A Geographically Weighted Regression Model to Examine the Spatial Variation of the Socioeconomic and Land-Use Factors Associated with Strava Bike Activity in Austin, Texas

Sirajum Munira
Texas A&M Transportation Institute
505 E Huntland Dr, Austin, TX 78752
Email: munira_silvy@tamu.edu
ORCID: 0000-0002-4953-2628

Ipek N. Sener, Ph.D. (Corresponding author)
Texas A&M Transportation Institute
505 E Huntland Dr, Austin, TX 78752
Email: i-sener@tti.tamu.edu
ORCID: 0000-0001-5493-8756

ABSTRACT

Despite evidence showing the spatial nonstationarity of the determinants of bike activity, very few studies have addressed the phenomena, probably due to the limited sample size of the traditional count data. To address this gap, this study demonstrated the applicability of Strava bike activity data by developing a geographically weighted Poisson regression (GWPR) model that can reveal how the influence of socioeconomic and land-use factors vary across a region. The city of Austin was selected as a case study, and Strava bike volume was gleaned from 1494 intersections. The representativeness of the Strava data was first examined by comparing those data with the video-based actual bicycle volume data from 43 intersections in the study area. Despite the high deviation in several locations, Strava volume exhibited moderate linear relationships with actual volume. The GWPR model developed in this study outperformed the traditional global model and revealed significant spatial variability of nine variables related to age, income, education, transit stops, hub locations, offices, schools, trails, and sidewalk facilities. Notable spatial variations on bike activity were observed across the study area in terms of magnitude, direction, and significance of the impact for all model variables. The analysis and discussion offer guidance to practitioners and policymakers in tailoring policies and programs that consider the spatial context. The study also provides insights for understanding the potential use of crowdsourced data in examining bike activity, especially when resources are limited.

Keywords: Bike activity, Strava data, spatial variations, intersection, geographically weighted Poisson regression

1. Introduction

Nonmotorized activities such as walking and biking offer a broad range of individual and community benefits related to health, environment, economics, and road safety (De Hartog et al. 2010; Jacobsen 2003; Schnohr et al. 2006). In recognizing the myriad benefits linked to increased nonmotorized activity, transport professionals and health advocates worldwide have been committed to promoting walking and biking over the past several years. In order to formulate strategies to encourage active mode usage as well as prioritize nonmotorized facility projects, transport professionals must understand the nonmotorized demand and its relationship with determinants across a spatial domain.

Research is replete with studies documenting the influence of several determinants on bike and pedestrian demand, or volume, at both the macro (e.g., census tract, city) and micro (intersection or street segment) levels, mostly employing various aspatial models. The traditional methods are largely based on the assumption that the relationship between the built environment or sociodemographic factors and walking and biking is spatially homogeneous, which is not always true. Nonmotorized activity may vary widely by location and time, even across similar traffic and environmental conditions within a region (Griswold et al. 2011). It is likely that one variable may have strong influence on bike demand in one location but may exhibit weaker association elsewhere (Yang et al. 2017). In addition, while some variables (such as population density) have been established as a consistent influence across studies, variation in the direction of influence in some variables has been observed. For example, income was found to have both negative (Hankey et al. 2017) and positive influences (Strauss and Miranda-Moreno 2013) on bike activity. Varying effects of school land use were also observed (Hasani et al. 2019; Strauss and Miranda- Moreno 2013). While the influence of variables largely depends on the local community and population, it may also vary based on the type of neighborhood (urban, suburban, and rural mixture conditions) within the same region (Qin and Ivan 2001). If such variations in influence exist, it is particularly necessary to quantify and analyze them in order to better inform policies and investment decisions. However, a considerable gap exists regarding the handling of spatial non-stationarity in nonmotorized studies.

Geographically weighted regression (GWR) is a well-known method that addresses the issue of spatial nonstationarity (Brunsdon et al. 1998). This localized regression technique, although suggested to be more exploratory rather than confirmatory in nature (Cromley and Hanink 2014; Mennis 2006), allows to reveal geographical variations in the relationships between a dependent variable and explanatory variables (Harris et al. 2010). Despite its limitations, the GWR method has been the most popular spatially varying coefficient model, mostly due to its relative simplicity (Murakami et al. 2019). The method has been widely applied in health and environmental studies (e.g., Gao and Li 2010; Nakaya et al. 2005) and in a few transportation studies, such as traffic and transit analysis (e.g., Cardozo et al. 2012; Selby and Kockelman 2013; Zhang and Wang 2014). However, very few research efforts have emphasized nonmotorized activity analysis, particularly bike ridership investigations at the micro level (e.g., Yang et al. 2017). One probable explanation for this gap is that spatial analysis requires large-scale spatial data that are most often not practical to collect through traditional counting methods. Even with the automatic counting technologies, it is only feasible to gather nonmotorized activity data from a limited number of locations. Recently, the growing proliferation of crowdsourced fitness app data (such as

Strava, Endomondo, or MapMyRide) has offered a solution to the issues of limited temporal and spatial coverage of bike data posed by the traditional counting methods. Big data sources, characterized by their volume, velocity, and variety (Laney 2001), offer great potential to understand the detailed spatiotemporal travel pattern of nonmotorized traffic on an unprecedented level of detail. Although such crowd-sourced data have often been blamed for lacking quality assurance and representativeness (Goodchild and Li 2012; Jackson et al. 2013) and are considered inadequate without validation from an actual location count (Jestico et al. 2016), they no doubt offer enhanced space-time resolution that can be leveraged to gain insights into the spatial variation of features affecting bike activity, whereas manual count data fall short. A few studies focusing on crowdsourced data (such as a bike-sharing system or fitness app) have incorporated spatial dependency or non-stationarity of the observations while investigating bike activity (e.g., Griffin and Jiao 2015; Ji et al. 2018; Lee and Sener 2019; Shen et al. 2018) and demonstrated the superiority of various spatial models. The findings underscore that investigation of the influence mechanisms of different features is needed for optimized resource allocation and focused policy efforts.

Among many crowdsourced data sources, Strava has been creating a huge dataset, gleaning user data from both pedestrians and bicyclists. Strava is one of the largest cycling fitness apps, having reported distance data of 8 billion miles from 48 million users across 195 countries in 2019 (Strava 2019). Based on the large number of users, a handful of researchers have investigated the use of Strava data sources to examine various issues of nonmotorized travel, including the impact of the built environment and socio-demographic features on cycling behavior (Griffin and Jiao 2015; Hochmair et al. 2019; Sun et al. 2017), accident risk (Saha et al. 2018; Sanders et al. 2017), and air pollution during cycling (Lee and Sener 2019; Sun and Mobasheri 2017). For example, Griffin and Jiao (2015) employed GWR in the Austin area to evaluate bike ridership gleaned from Strava data (at the census block group level) with respect to residential and employment density, land-use diversity, bicycle facilities, and terrain. Readers are referred to Lee and Sener (2020a) for an extensive review of Strava Metro data and their use in active transportation literature. Moreover, a number of studies in the field analyzed OD-level Strava data, which required to aggregate the trips at a larger scale, for instance, at city level (Selala and Musakwa 2016) or block group level (Hochmair et al. 2019). Prior studies high- lighted the importance of conducting a more disaggregate-level analysis, such as at an intersection level, for better insights into the cycling patterns (e.g. Selala and Musakwa 2016).

In addition, while investigating the representativeness of the Strava data, a few studies have compared Strava bike volume and manual count volume for their study area. For example, Jestico et al. (2016) examined the representativeness of the Strava data on total bicycle volume in Victoria, Canada, and based on their results, suggested the use of Strava volume as an indicator of the actual bicycle volume. A similar study (Sanders et al. 2017) that developed a bicycle exposure model for the Seattle area also suggested that Strava volume can be a reasonable proxy of the total bicycle volume in certain circumstances. Contributing to the growing research in the use of crowdsourced data, and in light of the aforementioned research gap in bicycling research, this study harnessed the enhanced resolution offered by the Strava data to develop a GWR model to better characterize the spatial distribution of bike activity and to quantify the spatially nonstationary relationships between the surrounding influencing features and bike demand in intersections. To the best of the researchers' knowledge, no study has explored Strava bike volume at the microscopic level (inter- section) for the purpose of developing a spatial model utilizing socio-economic and built environment features. Moreover, this study also contributes to the efforts of city officials because a large proportion of crashes occur at intersections in the study area (Texas Department of Transportation (TxDOT), 2016), and gathering insights into the inter- section-level bicycle volume is essential in developing focused policy efforts and effective safety implementation plans.

2. Study area and data

2.1. Study area

With an area of 326 square miles, the city of Austin accommodates a population of over 996,369 (City of Austin Planning and Zoning 2020). Downtown Austin, which is located on the north bank of the Colorado River, is the central business district of the city. The University of Texas (UT) at Austin, accommodating over 50,000 students, is located north of the downtown area. Although the eastern part of the city is flat, the western part contains some hilly terrain. The city is also home to several natural and man-made lakes.

Austin makes an excellent case study for this research for multiple reasons. First, the city has experienced tremendous population growth, especially since 2000. While the neighborhoods on the city's edge (far south Austin) have observed dramatic increase, growth within the city's inner regions (downtown, Montopolis, and Pleasant Valley) is also notable (Hedman et al. 2017). Second, by its very nature, the city is diverse in terms of age, culture, income, and built environment characteristics, and it has experienced a steep rise in the degree of socio-economic spatial separation over the last few decades. Since 2000, the city's Hispanic and Latino population share has increased and is mainly concentrated on the east side (Hedman et al. 2017). There has also been a rise in younger, single-person households. The growing economic development has attracted a young, well-educated, and environmentally and health-conscious population to the city. Third, the diversity in the land-use characteristics across the city is also notable. Despite being heavily car-dependent, especially in suburban neighborhoods, the city has observed a significant increase in bicycle commuters in the last few years. Moreover, the region has a total of 267.5 miles of bicycle facilities, including protected and buffered bicycle lanes and urban trails (shared-use paths; City of Austin 2018a). Although the improved facilities have spurred increases in bike activities in many locations, some areas have not observed significant bike activity, which might be attributed to the local socioeconomic or land-use characteristics of a neighborhood.

2.2. Bicycle volume data

The primary data, intersection bicycle volume, utilized in this study were obtained from Strava Metro® (proprietary source) through the Texas Department of Transportation (TxDOT). Strava Metro is a data service that produces anonymized and aggregated activity data from users of the Strava app, which allows cyclists and runners to track their activities (such as rides, runs, and walks) on a smartphone or other GPS device.

The obtained dataset contains three subsets in three formats: streets, origin-destination, and nodes. Because the spatial unit for this study was the intersection, node-level data (street intersections) were extracted. Strava Metro reports both all-purpose cycling activity and commuting activity counts for the nodes. This research processed the total bicycle volume count for all nodes for the year 2017 since the actual volume data were available for the same year. In order to overlay the Strava nodes with the street intersections, the bicycle street network for the study area was used, as provided by the Austin Department of Transportation. The process extracted 2520 intersections, and Strava annual (2017) bicycle volume at various intersections ranged from 5 to 4005, meaning that some locations observed only 5 bicyclists in an entire year. However, the observations might have been influenced by two factors: (a) the actual volume might have varied greatly from the Strava volume in that location, and/or (b) the node/intersection might have represented excessive nodes of the network, such as alleys, driveways, or off-street pedestrian paths. Therefore, those data points were identified as unreasonable or outliers—a common phenomenon for crowdsourced, or in particular Strava data (Wang et al. 2016). Thus, to mitigate possible bias in the model exhibiting the relationship between bike volume and various explanatory variables, a data filtering task was performed. To remove intersections with very small annual Strava bike volumes, only intersections experiencing at least an annual Strava bike volume of 1000[1] (i.e., around three bike activities per day) were selected. The final dataset contained 1494 intersections in the study area.

To examine the representativeness of the Strava data, the actual bicycle volume data was also estimated at the intersections in the study area using two types of bicycle count data: short count data (24-hour data at 43 intersections collected in 2017) and continuous count data. The short count data were obtained from the City of Austin Transportation Department, which collected the count data by a video recorder, at each intersection, on typical weekdays over a 5 month- period (April, May, June, August, and October) in 2017. The continuous location counts, which are available since 2012, were obtained for 11 locations in the study area from Eco-Counter, a company that assists with continuous data collection for pedestrians and bicyclists in specific locations across cities around the world (Eco-Counter 2019). The continuous count data were utilized to estimate the daily and monthly factors that can be used to scale the short count data to a representative value (Nordback et al. 2013). Following standard guidelines of factor adjustment process for representative volume estimation (Johnstone et al. 2017; Nordback et al. 2013; Turner et al., 2018), annual average daily bicycle volumes were estimated for 43 intersections in the study area.

2.3. Explanatory variables

In order to gather data related to explanatory variables for the study, prior studies were reviewed in detail focusing on insights capturing the association between bicycle traffic and socioeconomic and land-use features (Chen et al. 2017; Dill 2009; Ewing and Cervero 2010; Hankey et al. 2017; Hasani et al. 2019; Strauss and Miranda-Moreno 2013; Tabeshian and Kattan 2014). This study sought to build a rich set of explanatory variables using insights from the earlier studies as well as the data available for the study area. The gathered variables in this study can be broadly classified into five categories: demographics, socioeconomics, pedestrian- or bicycle-specific infrastructure, transit facilities, and land use (for more details on the explanatory variables, see Munira and Sener 2017). Buffer-zone radii of 0.1 mile, 0.5 mile, and 1 mile were created; the sizes were derived from previous studies (Hasani et al. 2019; Tabeshian and Kattan 2014).

[1] Cutoff values ranging from 500 to 2000 annual bike volume were also assessed but are not reported due to the brevity of the paper. Based on the model performance, the researchers selected the cutoff value of 1000 for final model building.

The demographic and socioeconomics variables included age, gender, education, race, household size and occupancy status, income, and commute mode and time of the surrounding population. The bicycle-specific infrastructure-related variables included different types of bicycle infrastructure based on the conditions and comfort level, which was developed by the City of Austin (2017), bike signal, intersection density, and bike-sharing stations. Various transit-facility-related variables were gathered, including frequency of transit stops, transit route length, and distance from hub locations. Land-use variables, such as number of schools, offices, industries, open areas, mixed-use developments, water areas, and bicycle-accessible bridges, etc. were also gathered based on available data.

Note that although the Strava bike volume does not represent the actual trip generation, it is expected to be related to the bicyclist population living/working around the region since bicycle trips are mostly short trips (e.g., Dill 2009 noted that median bicycle trip length is below 3 miles). Therefore, the sociodemographic variables were added into the models by taking empirical evidence from previous studies and assuming that bike volume is influenced by the characteristics of the population living near the location.

Data were gathered from both publicly available sources and private communications with the City of Austin Transportation Department. Public data were gathered from the City of Austin data portal, ACS 2017 survey, City of Austin Planning and Development Review Department, Texas Education Agency, Austin Transportation Department Arterial Management Division, and Capital Metro data portal. Moreover, an updated (2019) bicycle infrastructure map was obtained from the Data and Technology Services of the City of Austin Transportation Department.

Since all of the raw datasets obtained from different sources were at different spatial scales, the datasets were cleaned and processed to bring them to homogenous spatial scales (buffer level). Over 140 variables of 30 distinct types were created for each of the buffer zones. Several new variables were created by aggregating the original variables when deemed meaningful and statistically necessary. For example, the variable “total population between the age of 18 to 34” was created by aggregating the population of six categories of ages for both males and females. Similarly, the variable “paved and unpaved trail length” was created by aggregating the paved and unpaved trail lengths for the buffer zones. The data processing and analysis were performed using R statistical software and ArcGIS.

3. Methodology

3.1. Comparison of Strava and actual volume data

The annual Strava volume data were compared with the actual annual average bicycle volume data at the same locations to explore the representativeness of the Strava data. The comparison was performed using three key statistics—average percent deviation (APD), average of the absolute percent difference (AAPD), and Pearson's correlation coefficient—following guidelines from the Transportation Research Board's *Guidebook on Pedestrian and Bicycle Volume Data Collection* (Ryus et al. 2014). The APD represents the overall divergence from the actual volume data. The AAPD is a measure of the source's consistency. Pearson's correlation coefficient is a measure of linear correlation between the actual bike volume and Strava cyclist volumes.

3.2. Model variable selection

The variable selection was completed through an extensive three-stage procedure. First, a simple OLS model was developed to analyze the relative strengths of relationships between each of the explanatory variables and the dependent variable (Strava bike volume). This included identification of the variables that had significant association—at a 90% confidence level—with bike volume. Second, Pearson's correlation coefficients were examined for each pair of explanatory variables to investigate the correlation between variables. This process yielded a large number of significantly correlated variables, from which highly correlated (at 0.7) variable pairs were recognized. Finally, by iterating several combinations of variables that are not highly correlated, a final model was selected based on parsimony and intuitive considerations (taking into account the local conditions), as well as the significance and statistical fit, using the corrected Akaike's information criterion (AICc) value that compared the goodness of fit of the models. A collinearity inspection of the final model variables was also conducted to ensure no multicollinearity. The same set of variables were used in the local and global models for comparison, which is consistent with previous literature (Blainey 2010; Yang et al. 2017).

Note that the process of variable selection may have also been performed by utilizing various state-of-the-art machine learning approaches, including Lasso or Random Forest. However, this study opted for the manual approach over those machine learning processes because they are often referred to as the black box approach with limited interpretability (Chen et al. 2017). The approach followed in this study promotes a profound understanding of the influence of individual variables and the dynamics of their relationship, given the local condition, in order to make informed decisions.

3.3. Geographically weighted Poisson regression modeling

To explore the local variations of the influence of socioeconomic and built environment features on the Strava bike activity at intersections, a GWPR model was developed for this study. The GWPR model can handle the discrete and dispersed count nature of the data while accounting for the spatial nonstationarity and spatial heterogeneity of parameter estimates.

The form of a GWPR model is as follows:

$$\ln y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \ldots\ldots + \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{1}$$

where y_i is the dependent variable; *i* denotes intersections of the study area; x_{ik} denotes the independent variables of the model, where coefficients $\beta_k(u_i, v_i)$ are varying conditionals on the location; (u_i, v_i) is the location of intersection i; and ε_i is the random error term.

The estimated coefficient, β at location i, can be obtained using the following equation (Fotheringham et al. 2002):

$$\beta_n(u_i, v_i) = (\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_i, v_i)\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{W}(u_i, v_i)\mathbf{Y} \tag{2}$$

where $\mathbf{W}(u_i, v_i)$ denotes an *n* by *n* spatial weight matrix that can be expressed as $\mathbf{W}(i)$:

$$\mathbf{W}(i) = \begin{bmatrix} w_{i1} & 0 & \ldots & 0 \\ 0 & w_{i2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & \ldots & w_{in} \end{bmatrix}$$

where the diagonal element $w_{ij} (j = 1 \ldots. n)$ is the weight matrix given to intersection j in the calibration of the model for intersection i, and off-diagonal elements are 0.

In this modeling framework, a regression equation is estimated for each intersection based on the observations in nearby intersections, which means the observations closer to intersection *i* have more of an influence on the parameter estimation of *i* than the intersections far from it. This effect gradually decreases as the distance between the two locations increases. To estimate the smoothed geographical variations in the parameters with a distance-based weighting scheme, GWPR employs a spatial kernel method. The kernel function is generally one of two types: Gaussian or bi-square (Fotheringham et al. 2002).

The kernel's bandwidth, cut off for the distance to assign weight, could either be fixed (based on distance) or adaptive (based on a specific number of neighbors). The GWPR modeling outcomes are sensitive to bandwidth selection (Yu and Peng 2019). A small bandwidth encompassing a small number of observations may result in unstable fits, while too large a bandwidth may introduce bias. Studies have indicated that adaptive kernels are suitable when observations are sparsely dis- tributed over space because some locations may only have a few neighbors if fixed methods are used (Chen et al. 2017; Feuillet et al. 2015). Given the irregular distribution of Strava bike activity across the study area, this study employed the adaptive kernel method coupled with a bi-square weighting scheme to ensure that each local regression encompassed enough regression points regardless of the surrounding density.

The bi-square weighting scheme can be specified as follows:

$$w_{ij} = \begin{cases} \left[1 - \left(d_{ij} / h_i\right)^2\right]^2 & if\, d_{ij} < h_i \\ 0 & Otherwise \end{cases}$$

where d_{ij} is the distance between two observations i and j, and h_i is the maximum distance to assign nonzero weights. Observations located beyond bandwidth h_i are assigned a weight of zero.

Using an adaptive bandwidth, the minimum number of neighbors to be included in calibrating the local models can be identified. The AICc (Akaike 1974) is generally utilized to determine the optimal bandwidth, in which the model with the lowest AICc is identified as the best model (Fotheringham et al. 2002; Nakaya et al. 2005). In general, models are not deemed significantly different if the difference of AICc values is less than 4 (Charlton and Fotheringham 2009).

In addition, some of the explanatory variables selected for the model may not be subject to local dependency. In those cases, instead of a fully local or a fully global model, a semiparametric model combining the global and local effects of the variables is more appropriate. For this study, a geographical variability test was performed to examine the spatial variability of each parameter and to determine which variables vary over space. For the geographic variability check, different semiparametric models were developed wherein the variables in question were kept as fixed (global) while the other variables were kept as local. The variables will vary over space if the original GWPR model performs better than the semiparametric model in terms of comparison criteria such as AICc (Nakaya et al. 2005). The test suggested that each of the model variables significantly varied over space. Therefore, this study built the fully local GWPR model because it was expected to best explain the varying relations between bike activity and the explanatory variables.

To perform the GWPR analysis, this study used GWR4, a Windows application for geographically weighted regression modeling developed by Nakaya (2012). Although evaluating a geographically weighted negative binomial regression model might have been beneficial for the study, the latest available software—GWR4.0—does not allow for the calibration of a geographically weighted regression model with a negative binomial structure. Thus, the model could not be developed for this research. However, since the local models are fitted using several similar adjoined observations, it is expected that the variance of bike volume will become much closer to the mean while estimating the local parameters in a GWPR setting (Li et al. 2013; Xu & Huang, 2015). Also, studies conducting similar analyses have noted that Poisson regression does not produce significantly different results compared to the negative binomial model since the coefficients of the models are similar for the two error distributions (Hadayeghi et al. 2010; Li et al. 2013).

4. Results and discussion

4.1. Comparison results of Strava and actual volume data

Although Strava provides a large sample size with enhanced temporal and spatial resolution, sampling bias is a well-recognized concern across emerging datasets, including Strava (Lee and Sener, 2020a,b). For example, Strava data tend to oversample male and fitness riders and undersample young, female, and novice bicyclists, as noted by other studies (Boss et al. 2018; Jestico et al. 2016). The results of this study confirmed the variations between Strava and actual volume data.

A comparison of the actual volume and Strava data for 43 inter- sections showed that the percentage deviation varied from −43% to −97%, which indicated that the divergence of Strava volume from the actual volume varied with space. The variations differed across geographic locations (Fig. 1), probably due to a difference in demographic characteristics and trip purpose distribution in different areas. The APD was −89%, which was fairly high. The AAPD was 89%, which indicated that Strava volume was less than the actual volume in all locations, as expected. Moreover, Pearson's correlation coefficient was $r = 0.63$ ($p < 0.0001$), which indicated a strong linear association between the volume data from the two sources. The associations were stronger when compared to other recent studies. For instance, Turner et al. (2019) reported a Pearson's of $r = 0.59$ and an AAPD of 92.49% when comparing Strava and actual volume data from eight counting locations in Austin. For a study in Miami-Dade County, Florida, Hochmair et al. (2019) compared Strava activity counts and video-based actual count data on 32 sites and reported a correlation of $r = 0.5$.

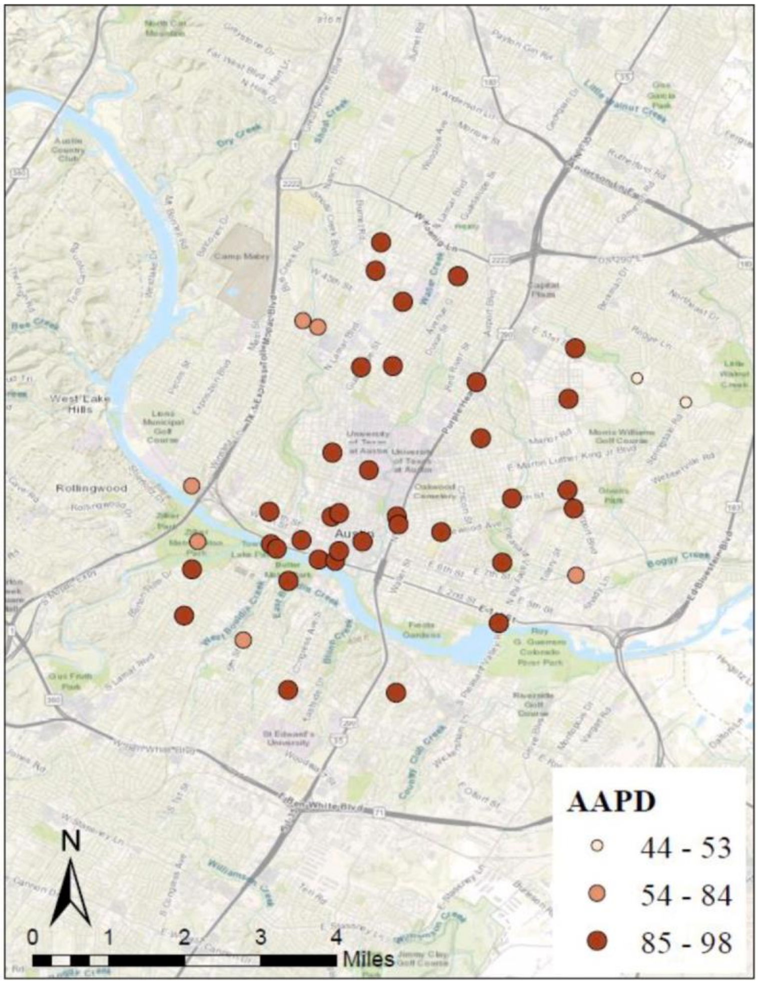


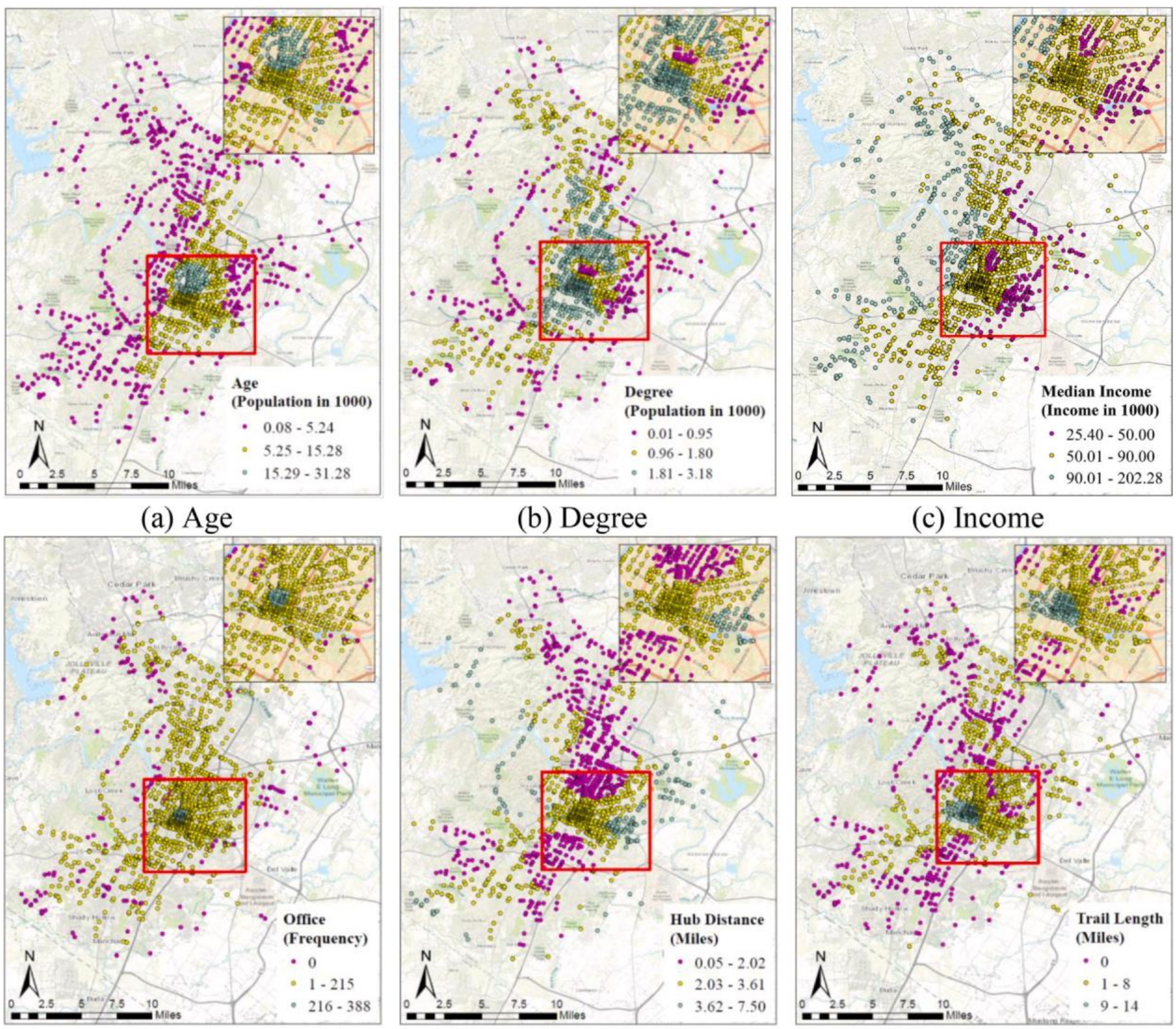
Fig. 1. AAPD of Strava bicycle volume in 43 locations in Austin.

4.2. Description of explanatory variables and spatial distribution

The final model contained nine explanatory variables at varying buffer zones: population age 18 to 34 (1.0 mile), population with at least a bachelor's degree (1.0 mile), median income (1.0 mile), frequency of school (0.5 mile), length of helpful sidewalk (0.1 mile), trail length (paved and unpaved; 1.0 mile), frequency of office establishment (0.5 mile), frequency of transit stop (0.5 mile), and distance from transit hub. A collinearity inspection of the nine variables indicated that all variance inflation factor values were less than 2.6, confirming no collinearity among the final model variables.

The age, education, and income variables were noted to have significant influences on bike activity in several other studies. It was also expected that the presence of public schools (primary and secondary regular, charter, and alternative schools) and office establishments would have an impact on bicyclists in an area. The transit hub refers to a site containing transit stations (such as bus and/or light rail) and other facilities, including a Park & Ride, a space dedicated for bikes, carshare, and more. It was expected that the decreasing distance from hub locations to the intersections would increase bike activity. The helpful sidewalk variable refers to sidewalk facilities next to a less- comfortable road (high traffic volumes and speeds, and little or no bicycle accommodations). Finally, the trail facility refers to the length of both paved and unpaved trails, which is expected to promote bike activity.

Fig. 2 illustrates the spatial distribution of the six variables that showed notable variation across the study area. For discussion, this study mainly followed the neighborhood definition provided by Hedman et al. (2017). The spatial distribution of demographic characteristics (Fig. 2a, b, and c) across the city exhibits notable variation. The high concentration of the young population around the city center can be attributed to the UT student population. Overall, the affluent populations are concentrated on the city's west side rather than on the east. A previous report (Hedman et al. 2017) also suggested that the less wealthy, more diverse communities are generally concentrated on the east side of the city.



(a) Age (b) Degree (c) Income

(d) Office establishment (e) Distance from hub (f) Trail length

Fig. 2. Spatial distribution of explanatory variables.

The intersections in the central region are surrounded by many office establishments (Fig. 2d). The concentration decreases as the distance from downtown increases. Moreover, intersections in the central, north, and southwest regions are served by hub facilities (Fig. 2e), while intersections in the southeast regions are located far from hub locations. Similarly, a majority of the intersections in the study area are served by multiple transit stops (bus and rail), except some intersections on the northeast side. Moreover, the length of trail facilities (both paved and unpaved) is highest around the central area near Zilker Park (Fig. 2f). A number of intersections in the northwest and southwest regions do not have any nearby trail facility.

4.3. GWPR model results

4.3.1. Model performance

The final model's performance was first compared with a global model, which is the same as the traditional Poisson model, in terms of R-square and AICc. As shown in Table 1, the final GWPR model has a significantly smaller AICc and a larger percent deviance explained (pseudo R-square) compared to the global model. The findings further emphasize the spatial variation in the relationships between the predictors and bike activity.

Table 1. Model performance comparison.

Model Type	AICc	Percent Deviance Explained
Global Model	47,925	0.15
GWPR Model	28,423	0.5

Note. Bandwidth size is 164.

Fig. 3 presents the spatial variation of R-square obtained from the GWPR model. Spatial variations of the local R-squared estimates illustrate the difference in the combined statistical impact of the final model variables on bike activity across the neighborhoods in the Austin area, from very low (0.1) to high (0.66). The highest predictive power of the model was observed in the central north, southwest, and downtown regions. The predictive power was especially lower in the northeast region, which emphasizes the importance of conducting an in-depth and more spatially focused analysis for the related areas. In general, the findings indicate that significant heterogeneity exists among different explanatory features when relating them to bike volume. This heterogeneity, however, cannot be captured by traditional global models.

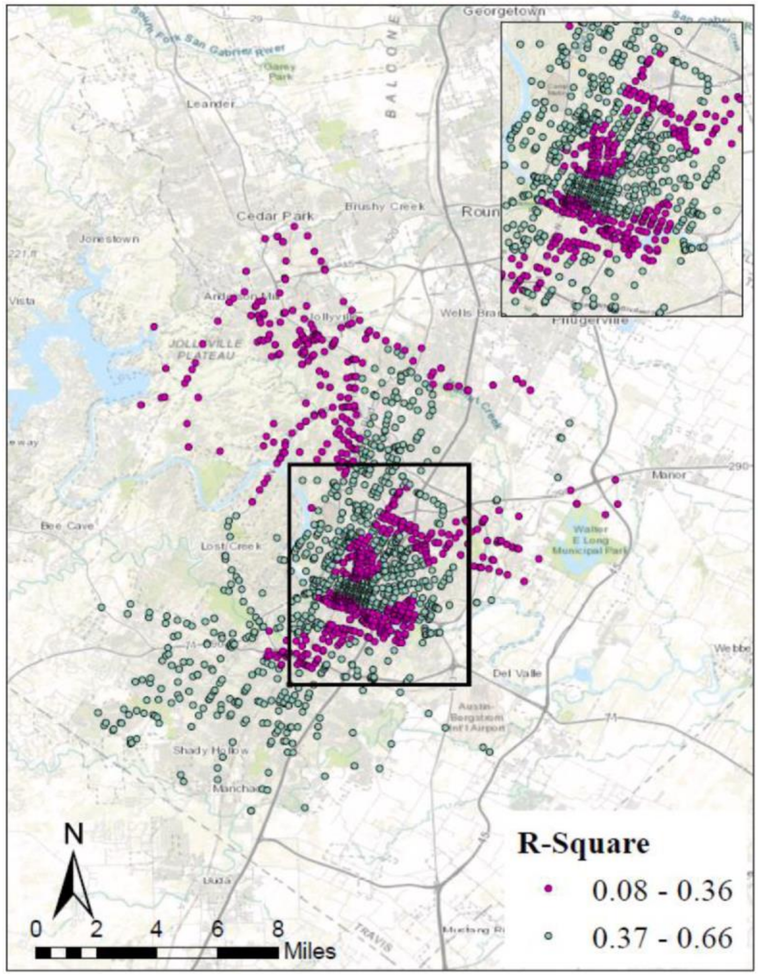


Fig. 3. R-squared values derived from the final GWPR model.

4.3.2. Model estimates

The coefficient estimates from the GWPR model exhibited notable variation in terms of magnitude, direction, and significance. A summary of the direction and significance of the relationship of the variables is presented in Table 2. The table shows that each of the nine variables had both positive and negative influences on bike volume, depending on the locations of the intersections. Each of the variables had a varying proportion of positive and negative significant estimates.

Table 2. Summary of the directions of relationships derived from the GWPR model.

Variable (buffer length)	Total Significance ($p < 0.1$)	Among Significant ($p < 0.1$)	
		Positive	Negative
Population age 18 to 34 (in 1,000; 1.0 mile)	18%	56%	44%
Population with bachelor or higher degree (in 1,000; 1.0 mile)	16%	57%	43%
Median income (in 1,000; 1.0 mile)	14%	50%	50%
Frequency of school (in 100; 0.5 mile)	20%	45%	55%
Frequency of office establishment (in 100; 0.5 mile)	19%	41%	59%
Distance from transit hub (in miles)	14%	51%	49%
Frequency of transit stop (in 100; 0.5 mile)	17%	61%	39%
Length (mile) of helpful sidewalk (0.1 mile)	19%	43%	57%
Paved and unpaved trail length (1.0 mile)	15%	53%	47%

Note. A total of 1,494 intersections; dependent variable: Strava bike volume.

Fig. 4 illustrates the spatial distribution, strength, and direction of coefficient values describing the relation between the nine model variables and bike activity. Due to space considerations and to bolster the argument, only significant coefficients of the final model are discussed in the following sections.

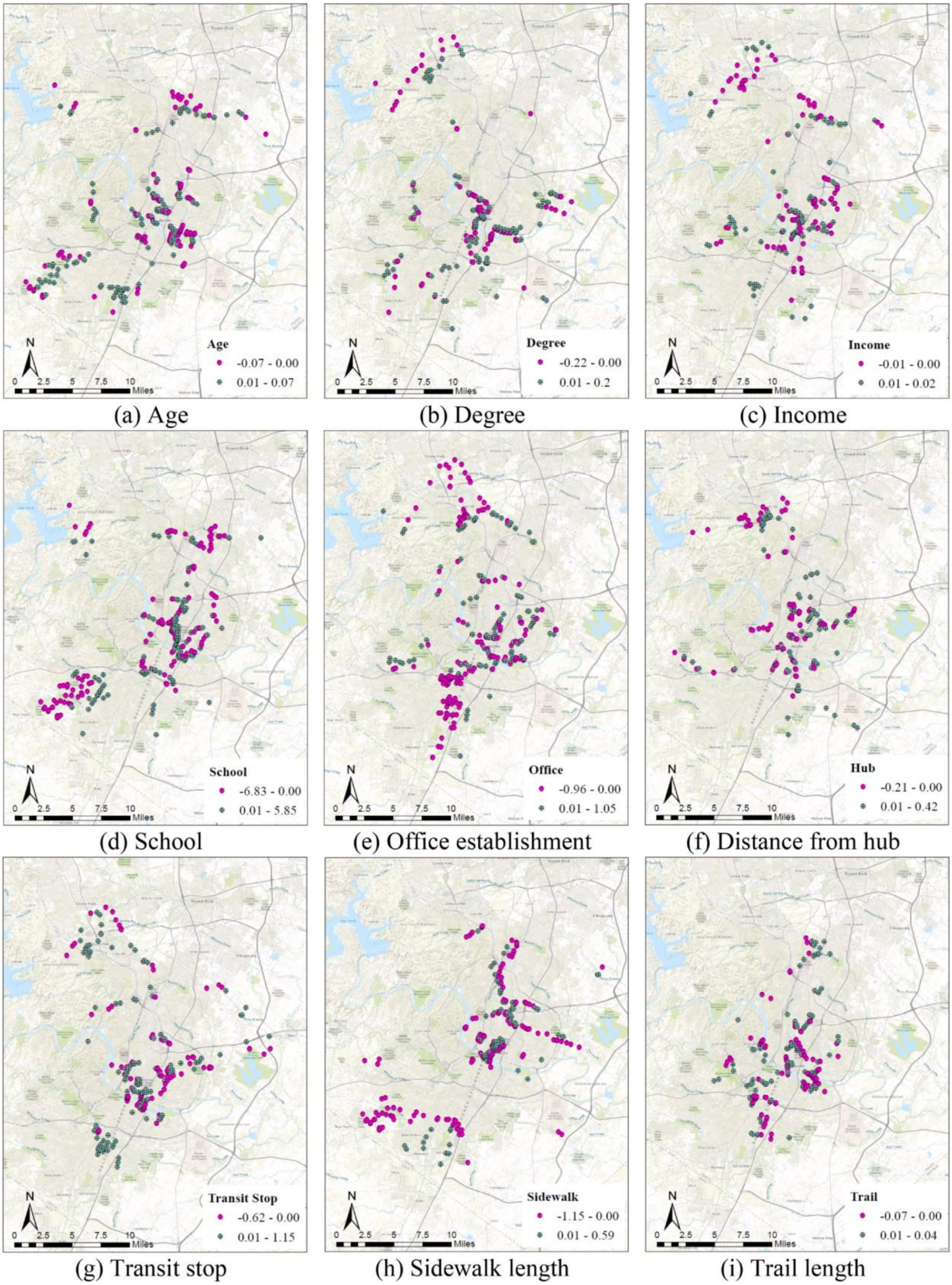


Fig. 4. Spatial distribution of the final GWPR model coefficients (significant at $p < 0.1$).

In terms of age, the young population (age 18 to 34) was associated with a greater bike activity at the nearby intersections. Similarly, the population with a higher degree had positive influences on bike volume at most intersections in the study area. These findings are consistent with previous studies indicating higher biking among young and educated populations (e.g., Sallis et al. 2013). However, the equal distribution of positive and negative influence of the income variable suggests that both low- and high-income populations contribute to the bike activity in the city of Austin.

In terms of land-use factors, a high frequency of school and office establishments mostly have a negative association with bike activity at nearby intersections. An interesting finding was observed in terms of the positive influence of office establishments, which were mainly concentrated in central Austin and decreased with increasing distance from downtown. This positive influence may likely be attributed to individuals living in the central/downtown region being more likely to opt for biking for their daily commute, whereas driving is a more likely choice for suburban commuters. On the other hand, the positive influence of schools was mainly observed near the university/college campuses of the study area, such as the UT campus in the downtown area as well as the north and south campuses of Austin Community College. This result suggests that although primary or high school students generally avoid biking to school, university students are more likely to bike to campus or nearby locations. A similar finding was reported by Schneider and Stefanich (2015), who indicated the positive influence of proximity to a university campus and mixed land uses on bicycling.

Further, a high frequency of transit stops was found to generally encourage bike activity at nearby locations, which was expected given that bicycles are allowed on buses and trains in Austin (CapMetro, 2017). However, the north and south regions, away from downtown, exhibited more positive influences of transit facilities on bicycling compared to downtown. This result indicates that bicyclists living in suburban regions are more likely to take advantage of the increased mobility and accessibility offered by integrated transit facilities. Furthermore, the transit hubs exhibited both positive and negative in- fluences on bike activity.

The observed relationship between a helpful sidewalk and bike activity was unexpected but insightful. Generally, it was seen that despite having a helpful sidewalk adjacent to high-speed and high-traffic roads, bicyclists tended to avoid those areas, although a few downtown intersections exhibited the usefulness of helpful sidewalk facilities for bicyclists. This finding is likely the result of bicyclists in suburban regions tending to feel less safe on roads, mostly due to motorized traffic, compared to those living in urban regions (National Highway Traffic Safety Administration 2008). Moreover, while most of the intersections demonstrated the positive influence of trail locations, results from some intersections, such as in the East Austin and Riverside neighborhoods, suggested that a trail facility is not a positive determinant of bike activity. A possible explanation for this result is the poor condition of the trails in that area (Buchele 2019). A recent survey conducted by the City of Austin (2018b) revealed local people's desire to improve the quality and connectivity of nonmotorized facilities in the noted area.

5. Conclusion

This study examined the spatial influence mechanisms of various socioeconomic and land-use features on intersection bike volume in Austin using Strava data. A GWPR model was developed utilizing nine variables that were selected from a rich set of explanatory variables. The model, which outperformed the global model, revealed significant spatial variability of the explanatory variables with relation to bike volume that could not be reflected by the traditional models.

In general, the location-specific estimates of the socioeconomic and built environment variables demonstrated how the determinants of bike activity vary between urban and suburban neighborhoods. One of the notable observations was that the central region, especially the area around the UT campus and downtown, exhibited the most variation. For example, both positive and negative influences of age, income, and degree on bike activity across this area indicated the varying socioeconomic characteristics of people biking in this area. While office and school land-use impede bicyclists across most of the study area, this specific region largely exhibited positive associations. Even the sidewalks next to the less comfortable roads (high traffic volumes and speeds, and little or no bicycle accommodations) mostly appeared to be a positive determinant of bike activity. This area, which is dominated by mixed land-use development, has comparatively higher and better nonmotorized facilities (City of Austin 2014). The overall positive im- pact of a trail and helpful sidewalk on the bike activity in this area suggests that balanced and mixed land use with continuous and connected nonmotorized facilities can encourage more bike activity among a population that reflects different demographics. On a similar note, the negative influence of some trail facilities along the east side (of Interstate 35) of the central region suggests that uncomfortable and poorly connected trail facilities may deter bicyclists even if those facilities are located in areas of high bike activity.

In summary, the contribution of this study is twofold. First, this study demonstrates the applicability of Strava-gleaned bike data to illustrate the variability of the determinants, which traditional count data with limited sample size cannot emulate. This might be particularly essential in cases where agencies are limited in resources (budget, time, staff, etc.) to collect additional data and need support from alternative, supplementary sources of data. Important to note that, while Strava has enriched bicycle research capabilities as exemplified in the current study, several challenges exist (e.g. under-representativeness of the general population, bias towards and away from certain groups) and hence researchers need to be cautious when generalizing the results as noted by Lee and Sener (2020a). Second, this study helps inform policy recommendations by illustrating how policies that encourage bicycle activity, including building or improving new infrastructures, should not be adopted uniformly throughout the city because the driving factors differ across neighborhoods. Instead, the characteristics and needs of the specific regions should be well understood and differences across regions should be recognized when making policy decisions. For example, people living in the downtown region may perceive sidewalks adjacent to high-speed roads as advantageous in reaching to destinations, while suburban populations may not feel comfortable being next to a busy, high-speed highway and tend to avoid those routes. Thus, efforts to increase the number of people commuting to offices and schools from suburban areas may warrant improved bike facilities with enhanced connectivity. In addition, this finding emphasizes the need for initiatives like Safe Routes to School (City of Austin 2018c) to encourage elementary and middle school students to bike to school. Furthermore, the study findings reveal that neighborhoods dominated by mixed-use development have more potential to increase bike mode share among populations of different demographics. This finding may have implications in future land-use planning of the city. The scope of this study can be extensively widened. Similar to other studies of nonmotorized activity models (Hochmair et al. 2019; Winters et al. 2010), this study utilizes cross-sectional data and hence the findings provide evidence of associations. Future research on multi-year panel data will be necessary to further build the body of evidence for causality (Xie and Levinson 2010). Furthermore, neighborhoods with low model fit merit further investigation, particularly to ascertain if additional variables that were not available in the current study—for example, number of students at a university or college, frequency of transit service—can explain the variability of bike activity in those locations. Moreover, since Strava also provides data at enhanced temporal resolution, future GWPR models can be extended to account for the temporal variation of bike activity. Finally, the GWR model of this study sought the 'best-on-average' scale (or single bandwidth) of relationship non-stationarity. Future studies would benefit from testing this assumption and examining the potential of incorporating flexible bandwidth (Leong and Yue 2017; Fotheringham et al. 2017) since some relationships (between the dependent and explanatory variables) may operate at a larger or smaller scale (Murakami et al. 2019). In addition, future research will be important to utilize diagnostic tools in order to investigate and account for collinearity and outliers if exists (Wheeler 2007; Harris 2019). Different model specifications can also be examined to test the sensitivity of this analysis.

Declaration of Competing Interest

None.

Acknowledgments

This research was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation's University Transportation Centers Program (Federal Grant Number: 69A3551747115). The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The U.S. Government assumes no liability for the contents or use thereof.

The authors would like to thank the City of Austin for its assistance in furnishing data required for this research, and TxDOT's Public Transportation Division for making the Strava Metro data available to Texas A&M Transportation Institute (TTI) researchers. The authors would also like to acknowledge TTI researcher Boya Dai for her assistance in preparation of the Strava data and TTI editor Dawn Herring for her editorial review. The authors also thank two anonymous reviewers and the editor for their insightful feedback.

References

Akaike, H., 1974. A new look at the statistical model identification. In: Selected Papers of Hirotugu Akaike. Springer, New York, NY, pp. 215–222.

Blainey, S., 2010. Trip end models of local rail demand in England and Wales. J. Transp. Geogr. 18 (1), 153–165. https://doi.org/10.1016/j.jtrangeo.2008.11.002.

Boss, D., Nelson, T., Winters, M., Ferster, C.J., 2018. Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. J. Transp. Health 9, 226–233. https://doi.org/10.1016/j.jth.2018.02.008.

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression. J. R. Sta. Soc. Ser. 47 (3), 431–443.

Buchele, M., 2019. Austin unveils design for a new bridge on the southern end of lady bird lake. Retrieved from KUT website: https://www.kut.org/post/austin-unveils-design- new bridge-southern-end-lady-bird-lakeCapMetro. In: (2017). *Get Ready to Ride*, Retrieved from https://www.capmetro.org/uploadedFiles/Capmetroorg/Schedules_and_Maps/get-ready-to-ride.pdf.

Cardozo, O.D., García-Palomares, J.C., Gutiérrez, J., 2012. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. Appl. Geogr. 34, 548–558. https://doi.org/10.1016/j.apgeog.2012.01.005.

Charlton, M., Fotheringham, A.S., 2009. Geographically Weighted Regression: White Paper. Retrieved from. http://www.geos.ed.ac.uk/~gisteac/fspat/gwr/arcgis_gwr/GWR_WhitePaper.pdf.

Chen, P., Zhou, J., Sun, F., 2017. Built environment determinants of bicycle volume: a longitudinal analysis. J. Transport Land Use 10 (1). https://doi.org/10.5198/jtlu. 2017.892.

City of Austin, 2014. 2014 Austin Bicycle Plan. Retrieved from. https://austintexas.gov/sites/default/files/files/2014_Austin_Bicycle_Master_Plan_Reduced_Size_.pdf.

City of Austin, 2017. Austin Texas Bike Map. Retrieved from. https://austintexas.gov/sites/default/files/files/Transportation/2017_Austin_Bike_Map_-_Side_2.pdf.

City of Austin, 2018a. The City of Austin Transportation Department 2018 Annual Report. Retrieved from. https://1d0d7cb9-bfde-4667-8eeb-20a5ee919bd6.filesusr.com/ugd/956239_455271f084ea4a4d9caf7bb098fb18ab.pdf.

City of Austin. 2018b. Longhorn Dam all public input. Retrieved from http://austintexas.gov/sites/default/files/files/Final_Longhorn_Dam_Data_Report.pdf.

City of Austin, 2018c. Safe Routes to School Infrastructure Report. Retrieved from. http://www.austintexas.gov/sites/default/files/files/Public_Works/2019_02_01_Austin_SRTS_District_10_Report_rev2.pdf.

City of Austin Planning and Zoning. 2020. Demographics. Retrieved from http://www.austintexas.gov/demographics.

Cromley, R.G., Hanink, D.M., 2014. Visualizing robust geographically weighted parameter estimates. Cartogr. Geogr. Inf. Sci. 41 (1), 100–110.

De Hartog, J.J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? Environ. Health Perspect. 118 (8), 1109–1116. https://doi.org/10.1289/ehp.0901747.

Dill, J., 2009. Bicycling for transportation and health: the role of infrastructure. J. Public Health Policy 30 (1), S95–S110. https://doi.org/10.1057/jphp.2008.56.

Eco-Counter. (2019). Products. Retrieved from https://www.eco-compteur.com/en/produits/multi-range/urban-multi/.

Ewing, R., Cervero, R., 2010. Travel and the built environment: a meta-analysis. J. Am. Plan. Assoc. 76 (3), 265–294. https://doi.org/10.1080/01944361003766766.

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. Int. J. Health Geogr. 14 (1), 12. https://doi.org/10.1186/s12942-015-0002-z.

Fotheringham, A.S., Brunsdon, C., Charlton, M.E., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester, England.

Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (MGWR). Ann. Am. Assoc. Geogr. 107 (6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480.

Gao, J., Li, S., 2010. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using geographically weighted regression. Appl. Geogr. 31, 292–302. https://doi.org/10.1016/j.apgeog. 2010.06.003.

Goodchild, M.F., Li, L., 2012. Assuring the quality of volunteered geographic information. Spatial Statistics 1, 110–120. https://doi.org/10.1016/j.spasta.2012.03.002.

Griffin, G.P., Jiao, J., 2015. Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. J. Transp. Health 2 (2), 238–247. https://doi.org/10.31235/osf.io/5gy3u.

Griswold, J., Medury, A., Schneider, R., 2011. Pilot models for estimating bicycle intersection volumes. Transp. Res. Rec. 2247, 1–7. https://doi.org/10.3141/2247-01.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using geographically weighted Poisson regression. Accid. Anal. Prev. 42 (2), 676–688. https://doi.org/10.1016/j.aap.2009.10.016.

Hankey, S., Lu, T., Mondschein, A., Buehler, R., 2017. Spatial models of active travel in small communities: merging the goals of traffic monitoring and direct-demand modeling. J. Transp. Health 7, 149–159. https://doi.org/10.1016/j.jth.2017.08.009.

Harris, P., 2019. A simulation study on specifying a regression model for spatial data: choosing between autocorrelation and heterogeneity effects. Geogr. Anal. 51 (2), 151–181.

Harris, R., Singleton, A., Grose, D., Brunsdon, C., Longley, P., 2010. Grid-enabling geographically weighted regression: a case study of participation in higher education in England. Trans. GIS 14 (1), 43–61.

Hasani, M., Jahangiri, A., Sener, I.N., Munira, S., Owens, J.M., Appleyard, B., ... Ghanipoor Machiani, S., 2019. Identifying high-risk intersections for walking and bicycling using multiple data sources in the city of San Diego. Journal of Advanced Transportation. https://doi.org/10.1155/2019/9072358.

Hedman, C., Elliott, D., Srini, T., Kooragayala, S., 2017. Austin and the state of low-and middle-income housing. Retrieved from Urban Institute website. https://www.urban.org/sites/default/files/publication/93781/austin_lmi_housing.pdf.

Hochmair, H.H., Bardin, E., Ahmouda, A., 2019. Estimating bicycle trip volume for Miami-Dade County from Strava tracking data. J. Transp. Geogr. 75, 58–69. https://doi.org/10.1016/j.jtrangeo.2019.01.013.

Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., Stefanidis, A., 2013. Assessing completeness and spatial error of features in volunteered geographic information. ISPRS Int. J. Geo Inf. 2 (2), 507–530. https://doi.org/10.3390/ ijgi2020507.

Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Injury Prevention 9 (3), 205–209. https://doi.org/10.1136/ip.9.3.205.

Jestico, B., Nelson, T., Winters, M., 2016. Mapping ridership using crowdsourced cycling data. J. Transp. Geogr. 52, 90–97. https://doi.org/10.1016/j.jtrangeo.2016.03.006.

Ji, Y., Ma, X., Yang, M., Jin, Y., Gao, L., 2018. Exploring spatially varying influences on metro-bikeshare transfer: a geographically weighted Poisson regression approach. Sustainability 10 (5), 1526.

Johnstone, D., Nordback, K., Lowry, M., 2017. *Collecting network-wide bicycle and pedestrian data: A guidebook for when and where to count* (Report No. WA-RD 875.1). Washington State Department of Transportation, Office of Research and Library Services, Seattle, WA.

Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. vol. 6. META Group Research Note, pp. 70.

Lee, K., Sener, I.N., 2019. Understanding potential exposure of bicyclists on roadways to traffic-related air pollution: findings from El Paso, Texas, using Strava metro data. Int. J. Environ. Res. Public Health 16 (3), 371. https://doi.org/10.3390/ijerph16030371.

Lee, K., Sener, I.N., 2020a. Strava Metro data for bicycle monitoring: a literature review. Transp. Rev. https://doi.org/10.1080/01441647.2020.1798558.

Lee, K., Sener, I.N., 2020b. Emerging data for pedestrian and bicycle monitoring: sources and applications. Trans. Res. Interdiscip. Perspectives. https://doi.org/10.1016/j.trip.2020.100095.

Leong, Y.-Y., Yue, J.C., 2017. A modification to geographically weighted regression. Int. J. Health Geogr. 16 (1), 11. https://doi.org/10.1186/s12942-017-0085-9.

Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using geographically weighted Poisson regression for county-level crash modeling in California. Saf. Sci. 58, 89–97. https://doi.org/10.1016/j.ssci.2013.04.005.

Mennis, J., 2006. Mapping the results of geographically weighted regression. Cartogr. J.

43 (2), 171–179.

Munira, S., Sener, I.N., 2017. Use of the Direct-Demand Modeling in Estimating Nonmotorized Activity: A Meta-Analysis. Texas A&M Transportation Institute, College Station, TX.

Murakami, D., Lu, B., Harris, P., Brunsdon, C., Charlton, M., Nakaya, T., Griffith, D.A., 2019. The importance of scale in spatially varying coefficient modeling. Ann. Am. Assoc. Geogr. 109 (1), 50–70.

Nakaya, T., 2012. GWR4: Windows Application for Geographically Weighted Regression Modelling. Retrieved from. http://gwr.maynoothuniversity.ie/gwr4-software/.

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically weighted Poisson regression for disease association mapping. Stat. Med. 24 (17), 2695–2717. https://doi.org/10.1002/sim.2129.

National Highway Traffic Safety Administration, 2008. National Survey of Bicyclist and Pedestrian Attitudes and Behavior—Volume I: Summary Report. U.S. Department of Transportation, Washington, DC.

Nordback, K., Marshall, W.E., Janson, B.N., Stolz, E., 2013. Estimating annual average daily bicyclists: error and accuracy. Transp. Res. Rec. 2339 (1), 90–97. https://doi.org/10.3141/2339-10.

Qin, X., Ivan, J., 2001. Estimating pedestrian exposure prediction model in rural areas. Transp. Res. Rec. 1773, 89–96. https://doi.org/10.3141/1773-11.

Ryus, P., Ferguson, E., Laustsen, K.M., Schneider, R.J., Proulx, F.R., Hull, R., Miranda-Moreno, L., 2014. NCHRP Report 797: Guidebook on Pedestrian and Bicycle Volume Data Collection. The National Academies Press, Washington, DC.

Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. Accid. Anal. Prev. 118, 166–177. https://doi.org/10.1016/j.aap.2018.02.014.

Sallis, J.F., Conway, T.L., Dillon, L.I., Frank, L.D., Adams, M.A., Cain, K.L., Saelens, B.E., 2013. Environmental and demographic correlates of bicycling. Prev. Med. 57 (5), 456–460. https://doi.org/10.1016/j.ypmed.2013.06.014.

Sanders, R.L., Frackelton, A., Gardner, S., Schneider, R., Hintze, M., 2017. “Ballpark” Method for Estimating Pedestrian & Bicyclist Exposure in Seattle: A Potential Option for Resource-Constrained Cities in an Age of Big Data. Paper presented at the Transportation Research Board 96th Annual Meeting, Washington, DC.

Schneider, R.J., Stefanich, J., 2015. Neighborhood characteristics that support bicycle commuting: analysis of the top 100 US census tracts. Transp. Res. Rec. 2520 (1), 41–51. https://doi.org/10.3141/2520-06.

Schnohr, P., Lange, P., Scharling, H., Jensen, J.S., 2006. Long-term physical activity in leisure time and mortality from coronary heart disease, stroke, respiratory diseases, and cancer: the Copenhagen City Heart Study. Eur. J. Cardiovasc. Prev. Rehabil. 13 (2), 173–179. https://doi.org/10.1097/01.hjr.0000198923.80555.b7.

Selala, M.K., Musakwa, W., 2016. The potential of Strava data to contribute in nonmotorized transport (Nmt) planning in Johannesburg. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLI-B2, 587–594.

Selby, B., Kockelman, K.M., 2013. Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. J. Transp. Geogr. 29, 24–32. https://doi.org/10.1016/j.jtrangeo.2012.12.009.

Shen, Y., Zhang, X., Zhao, J., 2018. Understanding the usage of dockless bike sharing in Singapore. Int. J. Sustain. Transp. 12 (9), 686–700.

Strauss, J., Miranda-Moreno, L., 2013. Spatial modeling of bicycle activity at signalized intersections. J. Transport Land Use 6 (2), 47–58. https://doi.org/10.5198/jtlu.v6i2.296.

Strava, 2019. Strava Releases 2019 Year in Sport Data Report [Web Log Post]. Retrieved from. https://blog.strava.com/press/strava-releases-2019-year-in-sport-data-report/. Sun, Y., Mobasheri, A., 2017. Utilizing crowdsourced data for studies of cycling and air pollution exposure: a case study using Strava data. Int. J. Environ. Res. Public Health 14 (3), 274. https://doi.org/10.3390/ijerph14030274.

Sun, Y., Du, Y., Wang, Y., Zhuang, L., 2017. Examining associations of environmental characteristics with recreational cycling behaviour by street-level Strava data. Int. J. Environ. Res. Public Health 14 (6), 644. https://doi.org/10.3390/ijerph14060644.

Tabeshian, M., Kattan, L., 2014. Modeling nonmotorized travel demand at intersections in Calgary, Canada: use of traffic counts and geographic information system data. Transp. Res. Rec. 2430, 38–46. https://doi.org/10.3141/2430-05.

Texas Department of Transportation (TxDOT), 2016. Texas Intersection Safety Implementation Plan: Preliminary Findings for Texas's Capital Area Metropolitan Planning Organization. Retrieved from. https://www.texasshsp.com/wp-content/uploads/2017/02/Preliminary-Findings_CAMPO_2016-06-15.pdf.

Turner, S., Benz, R., Hudson, J., Griffin, G., Lasley, P., Dadashova, B., Das, S., 2019. Improving the Amount and Availability of Pedestrian and Bicyclist Count Data in Texas. Texas A&M Transportation Institute, College Station, TX.

Wang, H., Wang, Y., Lowry, M.B., Chen, C., Pu, Z., 2016. Bicycle Safety Analysis: Crowdsourcing Bicycle Travel Data to Estimate Risk Exposure and Create Safety Performance Functions. US Department of Transportation Research and Innovative Technology Administration (RITA).

Wheeler, D.C., 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environ. Plan. A 39 (10), 2464–2481.

Winters, M., Brauer, M., Setton, E.M., Teschke, K., 2010. Built environment influences on healthy transportation choices: bicycling versus driving. J. Urban Health 87 (6), 969–993.

Xie, F., Levinson, D., 2010. How streetcars shaped suburbanization: a granger causality analysis of land use and transit in the twin cities. *Journal of Economic Geography*, 10(3), 453-470.Xu, P., & Huang, H. (2015). Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16–25. https://doi.org/10.1016/j.aap.2014.10.020.

Yang, H., Lu, X., Cherry, C., Liu, X., Li, Y., 2017. Spatial variations in active mode trip volume at intersections: a local analysis utilizing geographically weighted regression. J. Transp. Geogr. 64, 184–194. https://doi.org/10.1016/j.jtrangeo.2017.09.007.

Yu, H., Peng, Z.R., 2019. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. J. Transp. Geogr. 75, 147–163. https://doi.org/10.1016/j.jtrangeo.2019.01.004.

Zhang, D., Wang, X.C., 2014. Transit ridership estimation with network Kriging: a case study of second avenue Subway, NYC. J. Transp. Geogr. 41, 107–115. https://doi.org/10.1016/j.jtrangeo.2014.08.021.