Connecting people and places: Exploring new measures of travel behavior

By Adie Tomer, Joseph Kane, and Jennifer S. Vey | September 2020

Methodology

Note: This methodological addendum will update as the Metropolitan Policy Program publishes additional research products within the project series.

The project combines geographically granular travel data and other local economic, social, and land use data to explore how neighborhoods interact within a given set of metropolitan areas. The project relies on a cutting-edge transportation dataset directly informed by anonymized GPS sensor and related data. This methodology describes the travel database in greater detail, the geographic extent of the data, how we combine travel data with other data sources, and the new set of performance measures derived from the complete database.

Trip Data

The explosion in digital transportation data—including individual traveler records from GPS devices, mobile phones, and other technologies—makes this analysis possible. While there are numerous private data providers emerging that track travel behavior, we rely on trip data available from Replica. Previously part of Sidewalk Labs—an Alphabet company focused on urban innovation—Replica is a "next-generation urban planning tool" that models trip movement based on de-identified mobile location data, population data, and other field data available from local government agencies. Replica estimates how individuals travel each day based on data from their smartphones, which is aggregated by a variety of different companies (e.g. Foursquare, Mapbox), anonymized to protect privacy, and corroborated with several other data sources to verify accuracy.

In particular, Replica provides a set of baseline trip data, including the total number of trips to/from a given neighborhood, when these trips take place (i.e. by time of day), and the purpose of these trips (i.e. commuting to work or shopping). It does so by collecting data on individuals' location history, using about 10 to 15 percent of a local population to create a travel model. This model estimates activities for a collection of "synthetic" people and households throughout the day, intended to approximate where, when, and why they are traveling. Since mobile location data are not available for most of a resident population—including many younger and older travelers—the model uses Census demographic data to create a representative sample of a region's travel patterns.

The model is further refined using other on-the-ground data, including traffic counts and transit records. Finally, additional computer simulations help generate a week of activities for this synthetic population, allowing for additional verification. Replica's estimates are continually evolving, with the aim to better capture a range of different trips, including those over shorter distances, and a range of different metropolitan areas nationally.

The most current data used for this analysis is for a typical Thursday in Fall 2018 (September through November). However, with timelier and geographically-granular data—updated every three months across a growing number of regions—Replica is

providing a new set of metrics that can inform new research efforts, planning decisions, and other policy interventions over time.

Key Terms

Trip: Movement of an individual from an origin to a destination. Based on one consecutive movement from start to finish, trips do not include specific route information; for instance, if an individual travels to the store and then travels home afterward, these are two distinct trips.

Distance: The physical extent of a given trip, measured in miles traveled.

Duration: The chronological extent of a given trip, measured in minutes traveled.

Neighborhood: Local geographic unit of analysis, equating perfectly with 2010 Census tracts.

Metropolitan area: Larger geography that captures many different neighborhoods. These include metropolitan statistical areas (MSAs), as defined by the Office of Management and Budget.

How does Replica take data privacy into consideration?

Replica sources anonymized third party data that has been further replaced with masked identifiers, so the data sets remain intact while reducing potential re-identification concerns, to create: a synthetic population that matches aggregate statistics of a region's real population; estimates of aggregate movement patterns across coarse-grained locations; and a travel model that represents who moves where, when, why, and how. These components yield a replica of likely transportation patterns across the region.

To protect privacy, Replica uses de-identified data and applies de-identification measures in its data platform. This minimizes the risk of re-identifying specific trips or people. Replica's "synthetic" data closely matches aggregated statistics, but is not intended to match any specific underlying person in the original data. The Brookings Institution's *Terms of Use* prohibit the authors from doing so as well. Independent models are built using various data sources to abstract-out identifying details of any given individual before combining with other data sources.

Geographic Coverage

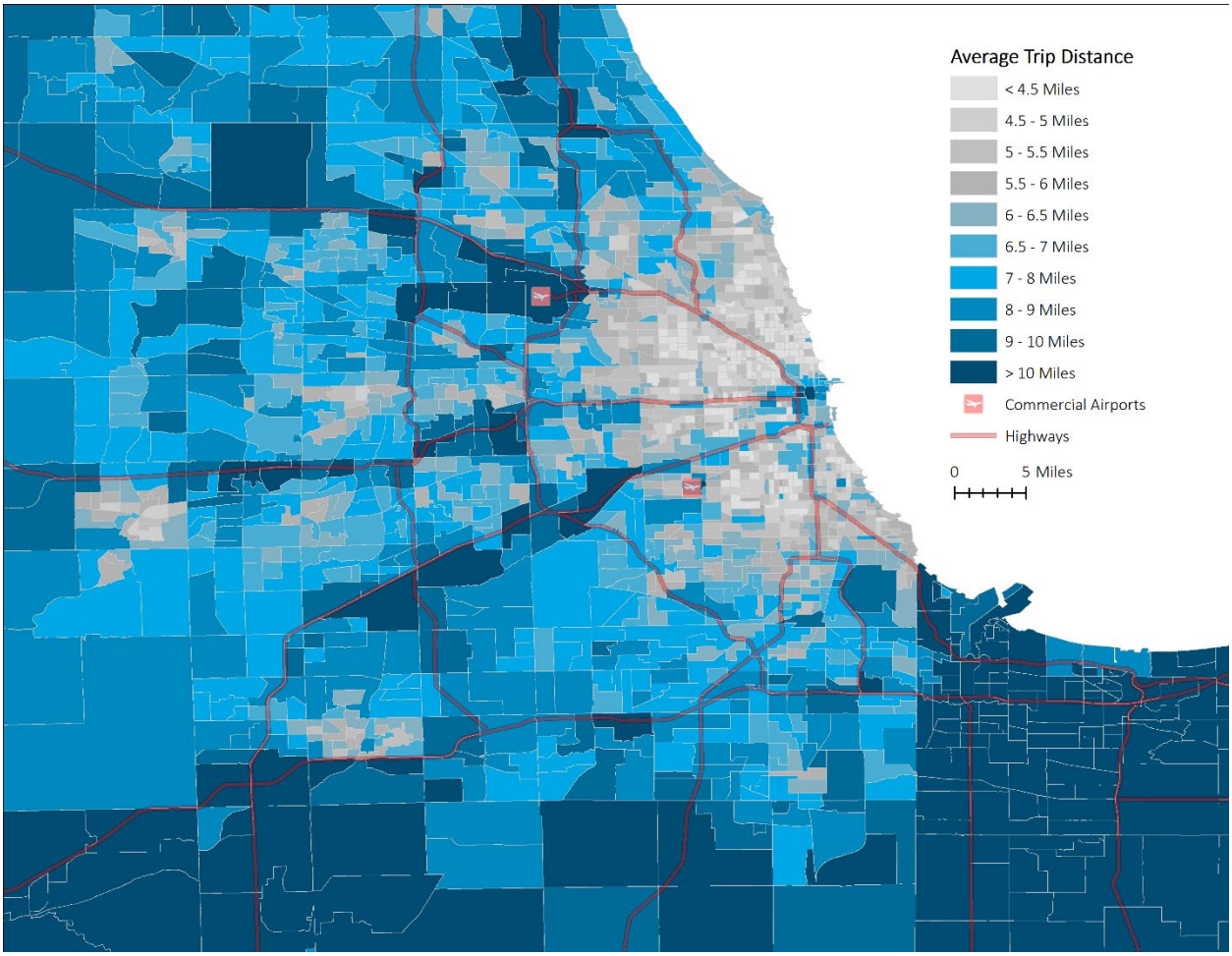
We analyze travel patterns at the neighborhood level, aggregating trends up to larger county, metropolitan, and other geographic designations.

"Neighborhoods" refer to the base level of local geography where individuals travel, defined at the Census tract level. Replica provides information for the total number of trips all individuals take to and from a given set of Census tracts, or what's known as origins and destinations (OD). For example, we analyze how many trips start in one tract and end in another throughout a typical weekday. Replica includes trips starting and ending in the same tract, or intra-neighborhood travel. Replica's internal database includes trips at even greater geographic granularity, down to specific routes. While Brookings did not have access to this level of detail, it allows our dataset to include OD data with highly specific travel statistics.

We combine neighborhood travel across all tracts in several different metropolitan statistical areas (MSAs), as defined by the Office of Management and Budget. The MSAs in this project contain anywhere from hundreds to thousands of different Census tracts, allowing for a detailed view of how whole regions and individual neighborhoods connect. For example, the Kansas City MSA contains 528 different tracts; we look at how all these tracts connect with another in terms of the total number and type of trips throughout a typical weekday. We also create summary statistics for the entire MSA based on these tract-to-tract connections. We exclude any trips starting or ending outside the OMB-designated metropolitan boundaries.

Initially, we analyzed travel patterns across five different MSAs: Kansas City, Chicago, Sacramento, Portland, and Dallas. In total, these MSAs include: a combined 5,011 tracts, a combined population of 23.3 million, and a combined 68.1 million trips. Over time, we will analyze data for several additional MSAs nationally as new Replica data become available. Figure A1 below illustrates the wide range of tracts and distances that can appear in a given MSA.

Figure A1. Example of tract-level detail available in the Chicago MSA



Source: Brookings analysis of Replica data.

Calculating Distance and Duration

Throughout the project series, we examine several different metrics to describe the nature and extent of trips. Two metrics—distance and duration—serve as the primary ways to gauge the physical and chronological extent of any given trip.

Distance is measured in the number of miles traveled from an origin to a destination. When estimating the daily travel habits of a synthetic population, Replica examines the location of activities (i.e. home, work, etc.) and the transportation network segments that individuals use to reach their destinations. Specific network segments are based on the particular infrastructure and mode used, including driving, walking, and riding transit. These network segments are flattened and organized into consistent travel records for a group of individuals at the tract level, which form the basis of the distance traveled. Trips at the tract level are assigned an average distance, median distance, minimum distance, maximum distance, and standard deviation, which allow for a detailed range of data points across different origins and destinations.

Duration is measured in the number of minutes traveled from an origin to a destination. Similar to distance, Replica estimates how long trips take based on the daily travel habits of a synthetic population. Transportation network segments are flattened and organized to create consistent travel records for a group of individuals at the tract level. Trips at the tract level are assigned an average duration, median duration, minimum duration, maximum duration, and standard deviation, which allow for a detailed range of data points across different origins and destinations.

Trip Details

In addition to describing trips by distance and duration, we will also investigate how trips vary by the time of day when they occur and by their purpose. Again, Replica estimates these trip details based on location data for individuals, and this information is expanded to cover a broader synthetic population across a whole region.

In total, we will look at trips that occur across five different time periods on a typical weekday: 12 a.m. to 6 a.m.; 6 a.m. to 10 a.m.; 10 a.m. to 4 p.m.; 4 p.m. to 8 p.m.; and 8 p.m. to 12 a.m. These times are based on when a trip starts, available by the trip origin and trip destination.

By purpose, we look at trips across three different categories: home, work, and other. Trips to/from home are related to residential locations; trips to/from work are related to employment locations; and trips for other purposes are related to locations for shopping, recreation, and miscellaneous needs. Replica also estimates the number of trips to/from school, but these data have not been separately analyzed.

At this time, we do not analyze travel by transportation mode.

External Data and Modeling

The Replica data only reveal so much about metropolitan or neighborhood travel patterns. Viewed in isolation, the Replica data simply tells the user how many people traveled between certain places and the general purpose of these trips. To complement the Replica data and more clearly investigate these associated factors, we have constructed an extensive database containing tract-level economic and built environment data. These external data come from a range of sources and cover a range of variables, as shown in Table A1 below. Some of the primary variables include:

- **demographic indicators** (race, income, poverty, and educational attainment) from the Census American Community Survey (ACS) program;
- **economic indicators** (jobs, job density, and job types), from other Brookings analyses and the Census Longitudinal Employer-Household Dynamics (LEHD) program;
- **housing indicators** (number of units, housing density, and type of housing), also from the ACS program; and
- **place-based indicators** (intersection density, land use, and health status, among others), from the Environmental Protection Agency, Center for Neighborhood Technology, Centers for Disease Control and Prevention, and multiple other data providers.

Table A1. Selected variables analyzed and data sources used

Variable	Source	Vintage
Race	ACS	2017
Income	ACS	2017
Poverty	ACS	2017
Number of Housing Units	ACS	2017
Housing Density	ACS	2017
Type of Housing	ACS	2017
Distance to CBD	ACS	2017
Population	ACS	2017
Population Density	ACS	2017
Housing/Transportation Costs	CNT	2017
Intersection Density	CNT	2017
Average Block Size	CNT	2017
Number of Jobs	LEHD/LODES	2015
Job Density	LEHD/LODES	2015
Jobs By Industry	LEHD/LODES	2015

We have also developed several different regression models to assess these variables of interest while controlling for other factors. These models aim to clarify the extent to which certain neighborhood conditions can affect travel patterns, both within and across different regions. We use Ordinary Least Squares (OLS) multiple regressions to investigate the association between demographic, economic, housing, and place-based variables and tract-level travel patterns across different MSAs.