Utility-Based Accessibility to Community Resources: An Application of Location-Based Services Data

Gregory Macfarlane[1,*], Emma Stucki[1], Michael Copley[1],

Abstract

Understanding who in a community has access to its resources – parks, libraries, grocery stores, etc. – has profound equity implications, but typical methods to understand access to these resources are limited. Travel time buffers require researchers to assert mode of access as well as an arbitrary distance threshold; further, these methods do not distinguish between destination quality attributes in an effective way. In this research, we present a methodology to develop utility-based accessibility measures for parks, libraries, and grocery stores in Utah County, Utah. The method relies on passive location-based services data to model destination choice to these community resources; the destination choice model utility functions in turn allow us to develop a picture of regional access that is sensitive to: the quality and size of the destination resource; continuous (non-binary) travel impedance by multiple modes; and the sociodemographic attributes of the traveler. We then use this measure to explore equity in access to the specified community resources across income level and minority status in Utah County.

Keywords: Accessibility Passive Data Location Choice

1. Introduction

NEED TO EXPAND, PROBABLY WITH SOME REFERENCES NOW IN LITERATURE REVIEW

Communities provide important resources to their members, and spatial exclusion from these resources can negatively affect both subjective measures of well-being (**?**) and economic opportunity. Measuring good or poor access to these resources is an important concern. But access involves more than merely distance or travel impedance: it is a function of the quality of the resource, and how many options for the resource are available.

In this paper, we consider utility-based access to parks, grocery stores, and libraries in Utah County, Utah. The utility preferences are estimated on location-based services data obtained from a third-party commercial data aggregator. We then use the model estimates to construct a composite accessibility measure and compare the measure with neighborhood-level sociodemographic characteristics.

The paper begins with a discussion of previous attempts to evaluate access to community resources. We then describe the methodology employed in this research, which makes use of novel third-party datasets

2. Literature

NEED TO SIMPLIFY / STREAMLINE

Research about accessibility in the recent years has placed a considerable focus on access to jobs and the corresponding negative impact that there is on a community when there is a lack of accessibility to quality jobs. A research study was done examining job accessibility of the poor in Los Angeles to determine whether it was a problem in accessibility that caused the employment distribution pattern that is present in

*Corresponding Author

Email addresses: gregmacfarlane@byu.edu (Gregory Macfarlane), stuckiemma@gmail.com (Emma Stucki)

the cities today. This research was done to address an issue that is caused by not having access to jobs, and they found similar results to much of the other research that has been done, in that there are jobs that are accessible in the poor city centers, but the number of jobs is declining. So although there is access, the lack of access is concerning when considering the impact it could have on communities that are already struggling (Hu, 2015). In a similar research study done in Australia that examined accessibility to jobs, they connected accessibility of jobs to sense of well-being and satisfaction with life. In this study they found that transport disadvantage is positively associated with social exclusion (where their definition of social exclusion means comparatively less access to employment, shops, and other entertainment) and social exclusion is strongly negatively associated with well-being, showing an overall conclusion that social exclusion contributes to poor well being and transport disadvantage contributes to social exclusion (**?**). Both of these research studies showed the effect of lack of access to jobs, but neither really discussed the effects of limited accessibility to other public resources such as parks and greenspace, grocery stores, and libraries.

There is comparatively little research that has been done on the subject, but one article does specifically address nonwork accessibility and its impact on vulnerable social groups. Grengs (2015) found that when looking at accessibility among vulnerable social groups such as African Americans, Hispanics, low-income households, and households in poverty, there is a substantially larger share of households with extreme levels of low accessibility and so they share a remarkable disadvantage in accessibility to shopping and supermarkets. This research used a gravity model and included the impedance factor from traveling between origin and destination as well as the attractiveness factor based on the number of opportunities in the destination zone. The model of accessibility is relatively robust in including attractiveness and impedance for the different zones, however, the attractiveness part of the equation was entirely based upon the number of opportunities in the destination. Even this research lacks a study based upon what amenity is most attractive to people for a particular resource, and whether or not they have access to that amenity. For example, in a grocery store, is it more desirable to have more fresh produce or lower prices? Then using that information, does the person have access to the resources that they want? These variables help us identify the potential reach accessibility needs to have for a particular resource. In our study we will attempt to analyze different variables that could contribute to the attractiveness of a resource for parks and greenspace, grocery stores, and libraries.

We have chosen to analyze these resources because of several reasons, including their popularity in the community, the availability of existing data to collect, and the negative or positive impacts having access to these resources has on physical and emotional well-being.

One of the resources that we are analyzing is libraries. They are a space that allows people to gather to learn, escape pressures of life, connect with others, and socialize in a way that is different than any other community building. In addition, libraries are a community resource that are almost entirely supported by the local community. (Kalikow Maxwell, 2008) Libraries are indoor spaces that are free for public use, and available in most communities. In a major study of residential environments, libraries were found to be more popular than any other amenity except a food or drugstore. This was true for every demographic. However, despite its popularity and social benefit, libraries are still not very prevalent. According to the statistical abstract there are 39,400 pharmacies, 67,000 supermarkets and only 16,192 public libraries in the U.S. For the third most desired community resource, the number of libraries is remarkably low, and thus it is essential to improve accessibility to this resource. Additionally, libraries provide a place to gather together and learn together with other members of the community. Many stories of the aftermath of the tragedy of 9/11 tell how much of the community went to the library in search of information and community feeling and to gather together in their loss. Because libraries are a free resource, they are available to every demographic and something that is important to all (Barclay, 2017)

Parks fill a niche that is similar to libraries: just as libraries provide places for community gatherings and self-improvement, so do parks. Some of the reasons parks are essential are because of their benefit to mental health and physical health. Parks and other greenspace have been the subject of research when comparing access to parks and the influence on mental health. In a study done on young adults and teens and their access to green space, it was found that an increase in access to greenspace corresponded to a decrease in likelihood of anxiety, depression, or another mental health disorder. (Madzia et al., 2019) This research noted that some parks are more desirable than others because of their environment or community.

People may not feel safe in a certain place and be less likely to frequent a park for that reason, and so may not receive the benefits of the park close by. Yet there also may be other reasons not to frequent a park, such as low upkeep, lack of shade, or lack of amenities and play equipment for children or adults. One aspect that we will be analyzing in this research is what qualities of parks are more desirable, such as more vegetation, or trails, or sports courts, or playgrounds. All of these factors contribute to what draws a person to a particular park and can help us identify how to improve parks that are existing and not used as much because of a previously unknown variable. In addition to improving mental health as was analyzed in this research, parks can also provide ways to exercise and become healthy. In a review done about the proximity and density of parks and physical activity in the United States it was found that several studies found a positive correlation between proximity to parks and physical activity, and in the studies that compared multiple measurements and used smaller buffer sizes there was a stronger correlation between parks and physical activity (Bancroft et al., 2015). Although this is somewhat inconclusive, it is a factor that could be a positive impact on health for those within access to parks.

The community resource that has many proven studies that show a correlation between health and accessibility are grocery stores. Research that has been done on this subject has termed lack of accessibility to fresh produce in grocery stores as 'food deserts'. These food deserts are often located in areas with a low-income demographic, or a high percentage of minority population. As a result, these groups have less access to healthy foods and are more likely to have negative health effects. In a study done comparing access to supermarkets and fruit and vegetable consumption, it was found that when only looking at distance to nearest grocery store there was not a significant correlation between shopping and fruit and vegetable consumption. However, this study also found that many people passed their nearest option to go to a different supermarket for their primary shopping, and those who shopped at less expensive grocery stores had a corresponding diet with fewer fruits and vegetables (Aggarwal et al., 2014). Therefore, in addition to access to healthy foods, it is also personal choice that perhaps causes those in lower economic classes to choose to forgo healthier options for cheaper options. These lower income demographics also frequently do not have easy access to locations that accept food stamps, or other places, such as food pantries, where they can get access to healthier food options at an affordable price for their income. In a study done on access to fresh produce in low-income neighborhoods in Los Angeles it was found that only 41% of food pantry clients were within walking distance of stores with fresh produce, 83% were within walking distance of stores with limited produce, and 13% were not within walking distance of either store type. (Algert et al., 2006) Grocery store accessibility is important for other demographic groups as well for the same reasons, and despite the seemingly common presence of grocery stores throughout a city, we can see from this study there are still locations and people that experience a lack of accessibility.

Because of the benefit of having these resources close, it is important to identify what exactly makes something accessible. There are many different ways of defining accessibility including isochrone, distance, community-based, and network based. The isochrone definition of accessibility is defined as being based on location, such as whether or not you are within a mile of a certain resource. Algert et al. used this basis for their accessibility model when determining accessibility of low-income neighborhoods to healthy foods in Los Angeles. They used a network distance model, tracing roads a distance of 0.8 km. in every direction originating from each store location. (Algert et al., 2006) This idea of accessibility is limited because it fails to include different modes of travel or routes to go to the grocery store, such as on the way back from work. In addition, it also fails to include time accessibility as well as different variables such as familiarity, price comparison, or availability of food groups, that may encourage or dissuade a person from visiting a particular store. Another accessibility measure is the distance model, which determines accessibility by how close the nearest amenity is to a certain location. This definition of accessibility is slightly more variable than the isochrone method because it includes multiple stores in the method and includes stores that are perhaps a little further away but could be reached using different modes of transportation. But like the isochrone method, this does not use individual level measures such as activity patterns or personal preferences. This method was used in research done by Clifton to determine food availability for low-income families in Texas. This study was able to determine and use different mobility strategies, such as auto, rides, transit, walking, borrowing, taxis, etc. They also included an additional variable that questioned the distance to preferred supermarkets over distance to nearest supermarkets.(Clifton, 2004) This variable adds an individual level to the model in addition to the simple distance model. However, it still lacked the whole individual space time environment and included just a few distinct variables.

The community-based model of accessibility is probably one of the most simplistic definitions of accessibility of these four. This model is primarily used to determine if a particular resource is located in a particular city or county. This model could be helpful for resources that are perhaps not quite as prevalent such as hospitals or libraries, or for relatively small cities. However, for resources that are more common like parks and grocery stores or for metropolitan areas this model is not able to accurately represent accessibility for specific resources. (I can't find an example, do you have one?)

Access via network determines accessibility based on network availability rather than a set distance or time factor. Because of this it is a measure that can be very useful when looking at social exclusion within accessibility. This measure is also frequently more difficult to calculate because it is based off of individual characteristics and circumstance. In a study done analyzing the role of social capital influence variables on travel it was found that these variables distinctly affected the amount of time spent traveling as well as the travel mode of choice (Ciommo et al., 2014). This implies that these social capital variables in an individual network have a large effect on travel and accessibility and are an important factor to include.

Despite the benefit and particular variables that these models of accessibility favor, none of them can easily address the quality or the attributes of the target resource. Additionally, people might not go to the nearest grocery store if there is a better one a bit further away, or if there is a transit route that makes accessing a different one easier. There are two accessibility measures that can include some of these qualities: time-space accessibility and utility-based accessibility.

Time-space accessibility uses measures of time restriction as well as space restriction in the model to identify a potential interaction space for a person to access a resource. This type of model was used successfully in research done by Widener et al. on the accessibility of grocery stores and supermarkets in Cinncinati, Ohio. They allowed for a 120 minute time budget, and for their space measures they used transit systems along the commute from work to home, or just directly from home, and every option that was available along that line within the time budget (Widener et al., 2015). In another study done in China, Chen and Yeh (2021) used three different space-time accessibility measures to determine how these factors affected shopping activity frequency as well as travel distance. In the smallest space-time constraint it was found that service-rich areas had more shopping activities and trips but tended to have a smaller travel distance. In the medium and large space time constraints it was found that increase of neighborhood service opportunities did not significantly increase service related activities. This study found that unequal space-time measures are not significantly affected by an increase in density of resources unless the original space-time measure was small in a service rich area. Thus, it is important to note that accessibility is increased more by improving space-time measures, rather than by increasing density of resources. These measures should be included in data collection on measures of accessibility.

Another measure of accessibility is utility-based accessibility which looks at the utility that a person derives from the location or service at the end of their trip. This model is described in research done by Dong et al. to be composed of two things: systematic utility, which is observable attributes of the resource, and the disturbance, which is the unobservable part of the resource or the individual's opinion of the resource. Using these parameters a multinomial logit model is formed and the overall maximum utility is found to hypothesis the most likely option for each individual (Dong et al., 2006). Using this model, the utility function derived can be used to identify an individual response after a certain change in choice attributes. This allows the model to be extremely variable in regard to specific utility measures, but, it is limited in its ability to compare to different utility functions and is frequently difficult to interpret (Handy and Niemeier, 1997)

The best way to accurately determine accessibility to resources is to determine which attributes are most attractive to the individual and determine the cost of traveling to obtain that attribute in a particular resource. This is obtained through the utility-based model, as described previously, but it is very difficult to collect the data required to estimate these models, especially for non-frequent purposes. However, the benefits of such an approach proved successful in research done in Alameda County, California (Macfarlane et al., 2020). We can rely on large-scale passive origin-destination data sets to estimate these models. We will also use previously collected attributes for each resource along with these large data sets to determine the most attractive features of each community resource. Using these measures and this model, we can determine the accessibility of resources to different block groups and demographic groups and use this information to identify potential solutions to improving accessibility to the most desirable resources.

3. Methods

In this section, we present a method to estimate utility-based access to community resources in Utah County, Utah.

3.1. Modeling Framework

In a destination choice modeling framework (Recker and Kostyniuk, 1978), an individual at origin i considering a destination j from a set of possible destinations J has a choice probability

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_{j' \in J} e^{V_{ij'}}} \tag{1}$$

where V_{ij} is a linear-in-parameters function representing the utility of destination j. The destination utility consists of two basic elements:

$$V_{ij} = \beta t_{ij} + X_j \gamma \tag{2}$$

where t_{ij} is a measure of the travel impedance between i and j, X_j is a vector of attributes of destination j, and β, γ are estimated parameters relating the travel impedance and the destination attributes to the utility. These parameters may be estimated by maximum likelihood given sufficient observational data.

The logarithm of the denominator of the choice probability given in Equation (1) is a quantity called the *logsum* and represents the total value — or accessibility A — of the choice set for individual i (Williams, 1977; Handy and Niemeier, 1997)

$$A_i = \log \left(\sum_{j' \in J} e^{V_{ij'}} \right) + C \tag{3}$$

where C is an unknown constant resulting from the fact that the utility represented in Equation (2) is not absolute, but rather relative to the utilities of the other options. The difference in logsum values between two different origin points could be compared to determine which location has "better" accessibility to the destinations in question, based on the elements included in Equation (2). Accessibility might be improved by lower travel impedance, or by improved amenities, or even by simply having more options available.

These other elements include attributes of the community resource relevant to the destination choice problem: the size of the resource, amenities available, the price of goods on sale, etc. Each of these variables has an importance weighted against the travel impedance t_{ij}, which might take various forms depending on the data available and the destination resource in question.

Simple measures such as the highway travel time or the walk distance might be more or less appropriate for particular resources. Another option commonly used in travel demand models is actually the logsum of a *mode* choice model with the utility of choosing each mode given by a set of utility equations. In this study we adopt generic mode choice utility equations

$$\begin{aligned} V_{ij\text{auto}} &= -0.028 * (t_{ij\text{auto}}) \\ V_{ij\text{transit}} &= -4 - 0.028 * (t_{ij\text{transit}}) - 0.056 * (wt_{ij}) - 0.372 * (at_{ij}) \\ V_{ij\text{walk}} &= -5 - 0.028 * (t_{ij\text{walk}}) - 1.12 * (d_{ij} | d_{ij} < 1.5) - 5.58 * (d_{ij} | d_{ij} \geq 1.5) \end{aligned}$$

where t_{ij} is the travel time in minutes from i to j by each mode (including only in-vehicle time for transit), wt is the transit wait and transfer time, at is the time to access and egress transit by walking, and d_{ij} is the walking distance in miles. The walking distance uses two different utility parameters depending on whether the walking distance is greater than 1.5 miles. The travel impedance logsum between i and j is then

$$MCLS_{ij} = \log \left(\exp(V_{ij\text{auto}}) + \exp(V_{ij\text{transit}}) + \exp(V_{ij\text{walk}}) \right) \tag{4}$$

Table 1: Block Group Summary Statistics

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
Density: Households per square kilometer	340	0	558.3	659.3	0.0	394.2	4741.9
Income: Median block group income	330	2	80309.1	31030.5	20588.0	77099.0	196458.0
Low Income: Share of households making less than $35k	329	1	16.6	13.4	0.0	12.7	70.4
High Income: Share of households making more than $125k	322	1	23.0	17.1	0.0	19.1	92.3
Children: Share of households with children under 6	333	1	24.2	12.3	0.0	22.1	84.6
Black: Share of population who is Black	116	0	0.5	0.9	0.0	0.0	7.4
Asian: Share of population who is Asian	205	0	1.4	2.3	0.0	0.5	20.3
Hispanic: Share of population who is Hispanic*	330	0	11.6	10.6	0.0	8.6	62.1
White: Share of population who is White	339	0	82.6	11.9	32.8	84.3	100.0

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

3.2. Data

Utah County, Utah, is among the fastest-growing urbanized regions in the United States, with formerly agrarian areas and open rangeland being converted to predominately suburban built environments. The population and economic center of the county is in Provo and Orem, home to two large universities (Brigham Young and Utah Valley), but the most rapid development in commercial and residential terms has been in communities north of Utah Lake, between Provo and Salt Lake City to the north. Interstate 15 travels the spine of the county, and a commuter rail system travels multiple times a day between Provo and Salt Lake City with stations in Orem, American Fork, and Lehi. A bus rapid transit (BRT) system connects the universities, two commuter rail stations, and the densest portions of Provo and Orem, and other local bus services operate in other communities in the region. Table 1 presents descriptive statistics of the block groups in Utah County obtained from the 2015-2019 American Community Survey (ACS) using the tidycensus package for R (Walker and Herman, 2021).

3.2.1. Resource Data

Figure 1 shows the locations of three types of community resources in Utah County: parks, grocery stores, and libraries. For each resource, and initial list of resources and elementary attributes was obtained by executing a relevant query to OpenStreetMap (OSM) CITE OSM.

Public parks and their attributes retrieved from OSM were verified and corrected using aerial imagery and some site visits. The attributes included the size of the park in acres, whether the park includes a playground, and whether the park includes facilities for volleyball, basketball, and tennis. The constructed dataset includes `nrow(park_polygons)` attributed parks.

Grocery stores were retrieved from OSM and validated using internet resources and site visits. The complete Nutritional Environment Measures Survey (NEMS-S) CITE NEMS was collected for each store,

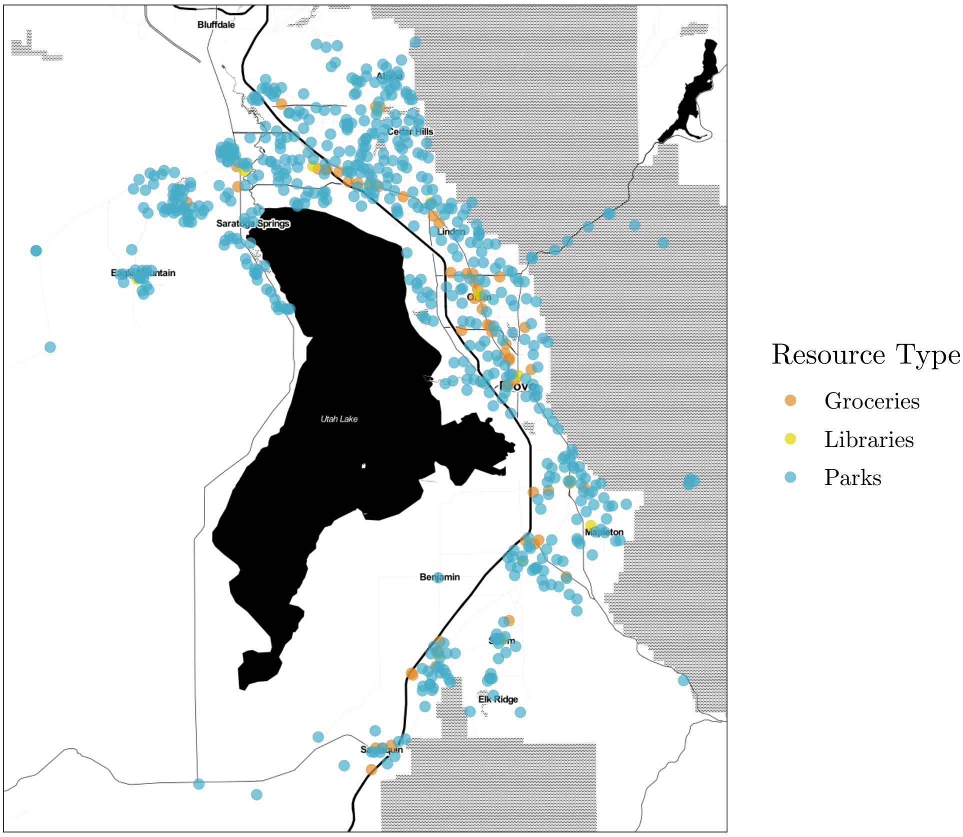

Figure 1: Community Resources in Utah County

but this preliminary analysis only includes cursory information on the stores including whether the store is a convenience store or some other non-traditional grocery, whether the store includes a pharmacy or other non-food merchandise, and the number of registers as a measure of the store's size. The constructed dataset includes `r nrow(groceries)‘ stores.

Libraries were retrieved from OSM, and validated using library websites and some site visits. The initial query returned university libraries and other specialty resources; though some of these libraries are open to those outside the university community, these were removed so that the resource list only includes libraries generally catering to the general public. The amenities available include whether the library offers additional classes and programs, and whether the library includes genealogical or family history resources. Other variables discussed in the literature such as the availability of computers and public wi-fi networks were present in every library and therefore cannot be included in the destination utility equations.

3.2.2. Mobile Device Data

Estimating the utility coefficients in Equation (2) requires observations of people making trips to the various resources. Regional household travel surveys could be used for this task, if a sufficiently large sample size allows an analyst to create a subsample of trips to the specific community resource. Unfortunately, low data availability requires that most discretionary activities be grouped into more general purposes CITE NCHRP 716.

In the last several years, various commercial data products developed from mobile device and location-based services (LBS) data have entered common use in transportation planning activities. Applications or websites that serve mobile content based on a users location will log this location information and sell the data to commercial third-party aggregators. These aggregators in turn will weight and anonymize the data before selling the prepared datasets to transportation planning agencies. These LBS datasets typically contain vehicle or person flows between spatially defined zones, sometimes segmented by inferred transportation mode, time of day, day of week, or imputed trip purpose. These datasets have been shown to accurately reflect visits to recreation areas and other land uses CITE MONZ, and are becoming a common part of transportation planning practice CITE TCRP, UDOT REPORT.

Macfarlane et al. (2021) present a method for estimating destination choice models from such data, which we repeat in this study. We provided a set of geometric polygons for each park, grocery store, and library to StreetLight Data, Inc., a commercial aggregator. StreetLight Data in turn provided data on the number of mobile devices observed in each polygon grouped by the inferred residence block group of of those devices on an average day in 2019. IS THIS CORRECT? We then created a simulated destination choice estimation dataset for each community resource by sampling 10,000 block group - resource "trips" from the StreetLight dataset. This created a "chosen" alternative; we then sampled ten additional resources at random (each simulated trip was paired with a different sample) to serve as the non-chosen alternatives. Random sampling of alternatives is a common practice that results in unbiased estimates, though the standard errors of the estimates might be larger than could be obtained through a more carefully designed sampling schemed CITE TRAIN.

3.2.3. Travel Times

Once the choice, alternatives, and attributes of the alternatives have been defined, the last element of the choice utility is the travel impedance between each block group and each resource. Using the `otpr` R interface (Marcus Young, 2020) to OpenTripPlanner CITE, we estimated the highway drive travel time, the walking route time, and the transit travel time for trips departing on October 1, 2021 at 8 AM. The time and date are most relevant for the transit path builder in OpenTripPlanner, which uses detailed transit path information stored in the Utah Transit Authority GTFS feed file for Fall 2021. The transit path contains separate measures of the total travel time, the time in the transit vehicle, transfer time, and access / egress time, allowing us full use of the mode choice utility equations and resulting logsum described in Equation (4).

For groceries and libraries, we queried from OTP the shortest time path on each mode from the population-weighted block group centroid to the centroid of the grocery store or library centroid. Because some parks in the dataset can be relatively large and the centroid might be far from the park access or use

Table 2: Park Destination Choice Utilities

	Car	MCLS	Attributes	All - Car	All - Logsum
Drive time	-0.215(-95.949)**			-0.209(-69.212)**	
Mode Choice Logsum		7.678(95.958)**			7.450(69.216)**
log(Acres)			1.308(77.120)**	1.300(46.869)**	1.301(46.858)**
Playground			4.567(33.939)**	4.476(30.127)**	4.477(30.118)**
Volleyball			-0.369(-9.580)**	-0.663(-11.065)**	-0.664(-11.067)**
Basketball			-0.669(-15.625)**	-0.534(-7.632)**	-0.535(-7.642)**
Tennis			-0.549(-13.065)**	-0.884(-14.678)**	-0.886(-14.693)**
Num.Obs.	8,984	8,984	8,984	8,984	8,984
Log Likelihood	-9,288.8	-9,284.7	-11,822.1	-4,774.9	-4,772.2
McFadden Rho-Sq	0.569	0.569	0.451	0.778	0.778

t-statistics in parentheses. * p < 0.5, ** p < 0.1

point, we instead sampled points along the boundary of the park polygon, and queried the shortest time path by each mode to the nearest boundary point.

4. Results

4.1. Destination Choice Models

Using the simulated trip choices assembled from the location-based services data, we estimate destination choice models with the `mlogit` package for R (R Core Team, 2021; Croissant, 2020).

Table 2 presents the model estimation results for five different specifications of park destination choice. The "Car" model includes only the network travel time by car as a predictor of park choice; the "MCLS" model similarly contains only the mode choice logsum as an impedance term. The signs on the coefficient indicate that people are more likely to choose parks with lower car distance or higher multi-modal access, all else equal. The "Attributes" model includes only information on the park attributes including size and amenities. On balance, people appear attracted to larger parks and parks with playgrounds, while somewhat deterred by various sports facilities. The "All" models include both the relevant travel impedance term as well as destination attributes.

For most block group-park pairs, the transit and walk travel disutilities are sufficiently high that choosing these travel modes is unlikely. As a result, the mode choice logsum is highly collinear with the car travel time. Nevertheless, there are small differences differences between the models with the different impedance terms. Using a non-nested likelihood statistic test presented by HOROWITZ, we can reject the null hypothesis that the two "All" models have equivalent likelihood (p-value of 0.00969), and infer that the mode choice logsum is a marginally better estimator of park choice than the vehicle travel time alone.

Table 3 presents the model estimation results for the grocery store models. As with the parks models in Table 2, the most predictive model contains both a travel impedance term and attributes of the destination grocery store. The number of registers suggests that people prefer larger stores, all else equal; ethnic markets, convenience stores, and other facilities are less preferred while stores with pharmacies and other merchandise (clothes, home goods, etc.) attract visitors. The ratio of the drive time and convenience store coefficients suggests that on average, people are willing to drive 6.86 minutes to reach a store that is not a convenience store. In terms of the travel impedance, there is not a sufficiently large gap in the model likelihoods to reject that the mode choice logsum and the drive time are equivalent predictors of grocery store choice.

Table 4 presents the model estimation results for the library destination choice models. As with parks and grocery stores, both travel impedance and destination attributes are significant predictors of library choice. In this case, however, the library attributes provide very little predictive power of library choice. This is perhaps because virtually all libraries in the dataset offer the same set of basic amenities, but also because each municipality in Utah County tends to operate its own library rather than having a system of

Table 3: Grocery Destination Choice Utilities

	Car	MCLS	Attributes	Size	All - Car	All - Logsum
Drive time	-0.206(-90.014)**				-0.217(-78.388)**	
Mode Choice Logsum		7.340(90.019)**				7.733(78.399)**
Convenience Store			-2.339(-11.310)**	-1.600(-7.684)**	-1.486(-6.765)**	-1.488(-6.773)**
Other non-standard			-1.894(-14.604)**	-1.255(-9.554)**	-1.055(-7.487)**	-1.056(-7.490)**
Has pharmacy			0.616(19.421)**	0.329(8.901)**	0.249(5.488)**	0.249(5.502)**
Ethnic market			-1.680(-16.846)**	-0.997(-9.750)**	-0.883(-8.072)**	-0.884(-8.078)**
Has other merchandise			1.523(48.309)**	0.769(19.144)**	0.881(17.631)**	0.882(17.660)**
Number of registers				0.073(42.117)**	0.083(36.312)**	0.083(36.294)**
Number of self-checkout				0.031(15.255)**	0.027(10.049)**	0.027(10.041)**
Num.Obs.	8,404	8,404	8,404	8,404	8,404	8,404
Log Likelihood	-11,861	-11,861.8	-16,898.4	-15,806.6	-8,802.4	-8,802.2
McFadden Rho-Sq	0.411	0.411	0.161	0.216	0.563	0.563

t-statistics in parentheses. * $p < 0.5$, ** $p < 0.1$

Table 4: Library Destination Choice Utilities

	Car	MCLS	Attributes	All - Car	All - Logsum
Drive time	-0.233(-95.379)**			-0.232(-89.281)**	
Mode Choice Logsum		8.306(95.361)**			8.270(89.266)**
Offers Classes			1.318(44.405)**	1.258(23.053)**	1.257(23.033)**
Genealogy Resources			-1.127(-44.021)**	-1.024(-25.610)**	-1.024(-25.601)**
Num.Obs.	9,816	9,816	9,816	9,816	9,816
Log Likelihood	-10,841.9	-10,840.3	-21,944.4	-10,322.5	-10,321.7
McFadden Rho-Sq	0.539	0.539	0.068	0.561	0.561

t-statistics in parentheses. * $p < 0.5$, ** $p < 0.1$

interconnected library branches as might be typical in larger cities or other regions. Additionally, there is no significant difference between the prediction power of the mode choice logsum versus the car travel time.

4.2. Accessibilities

Using the results of the "All - Logsum" models estimated for each community resource in the last section, we calculate the total utility-based accessibility measure for each block group in Utah County. For comparison to a more traditional measure, we also created buffer-based accessibility terms where a block group has "access" to a grocery store if there is one within a 5-minute drive, a park if there is one within a five-minute walk, and a library if there is one within a ten-minute drive.

Figure 2 spatially presents the difference between the buffer-based measure and the logsum-based measure. The two measures largely show the same basic shape: block groups along the spine of the county tend to have binary access in the buffer and also have a higher logsum value. The difference is at the margins, where the discontinuity of the buffer measure is replaced by a smoother access surface, more spatially reflective of what people are likely to experience.

The discontinuity of the buffer measure is perhaps more clearly seen with in Figure 3, which plots the logsum against the travel time in minutes (drive time for grocery stores and libraries; walk time for parks), for block groups in the study region. For block groups with equivalent travel time to a particular community resource, the accessibility logsum value varies substantially. This variance might be due to a travel time differential between drive, walk, and transit modes captured in the mode choice logsum, or it could also be because the resources available near the set of block groups have substantial variance in their amenities. Being near a single poor-quality grocery store is not the same thing as being near multiple high-quality groceries, and the logsum value can capture this variance in its construction.

What this distinction between access estimation methods might mean for policy analysis is yet to be determined, and future research is required. In this analysis, we estimate that 2,385 households live in

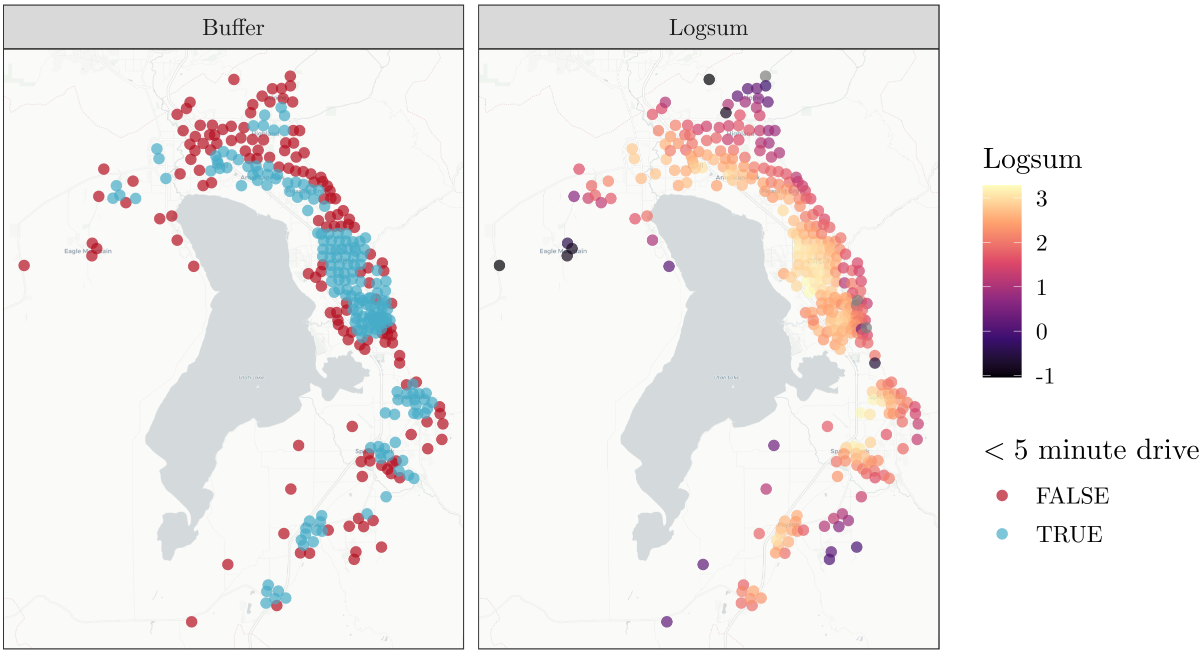

Figure 2: Spatial comparison of grocery access buffer versus logsum.

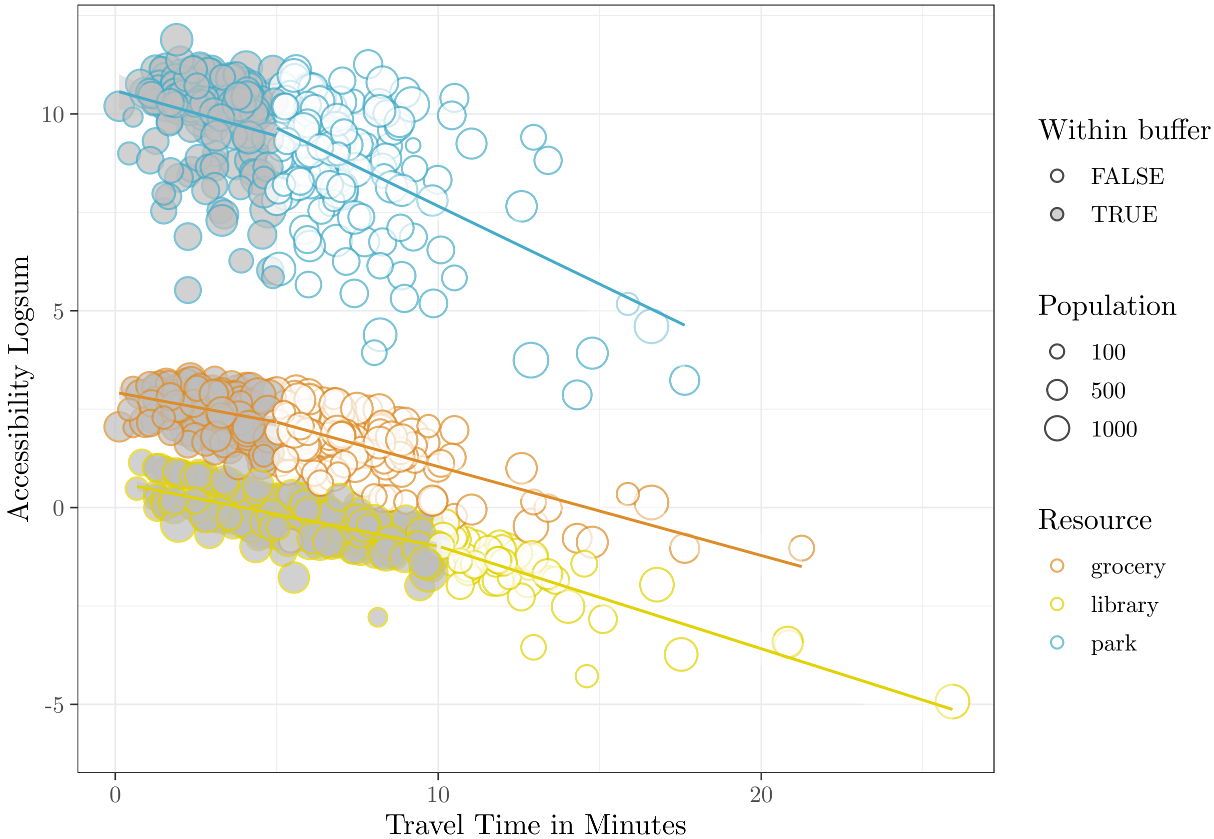

Figure 3: Relationship between travel time, travel-time based buffer, and logsum access value for block groups in Utah County.

block groups outside the boundary of all three resource buffers. At the same time, 1,633 households live in block groups that are beneath the regional mean utility-based access to all three resources. The population-weighted mean of the block group median incomes of these two groups is almost identical however, with $98,848 for the buffer-based measure and $97,929 for the utility-based version. This indicates that the areas of Utah County that are without access to these community resources tend to be exurban regions on the edges of the county, with relatively high income areas.

5. Limitations and Future Research

The location-based services data reveals the likely home location of devices observed within a geographic polygon, within some measurement error. It cannot tell us whether the device holder actually accomplished the assumed activity; that is, there may be a reason why a device was observed near a library even though the person did not actually patronize the library. Additionally, the method we use to compile the estimation dataset presumes that the choice to make a trip to the community resource has already been made. Though it can suggest how the accessibility of a neighborhood to these resource would improve were transportation impedance decreased or the resources expanded or improved, it cannot tell us how many more people might take advantage of the resource in that case.

In this research, all simulated trips were grouped into a single pooled model for analysis. This implies that the effect of amenities and travel impedance on destination choice is similar for all neighborhoods. A segmented model where, for example, low-income block groups and high-income block groups were estimated separately could allow for flexibility in these estimates and reveal differences in preferences among residents of the different neighborhoods. Some neighborhoods might show a particular preference for access utilities by transit, or for specific park amenities. A latent class choice model CITE JOAN would go further in potentially informing which demographic variables are meaningful in defining possible data segmentation schemes.

A necessary assumption made when constructing the estimation dataset is that people experience access from their home neighborhood. This may not always be true; for instance, people may choose to shop at grocery stores or visit libraries that are near their workplace, or that are between their homes and some other frequent destination. Methods to account for access and destination choices experienced at other points in the day would be a useful and interesting extension. Similarly, we assumed that the distance between a home and a community resource was represented by the distance between the block group centroid and the resource. For some block groups in less dense areas of the county, the error in measured travel time between the block group centroid and the actual home location might be larger than the total travel time. It might be possible to simulate home locations within each block group and use those locations in the travel time calculations. Alternatively, it might be possible to estimate the model using block group data as in this study but apply the model at a more fine resolution (e.g. block) when investigating accessibilities and conducting policy analyses.

This paper presents preliminary model estimates using plausible destination choice utility values. Several additional variables might be further explored, particularly in regards to the grocery resources. The NEMS-S survey is a highly detailed picture of the offerings of a particular grocery store, including information on the availability of relatively healthier or fresher foods and their prices. This study was only able to explore a few key size variables, but a deeper investigation into grocery store amenities and offerings preferences — and how they might influence a collective understanding of nutrition access more broadly — is needed.

Ultimately, the purpose of any accessibility measure to a community resources is to enable a subsequent analysis of some metric of well-being. CITE EP:B PAPER suggest that a utility-based access to parks measure is more predictive of physical health outcomes than a buffer-based measure. Is this true for more community resources? Would using a more subtle or nuanced measure of access to libraries help in understanding a link between community form and social isolation or mental health? A key benefit of this method is that is provides a way to evaluate the benefit of investments in resources against the benefits of investing in the transportation system. Will a community benefit more from a new grocery store nearby, or expanded options at an existing grocery store, or from improving bike or bus connections to that existing store? An examination of this question is left for future research, but this paper presents a method for how this could be approached.

6. Conclusions

NEED TO WRITE

Acknowledgements

The authors are grateful to Alisha Redelfs, Lori Spruance, Kaeli Monahan, and Mali Smith for their help in gathering the grocery store information data. Connor Williams gathered the parks data. Tables and figures in the article are produced using a variety of packages for R (Arel-Bundock, 2021; Dunnington, 2021)

References

Aggarwal, A., Cook, A. J., Jiao, J., Seguin, R. A., Vernez Moudon, A., Hurvitz, P. M., and Drewnowski, A. (2014). Access to supermarkets and fruit and vegetable consumption. *American journal of public health*, 104(5):917–923.

Arel-Bundock, V. (2021). *modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. R package version 0.8.0.

Bancroft, C., Joshi, S., Rundle, A., Hutson, M., Chong, C., Weiss, C. C., Genkinger, J., Neckerman, K., and Lovasi, G. (2015). 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.

Barclay, D. A. (2017). Space and the social worth of public libraries. *Public library quarterly*, 36(4):267–273.

Chen, Z. and Yeh, A. G.-O. (2021). Effects of built environment on activity participation under different space-time constraints: A case study of guangzhou, china. *Travel Behaviour and Society*, 22:84–93.

Croissant, Y. (2020). Estimation of random utility models in R: The mlogit package. *Journal of Statistical Software*, 95(11):1–41.

Dong, X., Ben-Akiva, M. E., Bowman, J. L., and Walker, J. L. (2006). Moving from trip-based to activity-based measures of accessibility. *Transportation Research Part A: policy and practice*, 40(2):163–180.

Dunnington, D. (2021). *ggspatial: Spatial Data Framework for ggplot2*. R package version 1.1.5.

Grengs, J. (2015). Nonwork accessibility as a social equity indicator. *International Journal of Sustainable Transportation*, 9(1):1–14.

Handy, S. L. and Niemeier, D. A. (1997). Measuring accessibility: an exploration of issues and alternatives. *Environment and planning A*, 29(7):1175–1194.

Macfarlane, G. S., Boyd, N., Taylor, J. E., and Watkins, K. (2020). Modeling the impacts of park access on health outcomes: A utility-based accessibility approach. *Environment and Planning B: Urban Analytics and City Science*, page 2399808320974027.

Macfarlane, G. S., Turley Voulgaris, C., and Tapia, T. (2021). If you build it who will come? equity analysis of park system changes during covid-19 using passive origin-destination data. *Journal of Transport and Land Use*. Under review, preprint available at https://byu-transpolab.github.io/alameda_destinationchoice/alameda_destinationchoice.pdf.

Marcus Young (2020). *otpr: An API wrapper for OpenTripPlanner*. R package version 0.5.0.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Recker, W. W. and Kostyniuk, L. P. (1978). Factors influencing destination choice for the urban grocery shopping trip. *Transportation*, 7(1):19–33.

Walker, K. and Herman, M. (2021). *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*. R package version 1.0.

Widener, M. J., Farber, S., Neutens, T., and Horner, M. (2015). Spatiotemporal accessibility to supermarkets using public transit: an interaction potential approach in cincinnati, ohio. *Journal of Transport Geography*, 42:72–83.

Williams, H. C. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and planning A*, 9(3):285–344.