New Open-Source Analyses of Transit Job Access and Transit Ridership

April 2022

A Research Report from the National Center for Sustainable Transportation

Marlon G. Boarnet, University of Southern California
David Flores Moctezuma, University of Southern California
James Gross, University of Southern California

National Center for Sustainable Transportation

METRANS Transportation Consortium USC | CSULB

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**
NCST-USC-RR-22-15	N/A	N/A

4. Title and Subtitle New Open-Source Analyses of Transit Job Access and Transit Ridership	**5. Report Date** April 2022 **6. Performing Organization Code** N/A
7. Author(s) Marlon G. Boarnet, Ph.D., https://orcid.org/0000-0002-0890-347X David Flores Moctezuma, https://orcid.org/0000-0002-4246-8459 James Gross, https://orcid.org/0000-0002-8405-7340	**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address University of Southern California METRANS Transportation Center University Park Campus, VKC 367 MC:0626 Los Angeles, California 90089-0626	**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747114
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590	**13. Type of Report and Period Covered** Final Research Report (August 2020 – December 2021) **14. Sponsoring Agency Code** USDOT OST-R

15. Supplementary Notes

DOI: https://doi.org/10.7922/G2862DSW
Dataset DOI: https://doi.org/10.7910/DVN/S2WRF2

16. Abstract

This research project examines the link between job access and stop/station level transit ridership. Job access, following recent literature, is measured as the number of jobs that can be reached within a 30-minute transit travel time, including transfers and walk time to access jobs once exiting a transit station. Cumulative opportunity job access measures of this sort—i.e., the number of jobs that can be reached within 30 minutes—have become common in the recent access literature, and those measures have often focused on access via transit. Yet there have been few studies that examine the link between transit job access and transit ridership, and of those none that examine the link at a station or stop level. This study uses station and stop level ridership data for the Los Angeles Metro bus and rail system and the BART rail system in the San Francisco Bay Area. The research team calculated transit job access as jobs that can be reached within 30 minutes, using the Remix software tool. Regression analysis of 1,000 randomly selected Los Angeles bus stops reveals a robust relationship between stop-level ridership and job access. The association between transit job access and bus stop ridership (embarkations and disembarkations at the stop) is statistically significant. Converting that association into an elasticity, if the number of jobs accessible within 30-minutes were to increase by 1 percent, on average stop-level ridership would increase between 0.6 to 0.8 percent. The same association, with similar magnitudes, exists for Metro rail stations and BART rail stations, but due the smaller sample sizes, those relationships are not statistically significant when control variables are added to the regression. The findings show that job access is closely related to ridership at the bus stop level, suggesting transit agencies can increase job access by increasing bus frequency, reducing transfers, siting lines that connect job concentrations to residents, and by improving bus stop/rail station access/egress times.

17. Key Words	**18. Distribution Statement**
Accessibility, Job Access, Employment, Transit Ridership, Boarding & Alighting, Travel Time, Public Transit Networks	No restrictions.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	50	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

About the National Center for Sustainable Transportation

The National Center for Sustainable Transportation is a consortium of leading universities committed to advancing an environmentally sustainable transportation system through cutting-edge research, direct policy engagement, and education of our future leaders. Consortium members include: University of California, Davis; University of California, Riverside; University of Southern California; California State University, Long Beach; Georgia Institute of Technology; and University of Vermont. More information can be found at: ncst.ucdavis.edu.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Acknowledgments

This study was funded, partially or entirely, by a grant from the National Center for Sustainable Transportation (NCST), supported by the U.S. Department of Transportation (USDOT) through the University Transportation Centers program. The authors would like to thank the NCST and the USDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project. The authors are grateful to Los Angeles Metro, which assisted by providing data on bus stop level boarding and alighting, and to Remix, Inc. (now a subsidiary of Via) for their assistance with the Remix software. All analysis, conclusions, and interpretations in this report are the responsibility of the authors alone.

NCST

New Open-Source Analyses of Transit Job Access and Transit Ridership

A National Center for Sustainable Transportation Research Report

April 2022

Marlon G. Boarnet, David Flores Moctezuma, and James Gross

Sol Price School of Public Policy, University of Southern California

[page intentionally left blank]

TABLE OF CONTENTS

List of Tables

List of Figures

New Open-Source Analyses of Transit Job Access and Transit Ridership

EXECUTIVE SUMMARY

The objective of this research project is to use regression analysis to explain transit stop on/off (boarding and alighting) as a function job access. We utilize Remix, a commercially available software tool, to obtain 30-minute job accessibility isochrones to determine the number of jobs that can be reached from transit stations/stops within 30 minutes during the weekday peak hour time of 8:00am. We proceed to test regression models to determine how average transit ridership at the station or stop level is associated with the number of jobs that are accessible within a 30-minute public transit trip (inclusive of walk time to/from stations or stops, and transfer time).

The previous literature focused on the relationship between transit system characteristics and ridership. Often, transit ridership studies focused on the operator's perspective. The primarily focus was fare levels, service quality, and frequency. More recently, there has been growing interest in the role of job accessibility as an organizing concept for transportation planning (Cui and Levinson 2019). We use a cumulative opportunities measure of job access—counts of the number of jobs that can be reached in 30 minutes via transit. This project's primary contribution is the spatial fidelity of job access at specific public transit stations to examine the relationship between transit ridership and job accessibility.

Our job accessibility analysis focuses on two public transit systems: the San Francisco Bay Area Rapid District (BART) and the Los Angeles County Metropolitan Transportation Authority (L.A. Metro). Daily boardings and alightings, averaged over the month of October 2019, were analyzed for the 13,962 bus stops and 96 light rail stations that are part of the LA Metro system, and the 48 stations that are part of the BART system. The job distribution within the County of Los Angeles and the five counties within which BART operates was measured using Remix, a transportation software tool.

In an initial step, we combined bus stops that are located within 60 meters of each other, as is the case when multiple bus stops converge on opposite corners of a street intersection. We then summed the average daily boarding and alighting data for all stops that were aggregated into a single combined stop. We used bus stop locations (or the centroid of the area created by bus stops within 60-meters of each other that were combined into a single stop) as our initial departure point from which a 30-minute transit travel time isochrone was created. We stratified the resulting 8,036 collapsed bus stops in the LA Metro system into quintiles based on boardings and alightings, and selected 200 stops randomly from within each quintile, for a stratified random sample of 1,000 bus stops. For the analysis of rail stations for both LA Metro and BART stations, no stations were combined, as each of those stations was one distinct location per station.

We used regression analysis to examine the association between station and stop level ridership (boardings plus alightings) and 30-minute transit job access, analyzing each rail system and the bus system separately.

The results suggest that our main regressor of interest, the number of jobs accessible via transit within a 30-minute trip, is positive and statistically significant in all regression models for bus stops—with and without other control variables. The bivariate relationship between job access and ridership is also statistically significant for rail stations, however, when adding additional controls to the LA Metro light rail and BART models, the results are no longer statistically significant. This can be explained in part by the fact that the L.A. Metro has 96 stations and BART has 48 stations with available data, reducing the sample size for the rail station analysis. For the LA Metro bus system, we find that a one-percent increase in job accessibility within 30-minutes is associated with, on average, stop-level ridership increases between 0.6 and 0.7 percent. The bivariate relationship for the rail systems show that a one percent increase in job accessibility is associated with station level ridership increases between 0.6 and 0.8 percent.

Our finding that job access is closely related to ridership at the bus stop and (in the bivariate relationship) station level is intuitive. However, to our knowledge, it has not been demonstrated in the literature. We acknowledge that changing the pattern of job locations and bus routes and stops, while possible, requires many factors. Still, route alignment to well connect residences to jobs is a vital part of transit system planning, as indicated by our regression analysis. Additionally, transit agencies should focus on other factors that increase stop/station level job-access, such as frequency, reducing transfers and transfer wait times, and improving access/egress times to/from stations and bus stops.

Introduction

Accessibility has become a popular concept in transportation planning (see, e.g., Wu, Levinson, and Owen, 2021). Intuitively, one would expect that the ridership on transit systems would depend on the accessibility provided by that system. In some cases, that concept has been studied for entire systems or transit lines. Yet, to our knowledge, no study has examined the link between transit job access and ridership at the stop or station level, despite the fact that the stop or station level is likely the level where an access-ridership relationship would be most evident and most important. We close that gap by studying the relationship between job access and ridership, at the stop and station level, for rail and bus stops in the Los Angeles Metro system and for the BART rail system. Our results confirm the hypothesized relationship between access and ridership, with elasticities that imply, for the Los Angeles Metro bus system, that a one percentage point increase in job access is associated with between 0.6 and 0.7 percentage point increases in bus stop ridership.

Literature Review

Transit Ridership

Prior literature overwhelmingly focuses on the relationship between transit system characteristics and ridership. The emphasis is typically on high level aggregations (i.e., city level or transit systems). Job accessibility is rarely examined as an explanatory variable. The work by Taylor et al. (2009) is one of the most well-known examples. Taylor's study conducted a cross-sectional analysis of transit use in 265 urbanized areas in the United States. The study tested multiple variables measuring regional geography, metropolitan economy, population characteristics, auto/highway system characteristics, and transit characteristics (Taylor, et al. 2009). Taylor's study finds that transit policies are statistically significant determinants of the overall level of transit use in an urbanized area. The observed range in both fares and service frequency accounts for almost doubling or halving transit ridership in a given urban area (Taylor, et al. 2009).

Handy et al. (2013) summarized several studies that examined the association between transit ridership and frequency, system expansion, and fares, virtually always at the system or city level. Other studies focus on finding the effects of fares, quality of services and income and car ownership (e.g., Paulley, et al. 2006). Specifically, Paulley et al. (2006) present fare elasticities to determine the ratio of the proportional change in patronage to the proportional change in fares. Currie and Delbosc (2011), is an exception to the literature that focuses on relationships among transit system average values. Their analysis focuses on bus rapid transit (BRT) systems in Australia. The analysis explores whether BRT design features increase ridership above and beyond the impact of service levels, using data for routes. The study utilizes data on 77 BRT and non-BRT bus routes in Australia. The research suggests that overall, some BRT infrastructure treatments have a significant impact on ridership. This includes factors such as right of way. (Currie and Delbosc 2011).

Overall, the literature on transit ridership has focused on fare levels, service quality, and (to a lesser extent) infrastructure. The relationship between accessibility and transit ridership, specifically accessibility to jobs, has been relatively under-emphasized. That is in part due to the fact that the literature on accessibility, while decades old, has only recently become popular, due in large part to analytical tools that allow more quick measurement of the accessibility provided by transportation systems.

Accessibility

Accessibility is defined as the potential for or ability to reach valued destinations (Hansen 1959). Hansen developed a basic use-based method. This method counts how many opportunities can be reached in a given time cost. A large literature has expanded on and implemented that concept, calculating how many opportunities (most commonly jobs) can be reached in a given travel time (see, e.g., Boer, et al. 2018, Cui and Levinson 2018a, Cui and Levinson 2018b, Deboosere, El-Geneidy and Levinson 2018, O'Kelly and Horner 2003, Srour, Kockelman and Dunn 2002, Wachs and Kumagai 1973). In contrast, Cui and Levinson (2019) identify a dual measure which is the travel cost of accessing a fixed number of opportunities (Cui and Levinson 2019). For our purposes, which is a general measure of accessibility, what Cui and Levinson (2019) call the "primal" method—the number of opportunities that can be reached in a given travel time—is useful.

We ask how the transit system provides access to jobs. This more classic (or "primal" per Cui and Levinson) measure of accessibility is a function of the spatial distribution of land uses and the characteristics of the transport network, such as transit routes and frequencies. (See, e.g., Paez, Scott, and Morency, 2012 for a conceptual discussion). System characteristics at the bus route and rail transit line level are associated both with transit job accessibility and ridership.

Recently, there has been a growing interest in the role of job accessibility as a central organizing concept for transportation planning (e.g., Cui and Levinson 2019). Boarnet et al., (2017) examined how to construct different measures of job access via transit using transit agency data, and from general transit feed system (GTFS) data (Boarnet, et al. 2017, Painter, et al. 2019).

Wu, Levinson and Owen (2021) examine the relationship between transit mode share and accessibility within 35 minute travel times, and they found that their job access measure explained much of the variation in transit mode share across 48 major metropolitan areas. Merlin and Hu (2017) examine four accessibility measures in the city of Los Angeles, two of which are more common "cumulative opportunities" measures based on the number of jobs that can be reached, and two of which are "competitive" measures that account for the competing labor supply that can reach the jobs. Merlin and Hu (2017) conclude that the competitive measures, which account for competing labor supply in a manner first popularized by Shen (1998), are more associated with employment outcomes. Merlin and Hu (2017) recommend using the competitive measures, due to their stronger association with employment outcomes. In this study, however, we use the simpler cumulative opportunity measure—counts of jobs which can be reached within a given travel time—because our focus is not on labor market outcomes but on the overall accessibility provided by the transit system. Our question is how many destinations one can reach via transit, and how that relates to transit ridership, and hence cumulative opportunities are appropriate.

The most common challenge in evaluating transit accessibility has been calculating travel times. Prior to 2005 and the introduction of the General Transit Feed Specification (GTFS), detailed transit schedules were not widely available (Owen and Levinson 2015). Since then, the popularity of the GTFS has somewhat revolutionized the use of access, allowing quick calculations of transit travel times based on route maps and frequencies for transit systems in many countries, and increasing the use of cumulative opportunity measures of transit access.

Painter, Boarnet and Swayne (2019), develop a job accessibility measure using GTFS data and Remix, a commercially available software tool. The results suggest that greater job accessibility by transit within 15, 30, and 45-minute travel times increased the likelihood of being part of the labor force (Painter, et al. 2019). Following the approach used by Painter, Boarnet, and Swayne (2019), the primary contribution of this study is the spatial fidelity of job access at specific bus transit stations and the relationship between average transit ridership and job accessibility (the number of jobs accessible within a 30-minute transit ride).

Returning to the literature on transit ridership, accessibility is a mix of what have sometimes been called internal and external factors. Transit agencies have a set of internal factors which can influence ridership (e.g., frequency, fares) and external factors (e.g., density levels, congestion on parallel automobile routes.) For a discussion of this distinction, see, e.g., Taylor and Fink (2003) and Alam, Nixon, and Zhang (2018). We will return to this point in the conclusion of the report.

Data Sources

The L.A. Metro Service Planning, Scheduling, and Analysis team provided data regarding Los Angeles Metro bus and rail boarding and alighting. The dataset covered daily boarding and alighting, averaged over the month of October 2019. This included weekday, Saturday and Sunday activity for all 14,058 bus stops and rail stations within the LA metro system. Of those 14,058 stops, 96 stops were rail stations, and 13,962 stops were individual bus stops. The San Francisco Bay Area Rapid Transit District (BART) provided monthly ridership reports through their online portal. The dataset covered actual daily boarding and alighting ridership. The BART data included average ridership for the month of October 2019 weekday, Saturday, and Sunday for 48 stations that were part of the BART transit system at the time (the Silicon Valley BART extension with two additional stations became operational in June, 2020). The Metro and BART systems are displayed in Figure 1.

For both L.A. Metro and BART, the average daily totals for the month of October 2019 included boarding (on) and alighting (off) at each station or stop. For L.A. Metro busses, this included the weekday, Saturday, and Sunday time periods split by each individual bus line that stopped at each designated bus stop. The latitude and longitude for each stop was available for each bus stop and rail station. The following information was also contained within the dataset: an individual LA Metro stop identifier (stop_num); the address location; and the number of lines that serve that stop. The format for L.A. Metro rail stations was similar. BART monthly files contain ridership counts in the format of what BART calls “entry-exit” matrices. This included the type of day (weekday, Saturday, and Sunday) in a matrix format that provided the boarding station and the alighting station.

The IPUMS National Historical Geographical Information System (NHGIS) provided demographic variables for 2019 (Manson, et al. 2021). Each dataset was joined to the County of Los Angeles census block group boundary shapefile. The variables of interest for this study are the percentage of the population that is African American; the percentage Asian; and the percentage of the population that is Hispanic (as per census definitions); the percent of the population comprised of persons who were born outside of the United States; the percentage of zero-vehicle households in the area; percentage of population with income below the poverty line; population density in number of persons per square mile; and percentage of population under eighteen years of age. In the next section the process by which each of the variables of interest was extracted will be explained in further detail. (Also please refer to Appendix B).

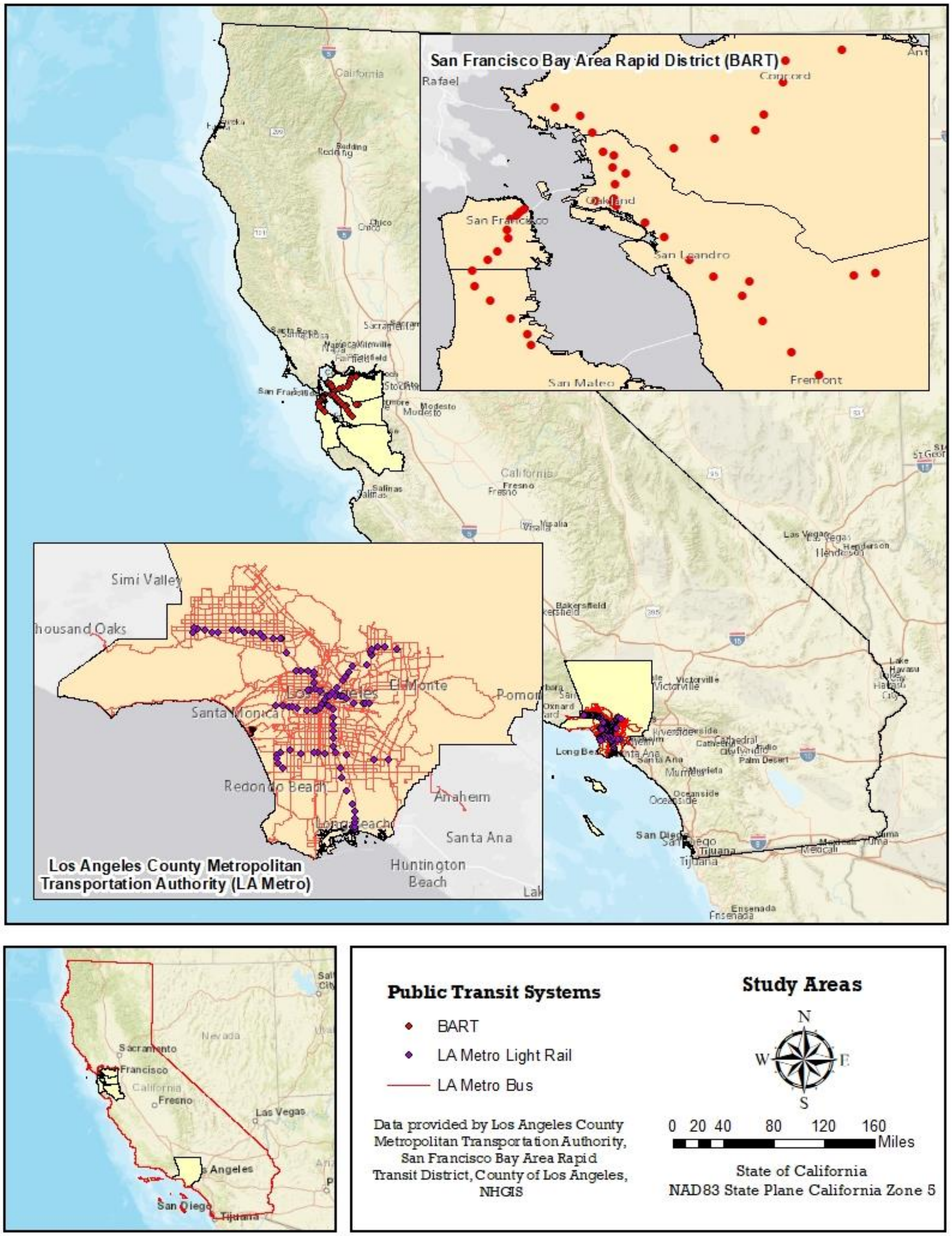

Figure 1. Public transit systems: Los Angeles County Metropolitan Transportation Authority (L.A. Metro) and San Francisco Bay Area Rapid Transit District (BART)

Methodology

Our objective is to use regression analysis to explain transit stop on/off (boarding and alighting) as a function of measures of job access. The dependent variable is the daily average of the transit stop ridership (boarding and alighting) for October 2019. The key independent variable of interest is a measure of the number of jobs that can be reached in a 30-minute transit commute (inclusive of walk time to/from stations and wait and transfer time). We use data for Los Angeles Metro bus and rail systems and the BART rail system. The method used to analyze the L.A. Metro bus system requires more explanation.

There are 96 rail stations and 13,962 bus stops (without rail) in the Los Angeles Metro service area. The Los Angeles Metro bus system is shown in Figure 2. The 13,962 bus stops often include stops diagonally across on opposite sides of the street, serving the same line in opposite directions. At times, there are other nearby stops. These include multiple stops on the same side of the street. Our criteria for identifying bus stops is to identify stops that are directly on opposite sides of the street from one another within a 60-meter threshold as one single stop. In cases where stops are within 60 meters of each other, the midpoint between both stops diagonally across or directly across from each other is considered the centroid for that set of stops. The average daily on/off boarding/alighting for each stop is aggregated and the middle distance point between the stops (centroid) is positioned as the collapsed stop. Bus stops within close proximity (specifically, within 60 meters) of each other are combined into one collapsed stop. For rail transit, the data from L.A. Metro and BART identified stations that were always clearly one station, rather than multiple entrances to the same station, and so we did not develop collapsed stops for rail transit stations.

In some cases, multiple bus stops can converge on opposite corners of a street intersection (within the 60 meters from one another). In this scenario, multiple stops would create a polygon. The average daily on/off boarding/alighting data for all stops was aggregated and the centroid served as the collapsed bus stop. We developed the 60-meter threshold for combining bus stops after detailed examination of the geographic pattern of bus stops. Larger distance thresholds (> 60 meters) would in some cases chain distinct bus stops along a street into one large collapsed stop spanning more than a block. We found that the 60-meter threshold identifies stops nearby or on opposite sides of a street that are, intuitively, one bus stop.

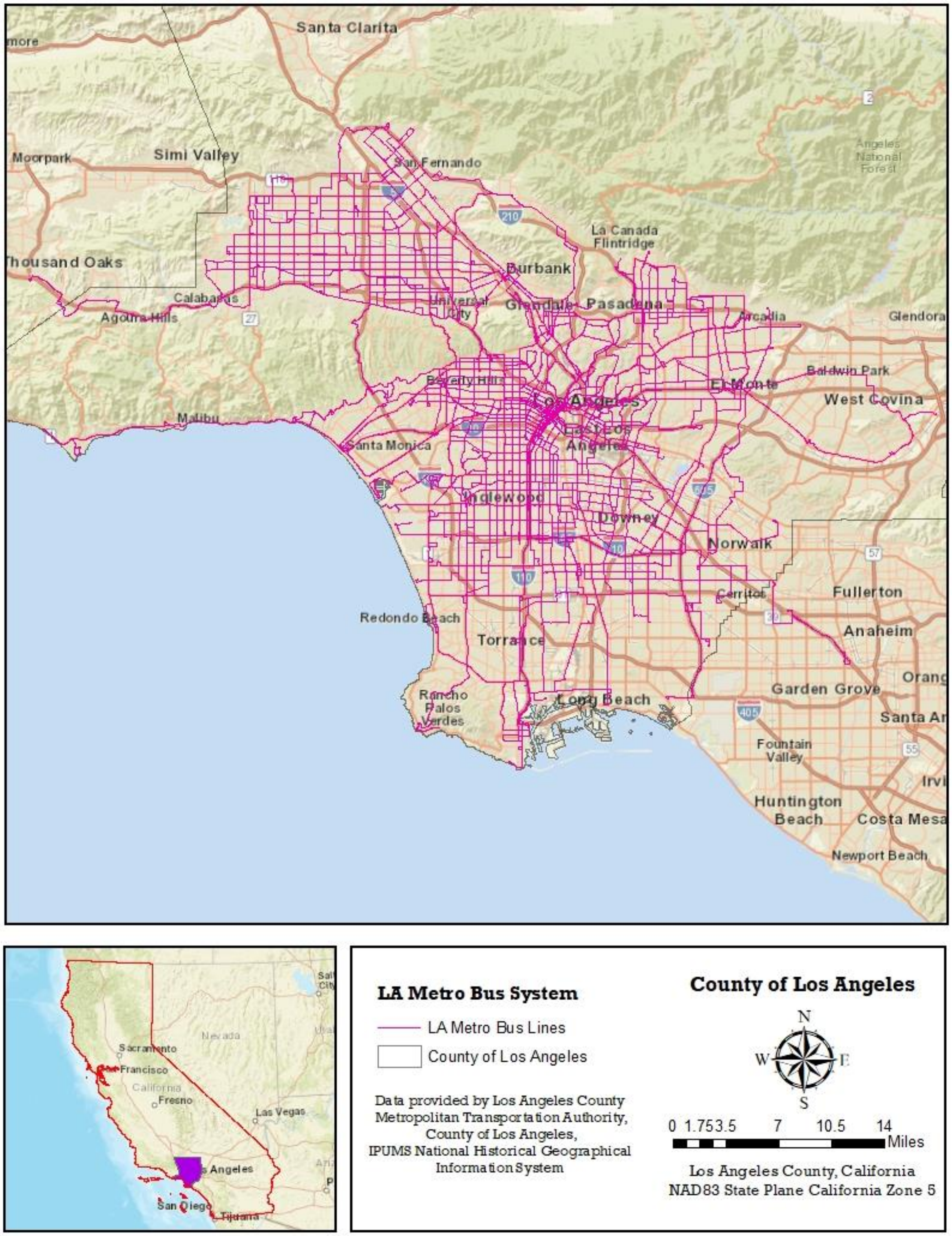

Figure 2. Map of the L.A. Metro bus system

Collapsed stops were formed with a GIS process that was initiated utilizing ArcMap 10.8. First a 60-meter buffer area was created around each individual bus stop. The result created a series of converging patterns, all within the same parameter. To avoid any sequential bonding of bus stops that would meet the required 60-meter threshold from one of the individual bus stops within a polygon region, any converging space was removed through a GIS extracting process (Please refer to Appendix A). This process eliminated the possibility that a bus stop would be part of two different collapsed stops. This is illustrated in Figure 3. The red boundaries in Figure 3 show the union of 60-meter boundaries around stops, with stops shown as black dots.

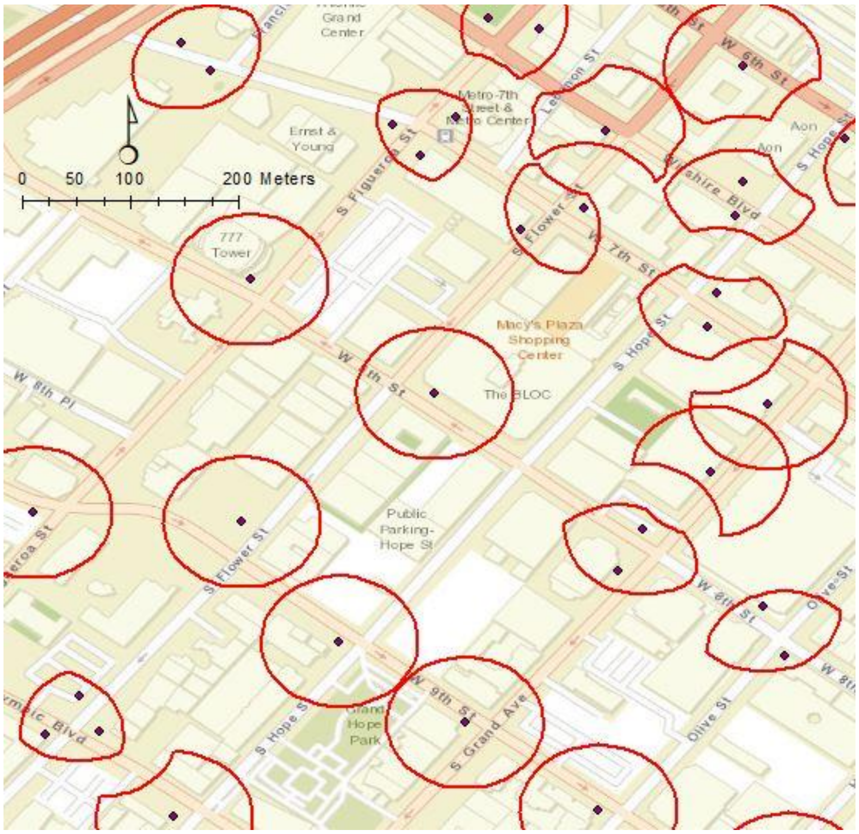

Figure 3. Example of collapsed bus stop in Downtown Los Angeles. Circles around stops are 60 meters. Collapsed stops are indicated with areas that are the union of 60-meter radius circles around individual bus stops

Table 1 shows descriptive statistics for ridership (the sum of boarding and alightings), for all 13,962 L.A. Metro bus stops, the 96 L.A. Metro rail stations, the 8,036 L.A. Metro collapsed bus stops, and the 48 BART rail stations.

Table 1. Descriptive statistics of L.A. Metro –all bus stops; L.A. Metro light rail system; L.A. Metro Collapsed bus stops; and BART rail stations

L.A. Metro - All Bus Stops

	Weekday	Saturday	Sunday
Mean daily (boarding plus alighting)	132.60	92.65	73.38
Minimum	0	0	0
25th percentile	14	9	7
Median	42	30	23
75th percentile	126	86	69
Maximum	12,223	7,181	5,798
Number of observations (stops)	13,962	13,962	13,962 [1]

L.A. Metro - Rail Stations

	Weekday	Saturday	Sunday
Mean daily (boarding plus alighting)	7,138.50	4,039.82	3,542.13
Minimum	888	317	244
25th percentile	3,206	1,619.75	1,444.5
Median	4,389	2,387	2,041
75th percentile	6,655	3,719	3,350
Maximum	56,073	28,981	25,019
Number of observations (stations)	96	96	96

L.A. Metro Collapsed Bus Stops

	Weekday	Saturday	Sunday
Mean daily (boarding plus alighting)	230.4	141.9	109.5
Min	0	0	0
25th percentile	20	7	5
Median	63	33	25
75th percentile	189	111	88
Max	16,678	10,628	8,150
Number of observations (stations)	8,036	8,036	8,036 [2]

[1] The original data provided by the L.A. Metro Service Planning, Scheduling, and Analysis team contained 110 bus stops that were missing or not reporting average weekday boarding and alighting information. Weekend service: 1,798 and 2,329 bus stops were not operating or reporting any data for Saturday and Sunday respectively.

[2] 64 Collapsed bus stops resulted in a zero-sum average weekly boarding and alighting.

BART Rail Stations

	Weekday	Saturday	Sunday
Mean daily (boarding plus alighting)	17,512	6,747.4	4,492.3
Min	2,672	705	539
25th percentile	8,067	2,841	2,037
Median	11,763.5	4,524	2,939
75th percentile	1,8038	8,135	5,506.8
Max	91,489	35,941	22,636
Number of observations (stations)	48	48	48

Stratification by Quintiles

We found that it was prohibitively time consuming to calculate job access measures for all 8,036 collapsed stops. For that reason, we used a stratified random sample of 1,000 collapsed bus stops, stratified by average weekday boarding and alighting. The boarding and alighting for each collapsed stop is the sum of boardings and alightings for the constituent stops that form the collapsed stops. Those 8,036 bus stops were grouped into quintiles based on average weekday boarding and alighting. Summary information for the quintiles is shown in Table 2. We sorted all 8,036 collapsed bus stops based on the sum of the average weekday boarding plus alighting from lowest to highest. Once stratified from lowest to highest, quintiles were formed with each quintile containing 1,607 collapsed bus stops with the exception of the fifth quintile which contained 1,608 collapsed bus stops.

Table 2. Descriptive statistics for stratified quintiles

Quintile Number	Min on+off	Max on+off	Mean on+off	Median on+off	Number of collapsed stops in quintile
First	0	15	7.10	7	1,607 [3]
Second	15	42	26.77	26	1,607
Third	42	99	66.13	65	1,607
Fourth	99	260	163.16	155	1,607
Fifth	260	16,678	1,001.73	588	1,608
Number of Stops					8,036

The location of each quintile of boardings+alightings for bus stops is shown in Figure 4 through Figure 8. As expected, the highest ridership (boardings+alightings) stops are closer to downtown and the lower quintile boarding+alighting stops are more distant from downtown.

[3] The original data from the L.A. Metro Service Planning, Scheduling, and Analysis team on some instances had bus stops not reporting or missing data. As a result, when collapsed, the sum average weekday boarding and alighting data for the first and second quintiles indicate very low ridership.

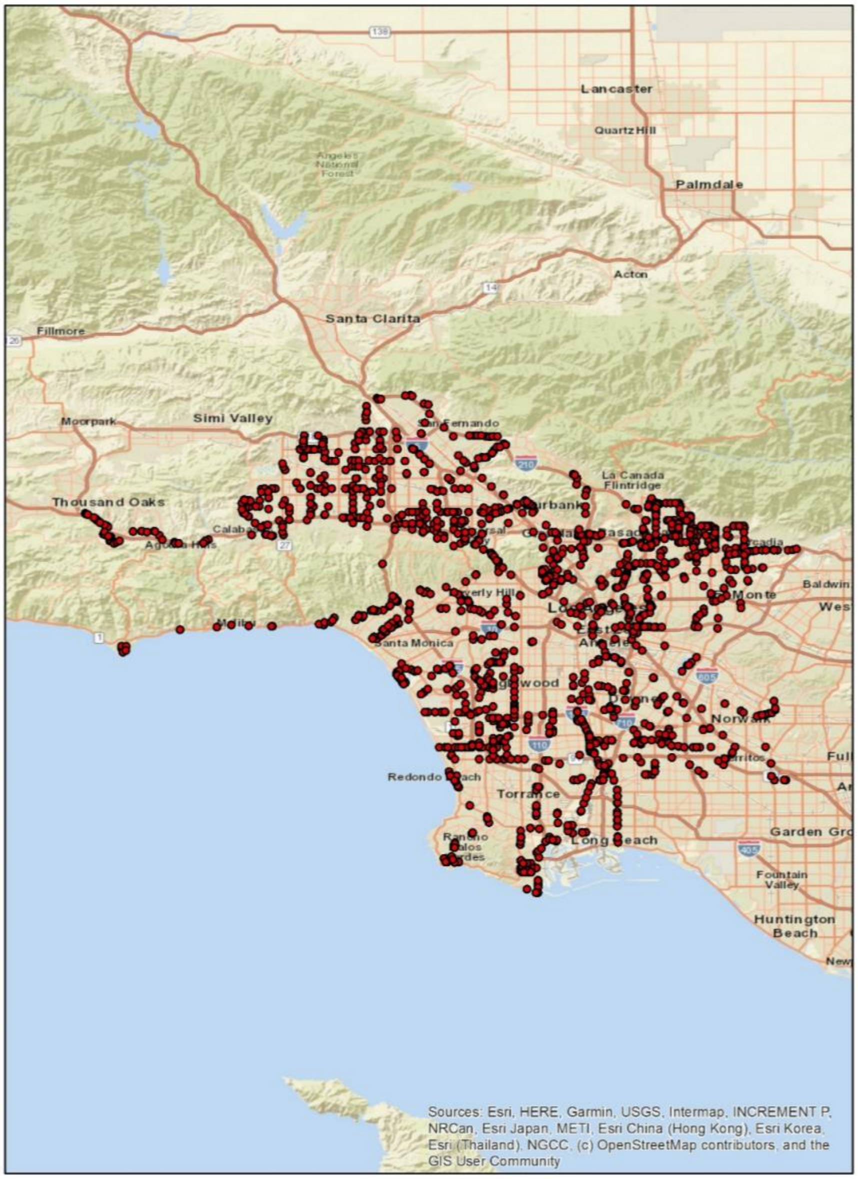

Figure 4. Location of first (lowest) quintile boarding + alighting L.A. Metro bus stops

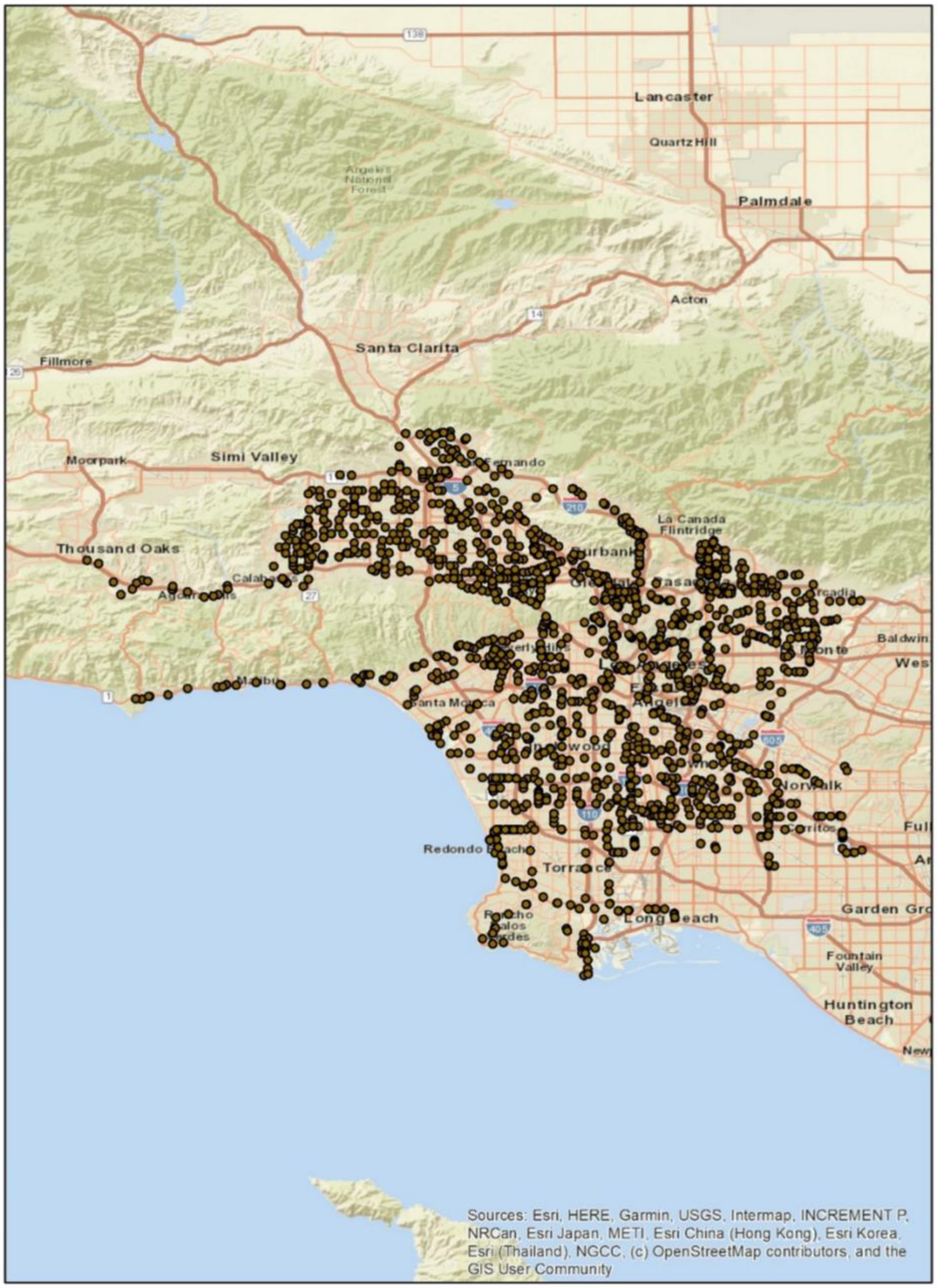

Figure 5. Location of second quintile boarding + alighting L.A. Metro bus stops

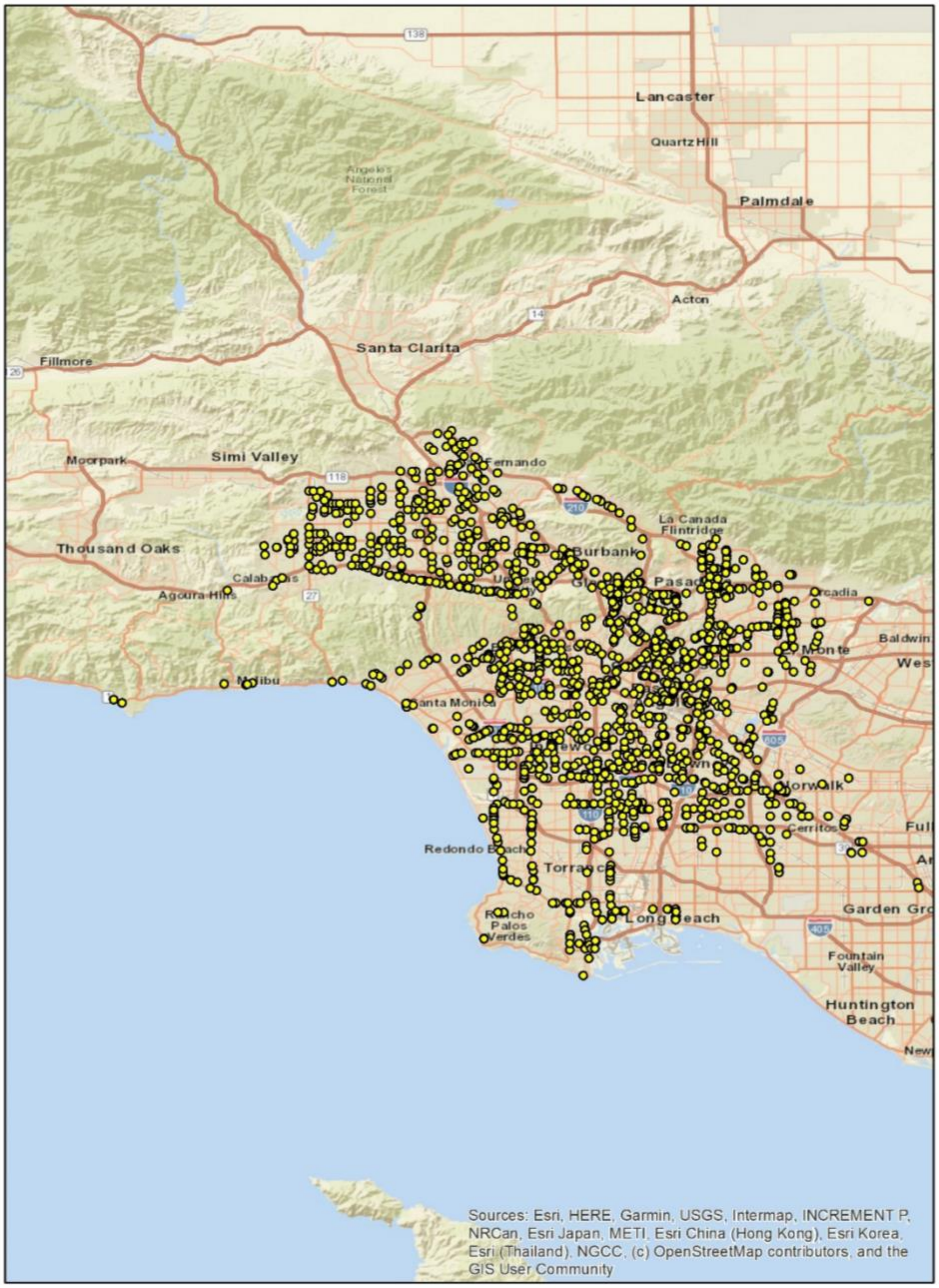

Figure 6. Location of third quintile boarding + alighting L.A. Metro bus stops

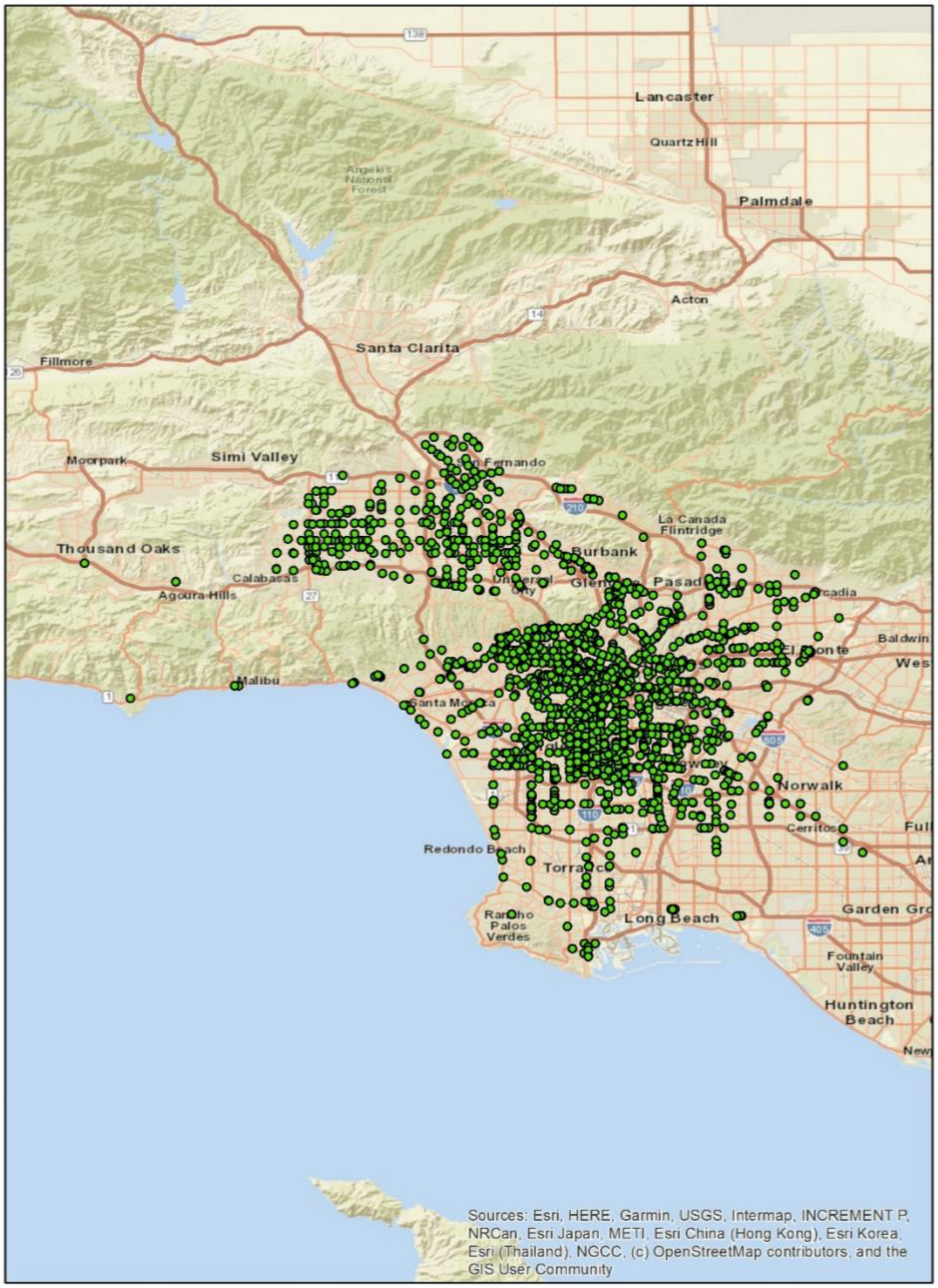

Figure 7. Location of fourth quintile boarding + alighting L.A. Metro bus stops

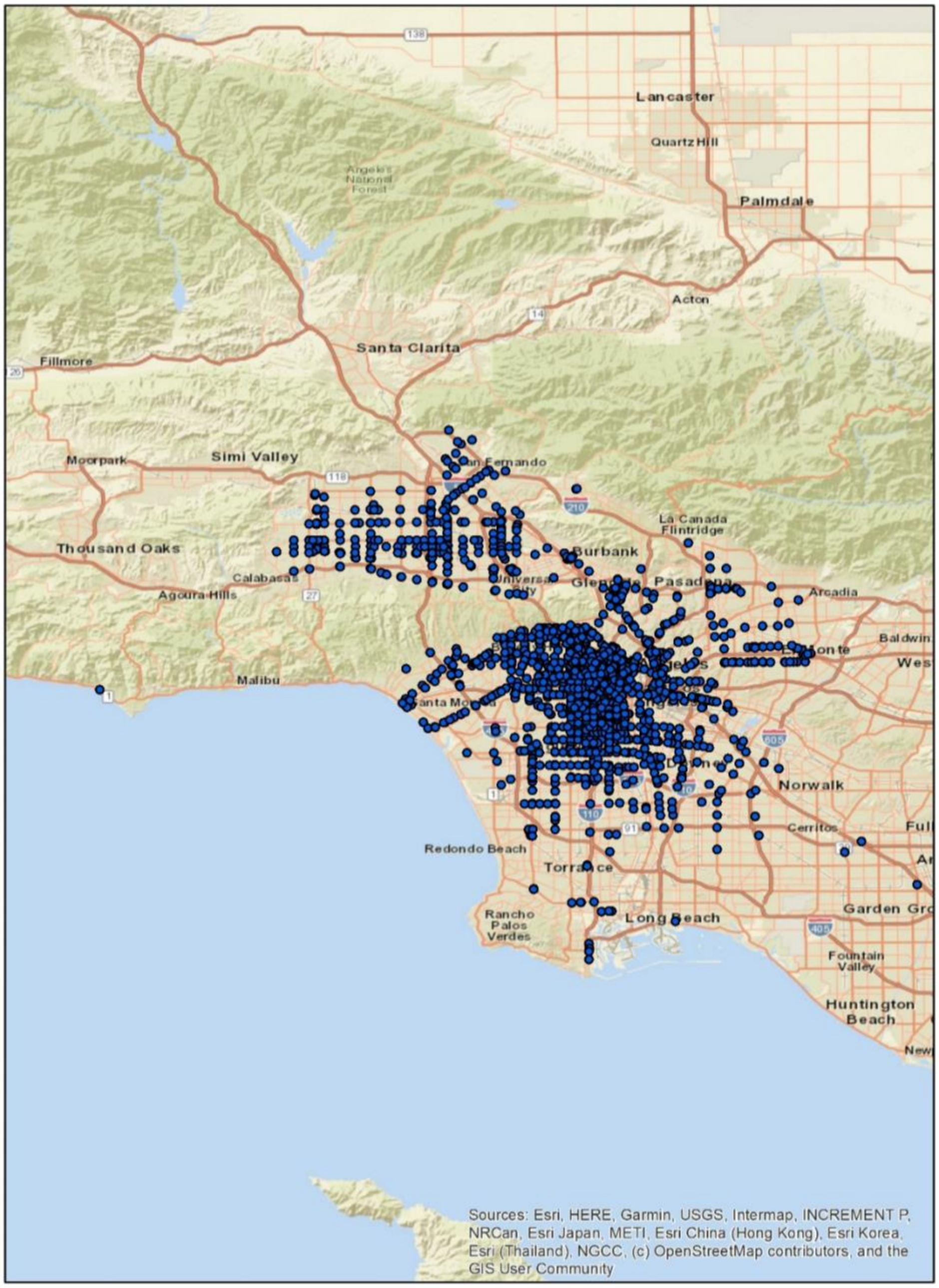

Figure 8. Location of fifth quintile boarding + alighting L.A. Metro bus stops

Once all 8,036 collapsed bus stops were placed into quintiles, we drew 200 collapsed bus stops at random from each quintile, each draw without replacement. The 200 random draws per quintile were made by using a random number generator that was assigned to each of the collapsed bus stops within each quintile. A total of 1,000 randomly drawn collapsed bus stops, 200 per quintile, were then utilized to obtain 30-minute travel time isochrones from the centroid of each collapsed bus stop, providing the job-accessibility variable for the regression analysis presented later in this report.

30-minute Job Accessibility Isochrones

We use the Remix commercial software to extract the number of jobs that are accessible within a 30-minute travel time. This allows us to obtain 30-minute travel time isochrones from the centroid of the 1,000 randomly selected LA Metro collapsed bus stops and 96 Metro rail transit stations, and the 48 BART rail transit stations. Remix transit maps are based on GTFS data from transit authorities. When a map is generated in Remix, the most recently updated GTFS data version in their system is used to draw the transit lines.[4] Once a map is built, the GTFS data in it remain static so that users can save and reuse a particular scenario of interest. Once a point is entered in the Remix tool, the user can determine specific parameters such as boarding time; type of transportation system; time traveled, etc. The subsequent result is an isochrone that is directly overlaid on a job accessibility dataset.[5] As illustrative examples, the 30-minute travel time isochrone map is shown for the Powell Street station in San Francisco in Figure 9. Figure 10 shows the 30-minute isochrone for the bus stops at Broadway and 7th Streets in downtown Los Angeles.

[4] Remix GTFS upload dates: L.A. Metro Light Rail, updated February 9, 2019; L.A. Metro Bus, updated March 18, 2019; BART, updated March 22, 2021 (for a summary of service changes to-date since the COVID 19 pandemic for BART please visit: https://www.bart.gov/news/articles/2020/news20200406)

[5] An alternative open source method for this type of analysis can be performed using the r5r library for R software. For a discussion on how to use this method please see, e.g., Pereira et al, 2021.

Figure 9. Example of a 30-minute job accessibility isochrones for the Powell Street Station (BART) in the San Francisco Financial District

We obtained isochrones with an assumption of transfer wait times of half headways, during the morning peak hour, at 8 a.m. (based on the Remix GTFS transit network upload dates that were current at the time each scenario map was originally generated, with upload dates in footnote 4). The geographic coordinate of each of the 1,000 collapsed bus stops and rail stations were entered into Remix. This allows us to obtain total job counts within each isochrone centered on rail and bus stops. This gives a cumulative opportunities measure—the number of jobs that can be reached within 30 minutes from each of the 1,000 collapsed bus stops in our analysis. We get the same cumulative opportunities measure for L.A. Metro rail stations and BART rail stations.

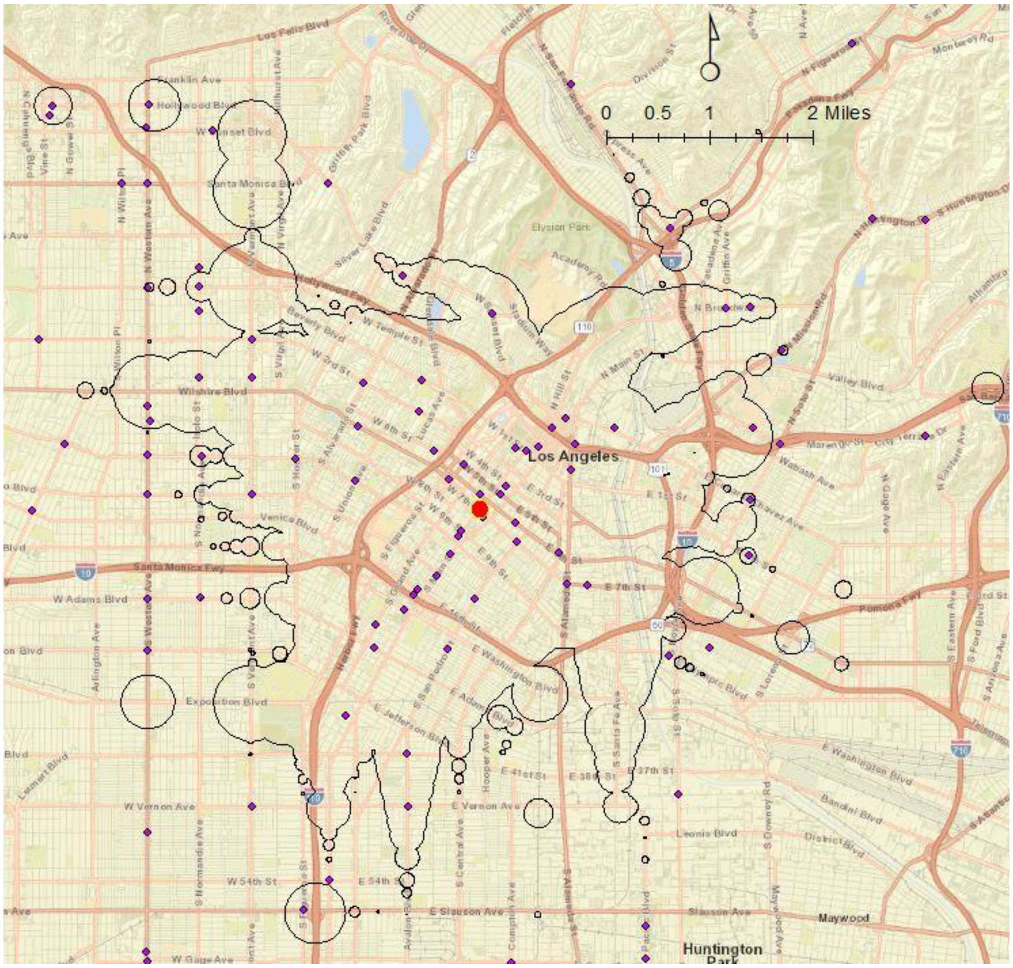

Figure 10. Example of a 30-minute job accessibility isochrone for the LA Metro collapsed bus stop near the corner of Broadway Street and 7th Street (Unique ID 4753) in Downtown Los Angeles

Independent Variables

We also control for a set of demographic and built environment factors that may influence ridership. For all variables other than the key variable of Job Access, we obtain them from IPUMS ACS 5-year series for 2015-2019 through the NHGIS data portal (Manson, et al. 2021). We first download the 2019 shapefile for census block groups in California and then join the respective dataset for each control variable with the Geopandas package in Python (Jordahl, et al. 2020). Since each dataset is produced from the ACS at the same block-group level, this requires no additional transformation. Next, we transform each variable as necessary for readability; for example, ACS total population is divided by square footage and multiplied by 5,280 squared to produce a measure of population per square mile, and the number of persons with access to a vehicle is divided by total number of persons for whom vehicle access is recorded in the ACS and multiplied by 100 to produce a percentage of persons with vehicle access for each block group (Please refer to Appendix B for Python script). For each stop or station, we use a geographically weighted average of the control variable values in each block group within a .75-mile radius of each station. This process is described in detail in the Results section below. The 0.75 mile buffers around Metro rail stations are shown in Figure 11.

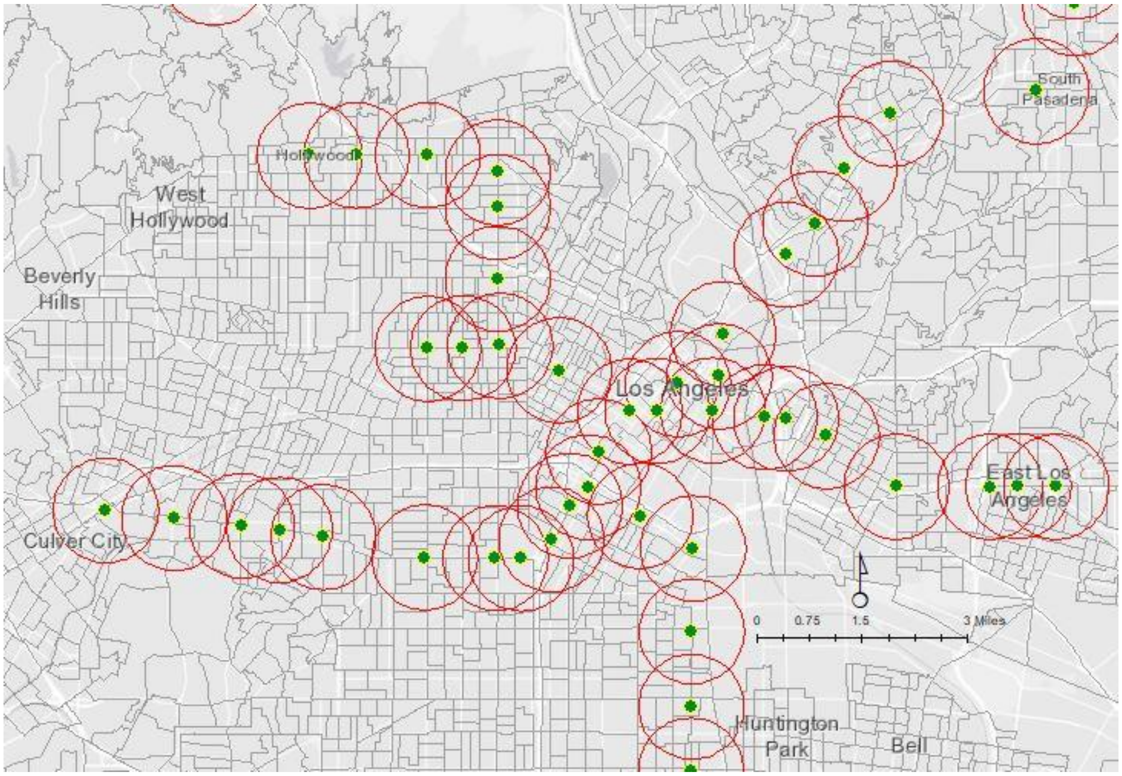

Figure 11. Example of a .75-mile radius buffer area around L.A. Metro Light Rail Stations

Results

Our OLS regression model is given by the formula:

$$Y_i = a + \beta_1 X_{1i} \dots \beta_n X_{ni} + \varepsilon_i$$

Where Y is the dependent variable of average ridership, a is the constant, ε is the error term, and β represents the coefficients. X represents our independent variables, including our variable of interest, Job Access, as well as a set of control variables gathered from the American Community Survey (ACS) 5-Year estimates (2015-2019), i indexes observations which include the collapsed set of LA Metro Bus stops, the LA Metro Rail stations, and the San Francisco BART stations, in respective models, and n represents the number of independent variables. For our Job Access variable, we use the Remix Job Access platform to gather the number of jobs accessible in 30 minutes from each stop or station. For the control variables, we take a .75-mile buffer around the centroid of each stop or station and take an aerial interpolation of each ACS variable at the block group level. In other words, we first use a buffer with a radius of .75 miles to create a circle around each bus stop or station, and take a weighted average of each block group based on the area which is covered by the circle. The control variables are the following demographic variables for the areas within 0.75 miles of a collapsed stop or rail station:

- Percent of persons African American, Asian, and Hispanic, per census definitions
- Percent of persons who were born outside of the United States out of the total census population
- Percent of households without vehicle access
- Percent of persons with income at or below 99 percent of the census-defined Poverty Line
- Population Density in number of persons per square mile
- Percent persons under 18 Years of Age

Descriptive statistics for the control variables and the key variable of interest, jobs within 30-minute morning peak transit travel, are shown in Tables 3 to 5 below.

Table 3. Descriptive statistics for L.A. Metro bus collapsed bus stops

	% Under Poverty Line	% Households w/o Vehicle	Population per Sq. Mi.	% Under 18	% African American	% Asian	% Hispanic	% Foreign Born	Jobs Accessible within 30 Minutes	Average Daily On/Off Ridership
Count	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Mean	14.96	9.39	122,876.25	20.33	10.19	13.68	43.90	33.50	74,422.02	281.94
St dev	7.09	6.26	56,148.20	4.16	12.36	11.80	23.33	8.12	108,567.84	773.97
Min	2.80	0.27	3,025.06	10.22	0.20	0.32	3.82	11.65	511.00	0.00
25th percentile	9.68	5.09	87,397.12	17.37	2.75	6.50	24.17	27.75	17,331.25	20.00
50th percentile	13.31	7.29	112,963.69	19.66	5.10	10.67	41.02	33.45	31,510.00	63.50
75th percentile	18.78	12.05	147,371.23	23.05	10.67	18.10	60.50	39.11	67,656.00	202.00
Max	41.29	30.40	318,181.91	31.47	62.50	63.04	95.28	56.25	543,229.00	12,223.00

Table 4. Descriptive statistics for L.A. Metro light rail

	% Under Poverty Line	% Households w/o Vehicle	Population per Sq. Mi.	% Under 18	% African American	% Asian	% Hispanic	% Foreign Born	Jobs Accessible within 30 Minutes	Average Daily On/Off Ridership
Count	96	96	96	96	96	96	96	96	96	96
Mean	19.29	12.95	146,367.88	20.06	10.95	13.20	48.35	32.86	231,380.17	7,138.50
St dev	9.49	7.79	57,917.17	4.67	10.65	8.26	20.78	7.37	172,616.14	9,244.41
Min	7.10	2.43	33,920.12	11.55	0.60	0.38	12.75	19.75	15,542.00	888.00
25th percentile	11.44	6.79	110,875.87	16.51	4.59	7.54	32.27	26.81	88,639.75	3,206.00
50th percentile	16.23	11.60	139,945.74	18.92	7.36	12.16	49.59	32.12	155,316.50	4,389.00
75th percentile	26.16	18.40	170,089.43	23.20	11.63	18.83	63.10	39.10	394,158.25	6,655.00
Max	48.24	30.73	318,161.41	30.35	43.99	42.96	90.86	49.46	563,356.00	56,073.00

Table 5. Descriptive statistics for BART rail system

	% Under Poverty Line	% Households w/o Vehicle	Population per Sq. Mi.	% Under 18	% African American	% Asian	% Hispanic	% Foreign Born	Jobs Accessible within 30 Minutes	Average Daily On/Off Ridership
Count	48	48	48	48	48	48	48	48	48	48
Mean	10.00	12.20	115,525.66	18.40	8.60	26.82	21.19	30.97	182,509.85	17,512.00
St dev	4.58	11.76	90,048.99	4.84	6.78	13.48	11.59	8.84	220,633.80	18,336.94
Min	2.58	0.84	2,876.05	7.81	0.42	7.58	4.74	13.30	3,875.00	2672.00
25th percentile	6.64	4.93	62,458.03	14.80	3.36	17.54	12.79	24.75	27,913.50	8,324.25
50th percentile	9.81	6.79	83,277.70	19.73	5.46	22.80	16.88	31.16	58,604.00	11,763.50
75th percentile	12.29	15.46	140,592.31	21.41	11.49	34.18	31.01	36.87	343,130.00	17,307.25
Max	22.42	47.90	348,974.29	28.82	28.95	77.51	48.89	54.21	645,715.00	91,489.00

Table 6 shows the regression results from our main models. Below each coefficient we include the standard error, in parentheses, and the elasticity at the mean, italicized. The elasticities are calculated by multiplying the coefficient by the mean of the independent variable and dividing by the mean of the dependent variable. In other words, while the coefficient can tell us the expected change in the dependent variable with a one-unit increase in the independent variable, the elasticity can tell us the expected percentage change in the dependent variable for a one-percent increase in the independent variable. We run separate regressions with and without control variables on the collapsed LA Metro Bus Stops, LA Metro Train Stations, and BART stations, respectively. In order to investigate whether our results are sensitive to outliers, we rerun the LA Metro Bus Stop models after removing all observations with average ridership over 4,000 (the 99^{th} percentile). These can be found in the third and fourth columns from the left in Table 6.

Table 6. OLS regression results

	Dependent variable: Average Ridership							
	Metro Bus		Metro Bus without Outliers		Metro Rail		BART	
Job Access	0.003***	0.003***	0.002***	0.002***	0.026***	0.015	0.063***	-0.013
	(0.0002)	(0.0004)	(0.0001)	(0.0003)	(0.005)	(0.016)	(0.008)	(0.013)
	0.70	*0.69*	*0.62*	*0.58*	*0.83*	*0.49*	*0.66*	*-0.13*
Constant	84.975***	-338.750*	87.743***	-283.549***	1,238.152	7,913.036	5,992.490**	-10,706.535
	(27.571)	(174.935)	(16.823)	(105.927)	(1,399.581)	(8,238.626)	(2,268.565)	(13,588.604)
Percent African-American		2.578		3.600**		26.433		-165.977
		(2.498)		(1.511)		(105.829)		(387.271)
		0.09		*0.16*		*0.04*		*-0.08*
Percent Asian		-5.059*		-4.067**		86.886		-367.955
		(2.927)		(1.771)		(236.194)		(388.324)
		-0.25		*-0.24*		*0.16*		*-0.56*
Percent Hispanic		-2.015		-1.813		52.631		-366.077
		(1.924)		(1.164)		(141.335)		(292.088)
		-0.31		*-0.34*		*0.36*		*-0.44*
Percent Foreign Born		9.005*		8.736***		-246.108		775.690
		(5.230)		(3.165)		(460.334)		(620.579)
		1.07		*1.26*		*-1.13*		*1.37*
Percent without a Vehicle		-12.89		-1.055		338.372		1,818.778***
		(11.309)		(6.862)		(351.014)		(351.359)
		-0.43		*-0.04*		*0.61*		*1.27*

	Dependent variable: Average Ridership							
	Metro Bus		**Metro Bus without Outliers**		**Metro Rail**		**BART**	
Percent under the Poverty Line		2.891		0.274		-164.516		-283.207
		(8.142)		(4.930)		(194.768)		(666.295)
		0.15		*0.02*		*-0.44*		*-0.16*
Population Per Sq. Mi.		$.002^{**}$		0.0006		0.021		-0.014
		(0.0007)		(0.0004)		(0.032)		(0.045)
		0.78		*0.30*		*0.43*		*-0.09*
Percent under 18 Years of Age		5.717		6.088		-220.49		422.596
		(8.794)		(5.318)		(412.882)		(528.209)
		0.41		*0.53*		*-0.62*		*0.24*
Observations	1,000	1,000	992	992	96	96	48	48
R^2	0.138	0.159	0.19	0.221	0.227	0.245	0.577	0.86
Adjusted R^2	0.137	0.151	0.19	0.213	0.219	0.167	0.568	0.816
Residual Std. Error	719.015	712.993	437.311	430.838	8,172.269	8439.734	12,058.597	7,872.605
F Statistic	159.540^{***}	20.797^{***}	232.834^{***}	30.872^{***}	27.562^{***}	3.109^{***}	62.682^{***}	24.109^{***}
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$							

Our main regressor of interest is the number of jobs accessible via transit within 30 minutes. In six models, we find this variable to be positive and significant. However, when we add controls to the BART and Metro Rail models, we find insignificant results. This is not necessarily surprising, due to the lower number of observations available in the BART and LA Metro Rail systems. In our Metro Bus models, we find that a one-percent increase in Job Access at the mean would be expected to associated with about a 0.7 percent (with outliers) or 0.6 percent (without outliers) increase in bus stop boarding and alightings, all else being held equal. Similarly, in the LA Metro Rail and BART systems, we find elasticities of about 0.8 and 0.7, respectively.

While the relationship between job access and ridership is generally positive and significant, it is possible that the relationship is nonlinear and not the same for high and low-ridership bus stops. In order to investigate this possibility, we implement a separate regression model for each quintile of ridership. See Table 7 for the results of the regressions by L.A. bus stop quintile.

Table 7. OLS regressions by bus ridership quintiles

	Dependent variable: Average Ridership				
	Lowest Ridership (1)	**(2)**	**(3)**	**(4)**	**Highest Ridership (5)**
Job Access	-0.000005	0.00001	0.00006*	-0.00003	0.002
	(0.000008)	(0.00001)	(0.00003)	(0.00006)	(0.001)
	-0.02	*0.02*	*0.04*	*-0.01*	*0.31*
Constant	7.120***	20.319***	60.750***	175.694***	82.139
	(2.118)	(4.022)	(8.815)	(27.220)	(1,019.853)
Percent African-American	0.009	0.112	0.043	-0.094	-0.376
	(0.031)	(0.070)	(0.125)	(0.346)	(11.186)
	0.01	*0.04*	*0.01*	*-0.01*	*0.00*
Percent Asian	-0.013	0.014	0.053	0.865*	-17.8
	(0.036)	(0.072)	(0.129)	(0.477)	(14.862)
	-0.03	*0.01*	*0.01*	*0.06*	*-0.21*
Percent Hispanic	0.047*	-0.054	0.207**	0.389	-2.668
	(0.024)	(0.043)	(0.089)	(0.294)	(10.693)
	0.25	*-0.09*	*0.14*	*0.11*	*-0.11*
Percent Foreign Born	0.115*	0.271**	0.143	-1.14	15.723
	(0.066)	(0.127)	(0.229)	(0.839)	(27.862)
	0.53	*0.32*	*0.07*	*-0.24*	*0.50*
Percent without a Vehicle	0.034	-0.207	0.337	3.123*	-36.822
	(0.189)	(0.285)	(0.627)	(1.658)	(46.506)
	0.03	*-0.06*	*0.04*	*0.20*	*-0.46*
Percent under the Poverty Line	0.036	0.037	-0.28	-2.624*	4.369
	(0.078)	(0.231)	(0.406)	(1.371)	(44.228)
	0.06	*0.02*	*-0.06*	*-0.26*	*0.07*

	Dependent variable: Average Ridership				
	Lowest Ridership				**Highest Ridership**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Population Per Sq. Mi.	-0.00002*	-0.00001	-0.00003	0.0002	0.006**
	(0.00001)	(0.00002)	(0.00005)	(0.0001)	(0.003)
	-0.25	*-0.04*	*-0.05*	*0.13*	*0.89*
Percent under 18 Years of Age	-0.217**	0.024	-0.411	-0.289	-3.985
	(0.109)	(0.202)	(0.445)	(1.429)	(46.543)
	-0.65	*0.02*	*-0.13*	*-0.04*	*-0.07*
Observations	200	200	200	200	200
R^2	0.09	0.064	0.104	0.08	0.105
Adjusted R^2	0.047	0.019	0.062	0.036	0.063
Residual Std. Error	3.823	7.457	16.057	44.85	1388.257
F Statistic	2.081**	1.433	2.453**	1.829*	2.476**
Note:	*p<0.1; **p<0.05; ***p<0.01				

Similarly, in Table 8, we run five regressions for L.A. Metro collapsed bus stops by quintiles of job access. Those job access quintiles are displayed in Figure 12. Again, we find the strongest relationship between Job Access and Ridership at the highest quintile

Table 8. OLS regressions by job access quintiles

	Dependent variable: Average Ridership				
	Lowest Job Access				**Highest Job Access**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Job Access	0.003	0.013**	-0.002	0.004	0.003**
	(0.002)	(0.005)	(0.006)	(0.003)	(0.001)
	0.53	*2.26*	*-0.34*	*0.72*	*1.12*
Constant	-23.921	-379.270**	-539.067*	-1209.103**	507.831
	(36.610)	(154.808)	(273.382)	(467.089)	(1,255.269)
Percent African-American	0.67	0.144	5.614**	3.845	9.669
	-0.896	-1.648	-2.241	-4.404	-29.862
	0.13	*0.02*	*0.31*	*0.14*	*0.09*
Percent Asian	-1.815**	-2.925	-3.163	-10.936	-26.883
	(0.785)	(2.319)	(2.755)	(6.929)	(25.198)
	-0.62	*-0.30*	*-0.27*	*-0.41*	*-0.54*
Percent Hispanic	0.042	-2.545*	-3.998*	-2.008	12.254
	(0.487)	(1.417)	(2.047)	(4.089)	(15.994)
	0.04	*-1.19*	*-1.06*	*-0.24*	*0.70*
Percent Foreign Born	3.018**	3.696	12.018**	13.963*	-7.426

	Dependent variable: Average Ridership				
	Lowest Job Access				**Highest Job Access**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	(1.433)	(4.251)	(5.336)	(7.894)	(49.832)
	2.15	*1.09*	*2.25*	*1.47*	*-0.36*
Percent without a Vehicle	7.596	-2.609	-6.131	-17.28	-51.227
	(5.121)	(8.418)	(11.880)	(21.594)	(61.126)
	0.95	*-0.18*	*-0.26*	*-0.49*	*-1.17*
Percent under the Poverty Line	-3.165	3.897	14.454**	5.74	-2.361
	(2.853)	(5.917)	(6.569)	(14.308)	(63.044)
	-0.81	*0.53*	*1.13*	*0.25*	*-0.06*
Population Per Sq. Mi.	-10736.043	20312.076	-12072.592	75016.867	164124.326
	(8,425.162)	(20,382.363)	(29,414.008)	(48,480.251)	(110,468.015)
	-0.76	*0.78*	*-0.26*	*1.02*	*1.40*
Percent under 18 Years of Age	-0.086	6.922	20.713**	39.337*	-36.561
	(1.833)	(6.352)	(10.279)	(20.007)	(60.646)
	-0.04	*1.38*	*2.38*	*2.47*	*-0.84*
Observations	200	200	200	200	200
R^2	0.091	0.103	0.141	0.124	0.077
Adjusted R^2	0.048	0.061	0.1	0.083	0.033
Residual Std. Error	73.143	223.575	319.089	620.702	1420.891
F Statistic	2.105**	2.428**	3.451***	2.999***	1.750*
Note:	*p<0.1; **p<0.05; ***p<0.01				

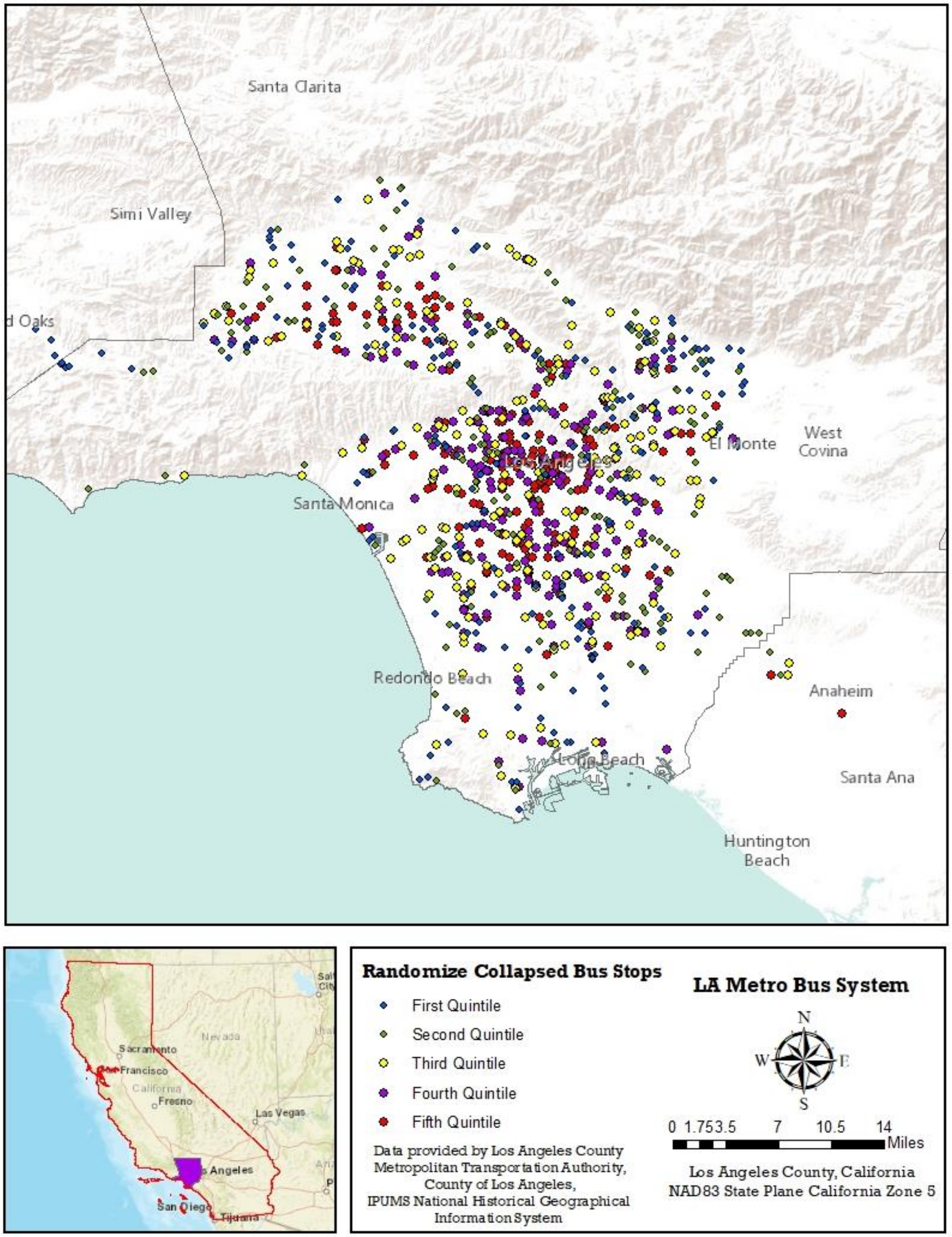

Figure 12. Job access quintiles for the 1,000 collapsed L.A. Metro bus stops used in the regression analysis

Discussion

The full sample results show consistent relationships between job access and stop/station ridership (boardings/alightings), with an elasticity in the range of 0.6 to 0.8, larger than the elasticity of any of the control variables except percent immigrant in the surrounding neighborhoods. Note that the control variables are generally not policy variables, and that the relationship between job access and ridership has a similar elasticity across the bus and rail systems.

The job access variable is less robust in the L.A. Metro and BART analysis. For L.A. Metro, the small sample size (n = 96) is likely part of the reason why job access is not statistically significant when the control variables are added (Table 6). But note that the job access variable loses magnitude when control variables are added. For BART (n = 48), the job access variable loses significance and changes sign when control variables are added (Table 6). That suggests that the loss of significance may be for reasons beyond sample size, although it is not possible to give a definitive answer. We note that the BART system was built as a high-speed, long-distance commuter rail system, and station ridership hence may have less of a relationship with job access in such a system. Note that for all three transit systems—BART, L.A. Metro rail, L.A. Metro bus—the bivariate relationship between 30-minute job access and station area ridership is strong.

The quintile regressions are less conclusive than the full sample. Tables 7 and 8 show the regressions for quintiles of bus stop ridership and bus stop 30-minute transit job access. The regressions for bus stop ridership quintiles (Table 7) show no relationship between ridership quintiles and job access, with the job access variable insignificant in those smaller sample size (n = 200) regressions. In Table 8, the 2nd and highest job access quintiles are statistically significant, both with elasticities larger than 1. This suggests that improving transit job access both in high access locations (the top quintile of job access) and in lower access locations (the 2nd quintile) can lead to ridership gains. Overall, our interpretation of the regressions sorted by quintiles is that we do not find strong evidence of differences across levels of stop-level ridership or job access, and we suggest that the full sample results are preferred.

Conclusion

The relationship between job access and stop/station ridership is robust in the bivariate regressions, although for L.A. Metro rail and BART not statistically significant when control variables are added. Our key finding is the relationship between 30-minute job access and bus stop ridership. If the number of jobs accessible within 30-minutes were to increase by 1 percent, our results suggest that on average stop-level ridership would increase by from 0.6 to 0.8 percent.

The job access measure depends on the distribution of jobs, and hence the siting of bus stops and lines, the frequency of morning peak hour service, and the pattern of transfers. Anything that increases bus travel time, including transfer and wait time, will increase our job access measure. Similarly, siting stops and lines in areas that are more job rich will increase transit access.

Decomposing job access into the location of bus stops, bus routes, frequencies, and transfers is beyond the scope of this research. Yet we do offer some suggestions. Job access is a mix of internal and external factors or, equivalently, a mix of factors that transit agencies can and cannot control. Changing the pattern of job locations will be slow in the near-term and outside of the realm of transit agencies, and hence is an external factor. Changing patterns of bus routes and stops, while internal factors, will require many decisions that go beyond job access. Hence bus agencies might find it more fruitful to focus on increasing job access by increasing bus frequency (which will reduce wait times at stops), reducing transfers or transfer wait times, and improving stop access/egress times (e.g., encouraging bicycle and car access/egress where possible.) We note that there are internal factors that are within transit agency control that can influence transit job access, and we suggest renewed focus on those factors.

Our finding that job access is closely related to ridership at the bus stop level is intuitive, but has not previously been demonstrated in the literature to our knowledge. With new tools available to measure job access more quickly and easily, we suggest that job access should have a prominent role in transit planning.

References

Alam, Bhuiyan Monwar, Hilary Nixon, and Qiong Zhang. 2018. "Factors Affecting Travel Demand by Bus: An Empirical Analysis at U.S. Metropolitan Statistical Area Level." *Transportation Research Record* 2672 (8): 817-826.

Boarnet, Marlon, Genevieve Giuliano, Yuting Hou, and Eun Jin Shin. 2017. "First/last Mile Transit Access as an Equity Planning issue." *Transportation Research Part A: Policy and Practice* 103: 296-310.

Boer, Cui, Genevieve Boisjoly, Ahmed El-Geneidy, and David Levinson. 2018. "Accessibility, equity, and the journey to work."

Cui, Mengying, and David Levinson. 2018. "Accessibility Analysis of Risk Severity." *Transportation* 45 (4): 1029-1050.

Cui, Mengying, and David Levinson. 2018. "Accessibility and the Ring of Unreliability." *Transportmetrica A: Transport Science* 14 (1-2): 4-21.

Cui, Mengying, and David Levinson. 2019. "Primal and Dual Access." *Geographical Analysis* (52) 3: 452-474.

Currie, Graham, and Alexa Delbosc. 2011. "Understanding bus rapid transit route ridership drivers: An empirical study of Australian BRT systems." *Transport Policy* 18 (5): 755-764.

Deboosere, Robbin, Ahmed M El-Geneidy, and David Levinson. 2018. "Accessibility-Oriented Development." *Journal of Transport Geography* 70: 11-20.

Handy, Susan, Kristin Lovejoy, Marlon Boarnet, and Steven Spears. 2013. *Impacts of Transit Service Strategies on Passenger Vehicle Use and Greenhouse Gas Emissions.* Policy Brief, Sacramento, CA: California Air Resources Board. https://ww3.arb.ca.gov/cc/sb375/policies/transitservice/transit_brief.pdf.

Hansen, Walter G. 1959. "How Accessibility Shapes Land Use." *Journal of the American Institute of Planners* 25 (2): 73-76.

Jordahl, Kelsey, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, and François Leblanc. 2020. "Geopandas/geopandas: v0.8.1 (Version v0.8.1)." *Geopandas Version v0.8.1.* Zenodo, July 15. http://doi.org/10.5281/zenodo.3946761.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2021. *IPUMS National Historical Geographic Information System: Version 16.0 [dataset].* Minneapolis. http://doi.org/10.18128/D050.V16.0.

Merlin, Louis A, and Lingqian Hu. 2017. "Does competition matter in measures of job accessibility? Explaining employment in Los Angeles." *Journal of Transport Geography* 64: 77-88.

O'Kelly, Morton E, and Mark W Horner. 2003. "Aggregate Accessibility to Population at the County Level: US 1940-2000." *Journal of Geographical Systems* 5 (1): 5-23.

Owen, Andrew, and David Levinson. 2015. "Modeling the Commute Mode Share of Transit Using Continuous Accessibility to Jobs." *Transportation Research Part A: Policy and Practice* 74: 110-122.

Painter, Gary, Marlon Boarnet, Madison Swayne, and M Miller. 2019. *Innovation on Job Accessibility and Transit Scenario Planning with General Transit Feed Specification (GTFS) Data.* Research Report, Washington, D.C.: Transportation Research Board.

Paulley, Neil, Richard Balcombe, Roger Mackett, Helena Titheridge, John Preston, Mark Wardman, Jeremy Shires, and Peter White. 2006. "The demand for public transport: The effects of fares, quality of service, income and car ownership." *Transport Policy* 13 (4): 295-306.

Pereira, Rafael H.M., Marcus Saraiva, Daniel Herszenhut, Carlos Kaue Vieira Braga, and Matthew Wigginton Conway. 2021. "R5r: Rapid Realistic Routing on Multimodal Transport Networks with R5 in R." *Findings*, March. http://doi.org/10.32866/001c.21262.

Shen, Qing. 1998. "Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers." *Environment and Planning B: Planning and Design* 25 (3): 345-365.

Srour, Issam M, Kara M Kockelman, and Travis P Dunn. 2002. "Accessibility Indices: Connection to Residential Land Prices and Location Choices." *Transportation Research Record* 1805 (1): 25-34.

Taylor, Brian D., and Camille NY Fink. 2003. "The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature." *University of California Transportation Center* (University of California). https://escholarship.org/uc/item/3xk9j8m2.

Taylor, Brian D., Douglas Miller, Hiroyuki Iseki, and Camille Fink. 2009. "Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas." *Transportation Research Part A* 43 (1): 60-77.

Wachs, Martin, and T. Gordon Kumagai. 1973. "Physical Accessibility as a Social Indicator." *Socio-Economic Planning Services* 7 (5): 437-456.

Wu, Hao, David Levinson, and Andrew Owen. 2021. "Commute mode share and access to jobs across US metropolitan areas." *Environment and Planning B: Urban Analytics and City Science* 48 (4): 671-684.

Data Summary

Products of Research

We collected data from three primary sources. (1) Los Angeles Metro and the Bay Area Rapid Transit system (BART) provided data on station/stop level boarding and alighting. For rail transit, those were available from public web pages for the operators. L.A. Metro provided the data for bus station boarding and alighting. (2) We calculated the number of jobs that could be reached in a 30-minute transit trip from each station/stop using the REMIX software tool. (3) We collected data from the U.S. Census. The method of data collection is described in the report, with Python code for census data in the report Appendix B.

Data Format and Content

We attach a file with the data used for the regression analysis. Variable names are listed in the top row.

Data Access and Sharing

The data are available in the Dataverse data repository: https://doi.org/10.7910/DVN/S2WRF2.

Reuse and Redistribution

Data that is restricted were not released. Data on Dataverse are available with proper citation. Suggested citation:

> Boarnet, Marlon ; Flores Moctezuma, David; Gross, James, 2022, “Replication Data for: New Open-Source Analyses of Transit Job Access and Transit Ridership”, https://doi.org/10.7910/DVN/S2WRF2, Harvard Dataverse, V1.

Appendix A. L.A. Metro Collapsed Bus Stops GIS Process

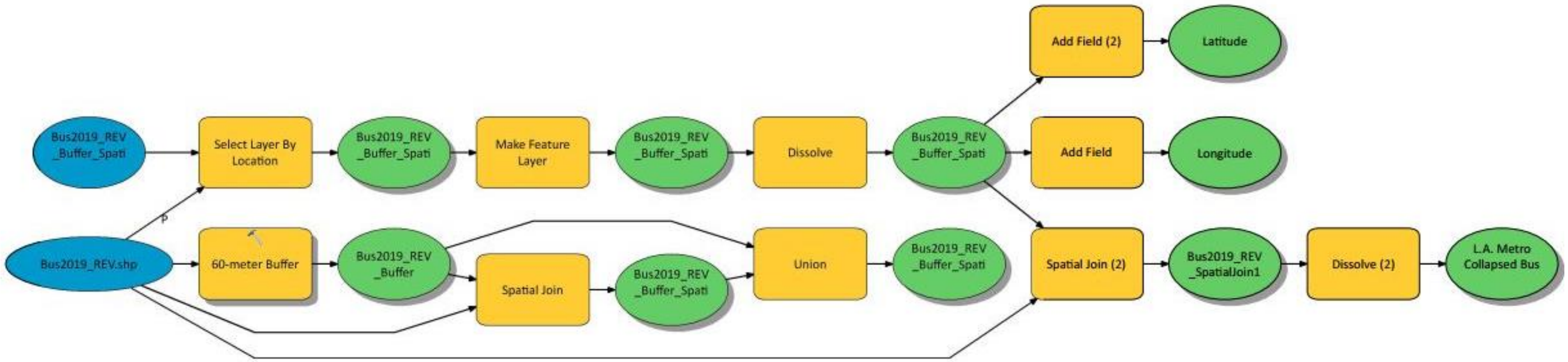

Figure 13. L.A. Metro Collapsed Bus Stops GIS Process Flow Chart

Appendix B. Independent Variables Python Script

```
#import necessary modules
import geopandas
import pandas
import numpy
import fiona


#Load bus stops csv
metrobus = pandas.read_csv('LAMetro/LA Metro Bus Stops/Collapsed_Master.csv')
C:\Users\james\miniconda3\lib\site-
packages\IPython\core\interactiveshell.py:3444: DtypeWarning: Columns
(129,148) have mixed types.Specify dtype option on import or set
low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)


#convert to gdf, set geometry
metrobus = geopandas.GeoDataFrame(metrobus,
geometry=geopandas.points_from_xy(metrobus.Longitude1, metrobus.Latitude1))
metrobus.crs = "EPSG:4326"#set to WGS 84, since that was the original
geometry that the lat and long came from
C:\Users\james\miniconda3\lib\site-packages\geopandas\array.py:275:
ShapelyDeprecationWarning: The array interface is deprecated and will no
longer work in Shapely 2.0. Convert the '.coords' to a numpy array instead.
  return GeometryArray(vectorized.points_from_xy(x, y, z), crs=crs)


#load files
bart = geopandas.read_file('BART/BART_sta_13.shp')
metrorail = geopandas.read_file('LAMetro/LA Metro Bus
Stops/LA_Metro_Light_Rail.shp')


#Assign a variable to indicate which group they belong to
bart['GROUP'] = 'bart'
metrorail['GROUP'] = 'metrorail'
metrobus['GROUP'] = 'metrobus'


#sort gdfs to ensure same order every time, assign unique ID such that Metro
Rail Unique ID starts at 10,000 and Bart starts at 20,000
metrorail = metrorail.sort_values(by='stopnum',axis=0,ascending=True)
bart = bart.sort_values(by='STATION',axis=0,ascending=True)
metrorail['Unique_ID'] = range(10000, 10000 + len(metrorail))
bart['Unique_ID'] = range(20000, 20000 + len(bart))


bart = bart.to_crs('epsg:3857')#NAD 83
metrorail = metrorail.to_crs('epsg:3857')
metrobus = metrobus.to_crs('epsg:3857')
bart.crs == metrorail.crs == metrobus.crs
True
```

```
merged = pandas.merge(bart, metrorail, how = 'outer')
merged = pandas.merge(merged, metrobus, how = 'outer')

merged['geometry'] = merged.geometry.buffer(3960)#buffer by 3960 feet
(about .75 mi)

#load ACS data and merge with bgroups

#load bgroups
bgroups = geopandas.read_file('RHS
Vars/GIS/nhgis0001_shapefile_tl2019_060_blck_grp_2019/CA_blck_grp_2019.shp')

#load data and transform columns
#pct poverty
poverty = pandas.read_csv('RHS
Vars/Poverty/nhgis0001_csv/nhgis0001_ds244_20195_2019_blck_grp.csv')
poverty['pct_poverty'] =
100*(poverty['ALWVE002']+poverty['ALWVE003'])/poverty['ALWVE001']
poverty = poverty[['pct_poverty','GISJOIN']]

#pct zero vehicles
vehicles = pandas.read_csv('RHS
Vars/Vehicles/nhgis0010_csv/nhgis0010_ds244_20195_2019_blck_grp.csv')
vehicles['pct_novehicle'] =
100*(vehicles['AL0NE003']+vehicles['AL0NE010'])/vehicles['AL0NE001']
vehicles = vehicles[['pct_novehicle','GISJOIN']]

#pop density
population = pandas.read_csv('RHS
Vars/Density/nhgis0004_csv/nhgis0004_ds244_20195_2019_blck_grp.csv')
population['population'] = (population['ALUBE001'])
population = population[['population','GISJOIN']]

#age
age = pandas.read_csv('RHS
Vars/Age/nhgis0005_csv/nhgis0005_ds244_20195_2019_blck_grp.csv')
age['under18'] =
100*(age['ALT0E003']+age['ALT0E004']+age['ALT0E005']+age['ALT0E006']+age['ALT
0E027']+age['ALT0E028']+age['ALT0E029']+age['ALT0E030'])/age['ALT0E001']
age = age[['under18','GISJOIN']]

#pct race AA
raceAA = pandas.read_csv('RHS
Vars/Race/nhgis0007_csv/nhgis0007_ds244_20195_2019_blck_grp.csv')
raceAA['pct_AA'] = 100*(raceAA['ALUCE003'])/raceAA['ALUCE001']
raceAA = raceAA[['pct_AA','GISJOIN']]
```

```
#pct race Asian
raceAsian = pandas.read_csv('RHS
Vars/Race/nhgis0007_csv/nhgis0007_ds244_20195_2019_blck_grp.csv')
raceAsian['pct_Asian'] = 100*(raceAsian['ALUCE005'])/raceAsian['ALUCE001']
raceAsian = raceAsian[['pct_Asian','GISJOIN']]

#pct hispanic
hispanic = pandas.read_csv('RHS
Vars/Hispanic/nhgis0008_csv/nhgis0008_ds244_20195_2019_blck_grp.csv')
hispanic['pct_hispanic'] = 100*(hispanic['ALUKE012'])/hispanic['ALUKE001']
hispanic = hispanic[['pct_hispanic','GISJOIN']]

#immigration
immigration = pandas.read_csv('RHS
Vars/Immigration/nhgis0011_csv/nhgis0011_ds244_20195_blck_grp.csv')
immigration['pct_immigrant'] =
100*(immigration['AL2OE005'])/immigration['AL2OE001']
immigration = immigration[['pct_immigrant','GISJOIN']]

#merge all RHS columns with bgroups
bgroups = bgroups.merge(poverty,on='GISJOIN')
bgroups = bgroups.merge(vehicles,on='GISJOIN')
bgroups = bgroups.merge(population,on='GISJOIN')
bgroups = bgroups.merge(age,on='GISJOIN')
bgroups = bgroups.merge(raceAA,on='GISJOIN')
bgroups = bgroups.merge(raceAsian,on='GISJOIN')
bgroups = bgroups.merge(hispanic,on='GISJOIN')
bgroups = bgroups.merge(immigration,on='GISJOIN')

#convert total population to population density
bgroups['popdensity'] =
bgroups['population']/bgroups.geometry.area*5280**2#converts sq ft to sq mi

#Since we will do a spatial join, make sure the crs's match
bgroups = bgroups.to_crs('epsg:3857')
bgroups.crs == merged.crs
True

#Drop missing values so they don't throw off the weighted averages of the
next cell
bgroups_cleaned = bgroups.dropna()

#finally, spatially join by weighted average of all overlayed block groups
stop_area_id_column = 'Unique_ID'

inter = geopandas.overlay(merged,bgroups_cleaned)
inter['area'] = inter.area
```

```
wm = lambda x: numpy.average(x, weights=inter.loc[x.index, "area"])
#https://stackoverflow.com/questions/31521027/groupby-weighted-average-and-sum-in-pandas-dataframe
poverty = inter.groupby(stop_area_id_column).agg({'pct_poverty':wm})
vehicles = inter.groupby(stop_area_id_column).agg({'pct_novehicle':wm})
popdensity = inter.groupby(stop_area_id_column).agg({'popdensity':wm})
age = inter.groupby(stop_area_id_column).agg({'under18':wm})
raceAA = inter.groupby(stop_area_id_column).agg({'pct_AA':wm})
raceAsian = inter.groupby(stop_area_id_column).agg({'pct_Asian':wm})
hispanic = inter.groupby(stop_area_id_column).agg({'pct_hispanic':wm})
immigration = inter.groupby(stop_area_id_column).agg({'pct_immigrant':wm})

#Finally, merge all RHS columns with the 'merged' dataset
#Merge RHS variables into one dataset
RHS = poverty.merge(vehicles, how='inner', left_index=True, right_index=True)
RHS = RHS.merge(popdensity, how='inner', left_index=True, right_index=True)
RHS = RHS.merge(age, how='inner', left_index=True, right_index=True)
RHS = RHS.merge(raceAA, how='inner', left_index=True, right_index=True)
RHS = RHS.merge(raceAsian, how='inner', left_index=True, right_index=True)
RHS = RHS.merge(hispanic, how='inner', left_index=True, right_index=True)
RHS = RHS.merge(immigration, how='inner', left_index=True, right_index=True)

#Set the Unique_ID as a column instead of index
RHS['Unique_ID'] = RHS.index
RHS.index.name = 'row'
RHS.Unique_ID = RHS.Unique_ID.astype('int')

#merge to 'merged'
merged.Unique_ID = merged.Unique_ID.astype('int')
merged = merged.merge(RHS,on='Unique_ID',how='inner')

#SAVE OUTPUT
temp = merged.copy()
temp.geometry = temp.geometry.centroid
temp.to_file('RHS.shp')

bgroups.to_file('bgroups_out.shp')

bgroups_cleaned.to_file('bgroups_cleaned_out.shp')

temp = pandas.DataFrame(merged)
temp.to_csv('RHS.csv')
```