Do Urban Parks Promote Racial Diversity? Evidence from New York City*

Timur Abbiasov†

Columbia University‡

October 24, 2020

Preliminary and incomplete, please do not cite without permission.

Abstract

While policymakers and urban planners often praise public space for promoting racial diversity, empirically it remains unclear whether these places play a role in creating diverse social environments for the city's residents. Focusing on parks in New York City as the exemplar of modern public space, I estimate the causal relationship between access to parks and individually experienced diversity. To do so, I introduce a measure of racial diversity that captures one's level of exposure to diverse others in places visited on a daily basis, utilizing a novel dataset featuring individual GPS tracking data for more than 60 thousand Twitter users in the New York metro area. My empirical strategy relies on obtaining a time-varying measure of access to parks that incorporates information about ongoing construction and repair works across the city. The results show that additional 10 acres of parkland within the 5km radius from home increase individual chances of encounters with other racial/ethnic groups by 1%. I also provide evidence to suggest that park accessibility affects the diversity of white and black residents differently: for parks located closer to home, the effect appears to be more pronounced for whites than blacks, suggesting that parks in the majority-white neighborhoods are able to attract a broader range of visitors compared to the local parks in black neighborhoods.

JEL Classification: J15, Q26, L83, R23
Keywords: Public space, parks, social interactions, diversity, racial segregation

*I am very grateful to Donald Davis, Reka Juhasz and David Weinstein for their invaluable guidance and continued support. I also thank Juan Francisco Saldarriaga, Lance Freeman, Dmitry Sedov, Dmitry Sorokin, Iain Bamford, Pablo Ernesto Warnes, Howard Zihao Zhang, Yue Yu and seminar participants at Columbia for helpful discussions and comments. All errors are mine.

†Email: t.abbiasov@columbia.com

‡Department of Economics, 420 W. 118th St., New York, NY 10027

Contents

1 Introduction

Economists have long been invested in the idea that social interactions play an important role in determining individual outcomes. Unfortunately, being surrounded by conducive neighbors or peers is a privilege not equally allowed to everyone in American cities. In fact, due to the enduring presence of residential segregation, this disparity is particularly harmful to racial and ethnic minorities, and especially so for African Americans (Boustan, 2012; Massey, 1990). While most of the empirical evidence examining policies designed to mitigate these adverse effects focuses exclusively on residential choices[1], some prominent scholars have argued that shared spaces such as parks, cafes, and libraries play ”a critical but underappreciated role in modern societies” (Klinenberg, 2018) and affect the lived experience of diversity by creating opportunities for meaningful interaction across ethnic lines (Anderson, 2011). Public parks, in particular, have attracted interest among city planners (Langegger, 2013; Low et al., 2009), and policymakers[2] as a viable investment offering the potential to integrate the city’s diverse communities and to promote social tolerance between their diverse members.

However, empirically it remains unclear whether urban parks play a role in creating diverse social environments for the city’s residents. In this paper, I investigate the relationship between access to park space and individually experienced diversity. To do so I exploit a novel dataset containing six months of GPS tracking data for more than 60 thousand Twitter users in New York City. This data allows me to obtain a measure of individual diversity that captures one’s level of exposure to diverse others in places visited on a daily basis (similar to Athey et al. (2020) and Xu et al. (2019)). To identify the effect of interest, I construct a time-varying measure of individual park access that incorporates data on various construction works that temporarily limit access to certain park areas across the city. This empirical strategy allows me to circumvent the endogeneity problem that arises in the cross-section setting due to residential sorting on unobservables affecting individual behavior.

[1] See Gatreaux (Popkin et al. (1993)) and Moving to Opportunity (Kling et al. (2007)) studies, also Oreopoulos (2003) and Vigdor (2002), who study the impact of re-locations arising from administrative assignment to public housing projects.

[2] For example, in the strategic plan for New York released in 2015 by Bill de Blasio and The City of New York it is outlined that ”Parks and public space are essential to [...] promoting interaction ...” (Bill de Blasio and The City of New York, 2015).

I establish three main results. First, my findings confirm that improved access to park space has a sizable effect on the individually experienced diversity. More specifically, the results indicate that additional 10 acres of parkland within the 5km radius from home increase individual chances of encounters with other racial/ethnic groups by 1%. In other words, an average-sized community park[3] increases the chances of encounters across racial or ethnic lines by 2-5%. Second, I document a non-monotonicity in the estimated effect with respect to distance from home. For an average user in my sample, parks located closer to the residence location (less than 1km away) contribute less to experienced diversity than parks located further away: the effect peaks around the 1-2 km range and then fades out rather quickly. Importantly, this finding emphasizes the importance of parks designated to serve wider geographic areas (as opposed to local parks in racially/ethnically more uniform neighborhoods) and suggests that location betweenness plays a role in determining the extent to which a certain park is able to promote diversity. Third, I find evidence that suggests that park accessibility affects the diversity of white and black residents differently: for parks located closer to home, the effect appears to be more pronounced for whites than blacks, meaning that parks in the majority-white neighborhoods attract a broader range of visitors compared to the local parks in black neighborhoods.

My findings have several important implications. First, I provide empirical evidence to support the claims made by many social scientists, policymakers and urban planners arguing that the provision of park space is essential for promoting opportunities for racially and ethnically diverse encounters. This paper hence emphasizes the role of parks – and, more generally, of public space – in nurturing diverse social environments and suggests a viable policy that can help cities reduce the racial and socioeconomic isolation without resorting to the complicated and costly measures that operate through residential choices. Furthermore, the obtained results indicate that the positioning of parks with respect to residential communities affects the extent to which they are able to promote racially diverse encounters: parks serving wider geographic areas and located slightly further away from everyone's residences have the highest impact on diversity, while parks located in direct proximity to the black users' homes appear to produce a substantially weaker effect.

[3]Based on average area of 20-40 acres.

The main dataset used in this paper is based on the collection of nearly all geotagged Twitter posts published between June and December 2014 and originating within the New York metro area.[4] The key advantage of focusing on this specific period comes from the fact that prior to 2015 every tweet tagged with some location, even something as broad as "New York City", automatically exposed the precise GPS coordinates of the device through the API – a policy that many users may remain unaware of (Wired, 2019). For each individual, I then use a commercial machine learning solution to process the profile image and obtain the perceived ethic or racial attributes. Combined with the information on residence locations – also inferred from the patterns of online activity – this approach essentially allows me to create a detailed representation of how different racial or ethnic communities co-locate daily in urban space. Section 2 offers a more thorough look at the construction of the final dataset.

To measure experienced diversity, I map each user's locations reported in a given month into a grid of geographic units that are approximately 150m wide and 150 long[5], thus setting the spatial and temporal resolution at which exposure to other people is measured. Accordingly, for each user, I define the monthly diversity index as the *expected share of other racial or ethnic groups in the total pool of people this individual is exposed to when not at home*, where the probabilities of visiting each place for every person are assigned using monthly visits. Essentially, in this context, two individuals are seen as more 'exposed' to each other when they visit the same locations more frequently in a given month (but not necessarily at the same time). While this measure does not capture the actual social connections, it provides a way to interpret how much of the same space different racial group cohabit on a daily basis. As previously noted by other authors, this outcome is of interest by itself, as it captures the opportunities for casual encounters and a sense of shared experience (Athey et al., 2020; Klinenberg, 2018; Anderson, 2011). Additional information on measuring experienced diversity is presented in Section 3.

The data I use in this study also comes with several limitations. First of all, not everyone uses Twitter, and hence the sample of individuals is not random. Nonetheless, when

[4]According to Kinder-Kurlanda et al. (2017), who provide the files necessary for re-downloading the geotagged posts data via the Twitter API, over 90% of all geotagged tweets in the US posted between June and December 2014 were exposed to their data collection algorithm.

[5]Known as *geohash7*

compared to the census population of age 18 to 45, the sample used in this study appears to be similar to the representative along a number of dimensions (see Section 2 for more details). Second, even if some users were unaware that including any geotag whatsoever in a public post automatically implies exposing their exact location, posting a tweet and attaching information about the location remains a voluntary decision. Consequently, the present approach allows identifying only a limited subset of the actual movements for each user in a given month. While the exact nature of this subset is not well documented, Drakonakis, Ilia, Ioannidis, and Polakis (2019), for example, have argued that it is rich enough to not only successfully identify the home address for the majority of the users, but also to uncover a significant amount of "sensitive" locations that users have visited, i.e. those pertaining to health, religion, and sex/nightlife.[6]

This work is closely related to recent studies in the literature that use phone or online activity data to measure social segregation. One of the first such studies by Davis, Dingel, Monras, and Morales (2019) examines consumption segregation in New York using data on restaurant visits obtained through the Yelp platform, and find that venue choices are only about half as segregated as residences. In another particularly relevant study, Athey, Ferguson, Gentzkow, and Schmidt (2020) demonstrate how anonymized location data from smartphones can be used to measure *experienced segregation*, capturing city residents' exposure to diverse others in the places they visit on a daily basis. Importantly, the authors find that experienced isolation tends to be substantially lower than the corresponding residential isolation measures. Finally, Xu, Belyi, Santi, and Ratti (2019) utilize a dataset of call detail records from Singapore to validate another similarity measure based on imputed SES levels, which takes into account both the physical 'co-location' and connections over the mobile network. Interestingly, they find that the degrees of individual isolation in the communication network and urban space are not tightly related.

This paper distinguishes itself from the mentioned studies in two important ways. First, in the present study, I am able to establish one of the causal mechanisms through which city planning affects individually experienced segregation. Second, I use a more technologically advanced approach for inferring personal attributes (such as race or ethnicity) that automates

[6] See also, Qian et al. (2017).

the classification of profile pictures with machine learning tools and allows me to conduct a large scale analysis without having to rely on indirect and arguably less reliable SES imputation methods (such as used in Athey et al., 2020; Xu et al., 2019, who use anonymized phone records).

Table 1: User Activity Statistics

Monthly, by User	Mean	Median	20% pct.	80% pct.
# of active days in a month	18.11	16.16	9.50	26.25
# of unique locations visited in a month (geo7)	17.18	12.33	6.67	24.00
# of unique locations visited in a month (geo6)	11.97	9.20	5.00	16.83
# of unique day-location pairs in a month (geo6)	26.46	20.67	11.00	38.00
Combined	Total			
# of tweets	15,622,601			
# of users	60,765			

Notes: (1) User statistics are calculated using monthly averages for each uses; (2) *geohash7* grid consists of rectangles approx. 150m by 150m; (3) for *geohash6* the dimensions are approx. 1.2km by 600m

2 Data

2.1 Twitter Posts and Users

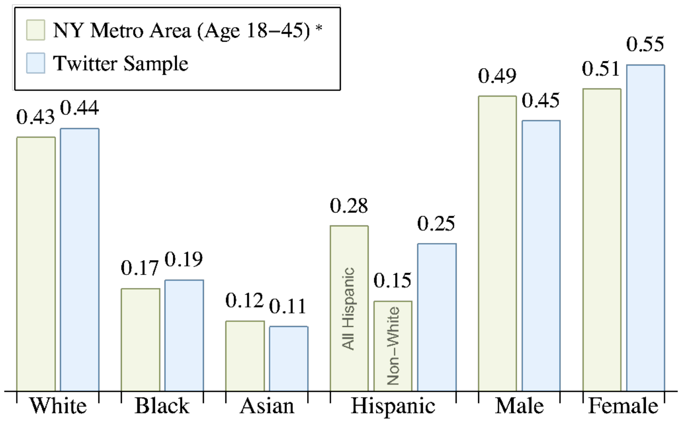
Twitter is a popular microblogging platform allowing people to exchange ideas, real-time information, and latest news in the form of short messages. In the last quarter of 2014 the number of its monthly active users in the US reached 63 millions. Due to the service popularity and the unprecedented amount of personally identifiable geospatial information it offers to researchers, Twitter data provides new opportunities for locating and measuring daily individual activities in the urban areas.

According to the company's policy, user messages (called *tweets*) can optionally include a *geotag* that provides some details about the location of the person at the time of posting. Importantly, prior to 2015 every message with a geotag, even something as coarse as 'Bronx' or 'NYC', revealed the precise GPS coordinates of the device through the the meta-data available to the developers.

In this paper I use a dataset of more than 15 million geotagged tweets posted between June 1 and December 1, 2014, originating within the New York metro area. In addition to the main text and a pair of GPS coordinates, the data associated with each message includes a time marker, user name and surname, a profile picture, and references to other tweets or users.

Table 1 summarizes the geospatial information collected about each user. On average, users are active during 18 days in a given month (conditional on reporting at least once in

Figure 1: Sample Description: Racial and Gender Attributes



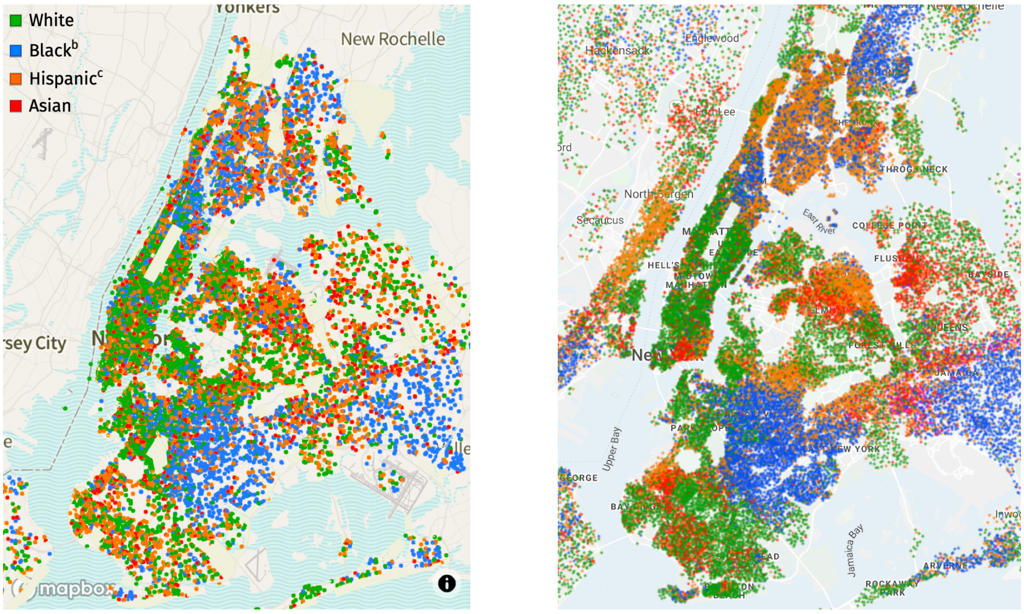
Note: New York metro aggregates are obtained from the American Community Survey 5-year estimates for 2011-2015.

that month) and according to this metric there's a substantial amount of variation between users. In terms of unique geographic locations, an average user visits about 10-20 distinct places in a single month, with an average of about 2-3 days spent in each location, and the frequency of days per location also varies substantially between users.

I use profile pictures in order to obtain the perceived racial or ethnic attributes for the users in my dataset. More specifically, each image is processed via the Clarifai[7] web-service that uses machine learning tools to predict the most likely racial/ethnic group for the person in the photo (see more details in the Appendix A.1). Figure 1 displays a summary of the resulting dataset, confirming that major racial and ethnic groups are reasonably well represented in the Twitter sample as compared to the census population of the New York metro area in the age of 18-45.

The analysis present in this study relies on using each user's residence location as a proxy for the most likely origin of his/her daily trips. However, such information is not publicly available and thus has to be inferred from the patterns of online activity. Similarly to the commonly adopted methodology,[8] I define the individual place of residence as the GPS location of the largest cluster of nighttime tweets (posted between 8PM and 6AM) that is reported at least five times over a period spanning at least two weeks between the last and first post. Figure 2 confirms the credibility of my data imputation strategy by offering a visual comparison of the identified home locations combined with imputed racial/ethnic attributes

[7] https://www.clarifai.com
[8] See, e.g. Athey et al. (2020), Wang et al. (2018), Xu et al. (2019).



(A) Twitter Data (B) Census 2010[9]

Figure 2: Home locations by race/ethnicity compared

from the Twitter data with a similar map created by The New York Times using the data from Census 2010.

In order to better understand the characteristics of a typical user's residence characteristics, I matched the information on inferred home locations with the tract-level data from the American Community Survey 5-year estimates for 2011-2015. Table 8 in the appendix describes the home-tract environment of a median user in my sample as compared to the representative tracts of the New York metro area, New York City and Manhattan, correspondingly. I find that the median income in those tracts where Twitter users reside is relatively high – about 77 thousand dollars, which is closer to that of Manhattan tracts (81.9 thousand) than to the level of the larger metro area (67.9 thousand). The median share of residents with a college degree or higher among Twitter users' tracts is 38%, for NY metro area it is 32%, while for Manhattan the corresponding number reaches almost 70%. When looking at the the living unit characteristics, it appears that Twitter users reside in places with a relatively high share of owner-occupied housing compared to the median tracts of both Manhattan and New York City. Finally, census tracts populated by Twitter users typically have more than a half of units with 2 or more bedrooms, which is similar to what I find for a representative tract in New York City.

[9] www.nytimes.com/interactive/2015/07/08/us/census-race-map.html

TABLE 2: PARKS

		NYC	Manhattan
Total number of parks		2101	417
Park size (acres)	Mean	8.255	6.787
	Median	0.664	0.437
	20% pct.	0.114	0.091
	80% pct.	2.126	1.751
# of parks, by avg. monthly visitors	# of unique Twitter users		
	[1-2)	940	169
	[2-5)	422	105
	[5-20)	152	68
	[20-100)	65	35
	[100-1000]	17	10

2.2 Parks and Capital Projects in New York

The data about parks in New York comes primarily from the ParkServe database that tracks urban park access nationwide and includes data for 13,931 cities and towns in the U.S.[10]. As summarized in Table 2, New York City accommodates about 2100 parks. According to ParkServe, 99% of the city's residents live within a ten minute walk from the nearest green-space area, making New York's park system one of the most accessible in the country[11]. The same table provides details on the distribution of park sizes in the city, for example, showing that the median park in New York spans about 0.7 acres, while in Manhattan parks are slightly smaller with a median area of 0.44 acres.

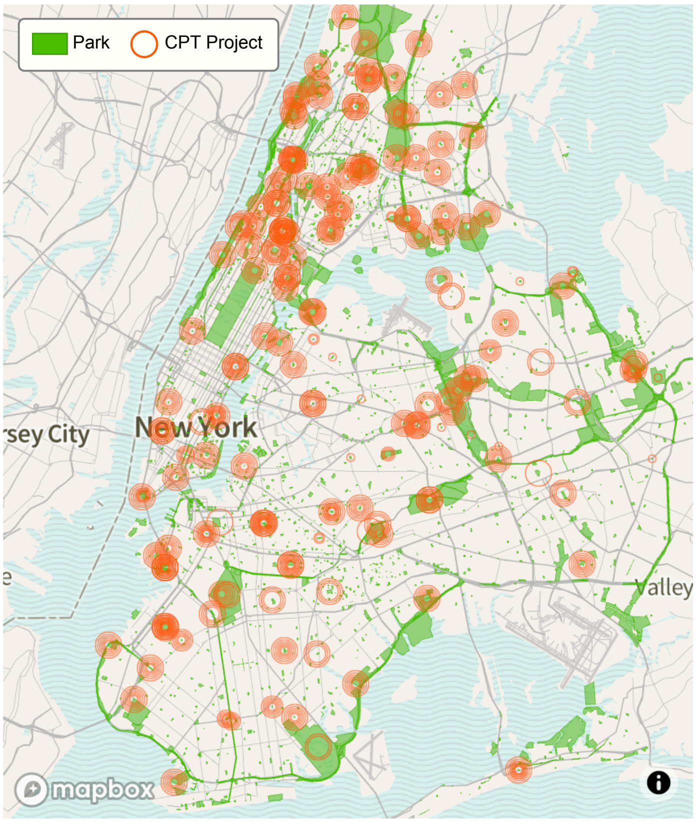
Table 2 also displays information about park visitors as reflected in the Twitter dataset. In New York City, I am able to identify 17 parks with a monthly average of more than 100 unique visitors, 65 parks with 20 to 100 visitors, about 150 parks with an average between 5 and 20 users, and for almost a thousand locations I can only detect between 1 and 2 monthly visitors. Parks in Manhattan exhibit a denser coverage, with about 25% of all parks having at least 5 monthly visitors during the period of Jun - Dec 2014.

All of the city's long-term infrastructure investments related to parks, including projects such as the construction or reconstruction of parks and playgrounds, installation of fencing

[10] https://www.tpl.org/media-room/trust-public-land-makes-park-information-database-and-platform-available-mill

[11] The Trust for Public Land, link: https://www.tpl.org/parkserve/

FIGURE 3: NEW YORK CITY PARKS AND CAPITAL PROJECTS



and benches, and various other repair and improvement works are being managed by the NYC Parks' Capital Projects division[12]. Provided by the City Parks Department, Capital Project Tracker is an open resource that allows to track the progress of every single capital project that took place in New York since the initiative was launched in the early 2014, including information about funding and the precise timeline for design, procurement and construction phases of each project. Figure 3 maps all of the capital projects that were in progress in New York City parks during the period of June - December 2014, displaying 235 projects in total.

[12] www.nycgovparks.org/capital-projects

3 Experienced Diversity

In this section I introduce a measure of individual diversity that captures each persons' exposure to others in the places visited on a daily basis. As further explained in Section 3.1 below, I define the monthly individually experienced diversity as the expected share of other racial or ethnic groups in the total pool of people the individual is exposed to outside of home, where the probabilities of visiting each geographic location in a given month are inferred from personal travel data. Section 3.2 demonstrates how the proposed diversity index can be estimated using the GPS tracking data collected from Twitter and Section 4.1 provides a brief descriptive analysis of the distribution of obtained estimates in the New York metro area.

3.1 Definition

Consider a setting with multiple individuals indexed by i and each belonging to one of the several groups indexed by g. Let's assume that on every day of the month each individual decides either to stay home or to visit one of the geographic locations indexed by j, and that the probability of visiting j on a given day, denoted as P_{ijt}, only varies between months.

Consequently, for an individual i from group g, I define the overall level of experienced diversity in places this person visited over the course of the month t as

$$ExperiencedDiversity_{it}(g) = \frac{\sum_{j \in \mathcal{J}} P_{ijt} \times \overline{V}_{jt-g}}{\sum_{j \in \mathcal{J}} P_{ijt} \times \overline{V}_{jt}}, \tag{1}$$

where $\overline{V}_{jt-g}$ is the average number of people from groups different than g who visit j on any given day in t, and $\overline{V}_{jt}$, correspondingly, is the overall monthly average of users visiting j on a typical day.

3.2 Estimation

Implementing the experienced diversity measure as defined in (1) requires obtaining the probability estimates for each user and location in a given month. To do so, I map the entire set of geolocated Tweets into rectangular geographic cells that are approximately 150m wide and 150m long (commonly denoted as *geohash7*s). Correspondingly, I compute the individual probability of visiting each such cell in two steps.

First, I consider a set of wider geographic destinations, each consisting of 16 adjacent *geohash7* boxes (and, in turn, constituting the *geohash6* grid), and define the probability of visiting each *geohash6* box on a given day as the proportion of all unique day-location pairs reported by the user from this cell. After that, for each user I estimate the conditional probabilty of visiting each smaller *geohash7* cell, given that the report comes from a particular *geohash6* location. This allows me to compute the probability of visiting each *geohash7* cell using the chain rule:

$$\widehat{P}_{ijt} = \widehat{P}_{it}\big[J = j\big] = \widehat{P}_{it}\big[J = j \,|\, J \in Nbhd_{geo6}(j)\big] \times \widehat{P}_{it}\big[J \in Nbhd_{geo6}(j)\big] \tag{2}$$

The main motivation for adopting this two-step approach comes from the fact that while Twitter users often report multiple adjacent *geohash7* locations on the same day, the probability term introduced in (1) assumes that only one location can be visited on each day. To see why this matters for computing individual diversity, consider an example case when the observed monthly data displays a particular user tweeting from 6 adjacent *geohash7* cells on every odd-numbered day, and also reporting from a single different, sufficiently remote *geohash7* on every even-numbered day. If one simply assigns to each *geohash7* cell the monthly share of unique location-day pairs in the travel history of this user, that person's presence in the locations visited on the odd days will be oversampled from the daily probability viewpoint. On the other hand, by following the two-step method described above one would obtain a more realistic conclusion that the chances of visiting each of the broader the destinations on a given day for this user are equal, and in the event of going to the first one, the user decides randomly between the 6 options within that destination.

Table 3: Experienced Diversity Summary Statistics

	All	White	Black[a]	Hispanic[b]	Asian
Number of users	46, 984	20,981	8,886	11,977	5,160
Avg. monthly diversity					
Mean	0.643	0.503	0.719	0.732	0.876
Median	0.654	0.498	0.776	0.754	0.879
St. Dev.	0.194	0.138	0.193	0.123	0.076
Benchmark diversity level[c]	-	0.553	0.811	0.745	0.89
% of users below	-	73.1	60.0	45.9	60.0
Avg. monthly diversity gap[d]					
Mean	-	-0.050	-0.092	-0.014	-0.015
Median	-	-0.055	-0.034	0.009	-0.011

Notes: (a) Black or African American; (b) Hispanic, latino, or Spanish origin; (c) Benchmark levels for each group reflect the diversity level under uniformly random mixing; (d) Gap levels reflect absolute distance from the benchmark;

4 Descriptive Statistics

4.1 Experienced Diversity

Table 3 provides a summary of the experienced monthly diversity levels estimated for the sample of Twitter users residing in the New York metro area, observed during June - December 2014. Overall, I find that white and black residents are experiencing the lowest exposure to diverse others compared to asian and hispanic users. In particular, the average monthly experienced diversity for black residents is equal to 0.719, which corresponds to 71.9% chance of encounters with other racial or ethnic groups. For blacks, this is 9.2% less than the benchmark probability under the uniform mixing scenario (i.e. no segregation). For whites the average monthly diversity is about 50% and the corresponding gap from the unform-mixing benchmark is roughly 3.5%. For hispanic and asian users I find the average diversity gap to be considerably lower: 0.9% and 1.1% correspondingly.

4.2 Differences in Access and Visits to Parks

Before proceeding to the main results I investigate whether the four major racial/ethnic groups exhibit differences in greenspace access and whether they differ in their decisions regarding park visits.

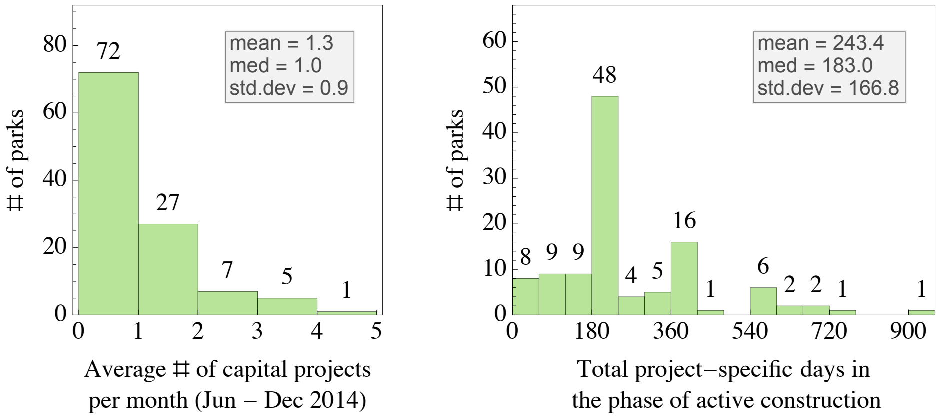
Table 9 describes the average total park acreage available to each user in the sample within two distance bands from the user's home location. The first (0-1km) distance range captures parks in close proximity, and the second one (0-5km) captures park access measured over a wider geographic area. I find that white and hispanic residents have significantly more park space in close proximity to home compared to blacks. Black residents on average have access to 241 acres of greenspace within a one-kilometer radius from home, while the average white and hispanic users have access to 266-267 of park space in the same distance range. Other pairwise comparisons between the mentioned groups, however, do not indicate any statistically significant differences.[13] These conclusions fall in line with the previously published works in the urban planning literature that examine spatial disparities in park and greenspace access and conclude that non-whites tend to live in neighborhoods with lower amounts of park space (Saporito and Casey, 2015; Heckert, 2013; Boone et al., 2009).

To further investigate the racial disparities in park usage, in Figure 5 I examine the composition and traveled distances for the visitors of parks located in either predominantly black or predominantly white residential neighborhoods. Interestingly, while parks in the majority-white neighborhoods tend to serve a significant population of black residents, who often travel more than 4-5 kilometers to these parks, the reverse is not true: white residents are far less likely to travel longer distances to visit parks in the majority black neighborhoods. It seems reasonable to interpret this as suggesting that parks in majority white neighborhoods have certain qualities that differentiate them from the parks in black neighborhoods and make them overall more attractive for a broader population of visitors. While previous literature offers no empirical evidence to shed light on the factors that make parks in white neighborhoods more widely attractive[14], speculatively, I find it likely that parks in more affluent neighborhoods offer a better variety of complementary amenities in their vicinity, are generally better funded and maintained, and are more likely to be situated in a locations with a high centrality index or good transport access.

[13] Difference-in-means tests are reported in panel 9b

[14] To the best of my knowledge, the only empirical work to date that examines differences in park usage is Gobster (2002), who conducted on-site surveys of visitors in Lincoln Park, Chicago. One of his key findings is that 80% of white park users came from neighborhoods closest the park, compared to only 60% of minority users.

Figure 4: Capital Projects Construction During June - Dec 2014



(a) monthly frequency (b) Number of days in construction phase

4.3 Construction Activity in Parks

In figures 4a and 4b I use the timelines provided in the Capital Project Tracker to examine how often each of the parks in my dataset had planned construction works. Overall, I am able to identify 112 of NYC's parks where at least one construction project was implemented during Jun - Dec 2014. Correspondingly, among the parks with active projects managed by the Capital Projects division, the average frequency of scheduled works is about 1.3 per month. Figure4b further reveals that slightly less than a half of these parks had ongoing construction for the entire period of observation, i.e. 183 days.

5 Main Results

5.1 Suggestive Evidence

To investigate the relationship between access to park space and individual experienced diversity, I first estimate a cross-section model where the independent variable combines measures of total accessible park acreage (in the vicinity of the user's home location) and the outcome of interest is the individual average monthly experienced diversity. More specifically, I estimate the following equation:

$$ExperiencedDiversity_i = TotalParkArea_i \beta + \varepsilon_i \tag{3}$$

where diversity is measured as the average monthly diversity index for individual i: $x_i = \bar{x}_{it}$, and $TotalParkArea_i$ is a vector corresponding to the total park area within consecutive distance bands from the user's home location.

Table 4 presents the estimated coefficients. When pooling of all of the city's residents together (column 1), I find that a one standard deviation increase in the total park acreage accessible to the user within a 5-km radius from home leads to a 2.1% increase in experienced diversity, or, in other words, improves the chances of individual's encounters with other racial or ethnic groups by 2.1 percent. In column 3 I demonstrate that the effect is even more pronounced (3.2%) in the subsample of white and black users. Furthermore, in columns 2 and 4 I document that the estimated coefficients behave non-monotonically with respect to distance from the user's residence location. For parks located in direct proximity (between 0 and 1 km) the effect is two to three times smaller than for parks in the mid-range distance (1 to 3 km): in fact, the effect peaks in the 1 to 2 kilometers range and fades out reaching zero at about 5 km distance. These results suggest that parks designated to serve wider geographic areas, which are generally located somewhat further away from the local communities, contribute more to creating opportunities for casual encounters across ethnic or racial lines than local parks.

Even though the estimates of the cross section specification offer several insights regarding the relationship between individual access to parks and exposure to diversity, they can

Table 4: Access to Parks and Diversity: Cross-Section Model

Park Area (standardized)	Experienced Diversity: All Residents (1)	(2)	Black or White (3)	(4)
total (acres)	0.021***		0.032***	
	(0.003)		(0.005)	
0-1km		0.006***		0.011***
		(0.002)		(0.003)
1-2km		0.015***		0.020***
		(0.002)		(0.003)
2-3km		0.012***		0.019***
		(0.002)		(0.003)
3-4km		0.010***		0.015***
		(0.002)		(0.003)
4-5km		0.004*		0.007*
		(0.003)		(0.004)
Observations	25,303	25,303	15,613	15,613
County FE	✓	✓	✓	✓
R^2	0.015	0.017	0.047	0.052

Note: Standard errors are robust to heteroskedasticity and clustering at the tract level. *p<0.1; **p<0.05; ***p<0.01

not be interpreted as causal. The primary endogeneity concern in estimating 3 is that users' unobservable characteristics affecting diversity are likely to be correlated with residential choices, and hence, access to park space. In the following section I outline an empirical strategy exploiting the panel dimension of my dataset, allowing me to estimate the effect of interest while accounting for static unobserved heterogeneity between users.

5.2 Construction Works and Access to Parks

One common feature among most of the capital projects is that ongoing construction tends to either reduce or completely prevent access to certain park areas or facilities. Appendix A.2 provides several illustrative examples from the Park Department's website page that lists notices about the upcoming works and warns potential park visitors about the scheduled disruptions in the operation of various park-related amenities. The aim of this section is to outline an approach that allows me to use the NYC Parks' data on construction timelines during the period of June - Dec 2014 to obtain a time-varying measure of the amount of parkland effectively available to the public in different locations across the city.

Table 5: Park Visits and Current Construction

	Dependent variable:
	Log daily visits
SizeCurrentConstruction	−0.025***
	(0.006)
Observations	114,730
Fixed effects	park, date
R^2	0.610

Note: The independent variable corresponds to the number of individual capital projects under construction on park's territory, and the measures of daily park visits are based on the Twitter dataset. Standard errors robust to heteroskedasticity and clustering at the tract level are reported in parentheses: *p<0.1; **p<0.05; ***p<0.01

The primary goal of the proposed procedure is to obtain a measure that reflects how much park space is available to each user and takes into account the spatial and temporal distribution of ongoing construction projects that limit access to certain parks. Investigating the central questions raised in this paper requires using a measure that is easy to interpret and that allows spatial aggregation (and disaggregation) across users and neighborhoods. Hence, for each park I propose to estimate the effectively served acreage: the total acreage multiplied by the park's relative visitor capacity that takes values between zero and one, depending on how much of the park's space is occupied with construction projects.

Let's start with assuming that c_{jt} is the unobserved variable that describes the capacity of a given park j at time t, defined as follows:

$$Capacity_{jt} \equiv c_{jt} = \frac{V_{jt}(x_{jt})}{V_{jt}(0)} = \frac{v_{jt}}{V_{jt}(0)} \tag{4}$$

where $V_{jt}(x)$ is the number of people who decide to visit park j at time t in the event when exactly x construction projects are taking place in that park. Note the $V_{jt}(x_{jt})$ is thus the observed number of visits, v_{jt}, and $V_{jt}(0)$ is the number of visits in the potential scenario when the park operates in full capacity (i.e. no construction at time t).

Let's further assume that on average an additional construction project reduces the park's capacity by a given percentage amount, namely:

$$\log(c_{jt}(x_{jt})) = x_{jt}\beta + \varepsilon_{jt}, \tag{5}$$

where $\mathbb{E}[\varepsilon_{jt} x_t] = 0$. If we denote the logarithm of the potential demand for park j at time t, $\log V_{jt}(0)$, as ψ_{jt}, (4) and (5) imply that

$$\log v_{jt} = x_{jt}\beta + \psi_{jt} + \varepsilon_{jt}.$$

Under the assumption that the park's potential demand (or, equivalently, ψ_{jt}) is independent of the number of currently active construction projects after controlling for park-level unobserved heterogeneity and seasonality, the effect of additional capital project on park's capacity, which I denote as β, can be estimated using the following panel specification with park and time fixed effects (θ_j and γ_t):

$$\log v_{jt} = x_{jt}\beta + \theta_j + \gamma_t + \epsilon_{jt} \tag{6}$$

Table 5 presents estimates for the model in described in Equation 6. The dependent variable in this case is the number of daily visits to each park in New York as inferred from the Twitter dataset, and the independent variable is the total number of construction projects operating within each park on a given day. I find that on average, a single construction project on a given day reduces the number of visitors by 2.5%. Moreover, these estimates allow me to obtain each park's predicted capacity at time t as a function of the observed construction activity:

$$\widehat{Capacity}_{jt} = \widehat{\log v}_{jt} = \exp(x_{jt}\widehat{\beta})$$

Therefore, I suggest to use the following approximation for estimating the area that each park can effectively serve to the public:

$$\widehat{AccesibleParkArea}_{jt} = \widehat{Capacity}_{jt} \times TotalArea_j \tag{7}$$

Most importantly, by defining access to parkland in this particular way, I am able to obtain a time-varying measure of park availability that is expressed in real units (acres) and can be easily aggregated over multiple locations and time intervals.

TABLE 6: INDIVIDUAL PARK VISITS AND ACCESS TO PARKS

	Dependent variable: Number of monthly visits to parks with active construction			
Accessible Park Area (acres)	Black or White (1)	Black Only (3)	White Only (5)	All Other Users (7)
Total $_{(0\text{-}5km]}$	0.013*** (0.003)	0.018*** (0.006)	0.010*** (0.004)	0.015* (0.008)
Observations	57,431	15,504	41,927	32,385
User FE	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
R^2	0.712	0.603	0.469	0.788

Note: Standard errors robust to heteroskedasticity and clustering at the individul level are reported in parentheses: *p<0.1; **p<0.05; ***p<0.01

5.3 Identification

Since individuals with different levels of access to parkland may differ in their unobserved characteristics that affect diversity, I propose to estimate the effect of interest via a panel specification with user and time fixed effects[15], as follows:

$$ExperiencedDiversity_{it} = AccessibleParkArea_{it}\beta + \theta_j + \gamma_t + \epsilon_{it}, \tag{8}$$

In the above model, the variation in the explanatory variable derives from the geographic distribution of park construction works (as explained in 5.2), and hence the identifying restrictions in this case require that the timing of the capital projects occurring in the vicinity of each user's home location is exogenous with respect to the time-varying individual unobservables.

Note that in order to be able to interpret the estimates of the proposed model as causal effects, one first needs to establish that the main regressor of interest ($AccessibleParklandArea_{jt}$) impacts individual visits to those parks in the city where construction is underway. If that is the case, then the coefficient of interest, β, can be understood as the change in individually experienced diversity that is facilitated by parks that serve as additional places

[15]Including time fixed effects is important to filter out the common seasonality part in diversity and scheduled park construction works.

of shared destination. Consequently, I test this assertion using a panel regression with fixed effects similar to (8):

$$ParkVisits_{it} = AccessibleParkArea_{it}\beta + \theta_j + \gamma_t + \epsilon_{it} \tag{9}$$

The results reported in Table 6 confirm that for both white and black or African American individuals it holds the total accessible park area has a positive effect on the frequency of visits to the parks with active CPT projects. However, for all other racial groups (i.e. asian and hispanic) I don't find sufficiently strong empirical evidence to verify that the temporal disruptions in park accessibility have an impact on the observed choices. Hence, taking into account that the degree of isolation experienced by white and black individuals is typically higher than that of hispanic and asian users in the sample (as suggested by Table 3) – rendering the central question of this study arguably less relevant for latter two groups – in estimating (8) I only focus on white and black individuals.

5.4 Individual Diversity and Access to Parks

In order to estimate the effect of the available parkland area on the individual experienced diversity, I use a monthly panel of Twitter users residing in New York City during the period between June 1 and Dec 1 2014. The key independent variable of interest is the effective amount of parkland available to each user within a 5 km radius from his or her home location. To compute it, I use the approach outlined in Section 5.2 to obtain the predicted daily capacities of each park using the timeline of construction works provided by the Parks Department. Then, for a given individual, month and radius r, I define:

$$AccesibleParkArea_{it}(r) = \sum_{\mathrm{dist}(i,j)<r} \overline{Capacity}_{jt} \times TotalArea_j,$$

where $\overline{Capacity}_{jt}$ denotes the average estimated capacity of park j during the month t, and $TotalArea_j$ denotes the area of the park measured in acres.

Main results are presented in Table 7. The estimates in the first row correspond to the effect of the total accessible park area within a 5 km radius from the user's home location. Using the entire sample of black and white individuals in NYC, I find that improved park

Table 7: Experienced Diversity and Access to Parks

Accesible Park Area (acres)	Dependent variable: *ExperiencedDiversity*$_{it}$					
	Black or White		Black Only		White Only	
	(1)	(2)	(3)	(4)	(5)	(6)
Total $_{(0\text{-}5km]}$	.0009**		.0004		.0014***	
	(.0004)		(.0007)		(.0004)	
(0-1km)		.0004		.0013		.0016*
		(.0007)		(.0012)		(.0009)
[1-2km)		.0009***		.0012**		.0010***
		(.0003)		(.0006)		(.0003)
[2km +)		.0002		.0002		.0003*
		(.0002)		(.0003)		(.0002)
Observations	57,431	57,431	15,504	15,504	41,927	41,927
User FE	✓	✓	✓	✓	✓	✓
Month FE	✓	✓	✓	✓	✓	✓
R^2	0.712	0.712	0.603	0.604	0.469	0.469

Note: Standard errors robust to heteroskedasticity and clustering at the individul level are reported in parentheses: *p<0.1; **p<0.05; ***p<0.01

access leads to higher experienced diversity at the individual level, indicating that on average additional 10 acres of parkland increase one's diversity index by 1 p.p., which corresponds to 1% higher chances of encounters with other racial or ethnic groups in a given month.

In line with the suggestive evidence obtained from the cross section model, the estimates for the pooled subsample of black and white residents (column 2) confirm the presence of non-monotonicity in the effects with respect to distance from the user's home location. The point estimate for the parks in the 1-2 km range is statistically significant and equals to 0.9 p.p. This is more than two times larger than the corresponding estimate for parks located closer to home (0.4 p.p), which, in addition, is not significant in the baseline specification.

Furthermore, I find evidence indicating that park accessibility is affecting the diversity of white and black residents differently. As shown in the second column of Table 7 for blacks the impact is significant only for the parks located between 1 and 2 kilometers from home. The relative magnitude of the point estimate for the parks closer to home is also smaller for black residents than for whites. In order to test whether these disparities are significant, I introduce several interaction terms into the main specification that allow me to differentiate the effects for black and white users. The last four rows in Table 11 in the appendix report the estimates for these interaction terms. Overall, it appears as though on average black residents indeed are less exposed to diversity in parks located in their own neighborhood (located less than 1 km away from the residence location). However, the evidence is not entirely conclusive, and as Table 11 demonstrates, when I include month-race fixed effects in the estimated equation, I find that the mentioned difference is only marginally statistically significant.

5.5 Discussion

My empirical approach allows me to establish three key results. First, I document that access to park space has a sizable effect on individual exposure to diversity. To put the estimates into perspective, I predict that an average-sized community park[16] within a 5-km radius from home increases individual chances of encounters with other racial or ethnic groups by 2-5%. Second, I find that parks located in direct proximity to one's residence on average offer less exposure to diversity than parks located slightly further away (within 1 to 2 kilometers). I interpret this finding as indicating that parks designated to serve wider geographic areas, such as community parks and flagship parks, are more successful in fostering racial diversity than the smaller local parks. Lastly, I offer evidence suggesting that the observed non-monotonicity in the estimated effects is partly driven by the parks specifically in close proximity to the residences of black users, implying that parks in the majority-white neighborhoods are able to attract a broader range of visitors compared to the local parks in black neighborhoods.

The employed reduced-form approach, however, does now allow me to identify the extent to which certain park attributes (such as quality, safety, or the variety of nearby businesses) affect the park's role in promoting racially diverse encounters. Hence, I believe that further research should be conducted to incorporate these different characteristics into a more parsimonious discrete choice model. More specifically, estimating the residents' preferences regarding park visits would allow the researchers to evaluate the impact of counterfactual park planning decisions on the interactions between different communities that populate the

[16] Using an average area of 20-40 acres based on the figures I obtained from NYC Open Data.

city, which would be directly valuable to urban planners willing to promote diversity in their city.

6 Conclusion

In this paper, I provide the first empirical evidence to support the claims made by many social scientists, policymakers, and urban planners arguing that the provision of park space allows the city to promote opportunities for racially and ethnically diverse encounters. Hence, the present work emphasizes the role of parks – and, more generally, of public space – in nurturing diverse social environments and suggests a viable policy that can help cities reduce the racial and socioeconomic isolation without resorting to the complicated and costly measures that operate through residential choices. Furthermore, my findings indicate that the positioning of parks with respect to residential communities affects the extent to which they are able to promote racially diverse encounters: parks serving wider geographic areas and located slightly further away from everyone's residences have the highest impact on diversity, while parks located in direct proximity to the black users' homes appear to produce a substantially weaker effect. More broadly this paper builds on the ideas outlined in the seminal works of Jane Jacobs (Jacobs, 1961) and serves as a contribution towards developing a framework for studying the empirical relationships between urban spaces and social phenomena through the lens of high-resolution human-generated data.

References

Anderson, E. (2011): *The Cosmopolitan Canopy: Race and Civility in Everyday Life*, W. W. Norton.

Athey, S., B. Ferguson, M. Gentzkow, and T. Schmidt (2020): "Experienced Segregation," Working Paper 27572, National Bureau of Economic Research.

Bill de Blasio and The City of New York (2015): *OneNYC.*, City of New York, Mayor's Office of Sustainability.

Boone, C. G., G. L. Buckley, J. M. Grove, and C. Sister (2009): "Parks and People: An Environmental Justice Inquiry in Baltimore, Maryland," *Annals of the Association of American Geographers*, 99, 767–787.

Boustan, L. P. (2012): "Racial Residential Segregation in American Cities," in *The Oxford Handbook of Urban Economics and Planning*, Oxford University Press.

Davis, D. R., J. I. Dingel, J. Monras, and E. Morales (2019): "How Segregated Is Urban Consumption?" *Journal of Political Economy*, 127, 1684–1738.

Drakonakis, K., P. Ilia, S. Ioannidis, and J. Polakis (2019): "Please forget where I was last summer: The privacy risks of public location (meta) data," *arXiv preprint arXiv:1901.00897*.

Gobster, P. H. (2002): "Managing urban parks for a racially and ethnically diverse clientele," *Leisure sciences*, 24, 143–159.

Heckert, M. (2013): "Access and Equity in Greenspace Provision: A Comparison of Methods to Assess the Impacts of Greening Vacant Land," *Transactions in GIS*, 17, 808–827.

Jacobs, J. (1961): "The Death and Life of Great American Cities." .

Kinder-Kurlanda, K., K. Weller, W. Zenk-Möltgen, J. Pfeffer, and F. Morstatter (2017): "Archiving information from geotagged tweets to promote reproducibility and comparability in social media research," *Big Data & Society*, 4, 2053951717736336.

KLINENBERG, E. (2018): *Palaces for the People: How To Build a More Equal and United Society*, Random House.

KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75, 83–119.

LANGEGGER, S. (2013): "Viva la Raza! A park, a riot and neighbourhood change in North Denver," *Urban Studies*, 50, 3360–3377.

LOW, S., D. TAPLIN, AND S. SCHELD (2009): *Rethinking urban parks: Public space and cultural diversity*, University of Texas Press.

MASSEY, D. S. (1990): "American Apartheid: Segregation and the Making of the Underclass," *American Journal of Sociology*, 96, 329–357.

OREOPOULOS, P. (2003): "The Long-Run Consequences of Living in a Poor Neighborhood," *The Quarterly Journal of Economics*, 118, 1533–1575.

POPKIN, S. J., J. E. ROSENBAUM, AND P. M. MEADEN (1993): "Labor market experiences of low-income black women in middle-class suburbs: Evidence from a survey of gautreaux program participants," *Journal of Policy Analysis and Management*, 12, 556–573.

QIAN, C., P. KATS, S. MALINCHIK, M. HOFFMAN, B. KETTLER, C. KONTOKOSTA, AND S. SOBOLEVSKY (2017): "Geo-tagged social media data as a proxy for urban mobility," in *International Conference on Applied Human Factors and Ergonomics*, Springer, 29–40.

SAPORITO, S. AND D. CASEY (2015): "Are There Relationships Among Racial Segregation, Economic Isolation, and Proximity to Green Space?" *Human Ecology Review*, 21, 113–132.

VIGDOR, J. L. (2002): "Locations, Outcomes, and Selective Migration," *The Review of Economics and Statistics*, 84, 751–755.

WANG, Q., N. E. PHILLIPS, M. L. SMALL, AND R. J. SAMPSON (2018): "Urban mobility and neighborhood isolation in America's 50 largest cities," *Proceedings of the National Academy of Sciences*, 115, 7735–7740.

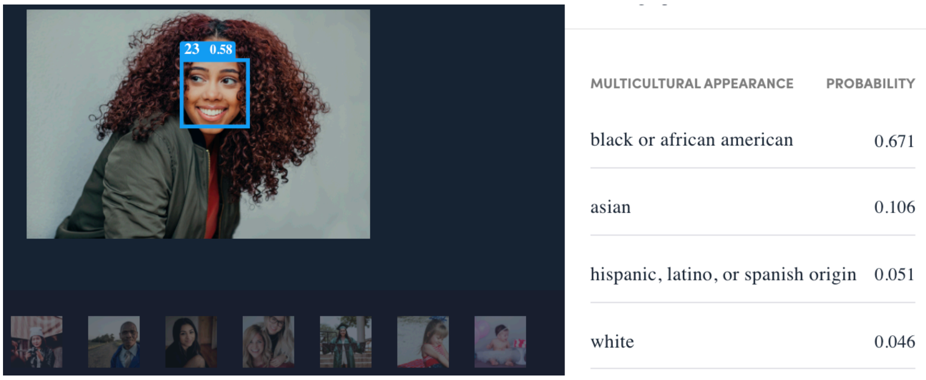
WIRED (2019): "Your Old Tweets Give Away More Location Data Than You Think," Link: https://www.wired.com/story/twitter-location-data-gps-privacy/.

XU, Y., A. BELYI, P. SANTI, AND C. RATTI (2019): "Quantifying segregation in an integrated urban physical-social space," *Journal of The Royal Society Interface*, 16, 20190536.

A Appendix

A.1 Clarifai.com

Below is an example from clarifai.com, demonstrating the predictions of a machine learning model trained to obtain racial or ethnic classification (called *multicultural apperance*) using profile images:



For the purposes of assigning a single race category to each user, I hence select the most likely label from the predictions of the model.

A.2 Capital Project Tracker

Below I include some example notices from www.nycgovparks.org/notices, demonstrating the cases when access to park facilities is either reduced or completely prevented by the ongoing capital projects:

`Riverside Skate Park`:

> `Riverside Skate Park is closed to reconstruct the existing skate park with new skate elements, fencing, benches, picnic tables, and landscaping. Please visit the Capital Project Tracker page for updates on the project.`

`ParkBronx River Parkway`:

> `Shoelace Park is currently under construction. During construction, some park entrances and paths may be temporarily closed. For more information about the project, please visit the Shoelace Park Reconstruction project page.`

`O'Neill Triangle`:

> `Plaza and seating area temporary closed for NYCDOT bridge removal work.`

Old Hickory Playground:

The playground will be closed until further notice for the repair and replacement of safety surfacing.

Van Cortlandt Park:

Due to construction at Broadway and West 242nd Street, the entrance near the subway is closed. Please access the park at the stairway adjacent to the public comfort station (Broadway & Manhattan College Parkway). There is an accessible entrance at the end of the block. The barbecuing area at Van Cortlandt Park has temporarily been moved behind the Nature Center. For more information on the progress of this construction project, please visit our Capital Project Tracker page.

Squibb Park:

Squibb Park will be temporarily closed during the removal and construction of the replacement of Squibb Bridge. We are closing Squibb Park out of an abundance of caution and apologize for any inconvenience. For updates on this project, please visit Brooklyn Bridge Park's Squibb Bridge page.

A.3 Figures and Tables

Table 8: Sample Description: Twitter sample compared to census tracts in New York

	Twitter sample	NY metro tracts	NYC tracts	Manhattan tracts
Median income (thousand $)	77.0	67.9	54.6	81.9
Share with BA degree or higher	0.383	0.323	0.278	0.699
Share owner-occupied units	0.520	0.586	0.330	0.214
Share hh. with 2 or more bedrooms	0.581	0.631	0.590	0.408

Note: Census aggregates are obtained from the American Community Survey 5-year estimates for 2011-2015.

Table 9: Park Access Differences By Race/Ethnicity

(a) Park Acreage Accessible within 0-1km and 0-5km Radius from Home

Race/Ethnicity	Parks Area (0,1km]			Parks Area (0,5km]		
	Mean	Med.	St.dev	Mean	Med.	St.dev
Asian	253.84	40.67	389.76	1818.98	1566.69	979.03
Black	241.62	44.76	398.92	1979.37	1618.94	1191.66
Hispanic	266.02	42.95	417.54	1929.17	1592.95	1089.13
White	267.41	43.54	413.07	1595.33	1502.18	795.42

(b) Difference in Means Test:
Park Acreage Accessible Within 0-1km Radius

	Asian	Black	Hispanic	White
p-value for H_0 $\mu_{Row} - \mu_{Col} < 0$				
Asian	0.500	0.940	0.120	0.061*
Black	0.060	0.500	0.001***	0.000***
Hispanic	0.880	0.999	0.500	0.379
White	0.939	1.000	0.621	0.500

Table 10: Park Access Differences By Race/Ethnicity

Race /Ethnicity	Category	Parks Area (0,1km] Mean	Med.	St.dev
Asian	Community Park	53.89	5.40	119.87
	Flagship Park	202.47	0.00	384.73
	Neighborhood Park	11.13	6.76	12.59
	Playground	1.97	1.13	2.30
Black	Community Park	58.34	6.73	118.38
	Flagship Park	183.11	0.00	401.47
	Neighborhood Park	10.77	5.10	13.48
	Playground	1.99	1.08	2.30
Hispanic	Community Park	59.73	4.93	130.99
	Flagship Park	210.14	0.00	421.07
	Neighborhood Park	11.06	6.20	14.56
	Playground	1.99	1.13	2.35
White	Community Park	51.52	6.51	112.95
	Flagship Park	225.03	0.00	406.39
	Neighborhood Park	11.53	8.03	15.04
	Playground	1.92	1.08	2.27

Figure 5: Distance of Travel to Parks by Neighborhood Composition

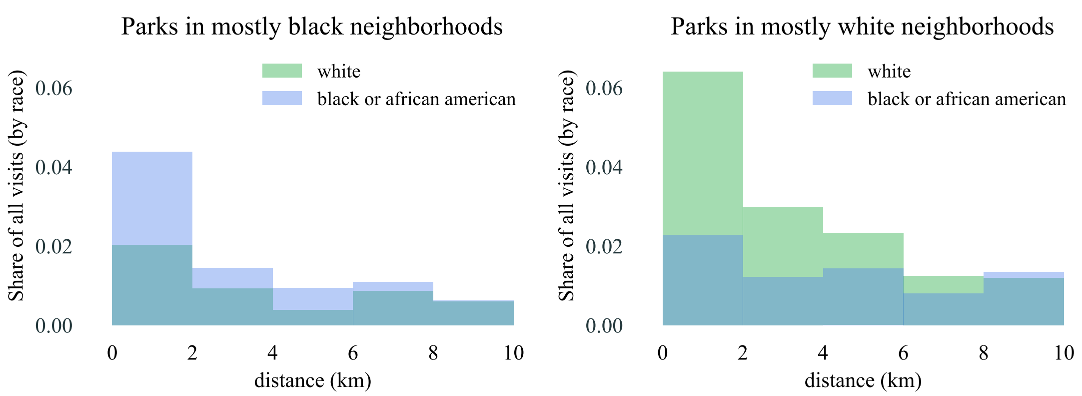


Table 11: Testing Differences in the Effects for Black and White Residents

Accesible Park Area (arces)	*$ExperiencedDiversity_{it}$*			
	Black or White		Black or White	
	(1)	(2)	(3)	(4)
Total $_{(0\text{-}5km]}$	.0020***		.0014***	
	(.0004)		(.0004)	
(0-1km)		.0020**		.0016*
		(.0009)		(.0009)
[1-2km)		.0013***		.0010***
		(.0003)		(.0003)
[2km+)		.0006***		.0003*
		(.0002)		(.0002)
Total × Black	−.0033***		.0011	
	(.0008)		(.0008)	
(0-1km) × Black		−.0042***		−.0030*
		(.0015)		(.0015)
[1-2km)× Black		−.0010		.0002
		(.0007)		(.0007)
[2+km) × Black		−.0012***		−.0005
		(.0003)		(.0003)
Observations	57,431	57,431	57,431	57,431
User FE	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
Race x month FE	X	X	✓	✓