Taylor & Francis
Taylor & Francis Group

International Journal of Sustainable Transportation

ISSN: 1556-8318 (Print) 1556-8334 (Online) Journal homepage: www.tandfonline.com/journals/ujst20

Bikesharing Trip Patterns in New York City: Associations with Land Use, Subways, and Bicycle Lanes

Robert B. Noland, Michael J. Smart & Ziye Guo

To cite this article: Robert B. Noland, Michael J. Smart & Ziye Guo (2019) Bikesharing Trip Patterns in New York City: Associations with Land Use, Subways, and Bicycle Lanes, International Journal of Sustainable Transportation, 13:9, 664-674, DOI: 10.1080/15568318.2018.1501520

To link to this article: https://doi.org/10.1080/15568318.2018.1501520

INTERNATIONAL JOURNAL OF SUSTAINABLE TRANSPORTATION
2019, VOL. 13, NO. 9, 664–674
https://doi.org/10.1080/15568318.2018.1501520

Taylor & Francis
Taylor & Francis Group

Check for updates

Bikesharing trip patterns in New York City: Associations with land use, subways, and bicycle lanes

Robert B. Noland, Michael J. Smart, and Ziye Guo

Rutgers University, Edward J. Bloustein School of Planning and Public Policy, New Brunswick, NJ, USA

ABSTRACT

As bikesharing systems have proliferated, few studies have examined the trips made on these systems. In this paper, we examine trips between origin-destination pairs during three months in 2015 on New York City's Citi Bike system. Findings suggest considerable variation across user types, across months, and across times of day. Principal findings indicate that bikesharing is used for transit access and egress during rush hours, and that stations located along the same high-quality bicycle route see far more trips than do other station pairs. Casual users complement subscribers' usage by using bicycles more frequently during midday and the evening, and between areas characterized by nearby recreational land uses. Loop trips to and from the same station also occur and are likely recreational trips. The data analyzed is essentially a form of "big data." That is, large data sets that are ubiquitously collected. The analysis suggests that in this case, "big data" that lacks the socio-economic data commonly collected and used in travel analysis can provide useful insights to planners.

ARTICLE HISTORY
Received 27 April 2017
Revised 12 April 2018
Accepted 13 July 2018

KEYWORDS
Bikeshare; big data; travel patterns

1. Introduction

Bikesharing is now established in hundreds of cities and locations throughout the world. The Citi Bike system in New York City is one of the largest, and some stations have well over a thousand customers each month. Previous work has linked the generation of these trips to demographic and land use factors, proximity to subway stations, and nearby bicycle infrastructure (Noland, Smart, & Guo, 2017a). In general, it is found that being near residential areas, employment zones, and subway stations leads to more trips, with some variation between weekday and weekend usage, as well as for those who are subscribers versus casual users. However, little is known about the patterns associated with these trips. That is, how frequently are specific trips taken? What type of user takes certain trips and during which day of the week or time of day?

While scholarship on bikeshare usage is in its infancy, a large number of studies have examined cycling behavior. In the United States, women are considerably less likely to cycle than are men, perhaps due to their more varied trip destinations and greater household labor burdens (Garrard, Handy, & Dill, 2012). As individuals age, the bulk of research suggests cycling rates drop off (Dill & Voros, 2007; Pucher, Komanoff, & Schimek, 1999). The relationship between income and cycling is murkier; some studies find that higher income is associated with less cycling (Guo, Bhat, & Copperman, 2007), while others find a positive correlation (Dill & Voros, 2007) and others find no relationship (Dill & Carr, 2003).

Of particular interest for our analysis is the role of the built environment. The provision of bicycle lanes and bicycle paths is associated with increased rates of cycling (Buehler & Pucher, 2012; Dill, 2009), and that this relationship appears stronger for women than for men (Garrard, Rose, & Lo, 2008). Environments characterized by a greater mix of land uses are also associated with more cycling, likely due to the greater proximity of diverse origins and destinations (Forsyth & Krizek, 2010). While behavior of cyclists and those using bikeshare may differ, we expect that there may be some similarities. For example, in our analysis we examine trips taken with bikeshare and associate these with the proximity of bikeshare trips to bike lanes, the land uses surrounding bikeshare stations, and links to the New York City subway.

Most analyses of cycling behavior rely on collection of survey data. Bikeshare data is ubiquitously collected; that is, "big data," as the data streams are continuously produced, generating very large datasets that allow one to examine the patterns of user behavior. This offers opportunities for detailed analysis, and urban planners are beginning to examine the patterns in these data (Batty, 2013). The main shortcoming is that these data tend to lack information on individuals, as opposed to surveys.[1] In this research we link the data on trips to spatial data. This is done without the detailed individual or household survey data usually

CONTACT Robert B. Noland rnoland@rutgers.edu Edward J. Bloustein School of Planning and Public Policy, Rutgers University, 33 Livingston Avenue, New Brunswick, NJ 08901

[1]The Citi Bike data distinguishes between subscribers and casual (one-time) users. The former includes self-reported gender and age, but we do not analyze these components, only differences between subscribers and casual users.

collected in transportation research. Our objective is to answer key questions about the trips taken with the bikeshare system, using only openly available data sources. Our primary objective is to examine the links between land use, subway stations, and bike lanes on the one hand, and bikesharing trips on the other; for this purpose, additional socio-demographic survey data would be of interest but not of crucial need.

Other bikeshare systems have also been analyzed and have provided theoretically consistent results. These include Washington, DC where it was found that bicycle lanes, greater proximity to bars and restaurants, and—surprisingly—higher levels of motor vehicle ownership in the area (though likely due to income levels) play a role in bikeshare usage (Buck & Buehler, 2012) as does weather (Gebhart & Noland, 2014). Seasonal variability affects the number of trips taken, but specific events, such as rain, also deter usage (Gebhart & Noland, 2014). In Minneapolis, there are positive associations with bicycle infrastructure and nearby levels of economic activity (Wang, Lindsey, Schoner, & Harrison, 2015). The system in London, UK, has also been analyzed. Of interest is the change in usage patterns when casual users (i.e. those not subscribing as annual members) were allowed to use the system. The result was an increase in weekend usage (Lathia, Ahmed, & Capra, 2012). Most analyses link bikeshare data to geographic variables to examine associations, and these are usually based on travel behavior theory.

Another approach to analyzing bikeshare data is to examine patterns visually. For example, a study of the London bikeshare data examined differences in usage between men and women (Beecham & Wood, 2014). Women were found to prefer safer routes and are less likely to use bikeshare for commuting (based on the pattern of flows from major rail hubs). Another visualization study examined data in Nanjing, China (Zhao, Wang, & Deng, 2015). Differences in how men and women use the bikes on different days of the week are examined. Of particular interest for our study is their finding that those trips that start and end at the same docking station (what we define as "loop trips") are most likely recreational. This is based on their finding that these are mainly taken by men on weekends.

The analysis conducted here takes a different approach than most prior analysis, which focused on trips generated from bikeshare stations or visualization patterns. Each trip is evaluated as a linkage between an origin and a destination; that is, where the bicycle is checked out and where it is returned, similar to visualization studies. Thus, the unit of analysis is the number of trips between origin-destination (OD) pairs, but we include a statistical analysis to evaluate factors associated with these, disaggregating by user type (subscriber vs. casual users), day of the week (weekday vs. weekend/holidays), and time of day (morning vs. evening peaks). Our data obviously lack the specific start and stop points of the full trip, as we only have the portion of the trip made with Citi Bike. We also do not have information on the specific route taken which would require access to GPS data, which was not available.

While the determinants of OD pairs among bikesharing systems have not been analyzed before, OD modeling is quite common (for a thorough overview of its application in transportation planning, see Antoniou et al., 2016). Origin-destination models are specific applications of the general concept of a spatial interaction model, which assumes that interactions between units of geography vary systematically by proximity (with closer areas having more interactions) and also by characteristics of the local population, economic conditions, and so forth (Sen & Smith, 2012). In addition to its widespread application in transportation planning, spatial interaction modeling has been used to estimate flows of information and innovation (Fischer, Scherngell, & Jansenberger, 2006), retail trade (Lee & Pace, 2005), and international trade (Porojan, 2001), among others.

The objective of this paper is to analyze the type of trips made, based on the origin and destination and the patterns in the data. The questions that are examined include: what correlates might be associated with OD pairs, how do they vary seasonally, what differences are there between subscribers and casual users, and how do trips vary throughout the day? Little is known about the type of OD pairs that are most common; that is, do customers generally take short trips? Do they take trips with a subway station as either an origin or destination? How does the provision of bicycle lanes influence their trip patterns? Understanding these issues is useful for those planning new bikeshare systems or extensions to existing systems. It also provides an understanding of bicycle travel patterns in a city, which can provide insight on where to provide bicycle infrastructure as well as how to link the bikeshare stations to public transit. This analysis also provides a useful demonstration of how data that lack information on individual socio-economics and the travel intent of users can reveal useful insights.

2. Background on Citi Bike

Citi Bike generates over one million trips per month. This provides a surfeit of data that allows us to study various travel patterns. As of November 2015 there were 475 stations, and thus over 225,000 possible OD pairs that can be accessed using a bikeshare bicycle. Similar to most bikeshare systems, Citi Bike sets their pricing to penalize longer trips. Citi Bike (as of April 2018) has three pricing options. Annual subscriptions are \$14.95/month or \$169.00/year. Subscribers pay no additional fee for the first 45 minutes of any trip, and then pay an additional \$2.50 per 15 minute period for any bicycle used over 45 minutes.[2] For non-subscribers, 24 hour passes are \$12.00 and 72 hour passes are \$24.00; with these, only the first 30 minutes of a given trip are free, and thereafter charges are \$4.00 per 15-minutes. Another option allows a single 30 minute trip for \$3.00, although this was not available at the time we downloaded the data (Citi Bike, 2018).

Figure 1 shows the service area of Citi Bike in New York City in October of 2015, covering Manhattan south of 86th

[2]At the time corresponding to our data, the penalty for longer trips was greater for annual subscribers. An additional \$6.50 for 76–105 minutes, and then an additional \$9.00 per half-hour was added for trips lasting longer than 105 minutes.

Figure 1. Citi Bike Service Area in New York City, October 2015.

Street and inner portions of the boroughs of Brooklyn and Queens. A small number of additional Citi Bike stations were also located across the Hudson River from the tip of Manhattan in Jersey City, New Jersey at this time, but we do not consider them in our analysis, as there is no reasonable direct connection between lower Manhattan and New Jersey for cyclists.[3]

3. Data

Citi Bike usage data were downloaded from the Citi Bike website (Citi Bike, 2016). Three months of data were obtained, February, July, and October 2015, providing seasonal variation in usage. This seasonal variation proxies for both differences in temperature and daylight hours, both of which can affect usage (Noland & Ishaque, 2006).[4] Average precipitation levels do not vary much by month, although in February this is often snow. The data include information on the origin and destination stations (i.e. where the bicycle was checked out and returned), the date and time of both check out and return, duration of the trip (in seconds), and

[3]A bicycle and pedestrian path is available on the George Washington Bridge, but this would require about a 20 mile bicycle trip from midtown Manhattan to Jersey City.

[4]Average monthly mean temperatures in New York City for February are high of 5 °C, low of −2 °C, July high of 29 °C, low of 21 °C, October high of 18 °C, and low of 10 °C (*Current results, weather and science facts*, 2018). The total daylight ranges from 10 hours, 5 minutes at the start of February, to 11 hours, 36 minutes at the end of February; for July the range is from 15 hours, 3 minutes, to 14 hours, 24 minutes; and, for October from 11 hours, 49 minutes, to 10 hours, 31 minutes (Timebie, 2018).

Table 1. Type of trips taken, by month.

	Subscriber trips	Casual trips	Weekday trips	Weekend/ holiday trips	Total trips
February 2015	194,463	2250	151,919	44,794	196,713
	98.9%	1.1%	77.2%	22.8%	
July 2015	904,267	178,454	826,935	255,786	1,082,722
	83.5%	16.5%	76.4%	23.6%	
October 2015	1,062,505	144,254	889,337	317,382	1,206,863
	88.0%	12.0%	73.7%	26.3%	

whether the user has a subscription or is a casual customer. The latitude and longitude of each station location is also included, allowing other spatial data to be linked to each station.

In February 2015, there were 328 active stations. In July, the system had 330 stations. In October, the number of active stations grew to 463 after an expansion of the network further north in Manhattan, plus additional stations in Brooklyn and some in Queens. The number of possible trips between unique bikeshare stations, n, is equal to $n^2–n$. Most stations, however, have some "loop" trips that start and end at the same station; thus there are n possible loop trips. The number of possible OD pairs therefore range from 107,256 in February, to 213,906 in October, which shows the potential increase in access that these systems can provide.

The data were purged of various outliers, based on the trip duration. Any trip in excess of three hours was eliminated from the data, as well as any trip of 60 seconds or less.[5] The former most likely represent cases where the user did not properly return the bicycle to a docking station, as it is unlikely that many users would opt to pay the high fees associated with keeping the bicycle for an extended time period. Those less than 60 seconds were likely cases where the bicycle was found to be in disrepair and was returned to the docking station. These latter are generally removed from the raw data on-line, but some were found in the processed data. Fewer than 0.5% of trips were removed from the data.

Table 1 shows the counts for each month of the year, broken down by user type and whether the trip was on a weekday or weekend/holiday. There are over five times more trips in both July and October than in February, though part of the increase in October is likely due to the system expansion that occurred. February 2015 was one of the coldest Februaries on record in New York City, likely explaining much of the reduction in usage. In fact, there were very few casual trips in February, with about 80 times more in July and 65 times more in October. This is perhaps related to the recreational nature of many casual users' bikeshare trips, contrasted with subscribers who may be more likely to use bikeshare for utilitarian trips year-round. One of our objectives is to tease out some of these differences in usage between the different user groups.

Data on land use, at the tax lot level, were obtained from the New York City Department of City Planning. Service areas for each bikeshare station were defined using Thiessen polygons. This method defines the service area boundaries around each bikeshare station, with space allocated to the nearest bike share station. The service areas are also truncated at one-quarter mile with the assumption that this is the maximum walk access distance; the land use data were then allocated to these polygons to obtain the percent land use in various categories around each station.

[5]These are now being removed from the dataset by Citibike prior to making the data available.

Table 2. Mean and median duration, in minutes, for loop and OD trips, by month.

	Loop trips		OD trips	
	Mean	Median	Mean	Median
February 2015	15.9	10.7	10.4	8.4
July 2015	29.8	22.7	14.1	11.0
October 2015	25.4	17.4	13.7	10.5

Subway station locations and bicycle routes were also linked to each station. For subways a dummy variable was included for every bikeshare service area containing a subway station. Another dummy variable was defined if the same high-quality bicycle route (for instance, the Hudson River Greenway) was within the service areas of both the origin and destination stations. Four high quality bicycle routes in Manhattan were specified: the Hudson River Greenway, the East River Greenway, 8th/9th Avenue, and 1st/2nd Avenue, the latter two being one-way pairs. This was based on evidence that some of these trips are quite frequent, as can be seen in Figure 2 which maps the top 100 most frequent trips in July 2015. In particular, OD pairs that could be reasonably traveled using the Hudson River Greenway have a substantial number of trips.

Counts of each trip were summed up for each possible OD pair, including by user type, weekday/weekend holiday, and by certain times of day.[6] There are a non-trivial number of trips that start and end at the same bikeshare station; we define these as "loop" trips, versus trips to another destination, which we define as OD trips. In February 1.5% of trips are loop trips; in July this rises to 2.6% of trips. Loop trips in October account for 1.9% of trips.

Trip durations also show seasonal variation, with loop trips being longer in all cases. February trips are much shorter than those in July and October. The trip duration mean and median values are shown in Table 2.

This preliminary analysis sets the stage for further in-depth exploration of the trip patterns that emerge from very large datasets. While these data lack the type of demographic and trip purpose information usually found in travel surveys, an exploration of the patterns provides some useful insights. Specifically, the analysis that follows focuses first on cross-tabulations of subscribers versus casual users, whether loop or OD trips are taken, the day of the week and time of day of these trip types. This is followed by statistical models to evaluate the associations between various OD trips with land use patterns, subway, and bicycle lane availability.

[6]Holidays were Presidents Day in February, Independence Day in July, and Columbus Day in October. In February there were 19 weekdays and 9 weekend/holiday days, in July this was 22 weekdays and 9 weekend/holiday days, and in October there were 21 weekdays and 10 weekend/holiday days.

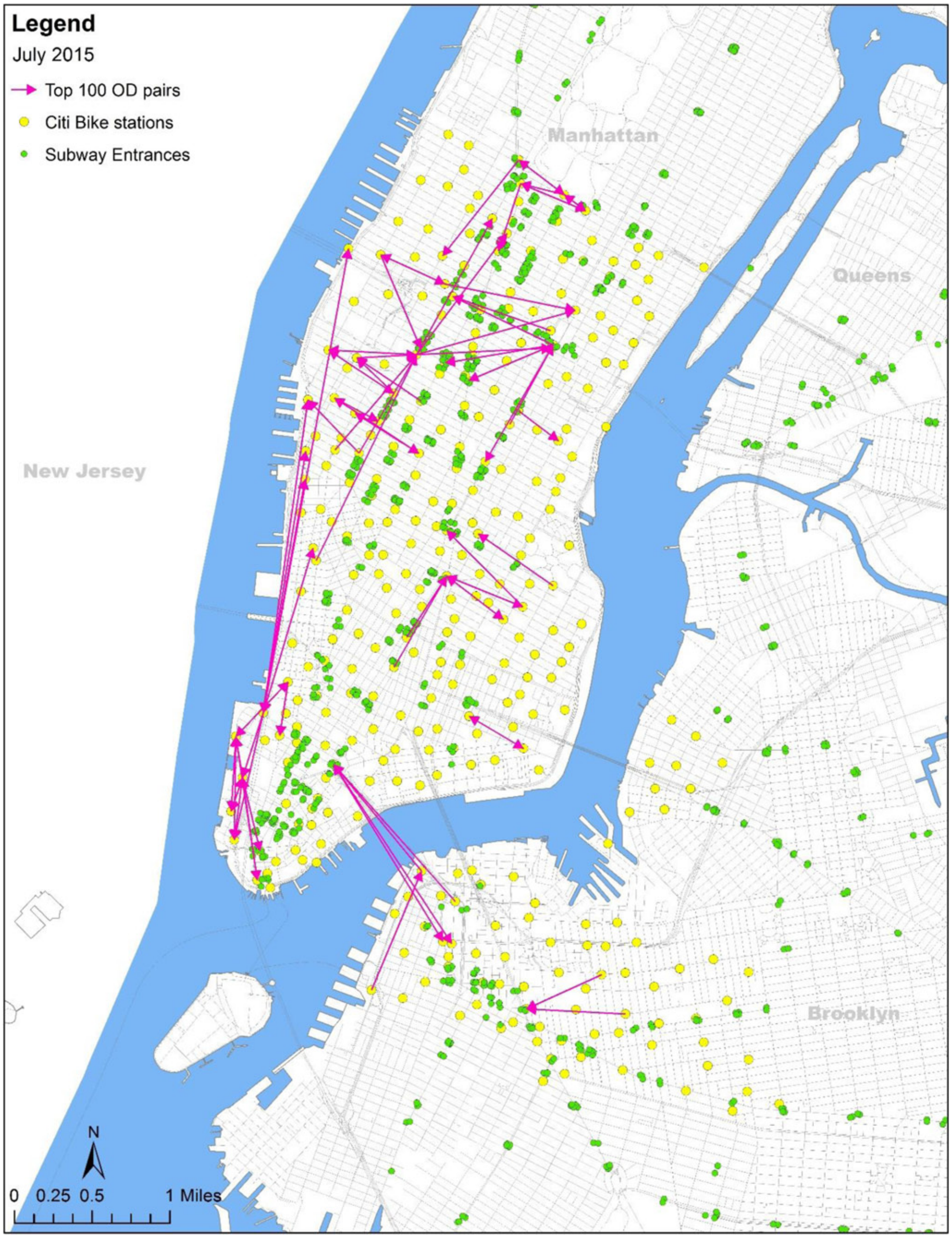

Figure 2. Top 100 most frequent origin-destination trips, July 2015.

4. Preliminary analysis and methods

While the purpose of each trip is not known, it is likely that loop trips are mainly recreational trips. In fact, the most frequent loop trip begins and ends at the Central Park South & 6th Avenue bikeshare station. This station is located at an entrance to Central Park, which offers plentiful recreational cycling opportunities. In July 2015, 3205 loop trips were taken from this station, in October 2015, this dropped to 1343 trips, while in February 2015, there were only 85 loop trips.[7]

To further analyze these differences in trip patterns, cross-tabulations of the average daily loop trips and OD trips are interacted with user type (subscriber versus casual users) and day of the week (weekday vs. weekends/holidays) for each month. Daily trip averages are based on weekdays and weekends/holidays for each month. These results are shown in Table 3.

The expected pattern is that loop trips are more likely made by casual users and on weekends/holidays. This is confirmed by the relatively large percent of loop trips made by casual users relative to subscribers (6–10% of all casual users' trips, vs. under 2% for subscribers), although the total number of trips is much less since there are far more subscriber relative to casual trips. Subscribers make more OD trips during weekdays, while casual users make more of these trips on the weekend/holidays. This suggests that

[7]It is certainly possible that some loop trips are for utilitarian trips, such as going to a grocery store. However, given the density of bikeshare stations in New York City and the pricing schedule, it seems unlikely that users would not return their bicycle to a nearby docking station. Further, the bicycles lack built-in locking mechanisms, meaning that parking a bicycle unattended outside a store poses a significant risk of theft and hefty fines for the user should the bicycle be stolen.

Table 3. Cross-tabulations of trips per day by trip type, user type, and day of the week.

	Subscriber				Casual			
	Weekday		Weekend/holiday		Weekday		Weekend/holiday	
	Loop trips	OD trips	Loop trips	OD trips	Loop trips	OD trips	Loop trips	OD trips
February 2015	99	7836	88	4768	5	56	14	108
July 2015	461	32,683	473	18,982	319	4124	691	8275
October 2015	442	38,129	497	24,754	210	3570	423	6064

Table 4. Cross-tabulations of AM, mid-day, and PM trips per day, by trip type, user type, and day of the week.

	Subscriber				Casual			
	Weekday		Weekend/holiday		Weekday		Weekend/holiday	
	Loop trips	OD trips	Loop trips	OD trips	Loop trips	OD trips	Loop trips	OD trips
February								
AM (7 am–10 am)	15	1831	7	447	1	7	1	6
Mid-day (11 am–2 pm)	20	914	21	1189	2	13	4	31
PM (4 pm–7 pm)	24	2239	21	1068	1	12	3	24
July								
AM (7 am–10 am)	59	7692	42	1509	22	367	42	347
Mid-day (11 am–2 pm)	84	3649	90	4096	70	800	168	1857
PM (4 pm–7 pm)	105	8992	110	4366	72	1067	179	2188
October								
AM (7 am–10 am)	67	9894	43	2426	13	319	24	281
Mid-day (11 am–2 pm)	90	4503	106	5488	61	917	122	1671
PM (4 pm–7 pm)	111	10,176	129	6189	45	895	105	1626

subscribers are potentially using the bikeshare for commuting and daily utilitarian trips. Casual users are more likely to use the bikeshare for other types of trips.

In Table 4, a breakdown by three 3-hour periods (AM peak, mid-day, and PM peak) is presented, again based on average number of trips per day for each time period. Table 4 displays these results as a percent of the total for each user type (i.e. subscriber vs. casual). Loop trips are generally more common later in the day and on weekends/holidays, though for the AM peak, loops are more common during the week. These could be recreational rides taken for exercise in the morning. For casual users fewer AM loop trips are made and far more are made during mid-day and PM time periods on weekends and holidays.

Subscribers' weekday OD trips follow a typical diurnal travel pattern; trips are fewest during the mid-day period, with peaks during the AM and PM periods, with the latter being a bit higher. This pattern is similar in each month, suggesting some regularity for what are likely commuting trips. This is confirmed by examining the weekend/holiday use patterns of subscribers, which have the fewest trips in the AM time period and somewhat similar mid-day and PM trips. Casual users follow the same pattern on both weekdays and weekends/holidays, and this is similar to the weekend/holiday use by subscribers. Thus it seems likely that in both cases these are not regular commute trips, but may be either utilitarian or recreational.

To examine factors associated with trip-making between each one-way directional OD pair, a multivariate analysis was conducted. Independent variables included two dummy variables for a subway station located near the origin and destination stations, variables for the percent of various land use types near each station, and a dummy variable indicating whether both the origin and destination station are along the same high-quality bicycle path. A dummy variable for a loop trip and a variable for street- and path-network distance between stations were also included.

Table 5. OD pairs with no trips, by month.

	OD pairs with zero trips	Total possible OD pairs	Percent not taken
February 2015	70,955	107,584	66.0%
July 2015	35,894	108,900	33.0%
October 2015	114,473	214,369	53.4%

Many OD pairs have zero trips. Therefore, as the data are count data and non-normally distributed, a zero-inflated negative-binomial model was estimated. Table 5 displays the variation in OD pairs with zero trips by month. More OD pairs do not have trips in February, most likely because of colder weather. October similarly has a larger fraction of OD pairs with no trips, possibly because some distances are now much longer given the expanded network. Our rational for using a zero-inflated model is that many of the possible OD pairs have no trips, mainly because of the distance between station pairs (up to 7.8 miles in the February data, and 9.0 miles in the October data, following station expansion). Thus, we control for distance to determine the likelihood of a zero in the inflated model.[8]

The land use variables were aggregated based on similar functions, as well as according to which land use types were collinear—that is, those that tended to exist in the same area; no other independent variables were collinear. Residential land uses are the omitted category; thus the

[8]Negative binomial models were also estimated. The zero-inflated models provided a slightly better fit based on Akaike Information Criteria. The prior results were presented in a paper at the Transportation Research Board (Noland, Smart, & Guo, 2017b).

Table 6. Zero-inflated negative-binomial models for February 2015.

	Subscriber trips Coef. (z-stat)	Casual trips Coef. (z-stat)	Weekday trips Coef. (z-stat)	Weekend/ holiday trips Coef. (z-stat)	All trips, 8–9 am Coef. (z-stat)	All trips, 5–6 pm Coef. (z-stat)
Loop trip	−0.48	1.97	−0.45	−0.54	−0.23	−0.21
	(−6.67)	(11.09)	(−5.88)	(−6.71)	(−1.52)	(−1.93)
Distance (miles)	−1.05	0.01	−0.93	−1.50	−0.38	−0.54
	(−101.60)	(0.19)	(−86.39)	(−143.99)	(−14.35)	(−22.32)
Percent mixed land use, commercial, office, public at origin	0.86	1.48	1.09	−0.09	0.06	1.92
	(24.81)	(8.23)	(28.83)	(−1.93)	(0.73)	(27.09)
Percent industrial, transportation, parking at origin	−0.66	0.76	−0.43	−1.28	−1.14	−0.06
	(−9.58)	(2.24)	(−5.75)	(−13.53)	(−6.22)	(−0.40)
Percent open space/recreational at origin	−14.21	−2.03	−13.72	−17.90	−10.82	−13.05
	(−17.88)	(−0.58)	(−15.86)	(−14.93)	(−5.40)	(−8.06)
Percent unknown land use at origin	1.69	47.05	5.53	−35.59	−44.91	66.35
	(0.30)	(1.89)	(0.90)	(−4.49)	(−3.10)	(6.19)
Percent mixed land use, commercial, office, public at destination	0.82	1.36	1.01	−0.02	2.80	0.43
	(23.67)	(7.55)	(26.89)	(−0.43)	(28.06)	(6.54)
Percent industrial, transportation, parking at destination	−0.48	0.74	−0.26	−1.09	0.02	−0.46
	(−6.97)	(2.16)	(−3.51)	(−11.64)	(0.09)	(−3.23)
Percent open space/recreational at destination	−14.48	2.05	−14.38	−15.63	−9.84	−18.88
	(−18.31)	(0.62)	(−16.68)	(−13.50)	(−4.77)	(−10.86)
Percent unknown land use at destination	12.51	8.08	18.15	−35.30	75.31	−44.74
	(2.19)	(0.31)	(2.94)	(−4.41)	(5.32)	(−3.89)
Origin near subway	−0.04	0.16	−0.03	−0.03	−0.37	0.11
	(−2.42)	(2.12)	(−1.95)	(−1.71)	(−10.02)	(3.85)
Destination near subway	0.02	0.27	0.02	0.03	0.16	−0.01
	(1.29)	(3.59)	(1.28)	(1.45)	(3.91)	(−0.39)
Along bicycle path	0.58	0.54	0.56	0.55	0.45	0.46
	(14.79)	(3.73)	(13.52)	(12.05)	(5.14)	(7.21)
Constant	1.49	−4.70	0.82	1.78	−1.93	−1.55
	(43.03)	(−20.18)	(21.87)	(39.37)	(−20.55)	(−22.66)
Inflated—logit model						
Distance (miles)	1.24	1.28	1.31	0.63	1.39	1.34
	(64.50)	(15.41)	(69.42)	(0.03)	(40.25)	(41.76)
Constant	−3.70	−0.89	−3.49	−16.03	−1.74	−2.18
	(−72.73)	(−3.16)	(−71.97)	(−0.03)	(−19.79)	(−29.70)
Dispersion parameter	1.46	4.47	1.56	1.62	3.77	2.00
McFadden's Pseudo R^2	0.059	0.028	0.050	0.046	0.024	0.027
N	107,584	107,584	107,584	107,584	107,584	107,584

parameter estimates on the other land use variables can be interpreted relative to the influence of residential land uses on trip-making.

Models are separately estimated for subscriber trips, casual user trips, workday trips, weekend/holiday trips, and peak morning and evening time of day (for all users): 8 am–9 am and 5 pm–6 pm. Results for the three months are shown in Tables 6–8. October has a larger range of potential values for the independent variables given the increase in bikeshare stations.

5. Results

Model results are displayed in Tables 6–8 for each of the three months analyzed. In all but one model, the distance coefficient for the zero-inflation logit equation is statistically significant. The model for weekend/holiday trips did not have a significant coefficient, and is equivalent to a simple negative-binomial model. All other coefficient estimates support the hypothesis that greater distance between the station pairs increases the likelihood of not having any trips (i.e, coefficients are all positive).

The models control for loop trips. While these are relatively infrequent compared to those trips with different origins and destinations, results for casual users are positive and statistically significant, as opposed to no statistical association with subscriber trips. This suggests that casual users have a greater likelihood of using Citi Bike for loop trips and this supports our hypotheses that these are likely recreational trips. In July and October, weekend and holiday trips are more likely to be loops, while in February these are less likely to be so. This again suggests that these are recreational, since we do not expect recreational trips in February, due to colder temperatures. Those trips in the AM peak period also have no association with loop trips, since these are likely commute trips. Surprisingly, those trips during the PM peak in July and October are also more likely to be loops, and again, these may be recreational trips.

Within the primary equation, distance is highly significant in all but one of the models. Greater distance between stations results in less frequent trips between them. The one exception is casual trips in February which has a coefficient estimate that is not statistically significant, though distance is highly significant in the zero-inflation equation in that model. This may be an artefact of very few casual users in February (see Table 4, summing the casual user trips gives a total of 105 trips for the entire month).

The land use variables are based on the percent of each land use within the station service area. These should be

Table 7. Zero-inflated negative-binomial models for July 2015.

	Subscriber trips Coef. (z-stat)	Casual trips Coef. (z-stat)	Weekday trips Coef. (z-stat)	Weekend/ holiday trips Coef. (z-stat)	All trips, 8–9 am Coef. (z-stat)	All trips, 5–6 pm Coef. (z-stat)
Loop trip	0.05	1.78	0.33	0.93	0.03	0.50
	(0.85)	(24.00)	(5.25)	(15.26)	(0.26)	(6.58)
Distance (miles)	−0.87	−0.57	−0.85	−0.75	−0.50	−0.64
	(−198.60)	(−69.05)	(−188.54)	(−114.83)	(−43.65)	(−69.09)
Percent mixed land use, commercial, office, public at origin	1.01	1.30	1.21	0.47	−0.15	2.17
	(42.80)	(35.80)	(50.54)	(16.31)	(−3.01)	(55.82)
Percent industrial, transportation, parking at origin	−0.69	0.50	−0.42	−0.64	−1.26	0.45
	(−15.39)	(7.03)	(−9.23)	(−11.32)	(−12.37)	(5.92)
Percent open space/recreational at origin	−9.20	2.99	−7.59	−3.83	−16.86	−6.84
	(−20.49)	(4.70)	(−16.92)	(−6.99)	(−14.84)	(−9.43)
Percent unknown land use at origin	63.64	80.05	69.23	63.91	−11.89	111.48
	(18.02)	(15.80)	(19.40)	(15.38)	(−1.51)	(20.59)
Percent mixed land use, commercial, office, public at destination	0.96	1.34	1.16	0.48	3.04	0.53
	(40.97)	(37.31)	(49.07)	(16.74)	(54.15)	(14.54)
Percent industrial, transportation, parking at destination	−0.62	0.57	−0.37	−0.54	0.92	−0.35
	(−13.92)	(8.02)	(−8.05)	(−9.60)	(8.37)	(−4.84)
Percent open space/recreational at destination	−8.97	0.13	−8.40	−5.06	−8.06	−12.10
	(−20.01)	(0.19)	(−18.57)	(−9.23)	(−8.26)	(−15.44)
Percent unknown land use at destination	68.73	87.76	71.30	77.83	126.24	33.72
	(19.54)	(17.31)	(20.03)	(18.72)	(16.48)	(6.06)
Origin near subway	0.01	−0.10	0.02	−0.09	−0.21	0.12
	(1.49)	(−6.63)	(2.04)	(−7.31)	(−9.82)	(7.71)
Destination near subway	0.02	−0.11	0.02	−0.09	0.28	−0.04
	(1.60)	(−7.33)	(1.87)	(−7.66)	(12.57)	(−2.51)
Along bicycle path	1.02	1.09	1.03	1.13	0.76	0.97
	(30.45)	(25.81)	(30.66)	(32.87)	(12.63)	(22.77)
Constant	2.37	−0.06	1.92	1.88	−1.04	−0.60
	(100.58)	(−1.65)	(80.38)	(64.40)	(−19.84)	(−15.99)
Inflated—logit model						
Distance (miles)	1.09	1.32	1.11	1.17	1.10	1.06
	(65.68)	(76.23)	(69.32)	(77.35)	(60.18)	(58.66)
Constant	−5.35	−3.92	−5.21	−4.01	−2.75	−3.48
	(−76.20)	(−68.80)	(−79.32)	(−79.86)	(−47.15)	(−55.45)
Dispersion parameter	1.22	1.70	1.23	1.14	2.90	1.58
McFadden's Pseudo R^2	0.077	0.042	0.076	0.055	0.035	0.047
N	108,900	108,900	108,900	108,900	108,900	108,900

interpreted relative to purely residential areas, our excluded category. Those stations with a higher fraction of mixed land uses, commercial, office, and public buildings at both the origin and destination points have positive coefficient values for both subscriber and casual user trips, as well as those trips made on weekdays, and this holds for all months. The main distinction is with weekend/holiday trips which have a smaller coefficient and in one case is not statistically significant. In particular, AM trips seem to have the least association with the origin being linked to these land uses, most likely as bike origin points are in more residential areas in the morning.

Trip associations with industrial, transportation, and parking land use fractions tend to be negative in most models for both origin and destination points. This likely reflects these areas being less attractive for cycling or simply having fewer activity sites to attract or produce trips. There are, however, some results that stand out. In particular, in February and July, the coefficients are positive for casual users. This may reflect a lack of accessibility by other modes, but the result does not stand up in October where the coefficient on the casual user model is negative. Coefficients at the destination are also positive in the AM peak models, though not significant in February, suggesting these are commute trip destinations. The July PM model also has a positive and significant value on the origin, while other months are not significant.

Those bikeshare station areas with more open space and recreational land uses generally have fairly consistent results for all months; the only positive association is for casual users at both the origin and destination point (except a small negative coefficient for February and not statistically significant at the destination in all months). The coefficients are negative in every other case. This adds some further evidence that casual users are more likely to be using the bicycles for recreational purposes as they are more likely to access bikeshare near recreational land uses.

Proximity to a subway station is also included for both origin and destination stations. For the subscriber model, the coefficient on the origin point is negative (February and October) and not significant in July. The negative coefficient suggests that these trips are less likely to start near a subway station. For the destination point, February and July are not significant while October is negative. The October coefficient values are also much larger, perhaps due to the expansion of the bikeshare system. We explore this by time of day below, and find important variation across time periods.

Table 8. Zero-inflated negative-binomial models for October 2015.

	Subscriber trips Coef. (z-stat)	Casual trips Coef. (z-stat)	Weekday trips Coef. (z-stat)	Weekend/ holiday trips Coef. (z-stat)	All trips, 8–9 am Coef. (z-stat)	All trips, 5–6 pm Coef. (z-stat)
Loop trip	−0.05	1.39	0.15	0.38	−0.25	0.41
	(−0.93)	(18.75)	(2.52)	(6.93)	(−2.55)	(6.00)
Distance (miles)	−0.94	−0.62	−0.90	−0.96	−0.60	−0.70
	(−246.02)	(−69.86)	(−222.84)	(−153.41)	(−64.37)	(−82.82)
Percent mixed land use, commercial, office, public at origin	1.32	0.64	1.44	0.59	0.01	2.34
	(70.00)	(20.16)	(73.78)	(26.57)	(0.23)	(75.28)
Percent industrial, transportation, parking at origin	−1.12	−0.74	−0.93	−1.14	−2.04	−0.11
	(−28.56)	(−11.12)	(−23.13)	(−24.28)	(−23.12)	(−1.58)
Percent open space/recreational at origin	−12.08	3.21	−10.51	−8.05	−20.00	−5.35
	(−23.12)	(4.00)	(−19.71)	(−13.16)	(−16.33)	(−6.52)
Percent unknown land use at origin	9.95	−8.62	12.31	−6.89	−41.28	52.02
	(3.99)	(−1.79)	(4.69)	(−2.29)	(−6.54)	(12.44)
Percent mixed land use, commercial, office, public at destination	1.52	0.79	1.64	0.74	3.52	0.67
	(80.13)	(24.87)	(84.01)	(33.44)	(80.08)	(22.68)
Percent industrial, transportation, parking at destination	−0.85	−0.62	−0.66	−0.95	0.96	−1.04
	(−21.62)	(−9.38)	(−16.34)	(−20.26)	(10.64)	(−15.49)
Percent open space/recreational at destination	−10.14	1.29	−9.11	−7.19	−2.78	−18.09
	(−19.53)	(1.58)	(−17.05)	(−11.85)	(−2.56)	(−19.67)
Percent unknown land use at destination	15.56	−8.21	18.11	−4.27	77.13	−16.60
	(6.23)	(−1.72)	(6.94)	(−1.42)	(14.37)	(−3.80)
Origin near subway	−0.10	−0.11	−0.09	−0.13	−0.38	0.00
	(−11.53)	(−6.92)	(−10.37)	(−12.64)	(−21.49)	(0.12)
Destination near subway	−0.10	−0.14	−0.10	−0.13	0.11	−0.13
	(−11.70)	(−8.83)	(−10.98)	(−13.09)	(5.56)	(−9.89)
Along bicycle path	0.76	0.99	0.71	1.01	0.48	0.70
	(24.34)	(21.09)	(22.39)	(31.51)	(8.41)	(17.43)
Constant	2.05	0.75	1.64	2.02	−0.99	−0.49
	(106.28)	(23.49)	(83.12)	(89.27)	(−24.38)	(−16.23)
Inflated—logit model						
Distance (miles)	0.97	1.28	1.00	0.95	1.04	1.01
	(91.76)	(83.89)	(96.74)	(73.96)	(72.26)	(70.94)
Constant	−4.46	−3.35	−4.32	−3.68	−2.62	−3.06
	(−103.07)	(−67.54)	(−106.05)	(−81.14)	(−54.54)	(−65.35)
Dispersion parameter	1.43	2.34	1.45	1.25	3.11	1.65
McFadden's Pseudo R^2	0.090	0.030	0.085	0.076	0.046	0.055
N	214,369	214,369	214,369	214,369	214,369	214,369

Casual and weekend/holiday trips tend to be negatively associated with subway access at both origin and destination points in July and October. This is not the case in the February model. This result suggests that many trips are made independent of the subway system and thus represent new opportunities to travel for users. Weekday trips have a weak association with origin and destination points near subway stations in July, are negative in October, and not statistically significant in February, suggesting no association with subway access. In most cases, regardless of significance levels, the coefficients have relatively small values, suggesting minor associations.

More consistent and stronger associations are seen for the subway proximity variables for AM and PM trips. In the morning, bicycle origin points tend to be negatively associated with subways. Thus, bicycles are being rented where there are no subways. Destination points, on the other hand, have a positive association with subway proximity for AM trips, suggesting that users transfer to a subway and are using the bicycle to access the subway station. The reverse occurs in the evening peak when coefficients tend to be negatively associated with subway proximity at the destination, and positively associated at the origin point, suggesting a return trip home to a location not near subway stations and using a subway for the first part of the trip. However, this result does not show up at the same level of significance in February and October.

Our final variable is for both starting and end points being located along the same bicycle path in Manhattan. In all cases this variable is positive and statistically significant, suggesting that those trips that can be taken on a bike path are more frequent than other trips. While we cannot know the specific route each user took, this is a highly suggestive result.

6. Discussion and conclusions

This analysis has shown the ability to extract insightful results from a "big" dataset lacking any socio-economic components on individual users. One of the key questions is understanding the difference between subscribers versus casual users. Based on the day of the week and time of day that these users make various trips it seems likely that casual users are more likely to use the bicycles for recreational trips, including loop trips.

Linking the data geographically with land use, subway and bike path data reveals how these factors are associated with the frequency with which certain trips are taken. There is a clear pattern of combining peak-hour Citi Bike trips with subway usage, with users accessing subway stations via bikeshare in the morning, and accessing bikeshare via the subway in the evening. Previous work on New York City's bikeshare has found a similar pattern (Noland et al., 2017a). Bikeshare can either complement or substitute for transit trips; our results suggest that the subway is a complement, helping to provide greater access and egress to stations, that is, the "last mile" problem. Another recent study suggested small reductions in bus use in New York City when bikeshare becomes available (Campbell & Brakewood, 2017). Our findings also suggest that most trips are made independent of subway stations; this shows the broad benefits of the increased access offered by the bikeshare system as the network of stations grows.

Bike paths are a major complementary component of any system and when near origin and destination points, they seem to lead to more frequent usage. This is consistent with studies of cycling behavior that suggest bicycle lanes and paths are critical infrastructure for encouraging increased usage (Buehler & Pucher, 2012; Dill, 2009).

Areas with more open and recreational space have the expected pattern; casual users are more likely to rent bicycles from these start locations as they are more likely to be recreational users. Mixed, commercial, office, and public building land uses tend to have positive associations for subscriber, casual, and weekday trips and less so on weekends and holidays. This suggests fewer work and other trips to these locations on non-workdays. AM trips also likely start near residential locations and show less association with these land uses. Industrial, transportation, and parking land uses generally have negative associations.

One of our objectives was to show the feasibility of extracting meaning from the type of trips most frequently made by bikeshare. Without any information on user demographics, the story seems nevertheless compelling; the analysis shows how the bicycles are used for what are likely commuting purposes, recreational trips, connections to subways, and the value of safe bicycle paths along major corridors. The overall mobility benefits offered by Citi Bike are large, providing access to many parts of the city with insufficient subway service.

The opportunities for further exploration of the data remain, particularly in examining gender differences among subscribers, as this is included in the subscriber—but not casual user—data. The data can also be sliced up further to examine more fine-grained elements of specific trips. Seasonal differences may also reflect both temperature and daylight conditions, and this could be further explored. However, despite the benefits shown here of using non-behavioral data, there is still a benefit to gathering survey data on bikeshare users to supplement this type of analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Robert B. Noland http://orcid.org/0000-0003-0775-0624
Michael J. Smart http://orcid.org/0000-0003-3739-2725

References

Antoniou, C., Barceló, J., Breen, M., Bullejos, M., Casas, J., Cipriani, E., … Marzano, V. (2016). Towards a generic benchmarking platform for origin–destination flows estimation/updating algorithms: Design, demonstration and validation. *Transportation Research Part C: Emerging Technologies*, 66, 79–98.

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274–279.

Beecham, R., & Wood, J. (2014). Exploring gendered cycling behaviours within a large-scale behavioural data-set. *Transportation Planning and Technology*, 37(1), 83–97.

Buck, D., & Buehler, R. (2012). Bike lanes and other determinants of capital bikeshare trips. In: *91st Annual Meeting of the Transportation Research Board*, Washington, DC, pp. 12–3539.

Buehler, R., & Pucher, J. (2012). Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation*, 39(2), 409–432.

Campbell, K.B., & Brakewood, C. (2017). Sharing riders: How bikesharing impacts bus ridership in New York City. *Transportation Research Part A: Policy and Practice*, 100, 264–282.

Citi Bike. (2016), Citi Bike System Data, https://www.citibikenyc.com/system-data, accessed Aug 13, 2018.

Citi Bike. (2018). *Pricing, choose your plan,* https://www.citibikenyc.com/pricing, accessed Aug 13, 2018.

Current results, weather and science facts, 2018. (2018, April 7). New York City Temperatures: Averages by Month.

Dill, J. (2009). Bicycling for transportation and health: the role of infrastructure. *Journal of Public Health Policy*, 30, S95–S110.

Dill, J., & Carr, T. (2003). Bicycle commuting and facilities in major US cities: If you build them, commuters will use them. *Transportation Research Record: Journal of the Transportation Research Board*, 1828, 116–123.

Dill, J., & Voros, K. (2007). Factors affecting bicycling demand: Initial survey findings from the Portland, Oregon, region. *Transportation Research Record: Journal of the Transportation Research Board*, 2031, 9–17.

Fischer, M. M., Scherngell, T., & Jansenberger, E. (2006). The geography of knowledge spillovers between high-technology firms in Europe: Evidence from a spatial interaction modeling perspective. *Geographical Analysis*, 38(3), 288–309.

Forsyth, A., & Krizek, K. J. (2010). Promoting walking and bicycling: Assessing the evidence to assist planners. *Built Environment*, 36(4), 429–446.

Garrard, J., Handy, S., & Dill, J. (2012). Women and cycling. In R. Buehler & J. Pucher (Eds.), *City cycling* (pp. 211–234). Cambridge, MA: MIT Press.

Garrard, J., Rose, G., & Lo, S. K. (2008). Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive Medicine*, 4(1), 55–59.

Gebhart, K., & Noland, R. B. (2014). The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, 41(6), 1205–1225.

Guo, J., Bhat, C., & Copperman, R. (2007). Effect of the built environment on motorized and nonmotorized trip making: Substitutive, complementary, or synergistic? *Transportation Research Record: Journal of the Transportation Research Board, 2010*, 1–11.

Lathia, N., Ahmed, S., & Capra, L. (2012). Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies*, 22, 88–102.

Lee, M., & Pace, R. K. (2005). Spatial distribution of retail sales. *The Journal of Real Estate Finance and Economics*, 31(1), 53–69.

Noland, R. B., & Ishaque, M. M. (2006). Smart bicycles in an urban area: Evaluation of a pilot scheme in London. *Journal of Public Transportation*, 9(5), 71–95.

Noland, R. B., Smart, M. J., & Guo, Z. (2017a). Bikeshare trip generation in New York City. *Transportation Research Part A: Policy and Practice*, 94, 164–181.

Noland, R. B., Smart, M. J., & Guo, Z. (2017b). What do people use bikesharing for? Analysis of origin and destination pairs in New York City. In *96th Annual Meeting of the Transportation Research Board*. Washington, DC: Transportation Research Board.

Porojan, A. (2001). Trade flows and spatial effects: The gravity model revisited. *Open Economies Review*, 12(3), 265–280.

Pucher, J., Komanoff, C., & Schimek, P. (1999). Bicycling renaissance in North America?: Recent trends and alternative policies to promote bicycling. *Transportation Research Part A: Policy and Practice*, 33(7–8), 625–654.

Sen, A. & Smith, T. (2012). *Gravity models of spatial interaction behavior*, Advances in Spatial and Network Economics, New York: Springer.

TimebieDaylight Hours of New York (2018). http://www.timebie.com/sun/newyorkny.php, accessed Aug 13, 2018.

Wang, X., Lindsey, G., Schoner, J. E., & Harrison, A. (2015). Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development*, 142(1), 04015001-1 to 04015001-9.

Zhao, J., Wang, J., & Deng, W. (2015). Exploring bikesharing travel time and trip chain by gender and day of the week. *Transportation Research Part C: Emerging Technologies*, 58, 251–264.