The Spatial Variability of COVID-19 in Massachusetts: Implications for Sustainable Development

Submitted by Peiyao Wang

In Partial Fulfillment of the Requirements for the Degree of Master of Science in

Environmental Policy and Planning

February 2021

Thesis Advisor: Sumeeta Srinivasan, Ph.D.

Thesis Reader: Professor Mary E. Davis, Ph.D.

Abstract

The goal of this thesis is to understand the spatial variability of COVID-19 incidence rates across Massachusetts cities and towns while controlling for socioeconomic and urban built environment factors. A dataset which included nine socio-economic indicators and eight built environment indicators was collated. These variables were used to explain the spatial variability of COVID-19 incidence using Ordinary Least Square (OLS) and Spatial Error (SE) regressions as well as Geographically Weighted Regression (GWR). Principle Component Analysis (PCA), were used to create new built environment variables that could help reduce multicollinearity.

The results of the analysis suggest that built environment, education and the percentage of renter-occupied housing units significantly predict COVID-19 incidence rates across the state. The GWR also suggested that these significant predictors of COVID-19 incidence rate varied locally within the state. Urban planners and policymakers need to work on addressing inequity of access not only in education and housing but also access to non-motorized vehicle and green spaces in communities which were found to be harder hit by the COVID-19 pandemic.

Table of contents

List of Figures

List of Tables

1. Introduction

1.1. Topic

Cities vary widely in the US. In some cities like Houston and Los Angeles, residents rely heavily on private cars in cities. Work and non-work destinations are relatively far for most residents, and roads are designed only for driving, but not for pedestrians or bicyclists. That means most residents in Houston or Los Angeles cannot live without a car. Other cities like Boston and New York City, on the other hand, are quite different. In Boston, many residents can rely on public transit to travel to work or non-work destinations. Roads are designed to be walkable and bicycle friendly. Urban design concepts such as transit-oriented development have led planners to create more environmentally friendly and sustainable cities that do not rely heavily on driving.

At the onset of the COVID-19 epidemic, some of those most sustainable cities in US became the first hot spots. Based on this, many started to blame the sustainable urban planning principles. Rocklöv and Sjödin (2020) note that populous urban areas make people more vulnerable to infectious diseases. And this may make many people unwilling to support policies dedicated to urban sustainability development.

However, from the perspective of a planner, we should not give up the idea of building sustainable cities just because of the novel Corona virus, since other more recent reports suggest that the virus is now widespread in the country regardless of density (Caron 2020). In this thesis I will use spatial statistical analysis to explore the relationship between the built environment and the spatial variability of COVID-19 incidence rates in the state of Massachusetts. In this thesis, COVID-19 incidence rate is defined as "the proportion of an initially disease-free population that suffers a disease during a specified time interval (You, Wu, and Guo 2020)." Thus, the COVID-19 incidence rates in Massachusetts is the ratio of accumulated confirmed cases in each city/town to the total population of that city/town.

1.2. Background

The novel Coronavirus epidemic of 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Mayo Clinic 2020). It was first identified in December 2019 in Wuhan, Hubei, China, and has resulted in an ongoing pandemic (Mayo Clinic 2020). The pandemic was confirmed to have reached the United States in January 2020. As of December 2020, the U.S. had the most accumulated confirmed active cases and deaths in the world (CSSE at JHU 2020). As of December 9, 2020, its death rate had become the fourteenth-highest rate globally (Coronavirus Statistics by Country 2020).

COVID-19 can be transmitted through respiratory droplets and contact, and appears to both spread easily and persist in the community (Chan et al. 2020). Because of the strong transmission ability of this novel infectious disease, some high density metro areas quickly became epicenters at the earliest stage of the pandemic, such as the

New York-New Jersey metro area and the Boston, Massachusetts metro area. In May 2020, Massachusetts was third in the U.S. for overall number of cases and third for cases per capita statewide. New cases per day peaked on April 24, 2020 at 4,946, which accounts for 13.2% of the total new cases in the whole country (CDC COVID Data Tracker 2020). However, after a few months, with quarantine and isolation measures in place, the epidemic declined in northeastern cities in the US. As of July 27 2020, Massachusetts had dropped to ninth in the U.S. for total cases statewide, and the new cases per day has stayed under 400 for over 40 days (COVID-19 Response Reporting 2020). Massachusetts started to experience a second wave of COVID-19 in the autumn 2020 (CDC COVID Data Tracker 2020), while the number of cases per capita is only ranked as the thirteen-nineth among all states (CDC). In contrast, some car-dependent states such as Texas, Florida and Illinois had case rates grow over 100% in June, and jumped to the top of the list by July 2020 (CDC COVID Data Tracker 2020). The pandemic in those states continued to be serious through December 2020 (CDC).

While the pandemic rages countrywide, people are trying to understand the factors that caused the spatial variability of incidence rate of COVID-19. Some studies have applied several regression models to examine some factors which may related to the infection and death rate of COVID-19. Studies focusing on the relationship between urban planning and COVID-19 incidence rate and death rate choose a variety of indicators, which include demographic characteristics, income, behavior, transportation access and city size (Sannigrahi et al. 2020; Cao, Hiyoshi, and Montgomery 2020; Cordes and Castro 2020; Mollalo, Vahedi, and Rivera 2020; You, Wu, and Guo 2020). After reviewing of the literature, I found that most of those selected indicators could be grouped into three categories: Environmental, Economic, and Social. However, they do not specifically focus on the built environment or implications for urban planners who are concerned with designing sustainable cities.

Sustainable planning is a fluid concept with the goal of devising policies that will improve the living and working conditions for present and future generations of city dwellers (Committee on Pathways to Urban Sustainability: Challenges and Opportunities et al. 2016). In the most general terms, urban sustainability can be thought of as the measurable improvement of near-and long-term human well-being achieved through actions across environmental (resource consumption and environmental impact), economic (resource use efficiency and economic return), and social (social well-being and health) dimensions (Committee on Pathways to Urban Sustainability: Challenges and Opportunities et al. 2016). Measuring progress towards urban sustainable or unsustainable development requires quantifying phenomena which represent such progress. This is done through urban sustainable indicator sets (Cutaia 2016). There are a variety of indicator sets based on different levels, such as local, state and national. Also, those sets are slightly different based on challenges faced by each specific place. However, they all typically span the three dimensions of urban sustainability: environmental, economic and social.

Based on the environmental, economic, and social dimensions of urban sustainability indicator sets, it is reasonable to select indicators from suitable sets as predictors for the spatial variability of COVID-19 incidence rates across Massachusetts cities and towns. By estimating spatial regression models and geographically weighted regression models that control for selected sustainability indicators to predict COVID-19 data, we may be able to get an understanding of the association between urban built environment and the spatial variability of COVID-19 incidence rate. The analysis results may be useful in suggesting implications to urban policymakers with respect to future infectious disease control.

1.3. Region of study

Massachusetts was chosen as the case study due to the availability of data at a more disaggregate level of city/town. To understand the effects of sustainability indicators on the spatial variability of COVID-19 incidence rate across the state, 351 cities and towns in Massachusetts were used as the unit of analysis (Massachusetts City and Town Websites 2020). Figure 1 shows the boundaries of all cities and towns in Massachusetts.

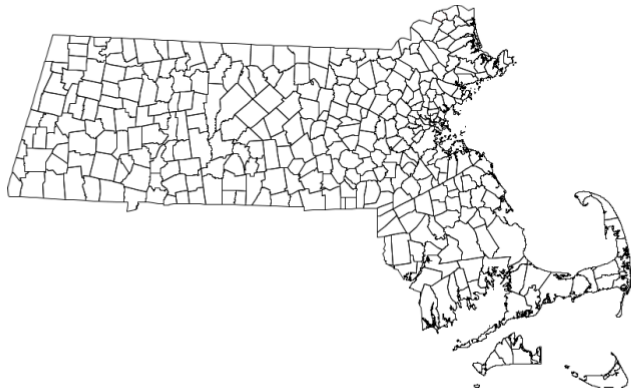

Figure 1. Massachusetts Cities and Towns

Massachusetts has been planning for sustainable development for several decades. The Massachusetts Office for Commonwealth Development (OCD) has published ten sustainable development principles for the built and natural environment through "the integration of energy, environmental, housing, and transportation agencies' policies, programs and regulations" (Massachusetts Office for Commonwealth Development n.d.). The state government has not only published the sustainability development principles, but also established several sustainability programs with state agencies and universities (10 sustainability programs in MA 2019).

1.4 Research questions

In this thesis I will answer the following questions:

1. What predicts spatial variability of COVID-19 incidence rate across Massachusetts cities and towns?

2. Among the urban built environment and socioeconomic predictors, which ones significantly predict the COVID-19 incidence rate across Massachusetts cities and towns?

3. What are the implications for policy makers planning for sustainable cities?

2. **Literature review**

This section summarizes articles based on three topics:

1) Spatial analysis of COVID-19, 2) Sustainable urban development in the context of public health crisis, 3) Urban sustainability indicator sets.

2.1. Spatial analysis of COVID-19

Since the COVID-19 outbreak, several articles have used different models to estimate the incidence and death rate due to the virus (Sannigrahi et al. 2020; Cao, Hiyoshi, and Montgomery 2020; Cordes and Castro 2020; Mollalo, Vahedi, and Rivera 2020; You, Wu, and Guo 2020; Adekunle et al. 2020; Guliyev 2020; Gao et al., 2020). These studies have found many variables that significantly predict the mortality due to COVID-19. You, Wu, and Guo (2020) pointed out that the continued monitoring of these factors can help to control the transmission of COVID-19, and benefits the policymakers to establish healthy cities. These studies have several shortcomings: 1) the spatial units are relative large, which may lead to bias due to the ecological fallacy, 2) the prediction variables may be correlated with each other, leading to data redundancy and collinearity problems, 3) there may be spatial autocorrelation in the residuals since the incidence rates are likely to have spatial dependency.

Mollalo et al. (2020) compiled a geodatabase of 35 environmental, socioeconomic, topographic, and demographic variables that could explain the spatial variability of

COVID-19 disease incidence at the county level across the U.S. They then employed several spatial models where income inequality, median household income, the proportion of black females, and the proportion of nurse practitioners were found to be significant explanatory variables for the variation in incidence of COVID-19 (Mollalo, Vahedi, and Rivera 2020). You et al. (2020) applied spatial regression analysis to explore the relationship between COVID-19 incidence rates and social, economic factors in 13 districts of Wuhan, China. The results showed that population density, construction land area proportion, value-added of tertiary industry per unit of land area, total retail sales of consumer goods per unit of land area, public green space density, and aged population density were positive related to COVID-19 incidence rate. Cordes and Castro's (2020) study showed a negative association between white race, education, and income with the proportion positive tests, and positive associations with black race, Hispanic ethnicity, and poverty in New York City. Sannigrahi et al. (2020) found that a strong positive association between income/total population and COVID cases/deaths exists in the context of European countries. Also, statistically significant associations were observed between COVID-19 case-fatality rate and population size and proportion of female smokers in a country-level worldwide spatial regression analysis led by Cao, Hiyoshi, and Montgomery (2020). Other studies also point out other positively related factors such as asthma, poverty in urban areas and unemployment in rural areas (Ramírez and Lee 2020). Other COVID-19 studies focus on the nexus between the number of confirmed cases, attributable deaths, and also recovered cases (Guliyev 2020; Adekunle et al. 2020), and find them to be correlated with each other.

2.2. Sustainable urban development in the context of public health crisis

Sustainable urban development includes the concepts of environmental sustainability such as reducing consumption, waste and harmful impacts on people and place while enhancing the overall well-being of both people and places. High population density, proximity to transit and walkability are all aspects of planning for sustainability.

However, in the recent COVID-19 epidemic, urban density has been understood by the general population as the cause of the spread of the virus (Keesmaat 2020). One reporter of the Boston Globe blamed city density for "people, pollution, and pestilence" as well as "crime and crowds" (Social Distancing Revives America's Suburban Instincts 2020). Even before COVID-19, urban density has often been a point of contention. According to headlines of some articles (Cox 2000), urban density intensified the environmental pollution problem. Also, dense cities are assumed to make people more vulnerable to infectious diseases and terrorist attacks (Otoole, 2020). Opponents of urban density want to use COVID-19 as an opportunity to return to low-density housing and private automobiles in order to practice social distancing and isolation.

However, most city planners believe urban density in and of itself does not condemn cities to high rates of infection (Keesmaat 2020). Cities differ along many dimensions — population size, age, education level, income level, religiosity, the kinds of work people do, levels of social capital, and more. All of these factors and others may affect their vulnerability to the coronavirus (Bloomberg 2020). Meanwhile, other sustainable cities with high urban density in Asia, such as Shanghai, Hongkong, and Seoul have also controlled the virus well. In sharp contrast to the previous perspective, proponents of density see it as an opportunity to encourage a healthy lifestyle—less reliance on automobiles, more walking and cycling—in order to combat chronic diseases (such as diabetes) that make residents more vulnerable to infectious diseases such as COVID-19 (Keesmaat 2020). According to this point of view, cities will remain vibrant and dynamic centers of economic and cultural activity in the future (Keesmaat 2020). Some statistical results also point to factors other than urban density affect the spread of COVID-19. Increasing evidence shows that some racial and ethnic minority groups and low income groups are being disproportionately affected by COVID-19 (CDC 2020). Other studies indicate that air pollution exposure is linked to higher COVID-19 cases and deaths (Ozgen, Strobl, and Cole 2020).

2.3. Urban sustainability indicator sets

Measuring progress towards sustainable or unsustainable urban development requires quantification with the help of suitable sustainability indicator sets (Verma and Raghubanshi 2018). Those sets can provide a comprehensive, easy to understand, and reliable picture of the sustainability conditions of a municipal area, city or country, with the intention of informing decision-making (Rinne, Lyytimäki, and Kautto 2013). Good indicator sets must clearly differentiate between sustainable and unsustainable development and results should be clearly stated without any confusion for policy making (Lee and Huang 2007). Also, indicators in each set should be limited in number, should be well founded, should use official data, and should have a broad coverage of urban development conditions (Verma and Raghubanshi 2018).

National, regional, and local actors around the world are approaching sustainability based on their own strengths and weaknesses. While urban sustainability performance is measured across the world, there is no single set of indicators that can be used for all the urban areas (Shen and Guo 2014). Thus, there exist several different urban sustainability indicator sets made by different governments and organizations.

The first set of sustainable indicators were published in 1996 by the United Nation Department of Economics and Social Affairs in the form of the Driving force – State – Response (DSR) framework (King 2016). In recent years, the UN published 17 Sustainable Development Goals (SGDs), which are the blueprint to achieve a better and more sustainable future for all (Neshovski n.d.). Those goals are designed to be achieved by 2030. Of the 17 SDGs goals, it is Goal 11 that mentioned the challenge of Sustainable Cities and Human Settlements, which aims to "make cities and human settlements inclusive, safe, resilient and sustainable" (Steiniger et al. 2020). According to Goal 11, the Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs) published a final list of proposed Sustainable Development Goal indicators (Official List of Proposed SDG Indicators 2016). The list includes 230 indicators on which general agreement has been reached.

While the United Nations has published a comprehensive indicator set, other agencies have also published many different sets based on local characteristics. Some selected influential sustainable indicator sets are listed as following: 1) The *Urban Sustainability Indicators* is a product of the European Foundation for the living and working conditions (Mega and Pedersen, 1998). It presents the urban sustainability indicators framework in the context of the Foundation's program on socioeconomic aspects of sustainable development; 2) The *North America Green City Index* assesses the environmental performance of 27 major U.S. and Canadian cities (The green cities challenge 2011); 3) The *Sustainable Cities Index* by Arcadis was tested with data on 50 world cities from 31 countries (Arcadis, 2018); 4) The *Sustainability Urban Development Indicators* is a product of the University of Pennsylvania and was commissioned by the Office of Policy Development and Research, U.S. Department of Housing and Urban Development. It focuses on U.S. urban areas (Office of Policy Development and Research n.d.). In this study I will use indicators related to the built environment as a proxy for sustainability measures of towns in Massachusetts.

3. Data and Methodology

In this chapter I will describe the data and methods used for the analysis.

3.1. Data description

This section describes the data sets as well as the process by which they were collected and prepared. The dependent variable was the COVID-19 incidence rate and the predictors were selected urban sustainability indicators for Massachusetts cities and towns.

3.1.1. COVID-19 incidence rate

The accumulated confirmed COVID-19 case raw data for 351 cities/ towns were collected from the *Weekly COVID-19 Public Health Report – October 22, 2020* on the Massachusetts Government website (Mass Gov 2020). According to the definition of COVID-19 incidence rate which is “the proportion of an initially disease-free population that suffers a disease during a specified time interval” (You, Wu, and Guo 2020), the initial case data was adjusted based on the population and calculated as COVID-19 incidence rate per 1000 people for each city/town (See Table 1. for details). A graphical summary of the COVID-19 incidence rate per 1000 people is shown in Figure 2. As expected, it is extremely right skewed. The log-transformed data is approximately normally distributed as seen in Figure 3.

Seven of the 351 towns had not reported a COVID-19 case until October 22, 2020 and were removed from the sample. The Massachusetts town shapefile was downloaded from Mass GIS, which includes each town's unique ID and the boundaries of all towns. The descriptive statistic results for the COVID-19 data are listed in Table 3.

Table 1. COVID-19 Data

Dependent Variable	Description	Unit	Source	Year
Accumulated COVID-19 Cases	Accumulated count of confirmed COVID-19 cases from January 1, 2020 to October 22, 2020 by city/town	Number	Mass.gov	2020
COVID-19 incidence rate per 1000 people	COVID-19 incidence rate per 1000 people	Number		2020
Ln COVID-19 incidence rate per 1000 people	Log-transformed COVID-19 incidence rate per 1000 people	Number		2020

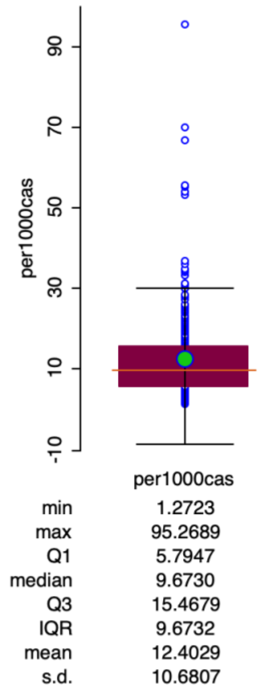

Figure 2.
COVID-19 incidence rate per 1000 people

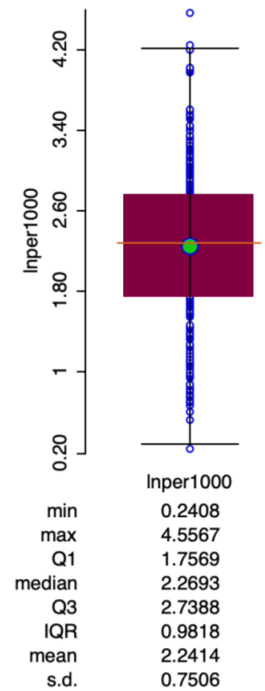

Figure 3.
Log-transformed
COVID-19 incidence rate per 1000 people

3.1.2. Urban sustainability indicator set

Based on the literature review results, two materials were used as resources to define the urban sustainability indicator set for Massachusetts cities and towns, which are *Committee on Pathways to Urban Sustainability: Challenges and Opportunities* and *Sustainable development principles for Massachusetts*. The first one is on the challenges and opportunities for urban sustainability in the US while the second was published by Massachusetts Office for Commonwealth Development (OCD) and for Massachusetts.

In the first material, a variety of urban sustainability indicator systems published by various European and American organizations in the last ten years have been evaluated. The research team conducted a meta-review of urban sustainability indicator systems which included about 60 indicators. Those indicators were classified in three categories, which are environmental indicators, economic indicators, and social indicators (See Appendix A for details). Focusing on Massachusetts, the mission of the OCD is to "care for the built and natural environment by promoting sustainable development through the integration of energy, environmental, housing, and transportation agencies' policies, programs and regulations (MA Sustainable Development Principles 2019)".

Based on the attributes suggested by the OCD mission and data availability, 17 indictors were chosen from the 60 indicators listed in the book - *Committee on Pathways to Urban Sustainability: Challenges and Opportunities* (See Appendix A for details about the 60 indicators). The indicator set included energy, environmental, housing, transportation and economic indicators, which can be resegregated as two groups: socio-economic indicators and built environment indicators. The unit, resource, year, description, and formatting method (See Chapter 3.2 for details) of each indicator are listed in Table 2. The descriptive statistic results for all indicators are listed in Table 3.

Table 2. Urban Sustainability Indicator Set

Type	Description	Unit	Year	Method
Predictor Variables				
Socio-economic Indicators				
Median Household Income [1]	Median household income in the past 12 months	Dollar	2014	Join
Households in Poverty[1]	Households in poverty/Households number	Percent	2014	Join
Gini Index[1]	Gini Index of Income Inequality; 0 = equality; 1 = inequality	0 to 1	2014	Join
Unemployment[1]	Unemployed/population 16 years and over	Percent	2014	Join
Without Health Insurance[1]	Population without health insurance coverage/Population	Percent	2013	Join
Population 25 years and over with Educational Level Below High School[1]	%No high school + %less than high school	Percent	2014	Join
Population that cannot Speak English Well[1]	Population that does not speak English well + don't speak English at all)/ Population 5 years old and over	Percent	2014	Join
Minority[1]	Other races other than white/ Population	Percent	2010	Join
Renter Occupied Housing Units[1]	Renter-occupied housing units/ Total housing units	Percent	2014	Join

Type	Description	Unit	Year	Method
Built Environment Indicators				
Commuters who Drive to Work[1]	Commuters who drive to work/ worker 16 years old and over	Percent	2014	Join
Large Households[1]	Households with three or more People/ Total households number	Percent	2014	Join
PM 2.5[2]	Estimates of outdoor concentrations for PM2.5 annual-average values at census track level	Micrograms per cubic meter	2015	Clip and Spatial Join
Walkability[1]	Length of walking trails	Feet	2019	Clip and Spatial Join
Bike-friendliness[1]	Length of bicycle trails	Feet	2019	Clip and Spatial Join
Carbon Emissions[1]	1-km resolution inventory of annual on-road CO2 emissions	Tons CO2 per Km2	2019	Zonal Statistic
Tree Canopy Coverage[3]	30 m raster geospatial dataset contains percent tree canopy estimates (0 to 100%)	Percent	2016	Zonal Statistic
Green Open Space[4]	GIS-calculated acreage of conservation lands and outdoor recreational facilities	Acreage	2020	Clip and Spatial Join

Data Sources: [1] DataCommon [2] Caces [3] MRLC [4] MassGIS

Table 3. Descriptive statistics for all variables

N = 344	Mean	St.Dev.	Min	Max	Median
Dependent Variables					
Ln Covid-19 incidence rate per 1000 people	2.24	0.75	0.24	4.56	2.27
Predicting Variables					
Socio-economic Indicators					
--Median Household Income (dollar)	90239.78	29949.33	26458.00	224784.00	84457.50
--%Households under Poverty	7.31	4.44	0.00	28.36	6.26
--Gini Index	0.43	0.05	0.32	0.60	0.42
--%Unemployment	4.74	1.85	0.00	14.89	4.53
--%Without Health Insurance	2.45	1.46	0.00	9.27	2.18
--% Educational Level Below High School	8.79	7.24	0.42	52.66	6.77
--%Population cannot Speak English Well	1.70	2.99	0.00	26.76	0.88
--%Minority	10.98	11.38	1.80	79.53	6.75
--%Renter Occupied Housing Units	23.02	13.98	0.00	73.87	19.49
Built Environment Indicators					
--%Commuters who Drive to Work	84.84	10.03	29.83	98.33	87.58
--%Large Households	0.40	0.08	0.12	0.58	0.41
--PM 2.5 (Micrograms per cubic meter)	5.90	0.51	4.76	7.22	5.87
--Walkability (Length of walking trails in feet)	120528.09	129737.95	0.00	729249.24	79639.77
--Bike-friendliness (Length of cycling trails in feet)	21131.11	88772.03	0.00	1508540.09	201.03
--Carbon Emissions (Tons CO2 per Km2)	1681832.94	2216976.11	17787.59	11265534.64	656622.52
--%Tree Canopy Coverage	53.85	15.62	4.23	81.38	54.95
--Green Open Space (Acreage)	6422.00	6051.20	99.68	38212.70	4271.93

3.2. Data formatting

All the data cleaning and formatting processes were done using Microsoft Excel and ESRI's ArcMap 10.8.1.

Parts of the indicator dataset were directly joined to the Massachusetts town shapefile since they were already at the town level. Others which were initially at census block level were clipped with the Massachusetts town shapefile first and then spatially joined to the town. In addition, other raster datasets were joined using zonal statistics with the Massachusetts town shapefile. Details for the method applied to each indicator's dataset are listed in Table 3.

3.3. Statistical analysis methods

The statistical analysis process included four steps: 1) Reducing data redundancy through Principle Components Analysis (PCA), 2) Estimating the spatial variability of COVID-19 incidence rates through multivariable regression and spatial regression, and 3) Locally modeling the spatial variability of COVID-19 incidence rates through Geographically Weighted Regression (GWR), 4) Comparing regression results.

All statistical analyses were performed via GeoDa 1.16.0 and ArcMap 10.8.1. Predictor variables were considered statistically significant at the threshold level of $P<0.05$.

3.3.1. Reducing data redundancy

Initially all 17 independent variables were used to predict the natural logarithm of COVID-19 incidence rate per 1000 people as the dependent variable in an Ordinary Least Square (OLS) model to explore the combined linear relationship between all 17 predicting variables and the dependent variable. The Adjusted R-square for this model is 0.56. However, the multicollinearity condition number of the model was 118.22. A condition number between 10 and 30 indicates the presence of multicollinearity and when a value is larger than 30, the multicollinearity is regarded as strong (Kim 2019). One crucial assumption of OLS is that observations should be independent. The extremely high condition number shows this assumption may not be satisfied as a consequence of the multicollinearity problem that resulted from including all of the related predictor variables in a single model. Thus, several indicators were discarded from the socio-economic group that were highly correlated. A principal component analysis (PCA) on the variables in the built environment group also helped reduce data redundancy. PCA is a statistical procedure designed to put as much information from the data into the fewest number of uncorrelated components. It is used to identify the set of variables that account for the most variation in the original dataset (Pearson 1901).

Among the nine indicators in the socio-economic group, Median household income, Unemployment, Percentage of households without health insurance, and Population percentage which cannot speak English well were all discarded since it was determined that they were measuring similar socio-economic characteristics as percentage of households in poverty. Five indicators: percentage of households in poverty, Gini index, percentage with equal to or less than high school education level, percentage of minority, and percentage of renter occupied housing were used for further analysis as they were not significantly correlated with each other. PCA was carried out with the eight built environment indicators in GeoDa using the singular Value Decomposition (SVD) method. The first four components accounted for 77.5% of the variation in the data. Detailed explanations for each component are described in the next chapter.

Finally, the data redundancy and collinearity problem were further assessed by the Exploratory Regression tool in ArcMap. After eliminating some of the initial predicting indicators, nine indicators were kept for further analysis, which were households in poverty, Gini index, low education level, minority, renter occupied housing, PC1, PC2, PC3, and PC4. The variance inflation factor (VIF) of the 9 exploratory indicators were assessed. VIF provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity. The VIF of all those nine indicators are listed as below: 1.63 for Gini index, 3.28 for percentage of household under poverty, 2,4 for percentage of low education, 3.07 for percentage of minority, 3.54 for percentage of renter-occupied housing unit, 2.35 for PC1, 1.07 for PC2, 1.01 for PC3, and 1.15 for PC4. They are all below or slightly above three, which suggested very low levels of multicollinearity.

3.3.2. Modeling spatial variability globally - Ordinary Least Squares and Spatial Error

Having identified the new set of explanatory variables described above, an Ordinary Least Squares (OLS) model and Spatial Error model were estimated with the reduced dataset. Regression analysis was used to assess the association between COVID-19 incidence rate and Built environment components while controlling for socio-economic indicators.

For both the OLS and the Spatial Error model, the 1st order queen contiguity criterion was used to define the spatial weights matrix. This assumes that towns that share a border are neighbors.

The OLS regression method estimates coefficients for the independent variables (Introduction to Residuals and Least Squares Regression 2017). However, in order to estimate coefficients, three assumptions need to be satisfied: 1) The errors (difference between actual dependent variables and the predicated ones) are distributed with zero mean and constant variance (homoscedastic). 2) Errors are independent, uncorrelated to each other. 3) Errors are normally distributed (Introduction to Residuals and Least Squares Regression 2017). Among those three assumptions, the randomness of error is problematic since the original dataset has spatial dependence. In addition, three VIFs of the 9 variables in exploratory regression are slightly higher than 3, which indicates that the collinearity problem of datasets may violate the independent-errors assumption of OLS.

The two main kinds of spatial dependence are Spatial Lag and Spatial Error. Spatial error violates the assumption that the error terms are uncorrelated, and OLS coefficients will be inefficient. Spatial lag violates the assumption that observations are independent (and uncorrelated errors), and OLS coefficients will be both biased and inefficient (Introduction to Residuals and Least Squares Regression 2017). In order to correct for spatial dependency, there are two methods to optimize the OLS model, which are the Spatial Lag model and Spatial Error model. The Spatial Lag method aims to solve the problem of dependent observations, while the spatial error method aims to solve the problem of correlated errors.

The classic OLS model was run first with diagnostics for spatial dependence. R^2 is the proportion of the variance in the dependent variable that is explained by the independent variables and shows the overall goodness-of-fit of the model. Each coefficient shows the predicted (average) change in COVID-19 incidence rate for a unit change in that indicator if all other indicators are held constant. Each t-statistic value indicates that the predicting indicator may or may not be a significant determinant of the COVID-19 incidence rate. In terms of the diagnostic part, the multicollinearity number is a diagnostic to suggest problems with the stability of the regression results due to multicollinearity. The value of the Lagrange Multiplier and Robust LM suggests which alternative specifications (Spatial Lag model or Spatial Error model) should be used to improve upon the OLS specification. I used Luc Anselin's spatial regression decision sequence to determine a final spatial regression model (Anselin 2010): between the Robust LM-Error test and the Robust LM-lag test results, the one with higher value should be selected as the method to improve the OLS result. According to the diagnostic for spatial dependence in the OLS report, the Spatial Error model was selected to improve the classic model.

3.3.3. Modeling spatial variability locally - Geographically Weighted Regression (GWR)

While the classic OLS and Spatial Error models measure the overall relationship between the dependent variable and the predictor indicator sets, the intrinsic difference of relationships across space might be ignored. Globally regression models such as OLS and Spatial Error models assume these relationships do not vary over space. To relax this assumption and allow for "parameters to vary spatially" (Wheeler D.C. 2014), GWR was introduced. The motivation for GWR is the idea that a set of constant regression coefficients cannot adequately capture spatially varying relationships between covariates and an outcome variable (Wheeler D.C. 2014). GWR is an "extension of general regression models and based on kernel-weighted regression," which allows parameters to be derived for each location separately instead of globally (Wheeler D.C. 2014). The research area of this study is the whole state including rural, urban, and suburban regions, therefore, the built environment and socio-economic conditions will vary widely. It is reasonable to use GWR as an extension of previous global models to locally examine spatial non-stationarity.

4. Results

4.1. Reducing data redundancy - Principal Component Analysis for built environment indicators

In order to reduce data redundancy and reduce the multicollinearity in the original predictors, a PCA was carried out for the 8 built environment variables, which were: percent of commuters driving to work, percent of households with more than three members, average value of PM 2.5 in micrograms per cubic, walkability as measured by total length of sidewalk in each city, bicycle friendliness as measured by total length of bike-trails in each city, average value of carbon emissions in tons CO2 per km^2units, average tree canopy coverage value, and green open space in total acreages . The PCA used the Singular Value Decomposition (SVD) method in GeoDa. The eigenvalues of the first four components are close to or greater than 1, and they cumulatively explained 78.84% of the variance in the attributes. Therefore, the first four components were used to estimate the spatial regression and they are explained in detail below.

4.1.1. Rurality component – PC 1

The first component loaded positively on green space (0.333), percentage of tree canopy coverage (0.500) and percentage of commuters drive to work (0.354). It also loaded negatively on carbon emissions (-0.442) and bike friendliness (-0.294). The map of PC1 (Figure 4.) shows the component is highly clustered in central and western Massachusetts in towns such as North Salem, Petersham, and Blandford. Most of those cities are rural with low population and large natural reserves.

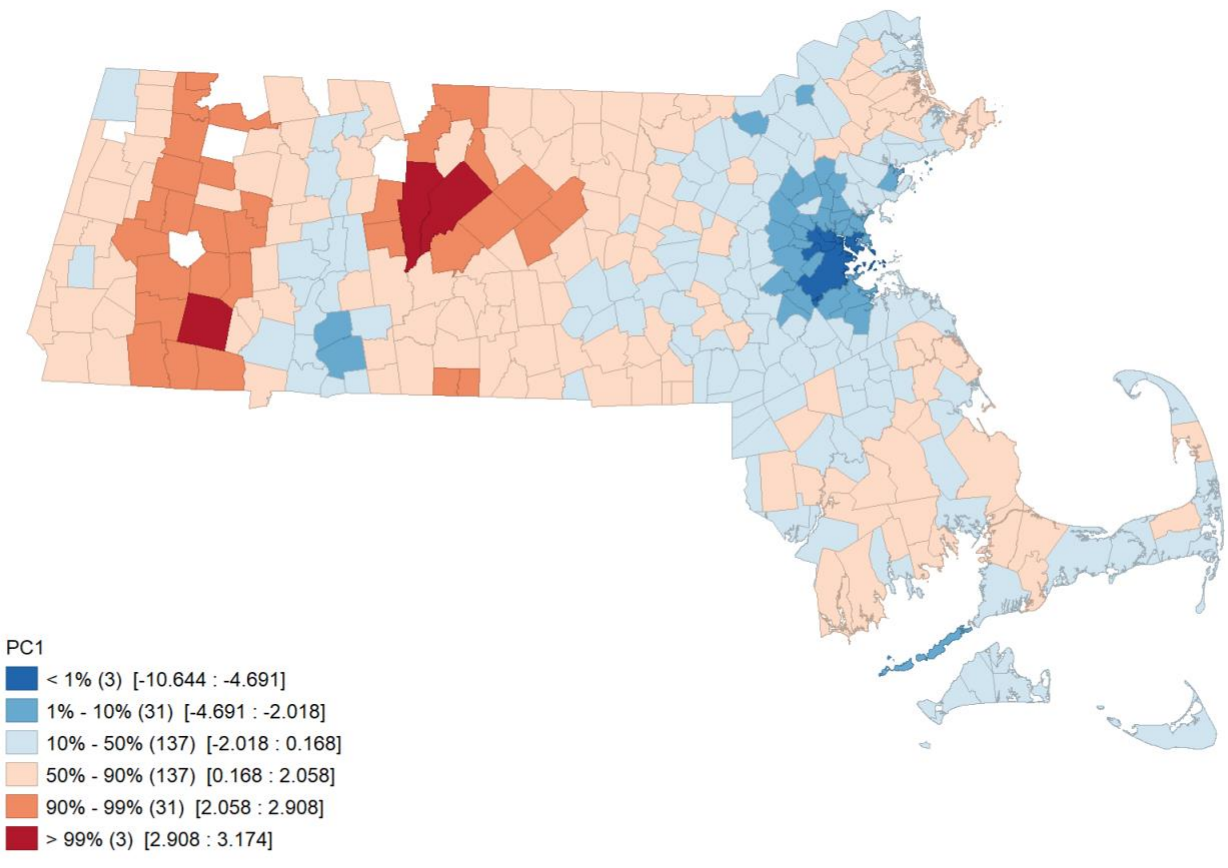

Figure 4. Rurality Component – PC 1

4.1.2. Green component – Non-motorized transport (NMT) friendly – PC2

The second component loaded positively on walkability (0.665), bike friendliness (0.472), and green open space (0.437). The map of PC 2 (Figure 5.) shows that both towns like Boston and Williamstown represent high values for this component. Both Boston and Williamstown have several higher education campuses even though they are quite different in terms of urbanization and population density. The surrounding cities of Boston such as Milton, Quincy, Newton and Cambridge also have high values for this component as do some university towns in the western part of the state.

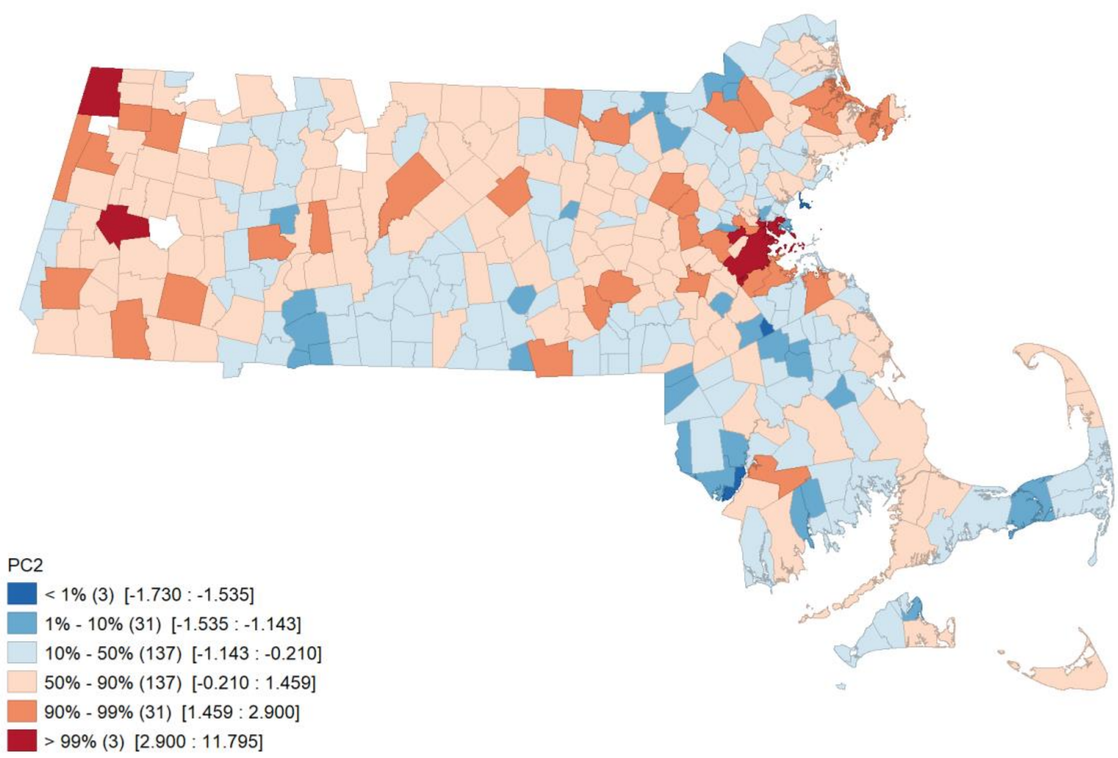

Figure 5. Green Component – Non-motorized transport (NMT) Friendly – PC2

4.1.3. Small household component – PC3

The only variable that significantly loaded highly on component 3 is percentage of the town that have large households (-0.990), and it was negative. Thus component 3 is classified as the small household component. The map of PC3 (Figure 6.) shows Quincy, Chelmsford, and Otis are representative of this component. Other cities such as Brookline, Yarmouth, Wellfleet (Cape Cod) etc. also have this characteristic, and they are randomly located across the state.

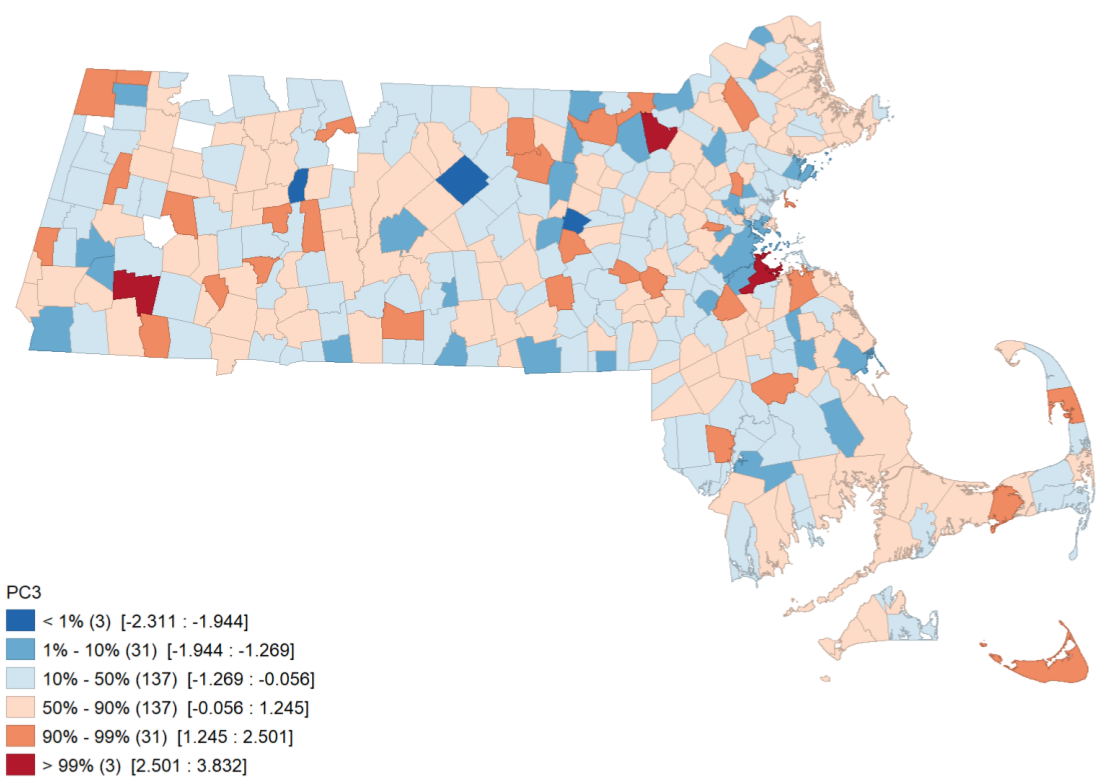

Figure 6. Small Household Component – PC3

4.1.4. Auto-oriented component – PC4

The variables that load highest positively on component 4 are percentage of commuters who drive to work (0.688), and carbon emissions (0.517). Based on these variables the component is classified as measuring the auto-orientedness of a town. The map of PC4 (Figure 7.) shows the high values of this component are clustered in the northern suburbs (Methuen, Andover, North Andover, Tewksbury), near Worcester (Auburn, Millbury, Grafton etc.) and in the suburbs north of Cambridge (Lexington, Burlington, Waltham, Woburn etc.) as well as Boston itself.

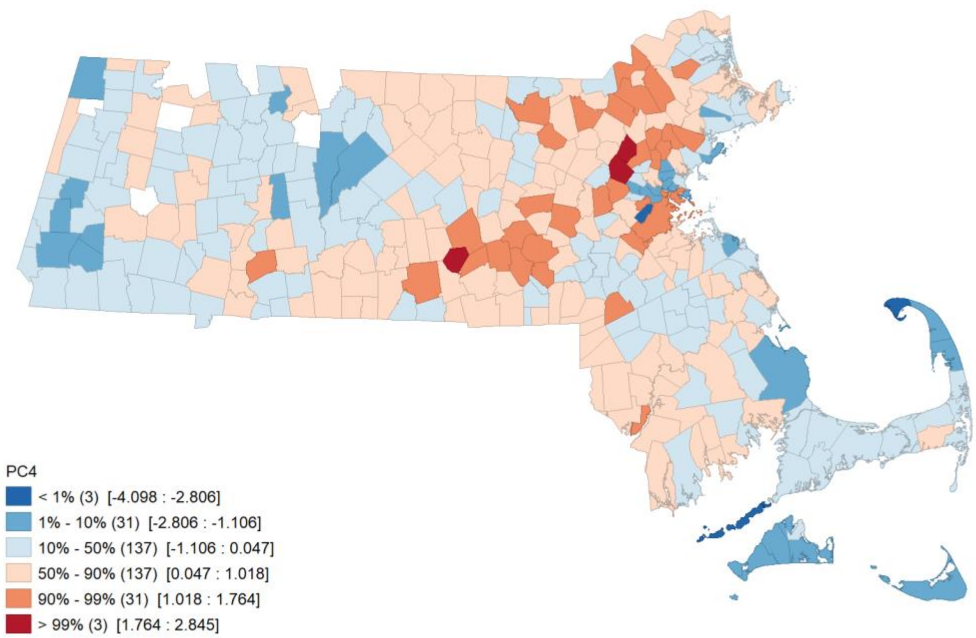

Figure 7. Auto Oriented Component – PC4

4.2. Global spatial regression models and analysis

The spatial regression models were run with the 1st order queen contiguity weight matrix in GeoDa.

Table 4. OLS Result

R^2	Adjusted-R^2	AICc	
0.5475	0.5352	522.131	
Regression Diagnostics			
Multicollinearity Condition Number		37.6214	
Test	MI/DF	Value	Prob
Moran's I (error)	0.2538	7.5881	0.00000
Lagrange Multiplier (lag)	1	40.7882	0.00000
Robust LM (lag)	1	3.0167	0.08241
Lagrange Multiplier (error)	1	50.7990	0.00000
Robust LM (error)	1	13.0275	0.00031
Lagrange Multiplier (SARMA)	2	53.8157	0.00000
Variable		***Coefficient***	***Probability***
Constant		2.4149	0.00000
Gini Index		-1.1656	0.12424
%Household under Poverty		-0.0226	0.04578
%Low Education		0.0251	0.00003
%Minority		0.0088	0.03898
%Renter Occupied Housing Unit		0.0074	0.04561
Rurality Component		-0.1941	0.00000
NMT Green Component		-0.1151	0.00000
Small Household Component		-0.0096	0.72842
Auto-oriented Component		0.1588	0.00000

From the OLS results (Table 4) the p-values suggest that among the nine explanatory variables, the coefficients of seven variables (%household under poverty, %low education, %minority, %renter occupied housing unit, rurality component - PC1, NMT-friendly-green component -PC2, auto-oriented component - PC4) are statistically significant. Thus, although the overall fit of the model is fairly good (Adjusted R^2= 0.54), some independent variables in this model are not significant. In addition, the relevant results show that multicollinearity is still a problem (37.62) and that heteroskedasticity is still a problem (multicollinearity condition number is more than 30, and the result of Koenker-Bassett test is not statistically significant). In terms of residuals, based on the Local Moran's I map (Figure 9.), there are some high value clusters around northeastern Massachusetts (Topsfield, Boxford etc.) and cities located to the west of Cape Cod Bay (Lakeville, Taunton, Raynham), as well as some low value clusters in northwestern Massachusetts (Plainfield, Windsor etc.) and Martha's Vineyard. Furthermore, Moran's I index (0.266 > 0.16, Figure 8.) also demonstrates the clustering trend of residuals. In brief, the random error assumption of OLS has been violated, and the multicollinearity number is high. Therefore, the spatial lag or spatial error models had to be estimated.

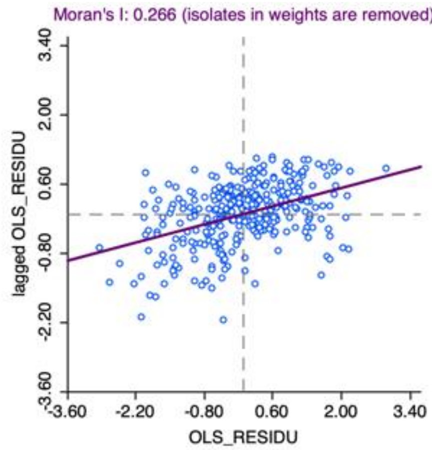

Figure 8. Moran's I Index for OLS residual

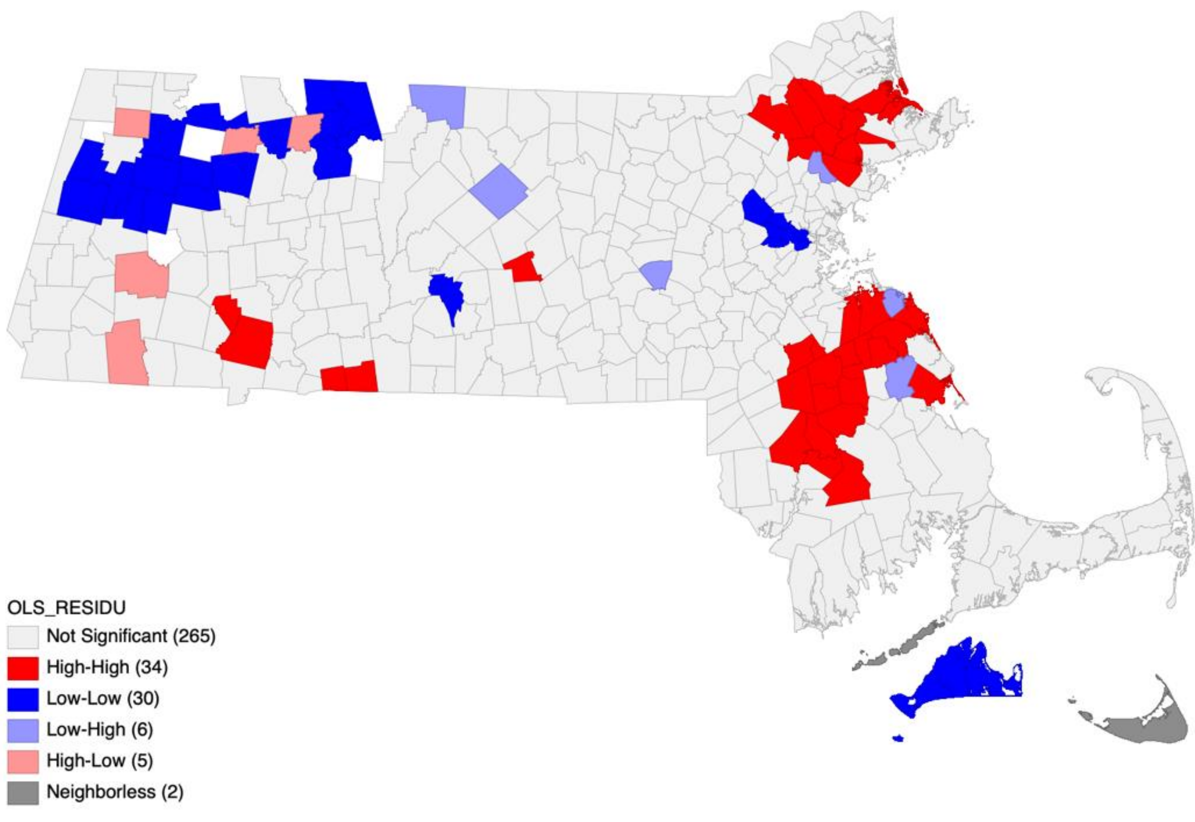

Figure 9. Moran's I Cluster Map for OLS Residual

According to the regression optimization methods from Anselin's Geoda Workbook (Figure 4.) and the OLS diagnostics report, since the results of Robust LM (error) has a higher value than the result of Robust LM (lag) and is statistically significant, the spatial error model is better at dealing with the spatial autocorrelation of residuals and the multicollinearity problem.

Table 5. Spatial Error Result

R^2	AICc		
0.6423	522.131		
Diagnostics for Spatial Dependence			
Test	DF	Value	Prob
Likelihood Ratio Test	1	55.2962	0.00000
Diagnostics for heteroskedasticity			
Test	DF	Value	Prob
Breusch-Pagan test	9	83.8163	0.00000

Variable	***Coefficient***	***Probability***
Constant	2.1947	0.00000
Gini Index	-0.7279	0.30656
%Household under Poverty	-0.0059	0.57922
%Low Education	0.0140	0.00968
%Minority	0.0063	0.12902
%Renter Occupied Housing Unit	0.0092	0.00551
Rurality Component	-0.1647	0.00000
NMT Green Component	-0.0897	0.00012
Small Household Component	0.0079	0.73383
Auto-oriented Component	0.08379	0.01629
LAMBDA	0.54601	0.00000

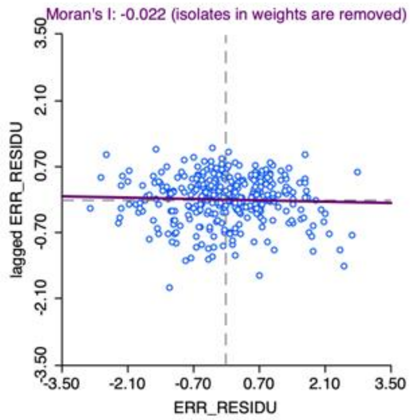

Figure 10. Moran's I Index for Spatial Error residual

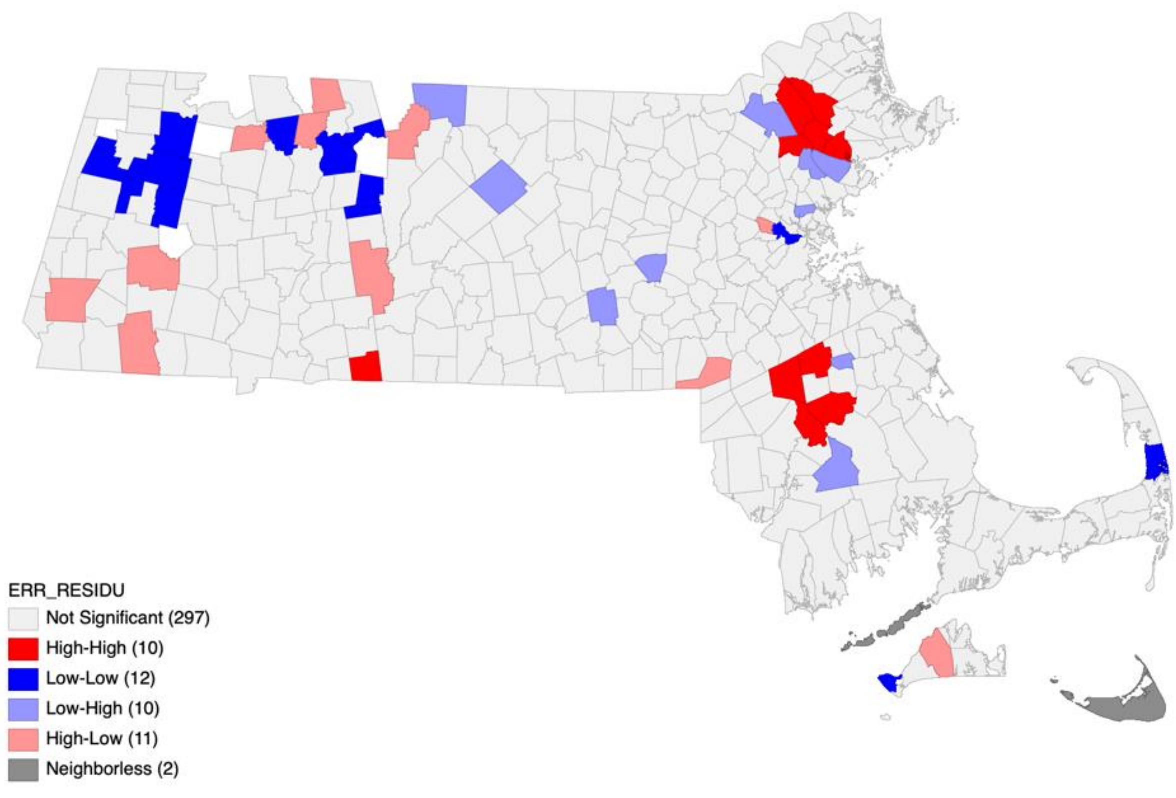

Figure 11. Moran's I Cluster Map for Spatial Error residual

As Table 5 shows, the relevant spatial error model estimates indicated that the overall fit of model improved (R^2= 0.64, which is higher than the result of OLS, meanwhile, Log-Likelihood, AIC and SC are all comparable to the OLS results). Further, as Figure10 and Figure 11 shows, the residuals of this regression model are no longer spatially autocorrelated (Moran's I index is -0.022).

The spatial error results and corresponding p-values suggest that among the nine explanatory variables, only the coefficients of five variables (percentage of low education, percentage of renter occupied housing unit, rurality component - PC1, NMT-friendly-green component -PC2, auto-oriented component - PC4) are statistically significant. The coefficients suggest that for every 1 percent increase of population 25 years old and over whose education level is below high school, the COVID-19 incidence rate per 1000 people increases by 2.76% ($e^{1.015}$); for every 1 percent increase of renter occupied housing units, the COVID-19 incidence rate per 1000 people increases by 2.74% ($e^{1.009}$). As for the principle components, the Rurality Component (PC1), and NMT-friendly-green Component (PC2) are negatively associated with COVID-19 incidence rate suggesting that more rural and more walk and bike friendly towns had lower incidence of COVID-19. The Auto-oriented Component (PC4) is positively associated with COVID-19 incidence rate suggesting that more auto-oriented towns had a higher incidence of COVID-19 in Massachusetts.

4.3. Local spatial regression model - Geographically Weighted Regression (GWR)

Geographically weighted regression (GWR) allows relationships in a regression model to vary over space (Wheeler D.C. 2014). In contrast to traditional linear regression models, which have constant regression coefficients over space, regression coefficients are estimated locally at spatially referenced data points with GWR (Wheeler D.C. 2014).

GWR was estimated first with all the predictor variables, however, the model didn't run successfully because of global multicollinearity in the dataset (from the OLS report, the multicollinearity number is above 30). Therefore, the Gini index was dropped as a predictor variable as it was not significant in either the OLS or Spatial error model and was assumed to be similar to the poverty indicator. After that, the GWR model with 8 predictor variables was estimated. The local R^2 of the model is shown in Figure 12.

Compared to the OLS result, the GWR had higher adjusted R^2 (R^2 = 0.6357 > 0.5352 and smaller AICc (AICc = 475.69 < 522.13), which suggests that the GWR model is a significant improvement upon the OLS specification. The Local R^2 map suggests that the GWR model fit best for the Metro-Boston area approximately (see Figure 12.), and it also fit relatively well in the suburbs surrounding Greater Boston area. The GWR model did not fit well in northwestern Massachusetts and towns near Cape Cod.

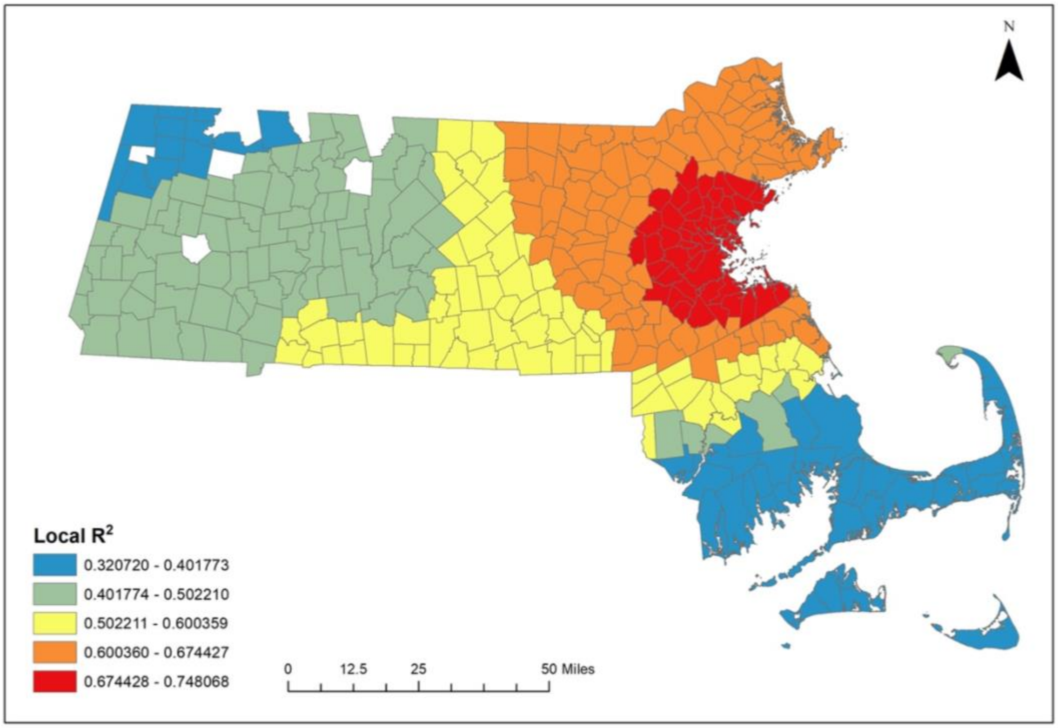

Figure 12. GWR Local R^2

Based on the T-statistic results of the GWR model, among the eight variables, seven are significantly related to the COVID-19 incidence rate locally in Massachusetts (percentage of households in poverty, percentage of low education level population, percentage of minority, percentage of renter occupied housing, rurality component - PC1, NMT-friendly-green component -PC2, auto-oriented component - PC4). However, the small household component – PC3 is not statistically significant. The local T-statistics results and local coefficients for the five indicators which are statistically significant both in the spatial error model and the GWR model are shown respectively in Figures 13 to 22.

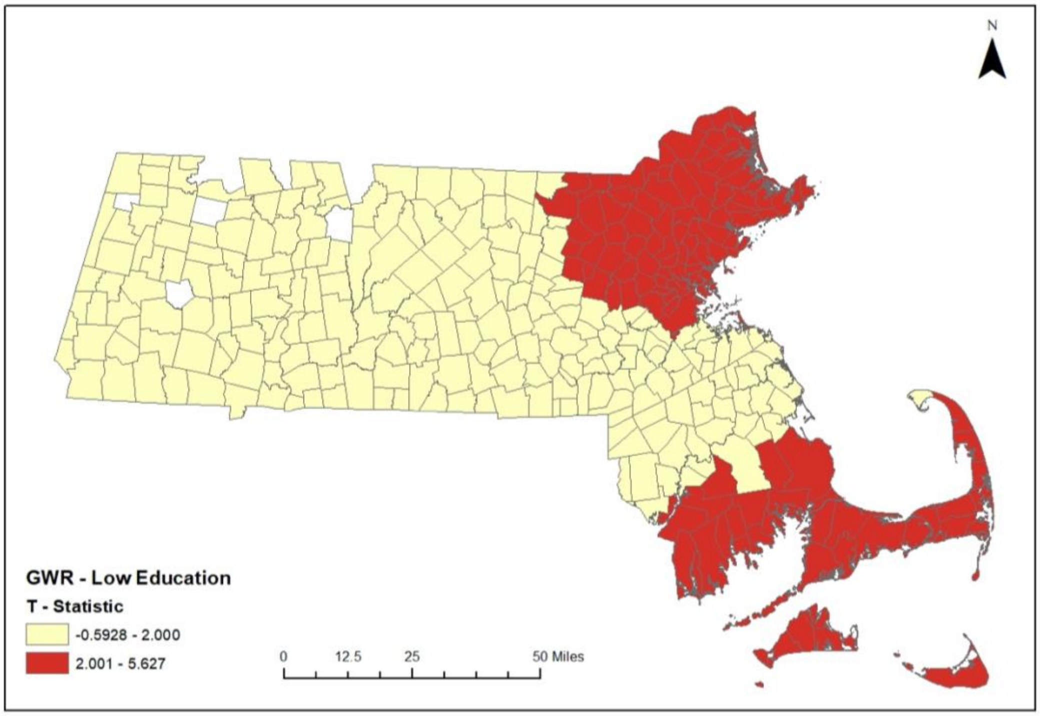

Figure 13. Local T – statistics Results for Percentage of Low Educated Population

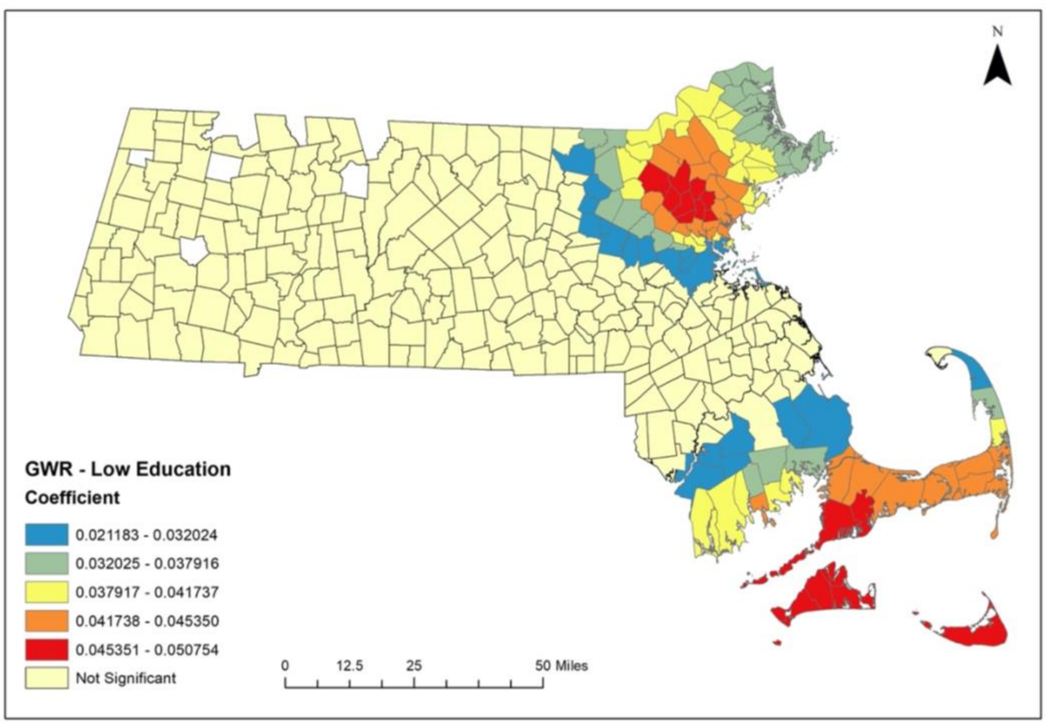

Figure 14. Local Coefficients for Percentage of Low Educated Population

As Figure 13. and Figure 14 show, the percentage of low educated population (population 25 and over whose education level is less than high school) in northern Massachusetts (cities including and to the north of Boston, Brookline, Newton, etc.) as well as cities in South Massachusetts surrounding the Cape Cod Bay, Nantucket Island and Martha's Vineyard are significantly positive in affecting the COVID-19 incidence rate. The highest values for the coefficients are also concentrated in the Islands and a few towns such as Billerica, Wilmington and Burlington.

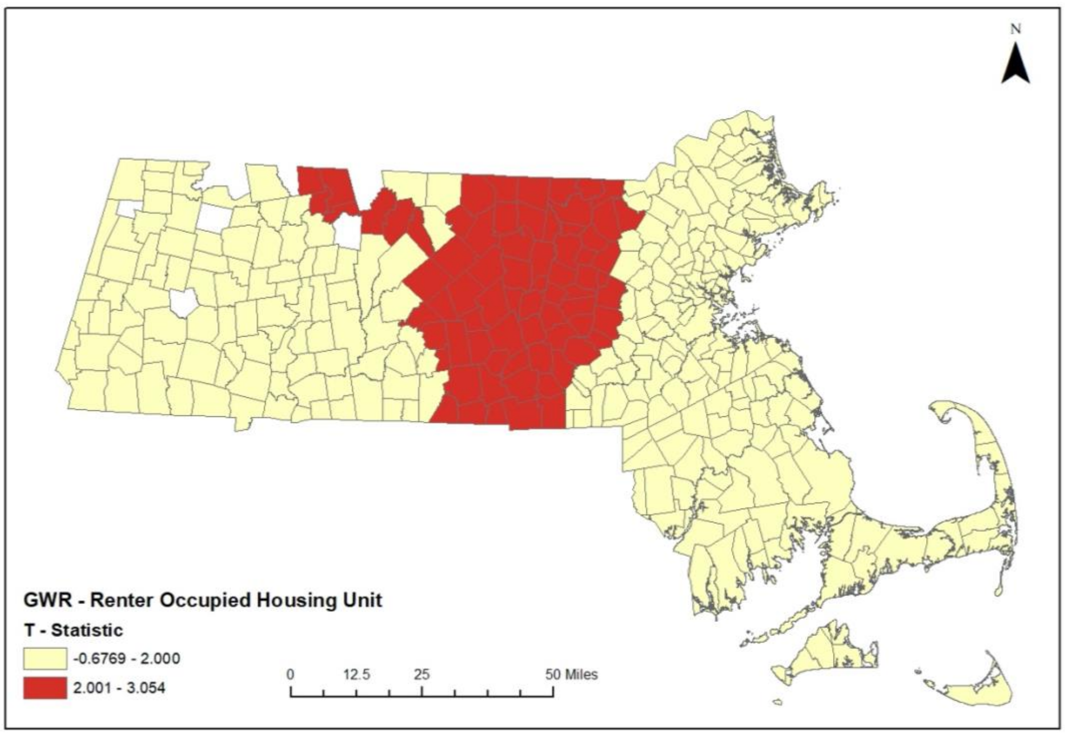

Figure 15. Local T- statistics Results for Percentage of Renter Occupied Housing Units

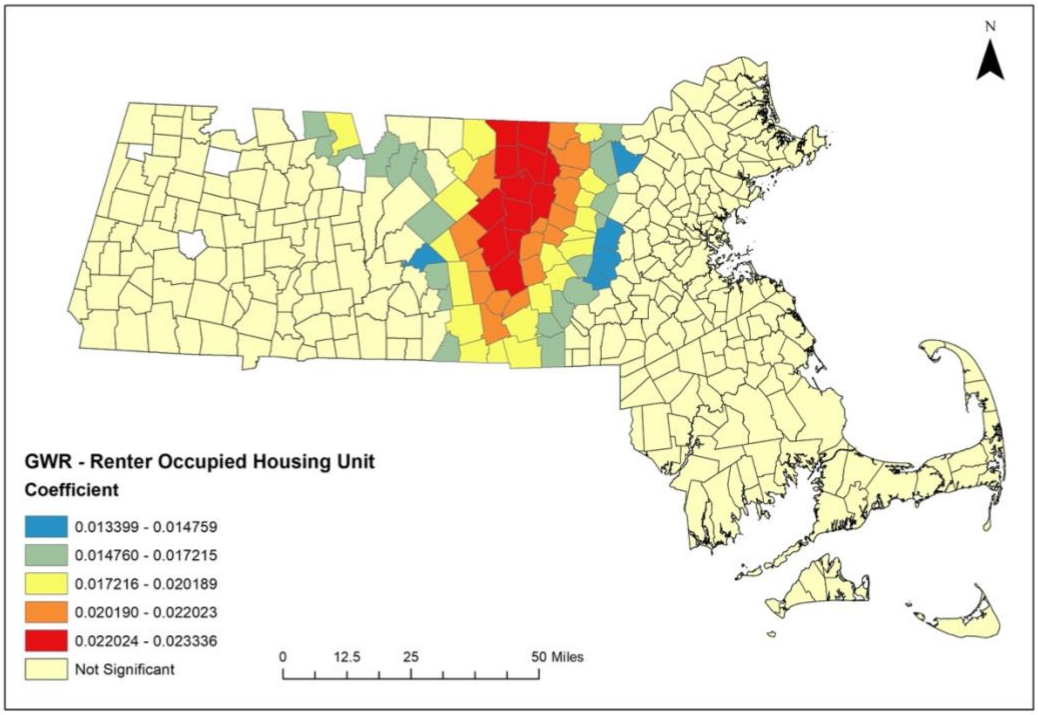

Figure 16. Local Coefficients for Percentage of Renter Occupied Housing Units

As Figure 15 and Figure 16 show, the percentage of renter occupied housing units in cities located in central Massachusetts (Hubbardston, Barre, Oakham, Rutland etc.) are significantly positive in affecting to the COVID-19 incidence rate but this is not true in Boston and its suburbs.

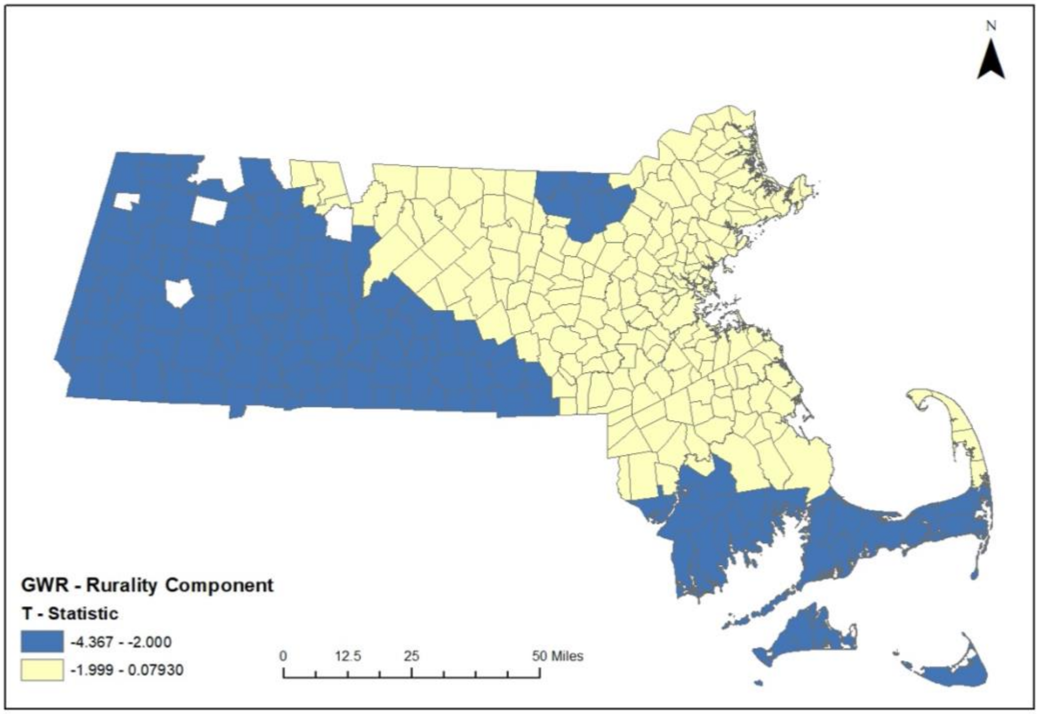

Figure 17. Local T- statistics Results for the Rurality Component

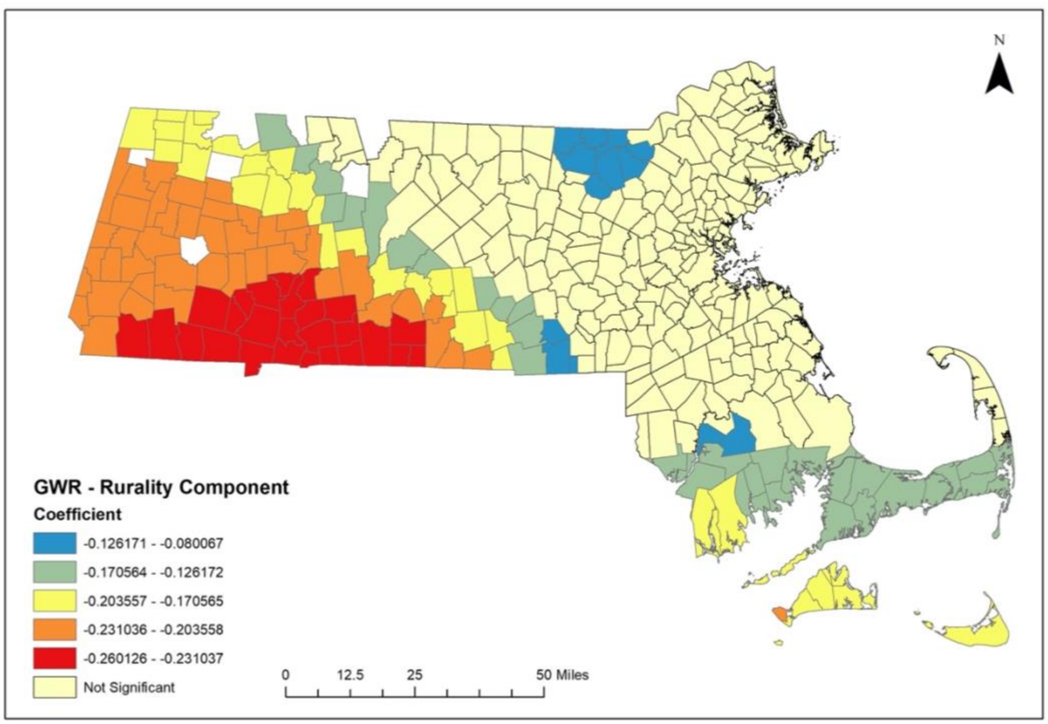

Figure 18. Local Coefficients for the Rurality Component

As Figure 17 and Figure 18 show, the rurality component is not statistically significant in Boston and its surrounding suburbs but has a negative coefficient in western Massachusetts where higher values suggest lower COVID-19 incidence rates. The highest coefficient values are found along the towns that border Connecticut. This suggests that rurality which is also a proxy for population density may not be significant in affecting COVID-19 incidence in the most populated parts of the state.

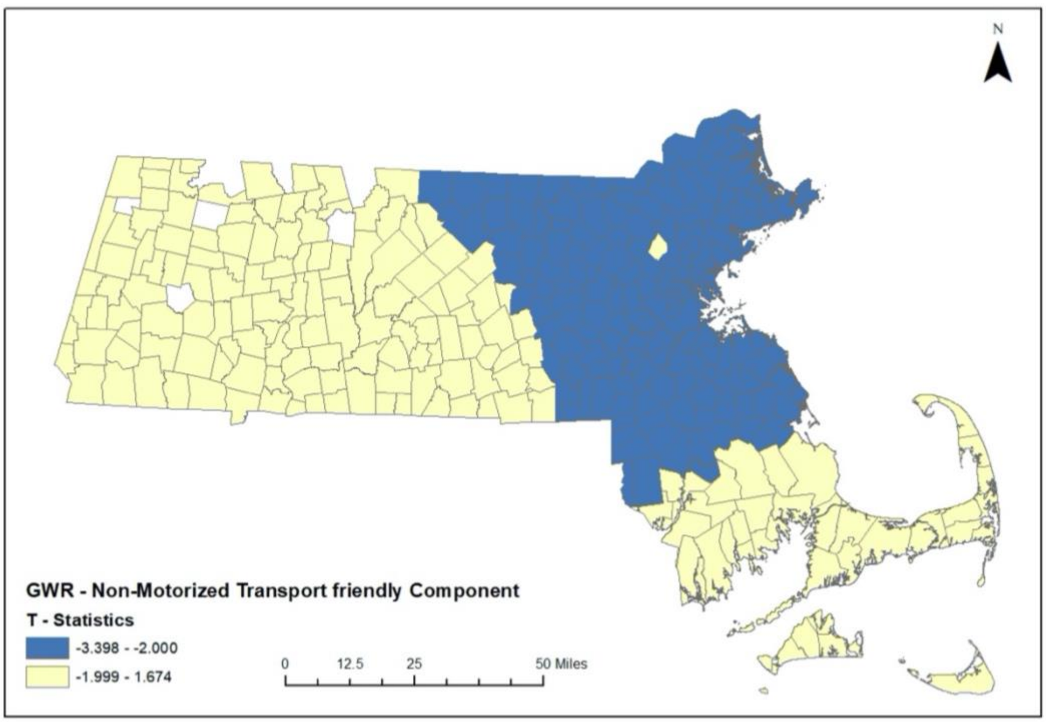

Figure 19. Local T- statistics Results for the non-motorized transport friendly green component

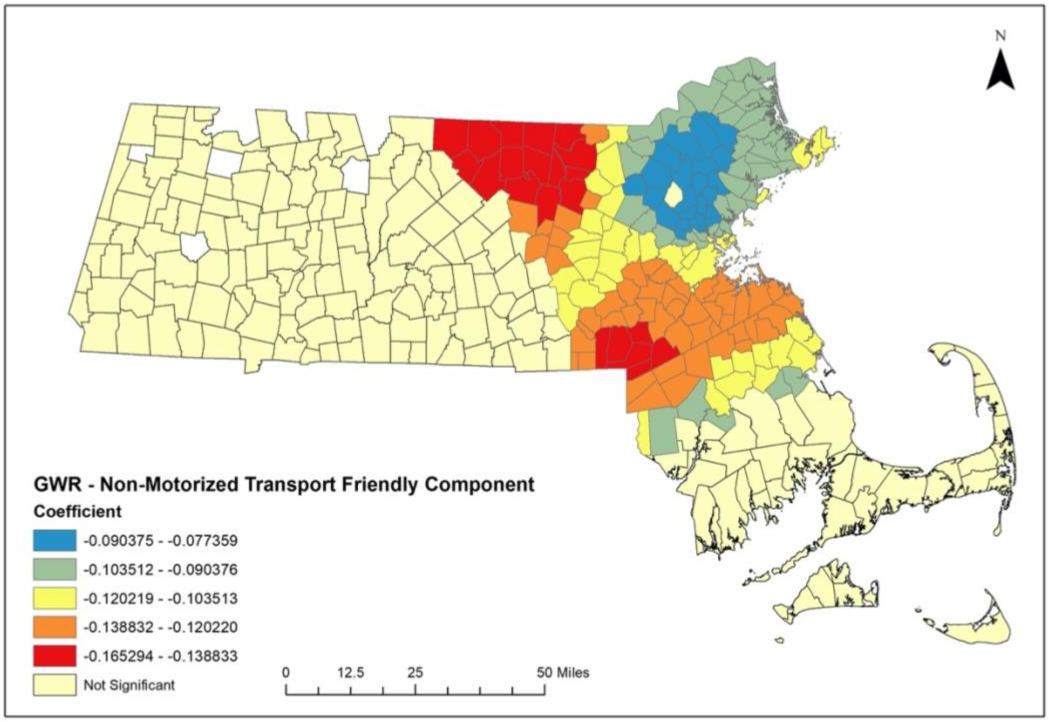

Figure 20. Local Coefficients for the non-motorized transport friendly green component

As Figure 19 and Figure 20 shown, the non-motorized transport friendly green component is significantly affecting the COVID-19 incidence rate locally in Boston and its surrounding suburbs. The sign of coefficients suggests that more bicycle and walk friendly locations in these towns have lower incidence rates. This makes sense in that Boston and its surroundings score better in terms of bicycle friendliness and walkability. However, it is also interesting to note that the locations with the highest coefficients are to be found in Winchendon, Ashburnham as well as Plainville and Wrentham, which are quite far from the most populous parts of the metro area.

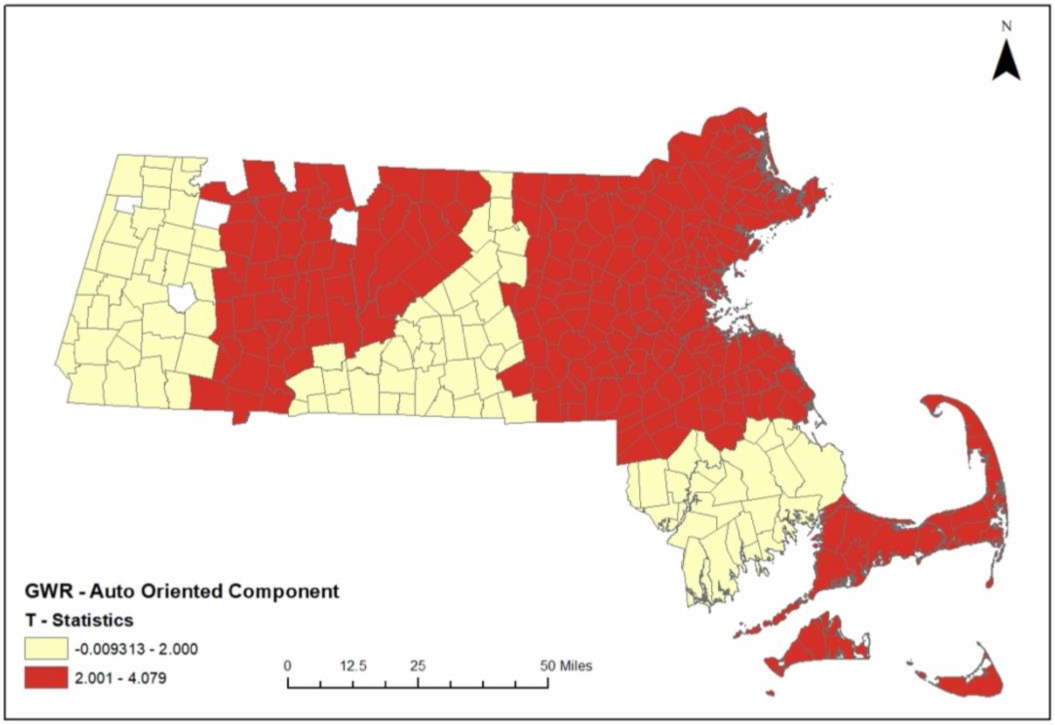

Figure 21. Local T- statistics Results for the Auto-Oriented Component

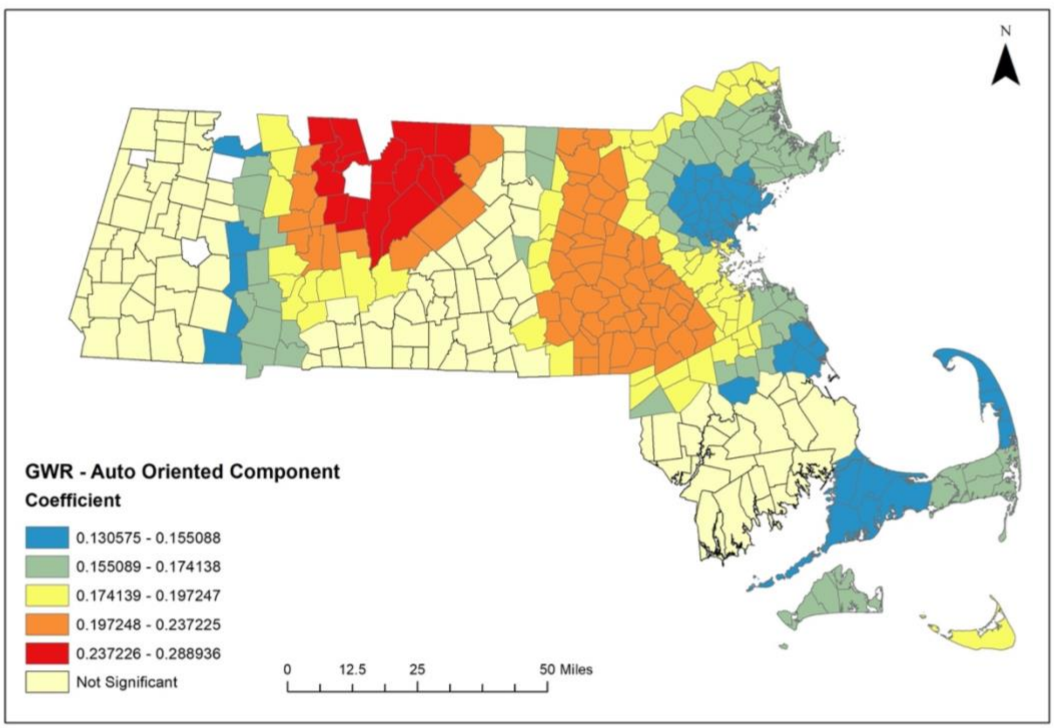

Figure 22. Local Coefficients for the Auto-Oriented Component

As Figure 21 and Figure 22 show, the auto-oriented component is significantly related to COVID-19 incidences rate for most Massachusetts cities, except for some cities located in western, and central Massachusetts, and many cities located near New Bedford. The sign of coefficients suggests that the towns in red shown in Figure 21 and 22 appear to show a high correlation between being more auto-oriented and higher rates of COVID-19 incidence. The coefficient was highest in towns near Royalston and Winchendon in north western Massachusetts which suggests that this was not as much of a factor in affecting incidence in Boston and its suburbs.

4.4. Comparison of Results

According to Table 6, both the spatial error model and GWR model outperformed the OLS model with higher adjusted R^2s and lower AICcs. The spatial error model obtained the highest adjusted R^2 and lowest AICc compared to OLS and GWR, suggesting that it is the best fit model to measure the spatial variability of COVID-19 incidence rate.

Table 6. Comparison of Results

Models	OLS	SE	GWR
Adj. R^2	0.5352	0.6423	0.6357
AICc	522.131	4 66.835	475.69

5. Conclusions, limitations and future directions for policies

5.1. Conclusions

In this GIS based research, I compiled variables that could potentially explain the spatial variability in the COVID-19 incidence rate at the town level across the Massachusetts state. These variables were grouped into two different themes, which were socio-economic and built environment. The variables were used to model the spatial variability of COVID-19 incidence rate using OLS, spatial regression and GWR models.

Based on the findings from the regression models (OLS & Spatial Error), five indicators: percentage of low educated population, percentage of renter-occupied housing units, rurality component, NMT- friendly component, and auto-oriented component, significantly explain the spatial variability of COVID-19 incidence in Massachusetts. The percentage of low educated population, percentage of renter-occupied housing units, and auto-oriented component are positively associated with the incidence rate while the rurality component and NMT-friendly component are negatively associated with the COVID-19 incidence rate.

The negative effect of education level and rental housing to COVID infection confirms that urban inequity problems are important factors that planners need to address. Meanwhile, in both global and local models, the Auto-oriented Component was significantly related to the disease incidence rate positively while the rurality component and NMT-friendly was negatively related. This suggests that the built environment (walkability, bike friendliness, proximity to green spaces, driving as the commute method) appeared to affect the incidence rate. Therefore, built environment could be key in preventing current disease spread and post-COVID planning. Since the small household component was not statistically significant in either the global and local models, it suggests that many people living together in a household may not be as influential in the context of COVID control in Massachusetts as previously thought.

The GWR model suggested that the predictors for the incidence rate vary spatially across the state. For example, the rurality component is significant in western Massachusetts while percentage of renter-occupied housing unit is significant in central Massachusetts. In addition, the Auto-oriented component, NMT component, and education level are significant in the northern suburbs of Massachusetts. Thus, in order to address the current public health crisis and promote sustainable development in the future, built environment and public health policies need to be localized based on the character of each region.

To summarize, this thesis found that supporting sustainable development by addressing the built environment through supporting NMT and incorporating green spaces could be viable strategies for controlling future pandemics like COVID-19.

Reducing inequities in education and access to housing will also continue to be key in ensuring public health safety.

5.2. Limitations

Compared to previous studies based on counties (Mollalo, Vahedi, and Rivera 2020) or European countries (Sannigrahi et al. 2020b), the spatial unit of this study – city/town – is relatively small, which helps estimate more fine grained results that can help city planners. However, both predictor and outcome variables vary within towns and may not be evenly distributed, which could affect the conclusions from the estimated models.

To best of my knowledge, the finest spatial unit data for Massachusetts COVID-19 cases that is publicly available is at the city level. In addition, most socio-economic and built environment data are not available in the most recent year. Therefore, data availability is one of the limitations of this study.

Some aspects of the built environment such as natural resource and waste management were not included as explanatory variables in this study because those data are not available at the town or smaller level. They should be incorporated in future studies to better understand the influence of these features of the built environment on COVID-19.

In addition, OLS, Spatial Error, and GWR assume the relationship between the dependent and independent variables are log-linear. However, this may not accord with reality. Also, the relationship being measured in this study is not able to determine a causal relationship. Thus, when it comes to the results, it is hard to estimate which predictor is substantially influential. Further studies should consider using more complex models , such as adding quadratic terms for predictors or interaction terms between predictors to predict the infection rate and compare the results.

Finally, this study used data based on reported illnesses and there may have been more cases than reported. Furthermore, the model results were significant in Massachusetts and other states may have very different results.

5.3 Policy Implications

This study has confirmed to an extent that education level and housing conditions as well as built environment factors could significantly predict COVID-19 incidence rate in Massachusetts. By contrast, large households did not appear to be a significant predictor in Massachusetts. The results suggest that urban planners and policymakers need to work on addressing inequity of access not only in education and housing but also in access to non-motorized vehicle based travel and green spaces in communities which appear to have been harder hit by the COVID-19 pandemic in Massachusetts. It is vital that policy makers continue to promote sustainable development by addressing the built environment under new norms for public health.

One finding of this study is that cities with a higher percentage of poorly educated population are more vulnerable to COVID-19. While it may be impossible to reduce the inequity in education in the short term, it is possible to build the COVID-19-related information sharing system based on different target audiences.

Results from the GWR model suggest that different indicators predict the incidence rate differently across the state. Acknowledging these local differences, municipal governments should focus on the variables that have significant impact on the local area. In the long-term, policymakers should support the development and implementation of local and regional plans that address public health.

Of the four built environment components derived, three were significant both in the overall regression models and locally in the GWR. These components measured the rurality, NMT-friendliness , and the Auto-orientedness of a town. Variables that loaded highly on those components included: green space, percentage of tree canopy coverage, percentage of commuters drive to work, carbon emissions, walkability, and bike friendliness. Improving green space access and tree canopy may be key to allowing residents to socially distance successfully. Other studies have suggested that pollution exposure may play a role in COVID-19 incidence (Ramirez and Lee, 2020) and our results appear to confirm this through the proxy for green spaces which are likely to be locations with lower pollution.

Non-motorized transport by walking and biking could also be significant ways in which residents could commute more safely when they must. Bike lane and sidewalk projects may be vital for safe socially distanced NMT based commuting. To summarize my thesis suggests that planners must address the built environment to improve both the health of urban residents and make cities more sustainable.

Appendix

A. Meta-Review of urban Sustainability Indicator Sets from *Committee on Pathways to Urban Sustainability: Challenges and Opportunities*

		Urban Sustainability Systems				
Measures		**I**	**II**	**III**	**IV**	**V**
Environmental Indicators						
Air quality	Air Quality index[II, IV, V]; criteria pollutant nonattainment[V]; nitrogen oxides emissions[I]; sulfur dioxide emissions[I]; $PM_{2.5}$ emissions[III, V]; PM_{10} emissions[I, III]					
Greenhouse gas emissions	Residential greenhouse gas (GHG) emissions[I, V]; commercial GHG emissions[I, V]; industrial GHG emissions[II, V]; total greenhouse gases (CO_2, CH_4, N_2O, and chlorofluorocarbons [CFCs]) [II, III, IV]; CO_2 emissions by the energy sector divided by the total electricity output[IV]; annual amount of carbon dioxide emissions divided by the city population[IV]	2	2	1	3	3
Water	Average annual precipitation per year[V]; example of a waterway applicable to the city[V]; number of waterways impaired[IV, V]; water leakages in water distribution system[I]; total water consumption[II]; water consumption per capita[I, IV]; drinking water quality[III]; water usage[IV, V]	2	1	1	3	4
Land	Green space[I, II, III]; existing tree canopy[IV, V]; landslide vulnerability[IV, V]; park acres per 1,000 residents[IV,V]; urban sprawl[I, III]	2	1	2	3	3
Waste	Total solid waste production[IV]; percent of municipal solid waste recycled [I, IV]; waste management indicator[II]; solid waste management[III]	1	1	1	2	—
Ecological footprint	Ecological footprint by Global Footprint Network[V]	—	—	—	—	1
Natural hazards vulnerability	Natural hazards vulnerability[IV, V]; natural catastrophe exposure[III]	—	—	1	1	1
Economic Indicators						
Income	City income[II]; gross domestic product (GDP) per capita[I, III]; median household income[V]	1	1	1	—	1
Price	Consumer Price Index[III, IV]; average residential electricity rate[V]	—	—	1	1	1
Unemployment	Unemployment[II, IV, V]; goods employment[I]; services employment[I]; employment by mix of economic sectors[V]	2	1	—	1	2
Energy	Energy consumption[II, III]; energy consumption per capita[IV]; residential energy intensity[IV, V]; commercial energy intensity[V]; industrial energy intensity[V]; electricity consumption per person[I]; electricity consumption per unit of GDP[I]; energy efficiency[III]; renewable energy consumption[IV, V]; share of renewable energy in energy mix[III]; system average interruption duration index[V]; light-emitting diode (LED) street lighting[V]	2	1	3	3	6
Financial health	G.O. Bond ratings 2014 or 2015 S&P ratings or *Moody's – AA+ is S&P equivalent to Aa1[V]; viability of the urban economy[II]; city fiscal deficit[II]; indicator tracks the performance of banks and thrifts in meeting the credit needs of the community by using Community[IV]; reinvestment Act lender ratings[IV]	—	2	—	2	1

	Measures	Urban Sustainability Systems I	II	III	IV	V
Transportation	Transportation mode share[IV, V]; share of workers traveling by public transit, bicycle, or foot[I, II, IV]; annual vehicle miles of travel per capita[IV, V]; public transportation ridership average[IV, V]; mean travel time to work in minutes[I, III, IV, V]; Walkscore.com[IV, V]; length of transport infrastructure[I, III]; congestion[V]; yearly delay[V]; excess fuel[V]; cost[V]; licensed drivers per 1,000 driving age population[V]	3	1	2	6	10
Social Indicators						
Demographics	Population[I, V]; population density[I, V]; race, ethnicity, age, and gender[V]	2	—	—	—	3
Education	High school, college, and bachelor's degrees[IV, V]; university rankings[III]; literacy[III]	—	—	2	1	1
Public health	Poor or fair health[V]; adult obesity[V]; premature age-adjusted mortality[V]; under age 5 mortality per 1,000 live births[IV]; life expectancy at birth[III]; percent of population with health insurance[IV]; total percentage of the population affected seriously by crime or traffic accidents[II, IV]; rate of violent crimes[IV, V]; roadway fatalities per hundred million annual vehicle miles of travel (VMT)[V]	—	1	1	4	5
Equity	Ratio of household income at the 80th% to income 20th%[V]; Gini coefficient[III, IV]; percentage of city population living in poverty[IV, V]; percentage of people affected by poverty, unemployment, lack of access to education, information, training, and leisure[III]; percentage of low-income households within 1/4 mile of a neighborhood center and a transit stop[II]; ratio of economically active population to economically inactive population[III]; children in poverty[IV, V]	—	1	3	3	3
Housing and buildings	Housing affordability[IV]; home ownership[V]; percentage of the homeless population[II]; percentage of the population affected by poor housing conditions[II, IV]; numbers of Leadership in Energy & Environmental Design (LEED)-certified buildings[I]; number of houses certified as energy efficient by certification organizations[IV]; median house size of new construction[IV]	1	2	—	4	1
Citizen participation	Percentage of people participating in local election or as active members in associations for urban improvement and quality of life[II, IV]; voter participation as a percentage of the population[IV]	—	1	—	2	—

NOTE: Superscripts indicate the urban sustainability system that uses each measure indicated in the right-hand columns: I = America Green City Index; II = Urban Sustainability Indicators; III = Sustainable Cities Index; IV = Sustainability Urban Development Indicators; V = The Academies' Pathways to Urban Sustainability Committee. The number in each cell represents the number of measures represented in each urban sustainability indicator set.

Bibliography

Adekunle, Ibrahim Ayoade, Abayomi Toyin Onanuga, Olanrewaju Olugbenga Akinola, and Olakitan Wahab Ogunbanjo. 2020. “Modelling Spatial Variations of Coronavirus Disease (COVID-19) in Africa.” *Science of The Total Environment* 729 (August): 138998. https://doi.org/10.1016/j.scitotenv.2020.138998.

Ali, Shahmir H, Joshua Foreman, Yesim Tozan, Ariadna Capasso, Abbey M Jones, and Ralph J DiClemente. 2020. “Trends and Predictors of COVID-19 Information Sources and Their Relationship With Knowledge and Beliefs Related to the Pandemic: Nationwide Cross-Sectional Study.” *JMIR Public Health and Surveillance* 6 (4): e21071. https://doi.org/10.2196/21071.

Cao, Yang, Ayako Hiyoshi, and Scott Montgomery. 2020. “COVID-19 Case-Fatality Rate and Demographic and Socioeconomic Influencers: A Worldwide Spatial Regression Analysis Based on Country-Level Data.” Preprint. Public and Global Health. https://doi.org/10.1101/2020.07.31.20165811.

Caron, Christina. 2020. “Are Cities a Safe Place to Live During a Pandemic?” *The New York Times*, December 7, 2020, sec. Well. https://www.nytimes.com/2020/12/07/well/live/coronavirus-cities-safe.html.

CDC. 2020. “Communities, Schools, Workplaces, & Events.” Centers for Disease Control and Prevention. April 30, 2020. https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html.

“CDC COVID Data Tracker.” n.d. Accessed July 29, 2020. https://www.cdc.gov/covid-data-tracker/#cases.

“Citizen Centric Cities The Sustainable Cities Index 2018.” 2018 Arcadis United States https://www.arcadis.com/en/united-states/our-perspectives/sustainable-cities-index-2018/united-states/

Committee on Pathways to Urban Sustainability: Challenges and Opportunities, Science and Technology for Sustainability Program, Policy and Global Affairs, and National Academies of Sciences, Engineering, and Medicine. 2016. *Pathways to Urban Sustainability: Challenges and Opportunities for the United States*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/23551.

Cordes, Jack, and Marcia C. Castro. 2020. “Spatial Analysis of COVID-19 Clusters and Contextual Factors in New York City.” *Spatial and Spatio-Temporal Epidemiology* 34 (August): 100355. https://doi.org/10.1016/j.sste.2020.100355.

“Coronavirus COVID-19 (2019-NCoV).” 2020 Accessed August 7, 2020. https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6.

"Coronavirus Disease 2019 (COVID-19) - Symptoms and Causes." 2020 Mayo Clinic. Accessed August 7, 2020. https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/syc-20479963.

"Coronavirus Statistics by Country." 2020 Coronavirus Dashboard. Accessed August 7, 2020. https://covid-019.com/countries.

Cutaia, F. n.d. *The Use of Landscape Indicators in Environmental Assessment BT -Strategic Environmental Assessment: Integrating Landscape and Urban Planning. In: F. Cutaia (Ed.). Cham: Springer International Publishing. Pp. 29–43. Doi: 10.1007/ 978-3-319-42132-2_4.*

"Five Ideas to Turn the COVID-19 Recovery into a Global Green New Deal." 2020 SEI. Accessed December 15, 2020. https://www.sei.org/featured/5-ideas-to-turn-the-covid-19-recovery-into-a-global-green-new-deal/.

Franch-Pardo, Ivan, Brian M. Napoletano, Fernando Rosete-Verges, and Lawal Billa. 2020. "Spatial Analysis and GIS in the Study of COVID-19. A Review." *Science of The Total Environment* 739 (October): 140033. https://doi.org/10.1016/j.scitotenv.2020.140033.

Gao, Song, Jinmeng Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. n.d. "Mapping County-Level Mobility Pattern Changes in the United States in Response to COVID-19," 11.

Guliyev, Hasraddin. 2020. "Determining the Spatial Effects of COVID-19 Using the Spatial Panel Data Model." *Spatial Statistics* 38 (August): 100443. https://doi.org/10.1016/j.spasta.2020.100443.

Hassan, Mohammad Mahmudul, Abul Kalam, Shahanaj Shano, Md. Raihan Khan Nayem, Md. Kaisar Rahman, Shahneaz Ali Khan, and Ariful Islam. 2020. "Assessment of Epidemiological Determinants of COVID-19 Pandemic Related to Social and Economic Factors Globally." Preprint. LIFE SCIENCES. https://doi.org/10.20944/preprints202006.0346.v1.

"How COVID-19 Will Shape Urban Mobility." 2020. BCG Global. July 19, 2020. https://www.bcg.com/publications/2020/how-covid-19-will-shape-urban-mobility.

"How the COVID-19 Crisis Can Accelerate Sustainable Urban Mobility." 2020 World Business Council for Sustainable Development (WBCSD). Accessed December 15, 2020. https://www.wbcsd.org/0ajkq.

"Introduction to Residuals and Least Squares Regression (Video)." n.d. Khan Academy. Accessed December 15, 2020. https://www.khanacademy.org/math/statistics-probability/describing-relationships-quantitative-data/regression-library/v/introduction-to-residuals-and-least-squares-regression.

Keesmaat, Jennifer. 2020. "The Pandemic Does Not Spell the End for Cities," May 28, 2020. https://www.foreignaffairs.com/articles/north-america/2020-05-28/pandemic-does-not-spell-end-cities.

Kim, Jong Hae. 2019. "Multicollinearity and Misleading Statistical Results." *Korean Journal of Anesthesiology* 72 (6): 558–69. https://doi.org/10.4097/kja.19087.

King, Lester O. 2016. "Functional Sustainability Indicators." *Ecological Indicators* 66 (July): 121–31. https://doi.org/10.1016/j.ecolind.2016.01.027.

"List of United States Cities by Population Density." 2020. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_United_States_cities_by_population_density&oldid=969836437.

Luc Anselin, Anil K. Bera 1998 "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics" https://dces.wisc.edu/wp-content/uploads/sites/128/2013/08/W7_AnselinBera1998.pdf

"Massachusetts City and Town Websites." n.d. Mass.Gov. Accessed July 29, 2020. https://www.mass.gov/lists/massachusetts-city-and-town-websites.

Mega, Voula, and Jørn Pedersen. 1998. *Urban Sustainability Indicators*. EF, 98,07 EN. Luxembourg: Off. for Off. Publ. of the Europ. Communities.

Mollalo, Abolfazl, Behzad Vahedi, and Kiara M. Rivera. 2020. "GIS-Based Spatial Modeling of COVID-19 Incidence Rate in the Continental United States." *Science of The Total Environment* 728 (August): 138884. https://doi.org/10.1016/j.scitotenv.2020.138884.

Neshovski, Robert. n.d. "Home." *United Nations Sustainable Development* (blog). Accessed August 4, 2020. https://www.un.org/sustainabledevelopment/.

"Official List of Proposed SDG Indicators." 2016 Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs) https://sustainabledevelopment.un.org/content/documents/11803Official-List-of-Proposed-SDG-Indicators.pdf

Ozgen, Ceren, Eric Strobl, and Matt Cole. n.d. "Air Pollution Exposure Linked to Higher COVID-19 Cases and Deaths – New Study." The Conversation. Accessed July 29, 2020. http://theconversation.com/air-pollution-exposure-linked-to-higher-covid-19-cases-and-deaths-new-study-141620.

Pearson, Karl. 1901. "LIII. *On Lines and Planes of Closest Fit to Systems of Points in Space*." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11): 559–72. https://doi.org/10.1080/14786440109462720.

Ramírez, Ivan J., and Jieun Lee. 2020. “COVID-19 Emergence and Social and Health Determinants in Colorado: A Rapid Spatial Analysis.” *International Journal of Environmental Research and Public Health* 17 (11): 3856. https://doi.org/10.3390/ijerph17113856.

Rinne, Janne, Jari Lyytimäki, and Petrus Kautto. 2013. “From Sustainability to Well-Being: Lessons Learned from the Use of Sustainable Development Indicators at National and EU Level.” *Ecological Indicators* 35 (December): 35–42. https://doi.org/10.1016/j.ecolind.2012.09.023.

Rocklöv, J, and H Sjödin. 2020. “High Population Densities Catalyze the Spread of COVID-19.” *Journal of Travel Medicine*, March. https://doi.org/10.1093/jtm/taaa038.

Sannigrahi, Srikanta, Francesco Pilla, Bidroha Basu, Arunima Sarkar Basu, and Anna Molter. 2020. “Examining the Association between Socio-Demographic Composition and COVID-19 Fatalities in the European Region Using Spatial Regression Approach.” *Sustainable Cities and Society* 62 (November): 102418. https://doi.org/10.1016/j.scs.2020.102418.

Shen, Li, and Xulin Guo. 2014. “Spatial Quantification and Pattern Analysis of Urban Sustainability Based on a Subjectively Weighted Indicator Model: A Case Study in the City of Saskatoon, SK, Canada.” *Applied Geography* 53 (September): 117–27. https://doi.org/10.1016/j.apgeog.2014.06.001.

“Social Distancing Revives America’s Suburban Instincts - The Boston Globe.” n.d. Accessed July 28, 2020. https://www.bostonglobe.com/2020/03/16/arts/social-distancing-revives-americas-suburban-instincts/.

Steiniger, Stefan, Elizabeth Wagemann, Francisco de la Barrera, María Molinos-Senante, Rodrigo Villegas, Helen de la Fuente, Alejandra Vives, et al. 2020. “Localising Urban Sustainability Indicators: The CEDEUS Indicator Set, and Lessons from an Expert-Driven Process.” *Cities* 101 (June): 102683. https://doi.org/10.1016/j.cities.2020.102683.

"The GREEN CITIES CHALLENGE." 2011.Canadian Business, Oct 24, 17-21. https://login.ezproxy.library.tufts.edu/login?url=https://search-proquest-com.ezproxy.library.tufts.edu/docview/918793861?accountid=14434.

Verma, Pramit, and A.S. Raghubanshi. 2018. “Urban Sustainability Indicators: Challenges and Opportunities.” *Ecological Indicators* 93 (October): 282–91. https://doi.org/10.1016/j.ecolind.2018.05.007.

“We Were Warned Not to Bunch Up | Newgeography.Com.” n.d. Accessed July 28, 2020. https://www.newgeography.com/content/006581-we-were-warned-not-bunch-up.

“What We Know About Density and Covid-19’s Spread - Bloomberg.” n.d. Accessed July 28, 2020. https://www.bloomberg.com/news/articles/2020-04-03/what-we-know-about-density-and-covid-19-s-spread.

Wheeler D.C. (2014) Geographically Weighted Regression. In: Fischer M., Nijkamp P. (eds) Handbook of Regional Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23430-9_77

You, Heyuan, Xin Wu, and Xuxu Guo. 2020a. "Distribution of COVID-19 Mortality Rate in Association with Social and Economic Factors in Wuhan, China: Implications for Urban Development." *International Journal of Environmental Research and Public Health* 17 (10): 3417. https://doi.org/10.3390/ijerph17103417.

"10 Sustainability Resources and Programs in MA" n.d Mass.gov. https://www.mass.gov/files/documents/2017/06/zk/10%2520Sustainability%2520Resources%2520and%2520Programs%2520in%2520MA.pdf