Appendix 1. Technical Methodology

This study combines detailed data on transit systems, demographics, and employment to determine the accessibility of jobs via transit within and across the country's 100 largest metropolitan areas.

Transit Data

For this analysis, an extensive original database was constructed to house information on routes, stops, and schedules for each eligible transit system within the nation's 100 most populous metropolitan areas—as defined by the U.S. Office of Management and Budget in 2008 and based on Census Bureau population estimates for that year.

Determining Transit System Eligibility

The most comprehensive source of information on transit agencies operating in the United States is the Federal Transit Administration's National Transit Database (NTD). The U.S. Code requires all agencies receiving federal support to report certain information to the NTD. Transit agencies and private providers that do not receive federal funding may also choose to be included in the database.[93] Every system registered in the NTD was eligible for inclusion in the analysis.

Systems in the 100 largest metro areas were identified based on two variables in the NTD: Urbanized Areas served and the ZIP code of the primary administrative office. That initial list was then analyzed using geographic information systems (GIS) mapping software to remove systems that do not actually operate in one of the 100 metro areas. Any systems that operate on a request basis as opposed to a scheduled route (e.g., vanpool and paratransit services) were also removed from the list.

Based on these eligibility criteria, 371 individual transit agencies were identified in the 100 largest metro areas, representing 10 modes of transit.[94] (Appendix 5 includes a full list of these providers.)

Collecting and Creating Model-Ready Transit Data

Data on routes, schedules, and stops for each of these 371 agencies were collected and cleaned over a period of several months, between May 2009 and February 2011.

Data collected for each transit agency were translated into a consistent format to ensure comparability. Google's open source format—General Transit Feed Specification (GTFS)—uses six to 12 text files that interact to provide a comprehensive representation of scheduled transit service.[95] Many of the country's largest transit systems publish their GTFS feeds. These pre-formatted feeds provided an important source of easily accessible data and significantly reduced the collection and cleaning workload.

Transit agencies without publicly available GTFS feeds were contacted directly to collect the relevant data. Some agencies provided the data in a GTFS format (though the feed had not been released publicly), but almost two-thirds did not. In these cases, they were asked to provide GIS shapefiles and digital schedules.

Some agencies without GTFS feeds maintain a comprehensive list of geocoded stop locations, stop order along routes, and associated digital schedules. This combination of digital schedules and shapefiles was only a few modifications away from a working GTFS feed.

When the shapefiles or digital schedules did not exist, GTFS feeds were created manually. In the case of digital schedules, this required transcribing PDF or paper schedules into an Excel file. Creating stop shapefiles was more challenging. Many agencies' data had stops without route or stop order fields. In these instances, those fields were created and populated manually in GIS. In scenarios without any stop locations, which invariably occurred with flag-a-bus systems, stops were drawn manually by following published route information, placing stops every two blocks and no more than every half mile in less populated areas.[96]

In the end, the database included 168 official system GTFS feeds and 203 manually created GTFS feeds.[97] Of the latter, over 130 systems included manually drawn stops.

Transit Model

Once data on schedules, routes, and stops were cleaned and standardized, the travel time between various origins and destinations was determined. This required specialized GIS software. Following an exhaustive survey of software alternatives, Brookings contracted with RAPIDIS, a Danish firm, to procure that firm's *Traffic Analyst* software. The software functions as an extension to ESRI's ArcGIS software suite.

Traffic Analyst offers multiple modeling features that made it an optimal choice for this analysis. First, it uses a GTFS import tool to model multiple transit systems within each metropolitan area and creates seamless interactions between different systems, effectively allowing riders to access all means of transit available to them within the region. Second, it enables the modeler to control elements like walking speed, number of transfer connections, time of day, and service date. Each of these elements is discussed in more detail below.

Origins and Destinations

The first step to modeling transit is to create origin-destination pairs. Within each metropolitan area, the analysis attempts to model travel via transit between every census block group (origin) and every census tract (destination).[98] The software uses block group population-weighted mean centroids (developed using Census 2000 block population data) as the origin point for each trip. Standard geographic centroids represent the census tract destinations.

Block groups, or "neighborhoods," not served by transit are removed from subsequent analysis. A block group is considered served if at least one transit stop falls within a 3/4-mile radius of its population-weighted centroid. The 3/4 mile radius is drawn around the block group centroid as a Euclidean buffer, which is a straight line distance that does not account for the street network or other geographic barriers.

While the literature sometimes uses 1/4 mile (bus) or 1/2 mile (rail) as the distance an individual will walk to access transit, consultation with experts for this project revealed that these distances are too short for many trips, especially for those individuals with no transportation alternatives.[99] Extending the distance to 3/4 mile better reflects the distances individuals will walk when faced with no transportation alternatives.[100] In addition, suburban block groups tend to be larger in land area than urban block groups. Because this analysis uses the centroids of block groups as the origin, shortening the distance of the "inclusion" buffer to 1/2 or 1/4 mile would disproportionately exclude suburban block groups that may cover more area yet still have transit options within walking distance for a large share of residents.

Walking Speed and Transfers

After establishing the origins and destinations, the analysis next connects each centroid to the nearest transit stop. These *connectors* are comparable to the time it takes an individual to walk from one's home to the closest transit stop, and then from one's destination transit stop to work. A connector is drawn from each origin centroid to any stop within 1,210 meters (roughly 3/4 mile). The same applies for destination centroids; however, if no stops are within that distance, then a connector is drawn to the single nearest stop of any distance.

The connectors are drawn in *Traffic Analyst* using Euclidean distances. Travel time for each connector is calculated using a formula of roughly *(5 km/hour) / 1.2*. The 5 km / hour is a generally accepted walking speed and the division by 1.2 controls for the use of Euclidean distances, which tend to be shorter than standard street network travel.

In addition to drawing origin and destination transit connectors, the model also generates stop-to-stop transfers. Using a limit of 400 Euclidean meters (roughly 3/4 mile), the model permits each stop to transfer to no more than eight neighboring stops. The transfers may exist between any system within the same metropolitan area, so there are no penalties for changing system operator. Finally, the model treats transfer walking speeds the same as connector walking speeds.

Travel Time

For each metropolitan area, the software models travel times over the course of a morning peak period commute—between 6:00 AM and 9:00 AM. According to the 2008 single-year American Community Survey (ACS) estimates, 65.6 percent of public transportation commuters in the 100 largest metropolitan areas leave for work during this three-hour period. In addition to the consistent time period, all travel is measured on a Monday, when transit service is typically more frequent than on weekend days. Any additional congestion due to morning rush hour commuting patterns, especially for bus routes, is accounted for in the system's published schedules.

Within the three-hour time period, the model uses a series of randomized "launches", or traveler departures, every five minutes.[101] This equates to 36 total launches from each covered block group to every tract in the three-hour window. These randomized launches mean that some travelers will "just catch the bus" while others just miss it. The final results of those 36 launches are then combined to determine an average travel time between each origin/destination pair.

Urban versus Rural

The 371 systems included in this analysis fall into two main categories of federal assistance as identified in the NTD—those in the Urbanized Area Formula Program (Urban) or the Rural Transit Assistance Program (Rural). The Urban systems are typically the major transit providers in these metropolitan areas (e.g., the New York Metropolitan Transit Authority (MTA)). The Rural systems are typically smaller providers on the outskirts of these large metropolitan areas (e.g., the Heart of Iowa Transit Agency near Des Moines).

Any Urban system operating scheduled fixed route service was coded and modeled in full as described above. If a Rural system included fixed commuter service to the major urban area, the database and model included those routes in full detail.[102] Any other Rural fixed route service was not modeled—meaning it was not included as an option for getting from origin to destination in the regional transit network—but was factored into final job access measures (see the ***Metrics*** section for more details).

Time versus Money

These models allow a rider to take advantage of all transit options available within his/her metro area, treating different systems as part of one metropolitan network. The result is a measure of cost in terms of the time it takes to get from each origin to each destination within the region. However, this analysis does not measure the monetary cost of these trips (and potential multiple transfers across different systems) as part of the accessibility metrics.

Demographic and Employment Data

Once transit service was fully modeled in each of the 100 largest metropolitan areas, the results were combined with detailed demographic and employment data to assess the extent to which these transit systems align with where people work and live.

Characteristics of Origins

To better understand the characteristics of neighborhoods served by transit, this analysis uses estimates from the Nielsen Pop-Facts 2010 Database. Working-age population is measured by the number of residents aged 18 to 64. Neighborhood income has one of three values. Low-income block groups have median household incomes below 80 percent of the metropolitan area's median household income (AMI).[103] Middle-income block groups have median incomes between 80 and 120 percent of the AMI, and high-income block groups are those with median household incomes above 120 percent of AMI.[104]

Characteristics of Destinations

Data on employment come from Nielsen Business-Facts Database and represent census tract job counts as of the second quarter of 2010. In addition to measuring the total number of jobs located in tracts reached via transit, the analysis also considers the types of jobs found in each tract, based on 2-digit Standard Industry Classification (SIC) codes included in the Business-Facts Database.

Jobs were grouped into three categories based on industry: low-, middle-, and high-skill. The analysis proxies the skill level of each industry by the ratio of workers in that industry who had less than a bachelor's degree to those with a bachelor's degree or higher.[105] Because the 2-digit aggregation for Services in the SIC captures both high-skill and low-skill services, the same approach—ratio of less-than-bachelor's to bachelor's and above—was used to break out services by type before assigning a final designation.[106] High-skill industries are those with a ratio less than 1; middle-skill industries have

Table A. Industry Skill Designations by Ratio of Workers without a BA to Workers with a BA

Industry	2000 Ratio of Non-BA Holders to BA Holders	Skill Designation
Services: business, legal, social, health, education	0.5	High
Public Administration	0.5	High
Finance, insurance, and real estate	0.5	High
Transportation, communications, and utilities	1.1	Middle
Wholesale trade	1.2	Middle
Manufacturing	1.5	Middle
Retail trade	1.6	Low
Mining	1.7	Low
Services: personal, entertainment, lodging, repair	1.8	Low
Agriculture, forestry, and fisheries	2.3	Low
Construction	2.4	Low

Source: Brookings Institution analysis of 1990 and 2000 Public Use Microdata Samples data

a ratio between 1 and 1.5; and low-skill industries are those with a ratio greater than 1.5. The resulting skill designations are presented in Table A.

Metrics

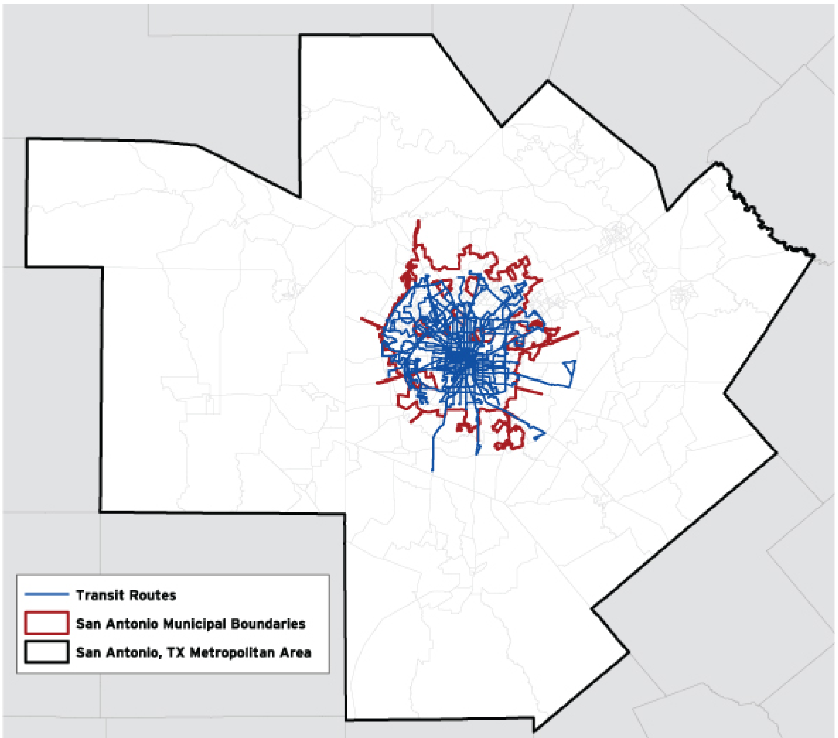
This analysis combines transit, demographic, and employment data to assess metropolitan performance in three key areas: transit coverage, service frequency, and share of metropolitan jobs accessible via transit.

Coverage—A metro area's transit coverage is measured by the share of working-age residents living in block groups that are considered "served" by transit (i.e. block groups with access to at least one transit stop within 3/4 mile of their population-weighted centroid).

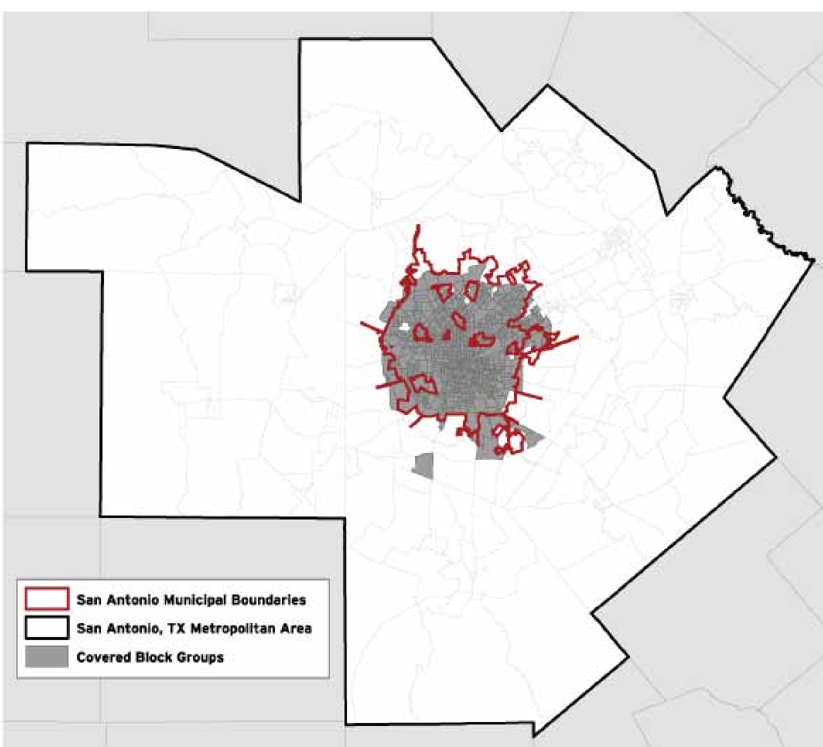
Service Frequency—For block groups considered "covered" by transit, this measure calculates the typical frequency at which that block group is serviced by any route during the morning rush hour, or how many minutes commuters must wait between buses or trains. The median frequency, or "headway," for each metropolitan area is then calculated based on the median headway in each covered block group, and weighted by block group working-age population. For example, many block groups in Midtown Manhattan see median headways of 3 minutes or less, while a neighborhood in Rockland County may see the typical wait time between buses extend more than two hours. When the headways in each covered block group in the region are taken into account, the median population-weighted frequency of service in New York's metropolitan area is 4.6 minutes.

Job Access—For block groups with transit service, this measure assesses the share of total metropolitan employment that can be reached via transit within 90 minutes.[107]

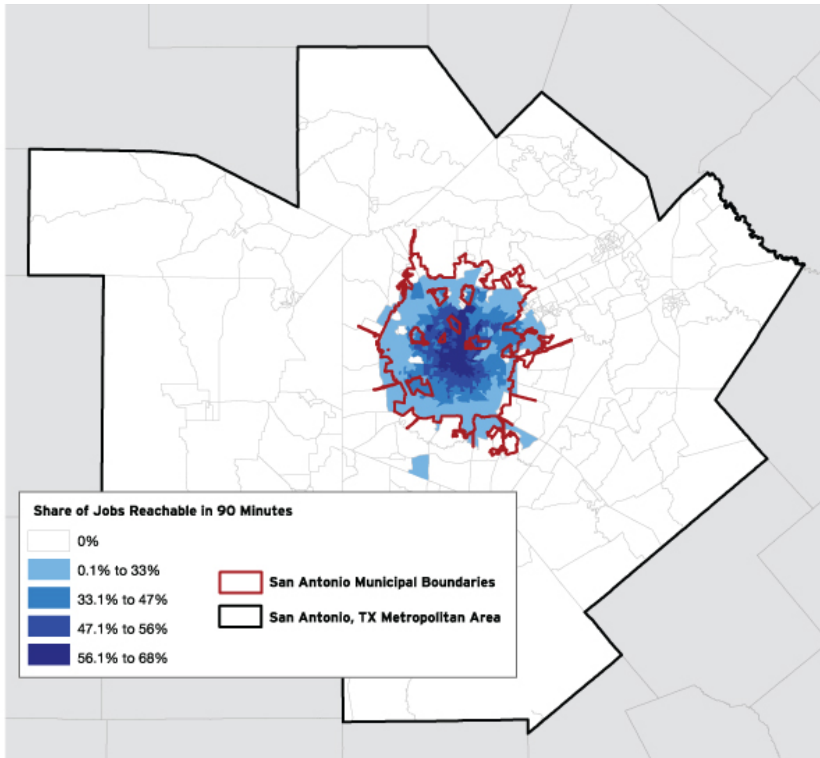
San Antonio Figure A.



San Antonio Figure B.



San Antonio Figure C.



For each block group, tracts reached within 90 minutes are flagged and the jobs in these census tracts are aggregated. That number is then divided by the total number of metropolitan jobs to determine the share of jobs that can be reached within 90 minutes from the relevant block group. (Jobs in a block group's home census tract are added to the reachable jobs total if not already assigned by the transit model, as are jobs in neighboring tracts considered reachable via the un-modeled rural transit services described above.)

Metro-wide job access is then calculated as the average share of jobs reachable within 90 minutes across all block groups, weighted by block group working-age population. (Note that covered block groups that cannot reach any destination within 90 minutes are removed from this measure.) In short, for neighborhoods that can reach at least one other destination within 90 minutes, this measure reflects the share of metropolitan jobs the typical working-age resident can reach via transit.

In addition to calculating the share of total jobs accessible via transit within 90 minutes, the analysis considers the share of low-, middle-, and high-skill jobs accessible via transit.

Each metric is also calculated by income group (i.e., low-, middle-, and high-income block groups), and separately for cities and suburbs. "Cities" are designated as the first city listed in the official metropolitan statistical area name, as well as any other city that appears in the name and has a population over 100,000. According to these criteria, there are 137 primary cities in the 100 largest metro areas. "Suburbs" make up the balance of these metropolitan areas outside of their primary cities.

Combined Access

Finally, to create one overall ranking of relative metropolitan performance, these metrics are incorporated into a combined score. Because frequency of service contributes to the number of destinations reachable in the job share measure, the final ranking relies on a metro area's coverage and job share measures. The share of working-age residents covered by transit is multiplied by the share of jobs reachable in 90 minutes, and the product is multiplied by 100. All metro areas are then ranked based on the outcome.