BICYCLE ROUTE CHOICE

GPS DATA COLLECTION AND TRAVEL MODEL DEVELOPMENT—YEAR 1 (2012–13)

FINAL PROJECT REPORT

by

Qing Shen, P.I.*; Peng Chen*; Peter Schmiedeskamp*; Alon Bassok*; Suzanne Childress†

*University of Washington
†Puget Sound Regional Council

for

Pacific Northwest Transportation Consortium (PacTrans)
USDOT University Transportation Center for Federal Region 10
University of Washington
More Hall 112, Box 352700
Seattle, WA 98195-2700

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Pacific Northwest Transportation Consortium and the U.S. Government assumes no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
2012-S-UW-0019	01538109	

4. Title and Subtitle	5. Report Date
Bicycle Route Choice: GPS Data Collection and Travel Model Development	September 15, 2014
	6. Performing Organization Code
7. Author(s)	**8. Performing Organization Report No.**
Qing Shen, Peng Chen, Peter Schmiedeskamp, Alon Bassok, Suzanne Childress	19-625083
9. Performing Organization Name and Address	**10. Work Unit No. (TRAIS)**
PacTrans Pacific Northwest Transportation Consortium University Transportation Center for Region 10 University of Washington More Hall 112 Seattle, WA 98195-2700 Department of Civil and Environmental Engineering University of Washington 201 More Hall Seattle, WA 98105 Puget Sound Regional Council 101 Western Avenue Suite 500 Seattle, WA 98104 (List matching agency name and address)	**11. Contract or Grant No.** DTRT12-UTC10
12. Sponsoring Organization Name and Address	**13. Type of Report and Period Covered**
United States of America Department of Transportation Research and Innovative Technology Administration	Research 9/1/2012-7/31/2014
	14. Sponsoring Agency Code

15. Supplementary Notes

Report uploaded at www.pacTrans.org

16. Abstract

Bicycle use is being promoted for a variety of social benefits. Because of the benefits associated with bicycling, jurisdictions across the central Puget Sound region and the nation have been investing in improvements to bicycle infrastructure. Academic and professional literature provides a basis for generally understanding bicycling behavior. However, less is known about the benefits of one facility type over another, or the potential inducement of new bicycle users when a policy intervention improves bicycling conditions.

This study will rely on GPS bicycle trace data collected by the Puget Sound Regional Council through the CycleTrack mobile application. The objectives of the study include improving the Puget Sound Regional Council's travel demand model to include bicycle route choice and assignment, which will allow for policy analysis and an improved understanding of the tradeoffs between facilities, the relationship between utilitarian and recreational bicycling, and an analysis of the utility of a number of bicycle facilities that will become operational over the course of the study.

17. Key Words	18. Distribution Statement
Bicycle, data, GPS	No restrictions.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.		NA

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

Contents

List of Figures

List of Tables

Executive Summary / Introduction

Bicycle use is being promoted for a variety of social benefits. Persons who choose to bicycle on a regular basis for commute or other purposes receive benefits in the form of lower obesity rates and other health improvements, and in lower transportation costs (Dill 2009; Frank et al. 2006; Sallis et al. 2004). In addition, society as a whole benefits from an increase in cycling. As people shift modes from automobiles to bicycles, additional capacity is available on the roadways relieving congestion. Further, as those who choose to bicycle become healthier, they present less of a burden on the health care system, reducing overall health care costs (Sturm 2002).

Because of the benefits associated with bicycling, jurisdictions across the central Puget Sound region and the nation have been investing in improvements to bicycle infrastructure. The improvements range from inexpensive tools such as painted sharrows on arterials and road rechannelization projects, to more costly greenways and separated off-road facilities. While investments in bicycle infrastructure are far less costly than traditional roadway or transit improvements, jurisdictions must decide how to best invest limited resources.

Academic and professional literature provides a basis for generally understanding bicycling behavior. However, less is known about the benefits of one facility type over another, or the potential inducement of new bicycle users when a policy intervention improves bicycling conditions (Dill and Gliebe 2008; Krizek, El-Geneidy, and Thompson 2007; Tilahun, Levinson, and Krizek 2007). Furthermore, analytical tools aimed at assessing travel demand, benefit-cost analysis, and travel behavior cannot be improved to capture the benefits of bicycling in the absence of 1) data on the use of bicycling facilities, and 2) a theoretical framework for understanding cyclist route choice decisions. This project begins to address these gaps by collecting GPS trace data for bicycle users in the central Puget Sound, and utilizing it for policy analysis and travel model improvements.

This project was successful in using a GPS smartphone application, CycleTracks, written by the San Francisco County Transportation Authority to collect revealed preference GPS data representing 2,750 trips taken by 165 unique bicyclists. The collected sample has limitations that preclude making robust, generalizable conclusions. However, in addition to validating the collection mechanism, this project has achieved a number of aims as it has worked toward developing a statistical model.

This project has contributed to knowledge transfer by reviewing the literature to identify factors important to bicyclists and their decision making; identifying from the literature an appropriate approach to choice set generation; and selecting / beginning implementation of a robust and repeatable approach to the data processing that more closely integrates with the IT infrastructure of the Puget Sound Regional Council.

As noted above, the data collection process did yield a dataset with limitations, in particular with regards to the non-random sampling methodology, and the somewhat smaller than expected sample size. The limitations posed by the sample could be addressed through a future data collection effort using the CycleTracks application. A recommendation of this report is to revisit the sampling approach in the future—perhaps teaming with the organizers of a region-wide bike-to-work event for recruitment and data collection. Finally, while the choice set generation proved more problematic than initially anticipated, the analytical approach identified and the data processing strategies begun in this project could be completed to carry out the analysis of such a future data set.

Chapter 1 Literature Review

This literature review considers the factors that affect a person's choice to use a bicycle. The review begins with a number of items that affect bicycle mode choice, but are not specific to route decisions. There variables are, for the most part, exogenous to route-choice modeling efforts and are also generally outside the control of engineering and urban planning interventions. This literature review then turns its attentions to built environment and other factors that can be used within route-choice modeling and are heavily influenced by design interventions.

1.1 Weather

Temperature and precipitation play a large role in whether people choose to use bicycles. When it is cold or very hot outside less people will use bicycles. Similarly, as precipitation increases, less people will use bicycles. This relationship is far more pronounced for choice users rather than commuters (Miranda-Moreno and Nosal 2011). That is, a rainy Saturday will see fewer riders compared to a sunny one as compared to a regular commute day. However, because there is generally little variation in temperature and rain along differing routes between an origin and a destination, weather has an affect on mode choice rather than route choice. And, for the purposes of modelling, it is unlikely that aggregate travel models can account for the fluctuations in temperature and precipitation.

1.2 Household size

The smaller the household, the more likely the persons residing there are to use bicycling as a mode of travel (Andrade and Kagaya 2012). This finding makes intuitive sense. Smaller households, especially those without children, do not have to make the same trips that those with children do—i.e. there is no need to drop off children at school or soccer practice, making it more feasible to not rely on a car.

1.3 Vehicle availability

Vehicle availability is not important in the choice for whether or not people choose to use a bicycle (Andrade and Kagaya 2012). When households own vehicles, they tend to use them. That is not to say that people who own vehicles may not just use them for one trip purpose and not bicycle for other purposes—e.g. use the car for grocery shopping and the bicycle for commuting. However, once a household has a vehicle available to them, they are more likely to use them. The likelihood of cycling is further reduced as the number of vehicles available increases. Conversely, if you move from areas where 100 percent of households own vehicles to places where only one in five households does so, cycling rates are increased by 30 percent (Meng et al. 2014).

1.4 Trip end amenities

The lack of provision of trip end amenities decreases a persons propensity to use bicycles (Nkurunziza et al. 2012). At a most basic level, trip end amenities include secure bicycle parking. This is important for experienced riders who have expensive bicycles and novice riders who need encouragement. Additional amenities include locker facilities, at least for the purpose of stowing bicycle related equipment—e.g. helmet and shoes. More advanced treatments include showers and locker room, which become much more useful for people commuting over longer distances, especially if their commute is also a part of their exercise regiment.

1.5 Attitude

People are more likely to try bicycling as a mode of transportation if there is general support for bicycling in their city. When there is not, there can be anger and frustration from drivers and pedestrians alike when they encounter a bicyclists in what they perceive to be their space, which can be intimidating to new bicyclists. In promoting new bicycle infrastructure projects or pro-bicycling programs, jurisdictions should work to *normalize* bicycling and work towards positive attitudes of everyone who uses the right-of-way, not just those most interested in bicycling (Castillo-Manzano and Sanchez-Braza 2013).

Because bicycling can have a communal aspect, and for many people a sense of self-identity (Chatterton and Wilson 2014) group rides or other events that promote the social aspect of bicycling will promote increase in bicycling across trip purposes. This is particularly important for adolescents as their early positive experience with bicycling and other people's attitudes regarding bicycling around them greatly influences their propensity to ride bicycles as adults (Sigurdardottir et al. 2013). And, the enjoyment of recreational cycling changes people's behaviour, such that bicycling becomes a habit (Schneider 2013) and a regular part of transportation options.

The remaining items discussed in the literature review all have the potential to affect bicycle route choice and are also factors that can be influence by planning, policy and engineering.

1.6 Distance

Although not strictly a factor that affects route choice, cyclists generally avoid longer trips, especially for commute trip purposes (Broach, Dill, and Gliebe 2012). From a modelling perspective, this has several implications. First, beyond a certain distance (likely 20 miles or more), there is little reason to believe people will choose bicycling (except for exercise or recreation). This suggests that origin and destination pairs beyond a threshold of 20 miles do not need to be considered. Secondly, because more modelling efforts are concerned with a typical weekday, and, because so many of the trips made would related to non recreational travel, it is again safe to only consider short origin and destination pairs.

1.7 Slope

Bicyclists generally prefer flat places to ride with locations with minimal slopes being preferable to places with steep hills (Broach, Dill, and Gliebe 2012). Similar to other variables, this preference changes by trip purpose. Commuters generally dislike hills (Hood, Sall, and Charlton 2011). Bicyclists who are riding for exercise purposes may prefer to have topographical changes for increased exertion, especially if they are training for a longer, more strenuous ride. Of course, if a trip end is on top of a hill, novice bicyclists may switch mode to avoid the climb, but in regards to route, would generally just attempt to find the one with smallest increase in slope.

It should be noted that flatter terrain is also preferred by other modes. In this way bicycles are similar to trucks and buses. All three are slow to start from stop, require more effort to slow down and have difficulty climbing hills. This is the reason that trucks are given climbing lanes on steep roads, and why bicycles need climbing lanes while climbing hills, but can more easily mix with traffic when traveling down a hill.

1.8 Transit

Access to transit is one of the most important influences as to whether people will take a bus or train. If a transit stop or station is not located conveniently to people's homes or trip ends, they are less likely to rely on transit. It is well understood in transit planning that people generally will not walk more that one quarter mile to a bus stop and not more that one half of a mile to a rail station. This *Last Mile* problem is partially what makes bicycle share programs so attractive in settings where there are gaps in walkable transit access and providing *door-to-door* service (Sayarshad, Tavassoli, and Zhao 2012; Xu et al. 2013).

Because of the short distance people are willing to walk to transit, trip ends further away from transit stops and stations make people more likely to use bicycles in trips where they would combine bicycling and transit use (Andrade and Kagaya 2012). Notably, this suggests that improving bicycle facilities will raise transit use and vice versa.

Indeed, in reviewing green infrastructure in several non-US cities, Cervero and Sullivan (2011) found that roughly 30 percent of suburban commuter transit trips were made initially by bicycle to get to the rail stations. Similarly Thakuriah et al. (2012) found that new bicycle facilities that were further away from established transit lines and locations with good transit access—e.g. CBDs were more heavily utilized and incentivize greater mode shift from single occupancy vehicle use. It should be noted that the focus in connecting transit and bicycling is primarily for the commute trip purpose (O'Brien, Cheshire, and Batty 2014), as other trip purposes likely do not involve a transit hub, and, the provision of bicycle parking at transit hubs further promotes use by commuters (especially in instances where bicycles are not allowed on transit vehicles).

1.9 Air quality

It has been demonstrated that cycling in an urban environment increases bicyclists exposure to black carbon, a proxy for diesel emissions (Nwokoro et al. 2012). While the average bicyclist is not likely to spend much time contemplating the exact parts per millions of noxious fumes they are inhaling, they are likely to note that they do not enjoy being stuck behind a bus or truck at an intersection. And, it has been shown that roadways with high truck volumes reduce bicycle level-of-service (Robertson and Hawkins 2013).

Given the preference, cyclists would choose a facility away from the larger vehicles, which also has other benefits in terms of increasing the perception of safety. To be fair, not all buses and trucks are polluting and many fleets are converting to cleaner technology. Nonetheless, roadway facilities with high volumes of larger vehicles can likely be considered poorer than with lower volumes.

1.10 Bicycle facilities

All else equal, bicyclists prefer to be on off-street paths and trails (Broach, Dill, and Gliebe 2012; Kang and Fricker 2013). This is more true for novice bicyclists than experienced commuters because if an off-street path is heavily used by pedestrians it may be difficult to ride a bicycle around them and easier to ride in the street with mixed traffic. Nonetheless, off-street paths are preferable to on-street facilities.

If the bicycle facility is on the roadway, protected lanes or cycle tracks are preferable to just a signed route, which is preferred over just being on the road in mixed traffic. In places where there is a provision for protected lanes, higher lane widths are preferred to lower ones (Carter et al. 2013). This is especially true in locations where you have bicyclists travelling in two directions.

Although one-directional cycle tracts have been shown to be safer than those that allow travel in both directions, especially at intersections, both considerably improve safety and more importantly, the perception of safety (Thomas and DeRobertis 2013). It is hard to imagine that most bicyclists are aware of the location of bicycle incidents (with the exception of the locations of *ghost bikes*. However, cyclists are aware of the places that feel safer to them, and generally those places are facilities that are separated from vehicular traffic.

The addition of new facilities can also have a positive impact from a psychological standpoint. People that do not regularly ride, may begin to do so if they think an improvement in infrastructure (a new facility or safety improvements to an existing facility) have been made, this is especially true when some sort of life event has spurred those people to try bicycling (Chatterjee, Sherwin, and Jain 2013)

1.11 Pavement quality

Bicyclists, like motorists have a more pleasant experience and prefer to use well paved facilities (Carter et al. 2013). However, unlike vehicles, bicyclists are much more sensitive to poor pavement conditions because they are not insulated and even with some shocks cannot absorb the jarring bumps of large pothole or cracks well.

1.12 Sidewalks

The presence of sidewalks and a bicyclists ability to share those facilities with pedestrians has been shown to have a positive influence on bicycle utilization (Andrade and Kagaya 2012). In Seattle, bicyclists are allowed to use the side walk facilities so long as they travel at slow, reasonable speeds. As such, ceteris paribus, the presence of sidewalks on one route and not on another should make the route with a sidewalk more attractive.

1.13 Turns

Fewer turns at intersections are preferable to routes with a greater number of turns (Broach, Dill, and Gliebe 2012). Turns can be points of conflict for bicyclists, especially if they have to either 1) make a left turn from being in the right lane or protected bike lane and have to move over to a center or left lane (particularly if it is during a green light cycle) and 2) when making right turns in areas with high pedestrian volumes or if vehicles attempt to turn right across the bicycle lane.

1.14 Traffic signals

Routes with fewer traffic signals may be preferable to ones with more signals, (Broach, Dill, and Gliebe 2012) and increase bicycle level-of-service (Robertson and Hawkins 2013). Roundabout and other traffic calming devices allow bicyclists to continue their trip without stopping. While stopping may increase time, a general detriment to trips, the real problems for bicyclists are the energy exertions in starting again and having an appropriate place to stop.

1.15 Traffic volume

Routes with lower traffic volumes are preferable to ones with higher volumes (Broach, Dill, and Gliebe 2012). The more vehicles that are on the road, there is a greater risk for conflict between the modes. This is especially frightening for more novice bicyclists.

Vehicle volumes also are a good indicator for higher pollution levels (less healthy or enjoyable to ride in places where you are inhaling fumes) and also for the potential of conflicts as both bicyclists and vehicles attempt to make turns.

Similarly, a high volume of bicyclists makes a facility more attractive and can stimulate additional cycling (Vandenbulcke et al. 2011). The notion of achieving a *critical mass* normalizes cycling and makes it more inviting for new cyclists to use a route.

1.16 Traffic speed

Cyclists generally prefer facilities where vehicles travel at lower speeds, which, goes along with roadways with fewer lanes (Hood, Sall, and Charlton 2011). It is well understood that at higher vehicle speeds, the severity of collisions is much worse. Roadways with many lanes generally have higher posted speed limits and higher actual speeds. Although there is a relationship between speed limits and speeds, drivers often drive the speeds that they are most comfortable driving. Narrower lanes and roadways with fewer lanes generally force drivers to drive more slowly, or at least closer to the posted speed limit. Regardless, roadways with lower vehicle speeds are preferred by bicyclists.

1.17 Parking

The less on-street vehicle parking is available the better (Carter et al. 2013). Unless vehicle parking is set up in a way that it does not conflict with a protected bicycle facility, the less parking, the more likely a route will be used. Parking creates several problems for bicyclists. First, drivers are often not aware of their surroundings when they exit the vehicle and occasionally *door* bicyclists as they attempt to exit the vehicle and don't notice a bicyclist next to their car.

The same is also true in locations where the bike lane is next to the sidewalk with vehicle parking immediately next to the bike lane. In this case, instead of the driver opening the door into a moving bicyclists, it is someone on the passenger side.

Finally, vehicles attempting to park or leave a parking space generally infringe on the bicycle lane, especially as bicyclists may attempt to stay on the right side of the lane in cases with higher traffic volumes.

Parking also matters at trip ends. As parking becomes more costly or is unavailable, more people will use bicycles (Miller and Handy 2012). This phenomenon, like many of the cost decisions generally influences mode choice far more than route choice, unless parking constraints occur in a trip that is a part of a larger tour—e.g. if there is a choice of where to stop for coffee or groceries. But, even in the case of mode choice, bicycles will be used for tours if the tours are related to subsistence activities (Li et al. 2013).

1.18 Summary

Table 1.1 summarizes the variables that may affect bicycle route choice and their anticipated impacts.

Table 1.1: Variables affecting bicycle route choice and their anticipated impacts

Variable	Expected impact
Distance	Shorter is better
Slope	Less is better
Transit	More than .5 mile from station or stop is better
Air quality	Being on facilities with lower truck and bus volumes is better
Bicycle facilities	Off-street paths are better than protected on-street, which are better than unprotected on-street with cars; wider bicycle lanes are preferred over narrower ones.
Pavement quality	Better condition stimulates use
Sidewalks	Availability and legal use is better
Turn	Fewer turns at intersections are preferable
Traffic signals	Fewer are better
Traffic volume	Lower volumes are preferable
Vehicle speed	The lower the better
Parking	Fewer on-street vehicle parking locations promote use

Chapter 2 Study Site / Data

2.1 Study Site

The study site for this project encompasses the Central Puget Sound Region, the jurisdictional domain of the Puget Sound Regional Council. As described by the Puget Sound Regional Council (2008):

> The central Puget Sound region is one of the principal metropolitan regions in the Pacific Northwest of the United States. It includes King, Kitsap, Pierce and Snohomish counties and their 82 cities and towns, covering an area of nearly 6,300 square miles (16,300 square kilometers). The region's geography is diverse, and includes urban, rural, and resource lands. Numerous hills, mountains, and lakes provide significant variety to the topography of the region, which ranges in elevation from sea level at Puget Sound to over 14,000 feet (more than 4,000 meters) at Mount Rainier.

Table 2.1 provides population numbers for the region, and Figure 2.1 provides a context map for the region.

Table 2.1: Population of Central Puget Sound Region (Puget Sound Regional Council 2014b)

Population	2000	2010	2013
King County	1,737,000	1,931,200	1,981,900
Kitsap County	232,000	251,100	254,000
Pierce County	700,800	795,200	814,500
Snohomish County	606,000	713,300	730,500
Region Total	3,275,800	3,690,900	3,780,900

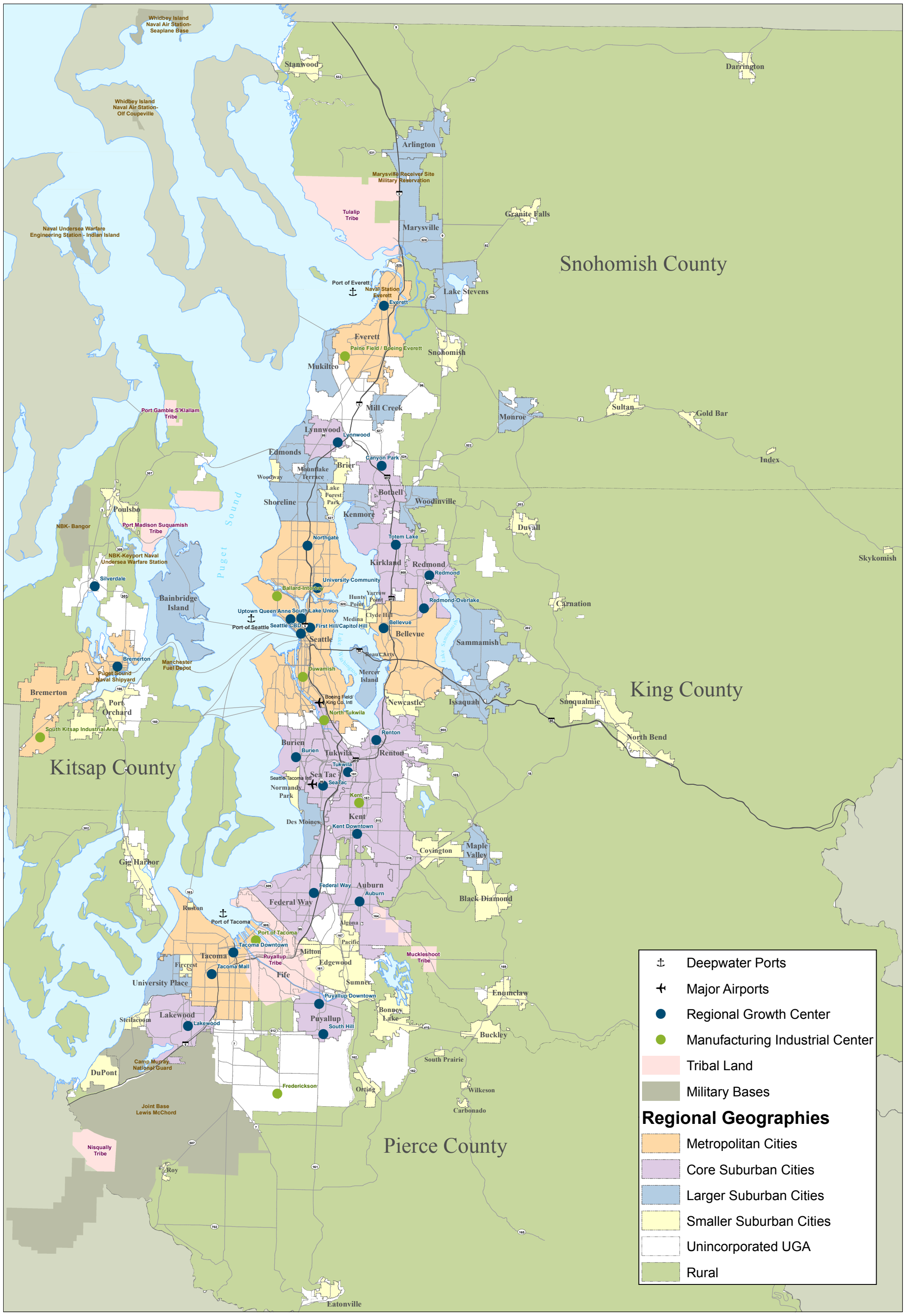


Figure 2.1: Map of the Central Puget Sound Region (Puget Sound Regional Council 2014a)

2.2 Data

This study made use of an original dataset containing GPS traces of bicycle trips collected by the Puget Sound Regional Council between April, 2012 and October, 2012. During that time, 165 unique users logged a total of 2,750 trips throughout the region.

These data include characteristics of the trip (Table 2.2) and characteristics of the individual recording the trip (Table 2.3).

Table 2.2: Trip-level variables and associated data types

Variable	Type
Unique trip identifier	Numeric
Unique user identifier	Numeric
Trace Information (each trace is a collection of points with the following characteristics):	
Latitude	Numeric
Longitude	Numeric
Altitude	Numeric
Horizontal Accuracy	Numeric
Vertical Accuracy	Numeric
Speed	Numeric
Recorded	Date-time

Table 2.3: Person-level variables and associated data types

Variable	Type
Unique user identifier	Numeric
Age	Numeric
Gender	Factor
Home ZIP code	Factor
School ZIP code	Factor
Work ZIP code	Factor
Cycling Frequency	Factor

In addition to bicycle trip data, a number of geospatial data layers were required for further analysis. These layers contain information about the region and are summarized in Table 2.4. The majority of these layers were pre-existing, and were extracted from the regional council's geo-database.

Table 2.4: Supplementary geospatial data layers

Layer	Type
PSRC road network (edges)	Polyline
PSRC road network (junctions)	Point
PSRC bicycle facilities (edges)	Polyline
PSRC bicycle facilities (junctions)	Point
PSRC network attributes	Table only
PSRC junction elevations	Table used only
PSRC edges elevations	Table used only
PSRC parcel	Multi-polygon

A full description of the sampling strategy, data collection methods, data collection tools, data cleanup, and data processing can be found in Chapter 3, Methods. A description of the sample characteristics is in Chapter 4, Results and a discussion of dataset limitations imposed by the data set is discussed in Chapter 5.

Chapter 3 Methods

3.1 Sampling and Recruitment

Individuals participating in the data collection effort were self-selected from a population of bicyclists in the Central Puget Sound region. The study was publicized through existing communications channels available to the Puget Sound Regional Council, primarily through its Bicycle and Pedestrian Advisory Committee (BPAC). Email announcements, paper flyers, and a website were developed describing the study goals and how people could participate. These materials, in turn, were sent out to to BPAC members with the request to redistribute as broadly as possible.

Other publicity venues included the regional council's Facebook and Twitter social media pages. The study was also publicized on the widely-read Seattle Transit Blog. A similar study in San Francisco by Hood, Sall, and Charlton (2011) had incentivized participation by offering prizes, however no such incentive was offered for this study.

To participate in the study, individuals were directed to download a smartphone-based survey and data collection application from either the Apple or Google application stores. Use of the application to log one or more bicycle trips constituted participation in the study.

3.2 Data Collection

Revealed preference GPS data were collected using a smartphone-based application called CycleTracks. While investigating options to collect these data from bicyclists, PSRC discovered that a similar bicycle route choice modeling effort by the San Francisco County Transportation Authority (SFCTA) had already developed an open source smartphone-based application, CycleTracks, that combined GPS data collection with additional survey questions. In addition, SFCTA had retained the back-end information technology services that store and query data collected from the smartphone application.

Based on the readiness and availability of the application, the application's zero cost to PSRC or study participants, the pre-existing back-end IT infrastructure maintained by SFCTA, SFCTA's willingness to make this service accessible to PSRC for a nominal fee, and the integration of relevant survey questions together with the tool, PSRC opted to collect these person and trip level data using the CycleTracks application. Other trip logging software including the popular RunKeeper application were considered, but due to a lack of associated sociodemographic, bicycling typing, and trip purpose data, PSRC only considered data collected via CycleTracks.

CycleTracks was designed to run on smartphones running Apple's IOS operating system as well as smartphones running Google's Android platform. The Android and IOS versions of CycleTracks differed somewhat in appearance and user interface, however, both versions collected similar data in terms of GPS traces and survey questions asked. On both platforms, as a one-time setup, users were asked to enter several questions about themselves. Person-level variables and data types are shown in Table 3.1.

In addition to the one-time setup of the application for personal characteristics, each trip required study participants to indicate the start of their trip (initiating GPS logging), the end of their trip, and the purpose of their trip. A complete set of screenshots of the Android and IOS versions of the application can be found in Appendix A.

At the conclusion of a successfully logged trip, the user's personal information, GPS trace, and trip purpose were uploaded to a server at SFCTA. The regional council was granted access to a web application at SFCTA from which these data were downloaded in a delimited text format.

Table 3.1: Person-level variables collected by CycleTracks

Variable	IOS Field Type	Android Field Type
Age	Short text	Short text
Gender	Short text	Radio button (male/female)
Cycling Frequency	Radio button (Less than once per month, several times a month, several times per week, daily)	Slider (Less than once per month, several times a month, several times per week, daily)
Home ZIP	Short text	Short text
Work ZIP	Short text	Short text
School ZIP	Short text	Short text
Email Address (was optional, not used in the Puget Sound Regional Council study)	Short text	Short text

3.3 CycleTracks Data Cleaning and Bad Trace Identification

An early step in creating the choice set used for modeling was to clean the raw cycle tracks data and remove traces that could not be used for analysis. We identified a number of issues with individual traces that needed to be dropped from the dataset. In many cases, the GPS traces were relatively clean, in that there were no clear discontinuities or other aberrations observed. These traces are characterized by clear alignment with roadways or other facilities, and a lack of single points lying far from others (see Figure 3.1).

In other cases, we observed traces that were nearly perfect, however appeared shifted from any discernible roadway or bicycling facility (see Figure 3.2). We suspect that these are the result of a facility being proximate to a hillside or other feature that may cause a reflected GPS signal. GPS is not perfectly accurate, and so some deviation from actual facilities is to be expected. The degree to which GPS traces are distant from facilities determines whether or not they can be included for analysis.

The urban canyon effect is particularly evident in traces that coincide with downtown Seattle. Reflected GPS signals results a great deal of "jumping" of GPS points (see Figure 3.3). These traces are problematic if identification of a specific route through the central business district is desired. The prevalence of this in downtown Seattle suggests treating a small area in downtown as a single destination.

One issue we encountered with the Apple IOS version of the CycleTracks application was that, at the time it was developed, IOS did not support background GPS operation. This meant that users might start logging a trip and then put their iPhone to sleep manually. For the time that the phone was asleep, it would not log GPS locations. As a result, many traces would start, and then jump to a new location when the iPhone was woken back up. Figure 3.5 illustrates an otherwise clean GPS trace with discontinuities believed to be caused by a user putting her iPhone to sleep for a portion of the trip. Trips with significant discontinuities were deemed unusable.

Sleeping was not the only condition that created discontinuities in GPS traces. Another example of this was when the smartphone would lose its GPS fix. Most smartphones use several mechanisms to locate themselves spatially including, GPS, cell tower triangulation, and proximate wi-fi access points. When a GPS signal is lost, many smartphones will devolve to triangulating an approximate location using cell phone tower signals. This results in short discontinuities where the user appears to “teleport” to a new location briefly (see Figure 3.5). If the discontinuity is brief, the individual bad points can often be removed, and the trace salvaged.

The last major case of trace issue we saw was where users started logging trips, but then immediately ended the trip (see Figure 3.6). Anecdotally, a number of users told us that they had done this out of curiosity to try out the interface of the application. These traces were dropped from the analysis.

Operationally, we identified several criteria by which a trace can be removed from consideration in our analysis:

1. Minimum bounding rectangle of less than $1024 m^2$. This threshold is somewhat arbitrary but covers the case where a user aborted logging early, but still uploaded the trip.

2. 10 or more GPS points exceed 15.65 m/s, indicating that the user spent some portion of his trip traveling by a motorized, non-bicycle mode. 10 points is an arbitrary threshold, however seems to be sufficiently high as to tolerate some incorrect data, but is sufficiently low as to exclude trips with bursts of speed deemed impossible for bicyclists to achieve.

3. Fewer than 20 GPS points exceeding 4.5 m/s. Trips in which the user did not, at any time, achieve a speed greater than 4.5 m/s were deemed most likely to be a non-bicycling trip.

4. Trips in which the time between recording GPS points exceeded 1 minutes. Many such trips exist, believed to be the result of iPhone users pausing the logging function on their phone in error. These traces are unusable because there are significant gaps in the record.

Appendix D provides example SQL code for the trace identification process.

In addition to the problem of bad traces, we noted that a number of users were very diligent in logging their trips, even when trips did not differ from day to day. Often they would log their daily commute repeatedly, which resulted in duplicative traces. Where user, origin, destination, and route are all the same, these are collapsed to a single record.

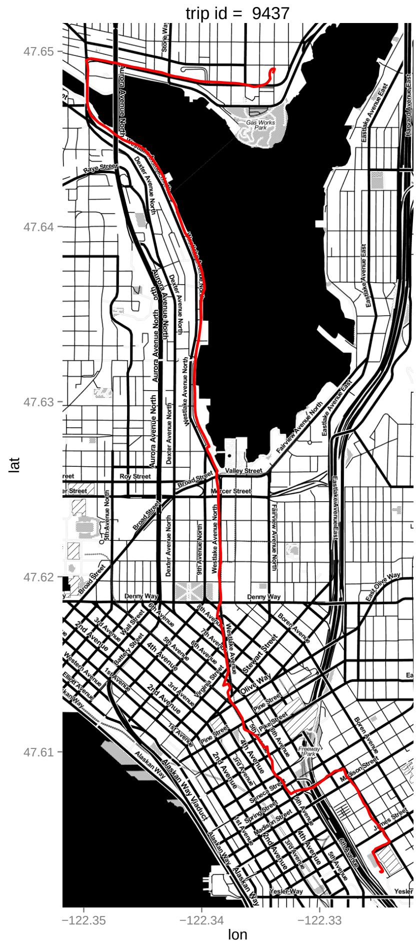


Figure 3.1: A clean GPS trace

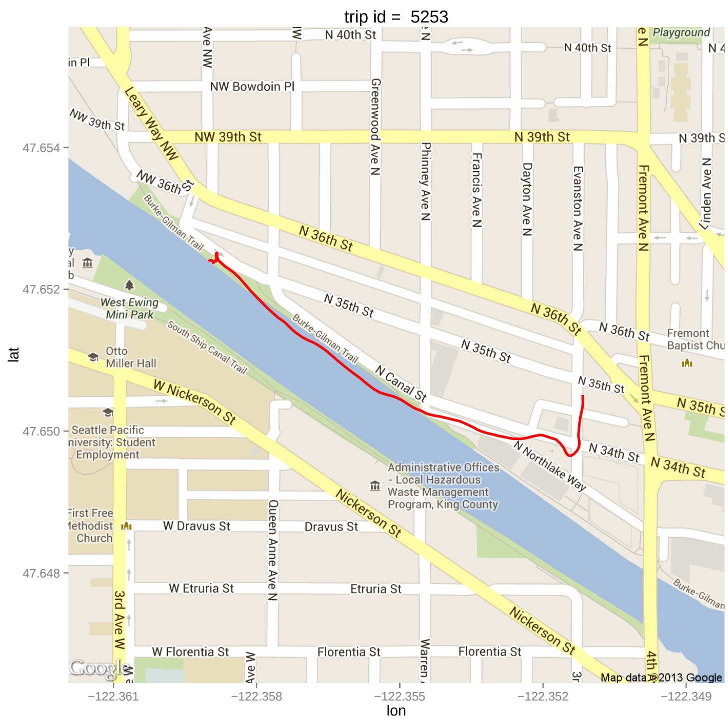


Figure 3.2: A nearly clean GPS trace paralleling a facility

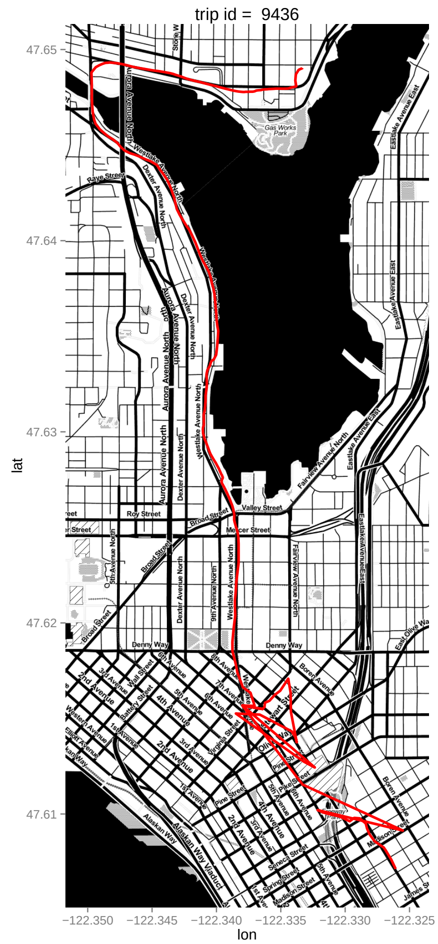


Figure 3.3: Urban canyon effect in an otherwise clean trace

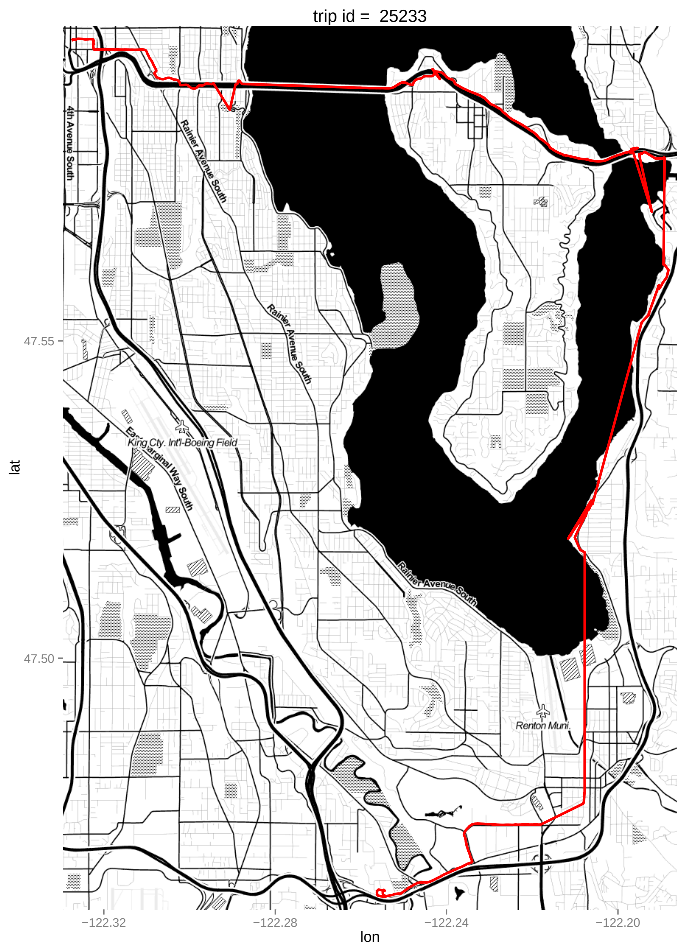


Figure 3.4: GPS stopped for part of a trace

Figure 3.5: Apparent "teleportation" where GPS loses its fix

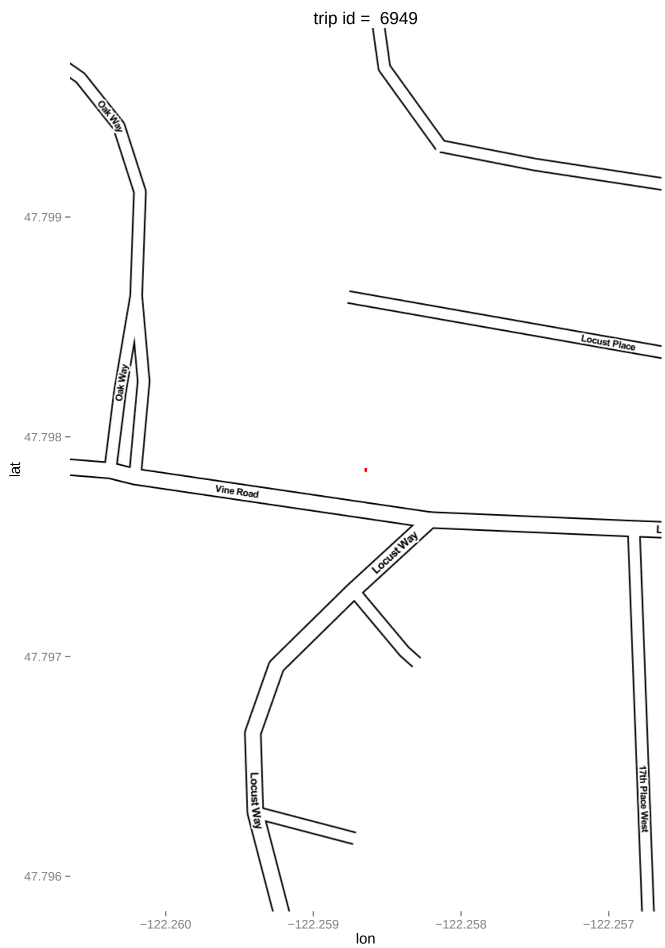


Figure 3.6: Trip logged with negligible distance traveled

3.4 Choice Set Generation

We chose to focus on developing a choice set based around the calibrated labeling method as proposed by Broach, Gliebe, and Dill (2010). As the authors of that study noted, this method has advantages both of computing tractability, as well as in providing results that may be more intuitively linked to bicyclist behavior as compared to other methods. We describe our approach in this section with the caveat that, as of this writing, choice set generation is ongoing.

The calibrated labeling method chosen for this project identifies meaningful criteria or labels for a given type of route. In the traditional labeled route method, routes are identified through a street network according to a minimization or maximization of the given label. For example, a labeled route might identify the route that maximizes the proportion of the trip ridden on on-street bicycle facilities. Conversely, another labeled route might be identified that minimizes the proportion of up-slope travel.

Typically these attributes of the network are included as a cost function, and routes are identified using an algorithm such as that developed by Dijkstra (1959). However, the calibrated labeling method extends this approach by "instead of generating a single optimal route for each attribute label, multiple optima are generated by varying the label cost function parameter. Second, the range over which the parameter varies is calibrated with the observed distribution of shortest path deviations" (Broach, Gliebe, and Dill 2010, 91).

Based on our review of the literature, considerations to limit the complexity of the model, and data availability, we identified the following set of labels for consideration:

- Minimize path distance
- Minimize proportion of up-slope per trip
- Minimize mean up-slope per trip
- Minimize total number of intersections
- Minimize number of left turns

- Minimize number of right turns
- Minimize share of trip up-slope
- Maximize share of trip on a bicycle facility
- Maximize the degree of land use mix along the route
- Minimize share of trip along commercial parcel

Procedurally, generation of the choice set entails a number of geospatial processing steps. First, the raw CycleTracks GPS data needs to be cleaned, and individual traces matched to the roadway network. Next, relevant characteristics of the road network must be determined, either from properties of the network itself (e.g. identification of left turns and right turns), or by combining other data such as parcel / land use data along corridor segments. Choice sets are ultimately created through a network process in which the labeled criteria are maximized or minimized as described previously. In addition, the calibrated labeling method requires that additional routes are identified by varying each of the labeled cost function parameters.

Our initial approach to generating a choice set involved a manual process using ESRI's ArcGIS product together with the Network Analyst toolbox. During this process we encountered several issues that ultimately caused us to seek a new approach to the geospatial processing portions of the choice set generation process. The first issue we encountered was related to the size of some of our datasets. In particular, where land-use characteristics were required, this involved the spatial intersection of roadway network segments together with the Puget Sound Regional Council's parcel layer, which spans four counties, and represents nearly a gigabyte of geometry data. In practice, we found that ArcGIS would simply crash or simply not complete many of the spatial intersect operations. The second issue we encountered was that of repeatability. Whether a defect in the tool, differences in the computer systems, or human error, we identified discrepancies in the results of our manual process that proved time consuming to resolve satisfactorily.

These issues, combined with a growing recognition that additional data collection might be undertaken in the future, and a desire to more tightly integrate the bicycle modeling effort with other aspects of the regional council's data processing and IT infrastructure, caused us to seek an automated, programmatic approach to choice set generation.

The approach we explored for these operations was to code the data processing tasks in the declarative SQL programming language, against data stored in a relational database. We chose this approach in part because we had already resorted to using the relational geodatabase Spatialite for those spatial intersection operations that we were not able to complete using ArcGIS. As a general approach for data analysis, Howe and Halperin (2012) note a number of benefits of using relational databases and specifying analyses declaratively in SQL rather than scripting procedures in other programming languages. In our early experience with this approach, we found many common spatial operations reasonably easy to specify. A benefit we found in this approach, is that large intersecting operations, such as those involving the regional council's parcel database, completed on the order of seconds rather than minutes or hours (see Appendix E for an example of a spatial intersection in SQL using Spatialite). These operations also appeared to yield consistent results, fitting with our desire for improved reproducibility.

This approach to the choice set generation is also consistent with the IT infrastructure of the regional council, and its existing geospatial data. The regional council was an early adopter of a centralized geodatabase infrastructure, and thus they have already committed to an infrastructure capable of these types of data processing tasks. Performing the analysis directly in the geodatabase has the additional benefit that it simplifies the workflow by removing the need to extract data from the geodatabase (potentially resulting in loss of topological consistency), and loading the data into another analytical platform.

One barrier to implementing these processing tasks in the regional council's database as it exists currently, is that PSRC's geodatabase is built on Microsoft SQL Server 2008, overlaid with ESRI's ArcSDE middleware product. Microsoft SQL Server 2008 lacks many of the standard geospatial datatypes and spatial processing functions that exist in newer databases. Subsequent releases of Microsoft SQL Server have adopted the Open Geospatial Consortium's (OGC) standard geodatabase extensions. In addition, all other major commercial and open source RDBMS vendors (e.g. PostGIS, Oracle) have also adopted these standards. When PSRC updates its geodatabase, either to another vendor's product or to a more recent version of Microsoft's RDBMS, the upgraded geodatabase will almost certainly support the OGC standards.

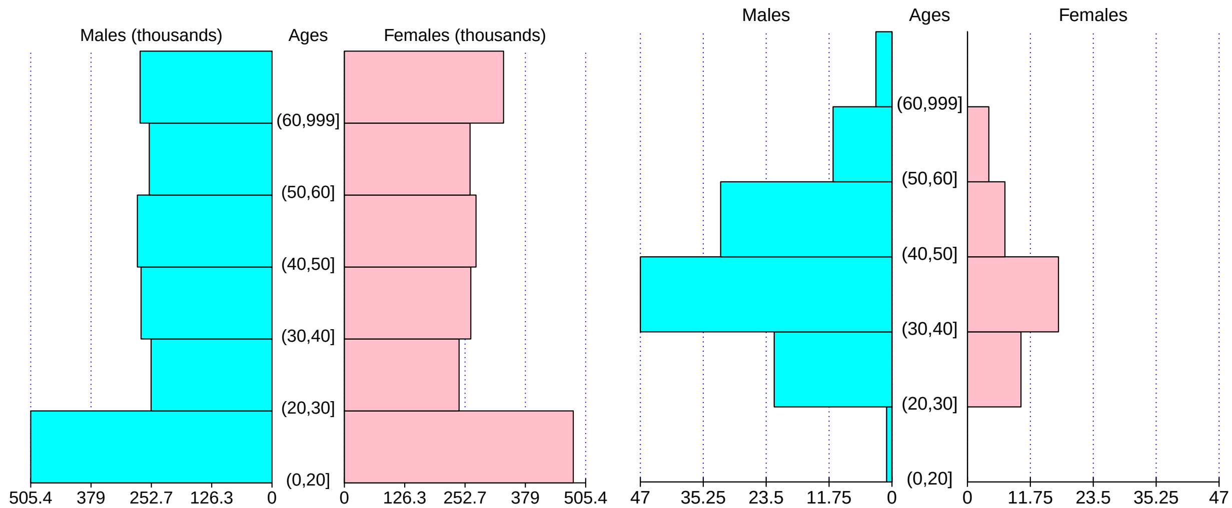
Chapter 4 Results

As the choice set generation and statistical modeling tasks are ongoing, this section is limited to a presentation of results from other phases of the study.

This study began with a literature review to identify factors influencing bicyclist route choice. In addition to the person-level characteristics collected via the CycleTracks application, from the list of factors identified in our literature review, we selected a set of labeled route criteria for choice set generation that include path distance, up-slope travel, number of intersections, number and type of turns, share of the trip on a bicycle facility, land use mix along the route, and the share of trip along commercial parcels.

Our data collection process resulted in a dataset that includes 165 unique users and a total of 2,750 trips, and somewhat fewer unique trips. From the outset of this study, we recognized that we would be working from a self-selected sample, and that there would be major limitations on our ability to generalize from our dataset. From the data collection phase of this study we observe that our final sample does not look much like the broader population of Central Puget Sound. Looking at the breakdown of ages and gender represented in our sample (see Figure 4.1), our sample may arguably look more like the population of bicyclists in the region, with the notable lack of individuals in the youngest age category. The relatively small number of female participants also poses a problem if we consider that females are an under-represented demographic in cycling, whose preferences may be important for policy makers wishing to expand rates of bicycling.

Also notable is that the vast majority of trips logged were either for work commuting purposes or school commuting purposes (see Figure 4.2). This result is not unexpected, however this makes this dataset less useful when considering facilities for recreational riding purposes.



(a) Central Puget Sound (b) CycleTracks Sample

Figure 4.1: Population Pyramids for Central Puget Sound Region vs. CycleTracks Sample

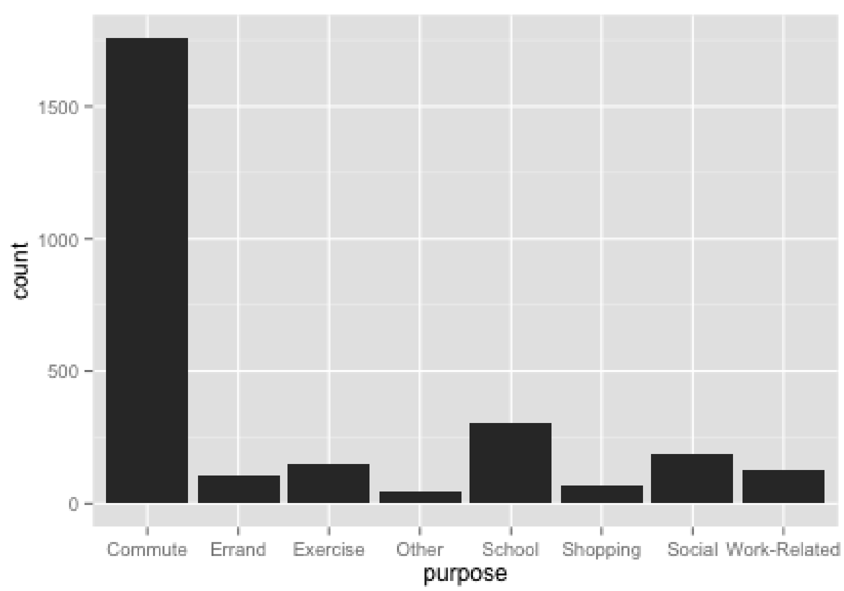


Figure 4.2: Breakdown of trips by purpose

As discussed in the context of bad trace identification and removal in Section 3.3, our data collection effort identified significant problems in consistency and cleanliness of the raw CycleTracks GPS data. In order to deal with those limitations we developed routines for use with a relational geodatabase for identifying and removing unacceptable traces.

In the process of generating a choice set for statistical modeling, we encountered issues with the data processing tasks which resulted in us identifying a process that uses a relational database to declaratively specify geospatial operations. We believe the identified process will result in a bicycle model that is reproducible, and that is better integrated with the geospatial data store of the regional council.

Chapter 5 Discussion

The dataset that resulted from our data collection effort poses significant limitations for subsequent statistical modeling. As discussed in previous sections, the self-selected recruitment of our sample is problematic if we intend to make generalizations from the data. In addition, the under-representation of females and individuals in several age categories makes it impossible to satisfactorily answer a number of questions relevant to policy makers who wish to expand bicycling among particular demographics.

The data collection process did, to an extent, validate the ability for a public agency to crowd-source data using a smartphone app. A reasonably large amount of data was collected at an extremely low cost. The cost of this data collection effort was no doubt thanks to the San Francisco County Transportation Authority's foresight in developing a robust platform for data collection, making the application free, and providing access to the back-end infrastructure to other government agencies for the marginal cost of its use. In its unmodified form, the CycleTracks application proved adequate for collecting the set of attributes that we were interested in.

One problematic aspect of the CycleTracks application was in its inability to run as a background application on Apple's IOS. This limitation, imposed by earlier versions of IOS, has since been lifted. If the data collection were repeated today, with an updated version of CycleTracks, we anticipate that we would see a better quality set of traces, given that the sleep issue was one of the more common issues with traces.

Creating the choice set for statistical modeling proved more complicated than we initially anticipated. Issues of reproducibility and dealing with data at scale forced us to abandon a more manual ad-hoc approach to the geospatial analysis, in favor of a more robust but more technically demanding process. Nevertheless, we believe this approach offers benefits beyond reproducibility, in that it will allow for better integration with future regional council geodatabase infrastructure. It will also allow for procedures generated here to be used in processing future datasets, or to more easily assess before-after performance of major new bicycling facilities.

Chapter 6 Conclusions and Recommendations

This project set out to advance the state of bicycle model development at the Puget Sound Regional Council through the collection of revealed preference bicycle GPS data, through the development of a theoretical framework for for understanding bicyclist route choice decisions, and through the creation of analytical procedures.

Through a review of the literature on bicyclist route preferences, we have identified a set of factors that the regional council can use for future modeling tasks. This set of factors includes the subset identified in previous chapters for use with the calibrated labeling method for choice set generation. In addition to being supported by the literature, these factors also take into account the availability of datasets at the regional council. We recommend that future modeling tasks consider the factors identified in this report.

The data collection process did yield a dataset containing revealed preferences for bicyclist route choice, however limitations in the sampling methodology will limit the usefulness of this dataset for modeling purposes. In this regard, the data collection process is a qualified success. The mechanisms and tools for data collection have proved successful. This is an important result given that this is the first attempt by the regional council to crowd-source a dataset in this manner. The choice (and the ability) to use a preexisting data collection tool such as CycleTracks was key to collecting this data in a cost-effective and relatively problem free manner. Future studies would do well to repeat this approach where possible.

The limitations posed by the sample could be addressed through a future data collection effort using the CycleTracks application, where users are recruited through random sampling, perhaps with some oversampling amongst demographics of interest. In practice, however, we would anticipate a significant non-response bias if this recruitment were not performed on an identified population of bicyclists. One strategy for obtaining a larger (and perhaps more realistic) sample in Central Puget Sound would be partner on a major general-audience bicycling event such as Cascade Bicycle Club's annual commuter challenge—perhaps augmenting CycleTracks to become a means for participants to log their mileage.

The unexpected difficulty in creating a choice set underscores the necessity to consider reproducible methods that can operate at scale, such as the one described in this report. Given our initial successes in carrying out portions of the geospatial analysis required to build a choice set, we recommend that the regional council carry forward the approach of building these choice sets directly in their geodatabase. Such an approach obviates exporting and reimporting geospatial data into another platform.

Bibliography

Andrade, Katia, and Seiichi Kagaya. 2012. “Investigating Behavior of Active Cyclists Influences on Bicycle Commuting.” *Transportation Research Record: Journal of the Transportation Research Board,* no. 2314:89–96. doi:10.3141/2314-12.

Broach, Joseph, Jennifer Dill, and John Gliebe. 2012. “Where Do Cyclists Ride? A Route Choice Model Developed With Revealed Preference GPS Data.” *Transportation Research Part A—Policy and Practice* 46, no. 10 (December): 1730–1740. doi:10.1016/j.tra.2012.07.005.

Broach, Joseph, John Gliebe, and Jennifer Dill. 2010. “Calibrated Labeling Method for Generating Bicyclist Route Choice Sets Incorporating Unbiased Attribute Variation.” *Transportation Research Record: Journal of the Transportation Research Board* 2197:89–97. doi:10.3141/2197-11.

Carter, Peter, Francisco Martin, Miguel Nunez, Sarah Peters, Leon Raykin, Julia Salinas, and Ronald Milam. 2013. “Complete Enough for Complete Streets? Sensitivity Testing of Multimodal Level of Service in the Highway Capacity Manual.” *Transportation Research Record: Journal of the Transportation Research Board,* no. 2395:31–40. doi:10.3141/2395-04.

Castillo-Manzano, Jose I., and Antonio Sanchez-Braza. 2013. “Can Anyone Hate The Bicycle? The Hunt For An Optimal Local Transportation Policy To Encourage Bicycle Usage.” *Environmental Politics* 22, no. 6 (November): 1010–1028. doi:10.1080/09644016.2012.740936.

Cervero, Robert, and Cathleen Sullivan. 2011. “Green TODs: Marrying Transit-Oriented Development And Green Urbanism.” *International Journal of Sustainable Development and World Ecology* 18 (3): 210–218. doi:10.1080/13504509.2011.570801.

Chatterjee, Kiron, Henrietta Sherwin, and Juliet Jain. 2013. "Triggers For Changes In Cycling: The Role Of Life Events And Modifications To The External Environment." *Journal of Transport Geography* 30 (June): 183–193. doi:10.1016/j.jtrangeo.2013.02.007.

Chatterton, Tim, and Charlie Wilson. 2014. "The 'Four Dimensions Of Behaviour' Framework: A Tool For Characterising Behaviours To Help Design Better Interventions." *Transportation Planning and Technology* 37, no. 1 (January): 38–61. doi:{10.1080/03081060.2013.850257}.

Dijkstra, E.W. 1959. "A Note on Two Problems in Connexion With Graphs." *Numerisch Mathematik* 1 (1): 269–271.

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

Dill, Jennifer, and John P Gliebe. 2008. "Understanding and measuring bicycling behavior: A focus on travel time and route choice."

Frank, Lawrence D., James F. Sallis, Terry L. Conway, James E. Chapman, Brian E. Saelens, and William Bachman. 2006. "Many Pathways from Land Use to Health: Associations between Neighborhood Walkability and Active Transportation, Body Mass Index, and Air Quality." *Journal of the American Planning Association* 72 (1): 75–87. doi:10.1080/01944360608976725.

Hood, Jeffrey, Elizabeth Sall, and Billy Charlton. 2011. "A GPS-Based Bicycle Route Choice Model For San Francisco, California." *Transportation Letters—The International Journal of Transportation Research* 3, no. 1 (January): 63–75. doi:10.3328/TL.2011.03.01.63-75.

Howe, Bill, and Daniel Halperin. 2012. "Advancing Declarative Query in the Long Tail of Science." *IEEE Data Eng. Bull.* 35 (3): 16–26.

Kang, Lei, and Jon D. Fricker. 2013. "Bicyclist Commuters' Choice Of On-Street Versus Off-Street Route Segments." *Transportation* 40, no. 5 (September): 887–902. doi:10.1007/s11116-013-9453-x.

Krizek, Kevin J, Ahmed El-Geneidy, and Kristin Thompson. 2007. "A detailed analysis of how an urban trail system affects cyclists travel." *Transportation* 34 (5): 611–624.

Li, Zhibin, Wei Wang, Chen Yang, and Guojun Jiang. 2013. "Exploring The Causal Relationship Between Bicycle Choice And Trip Chain Pattern." *Transport Policy* 29, no. SI (September): 170–177. doi:10.1016/j.tranpol.2013.06.001.

Meng, Meng, Chunfu Shao, Jingjing Zeng, and Chunjiao Dong. 2014. "A Simulation-Based Dynamic Traffic Assignment Model With Combined Modes." *PROMET—Traffic & Transportation* 26 (1): 65–73. doi:10.7307/ptt.v26i1.1252.

Miller, Joshua D., and Susan L. Handy. 2012. "Factors That Influence University Employees to Commute by Bicycle." *Transportation Research Record: Journal of the Transportation Research Board,* no. 2314:112–119. doi:10.3141/2314-15.

Miranda-Moreno, Luis F., and Thomas Nosal. 2011. "Weather or Not to Cycle Temporal Trends and Impact of Weather on Cycling in an Urban Environment." *Transportation Research Record: Journal of the Transportation Research Board,* no. 2247:42–52. doi:10.3141/2247-06.

Nkurunziza, Alphonse, Mark Zuidgeest, Mark Brussel, and Martin Van Maarseveen. 2012. "Examining The Potential For Modal Change: Motivators And Barriers For Bicycle Commuting In Dar-Es-Salaam." *Transport Policy* 24 (November): 249–259. doi:10.1016/j.tranpol.2012.09.002.

Nwokoro, Chinedu, Clare Ewin, Clare Harrison, Mubin Ibrahim, Isobel Dundas, Iain Dickson, Naseem Mushtaq, and Jonathan Grigg. 2012. "Cycling To Work In London And Inhaled Dose Of Black Carbon." *European Respiratory Journal* 40, no. 5 (November): 1091–1097. doi:10.1183/09031936.00195711.

O'Brien, Oliver, James Cheshire, and Michael Batty. 2014. "Mining Bicycle Sharing Data For Generating Insights Into Sustainable Transport Systems." *Journal Of Transport Geography* 34 (January): 262–273. doi:10.1016/j.jtrangeo.2013.06.007.

Puget Sound Regional Council. 2014a. "Map of Central Puget Sound Region." Accessed April 18. http://www.psrc.org/assets/10749/PSRCRegion.pdf.

———. 2014b. "Regional Data Profile: Population & Households." Accessed April 18. http://www.psrc.org/data/regionalprofile/regionalprofile-pop/.

———. 2008. "Vision 2040." April. Accessed April 18, 2014. http://www.psrc.org/assets/366/7293-V2040.pdf.

Robertson, James, and H. Gene Hawkins. 2013. "Shared Roadway Implementation Guidance." *Journal Of Transportation Engineering* 139, no. 8 (August): 833–839. doi:10.1061/(ASCE)TE.1943-5436.0000563.

Sallis, James F, Lawrence D Frank, Brian E Saelens, and M.Katherine Kraft. 2004. "Active transportation and physical activity: opportunities for collaboration on transportation and public health research." *Transportation Research Part A: Policy and Practice* 38 (4): 249–268. doi:10.1016/j.tra.2003.11.003.

Sayarshad, Hamidreza, Sepideh Tavassoli, and Fang Zhao. 2012. "A Multi-Periodic Optimization Formulation For Bike Planning And Bike Utilization." *Applied Mathematical Modelling* 36, no. 10 (October): 4944–4951. doi:10.1016/j.apm.2011.12.032.

Schneider, Robert J. 2013. "Theory Of Routine Mode Choice Decisions: An Operational Framework To Increase Sustainable Transportation." *Transport Policy* 25 (January): 128–137. doi:10.1016/j.tranpol.2012.10.007.

Sigurdardottir, Sigrun Birna, Sigal Kaplan, Mette Moller, and Thomas William Teasdale. 2013. "Understanding Adolescents' Intentions To Commute By Car Or Bicycle As Adults." *Transportation Research Part D—Transport And Environment* 24 (October): 1–9. doi:10.1016/j.trd.2013.04.008.

Sturm, Roland. 2002. "The Effects Of Obesity, Smoking, And Drinking On Medical Problems And Costs." *Health Affairs* 21 (2): 245–253. doi:10.1377/hlthaff.21.2.245.

Thakuriah, Piyushimita (Vonu), Paul Metaxatos, Jane Lin, and Elizabeth Jensen. 2012. "An Examination Of Factors Affecting Propensities To Use Bicycle And Pedestrian Facilities In Suburban Locations." *Transportation Research Part D—Transport And Environment* 17, no. 4 (June): 341–348. doi:10.1016/j.trd.2012.01.006.

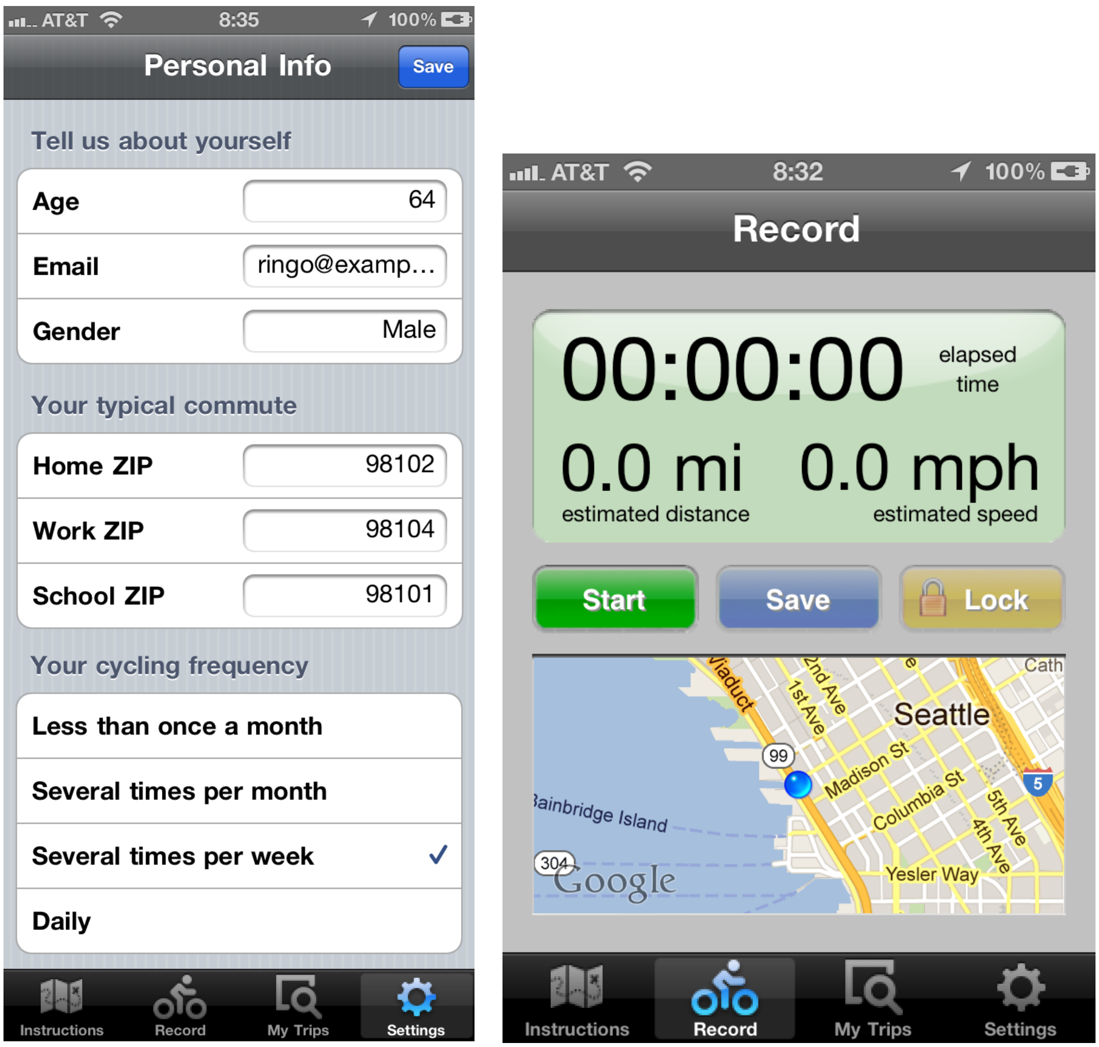
Thomas, Beth, and Michelle DeRobertis. 2013. "The Safety Of Urban Cycle Tracks: A Review Of The Literature." *Accident Analysis And Prevention* 52 (March): 219–227. doi:10.1016/j.aap.2012.12.017.

Tilahun, Nebiyou Y., David M. Levinson, and Kevin J. Krizek. 2007. "Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey." *Transportation Research Part A: Policy and Practice* 41 (4): 287–301. doi:10.1016/j.tra.2006.09.007.

Vandenbulcke, Gregory, Claire Dujardin, Isabelle Thomas, Bas de Geus, Bart Degraeuwe, Romain Meeusen, and Luc Int Panis. 2011. "Cycle Commuting In Belgium: Spatial Determinants And 'Re-Cycling' Strategies." *Transportation Research Part A—Policy And Practice* 45, no. 2 (February): 118–137. doi:10.1016/j.tra.2010.11.004.

Xu, Haitao, Jing Ying, Hao Wu, and Fei Lin. 2013. "Public Bicycle Traffic Flow Prediction based on a Hybrid Model." *Applied Mathematics & Information Sciences* 7, no. 2 (March): 667–674.

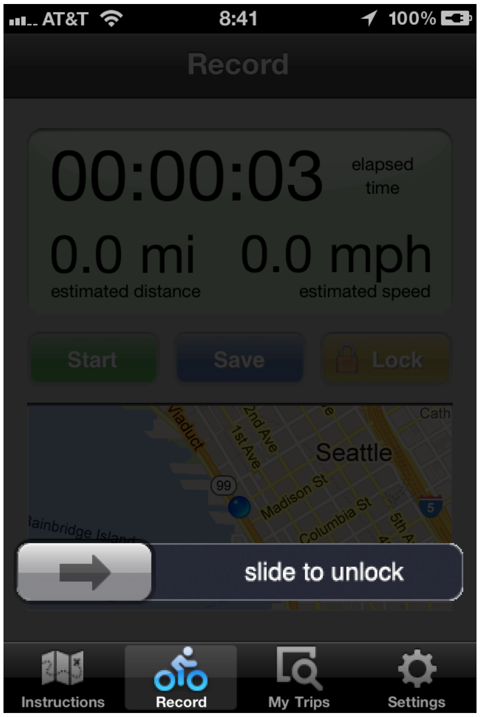
Appendix A CycleTracks Survey Tool Interface



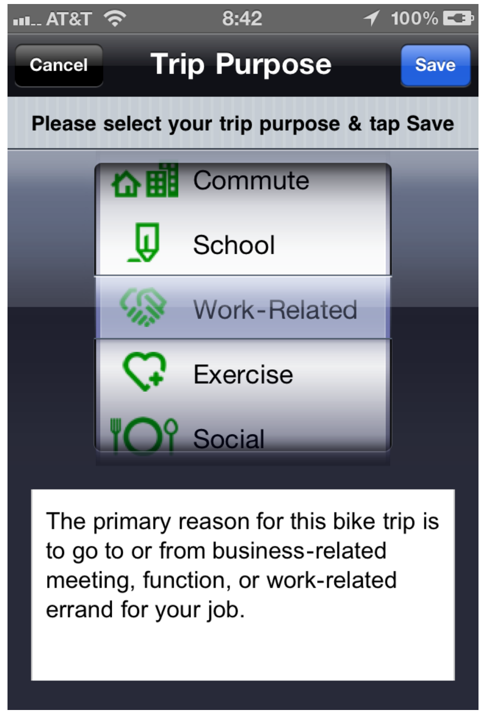
(a) Collecting personal information

(b) Trip recording interface

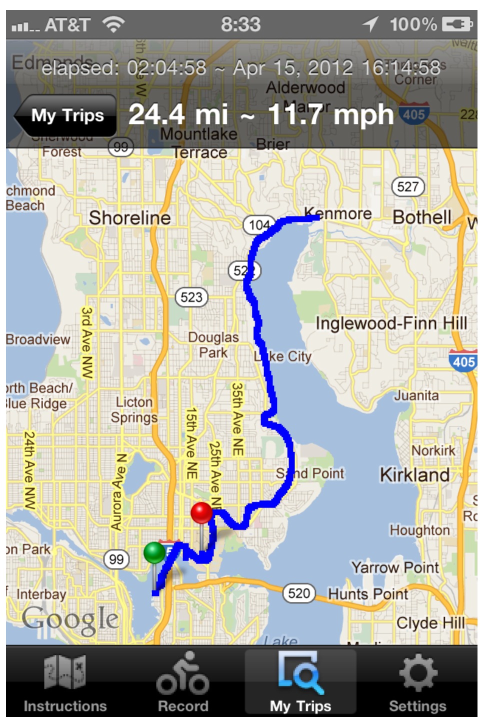
Figure A.1: Personal information and trip recording screens from IOS CycleTracks Application



(a) Lock screen

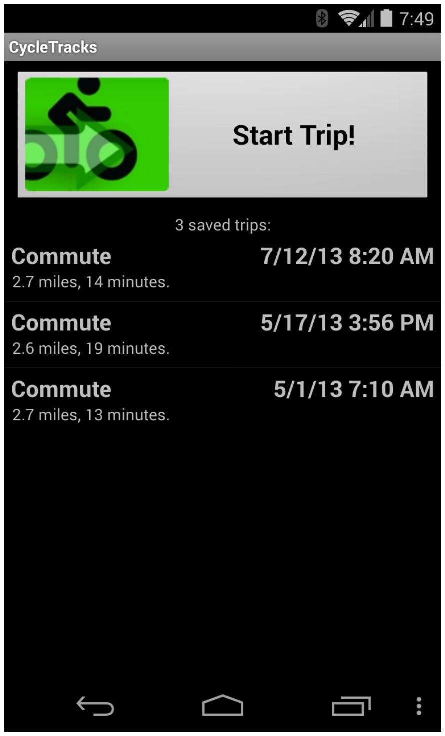


(b) Trip purpose

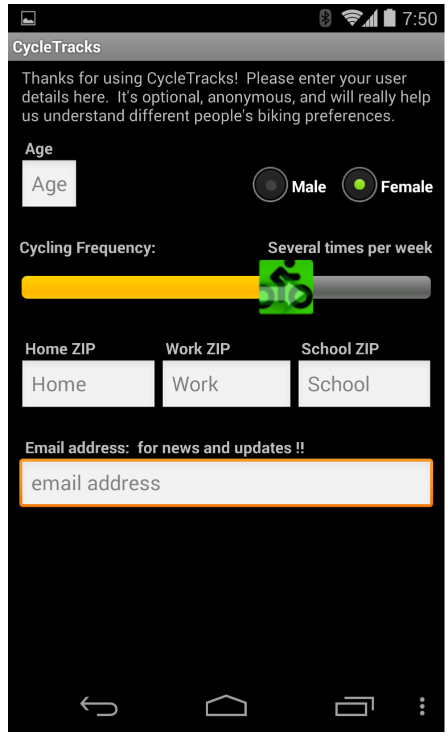


(c) Completed trip trace

Figure A.2: Lock screen, trip purpose, and trace screens from IOS CycleTracks Application



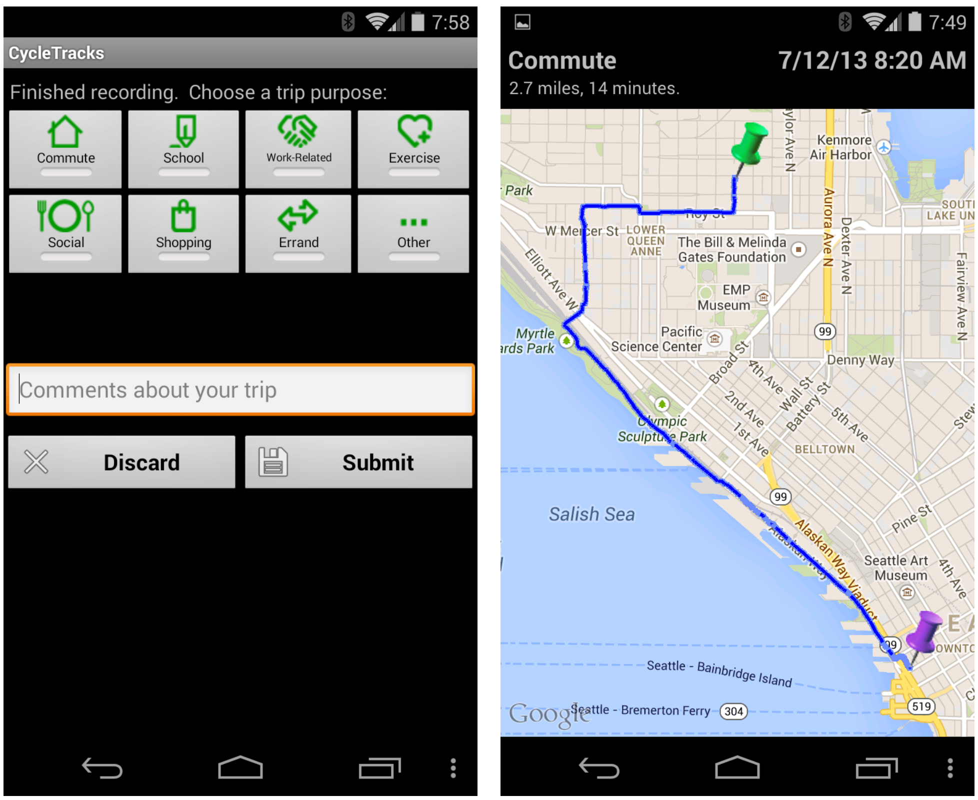
(a) Main screen



(b) Personal information

(c) Trip recording

Figure A.3: Main screen, personal information, and trip recording screens from Android CycleTracks Application



(a) Trip purpose (b) Completed trace

Figure A.4: Trip purpose and completed trace screens from Android CycleTracks Application

Appendix B Geospatial Data Extract Transform and Load Procedures

Geospatial data were originally provided in a variety of formats, including shapefiles and ESRI personal geodatabases. The following shell script demonstrates the use of ogr2ogr for loading layers from a variety of sources, transforming those layers into a consistent datum, and then loading the layers into a geospatial database—in this case, Spatialite.

```bash
#!/bin/bash
# Extract transform and load procedures

# Make for nominal portability to PostGIS
ofmt="-f SQLite"
basepath=~/Desktop/ct-postproc/data
orig=$basepath/original
opath=$basepath/cycletracks.sqlite
flags="-t_srs EPSG:3857 -a_srs EPSG:3857 -dsco SPATIALITE=YES -gt 65536 -nlt PROMOTE_TO_MULTI"
append=$flags" -append -update"

rm $opath

# Full PSRC roads network (TransRefEdges and TransRefJunctions)
echo "psrc network"
ogr2ogr $ofmt $opath $orig/PSRC_Net.gdb $flags

# Bicycle facilities - FAILS (need to bring up to speed in ArcCatalog)
echo "bike facilities"
ogr2ogr $ofmt $opath $orig/bikefacility/bikefacilities.gdb \
```

```
transrefedges_bikefacil -nln bike_transrefedges $append

ogr2ogr $ofmt $opath $orig/bikefacility/bikefacilities.gdb \
transrefjunctions -nln bike_transrefjunctions $append

# CycleTracks Traces. This is a little tricker. We create a virtual
# OGR layer in order to project WGS84 lat/lon on the fly.
vrtrace=$basepath/traces.vrt
echo "traces"
echo "<OGRVRTDataSource>
    <OGRVRTLayer name='traces'>
        <SrcDataSource>$orig/cycletracks/traces.csv</SrcDataSource>
        <GeometryType>wkbPoint</GeometryType>
        <LayerSRS>WGS84</LayerSRS>
        <GeometryField encoding='PointFromColumns' x='Longitude' y='Latitude
   '/>
    </OGRVRTLayer>
</OGRVRTDataSource>" > $vrtrace

ogr2ogr $ofmt $opath $vrtrace $append
rm $vrtrace

# CycleTracks Trips (fake SRS again)
echo "trips"
ogr2ogr $ofmt $opath \
$orig/cycletracks/trips.csv $append -s_srs EPSG:3857


# Mode attributes - table only, fake input srs
echo "mode attributes"
ogr2ogr $ofmt $opath $orig/NetAtts.gdb \
$append -s_srs EPSG:3857 -nln network_attributes
```

```
# Junction elevations
echo "junction elevations"
ogr2ogr $ofmt $opath $orig/network/junction_elevation.gdb \
$append -nln elevation_junctions

# TransRefEdges with elevation
echo "edges elevations"
ogr2ogr $ofmt $opath $orig/network/peters_transrefedges.gdb \
$append -nln elevation_edges

# Parcels; appending -skipfailures, but should validate we're not
# losing much valid data, though looks like only one geometry.
echo "parcels"
ogr2ogr $ofmt $opath $orig/parcel/prcl05/prcl05.shp \
$append -skipfailures -nln parcels_2005
```

Spatialite, while convenient for exploratory analysis by an individual researcher may be less reflective of an enterprise geospatial RDBMS than a client-server oriented system. The following demonstrates the same extract transform and load procedures using PostGIS, which may more closely mirror PSRC's MS SQL Server deployment.

```
#!/bin/bash
# Extract transform and load procedures

# First drop the cycletracks database if it exists. This is a clean
# load of the data.
psql -d postgres -c "DROP DATABASE cycletracks;"
createdb cycletracks
psql -d cycletracks -c "CREATE EXTENSION postgis;"

ofmt='-f PostgreSQL PG:dbname=cycletracks'
basepath=~/Desktop/ct-postproc/data
orig=$basepath/original
```

```
flags="-t_srs EPSG:4326 -a_srs EPSG:4326 -nlt PROMOTE_TO_MULTI"
append=$flags


# Full PSRC roads network (TransRefEdges and TransRefJunctions)
echo "psrc network"
ogr2ogr $ofmt $orig/PSRC_Net.gdb $flags

# Bicycle facilities
echo "bike facilities"
ogr2ogr $ofmt $orig/bikefacility/bikefacilities.gdb transrefedges_bikefacil \
-nln bike_transrefedges $flags

ogr2ogr $ofmt $orig/bikefacility/bikefacilities.gdb transrefjunctions \
-nln bike_transrefjunctions $flags

# CycleTracks Traces. This is a little tricker. We create a virtual
# OGR layer in order to project WGS84 lat/lon on the fly.
vrtrace=$basepath/traces.vrt
echo "traces"
echo "<OGRVRTDataSource>
    <OGRVRTLayer name='traces'>
        <SrcDataSource>$orig/cycletracks/traces.csv</SrcDataSource>
        <GeometryType>wkbPoint</GeometryType>
        <LayerSRS>WGS84</LayerSRS>
        <GeometryField encoding='PointFromColumns' x='Longitude' y='Latitude
    '/>
    </OGRVRTLayer>
</OGRVRTDataSource>" > $vrtrace

ogr2ogr $ofmt $vrtrace $flags -lco \
COLUMN_TYPES:"enabled=boolean,latitude=numeric,longitude=numeric,altitude=
    numeric,haccuracy=numeric,vaccuracy=numeric,speed=numeric,recorded=
```

```
    timestamp"
rm $vrtrace

# CycleTracks Trips (fake SRS again)
echo "trips"
ogr2ogr $ofmt \
$orig/cycletracks/trips.csv $flags -s_srs EPSG:4326


# Mode attributes - table only, fake input srs
echo "mode attributes"
ogr2ogr $ofmt $orig/NetAtts.gdb \
$flags -s_srs EPSG:4326 -nln network_attributes

# Junction elevations
echo "junction elevations"
ogr2ogr $ofmt $orig/network/junction_elevation.gdb \
$flags -nln elevation_junctions

# TransRefEdges with elevation
echo "edges elevations"
ogr2ogr $ofmt $orig/network/peters_transrefedges.gdb \
$flags -nln elevation_edges

# Parcels; appending -skipfailures, but should validate we're not
# losing much valid data, though looks like only one geometry.
echo "parcels"
ogr2ogr $ofmt $orig/parcel/prcl05/prcl05.shp \
$flags -skipfailures -nln parcels_2005
```

Appendix C Identifying GPS traces on a roadway network

```sql
-- Clear out old versions if they exist
SELECT DisableSpatialIndex('transrefjunctions_50m', 'Geometry');

DROP TABLE transrefjunctions_50m;
DROP TABLE junctions_intersect_tracepoints;

-- Create a buffered version of transrefjunctions and enable RTree
-- spatial index.
CREATE TABLE transrefjunctions_50m AS
SELECT psrcjunctid, ST_Buffer(Geometry, 50) Geometry
FROM transrefjunctions;

SELECT RecoverGeometryColumn('transrefjunctions_50m', 'Geometry',
 3857, 'POLYGON', 2);

SELECT CreateSpatialIndex('transrefjunctions_50m', 'Geometry');

-- Test for intersection of buffered transrefjunctions and gps points
CREATE TABLE gps_intersect_transrefjunctions_50m AS
SELECT b.trip_id trip_id, a.psrcjunctid psrcjunctid,
        b.OGC_FID OGC_FID
FROM transrefjunctions_50m a, traces b
WHERE ST_Intersects(a.Geometry, b.Geometry)
AND b.ROWID IN
     (SELECT ROWID FROM SpatialIndex
     WHERE f_table_name = "traces" and search_frame = a.GEOMETRY);
```

```
-- Get the number of GPS points by trip and intersected junction
CREATE VIEW points_by_trip_junction AS
SELECT trip_id, psrcjunctid, count(OGC_FID) npoints
FROM gps_intersect_transrefjunctions_50m
GROUP BY trip_id, psrcjunctid;

-- Order junctions
CREATE TABLE ordered_junctions AS
SELECT b.trip_id trip_id, a.psrcjunctid psrcjunctid,
        count(b.OGC_FID) npoints,
        min(strftime('%s', b.recorded)) mintime,
        max(strftime('%s', b.recorded)) maxtime
FROM transrefjunctions_50m a, traces b
WHERE ST_Intersects(a.Geometry, b.Geometry)
AND b.ROWID IN
    (SELECT ROWID FROM SpatialIndex
     WHERE f_table_name = "traces" AND search_frame = a.GEOMETRY)
GROUP BY b.trip_id, a.psrcjunctid
ORDER BY mintime;
```

Appendix D Identification of Bad CycleTracks GPS Traces (Spatialite)

```sql
-- Drop traces_filter table if it already exists
DROP VIEW traces_filter_mbr;
DROP VIEW traces_filter_hi_speed;
DROP VIEW traces_filter_lo_speed;
DROP VIEW traces_filter_gaps;

DROP TABLE traces_bad;

-- -- Remove the spatial views that have already been registered
-- DELETE FROM views_geometry_columns
-- WHERE view_name = 'traces_bad';

-- Identify those traces:
-- 1.) whose minimum bounding box area is < 1,024m
-- 2.) that contain > 10 points exceeding 15.65 m/s (NOTE: review
-- thresholds!)
-- 3.) that contain < 20 points exceeding minimum 4.5 m/s threshold
-- 4.) that have gaps in the recording of points that exceed 1 minute.
CREATE VIEW traces_filter_mbr AS
SELECT trip_id, Area(Extent(Geometry)) < 1024 AS mbr_lt_1024
FROM traces
GROUP BY trip_id;

CREATE VIEW traces_filter_hi_speed AS
SELECT trip_id, count(trip_id) > 10 AS excess_speed
FROM traces
```

```sql
WHERE speed > 15.65
GROUP BY trip_id;

CREATE VIEW traces_filter_lo_speed AS
SELECT trip_id, count(trip_id) < 20 AS insufficient_speed
FROM traces
WHERE speed > 4.5
GROUP BY trip_id;

CREATE VIEW traces_filter_gaps AS
SELECT DISTINCT trip_id,
(SELECT strftime('%s', b.recorded)
     FROM traces b
     WHERE b.OGC_FID = a.OGC_FID + 1
     AND a.trip_id = b.trip_id) - strftime('%s', a.recorded) > 60 gaps
FROM traces a
WHERE gaps = 1;

-- Create a table containing all the misfit links for inspection
-- purposes (i.e. weed out false positives)
CREATE TABLE traces_bad AS
SELECT a.rowid AS rowid, a.Geometry AS Geometry,
    a.trip_id AS trip_id,
    b.mbr_lt_1024 AS mbr_lt_1024,
    c.excess_speed AS excess_speed,
    d.insufficient_speed AS insufficient_speed,
    e.gaps AS gaps
FROM traces AS a
LEFT JOIN traces_filter_mbr AS b USING (trip_id)
LEFT JOIN traces_filter_hi_speed AS c USING (trip_id)
LEFT JOIN traces_filter_lo_speed AS d USING (trip_id)
LEFT JOIN traces_filter_gaps AS e USING (trip_id)
WHERE b.mbr_lt_1024 = 1
```

```
OR c.excess_speed = 1
OR d.insufficient_speed = 1
OR e.gaps = 1;

SELECT RecoverGeometryColumn('traces_bad', 'Geometry', 3857,
'POINT',2);
```

Appendix E Example Spatial Intersection Operation in SQL (Spatialite)

```sql
SELECT *
FROM layer1, layer2
WHERE ST_Intersects(layer1.GEOMETRY, layer2.GEOMETRY)
AND layer2.ROWID IN (SELECT ROWID FROM SpatialIndex WHERE f_table_name="layer2" AND search_frame=layer1.GEOMETRY)
```