GEORGIA DOT RESEARCH PROJECT 14-39

FINAL REPORT

USING CROWDSOURCING TO PRIORITIZE BICYCLE NETWORK IMPROVEMENTS

Georgia Department of Transportation

OFFICE OF RESEARCH
15 KENNEDY DRIVE
FOREST PARK, GA 30297-2534

This page intentionally left blank.

GDOT Research Project RP14-39

Final Report

Using Crowdsourcing to Prioritize Bicycle Network Improvements

By
Dr. Kari E. Watkins
Assistant Professor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Chris LeDantec
Assistant Professor
School of Literature, Media and Communication
Georgia Institute of Technology

Contract with

Georgia Department of Transportation

In cooperation with
U.S. Department of Transportation
Federal Highway Administration

April 2016

This page intentionally left blank.

<table>
<tr><td>1.Report No.:
FHWA-GA-16-1439</td><td>2. Government Accession No.:</td><td>3. Recipient's Catalog No.:</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle:
Using Crowdsourcing to Prioritize Bicycle Network Improvements</td><td>5. Report Date:
April 2016</td></tr>
<tr><td>6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s):
Dr. Kari E. Watkins, PE (P.I.), Dr. Chris LeDantec (co-P.I), Aditi Misra, Mariam Asad, Charlene Mingus, Cary Bearn, Alex Poznanski, Anhong Guo, Rohit Ammanamanchi, Vernon Gentry, Aaron Gooze</td><td>8. Performing Organ. Report No.:</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address:
Georgia Institute of Technology
School of Civil and Environmental Engineering
School of Literature. Media and Communication</td><td>10. Work Unit No.</td></tr>
<tr><td>11. Contract or Grant No.:
GDOT Research Project No. 0013114
(RP 14-39; UTC Sub-Project)</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address:
Georgia Department of Transportation
Office of Materials & Research
15 Kennedy Drive
Forest Park, GA 30297-2534</td><td>13. Type of Report and Period Covered:
Final; May 2014- April 2016</td></tr>
<tr><td>14. Sponsoring Agency Code:</td></tr>
<tr><td colspan="3">15. Supplementary Notes:
Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration.</td></tr>
<tr><td colspan="3">16. Abstract:
Effort to improve the bicycle route network using crowdsourced data is a powerful means of incorporating citizens in infrastructure improvement decisions, which will improve livability by maximizing the benefit of the bicycle infrastructure funding and empowering citizens to be more active in transportation decisions. This research developed a free, GPS-enabled smartphone application to collect socio-demographic and route data of cyclists in Atlanta. The crowdsourced data were then used to model the factors influencing bicycle route choices of different types of cyclists as defined by their perceived safety and comfort with a facility. Finally, this research refined a quality-of-service measure for bicyclists based on the perceived level of traffic stress (LTS) that the users attach to the facility. The developed quality-of-service measure can be used by transportation professionals to compare alternative roadway and bikeway designs using quantifiable variables such as speed limit, traffic volume, and number of through lanes.</td></tr>
<tr><td colspan="2">17. Key Words:
Crowdsourced data; smartphone app; bicycle infrastructure; bicyclist route choice; level of traffic stress (LTS)</td><td>18. Distribution Statement:</td></tr>
<tr><td>19. Security Classification (of this report):
Unclassified</td><td>20. Security Classification (of this page):
Unclassified</td><td>21. Number of Pages: 217
22. Price:</td></tr>
</table>

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Bicycling has been identified as a critical component of livable communities, as it offers an environmentally friendly, cost-effective, congestion-reducing, and health-promoting mode of transportation for short trips. According to the National Household Travel Survey (NHTS), nearly 40% of all personal trips in the U.S. are two miles or less, a reasonable bicycling distance. However, only about 1% of all such trips are made on bicycles, although it has been widely acknowledged that bicycling is a healthier and non-polluting mode of transportation. A major reason frequently cited for not adopting bicycling is a perceived lack of safety in shared facilities having high traffic speed and volume. Towards mitigating that concern, separate bicycling facilities need to be built along corridors that may provide a short route to a destination but are avoided by cyclists due to such factors. However, bicyclists being a small and dispersed group, it is difficult to get data on their travel patterns through traditional traffic counts and hence regional transportation agencies often follow heuristics or stated preference surveys to assign cyclists to the city street network. While heuristics are entirely subjective and dependent on the person modelling the traffic flow, stated preference surveys often suffer from recall bias and selective preference of the survey participant leading to incorrect understanding of the actual route choice of the cyclists.

This research was carried out in multiple related areas: (1) creating a freely available GPS enabled smartphone based application to collect revealed preference data from cyclists of Atlanta, (2) developing an open source data cleaning and map matching procedural standard, (3) using the data collected via the smartphone application to understand the influence of socio-demographics of cyclists on route preferences, (4) developing a route choice model for the planners and transportation decision makers of Atlanta to understand if and how much cyclists deviate from the shortest route between origin and destination and finally, (5) developing a link level stress measure for cyclists based on their level of confidence and comfort with cycling infrastructure.

Cycle Atlanta, a Geographical Positioning System (GPS) based smartphone application (app), was developed at Georgia Institute of Technology (Georgia Tech) in collaboration with the City of Atlanta to collect revealed preference route choice data of cyclists in Atlanta. Along with recording routes, the app provides the users the option to input demographics like age, gender, ethnicity and income, and rider characteristics like rider type, rider experience, and riding frequency while recording their trips. This research uses the data collected through the Cycle Atlanta app to understand the route choice of the cyclists and how the choice is influenced by rider and route characteristics.

The first part of data analysis shows that socio-demographic variables and riding patterns are significant predictors of a cyclist's probability of self-classifying himself/herself into a particular category based on his/her comfort with presence or absence of cycling infrastructure and his/her interest in cycling. In particular, gender, rider history, and cycling frequency are significant in all the models. The results indicate that knowing a cyclist's demographic information can potentially help in classifying the cyclist into a particular rider type. In the future, this can help researchers to streamline surveys by replacing sociodemographic questions by a single rider type classification question. Alternatively, knowing the socio-demographics characteristics commonly available through census data and other surveys, researchers will also be able to predict the rider type and hence infrastructure preferences of people without having to undertake a new survey design for cyclists only. It will also help in understanding infrastructure and facility need of future cyclists who are not yet cycling and hence, there is no revealed preference data on their preference currently.

The results also direct attention to the requirement of segmented route and facility preference decision models for different cyclist types. Since the purpose of the route and facility preference analysis is to understand the requirements by rider types, segmented models based on rider type may enable a planner to better predict the choices of a future cyclist based solely on demographic information of the cyclists. Future route decision model research may therefore explore segmentation of the dataset to achieve better predictability.

From the second part of the analysis, it is evident that most route perception issues and facilities are viewed on a similar scale by cyclists as the mean scores on those facilities are quite similar across rider types. Other results indicate that sociodemographic attributes and confidence levels influence infrastructure and facility preference. However, the model fits are substantially low indicating that rider level data are not sufficient to predict the route level decision process. Further investigation is necessary, as the literature shows that choice of route depends on route characteristics as well as rider characteristics like age and gender.

The final part of the study proposes a quality-of-service measure for bicyclists based on the perceived level of traffic stress (LTS) that the users attach to the facility. This research proposes a modified LTS measure which is based on a LTS measure developed at Mineta Transportation Institute and uses traffic and roadway characteristics data that are readily available to most transportation agencies and that has been validated by the literature.

The modified LTS measure can be used by transportation professionals to compare alternative roadway and bikeway designs using quantifiable variables such as speed limit, traffic volume, and number of through travel lanes. The modified LTS measure also provides results which can easily be understood by the public and decision makers. A case study conducted using the modified LTS measure demonstrated the effectiveness of the measure. However, since the purpose of the study was to propose a generalized measure, some of the data that may affect perceived LTS was intentionally not included in this study. In the future, the modified LTS may be updated to include intersection LTS (signalized separated turning movements, vehicle entry point for bicycle lanes and protected cycle tracks, bike boxes, left-turn queue and unsignalized intersection crossings) and bicycle boulevards depending on data availability and sufficiency.

In summary, this research will improve opportunities for bicycling, and therefore community livability, in urban and suburban areas by evaluating the bicycle network and identifying the routes/ links that will have the most impact on cycling ridership should they be improved. This research is innovative as the analysis is based on crowdsourced observed and perception data collected in real time from actual cyclists, a new form of data collection. Additionally, the research also includes modeling the factors influencing bicycle route choices in urban and suburban areas, which have also not been compared before. Both the crowdsourced data and the bicycle route choice model were used to develop a stress metric to describe bicycle network link importance. City planners and engineers will be able to use the results from the LTS research to a) identify critical segments or routes along the existing or proposed bicycle network that would benefit the most bicyclists if they were improved and b) identify new bicycle routes that, if built, may encourage more cycling. Furthermore, effort to improve the bicycle route network using crowdsourced data is a powerful means of incorporating citizens in infrastructure improvement decisions, which will improve livability by maximizing the benefit of the bicycle infrastructure funding and empowering citizens to be more active in transportation decisions.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the City of Atlanta; Georgia Department of Transportation; Southeastern Transportation Research, Innovation, Development, and Education Center; GVU Center at Georgia Institute of Technology; and Institute for People and Technology for their funding on the project. Any errors or omissions are those of the authors and not the funding agencies. We would also like to acknowledge the assistance provided by Atlanta Regional Commission, Atlanta Bicycle Coalition, and Midtown Alliance at various times during the course of the project. Finally, we owe a world of thanks to the users of Cycle Atlanta who have shared their trips with us to enable this research.

CHAPTER 1. BACKGROUND AND MOTIVATION

Traditionally, transportation planning in the U.S. has been automobile focused, resulting in marginalization of healthy and active modes of transportation like cycling and walking. Environmentally, this has led to increased air pollution; economically, this has made the country dependent on international fuel economy; and socially, this has brought about an alarming increase in obesity, heart disease and asthma among both adults and children (Sallis 2004).

As a mode choice, bicycling can reduce overall congestion, air pollution and energy consumption while at the same time enable an active lifestyle and a low cost, equitable means of transportation. In view of all these, recently, the federal government has geared its policies towards promoting biking and walking, and several state and local transportation planning agencies have incorporated a bicycle planning module in their long term vision for the region. However, literature shows that although 40% of the trips made in U.S. are of bike-able distance, only 1.8% of such trips are bicycle trips (Pucher et al. 2011). This low subscription to bicycling has been generally attributed to safety issues (AASHTO 2012) with major factors contributing to negative safety perceptions being high speed limits, high traffic volumes, last mile disconnect in the network and absence of dedicated facilities for cyclists that can provide a physical separation from the vehicular traffic (Dill and Carr 2003, Buehler and Pucher 2011).

Studies reveal that a substantial increase in the number of bicyclists can be achieved by providing facilities for safe riding (Pucher and Buehler 2007) and therefore, it is important for the planning agencies to know where the cyclists prefer to bike and possibly their ‘willingness to pay’ for an added facility. Cities often try to organize the route network by balancing the connectivity of the network, shortest travel distances, parking locations, and traffic volumes (Dill 2004) – but this task is often difficult as perception of safety and comfort varies across level of experience of the cyclists, age, traffic characteristics and several other factors. Therefore, a better approach to understand cyclist route choice is to collect the revealed preference data on the routes where the cyclists actually travel and then model the factors that influence the route choice decision of the cyclists.

Bicycling trip data are sparsely available for at least four reasons – first, bicycling trips often use by-lanes and short-cuts that are not manned during traditional traffic counts; second, bicycling trips also tend to happen during non-peak hours of commute, thus again not being counted during traffic counts; third, the automated counters are designed to detect vehicular metallic mass and therefore tend to underestimate bicycling trips; and fourth, since bicyclists constitute a marginal proportion of the total traffic, there are rarely separate count efforts employed for cycling trip counts. As a solution to such issues of data collection, smartphone applications have been developed to enable users to record their trips by themselves. The earliest example of such an effort is the CycleTracks application developed at San Francisco County Transportation Authority (Hood et al. 2009) which has now been adopted by over a dozen cities across the U.S.

Recently, the City of Atlanta started expanding its network of bicycle facilities to encourage people to bicycle more often. In doing so, they needed to understand the most travelled corridors, as well as particular streets that are avoided by cyclists even when those streets are the shortest connectors between any two points en route. For the purpose of data collection, collaboration was set up between the Georgia Institute of Technology and the City of Atlanta's planning office to develop a smartphone application that would help in collecting data from bicyclists who use these corridors and other city streets. The project was further facilitated by support from Atlanta Regional Commission who viewed the project as a means to foster “extensive public involvement by neighborhood residents, business owners, and the citywide cycling community” (The City of Atlanta, 2011).

The smartphone application created for this initiative was named Cycle Atlanta, after the name of the project, and was developed by an interdisciplinary team of researchers at the Georgia Institute of Technology. The application was based off of CycleTracks, although Cycle Atlanta was substantially updated to make better use of current features available in Apple Inc.’s proprietary mobile operating system (iOS) and Android as well as to include features that the City and local bicycle advocacy groups wanted in the application. The basic feature is trip recording, where the application uses the Global Positioning System (GPS) of the phone to record the location of the user once per second (Figure 1). At the end of the trip, the user is given the option to ‘Save’ the trip and only after the user saves the trip, the trip and related data are uploaded to a secure server. Once the trip is saved, the user can also specify the trip purpose and any related free-form note. Trip purposes have been categorized as commute, school, work-related, exercise, social, shopping, errand, or other, enabling data users to segregate routes based on purpose as the infrastructure requirements and preferences may differ by the purpose of the trip. The free-form notes inform the city about the concerns of the users regarding particular routes and help in initiating correctional measures sooner.

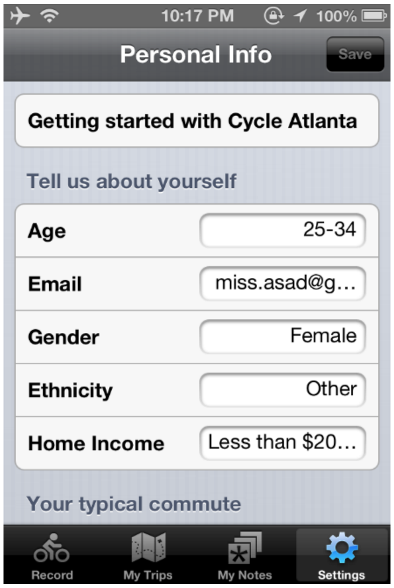


Fig. 1(a)

Fig. 1(b)

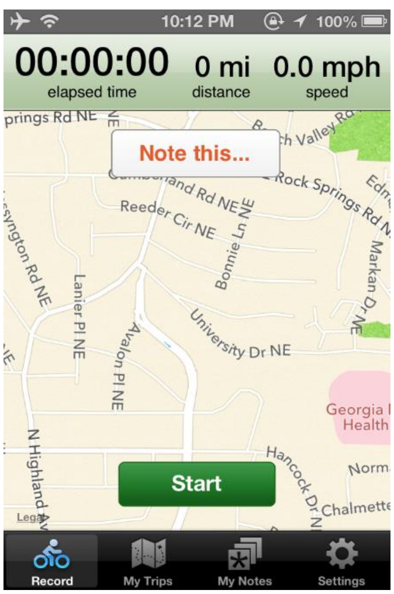


Fig. 1(c)

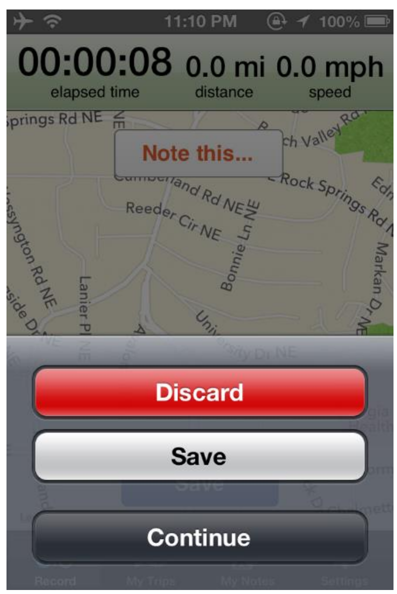


Fig. 1(d)

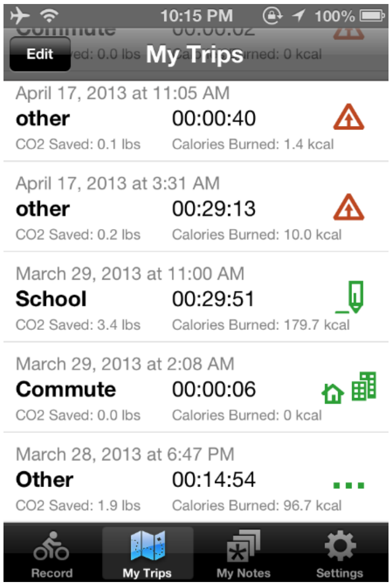


Fig. 1(e)

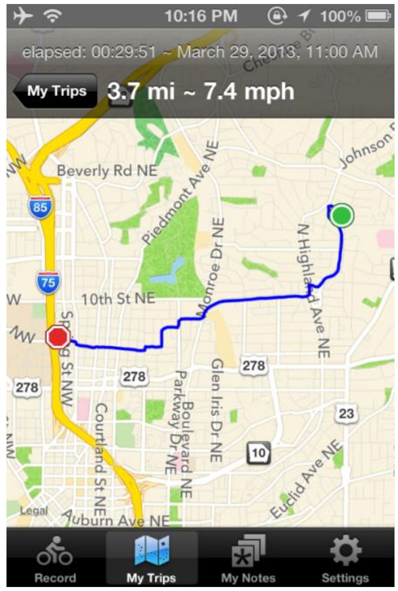


Fig. 1(f)

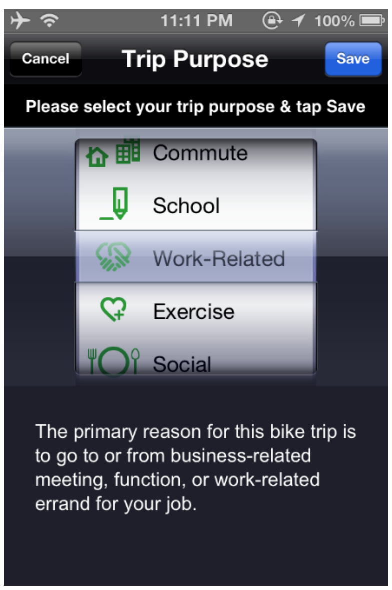


Fig. 1(g)

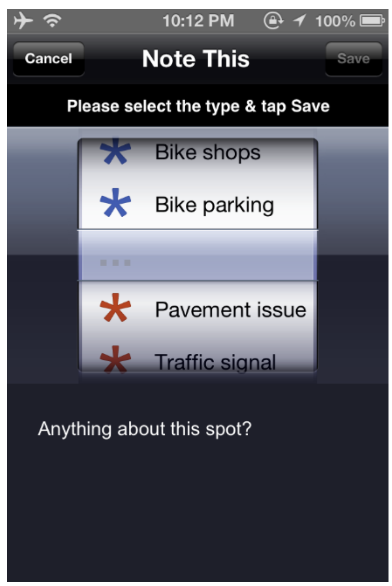


Fig. 1(h)

Figure 1. Cycle Atlanta User Interfaces: (a) Socio-demographic information, (b) Riding characteristics information, (c) Trip start screen, (d) Trip end and record screen, (e) &(f) Viewing trips, (g) Choosing a trip purpose, (h) Adding a note on anything about a particular spot

In addition to tracking cyclists' trips, the app also provides options to enter personal information, including age, email address, gender, ethnicity, home income, zip codes (home, work, and school), cycle frequency, rider type, and rider history. Figure 2 shows the group breakdown of the demographic and rider type categories in Cycle Atlanta. The breakdown of age, gender, income and ethnicity was kept similar to the breakdown as found in the household travel survey conducted by Atlanta Regional Commission while the rider type and rider history categories are exclusive and unique to the design of Cycle Atlanta. The rider history field allows users to specify how long they have been cycling and can choose from categories like ‘since childhood’, ‘several years’, ‘one year or less’ and ‘just trying it/just started’. This rider attribute can be used to see if people who biked from childhood are more likely to adapt to bicycling as a mode choice or if there is any relation between biking preferences and the years of experience the bicyclist has.

Age	Gender	Ethnicity	Home Income
Less Than 18	Male	White	Less than $20,000
18-24	Female	African American	$20,000-$39,999
25-34		Asian	$40,000-$59,999
35-44		Native American	$60,000-$74,999
45-54		Pacific Islander	$75,000-$99,999
55-64		Multi-racial	$100,000 or greater
65+		Hispanic/Mexican/Latino	
		Other	

Cycle Frequency	Rider History
Daily	Since childhood
Several times per week	Several years
Several times per month	One year or less
Less than once a month	Just trying it/just started

Figure 2. Cycle Atlanta Demographic Categories and Rider Characteristics

The other parameter, cyclist type, was adopted from a Portland State study (Dill and McNeil 2012) and modified to suit the needs of Cycle Atlanta. At its present form, there are four cyclist categories that the user can choose from – 'strong and fearless', 'enthused and confident', 'comfortable but cautious', and 'interested but concerned'. At its core, this new parameter actually represents an interaction term between the attitude of the rider and his/her comfort level with the city road network. This new parameter is thus best capable of presenting the varying need of different cyclist types and can actually be used as a proxy for risk aversion attitude for modeling bicyclist route choice.

Providing all such details is entirely optional on the part of the application user and is strictly protected by the privacy protection provisions of Institutional Review Board (IRB). Therefore, users are not required to enter information into any of the fields, and can still use the app to record trips if they choose not to share their personal information.

The motivation of this research is derived from two main issues that it attempts to address. First, the research uses smartphone-application-based crowdsourced data for understanding the infrastructural need of the cyclists and their behavioral preferences. The advantages of using crowdsourced data is that it provides a less costly and labor intensive method of sampling for the planning agencies; while for the participants, it provides them with the flexibility of participation without any time and locational constraint. However, since the process is based on voluntary participation, there is a possibility of having self-selection bias in the collected data; i.e. any trend in the data may be heavily influenced by motivated and enthused cyclists rather than the infrequent and casual cyclists who are in more need of cycling infrastructure than the avid cyclists. The data may also suffer from systematic bias towards a young and high income generation, because of its dependence on technology for data input.

Therefore, the ability of this research to realistically predict the behavioral preferences of the cyclists of Atlanta will add to the literature on the suitability and reliability of crowdsourced data for planning purposes.

This research is also motivated by earlier studies that addressed cycling route choice problems, but did so without any consideration for attitudinal differences between cyclist types based on rider characteristics (for example, between ‘strong and fearless’ cyclists versus ‘comfortable but cautious’ cyclists), cycling experiences or socio-demographics like age or gender. It has been speculated that perception of safety varies across all these categories, but the effect of varying perception of safety on route choice decisions is a poorly studied area that needs further exploration. The Cycle Atlanta app collects optional demographic data from its users including age, gender, ethnicity, cycling frequency, rider type (comfort level), rider history (cycling experience), home income and zip codes (home, work and school). As will be evident from the literature review, the cycling experience and the type of cyclists are new features exclusive to the Cycle Atlanta dataset and inspire this research to identify how route choice preferences differ across rider type, age, gender, experience and any interaction thereof.

Finally, the aim of this research is to create a practical and useable tool that the planners and engineers can use to identify links that are critical for having a connected bicycling network between more frequently used origins and destinations. Research has shown that the decision to bicycle depends significantly on having a connected shortest route between origin and destination. However, as mentioned earlier, since this connection is related to perception of safety, it is important to understand how the definition of a connected network varies between different types of cyclists. This research contributes in that area by developing a quality-of-service measure based on perceived level of traffic stress as it varies across rider types.

CHAPTER 2. PARTICIPATORY PLANNING AND CROWDSOURCED DATA

INTRODUCTION

Researchers have long emphasized the importance of public participation in the planning process as a critical component to the successful implementation of any plan *(*Innes 1998, Burby 2003, Slotterback 2010*)*. Broad public participation leads to "greater legitimization and acceptance of public decisions, greater transparency, and efficiency in public expenditures, and greater citizens' satisfaction" (Burby 2010). According to Burby, inclusion of stakeholders with varied interests and different backgrounds makes a plan comprehensive, acceptable, and more easily implementable (Burby 2010). Moreover, a participatory planning process effectively recognizes that "society is pluralist and there are legitimate conflicts of interest that have to be addressed by the application of consensus building methods" (Hague et al. 2003). With these traits in mind, participatory planning has the potential to involve broader and more diverse groups of people into a planning dialogue and hence, can bring in newer perspectives and ideas to the planning problem at hand (Rabinowitz 2013).

Recent research, however, suggests that citizen involvement at different stages and levels of planning is steadily declining in the U.S. (Skocpol and Fiorina 1999, Galston 2004, Pew Research Center 2013). This seems counterintuitive given the fact that over the last few decades, information accessibility and remote participation has been facilitated and made easier through the ubiquitous use of the internet and web-based social media. A wealth of emerging technologies have brought about significant new forms of communication and interaction, providing diverse new ways of documenting, sharing, and reflecting on the world at a truly global scale.

One possible reason for this decrease in citizen involvement may be that planners and policy makers have yet to embrace technology-mediated forms of participation and instead still rely on methods that require the physical presence of the participant. These methods limit the availability of the planning process by placing time and location constraints on participation and may also alienate or further disadvantage citizens for whom travelling to a planning meeting is neither physically nor financially viable.

One strategy for overcoming limited participation from interested stakeholders is to implement multiple methods of participation that participants can choose from, depending on their level of comfort and accessibility (Wagner 2012). Slotterback (2010) proposed that along with the traditional methods of public hearings and open-house meetings, more accessible modes of communication like project websites, web-based meetings and discussions may be adopted as a means of increasing public participation in the planning process. Toward that end, the purpose of this paper is to encourage the use of crowdsourcing platforms as a possible means of involving people from diverse walks of life to effectively participate in planning for transportation systems without putting additional financial burden on the transportation agency. This chapter highlights the successful use of crowdsourcing in a few transportation projects, providing examples of projects that have overcome many of the initial challenges of adopting crowdsourcing in transportation planning and establishing a robust starting point for future work.

This chapter is organized as follows: first, the concept of crowdsourcing is discussed along with a commentary on the existing platforms and types of crowdsourcing and the issues associated with crowdsourcing in general. Then, the crowdsourcing case studies in transportation planning are presented with reference to the different genres of crowdsourcing. The first group of case studies focuses on receiving feedback from users while the second group focuses on use of crowdsourcing for data collection. A standalone example is provided at the end of the case studies sub-section as it deserves special mention because of its use of data quality editors to ensure data usability and validity, thereby addressing one of the biggest issues of crowdsourced data collection.

CROWDSOURCING: CONCEPTS, PLATFORMS AND ISSUES

At its conception, social computing focused mainly on building a network of collaborators and facilitating online communication between groups. This has eventually given rise to open source platforms and forums where people with similar motivation and outlook can come together to solve issues and to find answers to problems that affect their community. Crowdsourcing is one such example where an organizer or an organization is able to use the network of collaborators to solve a problem that would otherwise be cost or labor intensive or for which the available expertise within a defined organization is unavailable or insufficient.

Crowdsourcing has been alternately defined as: the outsourcing of a job (typically performed by a designated agent) to a large undefined group in the form of an open call (Howe 2006); a process that "enlists a crowd of humans to help solve a problem defined by the system owners" (Doan et al. 2011); or "a sourcing model in which organizations use predominantly advanced Internet technologies to harness the efforts of a virtual crowd to perform specific organizational tasks" (Saxton et al. 2013). Common across these alternate definitions is the notion that crowdsourcing invites all interested people to form an open forum of ideas that can eventually lead to a solution of the assigned problem. As Howe (2006) states, crowdsourcing utilizes the "latent potential of crowd" to achieve a solution to a problem that the crowd can relate to.

According to Saxton et al., crowdsourcing systems are characterized by three main features – the process of outsourcing the problem, the crowd, and a web-based platform for collaboration (Saxton et al. 2013). Outsourcing a problem generally implies getting a task done by outside sources even when it could have been performed by people within a system; in crowdsourcing, outsourcing is done in cases where either the in-house expertise has failed to produce a solution, or is an expensive means to produce a solution, or there is no in-house expertise available to use for solving the issue. Crowdsourcing systems also rely largely on an anonymous unidentified group of people ("the crowd") to come together willingly instead of the business sub-contract model of outsourcing where the task is performed by a previously identified and designated group of people or a company (Saxton 2013).

An important subset of the general crowdsourcing idea is the concept of citizen science, in which amateurs contribute to research projects in conjunction with the professional scientists. Goodchild used the term "citizen science" in describing crowdsourced geo-mapping, referring to the fact that information generated through crowdsourcing, although not of the level of a professional, helps in expanding the reach of science (Goodchild 2008). The nature of participation of the people in citizen science projects takes different forms depending on the type of the project and can range from data collection to data analysis, from instrument building to taking part in scientific expeditions. Recent citizen science projects tend to focus on utilizing the ever increasing reach and availability of electronic gadgets, particularly mobile phones and sensors, for data collection and monitoring purposes. In their experiments, Kuznetsov and Paulos (2010) and Kuznetsov et al.(2011) provided citizen scientists with sensors to monitor air and environmental quality, while the CycleTrack project in San Francisco used GPS enabled mobile devices to record cyclist trip data (Hood et al. 2011). Citizen science projects are gaining popularity as an alternative to cost intensive data collection efforts, particularly in cases where the information needed is global in character, and are thus being increasingly used for planning and monitoring purposes.

Existing Crowdsourcing Platforms and Systems

Despite the advantages discussed in the previous section, crowdsourcing can only be successful if a platform exists that can provide open access to incorporate, modify, and synthesize data. There are four different versions of this shared platform – the wiki system, open source software, geocrowd mapping, and mash-ups using crowdsoucring data (Kitchin and Dodge 2011). Wiki systems are mainly centered on authoring information; open source software provides a platform to share and co-develop program source code; geocrowd mapping entails collecting, cleaning, and uploading GPS data; and mash-ups are combinations of some or all of these. While maintaining coordination among people coming from different backgrounds and motivations is a significant challenge, this voluntary coming together of a mass of people for a purpose is particularly useful in tackling problems that are large scale, e.g., mapping of a country.

Beyond the fundamental concept of providing an open access and participatory platform for a large group of people, crowdsourcing projects can be markedly different depending on the purpose of the project, the nature of involvement required, or if some special expertise is required for participation. Figure 3 schematically represents the different categorizations of crowdsourcing systems which are further discussed herein. Based on the nature of involvement of the participants in solving the problem, Doan et al. (2011) classified crowdsourcing systems as either explicit or implicit systems (Figure 3). Explicit systems are standalone systems where users participate and collaborate in executing a stated problem like answering questions via the web, testing software and writing web content (e.g., Wikipedia). Within explicit systems there are four different types of tasks that users generally perform: (i) evaluating (e.g., book review), (ii) sharing (e.g., feedback on system performance), (iii) building artifacts (e.g., designing T-shirts at Threadless.com), and (iv) executing tasks (e.g., collaborating on finding gold mining spots). Implicit systems can be standalone or piggyback depending on projects. In standalone implicit crowdsourcing systems, the system owners benefit from the indirect input provided by the users; the direct user input is used to solve a problem that is related to but not the same as the issue that the users of the system respond to. For example, although humans are more efficient at image recognition than computers, they are not necessarily willing to perform this task unless it is packaged in a form that attracts them. In the Extra Sensory Perception (ESP) game, the participants are shown images and asked to guess common words to describe those images as part of playing the game. Those words are then used to label the image (Doan et al. 2011). In piggyback crowdsourcing systems, the traces of the users are collected from an entirely different system – ad keywords generated based on Google and Yahoo search traces are examples of piggyback implicit crowdsourcing systems.

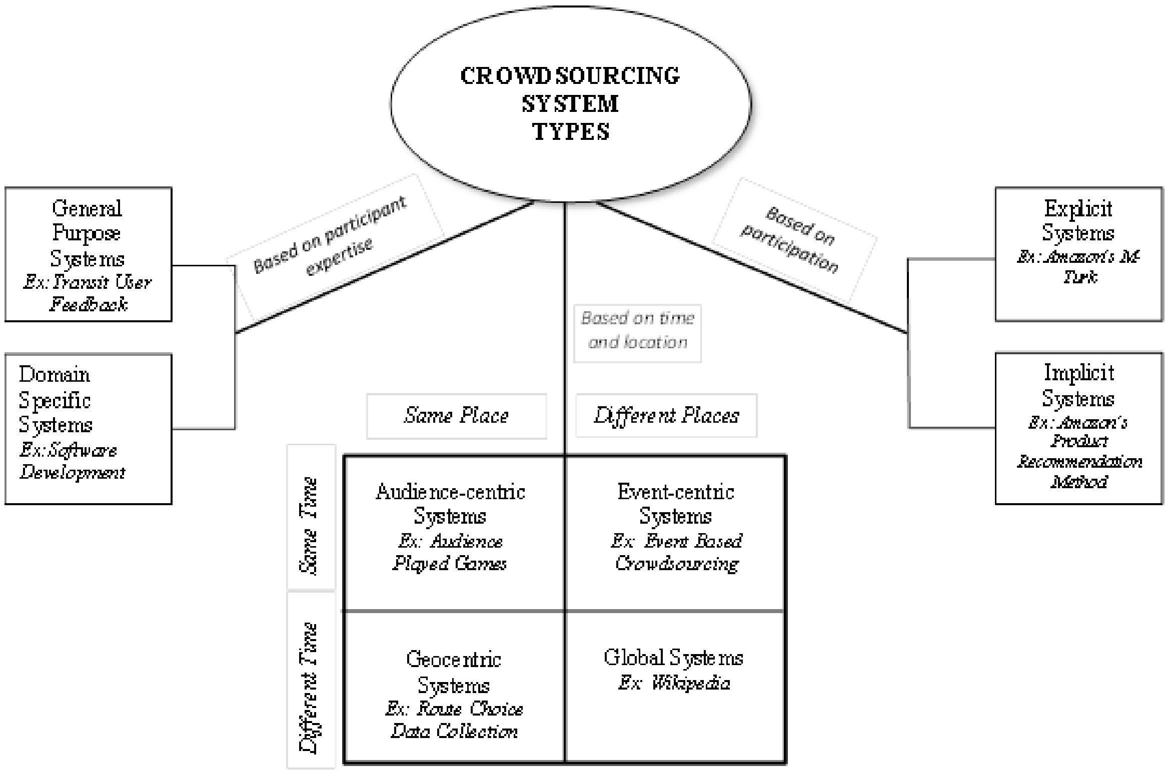


Figure 3. Classifications of Crowdsourcing Systems (Doan et al. 2011, Steinfield et al. 2013, Erickson 2010)

Steinfield et al. (2013) categorized public participation as either general purpose or domain specific systems. General purpose systems do not require any special expertise from the contributors and are not targeted to any user group in particular, while domain specific systems are designed for a special purpose user group (Figure 3). For example, most crowdsourced service quality feedback does not require any special expertise on the part of the participants and are hence, general purpose systems. Conversely, developing or beta-testing open source software through crowdsourcing requires expertise in particular programming languages and platforms and are hence, domain specific systems.

Crowdsourcing systems are further classified based on whether the system is local or global in scope and whether the system is time bound or not (Erickson 2010) (Figure 3). For crowdsourcing systems where the participants are at the same place at the same time, the system is termed as audience-centric (e.g., clickers used in class discussions). For systems where participants can be at different places but the crowdsourced event is time bound i.e., it has a start and end time between which the collaboration has to happen, such systems are termed as event-centric. An example of event-centric crowdsourcing is organized online brainstorming sessions triggered by an event and spanning over a limited period of time. Systems where collaboration can happen between people from different places and over an indefinite period of time are termed global crowdsourcing systems (e.g., Wikipedia). Finally, systems where people are at the same place but the crowdsourcing is an ongoing process are termed as geo-centric crowdsourcing – an example is bicycle route choice data collection for a city.

Crowdsourcing Issues

As crowdsourcing keeps evolving and gaining popularity, different and larger systems are being experimented with and the issues uniquely associated with the characteristics of the systems are gradually surfacing. For example, domain specific systems automatically reduce the crowd size by requiring some expertise from the participants while implicit systems have the issue of not having explicit participant consent in using their contribution for the actual purpose of the project. *A priori* understanding of the project characteristics and hence its category can often largely help in setting up plans early to overcome such issues. The last case study presented in this paper is one such example where instead of making the system domain specific, an expert group is used as data quality auditor. This helps in retaining a larger participant base as well as provides the necessary check on the usability of the data collected through a general crowdsourcing system. As crowdsourcing gets applied to different domains, and as the scale and scope of crowdsourcing systems increases, additional techniques for addressing these system specific issues need to be developed based on the requirement of the projects.

Beside the unique issues of the systems, operation and maintenance of crowdsourcing systems in general suffer from four major issues – (i) how to recruit and retain the participant base, (ii) user capabilities, (iii) how to aggregate the information provided by the users and (iv) how to evaluate the contribution of the users (Doan et al. 2011). The problem of recruiting and retaining participants is a major issue in adopting crowdsourcing for any project. Depending on the purpose of the project, it is often important that feedback is obtained from users with particular skills or expertise. Furthermore, retaining participants is often important for understanding a trend over time – to allow the crowd's understanding of the problem to evolve throughout the process. The use of recurring campaigns and marketing strategies at frequent intervals (along with new releases of apps) is suggested where applicable so that people remain curious about the project and the developers can help maintain a participant base over time (Priedhorsky et al. 2007). Using incentives in the form of material benefits as well as acknowledgement of contribution in the form of gratification announcements at project sites make people feel encouraged to participate in the project and can help recognize diverse kinds of contributions from the crowd (Doan et al. 2011).

Dealing with user capability is an important issue in citizen science projects and in problem solving projects where participants are required to have some background to appreciate the assigned task. While participatory planning may not generally require special skill sets, in cases where the planning process targets a special group, it is important that the participants are aware of the specific problems of that group (e.g., planning for bicyclists' needs requires presence of people who bike in that area so that the relevant problems and issues are brought up on the table). In such cases, the crowdsourcing process may be most successful if it is designed as a domain specific system – rather than a general purpose one – where specific tools and capabilities are made available to develop and maintain relevant user capabilities.

Problems with data quality and challenges with data aggregation are two important issues that often undermine the benefits of crowdsourcing systems. Regarding the importance of data quality, Heipke (2010) assessed that “quality issues have been a primary point of debate since crowdsourcing results started to appear”. From that perspective, a degree of loose hierarchical authority is needed to ensure that the data is useful for its intended purpose. Additionally, aggregation of the data from crowdsourcing is often a complicated task given the volume of responses received from a diverse pool of crowd participants. Coping with data issues is either often labor intensive as large data sets need to be manually cleaned, or more cost intensive as complex data management systems and processes need to be put into place in an attempt to reduce sources of human error.

Evaluating the contribution of the user is commonly accomplished by setting up an automatic screening program to evaluate the validity of user-submitted information based on predefined criteria. The screening program rejects any input that does not follow the set criteria and thus only valid information is retained. However, this kind of automation is possible only in cases where the input is sufficiently normalized to be evaluated programmatically – in cases where the responses are descriptive or subjective, there needs to be a manual evaluation stage where each response is evaluated based on its potential contribution to the project. Such manual processes are labor and cost intensive and are prone to subjective biases of the evaluator, but are also much needed in order to ensure data quality for the project.

CROWDSOURCING AND ITS USE IN TRANSPORTATION

The characteristic of crowdsourcing that makes it suitable and useful for transportation planning is that it voluntarily brings together a large group of people on the same platform to address common issues that affect them. The use of crowdsourcing works successfully for local purposes through localized knowledge and acquired experiences (Brabham 2009) because people in a region tend to identify themselves with the region where they live, work, and socialize, and are generally more interested in the systems that affect them (Erickson 2010).

A survey of existing transportation systems which use crowdsourcing reveals that the predominant purposes of using crowdsourcing in these projects are either data or feedback collection from the users. For example, one popular use of crowdsourcing is in collecting route choice data from bicyclists using the GPS functionality of the user's cell phone – such data are not readily available through the standard data collection procedures and designing a separate survey for a small population of users is often not cost effective for regional planning agencies. Crowdsourcing in this case helps the geographically dispersed and diverse population of cyclists to work together on a common interest without financially burdening the planning agencies. Similarly, crowdsourcing can also help in collecting feedback from a socio-demographically diverse range of users of any transit system that can be immensely useful for improving transit service quality and standards.

Transportation related crowdsourcing systems designed to date can be implicit or explicit standalone systems as defined by Doan et al. (2011) and discussed in the previous section. They may also be either geocentric systems where only local users are engaged or global systems where any person can contribute to the system. Extending the categorization of public participation as defined by Steinfield et al. (2013), transportation crowdsourcing systems may be further classified as either general purpose or domain specific systems. General purpose crowdsourcing systems do not require any special expertise from the contributors and are not targeted to any user group in particular, while domain specific systems are designed for a special purpose user group.

Examples of transportation related crowdsourcing are presented below with reference to the above mentioned classification systems: the first group of examples focus on receiving feedback from users while the second group of examples focuses on use of crowdsourcing for data collection. A standalone example is provided at the end of the sub-section as it deserves special mention for its use of data quality editors to ensure data usability and validity and at the same time, maintaining a broad user base, thereby addressing one of primary challenges of crowdsourced data collection. The section is followed by a discussion on the advantages and disadvantages of crowdsourcing systems.

Crowdsourcing Case Studies

(i) *User Feedback Based Crowdsourcing Systems* :

Three seminal examples of general purpose user-feedback systems are SeeClickFix (http://seeclickfix.com), PublicStuff (http://www.publicstuff.com) and FixMyStreet (http://www.fixmystreet.com), all of which rely on public feedback about neighborhood issues and have been successful in mobilizing communities to take up the task voluntarily. While FixMyStreet is essentially for users to report road maintenance issues, the developers have a similar transit-based tool called FixMyTransport (http://www.fixmytransport.com). SeeClickFix and PublicStuff can be used to report "any non-emergency issue anywhere in the world that a user wants to be fixed" (seeclickfix.com), be it infrastructural or governance related. In SeeClickFix, users can also set up neighborhood watches where they monitor and report local community issues which are then taken up by advocacy groups or elected officials, and solutions are proposed publicly. It is evident from the nature of the participation in these cases that no special expertise is expected from the users. It is interesting to note that the majority of the reported issues are local and community oriented in nature, reinforcing the concept that crowdsourcing can be successful in addressing local and regional issues, making it suitable for transportation planning.

Shareabouts is another example of a general purpose crowdsourcing system utilizing an innovative approach. Shareabouts *(*http://www.shareabouts.org*)* is a web-based system that uses maps to generate user feedback on preferred location of facilities and amenities. A few ongoing projects that use Shareabouts are (i) Chicago Bikeshare where people pin preferred bikeshare locations on the map provided, (ii) North Carolina Alternative Bike Route Plan where people can vote for preferred alternatives as well as mark any segment that they think might be an inappropriate alternative, and (iii) Philadelphia Bike Parking Survey where crowdsourced information is collected for estimating the bike parking capacity of the existing stations and plan for future expansion. In Boston, Street Bump (http://streetbump.org) is a mobile application that uses a smartphone's accelerometer to detect potholes and other street hazards as people drive around the city – the geo-located street quality data collected through crowdsourcing is automatically uploaded and integrated with the city's process for locating and fixing pavement quality issues.

A transit project using a general purpose crowdsourcing system, OneBusAway was created to address the reliability issues with on time performance of transit systems in Seattle and to expand upon existing transit tools in the region. OneBusAway provides several feedback mechanisms (email, Twitter, blog, bug tracker) that allow users to make comments or suggestions about the tools (Ferris et al. 2010). The design of the various tools, along with development of new features, has been further shaped by feedback from users via several user studies and the IdeaScale feedback platform (another general use tool that can be applied to transportation). Because OneBusAway is open source software, users have also submitted improvements of their own to the code. Thus, users eventually become partners in development and design of the OneBusAway program, which promotes a sense of community among the transit riders in the region and a sense of ownership of the program. This ownership is an important factor in maintaining the user base for the program (Ferris et al. 2010).

Another general purpose crowdsourcing project related to transit systems is Tiramisu transit (Zimmerman et al. 2011), a user feedback based real time information system for public transportation in Pittsburgh. Tiramisu Transit, a ‘crowd-powered transit information system’, uses riders as the human equivalent of automated vehicle location (AVL) thereby providing an innovative alternative to more traditional cost intensive data collection. Tiramisu Transit is a smartphone app developed by researchers at Carnegie Mellon University to improve users' transit experiences and transit accessibility (Zimmerman et al. 2011) Upon activation, the app shows a list of buses or light rail vehicles scheduled for arriving at that time – this list is based off past arrival data as well as real time data sent by riders on the vehicle. Tiramisu provides an option for the rider to indicate the level of fullness of the bus, which aids people with disabilities to choose the bus they want to access. Once aboard, the rider can use Tiramisu to find out which stop is next and to report problems, positive experiences and suggestions. Use of Tiramisu is motivated by the rider’s ability to use the same real-time arrival and fullness information they are reporting.

(ii) Crowdsourcing Systems for Data Collection

While issue reporting crowdsourcing systems like SeeClickFix and FixMyStreet do not call for any specific expertise from the user, there may often be systems where data and information are needed from a group with specific expertise or purpose, termed domain specific systems (Erickson 2010). Domain specific systems may be nested under a general purpose system, such as the bike projects undertaken using ShareAbouts. While all of these projects use the same crowdsourcing platform, the information is collected for one specific region, because it is more useful if it comes from the cyclists who use the facilities on a regular basis. Examples of standalone domain specific systems are the crowdsourced bike route data collection projects undertaken in San Francisco, Minneapolis, Atlanta, and Austin. These projects focus on developing smartphone apps and websites for cyclists to record their trips so that region-specific bikability maps can be created and facilities can be constructed on route segments as required.

CycleTracks (Hood et al. 2011) and Cycle Atlanta *(*www.cycleatlanta.org*)* are both projects for collecting bike route choice data through GPS enabled smart phones. The creation of CycleTracks by the San Francisco County Transportation Authority (SFCTA) in late 2009 was motivated by the lack of data on cyclists, cycling infrastructure, and eventually cyclist route choices. Traditionally, such data would be collected through public meetings because cyclists represent only 1-2 percent of commuters making vehicle count methods less useful. CycleTracks made participation in data collection for cyclists more accessible by moving data collection to the increasingly common smartphone use. In CycleTracks, first time users are asked optional information to determine cycling habits, such as riding frequency, age, gender, and zip codes for home, work, and school. Users record their trips by starting the app when they set out on a ride and then saving and uploading their data once they've reached their destination. The app records bicycle trip route, time, distance, and average speed, along with user-reported trip purpose and notes. The trip data are wirelessly uploaded for analysis of cyclist route choice and is later used for planning facilities along the predicted routes (Hood et al. 2011).

Cycle Atlanta, a similar smartphone app for collecting data about cyclists and their routes within the city of Atlanta, was built off the open source codebase of the CycleTracks app. Cycle Atlanta also uses the GPS capabilities of smartphones to save and upload routes to provide basic data on how cyclists navigate the city, but the project team added features to the app including the ability to note with photos and textual descriptions of specific locations as either issues (pavement issues, traffic signal, enforcement, etc.) or amenities (bike parking, public restrooms, water fountains, etc.). The app also includes the collection of additional demographic data including cyclist ability and history as an indicator of comfort level to allow analysis of route data around an established taxonomy of urban cyclists (Dill and McNeil 2013), and to enable correlation with existing cyclist count and census data. As a distinctly different approach from CycleTracks, Cycle Atlanta categorizes cyclists into groups based on their cycling comfort level. The categories include strong and fearless, enthused and confident, comfortable but cautious, and interested but concerned. This categorization helps in understanding the preferences of different types of cyclists in choosing routes and hence can be immensely informative in creating a tailored application like bike maps for any particular group of users. Since the apps were launched in early October 2012, Cycle Atlanta has been used by over 1500 cyclists in Atlanta who have recorded more than 20,000 rides – represented by over 30 million individual data points. These data are the core piece of the City of Atlanta's effort to facilitate more streamlined communication between planners and cyclists.

A significant role of domain-specific crowdsourcing is in providing information from an otherwise unrepresented or underrepresented community. For example, due to the small size of the cycling community, bicycle maps are not commercially attractive and hence, are rare. Therefore, crowdsourced maps and geowikis are particularly suitable for understanding bicycle routes and for developing bicycle route maps (Masli 2011). Also, cyclists can benefit from regularly updated information, which is easy to maintain through "delegated responsibility among a motivated community with common purpose" (Masli 2011). Cyclopath (http://www.cyclopath.org), a crowdsourced geowiki-based bicycle map developed by researchers at the University of Minnesota, provides an example of a domain specific use of crowdsourcing in transportation. Cyclopath maintains an active database of user-contributed bicycle routes and trails within the Minneapolis – St. Paul metropolitan area. The users of Cyclopath can add, modify, and delete roads and bike trails, segments thereof, points of interest, and neighborhoods. In addition, Cyclopath allows users to add notes and tags describing any feature on the map, such as 'bumpy' or 'closed'. Revisions are public and tagged to user logins for transparency and accountability. Cyclopath also has features that help the community to moderate itself. A list of 'Recent Changes' is also maintained, so that other users can identify and undo malicious modifications to the geowiki. Finally, Cyclopath allows a user to rate bike routes on a five-point qualitative scale (excellent, good, fair, poor, and impassable) for their own use and for aggregation to enhance bikability ratings. The Cyclopath community has made more than 13,000 revisions since release (cyclopath.org).

(iii) Standalone Crowdsourced Data Quality Auditor System

Along with generating data from underrepresented groups, domain-specific crowdsourcing also helps in data quality management, which is an issue with self-reported data in crowdsourced systems. As a study by Wiggins and Crowston (2013) revealed, most of the systems that use voluntary public participation include some form of expert control over the data. An expert user group can act as a bridge between general users and the system by filtering required information from general information and then by translating back the feedback from the system to the general users in a meaningful way. This helps in maintaining a feedback loop that is important in retaining participants and also prevents losing the critical mass which is often the case if the entire process is domain specific.

A standalone example of such an effort in transportation systems is the transit ambassador program initiated by the OneBusAway, Seattle program (Gooze 2013). The transit ambassadors are a *super user group*, with a solid understanding of the transit network and basic computational and analytical skills. Their role is to filter the incoming general purpose crowdsourced information and channel it to the respective departments within the transit agency for necessary action. Three core goals of the program development included addressing problem resolution, engaging the community, and improving agency-rider communication. Beginning in the fall of 2011, a number of errors with the real-time transit prediction data surfaced, affecting over 77% of a survey of riders (Gooze et al. 2013). While the OneBusAway mobile application included an error reporting function to allow users to identify errors experienced, the amount and quality of the crowdsourced reports began to overwhelm the OneBusAway administrators. Oftentimes, reports were duplicates of previously reported errors or the information submitted was incomplete and required additional effort to utilize it. With upwards of 500 errors reported on a weekly basis, the time required to evaluate these reports and any attempt to leverage them in order to resolve underlying problems with the real-time system would have required an effort from a collection of individuals. In contrast to previously described crowdsourcing programs, this was not an issue of data collection, but rather a problem with information management. The management of the errors required the coordination between the agency, the OneBusAway administrator and the riding community; however, due to the constrained resources of each organization, there was no single contact to coordinate between these entities. This role fell to a collection of volunteer super users, or OneBusAway Transit Ambassadors. Figure 4 provides a visual summary of the flow of information established within the program and the role of the Ambassadors in coordination of the process.

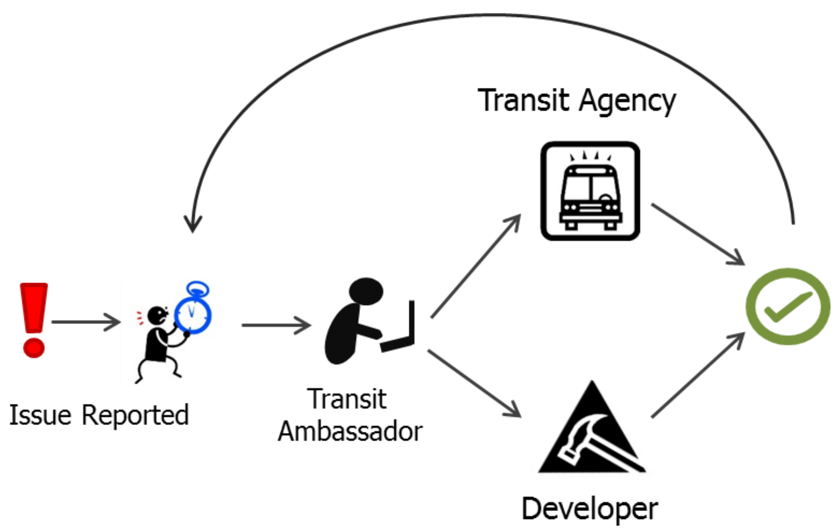


Figure 4. Information Flow of the Transit Ambassador Program

An initial group of three Transit Ambassadors were recruited from the rider community via blog solicitation and email outreach. The Ambassadors were provided resources such as transit schedule data, agency alert information and an “Error Decision Matrix” to assist in categorizing the crowdsourced error reports. All error reports were collected into an online database that allowed the Ambassadors to not just validate the error but to identify the nature and possible cause. This action of validation was a necessary and vital step in transforming the overwhelming amount of crowdsourced information from varying noise into usable knowledge. Finally, the Ambassadors aggregated the information to forward onto the transit agency a clear and concise summary of notable issues reported by riders. For example, the summary of errors by vehicle and route provided the transit agency with valuable supporting information to help target potential actions to improve the real-time information system. The overarching role of the Ambassadors provided a level of expertise that could accurately evaluate the incoming error reports and thus efficiently triage and divert any relevant issues to the appropriate organization.

Providing a behind-the-scenes look at the underlying issues confronting the transit agency allowed the Ambassadors to relay that information to the rider community and to provide some context to the errors that everyone was experiencing. For example, a typical public relations response by the agency would have been interpreted far differently as compared to the Ambassadors relaying this information out in the community, which provided an enhanced level of trust. While some underlying real-time issues could not be resolved by the agency, the Ambassadors provided a means to explain to riders why an issue could not be fixed and how they could best adjust to the situation.

The success of the outreach exhibited by the Ambassadors and their role in representing not just the agency but the riders themselves gave validity to the potential that a fully deployed Ambassador program has within any real-time information system. With the proper adjustments to the available agency support and an expansion of the amount of Ambassadors, a Transit Ambassador program can effectively accomplish the core objectives and serve as not only a means for improving the real-time information product but serve as a mechanism for an agency to fully engage its riding community in a method that improves the overall functionality and quality of the transit service provided.

Summing It Up

Despite the fact that crowdsourcing has been used in transportation planning only recently, it is evident from the case studies presented that it has immense potential in augmenting/replacing traditional survey methods, particularly for groups of stakeholders who have a small user base in the transportation system. As seen with all systems, crowdsourcing also has its own issues that need to be addressed through proper planning and understanding of the system. Although there are criticisms with respect to data quality and data management issues, it is undeniable that crowdsourcing has been successful in engaging groups of people in solving a problem that affects their community. Crowdsourcing for bike route choice data has successfully solved the issue of data aggregation, defining a role of the users and linking their contribution to the final goal of the project by developing facilities for the bicyclists in San Francisco, Minneapolis, Atlanta and Austin. Meanwhile, transit information systems like Tiramisu Transit and OneBusAway have been very successful in redefining the role of their users in monitoring service standards and quality. The OneBusAway transit ambassador program has the potential to address the data quality issues associated with crowdsourcing by filtering and validating the data received from participants before the data reach the agency.

Most of the crowdsourcing systems use devices and technologies that are readily available and low cost – often crowdsourcing is based off devices that are owned by individuals (as in cycling data collection in CycleTracks and CycleAtlanta), involving no major financial investment on the part of the system. In an exemplary case, the Tiramisu project, described previously, uses crowdsourcing to actually replace the requirement of high cost AVLs. Tiramisu provides an example of ideal civic engagement in transit planning and operation where riders take care of other riders without the direct involvement of the transit agency and create an information sharing legacy that is beneficial to both the users and the agency. With current funding limitations, crowdsourcing can be a preferred alternative to involve the public despite limited resources.

It should be noted however, that CycleTracks and CycleAtlanta are based on the widespread popularity and reach of the smartphone technology for crowdsourcing. While smartphones are easy to carry powerful devices that provide an inexpensive means of data collection, the usage of smartphones is restricted among some groups, such as those above 40 years old – thus, using smartphones for data collection comes with the issue of bias towards the input from these populations *(Windmiller et al. 2014)*. Further research into possible biases arising from smartphone data collection is underway *(Windmiller et al. 2014)* and preliminary results show that age, income and ethnicity are the major factors that should be considered in smartphone data collection. This, however, can be addressed using proper outreach efforts and using supportive traditional methods for people who are not currently smartphone users.

CONCLUSION

Crowdsourced transportation projects bear evidence that crowdsourcing has the potential to bring together a large group of people on the same platform when there is an issue that affects them all. Systematic use of information and feedback from users for the purpose of transportation planning or for improving service standards is receiving significant attention recently and smart technology based crowdsourcing provides an ideal platform for engaging a broad group of users with limited additional financial burden on the system or the agency – possibly even replacing costly equipment. Crowdsourcing for data collection is found to be financially most effective in cases where the user base is small but enthused and motivated as in the case of bicyclists – in such cases crowdsourcing has a huge potential in augmenting the standard data collection procedures by including the requirements of the otherwise marginalized groups of users. Examples of a few potential transportation related cases where crowdsourcing can be used are traffic data collection, getting user feedback for different systems, pavement and sidewalk quality monitoring and in understanding people's opinion in creating new facilities.

Crowdsourcing issues are mostly concentrated around problems with data quality, accuracy and data aggregation. However, these issues may be addressed through proper planning and with an understanding of the final goal of the crowdsourcing project. Further research and implementation of such strategies in real life projects are needed to establish a generic framework of crowdsourcing for transportation planning.

CHAPTER 3. DATA CLEANING AND MAP MATCHING

INTRODUCTION

Traditional travel behavior related data collection methods depend on surveys where the respondent is either required to recall an experience or incident or extrapolate present day experience to a future scenario. In both cases, the collected data suffers from respondents' personal bias and is dependent on recall efficiency of the respondents. In particular, surveys that span for multiple days and record multiple events are more likely to suffer from missing data on events that the respondent did not consider important for the reporting purpose. For example, household travel diary surveys that are generally used to record all household trips over a week typically tend to under report short trips and connection trips (walk to transit, errand trip during lunch). The other problem with multi - day surveys is that respondents suffer from survey fatigue towards the end of the survey period, leading to high rates of attrition and often, a low response rate to begin with.

Replacing or complementing traditional survey methods with data collected via advanced technologies like sensors and GPS enabled devices is gradually becoming popular among transportation researchers across the world (Shen and Stopher 2014, Du and Aultman-Hall 2007, NCHRP 2014). The ability of such technologies to record data without direct effort of the participant as well as the ability to capture revealed preference data of the participants in real time makes these technologies immensely usable for intelligent transportation systems where routing decisions are dynamic and made in real time. As recall effort is removed, the data quality is also improved substantially even though data collected through GPS is not always completely accurate. Multiple studies have shown that GPS devices capture 20% – 30% more trips than traditional survey methods (Bricka & Bhat, 2006; Stopher & Greaves, 2009; Stopher & Shen, 2011)

The use of GPS for data collection becomes particularly useful for user groups that are small in number but have distinct trip characteristics and hence infrastructure preferences such as pedestrians or bicyclists. Planning for such groups without data from the users themselves can be seriously flawed and has proven to be significantly less effective for encouraging people to use non-motorized transport. The regional planning agencies, however, are often unable to allocate sufficient funds and human resources to conduct a separate data collection effort for users like cyclists who constitute only about 1 % of all transportation system users. Towards that end, there has been a shift recently towards designing participatory planning processes that help people come together on a virtual platform or to provide indirect input to the planning process through data contribution. These methods rely on people’s willingness to participate in the planning process and to share data voluntarily, and passive data collection technologies play a significant role in facilitating such platforms. These platforms, if successful, can remove additional data collection burdens for the planning agencies and can bridge the gap between data need and data availability. Cycle Atlanta is one such platform which was created for cyclists to voluntarily come together to better inform the City of Atlanta of their preferences and requirements.

The primary purpose of the Cycle Atlanta app is trip recording, in which the app uses the GPS capability of the phone to record the location of a user on a second by second basis and the trip is uploaded when it is complete. This offers a high level of precision for recording trips over time, allowing for detailed research on route preferences not previously available on mass-scale data. However, to enable studies involving route choice and decision making models, these GPS points must be processed so that a trip follows the existing system of linear paths which comprise the road network. This task is non-trivial, particularly for the scale of Cycle Atlanta, in which we are analyzing about 20,000 trips with each trip consisting of about 1000 geocoded points.

In this chapter, we present the preprocessing routines as well as two map matching procedures that were designed and used to process the Cycle Atlanta data. Parts of the preprocessing routines were borrowed from existing literature while parts were designed in house and proved to be very effective in reducing the computational burden at later stages of analysis. The map matching algorithms were developed as a combined outcome of a decision process on the part of the user as well as available network characteristics. The entire process of cleaning the GPS data and matching trips to network was carried out using open source software R and the code will be freely available in the GitHub repository of Cycle Atlanta.

BACKGROUND

This research is based off data collected via GPS enabled smartphones and hence, the literature review presented here is focused on issues that are commonly found in the data collected through GPS. It should be noted though that there are other passive data collection devices like stationary sensors and in-vehicle information systems which have their own advantages and disadvantages that are beyond the scope of this paper and hence not covered here.

GPS enabled devices are capable of recording the latitude, longitude, heading, speed, altitude, and timestamp of a user at intervals of one second, three seconds or five seconds. Generally, data are collected per second but Shen and Stopher (2009) compared data adequacy for trip identification and mode detection across different time intervals and concluded that a five second interval provides sufficient data for trip and mode identification while reducing the number of GPS points to be dealt with. The location point is recorded via triangulation of signals from at least 3-4 satellites and issues with GPS data are mainly related to loss of signals or reflection of signals across high rise buildings.

There are two different types of GPS devices that are used for data collection – (1) GPS units either connected to a hand held Personal Data Assistant (PDA) or carried separately by the participant – in both cases, the GPS unit is dedicated to the purpose of data collection and it is the respondents' responsibility to monitor the unit and, often, record additional data such as sociodemographic data that the GPS unit cannot capture (Du and Aultman-Hall 2003, NCHRP 2014) and (2) GPS enabled devices that passively collect location data via some application – the primary purpose of the devices is not collection of location data but having GPS capability enables it to record location data of the user with timestamps. The respondent burden is hugely reduced with these devices as the users do not have to take part directly in the data collection system.

The data collected via passive GPS enabled devices often suffer from noise and uncertainty that require a substantial amount of post processing efforts to render the data usable (Shui and Shalaby 2007, Quddus et al. 2007, Pyo 2009, Auxhaussan 2012). The inaccuracy occurs from multiple sources - loss of signals at certain locations, particularly at the start and end of the trip (cold start/warm start), reflection of signals between tall buildings before it reaches the device (urban canyon effect), not having enough satellites for accurate triangulation, and interference of signals at intersections. The data cleaning methods developed to deal with these problems are generally rule-based and use the number of satellites, speed and heading change, as well as position jump to identify points that are part of the trip (Stopher et al. 2005, Lawson et al. 2010, Shen and Stopher 2014, Wolf et al. 2001).

Trip Start and End

One of the most difficult parts of GPS data cleaning is identifying trip start and trip end points. At the start of the trip, the GPS device requires some time to acquire satellite signal, stabilize and then start recording the trip. Until that time, the signal jumps, resulting in a scattered cluster of points which make it difficult to identify the actual start or end point of the trip. Since the GPS records data at a regular interval, the most intuitive approach to identify trip end points is to identify points where the time interval between any two consecutive points is more than a specified dwell time. Multiple studies have used this to identify trip start and end points, albeit with different dwell time criteria. The most commonly used time gap is 120 seconds, the Highway Capacity Manual prescribed maximum signal timing, so that trips are not terminated when they are stopped at the signals (Schonfelder et al . 2002, 2003, Wolf 2001, Du and Aultman-Hall 2007, Stopher et al. 2005). Other studies have used dwell time from 45 seconds (Pearson 2001) to 3 minutes (Doherty et al. 2000). However, in most cases, dwell time based identification is supplemented by other criteria like zero speed, zero change in bearing (Doherty et al .2000, Schussler and Auxhausen 2009, Lawson et al. 2010), difference in latitude and longitude (Stopher et al. 2005) and point density (Schussler and Auxhausen 2009). Since most of the methods were developed for vehicular traffic, Schussler and Auxhausen 2009, Doherty et al. 2000, and Stopher et al. 2003 used engine stop and start time difference as a measure of trip end and start identification too.

Du and Aultman-Hall (2005) used a maximum and minimum dwell time criterion along with distance from network and bearing changes to identify trip ends. They used a buffer distance of 15 meters from road centerline and GPS points outside this buffer were discarded. For the points within that 15 meter buffer zone, dwell time and bearing changes were used to identify trip start and end points. Du and Aultman-Hall (2005) experimented with multiple combinations of minimum and maximum dwell times and the algorithm was tested for minimum dwell times of 20, 40 and 60 seconds while maximum dwell times tested were 60, 100 and 140 seconds. They found that any dwell time between the maximum and the minimum is generally associated with a 180° bearing change.

Besides identifying trip start and end points, there are also issues with signal loss and signal noise. NCHRP 775 provides a comparison of three methods for dealing with GPS noise filtering (Stopher et al.2005, Schussler and Auxhausen 2009, Lawson et al. 2010) and compares the results with a base case where the actual trip is known. All the three methods use dwell time threshold, number of available satellites and a threshold value of horizontal dilution of precision for noise filtering, along with zero speed and zero heading change. The report classifies the error of not removing an invalid point that has been removed in the base case as a Type 1 error while removing a point not removed in the base case as a Type 2 error. The analysis of three major data cleaning methods show that all the methods tend to have more type 1 errors than type 2 errors, whereby the methods tend to retain more points than making the error of removing a point that is part of the trip (NCHRP 2014).

Map Matching

Map matching is the process of relating input data from global positioning systems to a spatial road network map to correctly identify the position of a vehicle on the road network (Quddus et al. 2007, Zhou and Golledge 2006, Auxhausan et al. 2009). For transportation related studies, as Zhou and Golledge (2006) mention, the map matching process is a means of transferring road attributes to the travelled route so that further inferences can be made about travel patterns and preferences.

Map matching can be either a real time process which is used for most ITS related applications or can be done as a post processing step where the vehicle tracking information is not essential to update network information instantaneously. For this research, map matching was done as a post processing step of data collection and cleaning. The process of map matching can be either a point to point matching, point to line matching, or curve to curve (polyline to polyline). The process of matching is significantly complicated and is highly prone to errors due to the compounded effect of uncertainty of GPS points as well as inaccuracy in the road network. Quddus et al. (2007) mentions the accuracy of the match depends both on the quality of the road network map as well as the algorithm used because different algorithms may provide different efficiency for the same map.

Map matching algorithms are primarily classified as geometric, topological, and probabilistic. Other advanced methods that have been used for map matching include adaptive fuzzy logic (Kim and Kim 2001) and Bayesian belief theory (Zhou and Golledge 2006). The geometric algorithms generally only take into account the distance of the GPS point to the road segments and return the nearest segment to the GPS point as the matched segment. The biggest issue with distance based matching is that if there are parallel segments that are sufficiently close to each other, then the GPS point can match to the wrong link. This is particularly true if the correct link has less nodes than the wrong link or if the matching buffer zones overlap. To overcome these issues, other parameters can be added to the matching criteria which include travel direction and bearing change, road attributes like one way lanes, speed limits and distances travelled on a segment (Najjar and Bonnifait 2003, Taylor 2001). In view of completing multiple parameter matching criteria, Quddus et al. (2003) suggested using a weighting approach to decide on the best match while Kim and Kim (2001) proposed an adaptive fuzzy network based training for the same purpose.

The issue with not using network information for the matching process often can result in matches that do not form a logical path which is important for transportation related purposes. The topological approaches to map matching restrict the matching using topological properties of the underlying network like link connectivity and altitude difference (Greenfield 2002, Ochieng et al. 2003). However, constructing a path from link connectivity depends on successive link matches which can result in a wrong path even if one of the links is matched wrong. To address this issue, Pyo et al. (2001) and Marchal et al. (2004) suggested keeping multiple candidate solutions for each GPS point along with some measure of goodness of fit for each candidate. At the end of the matching process, the measures are aggregated and a match is decided based on that score.

Statistical map matching methods have used linear regression models to fit the GPS points to a road network (Lakakis 2000). Bierlaire et al. (2013) proposed a map matching algorithm that first generates a path between origin and destination points and then calculates the likelihood that the GPS points are generated along that path using geographical and temporal information. However, since the path is generated based on shortest distance between the origin and destination points, if the vehicle uses any other criteria for path choice, it will be difficult to find an appropriate match.

METHODOLOGY

In this section, we discuss the methods we used to collect, prepare and match the GPS data to the road network map of Atlanta. Multiple platforms and algorithms are available for most data processing steps and the related ones were tried and tested within this research. However, because bicycling trips tend to differ from vehicular trips, using algorithmic approaches developed for car traffic proved to be not very effective for either cleaning or matching. For example, since bicyclists often trade off shortest path for safest path or other considerations like scenery, any map matching algorithm based on the shortest path approach could not be used. Some algorithms could not be used because of the scale of application – this research used 15 million GPS points and a road network with more than 18,000 links. In addition, bicyclists are likely to have more options at each stage of the trip and may not be required to follow the conventional routes of vehicular traffic, making it more ideal to be modeled as a decision process at every intersection than a predetermined route between origin and final destination. Therefore, most of the methods used and described here were designed keeping in mind the particular nature of the dataset and were modified as required iteratively during the process.

Data Collection

The GPS data used for this study were collected via the smartphone application Cycle Atlanta. Launch of the app in October 2012 was announced by the Mayor of the City of Atlanta and the app was widely publicized through various cycling advocacy groups and social media. Participation in using the app is voluntary and no reward was offered to record trips. The app is designed for both Android and iPhone GPS-enabled smartphones and is freely available for download from the app stores. The user has to turn on the app at the start of the trip and geolocation of the user is recorded from that point until the user indicates a trip end. The trip is not saved unless the person wants to save the trip which s/he can indicate via the ‘save’ button. At that point, the trip is uploaded to the secured database maintained by Georgia Tech. For each trip, the app records latitude, longitude, altitude, speed, time, and horizontal and vertical accuracy at an interval of 1 second. Figure 5(a) and 5(b) show an example of the original uncleaned data from the Cycle Atlanta app.

(a) (b)

Figure 5. Original Uncleaned Data: (a) Raw GPS Points (b) Trip Lines Constructed from GPS points

Data Cleaning and Noise Filtering

The data issues found were classified as (1) systemic, (2) operational and (3) random. Systemic errors include issues that occur because of the use of GPS capability and are general in nature across all studies using GPS data. For example, cold and warm start problems, signal loss issues and urban canyon effects will be classified as systemic errors within our classification system. Operational errors are often errors introduced in the system by the users. These issues include forgetting to turn off the app after trip completion, using the app for non- cycling trips, using shortcuts and bylanes that are not part of the street network, etc. These errors will depend on the purpose of data collection and consequently on the participants. Random errors are most often related to systemic errors brought into the data due to use of GPS, but the nature of the errors are specific to each instant of recording and, hence, no standardized method can be applied to remove such errors.

The data cleaning was done following established practices from the literature. However, knowing the difficulty of map matching with noisy data, a lot of effort was put into early cleaning stages before running the snapping algorithms and therefore, the standard practices were modified and customized to suit our needs. Some additional criteria were also implemented keeping in mind the specific nature of the dataset. Efforts were made to attain a balance between retaining as much necessary information as possible in contrast to retaining data that is erroneous and can increase the computational burden for a later stage of analysis. It should also be noted that the app did not report the number of satellites, so that information could not be used for data processing in our case.

Operational Error Handling

As the study focused on bicyclists in Atlanta, at first, the data were checked for geographical limits – since the app is freely available to anyone owning a smartphone, it was suspected that the data might have trips that are not Atlanta based. Therefore, any point with latitude and longitude beyond the latitude and longitudinal boundaries of Atlanta [NW: 33.886823, -84.551068; SE: 33.647808, -84.28956] was removed from the dataset. Some trips were recorded over multiple days which can happen if the user forgets to turn off the app at the end of a trip and the app continues to record trips as continuation of the first trip until it is turned off. In such cases, the day with maximum number of recorded points was retained and data from other days were discarded.

Random Error Handling

Duplicate removal and basic data filtering: Two types of duplicates were identified: (1) points within same trip having same timestamp but different latitude and longitude and (2) identical latitude, longitude, timestamp and user id but different trip id. So, while in the first case, all points except the first point are removed, in the second case, the trip with the lower trip id is retained and the duplicates are removed. Some points were recorded with invalid timestamp (0000-00-00, 00:00:00) – these points were also removed during this step.

Horizontal Accuracy: As mentioned in NCHRP report and used in other research, the horizontal accuracy (haccuracy) threshold could be between 5 and 20 for a point to be a valid point. For this research, haccuracy limit was set to 30 – any point with horizontal accuracy more than 30 was removed from the database. The higher-than-standard limit was set after experimenting with haccuarcy values of 10, 20 and 30. Since the data are from cyclists who tend to use bylanes, cut throughs and underpasses which do not always have a good signal, setting a higher accuracy expectation resulted in removing too many points and created connectivity issues as well as sparse data problem for shorter trips.

Systemic Error Handling

Speed, Distance and Heading: The app recorded instantaneous speed at each point as well as latitude and longitude. Since the app is designed for cyclists, points with instantaneous speed more than 12 mph were discarded. Points with zero speed were further checked for distance and bearing from a point preceding 10 points upstream and the point succeeding 10 points downstream. If either distance or bearing change remained zero, the point was removed from the database.

Sparse Data: Some trips were found to have too few points for proper identification. The threshold ratio of distance to number of points was set such that speed between two consecutive points should not exceed 100 feet per second. If more than 50% of a trip consisted of points that did not match this criterion, the trip was discarded.

Noise Filtering: To filter points that are mainly signal jumps, a criterion similar to sparse data was used. If the distance from the point 10 steps before and/or 10 steps ahead of the point being checked is such that it cannot be traversed in the time between the timestamps at a speed of 70 feet per second, then that point is removed from the dataset. An additional check, if a large group of 10 or more points are major deviations, was used to remove any GPS point that was over 5,280 feet from the point that is 10 positions prior to it.

Data Reduction: The Cycle Atlanta dataset consists of about 15,000 trips, with each trip on average recording more than 1000 GPS points. One of the concerns was using such a large amount of data for map matching and our initial experiments of map matching in ArcGIS and R proved to be significantly slow and often problematic. Therefore, we decided to apply the Douglas-Peukar algorithm to remove points for a trip that aren't necessary to identify its true shape and distance. The algorithm first identifies the starting and ending point. Then it finds the point in the line that is furthest perpendicularly from that line. If that distance from the point to the line is greater than the tolerance, then that point is kept and it remaps the "line" from the starting point to that furthest point. That new line then finds the point that is furthest from itself and does the same check. If it is within the tolerance, then that point is dropped and the algorithm checks for the next furthest point. It iterates over the whole line until all points have been checked. The tolerance used for our purposes was 5 feet, with the projection of the NAD83, UTM18 (North American Datum 1983, Universal Transverse Mercator, Zone 18). This means that any point that varied by more than 5 feet from the line between the points before and after, was kept, and any point that was under 5 feet was removed. This struck a good balance between ensuring that much of the route shape was kept while limiting the number of points needed. In addition, it ensured that for snapping purposes, no streets were skipped that were clearly traveled on. For a street to be snapped, there had to be a point near it. Therefore, reducing the number of points with too large of a tolerance would have resulted in long straight segments of a path with no points kept. The 5 foot tolerance allows for enough precision while clearly reducing the number of points required.

With this simplified line, we can then interpolate the points in it for snapping purposes and determine whether there are any path duplicates. The function ST_DumpPoints in PostGIS takes the simplified line and returns the points of that line, thus reducing the number of points to snap from roughly 15 million to 2 million.

Map Matching

There were two processes involved in the map matching part of the project. First, the network to which the GPS points to be snapped had to be cleaned and processed for the purpose of matching bicycling trips which are quite different from vehicular trips. For example, bicycling trips do not happen on freeways and keeping the freeways in the map might result in some nearby trips snapping wrongly to freeway segments. Therefore, we preprocessed the network map to better suit our purpose.

Three data sources were used to create the road network map. The Atlanta Regional Commission's street network shapefile (RC_ROUTES) was obtained from the travel demand modeling group of Atlanta Regional Commission (ARC). It is a modified version of the roadway database maintained by the Georgia Department of Transportation (GDOT) and focuses on state managed roadways rather than locally managed roadways and bikeways. However, it contains the most comprehensive inventory of roadway characteristics like speed limit, annual average daily traffic (AADT), number of lanes, truck volume etc. which are useful information for route choice modeling at a later stage. The second data source used is Open Street Map's (OSM) bicycle map for Atlanta. The OSM map has local roads and locally managed facilities which were not present in the RC_Routes map. The two maps were spatially joined based on a buffer distance to get a more complete map of the road network of Atlanta. The resulting map was then cleaned for non-bicycling facilities like freeways. The final data source was the Metro Atlanta Bicycle Facility Inventory. The location of on street parking on roadways with conventional bicycle lanes and buffered bicycle lanes was manually coded in ArcGIS using Google Earth imagery. The treatment of intersection approaches with right turn only motor vehicle lanes that connect to links with conventional bicycle lanes, buffered bicycle lanes, or protected cycle tracks were also manually coded in ArcGIS using Google Earth Imagery. As a final measure, the trips were plotted on the map and checked for links traversed by cyclists but missing in the network. Such links were manually added where more than 2 bicycle trips were found to follow a path but the path was not marked as a link in the network. This was assumed to be mainly because of tendency and ability of bicyclists to use cut-thrus and private alleys which are not marked in regional network maps. However, shortcuts through parking lots were not added as links although there were multiple such cases.

Rebuilding the network file was done in ArcGIS partly because it was easier to merge multiple shapefiles in ArcGIS and also because we felt it was necessary to have a visual check on the merging and link imputation processes. The final shapefile was then imported to R using rgdal [shapefile to .ogr] (Bivand et al. 2015) and shp2graph [.ogr to igraph object] (Lu 2014) packages within R. The map was imported with the option of retaining all the associated properties as dataframes i.e. all the link characteristics were imported into R along with the shapefile. The graph object was then used for map matching using the packages igraph (Csardi et al. 2015) which is a fast and efficient package for handling large networks and spatstat (Baddaley et al. 2015) which is an advanced spatial statistical package for analyzing spatial patterns.

Once the network map was ready, we used two different methods to snap the trip data onto the network. The first base case used a combination of geometric and topological approach while the second method was designed to make use of the adjacency properties of network elements and reduce the computational burden of network search at every single instance of GPS recording.

Scenario 1: Base case

In this method, a spatial cross distance matrix between each trip point and each node in the network was created using '*spatstat*' package. The matrix was then sorted to get the minimum distance node for each trip point. The list of the nodes thus generated was first cleaned to keep one instance of a node occurring for multiple times *consecutively*. It should be noted that separate instances of the same node were retained in the list where sequences were broken by instances of other nodes occurring in between sequences of the same nodes. For example, a node sequence {4, 5, 5, 2, 5, 5} would be filtered via this process to {4, 5, 2, 5}. Then the list was filtered for oscillation of nodes – if a different node was visited after and before two instances of same node, that node was removed from the list i.e., after this step, the list showed earlier should become {4, 5, 5}. Finally, the list was filtered for unique values as one node could be the nearest node of multiple trip points and the list would contain multiple entries of that same node – at this stage, the list should only contain {4, 5}. The unique nodes were then *successively* checked for connectivity using the '*igraph'* package. As identification of a trip was dependent upon proper identification of the start point, a special check was introduced to identify the start point. The first node in the filtered list was checked to see if it was the start node for at least two consecutive links i.e. if at least three consecutive points after the first point were successively connected. Upon failure, the next point was checked for similar criteria and the process was repeated unless a match was found. In all cases, a match was found within the first three nodes. The connectivity check was then run successively on each node starting from the matched start node and the connected nodes were retained in a list. At any point during the process, a list was terminated when a point was found to not be connected to any of its three consecutive nodes and a new list was started with the first point that was not connected to its preceding point. So, if the starting list was c = {1, 2, 5, 6, 7, 8, 9, 10}, node 1 was first checked if it was connected to node 2; if true, node 2 was then checked for connectivity with node 5 and then node 5 with node 6. If node 5 was not connected to node 6, it was checked if node 5 was connected to node 7 or node 8. If it was not connected to either node 6, node 7 or node 8, then list was terminated as c = {1, 2, 5} and a new list d = {6} was started. Next it was checked if node 6 was connected to node 7 – if true, then node 7 was added to list d such that d = {6, 7} and the check continued with addition of connected nodes to the list and termination of a list whenever the last node of the list had three consecutive nodes to which it was not connected. However, random checks on trips indicated that loss of connectivity either occurred at the beginning or at the end of the trips, implying that the points were GPS errors rather than actual trip points. The process would give multiple lists of connected nodes which could then be aggregated and used to find the actual trip once the last node was reached. Any list with 3 nodes or less was not considered for the purpose of aggregation as most of the road segments in the original map had more than 3 nodes (consisting of 2+ polylines). When aggregating, a check was introduced for a common node in the adjacency list of the last node of the first list and the first node of the second list and that node was added into the list between the other two nodes to get a continuous path. One of the biggest advantages of this method would be that it inherently adjusts for sparse data through aggregation of links and via imputation of the node from the adjacency matrix which could be missed due to sparse data. So, if there are two separate lists c = {1, 2, 5, 8} and d = {6, 7, 10, 11}, upon aggregation, the method checked the adjacency list of node 8 and the adjacency list of node 6. If the adjacency of list of node 8 was {12, 13, 15} while the adjacency list of node 6 was {16, 13 and 9}, then node 13 was added to the end of list c and then list c and list d were aggregated.

The connecting links for successive pairs of nodes were then retrieved with their associated properties and stored as the chosen path for the trip.

Pseudo Code:

$\boldsymbol{while}\ trip - id\ \neq \emptyset\ \{$
$d\ \leftarrow \boldsymbol{crossdist}(points, nodes);$
$c \leftarrow \{\boldsymbol{for\ each}\ col\ in\ d\ get\ row_{id}\ with\ \mathbf{min}\ cell\ value\};$
$\boldsymbol{filter}\ c\ by:$
$\boldsymbol{for\ i\ in\ c}[\boldsymbol{n}]$
$\{$
$\boldsymbol{if}\ \{$
$c(i) = c(i+1)$
$c[n] = c[n] - c(i+1);$
$\};$
$i = i + 1;$
$\}\ end\ for$
$\boldsymbol{return\ c};$
$\boldsymbol{for\ i\ in\ c}[\boldsymbol{m}]$
$\{$
$\boldsymbol{if}\{$
$c(i-1) = c(i+1);$
$c[m] = c[m] -\ c(i)$

$\};$
$i = i + 1;$
$\}$
$\boldsymbol{return\ \ c};$
$\boldsymbol{filter\ c\ by}:\ unique\ row_{id};\ return\ c;$
$\boldsymbol{check}: \boldsymbol{for\ j\ in\ c[k]}\ \{$
$\boldsymbol{if}\ \{$
$\boldsymbol{are.connected}(c(j), c(j+1)) == \boldsymbol{T}$
$j = j + 1;$
$\boldsymbol{elseif}$
$\{$
$\boldsymbol{are.connected}\big(c(j), c(j+2)\big) == \boldsymbol{T},$
$\boldsymbol{check}: \boldsymbol{if}\ \{$
$\boldsymbol{are.connected}\ \big(c(j+1), c(j+2)\big) == \boldsymbol{T}$
$j = j + 2;$
$\}$
$c[k] = c[k] - c(j+1); j = j + 1;$
$\}$
$\boldsymbol{elseif}$
$\{$
$\boldsymbol{are.connected}\big(c(j), c(j+3)\big) == \boldsymbol{T}$
$\boldsymbol{check}:$
$\boldsymbol{if}\{$
$\boldsymbol{are.connected}\big(c(j+2), c(j+3)\big) == \boldsymbol{T}$
$\boldsymbol{if}\{$
$(\boldsymbol{are.connected}\big(c(j+2), c(j+1)\big) == \boldsymbol{T}$
$j = j + 3\};$
$c = c[k] - c(j+1); j = j + 2\}$
$c = c[k] - c(j+1) - c(j+2); j = j + 1;$
$\}$
$j = j + 3; \}$
$counter = 1$
$\boldsymbol{truncate}\ c\ as\ list\ [counter] = c[j];$
$counter = counter + 1$
$\}$
$c\ \ = c[k] - c[j];$
$\boldsymbol{repeat\ process}$
$\}$
$\boldsymbol{Check}:$
$adjacency\ list\ of\ last\ element\ of\ list(i)\ with\ adjacency\ list\ of\ first\ element\ of\ list(i + 1)$
$p\ \leftarrow \{\ common\ node\};$
$final_{list} \leftarrow unlist\ all\ lists$
$\}$
$trip - id < - next(trip - id)$

Validation:

Five random trips were chosen and the snapping algorithm was run on them. In all the five cases, the removed nodes were the last few nodes in the list. We also ran a Djikstra's shortest route between the first and last node before removal and in all the cases, the program could not find a valid route between them. However, when the same program was run to find the shortest route between the revised first and last node, it was able to find valid routes between them which indicates that removed points have a high probability of being GPS errors rather than valid trip identification points.

We did not have the issue of sparse data, so it was not possible to validate the effectiveness of the adjacency matrix approach for finding missing links. Since the proposed algorithm only checks for one missing node, it might have issues when there are multiple missing links in between two valid nodes. In such cases, it may be worthwhile to assume that the trip took the shortest path between the last node of the first list and the first node of the second list and then retrieved the nodes associated with that path. The final and complete node list then will consist of the two original lists with the nodes on the shortest path added in between the last node of the first list and the first node of the second list.

Scenario 2: Map matching using adjacency matrix

In this case, first we created an adjacency matrix for all the nodes in the graph. Then, only the first five GPS points of a trip were selected and their nearest nodes were searched for in the entire network. The nearest nodes were then checked for connectivity and the first point that led to three consecutive connected links was flagged as the trip origin. We next created an adjacency list for the selected node and added the selected node to that list. For the next GPS point, we searched the nearest node from that list. Once that node was selected, we took the adjacency list of that node, added the node back itself and searched for the nearest neighbor for the next GPS point and the process was repeated until we reached the trip end. The final list was then first filtered for repeated occurrences of the same node consecutively and then it was filtered for instances where one different node occured between two instances of same node.

Pseudo Code:

$while\ trip - id\ \neq \emptyset\ \{$
$\#\#\ \boldsymbol{get\ the\ start\ point}\ \#\#$
$points\_sub\ \leftarrow points[1:5]$
$d\ \leftarrow crossdist(points\_sub, nodes);$
$c[m] \leftarrow for\ each\ col\ in\ d\ get\ row_{id}\ with \min cell\ value\ ;$
$c[n] \leftarrow filter\ c[m] by\ unique\ row_{id};$
$check\ if\ c[n] is\ single\ element, then\ return\ that\ node$
$for\ i\ in\ c[n]\{$
$a[j] \leftarrow adjacency\ list(c(i));$
$check:$
$if\ c(i+1) is\ in\ a[j] == \ true; d[\ ] = [c(i+1), ..]$
$i = i + 1\ ;$
$elseif\ c(i+1)\ is\ in\ a[j] == \ false,$
$f[\ ] = c(i+1)$
$i = i + 1; \}\ end\ if$
$check:$
$if\ len\ [d] > len[f]$
$nodes = nodes - [f\ ]$
$else\ nodes = nodes - [d\ ]$
$\#\#\ \boldsymbol{End\ Start\ Point}\ \#\#$

$\#\#\ \boldsymbol{Map\ Matching}\ \#\#$
$for\ i\ in\ nodes\ \{$
$a[j] \leftarrow adjacency.list\ [nodes(i)];$
$c[k] \leftarrow concatenate\ (a[j], nodes(i)]$
$check:$
$min.distance(nodes(i+1), c[k]);$
$select\ c(k) with \min distance\ from\ nodes(i+1);$
$d[i] \leftarrow \{c(k), ..\}$
$i = i + 1; \}\ end\ for$
$\#\#\ \boldsymbol{Post\ Processing}\ \#\#$
$for\ i\ in\ d[i]\{$

$check$:
if {
$d(m) == d(m+1)$,
$d[i] = d[i] - d(m)$;
$m = m + 1\}$ $endif$
$return\ d\}$;
$for\ i\ in\ d[i]\{$
$check$:
$if\ \{d(j-1) == d(j+1)$,
$d[i] = d[i] - d(j)$
$\}$
$return\ d; \}\}$

There are multiple advantages to the second method of map matching as compared to the first one. First, it is computationally efficient as it has to search the entire network only for the first few GPS points instead of all the GPS points in a trip. In the case of large network files like that of Atlanta, this is a huge advantage. Second, it does not require storing a huge distance matrix for further processing which renders the process memory efficient. Third, by using an adjacency list, it already asserts connectivity and no further check is required. Fourth, without any nearest distance threshold, it will always find an adjacent node for a GPS point which reduces the risk of prematurely ending a trip. The threshold distance value for the nearest node can be modified as needed for the research. Finally, the biggest advantage of this algorithm is that with slight modifications, it can be used to probabilistically determine a route when GPS data are sparse. By adding link characteristics to the adjacency list, the algorithm can use the utility maximization or cost minimization concept to identify which link will be chosen by the user probabilistically.

However, the problem with this method is that it is contingent upon correctly identifying the previous node. If any node identified within the process is wrong, there are chances that the trip will be identified completely wrong. This is particularly true if no threshold value for distance between the GPS point and the nearest node is set: in that case, every GPS point will return a near node however far it may be, and the trip will be a set of nodes that do not form a feasible path. On the other hand, this can also be seen as a second step of data cleaning where GPS points that are actually errors are forced to follow the actual route instead of having to do multiple checks as in the previous algorithm.

CONCLUSION

In this part of the research, a standardized data cleaning and map matching procedure was developed which can be useful for any related GPS based data collection effort. The code will be made open source and will be available for any future studies. It was found during this stage that having a complete and connected street network map is essential for proper execution of any snapping algorithm. Multiple platforms, both GIS based and script based, were tested and it was found that methods based on scripting languages like ‘R’ are computationally more efficient – however, there needs to be a procedure that will enable the results of scripts to be displayed in any GIS based software so that a visual check can be performed. In the future, this research will be undertaken by the current study’s researchers.

CHAPTER 4 GEOGRAPHICAL DISTRIBUTION OF CYCLE ATLANTA USERS

This chapter looks into the spatial distribution of Cycle Atlanta users. The smartphone app provides the users with options to provide their home zip code, school zip code (in cases where the user is a student) and work zip code. The reported zip code was used to plot the distribution of the users within Atlanta Metropolitan area. Further, the user distribution was plotted against ethnicity, age, income and population density distributions within Atlanta to identify any spatial correlation between user self - selection and user's home/work/school zip code.

METHODOLOGY

Users who indicated that they lived outside of the Atlanta metropolitan area were purged from the database. This was done by sorting the table of users by the home zip code they reported and deleting the records that contained zip codes outside of the Atlanta area.

For geographic analysis using ArcGIS, a shapefile of Atlanta zip codes was obtained from the Atlanta Regional Commission (ARC). However, ARC's zip code shapefile did not contain all of the zip codes reported by Cycle Atlanta users. For example, the zip code 30332, which contains part of Georgia Tech's campus, was not part of the ARC zip code shapefile. To rectify this, missing zip codes were drawn into the shapefile using Google Maps and a shapefile of city streets for guidance. The chosen study area comprised of zip codes located either completely or partially within Atlanta city limits and/or the Perimeter (I-285), as shown by the red shading in Figure 6.

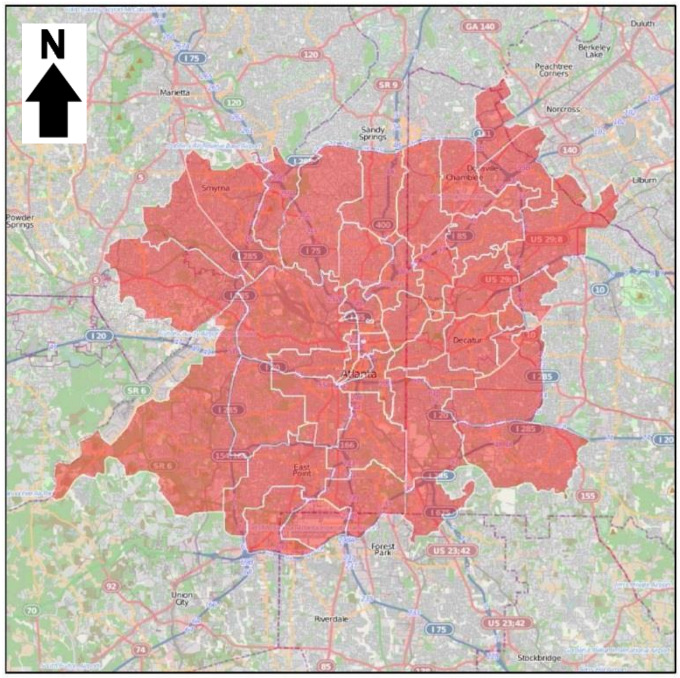


Figure 6. Zip Codes completely or partially within Perimeter (I-285) or City of Atlanta limits

The Cycle Atlanta datasets were queried to return a list of each discrete home zip code in that dataset as well as how many people in the dataset reported that zip code as their home zip code. The datasets were then joined to the study area zip code shapefile using home zip code as the common field. This resulted in a data table containing fields for home zip code and number of Cycle Atlanta users living in the zip code.

A map was created in ArcGIS to show the percent of cyclists within the study area who reported each zip code as the one they resided in. The map was shaded in such a way that darker zip codes had a greater percentage of the dataset's cyclists (within the study area). For context, an OpenStreetMap basemap was added to each map.

To analyze the relationships between Cycle Atlanta user home zip codes and demographic traits associated with those zip codes, four maps were generated using census data. For each map, the shade of the zip code polygon represents the demographic variable (zip code median age, median annual income, percent of non-white residents, and population density). The size of the black dot over a zip code represents the percent of

Cycle Atlanta users residing there. The median age and percent non-white data were obtained from American Community Survey table DP05, "Demographic and Housing, 2007-2011 5-Year Estimates". The median income data were obtained from American Community Survey table S1903, "Median Income in the Past Twelve Months (In 2011 Inflation-Adjusted Dollars), 2007-2011 5-Year Estimates". The population density data were obtained from American Community Survey table B01003, “Total population, 2007-2011 American Community Survey 5-Year Estimates".

Figure 7 shows that cyclists are concentrated in the "intown" part of Atlanta, near the center of the Perimeter. Specifically, zip codes east of the Downtown Connector (the north-south running Interstate near the center of the study area) have the highest percentages of cyclists living within them.

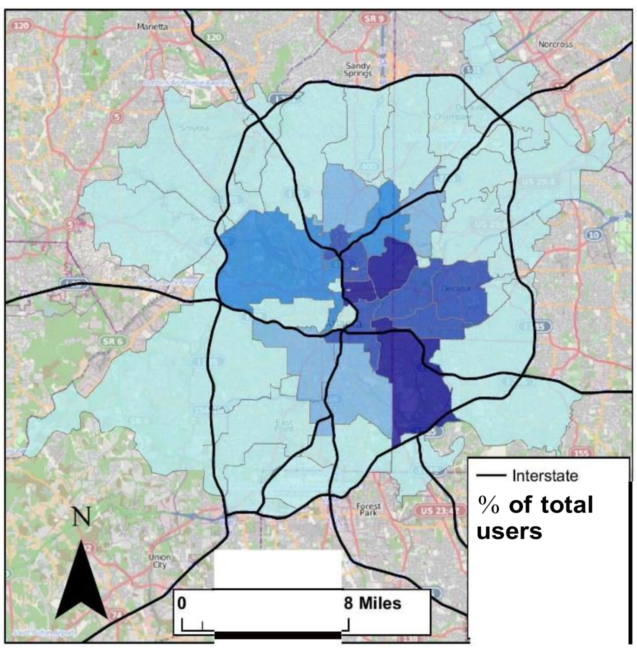


Figure 7. Distribution of Cycle Atlanta Users by Home Zip Code

SPATIAL CORRELATION

This section examines the correlation between the percent of Cycle Atlanta users residing in a zip code and several demographic characteristics of the overall population in that zip code – median income, median age, non-white population, and population density.

Figure 8 shows a comparison between the percent of Cycle Atlanta users living in a zip code and the percent of non-white, non-Hispanic residents living in that zip code. The darker the zip code, the greater the percentage of non-white residents; the bigger the black dot over a zip code, the higher the percentage of Cycle Atlanta users living there. It is difficult to see a clear relationship between the two variables. Some zip codes have a low percentage of non-white residents and a high percent of Cycle Atlanta users living there such as 30306 and 30307 (located between E4 and E5). However, some zip codes have a high percentage of non-white residents and a high percentage of Cycle Atlanta users, such as 30316 (located between E5 and E6).

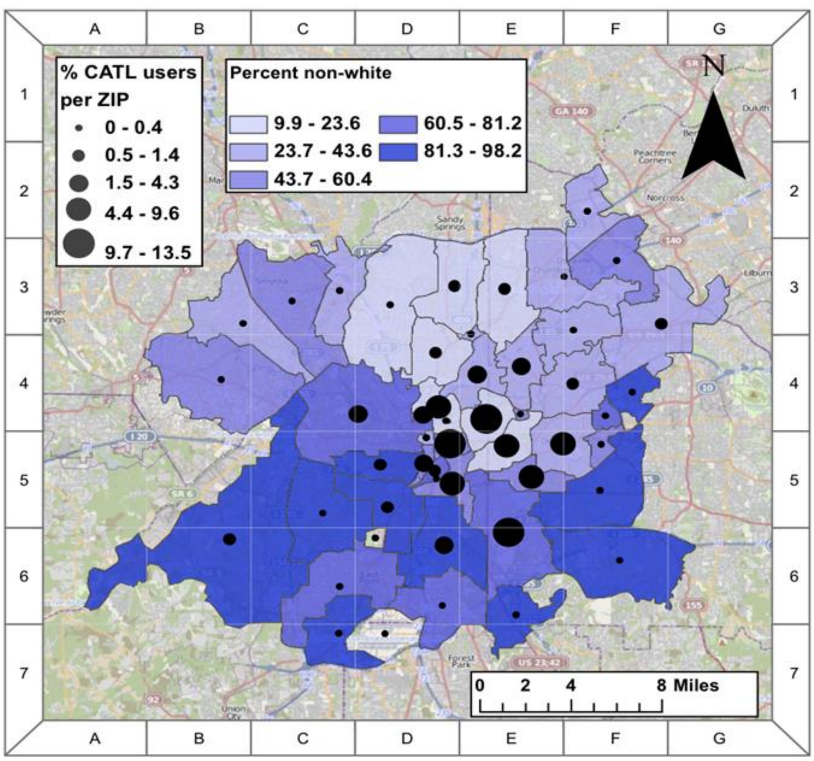


Figure 8. Cycle Atlanta Users Home Zip Code Distribution across Ethnicity Distribution in Atlanta

Figure 9 shows a comparison between the percent of Cycle Atlanta users living in a zip code and the median income of households in the zip code. The darker the zip code, the greater the median incomes of households there; the bigger the black dot over a zip code, the higher the percentage of Cycle Atlanta users living there. Although a high percentage of Cycle Atlanta users are from the high income group (greater than $100,000), that is not reflected in the geographical representation. Zip code 30327, for example, has the highest median income of any zip code (between $100,000 and $130,270, the income group that had the greatest number of Cycle Atlanta users in it). However, 30327 also has one of the lowest percentages of Cycle Atlanta users residing in it, at less than 0.353.

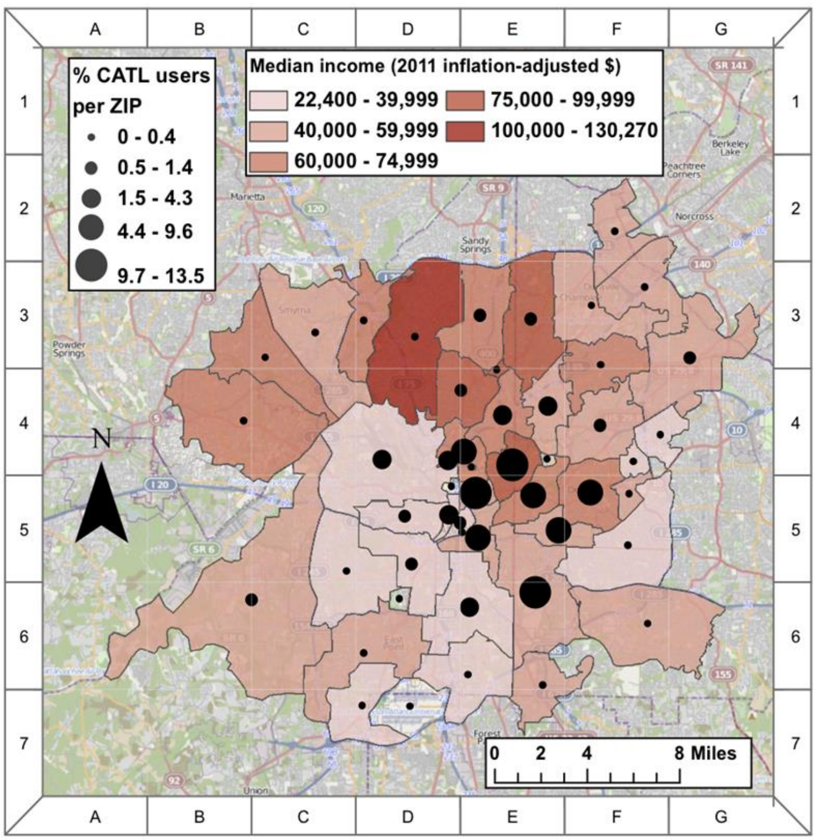


Figure 9. Cycle Atlanta Home Zip Code Distribution across Median Household Income Distribution in Atlanta

Figure 10 shows a comparison between the percent of Cycle Atlanta users living in a zip code and the median age of people living in that zip code. The darker the zip code, the greater the median age of people living there; the bigger the black dot over a zip code, the higher the percentage of Cycle Atlanta users living there. The researchers could expect zip codes with median ages between 25 and 34 to have the greatest percentage of Cycle Atlanta users residing in them, since this was the age category with the greatest percent of Cycle Atlanta users. While this is somewhat true, it appears that zip codes with median ages between 35 and 44 have greater percentages of Cycle Atlanta users living in them than zip codes with median ages between 25 and 34 (which is the age category that has the second highest percentage of Cycle Atlanta users in it).

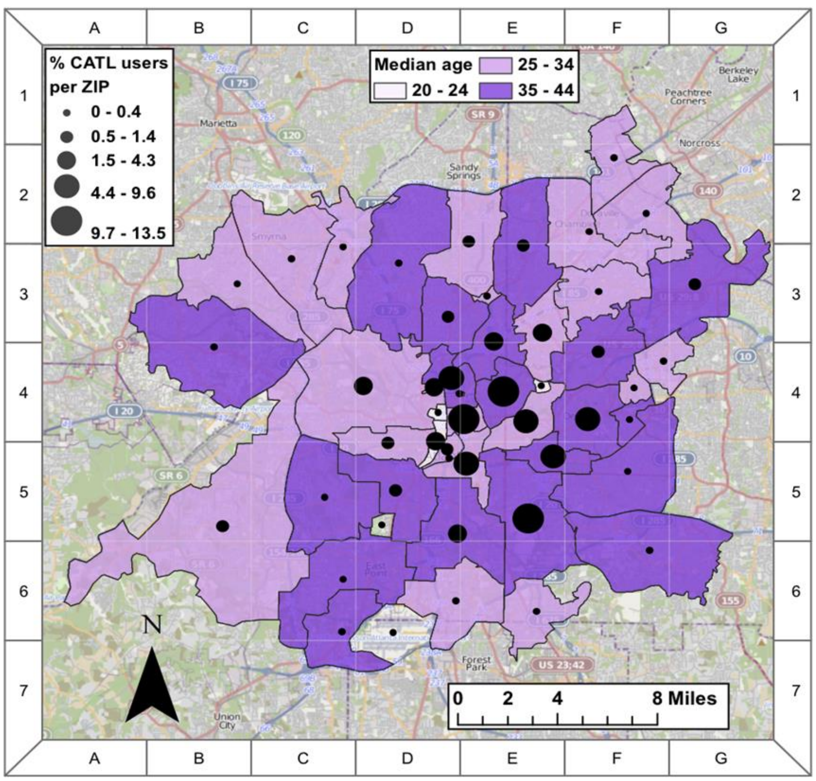


Figure 10. Cycle Atlanta Users Home Zip Code Distribution across Median Age Distribution in Atlanta

Figure 11 shows a comparison between the percent of Cycle Atlanta users living in a zip code and the population density of that zip code. The darker the zip code, the greater the population density; the bigger the black dot over a zip code, the higher the percentage of Cycle Atlanta users living there. Aside from a few outliers, this map suggests that as the population density of a zip code increases, so does the percentage of Cycle Atlanta users living in that zip code. This makes sense, since high-density urban areas are often the most bikeable. One note-worthy outlier is the zip code 30316, located between E5 and E6. This zip code contains dense areas such as East Atlanta Village and Reynoldstown in the northern part, but also less dense areas such as Gresham Park in the southern part. It is likely that if this zip code were separated into a north part and a south part, the north part would show high density as well as a high percentage of Cycle Atlanta users residing in it, and the south part would show low density and a low percentage of Cycle Atlanta users.

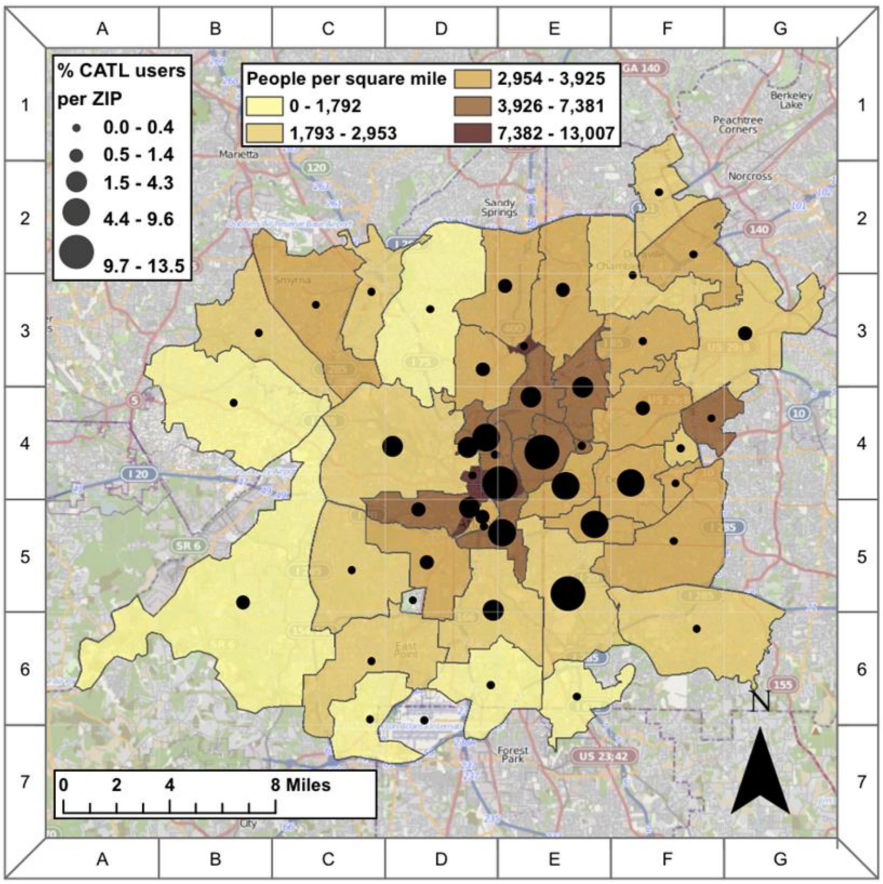


Figure 11. Cycle Atlanta Users Home Zip Code Distribution across Population Density Distribution in Atlanta

CONCLUSION

The end result of this analysis was an awareness of the differences between Cycle Atlanta users and characteristics of the population in Atlanta. Subsequent to this analysis, additional recruitment efforts were undertaken to add Cycle Atlanta users from diverse communities around Atlanta. It is of note that less data exist about the cycling population in Atlanta, therefore comparisons to the cycling population specifically could not be made.

It could be argued that zip codes are not an adequate geographic unit for analytical purposes. Their boundaries tend to be arbitrary, as do their sizes. In the study area used for this analysis, some zip codes were very large, while some were very small. Some followed neighborhood boundaries, while others crossed them. Because of these inconsistencies, it is difficult to draw conclusions from an analysis that uses zip codes as the geographic unit (although the zip codes do allow general trends to be observed). Zip codes were chosen for this analysis because that is how Cycle Atlanta users indicate their home, work, and school locations when using the app, in absence of other more refined boundaries. To protect user privacy, home and work address were not specified in the app.

CHAPTER 5. SOCIO-DEMOGRAPHIC INFLUENCE ON RIDER TYPE CLASSIFICATION AND INFRASTRUCTURE PREFERENCE

INTRODUCTION

Popular adoption of bicycling as a mode of transportation can reduce overall congestion, air pollution and energy consumption while at the same time enabling an active lifestyle and providing users with a low cost, equitable means of transportation (Sallis et al. 2004, Damant-Siriois et al. 2014). In view of all these prospective benefits, the federal government has recently reoriented its policies towards promoting biking and walking (FHWA). Additionally, several state and local transportation planning agencies have incorporated a bicycle planning module in their long term vision for their regions, including Atlanta (www.atlantaregional.com/plan2040). Despite this recent interest, research shows that although 40% of the trips made in the U.S. are less than 3 miles and may therefore be a bikeable trip due to the short distance, only 1.8% of such trips are bicycle trips (Pucher et al. 2011). This low usage of bicycling has been generally attributed to safety issues (AASHTO 2012), with major safety perception factors including high speed limits, high traffic volumes, last mile disconnect in the network, and an absence of physically separated facilities for cyclists (Dill and Carr 2003, Buehler and Pucher 2012).

Studies reveal that a substantial increase in the number of bicyclists can be achieved by providing facilities for safe riding (Pucher and Buehler 2009), and therefore it is important for planning agencies to know where cyclists prefer to bike and their desire for dedicated facilities. Cities often try to organize the route network by balancing the connectivity of the network, shortest travel distances, parking locations, and traffic volumes (Dill 2004) – but this task is often difficult as perception of safety and comfort may vary across the level of experience of the cyclists, age, gender, traffic characteristics, and other factors. This difference in perception is further complicated by a lack of data on individuals who are not cycling at present, but who may otherwise choose bicycling as a mode if proper infrastructure and environment are provided. Geller (2006) suggested that the majority of Americans belong to this group, and that large scale adoption of bicycling for transportation is dependent on making bicycling a viable and acceptable option for such riders. He hypothesized, based on his experience working as a bicycle planner with the City of Portland, that infrastructure preferences are different across the population of Portland cyclists. According to Geller (2006), these preferences are reflected in their level of comfort and willingness to bicycle given certain combinations of road characteristics and bicycle infrastructure.

Multiple studies have attempted to group the diverse range of cyclists and their perceptions into categories based on person-level attributes as well as trip-level attributes so that preferences and perceptions of cyclists and non-cyclists can be predicted even when data is sparse. In addition, there have been studies relating socio-demographic characteristics to cycling propensity, although results do not always agree. Being female appears to be the only consistent attribute which shows a negative propensity towards cycling (Krizek et al. 2005, Gerard 2006, Gerard et al. 2007, Emond et al. 2009, Akar et al. 2013, Segdahlia and Sanchez 2014b). However, even that result holds true only for cyclists in the U.S. and Australia – the proportion of cyclists that are female is higher in the Netherlands and Denmark, while female cyclists in Germany make more trips on average than their male counterparts (Gerard 2006). Studies by Gerard et al. (2007) and Krizek et al. (2005) point towards a greater safety concern and absence of cycling infrastructure as reason for this lower adoption rate of cycling among females in the U.S. and Australia. They suggest that aversion to cycling in absence of cycling infrastructure may stem from the more risk averse nature of women than men under similar situations. On the other hand, drawing parallels from health behavior research, Emond et al. (2009) suggested that there are multiple attitudinal factors that influence propensity to bicycle including self-efficacy or a person's ability to confidently engage in an activity.

In this research, we use data collected from cyclists in the Atlanta region to answer two questions: (i) whether self- efficacy, described here as the level of comfort and confidence in bicycling, varies across socio-demographic attributes of the cyclists and (ii) if infrastructure preference is influenced by the confidence and/or socio-demographic attributes of the cyclist.

Accordingly, the analysis is presented in two parts: in the first part, we use different logistic regression models to understand the dependence of self-classified rider type on socio-demographic variables and riding characteristics like riding frequency and riding history. In the second part, we use multiple stated preference survey datasets to understand if there are any preferences for infrastructure type (bike lane vs bike path) or road characteristics (slopes or traffic conditions) that vary distinctly across the rider types. The survey data were analyzed using factor analysis and regression models were constructed to understand the correlation between confidence levels and infrastructure preferences modeled as different factors.

LITERATURE REVIEW

One of the earliest studies on route and infrastructure preference of cyclists was done by Aultman-Hall et al. (1997) for the city of Guelph, Ontario, Canada. The user data relied on both a mail out community survey and a user survey distributed to cyclists at cycle shops and on the road. The study revealed that men bicycle for statistically significantly longer distances than women. Cyclists were found not to choose an off road path or trail until the quality of such road or trail was significantly better than the city streets. Collector roads were universally favored over arterials, although speed limit of the link did not seem to influence route choice. For the following characteristics, a statistically significant difference was found between the chosen route and the shortest route: (i) grade – cyclists prefer routes with less grade than a shortest route would have; (ii) traffic signals – cyclists tend to use routes with more signals than the shortest route and they particularly use traffic signals for any turning movement; (iii) number of turns – cyclists seem to prefer routes with fewer turns than the shortest route, and finally, (iv) cyclists tend to avoid links with more than two buses per hour.

Another stated preference survey study, conducted in St. Paul – Minneapolis by Krizek et al. (2006), revealed that cyclists prefer on-street bicycle lanes to off-street trails. Streets with bike lanes were also found to be preferable to streets with no on-street parking but no bicycle lane as well. Tilahun et al. (2006) conducted an adaptive stated preference survey to determine cyclists' preference between off-road facilities, designated bicycle lanes with parking and with no parking, and shared facilities with varying levels of traffic. The results indicate that higher income households have higher odds of choosing better facilities, and age and sex are not significant, although females are more likely to choose safe facilities than men.

Sener et al. (2008) used a web-based stated preference survey to collect data on cyclist route choice in Texas and used a panel mixed multinomial logit model to estimate route choice parameters. Of the chosen attributes, they found that cyclists expressed a sensitivity to travel time and preferred streets without on-street parking. Moderate hills were preferred over flat terrain and cyclist characteristics entered the model specification only as interaction terms between parking and cyclist age – no significant influence of any other cyclist characteristic (experience, for example) was noted. Titze et al. (2008) performed a study on 1000 bicyclists in Graz, Austria where the participants were asked to fill out a survey related to physical environment, home and destination environment, social environment, living quarters and attitudes. Their study reported that presence of bike lane connectivity, social support, and perceived benefit of rapidity have a positive influence on bicycling while perceived barriers of impractical mode of transport and physical discomfort act as major deterrents to bicycling.

Xing et al. (2008) conducted an online survey among residents of five cities similar to Davis, California. Using individual level factors such as socio-demographic and attitudinal attributes, social environment factors such as the perception about bicyclists and bicycling, and physical environment variables such as presence of bicycle infrastructure, they sought to understand how these factors influenced owning and using a bicycle. The study showed that attitudes such as "I like biking" are significant in influencing a decision to own or use a bicycle. Similarly, a social perception of who the cyclists really are can significantly impact the possibility of owning and using a bicycle. Emond et al. (2009) studied the influence of gender on a binary dependent variable indicating whether the participant bicycled in the last 7 days or not. They used the same ecological model and the same variables as those used by Xing et al. (2008). The results indicate that there are significant differences in attitudes and preferences between men and women cyclists. While male cyclists find living in a bicycle-friendly community to be a positive reinforcement and are more likely to bicycle if they bicycled in their youth, female riders are mainly motivated by safety factors and are much more affected if the size of the street increases or if there are no bike lanes. Comfort is the most important factor for women bicyclists and their household responsibilities possibly deter them from bicycling, as reflected in the significance of the factor 'I need a car to do many of the things that I like to do' for women cyclists. Social perception (bicyclists are rich/ bicyclists are poor) and access to safe destinations influence both genders similarly and so do socio-demographic variables like education and child assistance, both of which positively influence propensity to bicycle. However, as with Xing et al. (2008), ‘I like biking’ is the most significant positive factor for both groups of cyclists. The study by Emond et al. (2009) points to the fact that instead of experience only, gender should be an important consideration when planning for bicycle infrastructure. The present bicycle compatibility index (BCI) used by FHWA for planning purposes assumes that cyclists become confident with experience and will be comfortable bicycling under most traffic situations. Therefore, separate infrastructures are planned for inexperienced riders and mostly for recreational purposes which often fails to provide access to services. As the study by Emond et al. (2009) shows, female riders have marked preference for separate infrastructure irrespective of experience and hence, such facilities should be designed more along the primary road network. The study also attributed the gender difference in bicycling to difference in perceived safety of bicycling and the difference in comfort level across different facility types between men and women cyclists.

Winter et al. (2011) conducted a survey among 1402 current and potential bicyclists in Vancouver, Canada to understand the ‘potential motivators and deterrents of cycling’. Respondents were grouped into potential (n =197), occasional (n = 617), frequent (n = 481), and regular (n = 107) cyclists. 73 potential motivators and deterrents to cycling were identified from literature and presented in the survey questionnaire as “how would [item X] influence your decision to cycle?” The responses could be marked on a 5 category behavioral intent scale with much less likely to cycle having an influence score of – 1 and much more likely to cycle having an influence score of +1 and intermediate categories being marked at increments of 0.5. From the mean response score, the top 3 motivating items were “the route is away from traffic noise & air pollution”, “the route has beautiful scenery” and “the route has bicycle paths separated from traffic for the entire distance" while the top 3 deterring items were "I need to carry bulky or heavy items", "the route has surfaces that can be slick when wet or icy when cold" and "the route is not well lit after dark". Safety factors, which included items like 'risk from motorists who don't know how to drive safely near bicycles' and 'risk of injury from car-bike collisions', had the highest factor scores as a deterrent to cycling has the highest negative mean factor score on influence on likelihood of cycling, followed by poor weather and darkness, interactions with motor vehicles and route surfaces. Mean scores and ranks were similar across the different groups. The factors conducive to cycling were ease of cycling, integration with transit, bike parking, and end-of-trip facilities. Lane marking and signage also scored substantially high as factors conducive to cycling. The same participants were provided with pictures of 16 different routes and asked to indicate which routes they would choose to cycle (Winter and Teschke 2010). Route preferences were found to be similar across all the cyclist types (potential, occasional, frequent and regular). Between 70% and 85% of the participants chose off-street paths, 70% of the participants would bike on physically separated routes next to major streets and 50-65% would bike on residential routes. Routes with bike lanes, paved surfaces, no on-street parking and traffic calming increased the likelihood of choosing the route from 12% to 37%.

Yang and Mesbah (2013) conducted a survey on a small sample at University of Queensland, Brisbane with 19 respondents, majority of whom were high school students and postgraduates. They found that distance, travel time, traffic safety and gradient are the most important factors in route choice of bicyclists. Akar et al. (2013) sent out an online survey for the student, staff and faculty of Ohio State University (OSU) through the OSU Transportation and Parking Service's (T&P) webpage. From about 2000 respondents who provided nearly complete data, results indicate that female respondents are more likely to overestimate their commute distances, do not feel safe in vehicular traffic and are more deterred from cycling by the absence of bike lanes/paths/trails. Female participants also cited the need to change clothes and carry things as major reasons for not bicycling. A significantly lower percentage of female cyclists considered themselves advanced cyclists (9%) as compared to male cyclists (35%). Female cyclists are also significantly more likely to feel biking and walking on campus after dark unsafe – about 30% of female participants agreed that they feel safe biking and walking on campus after dark as compared to about 70% of male participants. Mode choice models indicate that being a female makes it less likely to be a cyclist, while having bike lanes and trails and feeling safe positively impact the possibility of choosing cycling for the commute.

Segadilha and Sanchez (2014a) conducted a survey on 49 frequent cyclists in the city of Sao Carlos, Brazil to identify the relative importance of multiple factors identified from the literature. Slope appeared to have the least importance in choice of route, while number of trucks, number of buses, traffic volume, and traffic speed score the highest in influencing route choice decisions. The authors also reported that stratification by age, gender, and cycling frequency showed that preferences may vary significantly across these categories. Sousa et al. (2014) conducted an email survey involving 380 students from 3 different cities in Brazil. The survey provided the participants with six statements on perceived barriers to cycling and asked them to rate how positively or negatively these barriers influence their decision to bicycle. The perceived barriers included absence of adequate infrastructure, traffic safety, distance, physical ability and experience, slopes, and climate. Lack of adequate infrastructure, lack of safety, and slopes were found to be the most important barriers to cycling, while climate was the least important. Wang et al. (2014) conducted a survey at the University of Auckland to understand the factors that influence a cyclist's decision to bicycle and his/her route choice. The study concluded that safety, low traffic volume and speed, separation from cars and pedestrians, and separate facilities are important factors in promoting cycling, along with connectivity and ability to carry the bicycles on public transport. Wang et al. (2014) also concluded that safety is a more important factor for female than for male cyclists.

Additionally, Barros et al. (2015) designed an online survey to understand the different factors that affected mode and route choice. Their findings suggest that presence of cycle lanes and bicycle parking encourage people to choose bicycling. Mertens et al. (2015) conducted a web-based survey with 389 respondents where participants were presented with photographs of two alternative cycling routes. The study was designed to understand how macro-environmental factors like residential density interacted with micro-environmental factors like speed limit and presence of bicycle facilities to affect participants' decision to bicycle. Mertens et al. (2015) found that while participants preferred low residential density over medium or high residential density, preference for a low speed limit and physically separated bicycle facility does not vary across choice of residential density.

On another subject, several studies have developed classification systems for cyclists based on person-level attributes as well as trip-level attributes. Damant-Siriois et al. (2014) and Dill and Voros (2008) provide comprehensive accounts of different cyclist type classifications based on both person- and trip-level attributes. Person-level attributes used in classifications include the attitude and comfort level of cyclists (Geller 2006), the behavioral perspective and value system of the individual (Paulssen et al. 2011) and a cyclist's preference for infrastructure (Larsen and Geneidy 2011). Trip-level attributes include trip purpose (Kroesen and Handy 2013) and whether trips depended on weather conditions (Bergstrom and Magnusson 2003). However, Geller's classification of cyclists into four different categories of *strong and fearless*, *enthused and confident*, *interested but concerned*, and *no way no how*, based on their comfort level in cycling, gained notable popularity and was used in planning for cycling infrastructure by multiple regional planning agencies in the last decade (Dill and McNeil 2012; Geller 2006). It should be noted though that this classification was devised on an ad-hoc basis and was not based on any survey or self-description of cyclists (Geller 2006). Subsequently, in a recent study, Dill and McNeil (2012) conducted a random phone survey of 908 adults in Portland, OR asking respondents about their comfort level in bicycling on non-residential streets, with and without bike lanes, to set up a basis for the categorization. This comfort level question was combined with an interest question which asked if respondents wanted to bicycle more than they are currently doing. The answers to these two questions were considered together to categorize riders into the classifications suggested by Geller (2006). The study categorized the riders who were comfortable bicycling on non-residential streets even without bike lanes as *strong and fearless* irrespective of their response to the interest question. Cyclists who were comfortable on non-residential streets only with bike lanes were classified as *enthused and confident* riders. The cyclists who were not comfortable on any facilities and/or have not bicycled for transportation for the last 30 days were categorized as *no way no how* while the *interested but concerned* group had cyclists uncomfortable on residential streets irrespective of their interest in bicycling.

NEW CYCLIST CATEGORIES

According to Geller (2006), the *strong and fearless* riders are cyclists who would bicycle irrespective of road and traffic conditions and whether separate facilities are present or not; the *enthused and confident* riders are cyclists who will choose cycling as a mode of transport even if bare-bones facilities are present or if it is not infeasible for them because of distance or road features; *interested but concerned* is the group of individuals who are currently very infrequent bicyclists or do not bicycle at all and will bicycle only when they have protected and separate facilities for the purpose i.e., these are the riders who are willing to bicycle if proper infrastructure is provided. The fourth category of riders, *no way no how*, includes the individuals who will not bicycle under any circumstances.

There are multiple ways in which this classification system can be improved. First, within a wide spectrum of cyclist 'types', the classification misses those people who are enthusiastic bicyclists but are not willing to bicycle with bare-bones cycling infrastructure. Most often, their concerns are more related to safety than confidence (for example, riding together with children). These cyclists possibly bike commute every day but using different routes than the enthused and confident group, and often undertake longer detours to find safer routes. On the other hand, while they prefer separate cycling infrastructure, these are the people who can also be motivated by traffic calming measures and do not require a physical separation from the traffic to be able to bicycle. Therefore, this group of cyclists belongs neither to the *enthused and confident* group, nor to the *interested but concerned* group as proposed by Geller.

Geller's classification also misses the captive riders of the system – the people who cycle because of a lack of alternatives. While these cyclists may make frequent and regular trips, they are not bicycle enthusiasts and generally associate a negative social image with cycling. It is much more difficult to retain such users in a cycling system without changes in the social perception that cyclists are either poor or rich and that cycling is not the natural normal mode of transportation (Xing et al. 2007, Emond et al. 2009). Efforts in that direction will involve not only building infrastructure, but also creating awareness and education, thus requiring a different policy approach than that for other groups.

Our research primarily focused on collecting data from cyclists via a GPS enabled smartphone application. We assumed that non-cyclists will not use the application and hence, the *no way no how* category was not applicable. Therefore, for our research, we modified the existing rider type classification and added a new group called *comfortable but cautious* to the existing system. This new category was designed to adequately represent the riders who differ in their view of safety from more aggressive riders but, at the same time, are similarly enthusiastic about bicycling. We also assumed that most captive riders will not be motivated enough to provide personal and cycling data voluntarily and therefore did not create a separate group for them. However, future work should consider their preferences as well. In its final form, the rider types suggested in this research consisted of four different groups:

(i) strong and fearless;

(ii) enthused and confident;

(iii*) comfortable, but cautious;*

(iv) interested but concerned;

While the other groups are expected to show similar attitudinal preference as the Portland study (Geller 2006, Dill 2012), the *comfortable and cautious* group of riders is hypothesized to include a greater proportion of female cyclists and/or individuals in higher age groups who are bicycle enthusiasts, but are less risk-taking in attitude and hence may appear to be less confident.

METHODOLOGY

The primary purpose of this study is to develop a model that can help us to relate readily available socio-demographic data to cyclists' stated preferences for infrastructure. Toward that end, we used multiple data sources and models to find the combinations of attributes that can best predict the infrastructure preferences of cyclists. First, we used socio-demographic data collected from riders who recorded their trips on a smartphone app called Cycle Atlanta. We hypothesized that the rider type classification can serve as a proxy for how cyclists of different ages, genders, incomes, and ethnicities perceive risk and comfort on the streets of Atlanta.

We then used stated preference survey data from two online surveys administered separately by two different groups and at a time gap of six months. The first survey was conducted by the Atlanta Regional Commission across the region, and the second was conducted by our research group and geared to the users of the Cycle Atlanta smartphone application. The survey questions related to socio-demographic information and infrastructure preferences of the participants were carefully designed to ensure that they had identical wording and choice order in both the surveys. The data on infrastructure preferences were then analyzed using factor analysis to group similar, correlated preferences under one factor. The factor scores were then regressed against socio-demographic variables to understand how they influence a participant's infrastructure preferences.

PART 1. Predicting Rider Type based on Socio-demographics and Riding Behavior

In the first part of the analysis, we directed our efforts towards identifying the relationship between stated rider type and other socio-demographic variables of participants.

Data Source: Cycle Atlanta

The first analysis uses the data collected through the Cycle Atlanta smartphone application, developed through a collaboration between the Georgia Institute of Technology and the City of Atlanta's planning office to promote cycling in Atlanta (The City of Atlanta, 2011).

The application was named Cycle Atlanta after the larger planning project for which the application was initiated, and was developed by an interdisciplinary team of researchers. The application was originally based on San Francisco's CycleTracks (Hood et al. 2011), although Cycle Atlanta was substantially updated to make better use of current features available in iOS and Android as well as to include features that the City and local bicycle advocacy groups wanted in the application. The basic feature is trip recording, where the application uses the GPS of the phone to record the location of the user once per second. In addition to tracking cyclists' trips, the app also provides options to enter personal information, including age, email address, gender, ethnicity, home income, zip codes (home, work, and school), cycle frequency, rider type, and rider history (Misra et al. 2014).

The breakdown of age, gender, income, and ethnicity was kept similar to the breakdown as found in the household travel survey. The age and income intervals as well as the gender and ethnicity subcategories were adopted from the household travel survey conducted by Atlanta Regional Commission (www.atlantaregional.com/transportation/travel-demand-model/household-travel-survey). The rider type and rider history categories are exclusive and unique to the design of Cycle Atlanta. The cycling experience field allowed users to specify how long they have been cycling and can choose from the categories 'since childhood', 'several years', 'one year or less' and 'just trying it/just started'.

As of June 2014, the Cycle Atlanta dataset consisted of 1529 unique users who could provide information on their age, gender, ethnicity, income, rider history and cycling frequency. Because there were only 6 cases in the age group of 65+, that group was merged with the age group of 55-64 years old and the new group is referred to as “age 55+” for the rest of the analysis. About 60% of the riders provided information on each of the socio-demographic categories. The users of Cycle Atlanta are predominantly male (about 75%), white (about 80%) and mostly from a high income group (>$75,000) (about 45%). Table 1 presents the basic statistics of the different socio-demographic variables considered in this study. The median age of the users is between 25-34 years, while the median income is between $60,000 and $74,999. The median rider type is an *enthused and confident* rider with median cycling frequency of several times per week and a median riding history of several years.

Analysis and Results

The goal of this part of the study was to understand the relationship between cyclist self-classification into different rider types and the socio-demographic make-up and riding pattern of the cyclists.

Multivariate Analysis

Except for ethnicity and gender, the socio-demographic variables considered in this study have an underlying order, although they are categorical. This led us to use methods and analyses relevant to ordinal variables instead of nominal variables. To understand degree of association between variables, polychoric correlation was used, which assumes an underlying continuous bi-variate normal distribution for discrete categorical variables with an ordinal scale. Figure 12 shows the correlation coefficients obtained from the analysis. Age and income are correlated with a measure of correlation in the range of 0.5. Rider type is correlated with gender, cycling frequency, and rider history, each with a correlation ~ 0.35.

Figure 13 shows the percentage of rider types across the different socio-economic variables as well as rider history and frequency. As hypothesized, higher proportions of *strong and fearless* and *enthused and confident* riders are in the age groups below 35 years and are male. They are also disproportionately present in high income groups, indicating that people in those income groups are possibly more confident and aggressive than those in other income groups. Cyclists with a history of less than a year are more represented in the *comfortable and cautious* and the *interested but concerned* groups than any other group. Among infrequent cyclists, a high proportion of people consider themselves to be either *comfortable but cautious* or *interested but concerned* riders.

Table 1. Basic Statistics for Socio-demographic Variables

Socio-demographic Variables(n = 1529)		
Age(n = 1001)	Count	Percentage
Less than 18	6	0.6
18-24	110	10.99
25-34	448	44.76
35-44	218	21.78
45-54	144	14.39
55-64	66	6.59
65+	9	0.9
Gender(n = 981)		
Female	240	24.46
Male	741	75.54
Income(n = 776)		
Less than $20,000	78	10.05
$20,000 to $39,999	133	17.14
$40,000 to $59,999	111	14.3
$60,000 to $74,999	95	12.24
$75,000 to $99,999	112	14.43
$100,000 or greater	247	31.83
Ethnicity(n = 955)		
African American	46	4.82
Asian	43	4.5
Hispanic / Mexican / Latino	53	5.55
Multi-racial	22	2.3
Native American	3	0.31
Pacific Islander	2	0.21
White	770	80.63
Other	16	1.68
Cycling Frequency(n = 546)		
Daily	158	28.94
Several times per week	260	47.62
Several times per month	113	20.7
Less than once a month	15	2.75
Rider Type(n = 989)		
Strong & fearless	187	18.91
Enthused & confident	443	44.79
Comfortable, but cautious	333	33.67
Interested, but concerned	26	2.63
Rider History(n = 985)		
Just trying it out / just started	59	5.99
One year or less	120	12.18
Several years	330	33.5
Since childhood	476	48.32

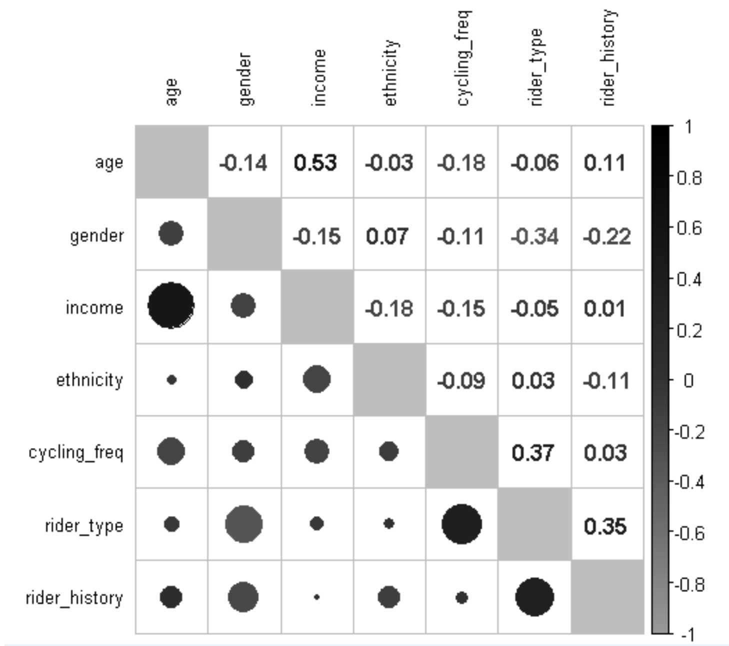
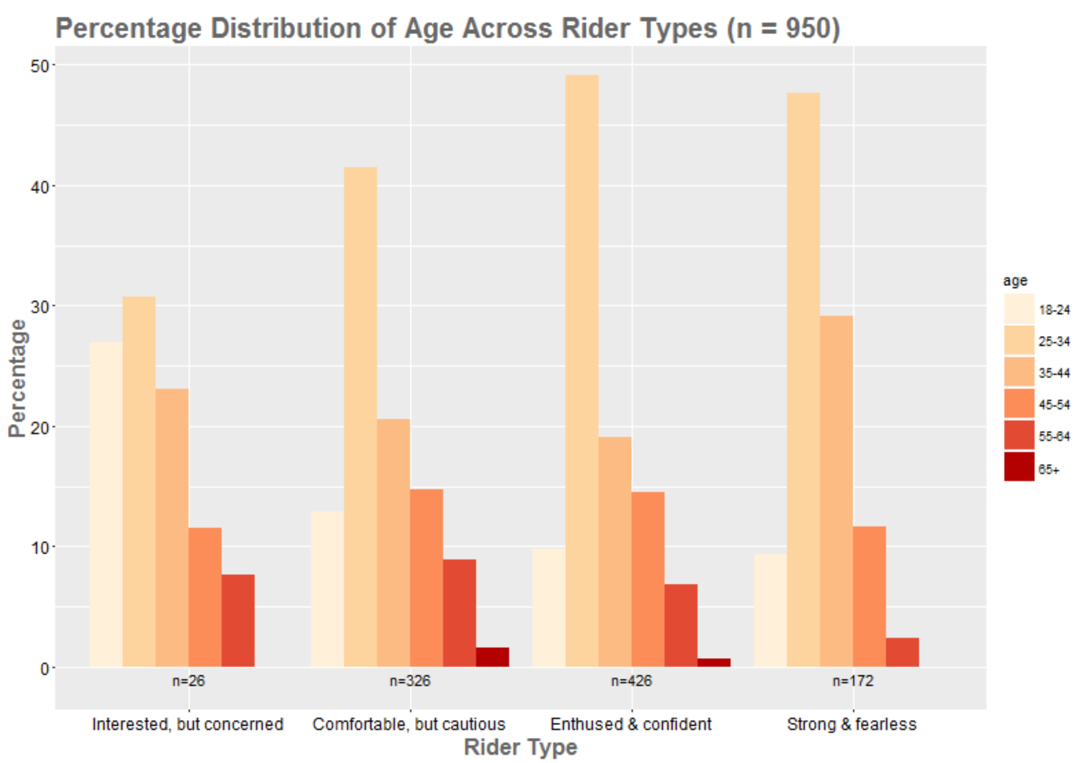
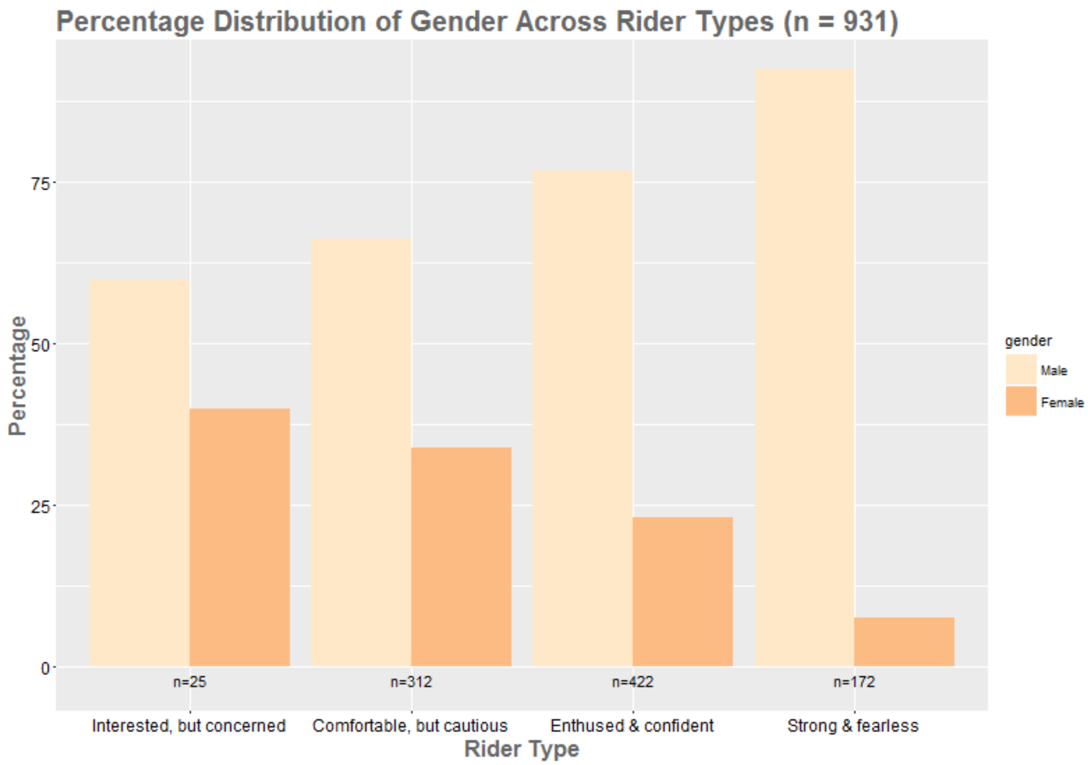
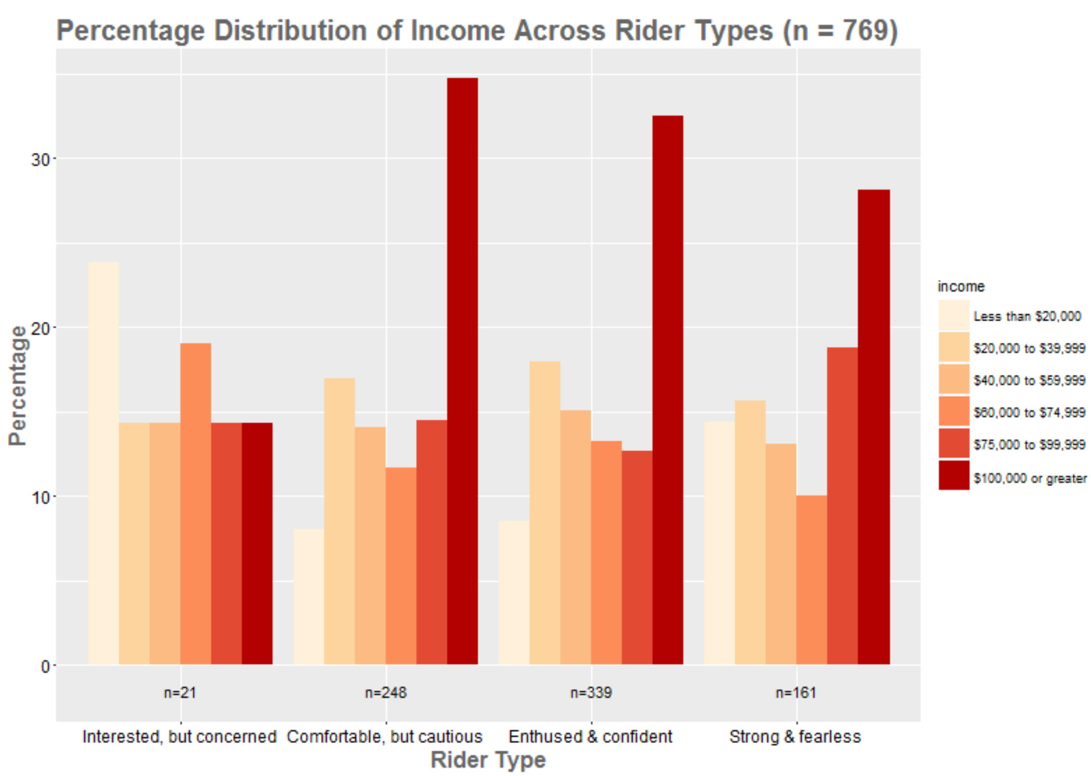
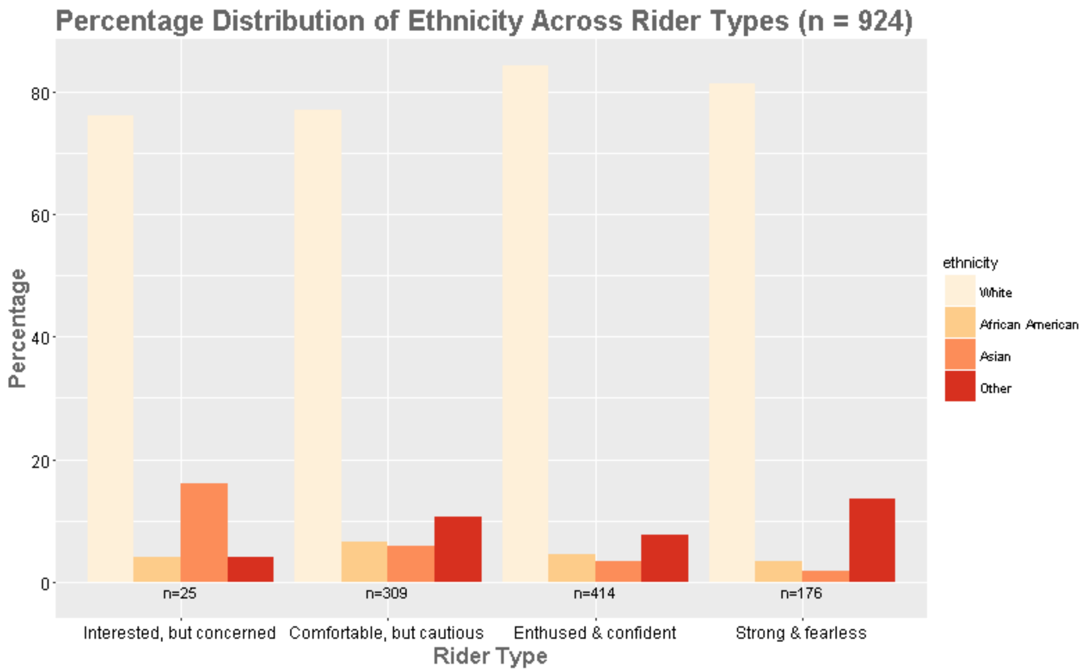


Figure 12. Measure of Association between Variables

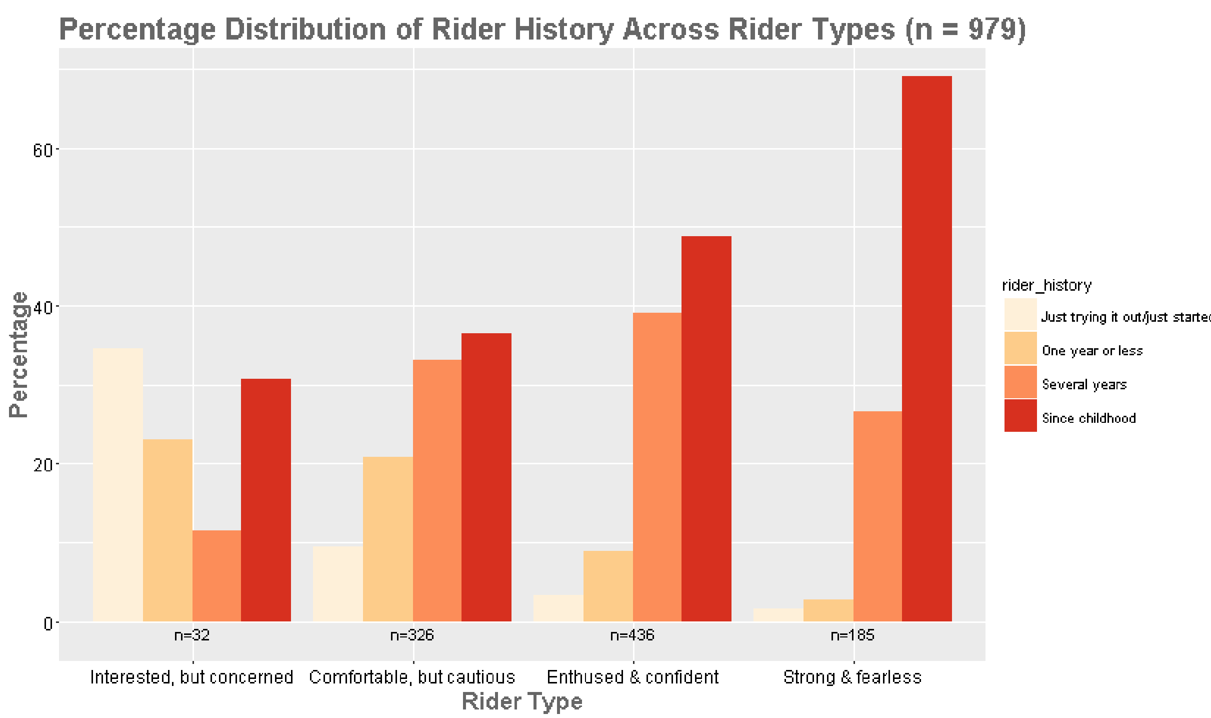



Percentage Distribution of Gender Across Rider Types (n = 931)
Percentage
0
25
50
75
n=25
n=312
n=422
n=172
Interested, but concerned
Comfortable, but cautious
Enthused & confident
Strong & fearless
Rider Type
gender
Male
Female


Percentage Distribution of Income Across Rider Types (n = 769)
Percentage
0
10
20
30
n=21
n=248
n=339
n=161
Interested, but concerned
Comfortable, but cautious
Enthused & confident
Strong & fearless
Rider Type
income
Less than $20,000
$20,000 to $39,999
$40,000 to $59,999
$60,000 to $74,999
$75,000 to $99,999
$100,000 or greater



Percentage Distribution of Ethnicity Across Rider Types (n = 924)
Percentage
80
60
40
20
0
n=25
n=309
n=414
n=176
Interested, but concerned
Comfortable, but cautious
Enthused & confident
Strong & fearless
Rider Type
ethnicity
White
African American
Asian
Other



Percentage Distribution of Rider History Across Rider Types (n = 979)
Percentage
60
40
20
0
n=32
n=326
n=436
n=185
Interested, but concerned
Comfortable, but cautious
Enthused & confident
Strong & fearless
Rider Type
rider_history
Just trying it out/just started
One year or less
Several years
Since childhood

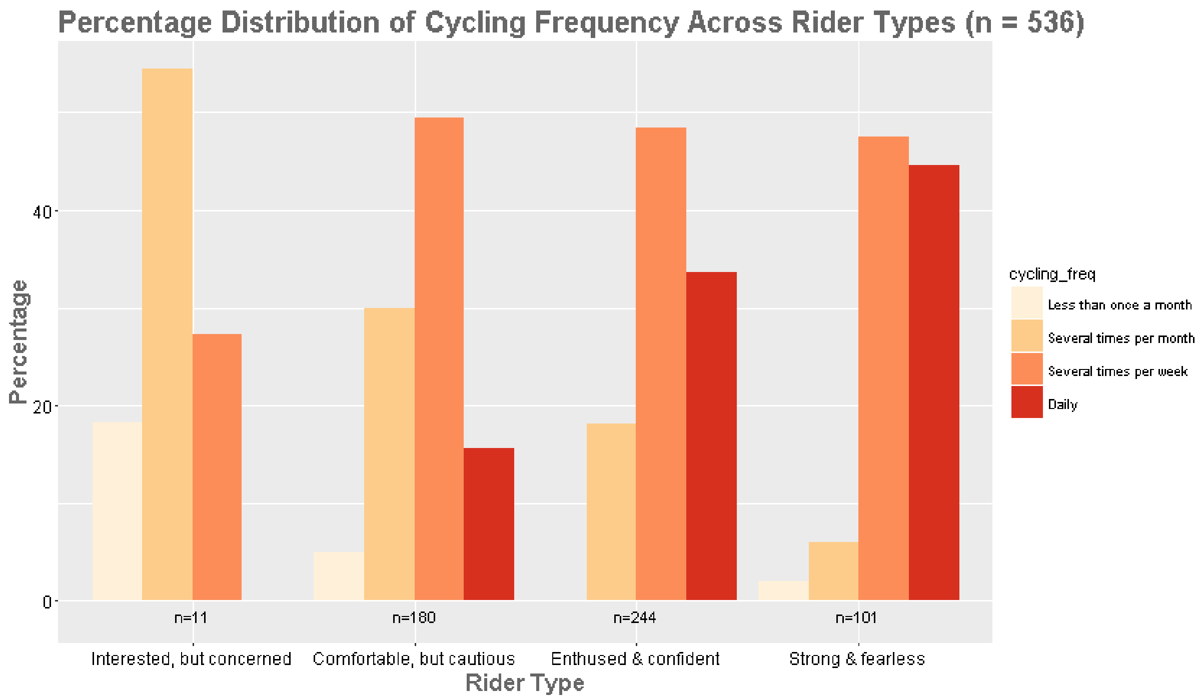


Figure 13. Socio-demographic and Riding Pattern Distribution of Cyclists across Rider Types

Logistic Regression Models

Two main types of variables were used in these models – the socio-demographics and the riding habit/pattern of the participant. The socio-demographic variables included age, gender, income, and ethnicity while the riding pattern variables included cycling frequency and rider history. From the distribution of age and gender across rider type, it was evident that there were very few participants in the age group above 45. So the age groups 45-54, 55-64 and 65+ were grouped into one category of 45+. The riding pattern was found to be distinctly similar across the age group of 25-34 years and 35-44 years and hence, these two groups were also merged to form a new group of 25-44 years. Similarly, different income categories were consolidated into 3 categories, and different ethnicity types were consolidated into 4 categories. For rider history, the ‘just started’ category was merged with the ‘less than a week’ category, resulting in 3 categories instead of 4.

Of the total 989 users who provided data on rider type, only 26 users classified themselves as *interested but concerned*. Cross tabulation of rider type across other variables showed *interested but concerned* riders having zero cell values with cycling frequency ‘less than once a month’ and small valued cells for age group 45+ (2 users) and ethnicity ‘African American’(1 user) and ‘Other’(1 user) thereby presenting a problem of quasi separation. Within cycling frequency also, there are only 13 users who have cycling frequency less than once per month and none of them are *enthused and confident* riders (0 users) which again presented the issue of separation. Quasi/complete separation implies a perfect prediction scenario where the dependent variable Y can be completely predicted by variable X when the separation is complete. In case of quasi complete separation, perfect prediction happens only for a subset of observations (Albert and Anderson 1984). For example, in this dataset, it can be predicted with absolute certainty that none of the riders who bicycle less than once per month will classify themselves as *enthused and confident*, although the same cannot be said about whether riders with cycling frequency less than once a month will classify themselves as *strong and fearless* or *comfortable but cautious*. Models estimated under quasi/complete separation are more likely to either not converge or give high co-efficient estimates and infinite standard error as the log-likelihood will be presumably flat (Zorn 2005). The most common way of dealing with quasi separation is to remove the problematic covariate which again might give specification bias if the covariate is strongly correlated. We ran models both by removing observations and by aggregating the sparsely populated group with its nearest neighbor. In case of cycling frequency, the last group, cycling frequency less than once per month was merged with the group which bicycles a few times per month, and the new group was named cycling frequency once or less per week. Models ran by removing the observations with cycling frequency less than once per month gave a much lower model fit than the aggregated models and hence, in this paper, models with aggregated data are presented. Similarly, for addressing the quasi separation problem related to rider type, two alternative model sets were designed – one where the *interested but concerned* group (26 users) was merged with its next higher group *comfortable but cautious* (333 users) and another where the *interested but concerned* users were removed from the sample space and models were estimated for the remaining three categories. The model estimates in either case were not significantly different and in keeping with our aggregation theme, in this report, the aggregated models are presented.

Based on these rider type distributions, logistic regression models were estimated for each rider type to understand how the self-described confidence level is affected by socio-economic variables as well as riding patterns of the cyclists. Several logistic regression models were explored to find the best way to represent the pertinent relationships. Since cycling frequency and rider type may have bi-directional causality, they were tested for explanatory power and likely association. A single variable ordinal model for rider type with cycling frequency as the explanatory variable gives a McFadden's ρ^2 of 0.48, but an ordinal model for cycling frequency with rider type as the explanatory variable gives a McFadden's ρ^2 of 0.07 (both unadjusted for sample size difference). Although it was found that cycling frequency has a greater explanatory power for rider type than rider type has for cycling frequency, in view of the simultaneity issue, models with cycling frequency and models without cycling frequency are both presented here.

Since the discrete observed rider type categories (*y*) were originally thought of as representing a latent continuous scale of confidence and comfort (*y**), two variations of the user's underlying decision process along that one dimensional scale were initially estimated. The first is where the self-classification process was thought of as representing a binary choice for each rider type (for example, "Am I *strong or fearless* or not?"). This process was estimated using binary logistic regression models where the rider classifies himself/herself into a category ($y = 1$) if he/she perceives himself/herself above a certain confidence level threshold ($y^* > \tau$); if the perceived confidence level is at or below the threshold ($y^* \leq \tau$), the rider does not choose that rider type category ($y = 0$). Four different binary logistic models were estimated – one for each rider type. For each of these four choices, several models were run with different variable combinations to balance model fit and parsimony. Age group 45+, gender male, income less than $40,000, rider history since childhood, and cycling frequency of daily were chosen to be the base categories for age, gender, income, rider history and cycling frequency variables respectively. Ethnicity was not included in the models due to its heavy bias towards white riders. Model fit statistics were calculated based off the corresponding equally likely model statistics (Mokhtarian 2016). In addition, even when not significant, variables with *t-statistic* >1 were kept in the models.

The first models were run with age and gender as explanatory variables which gave model fits in the range of 0.2 - 0.3 (with base equally likely). Age group 25-44 and gender were significant for *strong and fearless group* and for the group including *comfortable but cautious and interested but concerned*. At the second stage, income was added to age and gender. While income itself was not significant, McFadden's ρ^2 for these models ranged between 0.3 and 0.45 although the sample size reduced to 932 from 742. Walden's t-test did not show significance of the income variable (p = 0.94, 0.32). Since the correlation between age and income was earlier found to be high (0.53), at the next step, an interaction term between age and income was introduced in the model. However, the model fit was not found to be significantly different from the previous model. In addition, introduction of interaction term led to perverse signs for the income variable. Therefore, age and income were included in the model as separate variables. Since models with age and income gave a better fit, we tested these models for multi-collinearity effect. The VIF (Variation Inflation Factor) test was performed on a linear version of the models, and the VIF was found to be less than 5 for all variables including income.

Rider history was added to the model at the next step and was found to be significant across all the models. Wald's test as well shows that rider history is a significant variable (p = 3.2 e-09) for the model. At this stage, the ρ^2 values for the models range between 0.4 and 0.5, and both age groups and gender are significant across the *strong and fearless* and the *comfortable but cautious* and *interested but concerned* group. Rider history is the only significant variable for the *enthused and confident* group at this stage. Cycling frequency was added at the last step of model building and was found to be significant by Wald's test (p = 0.013). McFadden's ρ^2 values for the models with cycling frequency are ~0.7 (with base equally likely). Since the model fits were quite high, it was hypothesized that cycling frequency determines, to a large extent, the propensity of a cyclist to self-classify himself/herself into a particular category. However, at this stage, model sample sizes were ~ 33% of the original sample sizes mainly because of missing data on income and cycling frequency. Since income was insignificant in all models, a final model was designed by removing income but leaving in cycling frequency which brought back the sample size ~ 50% of the original. The ρ^2 for this model was found to be slightly lower than the earlier model, but in absence of income, age group

25-44 was found to gain significance. Age, gender, rider history, and cycling frequency were found to have significant influence on whether a cyclist classifies himself or herself into the categories of *strong and fearless* as well as *comfortable but cautious* and *interested but concerned*. The only significant predictor for the *enthused and confident* group was found to be cycling frequency and therefore, a model with only rider history and cycling frequency was built for this group and the ρ^2 was found to be ~ 0.6. A model with cycling frequency as the only exogenous variable was found to provide a ρ^2 of 0.48 indicating that the propensity of cyclist classifying himself/herself into the *enthused and confident* category is well specified by his/her cycling frequency alone. It may therefore be suggested that cyclists who self-classify themselves into this category mostly do so because of their riding frequency rather than their self- perception on a confidence scale. As mentioned earlier, for all the categories, two final models are presented: one without cycling frequency and one with cycling frequency. Table 2a presents the model results for binary logistic models.

The second variation on user's decision process was modeled using ordinal logistic models where the riders are thought of as classifying themselves into different categories *(y)* based on ordered partition of a latent continuous one dimensional confidence scale*(y^*)* (*$y = k$, if $\tau_{k-1} < y^* \leq \tau_k$ where k = rider type categories in an ordered scale of 1 through 4, with 1 being least confident and 4 being most confident*). The model building exercise was the same as that for binary models and the results for the ordinal models are presented in Table 2b.

Both the binary and ordinal logistic models are parsimonious and efficient as the choice is modeled on a single dimensional latent continuous variable. However, as mentioned by Bhat and Pulugurta (1997), it might be oversimplification of the actual decision process where the user is actually choosing among many alternatives the one alternative that he/she feels best satisfied with. In this case, the user has a *k*-dimensional choice space where *k* represents the number of choices faced by the user and estimating an unordered response using an ordered response model can lead to biases in estimating probability of the choices (Bhat and Pulugurta 1997, Amemiya 1985). Therefore, the next set of models estimated were multinomial logistic regressions where the user was thought of as having to choose between the four rider type categories simultaneously ("Am I *strong and fearless* or *enthused and confident* or *comfortable but concerned*, etc."). The same model building exercise was followed in this case as with the binary logit models with the *comfortable but cautious* category treated as the base category. The first model included only age and gender and gave a McFadden's ρ^2 of 0.15. The final model, without income, included age group 18-24 and 25-44, gender, rider history, and cycling frequency and gave a McFadden's ρ^2 of 0.6. The model with income and cycling frequency gave a model fit of 0.7 (unadjusted for model sample size). Age group 25-44 was found to be significant for the enthused and confident group when base group was changed to age group 18-24 indicating that cyclists in the age group of 25-44 behave significantly different in self-classifying themselves into enthused and confident group as compared to the age group 18-24. Chi-squared tests for model comparisons could not be performed due to unequal sample sizes. Models with cycling frequency gave a higher McFadden's ρ^2 than the models without cycling frequency but were estimated on a much smaller sample size, potentially removing a considerable amount of variation present in the dataset that was used for estimating the other models. Therefore, it cannot be definitively concluded that the models with cycling frequency are better models than their counterparts and hence, both types of models are presented in this report. The multinomial logistic (MNL) models are presented in Table 2c. Table 3 presents the odds ratio for the multinomial and the ordinal models both with and without cycling frequency.

Table 2(a). Binary Logistic Regression Models

Co-efficients	Strong and Fearless Estimates *(t-stat)*		Enthused and Confident Estimates *(t-stat)*		Comfortable but Cautious & Interested, but concerned Estimates *(t-stat)*	
	Model 1 N= 740	*Model 2 N= 496*	*Model 1 N= 740*	*Model 2 N= 499*	*Model 1 N= 740*	*Model 2 N= 496*
	0.239* **	**0.329 *** **	**0.468 *** **	**0.531 *** **	**0.293 *** **	**0.167 ** **
Intercept	*(5.832)*	*(6.846)*	*(8.929)*	*(11.406)*	*(6.077)*	*(2.832)*
Age			**Base: Age 45+**			
18-24	**0.102 .**	0.0267	-0.082		-0.02	-0.05
	(1.763)	*(0.429)*	*(-1.112)*		*(-0.286)*	*(-0.654)*
25-44	**0.11** **	**0.066 .**	-0.002		**-0.108 ***	**-0.1 ***
	(2.986)	*(1.665)*	*(-0.05)*		*(-2.475)*	(-1.978)
Gender			**Base: Male**			
Female	**-0.155*** **	**-0.168 *** **	0.01		**0.145 *** **	**0.158 *** **
	(-4.618)	*(-4.452)*	*(0.243)*		*(3.467)*	*(3.428)*
Income			**Base: Income < $75,000**			
Income>= $75,000	0.007		-0.045		0.037	
	(0.248)		*(-1.142)*		*(1.03)*	
Rider history			**Base: Since Childhood**			
One year or less	**-0.235*** **	**-0.169 *** **	**-0.14 ** **	-0.097	**0.375 *** **	**0.269 *** **
	(-5.779)	*(-3.644)*	*(-2.685)*	(-1.573)	*(7.81)*	*(4.733)*
Several years	**-0.143*** **	**-0.135 *** **	**0.081 ***	0.063	**0.063 .**	*0.069*
	(-4.512)	*(-3.62)*	*(1.982)*	*(1.258)*	*(1.671)*	*(1.503)*
Cycling Frequency			**Base: Daily**			
Several times/week		**-0.071 .**		-0.08		**0.144 ** **
		(-1.792)		*(-1.517)*		*(3.0)*
Once or less/week		**-0.181 *** **		**-0.183 ** **		**0.357 *** **
		(-3.88)		*(-2.965)*		*(6.252)*
Model Statistics						
Market Share of Group in the Model Dataset	*150 (20.27%)*	*90 (18.15%)*	*328 (44.32%)*	*226 (45.29%)*	*262 (35.44%)*	*180 (36.29%)*
Market Share of Other Groups in the Model Dataset	*590*	*406*	*412*	*273*	*478*	*316*
McFadden's ρ2 (Full model, base EL)	*0.368*	*0.630*	*0.460*	*0.636*	*0.473*	*0.660*
McFadden's ρ2 (MS model, base EL)	*0.177*	*0.177*	*0.292*	*0.292*	*0.251*	*0.251*
LL(0)	*-536.232*	*-536.232*	*-967.031*	*-967.031*	*-876.397*	*-876.397*
LL(MS)	*-441.357*	*-441.357*	*-684.42*	*-684.42*	*-656.533*	*-656.533*
LL(Full Model)	*-339.0786*	*-198.181*	*-522.546*	*-352.327*	*-462.058*	*-298.22*
G2=-[2(LL(Null)-LL(Full Model))]	*394.3068*	*676.102*	*888.97*	*1229.408*	*828.678*	*1156.354*

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Table 2(b). Ordinal Logistic Regression Models

Co-efficients	Estimates (t-stat) Model 1 N= 740	Model 2 N= 496
Intercepts	**Base: Comfortable,but cautious & Interested, but concerned**	
Comfortable, but cautious & Interested, but concerned| Enthused and confident	**-0.938** ***	**-1.755** ***
	(-4.53)	*(-6.357)*
Enthused and confident|Strong and Fearless	**1.26** ***	**0.692** **
	(6.005)	*(2.62)*
Age	**Base: Age 18-24**	
Age 18-24	0.316	0.151
	(1.065)	*(0.451)*
Age 25-44	**0.622** **	**0.448** *
	(3.357)	*(2.072)*
Gender	**Base: Male**	
Female	**-0.823** ***	**-0.939** ***
	(-4.847)	*(-4.546)*
Income	**Base: Income < $75,000**	
Income >= $75,000	-0.058	
	(-0.385)	
Rider History	**Base: Since Childhood**	
One year or less	**-1.791** ***	**-1.388** ***
	(-8.127)	*(-5.209)*
Several years	**-0.554** **	**-0.596** **
	(-3.532)	*(-2.994)*
Cycling Frequncy	**Base: Daily**	
Several times per week		**-0.638** **
		(-3.045)
Several times per month		**-1.68** ***
		(-6.361)
Model Statistics		
McFadden's ρ2 (MS model, base EL)	*0.093*	*0.093*
McFadden's ρ2 (Full model, base EL)	*0.33*	*0.58*
LL(Null Model)	*-1086.527*	*-1086.527*
LL(MS Model)	*-985.081*	*-985.081*
LL(Full Model)	-723.306	*-459.281*
G2(Full Model, base EL)	*726.442*	*1254.492*

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Table 2(c). Multinomial Logistic Models

Co-efficients	Enthused and confident Estimates *(t-stat)*		Strong and fearless Estimates *(t-stat)*	
	Model 1 *N= 740*	*Model 2* *N= 496*	*Model 1* *N= 740*	*Model 2* *N= 496*
Intercepts	**Base: Comfortable, but cautious & Interested, but concerned**			
Enthused and Confident	**0.51*** *(2.09)*	**1.183 **** *(3.546)*		
Strong and Fearless			-0.274 *(-0.84)*	**0.829 *** *(1.933)*
Age	**Base: Age 45+**			
Age 18-24	-0.122 *(-0.355)*	0.202 *(0.513)*	0.363 *(1.392)*	0.297 *(0.519)*
Age 25-44	**0.348 .** *(1.6)*	**0.394 .** *(1.576)*	**0.945 **** *(3.149)*	**0.731*** *(2.041)*
Gender	**Base: Male**			
Female	**-0.413*** *(-2.14)*	**-0.478 *** *(-2.072)*	**-1.64 ***** *(-4.833)*	**-2.199 ***** *(-4.39)*
Income	**Base: Income < $75,000**			
Income >= $75,000	-0.221 *(-1.19)*		-0.1 *(-0.421)*	
Rider History	**Base: Since Childhood**			
One year or less	**-1.305 ***** *(-5.38)*	**-1.053 **** *(-3.576)*	***-2.61* ***** *(-6.09)*	**-2.07 ***** *(-4.176)*
Several years	-0.061 *(-0.315)*	-0.156 *(-0.648)*	***-0.921***** *(-3.675)*	**-1.077 **** *(-3.135)*
Cycling Frequncy	**Base: Daily**			
Several times/week		**-0.771 **** *(-2.767)*		**-0.954 **** *(-2.789)*
Once or less/week		**-1.556 ***** *(-4.949)*		**-2.547 ***** *(-5.305)*
Model Statistics				
McFadden's ρ2 (MS model, base EL)	*0.09*			
McFadden's ρ2 (Full Model, base EL)	*0.34*	*0.58*		
LL(Null Model)	*-1080.56*	*-1080.56*		
LL(MS Model)	*-985.08*	*-985.08*		
LL(Full Model)	*-716.136*	*-452.958*		
G2(Full Model, base EL)	*537.89*	*1064.24*		

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Table 3. Odds Ratio for Multinomial and Ordinal Models with and without Cycling Frequency

	MNL Model 1		MNL Model 2		Ordinal Model 1	Ordinal Model 2
	Enthused and confident	Strong and fearless	Enthused and confident	Strong and fearless		
Age 18-24	0.884	1.89	1.223	1.345	1.372	1.163
Age 25-44	1.416	2.574	1.482	2.077	1.863	1.565
Gender	0.66	0.193	0.62	0.11	0.439	0.39
Income > = $75,000	0.8	0.904			0.943	
Rider history less than a year	0.27	0.0735	0.349	0.126	0.167	0.249
Rider history several years	0.94	0.398	0.859	0.341	0.575	0.551
Cycling frequency several times per week			0.462	0.385		0.529
Cycling frequency once or less per week			0.211	0.078		0.186

Overall, some distinct patterns were visible across all the models that we experimented with:

(1) Gender was significant in all the models with a negative sign implying that female cyclists are more likely to classify themselves into low comfort low confidence groups. The negative coefficients increase in value as we move from the *comfortable but cautious* and *interested but concerned* group to *strong and fearless* group which strengthens the previous inference. For the ordinal logit models, the odds ratio is ~ 0.4 which means that being female decreases the probability of being in higher confidence groups by about half. From the MNL models, being a female rider decreases the chance of being an *enthused and confident* rider as compared to *comfortable but cautious* rider by more than

30% while the chance of being a *strong and fearless* rider as compared to comfortable but cautious rider is decreased by about 80%.

(2) Cyclists in the age group of 25-44 and 18-24 are more likely to be more confident riders than the cyclists in the age group of 45+. From the ordinal model without cycling frequency, cyclists in the age group of 25 to 44 are ~ 86% more likely to classify themselves into more confident categories as compared to the cyclists in the age group of 45+ while from the model with cycling frequency, riders in the age group of 25-44 are about 56% more likely to classify themselves into higher confidence groups; cyclists in the age group of 25-44 are also more likely to classify themselves into higher confidence groups than cyclists in the age group of 18-24. This may be due to the inherent construct of the dataset where most users in the age group of 18-24 are students and use bicycle because they do not have access to a car. Intuitively, they may be less bicycle enthusiasts than riders in the age group of 25-44, who, being in the higher income group (also a construct of this dataset), may have access to an automobile but still choose cycling as a mode of commute.

(3) Income is not significant but income greater than $75,000 is positively related to classifying oneself into *strong and fearless* and the *comfortable but cautious* and *interested but concerned* group and is negatively related to classifying oneself into *enthused and confident* group.

(4) Riders with more experience are likely to be more confident as is captured by the negative coefficients of rider history of several years and rider history of one year or less as compared to the riders riding from childhood. Riders in the several years category

show odds ratios of 0.55 and 0. 57 for the ordinal model implying that such riders are ~45% less likely to be as confident as the riders riding from childhood while following the same logic, the new riders are ~75-85% less likely to be as confident as those riding from childhood.

(5) Cycling frequency is a significant determinant of rider type, and higher frequency of cycling implies a more confident cyclist. Cyclists with cycling frequency several times per week and cycling frequency once or less per week are both less likely to be more confident than cyclists with cycling frequency daily. However, the magnitude of the coefficient is higher in the once or less per week category than several times per week implying that cyclists in that category are even less likely than the cyclists in the several times per week category to be more confident riders. Cyclists who bicycle several times per week are about 50% less likely to rate themselves into higher confidence categories than riders who bicycle daily. Similarly, cyclists with cycling frequency once or less per week are about 80% less likely to classify themselves into higher confidence categories as compared to daily cyclists.

(6) Since the ρ^2 are similar across binary, multinomial, and ordinal models, it is difficult to justify the use of any one particular type of model for the purpose of cyclist classification. However, ordinal models impose an inherent restriction on the estimation process by assuming that the effect of the explanatory variables are the same at different category levels, i.e., how gender influences in self-classifying someone as a *comfortable but cautious* rider rather than an *enthused and confident* rider is the same as the influence of gender on being *enthused and confident* rather than *strong and fearless.* This may not hold true if the perceived difference in confidence between being *strong and fearless* and

enthused and confident is smaller than the difference between *comfortable but cautious* and *enthused and confident*. Gender may have a much more pronounced effect on choosing whether a rider is *comfortable but cautious* as compared to *enthused and confident* than in choosing between *strong and fearless* and *enthused and confident* rider type. Therefore, conceptually, MNL models seem to be more appropriate for the purpose of this research.

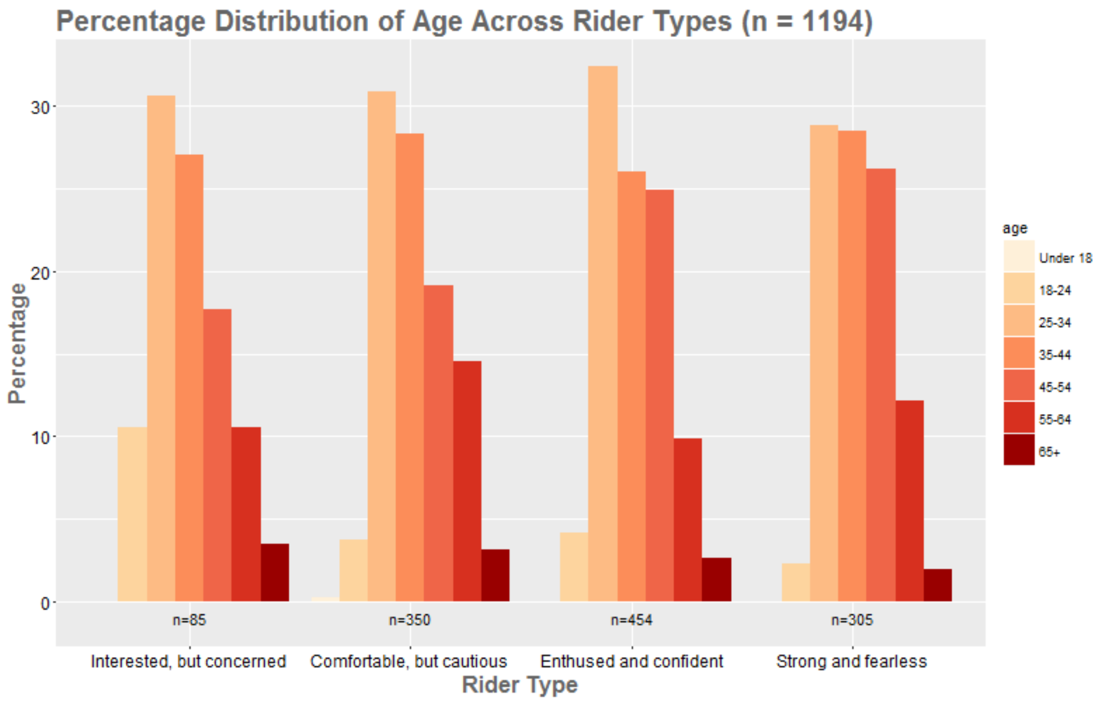
PART 2: Understanding Infrastructure Preference of Cyclists

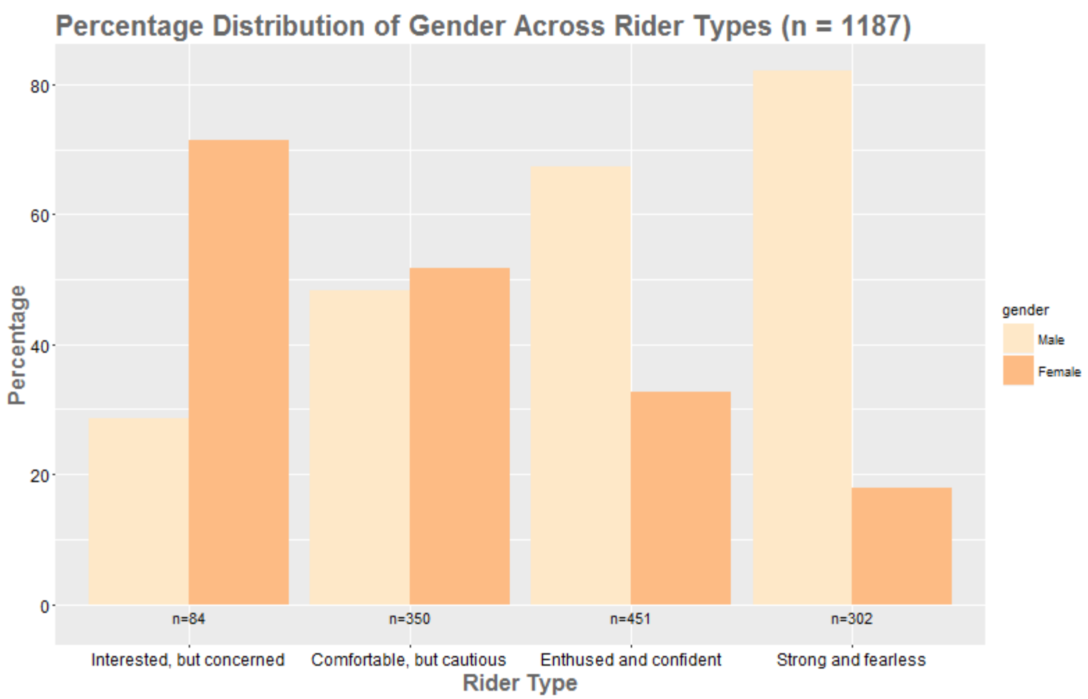
In addition to the influence of socio-demographics on cyclist types, it is important to understand how cyclist type influences preferences for infrastructure. The basic premise of this part of the research is to understand whether route preference is perception dependent and if that perception is a construct of the socio-demographic background of the cyclist. Based on model results presented in the first part of this paper, we hypothesized that female riders and riders in the age group above 44 years are more likely to prioritize safety over shortest route. Based on literature review (Hood et al. 2011, Sener et al. 2004), we also hypothesized that route impediments like high slope or poor pavement conditions are more likely to deter female cyclists from choosing that route.

Data Source: Cycle Atlanta User Survey and Atlanta Regional Commission Survey

The second part of the study is based on an online survey that was conducted among the users of the Cycle Atlanta application in the spring of 2014. The survey was sent to current application users via email addresses provided on the same user information screen asking demographics questions. The survey was divided into a few segments: (i) users' feedback on the Cycle Atlanta application, (ii) user feedback on the level of civic engagement and public participation in planning achieved through the Cycle Atlanta application, (iii) user feedback on which factors made them more likely to choose cycling as a mode of transport and (iv) socio-demographic and cycling related information of the respondents including self-classified rider type. The survey was sent to 697 application users and the particular question considered for this part of the analysis had 127 responses, a response rate of approximately 18%.

To increase the number of observations for this analysis, we appended the Atlanta Regional Commission's Bicycle User survey dataset to the existing Cycle Atlanta survey dataset. Both the Cycle Atlanta survey and the regional bicycle user survey were web based surveys advertised through the same channels – the Cycle Atlanta survey was designed in accordance with the regional survey to preserve the comparability of the datasets. Since the regional survey went out to all the bicyclists in the Atlanta region, there is a possibility that it would include the Cycle Atlanta users as well and some respondents may be present in both the datasets. Euclidean distance matrices, which quantify the dissimilarity between rows of sample data, were calculated individually for the Cycle Atlanta and the regional survey data and also after appending the datasets. The distance measures were found to be similar for both the cases and therefore, it was assumed that even though combining the two datasets may result in some data overlap, it will not significantly affect the results and the interpretation of the results. In addition, to maintain compatibility with Cycle Atlanta users, only the bicyclists in the regional survey having access to smartphones were included in the analysis. The non-smartphone respondents in the regional survey were instead used to verify that smartphone ownership does not bias results. Figure 14 shows the sociodemographic distribution of the pooled survey respondents across different rider types.


Percentage Distribution of Age Across Rider Types (n = 1194)
Percentage
0
10
20
30
n=85
n=350
n=454
n=305
Interested, but concerned
Comfortable, but cautious
Enthused and confident
Strong and fearless
Rider Type
age
Under 18
18-24
25-34
35-44
45-54
55-64
65+


Percentage Distribution of Gender Across Rider Types (n = 1187)
Percentage
0
20
40
60
80
n=84
n=350
n=451
n=302
Interested, but concerned
Comfortable, but cautious
Enthused and confident
Strong and fearless
Rider Type
gender
Male
Female

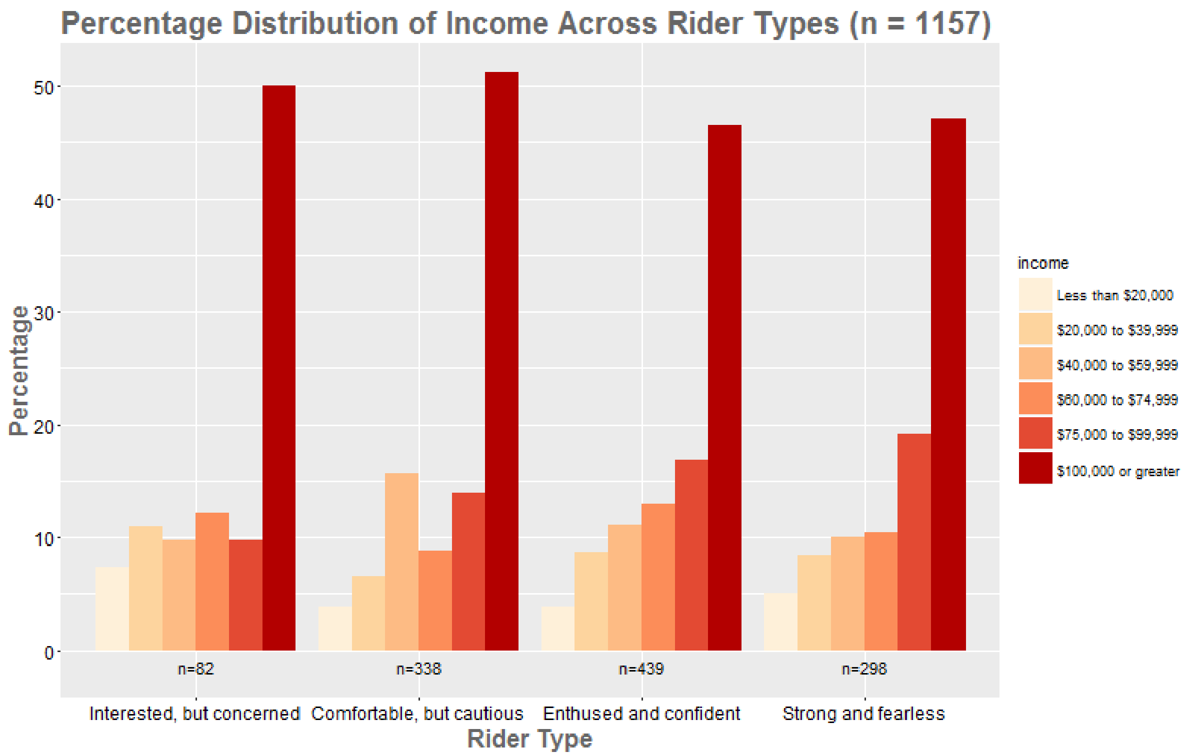


Figure 14. Socio-demographic Distributions of Pooled Survey Respondents across Rider Types

Table 4. Means and Standard Deviations of Item Responses on Road Conditions and Facilities by Rider Type

	Strong and Fearless		**Enthused and Confident**		**Comfortable but Cautious**		**Interested but Concerned**		**Significance in difference in mean scores (ANOVA)**
Conditions	Mean Ratings	Std. Deviation	Mean Ratings	Std. Deviation	Mean Ratings	Std. Deviation	Mean Ratings	Std. Deviation	
Bike Lane	3.13	*2.02*	3.41	*2.08*	3.47	*2.02*	3.29	*1.89*	***
Separate Path	3.00	*1.98*	3.39	*2.09*	3.65	*2.08*	3.72	*2.00*	***
Heavy Traffic	1.44	*1.35*	1.07	*1.25*	0.64	*1.10*	0.43	*0.98*	***
High Speed	1.03	*1.28*	0.60	*1.08*	0.47	*0.96*	0.31	*0.83*	***
Safe Routes	3.06	*2.08*	3.08	*2.12*	3.04	*2.10*	2.78	*2.10*	*
Directness	3.03	*1.96*	2.98	*1.93*	3.07	*1.95*	3.18	*2.03*	
Poor Pavement	1.57	*1.31*	1.42	*1.21*	1.20	*1.25*	1.11	*1.25*	***
Steep Hill	2.31	*1.43*	2.16	*1.41*	1.85	*1.40*	1.28	*1.22*	***
Parked car	2.23	*1.26*	2.18	*1.20*	1.97	*1.25*	1.68	*1.24*	***
Traffic Signal	2.15	*1.28*	2.20	*1.22*	2.00	*1.34*	1.79	*1.45*	*
Attractive Scenery	2.82	*1.83*	2.86	*1.76*	2.96	*1.78*	2.84	*1.79*	
Note:	*Responses were coded as a five point Likert scale (Much less likely =1, Much more likely = 5). Scores higher than 3.00 indicate a preference for that facility or condition while scores less than 3.00 indicate a negative impact of that facility or condition on choosing bicycling as an option. *** indicates significance at 0.001 level, ** indicates at 0.01 level, and * indicates significance at 0.05 level.*								

Table 5. *p*–values for Pairwise t-test on Respondents' Ratings on Influence of Road Conditions and Facilities on Bicycling Propensity, Paired by Rider Type

Road Condition	Rider Type	SF	EC	CC
Bike Lanes	EC	0.151		
	CC	**0.008**	0.048	
	IC	0.657	0.127	**0.000**
Separate Paths	EC	0.70421		
	CC	0.45595	0.552	
	IC	0.05538	**0.00033**	**1E-05**
Heavy Traffic	EC	0.16		
	CC	**1.00E-05**	**7.60E-07**	
	IC	**1.10E-08**	**1.20E-11**	0.014
High Speed	EC	0.1718		
	CC	**0.0037**	0.0121	
	IC	**2.30E-06**	**1.10E-07**	**0.0013**
Safe Route	EC	0.0439		
	CC	**0.0039**	0.1687	
	IC	**0.0154**	0.492	0.5504
Directness	EC	0.988		
	CC	0.214	0.036	
	IC	0.683	0.507	0.189
Poor Pavement	EC	0.302		
	CC	**0.004**	**0.002**	
	IC	**0.003**	**0.002**	0.750
Steep Hills	EC	**0.00058**		
	CC	**7.40E-09**	**0.00014**	
	IC	**1.90E-09**	**3.50E-05**	0.47
Parked Cars	EC	0.02712		
	CC	**2.40E-05**	**1.02E-03**	
	IC	**1.10E-04**	**7.71E-03**	0.73144
Traffic Signals	EC	0.1818		
	CC	**0.0029**	**0.0074**	
	IC	0.0345	0.2100	0.2097
Attractive Scenary	EC	0.196		
	CC	0.016	0.072	
	IC	0.143	0.764	0.157

Note: Rider Type: SF = Strong and fearless, EC = Enthused and confident, CC = Comfortable, but cautious, IC = Interested, but concerned. Each cell value represents p-value for pairwise t-test between the row-column pair of rider type. Significant values are marked in bold.

Basic Statistics

In both the surveys, the respondents were asked to indicate on a five point scale of less likely to highly likely, how the presence of route choice related factors may influence their decision to choose bicycling as a mode of transportation. The mean scores and standard deviations of each option were further calculated for each rider type and are shown in Table 4 along with the significance in difference of scores across rider types. Scores higher than 3.00 indicate a preference for that facility or condition while scores less than 3.00 indicate a negative impact of that facility or condition on choosing bicycling as an option.

In general, bike lane and separate paths have a very high score across all rider types implying that all riders prefer dedicated facilities. Although this is generally not surprising, it is counter to the opinions expressed by some vehicular cycling enthusiasts (Forester 2012). Similar high scores are noted for safety and directness indicating that dedicated bike facilities along shortest routes are preferred by all riders. Negative conditions like poor pavement, steep hills, parked cars, and traffic signals negatively affect the decision to bicycle, but to a much lesser degree than traffic speed and volume. Traffic stress stands out as the most deterring factor preventing people to decide in favor of bicycling.

While average scores for individual conditions and facilities are mostly similar across the first three rider types, an ANOVA done on the scores show significant difference across rider types. To understand which groups actually differed significantly, a pairwise t-test was conducted. The p-values are shown in Table 5. For neighboring categories, there is no significant difference in mean scores across rider types strong and fearless and enthused and confident for any item while interested but concerned and comfortable but cautious groups have significant difference for traffic speed and separate facilities. There is significant difference across groups particularly for heavy traffic, high speed, poor pavements, steep hills and parked cars. There is no significant difference in item scores for directness of route and attractive scenery on route.

Further exploratory analysis was performed to validate the hypothesis that reported preference of infrastructure and facilities depends on the socio-demographic attribute of the users. First, a factor analysis was performed to group the 10 road conditions and facility preferences into fewer factors. Then, regression analyses were done with each factor as the dependent variable and the sociodemographic attributes of the respondents as the explanatory variables.

Factor Analysis

Factor analysis is used to address underlying correlation among all or some of the observed variables such that there will be multi-collinearity issues if the variables are treated individually in the data (Thompson 2004, Hurley et al. 1997). Factor analysis, thus, helps in reducing the dimensionality of the dataset as well as in identifying the correlated structure of observed variables. For example, among the 10 variables influencing the decision to bicycle, high traffic speed and heavy traffic volume may be correlated (that is people who are averse to high traffic speed are also likely to be averse to heavy traffic volume). By using factor analysis, we may be able to group these two variables together to form a new factor variable which can then be used for regression. This will reduce the number of variables from 10 to 9 and will remove the collinearity that would occur if both the variables are treated separately in regression.

Factor analysis can be exploratory or confirmatory, the latter being used to test a pre-determined hypothesized correlation among some of the variables (Hurley et al. 1997). In our case, no correlation structure was initially hypothesized, and hence, an exploratory factor analysis was performed. The variables were allowed to load into all factors, irrespective of their score, and an orthogonal rotation was used. A scree plot was used to determine the optimum number of components; and two models, one with two factors and another with three factors, were tested. The three factor model was used for further analysis. Table 6 presents the results of the three factor model.

Table 6. Exploratory Factor Analysis: Loadings

Loadings:	Factor1	Factor2	Factor3
Bike lanes	0.884	0.337	
Separate paths	0.89	0.316	
Safe route	0.767	0.389	
Directness	0.797	0.357	
Attractive scenery	0.76	0.405	
Steep hills	0.445	0.577	
Parked cars	0.43	0.714	
Traffic signals	0.443	0.645	
Heavy traffic			0.778
High traffic speed			0.782
Poor pavement	0.399	0.42	0.386
	Factor1	Factor2	Factor3
Sum of Square loadings	4.173	2.164	1.763
Proportion Var	0.379	0.197	0.16
Cumulative Var	0.379	0.576	0.736

Table 6 shows that factor 1 has high loading on bike lanes, separate paths, safe route, directness, and attractive scenery. This factor was named Protected Environment as the preference of people scoring high on this factor appears to be direct and safe facilities. The second factor has moderately high scores on steep hills, parked cars and traffic signals and was therefore named Route Impedance implying that people who score high on this factor prefer routes with less disruption. The third factor has the highest loading on heavy traffic and high traffic speed and was named Route Stress indicating that people who score high on this factor are averse to traffic stress: their decision to bicycle is largely determined by the traffic speed and volume in the corridor.

Regression Analysis

In the second stage, to understand if sociodemographic attributes of riders influence the infrastructure preference, the three factors were used in regression equations with age, gender, income, and rider type as explanatory variables and the factors as the dependent variables. Table 7 provides the details of the regression analysis.

For the Protected Environment factor, gender, income, and rider type are significant implying that females and those in high income groups prefer facilities. The rider type variable has a negative coefficient indicating that people with lower confidence levels prefer separate facilities. Both Route Impedance and Route Stress are factors with negative connotations and the regression results should be interpreted accordingly. Age, income, and rider type are significant for the Route Impedance factor, which includes steep hills, parked cars, and delayed traffic signals. Age shows a negative sign implying that older riders have a stronger aversion to route impedances like steep hills or parked cars, while rider type shows a positive coefficient implying that less confident riders are more averse to route impedance.

For the Route Stress factor, age, gender, income, and rider type are significant variables. Age, gender, and income have negative coefficients while rider types has a positive coefficient. The result can be interpreted as older and female riders, as well as riders in the high income group, are less likely to decide to bike under traffic stress while riders in the low confidence category are also deterred from bicycling because of high traffic speed and volume.

It should however be noted that all the regression models have low R-squared values, ranging from 0.04 (Protected Environment and Route Impedance) to 0.1(Route Stress). Therefore, while it can be implied that infrastructure preference affects the decision to bicycle differently for older, female, and less confident riders, it is also imperative that there are other factors that influence the decision to bicycle which are not included in these models.

Table 7. Regression Analysis for Protected Environment, Route Impedance and Route Stress

Coefficients	Protected Environment				Route Impedance				Route Stress			
	Estimate *(t-stat)*	Sig.	Estimate *(t-stat)*	Sig.	Estimate *(t-stat)*	Sig.	Estimate *(t-stat)*	Sig.	Estimate *(t-stat)*	Sig.	Estimate *(t-stat)*	Sig.
(Intercept)	**-0.304** *(-2.304)*	*	**-0.964** *(-6.138)*	***	**0.48** *(3.702)*	**	**-0.231** *(-1.488)*	.	**0.172** *(1.595)*	.	**-0.501** *(-3.888)*	**
Age	**-0.106** *(-4.372)*	***	**-0.085** *(-3.491)*	**	**-0.082** *(-3.329)*	**	**-0.066** *(-2.721)*	**	**-0.047** *(-2.282)*	*	**-0.029** *(-1.45)*	.
Gender	**0.157** *(2.833)*	**	**0.187** *(3.273)*	**	**-0.197** *(-3.547)*	**	-0.042 *(-0.744)*		**-0.077** *(-1.672)*	.	0.05 *(1.059)*	
Income	**0.0723** *(3.791)*	**	**0.062** *(3.323)*	**	-0.002 *(-0.1)*		0.003 *(0.154)*		0.010 *(0.658)*		0.013 *(0.84)*	
Strong and fearless			**0.443** *(4.735)*	***			**0.783** *(8.482)*	***			**0.677** *(8.829)*	***
Enthused and confident			**0.821** *(9.424)*	***			**0.476** *(5.527)*	***			**0.53** *(7.418)*	***
Comfortable but cautious			**0.8** *(8.994)*	***			**0.28** *(3.189)*	***			**0.336** *(4.602)*	***
Model Statistics	Multiple R-squared: 0.030, Adjusted R-squared: 0.028 F-statistic: 12.2 on 3 and 1181 DF, p-value: 7.261e-08		Multiple R-squared: 0.116, Adjusted R-squared: 0.111 F-statistic: 25.48 on 6 and 1167 DF, p-value: < 2.2e-16		Multiple R-squared: 0.019, Adjusted R-squared: 0.016 F-statistic: 7.552 on 3 and 1170 DF, p-value: 5.26e-05		Multiple R-squared: 0.084, Adjusted R-squared: 0.08 F-statistic: 17.93 on 6 and 1167 DF, p-value: < 2.2e-16		Multiple R-squared: 0.006, Adjusted R-squared: 0.004 F-statistic: 2.503 on 3 and 1170 DF, p-value: 0.0579		Multiple R-squared: 0.076, Adjusted R-squared: 0.071 F-statistic: 16 on 6 and 1167 DF, p-value: < 2.2e-16	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

DISCUSSION OF RESULTS

A substantial component of the self-classification of rider types may be correlated with the socio-demographic make-up of a cyclist. In particular, gender and age have a demonstrated effect on an individual's attitude towards safety, comfort, and confidence. Studies on the effect of gender on confidence have shown that females are much less likely to undertake risky tasks and more likely to report themselves to be less confident than their male counterparts even when performing identical tasks (Kray et al. 2001). A study by Byrnes et al. (1998) showed that this gender gap increases with increasing age although the level of gap is decreasing over time. Emond et al. (2009) also emphasized the difference in individual and social perceptions between the two genders and its strong influence on self-efficacy.

This study was undertaken to provide a data driven answer to the question whether cycling infrastructure preference depends on the socio-demographic characteristics of the cyclists. We also aimed to validate the generalizability of the existing popular classification of cyclists based on their comfort and confidence level via information provided by the cyclists themselves. Results of our first analysis show that age, gender, and rider history influence self-classification of the riders into rider types. Overall, the confidence and comfort level decreases with age and is significantly lower for female riders as compared to their male counterparts. Cycling frequency and riding history both have a significant role in determining rider type but are positive reinforcement only through an adoption phase. The binary logistic regression model fits are comparatively lower than the ordinal logistic and ordered probit models indicating the validity of an underlying scale of confidence and comfort in the construct of rider type categories. Future research should aim to validate the claim based on revealed preference route choice data.

The purpose of the second part of the analysis was to understand if perceptions about route characteristics and safety are related to a cyclist's sociodemographic make up and self-perception as a particular rider type from a stated preference dataset and without any input from revealed route choice preferences. The results indicate a preference of dedicated facilities across all rider types and also point toward high speed traffic and high volume traffic as the factors that most negatively affect willingness to cycle. Further analysis reveals that female riders are more likely to prefer separate facilities while elderly riders are more likely to be averse to steep hills and other route impedance. Less confident riders prefer both separate facility and less impedance in their routes. Higher income riders prefer protected environments but are not as deterred by presence of slopes or traffic signals on their routes. High traffic speeds and heavy traffic volumes deter females, older riders, and less confident riders from bicycling.

LIMITATIONS

The innovative use of smartphone based application to collect revealed preference cyclist route choice data has its own caveats. The Cycle Atlanta data suffer from the issues of self-selection bias as it is a crowd sourced data collection system and people who participate are those who are sufficiently interested in the project, willing to share data, and invest time without any personal gain. As a result, the Cycle Atlanta dataset is heavily dominated by male white cyclists in the age group between 25-44 years. This is not a representative sample of the population of Atlanta where 50% of the population is female and about 54% is African American (ACS 2012). However, there is currently no reliable estimate on the makeup of the cycling population of Atlanta (Poznanski 2013), and hence it is difficult to comment on the representativeness of the Cycle Atlanta data with regard to the cycling population of Atlanta. We compared the sociodemographic distribution of the Cycle Atlanta users to the participants of the Atlanta Regional Commission data and found no statistically significant difference. While this may mean that the cycling population of Atlanta is fairly homogenous, it may also be due to the reason that Atlanta Regional Commission's survey was advertised through pro bicycle channels that typically reach out to people with similar attitudinal preferences as the Cycle Atlanta users. It is an ongoing future research debate as to whether weighting the data by Atlanta population proportion should be considered as that may interfere with the representativeness of the collected data.

The other issue associated with the data is that by design, the data systematically miss the cyclists who do not own a smartphone. A study on smartphone ownership (Windmiller et al.

2014) has shown that this systemic bias affects people in older age groups, certain ethnic groups, and sometimes people in lower income groups. The Cycle Atlanta dataset is sparse in all of these categories – the users are mostly White and high income in the age group of 25- 44. It is difficult to estimate how much of the sparsity is caused by use of smartphone for data collection and how much is due to characteristics that define cyclists of Atlanta. However, the number of non-smartphone owning Atlanta Regional Commission participants and Cycle Atlanta participants were similar in number, therefore making up a similar portion of the sample. The models were run for the non-smartphone owners as a separate group and no statistically significant difference in infrastructure preference was noticed. Therefore, in spite of the biases in the data collected via smartphone app, any infrastructure requirement predicted based on Cycle Atlanta data can be assumed to hold true for non-smartphone users as well.

CONCLUSION AND FUTURE RESEARCH

This research provided an analytical approach to understand the characteristics and preferences of the different types of cyclists. Cyclists using the Cycle Atlanta tracking application were given the option to self-classify themselves into rider types, and socio-economic data were collected to understand the basis for riders' choice of rider type. The rider type classification was also used in a stated preference survey on route choice attributes to understand if such preferences are influenced by rider types.

The first part of the analysis shows that socio-demographic variables and riding pattern are significant predictors of a cyclist's probability of self-classifying himself/herself into a particular category. In particular, gender, rider history, and cycling frequency are significant in all the models. The results indicate that knowing a cyclist's demographic information can potentially help in classifying the cyclist into a particular rider type. In the future, this can help us to streamline surveys by replacing sociodemographic questions by a single rider type classification question. Alternatively, by knowing the socio-demographics characteristics commonly available through census data and other surveys, we will also be able to predict the rider type and hence infrastructure preferences of people without having to undertake a new survey design for cyclists only. It will also help in understanding infrastructure and facility need of future cyclists who are not yet cycling and hence, there are no revealed preference data on the preference of such future cyclists currently.

The results also direct attention to the requirement of segmented route and facility preference decision models for different cyclist types. Since the purpose of the route and facility preference analysis is to understand the requirements by rider types, segmented models based on rider type may enable a planner to better predict the choices of a future cyclist based solely on the demographic information of the cyclists. Future route decision model research may therefore explore segmentation of the dataset to achieve better predictability.

From the second part of the analysis, it is evident that most route perception issues and facilities are viewed on a similar scale by cyclists as the mean scores on those facilities are quite similar across rider types. Other results indicate that sociodemographic attributes and confidence levels influence infrastructure and facility preference. However, the model fits are substantially low indicating that rider level data are not sufficient to predict route level decision process. Further investigation is necessary, as the literature shows that choice of route depends on route characteristics as well as rider characteristics like age and gender. Therefore, in future, for further insight, we plan to augment user data by revealed preference route choice data to make any definitive conclusion about the preference and requirements of cyclists.

CHAPTER 6. ROUTE CHOICE MODELING

INTRODUCTION

In recent times, bicycling as an alternative mode of transportation has been promoted both at the federal as well as local government level. Cycling is a healthy, green, and affordable mode of transportation that can provide easy accessibility to a multitude of destinations that are not within walking distances. However, the adoption of cycling as an alternative mode of commute is considerably hampered by a perceived lack of safety on the part of users, a major component of which arises from a lack of dedicated facilities. For the regional planning agencies, building dedicated infrastructure comes with substantial investment requirements and often with decisions to convert vehicular traffic lanes to bicycle facilities, both of which require strong justification that indicates potential benefits of such construction.

In cities like Atlanta, where cyclists are traditionally few in number, in addressing issues on cycling infrastructure, regional planning agencies face the additional problem of lack of data on cyclists and their preferences. Most often, this lack is addressed by conducting surveys where participants are asked to indicate factors that influence their decision to bicycle or to map their latest rides. However, such stated preference surveys come with the issues of recall bias and often, a relatively small sample size that may not sufficiently represent the cycling community.

A common alternative to stated preference surveys is revealed preference where data on vehicle trajectories are collected via vehicle tracking devices. Vehicle route choice preferences are then modeled based on revealed chosen trajectory. Although frequently used for collecting data on automobiles, this approach has been rarely used for cyclists because of multiple issues. First, until recently, tracking devices were prohibitively costly and hence, were only used for cases of primary importance to travel demand modeling and traffic flow management. Second, computational effort required to model route choice is significant and its effectiveness is dependent on availability of a high resolution network. Often, street networks are not updated to include bicycle facilities recently constructed and bicyclists tend to use by-lanes and cut-thrus that are rarely found in street networks, both of which render route choice modeling for bicyclists much less effective and much more complicated. Finally, bicyclists are much less likely to optimize routes based on travel time which is the standard optimization algorithm used for vehicular traffic. Developing and using algorithms suited for modeling bicyclist route choice require separate efforts than the standard practice and agencies are often restricted by budget to allocate separate resources for cycling.

Recently, integration of GPS capabilities into hand held devices and smartphones have opened up a new dimension in low cost real time data collection, and bicycling research has gained significantly from such advances. Hood et al. (2011), Broach et al. (2012) and Menghini et al. (2011) used either standalone GPS device-based data or GPS enabled smartphone-based data to analyze and model route choices of cyclists in San Francisco, Portland, and Zurich, respectively. New computationally efficient algorithms have been designed and proposed to generate route alternatives that consider different optimization objectives like slope, scenery, traffic speed, and presence of facilities. However, availability of high resolution network data and matching GPS data to that network still remains an issue that hinders route choice modeling of cyclists.

Parallel to route preference, research on cyclists suggests that preference of cycling infrastructure may depend on comfort and confidence level of cyclists which is popularly categorized into four different categories of *strong and fearless*, *enthused and confident*, *interested but concerned*, and *no way no how* (Geller 2006). In our previous research, we slightly modified these categories to *strong and fearless, enthused and confident, comfortable but cautious*, and *interested but concerned* and related them to the socio-demographics of the cyclists (Misra et al. 2015). We further used stated preference surveys to understand if infrastructure preference varied significantly by the rider types mentioned above. In this chapter, we extend that research using revealed preference data collected via the smartphone application Cycle Atlanta (Misra et al. 2014). In particular, we model the likelihood of choosing the shortest path between origin and destination as explained by socio-demographics and confidence level of the riders. In the future, this analysis will be extended to include route characteristics like traffic speed, traffic volume, number of lanes, and presence of facilities.

BACKGROUND AND MOTIVATION

There are three aspects to the route choice problem: (1) path enumeration or generating a set of travel alternatives for any chosen route between a pair of origin and destination, (2) estimating a disaggregate demand model based on individual route choices, and (3) predicting choice probabilities using the route choice model developed for different planning purposes. Route choice modelling presents unique challenges in all of the above categories. In this section, we present a brief contextual review of literature in the area of path enumeration and route choice modeling.

Path Enumeration/Choice Set Generation

For the path enumeration step, a region wide street network can theoretically provide an infinite set of alternatives for any chosen route between any pair of origin and destination. While that makes solving the problem computationally infeasible, at the same time, limiting choice set to any arbitrary number of alternatives for ease of computation leads to poor prediction performance for some model specifications (Horowitz and Louviere (1995), Bekhor and Prato

(2006), Prato and Bekhor (2007), Bliemer and Bovy (2008)). Therefore, a trade-off is needed in the use of a search algorithm that can produce a reasonably competitive set of alternatives while being computationally efficient. Figure 15 shows the different path generation algorithms popularly used in route choice modeling.

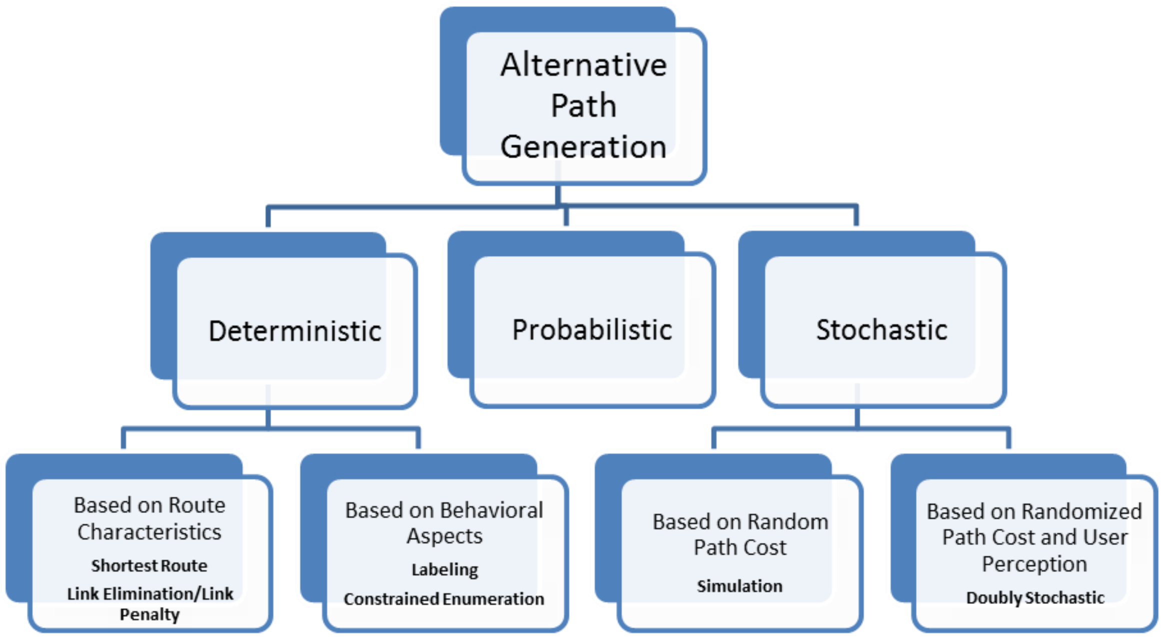


Figure 15. Different types of Path Generation Algorithms

Traditionally, generation of a set of alternatives to a chosen route is based on shortest route algorithms (Prato 2009, Broach 2010). Several modifications and variations of the shortest route algorithm have been proposed in the literature over time and interested readers may find a summary of such advances in Prato (2009). From the behavioral aspect, Ben-Akiva et al. (1984) proposed a labeling algorithm where each label represented a particular route characteristic that a user might want to minimize or maximize – for example, a label may be the shortest distance or the least congested path. Ramming (2002) used the labeling algorithm to search for the shortest route based on 16 different labels while Prato and Bekhor (2006) used 4 attributes to generate alternatives. The results show that the success of the labeling algorithm is dependent on choice of labels, which is up to the discretion of the analyst and based on the understanding of the user preference and behavior.

Recent choice set generation literature shows incorporation of randomness of individual preferences and is based on importance sampling – the probability of choosing a route depends on the importance of route characteristics like distance or congestion. Prato (2009), Freijinger (2007), Bovy et al. (2009), and Frejinger et al. (2009) suggest using a sampling correction term in the route choice model when choice alternatives are generated using these probabilistic methods.

In the simulation approach, it is assumed that users perceive path cost with some errors – so the cost function is assumed to be from a distribution and the results largely depend on the choice of the distribution from which the cost function is extracted. The doubly stochastic method is based on the assumption that both the path cost and the perception of the path cost vary among users – i.e. each user perceives path cost differently and with some error (Bovy 2009). The probabilistic methods are good at producing a heterogeneous set of alternatives and have been shown to replicate the observed route more frequently than other alternatives.

Two other methods not used very often in route choice are the constraint enumeration method and probabilistic method (Prato and Bekhor 2006, Bekhor and Prato 2009, Friedrich et al. 2001). The constrained enumeration method uses branch and bound algorithms and is based on the idea that instead of least cost, users often choose based on personal preferences. The algorithm was found to replicate the observed route closely but the computation time increases exponentially with the depth of the branching tree which is significant for a large network and is therefore limited in scope to small networks. The probabilistic method assigns a generated probability to each route and is therefore again computationally prohibitive to be carried out for a large regional network.

Route Choice Modelling

One of the predominant issues with route choice modelling cascades down from the previous step of choice set generation. Bekhor et al. (2006), Prato and Bekhor (2006), Freijinger (2007), and Bovy (2009) show that size and composition of the choice set effects the objective function and convergence rate of models – therefore, much of the prediction performance of the models will depend on the generated choice set and the alternatives contained therein.

The other issue with route choice modeling is that in a large network, alternative routes are often very similar and have overlapping links. This leads to the violation of the independence of irrelevant alternatives (IIA) property of logit structures and warrants either (i) use of models with explicit inclusion of correlation terms like the Generalized Extreme Value (GEV)family or probit models which are computationally expensive and complicated or (ii) use of some correction factor along with the multinomial logit structure to account for the path overlap. Table 8 shows the most common choice models for route choice modelling.

Table 8. Most Common Choice Models Used in Route Choice Modelling

Name	Basic Structure	Modification	Formulation
C-Logit (Cascetta et al. 1996)	Multinomial Logistic	Additive Similarity Measure	$P_k = \frac{\exp(V_k + \beta_{CF}.CF_k)}{\sum_{l \in C} \exp(V_l + \beta_{CF}.CF_l)}$
Path Size Logit (PSL) (Ben-Akiva and Bierlaire 1999; Ramming 2002)	Multinomial Logistic	Path Size Modification for Path Overlap	$P_k = \frac{\exp(V_k + \beta_{PS}.\ln PS_k)}{\sum_{i \in C} \exp(V_l + \beta_{PS}.\ln PS_l)}$
Path Size Correction (Bovy et al. 2008)	Multinomial Logistic	Path Size Correction for Path Overlap	$P_k = \frac{\exp(V_k + \beta_{PSC}.PSC_k)}{\sum_{i \in C} \exp(V_l + \beta_{PSC}.PSC_l)}$
Multinomial Probit (Daganzo and Sheffi 1977, Sheffi and Powell 1982), **Logit Kernel with Random Co-efficients/Mixed Logit model** (Ben-Akiva and Bolduc 1996, McFadden and Train 2000, Lam and Small 2001, Nielson 2002, 2004) **Logit Kernel with Factor Analytic Approach** (Bekhor et al. 2002, Frejinger and Bierlaire 2007)	Non GEV/Non Multinomial Logistic	Inherently do not have the IIA property	$P_{nk}(\beta_n) = \frac{\exp(\beta_n' X_{nk})}{\sum_{i \in C} \exp(\beta_n' X_{nl})}$ $P_{nk}(\beta_n) = \int \frac{\exp(\beta_n' X_{nk})}{\sum_{i \in C} \exp(\beta_n' X_{nl})} f(\beta) d\beta$

The first introduction of a correction factor in a logit structure is attributed to Cascetta et al. (1996). The model proposed by Cascetta et al. (1996), called C-logit, is based on MNL structure and involves estimating an additional term that is a similarity measure between a route and the other routes in a choice set. However, the most commonly used model specification is the path size logit (PSL) presented by Ben-Akiva and Bierlaire (1999) and modified later by Ramming (2002). The probability of choosing path k according to the PSL structure is

$$P_k = \frac{\exp(V_k + \beta_{PS}.\, lnPS_k)}{\sum_{i \in C} \exp(V_l + \beta_{PS}.\, lnPS_l)}$$

where PS_k and PS_l are the path sizes of routes *k* and *l* respectively and β_{PS} is the parameter to be estimated. V_k and V_l are the observed utilities of path *k* and *l* respectively ($l \neq k$). *C* is the set of path choice alternatives of which *i* is any alternative. Path sizes are formulated differently by Ben-Akiva and Bierlaire (1999) and Ramming (2002) and are given respectively as

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \delta_{al}} \quad \text{(Ben- Akiva and Bierlaire, 1999)}$$

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \left(\frac{L_k}{L_l}\right)^{\gamma_{PS}} \delta_{al}} \quad \text{(Ramming 2002)}$$

where L_a is the length of link *a*, L_k is the length of path *k*, L_l is the length of path *l*, $\delta_{al} = 1$ if link *a* is traversed by alternative *l* and 0 otherwise, and γ_{PS} is a scale factor.

While path size logit provides a computationally simple model, it only accounts for a part of the correlation and within MNL error structure. Frejinger (2007) estimated the PSL model with sampling correction while considering the full choice set of paths as the actual choice set, and results show that unbiased estimates are obtained only when the correction term is calculated on the full choice set.

Bovy et al. (2008) suggested a similar formulation of the problem but with modification of the path size term, and this model came to be known as Path Size Correction Logit (PSCL). The probability of choosing route k is given by

$$P_k = \frac{\exp(V_k \ + \beta_{PSC}.PSC_k)}{\sum_{i\in C}\exp(V_l + \beta_{PSC}.PSC_l)}$$

And the path size correction is given by:

$$PSC_k = -\sum_{a\in\Gamma_k}\frac{L_a}{L_k} ln \sum_{i\in C}\delta_{al}$$

where all terms have same interpretations as the previous equation (PSC_k and PSC_l are the path sizes of routes k and l respectively and β_{PSC} is the parameter to be estimated)

A completely different set of models not based on the logit formulation have been proposed to account for the correlation explicitly. These models are formulated based on the GEV structure and include Paired Combinatorial Logit (Prashkar and Bekhor 1998, 2000, Koelman and Wen 1998), Cross Nested Logit (Vovsha 1997, Prashkar and Bekhor 1998), and Generalized Nested Logit (Bekhor and Prashkar 2001, Wen and Koppelman 2001). While these models are good at accounting for link overlap issues, they are not used frequently because of the computational complexity and cost which far outweigh the benefits (Prato 2009).

Three different non-GEV and non-logit based models have been used – the multinomial probit (Daganzo and Sheffi 1977, Sheffi and Powell 1982), the Logit Kernel with Random Co-efficients or Mixed Logit model (Ben-Akiva and Bolduc 1996, McFadden and Train 2000, Jou 2001, Lam and Small 2001, Nielson et al. 2002, Nielson 2004), and Logit Kernel with Factor Analytic Approach (Bekhor et al. 2002, Frejinger and Bierlaire 2007). The computational costs for all of these models are significantly high, particularly for large networks which prevent their frequent use in route choice modelling problems.

Bicycle Route Choice

Bicyclist route choice is uniquely different from vehicular route choice in preferring safety over travel time or congestion and in using by-lanes and cut-thrus instead of the street network. Therefore, even if similar frameworks are used to model cyclist route choice and vehicular route choice, cycling route choice requires special attention in formulating choice preferences and in matching travelled roads to the street network. The first relevant study in this direction was done by Hall et al. (1997) for the city of Guelph, Ontario, Canada and since then, a few other publications have also addressed the issue of bicyclist route choice: Table 9 summarizes four such publications that are relevant to this research.

Table 9. Salient Bicycle Route Choice Literature Highlights

Article	Study Location	Choice Set Generation	Choice Model	Major Findings
Hall et al. (1997)	City of Guelph, Ontario, Canada	Shortest Distance	NA	(i) Grade (ii) traffic signals (iii) number of turns (iv) no. of bus routes per hour
Broach et al. (2009)	Portland, Oregon, U.S.A.	Labeling algorithm with 20 alternatives	Path Size Logit	Distance, slope, no. of turns, traffic volume are disutilities; bike facilities are useful
Menghini et al. (2010)	Zurich, Switzerland	Search and bound algorithm	Path Size Logit	Distance, average gradient, marked bike facility influence choice
Hood et al. (2009)	San Francisco, California, U.S.A.	Doubly Stochastic	Path Size Logit	Shorter routes, fewer turns are preferred; upslope is always a disutility; female cyclists avoid slope; bike facilities are preferred by new cyclists

The study by Hall et al. (1997) was followed up by a series of bicycle route choice studies – these studies overcame the data issues of the previous study by using a GPS-based automatic route data collection system which minimized user induced bias in the data, had the advantage of recording multiple trips for each individual, and also significantly increased the number of participants as now the participant either only had to carry an instrument that recorded the trip automatically or had to download and turn on a smartphone app to record his/her trip.

Two parallel data collection methods emerged using GPS-based route data for bicycle route choice modeling around the same time. In the first case, GPS enabled devices were distributed among the participants at the beginning of the study for a scheduled period of time, and the participant either had to carry it with him/her or had to mount the device on his/her bicycle. The data recorded were locally stored in the memory card of the device and at the end of the scheduled period, the devices were called back from the participants and the data were retrieved and analyzed. Proponents of this approach include the Oregon bike research group (Broach et al. 2009, Dill et al. 2008) and the travel demand modeling group headed by Dr. Kay Auxhausen at Eidgenossische Technische Hochschule, Zurich (Menghini et al. 2010) who have used this approach for data collection and route choice analysis.

The other method of data collection is to use the GPS enabled smartphones owned by participants for recording data and directly uploading the data wirelessly from the phone to a central server. The advantage in this method is that it is cheap and does not require any investment in equipment while data quality is comparable to the GPS devices. Pioneering work in this area was done at San Francisco County Transportation Authority (SFCTA) (Charlton et al. 2009, Hood et al. 2009) and was later extended by the Cycle Atlanta group (Poznanski 2013). Hudson et al. (2012) also did a similar route choice study in Austin.

In the following section, the research using these two different technologies is discussed in detail, and a critique of the models developed in these studies is also presented.

The GPS Device Based Data Collection Efforts

Dill et al. (2008) and Broach et al. (2009) both based their research on the same study of 162 cyclists in Portland, Oregon. While Dill et al. (2008) used the collected data to compare chosen route against the shortest path, Broach et al. (2009) extended the study to develop a path choice model for cyclists in Portland, Oregon. The studies included both utilitarian and recreational trips and participants were chosen through stratified sampling from respondents to an online survey. The demographic and personal characteristics used for stratification were cycling frequency (frequent vs. infrequent), home location (Portland vs. remainder), age, and gender. As mentioned earlier, at the beginning of their study, GPS devices were distributed to the participants, and at the beginning of each trip, the participant had to tap the screen to turn the device on. The user also had to enter a few other pieces of information by choosing from drop down menus provided: including the trip destination category (home, school, work, etc.), the weather (sunny, cloudy, rain, etc.), temperature (hot, moderate, etc.) and wind (heavy, light, or no wind). Enroute, the GPS unit recorded location data at every 3 seconds, and these data were stored in the device. At the end of the study period, the data were retrieved and each individual trip was mapped. Participants were then asked to log on to their maps and identify any trip that was recorded erroneously – this was supplemented by a questionnaire to validate the correctness of the data collected and to understand the reason behind the mode choice.

As with the study by Hall et al. (1997), snapping the chosen route to the city network required augmentation of the network with links from the route data collected and validation through aerial photography. Map matching algorithms were developed by the researchers to account for GPS related errors like snapping onto adjacent roads instead of the actual route, data point clouds at intersections, and at start and stop, erroneous turns, etc. The final cleaned and matched network data were used for the analysis.

A statistical analysis of the data revealed that women traveled less distance than men and also ranked facilities higher than men. The most important factor in choosing a route was stated to be minimum time followed by low traffic volume and presence of a bike lane. No significant relationship was found between route choice and slope. A comparison between shortest route and the actual route showed that people spent more time on bicycle facilities and low traffic streets than predicted by the shortest route and that the deviation from shortest route increased with length of trip.

Broach et al. extended the study by Dill et al. (2008) to develop a multivariate discrete choice model of bike route choice of cyclists in Portland. In doing so, Broach et al. overcame the issue of comparing actual route choice with only the shortest distance – the model now was capable of predicting marginal utilities of different attributes and handling any interaction between them. One of the difficulties in developing a discrete choice model is generating a feasible set of not-chosen alternatives. In a city road network, this has infinite possibilities and hence, a sorting algorithm needs to be used to restrict the choice set to a finite number. In this case, a labeling algorithm was used after trial and error with a few other algorithms. On average, 20 different alternatives were generated for every route. For the choice model, a multinomial logit model was used with correction for path choice overlapping using path size correction term from Ben-Akiva and Bierlaire (1999). The path attributes used for the model were distance, slope, turns, traffic volume, signals, and bike facility type. With all other parameters held constant, log distance was the most important factor in route choice, implying that for a short commute, a cyclist will be less willing to take the same detour as he/she would be if the commute was longer. Distance was found to be strongly correlated with travel time and hence not included in the study. Slopes and turns consistently had negative coefficients implying a disincentive attached to routes with high slopes or a significant number of turns. Traffic volume also proved to be a disutility, while traffic signals had a positive utility when the cross traffic was high and had a disutility for low traffic streets. Bike boulevards and paths were strongly preferred while the utility associated with bike lanes was just enough to offset the disutility of traffic volume in that link. Therefore, bike lanes are preferred in streets with high traffic and over busy arterials without any bike lane, but they do not add any separate value to the cyclists by themselves. The route choice model developed by Broach et al. is being incorporated into the regional travel demand model of Metro, the Portland area municipal planning organization (MPO) in an effort to better predict where cyclists travel and what type of facilities they prefer, so that optimized investment decisions can be made.

Menghini et al. (2010) did a similar study on bicyclists in Zurich – however, they did not directly conduct the data collection for the study but rather received a multimodal travel dataset from a private agency in Zurich. The unique contribution of this study is in developing a GPS data cleaning algorithm for large datasets without any other information. The choice set was generated under MATSim (Multi-agent Transportation Simulation), using a search and bound algorithm, which generated about 60 alternatives to each route. The route choice model selected is the multinomial logit model with path choice overlap correction as in the study by Broach et al. (2009). The parameters used for estimation are maximum and average gradient, length of trip, percentage of marked bike facility, and number of traffic lights. Length was found to be the single most important criterion followed by average gradient and percentage of marked bike facility. Number of traffic lights and maximum gradient did not have any significant impact on route choice.

Limitations of the Studies

While the study by Broach et al. (2009) was one of the first to use revealed preference data for creating a route choice model for cyclists, a few limitations still remain that require further research. For instance, the study was based in Portland, a very bike friendly city with more bike facilities than can be expected in an average U.S. city. Therefore, it is difficult to translate the findings regarding preference of bike facility and the willingness to travel an extra distance to avail a facility to other regions. Second, the GPS data cleaning was done manually, which was possible because of a low participant base and a low number of trips (1559 trips) used in the study. For larger datasets, it will not be possible to clean GPS data manually, and some algorithms and scripts will be necessary. But the most important issue with the studies by Dill et al. (2008) and Broach et al. (2009) is similar to that noted in the study by Hall et al. (1997) - all these studies lack segmentation of the cyclist population based on experience, comfort level, or attitude, although the literature has always emphasized the impact of rider characteristics on route choice decisions (Pucher and Buehler 2007, Krizek 2007). During data collection, cyclists were categorized as frequent and infrequent cyclists but no separate analysis was performed, possibly because the number of infrequent cyclists was very low, but this defeats one of the major purposes of modeling as mentioned earlier.

Menghini et al. (2010) acknowledge this limitation of their study and suggest using socio-demographic characteristics like age, gender, and a measure of risk aversion of the riders to overcome the issue. As the remaining part of the literature critique will show, although age and gender have been used in one of the models, a measure of risk aversion and the level of experience of the cyclist still remain to be modeled into route choice studies.

The GPS Enabled Smartphone Based Data Collection Effort

The data collection effort of this group was based off a free smartphone app called CycleTracks that was created and developed at San Francisco County Transportation Authority (SFCTA). A user has to only download the app and turn it on at the start of the trip - the app then records route GPS data for every second of travel and stores it locally. On completion of the trip, the user is given the option to upload the trip or discard it. Upon uploading, the trip data are stored in a central server before that can be used for analysis (Charlton et al. 2009). As the goal of the study was to relate cyclist route choice to personal, trip-based, and network characteristic-based factors, the app comes with an optional provision where participants can provide their age, gender, cycling frequency, and the purpose of the trip. There is also a provision to enter the participant's email address, should he/she choose to do so. For maintaining anonymity of data, this field is completely scrubbed off during data analysis and is only stored for the purpose of future correspondence or survey needs.

The purpose of the CycleTracks research was to develop a bicycle route choice module for the existing tour-based travel demand model SF-CHAMP for San Francisco and the Bay Area. The data collected through CycleTracks were used to develop a multinomial logit model for cyclist path choice from which logsums are fed into SF-CHAMP, enabling it to assign the generated trips to the city road network. Hood et al. (2009) used the data cleaning and map matching algorithm developed at ETH by Dr. Kay Auxhausen and Dr. Nadine Schussler

(Schussler and Auxhausen 2009) and used the same multinomial logit model with correction for path size overlap as was done by Broach et al. (2009) and Menghini et al. (2010). The contributions of this study are in using a different algorithm for the choice set generation, including gender and cycling frequency as model parameters, using panel data for model validation, and in extending the modeling exercise into a benefit cost analysis of possible new facility construction. The choice set generation algorithm used is a doubly stochastic genetic search algorithm that generates the choice set through randomizing both the link attributes and the beta coefficients of the cost function. The parameters that were estimated in the model are length, number of turns per km, upslope, type of bike facility, and travel in the wrong direction. The results suggest that cyclists prefer shorter routes and fewer turns, whereas upslope is always a disutility. However, bike lanes were found to be preferred over shared use lanes, and infrequent bicyclists were more likely to prefer a bike lane than shared lane. Slopes were particularly avoided by female cyclists and during a commute trip. A holdback sample of 202 cyclists was used to validate the results of the model.

The CycleTracks app was adopted for a similar study in Austin, Texas by Hudson et al. (2012). The data collected using the app were used to develop a cycling route choice map for the region, but was not extended to modeling route preferences.

Limitations of the Studies

Although the study by Hood et al. (2009) addressed most of the issues discussed previously, it still falls short of including a risk aversion attitude measure into the route model. Another issue in this study was the disproportionate share of trips recorded by users. To solve the problem, in the log likelihood function, each observation was weighted by the inverse of the number of observations for an individual to have equal weight for all observations. However, this raises the question of if a cyclist who used a route once should have equal importance in model estimation as someone who uses the route regularly. Further research is needed to ascertain a weighting function that answers such questions. Also, the benefit-cost analysis done in this study is based on a national cost estimate and calculated the user benefit or willingness to pay only based on value of time while several other factors like health benefits and environmental benefits are more important for bicyclists. Therefore, further research is needed in developing a benefit cost analysis framework for cycling facilities that include all relevant factors.

A particular study, though not based on revealed preference data, deserves mention because of its use of a different model for bike route choice modeling and for being the only study having on-street parking as a route attribute in the proposed model. Sener et al. (2008) used a web based stated preference survey to collect data on cyclist route choice in Texas and used a panel mixed multinomial logit model to estimate route choice parameters. Of the chosen attributes, they found that cyclists were particularly sensitive to travel time and preferred streets with no on-street parking. Moderate hills were preferred over flat terrain and no significant influence of any other cyclist characteristic (experience, for example) was noted except for cyclist age as interaction between parking and cyclist age.

ANALYSIS AND RESULTS

For the purpose of this research, the alternative to the chosen route was taken to be the shortest route path generated by the A-star algorithm. The binary logistic choice that was modelled was whether the riders chose the shorter of the two routes depending on their age, gender, and what type of rider they are. Additional regression models were also constructed to understand (i) the relationship between trip length and rider characteristics and (ii) the percent deviation of the chosen route from the predicted shortest route based on rider characteristics.

Three data sources were used to create the road network map. The Atlanta Regional Commission's street network shapefile (RC_ROUTES) was obtained from the travel demand modeling group of Atlanta Regional Commission (ARC). It is a modified version of the roadway database maintained by the Georgia Department of Transportation (GDOT) and focuses on state managed roadways rather than locally managed roadways and bikeways. However, it contains the most comprehensive inventory of roadway characteristics like speed limit, annual average daily traffic (AADT), number of lanes, truck volume, etc. which are useful information for route choice modeling at a later stage. The second data source used was Open Street Map's (OSM) bicycle map for Atlanta. The OSM map has local roads and locally managed facilities which were not present in the RC_Routes map. The two maps were spatially joined based on a buffer distance to get a more complete map of the road network of Atlanta. The resulting map was then cleaned for non-bicycling facilities like freeways. The final data source was the Metro Atlanta Bicycle Facility Inventory. The location of on street parking on roadways with conventional bicycle lanes and buffered bicycle lanes was manually coded in ArcGIS using Google Earth imagery. The treatment of intersection approaches with right turn only motor vehicle lanes that connect to links with conventional bicycle lanes, buffered bicycle lanes, or protected cycle tracks were also manually coded in ArcGIS using Google Earth Imagery. As a final measure, the trips were plotted on the map and checked for links traversed by cyclists but missing in the network. Such links were manually added where more than 2 bicycle trips were found to follow a path, but the path was not marked as a link in the network. This was assumed to be mainly because of tendency and ability of bicyclists to use cut-thrus and private alleys which are not marked in regional network maps. However, shortcuts through parking lots were not added as links although there were multiple such cases.

Figure 16(a) shows the number of trips recorded by each rider. Figure 16 (b) shows the trips by purpose. Figure 16(c) shows the trip purpose across age – riders in the age group > 45 years use cycling for exercise than any other group. Figure 16(d) shows trip purpose by gender, and since the data are heavily dominated by male cyclists, they are the dominating group in all trip purpose categories. However, female riders have almost a similar share of shopping trips in spite of being a small fraction of the riders. This calls for particular consideration in land use planning to allow women to do trip chaining comfortably and easily. Figure 16(e) shows trip purpose by rider type. The strong and fearless riders make more social and shopping trips by cycling than other types of riders, while enthused and confident riders using cycling for commute more than any other rider type. Comfortable but cautious riders use cycling for exercise more than other rider types. Figure 16(f) shows the frequency distribution of trip length. The mean trip length, marked by dashed red line was found to be about 3.75 miles (about 5.5 Kms). The majority of the trip lengths were within 4-6 miles which is a standard commute distance. Figure 16(g) shows the trip length by age. It should be noted that the highest frequency of trips for younger riders are at a shorter distance than that of senior riders which is initially counter intuitive. However, one of the reasons may be that senior riders are less likely to choose shorter routes if that does not provide sufficient safety and comfort while younger riders may prefer shorter distance to a detour for a bike facility. The younger riders are also dominated by college students, and their commutes may be much shorter in length. Figure 16(h) shows trip length across gender, and we see a similar trend as age here – the highest frequency of trip lengths for women are longer than that of men. Figure 16(i) shows trip length across rider type, and enthused and confident riders are seen to have shorter trips than comfortable but cautious riders. Strong and fearless riders have slightly longer mean trip length than enthused and confident riders, but that may be because they bicycle longer distances.

For the purpose of this study, we considered only the “primary” trip of each user and therefore restricted the analysis to work or school trips (trip purpose = “work”/ “school”), thus reducing the number of trips to be considered from about 20,000 to about 12,000. The trips were further restricted to be greater than 1 mile and less than 8 miles, resulting in about 10,000 trips.

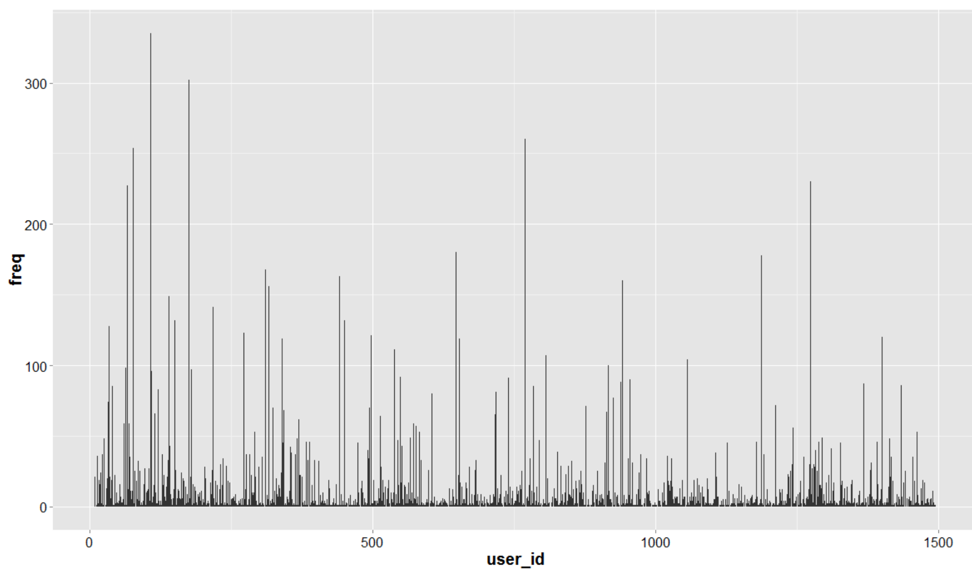


Figure 16(a) Number of Trips Recorded by Users

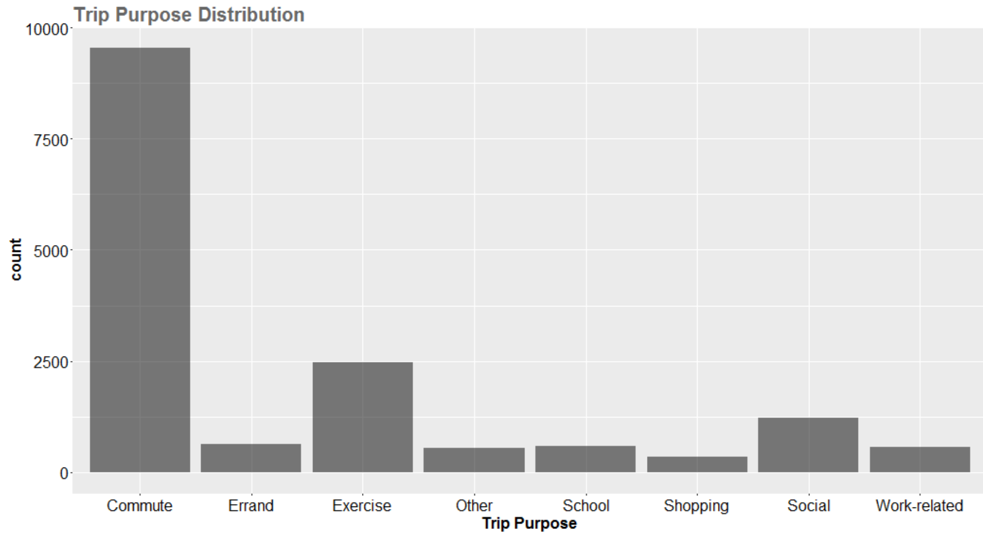


Figure 16(b) Trip Purpose Distribution

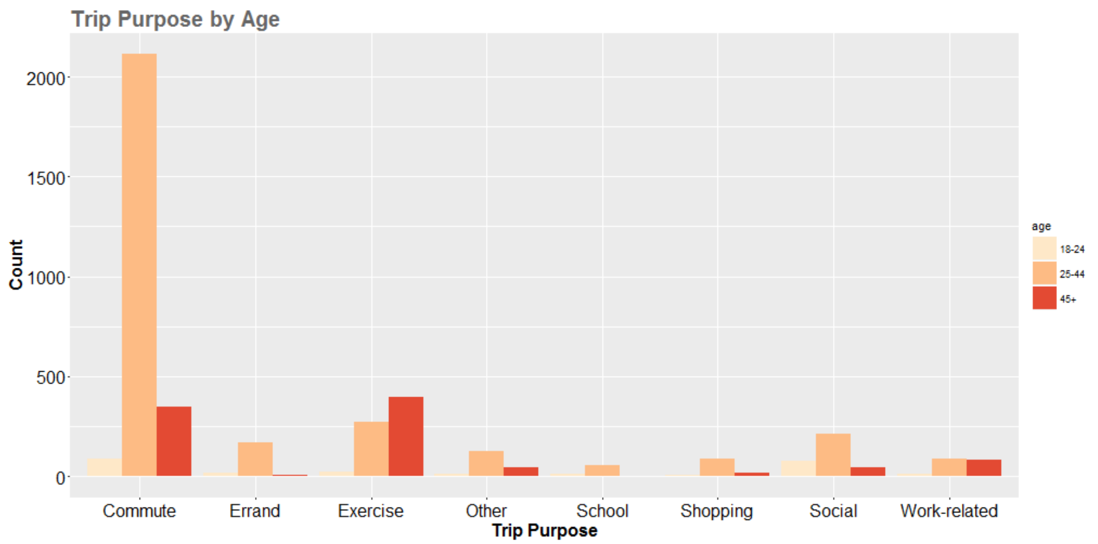


Figure 16(c) Trip Purpose Distribution across Age

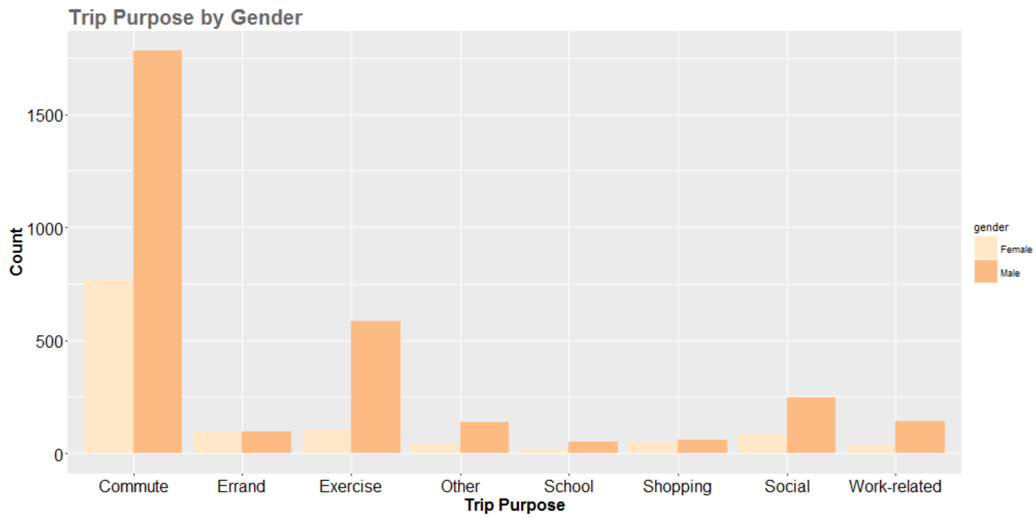


Figure 16(d) Trip Purpose Distribution across Gender

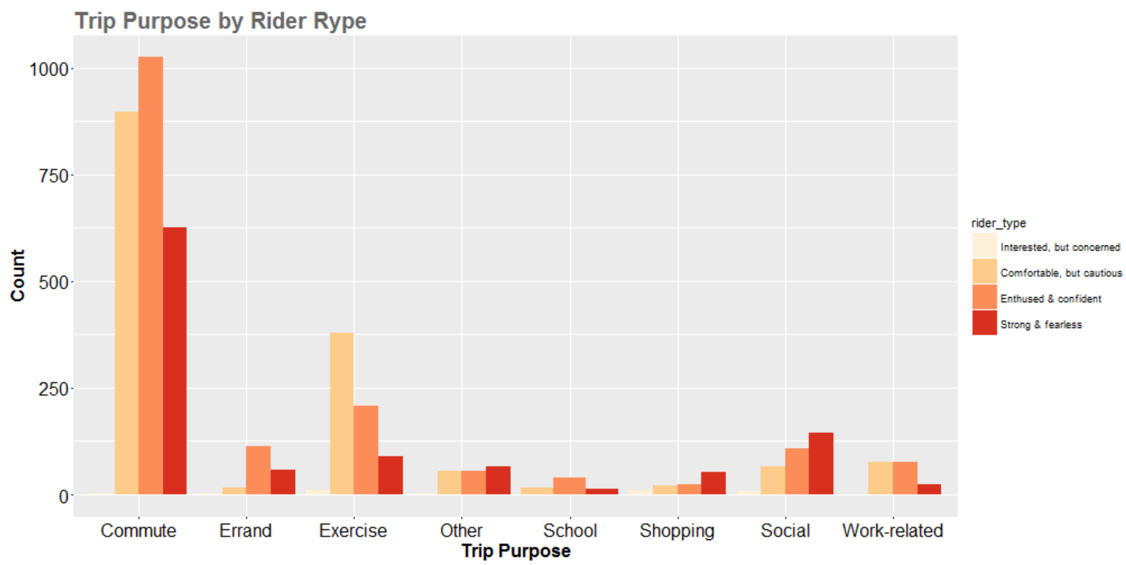


Figure 16(e) Trip Purpose Distribution across Rider Type

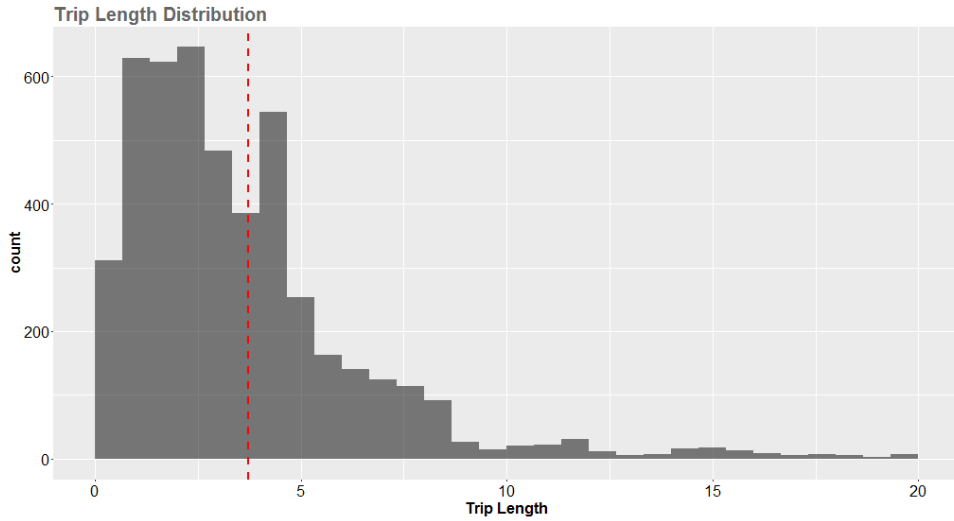


Figure 16(f) Trip Length Distribution

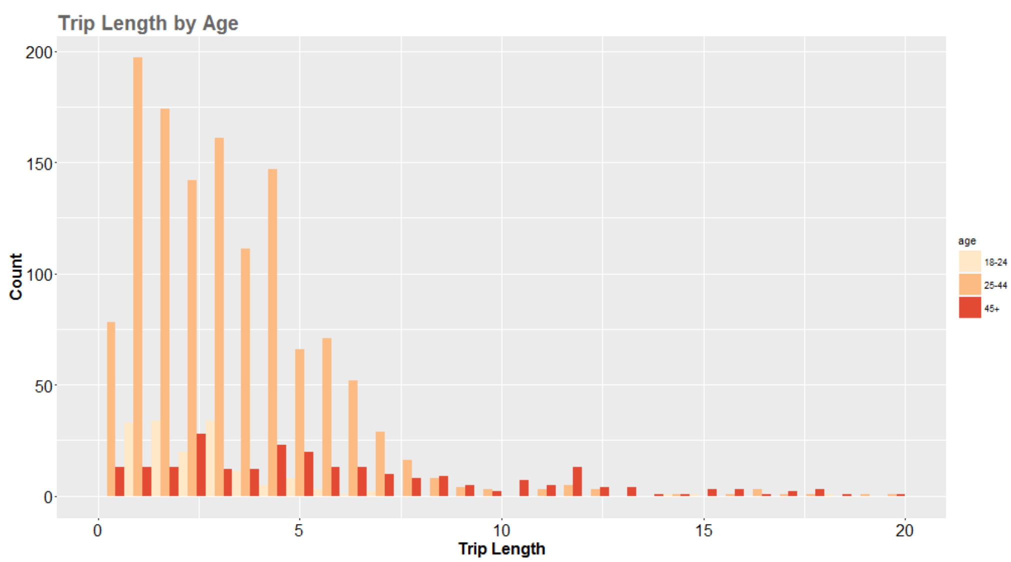


Figure 16(g) Trip Length across Age

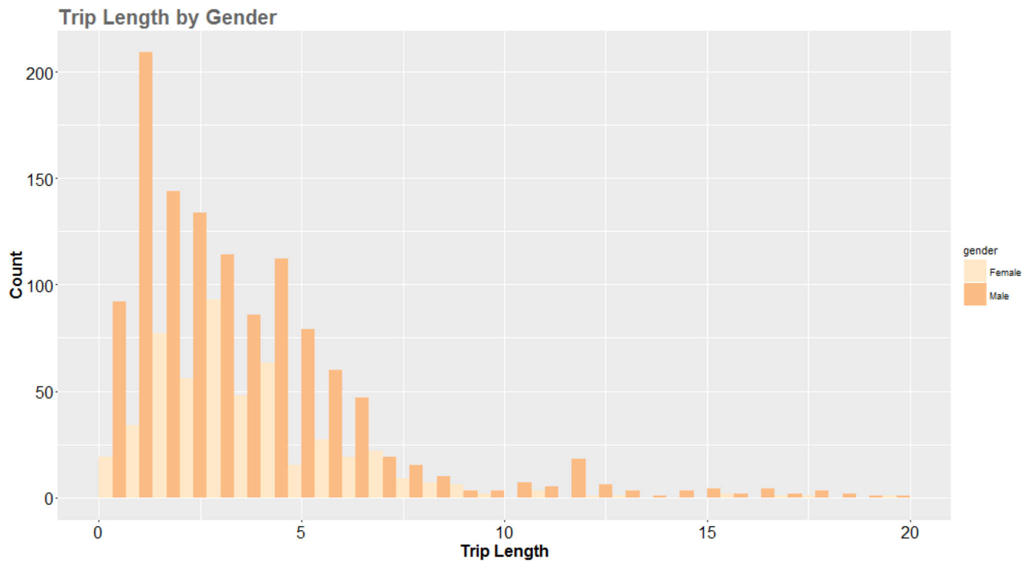


Figure 16(h) Trip Length across Gender

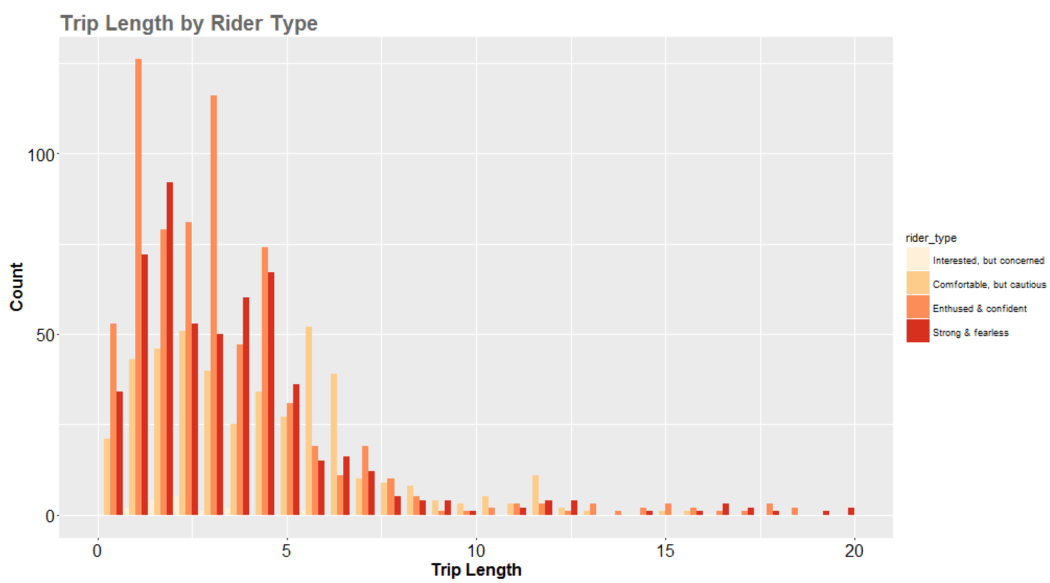


Figure 16(i) Trip Length across Rider Type

Figure 16. Cycle Atlanta Trips (a) Number of Trips Recorded by Users (b) Trip Distribution by Purpose (c) Trip Purpose Distribution across Age (d) Trip Purpose Distribution across Gender (e) Trip Purpose Distribution across Rider Type (f) Trip length Distribution (g) Trip Length across Age (h) Trip Length across Gender (i) Trip Length across Rider Type

The trip data were then linked to the sociodemographic data via the user id key. As with the models in Chapter 4, the *comfortable but cautious* group and the *interested but concerned* group were merged into one group. One random trip was then chosen per user to perform the shortest route analysis. The first model estimated is a linear regression model to understand the relationship between trip length and age, gender, and rider types. For all the models, the *comfortable but cautious & interested but concerned* group was chosen as the base group as was age 18-24 implying that all results should be interpreted in a comparison to that category. Table 10 presents the results of the regression model on trip length as function of sociodemographic characteristics of the riders. Age has a positive relationship with trip length and male riders are also more likely to ride longer distances. *Enthused and confident* riders are less likely to take longer trips than *comfortable but cautious* riders, but *strong and fearless* riders are more likely to take longer trips. This may be because *enthused and confident* riders are more inclined to use shortest routes even if there are no bicycle facilities which renders their trip short compared to *comfortable but cautious* riders. On the other hand, *strong and fearless* riders are more likely to naturally undertake longer trips than any other categories.

Table 10. Trip Length as Function of Socio-demographic Characteristics

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	2.4334	0.17233	14.12	< 2e-16	***
age25-44	0.96803	0.15798	6.128	9.67E-10	***
age45+	2.37601	0.1906	12.466	< 2e-16	***
genderMale	0.10798	0.10187	1.06	0.289	
Enthused and confident	-0.52289	0.08236	-6.349	2.38E-10	***
Strong and fearless	0.04501	0.07	0.643	0.52	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

The second model was used to understand the relationship between deviations from the network based on the shortest route depending on their socio-demographic characteristics. For the majority of the trips, the network based shortest route is shorter than the actual trip length, therefore this model may serve as a proxy to understand if any rider group is systematically choosing a longer route possibly because of factors not yet known to us. Table 11 presents the results of the model estimate. Gender is only significant in this model and male riders are less likely to deviate from shortest route as compared to female riders. Similarly, *enthused and confident* riders and *strong and fearless* riders are also less likely to choose longer routes over shortest routes as compared to *comfortable but cautious* riders, with *strong and fearless* riders more likely to choose shortest routes than *enthused and confident* riders. People in the age group >45 are less likely to choose the shortest route than riders in the age group of 18-24 while people in the age group of 25-44 are more likely.

Table 11. Deviation from Network based Shortest Route as Function of Socio-Demographic Characteristics

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	6.7953	8.9253	0.761	0.4465	
age25-44	-0.1981	8.1843	-0.024	0.9807	
age45+	2.3757	9.8707	0.241	0.8098	
genderMale	-9.3924	5.273	-1.781	0.0749	.
Enthused and confident	-2.4476	4.2633	-0.574	0.5659	
Strong and fearless	-5.2741	3.6233	-1.456	0.1456	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Finally, a binary logistic choice model was used to understand whether a rider chose a shorter of the two available alternatives depending on the socio-demographic characteristics. The model estimates are presented in Table 12(a).

Table 12(a). Choice of Shorter Route Based on Socio-demographic Characteristics

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	0.618	0.0303	20.407	< 2e-16	***
age25-44	-0.072	0.0278	-2.593	0.00956	**
age45+	-0.150	0.0335	-4.486	7.42E-06	***
genderMale	-0.012	0.0179	-0.681	0.4956	
Enthused and confident	0.058	0.0145	3.989	6.73E-05	***
Strong and fearless	0.007	0.0123	0.595	0.55207	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

The results show that senior riders are more likely to not choose shortest route while *enthused and confident* riders as well as *strong and fearless* riders are more likely to choose shortest routes. However, male riders are also more likely not to choose shortest routes, which is counterintuitive. This may be because in general, male riders undertake longer trips, and hence, there is not much difference from the shortest route and the actual route. Another model was estimated by including trip distance to see if distance is a reason for the counterintuitive sign of this model. The results are presented in Table 12(b). With the introduction of trip length, age loses its significance indicating that trip length is related to age of a rider. However, trip length is significant and has a negative sign indicating that longer the trip is, riders are less likely to choose shortest routes possibly because either the difference is not significant enough or because longer trips require being comfortable for a longer time and people are more likely to choose facilities that maximize that perceived comfort.

Table 12(b). Choice of Shorter Route Based on Socio-demographic Characteristics and Trip Distance

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	0.7549	0.0293	25.7620	<2e-16	***
age25-44	-0.0179	0.0264	-0.6770	0.4986	
age45+	-0.0171	0.0322	-0.5300	0.5959	
genderMale	-0.0061	0.0169	-0.3630	0.7169	
Enthused and confident	0.0283	0.0138	2.0590	0.0396	*
Strong and fearless	0.0098	0.0116	0.8430	0.3995	
Trip length	-0.0562	0.0025	-22.7850	<2e-16	***

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

CONCLUSION

In this chapter, we analyzed the Cycle Atlanta trips based on the socio-demographic characteristics of its users. We also estimated the likelihood of a cyclist to choose a longer route over a network based shortest route depending on his/her socio-demographic makeup and confidence level. The results show that female cyclists are more likely to make trips that are shorter than their male counterparts, have almost the same number of shopping trips as male riders, and are more likely to take longer routes than the network based shortest routes. This indicates that instead of classifying cyclists as new or inexperienced cyclists and constructing infrastructure accordingly, it will be more effective if the gender distribution of the locality is taken into consideration when planning for cycling infrastructure. Since the only trip purpose that has more female cyclists than male is shopping, it may require a serious reconsideration of land use planning to have more female cyclists and more trips from existing female cyclists. Distance is not the only factor that determines route choice for bicyclists, and future research will look into incorporating other factors like traffic stress, presence of facilities, scenery, and slope into the route choice model of cyclists in Atlanta.

CHAPTER 7. LINK BASED QUALITY-OF-SERVICE MEASURE USING BICYCLIST PERCEIVED LEVEL OF TRAFFIC STRESS

INTRODUCTION

Levels of bicycling in the United States remain low compared to its international peers. However, cities throughout the country are interested in attracting new riders since bicycling is a healthy and environmentally friendly mode of transportation that has the potential to provide more equitable transportation options for all sections of society (Buehler and Pucher 2008). Multiple studies have been conducted to understand the barriers to bicycling. The findings indicate that topography (Buehler and Pucher 2008, Dill and Gliebe 2008, Winters et al. 2010, Krenn et al. 2014, Broach et al. 2012), weather (Buehler and Pucher 2008, Godefroy and Morency 2012, Parkin et al. 2007, Nankervis 1999, Brandenburg et al. 2004, Miranda-Moreno and Nosal 2011), city size and population density (Buehler and Pucher 2008, Godefroy and Morency 2012, Pucher et al. 1999, Cervero and Duncan 2003, Winters et al. 2010), income and sociodemographics (Buehler and Pucher 2008, Godefroy amd Morency 2012, Pucher 1999, Dill and Carr 2003, Smart et al. 2014), and relative cost of motor vehicles and public transit (Pucher et al. 1999) can influence the decision to bicycle. After a review of the literature, it was determined that this study would focus on traffic, roadway, and bikeway characteristics since they have been found to have a significant influence on the decision to bicycle based on their effect on the perceived safety of a facility (Buehler and Pucher 2008, Pucher et al. 2010, Winters et al. 2011, Misra et al. 2014, Urban et al. 2014).

Bicycling mode share can be increased by building well connected bicycle facilities, which address safety concerns and appeal to the majority of the public. Research has shown that Americans have a range of tolerances for perceived traffic stress with the majority of the population tolerating only low levels of traffic stress such as low traffic volumes and motor vehicle speeds while bicycling (Winters et al. 2010, Dill and McNeil 2013). To plan for expansion of the bicycle network, it is important for cities to know how existing facilities are performing and set up measures of performance such as quality-of-service measures. Existing bicycle quality-of-service measures include the Highway Capacity Manual Bicycle Level of Service (HCM BLOS), Bicycle Compatibility Index (BCI), Bicycle Environmental Quality Index (BEQI), and the Mineta Transportation Institute Level of Stress (MTI LTS) (HCM 2010, Harkey et al. 1998, San Francisco Department of Public Health, n.d., Mekuria et al. 2012).

The above mentioned existing bicycle quality-of-service measures often require data, which are not readily available and which can only be obtained through extensive and costly field research. In addition, all of these measures except for the MTI LTS lack the ability to measure the perceived stress related to a facility, a measure which is critically important to bicycle mode choice. The MTI LTS was chosen by this study as a quality-of-service model to build upon due to its many strengths including requiring more easily accessible data, being more intuitively understandable to the public, consideration of both current and potential bicyclists through the use of Geller’s four types of bicyclists, and ability to analyze innovative bicycle facilities such as protected cycle tracks.

The objective of this study is to modify the Mineta Transportation Institute’s LTS measure using traffic and roadway characteristics data that are available to most planning and engineering agencies and also take into account the perceived traffic stress that bicyclists associate with a facility. The modified LTS seeks to build upon the MTI LTS by updating the bicyclist typology used to the Cycle Atlanta typology, grounding the criteria used to measure the perceived stress related to a facility in the literature, and refining the required data to more easily accessible data while maintaining strength of analysis.

The purpose of the modified LTS is to build a quality-of-service measure which will create a standardized system of designating streets into different categories using data that are readily available to transportation professionals. Such a standardized quality-of-service measure will allow planners and engineers to assess the suitability of different bicycle facility options and assist in project selection. The proposed measure was used in a case study area around the Atlanta Eastside Beltline Trail and was found to be useful in determining the perceived stress of different roadway and bikeway facilities for current and potential bicyclists.

LITERATURE REVIEW/BACKGROUND

The HCM BLOS measures the quality-of-service experienced by bicyclists by stratifying multiple performance measures to determine levels of service ranging from Level-of-service (LOS) F as the worst condition to LOS A as the best operating condition (HCM, 2010). BLOS is calculated with the use of a linear function with weights assigned to independent variables and produces a numerical score ranging from 0 to 6 with the numerical score relating to the LOS A - F grade as follows: A 2.00, 2.00 < B 2.75, 2.75 < C 3.50, 3.50 < D 4.25, 4.25 < E 5.00, and F > 5.00 [27]. The HCM BLOS model was developed by showing videos of various bicycle facilities to participants who were asked to rank how satisfied they were with the bicycle facilities on a six-point scale ranging from “very dissatisfied” to “very satisfied” (Parks et al. 2013).

The HCM BLOS model considers width of outside lane, width of bike lane, width of shoulder, proportion of occupied on-street parking, vehicle traffic volume, vehicle speeds, percentage of heavy vehicles, pavement condition, and number of through lanes (HCM 2010).

HCM BLOS strengths include its basis in bicyclists' perception of facility characteristics, its focus on facility design that is directly under the influence of operating agencies, and that it is directly measurable in the field. However, HCM BLOS has multiple weaknesses: lack of transparency for the public and decision makers, a focus on arterial roadways over local roadways, lack of sensitivity to driveway type, lack of consideration of innovative bicycle facilities, intersection LOS that requires further refinement, acceptance of wide outside motor vehicle lanes, and limited validation with surveys (Parks et al. 2013). While HCM BLOS A – F is familiar to transportation professionals, it is not well understood by the public and decision makers, which limits the quality-of-service measure's ability to assist in the selection of the appropriate bicycle facility for a specific location. In addition, HCM BLOS focuses on assessing arterials and collectors and does not consider local roads, which research has shown are preferred by cyclists, even if the distance was up to 10% longer (Winters 2011). By focusing only on BLOS for arterials and collectors, there is a potential to bias the allocation of new bicycle facilities to arterials and collectors with a lack of consideration for how local streets fit into the network.

Another bicycle quality-of-service measure is the Bicycle Compatibility Index (BCI), which predates HCM BLOS. Bicyclists' perception of comfort is rated like the HCM LOS A – F with A representing the most comfortable and F the least comfortable. The regression model used by BCI includes the following significant variables: number of lanes, directions of travel, curve lane, bicycle lane, paved shoulder, parking lane width, gutter pan width, traffic volume, speed limit, 85^{th} percentile speed, driveway density, presence and type of sidewalks, presence and type of medians, and type of roadside development. Adjustment factors are also developed for the presence of large trucks or buses, vehicles turning right into driveway, and vehicles pulling into or out of on-street parking spaces. The three primary weaknesses of the BCI LOS are reliance on skill level typology, reliance on data which must be gathered through field analysis, and inability to analyze segments with varying geometric and operation characteristics.

A third bicycle quality-of-service measure is the BEQI model which was developed by the San Francisco Department of Public Health's Environmental Section through a survey of transportation professionals and members of the bicycling community. Survey participants were asked to weigh the most important variables that affected their perception of bicycle facility quality. Based on the responses, the variables, "were combined in an index that ranged from 0 to 100" (San Francisco Department of Public Health, n.d.). The BEQI tool includes 22 variables, however, the factors with the highest weight in the BEQI tool are bicycle facility type, bicycle facility width, pavement type, pavement condition, slope, pavement markings, connectivity, driveway cuts, and presence of trees (San Francisco Department of Public Health, n.d.). The primary strength of the BEQI tool is the software that is publicly available to execute the tool. The weaknesses of the BEQI tool include the tool being difficult to implement outside of San Francisco, requiring a high number of variables, requiring data that needs to be gathered manually, and being developed with the input of transportation professionals and current bicyclists with no input from potential bicyclists.

The quality-of-service measure that is most relevant to this research is the MTI LTS. The Mineta Transportation Institute study classified roadways and bikeways into four levels of traffic stress according to a modified version of Geller's four types of bicyclists. LTS 1 included facilities suitable for children; LTS 2 facilities characteristics were based on the Dutch CROW (Center for Research and Contract Standardization in Civil and Traffic Engineering (Netherlands)) Design Guide and were intended to be comfortable for most adults; and LTS 3 and LTS 4 present tolerance for characteristics of higher stress (Mekuria et al. 2012). LTS criteria were developed for the following facility types: physically separated bikeways, bike lanes, and shared travel lanes. LTS criteria were developed for right-turn only motor vehicle lanes and unsignalized intersections also. High stress roadways at unsignalized intersections and limited access roadways were identified as the main barriers to low stress bicycling.

The MTI LTS takes into consideration the following variables; number of through lanes, bicycle facilities, posted speed, width of bike lane, width of parking lane, bike lane blockage, right turn lane geometric information, on street parking (alongside bicycle facilities), signalized intersections, and median (Mekuria 2012). The two main strengths of the MTI LTS are being more intuitively understandable to the public and decision makers and considering both current and potential bicyclists. MTI LTS has already been deployed in numerous bicycle and pedestrian plans. The MTI LTS requires the most readily available data out of the quality-of-service models discussed here. Requiring easily accessible data makes the analysis of roadways and bikeways much easier for jurisdictions. Unlike other quality-of-service tools, the MTI LTS categorizes facilities based on the preferences of the entire adult population who currently bike and who would consider biking.

The MTI LTS has two primary weaknesses: data that requires manual collection and lack of research used to validate traffic and roadway characteristics that affect perceived stress. Another weakness is the approximation of bike lane blockage by assuming that bike lane blockage is frequent in commercial areas and rare in all other areas when it is unknown how effective this method is for approximating bicycle lane blockage by motor vehicles (Mekuria et al. 2012). Manual data collection is required to measure bicycle lane and parking lane width since most jurisdictions do not collect these data. Manual data collection can be very time consuming and may not be feasible. The majority of the criteria used to classify roadways and bikeways by LTS level were based on Dutch bicycle design criteria and not through research measuring the perceived stress or comfort of roadway, bikeway, and traffic characteristics for U.S. current and potential bicyclists.

As the literature review illustrates, existing bicycle quality-of-service measures often require data that are labor intensive and costly to obtain, lack transparency and are difficult for the public and decision makers to read, and are unable to analyze innovative bicycle facilities such as protected cycle tracks. To help agencies and decision makers have access to a quality-of-service tool that is easily understood and not data intensive, yet effective, this study proposes a modified quality-of-service measure which can be easily implemented throughout the United States.

The modified LTS is built based on the concept that facilities may be associated with different levels of perceived safety, and the perception depends on the type of bicyclist and his/her tolerance level for traffic stress. There have been several studies that have classified bicyclists into different categories based on their skill level (Dill and McNeil 2013, AASHTO 2012) and bicycling frequency (Winters et al. 2011, Dill and Voros 2007, Sanders 2013, Ahmed et al. 2013). However, this study uses the bicyclist classification introduced by Roger Geller (Geller 2006) and later modified by Misra et al. (2014).

Geller (2006) categorized current and potential bicyclists of Portland by their level of comfort riding on different types of roadway and bikeway facilities. The four bicyclist types suggested by Geller are (i) Strong and Fearless (less than one percent of bicyclists), (ii) Enthused and Confident (seven percent), (iii) Interested but Concerned (60 percent), and (iv) No Way No

How (33 percent).The Cycle Atlanta typology is a modified version of the Geller typology, in which the No Way No How type was dropped, because the typology includes only descriptions of those who are currently bicycling or who are interested in bicycling. In addition, the Interested but Concerned type used in the Geller typology was split into two types with Comfortable but Cautious category intended to include bicyclists such as females and/or older travelers who are bicycle enthusiasts, but may be more risk adverse (Misra et al. 2014). See Table 13 for descriptions of all four Cycle Atlanta types. People who identify as LTS 2 Comfortable but Cautious are estimated to be the largest type present in the population and will not bike on shared roadways with high motor vehicle speeds and traffic volume, will only bike on roadways with low speeds and low traffic volumes like local or neighborhood roads, and prefer to bike on bicycle or shared-use paths. The Cycle Atlanta typology is used in this research as the basis for the modified LTS roadway and bikeway criteria which are discussed in more detail later.

Table 13. Cycle Atlanta LTS Typology

LTS Type		Description
LTS 1	Interested, but concerned	I have heard a lot about cycling and I am curious to try it, but I require facilities geared to cyclists before I would do so
LTS 2	Comfortable but cautious	I am comfortable on most roads, but strongly prefer facilities geared to cyclists and will choose another mode depending on facilities
LTS 3	Enthused and confident	I am confident sharing the road with vehicles but prefer facilities geared to cyclists
LTS 4	Strong and fearless	I am willing to bike in any situation and being a cyclist is part of my identity

MODIFIED LTS MEASURE

The modified LTS quality-of-service measure builds upon the MTI LTS and classifies roadways and bikeways by one of four levels of traffic stress based on traffic and geometric characteristics such as traffic volume, posted speed limit, number through lanes per direction, presence of on street parking, and bicycle facility type. Roadways and bikeways categorized at LTS 1 are the least stressful and have low traffic volumes and low speed limits, while roadways and bikeways categorized as LTS 4 are the most stressful and have the highest traffic volumes and speed limits. It is estimated that the majority of current and potential bicyclists find LTS 1 and LTS 2 facilities comfortable. Table 14 provides a description of the characteristics of roadways and bikeways for each LTS. This table is a modified version of a similar table used by the Mineta Transportation Institute to describe the roadway and traffic characteristics of its LTS measure. MTI LTS classifies protected shared paths, cycle tracks, and side paths as LTS 1; however, the modified LTS re-classified protected cycle tracks and side paths as LTS 2 due to the increased presence of conflict zones such as driveways and intersections for these facilities as opposed to the presence of few conflict zones for most shared paths. MTI LTS considered LTS 1 facilities suitable for children; however, the modified LTS does not make assessments for children since there is very limited research on perceived stress for children. The modified LTS also introduced buffered bicycle lanes as a facility type since this facility type was not considered by the MTI LTS.

Table 14. LTS Roadway and Bikeway Characteristics

LTS Level	Modified LTS Roadway and Bikeway Descriptions
LTS 1	Considered comfortable and low stress by almost all cyclists. Includes shared paths which separate cyclists from motor vehicle traffic and present few conflict zones such as intersections and driveways. Shared travel lanes are only tolerable if traffic volume is so low that cyclists only occasionally interact with motor vehicles and there is little difference in travel speed between cyclists and motor vehicles due to a posted speed limit of 25 mph or below. Intersections are low stress to approach and cross.
LTS 2	Considered low stress by all cyclists except for people who identify as LTS 1. Includes side paths and protected cycle tracks which are low stress, but present some conflict zones at driveways and intersections. Shared travel lanes can only have one lane per direction, a speed limit of 30 mph or below, and must be classified as local. Conventional bike lanes and buffered bike lanes allow for slightly higher traffic volume, speed, and classification as local or collector.
LTS 3	Conventional bike lanes or buffered bike lanes are located on roadways with moderate traffic volume and speed and can be classified as minor arterial or lower. Shared travel lanes must be classified as collector or lower and 35 mph or lower. Roadways of LTS 3 can have 2 lanes or less per direction.
LTS 4	Any level of stress beyond LTS 3 excluding limited access roadways. Includes all roadways with a posted speed limit above 40 mph and/or 3 or more lanes per direction with or without bicycle lanes.

Criteria Used for Calculating Level of Traffic Stress

The details of traffic stress classification for separated bicycle facilities are presented below. The criteria tables for shared travel lanes and on-road bicycle facilities are also given. Note that criteria tables follow the rule that the aspect of a link with the highest LTS determines the LTS of that segment. For example, a conventional bicycle lane with no adjacent motor vehicle parking (see Table 15) with one through lane per direction (LTS 1), a posted speed of 35 mph (LTS 3), a functional class of collector (LTS 2), and a traffic volume of 10,000 vehicles per day (LTS 2) would be classified as LTS 3 for the link as a whole. The notation "(no effect)" means that the factor does not cause an increase to that LTS. Table 15 through Table 19 provide the criteria used in developing the proposed LTS.

Criteria for Separated Bicycle Facilities

Research has shown that people prefer separated bicycle infrastructure (Winters et al. 2010, Kremm et al. 2014, Broach et al. 2012, Misra et al. 2014, Dill and Voros 2007, Monsesre et al. 2014). MTI LTS classified all separated bicycle facilities (shared-use paths, side paths, and protected cycle tracks) as LTS 1. However, this method does not consider the potential stress of bicycle and motor vehicle interaction at driveways, intersections, and loading areas. Therefore, in this study, separated bicycle facilities or shared-use paths, which are the most separated from motor vehicle traffic, are classified as LTS 1. Protected bicycle facilities such as side paths, one and two way cycle tracks, and raised cycle tracks are classified as LTS 2 due to the potential interaction of motor vehicles and bicycles at midblock driveways, intersections, and loading bays.

Traffic, Roadway, and On-Road Bikeway Characteristics

The roadway and traffic characteristics which are considered include: number of through lanes per direction, traffic volume or annual average daily traffic (AADT), functional class, and posted speed limit. The focus on traffic volume and speed is supported by Winters' survey of current and potential bicyclists in Metro Vancouver. This study found that high traffic volume and traffic speed were major deterrents from riding (Winters et al. 2011). Thus, for conventional bicycle lanes, buffered bicycle lanes, and shared travel lanes, the level of traffic stress for a link increases as those variables increase. The perceived stress caused by the presence of or lack of on street motor vehicle parking was also considered.

Traffic Volume or Annual Average Daily Traffic (AADT) and Functional Class

MTI LTS does not include traffic volume or functional class when classifying facilities. However, research has shown that the majority of people who want to bicycle more list “too much traffic” as the top environmental barrier (Dill and Voros 2007). Therefore, traffic volume and functional class were included in this study. Number of travel lanes and functional class have a strong relationship, as the USDOT FHWA Highway Functional Classification Concepts, Criteria, and Procedures states, “roadways are designed and constructed according to their expected function” (USDOT 2013). For example, an arterial is designed to be a high capacity roadway and would likely have more travel lanes, while a collector would likely have less travel lanes than an arterial and a local road even less travel lanes than a collector. Research by Winters et al. (2010) also found that when comparing shortest route to actual route, bicyclists traveled significantly less along arterial roads than predicted by the shortest route model and significantly more along local roads.

Number of Through Lanes per Direction

Multilane streets, as opposed to those with one lane in each direction, promote higher motor vehicle traffic speed and decreases the visibility of bicyclists for left-turning and cross motor vehicle traffic at intersections and driveways (Mekuria et al. 2012). The MTI LTS based its LTS criteria for number of lanes on the Dutch CROW Design Manual and modified the Dutch standards by allowing more lanes per direction if the roadway had a median. This study did not consider medians due to the lack of data on the location of medians in the case study area. However, roadways were categorized using the basic number of through lanes criteria used by MTI.

Posted Traffic Speed

High motor vehicle travel speeds have been rated by current and potential bicyclists as a deterrent to bicycling (Winters et al. 2011). Measures of observed speed when available are the best data to use especially when observed traffic speed and the posted speed limit differ. However, observed traffic speed is typically not available. Data on posted speed limit are readily available and for this reason, were used in the study. The posted speed limit criteria used in this study follow the methodology used by MTI for conventional bicycle lanes. This study modified the conventional bicycle lane criteria table to create a buffered bicycle lane table since MTI did not include criteria for buffered bicycle lanes in its analysis. The criteria table for buffered bicycle lanes allows for a slighter higher posted speed limit and functional classification; however, the AADT and number of through lanes per direction remain the same.

On Street Parking

Winters' survey of Metro Vancouver residents found that respondents preferred streets without on street parking to those with on street parking (Winters 2011). It would be preferable to consider if the width of the bicycle lane and parking lane were adequate to reduce perceived stress due to the potential of "dooring". However, parking and bicycle lane width data are typically not readily available. Data collection for on street parking was limited to conventional bike lanes and buffered bike lanes due to the potential that these facilities would position riders in the "dooring" zone.

Table 15. Criteria for Bike Lanes Not Alongside Parking Lane

	LTS ≥ 1	LTS ≥ 2	LTS ≥ 3	LTS ≥ 4
Through lanes per direction	1	(no effect)	≤ 2	(no effect)
Traffic Volume (AADT)	≤ 6,300	> 6,300 - ≤ 14,000	> 14,000 - ≤ 27,000	> 27,000
Functional Class	Local	Major or Minor Collector	Minor Arterial	Principal Arterial
Speed Limit	≤ 25 mph	30 mph	35 mph	≥ 40 mph

Note: (no effect) = factor does not trigger an increase to this level of traffic stress.

Table 16. Criteria for Bike Lanes Alongside Parking Lane

	LTS ≥ 1	LTS ≥ 2	LTS ≥ 3	LTS ≥ 4
Through lanes per direction	1	(no effect)	≤ 2	(no effect)
Traffic Volume (AADT)	≤ 3,000	>3,000 - ≤ 6,300	> 6,300 - ≤ 14,000	> 14,000
Functional Class	Local	(no effect)	Major or Minor Collector	Minor Arterial
Speed Limit	≤ 25 mph	30 mph	35 mph	≥ 40 mph

Note: (no effect) = factor does not trigger an increase to this level of traffic stress.

Table 17. Criteria for Buffered Bike Lanes Not Alongside Parking Lane

	LTS ≥ 1	LTS ≥ 2	LTS ≥ 3	LTS ≥ 4
Through lanes per direction	1	(no effect)	≤ 2	(no effect)
Traffic Volume (AADT)	≤ 6,300	> 6,300 - ≤ 14,000	> 14,000 - ≤ 27,000	> 27,000
Functional Class	Local or Major or Minor Collector	(no effect)	Minor Arterial	Principal Arterial
Speed Limit	≤ 30 mph	35 mph	≥ 40 mph	(no effect)

Note: (no effect) = factor does not trigger an increase to this level of traffic stress.

Table 18. Criteria for Buffered Bike Lanes Alongside Parking Lane

	LTS ≥ 1	LTS ≥ 2	LTS ≥ 3	LTS ≥ 4
Through lanes per direction	1	(no effect)	≤ 2	(no effect)
Traffic Volume (AADT)	≤ 3,000	>3,000 - ≤ 6,300	> 6,300 - ≤ 14,000	> 14,000
Functional Class	Local	Major or Minor Collector	Minor Arterial	Principal Arterial
Speed Limit	≤ 25 mph	30 mph	35 mph	≥ 40 mph

Note: (no effect) = factor does not trigger an increase to this level of traffic stress.

Table 19. Criteria for Shared Travel Lanes

	LTS ≥ 1	LTS ≥ 2	LTS ≥ 3	LTS ≥ 4
Through lanes per direction	1	(no effect)	≤ 2	(no effect)
Traffic Volume (AADT)	≤ 2,000	>2,0000 - ≤ 6,000	> 6,000 - ≤ 14,000	> 14,000
Functional Class	Local	(no effect)	Major or Minor Collector	Minor Arterial
Speed Limit	≤ 25 mph	30 mph	35 mph	≥ 40 mph

CASE STUDY 1 - Beltline

The modified LTS measure was used to classify roadway and bikeway facilities within a six-mile buffer of the Atlanta BeltLine Eastside Trail. The Eastside Trail is a small part of a much larger transportation and economic development project which will provide parks, shared use paths, and transit along a 22-mile historic railroad corridor in Atlanta, Georgia (Atlanta Beltline 2015). The completed Atlanta BeltLine will connect 45 neighborhoods. Four sections of the trail are currently completed, and the Eastside Trail, which is the focus of this case study, was the first segment to be completed (Atlanta Beltline 2015). The case study area was limited to six-miles around the Eastside Trail as research has shown that routes over six miles are perceived as a strong deterrent in the choice to bicycle for many people (Winters et al. 2011).

Data

Three primary data sources were used in this analysis. The NAVTEQ (a company name) Streets 2014 shapefile was obtained by Atlanta Regional Commission (ARC) from the company HERE. It includes a comprehensive inventory of roadways, especially local roadways that are omitted from other data sources. The other roadway database used in the research, RC_ROUTES_ARC, is a modified version of the roadway database maintained by the Georgia Department of Transportation (GDOT) and focuses on state managed roadways rather than locally managed roadways and bikeways. The third data source was the Metro Atlanta Bicycle Facility Inventory, which was compiled from information provided by local governments in the region and verified with Google Earth and Bing Imagery. The location of on street parking on roadways with conventional bicycle lanes and buffered bicycle lanes was manually coded in ArcGIS using Google Earth imagery.

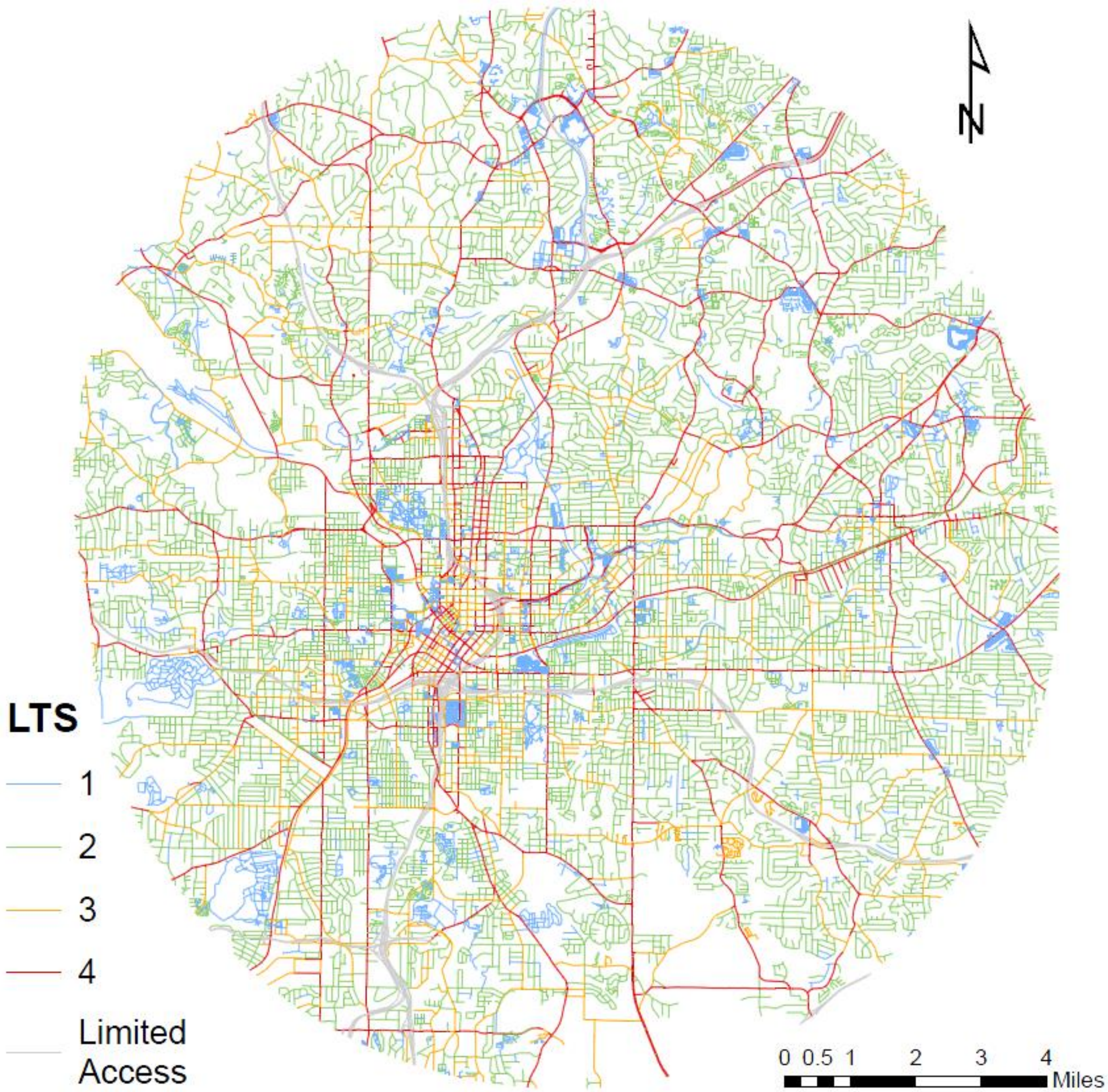


Figure 17. LTS Measure Applied in Case Study Area

An overview of the case study area with the modified LTS measure applied can be seen in Figure 17. LTS is coded by color with blue = LTS 1, green = LTS 2, orange = LTS 3, red = LTS 4, and grey indicating limited access roadways. While only 15% of facilities are LTS 1, a little over half, 54%, of the facilities are LTS 2. The robust presence of LTS 2 facilities in general was also noted in the MTI study (Mekuria et al. 2012) and indicates the prevalence of local or neighborhood streets in the case study area. Approximately 69% of the roadways and bikeways in the case study area are classified at LTS 1 or LTS 2 which are considered low stress for the majority of current and potential bicyclists. As can be seen in Table 20, 344 miles of facilities are LTS 1 and 1,223 miles are LTS 2 in the case study area for a total of 1,567 miles of low stress facilities. Of the 1567 miles of low stress facilities, 1524 miles are local roads while the remaining 43 miles are arterials or other connectors. The shared travel roadway criteria can be seen in Table 20. For a shared travel roadway to be classified as low stress, it must be a local street with a maximum speed limit of 30 mph, have a traffic volume of 6,000 vehicles per day or less, and have a maximum of one through lane per direction of travel.

Table 20. Distribution of Centerline Miles by Level of Traffic Stress and Facility Type

	LTS 1	LTS 2	LTS 3	LTS 4	N/A	Total Miles	Total %
Conventional Bicycle Lanes	0%	16%	**59%**	25%	-	35	100%
Buffered Bicycle Lanes	0%	6%	12%	**82%**	-	2	100%
Shared Travel Roadways	16%	**59%**	12%	13%	-	2028	100%
Side Paths	**100%**	-	-	-	-	10	100%
Protected Cycle Tracks	-	**100%**	-	-	-	1	100%
Shared-Use Paths	-	**100%**	-	-	-	27	100%
Limited Access Roadways	-	-	-	-	**100%**	164	100%
Total Miles	344	**1223**	270	266	164	2267	

Figure 18 presents a zoomed-in version of Figure 17 to provide a more detailed image of the LTS classification of roadways and bikeways around the Atlanta BeltLine Eastside Trail, the focus of this case study.

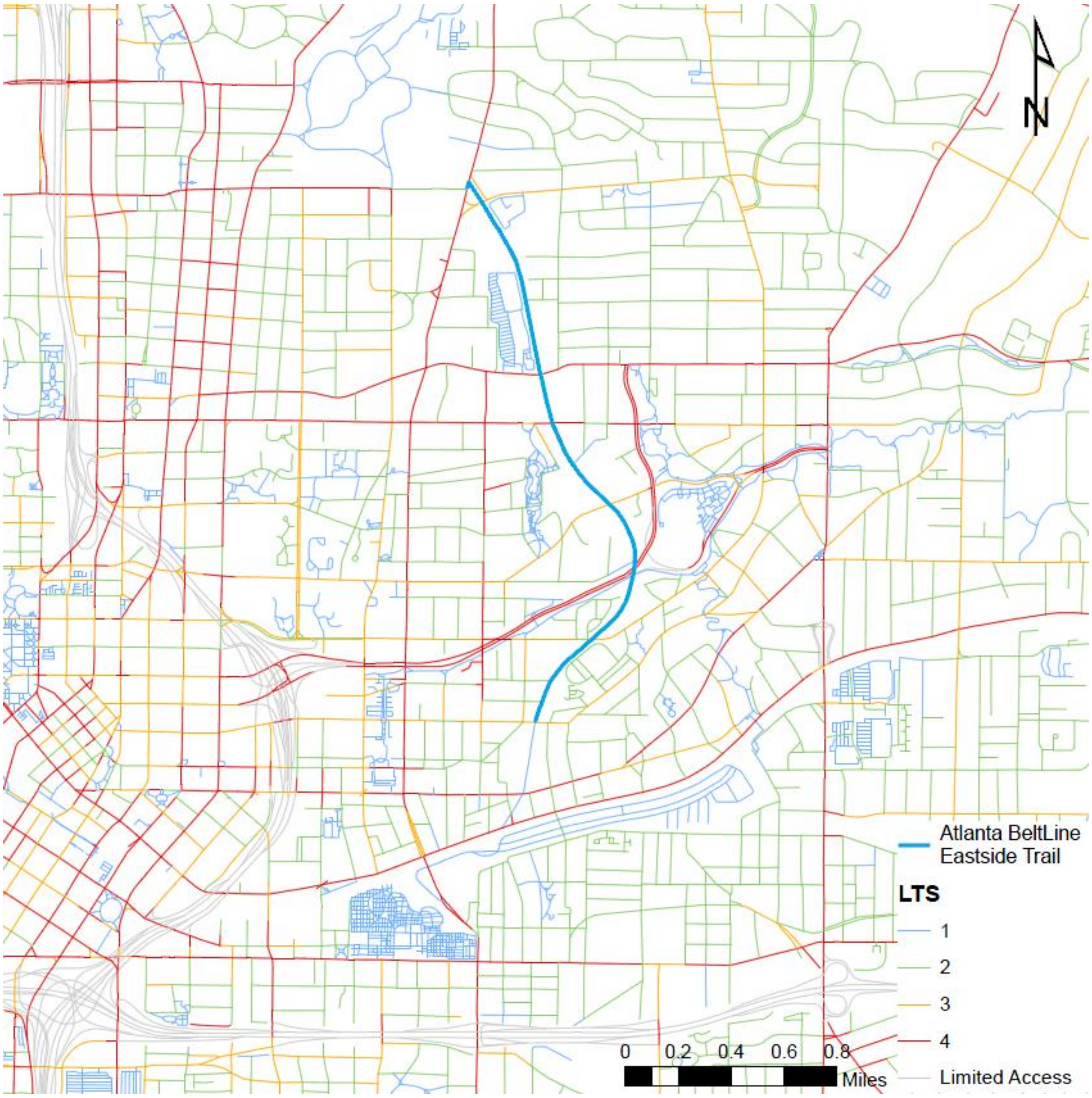


Figure 18. Closer View of LTS in Case Study Area (Atlanta Beltline)

While the majority of shared travel roadways are categorized as low stress facilities (LTS 1 or 2), any collector or arterial functional class roadway without a bicycle facility is categorized as high stress (LTS 3 or 4) based on the criteria in Table19. The majority of the conventional bicycle lanes were categorized as high stress due to high traffic volume and high number of through lanes per direction. The buffered bicycle lane criterion allows for a higher threshold for traffic volume and number of through lanes per direction; however, the two miles of facilities which have currently been built in the case study area are on roadways with traffic volumes that exceed the LTS 1 and 2 threshold.

Even though a majority of the facilities are classified as low stress by the modified LTS measure, it does not mean that the facilities are well connected, as shown in Figure 18. Connectivity in the study area is reduced as a result of two factors; limited access roadways which do not allow bicycle traffic and collector and arterial functional class roadways which trigger the high stress classification. A total of 164 miles of limited access roadways exist throughout the case study area. While there are only 419 miles of collectors and arterials in the study area, they present barriers to a connected bicycle network. Investment in strategic bicycle facilities may be needed to create connected low stress facilities across interstates and other limited access roadways.

A map of roadways and bikeways classified as LTS 1 or LTS 2 is shown in Figure 19. This map reveals that while a majority of the roadways and bikeways in the study area are classified as LTS 1 and LTS 2, these facilities appear to not be well connected. This concept is explored further in the map in Figure 20 where the Atlanta BeltLine Eastside Trail's bikeshed is considered for LTS 1 and LTS 2 facilities. The overview map, Figure 20, shows that the bikeshed does not spread very far outward and includes gaps within the bikeshed. Figure 21 shows a closer view of the previous map.

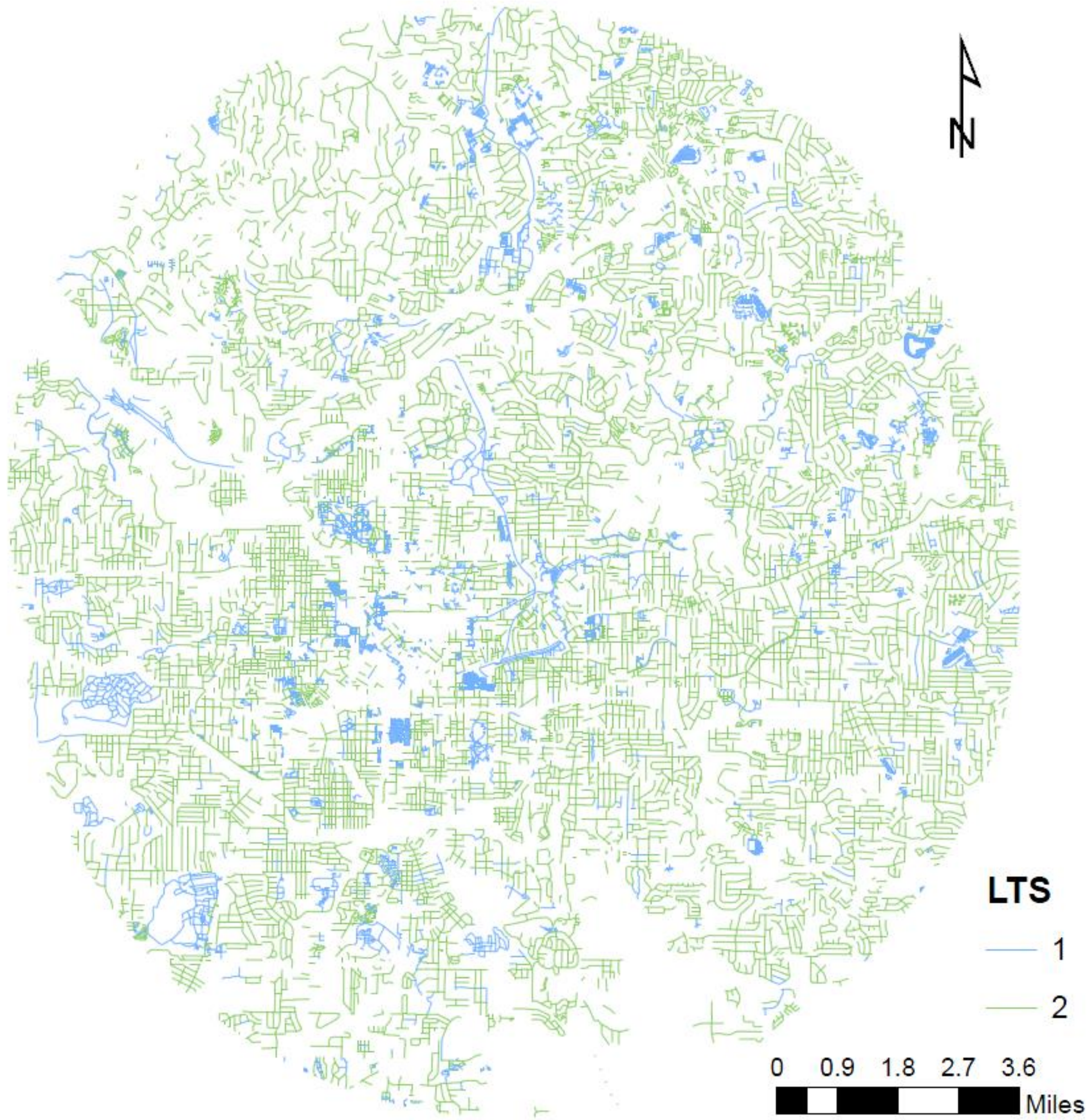


Figure 19. Case Study Area LTS 1 and LTS 2 Facilities Only

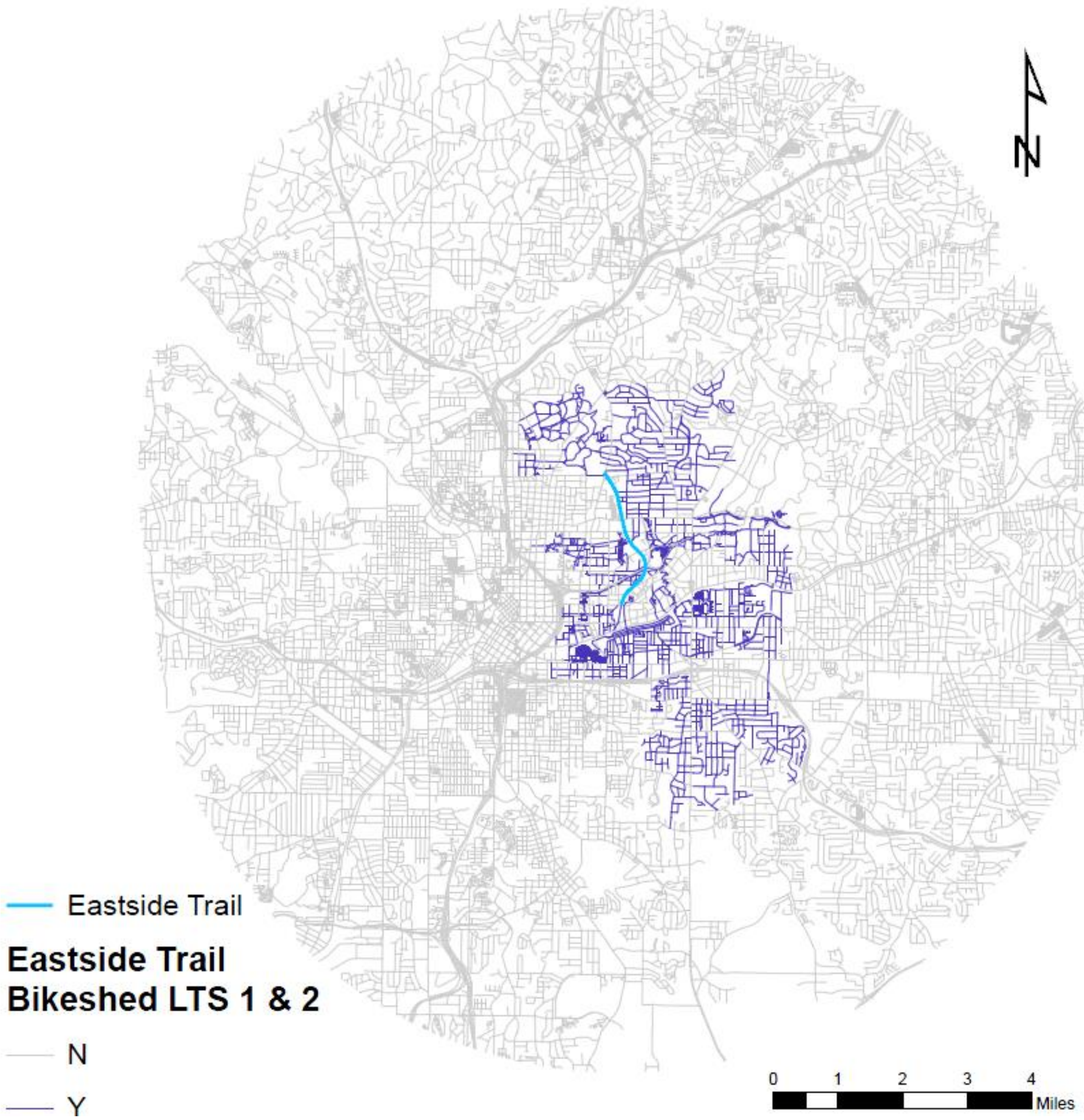


Figure 20. Eastside Trail Bikeshed with LTS 1 and LTS 2 Facilities Only

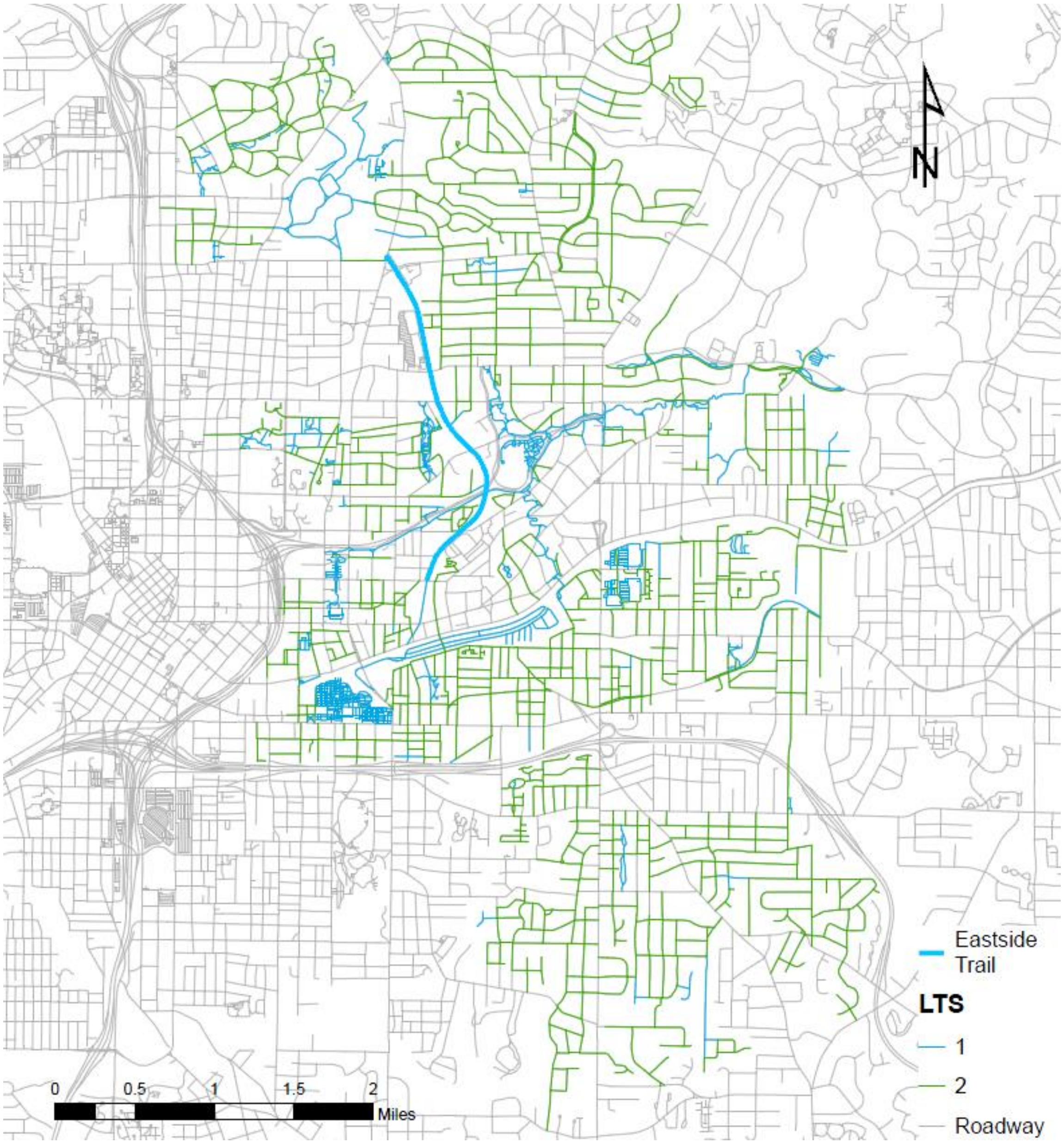


Figure 21. Closer View of Eastside Trail Bikeshed with LTS 1 and LTS 2 Facilities Only

Although a cursory overview of the study area reveals a large amount of LTS 1 and 2 roadways and bikeways, further analysis reveals that the facilities are not well connected. This means that for people who are comfortable using facilities at LTS 1 and LTS 2, estimated to be a majority of current and potential bicyclists, the bike network is disconnected. The case study also shows that while local roadways are an important part of a low stress bicycle network, a well-connected bikeway network cannot be achieved with local streets alone. Collector and arterial roadways provide the connectivity of a roadway network, yet they are too stressful for the majority of current and potential bicyclists without bicycle facilities that provide separation from motor vehicle traffic. Conventional bicycle lanes and buffered bicycle lanes are appropriate to install on collector and arterial roadways when the traffic volume is lower. However, collector and arterial roadways with high traffic volume require greater separation through the use of protected cycle tracks or side paths.

CASE STUDY 2 – MARTA Stations

A further investigation was conducted to demonstrate the utility of using LTS methodology in evaluating the impact of bicycle infrastructure investments. In this case, a similar analysis was undertaken for a 3 miles buffer around the MARTA West End, Oakland City, and Lakewood/Ft. McPherson stations. Improving the bicycle network around MARTA stations can directly increase the bike catchment area for that station and, as a result, could substantially change the commute environment around that station. These stations were chosen specifically for the current development strategies based on market strength and social equity.

To evaluate the low stress bike networks accessing the West End, Oakland City, and Lakewood/Ft. McPherson MARTA stations, three low stress (LTS 1-2) networks as well as the entire (LTS 1-4) bike network were compared based on total network length, accessible area, and accessible population. The accessible area and population were determined based on the 2010 census blocks that intersected each network. The 2010 census was used instead of the 2009-2013 5-year American Community Survey (ACS) estimates because the 5-year estimates are only available at the block group level. The study area population was only 1.7% larger based on the 2013 5-year ACS census block group estimates compared to the 2010 census blocks, and so the 2010 census blocks were chosen for the analysis to allow higher precision. The block group was not granular enough to provide a precise enough definition of the study area.

The low stress networks analyzed were based on the existing low stress infrastructure, proposed improvements in the area, and select key improvements based on the LTS analysis. The final entire LTS bike network included the entire bike network and represented the network available to the most stress-tolerant bicyclists. For each of these analyses, the LTS network was converted into a Network Dataset in ESRI ArcMap. The service area tool identified the streets that were within a network distance of 3 miles from each of the study area MARTA stations.

Figure 22 shows the LTS 1-2 area accessible to each of the study area MARTA stations by network distance.

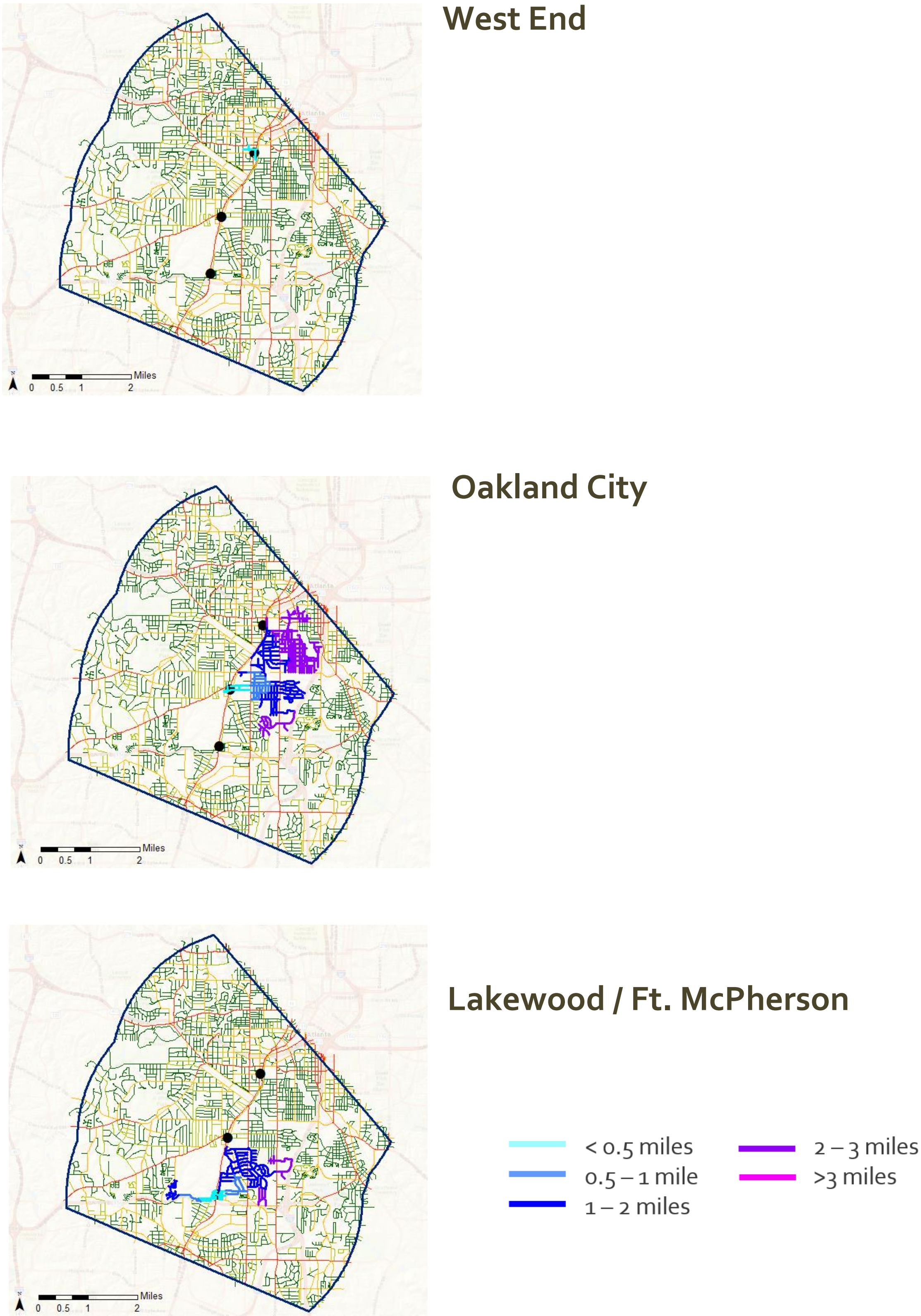


Figure 22. Service Area Analysis based on Existing Conditions LTS 1-2 (Blue) and LTS 1-4 (orange) Network

Proposed Improvements – Low Stress Network

Figure 23 highlights the location and LTS classifications for the proposed improvements (the thick line shows the improved LTS and the superimposed thin line shows the original LTS for the same link). The specific improvements are concentrated in the around West End MARTA station. The addition of the Southwest portion of the beltline trail and the proposed multi-use trail along Peters Street and Lee Street are the most impactful improvements. Figure 24 shows the bike-able network based this proposed network, restricted to a 3 mile network distance from each of the study area MARTA stations.

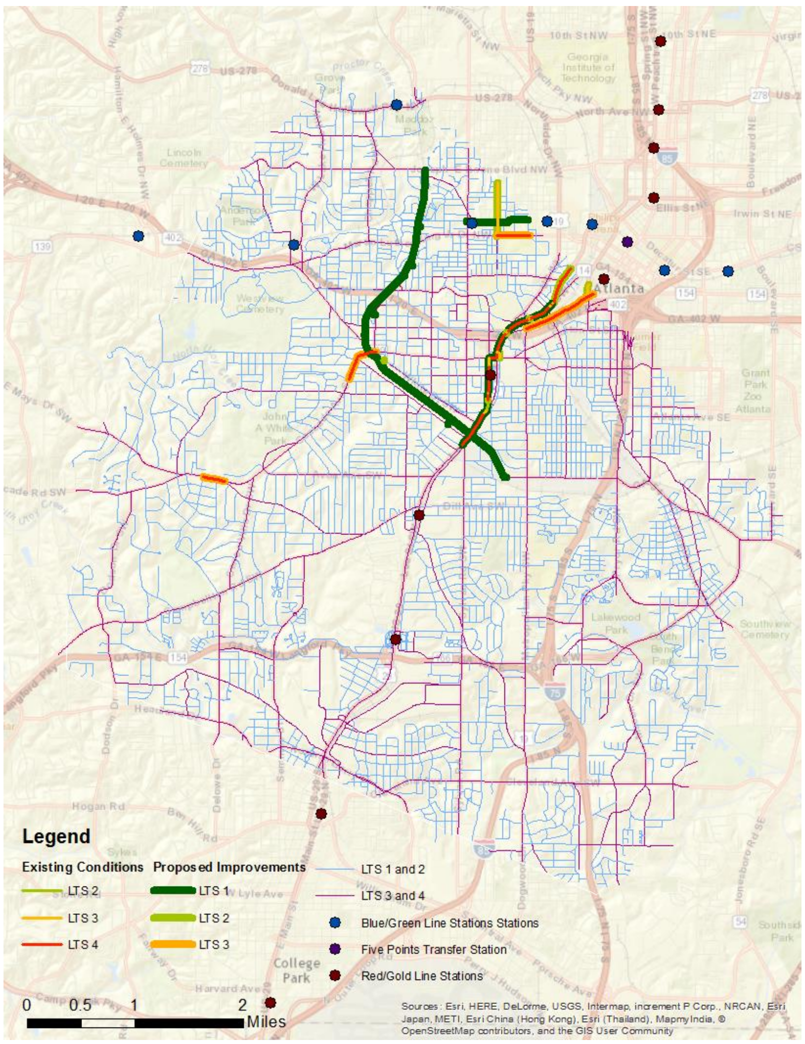


Figure 23. LTS for Links with Proposed Improvements (thick line) and Previous LTS (thin line)

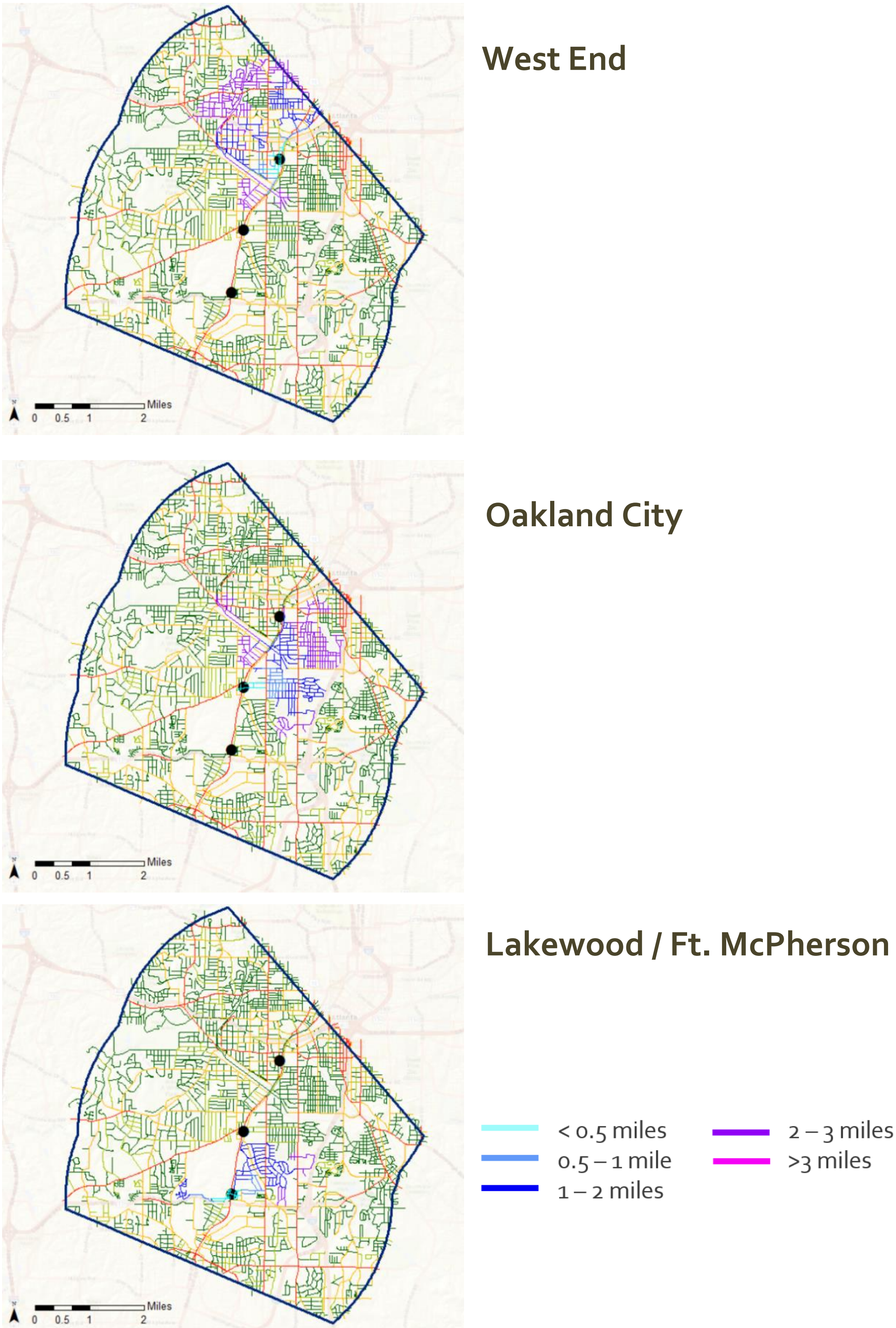


Figure 24. Service Area Analysis based on Proposed Conditions based on Cycle Atlanta Phase 1.0 Plan, Infrastructure Bond, and Southwest Beltline Access Points.

Select Key Improvements – Low Stress Network

In addition to the proposed improvements network, select key improvements were modeled as a demonstration of how this network analysis could be used to identify priority improvements. Potential key improvement locations were identified based on the existing network (Figure 25). These key improvements modeled in the analysis serve as an example of how select targeted improvements can provide major improvements in accessibility. This simple demonstration includes less than 4 miles of high quality improvements (primarily cycle tracks and/or side paths) (Figure 25, thick blue lines represent these key improvements). Figure 26 shows the effect of select key improvements in the study area considered in this analysis.

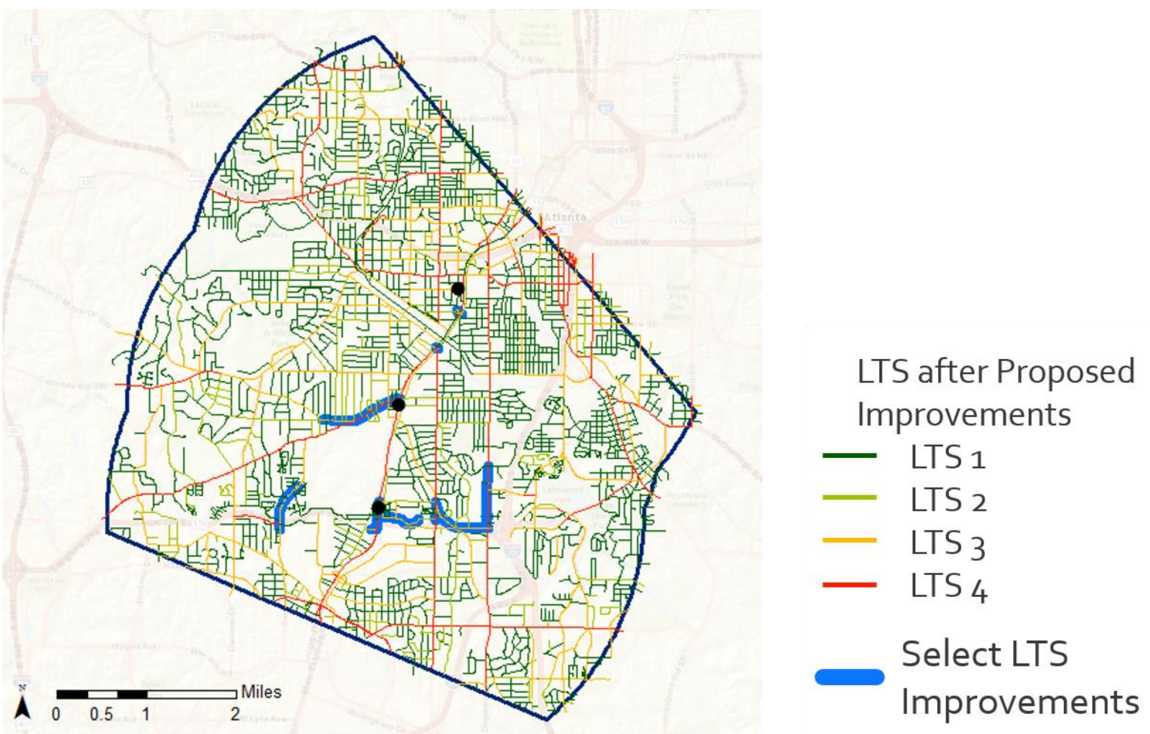


Figure 25. Existing Network with Possible Key Improvements

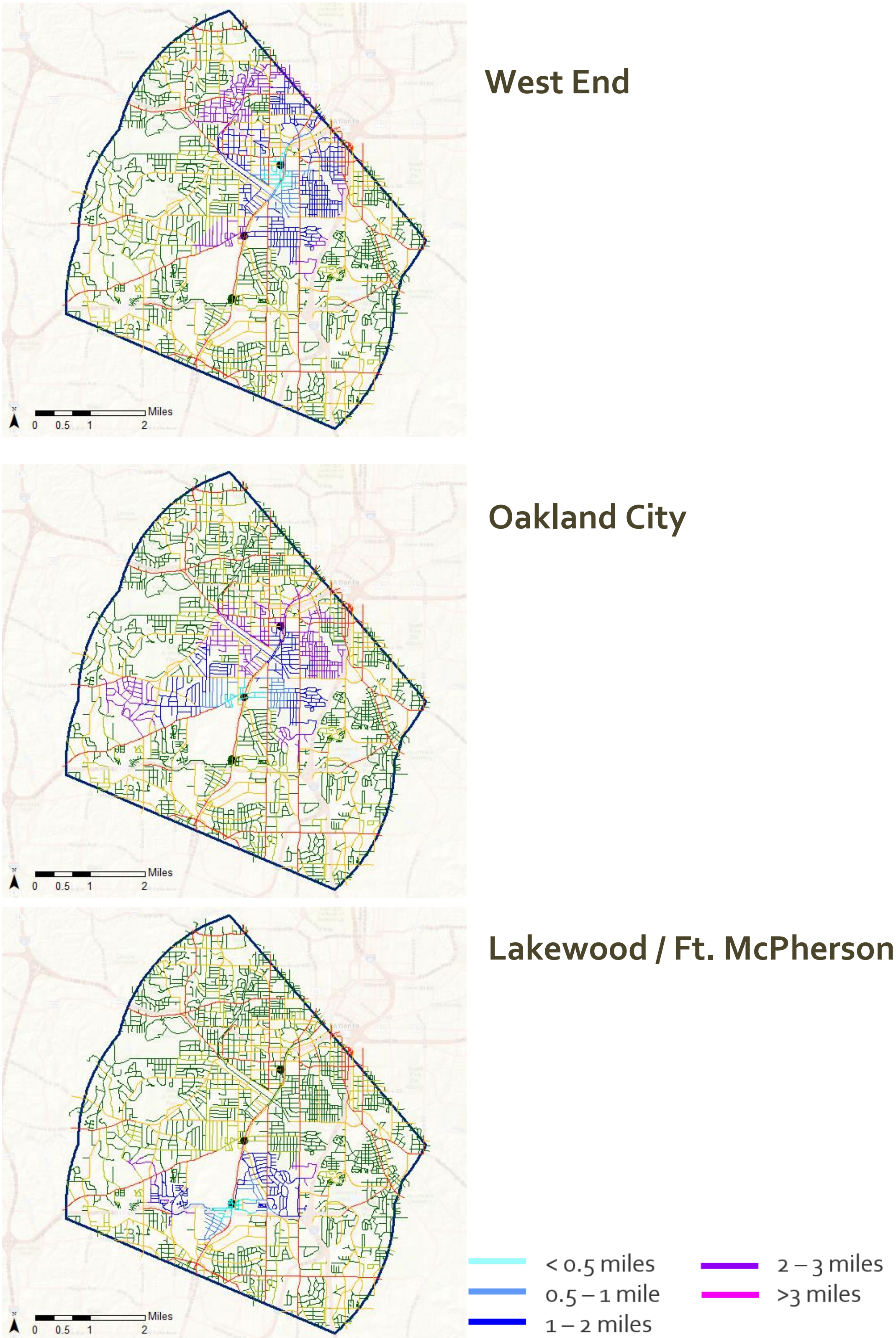


Figure 26. Service Area Analysis based on Select Key Improvements.

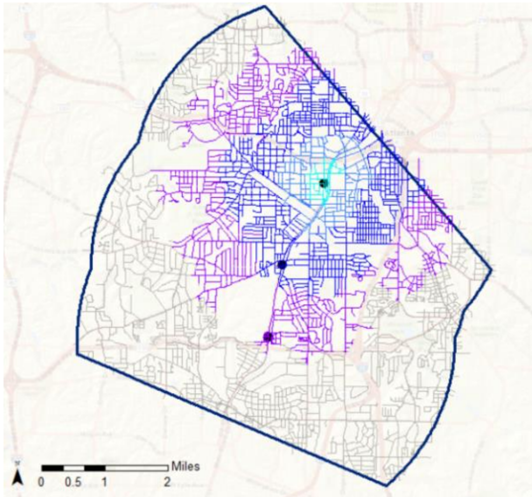
Entire Bike Network

The final network considered under this analysis includes the entire bike network. This includes all infrastructure that a bicyclist is legally permitted to use (i.e. the network excluding highways and restricted access roadways), including roadways classified as LTS 1-4. This network is shown in Figure 27.

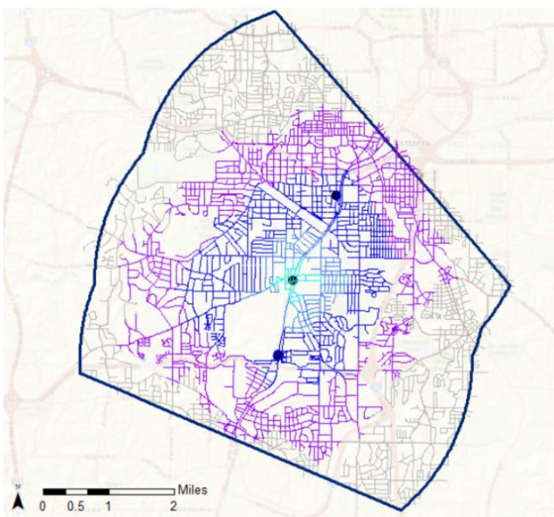
Results

The accessible area for West End, Oakland City, and Lakewood/Ft McPherson were determined based on a service area analysis with a maximum of a 3 mile network distance. This service area analysis was conducted separately for four different networks: existing conditions LTS 1-2; proposed improvements LTS 1-2; select key improvements LTS 1-2; and the entire bike-able network for *strong and fearless* users (LTS 1-4). Figure 28 shows that as the network improved, the accessible area also expands.

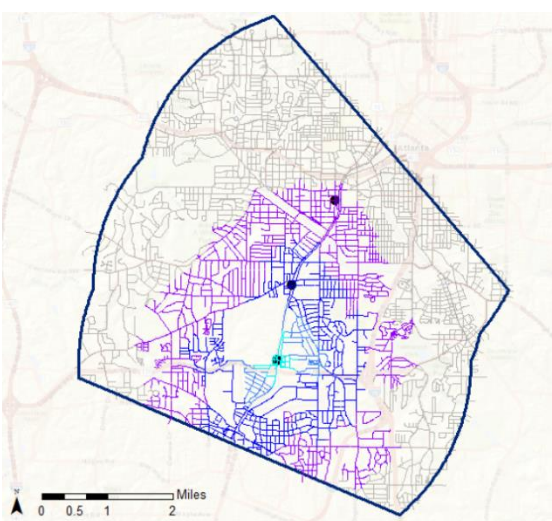
The tables included within Figure 28 show that the overall network distance increases with improvements in the bicycle infrastructure. Figure 29 shows that there are increases in the relative makeup of each network by distance category. The proposed improvements are associated with the largest increase in accessibility at 2-3 miles (158% increase compared to the existing network (Figure 29). The select key improvements, when compared to proposed improvements, are associated with an additional 149% increase in network length within 0.5 – 1 mile of the stations and an additional 90% increase and 85% increase within 0.5 miles and 1-2 miles respectively. This analysis indicates that in addition to expanding the overall network, the select key improvements are associated with increasing the length of the network within 2 miles of the stations.



West End



Oakland City



Lakewood / Ft. McPherson

Figure 27. Service Area Analysis based on Entire Bike-able Network (LTS 1-4)

Existing Conditions

	Total Network Distance (miles)
0-0.5 miles	4.4
0.5-1 mile	6.4
1-2 miles	25.7
2-3 miles	16.2

Proposed Improvements

	Total Network Distance (miles)
0-0.5 miles	6.0
0.5-1 mile	9.9
1-2 miles	41.9
2-3 miles	41.8

Select Key Improvements

	Total Network Distance (miles)
0-0.5 miles	11.4
0.5-1 mile	24.7
1-2 miles	77.5
2-3 miles	36.4

Level of Traffic Stress 4

	Total Network Distance (miles)
0-0.5 miles	17.2
0.5-1 mile	54.9
1-2 miles	150.9
2-3 miles	109.8

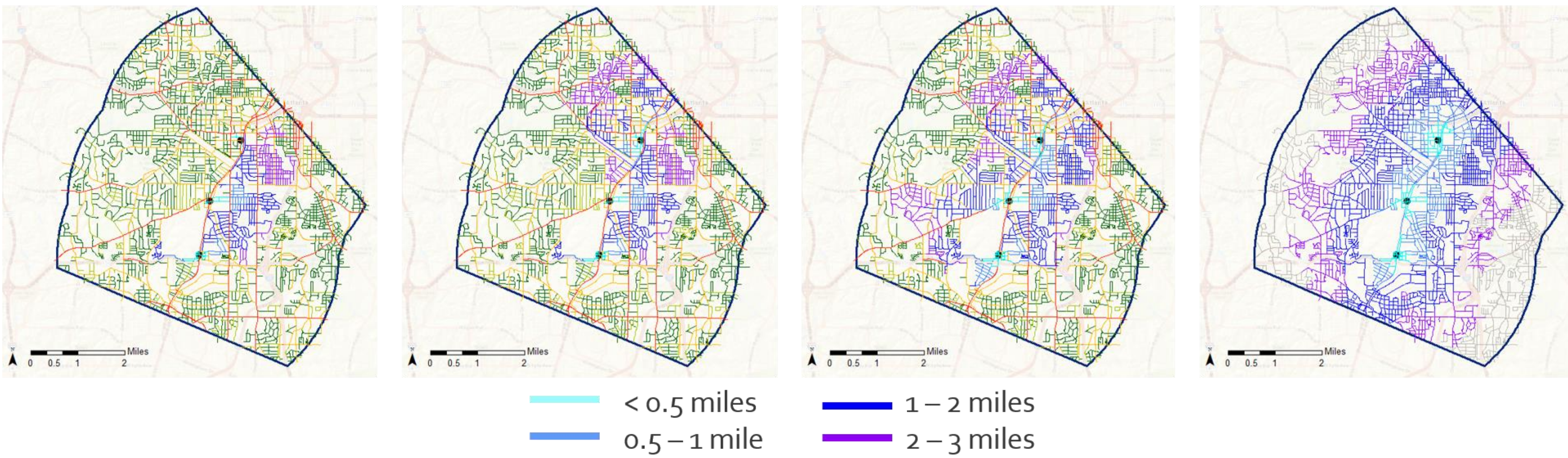


Figure 28. Bike Accessibility by Network Distance for Each of the Four Modeled Networks.

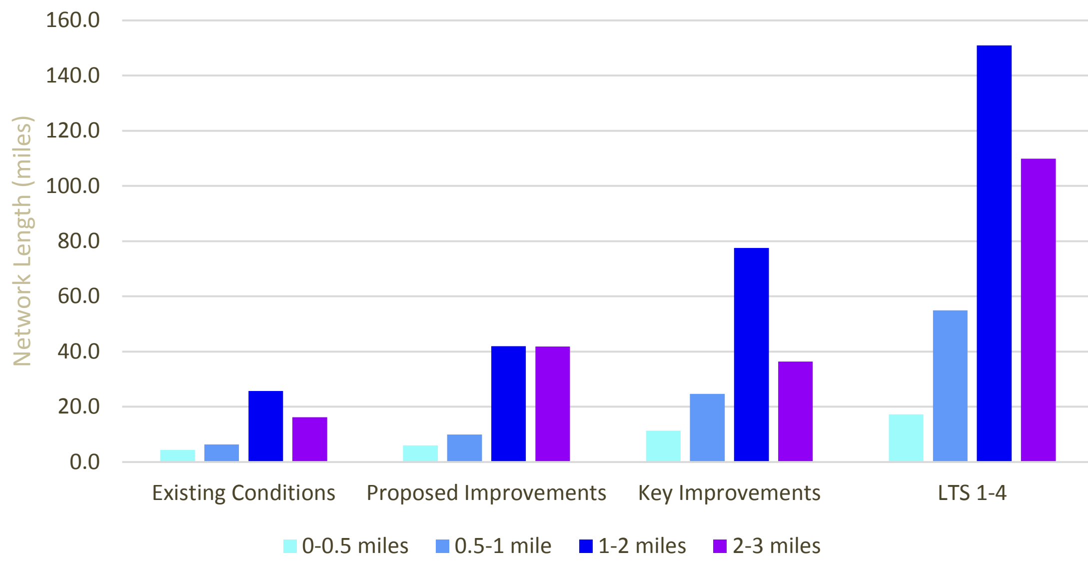


Figure 29. Total Network Length by Distance from the Study Area Stations

To evaluate accessible population and bike-able area, the census blocks that intersect each network were identified and compared to the entire study area population, area, and network distance. The entire study area network is 443 miles long. The LTS 1-4 network is 333 miles long, representing 75% of the network distance, 78% of the study area, and 84% of the population. Under the existing conditions, the low stress bike-able network is only 53 miles long, representing 12% of the network distance, 13% of the study area, and 15% of the population. The proposed improvements define an accessible network that is 101 miles long and represents 23% of the network length, 23% of the study area, and 23% of the population. In addition to the proposed improvements, the select key improvements add an additional 49 miles to the network length (34% of the total study area network) and provide bike-able access to 50% of the population within the study area.

The proposed improvements increase the population that can access MARTA through a low stress bike network by 55%, while the select key improvements increase the accessible population by an additional 116% (Table 21). There are above average increases in the accessible African American population and the population over 45 years old with the proposed improvements. Comparing the proposed improvements network to the select key improvements network, there are additional, above average increases in the accessible African American population, the 18-24 year old population, and (to a much lesser degree) the female population.

Table 21. MARTA Access Demographics based on a 3 mile Biking Distance and Different Levels of Stress and proposed bicycle improvements(Data: 2010 Census)

	Low Stress Existing Conditions	Proposed Improv.	Select Key Improv.	LTS 1-4	Study Area	Proposed increase in Access	Additional Select Key Improv. Increase in Access
Network Length	53.0	100.8	150.0	332.9	443.3	90%	49%
Land Area (sq mi)	4.1	7.0	13.1	24.0	30.9	71%	89%
Total Population	14,656	22,649	48,877	83,142	98,597	55%	116%
White Alone	1,120	1,335	2,144	5,059	6,293	19%	61%
Black Alone	12,991	20,516	45,305	74,936	88,192	**58%**	**121%**
Non White/Black Alone	545	798	1,428	3,147	4,112	46%	79%
Under 18 years	3,790	5,629	11,007	19,565	23,607	49%	96%
18-24 years	1,692	2,521	9,393	13,297	14,874	49%	**273%**
25-34 years	2,356	3,435	6,443	11,412	13,493	46%	88%
35-44 years	1,834	2,784	5,305	95,71	11,482	52%	91%
45-54 years	2,001	3,155	6,206	10,862	13,054	**58%**	97%
55-64 years	1,557	2,552	5,248	9,111	10,927	**64%**	106%
Over 65 years	1,426	2,573	5,275	9,324	11,160	**80%**	105%
Female	7,688	11,884	25,912	44,276	52,278	55%	**118%**
Male	6,968	10,765	22,965	38,866	46,319	54%	113%

These results show that the proposed improvements in the study result in a considerable expansion of the bike-to-transit access area. These improvements in accessibility resulting from the proposed improvements are exclusively a result of the investment in infrastructure around West End Marta stations. These improvements would result in a disproportionately large increase in the bike to transit access for African American, adult, and aging populations (Table 21). These results show that the stated intentions of the bicycle planning efforts in Atlanta to improve overall access with specific interest in the minority and aging population are consistent with this analysis of the low stress bicycle network.

The strategic key improvements were identified based on a visual identification of choke points in the low stress network. The improvements resulted in an additional 116% increase in transit access to population with low stress bike conditions. This increase was primarily around Oakland City and Lakewood/Ft McPherson stations with the largest increase in access among 18-24 year olds (273%, Table 21).

DISCUSSION

With the very limited network, only15% of the population in the study area can bike along a low stress network to a MARTA station.

The proposed improvements are associated with dramatic increases in low stress bike access to the transit stations, specifically in the area north west of the West End station. Under the existing conditions, low stress bike access to/from the West End station is prevented because high stress arterials surround the study area. However, the proposed improvements along Lee Street and the access to the South West portion of the beltline provide low stress access to the West End station. This access in the area immediately surrounding the station connects to local residential streets which extend north and west, expanding the access to low stress bike network by 90% and the accessible population by 55% (Table 21).

The select key improvements were identified solely with the intent of expanding the low stress bike access to the MARTA station in the study area. The select improvements are intended to improve low stress bike access to the Oakland City MARTA station from the west side of the study area, improve access to the Lakewood/Ft. McPherson station, and allow low stress East-West connection across the rail corridor (the East-West connections between Oakland

City and West End were all categorized as high stress links). These targeted improvements expanded the network by 49% and increased the accessible population by 116% (Table 21). The majority of the improved low stress bike access to MARTA resulting from the select key improvements are to the west of the Oakland City Station.

ANALYSIS OF LTS BY CRITERIA

In addition to conducting the low stress bike network accessibility analysis, it was important to understand first what components of the LTS criteria may be driving the overall LTS. The development of the LTS criteria was based in an analysis of existing literature. This analysis is an attempt to understand whether or not any single criterion component was driving the overall LTS designation.

To better understand how each criterion impacted the overall LTS score, the LTS of each link was identified based only on a single criterion, and each link was given 4 “LTS by Criteria” scores: Lane LTS, AADT LTS, Functional Classification LTS, and Speed Limit LTS. Figure 30 visualizes each of the four LTS by criteria in three different maps. Each row includes only the links that were scored as a specific overall LTS (2, 3, or 4) based on all the criteria. Each column shows the LTS score according to a specific criterion (by column). For example, all the maps in the Overall LTS 3 row visualize the same links, but each link is colored according the LTS by Criteria. The map in the Overall LTS 3 row and the Functional Classification column shows the LTS as it would have been determined by the Functional Classification of the link (for only links that were given an overall LTS of 3). If a single criterion map in the LTS 2 row is predominantly light green (LTS 2), then that criterion is driving most of the LTS 2 designation. In this case, LTS 2 is driven mostly by the speed limit designation.

For the low stress bike network analysis, LTS 2 was considered bike-able and LTS 3 was considered too stressful. As a result, the specific criteria driving the jump from LTS 2 to 3 are the most meaningful in the analysis. Figure 30 shows that the two criteria that may be driving a link getting categorized as LTS 3 are functional classification and speed limit. Figure 31 shows the cases in which speed limit and functional classification are the sole determiners of a link being classified as LTS 3 instead of LTS 2.

The classification of a link as LTS 3 compared to LTS 2 solely because of the speed limit seems a legitimate upgrade in LTS. The BLOS and the BCI research both show that a bicyclist is able to perceive a difference in speed. The 85th percentile speed of a link in the BCI index will increase the overall BCI by 0.16 for every 5 mile per hour increase in speed (Harkey et al., 1998). The impact of speed limit on BLOS is less obvious as the factor included in the BLOS equation is: $[1.1199 \ln(SPD - 20) + 0.8103] * (1 + 10.38 * HV)^2$, where HV stands for the proportion of heavy vehicles and SPD stands for prevailing speed. It is not obvious what the exact speed limit threshold should be, but it is intuitive that there is one that exists.

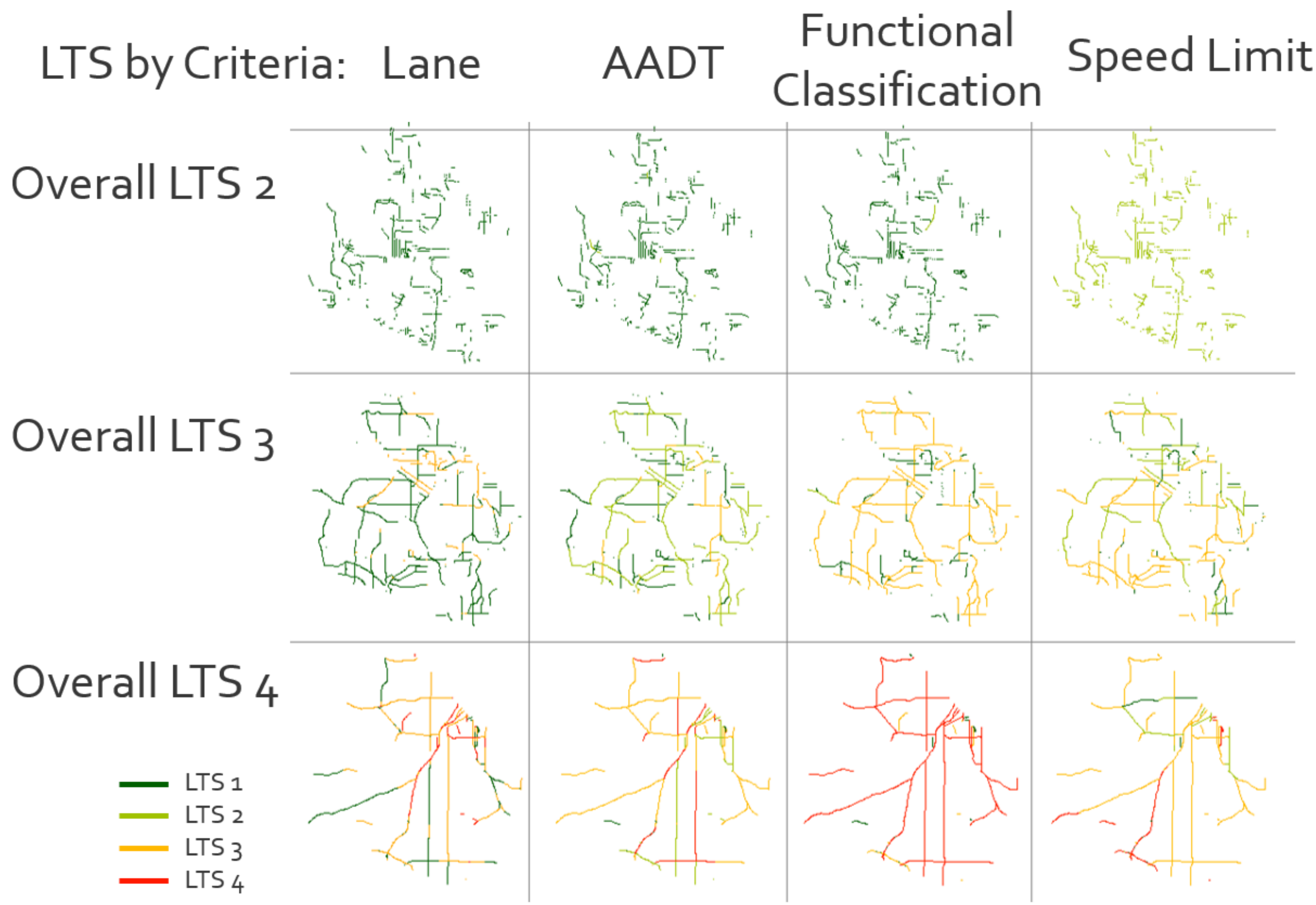


Figure 30. Overall Relevance of Specific Criteria for Determining Overall LTS

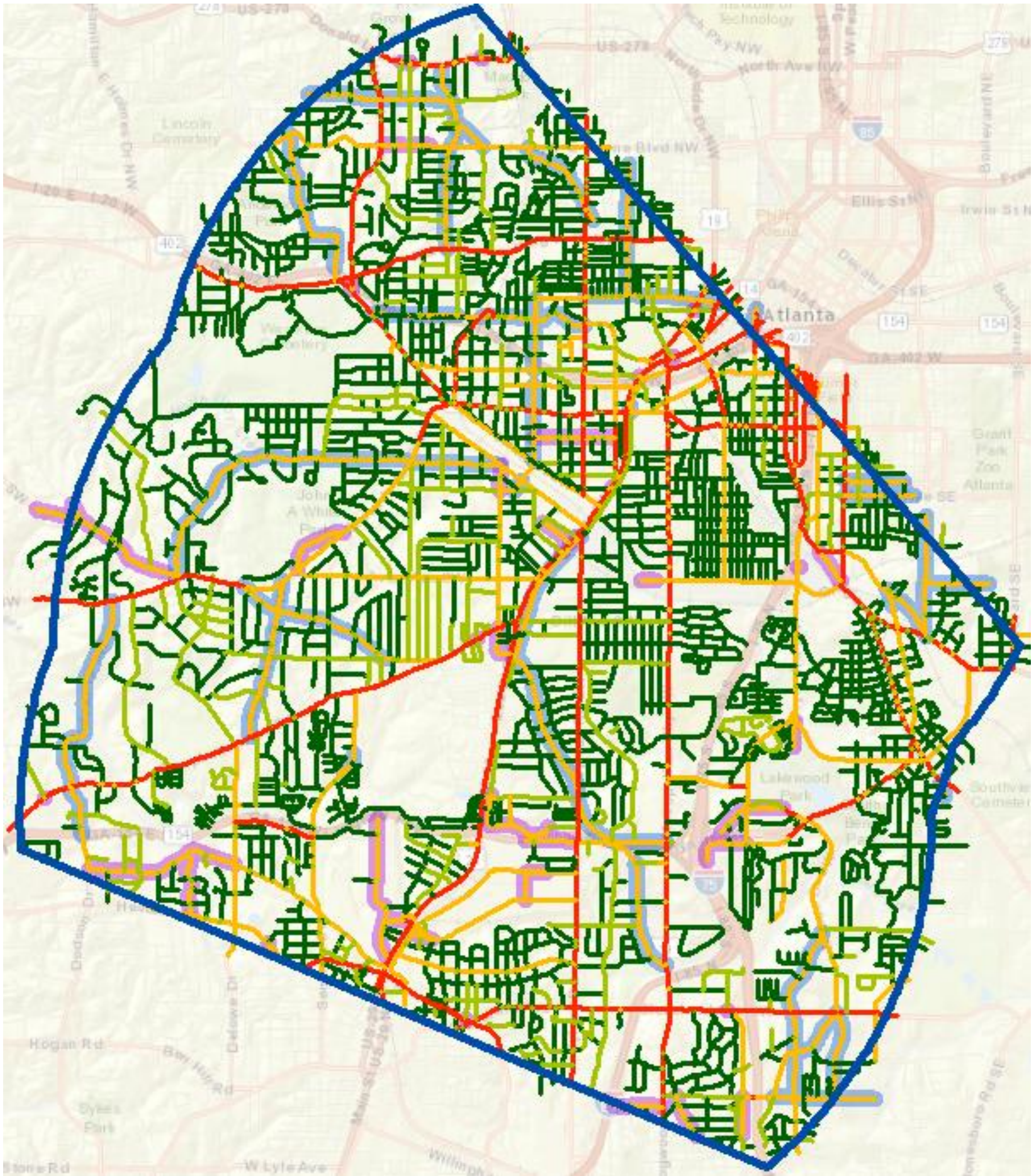


Figure 31. Infrastructure with LTS 3 because of Speed Limit (highlighted in pink) and Functional Classification (highlighted in blue).

It gives pause that links with widths, volumes, and speed limits that classify as LTS 2 should be bumped up to LTS 3 solely because of the functional classification. The functional classification is not included in BCI, BLOS, or original LTS methodologies, and it is unclear whether or not this criterion alone should prevent a link from being included in the low stress network. However, without further research identifying whether or not the functional classification of the road is perceived by bicyclists, there is no justification for eliminating this criterion from the analysis. Therefore, the analysis proceeded with all four LTS criteria.

CONCLUSION

The increased interest in robust bicycle networks by cities throughout the United States emphasizes the need for a standardized bicycle quality-of-service measure that will allow the designation of streets into different categories. This study proposes a quality-of-service measure for bicyclists that is based on the perceived level of traffic stress that the users of differing typologies attach to the facility. The modified LTS measure uses traffic and roadway characteristics data that are readily available to most transportation agencies and has been validated by the literature.

The modified LTS measure can be used by transportation professionals to compare alternative roadway and bikeway designs using quantifiable variables such as speed limit, traffic volume, and number of through travel lanes. The modified LTS measure also provides results that can easily be understood by the public and decision makers. The case studies presented here demonstrate that the LTS methodology in conjunction with a simple connectivity analysis can be used to evaluate and compare bike accessibility within the network and to access transit stations.

The methodology for comparing access presented here could also be applied to a comparison of specific bike infrastructure investment alternatives. The analysis could be used to answer questions like: Could more people access transit through low stress bike infrastructure with a 5-mile buffered bike lane on street X, a 2-mile cycle track along street Y, or 6-0.1 mile side paths targeting specific holes in the low stress network? Of course, the question is specific to low stress bike access, and before making any infrastructure investment, it is important to consider the larger planning context.

Overall, the case study was successful in evaluating the low stress bike access to MARTA stations and comparing this access based on different bicycling infrastructure improvements. However, it is important to understand that the LTS methodology itself has yet to be validated through any user studies. Although the specific criteria thresholds defining each LTS level are supported by the literature, they were developed based on the expert opinions of several researchers.

Furthermore, some of the data that may affect perceived LTS were intentionally not included in this study. In the future, the modified LTS may be updated to include intersection LTS (signalized separated turning movements, vehicle entry point for bicycle lanes and protected cycle tracks, bike boxes, left-turn queue and unsignalized intersection crossings) and bicycle boulevards depending on data availability and sufficiency.

The next steps in this research must be to validate the LTS methodology. The case study analysis shows there is potential value gained from using the current iteration of the Atlanta LTS methodology to compare potential bicycle investments. However, before the method becomes too established in practical applications, it is essential that efforts are made to validate the LTS methodology.

CHAPTER 8. CONCLUSIONS AND RECOMMENDATIONS

The short term goal of this study was to develop a smartphone app to record cyclist trips in Atlanta to obtain data about infrastructure preferences. These data can eventually be used to create a bicycle route choice model to better understand the need for new cycling facilities in Atlanta. Currently, the activity based model of the Atlanta Regional Commission is capable of predicting the number of bike trips in the region, but lacks the capability of assigning the trips onto the city network. The next phase of this research will include a route choice model with logsums from the model estimate that can be fed back to the regional demand model to forecast where the generated trips are most likely to occur. This will greatly help the agency to plan for infrastructure and in improving travel demand forecasting for the region.

The materials produced in this project have been used to give multiple professional seminars and have been included in an undergraduate course titled Multimodal Transportation and a graduate course titled Complete Streets at Georgia Tech. Presentations that have results from this work include:

- Watkins, K., R. Ammanamanchi, J. LaMondia, and C. LeDantec, “Comparison of Smartphone-based Cyclist GPS Data Sources”, *Transportation Research Board 2016 Annual Meeting*.
- Watkins, K. *Integrating Ped/Bike Concepts into University Courses*, Ped-Bike Information Center (part of UNC Highway Safety Research Center) webinar, August 2015.
- Watkins, K. *Crowdsourcing Cyclist Data*, Unraveling Cycling and Pedestrian Flows (European Research Council Grant Kick-off Conference), Amsterdam, Netherlands, July 2015.
- LeDantec, C., K. Watkins, R. Clark, and E. Mynatt, “Cycle Atlanta and OneBusAway: Driving innovation through the data ecosystems of civic computing”, *HCI International 2015*.

- LeDantec, C., M. Asad, A. Misra, and K. Watkins, “Planning with Crowdsourced Data: Rhetoric and Representation in Transportation Planning”, *ACM Conference on Computer-Supported Cooperative Work and Social Computing, 2015.*
- Watkins, K. and C. LeDantec. “Cycle Atlanta”. *Presentation in the Bicycling Atlanta session at the Atlanta Studies Symposium*, March 2015.
- Misra, A., K. Watkins, and C. LeDantec, “Socio-demographic Influence on Cyclists’ Self Classification by Rider Type”, *Transportation Research Board 2015 Annual Meeting.*
- Watkins, K. Cycle Atlanta: Creating a Cycling City, *American Planning Association National Conference*, 2014.
- Watkins, K. Atlanta’s Innovative Planning Technology (mobile workshop), *American Planning Association National Conference*, 2014.
- Misra, A., K. Watkins and C. LeDantec, “Cycle Atlanta – Facilitating GPS Based Data Collection for Bicyclists in Atlanta” *North American Travel Monitoring Exposition and Conference,* 2014.
- Watkins, K. *Teaching Pedestrian and Bicycle Education*, Transportation Research Board Pedestrian and Bicycle Education Subcommittee, Washington DC, January 2014.
- Misra, A., A. Gooze, K. Watkins, M. Asad and C. LeDantec. “Crowdsourcing and Its Application to Transportation Planning” *Transportation Research Board 2014 Annual Meeting*.
- Watkins, K. *Cycle Atlanta: Increasing Public Participation in Cycling Infrastructure Decisions,* Programs for Promoting Cycling, Alabama Transportation Conference, February 2013.
- Misra, A. and K. Watkins*, “Cycle Atlanta: Mapping the Ride to a Better Atlanta” Georgia Transportation Institute poster session,* September 2013, September 2014 and September 2015.

Dr. Watkins offered an undergraduate-level course in the Fall 2013, Fall 2014, and Fall 2015 semester on Multimodal Transportation at Georgia Tech. In Fall 2014 and Fall 2015, this was paired with a graduate course in Complete Streets. The first half of the Multimodal Transportation course and the entire Complete Streets course included topics such as integrating bike and pedestrian infrastructure into street design. A module included the data from Cycle Atlanta with mapping obtained from the apps to facilitate discussion about data sources to understand cyclist movements. Using these materials, Dr. Watkins is available to teach a short course about cyclist GPS data collection uses as desired by any local agency.

CONTRIBUTIONS AND RECOMMENDATIONS

The primary contribution of this research lies in creating a database for cyclists of Atlanta with data from cyclists themselves. As we mentioned earlier, before Cycle Atlanta, there was no benchmark data collection effort that is representative of the cyclists of Atlanta and hence could be used for data-based infrastructure planning. Cycle Atlanta bridges that gap by providing route and sociodemographic data from real cyclists and thus helps planners to make more informed decisions.

Cycle Atlanta data can also play a major role in bridging the gap between National Household Travel Survey data collection efforts and data required in planning for cyclists. As mentioned in the literature, preference for infrastructure and hence the decision to bicycle depends largely on the confidence and comfort of the cyclist with a particular corridor. Riders have been classified into four different categories of comfort and confidence in our research, and we linked these categories to the sociodemographic make-up of the cyclists. The sociodemographic distribution of any region is readily available via census data, and hence, the predominant rider type of that locality can be estimated to make decisions regarding infrastructure.

Cycle Atlanta is unique in its approach to connect sociodemographics of the cyclists to the actual route choice and in comparing stated preference of cyclists versus their actual revealed preference. The stated preference survey indicated that separate facilities are preferred by all cyclist types irrespective of how confident of a rider they are. However, actual trip analysis shows that more confident riders have shorter trip lengths and are more likely to choose shortest routes rather than detour for safer facilities. Similar trends are noted across age and gender. Therefore, to attract less confident riders and female or older riders, it is necessary to have low-stress physically separated infrastructure.

This research also looked into refining the quality of service measure bicycle level of traffic stress (LTS) that can be used to understand the suitability of any corridor for a particular rider type. Several case studies show that depending on level of confidence, the network can appear to be severely disconnected to some cyclists. Our analysis may serve as a starting point for GDOT to plan for bicycle infrastructure, and in the future, as infrastructure is built, the LTS can be used as a measure of cyclist network connectivity.

REFERENCES

1. Innes, J. E. Information in Communicative Planning, *Journal of the American Planning Association*, Vol 64, No.1, pp.52-63,1998.
2. Burby, R.J. Making Plans that Matter, *Journal of the American Planning Association*, Vol.69, No.1, pp.33-49, 2003.
3. Slotterback, C.S. Public Involvement In Transportation Project Planning and Design, *Journal of Architectural and Planning Research*, Vol.27, No.2, pp.144, 2010.
4. Insua. R. D., Kersten E.G., Rios J. and Grima C. Towards decision support for participatory democracy,*ISeB*, Vol. 6, pp.161–191, 2008.
5. Hague, C., Kirk, K., Higgins, M., Prior, A., Jenkins, P., Smith, H. and Grimes, W. Participatory planning for sustainable communities. 2003.
6. Rabinowitz, P. Participatory Approaches to Planning Community Interventions, *http://ctb.ku.edu/en/tablecontents/sub_section_main_1143.aspx*, 2013. Accessed May 2013.
7. Skocpol T., and Fiorina, M. (eds.) Civic Engagement in American Democracy. Brookings Institution Press,1999.
8. Galston,W. A. Civic education and political participation. *PS: Political Science & Politics*, Vol.37, pp. 253–266, 2004.
9. Pew Research Center for the People and the Press. Cable and Internet loom large in fragmented political news universe. *http://www.people-press.org/2004/01/11/cable-and-internet-loom-large-in-fragmented-political-newsuniverse/*, 2004. Accessed May 2013.
10. Wagner, J. Measuring the Performance of Public Engagement in Transportation Planning: Three Best Principles. In *TRB 2013 Annual Meeting*, Vol. 954, 2012.
11. Howe, J. The Rise of Crowdsourcing. *Wired, Conde Nast Digital*, 14, 6, Jun, 2006.
12. Doan, A,. Ramakrishnan, R., Halevy, A.Y. Crowdsouring Systems on the World-Wide Web, *Communications of the ACM*. Vol. 54, No. 4, pp. 86-96, 2011.

13. Saxton, G. D., Oh, O.,and Kishore, R. Rules of Crowdsourcing: Models, Issues, and Systems of Control, *http://www.acsu.buffalo.edu/~rkishore/papers/Saxton-et-al-Crowdsourcing-ISM-Forthcoming.pdf* . Accessed May 2013.

14. Goodchild, M. Assertion and authority: the science of user-generated geographic content. *Earth*, 2008, pp. 1-18.

15. Kuznetsov,S. and Paulos, E. Participatory Sensing in Public Spaces: Activating Urban Surfaces with Sensor Probes, *ACM Designing Interactive Systems (DIS)* 2010.

16. Kuznetsov, S., Davis, G. N., Cheung, J. C. and Paulos, E. Ceci N'est Pas Une Pipe Bombe: Authoring Urban Landscapes with Air Quality Sensors, *ACM SIGCHI* 2011.

17. Hood,J. , Sall, E., Charlton, B. A GPS-based bicycle route choice model for SanFrancisco, California. *Transportation Letters: The International Journal of Transportation Research*, Vol.3, Jan. 2011, pp. 63-75

18. Kitchin, R. and Dodge, M. Rethinking Maps, In *The Map Reader: Theories of Mapping Practice and Cartographic Representation (eds M. Dodge, R. Kitchin and C. Perkins)*, John Wiley & Sons, Ltd, Chichester, UK.

19. Steinfield, A., Zimmerman, J. and Tomasic, A. Bringing Customers Back into Transportation: Citizen-Driven Transit Service Innovation via Social Computing, In *Best Practices for Transportation Agency Use of Social Media*, (*eds. S. Bregman and K. Watkins)*, CRC Press, 2013

20. Erickson, T. Geocentric Crowdsourcing and Smarter Cities: Enabling Urban Intelligence in Cities and Regions. *A position paper for the 1st International workshop on ubiquitous crowdsourcing. UbiComp'10*, Copenhagen, Denmark, 2010.

21. Priedhorsky, R., Jordan, B., and Terveen, L. How a Personalized Geowiki Can Help Bicyclists Share Information More Effectively. *Proc. 2007 Int'l Symposium on Wikis, ACM,* 2007, pp. 93-98.

22. Heipke, C. Crowdsourcing Geospatial Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 65, No. 6, pp. 550-557, 2010.

23. Brabham, D. C. Crowdsourcing the Public Participation Process for Planning Projects. *Planning Theory*, Vol. 8, No. 3, Jul. 2009, pp. 242-262.

24. SeeClickFix. *http://seeclickfix.com/*, 2013. Accessed May 2013.

25. Ferris, B., Watkins, K., and Borning, A. OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. *Proceedings of CHI 2010.* Atlanta, GA, USA, April 10 – 15, 2010.

26. Zimmerman, J., Tomasic, A., Garrod, C., Yoo, D., Hiruncharoenvate, C., Aziz, R., Thiruvengadam, N.R., Huang, Y. and Steinfeld, A. Field trial of Tiramisu: crowd-sourcing bus arrival times to spur codesign. *In Proceedings of CHI '11. ACM*, New York, NY, USA, 1677-1686, 2011.

27. Cycle Atlanta. *http://www.cycleatlanta.org*, 2013. Accessed May 2013.

28. Dill, J. and McNeil, N. Four Types of Cyclists? Examining the Typology to Better Understand Bicycling Behavior and Potential, *92nd Annual Meeting of the Transportation Research Board*, 2013.

29. Masli, M. Crowdsourcing Maps, *Computer,* Vol. 44, No. 11, 2011, pp. 90-93.

30. Cyclopath. *http://cyclopath.org*, 2011. Accessed May 2013.

31. Wiggins, A. & Crowston, K. From conservation to crowdsourcing: A typology of citizen science,In *Proceedings of the Forty-fourth Hawai'i International Conference on System Science (HICSS- 44),* Koloa, HI, 4–7 January.

32. Gooze, A.I. Real-time Transit Information Accuracy: Impacts and Proposed Solutions, M.S Thesis, Georgia Tech, *http://hdl.handle.net/1853/47638*, 2013. Accessed May 2013.

33. Gooze, A, Watkins, K., Borning, A. Benefits of Real-Time Information and the Impacts of Data Accuracy on the Rider Experience. To appear in *Transportation Research Record: Journal of the Transportation Research Board*, 2013.

34. Windmiller, S., Hennessy, T. and Watkins, K. “Communication Technology Usage and the Rider Experience: A Case Study of St. Louis Metro”, submitted to Transportation Research Board Annual Meeting, 2014.

35. AASHTO. (2012) Guide for the Development of Bicycle Facilities.

36. Adam, N.R., and Wortman, J.C. (1989) Security-control methods for statistical databases: A comparative study. ACM Computing Surveys, 21:515-556.

37. Ardagna C.A., Cremonini M., Damiani E., Vimercati S.C, Samarati P. (2007) Location Privacy Protection Through Obfuscation-based Techniques, In, Data and Applications Security XXIV 4602 Lecture Notes in Computer Science E Barker, Steve.E Ahn, Gail-Joon.

38. Aultman-Hall, et al (1997) Analysis of Bicycle Commuter Routes Using Geographic Information Systems. Transportation Research Record No. 1578, Transportation Research Board.

39. Bayardo, Jr. R. J. and Agrawal, R. (2005) Data Privacy through Optimal k-Anonymization. In ICDE, 2005.

40. Bekhor, S., Ben-Akiva, M.E., Ramming, S., (2006). Evaluation of choice set generation algorithms for route choice models. Annals of Operations Research, 144(1), 235-247.

41. Bekhor, S., Prato, C.G., (2006). Effects of choice set composition in route choice modeling. Proceedings of the 11th International Conference on Travel Behavior Research, Kyoto, Japan.

42. Ben-Akiva, M. & Bierlaire, M. (1999), “Discrete choice methods and their applications to short term travel decisions,” in R. Hall, ed., *Handbook of Transportation Science*, Kluwer Academic Publishers, Norwell, MA, chapter 2,

43. Ben-Akiva, M.E., Bergman, M.J., Daly, A.J., Ramaswamy, R., (1984). Modeling inter-urban route choice behaviour. In: Volmuller, J., Hamerslag, R. (Eds.), Proceedings of the 9th International Symposium on Transportation and Traffic Theory. VNU Science Press, Utrecht, The Netherlands, 299-330.

44. Ben-Akiva, M.E., Bolduc, D., (1996). Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper, Massachusetts Institute of Technology, Cambridge, USA.

45. Beresford, A. R. and Stajano, F. (2003) Location Privacy in Pervasive Computing. IEEE Pervasive Computing, 2(1):46–55, 2003.

Mokbel, M. F. (2006) Towards Privacy-Aware Location-Based Database Servers. In Proceedings of the ICDE International Workshop on Privacy Data Management, PDM, 2006.

46. Bierlaire, M., Frejinger, E., (2005). Route choice models with subpath components. Proceedings of the 5th Swiss Transport Research Conference, Ascona, Switzerland.

47. Bierlaire, M., Frejinger, E., (2008). Route choice modeling with network-free data. Transportation Research Part C, 16(2), 187-198.

48. Bliemer, M.C.J., Bovy, P.H.L., (2008). Impact of route choice set on route choice probabilities. Transportation Research Record, 2076, 10-19.

49. Bovy, P.H.L., (2009). On modelling route choice sets in transportation networks: a synthesis. Transport Reviews, 29(1), 43-68.

50. Bovy, P.H.L., Bekhor, S., Prato, C.G., (2008). The factor of revised path size: an alternative derivation. Transportation Research Record, 2076, 132-140.

51. Bovy, P.H.L., Bekhor, S., Prato, C.G., (2009). Route sampling correction for stochastic route choice set generation. Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

52. Bovy, P.H.L., Fiorenzo-Catalano, S., (2007). Stochastic route choice set generation: behavioral and probabilistic foundations. Transportmetrica, 3(3), 173-189.

53. Broach, J, et al (2010) Bicycle route choice model developed using revealed preference GPS data. Transportation Research Board Annual Meeting Compendium.

54. Buehler, R. and Pucher. J. (2011) Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes, http://policy.rutgers.edu/faculty/pucher/bikepaths.pdf, accessed August 2013.
55. Cascetta, E., (2001). Transportation Systems Engineering: Theory and Methods. Kluwer Academic Publishers, Dordrecht, The Netherlands.
56. Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., (1996). A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Lesort, J.B. (Ed.), Proceedings of the Thirteenth International Symposium on Transportation and Traffic Theory. Pergamon, Lyon, France, 697-711.
57. Charlton, B, et al (2010) CycleTracks – a Bicycle Route Choice Data Collection Application for GPS-Enabled Smart Phones. Innovations in Travel Modeling paper compendium, Transportation Research Board.
58. Chu, C., (1989). A paired combinatorial logit model for travel demand analysis. Proceedings of the 5th World Conference on Transportation Research, Ventura, USA, 295-309.
59. Daganzo, C.F., Sheffi, Y., (1977). On stochastic models of traffic assignment. Transportation Science, 11, 253-274.
60. De la Barra, T., Perez, B., Anez, J., (1993). Multidimensional path search and assignment. Proceedings of the 21st PTRC Summer Annual Meeting, Manchester, England, 307-319.
61. Dial, R.B., (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation Research, 5(2), 83-111.
62. Dijkstra, E.W., (1959). A note on two problems in connection with graphs. Numerical Mathematics, 1, 269-271.
63. Dill, J, et al (2008) Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route Choice. Oregon Transportation Research and Education Consortium, final report OTREC-RR-08-03.
64. Dill, J. (2004) Measuring Network Connectivity for Bicycling and Walking.

65. Dill, J. and Carr, T. (2003) Bicycle Commuting and Facilities in Major U.S.. Cities: If You Build Them, Commuters Will Use Them, Transportation Research Records, Volume 1828, pg 116-123.

66. Dill, J. and McNeil, N. (2012) Four types of cyclists? Examining the typology to better understand bicycling behavior and potential., in 92nd Annual Meeting of the Transportation Research Board.

67. Dougherty, M., (1995). A review of neural networks applied to transport. Transportation Research Part C, 3(4), 247-260.

68. Eppstein, D., (1998). Finding the *K* shortest paths. Journal of the Society for Industrial and Applied Mathematics, 28(2), 652-673.

69. Fiorenzo-Catalano, S., Van der Zij, N.J., (2001). A forecasting model for inland navigation based on route enumeration. Proceedings of the European Transport Conference, PTRC Education and Research Services Ltd., London, 1-11.

70. Frejinger, E., (2007). Random sampling of alternatives in a route choice context. Proceedings of the European Transport Conference, Leeuwenhorst, The Netherlands.

71. Frejinger, E., Bierlaire, M., (2007). Capturing correlation with subnetworks in route choice models. Transportation Research Part B, 41(3), 363-378.

72. Frejinger, E., Bierlaire, M., (2007). Capturing correlation with subnetworks in route choice models. Transportation Research Part B, 41(3), 363-378.

73. Frejinger, E., Bierlaire, M., Ben-Akiva, M.E., (2009). Sampling of alternatives for route choice modeling. Transportation Research Part B,

74. Friedrich, M., Hofsäß, I., Wekeck, S., (2001). Timetable-based transit assignment using branch & bound. Transportation Research Record, 1752, 100-107.

75. Gliebe, J.P., Koelman, F., Ziliaskopoulos, A., (1999). Route choice using a paired combinatorial logit model. Proceedings of the 78th Annual Meeting of the Transportation Research Board, Washington, D.C.

76. Hoh B., Gruteser M., Herring R., Ban, J., Work, D. Herrera, J-C, Bayen, A, M., Annavaram, M., Jacobson, Q. (2008) Virtual Trip Lines for Distributed Privacy-Preserving Traffic Monitoring, *MobiSys'08*

77. Hood, J. et al. (2011) A GPS-based bicycle route choice model for San Francisco, California. Transportation Letters: the International Journal of Transportation Research 3: 63-75.

78. Horowitz, J.L., Louviere, J.J., (1995) What is the role of consideration sets in choice modeling? International Journal of Research in Marketing, 12, 39-54.

79. Hudson, J, et al (2012) Using Smartphones to Collect Bicycle Travel Data in Texas. University Transportation Center for Mobility, Texas Transportation Institute, Project report UTCM 11-35-69.

80. Hunt, D.T., Kornhauser, A.L., (1997). Assigning traffic over essentially-least-cost paths. Transportation Research Record, 1556, 1-7.

81. Jan, O., Horowitz, A., Peng, Z., (2000). Using GPS data to understand variations in path choice. Transportation Research Record 1706, 145-151.

82. Koelman, F., Wen, C., (1998). Alternative nested logit models: structure, properties and estimation. Transportation Research Part B, 32(5), 289-298.

83. Krizek K. (2007) Two Approaches to Valuing Some of Bicycle Facilities' Presumed Benefits: Propose a session for the 2007 National Planning Conference in the City of Brotherly Love. Journal of the American Planning Association. 72(3): 309-320.

84. Kuby, M., Zhongyi, X., Xiaodong, X., (1997). A minimax method for finding the k-best differentiated paths. Geographical Analysis, 29(4), 298-313.

85. Lam, T.C., Small, K., (2001). The value of time and reliability: measurement from a value pricing experiment. Transportation Research Part E, 37(2-3), 231-251.

86. League of American Bicyclists. Bicycle Commuting Data, http://www.bikeleague.org/

87. Liu, S., Araujo, M., Brunskill, E., Rossetti, R., Barros, J., and Krishnan, R. (2013. Understanding sequential decisions via inverse reinforcement learning. In Mobile Data Management (MDM), IEEE 14th International Conference on, volume 1, pages 177– 186. IEEE, 2013.
88. Lombard, K., Church, R.L., (1993). The gateway shortest path problem: generating alternative routes for a corridor location problem. Geographical Systems, 1, 25-45.
89. M. Gruteser and D. Grunwald. (2003) Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In MobiSys, 2003.
    Chow, C-Y, Mokbel, M.F. and Liu, X. (2006) A Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-Based Services, ACM-GIS'06
90. Ma. C.Y.T, Yau, D.K.Y, Yip N. K., Rao, N S V (2010) Privacy Vulnerability of Published Anonymous Mobility Traces, MobiCom 2010.
91. McFadden, D., Train, T., (2000). Mixed MNL models for discrete response. Journal of Applied Econometrics, 15(5), 447-470.
92. Menghini, G. et al (2010) Route choice of cyclists in Zurich. Transportation Research Part A 44 754–765.
93. Nielsen, O.A., (2004). Behavioural responses to pricing schemes: description of the Danish AKTA experiment. Journal of Intelligent Transportation Systems, 8(4), 233-251.
94. Nielsen, O.A., Daly, A., Frederiksen, R.D., (2002). A stochastic multi-class road assignment model with distributed time and cost coefficients. Networks and Spatial Economics, 2, 327-346.
95. Opportunities for collaboration on transportation and public health research. Transportation Research Part A 28(4): 249-268.
96. Park, D., Rilett, L.R., (1997). Identifying multiple and reasonable paths in transportation networks: a heuristic approach. Transportation Research Record 1607, 31-37.
97. Poznanski, A. J. (2013), Analyzing demographic and geographic characteristics of "Cycle Atlanta" smartphone application users. MS thesis, Georgia Tech.

98. Prashker, J.N., Bekhor, S., (1998). Investigation of stochastic network loading procedures. Transportation Research Record, 1645, 94-102.

99. Prashker, J.N., Bekhor, S., (2000). Congestion, stochastic, and similarity effects in stochastic user equilibrium models. Transportation Research Record, 1733, 80-87.

100. Prashker, J.N., Bekhor, S., (2004). Route choice models used in the stochastic user equilibrium problem: a review. Transport Reviews, 24(4), 437-463.

101. Prato, C. G. (2009) Route choice modeling: past, present and future research directions, *Journal of Choice Modelling,* 2(1), pp. 65-100

102. Prato, C.G., (2005). Latent factors and route choice behaviour. Ph.D. Thesis, Turin Polytechnic, Italy.

103. Prato, C.G., Bekhor, S., (2006). Applying branch & bound technique to route choice set generation. Transportation Research Record, 1985, 19-28.

104. Prato, C.G., Bekhor, S., (2007). Modeling route choice behavior: how relevant is the choice set composition? Transportation Research Record, 2003, 64-73.

105. Pucher, J. and R. Buehler. (2007) Making Cycling Irresistible: Lessons from the Netherlands, Denmark and Germany. Transport Reviews: A Transnational Transdisciplinary Journal. 28(4): 495-528.

106. Ramming, S.,(2002). Network knowledge and route choice. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, USA.

107. Ruphail, N.M., Ranjithan, S.R., ElDessouki, W., Smith, T., Brill, E.D., (1995). A decision suort system for dynamic pre-trip route planning. Applications of advanced technologies. In: Transportation Engineering: Proceedings of The Fourth International Conference, 325-329.

108. Sallis, JF, Frank LD, Saelens BE, Kraft MK (2004) Active transportation and physical activity:

109. Samarati P. and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical Report SRI-CSL-98-04, Computer Science Laboratory, SRI International,

110. Schussler, N. & Axhausen, K. (2009), "Processing raw data from global positioning systems without additional information," Transportation Research Record: Journal of the Transportation Research Board 2105, 28–36.

111. Scott, K., Pabon-Jimenez, G., Bernstein, D., (1997). Finding alternatives to the best path. Proceedings of the 76th Annual Meeting of the Transportation Research Board, Washington, D.C.

112. Sener, I., Eluru, N. & Bhat, C. (2009), "An analysis of bicycle route choice preferences in Texas, US, *Transportation* 36, 511–539. Transportation Research Board Annual Meeting Compendium

113. Sheffi, Y., (1985). Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall, Englewood Cliffs, USA.

114. Sheffi, Y., Powell, W.B., 1982. An algorithm for the equilibrium assignment problem with random link times. Networks, 12, 191-207.

115. Swait, J., (2001). Choice set generation within the generalized extreme value family of discrete choice models. Transportation Research Part B, 35(7), 643-666.

116. Swait, J., Ben-Akiva, M.E., (1987). Incorporating random constraints in discrete models of choice set generation. Transportation Research Part B, 21(2), 91-102.

117. The City of Atlanta, Department of Planning and Community Development Department (2011) CycleAtlanta:Phase1.0 http://documents.atlantaregional.com/lci/2012Applications/Innov_Atlanta_CycleAtlanta.pdf, accessed November 2013.

118. Van der Zij, N.J., Fiorenzo-Catalano, S., (2005). Path enumeration by finding the constrained K-shortest paths. Transportation Research Part B, 39(6), 545-563.

119. Vovsha, P., (1997). The cross-nested logit model: application to mode choice in the Tel Aviv metropolitan area. Transportation Research Record, 1607, 13-20.

120. Wen, C., Koelman, F., (2001). The generalized nested logit model. Transportation Research Part B, 35(7), 627-641.

121. Zhou, Z., Chen, A., (2003). Stochastic user equilibrium problem: a comparison between length-based and congestion-based C-Logit models. In: Loo, B.P.Y., Lam, S.W.K. (Eds.), Proceedings of the 8th Hong Kong Society of Transportation Studies Conference: Transportation and Logistics. Hong Kong, China, 244-253.

122. Ziebart, B. D.,Maas, A. L,, Bagnell, J.L and Dey, A. K., (2008).Maximum entropy inverse reinforcement learning. In Artificial Intelligence, AAAI 23th Conference, 1433–1438.

123. R. Buehler and J. Pucher, “Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes,” Transportation, vol. 39, no. 2, pp. 409–432, Mar. 2012.

124. J. Dill and J. Gliebe, “Understanding and measuring bicycling behavior: A focus on travel time and route choice,” 2008.

125. M. Winters, K. Teschke, M. Grant, E. M. Setton, and M. Brauer, “How Far Out of the Way Will We Travel?: Built Environment Influences on Route Selection for Bicycle and Car Travel,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2190, no. -1, pp. 1–10, Dec. 2010.

126. P. J. Krenn, P. Oja, and S. Titze, “Route choices of transport bicyclists: a comparison of actually used and shortest routes,” Int. J. Behav. Nutr. Phys. Act., vol. 11, no. 1, p. 31, 2014.

127. J. Broach, J. Dill, and J. Gliebe, “Where do cyclists ride? A route choice model developed with revealed preference GPS data,” Transp. Res. Part Policy Pract., vol. 46, no. 10, pp. 1730–1740, Dec. 2012.

128. J. Hood, E. Sall, and B. Charlton, “A GPS-based bicycle route choice model for San Francisco, California,” Transp. Lett. Int. J. Transp. Res., vol. 3, no. 1, pp. 63–75, Jan. 2011.

129. F. Godefroy and C. Morency, “Estimating Latent Cycling Trips in Montreal, Canada,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2314, no. -1, pp. 120–128, Dec. 2012.

130. J. Parkin, M. Wardman, and M. Page, “Estimation of the determinants of bicycle mode share for the journey to work using census data,” Transportation, vol. 35, no. 1, pp. 93–109, Nov. 2007.

131. M. Nankervis, “The effect of weather and climate on bicycle commuting,” Transp. Res. Part Policy Pract., vol. 33, no. 6, pp. 417–431, 1999.

132. C. Brandenburg, A. Matzarakis, and A. Arnberger, “The effects of weather on frequencies of use by commuting and recreation bicyclists,” Adv. Tour. Climatol., vol. 12, pp. 189–197, 2004.

133. L. F. Miranda-Moreno and T. Nosal, “Weather or Not to Cycle: Temporal Trends and Impact of Weather on Cycling in an Urban Environment,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2247, no. -1, pp. 42–52, Dec. 2011.

134. J. Pucher, C. Komanoff, and P. Schimek, “Bicycling renaissance in North America? Recent trends and alternative policies to promote bicycling,” Transp. Res. Part Policy Pract., vol. 33, no. No. 7/8, pp. 625–654, 1999.

135. R. Cervero and M. Duncan, “Walking, bicycling, and urban landscapes: evidence from the San Francisco Bay Area,” Am. J. Public Health, vol. 93, no. 9, pp. 1478–1483, 2003.

136. M. Winters, M. Brauer, E. M. Setton, and K. Teschke, “Built Environment Influences on Healthy Transportation Choices: Bicycling versus Driving,” J. Urban Health, vol. 87, no. 6, pp. 969–993, Dec. 2010.

137. J. Dill and T. Carr, “Bicycle commuting and facilities in major US cities: if you build them, commuters will use them,” Transp. Res. Rec. J. Transp. Res. Board, vol. 1828, no. 1, pp. 116–123, 2003.

138. M. J. Smart, K. M. Ralph, B. D. Taylor, C. Turley, and A. E. Brown, “Honey, Can You Pick-Up Groceries on Your Way Home? Analyzing activities and travel among students and in non-traditional households,” UCTC-FR-2014-07, 2014.

139. J. Pucher, J. Dill, and S. Handy, “Infrastructure, programs, and policies to increase bicycling: An international review,” Prev. Med., vol. 50, pp. S106–S125, Jan. 2010.

140. M. Winters, G. Davidson, D. Kao, and K. Teschke, “Motivators and deterrents of bicycling: comparing influences on decisions to ride,” Transportation, vol. 38, no. 1, pp. 153–168, Jan. 2011.

141. A. Misra, K. Watkins, and C. A. Le Dantec, “Socio-demographic Influence on Cyclists’ Self Classification by Rider Type,” presented at the Transportation Research Board 94th Annual Meeting, 2014.

142. M. S. Urban, C. D. Porter, K. E. Proussalgoglou, R. Calix, and C. Chu, “Modeling the Impacts of Bicycle Facilities on Commute and Recreational Bicycling in Los Angeles County,” in Transportation Research Board 93rd Annual Meeting, Washington, D.C., 2014, vol. No. 14–3904.

143. J. Dill and N. McNeil, “Four Types of Cyclists?: Examination of Typology for Better Understanding of Bicycling Behavior and Potential,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2387, no. -1, pp. 129–138, Dec. 2013.

144. HCM 2010 Highway Capacity Manual, vol. 3, 4 vols. Transportation Research Board, 2010.

145. D. L. Harkey, D. W. Reinfurt, and M. Knuiman, “Development of the bicycle compatibility index,” Transp. Res. Rec. J. Transp. Res. Board, vol. 1636, no. 1, pp. 13–20, 1998.

146. San Francisco Department of Public Health Environmental Health Section, “Bicycle Environmental Quality Index Data Collection Manual,” n.d.

147. M. C. Mekuria, P. G. Furth, and H. Nixon, “Low-stress bicycling and network connectivity,” Mineta Transportation Institute, 11-19, 2012.

148. HCM, HCM 2010 Highway Capacity Manual, vol. 1, 4 vols. Transportation Research Board, 2010.

149. J. Parks, A. Tanaka, P. Ryus, C. M. Monsere, N. McNeil, and M. Goodno, “Assessment of Three Alternative Bicycle Infrastructure Quality-of-Service Metrics,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2387, no. -1, pp. 56–65, Dec. 2013.

150. M. Winters, “Improving Public Health Through Active Transportation: Understanding the Influence of the Built Environment on Decisions to Travel by Bicycle,” The University of British Columbia, 2011.

151. American Association of State Highway and Transportation Officials, “Guide for the Development of Bicycle Facilties,” AASHTO, Washington, D.C., 1999.

152. AASHTO, Guide for the Development of Bicycle Facilities, 4th Edition, 4th ed. 2012.

153. J. Dill and K. Voros, “Factors affecting bicycling demand: Initial survey findings from the Portland region,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2031, no. 1, pp. 9–17, 2007.

154. R. L. Sanders, “Dissecting perceived traffic risk as a barrier to adult bicycling,” in Proceedings of the 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.

155. F. Ahmed, G. Rose, and C. Jakob, “Commuter Cyclist Travel Behavior: Examination of the Impact of Changes in Weather,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2387, no. -1, pp. 76–82, Dec. 2013.

156. R. Geller, “Four Types of Cyclists.” Portland Office of Transportation, 2006.

157. U.S. Department of Transportation Federal Highway Administration, “Highway Functional Classification Concepts, Criteria and Procedures, 2013 Edition,” USDOT FHWA, 2013.

158. C. Monsere, J. Dill, N. McNeil, K. Clifton, N. Foster, T. Goddard, M. Berkow, J. Gilpin, K. Voros, D. van Hengel, and others, “Lessons from the Green Lanes: Evaluating Protected Bike Lanes in the US,” 2014.

159. Atlanta BeltLine, “Atlanta BeltLine Overview.” 2015.