Cycling Willingness: Investigating Distance as a Dependent Variable in Cycling Behavior Among College Students

Thomas Wuerzer  
Boise State University

Susan G. Mason  
Boise State University

Publication Information
Wuerzer, Thomas; and Mason, Susan G. (2015). "Cycling Willingness: Investigating Distance as a Dependent Variable in Cycling Behavior Among College Students". Applied Geography, 60, 95-106. http://dx.doi.org/10.1016/j.apgeog.2015.03.009

NOTICE: This is the author's version of a work that was accepted for publication in Applied Geography. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Applied Geography, vol. 60, June 2015. doi: 10.1016/j.apgeog.2015.03.009
Cycling Willingness: Investigating Distance as a Dependent Variable in Cycling Behavior Among College Students

Thomas Wuerzer
Boise State University

Susan G. Mason
Boise State University

Abstract

We present a novel approach to understanding distance as a barrier to cycling and its use as a dependent variable in multinomial logistic regression. In doing so, this study explores distances in relation to spatially and relevant human factors such as gender and propensity to cycle among college students. College students (N= 949) participated in a health survey and stated possible predictors of cycling based on their cycle usage and preferences in the previous 30 days. While utilizing GIS in a bicycle-friendly network, we created geo-statistical GIS-groupings and performed multinomial logistic regression analysis. We examined college students to discover how their demographic and personal characteristics may mediate the deterrent properties of distance when considered as a dependent variable in cycling to a college campus. Age and propensity for cycling for transportation mediate the negative effect of distance on the likelihood of cycling. The findings also suggest that infrastructure improvements could lessen the impact of distance as a barrier to cycling and increase the likelihood of cycling for commuting.

1. Introduction

The trend toward cycling as an important and healthy mode of transportation in larger cities has increased (League of American Bicyclists, 2013), and more and more communities are encouraging active transportation opportunities (Dujardin et al., 2012; Shan, 2014; Snizek et al., 2013). In the U.S., this trend is mostly seen in cities in the East and in Western states (League of American Bicyclists, 2013). Additional cities with the highest percentage of change in cycling between the years 2000 and 2011 have also established bike share programs (García-Palomares et al., 2012; Rybarczyk & Wu, 2010), initiated complete street programs (National Complete Streets Coalition, 2013), have a large campus population, or have shown an increased awareness of the need for sustainability in transportation (Schneider, 2013).

The physical built environment influences the degree of cycling for transportation and recreation (McGinn et al., 2007). Others have pointed out that the built environment, personal demographics, perceptions, and attitudes about physical activity can be barriers to or facilitators for cycling (Charreire et al., 2012). More detail on trip distribution, patterns and daily activities of college students is provided by Chen (2012). Utilizing trip diaries and surveys, Chen finds destinations with the highest frequency are labeled “home” and “academic activities”. Chen (2012) also notes that the surveys were geocoded but does not explain whether a bicycle-friendly network instead of a street-based network was explicitly used to compute biking and distances on dedicated bike paths. Khattak et al. (2011) and Nayar (2012) not only present a campus-oriented survey in their studies but also demonstrate the different behaviors among students living on-campus versus off-campus. This includes dissimilar statistical patterns found in transportation behavior in traveling to and from an urban campus compared to that occurring in the setting of a suburban campus in which personal vehicles, public transit, or bicycles were used. Both studies included mixed modes of transportation, and in Nayar’s (2012) study it is not clear whether distance was considered.

The literature on cycling primarily provides studies (Buehler, 2012; Buehler & Pucher, 2011; Chen, 2012; Krizek & Johnson, 2006) which emphasize general population and function (i.e., distance to trailheads or bike facilities) and rarely specify targeted groups. Some studies present distance as an independent variable in their statistical models and as a supporting incentive in the larger process of deciding to bike or not to bike (Emond & Handy, 2012;
Gatersleben & Appleton, 2007; Handy et al., 2010; Krizek & Johnson, 2006; Xing et al., 2010). Heinen et al. (2010) present an interesting observation that “most research into bicycle use identifies distance as a significant factor” (p.61). Mullan also found in her studies on views of cycling for transportation (Mullan, 2013) and also cycling for recreation (Mullan, 2012) that distance, in terms of the length of the journey as well as trip purpose, are important in identifying willingness to cycle. For those who use cycling for transportation, the decision to do so was rarely based on health or environmental factors. Factors such as time, distance, and weather were more important drivers in the decision-making process (Mullan, 2013). These findings suggest distance creates a more significant barrier to some people than others. As such, we transpose the approach to studying distance as an independent variable to studying it as a dependent variable. We use spatial statics to identify clusters of cyclist which in turn permits additional multinomial logistic regression analysis. We examine demographic and behavioral characteristics that may mediate different distances affecting college students’ propensity toward cycling to campus. By knowing what specifically contributes to distance being an impediment to cycling, we hope to gain insights on how to influence policy that promotes cycling over a variety of distances and make such distances less of an impediment.

This study focuses on distance-relevant cycling behavior among college students at [University in the West], "bicycle-friendly university" named by the League of American Bicyclists (2011b), and within the region of [Removed for Review Purposes] where approximately 4% of its population regularly commute to work by bike (League of American Bicyclists, 2011a) and which has a richly-developed bike path network along a river. We examine several human factors that mediate distance as a barrier to or facilitator of engaging in cycling such as gender, age, biking usage, and car ownership, as well as spatial travel patterns to help explain the impact of distance on the decision of college students to bike. We use the application of bicycle-friendly networks in a GIS analysis to examine the patterns of cycling to the center of campus (the centrally located library) with statistical groupings in GIS and, thereafter, use multinomial logistic regression analysis to identify factors having significance for distance. We believe this combination of spatially-driven research along with a statistical analysis presents a novel approach and platform to study the cycling-based activity of college students.

2. College Students, Human Factors and Distance

2.1 Cycling Among College Students

Many studies examine choice of transportation as an integral focus among children in kindergarten, elementary school, and junior high schools (Evenson et al., 2003; Ewing et al., 2004; Hoffman et al., 2014; Lang et al., 2011; McDonald, 2008; Schlossberg et al., 2006). Faulkner et al. (2009) reviewed existing research finding students who walk, bike or use other non-motorized transportation to get to school tend to be overall more active than others. In regard to college-students, other scholars (Balsas, 2003; Bopp et al., 2011; Chen, 2012; Kamruzzaman et al., 2011; Khattak et al., 2011; Nayar, 2012) present valuable insights into behavior, modal choice, and cycling activities on and in the immediate proximity of college campuses.

We focus on college students and their campuses to ascertain whether campus transportation policies might influence the behavior of students both within and outside those boundaries. Balsas develops the argument that communities in close proximity are affected by universities and their “parking, traffic, service access and off-campus housing” (Balsas, 2003, p. 36). With more insights into cycling behavior among college students which influences future campus policies, and can ultimately transform neighboring communities. Schneider (2013) notes the efforts of communities to shift mode choice from automobiles to walking and bicycling as they frame a discussion on routine mode choice decisions. This discussion, on the steps needed to make such a shift happen, has value for larger college campuses and cities.

Our focus on a college population not only informs policy and decision makers but also provides insight into reorganizing, for example, college campus policies based on transportation and housing needs. Balsas (2003) explicitly states that college campuses are unique places whose population (students, staff, and faculty) turnover at a much higher frequency than the general population. This is generally supported by Bopp et al. (2011); however, in the context of active commuting, these researchers provide little knowledge about how a campus and its cycling environment influence student travel patterns and behaviors.
Campus cycling policies affect their population and adjacent communities, but there is little empirical work available on the subject of cycling among college students and their behavior when they live off-campus. Understanding these needs and then better addressing them might help to build more efficient and cost-effective transportation policy in the context of college campuses. It is also possible to imagine that the routines and behaviors adopted in college, such as cycling to work or school, can be carried on later in life, as Balsas (2003) suggests.

2.2 Factors of Age, Gender, Cycling for Recreation and Car Ownership

Shahan (2007) found rates of cycling in both Maryland and the Netherlands more often positively associated with age and negatively associated distance than with factors such as bike infrastructure and amenities. Pucher, Garrard, et al. (2011) found that cyclists in the outer suburbs of Melbourne and Sydney tend to be under the age of 20 and primarily cycle for recreation on weekends. In the US, Pucher, Buehler, and Seinen (2011), find that the number of 40 to 64-year old cyclists increased the most of any age group they studied, more than doubling their share of bike trips between 2001 and 2009.

There is evidence that “women cycle shorter distances to work than men” (Heinen et al., 2011, p. 62). This was also found to be the case in Copenhagen: persons commuting 8 kilometers (approximately 5 miles) or more are more likely to be male. Additionally, long distance commuters in general are simply less likely to be women (Hansen & Nielsen, 2014). This is notable, since Copenhagen is well known for its extensive cycling infrastructure available to riders. Akar et al. (2013) cite many other studies that demonstrate women are likely to cycle less than men. Pucher, Buehler, and Seinen (2011) note that in the U.S. the share of women cyclists is 25% and the percent of trips women took by bike dropped from 33% to 22% between 2001 and 2009. Finally, Buehler (2012) specifically finds that men have a 2.56 greater likelihood of cycling to work than do women.

In the US, more than two-thirds of bike trips are taken for recreation (Pucher & Dijkstra, 2000). With such a large percentage of cycling in general for recreation purposes, it is reasonable to consider whether one’s predilection toward cycling for recreation has an influence on one’s decision to cycle for transportation. Stinson and Bhat (2004) found a relationship between cycling in general and cycling for commuting. All the while, Xing et al. (2010) found that 90% of those who cycle for transportation are cycling for other purposes as well. While this suggests a connection in cycling behavior, it may not be that recreational cycling influences commuting cycling habits. Kroesen and Handy (2014) find that the propensity to cycle for commuting has an influence on cycling for recreation and that it is slightly greater than the influence of recreational cycling on commuting cycling behavior.

Car ownership has repeatedly been shown to have an impact on the likelihood of cycling. Kroesen and Handy (2014) found that although 63% percent of cyclist own one or more cars there is a strong negative influence of car ownership on the probability of being a cyclist or not. Buehler (2012) presented that for each additional car per household the likelihood of cycling decreases by 77%. Pucher et al. (2010) showed the frequency of cycling to be more than twice as high for households without cars as for households with three or more cars.

2.3 The “Issue” of Distance as a Measure in Cycling/Transportation Studies

In reviewing current scholarly work, we find authors collectively present distance as a factor (measured by length, driving duration, or time-competition between modes) in their surveys or statistical models as part of a larger explanation of biking behavior. However, Xing et al. (2010) and others do not examine why distance represents a greater barrier to some or serves to facilitate cycling for others. In terms of walking, Shan (2014) recently presented that “20 min walking distance is a critical divide”(p.32) in utilizing urban green spaces.

If the decision to engage in physical activity and use cycling for transportation is affected by distance, then we would like to know for whom distance is a barrier. Is it for everyone or specific groups? Everyone equally? Women more than men? Old versus young? These questions call for applying geographic information systems (GIS) paired with approaches to spatial statistics and spatial groupings and statistical analysis on human factors such as age and gender. Work by Iacono et al. (2008) estimates distance decay factors and their influence on travel behavior but does not treat distance as a dependent variable. In Garcia-Palomares et al.’s (2013) work, distance decay functions are used to compare the sensitivity among, i.e., age groups and gender toward their willingness to walk. Zhao et al. (2003) use similar distance decay application to show the deteriorating influence of increasing distance on walking to transit services. Overall, prior studies establish distances as part of a regression model or, i.e., binary-logit models.
on modes of transportation choices (Emond & Handy, 2012; Ewing et al., 2004), travel behavior (Bopp et al., 2011; Chen, 2012). This limited the application of its measure as similar to monetary values such as housing values. Distance is a linear measure across space and should be treated as such in any analysis, along with other statistical procedures.

Specifically relevant to this study, (Redacted for Review, 2013) demonstrated the application of barrier and facilitator indexes as precursors for biking behavior among college students. Barriers such as urban form, gender, variation on distance, or network layout have also been noted by Heinen et al. (2010) but not tested as to whether they apply to a spatial distribution as found in our groupings. Heinen et al. (2010) also state that distance appears to have a negative influence on decisions to bike, and that “little is known about the effect of distance on cycling frequency” (p. 62). Krizek and Johnson (2006) presented proximity to trails and retail (i.e., 0-200, 201-400 meters) for bicycle (and walking) as influential factors but did not report whether their study applied groupings of subjects across space and intensified the influence of location (therefore, distance) in their models.

3. Research Aims and Hypothesis

Boussauw et al. (2014) and their work on short trips and the distance between homes and schools emphasizes trip length and elementary schools in the context of central places as destinations. We see a college campus, similar to the concepts of central place theory (Christaller, 1933) or Perroux’s field of influence (growth poles) (Parr, 1999), functioning as the nucleus of new sustainable policy and practices (Balsas, 2003). We suggest that college students are an integral part of active transportation efforts that are in need of more explicit research dedicated to cycling preferences (Pucher, Buehler, Merom, et al., 2011; Shan, 2014).

We initially focus on treating distance as the dependent variable and then turn to looking at the factors that may contribute to making distance a barrier to cycling for transportation in a college student population. A typical postulation in transportation research is that as distances to a destination increases, we intuitively follow the assumption that the likelihood of cycling to the destination decreases. That said, we argue that users of cycling, when compared among each other, have certain characteristics by which we can group and compare them. Heinen et al. (2011) find that over longer distances, workers are not affected by their social environment in their decision to commuter-cycle but rather base that decision on individual considerations. We argue that cycling behavior is built on preferences and examine the way that the relevance of distance is specifically mediated by factors of age, gender, healthy weight, and propensity for cycling as transportation.

We are basing our hypothesis, that distance no longer be treated as an independent variable in a statistical model but rather as a “pricing”-function, conceptually on the underlying logic of hedonic modeling (Rosen, 1974). Hedonic models, usually applied in real estate and related disciplines, measure, for example, property values as a pricing function including positive and negative influences of their utility bearing attributes such as property size, home size, number of rooms, or socio-demographic parameters (Paterson & Boyle, 2002; Zhou et al., 2006). Most recent planning-related applications include the impacts of parks or TODs on housing values (Duncan, 2011), or the influence of mixed-use developments on nearby residential properties (Song & Knaap, 2004). We argue that using distance as a dependent variable (here as distance from home to campus), is similar to the hedonic pricing function, expressed through the model’s characteristics and dependent on other variables such as age, gender, car ownership, and recreational cycling behavior. We recognize that using the perspective of hedonics models can show that user-specific characteristics are likely to manifest utility-bearing attributes for the cyclist. Given this view, we use the propensity to cycle for recreation, having a positive influence on outcomes, and car ownership as having a negative influence as control variables for the dependent variable - the cycling distance.

We hypothesize that these spatial patterns demonstrate that distance can be a greater or smaller barrier to some users than to others. Some could argue that the principle of location-efficient mortgage (LEM), in which one would accept borrowing more money for a home mortgage due to the potential savings benefits of reduced transportation costs (or transit) could be applied (Krizek, NA). However, the distance from such a LEM-location, in this case, a campus, would be a dependent variable (outcome) of a prior decision process. This is shown in the case of walking and distances in a study by Manaugh and El-Geneidy (2013); when asked about location choice, most of their participants answered with “Ability to walk to campus”. Specifically, we hypothesize that older persons, males, and persons in relatively good shape, and having a propensity to cycle for recreation play a role in predicting the distance one is willing to cycle to campus, regardless of the distance of the cluster the person resides.
To test our argument, we use survey data we collected from college students attending [University in the West] to run spatial and statistical analyses in GIS and SPSS while exploring a) distance as a measure of cycling behavior in general, and b) using distance and recent cycling activity (in the previous 30 days) as a grouping factor while comparing the characteristics of users in the clusters.

4. Methodological Approaches

To achieve the aims of this study, we conducted a survey among [University in the West’s] students (N = 949) as an extension of the National Collegiate Health Assessment for the American College Health Association (NCHA-ACHA II). NCHA-ACHA II examines topics such as substance use and abuse, and mental and physical health. We added questions for our particular university to the national survey. Students were asked additional questions about their cycling behavior in transportation and recreation, and about barriers to and facilitators of cycling in the context of spatial patterns. In the second tier of this study using GIS, distances that students biked were computed from survey responses, and then, using a spatial analysis, we were able to group them into clusters. In the third tier, these groupings or clustered relationships were used in multinomial logistic regression.

4.1 Cycling Among College Students – Survey Instrument

Utilizing the NCHA-ACHA II as an assessment platform, we added additional questions about cycling-specific behavior. The NCHA-ACHA II is a national survey targeting college students as subjects, their health-related behaviors, and their perceptions. NCHA-ACHA II is conducted every two years on participating college campuses across the US. The survey instrument contains 66 items examining and assessing college students’ consumption of alcohol and drugs; habits of nutrition, weight, and physical activity; sexual behaviors, personal safety, and violence; and health education and academic performance. In addition to the NCHA-ACHA II items, this research added questions that asked students to identify their behaviors and barriers to and facilitators of cycling as a mode of transportation and as recreational biking. Such predictors of cycling can be based on personal features such as gender and age, or environmental characteristics such as cycling infrastructure, bike lanes, availability of showers at the target destination, or current weather conditions. Aligned with NCHA-ACHA II sampling methodology, students 18 years or older, with equal or more than four credits enrollment, and no dual enrollment (students in high school taking college courses) were randomly selected, resulting in 4,450 students of the approximate 20,000 being sent the survey to be completed between November 7th and 21st of 2011. The survey reflects cycling behavior in the previous 30 days. A total of 949 surveys (approximately 21 percent of the invited students) were returned, resulting in a margin of error +/- 3% at a 95% confidence level interval.

4.2 GIS Methodology and Dependent Variable

For geocoding, we asked the students about the closest intersection (cross streets) to their homes and associated zip code. Using this closest intersection as an “address” location balances expected spatial bias and the known issue of block aggregation (Apparicio, 2007; Hewko et al., 2002), and the central advantage that participants are more likely to give an intersection as an answer than their personal address. We believe that using intersection addressing has the potential to increase responses and survey returns in assessing this critical spatial component. We removed typos and spelling errors of street names and geocoded the data. This process resulted in 697 of the original 949 data points for mapping and analysis. The geocoded data points were used for network analysis with the center of campus - the library - being the destination for all routes.

4.3 Bicycle-Friendly Network Analysis

This study accounts for the bicycle-friendliness of the City of [Removed for Review Purposes] and its metropolitan region rather than using a regular street network in the GIS analysis as other scholars have done (Akar & Clifton, 2009; Aultman-Hall et al., 1997; Krizek & Johnson, 2006). The 25-mile bike path along the river connects cycling infrastructure such as bike lanes and bike routes on local and minor streets in the area. In creating a bicycle-friendly network for analysis, we used existing street networks provided by the local planning agency and added the 25-mile bike path and other bike routes. We did not test for landscape influences, such as those noted by Cervero et al. (2009), due to the predominantly flat terrain of the biking environment.
We used a hierarchy-based ranking system in which the bike path itself received the highest ranking. Streets containing bike routes and bike lanes as well as local streets were assigned the same ranks and therefore considered as bike-friendly. Major roads and interstate highways were allotted the lowest ranking values, setting them apart as being almost bike-unfriendly and less likely to be chosen by a bicyclist.

We introduced the hierarchy-based network to a “hierarchy-over-length” index as a cost factor, dividing such bike-friendliness rank by the length of each GIS segment. For computational purposes, the major bike path as a stand-alone bike path equals a value of 1, and the values range to nearby interstate highways, which are set at a value of 12, with other road categories somewhere in between. This allows trip length to be calculated as a potential cost of cycling, based on our index, while integrating the variation in the friendliness of the bike network as part of the computed route’s value. For example, a one-mile (5250ft) GIS segment would have a value of 1/5280 (or 0.00019) for bike path versus the near-by interstate highway (12/5280 or 0.00223). In short, this calculation creates a measure that simulates attractiveness of a street segment and expresses the likelihood of its being chosen. As such, the “hierarchy-over-length” calculation elegantly functions as a cost factor and accounts for the possibility for a network segment to be selected based on its bike friendliness and not just its length. Network routes based on the “hierarchy-over-length” index were then computed with ArcGIS, the scenario GIS-platform CommunityViz, ESRI’s Network Analyst, and utilized in SPSS and STATA for further analysis.

4.4 Accounting for Spatial Limitations and Errors

We are aware of the potential limitations of the assumed-route measure. It is based on the intersection named by the student while measuring the distance along the bike-friendly network with our index. Only intensive cycling diaries (Chen, 2012) or drawing of daily routes (Snizek et al., 2013) would give a more defensible representation of chosen routes. This would also include the issue of perception errors in mapping and manifesting a bias between observed and computed durations (Kamruzzaman et al., 2011). We presume that the assumed-route approach will function as a proxy and be more likely to accurately reflect the spatial cycling patterns used in the bike-friendly network. This assumption is strongly supported in statements by Buehler and Pucher (2012, p. 428) “that bike lanes and paths encourage cycling” when compared to traditional road network analysis. Houston (2014) outlines neighborhood effects on physical activity (walking) and studies the potential zoning and scaling issues of themodifiable areal unit problem (MAUP). We anticipated that encouragement through our bike-friendly physical built environment (Buehler, 2012) should have very few influences on MAUP, if they exist at all.

Further, analyzing spatial patterns requires testing to determine whether the data is spatially correlated, thus causing biases in the results, first, on blank values (locations only) and, second, with computed measures such as the distance to campus (DIS2CAM). Geocoding at the intersection level will automatically cause clustering of data points; hence, multiple survey responses will be ‘stacked’ on top of each other at one intersection. Tests for spatial autocorrelation with Global Moran’s I, used for measuring autocorrelation between a feature’s value and its specific location, indicate a shift from clustered to disperse distributions with increasing distance. Testing with Ripley’s K Function, that is, measuring spatial dependence across multiple features supports statistically significant clustering at smaller distances between neighboring points. This was expected due to the overlapping caused by our intersection geocoding and is likely to be causing an intersection bias or ‘aggregation errors’ as originally noted by Hewko et al. (2002) and Apparicio (2007). Such bias is considerable in our context due to the grid-style street layout in the suburban parts of our study area and the fact that the survey requests the nearest intersection. To control for aggregation bias and MAUP-related issues (Jacobs-Crisioni et al., 2014), we applied a test for incremental spatial autocorrelation. While increasing the distance incrementally between single data points and their neighbors, this method returns a z-score that echoes the intensity of spatial clustering at a certain distance band (Environmental Systems Research Institute, 2013). The resulting z-scores demonstrate that relative high clustering occurs in lower distances and that clustering drops stepwise (i.e., 3-4,000 ft., 5-6,000 ft., and approximately 7-8,000 ft.) to lower a z-score. We explain this result, the expected clustering, as being due to residential subdivisions following the regional one-mile (5280 ft.) street grid and intersection-geocoding [see also Figure 1]. In short, all tests on spatial limitations and errors indicated that further statistical analysis is appropriate.
4.5 Analyzing Patterns/Groupings of Cycling Behavior

To test our hypothesis that distance should be treated as a dependent variable, we applied the ArcGIS methodology of Grouping Analysis in preparation for statistical analysis. This GIS tool processes input data and selected attributes (variables) and returns recommendations for potential groupings based on feature, attributes, and optional spatial/temporal constraints. It applies a statistical method identifying groups in spatial patterns based on locations and values. This test returns a so-called pseudo-R2 indicating how much influence the selected variable exerts upon the suggested grouping. The higher the pseudo-R2, the higher this corresponding variable is in discriminating among features (Environmental Systems Research Institute, 2013). With known, but controlled for, spatial autocorrelation bias, we ran multiple iterations to find the best groupings among the observations. Specifically, for the grouping analysis we used the noted three distances (DIS2PATH (home to closest bike path), DIS2GREEN (home to regional bike system), DIS2CAM (home to campus library)) and the responses as to whether cycling was used for transportation or recreation.

5. Results and Findings

We determined the optimal number of groups with spatial constraints such as eight nearest neighbors and random location seeding. Random seeding also integrates influences that can be found in a sensitivity analysis (Environmental Systems Research Institute, 2013). Table 1 shows the pseudo-R2 values and how spatial constraints are responsible for the within grouping process: Using a Pseudo-F score, the GIS internally evaluates these groups, where the higher the Pseudo-F score, the more reliable the grouping outcome. Our GIS grouping analysis suggested seven groups (F-score: 360.82) that are acceptable given the spatial clustering of the geocoded locations.

5.1 GIS Analysis of Clusters/Groupings

Table 1 shows further descriptive statistics for the distances from a student’s home to the closest bike-friendly network segment (DIS2PATH), closest distance along the network to the major bike path network along the river (DIS2GREEN), and the total distance from named intersection to the campus library (DIS2CAM). These measures are distinct from most of those in the reviewed literature, which uses a straight line, or categories of distances (Krizek & Johnson, 2006; Millward et al., 2013; Sander et al., 2010). Using a network gives the benefit of topographic and built environment barriers (i.e., near-by interstate or large river) being represented in this continuous measure. In general, the majority of the students live in close proximity to a bike-friendly street or path, as seen later in Cluster 7 (see Figure 1).

The grouping analysis process produced group-amounts of 31, 66, 14, 73, 101, 1, and 410 students as seen in Table 2.

<table>
<thead>
<tr>
<th>Table 1: Variables in GIS groupings process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable*</td>
</tr>
<tr>
<td>Distance to bike path (DIS2PATH, in feet)</td>
</tr>
<tr>
<td>Distance to regional bike path (DIS2GREEN, in feet)</td>
</tr>
<tr>
<td>Distance to campus (DIST2CAM, in feet)</td>
</tr>
<tr>
<td>Cycling for Transportation or Recreation (binary)</td>
</tr>
<tr>
<td>Min</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>1096</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Table includes extreme outliers that have been flagged during GIS statistical grouping but are still reported here. **Pseudo-R2 for GIS grouping
Cluster | Min  | Max  | Mean  | Std. Dev. | N  
---|------|------|-------|-----------|---
1  | 95836.62 | 220997.95 | 146069.43 | 27396.72 | 31 
2  | 14913.55 | 66205.74 | 38720.79 | 15404.28 | 66 
3  | 93531.90 | 216679.53 | 139588.81 | 40156.25 | 14 
4  | 33216.94 | 122113.16 | 71492.90 | 19197.55 | 73 
5  | 8038.51 | 119459.24 | 45017.06 | 18795.15 | 101 
6  | 97006.27 | 97006.27 | 97006.27 | 0.00 | 1 
7  | 1096.38 | 41350.96 | 9924.41 | 8464.30 | 410 

*Table indicates the distance to campus as noted in variable DIS2CAM.

Figure 1 shows the parallel box plot of the variables used in our GIS grouping analysis. The boxplot provides a comparison of the center, symmetry, spread and outliers of the variables used in the grouping process. The individual analysis for each cluster and the shown boxplot support the removal of cluster 6 as extreme outlier. Cluster 6 contains only one student and its measure to the closest bike path (DIS2PATH) is more than 25 units of standardized values offset. For the purpose of the spatial discussion on cycling behavior among college students, six groups have been chosen for further analysis.

**Figure 1:** Parallel Box Plot on all GIS groupings

*NQ67 denotes “Cycling for Transportation”*
As seen in Figure 2, one notable observation is the likely influence of the physical built environment, such as the extensive bike path network along the river and major regional roads. Users in Cluster 7 benefit from a SW-NE orientation of such network as well as an N-S oriented path system. Clusters 2 and 5 have similar bike environs but contain wide roads; the downtown-to-interstate connector and the interstate itself and could function as physical barriers and hinder pleasant cycling for transportation.

Figure 2: Clusters Overview

5.2 Independent and Control Variables

As seen in Table 3, and drawn from the previously discussed literature and survey data, the independent variables include the interval level variable of age, which averages 25.71 years, and several dummy variables. The first dummy variable is female (gender), which represents 68% of the population. As noted, literature suggests that females will be less likely to cycle than males. The second dummy variable is healthy Body Mass Index (BMI) where 55% of the respondents indicate they are of healthy weight. BMI is included for a reason not unlike Faulkner’s (2009) finding that students who actively commute to school are thought to be more active in general. In short, we are testing whether people with healthy body weights are more likely to be cycling for transportation since they may already be more active. The third dummy variable is whether the respondent cycled in the previous 30 days for transportation, which is similar to the concept of habit that Verplanken and Aarts (1999) ascribe foster behavior. Sixty-two percent of study respondents indicated they do cycle for transportation.
Table 3: Descriptive Statistics of Data Used in Analyses, Survey Sample with University Population in Parentheses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>N</th>
<th>Survey Sample Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clusters*</td>
<td>1</td>
<td>7</td>
<td>5.57</td>
<td>1.95</td>
<td>695</td>
<td>NA</td>
</tr>
<tr>
<td>Age</td>
<td>18</td>
<td>62</td>
<td>25.71</td>
<td>8.26</td>
<td>941</td>
<td>(17694)</td>
</tr>
<tr>
<td>Gender: Female =1</td>
<td>0</td>
<td>1</td>
<td>0.68</td>
<td>0.47</td>
<td>695</td>
<td>647 (9367)</td>
</tr>
<tr>
<td>Health (BMI) = 1 (healthy)+</td>
<td>0</td>
<td>1</td>
<td>0.55</td>
<td>0.5</td>
<td>686</td>
<td>925</td>
</tr>
<tr>
<td>Cycled for either transportation or recreation*</td>
<td>0</td>
<td>1</td>
<td>.59</td>
<td>.5</td>
<td>687</td>
<td>943</td>
</tr>
<tr>
<td>Cycled for transportation in last 30 days*</td>
<td>0</td>
<td>1</td>
<td>0.62</td>
<td>0.47</td>
<td>431</td>
<td>555</td>
</tr>
<tr>
<td>Cycled for recreation in last 30 days*</td>
<td>0</td>
<td>1</td>
<td>0.43</td>
<td>0.5</td>
<td>432</td>
<td>556</td>
</tr>
<tr>
<td>Own a car*</td>
<td>0</td>
<td>1</td>
<td>0.88</td>
<td>0.32</td>
<td>693</td>
<td>693</td>
</tr>
</tbody>
</table>

+ healthy weight = BMI 18.5 to 24.9 kg.m-2; Underweight = BMI < 18.5 kg.m-2 or Overweight/Obese = BMI > 25 kg.m-2
* count of cluster

The control variables are dummy variables drawn from the survey. The control variables include whether the respondents indicated they cycled in the previous 30 days for recreation and their car ownership status (see Table 3). Only 42% of the respondents indicated they were cycling for recreation. This control variable is not in the model as a variable relating to the concept of habit-like such as “cycling for transportation.” The reason for including this variable is that being familiar and comfortable with cycling is seen as likely to promote cycling for transportation. Studies have found that cycling for recreation is associated with cycling for transportation, as previously noted, but also as noted in the literature, the direction of the relationship still needs to be better determined. Finally, a dummy variable on car ownership is included because car ownership has repeatedly been found to be negatively related to the propensity to cycle. A vast majority of our study population (88%) own cars.

5.3 Statistical Analysis of Cycling Groupings

We used multinomial logistic regression to determine if there are differences between groups, as they relate to their distance from the center of the campus, and what role the independent variables of age, gender, health, and cycling for transportation played while controlling for cycling for recreation and car ownership. Prior to conducting the regression analysis, a test of Pearson’s r correlation among the dependent and independent variables was conducted. The highest correlation (.369) was between cycling in the previous 30 days for transportation and the GIS clusters, indicating there were no concerns with multicollinearity among the variables. We then used a multinomial logistic regression with Cluster 7, the cluster nearest to the university as the base category for comparison. In the analysis, we calculated the relative coefficient ratios. This permits interpreting the odds of each variable as being more likely to be in a particular cluster over Cluster 7.

Results as shown in Table 4 reveal that age, cycling for transportation, and car ownership are significant in most models when compared to Cluster 7. Cluster 1 was an exception with only car ownership demonstrating a significant impact. Asymptotic z tests concluded that age (LR = 29.79; p < .000), cycling for transportation (LR 57.84; p < .000) and owning a car (LR 11.23 p < .000) do significantly predict the cluster groups. Additionally, as seen in Table 4, the relative coefficient ratios indicate the odds that a student is in Cluster 2 or 5 which are the closest to Cluster 7 as result of their age and controlling for car ownership increased by 8 and 14 % respectively when compared to Cluster 7. The odds that a student is in Cluster 4 and 3, which are further away from Cluster 7 due to age and controlling for
car ownership, increased by 6% and 7% respectively, when compared to Cluster 7. The effect of cycling for transportation on a student being in any particular cluster, compared to Cluster 7 after adjusting for age and controlling for car ownership, decreased anywhere from 93% for Cluster 3 to 88% for Cluster 2.

This means that when we take into account someone having cycled for transportation in the previous 30 days, cycling for transportation reduces the explanation as to why that person is in any particular group. Table 5 describes the mean age of each group and reveals that those students closest to the university on average were considerably younger with a mean age of 22 years. Students in other clusters ranged in age from 27 to 29 years old. Finally, car ownership was a significant factor in explaining distance from the campus for all groups except Cluster 2.

### Table 4: Statistically Significant Multinomial Logistic Regression Estimates of Age, Gender, Healthy Weight, Cycling for Transportation, Cycling for Recreation on 5 GIS Clusters with Cluster 7 as the Base Cluster and Relative Coefficient Ratios (Standard Errors in Parentheses)

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.06</td>
<td>1.08**</td>
<td>1.07+</td>
<td>1.06**</td>
<td>1.14**</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Gender: Female</td>
<td>1.03</td>
<td>0.84</td>
<td>0.9</td>
<td>0.95</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>(0.92)</td>
<td>(0.28)</td>
<td>(0.70)</td>
<td>(0.33)</td>
<td>(0.57)</td>
</tr>
<tr>
<td>Healthy Weight (not over or under weight)</td>
<td>0.7</td>
<td>0.97</td>
<td>0.26</td>
<td>0.7</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>(0.60)</td>
<td>(0.32)</td>
<td>(0.27)</td>
<td>(0.23)</td>
<td>(0.34)</td>
</tr>
<tr>
<td>Cycle for Transportation</td>
<td>0.22</td>
<td>-.22**</td>
<td>-.07*</td>
<td>-.15**</td>
<td>-.12+</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Cycle for Recreation</td>
<td>0.85</td>
<td>1.3</td>
<td>0.68</td>
<td>1.32</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>(0.80)</td>
<td>(0.42)</td>
<td>(0.58)</td>
<td>(0.45)</td>
<td>(2.08)</td>
</tr>
<tr>
<td>Own a Car</td>
<td>1.78e+08**</td>
<td>4.62</td>
<td>6.77e+07***</td>
<td>1.45e+08**</td>
<td>4.48e+07**</td>
</tr>
<tr>
<td></td>
<td>(-3.06E+08)</td>
<td>(-4.81)</td>
<td>(-9.88E+07)</td>
<td>(-1.01E+08)</td>
<td>(-9.09E+07)</td>
</tr>
<tr>
<td>Pseudo R ²</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR Chi²</td>
<td>121.33**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>422</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1 p \leq .05 \text{ (one-tailed)}; \ * p \leq .05; \ ** p \leq .01 \text{ (two-tailed)}\)

### Table 5: Clusters by mean Age (in years)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Mean Age</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.48</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>29.72</td>
<td>66</td>
</tr>
<tr>
<td>3</td>
<td>27.79</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>27.74</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>29.39</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>22.42</td>
<td>409</td>
</tr>
</tbody>
</table>
5.4 Distance Decay Functions

Further, in plotting select distance decay functions, we gained a likely indication on the span of distance and insights into the willingness of bike users to bike farther distances. We calculated distance bands of 1000ft in using the proportion of bike users per band over the total respondents and plotted normalized values with a fitted curve similar to that utilized by Zhao et al. (2003) and García-Palomares et al. (2013). Figure 3 shows graph of distance decay in the case of transportation for biking ($r^2=.74$) and, in comparison, car owners ($r^2=.74$) that used their bikes as well. Based on the decay function, car owners are likely to be less willing to bike to campus, and if they do, to travel shorter distances. Marquet and Miralles-Guasch (2014) report similar effects of car ownership on walking (p.216). In support of our suggestion of 3-4 miles, the graph shows slight peaks in the 3-4 mile distance ranges.

We have similar findings when plotting distance decay for the significant variable age. Using the mean age of 25.71 years, we split data into two groups with ages above and below the mean for visualization. Figure 4 shows the scatter plot of data points and estimated trend lines. It is evident, that the willingness to cycle farther declines more rapidly for those of younger age ($r^2=.63$) than those above the mean age ($r^2=.27$). This low $R^2$ for the above mean age is not surprising as the mean ages are not distribute uniformly with distance as seen in Table 5 and in comparison with the spatial configurations of clusters in Figure 1.

**Figure 3:** Distance decay functions – Cycling for Transportation and Car owners
6. Discussion and Conclusions

We believe our study successfully presents value in considering distance as a dependent variable when investigating cycling behavior. Further, we specifically find that students who cycled for transportation are more likely to cycle regardless of the distance. In general, students that reported cycling activities tended to be younger and live closer to campus: Cluster 7. The fact that car ownership was significant across nearly all clusters suggests that it has an impact, and we note that the impact appears to be greater the farther out from campus that one lives. Additionally, in the case of Cluster 2 which is nearest to Cluster 7, car ownership was not significant, suggesting that there is a span of distance that one is likely to cycle whether they own a car or not. The analysis suggests that this distance appears to be about 3-4 miles (approx. 15,840 – 21,120 ft.). This covers the distances in the ranges of Cluster 2, 5, and 7 in our GIS analysis. In the case of Cluster 7, it is possible that it contains proximity trips, which could be determined by applying Marquet and Miralles-Guasch’s term (2014) for walking trips inside neighborhood limits (around campus; within the cluster) and as an “indicator of local activity”. Cycling extends this activity space through faster commutes (St-Louis et al., 2014) but also could increase health factors as being a motive for active transportation. In the end, factors such as age and propensity to cycle for transportation can help explain the impact of distance, if the distance is not too great as was the case in the farthest outlying group, Cluster 1. Car ownership also appears to play a role in determining the distance one is willing to cycle to campus, except in the case of respondents in Cluster 2. This supports the argument we raised earlier that prior cycling behavior is likely reflected in the off-campus housing allocation due to the intent to commute to campus despite car ownership and that the willingness to cycle for transportation is moderated by other trade-off factors and personal characteristics.

Distance is less distinct as an explanation for understanding the spatial cluster of cyclists when we consider age, those that already cycle for transportation (in the previous 30 days) and those who own cars. The findings on car ownership and distance decay also concur with prior findings by Balsas (2003), Bopp et al. (2011), and Nayar (2012) that revealed the importance of a cycling policy on a university campus due to the influence on nearby communities and changes in behavior due to education. Increasing cycling rates among residents holds the promise of helping both individuals and communities. However, caution is warranted to not overstate the potential impact of
cycling on the transportation sector. Pucher, Garrard, et al. (2011) found those most likely to cycle will also use transit for shorter distance if the option is available, faster and cheaper than cycling. Similarly, Park et al. (2014) found that when comparing those that commute to work by car and those that use transit, the transit users are more likely to choose cycling for their commute. Finally, Xing et al. (2010) found cyclists that also enjoying driving, decreased the miles they chose to cycle for transportation.

Gender and general health were not significant predictors as expected in any of the cluster distances. One potential explanation to the insignificance of gender is that the bicycle-friendly nature of the terrain and of the community make cycling to campus appealing to both genders, students at all distances, and those within varying levels of physical fitness. The fact that one’s personal health, in terms of being a healthy weight, was not significant in any of the clusters of biking for transportation may be similar to a finding by Xing et al. (2010) in which people reporting themselves as healthy but associated themselves with cycling for recreation and not for transportation. Additionally, the fact that cycling for recreation was not statistically significant is somewhat surprising as it is reasonable to believe that people who are more comfortable with cycling would also be more likely to cycle for transportation; however, again, this may be related to Xing et al.’s (2010) findings. They suggest that recreational cyclists may be more experienced and prepared and as such, more likely to go longer distances than cyclists would go for transportation. This may result in recreational cyclists being influenced more by their own personal characteristics or behaviors when considering the perceived effect of distance on their decision to cycle than other factors. The findings of St-Louis et al. (2014) and their approach on “trip satisfaction” among commuters and potential “higher enjoyment” (p.16) using a bike for the commute also suggest this is the case as well. On the other hand, the direction of the relationship between the two types of cycling may be that of commuting practices influencing recreational use, as pointed out recently by Kroesen and Handy (2014). The perception of safety is also often cited as a barrier to cycling for transportation (Bauman et al., 2008) and longer distances can exacerbate perception of safety, with farther distances being perceived as less safe when cycling for transportation (Xing et al., 2010). This may suggest one element not tested here, safety of the route or perception of safety, may also play a mediating role in the one’s willingness to cycle specific distances for some types of riders.

Overall, our findings based on the concept of hedonic models reveal distance can be treated as dependent variable in cycling behavior studies. Distance matters to cyclists as many of the presented studies found (Emond & Handy, 2012; Handy et al., 2010; Heinen et al., 2010). However, treating distance as a dependent variable on cycling behavior creates a specific framework to consider for policy and infrastructure applications. It helps explain the way variables such as age, gender, car ownership, or personal health moderate or not the impact of distance for cycling for transportation. The findings enable policy makers to target the specific, distance-based, areas with targeted programs for, residents of certain age and economic means (here, students) with options such as bike share, park’n’ride, or enhanced public transit connectivity to bridge the ‘gap’ between distance bands as found in some of our clusters. One recommendation drawn specifically from these findings could address the need for parking facilities, their maintenance, and the resulting costs of parking on campus. By creating park-n-bike facilities with bike share stations at the approximate distance of Cluster 2, which is 3-4 miles, it may be possible to reduce the parking facilities needed on campus as well as encouraging more active life-styles among students of all ages, car-owner, etc. However, it should be noted that a bike-friendly network may need to be in place or established prior to realizing the benefits of this policy. There is evidence that bike infrastructure along with distance are important to outcomes when comparing Cluster 5 to Cluster 2. Cluster 5, which lacks the bike-friendly infrastructure found in Cluster 2, did not result in outcomes similar to those found in Cluster 2; however, Cluster 5 is located in approximately the same distance from campus as Cluster 2.

Another policy outcome may be the integration of campus bike planning into local government plans so that more cycling to campus could be accommodated by a wider variety of students, regardless of the distance they live from campus. In our local case, it appears that the City of [Removed for review] could increase the bicycle-friendly path network not just around campus but also in other clusters. For example, Cluster 7 that is associated with the immediate adjacent community to campus may benefit from additional planning for cyclist to advance existing infrastructure policy and plans. In addition, the results found for Cluster 2 and 5 present intermediate proximity and significant cycling patterns (in terms of significant variables) and have the potential to be addressed by the university and community. Efforts such as this could help alleviate perceived barriers to cycling and in turn increase the likelihood that a wider demographic and more spatially diverse groups cycle. Additionally, it is possible that the more cycling infrastructure for commuters could have a multiplier effect on cycling in general. Considering the findings and using the concept of location-efficient mortgage where distance is already included in the individual’s
decision to live in a specific area it appears public policies can influence cycling for transportation. Whereas Manaugh and El-Geneidy (2013) presented the ability to walk to campus as reason of location choice, our analysis suggest that Cluster 2, 5, and 7 are preferred locations – in line with proximity and cycling friendliness. As Balsas (2003) pointed out, campus policy impacts nearby communities. Our analysis demonstrates that communities in the nearby proximity and those farther out could benefit from such policies.

Ultimately, the findings here suggest that, at least in terms of cycling for longer distances, bike networks and other infrastructure that reduce effort, time or congestion can help cycling be a more competitive mode of transportation. One finding is that infrastructure improvements such as smooth transitions across bike networks and/or having infrastructure that promotes cycling efficiency can increase the likelihood of cycling for commuting. These types of improvements could be a means of actually competing with as well as being a substitute for at least a portion of an otherwise less sustainable commute. Bike share programs could also be part of the partial segment transition option, with bike share stations located at park-n-ride locations or park-n-bike areas created for just that purpose.

In the end, understanding the factors that make distance more or less of a factor in the decision or willingness to be a commuter cyclists could go a long way toward motivating people who were previously not likely to cycle for transportation. Studies suggest bike amenities such as showers at work are helpful in promoting commuting by bike, but research suggests that these are not necessarily sufficient types of infrastructure for overcoming the impact of distance on the willingness to commuter cycle. Dedicated bike paths, cycle-tracks, bike lanes, and/or park-n-ride/bike facilities can render cycling more efficient and attractive for commuting by enabling cyclists to overcome farther distances with ease and in feasible durations.

ROLE OF FUNDING SOURCES DISCLOSURE
This research and/or the preparation of the manuscript were supported by funding from the Department of Public Policy and Administration and the Department of Kinesiology at [removed].

CONFLICT OF INTEREST DISCLOSURE
The authors have no conflicts of interest to report. The authors confirm that the research presented in this article met the ethical guidelines, including adherence to the legal requirements, of the United States and received approval from the Institutional Review Board of [removed].
References


Hansen, K. B., & Nielsen, T. A. S. (2014). Exploring characteristics and motives of long distance commuter cyclists. *Transport Policy, 35*(0), 57-63. doi: [http://dx.doi.org/10.1016/j.tranpol.2014.05.001](http://dx.doi.org/10.1016/j.tranpol.2014.05.001)

Heinen, E., Maat, K., & van Wee, B. (2011). The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. *Transportation Research Part D: Transport and Environment, 16*(2), 102-109. doi: [http://dx.doi.org/10.1016/j.trd.2010.08.010](http://dx.doi.org/10.1016/j.trd.2010.08.010)


Jacobs-Crisisoni, C., Rietveld, P., & Koomen, E. (2014). The impact of spatial aggregation on urban development analyses. *Applied Geography, 47*(0), 46-56. doi: [http://dx.doi.org/10.1016/j.apgeog.2013.11.014](http://dx.doi.org/10.1016/j.apgeog.2013.11.014)


