Where does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus

Authors:

Greg P. Griffin *, Junfeng Jiao b

a Texas A&M Transportation Institute, Public Engagement Planning Program, 505 E. Huntland Drive, Suite 455, Austin, TX 78723
b The University of Texas at Austin, Community and Regional Planning, 310 Inner Campus Drive, Austin, TX 78712-1009, U.S. Tel.: +1 (512) 475-6158

*Correspondence to: Texas A&M Transportation Institute, 505 E. Huntland Drive, Suite 455, Austin, TX 78723, U.S. Tel.: +1 (512) 407-1111

E-mail addresses: g-griffin@rtimail.tamu.edu (G. Griffin), jjiao@austin.utexas.edu (J. Jiao)

Abstract:

Research on the role of bicycling for health through physical activity has been limited by the lack of information on where bicyclists ride. New big data sources available through smartphone-based applications provide a rich source to provide bicycle volume data more comparable to the scale of information available for automotive and public transit modes. In the case of smartphone apps for fitness tracking, results of this data can be used similar to the growing application of global positioning systems for automotive travel surveying. The authors evaluate data from Travis County, Texas for the purpose of determining where bicyclists ride, primarily for fitness purposes. Ride trip volumes are evaluated with residential and employment density, land use diversity, bicycle facilities and terrain to characterize places chosen for bicycling for health. Though limited to bicycle rides and routes voluntarily logged using the smartphone app, this method provides promise for applications in multi-modal transportation planning and health impact assessment studies.

Keywords: bicycling, health, crowdsourcing, VGI, GIS, built environment

NOTICE: this is the author’s version of a work that was accepted for publication in Journal of Transport & Health. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Journal of Transport & Health, [forthcoming (2015)] DOI: 10.1016/j.jth.2014.12.001
1. Introduction

Research on bicycle travel choices are beginning to connect volume data with routing information, recently bolstered by the use of global positioning systems (GPS). This study evaluates a burgeoning big data source of volunteered geographic information (VGI) through a popular smartphone application with the combination of land-based activities and complex of transport networks that Levinson and Krizek dubbed *Place and Plexus* (2008). They use the term plexus to refer “...to the complex of networks that connect people and places”, which is an important distinction from merely speaking of a given roadway segment or group of segments (2008, p. 1). In transport for health as in other purposes, a full complex of networks provides the combination of access and mobility that allow pursuit of goals. In the case of bicycling for health, this could include seeking routes for physical challenge and fun as much as personal safety (Buehler & Pucher, 2011; Lusk, Wen, & Zhou, 2014; Winters, Teschke, Grant, Setton, & Brauer, 2010). However, few studies have evaluated revealed bicycling behaviours with a large-scale data set.

This study aims to uncover characteristics of fitness-oriented bicycling route and context choices as recorded through the Strava smartphone application. The study also evaluates the use of passively-collected mobile application data as resource for transportation planning. Researchers use a combination of network-based roadway characteristics anticipated to attract bicycling for health, along with contextual attributes of density, diversity, and design that have been associated with active transportation, (Cervero & Kockelman, 1997; Ewing & Cervero, 2010).

Though the use of smartphone applications to evaluate mobility is relatively new, observational methods were popularized in the field of planning beginning in the 1960s (Gehl & Svarre, 2013; Jacobs, 1961; Whyte, 1980). The use of GPS provided a framework for digital analysis of personal paths on a new level of detail and breadth simultaneously, enabling analysis of bicycling beginning in Portland, Oregon in 2007 (Broach, Dill, & Gliebe, 2012; Dill, 2009; Hood, Sall, & Charlton, 2011; Hudson, Duthie, Yatinkumar, Larsen, & Meyer, 2012). Each of these studies specifically recruited participants, and used specialized equipment or software. This study is the first to passively collect GPS data for analysing bicycling trips, focusing on a sample dataset in Travis County, Texas.
1.1. Monitoring Bicycling as a Form of Monitoring Community Health

Though the relationship between bicycling as active transportation and positive health outcomes is well established (Calise, Dumith, DeJong, & Kohl, 2012; Edwards & Mason, 2014; Rojas-Rueda, de Nazelle, Tainio, & Nieuwenhuijsen, 2011; Sallis, Frank, Saelens, & Kraft, 2004), the use of individual monitoring methods to evaluate health impacts was first employed by Dill using GPS (2009). Most of the trips in Dill’s study were for transportation rather than fitness purposes, and 59% of the participants recorded at least 150 minutes of bicycling over a 7-day period. They averaged 16-21 kilometres per hour, equating to 6-8 metabolic equivalents (Dill, 2009, p. S101). Bicycling’s health benefits extend to physical and social health, as well (Underwood, Handy, Paterniti, & Lee, 2014), and planners have an important role in fostering health through transportation infrastructure and programmes (Cohen, Boniface, & Watkins, 2014).

Monitoring health at the level of a state or nation often incorporates aggregation of statistics on health outcomes directly from the medical community, but the behavioural risk factors more commonly employ residential surveys. The world’s largest on-going health survey system is the United States’ Behavioral Risk Factor Surveillance System (BRFSS), which includes a core set of questions from the Centers for Disease Control and Prevention, and states may add questions of local interest as well (U.S. Centers for Disease Control and Prevention, n.d.-a). The BRFSS includes several components of interest to physical activity researchers, such as obesity and diseases related to inactivity. The Texas edition includes a single question related to exercise: “During the past month, did you participate in any physical activities?” to which 80.2% of respondents in the region of this study, the Austin-Round Rock, TX Metropolitan Statistical Area, responded affirmatively (U.S. Centers for Disease Control and Prevention, n.d.-b). Surveys of travel provide additional insights on physical activity from bicycling and walking. The American Community Survey also asks a single question related to active travel: “How did this person usually get to work LAST WEEK?” (McKenzie, 2014), which obscures active commutes taken under half of the days in a workweek or active travel for any other purpose (Griffin, Nordback, Götschi, Stolz, & Kothuri, 2014). The National Household Travel Survey (NHTS) provides the most in-depth standard survey in the United States, asking questions related to the frequency of bicycling or walking for multiple trip purposes, trip distance, time, and other factors (Federal
Highway Administration, 2010). The 2012 National Survey of Bicyclist and Pedestrian Attitudes and Behavior provides the most comprehensive and recent nationwide study of active travel, including the extent to which respondents engaged in bicycling and walking outdoors; the availability and use of bike paths and lanes in the community; and changes in bicycling and pedestrian behavior and attitudes since 2002 (Schroeder & Wilbur, 2013). Several other valuable surveys and methods are reviewed in “The Colorado Mile Markers” report (Charlier Associates, Krizek, & Forsyth, 2012). Each of these surveys can be a source of population-level monitoring with substantial time gaps, but local methods need to be included to understand facility-level behaviors (Lindsey, Nordback, & Figliozzi, 2014).

Though regular bicycle traffic monitoring is not yet regularly practiced by most transport agencies, many are beginning to do so for multiple reasons, such as evaluating the sufficiency of infrastructure; to provide justification for funding allocation; to evaluate crash rates; to compare traffic volumes in various cities and other reasons (Luukkonen, 2011). Traditional means of monitoring bicycling include manual counts (Nordback, Marshall, & Janson, 2013), intercept and mail-back surveys (Forsyth, Krizek, & Agrawal, 2010), pneumatic tube counters similar to vehicle counters but calibrated for bicyclists (Hyde-Wright, Graham, & Nordback, 2014) and others (Griffin et al., 2014). Each of these methods has its advantages, but each is almost impossible to accomplish over a broad area simultaneously, which is why crowdsourced methods are gaining interest in planning.

1.2. Crowdsourcing Bicyclist Traffic and Participation in Planning

Recent advancements in information and communication technologie (ICT) have created new opportunities to study bicyclist behaviour and improve transportation plans with their contributions. San Francisco County first developed a smartphone application specifically for bicycle travel surveys (Hood et al., 2011), which was adopted by several others for use in research and bicycle transportation planning (e.g. Hudson et al. 2012; Duthie and Unnikrishnan 2014). Third-generation bicycle sharing systems employing electronic check-out systems provide a rich source for origin and destination data, in addition to volume flows (DeMaio, 2009). Several of these existing and emerging traffic volume methods are reviewed in a recent
Transportation Research Board Circular (Griffin et al., 2014), and there is at least one ongoing research project to advance crowdsourced bicycle traffic as of this writing (Figliozi, 2014).

For all its apparent advantages, researchers also offer cautions on the use of volunteered geographic information. First is the potential for the digital divide to create or foster inequities in planning (Sui, Goodchild, & Elwood, 2013). Any volunteered dataset by its nature only represents the input of the participants, and planners need to understand the representation of each dataset used, and its limits for a particular application, such as bicycle transportation planning and health. Also, each dataset can be expected to have different levels of individual privacy, and researchers and planners should identify personally identifiable information that can be gleaned from the data, and follow applicable protocols to protect human subjects (Goranson, Thihalolipavan, & Tada, 2013). Part of respect for humans and participants for planners also includes using data appropriately to provide equitable analysis for different populations and travel modes, and new datasets offer additional opportunities to more effectively plan for diverse, healthy and sustainable communities.

Various aspects of crowdsourcing and “Web 2.0” technologies including social media have been studied in transportation planning (Evans-Cowley & Griffin, 2012; Krykewycz, Pollard, Canzoneri, & He, 2011; Twitchen & Adams, 2011), and a few studies are emerging in the health and informatics fields (Kamel Boulos & Al-Shorbaji, 2014; Kamel Boulos et al., 2011; Kass-Hout & Alhinnawi, 2013). The transportation studies have primarily focused on increasing public involvement or crowdsourcing problem areas, rather than technologies for estimating bicycle volumes and locations. There are also several studies on the use of social media for advanced disease surveillance (Moorhead et al., 2013; e.g. Stoll, Quarrarone, & Torres-Urquidy, 2013), but few have yet to mention the prospect of using personal monitoring devices or apps for studies of large areas or populations. One such study posits “…Internet-connected, gamified fitness gadgets and apps are good examples of this ‘connected health/care’ vision realised through IoT [Internet of Things] technology” (Kamel Boulos & Al-Shorbaji, 2014). Rooksby et al. see many of these personal tracking methods as ‘lived informatics’ that may have utility not only for the individual, but also for activity research (Rooksby, Rost,
Morrison, & Chalmers, 2014). The present study employs a novel analysis of data aggregated from a smartphone app with more traditional geographic data, as described in the next section focusing on methods.

2. Integrating Volunteered Bicycling Data and the Built Environment

This study makes use of bicycle volume data based on an interconnected complex of networks, described as a “plexus”, and other attributes tied to block group areas denoted as “place”. Geographic information systems (GIS) provide an analytical environment for large, spatially referenced datasets, and contemporary statistical methods. Researchers used ArcGIS version 10.2 from ESRI, Inc., with its Geostatistical Analyst extension to develop the datasets and perform analysis in this study.

2.1. Strava Fitness App

Volunteered geographic data as a general source of information for studying transport and health are relatively new (Elwood, Goodchild, & Sui, 2011), and industrial innovation is developing novel techniques for applying gathered data to be of use to planners and researchers. A firm newly marketing their data for this purpose is Strava, Inc. (Albergotti, 2014), which offers a popular smartphone-based, GPS-supported fitness app. Users of the app contribute to the Strava dataset with every upload (Strava Inc., 2014a). The data product that aggregates the users’ rides is called Strava Metro, which the company describes can help “departments of transportation and city planners, as well as advocacy groups and corporations, [to] make informed and effective decisions when planning, maintaining, and upgrading cycling and pedestrian corridors” (Strava Inc., 2014b). A philosopher, Greenhalgh-Spencer, describes Strava as part of a new breed of technologies creating a hybrid social place in which “…runners and riders participate in a moment that is concurrently online and in material space” (2013, p. 320). Bicyclists using this app, whether for health-oriented riding, work commuting or other purposes, use speed, route and other information to record their own efforts (including calories burned), in addition to evaluating their performance against others in their community. Strava aggregates individual trips along roadway or trail segments, removing any personally identifiable information (Strava Inc., 2014a).

A glimpse of the bicycle route volumes recorded with the Strava app shown in Figure 1 indicates varying levels of bicycling volume, even though this dataset includes only one week from August 9-15, 2013
covering Travis County, contains the capital city of Texas, Austin. Bicyclists recorded 16,278 trips during the week, averaging 34 kilometres in length. Strava estimates 22% of these trips were commuting rides, which they record as trips beginning and ending at different locations (personal communication, Brian Riordan, February 19, 2014). The Strava trip purpose estimation method has not been independently evaluated as of this writing. Therefore, this study uses all trips from the sample dataset without distinguishing trip purpose.

The routes reflected in these results are visually associated with more densely populated areas, with several popular bicycle training routes reflected, such as the “Mansfield Dam Loop Ride” (Bicycle Sport Shop, 2010) along the state highways with wide shoulders west of Austin, Texas. Trips taken to the east reflect the wider variation of choices with a lot of rural, low traffic roads to choose from, without topographic barriers.
Figure 1. Bicycling volume in Travis County, Texas from Strava, August 9-15, 2013
As a fitness-oriented app, this dataset only captures its users, representing a subset of the population who fall in two exclusive categories of the population. First, approximately 57% of American adults use their cell phone to go online (Duggan & Smith, 2013), and second, the 20% or less of the population in Austin, Texas who are comfortable riding on streets, though most of them would only do so if bike lanes are present (Halff Associates Inc., 2014). Therefore, this dataset captures a quite small portion of the population as Strava users, but its coverage of streets is many times greater than the few monitored by the region’s bicycle and pedestrian counting program (Griffin, 2011). With 2,701 users in Austin as of August, 2013, this Strava dataset has a larger user base (n=2,701) than the participants in previous GPS-based bicycle travel surveys in the previous 5 GPS-based bicycling studies: Portland, Oregon - 164 cyclists (Broach et al., 2012), Austin, Texas - 317 cyclists (Hudson et al., 2012), Waterloo, Ontario – more than 400 cyclists (Casello & Usykov, 2014), San Francisco, California - 952 cyclists (Hood et al., 2011) and Zurich, Switzerland – 2,435 cyclists (Menghini, Carrasco, Schüssler, & Axhausen, 2010). It should be noted that the GPS surveys integrating trip purpose, demographic and other questions are not achievable through smartphone apps that do not integrate these specific questions. Figure 2 shows that Strava users do not represent the population as a whole. Research is lacking to determine demographic differences between Strava users and active bicyclists as a segment of the population, so how well the Strava data may represent the bicycling population cannot be determined at this time. The Strava users who reported their sex and age in the Travis County, Texas sample are heavily skewed toward the male sex, and most are between 25 and 54 years of age.

Figure 2. Travis County Population Estimate by age and sex at left (U.S. Census Bureau, 2012), Strava fitness app bicyclists at right (Strava Inc., 2014c)
Results from previous GPS-based surveys reveal relationships with very specific roadway factors that
can be more difficult to relate from other survey methods. Each of the previous studies reviewed includes
different variables related to route choice, but common attributes and their respective coefficient values are
summarized in Table 1. Different definitions used by the researchers are described in parentheses. Where
researchers published multiple model results, those included below were the results from the tables described
as most independent from other model interactions or appropriate for generalization.

Table 1. Bicycle route choice model summary

<table>
<thead>
<tr>
<th></th>
<th>Slope (definition from study)</th>
<th>Distance</th>
<th>Bicycle Paths (definition)</th>
<th>Bicycle Lanes (definition)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Broach et al., 2012)</td>
<td>-2.85 (proportion upslope 2–4%)</td>
<td>-5.22</td>
<td>1.57 (proportion bike paths)</td>
<td>1.81 (bridge w/ bike lane)</td>
</tr>
<tr>
<td>(Casello &amp; Usykov, 2014)</td>
<td>-1.4864 (positive gradient)</td>
<td>-0.1818</td>
<td>not included</td>
<td>4.3081</td>
</tr>
<tr>
<td>(Hood et al., 2011)</td>
<td>-0.50 (average up-slope)</td>
<td>-1.69</td>
<td>1.89 (proportion bike paths)</td>
<td>2.15</td>
</tr>
<tr>
<td>(Menghini et al., 2010)</td>
<td>-25.80 (maximum gradient)</td>
<td>-2.97</td>
<td>2.69</td>
<td>not included</td>
</tr>
</tbody>
</table>

Despite the fact that researchers may have used different definitions of the variable, model
estimation methods, and that each study was performed in different cities (2 European, 2 North American),
each of the signs indicate the same directions. Bicyclists tended to choose routes that minimized slope and
distance, but sought separation from vehicles on bicycle lanes or off-street paths.

Additional data such as direction of travel, time (and therefore speed), and other data provided in the
Strava product are of interest for further research, but are outside the scope of this study. Such a
crowdsourced dataset offers the distinction of gathering bicycle volume simultaneously, which would be
prohibitively expensive for most jurisdictions to consider using traditional methods. The bicyclist volume data
is attached to Open Street Map’s network using a patented alignment process (Kitchel & Riordan, 2013), but
this data lacks other roadway data useful for planning, such as the presence and width of bike lanes for
analysis.

2.2. Smart Location Database
A growing body of literature covering transportation and health integrates population-based statistics with built environment information affecting travel choices. As an active transportation mode, bicycling serves a public health function in addition to transportation (Garrard, Rissel, & Bauman, 2012; Rojas-Rueda et al., 2011). The U.S. Environmental Protection Agency developed the Smart Location Database to integrate many of the most valuable statistics for analysing urban planning issues at the Census block group level (Ramsey & Bell, 2014). In addition to demographic and employment statistics, the database compiles residential and employment density, land use diversity, design of the built environment, access to destinations, and distance to transit; each playing a role in transportation choices (Ewing & Cervero, 2010).

2.3. Connecting Networks and Places

Geostatistical analysis of demographic and built environmental variables requires the roadway network to be integrated in the Smart Location Database’s block group geography. The Strava Metro data product identifies bicyclist counts, which were multiplied by segment length to compute bicycle kilometres travelled (BKT). In one week of data collection in August 2013, Strava users logged 85,555 kilometres in Travis County, Texas.

2.4. Applying VGI for Health Behavioural Analysis

Ewing and Cervero (Ewing & Cervero, 2010) point out that the relationship between built environment variables and travel are complex. In many studies, variables considered as strong predictors of travel behaviour (such as density on transit) may not be expressed in a given context. Other results are confounded by self-selection bias. The relationship between these variables and bicycling have been lightly studied to date, and the authors are not aware of studies specifically addressing observed bicycling for health behaviours related to built environment variables.

Jennifer Dill’s early application of GPS travel studies for bicycling in Portland, Oregon provided the first large-scale analysis of revealed preferences through natural observation (Dill, 2009). Nearly all of the trips recorded identified a utilitarian purpose, rather than exercise. Though several bicyclist choice behaviours can be expected to be similar for bicycling trips, such as avoiding streets with lots of traffic and preferring routes with bicycle facilities, at least two exceptions to Dill’s findings are expected. In her study, utilitarian
cyclists avoid hills and seek to minimize the total distance on a given ride. Since duration and slope provide more physical challenge, these are anticipated to be sought by users of the Strava fitness app. Though the marketing focus of the Strava app is oriented towards fitness, it is likely that its users log trips for other purposes, as well.

Explanatory variables of bicycle routes from earlier studies were used to identify likely correlates with Strava trips, and these were extended to include additional built environmental measures in the Smart Location Database anticipated to play a role in bicycle route choice. Shoulder width has been shown to have a similar, but less pronounced attractiveness than bike lanes (Harkey & Stewart, 1997; Krizek & et al, 2006). State highways in the Travis County area often include very wide shoulders that are locally known as popular choices for large group rides. Each variable is calculated for 528 census block groups in Travis County, Texas.

Place-based variables impact behavioural choices directly and indirectly through a sense of place (Rosa, 2007). The first two variables are taken from the Smart Location Database, and characterize density and diversity; two of the 5 groups of “D” variables known to impact travel choices (Ewing & Cervero, 2010). Gross activity density is the number of employees and housing units on unprotected land (Ramsey & Bell, 2014). Regional diversity is the deviation of census block group ratio of jobs to population from the regional average ratio of jobs to population (Ramsey & Bell, 2014). The average percent slope was calculated form the United States Geological Survey’s National Elevation Database at a resolution of 30 meters, and characterizes terrain over the block group. Though negatively correlated in studies of bicycling for transportation (Broach et al., 2012; Casello & Usykov, 2014; Dill, 2009; Hood et al., 2011; Menghini et al., 2010), terrain could have the inverse relationship for fitness bicycling. This could be attractive to bicyclists for the added physical challenge, enjoyment of downhill descents, the aesthetic scenery or a combination thereof. The next section summarizes results of the regression analysis of the environmental variables and bicycle kilometres travelled.

3. Results

Travis County varies significantly in its density and diversity of land uses, terrain, and bicycle infrastructure. Table 2 includes the descriptive statistics from the variables used, including the network (plexus-based) attributes calculated as a percentage of all network lengths within each block group.
Table 2. Descriptive statistics of variables by block group

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Mean</th>
<th>Maximum</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle kilometres travelled (BKT) per kilometre of roadway</td>
<td>0</td>
<td>9.40</td>
<td>70.82</td>
<td>10.85</td>
</tr>
<tr>
<td>Place-based Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross activity density</td>
<td>0.02</td>
<td>7.26</td>
<td>208.51</td>
<td>12.30</td>
</tr>
<tr>
<td>Regional diversity</td>
<td>0</td>
<td>0.26</td>
<td>0.99</td>
<td>0.26</td>
</tr>
<tr>
<td>Average Percent Slope</td>
<td>0</td>
<td>3.84</td>
<td>20.00</td>
<td>3.30</td>
</tr>
<tr>
<td>Plexus-based Variables*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle lanes</td>
<td>0</td>
<td>8.42</td>
<td>100</td>
<td>12.12</td>
</tr>
<tr>
<td>Bicycle paths</td>
<td>0</td>
<td>2.79</td>
<td>71.68</td>
<td>8.35</td>
</tr>
<tr>
<td>Roadway shoulder (&gt;1.2m in width)</td>
<td>0</td>
<td>5.26</td>
<td>100</td>
<td>13.89</td>
</tr>
</tbody>
</table>

*Each of the plexus-based variables pertains to the roadway network, and is presented as a percent of roadway length within each block group.

At a global level, bicycle kilometres of travel can be explained by three place-based variables and three plexus-based variables depicted in Table 3. This grouping is separated to identify variables that may not be directly changeable from an engineering perspective (place-based), versus roadways that are within direct management of transportation authorities (plexus-based).
Table 3. Predictors of Bicycle Kilometres Travelled by Census Block Group

<table>
<thead>
<tr>
<th>Place-based Variables</th>
<th>Ordinary Least Squares</th>
<th>Geographically Weighted Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (upper, lower)</td>
<td>SE</td>
</tr>
<tr>
<td>Gross activity density</td>
<td>0.068877*</td>
<td>0.1373, 0.0044</td>
</tr>
<tr>
<td>Average Percent Slope</td>
<td>1.046875**</td>
<td>1.2889, 0.8049</td>
</tr>
</tbody>
</table>

| Plexus-based Variables                 |                        |                     |
| Bicycle lane length in meters          | 0.000956*               | 0.0028, -0.0009     | 0.000956, 0.000695 | 0.0029, -0.0015 |
| Bicycle path length in meters          | 0.001914**              | 0.0025, 0.0013      | 0.000307, 0.002081 | 0.0037, 0.0004  |
| Roadway shoulder (>1.2m wide) length in meters | 0.000162               | 0.0005, -0.0002    | 0.000159, 0.003828 | 0.0067, 0.0009 |

Probability: *p<0.05; **p<0.01

Ordinary least squares (OLS) regression of these six variables resulted in a multiple R-squared of 0.44501, indicating these six variables explain over 2/5 of the variation in bicycling volumes as evidenced through recorded trips using Strava. The low Variance Inflation Factor (VIF), near one for each variable, shows little redundancy between the explanatory variables. An accepted upper bound for the VIF is 7.5, so each variable was retained. The dependent variable, bicycle kilometres travelled (BKT) is significantly spatially autocorrelated. Moran’s index for BKT is 0.22, with a z-score of 22.08, (p=0.0000), indicating a less than 1% likelihood that its spatial clustering is random. When spatial autocorrelation is present, it can lead to misspecified models that do not account for this type of clustering, and geographically weighted regression (GWR) may help control for autocorrelation and nonstationarity (Kent & Capello, 2013; Miller, 2012). Since the spatial distribution of bicycle volumes demonstrates nonstationarity and spatial autocorrelation, and pass checks of regularity for global analysis, geographically weighted regression can be used to improve the model with localized coefficients.

3.1. Geographically Weighted Regression

Using the same variables, GWR produces a better fitting model by adjusting for local variation in the relationship between the dependent and explanatory variables. This sample of Strava data varies significantly
in density of data available over the study area. This situation merits use of an adaptive bandwidth based on the relative density of dependent variable data, rather than a fixed bandwidth distance to support the estimation of local parameters with adequate confidence (Nowrouzian & Inivasan, 2013, p. 111). GWR using an adaptive bandwidth approach considers the spatial relationship of each block group with its neighbour using spatial weights. The GWR model resulted in an AICc of 7,571, and an R squared value of 0.7877 (adjusted R squared 0.7458). Figure 3 shows the results of the local GWR coefficients. The more rural areas in the west and east ends of the county are highly correlated with bicycling activity, as shown in the density and diversity variables. The block groups with the greatest correlation with slope are expectantly, hilly. Urban bicycle lanes had relatively little impact on this sample’s Strava users, though roadway shoulders were highly correlated to the west, where there roadway connectivity is lower, and cyclists are forced to share roads with high-speed traffic. Bicycle paths were highly correlated in areas that they are available. The bands of low correlation are in areas without significant off-street paths as bicycling options.
Figure 3. Geographically Weighted Regression of Bicycle Kilometres Travelled with 3 Place and 3 Plexus Variables
4. Discussion

Results from both regression methods demonstrate that though the six explanatory variables provide moderately good correlation with predicted bicycle volume, this relationship varies significantly over the study area. Though more bicycle lanes are provided in and near downtown Austin, the map of correlation with gross activity density shows bicyclists using Strava did not necessarily seek out these areas. This could be related to a preference of fitness-oriented cyclists to seek roadways that are uninterrupted by stop lights. If a cyclist is seeking a route for the purpose of accomplishing training goals for distance, speed, or wattage, the traffic inherent in urban areas will tend to discourage riding in urban areas where bike lanes are most common. Therefore, this result should not be taken to mean that bike lanes are not desirable by fitness-oriented cyclists, but that they are generally not provided in more rural areas that more likely suit their training desires. This is evidenced by the correlations for roadways with shoulders in Figure 3, where a positive correlation is found throughout the county, except in the east, where many low-traffic roads, often in agricultural areas, are available for cycling. Regional diversity refers to the balance of jobs and housing, and the generally high correlation may be a tautological relationship rather than potentially causal, since most bicycling trips are likely to be near a place of residence or business.

The positive correlation of bicyclist volumes with hillier areas may have implications for the provision of bicycle infrastructure that contributes to public health. Though it may seem terrain is an uncontrollable variable, transport agencies can affect access to hillier areas for fitness and recreational riding with bicycle infrastructure, which could result in positive health outcomes over time if use and intensity are increased. Percent slope was the second highest coefficient in this study for predicting variation in bicyclist volumes, and this finding from Travis County, Texas could be analysed elsewhere to seek additional insights on this issue.

Crowdsourced forms of activity gathering offer promise as a method for filling information gaps in transportation planning, and for health impact assessments. Following methods similar to Rojas-Rueda et al.’s health impact assessment of Barcelona’s bikeshare program (2011), the increased automated digitization of activity patterns provide a ready-made data source for evaluating impacts and opportunities related to active
transportation. As described previously, the lack of bicycling volume data for estimating demand in planning studies may be partially remedied by these new data sources, but analysts are cautioned that the contributors to each source may not represent a broader population of interest.

4.1. Potential Implications of Personal Monitoring for Health

The impact of this type of monitoring could have far-reaching effects that include increased volume and spatial granularity of health-related information, and either real or perceived invasions of privacy with the long-term storage and availability of personal data. Though the data sold through this particular application is aggregated to a level that these authors cannot see how it might be mis-appropriated, the potential exists. Also, Rooksby and others suggest that the use of ‘lived informatics’ data will vary not only across individuals, but also over time for one person. Writing about physical activity tracking devices, they state, “people will pick and choose trackers, switch over time, and will have times when they track a lot and times when they track just a little. Consider collecting metadata on app use alongside activity data” (Rooksby et al., 2014). If data from commercially popular apps become increasingly sought for public health or transportation planning purposes, additional features to record metadata may be desirable to report precision and accuracy of the data.

4.2. Limitations

This crowdsourced dataset represents a larger population than previous GPS-based bicyclist surveys, and this type of dataset may be expected to increase in popularity over time. However, this user group is currently a small portion of any given area’s population, and the strong demographic bias towards young and middle-aged males needs to be considered. Additionally, Strava data is logged whenever the user chooses to begin a trip; this could be from the person’s residence, work, or she may have driven a vehicle to a remote bicycling location to begin a ride. Therefore, the data represents where rides are happening, but characteristics of the home or work communities of its riders are currently available only through inference. Though there are several limiting factors for the use of this data, it represents a relatively large volume of users and solves several challenges of previous methods of monitoring bicyclists.

5. Conclusions
Bicycling for health and fitness is not well understood from the spatial perspective, but this study provides insights on bicyclists revealed preferences for roads with bicycle lanes, shoulders, paths, challenging terrain, in and near populated places with businesses. Most of these characteristics are common with utilitarian bicycling, but Strava users, oriented towards bicycling for fitness, appear to be seeking steep hills for the challenge, rather than avoiding them for efficiency.

Even recently, Ewing and Cervero noted the “dearth of solid research” on several relationships between transport and the built environment, including the choice of bicycling as a travel mode (2010, p. 277). GPS travel survey methods including crowdsourced data sources such as Strava are beginning to bolster this research gap, though much remains to be done. The granularity and scalability of newer, bigger data helps overcome many of the obstacles common to traditional travel surveys, including response rates, memory and biases (Bricka, Zmud, Wolf, & Freedman, 2009; Shen & Stopher, 2014).

This new data source provides an opportunity to analyse bicycling routes and volumes at a larger scale than previously possible, and appears to be consistent with variables shown previously to be related to bicycling volumes. Though it only represents users of the Strava bicycling app, this represents a ready-made population that is larger than previous studies with active volunteers. As a ‘passive’ data collection method, this technique offers promise for analysis of spatial trends at a local level, and further research may demonstrate its applicability for comparing cities or states in terms of policy and infrastructure implementation. Researchers and planners are advised to seek ways to use volunteered geographic information in a manner that respects individuals and enhances the prospects for communities’ present and future health and quality of life.

Crowdsourced bicycle volumes provide an additional tool for empirical assessments of transport in planning and health, but big data and ‘smart city’ trends are not without critique. Wyly argues that the growth of privatized data used in social science is part of an “unstable, highly-capitalized automated epistemology” that has the potential to inappropriately replace direct observation in positive studies (Wyly, 2014, p. 686). This paper argues that the bicycle mode particularly has been under-observed, and additional resources to understand and improve transport and health are welcomed improvements. To observe as a positivist and a
humanist requires perspective, and researchers searching for better answers will have to grapple with the role of an expanding set of data choices in order to separate rigor from rhetoric.

Comparison of bicycle route choices from previous studies summarized in Table 1 with the present study of cyclists using the Strava app in Table 3 indicates that cyclists riding to track statistics for fitness may be less sensitive to bicycle-specific infrastructure. They may ride at a speed closer to that of the automotive traffic in urban areas, and they may choose routes that are less congested by vehicles in general as well. Findings from this study suggest that this dataset can be a useful in health assessments and transportation studies as a supplement, but not replacement, of more traditional data sources such as counts performed with pneumatic tubes.

5.1. Potential Applications and Future Work

This data and similar methods have limitations, but provide an immediately applicable information source for analysing and planning bicycle facilities and programmes. The representativeness of the Strava user base (or other data purveyors) is important for proper use of the data, considering its context of bicycling purpose, trip lengths, and other factors. Recognizing its limitations, this dataset can be used for many of the purposes traditional monitoring methods address, such as transport planning and health impact assessments, but additional research is needed to understand the differences between Strava users and all bicyclists in a given area. Surveying active users and comparing responses to the larger bicycling community and general population will provide further insights on the applicability of this dataset and others like it. Calibrating volunteered volume data with actual counts through manual or automated means, combined with an intercept survey will allow further understanding of the differences in trip behaviour characteristics between crowdsourced data participants and the total bicycling population in an area. Similar to Figliozzi et al.’s evaluation of bicycle counting durations (2014), future studies could use different durations of summarized bicycle volumes; a longer duration may provide stronger predictive power in areas with relatively low volumes. Private sector innovations provide an opportunity for improvements in the way health impact assessments and transport planning are performed, but rigorous analysis and objective perspectives are needed to improve the state of the practice.
Acknowledgements

The authors appreciate Brian Riordian of Strava Inc. providing early access to Strava Metro data used in this study, and the insightful and critical comments received on the original version of this article from its volunteer reviewers.
References


Halff Associates Inc. (2014). *City of Austin Urban Trails Master Plan (Working Draft for Review and Discussion).*


Ramsey, K., & Bell, A. (2014). *Smart Location Database*. Washington, D.C.


Strava Inc. (2014c). *Strava Metro Brief*. Hanover, NH.


