If you build it who will come? Equity analysis of park system changes using passive origin-destination data

Gregory Macfarlane[a,*], Teresa Tapia[c], Carole Turley-Voulgaris[b]

[a] *Brigham Young University, Civil and Environmental Engineering Department, 430 Engineering Building, Provo, Utah 84602*
[b] *Harvard Graduate School of Design, 48 Quincy St, Cambridge, Massachussetts 02138*
[c] *StreetLight Data, Inc., San Francisco, California*

Abstract

Parks provide important benefits to those who live near them, in the form of improved property values, health outcomes, etc.; nevertheless, measuring and understanding who lives near a park is an open research question. In particular, it is not well understood which park individuals will choose to use when given a choice among a set of nearby parks of varying sizes and at varying distances from their home. In this paper we present a park activity location choice model estimated from a passive origin-destination dataset — supplied by StreetLight Data, Inc. — representing trips to parks and green spaces in Alameda County, California. The estimated model parameters reveal heterogeneous preferences for park size and willingness-to-travel across block-group level socioeconomic segmentation: Specifically, high-income block groups appear more positively attracted to larger parks, and block groups with a high proportion of ethnic minority individuals are more likely to select nearby parks. The findings have importance for understanding recreational access among different populations, and the methodology more generally supplies a potential template for using passive data products within travel modeling.

Keywords: Accessibility Passive Data Location Choice

1. Introduction

Parks and other green spaces generate immense value for the public who are able to access them. The City Parks Alliance (2019) categorizes the observed benefits of urban parks as encouraging active lifestyles (Bancroft et al., 2015), contributing to local economies, aiding in stormwater management and flood mitigation, improving local air quality, increasing community engagement (Madzia et al., 2018), and enhancing public equity.

Nevertheless, understanding and quantifying these benefits depends in many cases on identifying who lives near the parks and is therefore able to access them. Many previous studies (e.g., Richardson et al., 2012) rely on comparison of total greenspace across metropolitan areas; this methodology may not adequately control for city-level fixed effects and it may ignore the potentially inequitable distribution of park space within a region. Studies focusing on access within metropolitan areas typically assume that people living within a certain distance or travel time threshold have access to a park, or examine the quantity of park space within one's own arbitrarily defined "neighborhood" (Mitchell and Popham, 2008, Stark2014). But these methods do not account for the fact that some people will travel to other parks to perform recreational activities. A more holistic measure that continuously measures access across multiple preference dimensions is desirable.

*Corresponding Author
Email addresses: gregmacfarlane@byu.edu (Gregory Macfarlane), teresa.tapia@streetlightdata.com (Teresa Tapia), cvoulgaris@gsd.harvard.edu (Carole Turley-Voulgaris)

An appealing solution would be to examine and model the activity location choices of park users. Such a model would help researchers understand how individuals of different backgrounds and preferences value different park amenities. Further, the logsums of a location choice model provide a continuous measure of accessibility that explicitly accounts for such variation (de Jong et al., 2007). Unfortunately, park choice models of this form are rare in the literature. Travel demand models built for infrastructure forecasting are a common way to generate such accessibility logsums, but these models group many different kinds of social and recreational trips together (National Academies of Sciences Engineering and Medicine, 2012). Further, the attraction term for such trip purposes is commonly a function of the retail or service employment or the number of households at the destination; a typical park or green space has neither employees nor residents. Finally, many regional household travel surveys are oriented towards an average weekday travel pattern, and many park trips occur irregularly or on weekends.

In this paper we present a park destination choice model where individuals living in Alameda County, California choose among parks in the same county. The individuals are constructed from passive data that was derived from mobile devices and processed using algorithms developed by StreetLight Data, Inc. The origin location points are inferred residence block groups for unique devices and the destination points are geofenced polygons representing green and open spaces. The individuals' choice of park location is conditioned on the distance from the block group to the parks in the choice set as well as the size of each park; market segmentation allows for heterogeneous responses between ethnic groups and income strata.

The paper proceeds in the following manner: A discussion of prior attempts to study park choice and employ passive origin-destination data in the literature is given directly. The Methodology section presents the data gathering and cleaning efforts as well as the econometric location choice model. The Results section presents the estimated model coefficients and a discussion of the findings, as well as a model validation exercise. After presenting limitations and associated avenues for future research, a final Conclusions section outlines the contributions of this study for recreational trip modeling and location choice modeling more generally.

2. Literature

2.1. Defining and measuring park accessibility

Handy and Niemeier (1997) identify three broad types of accessibility measures: cumulative opportunity measures, gravity-based measures, and utility-based measures. Geurs and van Wee (2004) classify cumulative opportunity and gravity-based measures into a single category that they refer to as location-based measures. Researchers have applied versions of location-based measures and, to a lesser degree, utility-based measures to quantify park accessibility

2.1.1. Location-based measures of park accessibility

Cumulative opportunity measures are calculated by counting the number of origins or destinations within a threshold travel cost of a location (where cost might be some combination of distance, travel time, and/or monetary cost of travel). A strength of cumulative opportunity measures lies in their simplicity and intuitive interpretation. However, they may be too simple, especially with regard to trip costs near the threshold. An example of a cumulative opportunity measure might be the number of parks within a ten-minute walk of a person's home, or the number of households living within ten minutes of a park. This measure would imply that a household living immediately adjacent to a park has the same access to it as one that lives nine minutes away, but that a household living eleven minutes away has no access to it.

ParkScore (Trust for Public Land, 2019), developed by the Trust for Public Land, is a popular measure of park accessibility that starts from a cumulative opportunity measure (the share of the population that resides within a 10-minute walk of a green space) and adjusts this value based on the total city green space, investment, and amenities weighted against the socioeconomic characteristics of the population outside of the 10-minute walk threshold. The resulting score is a convenient quantitative tool in estimating the relative quality of green space access across cities (Rigolon et al., 2018). It may be less useful at identifying the comparative quality of access within a city, particularly since more than 95% of residents in many dense metropolitan areas like San Francisco and New York may live within the binary 10-minute walk threshold [citation needed]. The Centers for Disease Control and Prevention (CDC) has developed an "Accessibility to Parks Indicator" along similar lines (Ussery et al., 2016), calculating the share of the population living within a half-mile of a park for each county in the U.S.

Gravity-based accessibility measures take a similar approach to cumulative opportunity measures, but theoretically include all possible destinations and weight them according to the travel cost that they impose, based on an impedance function. Cumulative opportunity measures may be considered a special case of gravity-based measures, where the impedance function takes the form of a binary step function that equals zero after a cutoff travel cost (which is why Geurs and van Wee (2004) classify them both as location-based). The inverse square function is a common form of impedance function for gravity-based accessibility measures.

A major advantage of gravity-based accessibility measures lies in their consistency with travel behavior theory: Gravity-based measures have their roots in the trip-distribution step of the traditional four-step travel demand forecasting method, where trips originating in a particular zone are distributed among destination zones, proportionate to each zone's gravity-based accessibility. Urban scholars have used gravity-based measures to explore the spatial distribution of park access across Tainan City, Taiwan (Chang and Liao, 2011) and to estimate the relationship between park access and housing prices in Shenzhen, China (Wu et al., 2017).

Some scholars have used location-based measures of park accessibility to evaluate equity in park access. Chang and Liao (2011) use a gravity-based measure to determine that low-income neighborhoods have less access to parks than higher-income neighborhoods in Tainan City, Taiwan. Bruton and Floyd (2014) conduct a neighborhood-level analysis of park amenities in Greensboro, North Carolina, and find that low-income neighborhoods tend to have parks with more picnic areas, more trash cans, and fewer wooded areas, but they do not address the question of the extent to which different populations might value these different amenities. Kabisch and Haase (2014) find that neighborhoods in Berlin with high immigrant populations and older populations likewise had less access to parks, and they pair these findings with survey results suggesting that these disparities are not consistent with the preferences expressed by those populations.

The question of whether uneven distribution of parks and park amenities might reflect local preferences could be more appropriately addressed using utility-based measures of park accessibility (as discussed below) and though qualitative studies (as discussed in the following section).

2.1.2. Utility-based measures of park accessibility

While traditional four-step travel demand models distribute zonal trips based on a gravity-based accessibility model, the travel demand modeling profession has shifted more recently towards a destination choice framework that distributes trips based on discrete-choice regression models. McFadden (1974a) first applied discrete choice models to urban travel demand to predict mode choice, and activity-based models apply them to all travel behavior choices, including to select among alternative routes or alternative destinations (de Dios Ortúzar and Willumsen, 2011). Though the application of random utility models to destination choice is not new (see Anas, 1983), the increasing availability of computing resources makes estimating and applying discrete choice models on large alternative sets in a practical context more feasible.

When applied to destination choice, discrete choice models estimate the likelihood of selecting a particular destination among a set of alternatives based on the relative attractiveness, or *utility*, of each alternative. Utility may be function of distance or travel time alone (in which case, a utility-based accessibility measure might be quite similar to a location-based measure), but it can also incorporate other destination characteristics that lead one destination to be more highly-utilized than another. For a utility-based measure of park accessibility, these might include park size or park amenities.

Destination-choice models have not commonly been used to measure park accessibility, scholars have acknowledged that park accessibility metrics should be linked with park use, since a park that has many visitors must by definition be accessible to those visitors. McCormack et al. (2010) provide a comprehensive review of this literature; it is sufficient here to note that most studies find park use to depend on a complicated interplay between park size, maintenance, facilities, and travel distance. Many of these attributes are incorporated into ParkIndex (Kaczynski et al., 2016), which estimates the resident park use potential within $100m^2$ grid cells, based on a household park use survey in Kansas City.

The park use potential measured by ParkIndex is not dissimilar from a park trip production potential, as used by transportation planners and engineers. Along these lines, the question of park use is a destination choice problem, where trip makers consider which park is most attractive to accomplish their recreation activity. The Institute of Transportation Engineers (ITE) Trip Generation Manual (2017) contains trip attraction rates for public parks that use the park acreage, number of picnic tables, employees, and other variables as attraction terms. As with many land uses in Trip Generation, the provided trip generation rates are based on a limited number of observational samples (Shoup, 2003) and may not represent large-sample behavior (Millard-Ball, 2015). Moreover, regression-based attraction rates isolated from trip production and travel behavior ignore the geographical and behavioral contexts in which people actually make trips to parks (Barnard and Brindle, 1987): Though more people may come to larger parks, planners cannot bring more people to a park simply by increasing its size.

There are limited examples of researchers using a destination choice model to predict recreation attractions. Kinnell et al. (2006) apply a choice model to a survey of park visitors in New Jersey, and estimate the relative attractiveness of park attributes including playgrounds, picnic areas, and park acreage weighed against the travel disutility and the relative crime rate at the destination. In a similar study, Meyerhoff et al. (2010) model the urban swimming location choice for a surveyed sample. In both studies, the researchers were attempting to ascertain which attributes of a recreation generated the most positive utility, and therefore which attributes should be prioritized for improvement. These studies have not to our knowledge been previously referenced in discussions of park accessibility.

One obstacle to estimating discrete-choice models to estimate park choice has been the lack of sufficiently detailed, trip-level data. The advent of large-scale mobile networks and the seemingly perpetual association of unique devices with unique users has given researchers a new opportunity to observe the movements and activity location patterns for large subsets of the population (Naboulsi et al., 2016). Such passively collected movement data — sometimes referred to as “Big Data” — is a by-product of other systems including cellular call data records (e.g., Bolla and Davoli, 2000; Calabrese et al., 2011), probe GPS data (Huang and Levinson, 2015), and more recently Location Based Services (LBS) (Roll, 2019; Komanduri et al., 2017). LBS use a network of mobile applications that obtain the users’ physical location. A variety of commercial vendors repackage, clean, and scale these data to population or traffic targets and sell origin-destination matrices to researchers and practitioners at relatively low prices. Monz et al. (2019), for example, demonstrate how passive device data can be used to accurately estimate trips to natural recreation areas.

Passive origin-destination matrices are beginning to inform trip distribution model development more directly as well. Kressner (2017) proposes one methodology, where passive origin-destination matrices serve as a probabilistic sampling frame for a simulated trip destination choice. Bernardin et al. (2018) employ a passive origin-destination matrix as a shadow price reference in an activity-based location choice model, iteratively adjusting the parameters of the choice utilities to minimize the observed error between the matrix and the modeled predictions. A similar method developed by Zhu and Ye (2018) uses the passive dataset directly, sampling 10,000 random trips from GPS traces of taxi trips in Shanghai and estimating a destination choice model. Employing the passive data set in this way provides the authors an opportunity to examine the choices of a large sample of a small population (taxi passengers) as well as sufficient data to estimate a “constants-rich” destination choice model with uniquely estimated coefficients for each origin-destination pair. The Zhu and Ye (2018) methodology suggests that a similar approach should apply in other contexts, including park choice. Applying these models to park choice allows us to not only predict potential park use, but also determine which park characteristics are most valuable to particular populations. The results of such an analysis offer an important complement to existing qualitative and survey-based research on park use by marginalized populations.

2.2. Sociodemographic variation in park utility

The idea that different racial, ethnic, or cultural groups have different recreational styles, and might thus have different needs and preferences for parks and open space, has been thoroughly discussed in the leisure studies literature, and Husbands and Idahosa (1995) offer a detailed review of that research as of the mid-1990s. In general, explanations for racial and ethnic differences in park use can be classified into two categories: those rooted in cultural and lifestyle differences, and those rooted in discrimination and marginalization.

Byrne and Wolch (2009) summarize literature in the former category, noting that African Americans have been described as preferring more social, sports-oriented spaces, relative to white people who prefer secluded natural settings (Washburne, 1978; Hutchison, 1987; Floyd and Shinew, 1999; Gobster, 2002; Payne et al., 2002; Ho et al., 2005); Asians are described as valuing aesthetics over recreational spaces; (Gobster, 2002; Payne et al., 2002; Ho et al., 2005), and Latinos are said to value group-oriented amenities like picnic tables and restrooms (Baas et al., 1993; Hutchison, 1987; Irwin et al., 1990).

Byrne and Wolch (2009) criticize such scholarship as having grossly exaggerated ethno-racial differences in park use and preferences, and suggest a model for explaining park use based on four elements: Sociodemographic characteristics; park amenities and surrounding land uses; historical/cultural context of park provision (including development politics and discriminatory land-use policies); and individual perceptions of park space (including safety and sense of welcome).

Byrne (2012) applies a cultural politics theoretical frame to why people of color are underrepresented among visitors to some urban parks. Focus groups of Latino residents emphasized the importance of parks to children and childhood. Participants described visiting parks with their children and the positive and negatives associations that parks evoked of their own childhood memories of parks and wilderness. Participants described barriers to visiting parks including distance, inadequate or poorly maintained facilities, and fear of crime. They cited a lack of Spanish-language signage not only as a barrier to understanding but also as a signal that a park was not intended to serve Spanish speakers. Participants also expressed that they did not feel welcome in parks located in high-income or predominantly white neighborhoods, either because they expected that other park users would have racist attitudes, or because they expected that a more boisterous Latino 'recreational style' would not be tolerated or that there would be other behavioral norms they were not aware of.

In an observational and survey-based study of park users in Los Angeles, Loukaitou-Sideris (1995) found a high-level of enthusiasm for park use among Hispanic people. While she found, consistent with prior research (Baas et al., 1993; Hutchison, 1987; Irwin et al., 1990), that Hispanic park users showed a preference for passive recreation, she found that to be the case for all other user groups as well. She also found that Hispanic park users were the most likely to actively appropriate and modify park space, for example, by bringing items from home. She found that Hispanic park users tended to visit parks as family groups; African American park users tended to visit parks as peer groups; Caucasian park users tended to visit parks alone; and Asian residents were least likely to visit parks, even in a predominantly Asian neighborhood. Interviews with local elderly Asian residents (Chinese immigrants) suggested that a lack of interest in American parks was rooted in perceptions of the ideal park as "an aesthetic element of gorgeous design," leaving them unimpressed with poorly landscaped American parks emphasizing recreational functions.

2.3. Changing park systems in response to a global pandemic

[placeholder for literature on COVID shifting streets - will draw heavily from Tab Combs' working paper]

The definition of an urban park is not well-established, though some might echo Justice Potter Stewart in arguing that we know one when we see one (196, 1964). If we define an urban park as a public space that is designated for the purpose of recreation, exercise, and social gathering, then the rapid reallocation of street space that occurred in response to the global COVID-19 pandemic in 2019 and 2020 could be characterized as a proliferation of small urban parks. During that same period, many municipalities closed specific park amenities including playgrounds and restrooms.

How might these combined actions have changed the distribution of park utility, benefits, or access among different populations of park users?

3. Methodology

We constructed a dataset on which to estimate park activity location choices for a sample of observed trips in Alameda County, California. Alameda County is one of the seven counties that constitutes the San

Francisco Bay Area metropolitan region in California. Alameda is the seventh most populous county in California with a population of 1.5 million (U.S. Census Bureau, 2019), and has 14 incorporated cities and several unincorporated communities. It is an economically and ethnically diverse county and hence it was an attractive area to use for this study. The racial makeup of Alameda County was (49.7%) White, (11.2%) African American, (1.0%) Native American, (38.7%) Asian, (1.0%) Pacific Islander, and (22.4% Hispanic or Latino (of any race). Alameda County has a diverse set of parks, ranging from local small community parks, urban and transit-accessible parks like the Lake Merritt Recreational area, accessible coastal access, and suburban recreational areas like Lake Chabot.

3.1. Data

We constructed an analysis dataset from a publicly-available parks polygons layer, a commercially acquired passive device origin-destination table representing trips between the parks and home block groups, and American Community Survey data for the home block groups.

We obtained a polygons shapefile layer representing open spaces in Alameda County, California from the California Protected Areas Database (GreenInfo Network, 2019). This dataset was selected because it included multiple different types of open space including local and state parks, traditional green spaces as well as wildlife refuges and other facilities that can be used for recreation. We removed facilities that did not allow open access to the public (such as the Oakland Zoo) and facilities whose boundaries conflated with freeway right-of-way – this prevents trips through the park from being conflated with park trips in the passive origin-destination data.

From this initial parks polygons dataset, we obtained park attribute information through OpenStreetMap (OSM) using the `osmdata` package for R (Padgham et al., 2017). Three attribute elements are considered in this analysis. First, we identify playgrounds using OSM facilities given a `leisure = playground` tag. The tagged facilities can be either polygon or point objects; in the former case we use the polygon centroid to determine the location of the playground.

Second, we consider sport fields of various kinds identified with the OSM `leisure = pitch` tag. This tag has an additional attribute describing the sport the field is designed for, which we retain in a consolidated manner. Soccer and American football fields are considered in a single category, and baseball and softball fields are combined as well. Basketball, tennis, and volleyball courts are kept as distinct categories, with all other sport-specific fields combined into a single "other." Golf courses are discarded. As with playgrounds, polygon field and court objects are converted into points at the polygon centroid.

Finally, we identified trails and footpaths using the `path`, `cycleway`, and `footway` values of the `highway` tag. A visual inspection of the resulting data revealed that the large preponderance of sidewalks and cycling trails within parks in Alameda County are identified properly with these variables. Trails are represented in OSM as polylines, or as polygons if they form a complete circle. In the latter case, we converted the polygon boundary into an explicit polyline object.

It is possible for each of these facilities to exist outside the context of a public park. For example, many private apartment complexes have playgrounds and high schools will have sports facilities that are not necessarily open to the general public. We spatially matched the OSM amenities data and retained only those facilities that intersected with the CPAD open spaces database identified earlier.

We provided the park boundaries layer to a commercial firm, StreetLight Data Inc., which develops and resells origin-destination matrices derived from passive device location data. The provider employs a proprietary data processing engine (called Route Science) to algorithmically transform observed device location data points (the provider uses in-vehicle GPS units and mobile device LBS) over time into contextualized, normalized, and aggregated travel patterns. From these travel patterns, the Route Science processing algorithms infer likely home Census block group locations for composite groups of people and converts raw location data points into trip origin and destination points (Pan et al., 2006; Friedrich et al., 2010).

For each park polygon, the firm returned a population-weighted estimate of how many devices were observed by home location block group over several months in the period between May 2018 and October 2018. We transformed this table such that it represented the weighted unique devices traveling between block groups and parks. We discarded home location block groups outside of Alameda County; the imputed

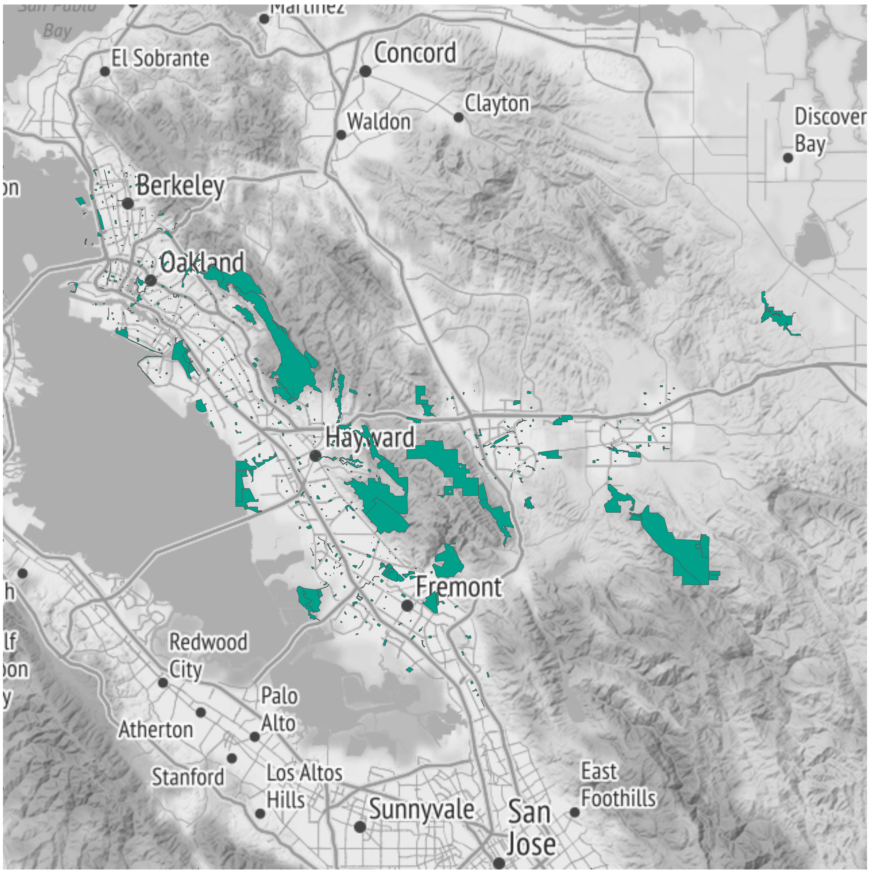

Figure 1: Location of parks in Alameda County.

Table 1: Park Summary Statistics

		Local Park (N=441)		Recreation Area (N=59)	
		Mean	Std. Dev.	Mean	Std. Dev.
Acres		59.8	370.5	125.6	505.1
Mobile Devices		1450.0	6685.4	2659.0	6161.0
		N	%	N	%
type	Local Park	441	88.2	0	0.0
	Local Recreation Area	0	0.0	57	11.4
	State Recreation Area	0	0.0	2	0.4
Access	Open Access	441	88.2	59	11.8
	No Public Access	0	0.0	0	0.0
	Restricted Access	0	0.0	0	0.0
Playground	FALSE	224	44.8	44	8.8
	TRUE	217	43.4	15	3.0
Any Sport Field	FALSE	276	55.2	39	7.8
	TRUE	165	33.0	20	4.0
Football / Soccer	FALSE	415	83.0	51	10.2
	TRUE	26	5.2	8	1.6
Baseball	FALSE	364	72.8	45	9.0
	TRUE	77	15.4	14	2.8
Basketball	FALSE	341	68.2	52	10.4
	TRUE	100	20.0	7	1.4
Tennis	FALSE	387	77.4	53	10.6
	TRUE	54	10.8	6	1.2
Volleyball	FALSE	434	86.8	57	11.4
	TRUE	7	1.4	2	0.4
Trail	FALSE	155	31.0	22	4.4
	TRUE	286	57.2	37	7.4

home locations can be far away from the study area for a small amount of trips and are unlikely to represent common or repeated park activities.

Table 1 presents descriptive statistics on the 500 parks assembled for this study, grouped according to the park type as defined on CPAD. A little more than half of the parks have identified paths, while around 40% of the identified parks have playgrounds and sport fields.

In order to understand the demographic makeup of the home block groups and potentially the characteristics of the people who make each trip, we obtained 2013-2017 five-year data aggregations from the American Community Survey using the `tidycensus` (Walker, 2019) interface to the Census API for several key demographic and built environment variables: the share of individuals by race, the share of households by income level, household median income, the share of households with children under 6 years old, and the household density. An important attribute of the choice model is the distance from the home block group to the park boundary. Because we have no information on where in the block group a home is actually located, we use the population-weighted block group centroid published by the Census Bureau as the location for all homes in the block group. We then measured the network-based distance between the park and the home block group centroid using a walk network derived from OpenStreetMap using the `networkx` package for Python (Hagberg et al., 2008),

Table 2: Block Group Summary Statistics

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
Density: Households per square kilometer	1041	0	1709.6	1527.8	0.0	1350.1	19490.0
Income: Median tract income	971	3	93246.3	44900.3	9821.0	85673.0	250001.0
Low Income: Share of households making less than $35k	980	0	18.4	14.0	0.0	15.1	91.7
High Income: Share of households making more than $125k	1005	0	32.5	20.3	0.0	30.4	100.0
Children: Share of households with children under 6	985	0	15.1	9.0	0.0	14.0	62.7
Black: Share of population who is Black	926	0	11.8	14.1	0.0	6.4	81.4
Asian: Share of population who is Asian	1012	0	25.8	20.4	0.0	20.1	90.4
Hispanic: Share of population who is Hispanic	1020	0	22.3	18.9	0.0	15.7	85.4
White: Share of population who is White	1033	0	34.2	23.0	0.0	29.1	100.0

[a] Hispanic indicates Hispanic individuals of all races; non-Hispanic individuals report a single race alone.

3.2. Model

In random utility choice theory, if an individual living in block group n wishes to make a park trip, the probability that the individual will choose park i from the set of all parks J can be described as a ratio of the park's measurable utility V_{ni} to the sum of the utilities for all parks in the set. In the common destination choice framework we apply a multinomial logit model (McFadden, 1974b, Recker and Kostyniuk (1978)),

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_{j \in J} \exp(V_{nj})} \tag{1}$$

where the measurable utility V_{ni} is a linear-in-parameters function of the destination attributes.

$$V_{ni} = X\beta \tag{2}$$

where β is a vector of estimable coefficients giving the relative utility (or disutility) of that attribute to the choice maker, all else equal. It is possible to add amenities of the park or the journey to the utility equation. However, as the number of alternatives is large, it is impractical to consider alternative-specific constants or coefficients and therefore not possible to include attributes of the home block group or traveler n directly. We can, however, segment the data and estimate different distance and size parameters for different segments to observe heterogeneity in the utility parameters between different socioeconomic groups.

The logarithm of the sum in the denominator of Equation 1 (called the logsum) provides a measure of the consumer surplus of the choice set (Williams, 1977),

$$CS_n = \ln \sum_{j \in J} \exp(V_{nj}) + C \tag{3}$$

where C is a constant indicating an unknown absolute value. But comparing the relative logsum values across choice makers, $CS_n - CS_{n-1}$ gives an indication of which choice maker has a more valuable choice set. Or, in this case of a park destination choice model, which choice maker has better access to parks. Such a "utility-based" accessibility term is thus continuously defined, dervied directly from choice theory, and can contain multiple dimensions of the attributes of the choice (Handy and Niemeier, 1997; Dong et al., 2006).

In the most typical cases, researchers estimate the utility coefficients for destination choice models from household travel surveys. As we have no knowledge of an appropriate survey on park access, we need to synthesize a suitable estimation data set. We do this by sampling 2×10^4 random discrete device origin-destination pairs from the commercial passive data matrix, weighted by the volume of the flows. This corresponds to a 4.3% sample of all the observed device origin-destination pairs.

The sampled origin-destination pair gives the home location as well as the "chosen" alternative for a synthetic person. In principle the individual's choice set contains all the parks in our dataset; in practice it can be difficult to estimate choice models with so many alternatives ($|J| = 500$). For this reason we randomly sample 10 additional parks to serve as the non-chosen alternatives for our synthetic choice maker. Such random sampling of alternatives reduces the efficiency of the estimated coefficients but the coefficients remain unbiased (Train, 2009). As the model has no alternative-specific constants, the standard likelihood comparison statistic against the market shares model ρ^2 is not computable. We instead use the likelihood comparison against the equal shares model ρ_0^2.

The resulting analysis dataset therefore contains 2×10^4 choice makers that select between 11 parks including the park they were observed to choose; the measured distance between the choice maker's block group and all parks in the choice set; and the acreage of each park in the choice set. We hold out a random sample of approximately 20% of choice makers for validation purposes. We use the `mlogit` package for R (Croissant, 2019; R Core Team, 2020) to estimate the multinomial logit models.

4. Results

We estimated multinomial logit park activity location choice models including coefficients for the distance between the park and the home block group and the acreage of the park. We applied a Yeo-Johnson transformation (Yeo and Johnson, 2000) to both distance and acreage; the Yeo-Johnson transformation replicates the constant marginal elasticity of a logarithmic transformation while avoiding undefined values ($YJ(0) = 0$). For simplicity, we call this transformation `log()` in the model results tables. Using a constant marginal elasticity is better reflective of how people perceive distances and sizes; a one-mile increase to a trip distance is more impactful to a one-mile trip than a ten-mile trip.

Table 3 presents the model estimation results for a series of models with different utility function definitions, each estimated on the complete set of synthetic choice makers. The first model — named "Network Distance" — only considers the distance to the park and the size of the park in the utility equation. The estimated coefficients are significant and of the expected sign: That is, individuals will travel further distances to reach larger parks. The ratio of the estimated coefficients implies that on average, people will travel 3.3089752 times further to reach a park twice as large.

The second and third models in Table 3 include a vector of park attributes. In the model labeled "Park Attributes," the presence of any sport field is considered with a single dummy value, and in the "Sport Detail" model this variable is disaggregated into facilities for different sports. The value of the size and distance coefficients change modestly from the "Network Distance" model, with the implied size to distance trade-off rising to 3.3089752. Examining the two amenities models — independently and in comparison with each other — reveals a few surprising findings. First, it appears that playgrounds and sport fields in general contribute *negatively* to the choice utility equation. This is both unintuitive and contradictory to previous findings in this space (e.g., Kinnell et al., 2006). Considering different sports separately, there is a wide variety of observed response with tennis and volleyball facilities attracting more trips, and football and basketball facilities attracting fewer, all else equal. Trails and walking paths give substantive positive utility in both models. The difference in likelihood statistics between the three models is significant (likelihood ratio test p-value 3.5989633×10^{-4}), and so in spite of the curious aggregate findings, we move forward with this utility specification.

It is worth investigating the heterogeneity in preferences that exist among different demographic groups. Though the income and ethnicity of the synthetic park visitors is not known, we can segment the estimation dataset based on the socioeconomic makeup of the visitors' residence block group. The models presented in Table 4 were estimated on segments developed in this manner. Models under the "Minority" heading include a race-based segmentation: simulated individuals living in block groups with more than thirty percent Black persons are included in the ">30% Black" model, an analogous segmentation for block groups with high Asian populations are in the ">30% Asian" model, and the "Other" model contains all other block groups. Another set of model segmentation relies on the share of the population in each block group with household incomes above or below certain thresholds. Again, we use the threshold definitions based on in 2.

Table 3: Estimated Model Coefficients

	Network Distance	Park Attributes	Sport Detail
log(Distance)	-1.354***	-1.394***	-1.390***
	(0.022)	(0.022)	(0.022)
log(Acres)	0.409***	0.353***	0.348***
	(0.011)	(0.012)	(0.012)
Playground		-0.437***	-0.551***
		(0.049)	(0.049)
Trail		0.555***	0.563***
		(0.052)	(0.053)
Sport Field		-0.324***	
		(0.050)	
Basketball			-0.237***
			(0.068)
Baseball			0.075
			(0.067)
Football / Soccer			-0.490***
			(0.095)
Tennis			0.231***
			(0.066)
Volleyball			0.607***
			(0.135)
Other Sport			-0.150*
			(0.087)
Num.Obs.	3971	3971	3971
AIC	11600.2	11306.5	11293.6
Log.Lik.	-5798.103	-5648.236	-5636.809
rho2			
rho20	0.391	0.407	0.408

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Estimated Model Coefficients with Block Group Segmentations

	Minority				Income			Children		
	> 30% Asian	> 30% Black	> 30% Hispanic	Other	> 30% Low income	> 50% High income	Other	> 25% children	< 5% children	Other
log(Distance)	-1.325***	-1.524***	-1.277***	-1.428***	-1.419***	-1.309***	-1.394***	-1.236***	-1.594***	-1.398***
	(0.038)	(0.064)	(0.050)	(0.040)	(0.051)	(0.051)	(0.029)	(0.059)	(0.083)	(0.026)
log(Acres)	0.372***	0.311***	0.337***	0.351***	0.338***	0.342***	0.354***	0.313***	0.372***	0.356***
	(0.020)	(0.034)	(0.026)	(0.022)	(0.027)	(0.028)	(0.015)	(0.030)	(0.044)	(0.014)
Playground	-0.516***	-0.389***	-0.308***	-0.880***	-0.515***	-0.499***	-0.585***	-0.431***	-0.879***	-0.540***
	(0.085)	(0.124)	(0.106)	(0.094)	(0.104)	(0.123)	(0.063)	(0.124)	(0.180)	(0.057)
Trail	0.655***	0.361***	0.269**	0.939***	0.306***	0.896***	0.612***	0.122	0.744***	0.638***
	(0.095)	(0.130)	(0.110)	(0.106)	(0.109)	(0.148)	(0.068)	(0.126)	(0.190)	(0.062)
Basketball	-0.009	-0.305	-0.458***	-0.413***	-0.187	-0.109	-0.300***	-0.316*	-0.301	-0.215***
	(0.110)	(0.187)	(0.153)	(0.136)	(0.153)	(0.160)	(0.088)	(0.172)	(0.262)	(0.078)
Baseball	0.126	0.151	0.090	-0.066	0.054	-0.108	0.139	0.127	-0.038	0.085
	(0.112)	(0.176)	(0.144)	(0.131)	(0.145)	(0.165)	(0.085)	(0.168)	(0.256)	(0.076)
Football / Soccer	-0.495***	-1.014***	-0.247	-0.465**	-0.886***	-0.263	-0.451***	-0.163	-0.964***	-0.538***
	(0.154)	(0.285)	(0.205)	(0.184)	(0.226)	(0.223)	(0.120)	(0.228)	(0.357)	(0.110)
Tennis	0.394***	-0.557***	0.114	0.452***	-0.096	0.659***	0.201**	0.348**	-0.059	0.229***
	(0.109)	(0.214)	(0.154)	(0.121)	(0.163)	(0.146)	(0.085)	(0.163)	(0.264)	(0.076)
Volleyball	0.702***	-0.077	0.698**	0.299	0.433	0.446*	0.593***	0.549*	0.424	0.610***
	(0.197)	(0.616)	(0.313)	(0.286)	(0.403)	(0.257)	(0.177)	(0.329)	(0.538)	(0.154)
Other Sport	-0.036	-0.201	-0.417*	-0.125	-0.495**	0.010	-0.133	0.268	-0.417	-0.223**
	(0.141)	(0.253)	(0.216)	(0.156)	(0.222)	(0.184)	(0.112)	(0.212)	(0.334)	(0.100)
Num.Obs.	1355	561	743	1312	777	758	2436	544	358	3069
AIC	3916.9	1675.3	2444.7	3182.5	2396.3	1908.9	6975.0	1810.6	895.4	8574.4
Log.Lik.	-1948.473	-827.668	-1212.336	-1581.241	-1188.142	-944.432	-3477.497	-895.280	-437.678	-4277.200
rho2										
rho20	0.400	0.385	0.320	0.497	0.362	0.480	0.405	0.314	0.490	0.419

* p < 0.1, ** p < 0.05, *** p < 0.01

The model estimates in Table 4 reveal that there is noticeable heterogeneity in the response among different socioeconomic groups. Park visitors living in block groups with a high proportion of Black and low-income residents show less affinity for trails and other walkways, but appear are also considerably to be more sensitive to the distance of a park. Visitors living in high-income neighborhoods are more attracted to the amenities of a park, or rather these visitors do not show significant negative coefficients for amenities, e.g. basketball courts and football fields.

Seeing that there is a difference in the response in the model segmentation, it is also worth considering the role of our segmentation thresholds in these findings. Figure 2 shows the estimated coefficients and confidence intervals for these different amenities at different threshold levels of segmentation. The threshold level means that at least that percent of the block group's population falls in that category. The confidence intervals widen as more observations are excluded from the model. The estimated coefficients for the different segmentations are identical when the share equals zero, and simply represent the "Sport Detail" model from Table 3.

Overall, increasing the segmentation threshold level reveals some additional information about user preferences. First, it should be noted that there is some inconsistency: for instance, block groups with at least 40% of low income households show a lower importance of distance than block groups with either 30% or 50% low income households. The increasing width of the confidence interval, however, means it is difficult to make generalized statements. Residents of block groups with a higher share of Asian or high income households both show relatively more affinity for tennis courts and trails. Block groups with high Black populations show a somewhat greater preference for playgrounds, as well appearing to be the most distance-sensitive group.

4.1. Model Application: Equity Analysis of COVID-19 Street Openings

In spring and summer 2020, cities across the world responded to the COVID-19 pandemic by converting city streets into temporary pedestrian plazas. The stated goals of this policy included providing recreational space that would allow people to walk and exercise outside with sufficient personal space and less risk of conflict from vehicle traffic. This policy created several dozen temporary open spaces in urbanized areas of the county. In this section, we apply the models estimated above to evaluate the benefits of this policy in terms of aggregate value, as well as the equity of the policy with respect to different income and ethnic groups.

We obtained the list of streets in Alameda County that were reported closed in the "Shifting Streets" COVID-19 mobility dataset (Combs et al., 2020). This dataset reports that 74 individual streets were closed to vehicle traffic (and thereby opened as public spaces); these streets represent 27.5730459 total miles across the cities of Berkeley, Oakland, and Alameda. For the purposes of this analysis, we represent each opened street as a single "park" without any sport facilities or playgrounds, but with a trail / walking path. The database provides the opened streets as polyline objects; we assert a 25-foot buffer around the line to represent a polygon with a measurable area. Finally, we measure the network-based distance from each population-weighted block group centroid to the nearest boundary of each new open space facility created by this policy.

Using this new dataset — augmented with parks added by street openings — we applied the "Sport Detail" non-segmented model to calculate park choice utilities and utility-based accessibility values for each block group in Alameda County. As shown with Equation (3), the difference in utility-based accessibility values with and without the opened streets is the consumer surplus provided by the policy. This surplus is given in a unitless utility, but it is possible to convert the surplus into monetary units by diving the surplus by a cost utility coefficient. The dataset used for this research does not have any information on travel costs or entrance fees, and such data would likely not be relevant in the context of urban parks. As a result, there is no direct link between the utility and a monetary cost in our estimated models.

As a substitution, we use an estimate of the cost coefficient obtained from the open-source activity-based travel demand model ActivitySim, which is itself based on the regional travel model employed by the Metropolitan Transportation Commission (MTC), the San Francisco Bay regional MPO. ActivitySim uses a cost coefficient of -0.6 divided by the each simulated agent's value of time to determine destination

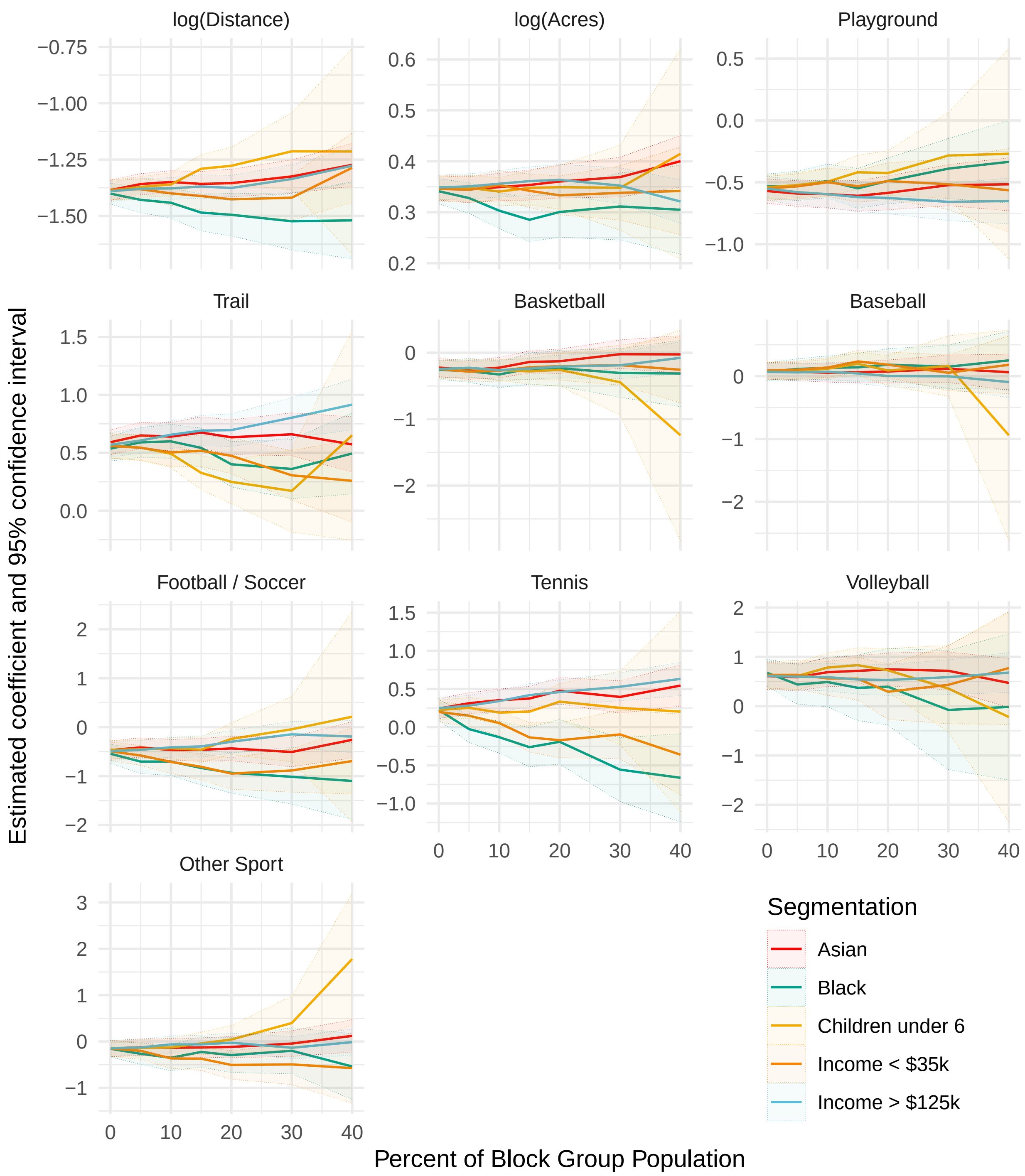

Figure 2: Estimated utility coefficients and 95% confidence intervals for park amenities at different socioeconomic threshold levels.

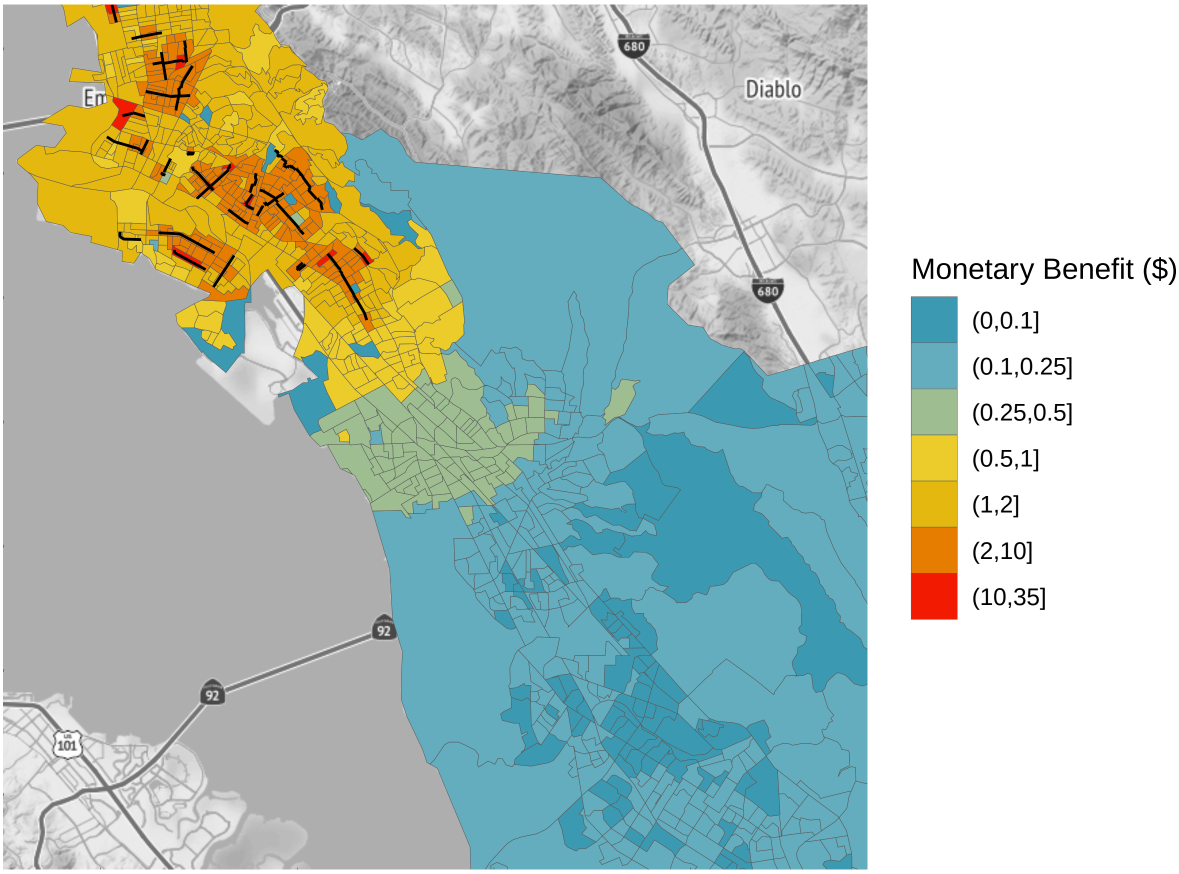

Figure 3: Monetary value of street opening to residents based on utility change.

choices for non-work trips.[1] In ActivitySim, as in most activity-based travel models, the value of time is considered to vary with an individual's income, but in this aggregate destination choice model, an aggregate value of time will suffice. The average value of time in the synthetic population for the Bay Area is \$7.75 per hour, resulting in a cost coefficient on the destination choice utility of -0.215. Dividing the difference in accessibility logsums by the negative of this value gives an initial estimate of the monetary value of the policy to each park user.

Figure 3 presents this monetary valuation spatially. Unsurprisingly, the benefits are concentrated in the block groups surrounding the opened streets. Most residents of central Oakland see a benefit of somewhere around \$1, while some zones see an equivalent benefit of as much as \$30. One property of logsum-based accessibility terms is that there is some benefit given for simply having more options, whether or not those options are attractive in any way. In this application, these benefits are small, on the order of 10 cents for most block groups away from where the street openings occurred.

More interesting than the total benefit or even its spatial distribution, however, is the social equity of its distribution among different population segments. If we assign the block-group level monetary benefit to each household in the block group, we can begin to allocate the distribution of benefits proportionally to households of different sociodemographic classifications. Specifically, if a block group with N total households has a measured consumer surplus δCS, then the share of the total benefits going to a particular population segment k is

[1] To be precise, this is the cost coefficient on the mode choice model for social, recreational, and other trip purposes, which influences destination choice through a logsum-based impedance term.

Table 5: Equity Distribution of Street Opening Benefits

Group	Benefit	Percent of Benefits	Households	Percent of Households
Households with Children under 6	$91,530	14.15	86,095	15.13
Income $<$ $35k	$157,608	24.36	102,580	18.03
Income $>$ $125k	$164,211	25.38	190,573	33.49
Black	$117,779	18.20	63,037	11.08
Asian	$124,652	19.27	157,980	27.76
Hispanic	$144,319	22.31	117,935	20.72
White	$220,039	34.01	196,736	34.57
All Households	$647,023	100.00	569,070	100.00

$$S_k = N * P_k * \delta CS \tag{4}$$

where and P_k is the proportion of the block group's population in segment k. There is some opportunity for confusion when some demographic variables we use (share of households with children, household income) are defined at the household level and other (ethnicity) are defined at the person level. It is similarly not clear whether the benefits of improved park access should be assigned at the person level, the household level, or the number of total park trip makers in each block group. For consistency and simplicity, we assert that the benefit is assigned to each household, and that persons receive a proportional share of the household benefit. For example, a block group with 30% Black individuals will receive 30% of the benefits assigned to all the households in the block group.

Table 5 shows the total benefit assigned to households in this way as well as the share of all monetary benefits in the region. In some cases, the policy of opening streets as public spaces had a pro-social benefit, as 18.7% of benefits went to Black individuals, even though only 11.4% of the population of Alameda County is Black. Similarly, roughly one-quarter of total benefits went to households making less than $35,000 per year even though only one-fifth of the households are in this category. On the other hand, a smaller than expected share of benefits is allocated to Asian individuals and households making more than $125,000 per year.

By this analysis, the policy to open streets as pedestrian plazas and public spaces appears to be a pro-social policy with substantial benefits to the community. There are some limitations and caveats that ought to be considered; for example, COVID-19 led to the closure of some park facilities that were not captured in this analysis. This policy would lead to a decrease in the consumer surplus for park access, which might overwhelm or at least change the distribution of benefits we measured here. A policy of permanently closing these streets to vehicle traffic would also have potentially deleterious effects on transportation access that would need to be considered against the benefits we measure here; in the case of the COVID-19 quarantine, the opportunity cost of closing a street to nonexistent vehicle traffic is basically zero.

5. Limitations and Future Directions

The ideal dataset for estimating individual choices would be a high-quality, large-sample household travel survey of real individuals. Such a survey would give details on whether an observed trip to a park was actually a recreation trip or rather a different activity entirely. The individual-level demographic data would also be valuable in understanding more clearly the observed heterogeneity in response among different income or ethnic groups. Additionally, the trends and correlations revealed in the presented models may reflect situational inequalities rather than true preferences. For example, the distinct observed parameters on size and distance for minority block groups may indicate that areas with large minority populations tend to have smaller parks that are more geographically distributed relative to other regions of the region. Transit access may also affect park choice and how far people are willing to travel to access a park. Preliminary analysis of our source data indicates a qualitative correlation between good transit access and diverse park use from both a geographic and demographic perspective.

We limited our analysis to home locations and parks in Alameda County, California. It is possible that some Alameda residents visit parks in neighboring counties, just as it is possible that parks in Alameda County attract trips from outside the county borders. This is most likely for block groups and parks on the north and south borders of the county. The lower measured accessibility in the area around Berkeley in the northern part of the county () is likely affected by the ommission of parks and residents in Contra Costa County.

Using Euclidean distance to represent the distance between the block group centroid and the border of the park leaves something to be desired: Depending on network topography and built environment characteristics, there may be a significant variation in perceived travel times between two parks with similar straight-line distances. That said, a more precise network-based measure might not overcome the inaccuracies resulting from our necessarily measuring distances from the block group centroid. As above, an individual-level survey where the home location is explicitly known would be preferable regardless of the distance method employed.

The activity location data used in this specific analysis treats all days of the week and day periods together; it is likely that weekend park choice is substantially different from weekday choice, as the activities performed may be the same. Also recall that the data consider each device-park pair as a unique trip. Repeated trips to the same park may not be properly considered in the data sample. A more precise time-of-day and day-of-week segmentation is warranted.

We applied a naive random sampling of the alternatives in our model estimation and validation; a more considered approach involving hierarchical destination sampling may yield more efficient estimates and therefore a clearer picture of the role of size, distance, and other amenities on the observed choices. The relatively weak predictive power of such a simple model formulation (size and distance only) indicates that there is potential to examine the role that additional park amenities — ball fields, playgrounds, water features, etc. — play in the relative attractiveness of parks for different market segments. The quality of park maintenance is another important feature identified in the recreation literature (Fletcher and Fletcher, 2003) that is not included here.

6. Conclusions

As transportation professionals seek to improve access to parks and better coordinate transportation and land use efforts — and as researchers more generally try to understand the role parks and open spaces play in public health and society — it is increasingly important to better understand how, when, and why individuals travel to parks. This intersection between recreation and transportation has received relatively little exploration, partially because travel survey data emphasizes weekday travel and because the role of parks in daily activities can be more complicated than with other land uses. This study contributes to the understanding of recreation access by presenting a method to develop access measures explicitly based on the observed choices of individuals. The resulting access measure is continuously defined and incorporates multiple dimensions of access, including the travel necessary to reach all nearby parks as well as the amenities of each of those parks. Further, the measure we have presented reveals heterogeneous preferences for travel and park size across market segments, a heterogeneity that could perhaps be incorporated into an understanding of accessibility.

With the growing availability of passive transportation data, there is a correspondingly increased opportunity to explore such data to develop a better understanding of travel patterns in more careful detail than is possible with household travel surveys. Capturing a sufficiently large survey to study trip patterns to a single park is an enormous undertaking, and doing such an exercise for an entire park system is prohibitively expensive and time-consuming. Passive data sets therefore enable analyses that would be unlikely or impossible by other means. Challenges to the representativeness and comprehensiveness of passive data products are in many cases fair, but this should not preclude their use in cases where traditional techniques are not practicable.

Acknowledgments

Graphics and tables were developed using several open-source R packages (Kahle and Wickham, 2013; Arel-Bundock, 2020; Ram and Wickham, 2018).

References

Jacobellis v. Ohio. *US Supreme Court*, 378(No. 11):184, 1964.

Alex Anas. Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological*, 17(1):13–23, 1983.

Vincent Arel-Bundock. modelsummary: Summary tables and plots for statistical models and data: Beautiful, customizable, and publication-ready, 2020. URL https://vincentarelbundock.github.io/modelsummary/. R package version 0.6.3.

John M Baas, Alan Ewert, and Deborah J Chavez. Influence of ethnicity on recreation and natural environment use patterns: Managing recreation sites for ethnic and racial diversity. *Environmental Management*, 17(4):523, 1993.

Carolyn Bancroft, Spruha Joshi, Andrew Rundle, Malo Hutson, Catherine Chong, Christopher C. Weiss, Jeanine Genkinger, Kathryn Neckerman, and Gina Lovasi. Association of proximity and density of parks and objectively measured physical activity in the United States: A systematic review. *Social Science & Medicine*, 138:22–30, aug 2015. ISSN 02779536. doi: 10.1016/j.socscimed.2015.05.034. URL https://linkinghub.elsevier.com/retrieve/pii/S0277953615003160.

P. O. Barnard and R. E. Brindle. A review and critique of current methods used to predict traffic generation with some accompanying suggestions on alternative approaches. *Transportation Planning and Technology*, 11(4):273–288, jun 1987. ISSN 0308-1060. doi: 10.1080/03081068708717349. URL http://www.tandfonline.com/doi/abs/10.1080/03081068708717349.

Vincent L Bernardin, John L Bowman, Mark Bradley, Jason Chen, Nazneen Ferdous, and Yuen Lee. Incorporating Big Data in an Activity-Based Travel Model: The Chattanooga Case Study. In *TRB Annual Meeting*. Transportation Research Board, 2018. URL https://trid.trb.org/view/1495930.

R. Bolla and F. Davoli. Road traffic estimation from location tracking data in the mobile cellular network. In *2000 IEEE Wireless Communications and Networking Conference. Conference Record (Cat. No.00TH8540)*, pages 1107–1112. IEEE, 2000. ISBN 0-7803-6596-8. doi: 10.1109/WCNC.2000.904783. URL http://ieeexplore.ieee.org/document/904783/.

Candice M Bruton and Myron F Floyd. Disparities in built and natural features of urban parks: comparisons by neighborhood level race/ethnicity and income. *Journal of Urban Health*, 91(5):894–907, 2014.

Jason Byrne. When green is white: The cultural politics of race, nature and social exclusion in a los angeles urban national park. *Geoforum*, 43(3):595–611, 2012.

Jason Byrne and Jennifer Wolch. Nature, race, and parks: past research and future directions for geographic research. *Progress in human geography*, 33(6):743–765, 2009.

Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. Estimating Origin-Destination Flows Using Mobile Phone Location Data. *IEEE Pervasive Computing*, 10(4):36–44, 2011. doi: 10.1109/mprv.2011.41. URL http://dx.doi.org/10.1109/mprv.2011.41http://hdl.handle.net/1721.1/101623http://creativecommons.org/licenses/by-nc-sa/4.0/.

Hsueh-Sheng Chang and Chin-Hsien Liao. Exploring an integrated method for measuring the relative spatial equity in public facilities in the context of urban parks. *Cities*, 28(5):361–371, 2011.

City Parks Alliance. Why City Parks Matter, 2019. URL https://cityparksalliance.org/about-us/why-city-parks-matter/.

Tabitha Combs, Carlos Felipe Pardo, Epiandes, MobilityWorks, and Datasketch. The "shifting streets" covid-19 mobility dataset., 2020. URL http://pedbikeinfo.org/resources/resources_details.cfm?id=5235. Accessed: 2020-12-10.

Yves Croissant. *mlogit: Multinomial Logit Models*, 2019. URL https://cran.r-project.org/package=mlogit.

Juan de Dios Ortúzar and Luis G Willumsen. *Modelling transport*. John wiley & sons, 2011.

Gerard de Jong, Andrew Daly, Marits Pieters, and Toon van der Hoorn. The logsum as an evaluation measure: Review of the literature and new results. *Transportation Research Part A: Policy and Practice*, 41(9):874–889, nov 2007. ISSN 0965-8564. doi: 10.1016/J.TRA.2006.10.002. URL https://www.sciencedirect.com/science/article/pii/S0965856407000316.

Xiaojing Dong, Moshe E. Ben-Akiva, John L. Bowman, and Joan L. Walker. Moving from trip-based to activity-based measures of accessibility. *Transportation Research Part A: Policy and Practice*, 40(2):163–180, feb 2006. ISSN 0965-8564. doi: 10.1016/J.TRA.2005.05.002. URL https://www.sciencedirect.com/science/article/pii/S0965856405000820?via{%}3Dihub.

Donna Fletcher and Harold Fletcher. Manageable predictors of park visitor satisfaction: Maintenance and personnel. *Journal of Park and Recreation Administration*, 21:21–37, 2003. URL https://pdfs.semanticscholar.org/f5cb/11ef839c6366e90654d1a0110f11cbf65334.pdf.

Myron F Floyd and Kimberly J Shinew. Convergence and divergence in leisure style among whites and african americans: Toward an interracial contact hypothesis. *Journal of Leisure Research*, 31(4):359–384, 1999.

Markus Friedrich, Katrin Immisch, Prokop Jehlicka, Thomas Otterstätter, and Johannes Schlaich. Generating Origin-Destination Matrices from Mobile Phone Trajectories. *Transportation Research Record: Journal of the Transportation Research Board*, 2196:93–101, 2010. doi: 10.3141/2196-10. URL https://journals.sagepub.com/doi/pdf/10.3141/2196-10.

Karst T. Geurs and Bert van Wee. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12(2):127 – 140, 2004. ISSN 0966-6923. doi: https://doi.org/10.1016/j.jtrangeo.2003.10.005. URL http://www.sciencedirect.com/science/article/pii/S0966692303000607.

Paul H Gobster. Managing urban parks for a racially and ethnically diverse clientele. *Leisure sciences*, 24(2):143–159, 2002.

GreenInfo Network. California Protected Areas Database, 2019. URL https://www.calands.org/cpad/.

Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

S L Handy and D A Niemeier. Measuring Accessibility: An Exploration of Issues and Alternatives. *Environment and Planning A*, 29(7):1175–1194, jul 1997. ISSN 0308-518X. doi: 10.1068/a291175. URL http://journals.sagepub.com/doi/10.1068/a291175.

Ching-hua Ho, Vinod Sasidharan, William Elmendorf, Fern K Willits, Alan Graefe, and Geoffrey Godbey. Gender and ethnic variations in urban park preferences, visitation, and perceived benefits. *Journal of leisure research*, 37(3):281–306, 2005.

Arthur Huang and David Levinson. Axis of travel: Modeling non-work destination choice with GPS data. *Transportation Research Part C: Emerging Technologies*, 58:208–223, sep 2015. ISSN 0968-090X. doi: 10.1016/J.TRC.2015.03.022. URL https://www.sciencedirect.com/science/article/pii/S0968090X15001072.

Winston Husbands and Paul Idahosa. Ethnicity and recreation behaviour: a review and critique of the literature. *Canadian Ethnic Studies= Études ethniques au Canada*, 27(1):84, 1995.

Ray Hutchison. Ethnicity and urban recreation: Whites, blacks, and hispanics in chicago's public parks. *Journal of Leisure Research*, 19(3):205–222, 1987.

Institute of Transportation Engineers. *Trip Generation Manual*. Institute of Transportation Engineers, Washington, D.C., 10th editi edition, 2017. ISBN 9781933452647.

Patricia Noonan Irwin, William C Gartner, and Carolyn C Phelps. Mexican-american/anglo cultural differences as recreation style determinants. *Leisure Sciences*, 12(4):335–348, 1990.

Nadja Kabisch and Dagmar Haase. Green justice or just green? provision of urban green spaces in berlin, germany. *Landscape and urban planning*, 122:129–139, 2014.

Andrew T. Kaczynski, Jasper Schipperijn, J. Aaron Hipp, Gina M. Besenyi, Sonja A. Wilhelm Stanis, S. Morgan Hughey, and Sara Wilcox. ParkIndex: Development of a standardized metric of park access for research and planning. *Preventive Medicine*, 87:110–114, jun 2016. ISSN 0091-7435. doi: 10.1016/J.YPMED.2016.02.012. URL https://www.sciencedirect.com/science/article/pii/S0091743516000517.

David Kahle and Hadley Wickham. ggmap: Spatial visualization with ggplot2, 2013. URL https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

J. C. Kinnell, M. F. Bingham, A. F. Mohamed, W. H. Desvousges, T. B. Kiler, E. K. Hastings, and K. T. Kuhns. Estimating Site Choice Decisions for Urban Recreators. *Land Economics*, 82(2):257–272, may 2006. ISSN 0023-7639. doi: 10.3368/le.82.2.257. URL http://le.uwpress.org/cgi/doi/10.3368/le.82.2.257.

Anurag Komanduri, Laura Schewel, Dan Beagan, and Dale Wong. Using Big Data to Develop Comparative Commercial Vehicle Metrics for Port Traffic at Major Ports in the U.S. In *TRB Annual Meeting*. Transportation Research Board, 2017. URL https://trid.trb.org/View/1438797.

Josephine D Kressner. Synthetic Household Travel Data Using Consumer and Mobile Phone Data. Technical report, Transportation Research Board, 2017.

Anastasia Loukaitou-Sideris. Urban form and social context: Cultural differentiation in the uses of urban parks. *Journal of Planning Education and Research*, 14(2):89–102, 1995.

Juliana Madzia, Patrick Ryan, Kimberly Yolton, Zana Percy, Nick Newman, Grace LeMasters, and Cole Brokamp. Residential Greenspace Association with Childhood Behavioral Outcomes. *The Journal of Pediatrics*, dec 2018. ISSN 0022-3476. doi: 10.1016/J.JPEDS.2018.10.061. URL https://www.sciencedirect.com/science/article/pii/S0022347618315683.

Gavin R. McCormack, Melanie Rock, Ann M. Toohey, and Danica Hignell. Characteristics of urban parks associated with park use and physical activity: A review of qualitative research. *Health & Place*, 16(4):712–726, jul 2010. ISSN 1353-8292. doi: 10.1016/J.HEALTHPLACE.2010.03.003. URL https://www.sciencedirect.com/science/article/pii/S1353829210000316.

Daniel McFadden. The measurement of urban travel demand. *Journal of public economics*, 3(4):303–328, 1974a.

Daniel L. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Paul Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974b.

Jürgen Meyerhoff, Alexandra Dehnhardt, and Volkmar Hartje. Take your swimsuit along: the value of improving urban bathing sites in the metropolitan area of Berlin. *Journal of Environmental Planning and Management*, 53(1):107–124, jan 2010. ISSN 0964-0568. doi: 10.1080/09640560903399863. URL http://www.tandfonline.com/doi/full/10.1080/09640560903399863.

Adam Millard-Ball. Phantom trips: Overestimating the traffic impacts of new development. *Journal of Transport and Land Use*, 8(1):31–49, 2015. doi: 10.5198/jtlu.2015.384. URL https://www.jstor.org/stable/pdf/26202700.pdf?refreqid=excelsior{%}3A6bf53b09919c223966f1dc8468372150.

Richard Mitchell and Frank Popham. Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet*, 372(9650):1655–1660, nov 2008. ISSN 0140-6736. doi: 10.1016/S0140-6736(08)61689-X.

Christopher Monz, Milan Mitrovich, Ashley D'Antonio, and Abigail Sisneros-Kidd. Using Mobile Device Data to Estimate Visitation in Parks and Protected Areas: An Example from the Nature Reserve of Orange County, California. *Journal of Park and Recreation Administration*, 0(0), sep 2019. ISSN 21606862. doi: 10.18666/JPRA-2019-9899. URL https://js.sagamorepub.com/jpra/article/view/9899.

Diala Naboulsi, Marco Fiore, Stephane Ribot, and Razvan Stanica. Large-Scale Mobile Traffic Analysis: A Survey. *IEEE Communications Surveys & Tutorials*, 18(1):124–161, 2016. ISSN 1553-877X. doi: 10.1109/COMST.2015.2491361. URL http://ieeexplore.ieee.org/document/7299258/.

National Academies of Sciences Engineering and Medicine. *Travel Demand Forecasting: Parameters and Techniques*. NCHRP 716. National Academies Press, Washington, D.C., 2012. ISBN 978-0-309-28255-0. doi: 10.17226/14665. URL http://www.nap.edu/catalog/14665.

Mark Padgham, Bob Rudis, Robin Lovelace, and Maëlle Salmon. osmdata. *The Journal of Open Source Software*, 2(14), jun 2017. doi: 10.21105/joss.00305. URL https://doi.org/10.21105/joss.00305.

Changxuan Pan, Jiangang Lu, Shan Di, and Bin Ran. Cellular-Based Data-Extracting Method for Trip Distribution. *Transportation Research Record: Journal of the Transportation Research Board*, 1945(1):33–39, jan 2006. ISSN 0361-1981. doi: 10.1177/0361198106194500105. URL http://journals.sagepub.com/doi/10.1177/0361198106194500105.

Laura L Payne, Andrew J Mowen, and Elizabeth Orsega-Smith. An examination of park preferences and behaviors among urban residents: the role of residential location, race, and age. *Leisure sciences*, 24(2):181–198, 2002.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

Karthik Ram and Hadley Wickham. wesanderson: A wes anderson palette generator, 2018. URL https://CRAN.R-project.org/package=wesanderson. R package version 0.3.6.

Wilfred W. Recker and Lidia P. Kostyniuk. Factors influencing destination choice for the urban grocery shopping trip. *Transportation*, 7(1):19–33, mar 1978. ISSN 0049-4488. doi: 10.1007/BF00148369. URL http://link.springer.com/10.1007/BF00148369.

Elizabeth A Richardson, Richard Mitchell, Terry Hartig, Sjerp de Vries, Thomas Astell-Burt, and Howard Frumkin. Green cities and health: a question of scale? *Journal of Epidemiology and Community Health*, 66(2):160 LP – 165, feb 2012. doi: 10.1136/jech.2011.137240.

Alessandro Rigolon, Matthew Browning, and Viniece Jennings. Inequities in the quality of urban park systems: An environmental justice investigation of cities in the United States. *Landscape and Urban Planning*, 178:156–169, oct 2018. ISSN 0169-2046. doi: 10.1016/J.LANDURBPLAN.2018.05.026. URL https://www.sciencedirect.com/science/article/pii/S0169204618304316.

Josh Roll. Evaluating Streetlight Estimates of Annual Average Daily Traffic in Oregon. Technical report, Oregon Department of Transportation, 2019. URL https://trid.trb.org/View/1630252.

D. C. Shoup. Truth in Transportation Planning. *Journal of Transportation and Statistics*, 6(1):1–12, 2003.

Kenneth E. Train. *Discrete Choice Methods with Simulation, 2nd Edition*. Cambridge University Press, Cambridge, mar 2009. doi: 10.1016/S0898-1221(04)90100-9. URL http://linkinghub.elsevier.com/retrieve/pii/S0898122104901009.

Trust for Public Land. 2019 ParkScore Index, 2019. URL https://www.tpl.org/parkscore.

U.S. Census Bureau. QuickFacts: Alameda County, California, 2019. URL https://www.census.gov/quickfacts/fact/table/alamedacountycalifornia/PST045218.

Emily Neusel Ussery, Leah Yngve, Dee Merriam, Geoffrey Whitfield, Stephanie Foster, Arthur Wendel, and Tegan Boehmer. The National Environmental Public Health Tracking Network Access to Parks Indicator: A National County-Level Measure of Park Proximity. *Journal of park and recreation administration*, 34(3):52, 2016. doi: 10.18666/JPRA-2016-V34-I3-7119. URL http://www.ncbi.nlm.nih.gov/pubmed/28868528http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5580831.

Kyle Walker. *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*, 2019. URL https://cran.r-project.org/package=tidycensus.

Randel F Washburne. Black under-participation in wildland recreation: Alternative explanations. *Leisure sciences*, 1(2): 175–189, 1978.

H C W L Williams. On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A: Economy and Space*, 9(3):285–344, mar 1977. ISSN 0308-518X. doi: 10.1068/a090285. URL http://journals.sagepub.com/doi/10.1068/a090285http://ideas.repec.org/a/pio/envira/v9y1977i3p285-344.html.

Chao Wu, Xinyue Ye, Qingyun Du, and Ping Luo. Spatial effects of accessibility to parks on housing prices in shenzhen, china. *Habitat International*, 63:45–54, 2017.

In-Kwon Yeo and Richard A Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87:954–959, 2000. URL https://www.jstor.org/stable/pdf/2673623.pdf?refreqid=excelsior%3A7f84252acd9abceb732e404db81fe53f.

Jiayu Zhu and Xin Ye. Development of destination choice model with pairwise district-level constants using taxi GPS data. *Transportation Research Part C: Emerging Technologies*, 93:410–424, aug 2018. ISSN 0968-090X. doi: 10.1016/J.TRC.2018.06.016. URL https://www.sciencedirect.com/science/article/pii/S0968090X18306454?via{%}3Dihub{#}b0085.