Modeling route choice of utilitarian bikeshare users with GPS data

Ranjit Khatri
Email: rkhatri@vols.utk.edu
The University of Tennessee, Knoxville
311 John D. Tickle Building
Knoxville, Tennessee 37996-2313

Christopher R. Cherry
Email: cherry@utk.edu
The University of Tennessee, Knoxville
321 John D. Tickle Building
Knoxville, Tennessee 37996-2313

Shashi S. Nambisan
Email: shashi@utk.edu
The University of Tennessee, Knoxville
320 John D. Tickle Building
Knoxville, Tennessee 37996-2313

Lee D. Han
Email: lhan@utk.edu
The University of Tennessee, Knoxville
319 John D. Tickle Building
Knoxville, Tennessee 37996-2313

Preprint please cite:

Khatri, R., C. Cherry, S. Nambisan, L. Han (2016) Modeling route choice of bikeshare users with GPS data. Transportation Research Record: Journal of the Transportation Research Board. 2587. Doi:10.3141/2587-17.

6,385 words + 2 figures (500 words) + 2 tables (500 words) = 7,385 words

ABSTRACT

Understanding bicyclist's route choice is a difficult problem given the many factors that influence attractiveness of different routes. The advent of low-cost GPS devices has made route choice analysis more precise. Bikeshare, with instrumented bikes, allows for better assessment of revealed route preference of a large sub-population of cyclists. In this paper, we used GPS data obtained from 9,101 trips made by 1,866 bikeshare users from Grid Bikeshare in Phoenix, Arizona. This unique bikeshare system relies on Social Bicycles' onboard telematics that allows non-station origins and destinations and operates on a grid street network, both enabling unique route choice analysis. The trips only include direct utilitarian trips; removing circuitous trips that could include multiple destinations or be recreational trips. The analysis focused on facility usage assessment and route choice behavior. The results were compared between two categories of bikeshare users, registered users and casual users. Registered users made shorter trips including roads with low volume and preferred bike specific infrastructure A Path Size Logit Model was used to model route choice. Riders were very sensitive to travel distance, with little deviation from the shortest path to utilize more bike-friendly infrastructure. Travel on the bike-specific facilities is equivalent to decreasing distance by 44.9% (53.3% for casual users). Left turns imposed higher disutility compared to right turns for casual users. The proportion of one-way segments, AADT, and length of the trip have a negative influence on route choice and a number of signalized intersections have a positive influence on selecting a route. The results were also compared with previous studies.

INTRODUCTION

Bicycle use has grown in most North American and European cities in the past decade. A 46% upsurge in bicycle commuting has been seen in the United States from 2005 to 2013 (*1*). This increase could be attributed to the growing concerns among public over the lack of physical activity, increased auto dependency resulting degraded air quality, and congestion that results in environmental, social and economic costs. Hence, investment on bicycling could result in health care cost savings, fuel savings, and reduced emissions (*2*) in addition to an increase in commute mode share (*3*).

In the course of understanding the riding behavior of cyclists, there have been several efforts to determine route choice behavior. Two main approaches to predict factors influencing route choice behavior of cyclists hinge on either stated preference (SP) data (*4-6*) or revealed preference (RP) data (*7-9*). Most of these studies have focused on the presence of various bike-specific infrastructure, route attributes, individual characteristics, land use and so on. There are numerous studies using SP data that exploits the advantages such as ease of data collection and simplicity in modeling. Typical SP surveys will allow the participants to rate different type of factors and choose among different side-by-side alternative routes. Most of these studies attempt to model behavioral intent and are inflicted by the possibility of bias from their actual behavior (*4*). The Recent development of low-cost GPS has made an accurate collection of the actual routes possible. Some studies collect data through GPS installed on the bike of the participants (*10*), whereas other studies make use of the smartphone applications (*7, 8*) to collect data. In addition to reduced burden on the participants to remember the route, RP study is low-cost, efficient in mapping the route and determining attributes of the route.

Previous route choice models, either based on SP or RP, have consistent findings on some of the factors that influence route choice decisions for bicyclists, like distance, safety, turn frequency, road grade, intersection control, traffic volumes, land use and aesthetics along the route (*5, 7, 9, 11, 12*). This behavior is inconsistent with the driver's route choice, who generally chose routes based primarily on distance and duration of travel. In general, travel time and suitability of the route remain two major objectives while selecting a route (*11*). Of all route attributes, provision of facilities dedicated to cycling have been portrayed as a factor that induces new cyclists, in addition to encouraging existing cyclists (*5, 13*). Some studies found bike lanes to be superior to all other bike facilities, from a user perspective (*12*), while others found off-street bike facilities were valued more than other bike facilities. Different from most of the studies, another study found that longer the length of the bike facility, higher is the deviation from shortest route to use them (*14*). Proximity to bicycle facilities was another factor that induces use of bicycle infrastructure (*14*).

Planners and engineers require quality bicycling data, preferably unbiased by self-selection, to understand the behavior of cyclists. These are complemented by new methods of real-time data collection using dedicated GPS devices or built-in GPS in smartphones. These have facilitated researchers and practitioners with new techniques to assess the route choice and behavior of cyclists on the road. Some cyclists are using smartphone applications such as Strava, MapMyRide, CycleMaps or other fitness tracking applications to record and track their data in order to encourage physical activity (*15*). However, those data sources are not usually accessible to planners. Leveraging this technology, some cities are utilizing GPS data collection techniques from open source applications like Cycle Tracks (*12*). These data collection techniques utilize built-in GPS capabilities of smartphones, which provides high quality revealed data at a reduced cost compared to SP surveys. The collected data is sent to remote servers without any requirement to go to field to retrieve the data. Several

applications are being used by the cities of US and Canada, like Cycle Tracks (San Francisco, Calif.), Cycle Atlanta (Atlanta, Georgia), CyclePhilly (Philadelphia, Penn.), My ResoVelo (Montreal, Quebec), and I Bike KNX (Knoxville, Tenn.). The data from these apps can inform transportation planning in these cities and allows for disaggregate analysis. One of the challenges with app-based data collection is that users have to opt-in and use the application for every trip.

In the last decade, bikesharing system has gained popularity in many North American cities along with the other major cities in the world. There are more than one million bicycles under bikesharing scheme in more than 500 cities of 49 countries (*16*). This allows individuals to use a bicycle for a short period of time between fixed bikeshare stations. Some, like Grid Bikeshare in Phoenix, Ariz., have facilitated the use of public racks as the bike stations, removing the requirement to return bikes to bikeshare stations. Bikeshare is meant for efficient short-distance travel, thus solving the "first/last mile problem" by connecting to other modes or providing urban circulation. Furthermore, bikeshare is meant for inducing individuals to cycle and increasing total bicycle trips in a city. Although it can be beneficial in reducing car use and increasing bicycle trips, some results suggest that bikeshare replaces mostly public transit trips and walk trips rather than car trips (*17, 18*). In addition to expanding docking stations and making convenient use of bikeshare, high substitution of car trips could be obtained by making the travel time of bikeshare trips competitive to that of car trip by achieving efficient routing or improving bicycle amenities (*18*).

Bikeshare systems are ripe for developing new data streams to understand bicycling behavior in cities. Several recent studies have mined bikeshare data to understand flows between stations and identify differences in user types (*19*). Bikeshare users are generally classified as registered users (frequent users who subscribe to a membership that usually includes unlimited use for the duration of the membership) and casual users (occasional users who pay for service as they use it, often tourists). Unlike the casual users, who primarily make recreational trips, commuting is the main purpose for registered users (*18*).

Most of the previous literature on bikeshare users focus on the demographics of users (*20*), or station or system performance (*19, 21*). Recent bikeshare systems have included vehicle-tracking telematics onboard the bicycle, which allows for a finer level of analysis, i.e., vehicle-level analysis instead of station-level of analysis. This has opened a new opportunity to investigate route choice, particularly as it relates to safety and comfort, of an entire sub-population of bicyclists, bikeshare users. This subpopulation is an important group because it constitutes a large portion of urban cyclists and represents an important part of the travel trip, generally short urban center trips. To the author's knowledge, there is no study based on the real-time GPS data of bicyclists in bikeshare systems. Although there are many route choice models trying to describe the conventional bicyclists' trip patterns, understanding the decision pattern of the bikeshare users is an important aspect of the route choice question. This study relies on data from one of the first GPS-enabled bikeshare systems in North America, the Grid Bikeshare system in Phoenix, Ariz. This system is unique because it relies on Social Bicycles' (SoBi) onboard telematics, it utilizes a more flexible station and pricing protocol (e.g., users are not required to return bikes to stations), and it is deployed in a city with a grid street network that provides many possible route choices. We investigate and model bikeshare riders' route choice and identify factors that influence that choice. This is done using GPS tracks for each trip and conducting GIS-based analysis to create alternative routes, similar to other studies, but with a more robust dataset and user type. The rest of the paper presents the methodology that describes the data and modeling approach, the results of the model, and conclusions and recommendations.

METHODOLOGY

Data Description

Data were collected from Grid Bikeshare, which began operation in Phoenix, Ariz. in 2014. The Grid Bikeshare system was installed in Fall 2014 and includes about 500 bikes and 39 stations (or hubs). The stations cover an area that is approximately 2.5 km East to West and 8 km North to South, covering downtown Phoenix. The system is also in the process of expanding to Tempe and Mesa. Although the system relies heavily on stations, users can also park bikes away from stations for a small fee. The target population for the study was all registered users cyclists who either register monthly/annually for the bikeshare or are casual users paying a marginal per-trip fee to rent a bike. The total dataset is segmented intro two general categories: registered and casual users. The data used in this study include trips made from November 2014 to May 2015, which includes 9,101 trips made by 1,866 users after data cleaning. The available GPS data, collected by the telematics system, did not have timestamps for each point and were not uniformly spaced in terms of time or distance. However, date and time of the start and end of each trip were known. The frequency of the GPS readings varied from 1 per minute to 25 per minute.

Data Cleaning

For each trip, data were collected using GPS devices. GPS data logging frequency varied but their sub-minute resolution allowed reasonable route assignment. Raw GPS data obtained were cleaned for further analysis in order to prevent any incorrect interpretation from the results of the study. The GPS data includes errors, which could be associated with urban canyons, unavailability of satellites, quality of GPS unit, and others. In addition to removing the "error points", another main objective of the data cleaning is to remove all the possible recreational trips. With a high number of trips made on weekends [Figure 1(c)], it becomes necessary to remove possible recreational trips. This was done for the current scope of analysis because bicycle trips for recreational purposes are very different from the utilitarian trips. For instance, recreational cyclists might be using longer route including bicycle-specific facilities without apparent destinations. Also, many recreational trips returned to the origin, or included loops, making route assignment and identification of alternate routes challenging at best. The following are the two basic criteria for the data cleaning process.

1. Trips with following criteria were removed
    a. Travel Time < 1 min
    b. Travel Time > 90 minutes
    c. Travel distance < 0.02 miles
    d. Travel distance > 10 miles
    e. Average velocity <1.5 mph
    f. Average velocity > 25 mph
    g. Trips having fewer than 10 GPS points

2. Trips based on the origin and destination and shortest distance were removed to eliminate circuitous tours that were not likely destined for a specific place.
    a. Trip distance > 3 × Euclidean O-D distance
    b. Trip distance > 2.5 × shortest possible travel distance between the O-D pair

There were 20,468 trips in the raw data. Using first criteria mentioned above, 3,925 trips (20% of 20,468) were removed. For the remaining 16,543 trips, criteria 2(a) removed approximately 25% of the remaining trips. There were only additional 71 trips deleted from

the criteria 2(b), as most of the trips satisfying criteria 2(b) also satisfied criteria 2(a), and were previously removed. There was a change in demographics of trips after data cleaning. For casual members, the percentage of users, total miles traveled, and the number of trips were reduced from 92% to 85%, 77% to 63% and 68% to 56%, respectively. For registered members, the percentage of a number of users, total miles traveled, and the number of trips increased from 8% to 15%, 32% to 44% and 23% to 37%, respectively. The majority of the trips removed were casual trips.

Completing the road network

The road network in a GIS environment was provided by the Maricopa Association of Governments. It included attributes for roadway segments that are of interest to this study (e.g., Average Annual Daily Traffic (AADT), geometry, and bike-specific facilities.). We supplemented that spatial dataset with crash data, speed limits and locations of signalized intersections. The GPS data were matched to the road network after cleaning. For that purpose, the road network had to be supplemented by additional links to predict the path of the cyclists. In contrast to the motor vehicle drivers, the path followed by bicyclists includes those links, which may not be present in the base network, such as parking facilities, alleys, or shared use paths. Hence, all the additional or missing links were added using an ArcGIS interface to include all the links used for bike travel.

Map matching

The raw GPS coordinates available from the bikeshare users were matched to the street segments to identify all the links that were traversed during the trip. However, it is difficult to estimate the path with high accuracy. Key reasons behind this are the inaccuracy of the data points and the use of the sidewalks, parking lots and alleys, which are not represented as the separate features on the map. The available methods for the map matching are geometric map matching, topological map matching, and advanced map matching. The method used for this study to match the GPS points is obtained from the study by Hudson et al. (*22*), which uses the ArcGIS model for predicting the actual path of the bicyclists based on an algorithm developed by Dalumpines and Scott (*23*). This algorithm successfully implements geometric and topological map matching procedure with the help of network functions in ArcGIS. The GIS model used the buffer of 250 feet around each GPS point for implementing the algorithm. This value of 250 feet is determined based upon trial and error with the sample trips, which provided the highest accuracy of matched trips. Two consecutive points should be less than 500 ft. (Euclidean distance) apart to create the continuous restriction feature while implementing the GIS model. Since the frequency of the GPS points was not uniform, out of 12,454 trips used for map matching, 9,101 trips that accurately matched to the street were used for final analysis.

Generation of Choice Sets

Alternative routes for the pair of origin and destination were created using the Network Analyst extension in ArcGIS10.1. In total, there were six alternatives: five non-chosen and one chosen alternative. Simple Labeled Route method was used to generate five non-chosen alternatives (*24*). In this method, the shortest path between origin and destination was determined such that certain attributes of the path were either maximized or minimized. The five alternatives were created by maximizing use of bicycle friendly infrastructure along the route and minimizing length, the number of signalized intersections, the proportion of one-way road segments and the number of junctions separately. These alternative routes were joined to the street networks of the study area to attain attributes along the route.

Discrete Route Choice Model

Discrete route choice methods empirically model and analyze the decision maker's preferences among a set of alternatives available to them (*7, 11-13*). The Multinomial Logit (MNL) model is the simplest among the family of logit models, for which the probability of choosing the alternative i among the alternatives available in the choice set C_n is given by

$$P(i|\, C_n) = \frac{\exp(V_{in})}{\sum_{j \in Cn} \exp(V_{jn})} \tag{1}$$

Where, C_n is the choice set of alternatives, i is the chosen alternative, j is any alternative within C_n, V_{in} and V_{jn} are the utility of the alternative i and j.

The Independence of Irrelevant Alternatives (IIA) property of the MNL model suggests that the alternatives should be mutually exclusive, i.e. the alternative routes should not have overlapping routes. If this is property is not considered, MNL will overestimate the overlapping paths. Hence, a correction is introduced in the model in the form of a Path Size (PS) factor given by following equation (*25*).

$$PS_{in} = \sum_{a \in \Gamma i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in Cn} (\frac{L_i}{L_j})^{\gamma} \delta_{aj}} \tag{2}$$

Where, l_a is the length of link a, L_i is the length of the alternative i, Γ_i is the set of the links of alternative i, $\delta_{aj} = 1$ if j includes the link a, 0 otherwise, and γ = long-path correction factor, which is considered 0 in our case. For this study, due to the few number of alternatives, there are not any very long alternatives in our choice set C_n. Hence, the above equation will be reduced to the basic Path Size Logit (PSL) model (*26*). After the correction factor of PSL, the resulting probability that the alternative i is chosen from choice set C_n is given by

$$P(i|\, C_n) = \frac{\exp(V_{in} + \ln\,(PS_{in})}{\sum_{j \in Cn} \exp(V_{jn} + \ln\,(PS_{in}))} \tag{3}$$

Where PS_{in} will have values between 0 and 1, and hence, the $\ln(PS_{in})$ is always negative. This implies that the utility decreases when there is more overlap between the alternatives, as we are introducing the penalty for the route by introducing the path size factor. The model was estimated through the freely available software Easy Logit Modeler (*27*). The new form of the deterministic part of the utility function will be:

$$U_n = \beta x_n + \beta_{PS} * \ln\,(\text{path size}) \tag{4}$$

Where, x_n is the vector of attributes of the route and β is the estimated coefficient.

Distance trade-off calculation

To aid in interpretation, we can estimate marginal rates of substitution between distance and other explanatory variables. The distance trade-off for a unit change in attributes can be determined after estimating the utility coefficients of the attributes from following equation for the non-unit changes:

$$Equivalent\ \%\Delta\ distance = \left(Exp\left(\Delta attribute * \frac{\beta_{\ attributes}}{\beta_{\ \ln(distance)}} \right) - 1 \right) * 100 \tag{5}$$

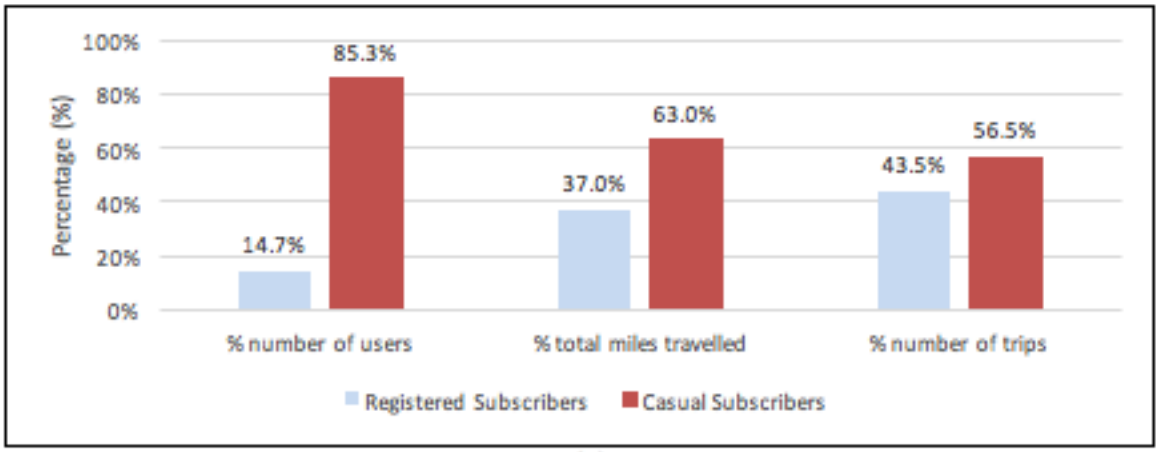
Where β is the coefficient of the attributes of the path estimated from the model.

RESULTS

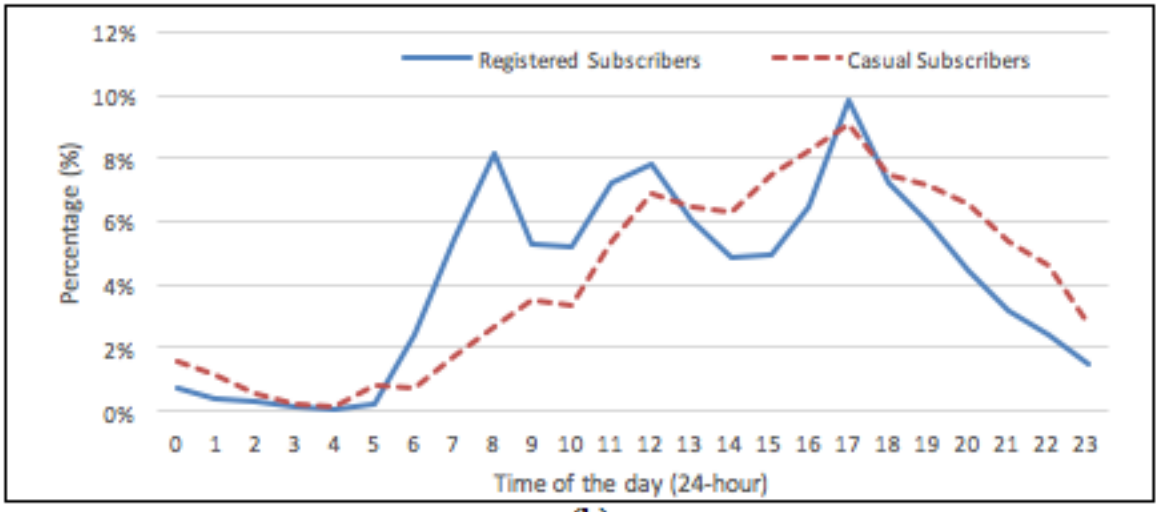
Descriptive Results

The registered users comprise approximately 15 % of the 1,866 users but account for 37 % of the total 10,476 miles traveled. After cleaning the data (i.e., removing recreational tours and erroneous GPS points), the final dataset was reduced to 9101 observations of which 43.5 % of the trips were made by registered users and 56.5 % of the trips made by casual users (Figure 1(a)).

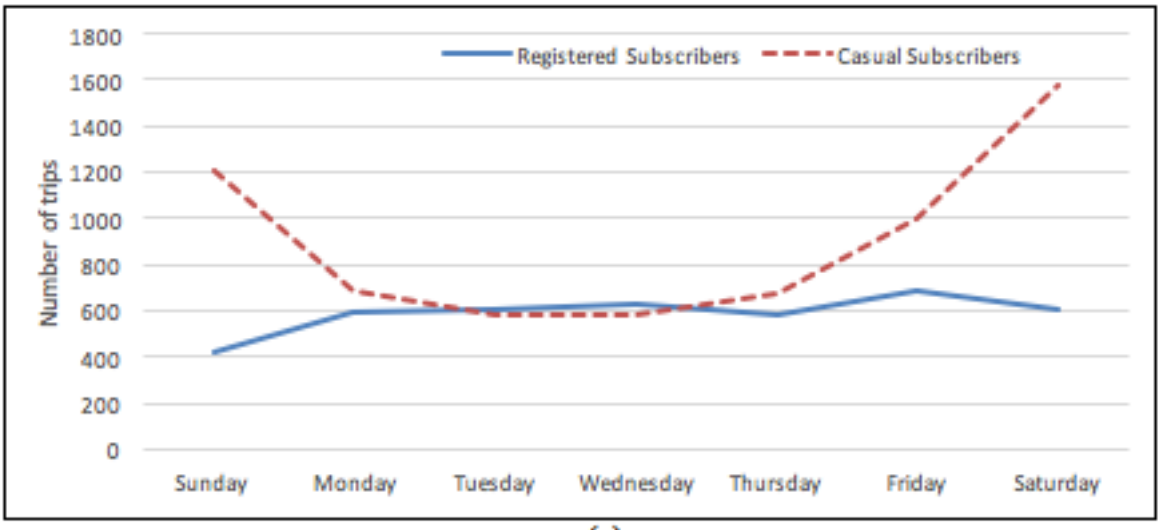
Figure 1(b) shows that trips by casual users increase steadily from the morning and peak at 5 pm, and then drop off into the night. However, trips by registered users peak at 8 a.m., 12 p.m. and 5 p.m. This shows commute nature of the trips made by registered users. Figure 1(c) shows that most of the casual trips are made during the weekend, with a total number of weekend trips being approximately equal to the total weekday trips. The variation in daily activity for registered users is small during the weekdays.



(a)



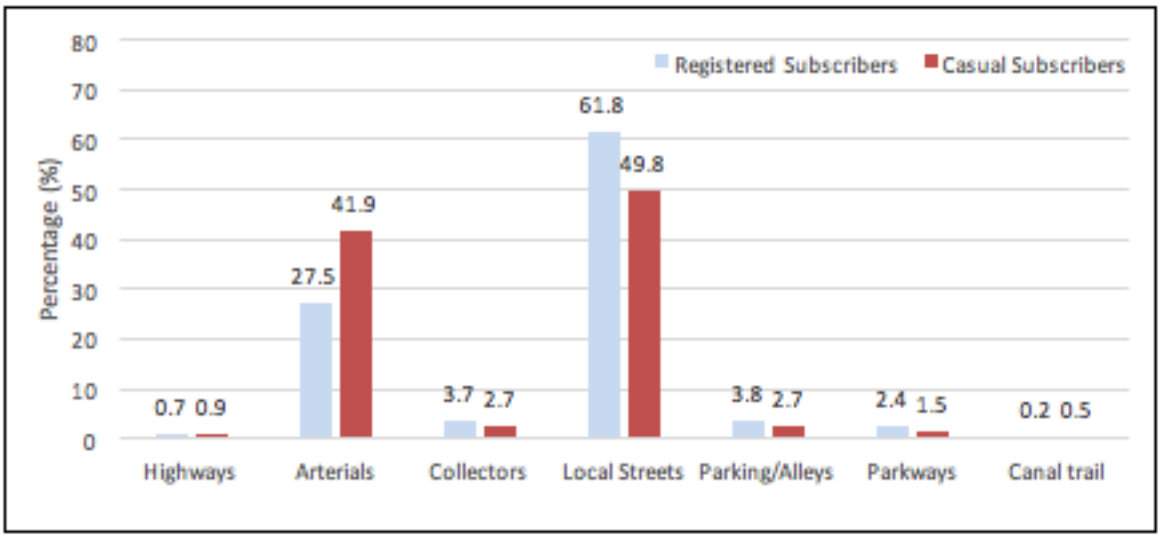
(b)



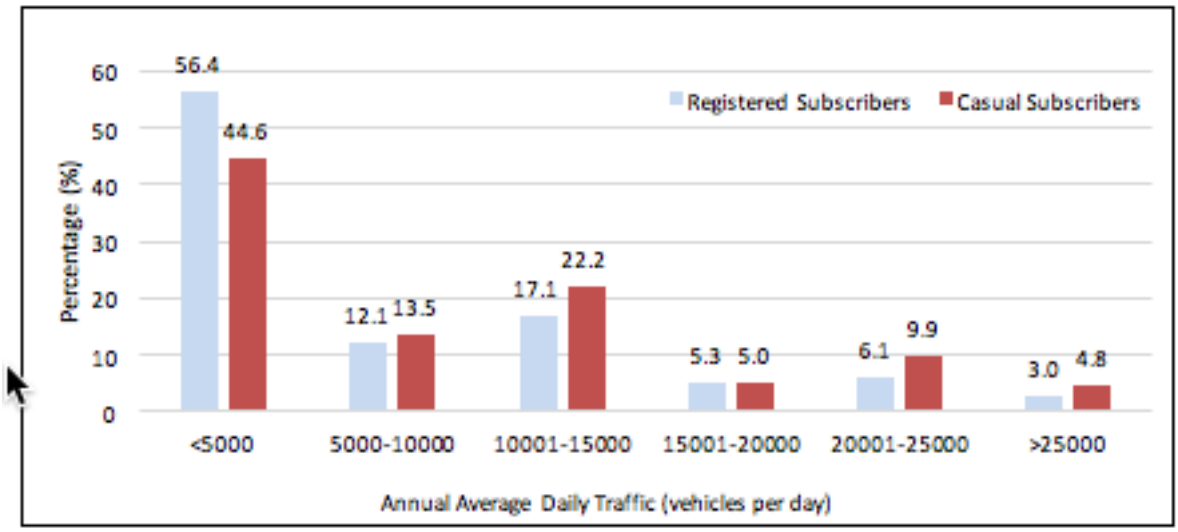
(c)

FIGURE 1 (a) Distribution of registered and casual users (b) Time of the day variation for a percentage of trips for registered and casual users (c) Day of the week variation of a number of trips for registered and casual users.

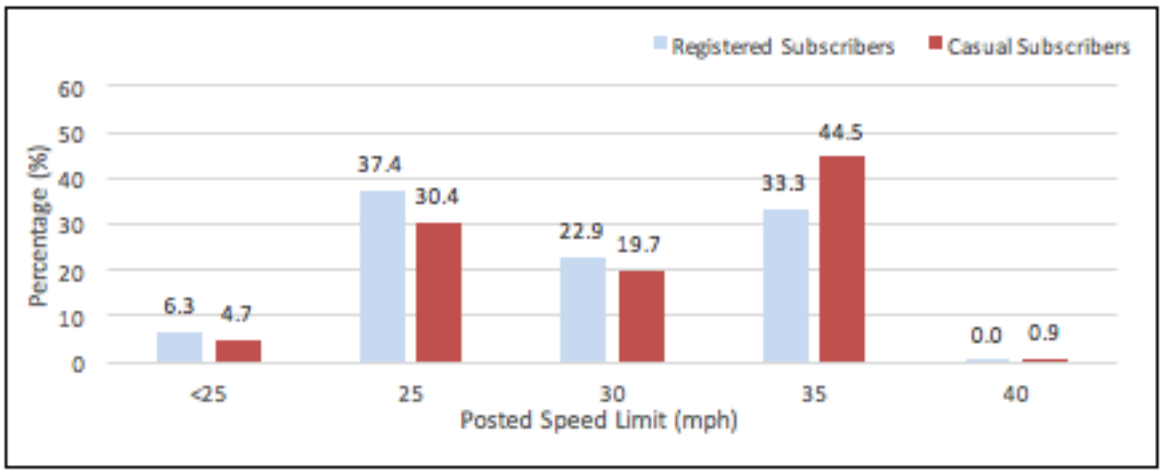
The trip behavior of the two user groups differed. The mean distance of the trips for registered and casual users were 1.0 (std. dev: 0.64) and 1.3 (std. dev: 0.95) miles, respectively; and similarly, the mean duration of the trip was 9.5 (std. dev: 7.2) and 14.5 (std. dev: 11.7) minutes, respectively. Registered users were making high percentage (69%) of trips with a travel time of less than 10 minutes. In contrast, 45% of casual user's trips are less than 10 minutes. Similarly, only 2% of the registered user's trips and 10% of the casual user's trips have traveled time greater than 30 minutes. The statistics of the usage of roadway infrastructure based on various characteristics of roadways provide insights on the behavior of these users. A high proportion of registered users preferred using lower volume and lower speed streets (Figure 2a, 2b, and 2c) implying familiarity with alternative routes on the network. In contrast, the casual users were more inclined to use the higher volume and higher speed roads.



(a)



(b)



(c)

FIGURE 2 (a) Distribution of travel distance over classifications of roads (b) Distribution of travel distance over a range of AADT (c) Distribution of travel distance over different posted speed limit.

Model Results

Before modeling, the correlation between variables was analyzed. The speed limit was found to be highly correlated with AADT along the route (Correlation Coefficient=0.60, p-value=0.000). AADT was included and the speed limit was excluded from a final model based upon likelihood ratio test. AADT was scaled to a smaller value dividing by 1000 to obtain more appropriately scaled parameter estimates. Furthermore, bicycle-vehicle-crash per mile was tested in the model because it was hypothesized that number of crashes would act as a proxy to the dangerous road. However, this variable was insignificant. Bikeshare users do not likely know the number of bicycle-motor vehicle collision (or risk proxy) and could not respond accordingly to avoid the routes with a high number of crashes.

Table 1 presents results for the estimation of the final Path Size Logit route choice model. Two models were developed, one for registered subscribers and one for casual

subscribers. The negative coefficient for distance variable supports the well-known fact that bicyclists prefer shorter routes among available alternatives unless there are other desirable attributes of other alternatives that outweigh the advantage of short distance. The magnitude of the coefficient suggests that registered users are more sensitive to the length of selected route compared to casual users. This is likely because registered users use bikeshare to make utilitarian trips in most of the cases, which is reinforced by the time of the day and week of the day variation shown in Figure 1(b) and 1 (c). The average length of the observed path for registered users and casual users is 6.9 % and 8.3 % higher than the average length of shortest path respectively. This statistics not only bolsters the difference in preference of these two groups over the length of the route but also points to other factors influencing route choice.

AADT is associated with negative utility for both categories of users (Table 1). The risks associated while traveling along high volume roads, which affects the perceived safety of the users, is likely a major reason for the disutility towards high volume roads. The extent of disutility is slightly higher for registered users. Registered users are those users who are committed to using bikeshare system, and subscribed bikeshare for a month or year. Hence, they are most likely to have information on which roads have a high flow of vehicles in the surrounding network, and avoid them as far as possible. AADT was interacted with the peak hour (7 am-9 am and 4 pm-6 pm) to test if the time of the day has an effect on route choice. However, the disutility of high AADT perceived by bikeshare users does not change significantly over the time of the day.

Both groups of users have a high preference towards including bike-specific facilities on their route. This finding supports previous literature that asserts the preference of bike lanes and bike routes among cyclists. Use of bike lanes, shared path or multiuse path have many inherent advantages such as separation from high-speed traffic and increase in perceived safety and freedom to ride at their preferred speed Travel on the bike-specific facilities is equivalent to decreasing distance by 44.9% (53.3% for casual users). In peak hours, registered users are more likely to use bike-specific facilities as compared to off-peak hours. As bicycle facilities provide efficient and fast travel, registered users prefer using bicycle facilities to avoid delay during peak hours. There is similar effect among casual users but they are less sensitive towards using bike-specific facilities than registered users.

Registered users avoid including one-way road segments in their trip as far as possible. The route choice behavior of casual users is not definitive regarding the inclusion of one-way road segments. Registered users are aware of the information on the competing routes, which allows them to choose a route that minimizes or avoids one-way street segments. On the other hand, casual users (sometime tourists and most of the time infrequent bicyclists) are unaware of the alternatives and could not avoid these segments.

A number of signalized intersections along the route is a positive factor for the selection of the route. The coefficients for the number of signals per mile are equal for both groups of bikeshare users. This result could be counterintuitive that the signalized intersection on the route decreases the utility of the route because these intersections add delay and potential risk. However, this might not be always true. Signalized intersections provide relatively safe, protected crossings of large roadways. It is also reasonable that, in a grid network, cyclists tend to ride on main arterials and tend to avoid routes that require them to cross un-signalized (e.g., midblock) minor street crossings. The downtown Phoenix area is highly signalized, making it difficult to avoid signals along various routes.

The number of turns along the route is another factor that a cyclist accounts for while choosing a route. Both, registered and casual users, valued routes with fewer left- and right-turns, but in a different manner as suggested from Table 1. The difference in the value of

coefficient shows that cyclists, in general, have a greater aversion to left turns compared to right turns, as expected. The higher delay associated with left turns, at signalized as well as

TABLE 1 Estimation of Utility Coefficients

Variable	Registered Subscribers			Casual Subscribers		
	Est. Coeff.	*t*-stat	Rand. Err.	Est.Coeff	*t*-stat	Rand. Err.
ln(length)	-4.64	-17.77	0.261	-2.83	-15.05	0.188
Proportion of bike facilities	2.77	17.78	0.156	2.16	19.32	0.112
Number of left turns per mile	-0.14	-10.22	0.014	-0.17	-13.61	0.013
Number of right turns per mile	-0.14	-9.91	0.014	-0.13	-10.66	0.012
Proportion of one way	-0.43	-3.63	0.119	0.11	1.20**	0.090
Numbers of signals per mile	0.25	17.99	0.014	0.24	20.81	0.012
AADT/1000	-0.16	-21.27	0.007	-0.08	-15.80	0.005
ln(length)*Peak hour	-3.97	-6.63	0.598	-2.45	-5.19	0.472
Proportion of bike * Peak hour	0.93	3.23	0.288	0.46	1.92*	0.239
Proportion of one way * peak hour	-0.57	-2.61	0.219	-0.49	-2.42	0.202
ln(PS)	1.26	15.77	0.080	1.17	18.82	0.062
Log Likelihood at Zero		-5587.23		-7533.47		
Log Likelihood at Convergence		-4140.59		-6284.41		
Adjusted Rho Squared w.r.t. Zero		0.2569		0.1643		
Number of Cases		3958		5143		

** - Insignificant
* - Significant at 90% Confidence Limit

un-signalized intersections, and additional safety risk associated left turns compared to right turns could be the main reasons for disutility of this variable for both users. The difference is more pronounced for casual cyclists, but registered users do not seem to differentiate much between left and right turns in terms of utility. We found that registered users mostly travel on low volume and low-speed roads. Since the left and right turns have a similar effect in terms of delay time or difficulty in maneuvering in low volume and low-speed roads mostly traversed by registered users, they might give equal priority to left and right turns. For each left turn, registered users would choose routes that were 3.2 % longer (6.3 % for casual users) (Table 2). Corresponding additional percentage of route length for additional one right turn is 3.1 % and 4.8 % for each additional right turn. This clarifies the comparison between left and right turns on the route made by registered and casual and users.

Effect of time of the day and day of the week was analyzed. For this, morning and evening peak hour (7 am to 9 am and 4 pm to 6 pm) are categorized as peak hour. This variable is interacted with other variables of the model. The proportion of bike-specific facilities, the proportion of one-way road segments and trip length had significantly different effects compared to off-peak hours. Both users were more likely to make shorter trip avoiding one-way segments and including bike specific facilities. Other variables did not have distinctly different effects at different times of the day. Travel on bike-specific facilities in peak hour is equivalent to decreasing distance by 20.9% for registered users (17.1% for casual users) compared to off-peak hours. Peak hour traffic and delays associated with travel

could be the major reason for selecting shorter routes by both users. This result also bolsters the sensitiveness of registered users towards shorter trip length compared to casual users.

TABLE 2 Marginal Rate of Substitution with Distance (%) for Unit Change in Attribute

Variable	Distance Trade-off (% distance)	
	Registered Subscribers	Casual Subscribers
Proportion of bike facilities	-44.9	-53.3
Number of left turns per mile	3.2	6.3
Number of right turns per mile	3.1	4.8
Proportion of one-way	9.8	-3.7
Number of signals per mile	-5.2	-8.2
Peak hour (Baseline: Off peak hour)		
Proportion of bike facilities	-20.9	-17.1
Proportion of one-way facilities	15.5	22.0

Similarly, the advantage of fast travel without being interfered by a large number of high-speed vehicles during peak hour, in addition to increased perceived safety, can explain the preference of bike specific facilities during peak hour as compared to an off-peak hour. A key unobserved factor that likely varies between times of day, demographics, likely affects the differences as well.

As the utility of overlapping paths is overestimated in the MNL, Path Size Correction is introduced to adjust the utilities for overlap. Since the value of PS lies between 0 and 1 (1 for the unique route), ln (PS) is always negative. This will reduce the deterministic utility of the route based on the degree of overlap. Hence, this corrects the overestimated utility of overlapping routes in a set of alternatives. Similarly, the estimate of ln(PS) should be positive and significantly different from 1, similar to our results (*28*).

DISCUSSION AND CONCLUSION

The main objective of the study was to use the real-time GPS data to determine the route choice preference of bikeshare users. Although, the number of bikeshare system has increased sharply in North America, thorough research on the behavior and route selection in unknown, especially for bikeshare users. This research is unique because it is among the first that looks at route-level analysis of bikeshare trip (enabled by GPS), it relies on a system without strict station origin and destinations, and it is operated in a city with a grid transportation system enabling many feasible alternative routes.

The findings from this paper can be compared across other studies of bicyclist route choice. Several consistencies exist between this and previous studies that examine conventional cyclists. Negative utility for the longer trips is consistent with all of the previous studies. The length of the route results in negative utility for all of the studies. Consistent positive preferences towards bike facilities can be seen in previous studies (*9, 12, 13*). A study in San Francisco, California (*12*) found cyclists are willing to add a mile on bike lanes in exchange for 0.5 miles of ordinary roads. This result is very near to the result of our study, which indicates that if there is one mile of road with bicycle facilities it is equivalent to 0.55 miles of normal road (0.47 miles for casual users). A number of turns per mile was found to

have distance trade-off value of 4.2% (commute trips) and 7.4% (non-commute trips respectively (*7*).Left and right turns are analyzed separately in this paper. For registered and casual users, distance trade-off is 3.2% for left-turn (3.2% for right turn) and 6.3% for left-turn (4.8% for right-turn) respectively. This gives a ground for comparability of registered and casual users with the commute and non-commute trips. Other studies, however, estimated 17 % distance value for one turn. This is significantly higher than the value estimated by this study (*12*). Signalized intersections were found to be used while crossing major roadways and turning, traffic volume has a consistently negative impact on route choice (*5, 13*), especially for left turns in the cross street with high AADT relative to right turns with the same AADT in cross streets (*7*). This result of negative utility for AADT is consistent with this study too. In spite of various studies that identify the possible risk factors for the cyclists making turns or making through movements at signalized intersections (*29*), our study found preferences towards signalized intersections. This might be attributed to either different behavior of bikeshare users or benefits of protected phases through signalized intersections.

Since most of the origin and destination were fixed, though this bikeshare system allows non-station origins and destinations, a future study could focus on the influence of the placement of these stations on the route choice model to balance placing stations on visible, busy streets that force users to ride on those streets for station access. The riders tended to value travel distance more than other factors and planners should focus on providing better alternative route information, especially to non-subscribing users, identifying station locations that allow direct access to bike-friendly routes, and improving the safety and operations of routes in the service area in regard to cycling (e.g., lowering speed limits on main corridors).

This study suffers from several limitations. Discrete choice modeling was used due to its simplicity and predictive power and precedence in route choice analysis. The future research could utilize the alternative model forms such as machine learning and compare the results with that of the discrete choice model. As a purely revealed preference study, we do not know anything about the actual alternative routes that are in the choice set of the users. We estimated five reasonable alternative routes. We also do not have any information on demographic factors, such as age, gender, income, cycling frequency etc. that could influence route choice or value of attributes. A Recent result from Capital Bikeshare member survey report has identified that bikeshare users tend to be young, high income, white, and male (*30*). We begin to understand some of these difference with the two different classes of users (casual and registered). This study only focuses on a subset of cyclists, bikeshare riders, and is not representative of many urban cyclists. This data, hence, suffers from self-selection bias, those who self-select into bikeshare systems and choose a route, not those who decide not to ride a bike because of the suitable route. Still, this paper provides a first attempt to estimate route choice for an important and growing sub-population of cyclists, bikeshare users.

ACKNOWLEDGEMENTS

The authors would like to thank Southeastern Transportation Center (STC), a Federal University Transportation Center, for funding this project and Social Bicycles for providing the GPS data from the Grid Bikeshare, in Phoenix Arizona.

REFERENCES

1. League of American Bicyclists, *Where we ride, analysis of bicycle commuting in American cities*. 2014.

2. Gotschi, T., *Costs and benefits of bicycling investments in Portland, Oregon.* Journal of Physical Activity and Health, 2011. **8**(1): p. S49-S58.
3. Barnes, G., K. Thompson, and K. Krizek. *A longitudinal analysis of the effect of bicycle facilities on commute mode share*. in *85th Annual Meeting of the Transportation Research Board. Transportation Research Board, Washington, DC*. 2006.
4. Stinson, M. and C. Bhat, *Commuter bicyclist route choice: analysis using a stated preference survey.* Transportation Research Record: Journal of the Transportation Research Board, 2003(1828): p. 107-115.
5. Sener, I.N., N. Eluru, and C.R. Bhat, *An analysis of bicycle route choice preferences in Texas, US.* Transportation, 2009. **36**(5): p. 511-539.
6. Segadilha, A.B.P. and S. da Penha Sanches, *Identification of Factors that Influence Cyclistś Route Choice.* Procedia-Social and Behavioral Sciences, 2014. **160**: p. 372-380.
7. Broach, J., J. Dill, and J. Gliebe, *Where do cyclists ride? A route choice model developed with revealed preference GPS data.* Transportation Research Part A: Policy and Practice, 2012. **46**(10): p. 1730-1740.
8. Dill, J. and J. Gliebe, *Understanding and measuring bicycling behavior: A focus on travel time and route choice.* 2008.
9. Broach, J., J. Gliebe, and J. Dill, *Development of a Multi-class bicyclist route choice model using revealed preference data.* 2009.
10. Langford, B.C., *A comparative health and safety analysis of electric-assist and regular bicycles in an on-campus bicycle sharing system.* 2013.
11. Ehrgott, M., J.Y. Wang, A. Raith, and C. Van Houtte, *A bi-objective cyclist route choice model.* Transportation research part A: policy and practice, 2012. **46**(4): p. 652-663.
12. Hood, J., E. Sall, and B. Charlton, *A GPS-based bicycle route choice model for San Francisco, California.* Transportation letters, 2011. **3**(1): p. 63-75.
13. Menghini, G., N. Carrasco, N. Schüssler, and K.W. Axhausen, *Route choice of cyclists in Zurich.* Transportation research part A: policy and practice, 2010. **44**(9): p. 754-765.
14. Larsen, J. and A. El-Geneidy, *A travel behavior analysis of urban cycling facilities in Montréal, Canada.* Transportation research part D: transport and environment, 2011. **16**(2): p. 172-177.
15. Klasnja, P. and W. Pratt, *Healthcare in the pocket: mapping the space of mobile-phone health interventions.* Journal of biomedical informatics, 2012. **45**(1): p. 184-198.
16. Franke, P., *The Adoption of Motorized Two Wheelers: A Study of Transportation development in Shanghai, China*, in *Department of City and Regional Planning*. 1997, University of California, Berkeley.
17. Shaheen, S., S. Guzman, and H. Zhang, *Bikesharing in Europe, the Americas, and Asia: past, present, and future.* Transportation Research Record: Journal of the Transportation Research Board, 2010(2143): p. 159-167.
18. Fishman, E., S. Washington, and N. Haworth, *Bike Share: A Synthesis of the Literature.* Transport Reviews, 2013. **33**(2): p. 148-165.
19. Wang, X., G. Lindsey, J.E. Schoner, and A. Harrison, *Modeling Bike Share Station Activity: Effects of Nearby Businesses and Jobs on Trips to and from Stations.* Journal of Urban Planning and Development, 2015: p. 04015001.
20. Buck, D., R. Buehler, P. Happ, B. Rawls, P. Chung, and N. Borecki, *Are Bikeshare Users Different from Regular Cyclists? A First Look at Short-Term Users, Annual Members, and Area Cyclists in the Washington, DC, Region.* Transportation Research Record: Journal of the Transportation Research Board, 2013(2387): p. 112-119.
21. Schoner, J. and D. Levinson, *Which Station? Access Trips and Bike Share Route Choice*. 2013.
22. Hudson, J.G., J.C. Duthie, Y.K. Rathod, K.A. Larsen, and J.L. Meyer, *Using smartphones to collect bicycle travel data in Texas*. 2012.
23. Dalumpines, R. and D.M. Scott, *GIS-based map-matching: Development and demonstration of a postprocessing map-matching algorithm for transportation research*, in *Advancing geoinformation science for a changing world*. 2011, Springer. p. 101-120.

24. Ben-Akiva, M., M. Bergman, A.J. Daly, and R. Ramaswamy. *Modeling inter-urban route choice behaviour*. in *Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Press, Utrecht*. 1984.
25. Ramming, M.S., *Network knowledge and route choice*. 2001, Massachusetts Institute of Technology.
26. NMCBen-Akiva, M. and M. Bierlaire, *Discrete choice methods and their applications to short term travel decisions*, in *Handbook of transportation science*. 1999, Springer. p. 5-33.
27. Santos, A., N. McGuckin, H.Y. Nakamoto, D. Gray, and S. Liss, *Summary of travel trends: 2009 national household travel survey*. 2011.
28. Frejinger, E. and M. Bierlaire, *Capturing correlation with subnetworks in route choice models.* Transportation Research Part B: Methodological, 2007. **41**(3): p. 363-378.
29. Wachtel, A. and D. Lewiston, *Risk factors for bicycle-motor vehicle collisions at intersections.* ITE Journal(Institute of Transportation Engineers), 1994. **64**(9): p. 30-35.
30. *2014 Capital Bikeshare Member Survey Report*. Available from: http://www.capitalbikeshare.com/assets/pdf/cabi-2014surveyreport.pdf.