Rapid Bus Service Allocation in Los Angeles:

Spatial Analysis of Transit Service Patterns

Lucas J Smith

University of Southern California, PPD-427

Introduction

In any area as populous as Los Angeles, transportation is always a significant policy issue. Like many cities, transportation has long been part of the narrative about Los Angeles; from the nostalgia for the days of Pacific Electric Red Cars to the popular notions of "car culture" and legendary traffic to the region's new push for rail in the past few decades. While the rail and subway openings may steal the spotlight, a great deal of the transit system's work is done by the bus network, and local agencies have been innovative in that realm as well, particularly with the Metro Rapid bus product.

Metro Background

The Los Angeles County Metropolitan Transportation Authority, or "Metro," operates bus and rail public transit in Los Angeles County. In addition to Metro, there are multiple municipal transit agencies that operate transit service in Los Angeles County cities. In some parts of the county these municipal carriers are the only transit operators, and in other areas Metro and municipal service is available. While the Los Angeles (City) Department of Transportation provides some bus service, the vast majority of transit service is provided by Metro buses and rail. The sometimes nebulous boundaries and overlap of the Metro district are significant to the analysis discussed in this paper, because it only considered transit service provided by Metro.

Metro provides a variety of different service types. The Metro Rail system consists of the Red and Purple Line subways; the Blue, Green, Gold, and Expo Line light rail lines; the Orange and Silver Line "Transitway" services consisting of buses operating in dedicated rights-of-way; many typical Local bus routes, and a handful of Metro Rapid bus routes.

Metro Rapid is special type of bus rapid transit service intended to provide a better quality of service than traditional local buses, on similar routes. Rapid routes, though they operate in mixed traffic with automobiles, stop less frequently than a typical bus and feature special branding at stops and in bus design. Some routes have signal priority at stoplights, and some operate partially in bus-only lanes. These upgrades over regular bus service allow for faster operation and improved rider satisfaction (Los Angeles County Metropolitan Transportation Authority, 2002). Since introducing Metro Rapid through a pilot project beginning in 2000, the Metro Rapid network has grown to 21 routes currently operating.

Question and Thesis

Among the transit products Metro provides, there is a clear hierarchy among the different modes in terms of the quality of service provided to customers. At the high end of this scale are the subway lines, and at the low end are buses. Among the bus routes, Rapid routes are certainly more desirable despite using essentially the same equipment as Local buses.

Metro has created a network of Rapid buses that spans most of the agency's service area, but is Rapid service allocated fairly? This paper and analysis attempts to quantify which parts of Los Angeles County are relatively underserved and "over-served" by Metro Rapid service. I attempt to assess which parts of the county deserve more Rapid service than they currently enjoy, and which areas Metro has been overemphasizing. To this end, I calculated the ratio of Rapid-or-better service, measured in number of arrivals per day of a Rapid bus or train, to the population density of each Census block group in Los Angeles County.

Methodology

Data Acquisition

The analysis drew on various datasets from different sources:

- Polygons representing Census Block Groups with demographic data from the American Community Survey from the Census Bureau.
- Points and polylines representing Metro bus and rail lines and stops, from Metro.
- Transit schedule data for all Metro bus and rail lines, in General Transit Feed Specification (GTFS) format, from Metro.

The raw data for this analysis was easily obtained. The U.S. Census Bureau provides geodata through its TIGER/Line websites. For this project, I downloaded a geodatabase containing California block groups joined with data from the 2008-2012 American Community Survey 5-year estimates, a dataset that contained far more survey data than was needed for this analysis but reduced the steps necessary to join population data to block group geometry (U.S. Census Bureau).

Metro provides a great deal of data to the public through its Metro Developer blog at developer.metro.net. From this site, I downloaded shapefiles for each Metro Rail line, all bus lines, and all bus stops and rail stations. The route data was in the form of polylines representing each route, with metadata detailing route numbers. The bus stops and rail stations were in the form of points for each station or stop, with metadata detailing stop ID numbers and every bus or rail route that served the stop. The Metro Developer blog was also the source for the GTFS files, which systematically describe the entire Metro transit system with a series of text files designed to be machine-readable.

The General Transit Feed Specification.

I assume readers of this paper are familiar with shapefiles, but I expect some explanation of the General Transit Feed Specification, or GTFS, is warranted. Originally developed in partnership between Google and the Tri-Met public transit agency in Portland, Oregon, GTFS has become the standard format for transit data. A GTFS feed, as the collection of files is called, can precisely describe all routes, stops, schedules, fares, transit centers, transfers, and every other aspect of a transit agency's operation. A feed is comprised of between six and 13 (depending on the level of detail encompassed by the feed) comma-separated text files saved with the .txt extension. Since transit agencies can have highly complex operations, the GTFS specification is designed to minimize repeating of data. This is useful for storage space, but it means that only in combination of agency.txt, stops.txt, routes.txt, trips.txt, stop_times.txt, calendar.txt, and in some cases calendar_dates.txt files can any human-readable information be gleaned from the feed. This means extensive processing is necessary for any comprehensible use of GTFS data.

Among Los Angeles area transit operators, most do not distribute GTFS data. Metro is one of the leaders, however, and offers a very complete feed. Metro's current GTFS feed contains nine files, which range in length from two lines of text (a 148 byte file) to 2,229,367 lines of text (a 154 megabyte file). The stop ID numbers in the GTFS feed are consistent with the stop ID numbers in Metro's shapefiles, which made this analysis possible.

Data Conversion and Refinement

Processing GTFS data with Python.

The purpose of including GTFS data in my analysis was to quantify the intensity of Rapid-or-better service at each stop in Los Angeles County, to then be generalized to intensity of Rapid-or-better service within each Los Angeles County block group. While technically possible to accomplish by hand using schedules, the GTFS feed makes it practically possible for a system the size of Metro's. Deriving the frequency/intensity of service data required data processing beyond the scope of ArcGIS or any other ready-built software I am aware of, so I wrote a custom Python script to execute the core of the processing. This section details the processing carried out by the script.

The script calls on Python's built-in CSV library, so the first step in preparing the GTFS text files is to change their file extension to .csv.[1] This is a manual step. Then, the script can be run on the files. In practice, I ran the script separately for each type of service (Rapid, Rail, Transitway) to reduce the chance of error and the amount of time the script was processing without giving the user feedback on failure or success.

First, the script generates a list of the Route IDs that correspond to the type of service (Rapid, Rail, Transitway) specified by the user. Due to a quirk of the Metro GTFS feed, all bus Route IDs are the route number plus "-13074" (i.e., the Route ID for route 754 is "754-13074"). Next, the script reads through every trip in trips.csv, filtering it down to a list of all Trips that correspond to the Route IDs of interest and are on weekdays. These Trips each consist of a full row from the trips.csv file, so they then are processed to extract the Trip IDs from each Trip. Next, the script reads through the

[1] This is not technically necessary as a function of the CSV library, but in using my script it is now necessary due to some hard-coded ".csv" filenames in the script's code.

stops.csv file to generate a list of Dictionaries of the Stop IDs for every bus and rail stop listed in stops.csv (i.e., every bus stop and rail station in the Metro system).

The next step in the process is the real meat of the data processing. The script opens the massive stop_times.csv file (over 2 million lines for Metro) and loops through each line of it. Each line in this file represents an arrival of a vehicle at a stop or station. Each arrival's Trip ID is checked against the list of Trip IDs of interest, and if there is a match, the Dictionary entry corresponding to the arrival's Stop ID and Route ID is incremented. This produces a list of Dictionaries that count the number of stops each route of interest stops at each stop. The CSV library's DictWriter method is called on this list of Dictionaries, and it saves a CSV table of the counts for each route.

The final step in the processing is conducted manually. Using Excel, I added a TOTAL column to each table, and summed the arrival counts across every route. This column was converted to numbers and the letter "r" was prepended to each route number header to prepare the data for seamless import into ArcGIS.

Refining stops and routes.

Metro's bus stop shapefile includes all stops for Rapid, Local, Express, and Transitway buses. For this analysis, I needed to create layers for Rapid stops and Transitway stops. This was accomplished using definition queries in ArcMap and exporting the results to new feature classes.

Trimming and culling TIGER/Line.

Before importing it into ArcMap, the block groups TIGER/Line + American Community Survey feature class was reduced in size; initial attempts to import the file resulted in crashes. Using ArcCatalog, I trimmed the feature class to Los Angeles County based on the field corresponding to county ID numbers, and I deleted a plethora of unneeded ACS survey fields from the attribute table.

Spatial Analysis & Geoprocessing

Conceptually, the analysis consisted of determining how many times a Rapid-or-better service stopped in the area accessible to each block group on a weekday, and comparing that amount to the relative population density of each block group. Representationally, this meant counting the number of arrivals of a Rapid-or-better vehicle at every stop, totaling the number of vehicle-arrivals at all stops within half a mile of the centroid of each block group, and dividing this number by the density (expressed as people per square mile).

With the source datasets in order, it was now possible to begin the real spatial analysis at the core of this project. At this point, the block groups geographic and demographic data was joined and ready, the relevant lines for transit service and points for transit stops were sorted from the irrelevant lines and stops, and the tables of service intensity had been distilled from the GTFS feed and were in a format that worked with ArcGIS.

The first step in ArcGIS was to join the intensity of service tables to the bus stops and rail stations layer. This was a simple join thanks to the consistent numbering of Stop IDs across datasets from Metro.

Next, I created half-mile buffers around each Rapid bus stop and rail station point, keeping the counts of total vehicle arrivals from the stops and stations. Often, quarter-mile buffers are used to estimate walksheds for transit stations, but I believe half-mile buffers were appropriate in this case. They would be combined with centroids of the block groups, so it seemed more likely that a quarter-mile buffer would miss a centroid it should have captured than a half-mile would capture a centroid it should not. Since this analysis was focused on comparing relative service intensity throughout the county using something that amounts more to an index than to a more specific statistic, I thought it made sense to systematically over-count walksheds than systematically undercount. These buffers overlapped one another quite a bit, and most of the county was covered by more than one buffer.

After this, centroids were generated for the block group layer. Because most block groups are relatively small in size, the use of centroids distorts the analysis less than it would if larger geographies were used. However, there are some large block groups in Los Angeles County, but since the Census creates more block groups in denser areas most large block groups were in non-dense areas that largely did not have Rapid-or-better transit service.

Next, the Spatial Join operation was employed to join the counts from the transit stop buffers to the block group centroids they captured. Due to idiosyncrasies of the different shapefiles and GTFS count fields, these joins had to be done once for each rail line, and once for Rapid buses, creating separate feature classes for each one, and multiple points for each captured block group within these feature classes.[2] I then merged these multiple feature classes into one "grand total" feature class. Finally, I used the Many-to-One Spatial Join tool to join the multiple transit count centroid points back to their original block groups.

---

[2] The multiple points per block group issue could have been avoided with better use of the Many-to-One setting in the Spatial Join tool, but I misunderstood the operation of this tool until it was too late. No harm was done, but it did affect the way subsequent steps had to be carried out.

Lastly, I created Area In Square Miles and Population Density fields in the block group layer, which were populated using Field Calculator (AREA and Population / Area, respectively).

Design and Symbology

The final step in preparing the map was to symbolize the information I had gleaned from the data. The final map I created symbolized each block group's index of service intensity normalized by its population density if it has any Rapid-or-better service, or its population density if it does not have any Rapid-or-better service. This allows viewers to see what areas have disproportionately low or high levels of service, as well as which areas should be high priorities for adding future lines. The block groups are overlaid with dotted grey lines showing Rapid bus lines and rail lines.

The normalized intensity-of-service index is symbolized by a green-yellow-red color ramp, divided into seven classes by quantile. Thus, a block group symbolized with yellow fill has an intensity of service relative to density that is near the median for the county. Those closer to green have more service relative to density, and those closer to red have less. The three block groups with the highest intensity-of-service index relative to density were not symbolized because they were outliers in large size and low population.

The areas that are currently not served by Rapid-or-better transit were symbolized with a light-to-dark grey color ramp, divided into six classes using Jenks Natural Breaks Optimization, manual rounded slightly. Denser areas are indicated with darker shades of grey, and the classification was conducted across all block groups in the county, not just those with no Rapid-or-better service.

Limitations, Assumptions, and Uncertainty

There are multiple limitations to this analysis, mostly due to assumptions made in its preparation and the data available. It is necessary for one to understand these limitations prior to drawing conclusions based on this analysis.

A major limitation is caused by the consideration of only Metro service in this analysis. As stated earlier, Los Angeles County is home to many municipal transit operators in addition to Metro, and their service areas overlap to varying degrees. Some of the municipal operators operate Rapid services (Santa Monica Big Blue Bus and Foothill Transit are notable examples) but many do not. It could be argued that there is still a dearth of Rapid-or-better service in these areas, but it is tenuous to suggest that they are necessarily underserved by transit simply because they do not have Metro Rapid service.

Uncertainty can be caused by the nature of the use of buffers and centroids. Centroids are by no means a perfect representation of the block groups, particularly for large or odd-shaped block groups, of which there are plenty in Los Angeles County. Perhaps a half-mile as the crow flies is not the perfect indication of how well an area is served by transit. The representational uncertainty stemming from the centroids and buffers is not insignificant.

Further, simple population density is certainly not the ultimate way to determine where Rapid bus service should be provided. While it may be a good first criterion to consider, it is easy to imagine scenarios where it is not nearly perfect. Since Rapid bus service was partially designed to provide a better service in higher traffic, it is plausible there are areas where there is not enough traffic to make Rapid buses essential. Perhaps a transit agency sees its driving mission to be providing service to transit-dependent residents, leading them to "underserve" certain dense areas that have high levels of automobile access.

Results

From the map generated by this analysis, certain patterns appear. Downtown Los Angeles and all areas along rail lines have high intensity of service relative to population density. Another swath of bright green covers much of West Los Angeles, where the 720 Wilshire Rapid provides a high level of service. Notable sections of orange and red are visible in South Los Angeles and certain parts of the San Fernando Valley, indicating areas that have some Rapid-or-better service but less of it, relative to their density, compared to much of the area covered by Metro Rapid.

Among areas with no Rapid-or-better service, there are a few "hot-spots" of high density that are enclosed but far from existing Rapid routes, indicating that they are areas that could be served in the Metro service area. Some of the densest unserved block groups in the county are immediately to the west of Downtown. There are also stretches of high-density unserved areas south of Downtown and in the eastern portion of the Valley.

Solutions

The best solution for these underserved areas would be for Metro to simply add more Rapid service along existing corridors and create new routes in areas that do not have them now. Of course, it's not that simple and additional service is always subject to funding and politics. Perhaps reallocating some service hours from high-intensity relative to density Rapid routes to those with less intensity relative to density would be appropriate. It seems that the Westside may be enjoying more than its share of service, at the expense of South L.A.

One recommendation that seems logical is to fill the service gap between Downtown and Vermont north of Wilshire. This is an area that is especially dense and is surrounded on all sides by many connections to Rapid-or-better transit but is lacking in that high quality service. Even a short connector route could be useful here to funnel riders to downtown transfers or Red Line stations.

I think it is also clear that South L.A. is not getting a great deal from Metro. Between adding service or converting some Local bus service hours to Rapid bus hours in this area, Metro has an opportunity to improve the product it provides to this constituency.

One area that has high levels of service throughout is Downtown, but that should not change. Downtown transfers are an important aspect of Metro's service. Even if Los Angeles' downtown is not currently the center of housing and employment that most downtowns are, having an area where many bus and rail routes overlap is good for the system. It facilitates transfers for a wide variety of trips, and its representation here as disproportionately high is simply a result of how the representation used in this analysis would represent any downtown and should not be taken as an indication of a misallocation of service.

References

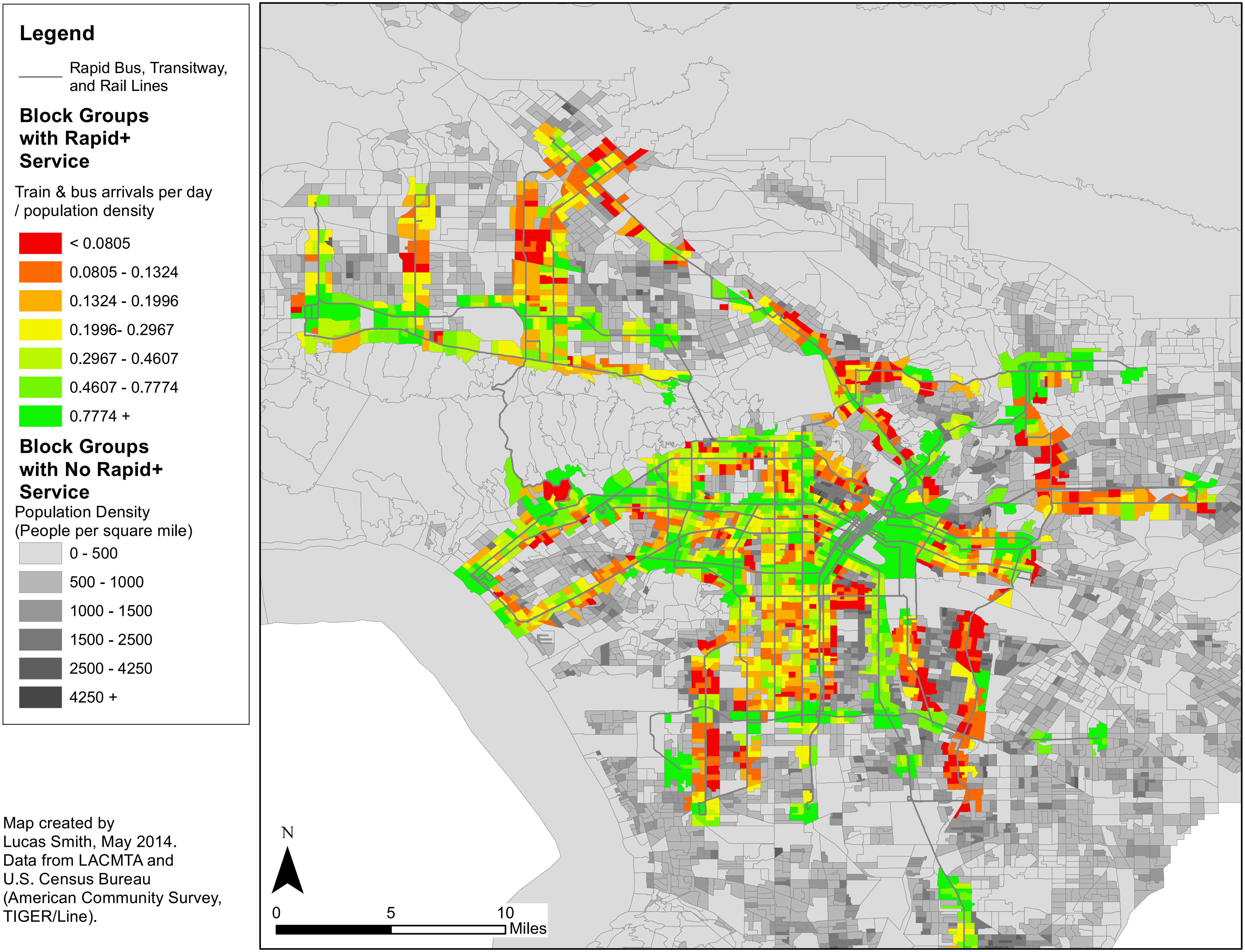
General Transit Feed Specification Reference. (2012). Retrieved from https://developers.google.com/transit/gtfs/reference

Los Angeles County Metropolitan Transportation Authority. (2002). *Final report: Los Angeles Metro rapid demonstration program.* Retrieved from http://media.metro.net/projects_studies/rapid/images/demonstration_program_report.pdf

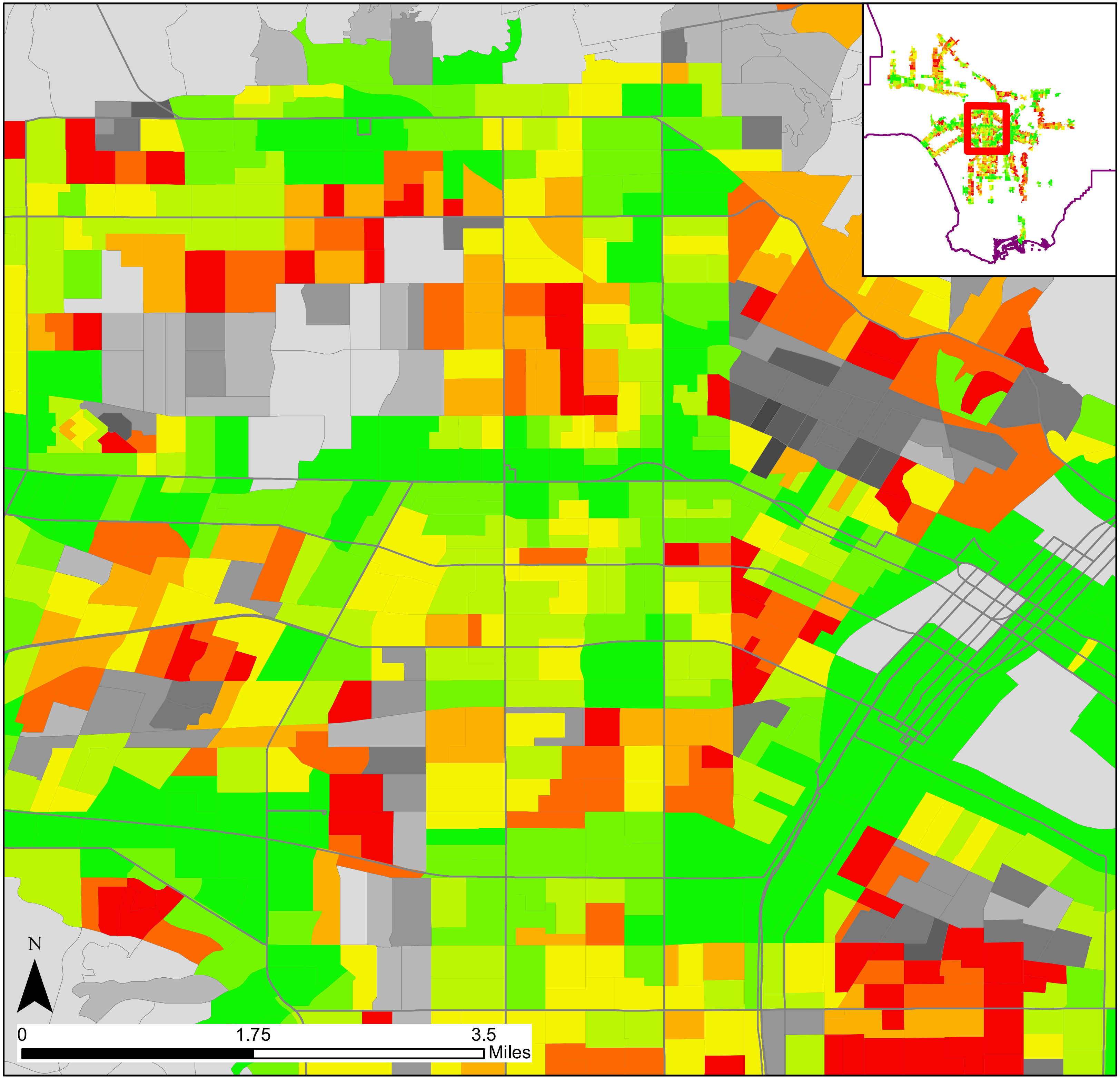
Metro developer: Metro's official blog of transit data and technology. Retrieved from http://developer.metro.net/

U.S. Census Bureau. *TIGER/Line Shapefiles pre-joined with demographic data.* Retrieved from https://www.census.gov/geo/maps-data/data/tiger-data.html

Roth, M. (2010). How Google and Portland's TriMet set the standard for open transit data. *Streetsblog SF.* Retrieved from http://sf.streetsblog.org/2010/01/05/how-google-and-portlands-trimet-set-the-standard-for-open-transit-data/



Map created by
Lucas Smith, May 2014.
Data from LACMTA and
U.S. Census Bureau
(American Community Survey,
TIGER/Line).



Map created by Lucas Smith, May 2014. Data from LACMTA and U.S. Census Bureau (American Community Survey, TIGER/Line).

Legend

Block Groups in Los Angeles County

Population Density
(People per square mile)

- 0 - 500
- 500 - 1000
- 1000 - 1500
- 1500 - 2500
- 2500 - 4250
- 4250 +

Map created by
Lucas Smith, May 2014.
Data from LACMTA and
U.S. Census Bureau
(American Community Survey,
TIGER/Line).