THE UNIVERSITY OF CHICAGO

Urban Neighborhoods and Depression: A Spatial Examination of Depression within Areas of Varying Sociodemographic Characteristics

By

Kaylah Thomas

June 2022

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Marc Berman

Preceptor: Elizabeth Huppert

Abstract

Urban neighborhood characteristics are vastly diverse and can help to explain social patterns and mental health outcomes. A granular examination of neighborhood sociodemographic factors and the patterns of mobility that are created by the residents within these areas can reveal influences on depression rates. The close ties between sociodemographic factors and neighborhood environmental characteristics, such as mobility, can serve as indicators for why some urban areas are more affected than others. Using data from the American Community Survey, as well as collected geolocational data and depression instances from Twitter users in Chicago and outlying areas, regression and spatial autocorrelational models were developed to determine sociodemographic factor effects on mobility and depression rate in urban neighborhoods. Analysis shows that the percent population of white individuals positively impacted the radius of gyration for a given area, while percent population Hispanic, higher level education, and a higher Gini wealth inequality index has a negative effect on radius of gyration. Models including mobility as an influencing factor of depression show that an increased radius of gyration has a significant effect on higher depression rates, while a higher Gini wealth inequality index also decreases depression rates. All variables were spatially significant, and areas that share similar sociodemographic, mobility, and depression patterns cluster in the city center as well as the outlying areas of Chicago. These findings can help illustrate which neighborhood characteristics should be closely examined when determining patterns of depression, as well as the identifying sociodemographic variables that can be monitored to help support areas with higher depression rates than others.

Acknowledgements

I would like to show my appreciation to all those that helped me throughout my journey to produce this thesis. Thank you to my Advisor Marc Berman and all those at the Environmental Neuroscience Lab for sharing their knowledge and advice with me. I would like to especially thank Andrew Stier for his guidance in the construction and writing of this paper as well as his direction regarding the computational processes that I learned along the way. I would also like to thank my friends in the MACSS program for their support and comradery, and finally my family and friends from home, Simone, David, David W., Ryan, and Ivy, who have supported me near and far throughout all my endeavors. Thank you.

Contents

Introduction

The effects of city living have been long discussed within many social science fields, such as psychology, sociology, and urban planning. Even within cities, there is vast diversity on a demographic and socioeconomic scale. It is important to consider these factors and their impacts on the mental health of people residing within the neighborhoods of urban areas. Depression is an especially concerning mental health issue due to negative consequences like exacerbating other illnesses, poor individual social and economic outcomes, and possible effects on individual morbidity (Gaynes et al., 2002, Simon, 2003; Mykletun et al., 2009). The levels to which the surrounding environment impacts people in urban areas is an important question to pose when considering the public health of cities. As a result, the present research makes an effort to understand the effects that local urban environment characteristics (e.g., mobility, race, education, income, and age) have on depression.

Previous literature has built a catalogue of features representative of cities or urban environments including density, air pollution, immigration, mobility, economy, race, and education that are all possible influences of the mental health outcomes for people in cities (Galea, Freudenberg, et al., 2005). Some of these factors are disputed as to whether they are negative or positive contributors to mental health. For example, population density increases can indicate a boom in resources, innovation, or business, but can also indicate overcrowding. These factors then positively or negatively impact the mental health of the individuals residing in these areas. Galea, Freudenberg, et al. (2005) also emphasize the presence of spatial patterns and neighborhood characteristics (race, education, and socioeconomic status) as important to understanding the differing influence that these high-level factors may have on a given area. In Stier et al. (2021), the authors demonstrate such variation, in that prevalence of depression decreases with city size. This effect was predicted and explained by a model that layered characteristics of cities' built environments, socioeconomic-networks, and the importance of social ties in depression. While this work showed first order effects of city population and social network density, it also opens up the possibility of examining which neighborhood characteristics contribute or correlate to urban depression — either positively or negatively — in excess of the first order effect.

While this previous study focused on cities holistically, the present research aims to analyze the relationship between urban environments and depression within cities by examining the sociodemographic differences and mobility levels that occur on the aggregated spatial level of neighborhoods. Using neighborhood level sociodemographic characteristics as a predictor of urban living conditions can help to enforce the correlation (Galea, Freudenberg, et al., 2005) between variables like mobility, which is found to influence mental health outcomes, and depression.

In an analysis of depression on a neighborhood scale, we may expect to see similar trends as determined in Stier et al. (2021), in which a more connected social network and the characteristics that might be associated with this densification of interactions are associated with lower levels of depression. Other urban living conditions, like mobility, may demonstrate similar associations, such as higher mobility in neighborhoods indicating lower rates of depression. However, variance of characteristics such as demographics or socioeconomic status between neighborhoods may have an effect on this prediction. Some studies correlate high density areas within urban environments to have higher levels of depression when controlling for socioeconomic and demographic factors (Echeverría et al., 2008; McKenzie, 2013; Domènech-Abella et al., 2018). When considering both findings, this leads to a prediction that sociodemographic characteristics that correlate to undesirable urban living conditions (low mobility) may indicate higher levels of depression than in urban areas with more external environmental resources.

Ultimately, I would like to determine, to what extent do sociodemographic characteristics of urban neighborhoods affect mobility and subsequently, depression? Using a sample collected from Chicago and outlying areas, I hypothesize that an analysis of this sample will display that depression will be more prevalent in neighborhoods where there are sociodemographic characteristics correlating to lower levels of mobility, such as minority race or lower socioeconomic status. Depressive episodes or symptoms affect between 4.7 and 18.5 percent of individuals aged 18 or older in the U.S. based on various reports (Substance Abuse and Mental Health Services Administration, 2020; National Center for Health Statistics, 2020; Centers for Disease Control and Prevention, 2020;), but studies show that there is a disproportionate display of depression within communities of lower socioeconomic status, as well as certain demographics such as the Hispanic and African American population (Lorant et al., 2003; Center for Disease Control and Prevention, 2020). Since sociodemographic characteristics are representations of the individuals that are located in a given area, examining neighborhood environmental characteristics such as sociodemographics and mobility can help to determine which patterns are most suggestive of depression in an individual. Understanding the factors which are most indicative of depression can aid in public policy, resource allocation, and psychological efforts to minimize environmental factors that influence depression.

In order to examine these effects, it is important to understand the background of city and urban environments on the mental health outcomes of the public. Once general influencing factors are established, examination of sociodemographic and external environment factors as key influences on mental health outcomes will be discussed. Then, the importance of examining these factors on a more granular neighborhood level will be considered. The present research will use shared social media posts in order to collect mobility, location, and depressive sentiment for a sample of users, and the sociodemographic characteristics of the neighborhood a user resides in will be collected to determine if there are discernable patterns between spatial location, key factors, and depression.

Background

City effects of mental health

There are many documented arguments about the influence that a person's environment has on mental health outcomes individually as well as on a collective scale. The increase of urbanization over time (United Nations, 2019) signals a need to examine the effects that sociodemographic characteristics in the city environment, as well as the neighborhood environment, may have on the mental health of individuals. Urban health is an important subject analyzing the health and wellbeing of individuals as they are affected by their environment, and there is a focus specifically on the effects of cities due to the increased migration of individuals to these areas. Comparison studies between urban and rural areas reveal varying results regarding whether the environment itself has an effect on mental health or physical health outcomes. Some US studies state that urban areas have higher instances of major depressive episodes (Blazer et al., 1985; Gruebner et al. 2017), while others have not found any significant overall effect of urban or rural environment as a contributing factor to mental disorders (Breslau et al., 2014). Most commonly, research has found that there are many environmental factors that can have an effect on the mental health outcomes of individuals within city or rural areas. That being said, examining the effects of predetermined characteristics may be more beneficial than binarizing areas as urban or rural and assuming the characteristics will be the same across the category (Judd et al., 2002). Psychologically, cities provide rich social networks and opportunities for growth, but it can also be a source of cognitive overload and stress on the mental state of city dwellers (Milgram, 1970). While recognizing that these benefits may occur at different rates depending on the city, factors driving urbanization like residential density, transportation, or access to parks, suggest that a consequence of these factors may be poorer mental health outcomes, but they are not necessarily the primary causal factors (Hartig & Kahn, 2016). When searching for effects that may also be contributors to mental health outcomes, Galea, Freudenberg, et al. (2005) suggests examining global, municipal, public health, and urban living conditions as they all influence the health outcomes of people residing in the area (Galea, Freudenberg et al., 2005, p.1020). Global factors include social, political, and economic trends observed over time. Migration and suburbanization have influenced the public health of those living in cities over time due to the increase of population, and the subsequent spread of resources that occurs when people move further from the crowded center of the city. Municipal factors ,such as government-provided services, are highly correlated factors to public health, but the variations between each built environment on a neighborhood level requires that public health be examined closer than the city level (Zakus & Lysack, 1998; MacQueen et al., 2001).

The most directly observed factors that contribute to the health outcomes of individuals living in cities are urban living conditions like the physical, social, and health environments. The environmental conditions of one’s housing, such as building condition or perceived condition of the surrounding area, are seen to have a significant impact on the mental health conditions of residents (Weich et al., 2002; Galea, Ahern, et al., 2005). Perceived damage to the external neighborhood, as well as individual living conditions common to the area, correlate to higher reported levels of both lifetime depression and six-month instances of depression (Galea, Ahern, et al., 2005). Additionally, factors such as air quality and mobility associated with cities had a significant effect on health outcomes, where better air quality improved childhood instances of asthma and increased mobility is associated with poorer access to health accommodations (Northridge et al., 2003). Especially in the case of mobility, there are associated factors that may influence these findings, which in turn influence mental health. Increased mobility typically means a more connected area, but if an individual has to travel in order to reach health accommodations, this negative outcome may be a product of their local environment rather than the city. This emphasizes the necessity to examine the more granular characteristics that are associated with these larger influences on health. Alongside these influencing factors, sociodemographic determinants of physical and mental health are closely correlated on an individual and neighborhood level.

What characteristics of a city are important to focus on?

Galea, Freudenberg, et al. (2005, p.1020) proposes *A conceptual framework for Urban Health* that acknowledges physical, social, economic, and political factors as an influence on the health of the individuals residing in areas of interest in research. The framework serves as a way to guide researchers towards making thoughtful decisions about possible comorbid factors, and the influences that higher level factors have on individual health outcomes. This model focuses on cities and mental health outcomes, from high level global influences to more observable and measurable influences that differ by city, as described in the previous section. For the present study, the measurable phenotypes of social and economic factors are the focus as potential predictors of depression. Sociodemographic factors refer to measures which tell us about the distribution of a population, and include age, gender, race, ethnicity, income, among others. These factors are important to consider when examining city effects on individuals because they are significant in influencing living conditions for entire communities (Tartaglia, 2013).

There is a unique ability to examine heterogenous levels of sociodemographic characteristics and their spatial patterns due to the high levels of diversity within cities. Racial and ethnic diversity are characteristic of cities themselves, but the overall diversity within a city may mask the segregation that occurs between neighborhoods in this area. Studies regarding residential segregation find that race and ethnicity are more indicative of spatial segregation than income or other factors in metropolitan areas (Acevedo-Garcia et al., 2003). On an individual level, it has been found that discrimination can have an adverse effect on psychological health, so institutionalized patterns of unfair treatment (e.g. based on race) such as forced segregation and economic disadvantages, may contribute to an overall increase of depression rates for a given area (Williams, 1999). Socioeconomic factors, especially wealth disparities prevalent in large urban areas, are also important to consider when examining the effects of a city on public health. For example, high Gini coefficients which signal large wealth disparities and lower incomes, are strongly associated with poor reported general health in metropolitan areas (Blakely et al., 2002). While there is not one measurement for each of these sociodemographic factors that is indicative of the conditions of all residents of an urban area, the different levels are interwoven with social dynamics that significantly affect the psychological health of residents.

Beyond sociodemographic factors, other variables that will be considered within this study include mobility factors. Public access to transportation within cities is often associated with people having low income or coming from areas within the city with low median income

(Glaeser et al., 2008). Additionally, urban areas characterized by low median income, multifamily households, and high-density areas have shorter trip distances and more access to local public transportation. While the association with sociodemographic factors may indicate that residents have more access to public transportation and are subsequently more mobile, there are significant associations between mobility as it directly impacts depression and urban condition (Chen & Akar, 2017). If an individual is not highly mobile and their area condition is disadvantaged, then they are seen to have higher levels of depression (Vallée et al., 2011). Valleé et al. (2011) and Northridge et al. (2003) both illustrate how it is necessary to examine the neighborhood and environment of an individual in order to understand underlying factors that contribute to mental health outcomes. The environmental condition of one's area is significantly associated with previously discussed sociodemographic characteristics, which can indicate how mobility can be applied on an aggregated scale in order to determine if there can be a measurable effect on depression using these factors.

Neighborhood level analysis

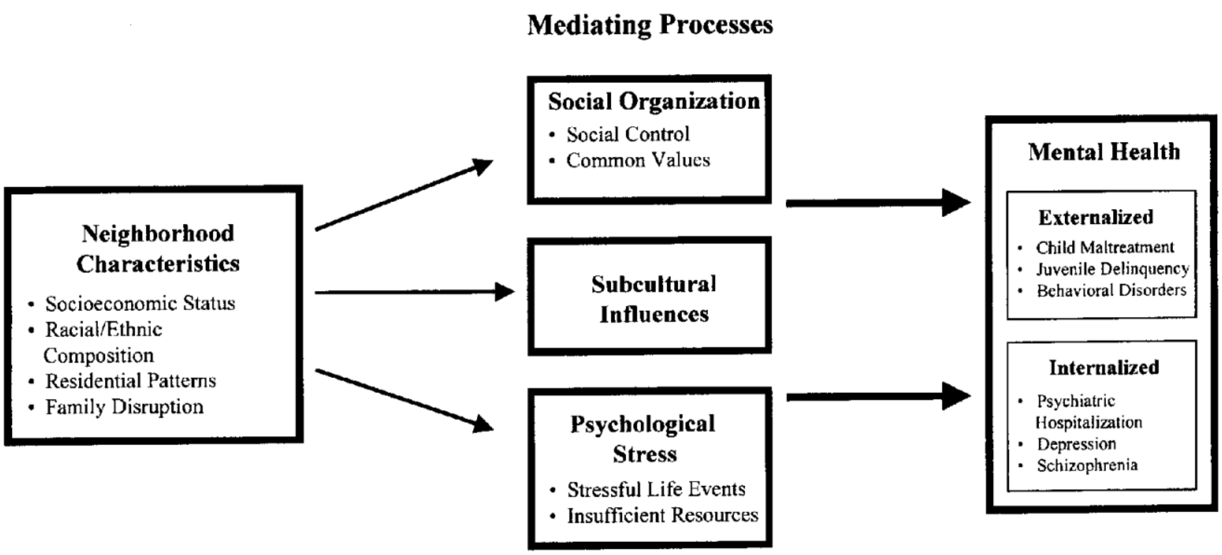
The mental health effects that cities have on depression rates have been examined in the previous work conducted by Stier et al. (2021), particularly the effects of social networks on depression. Social networks also act as urban living characteristics, since sociodemographic patterns can be detected in neighborhood defined areas that have a closer social network (Bailey et al., 2020). As discussed in the above sections, cities are often described using generalized characteristics, but this does not account for the variation that can be observed when examining neighborhoods within these cities. This variation suggests that considering the different race, ethnicity, and income factors of smaller areas may contribute to understanding patterns of mental health in cities. While it is beneficial to examine the overall impacts of urbanization, these sociodemographic factors are best analyzed in a more granular level (Zakus & Lysack, 1998; MacQueen et al., 2001). Using neighborhood level measures of these variables can help to make stronger correlations between the collected sociodemographic measures and assumptions about individual outcomes of depression. Spatial patterns of neighborhood characteristics are significant in predicting mobility and mental health outcomes within cities (Graif et al., 2016). Areas with significant perceived disadvantage adjacent to areas of high wealth are seen to have significant instances of depression (Pearson et al., 2013), which supports the need to make a closer examination.

On a neighborhood level, the classical *Structural Characteristics Model* (Figure 1, Wandersman & Nation, 1998, p.648) helps to explain the social outcomes that may arise when influenced by neighborhood level. Socioeconomic status, racial and ethnicity distribution, and family disruption are cited as contributing characteristics to psychological stress and thus reflected as negative mental health outcomes (Wandersman & Nation, 1998). Incorporating mediating factors, such as mobility, into analysis can be a way to further strengthen the correlations between mental health and sociodemographic characteristics.

Racial and ethnic diversity are characteristic of cities themselves, but the overall diversity within a city may mask the segregation that occurs between neighborhoods in this area. Studies regarding residential segregation find that race and ethnicity are more indicative of spatial segregation than income or other factors in metropolitan areas (Acevedo-Garcia et al., 2003). On an individual level, it has been found that social phenomenon such as discrimination can have an adverse effect on psychological health, so institutionalized patterns of unfair treatment such as forced segregation and economic disadvantages may contribute to aggregated levels of depression based on characteristics such as race (Williams, 1999). Neighborhood level socioeconomic definitions are strongly correlated to the mental health outcomes of individuals residing in a given area even when controlling for individual economic status (Pickett & Pearl, 2001). Education is often used as an indicator of socioeconomic status, and thus is also a

Figure 1

Structural Characteristics Model



Note. From "Urban Neighborhoods and Mental Health: Psychological contributions to understanding toxicity, resilience, and interventions" by A. Wandersman & M. Nation, 1998, *American Psychologist, 53(6)*, 648

significant factor in predicting depression (Akhtar-Danesh & Landeen, 2007).

This background introduces a motivation to examine language use in geo-located social media posts indicative of depressive symptoms and determine how neighborhood sociodemographic characteristics may correlate to the frequency of these experiences in certain areas. Considering this information in the context of Chicago, IL, which experiences a high level of disparity between area sociodemographic characteristics (Sandoval, 2011; Lee, 2009), may help to further illustrate findings that may not be detected in areas with more homogenous characteristics between areas.

Methods

Data

The data collected for this research includes mobility and depression data that were scraped from online social media sources that provide geolocation information and are interpreted on a census tract level, as well as sociodemographic variables collected from the US Census Bureau American Community Survey (2019). These two datasets were combined in order to perform statistical analyses relating neighborhood characteristics with social media instances of depression and mobility in a given area.

The Sociodemographic variables were collected via the US Census Bureau American Community Survey conducted in 2019. Variables relevant to Race, Income, Education, Age, and Gender were selected from the survey and examined on a census tract level, in order to allow for possible aggregation to larger spatial scales over time. The variables were collected for each census tract that is encompassed within our selected Metropolitan Statistical Area (MSA) using the TidyCensus API (1.1.9.9000) available within R. The spatial measure of MSA was selected because this spatial boundary defined by the occurrence of a city with at least 50,000 or more residents and reflect census defined areas of urbanization. The selected MSA is the Chicago-Naperville-Elgin, IL-IN-WI Metro MSA, henceforth referenced as the Chicago MSA. The total estimated population and the margin of error was collected for the desired area. The total summary statistic for each relevant feature category was also collected and the percent estimate was determined based off of these values. Race variables included the percent estimates for White, Black, Asian, Hawaiian and Pacific Islander, Native American, and Hispanic populations. Variables used to determine socioeconomic status were the Gini wealth inequality index, median income, and the percent estimate of individuals receiving public assistance or not receiving public assistance. The Gini index is a measure that quantifies the variability of a measure. This inequality has been calculated for the identified census tract and zip code areas in the context of wealth, in which a value approaching 1 indicates more inequality of the wealth distribution in an area. Education variables were split into percent estimates of individuals who have not completed any college or received a college degree. All age variables were categorized into bins ranging from under 18, 18 to 24, 25 to 34, 35 to 34, 45 to 54, 55 to 64, 65 to 74, and over 75. Finally, the percent estimate of male and female gender was gathered. When necessary, the census tracts were aggregated to represent Zip Code areas via a census tract to zip code comparison data file (US Census Bureau, 2022). This enabled examination of the total Estimate and Percent estimate of our variables on varying scales (Table 1).

The Mobility variables and Depression Rate were collected through Twitter's Academic Research API (Twitter, 2022). Tweets with geolocation tags were used for this research. Users with the majority of tweets occurring within the Chicago MSA were selected for this study. Users with tweets within June, July, and August of 2019 were included in this analysis. In order to calculate mobility variables, the radius of gyration was determined for each user included in the dataset. Radius of gyration is the characteristic distance between tweets within the same area (Gonzalez et al., 2008). This measure was aggregated to the appropriate spatial level for analysis. The dataset included over 2.2 million tweets collected for the Chicago MSA and 10,624 users were included in the analysis, see Figure 2 for the population distribution of the sample.

Table 1

Distribution of Sociodemographic, Mobility, and Depression variables for Chicago MSA Zip Codes

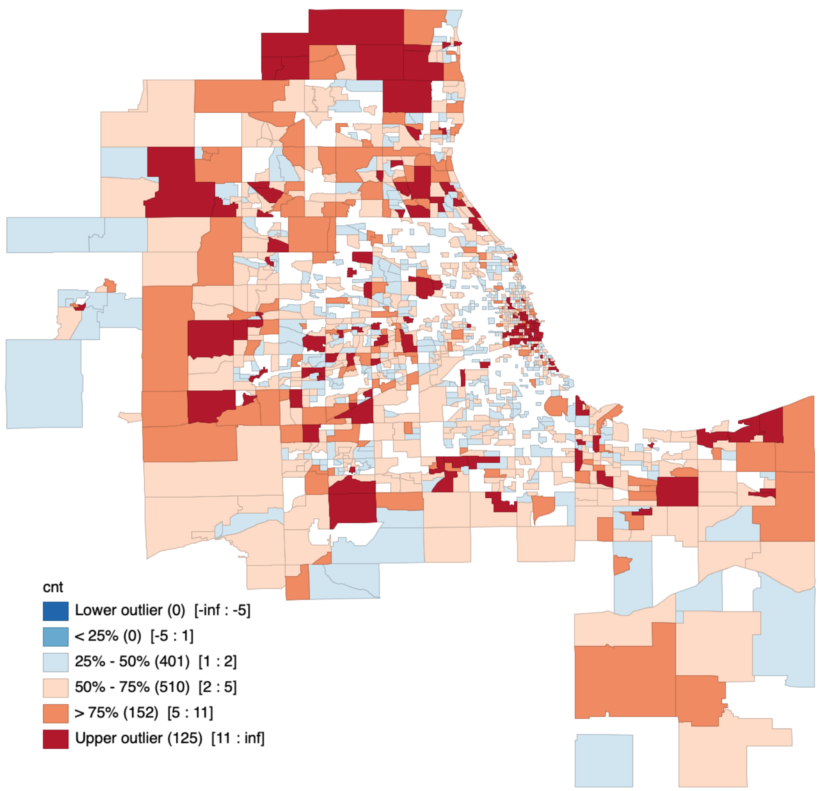
Variables	Mean	SD	Min	Max
		Mobility		
Population sample	35.189	99.037	1.000	741.000
Radius of gyration	37642.743	18193.600	5958.776	98416.709
Depression rate	.091	.139	.007	1.000
		Sociodemographic Variables		
Race: White †	.588	.267	.010	.957
Race: Black †	.149	.242	0.000	.954
Race: Native †	.001	.004	0.000	.066
Race: Asian †	.066	.079	0.000	.432
Race: HIPI †	.000	.001	0.000	.005
Race: Hispanic †	.176	.170	0.000	.896
Education: No College †	.098	.078	0.000	.408
Education: College †	.463	.198	.139	.978
Age: Under 18 †	.224	.054	0.000	.341
Age: 18 to 24 †	.090	.046	0.000	.629
Age: 25 to 34 †	.139	.067	0.000	.470
Age: 35 to 44 †	.130	.027	0.000	.329
Age: 45 to 54 †	.135	.024	.002	.195
Age: 55 to 64 †	.131	.028	.017	.231
Age: 65 to 74 †	.088	.038	.005	.573
Age: Over 75 †	.063	.032	0.000	.228
Gender: Female †	.507	.038	.093	.578
Gender: Male †	.493	.038	.422	.907
Income: Public Assistance†	.0195	.015	0.000	.099
Income: No Public Assistance†	0.981	.0154	.901	1.000
Income: Gini Inequality	.430	.057	.277	.636
Income: Median Income	82352.539	32723.820	22158.000	248243.000

Note. † indicates that the values are being reported as percent estimates of the total population. n = 296.

The user tweets were also collected for text processing, allowing for the calculation of depression rate for each user as specified by Stier et al. (2021). The text was prepared using previously implemented natural language processing exploring prevalence of depression (Stier et al., 2021; Yazdavar et al., 2017). In order to calculate instances of depression within tweets, seed words relating to depression and depressive symptoms collected from the PHQ-9 were used, and a Latent Dirichlet Allocation (LDA) model was built. This model allowed for the observation of topics referenced in user tweets are related to depression. Depression rate of an area was calculated as the proportion of collected users with a greater than zero instance for any of the topics that contain words selected using the PHQ-9 criteria (Stier et al., 2021).

Figure 2

Population Distribution of Twitter Users



Note. The legend provides the number of areas within the quantile, and the range of values that the quantile encompasses.

The mobility measures calculated from this sample include radius of gyration. Using the geolocated tweets that occur within the Chicago MSA, the trajectory of tweets collected within the three-month time period were determined and evaluated using Equation 1 (González et al., 2008, suppl. pp. 5):

$$r_g^a(t) = \sqrt{\frac{1}{n_c^a(t)}\sum_{i=1}^{n_c^a}(r_i^{a\rightarrow} - r_{cm}^{a\rightarrow})^2} \tag{1}$$

In which $r_i^{a\rightarrow}$ represents the positions recorded for a given user and $r_{cm}^{a\rightarrow} = 1/\frac{1}{n_c^a(t)}\sum_{i=1}^{n_c^a} r_i^{a\rightarrow}$, is the centroid found from the geolocated tweets of a user.

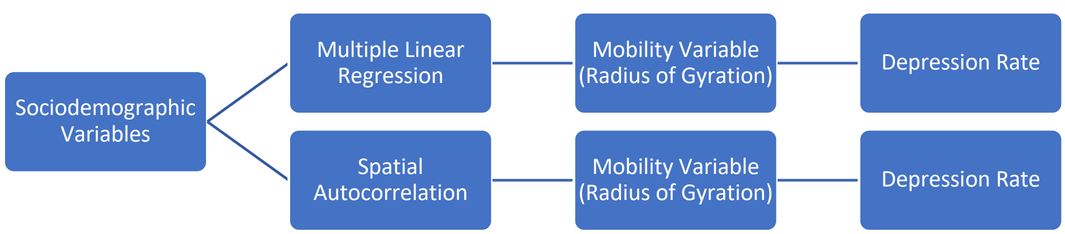
Tools of analysis

After collecting this data, a multiple regression analysis was performed using all the selected sociodemographic variables to determine possible correlations between neighborhood characteristics and radius of gyration (the mobility variable), as well as depression rates. The correlation between sociodemographic variables and mobility was included in the analysis due to justification from the structural characteristics model (Wandersman & Nation, 1998, p.648), which dictates that social organization or psychological stress can act as a mediating process between neighborhood sociodemographic characteristics and mental health outcomes. Determining a significance between this relationship will strengthen the claim that sociodemographic characteristics influence depression rates. The analysis for this can be found at the following GitHub repository: https://github.com/kthomas14/Neighborhoods_and_depression. Global and Local spatial autocorrelation was also performed to determine if there are significant spatial patterns relating sociodemographic, mobility, and mental health outcomes. Figure 3 illustrates the flow of analysis which were used to inform the following results.

Linear regression analysis is an important methodological tool that helps to identify trends in the data as they relate to the variables of interest. By utilizing multiple linear regression, the effects of the other variables of interest can be controlled and examined. Adding lasso regression techniques to this model can also help to determine which factors have a large effect on the variable of interest, as well as which factors might add unnecessary noise to the model. Determining the accuracy of these models can help to illustrate how well patterns of depression can be detected using the model.

Figure 3

Workflow for Analysis of Sociodemographic, Mobility, and Depression Variables



Note. For the analysis, multiple linear regression and spatial autocorrelation will be implemented using sociodemographic variables as predictors of mobility. Following this analysis, mobility will join the sociodemographic predictors to build a model for depression rate.

Pairing the variable selection and accuracy alongside spatial predictions of the factors of interest can also be a very helpful way to assess the strength of patterns, as well as the distribution of depression within an area. Global spatial autocorrelation will be initially used in order to assess if there are significant spatial patterns for the given variable. The method of analysis for this assessment will be Global Moran's I. This measure will yield a pseudo-p value computed using random permutations that suggest whether the spatial units within the sampled area follow a pattern more observable than random. Ensuring that there is significant global spatial autocorrelation affirms the inclusion of the tested factors in a local spatial autocorrelation. A local examination of autocorrelation was performed using a Multivariate Local Geary measure. This spatial analysis displays significant patterns of attributes within local clustering. By adding all relevant variables into this measure, we are able to assess whether the selected sociodemographic variables and the dependent variables of interest in one area, have similar levels or distributions as the neighbors included in the local cluster. Bivariate examination of the dependent variables also ensures that there are patterns within the cluster for the independent variables that align with the local cluster's y variable. This approach will be particularly useful in explaining the relationship between mobility and depression, at the end of the proposed workflow in Figure 3.

These methods of analysis are complementary to one another due to the concentration on urban characteristics and spatial patterns in the present research. While the multiple regression analysis will help to develop a model that will predict the distribution of the sociodemographic variables as they may affect mobility or depression, ultimately these patterns must also have a spatial reference. Areas that share sociodemographic values which indicate similar mobility measures or depression rates might be located in completely different areas. Examining spatial patterns help to highlight possible outliers in the data, as well as ensure that the model is also showing how these patterns are significant to urban areas.

Results

Multiple linear regression

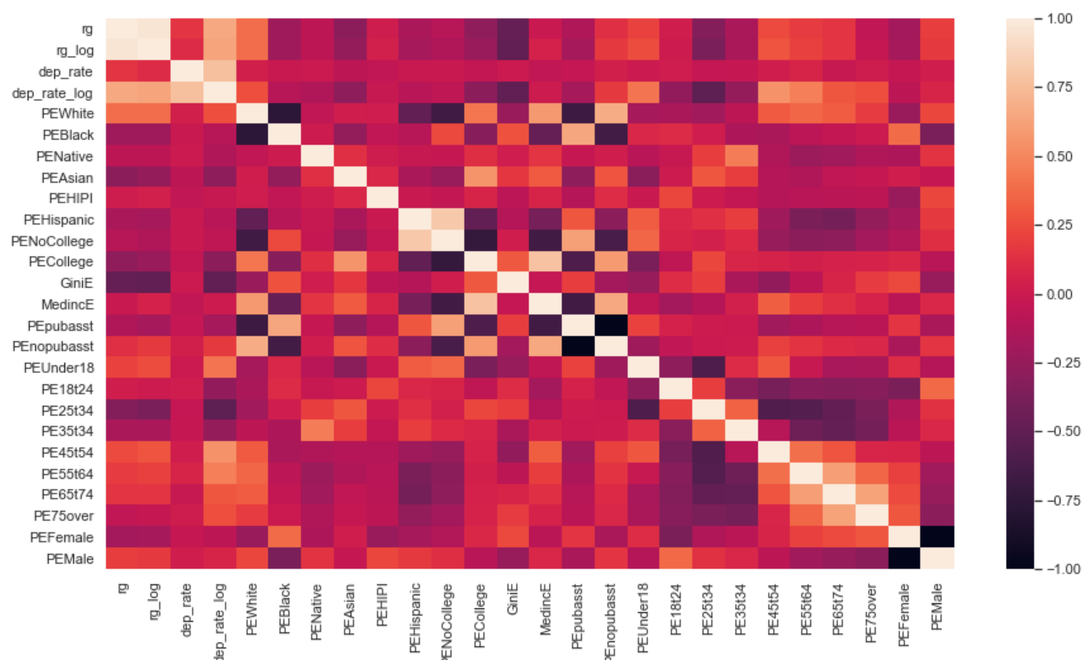
For analysis, I used multiple linear regression with Lasso regularization in a 20-fold cross-validation. The dependent variables were logarithmically transformed to ease interpretability. A linear model was ultimately chosen after testing many different supervised learning methods that handle regression. Polynomial regression was considered to determine if collinearity will have a significant effect on the model, but the resulting testing score measured by the Coefficient of Determination indicated that a simpler model will be superior in predicting the dependent variables.

After examining the distribution of the census data and making comparisons with the mobility and depression variables through correlation matrices (see Figure 5), all variables were initially included in the model in order to proceed with variable selection and model tuning. The variables were standardized using standard scaling to make comparisons between variables that use specific estimates as well as variables measured using percent estimates. After, the variable set was normalized using Lasso regularization. This method was selected because I was interested in dropping variables from the dataset that increase noise and prove to be redundant to the interpretation. Ultimately, this regularization eliminated varying numbers of sociodemographic variables depending on whether the dependent variable was depression rate or radius of gyration. This may be due to the proportional relationship between features within the respective categories, which are largely percent estimate variables. After examining residual plots created for radius of gyration and depression rate derived from the created Lasso regression models, it became clear that these variables should be transformed to better fit a linear model. Both plots displayed heteroscedasticity that became resolved by performing log transformations on the dependent variables.

The finalized regression models were created using the same standardization and lasso methods, but all models included the logarithmically transformed radius of gyration and depression rate variables. Once the appropriate variables were eliminated from this process, a new multiple regression model was fit using the remaining variables and the significance of each variable was assessed in Table 2. The resulting models were evaluated using the Coefficient of

Figure 4

Correlation Matrix



Note. Both Radius of Gyration and Depression rate are represented in the correlation matrix as well as the logarithmically transformed variables.

Determination, or R^2 scoring. This was selected because it is a standardized approach to interpreting the seen error that cannot be explained by the selected independent variables. The entire sample was included fitting in the regression model in order to mitigate any skew that could result from not including all locations in the model. This does cause a slight overfit of the final model, but measures such as cross validation were used to ensure that the model would be sufficiently generalizable to all urban environments within the sample (see Figure 5 for workflow of the regression analysis). The significance and results of the models are shown in Table 2 and will be discussed next.

Figure 5

Workflow of Multiple Regression Analysis

The best created model to predict radius of gyration had an R^2 score of 0.505, indicating that the accepted neighborhood characteristics from zip code areas explain 51% of the variation perceived in the data. After the fit of the lasso regularized regression plots were examined, a multiple linear regression model was created using the variables that were not dropped in the

Table 2

Multiple Linear Regression Significance Table for Sociodemographic Variables Predicting Radius of Gyration and Sociodemographic and Mobility variables Predicting Depression Rate

Variables	Dependent Variable: Radius of Gyration		Dependent Variable: Depression rate		
	Lasso (α = .029)	Multiple Regression	Lasso (α = .079)	Multiple Regression	Multiple Regression (No Radius of gyration)
Race: White	.000	5.394 ***	.000	—	—
Race: Black	.414	—	.000	—	—
Race: Native	.000	1.917 *	.000	—	—
Race: Asian	.700	—	.000	—	—
Race: HIPI	.000	—	.000	—	—
Race: Hispanic	-.524	-4.718 ***	.000	—	—
Education: No College	.000	—	.000	—	—
Education: College	-2.190 *	-5.408 ***	.000	—	—
Age: Under 18	.000	3.381 ***	.000	1.322	1.823
Age: 18 to 24	.000	1.481	.000	-1.150	-.019
Age: 25 to 34	.000	0.151	.000	—	—
Age: 35 to 44	.000	-2.859**	.000	—	—
Age: 45 to 54	.000	—	.000	.473	.562
Age: 55 to 64	.000	—	.000	.855	2.115 *
Age: 65 to 74	.000	—	.000	—	—
Age: Over 75	.000	-1.417	.000	1.264	.803
Gender: Female	.000	—	.000	—	—
Gender: Male	.000	—	.000	—	—
Income: Public Assistance	.000	-1.301	.000	—	—
Income: No Public Assistance	.000	—	.000	—	—
Income: Gini Inequality	-5.380	-5.045 ***	-.705	-1.252	-3.308 ***
Income: Median Income	.000	—	.000	—	—
Radius of Gyration	—	—	2.99 **	4.049	—
R^2	.505	.519	.531	.548	.447

**p < .05. **p < .01. ***p < .001*

lasso regression. The best created lasso model for depression rate yielded an R^2 score of 0.530, indicating that neighborhood characteristics accounted for 53 percent of the variation found in the model.

The multiple regression model fit with the variables remaining after lasso regularization resulted in an R^2 score of 0.519 when the predictor was the logarithmic transformation of radius of gyration. This indicates that the sociodemographic characteristics of a given area explain 52% of the variation perceived in the model. The variables included in this model were Percent estimates for White, Hispanic, Native American, received a college degree, receiving public assistance, the age ranges Under 18, 35 to 44, 18 to 24 25 to 34, 34 to 44, over 75 and the Gini wealth inequality index. At $\alpha > 0.001$, significant variables include percent estimates for White, Hispanic, received a college degree, under 18 and the Gini index. At $\alpha > 0.01$ the age range 34 to 44 is significant. This suggests that sociodemographic factors from the general categories of race, income, education, and age are all significant to interpreting patterns of mobility in urban areas.

The final multiple regression model created using all sociodemographic variables and the logarithmically transformed radius of gyration variable to predict the logarithmic transformation of depression rates in a neighborhood yielded an R^2 score of 0.548. The regression model was built using the variables: radius of gyration, the Gini inequality index, and the percent estimates for the age ranges under 18, 18 to 24, 45 to 54, 55 to 64, and over 75. At $\alpha > 0.001$, radius of gyration is the only variable considered significant. No other variables were considered significant. In order to determine if radius of gyration is masking the sociodemographic variables included in the model, an additional multiple regression model was created with the same variables accepted in the lasso regression model, but excluding radius of gyration. In this model, an R^2 score of 0.447 was recorded, and Gini index and the under 18 age range were significant at $\alpha > 0.01$. This finding suggests that radius of gyration is a powerful predictor that masks the effects of other variables when predicting patterns of depression.

Spatial Analysis

Spatial analysis was conducted for all variables to avoid making deductions about the correlation between spatially insignificant variables. The target area, Chicago and outlying areas within the dedicated MSA code, were examined at the census tract level. For each variable, global spatial autocorrelation was performed using distance weights (see Table 3). While the data was spatially aggregated on the census tract level, distance bands were created based on the average size of city zip codes represented in the data set to ensure an even distribution of the data within each cluster. This weighting technique allows for appropriately sized neighborhoods in which to compare the dependent variables. Each census tract became clustered with any tracts within a radius of 23.463810km. The Chicago MSA is roughly 28,120 km^2 and on average, 502 census tracts were included in a cluster. An initial analysis using Global Moran's I shows a significant spatial pattern associated with all variables except for depression rate.

While all variables were considered significant, local spatial autocorrelation was built using only variables included in the Lasso Regression model for the respective dependent variable (see Table 2 and previous *Results* section for included variables). Local spatial autocorrelation modeling determines if there are any significant clusters within the target city indicating patterns with the adjacent areas. It must also be noted that the dependent variables were not logarithmically transformed for this analysis. A multivariate local Geary model was created using relevant sociodemographic variables and radius of gyration to visualize if there is

Table 3

Global Moran's I Spatial Autocorrelation for all Variables

Variable	Global Moran's I	Z-stat (9,999 permutations)	Pseudo p-value
Race: White	.195	50.246	.000***
Race: Black	.208	53.246	.000***
Race: Native	.021	5.692	.002**
Race: Asian	.128	32.953	.000***
Race: Hipi	.003	1.099	.101
Race: Hispanic	.020	5.288	.000***
Education: No College	.032	8.471	.001***
Education: College	.203	51.575	.000***
Age: Under 18	.090	23.071	.000***
Age: 18 To 24	.029	7.638	.000***
Age: 25 To 34	.232	59.868	.000***
Age: 35 To 44	.040	10.177	.000***
Age: 45 To 54	.117	29.430	.000***
Age: 55 To 64	.107	27.247	.000***
Age: 65 To 74	.037	9.541	.000***
Age: Over 75	.020	5.357	.000***
Gender: Female	.017	4.582	.000***
Gender: Male	.024	6.343	.000***
Income: Public Assistance	.070	18.137	.000***
Income: No Public Assistance	.005	1.578	.044
Income: Gini Inequality	.180	45.369	.000***
Income: Median Income	.168	42.866	.000***
Radius of gyration	.678	172.675	.000***
Depression rate	-0.003	-0.493	0.307

Distance weighting: 14.58mi/23.46km
Neighbor min/man/avg: 1/502/238
Non-zero: 20.08%
**p* < .05. ***p* < .01. ****p* < .001

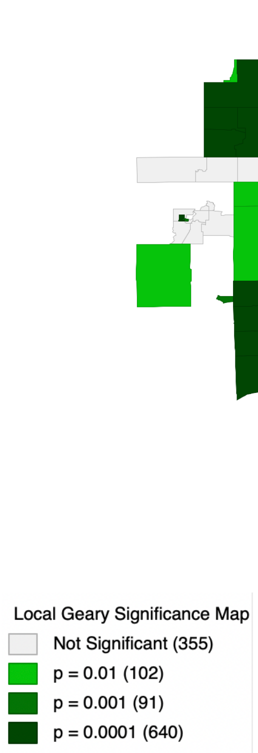
any significant clustering affirming that all of these neighborhood effects have a significant relationship (see Figure 6*)*. The resulting visualization was iterated with 9,999 permutations to detect significance of the spatial relationship between all the variables at α significance from 0.01 to 0.0001. The majority of tracts had significant clustering where $p \leq 0.0001$. Significant spatial areas were clustered in the outer Chicago metropolitan area, as well as in the very center of the city. The overwhelming majority of clusters had positive relationships (see Figure 7), illustrating that positive trends in the variables included in the Local Geary autocorrelation for a given tract indicate positive trends for the neighboring clusters as well.

Next, a multivariate local Geary model was created with all the variables included in the multivariate regression analysis for depression rate. Local Geary Cluster mapping with a 0.0001 significance filter and 9,999 permutations on an analysis of the target areas reveals that the majority of census tracts have significant clustering at $\alpha > 0.0001$, following a similar spatial pattern as the previous model (Figure 8). This autocorrelation also shows significant positive clustering within the city, with very little significant negative clustering (Figure 9). This suggests that the created clusters for each significant area shares features within similar attribute space. These comparisons were strong within the city center, and in the outermost areas of the city boundaries, similar to the analysis not including depression rate.

A bivariate local LISA analysis comparing the spatial relationship between the two variables of interest, radius of gyration and depression rate was conducted to see if mobility is an appropriate variable to include in the spatial analysis. A global spatial autocorrelation using a bivariate analysis of Moran's I yields a value of 0.017, with a significance of 0.026 and z-value of 1.99 when conducted with 9,999 permutations. A local spatial autocorrelation analysis using bivariate local Moran's I for radius of gyration as a predictor of depression rate did not yield many significant results as shown in Figure 10. Bivariate analysis detects a relationship between x and the spatial lag of y, so significance of a cluster suggests that the present value of X has a strong spatial pattern with the Y values within the neighboring areas. The resulting bivariate analysis shows that radius of gyration within the clusters is not indicative of any spatial patterns of the depression rate of neighboring areas. A bivariate analysis of the reverse relationship was also conducted. The resulting spatial autocorrelation contained many significant census tracts at $\alpha > 0.0001$ (Figure 11). Clustering in the city center indicates that areas with a low depression rate have neighboring values that have a low average radius of gyration. Clustering patterns on the outer edge of the areas of interest suggest that areas with a low depression rate have neighbors with a higher average radius of gyration.

Figure 6

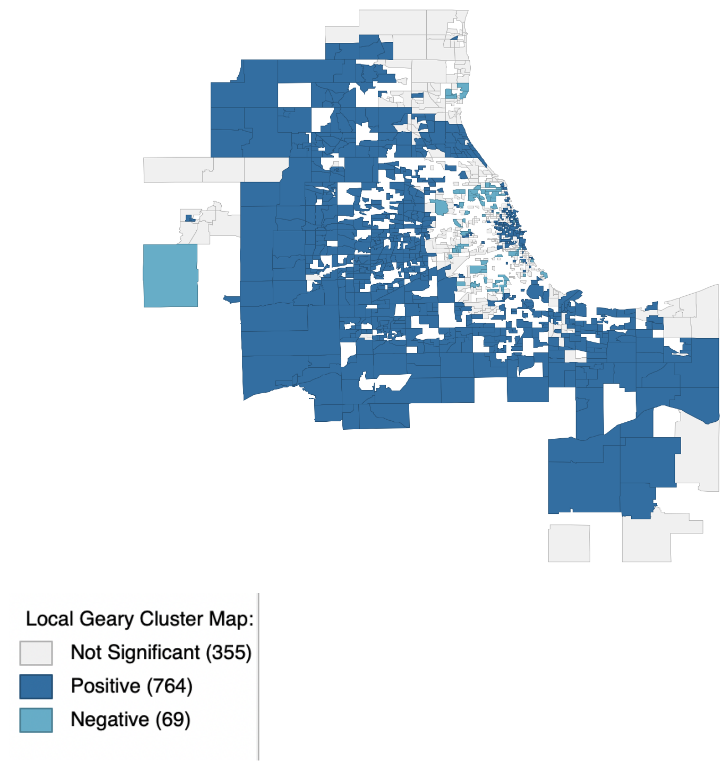
Local Geary Spatial Autocorrelation analysis of patterns between sociodemographic variables and radius of gyration, Significance Map



Note. The majority of census tracts have significance at p = 0.0001, indicating that neighboring tracts within the cluster share similar attribute space with the core.

Figure 7

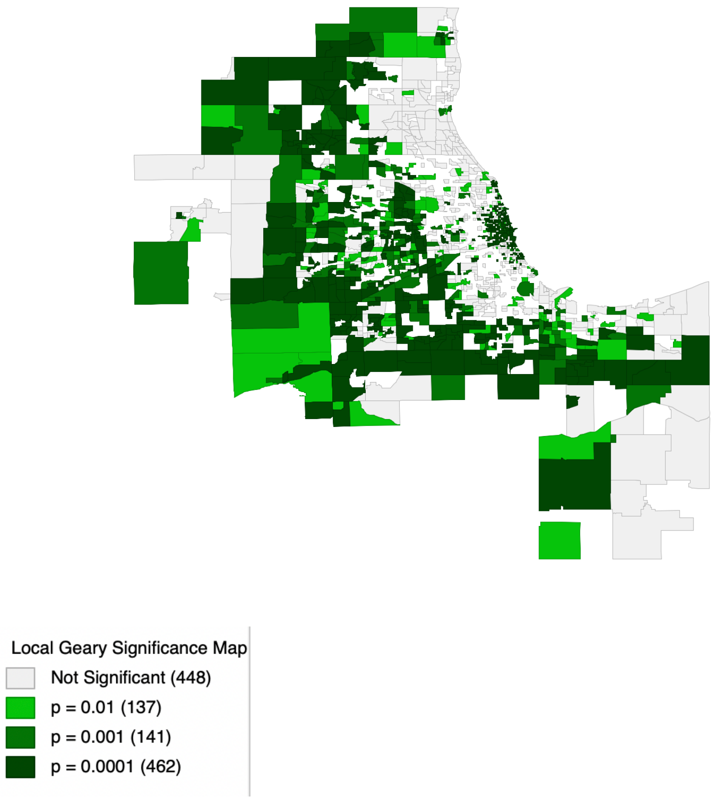
Local Geary Spatial Autocorrelation analysis of patterns between sociodemographic variables and radius of gyration, Clustering map



Note. This figure illustrates the directional relationship between the selected census tract and its neighbors within the cluster. The large majority of census tracts are surrounded by other tracts that have similar features to the core.

Figure 8

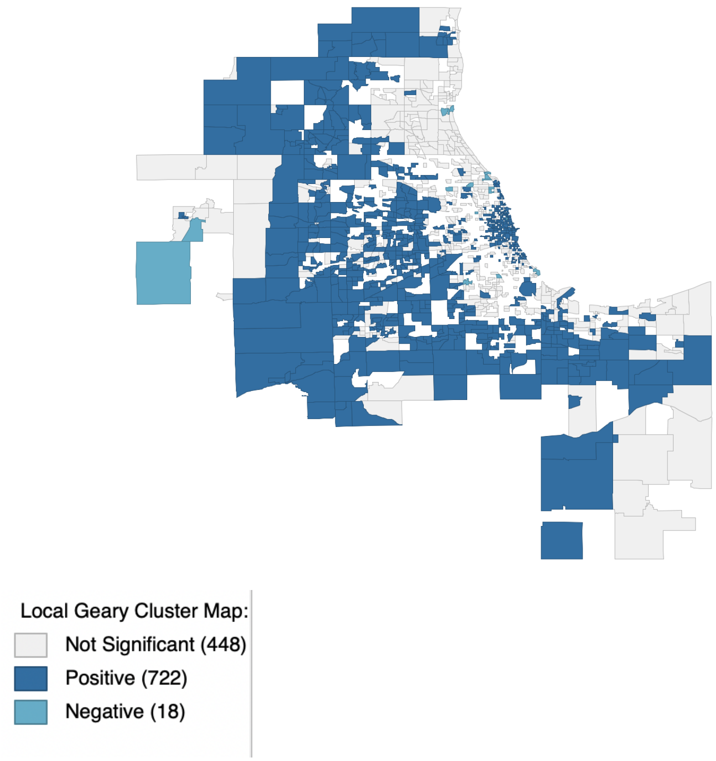
Local Geary Spatial Autocorrelation analysis of sociodemographic variables, radius of gyration, and depression rate, Significance map



Note. There are clusters within the center of the city as well as the outer edges of the dedicated MSA area that share significant patterns with the neighboring census tracts at $p > 0.01$.

Figure 9

Local Geary Spatial Autocorrelation analysis of patterns between sociodemographic variables, radius of gyration, and depression, Clustering map



Note. Of all significant census tracts, the neighboring tracts within the clusters share a similar attribute space to the core when including sociodemographic, mobility, and depression rate variables.

Figure 10

Bivariate Local Moran's I Spatial Autocorrelation, Radius of Gyration as a predictor of Depression Rate

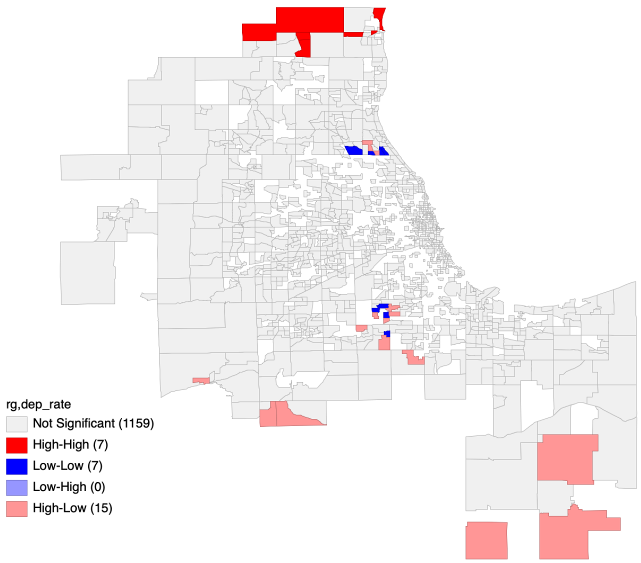
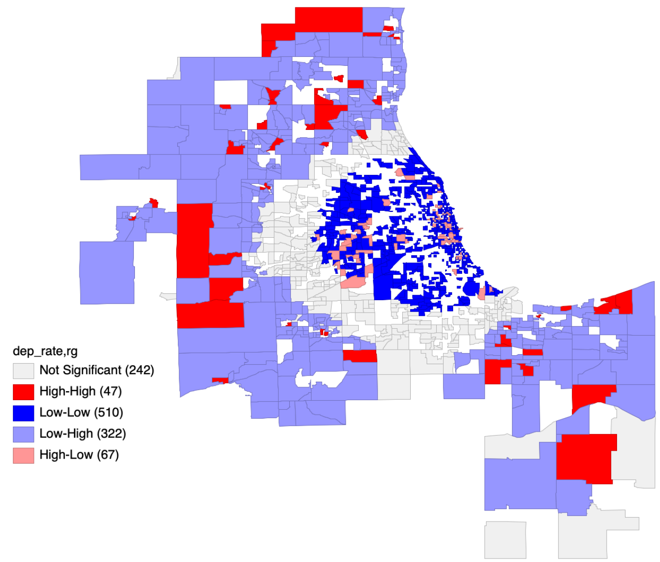


Figure 11

Bivariate Local Moran's I Spatial Autocorrelation, Depression Rate as a predictor of Radius of Gyration



Discussion

Sociodemographic characteristics and mobility

The lasso regression analysis of neighborhood sociodemographic factors and radius of gyration revealed that the percent estimate for White, Hispanic, received a college degree, and the age ranges Under 18, and 35 to 44 are significant to the model, as well as the Gini inequality coefficient (Table 2) . The removal of some sociodemographic factors by lasso regularization may be due to other percent estimate variables present within the category also being representative of the distribution. In this case, having several proportional measures such as percent estimates may add noise to the model and cause it to become removed. The purpose of creating this model aligns with the structural characteristics model, in which sociodemographic variables, or neighborhood factors, influence mobility, a mediating factor. Mediating factors more directly affect mental health outcomes. The final model indicates that the Gini index, college education, percent estimate Hispanic, and the age ranges 34 to 44 and over 75, are all negatively correlated with radius of gyration. An increase in these variables correlate with a decrease in radius of gyration in urban neighborhoods. A decrease in the Gini wealth index indicates homogeneity within an area. This factor being associated with lower radius of gyration may be explained by areas with low inequality, in which areas with higher overall associated socioeconomic environments have positive associations with commuting (Feuillet et al., 2015). Evidence has also been found that areas with lower overall socioeconomic environments use public transportation and commute more often than other areas (Glaeser, et al., 2006), which can explain homogeneity of the wealth dispersion on the other end of the spectrum in which most people have a low median income. Further examination of the effects of the Gini index on mobility can be examined using other mobility measures. For example, public transport or local commute is more often associated with lower income communities in urban areas, but higher income areas may have a higher total distance traveled if commuting from suburbs, or accounting for movement outside of the local area. Incorporating measures like "total distance traveled" in this analysis may help support the pattern of lower Gini index with an increase in radius of gyration.

The negative trend in the percent estimate of individuals with a college degree and radius of gyration is contradictory to previous findings that higher education levels would indicate higher mobility (Machin et al., 2012). But, supporting research has also emerged which find that urban areas in Spain indicate that individuals with occupations requiring higher education are more often residing in the center of large cities, reducing average mobility (Mohino et al., 2016). These findings suggest that for the present data and methods, patterns can be examined and reported on a local level to determine if education and mobility trend in a similar manner rather than running a large scale analysis. Further research about the impact of remote work in the present era could also influence the relationship between education and mobility. The percent distribution of Hispanic individuals within an aera and a negative correlation with mobility can be explained by traditional associations with other sociodemographic variables. Neighborhoods with lower median income and predominantly minority populations, such as the Hispanic population, are traditionally less likely to travel outside of one's neighborhood (Wang et al., 2018; Bora et al., 2014). Age trends for older ranges indicate that mobility is often more challenging as one grows older, as well as retirement generally decreasing the need for a daily commute.

The positive association of radius of gyration and the percent estimate White in a neighborhood can be correlated with historical patterns of predominantly white neighborhoods having higher education rates and higher median income, which can explain possible influences like migration to suburban areas and longer commutes or personal cars that enable a higher level of mobility from White populations, especially from suburban areas (South & Crowder, 1997).

These observed patterns from the fitted model align with the findings from Valée et al. (2011), in that negative correlations with the significant sociodemographic characteristics are indicative of the urban living condition of less mobility. While the centers of cities are often associated with more mobility due to high access to transportation (Glaeser, et al., 2006), it is also true that wealthier suburban areas in which individuals may travel to work might have very high levels of mobility. This could help address why characteristics associated with outer suburban areas are also associated with a high radius of gyration. Incorporating network analysis to determine social and occupational mobility may help further address why these mobility factors do not have strongly discernable trends.

Spatial patterns of these variables were largely significant in the center and edge of the target city. This may be explained by homogeneity within suburban and inner-city characteristics as detected when identifying global spatial autocorrelation trends in the collected sociodemographic areas. The significance of all of these variables on a global scale explains that there are spatial trends in the Chicago MSA area and that neighboring areas have like-characteristics. The results of the local multivariate analysis indicated almost exclusively positive clustering within significant areas, which illustrates that surrounding areas share similar patterns in the variables of interest. Many of these clusters remained significant at the $p=0.0001$ level, signifying strong correlations within attribute space. The results of this spatial autocorrelation suggest that neighborhood clusters have homogenous sociodemographic and mobility traits. It has been found that, in areas with poor socioeconomic status or poor natural environment, it is less likely for individuals residing in areas with surrounding sociodemographic similarities to travel outside of their respective neighborhoods (Sampson & Levy, 2022). The significant spatial clustering in this model along with the multiple regression analysis reveal that while neighborhoods have similar variables within their attribute space, there must be a more granular analysis to focus on what variables influence the multivariate autocorrelation and how these patterns behave. The multiple regression analysis of radius of gyration indicate that the patterns are not overwhelmingly present, so an extensive examination of spatial relationships may help explain how sociodemographic characteristics influence one another and act as predictors of mobility.

Sociodemographic characteristics, mobility, and depression

The multiple regression model including sociodemographic and mobility variables as predictors of depression yielded a 0.548 R^2 score, which provides an explanation for the 55% of variance observed within the model. This lasso regression eliminated all variables except radius of gyration, the Gini inequality index, and the percent estimates for the age ranges under 18, 18 to 24, 45 to 54, 55 to 64, and over 75. None of the variables except radius of gyration were significant to this model. According to the multiple regression model adjusted to unmask sociodemographic affects obscured by radius of gyration, the Gini inequality index is the only significant predictor of depression rate in urban neighborhoods. Similar explanations for homogeneity in lower Gini values indicating lower radius of gyration can be applied to depression rates in urban neighborhoods. Depression can be correlated to the neighborhood characteristics of sociodemographic variables, especially economic factors (Akhtar-Danesh & Landeen, 2007; Eaton et al., 2001; Gutiérrez et al., 2002; Lorant et al., 2003). The literature on the subject is very consistent, but decreasing Gini inequality correlating with higher rates of depression in this analysis calls for a more localized approach to determine exactly what features in income-homogenous areas may also contribute to these depression rates. For example, homogenous areas of high wealth may have different sociodemographic influences than areas with very low wealth.

Additionally, an increase in the logarithmically transformed variable for radius of gyration indicates an increase in depression rate for an area. This may be explained by increased homogeneity of sociodemographic characteristics contributing to depression, specifically as seen in Hispanics and Asians, although race was ultimately removed from my analysis of depression rate by the regression model (Pescosolido et al., 2020). The finding that radius of gyration heavily influences the significance of other predictors in the model illustrates the role of these urban living conditions as directly influencing the mental health outcomes of individuals, which is iterated in the structural characteristics model.

The spatial analysis of all variables, including depression, indicates similarly strong results as the spatial analysis focusing on mobility. Emphasizing the correlation that Wandersman & Nation (1998) describe in their structural characteristics model, it would be expected to see similar patterns within the model focusing on mediating processes (mobility) and health outcomes (depression rates). This model contains largely positive correlations between depression and all other variables within attribute space and neighboring space. This means that census tracts within the defined cluster for a single census tract have similar attributes. The high level of correlation between all of the variables in attribute space also enforces defining urban neighborhoods as the size of the mean zip codes in an area, since we would like to define neighborhoods as areas sharing similar characteristics. The significance of sociodemographic variables in the regression model for depression suggests that while there are strong similarities within areas in the Chicago MSA, further analysis to determine which sociodemographic variables might be more powerful predictors when paired with others could help to reveal a more salient relationship linking sociodemographic factors, mobility factors, and depression rate.

A bivariate spatial analysis of radius of gyration and depression rate was also included in the analysis of this data. This revealed that radius of gyration, while a stronger overall predictor than sociodemographic variables in the multiple regression analysis, was not very significant in predicting spatial trends of depression rate in neighboring clusters. Local spatial autocorrelation conducted to determine if there is a reverse relationship present was very significant in predicting the radius of gyration of other census tracts. This relationship can be explained by depression as an influencer of mobility. Often these external symptoms or habits, like mobility, can have both a influencing effect and occur as a result of depression. The bivariate analysis suggests that depression might be more significant in influencing mobility, rather than a lack of movement resulting in depression, although there is evidence of both of these effects (Hirvensalo et al., 2007; Mitsue & Yamamoto, 2019).

Conclusion

The result of this research reveals that there are some spatial patterns that are consistent within sociodemographic variables, mobility, and depression rates in the Chicago MSA area, but it seems that relational patterns are not able to be determined using linear regression approaches. The multiple linear regression models highlight the necessity to make locally informed assessments about the influences of environmental characteristics on depression, while making suggestions about the directional relationships between the selected factors and mobility or depression. This research serves as a preliminary examination of the factors that can contribute to the analysis of depression in urban areas. The ability to use computational approaches and national data to make an almost real-time analysis of the present condition of depression in an area and determine correlations between possible mitigating factors can influence urban planning and policy change in order to provide certain areas with the required support to lower levels of depression.

Possible limitations to this analysis include the inclusion of one MSA in the analysis for one time period. Collection speed and data filtering were a substantial contributor to the analysis process. Including cities with a higher variety of global characteristics, such as climate, overall income differences, and population density could help to find discernable patterns in the sociodemographic consistencies within urban neighborhoods. Additionally, a time series analysis may also reveal different patterns of depression that align with seasonality, or as influenced by the COVID-19 pandemic. Mobility variables can also become more ecologically valid using distance calculations created with street map paths, rather than a general radius of mobility. By utilizing a more realistic distance of mobility, there may be increased distances for individuals residing in city centers. While the radius of their mobility may be small, traversing actual paths traveled may reveal more similarities in length to those commuting from suburban areas. Incorporating mediating factors other than mobility that can also be collected computationally is a way in which the complex relationship between sociodemographic and mental health outcomes can be further examined. Building on this already created dataset could provide more robust and supported evidence for the individual effects depression may have in certain areas.

These findings suggest the beginning of an analysis of the complex urban conditions that may influence a presence, or lack thereof, of depression in neighborhoods. Local examinations of a neighborhood's sociodemographic characteristics may continue to provide insight to the types of individuals that may be more susceptible to depression as a result of urban living conditions, as evidenced by existing research about general sociodemographic variables affecting depression. Spatial patterns will help to support this claim and can illustrate the influences sociodemographic factors have on mediating factors, like mobility. This research can eventually aid in describing the complex and interactive relationship between urban environments and depression.

References

Acevedo-Garcia, D., Lochner, K. A., Osypuk, T. L., & Subramanian, S. V. (2003). Future directions in residential segregation and health research: a multilevel approach. *American journal of public health*, *93*(2), 215-221

Akhtar-Danesh, N., & Landeen, J. (2007). Relation between depression and sociodemographic factors. *International journal of mental health systems*, *1*(1), 1-9.

Bailey, M., Farrell, P., Kuchler, T., & Stroebel, J. (2020). Social connectedness in urban areas. *Journal of Urban Economics*, *118*, 103264.

Blakely, T. A., Lochner, K., & Kawachi, I. (2002). Metropolitan area income inequality and self-rated health—a multi-level study. *Social science & medicine*, *54*(1), 65-77

Blazer, D., George, L. K., Landerman, R., Pennybacker, M., Melville, M. L., Woodbury, M., ... & Jordan, K. (1985). Psychiatric disorders: A rural/urban comparison. *Archives of general psychiatry*, *42*(7), 651-656.

Bora, N., Chang, Y. H., & Maheswaran, R. (2014, April). Mobility patterns and user dynamics in racially segregated geographies of US cities. In International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (pp. 11-18). Springer, Cham.

Breslau, J., Marshall, G. N., Pincus, H. A., & Brown, R. A. (2014). Are mental disorders more common in urban than rural areas of the United States?. *Journal of psychiatric research*, *56*, 50-55.

Chen, N., & Akar, G. (2017). How do socio-demographics and built environment affect individual accessibility based on activity space? Evidence from Greater Cleveland, Ohio. *Journal of Transport and Land Use*, *10*(1), 477-503.

Clarke, C. C., Schiller, J. S., Boersma, P. National Center for Health Statistics. (2020). *Early Release of Selected Estimates Based on Data From the 2019 National Health Interview Survey*.

Domènech-Abella, J., Mundó, J., Leonardi, M., Chatterji, S., Tobiasz-Adamczyk, B., Koskinen, S., ... & Haro, J. M. (2018). The association between socioeconomic status and depression among older adults in Finland, Poland and Spain: A comparative cross-sectional study of distinct measures and pathways. *Journal of affective disorders*, *241*, 311-318.

Eaton, W. W., Muntaner, C., Bovasso, G., & Smith, C. (2001). Socioeconomic status and depressive syndrome: the role of inter-and intra-generational mobility, government assistance, and work environment. Journal of health and social behavior, 42(3), 277.

Echeverría, S., Diez-Roux, A. V., Shea, S., Borrell, L. N., & Jackson, S. (2008). Associations of neighborhood problems and neighborhood social cohesion with mental health and health behaviors: the Multi-Ethnic Study of Atherosclerosis. *Health & place*, *14*(4), 853-865.

Feuillet, T., Charreire, H., Menai, M., Salze, P., Simon, C., Dugas, J., ... & Oppert, J. M. (2015). Spatial heterogeneity of the relationships between environmental characteristics and active commuting: towards a locally varying social ecological model. *International Journal of Health Geographics*, *14*(1), 1-14.

Galea, S., Ahern, J., Rudenstine, S., Wallace, Z., & Vlahov, D. (2005). Urban built environment and depression: a multilevel analysis. *Journal of Epidemiology & Community Health*, *59*(10), 822-827.

Galea, S., Freudenberg, N., & Vlahov, D. (2005). Cities and population health. *Social science & medicine*, *60*(5), 1017-1033.

Gaynes, B. N., Burns, B. J., Tweed, D. L., & Erickson, P. (2002). Depression and health-related quality of life. *The Journal of nervous and mental disease*, *190*(12), 799-806.

Glaeser, E. L., Kahn, M. E., & Rappaport, J. (2008). Why do the poor live in cities? The role of public transportation. *Journal of urban Economics*, *63*(1), 1-24.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *nature*, *453*(7196), 779-782.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *nature*, *453*(7196, Suppl.), S1 – S17.

Graif, C., Arcaya, M. C., & Roux, A. V. D. (2016). Moving to opportunity and mental health: Exploring the spatial context of neighborhood effects. *Social science & medicine*, *162*, 50-58.

Gruebner, O., Rapp, M. A., Adli, M., Kluge, U., Galea, S., & Heinz, A. (2017). Cities and mental health. *Deutsches Ärzteblatt International*, *114*(8), 121.

Gutiérrez-Lobos, K., Scherer, M., Anderer, P., & Katschnig, H. (2002). The influence of age on the female/male ratio of treated incidence rates in depression. BMC psychiatry, 2(1), 1-8.

Hartig, T., & Kahn Jr, P. H. (2016). Living in cities, naturally. *Science*, *352*(6288), 938-940.

Hirvensalo, M., Sakari-Rantala, R., Kallinen, M., Leinonen, R., Lintunen, T., & Rantanen, T. (2007). Underlying factors in the association between depressed mood and mobility limitation in older people. *Gerontology*, *53*(3), 173-178.

Jones-Webb, R., & Wall, M. (2008). Neighborhood racial/ethnic concentration, social disadvantage, and homicide risk: an ecological analysis of 10 US cities. *Journal of Urban Health*, *85*(5), 662-676.

Judd, F. K., Jackson, H. J., Komiti, A., Murray, G., Hodgins, G., & Fraser, C. (2002). High prevalence disorders in urban and rural communities. *Australian & New Zealand Journal of Psychiatry*, *36*(1), 104-113.

Lee, M. A. (2009). Neighborhood residential segregation and mental health: a multilevel analysis on Hispanic Americans in Chicago. Social science & medicine, 68(11), 1975-1984.

Lorant, V., Deliège, D., Eaton, W., Robert, A., Philippot, P., & Ansseau, M. (2003). Socioeconomic inequalities in depression: a meta-analysis. *American journal of epidemiology*, *157*(2), 98-112.

Machin, S., Salvanes, K. G., & Pelkonen, P. (2012). Education and mobility. Journal of the European Economic Association, 10(2), 417-450.

MacQueen, K. M., McLellan, E., Metzger, D. S., Kegeles, S., Strauss, R. P., Scotti, R., ... & Trotter, R. T. (2001). What is community? An evidence-based definition for participatory public health. *American journal of public health*, *91*(12), 1929-1938.

McKenzie, K., Murray, A., & Booth, T. (2013). Do urban environments increase the risk of anxiety, depression and psychosis? An epidemiological study. *Journal of affective disorders*, *150*(3), 1019-1024.

Miles, R., Coutts, C., & Mohamadi, A. (2012). Neighborhood urban form, social environment, and depression. *Journal of Urban Health*, *89*(1), 1-18.

Milgram, S. (1970). The Experience of Living in Cities. *Science*, *167*(3924), 1461-1468.

Mitsue, S., & Yamamoto, T. (2019). Relationship between depression and movement quality in normal young adults. *Journal of Physical Therapy Science*, *31*(10), 819-822.

Mohino, I., Urena, J. M., & Solis, E. (2016). The influence of education on work-related travel in rural metro-adjacent regions: The case of Castilla-La Mancha (Spain). The Open Transportation Journal, 10(1).

Mykletun, A., Bjerkeset, O., Prince, M., Dewey, M., & Stewart, R. (2009). Levels of anxiety and depression as predictors of mortality: the HUNT study. *The British Journal of Psychiatry*, *195*(2), 118-125.

Northridge, M. E., Sclar, E. D., & Biswas, P. (2003). Sorting out the connections between the built environment and health: a conceptual framework for navigating pathways and planning healthy cities. *Journal of urban health*, *80*(4), 556-568.

Pearson, A. L., Griffin, E., Davies, A., & Kingham, S. (2013). An ecological study of the relationship between socioeconomic isolation and mental health in the most deprived areas in Auckland, New Zealand. *Health & place*, *19*, 159-166.

Pescosolido, B. A., Lee, B., & Kafadar, K. (2020). Cross-level sociodemographic homogeneity alters individual risk for completed suicide. *Proceedings of the National Academy of Sciences*, *117*(42), 26170-26175.

Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology & Community Health*, *55*(2), 111-122.

Sampson, R. J., & Levy, B. L. (2022). The Enduring Neighborhood Effect, Everyday Urban Mobility, and Violence in Chicago. University of Chicago Law Review, 89(2).

Sandoval, J. O. (2011). Neighborhood diversity and segregation in the Chicago metropolitan region, 1980-2000. Urban Geography, 32(5), 609-640.

Simon, G. E. (2003). Social and economic burden of mood disorders. *Biological psychiatry*, *54*(3), 208-215.

South, S. J., & Crowder, K. D. (1997). Residential mobility between cities and suburbs: Race, suburbanization, and back-to-the-city moves. *Demography*, *34*(4), 525-538.

Stier, A. J., Schertz, K. E., Rim, N. W., Cardenas-Iniguez, C., Lahey, B. B., Bettencourt, L. M., & Berman, M. G. (2021). Evidence and theory for lower rates of depression in larger US urban areas. Proceedings of the National Academy of Sciences, 118(31).

Substance Abuse and Mental Health Services Administration. (2020). *Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health* (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55). https://www.samhsa.gov/data/

Tartaglia, S. (2013). Different predictors of quality of life in urban environment. *Social Indicators Research*, *113*(3), 1045-1053.

Twitter (2022). Twitter API (v2). Twitter. https:/developer.twitter.com/en/solutions/academic-research

United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Urbanization Prospects: The 2018 Revision* (ST/ESA/SER.A/420). New York: United Nations.

US Census Bureau (2022). Relationship files. Census.gov. Rhttps://www.census.gov/geographies/reference-files/2010/geo/relationship-files.html#par_textimage_674173622

Vallée, J., Cadot, E., Roustit, C., Parizot, I., & Chauvin, P. (2011). The role of daily mobility in mental health inequalities: the interactive influence of activity space and neighbourhood of residence on depression. *Social science & medicine*, *73*(8), 1133-1144.

Villarroel, M. A., & Terlizzi, E. P. (2020). *Symptoms of depression among adults: United States, 2019*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Walker K, Herman M (2022). tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames. R package version 1.1.9.9000, https://walker-data.com/tidycensus/.

Wandersman, A., & Nation, M. (1998). Urban neighborhoods and mental health: Psychological contributions to understanding toxicity, resilience, and interventions. *American Psychologist*, *53*(6), 647.

Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. Proceedings of the National Academy of Sciences, 115(30), 7735-7740.

Weich, S., Blanchard, M., Prince, M., Burton, E., Erens, B. O. B., & Sproston, K. (2002). Mental health and the built environment: Cross–sectional survey of individual and contextual risk factors for depression. *The British Journal of Psychiatry*, *180*(5), 428-433.

Williams, D. R. (1999). Race, socioeconomic status, and health the added effects of racism and discrimination. *Annals of the New York Academy of Sciences*, *896*(1), 173-188.

Yazdavar, A. H., Al-Olimat, H. S., Ebrahimi, M., Bajaj, G., Banerjee, T., Thirunarayan, K., ... & Sheth, A. (2017, July). Semi-supervised approach to monitoring clinical depressive symptoms in social media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 1191-1198).

Zakus, J. D. L., & Lysack, C. L. (1998). Revisiting community participation. *Health policy and planning*, *13*(1), 1-12.