Spatial Analysis of Factors Predicting Bicycle Ridership

Andrew Harvey
Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, MN 55987

Keywords: Bicycling Economics, Urban Economics, Transportation Efficiency, Community Health

Abstract

As cities become more congested and city budgets continue to be strained, both commuters and cities are increasingly looking to bicycles as a viable mode of transportation. At present, bicycle commuting is not very well studied. This study is dedicated to one method that could help target investment helping to increase bicycling as a viable form of transportation in the eyes of the general public. Ordinary Least Squares analysis was applied to variables identified in literature as important factors that potentially correlate with a higher percentage of bicycle commuting ridership. The study area was the extent of Nice Ride Stations located within the City of Minneapolis, Minnesota.

Introduction

Significance of Research

A majority of American transportation infrastructure has been built for the automobile (Bureau of Transportation Statistics [BTS], 2016). As of 2013, there were 7,731 miles of commuter rail, 1,622 miles of heavy rail (subway like transportation), and 1,836 miles of light rail (BTS, 2016). At the same time, there were 8,656,070 lane miles of public roads (BTS, 2016). Over the course of the next several years a significant investment in the upkeep of this aging infrastructure will be necessary (BTS, 2016). There were roughly 50,000 structurally deficient bridges as of 2014 (BTS, 2016).

2,677,771,000,000 miles were travelled in 2013 just in passenger cars alone (National Highway Traffic Safety Administration (NHTSA), 2017). However, there has been a significant reduction in the increase of miles driven since the recession of 2008 (NHTSA, 2017). Despite the comparatively low number of transit miles and comparatively low level of investment in transit infrastructure, transit ridership numbers are on the rise (BTS, 2016). Ridership numbers have increased from below 8 billion passenger trips per year to over 12 billion in 2013 (BTS, 2016). Yet in 2013, vehicles driven alone remained the lion's share of the daily commute at 76.5%, and at a distant second was carpooling at 9.2% (BTS, 2016).

Given that, decisions may be made to decide whether to reinvest in this crumbling road infrastructure (BTS, 2016) with the question to be asked is: are there modes of transportation with a better return on investment? In multiple studies, bicycling is demonstrated to be of a net benefit per mile ridden, while automobile use has a net cost to society (Blue, 2016). So why would people still drive? Because it is the easy choice; rather than expending the mental and physical energy it takes to ride, just turn the key and the oil does the work for you. However, are there

Harvey, Andrew. 2017 Spatial Analysis of Factors Predicting Bicycle Ridership. Volume 20, Papers in Resource Analysis. 12 pp. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. Retrieved (date) from http://www.gis.smumn.edu

impediments that could be removed or incentives that could be offered that would increase the share of commuters taking to the roads on bicycles, thereby reducing the burden on society of the costliest transportation option?

Delimitations of the Problem

Delimitations consist of testing variables that literature suggested are linked to bicycle commuting ridership via Ordinary Least Squares Regression analysis, also known as OLS. Literature suggested bicycle commuting data is hard to come by. Data was mined from the Nice Ride bikeshare system's ridership data from 2016 and tested against various independent variables. The goal was to find several factors that are correlated with ridership and find the most cost-effective infrastructure to invest in. At the same time, variables were added to address the concern of confounding variables. These variables were added to help reduce the risk of correlation based on multicollinearity with some other factor.

Data Collection Techniques

Funding for bicycle commuter studies has traditionally been out of step with the funding provided for similar studies of automobile traffic. That is to say traditionally the funding is just not there. Financial outlays for bicycle infrastructure have been low relative to other forms of transportation infrastructure. Traditional bicycle infrastructure such as bike lanes, or sharrows, cost so little that the problem is less about funding than it is about taking space from other uses, usually parking (Blue, 2016). The more costly pieces of bicycle infrastructure, such as off-street bicycle trails, many of which are placed in former railroad right of ways (e.g., Midtown Greenway), were built more as recreational amenities. Those trails though, such as the Midtown Greenway and Cedar Lake Trails in Minneapolis, are purpose-built commuter trails, although they do also function for recreation. In fact, it is faster to get across South Minneapolis on the Midtown Greenway on a bicycle than it is to traverse the main automobile thoroughfare, Lake Street, by automobile; not just by a little bit, but by nearly half (Midtown Greenway Coalition 2017). The Cedar Lake Trail, shown in Figure 1 was built in 1995 and has three lanes to help improve traffic flow. One lane is for pedestrians, and the other two are for bicycles, one lane for each direction.

Figure 1. Cedar Lake Trail (Flynn, 2017).

The most common ridership data collection method is the one the City of Minneapolis uses. The city selects a day in the fall – usually in September – and posts trained "counters" at random places which are rotated each count year as well as pre-selected places that are counted year after year (Minge, Falero, Lindsey, Petesch, and Vovick, 2017). A sample of the City of Minneapolis ridership counts can be seen in Figure 2.

Figure 2. Sample bicycle counts (red dots) represents points collected every year. Yellow dots represent areas collected every three to four years (City of Minneapolis Department of Public Works, 2017).

New methods are emerging that could increase the effectiveness of analysis on bicycle commuting data. There is an increasing amount of road sensors that can detect the particular characteristics of bicycles and count them as they ride by. These are similar counts as would be obtained by the typical rubber hose counters, but will not be confused by cars that may also be passing over the sensors. There is a particularly interesting method in Minneapolis called Zap. The product uses an RFID sensor the user attaches to your bicycle. When a bicycle passes by Zap sensors, located throughout the city, the sensor will count the bike as having ridden past the checkpoint. Dero, the company producing Zap, partnered with other businesses to provide the additional incentive of drawings and business discounts throughout the city to those users who have passed the sensors most often.

The most complete and accessible open source dataset is the data available from Nice Ride Minnesota and other bike shares throughout the country and the world. Nice Ride data was chosen to answer the question: What are the factors that have the highest potential to increase commuting ridership in the City of Minneapolis?

Nice Ride Data was parsed into several different subsets, which will be covered in further detail. The main independent variables tested were educational attainment, canopy cover, educational enrollment, distance to the next nearest station, and distance to the nearest bicycle infrastructure of different types.

Methods

Nice Ride makes their ridership data freely available. The data contains start station, end station, member status, and time stamps for both the start and end of the trip in tabular format. An example of a few Nice Ride raw records can be seen in Figure 3. In addition, Nice Ride provides the geographic locations of their stations in tabular format with XY coordinates.

Start date	Start station	Start stati
11/6/2016 22:09	Willey Hall	30195
11/6/2016 21:53	Sanford Hall	30182
11/6/2016 21:52	11th Ave S & S 2nd Street	30154
11/6/2016 21:52	11th Ave S & S 2nd Street	30154

Figure 3. Example of raw Nice Ride records (dependent variable source).

The 2016 dataset from Nice Ride contained over 1 million rider trips. These data formed the backbone of the dependent variables. These individual trips can be consolidated into ridership totals at each station and then subtotals can be extracted based upon numerous factors, such as time of day, member vs casual ridership, etc. Each subset represents potential differences in ridership. For example, the people riding Nice Rides at 9AM may have different habits and routines than those who are riding at 1:45PM. There is likely more of a chance that the 9AM rider is commuting to work, while perhaps there is a higher chance that the afternoon rider may be riding to class, or home from school. Extracting subsets of the data allows the potential for the statistical analysis of those factors.

Following are the subsets of rider trips extracted from the ridership dataset. Each dependent variable dataset was placed into a single Esri geodatabase awaiting conflation of data from the independent variables listed later in the study.

Dependent Variables

Time of Day

Time of day could be a significant contributor to different patterns of bicycle traffic. Traffic patterns in automobiles can be very different throughout the day and bikes likely follow a similar rush hour pattern. Although all times of day were analyzed for statistical correlations, the morning and evening rush hours were the most important to this study, due to the target of the study being commuter trips. Using the timestamp provided by Nice Ride data, four categories of time of day were created using Python code. The four time periods were as follows: Morning rush (7am to 9am), Midday (9am to 4pm), Evening rush (4pm to 6pm), and all other times (6pm until the start of the next morning rush at 7am).

Member or Casual Rider

In addition to time of day, two additional dependent variable datasets were created from the Nice Ride data. There are two categories listed for user types in the Nice Ride data: member and casual. The "member" type in the Nice Ride data refers to users who have a monthly or yearly pass that allows them quick access to bicycles using a special code or key. These memberships usually indicate that these are frequent riders and the hypothesis would be that they would have more consistent patterns in their riding. The "casual" rider type purchases a rental on an hourly or daily basis. They typically do not ride as frequently, and judging by the locations and timestamps of these users, more of these trips appear to be recreational trips (e.g., a trip that both begins and ends at the Lake Calhoun station from 7 to 8 pm). Given the frequent use of bicycles by members, as compared to the casual user, this is more likely a dataset with a higher subset of commuters, who have more predictable patterns than the casual rider. Consequently, the hypothesis was made that the member attributed riders were more likely to produce reliable statistical analysis than the casual rider type. However, both types were analyzed separately using the OLS methods.

All Nice Ride Start and End Stations

In addition to these subsets described, a full dataset was created for use as a dependent variable, which contained all times and all user types. This produced two datasets, one for all start stations and one for all end stations, which were analyzed separately.

Independent Variables

Educational Attainment

Heesch, Sahlqvist, and Garrard (2012) suggested that a higher level of education attainment may correlate to higher ridership. Educational attainment is available from the 2015 American Community Survey reported by the number of people with a given attribute for each block group.

Educational Enrollment

Educational enrollment is also available from the 2015 American Community Survey, which is reported by the number of people with a given attribute for each block group (United States Census Bureau, 2016). The educational categories within the census data that were used for this study were High School Graduate, Some College, Bachelor's Degree, Master's Degree, and Doctoral Degree. For educational enrollment and education attainment, the Bachelor's Degree, Master's Degree, and Doctoral Degree numbers were summed to arrive at a single number for analysis purposes. Literature suggests there is a correlation between educational attainment and a higher level of bicycle ridership. Also, there tends to be a lower incidence of car ownership among the age cohort 18-22, typically enrolled in college (Heesch, Giles-Corti, and Turrel, 2015). These data were gathered at the block group level.

Distance to Nearest Station

One factor that was noticed as research progressed is that there are stations located very close together in areas with high ridership. For example, there are many stations clustered at the University of Minnesota, which is by far the most densely packed area of Nice Ride stations in the city.

Having stations with very high ridership next to stations with a very low ridership was a factor to be accounted for and explored in the study. Figure 4 illustrates an area near the University of Minnesota where there are many stations located very close together. Although located close together, not every one of these stations has an equivalent level of ridership; there is significant variation from station to station. Using the distance to the next station as a variable helps to account for the fact that this is a cluster of stations. Using this variable, the OLS accounts for a region with high ridership but also a high density of stations to choose from. As a result, distance to the next closest station was a variable introduced by running the script referred to later in the Methods section . The script used the Near tool to find the location of the next nearest station.

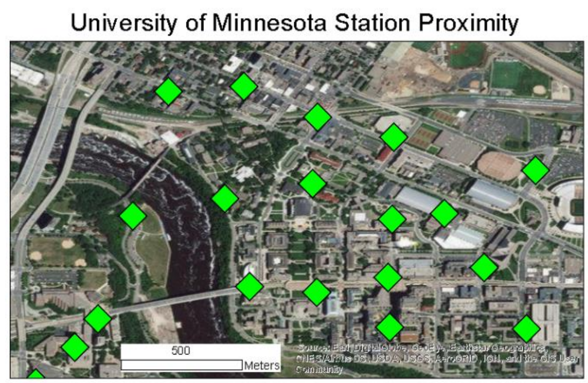

Figure 4. Density of Nice Ride Stations in green in the University of Minnesota area (DigitalGlobe, 2016 and Nice Ride Minnesota, 2016).

Income

Literature suggested in some cases income levels of a neighborhood may be a factor in ridership levels (Heesch *et al.*, 2012). Heesch *et al.* (2012) suggested there may be a positive correlation between income and ridership levels. However, in a later study, Heesch *et al.* (2015) suggested studies found negative correlations in other places, meaning that people with lower incomes might have higher ridership levels. Income data used for this study was gathered from the census block data in the American Community Survey of 2015. The average income for each block group polygon was calculated by joining the ACS data to the appropriate polygons.

That value was then applied to each Nice Ride Station using the OLS Variable Calculator mentioned later in this paper.

Distance to Major Activity Centers

Literature has also cited downtowns and other major activity centers as significant factors that influence biking (Heesch *et al.*, 2015). Specifically, an inverse correlation between distance from the downtown core and ridership has been shown; thusly, the closer to downtown and other major centers, the more likely you are to choose a bicycle as your vehicle. Due to the very high concentration of destinations at the University of Minnesota, it was included as an additional major center. As a result, Minneapolis had two major activity centers identified in the study area. The first was a point placed at the heart of downtown Minneapolis at 7th and Nicollet, and the second point was placed at the heart of the University of Minnesota campus in front of Coffman Memorial Union. Distance from the closest of these points was calculated using the OLS Variable Calculator.

Bikeways

Bikeway data was provided by the Minnesota Department of Transportation (MNDOT); however, attribution was partially modified for use in the study. Bikeways exist in many shapes and sizes in the United States as there is not yet a uniform code for constructing them as there is with other forms of transportation, like roads and rail networks. Therefore, the quality of bikeways and the experience of safety a rider feels while riding on infrastructure varies widely. Some bike lanes are no more than 4 inch lines of paint on very busy roads with no protection from cars; others are robust "bicycle highways" like the Cedar Lake Trail or the Midtown Greenway. Using the type field in the MNDOT data, the Bikeways feature class was queried to construct several sub datasets. The first was only those of the 'Paved Trail' type in the dataset; these are any off-street paths. The second was a subset of the 'Paved Trail' types consisting solely of the several "bicycle freeways" located around the City of Minneapolis (shown in Figure 5 as Major Bikeways). The Midtown Greenway and the Cedar Lake Trail are two jewels of the Minneapolis Bicycle commuting system; however, there are a few others as well which were included in this dataset.

Figure 5. Major Bikeways in the study area are shown in green. Esri Basemap.

The third subset of bikeways was inclusive of both the 'Bike Lane' type and the 'Paved Trail' type. These were all divided into different feature classes for use in the OLS Variable Calculator.

Canopy Cover

Heesch *et al.* (2015) suggests areas with low levels of canopy cover would have a reduction in the amount of recreation biking done in a survey conducted among a random sampling of people in Brisbane, Australia. In their survey, canopy cover was a self-reported variable.

A more rigorous GIS approach to canopy cover was taken as a part of this research. The University of Minnesota Department of Forestry conducted a survey using LIDAR and satellite imagery to derive a high-resolution land classification raster of the entirety of the City of Minneapolis (Bauer, Kilberg, Martin, and Tagar, 2011). These data (Figure 6) needed to be extracted to help index the “feel” of a street.

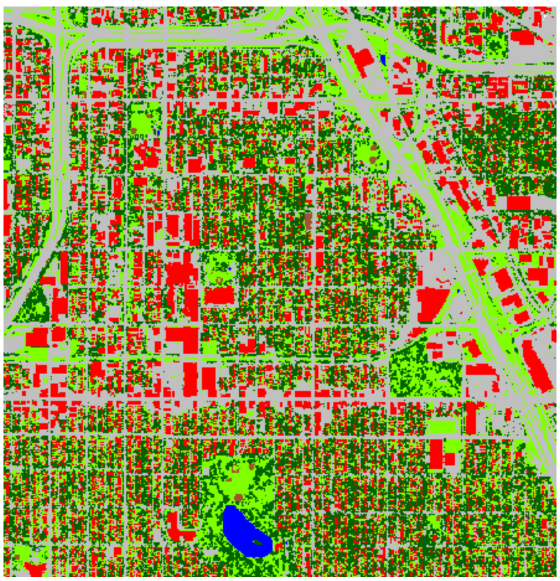

Figure 6. Raw land cover data with canopy cover in green (Bauer *et al.,* 2011).

Literature suggests a low amount of tree cover would have a negative impact on the ridership of an area. The raster of the city was reclassified into a binary raster: 1 being trees and 0 being no trees. From there the Hennepin County street centerline files were buffered by 10 m on each side to approximate an average size right of way of the road and sidewalks of an average city street. These buffered roads were then processed using Esri’s Zonal Statistics to calculate the number of pixels within each zone/street segment that were covered with tree canopy. Using Near analysis, each Nice Ride station was given the percentage of tree cover from the station’s nearest street segment. The result of the above analysis can be seen in part in Figure 7.

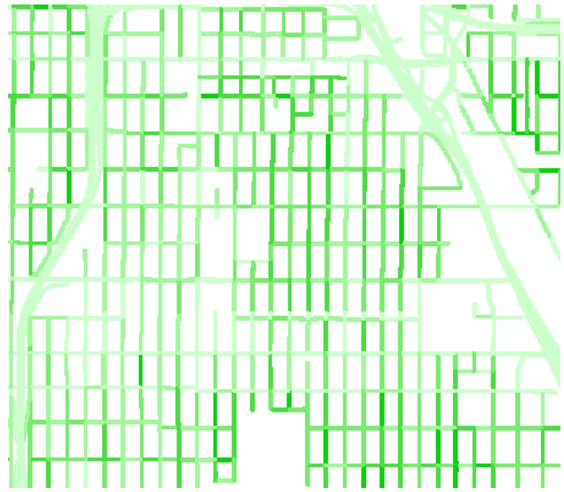

Figure 1. Canopy Cover Index.

Formatting Raw Nice Ride Data

The first step in building Nice Ride data into GIS ready data was to format the Nice Ride data correctly. The Nice Ride data from 2016 was available in two comma separated files. The first file was the latitude and longitude of every Nice Ride station in the system. The second comma separated file contains the start station, end station, start time, end time, user type, and length of trip in minutes for every single trip taken during the calendar year of 2016. The first step was to use the Display XY Data tool available in ArcMap to obtain a point feature class that would represent the locations of each station, first as an event layer, and then exported to a point feature class in the geodatabase. The data needed to be simplified so that statistics could be run on an amalgamation of station trips rather than single point to point trips. First the trips were joined to the station. Then, the Collect Events tool in the Spatial Statistics suite was used to count each event and add that event to the station point. This was conducted for the full start/stop dataset, the time of day dependent variables, and the member type dependent variable. Each dependent variable was output to its own shapefile for use in statistical analysis. Each of these tables was designed so that the OLS Variable Calculator could be run to “join” the independent variables that were later processed.

Splitting by Time of Data

The hypothesis posits that there may be different types of riders riding at different times of day. Using Excel the data was subdivided by time of day: the morning rush (7am to 9am), midday (9am to 4:30pm), the evening rush (4:30pm to 6:30pm), and all other times of day; each subset was considered a separate dependent variable. The Excel sheet had a column added that would calculate a simple hour of the day that the trip started. The data were then joined to the stations, and definition queries were used to run the Collect Events tool on each subset of trips.

Getting Data into One Place

The most important piece of the project was the script devised to facilitate future regression analysis projects. All Ordinary Least Squares regression analyses must be performed by compiling all data into the same table. One column functions as the dependent variable, and then all the independent variable values that are analyzed with it are compiled into the same table. However, these values must correspond in some way to the location or value of the dependent variable to be of any use. The script allows this process to be done in a straightforward manner using a properly formatted database. The file structure pictured below in Figure 8 details how the data must be compiled.

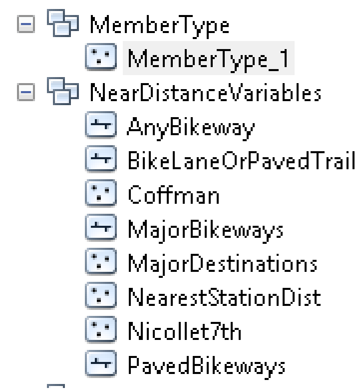

Figure 2. File Structure of OLS variable calculator.

The script uses one of two methods depending on where the independent variable is stored. The first method is for those values where the distance to the nearest feature is the value that is desired to be the independent variable. This includes, in this study, distance to bikeways, distance to bike lanes and bikeways, distance to any bike infrastructure, distance to major destinations, and distance to the nearest Nice Ride station. A list of independent variables that were used in the statistical analysis can be found in Table 1.

The second method is used when a corresponding value from the nearest feature is what is desired. For example, if the speed limit of the nearest road is desired, the tool first finds the nearest feature. The nearest feature ID is assigned, and then the script searches through the independent variable dataset for the column title “Iterate.” In this example, the iterate column is used to hold the speed limit. The script uses the “Iterate” column to hold the value that is the target value to be appended to the dependent variable feature class. The “Iterate” column in the target data set is then named the feature class name from the source feature class. Documentation is required to use the tool and was located on GitHub and freely available for academic use.

Table 1. List of independent variables in the statistical analysis.

Independent Variables
Distance to Major Destinations
Distance to University of Minnesota
Educational Enrollment
Educational Attainment
Distance to Downtown Minneapolis
Distance to Any Bikeway
Distance to Bike Lane or Paved Trail
Distance Major Bikeways
Income
Distance to Paved Bikeway
Speed Limit
Percentage Canopy Cover
Population Density

OLS/Exploratory Regression - How it Works

OLS analysis is an important method to determine predictive formulas for what potential factors contribute most to the future growth of bicycling. By utilizing existing values, for example income vs number of commuters, OLS will create a linear regression formula that will then be evaluated against itself to determine how much error existed between its predicted values and the observed dependent variable values (Zar, 2010). Once the regression, the line of best fit, is built, OLS calculates predicted values using the regression formula, and then OLS evaluates the delta, known in this type of analysis as the residual, which is the difference between the predicted value and the actual value. For each of the eight dependent variables (all trips start, all trips end, four times of day, member, and casual), a regression formula was created and then the regression equation was evaluated. The two most important statistical tests were the adjusted R-squared test and the Moran's I p-value test.

Adjusted R-Squared Test

Adjusted R-squared is a value ranging from 0 to 1. A value of 0 indicates perfect randomness, and 1 indicates a perfect correlation between two variables. In Figure 9 the R-squared value using the linear regression tools in Microsoft Office Excel 2016 is 0.2493. Adjusted R-squared differs from a standard R-squared value in that the complexity of the model is taken into account. That is, a model with more variables involved will likely contain more error so the R-squared value is adjusted to a lower value to compensate (Esri, 2015a).

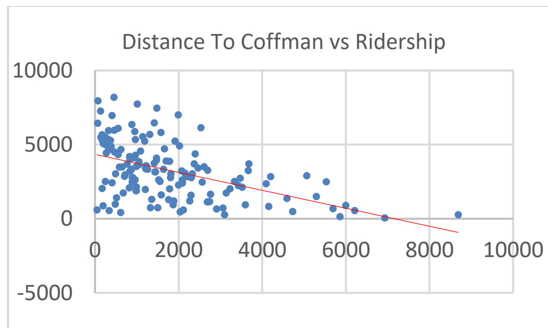

Figure 3. Difference between observed values (in blue) and the red regression line.

Global Moran's I p-Value

The Global Moran's I p-value is a test for spatial autocorrelation. Spatial autocorrelation is a measurement of how clustered data values are in a geographic area (Chun and Griffith, 2013). Essentially some correlations can exist due in a larger part to proximity to another nearby place. Knowing if a relationship is spatially autocorrelated can potentially suggest there are factors in the model that are not yet being accounted for.

Results

Each of the eight dependent variable datasets, including all independent variables, were first analyzed using the Exploratory Regression tool in the Spatial

Statistics toolbox of the ArcGIS Suite. It was deemed that the most consistent dataset, yielding the highest R-squared values was the subset of user type "member." It is useful to note that with the low R-squared values found in the other subsets, it is a finding of this study that the user type "member" was found to be statistically different than the member type "casual." The rest of the findings are based on regression analysis using the ridership of the "member" user type as the dependent variable.

Exploratory regression found there were several factors that seem to be correlated to a higher ridership in the member category of the Nice Ride dataset. However, constructing more complex models and adding additional variables to the regression algorithm mostly failed to produce higher R-squared values.

Each of the independent variables were also analyzed for their individual correlation to member ridership using exploratory regression. Table 2 reports each individual variable and its contribution to explaining member ridership. Using exploratory regression, it was found that the one factor that was most highly correlated with ridership was the distance to major destinations. The second highest correlated factor was just the distance to the University of Minnesota, and the third was the level of educational enrollment.

There are a few problems with these three factors being used in OLS analysis. The goal of the OLS is to build a model that achieves the greatest correlation factor with the least number of variables possible. One of the ways that OLS prevents overly complicated models is by testing for multicollinearity. In the case of these three factors, there is significant overlap, and thus a high level of multicollinearity. The variance inflation factor (VIF) indicates this; the Major Destinations variable had a summary VIF value of 10.25 and violated the VIF threshold of 7.5 in 1031 model iterations. According to the Esri Spatial Statistics guide, a VIF higher than 7.5 should be removed from a regression formula (Esri, 2015b). The model with the highest correlation can be found in Table 3.

Table 2. R-squared values * 100 showing significance of individual variables analyzed in isolation in comparison to the number of member riders.

Variable	*% Significance*
Major Destinations	33.32
University of Minnesota	28.43
Educational Enrollment	24.15
Downtown Minneapolis	7.34
Any Bikeway	6.25
Bike Lane or Paved Trail	6.25
Major Bikeways	2.46
Income	1.74
Educational Attainment	1.43
Paved Bikeways	0.63
Speed Limit	0.07
Canopy Cover	0.03
Population Density	0.00

Table 3. The model of best fit using the model with the highest R-Squared value.

Model of Best Fit Equation	
y = 2124.342+.0177U-.3764M+1571.94X	
R-Squared Value	0.320109
Adjusted R-Squared	0.305745
U= Distance From University of Minnesota	
M = Distance from Major Destinations	
X=Educational Enrollment %	

There are several reasons why these variables were not explanatory to the desired R squared value of at least 50%. There were additional tests unique to spatial statistics that failed, one being the Moran's I spatial auto correlation test. The test found that these data would not pass a spatial autocorrelation test. There is one interesting factor here; that is, the spatial autocorrelation showed a significant autocorrelation of ridership at the University of Minnesota.

Discussion

Polyline Vector Data Availability

In future studies there would be a strong benefit to the use of vector polyline routes that could be obtained from a dataset like Strava or MapMyRide that would provide a more robust dataset than that of the Nice Ride dataset. In addition, there are often user comments in these data that might be parsed for indicating whether or not these rides were commutes. Vector polylines would give a better analysis of the routes that are being used most once the riders leave the station. Each segment of road or trail could be analyzed with methods similar to those applied to the station points. For example, there are over a million rides that took place last year. Although many of those rides may have taken similar routes, there are routes that may have hardly been used. With the Nice Ride data, there is really no way to know where folks rode once they left the station.

Conclusion

This study shows that the quality and type of data being collected for bicycle ridership has, at present, limit use for statistical analysis purposes. With additional investment, and collaboration with partners like Strava and MapMyRide, more detailed datasets can be made available, which can help lead to better decision making. With more data will come more statistically confident measures of success and failure. It does seem that there is a link between education and the level of ridership, but that question will have to be settled through future studies.

Acknowledgements

I would like to thank most of all my wife for allowing me the time and resources to conduct this study. I would like to thank my advisor Greta Poser without whom many of the most important pieces of knowledge required for this analysis would not have been possible. I would also like to thank Saint Mary’s University of Minnesota for being a place of true learning.

References

Bauer, M., Kilberg, D., Martin, M., and Tagar, Z. 2011. Digital Classification and Mapping of Urban Tree Cover: City of Minneapolis. University of Minnesota Department of Forest Resources. Retrieved from https://rs.umn.edu/ utcmplsfinalclassification2009metadata.

Blue, E. 2016. Bikenomics. Portland, OR: Microcosm Publishing.

Bureau of Transportation Statistics (BTS). 2016. Pocket Guide to Transportation. United States Department of Transportation. Retrieved from https://www.rita.dot. gov/bts/sites/rita.dot.gov.bts/files/Pocket %20Guide%202016.pdf.

Chun, Y., and Griffith, D. 2013. Spatial Statistics and Geostatistics. New York: Sage Publications.

City of Minneapolis Department of Public Works. 2017. Bicyclist and Pedestrian Count Report 2016. Retrieved from http://www.minneapolismn.gov/www/groups/public/@publicworks/documents/webcontent/wcmsp-197773.pdf

DigitalGlobe. 2016. WorldView 2 Satellite Imagery. Esri Basemap. Retrieved from https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9

Esri. 2015a. Regression Analysis Basics. Esri. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/regression-analysis-basics.htm.

Esri. 2015b. Interpreting OLS Results. Esri. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/ordinary-least-squares.htm.

Flynn, S. 2017. Cedar Lake Regional Trail. Retrieved from http://metrobiketrails.weebly.com/hennepin-county.html.

Heesch, K., Giles-Corti, B., and Turrel, G. 2015. Cycling for Transport and Recreation: Associations with the Socio-Economic, Natural and Built Environment. *Health and Place, 36*, 152-161.

Heesch, K., Sahlqvist, S., Garrard, J., 2012. Gender Differences in Recreational and Transport Cycling: A Cross Sectional Mixed-Methods Comparison of Cycling Patterns, Motivators, and Constraints. *International Journal of Behavioral Nutrition and Physical Activity*, *9*, 106.

Midtown Greenway Coalition. 2017. About the Greenway. Retrieved from http://midtowngreenway.org/about-the-greenway/.

Minge, E., Falero, C., Lindsey, G., Petesch, M., Vovick, T. 2017. Bicycle and Pedestrian Data Collection Manual. Minnesota Department of Transportation. http://midtowngreenway.org/about-the-greenway/.

National Highway Traffic Safety Administration (NHTSA). 2017. Traffic Safety Facts. United States Department of Transportation. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812376.

Nice Ride Minnesota. 2016. Nice_Ride_trip_history_2016_season.csv. Retrieved from https://www.niceridemn.org/data/.

United States Census Bureau. 2016. American Community Survey, 2015 American Community Survey 5-year Estimates, Educational Attainment.

Zar, J. 2010. *Biostatistical analysis*. (5th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.