Bike sharing, bike infrastructure and congestion - Evidence from NYC's Citibike

Stefanie Braun, Stephen Rouse, and Hendrik Wolff

November 4, 2018

1 Introduction

Despite the widespread introduction of bike sharing programs in European cities over the last two decades, bike sharing systems have only recently started in North American cities. For instance, over the last few years, cities such as Washington D.C., Boston, San Francisco, Vancouver and New York City have all launched their own public ride share programs. Due to this large uptake in the popularity of bike sharing systems, the interest in examining their policy implications in the context of North American cities has been reignited. Specifically, there is significant interest in the environmental benefits they may induce in concentrated urban areas. While these biking numbers have dramatically increased in recent years, we want to better understand what factors are driving the changes. More specifically, we want to know if commuters are switching their modes of transportation away from private car transportation, and therefore reducing congestion and pollution as a result (Hamilton and Wichman, 2018).

The greenhouse gas emissions of the transportation sector accounted for 28.5 percent of the total U.S greenhouse gas emissions in 2016. Notably, between 1990 and 2016 the amount of miles traveled by motor vehicles increased by around 45%, induced by a number of factors: population growth, urban sprawl and economic growth.[1] The problems that congestion and air pollution create can become even more intensified in highly congested urban areas. Aside from the health concerns that arise out of increased levels of air pollution, congestion can lead to increased commuting costs and travel times (Schrank et al., 2015). We aim to investigate how bike sharing systems can improve the accessibility of sustainable transportation options in urban areas. For this, we are primarily interested in whether bike sharing can curtail the growing problems that highly polluted and congested cities face (Midgley, 2011). To identify the causality of our main concern, we will try to answer two questions: i) Does the implementation of a bike sharing program induce people to switch to "greener" modes of transportation, i.e. biking and bike sharing? ii) Might such a bike sharing program, which possibly encourages people to bike instead of drive, be more successful within an improved bike infrastructure network? To assess these questions and identify the precise causal effect of bike sharing on traffic congestion, a rigorous identification strategy must be proposed.

We begin our approach with an analysis of the transportation modes around the East River Bridges to and from Manhattan. To complement this mode share analysis, we leverage data on whether bike sharing stations, in combination with proper bike infrastructure, can induce individuals (at least during the warmer months of the year) to switch from driving a car or even from taking public transportation to a "greener" means of transport, i.e. biking. As noted in the New York City Department of Transportation's Mobility

[1] https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

Report (Mobility Report, June 2018), the number of daily cycling trips have increased from roughly two hundred fifty thousand bicycle trips a day in 2010 to four hundred and sixty thousand by 2016.

In the next step of the paper we aim to identify the causal effect between the installation and modification of bike infrastructure and the popularity of biking in Manhattan. To capture the popularity of biking in Manhattan, we will leverage the open data from New York City's bike sharing program, Citibike. This causal analysis will allow us to better understand the importance of having an effective bike infrastructure network that exists in conjunction with a bike sharing program. Such a partnership can be critical as a tool to help reduce congestion and therefore air pollution within highly crowded urban areas.

However, the primary focus of this paper will be on the impact of bicycle infrastructure on motor vehicle traffic induced congestion in New York City (NYC). We will support this analysis of bicycle-induced changes in congestion with our previous analysis of how the introduction of new Citibike (bike share) stations could reduce congestion. NYC is a great case study for the impact of bike sharing on congestion in urban areas, for three major reasons. First, according to Schrank et al. (2015), NYC ranks first in terms of total congestion costs among U.S. cities in 2014.[2] Second, the city has both an effective public transport system and an expansive bicycle lane network. The former is especially important in the sense that bike sharing is often seen as augmenting the 'last mile' of commuters' trips, in combination with public transportation (Midgley, 2011). Further, NYC is home to the largest network of bike routes in North America. Bicyclists can travel on over 1,000 miles of bike lanes.[3] Lastly, biking is a popular way of commuting in this city. The number of New York residents who ride a bike several times a month increased by 30% between 2011 and 2016. During those years daily cycling increased by 70%, with daily Citi Bike ridership increasing by 16% between 2016-2017.[4]

Similar to Hamilton and Wichman (2018), we seek to identify the causal effect of bike sharing infrastructure (Citibike) on motor vehicle induced congestion. More specifically, we focus on two major expansions of the Citibike bike sharing program and examine the effect of bike station locations on traffic congestion. We investigate the effect of two separate expansions of both the number of bike stations and bikes in Manhattan during August 2015 and August 2016. The discrete jumps in the number of bike stations at each roll-out (approximately 100 more stations for each, see Figure 3) and the fact that the new stations are spread evenly throughout Manhattan allow us to identify the changes in congestion within a Census block (group) over time. We spatially and temporally match those bike share stations with traffic data on Census block and Census block group level throughout Manhattan.

Empirically, we aim to capture the effect of bike share programs on traffic congestion using a fixed effects framework, in order to control for time-invariant unobservable variables within a spatially defined area over time. Moreover, we test for whether an intensified bike infrastructure within a Census block (group) increases the effect of congestion on bike sharing. While we find the presence of a bike station decreases congestion in both the Census block and Census block group levels, an additional bike station added to the group level actually marginally increases congestion. Since the impact of bike infrastructure on congestion may vary across different levels of congestion, we conduct quantile regressions to test for whether blocks with markedly different congestion levels respond differently to the introduction of stations. We conclude that especially highly congested blocks and block groups benefit even more from the presence of bike share stations. To investigate whether the popularity of certain bike stations further reduces congestion levels, we account for the number of arriving rides to each station.

Before moving on to the methodology, the next section gives an overview of the NYC Citibike program

[2]"Congestion Cost - Value of delay and fuel cost (estimated at $17.67 per hour of person travel, $94.04 per hour of truck time and state average fuel cost)" Schrank et al. (2015)

[3]http://www.nyc.gov/html/dot/html/bicyclists/bicyclists.shtml

[4]http://www.nyc.gov/html/dot/html/bicyclists/cyclinginthecity.shtml

and the operating principles on which bike sharing systems tend to operate in general. After that, we review related literature. The subsequent sections then introduce our empirical approach and outline our results, including several robustness checks. The final section provides some important points to discuss and consider moving forward. We moved a detailed description of our various data sources to the appendix.

2 Background on Citibike

First announced in 2011, the NYC Citibike program was launched in May 2013, starting with 6000 bikes docked across several hundred stations in Manhattan and Brooklyn.[5]
Figure 3 depicts the growth in ridership and bike share stations over time. The increase in the number of stations and number of trips reflects the overall increase in popularity of biking as a transportation mode in NYC between May 2013 and August 2017. Between August 2015 and 2016 Citibike expanded extensively by 280 additional stations and 4000 new bikes and in the course of this spreads further out throughout the city. A further expansion followed in September 2017 by 142 additional stations and 2000 further bikes.[3] Apart from extending the bike sharing program the New York government enhanced the city's on-street bike network by around 330 miles during the last five years.[4]

Identical to modern bike sharing systems, Citi Bike allows members to pick up and return bikes at docking stations. When a member has signed up/ purchased a particular plan (single ride, day pass or annual membership) it is able to unlock bikes at any station with a personal key. Typically, short rides, of up to 30-45 minutes are free. However, overage fees are usually charged for longer trips. [6]

The pricing strategy suggests that bike sharing programs are predominantly meant to provide an alternative to common means of transport such as motor vehicles or walking for short and medium distances. That said, the major objective is to increase mobility, reduce transportation and transportation infrastructure cost, as well as traffic congestion and fuel use. A further goal is to increase the use of public transport. With regards to this, bike sharing systems are meant to complete existing public transport by linking major routes and providing an alternative to walking or driving to important public transit stations (Martin and Shaheen, 2014; Pucher and Buehler, 2005; Shaheen et al., 2010).

3 Literature Review

Broadly defined, our paper adds to the economics literature on transportation. Recently there has been an abundance of research on policy debates concerning the optimal mechanisms to reduce environmental externalities caused by motor vehicle induced congestion. These studies typically focus on studying changes in the behavior of drivers or on the implications of public transport investments. Hamilton and Wichman (2018) are the first that examine the impact of bike sharing infrastructure as a new and more cost-effective transportation alternative and a possible supplement to increase the effectiveness of public transit. More specifically, they study the effect of bike sharing infrastructure on congestion.

From a big picture perspective, recent literature can be split into literature (i) that analyses policy impact on the consumer's decision making process within the existing transportation network and (ii) impacts of investment in new/additional means of transportation on the behavior of consumers.

The former mostly studies the impact of gasoline price changes and optimal taxation policies of gasoline on motor vehicle driver's behavior. The impact of an increase in U.S. gasoline taxes, i.e. higher gasoline prices on gasoline consumption is for example analysed by Bento et al. (2009) and West and Williams (2005).

[5]https://www.citibikenyc.com/about
[6]see for example: https://www.citibikenyc.com/pricing, https://www.mobibikes.ca/en/offers-subscription

Those studies often find relatively inelastic price elasticities of demand. Changes in gas prices are studied by Currie and Phung (2007) and Spiller and Stephens (2012) to see if consumers opt to substitute private car transportation for public transportation when gas prices rise. Both find evidence that public transit demand increases after a rise in gas prices. Burger and Kaffine (2009) find little evidence for speed reduction on Los Angeles freeway routes due to a rise of gasoline prices. This suggests that households do not typically make small adjustments in their driving behavior, even as fuel costs increase.

While taxes on gasoline are often designed to reduce the Vehicle Miles Traveled (VMT) and thus reduce the use of fuel, there is additional evidence that they can have substantial local environmental impacts. As noted by (Parry et al., 2007), there is strong evidence that gasoline taxes can limit the environmental damages that are induced by traffic congestion in particular. This implies that specific congestion reducing policies may be another important way to reduce environmental externalities by motor vehicle usage. Barth and Boriboonsomsin (2009) find motor vehicle emissions to be lowest for intermediate speeds, further motivating congestion reducing policies.

In terms of reducing congestion, there is a large amount of literature that studies the installation of new transportation options and infrastructure. While bike sharing has yet to be widely studied in this context, there is a wealth of literature on investment in public transportation infrastructure as a more environmentally friendly, congestion reducing, and more cost-effective means of transport. Beaudoin et al. (2015) provide a review of studies of the transit investment effects on air quality and congestion. Beaudoin and Lin Lawell (2016) find no empirical evidence that improvements in air quality are induced by increased transit supply at the margin, given existing urban travel regulations. However, Beaudoin et al. (2014) find increases in the public transport network reduces congestion levels. More specifically, they find a 10% increase in public transit capacity reduces traffic congestion by 0.8%. Several other studies also find limited or non-existent impacts from public transit investments on congestion levels. For example, Winston and Langer (2006) evaluate whether government spending on highway construction reduces congestion. They find it to be a very cost-ineffective way to reduce congestion levels. Their analysis observes that one dollar of highway spending reduce congestion costs to road users only by eleven cents in a corresponding year. Winston and Maheshri (2007) find urban rail transit systems in the U.S. to be welfare reducing (based on demand for relative to cost of the service), expect for the San Francisco BART. Other studies are Anderson (2014), Beaudoin and Lin Lawell (2016) and Duranton and Turner (2011). Anderson (2014) uses a public transit strike in Los Angeles to test his hypothesis that public transit users are most likely to be people that would otherwise travel along higher congested roadway. He indeed finds evidence for an 47% increase in travel delay when public transit is discontinued.

In conclusion, although investment in public transportation is a widely discussed topic in politics the findings indicate that the number of commuters using those means of transport is quite small and therefore do not largely contribute to reducing congestion and air quality in a lot of cases. Our paper contributes to this literature, as bike sharing is often seen as a way of public transportation to start and/or finish a trip with motorized public transportation and could overall increase the number of commuters using public transportation.

There are also some studies that go a step further and use novel identification strategies to evaluate the impact of congestion on air quality. Mixed results are found for example by Sexton (2012), Cutter and Neidell (2009) and Chen and Whalley (2012). While Sexton (2012) uses a general equilibrium approach and finds fare-free-days to increase local traffic and air pollution. In contrast, Cutter and Neidell (2009) and Chen and Whalley (2012) find positive evidence for reduced congestion and increased air quality by for example the introduction of a rail line in Taipei (Chen and Whalley, 2012). Overall, existing research suggests that a reduction in congestion improves air quality.

Although often seen as a completion of public transport, which does appear to have only small effects on the reduction of congestion due to a relative small number of commuters that use public transportation, recent research finds evidence for bike sharing to diminish congestion levels. Given the studies on the impact of congestion on air quality, this indicates that expanding bike sharing infrastructure in urban areas could be an effective way to reduce environmental externalities.

The only studies known by us that deal with the impact of bike sharing systems on congestion are Hamilton and Wichman (2018) and Wang and Zhou (2017). While Wang and Zhou (2017) use a difference-in-difference approach to evaluate whether or not bike sharing programs have an impact on city wide congestion levels in U.S. cities, Hamilton and Wichman (2018) examine the impact of a bike sharing program in Washington D.C on congestion levels. Wang and Zhou (2017) use panel data for 96 urban areas in the U.S. for 2005-2014 and Hamilton and Wichman (2018) bike share and real time traffic data for Washington DC. Both find a positive impact of bike sharing programs on congestion reduction in urban areas.

Our paper is unique in several ways. First, we identify the effect of intensified bike sharing infrastructure on congestion, by measuring the impact of an added bike station on congestion levels as well as by investigating whether the popularity of biking within a neighborhood amplifies the effect of the presence of a bike station on congestion. Second, we differ from previous studies that measure the impact on bike sharing on congestion as we both include a mode share analysis as well as a joint analysis of bike sharing and bike route infrastructure to support our findings. Finally, our analysis further contributes to the existing literature by supplementing the findings of Hamilton and Wichman (2018), through the application of their approach to New York City, therefore suggesting their findings may be more generalizable.

The following sections outline our data sources and present the empirical methodology in the spirit of Hamilton and Wichman (2018).

4 Did biking as a ”green” means of transport became more popular in NYC since the implementation of Citibike?

Prior to becoming a means to reduce congestion, a bike sharing program must have sustained adoption by the population. For this reason, we conduct an analysis of the popularity of different modes of transport before and after the implementation of Citibike. We chose the concentrated areas around the East River bridges of Manhattan for three reasons: i) focusing on a spatial area where daily commuters regularly travel through to get to work enables us to capture the causal changes in mode choice by inhabitants of NYC. ii) Given that we can access bike count data from the bridges, we not only have information on bike sharing rides but also on the usage of private bikes. iii) the East River Bridges are equipped with good bike infrastructure, which makes them a better test case for the behaviour of bicycle commuters. In addition to data on bike rides we use information on the ridership of yellow cab taxis and public transportation in these areas. Leveraging this expansive collection of data sources, we are able to identify potential changes in transportation mode choice over time. Moreover, we will be able to spatially narrow our analysis in 6 to East River Bridges in order to get a more accurate assessment of the causal effect of bike sharing stations on congestion. With this we are able to overcome our lack of data on traffic sensors and therefore congestion measures throughout census blocks of Manhattan.

As we discussed earlier, it’s not clear what factors are actually driving the increase in bike sharing ridership. As described in the NYC DOT’s Mobility Report (Mobility Report, June 2018), while bus ridership has fallen virtually every year since 2010, ride-hailing services recorded an all-time high of 92.5 million trips in 2016. We know that more people are cycling to work every year but it remains unclear what modes these commuters took before, or whether the larger ride counts every year are actually from new commuters. In this sense, we could just be seeing the same general pool of bike commuters as in past years, but with more frequent bike commutes. While less revolutionary of a finding, this remains a very interesting case study for our analysis.

Has Citibike introduced new stations in areas that provide a critical convenience boost for commuters? For example, have new stations on the east side of the Manhattan and Brooklyn bridges encouraged commuters to park and ride the last mile to their destination? More stations in areas such as those can provide seamless connections for commuters that want to avoid high congestion areas for car traffic (on the bridges), or the high costs of parking in Manhattan. It certainly seems as though there would be significant room for bike share adoption as a public transit extension, with roughly 95 percent of NYC residents currently walking to transit and 97 percent walking from transit (Mobility Report, June 2018).

5 Does the coexistence of bike sharing with an improved bike lane network make your city ”greener”?

Before investigating the impact of bike sharing on congestion within Manhattan, we sought to investigate whether an improvement in bike route infrastructure could better encourage biking as a mode of transportation. Through the focus on bike sharing with Citibike, this finding would emphasize the hypothesis that bike sharing programs require a well-developed bike route network in order to properly function.

For this, we graphically analyze the monthly bike share rides on each bike street segment over time in order to identify potential jumps in bike ridership after the installation or modification of individual bike lanes. As explained in section A.2 we first identify the census blocks intersected by a bike route and assign them the corresponding bike route’s information. Secondly, we remove all census blocks that do not contain a bike station after any of the bike share expansions. Finally, we aggregate the monthly rides of the census blocks so that they correspond to individually matched bike lanes. To focus on the possible increase in bike rides due to the modification or installation of a bike lane rather than accidentally capturing also the effect of bike station introductions we further disaggregate our dataset by focusing on the bike lanes with census blocks that contained bike ride stations after the introduction of Citebike in 2013 (and repeat the same procedure for the bike lanes with census blocks that contained bike stations after the Citebike extensions in August 2015, 2016 and September 2017). For the same reason we restrict the dataset to installation dates after 2013. We plot the aggregated monthly rides over time for each bike lane that intersects a census block containing a bike share station. To consider the effects before and after a bike lane installation or modification, we identify these dates in the graph and observe whether there are any observable spikes for each bike lane. To help identify spikes, we restrict our analysis to summer months, since the number of bike rides fluctuate quite heavily in the winter months from year to year. For instance, spikes might represent otherwise be obscured because of weather events like blizzards. Moreover, we wondered whether the type of the bike lane could influence the number of bike rides. To evaluate this, we repeat the procedure described above, but further subset our data by the bike lane’s type (focusing on Sharrows and Protected Paths for now).

Overall, aside from seasonal variations, the bike rides for each road segment tend to increase over time. Although we do not find any significant spikes after the installation/modification of bike routes, one can see that the upward trend over time is steeper for protected paths compared to shared lanes on streets (so called ”Sharrows”). We interpret protected bike paths as ”better” bike infrastructure than shared lanes, as they provide more security for bike riders. Looking at the number of rides between protected and shared bike routes, we can see that riders tend to prefer protected bike routes. Most of the shared bike lanes in our filtered dataset have rides per month that vary between around 100 and 2000. Protected bike lanes on the other hand, tend to have between 1000 to 5000 bike rides per month. The graphs for different Segment ids can be found in the Appendix in Figures 7 to 12.

6 Methodology - The impact of bike sharing stations on congestion

The goal of our empirical approach is to identify the causal effect of bike sharing on traffic congestion. One can think of different mechanisms why bike sharing may reduce traffic congestion. Automobile drivers might opt to use a bike instead of driving, because they expect time savings by avoiding traffic congestion or other utility increases related to biking, for example health related effects. Due to the extension of the existing public transportation network bike sharing might become more attractive than driving (Hamilton and Wichman, 2018). Other reasons Hamilton and Wichman (2018) mention are the following: If bike sharing users switch from driving, then clearly this should lead to less congestion on roads. If bike sharing users switch from other public transport or use it as a multi-modal supplement to complete a public transport trip then traffic should at least not increase. A possible factor that may increase congestion is the interference of bikes and cars on roads. However, as in Hamilton and Wichman (2018) we only seek to answer the combined effects of these mechanisms on congestion.
Controlling for unobserved effects on the block (group) level, we first estimate the relationship between traffic congestion and the existence of bike stations within a block (group). The baseline reduced form model is specified as

$$lnCONG_{jhdmt} = \alpha + \delta_j + \nu_h + \mu_m + \nu_t + \eta Temp_d + \phi AWND_D + \beta Station_{jhdmt} + \epsilon_{jhdmt} \tag{1}$$

where $CONG_{jhdmt}$ is the average congestion among all road sections with sensors in Census block (group) j during the hourly period h at day d in month m and year t. Besides from excluding all days with precipitation, we include the variables $Temp_dmt$ and $AWND_dmt$ to control for unfavorable weather circumstances, where bikers are less likely to be on the road. The variables δ_j,ν_h,μ_m and ν_t represent dummy variables for block group, hour, month and year fixed effects, respectively. $Station_{jhdmt}$ denotes the indicator variable for whether a station is available in the Census block (group) j in the time period specified by h,d,m,t. Hence, β is our primarily coefficient of interest. Standard errors are clustered by day to account for correlation across time but not for spatial correlation, as we expect congestion between neighboring blocks to be correlated.

The identification strategy in this paper is unique in that we have two different introductions of new bike stations after the initial launch of the bike sharing system in June 2013. These subsequent increases come in August 2015 and again in August 2016, increasing the total amount of bike stations throughout Manhattan by approximately 100 each time. As explained in section A, to estimate the effect of bike infrastructure on traffic congestion, we are associating each of the bike stations with the census blocks that they are located within. In this sense, we are able to identify treatment effects for each of the census blocks, depending on whether or not they have a station within them at each of the successive station updates. Similarly, we locate the traffic data sensors in order to merge the congestion data for each traffic sensor with the census blocks that it's contained inside.

As we expect the effect to be bigger in Census block (groups) that are more congested we prove the robustness of our results by conducting quantile regressions as well as splitting our data into geographical areas that have a congestion above the median and those with congestion below the median. This allows us to explore the heterogeneity in the impact of bike share stations. More clearly, we seek to identify whether the reduction of congestion is concentrated among Census blocks with relatively high congestion levels.

Moreover, we are aware of the fact that some Census block groups are more densely populated by bike share stations than others. We expect that this could amplify the decreasing effect on congestion in those Census block groups. However, we may on the other hand expect that a much higher number of bikers on the roads could lead to increased congestion as all traffic participants have to pay more attention to different transportation modes on a single road segment. Hence, in order to investigate whether a more dense bike infrastructure within a given geographical unit further increases the effect on congestion, we adjust our model by replacing the station indicator by a measure for the amount of bike stations within a Census block (group). Additionally, we seek to identify the effect of an intensified use of bike infrastructure within a spatially defined area by building a measure that accounts for the number of arriving bike sharing trips from the bike stations within a Census block (group). This allows us to identify whether the effect on congestion is stronger in geographic units that are more popular for biking. For this, we include a variable that represents the sum over all departing rides from all bike stations within a Census block (group) in addition to the station indicator in Equation 1. More clearly, we focus on the arriving rides as we would expect this measure to be more representative for the morning rush hour in Manhattan as the majority of commuters will bike towards Manhattan rather than from Manhattan to the outer Boroughs of NYC. Our hypothesis is supported by Figure 5, which clearly implies that during the morning rush hour between 6 a.m. and 10 a.m. the major amount of bike trips goes from outer boroughs to Manhattan, while during the afternoon rush hour the spike is depicted for the reverse direction.

Importantly, for our analysis we decide to focus on the borough Manhattan for several reasons: i) it is the borough with the by far largest amount of bike stations and traffic sensors given our dataset ii) congestion in Manhattan is a serious problem faced by NYC government iii) bike stations are widely spread throughout major parts of the borough which allows us to argue that bike stations are spread relatively evenly across more and less congested parts of Manhattan as well as higher and lower income blocks on the island[7]. The latter, leads us to assume that the bike stations across Manhattan are sited roughly randomly and prevents us from possible reverse causation problems.

7 Estimation Results

The results produced by our baseline model are displayed in Table **??** for blocks and in Table **??** for block groups. Our results suggest that the point estimates for the bike share station indicator are significant and negative. That means the results suggest that congestion is reduced by the presence of a bike station within a Census block (group). Moreover, the effect is smaller on block group level compared to the smaller unit, Census blocks.

We control for the popularity of the bike lanes within Census blocks by including the average number of bike rides taken on each bike lane for that month. The point coefficient of the bike ride variable itself however is close to zero as it measures the marginal impact of one additional ride. The results can be found in Table **??**. We made the decision to control for the number of rides on a bike lane because we believe it acts as the best proxy for both the type of that bike lane (protected lanes have much more riders) and simply how popular it is as well. We additionally control for the presence of subway stations and the count of total bike rides taken across each of the east river bridges, but these controls have been excluded from our regression.

8 Discussion

In general, our coefficient estimates imply a causal link between the presence of bike share stations and a reduction in congestion. In the context of extremely large costs of congestion, the reduction in congestion of approximately 17% on Census block level and around 10% on Census block group level can generate extensive welfare improvements (Schrank et al., 2015).

As estimated by Schrank et al. (2015), there are extremely large costs of congestion. Our analysis suggests, that an upgrade of the existing transport network by the introduction of a bike sharing program can noticeably reduce these costs. Using the estimates of Schrank et al. (2015) for 2014 and our estimation results for Census

[7]http://www.businessinsider.com/new-york-city-income-maps-2014-12

blocks, the reduction of around 2%, in the block groups that contain a bike station (and a traffic sensor) would reduce annual congestion costs if applied to the whole NYC area by approximately $35 per auto commuter and total costs by $294 million. The reduction of congestion costs are estimated as a combination of reduced travel time delays and less wasted fuel. Moreover, a reduction in congestion of around 2%, would imply social benefits in the form of CO_2 reduction. The estimates of Eisele et al. (2013) together with our results would imply reductions of around 11 pounds per commuter and 103 million pounds of congestion-induced CO_2 emissions. As these estimates ignore for example environmental benefits from improved air quality and possible private cost-savings and health benefits from mode switching that could arise for bicycle commuters one could expect the actual monetary benefits to be much greater (Hamilton and Wichman, 2018).

Controlling for the popularity of bike lane ridership within a Census block (group), we find that the congestion reducing effect is dramatically increased. This indicates that the more popular the station is the larger the congestion-reducing effect is.

Despite seeing consistent congestion decreases in our block level specifications, it is possible that our results may actually be underestimating the true congestion reducing effect of the presence of stations. On the smaller Census block level, decreasing congestion over time could in theory prompt changes to the commuting routes of drivers. In this sense, we may expect to see congestion decreases at first but smaller effects over time as drivers shift their routes from neighboring roads. The fact that we see persistent congestion decreasing effects over time in spite of this effect is very encouraging.

9 Conclusion

In our empirical analysis we explore the causal effect of bike sharing programs on traffic congestion. Our results suggest immense savings of congestion induced costs, that translate into considerable welfare gains. Using the NYC Real Time Traffic Speed Data Feed that contains traffic information from the NYCDOT speed detectors, of mostly arterials and highways within the City limits, we are able to construct a relatively fine spatially defined congestion measure. Our resulting panel dataset of hourly traffic observation on Census block (group) level suggests a reduction of traffic congestion in block groups with bike stations. The produced results indicate that the presence of bike share stations translates into approximately 11% reduction in traffic congestion on block group level and around 17% on Census block level. Considering bike sharing as a trip augmenting mode of transportation, the investigation of individuals' transportation substitution patterns towards bike sharing shows that there can be significant congestion-reducing effects through the better utilization of bike-share services. This can deliver important insights for policy makers to optimally allocate public funding across different types of public transportation. A further interesting aspect to study would be the relationship between the availability of parking spaces on the use of bike share programs. A lack of public available parking lots or new introduced restrictive regulations for public parking during our time span of interest might induce individuals to opt for public transport or biking instead of driving their car. This could further reduce congestion and can deliver interesting policy implications.

Over the course of writing this paper we weren't able to focus on all aspects we wanted. Thus, in the future we would first like to expand our analysis beyond Manhattan. Our plan is to include all parts of the city that include bike stations. We focused on Manhattan in our analysis as we argue that given the reasons mentioned in section 6 we would not necessarily have to employ the propensity score matching to account for non random siting of bike stations, and the fact that propensity score matching would be beyond the scope of this paper. However, for a spatially extended version of our dataset beyond the bridges of Manhattan we suggest the propensity score matching technique employed by Hamilton and Wichman (2018) to account for non random siting of bike station within more highly congested and higher income Census block (groups) rather than evenly distributed bike stations across the city.

Second, if possible we would like to remove a limitation of our dataset. A more extensive dataset of traffic data that delivers a larger amount of traffic sensors throughout the city would allow us to include a larger percentage of Census blocks and therefore would deliver more reliable results. Another possibility would be to use the dataset of yellow taxis' trips in NYC and construct a congestion measure given the taxi rides and trip length over the study period.

References

Anderson, M. L. (2014). Subways, Strikes, and Slowdowns: The Impact of Public Transit on Traffic Congestion. *American Economic Review*, 104(9):2763–2796.

Barth, M. and Boriboonsomsin, K. (2009). Traffic Congestion and Greenhouse Gases. *Access Magazine*.

Beaudoin, J., Farzin, Y. H., and Lin Lawell, C.-Y. (2014). Evaluating public transit investment in congested cities.

Beaudoin, J., Farzin, Y. H., and Lin Lawell, C. Y. (2015). Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality. *Research in Transportation Economics*, 52:15–22.

Beaudoin, J. and Lin Lawell, C.-Y. C. (2016). Is Public Transit's "Green" Reputation Deserved? Evaluating the Effects of Transit Supply on Air Quality. *Working Paper*.

Bento, A. M., Goulder, L. H., Jacobsen, M. R., and Von Haefen, R. H. (2009). Distributional and efficiency impacts of increased US gasoline taxes. *American Economic Review*, 99(3):667–699.

Burger, N. E. and Kaffine, D. T. (2009). Gas Prices, Traffic, and Freeway Speeds in Los Angeles. *Review of Economics and Statistics*, 91(3):652–657.

Chen, Y. and Whalley, A. (2012). Green Infrastructure: The Effects of Urban Rail Transit on Air Quality. *American Economic Journal: Economic Policy*, 4(1):58–97.

Currie, G. and Phung, J. (2007). Transit Ridership, Auto Gas Prices, and World Events: New Drivers of Change? *Transportation Research Record: Journal of the Transportation Research Board*, 1992:3–10.

Cutter, W. B. and Neidell, M. (2009). Voluntary information programs and environmental regulation: Evidence from 'Spare the Air'. *Journal of Environmental Economics and Management*, 58(3):253–265.

Duranton, B. G. and Turner, M. A. (2011). Fundamental Law of Road Congestion: Evidence from U.S. Cities. *American Economic Review*, 101(October):2616–2652.

Eisele, W. L., Fossett, T., Schrank, D. L., Farzaneh, M., Meier, P. J., and Williams, S. P. (2013). Estimating and Incorporatin CO2 Emissions and Associated Fuel Consumption into the Urban Mobility Report. Technical report, University of Wisconsin, Madison.

Hamilton, T. L. and Wichman, C. J. (2018). Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare. *Journal of Environmental Economics and Management*, 87:72–93.

Martin, E. W. and Shaheen, S. A. (2014). Evaluating public transit modal shift dynamics in response to bikesharing: A tale of two U.S. cities. *Journal of Transport Geography*, 41:315–324.

Midgley, P. (2011). Bicycle-Sharing Schemes: Enhancing Sustainable Mobility in Urban Areas. *United Nations Department of Economic and Social Affairs: Commission on Sustainable Development*, (Background Paper No. 8).

Parry, I. W. H., Walls, M., and Harrington, W. (2007). Automobile Externalities and Policies. *Journal of Economic Literature*, 45(2):373–399.

Pucher, J. and Buehler, R. (2005). Cycling Trends & Policies in Canadian Cities. *World Transport Policy & Practice*, 11(1):43–61.

Schrank, D., Eisele, B., Lomax, T., and Bak, J. (2015). 2015 Urban Mobility Scorecard. *Texas A&M Transportation Institue*, 39(August):5.

Sexton, S. E. (2012). Paying for Pollution? How General Equilibrium Effects Undermine the "Spare the Air" Program. *Environmental and Resource Economics*, 53(4):553–575.

Shaheen, S. A., Guzman, S., and Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future. *Transportation Research Record*.

Spiller, E. and Stephens, H. (2012). *Does the Substitutability of Public Transit Affect Commuters' Response to Gasoline Price Changes?* Number July.

Wang, M. and Zhou, X. (2017). Bike-sharing systems and congestion: Evidence from US cities. *Journal of Transport Geography*, 65:147–154.

West, S. E. and Williams, R. C. (2005). The cost of reducing gasoline consumption. *American Economic Review*, 95(2):294–299.

Winston, C. and Langer, A. (2006). The effect of government highway spending on road users' congestion costs. *Journal of Urban Economics*, 60(3):463–483.

Winston, C. and Maheshri, V. (2007). On the social desirability of urban rail transit systems. *Journal of Urban Economics*, 62(2):362–382.

A Data

A.1 Citibike system data

Bikesharing system data is publicly available from Citibike. The dataset contains data on every ride taken on a CB-bike since its introduction on June 1st, 2013. More precisely, this includes exact trip duration (including the second that the trip started and ended), the latitude and longitude of the start and end stations, the users' type (subscriber or customer), the age of the rider, and their gender.

We are primarily interested in more aggregated information of the bike share data as we are concerned whether there are observable changes in transportation mode choice due to extension of bike infrastructure in NYC as well as the impact of the existence of bike sharing on congestion and mode choice, especially in spatial proximity to the east river bridges. For these reasons we decide to use the existence of bike share stations in spatial proximity as an indicator for bike sharing in that area and are therefore able to analyse aggregated bike trip information on station level. As the locations of the bike share stations are only available within the system data set, we have to back out the geographic locations of the bike share stations as well as the date of installation of those stations from the system dataset. Figure 3 presents the number of bike share stations at different dates of our time period of interest. Importantly, a large number of the stations were established in the course of expansions during our study period from April 2013 to August 2017. More clearly, we create an dummy variable that indicates the presence of a bike station in a given census block at a given point in time. This enables us on the one hand to generate traffic congestion observations within the same census block (group) before and after new bike share stations were installed. On the other hand we are able to identify the number of bike rides departing from and arriving at a given station.

A.2 Bike lane data and bike counts

The NYC Department of Transportation (DOT) biannually publishes information on all bike lines throughout NYC. The data contains information on the type and direction of the bike lane as well its spatial coordinates and the installation/modification dates. Using this data in combination with our information on the existence of bike sharing stations in given census blocks throughout Manhattan we are able to identify possible spikes in the number of bike rides after the installation or modification of a bike lane. Moreover, we can identify whether certain types of bike lane types (e.g. on/off street) are more popular/more in use than others. To spatially link the bike lane data to our stations within census blocks dataset, we assign the census blocks that are intersected by a given bike lane.

Additionally, since 1987 the NYC DOT publishes monthly bike counts for the east river bridges towards Manhattan. In 2014, NYC DOT installed automated counters. These enable continuous 24 hour data every day of the year, which is averaged on a monthly basis. Despite installing the automated counters in 2014, the data on bike counts wasn't released at monthly intervals until 2015 (yearly averages before then), and they only display the monthly counts for April to October. As a result, we only considered the counts for those months, from 2015 until 2018.

A.3 Other data sources

To account for the presence of a public transportation station nearby, we control for the precise location of each subway station. When there is a subway station included in that census block or block group, there is a dummy variable of one.

A.4 Traffic data

To calculate our congestion measures we access publicly available real time traffic data from BetaNYC, which is a not-for-profit fiscally sponsored organization and an organizing force for local civic engagement.[8]

Archived data from the NYC Real Time Traffic Speed Data Feed is available in five minute intervals starting from April 2015 and publicly downloadable in monthly files. The data contains traffic information from the NYCDOT speed detectors, mostly major arterials[9] and highways within the City limits. In detail, it contains average speed a vehicle traveled in the corresponding interval, average travel time, longitude and latitude of the sensors' location, as well as a description of start and end points of the road section covered by the detector. Important to note is that the variable linkTimeStamp indicates when data was last received from the sensor. If the data is other than the current date it indicates an error with that sensor and that data should not be used.[10] The fact that this data contains also observation from arterial roads within the city allows us to not only focus on highways as in Hamilton and Wichman (2018). This is important for the fact that bike stations are more likely to be accessible from arterial roads rather than highways.

However, as this data is only available starting from April 2015, rather than using the period before and after the launch of Citi Bike we focus on the period before and after the first extensive expansions of the bike sharing program. This expansion took place between August 2015 and 2016. Citi Bike expanded by 280 additional stations and 4000 new bikes during the time span (Section 1). Thus, using traffic data between April 2013 and August 2017 we receive traffic observations before and after the installation of bike stations. This extensive expansion ensures that we are able to conduct a quasi-experimental approach.

Each road segment is covered by several speed sensors, separating a long segment into shorter ones. To construct our congestion measure we take the location of each short road segment's sensor and match it with the corresponding census block the sensor is located in.

As we are interested in traffic congestion rather than observed speed as a primary variable of interest we have to construct a measure of congestion. For this we have to define a reference speed for each road segment. We define the reference speed, $Speed_j^R$, as the average of speeds during the early morning hours 3 a.m.- 4 a.m. , where we implicitly assume to have less traffic on the roads and therefore no congestion. This assumption can be supported by Figure 1. The Figure shows monthly average congestion throughout a day from Midnight to 4 p.m. It shows congestion levels to be lowest during the early morning hours with increasing congestion starting from around 6 a.m. Defining $Speed_{jt}^O$ as the observed speed at a time t congestion is measure as

$$CONG_{jt} = \frac{Speed_j^R}{Speed_{jt}^O}, \tag{2}$$

which is decreasing with the speed observed.

Similar to Hamilton and Wichman (2018), we focus on the morning rush hour from 6-10 a.m. to measure congestion in hourly intervals for two reasons: i) during this time of the day it is likely to focus on actual commuters that bike on a regular basis rather than for example tourists that do not have the outside option to take a car and iii) aggregating and filtering the data makes the huge amount of data more tractable in terms of available processing power. The former point is supported by Figure 2, which depicts a tremendous spike during the time span between 6-10 a.m. Moreover, to improve the likelihood of focusing on commuters we remove all weekend days. As Figure 2 indicates, the usage pattern of Citi Bike is significantly different

[8] http://data.beta.nyc/about

[9] according to http://www.nyc.gov/html/dot/downloads/pdf/nycdot-streetdesignmanual-interior-07-glossaryandappendices.pdf, an arterial street is the part of the roadway system serving as the principal network of through-traffic flow. The routes connect areas of principal traffic generation and important rural highways entering the cities.

[10] http://data.beta.nyc/dataset/nyc-real-time-traffic-speed-data-feed-archived/resource/6b126292-b251-4e59-bfba-6844a633a3a2

during the weekend and so we focus only on the weekday riders. Additionally, we can see from Figure 2 that the weekday rides start to increase at 6 a.m. and fall back down to relatively normal levels by 10 a.m. Further, to account for seasonality we focus on specific months, where arguably biking is a reasonable option to commute (Figure 4). Thus, we focus on the months April to October during our time span of interest. Hamilton and Wichman (2018) argue that an increased use of bike sharing by tourists would lead to biased estimates during the summer months (June-April). We abstract from this argument as New York can be seen as a whole year travel destination. [11] Lastly, we do not include the Citibike expansion that took place in September 2017, as our dataset would only include two month after this expansion (Oktober 2017 and April 2018).

A.5 Census blocks and Census block groups

In order to spatially link our variables of interest - bike share stations and congestion - we use census blocks and census block groups as geographic identifiers. U.S. Census block groups are the second smallest geographic unit of observation defined by the U.S. Census. According to Hamilton and Wichman (2018), the Census block groups may be preferable area identifier instead of the smallest designation, census blocks, for the following reasons. First, census blocks are too small in the sense that a sensors of a road segment might refer to several census blocks, although its sensor is in only one. Second, to capture the relationship of transportation mode choices more accurately as "the size of census blocks makes it likely that individuals move across blocks for their commuting choice (e.g., an individual might walk to a bikeshare station in a bordering block or bike through a block that borders that of the usual car commute)." (Hamilton and Wichman (2018), p.77) However, we also focus on the smaller designation in order to start from a more spatially refined identification. In this sense, a measure of the count of Citibike stations within a census block group may do very little to explain how accessible the stations are to people inside those block groups. The census blocks, by contrast, provide us with a unique look into the distribution or spread of the stations within each block group.

Technically the bike stations and therefore the number of stations within a block group are easily linked to their corresponding census block groups. As explained above the middle point of each traffic speed monitored road segment are linked to the census block data set. Figure 6 depicts the bike stations' locations throughout Manhattan within the Census blocks that contain a traffic sensors. More explicitly, green depicts areas with one bike station, yellow areas with two bike stations, areas that contain three bike stations are orange and areas containing four and five bike stations are characterized by different light and dark red, respectively.

A.6 Climate controls

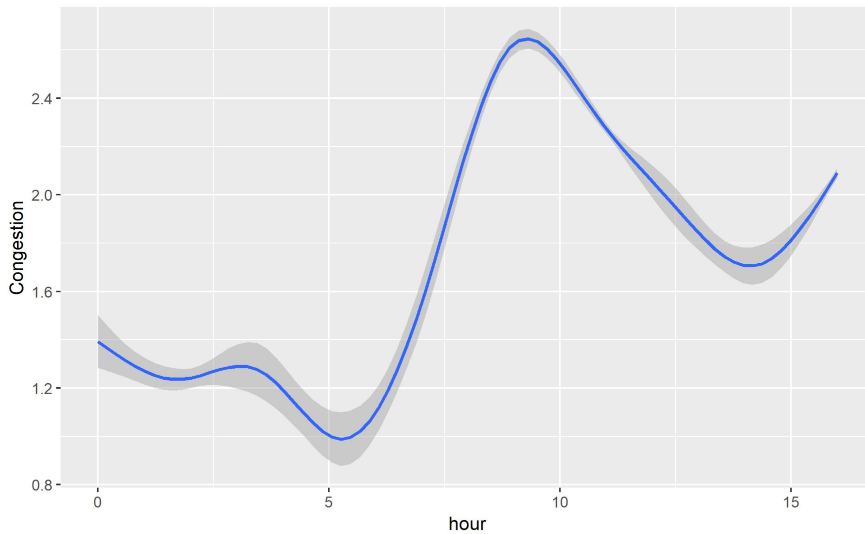
Furthermore, we access daily climate summaries for the weather station at Central Park New York City from 2013 to 2017 from the NOAA National Centers for Environmental Information[12]. Arguing that biking is predominantely an option on days without precipitation, we remove all days with precipitation above zero. We also include average temperature and average wind speed to account for daily fluctuations in weather.

[11] https://www.nycgo.com/assets/files/pdf/new_york_city_travel_and_tourism_trend_report_2017.pdf

[12] https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00094728/detail

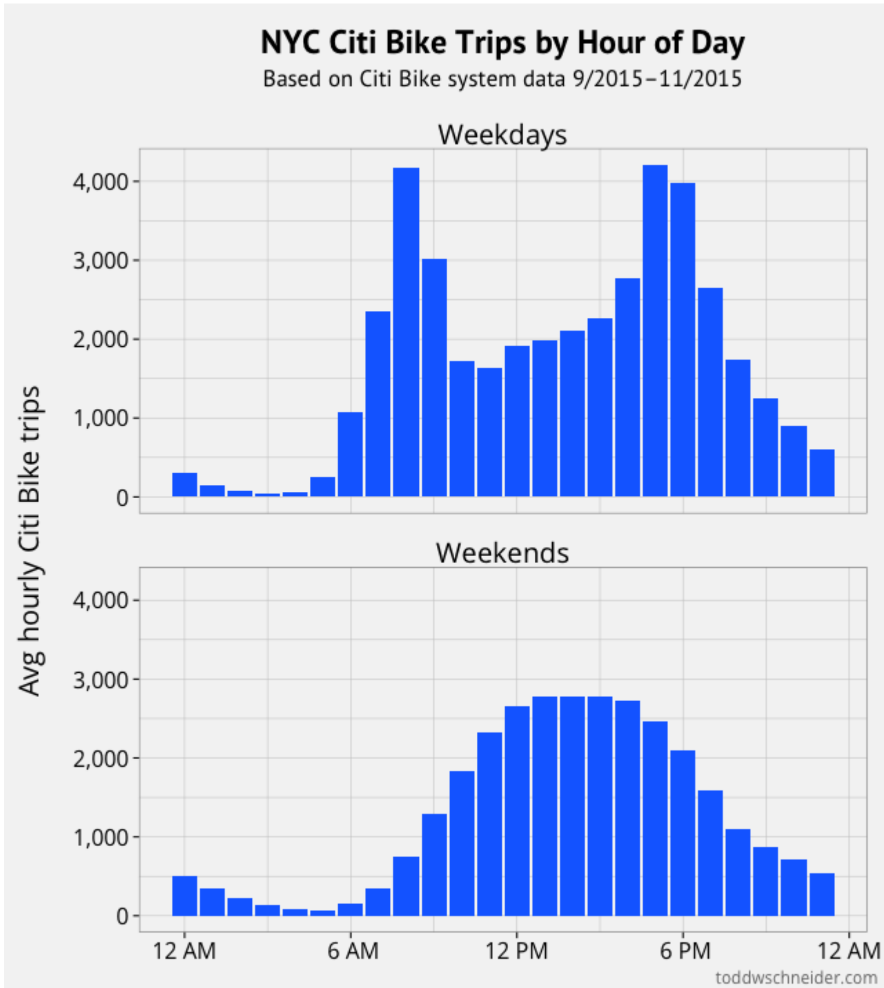
B Figures

Figure 1: Average congestion pattern throughout day for July 2017



Note: The graph depicts the average congestion pattern between Midnight and 4pm of July 2017

Figure 2: NYC Citi Bike Trips by Hour of Day



Source: http://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-system/

Figure 3: Trend in number of Citi Bikeshare trips and number of stations between May 2013 and August 2017

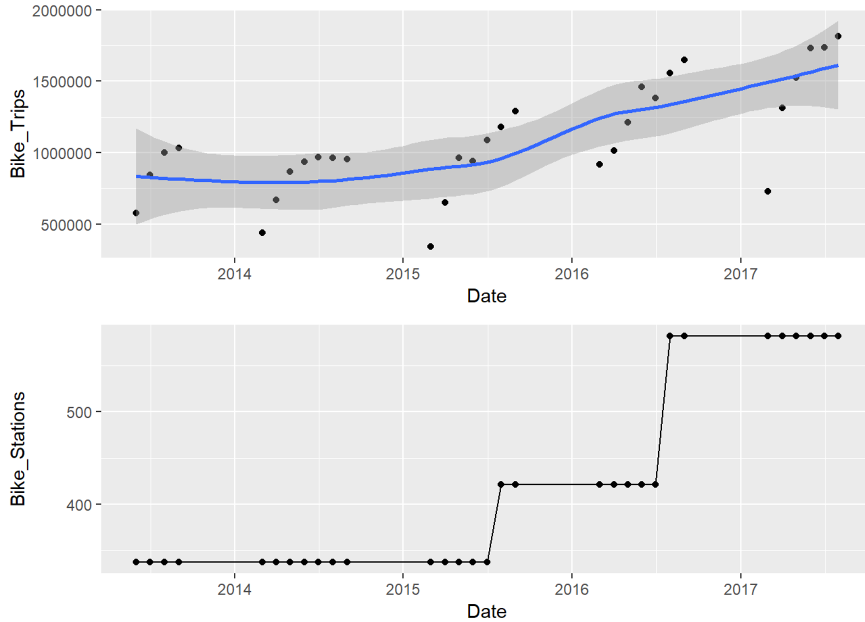
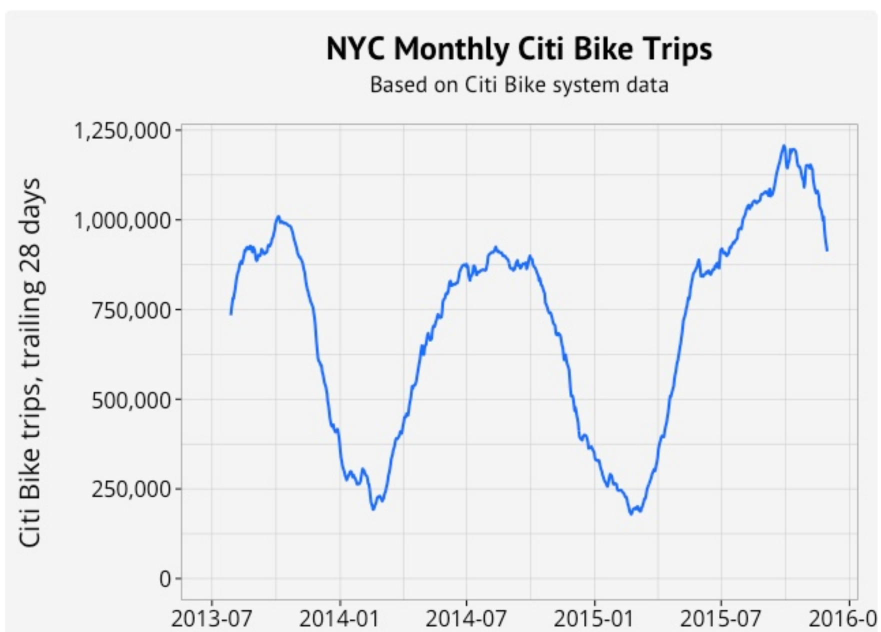
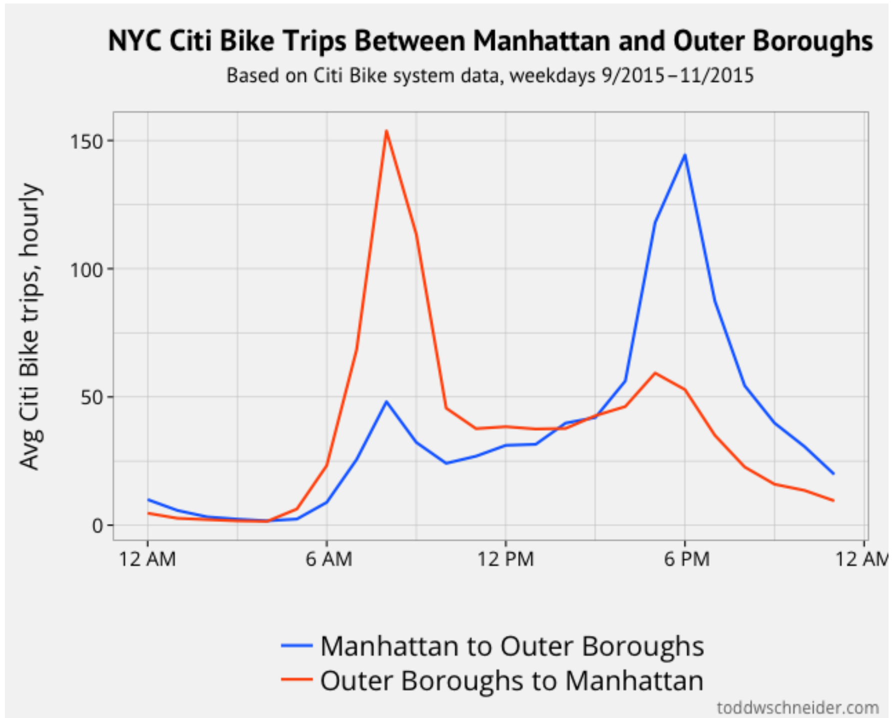


Figure 4: NYC Monthly Citi Bike Trips



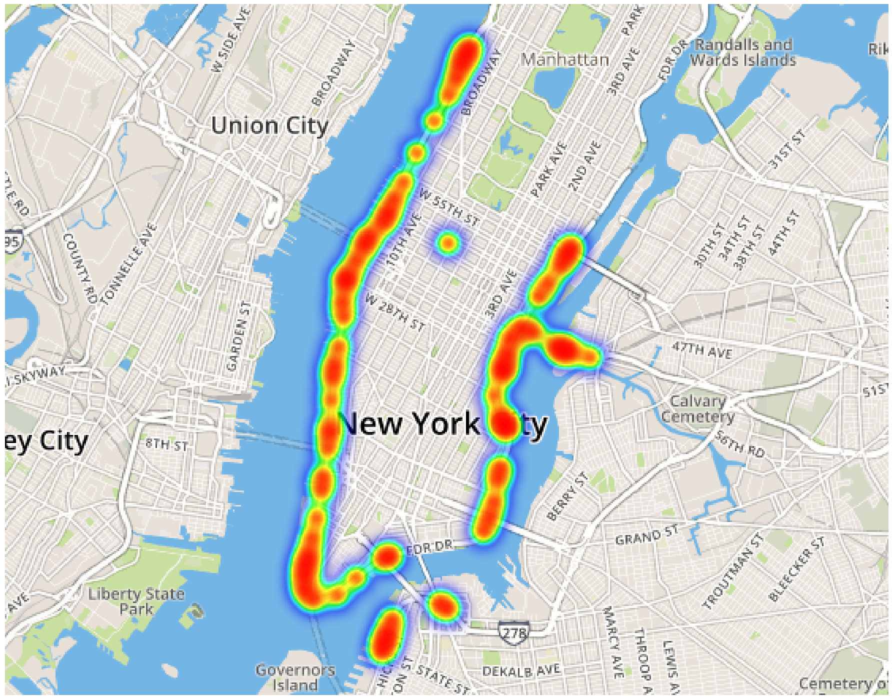
Source: http://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-system/

Figure 5: NYC Citi Bike Trips Between Manhattan and Outer Boroughs



Source: http://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-system/

Figure 6: Citibike station in Census blocks with Traffic sensors in August 2017



The above heatmap is coloured according to the number of stations in each census block group. Block groups with one station are green, those with two are yellow, three are orange, and those with four and five stations are light and dark red.

Figure 7:

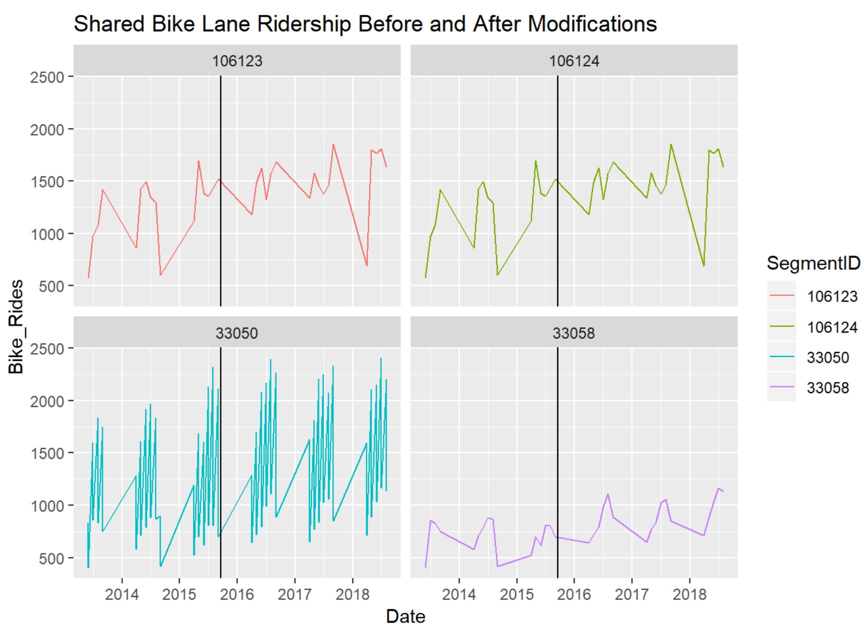


Figure 8:

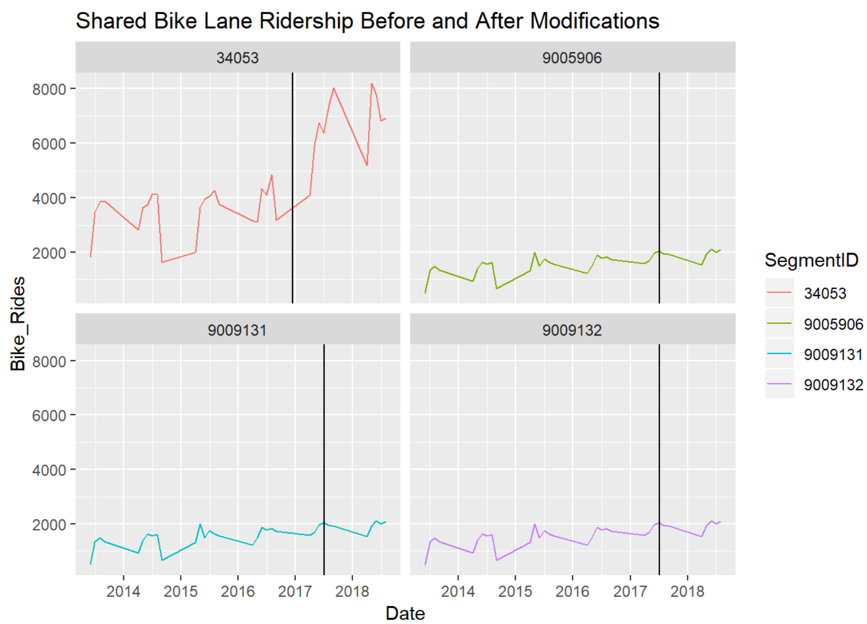


Figure 9:

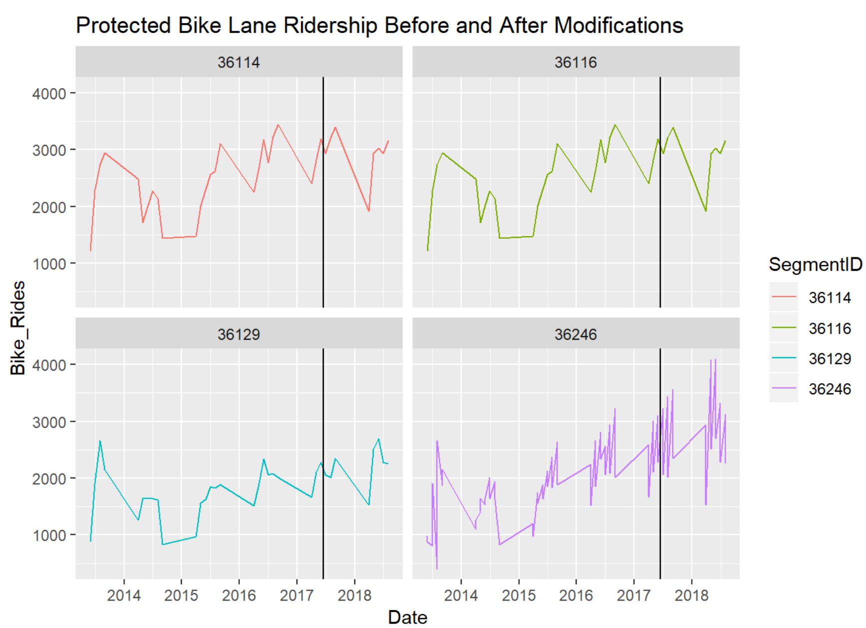


Figure 10:

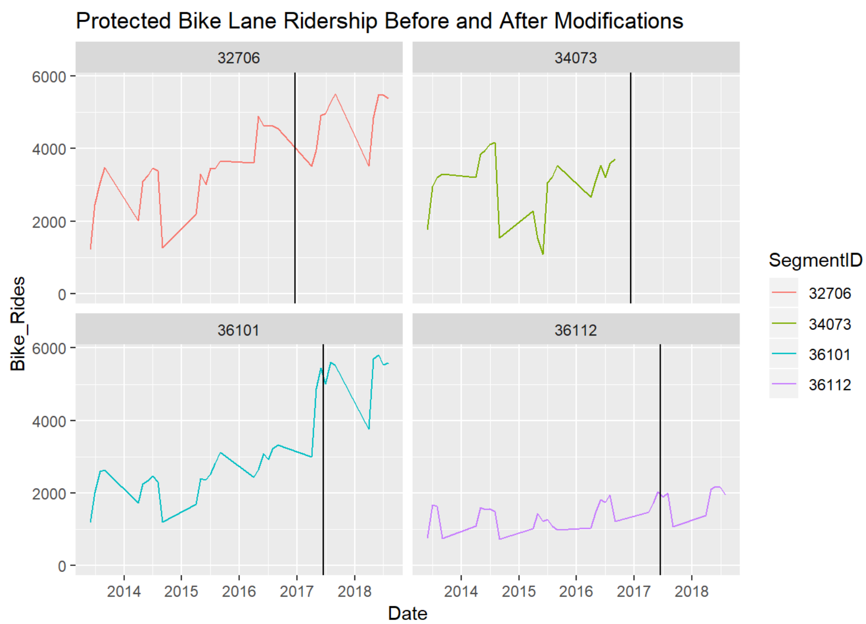


Figure 11:

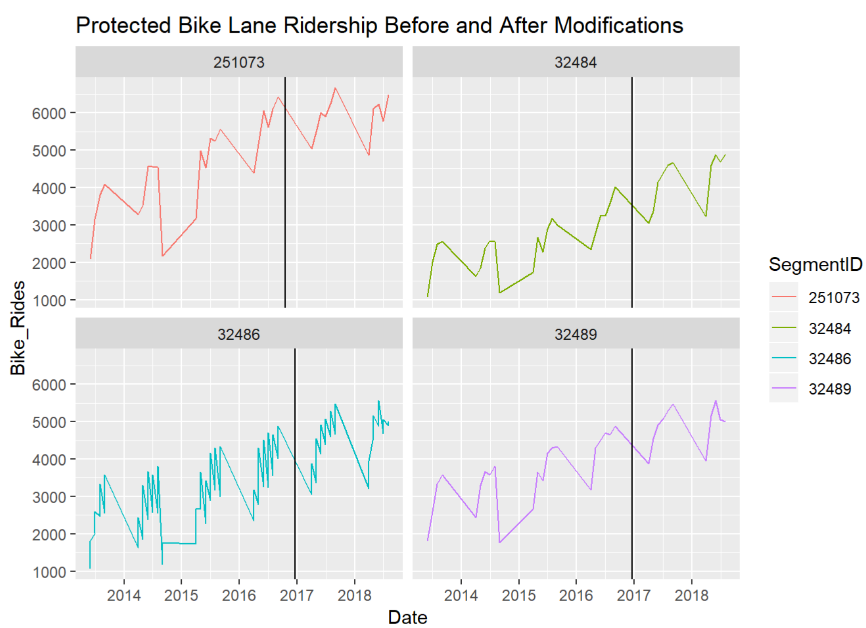
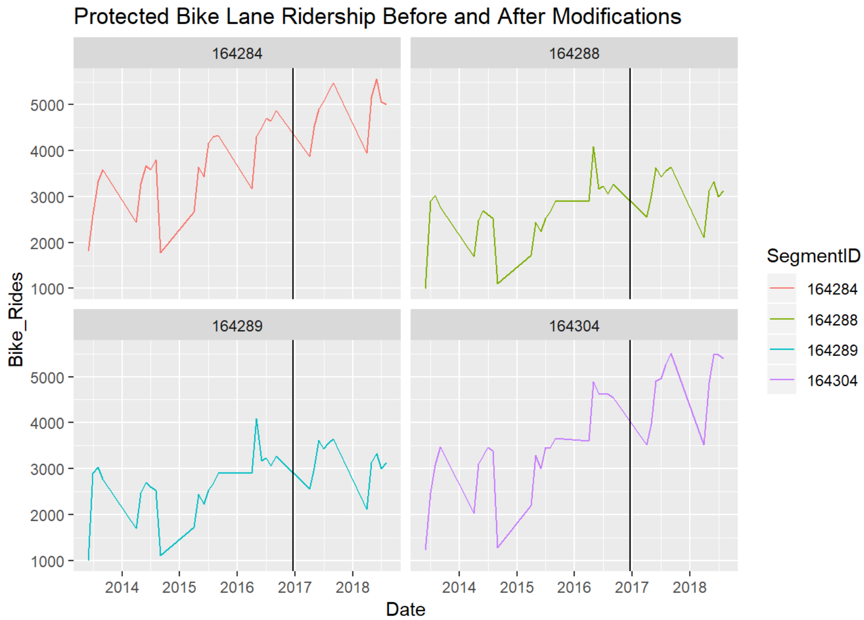


Figure 12:



Protected Bike Lane Ridership Before and After Modifications
164284
164288
164289
164304
5000
4000
3000
2000
1000
Bike_Rides
2014
2015
2016
2017
2018
Date
SegmentID
164284
164288
164289
164304

C Tables

Table 1: Census Block Fixed Effects

	Dependent variable:
	Congestion
station_count	-0.172^{***}
	(0.001)
Bike_Lane_Rides	-0.003^{***}
	(0.001)
linkId	-0.0001^{***}
	(0.00000)
Time_hour	0.115^{***}
	(0.002)
Date	0.0002^{***}
	(0.00001)
TAVG	0.001^{***}
	(0.0002)
AWND	0.010^{***}
	(0.001)
Constant	171.390^{***}
	(8.860)
Observations	30,307
R^2	0.175
Adjusted R^2	0.174
Residual Std. Error	0.421 (df = 30300)
F Statistic	$1,068.124^{***}$ (df = 6; 30300)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Census Block Group Fixed Effects

	Dependent variable:
	Congestion
station_count	-0.113^{***}
	(0.002)
Bike_Lane_Rides	-0.007^{***}
	(0.00003)
linkId	0.00000^{***}
	(0.00000)
Time_hour	0.110^{***}
	(0.001)
Date	0.0003^{***}
	(0.00000)
TAVG	0.001^{***}
	(0.0001)
AWND	0.014^{***}
	(0.001)
Constant	-16.476^{***}
	(0.137)
Observations	160,720
R^2	0.179
Adjusted R^2	0.179
Residual Std. Error	0.483 (df = 160713)
F Statistic	$5{,}833.843^{***}$ (df = 6; 160713)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$