Do It Yourself: Obtaining Updated Transit Stop and Route Shapefiles in Urban and Nonurban Areas

Seva Rodnyansky
University of Southern California

Abstract

Research that combines housing and transportation aims to jointly understand the elements of neighborhood accessibility, affordability, and sustainability. Access to high-quality public transit and nonmotorized transportation helps reduce emissions and transportation costs for all households, including those with lower incomes. Transit access also expands the range of community destinations and shopping opportunities for those without cars. However, researchers often struggle to obtain accurate, geo-coded data—especially in suburban and nonurban areas—on transit station locations, routes, and schedules. This article highlights a newer tool, the General Transit Feed Specification (GTFS) from Google, which provides an open source database of updated transit data. This free data source combines static and dynamic transit data and can be incorporated into analysis using geographic information system, or GIS, software. It also significantly eases cross-sectional, rural, and metropolitan-areawide analyses of housing using transportation as a key input. This article summarizes the GTFS data type, gives an overview of methods for using the data, explores current uses of the data, and suggests future applications.

Introduction

Accessibility to employment and amenities is a primary input to a household's choice of residential location. In the monocentric city model, households commute to jobs in the central business district and select housing locations by trading off the cost of commuting longer distances versus the higher cost of housing closer to the city center (Alonso, 1964; Brueckner, 1987). Although many U.S. metropolitan areas are less monocentric today than they were in the middle of the 20th century, many households still commute to downtowns or to local employment centers and must thus factor transportation into their location decision (Giuliano and Small, 1991; Redfearn, 2007). The same is true of other amenities households use—shopping, entertainment, educational, medical, and so on—which tend to cluster in particular locations, necessitating transportation.

Location accessibility depends on the transportation mode available and chosen. For example, a housing development near an interstate highway exit, but with no public transit, may have high accessibility for households with car access but low accessibility for those with no or low car access. However, if this highway is routinely congested, the traffic may lower the site's relative accessibility. If the same housing development had nearby bus access, its relative accessibility would still depend on the distance to the bus stop and on the frequency of bus service. Thus, while both the highway and the bus service may appear nominally accessible, in reality, traffic and transit service constraints may decrease the location's accessibility. Hence, when making residential location choices, households must optimize housing cost, job and amenity accessibility, mode choice, traffic, and transit service.

Researchers and policymakers who design and evaluate housing policies also need to take accessibility into account. Combined housing and transportation research aims to understand the combined elements of accessibility, affordability, and sustainability at both the household and neighborhood levels (Haas et al., 2013).

Measuring accessibility to account for transit service and traffic can be complex. Prior measures of location or neighborhood accessibility (for example, Alonso, 1964), used linear (Euclidean) distance, which gives only an approximation and is less suitable for cities with irregular topography or with grid street layouts, which includes many U.S. cities. More recent commercial Global Positioning System (or GPS)-based tools, such as Google Maps, take street network and congestion into account and have improved accessibility measurements for automobile, public transit, biking, and walking modes. Although Google Maps and similar tools work well for individuals, researchers and policymakers need data to be aggregable and analyzable over various time periods. For driving modes, this goal has been accomplished by using street network analyses in geographic information system (GIS) packages and by aggregating road sensor data, such as the Archived Data Management System (Giuliano, Chakrabarti, and Rhoads, 2016) in Los Angeles, California, but this method has helped enhance accessibility measurements for only automobiles, not other modes like public transit, walking, and biking.

Data on public transit stop location, service, and performance on an aggregate basis over multiple time periods are scant, limited to a few of the largest and most sophisticated transit operators in the United States, housed on multiple websites and in a variety of data types. A new data source, the General Transit Feed Specification (GTFS), has solved many of these issues simultaneously by providing a centralized database of current and historical transit stop locations, service times, and performance using the same file type. GTFS enables public transit accessibility measurements that were previously impossible or impractical. In a recent research project, which required obtaining the location of every bus stop in California, GTFS reduced data-gathering time from 3 months to 1 week and increased data completeness from 49 to 88 percent of counties (Bostic and Rodnyansky, 2016). The remainder of this article gives background and tips on using GTFS, showcases relevant research using GTFS, and provides ideas for future use in housing and neighborhood research.

GTFS Background

GTFS is an open-data tool connecting transit operators and users of transit data. GTFS is a unified standard file format for sharing transit route, stop, schedule, and performance data, interoperable among transit operators worldwide, regardless of size, language, or transit type.[1] The GTFS file format was developed through a partnership between TriMet, the transit operator in the Portland, Oregon metropolitan area, and engineers at Google, Inc. (McHugh, 2013). This initial project helped launch and integrate transit tracking and transit time measurement into Google Maps (McHugh, 2013). The public and institutional success of the Portland example encouraged other U.S. and international transit agencies to adopt the GTFS file format and provide their data to the public. Since the launch of the Portland project, data from at least 1,000 agencies worldwide have become available through GTFS. Moreover, an extension of GTFS called "GTFS-realtime" expands the available data to include real-time trip updates, service alerts, and vehicle positions, enabling more nuanced analyses.[2]

GTFS Do-It-Yourself

GTFS is a simple and accessible tool, by design. Researchers, policymakers, and others interested in transit data can download and perform analyses on GTFS data with tools as simple as spreadsheet software (for example, Microsoft Excel), text reader (for example, Windows Notepad), and GIS software (for example, Esri ArcGIS, PitneyBowes MapInfo, or open source QuantumGIS).

Each participating transit agency uploads as many as 13 files describing the various aspects of their transit operations. These files include stop locations and times, service frequencies, routes and route shapes, trips, fare attributes and rules, transfers, service calendars and off days, and agency and feed descriptions.[3] Those agencies participating in GTFS-realtime include additional files describing in-time vehicle positions, trip updates, and service alerts. Note that not all agencies choose to upload every file, and not all agencies update their files with every service change.[4] The files for each participating agency are downloadable in text, comma separated value, shapefile, or a combination of the three formats, depending on the agency and file type.

After downloading the data, users have several options in operationalizing it, depending on their purpose. At the simplest level, users can examine their data in a spreadsheet or a text reader, if they know the specific route or stop they seek. Most users, however, will want the view of all the stops and all the routes. To enable proper location of stops and routes on a map, GTFS data provide latitude and longitude coordinates. For agencies that provide GIS shapefiles, coordinates and projections will appear automatically once opened in GIS. For agencies that provide text files, users should import the text file into GIS, use the system's coordinate reader, and set a projection.[5] These

[1] https://developers.google.com/transit/gtfs/.

[2] https://developers.google.com/transit/gtfs-realtime/guides/feed-types.

[3] https://developers.google.com/transit/gtfs/examples/gtfs-feed.

[4] Some agencies update their GTFS data frequently. For example, the Sacramento Rapid Transit District has GTFS data posted from six time periods from 2013 through 2017 (SACRT, n.d.).

[5] Vance (2016) provides a handy tutorial for QGIS using the Cook County, Illinois Pace bus service. This tutorial readily generalizes to ArcGIS and other GIS software.

files can now be integrated with other geographic data. Alternatively, GTFS files can also be converted to a KML file for use with Google Earth.[6] To take advantage of the timetables and schedules provided via GTFS, users can plug GTFS into a network dataset for use with Esri ArcGIS Network Analyst, using a custom-written toolbox "Add GTFS to a Network Dataset" (Morang, n.d.).

The U.S. Department of Transportation (DOT)'s National Transit Map project, through the Bureau of Transportation Statistics, provides a GTFS-derived national map of all transit stops and all agencies whose data are represented (DOT, n.d.). A GIS shapefile of 0.5- and 0.25-mile buffers around fixed-route transit stops is also provided and may be especially useful for those studying transit-oriented development and accessibility at the national level.

For users looking for data from a specific transit agency, no single website provides all GTFS-participating agencies, due to the data's open-source nature. However, several sources provide overlapping lists of participating agencies and clickable links to download data directly or to the transit agency's webpage on which the data are hosted (exhibit 1). Housing and neighborhood policy researchers can use these sources to complement their analyses with realistic and detailed portrayals of transit accessibility.

To demonstrate an example, I set out to find and display all the bus stop locations in Fresno and Madera Counties in the San Joaquin Valley of California. Fresno County, with a population of 950,000, contains the Fresno metropolitan area, and Madera County, with a population of 150,000, contains the city of Madera. Fresno County has two main transit operators—Fresno County FAX and Clovis Transit—regularly operating 20 routes, and Madera County has one—Madera County Transit, with three routes. Using the GTFS Data Exchange listing, I sourced and downloaded the data from each county's transit feed;[7] both were already in shapefile format upon download. I imported both shapefiles into open-source QuantumGIS, in addition to layers showing county and city boundaries and a layer showing major roads. Exhibit 2 visualizes both counties' transit stops with dots demarcating stop locations. Researchers can readily replicate and extend such an analysis and visualization with the available GTFS data.

Exhibit 1

Sources To Find GTFS Data for Specific Transit Agencies

Source Name	Total Transit Agencies Listed	Website
TransitLand	2,090; about 1,000 have GTFS	https://transit.land/feed-registry/
GTFS Data Exchange	1,000	http://www.gtfs-data-exchange.com/agencies
Transitfeeds	550	http://transitfeeds.com/feeds
Transitwiki.org "Publicly accessible public transportation data"	401	https://www.transitwiki.org/TransitWiki/index.php/Publicly-accessible_public_transportation_data
Google Code Archive: googletransitdatafeed—PublicFeeds.wiki	256	https://code.google.com/archive/p/googletransitdatafeed/wikis/PublicFeeds.wiki
Trillium	150 or more	https://trilliumtransit.com/gtfs/our-work/

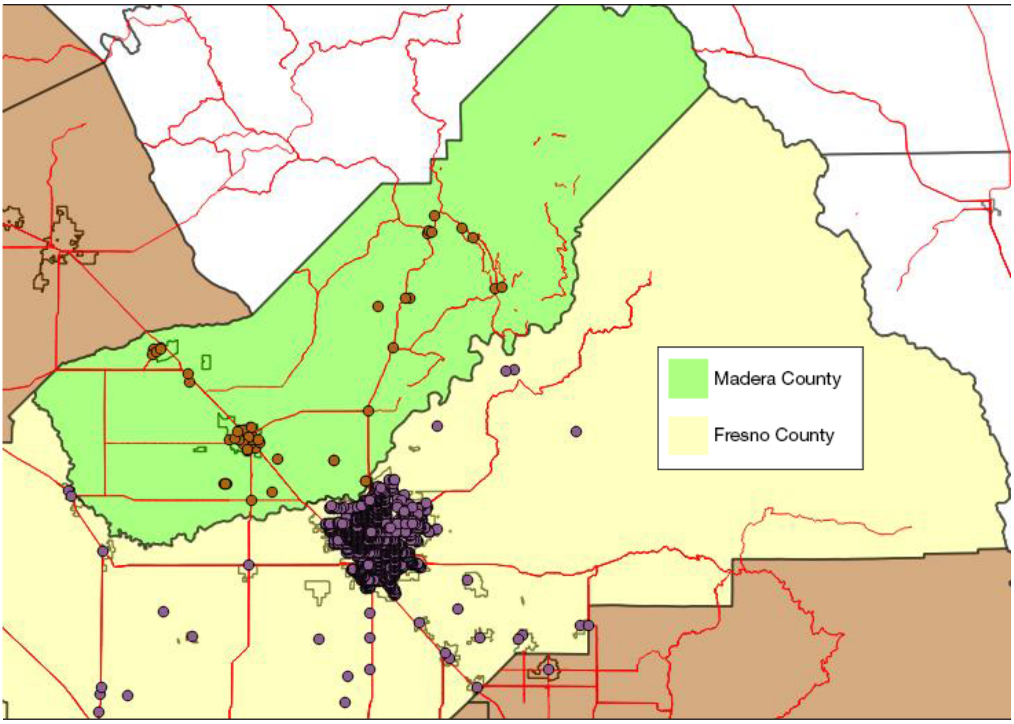
GTFS = General Transit Feed Specification.

[6] See Antrim (2015).

[7] Fresno County data feed: http://data.trilliumtransit.com/gtfs/fresnocounty-ca-us/; Madera County data feed: http://data.trilliumtransit.com/gtfs/madera-ca-us/.

Exhibit 2

Visualization of Bus Stop Locations in Fresno and Madera Counties Using GTFS Data



GTFS = General Transit Feed Specification.

Current Uses of GTFS

Current users of GTFS range widely from transit agencies, app and website developers, transit planners, researchers, and others. DOT's Federal Transit Administration reports that the "GTFS format is used by many transit agencies to communicate their schedules to online mapping programs and smartphone/tablet applications that travelers use to plan their transit trips" (DOT, 2016). GTFS.org, a website supported by the nonprofit Rocky Mountain Institute, reports six major applications of GTFS: (1) trip planning and maps, (2) timetable creation, (3) accessibility for the disabled, (4) planning and analysis, (5) real-time transit information, and (6) public information displays (GTFS.org, n.d.).

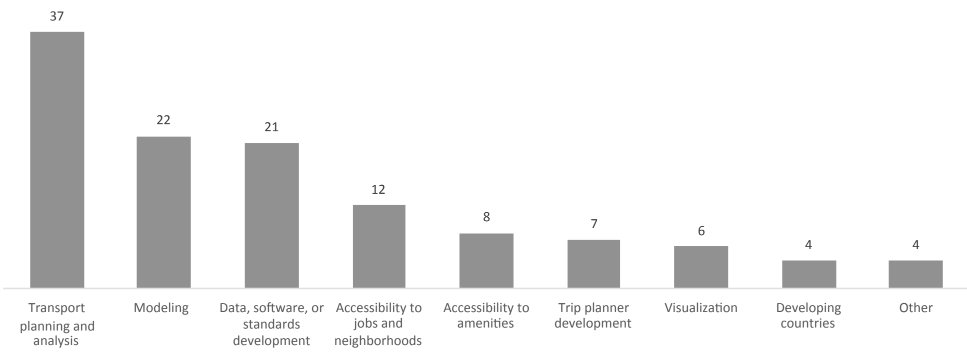
The previously mentioned categories intuitively make sense but do not give an understanding of the use of GTFS in research. To assess the depth and breadth of GTFS penetration into research, I conducted a literature scan of GTFS-related scholarly publications, agency reports, and unpublished working papers and theses. I used the GoogleScholar and Google search engines with a timeline from 2005[8] to the present and a targeted list of search terms: GTFS, General Transit Feed Specification, Google Transit Feed Specification, GTFS and geography, GTFS and planning, GTFS

[8] The earliest development of GTFS by TriMet and Google was in 2005.

and urban, GTFS and rural, and GTFS and transit. The literature scan yielded 121 relevant results, most from 2013 through 2017, published in a variety of journals and fields. Studies were categorized based on their main research angle. More than one-half of the research was specific to the transportation field: transportation planning and analyses, improved transportation modeling, and developing trip planners (exhibit 3). Another 21 studies focused on software, standards, or data development stemming from GTFS, and 6 studies used GTFS for visualization. Finally, less than 20 percent of studies focused on accessibility to jobs, neighborhoods, or amenities—our topic of interest. These accessibility-related studies are most relevant for housing, planning, and neighborhood researchers and set a precedent for incorporating GTFS in such research.

Exhibit 3

Counts of GTFS-Related Research by Category



GTFS = General Transit Feed Specification.

Accessibility to Amenities

Select scholars have used GTFS data to more accurately measure public transit accessibility to amenities including healthcare facilities, grocery stores, schools, and retail. Given the reliance of many low-income households on public transit, these studies help inform policymakers on the disparities in amenity access between driving and transit modes. A 2015 study in Melbourne, Australia, was among the first to demonstrate the potential of GTFS to measure amenity access. Rocha et al. (2015) assessed the transit accessibility of emergency dental care by differentiating patients by socioeconomic status and proximity to high-frequency bus stops. Using GTFS, they found that households living in areas with no high-frequency bus stops were no less likely to seek emergency dental care than those living in areas with high-frequency bus stops. A study of shopping amenity access in Montgomery and Prince George's Counties in Maryland found, using GTFS data, store locations to be "strongly influenced by access to transportation facilities, especially bus and light rail transit stops" (Ma, Knapp, and Knapp, 2014: 2). Mechaber (2015) found that access to selective enrollment and magnet high schools in Chicago was inequitable, when taking into account GTFS-derived public transit travel times, because it takes longer to get to these schools by transit than by car. School locations were found to be inequitable with respect to minority status and income, because lower-income and minority pupils both lived farther away from the schools and were more reliant on transit to get to school.

Another set of studies compared spatiotemporal accessibility to grocery stores via public transit versus driving, using GTFS data from the Cincinnati, Ohio metropolitan area. They found that food deserts for transit-dependent households change shape depending on the time of day and day of the week, due to the transit schedule, while car-dependent households have a more fixed food desert definition (Farber, Morang, and Widener, 2014). Further, Widener et al. (2015) found that many Cincinnati-area residents have improved accessibility to supermarkets if they access them on their return trip from work rather than departing from home. Widener (2017) extended this research to show that transit-dependent households have poorer spatial access to healthy food, as measured by the cumulative access to multiple grocery stores, and that transportation to obtain the healthy food is more costly using transit compared with driving modes. A followup study in the greater Toronto, Canada area finds that grocery store access in late nights and early mornings is lower for transit users than for car users (Widener et al., 2017). This finding is relevant because many lower-wage service employees have unconventional schedules and lower access to cars, necessitating transit accessibility to grocery stores at very late or very early hours. These studies barely scratch the surface of what is possible with GTFS in measuring amenity accessibility.

Accessibility to Jobs and Neighborhoods

A limited number of studies have used GTFS to study neighborhood and employment accessibility—major topics of interest for housing, planning, and neighborhood researchers. A cross-sectional analysis by Owen and Levinson (2014) used GTFS to rank 46 of the top 50 most populous U.S. metropolitan areas by average transit accessibility to jobs. They provided a realistic weighted average transit-travel time between residences and employment areas, from 7:00 to 9:00 a.m., including getting to and from public transit and any necessary transfer. New York, New York; San Francisco and Los Angeles, California; Washington, D.C.; and Chicago, Illinois, top the list in job accessibility by transit (Owen and Levinson, 2014). A San Francisco Bay Area simulation analysis found that the region has high job accessibility by walking and by transit, but disparities existed between census blocks in poverty and not in poverty (Blanchard and Waddell, 2017). Ma and Knaap (2014) demonstrated the use of GTFS and Open Street Map, another open data source, on neighborhood-level job accessibility for neighborhoods surrounding the proposed Purple transit line in the Maryland suburbs of Washington, D.C. Their model showed outcomes for two planned stations. Langley Park, a neighborhood with a high proportion of low-skilled workers, would see an 80-percent increase in low-skilled jobs accessible by transit when the transit line opens. For Bethesda, a regional employment center, 70,000 more employees would be within 1 hour on public transit once the line opens.

Several studies addressed transit equity at the neighborhood level by assessing differences in transit supply versus transit demand. Jiao and Dillivan (2013) found “transit deserts”—neighborhoods that lack adequate transit service but maintain high proportions of transit-dependent populations—near the historic downtowns and isolated rural areas of the Charlotte, North Carolina; Chicago; Cincinnati; and Portland metropolitan areas. Kahrobaei (2015) found that low-income neighborhoods with a high proportion of commuting workers have low bus frequency during the morning rush, necessitating a high degree of car ownership. Others proposed and evaluated a Gini coefficient-like measure of transit supply distribution equity, tested using GTFS data (Bertolaccini, 2013; Bertolaccini and Lownes, 2013).

Research relating housing to transit accessibility using GTFS has been limited. Zhong et al. (2017) harnessed GTFS to derive a model for optimally siting affordable housing to maximize residents' access to public transit and reduce geographic clustering of affordable housing, subject to land acquisition and construction budget constraints. Tested for Tempe, Arizona, the model "could provide insightful spatial decision support to affordable-housing providers or tax-credit administrators, facilitating the design of flexible strategies that address multiple social goals" (Zhong et al., 2017: 1). A study of California low-income housing tax credit affordable housing sites showed no statistical difference in transit accessibility between allocated (9-percent) and tax-exempt bond-derived (4-percent) tax-credit projects or between funded and nonfunded 9-percent tax-credit projects (Bostic and Rodnyansky, 2016). More research relating housing and transit accessibility is needed for both affordable and market-rate housing, and GTFS makes such research more attainable.

Conclusion

Transit accessibility is an important topic for housing researchers and policymakers at the national, regional, and local levels. GTFS data improve the geographic coverage, depth, and accuracy of measuring transit accessibility. Researchers and practitioners can use GTFS in a do-it-yourself manner, which was previously unavailable, to include better measures of transit accessibility into their analyses. These better measures are especially needed in housing-related research, which is underrepresented in its use of GTFS.

Acknowledgments

The author thanks session participants at the American Association of Geographers 2015 conference for helpful comments and Anna Sporrong for excellent research assistance.

Author

Seva Rodnyansky is a Ph.D. candidate in urban planning and development at the University of Southern California, Sol Price School of Public Policy.

References

Alonso, William. 1964. *Location and Land Use*. Cambridge, MA: Harvard University Press.

Antrim, Aaron. 2015. "Importing GTFS Data Into ArcGIS." Trillium. https://trilliumtransit.zendesk.com/hc/en-us/articles/201382609-Importing-GTFS-data-into-ArcGIS.

Bertolaccini, Kelly L. 2013. "Assessing the Equity of Transit Supply Distribution in Metropolitan Areas Using Lorenz Curves and Gini Coefficients." Master's thesis, civil engineering, University of Connecticut. http://digitalcommons.uconn.edu/gs_theses/483.

Bertolaccini, Kelly L., and Nicholas E. Lownes. 2013. "Effects of Scale and Boundary Selection in Assessing Equity of Transit Supply Distribution," *Transportation Research Record: Journal of the Transportation Research Board* 2350: 58–64.

Blanchard, Samuel D., and Paul Waddell. 2017. "Assessment of Regional Transit Accessibility in the San Francisco Bay Area of California With UrbanAccess," *Transportation Research Record: Journal of the Transportation Research Board* 2654: 45–54.

Bostic, Raphael W., and Seva Rodnyansky. 2016. Leveraging Federal Tools To Advance Local Goals: The Low Income Housing Tax Credit as a Case Study. Unpublished paper.

Brueckner, Jan K. 1987. "Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model." In *Handbook of Regional and Urban Economics*, Vol. 2, edited by Edwin S. Mills. Amsterdam: Elsevier: 821–845.

Farber, Steven, Melinda Z. Morang, and Michael J. Widener. 2014. "Temporal Variability in Transit-Based Accessibility to Supermarkets," *Applied Geography* 53: 149–159.

Giuliano, Genevieve, Sandip Chakrabarti, and Mohja Rhoads. 2016. "Using Regional Archived Multimodal Transportation System Data for Policy Analysis: A Case Study of the LA Metro Expo Line," *Journal of Planning Education and Research* 36 (2): 195–209.

Giuliano, Genevieve, and Kenneth Small. 1991. "Subcenters in the Los Angeles Region," *Regional Science and Urban Economics* 21 (2): 163–182.

GTFS.org. n.d. "Getting Started." http://gtfs.org/getting-started/#gtfs-applications.

Haas, Peter, Stephanie Morse, Sofia Becker, Linda Young, and Paul Esling. 2013. "The Influence of Spatial and Household Characteristics on Household Transportation Costs," *Research in Transportation Business & Management* 7: 14–26.

Jiao, Junfeng, and Maxwell Dillivan. 2013. "Transit Deserts: The Gap Between Demand and Supply," *Journal of Public Transportation* 16 (3): 2.

Kahrobaei, Sina. 2015. "Spatial and Temporal Measures of Mismatch Between Transit Supply and Employment for Low Income and Auto Dependent Populations." Master's thesis, civil engineering, University of Connecticut. http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1918&context=gs_theses.

Ma, Ting, Eli Knaap, and Gerrit-Jan Knaap. 2014. *Retail Location and Transit: An Econometric Examination of Retail Location in Prince George's and Montgomery County, Maryland*. College Park, MD: University of Maryland, National Center for Smart Growth. http://smartgrowth.umd.edu/assets/documents/research/retail_location_and_transit_100514_ek_gk.pdf.

Ma, Ting, and Gerrit-Jan Knaap. 2014. "Analyzing Employment Accessibility in a Multimodal Network Using GTFS: A Demonstration of the Purple Line, Maryland." College Park, MD: University of Maryland, National Center for Smart Growth. http://smartgrowth.umd.edu/assets/documents/research/acsp_tingma_20141029.pdf.

McHugh, Bibiana. 2013. "Pioneering Open Data Standards: The GTFS Story." *Beyond Transparency*. http://beyondtransparency.org/chapters/part-2/pioneering-open-data-standards-the-gtfs-story/.

Mechaber, Elias. 2015. "When Choice Is Not a Choice: A Geographic Accessibility Approach to Spatial Inequalities Concerning Magnet and Selective Enrollment Public High Schools in the City of Chicago." Undergraduate thesis, geography, University of Chicago. http://eliasmechaber.com/wp-content/uploads/2015/04/Mechaber-Elias_Thesis-Geography_2015.pdf.

Morang, Melinda. n.d. "Add GTFS to a Network Dataset—Overview." http://transit.melindamorang.com/overview_AddGTFStoND.html.

Owen, Andrew, and David M. Levinson. 2014. *Access Across America: Transit 2014*. CTS 14-11. Minneapolis: University of Minnesota, Center for Transportation Studies Research. http://hdl.handle.net/11299/168102.

Redfearn, Christian L. 2007. "The Topography of Metropolitan Employment: Identifying Centers of Employment in a Polycentric Urban Area," *Journal of Urban Economics* 61: 519–541.

Rocha, Carla M., Estie Kruger, Shane McGuire, and Marc Tennant. 2015. "Role of Public Transport in Accessibility to Emergency Dental Care in Melbourne, Australia," *Australian Journal of Primary Health* 21 (2): 227–232.

Sacramento Rapid Transit District (SACRT). n.d. "GTFS." http://iportal.sacrt.com/GTFS/SRTD/.

U.S. Department of Transportation (DOT). 2016. "STOPS—General Transit Feed Specification (GTFS) Data." Federal Transit Administration. https://www.transit.dot.gov/funding/grant-programs/capital-investments/stops-%E2%80%93-general-transit-feed-specification-gtfs-data.

———. n.d. "National Transit Map (Data, Maps and Apps)." Bureau of Transportation Statistics. https://www.rita.dot.gov/bts/ntm/map.

Vance, Steven. 2016. "Converting a Transit Agency's GTFS to Shapefile and GeoJSON with QGIS." Steven Can Plan. http://www.stevencanplan.com/2016/02/converting-a-transit-agencys-gtfs-to-shapefile-and-geojson-with-qgis/.

Widener, Michael J. 2017. "Comparing Measures of Accessibility to Urban Supermarkets for Transit and Auto Users," *The Professional Geographer* 69 (3): 362–371.

Widener, Michael J., Steven Farber, Tijs Neutens, and Mark Horner. 2015. "Spatiotemporal Accessibility to Supermarkets Using Public Transit: An Interaction Potential Approach in Cincinnati, Ohio," *Journal of Transport Geography* 42: 72–83.

Widener, Michael J., Leia Minaker, Steven Farber, Jeff Allen, Brigitte Vitali, Paul C. Coleman, and Brian Cook. 2017. "How Do Changes in the Daily Food and Transportation Environments Affect Grocery Store Accessibility?" *Applied Geography* 83: 46–62. https://doi.org/10.1016/j.apgeog.2017.03.018.

Zhong, Qing, Alex Karner, Michael Kuby, and Aaron Golub. 2017. "A Multiobjective Optimization Model for Locating Affordable Housing Investments While Maximizing Accessibility to Jobs by Public Transportation," *Environment and Planning B: Urban Analytics and City Science*. DOI: 2399808317719708.