IMPACTS OF CYCLING INFRASTRUCTURE ON BIKE-SHARING SYSTEMS

CEE 4620 Final Project, May 2018

Abstract

Bike-sharing systems promise physical activity and cost savings for the urban public, as well as congestion relief and pollution reduction for cities. However, previous studies have shown that simply installing a bike-sharing system does not guarantee these benefits. This study examines the effect of adding cycling infrastructure (e.g. bike lanes) to the streets between pairs of bike-sharing stations on the number of trips taken between those stations. Using multiple linear regression, this study found that completely covering the path between a pair of bike-sharing stations with cycling infrastructure corresponds with an increase in the number of monthly trips between that pair of stations by 3.2 (95% CI: 0.2, 6.1). This result corroborates the results of previous studies, which have also found weak positive correlations between additional cycling infrastructure and additional bike-sharing trips.

Fletcher Passow

fhp7@cornell.edu

Introduction

The recent rapid adoption of bike-sharing systems in many cities throughout the world is a reaction to the immense promise that bike-sharing shows to achieve many highly-desirable social outcomes. Travel by bicycle, at least in theory, provides “flexible mobility” and physical activity for urban-dwellers, produces lower emissions than travel by bus or car, reduces congestion, and saves individuals money. However, little research has been done to show that cities actually experience these benefits after bike-sharing systems are installed.[1] In fact, de Chardon et al. revealed in 2017 that out of 75 bike-sharing systems examined worldwide, one-third have less than one trip per bike per day on average.[2] Because of this, it seems evident that simply installing a bike-sharing system is not enough to experience the touted benefits.

To be able to understand what makes a bike-sharing system work, a good first step is to examine what makes a bike-sharing *station* attractive to users. Several studies have been carried out in this area examining factors such as weather, the station’s elevation, the number of nearby stations, the capacity of nearby stations, the density of population and jobs near the station, the station’s proximity to key types of businesses, and the proximity of the station to cycling infrastructure.[3], [4] One of these studies, by Buck and Buehler, calls for more case studies to more carefully examine the impacts that cycling infrastructure has on bike-sharing systems.[3] This call is echoed in at least two other recent papers.[2], [5]

Cycling infrastructure (hereafter “CI”) is important because it increases the real and perceived safety of cyclists on roadways. Fishman et al. report that “Safety concerns are a major barrier to bicycling in Australia, the UK, and North America, and these concerns appear to hold true for bike share participation.”[1] CI refers to a whole range of installations which increase cyclists’ safety. The most limited type of CI is a “sharrow”, a symbol painted on the road reminding motorists to share the road with cyclists. More protected forms of CI include “exclusive lanes” which are separated from motor-vehicle traffic by painted lines and “protected lanes” which are separated from cars and trucks by vertical barriers or parked cars. The most protected form of CI is, of course, a dedicated off-road path.[6] Whatever form it comes in, CI requires precious space on roadways and money to build and maintain, which can make it politically unpopular.[2] This makes an accurate understanding of the value of CI essential for the progress of cycling and bike-sharing as viable urban modes of transportation.

Previous work exploring the relationship between CI and bike-sharing system use has focused on the bike-sharing station as the unit of analysis.[3], [4] While this choice made sense for the previous studies because of the factors besides CI they were considering, studying CI in this way does not make very much sense. Users of bike-sharing systems are fundamentally using the system to make trips between pairs of stations, so the properties of a single station probably aren’t what the user is focusing on when they decide whether to take a bike-share or a private car to work, for example. The user will probably focus more on the route between two stations when they are making their mode choice, so exploring the impact of CI on bike-sharing station activity from the perspective of paths and networks may be complementary to previous work.

This study seeks to answer the question: “How does adding cycling infrastructure between two bike-sharing stations change the number of trips between those stations? To accomplish this, the study used multiple linear regression and a before-and-after style analysis on data from New York City's CitiBike program. Launched in May 2013, CitiBike has grown to encompass 750 stations and 12,000 bikes as of May 2018. In October 2017, the 50 millionth CitiBike trip was taken. The CitiBike system is operated by Motivate International, Inc. the company that also operates bike-sharing systems such as Boston's BlueBikes, Washington D.C.'s Capital Bikeshare, the California Bay Area's Ford GoBike, and others.[7]

Methods

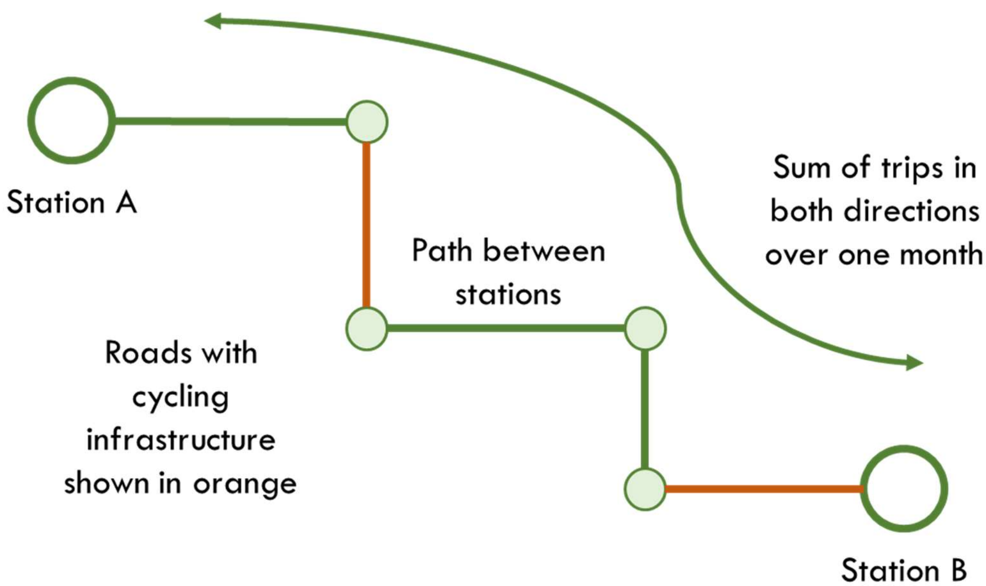
Unit of Analysis

The fundamental unit of analysis for this study is a pair of bike-sharing stations over the period of one month (Figure 1). This is an important departure from previous studies which used single stations for their units of analysis.[3], [4] For this study, each pair of stations is characterized by several factors, which can change depending on the month of interest.

First, each station-pair has a path over which cyclists travel to move from one station to the other. This study assumes that all cyclists traveling between a pair of stations will use the same path. Additionally, this study assumes that path will be the shortest path between the two stations on the road network. This path is also assumed to stay the same over time, not changing due to road closures or road improvements. The path is further characterized by the segments of the path which do or do not have CI installed along them. In the example given in Figure 1, cyclists travel along five different roads to move between stations A and B. Two of those roads, shown in orange on the diagram, have CI installed along their length. In contrast to the assumptions of static paths, this study does allow for the CI status of a road to change over time. The monthly change in the length of the path between a pair of stations which is covered by CI is the primary independent (predictor) variable for this study.

Second, each station-pair is characterized by a total number of trips in a month. This total includes the trips in both directions. For two hypothetical stations A and B, the total number of trips in a month is the sum of the trips from station A to station B and the trips from station B to station A. Directionality of the trips is not important because of the previous assumption that cyclists all take the same path between the pair of stations, regardless of which station is their origin and which station is their destination. Because of this, a change in CI along the route between the two stations will affect cyclists traveling in both directions equally. The monthly change of the number of trips between a pair of stations is the dependent variable for this study. Since each observation for this study contained data from a pair of station over a one-month period, the observations will hereafter be referred to as "station-pair-months".

Figure 1: Depiction of the unit of analysis, the station-pair-month



Data

Data Sources

To carry out this study, three main datasets were required. The first was a list of all the bike-sharing trips taken on CitiBike in New York City, including the stations between which the trips took place. The second was a list of all the cycling infrastructure in the city, including the time when it was installed. The third was a topologically correct representation of New York City's road network. While several cities other than New York, including Boston and Washington D.C., had the necessary data available, New York City's data had the most precise installation dates of CI.[8]–[11] This added precision was why New York City was selected as the location for this study's analysis.

The first dataset, a list of bike-sharing trips, was the raw data from which this study's dependent variable was derived. For New York City, this data is supplied to the public for free by Motivate International, Inc. (hereafter "Motivate"), the company that operates CitiBike.[9] This study used the data from July 2013, when records began, up until March 2018. Each record in the dataset represents a single trip and includes the following information relevant to this study:

- Start time and end time of the trip, reported with second-level precision
- Identification numbers of the start and end stations
- Geographic coordinates of the start and end stations

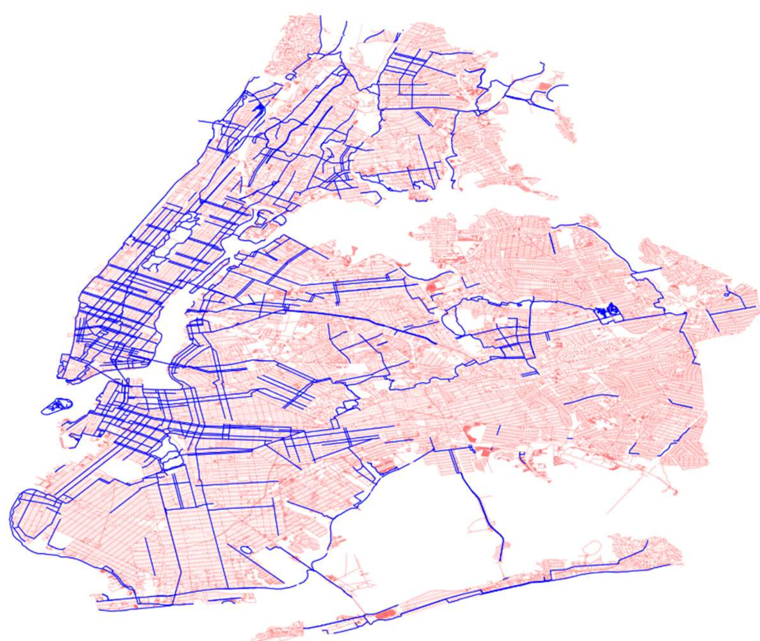
This dataset was probably highly reliable, because it was, and still is, automatically collected as an integral part of the day-to-day operations of the CitiBike system. If Motivate did not know which station their bikes were at or when the bikes were checked out and checked back in, they would not be able to collect the fees for their services or manage rebalancing. These two operations are essential for Motivate's business, so they are incentivized to measure and track this data well.

The second dataset, which described New York City's cycling infrastructure (Figure 2), was one part of the data used to derive this study's independent variable. The New York City Department of Transportation provides this data to the public for free through the New York City Open Data Portal.[8] Each record in the dataset represents a piece of CI, and the following data relevant to this study is included in each record:

- Geographic geometry of the infrastructure for mapping purposes
- Date the infrastructure was installed and the date the infrastructure was last modified
- Street the infrastructure is positioned on
- Infrastructure position, whether on-street or off-street
- Class of the cycling infrastructure (like the types of infrastructure described in the **Introduction**)

The primary weakness of this dataset, as far as this study was concerned, was that each bit of CI, usually in the form of a lane, was split up into many tiny geometries. For example, one "lane" that ran along a street for many blocks would be broken up into several separate observations in the dataset. Each observation would have the exact same attributes for lane class, on/off-street, etc., but would contain a different tiny geometry only one or two blocks long. The author overcame this by grouping all the rows with similar attributes (lane class, on/off-street, etc.) together and merging all their simple geometries into fewer, more complex geometries. This was achieved using the *GeoPandas* python package.[12]

Figure 2: New York City cycling infrastructure (in blue) and street network (in red)



The third dataset, a topologically correct road network for New York City, was another part of the data used to derive this study's independent variable (Figure 2). The data was acquired from OpenStreetMap (hereafter "OSM") using the *OSMnx* package.[13] This dataset encodes streets as network edges and intersections as network nodes. The attributes contained in this dataset include:

- the length of each edge
- the geographic coordinates of each node and edge
- other information about bridges, tunnels, street lights, etc.

This study acquired the network through *OSMnx* by using a location-based query to *OSMnx* for "New York, New York, United States". This query returned the road network for Manhattan, the Bronx, Brooklyn and Queens, but not for Staten Island. Because there are very few CitiBike stations and little CI on Staten Island, the borough was left out of the remainder of this study's analysis.

Preparing the dependent variable

This study used the CitiBike trip data to calculate the dependent variable: the difference in the number of trips between each station pair in each month. To accomplish this, the data were grouped by start station ID, end station ID, and the month of the trip start time. Then the number of records in each group was counted to find the total number of trips in that group. With this grouping and counting completed, the sums of the trips in both directions between each pair of stations in each month was taken. With the data for each station-pair-month in place, the differences between months could be taken to yield the final dependent variable. The differences were computed using a forward finite-difference method. That means that the difference in trips recorded for July 2016, for example, was equal to the number of trips in August 2016 minus the number of trips in July 2016.

Preparing the independent variable

This study used all three datasets listed above to derive the primary independent variable: the monthly change in the fraction of the length of the path between two nodes that is covered by CI. To prepare this variable, paths between stations had to be generated along a topologically correct road network that was "bike-conscious". By "bike-conscious", the author means that the network has information about not just the car infrastructure on each street, but also about the CI.

The OSM network served as the topologically correct network, but in its original form it did not include any data about CI along New York City's streets. To make the OSM network "bike-conscious", CI and CitiBike station data had to be transferred onto the appropriate nodes and edges of the OSM network. To accomplish a "join" like this, the datasets to be joined must have data in common. One piece of data that all the necessary datasets had in common was geographic coordinates. Each CitiBike trip observation included the geographic coordinates of the start and end stations, each piece of CI had a list of the coordinates of each of its vertices, and each OSM node had geographic coordinates associated with it. This study, therefore, joined the "bike-conscious" data onto the OSM network based on geographic coordinates.

Unfortunately, the geographic coordinates of each CitiBike station, CI vertex, and OSM node did not always line up with one another, even if they were all meant to represent the same intersection. To overcome this, this study used the *Shapely* python package to perform snapping operations.[14] A snapping operation changes the coordinates of a geographic object to exactly match the coordinates of another, nearby geographic object. For this study, CitiBike stations were snapped to the nearest OSM node (that is, a street intersection), and vertices of the NYC CI elements were snapped to the OSM nodes they passed near.

While the CI vertices were in general very close to the OSM nodes they passed by, CitiBike stations could be quite far away from intersections. While simply snapping over long distances seemed like a good option at first, it was soon discovered that *Shapely* did not guarantee that CitiBike stations would be snapped to the nearest OSM node if multiple OSM nodes existed within the snapping radius. To overcome this difficulty, an iterative snapping process was used for snapping CitiBike stations. First, all stations within five meters of an OSM node were snapped to the nearest node. Then, all stations within ten meters of an OSM node which had not been previously snapped were snapped to the nearest OSM node. The snapping distance was increased in five-meter increments until it reached 100 meters, at which point almost all the CitiBike stations had been snapped to OSM nodes.

Once the CitiBike stations and CI elements were all concurrent on the OSM nodes, this fact was used to join the station and infrastructure element information onto the OSM network. For the stations this was performed easily using the spatial join function from the *GeoPandas* python package.[12] For the CI data, however, using the *GeoPandas* spatial joining function produced erroneous results. Performing a spatial join by the intersection method joins the data only if the geometries contained in the data intersect. Since the vertices of the CI data had previously been snapped onto the OSM nodes, all streets coming to an intersection would 'intersect' with any bike lane on any street that came to that intersection (Figure 3). This vastly overestimated the amount of CI, because the CI information was added not only to streets which had CI on them, but also to streets which crossed the CI perpendicularly.

Figure 3: Erroneous joining of bike lane data to OSM edges

To correct this error, the author wrote a function which tagged OSM network edges (representing a street) if **both** ends of the edge intersected nodes which had a CI vertex snapped on top of it. The tagging function worked by travelling along a CI geometry, vertex by vertex. If a vertex was concurrent with an OSM node, it checked if there was an OSM edge between the current OSM node and the last OSM node encountered. If such an edge existed, then that edge was tagged with the ID number of the CI geometry the function was traversing. After this tagging had been performed for all CI elements, the CI data was joined to the OSM network using the tags that had been added.

With a “bike-conscious” network prepared, the independent variable could be calculated. This was accomplished by first finding the path for each station-pair, then traversing that path in each month to find the fraction of its length covered by CI. The path taken by cyclists between two CitiBike stations was assumed to be the shortest path, and was calculated using Dijkstra’s algorithm as implemented in the *NetworkX* python package.[15] These shortest paths were assumed to stay the same for all months within the period of interest (see the **Unit of analysis** section).

Once the shortest paths had been calculated, the fraction of the length of each path covered by CI was found by traversing each path in each month. As the path was traversed, two counters were incremented. The first, a counter of the total length traversed, kept a running total of the length of all the edges traversed up to that point. The second, a counter of total CI length traversed, kept a running total of the length of all the edges traversed up to that point that had CI on them. Since New York City’s CI was installed in stages, the CI length counter only counted CI that had been installed before the month for which the traversal was taking place. Once each traversal was complete, the CI length counter was divided by the total length counter to yield the fraction of the length of the path covered by CI in that month.

To produce the final independent variable for each station-pair-month, the change in the fraction of the length of the path between a pair of stations covered by CI, the differences between this length fraction for adjacent months were taken. As with the dependent variable, the differences were computed using a forward finite-difference method.

Final Dataset

The final dataset included only a subset of all the station-pair-month observations. The subset was found by applying the following inclusion criteria to the list of all possible station-pair-month observations.

- Both members of the station-pair must have been successfully matched with an OSM node
- The stations in the station-pair must be different from one another
- The total number of trips between the stations for the lifetime of the station-pair must have exceeded 500
- An infrastructure change must have taken place between the stations in the month of interest

The first and second criteria were necessary to guarantee that the independent variable could be calculated for each station-pair-month. For station-pair-months without successful OSM node matching of stations, no path could be found at all. The stations without successful matching were located in areas where the OSM network was not known, such as islands in the Hudson River, in New Jersey, and on Staten Island. For station-pair-months with identical start and end stations, it would be impossible to calculate a fraction of path length covered by CI, because the total length of the path (the denominator of the fraction) would be zero.

The third criterion was applied to speed computation and identify stations-pairs with consistent trips. Of the 213,913 station-pairs with a non-zero number of trips over this study's period of interest, only 58 percent had more than 10 trips, 27 percent had more than 100 trips, 11 percent had more than 500 trips, and 6 percent had more than 1000 trips. Limiting the number of station-pairs considered sped up computations of station-pair paths and CI length fractions, while only sacrificing station-pairs with very few and probably intermittent trips between them.

The fourth criterion was applied to make the final regression model only cover station-pair-months which had a non-zero infrastructure change. Out of 1.11 million station-pair-months with non-zero trips between CitiBike stations which had passed the previous criteria, 8293 were found which had an infrastructure change. This criterion prevented the regression from being overwhelmed by data with zero-infrastructure change, allowing analysis to focus on the impact of CI additions when they did occur.

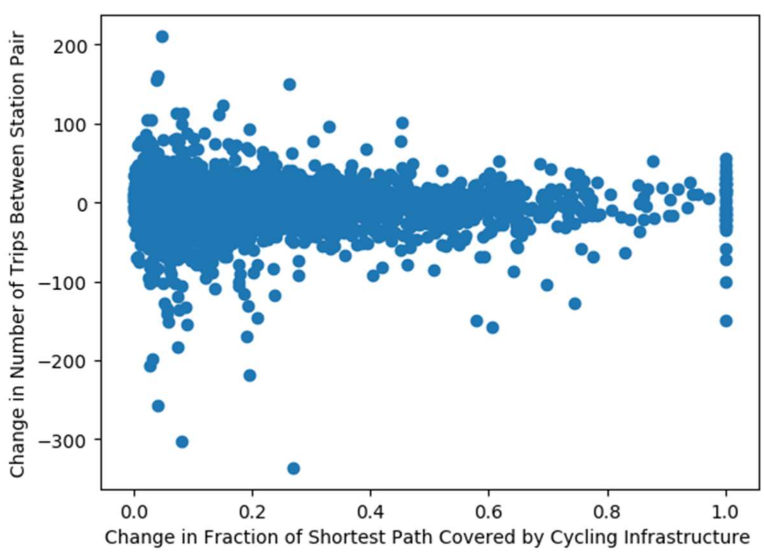
Model

This study used a multiple linear regression model to relate variables from the final dataset of 8293 station-pair-months just described. Three columns of the dataset were used in the final model, two as independent variables and one as the dependent variable (Table 1). See **Appendix I** for histograms of single variables. Figure 4 below shows a scatter plot relating the primary dependent and independent variables, showing that the expected relationship was weak, even the regression coefficients were numerically estimated.

Table 1: Descriptive statistics for regression variables

Variable	Min	Max	Mean	Standard Deviation
Change in Number of Trips Between Station Pair [Δ number of trips per month between stations]	-336.0	211.0	-2.70	21.39
Change in Fraction of Shortest Path Covered by Cycling Infrastructure [Δ fraction of path length covered by CI]	0.0007	1.0000	0.1235	0.1523
Change in Total Number of Trips in the CitiBike System [Δ CitiBike System trips per month]	-15240.0	11940.0	-176.5	8440.9

Figure 4: Illustration of the Relationship Between Change in Number of Trips Between Station Pair and Change in Fraction of Shortest Path Covered by Cycling Infrastructure



Multiple linear regression is a widely used statistical technique in many branches of science and engineering. Its purpose is to find a best-fit linear relationship between the dependent variable and many independent variables. The "best-fit" line is the line which minimizes the sum of the squares of the distances between the proposed line and every data point. The regression coefficients are estimated using an equation of the form given in Equation 1. The use of this equation assumes that the errors are normally distributed above and below the best-fit line.

Equation 1: Estimating regression coefficients using the least-squares solution

$$X^T X\beta = X^T y$$

For this study, the target dependent variables values vector y contained all the differences between numbers of trips for each of the 8293 station-pair-months in the final dataset. The design matrix, X, was 8293×3 and contained a column of one's, a column of the differences in shortest path fraction, and a column of the differences in total number of CitiBike system trips. The regression coefficients vector, β, was 3×1 and contained the regression coefficients being solved for.[16] Once the regression coefficients had been estimated, the model was most intuitively described and used in the form shown in Equation 2.

Equation 2: Multiple linear regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

For the data used in this model, the estimated regression coefficients, as well as the R^2 value for the model, are shown in Table 2. The estimation was carried out with the *statsmodels* python package.[17]

Table 2: Regression model estimated coefficients and R^2 value

Measure	Value	95% Confidence Interval
β_0: Intercept [Δ trips per month between stations]	-2.945	[-3.509, -2.378]
β_1: Beta for Change in Fraction of Shortest Path Covered by Cycling Infrastructure [Δ trips per month between stations/additional CI]	3.166	[0.234, 6.098]
β_2: Beta for Change in Total Number of Trips in the CitiBike System [Δ trips per month between stations/additional CitiBike System trip per month]	8.422E-4	[0.001, 0.001] (due to rounding)
R^2: The amount of variation in the dependent variable explained by the independent variables	0.108	N/A

Discussion

The primary finding of this study is that the R^2 value for the proposed model is quite low, which means that the independent variables explain very little of the variation in the dependent variable. In addition, the regression coefficients are also very small. For example, the meaning of the β_1 coefficient is that the number of trips for a station-pair-month would only go up by 3 if the path between the station-pair went from having no bike lanes at all in the current month to being completely covered by bike lanes along its entire length in the following month, all else remaining equal.

While these results are surprising from an intuitive point of view, similar results were found by Buck and Buehler and Faghih-Imani et al. Both studies found weak positive correlations between the sum length of cycling infrastructure within a radius of bike-sharing stations and the number of average daily bike-share checkouts at those stations. Buck and Buehler's model results suggested that for every additional kilometer of CI installed within a half-mile radius of a bike-sharing station, the average number of bike checkouts per day at that station would increase by 0.86.[3] Faghih-Imani et al. found that for every additional kilometer of CI within a 250-meter radius of a bike-sharing station, the number of hourly departures from the bike-sharing station over that station's capacity increases by 0.0361.[4] Multiplying Faghih-Imani's hourly number of bike-sharing station departures by 24 to get a rough estimate of daily departures yields a coefficient of 0.87 additional daily departures from a bike-sharing station for every additional kilometer of CI within a 250-meter radius of the bike-sharing station.

Conclusions

This study's results corroborate those of past studies, suggesting that adding cycling infrastructure has little impact on the usage of bike-sharing systems. Since the benefits are small, installing CI would only be warranted if the costs of installing it were similarly small, which is often not the case because CI takes up valuable space that could have been used for roadway or parking. However, given the limitations of this and past studies, further investigation may still be warranted. Future researchers in this area should consider the following expansions on this study's analysis which could improve the quality and defensibility of results.

First, the results of this study should be tested for sensitivity to the before-and-after time increment. While this study used a one-month increment, it is plausible that it takes bike-share users much longer than one month to react to CI installations. While weather and seasonal variability may become a concern as the time increment lengthens, high-quality weather data is available, so controlling for seasonal variability could be feasible.

Second, this model could be improved by using more realistic paths instead of assuming the shortest path between each pair of stations. This could be achieved by using Dijkstra's algorithm on a network with adjusted edge lengths. On this adjusted network, the length of edges with CI on them could be discounted by multiplying the true edge length by a discounting factor between zero and one. This discounting would incentivize Dijkstra's algorithm to select slightly longer paths with more CI along their length. The selection of the discounting factor would be an important choice if this type of analysis were being conducted. The discounting factor would depend on the aggregate decision-making patterns of all the bike-sharing users in a city, and should therefore be derived using an econometric analysis.

Third, the analysis used in this study, or an analysis like it, could be carried out on cities other than New York. Motivate, the operator of CitiBike, also operates bike-sharing systems in Boston, the Bay Area of California, Washington D.C., and other major U.S. cities. Motivate publishes data for all those cities free-of-charge online for the public to access in almost the same format as the data for New York City described in this paper. In addition, OpenStreetMap publishes network data for all the cities where Motivate operates bike-sharing systems. Thus, if the CI data is available from the city government itself, then this study's analysis could theoretically be performed on any of the cities where Motivate operates. The author of this study would be particularly interested in seeing data from the bike-share in the Bay Area of California, because seasonal weather variability is much less extreme there than in New York City.

Further research should be conducted, either expanding on this research in the ways suggested above, or through alternative means, to answer the important question: "What impact does cycling infrastructure have on the activity of bike-sharing systems?". A better answer to this question would inform planners and politicians, helping them to more accurately understand the true value of the CI they currently operate or might install. If, as this and past studies suggest, CI is not very effective at improving the usage of bike-sharing systems, then planners and politicians might be well-advised to spend their scarce resources elsewhere.

Works Cited

[1] E. Fishman, S. Washington, and N. Haworth, “Bike Share: A Synthesis of the Literature,” *Transp. Rev.*, vol. 33, no. 2, pp. 148–165, Mar. 2013.

[2] C. Médard de Chardon, G. Caruso, and I. Thomas, “Bicycle sharing system ‘success’ determinants,” *Transp. Res. Part Policy Pract.*, vol. 100, pp. 202–214, Jun. 2017.

[3] D. Buck and R. Buehler, “Bike Lanes and Other Determinants of Capital Bikeshare Trips,” in *TRB 91st Annual Meeting Compendium of Papers*, Washington D.C., 2012.

[4] A. Faghih-Imani, N. Eluru, A. M. El-Geneidy, M. Rabbat, and U. Haq, “How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal,” *J. Transp. Geogr.*, vol. 41, pp. 306–314, Dec. 2014.

[5] I. Mateo-Babiano, R. Bean, J. Corcoran, and D. Pojani, “How does our natural and built environment affect the use of bicycle sharing?,” *Transp. Res. Part Policy Pract.*, vol. 94, pp. 295–307, Dec. 2016.

[6] Boston Bikes, “Boston Bike Network Plan,” Boston Department of Transportation, Boston, 2013.

[7] Motivate International, Inc., “CitiBike,” *CitiBike*, 2018. [Online]. Available: https://www.citibikenyc.com/. [Accessed: 11-May-2018].

[8] New York City, “Bike Routes,” *NYC Open Data*, 2018. [Online]. Available: https://www.citibikenyc.com/system-data. [Accessed: 17-Apr-2018].

[9] Motivate International, Inc., “Motivate Developer Resources,” *Motivate*. [Online]. Available: https://www.motivateco.com/use-our-data/. [Accessed: 17-Apr-2018].

[10] District of Columbia, Department of Transportation, “Bicycle Lanes,” *DC.gov*, 2018. [Online]. Available: http://opendata.dc.gov/datasets/bicycle-lanes. [Accessed: 18-Apr-2018].

[11] Boston Maps, “Existing Bike Network,” *Analyze Boston*, 2018. [Online]. Available: https://data.boston.gov/dataset/existing-bike-network. [Accessed: 17-Apr-2018].

[12] K. Jordahl, J. Van Den Bossche, and J. Wasserman, *GeoPandas*. 2018.

[13] G. Boeing, “OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks,” *Comput. Environ. Urban Syst.*, vol. 65, pp. 126–139, Sep. 2017.

[14] S. Gillies, M. Toews, and J. Arnott, *Shapely*. 2018.

[15] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring Network Structure, Dynamics, and Function using NetworkX,” in *Proceedings of the 7th Python in Science Conference*, Pasadena, CA USA, 2008, pp. 11–15.

[16] D. C. Lay, S. R. Lay, and J. McDonald, *Linear algebra and its applications*, Fifth edition. Boston: Pearson, 2016.

[17] S. Seabold and J. Perktold, “Statsmodels: Econometric and statistical modeling with python,” presented at the 9th Python in Science Conference, 2010.

Appendix I: Descriptive plots for selected regression variables

Figure 5: Histogram of Change in Number of Trips Between Station Pairs

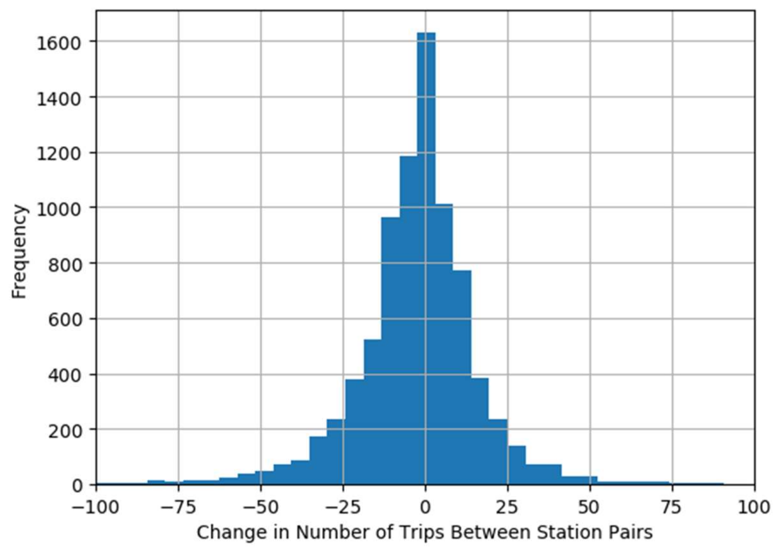


Figure 6: Histogram of Change in Fraction of Shortest Path Covered by Cycling Infrastructure

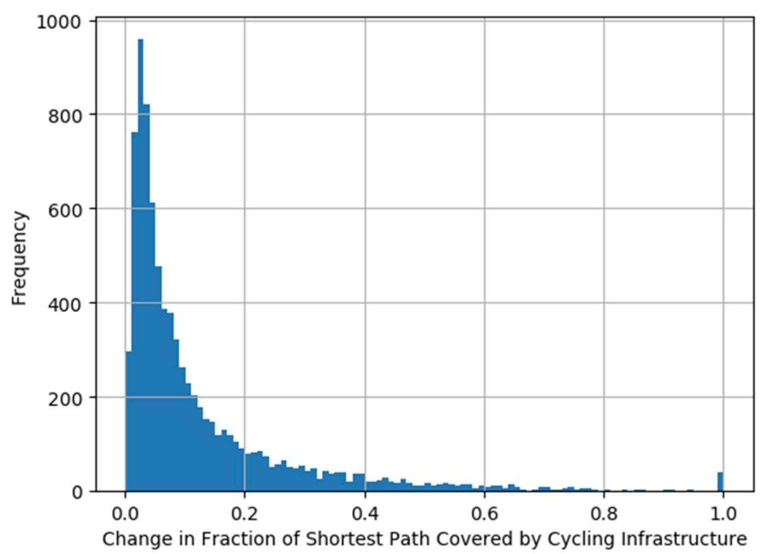


Figure 7: Illustration of the Relationship Between Change in Number of Trips Between Station Pair and Change in Total Number of Trips in the CitiBike System