Station-Level Forecasting of Bike Sharing Ridership:
Station Network Effects in Three U.S. Systems

R. Alexander Rixey
Transportation Planner
Fehr & Peers
600 Wilshire Boulevard, Suite 1050
Los Angeles, CA 90017
Tel: +1 (213) 261-3055
Fax: +1 (310) 394-7663
arixey@fehrandpeers.com

Original Submission Date: August 1, 2012
Revised Submission Date: November 15, 2012
Word Count: 5,989
Figure Count: 2
Table Count: 4

ABSTRACT

This study investigates the effects of demographic and built environment characteristics near bike sharing stations on bike sharing ridership levels in three operational U.S. systems. While previous studies have focused on the analysis of a single system, the increasing availability of station-level ridership data creates the opportunity to compare experiences across systems; particular attention is paid to data quality and consistency issues raised by a multi-city analysis. This project also expands on previous studies by including the network effects of the size and spatial distribution of the bike sharing station network, contributing to a more robust regression model for predicting station ridership.

The regression analysis identifies a number of variables as having statistically significant correlations with station-level bike sharing ridership: population density; retail job density; bike, walk, and transit commuters; median income; education; presence of bikeways; non-white population (negative association); days of precipitation (negative association); and proximity to a network of other bike sharing stations. Proximity to a greater number of other bike sharing stations exhibits a strong positive correlation with ridership in a variety of model specifications and while controlling for the other demographic and built environment variables, suggesting that access to a comprehensive network of stations is a critical factor supporting ridership. Relative to previous models, this model will be more widely applicable to a diverse range of communities and help those interested in adopting bike sharing systems to predict potential levels of ridership and identify station locations that will serve the greatest number of riders.

INTRODUCTION

Public bike sharing systems, which provide users with short-term access to bicycles through an automated kiosk, are becoming increasingly prevalent both in the United States and around the world (*1*) – U.S., IT-based systems have expanded from two systems in 2009 to 17 systems in 2012, with another 17 systems planned to launch in the U.S. during 2012 (*2*). Members typically pay a membership fee to join for a year, month, or day and may take an unlimited number of free trips as long as each trip lasts less than 30 minutes. For each trip longer than 30 minutes, an hourly user fee typically applies.

As more jurisdictions and, increasingly, private companies plan to implement bike sharing systems, the question of feasibility is at the forefront. In order to assess feasibility, planners need to determine an appropriate service area and the number, size, and location of bike sharing stations. With this information they can develop reasonable estimates of capital and operating costs. Accurate estimates of ridership at potential station locations can help planners to locate stations to maximize system-wide ridership.

This study investigates the effects of demographic, built environment, and bike sharing network characteristics near bike sharing stations on ridership levels at stations in three operational systems: Capital Bikeshare in Washington, D.C.; Nice Ride MN in Minneapolis/St. Paul, Minnesota; and Denver B-Cycle in Denver, Colorado.

PREVIOUS BIKE SHARING RIDERSHIP STUDIES

With the spread of bike sharing systems and the growth of bike sharing ridership, a number of studies have attempted to assess the feasibility of bike sharing systems, identify appropriate service areas, and forecast station-level and system-wide ridership. These studies provide a methodological basis for identifying service areas and forecasting ridership and identify a variety of relationships between demographic and built environment variables and bike sharing ridership.

Several studies apply demographic and spatial variables to create a weighted sum raster analysis, or "heatmap," that forecasts relative levels of ridership in a hypothetical service area by overlaying a grid on the study area, aggregating the input variables into a suitability score, and selecting a service area based on relative suitability values. Krykewycz et al (*3*) conducted such a weighted sum raster analysis for the Philadelphia, PA market, selecting input variables based on an understanding of theory, rather than an empirical analysis of relationships to ridership. For Seattle, WA, Gregerson et al (*4*) also created a weighted sum raster analysis of twelve indicator variables, including population and job densities, companies with commute trip reduction programs, tourist attractions, parks and recreation areas, topography, local and regional transit stations, bicycle friendly streets, and streets with bicycle lanes. These variables were selected based on their inclusion in Krykewycz et al (*3*) and supplemented with the topography and commute trip reduction program variables to reflect researcher understanding of conditions in Seattle (*4*). Olson et al (*5*) apply a similar approach to Providence, RI, including some of the variables above as well as population density of 20-49 year olds, and proximity to colleges, libraries and historic places. Again, however, it is not evident that the authors empirically tested their weighted sum raster analysis variables against actual ridership data. Finally, Krykewycz et al (*3*) and Gregerson et al (*4*) estimate system ridership by applying diversion rates from other modes to bike sharing based on data from Lyon, Paris, and Barcelona. Maurer (*6*) notes that this diversion rate approach has also been applied to Vancouver and New York City. Olson et al (*5*) do not provide a ridership estimate.

Other studies have undertaken empirical analyses of bike sharing ridership determinants. Buck and Buehler (*7*) study daily bike sharing ridership in Washington, D.C. at the station level, finding that total population, the supply of bike lanes, and the number of liquor license holders (a measure of retail destinations) are positively associated with ridership, while the percentage of households without access to a car is, in contrast with theory and intuition, negatively associated with ridership. Daddio (*8*) also examines Washington, D.C. at the station level including, among other variables, the distance from the ridership-weighted average center of the bike sharing system, one measure of the effect of the bike sharing station network. Daddio (*8*) finds that this variable has a strong significant association with ridership – the farther from the system center, the lower ridership levels are. This variable might not fully reflect access to a network of bike sharing stations, however; for example, two stations equidistant from the system center could have different concentrations of nearby bike sharing stations. Hampshire and Marla (*9)* study trip generation and attraction factors in Barcelona and Seville, Spain at an hourly, "sub-city district" level, finding that the number of bike stations within a district, population density, and labor market size are strong indicators of trip generation and attraction. Although the study uses a fine temporal scale, each observation represents the arrival rate or departure rate for a given hour and district, not necessarily an individual station; thus, it is unclear whether the relationship between the number of bike sharing stations in a district and departure rate suggests a network effect, or simply the aggregation of more locations from which trips are made.

Finally, Maurer (*6*) combines empirical analysis of existing bike sharing ridership in Minneapolis/St. Paul, MN with the weighted sum raster approach applied to Sacramento, CA. The regression analysis incorporates 16 independent variables, collected at the station level, and is refined to maximize total model R^2; significance of individual variables was not emphasized. As a result, some independent variables have counterintuitive coefficients. The number of total jobs has a negative coefficient, suggesting, contrary to theory and intuition, that dense employment centers would be poor locations for bike sharing. The total jobs variable may be offset partially by retail jobs and high-income jobs variables in some cases. Similarly, the total population variable has a negative but not significant relationship with ridership. Finally, the presence of bikeways has a negative but not significant relationship with ridership. These counterintuitive results might be attributable to multicollinearity among the variables and a large number of independent variables relative to the number of observations (n=65). Maurer does not include a network effects variable, but acknowledges the importance of network effects among bike sharing stations and recommends careful consideration of the complex interactions among stations (*6*).

The current project builds upon these methodologies and addresses the following limitations of the previous studies: 1) the lack of an empirical analysis of input variables; 2) the study of European systems, which might be less applicable to communities in the U.S. than studies of other operational U.S. systems; 3) the analysis of only a single system; 4) limited measures of bike sharing station network effects; and 5) the inclusion of variables with relationships to ridership that are counterintuitive or in conflict with theory.

METHODOLOGY

A regression analysis was performed using stations in the Capital Bikeshare, Denver B-Cycle, and Nice Ride MN systems as observations (n=264) and the natural log of average monthly rentals by station as the dependent variable. A consistent dataset of independent variables was collected across all three systems and compiled using Quantum GIS and a custom toolbox developed in ESRI ArcMap 10.1; variables selected are widely available so that this analysis can be expanded as additional ridership data become available. Bivariate correlations between each independent variable and the dependent variable were conducted in IBM SPSS Statistics 19 to determine which variables should be included for regression analysis. A multivariate linear regression was then refined to establish a predictive model of ridership in the three input systems. Finally, bivariate regression of the most significant network effects variable against the dependent variable was conducted to explore the robustness of the relationship, both within each system and across all three.

This study expands upon the general regression methodology of Buck and Buehler (*7*), Daddio (*8*), and Maurer (*6*), but differs in a few key respects. First, this study incorporates a consistent set of data from three bike sharing systems, rather than a single system alone, improving the robustness of the regression results. Second, this study focuses on maintaining the intuitive direction and statistical significance of the independent variables used in the regression, rather than only maximizing total model R^2. Finally, this study includes a measure of the network of bike sharing stations – the number of stations within a given distance of the station being analyzed. Although other studies have included rough measures of network effects (*8*),(*9*), this study's approach of centering the network effect variable on each analyzed station and investigating a range of distances helps to differentiate the stations and provides a comparable measure across systems. Like Buck and Buehler (*7*), Daddio (*8*), and Maurer (*6*), the present study uses inputs from operational, U.S. systems to improve the applicability of the model to other U.S. communities interested in pursuing a bike sharing system.

DATA

This section defines the variables tested in the regression analysis, discusses the process of compiling the regression dataset, addresses data quality, consistency and limitations, and presents descriptive statistics of the data.

The natural log of first season monthly average rentals, by station, served as the dependent variable. The natural log was selected, rather than directly using monthly average rentals, in order to help linearize the variable, to improve the continuity of a discrete count variable, and to address the positive skew of the monthly average rentals variable (*10*). The independent variables address a variety of demographic, built environment, and transportation network factors, collected for all three cities so that a consistent dataset could be created across the systems. The system-specific factors are the same for all stations within a given system, and are included to account for attributes specific to each city or system that could not be accounted for by the other variables.

Table 1 presents definitions of all variables considered for the regression analysis. Unless otherwise specified, variables are based on a 400-meter buffer around each bike sharing station to account for a catchment area of users likely to walk to the station.

TABLE 1 Variable Definitions

Variable	Definition	Source
Dependent		
ln(Monthly Rentals)	Natural log of the number of rentals during each system's first operating season, by station; normalized by number of months in first operating season	Bike sharing system operators
Independent		
<u>Demographic Factors</u>		
Population [1]	Total population (in 100s of persons)	U.S. Census Bureau, 2010
Jobs [1]	Total jobs (in 100s), by work area	Longitudinal Employer-Household Dynamics, 2010
High-Income Jobs [1]	Number of jobs (in 100s) paying more than $3,333 per month, by work area	Longitudinal Employer-Household Dynamics, 2010
Retail Jobs [1]	Total retail jobs (in 100s)	Longitudinal Employer-Household Dynamics, 2010
Alternative Commuters [2]	Proportion of workers who commuted by bicycle, walking, or public transportation (100s of workers)	U.S. Census Bureau, 2010
Median Income [2]	Median household income (in 1,000s of dollars)	U.S. Census Bureau, 2010
Non-White Population [2]	Proportion of population that is of a race other than "white alone"	U.S. Census Bureau, 2010
Low-Vehicle Households [2]	Proportion of workers with access to zero or one vehicles.	U.S. Census Bureau, 2010
Bachelor's Degree [2]	Proportion of population over the age of 25 whose highest educational attainment is a bachelor's degree	U.S. Census Bureau, 2010
Graduate Degree [2]	Proportion of population over the age of 25 whose highest educational attainment is a graduate or professional degree	U.S. Census Bureau, 2010
<u>Built Environment Factors</u>		
College	1 if a college is located within 400 meters, 0 otherwise	U.S. Census Bureau TIGER/Line Shapefile 2009 – Area Landmarks
Park	1 if a park is located within 400 meters, 0 otherwise	DC Office of the Chief Technology Officer; Open Street Map
<u>Transportation Network Factors</u>		
Bikeways	Length of existing bike lanes and paths (in 100s of meters)	District Department of Transportation; Denver GIS; Minnesota Department of Transportation
Bus Stops	Number of bus stops (in 10s of stops)	Washington Metropolitan Area Transit Authority; District Department of Transportation; Denver GIS; Metropolitan Council GIS
Stations Within [X] Meters	Number of bike sharing stations within [X] meters	Bike sharing system operators
<u>System-Specific Factors</u>		
DC Flag	1 if station is in Capital Bikeshare system, 0 otherwise	
DN Flag	1 if station is in Denver B-Cycle system, 0 otherwise	
MN Flag	1 if station is in Nice Ride MN system, 0 otherwise	
Precipitation Days	Average days per system operating month with precipitation 0.01 inches or more	National Climatic Data Center

[1] Summed proportionally by area intersecting 2010 Census Blocks
[2] Weighted average by area of buffer intersecting 2010 Census Tracts

Data Compilation, Quality, Consistency, and Limitations

Developing a multi-city dataset presented several challenges in gathering comparable variables across the three systems. This section discusses 1) the approach used in preparing each group of variables for the regression dataset, 2) concerns regarding the quality and consistency of the data, and 3) potential implications of the data concerns for the model. A custom geoprocessing toolbox was created to speed the compilation of the regression dataset; the tools were used for all variables except the system-specific factors.

Bike Sharing Rentals

Bike sharing station rental data were collected from the bike sharing operators. Information technology connected to the docks themselves ensures consistent records of the number of times bikes have been checked out from each station. Rental data from each system's first season of operations was used – Capital Bikeshare data spanned October 2010 through September 2011, Denver B-Cycle data spanned March 2011 through September 2011, and Nice Ride MN data spanned April 2011 through October 2011. The total number of checkouts over the course of the season for each station was divided by the number of operating days in the season and reported on a monthly basis; where data were available regarding the opening date of a station, that information was taken into account in the monthly average as well.

The temporal nature of the rental data presents some challenges, however. Because the three systems launched at different times and two of the systems, Nice Ride MN and Denver B-Cycle, closed their systems for the winter, it was not possible to use data from precisely the same time period. After system launch, ridership of bike sharing systems tends to increase over time as awareness of the system grows and more users are able to become long-term members. Using only data during the period of overlap among the three systems (April through September) would exclude the first six months of Capital Bikeshare's operations, when ridership was relatively low and would thereby overstate average monthly ridership in that system. On the other hand, including only the first seven months of operation for each system would understate Capital Bikeshare ridership by including only the months of October through April, which might generate lower ridership than the milder months following the launch of the other two systems. Faced with these concerns, this project uses the first full season (or year) for each system, to best reflect the ongoing pattern of operations. Rental levels for Capital Bikeshare might still be slightly overstated relative to the other two systems, since the average includes months toward the end of the first year, giving the Capital Bikeshare system more time to attract riders and grow membership.

Census Block-Level Data

The Population, Jobs, High-Income Jobs, and Retail Jobs variables were collected from 2010 Census and Longitudinal Employer-Household Dynamic data at the Census Block level. The data were aggregated to the 400-meter buffer surrounding each station with a sum weighted by the proportion of the area of the intersection of the buffer and each Census Block to the entire area of each intersected Census Block. Because of the fine spatial granularity of the Census Blocks, each 400-meter buffer intersects multiple Census Blocks; many Census Blocks are entirely contained within a buffer (see Figure 1). Because of this fine scale, the Census Block aggregation process accurately reflects conditions within 400 meters of the bike sharing station. The 2010 date of the Census data also accurately reflects conditions during the 2010-2011 period represented by the station ridership data. Finally, the use of a single data source helps to ensure consistency across the three cities, and makes this analysis readily scalable to include other cities and bike sharing systems.

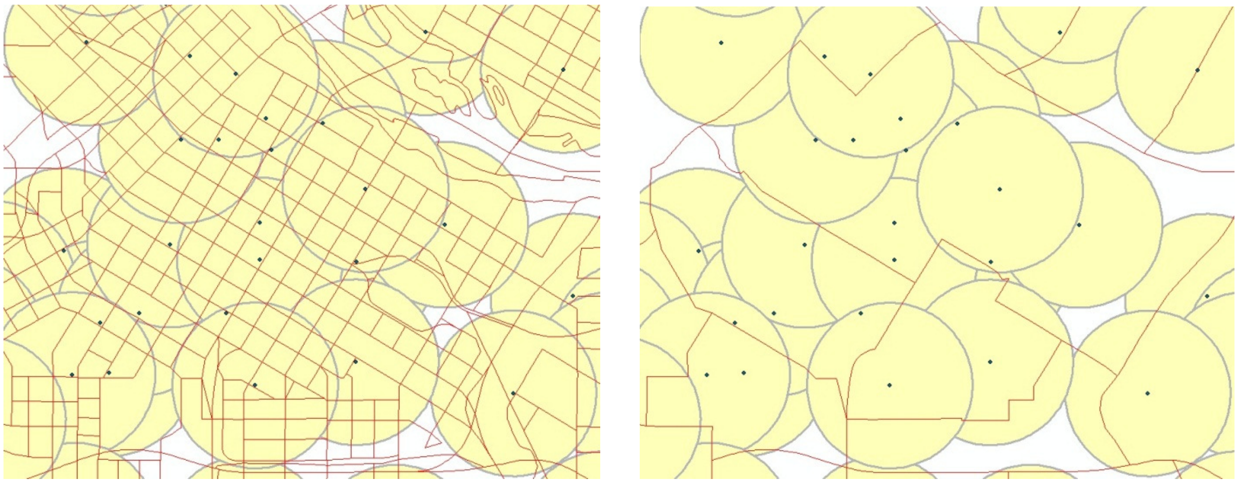

Figure 1 Nice Ride MN Station Buffers, Census Blocks (left), and Census Tracts (right).

Census Tract-Level Data

The Alternative Commuters, Median Income, Non-White Population, Low-Vehicle Households, Bachelor's Degree, and Graduate Degree variables were collected from 2010 Census data at the Census Tract level. The data were aggregated to the 400-meter buffer surrounding each station with a mean weighted by the proportion of the area of the intersection of the buffer and each Census Tract to the sum of the areas of each intersected Census Block. Like the Census Block-level data, these variables reflect the 2010-2011 ridership data period well and are consistent across the three cities; however, the coarse spatial granularity of the data result in proportional averages that might be less reflective of the actual conditions in the 400-meter buffers surrounding the bike sharing stations (see Figure 1). Furthermore, using an average introduces the issue of census tracts with zero or extremely low values of the variables. To address this issue, some unpopulated areas and other outliers were removed from the data – for example, the Census Tract covering the National Mall in Washington, DC was removed from the dataset; however, review of every Census Tract was not possible given resource constraints and limitations on local knowledge, so some outlying values may remain.

Built Environment Factors

The built environment factor data for the Colleges and Parks variables were collected as polygons and intersected with the station buffers to determine whether a college or park fell within 400 meters of a bike sharing station.

The shapefile for colleges was collected from a single dataset, enabling a consistent data collection methodology across the three cities. Colleges were identified by reviewing the attributes and searching for appropriate terms (e.g., "college," "university"). The data do not differentiate among institutions based on size of student population, type of institution (community college, four-year institution, or major research university), or whether students are predominantly residents or commuters, all factors which could influence bike sharing ridership.

The parks shapefiles were gathered from government agencies (where available) or from Open Street Map data, screened for parks and recreational facilities. Although the shapefiles appear to be consistent with the expected locations of parks based on a review of Google Maps for each city, the different sources of park shapefiles introduce the possibility of inconsistencies among the cities in the comprehensiveness of data or the types of facilities included.

Bikeways and Bus Stops

Bikeway data were collected from government agencies in each city. The total length of all bikeways within 400 meters of the bike sharing station (in 100s of meters) was summed to create the Bikeways variable. A bikeway was included in the analysis if a review of the shapefile's attribute table suggested it was a Class I (separated exclusive right of way for bikes alone or bikes and pedestrians with minimal cross traffic) or Class II (on-street striped lane designated exclusively for bike travel) facility. The quality of descriptive attributes was not entirely consistent across cities, particularly in terms of paving treatment and whether or not a bikeway is separated from traffic. The extent to which each jurisdiction's bikeways shapefile might be out of date or incomplete introduces additional potential inconsistencies.

Similarly, bus stop locations were collected from multiple agencies, introducing potential inconsistencies in the completeness of data and the types of bus stops (e.g., express, local, commuter, buses from other jurisdictions) included in the dataset. The number of bus stops within 400 meters of the bike sharing station (in 10s of stops) constitutes the Bus Stops variable.

Network Effects

The Stations Within [X] Meters variables were created using point shapefiles of the locations of bike sharing stations in each of the systems. Buffers with radii of 200, 400, 600, 800, 1200, 1600, 2400, 3200, 4000, 4800, 5600, and 6400 meters were created around each station. The count of bike sharing station points falling within each buffer was then recorded for each buffer radius. These variables provide a way to assess the availability of destination bike stations from a given bike station at a variety of scales.

Precipitation Days

The Precipitation Days variable is based on a 60-plus year average of climatic conditions from the National Climatic Data Center (NCDC). The NCDC data present the average number of days with precipitation greater than 0.01 inches by month. The Precipitation Days variable takes a simple average of these values for the months during which the system is in operation. To the extent that precipitation might have been abnormal during 2010-2011, these data might not represent actual conditions.

Descriptive Statistics

The stations in the Capital Bikeshare (n=98), Denver B-Cycle (n=51), and Nice Ride MN (n=115) systems provided a total of 264 observations with comparable data. The Arlington, VA stations in the Capital Bikeshare system were excluded from the analysis due to a lack of available bike infrastructure, bus stop, and parks data.

Figure 2 provides descriptive statistics of selected regression variables for each system individually and for the overall regression dataset. The variables that differed most among the systems are featured. The vertical lines of each box represent the first quartile, median, and third quartile values of the variable, while the left and right whiskers indicate the minimum and maximum values. Monthly Rentals are presented before taking the natural log for more intuitive interpretation; all other variables are presented as defined in Table 1.

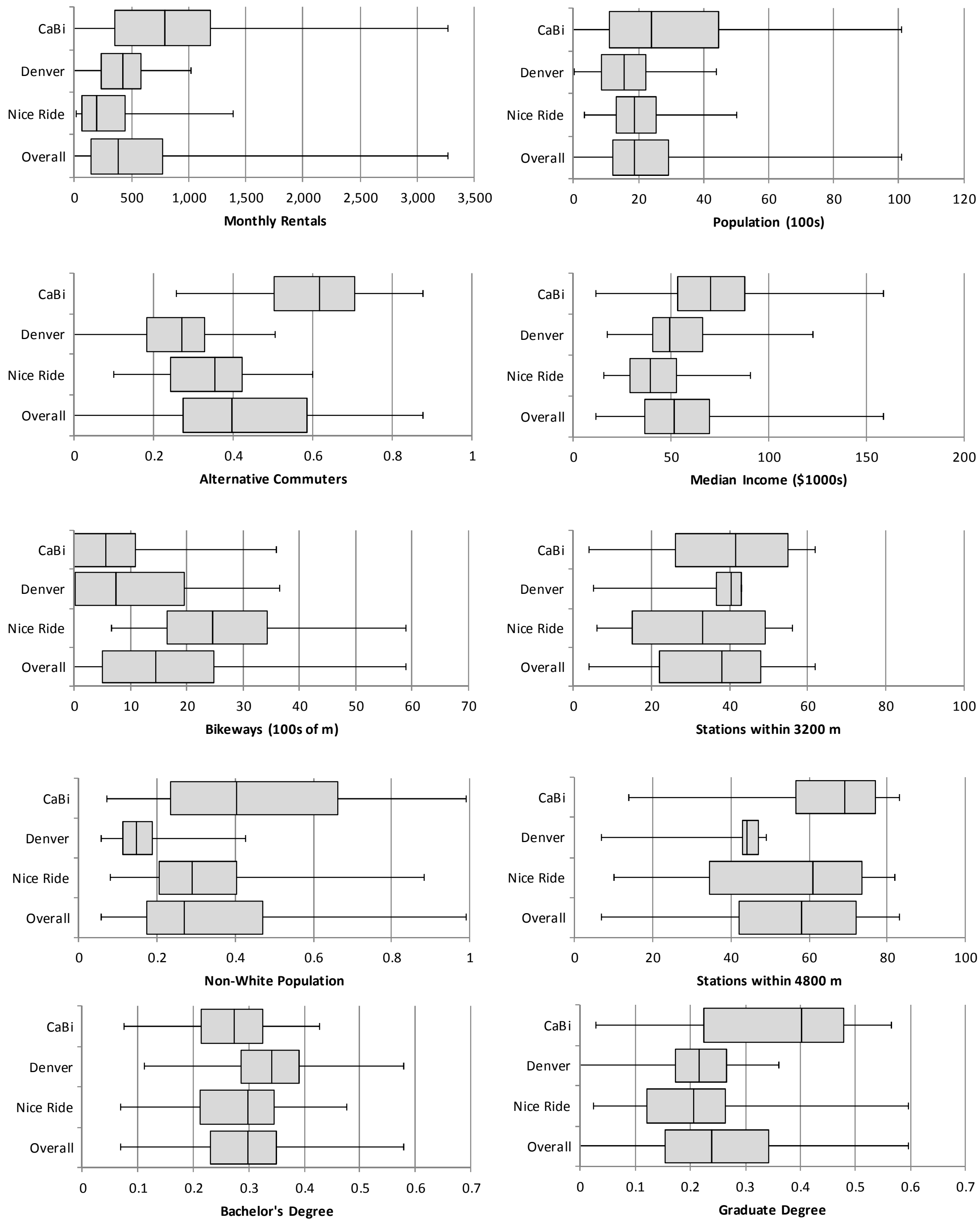

Figure 2 Box and Whisker Plots of Selected Regression Variables (n=265)

Capital Bikeshare reported the highest monthly average total rentals, as well as the highest median and maximum per-station values. Although Denver B-Cycle reported a higher median monthly, per-station rental level than Nice Ride MN, Nice Ride reported higher monthly average total rentals due to having more than twice as many stations as Denver.

Capital Bikeshare also had the highest median values of the Population, Alternative Commuters, Median Income, Graduate Degree, and Stations Within 4800 Meters (among other distances) variables. On the other hand, Capital Bikeshare had the lowest values of the Bikeways variable.

RESULTS

This section describes the process of identifying regression variables and developing a regression model of bike station ridership and discusses the results of the model.

Identification of Regression Variables

Regression variables were selected based on a review of bike sharing ridership estimation literature (*3*),(*4*),(*5*),(*6*),(*7*),(*8*),(*9*), intuition regarding relationships to ridership, and availability of consistent data sources across the three cities selected for analysis. The station network variables (“Stations Within [X] Meters”) were included to test the effects of bike sharing station network density, distribution and size on ridership. Bivariate correlation analysis shows that all of the independent variables have the expected relationship with bike sharing rentals (see Table 2). The station network variables were all significantly correlated with ridership at the 1% level, as were the majority of the other independent variables tested. Only the Park and DN Flag variables do not show a significant correlation with ridership.

TABLE 2 Bivariate Correlations with ln(Monthly Rentals) (n=265)

Variable	Pearson Correlation	Significance (2-tailed)
Population	0.346	0.000***
Jobs	0.402	0.000***
High-Income Jobs	0.390	0.000***
Retail Jobs	0.268	0.000***
Alternative Commuters	0.527	0.000***
Median Income	0.405	0.000***
Non-White Population	-0.442	0.000***
Low-Vehicle Households	0.503	0.000***
Bachelor's Degree	0.512	0.000***
Graduate Degree	0.575	0.000***
College	0.129	0.036**
Park	0.061	0.321
Bikeways	0.106	0.085*
Bus Stops	0.399	0.000***
Stations Within 200 Meters	0.189	0.002***
Stations Within 400 Meters	0.325	0.000***
Stations Within 600 Meters	0.504	0.000***
Stations Within 800 Meters	0.508	0.000***
Stations Within 1200 Meters	0.588	0.000***
Stations Within 1600 Meters	0.623	0.000***
Stations Within 2400 Meters	0.667	0.000***
Stations Within 3200 Meters	0.686	0.000***
Stations Within 4000 Meters	0.679	0.000***
Stations Within 4800 Meters	0.643	0.000***
Stations Within 5600 Meters	0.596	0.000***
Stations Within 6400 Meters	0.542	0.000***
DC Flag	0.281	0.000***
DN Flag	0.094	0.126
MN Flag	-0.349	0.000***
Precipitation Days	-0.309	0.000***

*, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Bike Sharing Station Ridership Multivariate Regression Models

Multivariate linear regression models were refined in order to 1) maximize the predictive power of the model as a whole, as measured by the model R^2, 2) incorporate a variety of independent variables, and 3) maintain statistical significance and intuitive direction of the included variables. Independent variables with a high degree of multicollinearity, such as Alternative Commuters and Low-Vehicle Households, or the multiple Jobs or Stations Within [X] Meters variables, were pared from the model to ensure that each included variable was statistically significant and of the theoretically expected direction. The preferred models are presented in Table 3.

TABLE 3 Multivariate Regression Results of Preferred Models (n=265)

	Coefficient			Standard Error			p-Value		
Variable	*Model 1*	*Model 2*	*Model 3*	*Model 1*	*Model 2*	*Model 3*	*Model 1*	*Model 2*	*Model 3*
Population	0.011	0.012	0.011	0.002	0.002	0.002	0.000***	0.000***	0.000***
Retail Jobs	0.026	0.032	0.026	0.010	0.011	0.010	0.011**	0.005***	0.013**
Alternative Commuters	1.309	1.723	1.558	0.268	0.284	0.478	0.000***	0.000***	0.001***
Median Income	0.006	0.008	0.006	0.002	0.002	0.003	0.000***	0.000***	0.019**
Non-White Population	-2.307	-2.037	-1.982	0.165	0.302	0.337	0.000***	0.000***	0.000***
Bachelor's Degree	–	0.014	0.008	–	0.009	0.008	–	0.095*	0.290
Stations Within 4800 Meters	0.030	0.026	0.029	0.003	0.003	0.003	0.000***	0.000***	0.000***
Bikeways	0.010	–	0.008	0.003	–	0.004	0.006***	–	0.042**
Precipitation Days	-0.743	–	–	0.093	–	–	0.000***	–	–
DC Flag	–	–	0.465	–	–	0.255	–	–	0.069*
DN Flag	–	–	0.944	–	–	0.121	–	–	0.000***
Constant	10.629	3.015	2.742	0.898	0.314	0.358	0.000***	0.000***	0.000***

	Model 1	*Model 2*	*Model 3*
Independent Variable	ln(Monthly Rentals)	ln(Monthly Rentals)	ln(Monthly Rentals)
R^2	0.808	0.760	0.809
Adjusted R^2	0.802	0.754	0.801

Pearson Correlation Coefficient

	Population	*Retail Jobs*	*Alternative Commuters*	*Median Income*	*Non-White Population*	*Graduate Degree*	*Stations Within 4800 Meters*	*Bikeways*
Population	1.000	-0.096	0.322	0.037	0.050	0.106	0.306	0.027
Retail Jobs	-0.096	1.000	0.108	0.170	-0.166	0.223	0.135	0.027
Alternative Commuters	0.322	0.108	1.000	0.129	0.161	0.525	0.634	-0.105
Median Income	0.037	0.170	0.129	1.000	-0.255	0.632	0.113	-0.136
Non-White Population	0.050	-0.166	0.161	-0.255	1.000	-0.435	0.048	-0.152
Graduate Degree	0.106	0.223	0.525	0.632	-0.435	1.000	0.224	-0.130
Stations Within 4800 Meters	0.306	0.135	0.634	0.113	0.048	0.224	1.000	0.280
Bikeways	0.027	0.027	-0.105	-0.136	-0.152	-0.130	0.280	1.000

*, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Discussion

All independent variables in the preferred regression expressed the theoretically expected sign, and all are statistically significant at the 1% level, except Retail Jobs which is significant at the 5% level in Model 1 and Bachelor's Degree, which is significant at the 10% level in Model 2. The values of adjusted R^2 for the models, 0.802 for Model 1 and 0.754 for Model 2, compare favorably with those of the single-city bike sharing ridership models, which range from 0.62 to 0.787 (*6*),(*7*),(*9*). For comparison, Model 3 includes the dummy variables for Denver and DC; because the Precipitation Days variable is a proxy for these dummies, it is correlated and therefore excluded from Model 3.

Although the Bikeways variable is significant in Model 1, it relies on the presence of another variable differentiating observations from the three systems, likely because the Nice Ride MN system has a relatively high coverage of bikeways, but relatively low per-station ridership (see Figure 2), while the other systems have a positive relationship between Bikeways and ridership. When system-specific "dummy" variables, such as DC Flag were included in preliminary model specifications, they showed statistical significance, suggesting that additional, city- or system-specific factors, such as operating season, membership and rental costs relative to incomes, and local bike culture, might also influence bike sharing ridership. These Flag variables are included in Model 3.

The Bikeways variable only becomes significant and positive when included with the Precipitation Days variable, which is specific to each system and is the same for all stations within a system. In this sense, the Precipitation Days variable acts as a rough ordinal variable only, since there are only three possible values across the entire dataset. This effect of the Precipitation Days variable in Model 1 is large relative to the other variables, but is offset by a large constant term. With the combined effects of the other variables, Model 1 does a better job of estimating station-level ridership within the three input systems; however, Model 1 is highly sensitive to the Precipitation Days variable, and will likely not yield reasonable results if applied to cities with Precipitation Days values different from those in the three input systems. Model 2 was developed to avoid this sensitivity to the Precipitation Days variable and other city-specific effects. It still has a relatively high R^2 and contains the other independent variables, but is more applicable to other contexts than Model 1.

The other independent variables are more robust than the Bikeways variable. Both Population and Retail Jobs are positively correlated with ridership – the more population or retail jobs are concentrated in the 400-meter buffer surrounding a station, the higher the ridership at that station tends to be. Total Jobs and High-Income Jobs were also tested in a variety of model specifications, but Retail Jobs was consistently more significantly related to ridership. The Retail Jobs variable could capture the effect both of high employment density (employees commuting by bike sharing) and of a high concentration of attractive retail destinations (shopping, entertainment, or leisure trips by bike sharing).

The Alternative Commuters, Median Income, and Graduate Degree variables were also positively correlated with ridership – the higher the proportion of commuters traveling by modes other than driving, or the higher the median income or education of populations surrounding a bike sharing station, the higher the ridership at that station tends to be. The proportion of Non-White Population, on the other hand, was negatively correlated with ridership. The relationships between ridership and income and race should not prevent the placement of stations in low-income communities or communities of color, however. Low income communities in particular may have difficulty acquiring debit or credit cards needed to access the system; these communities may warrant additional outreach, such as the Bank on DC program, which provides unbanked or underbanked bikesharing users with a discounted bikesharing membership and a debit or credit card (11).

Finally, the Stations Within 4800 Meters variable had a strong, positive correlation with ridership, significant at the 1% level in both preferred models. The variable was also significant in all other specifications tested, alone and controlling for other demographic and spatial variables, and even when Models 1 and 2 were tested on subsets of the regression data specific to each system. The robustness of the network effects variable was also tested at a variety of spatial scales by substituting other Stations Within [X] Meters variables into Models 1 and 2. Distances of 1200, 1600, 2400, 3200, 4000, 5600, and 6400 meters were tested; all were also significant at the 1% level. Of these network effects variables, the Stations Within 4800 Meters variable was selected for inclusion in the preferred regressions because it contributed to a high model R^2 and could be easily interpreted – 4800 meters is approximately 3 miles. The network effects variables at distances of 200, 400, 600, and 800 meters were not significant in all models, perhaps because these areas are too small to constitute an entire network or because stations located too close together served as substitutes more than complements to each other; there was less variation among stations in the network effects variables below 1200 meters.

Network Effects – Bivariate Regression

Bivariate regressions were performed to further test the robustness of the relationship between the number of bike sharing stations within 4800 meters of a given station and the natural log of the number of monthly rentals. Regressions were performed to test this relationship for the stations in each of the individual systems and for all stations as a group. In each regression, the two variables were positively correlated. The coefficient was also statistically significant at the 1% level in each case. Comparable regressions with Monthly Rentals as the independent variable instead of ln(Monthly Rentals) also found Stations Within 4800 Meters to be positively correlated with ridership and statistically significant at the 1% level.

TABLE 4 Bivariate Regressions of Stations Within 4800 Meters and ln(Monthly Rentals)

	All Systems		*Capital Bikeshare*		*Denver B-Cycle*		*Nice Ride MN*	
Variable	**Coefficient**	**p-Value**	**Coefficient**	**p-Value**	**Coefficient**	**p-Value**	**Coefficient**	**p-Value**
Stations Within 4800 Meters	0.041	0.000***	0.068	0.000***	0.027	0.001***	0.034	0.000***
Constant	3.439	0.000***	1.931	0.000***	4.834	0.000***	3.318	0.000***
Observations	264		98		52		115	
Adjusted R^2	0.411		0.657		0.194		0.442	

*** indicates significance at the 1% level.

CONCLUSIONS AND RECOMMENDATIONS

This study established statistically significant relationships between several independent variables and bike sharing ridership, based on the first-season experiences of three U.S. bike sharing systems. The study also developed a regression model that can be applied directly to other communities interested in pursuing bike sharing based on consistent and widely-available Census data.

The study results suggest that bike sharing station network effects are extremely important to ridership levels, with a robust, statistically significant relationship within systems, across systems, independent of other variables, and when controlling for other demographic and spatial variables. Population density, retail job density, median income levels, and the share of alternative commuters and non-white population are also critical factors in estimating bike sharing ridership. The presence of bicycle infrastructure, however, was less significant to bike sharing ridership across Capital Bikeshare, Denver B-Cycle, and Nice Ride MN than was suggested by Buck and Buehler's (*7*) analysis of the Capital Bikeshare system alone.

Bike sharing system planners should consider the importance of a comprehensive network of potential destinations within biking proximity of each other when determining the extent and spatial distribution of bike sharing stations. Operators of existing bike sharing systems might consider relocating underused, isolated stations to be closer to a central network of stations; however, to the extent that these peripheral stations address equity concerns by providing access to underserved communities, they might be better integrated by installing new stations that form a more continuous connection with the broader network.

Finally, practitioners should be aware that these results are based on early-adopting user populations and may change as bikesharing matures.

SUGGESTIONS FOR FURTHER RESEARCH

The experience of conducting this study suggests several directions for further research. First, as bike sharing ridership data become available for more U.S. systems, researchers should expand this analysis to include a more diverse range of systems in terms of size and context – research on small towns and suburban areas is notably lacking. Furthermore, as systems mature, ridership data can also be collected across a longer period. Using data beyond the first season of operation would help researchers to more accurately estimate the equilibrium level of ridership.

Second, the spatial scale at which the demographic and built environment variables are compiled could be refined. This study used 400-meter buffers for every variable except the network effects variables; however, different spatial scales could be more relevant for different variables. Researchers should test variables at other scales (e.g., Population within 3200 meters, or Retail Jobs within 1600 meters) to examine both local station characteristics and measures of the wider bike sharing environment.

Third, researchers could further refine the network effects variables. The current study uses a simple, linear buffer at a variety of distances to capture the number of bike sharing stations near a given station. Researchers could improve these variables by using network analysis to account for the actual street network that pedestrians and cyclists must travel to reach a bike sharing station. Bikeways and bike friendly streets could even be included in this variable as weights or impedances to the network travel. These improvements would more accurately reflect the accessibility of a given station to a surrounding pedestrian walkshed and to a network of other bike sharing stations.

Finally, ridership should be explored from a longitudinal or time series perspective. The use of a panel dataset, such as the one analyzed in the current project, precludes the exploration of several interesting variables that have temporal components. Observations could be broken down to the station-month or station-day level for longitudinal analysis. Researchers could then explore issues such as the time since system launch, which might explain upward trends in ridership as users adopt the program, or seasonality, including time of year, heating degree or cooling degree days, and precipitation. These improvements would certainly increase the data-intensiveness of an already data-hungry process, but would present opportunities to study additional ridership determinants, produce more accurate ridership estimates, and potentially explore relationships that affect not only bike sharing ridership, but bicycling in general. Longitudinal analysis would also allow researchers to explore the effects of expansions in the system, such as the installation of new stations, the expansion or relocation of existing stations, or the addition or removal of bikes within the system.

ACKNOWLEDGMENTS

The author would like to express his gratitude to the operators of Capital Bikeshare, Denver B-Cycle, and Nice Ride MN for making their ridership data available, and to Fehr & Peers for the opportunity to explore this interesting topic. Thanks also to Amy Smith for creating the custom GIS tool that made compiling the regression data possible. Lastly, many thanks to Lindsey Hilde, Gustavo Jimenez, Ian Moore, Leah Ramos, and Leon Raykin for their invaluable contributions: stimulating discussions of bike sharing and statistics, thoughtful edits to this paper, and general encouragement throughout the research process.

REFERENCES

1. Shaheen, S.A., S. Guzman, and H. Zhang. Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2143, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 159-167.
2. Shaheen, S.A., E.W. Martin, A.P. Cohen, and R.S. Finson. *Public Bikesharing in North America: Early Operator and User Understanding*. MTI Report 11-26. Mineta Transportation Institute, 2012.
3. Krykewycz, G.R., Puchalsky, C.M., Rocks, J., Bonnette, B., and F. Jaskiewicz. Defining a Primary Market Area and Estimating Demand for a Large-Scale Bicycle Sharing Program in Philadelphia. In *Transportation Research Record: Journal of the Transportation Research* Board, No. 2143, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 117-124.
4. Gregerson, J., M. Hepp-Buchanan, D. Rowe, J. Vander Sluis, E. Wygonik, M. Xenakis, and E. McCormack. Seattle Bicycle Share Feasibility Study. University of Washington College of Built Environment Department of Urban Design and Planning, 2011.
5. Olson, J., P. Goff, and S. Simms. City of Providence Bike Share Feasibility Study Final Report. Prepared by Alta Planning + Design for City of Providence. May 2011.
6. Maurer, L.K. Feasibility Study for a Bicycle Sharing Program in Sacramento, California. Presented at the 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
7. Buck, D. and R. Buehler. Bike Lanes and Other Determinants of Capital Bikeshare Trips. Presented at the 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
8. Hampshire, R.C. and L. Marla. An Analysis of Bike Sharing Usage: Explaining Trip Generation and Attraction from Observed Demand. Presented at the 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
9. Daddio, D.W. Maximizing Bicycle Sharing: An Empirical Analysis of Capital Bikeshare Usage. University of North Carolina at Chapel Hill, 2012.
10. Osborne, J. Notes on the Use of Data Transformations. Practical Assessment, Research & Evaluation, 8(6), 2002. http://PAREonline.net/getvn.asp?v=8&n=6. Accessed Apr. 14, 2008.
11. Capital Bikeshare. Capital Bikeshare Launches Bank On DC Partnership Program. December 16, 2011. http://www.capitalbikeshare.com/news/2011/12/16/1140. Accessed Nov. 13, 2012.