Pedestrianization and Business Visits: Evidence from NYC Open Streets*

Timur Abbiasov[1], Iain Bamford[2], and Pablo Warnes[3]

[1]Senseable City Lab, MIT
[2]Uber
[3]Department of Economics, Aalto University / Helsinki GSE

March 14, 2024

Abstract

There are significant debates in urban planning on the use of road space in cities. Should (some) streets be pedestrianized? Critics suggest closing streets to vehicles can harm local businesses by reducing access. The effect of pedestrianization on business visits has been difficult to assess due to the lack of an appropriate experiment and lack of systematic data on foot traffic. We examine a unique recent experiment, New York City's Open Streets program, which closed hundreds of street segments to cars, and utilize new anonymized cellphone geodata to measure visits to businesses. Using a spatial difference-in-difference design in which sites for counterfactual candidates for treatment are selected using convolutional neural networks, we find that an opening of an open street results in a 9% higher number of visits to businesses located on that street, contradicting critics' predictions. We find a more pronounced 12% increase in visits in neighborhoods with a low share of the workplace population. Analysis by industry reveals that restaurants, bars and amusement & recreation establishments are the most likely to benefit from the program. We also find evidence of positive spillover effects to neighboring streets.

*Corresponding author: pablo.warnes@aalto.fi.

1 Introduction

There are significant debates in urban planning on the use of road space in cities. In recent years, some cities have attempted to promote a shift away from car usage in favor of alternative transport means, such as walking and cycling (Buehler et al., 2017). These planning interventions often involve changing the designated uses of existing road space, by creating bicycle lanes, bus lanes, or banning car usage entirely. The claimed benefits of such interventions are several: reductions in pollution (Chiquetto, 1997), reductions in traffic accidents (Kaygisiz et al., 2017), improved physical health through exercise (Wolf et al., 2015) and improved subjective well-being (Singleton, 2019).[1]

Arguments against road space transformation such as pedestrianization[2] often center around the concerns of local businesses (Kumar and Ross, 2006; Wooller et al., 2012). By making car-based access to businesses more difficult, pedestrianization programs can, in principal, reduce revenues and visits to businesses, a concern reported by business owners themselves (ElFouly and Ghaly, 2017). In this paper, we study the effects of a large-scale pedestrianization program, New York City's Open Streets, on business visits.

New York City's Open Streets program closes city blocks to cars. The program operates by erecting barriers to Open Street city blocks during their times of operation. Broadly, during these times, use of the roadway is restricted to pedestrians and bicyclists. The program began in May 2020. Hundreds of city blocks have been part of the program at some point since its inception.[3]

To study the effects of the program on visits, we combine two main data sources. First, we collect web-scraped information on the city blocks that were part of the program and their dates of participation. Second, we measure business visits using anonymized cellphone geodata provided

[1] See Soni and Soni (2016) for a review.

[2] Pedestrianization refers to the conversion of an area, such as a road, into a pedestrian-only zone by restricting the access of vehicles. In practice, bicycles and other "human-powered" vehicles are often allowed in the pedestrianized zone, as is the case with the program studied here.

[3] Using our scraped program data, we find 718 street segments (as defined by the US Census Bureau) have been part of the program as of June 2021. These segments are often equivalent to the colloquial understanding of a city block, but are sometimes smaller.

by SafeGraph. SafeGraph aggregates raw cellphone ping coordinates into visits over time at the business level, and provides the street address of each business, allowing us to match businesses to the program participation of the city block on which they are located.

Our estimation strategy employs a matched difference-in-difference design in which sites for counterfactual candidates for open streets[4] are selected using convolutional neural networks (CNNs), based on an approach proposed by Pollmann (2020). This method allows us to handle the inherent high complexity of spatial features that may impact the choice of locations for the open streets. To obtain a pool of candidate counterfactual locations, we trained a classification algorithm similar to computer vision algorithms to predict sites where the opening of an open street is most likely based on the spatial configuration of multiple covariates, including the relative locations of businesses and population demographics. After obtaining the candidates for the Open Streets program, we use propensity score matching to match each business that belongs to an actual open street with a corresponding counterfactual business on one of the candidate streets. Compared to the popular approach of defining the treatment and control groups to be in the corresponding inner and outer rings around the location of the intervention, this method relies on control businesses located at the same distance from the counterfactual candidate sites as the treatment group. This yields an estimator that is formally valid under the quasi-experimental variation in treatment location in cases where the geographic extent of the effects is not known (Pollmann, 2020). Consequently, we are also able to estimate spatial treatment spillover effects by selecting counterfactual businesses at various distance ranges from the open street candidates.

With our matched sample in hand, we analyze the effect of the program using a stacked-in-event-time dynamic difference-in-differences design. We find significant positive effects of the program overall, with more pronounced effects for particular sub-groups of Open Streets and business types. Using the whole matched sample, we find that an opening of an open street results in a 9% higher number of visits to businesses located on that street. This result stands in contrast to the typical concern of significantly reduced visits due to restricted access. To examine the

[4] Although the official name of the program is "Open Streets," for the sake of brevity, we will call a street segment that was designated for the program an "open street."

heterogeneity in the effects, we next examine Open Streets in neighborhoods with a low share of workplace population, and find significantly positive effects on visits: our estimated effect rises to 12% for businesses in census block groups with less 75% of workplace population. We also consider effects by industry and find that it's mainly restaurants, bars, and amusement & recreation establishments that benefit from the Open Streets program. Dynamic analysis suggests the gains from an Open Streets designation accumulate over time, generally increasing slightly after the first month and remaining stable in the next four months. Additionally, we provide evidence to suggest that an Open Streets designation has positive effects on visits to restaurants and bars on neighboring streets, with significant spillover effects for establishments located between 40 to 80 meters from an open street, but not for the ones further away. The effects for all businesses in census block groups with a low share of the workplace population show a similar spatial pattern. Counter-intuitively, the spillover effects in the full sample are insignificant, except for a 7.5% increase in visits for businesses located between 160 and 200 meters away from an open street, with point estimates increasing with distance.

We contribute primarily to a literature on the effects of pedestrianization. Studies in this literature typically perform case studies of a particular pedestrianization episode, using qualitative methods or small-scale surveys (see e.g., Amistad, 2010; Chaudhuri and Zieff, 2015; Uzunoğlu and Uzunoğlu, 2020; Wolf et al., 2015). This literature has faced two main challenges in estimating the causal effects of pedestrianization. First, typical episodes of pedestrianization are small-scale, usually only affecting one or a handful of streets. This "small sample" problem makes the econometric identification of causal effects challenging. Second, to measure visits, the literature has relied on surveys. The cost of obtaining data for a wide "control group" often restricts the analysis to a before-after comparison of the pedestrianized area. Our work makes progress on both fronts. First, the New York City Open Streets program has pedestrianized hundreds of city blocks, allowing us to econometrically examine the effects of the program and examine heterogeneity in its effects across different types of streets. Second, newly available cellphone geodata allows us to systematically measure visits without relying on costly surveys, meaning we can build a broad comparison group for analysis. In complementary work, Yoshimura et al. (2022) study changes in pedestrian area and business revenues in Spain by analyzing land-use changes across fourteen Spanish cities. Compared to their related work, we study a particular pedestrianization program, rather than land-use changes generally, and assign businesses to treatment only if their street segment is pedestrianized, rather than if it is in the vicinity of an area in which land-use change has occurred. This sharp definition of treatment allows us to directly measure the effects of changing allowed uses of road space, rather than the bundle of changes associated with land-use change, which may include, for example, park construction.

Methodologically, we follow a set of recent papers employing difference-in-differences designs combined with matching-based control methods (see e.g., Deibler (2021) and Tsai (2018) for difference-in-differences combined with propensity score matching, and e.g., Wiltshire (2021) for synthetic control). Recent work has emphasized the potential pitfalls of traditional difference-in-differences designs in the presence of variation in treatment timing. We show that a stacked-in-event time matched difference-in-differences design such as that undertaken here has a simple interpretation: the estimated treatment effect is a weighted average of cohort-specific treatment effects (which compare units treated in a particular period to those never treated), where the weights are the share of treated units in each cohort. We therefore avoid the potential "negative weighting" issues emphasized in the recent literature on traditional difference-in-differences designs.

The paper is structured as followed. In Section 2 we describe the Open Streets program in more detail. In Section 3 we outline the data we utilize, in Section 4 we set out our empirical design, and in Section 5 we present our results for the direct effect of the Open Streets program, while in Section 6 we present our results on the spillover effects of the Open Streets program. Section 7 concludes.

2 Background

In May of 2020, as a direct reaction to the COVID-19 pandemic, City of New York implemented the Open Streets program. At that time, the program was seen as a way to encourage the use of public spaces while ensuring proper social distancing (Wamsley, n.d.). The program closed designated city blocks to vehicles, and was designed to transform streets into "public space open to all".[5] It operates by erecting barriers in the roadway at entrances to Open Streets areas which block large vehicles and typically display signs that state that only bicycles are allowed to pass through. The Open Street program is operated by the New York City Department of Transport. Both the designation of Open Street locations and their operation are determined with input from community-based organizations and private business groups, typically Business Improvement Districts. Throughout the program there have been several different "types" of Open Streets. Typical variations include the extent of vehicular restrictions (no vehicles permitted whatsoever vs local access and parking only) and times of operation (varying from 24h/day on all days of the week to a few hours on a single day of the week). At various points in time several specialized subprograms have existed, including "Open Streets Restaurants", which allowed restaurants to place tables on the roadway in an Open Street and "Open Streets Play", which emphasized programming for children. Our main results consider all Open Street types, but we also present our results omitting Open Streets that constituted bike lane creation. Information about the location and operating months of Open Streets are given in Section 3.

3 Data

3.1 Area and time frame of study

Throughout, we consider businesses and Open Streets in Manhattan and Brooklyn only, omitting the three remaining boroughs of New York City (Queens, the Bronx, and Staten Island). In-

[5] https://www1.nyc.gov/html/dot/html/pedestrians/openstreets.shtml

formal surveys suggest enforcement of Open Streets was particularly low in the boroughs we omit, complicating analysis if they were to be included.[6] We collect information on business visits from January 2019-June 2021 and information on Open Streets from the beginning of the program in May 2020 through June 2021.

3.2 Tiger Line road lines

We use Census 2020 Tiger Line road lines to assign businesses in the SafeGraph data to street segments, and associate those street segments with their Open Streets status. The Tiger Line road line data gives an identifier for each stretch or road similar to a city block ("segment", the stretch of a street between two cross streets) and records the name of the street, the smallest and largest address numbers on the segment, and the postal code(s) associated with the segment.

3.3 SafeGraph Visits data

We measure business visits using anonymized cell phone geodata provided by SafeGraph. Safegraph is a commercial provider of anonymized location data, obtained from a variety of mobile apps, and collected with consent from more than 40 million mobile devices in the United States.[7] This data has become popular recently in papers in economics (see e.g., Goldfarb and Tucker, 2020; Goolsbee and Syverson, 2021; Sedov, 2022). Safegraph aggregates raw cellphone ping coordinates into visits-over-time information at the "point of interest" level – retail locations and other entities, such as churches. We utilize SafeGraph's "Patterns" data at the monthly level. This provides us with a panel of number of unique visitors to each location in each month. We use data from January 2019 to June 2021. The data contain both street address information and lat-long information. We assign points of interest to Tiger Line street segments using ArcGIS geocoding. This procedure exploits the fact that we observe the street address of each point of interest in the SafeGraph data and the smallest and largest address numbers (as well as the street name) of each

[6] https://www.transalt.org/open-streets-forever-nyc
[7] https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset

street segment in the Tiger Line data. As well as number of unique visitors, SafeGraph also records median dwell time (median number of minutes devices spend at a location) and number of visitors split by operating system (iOS and Android).

An example map of points of interest in the SafeGraph data is given in Figure 1. This shows all points of interest in the data for a region of Manhattan and Brooklyn. For each point of interest, SafeGraph records the six-digit NAICS code associated with the location. In Figure 2 we plot the locations of all bars and restaurants in the data (locations with three-digit NAICS code 722 "Food Services and Drinking Places"). Summary stats for our sample are given in Table 1.

3.4 Open Streets

We collect information on Open Streets in operation from the official program page hosted on `nyc.gov`.[8] This page lists each Open Street in operation, with information on (i) the street the Open Street operates on (ii) the cross-streets the Open Street stretches between and (iii) information about the type of Open Street, including the days/times of operation. We accessed archived versions of the Open Streets program page hosted on Internet Archive's Wayback Machine. An example screen capture is shown in Figure 3. We accessed 152 historic snapshots of this site between May 2nd, 2020 and June 4th, 2021 and scraped the HTML contents using the "`read_html`" function included in Python's *pandas* software library. With the scraped data, for each unique geography of Open Street (street name, beginning and ending cross-street name combinations) in Brooklyn and Manhattan, we manually record which Tiger Line street segments are part of the Open Street. Using the multiple snapshots of the site, we build a monthly panel of Tiger Line street segments, recording whether an Open Street was in operation during each month in each Tiger Line street segment.

We map the street segments that were ever part of an Open Street in Figure 4. While Open Streets can be found in large swathes of Manhattan and Brooklyn, the street segments are disproportionately in the central areas of the city – the southern parts of Manhattan and north-western

[8] https://www1.nyc.gov/html/dot/html/pedestrians/openstreets.shtml

parts of Brooklyn. Open streets vary in length, ranging from one city block to several contiguous city blocks.

The Open Streets program began in May 2020. Different Open Streets began at different times. In Figure 7 we plot the number of street segments that first had an Open Street in each month. Approximately half of the street segments that belonged to an Open Street at some time by June 2021 entered the program in May 2020, with most others beginning in the later summer months of 2020. We plot the street segments that had an open street that began in May 2020 in Figure 5, and those that first had an open street that began after May 2020 in Figure 6. In Figure 8 we plot the number of street segments that were part of an operating Open Street segment in each month. Measured by number of street segments covered, the program increases in scale between May and October 2020, reduces in scale between October and 2020 and January 2021, then remains relatively constant thereafter.

Throughout, we define a business as treated if it is on a street segment that is part of Open Streets – that is, the street segment the business is on is pedestrianized. Implicitly, businesses *close* to an Open Street but not on one are defined as untreated. By specifying treatment in this way we may miss positive spillovers from Open Streets participation, and also measure (as is standard in difference-in-differences approaches) an overall effect that may be comprised of both reallocation of visits towards Open Streets and an aggregate increase in visits overall.

3.5 Longitudinal Employer-Household Dynamics (LEHD)

In order to study the heterogenous effects of the Open Streets program on areas of the city with a higher share of residential (or nighttime) population, and areas of the city with a higher share of workplace (or daytime) population, we will use the LEHD Longitudinal Origin-Destination Employment Statistics (LODES), produced by the Center for Economic Studies of the US Census Bureau. The LODES dataset provides yearly information at the census block level of the number of residents commuting *to* work to each census block (which we will call workplace population) and the number of residents commuting *from* each census block, which we will call residential population. We will aggregate this information at the census block group level (which is in between the census block and the census tract in size) and we will take the LODES data for 2019 only.

The LODES data allows us to construct a measure of the share of daytime population in a given location, relative to the total population (daytime and nighttime) that spend time in that location. Specifically, we define the workplace population share for a given census block group (CBG) as the population commuting to that CBG divided by the sum of the population commuting to the CBG and the population commuting from that CBG:

$$\text{workplace share} = \frac{\text{workplace population}}{\text{workplace population + residential population}}.$$

4 Empirical design

4.1 Defining treatment

While Open Streets status is not an absorbing state (segments with Open Street status may not have Open Street status in a later month), we restrict the analysis such that we consider only the first treated spell for each business. For each business on an Open Streets street segment, we first determine the month in which it was first treated. We define the sample of treated units as those (i) observed with positive visits for four months before and four months after the month in which they are first treated (ii) treated for at least four continuous months after first month treated. For a treated unit, the pool of potential control units that can be matched are those observed with positive visits for four months before and four months after the month in which the treated unit is first treated. We define a treated cohort as units that were first treated in the same calendar month.

4.2 Selecting counterfactual open streets using CNNs

In an ideal scenario, identifying the causal effects of converting a street into an open street on nearby businesses could be done by running the following experiment: first, selecting a pool of candidate streets, and then randomly dividing this pool into streets that are converted and those that are not, thereby defining the treatment and control locations. Then, assuming that candidate streets are located sufficiently far from each other, one could estimate the causal effects of opening an open street by comparing the changes in foot traffic among businesses in the treatment group with the corresponding changes in foot traffic among businesses in the control group.

However, in reality, open streets are not randomly assigned and hence, causal inference should rely on selecting a set of counterfactual locations, for which it could be argued that an open street could have been opened there just as likely as in the set of actual open street locations. To identify a set of such counterfactual locations we follow the recent literature on spatial causal inference (Pollmann, 2020) that relies on convolutional neural networks (CNNs), borrowing from image classification algorithms. In this approach, the data about open street locations and the relevant covariates, such as locations and types of businesses and neighborhood demographics, is represented on a discretized map that divides the space into grid cells. We set the width of each grid cell to be 0.025 miles (40.2 meters), and Table 2 summarizes the key variables used as inputs to train the CNN model. The idea is to use complex spatial information embedded in a map, including the relative position of open streets with respect to other covariates, to pick the most likely counterfactual locations, where open streets could have opened.

In the first step of the classification, the CNN considers broader 10x10 grids of cells and has to distinguish between two classes of grids. Intuitively, we want to train the classification to identify locations where there should be an open street but it is "missing". Hence, the first class is the set of grids where at least one of the 10x10 cells is treated, but it is intentionally removed from the data to appear as if it is not treated. The correct classification of such grids should be in the category where the real treatment location is missing. The second class combines the real treated grids without any data removed and the set of untreated grids, where no cells contain an open street. Both should be classified into grids that are not missing any real treatment locations. In the second step of the classification, if a grid is classified as having a missing treatment location, the network has to predict which of the 10 x 10 cells contains the missing treatment.

The prediction step of the convolutional neural network gives us an assessment of how similar each grid cell and its neighborhood are to the cells that contain an open street and their surrounding areas. In the next step, we use these predictions to obtain a smaller pool of candidate counterfactual locations for open streets with each location corresponding to the centroid of the grid cell used in the training step. This is done by taking all real treated locations and matching each one of them with the most similar untreated location based on the activation scores and conditional probabilities predicted by the CNN. This gives us a pool of candidate untreated locations that appear most likely to belong to an open street. Figure 9 shows the map of real treated locations and the matched pool of candidate locations. Then, we consider all street segments that contain at least one candidate location and are located at least 400 meters away from any of the real treatment locations and denote them as our candidates for open streets.

Finally, for our estimation of direct effects on businesses located on an open street, we define the businesses in the counterfactual treatment group using propensity score matching to match each business that belongs to an actual open street with a corresponding business on the candidate streets. To identify spatial spillover effects from the Open Streets program, we further define several additional treatment and counterfactual groups at different distance bands ranging from 40 to 200 meters away from the corresponding treated and counterfactual open streets. Panel a in Figure 10 displays an example of an open street segment with treated businesses located in the 40-80m range, and Panel B demonstrates how we pick the corresponding counterfactual businesses in the same distance range around a candidate street that we use for estimating spillover effects.

4.3 Matching treated businesses to candidates

Although the selection of counterfactual streets implies that we are comparing street segments that are similar in many characteristics, this selection process does not do a perfect job at balancing all observable characteristics between streets that are part of the Open Streets program and counterfactual streets. In particular, street segments that became Open Streets differ in some important ways from counterfactual street segments. As shown in Table 3, businesses on Open Streets segments are, for example, more likely to be in Manhattan, have a higher share of restaurants and bars on their segment in 2019, and a higher share of establishments in their census block group in 2019. These differences are important to the extent to which trends in outcomes may be determined by them. In Figure 11 we compare the dynamics of average visits between Manhattan and Brooklyn. While average visits was approximately equal in the months before the COVID-19 pandemic, visits dropped by significantly more in Manhattan, which contains the Central Business District and a large share of overall employment in New York City. Similarly, In Figure 12 we show the dynamics of average visits by two popular three-digit NAICS codes over our sample period: those referring to restaurants and bars (code 722, "Food Services and Drinking Places") and those referring to grocery and convenience stores (code 445, "Food and Beverage Stores"). While restaurants and bars enjoyed significantly higher average visits in the months before the pandemic, they suffered a sharper decline and recovery in the early months of the pandemic. Given the disproportionate specialization of Open Streets blocks in these categories, trends in future outcomes during the program may reflect these differing characteristics rather than a causal effect of the policy. In order to address these concerns, we match each treated business to a business on a counterfactual street segment.

We perform matching using a combination of exact matching and propensity score matching (Rosenbaum and Rubin, 1983). We match businesses on Open Streets street segments to businesses on the candidate street segments picked by the CNN using exact matching on borough and industry classification (6-digit NAICS classification), and inexact (through the propensity score) on a set of covariates.

For each treated cohort, we perform matching in two steps. We first run a probit regression (estimated separately for each treated cohort) that relates treatment status of the business to the number of businesses on the street segment (averaged over the months in 2019), the share of businesses on the street segment that are restaurants and bars (averaged over the months in 2019), the distance of the business to the CBD (measured as kilometers to the Empire State Building), the square of this distance, the workplace share (as defined in subsection 3.5), and the median dwell time for businesses in that street segment, averaged across 2019. We use the log odds of the predicted values from this probit regression to assign each treated and candidate unit a propensity score. Next, for each treated unit we match it to at most one control unit by searching for candidate units that share the same borough and six-digit industry classification, and choosing the control unit with the closest linearized propensity score. We perform this matching with replacement, meaning one control unit can be matched to more than one treated unit. Due to the fact we perform matching exactly within certain characteristics and the fact we apply a propensity score caliper (which requires a treated unit and matched control to have an estimated linearized propensity score within 0.1 standard deviations), not all treated units may find a successful match.

4.4 Covariate balance

We first note that, by construction, the matched treated units and matched controls are exactly balanced by borough and six-digit NAICS code combination. We compare matched treated units to their matched controls in Table 4. Compared to the full sample of treated and non-treated units in the counterfactual street segments (Table 3), matched treated units and controls are significantly more balanced across both characteristics used in the matching and those not used. Beginning with variables used in the matching, while in the full sample treated units had more than double the number of businesses on their street segment than non-treated units (17.22 vs 7.44, a difference of 9.78), in the matched sample this difference is reduced five-fold to 1.89.[9] In terms of segment-level share of restaurants and bars, in the full sample treated units had a 22 percentage point higher share, in the matched sample, the difference is completely eliminated. In terms of distance to the city center (CBD), in the full sample treated units were 6.35km closer on average, in the matched sample they are 0.16km further. The median dwell time for people visiting each establishment is still higher for the matched control establishments than for the matched treated sample (on average), but the difference is reduced by more than fifty per cent when compared to the full sample of businesses in counterfactual street segments (from a 48-minute difference to a 20-minute difference). Lastly, while in the full sample treated units were 48 percentage points more likely

[9]However, the difference is still statistically significant.

to be in Brooklyn, in the matched sample this difference is zero, as units are exactly matched by borough. We show a kernel density of the distribution of estimated propensity scores for treated units, matched controls, and all establishments in the counterfactual street segments in Figure 13. Although the distribution of propensity scores do not align perfectly between treated and control units, compared to the full sample of establishments in the counterfactual segments, the distribution of propensity scores for control units is much closer and shares a common support with the distribution of propensity scores for the treated units.

The identifying assumption for our matched difference-in-differences design requires treated and non-treated units would have followed parallel trends in the absence of intervention. The main objective of matching is to obtain a control sample that follows similar trends in visits to the treated units in the pre-treatment periods. The matching procedure does not use any information about the outcome variable to perform the matching, and so whether the matching used is successful in obtaining parallel pre-trends is an empirical question. In Figure 14, we plot visits over time to treated and non-treated units in the months before the program. In panel (a), we compare treated units to all non-treated units in the SafeGraph data. While in 2019, treated units enjoy higher average visits, this situation reverses in March and April of 2020 (the first months of the COVID-19 pandemic). These differential trends would complicate any subsequent difference-in-difference analysis. In panel (b), we compare all treated units to all units in counterfactual segments. We can see from this graph that, without the business level matching, there are still differential trends for the first months of 2020. In panel (c), we compare the same dynamics for matched treated units and their matched controls. In this case, the ordering of average visits remains the same throughout the entire period, with the magnitude of the gap remaining approximately constant. These broadly parallel dynamics in the months before the program aid the legitimacy of the identifying assumption; we will test for non-parallel pre-trends formally in the event study analysis described in subsection 5.3.

4.5 Estimating equations

We estimate the effects of Open Streets on business visits using a stacked-in-event-time difference-in-differences design. The design allows for comparison in outcomes for the treated group and matched control group before and after intervention, and combines all treated cohorts and their matched controls into a single specification. In particular, we estimate the following regression,

$$y_{it} = \alpha_i + \sum_{\substack{k=-4 \\ k\neq -1}}^{5} \gamma_k \mathbb{1}\{t - E_i = k\} + \sum_{\substack{k=-4 \\ k\neq -1}}^{5} \gamma_k^{Treated} \mathbb{1}\{t - E_i = k\} \times Treated_i + \epsilon_{it} \tag{1}$$

where y_{it} denotes unit i's outcome in month t, α_i is a unit-specific fixed effect, E_i is the month in which unit i is first treated if i is a treated unit and the month in which unit i's matched treated unit is treated if i is a control, $Treated_i$ is an indicator for if unit i is treated, and ϵ_{it} is the error term. The coefficients of interest are $\gamma_k^{Treated}$. We sometimes report a simplified version of equation 1, which collapses the treatment interaction to admit a single estimated treatment effect

$$y_{it} = \alpha_i + \sum_{\substack{k=-4 \\ k\neq -1}}^{5} \beta_k \mathbb{1}\{t - E_i = k\} + \beta^{Treated} \mathbb{1}\{t - E_i \geq 0\} \times Treated_i + \varepsilon_{it} \tag{2}$$

where $\beta^{Treated}$ is the coefficient of interest. Throughout, we cluster standard errors at the level at which we performed the exact match – five digit zip codes by six-digit NAICS codes.[10]

Several recent works have highlighted flaws in traditional two-way fixed-effect estimators when both (i) treatment timing is staggered (ii) treatment effects change over time since treatment (see e.g., Goodman-Bacon, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Borusyak et al., 2021). In such specifications, biases can arise due to the fact such a specification implicitly uses earlier-treated groups as controls after beginning treatment. If the treatment effect is not constant over time, the interpretation of estimates is thus complicated.

[10]The only exception will be for estimations on a subsample of establishments in one 3-digit NAICS subsector. In these cases, clustering at the 6-digit NAICS code would lead to too few clustering groups, which can lead to biased standard errors and over-rejection (Ibragimov and Müller, 2016).

Here, (i) analysis is cast in event time (ii) each treated unit is observed for the same number of months before and after treatment, and (iii) each treated unit is matched to a single control unit observed over the same time periods. This results in the estimates $\gamma_k^{Treated}$ having a clear interpretation: $\gamma_k^{Treated}$ is the weighted average of estimates from cohort-specific event study estimates (which compare treated units all treated at the same time to never treated units), where the weights are the number of treated units in each treated cohort. This formalizes the intuition in recent work using matching-based control strategies cast in event time that such studies avoid negative weighting issues in traditional two-way fixed effect designs (Deibler, 2021; Wiltshire, 2021).

5 Results: Direct Effects of Pedestrianization

We start by estimating equation 2 for the full matched sample and using the log of the monthly number of visitors to each establishment as the outcome variable. Column 1 in Table 5 reports the resulting coefficient $\beta^{Treated}$ for this difference-in-differences estimation[11]. For this sample of businesses, the average treatment effect of becoming part of an Open Streets segment is a relative increase in the number of visitors of 9%. This results is significant at a 1% level.

5.1 Effects by Workplace Population Share

During the first months of the Open Streets program, there was a substantial decrease in economic activity in the entire city. However, this decrease was particularly pronounced in areas of the city that had a high share of daytime population, but a low share of nighttime (or residential) population. Figure 15 plots the workplace population share in 2019 for each census block group (defined in subsection 3.5) and the yearly change from 2019 to 2020 of the total number of visits to each CBG (according to the SafeGraph "patterns" data). We can see from this figure that the locations with the highest share of workplace population share saw the largest declines in the number of visits to establishments. This is consistent with the large increase in remote work and the

[11]Because we discard treated units for which we cannot find a nearest-neighbor match that fits our matching criteria, the resulting effect should be interpreted as the average treatment effect in the remaining matched sample (ATM).

presence of lock-downs for part of 2020 in the city.

Miyauchi et al. (2021) find that consumption-related trips are more likely to originate from home, they are frequently part of a trip chain (which might originate from work or from home) and they became more local (at least in the Tokyo area) during the COVID-19 pandemic. Given these results, one can imagine that creating pedestrian areas in locations with a very high workplace population share might have had a lower impact on the number of visitors that visit these establishments (since there will likely be less consumers willing to travel to these locations in the first place). In order to test this hypothesis, we divide our sample into businesses in CBGs with a workplace population share lower than 75% and businesses in CBGs with a workplace population share higher than 75%. Columns 2 and 3 of Table 5 show the results of estimating equation 2 on these two samples. We see that, indeed, becoming part of the Open Streets program leads to a 12% relative increase in the monthly number of visitors for businesses in CBGs with lower workplace share population, but only a 6% increase (which isn't statistically significant at a 10% level) increase in areas with a high workplace share.

5.2 Effects by Industry Subsector

We next consider the heterogeneity in the effects of this pedestrianization policy by the type of industry of the establishments in our data. In particular, Figure 16 plots the results of estimating $\beta^{Treated}$ from equation 2 for the subsample of businesses in each of the five most popular NAICS 3-digit industry subsectors in our data. We can see that the effects of the Open Streets program are particularly large and significant at a 5% level for the Food Services and Drinking Places establishments and for the Amusement, Gambling, and Recreation Industries. Table 6 provides the numerical values of these estimations for these two industries. These positive and significant coefficients are consistent with pedestrian zones allowing for businesses to cater to consumers on the sidewalks and on the pedestrianized streets. This might have had a particularly important effect during the pandemic years, due to the social distancing requirements.

5.3 Event Study Dynamics

Our baseline event study is based on estimating equation 1 with the log of the monthly number of visitors as the outcome of interest for the whole sample of matched establishments. The results of this difference-in-differences estimation are represented in Figure 17, which plots the estimated coefficients $\gamma_k^{Treated}$ and their 95% confidence intervals. Each pre-treatment month has a coefficient statistically indistinguishable from zero, while coefficients associated with one, two and three months after treatment are statistically significant at a 95% level, with point estimates that are reasonably stable at around 10%. There do not seem to be any clear pre-trends for the months before treatment, and the effects seem to be reasonably persistent (at least after the first month of treatment).

We next consider the subset of establishments that are in census block groups with a workplace share lower than 75%. We show the event study using this sample in Figure 18. Each pre-treatment month has a coefficient statistically indistinguishable from zero, but coefficients for the months of treatment after the first month are positive and significant, and all post-intervention point estimates are positive. The point estimates for all months after the first month of treatment seem to be stable at a value slightly higher than 10%.

In contrast, Figure 19 shows the results from estimating equation 1 using the sample of firms in census block groups with a workplace share higher than 75%. In this case, as was the case in Section 5.1, we do not find any statistically significant effects after treatment. Reassuringly, there do not seem to be any pre-trends either, with all coefficients for the months before treatment also being statistically indistinguishable from zero at a 90% level.

Finally, we estimate the same event study regression for the subsample of businesses that are both located in census block groups that have a lower workplace share (less than 75%) and that are from one of the industries that is arguably particularly likely to benefit from the intervention: bars and restaurants (NAICS code 722). Figure 20 shows the result of this estimation. In this case, although the coefficient associated with the fourth month before treatment is statistically significant, the two months directly preceding the baseline period have effects that are very close to zero and statistically indistinguishable from zero at a 90% level. Consistent with our results in subsections 5.2 and 5.1, all the point estimates associated with the months after treatment are positive, and the coefficients for months one to four are statistically significant at a 95% level, with a point estimate that is stable in time and around 15%.

Given our positive estimated effects of pedestrianization on the number of visitors, what explains the resistance to pedestrianization from business owners cited in the introduction? There is some evidence for misguided beliefs among business owners on the transport modes of their customers. O'Connor et al. (2012) performs a survey that compares actual transport modes to the perceived beliefs of business owners. Their results suggest owners overestimate the share of their customers who use cars and underestimate the number who walk. Another possibility is status quo bias (Fernandez and Rodrik, 1991), in which there is less uncertainty in the car customers to be lost than the pedestrian customers to be gained.

6 Results: Spillover Effects to Nearby Establishments

How does pedestrianization of a street segment affect establishments that are nearby but not directly on the pedestrianized street? We first note that, as is argued by Pollmann (2020), we cannot compare establishments on an Open Street with establishments nearby as a way to study the spillover effects, since any difference in the number of visitors between these two types of establishments could be due to a combination of spillover effects and the direct treatment effect. Instead, we will rely on our counterfactual street segments. These segments function as plausible locations for an Open Street segment, that did not receive one and that are far apart enough to reasonably assume that the businesses located on them are not directly affected by any Open Street.[12]

Using these counterfactual locations, we can define treatment and control groups based on distance bands relative to either the directly treated segments (segments that were pedestrianized) or the counterfactual segments. Figure 9 shows an example of establishments selected as either

[12]We only select counterfactual segments that are at least 400 meters from an Open Streets segment.

treated or part of the control group for a distance band of 40 to 80 meters. Once we have selected all the establishments that are within a certain distance band (e.g., more than 40 meters but less than 80 meters) from either an Open Street segment or a counterfactual segment, we proceed to match treated businesses to businesses in the control group using the same procedure described in section 4.3. We then estimate a $\beta^treated$ from equation 2 for each distance band with the log of the monthly number of visitors as the outcome. This coefficient can be interpreted as the effect of being at each distance band from an Open Street segment.[13] We repeat this estimation for four distance bands: 40 to 80 meters, 80 to 120 meters, 120 to 160 meters and 160 to 200 meters.

Figure 21 plots the results from estimating $\beta^treated$ from equation 2 for each distance band on the full sample of establishments. We do not find significant spillover effects (at a 90% level) for distances between 40 to 160 meters in this case, but we do find positive and significant spillover effects for establishments at 160 to 200 meters from a pedestrianized street. The magnitude of this effect is roughly half of the direct effect, suggesting that creating pedestrian zones might attract more foot traffic to an area, and some of that increased pedestrian traffic may end up visiting establishments nearby but outside the pedestrian zone.

In Figure 22, we select the subsample of businesses that are in census block groups with a lower workplace share (less than 75%). In this case, we observe a positive and statistically significant spillover effect for establishments between 40 and 80 meters from an Open Street segment. The point estimates for the other distance bands are also positive, but not statistically significant. If we focus on the subsample of establishments that are part of the "Food Services and Drinking Places" industry subsector (NAICS 722) and are also in census block groups with a lower workplace share, we see a very similar pattern. Only the bars and restaurants that are between 40 and 80 meters from a Open Street segment seem to have positive and statistically significant spillovers in terms of the number of visitors.

[13] As in the case of the direct effects estimated in section 5, this effect should be interpreted as the average treatment on the matched sample (ATM), since we are not able to successfully match all treated establishments to a suitable control.

7 Conclusion

We study the effects of pedestrianization on business visits using a unique policy experiment: New York City's Open Streets program. The Open Streets program closed hundreds of city blocks to cars. We combine cellphone geodata with web-scraped information on the program, allowing us to measure visits to businesses that were part of the program before and during their participation.

We find that the Open Streets program lead to a relative increase in the number of visitors to establishments on these pedestrianized streets of 9% on average. These effects were more pronounced for areas of the city with a low workplace population share (12% on average), and for bars and restaurants (10% on average), and amusement, gambling and recreation industries (19% on average). We find that these effects are persistent in the short run, but take around a month to achieve their maximum value.

When looking at spillover effects, we find evidence of positive spillover effects for establishments that are nearby, although the distance at which these effects can be detected seems to depend on the type of establishment and on how residential is the area in which it is located. When these spillover effects are statistically significant, their magnitude is roughly half of what the direct effect of being on an Open Street segment is. These findings could be explained by an increase in pedestrian traffic to the area, which then leads to a fraction of these consumers venturing out to establishments near the pedestrian zones, but not directly on them.

We see our paper as contributing to the study of urban planning decisions using novel data sources, such as cellphone geodata, that allows new ways to see how individuals "use" the city. We show that this data can inform both general trends (including, in our case, the differential effect of the COVID-19 pandemic on different parts of the city and business types) as well as the effects of particular policies. While surveys are costly to implement, novel data such as cellphone geodata allows systematic measurement of patterns, meaning the construction of broad "control groups" is feasible. Understanding the causal effects of past planning policies is essential to their future design.

We leave several important issues to future work. First, our study naturally takes place in a time period affected in fundamental ways by the COVID-19 pandemic. If the Open Streets program in New York City and similar programs in other cities continue for years to come, their effects may shift as use of the city changes. Second, while we believe measuring the effects of pedestrianization across different types of city blocks, as we do here, is an important step in determining an "optimal" pedestrianization program, we leave a full investigation of that issue to future work.

References

Amistad, F. T. (2010), 'Assessment of the pedestrianization policy in vigan city: Unesco world heritage site', *Journal of urban planning and development* **136**(1), 11–22. Cited on page 4.

Borusyak, K., Jaravel, X. and Spiess, J. (2021), 'Revisiting event study designs: Robust and efficient estimation', *arXiv preprint arXiv:2108.12419* . Cited on page 16.

Buehler, R., Pucher, J., Gerike, R. and Götschi, T. (2017), 'Reducing car dependence in the heart of europe: lessons from germany, austria, and switzerland', *Transport reviews* **37**(1), 4–28. Cited on page 2.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230. Cited on page 16.

Chaudhuri, A. and Zieff, S. G. (2015), 'Do open streets initiatives impact local businesses? the case of sunday streets in san francisco, california', *Journal of Transport & Health* **2**(4), 529–539. Cited on page 4.

Chiquetto, S. (1997), 'The environmental impacts from the implementation of a pedestrianization scheme', *Transportation Research Part D: Transport and Environment* **2**(2), 133–146. Cited on page 2.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–96. Cited on page 16.

Deibler, D. M. (2021), 'The effect of outsourcing on remaining workers, rent distribution, and inequality'. Cited on pages 5 and 17.

ElFouly, H. A. and Ghaly, A. A.-G. (2017), 'The perceived impact of pedestrianization on local businesses in al-muizz egypt: A case study', *International Journal of Development and Sustainability* **6**(7), 399–411. Cited on page 2.

Fernandez, R. and Rodrik, D. (1991), 'Resistance to reform: Status quo bias in the presence of individual-specific uncertainty', *The American economic review* pp. 1146–1155. Cited on page 20.

Goldfarb, A. and Tucker, C. (2020), Which retail outlets generate the most physical interactions?, Technical report, National Bureau of Economic Research. Cited on page 7.

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277. Cited on page 16.

Goolsbee, A. and Syverson, C. (2021), 'Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020', *Journal of public economics* **193**, 104311. Cited on page 7.

Ibragimov, R. and Müller, U. K. (2016), 'Inference with few heterogeneous clusters', *Review of Economics and Statistics* **98**(1), 83–96. Cited on page 16.

Kaygisiz, Ö., Senbil, M. and Yildiz, A. (2017), 'Influence of urban built environment on traffic accidents: The case of eskisehir (turkey)', *Case studies on transport policy* **5**(2), 306–313. Cited on page 2.

Kumar, S. and Ross, W. (2006), 'Effects of pedestrianisation on the commercial and retail areas: study in khao san road, bangkok', *Splintered urbanism* . Cited on page 2.

Miyauchi, Y., Nakajima, K. and Redding, S. J. (2021), 'The economics of spatial mobility: Theory and evidence using smartphone data', *NBER Working Paper* . Cited on page 18.

O'Connor, D. et al. (2012), 'Shopping travel behaviour in dublin city centre survey.'. Cited on page 20.

Pollmann, M. (2020), 'Causal inference for spatial treatments', *arXiv preprint arXiv:2011.00373* . Cited on pages 3, 11, and 20.

Rosenbaum, P. R. and Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55. Cited on page 13.

Sedov, D. (2022), 'Restaurant closures during the covid-19 pandemic: A descriptive analysis', *Economics Letters* p. 110380. Cited on page 7.

Singleton, P. A. (2019), 'Walking (and cycling) to well-being: Modal and other determinants of subjective well-being during the commute', *Travel behaviour and society* **16**, 249–261. Cited on page 2.

Soni, N. and Soni, N. (2016), 'Benefits of pedestrianization and warrants to pedestrianize an area', *Land use policy* **57**, 139–150. Cited on page 2.

Tsai, W.-J. (2018), 'Mandatory retirement and older worker employment decisions: Evidence from a matched difference-in-differences estimator', *Pacific Economic Review* **23**(4), 590–608. Cited on page 5.

Uzunoğlu, K. and Uzunoğlu, S. S. (2020), 'The importance of pedestrianization in cities-assessment of pedestrianized streets in nicosia walled city', *European Journal of Sustainable Development* **9**(2), 589–589. Cited on page 4.

Wamsley, L. (n.d.), 'New york city is latest to close some streets to cars, making more space for people', *NPR* .

URL: *https://www.npr.org/sections/coronavirus-live-updates/2020/05/04/850357743/* Cited on page 6.

Wiltshire, J. C. (2021), 'Walmart supercenters and monopsony power: How a large, low-wage employer impacts local labor markets', *Job market paper* . Cited on pages 5 and 17.

Wolf, S. A., Grimshaw, V. E., Sacks, R., Maguire, T., Matera, C. and Lee, K. K. (2015), 'The impact of a temporary recurrent street closure on physical activity in new york city', *Journal of urban health* **92**(2), 230–241. Cited on pages 2 and 4.

Wooller, L., Badland, H. and Schofield, G. (2012), 'Pedestrianisation: Are we reading from the same page? perspective from key stakeholders in takapuna, auckland'. Cited on page 2.

Yoshimura, Y., Kumakoshi, Y., Fan, Y., Milardo, S., Koizumi, H., Santi, P., Arias, J. M., Zheng, S. and Ratti, C. (2022), 'Street pedestrianization in urban districts: Economic impacts in spanish cities', *Cities* **120**, 103468. Cited on page 5.

8 Figures

Figure 1: Locations in the SafeGraph data: example

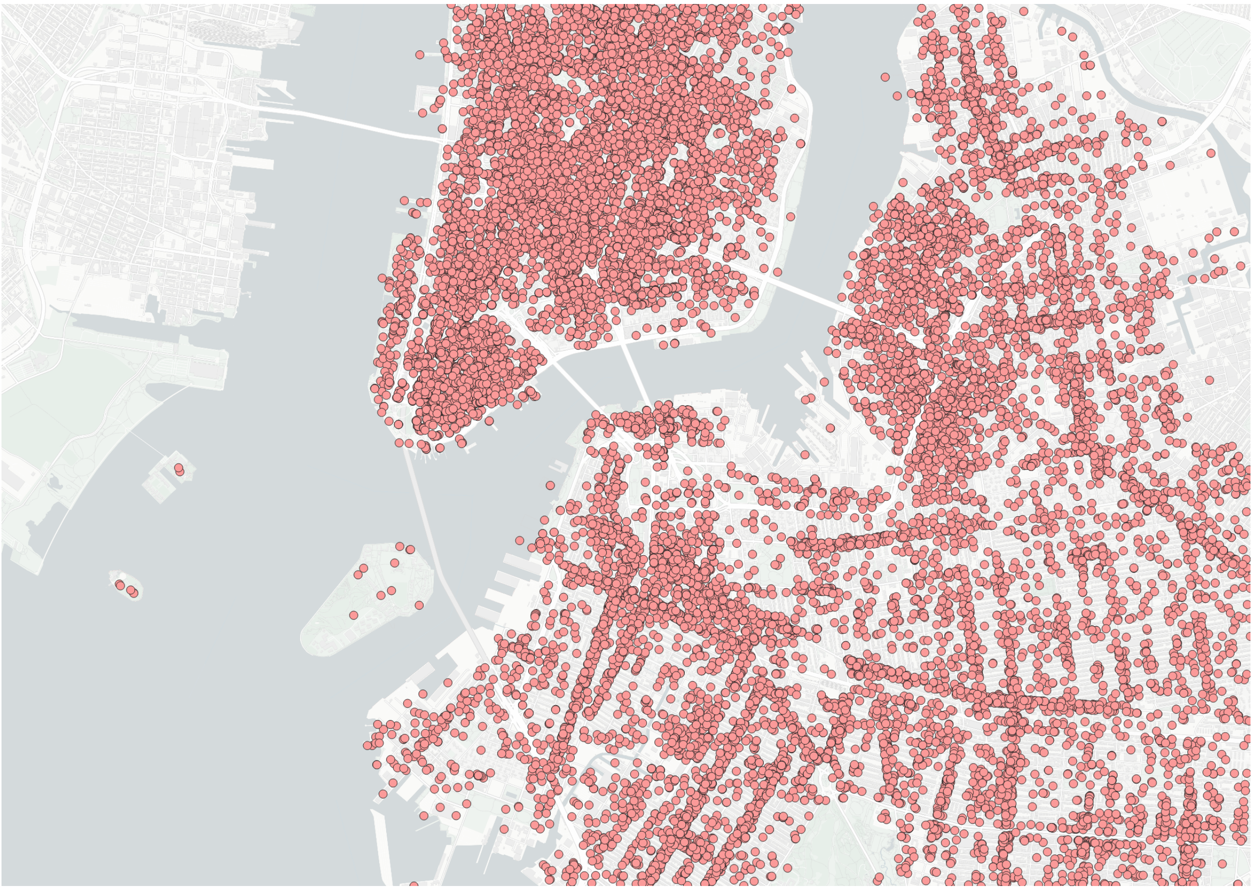


Figure: Each circle represents a location in the SafeGraph data.

Figure 2: Bars and restaurants in the SafeGraph data: example

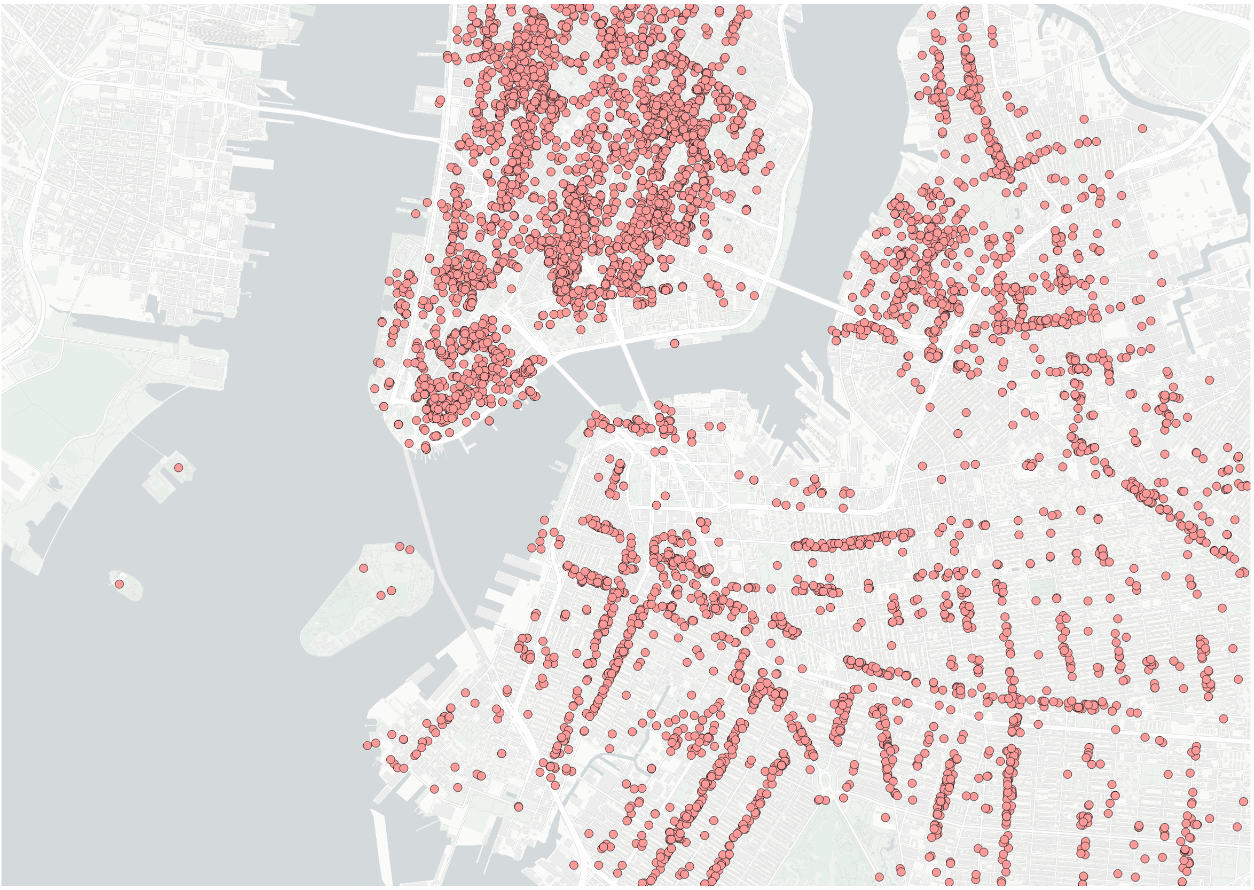
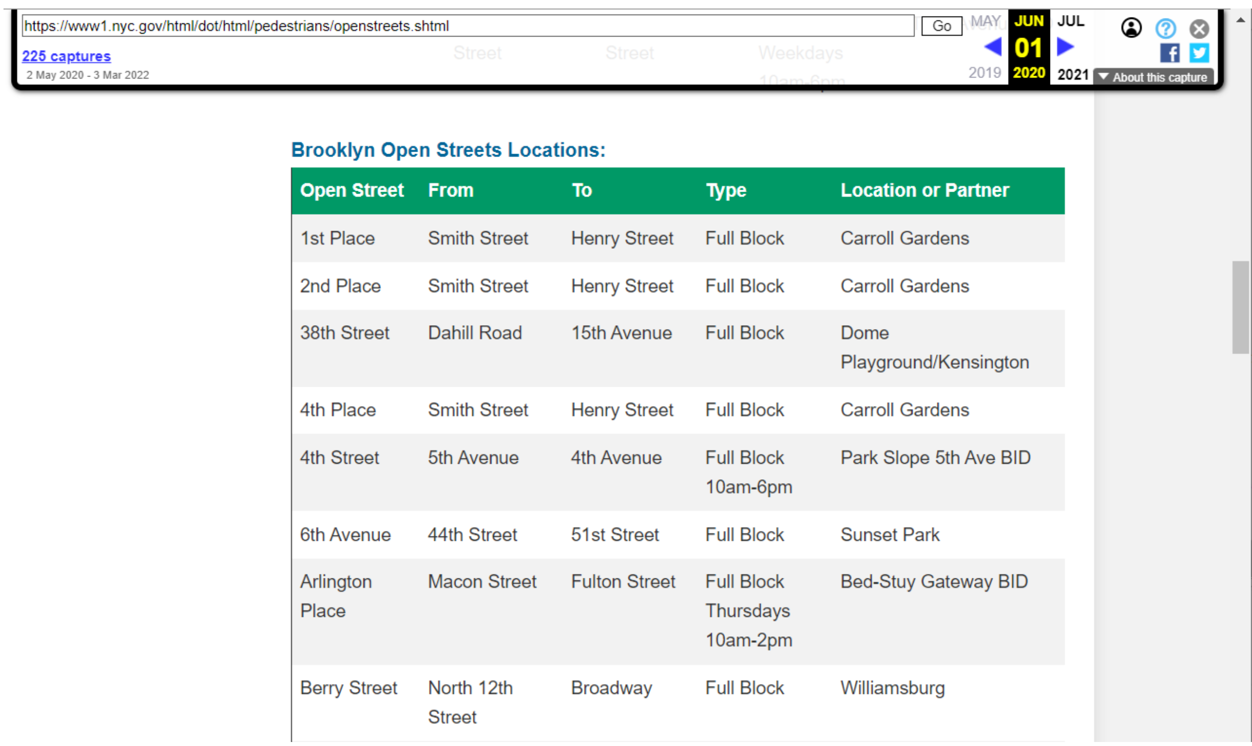


Figure: Each circle represents a location in the SafeGraph data with three-digit NAICS code 722 ("Food Services and Drinking Places").

Figure 3: Screen capture of example archived nyc.gov page



Brooklyn Open Streets Locations:

Open Street	From	To	Type	Location or Partner
1st Place	Smith Street	Henry Street	Full Block	Carroll Gardens
2nd Place	Smith Street	Henry Street	Full Block	Carroll Gardens
38th Street	Dahill Road	15th Avenue	Full Block	Dome Playground/Kensington
4th Place	Smith Street	Henry Street	Full Block	Carroll Gardens
4th Street	5th Avenue	4th Avenue	Full Block 10am-6pm	Park Slope 5th Ave BID
6th Avenue	44th Street	51st Street	Full Block	Sunset Park
Arlington Place	Macon Street	Fulton Street	Full Block Thursdays 10am-2pm	Bed-Stuy Gateway BID
Berry Street	North 12th Street	Broadway	Full Block	Williamsburg

Figure: Source: https://web.archive.org/web/20200601184936/https://www1.nyc.gov/html/dot/html/pedestrians/openstreets.shtml, accessed March 18th, 2022.

Figure 4: All street segments that were ever Open Streets in our sample

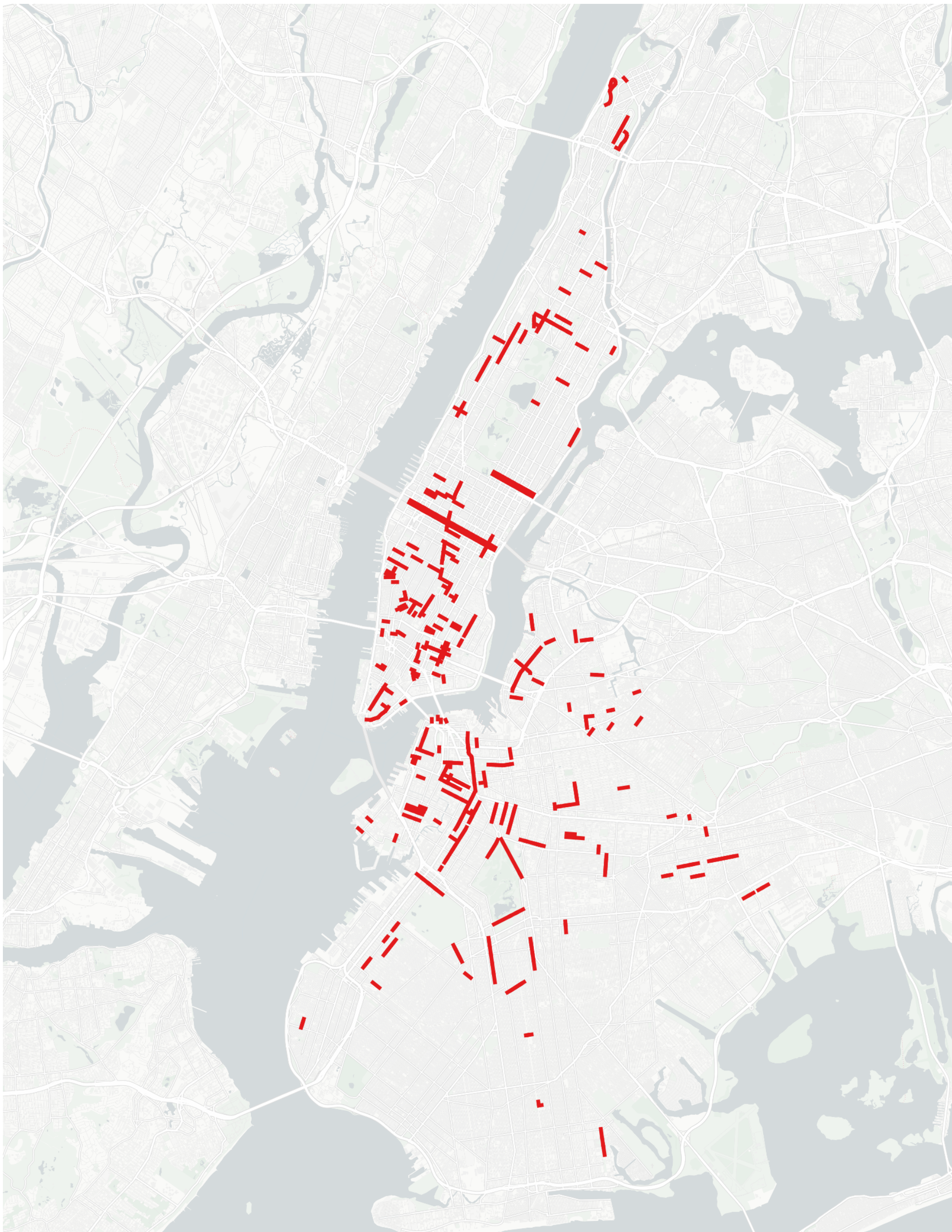


Figure: Tiger Line street segments that we record as ever having an Open Street, based on our scraped program data.

Figure 5: All street segments that first had Open Street in May 2020

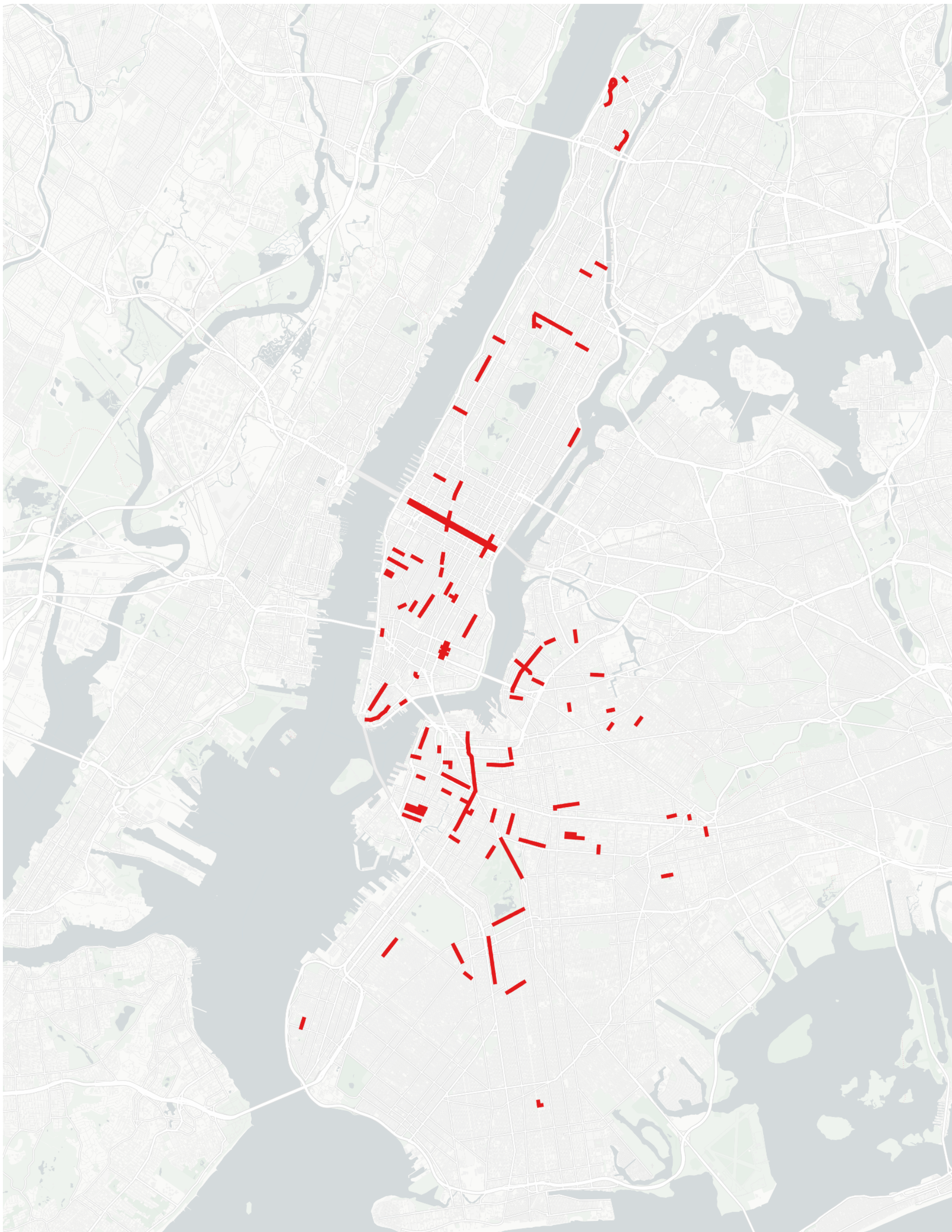


Figure: Tiger Line street segments that we record as having an Open Street that first began in May 2020, based on our scraped program data.

Figure 6: All street segments that first had Open Street after May 2020

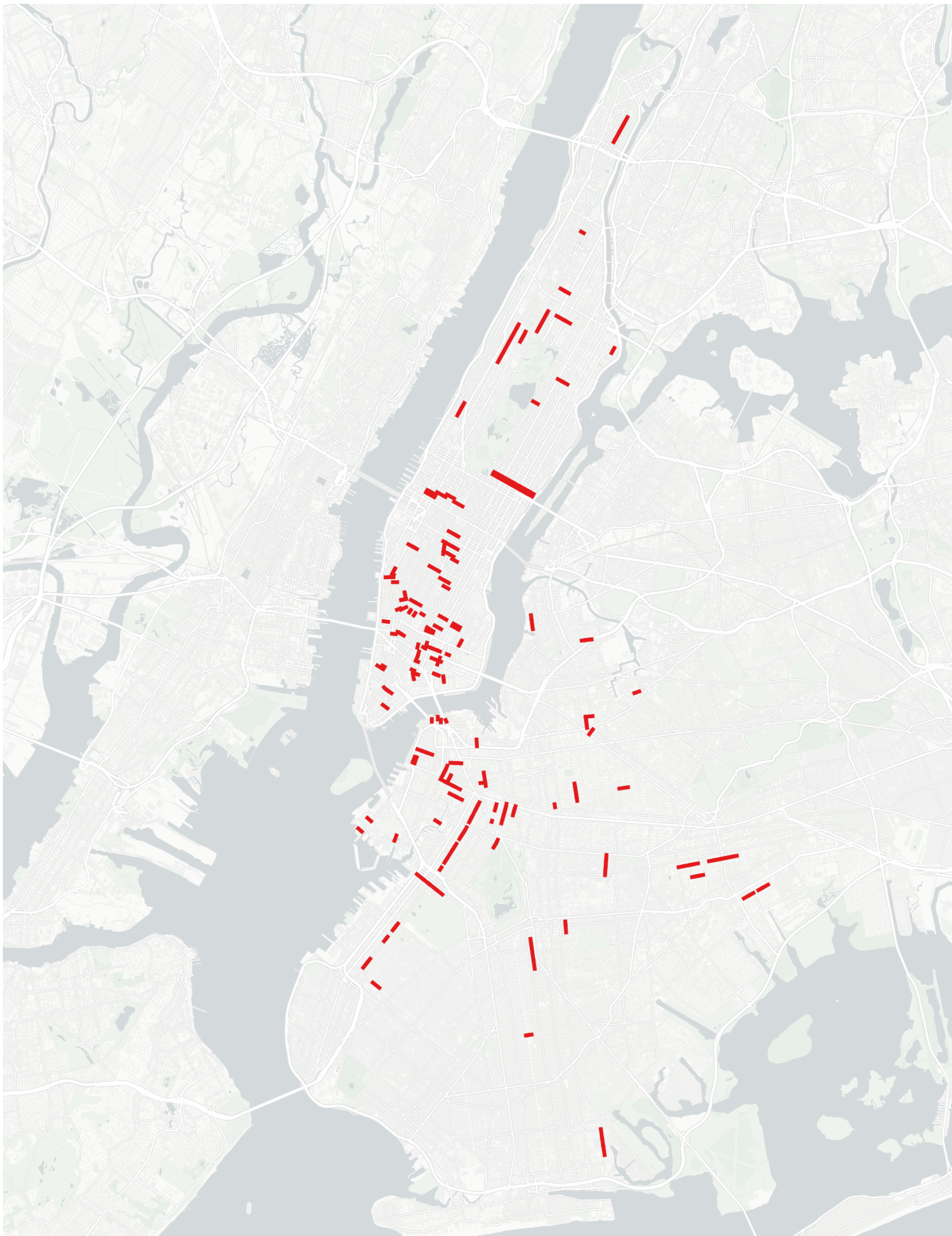


Figure: Tiger Line street segments that we record as having an Open Street that first began after May 2020, based on our scraped program data.

Figure 7: Number of street segments by first month part of an Open Street

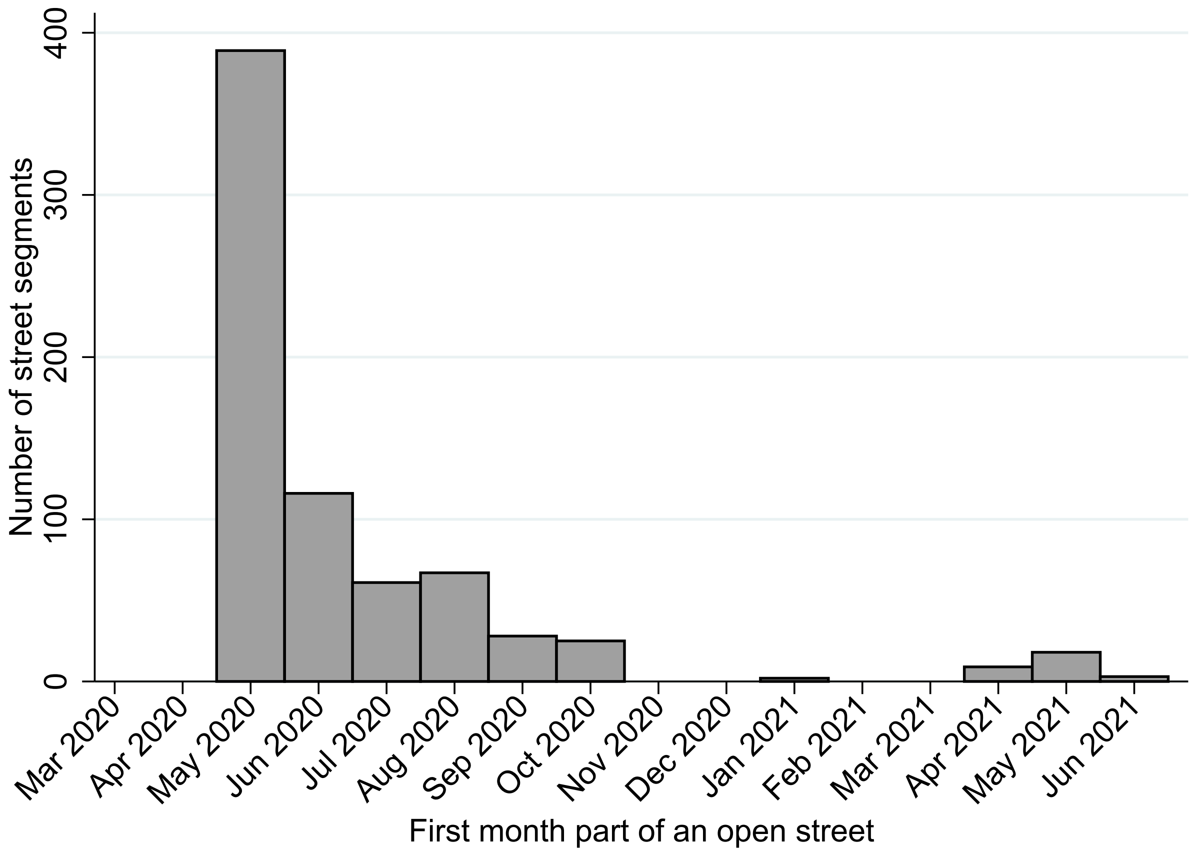


Figure: Based on scraped program data.

Figure 8: Number of street segments part of an Open Street by month

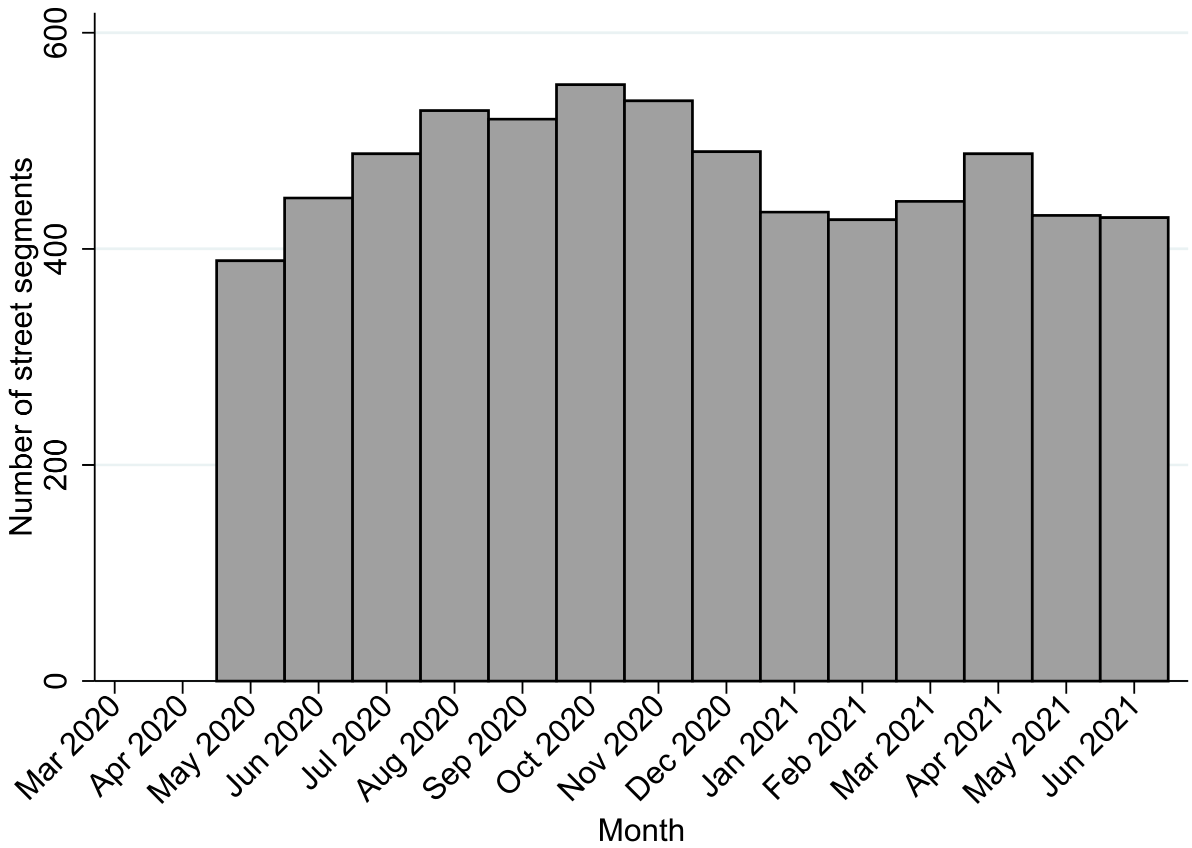


Figure: Based on scraped program data.

Figure 9: Businesses in counterfactual streets and in Open Street segments

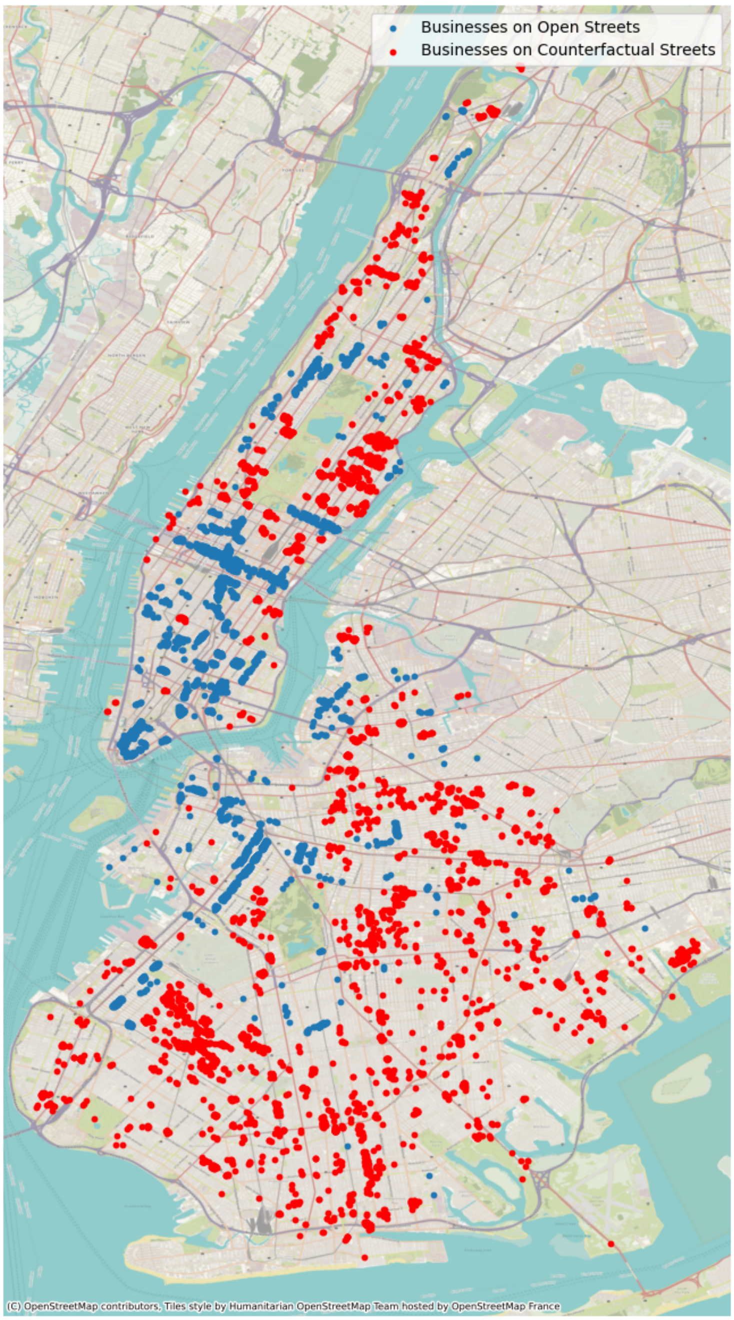
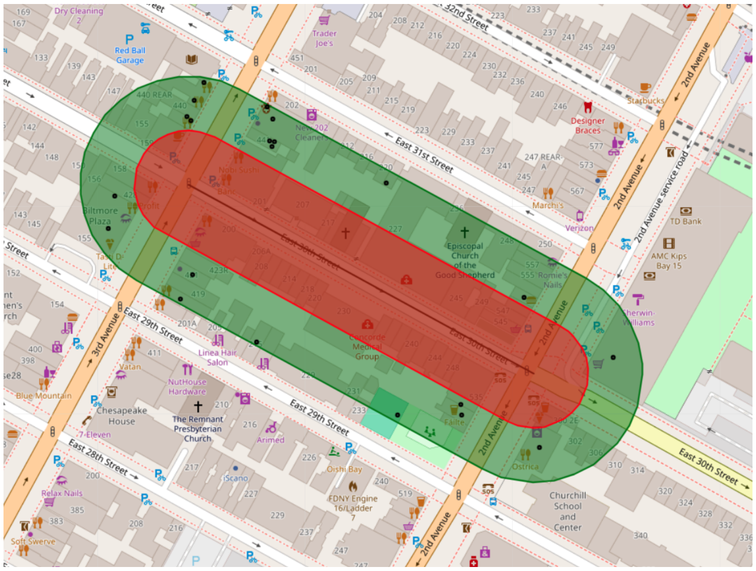


Figure: Map of Brooklyn and Manhattan with businesses on Open Streets (in blue) and businesses on counterfactual "candidate" street segments (in red), which were selected through the process explained in subsection 4.2.

Figure 10: Example of businesses selected for spillover analysis

(a) Businesses at 40-80m from an Open Street segment



(b) Businesses at 40-80m from a counterfactual “candidate” street segment

Figure: Example of businesses (represented by their centroids in black) at more than 40m and less than 80m from either an Open Street segment or a counterfactual street segment. The red buffer shows everything that is at less than 40m from the street segment, and the green buffer shows everything that is more than 40m and less than 80m from the segment.

Figure 11: Average visits over time by county

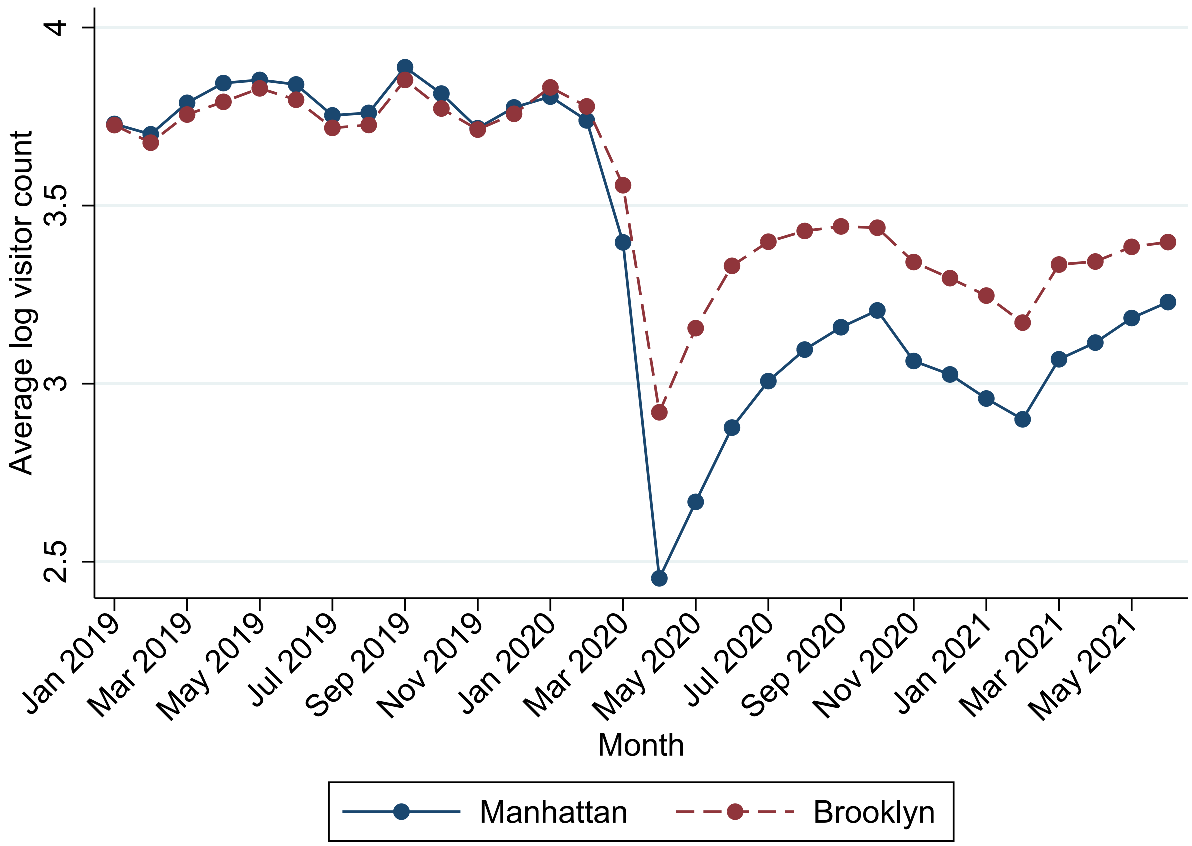


Figure: Comparison of visits over time among units in Manhattan and Brooklyn.

Figure 12: Average visits over time by industry code

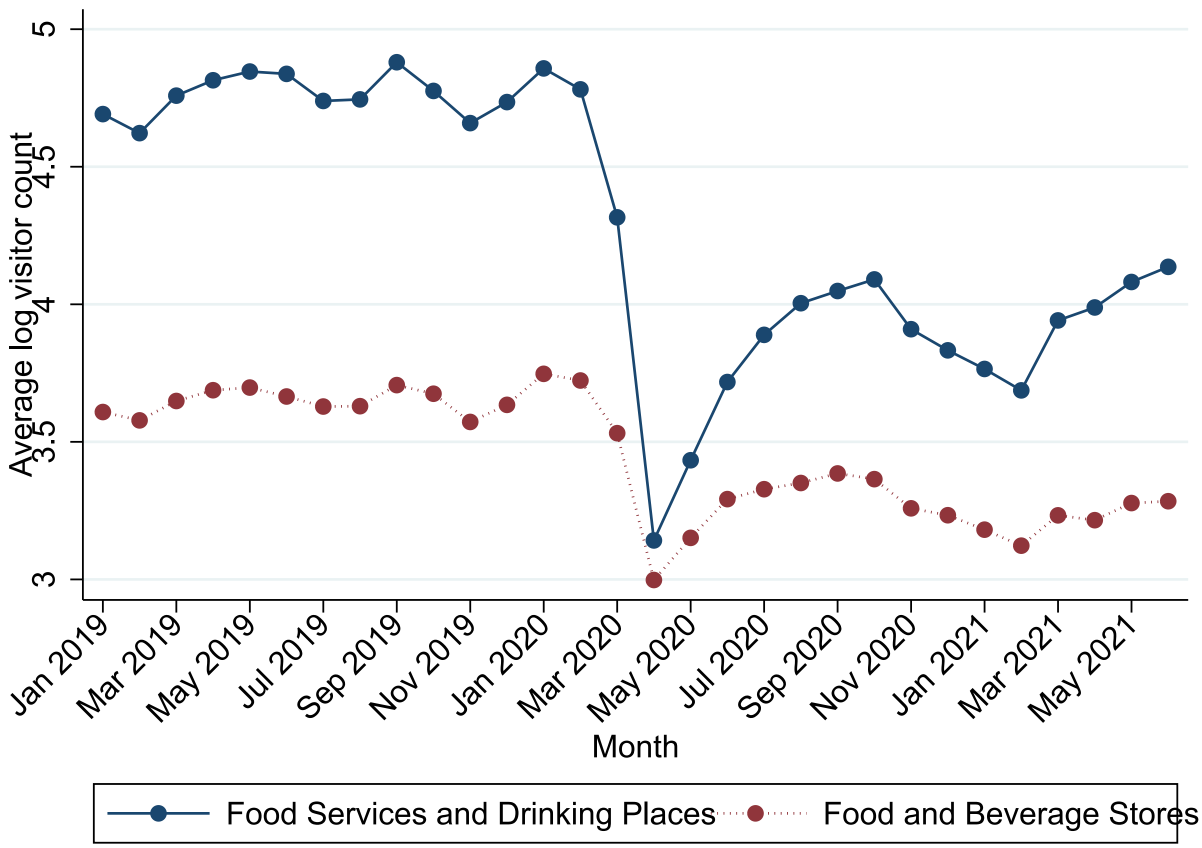


Figure: Comparison of visits over time among units with three-digit NAICS code 722 ("Food Services and Drinking Places") and those with three-digit NAICS code 445 ("Food and Beverage Stores").

Figure 13: Distribution of estimated propensity score

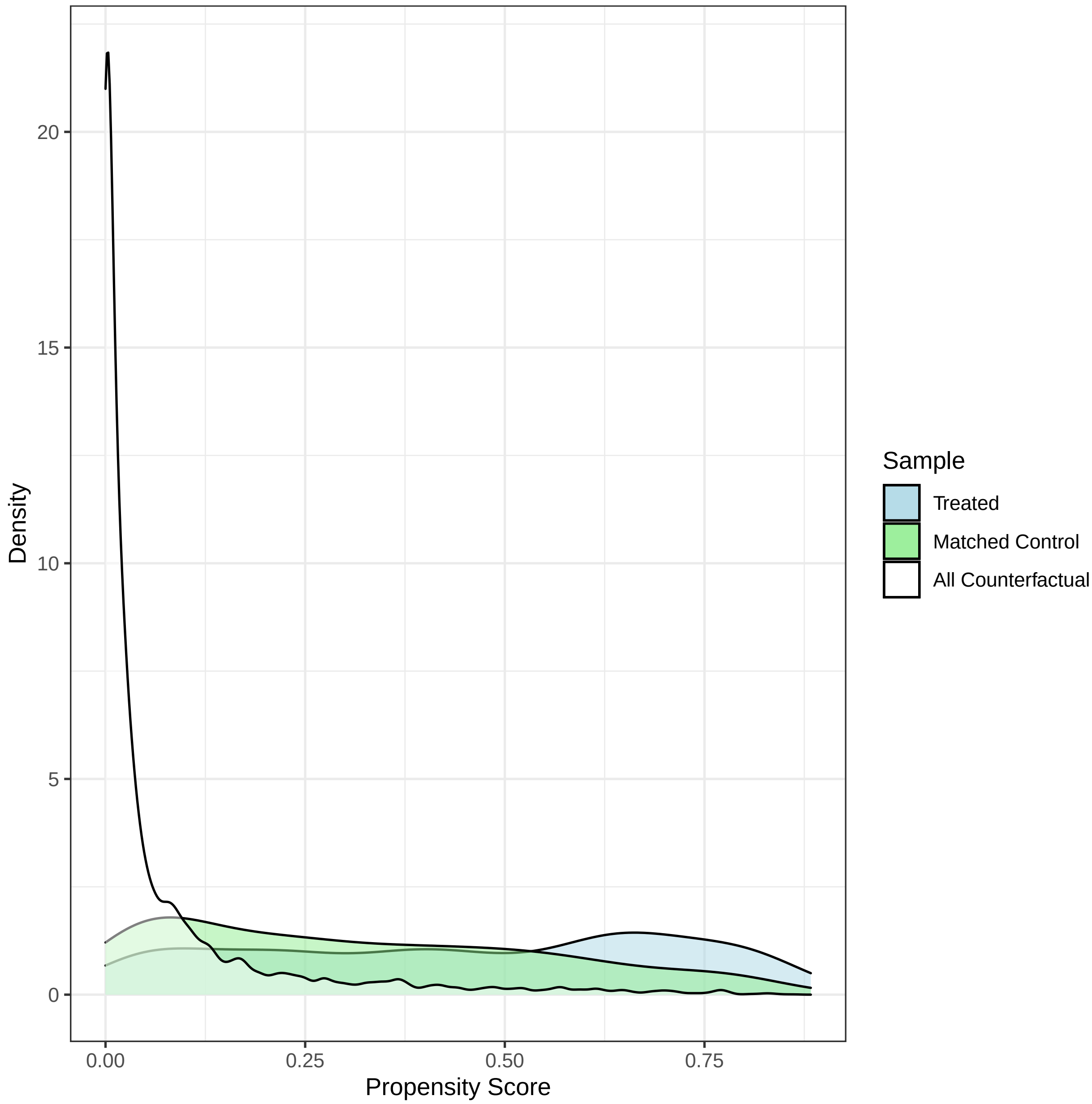
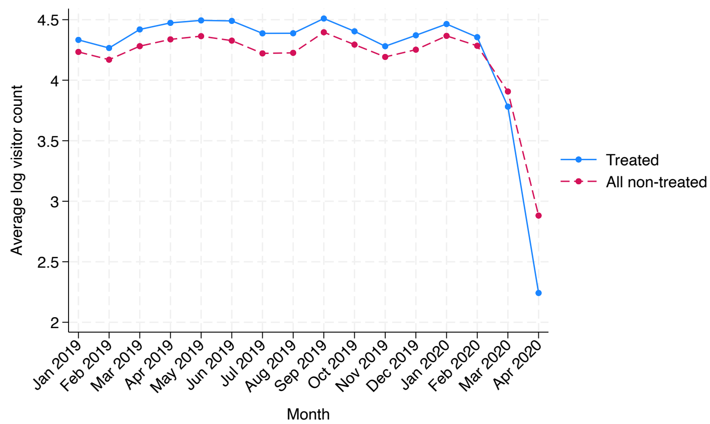


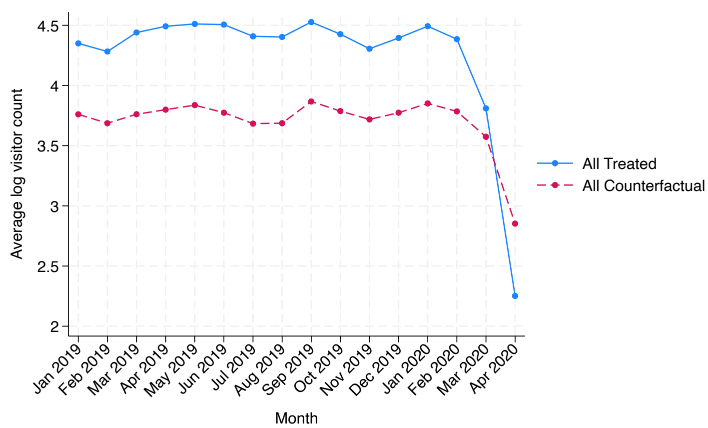
Figure: Kernel density of the propensity score for matched treated, matched control and all counterfactual units.

Figure 14: Pre-treatment dynamics in visitors

(a) Treated vs all non-treated



(b) All treated vs all counterfactual



(c) Matched treated vs matched control

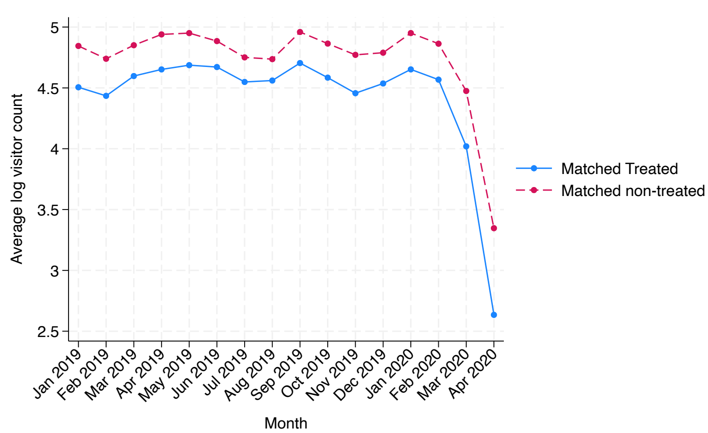


Figure 15: Workplace population share and changes in visits 2019-2020

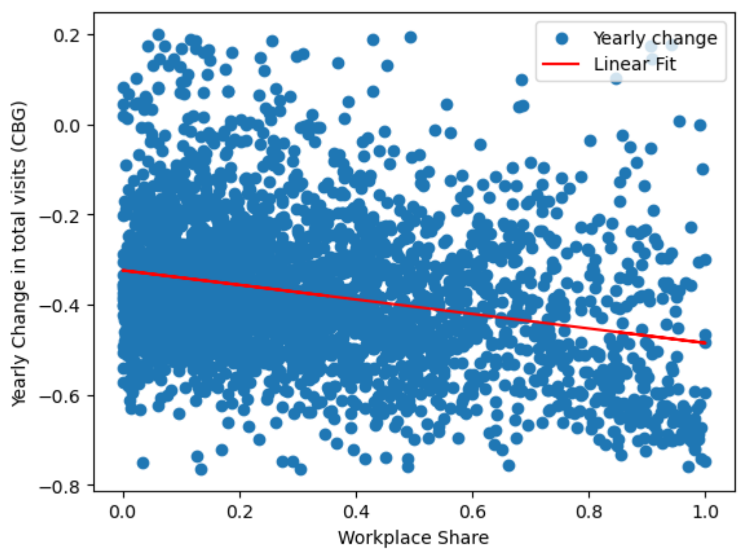


Figure: Scatter plot of the yearly proportional change in the total number of visits to establishments at the census block group (CBG) on the vertical axis and the 2019 workplace share on the horizontal axis. The workplace share is calculated as the number of people who commute into a given CBG (workplace population) divided by the sum of the number of people who commute into the CBG and the number of people who commute from that CBG to another CBG (residential population). The slope of the linear fit is -0.16, which is significant at a 1% level. We exclude yearly changes above the 99th percentile and below the 1st percentile to reduce the bias from large changes in the composition of establishments within a CBG. Sources: SafeGraph "Patterns" data and the Longitudinal Employer-Household Dynamics (LEHD).

Figure 16: Effect of Open Streets designation by industry

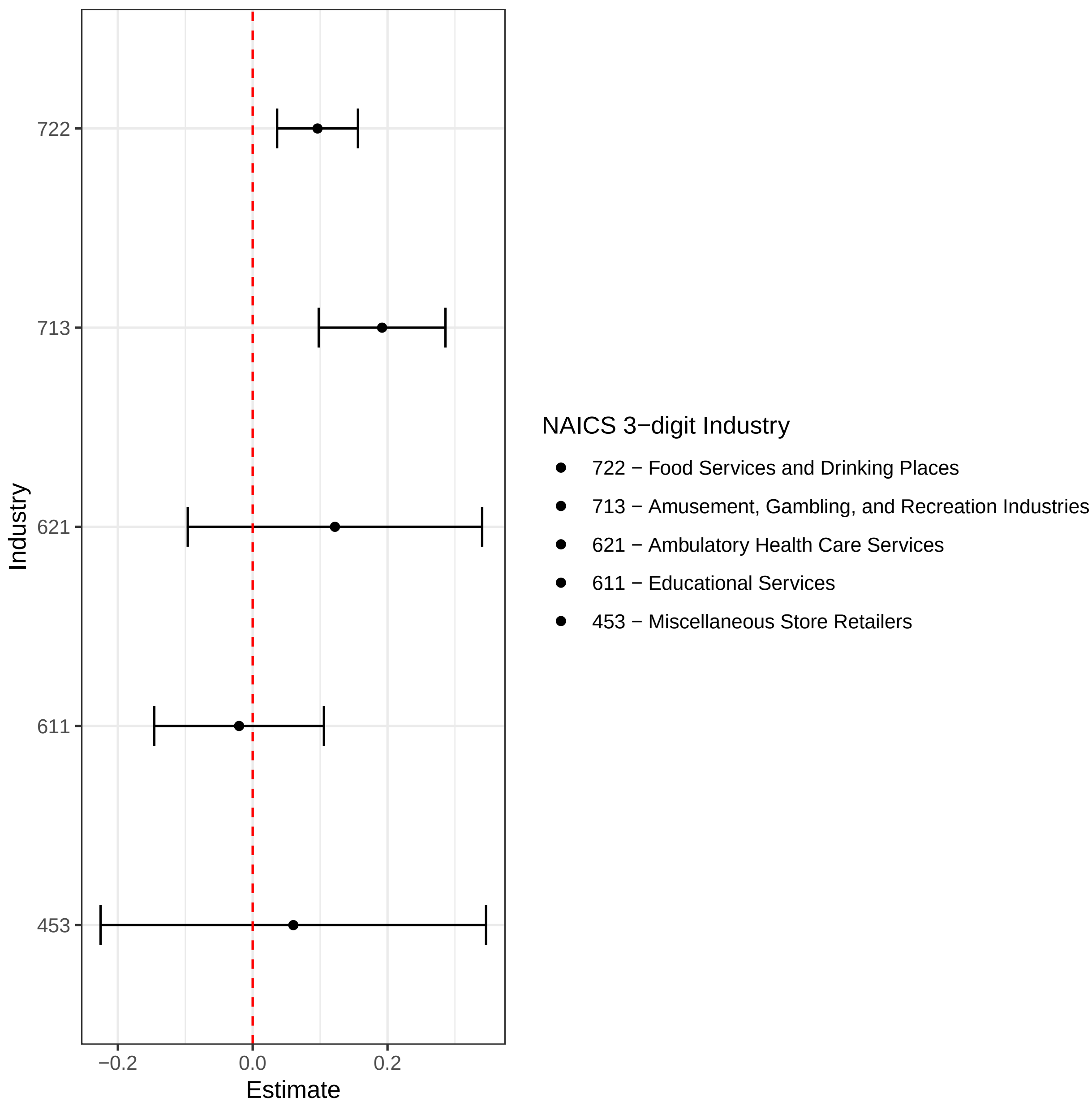


Figure: Shows regression coefficient and 95% confidence intervals from the estimtion of equation 2 with the log of the monthly number of visitors as the outcome of interest. Restricting to the subsample of businesses in each of the five largest 3-digit industry subsectors of the North American Industry Classification System (NAICS) in the matched sample.

Figure 17: Dynamic effects of Open Streets designation on visits: whole sample

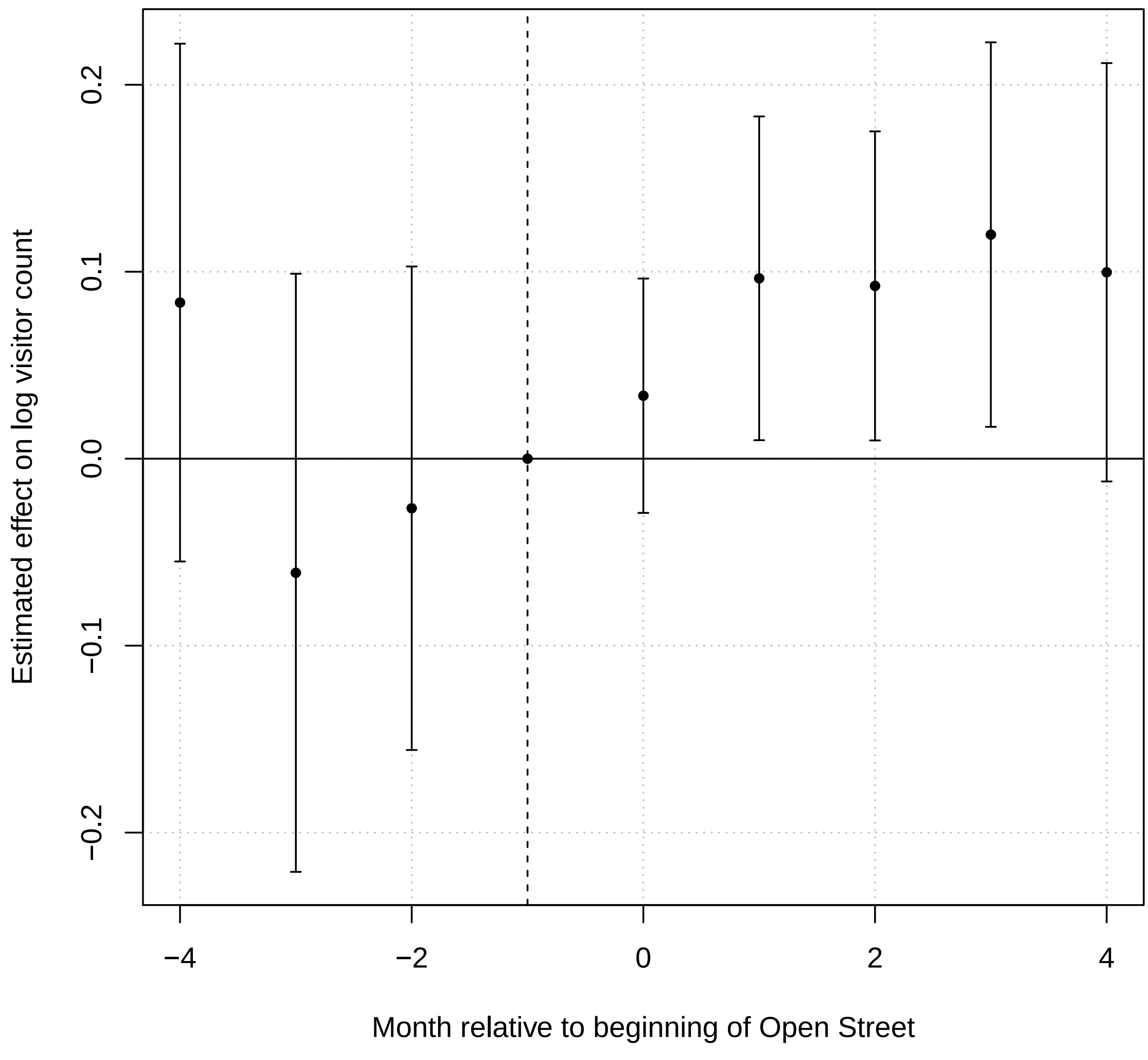


Figure: Shows regression coefficients $\gamma_k^{Treated}$ and 95% confidence intervals for each $k \in \{-4, 4\}$, with the log of the monthly number of visitors as the outcome of interest. Coefficients are normalized to month before treatment, $k = -1$. Standard errors clustered by six-digit naics code-borough interaction.

Figure 18: Dynamic effects of Open Streets designation on visits: lower workplace share ($\leq$0.75)

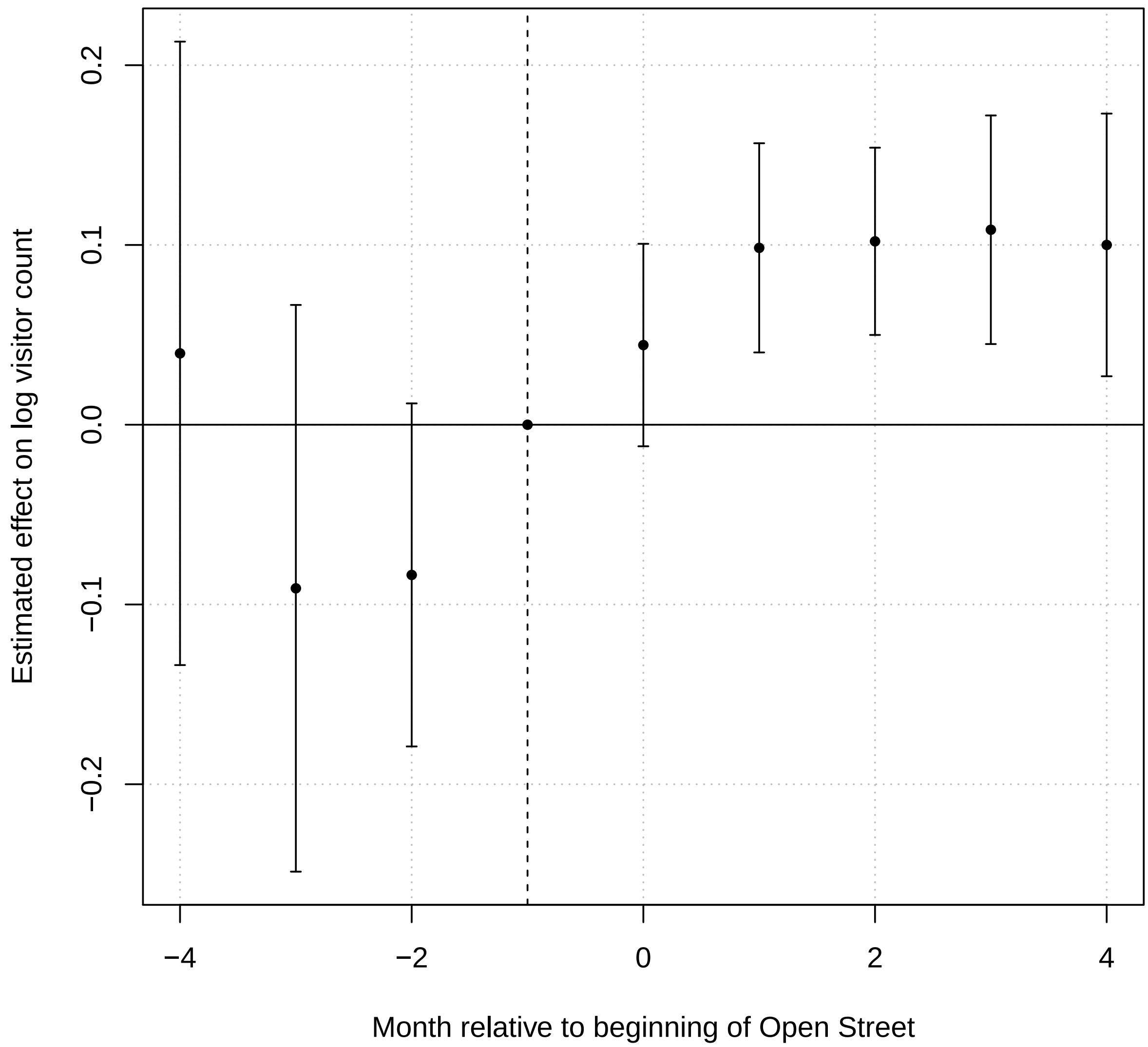


Figure: Sample restricted to only businesses in census block groups with a workplace share of at most 0.75. Shows regression coefficients $\gamma_k^{Treated}$ and 95% confidence intervals for each $k \in \{-4, 4\}$, with the log of the monthly number of visitors as the outcome of interest. Coefficients are normalized to month before treatment, $k = -1$. Standard errors clustered by six-digit naics code-borough interaction.

Figure 19: Dynamic effects of Open Streets designation on visits: high workplace share (>0.75)

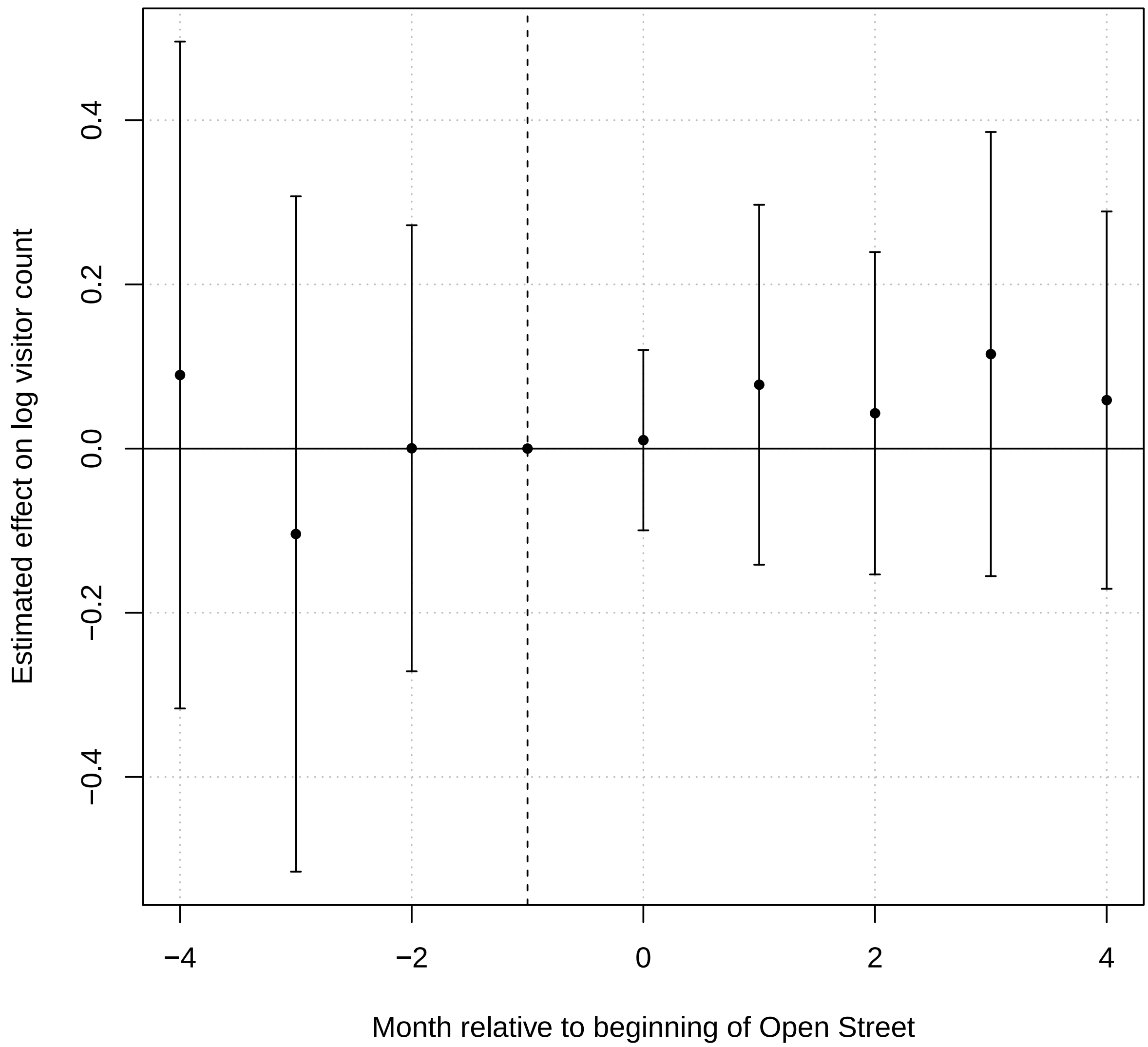


Figure: Sample restricted to only businesses in census block groups with a workplace share of more than 0.75. Shows regression coefficients $\gamma_k^{Treated}$ and 95% confidence intervals for each $k \in \{-4, 4\}$, with the log of the monthly number of visitors as the outcome of interest. Coefficients are normalized to month before treatment, $k = -1$. Standard errors clustered by six-digit NAICS code-borough interaction.

Figure 20: Dynamic effects of Open Streets designation on visits: lower workplace share (≤0.75), bars and restaurants only

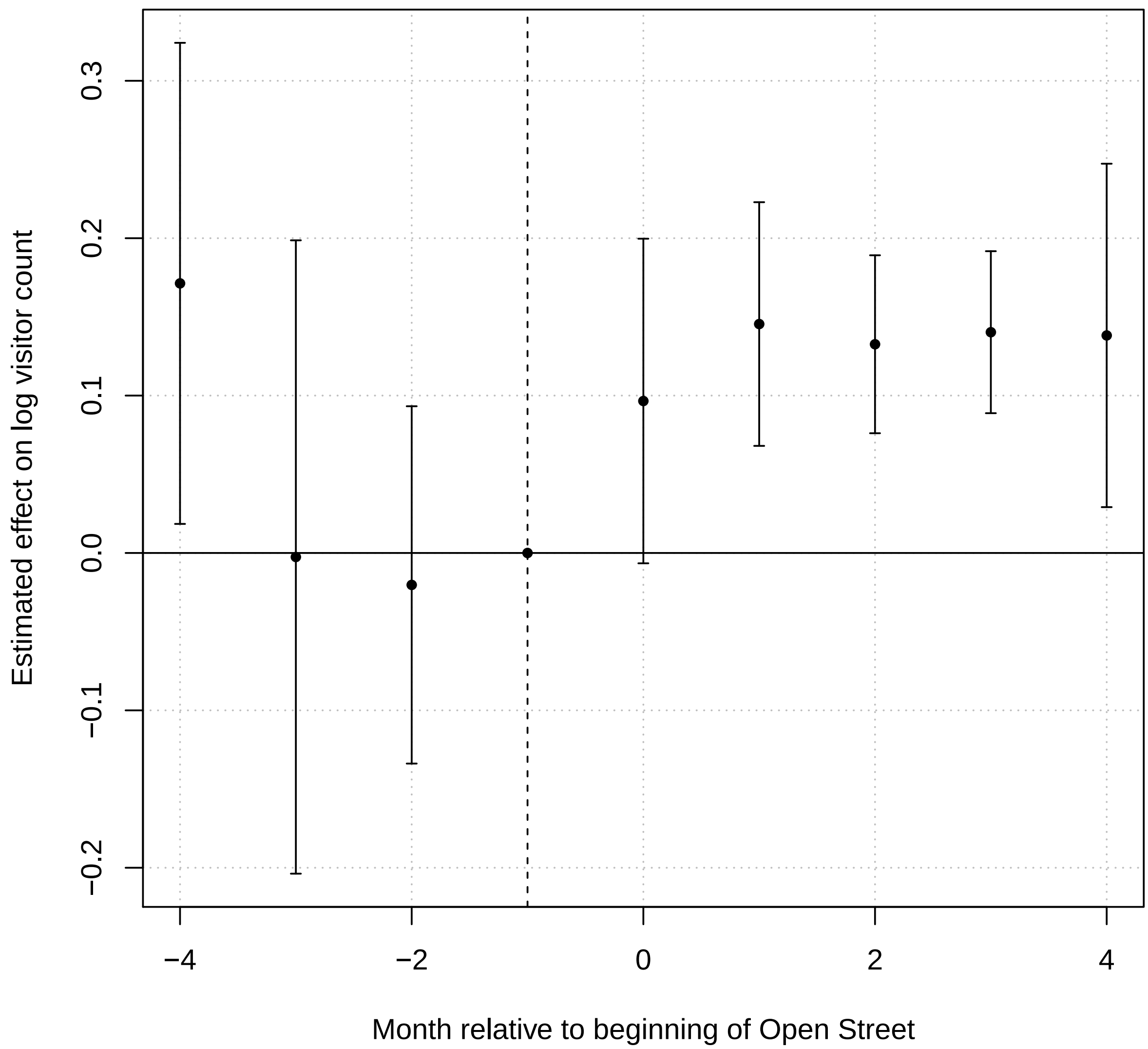


Figure: Sample restricted to only businesses in census block groups with a workplace share of at most 0.75 with NAICS code 722 (Food Services and Drinking Places). Shows regression coefficients $\gamma_k^{Treated}$ and 95% confidence intervals for each $k \in \{-4, 4\}$, with the log of the monthly number of visitors as the outcome of interest. Coefficients are normalized to month before treatment, $k = -1$. Standard errors clustered by six-digit naics code-borough interaction.

Figure 21: Spillover effects by distance to Open Street segments: full sample

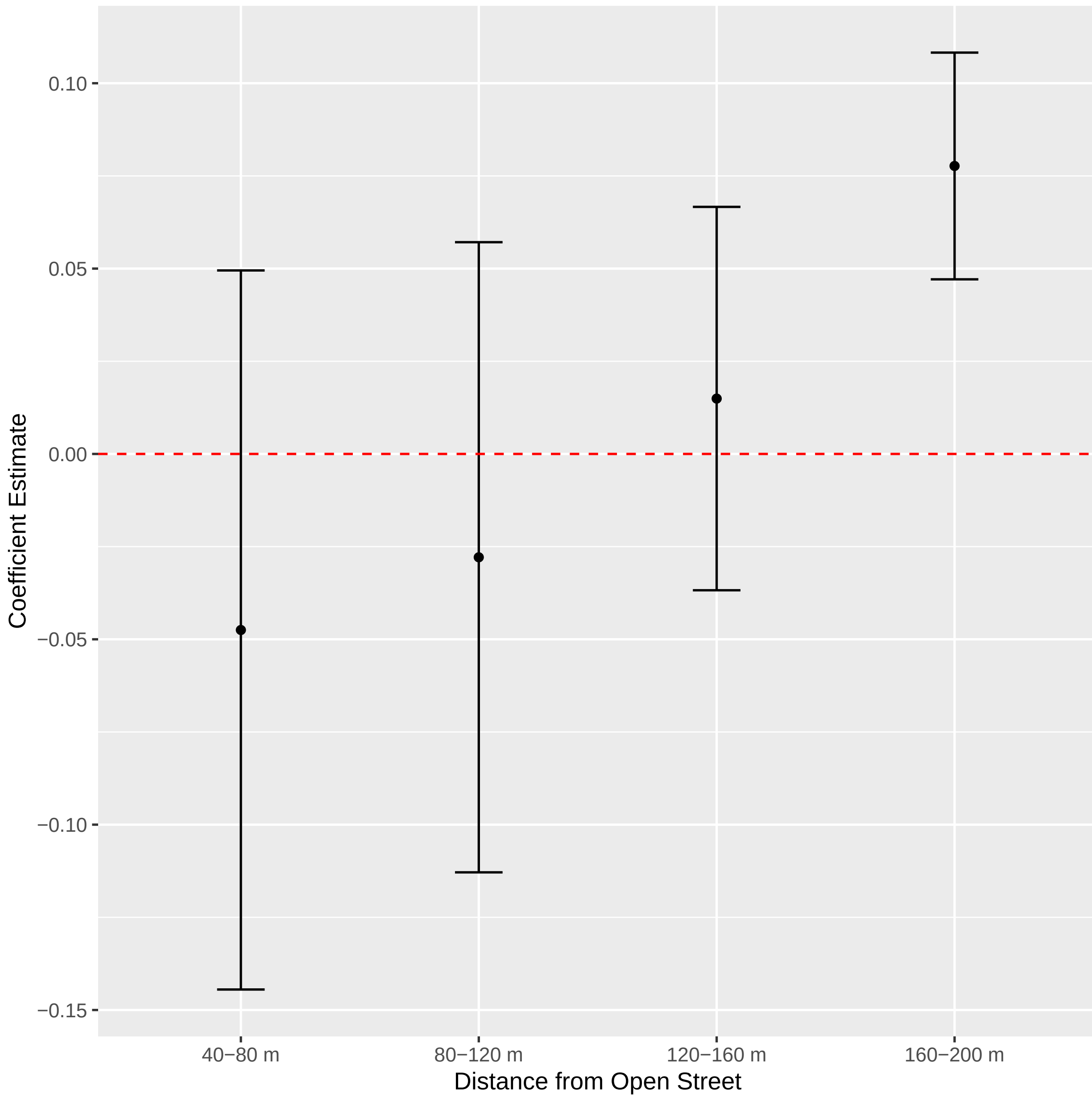


Figure: Spillover effects estimated from equation 2 with the log number of visitors as the outcome of interest, but where the treatment is defined as being at a certain distance band from an Open Street, and the control group is selected from businesses within the same distance band of a counterfactual candidate segment. Standard errors are clustered by six-digit NAICS code-borough interaction. The 95% confidence intervals are reported.

Figure 22: Spillover effects by distance to Open Street segments: lower workplace share

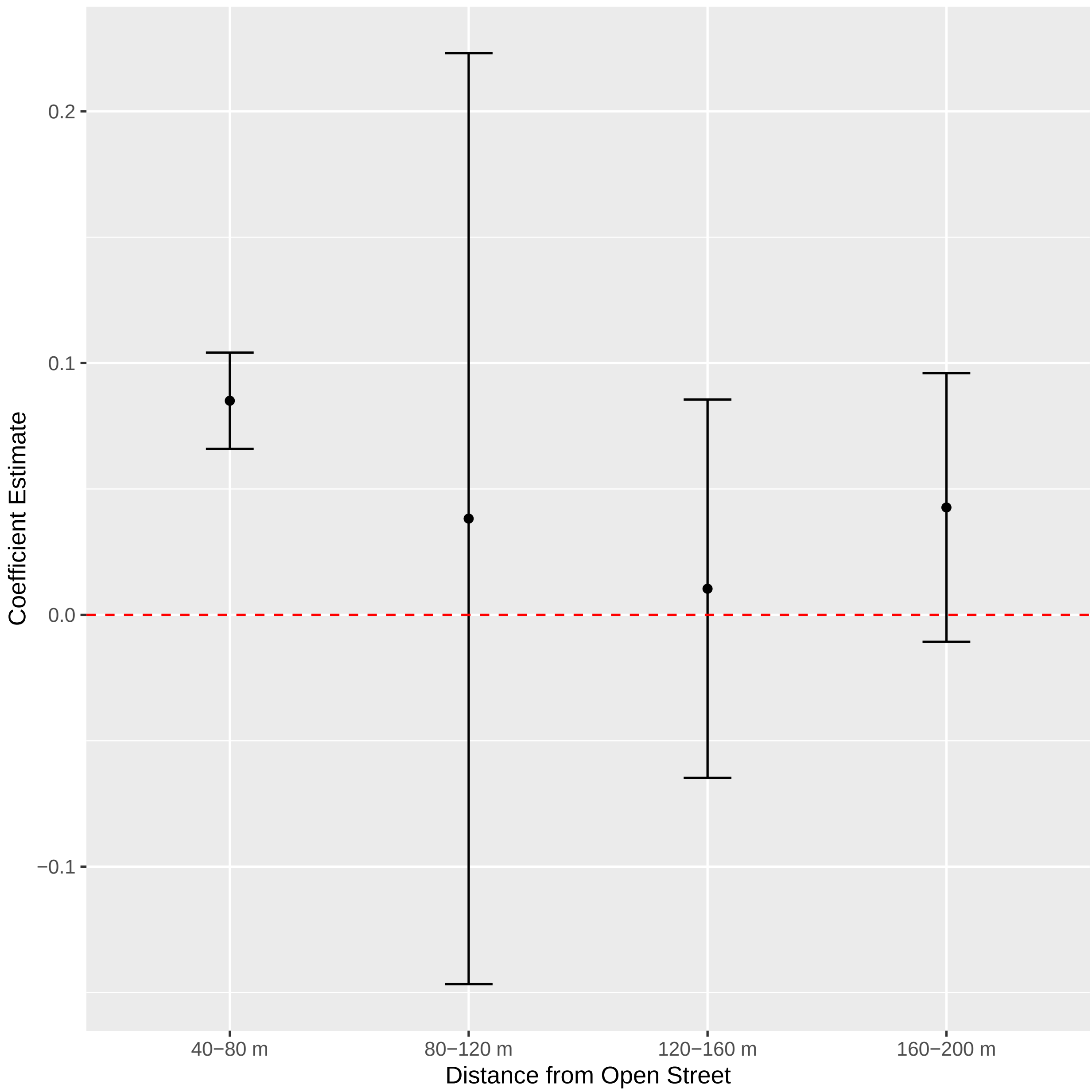


Figure: Spillover effects estimated from equation 2 with the log number of visitors as the outcome of interest, but where the treatment is defined as being at a certain distance band from an Open Street, and the control group is selected from businesses within the same distance band of a counterfactual candidate segment. Sample restricted to only businesses in census block groups with a workplace share of more than 0.75. Standard errors are clustered by six-digit NAICS code-borough interaction. The 95% confidence intervals are reported.

Figure 23: Spillover effects by distance to Open Street segments: lower workplace share and Food Services and Drinking Places

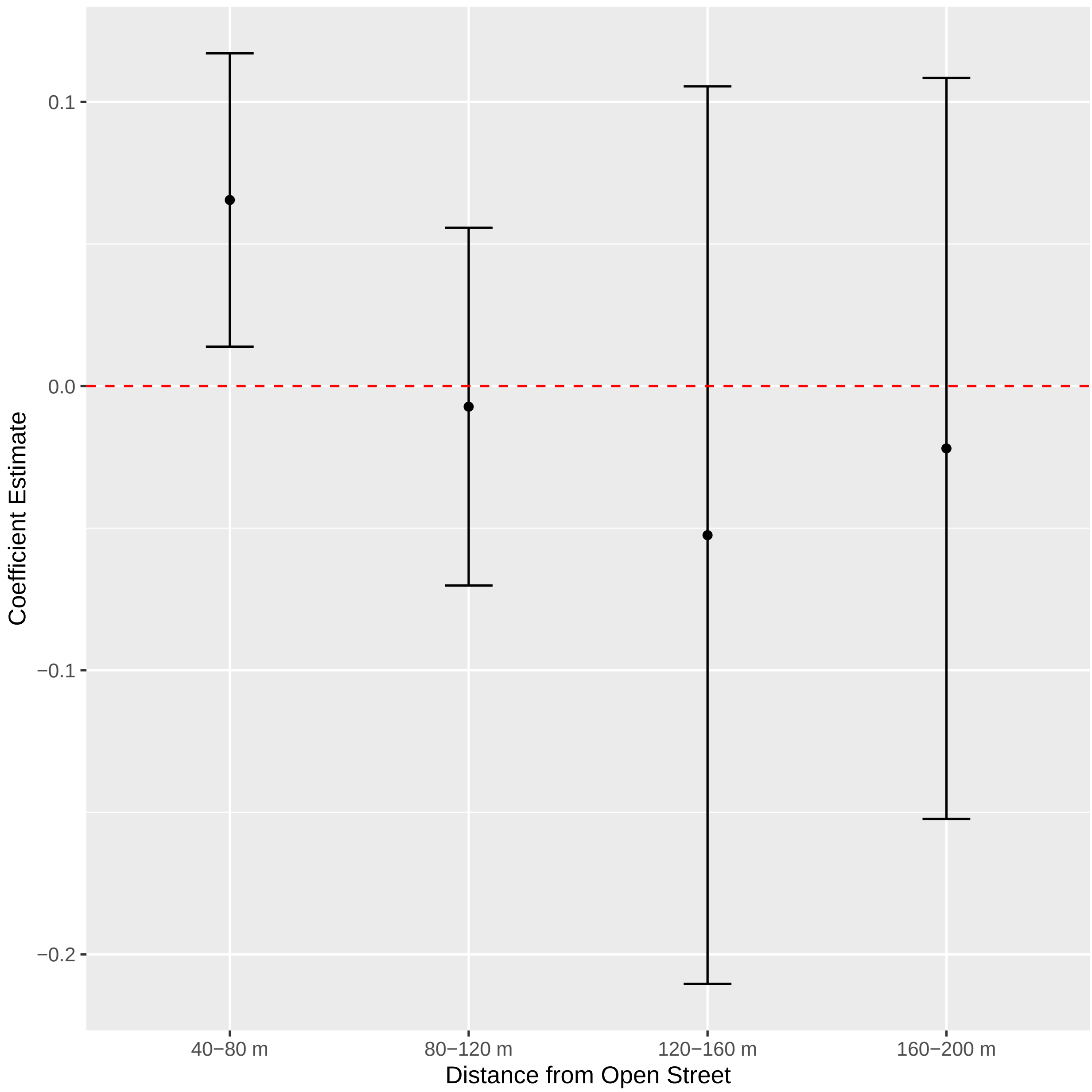


Figure: Spillover effects estimated from equation 2 with the log number of visitors as the outcome of interest, but where the treatment is defined as being at a certain distance band from an Open Street, and the control group is selected from businesses within the same distance band of a counterfactual candidate segment. Sample restricted to only businesses in census block groups with a workplace share of at most 0.75 with NAICS code 722 (Food Services and Drinking Places). Standard errors are clustered by six-digit NAICS code-borough interaction. The 95% confidence intervals are reported.

9 Tables

Table 1: Visits data summary stats

	Mean	Median	s.d.
Visit count	216.42	66.00	1,157.51
Unique visitor count	123.72	42.00	682.61
Median dwell time in minutes	107.34	46.00	162.48
Share of iOS devices	0.48	0.49	0.23
Manhattan	0.52	1.00	0.50
NAICS Food Services and Drinking Places	0.29	0.00	0.46
NAICS Ambulatory Health Care Services	0.09	0.00	0.29
NAICS Religious, Grantmaking, Civic, Professional, and Similar	0.08	0.00	0.27
NAICS Clothing and Clothing Accessories Stores	0.05	0.00	0.22
NAICS Food and Beverage Stores	0.05	0.00	0.22
NAICS Health and Personal Care Stores	0.04	0.00	0.21
NAICS Amusement, Gambling, and Recreation Industries	0.04	0.00	0.20
NAICS Miscellaneous Store Retailers	0.04	0.00	0.20
NAICS Elementary and Secondary Schools	0.04	0.00	0.19
Social Assistance	0.03	0.00	0.17

Summary stats of visits data at the business-month level. Number of observations 1,565,444. Number of businesses 63,116.

Table 2: Inputs for Convolutional Neural Net (CNN)

Inputs	**Source**
Number of open street edges in cell	SafeGraph
Number of businesses in cell	SafeGraph
Average visits to businesses in cell	SafeGraph
Number of Restaurants and other Eating Places (NAICS 7225) in cell	SafeGraph
Number of Drinking Places (Alcoholic Beverages) (NAICS 7224) in cell	SafeGraph
Number of Sporting Goods, Hobby, and Musical Instrument Stores (NAICS 4511) in cell	SafeGraph
Number of Clothing Stores (NAICS 4481) in cell of industry code	SafeGraph
Number of Other amusement and recreation industries (NAICS 7139) in cell	SafeGraph
Number of Offices of Physicians (NAICS 6211) in cell	SafeGraph
Number of Health and Personal Care Stores (NAICS 4461) in cell	SafeGraph
Total population in census block group	ACS 2019
Median household income in census block group	ACS 2019
Female population in census block group	ACS 2019
White population in census block group	ACS 2019
Population of ages 1-17 in census block group	ACS 2019
Population over 25 years of age in census block group	ACS 2019
Population over 25 years of age with at least College education in census block group	ACS 2019

This table summarizes the variables used to build the input layer in the convolutional neural net (CNN) which is then trained to predict the presence of an Open Street segment. The variables were constructed from two sources: SafeGraph's "Patterns" data, and the American Community Survey from 2019 at the census block group level.

Table 3: Balance table: treated units vs all businesses in counterfactual streets

Variable	(1) All Counterfactual	(2) All Treated	(3) Difference
Number of businesses, average 2019	7.44	17.22	9.79***
	(7.36)	(17.86)	(0.36)
Share Restaurants and Bars, average 2019	0.23	0.46	0.22***
	(0.28)	(0.33)	(0.01)
Distance to CBD	9.92	3.57	-6.35***
	(4.74)	(3.13)	(0.13)
Median dwell time, average 2019	130.04	82.40	-47.64***
	(162.49)	(76.61)	(4.23)
Brooklyn indicator	0.67	0.19	-0.48***
	(0.47)	(0.39)	(0.01)
Observations	3,407	1,629	5,036

Table compares the characteristics of treated units (on Open Streets) that meet the inclusion criteria described in section 4, and all units in counterfactual streets. Standard errors in parentheses.

Table 4: Balance table: treated units vs matched control

Variable	(1) Matched Controls	(2) Matched Treated	(3) Difference
Number of businesses, average 2019	7.93	9.82	1.89***
	(8.26)	(9.49)	(0.44)
Share Restaurants and Bars, average 2019	0.51	0.51	-0.00
	(0.35)	(0.35)	(0.02)
Distance to CBD	4.39	4.23	-0.16
	(3.20)	(2.92)	(0.15)
Median dwell time, average 2019	91.20	71.51	-19.68***
	(109.63)	(68.23)	(4.47)
Brooklyn indicator	0.23	0.23	-0.00
	(0.42)	(0.42)	(0.02)
Observations	833	833	1,666

Table compares the characteristics of treated units and matched control units, matched using the procedure as described in 4. Standard errors in parentheses.

Table 5: Difference-in-differences estimates: Full Sample and by workplace share

	(1) Full sample	(2) Low workplace share	(3) High workplace share
Post x treated	0.09***	0.12***	0.06
	(0.03)	(0.04)	(0.09)
Constant	3.81***	3.74***	3.93***
	(0.01)	(0.01)	(0.03)
Event-time month fixed effects	Yes	Yes	Yes
Business fixed effects	Yes	Yes	Yes
Observations	14994	9756	5238

Results based on estimates of equation 2 with the log of the monthly number of visitors as the outcome of interest. Standard errors clustered by six-digit NAICS code-borough interaction in parentheses. Column (1) reports the estimation for the full sample. Columns (2) reports the estimation results for the subsample of units in a census block group with a workplace share that is at most 0.75. Column (3) reports the estimation results for the subsample of units in a census block group with a workplace share that larger than 0.75.

Table 6: Difference-in-differences estimates by industry

	(1) NAICS=722	(2) NAICS=713	(3) All other
Post x treated	0.10***	0.19**	0.06
	(0.02)	(0.08)	(0.03)
Constant	3.88***	4.18***	3.59***
	(0.01)	(0.03)	(0.01)
Event-time month fixed effects	Yes	Yes	Yes
Business fixed effects	Yes	Yes	Yes
Observations	9090	1008	4896

Results based on estimates of equation 2 with the log of the monthly number of visitors as the outcome of interest. Standard errors clustered by six-digit NAICS code-borough interaction in parentheses. Column (1) reports the estimation results from a regression using only the subsample of units that are in the **Food Services and Drinking Places** subsector (NAICS 722). Column (2) reports the estimation results from a regression using only the subsample of units that are in the **Amusement, gambling and recreation industries** (NAICS 713). Column (3) reports the estimation results from a regression using the subsample of units that are in the remaining industries (different NAICS classification than 722 and 713).