Walking to transit – using big data to analyze bus and train ridership in Los Angeles

By

Ng Yim Chew Klo'e

Submitted to the Department of Urban Studies and Planning

in partial fulfillment of the requirements for the degree of

Master of Science in Urban Studies and Planning

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Author____

Department of Urban Studies and Planning

(16 August 2021)

Certified by ____

Professor Andres Sevstuk

Associate Professor of Urban Science and Planning

Thesis Supervisor

Department of Urban Studies and Planning

Accepted by____

Ceasar McDowell

Professor of the Practice

Chair, MCP Committee

Department of Urban Studies and Planning

Walking to transit – using big data to analyze bus and train ridership in Los Angeles

by

Ng Yim Chew Klo'e

Submitted to the Department of Urban Studies and Planning

on 16th August 2021

in partial fulfillment of the requirements for the degree of

Master of Science in Urban Studies and Planning

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ABSTRACT

Los Angeles passed one of the largest sales taxes in the country in 2016, which will give the county unprecedented financing in improving public transportation. Public transit ridership has been declining despite hefty investments, and it is important to understand why transit has not picked up. Studying current pedestrian-induced ridership is crucial as walkability is key in affecting ridership. Many prior studies assume linear relationships with established variables or explore transformed variables which have constrained assumptions. Machine learning models have the potential to discover nonlinear relationships such as step function and curvilinear relationships, which will help planners and policy makers make effective development decisions.

Thesis Supervisor: Andres Sevstuk

Title: Associate Professor of Urban Science and Planning

Introduction

In view of established benefits of walkable cities and parallel efforts to mitigate deficiencies caused by automobiles, planning authorities around the world have a renewed interest to prioritize urban environments that are more transit and pedestrian friendly. In Los Angeles, the city of freeways, this interest has been accompanied by decades of a sense of urgency and political controversy. Los Angeles' dependency on automobile date back to the 1920s, where a combination of favorable climate, widely scattered low-density population, and almost universal housing in detached single-family dwellings encouraged its widespread use (Bottles, 1987). Over the years, the city has seen various proposals to fund subways and elevated railways to undo problems rampant in car-centric cities — bad air quality, congestion, inequity, traffic deaths to name a few. The most recent initiative — Measure M, which was approved in 2016, is a full cent sales tax for Los Angeles County to finance new transportation projects and programs (Metro, 2017). This would generate $200 billion over 40 years for the L.A. Metropolitan Transportation Authority to expand rail, rapid bus, and bike networks and radically transform the public transportation system.

Faced with the nation's largest-ever infrastructure investment, Los Angeles has lofty plans to reinvent itself as the 'third-L.A.', one that is more transit-oriented and pedestrian-friendly. While this may bode well for a city synonymous with high car ownership and plagued with problems such as insufferable traffic conditions, pundits are less optimistic — and rightfully so, having had a history of multiple failed attempts at changing rooted Angelenos' behavior. L.A. has had a history of passing three major sales tax measures between 1980 and 2016 and building over 110 miles of rail in hopes of revamping the public transportation system. Despite these efforts, ridership on the transit system has been on the decline, with auto ownership and mileage travelled via private cars increasing at the same time. Walkability is key for transit use, and in a city that its dwellers are known to be resistant to walking and taking public transportation, herein lies an interesting question — who currently walks to take public transit in the city? Specifically, what are the factors affecting walking to transit stops and how do these differ between train and bus ridership? This paper uses big data to analyze pedestrian behavior in relation to public transit (bus and train) ridership and discusses the results using different computational methodologies. Data for modelling pedestrians in Los Angeles is collected from public data repositories as well as Streetlight Data, a data analytics company that provides actual location data collected from mobile devices. Next, Machine Learning methods are applied and compared against conventional models. Finally, implications of the findings and analysis are examined in the context of developing better policy recommendations and interventions for the city of Los Angeles.

Background

Long before the current smog-ridden and traffic-incapacitated dystopia Los Angeles is known for today, the city was once home to one of the largest mass transit systems in the world. The first electric trolley and streetcar lines emerged in 1887 in Los Angeles and expanded rapidly in the decade thereafter. Before the beginning of the last century, and for a half-century thereafter, streetcars dominated urban mass transit in Los Angeles (Rasmussen, 2003), and by 1920s the city had the best public streetcar system in the country.

The city grew in population and attracted real estate developers and entrepreneurs. Amongst them was Harry Huntington, one of the most prominent developer that was buying up land and developing housing outskirts. He supplemented these housing with streetcar lines to bring people around, and the motivation behind the expansion of streetcars was motivated by profiting from the real estate sales instead of building an efficient and economically sustainable public transportation network. The density for the streetcars simply did not make sense and was unprofitable.

In that same period, L.A. saw an increased dependency on automobiles. The preference for automobiles stemmed largely from widespread public dissatisfaction over streetcars, seen increasingly as undependable, overcrowded, and the main culprit for road congestions and accidents (Novak, 2013). Its popularity was also contributed by the improvement of local roads, emerging development of single-family tract houses and general affordability of cars. As a result, L.A.'s population of 600,000 into the 1920s people doubled in the decade and the number of registered cars quadrupled from 160,000 to 800,000 (Meares, 2019). By this time, Los Angeles had the highest ratio of automobiles per capita of any large city in the United States at about one auto per nine people, cementing a reputation of being decentralized, low-density and car dependent (Wachs, 1984). Concurrently, "voters and elected officials rejected several initiatives for the expansion and improvement of the regional rail network..in part because they did not wish to pay higher taxes and fares to support a crumbling transit system just as they were acquiring automobiles for the very first time" (Wachs, 1993). Over time, the dilapidated trolley lines went into decline and by the 1950s halted all operations. The streetcars, on the other hand, made their final run in 1963 and were replaced by diesel buses (Hobbs, 2014).

In 1947, the Master Plan of Metropolitan Los Angeles Freeways was adopted by the Regional Planning Commission and the construction of extensive freeway network commenced in 1950. In his seminal work "L.A. Freeway: An Appreciative Essay", Brodsly waxed lyrical about the freeway system being "the city's great synecdoche…employed to represent the totality of metropolitan Los Angeles and is one of the few parts capable of standing for the whole". The wide boulevards and freeways signaled the death of the streetcar system and further encouraged car ownership and usership. The reliance on cars subsequently brought about its own issues, with central L.A. becoming increasingly gridlocked. Although these freeways occupied an insignificant portion of land in Los Angeles area, freeways dominate the physical and psychological landscape and serving as the primary means of connections between the region's towns, suburbs and neighborhoods (Brodsly, 1981). As seen in a historic map dated 1963 (Figure 2), the physical structure of Los Angeles was pretty much shaped by the freeways back then.

The importance of having a rapid transit system for the future prosperity of Los Angeles was identified as early as 1906, but often took a backseat to highway investments. Unlike highway improvements that could be implemented in a piecemeal, less capital upfront fashion, transit developments were of a greater scale and cost. That did not sit well with the public that was already critical of private transit companies. Major rail transit initiatives did not take off due to a population that saw little value its construction (Adler 1987; Wachs 1993), unwillingness to allow private companies profit from raised fares, and a vulnerable city planning authority that gave in to political criticism (Wachs, 2007).

The Los Angeles Metropolitan Transit Authority was established in 1951 as a public transit planning agency for the region. The shift to public ownership expanded transit's mandate beyond maximizing profits. By the 1970s, automobile ownership has grown to a point where there are more registered vehicles than licensed drivers in L.A. Constantly expanding and building freeways were no longer sustainable for its car population growth and the dwellers became cognizant of issues surrounding air quality, energy, and quality of life. In 1980, majority of Los Angeles County voters (54.3%) approved Proposition A, a half-cent sales tax for transportation improvements with 35% of its revenue dedicated for rail constructure. A second half-cent sales tax for transportation, Proposition C, was approved again in 1990. Although the first ballot called for a temporary reduction in bus fares, the measures largely focused on only rail developments which detractors largely blamed the failure of the measures on.

In November 2016, 72 percent of votes in Los Angeles County approved of Measure M, a sales tax measure set to generate $200 billion over 40 years to expand rail, rapid bus, and bike networks (Bliss, 2019). This will give Los Angeles one of the biggest investments in the country to advance public transportation infrastructure. Despite being symbolically momentous, critics have also pointed out that voting for an improvement in transit system does not necessarily translate to intent to use it. Manville (2018) found in research conducted with 1,450 adults and concluded that "Few Angelenos viewed transit as an amenity that directly benefited them: They voted for Measure M as an expression of their political beliefs and in support of a broader social good", that someone else will benefit from the public service. Taylor & Morris

(2015) cite two policy loopholes with current transit governance structure. First, transit board members are appointed by the location that the transit agency serves, hence serving the voting public instead of actual transit riders requiring the services. Second, directors are typically drawn from an agency's transit service area which result in equal representation for both places with high and low transit ridership. For example, in Los Angeles County, four small suburbs have the same collective voting weight as Los Angeles city, which has roughly ten times the population and more than ten times the number of riders. Given that majority of transit riders typically are more vulnerable (ethnic minorities/lower household income), passing measures to improve transit ridership may not directly lead to a conversion in transit ridership.

Transport systems have and continue to be a central agenda of policy makers throughout the Termed the evolution of Los Angeles. Currently dubbed the 'Third L.A.' by Christopher Hawthorne, chief design officer for L.A., the city is striving to achieve a new civic identity that emphasizes on walkability and pedestrian amenities, regional mass transit and multi-family affordable residential projects.

The current Los Angeles public transportation agency, Los Angeles County Metropolitan Transportation Authority (LACMTA) or more commonly referred to as 'Metro', operates bus, light rail and subway services regionally. The Los Angeles Metro Rail, owned by the Los Angeles County Metropolitan Transportation Authority (Metro), currently has six lines, including two subway lines (B and D) and four light rail lines (A, C, L and E lines) , running across a total of 105 miles and serving a total of 92 stations (Wiki, 2021). As of 2018, the B & D lines serving Downtown Los Angeles, North Hollywood and Koreatown has the highest ridership, followed by the A-line serving Los Angeles and Long Beach, C-Line serving Redondo Beach and Norwalk, E-Line serving Downtown Los Angeles and Santa Monica and lastly the L-Line serving Azusa to East Los Angeles. Metro also owns an impressive fleet of buses, the third largest in North America as of September 2019 at 2,548 buses serving 122 bus routes across 13,283 bus stops. By 2047, Metro is expected to extend all the existing five light rail lines into the greater metropolitan region and potentially implement three to four new lines with dozens of new stations.

Data from the U.S. Census American Community Survey from years 2005 to 2017 show that driving alone remains the top transportation mode in L.A. County, increasing slightly from 76% to 79% over the years. The median household income for bus riders Is $17,154 and for train, $32,634, reflecting that public transit is crucial to residents in the lower income brackets, especially for bus riders.

Public transit ridership remains around 7% of all transportation modes despite the addition of 2 heavy rail, 4 light rail and 2 BRT lines since 1990 (Matute et al. 2016), showing that the extension of rail network did not attract new riders. Less than a quarter of residents take 84% of transit rides. Boardings for Metro dropped around 17 percent in the past five years even pre-pandemic—from just under 473 million in 2013 to around 391 million in 2018. Low ridership is associated with three main reasons: (1) lower income latino immigrant families, the predominant transit rider group in the city, having higher car ownership rates, (2) riders more dependent on network of buses than rail and yet investments from key bus lines diverted into investments for rail and (3) increasing dominance of Transit Network Companies such as Lyft and Uber which has started to eat significantly into the city's transit mode share (Sevtsuk, 2019). The decreasing public transit ridership numbers in absolute terms is concerning given the substantial financial commitment made in new transit investments.

Literature Review

Importance of Walkability

Pedestrians and walking are vital components of urban life, and are important for the social, economic, and environmental well-being of cities and their residents (Jacobs 1961; Mumford 1981; Whyte 2009; Speck 2012; Appleyard 1981; Leinberger 2007). The literature backing benefit of walkability is aplenty. It reduces the likelihood of obesity and chronic disease, improves mental health and happiness, allows for less noise and air pollution as compared to cars, and have been proven to support local businesses and promote tourism (Arup, 2016). Moderate walking has been found to benefit people across all ages, genders, and races, highlighting its importance to the public health community. Gehl (2004) generally categorized pedestrian activity to be induced by three main categories of motive: (1) necessary, which depicts activities that are carried out essential and trips that compelled to be taken; (2) optional – trip that are motivated by good urban conditions and (3) social – activities that are an occurrence of movement in spaces.

Walking is critical not only as a mode that has established benefits for health and environment, but as a gateway mode to public transit (Joh et al, 2015). It is a critical decision factor that affects one's decision to choose public transit (Murray and Wu, 2003). Every city resident is a pedestrian at some point in the day, and most public transit trips begin or end with walking (Edwards & Tsouros, 2008; Litman, 2011). American cities with larger numbers of rail and bus commuters also have more pedestrian commuters (Speck, 2012) and the relationship between transit-riders and pedestrians have been established to go hand in hand, with data reflecting more than 10% pedestrian footfall when more than a quarter of workers take transit (Freemark, 2010). The American Public Transportation Association (APTA) found in a 2017 report that more than 69% of transit users walk to their stop or station. From a planning perspective, redirecting drivers who park-and-ride to walk-and-ride would reduce number of parking lots around the station, reduce the separation of land uses and free up land for infill development (Cervero, 2001).

Much of prior studies investigating transit-associated walking times in the past were reliant on data obtained from National Household Travel Survey (Besser & Danneberg, 2005; Freeland et al, 2013, Tribby et al, 2020), a telephone survey administered by the US Department of Transportation to examine travel behavior in the United States. These surveys have been known to carry biases, often undercounting nonmotorized travel by ignoring short trips, non-work travel, travel by children, recreational travel and nonmotorized links (Litman, 2018). In this thesis, instead of using data points from stated/revealed preference surveys, big data collected regarding traffic counts, tree counts etc. is used as an estimation for pedestrian footfall.

Modeling Pedestrian Choices – Factors influencing walking

Human behavior modelling represents a complex task interspersing multiple disciplines, and the interest in studying and modelling pedestrian choices is not new. Discrete choice models (DCM) have precedented the field, where individuals' behavior can be modelled after their choice between different options presented to them (using stated or revealed preference data) by maximizing a perceived utility function (Ben-Akiva and Lerman, 1985). Factors influencing pedestrian activity level have been widely studied and summarized in the table below.

Factors	Literature	Findings
Socioeconomic factors	Hsiao, Lu, Sterlin7, & Weatherford, 1997; Loutzenheiser, 1997; Weinstein Agrawal et al., 2008; Freeland et al. 2015	Racial minorities, lower income households, blue collar neighborhoods more likely to walk
Vehicle Ownership	Hsiao et al., 1997; Weinstein Agrawal et al., 2008	Ownership has negative impact on walking, but positively to walking distance
Street Patterns	Hsiao et al., 1997; Jiang et al., 2012	Grid street pattern provides for more pedestrian linkages and increases likelihood to walk
Urban Factors/Built Environment	Hsiao et al. 1997; Loutzenheiser 1997; Zhao et al. 2003; Jiang et al., 2012; Park, Choi, & Lee, 2015); Cervero (1997)	Low traffic, higher population, land use mix and dwelling density have positive effects on walking

		3Ds: Density, Diversity and Design, with Design having the least impact
Quality of Transit Services	Kuby et al. 2004; Zhao et al. 2003	Higher number of transit lines at a stop or station

Table 1: Summary of prior literature with factors affecting pedestrian activity

The effects of design treatments, like aligning shade trees along sidewalks and siting parking lots in the rear of stores, on travel demand are thought to parallel the influences of density and diversity (Cervero & Kockelman, 1997). Under the current leadership of L.A. Mayor Eric Garcetti, L.A. is on a mission to plant 90,000 trees as part of the Green New Deal. This can reduce temperatures by up to 50% and encourage exercise and active transportation by making it more comfortable for people to walk and bicycle (Carpenter, 2021). Studying the current roles of trees in promoting walkability to public transit is thus important as L.A. advances on the Urban Trees Initiative.

Transportation and welfare studies show that without adequate transportation, welfare recipients face significant barriers in trying to move from welfare to work (Sanchez, 2008). Generally, 38% of all transit riders came from households with below $20,000 (Pucher and Renne, 2003), with significant disparities in income and ethnicity. While much attention has been given to expensive rail projects, research have also shown that the sole focus on it brings about consequential effects when addressing equity considerations. Specifically, bus riders across different cities have been found to be less wealthy as compared to rail riders, with poor and minority races traveling on buses in much greater proportions than whites (Fearnley 2006; Scauzillo 2018; Pucher et al. 2003) and this disparity is growing over time (Taylor & Morris, 2014). Metro has long been criticized for being rail-centric when allocating investment dollars, and by neglecting the improvement of bus systems, forget about most the city’s transit riders who rely on the bus. It is thus important to study the model differences between rail and bus ridership to understand the nuances of transit ridership within Los Angeles.

Nonlinear relationships affecting transit ridership

Many previous literature generally assume a linear or log-linear relationship between built environment variables and transit ridership (Ding et al., 2019) and few investigate the possibility that the built environment influences transit ridership in a non-linear manner (van Wee and Handy, 2016).

Studies that have explored the potential nonlinearity of variables and walking have utilized unordered logistic models and log-linear models (Schoner and Cao 2014; Wijk et al.2017; Heesch, Giles-Corti and Turrel 2015), but is limited precisely due to the assumptions imposed upon the models. In a more recent work, Tao et. al (2020) explored the nonlinear relationship between the built environment and active travel in the Twin Cities using a Gradient Boosting Decisions Tree and found that generally, built environment has more predictive power than demographics, and parks, proximity to downtown, and transit access have important influences. A similar study using Gradient Boosting Decisions Tree by Ding et al. (2019) also found that station-area built environment characteristics collectively contribute to 34% of the predictive power for Washington's Metrorail ridership, after controlling for transit service factors and demographics.

This paper differs by (1) exploring variables beyond built environment and (2) exploring a few different machine learning models to compare instead of only using Gradient Boosting Decisions Tree model and (3) exploring the differences between bus and train ridership.

Data and Method

Data Source

The dataset is assembled from various big datasets available online, with four different categories of variables based on prior literature:

Density of development: Includes information such as number of jobs, residents, Gross Floor Area retrieved from ESRI Business Analyst.

Demographics: Includes information such as age, gender, race, income, education level retrieved from census data as well as the Esri Tapestry Segmentation Data.

Utility value of the metro station: Calculated using the General Transit Feed System (GTFS) number of lines in each train station/number of buses passing through each station daily.

Quality of Urban design for pedestrian streets which reflects conditions to walk: Presence of ground amenities, number of trees along each walk and traffic density near each walk.

The sum of Metro's daily average boarding and alighting data around each station is used as an outcome variable.

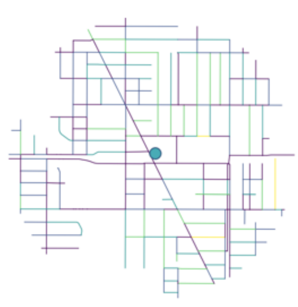
Variable Name	Description
empnum	Number of employees within 400m buffer of bus stop, obtained from ESRI business analyst
numbusstops	Number of other bus stops within 150m buffer of bus stop
numbusesthroughstop	Number of buses that pass through the bus stop, obtained from GTFS data
totalpop	Total population within 400m buffer of bus stop, obtained from Social Explorer

female_prop	Female population within 400m buffer of bus stop, obtained from Social Explorer
white_prop	White population within 400m buffer of bus stop, obtained from Social Explorer
african_am_prop	African american population within 400m buffer of bus stop, obtained from Social Explorer
college_prop	College population within 400m buffer of bus stop, obtained from Social Explorer
medianhhin	College population within 400m buffer of bus stop, obtained from Social Explorer
traffic	Traffic count within 400m buffer of bus stop, obtained from GeoHub Los Angeles
tree_count	Tree count within 400m buffer of bus stop, obtained from GeoHub Los Angeles
ground_amenities	Ground Amenities within 400m buffer of bus stop, obtained from ESRI business analyst
streetlight	Anonymized location records from smart phones collected at midpoints of streets around train stations

Table 2:Variables used in models

Standard linear regression, which uses the ordinary least squares estimator, assumes linear relationship between the dependent walking access variable and independent explanatory variables. This methodology has been extensively used to discover explanatory factors in motivating walking. For the first part of the thesis, I adopted a simple linear regression model but with two different approaches (radius vs betweenness) in computing urban design variables to compare against.

RADIUS: Average of variables in 400m radius catchment area



BETWEENNESS: Weighted urban design features based on footfall projection using factors such as number of population from origins

Linear Regression: Radius Method

A common approach in measuring spatial coverage of transit station catchment areas is the usage of a 400m standard distance as a measure of the buffer or radial service area from a bus stop (Biba, Curtin and Manca 2010; Hsiao et al. 1997; O'Neill, Ramsey and Chou 1992). A service area around a transit stop is broadly defined as the catchment area that draws ridership. Accurate understanding of the catchment size is critical in helping transit companies, planning authorities and policy makers in determining optimal stop spacing, identify redundancy and gaps at the route and system levels and understand and predict demand for transit (El-Geneidy et al., 2013).

Linear Regression: Betweenness Method

While studies have traditionally focused on a ¼ mile (400 meters) boundary limit when studying the effects of property value and transit ridership, recent studies (Ko et al, 2013; Nelson et al, 2015) have also found that transit-oriented price premium can be detected up to almost a mile beyond light-rail transit stations. This suggests that the perceived proximity to public transit is not bounded by fixed radii but driven by local urban street conditions, and a one-size-fits-all approach to define transit catchment areas is unlikely to be effective (Zhao et al., 2003). Traditional buffer methods to determine service areas have been questioned due to its inability to consider various factors pertaining to walking access. It also addresses inaccuracy derived from assuming land-use uniformity within the buffer. Originally proposed by Freeman (1977) and adopted in recent years for intra-urban traffic flow (Crucitti et al. 2006; Kazerani and Winter 2009; Sevtsuk and Mekonnen 2012; Ye, Wu et al. 2016, Cooper 2017, Sevtsuk 2021), the betweenness centrality of a node in a network is defined as the fraction of shortest paths between pairs of Origins and Destinations that pass by a particular location (UNA user guide, 2018).

$$Betweenness[i]^r = \sum_{j,k \in G-\{i\}, d[j,k] \leq r} \frac{n_{jk}[i]}{n_{jk}} \cdot W[j]$$

Where $Betweenness[i]^r$ measures the betweenness value of building *i* within the Search Radius *r*. Betweenness value of Bus Stops is measured with a 400m radius while Train Stations have a 1500m radius input. The distance decay parameter, β, considers people's willingness to walk, and as the distance they must walk to get to the station increases., their willingness to walk decreases. It is thus able to consider more accurate factors when modelling pedestrian flows, ignoring paths that will not be utilized in real life due to negative urban conditions. It is implemented via the Urban Network Analysis which takes in Census Blocks as Origins and Bus Stops or Train Stations as Destinations.

Summary of Models (Bus Stops)

Model 1: All variables against sum of boardings and alightings

Model 2: All variables against sum of boardings and alightings. Urban design variables (Traffic, Tree count and Ground Amenities) computed differently, using Betweenness Index

Model 3: Significant variables against sum of boardings and alightings

Variable	Model 1 OLS	Model 2 OLS	Model 3 OLS
Number of employees (400m)	0.0004	0.0006*	
Number of bus stops (150m)	1.9068***	2.221***	1.9320***
Number of buses passing through stop	2.5852***	2.6171***	2.6726***
Total Population (400m)	0.0257***	0.0142**	0.0240***
Female Proportion (400m)	-521.8486***		-433.3962***
White Proportion (400m)	85.1666		
African American Proportion (400m)	-99.2163	-183.4226	
College Proportion (400m)	612.4665**		752.0725***
Median Household Income (400m)	0.0008	-0.0002	
Traffic Count (400m)	-9.28E-05***	-0.0001***	-8.661e-05***
Tree Count (400m)	0.0182		
Number of Ground Amenities (400m)	0.0391**	0.0585**	0.0531***
Actual pedestrian count around station (100m)	0.0225***	0.0208***	0.0228***
Adjusted R-squared:	0.555	0.549	0.555
F-statistic:	223.2	315.3	361.4
Significance level *p<0.1, **p<0.05, ***p<0.01			
Cell entries reflect coefficients			

Table 3: Results, Linear Regression (Bus Stops)

I first tested a full model (Model 1) with all the selected variables (excluding variables which were co-linear). Models 2 & 3 had some variables dropped, and the urban design variables for Model 2 were computed via the betweenness impact. The following reports outcomes for Model 3.

Unlike other many U.S. urbanized areas, Los Angeles has no Central Business District in the traditional sense (Fraade, 2016) and it is estimated that only 2-3 percent of the regional labor force works Downtown versus 20 percent in New York and 10 percent in San Francisco (Taylor, 2016). Given that the job centers are scattered across, it is no surprise the number of employees is not a significant variable in the model.

As expected, having more bus stops within a 150m vicinity from the bus station around, as well as having a higher 'utility' value of having more buses passing through the stop would increase bus ridership. Having a higher population, college degree and having more ground amenities along the walk to the bus stop also positively contributes to bus ridership. More traffic around the paths discourages bus ridership.

Interestingly, the urban design factors computed via the Betweenness Index were insignificant, despite earlier expectations that the methodology would be more precise in estimating pedestrian footfall and in turn contribute to an improvement in the linear model. A few reasons may account for this. Firstly, the radius (400m for bus and 1500m for train) may not be the most optimal – different buffer distances can be tested and compared against. Secondly, transit riders are often formed by low-income households, but the origins considered were from all census blocks instead of only low-income housing areas. One improvement to the model can be to weight census blocks by income level.

Summary of Models (Train Stations)

A 1500m buffer around train stations was used, following studies that the transit catchment area for train stations is typically higher than bus stops.

Model 1: All variables against sum of boardings and alightings

Model 2: All variables against sum of boardings and alightings. Urban design variables (Traffic, Tree count and Ground Amenities) computed differently, using Betweenness Index

Model 3: Significant variables against sum of boardings and alightings

Variable	Model 1 OLS	Model 2 OLS	Model 3 OLS
Number of employees (1500m)	-0.0145	0.0704	
Number of bus stops (400m)	21.6973**	12.3543	25.5032***
Number of lines passing through station	1731.0321**	1708.5536***	982.5869*
Total Population (1500m)	-6.9585***	-8.8713***	-6.1616***
Female Proportion (1500m)	8.3168**	11.5992***	6.6362*
White Proportion (1500m)	3.4308*	4.319***	2.8705***
African American Proportion (1500m)	-0.7143	-0.4137	
College Proportion (1500m)	9.4135*	9.304**	7.4576**
Median Household Income (1500m)	-0.0218	-0.0535**	
Traffic Count (1500m)	-0.0004	-5.27E-05	
Tree Count (1500m)	-0.1207	35.6872	
Number of Ground Amenities (1500m)	2.3491	24.085	
Actual pedestrian count around station (100m)	0.0144*	0.0135**	0.0181***
Adjusted R-squared:	0.792	0.805	0.800
F-statistic:	15.90	17.19	30.17
Significance level *p<0.1, **p<0.05, ***p<0.01			
Cell entries reflect coefficients			

Table 4: Results: Linear Regression, Train Stations

Urban design variables (Traffic, Tree count, Number of ground amenities) are insignificant, which suggests that most people do not walk to take the train, and ridership may be contributed by driving to access trips to train stations. Interestingly, total population contributes negatively to train ridership which is counterintuitive to traditional planning theories which promotes high density to encourage high ridership. One reason may be attributed to the fact that the study looks only at the immediate surroundings of train stations, which is often built with parking spaces around.

The second simply reflects a breakdown in a typical Transit-Oriented Developments (TOD) model. The concept of TODs was first made popular in the late 1980s by urban planner Peter Calthorpe and is generally defined as “a mixed-use community that encourages people to live near transit services and to decrease their dependence on driving”. Transit-oriented development, for instance, has been shown to reduce vehicle mile travelled while increasing walking for transport and public transportation use (Cervero and

Kockelman 1997). Following its popularity around the world, the L.A. metro has adopted the concept of Transit-Oriented Developments (TOCs), touted as the improved version of TODs. Whilst the former often took the form of silo projects in the form of high-density, mixed-use buildings near transit stations, TOCs “take a more holistic view recognizing that neighborhoods surrounding transit stops are complex ecosystems that deal in physical form (buildings and infrastructure), mobility dynamics (how people get around), and finally social resiliency (community justice)” (Gensler, 2019). Successful TOCs would encourage environments that allow for residents to reach a transit station within 15 minutes. Despite the celebrated success of Transit-Oriented planning approaches, multiple studies have established the link between exacerbating house and rental prices within half-mile of transit facilities as compared to other neighborhoods that are further away (NJTOD, 2019). The deregulation of zoning and environmental laws has led to a boom in luxury housing market, especially along long commercial corridors like Ventura and Pico Boulevards that allow density bonuses (Platkin, 2020). This has also raised questions regarding equity and affordability, where low and moderate-income residents who would benefit most from proximity to transit are displaced or excluded. Luxury housing is often built near train stations, where wealthy residents who do not use the transit systems are attracted to the amenities that are often built alongside transit stations. The provision of affordable housing, especially near transit stations, is crucial in helping to achieve equitable transit-oriented communities.

Difference in results between Bus Stops and Train Stations

While the ambition to shrug away its smoggy image as a car-oriented city requires an increase in all other transportation modes (transit, walking, biking, micro-mobility), understanding the difference in ridership between bus and train is essential. Despite known benefits of improving bus services, billion-dollar rail projects have always been preferred because of political incentives.

For the bus models, proportion of female as a variable was negative while for the train models, proportion for women was positive. Studies have found consistent gender differences in participation in walking for some purposes, including for leisure (Pollard and Wagnild 2017) and may explain for the phenomenon here since amenities usually vary across transit modes.

Other Machine Learning Methods

Rapid technological advances, together with unprecedented computation powers in computers have led to the surge in popularity in machine-learning models in almost every field imaginable. This is applicable in field of Urban Planning as well, where planners and policy makers can do away with labor-intensive and costly on-site, manual data collections.

While Linear Regression is traditionally a great tool to implement for variable coefficients interpretations, as its name suggests, assumes a linear relationship. Machine Learning can provide model estimates that are more accurate, interpretable, and practical than linear regression. A careful design of hyperparameter tuning and flexible data splitting and validations is crucial to obtain reliable and stable results.

Summary of Models (Bus Stops)

I attempted 5 different models (Decision Trees, Random Forest, XGBoost, Ridge and Lasso Regression). Random Forest Model performed the best for bus stops, while XGBoost model performed the best for train stations, most likely due to a smaller sample size (n=96). The top 5 feature importance for buses were: (1) number of buses passing through stop, (2) proportion of white, (3) total population, (4) proportion of female and (5) number of other bus stops within 150m vicinity, while for trains (1) number of ground amenities, (2) traffic count, (3) total population, (4) median household income and (5) number of bus stops within 150m vicinity.

Partial dependence plots demonstrate the functional shape of effects and how each feature is related to model predictions. It helps to explore non-linear relationships between each variable and transit ridership while considering the effects of other variables in the model.

Partial dependence function:

$$\hat{f}_{x_S}(x_S) = E_{x_C}\left[\hat{f}(x_S, x_C)\right] = \int \hat{f}(x_S, x_C)\, d\mathbb{P}(x_C)$$

where x_S are the features for which the partial dependence function should be plotted and x_C are the other features used in the machine learning model $\hat{f}$.

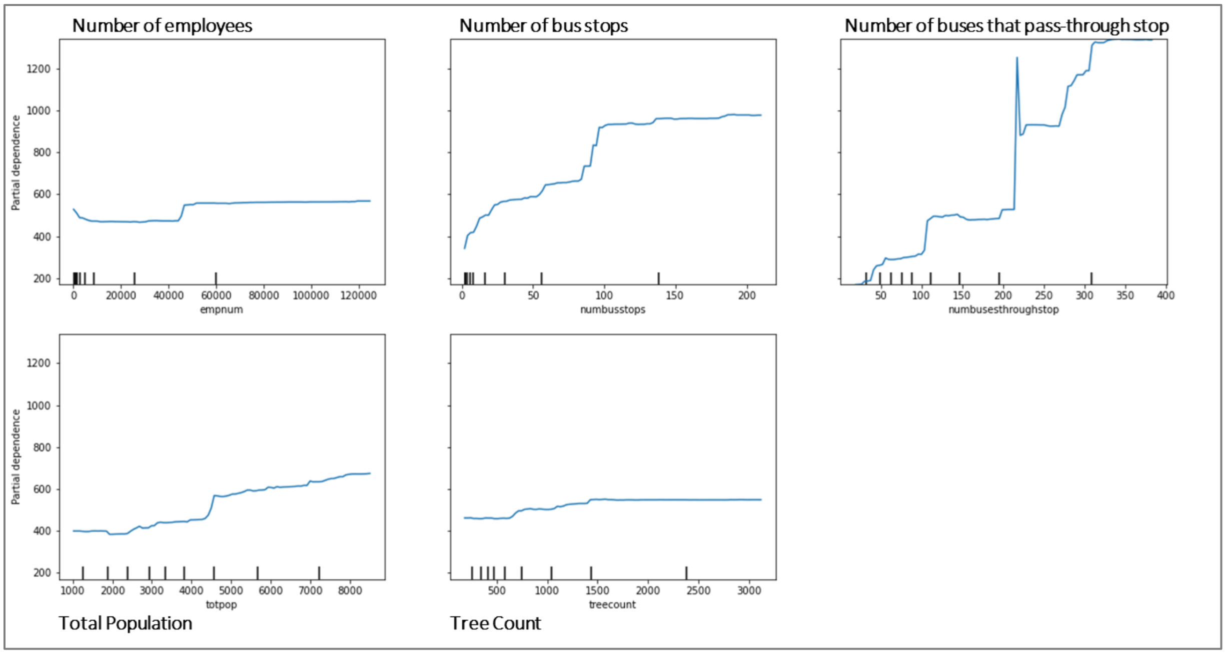


Figure 1: Partial dependence plots (Bus Stops)

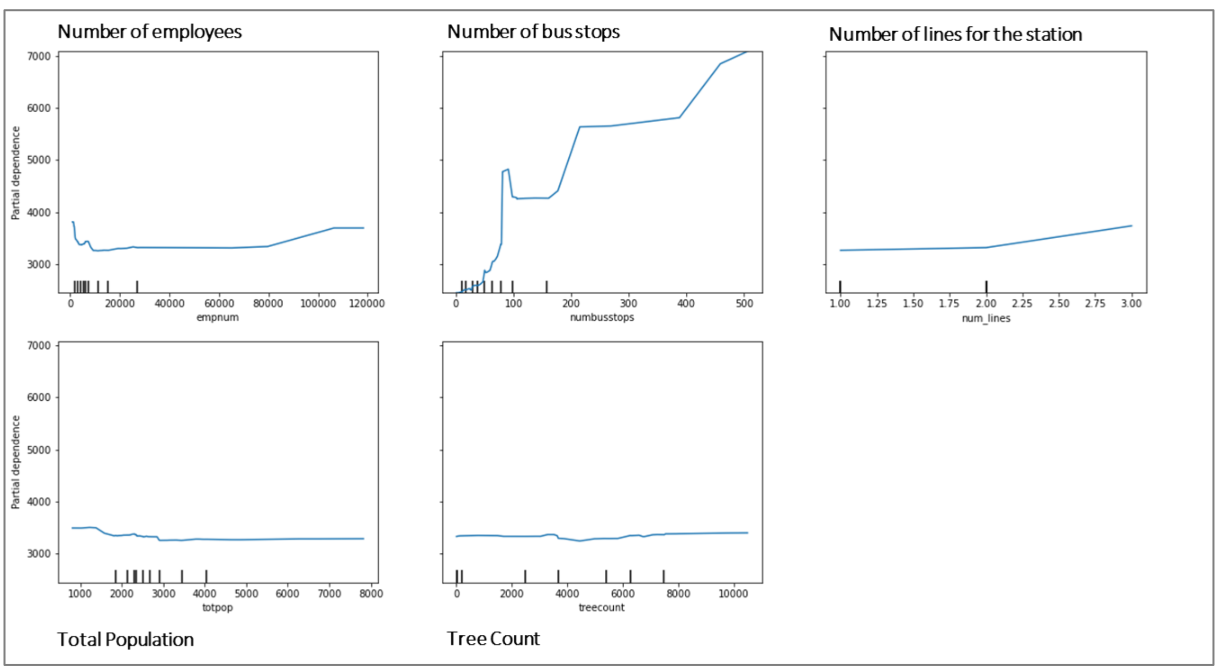


Figure 2: Partial dependence plots (Train Stations)

The number of bus stops in the vicinity has a positive association with both bus and train ridership. As seen from the plots, total population and tree count has flipped relationships for bus stops and train stations. Similar to the linear regression model, total population is negatively associated with train ridership, which is a unique scenario that is not often replicated in other cities. Prior literature often report that as population density increases, active travel increases as well (Ewing and Cervero 2010; Tao 2010). Both variables are associated positively with bus stops, yet negatively associated with train stations. Number of employees is associated positively for bus stops, and only positive for train stations beyond a certain threshold at around 80,000 employees.

Limitations and Conclusion

The models do not consider the intrinsic value of public transportation itself, such as pricing and rebate strategies. The variable inputs are also limited by data availability, such as sidewalk widths and weather specific data. Lastly, missing data points are imputed with zero values which may have skewed some findings.

Variables influencing walking to transit have been widely studied, and the results for linear regression in this study have been generally expected. Machine learning models, however, can reflect relationships that are non-linear or have a stepwise like relationship which would help planners transform into more actionable policies. For example, while trees have been known to generally contribute to positive walking experience and in turn contribute to ridership. However, as seen in the partial dependence plot for Tree Count for bus stops, there is an inflexion point at around 1,500 trees within a 400m buffer of bus stops which informs that constant tree planting beyond 1,500 trees may not necessarily be the most effective in inducing walking to ride buses. It allows planners to figure out the threshold for each increase in variable and weigh it against development costs.

Secondly, street level variables did not play a significant role in the outcomes, and further study may be needed to ascertain the impact of these variables and walking to take transit. Lastly, results have shown that some variables have an inverse relationship for bus and train ridership, which would raise questions about investment strategies. Rail transit does have its benefits, but Metro needs to strike a better balance between building rail and serving the majority of the city's transit riders, who rely on the bus and are disproportionately low-income minorities (Fraade, 2016).

References

Amy L. Freeland, Shailendra N. Banerjee, Andrew L. Dannenberg, and Arthur M. Wendel, 2013: Walking Associated With Public Transit: Moving Toward Increased Physical Activity in the United States American Journal of Public Health 103, 536_542, https://doi.org/10.2105/AJPH.2012.300912.

Besser, L. M., & Dannenberg, A. L. (2005). Walking to public transit: steps to help meet physical activity recommendations. American journal of preventive medicine, 29(4), 273-280.

Ben-Akiva, M., & Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand.

Biba, S., Curtin, K. M., & Manca, G. (2010). A new method for determining the population with walking access to transit. International Journal of Geographical Information Science, 24(3), 347-364.

Bottles, S. (1987). The Power of Consensus. In Los Angeles and the Automobile: The Making of the Modern City (pp. 92-121). University of California Press. Retrieved January 5, 2021, from http://www.jstor.org/stable/10.1525/j.ctt1pp0kb.7.

Broach, J. 2015. "Where do people prefer to walk? A pedestrian route choice model developed from GPS data." Accessed April 6, 2018. https://activelivingresearch.org/where-do-people-prefer-walk-pedestrian-route-choice-model-developed-gps-data.

Brian D. Taylor, "Congested Development? Rethinking the Causes and Consequences of Congestion" (lecture presented at Urban Planning 255, University of California Los Angeles, April 4-6, 2016).

Brodsly, David. L.A. Freeway: An Appreciative Essay. Berkeley: University of California, 1981.

Calvin P. Tribby, Barry I. Graubard, David Berrigan, National and metropolitan trends in public transit use, transit-related walking, and ridesharing between 2009 and 2017, Journal of Transport & Health, Volume 19, 2020, 100918, ISSN 2214-1405, https://doi.org/10.1016/j.jth.2020.100918.

Cervero, R. 2001. Walk-and-ride: Factors influencing pedestrian access to transit. Journal of Public Transportation, 3(4): 1–23.

Cervero, R., and Kockelman, K. (1997). "Travel demand and the 3Ds: Density, diversity, and design." Trans. Res. Part D, 2(3), 199–219.

Cooper, C. H. (2017). Using spatial network analysis to model pedal cycle flows, risk and mode choice. Journal of transport geography, 58, 157-165.

Crucitti, P., Latora, V., & Porta, S. (2006). Centrality measures in spatial networks of urban streets. Physical Review E, 73(3), 036125.

Ding, C., Cao, X., & Liu, C. (2019). How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds. Journal of Transport Geography, 77, 70-78.

Edwards P, Tsouros AD. A Healthy City is an Active City: A physical activity planning guide. Copenhagen, Denmark: World Health Organization Europe; 2008. Available at: http://www.euro.who.int/InformationSources/Publications/Catalogue/20081103_1.

El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., & Surprenant43 Legault, J. (2014). New evidence on walking distances to transit stops: Identifying redundancies and gaps 44 using variable service areas. Transportation, 41(1), 193-210.

Fearnley, N. (2006). Public transport subsidies in the UK: evidence of distributional effects. World Transport Policy & Practice, 12(1), 30-39.

Freeland, A. L., Banerjee, S. N., Dannenberg, A. L., & Wendel, A. M. (2013). Walking associated with public transit: moving toward increased physical activity in the United States. American journal of public health, 103(3), 536-542.

Freemark Y. Transit mode share trends looking steady; rail appears to encourage non-automobile commutes. The Transport Politic. October 13, 2010.

Gan, A. (2003). Forecasting Transit Walk Accessibility: Regression Model Alternative to Buffer Method. Transportation Research Record, 1835(1), 34–41. https://doi.org/10.3141/1835-05.

Heesch, K. C., Giles-Corti, B., & Turrell, G. (2015). Cycling for transport and recreation: associations with the socio-economic, natural and built environment. Health & place, 36, 152-161.

Hsiao, S., Lu, J., Sterling, J., & Weatherford, M. (1997). Use of geographic information system for analysis of transit pedestrian access. Transportation Research Record, 1604(1), 50-59.

Jacobs, J. (1961). Jane jacobs. The Death and Life of Great American Cities.

Joh, Kenneth & Chakrabarti, Sandip & Boarnet, Marlon & Woo, Ayoung. (2015). The Walking Renaissance: A Longitudinal Analysis of Walking Travel in the Greater Los Angeles Area, USA. Sustainability. 7. 8985-9011. 10.3390/su7078985.

Kazerani, A., & Winter, S. (2009, June). Can betweenness centrality explain traffic flow. In 12th AGILE international conference on geographic information science (pp. 1-9).

Ko, Kate & Cao, Jason. (2013). The Impact of Hiawatha Light Rail on Commercial and Industrial Property Values in Minneapolis. Journal of Public Transportation. 16. 47-66. 10.5038/2375-0901.16.1.3.

Lilah M. Besser, Andrew L. Dannenberg, Walking to Public Transit: Steps to Help Meet Physical Activity Recommendations, American Journal of Preventive Medicine,Volume 29, Issue 4, 2005, Pages 273-280, ISSN 0749-3797, https://doi.org/10.1016/j.amepre.2005.06.010.

Litman T. Evaluating public transportation health benefits. American Public Transportation Association Website. Available at: http://www.vtpi.org/tran_health.pdf. Updated June 8, 2011

Manville, M. (2018). Measure M and the potential transformation of mobility in Los Angeles.

Matute, J., Bains, J., Fraade, J., Gahbauer, J., Lu, R., Pinski, M., ... & Wickland, T. (2017). California Statewide Transit Strategic Plan: Recommendations Report.

Murray, A. T., & Wu, X. (2003). Accessibility tradeoffs in public transit planning. Journal of Geographical Systems, 5(1), 93-107.

Nelson, A.C., Eskic, D., Hamidi, S., Petheram, S.J., Ewing, R., & Liu, J.H. (2015). Office Rent Premiums with Respect to Light Rail Transit Stations: Case Study of Dallas, Texas, with Implications for Planning of Transit-Oriented Development. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2500, pp. 110-115.

Pucher, J., & Renne, J. L. (2003). Socioeconomics of urban travel. Evidence from the 2001 NHTS.

Rasmussen, C. (2003). LA then and now: Pasadena's Gold Line will travel a history laden route'. Los Angeles Times: New York.

Sanchez, T. W. (2008). Poverty, policy, and public transportation. Transportation Research Part A: Policy and Practice, 42(5), 833–841. https://doi.org/10.1016/j.tra.2008.01.011.

Sevtsuk, A., & Mekonnen, M. (2012). Urban network analysis. Revue internationale de géomatique–n, 287, 305.

Sevtsuk, A. (2021). Estimating pedestrian flows on street networks: Revisiting the betweenness index. Journal of the American Planning Association, 1-15.

Speck, J. (2013). Walkable city: How downtown can save America, one step at a time. Macmillan.

Tao, T., Wu, X., Cao, J., Fan, Y., Das, K., & Ramaswami, A. (2020). Exploring the nonlinear relationship between the built environment and active travel in the twin cities. Journal of Planning Education and Research, 0739456X20915765.

Taylor, B. D., & Morris, E. A. (2015). Public transportation objectives and rider demographics: are transit's priorities poor public policy?. Transportation, 42(2), 347-367.

Tribby, C. P., Graubard, B. I., & Berrigan, D. (2020). National and metropolitan trends in public transit use, transit-related walking, and ridesharing between 2009 and 2017. Journal of Transport & Health, 19, 100918.

Wachs, M. (1993). Learning from Los Angeles: transport, urban form, and air quality. Transportation, 20(4), 329-354.

Wachs, M. (1984). Autos, transit, and the sprawl of Los Angeles: The 1920s. Journal of the American Planning Association, 50(3), 297-310.

Ye, P., Wu, B., & Fan, W. (2016). Modified betweenness-based measure for traffic flow prediction of urban road. Transport Reseearch Record.

Zhao, F., Chow, L.-F., Li, M.-T., Ubaka, I., & Gan, A. (2003). Forecasting transit walk accessibility: Regression model alternative to buffer method. Transportation Research Record: Journal of the Transportation Research Board, 1835(1), 34–41. doi: 10.3141/1835-05.