Crowdsourcing Bicycle Volumes: Exploring the Role of Volunteered Geographic Information and Established Monitoring Methods

Greg P. Griffin and Junfeng Jiao

Abstract: The recent interest in performance measures and new bicycle infrastructure development has triggered rapid advancements in monitoring methods for active transportation, but comprehensive monitoring programs for the bicycle mode are far from ubiquitous. This study evaluates the use of GPS survey data and a new crowdsourced volume dataset that may offer promise to extend the reach of limited counting programs. The authors integrated count data from five separate trail locations in Austin, Texas, with a previous survey using the CycleTracks smartphone application and a new data product derived from a larger-scale use of the Strava fitness application. New crowdsourced methods offer a prospect of expanding the relative time and geography of bicycle traffic monitoring, but do not currently offer many other attributes about trips obtainable from other methods. Further studies involving the combination of high-accuracy monitoring points with crowdsourced datasets may improve the efficiency of monitoring programs over large areas.

INTRODUCTION
Currently, bicycle volume data is not typically gathered by transportation agencies, despite the fact that several new methods to count bicycle volumes have been developed in recent years (Nordback and Janson 2010, Ryan and Lindsey 2013). Methods to incorporate nonmotorized volume data as a regular practice only were introduced to the Federal Highway Administration’s Traffic Monitoring Guide in 2013 (U.S. Department of Transportation Federal Highway Administration 2013), and the cost and perceived importance of counting this mode may continue to be a barrier to increased data development. This paper begins with a review of these challenges and opportunities reflected in the literature, and employs temporal and geographic analysis to review two community-contributed geographic datasets developed in the Austin, Texas, area, with actual count data from bicycle traffic counters on five locations of the downtown trail system. The method demonstrates that some data sources are likely to vary in the proportion of actual volumes.

This study addresses a need to evaluate current datasets that may extend the value of bicycle counting programs through volumes contributed by other means. Significant progress in institutionalizing bicycle and pedestrian monitoring has been made in Colorado, Minnesota, and Oregon, and agencies are searching for efficient solutions to fill this need (Lindsey, Nordback, and Figliozzi 2014; Nordback and Janson 2010; Nordback 2014). A recent Transportation Research Board circular summarized a number of recent and ongoing studies in the realm of active transportation monitoring methods and identified a need to continue research in crowdsourced volume data sources (Griffin, Nordback, Götschi, Stolz, and Kothuri 2014).

The concept of crowdsourcing information on improving bicycling conditions is not new, but the datasets and tools for their analysis are continuing to expand. Indeed, the Federal Highway Administration’s Bicycle Compatibility Index, developed in the late 1990s, is predicated on multiple users’ perspective of a video-based evaluation of roadways (Harkey, Reinfurt, and Knuiman 1998). Krykewycz et al. expanded the bicycle level of service concept with Web-based crowdsourcing, allowing a localized valuation of “bikeability” (Krykewycz, Pollard, Canzoneri, and He 2011). This very concept of asking for public input on the fundamental components of a decision-making process that is traditionally technically driven (e.g., roadway conditions, traffic volumes, etc.) is by itself a paradigm shift (Brabham 2009; Connors, Lei, and Kelly 2011; Elwood, Goodchild, and Sui 2013; Schlossberg and Brehm 2009; Schlossberg and Shuford 2005; Tulloch 2008). The idea of crowdsourcing may be merging with transportation practitioners’ notions of survey-based input, both for travel surveys and public engagement in planning (Akar and Clifton 2009, Shen and Stopher 2014).

Two principal research questions posed in this study are:
1. At a given location, what proportion of total bicycle volumes are reflected in a global positioning system (GPS) travel survey and a new crowdsourced dataset?
2. How do GPS surveys and crowdsourced methods vary in terms of trip purpose?

These questions are explored by corroborating trail count values from recording equipment, along with comparison of a smartphone-based GPS survey and a crowdsourced bicycle volume dataset.
DATA SOURCES

In addition to bicycle traffic counts recorded with pneumatic tubes, this study takes advantage of a previous study using the smartphone application CycleTracks (Sall n.d.) and a new crowdsourced data product called Strava Metro (Albergotti 2014). Recent availability of accurate trail counts for the Austin area, along with the recent CycleTracks study and new Strava Metro data allow unique comparisons.

Because all five trail locations include at least one week’s worth of data in 15-minute increments, evaluation of bicycle volumes by time of day and day of week is possible. The trail counts were gathered with Eco-counter pneumatic tubes, and the counts on the Lance Armstrong Bikeway were gathered using an automatic traffic recorder (ATR) developed by Eco-counter that uses inductive loops to count passing bicycles.

CycleTracks, a smartphone application developed by the San Francisco County Transportation Authority, has been employed and adapted for use in several other cities (Sall n.d.). It was applied in an Austin area study in 2011, and the current study incorporates its summarized traffic volumes (Hudson, Duthie, Yatinkumar, Larsen, and Meyer 2012). CycleTracks is essentially a GPS-based travel diary that gathers additional valuable information such as trip purpose and demographics that passive data collection techniques do not.

Strava is a popular smartphone-based, GPS-supported fitness application that records route volume information now sold as Strava Metro, which the creators describe as “... aggregate[ing] collective real world cycling preferences in real time with the purpose of enabling state and local departments of transportation as well as advocacy groups and corporations to make informed and effective choices when improving and repairing alternative transportation corridors” (Strava Inc. 2014b). Though this dataset includes a trip purpose attribute derived from a proprietary algorithm addressing factors such as whether a trip had a coincident origin and destination (personal communication, Brian Riordan, February 19, 2014), all trips were used in this assessment without analyzing trip purpose.

Each of these data sources provides a type of potential bicycle volume data, but they must be normalized in terms of the type of volume data to be comparable, and they must overlap spatially to remove spatial effects on traffic volumes. Table 1 describes the locations and times of the three separate data sources.

The crowdsourced dataset covers the entire roadway network in a given area, which is a major change from existing monitoring methods focused on individual points. Previously, bicycle volumes covering entire networks only could be estimated through modeling. Though a variety of cursory sketch-level and more sophisticated models have been developed, very few transportation agencies apply them, partially because of a lack of traffic counts (Griffin et al. 2014; Griffin 2009; Kuzmyak, Walters, Bradley, and Kockelman 2014). The crowdsourced dataset, depicting traffic volumes during the week of August 9 to 15, 2013, in Austin, Texas, is depicted in Figure 1. Strava users are bicycling on a variety of Austin roadways, but the highest volumes are found on state highways with wide shoulders in Austin’s western, hillier areas. A great deal of the other high-volume bicycling is distributed throughout the city, most having either bicycle lanes or low traffic volumes and speeds. As described later, this dataset is limited for it characterizes only Strava users, and taking a first glimpse at the differences between these crowdsourcing contributors and all bicyclists is a central question of this study.

While the bicycle traffic counts by definition include all bicyclists passing a given point, the survey and crowdsourced datasets are merely a sample of the entire population. Of those that reported a gender in the CycleTracks study, 30 percent were female, which is seven percent less than a national study of all bicyclists from 2002 (Hudson et al. 2012). Participants also skewed younger than the general population, though the study cannot reflect the population because it was targeted towards adults. Similarly, the users of Strava Metro are less than one-quarter female and skew younger than the general population.

Comparison of American Community Survey demographic estimates within Travis County, Texas, with the Strava Metro community in August of 2013 in the same county is depicted in Figure 2. Though the Strava community is strongly skewed male, a gender study of the included trails necessary for precise qualification was not part of this study.

Table 1. Bicycle volume sources

<table>
<thead>
<tr>
<th>Technology</th>
<th>Location(s)</th>
<th>Time</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic traffic recorder (counts)</td>
<td>Lance Armstrong Bikeway at Waller Creek</td>
<td>2011 to present (with minor breaks)</td>
<td>n/a</td>
</tr>
<tr>
<td>Pneumatic tube counters (counts)</td>
<td>Johnson Creek Trail Lady Bird Lake Trail Shoal Creek Trail</td>
<td>August 9 to 15, 2013</td>
<td>n/a</td>
</tr>
<tr>
<td>CycleTracks (GPS survey)</td>
<td>Austin, Texas area</td>
<td>May 1 to October 31, 2011</td>
<td>316</td>
</tr>
<tr>
<td>Strava Metro (crowdsourced)</td>
<td>Travis County, Texas</td>
<td>August 9 to 15, 2013</td>
<td>2,701</td>
</tr>
</tbody>
</table>
Figure 1. Bicycle volumes reported by Strava, August 9 to 15, 2013

Figure 2. Travis County population estimate by age and sex at left (U.S. Census Bureau 2012), Strava fitness application bicyclists at right (Kitchel and Riordan 2014)
STUDY AREA
Austin is used as a case study location because of the availability of the previous CycleTracks study, a newly available Strava Metro sample dataset, and bicycle volumes from portable pneumatic tube counters and two automatic traffic recorders (ATRs) positioned on the Lance Armstrong Bikeway at its crossings of Shoal Creek and Waller Creek (see Figure 3). The ATRs began recording bidirectional bicycle and pedestrian volumes in 15-minute increments in January of 2011; however, the one at the crossing of Shoal Creek has been removed because of construction in the area.

Downtown Austin, Texas, is a relatively dense, mixed-use district with an on-street and off-street network of bicycle facilities connecting a multitude of destinations, including restaurants, parks, and offices (Halff Associates Inc. 2014). Austin launched a bicycle sharing system on December 21, 2013, and has expanded to 40 stations as of spring of 2014 (B-Cycle 2014). The data used in this study all preceded the launch of this system, so this system was not a factor in the traffic volumes reviewed.

The off-street bicycle network includes a variety of materials, but most are at least three to four meters (10 to 13 feet) in width. The trails surrounding and connecting to Lady Bird Lake are predominantly surfaced with decomposed granite, and are more popular for running and walking than as specific bicycle routes. Construction on the Lance Armstrong Bikeway began in 2007, and the last section of its almost ten-kilometer (six-mile) downtown is a cycle track facility under development as of this writing.

Researchers chose five specific monitoring locations shown in Figure 3 because they found the city of Austin had recorded bicycle counts at these locations during the same duration as the available Strava sample dataset, enabling synchronous comparison. These locations happen to be located along each side of popular entrances to downtown, and were anticipated to provide options for a range of potential trip purposes. All five monitoring locations in this study are off-street, shared-use paths in downtown Austin, Texas, that are on facilities accessible to a wide range of bicycling skill levels, in areas that can serve business, shopping, and recreation destinations.

METHODOLOGY
The traffic counts, survey, and crowdsourced data were compiled in a geographic information system (GIS) to coordinate comparisons of values at the five study points.

Temporal analysis of the bicyclist counts reveals they have differing traffic patterns over the course of a day and throughout the week. Figure 4 illustrates bicyclist volumes by hour for a full week, with most sites exhibiting a pattern generally associated with utilitarian riding with stiff peaks during morning and evening commute times. Because these sites also are located within an urban context near live music and nightlife, there is

![Figure 3. Bicycle count locations](image-url)
also a distinct, but smaller peak at 2 a.m., when Austin bars are required to close by ordinance. Visual analysis of the graphed datasets were reviewed according to Turner and Lasley (2013), finding no significant erroneous data.

The Lance Armstrong Bikeway crossing Waller Creek has a unique pattern, as it provides several services for bicyclists in Austin. In addition to being a comfortable facility serving the downtown commuter rail station, it is the only crossing of Interstate 35 that is separated from vehicle traffic within the city. Therefore, it is likely that this facility is highly used for a wider variety of purposes than are similar paths.

The weekly spread of bicyclists also is associated with various trip purposes identified in the literature (Lindsey, Hankey, Wang, and Chen 2013; Roll 2013). Previous studies generally indicated volumes gradually increased during the week towards Friday, often with more traffic on weekend days. The fairly flat patterns in Figure 5 are associated with multiple trip purposes spread over the week. The strong increase of the Lady Bird...
Lake Trail and Shoal Creek Trail sites on Saturday are likely related to their settings within larger downtown parks and may reflect more emphasis on recreation.

Miranda-Moreno et al. developed classification factors for trip purpose based on the day of week and time of peak travel to estimate the predominant travel use of a given facility (2013). The following formulas offer comparability on trail traffic with recent work by Hankey, Lindsey, and Marshall (2014). The first formula calculates an index of weekday versus weekend traffic, noted as WWI. The second is the average morning commute period (7 to 9 a.m.) divided by the midday (11 a.m. to 1 p.m.) traffic, AMI.

This line of research logically assumes that sites with a higher ratio of weekend traffic versus weekday can reasonably be considered to have a predominantly recreational use. This temporal factor is further bolstered by examining the ratio of morning traffic to midday traffic, following that a higher AMI (relative morning trips versus midday) is associated with commute or utilitarian traffic. Because this study includes results from a GPS survey using the CycleTracks application, the five sites in Austin have actual trip purposes allocated to individual routes, which are reviewed in the next section by comparing WWI and AMI against actual recorded trip purposes.

To evaluate the overall relationship between the counted volumes and the CycleTracks survey and Strava Metro data, the final method applies ordinary least squares regression.

RESULTS

To compare the surveyed and crowdsourced traffic volumes with the total traffic volume, the researchers coordinated the datasets over time and location. The CycleTracks and Strava Metro datasets were collected over different time spans, May 1 to August 31, 2011, and August 9 to 15, 2013, respectively. The Lance Armstrong Bikeway was counted using the automatic traffic recorders during the CycleTracks period, and the remaining trails were counted by the city of Austin using pneumatic tubes during the Strava Metro dataset recording time. The average daily bicycle (ADB) volumes are several times higher on the Lance Armstrong Bikeway through downtown than on the trails nearest Lady Bird Lake.

In terms of proportions of the full traffic as reflected in the counts, Table 2 shows the CycleTracks survey included a very small proportion of the total volume, less than a tenth of one percent of the volume of the two traffic recorders on the Lance Armstrong Bikeway. As the original study in Austin (Hudson et al. 2012) and this evaluation shows, the advantage of a GPS survey is its balance of breadth of coverage with depth of information. However, the Strava Metro dataset includes a much larger proportion of the total bicycle traffic on the three trail counting sites that coincided with the Strava Metro data from August 9 to 15, 2013, including almost five percent of the total traffic.

Similar to Hankey, Lindsey, and Marshall (2014), this study has access to a small sample size of five locations with counts at least a week long, thus factoring trip purposes is limited. Shown in Table 3, results of the relative weekend versus weekday traffic (WWI) is not consistent with the index of morning to midday traffic (AMI). However, they all include midrange values, reflecting multiple uses of the trail system. This is consistent with the downtown location in parkland, but surrounded by employment, densifying residential areas, with retail and other uses.

At this stage of refinement of bicycle factor groups and trip purposes, no known studies of volume-based factor group classifications and user-surveyed trip purposes exist. A previous study by Hudson, et al. using the CycleTracks application (Hudson et al. 2012) and the new Strava Metro data product (Kitchel and Riordan 2014) allows comparison on the actual trail sites, though separated in time.

Table 3 summarizes results of the trip purpose comparison, though the trail counts, CycleTracks GPS survey, and Strava Metro crowdsourced data all vary in total counts and time periods. Though the CycleTracks data is the smallest sample in this study, it provides the only direct insight into trip purpose, for it asks participants to categorize each trip among one of six purposes: commute, errands, school, exercise, social, and other. Strava Metro includes a number of commute trips that is inferred through GIS-based rules such as the trip beginning and ending at locations more than a certain distance apart, though the full algorithm is proprietary (personal communication, Brian Riordan, February 19, 2014). Because Strava is a fitness-oriented application (Strava Inc. 2014a), this study allocates all trips not designated as commute purposes as recreational. In each of the four trip purpose categorization methods, the three trails in park settings had the highest proportion of recreational trips. This finding is compatible with the authors’ familiarity with each of the five sites.

Because the Lance Armstrong Bikeway is designed as a concrete facility with proper width and sightlines for all bicycling purposes, and it directly connects multiple destinations, it is likely used more for commuting and other utilitarian purposes, in addition to recreation. Its western end at Shoal Creek (closed of this writing for construction) is located at the edge of parkland reaching in to the western side of downtown. This corroborates its more recreation-oriented values in each of the methods, as compared with the Lance Armstrong Bikeway’s crossing of Waller Creek on the east side of downtown Austin.

The overall relationship of the five counted locations with the survey and crowdsourced data was evaluated using ordinary least squares regression and additional measures of spatial autocorrelation using OpenGeoDa software (Anselin, Syabri, and Kho 2006). Inclusion of both the CycleTracks and Strava Metro as independent variables resulted in a 0.95 R-squared (0.90 adjusted R-squared), but neither variable met statistical significance tests for a 95 percent confidence interval, likely because of the small number of sites involved. Spatial autocorrelation is indicated with a Moran’s I value of 3.49 and with an associated probability of 0.0005. Tests of both the Breusch-Pagan and Koenker-Bassett have insignificant probabilities at 0.6 and 0.5, respectively, indicating no significant heteroskedasticity.
These values are consistent with comparison of the values on the trails near each other in relatively similar urban contexts. Additional monitoring points would need to be included to draw additional conclusions based on regression methods.

CONCLUSION

Through in-depth analysis of three bicycle volume data sources at five locations in Austin, Texas, this study demonstrates that crowdsourced datasources, sometimes included in the “big data” discourse, show promise as a legitimate tool for evaluating volumes of bicyclists. The relative sparseness of bicycle volume counts performed restricts analysis of various performance measures addressing safety, health, air quality, livability, and other factors. Additional crowdsourced data sources and the increasing volume and opportunities for temporal comparison through time will indicate this study may serve an important role in the growing field of active transportation monitoring.

The growing literature on bicycle traffic monitoring has demonstrated solid methods for counting and evaluating volumes on a range of infrastructure types, but no studies have previously addressed the challenge of how to fill enormous gaps of unfunded data needs to acquire volume data on the majority of roadways. Figure 1 shows the extent of geographic coverage offered by this method, and this study begins to characterize the differences between users of a fitness-oriented application that crowdsources bicycle volumes and the larger population in

| Table 2. Counted GPS survey (CycleTracks), and crowdsourced (Strava Metro) bicycle volumes |
|-----------------------------------------------|------------------|------------------|------------------|------------------|
| May 1 to August 31, 2011                        | ADB | Count | CycleTracks | Percent of Count |
| L. Armstrong Bikeway at Waller Cr.             | 664 | 81707 | 9           | 0.0110            |
| L. Armstrong Bikeway at Shoal Cr.              | 824 | 101363| 102         | 0.1006            |
| Total                                         | 1488| 183070| 111         | 0.0606            |
| August 9 to 15, 2013                           | ADB | Count | Strava Metro | Percent of Count |
| Johnson Creek Trail                           | 141 | 990   | 87          | 8.7879            |
| Lady Bird Lake Trail                          | 231 | 1615  | 46          | 2.8483            |
| Shoal Creek Trail                             | 252 | 1767  | 76          | 4.3011            |
| Total                                         | 625 | 4372  | 209         | 4.7804            |

| Table 3. Volume-based factor group classifications and CycleTracks trip purpose |
|-----------------------------------------------|---------------------------------|---------------------------------|------------------|------------------|
| WWI (Classifica- | AMI (Classification) | CycleTracks Percent Recreational (# of total trips) | Strava Metro Percent Recreational (# of total trips) |
| L. Armstrong Bikeway at Waller Cr. | 0.935 | Mixed-utilitarian | 0.738 | Mixed- recreational | 13 (102) | 44 (163) |
| L. Armstrong Bikeway at Shoal Cr. | 0.919 | Mixed-utilitarian | 0.794 | Mixed- recreational | 80 (10) | 61 (216) |
| Johnson Creek Trail | 1.258 | Mixed-recreational | 1.401 | Mixed- utilitarian | 100 (1) | 70 (87) |
| Lady Bird Lake Trail | 1.582 | Mixed-recreational | 1.081 | Mixed- utilitarian | 100 (1) | 91 (46) |
| Shoal Creek Trail | 1.297 | Mixed-recreational | 1.921 | Utilitarian | 100 (3) | 62 (68) |
| Average | 1.198 | Mixed-utilitarian | 1.187 | Mixed- utilitarian | 23 (weighted avg.) | 60 (weighted avg.) |
an area. Crowdsourced methods offer prospect to help fill the dearth of bicycle volume data, and this study demonstrates the percentage of one proprietary source of data ranging between 2.8 percent and 8.8 percent of all users on trails in downtown Austin, Texas. This study’s methods could be replicated in other locations to determine locally consistent factors of crowdsourced data contributors. Considering that the only previously available method to estimate volumes on most roadways was implementation of sophisticated nonmotorized travel models, crowdsourced traffic volumes offer significant opportunities to improve large-area data used in planning for active transportation, though subject to significant limitations at this time.

In terms of time-based factoring and trip purposes, this study provides further evidence that the bicycle factor groupings developed by Miranda-Moreno et al. (2013) later extended to mixed-mode trail traffic by Hankey et al. (2014) generally are supported by the CycleTracks trip purpose responses. Researchers and practitioners are urged caution in the application of a potentially growing field of crowdsourced datasets, for they may or may not represent the given population that is considered in a transportation planning context. Furthermore, the performance measures facilitated by traffic volumesstanding alone should not be mistaken for a comprehensive and ethical planning process (Appleyard, Ferrell, Carroll, and Taack 2014).

Notwithstanding the limitations discussed in the following section, crowd-sourced so-called “big data” offers great promise but has to be put in its place. Richard Shearmur (2015) describes the issue succinctly, “however big the data, Big Data are not about society, but about users and markets” (p. 3). The authors are trying to highlight the discrepancy, cognizant that the data skewing generally matches the male orientation towards fitness cycling in the United States. There are other useful evaluations of the use of big data, but the novelty of this work lies in the triangulation of bicycle volumes with three data sources at the same place.

**LIMITATIONS AND FURTHER RESEARCH**

The number of available weeklong counts restricts the number of factor groups and, therefore, the accuracy of expansion of crowdsourced volumes. In addition, the relatively small number of trips recorded through the CycleTracks GPS survey does not meet statistical thresholds to represent the locations in this study, perhaps with the exception of the Lance Armstrong Bikeway at Waller Creek (n = 102). An intercept survey with a question on trip purpose such as that developed by Forsyth et al. (Forsyth, Agrawal, and Krizek 2012) should be implemented to further evaluate this study’s findings within other contexts. Acquisition of additional location site counts of at least one-week duration following recent guidance (Figlozi, Johnson, Monsere, and Nordback 2014; Nordback and Janson 2010; Turner and Lasley 2013; U.S. Department of Transportation Federal Highway Administration 2013) could allow spatial extrapolation of counted locations with crowdsourced datasets, yielding accuracy characteristics of focused counts along with the spatial extent of crowdsourcing.

**About the Authors**

**Greg Griffin** is a Ph.D. student of community and regional planning, and researches crowdsourcing methods and active transportation with the Public Engagement Planning Program of the Texas A&M Transportation Institute. He is a member of the American Institute of Certified Planners and a lifelong bicyclist.

**Corresponding Address:**
The University of Texas at Austin School of Architecture
310 Inner Campus Drive Austin, TX 78712
gregpgriffin@utexas.edu

**Dr. Junfeng Jiao** received three master degrees and a Ph.D. in architecture, geographic information systems, transportation engineering, and urban planning from China, Netherlands, and the United States. He is an assistant professor in the School of Architecture at the University of Texas, Austin, and interested in the applications of GIS, GPS, and Big Data in urban environments.

**Corresponding Address:**
The University of Texas at Austin School of Architecture
310 Inner Campus Drive Austin, TX 78712
Telephone: (512) 471-1922
Fax: (512) 471-0716
jjiao@austin.utexas.edu

**Acknowledgements**

The authors appreciate Aleksiina Chapman of the City of Austin Transportation Department for providing the trail bicycle counts and Brian Riordian of Strava Inc. for providing early access to the Strava Metro data used in this study. No external funding supported the study.

**References**


Halff Associates Inc. 2014. City of Austin urban trails master plan, working draft for review and discussion. Austin, TX: City of Austin.


Shearmur, R. 2015. Dazzled by data: Big data, the census and urban geography. Urban Geography, DOI: 10.1080/02723638.2015.1050922.


