UTCM

Improving the Quality of Life by Enhancing Mobility

University Transportation Center for Mobility™

DOT Grant No. DTRT06-G-0044

Using Smartphones to Collect Bicycle Travel Data in Texas

Final Report

Joan G. Hudson, Jennifer C. Duthie, Yatinkumar K. Rathod, Katie A. Larsen, Joel L. Meyer

Performing Organization
University Transportation Center for Mobility™
Texas Transportation Institute
The Texas A&M University System
College Station, TX

Sponsoring Agency
Department of Transportation
Research and Innovative Technology Administration
Washington, DC

Texas Transportation Institute

UTCM Project #11-35-69
August 2012

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UTCM 11-35-69		

4. Title and Subtitle	5. Report Date
Using Smartphones to Collect Bicycle Travel Data in Texas	August 8, 2012
	6. Performing Organization Code Texas Transportation Institute
7. Author(s) Joan G. Hudson, Jennifer C. Duthie, Yatinkumar K. Rathod, Katie A. Larsen, Joel L. Meyer	8. Performing Organization Report No. UTCM 11-35-69
9. Performing Organization Name and Address University Transportation Center for Mobility™ Texas Transportation Institute The Texas A&M University System 3135 TAMU College Station, TX 77843-3135	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTRT06-G-0044
12. Sponsoring Agency Name and Address Department of Transportation Research and Innovative Technology Administration 400 7th Street, SW Washington, DC 20590	13. Type of Report and Period Covered Final Report January 1, 2011–May 31, 2012 14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the US Department of Transportation, University Transportation Centers Program

16. Abstract

Researchers believed that if smartphones could prove to be an effective tool for collecting bicycle travel data, the information could be used for aiding decision making as to what types of facilities users prefer and guiding decisions about future facilities. If adequate facilities were provided, the mode share of bicyclists would increase and lead to a reduction in congestion. Thus, researchers used an existing smartphone application, CycleTracks, developed by the San Francisco County Transportation Authority, to develop this study. Austin-area bicyclists were targeted to test the application. Austin's strong cycling culture, its known bicycle friendliness, and the presence of several universities including the University of Texas made it an ideal test environment. Bicycle route data was collected between May 1 and October 31, 2011 during which time over 3,600 routes were recorded. About 300 bicyclists provided their age, gender, bicycling frequency, home zip code, work zip code, and school zip code. Eighty-three percent of these participants indicated that they bicycle daily or several times per week. Most participants live and work in the central area of Austin. Seventy percent of the participants were male and 30 percent female. There were slightly more participants in the 20-29 age range than the 30-39 and 40-49 age ranges. Many defined the purpose of the bicycle trip: 85 percent of the trips were for the purpose of transportation vs. recreation.

Using algorithms within ArcGIS, researchers were able to match almost 90 percent of the bicycle routes. The collected dataset provided a rich set of bicyclist and route attributes useful for identifying route choice decisions. Despite the manageable challenges of the data cleaning, network completion, and map-matching process, the amount of information provided by the use of CycleTracks far exceeds what would be available using other data collection methods.

17. Key Word	18. Distribution Statement
Data Collection, Bicycle Route Data, Bicycle Travel, Bicycling, Smartphones, Bikeways, Travel Data, Texas	Public distribution

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	80	n/a

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Using Smartphones to Collect Bicycle Travel Data in Texas

by

Joan G. Hudson, P.E.
Associate Research Engineer
Texas Transportation Institute
Texas A&M University

Jennifer C. Duthie, Ph.D.
Center for Transportation Research
The University of Texas at Austin

Yatinkumar K. Rathod, P.E.
Associate Transportation Researcher
Texas Transportation Institute
Texas A&M University

Katie A. Larsen, E.I.T.
Center for Transportation Research
The University of Texas at Austin

Joel L. Meyer
Center for Transportation Research
The University of Texas at Austin

Project Title:
Using Smartphones to Collect Bicycle Travel Data in Texas
UTCM Project 11-35-69

August 2012

Sponsored by the
University Transportation Center for Mobility™
Texas Transportation Institute
The Texas A&M University System
College Station, Texas 77843-3135

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use.

Acknowledgments

Support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers Program to the University Transportation Center for Mobility™ (DTRT06-G-0044) with additional funding from the Austin District of the Texas Department of Transportation and in-kind services provided by the City of Austin. Researchers would like to express appreciation to Evan B. at the Environmental Systems Research Institute (ESRI) for his assistance with the map matching. Many people within the community assisted researchers in spreading the word to bicyclists about the project to encourage them to download the application and use it when riding, and for this help the researchers are grateful. Deserving particular appreciation are the following groups: the Austin Cycling Association, the City of Austin Bicycle Program, and the Mamma Jamma Ride organizers.

By far, the organization researchers owe the most gratitude is the San Francisco County Transportation Authority. Both Billy Charlton and Lisa Zorn were instrumental in this research project, both by housing the data on their server and ensuring access to the data.

Table of Contents

List of Figures

NOTE: Color figures in this report may not be legible if printed in black and white. A color PDF copy of this report may be accessed via the UTCM website at http://utcm.tamu.edu or on the Transportation Research Board's TRID database at http://trid.trb.org.

List of Tables

Executive Summary

As agencies look for ways to gather critical data surrounding bicycling, they often seek inexpensive and efficient means to understand where people are riding so that limited dollars are spent wisely. Smartphones, which many adults carry, are one way to collect bicycle route data. For this project, researchers evaluated a smartphone application developed by the San Francisco County Transportation Authority (SFCTA). Called CycleTracks, the application is available on both iPhones and Android-based smartphones. Using Austin as the case study, researchers collected bicycle route data during a six month period between May 1 and October 31, 2011. Over 3,600 routes were recorded in Austin and stored on the SFCTA servers. Researchers retrieved the global positioning system (GPS) location data from the servers, cleaned the data, entered missing links into the street network, and tested several methods of mapping the data. An important feature of the application is the ability to gather information about the bicyclist and the purpose of the trip. Participants were asked but not required to enter demographic information when downloading the application. About 300 bicyclists provided their age, gender, bicycling frequency, home zip code, work zip code, and school zip code. Following a trip, users were given the opportunity to define the purpose of the bicycle trip. The collected dataset provided a rich set of bicyclist and route attributes useful for identifying route choice decisions.

About 300 bicyclists provided their age, gender, bicycling frequency, home zip code, work zip code, and school zip code. A very high percentage (83 percent) of these participants indicated that they bicycle at least several times per week. Most participants live and work in the central area of Austin. Seventy percent of the participants were male and 30 percent were female. The highest percentage of participants was 20-29 years old. The majority of participants rode a bicycle at least several times per week and 70 percent were male. Many defined the purpose of the bicycle trip: 85 percent of the trips were for the purpose of transportation vs. recreation.

Using algorithms within ArcGIS, researchers were able to match almost 90 percent of the bicycle routes. The collected dataset provided a rich set of bicyclist and route attributes useful for identifying route choice decisions. Despite the manageable challenges of the data cleaning, network completion, and map-matching process, the amount of information provided by the use of CycleTracks far exceeds what would be available using other data collection methods.

This report summarizes the many processes employed as part of this study, from marketing to data analysis. Detailed descriptions about using ArcGIS and other methods are included along with advantages and disadvantages of the various approaches. Recommendations are provided for communities looking to utilize smartphone applications for data collection. Descriptions of the participants and an understanding of who is riding where can answer key questions for a region considering an investment in bicycling infrastructure, education, and encouragement. It is the hope of the researchers that communities will utilize the information provided here to expand the discussion and implement programs for furthering bicycle accommodations and safety. Having data to guide the development and evaluation of programs and projects is a critical step in understanding the successes and opportunities. The smartphone is a useful data collection tool that should be considered when deliberating inexpensive ways to gather critical information.

Introduction

At an increasing rate, smartphones are being used to collect travel data across various modes. A GPS sensor that quickly determines the latitude and longitude of the user's location is the most important smartphone technology for understanding speed, route, and location-specific details. The City of Boston, Massachusetts, encourages citizens to help document where potholes are located using a smartphone application called Street Bump, which employs the phone's accelerometer to detect "when a bump is hit while the GPS determines location" (Schwartz, 2012). Another smartphone application called PaceLogger was developed as a way to gather trip data that adds value to the traditional household travel survey trip data (Rodriguez, 2012).

As smartphone technology is improving, so is its popularity. The use of smartphones by the U.S. population has been tracked by the Pew Research Center over the years. As of February 2012, nearly half (46 percent) of all American adults owned a smartphone, which is an 11 percent increase from May of 2011. Considering that 41 percent of all American adults own cell phones (not smartphones), there are now more smartphone owners than basic mobile phone owners. Only 12 percent of adults do not own a cell phone. Figure 1 shows the results of the recent research effort: 60 percent of all college graduates, 67 percent of 18-35 year olds, and 68 percent of those with an annual household income of $75,000 or more have a smartphone. Almost half (49 percent) of African-Americans and half (49 percent) of Latinos are smartphone owners, while 45 percent of Caucasians own smartphones. Most of the smartphone owners use an Android (20 percent) or iPhone (19 percent) device, while a smaller percentage (6 percent) own a Blackberry (Smith, 2012).

Answering the question of where people bike and for what reasons motivated this research project. Researchers chose to test the method of tracking cyclists who volunteered to download an application called CycleTracks, which is specifically designed for smartphones with GPS technology. As communities seek ways to combat obesity, congestion, and pollution, they look to increase the number of people cycling. As a result, agencies seek to do more to improve safety and accommodations for these vulnerable transportation system users. The U.S. Department of Transportation (USDOT) supports these sentiments through its Livability Initiative, which provides policy guidance and improved funding mechanisms for livability concepts including bicycle and pedestrian transportation improvements (USDOT, 2011). Collecting data on cycling route choices and analyzing those choices with attributes of the routes, demographic information, and other user data helps inform decisions about what infrastructure improvements to make to support cycling.

However, there is very little data on levels of bicycle use and preferred facilities. Very few transportation agencies collect widespread data on bicycle travel. Transportation agencies that collect cycling route preference data typically use either stated preference surveys or revealed preference methods. Intercept-survey and count stations can be expensive and time consuming. To address the need for data, one agency in California turned to smartphones for help. Their smartphone application has the benefit of offering an inexpensive and dynamic alternative to traditional bicycle data collection. Researchers tested this product in Texas using the Austin area as a case study.

Smartphone ownership within demographic groups, 2011-2012

% of adults within each group who own a smartphone (indicates statistically significant difference between 2011 and 2012):*

	May 2011	February 2012	Change
All adults	**35%**	**46%**	**+11***
Gender			
Men	39	49	+10*
Women	31	44	+13*
Age			
18-24	49	67	+18*
25-34	58	71	+13*
35-44	44	54	+10*
45-54	28	44	+16*
55-64	22	31	+9*
65+	11	13	+2
Race/Ethnicity			
White, non-Hispanic	30	45	+15*
Black, non-Hispanic	44	49	+5
Hispanic	44	49	+5
Household Income			
Less than $30,000	22	34	+12*
$30,000-$49,999	40	46	+6
$50,000-$74,999	38	49	+11*
$75,000+	59	68	+9*
Education level			
Less than High School	18	25	+7
High School Grad	27	39	+12*
Some College	38	52	+14*
College+	48	60	+12*
Geography			
Urban	38	50	+12*
Suburban	38	46	+8*
Rural	21	34	+13*

Source: Pew Research Center's Internet & American Life Project April 26-May 22, 2011 and January 20-February 19, 2012 tracking surveys. For 2011 data, n=2,277 adults ages 18 and older, including 755 interviews conducted on respondent's cell phone. For 2012 data, n=2,253 adults and survey includes 901 cell phone interviews. Both 2011 and 2012 data include Spanish-language interviews.

Figure 1. Smartphone Ownership within Demographic Groups

This project used an innovative revealed preference method that gathers data from volunteer cyclists using the CycleTracks smartphone application developed for SFCTA that tracks cyclists' location on a bike trip using GPS technology embedded within their personal smartphones (specifically, iPhones and Android phones). The inventive use of existing technology already owned and carried by the cyclists makes this a low-cost approach for gathering cycling route choice data.

By understanding the routes, trip purpose, and demographic information of bicyclists, planners and engineers can prioritize projects, plan new bicycle accommodations, understand route choices, and better address the needs of this mode of non-motorized travel. With better accommodations, more people are expected to bicycle, which will remove motor vehicles from the roadways, reduce congestion, and improve overall health.

This research is not the first to collect and analyze bicyclist route information. Previous research has used both stated preference surveys and revealed preference methods, including distributing intercept surveys, conducting counts, asking participants to draw their route on paper or on a website, and distributing global positioning system units to participants.

This report summarizes the outcome of the pilot CycleTracks data collection program implemented in the Austin, Texas, area. Following a review of the literature on stated and revealed preference methods for understanding bike route choices and bicycle level of service concepts, this report explains the details of the CycleTracks application, volunteer recruitment methods, data preparation, and data analysis conducted for this pilot program.

Literature Review

Stated Preference Methods

Stated preference surveys gather route information by asking bicyclists to pick one of several route choices that they would prefer. Stinson and Bhat (2003) conducted a nationwide Internet survey with over 3,000 respondents, finding travel time to be the most important factor for bicyclists when choosing a route. Respondents' age and residential location had effects on the results, but income had no effect. Krizek (2006) described a study in the Twin Cities where bicyclists were shown 10-second video clips of a facility, taken from a bicyclist's perspective. The survey, which was administered on a computer, was adaptive in terms of the choice set based on a respondent's previous responses. The researchers' goal was to determine the tradeoffs each respondent was willing to make between travel time and facility quality. Facilities with bicycle lanes were found to most likely convince a bicyclist to choose a longer travel time path. Neither income nor gender was statistically significant, again lending credence to the use of the smartphone for data collection possibly biased toward higher-income individuals.

Tilahun et al. (2007) studied the tradeoffs bicyclists make when choosing routes, concluding that bike lanes are what are most desired for both respondents who do bicycle and those who do not. The researchers found higher-income households to be more likely than lower-income households to choose the better facility, all else constant. Hunt and Abraham (2007) studied the survey responses of 1,128 participants in Edmonton, Canada. Key findings indicated that the sensitivity to trip time varied with facility type and experience level, and secure parking and showers had a positive impact on the attractiveness of bicycling. The researchers did not find significant differences in behavior across the different segments of bicyclists. Sener et al. (2009) found motor vehicle traffic volume to be a significant factor in route choice as well as travel time for commuting bicyclists.

Stated preference surveys have the advantage of allowing for efficient data collection since the same (or similar) hypothetical scenarios can be presented to each respondent. However, problems with stated preference surveys are well documented (e.g., Bradley, 1988), namely that there is not always a direct correspondence between one's stated preference and his or her actual (revealed) preference.

Revealed Preference Methods

Researchers can find out travelers' actual (revealed) preferences through several methods. A common technique is intercept surveys (e.g., Shafizadeh and Niemeier, 1997; Howard and Burns, 2001; and Krizek et al., 2007), whereby researchers stop travelers who are passing by a certain point and ask them to fill out a survey.

Shafizadeh and Niemeier (1997) analyzed data from a mail-back intercept survey in Seattle, Washington. Higher-income bicyclists were found to have longer commute travel times than lower-income bicyclists, but this effect was reversed for suburb-to-suburb commuters. Intercept surveys can be an effective method for getting a representative sample of the population of travelers that pass by these points; however, they can be costly.

Aultman-Hall et al. (1997) analyzed 397 routes used by commuter cyclists in Guelph, Ontario, and saw an average deviation of 0.25 mi but found no clear relationship between shortest-path distance and percent route deviation. Guelph had no bicycle lanes at the time the study was conducted, but it did have an extensive network of off-street paths and trails. Hyodo et al. (2000) used data from a bicycle trip survey to estimate a route choice model nested within a destination choice model, where the potential destinations were railway stations in Japan. In the route choice model, bicyclists sought a perceived shortest path, assuming that links with desirable characteristics (e.g., with a bicycle-only lane) are perceived as shorter than they actually are. The station choice model considered only perceived distance to the station, whether the bicycle parking is on the same side of the station as the traveler's origin, and station-specific variables. Howard and Burns (2001) received 150 responses from bicyclists asked to report the route they took on their most recent commute to work. The researchers then compared the route to alternative routes optimized for safety (using the method developed by Sorton and Walsh [1994]), distance, and time. Average actual route length was 10 percent greater than the length of the shortest route. Of the three alternatives, actual routes most closely resembled shortest-distance paths. Winters et al. (2010) surveyed 174 bicyclists in the Vancouver, Canada region and asked for a description of a typical route for a specific trip. The researchers found 75 percent of trips in their sample within 10 percent of the shortest-path distance. Actual routes had, on average, more traffic-calming features, markings, and signs than the shortest-distance route.

More recently, electronic data collection methods have been used to reduce the user error in reporting route information. One method used a web survey that attracted over 1,000 participants to mark their top three most-used bicycle routes using Google Maps as part of a Texas Transportation Institute (TTI) study for the Austin District of the Texas Department of Transportation (TxDOT; Hudson et al., 2011). Another method employed the distribution of GPS devices to bicyclists, and the devices tracked the exact routes taken. Researchers were able to control the sample that was selected for the study, but one downside of this approach was the high cost of the units. Using this same method, Menghini et al. (2009) analyzed 73,493 bicycle trips made in Zurich, Switzerland, collected via GPS units. Shortest-path alternative routes were generated for the purpose of building a route choice model. Bicyclists were found to prefer direct and marked routes and to avoid steep gradients and signal-controlled intersections. Broach et al. (2011) used multi-day revealed preference data from 164 bicyclists in Portland, Oregon (collected electronically by Dill and Gliebe, 2008) to formulate a model that estimated the influence key variables have on route choice. Study participants were slightly older, had higher incomes, and were more likely to be female than the population of bicyclists included in a larger phone survey. Key results showed that commuters were more sensitive to distance and less sensitive to other variables (e.g., volume); bike boulevards and off-street paths were preferable to bike lanes; and half of all trips were less than 10 percent longer than the shortest-path distance.

The newest method of electronic data collection does not have the same high cost as GPS unit-based studies because it uses technology owned by the user—smartphones. Smartphones (e.g., iPhones, Android-based phones, Blackberries) are GPS enabled and allow users to access applications such as those developed to track bicycle routes. Another advantage of using smartphones to collect route data is that data collection can be ongoing for long periods since there is no cost for the collection of data. All of the cost is in building the application, marketing the use of the application, and analyzing the data. The downside is the response bias—likely toward avid and higher-income riders—that may occur. The first such smartphone application, CycleTracks, was developed by SFCTA and was the one used in this study.

The CycleTracks application is described in Charlton et al. (2010). Hood (2010) used the data collected by SFCTA to estimate a route choice model that quantified user preference for specific facility characteristics based on their demographic information. After data cleaning, the researchers obtained 2,777 routes uploaded by 366 users. A comparison with the information obtained from bicyclists in the 2000 Bay Area Travel Survey revealed no significant difference in mean age but a significantly lower proportion of females in the smartphone sample. The researchers hypothesized that route choice was not very sensitive to demographic characteristics. Key results indicate that 1) bicyclists prefer routes with fewer turns, 2) women and commuters prefer to avoid hills, and 3) infrequent bicyclists have a stronger preference for bicycle lanes.

Other entities have adapted the CycleTracks app to study travel patterns. AggieTrack was deployed by Texas A&M University (TAMU) and the Bryan-College Station Metropolitan Planning Organization (MPO) in Spring 2011 to analyze the travel patterns of TAMU students. Also, PTV NuStats developed the RouteScout app for the purposes of collecting travel data and the RideTrack app for transit on-board surveys. The Singapore-MIT Alliance for Research and Technology has developed a smartphone-based travel survey method with a pilot that took place in spring of 2012 and full implementation scheduled for summer and fall of 2012. Resource Systems Group is also working on a method for using smartphones for travel surveys and building origin-destination trip tables.

Bicycling Level of Service

Related to the collection of route choice data is the determination of level of service or suitability for bicycling facilities. Level of service (LOS) can be found for existing facilities and can provide goals (e.g., LOS “A”) when designing new facilities. Turner et al. (1997) and Moudon and Lee (2003) both contain thorough reviews of the literature in this area, so the review here will focus on the most relevant literature. To date, level of service has been determined based on user perceptions rather than actual routes taken. Harkey et al. (1998) developed a bicycle compatibility index (BCI) as part of a study sponsored by the Federal Highway Administration. Two hundred and two participants—ranging in age, experience level, and gender—watched videos of 67 locations and were asked to rank each site four times (based on volume, width, speed, and overall) on a six-point scale to express their comfort level. The following factors were found to be significant in predicting mean comfort level: presence of a bicycle lane or paved shoulder, bicycle lane width, curb lane width, curb lane volume, other lane(s) volume, 85th percentile speed, presence of a parking lane with more than 30 percent occupancy, and presence of residential roadside development. The comfort level ratings were then mapped to level of service designations based on the percent of riders who would use a given facility (e.g., level of service

“A” is equated with comfort level 1.5, the level at which 95 percent of riders would use the facility). Landis et al. (2003) guided approximately 60 bicyclists, selected to represent a cross section of the bicycling population, through 18 signalized intersections. Immediately after riding through the intersection, bicyclists graded it from “A” (most safe) to “F” (most unsafe). These data were then regressed on characteristics of the intersections to find the statistically significant factors, which were found to be motor vehicle volume per lane crossing the intersection, outside lane width, and intersection crossing distance. Landis et al. (1997) conducted a similar study for roadway segments.

The CycleTracks Smartphone Application

The CycleTracks application is free, quick to download and install, and easy to operate, with minimal user interaction needed to start logging trip data. It uploads the recorded trip immediately to the SFCTA server and is carefully designed not to drain the battery completely while in use. In addition to recording and sending trip data to SFCTA servers (see Appendix A for details on data that is stored), the CycleTracks app allows users to view maps of all the trips they have recorded and track their distance and average speed of each trip.

The user experience of the application is designed to be as unobtrusive as possible. The first time users start the application, it asks for some optional information to know about their cycling habits, along with some personal information that is also optional and does not limit the functionality of the application. The application asks for age, gender, cycling frequency, and home, work, and school zip codes, as well as an email address if the user is interested in hearing about similar projects in the future.

The first screen after the initial information screen shows a Start Trip button; tapping once starts the trip recording. The Recording screen shows parameters such as distance, time elapsed, current speed, and average speed. If the application is not able to obtain a GPS signal during the time it is running, when the user clicks the Finish button, nothing will be submitted to the SFCTA server. If the application is able to get a GPS fix, it will start recording the trip information, and when the user finishes the bike trip, he or she needs to tap the Finish button, choose a trip purpose from eight options (commute, work-related, school, social, shopping, errands, exercise, other), and then tap the Save button to upload the route to the SFCTA server. The user also has the option of discarding the trip instead of saving it.

The GPS data are saved locally on the device throughout the trip and only uploaded at completion, so a live data connection is not required during recording of the trip. Any trips that do not successfully upload for any reason are marked with an exclamation point and can be re-uploaded later. The user will be prompted to upload those trips later, or they will be automatically uploaded the next time a trip is recorded. Users can view all their saved trips as a Google Map overlay by tapping the View Saved Trips button. CycleTracks also has a reminder feature built in that, after 15 minutes, reminds the user at five-minute intervals that the application is still collecting data to make sure the user is aware of the application running in the background. The application also turns itself off if the battery life of the phone gets below 10 percent of its capacity to preserve the battery for emergency use.

Figure 2 presents snapshots of the various screens for the CycleTracks application for the iPhone and Android smartphone devices.

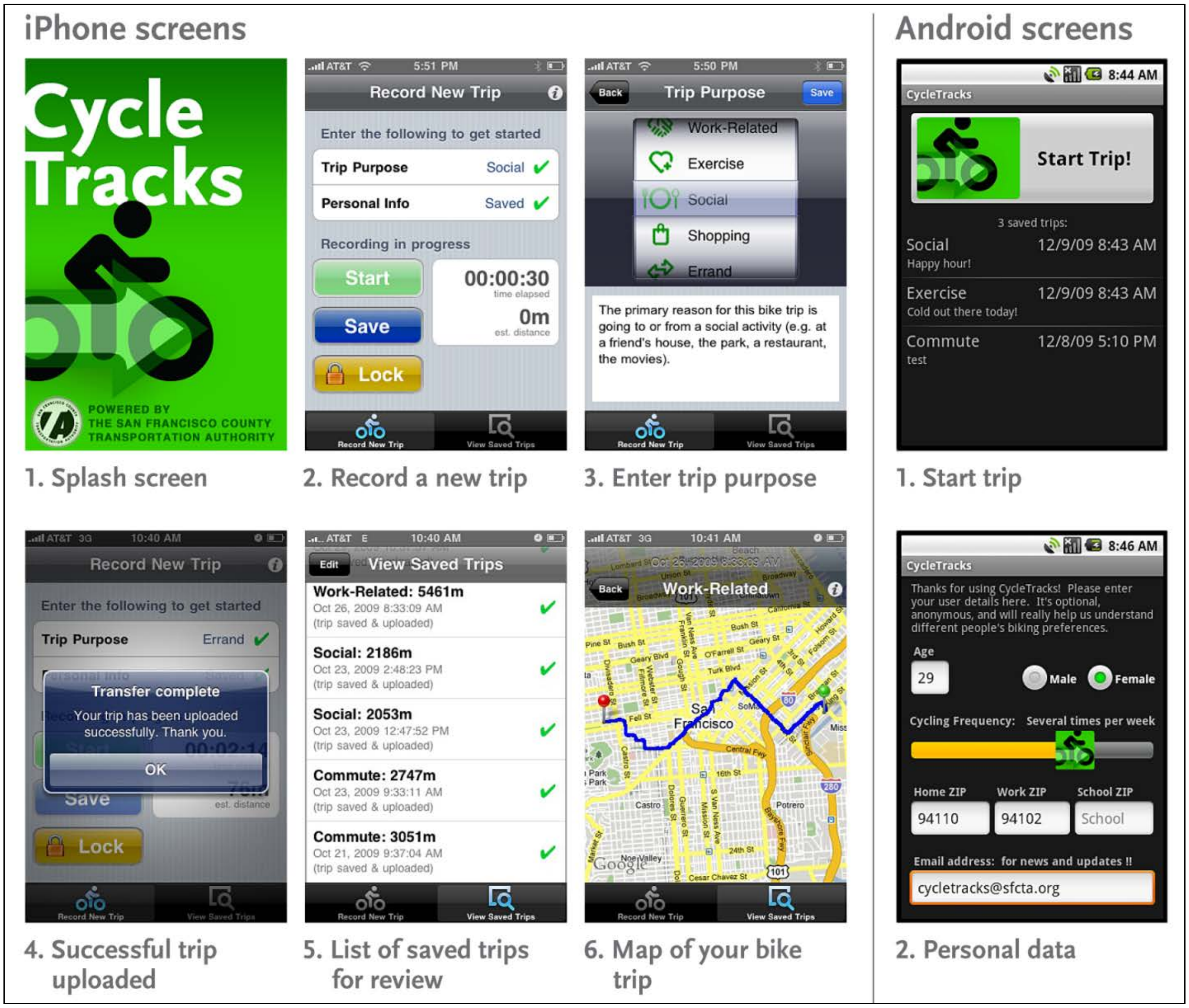


Figure 2. CycleTracks Application Screens

Recruiting Participants

For this project, researchers determined they would recruit volunteers to track the routes they took while cycling using the GPS-tracking CycleTracks smartphone application. As with any project that involves human subjects, receiving the Texas A&M University's Institutional Review Board approval was a critical step before any communication occurred. Upon receiving approval in late April 2011, researchers put into place the plan to recruit participants. Though unsure of the benefits it would bring, it was decided that a website would be a good first step in directing the public to information about the project. The website, www.cycletracksaustin.com, turned out to be a good central location to put the latest news, background information, and download instructions. Figure 3 illustrates the look of the website. The website address was placed on all printed publications, and quick response (QR) codes were printed on the posters for quicker access to instructions. A side benefit was that the number of visits to the website could be tracked. Counting the number of hits to the website was an easy way to measure the effectiveness of the outreach effort.

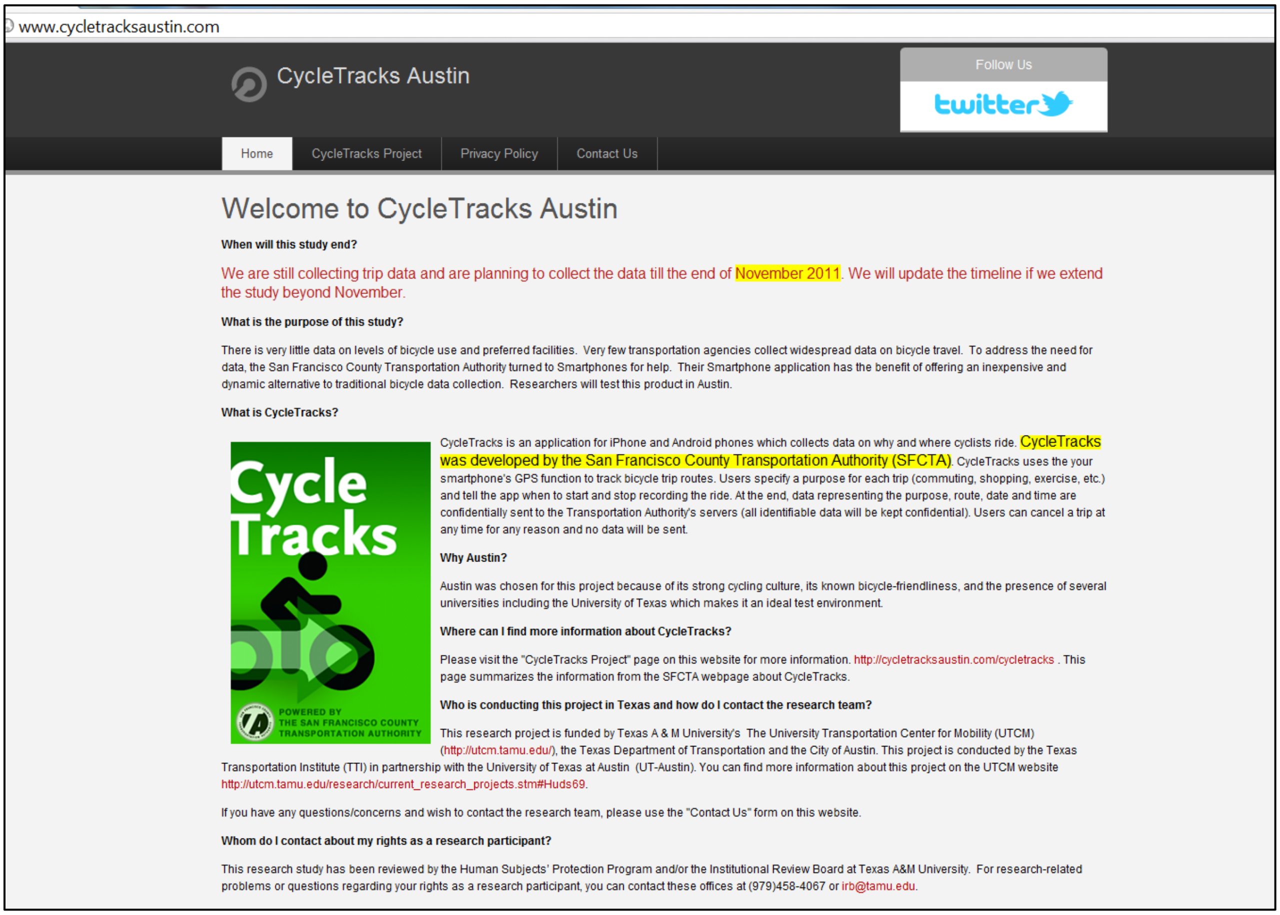


Figure 3. CycleTracks Austin Website Screen Shot

Extensive recruitment methods were enacted with the help of the Texas Transportation Institute's marketing and communications staff members, who were in the process of getting more involved in social media outlets at the time of this study. As such, the study was promoted on the TTI Facebook page and Twitter account. Others promoted the study via electronic media by posting on personal Twitter and Facebook pages, as well as encouraging retweets and shares by account followers. TTI Communications Services staff wrote a press release in early May 2011 that was sent to 130 media contacts (see Appendix B). The timing of the release coincided with National Bicycle Month, which garnered additional interest in the project. An extra push was given on Bike to Work Day (Friday, May 20, 2011), when postcards were placed at each of the 40 stations around Austin (see map in Figure 4). Several media outlets contacted TTI and conducted interviews with researchers. These included KEYE-42 (CBS Affiliate), KXAN (NBC Affiliate), YNN (Your News Now), KLBJ AM radio, and the *Austin Chronicle* and *Daily Texan* newspapers.

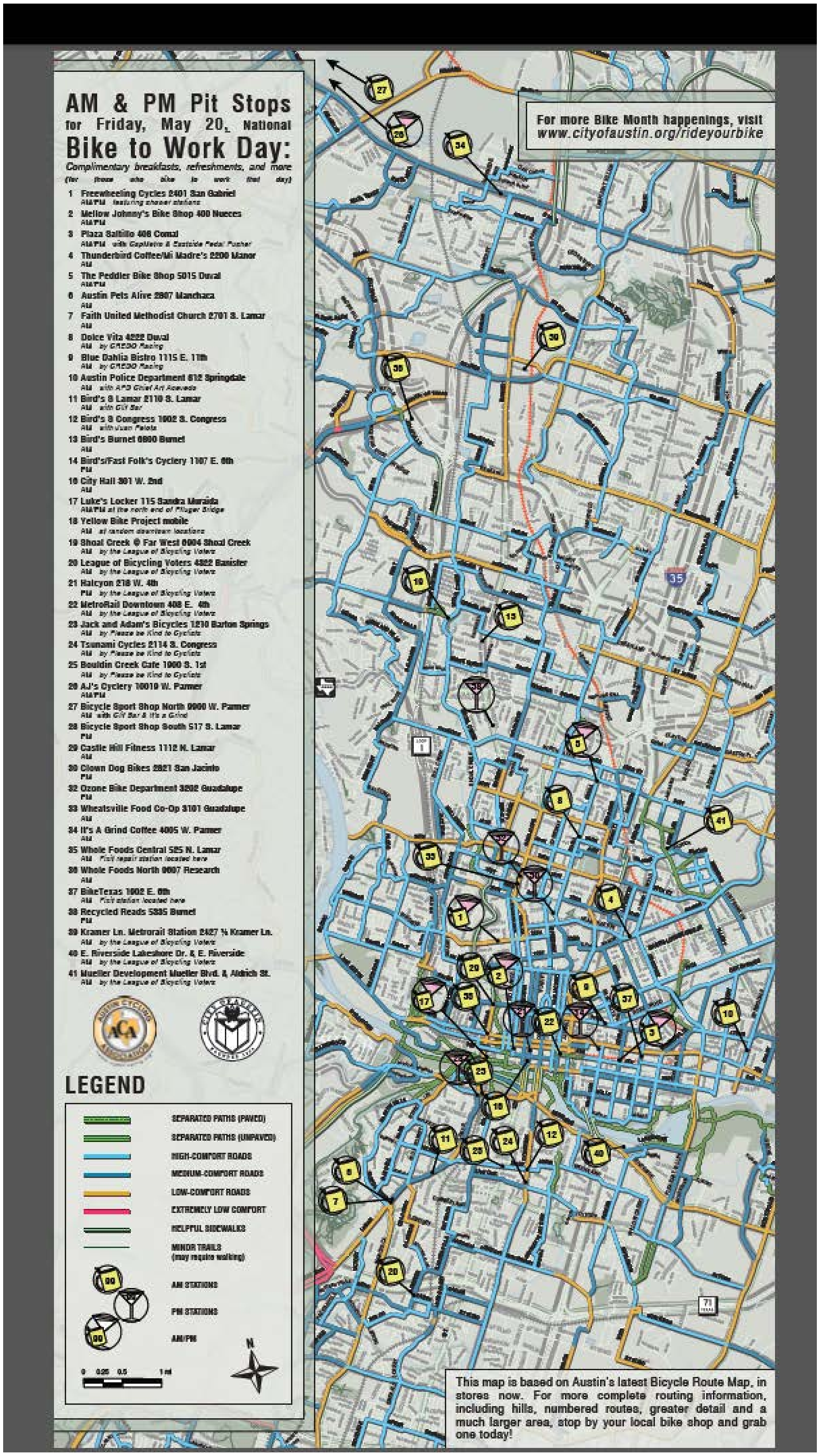


Figure 4. The City of Austin 2011 Bike to Work Day Map

Additionally, researchers emailed participants from a previous bicycle use survey who had indicated interest in participating in future studies and contacted cycling groups to request that they post the study recruitment information on their email lists, forums, and websites. Local government and business entities, such as the Hispanic Chamber of Commerce, Capital Area Metropolitan Planning Organization (CAMPO), and City of Austin Bicycle Program, were enlisted to promote the study in their email lists and newsletters, as well their social media outlets. The Texas Bicycle Coalition also spread the word through its listserv.

Recruiting efforts also included placing flyers in retail establishments that were known to be frequented by bicyclists, including most of the bicycle shops within the city of Austin. Other establishments included restaurants such as Freebirds in the Hancock shopping center, which is located near the Hyde Park neighborhood; student housing and a University of Texas (UT) shuttle bus stop; and outfitters such as the REI location in Round Rock, which sells bicycles in addition to other outdoor activity supplies. Both of the cited locations have public bulletin boards available, with REI even having a specific area set aside for bicycle-related activities. Numerous coffee shops were also targeted for posting information. Figure 5 shows the postcards that were developed for publicizing the project and requesting help.

Researchers approached colleges within Austin about promoting the project on their campuses. The Parking and Transportation Services Department at the University of Texas assisted in spreading the word by distributing flyers to students and faculty registering their bicycles and at events like their annual bike auction. Flyers were placed on bulletin boards around the UT campus as well. Finally, the UT bicycle coordinator emailed information to his listserv. In June, *The Daily Texan* (UT's free campus newspaper) printed an article about the project (see Appendix B).

Other Austin-area universities and colleges were also contacted. Phone calls were made to inquire about posting flyers at St. Edwards University, Huston-Tillotson University, and the Austin Presbyterian Theological Seminary, as well as at Austin Community College (ACC) campuses. Not every campus that was contacted responded to the request, but several, including the Eastview, Northwest, and Pinnacle campuses of ACC, did agree to accept a flyer to review before posting. During the review process at one of the campuses, it was discovered that the ACC Institutional Review Board would need to approve the flyers before posting. Given the timeframe of the study, researchers chose not to pursue posting flyers at those locations at that time. However, a second push for bicycle route data was conducted in September and October, and the ACC Rio Grande campus was contacted to see about placing flyers around their ideally located downtown campus. They agreed to post flyers and distribute postcards.

In September, when the second push was conducted, 500 postcards were placed in rider packets for the Mamma Jamma Ride in Austin (a bicycle ride that benefits thousands of Central Texans coping with breast cancer). The Austin Cycling Association agreed to hand out postcards at its training clinics and encouraged participation at its monthly meetings. Researchers attended a Thursday Night Social Ride to personally visit with bicyclists and hand out information.

Newspaper ads were also purchased during this second push in order to get the information on popular bicycle blogs, online website promotions, and newspapers. These included *Austin on Two Wheels*, League of Bicycling Voters, *Southwest Cycling News*, and *The Austin Chronicle*.

Figure 5. Postcards Distributed in May and September of 2011 to Area Bicyclists

Contact was made with Downtown Austin Alliance and Capital Metropolitan Transportation Authority (CapMetro) to further spread the word. The Austin Neighborhood Council posted the information on its listserv, and the Windsor Park Neighborhood Association sent an email to residents encouraging participation in the project.

Posters similar to the postcards but 11 inches by 17 inches were placed in areas coming into Zilker Park during the Austin City Limits (ACL) Music Festival, for which thousands of bicyclists rode to attend the weekend of live music in mid-September. Figure 6 shows a photo of the bicycles parked at the ACL festival.

Figure 6. ACL Music Fest 2011 Bike Parking

Postcards were placed on bicycles around Austin at key locations. Also targeted was the City of Austin bicycle parking area in One Texas Center, which has several bicycle racks and a locked bicycle parking cage. Commute Solutions had CycleTracks postcards available on the CAMPO table at Blues on the Green, which was a summer concert series held weekly at Zilker Park.

It was clear that bicyclists had knowledge of the project following these outreach efforts. The project website developed in May received many hits initially, decreasing through the heat of the summer. In September, after renewed efforts to spread the word were made, a significant increase in website visits was seen. These then dropped to around 100 hits per day on average down to 20 hits per day trickling off in November. Figure 7 illustrates the number of website hits per week from April to November. These numbers indicate high interest in the project during marketing efforts. However, once these efforts concluded the interest declined rapidly.

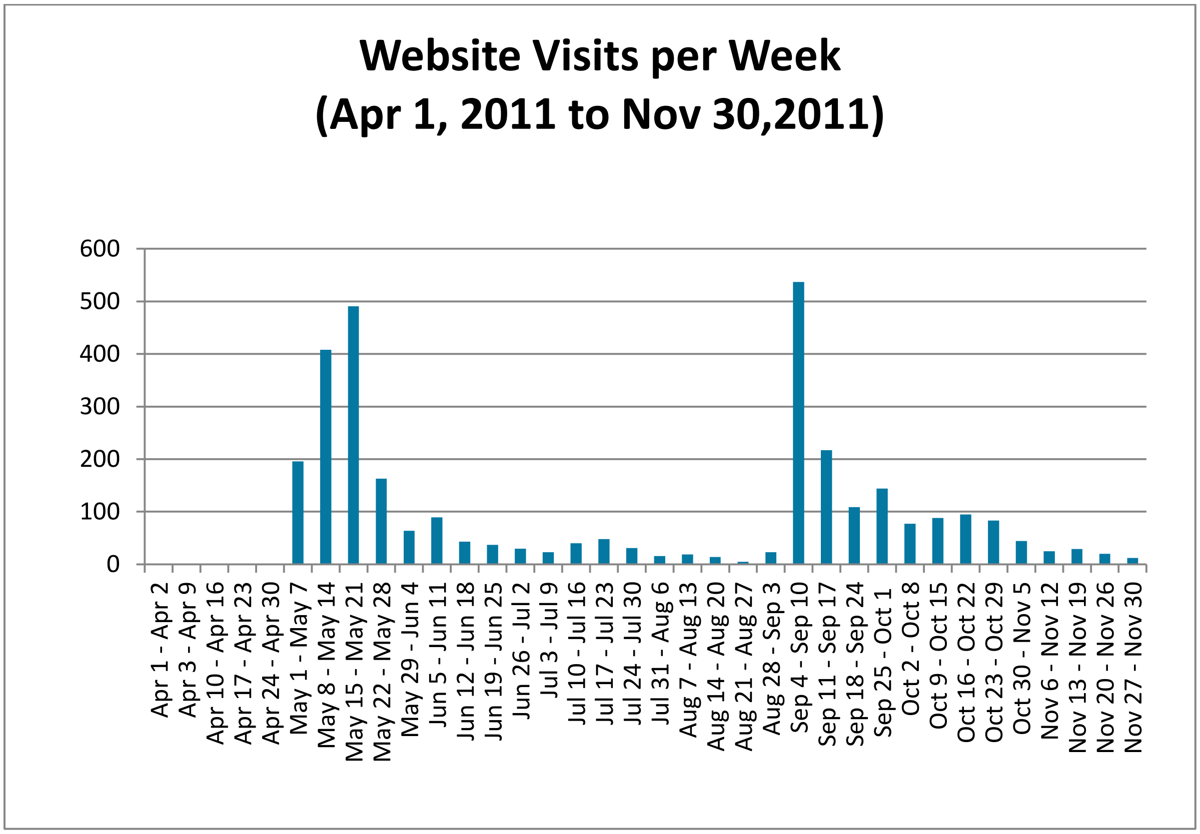


Figure 7. Website Visits per Week

Researchers compared the number of website visits per week to the number of trips recorded per week and found similarities. Coinciding with the peaks seen in Figure 7 when the marketing efforts were underway, the trips recorded had similar peaks as seen in Figure 8. The low number of trips recorded in August could be due to fewer people bicycling because of the extreme heat in Austin where almost 80 days of over 100 degree temperature occurred.

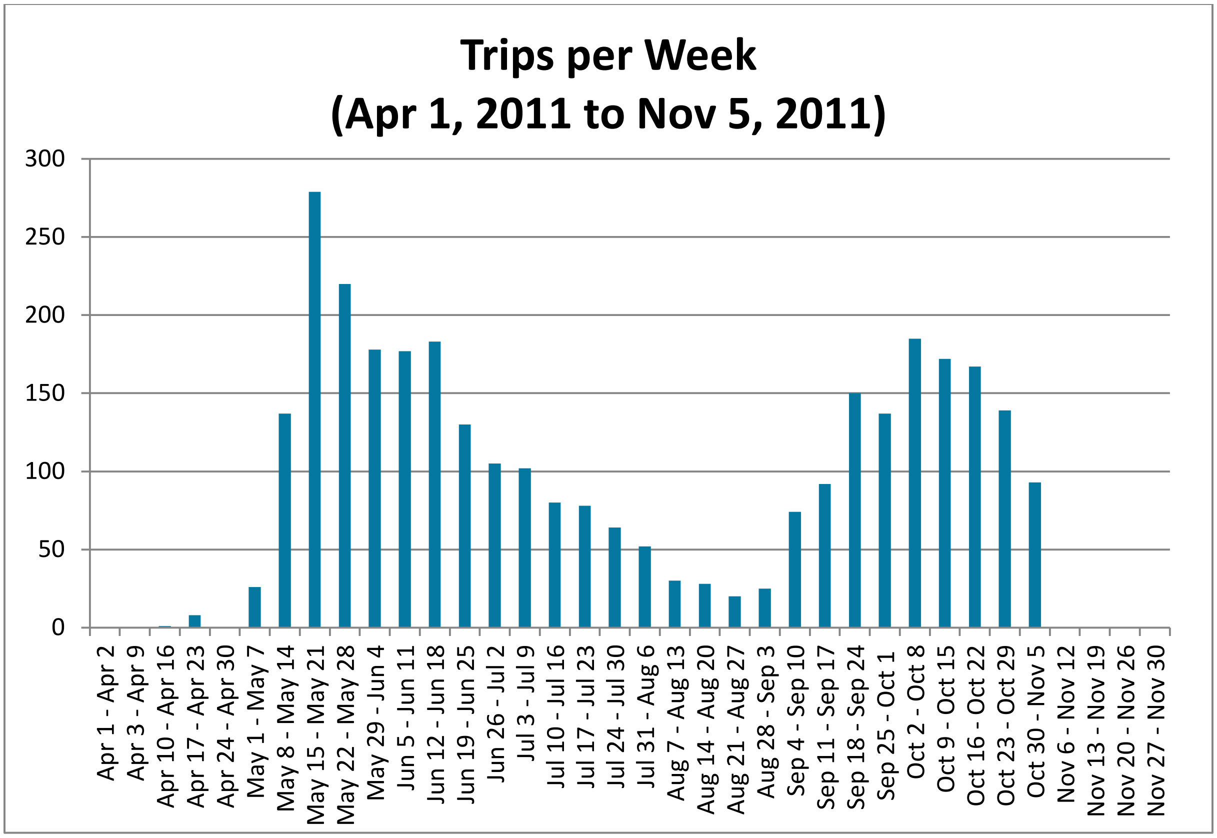


Figure 8. CycleTracks Recorded Bicycle Trips in Austin per Week

Data Preparation

Three critical components of processing GPS data include: (a) cleaning the data (e.g., removing outlying signals, signal errors, or very short segments), (b) creating a complete bicycling network that includes the network of streets and other links cyclists may use (e.g., park trails, parking lots, and driveways), and (c) matching the GPS points collected for each bike trip to the correct network links. Each of these components is discussed in this section.

Cleaning the Data

The first step in processing the GPS data was data cleaning. Data problems included signal interruptions which can occur because of obstructions (e.g., heavy tree foliage and multi-story buildings; Duncan and Mummery, 2007). Figure 9 shows GPS traces in the downtown area of Austin with one type of problem commonly encountered in the data: segments of a route where outlying GPS points cause the GPS trace to “fly off” to another area.

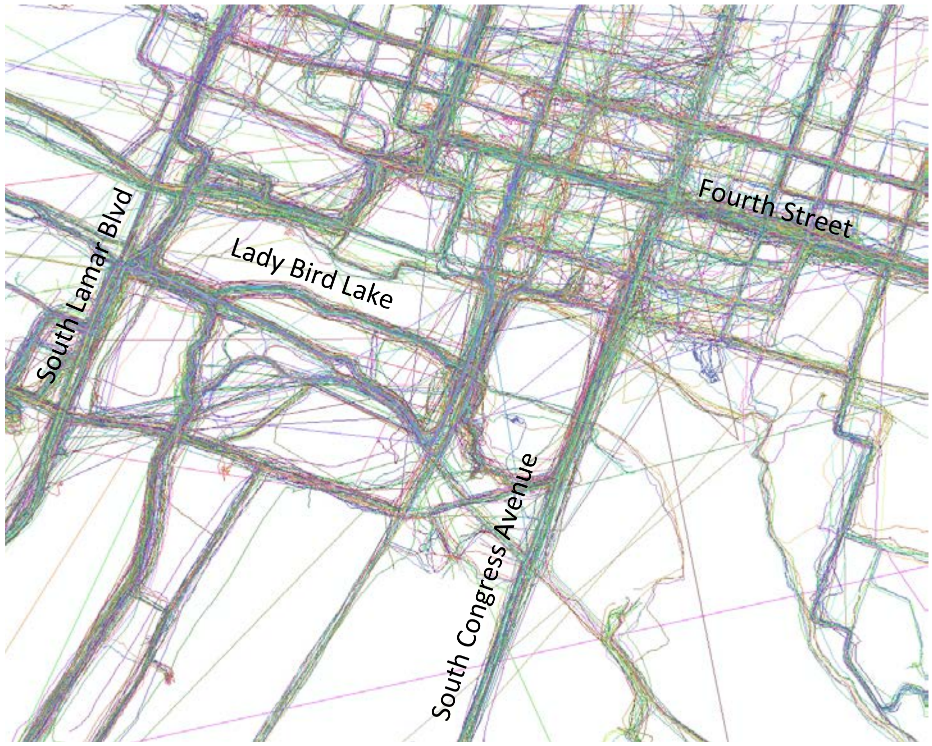


Figure 9. Excerpt of Map of Pre-Processed GPS Traces in the Downtown Austin Area

For the first step of data cleaning, several columns were added to the GPS Coordinates Table to note the changes between each pair of GPS points collected. These were (a) distance traveled since last point captured, (b) change in time, (c) speed, (d) change in altitude, and (e) slope. A trip was deleted from the dataset if it was taken outside of the study period (May 1 to October 31, 2011) or outside of the Austin study boundary. Points within a trip were deleted if (a) the horizontal or vertical accuracy measurement was greater than 100, or (b) the speed was greater than 30 mph or less than 2 mph. Once these points were removed, the new column data (representing changes between points) were recalculated. To account for trip chaining, a trip was split into multiple trips if there was more than three minutes or more than 1,000-ft between points. Finally, trips with fewer than five collected points were removed from the dataset.

If a trip was deleted from the GPS Coordinates Table, it was also deleted from the Trip Table. Similarly, if a user ID no longer had any trips associated with it, then the user was removed from the Person Table. Other minor changes were made to the dataset to ensure consistency, such as defining strict categories for gender (i.e., male, female, or other). The researchers recommend that future versions of the app offer defined choices for each demographic input rather than allow for open-ended responses.
After cleaning, the following problems remained:

1. GPS traces across parks, parking lots, driveways, campuses, and other places not coded as a link in the City of Austin's street network geographic information system (GIS) file.
2. GPS traces not aligning exactly with the network links.

The first problem was addressed by additional network coding and discussed in the following section. The second problem was managed through the choice of a map-matching algorithm, discussed in the section called Matching the GPS Points to the Network.

A total of 3,615 trips was collected by 317 participants, but after data cleaning only 3,198 trips remained to be input into the map-matching process.

Creating a Complete Network

Of interest in tracing the routes of cyclists is to find out the characteristics of the routes chosen. Schuessler and Axhausen (2009, p. 11) stressed that an "important requirement for each map-matching algorithm to work properly is a correct, consistent and complete representation of the real network by the network used for the map-matching. Unfortunately, hardly any network currently available can guarantee this requirement." This limitation is especially the case with tracking cycling routes because cyclists, like pedestrians, do not necessarily constrain their movement to the existing roadway network.

Schuessler and Axhausen (2009) used one of their map-matching algorithms to locate the routes of 3,932 car trip segments on a Swiss NAVTEQ™ network, a high-resolution proprietary navigation network covering Switzerland. Only between 2,065 and 2,088 car trip segments were successfully matched to the network, depending on the value set for a parameter in the algorithm. In evaluating why the number of matched routes was so low, a manual check revealed that for more than 80 percent of the unmatched car trip segments, links that should have existed were missing from the network (more than 40 percent), or travel occurred off the network (40 percent). Their results illustrate the importance of having a complete network. The following sub-sections discuss available networks and the process taken in this research project to complete the City of Austin network.

Available Public and Proprietary Networks

While it is unlikely that a network exists that includes all possible bicycle routes, several networks in the public and private domains offer a good starting point. Within the public domain, Zielstra and Hochmair (2012) make a distinction between authoritative data and volunteered data, with the former provided by professional organizations and the latter by volunteer collaborators.

As a source of authoritative data, cities typically offer free to the public a GIS representation of their roadway network created in-house by professional planners and GIS analysts, and they update it as new roads are added or deleted. The U.S. Census Bureau also maintains and offers free Topologically Integrated Geographic Encoding and Referencing (TIGER) System/line data in ESRI ArcGIS shapefile form that includes roads, railroads, rivers, lakes, and political, census, and natural boundaries (U.S. Census Bureau, 2012). The attributes of the road features include address ranges, road classification, geometry, length, street name, and zip code.

In addition to public agencies such as cities and the U.S. Census Bureau providing public data, members of the public can directly create network datasets through web applications that allow anyone to volunteer data to build and correct a network dataset online. Open Street Map (OSM, http://www.openstreetmap.org), a well-known example of such an effort, relies on volunteers to create a detailed map of communities and includes roadways and their attributes, in addition to locations of interest. In the United States, OSM began with the TIGER/line data from the U.S. Census (Zielstra and Hochmair, 2012). OSM network datasets must be converted to GIS format using online tools (e.g., http://code.google.com/p/osm2shp/) or from companies such as Geofabrik.de and Cloudmade.com (Zielstra and Hochmair, 2012).

Proprietary network datasets available for purchase also exist. NAVTEQ™ maps provide representation of the road network with up to 260 attributes, such as turn restrictions, physical barriers and gates, one-way streets, restricted access, and relative road heights (NAVTEQ, 2012). Other proprietary providers of networks include TeleAtlas and ESRI ArcGIS. Users of ArcGIS with a license have access to a proprietary road network dataset from TomTom that represents streets, interstate highways, and major roads within the U.S. and Canada in 2007 (ESRI, 2012).

Regardless of the source of the network dataset, the network will most likely be incomplete for analyzing cycling routes because the routes can include travel in areas other than the street network, such as parking lots and parks. As described in the next section, a network dataset that includes links for off-street paths was not available for Austin; therefore, links were manually added or appended from other datasets.

City of Austin Network

For this Austin CycleTracks study, the City of Austin's freely available ArcGIS shapefile of the roadway network was selected for use because the file provided the attributes needed (e.g., speed, road classification, one-way designations, and grade separation information). The process used for importing the GPS data into ArcGIS is in Appendix C. A review of the GPS traces revealed quite a few areas throughout Austin and surrounding area where the cyclists participating in the study strayed from the street network (see Figures 10, 11, and 12 where the gray lines represent the City of Austin street network and the red lines represent the GPS traces for each bike trip recorded in the study). The next subsection discusses the research team's effort to complete the network.

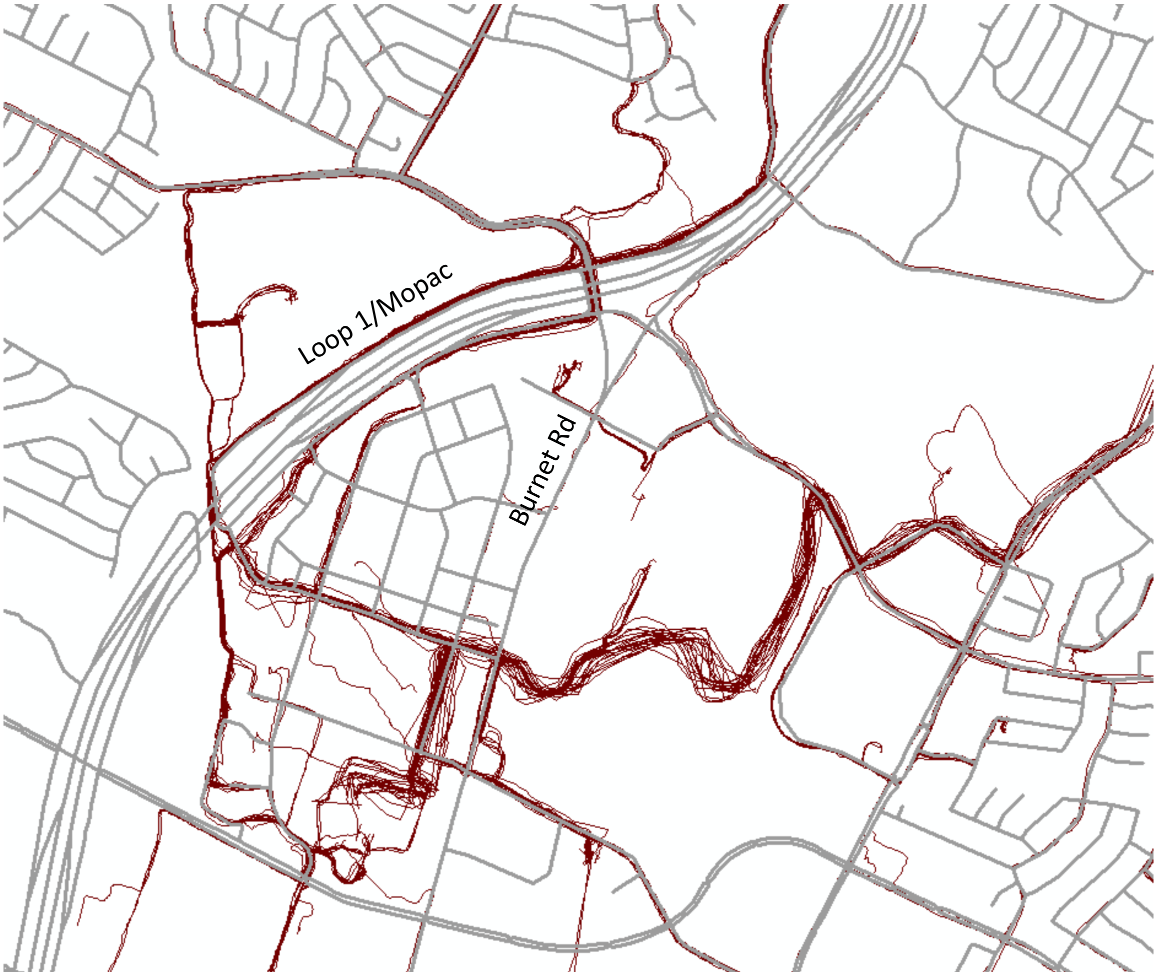


Figure 10. Bike Trips Traveling off the Street Network (between North Mopac and Burnet Road)

Figure 11. Bike Trips Traveling off the Street Network (Auditorium Shores)

Figure 12. Bike Trips Traveling off the Street Network (Central Park and Shoal Creek Park)

Completing the Network

Schuessler and Axhausen (2009, p. 12) dealt with the issue of an incomplete network by making the assumption that there are "no systematic errors in the network coding but that missing links are randomly distributed throughout the network." However, a review of the map reveals that assumptions should not be made for matching GPS traces made by cyclists. Rather than being randomly distributed, the missing links tend to be parking lots, driveways, and parks frequently used by cyclists. Assigning the GPS traces to nearby existing roads instead of those alternative facilities would give the wrong impression that cyclists are using existing street facilities, when in fact the cyclists most likely avoid the street facilities to shorten travel time, avoid unsafe conditions, or create a more pleasant travel route.

To avoid losing information about the facilities being used by cyclists, the time-consuming process of creating additional links in the network was undertaken for this Austin case study to create as complete a network as possible in the time provided. Zhou and Golledge (2006) noted that "compared to the rapid advances of positioning technology, improving map accuracy is more of a long-term, energy-consuming task . . . the map details have not been able to extend to the details of lanes and cover particular types of roads such as bike paths or sidewalks."

Examples of non-street facilities observed to be used by cyclists in the Austin CycleTracks study include parks and other off-street paths, campus sidewalks, parking lots, driveways, open fields, and alleyways. The existing City of Austin street network dataset included a field for road class with possible values ranging from 0 to 16. Additional road class values (17-26) were created by the research team to allow for more network coding options that better describe new links in the network. The Snap editing tool in ArcGIS 10.0 provided the means of ensuring that the new links directly connected to the existing network. In cases where the new link did not connect to some of the links, such as when a path existed underneath an elevated highway, the existing street dataset from the City of Austin included a field called ELEVATION1 that took a value of 1 for links at different elevations from other intersecting links. Table 1 lists the road classes as well as their description and percentage, by length in miles, within the entire coded network after the addition of links.

The network completion process resulted in researchers manually adding 923 additional links representing a wide variety of classes, appending 310 links from the existing Capital Area Council of Governments (CAPCOG) street dataset, and appending 343 links representing multi-use paths (Class 20) from the City of Austin park trail dataset. A lesson learned is that effort must be undertaken to ensure that the appended links are connected fully to the existing network.

Most of the successes of the map-matching algorithms discussed in the next section are dependent on the accuracy and completeness of the network.

Table 1. Road Classes Coded in Network Dataset

Road Class (ROAD_CLASS)	Description	Miles in Total Network	% in Total Network
Original			
1	Interstate highway, expressway, or toll road	195.53	2.75%
2	U.S. and/or state highway	186.40	2.62%
3	n/a (not described in metadata)	36.17	0.51%
4	County roads (RR, RM, FM, etc.) and/or major arterial	801.14	11.28%
5	Minor arterial	228.06	3.21%
6	Local city/county street	4481.29	63.08%
8	City collector	767.89	10.81%
9	n/a (not described in metadata)	0.12	0.00%
10	Ramps and turnarounds	138.91	1.96%
12	Driveway	49.20	0.69%
14	Unimproved public road	0.30	0.00%
15	Private road	57.03	0.80%
New			
17	Off-street path in park	8.57	0.12%
18	Off-street path on vacant or private property, or non-park open area	12.41	0.17%
19	Parking lot (to differentiate this category from driveway, this category is used for when the link passes mostly by parking stalls)	45.84	0.65%
20	Multi-use path for pedestrians and cyclists (this also includes the city-designated paths such as the Town Lake Hike and Bike Trail and the Lance Armstrong Bikeway, from the City of Austin's GIS file)	81.33	1.14%
21	Building walkway	0.56	0.01%
22	Paved sidewalk in park, campus, or open area	8.48	0.12%
23	School track or designated veloway	3.39	0.05%
25	Transit platform/station (e.g., Metrorail station area)	0.05	0.00%
26	Alleyways	1.44	0.02%

RECOMMENDATIONS

- During the data collection phase, fill in the network by adding missing links most likely to be used by cyclists.
- Make sure all layers used to create the street network are fully connected to one another
- Code the network with the type of facility the link represents to differentiate the types of facilities cyclists use on their routes.
- Use the ArcGIS snap tool during editing to guarantee continuity of the network.

Matching the GPS Points to the Network

The process of associating or transferring geographic data to another set of geographic data is generally referred to as map-matching. In the case of bicycle transportation, the GPS points for a bike trip (the sequence of which forms a GPS trace) indicating the route taken by bike are matched to the network links, representing roads, trails, and other facilities coded into the network as links. Unfortunately, GPS traces rarely align perfectly with the centerline of network links; therefore, procedures (i.e., algorithms) must be implemented to match the GPS trace to the most likely network links used by cyclists. The published literature offers a variety of map-matching algorithms to choose from, each with advantages and disadvantages. They are generally categorized as taking a geometric, topological or advanced approach or a combination of approaches. Appendix D provides an overview of the different categories of map-matching procedures.

In an effort to make the map-matching process as accessible to local agencies as possible, map-matching algorithms that could be easily implemented in ArcGIS 10.0 were tested for this project. ArcGIS is a GIS application commonly used by local planning agencies. Alternatives to ArcGIS for map-matching were tested and are described in Appendix E.

Both ArcGIS-based geometric and topological procedures were used to match the GPS data to the network. The first map-matching attempt used a purely geometric approach of snapping the GPS trace for each bike trip to the closest network links and then spatially joining the network link attributes to the user attributes. This geometric approach proved problematic. Though simple to implement, use of the Snap tool introduced errors for data analyses (Zhou and Golledge, 2006; Pyo et al., 2001; Schuessler and Axhausen, 2009). The tool does not take into account the feasibility or continuity of a route. For instance, a GPS trace starting on Street 1 may be closer to Street 2 than Street 1, but Street 2 is not accessible from Street 1. Additionally, GPS traces contain quite a bit of noise, and so portions of a GPS trace may be closer to a network link other than the one actually used. Around intersections, the problems with snapping to the closest link became quite evident. Figure 13 presents the results of snapping (using line-to-line snapping) of the GPS trace to the network links. The intersecting street link is closer to the GPS trace as the trace approaches an intersection (the light blue line is an example of an unsnapped GPS trace for a bike trip). The lack of consideration of feasible routes between an origin and destination pair and the inclusion of intersecting streets made the resulting spatial join of the network data suspect.

Figure 13. Post-Snap Results Showing Problems With Using Geometric Approach

Some of the disadvantages of geometric map-matching algorithms were overcome by incorporating consideration of the topology of the network (e.g., the allowable flow on a link) to match GPS traces so that feasible routes were generated.

An algorithm developed by Dalumpines and Scott (2011) uses geometric and topological map-matching approaches using ArcGIS's Network Analyst tools. Since their algorithm avoided the problems with a geometric-only approach and was implementable in ArcGIS it became the algorithm of choice for matching the Austin CycleTracks GPS traces to the network. As mentioned, Appendix D includes a description of the Dalumpines and Scott (2011) algorithm and Appendix F provides more information about how the researchers for this project implemented the Dalumpines and Scott (2011) algorithm in ArcGIS. Processing of the data in ArcGIS took several days to complete.

Of the 3,198 bike trips, the algorithm was able to generate routes on the network links for 2,820 of the trips, for an 88 percent match rate. Interestingly, this rate is the same match rate achieved by researchers in the Dalumpines and Scott study. Figure 14 compares the matched routes with the original GPS bike trip traces. Unlike the Dulumpines and Scott study, this CycleTracks study does not have a second source of location information (e.g., travel diary) to verify the accuracy of the map-matching algorithm.

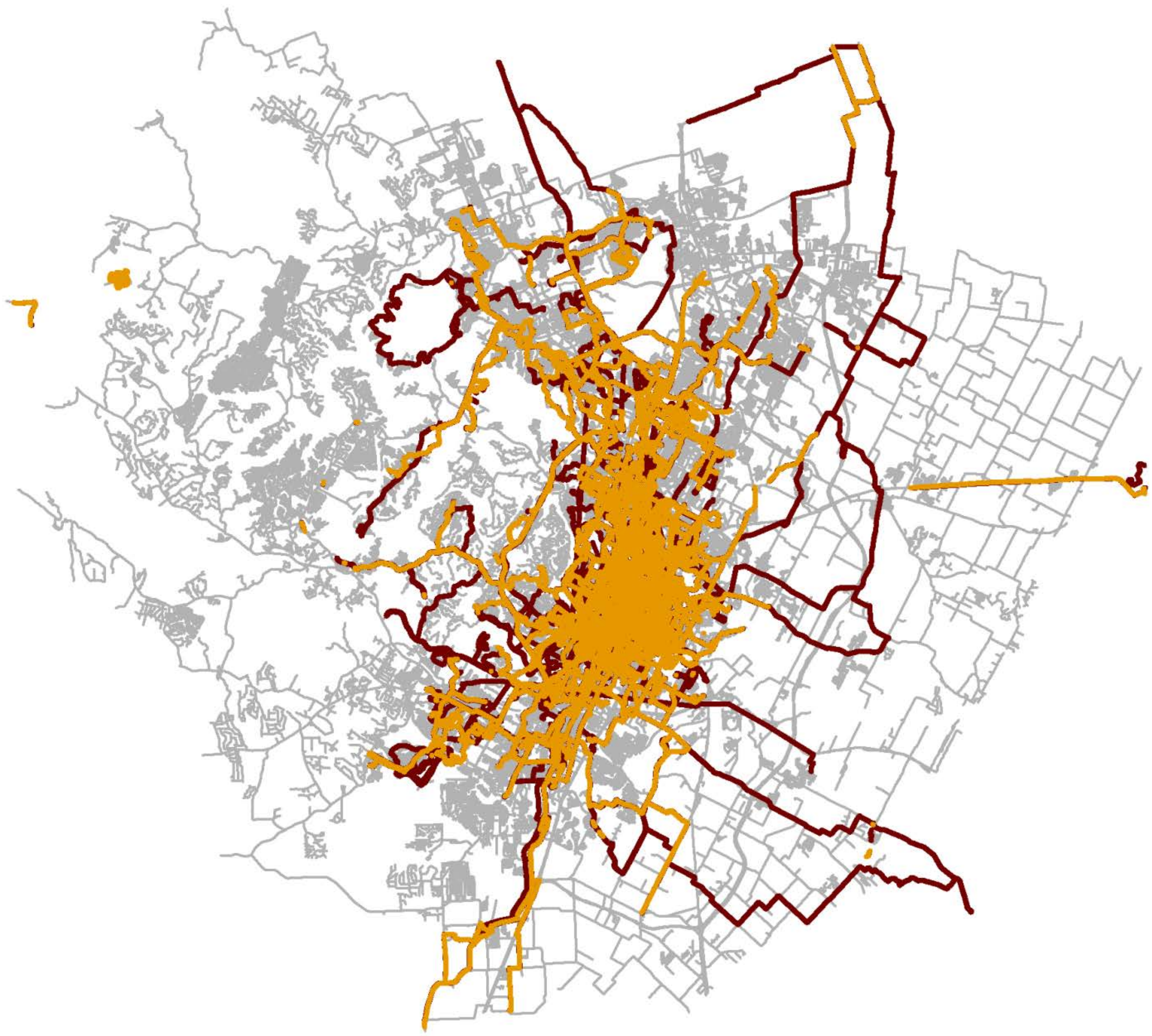


Figure 14. Original GPS Traces of Bike Trips (Maroon) and the Routes Matched within 200-ft Buffer around GPS Trace (Orange)

One of the drawbacks to the Dalumpines and Scott (2011) algorithm is that it requires the user to specify a value for a parameter – a buffer distance around the GPS trace – used to constrain the search for a feasible path in the network for the GPS trace. The specified buffer distance has an influential impact on the success rate of the matching process and has the potential to introduce errors in map-matching. The researchers for this project experimented with modifying the Dalumpines and Scott (2011) algorithm to become a parameter-free map-matching algorithm that matches GPS traces only to the closest network links and only if the route formed by those links is feasible. Because it is more restrictive, the number of GPS traces matched to the network though was less than that of the Dalumpines and Scott (2011). To maximize the data used for the data analysis, the Dalumpines and Scott (2011) algorithm results are used.

RECOMMENDATIONS

- Take advantage of ArcGIS's Network Analyst make Route Layer tool to identify feasibility of routes.
- Consider trying and using a combination of map-matching algorithms to capture as many feasible likely routes as possible and seek out map-matching algorithms with very few or no parameters.

Data Analysis

Data analysis was conducted before and after the map-matching process. The analysis before map-matching focused solely on describing the sociodemographic attributes of the participants to assess how closely the participant pool represented the population. Only 316 participants with usable GPS traces (3,264 in total) were included in this analysis.

The data analysis after the map-matching process sought to identify the characteristics of the bike routes and of the participants who chose the routes. Since the map-matching process with the Dalumpines and Scott (2011) algorithm did not match all the GPS traces, the analysis uses only the data associated with the 2,820 GPS traces matched to the network.

Description of Participants

As mentioned previously, when downloading the CycleTracks app, users were asked to answer questions about their gender, age, cycling frequency, home zip code, and work/school zip code. Tables, graphs and maps in this section describe the demographic information provided by the participants (all of whom were adults) with usable GPS traces.

Of those that did report a gender, 70 percent were male and 30 percent were female (see Figure 15). These proportions are very similar to the proportions of bicyclists by gender specified in the 2002 National Survey of Pedestrian and Bicyclist Attitudes and Behaviors (USDOT, 2002): 63 percent male and 37 percent female. They were even more similar to the proportions of bicycle commuters by gender in the 2012 Benchmarking Report, which used a weighted average of data from the American Community Surveys of 2007 to 2009: 72 percent male and 28 percent female (Alliance for Biking and Walking, 2012).

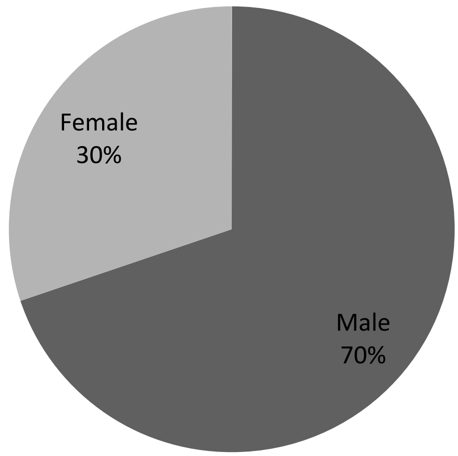


Figure 15. Gender of Participants (n = 302)

The age of the participants is shown in Figure 16, for those who provided this information. Over one-third of the participants were between the ages of 20 and 29, while under one-third were 30-39 years old and 22 percent were in the 40-49 age range. According to the 2009 National Household Travel Survey, 39 percent of bicyclists are under age 16, 54 percent are between the ages of 16 and 65, and 6 percent are over 65 years of age (USDOT, 2009). These results are not directly comparable to study results since the study was targeted to adults only.

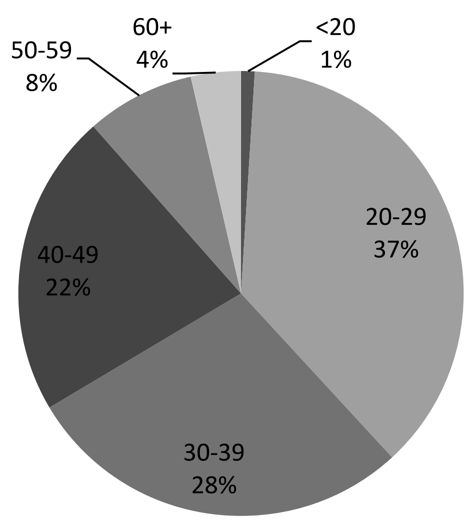


Figure 16. Age of Participants (n = 304)

As shown in Figure 17, by far the majority of participants rode a bicycle at least several times per week. If cycling frequency is truly a proxy for level of expertise as the SFCTA staff purported, then the study targeted expert bicyclists. This is important to consider when analyzing the results since the routes selected by novice bicyclists (or those who do not cycle frequently) may differ from the routes selected by more experienced bicyclists.

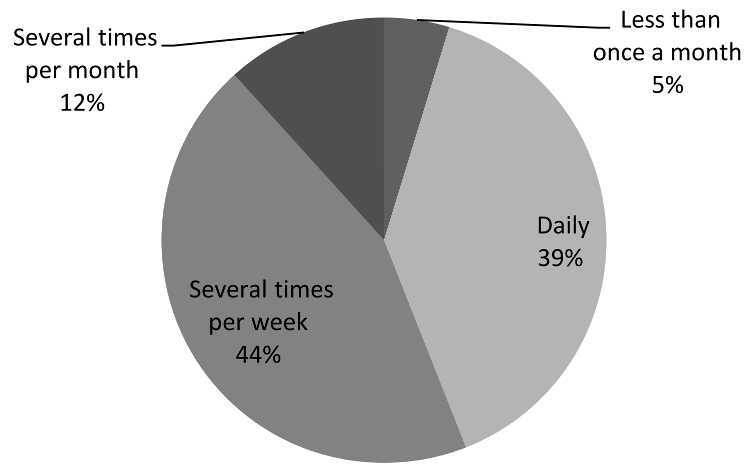


Figure 17. Participants' Cycling Frequency (n = 316)

Trip purpose is also important to consider when analyzing results because bicyclists may choose a more direct path when operating under a time constraint (e.g., need to arrive at work on time) and may choose a more circuitous path when bicycling for exercise. Nearly half of the trips logged were commute trips, as illustrated in Figure 18. However, there is overlap in the list of options provided in the application as far as trip purpose. A commute trip might be viewed as a trip to school or work. When speaking of bicycling for transportation versus recreation, one could combine trip purposes. For example, trip purposes defined as commute, errand, work-related, shopping, social, and school should be combined to give a picture of bicyclists who are riding for transportation. About 85 percent of the trips recorded were for reasons of transportation.

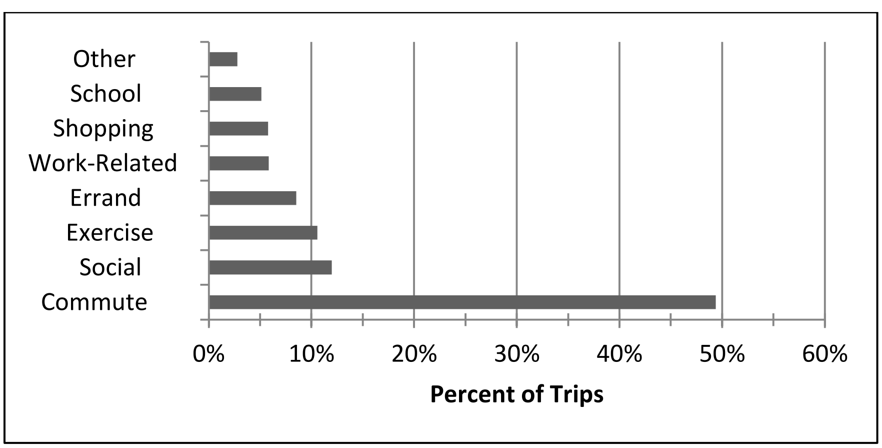


Figure 18. Participants' Trip Purposes (n = 3,264 trips)

There was a slight difference in cycling frequency when compared across genders, as shown in Figure 19. Of the male participants, a higher proportion bicycled at least several times per week compared to the female participants. Interestingly, a high proportion of participants who chose not to specify their gender bicycled less than once a month.

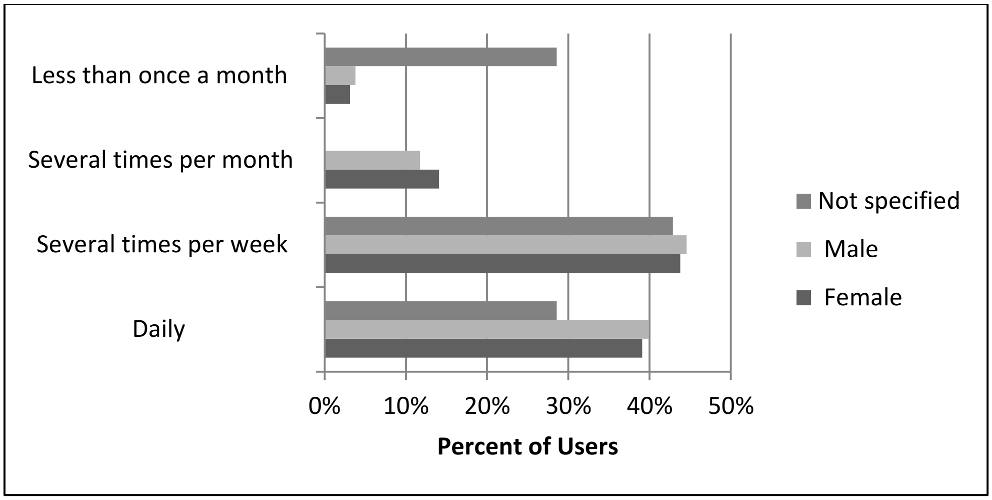


Figure 19. Cycling Frequency by Gender

Table 2 shows that the proportion of trips logged for each trip purpose did not differ much between male and female participants, with the exception of exercise as a trip purpose, where the percent of trips males took for exercise was twice that of females. Work-related trips were also higher as a percentage for males.

Table 2. Trip Purpose by Gender

Purpose	Total of Trips (Trip IDs)	% Total of Trips (Trip IDs)	Male	% Male	Female	% Female
Commute	1612	49%	1198	49%	370	51%
Errand	278	9%	196	8%	71	10%
Exercise	345	11%	283	12%	46	6%
Other	91	3%	64	3%	27	4%
School	167	5%	110	5%	57	8%
Shopping	189	6%	138	6%	36	5%
Social	391	12%	256	11%	107	15%
Work-Related	191	6%	179	7%	9	1%
TOTAL	**3264**	**100%**	**2424**	**74%**	**723**	**22%**

Figures 20, 21, and 22 show the geographic distribution of the participants' home, work, and school locations by zip code. Most participants lived or worked in central Austin—north of Ben White Boulevard/US 290, south and west of US 183, and east of Loop 360.

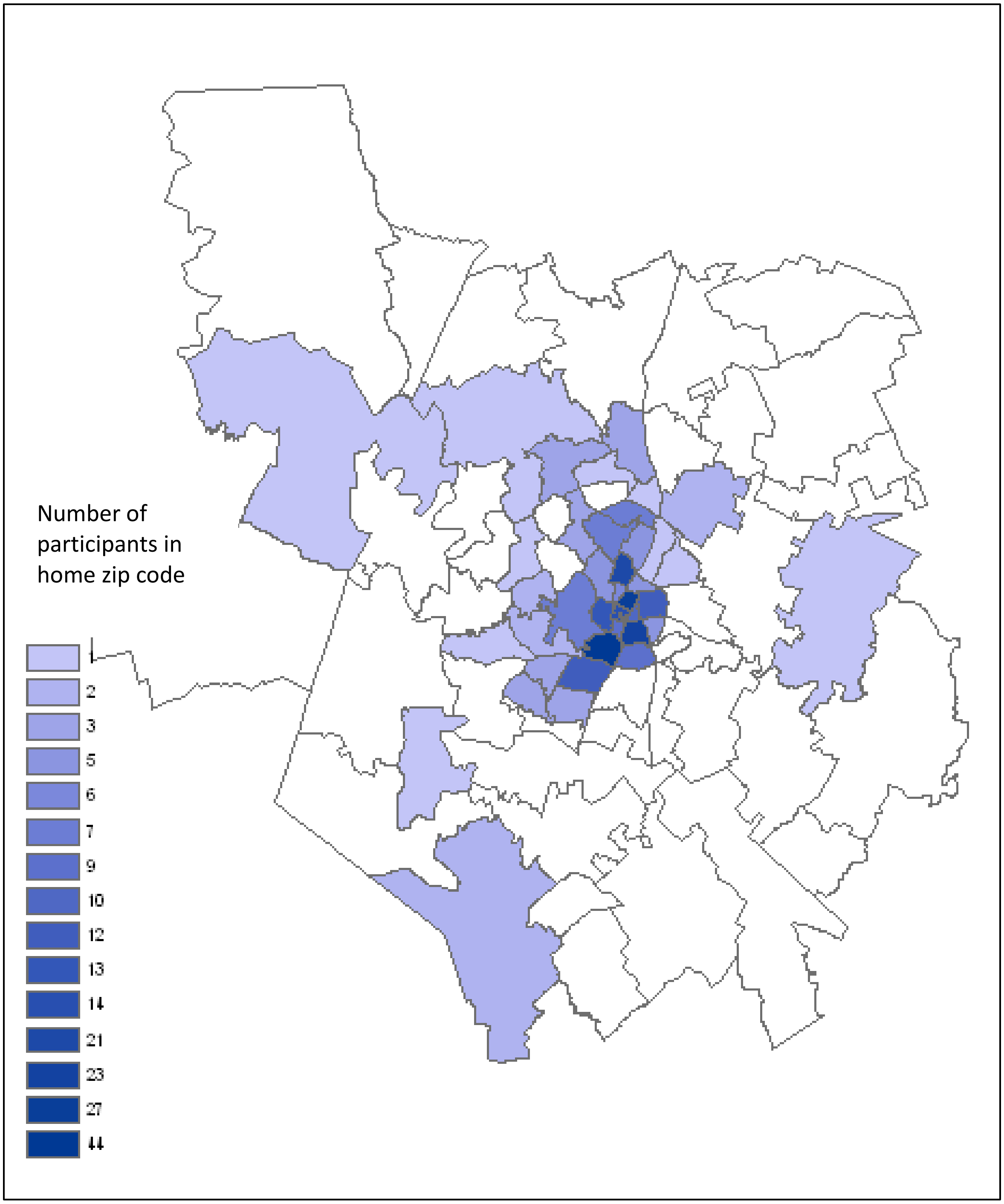


Figure 20. Count of Participants by Zip Code of Their Home

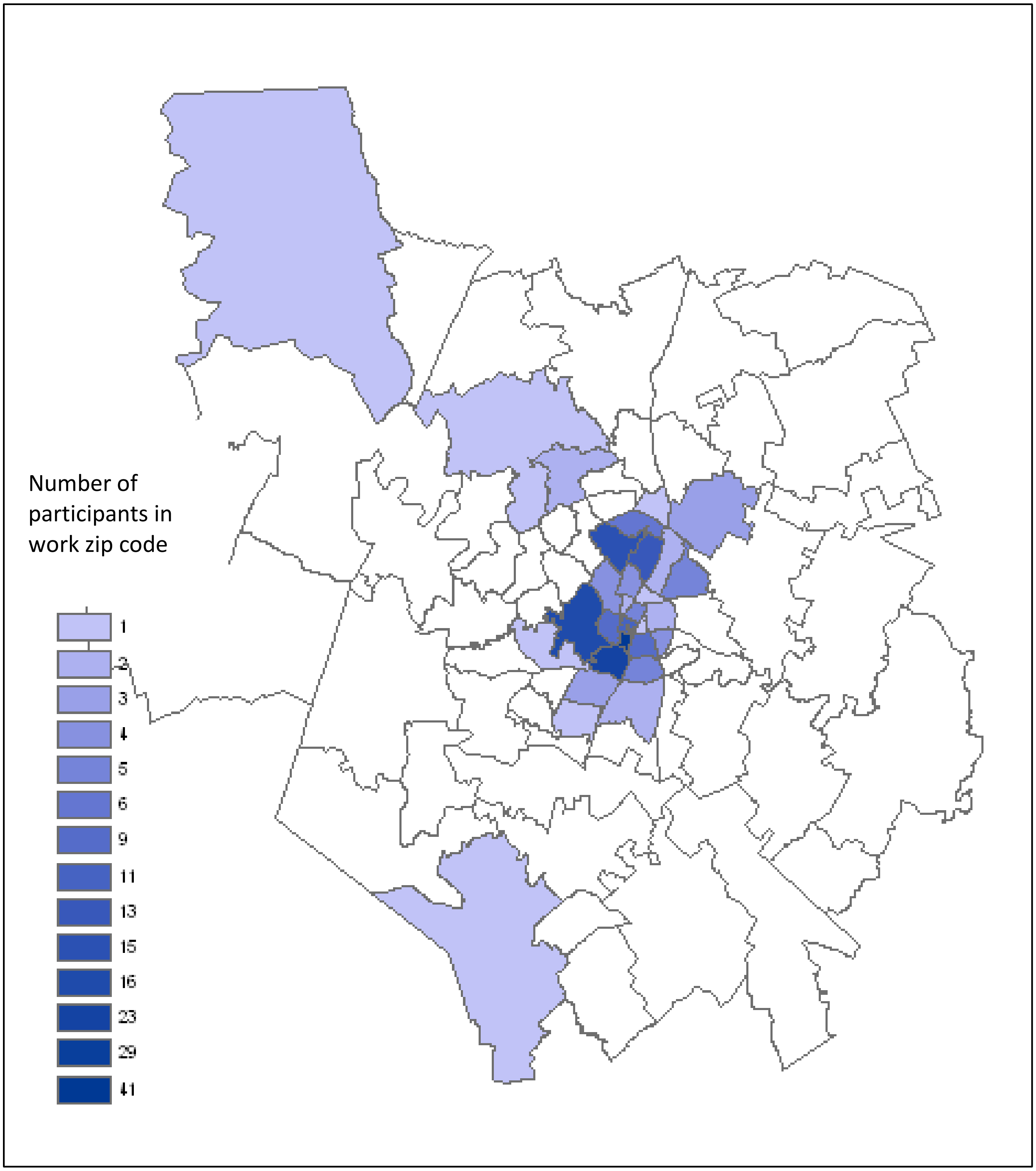


Figure 21. Count of Participants by Zip Code of Their Workplace

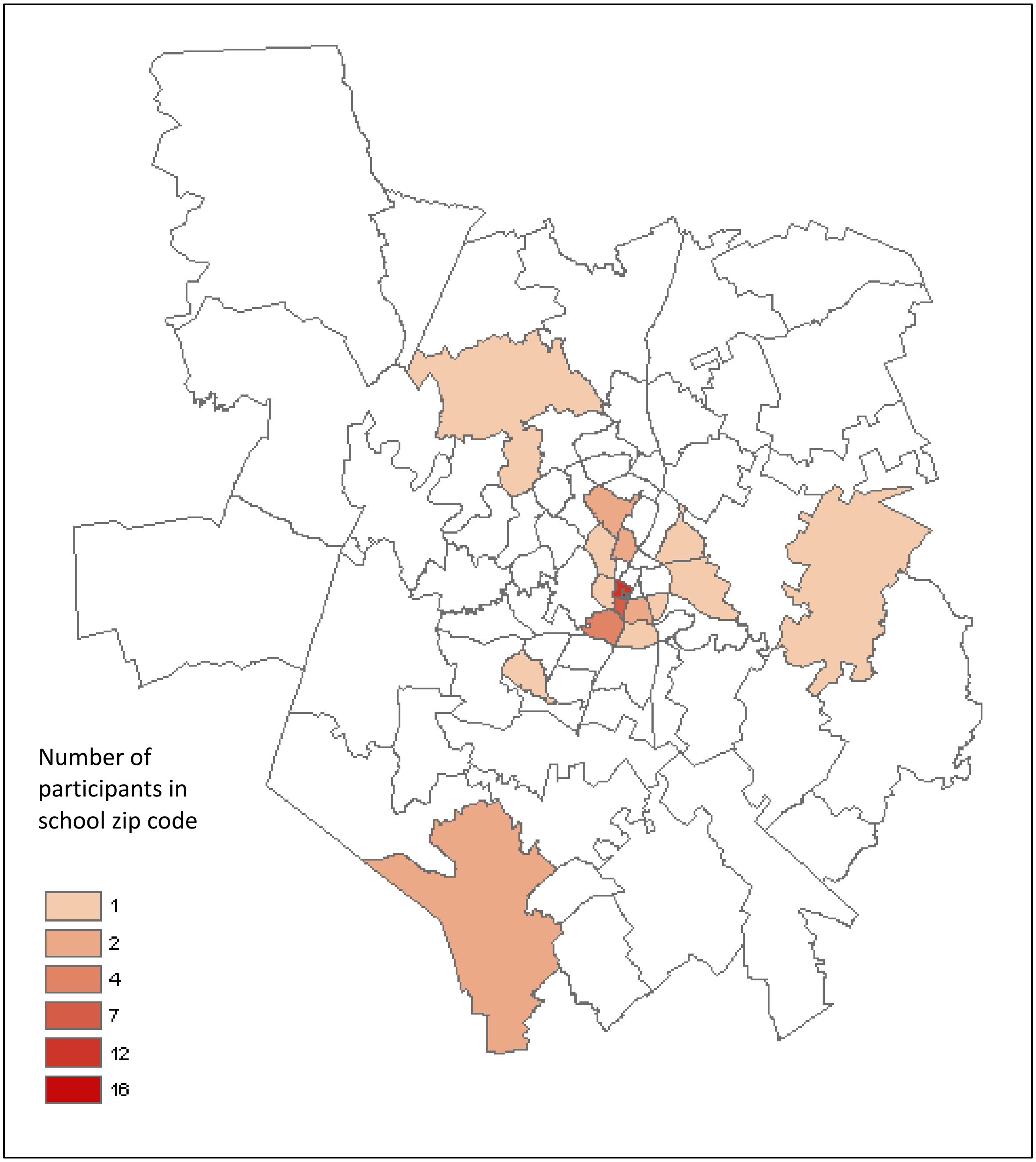


Figure 22. Count of Participants by Zip Code of Their School (If Applicable)

The above figures indicate that bicycle route data is achievable via Smartphone applications. The participants in this study made up a representative sample of the bicycling community in Austin.

Description of Bicycle Routes

This section presents the data analysis of the 2,820 bike trips matched to the network using the Dalumpines and Scott (2011) map-matching algorithm applied with line barriers located 200-ft on both sides of the bike trip GPS trace. The 200-ft buffer was used to constrain the area the algorithm searched to match the GPS traces. Since participants were not asked to report the actual route taken, the accuracy of the map-matching process is unknown. The best estimate of accuracy can be obtained by comparing the matched routes with the GPS trace for each bike trip (see Appendix F for examples comparing the matched bike trip with the actual GPS trace).

Table 3 presents descriptive statistics on the length of the 2,820 GPS traces matched to the network. Due to GPS signal interruptions, the original GPS traces had to be divided into shorter, continuous GPS traces for analysis. The results of the map-matching process are for those segments rather than for the entire GPS trace in most cases. Therefore, the mean and maximum presented are lower than the distance of the actual bike trips. In total, the routes matched to the network covered 8,528 miles.

Table 3. Distance Description of Matched Routes

n=2,820	Total Miles
Mean	3.02
Maximum	23.58
Sum	8527.60

Using a combination of spatial joins in ArcGIS 10.0 and queries and cross-tabulations in Microsoft Access, network, land use, railroad and self-reported user data were joined to the bike trips matched to the network.

User and Network Attributes

To demonstrate the variety of information extracted when joining the network and user data to the matched bike trips, a series of analyses were conducted to evaluate the route characteristics by gender, cycling frequency, trip purpose and age.

Gender and Road Class

Table 4 presents the percentages of the total route length by network segment type and gender. The male participant map-matched bike routes included more U.S. and/or state highways (Road Class 2) than females. A search for the bike trips matched to Road Class 2 showed Capital of Texas Highway, Ed Bluestein Boulevard, Ben White Boulevard, US Highway 183, and US Highway 290 as the roadways used. Most of the mileage occurred on Capital of Texas Highway, a popular highway for cyclists.

Table 4. Percentage of Route Length by Road Class and Gender

Road Class	Description	Males, Total Miles	% Within Males	Females, Total Miles	% Within Females
Original					
1	Interstate highway, expressway, or toll road	0.21	0.00%	0.00	0.00%
2	U.S. and/or state highway	132.93	2.08%	1.45	0.12%
3	n/a	0.62	0.01%	0.00	0.00%
4	County roads (RR, RM, FM, etc.) and/or major arterial	1402.53	21.94%	198.42	16.01%
5	Minor arterial	520.25	8.14%	92.51	7.46%
6	Local city/county street	1395.11	21.82%	268.51	21.66%
8	City collector	2188.09	34.23%	516.42	41.66%
10	Ramps and turnarounds	20.74	0.32%	0.67	0.05%
12	Driveway	74.01	1.16%	11.48	0.93%
14	Unimproved public road	0.64	0.01%	0.34	0.03%
15	Private road	28.25	0.44%	0.87	0.07%
New					
17	Off-street path in park	7.34	0.11%	9.29	0.75%
18	Off-street path on vacant or private property, or non-park open area	9.84	0.15%	1.45	0.12%
19	Parking lot	131.40	2.06%	10.16	0.82%
20	Multi-use path for pedestrians and cyclists	403.96	6.32%	90.79	7.32%
21	Building walkway	2.65	0.04%	0.00	0.00%
22	Paved sidewalk in park, campus, or open area	66.43	1.04%	28.46	2.30%
23	School track or designated veloway	2.63	0.04%	0.47	0.04%
25	Transit platform/station (e.g., Metrorail station area)	0.65	0.01%	0.00	0.00%
26	Alleyways	4.80	0.08%	8.36	0.67%
	Total	6393.1	100%	1239.7	100%

Gender and Road Description

The City of Austin network dataset used for this study contains a field that classifies roadways in a slightly different way (the classification excludes off-street facilities such as parks and parking lots). Table 5 presents the cross-tabulation of gender and the alternative roadway description, revealing a pattern similar to that seen in Table 4. The female participants, as a percentage of total route mileage, mostly traveled on local city, county, and neighborhood roads (63 percent), slightly more by percentage than male participants (56 percent).

Table 5. Percentage of Total Roadway Miles by Road Description and Gender

Road Description	Males Total Miles	% Males	Females Total Miles	% Females
No classification	775.52	12.13%	169.82	13.70%
A10—Interstate	0.21	0.00%	0.00	0.00%
A20—U.S. highways	128.85	2.02%	1.45	0.12%
A30—County roads and Austin Metropolitan Area Transportation Plan (AMATP), major arterial divided (MAD), and major arterial undivided (MAU)	1402.44	21.94%	198.42	16.01%
A31—Minor arterials	510.99	7.99%	91.01	7.34%
A40—Local city and county streets	1371.37	21.45%	261.87	21.12%
A45—Local neighborhood rural city streets and roads	2188.09	34.23%	516.42	41.66%
A63—Ramps, turnarounds, and cloverleaves	15.60	0.24%	0.67	0.05%
Total	6393.1	100%	1239.7	100%

Gender and Speed Limit

As expected because of the gender and road class cross-tabulation findings, a review of a cross-tabulation of gender with speed limits of the roadway facilities revealed that the female participants included lower speed facilities (35 mph or below) in their bike routes (76 percent) by route mileage more so than the male participants (67 percent), as seen in Table 6.

Table 6. Percentage of Total Route Length by Roadway Speed Limit and Gender

Speed (mph)	Males Total Miles	% Males	Females Total Miles	% Females
30 and Below	2007.99	31.41%	422.81	34.11%
35	2253.51	35.25%	522.97	42.19%
40	573.49	8.97%	95.76	7.72%
45	1182.83	18.50%	190.34	15.35%
50	190.32	2.98%	7.67	0.62%
55	97.68	1.53%	0.00	0.00%
60	62.65	0.98%	0.00	0.00%
65 and above	24.61	0.38%	0.095	0.01%
Total	6393.1	100%	1239.7	100%

Gender and Trip Purpose

The results in Table 7 show that the total female and male route mileage consisted mostly of commute trips. The only trip purpose where males and females differed significantly was in work-related trips.

Table 7. Percentage of Route Miles by Trip Purpose and Gender

Trip Purpose	Total Route Miles, Male	% Male	Total Route Miles, Female	% Female
Commute	4140	64.80%	765	61.70%
Errand	329.4	5.20%	72.9	5.90%
Exercise	441.4	6.90%	84.4	6.80%
Other	101.2	1.60%	37.4	3.00%
School	159.6	2.50%	58.5	4.70%
Shopping	221.3	3.50%	45.6	3.70%
Social	540.1	8.40%	171.8	13.90%
Work-Related	460.2	7.20%	4.1	0.30%
Total	6393.1	100.00%	1239.7	100.00%

Age and Road Class

Another user attribute joined with the bike trip and network data for analysis was age. Bike trips matched to highways revealed participants between the age of 35-50 and 25 and under. The bike trips from other age groups did not have routes on highways.

Figure 23 shows the distribution of route mileage within each age group for the five most popular road classes. As with highways (Road Class 1), the participants in the 25 and under and 35-50 age groups have a higher percentage of route miles on county roads or major arterials (Road Class 4) compared to other age groups.

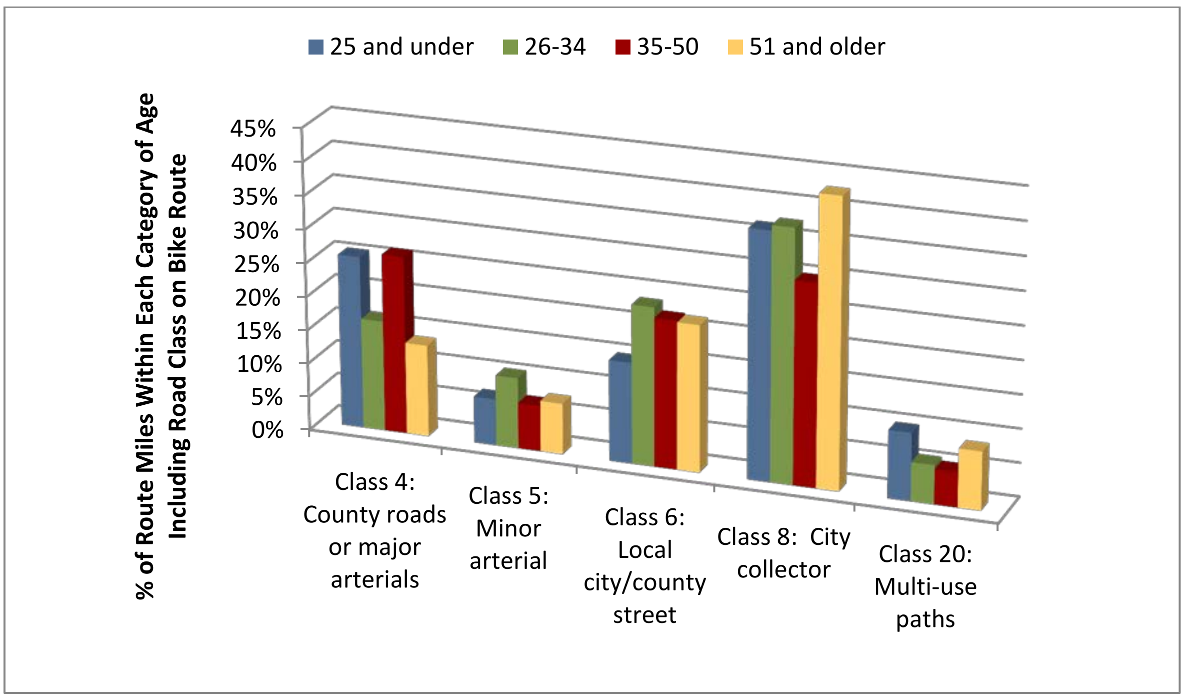


Figure 23. Percentage of Total Route Mileage on Road Classes for Each Age Category

Cycling Frequency and Road Class

Analysis of the data revealed the possibility that the choice of route may depend on how frequently the participants ride their bikes. Figure 24 illustrates the results of the cross-tabulation of cycling frequency with road class for the five most popular road classes, 4-8 and 20 (refer to Table 4 for a description of each road class).

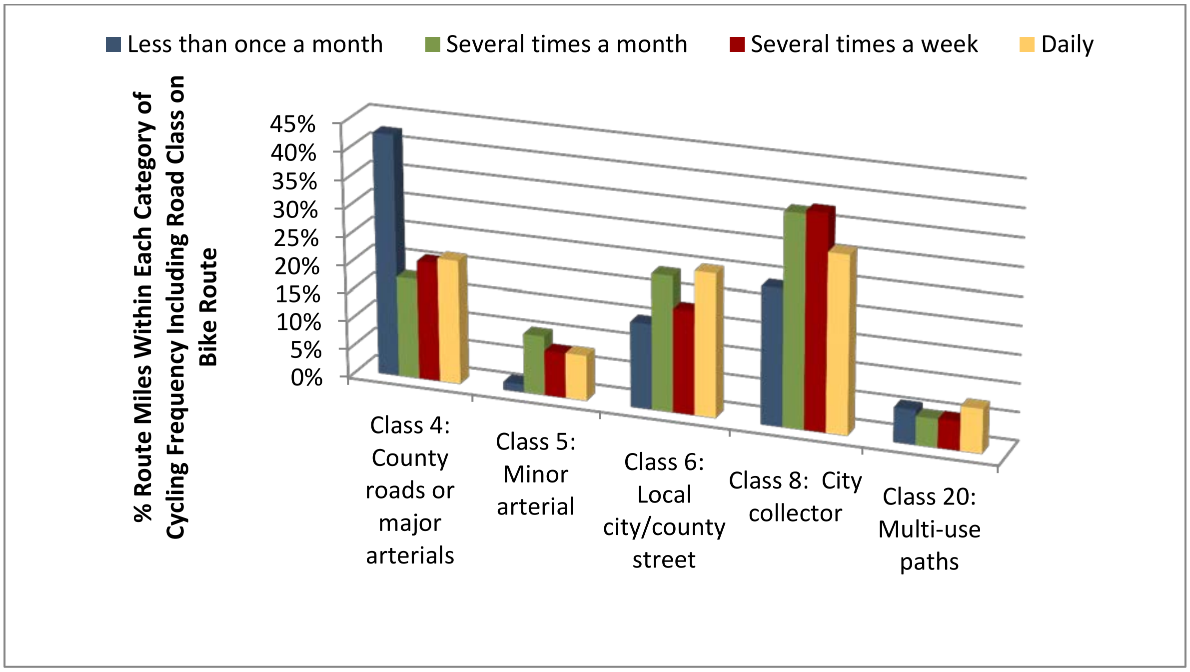


Figure 24. Road Class Choice by Cycling Frequency

Interestingly, for those that cycled less than once a month, route mileage consisted primarily of major arterials and county roads (Road Class 4) and city collectors (Road Class 8), however the total mileage of those cycling less than once a month was small (116.8 mi) compared to the other frequency categories (see Table 8). Only 5percent of the participants indicated bicycling less than once a month. About the same percentage of the cyclist route miles are on multi-use paths (Road Class 20) across all age groups.

Though not shown in Figure 24, participants riding daily or several times a week included more highways as a percentage of total mileage within each category of cycling frequency (Road Class 1), suggesting more frequent cyclists become more comfortable riding with higher motor vehicle volumes and speeds. An alternate explanation could be a cyclist may bike frequently because of a lack of transportation alternatives, therefore more likely to have to ride on the higher-speed, higher-volume roadways. Sidewalk riding may be more prevalent on these roadways but cannot be verified using the GPS traces.

Table 8. Total Route Miles by Cycling Frequency

	Less than once a month	Several times a month	Several times a week	Daily
Total miles	116.80	644.80	3703.11	3530.77

Land Use

The City of Austin offered a GIS shapefile consisting of polygons outlining the collection of parcels sharing the same land use code. Table 9 lists the land use codes, the description of the land use, the percentage of that land use in the Austin land area and the percentage of route miles within each gender for each land use within 200 feet of the routes. Joining the land use data to the matched bike trips helps characterize the areas along the length of the trip. Large percentage differences between male and female participants were not evident; most bicycled near single-family residential land uses.

Table 9. Land Use Statistics

Land Use Code	Land Use Description	Percentage of Total Land Area in Austin, Texas	Percentage of route miles within 200 feet of land use	
			% Males	% Females
100	Single Family	12.69%	61.17%	57.03%
113	Mobile Home	2.21%	0.21%	0.55%
150	Duplexes	0.30%	5.23%	5.68%
160	Large-lot Single Family	6.87%	0.02%	0.05%
210	Three/Fourplex	0.04%	0.54%	0.66%
220	Apartment/Condo	1.07%	3.61%	4.38%
230	Group Quarters	0.02%	0.20%	0.35%
240	Retirement Housing	0.04%	0.07%	0.08%
300	Commercial	1.60%	8.69%	8.08%
400	Office	0.69%	5.36%	6.61%
510	Manufacturing	0.39%	0.39%	0.16%
520	Warehousing	0.58%	1.01%	1.31%
530	Miscellaneous Industrial	0.24%	0.13%	0.60%
560	Resource Extraction (Mining)	0.75%	0.01%	0.02%
570	Landfills	0.22%	0.00%	0.00%
610	Semi-institutional Housing	0.05%	0.02%	0.01%
620	Hospital	0.02%	0.02%	0.01%
630	Government Services	0.21%	0.27%	0.29%
640	Educational	0.74%	0.49%	0.54%
650	Meeting and Assembly	0.43%	1.15%	1.65%
670	Cemeteries	0.06%	0.07%	0.05%
680	Cultural Services 7	0.01%	0.04%	0.06%
710	Parks/Greenbelts	2.53%	2.95%	2.36%
720	Golf Courses	0.66%	0.02%	0.02%
730	Camp Grounds	0.06%	0.00%	0.00%
750	Preserves	6.91%	0.10%	0.00%
810	Railroad Facilities	0.07%	0.33%	0.31%

Table 9. Land Use Statistics (Continued)

820	Transportation Facilities	0.02%	0.01%	0.02%
830	Airports and Aviation Facilities	0.42%	0.00%	0.00%
840	Marinas	0.02%	1.76%	2.06%
850	Parking	0.07%	1.25%	1.15%
860	Streets and Roads	6.09%	0.32%	0.48%
870	Utilities	0.44%	4.20%	4.78%
900	Undeveloped	13.13%	0.25%	0.48%
910	Agricultural	38.11%	0.12%	0.14%
940	Water	2.26%	0.00%	0.00%
999	Unknown	0.02%	0.00%	0.00%

Railroads

The Austin area does not have many railroads; however, there are numerous roadway/railroad intersections cyclists cross. An intersect spatial join was conducted to determine how many of the cycling routes intersect railroads. Figure 25 illustrates some of the railroad crossings with a red mark. Some of the red marks running east/west through downtown are probably not bicycle crossings because the Lance Armstrong Bikeway provides a way to travel along the side of the rail tracks. Additionally, the crossings on Mopac are most likely over or under the railroad.

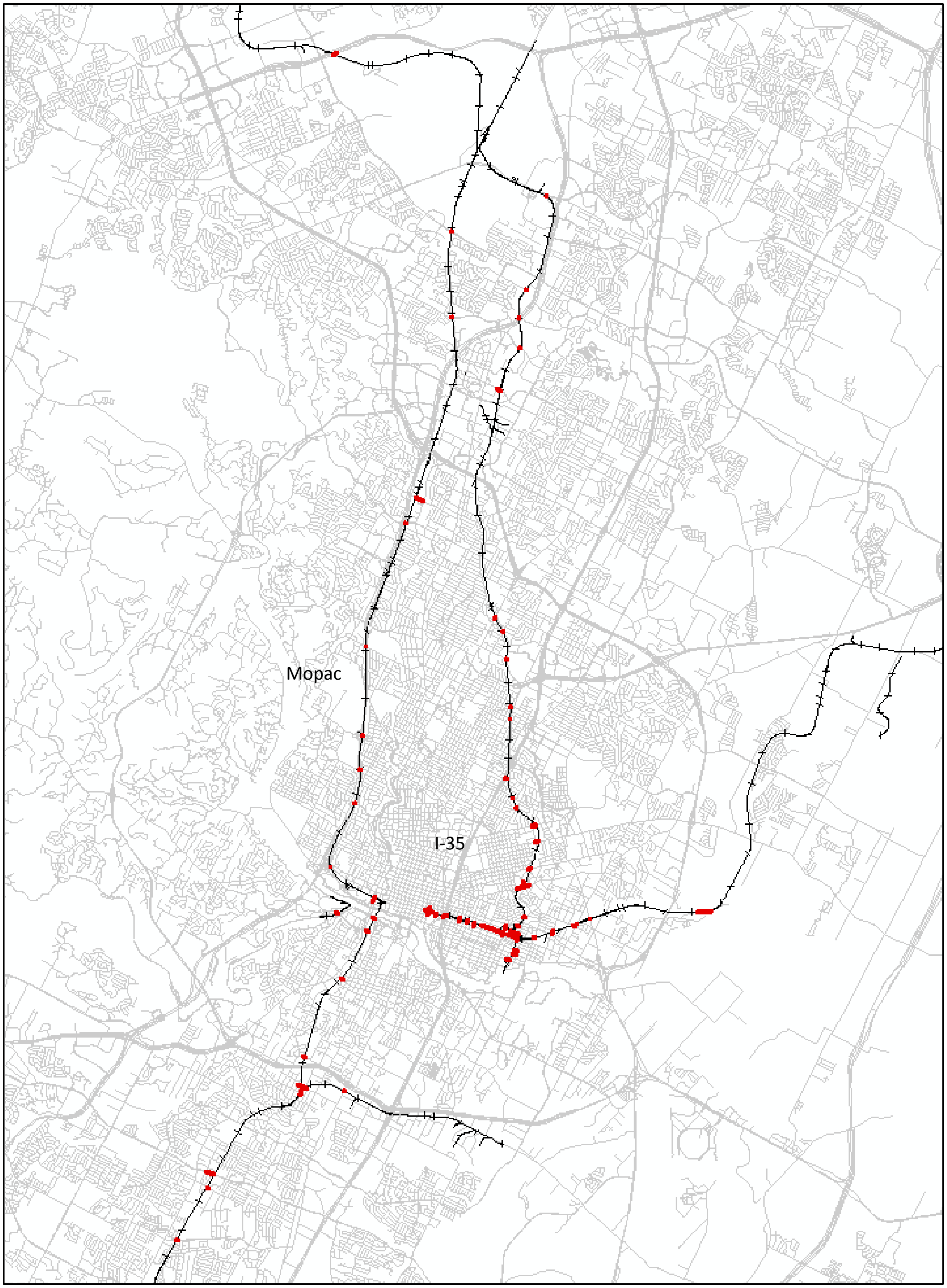


Figure 25. Bicycle Trip and Railroad Intersection (Crossings—Red; Railroad Tracks—Black)

The use of GIS makes it possible to identify which railroad tracks cyclists crossed the most. This useful information helps identify which crossings to examine for potential safety improvements. A spatial join of the railroad line features with the matched bike routes indicated 2,180 railroad crossings, although some crossings were made above or below the railroad (e.g., Mopac, West Riverside Drive, and Barton Springs Road) or were likely not crossings but sections adjacent to the railroad (such as along the Lance Armstrong Bikeway downtown). Table 10 shows the crossing locations where 50 or more crossings were made by the participants.

The highest number of crossings occurred on East 4th Street near I-35 east of downtown. Figure 26 shows an aerial image of one of several East 4th Street crossings. The railroad belongs to Capital Metropolitan Transportation Authority (CapMetro), and its Metro Red Line Commuter Train operates on the tracks. The second most frequently crossed railroad track is located on East 12th Street, south of the MLK Station near Walnut Avenue for the Capital Metro Red Line (see Figure 27).

To validate this method of data collection with a smartphone, one can compare the time and effort involved in identifying heavily used bicycle/rail crossings with traditional methods. Manually collecting the volume of bicyclists at railroad crossings in Austin would have been quite time consuming. A quick search of the matched route data in ArcGIS answered the question of which railroad crossings had the most number of cyclist crossings by those participating in the CycleTracks study.

Table 10. Count of Bike Trips Crossing Railroads By More Than 50 Matched Bike Trips

Roadway Crossing Railroad Tracks	**Number of Trips Crossing**
E 4th St (near and just east and west of I-35)	197
E 12th St (just south of the MLK Red Line station near Walnut Avenue)	63
E 51st St (at Airport Blvd)	56
West Rundberg Lane (near Metric Blvd)	56
Vinson Dr. (near Philco Road and St. Elmo Road)	55

Figure 26. Railroad Crossing at East 4th Street at I-35

Figure 27. Railroad Crossing at East 12th Street

City of Austin Bike Network

The City of Austin assigned bicycle user ratings to city-designated routes and other roadway links (see Figure 23). Table 11 defines most of the rating acronyms and presents the percentage of user ratings in the bicycle system. One of the project tasks identified in the proposal stage was looking at the Bicycle Compatibility Index of street segments used by participants. However, the City of Austin Bicycle Program staff suggested that their rating system may give a better condition of route characteristics.

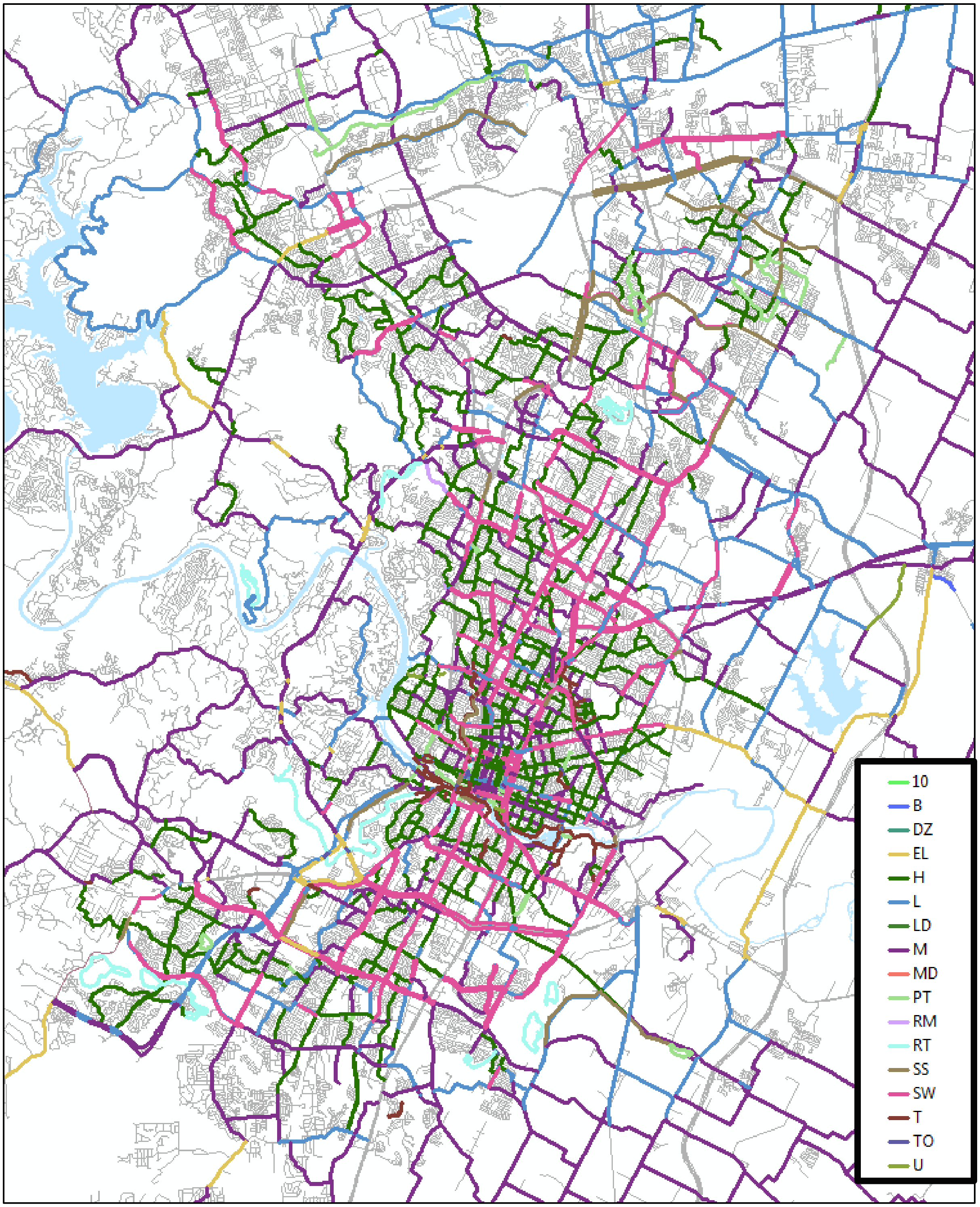


Figure 28. Excerpt of Map Showing Cyclist User Ratings for Austin, Texas, Area

Table 11. Cycling User Rating by Percentage of All Rated Segments

User Rating	Description from City of Austin	Miles	Percent of Total
None	No rating assigned	107.50	1.80%
EL	Extremely Low Ease of Use	143.52	2.41%
H	High Ease of Use	431.54	7.23%
L	Low Ease of Use	1286.84	21.57%
LD	Dirt Road—Low Ease of Use	5.31	0.09%
M	Medium Ease of Use	2139.00	35.85%
MD	Dirt Road—Medium Ease of Use	25.29	0.42%
PT	Paved Trail	32.12	0.54%
SS	Helpful Sidewalks	51.71	0.87%
SW	Sidewalks along "Low Ease of Use" Major Street	289.60	4.85%
T	Unpaved Trail	24.91	0.42%
U	Unrated	1374.75	23.04%
Total Miles		5967.08	

Note: Other rating abbreviations seen in Figure 28 are not shown in this table since they are not described and do not represent more than 1 percent of the total segments.

Unfortunately, the bicycle route data could not be matched to the street network, and therefore the GPS points were not matched. As can be seen in Figure 29, some of the bicycle routes were offset from the roadway centerlines, and even beyond the right-of-way (ROW) of the roadway. A less precise spatial join, using a 100-ft search radius, was attempted to capture the most likely user rating for a network link. Using a small radius, however, did not capture some of the bicycle user rating line features parallel to a network link. The obvious alternative, using a larger radius, would have contributed to additional error since the spatial join would have picked up bicycle user ratings for other irrelevant network links. Ideally, the City of Austin GIS feature dataset for link user rating should be tied directly to the City of Austin's network dataset, rather than being a feature dataset with separate line features and field values.

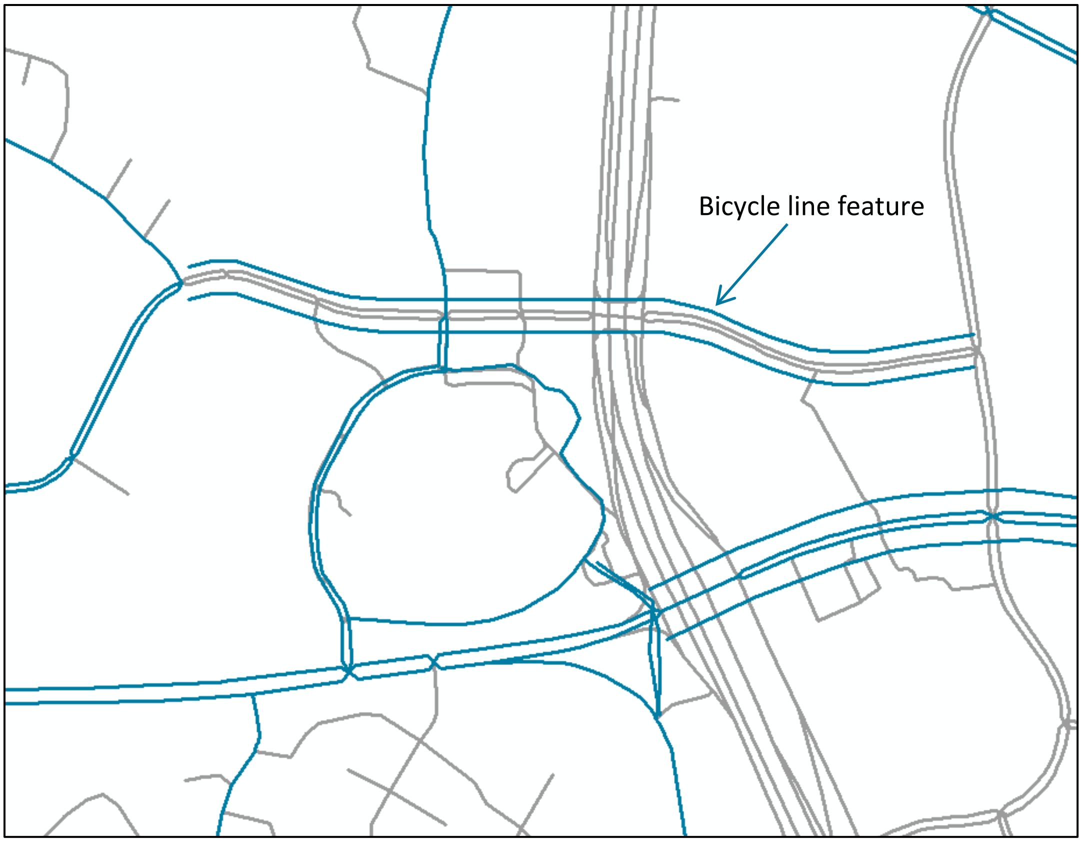


Figure 29. City of Austin Bicycle Routes (Blue Lines) and the Network Link Dataset (Gray Lines), Northwest Austin (183/Mopac)

RECOMMENDATIONS

- Structure feature datasets with network link information to correspond to the network dataset (so a one-to-one match can be made between the network link and the attributes of interest to avoid including intersecting line features representing links not along the route).

Continued study of the CycleTracks data should include joining additional attribute information, such as elevation, link traffic counts, air quality, crime, and bus routes, to the user and matched bike route data. Additional user data available from the CycleTracks dataset that would also be useful for analysis includes the day (e.g., weekday or weekend) and time of travel (e.g., day or night). Hood's (2010) analysis of the San Francisco CycleTracks data included attributes assigned to each link of the network, such as length, traffic volume, free-flow speed, number of lanes, bike facility class, up-slope, and crime rate.

The analysis of the matched bike trip data must come with the disclaimer that the map-matching process may not have accurately mapped the GPS traces to the network (e.g., the routes matched may not necessarily be the route actually taken). Additional work continues in the field of map-matching so that the highest number of trips is matched to the most likely network links.

The analysis in this report is best thought of as providing an example of the type of data analysis available to agencies that use the CycleTracks application for collecting cycling route choice information. Despite the manageable challenges of data cleaning, network completion, and map-matching, the amount of information provided by use of CycleTracks far exceeds what would be available using other data collection methods.

Conclusion

This report summarizes the outcome of the data collection pilot program in Austin, Texas, using the smartphone app called CycleTracks. The purpose of this project was to demonstrate the usefulness of gathering bike route choice data from GPS units in smartphones. Key to using the information was being able to process the data. Methods of processing the CycleTracks data were developed using ArcGIS because, though proprietary, it has a high market share and is software commonly used by local governments. Other methods were tested as well, such as Google Earth and MATLAB. The research was not the first to collect and analyze bicycle route information with the CycleTracks smartphone application; however, it was the first for the Austin area and the first to use ArcGIS software to develop a means to generate routes from the network and analyze the resulting cycling routes.

Since GPS traces contain random and systematic errors, the data collected from the CycleTracks application requires extensive cleaning and a map-matching algorithm able to match the GPS traces to the network as accurately as possible. Additional testing of map-matching algorithms using the ArcGIS software is recommended to be able to develop an algorithm with a high success at matching GPS traces to a network. Also key to projects like these is having a complete network.

The analysis of the routes matched to the network using the Dalumpines and Scott (2011) algorithm was found to be the best method for agencies using the CycleTracks application for collecting bicycle route choice data. Researchers spent significantly more time than was originally expected on data cleaning, completing the network, and map-matching. Nevertheless, as mentioned previously the potential route data and the wide variety of uses for the data made available through CycleTracks or other smartphone application far surpasses the challenges faced when considering other data collection alternatives.

The dataset from this project provides a wealth of information worthy of additional analyses. First, further analysis of revised and different map-matching methods is recommended and should incorporate additional data to join to the routes for a better understanding of routes chosen. This research should include the preparation of other datasets, such as the bicycle user rating GIS feature dataset, in a form more useful for analysis with the CycleTracks data. Second, another potential avenue for research is examining how well the routes chosen match up with bicycling level of service criteria and higher user ratings. This comparison would be useful in determining whether the bicycle level of service method is beneficial for making investment decisions.

Future extensions of this research could include modifying the CycleTracks application to match the GPS trace in real time. This could potentially reduce the data cleaning effort and save time compared to batch processing later. A disadvantage of real-time map-matching is the potential of incorrect mapping because real-time algorithms do not have the benefit of knowing the final trip outcome and must estimate the position with less information.

With the growing ubiquity of smartphones owned by individuals and the increased use of ArcGIS at local agencies, the two applications provide the means for efficiently gathering information about route choices made by cyclists. This invaluable information will help communities understand who is cycling where in order to determine what types of facilities are most appropriate. Improving accommodations for bicyclists will encourage use of one of the most efficient and sustainable forms of transportation thereby improving public health and traffic congestion.

References

Alliance for Biking and Walking. (2012). *Bicycling and Walking in the U.S.: 2012 Benchmarking Report*. http://www.peoplepoweredmovement.org/site/index.php/site/memberservices/2012_benchmarking_report/. Last accessed June 25, 2012.

Aultman-Hall, L., Hall, F., & Baetz, B. (1997). Analysis of Bicycle Commuter Routes Using Geographic Information Systems: Implications for Bicycle Planning. *Transportation Research Record: Jounral of the Transportation Research Board, 1578*, 102-110.

Bradley, M. (1988). Realism and Adaptation in Designing Hypothetical Travel Choice Concepts. *Jounral of Transport Economics and Policy, 22*(1), 121-137.

Broach, J., Gliebe, J., & Dill, J. (2011). Bicycle Route Choice Model Developed Using Revealed Preference GPS Data. *90th Transportation Research Board Annual Meeting Compendium of Papers*. Washington, D.C.

Charlton, B., Schwartz, M., Paul, M., Sall, E., & Hood, J. (2010). CycleTracks: A Bicycle Route Choice Data Collection Application for GPS-Enabled Smartphones. *Presented at the 3rd Conference on Innovations in Travel Modeling, a Transportation Research Board Conference.* Tempe, AZ, May 10-12, 2010.

Dalumpines, R., & Scott, D. M. (2011). GIS-Based Map-Matching: Development and Demonstration of a Post-Processing Map-Matching Algorithm for Transportation Research. *Transportation Research Board 2011 Annual Meeting Compendium of Papers.*

Dill, J., & Gliebe, J. (2008). *Understanding and Measuring Bicycle Behavior: A Focus on Travel Time and Route Choice Final Report OTREC-RR-08-03.* Oregon Transportation Research and Education Consortium.

Duncan, M. J., & Mummery, W. K. (2007). GIS or GPS? A Comparison of Two Methods for Assesing Route Taken During Active Transport. *American Journal of Preventative Medicine, 33*(1), 51-53.

ESRI. (2012). *ESRI TomTom Road Network*. Retrieved from http://www.arcgis.com/home/item.html?id=f38b87cc295541fb88513d1ed7cec9fd

Hood, J. N. (2010). *A GPS-Based Bicycle Route Choice Model for San Francisco, California.* Professional Report, University of California, Berkeley, Berkeley, California.

Howard, C., & Burns, E. (2001). Cycling to Work in Phoenix: Route Choice, Travel Behavior, and Commuter Characteristics. *Transportation Research Record: Journal of the Transportation Research Board, 1773*, 39-46.

Hunt, J. D., & Abraham, J. E. (2007). Influences on Bicycle Use. *Transportation, 34*(4), 453-470.

Hyodo, T., Suzuki, N., & Takahashi, K. (2000). Modeling of Bicycle Route and Destination Choice Behavior for Bicycle Road Network Plan. *Transportation Research Record: Journal of the Transportation Research Board, 1705*, 70-76.

Krizek, K. J. (2006). Two Approaches to Valuing Some of Bicycle Facilities' Presumed Benefits. *Journal of the American Planning Association, 72*(3), 309-320.

Krizek, K. J., El-Geneidy, A., & Thompson, K. (2007). A Detailed Analysis of How an Urban Trail System Affects Cyclists' Travel. *Transportation, 34*(5), 611-624.

Landis, B. W., Vattikuti, V. R., & Brannick, M. T. (1997). Real-Time Human Perceptions: Toward a Bicycle Level of Service. *Transportation Research Record: Journal of the Transportation Research Board, 1578*, 119-126.

Landis, B. W., Vattikuti, V. R., Ottenberg, R. M., Petritsch, T. A., Guttenplan, M., & Crider, L. B. (2003). Intersection level of service for the bicycle through movement. *Transportation Research Record: Journal of the Transportation Research Board*, 101-106.

Marchal, F., Hackney, J., & Axhausen, K. W. (2005). Efficient Map Matching of Large Global Positioning System Data Sets: Tests on Speed-Monitoring Experiment in Zurich. *Transportation Research Record: Journal of the Transportation Research Board, 1935*, 93-100.

Menghini, G., Carrasco, N., Schussler, N., & Axhausen, K. (2009). Route Choice of Cyclists: Discrete Choice Modelling Based on GPS-Data. *Arbeitsberichte Verkehrs-und Raumplanung, 544*.

Menghini, G., Carrasco, N., Schussler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A, 44*, 754-765.

Moudon, A. V., & Lee, C. (2003). Walking and bicycling: An evaluation of environmental audit instruments. *American Journal of Health Promotion, 18*(1), 21-37.

NAVTEQ. (2012). *NAVTEQ Networks*. Retrieved from http://www.navteq.com/

Newsom, P., & Krumm, J. (2009). Hidden Markov Map Matching Through Noise and Sparseness. *ACM GIS 2009.* Seattle, WA.

Nielsen, O. A., Wurtz, C., & Jorgensen, R. M. (2004). Improved map-matching algorithms for GPS-data-methodology and test on data from the AKTA road pricing experience in Copenhagen. *19th European Conference for ESRI Users, November 2004.* Copenhagen.

Pyo, J.-S., Shin, D.-H., & Sung, T.-K. (2001). Development of a map matching method using the multiple hypothesis technique. *2001 IEEE Intelligent Transportation Systems Conference Proceedings* (pp. 23-27). Oakland, CA: IEEE.

Quddus, M. A., Ochieng, W. Y., & Noland, R. B. (2007). Current map-matching algorithms for transport applications: State-of-the-art and future research directions. *Transportation Research Part C, 15*, 312-328.

Schuessler, N., & Axhausen, K. W. (2009). *Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT).* Swiss Federal Institute of Technology Zurich.

Schuessler, N., & Axhausen, K. W. (2009b). Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board, 2105*, 28-36.

Sener, I. N., Eluru, N., & Bhat, C. R. (2009). An Analysis of Bicycle Route Choice Preferences in Texas, U.S. *Transportation, 36*(5), 511-539.

Shafizadeh, K., & Niemeier, D. (1997). Bicycle Journey-to-Work: Travel Behavior Characteristics and Spatial Analysis. *Transportation Research Record: A Journal of the Transportation Research Board, 1578*, 84-90.

Sorton, A., & Walsh, T. (1994). Bicycle Stress Level as a Tool to Evaluate Urban and Suburban Bicycle Compatibility. *Transportation Research Record: Journal of the Transportation Research Board, 1438*, 17-24.

Stinson, M. A., & Bhat, C. R. (2003). Commuter Bicyclist Route Choice: Analysis Using a Stated Preference Survey. *Transportation Research Record: Journal of the Transportation Research Board, 1828*, 107-115.

Tilahun, N. Y., Levinson, D. M., & Krizek, K. J. (2007). Trails, Lanes, or Traffic: Valuing Bicycle Facilities with an Adaptive Stated Preference Survey. *Transportation Research Part A- Policy and Practice, 41*(4), 287-301.

Turner, S. M., Shafer, C. S., & Stewart, W. P. (1997). *Bicycle Suitability Criteria: Literature Review and State-of-the-Practice Survey.* Texas Department of Transportation Report Number TX-97/3988-1.

U.S. Census Bureau. (2012). Retrieved from http://www.census.gov/geo/www/tiger/

U.S. Department of Transportation (U.S. DOT). (2011). *DOT Livability*. Retrieved February 2011, from http://www.dot.gov/livability

Zhou, J. (., & Golledge, R. (2006). *A Three-Step General Map Matching Method in the GIS Environment: Travel/Transportation Study Perspective.* Manuscript for International Journal of Geographical Information Systems.

Zielstra, D., & Hochmair, H. H. (2012). *An Introduction to Freely Available Street Network Data*. Retrieved May 2012, from The Institute of Food and Agricultural Sciences (IFAS) University of Florida: http://edis.ifas.ufl.edu/fr343

Appendix A: Trip Data from CycleTracks

The uploaded trips from each user were stored in a database (MySQL) on the SFCTA server. This database contained three tables: Persons, Trips, and GPS Coordinates. The trip upload process from a phone typically took just a few seconds. To identify each user uniquely, the application collected the International Mobile Equipment Identity (IMEI) number from the phone. This number was only used to assign a unique user ID to the phone user so that all the uploaded trips from that phone could be tagged as that user ID. The IMEI number was discarded in the final database and was not shared with anyone to respect and protect user privacy. Figure 31 from the Transportation Research Board (TRB) presentation by Charlton et al. (2010) describes each field in the database in detail.

Table 1: Person Table

User ID	Numeric identifier for the person record
Created	Creation date/time for this user record
Device ID	IMEI number of phone hardware (stripped from final data)
Home ZIP	Home ZIP code*
School ZIP	School ZIP code*
Work ZIP	Primary workplace ZIP code*
Gender	Male/Female*
Age	Age in years of this person*
Biking Frequency	Daily / Several per week / Several per month / <1 per month*
Email	Email address for raffle & future contact*

* Fields with an asterisk are optionally provided by user

Table 2: Trip Table

Trip ID	To match to GPS Coordinates Table
User ID	To match to Person Table
Start Time	Time stamp for when user taps "start"
End Time	Time stamp for when user taps "stop"
Number of Coordinates	Number of non-null coordinates in this trip
Trip Purpose	Selected and confirmed by user (see Figure 1c)

Table 3: GPS Coordinates Table

Trip ID	To match to Trip Table
Time	Time stamp of when the GPS coordinate was taken. In addition to travel time and speed calculations, it is used to order points to determine route.
Latitude	Latitude and Longitude are stored with ten decimal points
Longitude	
Altitude	In meters
Estimated Accuracy	Provided by GPS system, in meters

Figure 30. Database Structure for CycleTracks Application Server
(Source: Bicycle Route Choice Data Collection—TRB Presentation, Charlton et al. [2010])

Appendix B: Media Coverage

Texas Transportation Institute

TTI Communications
979/845-1734
FAX 979/845-7575
http://tti.tamu.edu

MEDIA ADVISORY

Study Hopes to Improve Austin Biking

Project Seeks Participation from Bicyclists

Researchers from The University of Texas and Texas A&M University are hoping volunteers with smart phones can help them gather better data for improving bicycle routes and services in the Austin area.

Using CycleTracks, a free smart phone application available for iPhones and Android-based phones, researchers from the Texas Transportation Institute (TTI) and the UT Center for Transportation Research will be able to track actual usage data for bike routes throughout the Austin area.

“The GPS system will allow us to gather information about bicyclists and their preferred bike routes,” says TTI Associate Engineer Joan Hudson. “By taking part in the Austin-based program, users are providing valuable data that tells us where we need to focus improvements. We will also be able to determine how effective this satellite data collection is, compared to traditional information gathering techniques such as surveys.”

The study coincides with the annual Bike to Work Week, which begins May 16, and Bike to Work Day on May 20.

Bicyclists who choose to participate can download the CycleTracks app directly to their phone. Researchers will gather rider information through the month of June. Participation is anonymous unless the user opts to provide personal information such as age, gender, zip code, and email address. Providing this information will allow the research team to compare participant demographics to regional demographic data.

Developed by the San Francisco County Transportation Authority, CycleTracks allows the bicyclist to view a list of saved trips with maps and simple statistics including distance, time and average speed.

Further information is available by emailing Joan Hudson at j-hudson@tamu.edu or by visiting the project website, http://www.cycletracksaustin.com.

Additional Contact:
Rick Davenport
Texas Transportation Institute
979-862-3763

CYCLISTS' TRAVEL PATTERNS MINED FOR RESEARCH DATA

Daily Texan

Published 9 Jun 2011 at 7:47 PM

By Diego Cruz

Researchers are tracking Austin cyclists to better understand how they travel through the city.

A student bikes down Dean Keeton Street on Thursday morning. Researchers from the University of Texas and Texas A&M are gathering bicycle route data for a smartphone application called CycleTracks.

The application CycleTracks records routes cyclists take using the GPS function of their smartphones, according to the study's website. Users also have the option to indicate the purpose of each bicycle trip.

Texas A&M University's Texas Transportation Institute is conducting a study along with UT's Center for Transportation Research using the application in Austin.

Billy Charlton, deputy director for technological services at the San Francisco County Transportation Authority, said the organization created and released CycleTracks in 2009. He said the application gathered data the county used to develop a computer model to determine the pros and cons of using road space for bicycle lanes.

"The main reason we wanted this data is to predict which streets cyclists would prefer," Charlton said.

Joan Hudson, an associate research engineer at A&M's transportation institute and the principal investigator for the project in Austin, said the researchers are collecting data through May and June, and will analyze it in the following months.

"Having knowledge about where people ride can certainly help in improving routes," Hudson said.

She said facilitating transportation for bicyclists could decrease congestion for motor vehicles, and possibly convince some people to switch from cars to bicycles.

The application collects demographic data, including age and sex, to determine trends, but Hudson said the privacy of participants is completely assured. She said the data will also help determine whether the sample taken is biased toward higher income groups because the information is collected through smartphones.

Jennifer Duthie, a research fellow at UT's Center for Transportation Research, said the goal of the study is to determine the role smartphones could play in collecting bicycle route data for transportation planners.

"While this study focuses on bicyclists, there is the greater question of how smartphone and GPS devices can change the way data is collected not only for transportation planning but a whole host of applications," Duthie said.

She said San Francisco County found CycleTracks to be an effective and inexpensive method for collecting bicycle route data useful for planning, and hoped it would mean the same for Austin.

Sean Hall, a rhetoric and writing, premed, predental and preveterinary senior, rides his bicycle regularly and said knowing which routes to take determines how safe the trip is.

"The worst is when you have a one-lane, no-shoulder lane that has high-speed traffic," Hall said.

He said downtown and central Austin provide the safest routes with more bicycle lanes and lower speed limits.

Matt Welch, communications director for the Austin Cycling Association, said Austin is cycling-friendly and getting better every day, but it is not perfect. Austin placed 11th in Bicycling Magazine's most recent ranking of the top 50 bike-friendly cities in the U.S.

"We've got work to do on infrastructure planning, on creating safe routes for commuters, on improving our recreational amenities and basically just transforming Austin into a completely cycling-friendly city," Welch said.

Appendix C: Importing GPS Data into ArcGIS

The procedure followed for importing the GPS point data into ArcGIS 10.0 is outlined in the ESRI ArcGIS technical article titled "HowTo: Import XY data tables to ArcMap and convert the data to a shapefile" (http://support.esri.com/en/knowledgebase/techarticles/detail/27589). ArcGIS offers several geodatabase options. The personal geodatabase size limit proved too small for this project, so a file geodatabase was used instead.

The CycleTracks application recorded the location by GPS using the WGS 1984 datum (the reference frame for the geographic coordinates). Since the City of Austin file used for the data analysis was NAD 1983 FTIP 4203, the GPS data were first transformed from WGS 1984 to NAD 1983 and then projected to the City of Austin FTIP 4203 using the NAD_1983_To_WGS_1984_5 transformation option within and recommended by ArcGIS 10.0. All ArcGIS files used for the project were projected to the following:

Projected Coordinate System: NAD_1983_StatePlane_Texas_Central_FIPS_4203_Feet
Projection: Lambert_Conformal_Conic
False_Easting: 2296583.33333333
False_Northing:9842500.00000000
Central_Meridian: -100.33333333
Standard_Parallel_1: 30.11666667
Standard_Parallel_2: 31.88333333
Latitude_Of_Origin: 29.66666667
Linear Unit: Foot_US

Geographic Coordinate System: GCS_North_American_1983
Datum: D_North_American_1983
Prime Meridian: Greenwich
Angular Unit: Degree

Appendix D: Categories of Map-Matching Procedures

The map-matching procedure (i.e., algorithm) used the GPS points to determine what route the cyclist took. Map-matching can be done during the GPS collection (real time) or after data collection (batch or post-processing). In the latter case, the map-matching benefits from having knowledge of both the origin and destination and the entire GPS-traced path between those points; this was the case for the CycleTracks data collected for the Austin project.

Schuessler and Axhausen (2009), developers of an advanced map-matching algorithm used by Hood (2010) for the study of CycleTracks data collected in San Francisco, classified the different map-matching procedures into three categories:

- Geometric procedures.
- Topological procedures.
- Advanced procedures.

Hybrid algorithms consisting of geometric and topological procedures were used for the analysis of the CycleTracks data collected in Austin for this research report. For background, this appendix briefly describes each of the three types of procedures and their use and applicability for this project. Zhou and Golledge (2006) provided a review of the evolution of the different approaches to matching travel routes to the network from the simpler geometric procedure of snapping GPS points to the closest network links to more advanced algorithms that take into account the GPS points over time and the feasible connectivity of the route on the network. Quddus et al. (2007) also provided an overview of different map-matching approaches, with particular emphasis on real-time map-matching.

Hood's (2010) map-matching of the CycleTracks GPS traces gathered in San Francisco used the advanced approach developed by Schuessler and Axhausen (2009). Hood (2010) noted that after processing, only 1,454 traces from 260 participants successfully matched to the network (presumably from the starting number of 2,282 traces), revealing the difficulties of finding a map-matching algorithm that will successfully match the GPS points or traces to the network.

Geometric Procedures

The geometric procedures can generally be further divided into how GPS data match to a network:

- GPS point to network point (e.g., vertex and node or point-to-point).
- GPS point to network line (i.e., point-to-curve).
- GPS line to network line (i.e., curve-to-curve).

Geometric approaches match routes according to the distance of the GPS points to the nearest network link/node or to the shape of the GPS trace and the shape of the network (e.g., curve-to-curve matching). The simplest example of a geometric approach consists of snapping the GPS points to the nearest street network link and/or nodes; however, the resulting routes typically contain errors. For instance, a series of GPS points along a road that then cross an intersecting road will snap to the intersecting road rather than the road the GPS points were following. In addition, snapping of GPS points to the closest network links could result in the formation of infeasible routes. In a study of car trips tracked by GPS, the use of the snapping approach resulted in 33 percent of the 3,000 trips being incorrectly matched to the 300,000 link network (Nielsen et al., 2004). Newsom and Krumm (2009) illustrated this problem well, as demonstrated in Figure 32 from their article, by matching measured locations to the road network in order to infer the vehicle's actual path instead of matching to the nearest roadway.

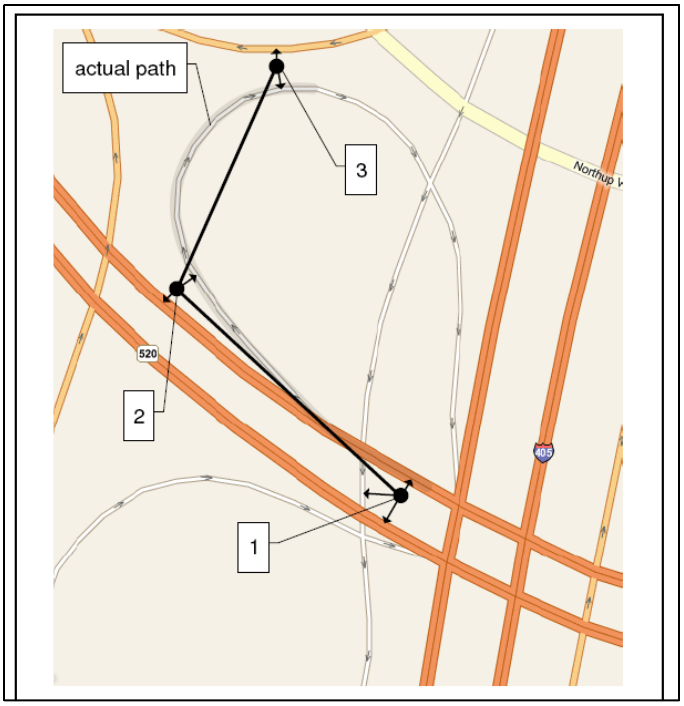


Figure 31. Problems with Matching to the Nearest Road Only (Newsom and Krumm, 2009)

Geometric algorithms that take into account the direction of the sequence of GPS points seek to select network links that closely follow the directional movement of the GPS trace. For instance, the San Francisco CycleTracks data were matched by Hood (2010) to the network using an algorithm developed by Schuessler and Axhausen (2009) that considered, among other criteria, the angle of changes in the sequence of GPS points to determine whether a trip reached the end of a network link (Figure 33). Implementation of the algorithm was done in the Java programming language. Efforts to implement this criteria into ArcGIS for this research project were considered; however, additional time is needed to determine how to efficiently incorporate this procedure fully within ArcGIS. Zhou and Golledge (2006) noted, though, that directional algorithms can be problematic for GPS traces for vehicles moving slowly or making brief stops, situations applicable to GPS tracking of bicyclists.

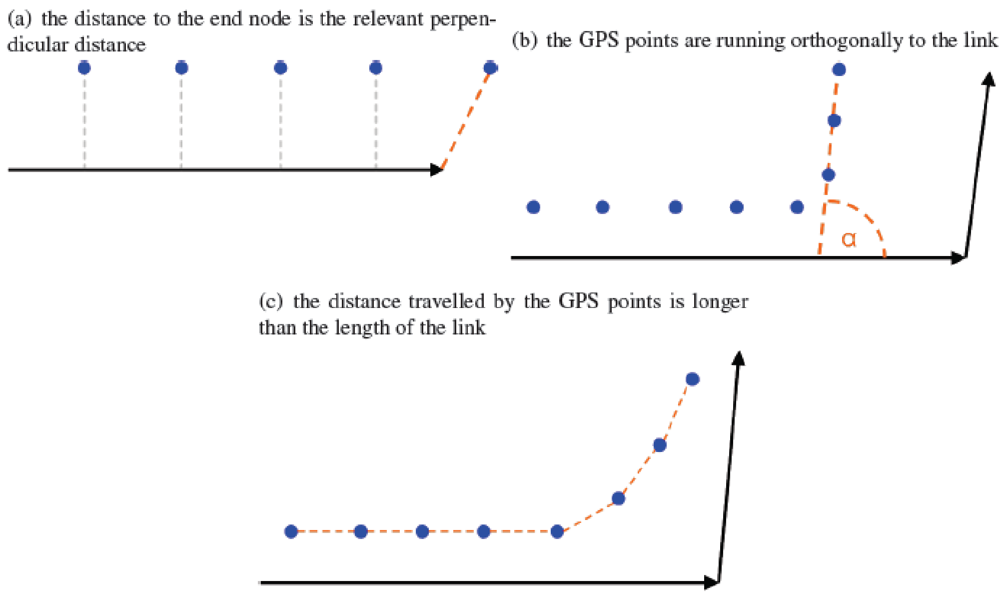


Figure 32. Three Criteria Used to Assess Reaching End of Link (Schuessler and Axhausen, 2009)

Topological Procedures

Topological procedures consider the feasibility of a path of a GPS trace. Coupled with geometric procedures, the topological procedures offer much better map-matching results than solely relying on geometric procedures because topological approaches consider the connectivity of the network in assessing the feasibility of a route.

Pyo et al. (2001) presented a map-matching method using the multiple hypothesis technique (MHT) that assigned likelihood probabilities to potential paths based on the position and heading of the GPS points and the topology of the road network. To improve computational time, path hypotheses with lower probabilities below a user-specified threshold were excluded from further consideration.

Of the hybrid geometric and topological procedures in the literature, the most relevant and directly applicable to the Austin CycleTracks study is the one developed by Dalumpines and Scott (2011) because of its reliance on the ArcGIS software, the software of choice for this research project. Their ArcGIS-based map-matching algorithm integrates existing geometric, buffer, and network analyst tools in Python code within the ArcGIS platform. They claimed their algorithm is the first map-matching algorithm in the literature to be entirely GIS-based for processing a batch of completed GPS trips. It is additionally unique for using a network dataset with information-providing attributes such as one-way streets, turn restrictions, road classification, and speed limits.

The ingenious aspect of the Dalumpines and Scott (2011) algorithm is the use of ArcGIS's Make Route Layer geoprocessing tool that uses Dijkstra's shortest-path algorithm to find the shortest path between an origin and destination. Their algorithm constrains the search for the shortest path within a buffer width established a certain distance away from each of the GPS traces. Dalumpines and Scott gave credit to Zhou and Golledge (2006) for mentioning the potential use of the shortest-path algorithm; however, Zhou and Golledge did not test the approach.

The tricky part of their algorithm, though, is determining what buffer width to use. Dalumpines and Scott (2011) found that too small of a buffer distance can prevent the matching of routes, while too wide of a buffer distance introduces the possibility of inclusion of incorrect links in the route. Dalumpines and Scott (2011, p. 11) stressed that the "accuracy of the GIS-based map-matching algorithm is sensitive to the buffer distance." Therefore, they had to conduct a sensitivity analysis to identify the optimum buffer distances by which to constrain the search for the feasible path along the GPS trace. However, their research benefited from additional traveler information that allowed them to compare the resulting constrained shortest-path routes with the actual routes. The sensitivity analysis consisted of testing buffer distances in increments of 5 m between 10 m and 100 m. With buffer distances below 25 m (on either side of the GPS trace), the shortest-path algorithm did not find any routes; however, with buffer distances above 60 m, inaccurate routes resulted.

Dalumpines and Scott (2011) listed the advantages of their algorithm as follows:

- Simple user interface.
- Adjustable and customizable parameters.
- Expandable performance functions.
- Accurate route generation within a reasonable amount of time.
- Portability (because of the platform independence of the Python script).

The Dalumpines and Scott (2011) algorithm, summarized in the following five steps, assumes no errors or gaps in the GPS traces and previous construction of a complete network dataset (in other words, the algorithm does not include data-cleaning steps). Dalumpines and Scott purchased a street network dataset from DMTI Spatial 27 CanMap® Route Logistics Version 2008.3 to match GPS traces tracking automobile movement to the street network.

1. Connect GPS points for each bike trip into a polyline feature. The first and last GPS points are designated as the origin and destination stops for use in the ArcGIS network analyst extension.
2. Create a buffer around the polyline feature with a user-defined distance. The buffer is used to delineate the barrier that constrains the search for the shortest path (see Figure 33).
3. Assign the origin and destination stops and the barriers for the route solver. If needed, additional stops in between the start and end points may be included.
4. Use ArcGIS's network analyst route solver to generate routes for each trip. The generated route is the shortest route (based on criteria such as travel distance) inside the buffer region.
5. Depending on the network dataset used, update the trip attribute table to include the number of left and right turns along the constrained shortest-path route calculated during the route solver operations.

ArcGIS 10.0’s Help offers the following image to explain how line barriers work in the process of finding the shortest path.

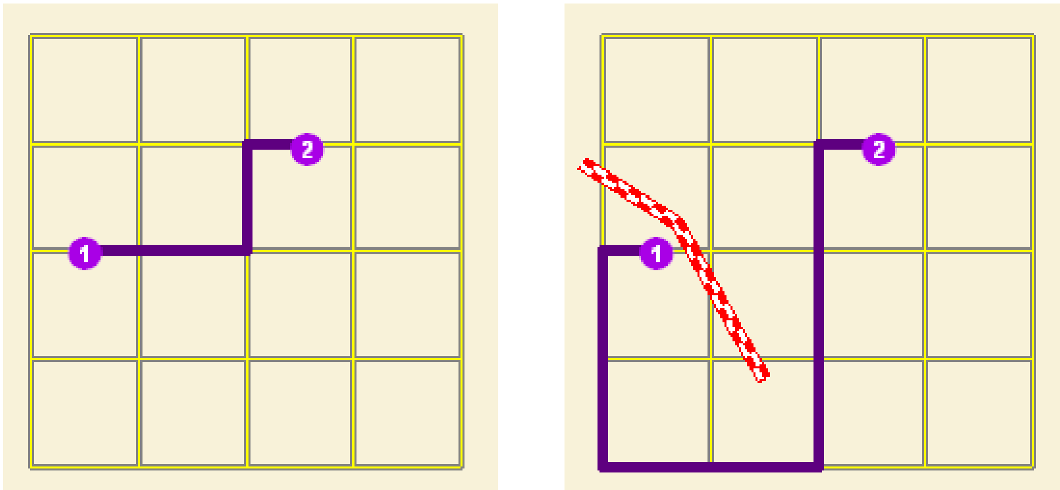


Figure 33. Restriction Line Barrier Prohibits Travel Anywhere the Barrier Intersects the Network (ESRI ArcGIS 10.0 Help)

Dalumpines and Scott (2011) indicated their algorithm correctly matched 91 out of the 104 total car trips sampled in their study, for a success rate of 88 percent. Travel diary records supplemented the GPS tracking and were used to validate their results and calibrate the user-specified buffer distance to obtain the best results.

As the first published algorithm to use ArcGIS for map-matching GPS traces to a network, the Dalumpines and Scott (2011) algorithm provided a promising approach for use in the Austin CycleTracks study. Appendix D describes how their algorithm was implemented in ArcGIS for this research project.

Advanced Procedures

Advanced map-matching procedures typically combine geometric and topological procedures with additional criteria, parameters, assumptions, probabilistic analysis, or optimization. A review of the literature for these types of algorithms revealed they are typically developed or applied to real-time map-matching for applications such as intelligent transportation systems (ITSs) and driver assistance tracking with GIS units in vehicles (Quddus et al., 2007).

For instance, some algorithms score potential routes based on a variety of criteria or probabilistic evaluation. Newsom and Krumm (2009) tested a map-matching algorithm that uses Hidden Markov Model (HMM) to find the most likely route on a network (Figure 35). The map-matching process assigned a probability for the distance measured between the GPS point and a network link (i.e., the measurement probability). Measurement probabilities were set equal to zero for network links more than 200 m away (a user-specified parameter) from a GPS point.

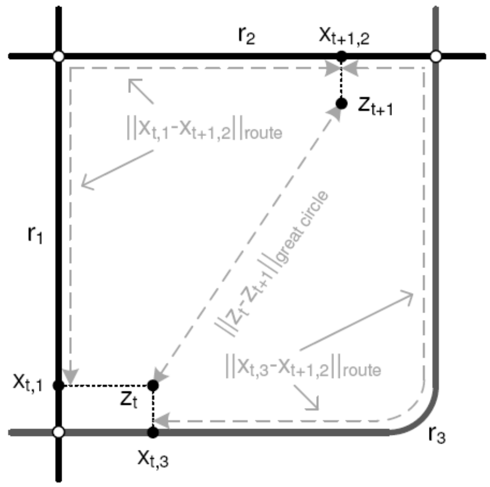


Figure 4: This shows an example of our notation. There are three road segments, r_1, r_2, and r_3, and two measured points, z_t and z_{t+1}. The first measured point, z_t, has candidate road matches at $x_{t,1}$ and $x_{t,3}$. Each match candidate results in a route to $x_{t+1,2}$, which is a match candidate for the second measured point, z_{t+1}. These two routes have their own lengths, as does the great circle path between the two measured points. Our data shows that the route distance and great circle distance are closer together for correct matches than for incorrect matches.

Figure 34. Hidden Markov Model Map-Matching Procedure (Excerpt from Newsom and Krumm [2009])

The Newsom and Krumm (2009) algorithm requires the specification ahead of time of the following key parameters, distributions, and assumptions: the distance threshold for setting the measurement probability equal to zero, the distribution or choice of initial state probabilities, and the assumption to favor transitions from one network link to another that are closest to the great circle distance between the GPS point measurements. The assumption of choosing the shortest great circle path does not take into consideration the noise associated with the use of GPS-tracking devices on slower moving objects such as cyclists.

The map-matching algorithm by Pyo et al. (2001) requires the user to specify minimum probability thresholds used to cut candidate links and paths from consideration and requires assumptions about the distribution of GPS position and GPS heading (direction) differences, as well as development of an error region around the GPS points.

The algorithm by Schuessler and Axhausen (2009) mentioned earlier as the one utilized by SFCTA requires the user to specify a search radius around the GPS point and a maximum number of route candidates, specifies a route score dependent on distance to link and speed at GPS point, and requires an assumption that a shortest-path route is taken. Schuessler and Axhausen refined the Marchal et al. (2005) algorithm.

Advanced map-matching algorithms tend to require more assumptions and user-specified values. This study focused on trying map-matching algorithms implementable using existing ArcGIS 10.0 tools to keep the procedures more accessible to staff at local agencies.

Appendix E: Alternatives to ArcGIS for Map-Matching

Consideration was given to matching and analyzing the GPS data using programs other than ArcGIS, such as MATLAB and Google Earth.

An earlier analysis of the data used MATLAB; however, ArcGIS was preferred because of its ubiquitous use by local governments and the likelihood of staff being more familiar with it. An attempt was also made to use Google Earth and Google Fusion Tables to match the GPS points to the Google Maps network.

The research team developed a software program using the PHP (Hypertext Preprocessor) language to break down the large GPS Coordinates data file into individual keyhole markup language (KML) files, which can be used in a variety of Google's products as well as in any prominent geographical data processing software. These KML files can be easily imported into Google Earth using its import functionality. Figure 36 shows a screen capture of a Google Earth window showing all the unprocessed trips on an aerial view. Researchers realized the limitation of using Google Earth to process the data because it does not allow any manipulation of the data. However, Google offers other programs such as Google Fusion Tables, which allows manipulation of imported data. Google Earth allows for exporting all the KML files into a single KML file, which can then be imported into the Google Fusion Tables program. Figure 37 shows a sample database view available in Google Fusion Tables, which allows quick deletion of any data after visual confirmation and a variety of ways to filter and sort the data. It also provides a quick count of various parameters, like unique number of users, number of total trips, number of trips per user, and number of trips without any user data. It also allows aggregation of data using any fields in the database. It can also be used to produce a frequency map, or a heat map, to find out the most active areas for trip generation during the data collection. Figure 38 shows a heat map for all the trips imported into the database. It shows that most of the trips were generated by the area west of IH 35W in Austin, and it shows a high amount of activity around the University of Texas campus.

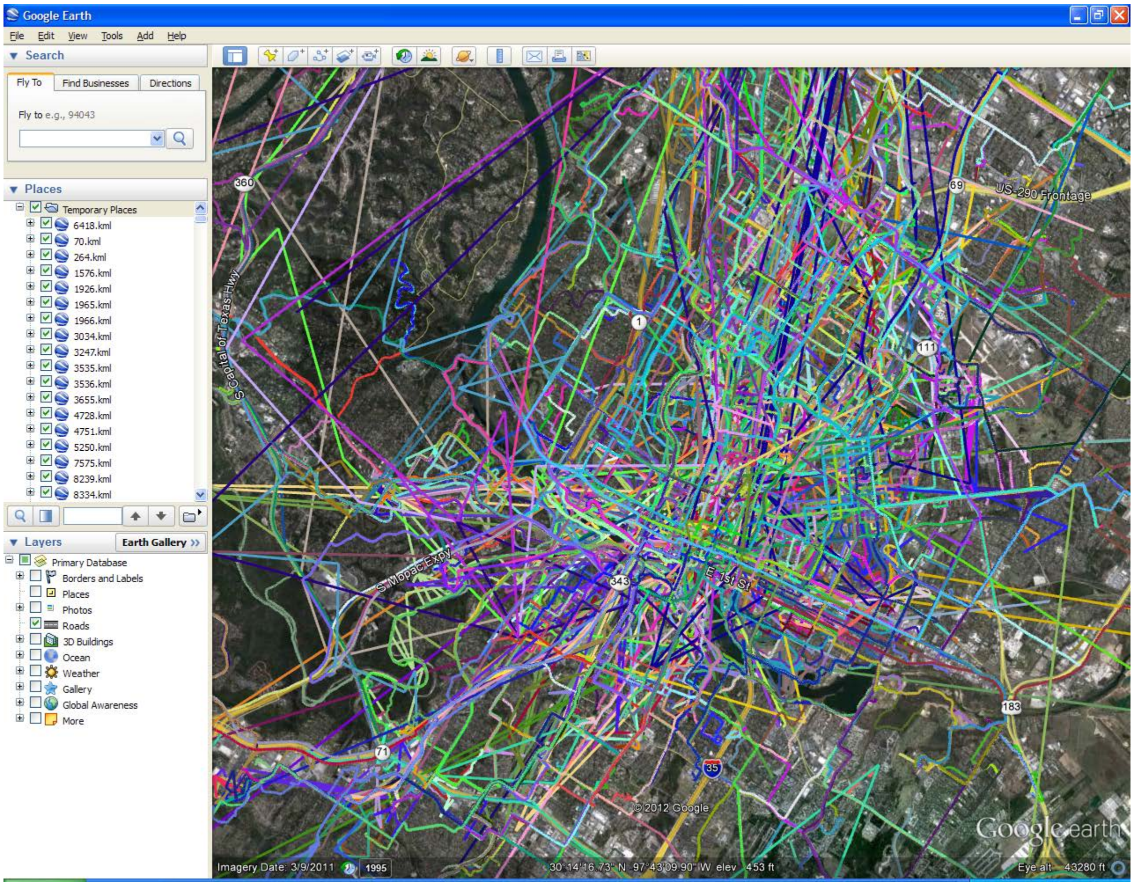


Figure 35. Google Earth Screenshot Showing Unprocessed Trips

File View Edit Visualize Merge Experiment

Showing **all rows** options

Description	Name	Age	Gender	Home Zip	School Zip	Work Zip
Age : 36 Cycling Frequency : Daily User ID : 2840 ...	19600.kml	36	M	78722		78701
Age : 33 Cycling Frequency : Several times per wee...	22871	33	F	78704	78704	78703
Age : Cycling Frequency : User ID : Gender :	22870					
Age : 51 Cycling Frequency : Several times per wee...	22869	51	F	78704		78721
Age : 29 Cycling Frequency : Several times per wee...	22868	29	M	78703	78712	
Age : 44 Cycling Frequency : Less than once a mont...	22865	44	F	78746		78723
Age : 48 Cycling Frequency : Daily User ID : 3250 ...	22867	48	M	78704		78746
Age : 26 Cycling Frequency : Daily User ID : 3308 ...	22864	26	Male	78702		78701
Age : 42 Cycling Frequency : Daily User ID : 3145 ...	22860	42	F	78704		78704
Age : 36 Cycling Frequency : Daily User ID : 2840 ...	22861	36	M	78722		78701
Age : 30 Cycling Frequency : Daily User ID : 3201 ...	22859	30	Male	78757	78701	
Age : 30 Cycling Frequency : Daily User ID : 3201 ...	22858	30	Male	78757	78701	
Age : 36 Cycling Frequency : Daily User ID : 2840 ...	22862	36	M	78722		78701
Age : 42 Cycling Frequency : Daily User ID : 3145 ...	22856	42	F	78704		78704
Age : 36 Cycling Frequency : Daily User ID : 2840 ...	22853	36	M	78722		78701
Age : 41 Cycling Frequency : Several times per wee...	22846	41	Male	78751		78701
Age : 34 Cycling Frequency : Several times per wee...	22844	34	M	78704		78712

Figure 36. Google Fusion Table—Data Table View

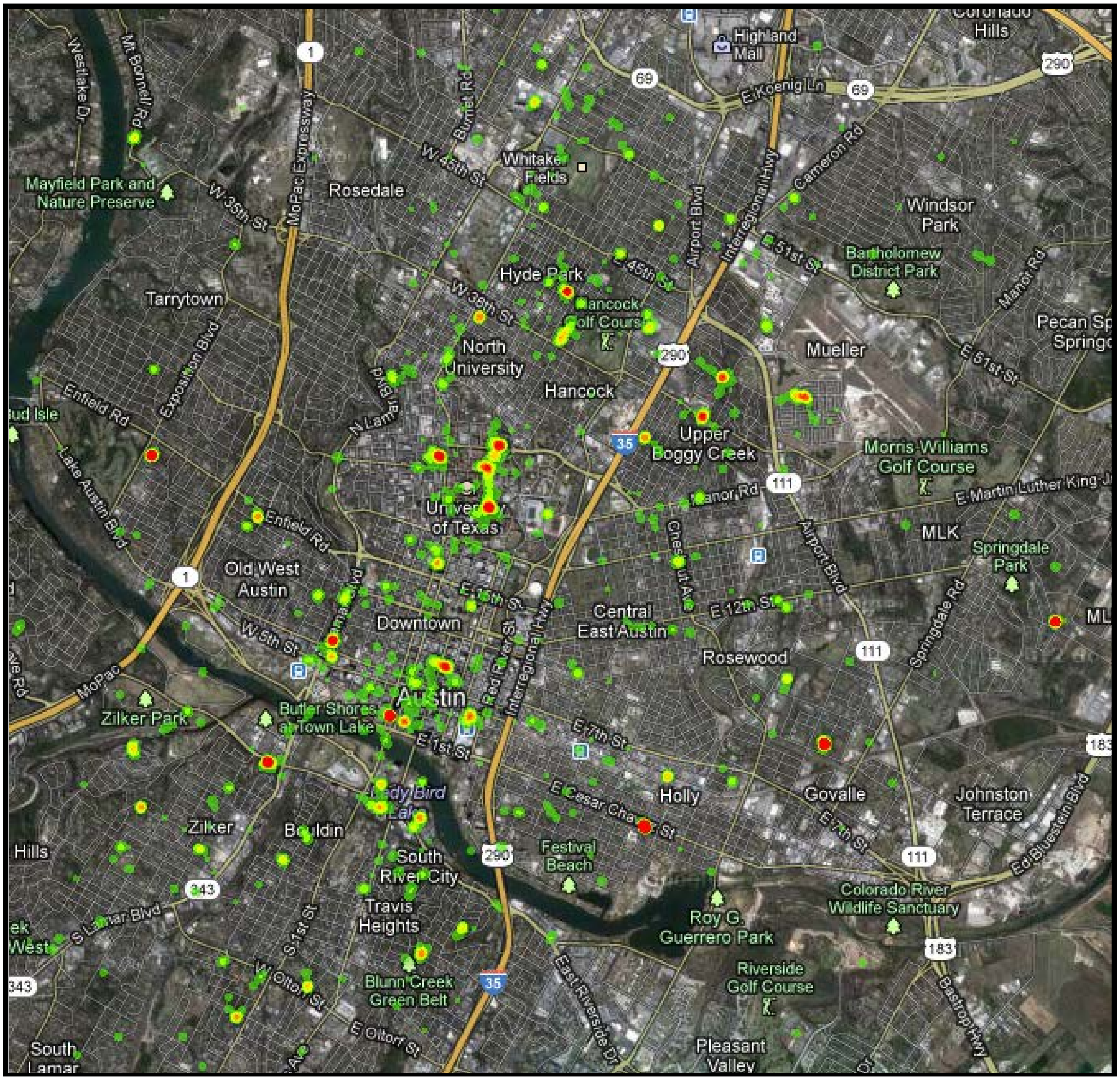


Figure 37. Heat Map from Google Fusion Table

While the Google Fusion Table program can be used to aggregate the user information effectively, it does not offer any tools to process the geographical data further. An attempt was made to use Google Maps API to perform reverse geocoding queries to match the latitude and longitude to the street network for cleaning up the trip routes.

Reverse geocoding is the process of back (reverse) coding of a point location (i.e., latitude and longitude) to a readable address or place name. This permits the identification of nearby street addresses, places, and/or subdivisions in a neighborhood, county, state, or country. The process of reverse geocoding takes a relatively high amount of server usage, which is the reason most of the services providing reverse geocoding have a limit on number of queries per day.

The researchers encountered the usage limit fairly quickly because of the vast amount of points in need of reverse geocoding. The Google Maps API service limits the reverse geocoding requests to 2,500 per day. This limit can be increased to 100,000 per day by subscribing to the business service, but the researchers were looking for a way that could be used by anyone at a minimal cost, so this approach was abandoned.

However, a new reverse geocoding approach that may be feasible to use in future projects was discovered toward the end of this project. MapQuest open platform web services allows the use of OSM data to handle many features, like reverse geocoding, without any query limits. The researchers tested a sample of latitude/longitude against this service and received encouraging results. A reverse geocoding request sent to the server returned a vast amount of data pertaining to the geolocation point, such as nearby building, road, city, county, state, country, and zip code. This service can be used by developing a software program to get the most used routes in the city or to route match the trips. Figure 39 shows a sample output returned by this service.

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
– <reversegeocode timestamp="Fri, 25 May 12 12:03:01 -0400" attribution="Data Copyright OpenStreetMap Contributors, Some Rig
  json_callback=renderExampleThreeResults&lat=30.2757217&lon=-97.7525246">
  – <result place_id="2174378940" osm_type="way" osm_id="135088177" lat="30.2757574067573" lon="-97.7524168360055">
      Whole Earth Provision Co., West 11th Street, Austin, Travis, Texas, 78703, United States of America
    </result>
  – <addressparts>
      <building>Whole Earth Provision Co.</building>
      <road>West 11th Street</road>
      <city>Austin</city>
      <county>Travis</county>
      <state>Texas</state>
      <postcode>78703</postcode>
      <country>United States of America</country>
      <country_code>us</country_code>
    </addressparts>
</reversegeocode>
```


Figure 38. A Sample Output Using the MapQuest Open Services API

Appendix F: ArcGIS ModelBuilder Implementation of Map-Matching Algorithm

The algorithm developed by Dalumpines and Scott (2011), described in Appendix C, was selected for this project because of its implementation within ArcGIS and its use of both geometric and topological procedures for map-matching. The algorithm mostly uses existing ArcGIS tools but requires some Python scripting to process in a batch all the bike trips. The standard Network Analyst© ArcGIS tools cannot process multiple bike trips with different origins and destinations and barriers. Fortunately, ArcGIS's ModelBuilder program within ArcCatalog provides a solution for handling the large number of bike trips and line barriers for each bike trip.

Dalumpines and Scott (2011) did not provide their Python script; however, the support staff at ESRI worked with the research team to develop a procedure using ArcCatalog's ModelBuilder tool that recreated the batch processing ability of the Dalumpines and Scott Python scripting. Figure 40 illustrates the steps involved in the Modelbuilder Model with the use of a flowchart. These steps are further defined in Table 14 and Figure 41.

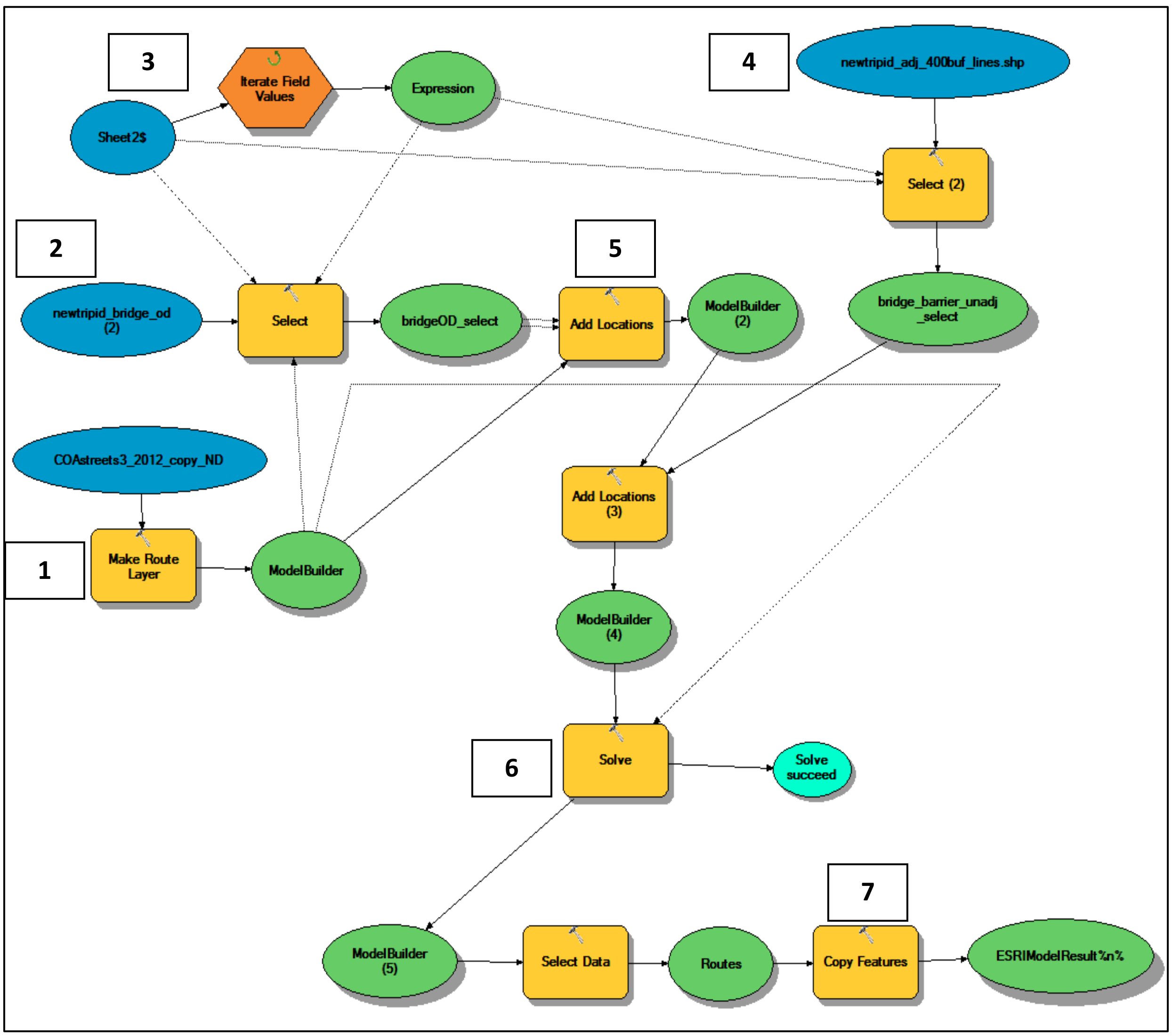
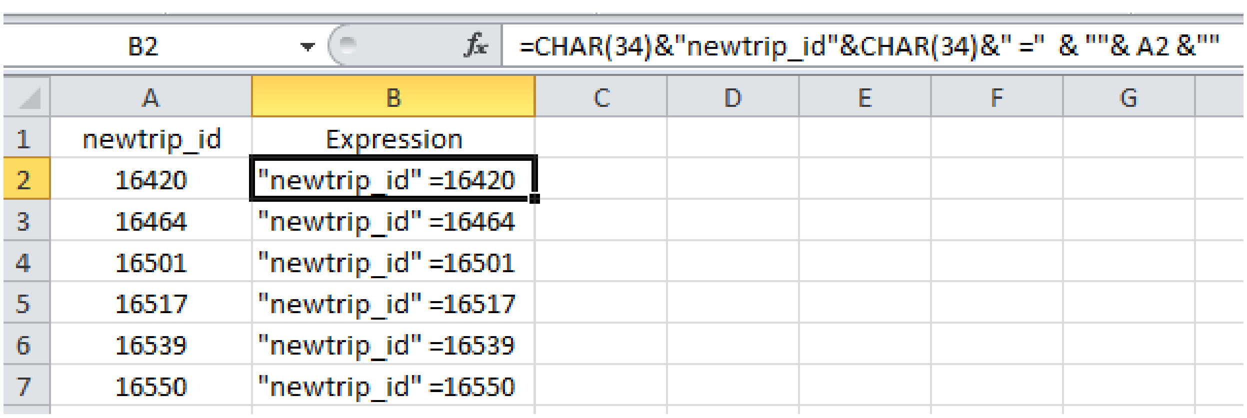


Figure 39. Model for Dalumpines and Scott (2011) Algorithm

Table 12. ModelBuilder Model to Replicate Dalumpines and Scott (2011) Algorithm

	Step	Description
1	Make route layer	Part of the Network Analyst toolbox, this command specifies the use of Dijkstra's shortest-path algorithm for creating a shortest-path route using the network dataset (e.g., City of Austin network).
2	Select	This command selects the origin and destination points for each bike trip.
3	Iterate field values	The iteration pulls from an Excel spreadsheet (labeled as Sheet2$ in Figure 38, and data entry shown in Figure 39). The expression is used to select the bike trip ID (e.g., "newtrip_id"= 16420 in Figure 35) from the origin and designation points file (newtrip_bridge_od) and the line barrier file (newtripid_adj_400buf_lines.shp). The model repeats the process of selecting origin and destination points and finding the shortest-path route within an area bounded by line barriers for each bike trip ID.
4	Select barriers	For each bike trip, this command selects the line barriers that define the feasible route area. The line barriers are created by using the ArcGIS 10.0 Buffer tool (using the option of merging the buffers that share the same bike trip ID) and then the Polygon to Line tool to convert the buffer polygons to multiple line features associated with each bike trip ID.
5	Add stop locations	This command adds the origin and destination points as the first and second stop, respectively, for the shortest-path route analysis.
6	Solve for route	This command solves for the shortest route path between the first and second stop within the area bounded by the line barrier.
7	Copy route to feature dataset	For each bike trip ID, a feature dataset is created with the shortest path on the network consisting of the network links.


B2 =CHAR(34)&"newtrip_id"&CHAR(34)&" =" & ""& A2 &""

	A	B	C	D	E	F	G
1	newtrip_id	Expression					
2	16420	"newtrip_id" =16420					
3	16464	"newtrip_id" =16464					
4	16501	"newtrip_id" =16501					
5	16517	"newtrip_id" =16517					
6	16539	"newtrip_id" =16539					
7	16550	"newtrip_id" =16550					

Figure 40. Excel Spreadsheet Entries for Iterations

The Dalumpines and Scott (2011) algorithm was run for the entire cleaned set of CycleTracks bike trip GPS traces to the network using a 200-ft buffer that prevented consideration of network links 200 feet away from the GPS trace. The smaller buffer was used to limit the search area to avoid inclusion of roadways most likely not part of the bike route taken. Too generous of a buffer width could result in routes that stray too far from the most likely route followed by the GPS trace; to narrow of a introduce the possibility of not capturing the needed network links.

The number of routes generated and matched to the network totaled 2,820, for a map-matching rate of 88 percent (i.e., the algorithm generated a route on the network for 88 percent of the GPS bike trip traces). Dalumpines and Scott (2011) indicated they correctly identified 88 percent of the routes; however, they benefited from having access to travel diary records to validate their results and calibrate the user-specified buffer distance to obtain the best results.

The algorithm used only the origin and destination points to match the routes. An attempt was made to improve the map-matching of trips by including additional points ("stops" in ArcGIS), especially for round trips. However, it was found that by adding more stops, the total number of matched routes declined (Figures 42 and 43); though matching could improve for some routes (Figure 44). Additional work will continue to identify how to improve the map-matching success.

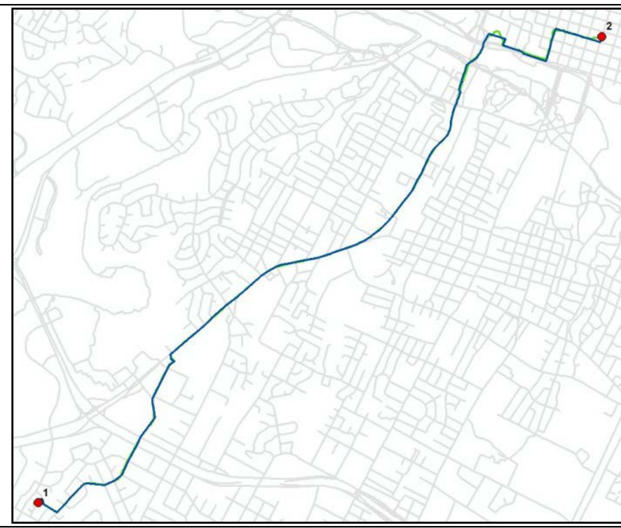
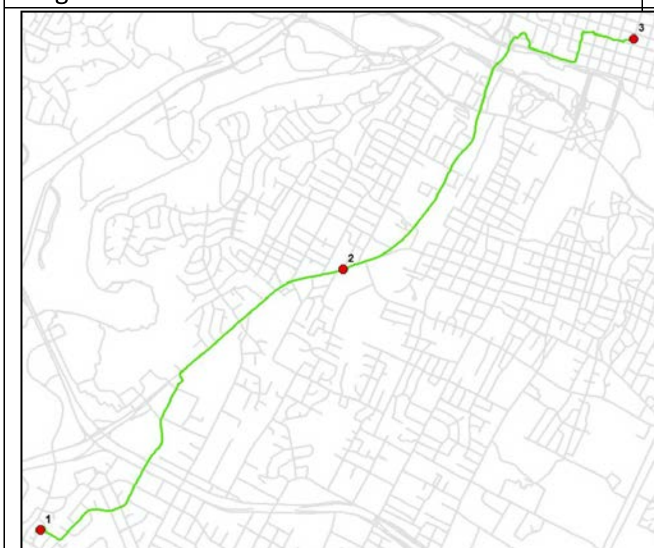
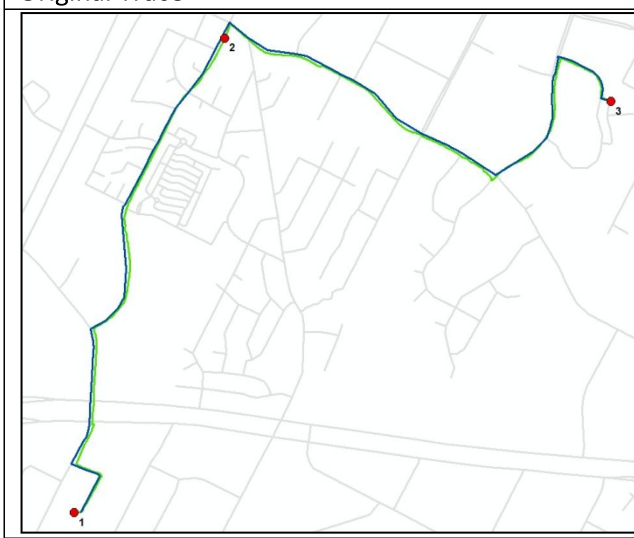
	
Original Trace	Two Points
	
Three Points (unmatched)	Five Points (unmatched route)
original GPS trace	
matched route	

Figure 41. Example of How Adding Points Results in Unmatched Routes

Original Trace

Two Points

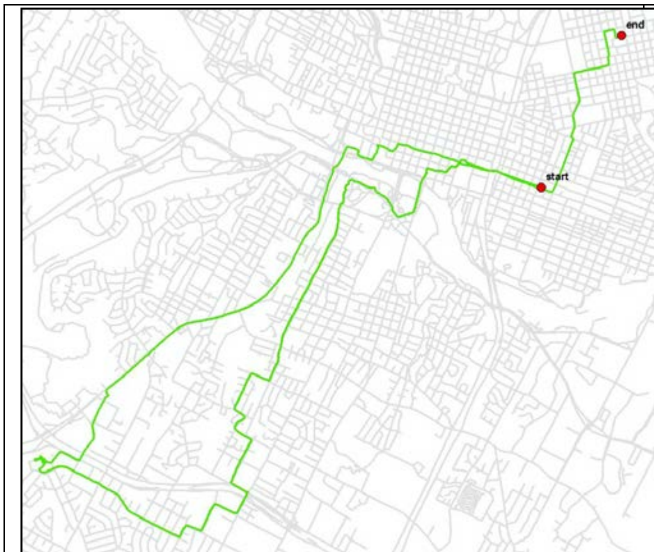


Three Points (unmatched)

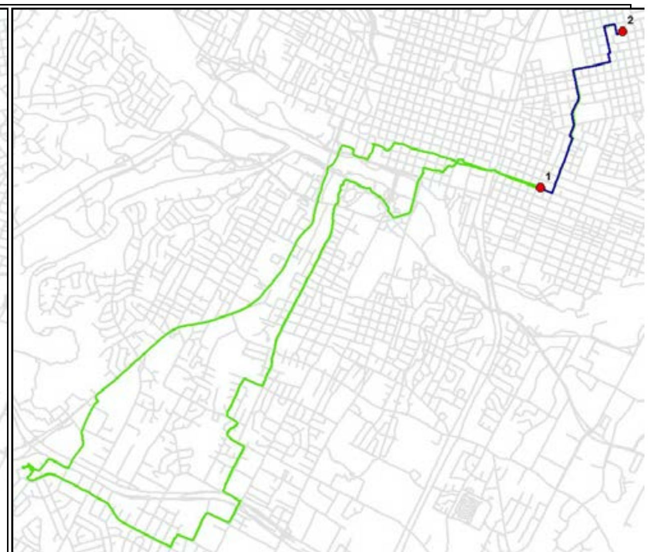
Five Points (unmatched route)

original GPS trace	
matched route	

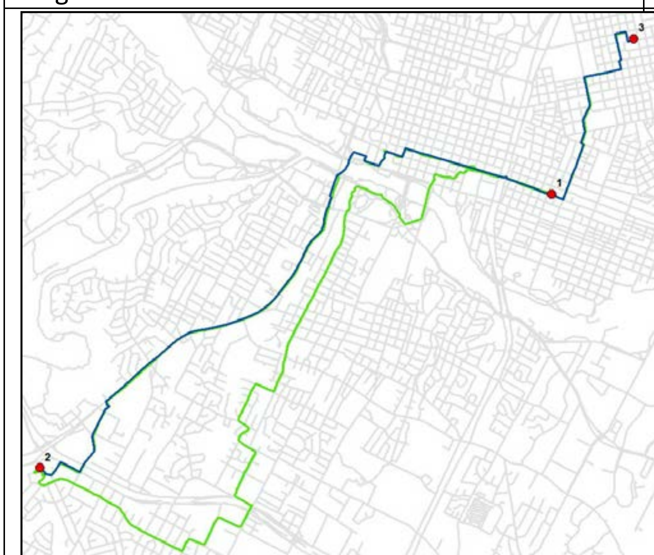
Figure 42. Another Example of How Adding Points Results in Unmatched Routes



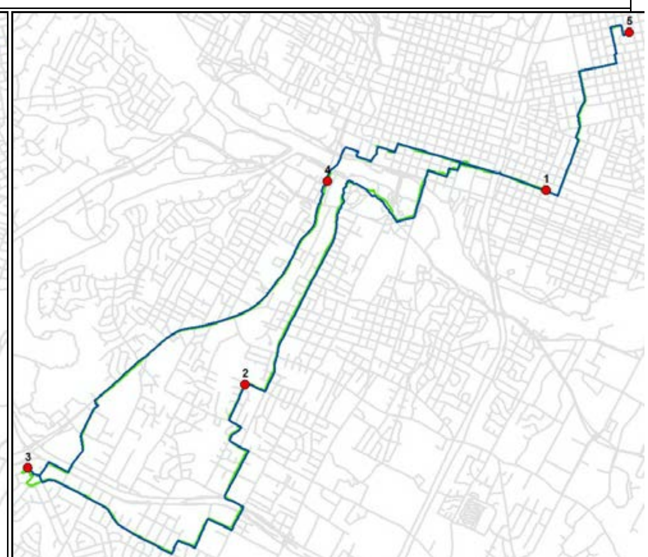
Original Trace



Two Points



Three Points (unmatched)



Five Points (unmatched route)

original GPS trace	
matched route	

Figure 43. Example of Round Trip Recognized by Including Five Points

University Transportation Center for Mobility™

Texas Transportation Institute

The Texas A&M University System

College Station, TX 77843-3135

Tel: 979.845.2538 Fax: 979.845.9761

utcm.tamu.edu

Texas Transportation Institute