivation, skill, and usage. The DDI ranges from 0 to 100 in value, where 100 indicates the highest level of digital divide. It consists of two scores, which both range from 0 to 100: the infrastructure/adoptions score (INFA) and the socio-economic score (S.E.). The INFA score groups five variables related to broadband infrastructure and adoption:

i) percentage of total 2018 population without access to fixed broadband of at least 100 Mbps download and 20 Mbps upload as of December 2018
ii) percent of homes without a computing device (desktops, laptops, smartphones, tablets, etc.)

iii) percent of homes with no internet access (have no internet subscription, including cellular data plans or dial-up)

iv) median maximum advertised download speeds

v) median maximum advertised upload speeds

The S.E. score groups four variables known to impact technology adoption:

i) percent population ages 65 and over

ii) percent population 25 and over with less than high school

iii) individual poverty rate

iv) percent of noninstitutionalized civilian population with a disability

To determine the overall DDI score, these two scores are combined (Gallardo, 2017).

Using Gallardo's proposed formula, I calculated the INFA and SE scores for all census tracts in United states to be included in the modeling process as control variables.

In order to investigate the calculated correlation at different scales, I employed MGWR to examine the relationship between the explanatory variable (normalized total number of tweets per county) and the dependent variables (normalized COVID-19 number of confirmed cases per county, INFA score per census tracts, and SE score per census tracts).

Results from the global regression model for 3142 observations (counties) with an R-Squared of 0.424 are summarized and presented in Table 15, to provide context for the MGWR results.

Table 15 Global Regression Results
### Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.000</td>
<td>0.014</td>
<td>1.000</td>
</tr>
<tr>
<td>Tweet rate</td>
<td>0.350</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>SE Score</td>
<td>-0.183</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td>INFA Score</td>
<td>-0.043</td>
<td>0.019</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The global model produces a relatively Moderate R2, indicating about 42% of the variation across covid-19 incidence rate can be accounted for by social media activity in this study. Based on a standard t-value threshold of 1.96 for a 95% confidence level, the tweet rate, INFA score, and SE score are statistically significant. The above results assume that the relationships are constant across the study area. In order to relax this assumption, determine unexplainable high levels of spatial heterogeneity, local multicollinearity, and concavity in the local subsets of the data (Oshan and Fotheringham, 2018; Oshan et al. 2020), and allow the processes to vary at different scales, it is necessary to employ MGWR. Calibrating an MGWR model produces a vector of optimal bandwidths that describes the spatial scale at which each process in the model varies (Oshan et al. 2020). MGWR was applied to the same set of the explanatory variable used in the global model, The R2 increased to 0.795 in the MGWR model from 0.424 in the global model, and the AIC decreased to 4430.295 in the MGWR model from 7190.708 in the global model. MGWR model obtained a high R2 (0.795), indicating that Twitter activity could explain about 80% of the total variations of COVID-19 incidence rates. In Table 16, the bandwidth related to the explanatory variable tweet rate is listed (theoretically, the global model assumes bandwidth of infinity). Association between covid-19 incidence rate and social
media activity seems to occur at a regional scale, with bandwidth implying several hundred nearest neighbors.

Table 16 MGWR results (The bandwidth is expressed as the number of neighboring census tracts)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bandwidth</th>
<th>Effective # parameters</th>
<th>Critical t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>842</td>
<td>209.837</td>
<td>3.679</td>
</tr>
<tr>
<td>Tweet rate</td>
<td>242</td>
<td>18.146</td>
<td>2.996</td>
</tr>
<tr>
<td>SE Score</td>
<td>2171</td>
<td>7.667</td>
<td>2.007</td>
</tr>
<tr>
<td>INFA Score</td>
<td>44</td>
<td>194.236</td>
<td>3.659</td>
</tr>
</tbody>
</table>

As expected, the Tweet rate has a positive non-zero parameter estimate and displays regional spatial variation (Figure 16). The tweet rate surface highlights significant role of twitter activity in explaining covid-19 incidence rate, it is clustered all over the country except for New

Figure 16 map of MGWR parameter estimate surfaces for Tweet rate.
England Texas, and Georgia, and the association is the strongest in midwestern and central states; the characterization of this cluster requires further investigation, but it is in agreement with the MGWR local R-squared. Spatial heterogeneity in the parameter estimate identifies hot spots of high and low Covid-19 incidence rates after controlling for the Twitter activity. These spatial patterns may include both the effect of geography and the effect of geographic patterning associated with potential omitted variables, as Coffee et al. (2020) highlight the role of location and spatial context as an essential component in shaping human behavior and socioeconomic status.

As shown in table 16, the socioeconomic score in MGWR is statistically non-zero and occurs at a regional scale with bandwidths implying nearest states; the infrastructure/adoption score in MGWR is statistically non-zero and occurs at a local scale with bandwidths implying nearest neighbors, suggesting the impacts of rural/urban divide in high-speed residential Internet access (Figure 17). As the residential Internet access moves toward high-speed connections, a rural-urban divide has been created in the U.S. regarding high-speed access. Rural–urban differences in people, place, and infrastructure are all possible causes of this high-speed digital divide. This highspeed digital divide is shaped by uneven geographic supply of internet infrastructure, demographic characteristics of users (age, income, educational attainment) and in network externalities (Whitacre & Mills, 2007; Greenstein, 2020). High speed internet infrastructure tends not to be available in low-density regions, while some areas lack any internet infrastructure (Forman et al. 2018). National patterns can obscure significant regional and inter-state differences (Pear et al., 2019). Existence of significant intercept with regional variation is an indication of local scale determinant alongside social media activity to
explain Covid-19 incidence rate. Disadvantaged groups are less likely to post crisis relevant information through social media (Xiao & Huang 2015); additionally, language barrier may affect marginalized communities from using information (Crawford and Finn 2015); different local and state governments policies might have led to regional variation of intercept. Thus, consideration of these determinants and follow-up investigations at a local scale is necessary for future multiscale analyses.

Figure 17 map of MGWR parameter estimate surfaces for INFA score.
Figure 18 illustrates the spatial distributions of local R2 values in the MGWR model. Several counties in Florida, California, the Tristate area, Texas, Michigan, Minnesota, Nevada, Washington, Carolinas, Utah and Louisiana show very high local R2, indicating a strong performance of the model in those counties. In contrast, the local R2 values were low in most of the counties in Montana, North Dakota, South Dakota, part of Wyoming, Kansas, and West Virginia, indicating a poor performance of the model across these states. R2 values distribution clearly shows that model works with high degree of accuracy in urban centers, coastal areas and cities while social media activity might not be a good explanatory variable to model covid-19 incidence rate in rural areas.
These findings provide explanations for the high variability of the disease incidence in
the continental United States. MGWR parameter surfaces indicate that there is a strong
correlation between COVID-19 cases and Twitter conversations. Our findings are thus in line
with that of Singh et al. (2020). Continued monitoring of social media activity can assist in
understanding the dynamics of disease spread. However, the predictive ability of the model is
somewhat limited by the granularity of the data. The finest spatial granularity at which
nationwide COVID-19 data is provided is at the county-level. Therefore, it is difficult to make
inferences at the sub-country and individual levels accurately.

Discourse Qualitative Analysis

To understand the nature of false information and its impacts, I undertook a rigorous
qualitative and quantitative analysis of tweets. Many scholars use misinformation,
disinformation, and downplaying interchangeably to signify false information. I argue that I can
gain much by operationalizing them distinctly from one another as subdivisions of false
information (Forati and Ghose, 2020). The distinction is a mixture of intent and content. First, I
conducted discourse analysis to organize all false information tweets into four major
categories: Downplaying, Misinformation, Disinformation, and Suspected Bot Activity (Table
17).

<table>
<thead>
<tr>
<th>False Information Tactic</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downplaying</td>
<td>Intentional minimization of risk, uncertainty, and/or effects of Covid-19</td>
</tr>
</tbody>
</table>

Table 17 False Information Categories and Definitions.
Misinformation | False information sharing without a clear agenda or personal gain; attributed to a lack of knowledge about Covid-19

Disinformation | Intentional spread of/ profiteering on uncertainty surrounding Covid-19 for personal or political gain.

Suspected Bot Activity | Repeated tweets from one account which contain false information with limited content on the user profile.

To gain a greater understanding of the content and motivation behind false information, I performed a topical content analysis to identify major topical areas of discussion in tweets. I noted four main themes of content in tweets: politically charged myths, health myths, religious myths, and Freedom rhetoric myths (table 18). Each theme was divided into subunits called myths highlighting the most prevalent false information being spread.

Table 18 Themes and Myths within topical content analysis framework Information Categories and Definitions

<table>
<thead>
<tr>
<th>Theme</th>
<th>Theme Definition</th>
<th>Myth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politically charged</td>
<td>Covid-19 false information with direct reference to rightwing political narratives</td>
<td>Demonizing Pro-Containment Supporters</td>
</tr>
<tr>
<td>Health</td>
<td>False information about the Covid-19 virus and related illnesses</td>
<td>Fake Cures, Vaccine is harmful</td>
</tr>
</tbody>
</table>
Next, through rigorous discourse analysis of tweets, I identified four main topical themes of discussion. I conducted a second round of discourse analysis to determine subcategories within each theme. I examined the content, intent, and history of tweets to determine these subcategories, known as myths (Table 19).

<table>
<thead>
<tr>
<th>Theme</th>
<th>Myth</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Religious</td>
<td>False information about Covid-19 with references to religious beliefs.</td>
<td>Immunity based on religious merit</td>
</tr>
<tr>
<td>Freedom Rhetoric</td>
<td>False information narratives that justify prioritizing individual freedom over public health.</td>
<td>Covid is Over</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Economic liberty over Safety</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Individual freedom over Safety</td>
</tr>
</tbody>
</table>
Politically charged Demonizing Pro-Containment Supporters Demonizing Pro-Containment Supporters, the myth targets both public health experts and government policymakers to undermine their pro-containment advocacy. Individuals and Democratic party members supporting health measures were targeted. This narrative is a known tactic for conservative rhetoric (Genosko, 2020)

Anti-state populist rhetoric Tweets reflect anti-government, populist, and conservative world views. The tweeters claim to be victims of
<table>
<thead>
<tr>
<th>Myth</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>media censorship and government</td>
<td>media censorship and government</td>
</tr>
<tr>
<td>persecution and claims Trump as</td>
<td>persecution and claims Trump as their true leader.</td>
</tr>
<tr>
<td>Racial Stereotype</td>
<td>Tweets express racism against China/Chinese, immigrants of color, and all</td>
</tr>
<tr>
<td></td>
<td>racial minorities in the US.</td>
</tr>
<tr>
<td>Covid as a hoax</td>
<td>Tweets promote Covid-19 as a fake illness invented by the deep state to</td>
</tr>
<tr>
<td></td>
<td>take control of individuals. It dismisses Covid-19 deaths or infection</td>
</tr>
<tr>
<td></td>
<td>rates as fake news.</td>
</tr>
<tr>
<td>Ridiculing experts/containment</td>
<td>This myth ridicules containment measures such as masks and safer-at-home</td>
</tr>
<tr>
<td></td>
<td>measures. It</td>
</tr>
<tr>
<td>Health</td>
<td>Fake Cures</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>also ridicules experts supporting containment measures.</td>
<td></td>
</tr>
<tr>
<td>Tweets promoted fake Covid-19 prevention methods and fake cures, a marketing tool for home remedies, the sale of immunity-boosting supplements, and holistic medicine classes.</td>
<td></td>
</tr>
<tr>
<td>Vaccine is harmful</td>
<td>This discourse was less prevalent as the study occurred before the vaccine for Covid-19 was approved.</td>
</tr>
<tr>
<td>Masks are harmful/ineffective</td>
<td>This category centered on erroneous claims of masks being either useless or harmful to your health</td>
</tr>
<tr>
<td>Incorrect info about social distancing</td>
<td>Tweets claimed that social distancing protocols at specific events were maintained while simultaneously including content that showed violations of health mandates.</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Flu comparison/not serious</td>
<td>The flu comparison undermines the seriousness of the pandemic and frames preventative measures as unnecessary or overvalued.</td>
</tr>
<tr>
<td>Freedom Rhetoric</td>
<td>Tweets indicated that the pandemic was over, despite a surge in infection.</td>
</tr>
<tr>
<td>Covid is Over</td>
<td>Tweets frame health mandates as bad for business. The rhetoric of</td>
</tr>
<tr>
<td>Economic liberty over Safety</td>
<td></td>
</tr>
</tbody>
</table>

127
<table>
<thead>
<tr>
<th></th>
<th>Individual freedom as a marketing technique is used here to sell various services.</th>
<th>Individual freedom over Safety Tweets champion individual freedom over public health and wellbeing. Tweets contain content depicting individuals partying in large groups or visiting family in defiance of containment measures.</th>
<th>Tweets display messages of religious exceptionalism, whereby those who are deemed righteous are protected by God.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual freedom over</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safety</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immunity based on religious merit</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Principal component analysis (PCA) is commonly applied in social science and epidemiology (Mohtashemi et al., 2007; Usman et al., 2012; Efimov et al., 2021) as it enables us to statistically investigate the geographic relationship/correlation between explanatory variables and identify similar subgroups (Lloyd, 2010). Therefore, PCA can transform a highly correlated multidimensional dataset into a manageable number of components with nonsignificant linear relationships. However, PCA assumes that the observations are independent and parameter estimates are constant across the entire study area. Therefore, PCA cannot model spatial dependence and heterogeneity in the associations and relationships amongst processes.

To examine the role of spatial dependence, spatial heterogeneity, and spatial scale in principal component analysis (PCA) for our geotagged tweets, I used Geographically weighted principal component analysis (GWPCA). It is a spatial extension of the PCA method where a series of local PCAs for each area is based on local variance-covariance matrixes, culminating in localized model parameter estimations (Harris et al., 2015).

Therefore, the extracted geographically weighted PCs (GWPCs) are highly explanatory and more accurate. Further, comparisons between local GWPCs can reveal whether and how associations between pandemic myths vary locally (Chen et al., 2020). Thus, GWPCA offers a more appropriate representation of the false information and investigates the local covid-19 pandemic myths that led to such a representation (Fotheringham et al., 2017; Oshan et al., 2019). In this study, the "GWmodel" package in R software was used for the GWPCA (Gollini et
al., 2013; Lu et al., 2014); the "princomp" function for PCA and ArcGIS (version 10.4) for map makings.

I examined the proportions of false information by (a) category and (b) topic in Figure 19. Figure 19a shows principal categories of false information: about 43% of tweets downplayed risk, uncertainty, and/or effects of Covid-19, about 40% propagated false information about the effectiveness of containment measures and covid-19 pandemic, and about 14% spread false information on uncertainty surrounding Covid-19 for personal or political gain, and finally about 3% of tweets identified as suspected bot activity by our research team.

Figure 19b sheds further light on the nature of misinformation. Here, I find that most tweets spreading falsehood contained some political rhetoric (56%). About 25% of tweets are centered on the prioritization of business activities and individual freedom over public health concerns. These tweets misused the socio-economic anxieties caused by the pandemic to promote business interests and economic prosperity over human wellbeing. At the same time, some used the discourse of freedom and self-expression to sell services and products antithetical to CDC mandates. About 14% of tweets downplayed the seriousness of Covid-19 and compared Covid-19 to the flu, propagated fake prevention techniques and cures, claimed masks as being either useless or harmful to health, made false claims of Safety while advertising key events, and were anti-vaccination in their rhetoric. Lastly, about 5% of tweets were
responsible for disseminating religious myths claiming exemption from Covid-19 on the grounds of religious exceptionalism.

Figure 20 shows the distribution of misinformation tweets across the conterminous USA. Respectively, the District of Columbia, Nevada, New Jersey, California, Oregon, Kansas, New York, Rhode Island, Florida, and Idaho are the top ten states with the highest number of tweets with false information. Significantly, Forati and Ghose (2021) have found that sites of Twitter misinformation later experienced a rise in the number of cases, highlighting the substantial impact of misinformation on the patterns of death.
Next, I examine these misinformation themes and myths quantitatively. Table 20 displays the descriptive statistical analysis of covid-19 pandemic myths. The mean and standard deviation values indicate that though all myths were prevalent across all states, some myths were clearly more dominant.

Table 20 Descriptive statistics of Covid-19 pandemic myths

<table>
<thead>
<tr>
<th>Myth</th>
<th>Sum</th>
<th>mean</th>
<th>std.dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonizing Prominent Actors</td>
<td>179</td>
<td>0.18</td>
<td>0.39</td>
</tr>
<tr>
<td>Anti-state populist rhetoric</td>
<td>225</td>
<td>0.23</td>
<td>0.43</td>
</tr>
<tr>
<td>Racial Stereotype</td>
<td>44</td>
<td>0.05</td>
<td>0.21</td>
</tr>
<tr>
<td>Ridiculing experts/containment</td>
<td>188</td>
<td>0.19</td>
<td>0.39</td>
</tr>
<tr>
<td>Myth</td>
<td>Count</td>
<td>Fraction</td>
<td>Cumulative Fraction</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------</td>
<td>----------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Covid as a hoax</td>
<td>99</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>Fake Cures</td>
<td>62</td>
<td>0.06</td>
<td>0.24</td>
</tr>
<tr>
<td>Vaccine is harmful</td>
<td>15</td>
<td>0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>Masks are harmful/ineffective</td>
<td>26</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td>Incorrect info about social distancing</td>
<td>36</td>
<td>0.04</td>
<td>0.19</td>
</tr>
<tr>
<td>Flu comparison/not serious</td>
<td>41</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>Covid is Over</td>
<td>63</td>
<td>0.06</td>
<td>0.25</td>
</tr>
<tr>
<td>Economic liberty over Safety</td>
<td>78</td>
<td>0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>Individual liberty over safety</td>
<td>182</td>
<td>0.19</td>
<td>0.39</td>
</tr>
<tr>
<td>Immunity based on religious merit</td>
<td>65</td>
<td>0.07</td>
<td>0.25</td>
</tr>
<tr>
<td>Demonizing Prominent Actors</td>
<td>179</td>
<td>0.18</td>
<td>0.39</td>
</tr>
<tr>
<td>Anti-state populist rhetoric</td>
<td>225</td>
<td>0.23</td>
<td>0.43</td>
</tr>
<tr>
<td>Racial Stereotype</td>
<td>44</td>
<td>0.05</td>
<td>0.21</td>
</tr>
<tr>
<td>Ridiculing experts/containment</td>
<td>188</td>
<td>0.19</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Forati and Ghose's (2021) research findings indicate a strong association between misinformation and covid-19 infection rate across US communities. Therefore, I conduct PCA to investigate the spatial connotations among covid-19 pandemic myths. The results of PCA and the number of principal components (PC) can also provide context for the following GWPCA. In Table 21, detailed results of global PCA are presented. The results of PCA revealed that the first 2 PCs had eigenvalues greater than one and cumulatively explained 87.6% of the observed variance in the covid-19 pandemic myths in social media.
The PCA loadings indicated that the first PC (PC1) was highly correlated with the myths belonging to Political and Health themes of misinformation. These include "Politically charged myths," "anti-state populist rhetoric" myths, "Demonizing Prominent Actors," and "Ridiculing Experts." Among the "Health disinformation theme," the most popular myths are "masks are harmful/useless," "vaccine is harmful," and various “Fake Cures." These categories explain the partisan right-wing politicization of the Covid-19 pandemic in the USA.

PC2 was highly correlated with the "Freedom Rhetoric" theme and its associated myths such as "individual liberty over safety," "Economic liberty over Safety," and "Covid is Over”, these highlights both economic and social anxieties. It should be noted that the covid-19 pandemic myths in social media are located, but the global correlation coefficient matrix was used for non-spatial PCA. As a result, the findings above may not accurately reflect the local structure of the covid-19 pandemic myths on social media in some places.

Table 21 The results of global PCA

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>eigenvalue</td>
<td>16.074</td>
<td>1.094</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>4.001</td>
<td>1.046</td>
</tr>
<tr>
<td>Proportion of Variance</td>
<td>0.820</td>
<td>0.056</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.820</td>
<td>0.876</td>
</tr>
<tr>
<td>Loadings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Politically charged myths</td>
<td>0.239</td>
<td>0.188</td>
</tr>
<tr>
<td>Myth</td>
<td>Score 1</td>
<td>Score 2</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Health Myths</td>
<td>0.237</td>
<td></td>
</tr>
<tr>
<td>Religious Myths</td>
<td>0.218</td>
<td>-0.168</td>
</tr>
<tr>
<td>Freedom Rhetoric Myths</td>
<td>0.228</td>
<td>-0.338</td>
</tr>
<tr>
<td>Demonizing Prominent Actors</td>
<td>0.234</td>
<td>0.185</td>
</tr>
<tr>
<td>Anti-state populist rhetoric</td>
<td>0.236</td>
<td></td>
</tr>
<tr>
<td>Racial Stereotype</td>
<td>0.224</td>
<td>0.212</td>
</tr>
<tr>
<td>Ridiculing experts/containment</td>
<td>0.232</td>
<td>0.251</td>
</tr>
<tr>
<td>Covid as a hoax</td>
<td>0.222</td>
<td>0.196</td>
</tr>
<tr>
<td>Fake Cures</td>
<td>0.236</td>
<td>0.145</td>
</tr>
<tr>
<td>Vaccine is harmful</td>
<td>0.232</td>
<td></td>
</tr>
<tr>
<td>Masks are harmful/ineffective</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Incorrect info about social distancing</td>
<td>0.111</td>
<td>0.229</td>
</tr>
<tr>
<td>Flu comparison/not serious</td>
<td>0.214</td>
<td>-0.260</td>
</tr>
<tr>
<td>Covid is Over</td>
<td>0.215</td>
<td>-0.295</td>
</tr>
<tr>
<td>Economic liberty over Safety</td>
<td>0.169</td>
<td>-0.531</td>
</tr>
<tr>
<td>Individual liberty over Safety</td>
<td>0.23</td>
<td>-0.254</td>
</tr>
</tbody>
</table>

I conducted a GWPCA to investigate the associations among the covid-19 pandemic myths, along with the goal of capturing the spatiotemporal patterns of tweets and examining their spatial heterogeneity. Following Harris et al. (2015), I use the Gaussian kernel function to explain the results, apply the exponential kernel function to check the results' robustness, and employ cross-validation (CV) to identify the best bandwidth.
Through cross-validation, the optimum adaptive bandwidth (number of nearest neighbors) for calibrating GWPCA was computed as 28 (CV score: 10.17736), which was less than the total number of states (i.e., 50). One possible explanation is that the relationships among the covid-19 pandemic myths had moderate spatial variability, and the regions with approximately homogeneous relationships were relatively small.

Figure 21 displays the cumulative percentage of the total variance (CPTV) and covid-19 pandemic myths in social media with the highest loadings for each component. In line with PCA results, the two first components of GWPCA were retained. Overall, most local CPTVs and average CPTV for the first two GWPCA components (92.27%) are larger than the corresponding PCA CPTV. The local CPTV for the first two GWPCA components ranged from 87.61% to 99.66%, with higher values on the east coast, west coast, and south of the United States, suggesting the presence of local spatial structure information among the covid-19 pandemic myths in social media which was ignored by global PCA.

The variables that emerged as most influential (winning variables) for two GWPCA components are depicted in the following visualization. Figure 21 illustrates the strong correlation of GWPCA Component 1 with particular misconceptions in different regions of the United States. The myth of Fake Cures is prevalent in Western states, including Washington, Oregon, Idaho, California, Nevada, Arizona, New Mexico, Colorado, Utah, and Wyoming. The notion of prioritizing Individual Liberty over Safety is specifically associated with Wisconsin. The Flu Comparison myth is common in the Midwest, encompassing Iowa, Missouri, Arkansas, Illinois, and Indiana. Religious Exceptionalism is a myth significantly correlated with Southern states, namely Florida, North Carolina, and Virginia. The belief that Covid is Over has high
correlations in Southern states such as Texas, Oklahoma, and West Virginia, as well as in
Midwestern states including Kansas, Michigan, and Ohio, and extends to Eastern states like
Pennsylvania, New York, New Jersey, Maryland, Delaware, Connecticut, Rhode Island,
Massachusetts, Maine, New Hampshire, and Vermont. The myth of Economic Liberty over
Safety is prevalent in the Midwest, particularly in North Dakota, South Dakota, and Minnesota,
and in Southern states including Alabama, Louisiana, Georgia, South Carolina, Tennessee, and
Kentucky. Lastly, the myth of Ridiculing Experts and Containment measures is notably present
in Montana. As such, GWPCA Component 1 serves as an indicator for the intensity of rhetoric
surrounding Freedom and Religious Myths across these states.

GWPC2 was highly correlated with "incorrect information about social distancing"
throughout the west and mountain states, while it was highly correlated with "Economic Liberty
over Safety" in the east of the Dakotas in the Midwest and in the South-East. GWPC2 is highly
correlated with "Covid as a Hoax" in the central Midwest and South and "flu comparison" in the
eastern Midwest. In New England, GWPC2 correlated highly with "Religious myths" Thus,
GWPC2 can be adopted to determine the intensity of Health and Politically Charged myths.
Discussion

Through a mixed-methods approach, our research qualitatively and quantitively examines Covid-19-related Twitter misinformation during the initial stages of the pandemic. Our results reveal that employing novel methodologies can provide significant insight into the collective and individual-level social media discourse, specifically falsehood. On the other hand, the Geographically Weighted (GW) framework is superior to conventional global PCA when it comes to exploring the geospatial distribution of different topics of social media discourse.
Hence, GWPCA can better address the spatial dependence and heterogeneity amongst associations between different categories of social media discourse.

Several studies showed that there is a strong spatiotemporal relationship between Twitter information flow and new cases of COVID-19 (Singh et al., 2020; Forati and Ghose, 2021). The visualization of GWPCA clearly showed that the covid-19 pandemic myths in social media vary locally; therefore, its components can be used to model Covid-19 infection rates. Additionally, the results of GWPCA, including CPTV and variables with the highest loadings for the GWPCs, also indicate a distinct spatial pattern, which is useful for public health planning and development.

Using the results of GWPC1 as an example, Freedom rhetoric and Religious Myths are of the most prominent factors leading to downplaying the potency of the Covid-19 virus in southern, midwestern, and east coast states and can be used to educate the public and be considered in public health policymaking processes. As another example, health-related myths like "Fake Cures" are at their highest level in the west coast, which can be further investigated in those certain communities to effectively improve public awareness about safe/scientifically proven prevention methods. Moreover, by combining the loadings of the variables and the scores of the GWPCs, I can evaluate and compare the covid-19 pandemic myths in social media from the perspectives of Politically charged myths, health myths, religious myths, and myths of freedom rhetoric.

Health misinformation can harm the public's health and is a critical threat to global public health (Chou et al., 2018), specifically COVID health misinformation is being propagated at an alarming rate on social media (Oxford Analytica, 2020), obscuring other credible healthy
behaviors like hand washing, social distancing, and promoting advocating erroneous behaviors that could accelerate the spread of the virus. In addition, such misinformation is drowning official public health advice on COVID-19 out, making it extremely difficult for healthcare professionals' voices to be heard, with potentially disastrous consequences as the virus spreads faster and more profoundly.

Conclusions

Social media conversations have a significant impact on public perception of COVID-19. Spatial modeling through MGWR explains 80% of the variability in the number of known COVID-19 cases in each county (based on Twitter activity and population), suggesting a strong association between these variables. Therefore, in the absence of other reliable indicators, analysis of Twitter conversations can help predict the spread and outbreak of COVID-19. Findings from our discourse analysis suggest that the six states of Kansas, Oregon, California, Washington, Hawaii, and Nevada, where people downplayed COVID-19 and propagated misinformation in May June the most, have experienced a massive surge in their known COVID-19 cases as of early.

As public opinion of COVID-19 is heavily influenced by social media conversations, I must investigate the nature and impact of the 'infodemic.' The findings of our study demonstrate how often, where, and how people are talking about the COVID-19 epidemic on social media. COVID-19-related tweets will continue to rise as the epidemic continues to affect individuals personally since people are more concerned about news that impacts them directly (Singh et al., 2020). Similarly, falsehood spreads farther, faster, deeper, and more widely than the truth.
(Vosoughi et al., 2018); thus, I need to investigate misinformation among the most affected people to deter potential-ly disastrous consequences of health misinformation on the public's health. To contain the COVID-19 'infodemic,' a persistent and coordinated effort from researchers, fact-checkers, social media platforms, independent media, news agencies, and government officials is required to adequately disseminate scientific knowledge and raise public awareness of the pandemic.

This research paid particular attention to the nature, topics, distribution, and roots of covid-19 pandemic myths on social media. Such research can assist in improved decision-making required for preparedness and response to public health crises. Our findings highlight the substantial impact of misinformation and resistance to containment measures on the patterns of illness and death during the pandemic. However, our study has several limitations. First, examining only geotagged tweets means that I examined only a subsection of all tweets. Second, the Twitter dataset exhibits several biases, such as the digital divide, demographic biases (Blank, 2017), and locational uncertainty. Third, our findings are constrained by data granularity issues. I aim to address these limitations in future research. Meanwhile, social media narratives containing misinformation and misinformation on disease spread have a significant impact on public awareness. A sustained and coordinated effort by fact-checkers, social media platforms, independent media, news agencies, and public authorities is needed to control the spread of misinformation about COVID-19 so that scientific information is effectively communicated, and public awareness of the pandemic is raised.
Driven by the impacts of climate change, there has been a surge in the frequency and severity of extreme weather events, including severe storms, floods, and hurricanes, which have resulted in significant damage to both people and places. The devastating and long-lasting effects of Hurricanes Katrina and Harvey on marginalized communities have underscored the necessity of addressing physical infrastructure vulnerabilities in conjunction with social and economic vulnerabilities in local policy responses. To achieve this, community-engaged approaches to policymaking are required, which involve examining the vulnerabilities of different communities, the inequitable distributional impacts of climate change, social justice concerns, and the connection between vulnerability and social inclusion/exclusion (Baum et al., 2019). Climate adaptation and resilience-building strategies must pay attention to both physical and socio-economic vulnerabilities, prioritizing a place-based, community-engaged approach in which community surveys, interviews, and discussions provide valuable local knowledge. By integrating local/experiential data with public data sets, policymakers can identify and respond to community-specific needs, vulnerabilities, and risks.

The use of social media feeds presents a unique opportunity to integrate user-generated content from affected individuals into studies (Sui and Goodchild, 2011). This data can serve as a substitute for individual and collective human behavior and complex social dynamics in geographic spaces (Cao et al., 2015), enabling us to analyze the dynamic interactions between

---

3 Portions of this chapter have been published in the International Journal of Disaster Risk Reduction, coauthored with Prof. Ghose:
human behavior and the environment (Shaw et al., 2016). Geotagged Twitter data offers rich individual narratives that can be spatially analyzed in conjunction with other data. The prevalence and extent of this data are particularly beneficial for analyzing fast-moving, high-impact events that require real-time data analysis. As a result, geotagged tweets have been extensively used to estimate real-time flood extents, provide immediate assistance, enhance situational awareness, and improve disaster management policy-making (Taylor et al. 2012; Schnebele & Cervone 2013; Fazeli et al., 2015; Li et al., 2018; de Albuquerque et al. 2019; Martín et al., 2020b). Nonetheless, this data is incomplete since only 1% of all tweets are geotagged (Sloan and Morgan, 2015). Additionally, the spatial distribution of geotagged tweets is further affected by the digital divide, as this depends on the demographic and socioeconomic characteristics of communities (Graham et al., 2014). The digital divide refers to the differences in internet access/ adoption among communities, which are influenced by socioeconomic characteristics such as income, education, and age. Despite these limitations, the benefits of utilizing geotagged tweets in disaster mitigation efforts are significant, leading to their adoption in disaster response planning activities.

In the United States, climate adaptation and resilience-building studies are focusing on frontline coastal communities and giving priority to the integration of individual and local knowledge with scientific datasets to address both biophysical and socio-economic vulnerabilities (Jurjonas and Seekamp, 2018; Johns et al., 2020). In order to gain valuable insights into ongoing resilience building efforts within affected communities, there is a need for the analysis of historical data, particularly individual accounts of past disasters.
Our research focuses on analyzing geotagged tweets during Hurricane Irma in 2017 in Pinellas County, Florida, which is currently undertaking climate adaptation and resilience building efforts (Johns, Dixon, Pontes, 2020). Pinellas County is located in a vulnerable area prone to hurricanes and climate change, and its population of 959,107 is characterized by racial and economic segregation (Mitchell & Chakraborty, 2014; Johns et al., 2020). The county suffered extensive damage during the category five Hurricane Irma in 2017, which swept through Florida and its islands. The severity of the hurricane prompted the largest evacuation in history (Bousquet and Klas, 2017) and caused damages estimated at USD 66.77 billion (Alam et al. 2018). Despite being the fifth most expensive hurricane in US history in terms of damages incurred (Costliest, U. S., 2018), scholarly research on Hurricane Irma has been relatively limited, with only 3,660 published academic articles compared to Hurricane Katrina’s 148,000 articles. The significant destruction caused by Hurricane Irma highlights the need for historical data analysis to enhance emergency response systems and prepare the community for future large-scale traumatic events (Shultz and Galea, 2017).

Our study aims to investigate the relationship between Twitter activity and Hurricane Irma in Pinellas County, Florida. Specifically, I will analyze Twitter activity on a daily basis during the period of September 8 to September 21, 2017, and examine the correlation between Twitter activity and the severity of the hurricane. In addition, I will identify and analyze major discussion themes related to the hurricane on Twitter. I will also explore the relationship between damage/flooding and social media activity, and investigate how demographic factors influence Twitter participation during a disaster. Furthermore, I will examine the impact of the digital divide on Twitter usage among vulnerable racialized communities in Pinellas County. By
answering these questions, I aim to provide insights into the use of social media during disasters and inform strategies for disaster response and management.

Our study employs a mixed-method approach to address the research questions. Firstly, I conduct a correlation analysis to investigate the potential association between Twitter activity and the severity of Hurricane Irma, with the aim of exploring whether individuals respond to natural disasters in proportion to their level of impact. This analysis is also intended to shed light on potential applications of social media in disaster management. Additionally, I employ discourse analysis to examine the individual narratives within the geotagged tweets. These narratives offer insight into the collective and individual reactions to the disaster and can be useful for disaster phase discovery and disaster management efforts.

Furthermore, I examine the spatial variations in Twitter activity during Hurricane Irma and investigate the effects of the digital divide on Twitter activity. Using spatial analysis and modeling through GIS, I aim to measure the spatial scale at which Twitter activity operates in relation to the level of hurricane damage and examine individual reactions to the disaster. I capture the spatiotemporal patterns of the geotagged tweets, discuss their spatial heterogeneity, and examine the experiences of historically marginalized minorities during the disaster.

The appropriate selection of the scale of analysis is crucial to geographical inquiries, and it constitutes a fundamental concern in this research. Additionally, it is important to examine the geographic scale that governs different processes (Fotheringham et al., 2017). Multiscale Geographically Weighted Regression (MGWR) is an innovative methodological approach that elucidates how geospatial patterns vary across scales. MGWR is capable of exploring the
potential spatial nonstationarity of relationships and providing a measure of the spatial scale that expounds which relationships occur at what scale (Oshan et al., 2019). By capturing the effect of scale in spatial processes, I can more accurately capture spatial heterogeneity, minimize overfitting, mitigate concurvity, and reduce bias in the parameter estimates (Yu et al., 2020).

In this study, I pay close attention to demographic disparities in neighborhood compositions and examine the socio-spatial contrasts between affluent, predominantly white coastal communities and impoverished inland communities predominantly inhabited by racial minorities. Our thorough investigation of the digital divide's role offers methodological and empirical insights for future research and contributes to alleviating inequalities in disaster response. Additionally, our application of MGWR in this research advances the scholarship on multiscale analysis.

We are cognizant of the limitations of utilizing geotagged Twitter data, as it only represents a fraction of Pinellas County residents at 27.55%. However, I have found that the demographic and socioeconomic profiles of these Twitter users are not significantly distinct from the general population of Pinellas County. Our research is aligned with ongoing resilience-building initiatives in Pinellas County and is responsive to its data requirements. Analyzing geotagged Twitter data from Pinellas County during Hurricane Irma enables policymakers to scrutinize the county's previous disaster response efforts. Additionally, our findings are incorporated into the current analysis of the county's physical and social vulnerabilities (Johns, Dixon, and Pontes, 2020). Therefore, our results complement the comprehensive community feedback that has been obtained through surveys, interviews, and participatory observations.
(Johns, Dixon, and Pontes, 2020). As a result, despite the data constraints, this study provides valuable insights for future disaster management efforts.

Literature Review

Social Media Data Applications

Annually, a substantial number of natural disasters occur worldwide, resulting in the loss of lives, displacement of people, and destruction of significant infrastructure and property, including buildings and roads, with associated economic costs amounting to billions of dollars (Altay and Green, 2006; Galindo and Batta, 2013). According to Swiss Re Institute Sigma (2017), natural disasters caused an estimated global loss of $175 billion in 2016. The impacts of these disasters can have long-lasting and detrimental effects on a nation's progress and development, potentially delaying these efforts for many years (Smith and Matthews, 2015; Huang and Cervone, 2016). Consequently, the growing number of natural disasters worldwide has led to increased efforts by countries to mitigate damage and manage disaster response operations more effectively (Akter and Wamba, 2017).

The formulation of appropriate techniques and methodologies for the collection, organization, and dissemination of real-time disaster information has become a national priority for efficient crisis management and disaster recovery tasks (Zheng et al., 2013). Analysis of user-generated data can aid in the development of the next generation of emergency response technologies (Mehrotra et al., 2013). Such data presents unique analytical opportunities that enable timely information exchange and promote community connectedness (Taylor et al., 2012); assist with the collection of data in near real-time (Triglav-Cekada and Radovan, 2013); provide complementary geospatial data in regions where
additional data are inadequate or absent (McDougall, 2011); and offer the ability of individuals to volunteer and participate from outside the impacted disaster location (Whittaker, McLennan, and Handmer, 2015). It has the potential to mitigate the effects of disasters, as an accurate and timely assessment of the situation can empower policymakers to make more informed decisions, take appropriate actions, and better manage the response process. As a result, over the past decade, the use of user-generated data in emergency and crisis management has increased tremendously, leading to the emergence of a new field of study called "crisis informatics". Within the field of crisis informatics, studies have investigated the application of social media in various disaster case studies, and developed different methodologies, tools, practices, and frameworks (Reuter et al., 2018; Reuter and Kaufhold, 2018).

The utility of social media in the context of disasters has been investigated by scholars. Taylor et al. (2012) conducted a content analysis of the Facebook page 'Cyclone Yasi Update' to investigate the types of user inquiries and responses during a disaster. Their findings suggest that a majority of users relied on social media for information during the event, with over a third of them providing general information or responding directly to specific inquiries, a quarter requesting for help, and half offering help or practical assistance. Additionally, over three-quarters of users posted messages of support and sympathy. The study provides compelling evidence to support the claim that social media can be an effective tool for providing psychological first aid and promoting community resilience in disaster situations.

Similarly, Mukkamala and Beck (2017) examined the nature of information shared on Twitter during two different natural disasters and how communities utilized technology to
respond to the hazards. Their study employed both manual and automated content analysis techniques to assess the value of user-generated content during disaster situations. They contend that social media platforms play a crucial role in facilitating collective-level situation awareness and providing relevant information to disaster management agencies.

Haworth (2018) conducted three studies investigating the application of social media in reducing community bushfire (wildfire) risk in Australia. His research underscores the importance of emphasizing the social aspects of user-generated data and prioritizing knowledge formation and dissemination through citizen science and mapping initiatives. Athanasis et al. (2018) developed an approach aimed at enhancing decision support tools for natural disaster management through information from the Twitter social network. Their approach integrates GIS modeling outputs with real-time information from Twitter to manage wildfire risks and monitor earthquakes in real-time. They highlight the significant role of social media data in promoting a more sophisticated transfer of knowledge among civil protection agencies, emergency response teams, and the affected population. Chowdhury et al. (2019) conducted a study on extracting keyphrases related to disasters from social media platforms such as Twitter. They argue that keyphrases can be highly valuable in filtering relevant tweets that can improve situational awareness during emergencies. From another perspective, Akter & Wamba (2017) examined the use of user-generated data in disaster management and found that the existing literature has primarily focused on descriptive (i.e., what happened) or diagnostic analytics (i.e., why did it happen). They suggest that user-generated data-driven crisis analytics platforms offer opportunities to apply predictive analytics (i.e., what will happen) in disaster management.
All disaster management phases require up-to-date and accurate information so that emergency responders know where to target aid and efficiently allocate personnel and emergency resources. One of the most devastating outcomes common across hurricane events is rapid flooding. Traditional data sources used for flood mapping include in-situ field sensor, including water stream gauges for monitoring of a variety of hazards parameters, such as rainfall, water level, shaking of ground (Chen et al., 2017), and remote-sensing systems such as satellite imagery for monitoring of the scope of floods (Tralli et al., 2005). There are limitations to these; water pressure gauges may be smashed due to very severe floods, and satellite images may not be available in severe weather conditions when clouds and smoke obstruct sensors. (Chen et al., 2017; Triglav-Cekada and Radovan, 2013). Thus, one of the biggest challenges for rapid flood mapping is the limited data availability throughout and immediately preceding flooding events. One solution is to incorporate georeferenced social media data to traditional data sources so that integrated knowledge can fill the data gaps. Flood extents can be reconstructed and mapped by incorporating user generated data with hydrological data sets (Fazeli et al., 2015). Other studies have applied user-generated data to augment traditional remote sensing approaches for calculating flood extents and identifying affected roads during a flood disaster (Schnebele & Cervone 2013; de Albuquerque et al. 2019; Taylor et al. 2012). Li et al. (2018) integrated water levels derived from information shared on social media (Twitter and Flickr) with stream gauges to generate inundation maps using a kernel-based approach, while accounting for tweet randomness. This approach enables rapid mapping of affected areas and is more responsive to directing aid to vulnerable areas. As Barkan and Pulido (2017) noted,
cartographic knowledge production "crystallizes recognition of injustice—even for people not interested" (Barkan and Pulido, p 38, 2017).

There are other challenges and limitations to the use of social media data. Social media usage is an individual act, is not universal and the data represents partial knowledge. Examples of reported challenges include issues of source credibility and data quality (Ostermann and Spinsanti 2011; Goodchild and Li 2012), information and personal security (Shanley et al. 2013), data management, perceived legal concerns associated with privacy and liability (Scassa 2013), and the underrepresentation of groups and individuals arising from digital divide (Sui et al., 2013; Van Dijk and Hacker 2013).

False information dissemination in social media is a significant problem. Gupta et al. (2013) highlight the role of Twitter in disseminating fake images during Hurricane Sandy. Their study identified and analyzed 10,350 tweets containing fake images that were circulated during Hurricane Sandy to understand the temporal aspects, social reputation, and influence patterns. Their findings indicate that at a time of crisis, users retweeted information from other users irrespective of whether they follow them or not. Further, their research indicates that automated techniques could be used to identify real images from fake images posted on Twitter. Vosoughi et al. investigated the differential diffusion of verified true and false news stories distributed on Twitter from 2006 to 2017. They found that falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced in false political news. They noted that fake news was more novel in nature than true news, indicating that people were more likely to share novel information. Whereas false stories inspired fear, disgust, and surprise in Twitter
replies, true stories inspired anticipation, sadness, joy, and trust (Vosoughi et al., 2018). Hung et al. (2016) used logistic regression to assess the credibility of user-generated flood incidents' data. They collected georeferenced user-generated data from two crisis mapping platforms and developed a binary logistic regression model for mapping the credibility scores of user-generated data instances.

Digital divide is a key concern, as Sui et al. (2013) report that two-thirds of humanity do not have access to the rapidly expanding digital world. Social media cannot represent 'the everybody' and favors 'the privileged,' or those with money, access, and time to utilize the technology (Haklay 2013; Crawford and Finn 2015). Studies examining social media participation indicate that certain communities (i.e., low income, low education, and elderly populations) may lack the resources, skills, and motivations to access social media. Therefore, they may be less likely to post-disaster relevant information through social media (Xiao & Huang 2015). Additionally, certain areas may be too badly damaged by the disaster, leading to meager participation in social media usage. Just 36% of the population had internet access in the Philippines when Typhoon Yolanda struck in 2013, thereby presenting a partial and skewed picture of the disaster through social media data (Crawford and Finn 2015). For those that are 'included,' the use of geospatial data from the crowd has been shown to enhance existing inequalities (Harvard Humanitarian Initiative 2010; Meier 2012; Ziemke 2012). For example, text messages sent to the Mission 4636 service during the 2010 Haiti earthquake crisis were translated into English and subsequently mapped and reported in English, preventing the Kreyòl speakers who texted for help from accessing the knowledge generated from the project, thus
reproducing unequal power relations between the poor Haitians and the rich who acted on the
information (Crawford and Finn 2015).

Social media access and participation during natural disasters have led to life saving
efforts in USA. During Hurricane Sandy, New York City staff not only tracked online social
networks and tweets for data but also responded directly to residents based on their requests.
Similarly, in Washington DC, Red Cross asked 23 staffers to monitor over 2.5 million social
media posts; and of which 4,500 were tagged for first responders to follow up on (Brooks,
2013). However, lack of participation excludes unrepresented individuals and communities
from receiving disaster assistance. Burns (2018) examined the social and political inequalities of
user generated data by foregrounding the struggles and variegations around data production
practices. Therefore, social media data should be seen not as reflections of on-the-ground
conditions but instead as a representational negotiation rooted in spatial inequalities. These
factors should be fully considered before such data can be leveraged to predict damage,
investigate impacted populations, and prioritize activities during disaster management (Haklay
2013; Crawford and Finn 2015; Zook et al. 2010; Burns 2015; Meier 2012; Ziemke 2012; Burns,
2018).

Spatial Modeling Approaches

Geographically weighted regression (GWR) has been recently employed to understand
how spatial processes like social resilience associated with disasters and disasters itself vary
across space (Chun, Chi, and Hwang, 2017; Purwaningsih, Prajaningrum, and Anugrahwati,
2018; Rifat and Liu, 2020). In order to understand the roles of environmental factors in
influencing the occurrence of pluvial floods, Wang et al. (2017) employed GWR to examine the
relationships between inundation frequency and some selected spatial explanatory factors. Their results showed that the GWR model could enhance accuracy and produce residuals that are insignificantly autocorrelated.

Johnson et al. (2019) implemented a GWR model to identify if accessibility to microfinance institutions which provides various financial services to poor households, helping them to cope with natural disasters and adapt to environmental changes in Bangladesh, was negatively affected by climate hazards, and investigate its spatial variation across 18 sub-districts. Mardianto et al. (2021) used GWR to estimate the number of flood disasters based on the influence of settlements along riverbanks. To propose appropriate mitigation efforts to reduce the number of disasters and the impact of losses.

Rifat and Liu (2020) examined disaster resilience influence on disaster losses among United States' coastal communities, and their results suggest that northeastern communities in the United States are comparatively more resilient than southeastern communities, and resilience components have a statistically significant impact on minimizing disaster losses. Additionally, their results highlighted the importance of the consideration of spatial variations of resilience components in the field of disaster resilience, especially in a larger geographic area where dimensions of resilience differ greatly. GWR adjusts for nonstationarity in relationships using a data-borrowing procedure in order to perform a series of local regressions for each area, which enables the estimation of model parameters at any number of locations in a study area in contrary to a traditional "global" ordinary least squares (OLS) regression model that estimates a single set of parameters, each of which is assumed to be constant across the entire study area. Comparison of local parameter estimates across space is beneficial because it shows
whether and how social media activity during disasters varies across geographic space, issues that are overlooked in a global model. Therefore, GWR offers a mechanism not only to explore whether a model is an appropriate representation but also to define the variables that lead to such a representation for specific locations (Fotheringham et al., 2017; Oshan et al., 2019).

Existing studies that utilize GWR to analyze disaster damage determinants have several limitations that make it difficult to interpret results or to gain collective insight about processes, thereby hampering practical policy implementations. GWR assumes that each determinant operates at the same spatial scale (i.e., the same kernel bandwidth for each variable). However, it is much more likely that the complex social, economic, and demographic factors associated with disasters may each vary at different scales (i.e., unique kernel bandwidths for each variable) (Oshan et al., 2019). When it is presumed that the same spatial scale applies to all these relationships, it is likely that the true trends across space are distorted since the model is misspecified. Consequently, when modeling complex spatial processes, it is essential to use a multiscale approach, such as multiscale geographically weighted regression (MGWR). MGWR is an extension of GWR that allows studying the relationships at varying spatial scales and achieves that by deriving an optimal bandwidth vector in which each element indicates the spatial scale at which a particular process takes place as opposed to a single, constant bandwidth for the entire study area (Fotheringham et al., 2017).

This dissertation examines the association between social media activity and disaster damage through an explicitly multiscale modeling approach (i.e., MGWR). It also demonstrates current best practices for building, interpreting, and reporting results for an MGWR model. By
doing so, this research overcomes the limitations of previous work and establishes a methodology for carrying out similar studies in the future.

Study Area

The study site, Pinellas County, is situated in the west-central Florida region on the Gulf coast. This area can be described as a peninsula bordered by Tampa Bay to the east and the Gulf of Mexico to the west. Over the period of 1960 to 2001, the county's population increased by 148%, indicating rapid urbanization and development (Xian & Crane, 2007). The county's population density of 1,264 persons per square kilometer is the highest in Florida, with more than three-quarters of its land area categorized as urban or impervious surface (Mitchell & Chakraborty 2014), which increases the likelihood of flooding. Furthermore, the effects of Sea Level Rise (SLR) are exacerbating the flooding issue in both inland and coastal areas, as it reduces the gradient for flow.

Our research is closely aligned with Pinellas County's ongoing objective of enhancing climate adaptation and resilience within marginalized communities. The settlement patterns of Pinellas County since the late 1800s have resulted in economically prosperous individuals residing in coastal areas, which renders them more susceptible to the adverse impacts of coastal flooding and storm surges. Conversely, socioeconomically disadvantaged groups are primarily located in inland industrial districts, which makes them less vulnerable to direct coastal flooding if adequate infrastructure maintenance is in place. This spatial distribution of residents has also led to marked racial disparities between coastal and inland communities, characterized by high levels of residential segregation between white and African-American
populations, as noted in Mitchell and Chakraborty's (2014) research. As predominantly white populations dominate waterfront areas, Pinellas County serves as an ideal case study for analyzing the intersection of social media activity and environmental inequalities.

Data Sources and Methodology

The present study offers innovative insights into the process of mapping flood damage in a near real-time and dynamic manner by utilizing Twitter data in conjunction with geospatial methods. More specifically, I conducted quantitative analyses of spatiotemporal patterns of tweets, as well as discourse analysis on extracted tweets, in order to identify the qualitative content that can inform long-term measures in the event of a disaster. Our approach represents a novel means of integrating both quantitative and qualitative methods in the analysis of Twitter data, with a particular focus on the spatiotemporal dimensions of social media discourse in the context of flood damage.

*Urban water body detection*

Urban water bodies detection is one of the critical components for urban health and sustainability, such as delineating inundation and water accumulation locations to prevent possible outbreaks of waterborne diseases. Satellite remote sensing is an effective approach to extract large-scale water bodies than conducting field surveys. In contrast to optical imagery, radar imagery, such as synthetic aperture radar (SAR) images, from sensors operating in longer wavelengths can penetrate clouds and alleviate the impact of cloud obscuration on satellite remote sensing. SAR images can distinguish water bodies from the radar waves' return strengths and provide opportunities for water body identification when in situ observations are
difficult to obtain under severe weather conditions. However, most SAR images (RADARSAT-2, TerraSAR-X, and COSMO-SkyMed) are costly, limiting their popularity and usage. Since the Sentinel-1 satellites from the European Space Agency (ESA) provide open and free SAR images worldwide, the budget for accessing radar images can be substantially reduced. Currently, Sentinel-1 SAR imagery has been widely used in water-related applications, such as flood mapping (Twele et al., 2016; Bioresita et al., 2018), soil moisture retrieval (Paloscia et al., 2013), and water change detection (Clement et al., 2018). Therefore, Sentinel-1 data for Pinellas County was obtained for two dates: 09/09/2017 and 08/17/2017. Previous studies have proposed various methods to extract water bodies from SAR images based on backscatter or the joint use of coherence (Chini et al., 2019), including thresholding (Martinis et al., 2009), change detection (Clement et al., 2018), or advanced algorithms for automatic mapping of water bodies (Martinis and Twele, 2010) and handling fuzzy areas (Pulvirenti et al., 2011). In this study, I employed thresholding to extract water bodies of Pinellas County on 09/09/2017 and flooded areas of hurricane Irma on 08/17/2017.

**Social media activity**

In 2018, Alam, Ofli, and Imran (2018) introduced CrisisMMD, multimodal Twitter corpora consisting of several thousands of manually annotated tweets and images collected from across the world during seven major natural disasters that took place in 2017. To construct our Twitter activity dataset, I collected Hurricane Irma's publicly available dataset from CrisisMMD. I used the Twitter Streaming API for data collection, as it enables the collection of data using tweets, texts, hashtags, user information, and location. To gather tweets based on provided tweet IDs in CrisisMMD, I used the Tweepy API (Roesslein, 2009). I
collected Hurricane Irma-related data from Twitter starting from September 6, 2017, to September 21, 2017, and the resulted collection consists of 3.5 million tweets and 176,000 images. After removing retweets and non-English tweets from the dataset, I filtered out 12,213 geotagged tweets worldwide, and out of all geotagged tweets, 5,489 tweets were in Florida, and out of all geotagged tweets in Florida, 191 tweets were in Pinellas County, our case study area.

In order to contextualize our analysis, I calculated descriptive statistics for the tweets collected within each census tract. There are a total of 4,245 census tracts in the state of Florida, with 246 located specifically within Pinellas County. As shown in Table 22, the average number of tweets per census tract in Florida is 1.29, while in Pinellas County, the average is 0.78 tweets per census tract. The range of tweet counts varies significantly, with Pinellas County exhibiting a range of 0 to 40 tweets per census tract, as compared to the wider range of 0 to 264 tweets per census tract across the entire state of Florida.

<table>
<thead>
<tr>
<th></th>
<th>Pinellas County</th>
<th>Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.78</td>
<td>1.29</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3.06</td>
<td>7.32</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>9.37</td>
<td>53.51</td>
</tr>
<tr>
<td>Range</td>
<td>0-41</td>
<td>0-264</td>
</tr>
<tr>
<td>Sum</td>
<td>191</td>
<td>5489</td>
</tr>
<tr>
<td>Number of census tracts</td>
<td>246</td>
<td>4245</td>
</tr>
</tbody>
</table>
Twitter's voluminous dataset of tweets and hashtags has popularized its use for quantitative and user generated data analysis (Zimbra et al., 2018). However, as different user groups have different social norms and idioms of practice (Gershon, 2010), generalizations made about one hashtag, tweet text, or network of users may not apply to another (Marwick, 2014). Qualitative research allows scholars to analyze specific user groups' practices and conduct textual analysis or discourse analysis of individual tweets. Therefore, I conducted discourse analysis on all available 191 geotagged Hurricane Irma-related tweets in Pinellas County as our case study dataset to investigate tweets topics and unveil people's concerns, difficulties, and conditions during hurricane Irma in Pinellas County.

Digital Divide

To address socio-spatial inequalities in our study area, I next calculated the Digital Divide Index (DDI). DDI was proposed by Gallardo (2017) to quantify internet physical access/adopter and socioeconomic characteristics that may negatively affect user motivation, skills, and usage. Using Gallardo's proposed formula, I calculated the DDI for all census tracts in the modeling process.

Multiscale Geographically Weighted Regression

To investigate the association between social media activity and flood rate at census tracts level, I calibrated a global model using regression, which assumes social media activity to be constant across the study area. Subsequently, an MGWR model was calibrated using a golden section search bandwidth selection routine to obtain optimal bandwidths. Following Oshan et al. (2020), MGWR maps were presented for social media activity in order to
investigate the parameter estimate spatial heterogeneity, then implications of results were discussed.

Results and Discussion

In order to investigate the correlation between social media activity and the severity of the hurricane, I created a timeline of Hurricane Irma and Twitter activity. To do so, the following phases of Hurricane Irma were identified (Cangialosi, Latto, and Berg, 2018):

- Hurricane preparation phase: (August 30 to September 9) On August 30, the National Hurricane Center noticed the formation of Hurricane Irma in the Atlantic Ocean and issued warnings about the potential incursion into the US, and the state had ordered people to begin evacuating on September 9 (Hong and Frias-Martinez, 2020).

- Hurricane landfall phase: (September 10 to September 11) Hurricane Irma lands on US territory on September 10, and On September 11, Irma downgraded to a Category 1 hurricane as it headed to Tampa. Twelve million people were without power. Irma was downgraded to a tropical storm as it hit the state of Georgia. (Cangialosi et al., 2018).

- Hurricane recovery phase: (September 12 to September 21), ten days after the hurricane’s landfall. After the hurricane’s departure, the evacuees start to return to their pre-hurricane locations (Wong, Shaheen, and Walker, 2018).
I obtained data on the daily average wind speed and precipitation levels during Hurricane Irma's timeline from the US National Oceanic and Atmospheric Administration (NOAA). Figure 22 illustrates the relationship between the daily average wind speed and the number of tweets that were geotagged to Hurricane Irma during the specified timeline in the United States.

The graph in Figure 22 reveals a remarkable linear relationship between the severity of Hurricane Irma, as measured by daily average wind speed, and the number of geotagged tweets related to the hurricane. To assess the strength of this relationship, I calculated Pearson's correlation coefficient, which yielded a value of +0.92. This indicates a strong positive correlation between disaster severity and social media activity. I can therefore conclude that social media activity is strongly associated with the severity of a disaster. Figure 22 demonstrates an upward trend in both the average wind speed and the number of tweets during the hurricane preparation phase, reaching its peak during the hurricane landfall phase on September 10, followed by a declining trend during the hurricane recovery phase. Thus, I can estimate the severity of a disaster and track the evolution of its various phases by analyzing patterns and trends in social media activity, as illustrated in Figure 22.
Next, I address the potential of social media discourse analysis and topic detection in disaster management and its association with different phases of disaster. Disaster-related social media activities are classified into four key phases. Prior to the disaster, the focus is on ‘mitigation’ and ‘preparedness’. Post-disaster, the focus is on ‘response’ and ‘recovery’ (McLoughlin, 1985). Following this approach, I conducted qualitative discourse analysis on 191 geotagged tweets, where tweets were coded manually and categorized into four main themes: preparedness, response, recovery, and personal (based on disaster phases). Preparedness refers to storm preparations and included tweets on home inventories, food stocking, guard
houses, equipment (storm shutters), news items, and warnings. Tweets concerning evacuation, food shortage, power outage, updates on wellness, and call for help were categorized as (storm) response. Tweets categorized as (storm) recovery included tweets containing search for accommodations, campaigns for help, updates on storm damages, updates on power outages, and reports of looting. Tweets that conveyed emotional support, sympathy, jokes or were uninformative or irrelevant were categorized as personal.

The results of our study reveal that within the temporal scope of September 8 to September 21, 29% of tweets pertaining to Hurricane Irma were classified as recovery-related, 25% as response-related, 17% as preparedness-related, and 28% as personal in nature. This pattern is further illustrated in Figure 23, which demonstrates a strong correlation between social media narratives and the various stages of the disaster. Specifically, during the hurricane preparation phase, the majority of tweets were either personal in nature or focused on storm preparedness. As the hurricane made landfall, tweets predominantly reflected responses to the storm. Subsequently, during the hurricane recovery phase, recovery-themed tweets emerged as the primary theme.

Table 23 provides an overview of the predominant discussion topics on Twitter during the various phases of the hurricane. In the preparation phase, approximately 45% of tweets centered around urging individuals to safeguard their properties by utilizing storm shutters and other protective equipment. Additionally, roughly 10% of tweets encouraged individuals to stock up on food and other essential supplies. Furthermore, a significant proportion (45%) of Twitter users utilized the platform to share news and alerts related to the hurricane,
corroborating the results of Valenzuela et al. (2017), who reported that Twitter is a prevalent source of information during natural disasters.

Figure 23 Tweets' topics categories during hurricane Irma
Table 23 Tweets’ topics in different phases of hurricane and example of tweets

<table>
<thead>
<tr>
<th>Stockpiling</th>
<th>Guarding houses</th>
<th>News and warning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparedness phase (33 tweets)</td>
<td>9.09% (3 tweets)</td>
<td>45.45% (15 tweets)</td>
</tr>
<tr>
<td><strong>Examples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“#StPete do you need water? CVS at corner of 66th Street and 22nd Avenue North @irma @TB_Times @wmnf”</td>
<td>“Before the storm. Making our equipment safe. #aikido #aikikai...”</td>
<td>“West bound Corey Bridge traffic halted without resident ID. Don’t live on the beach? Don’t go today. #irma”</td>
</tr>
<tr>
<td>Evacuation</td>
<td>Call for help and food</td>
<td>Power outage</td>
</tr>
<tr>
<td>Response phase (48 tweets)</td>
<td>16.67% (8 tweets)</td>
<td>8.33% (4 tweets)</td>
</tr>
<tr>
<td><strong>Examples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Everyone in the path, please stay safe 🙏 Hurricane Irma #pray #family #friends #blessings...”</td>
<td>“Just 20 of the 40 gallons of filtered water and provisions we have for Hurricane Irma...”</td>
<td>“The new normal since power is gone... trying to find peace in the insanity. #irma #hurricaneirma...”</td>
</tr>
</tbody>
</table>

Hurricane Irma made landfall on September 10, 2017, in the Florida Keys as a Category 4 hurricane, then hit southwestern Florida at Category 3 intensity. At the peak of the power outages following Hurricane Irma, over 36% of all Twitter accounts in Florida were without electricity (Mitsova et al., 2018). Despite power outages, communications via social media tend to continue during and after the storm in areas without power (Sadri et al., 2018). During this time, I find that 27% of tweets were about power outage (13 tweets), 17% about evacuation (8 tweets), 8% involved calls for help and food (4 tweets), and about 48% belong to the category...
of updates on wellbeing (23 tweets). The dominance of power outage as conversation theme underscores its significance on the lives of people. Power outages have profound impacts on flood control equipment, wastewater systems, transportation, and communication systems, with crippling impacts on the disaster management services (Mitsova et al. 2018). As Mitsova et al. (2018) note, power restoration after hurricanes is affected by communities' socioeconomic vulnerabilities; examination of power outage-related tweets can address inequalities in power restoration and highlight the impacts of blackouts on vulnerable communities. During the Hurricane recovery phase, I find that campaigning for repairs, raising money to help affected communities, looting concerns, and giving updates on damage and power outages are the main discussions in Twitter. These insights are significant for the county’s resilience building effects, as the study of evacuation-related tweets can advance understudied aspects of evacuation behavior and augment traditional evacuation behavior research approaches (Martín et al., 2020).

In line with other studies, our findings support the assertion that social media discourse during natural disasters can support community resilience (Taylor et al., 2012), facilitate collective-level situation awareness (Mukkamala and Beck, 2017), assist with the real time data collection (Triglav-Cekada and Radovan 2013; McDougall 2011), and social media topic detection has tremendous potential in the factual, organizational and psychological dimension of all stages of the disaster management (Gründer-Fahrer et al., 2018; Wukich, 2016).

Simultaneously, I am attentive to the impacts of socio-spatial inequalities on social media usage. To examine the effect of digital divide on twitter use, I employed Multiscale Geographically Weighted Regression (MGWR) with the dependent variable: flooded area per
total area per census tracts, and the explanatory variables: Empirical Bayesian normalized number of geotagged tweets per census tracts, infrastructure/adoption score (INFA) score per census tracts, and the socioeconomic score (SE) per census tracts.

Following Brakenridge, Anderson, and Caquad (2006), a standard generic preprocess workflow for Copernicus Sentinel-1 GRD data within the Sentinel application platform (SNAP), a common architecture for all Sentinel satellite toolboxes, has been conducted. After an extensive process of trials and error checking, I found "-16.5" as an acceptable threshold between land and water. I applied this threshold to map the water bodies and flooded areas in Pinellas County.

![Map of geographic distributions of tweets in Pinellas County in conjunction with calculated flood rate](image)

Figure 24 Geographic distributions of tweets in Pinellas County in conjunction with calculated flood rate

Figure 24 displays the geographic distributions of tweets in Pinellas County in conjunction with calculated flood rates. Although there is concentration of tweets in urban
areas (St. Petersburg, FL), it seems all areas in Pinellas County are represented in tweet dataset. Further, Figure 24 shows that the hurricane’s impacts were acutely felt in both coastal areas and inland areas.

Next, I calculated the demographic and socioeconomic characteristics of census tracts with tweets (weighted based on number of tweets) and of all census tracts in Pinellas County. Although geotagged tweets dataset represents just 27.55% of Pinellas County residents, I find that the demographic and socioeconomic characteristics of these twitter users are not significantly different from those of the Pinellas County overall (Table 24).

<table>
<thead>
<tr>
<th></th>
<th>Pinellas County</th>
<th>Census Tracts with Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population</td>
<td>961608</td>
<td>264968</td>
</tr>
<tr>
<td>Median age (years)</td>
<td>48.76</td>
<td>45.39</td>
</tr>
<tr>
<td>Median income (dollars)</td>
<td>74813</td>
<td>71324</td>
</tr>
<tr>
<td>European American</td>
<td>84.55%</td>
<td>82.59%</td>
</tr>
<tr>
<td>African American</td>
<td>10.30%</td>
<td>13.14%</td>
</tr>
<tr>
<td>Asian</td>
<td>3.42%</td>
<td>2.52%</td>
</tr>
<tr>
<td>Other Races</td>
<td>1.74%</td>
<td>1.75%</td>
</tr>
<tr>
<td>Hispanic or Latino (of any race)</td>
<td>9.14%</td>
<td>10.12%</td>
</tr>
</tbody>
</table>

Results from the global model (Ordinary Least Square) for 244 Observations (census tracts in Pinellas County, Florida) are summarized and presented in Table 25.
The global model produces a moderately low R-Squared of 0.212 with AIC of 655.615, indicating that 21% of the variation across flood rate can be accounted for by social media activity. Multicollinearity does not seem to be an issue since the VIFs for each explanatory variable are all under 10 (O’brien, 2007). Based on a standard t-value threshold of 1.96, I find all the variables to be statistically significant. Table 4 indicates that digital divide index components followed by twitter activity are significant determinants of flood rate, highlighting the fact that inclusion of social and spatiotemporal inequalities in modeling disaster damage is necessary to gain better understanding of damage in real time during natural disasters. On the other hand, the moderately low R-Squared of global model, underscores the importance of spatial heterogeneity and dependency in flood modeling.

Ordinary Least Square assumes that the relationships are constant across the study area, in order to relax this assumption, deter unexplainable high level of spatial heterogeneity,
and allows the processes to vary at different scales (Oshan and Fotheringham, 2018; Yu et al., 2020), MGWR was conducted on the same set of variables. The R-Squared increased to 0.68 in the MGWR model, and the AIC decreased to 531.02 (Figure 25). This indicates that 68% of the variability in flood and damage can be explained by Twitter activity while considering the effects of the digital divide.

Figure 25 Multiscale Geographically Weighted Regression Local R squared.

Pinellas County, surrounded by water on three sides, is vulnerable to hurricanes, flooding, and other disasters. Residential neighborhoods in the county are segregated along class and race lines, where coastal areas are composed of affluent, predominantly white communities while inland areas are comprised of socioeconomically marginalized racial minorities (Johns, Dixon, and Pontes, 2020). As seen in Figure 25, MGWR was able to model flooded areas/damages in coastal communities (wealthy population centers like Gulf Port, Largo, Clearwater) a lot better than socioeconomically marginalized inland areas, as in some part of inland areas, MGWR R-squared is not significantly different from zero. This indicates
that racialized communities are further marginalized by having no voice in social media during Hurricane Irma. Lack of access to participation in social media further marginalizes racialized communities from receiving timely assistance.

In Table 26, the bandwidths related to each explanatory variable are listed. Association between flood rates and social media activity seems to occur at a regional scale with bandwidths implying about a hundred nearest neighbors.

Table 26 Multiscale Geographically Weighted Regression results (The bandwidth is expressed as the number of neighboring census tracts)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bandwidth</th>
<th>Effective parameters</th>
<th>Critical t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20</td>
<td>37.961</td>
<td>3.250</td>
</tr>
<tr>
<td>Tweets Rate</td>
<td>101</td>
<td>1.862</td>
<td>2.227</td>
</tr>
<tr>
<td>SE Score</td>
<td>29</td>
<td>15.311</td>
<td>2.971</td>
</tr>
<tr>
<td>INFA Score</td>
<td>105</td>
<td>1.999</td>
<td>2.255</td>
</tr>
</tbody>
</table>

Next, I examined regional variations in tweet rate through MGWR. As expected, the Tweet rate has a positive non-zero parameter estimate and displays regional spatial variation (Figure 26). As MGWR local R-squared surface, the tweet rate surface is clustered in the dominantly white, wealthy coastal areas, and inspection of tweet rate surface shows that Twitter activity does not have a significant association with flood rate in inland areas where socioeconomically disadvantaged communities reside. The characterization of twitter activity and its association with flood modeling requires further investigation.
Spatial heterogeneity in the parameter estimate identifies hot spots of high and low flood rates after controlling for the Twitter activity. Coffee et al. (2020) highlights the role of location and spatial context as an essential component in shaping human behavior and socioeconomic status. Therefore, I examine the effects of digital divide and socioeconomic characteristics through MGWR. As shown in Figure 27, the digital divide scores (INFA and SE) have positive non-zero parameter estimates; meaning there is a direct relationship between the digital divide components and flood rate. Moreover, socioeconomic component displays local spatial variation, highlighting the importance of neighborhood/community level effect of...
socioeconomic characteristics such as age, educational attainment, poverty, and public health outcomes on social media participation and eventually flood modeling. On the other hand, infrastructure adoption component displays regional spatial variation, which means differences and inequalities in broadband infrastructure and adoption happens and impacts the social media participation and flood modeling at regional scale (about 100 census tracts).

On the other hand, the intercept in MGWR is statistically non-zero, suggesting that identifying additional determinants and follow-up investigations are needed. Recent literature suggests disadvantaged groups are less likely to post disaster relevant information through social media (Xiao & Huang 2015); additionally, language barriers can create an information gap between the marginalized communities and the wealthy, thereby further reproducing power inequalities (Crawford and Finn 2015). The urban core of Saint Petersburg, home to socioeconomically vulnerable communities, can be identified as a region that requires more resources. Thus, consideration of these determinants is necessary for future multiscale analyses of social media activity during disasters.
Figure 27 Maps of Multiscale Geographically Weighted Regression parameter estimate surfaces for a) INFA score which tend to show regional patterns of spatial heterogeneity b) SE score which tend to show local patterns of spatial heterogeneity c) Intercept score which tend to show local patterns of spatial heterogeneity.
Conclusion

Racialized communities in the USA are disproportionately affected by environmental injustices and require improved and effective policy responses. Disasters represent relatively quick-paced, high-impact events that required real-time data for damage mitigation from emergency services. Because of their unique individual qualitative data and quantitative metadata, location-based social media data can be used in various spatial studies. The integration of social media data containing personal narratives with other spatial data sets enables nuanced investigations, as extensive research on citizen science has proved the usefulness of involving the public in environmental monitoring and spatial planning. Social media provides an innovative source of qualitative and quantitative data and provides new insights into human-environment interactions. However, the role of the digital divide must be considered to understand the partial representations in social media, as the exclusion of vulnerable historically marginalized communities in such analysis can further perpetuate existing socio-spatial inequalities.

Our focus on the impacts of Hurricane Irma among racial minorities in impoverished neighborhoods will help to address inequities in future policy responses. Such research can assist in improved decision-making required for disaster preparedness and emergency response by providing valuable and updated information for supporting critical tasks and activities. This research created greater empirical and methodological insights to address the environmental injustices experienced by racialized communities. Empirically, the insights gained from this study can be utilized for effective public policy formulation. Methodologically, this study advanced the research on incorporating social media data into GIScience research.
As in any study, there are limitations as geotag tweets do not represent the entire Twitter community, and careful examination of the variability/heterogeneity in geotagged tweet datasets is required. Our study is mindful of these limitations. However, our study site, time frame and data use were shaped by the County’s climate adaptation policy needs. Alternately, a statewide study can be conducted using a larger sample of tweets, resulting in more comprehensive conclusions. It is also possible that different conclusions would emerge if the study was conducted over a different time period. It is crucial to keep in mind the access constraints of online social networks, as the use of social media varies by demographic and geographical area. Further, online social networks are just one part of a bigger communication paradigm, and it should be investigated within the context of other information.
Chapter 5: Conclusion

This conclusion section synthesizes the findings of three studies focused on examining the effects of opioids overdose deaths, social media conversations on COVID-19, and Hurricane Irma on racialized communities in the USA. The studies were conducted using different research methodologies, including the spatial modeling of geospatial data, discourse analysis of Twitter data, and citizen science. The studies aimed to provide a better understanding of the factors that affect public health crises, such as the opioid crisis, COVID-19 pandemic, and natural disasters, on vulnerable populations. The studies also examined how social media and other novel data sources can provide insights into public health challenges and contribute to decision-making in disaster management.

The first study examined the factors that influence opioid overdose deaths in Milwaukee County, highlighting the variability in relationships between White communities and communities of color. The study recommended community-level solutions based on the community-level factors that contribute to overdose risk. The study emphasized the value of precision epidemiology using MGWR analysis for defining and guiding responses to public health challenges. The findings highlight the need for a better understanding of contributing factors to guide interventions at local, regional, and national scales. GIS-based analytical approaches can reveal important insights into the community and individual factors that contribute to the risk for geographically discordant overdoses and provide actionable data for community agencies and organizations as they formulate strategies for addressing the drug overdose crisis.
The second study demonstrated the significant impact of social media conversations on public perception of COVID-19. The study showed that discourse analysis of Twitter data can help predict the spread and outbreak of COVID-19. However, the study also found that the spread of misinformation and resistance to containment measures had a substantial impact on the patterns of illness and death during the pandemic. The study recommended a persistent and coordinated effort from researchers, fact-checkers, social media platforms, independent media, news agencies, and government officials to contain the COVID-19 'infodemic.'

The third study focused on the impacts of Hurricane Irma on racial minorities in impoverished neighborhoods. The study recommended improved and effective policy responses to address environmental injustices experienced by racialized communities. The study demonstrated the usefulness of social media data containing personal narratives in various spatial studies and provided new insights into human-environment interactions. However, the study also highlighted the role of the digital divide in perpetuating existing socio-spatial inequalities.

Collectively, the dissertation research emphasizes the criticality of community-specific solutions and highlights the importance of diverse stakeholder involvement in formulating public health strategies. It advocates for interventions tailored to the unique circumstances of different communities, rather than generic approaches, with a focus on addressing health disparities.

Furthermore, the dissertation research illustrates how non-traditional data sources, like social media, can inform public policy and disaster management. By integrating these data
sources, the dissertation offers new perspectives on the socio-environmental determinants of health and disaster response.

Finally, the research calls attention to the socio-spatial inequalities that exacerbate the vulnerability of certain populations to public health crises and natural disasters. It urges the inclusion of marginalized communities in the decision-making process to ensure equitable and effective disaster management.

In essence, the dissertation advocates for an interdisciplinary approach to tackling public health and environmental challenges, promoting the integration of innovative methodologies and diverse data sources to enhance policy-making and advance the pursuit of equity and inclusion in public health practice.


Ajilore, O. (2020, July 17th). Rural America Is Starting To Feel the Impact of the Coronavirus. Retrieved from center for American progress:


https://doi.org/10.2196/18932


Bernasco, W., Ruiter, S., & Block, R. (2017). Do street robbery location choices vary over time of day or day of week? A test in Chicago. Journal of research in crime and delinquency, 54(2), 244-275.


Bousquet, S., & Klas, M. E. (2017). Millions of Floridians who fled Irma are eager to get home. Patience will be necessary. Miami Herald, 11.


https://doi.org/10.1007/s10708-014-9599-x


https://doi.org/10.1016/j.compenvurbsys.2015.01.002


Centers for Disease Control and Prevention. Available at:


Centers for Disease Control and Prevention. National Center for Health Statistics. Multiple Cause of Death 1999-2018 on CDC WONDER Online Database released in 2020. Data


http://dx.doi.org/10.1080/00045608.2011.627058


https://doi.org/10.5623/geomat-2010-0005


Laditka JN, Laditka SB, Probst JC. Health care access in rural areas: evidence that hospitalization for ambulatory care-sensitive conditions in the United States may increase with the level of rurality. Health & place. 2009 Sep 1;15(3):761-70.


Maxouris, C., Yan, H., & Vera, A. (2021). Coronavirus variant will likely drive a new wave of transmission come spring, some experts say. Retrieved from


http://dx.doi.org/10.1177/0049124108330005


https://doi.org/10.1080/13658816.2016.1164317


Swiss Re Institute Sigma, 2017. Global insured losses from disaster events were USD 54 billion in 2016, up 42% from 2015. Swiss Re Institute Sigma.


Townsley, M., & Sidebottom, A. (2010). All offenders are equal, but some are more equal than others: Variation in journeys to crime between offenders. Criminology, 48(3), 897-917.


https://doi.org/10.1080/15420353.2012.662471
