Master Degree in Specialized Economic Analysis

Measuring the impact of bike lanes in the number of bikeshare trips in Barcelona

Vitor Quintanilha Barbosa

Directors: Hannes Mueller and Joan Llull

June 2017
ABSTRACT:

We analyze the impact of the construction of bike lanes in Barcelona during 2015 and 2016 on the number of trips of a public bicycle-share system. We use different specifications of control groups and also different distances of impact for the new lanes. Using different econometric models we find robust effects when considering short distances (200m) that goes from 3.8% to around 5.7% in the number of trips. We find less convincing evidence for higher distances, what could indicate either absence of effect or the possibility of a displacement effect between stations.

ABSTRACT IN CATALAN:

Analitzem l'impacte de la construcció de carrils bici a Barcelona durant els anys 2015 i 2016 en el nombre de viatges d'un sistema públic de bicicletes compartits. Utilitzem diferents especificacions dels grups de control i també distàncies d'impacte diferents per als nous carrils. Mitjançant models economètrics diferents, es constaten uns efectes robusts al considerar distàncies curtes (200 m) que van del 3,8% al 5,7% en el nombre de viatges. Determinem proves menys convincinges per a distàncies més altes, el que podria indicar una ausència d'efecte o la possibilitat d'un desplaçament entre les estacions.

KEYWORDS IN ENGLISH: Bike Lane Impact, Cycling Infrastructure, Bicycle-Share System.

KEYWORDS IN CATALAN: Impacte carril bici, Infraestructura de ciclisme, Sistema de bicicletes compartides.
Measuring the impact of bike lanes in the number of bikeshare trips in Barcelona

Vitor Quintanilha Barbosa
2017/05

Abstract:

We analyze the impact of the construction of bike lanes in Barcelona during 2015 and 2016 on the number of trips of a public bicycle-share system. We use different specifications of control groups and also different distances of impact for the new lanes. Using different econometric models we find robust effects when considering short distances (200m) that goes from 3.8% to around 5.7% in the number of trips. We find less convincing evidence for higher distances, what could indicate either absence of effect or the possibility of a displacement effect between stations.

JEL: R22, C20.
Keywords: Bike Lane Impact, Cycling Infrastructure, Bicycle-Share System.

* Practically all software used in the development of this paper are high-quality open-source programs and packages freely available and maintained by an impressive community of developers. This paper would not have been feasible without the help of these thousands of contributors. All credits are given on the “References” section, but I would like to send special thanks to the developers of: Slackware (Linux distribution), R (statistical programming language), ggplot2 and ggmap (R packages), rGeos (R package and Geometry Engine), QGIS (GIS Software), sp (R package), Emacs (much more than a text editor) and ESS (Emacs Speacs Statistics). Doing research using these fantastic tools is to be sit on the “shoulders of a giant”. And for free!

† This paper would not have been possible without the indirect support from Bicing’s administration and “Open Data Barcelona”, who kindly provided all the data used in this project. Thanks!

‡ Master Candidate on the Economics of Public Policy Program at the Barcelona Graduate School Of Economics in 2016/2017.
1 Introduction

Bicycle commuters contributes to the social well being by avoiding all externalities that are common to motorized means of urban transportation. Furthermore, there has been an increasing perception and empirical evidence that the health benefits of bike users far exceed the health costs associated with higher injury risks [Woodcock et al., 2014].

As a response to this perception and to the challenges currently presented to urban centers, there has also been an increasing effort from many different cities to stimulate the use of bicycle as a mean of urban transportation. Among others, common public policies implemented include public bike-share systems and construction and maintenance of cycling infrastructure ¹ which increases the perception of safety by segregating cyclists from ordinary traffic.

Even though the number of studies on the impacts of different urban policies focused on promoting bicycle usage in urban center has been growing, there is still lack of clear empirical evidence to support decisions on the most effective policies, such as the exact impact of construction of bike lanes on bike usage.

Among the limitations the lack of quality longitudinal data and the perverse potential of reverse-causality for the decision on the location of new bike lanes appear as two special issues on this topic. Most of the studies on the current literature reach the consistent result of a high level of association between bike lanes availability and bike usage (see for example [Dill and Carr, 2003]), but almost all authors raise warnings on the causal interpretation of this correlation ².

In an attempt to contribute to this literature we use panel data for the public bike share system of Barcelona to analyze the impact of new bicycle lanes that were constructed during 2015 and 2016 on one of the city’s neighborhood in the demand for bicycle trips.

Barcelona has shown over the past few years an increasing number of pro-bicycle programs and policies, in part as a response for the high level of contamination in the city that have been failing to reach the contamination goals set by the European Union ³. Among the different public policies that have been implemented, the city’s administration intends to expand the availability of bike lanes from around 116km up to more than 300km until 2018 ⁴, with the stated goal to allow the majority of its citizens to have a bike lane within a distance of 400m from their houses.

By the other hand, Barcelona has a 10 years old public bike share program which operates with around 420 stations and 6,000 bikes. It has around 105,000 registered users and is responsible for more than 1.1mi trips per month.

¹See for example Midgley, 2011
²For a review on the literature, see J. Puncher and Handy, 2009
⁴From “Dossier de Premsa: Estratgia de la bicicleta”, document from Barcelona’s municipal administration available at https://www.slideshare.net/Barcelona_cat/estratgia-de-la-bicicleta (accessed on 2017/05/22)
accounting for approximately 25% of Barcelona’s commutes on bicycle. This scenario gives us a good opportunity to use geo-informative administrative data collected by the administration of Bicing to estimate the impacts of the constructed bike lanes into the demand for bicycle trips, with panel data that contains 24 months.

It is worth to consider the distinctions between analyzing data that comes from a bike share system and data from the actual number of trips (or stated preferences, like surveys used on some other studies). By one hand, data from bike share programs are more reliable and complete, both because they are company records with important value (such as charging costumers if they exceed a time limit) and therefore have to be treated with more care and because they can very easily be tracked with geo-information. However, there are at least two clear limitations (one general and one specific to our case). The first is that we can no longer talk about estimating “the city’s demand for bicycle commutes” and instead we must somehow focus our attention on the sub-population that use those services, which do not have to be representative of the city’s population. Secondly, it raises questions on whether the results could be driven by the dynamics of the service and not by the response from the user.

We proceed as follows. The next section discusses the two main sources of data we use to construct our analysis and what are their limitations. Section 3 offers some descriptive statistics of the bike lanes and also on the data available for the Bicing’s stations. Section 4 discusses how we construct our treatment and control groups and which lanes we select as treatment, which are key elements to our analysis. Section 5 presents econometric model and estimates considering a single treatment period for all stations, while the following section extends the analysis trying to capture different treatment periods for different stations depending on the inauguration of lanes. Finally, section 7 give some final considerations regarding our results and section 8 concludes.

2 Data

Two different types of data were merged based on their geographic information: the data on bike lanes and the data on Bicing’s stations.

The first set of data was made available by the “Open Data Barcelona” website and consists on geo-referenced data of all bicycle lanes in Barcelona. Specifically, we have monthly data of the number of movements (takes + return) for all 420 mechanical bicing stations.

The second data-set contains data for all 420 mechanical bicing stations. Specifically, we have monthly data of the number of movements (takes + return) for the Bicing’s stations. Section 3 offers some descriptive statistics of the bike lanes and also on the data available for the Bicing’s stations. Section 4 discusses how we construct our treatment and control groups and which lanes we select as treatment, which are key elements to our analysis. Section 5 presents econometric model and estimates considering a single treatment period for all stations, while the following section extends the analysis trying to capture different treatment periods for different stations depending on the inauguration of lanes. Finally, section 7 give some final considerations regarding our results and section 8 concludes.
on every station on the city for all months of 2015 and 2016. The data also had to be complemented with external information. For the bike lanes, we made extensive research to match all the lanes that were constructed and inaugurated between 01/01/2015 and 31/12/2016, trying to capture its exact inauguration date with as much precision as possible. Unfortunately, there is no way to be sure on how precise the dates are. Even if we consider that we were able to match all official inauguration dates, many times the lanes under construction have partial-inaugurations as the work advances and trams of the new bike lanes are ready to use. Furthermore, even if the official inauguration has not happened, many times the construction process creates a segregated (informal) path for cyclists that they can take advantage off. This will have important implications on the analysis of section 6 and lead us to devote a full section to an annual comparison (Section 5) robust to those potential failures.

For the data on the number of movements in all bike-share stations on the city we had to do two types of transformation. The first was to merge this data with the data for the elevation (measured as height in meters compared to the sea level). This was achieved using “Google Maps API” and the coordinate of the stations, which allowed us to get a very good approximation of the elevation for each particular station.

The second type of transformation is more complex. During the months our study took place, several stations suffered from temporary maintenance caused by a diversity of reasons that made then unavailable. Importantly, if a station was out-of-service for a full month we have no data for this station at that time, but if a station had partial services on any month we have the number of movements for this station for the period inside that month that it was operative. Even more worrying is that some stations had their location changed during this time span.

This could lead to different problems, including the existence of many “outliers” and an increase in the volatility of our data, but also the possibility that those stations that were kept “out of service” had a systematic relation with (a) the demand for bicycles and/or (b) the geographic region we are studying, inducing even more worries regarding possible endogeneity problems.

---

9 The data on the number of movements per station was granted by the Bicing’s Administration under some confidentiality terms. The data containing the location of the stations is open and available at Bicing’s official website or “Open Data Barcelona” website (see previous footnote).

10 The primary source of information for this goal was the city’s official website for investments in urban infrastructure [http://ajuntament.barcelona.cat/obres/ca]. This web contained a summary of the construction projects of all lanes constructed in the period of study and, more importantly, the estimated (predicted) inauguration dates of the lanes. The information from this web was considered as a baseline. Further information was matched using posts on social networks (twitter and instagram, specially from user @BicixBCN) which contained photos or videos of the inaugurated lanes that allowed us to be sure that the lanes were inaugurated at least on the publication date. Furthermore, news and official reports from “Ajuntament de Barcelona” were used and given priority when available.

11 For more information on this services go to [https://developers.google.com/maps/documentation/elevation/intro](https://developers.google.com/maps/documentation/elevation/intro) (accessed on 2017/05/22)
To account for that fact we did an extensive research on the historic of public announces from Bicing’s website\footnote{Available at \url{https://www.bicing.cat/ca/avisos/} (accessed on 22/05/2017)}, where all incidents are reported. We checked for any important and long-lasting incidence with stations and checked how those incidents appeared on our database. We dropped all observations where these time spans were correlated with high drops in the movements of the affected stations. Importantly, very few of the drops were located at the geographic region of interest.

Finally, all calculation of geometric distance between any pair of stations and lanes (or length of bike lanes) were made using the tools available at the R package rGeos \cite{Bivand:2017}.

3 Descriptive Statistics

3.1 Lanes

The database of lanes consisted of approximately 138.9km of lanes.

First, lanes were analyzed based on their construction date (before 2015, during 2015 and 2016, after 2016). Once we had matched all lanes that were constructed over the period for which we had available data (2015 and 2016), we selected the lanes that were going to be studied. Our only criteria for the selection of lanes was that it should represent relevant improvement in the connectivity of Bicing’s stations on its vicinity. This process lead us to the situation described in Figure\ref{fig:lanes}.
A representation of all Bike Lanes in Barcelona approximately in January of 2017. The information of lanes constructed during 2017 might not be exact, since it was not the intention of this study. Credits for the background image to Google: [https://developers.google.com/maps/](https://developers.google.com/maps/) for the plot to “ggplot2” [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en). The inauguration dates was researched by the author, which among other sources referred to [http://ajuntament.barcelona.cat/obres/ca](http://ajuntament.barcelona.cat/obres/ca).
Now we comment all categories of lanes that appears on the map:

- **Constructed Before 2015, NOT selected**: those are the lanes that were constructed before 2015 and that are not selected in our study. Those lanes will still be important when considering the control groups, as described in next sections.

- **Constructed Before 2015, selected**: those lanes existed before the time of our study (2015/2016). However, they are clearly related to other lanes that are selected by us. Therefore it would make no sense to treat other lanes as selected and leave these lanes as not selected, since those stations clearly had their connectivity improved by the other lanes. Therefore those stations are treated as selected stations.

- **Constructed in 2015/2016, NOT selected**: Those are the lanes that were constructed during the years of our study but that did not represent a significant connectivity improvement to any neighborhood of the city (nor to any Bicing station). Therefore, those stations are not considered, since they do not change our analysis nor create interest variance in connectivity that can be exploited.

- **Constructed in 2015/2016, selected**: Those are the lanes that did not exist and were fully constructed during 2015 and 2016, and the core of our analysis. They represent a great improvement in the connectivity for the neighborhood where they were constructed (called Les Corts).

- **Improved during 2015/2016, selected**: Those are the lanes that existed before 2015 but that are (a) directly related to the new lanes constructed during this period and (b) were target of renovations and improvements undertaken by the municipality. Therefore these lanes are also selected to our study, since they match our criteria of representing connectivity improvement to this neighborhood.

- **Constructed after 01/12/2016**: Those are the lanes that were constructed during December 2016 and during 2017. They were not considered on our work since we did not have data for the period when they were active.\(^{13}\)

Most of the lanes constructed were two-way bike lanes segregated from motorized traffic by obstacles.

The exact evolution of the lanes, with their respective construction dates, can be seen on figure 2. The evolution of the length of selected lane by date can be seen on Figure 3.

\(^{13}\)Some of those lanes were inaugurated at the beginning of December of 2016, therefore with a 1 month interpolation with our data. However, as there were few lanes and they would intervene with our data in one single month, they were not considered in our analysis.
A representation of Les Corts Neighborhood with the dates of construction of the selected lanes (title of each sub map in the format YYYYMMDD). Credits for the background image to Google: https://developers.google.com/maps/ for the plot to “ggplot2” [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en). The dates were researched by the author (see Section 2).
Even though the affected region was not directly connected to the network of bike lanes in the city, it could be questionable whether the urban infrastructure in this area of the city already allowed for a relatively safe bike usage before the interventions. If that is true, we should expect the impact of the bike lanes to be much lower than it would have been otherwise.

Indeed, most of the roads of this neighborhood was already considered as “pacified” roads by the municipality administration. These roads contain vertical signs stating preference for bicycle over vehicles and have their speed limit set to 30km per hour. The zone had already suffered these transformations on years previous to the implementation of the bike lanes. Figure 12 (Annex A) shows a map of the affected region with those roads.

It is clear that there was already some level of connectivity between the neighborhood and the network of bike lanes before the creation of the selected lane. Therefore, we should consider the treatment we try to estimate in this paper as an “improvement” in this connection, what might be useful to interpret our estimations.

Finally, an important issue for the identification strategy of this paper is what led the municipality to choose this area of the city to expand the network of bike lanes. If this choice was made by a prediction that in this area the demand for bicycle trips would increase more than in the rest of the city it would be impossible with our data to separate which part of the effect comes from the bike-lanes implemented or from this pre-existent tendency towards higher bicycle usage.

However, there are no signs that there was any special reason to expand the bike lanes in this region due to a higher expected demand for bicycle commutes. The strategy of the municipality administration appears to be a generalized expansion of the network, as stated in section 1. As the map of lanes show, this city was a very good candidate to start this expansion, since it was already partially connected to the network by a few lanes. Therefore we have no signs of endogeneity in the choice of this region.

### 3.2 Bicycle Stations

In all our analysis our dependent variable is the number of movements on each bicycle station of the city at each month.

As movements we understand all takes and devolutions of public bicycles in any station, both summed up without any distinction. Importantly, not all bicycles moved from one station to another represents one trip made by one user of Bicing. The company has some trucks with employees that move the bicycle around the city trying to match supply and demand, what accounts for roughly 20% of the total movements according to Bicing’s administration. In our data we can not distinguish whether a movement was done by an user of by one of Bicing’s employees.

This non-user movements would introduce potential challenges to identification of a treatment effect for the stations near the new bike lane if there was a coordinated policy of redistributing more bicycles to stations connected by a
Figure 3: Evolution of KMs of selected lane by date

The black line represents the accumulated sum of kilometers of selected lane that were constructed or modified in each date (on the date or before). The length of the horizontal bar goes from January/2015 up to December/2016. The blue horizontal line dividing the graph represents the first day of 2016. The red dotted horizontal lines represents the inauguration dates of the selected lanes.
However, Bicing’s only purpose with the redistribution of the bicycles around the city is to increase the likelihood that a new potential user will find a bicycle in its origin and an empty spot in its destination station.

In other words, the redistribution effort is a response to the demand for bicycles instead of a confounding factor systematically correlated to the demand and the bike lanes. Therefore, it plays in favor of our identification strategy.

On figure 4 we plot a sequence of boxplots, one for each month. In the graph we can see the distribution of the movement in the stations along the year, with important seasonal patterns in the months close to August and to the end (or beginning) of each year. In figure 5 we also plot the average movement with the same patterns repeated.

There is a total of 420 mechanical stations spread across the city. Figure 8 (Annex A) plots a heatmap with the average number of movements for each station for the year of 2015. Figure 9 (also in Annex A) plots the evolution of the number of movements in each station comparing the second semester of 2015 with the second semester of 2016.

On the first graph we can see that the are close to the selected lanes are an area of lower movement when compared to the center of the city. This might be considered a signal that those stations were not supply-constrained and could absorb extra demand possibly created by the intervention on the selected lanes. We can also check that the selected lanes are responsible for connecting one of the few parts of the city that had stations and were not integrated into the network of lanes.

The second heatmap showing the evolution between the second semester of 2015 and 2016 has a much less clear pattern. There appears to be some correlation between the selected lanes and stations with higher increases, but no evident conclusion could be reached by a graphical analysis.

4 Treatment and Control Groups

We now focus in describing how we define the groups of stations that we consider as “treated” and the (different) groups of stations we consider as controls. This is a crucial part of our analysis since most identification hypothesis will be made upon comparing the treated and control groups over some dimensions or characteristics.

---

14 This would cause a challenge because if Bicing decided focus its bicycles redistribution efforts in the region of study as a consequence of the construction of the lanes, it would be impossible (at least without more data) to distinguish which part of the potential increase in the usage of those stations was a result of this redistribution effect and which part was due to the new connected lanes.

15 To see why, imagine that the stations in the region under study were over-demanded before the implementation of the new lanes. In a scenario with no redistribution, we would not capture any effect, since even if the lane increased the demand for trips the supply could increase no further. However, if Bicing responds to excess of demand with redistribution, it is really allowing for the supply to increase (a little) as demands also increase.

16 Bicing also has electric bicycle Stations, not considered in this study.
The y-axis represents the total number of movements in a station. The x-axis represents the months in the format YYYYMM. Each boxplot represents the dispersion of the number of movements in the stations in one month. The outliers are categorized by being further to the box than 1.5 times the interquartile range.
Figure 5: Monthly average of the number of movements on bicycle stations

The y-axis represents the average number of movements in a station. The x-axis represents the months in the format YYYY/MM. The red vertical lines represent the months of December, while the blue represent the months of August.
One important concept to introduce is what we call over the next sections as **connection distance** (or, when implicit, just distance). In general terms, this is the maximum linear distance in meters from a station to a lane as to say that this station was **connected** by this lane. Therefore, for every lane in the city, we can imagine a buffer (polygon) around this lane with margins that are exactly \( m \) meters distant from the closest point of the lane. All stations inside this buffer are said to be **connected** to (or by) this lane.

Selecting an adequate distance is not straightforward nor has much consensus in the literature. For Barcelona one possible candidate would be 400 meters, since one of the city’s goals is that 95\% of it’s population to have a lane within this distance from their homes. However, a result obtained considering this distance could be dramatically different than another one considering 200m or 800m, what could lead us to achieve wrong conclusions.

Our strategy in this delicate topic is to run every analysis we do with four different connection distances: 200m, 400m, 600m and 800m. The exact usage and impact of those distances will become clear over the next sub-sections.

There are at least two important consequences of our choice of re-running the experiment for different distances. The first is that it makes the exposition of the results more complex and the interpretation harder. The second (and maybe most important) is that it creates a source of variation that sometimes can affect our results from different channels that can not be easily isolated from each other.

### 4.1 Treatment Group

For defining the treatment group we first consider one further distinction between the lanes in the city. We split all lanes in the city into a category of **connected lanes** and a category of **unconnected lanes**. This distinction tells us whether each lane was uninterruptedly connected to the main network of lanes in the city. The map resulting from this categorization can be seen on the Annex (Figure 10).

With this distinction in mind, as well as the definition of a connection distance, we can write a definition for a **treated station**. Consider a connection distance of \( m \) meters. A treated station is any station that, considering all selected lanes (i.e. considering the last month when all stations are constructed), satisfies at least one of these two conditions:

1. Be closer to a selected lane than \( m \) meters **and** be closer to a a selected lane than from any other non-selected and connected lane.
2. Be closer to a selected lane than 100\( m \), regardless of any other lane.

The intuition behind the definition is simple: we want to capture all stations that had their connectivity to the city improved as a consequence of the selected lanes. The reason why we drop stations that are closer to other connected lanes than to a selected lane is because this station was most probably already connected before. And the reason we create a cutoff of 100\( m \) within which all
stations are treated is to allow for a lane that is constructed just close to a station to affect this station regardless of its previous status. What this cutoff does in practice is to allow for some stations in the “fronteer” between the selected and the old lanes in Figure 8 (Annex A) to be considered as treated.

From now on, we call the treated (or treatment) group the group of all treated stations.

Notice that for each definition of distance $m$ we have a different treated group, and as $m$ grows the number of stations in the treatment group also grows, always by adding new members to the set of treated.

A second important concept is the concept of treatment period. A definition of treatment period for a station $i$ is the first period for which the station $i$ meets the definition of a treated station considering only selected lanes with inauguration date previously than this period.

Therefore another important consequence of increasing the connection distance is that the treatment periods potentially change for the same stations in treatment groups with different distances.

Table 1 shows the number of stations in the treatment group in each date considering different connection distances. The Figure 11 (Annex A) shows the composition of the treatment groups with different connection distances.

4.2 Control Group

We now turn our attention to the control group. For robustness, we propose three different definitions of control groups and run all our analysis with the three different definitions.

4.2.1 Control Group 1: All

The first control group (referred as “All”) contains every single station that does not belong to the treatment group.

This control groups provide us few variation as the connection distance increases. The only change that happens when a higher distance is considered is that some stations leave the control group and go to the treatment group.

Maps with the group composition are available at Figure 13 in the Annex A.

---

17 There is on subtle point in this definition. For the inauguration of lanes we have the exact date of inauguration, while for the demand we only have montly data. To come from the period in day to a period in month that can be used to define regression dummies we do the following: for each treated station we calculate the treatment period $p$ (day $d$, month $m$) following the definition we just described. If the day is before or equal to the 15th day of the month ($d \leq 15$), we consider month $t$ as the treatment month. If the day is after the 15th day of the month we consider the following month ($t + 1$) as the treatment month.

18 For example, consider a station that with a 200m connection distance is only treated (using our previous definition) to a lane inaugurated at 2016/09. But if a distance of 400m is considered, the treatment definition also matches another lane inaugurated at 2016/05. Therefore the period of treatment for this station was 2016/09 for the group formed with a 200m buffer and 2016/05 for the group with 400m.
Table 1: Number of treated stations per date considering different connection distances

<table>
<thead>
<tr>
<th>date</th>
<th>200m</th>
<th>400</th>
<th>600</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td>20150115</td>
<td>0 stations</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20150215</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20150315</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20150415</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20150515</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20150615</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20150715</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20150815</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20150915</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20151015</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20151115</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20151215</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>20160115</td>
<td>8</td>
<td>14</td>
<td>21</td>
<td>33</td>
</tr>
<tr>
<td>20160215</td>
<td>8</td>
<td>14</td>
<td>21</td>
<td>33</td>
</tr>
<tr>
<td>20160315</td>
<td>8</td>
<td>14</td>
<td>21</td>
<td>33</td>
</tr>
<tr>
<td>20160415</td>
<td>8</td>
<td>14</td>
<td>21</td>
<td>33</td>
</tr>
<tr>
<td>20160515</td>
<td>12</td>
<td>22</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>20160615</td>
<td>12</td>
<td>22</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>20160715</td>
<td>12</td>
<td>22</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>20160815</td>
<td>12</td>
<td>22</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>20160915</td>
<td>27</td>
<td>41</td>
<td>44</td>
<td>51</td>
</tr>
<tr>
<td>20161015</td>
<td>27</td>
<td>41</td>
<td>44</td>
<td>51</td>
</tr>
<tr>
<td>20161115</td>
<td>27</td>
<td>41</td>
<td>44</td>
<td>51</td>
</tr>
<tr>
<td>20161215</td>
<td>27</td>
<td>41</td>
<td>44</td>
<td>51</td>
</tr>
</tbody>
</table>

Dates in the first column are in the format YYYYMMDD. The 15th day of the month is considered because according to our definition any treatment day that occurs after the 15th day of a month is computed for the following month (see footnote on the previous page). The name of the columns are the connection distance in meters considered in the formation of the treatment group.
4.2.2 Control Group 2: Unconnected

The second definition of control group is the set of stations in the city that (a) is not part of the treatment group and (b) is unconnected to any non-selected and connected lane in the city (further away than \( m \) meters). As a consequence, for each different connection distance considered for the formation of this group, a different group of stations will be selected, which is necessarily a subset of the previous group. Stations leave this group as \( m \) increases from \( m_1 \) to \( m_2 \) by two reasons: either because they become treated stations when using \( m_2 \) as a connection distance or because they become connected to some lane when considering \( m_2 \) instead of \( m_1 \). On this first case the station is included in the treatment group, while in the second case the stations is just dropped from the sample.

Some remarks are worth making for this group. The intuition behind its definition is that it makes more sense to compare our treated stations with other stations that have always been unconnected than compare stations that had been connected before. However, as the connection distance used to construct this group grows, the number of observations and the geographic dispersion of the stations increase, raising questions on whether this would be a good comparison group for high connection distances.

Finally, notice that we use the same distance \( m \) to compute the treatment and this control group, so and increase in \( m \) leads to simultaneous changes in the treatment group and in this group.

Maps with the group composition are available at Figure 14 in the Annex A.

4.2.3 Control Group 3: Neighbors

Our last control group is refereed as the “Neighbors” group. This control group is constructed following these steps:

1. With a connection distance \( m \), prepare the treatment group.

2. With the treatment group, find the number \( n \) of stations that belong to this group.

3. Compute the distance from all other non-treated stations to their closest treated stations.

4. Select the \( n \) stations with the lower calculated distance.

Therefore, what we do is simply to create a set of stations that are neighbors to the treated stations. We force both groups to be of same size, so high connection distances will create larger treated groups that will be compensated by a larger control group of “Neighbors”.

As the connection distance considered grows, what happen with this control group is that most of the stations that were included as control migrate to the treated group, and new stations are added to the control to compensate.
Maps with the group composition are available at Figure 15 in the Annex A.

5 Models with Single Treatment Period

Before allowing for more complete models to estimate the treatment effect we start with an intuitively diff-in-diff analysis with a single treatment period followed by a fixed-effect models also with a single treatment period. These analysis will serve as baseline results against which other models will be compared.

We use $i$ to index stations and $t$ to index time periods. Let $D_i$ be a dummy variable for station $i$ that tells us if it belongs to the treated group and $d_t$ be a time dummy variable that has value 1 if period $t$ is after the treatment period.

Recall that the last selected lane had its modification date at September of 2016. Therefore for the rest of this section we define out dummy for period of treatment to be:

$$d_t = \begin{cases} 
1 & \text{if } t > 2016/09 \\
0 & \text{otherwise} 
\end{cases} \quad (1)$$

The intuition behind this first definition is to consider all the lanes as a single “treatment” instead of identifying different treatment dates for each treated station based on the dates of inauguration of the lanes. This single treatment represents the first date when all lanes were constructed. Even though this strategy does not make use of all information available at our database it has at least two advantages.

The first and powerful advantage is its simplicity in definition and interpretation of results. This allows, for example, a first approach to understand how the treatment effect changes among the different definitions of treatment and control groups (and its combinations).

The second advantage is that by treating the inauguration of all lanes as a single treatment we simplify the understanding of treatment. Instead of considering treatment as “being close to a selected bike lane” we can now define treatment as “being connected to the city’s network of bike lanes”. There may be an important difference in those two definitions, since the simple existence of a lane does not guarantee a high degree of connectivity.

Importantly, notice again from Figures 2 and 3 that almost all selected lanes have their modifications very close to September/2016. Therefore there is a clear connectivity gain around this date, which should be reflected in the movement of the stations if the lanes had a real impact on them.

Most of the results in this section depends on the common trend assumption. In other words, we need that in the absence of treatment both group of stations (treated and untreated) would have followed the same trend on the number of movements. A graphical representation of this hypothesis is present on graph 6.

---

19See Figures 2 and 3
where we plot the evolution of movements (as index number) for the treatment
and control groups using different connection distances and specifications.

From the figure we can perceive that the common trend is more reliable
for some groups than for other. As expected, the control groups “Neighbors”
seem to be the ones that best fit the evolution of the treated group up to the
treatment date.

Importantly, we can see that there are no graphical signs of different trends
before the treatment date we are considering in this section, what plays in favor
of the understanding that the treatment effect we are considering might be
important in increasing the connectivity of the region.

Finally, it is clear in the graph that there are important seasonality patterns
that differ among the groups. This point will be further discussed when defining
our econometric models, which will attempt to control for different seasonality
patterns for different groups of stations.
Figure 6: Evolution of movements in treatment and control groups (2015/01/01 = 1)

Evolution of movements per group definition and connection distance. The vertical split of the grid represents different control groups (“All”, “Neighbours” and “Unconnected”). The horizontal split represents different connection distances used to construct each group (200m, 400m, 600m). The vertical red line represents the treatment date (2016-09-01). All lines are in index points showing evolution, with base 1 = 2015-01-01
5.1 Differences in Differences

Our first approach is to compare the total number of movements in the stations in the last 4 months of 2016 (i.e. when all stations in the treatment group are connect by lanes) with the last 4 months of 2015 (i.e. when almost all stations are still unconnected). As described in previous sections, less than 400m of selected bike lanes were constructed in 2015. It is therefore reasonable to assume that in the last months of this year the connectivity of Les Corts had increased very little.

Let SY it represent the sum of the movements in the 4 last months of each year (t equals 2015 or 2016) for station i. Here dt is a dummy with value 1 if observation is after the treatment date (therefore, equals 1 for 2016 and 0 for 2015). Therefore we implement a simple differences in differences strategy using the following regression specification:

\[ \ln(SY_{it}) = \beta_0 + \beta_1 D_i + \beta_2 d_t + \beta_3 (D_i \ast d_t) + \epsilon_{it} \]  

The correct identification of the causal treatment effect in this setting depends directly on the common trend assumption between control and treatment groups, meaning that in the absence of treatment both groups would have followed a similar trend.

The coefficient of interest is \( \beta_3 \), which we report altogether with heteroskedastic robust standard errors in table 2 for the different specifications of control groups, constructed using different connection distances reported in the first column.

In this simple framework, and because we have simplified our data to have two periods, the reported coefficients represents the classical differences in differences framework, reporting the difference in the evolution of the means of the treated and control groups of each specification. As out dependent variable is in natural log, the effect can be interpreted approximately as percentage points.

The first point to notice is that the standard errors are very large in all specifications, what is probably due to a lack of explanatory power of the model. None of all estimated coefficients have significance at any acceptable level.

All estimates in all groups and for all distances have positive values, which tells us that we are looking to a region of the city that, in this annual comparison, have grown more than any of our control regions. The point estimates goes from close to 2% up to 6.6%. The effects are somewhat larger and robust for the first control group, and more volatile for the other groups. This comes with no surprise, since the first control group is the one less affected by the change in the connection distance considered.

20) The decision to sum the number of movements by station in the same months of different years is, again, an effort to pursue simplicity. By doing so we are comparing the same months of different years and as a consequence we do not have to care about any seasonality effects that may occur in specific months.

21) We drop from our sample any station that do not have data for any month in these two periods. In the whole city we loose observations on 30 stations, which represents 7% of our sample.

22) Most Adjusted \( R^2 \) are very low, lower than 0.05.
Table 2: Differences in Differences with single treatment for different control groups and connection distance

<table>
<thead>
<tr>
<th>Distance</th>
<th>C. All</th>
<th>C. Unconnected</th>
<th>C. Neighbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>d200</td>
<td>5.53 [15.96]</td>
<td>5.90 [18.58]</td>
<td>6.64 [23.88]</td>
</tr>
<tr>
<td>d800</td>
<td>5.07 [15.83]</td>
<td>3.19 [22.52]</td>
<td>4.61 [22.41]</td>
</tr>
</tbody>
</table>

P-values: (***) Lower than 0.01; (*) Lower than 0.05, (.) Lower than 0.1

Estimated treatment effect from equation (2). The dependent variable is the log of the sum of movements for each station in the 4 last months of each year (2015 and 2016). We report the value for the variable \( \beta_3 \) that represents the treatment effect, and the standard errors (robust to heteroskedasticity) in parentheses. Both are multiplied by 100. The first column reports the “connection distance” considered to create each group (in meters), as described in previous sections. For the definition of the groups, see Section 4.

This table gives us a first piece of evidence suggesting that the stations affected by the new lanes have a higher growth rate, and that this effect seems robust to different specifications. However the standard errors also suggest that the data seem to suffer from a high volatility, not explained in this simple framework.

5.2 Fixed Effects Model

Now we turn to estimate a model controlling for time and station fixed effects. Let \( t \) represent each month of 2015 and 2106 once again. Let \( Y_{it} \) represent the movement of station \( i \) at time \( t \). We estimate the following model:

\[
\log(Y_{it}) = \beta_3(D_i \times d_t) + \gamma_t + \alpha_i + \epsilon_{it} \tag{3}
\]

Where \( \gamma_t \) and \( \alpha_i \) are estimated using two sets of dummies for each month and for each station.

Also, because our data is made of monthly data and there appear to be clear patterns of seasonality, we also expand the model to allow the treated stations to have a different seasonality patterns.

It is worth to make some considerations on the seasonality patterns. Among different reasons for seasonality, such as vacations, climate, etc, one important element to understand seasonal patterns is the presence of the sea and the elevation of the city. Barcelona has a clear pattern of increasing elevation when coming from the sea into direction of the mountains in the back of the city. In the months closer to summer, there is a clear pattern of very high increase in the usage of the stations that are closer to the sea (lower elevation) and reduced demand for stations that are further away from it (higher elevation).
Table 3: Fixed-Effects estimates with single treatment period for different control groups and connection distances

<table>
<thead>
<tr>
<th>Distance</th>
<th>C. All</th>
<th>C. Unconnected</th>
<th>C. Neighbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>d200</td>
<td>4.18* [1.84]</td>
<td>4.45* [2.04]</td>
<td>5.69* [2.41]</td>
</tr>
<tr>
<td>d400</td>
<td>2.27 [1.65]</td>
<td>0.56 [2.06]</td>
<td>2.26 [2.21]</td>
</tr>
<tr>
<td>d600</td>
<td>1.75 [1.59]</td>
<td>-0.32 [2.21]</td>
<td>0.75 [2.12]</td>
</tr>
<tr>
<td>d800</td>
<td>2.86. [1.63]</td>
<td>1.67 [2.46]</td>
<td>2.22 [2.08]</td>
</tr>
</tbody>
</table>

p-values: (***) Lower than 0.01; (*) Lower than 0.05, (.) Lower than 0.1

Estimated treatment effect from equation 3. We report coefficient $\beta_3$ with heteroskedastic-robust standard errors clustered by station in parenthesis. Both are re-scaled (multiplied by 100)

To illustrate this point we include Figure [16] (Annex A) that shows the relative movements in the stations of the city comparing August (vacations and summer month) with April. There appears to be a visual relation between distance from the sea (and, therefore, elevation) and the relative movements in stations comparing the two months.

As our treatments happens on specific months at the year, it is important to take this seasonality pattern into account. We do so by including a polynomial equation of elevation (in meters from the sea level) interacted with a dummy variable for each month. Let $MONTH_{m,t}$ be a dummy variable that equals 1 if observation $t$ belongs to month $m$ and let $E_i$ be the elevation of station $i$ in meters. Therefore we add the following term to equation 3:

$$\sum_{m=1}^{12} \sum_{k=1}^{3} \pi_{m,k} * E_{i}^{k} * MONTH_{m,t}$$  \hspace{1cm} (4)

This specification allows seasonality effects to be different for each station depending on their elevation.

The identification on this setting depends on weaker assumptions than in the differences in differences estimation. Here we need the treatment to be independent of any other station characteristic that changes over time. If, for example, the decision for the location of the new bike lanes was based on time-fixed geographic characteristics that also affected the number of bicycle trips (e.g. elevation, roads design, neighbor infrastructure, etc.) those effects would be captured by the stations fixed effects and would not introduce bias in our estimators.

The results for $\beta_3$ (multiplied by 100) for equation 3 are reported on table 3 and at table 4 for the model with seasonality dummies.

The standard errors for this model are much lower than the previous model. The coefficients keep their positive sign in most specifications. In all specifications of control groups the only statistically significant coefficients are the one related to the groups formed by a connection distance of 200m. The results for this distance are robust and consistent, with the point estimates between 4%
and 6% in all specifications.

The specification allowing for different seasonal patterns for the treated stations have very little impact in all coefficients and also helps giving robustness to our estimates.\(^{23}\)

In summary, all results in this chapter points towards two robust conclusions. Using a single treatment period for all stations leads to a significant point estimate above 4.2% when considering very short connection distances. For higher distances, there is no clear sign of any effect.

6 Regression Models With Multiple Treatment Periods

Now we allow different stations to have different treatment dates. For careful definition of treatment period see section 4.1. Now we can use the notation \(D_{it}\) as a dummy variable that has value 1 if the treatment period of station \(i\) is before or equal to period \(t\).

By using this definition we must now understand treatment as the effect of a lane on nearby stations. We can expect the measured treatment effect to be lower than in the previous section, when we considered all lanes as a single treatment.

6.1 Fixed Effects with Multiple Treatments

The first model we consider is the same fixed effect model used with a single treatment period (Equation 3), but where we substitute the treatment definition to the new treatment dummy.

\[
\log(Y_{it}) = \beta_3 \times D_{it} + \alpha_i + \gamma_t + \epsilon_{it} \tag{5}
\]

\(^{23}\)In all specifications the coefficients related to the included polynomials are highly significant.
Table 5: Fixed effects regression with multiple treatment periods for different connection distances and control groups

<table>
<thead>
<tr>
<th>Distance</th>
<th>C. All</th>
<th>C. Unconnected</th>
<th>C. Neighbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>d200</td>
<td>3.92* [1.6]</td>
<td>3.79* [1.75]</td>
<td>4.37* [1.99]</td>
</tr>
<tr>
<td>d400</td>
<td>1.29 [1.66]</td>
<td>-0.53 [1.79]</td>
<td>1.23 [1.85]</td>
</tr>
<tr>
<td>d600</td>
<td>-0.31 [1.44]</td>
<td>-2.42 [1.69]</td>
<td>-1.04 [1.66]</td>
</tr>
<tr>
<td>d800</td>
<td>2.79 [2.14]</td>
<td>0.85 [2.22]</td>
<td>2.06 [2.22]</td>
</tr>
</tbody>
</table>

p-values: (**) Lower than 0.01; (*) Lower than 0.05, (.) Lower than 0.1.

Point estimates of treatment effect ($\beta_3$) in equation 5 and heteroskedastic-robust standard errors clustered by stations in parenthesis, both multiplied by 100.

Table 6: Fixed effects model for multiple treatment periods (with control for heterogeneous seasonal effects) for different connection distances and control groups

<table>
<thead>
<tr>
<th>Distance</th>
<th>C. All</th>
<th>C. Unconnected</th>
<th>C. Neighbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>d200</td>
<td>4.34** [1.62]</td>
<td>4.13* [1.78]</td>
<td>4.40* [2.18]</td>
</tr>
<tr>
<td>d400</td>
<td>2.27 [1.64]</td>
<td>0.51 [1.77]</td>
<td>1.98 [1.87]</td>
</tr>
<tr>
<td>d600</td>
<td>0.91 [1.45]</td>
<td>-1.42 [1.66]</td>
<td>-0.11 [1.68]</td>
</tr>
<tr>
<td>d800</td>
<td>3.85 [2.2]</td>
<td>2.29 [2.28]</td>
<td>3.72 [2.26]</td>
</tr>
</tbody>
</table>

p-values: (**) Lower than 0.01; (*) Lower than 0.05, (.) Lower than 0.1.

Same as Table 5, but controlling for season dummies interacted with the dummy for treatment group.

The assumptions needed to interpret the estimation of the treatment effect are the same as the fixed effects models for the single treatment period.

As before, the equation is estimated using dummies for each station and time period, and the standard errors are robust to heteroskedasticity and clustered by station. We also expand the model to allow for heterogeneous seasonality effects, defined in the same way as in the previous section (see Equation 4). The results are reported on table 5 and 6 for the model with season effects.

The results follow exactly the same pattern as when considering a single treatment effect. The only significant results at a 5% confidence level are at the 200m distance, and it is robust to all control group specifications, lying close to 4%. In this case the inclusion of the seasonality dummies as controls increase a little the estimated effects. All specifications of control groups leads to approximately the same conclusions.

6.2 Test for Granger Causality

Now we move to a model in the spirit of Granger Causality. Let $d_{it}$ be a dummy variable indicating whether station $i$ was treated exactly at time $t$. Also, let $d_{i(t-s)}$ be the value of $d_{it}$ for the same station $i$ $s$ months before and let $d_{i(t+s)}$
be the value for $d_{it} s$ months after. Finally, let $d_{i(t-3+)}$ be a dummy equal to 1 if the station was treated 3 months ago or more.

We estimate a model similar to the fixed effects model on the previous section (equation 5), but we add lags and leads for the station-specific treatment period dummy and the controls for seasonality effects.

\[
\log(Y_{it}) = \beta_0 d_{i(t-(3+))} + \beta_1 d_{i(t-2)} + \beta_2 d_{i(t-1)} + \beta_3 d_{it} + \beta_4 d_{i(t+1)} + (\ldots) + \beta_7 d_{i(t+4)} + \sum_{m=1}^{12} \sum_{k=1}^{3} \pi_{m,k} \ast E_{k}i \ast MONTH_{m,t} + \alpha_i + \gamma_t + \epsilon_{it} \quad (6)
\]

We would expect the leads ($d_{i(t+s)}$) to have no explanatory power in this regression, since there is no treatment at this period for this station. Furthermore, we would expect that the coefficients for the lags ($d_{i(t-s)}$) to represent the treatment effects for the months after treatment, therefore they should have a positive effect. This approach allows us to have some intuition on how the effect is spread over time for the treated stations. We plot the estimated coefficient for each lag and lead, as well as the coefficient for $d_{it}$, on graph 7.
Figure 7: Coefficients for treatment lags and leads in spirit of Granger Causality

The vertical division of the grids represents different controls groups (All, Neighbors and Unconnected, in this order). The horizontal splits of the grid represents different connection distances used to construct the control and treatment groups (200, 400, 600 and 800). The points are the estimated coefficients for \( d_{i(t+4)} \), ..., \( d_{i(t+1)} \), \( d_{it} \), \( d_{i(t-1)} \), \( d_{i(t-2)} \), \( d_{i(t-(3+))} \) from equation 6 (in this order). The red horizontal lines represents the 0 y intercept. The confidence interval around the point estimates have 95% confidence level and were constructed with heteroskedastic-robust standard errors clustered at the stations.
We can see that the estimation for most of our specifications do not reveal any clear pattern. The expectation of positive estimates for the coefficients regarding the treatment period and treatment lags only appear in a few graphs, such as the ones considering a short distance (200m), while in most other there appears to be no sign of this effect. This is however consistent with our conclusions so far that points towards a robust treatment effect for very short connection distances and no clear effect for higher distances.

Looking at the graphs for a 200m connection effect, we can see that all estimates for the lags are significant at the 5% confidence level, what follows our expectation. Therefore we focus our analysis in those graphs.

The coefficient for the treatment period is usually not significant. Two explanations are possible: there is more effect in subsequent months than in treatment month or this is a consequence of small error on the measurement of the exact treatment date. We can not distinguish these effects.

For some specifications leads are also significant and even higher than posterior months, what would not be expected. Here we can find other two possible explanations. First, it could be the case that these coefficients are capturing the effect of lanes constructed before the station-specific treatment period that marginally affected other stations considered non-treated by this lane. Also, it can (again) be explained by small mismatches between actual treatment date and the treatment date we are considering (see section 3.1).

### 7 Further Considerations

All results of this paper suggest positive effects higher than 4% for the stations near the selected lanes when considering very short connection distances. Some considerations on this results are welcome.

First, there is some evidence in the literature that the construction of a lane not only creates demand but displaces demand from other paths into the lane. Following this argument, it is possible that the construction of lanes would induce users of the bike share system to use a different station to take/return their bicycles. If users give preferences to paths with a bicycle lane, they might as well give preferences for stations that are close to them.

A consequence of this fact is that we can not distinguish whether this robust effect of a bike lane into the demand for close stations is simply a displacement effect or a creation of demand for bike trips. If anything, we could argue that we have some evidence of a displacement effect.

The first evidence is that we only find impact for the shortest distance considered, which is consistent with the displacement hypothesis because the treated stations close to the lanes would “steal” demand from others. Once the connection distance increases, this effects balances out and we find no treatment effect.

---

24 To better understand this argument take another look at Figure 3. It could be the case (for example) that the lanes inaugurated in 2016/05 had some effects on the stations treated on 2016/08. If that is true, the leads could be capturing this effect.
A second consideration in the same direction is to look closer to the results using the control group “Neighbors”. We can see that in almost all specifications this group has higher estimated effects than the other groups (for all distances). This is, again, consistent with the displacement effect, since in this group we are comparing the stations that would receive demand (treatment) against the stations that would lose demand (control).

8 Conclusion

We have analyzed the effect of the construction of bike lanes in Barcelona over the number of trips in the public bike-share system. In the pursuit of robust results we considered three different specifications for control groups and four different connection distances, as well as different econometric models. In the first part of our empirical strategy we show that the movements on the stations close to the lanes had grown more than on the respective control groups in an annual comparison. We then tried to improve the explanatory power of our model and to relax some assumptions, checking whether the effect was significant and exploring the time variance of our data. Finally we moved to a different definition of treatment, considering lane-specific impacts.

The most robust results are found when considering very short connection distances (200 meters), where virtually all different estimation techniques and control groups specifications led us to find positive and significant treatment effects. The value of the point estimates were also robust to those specifications, with effects varying from 3.8% to 5.7%.

For higher connection distances the estimations become less precise. There is a common and robust pattern of a higher decrease in the estimated treatment effect as connection distance increases from 200 to 400 or 600m. The same patterns shows that the estimation increases again for an connection distance of 800m, but without statistical significance.

A possible explanation to this pattern is that the creation of a bike lane not only increase the demand for the stations nearby but also displace some of the demand from other stations at the neighborhood. Therefore very short connection distances would capture only the stations that receive an increase in the demand, while average connection distances (400 and 600 meters) would capture both the stations that lost demand and the stations that received the same demand, leading to null effects.

Considering the different specifications for control groups, the group of unconnected stations on the city appeared to give the most volatile results. As argued before, this is probably due to a high geographical dispersion and reduced number of stations considered when the connection distance grows. It seems that among the three groups considered the results using this group should be treated with more caution than the others.

As a conclusion, we have very strong evidence of impact for short connection distances, but the true source of this treatment effect is still unknown. Further research with more detailed data (e.g. with the station-to-station trips) would
allow for more detailed conclusions.

References


9 Annex A: Figures and Maps
Figure 8: Average month movement on stations in 2015

All bicing stations in Barcelona with their color representing the average movements for each station in 2015. Credits for the background image to Google: [https://developers.google.com/maps/] for the plot to “ggplot2” [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en).
Figure 9: Evolution on monthly movements on stations between the second semester of 2015 and 2016

All bicing stations in Barcelona with their color representing the evolution in the average movements for each station in 2016 compared to 2015 (base). The black dots are stations with missing data for one or both periods. Credits for the background image to Google: https://developers.google.com/maps/ for the plot to “ggplot2” [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en).
Figure 11: Stations belonging to the treatment group considering different connection stations

Each map has the stations belonging to the treatment group constructed with the distance indicated in the maps sub titles. Credits for the background image to Google: https://developers.google.com/maps/; for the plot to “ggplot2” [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en).
Figure 12: Bike lanes and bike-friendly zones in the affected region

Each map has the stations belonging to the control group “All” constructed with the distance indicated in the maps sub titles. Credits for the background image to Google: [https://developers.google.com/maps/](https://developers.google.com/maps/) for the plot to “ggplot2” [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en).
Each map has the stations belonging to the control group “Unconnected” constructed with the distance indicated in the maps sub titles. Credits for the background image to Google: https://developers.google.com/maps/ for the plot to “ggplot2” [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en).
Figure 15: Stations belonging to the control group “Neighbors” considering different connection stations.

Each map has the stations belonging to the control group “Neighbors” constructed with the distance indicated in the maps sub titles. Credits for the background image to Google: [https://developers.google.com/maps/](https://developers.google.com/maps/) for the plot to “ggplot2” [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en).
Figure 16: Relative movement per station in August and April (averages for 2015 and 2016)

Each dot represents a station and the color represents the ratio of the average movement in August divided by the average movement in April for the years of 2015 and 2016 (\(\text{avg}(\text{movements}(\text{August})) / \text{avg}(\text{movements}(\text{April}))\)). Grey points represents missing data. Credits for the background image to Google: [https://developers.google.com/maps/](https://developers.google.com/maps/) for the plot to "ggplot2" [Wickham, 2009] and for the geo-coordinates to Open Data Barcelona (http://opendata-ajuntament.barcelona.cat/en).