

## **Modeling Social Justice Issues in Urban Zones Using Big Data: Evidence from Bikeshare Systems**

### **David A. King, Ph.D.**

Assistant Professor of Urban Planning  
Graduate School of Architecture, Planning and Preservation  
400 Avery Hall  
MC 0356  
1172 Amsterdam Avenue  
Columbia University  
New York, NY 10027  
[dk2475@columbia.edu](mailto:dk2475@columbia.edu)

### **Jonathan R. Peters, Ph.D.** (corresponding author)

Professor of Finance  
The College of Staten Island  
and  
Research Fellow,  
University Transportation Research Center – Region II  
The City College of New York

Room 3N-207  
2800 Victory Blvd.  
Staten Island, New York, 10314  
Phone: 718-982-2958  
Fax: 718-982-2965  
Email: [jonathan.peters@csi.cuny.edu](mailto:jonathan.peters@csi.cuny.edu)

### **Adam Davidson, M.C.P.**

Ph.D. Program in Earth and Environmental Science  
The City University of New York  
365 Fifth Avenue  
New York, New York 10016

### **Juan Francisco Saldarriaga**

Associate Research Scholar  
Adjunct Assistant Professor of Urban Planning  
[Spatial Information Design Lab](#)  
GSAPP | Columbia University  
(212) 854-2783

## **Abstract**

Cities across the United States are developing and installing bike share systems to improve access to cycling. The largest system in the country is in New York City, with almost 6,000 bikes across more than 300 stations, but most systems are a few hundred cycles sprinkled throughout a limited number of docking stations. Though modern bike share represents a new transportation technology, these systems quickly became fixtures in the cities where they operate. Despite their popularity, these programs require municipal support, which may include direct financial support, allocation of street spaces for docking stations, or other types of direct and in-kind assistance. The public involvement raises concerns about how the systems are designed and who has access to the bikes. More succinctly, do publicly supported bike share systems adequately serve a city's population?

We explore the distribution of bike share in three ways. First, we examine the spatial networks of bike share systems in New York City, Boston, Chicago, Washington, D.C. and California's Bay Area. We use GINI coefficients to show that the socio-demographic characteristics in areas served by bike share networks are similar to those areas served by public rail transit networks. Second, we use demographic splits to show how New York's Citibike program tilts heavily toward male users, but the fee structure is somewhat biased against women. Third, we proposed a credit-based system rebalancing program that both reduces rebalancing costs and improves gender equity. Overall this research demonstrates how new ("Big") data can be incorporated with more traditional form of analysis to both evaluate and improve the equitable distribution of publicly

available assets. We conclude with a discussion of the implications of this work for subsidy policy and local transportation planning.

## **Introduction**

The recent success of bike share systems is part of a larger trend toward multimodal transportation systems in cities. City officials and planners focus on public health benefits of active transportation, the low cost of access relative to other modes of travel, the imperative to promote sustainable transportation policies, and the increasing size of bicycle lane networks[1, 2]. These trends help shift transport planning away from auto-orientation and toward transit, walking, cycling and other services.

However, there have been limited studies of effect of bikeshare programs on equity needs. A 2013 evaluation of Philadelphia's 2010 Bikeshare Concept Study focused on the potential public health impacts of a system that still has not been launched [3]. The research noted high concentrations of asthma, obesity, heart disease, and other health concerns when viewed by resident geography, but there are only small overlaps of proposed bike share stations with neighborhoods with the poorest health. The author concludes that "the program's planners must take additional steps to ensure that those who are most likely to develop the health conditions against which biking can offer protection have access to the program"[3].

Early evidence suggests that planning bike share with an eye toward public health has either not been taken or do not provide an appreciable result. A survey of Washington, D.C.'s Capital Bikeshare users showed that riders were more likely to be

young, white, male, well educated, and only slightly less affluent than the regional average [4]. In a Masters thesis that focused on modeling the determinants of usage for Capital Bikeshare, Daddio found a similar correlation by examining the demographic makeup of a station's catchment area relative to trip departures per station. His most significant results suggest that nearly 80% of the variation in ridership can be explained by five variables: the positive effects of being proximate to a young population (age 20 – 39), proximity to a Metro station and a retail area; and the negative effects of being distant from the system center and proximate to a non-white population [5]. In short, Downing's conclusion is that the communities with the most to gain are the ones that Daddio finds use the system least.

Other evidence suggests that bike share systems can act as complements to conventional transit services and substitutes for some auto trips [6].

#### Data

Bike share systems in the United States use similar technologies, often from the same suppliers, and generally are supportive of open data standards [7-11]. Overall, bike share data used comes from two sources: historical data on individual trips and stations provided through each company's website and live feeds displaying individual station data. Not all of the systems use both sources. While some of them update and offer their data permanently, others, such as Divvy and Bay Area Bike Share have only made their data available in the context of data visualization challenges open to the public. This being said, the availability of the data is very good and as it always comes with geographic coordinates for the bike share stations, it is very simple to place them on a map using ArcGIS software.

Below we describe data used for each system.

- CitiBike (New York) provides both individual trips and a live feed updated every two minutes in json format. The location of the stations can be derived from both, although the live feed includes stations that might not be in use. With the individual trips file, you can use only the stations that have actually been active during that month. The data is for the period between July 2013 and May 2014.

- Divvy (Chicago) made their trip and station data for a data visualization challenge earlier this year. The data is for 2013 and it includes separate csv files for individual trips and stations and a GIS shapefile with the stations.

- Hubway (Boston) also made its data available in the context of a data visualization challenge in 2012. However, that dataset has been updated and it now includes individual trip data from October 2012 to November 2013. The dataset also includes a file with individual stations. All files come in csv format.

- Bay Area Bike Share (Bay Area) also made its data available in the context of a visualization challenge in early 2014. The dataset includes individual trip data from August 2013 to February 2014, as well as individual station data and rebalancing data, all in csv format.

- Capital Bikeshare (Washington D.C.) provides their individual trip data on their website

as well as an xml station status feed. The trip data includes individual rides from September 2010 through March 2014.

## Overview of the Bikeshare Networks

In this section we focus on five systems currently operating in the United States. Figures 1-5 show the spatial distribution of docking stations for New York City, Chicago, the San Francisco Bay Area, Boston and Washington, D.C. Each city's system shows distinct spatial characteristics of coverage, network density and connectivity. New York's network of docking stations is very dense, where most stations are within a few blocks of another. Boston's and Chicago's systems feature stations much more spread out than New York's, but both are contiguous. Washington and the Bay Area have spatially dispersed stations and feature spatial discontinuity where clusters of stations operate somewhat independently.

The network design of bike share systems has implications for usage, rebalancing issues, and equity. Nodes and stations within bike share networks are deliberately placed by urban planners and local officials. CitiBike in New York, in conjunction with the city's Department of Transportation, encouraged the public to submit station requests through an online portal [12]. Yet despite the enthusiasm of the crowd across the city, a dense network centered on lower Manhattan and Brooklyn was implemented and most of the proposed station areas were left without. While most of the city is far away from the bikes, the distribution closely aligns with wealthier parts of the city. In areas of Brooklyn to the east of the shaded area on the map the bike share stations do not serve lower income neighborhoods. The deliberate spatial distribution raises concerns that the bike share system is a new alternative for wealthier New Yorkers. While such a statement is factually true—CitiBike is distributed in wealthier parts of the city—the policy implications are ambiguous.



# NEW YORK - CITIBIKE

AREA WITHIN HALF A MILE OF CITIBIKE STATIONS

- CITIBIKE STATION
- //// HALF A MILE RADIUS

0 1 2 3 4 5 Miles  
0 1 2 3 4 5 Km

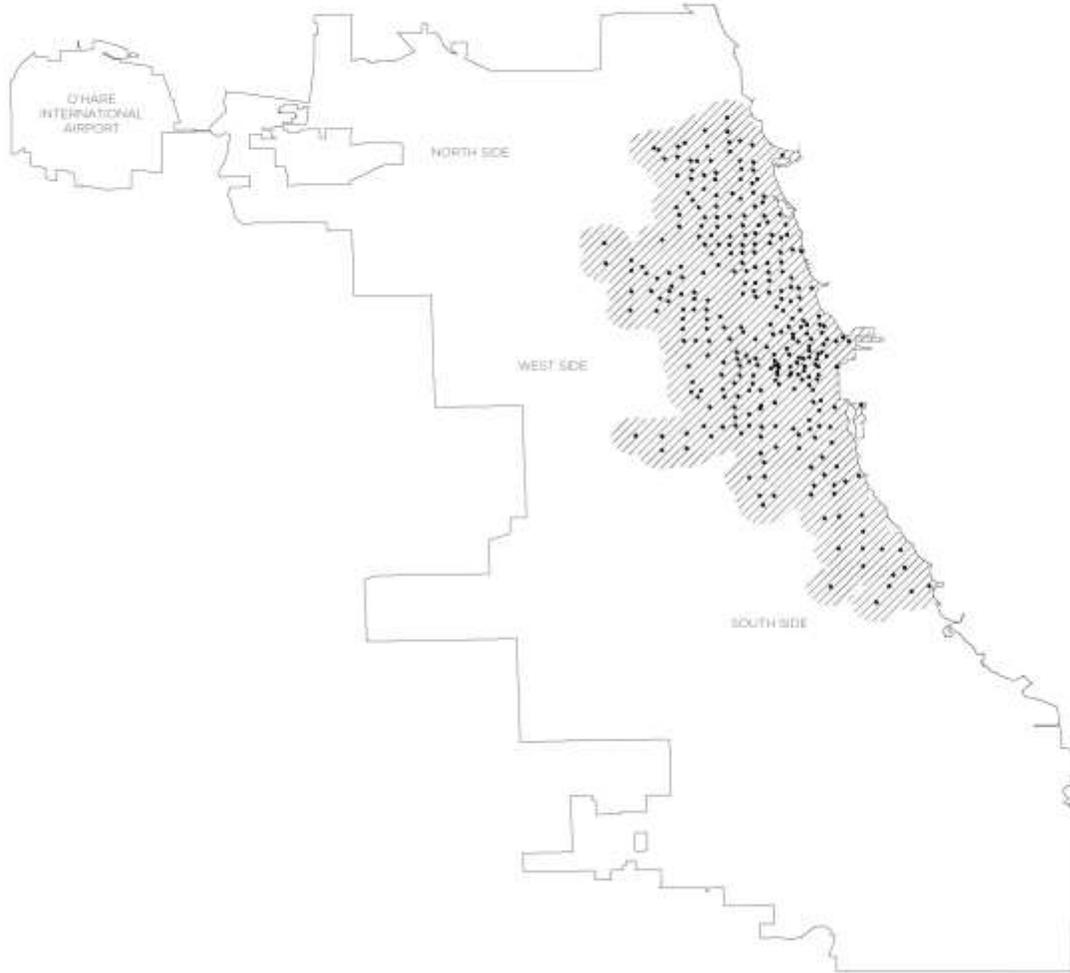


# CHICAGO - DIVVY

AREA WITHIN HALF A MILE OF DIVVY STATIONS

- DIVVY STATION
- //// HALF A MILE RADIUS

0 2.5 5 Miles

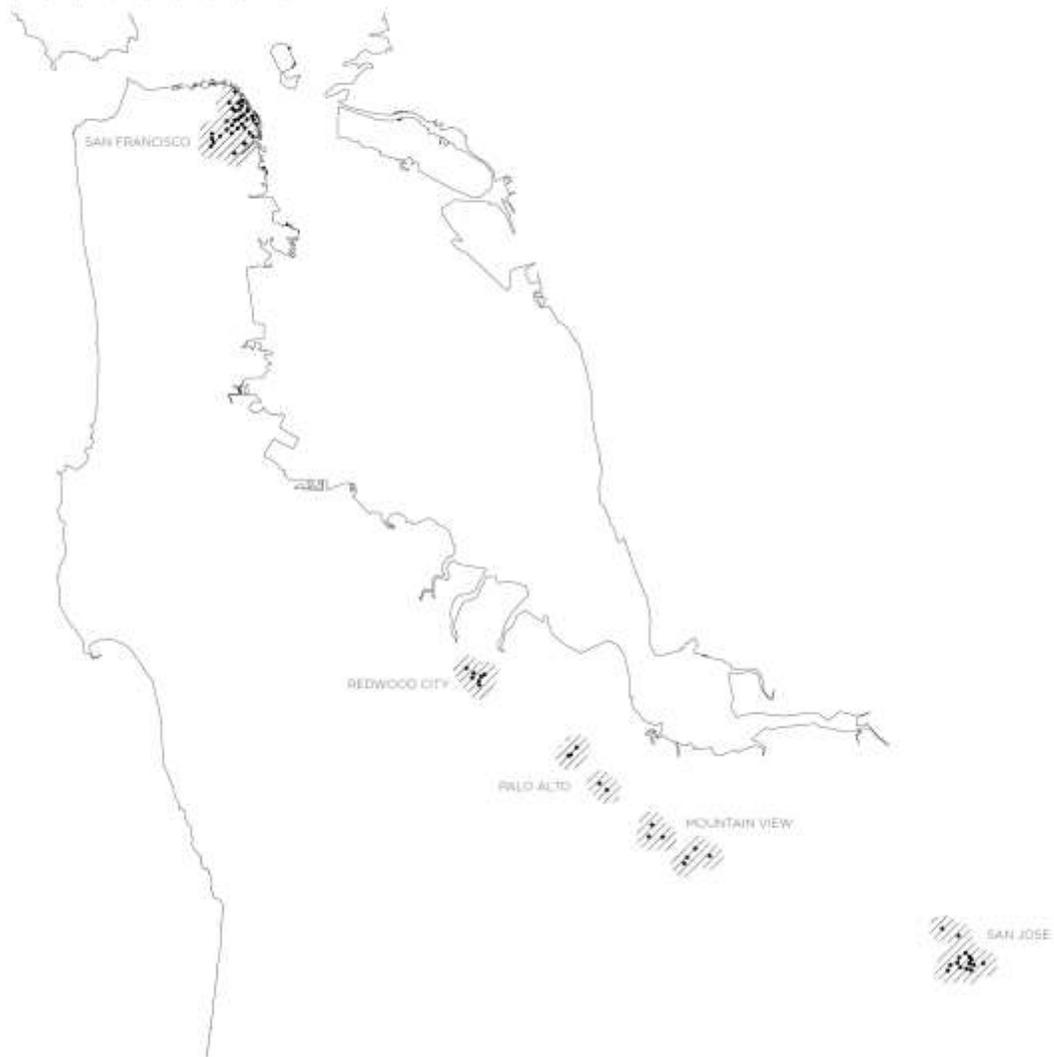


# BAY AREA - BIKE SHARE

AREA WITHIN HALF A MILE OF BAY AREA BIKE SHARE STATIONS

- BAY AREA BIKE SHARE STATION
- //// HALF A MILE RADIUS

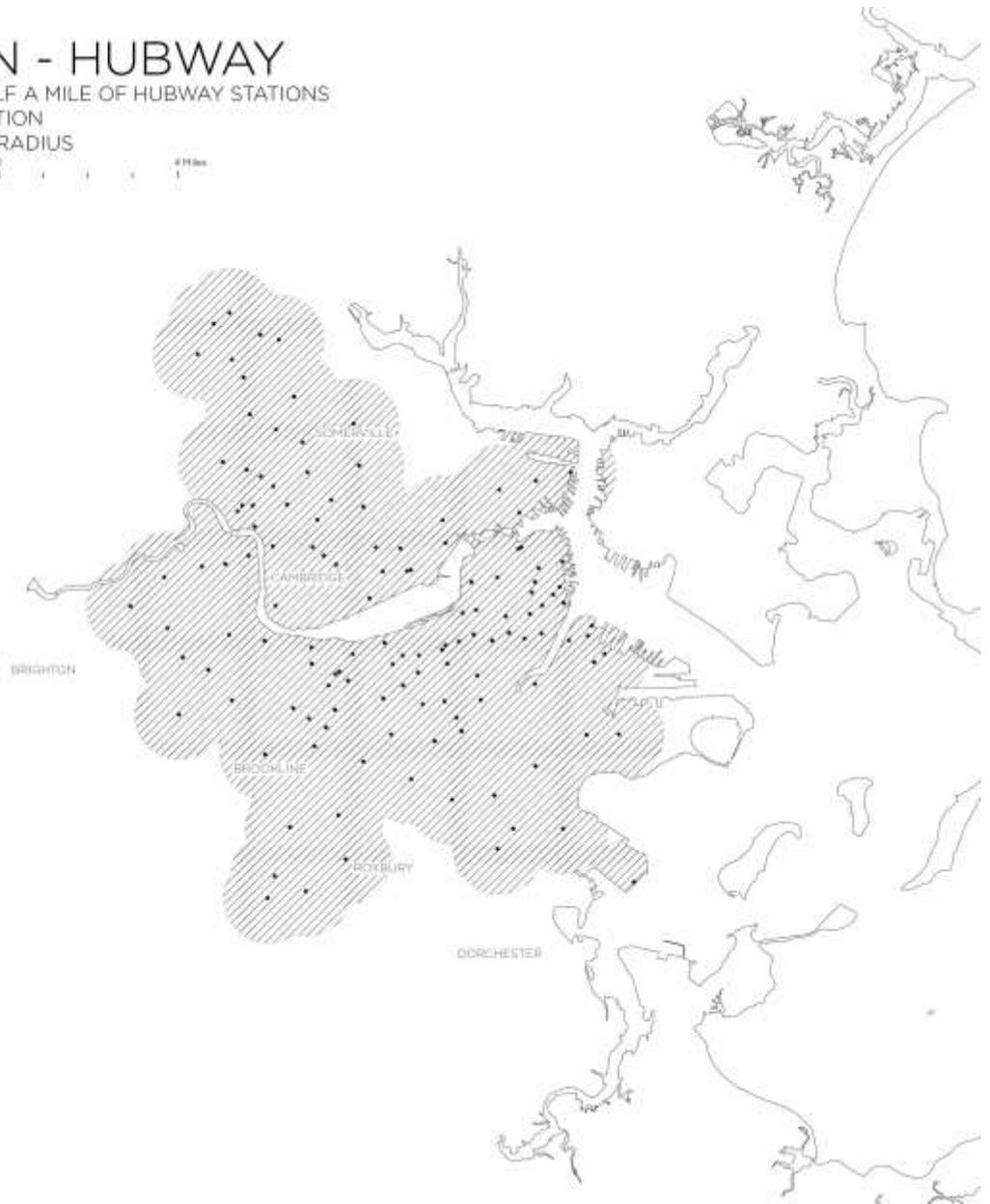
0 2.5 5 10 Miles



# BOSTON - HUBWAY

AREA WITHIN HALF A MILE OF HUBWAY STATIONS

- HUBWAY STATION
- //// HALF A MILE RADIUS



# WASHINGTON D.C. - CAPITAL BIKESHARE

AREA WITHIN HALF A MILE OF CAPITAL BIKESHARE STATIONS

• CAPITAL BIKESHARE STATION

//// HALF A MILE RADIUS

0 1.25 2.5 5 Miles



Overall bikeshare networks are centered on central business districts rather than residential neighborhoods. However, in some instances the networks are more proximate to non-whites than the more extensive rail system. We use U.S. Census data from the American Community Survey and the Longitudinal Employer-Household Dynamic (LEHD) to examine the spatial relations between bike share, households and employment. To keep comparisons consistent, Boston is also left out of some summary

data. As complete metropolitan areas these regions (excluding Boston) contain more than 40.5 million people and nearly 17 million primary jobs.<sup>1</sup>

The strength of using the Census and LEHD data is that we can compare & contrast residential and worker populations, thus giving a more complete picture of who has potential geographic access to transportation infrastructure throughout the course of their day. However the datasets provide some general, but not exact comparisons.

Detailed census data is mostly available at the tract level and above, while LEHD data is based on the census block group of the employer. Census data is on the whole more complete, providing detailed income, age, and racial data. LEHD data uses broader bins to group income, age, and race than the census. For example LEHD income is grouped into counts of three categories: “Below \$1,250 / month”, “Between \$1,251 - \$3,333 / month”, and “Above \$3,333 / month.” Census data provides median income by tract, as well as counts at much finer and longer gradations.

With these caveats in mind the authors looked at background population and employment characteristics of areas within a half mile of the bikeshare systems and then compared that to areas within a half mile of metro or subway style rail transit. This comparison is based on the common claim that bikeshare serves as a complement to transit, since it helps with the ‘last mile’ issue, short trips and circulation. For these reasons, areas with only commuter rail were not included since it serves longer distances and nearly exclusively the work trip. Likewise, areas only near bus service were also not included as many bus lines are perceived to suffer from speed, frequency, and reliability and thus may not be perceived as complements or substitutes to bike or rail transit by

---

<sup>1</sup> The state of Massachusetts did not submit LEHD encoded data, therefore it cannot be included in employment tabulations.

many potential riders. The resulting figures were then normalized as a percent of the total population category within the MSA.

Across selected regions (excluding Boston) rail is within a half mile of 18% of the residential population. However, if we also drop New York we bring that average to 11.5% (if we then add in Boston the number is only slightly raised to 12%). By comparison bikeshare is only accessible to about 6% of the residential population, but that ranges from a low of 3.6% of metropolitan San Franciscans, to a high of 13% of metropolitan Washingtonians. For Washington, that is actually greater coverage than the rail system provides. However, these numbers are low when compared to the number of jobs accessible to rail and bikeshare. Averaging the five regions where job numbers are available, we find that 33% of jobs are within a half mile of rail compared to 25% for bikeshare. In Washington and Chicago, the numbers are nearly even with bikeshare trailing rail by just a couple of points.

Table 1:

	BOS*	CHI	DC	NYC	SF	SJO	Total
AllPop (100%)	4,552,402	9,461,105	5,582,170	18,897,109	4,335,391	1,836,911	40,112,686
AllJobs (100%)		3,936,244	2,644,650	7,677,972	1,849,481	815,753	16,924,100
Pop_NonWhite	964,862	3,277,224	2,523,350	7,719,445	2,095,872	965,114	16,581,005
Pct_Pop_NonWhite	21%	35%	45%	41%	48%	53%	41%
Job_NonWhite		837,339	808,267	2,245,480	648,368	306,854	4,846,308
Pct_Job_NonWhite		21%	31%	29%	35%	38%	29%
Pop_Young	1,253,403	2,649,924	1,642,101	5,266,518	1,254,495	537,456	11,350,494
Pct_Pop_Young	28%	28%	29%	28%	29%	29%	28%
Job_Young		879,612	604,973	1,672,195	365,635	152,054	3,674,469
Pct_Job_Young		22%	23%	22%	20%	19%	22%

\*Does not have LEHD employment data, so is consequently left out of the Summary column

Table 2: Size of Systems

	Square Mileage Comparison					Bikeshare Size		
	Metro	Rail	Bikeshare	Rail / Metro	Bikeshare / Rail	Bikeshare stations	Bikes	Docks
BOS	3487	43.1	25.5	0.012	0.591	131	1300	2287
CHI	7197	72	44.1	0.01	0.613	300	3000	5040
DC	5598	52.3	71.4	0.009	1.363	321	2500	5290
NYC	6687	120	16.8	0.018	0.14	330	6000	11300
SF	2471	43.4	6.2	0.018	0.144	42	700	782
SJO	2679	34.6	8.5	0.013	0.246	28		456
Summary	28119	365.4	172.5	0.013	0.472	1152	13500	25155

### *Evaluating Equity with GINI Coefficients*

Work by Gordon and Peters address the issue of incongruity of users. In examining survey data of tolled bridges in the New York City region, they find that facilities with alternatives (good mass transit or free nearby routes) show higher levels of income inequality than their surrounding area, while facilities with few alternatives better reflect the background demographics of the area that they are situated in [13]. Thus facilities that are either inherently accessible or have few alternatives will share the demographic characteristics of their surroundings. Applied to bikeshare, we would not expect bikeshare to reach lower income or underrepresented groups if it is not first made geographically accessible to them. This adds an additional tension to Daddio's findings by indicating that there is a tension between maximizing ridership and social equity.

Gordon and Peters explored the use of geographic information systems as a tool to use for the evaluation of social equity. They estimated the relative burden of tolling on various user groups and estimated the general markets for transportation services. Using Lorenz curves and GINI coefficients, Gordon and Peters are able to estimate the income profiles of the general population that are geographically located near the toll facilities and the actual users of particular toll bridge facilities.

Likewise, the authors propose to estimate the likely users of bikeshare – that is, the population that is proximate to the bikeshare stations and following Daddio's findings, we expect that proximity to the system and centrality to the geographic center of the bikeshare network represents the users who have the best opportunity to utilize the bikeshare system. The authors pulled the Census 2010 data from the existing bikeshare stations in the four study cities and estimated the relative income distribution of the bikeshare systems in New York City.

Location	Buffer	Gini
DC	Capital Bikeshare	0.4535
DC	Bus Stops	0.4326
DC	Metro Stations	0.4443
Chicago	Bikeshare	0.4949
Chicago	Bus Stops	0.4809
Chicago	Rail Stations	0.5020
Boston	Bikeshare	0.5028
Boston	MBTA	0.4965

### Revenue Generation from Various Fees – Social Equity Conditions

New York’s Citibike System has been the subject of considerable interest regarding the financial needs of the system. Given that the New York system is not subsidized in terms of direct system operational support, the financial need to create enough revenue to cover existing costs has been a matter of major concern. Recently, the system has reported an ongoing operational loss. One consideration is the ability to focus additional fees for service to address the need for revenue and consider options as to how to raise revenue from the various classes of users – while avoiding any negative impacts on the social equity of the system. As highlighted above, the system network have in some cases a more limited footprint in terms of low income users as compared to the general metropolitan area (Washington DC excepted).

Social Justice analysis is mandated by Executive Order 12898 of 1998 and by Title 6 of the Civil Rights Act of 1964. These federal requirements are intended to assure that disadvantages classes are not discriminated against in terms of various government services. In the transportation sector, these rules are general understood to require that agencies address social justice concerns in the planning and implementation of transportation systems. Given that bikeshare systems are deployed under municipal contract, utilize public road space for stations and in many cases receive public subsidies as stated above, it is clear that social justice analysis is appropriate and may well be mandated by federal authorities. As such, consideration of how a change in rates should be applied or what aspect of costs subsidized should be evaluated for social justice issues.

In particular, the key metrics of analysis of social equity analysis are as follows:

- 1) Analyze the needs of protected classes
- 2) Included in the transportation planning process
- 3) Protected classes are not overly burdened with the cost of services
- 4) Protected classes are not denied the benefit of transportation services

In terms of protected classes, one generally considers the following as protected classes:

- 1) Low income individuals
- 2) Minority individuals
- 3) Disabled individuals

- 4) Senior Citizens
- 5) Rural populations

In addition, the impact on gender equality needs to be considered as we discuss changing various aspects of the system. Given the existing demographic data collection on annual users (subscribers) as opposed to daily and weekly users (customers), we can evaluate the existing subscribers in terms of gender and age issues. The authors developed a detailed record by record pricing algorithm to price the individual rides on the Citibike system in terms of overtime fees. We also utilized the demographic information and reported subscriber rates by gender to understand the fee structure aspects of bikeshare use. We evaluated 5,561,840 rides from June 2013 to February 2014. In addition, we also explored the cost of annual membership fee.

New York’s bikeshare system has benefited from high usage and enrollment as compared to other systems. However, the data indicates that a large amount of rides for subscribers are provided at no additional cost to the users other than the annual fee. Table 2 provides an overview of the revenue and usage components by gender for the trips reported from the system data.

The data reports a considerable amount of variation in usage pattern and fee structure by gender. Women represent 38% of Citibike subscribers but only 23.2% of rides. This strong imbalance provides one with an argument that additional fees are better applied to a per ride costs (say a flat fee of \$.25 per ride) as opposed to increases in the annual subscription fee. Further, additional subsidies are best applied to lowering the annual fee as opposed to general subsidy on all rides.

As a second consideration, one could consider increasing the overtime fee for rides that exceed the 45 minute basic ride limit. Examining the mix of overtime fee payments, we find that women pay 30.3% of the overtime fees – as they have a greater frequency of overtime rides with 1.17% of rides by women having an overtime fee as compared to .79% of rides by men. Given this disparity, increases in overtime fees also appears to have a significantly disparate impact on female users of this bikeshare system.

Taken as a whole, it appears due to variation in usage patterns that changes in the fee structure of the New York Citibike system should be evaluated for social justice impacts in terms of gender. If additional information was collected on racial and income data from subscribers one could further evaluate additional environmental justice issues.

Table 2 – New York City Bikeshare Data by Gender

	Subscriber	Rides	OT Fee	Rebalance Net
Subscribers - Male	62%	76.8%	69.7%	94%
Subscribers - Female	38%	23.2%	30.3%	6%

A further matter of considering additional fees follows as we examine the potential of using the pricing system to manage the distribution of bicycles in the system.

### **System Rebalancing**

One aspect that can be informed by system performance metrics is process and operational improvements. Bikeshare systems, while deployed for a considerable amount of time in a number of cities, still has a number of areas of operation practice that could benefit from additional analysis and exploration of usage patterns. One area of interest is the problem of system rebalancing. It appears to be a rather common issue that bikeshare system can suffer from bicycle shortages and surplus conditions in various stations at different times during the day and on specific days of the week. If we examine these patterns of usage, we find that these patterns appear to be stable and occur repeatedly at key stations in the network.

These oversupply or shortage conditions potentially can restrict system usage and also can create difficulties for users who may want to check in equipment at overloaded stations. The current solution to these problems is to physically redeploy the bicycles from surplus stations to shortage stations. Yet information on this aspect of bikeshare systems is difficult to examine. It is known through anecdotal comments that systems apparently have these problems, but the reporting on the amount of rebalancing needed for a given system is generally not reported in existing data sources. Dantos (2012) produced some analysis that indicated a greater propensity for users to move downhill as opposed to uphill from Capital Bikeshare stations. His work indicates a greater outflow from stations in Northwest Washington (an uphill area) as compared to other regions. The net result of this is a need to physically redistribute the bikes as needed to address this asymmetrical demand and supply conditions. This is similar in practice to the need to deadhead buses or trains in a transit system to redeploy assets as needed to address system load condition.

The scale and magnitude of these redeployments may be the subject of internal consideration for bikeshare operators, but public discussion and policy aspects appear to be understudied. Clearly, the need for excessive manual movements by trucks of bikeshare equipment impact the carbon footprint of the system and the cost of operations. Yet some information is available and provides tantalizing insight into the rebalancing issue. Recently, Alta, the CitiBike contract provider was subject to a lawsuit for back wages for rebalancing drivers and mechanics. As reported in the *Oregonian*, one employee stated that "A large scale bike-share program doesn't work without a fleet of trucks to keep the inventory balanced,". Further analysis of this problem is clearly merited. In most systems, the additional bike movements caused by rebalancing is scrubbed out of the data that is reported for public use.

The authors examined this question utilizing the existing volume data and developed metrics of oversupply and shortage by station for the Citibike System. As reported by

others – such as Dantos – many stations had reasonable balance in terms of bike inflows and outflows. These stations required little rebalancing on a daily basis and as such incurred less operational costs for this aspect of system. We examined all 330 station in the CitiBike system and located a number of stations that had significant imbalances. It is interesting to note that while many of the inbalance hours occurred during the morning and afternoon peak, some inbalance conditions existing for extended periods of time. In addition, during the same period some stations exhibited surplus conditions and others shortage conditions. As such, these conditions appear to be linked to various user needs and the regional job and home relationships.

One method to deal with this structural imbalance is to alter the price of dropping off or checking out a bike based upon the general daily pattern of usage by station. In our first example, we apply a fee of \$1.00 to take a bicycle out of a station during a period of shortage. Further, we apply a credit of \$1.00 for checking a bicycle into the same station during the same period – thereby increasing the supply of bikes at a deficit station. Correspondingly, we apply a credit of \$1.00 to user accounts if they check a bike out of a station during a surplus period and a fee of \$1.00 for checking in a bike at a surplus station during surplus demand hours.

The authors developed a pricing algorithm that reviewed the trip records and applied the fee or credit to three test stations which have very strong imbalance problems. The three stations have varying patterns of demand, with one having a surplus condition in the AM Peak (6:00 to 10:00 AM) and deficits conditions in the afternoon/evening. The other two stations have deficits in the AM peak and surplus conditions in the afternoon/evening. There were 140,539 trips that had a fee or credit applied producing 44,830 dollars in credits and 87,495 in fees for a net subscriber cost of \$42,665 in net revenue over 7 months. This revenue was generated in a very asymmetrical way in terms of gender, with 94% of the net revenue generated from male subscribers and 6% from female subscribers.

The response of users to these applied fees and credits is yet to be determined – however the application of these fees could be tested in a field trial. One significant benefit of the continuous collection of usage data is that we can alter the pricing pattern in a stepwise fashion, until the desired outcome is obtained. If users are responsive to fees, then the imbalances are solved without cost to user – as their behavior change in using the bikes would reduce the physical imbalances in the system. If users do not respond to the fee structure in terms of changing demand, then the cost of rebalancing would be applied to the users that as a collective create the structural imbalance. Users with cheap patterns of travel (using from and to stations without an imbalance) would be excused from any additional cost – and in our test case, that was women to a high degree.

## **Conclusions**

Large scale data offers the policy analyst the opportunity to fine tune the practices and standards of modern transportation systems in close to real time. In addition, the continuous collection of data allows the policy makers to adjust in a more fine grained fashion and progressively address system challenges through a number of minor actions as opposed to gross actions.

Today, if we compare our road or transit system data as compared to bikeshare systems, it is obvious that the bikeshare systems are feeding back useful and instructive data on usage and fee payment. In stark contrast, our road and transit systems provide little trip level data that has any demographic or social data that would allow policy makers to tailor their solutions to promote environmental justice. Our results to date indicate that there are significant environmental justice issues with respect to gender and income in the bikeshare systems studied. As such, we encourage policy makers to examine this detailed data on a regular basis to address the needs of protected classes.