

Portland State University

PDXScholar

TREC Datasets and Databases

Transportation Research and Education Center
(TREC)

6-2018

Nationwide Worker Database to Analyze Transit Outcomes Using an Economic and Demographic Lens

Robert Hibberd
University of Arizona

Arthur C. Nelson
University of Arizona

Follow this and additional works at: https://pdxscholar.library.pdx.edu/trec_data



Part of the [Urban Studies Commons](#), and the [Urban Studies and Planning Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Hibberd, Robert & Arthur C. Nelson. Data From: Nationwide Worker Database to Analyze Transit Outcomes Using an Economic and Demographic Lens. NITC-DP-1103. Portland, OR: Transportation Research and Education Center (TREC), 2018. https://doi.org/10.15760/TREC_datasets.01

This Dataset is brought to you for free and open access. It has been accepted for inclusion in TREC Datasets and Databases by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Nationwide Worker Database to Analyze Transit Outcomes Using an Economic and Demographic Lens



Robert Hibberd & Arthur C. Nelson

College of Architecture, Planning and Landscape Architecture, University of Arizona, Tucson, AZ

Abstract

This extensive database project provides demographic data for workers, both in terms of where they live and where they work, for 39 U.S. Metropolitan Statistical Areas (MSAs). Planners and researchers can use this database to assess the extent to which transit stations and station areas are associated with economic and demographic change, and to forecast similar magnitudes of change through proposed new or expanded transit systems. Specific data elements include job numbers based on job sector, earnings, race, education, sex, and distance of census blocks to transit separated by mode type. Data were extracted as part of a larger project that examines the impact of transit on a large number of societal factors, including economic development and demographic dynamics

(<https://nitc.trec.pdx.edu/research/project/1103/>).

Description of Database Content & Methodology

The data project contains a collection of tables, which provides U.S. census block data of the United States Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) data from 2002 to 2015. The LEHD LODES files include data for Residential Area Characteristics (RAC) and Work Area Characteristics (WAC). Provided data also include distances from each census block centroid or edge to the nearest transit station, separated by mode type. These have been exported from GIS geodatabases created from a combination of census blocks and the LODES tables. Data from each city were merged into one file per city that contains all years for both the RAC and the WAC data, which are combined into 35 separate data sets and are made available as csv and txt format.

The data for the project were downloaded from multiple sites. The GIS census block shapefiles (a GIS file format) are 2010 census blocks and come from the Integrated Public Use Microdata Series (IPUMS) National Historical GIS (IPUMS NHGIS) website.¹ The NHGIS census block data sets are useful due to the lack of availability of these data in a ready-made GIS format from the U.S. Census Bureau. The NHGIS has requested that all research utilizing their data sets include the citation they provide. Please refer to the required citation provided by the NHGIS at their website. The LEHD LODES data were downloaded as csv tables from the U.S. Census Bureau's download site² and joined to the census block data, which were converted from a shapefile to a geodatabase format for a more robust product.³ The Origin-Destination (OD) files are not included, because they are outside of the scope of this project.

Each geodatabase includes both the Residential Area Characteristics (RAC) and Work Area Characteristics (WAC) files for each of the years 2002-2015, where available. Each file has a concatenated name, combining the census block name and the LODES table name. For example, for Austin, TX, the 2002 RAC file is named "TX_block_2010_tx_rac_S000_JT00_2002_Austin." The first section, from the 2010 census block GIS data, is "TX_2010_block." The second section, from the LODES RAC table, is the remainder of the name, except for the name of the MSA at the end of the file name. The other portions of the LODES table name are described in the technical documentation.⁴ The concatenated name indicates that the LEHD LODES tables are joined to the census block data, which was accomplished through a table join in ArcGIS.

The distance from the census block edge to the nearest transit station point is provided in each file, as well, by transit mode (e.g., CRT or SCT). Distances were measured as euclidean distances⁵ using a Near analysis in ArcGIS. In order to calculate these distance measurements, files are projected to the North America Albers Equal Area Conic projection, a secant projection that minimizes the Scale Factor across the projection, ensuring that distortion from

the projection process is negligible overall.⁶ A number of the files are projected instead to the State Plane Coordinate System (SPCS) for the local area. These projections are frequently used as a standard by government and industry analysts.⁷ The GIS census blocks included in the data set all fall within the Metropolitan Statistical Area boundary for each city included in the data set. Care was taken to include all census blocks falling within the counties that fall within each of the MSAs, which are enumerated below (Table 1). Some files include data for combined MSAs, such as Seattle-Tacoma and San Diego-Oceanside-Escondido, because of the proximity of some MSAs to each other.

Content of Files

The data for each MSA are provided as separate feature classes (a logical file structure for GIS similar to shapefiles) for each MSA, and are housed inside a geodatabase. The feature classes are separated between three geodatabases in order to reduce the file size of each database for ease of use across the internet (Table 1). In addition to providing the geodatabases for use in GIS software, we have provided their contents as exported tables in csv and text format for ready use in statistics software.

Three fields, found in the geodatabase and the exported csv and text tables, organize the data in accordance with their original file structure. The extra field names are as follows: LEHD_Year, WAC_RAC, and City. These fields were necessary to keep data organized by the year, whether the data are from the WAC or RAC files, and to what city they belong.

While this process does result in a sizeable number of null values in the provided tables, it nevertheless made data much more manageable in size, by allowing fewer files. This format also greatly increases processing efficiency.

Metropolitan Areas included in the Database Project

Table 1: Metropolitan Areas (MSAs) grouped by geographic region and matched to the geodatabase folder they are stored in.

Region	Metropolitan Area(s)	Geodatabase folder Containing Data Files
U.S. West Coast	Seattle-Tacoma, WA	LODES_CompiledC.gdb
	Portland, OR	LODES_CompiledB.gdb
	Eugene-Springfield, OR	LODES_CompiledA.gdb
	Sacramento, CA	LODES_CompiledC.gdb
	San Jose-San Francisco-Stockton, CA	LODES_CompiledC.gdb
	San Diego-Oceanside-Escondido, CA	LODES_CompiledC.gdb
U.S. West	Salt Lake City, UT	LODES_CompiledC.gdb
	Las Vegas, NV	LODES_CompiledB.gdb
	Reno, NV	LODES_CompiledC.gdb
	Phoenix, AZ	LODES_CompiledB.gdb
	Tucson, AZ	LODES_CompiledC.gdb
	Albuquerque-Santa Fe, NM	LODES_CompiledA.gdb
	Denver, CO	LODES_CompiledA.gdb
U.S. South	Austin, TX	LODES_CompiledA.gdb

	Dallas-Fort Worth, TX	LODES_CompiledA.gdb
	San Antonio, TX	LODES_CompiledC.gdb
	Houston, TX	LODES_CompiledA.gdb
	Little Rock, AR	LODES_CompiledB.gdb
	Charlotte, NC	LODES_CompiledA.gdb
	Atlanta, GA	LODES_CompiledA.gdb
	Nashville, TN	LODES_CompiledB.gdb
	New Orleans, LA	LODES_CompiledB.gdb
	Norfolk, VA	LODES_CompiledB.gdb
	Miami, FL	LODES_CompiledB.gdb
	Orlando, FL	LODES_CompiledB.gdb
	Tampa, FL	LODES_CompiledC.gdb
U.S. Midwest	Minneapolis-St. Paul, MN	LODES_CompiledB.gdb
	Kansas City, MO	LODES_CompiledA.gdb
	St. Louis, MO	LODES_CompiledC.gdb
	Cleveland, OH	LODES_CompiledA.gdb
	Cincinnati, OH	LODES_CompiledA.gdb
	Buffalo, NY	LODES_CompiledA.gdb
U.S. East Coast	Philadelphia, PA	LODES_CompiledB.gdb
	Pittsburgh, PA	LODES_CompiledB.gdb
	Washington, DC	LODES_CompiledC.gdb

Using the Data

There are four zip files provided. Listed below are their contents and the required software for their use:

- One zip file that contains a folder of csv and text files, with the extensions “.csv” or “.txt.” This folder contains csv and text tables for all 35 separate data sets, which provide the U.S. Census Bureau’s longitudinal employment data (LODES) at the census block level for the years 2002 to 2015. These files can be used in such software as Excel or standard statistical packages such as SPSS or STATA. This folder is called:
 - **LODES_Compiled_Tables.zip**
- Three zip files that each contain a geodatabase with the data file extension “.gdb.” Together, these three geodatabases contain the U.S. Census Bureau’s longitudinal employment data (the LODES project) at the census block level for the years 2002 to 2015 for all 35 separate data set listed above (Table 1). ArcGIS, GDAL and QGIS software packages can read and use these files. These are called:
 - **LODES_CompiledA.gdb.zip***
 - **LODES_CompiledB.gdb.zip***
 - **LODES_CompiledC.gdb.zip***

*Each data zip file is large and can take up to 4 hours to download.

Recommended Citation for Database Project

Hibberd, Robert & Arthur C. Nelson. Nationwide Worker Database to Analyze Transit Outcomes Using an Economic and Demographic Lens. NITC-DP-1103. Portland, OR: Transportation Research and Education Center (TREC), 2018. https://doi.org/10.15760/TREC_datasets.01

Endnotes

¹ see <https://www.nhgis.org/>.

² See <https://lehd.ces.census.gov/data/lodes/LODES7/>.

³ More information on this data can be found at <https://lehd.ces.census.gov/>.

⁴ See <https://lehd.ces.census.gov/data/lodes/LODES7/LODESTechDoc7.2.pdf>.

⁴ Zhao, F., Chow, L. F., Li, M. T., Ubaka, I., & Gan, A. (2003). Forecasting transit walk 406 accessibility: Regression model alternative to buffer method. Transportation Research Record: 407 Journal of the Transportation Research Board, (1835), 34-41.

⁵ See the U.S. Environmental Protection Agency’s Smart Location Database technical documentation:

https://www.epa.gov/sites/production/files/2014-03/documents/sld_userguide.pdf.

⁶ Details of use for the SPCS can be found at: <http://desktop.arcgis.com/en/arcmap/10.3/guide-books/map-projections/state-plane-coordinate-system.htm>.

About the National Institute for Transportation and Communities

This project was funded by the National Institute for Transportation and Communities (NITC), one of five U.S. Department of Transportation national university transportation centers in the country. Hosted at the Transportation Research and Education Center (TREC) at Portland State University (PSU), the consortium is a PSU-led partnership with the University of Oregon, Oregon Institute of Technology, University of Utah and new partners University of Arizona and University of Texas at Arlington. We pursue our theme — improving mobility of people and goods to build strong communities — through research, education and technology transfer. Learn more at: <https://nitc.trec.pdx.edu/>