

NCHRP 08-36, Task 127

Employment Data for Planning

A Resource Guide

Requested by:

American Association of State Highway and
Transportation Officials (AASHTO)
Standing Committee on Planning

Prepared by:

Cambridge Systematics, Inc.
101 Station Landing, Suite 419
Medford, MA 02155

December 2017

The information contained in this report was prepared as part of NCHRP Project 08-36, Task 127, National Cooperative Highway Research Program (NCHRP).

Special Note: This report **IS NOT** an official publication of the NCHRP, the Transportation Research Board or the National Academies.

Acknowledgements

This study was conducted for the AASHTO Standing Committee on Planning, with funding provided through the National Cooperative Highway Research Program (NCHRP) Project 08-36, Research for the AASHTO Standing Committee on Planning. The NCHRP is supported by annual voluntary contributions from the State Departments of Transportation. Project 08-36 is intended to fund quick response studies on behalf of the Standing Committee on Planning. The report was prepared by Evan Enarson of Cambridge Systematics, Inc. The work was guided by a technical working group that included:

Ju-Yin Chen, Virginia Department of Transportation
William A. Christian, Fargo-Moorhead Metropolitan Council of Governments
Ed Christopher, MKC Associates
Ross M. Friedman, East-West Gateway Council of Governments
Charles Gorugantula, WSP USA
Samuel C. Granato, Ohio Department of Transportation
Shana Retherford Johnson, Foursquare Integrated Transportation Planning, Inc.
Phillip J. Mescher, Iowa Department of Transportation
Bradley M. Sharlow, Michigan Department of Transportation
Penelope Weinberger, American Association of State Highway and Transportation Officials

The project was managed by Lawrence D. Goldstein, NCHRP Senior Program Officer.

Disclaimer

The opinions and conclusions expressed or implied are those of the research agency that performed the research and are not necessarily those of the Transportation Research Board or its sponsoring agencies. This report has not been reviewed or accepted by the Transportation Research Board Executive Committee or the Governing Board of the National Research Council.

NCHRP 08-36, Task 127 Employment Data for Planning A Resource Guide

Requested by:
American Association of State Highway and
Transportation Officials (AASHTO)
Standing Committee on Planning

The information contained in this report was prepared as part of NCHRP Project 08-36, Task 127, National Cooperative Highway Research Program (NCHRP).
Special Note: This report **IS NOT** an official publication of the NCHRP, the Transportation Research Board or the National Academies.

Prepared by:
Cambridge Systematics, Inc.
101 Station Landing, Suite 410
Medford, MA 02155
December 2017

CONTENTS

SUMMARY	1
Objectives	2
CHAPTER 1 Overview of Employment Data Sources	4
Applying Employment Data in Transportation Planning	4
Summary of Major Employment-Related Data Sources	6
Understanding Differences and Distinctions in Employment Data	9
CHAPTER 2 Frequently Asked Questions.....	13
What Are the Differences Between Jobs, Workers, and Employment?	13
How Does a Company Differ from an Establishment?	13
Who Is Counted as a Worker at an Employment Site?	14
What Does Labor Force Mean?.....	15
What Are Some Common Issues that Affect the Accuracy of Employment Data?	15
How Can I Validate Data and Check for Data Weaknesses or Issues?	16
How Can I Combine or Reconcile Disparate Employment Estimates?.....	17
What Are the Advantages of Commercial Vendors Compared to Free Public Sources?.....	18
How Do Different Data Sources Address Persons Who Are Self-Employed or Who Work from Home?	18
How Are Hard-to-Reach or Transitional Employment such as Migrant Farmworkers or Military Service Members Accounted for?	19
CHAPTER 3 Quick Reference Guide	20
CHAPTER 4 Data Sources and Applications.....	25
4.1 Current Population Survey	25
4.2 Current Employment Statistics	27
4.3 Nonemployer Statistics	29
4.4 Quarterly Census of Employment and Wages	30
4.5 County Business Patterns.....	31
4.6 American Community Survey.....	33
4.7 Census Transportation Planning Products	34
4.8 Longitudinal Employment Household Dynamics	35
4.9 Quarterly Workforce Indicators	38
4.10 InfoUSA	40
4.11 Dun & Bradstreet	41

4.12	YourEconomy Time Series	43
4.13	Woods & Poole	44
4.14	Moody's Analytics	45
4.15	IHS Global Insight	46
CHAPTER 5 Resources and Links.....		48
	Transportation Research Board Resources	48
	Current Population Survey	48
	Current Employment Statistics	48
	Nonemployer Statistics	49
	Quarterly Census of Employment and Wages	49
	County Business Patterns	49
	American Community Survey	49
	Census Transportation Planning Products	49
	Public Use Microdata Sample	50
	Longitudinal Employer Household Dynamics Program.....	50
CHAPTER 6 Glossary		51
	List of Acronyms	54
APPENDIX A Practitioner Survey Results		56

LIST OF TABLES

Table 1-1. Comparison of total employment counts of major public data sources.....	10
Table 1-2. Example employment and establishment counts, public and private data sources.....	11
Table 3-1. What geographies are available?	21
Table 3-2. How frequently is data updated?	22
Table 3-3. What data coverage is available?.....	22
Table 3-4. What data characteristics are available?	23
Table 3-5. What questions are being asked?	24

LIST OF FIGURES

Figure 1-1. Example spatial distribution and employment patterns, public and private data sources.	12
Figure 4-1. LEHD processing steps.	37
Figure A-1. Survey responses received by state and regional agency.	57
Figure A-2. Question 1A: What types of projects do you use employment data for in long-range planning?.....	58
Figure A-3. Question 1B: What types of projects do you use employment data for in travel demand forecasting?	59
Figure A-4. Question 1C: What types of projects do you use employment data for in local project planning?.....	59
Figure A-5. Question 1D: What types of projects do you use employment data for in modal planning? ..	59
Figure A-6. Question 7: What data sources do you use most frequently?	66

Summary

High-quality employment data are critical to transportation planning, modelling, and policy analyses. These data convey important information on workplace locations, industry types, and total number of employees; worker demographic characteristics; and geographic distribution of home-to-work trips. Without these data, we would not be able to answer questions, such as: Where do workers live in a region? How do workers travel? How many trips are generated in an area? How accessible are employment centers? Or, how well are investments supporting jobs, freight industries, and economic competitiveness?

To answer these questions, employment data are utilized by transportation planning practitioners in a wide variety of applications. Data may be used to plan locations of future corridors and transit-oriented development centers, to update travel demand models, or to track the impacts of policies and investments. Increasingly, employment data are being used as key indicators in performance management approaches and to estimate demand for multimodal services or bike-sharing programs. These real-world applications carry significant implications as data may directly influence investment decisions, right-of-way acquisition, future needs estimates, and information provided to decisionmakers and the public.

The application and interpretation of employment data varies widely across transportation agencies and planning purposes. Federal and State agencies and private vendors provide a variety of sources of employment data—each with different measurement or estimation procedures, sample universes, time periods, geographies, and other information linked to employment. These data sources can produce significantly different estimates of total employment or distribution patterns of workers and employment locations within an area. These complexities present a challenge for many agencies, and the differences between data sources and employment concepts may not be well understood by end users and planning practitioners.

This research and technical guide provides a single point of reference for information on employment data. This information presented here is structured to provide summary-level, practical information to help planning professionals understand the range of available data sources, how data may be applied, and what data source may be best suited for transportation planning applications. This resource guide is not intended to duplicate or replace technical information on specific datasets that are well documented and available from Federal agencies, such as the Bureau of Labor Statistics and U.S. Census Bureau.

This guide provides a starting point for understanding the universe of available data sources and presents common applications of those data. A practical self-assessment guide enables practitioners to skip directly to topics of interest and quickly identify appropriate data sources, given an agency's geography, capacity, resources, and the intended application of data. Commonly used technical terms and data concepts are defined and frequently asked questions are answered in this guide. Detailed information on major public data sources and available information on private data sources are provided to summarize data basics, such as sample size and universe, geographic resolution, reporting and update frequency, and limitations and data cautions.

The resource guide will assist transportation planning practitioners to better understand differences, limitations, solutions, and technical concepts among available public and private employment data sources.

Objectives

High-quality employment data drawn from broad sectors of the economy are necessary for transportation planning, whether used for travel demand modeling or policy analysis. Availability of these data varies widely, however, as do the methods that the transportation community uses to combine, adjust, and manipulate employment data drawn from multiple sources. Transportation practitioners also vary widely in their understanding of the sources and attributes of employment data. The transportation industry does not have a recognized standard for distinguishing, or even understanding the differences among employment, jobs, labor force, and workers. Additionally, data are being manipulated to be used for purposes for which they were not gathered. In the current climate of confusion, clarity and information are more pressing than ever. As agencies move towards greater accountability in transportation planning, they need to have confidence in the information they produce, and they need to know their data sources are reliable. In other words, the transportation community needs to examine potential data sources in an appropriate context for their planned use.

This guide was developed with the objective of providing a resource for selecting and using employment data. Intended for transportation practitioners, the guide explains the various types of employment data that exist and that are commonly used for transportation planning purposes. This research further develops both the U.S. Census Bureau's own research on the interactions between the worker estimates of American Community Survey and Current Population Survey and the information available in the National Cooperative Highway Research Program (NCHRP) 08-36, Task 98: Improving Employment Data for Transportation Planning report that examines publicly available sources.

Methodology

The development of this guidebook was directed by the NCHRP 08-36, Task 127 Project Panel. An online survey of transportation planning practitioners provided information on real-world applications, challenges, and solutions and the frequently asked questions presented in this guide. Interviews with a select number of large and small, urban and rural transportation planning agencies provided additional detail on data uses, issues, and common challenges. Review of existing academic literature, Web sites, and technical documentation from Federal agencies, and independent review and evaluation of documentation and data provided by commercial vendors helped inform the data source summaries presented here.

Overview of this Resource Guide

This guide is intended to serve as a practical and informative resource for transportation planners that provides a single point of reference for information on employment data. The guide includes:

- An overview of the primary and most commonly used public and private employment data sources and their potential applications in transportation planning (Chapter 1).
- Frequently asked questions and recommendations for transportation agencies when using employment data and critical considerations and caveats for each primary employment data source (Chapter 2).
- A self-assessment to enable practitioners to skip directly to topics of interest and quickly identify appropriate data sources, given an agency's geography, capacity, resources, and the intended application of data (Chapter 3).
- More detailed information on the primary sources of employment data (Chapter 4), including:
 - Covered occupation and industry categories.
 - Sample size and framework.
 - Geographic resolution.

- Update frequency.
 - Cost of acquisition or required processing resources.
 - Limitations and data cautions.
- Links to online resources where users can obtain additional information, view best practice examples, or download data (Chapter 5).
- List of key concepts and glossary of key terms (Chapter 6).
- Findings from a survey of practitioners on the use of employment data in planning, policy, and modelling applications (Appendix A).

CHAPTER 1

Overview of Employment Data Sources

A wide range of data is available from public agencies and commercial vendors that enable various counts, summaries, analyses, and mapping of employment data. For transportation planning purposes, employment data sources provide a full set of counts, indicators, estimation measures, and analytical tools to identify needs and issues. These data can inform policies, strategies, and decisions within statewide or regional long-range plans down to corridor or local area plans. Employment data are critical for developing and applying travel demand and activity-based models. Data also can be applied to inform policy choices, program evaluations, and economic analyses.

Applying Employment Data in Transportation Planning

An overview of common applications and analyses that can be informed using employment data are provided below, along with some of the data sources most commonly used.

- Transportation and Transit Demand Analysis**—Employment data can provide proxy measures of overall demand for new or expanded transportation services without significant modeling capabilities. Data sources that allow for relatively granular geographic analysis can be used to identify potential customer markets by examining total employment and employment by industry. Data sources that provide current commuter characteristics and origin-destination statistics can provide more in-depth information on potential demand and current mode shares. Public data sources, such as the *Quarterly Census of Employment and Wages* and the *American Community Survey*, provide detailed employment information. American Community Survey data can be drawn down to census tract or block group to enable rough estimates of commuter demand. The *Longitudinal Employer-Household Dynamics (LEHD)* OnTheMap web-based analysis tool can be used to draw custom geographies and estimate commuter sheds. For example, estimates of potential customers for a bike sharing program could be developed by pulling counts of workers who both live and work in the same geographic boundaries. Using the same tool, transit demand between a given employment center and given housing centers could provide rough gauges of bus rapid transit demand potential.
- Trends and Forecasts**—Using employment data to show how the industry composition of an area has changed over time, or might change in the future, can help transportation planners develop plans and identify needs. Long-term trend employment data for States and regions is readily accessible from data sources, such as the *Current Employment Statistics*, while more detailed geographic data from the *Quarterly Census of Employment and Wages* can provide snapshots of the past decade. The new *YourEconomy Time Series* data source provides a new and effective source for generating employment and industry trends at the State, county, and regional levels. Forecasts of employment data and future industry shifts are generally only available through private providers, such as *IHS Global Insight*, *Woods & Poole*, or *Moody's Analytics*.
- Economic Analysis**—Employment data sources provide additional measures beyond worker and job counts. Measures of wages, earnings, hours worked, minutes of commute time, personal income, gross product, and industry revenues can be used in economic analyses for transportation planning purposes.

Benefit-cost and economic impact analyses draw on these data, and more general analyses of the cost of travel and commute times may be estimated. *American Community Survey*, *Census Transportation Planning Products*, and *Quarterly Census of Employment and Wages* data source provide additional information and measures beyond workers, jobs, and employment site counts.

- **Transportation Performance and Policy**—Employment data that provides detailed characteristics of workers and commuters can be used to identify and assess potential transportation policy initiatives or strategies. *American Community Survey* data can be used to demonstrate the impacts of transportation, housing, and job availability on specific segments of the working population, including by occupation, gender, and race or ethnicity. These data can help assess the impact of policies that might change commute patterns or programs that are geared toward certain segments of the working population. The same data also can be used to develop proxy performance measures, or to develop tracking and progress measures for strategies. Indicators such as mode share, commute time, and aggregate commute minutes provide reasonable tracking measures for a wide variety of geographies.
- **Commute Flows and Worker Origin-Destination Patterns**—Detailed information on commute patterns is available from data sources, such as the *American Community Survey* and *LEHD Origin-Destination Employment Statistics*. These data can be segmented, visualized, and analyzed in a variety of ways, from simple directional commuter flows in and out of an area to examining the travel distances of workers by earnings or by industry. LEHD data can be used to identify broad connected economic regions, or to drill down into worker flows within specific study areas. American Community Survey data can show transit utilization or bike and pedestrian job accessibility at census block group level or at the regional or State level. These data are not often fully utilized in transportation planning, and can provide contextual information when travel demand models or independent household travel survey data are not available.
- **Freight and Industry Economic Patterns**—Specific commodity flow data from public and private providers, such as Surface Transportation Board Waybill Samples or IHS Transearch data, are useful in determining what goods are moving and how. Employment data sources also can be used in freight and multimodal planning to supplement commodity-specific data and estimate economic patterns based on industry employment. Employment in freight-reliant industries, such as natural resources, wholesale and retail trade, construction, and manufacturing, can provide indicators of the general level of goods movement needs. Mapping the location of freight-reliant industry clusters also can be used to highlight transport needs within a specific area. Using long-term employment data to show shifts in industrial activity and the industry composition of a study area can be used to show how the transportation needs of businesses and workers are changing. Private data sources, such as *InfoUSA* or *Dun & Bradstreet*, are typically necessary to gather location-specific information on business and industry activity. Public data sources, such as the *Quarterly Census of Employment and Wages* or *Current Employment Statistics*, can be used to provide detailed employment estimates for various industries and to show trends over time.
- **Modelling and Estimation**—Employment data, typically from commercial vendors or ES-202-based data from State agencies, are used as a key input in travel demand models to create socioeconomic data and assign employment to geographic zones. Point-specific work location or establishment location information is critical for assignment of employment within transportation analysis zones. Typically, employment also is categorized into industry groupings for broad economic sectors, such as retail; service; and basic (mining, utilities, construction, manufacturing, wholesale trade, transportation, etc.). *Info USA*, *Dun & Bradstreet*, and *Woods & Poole* commercial datasets are often used, among others, as a source of employment data with geographic locations. When available through a State agency, ES-202 data, which is the source of the *Quarterly Census of Employment and Wages* data, also may be used within models. For model calibration and validation, some agencies have experimented with

using *LEHD Origin-Destination Employment Statistics* as a source of workplace and home-to-work flow data. However, this practice is not as common as relying on household and employer travel surveys.

- **Communications and Education**—Data can tell powerful stories, and help communicate transportation needs to public and elected audiences. Employment data can be used to illustrate average commute times, means of transportation to work, and commuter flows in and out of activity centers. For long-range plans, corridor studies, and even local area plans, this information can be effectively visualized and displayed to highlight potential issues and needs. Data sources that provide commute information and characteristics include *Census Transportation Planning Products*, *American Community Survey*, and *LEHD Origin-Destination Employment Statistics*.

Summary of Major Employment-Related Data Sources

The following section provides a brief summary of the major public and private employment data sources. Background and methodology information, geographic availability, key data characteristics, and supplement information are summarized. Potential interest to transportation planning practitioners and potential applications are addressed. More information on each of these data sources can be found in Chapter 5.

- **Current Population Survey (CPS)** is a joint effort by the U.S. Census Bureau (Census) and the U.S. Bureau of Labor Statistics (BLS), and serves as the primary source of labor force statistics in the United States. The CPS is a household-based demographic and economic survey with one of the most complete response rates of national surveys. The CPS provides employment characteristics and information that can help planners and decision makers to understand the composition of the labor force at the national and State level, as well as for 12 of the largest metropolitan statistical areas. Employment status, employment characteristics, wages and earnings, occupation and industry, and additional special supplemental data are available. Micro use and historical data are available through third-party resources, including the National Bureau of Economic Research. CPS is most useful for high-level economic and employment data and does not provide granular geographic information, workplace location, or journey-to-work information.
- **Current Employment Statistics (CES)** program of the BLS is an establishment-based payroll survey, which is conducted monthly. The CES provides estimates of nonfarm employment, hours, and earnings at the national, State, and 450 metropolitan areas. Hours and earnings are published only for private-sector jobs. The CES provides detailed historical series information from 1990 on, and less detailed information from 1939 on. CES is most interesting to planners to illustrate long-term changes in employment, demand (hours) and earnings. CES data are available by North American Industry Classification System (NAICS) code to track changes in industry composition of State and regional economies. Monthly data availability also makes CES among the datasets most frequently updated.
- **Quarterly Census of Employment and Wages (QCEW)** program (formerly known as the ES-202 Program) is a cooperative program between the BLS and State Employment Security Agencies (SESA). The QCEW utilizes employment and wage data reported quarterly by all employers subject to unemployment insurance (UI) requirements. These requirements cover more than 95 percent of U.S. jobs. Data are available at the county, MSA, State, and national levels by detailed NAICS industry. Data products include summary tabulations of the number of establishments, number of employees, and total and average wages paid. The data are available approximately five to six months after the end of each quarter. In some states, State and regional agencies may request ES-202 records directly from their respective SESAs, which can include additional detail and geographic location information, which is not available through the Federal program. The QCEW is most helpful for planners looking for detailed information on the number of jobs at county and regional levels, including industry information.

- **County Business Patterns (CBP)** is an annual series produced by the U.S. Census that provides business establishment and selected employment data at a county, congressional district, and ZIP-code level by NAICS industry. CBP data include the number of establishments, total employment, first quarter payroll, and annual payroll. These dataset also includes information on employment by establishment size for industries, allowing planners to understand the size of businesses in their area. CBP also provides more granular geographic data than other Federal employment establishment-based data sources. County and ZIP code historical data are available from 1998 on, and congressional district data are available from 2013 on.
- **Nonemployer Statistics (NES)** is an annual series produced by the U.S. Census that provides economic data for businesses that have no paid employees and are subject to Federal income tax. Data include the number of businesses and total receipts by industry. Most nonemployers are self-employed individuals operating unincorporated businesses and are not typically counted in establishment-based datasets. Data are available at the national, State, and county levels. Nonemployer data can help transportation planners estimate total workers in an area, and can be used to supplement other employment data sources.
- **Economic Census** products include the U.S. Census annual Economic Census, as well as additional related products, such as the Commodity Flow Survey and Statistics of Foreign Trade Program. The Economic Census dataset provides industry-level detail and supplemental information and the national, State, county, and subcounty levels. For major NAICS industries, data are available on the number of establishments; value of sales; shipments; receipts; and revenue, payroll, employees, and nonemployers. The Statistics of Foreign Trade Programs is an annual U.S. Census program that tracks the value, commodity type, and overseas shipment mode of international exports and imports. Data are available at the national and State levels, and to a lesser extent for metropolitan statistical areas. Employment supported by export activity is estimated by the U.S. Census for States. These data can help transportation planners understand the movement of goods within geographic areas.
- **American Community Survey (ACS)** is a continuous U.S. Census survey program that provides illustrative and detailed demographic and economic information on an annual basis. The ACS collects demographic, employment, income, commuting, housing, and a number of other social data characteristics from a sample of U.S. households. Data geographies available include nation, State, county, county subdivision, census tract, block group, place, consolidate city, congressional district, school district, metropolitan and metropolitan areas and principal cities, and ZIP code tabulation areas.

ACS data have been extensively used by transportation planners to understand household demographics and employment at a wide range of geographic levels. The ACS provides information based on both place of residence and place of work. Journey-to-work information is helpful in understanding commuting patterns and includes data on overall travel time, time of departure, and means of transportation. These data are available for any number of demographic, economic, and social variables. ACS data are useful in calibration and validation of travel demand models; and to inform trends and policies in State, regional, community, and corridor transportation plans.
- **Census Transportation Planning Products (CTPP)** is a set of special tabulations of the ACS journey-to-work questions designed by transportation planners specifically for transportation planning applications. The most current CTPP data tables are based on the 2006 to 2010 (five-year) ACS data. Although additional ACS surveys have been conducted since 2010, the CTPP special tabulations do not yet incorporate data from these more recent surveys. The CTPP tables are available for residence, workplace, and flows between residence and workplace—including trip characteristics, such as means of travel and travel time. CTPP data includes nearly 200 residence-based tables, 115 workplace tables, and 39 flow tables for different levels of geographic resolution ranging from States to metropolitan areas, counties, places, census tracts, and transportation analysis zones, where available.
- **LEHD** is a program developed and administered by the U.S. Census Bureau that combines administrative records, which are collected by the Census Bureau, BLS, Internal Revenue Service (IRS),

and other Federal and State agencies, to produce a unique system of data on employer-worker interactions that is not available from any other source. Unlike most other employment databases, the LEHD Program integrates data covering both employers and individual workers. This enables unique estimations of commute characteristics. Statistical noise and sampling techniques are used to prevent disclosure of identifying information, but the resulting data are robust and offers many applications in transportation planning.

The LEHD produces two principal data products: 1) the Quarterly Workforce Indicators (QWI) and 2) the LEHD Origin-Destination Employment Statistics (LODES). QWI is a set of economic indicators that provide information on employment, job creation, wages, and worker turnover at various levels at State, county, and metropolitan areas by industry and for some worker characteristics. LODES provides information on the geographic locations of employer worksites, worker residences, and residence to worksite flows, as well as key employer and worker characteristics, including industry type, firm size, worker age, gender, and income level.

Of most interest to transportation planners, LODES data include characteristics of workers and employers summarized for various levels of geography that define worker residence locations, workplace locations, and flows of workers between their residence and workplace location. LODES data are organized by State and include residence area, workplace area, and origin-destination information. This data are most readily accessible through the LEHD OnTheMap Web site, which allows users to define various geographies, draw custom areas, and link to home and work areas. These dataset can be used to develop summary information on commute flows by direction, commute flows to and from home residence areas and workplace areas, and to estimate the number of jobs within small geographic units.

- **YourEconomy Time Series (YTS)** is a new time series database developed by the Business Dynamics Research Consortium (BDRC). BDRC is a public-private, nonprofit research institute, affiliated with the University of Wisconsin-Extension, and devoted to the study of business performance and economic growth. YTS uses annual snapshots from the InfoUSA business establishment data as its primary data source, but combines the annual data into a consolidated, multiyear database; verifies and corrects attributes using other data sources; and appends additional attribute fields, such as congressional district, identification of new businesses, or summary measures based on geographic or temporal aggregations. These data are free and open-source and available at the State, county, and combined metropolitan statistical area. The unique trend information, offered by YTS, can inform State, regional, and local plans and help illustrate changes in the economy of a given area.
- **InfoUSA** is a commercial provider of business and consumer information products. InfoUSA maintains a proprietary database that includes establishment information for industry type, number of employees, and business location, among other attributes. Industries are classified by industry name, Standard Industrial Classification (SIC) code, and NAICS code. InfoUSA provides data in geographic information system (GIS) format for geocoded business addresses. For transportation planners and for model estimation and validation, latitude and longitude information for business locations can be particularly useful. Data are available through a paid subscription and subject to licensing and reuse restrictions.
- **Dun & Bradstreet, Inc. (D&B)** is a business services company that provides commercial data and analysis services to businesses. In 1963, D&B introduced the Data Universal Numbering System (DUNS), which is a unique nine-digit numeric identifier used as the de facto standard in national reporting. DUNS numbers are assigned to each business location in the D&B master business database. The proprietary database includes information on industry type, number of employees, and business location in GIS. Subscribers can request records for custom geographies at any level. Data are available through a paid subscription and subject to licensing and reuse restrictions.
- **Woods & Poole** is a business service provider that maintains a comprehensive database of employment data, along with demographic information for every State, region, county, and metropolitan area in the

U.S. for every year from 1970 to 2050. Woods & Poole county projections are updated annually and utilize projection models that take into account specific local conditions based on historical data. Data are available through a paid subscription and subject to licensing and reuse restrictions.

- **Moody's Analytics** provides historical and forecast data for the U.S. and States, including key economic variables, such as employment, gross product, wages by industry, labor force and unemployment rate, population, personal and household income, housing stock and prices, and retail sales. Forecasts include annual changes out to a 30-year time horizon, and are updated on a monthly basis to reflect the most current economic data, conditions, and expectations. Data are available through a paid subscription and subject to licensing and reuse restrictions.
- **IHS Global Insight** provides economic analysis, risk assessment, forecasting, and scenario planning. Their Business Market Insights (BMI) database provides both historical trends and 25-year forecasts of employment, number of establishments and sales volumes, by industry, for States, counties, metropolitan areas, and Census regions. Data are available through a paid subscription and subject to licensing and reuse restrictions.

Understanding Differences and Distinctions in Employment Data

Total Employment Counts

One of the most frequent observations of users of employment data is the seemingly wide discrepancies in total employment counts from available public sources. Differences in employment numbers are due to base data source, methodologies, excluded populations, and reporting or surveying time-periods.

The CPS is the official source of monthly 'jobs' numbers that are reported by Federal agencies and often seen in the media. Jobs numbers is inaccurate as this source only counts employed workers once, regardless of whether they hold more than one job. The CPS counts also are based on the broadest definition of employment and based on more complete survey samples than other household-based survey programs.

For example, under the CPS series the following types of workers are included in total employment counts: persons who worked part-time (for at least one hour); persons working in their own businesses, professions, or on their own farms; unpaid family workers working 15 hours or more a week; persons not actively working, but who have jobs and are temporarily absent (due to vacation, illness, weather, childcare, maternity or paternity leave, labor-management dispute, job training, or other family or personal reasons, etc.), whether or not time off is paid; and employed citizens of foreign countries who are temporarily in the United States, but not living on the premises of an embassy are included.

Employment data based on business databases maintained by the Census Bureau or by State employment agencies tend to exclude more workers than household-based surveys, but may produce more accurate counts of jobs tied to certain industries or workplaces. These data sources include the CES, QCEW, CBP, and Nonemployer Statistics. These data are not estimates, but tabulated counts based on administrative records. Records may be subject to reporting errors and differences in State-to-State database quality. Some employment data sources, such as CBP, introduce noise to lessen data suppression and avoid disclosures or may produce estimates of employment for establishments not reporting in the current quarter.

Table 1-1 on the following page highlights some of the differences in total employment counts among major public employment data sources. Generally, household-based surveys produce higher employment counts than data developed from administrative UI or payroll records. Differences in reporting methodology exist even among data sources that are drawn from the same base administrative source records. The CES and QCEW, for example, produce different tabulations from the same business database.

Table 1-1. Comparison of total employment counts of major public data sources.

Data Source	Current Population Survey	American Community Survey	Current Employment Statistics	Quarterly Census of Employment and Wages	County Business Patterns
Employment Count	2,982,000	2,624,436	2,598,300	2,553,198	2,253,795
Source Data	Survey	Survey and estimates	Sample	Tabulation and survey	Tabulation with noise infusion
Source Basis	Household survey	Household survey	Administrative establishment records	Administrative establishment records	Administrative establishment payroll
Reference Period	Monthly surveys	Annual surveys	Monthly reporting	Monthly reporting	1 st quarter reporting
Population Universe	Adult civilian workers	Adult civilian employed workers	Nonfarm employees, not seasonally adjusted	Total covered employees	Paid employees
Who is counted?	<ul style="list-style-type: none"> – Civilian workers 16 years and older – Part-time workers – Self-employed business or farm workers – Unpaid family workers – Persons temporarily absent from jobs – Noncitizen foreign employees 	<ul style="list-style-type: none"> – Civilian workers 16 years and older – Part-time – Self-employed – Family workers 	<ul style="list-style-type: none"> – Civilian workers – Workers actively on establishment payrolls – Part-time paid workers 	<ul style="list-style-type: none"> – UI-covered employees 	<ul style="list-style-type: none"> – Civilian workers – Paid workers
Who is not counted?	<ul style="list-style-type: none"> – Active duty military 	<ul style="list-style-type: none"> – Active duty military 	<ul style="list-style-type: none"> – Unpaid workers – Self-proprietors – Self-employed – Unpaid family or volunteer workers – Some farm workers – Household workers – Active duty military 	<ul style="list-style-type: none"> – Unpaid workers – Self-proprietors – Self-employed – Unpaid family workers – Some farm workers – Railroad employees – Elected and judiciary officials – Active duty military 	<ul style="list-style-type: none"> – Unpaid workers – Self-employed – Unpaid family workers – Railroad employees – Agricultural production workers – Most government employees
Duplicate Counts	Employed persons are counted only once, even if they hold more than one job.		Workers on the payroll of more than one establishment are counted in each establishment.		

Note: Total employment example data are for the State of Colorado, 2015.

Public and Private Data Sources

Private employment data sources range considerably in coverage and data quality. Vendors, such as InfoUSA, Dun & Bradstreet, Woods & Poole, Moody's Business Analytics, IHS Global Insight, and others, utilize a variety of administrative records, public filings, and online and in-person verification to compile business records. These databases are subject to reporting and data entry errors, and may contain incomplete records for some establishments.

Comparing public and private data sources can result in significant differences in employment totals. Total employment count may vary based on vendor information, self-reporting of information from an establishment, and classification of employment totals at a particular worksite location. Public employment data sources, such as the QCEW, rely on Multiple Worksite Reports filed as part of administrative UI records to distinguish employee counts among different establishments operated by the same business entity. Private vendors verifying data or utilizing public records to compile information may not have access to the same level of distinguishing information. Industry classification also may vary between public data sources and between private and public data sources. The QCEW program, for example, classifies employment at an establishment under the predominant industry activity, while other programs may split employment into several NAICS classifications based on multiple activities. Private vendors are likely to report total employment based on predominant activity or SIC code, and may also report employment in aggregate for companies that provide temporary workers, employee leasing, payroll services, or other administrative headquarters establishments.

Table 1-2 below illustrates differences in manufacturing employment and establishment counts in 2016 for a county in Colorado for the QCEW program and a major private vendor. Compared to the administrative record based counts from QCEW, the private data source includes significantly more establishments and fewer total employees.

Table 1-2. Example employment and establishment counts, public and private data sources.

Manufacturing (Boulder County, Colorado, 2016)	QCEW	Private Vendor
Employment	17,600	14,424
Establishments	629	880

Private vendor data may include more business establishments because databases are compiled from many different sources, including public legal records, credit rating agencies, sales and trade transactions, banking data, real estate transactions, and corporate filings, among other sources. Some of these records may be duplicates or include corporate entities with no employees. Exact employment data is not available for all records. In the example above, of establishment records in this database, 90 percent contain specific employment figures. Most private vendors that provide individual business information also include employment totals in ranges, or as rounded totals that may be useful for mapping applications and planning purposes, though less helpful in estimating specific employment figures.

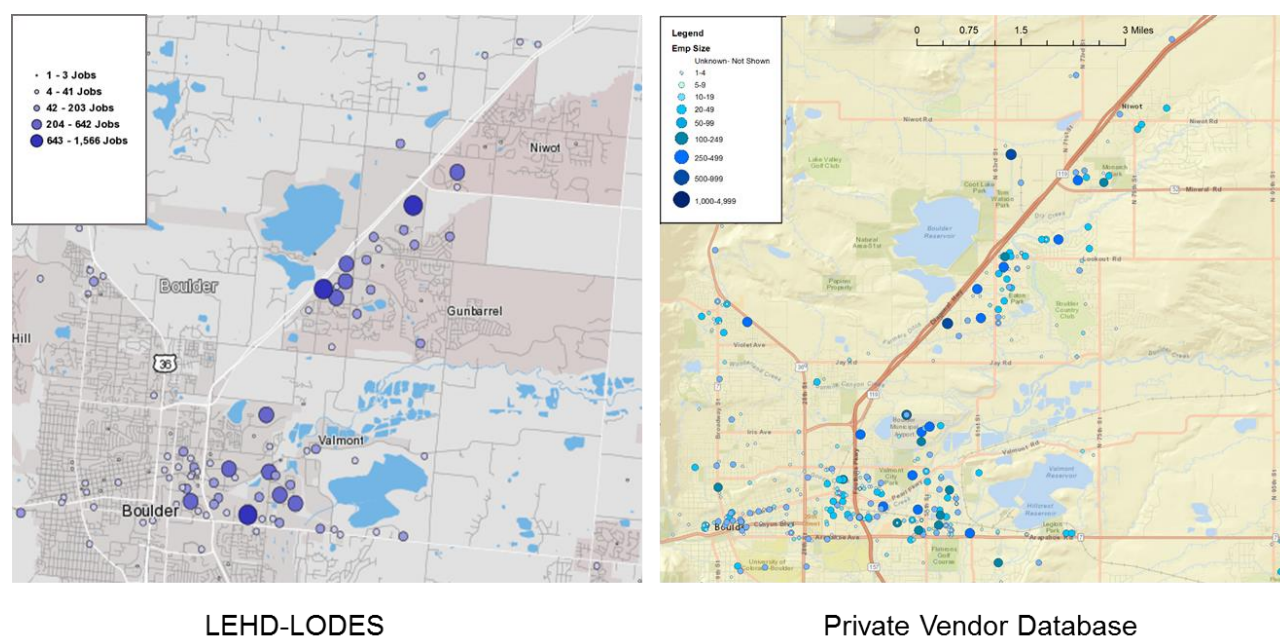
Spatial Distribution of Employment

Public and private data sources enable planning practitioners to map employer establishments, to assign employment to traffic analysis zones, or to examine spatial distribution patterns. However, all data sources are imperfect. Private data vendor information may not be accurate or address and latitude/longitude

information contained within these databases may not be correlated to the actual business location. Public data sources, such as LEHD-LODES and the data available within the OnTheMap application, introduce noise and use statistical sampling techniques so that locations of establishments are not exact representations. Generally, employment totals for a given geography are not impacted by noise infusion techniques. Location differences can be significant depending on the geography being examined. Even with noise infusion and inexact location data, these data sources may be acceptable for high-level planning and forecasting.

The maps in Figure 1-1 compare the location and employment size classification of manufacturing establishments in a portion of Boulder County, Colorado, from the LEHD-LODES OnTheMap application and a private vendor database for the same base year. While the general patterns and spatial distribution of employment is similar and activity centers are identifiable, there are significant differences.

Figure 1-1. Example spatial distribution and employment patterns, public and private data sources.



Source: Analysis of LEHD-LODES OnTheMap data and private vendor data.

The private vendor database includes records for 880 establishments supporting 14,424 jobs. The LODES database provides primary job counts total 18,525 for the same county area. Establishment counts are not available through LODES. Employment size ranges differ between the two sets. The private database suggests a larger number of businesses operating in the area. Many of these records are small employers with just one to four employees. In the private database, physical locations of these smaller establishments are often in primarily residential areas, or areas where manufacturing activity would not be likely. However, LODES physical locations are intentionally diffused with statistical sampling technique to avoid disclosure—particularly of small establishments in small geographic areas. These examples illustrate the care that should be taken when utilizing both public and private data sources to examine spatial distribution of employment and establishment locations.

CHAPTER 2

Frequently Asked Questions

The following chapter provides an overview of frequently asked questions and commonly encountered challenges when using employment data to inform planning activities. These questions were developed with input from practitioners across State departments of transportation (DOT) and regional organizations. While not exhaustive, this information and additional resources and links provide a starting point to determine what data and information can be used and when.

What Are the Differences Between Jobs, Workers, and Employment?

When it comes to counting employment in a given area, different datasets use differing definitions. For datasets that compile data at the establishment level, an individual worker working more than one job can be counted more than once. For household-level datasets that count employed persons, each worker is counted only once. Household survey programs count employed persons at their place of residence, while establishment-based programs count jobs at the place of work.

For household survey products, such as the American Community Survey, workers include all civilians 16 years old and over who either are paid employees, worked in their own business or profession, worked on their own farm, or who worked as unpaid workers in a family farm or business, as well as workers temporarily absent from their jobs. Excluded from the employed are people whose only activity consisted of work around the house or unpaid volunteer work for religious, charitable, and similar organizations; also excluded are all institutionalized people and people on active duty in the U.S. Armed Forces.

For most establishment-level products prepared by the BLS, jobs numbers include full-time, part-time, or temporary jobs. Datasets based on UI records, such as the QCEW or LEHD, exclude self-employed workers, farm workers, members of the U.S. Armed Forces, elected officials in most States, most employees of railroads, some domestic workers, most student workers at schools, and employees of certain small nonprofit organizations. The LEHD Origin-Destination Employment Statistics, which provides data on total jobs in an area, also provides a concept of workers by reporting data based on ‘primary job’. A primary job is the highest paying job for an individual worker for the year.

For most planning purposes, including examining origin and destination commute flows or estimating travel demand models, establishment-based job counts are preferable as they can be linked to physical locations. For reporting or performance indicators, employment data can provide a picture of the total number of employed, or unemployed, workers in a given area.

How Does a Company Differ from an Establishment?

Many employment datasets compile data at the establishment level, which can be commonly misinterpreted as a single business, company, or enterprise. Establishments are physical locations with commercial activity, and can be owned and operated by a single company or can be a subsidiary location of a larger company or enterprise that operates one or more establishments.

Establishments typically are a single physical location that is engaged in a primary type of economic activity for which a single industrial classification code can be applied. A company may consist of one or more establishments with each establishment typically engaged in the same economic activity or industrial classification. An enterprise is a larger unit of analysis, and is usually a company that engages in economic activities, which are classified into multiple industries. Franchise restaurants or retail stores are common examples of challenges using employment data to derive counts of companies in a given area. How these franchises are set up as corporations dictates how they are counted—either as independently owned and operated companies, or as establishments of a single corporate parent company.

Datasets from the BLS, including the CPS and QCEW, generally compile and report employment data by establishments. Note that the QCEW program attempts to collect information at the establishment level; however, some companies, with multiple establishments, may report as a single entity. For company or business-level data, the Business Employment Dynamics program reports information by company.

According to the BLS, most companies are single establishment. For planning applications and decision making, establishment-level data are more suggestive of real-world commercial activities and employment tied to a physical location.

Who Is Counted as a Worker at an Employment Site?

For many planning and modelling purposes, the data of most interest is that which links workers to particular geographic locations or employment sites and industry types. This information enables estimates of trip generation, commute flows, identification of activity centers, or prioritization of transportation needs. Most commonly used employment datasets generate reliable estimates of workers at particular locations; however, there are some exceptions.

Datasets such as the QCEW, LEHD, and CTPP, use a combination of company surveys and required reporting of administrative records to assign workers to establishments, and to classify establishments within industries.

Most employers have only one establishment or place of business activity. For companies with more than one establishment, data are collected and reported for workers and wages at each location. There are some exceptions. For employers with several very small establishments, or for employers with one large establishment and a single smaller establishment, workers are reported for only one location. Some companies cannot disaggregate and report establishments separately, and instead are reported for consolidated locations. Reporting requirements can vary between States, and discrepancies in reporting single or multi-establishment data can lead to over or undercounts of workers and may need to be independently verified.

Establishment-level datasets report only aggregate numbers of workers and do not distinguish among occupations, job types, or employment status. Jobs reported as associated with an establishment include most corporate officials, executives, supervisory personnel, professionals, clerical workers, wage earners, piece-workers, and part-time workers. Jobs not covered in establishment counts include sole proprietors, many unincorporated self-employed workers, unpaid family members, and certain farm and domestic workers.

Until 2012, Federal workers and self-employed jobs were not available in the LEHD jobs data, due to their exclusion from State UI programs. LEHD began integrating two additional data sources into the LEHD microdata: 1) Office of Personnel Management (OPM) records on the Federal workforce and 2) tax data on the self-employed. The OPM data have been successfully integrated into the LEHD microdata and were included in the 2012 release for LODES. Inclusion of the self-employed workforce in LEHD data reveals the degree of interconnection between the wage and salary workforce and those who start their own businesses. The expanded frame permits a separate characterization of workforce dynamics for this group

and makes it possible to produce statistics that capture the flows between wage and salary work and business ownership.

What Does Labor Force Mean?

Labor force statistics are the basis of most commonly reported employment and unemployment data at all geographic levels. The labor force includes all persons 16 years old and over who are either employed or unemployed or on active duty in the Armed Forces. Not included in the labor force are all persons 16 years old and over who are not currently working. Not included in the labor force are commonly students, homemakers, retired workers, seasonal workers who were not looking for work, institutionalized people, and persons performing incidental unpaid family work (less than 15 hours a week).

A subset of persons in the labor force is considered marginally attached or discouraged workers—sometimes referred to as underemployment. Marginally attached persons are not actively in the labor force, but who are available and want to work and have looked for a job in the prior 12 months. These persons are not counted as unemployed because they had not searched for work in the four weeks preceding the time that they were surveyed by the BLS. Among the marginally attached, discouraged workers currently were not looking for work, specifically because they believed no jobs were available for them or there were none for which they would qualify.

What Are Some Common Issues that Affect the Accuracy of Employment Data?

All employment data sources involve some level of surveying, sampling, estimation, and even intentional data suppression. Data gathered is imperfect and subject to survey response influences; administrative reporting errors; and limitations on how data is compiled, verified, categorized, and classified. All major public data source programs include quality assurance and verification steps in the data compilation process. Limitations of original source records and assignment and classification errors can create issues that affect the accuracy of employment data.

Multiple Site Employment Reporting

The underlying source data for many public data sources are drawn from UI records which in turn rely on employers to self-report employment associated with an enterprise location, including multiple site workplaces. Not all employers are able or required to report employment by location, and State rules governing reporting can vary. This may result in data issues where employment is over reported for corporate headquarters or addresses associated with the administration of payroll records, and underreported for actual work place locations. Most multi-location employers with a total of 10 or more employees combined in their secondary locations or establishments are required or requested to report multiple worksite reports. An employer's primary location is the location with the most employment in a State and all other locations within a State are considered secondary locations. The LEHD program researchers continue to introduce verification steps in data processing to correct known issues; however, the output data is only as good as the initial State-level reported data.

Industry Classification of Establishments

Data source programs also may classify employment at the establishment and within the establishment differently, particularly when classifying primary industry activities. For example, the QCEW program only reports the predominant economic activity at an establishment, while other data programs or commercial vendors also may include non-predominant economic activity within counts. An example provided by the BLS is a metal parts manufacturing factory that would be classified under manufacturing by the QCEW

program. This example factory also may perform metal plating on the parts as part of a package or deal for end-users. Under the QCEW count, this establishment would be counted only under manufacturing, but not plating because it is not the factory's predominant economic activity or output.

Data Suppression, Disclosure, and Estimation

All data publicly available is subject to data reporting standards, and published records may be suppressed or anonymized to avoid revealing confidential information. For example, the QCEW data are regularly suppressed when releasing data would enable a data user to estimate the value of an individual reporting entity too closely. These standards result in data not disclosed particularly in small geographic areas or for detailed industry classifications with only a few establishments operating. Data at national and State levels are not suppressed and include full counts of establishment records.

The U.S. Census Bureau's LEHD program uses administrative records to match employee and employer, using sensitive information. Noise infusion techniques are purposely introduced into LEHD datasets to allow for the release of data that might otherwise be considered confidential. Noise infusion results in modifications to data for every record. For a single workplace, data is always modified in the same direction and by the same proportion as prior data releases. The LEHD Origin-Destination Employment Statistics data source is a partially synthetic data release. Workforce characteristics by place of residence are synthesized based on underlying microdata. Workforce characteristics by place of work are not synthesized.

Commercial Vendor Data

Most business databases provided by commercial vendors are built on public records or administrative databases required at the State or local level. These databases may include business registrations; legal filings; corporate or doing business as name registers; and even public utility, tax records, and insurance filings. The quality, accuracy, and availability of information in these business databases across States and local jurisdictions may vary. Some States require annual business registrations, while other States or local governments may require periodic registration or may not adequately purge records of entities no longer in business. These data sources may result in a number of entities that are not active, not accurate, or not employers. Commercial vendors verify records through other databases, surveys, and live phone callers to ensure quality and accuracy. However, data entry errors, self-reporting errors from live verification, and database issues can result in unlikely, duplicative, or erroneous business records.

How Can I Validate Data and Check for Data Weaknesses or Issues?

Checking employment data for apparent issues, weaknesses, or inaccuracies can be challenging for many transportation agencies. Validating individual establishment records may require significant staff resources, or be improbable to accomplish under project or program timelines.

Smaller geographic units tend to have higher error probabilities, disclosure, or confidentiality issues. Checking total jobs counts in these areas can be accomplished by comparing datasets for reasonableness, acknowledging differences in worker counts and data source origin. Checking against trend and historical data of the same data source can be used to highlight potential reclassification issues or potential errors. When updating travel demand model analysis areas with new data or with a different data source, simple checks of the percent of workers located in certain areas can highlight significant discrepancies. American Community Survey data and other household survey-based data report margin of error brackets that should be examined for particularly significant errors. Margin of errors in survey and estimated data tend to be large for smaller sample sizes, particularly in rural areas, small geographic areas, or for detailed data segmentation and cross tabulations.

Employment data weaknesses can be particularly apparent when point-specific establishment locations are available, either through data purchased from commercial vendors or unique regional databases developed from state agency ES-202 data. Once mapped, business street address or coordinates can be verified through satellite imagery and web-based or public records searches. GIS tools can be effective to eliminate business locations where significant employment activity is likely to be located (e.g., parking lots, open space or undeveloped properties, residential areas). Local knowledge is instrumental in basic checks, and planners from local agencies or governments can be engaged to spot-check data. For data identifying major employers in an area, phone calls can be made to verify basic employment size, and even industry classification.

Commercial datasets that provide business names and street addresses can show significant errors for particular business activities. A 2014 study by the Initiative for a Competitive Inner City in Boston highlighted issues with a commercial vendor database. Research at the neighborhood level in Boston found that 43 percent of businesses in a leading commercial database were not found in public city or State databases. A walking inventory of commercial districts in inner city Boston revealed that 30 percent of businesses in the same commercial database did not exist, and 380 businesses were identified that were not included in the database. Bank branch locations and financial services providers (e.g., H&R Block) may show as ATMs or self-service kiosks where no employment activity takes places. Similarly, major franchise operations in retail can be examined in more detail. Business with one or few employees may often need additional scrutiny as records may not be as readily verified, and many incorporated businesses may not actually be engaged in business activity. Drawing samples from commercial vendors and attempting web-based or phone verification may be practical for smaller datasets. For larger datasets, consistency checks with publicly reported employment data and spatial analysis may help eliminate significant errors.

How Can I Combine or Reconcile Disparate Employment Estimates?

Each of the employment data sources profiled in this guide are distinct in survey and estimation methods, population universe and employment coverage, survey time period, industry classification, treatment of establishments with more than one primary industry activity, and other critical factors. As a result, worker or employment counts for the same area can generate different employment estimates.

Counts from different data sources should not be aggregated, but can be reported separately with major differences acknowledged. For example, worker counts from the QCEW can be reported along with U.S. Census Nonemployer Statistics alongside counts of farmworkers from the Census of Agriculture, but they should not necessarily be summed. Similarly, measures such as industry employment, class of worker, and occupation of worker are all different concepts and should not be considered together. Discrepancies between similar measures or employment concepts drawn from different data sources are typically relatively small in total, but can be significant at industry levels or at discrete geographies. Reclassification of establishments or reporting practices by entities also can shift employment records from year to year or between geographies. For most planning and modeling purposes, a consistent baseline dataset can be chosen. When integrating new data sources into a model, for example, standard deviations can be established to test data comparability and enable comparisons.

The CPS provides the highest response rates among government household surveys, averaging around 90 percent. These data are only available at the State level and track workers, rather than jobs. The QCEW program is well established and relies on business reporting records for UI requirements. Employment covered by this program represents about 97 percent of all wage and salary civilian employment in the U.S.

What Are the Advantages of Commercial Vendors Compared to Free Public Sources?

Generally, commercial data sources provide information that is not otherwise available from public sources. However, the reliability and accuracy of commercial vendors are not proven and may vary depending on geography, information or data detail, and the vendor providing the data. Most major commercial business databases utilize public records on incorporation or DUNS numbers as a foundation for databases. These records are then verified by phone or email on a periodic basis. Business addresses found in these databases may not match actual business or establishment locations. Depending on the time of record verification, some temporary establishments (e.g., seasonal tax preparation or holiday themed retail stores) may remain in the database. In addition, franchise stores, ATMs, mobile rental centers or kiosks, and other nonemployment activity business addresses may be included in databases drawn from public records.

Some commercial vendors provide street address or latitude and longitude coordinates that can be utilized in GIS to perform analysis of employment centers or industry clusters. This level of detail is not typically available from public data sources. Some SESAs that administer the ES-202 program do offer microdata for agency use. Commercial vendors also may provide additional data not available from public sources, which include detailed employment and industry forecasts, as well as other data characteristics that could be utilized, including SIC codes and self-reported business classifications.

How Do Different Data Sources Address Persons Who Are Self-Employed or Who Work from Home?

Capturing reliable data on the number of self-employed workers in an area is challenging and problematic. Self-employed workers are often defined by legal arrangements of a business, including incorporated businesses, limited liability corporations, and partnerships. Self-employed workers also are sometimes counted as single employee employers depending on legal arrangements.

Self-employed workers are defined by the U.S. Census Bureau within the American Community Survey and CTPP datasets as employed in their own business that is not incorporated. If a business is incorporated, that worker is likely counted as an employee. For payroll and establishment-based data sources, such as QCEW, self-employed persons are not counted. It is unclear how private vendors of commercial databases may treat incorporated or unincorporated businesses with a single employee. Most commercial datasets rely on incorporation of business records as the foundation of databases. The U.S. Census Nonemployer Statistics data source provides comprehensive accounting of nonemployers—these are businesses that have no paid employees and are subject to Federal income tax. Most nonemployers are self-employed individuals operating unincorporated businesses (known as sole proprietorships), which may or may not be the owner's principal source of income. The majority of all business establishments in the United States are nonemployers.

For workers working from home, the American Community Survey means of transportation to work datasets provide an option for percent of workers working from home by various demographic characteristics. This is the only reliable time-series data on telecommuters available for most geographies. An increasing number of workers also are considered part of the gig economy and may not appear in data sources, even those counting nonemployers. These workers file IRS 1099 Misc Income reports and can include occupations, such as drivers for ride-hailing services, freelance journalists, couriers, on-demand service providers, and other 'employees' of businesses who are treated as contractors. Currently, no data source exists to track on-demand workers as a distinct covered population.

How Are Hard-to-Reach or Transitional Employment such as Migrant Farmworkers or Military Service Members Accounted for?

Unfortunately, most employment data sources do not account for transitional employment, including farmworkers and military service members. Establishment-based data sources drawn from payroll and UI records generally do not account for active duty U.S. Armed Forces personnel, self-employed workers, most agricultural workers on small farms, elected officials in most States, most employees of railroads, some domestic workers, most student workers at schools, and employees of certain small nonprofit organizations. Generally, these populations make up a small proportion of total employment in a given area.

However, military bases and areas with significant agricultural activities can see large changes in staffing and seasonal employment. Reservists in the armed forces are counted among workers in datasets if they are also employed, full or part-time at a reporting enterprise. If reservists are self-employed, farm workers, or unpaid family workers, they would not be counted. The U.S. Census develops special supplemental reports to decennial census surveys that provide counts of active duty personnel. Personnel strength annual reports by State and service branch are available from the Department of Defense, Defense Manpower Data Center. The Department of Agriculture administers the Census of Agriculture once every five years to gather information on workers, land use, ownership, production, income, and other statistics. These data source is available at the State and county levels.

CHAPTER 3

Quick Reference Guide

This chapter provides users with a quick reference guide to readily identify relevant definitions, methods, and considerations within this Resource Guide, based on what goal, geography, data detail, and level of analyses are being considered. The following set of tables (Table 3-1 through Table 3-5) can be used to determine which data source may be the best fit for a desired analysis or geography. Not all data sources provide the same level of information at all geographies, and differences exist in data coverage and characteristics available. In many cases, it may be desirable to use different data sources to answer different data questions. However, combining or aggregating data on counts of establishments, jobs, or workers is problematic and should be avoided.

The common abbreviations below are used in the quick reference tables in the following sections.

- (CPS) Current Population Survey
- (CES) Current Employment Statistics
- (QCEW) Quarterly Census of Employment and Wages
- (CBP) County Business Patterns
- (NES) Nonemployer Statistics
- (ACS) American Community Survey
- (CTPP) Census Transportation Planning Products
- (QWI) Quarterly Workforce Indicators
- (LODES) LEHD Origin-Destination Employment Statistics
- (YTS) YourEconomy Time Series
- (USA) InfoUSA
- (D&B) Dun & Bradstreet, Inc.
- (W&P) Woods & Poole
- (MAS) Moody's Analytics
- (IHS) IHS Global Insight

Table 3-1. What geographies are available?

Geography	CPS	CES	QCEW	CBP	NES	ACS	CTPP	QWI	LODES	YTS	USA	D&B	W&P	MAS	IHS
Nation	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
State	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Metropolitan Statistical Area	<input checked="" type="checkbox"/> a	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Congressional District				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
County			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Subcounty (e.g., City, Place, Zip Code etc.)				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
Analysis Zone (e.g., TAZ, Census Tract, Census Block)						<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
Point-Specific (e.g., Latitude/Longitude, GIS)											<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			

CPS—Current Population Survey, CES—Current Employment Statistics, QCEW—Quarterly Census of Employment and Wages, CBP—County Business Patterns, NES—Nonemployer Statistics, ACS—American Community Survey, CTPP—Census Transportation Planning Products, QWI—LEHD Quarterly Workforce Indicators, LODES—LEHD Origin-Destination Employment Statistics, YTS—YourEconomy Time Series, USA—InfoUSA, D&B—Dun & Bradstreet, Inc., W&P—Woods & Poole, MAS—Moody's Analytics, and IHS—IHS Global Insight.

^a CPS data are limited to the 12 largest metropolitan statistical areas.

Table 3-2. How frequently is data updated?

Frequency	CPS	CES	QCEW	CBP	NES	ACS	CTPP	QWI	LODES	YTS	USA	D&B	W&P	MAS	IHS
Periodically							☑								
Annually				☑	☑	☑			☑	☑					
Quarterly			☑					☑							
Monthly	☑	☑													
Unknown or Not Available											☑	☑	☑	☑	☑

CPS—Current Population Survey, CES—Current Employment Statistics, QCEW—Quarterly Census of Employment and Wages, CBP—County Business Patterns, NES—Nonemployer Statistics, ACS—American Community Survey, CTPP—Census Transportation Planning Products, QWI—LEHD Quarterly Workforce Indicators, LODES—LEHD Origin-Destination Employment Statistics, YTS—YourEconomy Time Series, USA—InfoUSA, D&B—Dun & Bradstreet, Inc., W&P—Woods & Poole, MAS—Moody’s Analytics, and IHS—IHS Global Insight.

Table 3-3. What data coverage is available?

Universe	CPS	CES	QCEW	CBP	NES	ACS	CTPP	QWI	LODES	YTS	USA	D&B	W&P	MAS	IHS
Workers	☑				☑	☑	☑	☑	☑						
Jobs		☑	☑	☑				☑	☑	☑		☑	☑	☑	☑
Military															
Unpaid Workers	☑														
Self-Employed	☑			☑ ⁽¹⁾	☑	☑	☑			☑					
Farm Workers	☑														
Private Employment	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑		☑	☑	☑	☑
Public Employment	☑		☑			☑	☑	☑	☑	☑		☑	☑	☑	☑

CPS—Current Population Survey, CES—Current Employment Statistics, QCEW—Quarterly Census of Employment and Wages, CBP—County Business Patterns, NES—Nonemployer Statistics, ACS—American Community Survey, CTPP—Census Transportation Planning Products, QWI—LEHD Quarterly Workforce Indicators, LODES—LEHD Origin-Destination Employment Statistics, YTS—YourEconomy Time Series, USA—InfoUSA, D&B—Dun & Bradstreet, Inc., W&P—Woods & Poole, MAS—Moody’s Analytics, and IHS—IHS Global Insight.

^a CBP includes employment data from establishments with 1 paid employee. A self-employed business owner may be an employer, though the majority of nonemployer business owners are self-employed.

Table 3-4. What data characteristics are available?

Characteristics	CPS	CES	QCEW	CBP	NES	ACS	CTPP	QWI	LODES	YTS	USA	D&B	W&P	MAS	IHS
Employment	☑	☑	☑	☑		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Unemployment	☑		☑			☑									
Establishments		☑	☑	☑						☑		☑	☑	☑	☑
Firms/Companies					☑			☑							
Industry Classification		☑	☑	☑	☑	☑	☑	☑	☑			☑	☑	☑	☑
Occupation		☑				☑	☑								
Demographics	☑					☑	☑	☑	☑				☑		☑
Income and Earnings	☑	☑	☑	☑		☑	☑		☑				☑		
Commute Time and Mode						☑	☑								
Origin-Destination						☑	☑		☑						

CPS—Current Population Survey, CES—Current Employment Statistics, QCEW—Quarterly Census of Employment and Wages, CBP—County Business Patterns, NES—Nonemployer Statistics, ACS—American Community Survey, CTPP—Census Transportation Planning Products, QWI—LEHD Quarterly Workforce Indicators, LODES—LEHD Origin-Destination Employment Statistics, YTS—YourEconomy Time Series, USA—InfoUSA, D&B—Dun & Bradstreet, Inc., W&P—Woods & Poole, MAS—Moody’s Analytics, and IHS—IHS Global Insight.

Table 3-5. What questions are being asked?

Questions	CPS	CES	QCEW	CBP	NES	ACS	CTPP	QWI	LODES	YTS	USA	D&B	W&P	MAS	IHS
How do workers commute?						<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>								
Where do workers live and work?						<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>						
How many workers are in an area?	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>						
How many jobs are in an area?		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
How many establishments are in an area?		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>						<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
What does the workforce look like?				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>						
What does business activity look like?			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
What industries are in an area?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
How has the economy changed?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>					
How will the economy change in the future?													<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Where are employment centers in an area?							<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Where are establishments located?											<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			

CPS—Current Population Survey, CES—Current Employment Statistics, QCEW—Quarterly Census of Employment and Wages, CBP—County Business Patterns, NES—Nonemployer Statistics, ACS—American Community Survey, CTPP—Census Transportation Planning Products, QWI—LEHD Quarterly Workforce Indicators, LODES—LEHD Origin-Destination Employment Statistics, YTS—YourEconomy Time Series, USA—InfoUSA, D&B—Dun & Bradstreet, Inc., W&P—Woods & Poole, MAS—Moody's Analytics, and IHS—IHS Global Insight.

CHAPTER 4

Data Sources and Applications

This chapter presents key elements and characteristics for common public and private employment data sources. For each data source, information on methods, coverage, geography, and acquisition is summarized, as well as high-level information on limitations, benefits, and potential applications.

The information included in this review is drawn from Web sites and technical documents available from key Federal agencies who collect and maintain employment data (e.g., BLS and U.S. Census Bureau), and from Web sites and marketing materials provided by various commercial database developers. A glossary of technical terms is included in Chapter 6. Words or phrases displayed in *italics* are defined in the glossary. The following data sources are summarized in this guide:

- Public data sources:
 - Current Population Survey
 - Current Employment Statistics
 - Nonemployer Statistics
 - Quarterly Census of Employment and Wages
 - County Business Patterns
 - American Community Survey
 - Census Transportation Planning Products
 - Public Use Microdata Sample
 - Longitudinal Employer Household Dynamics Program
 - Quarterly Workforce Indicators
 - LEHD Origin-Destination Employment Statistics
- Commercial data sources:
 - InfoUSA
 - Dun & Bradstreet
 - YourEconomy Time Series
 - Woods & Poole
 - Moody's Business Analytics
 - IHS Global Insight

4.1 Current Population Survey

The CPS is a joint effort by the U.S. Census Bureau (Census) and the U.S. BLS, and serves as the primary source of [*labor force*](#) statistics for the population of the United States. The CPS collects extensive demographic information that helps planners, modelers, and decisionmakers to understand the demographic composition of the labor force at the national, State, and sub-State levels. The BLS uses the CPS data to develop its monthly labor force statistics. The monthly files are available within 30 to 45 days after data collection.

Data Coverage

The CPS survey universe includes all noninstitutionalized civilians 16 years of age or older, regardless of their employment status, type of employment, or type of industry. The survey sampling unit is a household or noninstitutional group quarter. The CPS does not survey uniformed military personnel or people residing in institutional group quarters. The survey collects demographic and employment information on each eligible household member. No information is collected on commuting behavior, such as workplace location, travel time to work, or means of transportation.

Data Collection Methodology

The current population survey is conducted by Census Bureau on a sample of about 60,000 occupied *housing units*. The CPS sample housing units are obtained from lists of addresses from the 2010 Decennial Census of Population and Housing. Independent samples are selected for each State and the District of Columbia. The samples for each State are tailored to the demographics and the labor market conditions that prevail in that State.

Each sampled household is surveyed once each month for four consecutive months, then dropped from the survey for eight months, and finally surveyed again once each month for four consecutive months of the following year. After that, the household is removed from the sample list. This means that each sample household is surveyed eight times over two consecutive years. The 4-8-4 sampling scheme has the added benefit of allowing the constant replenishment of the sample without excessive burden to respondents.

The survey samples are weighted to represent the universe of the population. The weight is the inverse of the probability of selecting that sample. Technical documentation of weighting and sample methodology is available on the U.S. Census Bureau CPS website.

Data Geography

CPS data provides reliable estimates at the State level and for large metropolitan areas (i.e., more than 500,000 population). Due to the sample size, it is not reliable at or below the county level. Furthermore, not all counties are sampled in the CPS. The data can be used for smaller metropolitan areas (with population less than 500,000), but should be used with caution due to high variability associated with the sample estimates. For these areas, estimates comparing percent distributions and ratios will provide data with less sampling variability than will estimates of absolute levels.

Data Content

The CPS data contains estimates of employment, unemployment, and characteristics of the labor force on a monthly basis. It provides the following information:

- Worker characteristics—Age, gender, race/ethnicity, income, family size/status, education level, and school enrollment.
- Employment type—Waged/salaried, Self-employed.
- Employment status—Full-time/part-time, Unemployed.
- Employment characteristics—Industry codes; occupation codes.

Industry codes represent the Census 2002 classification system, and were developed using the NAICS. Occupation codes represent the Census 2010 classification system, and were developed using the Standard Occupational Classification (SOC) Manual: 2010.

Data Acquisition and Sources

Current CPS data tables can be downloaded from Census Bureau's CPS Web site:

<http://www.census.gov/programs-surveys/cps.html>

This Web site also includes a link to the DataWeb FTP site, where users can download past CPS data tables by month. It also provides links to two data selection tools: "DataFerret" and "The CPS Table Creator." These tools enable users to create custom datasets and tables based on their specific needs. The data are available in both seasonally-adjusted and seasonally-unadjusted forms.

Data Application

The CPS provides employment estimates at the national, State, and large metropolitan area levels. Therefore, it can be used to understand the demographic nature of employment at these broad geographic levels. Since the data are published on a monthly basis, the CPS provides the most current employment estimates. The BLS uses CPS data to estimate monthly unemployment statistics. Since the data contain demographic information on the labor force, it is useful for analyzing employment trends of different demographic groups.

The data provide all estimates based on residence location. They do not include any information with respect to place of work.

4.2 Current Employment Statistics

The CES program of the BLS is a voluntary employer establishment-based payroll survey, conducted monthly. The CES provides estimates of nonfarm employment, hours, and earnings at the national, State, and metropolitan area levels. Typically, the monthly estimates are released on the third Friday of each month.

Data Coverage

The CES survey universe includes all nonagricultural private businesses and government agencies at the Federal, State, and local levels. The CES employment definition covers all persons on the employers' payroll during the reference week of the survey, including full-time, part-time, temporary, and intermittent workers. It excludes volunteer, farm, domestic, and unpaid workers. A worker who works for several employers may be counted multiple times by the CES. Uniformed members of the Armed Forces and self-employed workers are excluded from the CES.

Data Collection Methodology

The CES survey collects data on employment, earnings, and hours from approximately 146,000 businesses and government agencies representing approximately 623,000 individual establishments or worksites. The sample worksites are drawn from a universe of unemployment insurance tax accounts in each State, covering roughly 9 million establishments. The CES samples are extracted from all 50 States and the District of Columbia. The entire sample is redrawn annually, and a supplemental sample of new business births is selected midway through each sample year. The sample selection weight is approximately the inverse of probability of selecting the sample from UI accounts.

Each month, BLS collects data on employment payroll and paid hours from the sampled establishments. The data are collected by four regional data collection centers through different data collection methods. The methods include Computer Assisted Telephone Interviewing, Touchtone Data Entry, fax, web, and

Electronic Data Interchange. Offering survey respondents a choice of reporting methods helps to sustain response rates for this voluntary survey.

The survey form asks questions on total employee count, women employee count, payroll, commission, and hours. The same questions are repeated for each of the 12 months of a year.

The estimation method considers the facts of opening and closing of firms. CES uses automated editing and screening techniques to identify potentially erroneous sample data. The respondents are recontacted to validate or correct the reported information. A seasonal adjustment is applied to CES data.

Data Geography

The CES data are available at the national, State, and metropolitan area levels.

Data Content

The CES provides the following data items stratified by geographic area and by NAICS codes:

- All employees.
- Production or nonsupervisory employees (depending on industry).
- Women employees.
- Average weekly hours.
- Average hourly earnings (constant dollar and current dollar).
- Average weekly earnings.
- Average overtime hours in manufacturing.
- Indexes of aggregate hours and payrolls.
- Diffusion index.

The level of industry detail varies across the national, State, and major metropolitan area levels, depending on the adequacy of the sample size. All data are available without seasonal adjustment; for some industries, seasonally adjusted data also are available.

Data Acquisition and Sources

Current and historical public use CES data can be downloaded from the BLS Web site:

<http://www.bls.gov/ces/data.htm>

Due to data confidentiality, the BLS does not provide public use microdata. Researchers can get access to BLS microdata under certain circumstances. Information about qualifying for the program and the application process through which access may be granted can be found on the BLS Web site:

<http://www.bls.gov/bls/blsresda.htm>.

Data Applications

The CES surveys employer establishments. Therefore, the CES estimates are based on workplace geography. The data are useful to understand monthly changes in nonfarm payroll employment, average weekly earnings, and average weekly work hours by industry.

The CPS data also provide monthly statistics on employment. However, there are important differences in two datasets. CPS estimates are based on surveys of workers at their place of residence (where they live); whereas, CES estimates are based on surveys of employers at their workplace locations.

4.3 Nonemployer Statistics

Nonemployer Statistics is an annual series published by the Census Bureau that provides economic data for businesses that have no paid employees and are subject to Federal income tax. Most nonemployers are self-employed individuals operating unincorporated businesses (known as sole proprietorships), which may or may not be the owner's principal source of income.

Data Coverage

Nonemployer Statistics includes all businesses that have no paid employees, but are subject to Federal income taxes, and have revenues of at least \$1,000.

Data Collection Methodology

The primary source for Nonemployer Statistics is the annual or quarterly income tax returns filed by self-employed individuals with the IRS and maintained in the Census Bureau's Business Register. The data are processed through various automated and analytical review to eliminate employers from the tabulation, correct and complete data items, remove anomalies, and validate geography coding and industry classification. Prior to publication, noise infusion is applied to protect individual businesses from disclosure. Data are published approximately 18 months after each reference year.

Data Geography

Nonemployer Statistics data are available at the national, State, county, and, metropolitan area levels of geography.

Data Content

Nonemployer Statistics publishes separate records for each level of geography and industry (NAICS) code. Each record includes the following data items:

- Total number of establishments.
- Total annual receipts (\$000).

If the number of establishments associated with a specific industry code falls below the minimum threshold deemed necessary to protect individual business from disclosure, both data items are suppressed.

Data Acquisition and Sources

Nonemployer Statistics can be downloaded from the Census Bureau Web site:

<http://www.census.gov/econ/nonemployer/index.html>

Data are currently available for the years 1997 to 2014.

Data Application

Nonemployer Statistics data provide a supplement to other Census and BLS employer-based data products (including CES, QCEW, and LEHD) that are based on UI records. When combined together, the two datasets provide a more complete summary of total employment within a specified geographic area.

4.4 Quarterly Census of Employment and Wages

The QCEW Program (formerly known as the ES-202 Program) is a cooperative program between the BLS and SESA. The program utilizes employment and wage data that are reported quarterly by all employers who are subject to UI laws. Data products include summary tabulations of the number of establishments, number of employees, and total and average wages paid. The data are available approximately five to six months after the end of each quarter.

Data Coverage

The QCEW program produces employment and wage information for workers covered by State UI laws and for Federal workers covered by the Unemployment Compensation for Federal Employees (UCFE) program for all 50 States, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. Workers covered by UI laws represent about 97 percent of all civilian waged/salaried workers in the country. Workers who are not subject to UI or UCFE programs include self-employed workers, agricultural workers on small farms, uniformed members of the Armed Forces, elected officials in most States, most railroad workers, some domestic workers, most student workers at schools, and employees of foreign governments and certain small nonprofit organizations. A complete listing of worker categories that are exempt from employer UI contributions can be found in *26 USC, Subtitle C, Chapter 23, § 3306 (c)*.

Data Collection Methodology

Employers subject to State UI laws submit quarterly contribution reports, quarterly wage reports, and multiple worksite reports to SESAs. The **quarterly contribution report** includes the employer's unique State employer identification number (EIN), the number of employees for each month of the reference quarter, and the total wages paid to all employees during each month of the reference quarter. The **quarterly wage report** includes each employee's name, social security number (SSN), and total wages paid to that employee during the quarter. The **multiple worksite report** provides information on the number of employees and total wages paid at each worksite of a multiworksite employer within the State. Multiple worksite reports provide information on the distribution of employment and wages at a geographic level finer than that of a State. Each employer is assigned a six-digit NAICS code based on the reported description of its industrial activity, and is grouped into one of four ownership categories—private, Federal, State, or local.

The SESAs collect and compile the quarterly reports from employers within their State. They then send those reports to BLS within 15 weeks after the end each quarter. BLS checks the data for accuracy and compiles the data to develop the QCEW data tables.

Data Geography

The QCEW reports data by workplace location. Public use data are available at the county, metropolitan area, State, and national levels. In many States, qualified government agencies and researchers may obtain QCEW data at finer levels of geographic resolution by signing confidentiality agreements with their SESAs. However, the terms and conditions of such agreements vary from State to State.

Data Content

The QCEW data tables provide the following information on a quarterly basis by industry, ownership type, and establishment size:

- Number of employees.
- Number of establishments.
- Total wages.
- Average weekly wage.
- Average annual pay.

Employers are categorized by 2012 NAICS coding system. To maintain the employers' confidentiality, county-level data are not always available by six-digit NAICS code. The ownership types include Federal, State and local government, and private.

Data Acquisition and Sources

Current and historical public use QCEW data can be downloaded from the BLS Web site:

<http://www.bls.gov/cew/data.htm>

The data can be downloaded from this site in both comma-separated value (CSV) and Microsoft Excel (.xlsx) formats. Requests for more detailed data should be made directly to State agencies.

Data Application

QCEW data are widely used by Federal statistical agencies, BLS surveys, and other public and private establishments as a basis for their statistics and research publications. QCEW data are used by businesses and by public and private research organizations for economic forecasting and regional employment trend analysis. QCEW data can be used to visualize the temporal distribution of employment by industry at different geographic levels. Since the data tables are developed based on complete *administrative records*, the data accuracy is higher than with survey data sources, which are based on sample populations.

4.5 County Business Patterns

CBP is an annual series that provides business establishment and selected employment data at a subnational level by industry. The data include the number of establishments, total employment during the week of March 12, first quarter payroll, and annual payroll. CBP provides annual data on businesses with paid employees for all States; the District of Columbia; Puerto Rico; and other U.S. island territories (Guam, American Samoa, the Commonwealth of the Northern Mariana Islands, and the U.S. Virgin Islands) at a more detailed level of geography and industry than other Federal employment establishment-based data sources. CBP data are available within 18 months following the end of each reference year.

Data Coverage

CBP covers most, but not all, employment categories. It excludes data on self-employed individuals, domestic employees, railroad employees, agricultural workers, and most government employees. With the exception of government employees, the excluded employment categories are similar to those excluded in other Federal government establishment databases that utilize UI administrative records. Additionally, businesses operating without an *employer identification number* and businesses with an EIN, but with no

paid employees, are excluded from the CBP universe. Business establishments in the CBP are categorized by their six-digit 2012 NAICS code.

Data Collection Methodology

The Census Bureau's Business Register (BR) is the primary data source for identifying and locating business establishments in the CBP. The BR is a multirelational database that contains a record for each known business establishment with paid employees located in the United States, Puerto Rico, and other U.S. island territories. It includes geographic location; organization type; industry classification; and operating data (e.g., payroll and employment) for each included establishment. The BR contains more than 160,000 multi-establishment businesses, representing 1.8 million affiliated establishments, 5 million single establishment companies, and nearly 21 million nonemployer businesses. Information contained in the BR is confidential; access to these database is restricted to Census employees sworn to uphold the confidentiality provisions of Title 13 and Title 26 of the U.S. Code.

Establishment data in the BR are updated continuously from multiple sources, including Census surveys of establishments such as the CES, administrative records program such as the QCEW, income tax records from the IRS, and the five-year Economic Census. Data updates are linked to individual business establishments through the EIN.

Data contained in the BR are not subject to sampling error, but are subject to various nonsampling errors, including inability to identify all cases that should be in the universe; definition and classification difficulties; errors in recording or coding the data obtained; and other errors of coverage, processing, and estimation for missing or misreported data.

Data Geography

Statistics are available on business establishments at the national, State, county, metropolitan area, ZIP code, and Congressional District levels. Data for Puerto Rico and the Island territories are available at the State and county equivalent levels.

Employers without a fixed location within a State (or of unknown county location) are included under a general "statewide" classification at the end of the county tables. This incomplete detail causes only a slight undercount of county-level employment.

Data Content

CBP data are tabulated by geographic area, six-digit NAICS industry, *legal form of organization* (U.S. and State only), and employment size class (nine categories). The data tables contain the following formation:

- Number of establishments.
- Number of paid employees during the pay period that includes March 12.
- First quarter payroll.
- Annual payroll.

Data Acquisition and Sources

Current and historic CBP datasets can be downloaded from the Census Web site:

<http://www.census.gov/programs-surveys/cbp/data.html>

The CBP data are available in CSV format, along with record layouts and reference guides, from 1986 to present. CBP data tables from 2004 to present also are available through the Census Bureau's American Factfinder Web site:

<http://factfinder.census.gov>.

Data Application

CBP statistics provide the only annual source of complete and consistent county-level employment data for U.S., Puerto Rico, and Island Areas business establishments with industry detail. Therefore, the data are useful to analyze the employment activities at the county level. CBP allows planners to understand the changes in economic activity pattern over time. The data also are useful for determining annual changes in employment.

4.6 American Community Survey

The ACS is a continuous survey that provides vital demographic and economic information about the U.S. population on an annual basis. It replaces the Census long form questionnaire that was conducted as part of the decennial Census through 2000. The ACS collects demographic, employment, income, commuting, housing, and other *social data* from a sample of U.S. households. The survey is conducted by the Census Bureau throughout the United States and Puerto Rico. Estimates of population characteristics, developed from the survey, are summarized over one-year and five-year time periods. ACS three-year summaries were published from 2007 to 2013, but have been discontinued. The Census Bureau publishes both one-year and five-year summary databases on an annual basis. The estimates are available at different geographic levels, depending on the summary time period.

Data Coverage

The ACS draws samples from *housing units* and noninstitutional group quarters in each county or county equivalent of United States and Puerto Rico. The survey includes both civilians and uniformed members of the Armed Forces. It includes both waged/salaried and self-employed workers. Because the sample unit is the *household*, rather than an employment establishment, all industry types are eligible for inclusion.

Household members include everyone who is living in the sampled housing unit for more than two months on the day of the interview. Children who move between houses under a joint custody agreement are considered to be living in the sampled housing unit where they are residing on the day of the interview. Workers who stay at a location closer to work, but return regularly to the sampled housing unit to be with their families, also are considered to be living in the family residence. The group quarter (GQ) sample universe includes all people residing in the GQ facility at the time of interview. Data are collected for all people sampled, regardless of their length of stay. Children (not attending college) staying at a GQ facility functioning as a summer camp are not considered GQ residents.

Data Collection Methodology

The ACS consists of two separate samples: 1) housing units (HU) and 2) residents of noninstitutional GQ facilities. These samples are drawn from the Census Bureau's Master Address File (MAF). The MAF is the Census Bureau's official inventory of the addresses of housing units and GQs in the United States and Puerto Rico. The survey samples are selected from each county of the United States. In 2013, 3.55 million addresses were surveyed. The sample data are collected continuously throughout the year and are combined to produce annual estimates. More detailed information on data collection methodology is available at the Census Bureau's Web site for the ACS.

Data Geography

The Census Bureau develops estimates of household characteristics based on one-year and five-year accumulations of ACS surveys, and publishes the data for different levels of geographic detail, depending on the population within those geographic units. For geographic units with populations of at least 65,000, which includes all States, most metropolitan areas, and some large counties and places, the Census Bureau publishes one-year estimates based on data collected during the preceding 12 months of the most recent calendar year.

For most other statistical, legal, and administrative geographic units, including census tracts, census block groups, and smaller counties and incorporated places, such as cities and towns, the Census Bureau publishes five-year estimates based on data collected during the preceding 60 months of the most recent five calendar years.

Data Content

The ACS provides estimates of the following characteristics:

- **Demographics**—Total population, group quarter population, age, sex, race, ethnicity, etc.
- **Housing**—Occupancy, vehicle availability, value of home, Internet access, etc.
- **Economic**—Class of worker, income, journey to work, industry, occupation, employment status, etc.
- **Social**—School enrollment, educational attainment, place of birth, citizenship, etc.

A complete list of all ACS tables and files can be downloaded from the ACS Web site.

Data Acquisition and Sources

The Census Bureau's American Fact Finder (AFF) is the primary resource for accessing and downloading the ACS data tables. It allows users to select data tables for different geographic levels, household demographic characteristics, industry types, and survey year and product type. The link to the AFF Web site is:

<http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t>.

DataFerrett is another option for downloading the data tables. This option tabulates data at the Region, Division, and State levels. ACS data are not available at finer geographic levels from DataFerrett. The link to the DataFerrett Web site is:

http://thedataweb.rm.census.gov/Products_Services/DataFerrett.html.

4.7 Census Transportation Planning Products

The **CTPP** is a set of special tabulations of the ACS Journey-to-Work questions designed by transportation planners specifically for transportation planning applications. The CTPP was jointly funded by the American Association of State Highway and Transportation Officials (AASHTO) and State DOTs through a pooled fund study administered by the Federal Highway Administration (FHWA).

The most current CTPP data tables are based on the 2006 to 2010 (five-year) ACS data. Although additional ACS surveys have been conducted since 2010, the CTPP special tabulations do not incorporate data from these more recent surveys. The CTPP tables are organized into three geographic categories:

- **Part 1: Residence**—summarizing worker and household characteristics.
- **Part 2: Workplace**—summarizing worker characteristics at the workplace.

- **Part 3: Flows between residence and workplace**—including trip characteristics such as means of travel and travel time.

The CTPP includes nearly 200 residence-based tables, 115 workplace tables, and 39 flow tables for different levels of geographic resolution ranging from States to metropolitan areas, counties, places, census tracts, and traffic analysis zones (TAZ), where available. Users can select and download specific tables and geographic levels of interest from the AASHTO CTPP Web site:

<http://ctpp.transportation.org/Pages/5-Year-Data.aspx>.

Public Use Microdata Sample

The ACS **Public Use Microdata Sample** (PUMS) files are a set of untabulated records about individual people or housing units. The Census Bureau produces the PUMS files so that data users can create custom tables that are not available through pretabulated (or summary) ACS data products.

PUMS data consists of individual records of households responding to the ACS, but with all personal identifying information removed, including residence location. Instead, detailed residence location for each household record is replaced by geographic codes representing the surveyed household's State and Public Use Microdata Area (PUMA). A PUMA is a geographic area containing at least 100,000 people; that is an aggregation of census tracts and/or counties; and nests within a single State and/or metropolitan area.

PUMS data are available for both one-year and five-year ACS data summaries, and can be downloaded from the Census Bureau's ACS Web site:

<https://www.census.gov/programs-surveys/acs/data/pums.html>.

Data Application

ACS data have been extensively used by transportation planners to understand household demographics and employment at different geographic levels. The ACS provides information based on both place of residence and place of work. The ACS also provides journey-to-work data, which helps transportation planners to understand commuting patterns. The data are useful in calibration and validation of travel demand models. Since the ACS provides annual estimates, it also is useful in the trend analysis of employment.

4.8 Longitudinal Employment Household Dynamics

LEHD is a program developed and administered by the U.S. Census Bureau that combines administrative records collected by the Census Bureau, BLS, and other Federal and State agencies to produce a unique system of data on employer-worker interactions that is not available from other sources.

The LEHD produces and disseminates two principal data products: 1) the QWI and the LEHD Origin-Destination Employment Statistics. The QWI is a set of economic indicators that provide information on employment, job creation, wages, and worker turnover at various levels of geography (e.g., county, State, metropolitan area); by industry and ownership type; and by specific worker characteristics (e.g., gender and age). The LODES provides information on the geographic locations of employer worksites, worker residences, and residence to worksite flows, as well as key employer and worker characteristics (e.g., industry type, firm size, worker age, gender, income level).

Data Coverage

One source of input data for the LEHD is the quarterly reports submitted by employers subject to UI laws to SESA. These are the same reports that BLS uses to create the QCEW; consequently, the LEHD represents the same universe of workers as previously described for the QCEW (e.g., approximately 97 percent of all civilian waged/salaried workers in the U.S.).

The Census Bureau works with each State SESA through its Local Employment Dynamics (LED) Partnership to obtain State employment data directly from the States. Through this LED Partnership, it has been able to publish employer-related data covering the period from 2002 to 2014 for most States. Some States are missing workplace data for specific earlier years; workplace data are not currently available for Puerto Rico or the U.S. Virgin Islands.

Data Collection Methodology

Unlike most other employment databases, which survey or compile administrative records on either employers or on workers, the LEHD Program integrates data covering both employers and individual workers. The key to linking employers and workers is the quarterly wage reports submitted by employers to their SESAs. Each quarterly wage report includes a unique identifier for each worker—the employee’s SSN, and a unique identifier for the employer—the EIN. These two identifiers allow the Census Bureau to match workers to employers, and provide links to additional employer and worker characteristics contained in other data sources.

Employer characteristics, including the physical locations of primary and secondary employer worksites, industry type, number of employees, and wages paid by quarter, come from the quarterly contribution and wage reports, and from the employer’s UI application; these reports are matched to individual employers through the EIN.

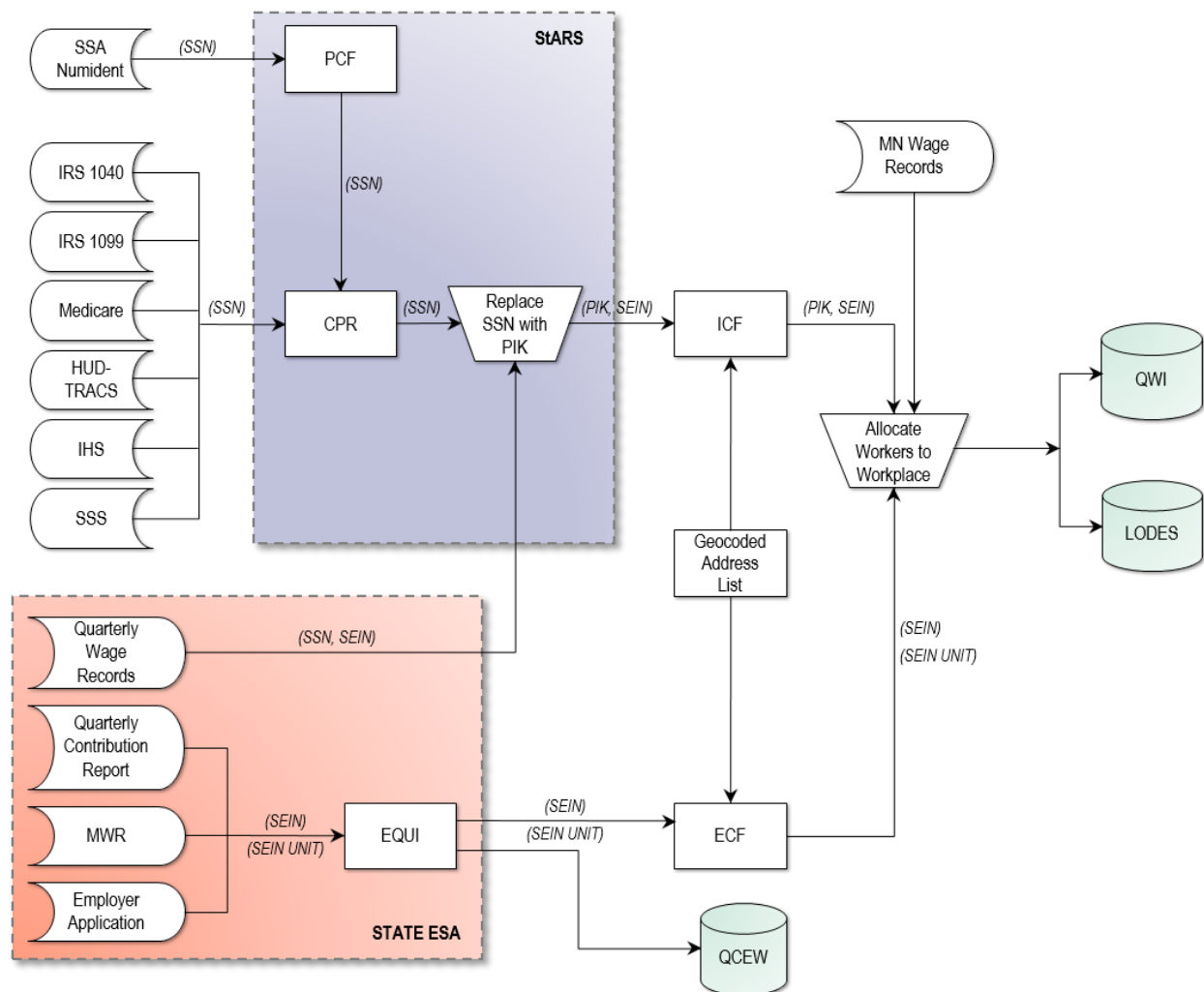
Characteristics of individual workers, including residence location, age, gender, income level, etc., come from the Statistical Administrative Records System (StARS). StARS is a centralized data repository, developed and maintained by the Census Bureau to support its various demographic and economic statistical programs. The Census Bureau compiles administrative records databases collected by other Federal and State agencies, including the IRS, Social Security Administration (SSA), and Medicare. Information from these databases can be matched to individual workers through their SSN. Under the StARS Program, Census Bureau staff replace each SSN with a protected identification key (PIK); remove personal identifiers (e.g., name); and create aggregate demographic records that can be used to improve the quality of its various statistical reports and databases.

Figure 4-1 presents a flow chart that summarizes the processing steps for the LEHD databases. First, the Census Bureau uses StARS to merge demographic data on individual citizens from administrative records data collected by other agencies. This Composite Person Record (CPR) database is then matched to the quarterly wage records to produce an Individual Characteristics File that includes demographic information and the home address for each worker, together with the employer’s EIN. A separate Employer Characteristics File (ECF) is created from the Enhanced Quarterly Unemployment Insurance (EQUI) reports collected by each State that contains key employer characteristics, the address for the employer’s primary workplace address, and each secondary workplace address, where available. The Census Bureau then uses its Master Address List (MAL) to geocode each address in the ICF and ECF databases to latitude/longitude coordinates, as well as to various levels of census geography (census block, tract, urbanized area, and county).

A key processing step in creating the LODS database is the linking of individual workers (and their home locations) to employer workplace locations. For those employers that have only a single workplace location within a State, workers can be linked directly to the workplace through the quarterly wage reports.

However, in all but one State (Minnesota), employers with multiple workplace establishments are not required to file separate quarterly wage reports for each workplace location. Consequently, there is no way to directly link workers of multi-workplace employers to a specific workplace establishment.

Figure 4-1. LEHD processing steps.



Source: U.S. Census Bureau, Center for Economic Studies

Although multi-workplace employers represent less than 3 percent of all employers, nationwide, they employ between 30 and 60 percent of all workers within a State. Therefore, the allocation of individual workers to specific workplace locations of multi-workplace employers represents a potentially significant source or error in using LODS data to estimate work trip flows between a worker's residence and workplace.

Using the Minnesota wage report data, the Census Bureau has calibrated a probability model of workers' home-to-work distance based on several variables, including length of employment at the firm, firm size and age, and number of secondary workplace locations. This model is used to impute the most likely workplace establishment for each worker at a multi-workplace employer.

To protect the identity and confidentiality of individual workers and employers, the Census Bureau applies a series of processing steps to "disclosure proof" the data. Three distinct processes are used. The

first step introduces a small amount of “noise” into the ECF at the establishment level (e.g., total number of employees, total wages paid) to prevent identification of individual employer characteristics.

The second step is introduced when worker residence and workplace locations are joined to create origin-destination work trip flows. For each census block containing one or more worker residences, the distributions of four worker or employer variables (i.e., industry code, employer ownership, worker age, monthly earnings) are synthesized by the Census Bureau in a way that prevents identification of any individual worker or employer. The perturbation process is done in such a way that, as the number of workers in a residence block increases, or at higher levels of Census geography, this synthetic distribution approaches the actual distribution of worker/employer characteristics.

The third disclosure proofing step is suppression of data for those residence or workplace census blocks that have fewer than three workers or employers. However, suppression is only used when the combination of noise and synthetic distributions still enable an individual worker or employer to be identified.

A more complete technical discussion of the LEHD database process can be found in the report, *LEHD Infrastructure Files in the Census RDC—Overview*. For more information on the disclosure proofing process used in the LEHD, see *Confidentiality Protection in the Census Bureau’s Quarterly Workforce Indicators, Technical Paper TP-2006-02*.

Data Geography

QWI data are reported based on detailed firm characteristics (geography, industry, age, and size) and worker demographic information (sex, age, education, race, and ethnicity); and are tabulated at the national, State, metropolitan area, county, and Workforce Investment Board (WIB) levels of geography.

LODES data files are available at the census block level of geographic detail. Census blocks can be easily aggregated to all other census geographies, including census tract, urbanized area, metropolitan area, county, and State. In many areas, census blocks also may be aggregated to TAZs for use in travel forecasting models. Due to the disclosure proofing procedures described above, employer and worker characteristics are not reliable at levels of geography smaller than a census tract or TAZ.

4.9 Quarterly Workforce Indicators

QWI data files include up to 32 indicators of workforce participation, turnover, wages, etc., which are reported on a quarterly basis for specific levels of geography and stratifications of worker and employer characteristics. More information on the data content of the QWI (Public Release) database, including data schema, record formats, and attribute definitions, can be found in the documentation, *Quarterly Workforce Indicators 101*.

These indicators can be downloaded by State as CSV files for various combinations of geography, worker demographics, and firm characteristics. The following are variables upon which QWI data may be stratified:

- **Geographic level**—County, metropolitan area, entire state, and workforce investment area.
- **Industry code**—All industries, NAICS sector, subsector, and industry group.
- **Firm ownership**—All ownership types and private ownership.
- **Worker gender**—All, male only, and female only.
- **Worker age**—All ages, 14 to 18, 19 to 21, 22 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 to 99.
- **Worker race**—All, White, Black, American Native, Asian, Pacific Islander, and multiple races.
- **Worker ethnicity**—All, Not Hispanic, and Hispanic or Latino.

- **Worker education**—All, less than high school, high school only, some college, and bachelor's or advanced degree.
- **Firm age**—All, 0 to 1 year, 2 to 3, 4 to 5, 6 to 10, and 11+.
- **Firm Size**—All, 0 to 19 employees, 20 to 49, 50 to 249, 250 to 499, and 500+.

The number of indicators and the time periods for which QWI data are available varies considerably by state. For some states, selected indicators are available by quarter as far back as the early 1990s.

LEHD Origin-Destination Employment Statistics

LODES data includes characteristics of workers and employers summarized for various levels of geography that define worker residence locations, workplace locations, and flows of workers between their residence and workplace location. Like the QWI, LODES data are organized by State. Within each State folder, the data are organized into three primary datasets:

1. **Residence Area Characteristics (RAC)**—The total number and worker/employer stratification of jobs summarized by the worker's place of residence. Because the unit of tabulation is jobs rather than workers, if a worker holds two or more jobs, (s)he will be counted for each job held. This could lead to situations where the total number of jobs exceeds the number of individual workers in a specified area. Each record in the database represents a specific Census Block. Each record includes the following data items:
 - a. Total number of jobs by workers residing in the Census Block.
 - b. Number of jobs stratified by worker age (under 29, 30 to 54, and 55+).
 - c. Number of jobs stratified by monthly earnings (under \$1,250, \$1,251 to \$3,333, and over \$3,333).
 - d. Number of jobs stratified by NAICS two-digit sector code (20 categories).
 - e. Number of jobs stratified by race (six categories).
 - f. Number of jobs stratified by ethnicity (two categories).
 - g. Number of jobs stratified by education level (four categories).
 - h. Number of jobs stratified by gender (two categories).
2. **Workplace Area Characteristics (WAC)**—The total number and worker/employer stratification of jobs summarized by place of work. Each record in the database represents a specific Census Block. Each record includes all of the data items in the RAC, plus the following additional data items:
 - a. Number of jobs stratified by age of firm (five categories).
 - b. Number of jobs stratified by size of firm (five categories).
3. **Origin-Destination (OD)**—The total number and worker/employer stratification of jobs whose workers reside in one Census Block and work in another Census Block (including the same Census Block). Each record in the database represents a unique pair of worker residence/workplace geographies. Each record includes the following reduced set of data items compared to the RAC and WAC:
 - a. Total number of jobs by workers residing in the origin Census Block whose workplace is located in the destination Census Block.
 - b. Number of jobs stratified by worker age (under 29, 30 to 54, and 55+).
 - c. Number of jobs stratified by monthly earnings (under \$1,250, \$1,251 to \$3,333, and over \$3,333).
 - d. Number of jobs stratified by general industry group (goods production; transportation, trade, and utilities; and all other sectors).

Within each dataset, LODES data files are further stratified by year and by job type (i.e., all jobs, primary jobs, private-sector jobs, and Federal jobs).

Data Acquisition and Sources

Both the QWI and LODES database files can be downloaded from the Census Bureau's LEHD Web site:
<http://lehd.ces.census.gov/data/>.

The Census Bureau also provides a web-based tool, [QWI Explorer](#), for displaying and graphing selected subsets of the QWI data, and a web-based map viewer, [OnTheMap](#), for displaying and conducting spatial-based analyses of the LODES data. More information on the data content of the LODES database, including data schema, record formats, and attribute definitions, can be found in *the LODES (Version 7.2) Technical Documentation*.

Data Application

The LODES database, in particular, has strong potential applications in transportation planning, either as a supplement or alternative to more traditional sources (e.g., large-scale household interview survey and CTPP) for determining the geographic distribution of home-to-work trips for use in metropolitan or statewide travel models.

4.10 InfoUSA

Founded in 1972, InfoUSA is a commercial provider of business and consumer information products, database marketing services, data processing services and sales and marketing solutions. InfoUSA maintains a proprietary database of 250 million consumers and 14 million businesses. The business database includes information on industry type, number of employees, and business location. Industries are classified by industry name, SIC code, and NAICS code. InfoUSA also provides data in GIS format by geocoding the business addresses.

Data Coverage

InfoUSA data covers businesses and public agencies in all 50 States and Washington, D.C. The data includes all industry types.

Data Collection Methodology

The database is continuously updated from more than 5,000 public sources, and more than 24 million phone calls are made per year to verify and collect additional information. In developing the database, 5,200 phone books, annual reports, and other business directories are continuously reviewed to find information. This includes public records data from county courthouse filings, Securities and Exchange Commission 10k filings, and Secretary of State data are used; and every month, the U.S. Postal Service National Change of Address, ZIP+4 and Delivery Sequence file are matched to standardize and keep addresses current.

Data Geography

Summary data are available by State, county, ZIP code, and city. Individual business locations can be obtained as GIS point features based on geocoded addresses.

Data Content

The InfoUSA business database includes the following information:

- Company name.
- Address of establishment location.
- Industry (SIC/NAICS code).
- Metropolitan area.
- Carrier route.
- Delivery point bar code.
- Phone number.
- Web site address.
- Number of employees.
- Sales volume.
- Credit rating score.
- Office size.
- Square footage.
- Headquarter/branch.
- Public/private.

Data Acquisition and Sources

Data acquisition information is available at the InfoUSA Web site:

<https://www.infousa.com>.

Cost of the data varies depending on the amount of data requested, which is influenced by both the coverage area and included attributes. Specific data tables can be customized and ordered from the Web site. The customization tool allows the extraction of data by geography, by industry, and by business size.

Data Application

InfoUSA provides the most current data on businesses. The data are available by industry type and size of business. This information can be used to determine the number of employees by industry within a geographic area. The data points or business locations in GIS format can be used to aggregate the data to any geographic level.

4.11 Dun & Bradstreet

Founded in 1841, Dun & Bradstreet, Inc. (D&B) is a business services company that provides commercial data and analysis services to businesses on credit history, business-to-business sales and marketing, risk exposure, and supply chain management. In 1963, D&B introduced the DUNS, which is a unique nine-digit numeric identifier assigned to each business location in the D&B master business database. Although not mandatory, DUNS has become the de facto worldwide standard for establishing a business credit file, which is used by lenders and potential business partners to help predict the financial stability of a company. The United States Government, the European Commission, and the United Nations require all grant applicants and contractors to have a DUNS number; and more than 50 global industry and trade associations recognize, recommend, or require that their members obtain a DUNS number. D&B is solely responsible for issuing DUNS numbers to businesses, and maintains the basic registration information about each business in its master database.

Data Coverage

The D&B master database contains information on each business establishment that has been issued a DUNS number. This includes more than 265 million business locations worldwide. Both private-sector

business and public agencies are included, but businesses or public agencies that do not have a DUNS number are excluded.

Data Collection Methodology

Basic registration data are obtained from each business that requests a DUNS number. These initial data include company name, mailing address, physical address for each business establishment location, contact information, and number of employees at each location. After a business is entered into the master database, D&B compiles additional data on each business using more than 30,000 public and proprietary data sources, including credit rating agencies, sales and trade transactions, banking data, corporate relationships and family trees, social media, economic census, etc. Source data are collected, edited, and verified using internal, proprietary methods; additional data items are derived from the source data (e.g., address locations are geocoded to produce latitude/longitude coordinates for each business establishment). D&B estimates that approximately 5 million updates are made per day to the master database. Updates are published on a monthly basis, at a minimum.

Data Geography

Authorized data users can select and download records of individual business establishments; or create summaries for selected standard geographies (e.g., counties, ZIP codes, metropolitan areas, and States). Inclusion of latitude/longitude coordinates enables users to create GIS point features for each business establishment.

Data Content

D&B's Hoover's subscription service provides authorized users access to more than 70 attributes business establishment in the master database. Attributes include the following:

- Company name (legal and "doing business as").
- Address of establishment location.
- Latitude/longitude coordinates of establishment address.
- DUNS number.
- Industry (NAICS code).
- Metropolitan area.
- Phone number.
- Primary contact name and title.
- Web site address.
- Number of employees.
- Sales volume.
- Credit rating score.
- Office size (square foot).
- Headquarter/branch.

Data Acquisition and Sources

D&B offers multiple data products and services, but the most relevant database for transportation planning applications is its Hoover's marketing and sales subscription service. Subscribers pay an annual fee to access the D&B master database through a web-based application. The application enables authorized users to select subsets of the more than 265 million data records based on various criteria, including geography, industry type (e.g., NAICS code); number of employees; etc. The selected subset of records can then be downloaded to the user's own computer for further analysis.

The cost of the subscription varies depending on the number of authorized named users and the total number of records that can be downloaded annually. Additional information about the subscription service is available at the D&B Web site:

<http://www.dnb.com/products/marketing-sales/hoovers-suite/hoovers.html>.

Data Application

The D&B database provides the most current data on business establishments and public agencies that have requested and obtained a DUNS number. The data are available as individual establishment records that can be aggregated to any level of geography, industry type, or size of business. This information can be used to determine the number of employees by industry within a geographic area.

4.12 YourEconomy Time Series

YourEconomy Time Series (YTS) is a new time series database developed by the BDRC, which is a public-private, nonprofit research institute, affiliated with the University of Wisconsin-Extension, and devoted to the study of business performance and economic growth. YTS uses annual snapshots from the InfoUSA business establishment data as its primary data source, but combines the annual data into a consolidated, multiyear database; verifies and corrects attributes using other data sources; and appends additional attribute fields, such as Congressional District, identification of new businesses, or summary measures based on geographic or temporal aggregations.

Data Coverage

YTS data includes all businesses and public agencies contained in the annual snapshots from the InfoUSA database.

Data Collection Methodology

The InfoUSA database is continuously updated from more than 5,000 public sources; and more than 24 million phone calls are made per year to verify and collect additional information. BDRC acquires annual snapshots of InfoUSA from 1997 to 2015, and merges these into a single master database. It then identifies and replaces missing data items for a small number of records, appends additional variables based on other sources, and generates summary variables based on geographic or temporal aggregations.

Data Geography

Summary data are available by State, county, ZIP code, and (combined) metropolitan area.

Data Content

The YTS database includes the following information:

- Employment.
- Number of establishments.
- Total sales.

Users may request this information for any combination of years from 1997 to 2015, for any available geographical region, and for any combination of two- to six-digit NAICS code.

Data Acquisition

YTS data can be purchased directly from BDRC. Additional information regarding database pricing and ordering procedures can be found on the BDRC Web site.

<http://exceptionalgrowth.org/our-databases.iegcc>.

Data Application

Historical time-series data on business establishments are useful for analyzing trends in economic growth for a region or subregion, identifying areas where commuter travel volumes may be changing, and identifying areas where targeted land use policies or investments in transportation infrastructure may be most effective.

4.13 Woods & Poole

Woods & Poole's database contains employment data, along with demographic information for every State, region, county, and metropolitan area in the U.S. for every year from 1970 to 2050. Woods & Poole county projections are updated annually and utilize projection models that take into account specific local conditions based on historical data.

Data Coverage

The database includes historical and forecasts of employment and demographic information for both civilian and noncivilian population of the United States from year 1970 to year 2050.

Data Collection Methodology

Historical population, demographics, and employment data are extracted from various Census databases, including the decennial Census, ACS, five-year Economic Census, and CBP. Current and future data are forecast from the historical data using proprietary projection models developed by Woods & Poole. The forecasts for each county in the United States are done simultaneously, so that changes in one county will affect growth or decline in other counties. For example, growth in employment and population in Houston will affect growth in other metropolitan areas, such as Dallas. This reflects the flow of economic activity around the country as new industries emerge or relocate in growing areas, and as people migrate in response to job opportunities. County projections are developed within the framework of the United States projection made by Woods & Poole.

Data Geography

The data are available at the county, metropolitan area, State, and regional levels. Counties, county-equivalents, and regions used in the Woods & Poole database are defined by the Bureau of Economic Analysis (BEA). Metropolitan areas in the database are those defined by the Office of Management and Budget (OMB).

Data Content

The primary employment data included in Woods & Poole database is number of jobs. It includes jobs from all employment types, including the following:

- Private and government.
- Farm and nonfarm.

Jobs are summarized by industry name, but not by specific SIC or NAICS code. Additional data available in the Woods & Poole database include the following:

- Population by age, race, and sex.
- Per capita income.
- Sales.
- Number of households.
- Household size and income.

Data Acquisition

Standard Woods & Poole databases can be purchased directly from their Web site:

<https://www.woodsandpoole.com/>.

Technical documentation and pricing information also are available on the Web site.

Data Application

The data are useful for understanding changes in employment and demographics over time. Since the data are projected to year 2050, they can provide key forecasting variables for long-range transportation planning.

4.14 Moody's Analytics

Moody's Analytics provides extensive historical and forecast data for the United States and Canada, including key economic variables, such as employment, gross product, wages by industry, labor force and unemployment rate, population, personal and household income, housing stock and prices, and retail sales. Forecasts include annual changes out to a 30-year time horizon; and are updated on a monthly basis to reflect the most current economic data, conditions, and expectations.

Data Coverage

Moody's Analytics offers several standard databases based on level of geography (counties, metropolitan areas, and States); industry detail; and whether or not historical data are included. Other specialized

database also are available, including historical trends and forecasts of housing supply and home prices at the metropolitan area level.

Data Collection Methodology

Historical population, demographics, and economic data are extracted from a variety of public and private sources, including the BEA, BLS, Census, Federal Reserve Bank, National Association of Realtors, and Dow Jones. The data are reconciled and summarized into the specified geographic area categories. Forecasts are developed using proprietary forecast models, and are updated monthly based on the most current data sources. Forecasts and alternative scenarios are produced yearly out to a 30-year time horizon.

Data Geography

Historical and forecast data are available at the county, metropolitan area, State, and national levels.

Data Content

Historical variables include employment, wage and salary income, and real and nominal output for five-digit NAICS categories, back to 1970.

Forecast variables include population, labor force and unemployment rates, total employment, wage and salary income by one-digit NAICS category, retail sales, and single- and multifamily housing stock.

Data Acquisition

Information on specific data products can be requested from the Moody's Analytics Web site:

<https://www.economy.com/products/data>.

Technical documentation and pricing information is also available on the Web site.

Data Application

The data are useful for understanding changes in employment and demographics over time. Since the data are forecast out to 30 years, they can provide key forecasting variables for long-range transportation planning.

4.15 IHS Global Insight

IHS Global Insight (a division of IHS Markit) provides economic analysis, risk assessment, forecasting, and scenario planning for business and government clients, both nationally and worldwide. Their BMI database provides both historical trends and 25-year forecasts of employment, number of establishments and sales volumes, by industry, for States, counties, metropolitan areas, and Census regions.

Data Coverage

The BMI database includes historical data and forecasts of number of establishments, employment, and sales volume by four-digit NAICS code for United States at several levels of geography, including counties, metropolitan areas, States, and Census regions.

Data Collection Methodology

Historical economic data are extracted from a variety of public sources, including BEA, BLS, and Census; and extend back to 1990 (1997 for sales volumes). Forecasts are developed using proprietary forecasting models, and are updated semiannually out to a 25-year time horizon.

Data Geography

Historical and forecast data are available at the county, metropolitan area, and State levels.

Data Content

Historical and forecast variables include number of establishments, employment, and total sales volume for four-digit NAICS categories.

Data Acquisition

Additional information on the BMI service can be obtained from the IHS Markit Web site:

<https://www.ihs.com/products/business-market-insights.html>.

Data Application

The data are useful for understanding changes in key business criteria over time. Since the data are forecast out to 25 years, they can provide key forecasting variables for long-range transportation planning.

CHAPTER 5

Resources and Links

The following published research reports and resources provide additional detailed information on the data sources covered within this resource guide. Published information on private data sources are not commonly available, but may be requested from private vendors. Information on public data sources includes applications for transportation planning, technical documentation, and detail on methodology. The following list is not exhaustive and new resources are continually available.

Transportation Research Board Resources

Census Data for Transportation Planning, TRB Subcommittee on Census Data for Transportation Planning, ABJ30(1), <http://www.trbcensus.com/index.html>.

NCHRP 08-36, Task 098 Improving Employment Data for Transportation Planning, [http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36\(98\)_FR.pdf](http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36(98)_FR.pdf).

NCHRP 08-36, Task 119 Transportation Users Guide to the Economic Census, [http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36\(119\)_FR.pdf](http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36(119)_FR.pdf).

NCHRP 08-36, Task 128 Data Visualization Methods for Transportation Agencies, <http://www.trb.org/Main/Blurbs/175902.aspx>.

NCHRP 08-36, Task 135, Addressing Margins of Error in Small Areas of Data Delivered through the American Fact Finder or the Census Transportation Planning Products Program, <https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4078>.

NCHRP 08-48 Report 588 Using American Community Survey Data for Transportation Planning, <http://www.trb.org/Publications/Blurbs/156802.aspx>.

NCHRP Synthesis 434: Use of the U.S. Census Bureau's Public Use Microdata Sample (PUMS) by State Departments of Transportation and Metropolitan Planning Organizations, <http://www.trb.org/Main/Blurbs/167250.aspx>.

Current Population Survey

CPS, Technical Documentation, <https://www.census.gov/programs-surveys/cps/technical-documentation.html>.

Current Employment Statistics

CES, Technical Documentation, <https://www.bls.gov/web/empst/cestn.htm>.

Mullins, J., One hundred Years of Current Employment Statistics—An Overview of Survey Advancements, Monthly Labor Review, U.S. Bureau of Labor Statistics, August 2016, <https://doi.org/10.21916/mlr.2016.39>.

Nonemployer Statistics

NES, Technical Documentation, <https://www.census.gov/programs-surveys/nonemployer-statistics/technical-documentation.html>.

Nonemployer Statistics Data User Guide, 2017, <https://www2.census.gov/programs-surveys/nonemployer-statistics/guidance/nonemployer-statistics-user-guide.pdf>.

Quarterly Census of Employment and Wages

QCEW Technical Documentation, <https://www.bls.gov/opub/hom/cew/home.htm>.

Clayton, R., Data Sharing: Progress and Challenges at BLS, Division of Administrative Statistics and Labor Turnover, U.S. Bureau of Labor Statistics, December 2014, https://www2.census.gov/adrm/fesac/2014-12-12_clayton.pdf.

Konigsberg, S., Talan, D., and Clayton, R., The Geospatial Distribution of Employment: Examples from the Bureau of Labor Statistics Quarterly Census of Employment and Wages Program, Division of Administrative Statistics and Labor Turnover, Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics, Monthly Labor Review, March 2007, <https://www.bls.gov/opub/mlr/2007/03/art4full.pdf>.

County Business Patterns

CBP, Technical Documentation, <https://www.census.gov/programs-surveys/cbp/technical-documentation.html>.

American Community Survey

ACS, Technical Documentation, <https://www.census.gov/programs-surveys/acs/technical-documentation.html>.

FHWA, ACS for Planning Archive, https://www.fhwa.dot.gov/planning/census_issues/american_community_survey/.

American Community Survey Content Test Evaluation Report: Journey to Work–Travel Mode of Commute and Time of Departure for Work, American Community Survey Research and Evaluation Program, September 2017, https://www.census.gov/content/dam/Census/library/working-papers/2017/acs/2017_McKenzie_01.pdf.

NCHRP 08-48 Report 588, Using American Community Survey Data for Transportation Planning, <http://www.trb.org/Publications/Blurbs/156802.aspx>.

Census Transportation Planning Products

Seo, J., T. Vo, S. Lee, F. Wen, and S. Choi, The CTPP Workplace Data for Transportation Planning: A Systematic Review, Commissioned Paper for: 2017 Applying Census Data for Transportation Conference, September 29, 2017, <http://onlinepubs.trb.org/onlinepubs/conferences/2017/censusdata/WorkplaceData.pdf>.

Public Use Microdata Sample

PUMS, Technical Documentation, <https://www.census.gov/programs-surveys/acs/technical-documentation/pums.html>.

NCHRP) Synthesis 434: Use of the U.S. Census Bureau's Public Use Microdata Sample (PUMS) by State Departments of Transportation and Metropolitan Planning Organizations, <http://www.trb.org/Main/Blurbs/167250.aspx>.

Longitudinal Employer Household Dynamics Program

LEHD, Technical Documentation, <https://lehd.ces.census.gov/learning/>.

LEHD, Research and Transportation Applications, <https://lehd.ces.census.gov/research/>.

CHAPTER 6

Glossary

This chapter provides definitions of commonly used terms in employment data sources.

Administrative Records: Data that are collected and/or maintained by Federal, State, or local government agencies as part of a legislative mandate or normal agency operations. Examples of administrative records include quarterly reports filed by employers, as mandated by the Federal Unemployment Tax Act, IRS tax returns, Social Security records, and Office or Personnel Management employee records.

Employed: A person who did any work for pay or profit during the survey reference week; or who did at least 15 hours of unpaid work in a family-operated business; or who was temporarily absent from their regular jobs because of illness, vacation, bad weather, industrial dispute, or various personal reasons.

Employer Identification Number (EIN): A unique nine-digit number assigned by the IRS to business entities operating in the United States for the purposes of identification. The **EIN** (also known as **Federal Employer Identification Number** or **FEIN**) is the corporate equivalent to a Social Security Number, although it is issued to anyone, including individuals, who have to pay withholding taxes on employees. It also is issued to entities, such as States, government agencies, corporations, limited liability companies, and any other organization that must have a number for a purpose in addition to reporting withholding tax, such as for opening a bank or brokerage account.

Establishment (BLS): An establishment, as defined by BLS, is a single physical location at which business is conducted, or services or industrial operations are performed. It is not necessarily identical with a company or enterprise, which may consist of one or more establishments. When two or more activities are carried on at a single location under a single ownership, all activities generally are grouped together as a single establishment. The entire establishment is classified on the basis of its major activity, and all data are included in that classification.

Group Quarters: A group quarters is a place where people live or stay other than the usual house, apartment, or mobile home, which are defined as housing units. Group quarters may have housing units on the premises for staff or guests. Two general types of group quarters are recognized:

1. **Institutional Group Quarters:** Facilities that house people who are primarily ineligible, unable, or unlikely to participate in the labor force while residents, such as adult correctional facilities, juvenile facilities, skilled-nursing facilities, and other facilities such as mental (psychiatric) hospitals and in-patient hospice facilities.
2. **Noninstitutional Group Quarters:** Facilities that house those who are primarily eligible, able, or likely to participate in the labor force while residents, such as college dormitories, military barracks, emergency and homeless shelters, adult group homes, and residential treatment centers.

Household: A household consists of all the people who occupy a housing unit. A household includes all related family members and any unrelated people, if any, such as lodgers, foster children, wards, or employees who share the housing unit. A person living alone in a housing unit, or a group of unrelated

people sharing a housing unit such as partners or roomers, also is counted as a household. The count of households excludes group quarters.

Housing Unit: A housing unit is a house, an apartment, a mobile home or trailer, a group of rooms, or a single room that is occupied, or, if vacant, is intended for occupancy as separate living quarters. Separate living quarters are those in which the occupants live separately from any other persons in the building, and which have direct access from the outside of the building or through a common hall.

Labor Force: Noninstitutionalized civilians who are 16 year or older and can be classified as employed or unemployed.

Legal Form of Organization: The Legal Form of Organization is derived from administrative records data sources. The following is a list of Legal Forms of Organization:

- **Corporation:** An incorporated business that is granted a charter recognizing it as a separate legal entity having its own privileges, and liabilities distinct from those of its members.
- **S-Corporation:** A form of corporation where the entity does not pay any Federal income taxes. The corporation's income or losses are divided among and passed to its shareholders. The shareholders must then report the income or loss on their own individual income tax returns.
- **Sole Proprietorships:** An unincorporated business with a sole owner.
- **Partnership:** An unincorporated business where two or more persons join to carry on a trade or business with each having a shared financial interest in the business.
- **Nonprofit:** An organization that does not distribute surplus funds to its owners or shareholders, but instead uses surplus funds to help pursue its goals. Most nonprofit organizations are exempt from income taxes.
- **Government:** A business that taxpayers primarily fund.

Metropolitan Area: A metropolitan statistical area (MSA), as defined by the Office of Management and Budget for the United States and Puerto Rico. Each MSA consists of one or more adjacent counties or county equivalents that have at least one urbanized area of at least 50,000 population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties.

Noninstitutionalized Civilians: Persons who do not live in institutional group quarters, and who are not on active duty in the Armed Forces.

Not in Labor Force: Persons 16 years old and over who are not classified as in the labor force. This category consists mainly of students, homemakers, retired workers, seasonal workers interviewed in an off-season who were not looking for work, institutionalized people, and people doing only incidental unpaid family work (less than 15 hours during the reference week).

Reference Week: The week for which the survey questions are asked.

Seasonal Adjustment: Over the course of a year, the size of the labor force, the levels of employment and unemployment, and other measures of labor market activity undergo sharp fluctuations due to such seasonal events as changes in weather, major holidays, and the opening and closing of schools. Because these seasonal events follow a more or less regular pattern each year, their influence on statistical trends can be eliminated by adjusting the statistics from month to month. These adjustments make it easier to observe the cyclical and other nonseasonal movements in the series. BLS regularly produces seasonally adjusted series for selected labor force data from the CPS.

Self-Employed Worker: A person who works for themselves and does not incorporate their business into any legal entity.

Size of Business: The size of business means the size of an organization by number of employees, sales volume, square footage, and business expense.

Social Data (ACS): The social data in the ACS include marital status, school enrollment, educational attainment, veteran status, disability status, place of birth, religion, language spoken, computer and Internet use, ancestry etc.

Unemployed: A person who did not have a job during the survey reference week, was actively looking for work for the past four weeks, and was available to work during the survey reference week.

Unemployment Insurance (UI): Insurance benefits paid by the State or Federal government to individuals who are involuntarily out of work in order to provide them with necessities, such as food, clothing, and shelter. In general, a tax on employers provides the funds to pay unemployment compensation. Each State establishes which employers are obligated to pay State unemployment taxes. Ordinarily a State will require payment of the tax from every individual, partnership, or corporation that pays wages to a specified minimum number of people to do work. Certain types of employment are excluded from mandated coverage, including some agricultural labor, certain charitable or nonprofit work, and some government work.

Waged/Salaried Worker: A person who works for a legal entity and receives payment for their work.

List of Acronyms

ACS	American Community Survey
BEA	Bureau of Economic Analysis
BLS	Bureau of Labor Statistics
CATI	Computer-Assisted Telephone Interview
CBP	County Business Patterns
CES	Current Employment Statistics
CPR	Composite Person Record
CPS	Current Population Survey
CSA	Combined Statistical Area
CSV	Comma Separated Value
CTPP	Census Transportation Planning Products
ECF	Employer Characteristics File
EDI	Electronic Data Interchange
EIN	Employer Identification Number
EQUI	Enhanced Quarterly Unemployment Insurance
FIPS	Federal Information Processing Standard
GQ	Group Quarters
HU	Housing Unit
ICF	Individual Characteristics File
IRS	Internal Revenue Service
LED	Local Employment Dynamics
LEHD	Longitudinal Employer Household Dynamics
LODES	LEHD Origin-Destination Employment Statistics
MAL	Master Address List
MDIV	Metropolitan Division
MICRO	Micropolitan Statistical Area
MSA	Metropolitan Statistical Area
NAICS	North American Industrial Classification System
NETS	National Employment Time Series

OD	Origin-Destination
OMB	Office of Management and Budget
PIK	Protected Identification Key
PUMS	Public Use Microdata Sample
QCEW	Quarterly Census of Employment and Wages
QWI	Quarterly Workforce Indicators
RAC	Residence Area Characteristics
SEIN	State Employer Identification Number
SESA	State Employment security Agency
SIC	Standard Industrial Classification
SOC	Standard Occupational Classification
SSN	Social Security Number
StARS	Statistical Administrative Records System
TAZ	Transportation Analysis Zone
TDE	Touchtone Data Entry
UCFE	Unemployment Compensation for Federal Employees
USPS	United States Postal Service
UI	Unemployment Insurance
WAC	Workplace Area Characteristics
WIB	Workforce Investment Board

APPENDIX A

Practitioner Survey Results

A survey of transportation planning practitioners was developed and administered to inform this National Cooperative Highway Research Program (NCHRP) 8-36.127 guidebook. Survey questions reflected several major themes and sought information on the use of employment data in real-world transportation and transit planning purposes. Major questions asked in the survey included:

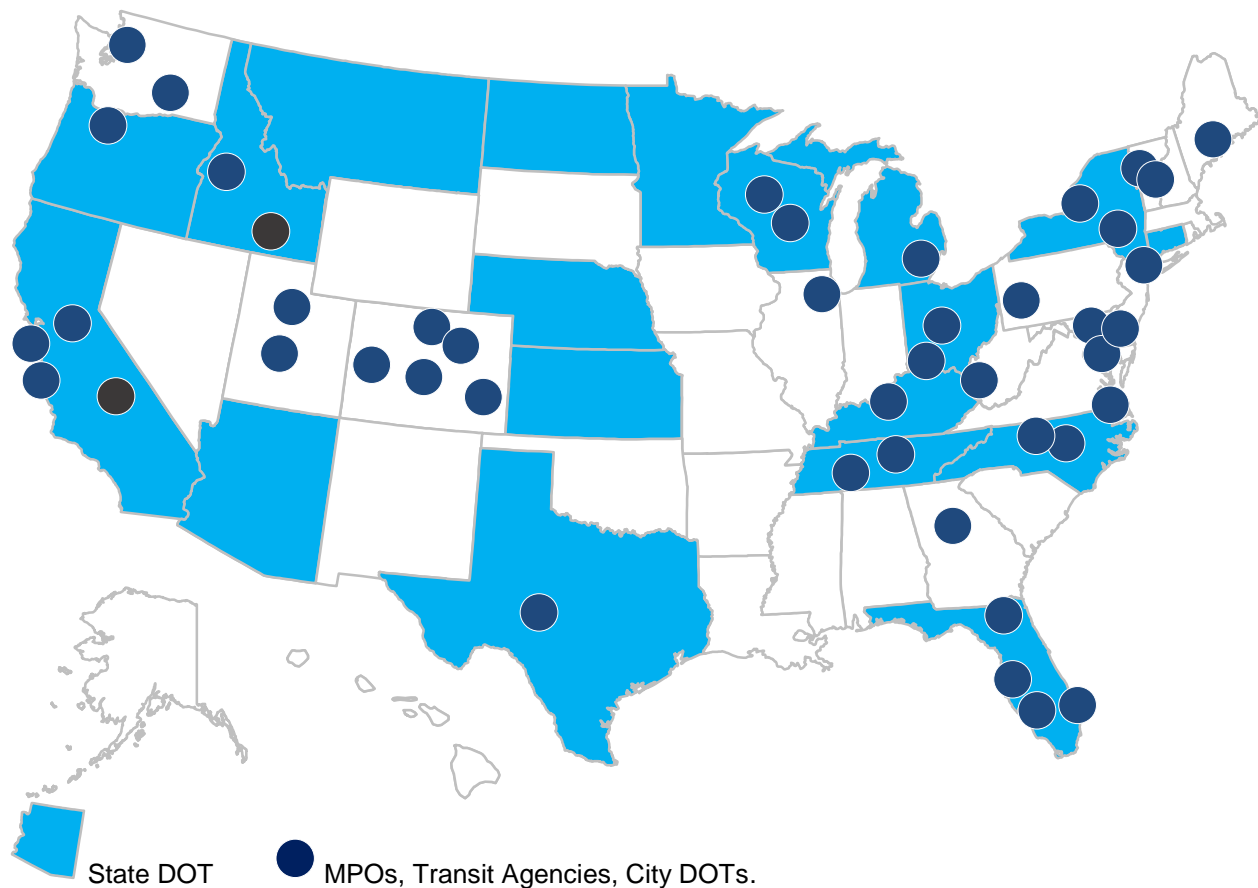
- How is employment data used?
- How are data issues resolved or worked around?
- How is employment data analyzed and communicated?
- What best practices should be highlighted?
- What topics and Frequently Asked Questions (FAQ) should be covered in the resource guide?
- What are the strengths and weaknesses of major datasets?

An online survey was distributed between December 2015 and February 2016.

- Direct email messages were sent to all 406 metropolitan planning organizations (MPO) contacts listed in the Federal Highway Administration's (FHWA) directory. This directory includes many smaller and mid-sized MPOs that are not members of national associations.
- Requests for distribution were sent to contacts at the American Association of State Highway and Transportation Officials (AASHTO), Association of Metropolitan Planning Organizations (AMPO), and National Association of Regional Councils (NARC). The survey was distributed to members of these organizations and resulted in responses.
- The survey link was posted for two consecutive months to the Travel Model Improvement Program (TMIP) and Census Transportation Planning Products (CTPP) list-servers. These postings resulted in relatively few, but quality responses.
- American Public Transportation Association (APTA) was engaged directly and agreed to distribute the survey on behalf of NCHRP. However, few transit agencies responded to the survey, which suggests that the survey link may not have been distributed.

Over 180 responses were received to the online survey with 112 complete or partially complete responses (i.e., most questions substantially answered). Figure A-1 shows the distribution of responding organizations across States and regions. About 19 State departments of transportation (DOT), two large urban city DOTs, 43 MPOs or regional organizations, four transit agencies, and 14 public or private parties participated. In many cases, respondents in different positions or divisions of the same agency provided information.

Figure A-1. Survey responses received by state and regional agency.



Survey respondents represent a range of roles in planning, forecasting, data management, and modeling. A selection of representative titles of survey respondents is included below. This highlights the broad range of perspectives engaged and professionals actively using employment data in transportation and transit planning and forecasting.

- Data and Mapping Administrator.
- Demographer.
- Economic and Systems Data Analyst.
- Economic Development Section.
- Economic Trends Planner.
- Executive Director.
- Geographic Information System (GIS) Coordinator.
- Information Systems Manager.
- Market Analysis, Fare Policy, Ridership Forecasting.
- Modeling and Statewide Data Development.
- Public Policy Analyst.
- Regional Planner.

- Research and Data Manager.
- Socioeconomic Program Manager.
- Traffic Analysis Engineer.
- Traffic Forecasting Branch Chief.
- Transit Planner.
- Transportation Analyst Policy and Planning.
- Transportation/Traffic Engineer.
- Transportation Planner/Analyst.
- Travel Demand Model Estimation, Calibration, and Validation.

Several public and private parties responded to the survey, including universities, research centers, data or professional services vendors, and transportation consulting firms. Generally, these respondents provided limited information and responses do not significantly change overall findings. In some cases, consultants indicated they were located onsite for an agency and/or responding on behalf of a State DOT.

Survey Results

The following sections break down responses to each survey question. Overall responses are summarized, and a selection of relevant verbatim comments are included.

Employment Data Applications

As shown in Figure A-2 to Figure A-5, this set of questions probed the use of employment data for a variety of projects or efforts within long-range planning, travel demand forecasting, local area or project planning, and modal planning. Generally, employment data is more frequently used at broad geographic levels (State, region, or corridor); and higher-level analysis, such as scenario planning, model development and forecasting, and modal analysis. Data may be used less frequently for specific project or detailed analyses (e.g., safety, environmental justice, station area, operations, and microsimulation). This pattern reflects common uses of data today and limitations of some data sources at local or subarea levels.

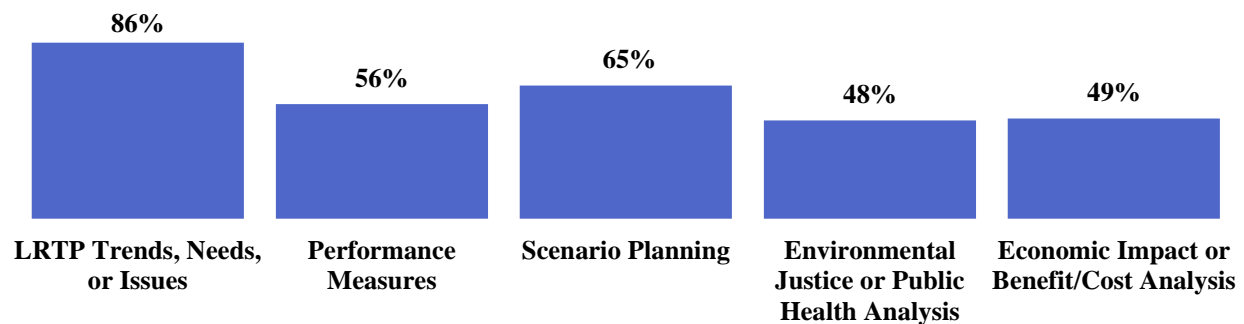


Figure A-2. Question 1A: What types of projects do you use employment data for in long-range planning?

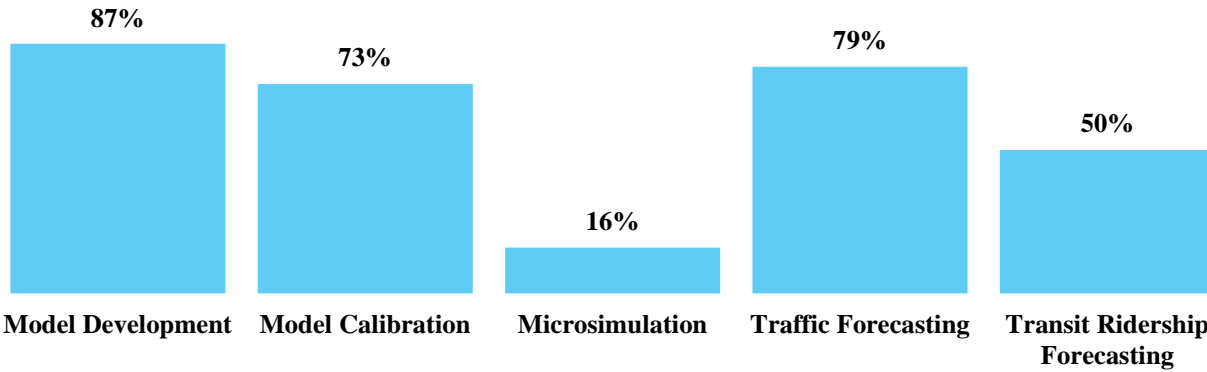


Figure A-3. Question 1B: What types of projects do you use employment data for in travel demand forecasting?

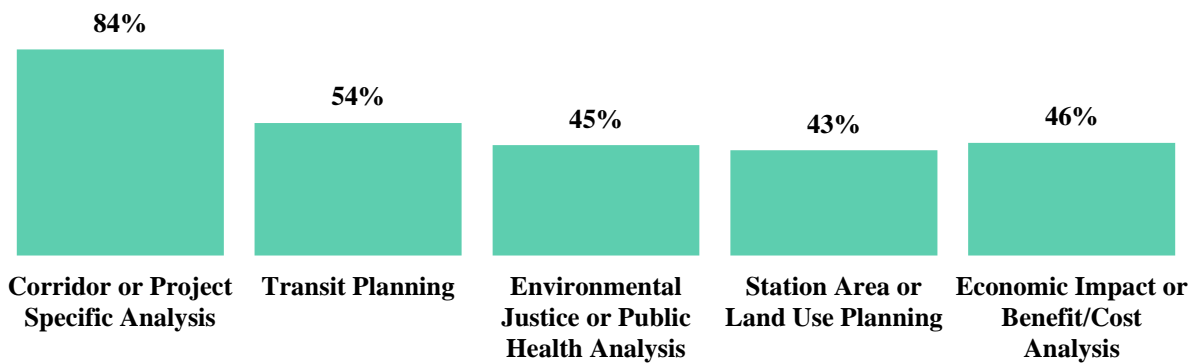


Figure A-4. Question 1C: What types of projects do you use employment data for in local project planning?

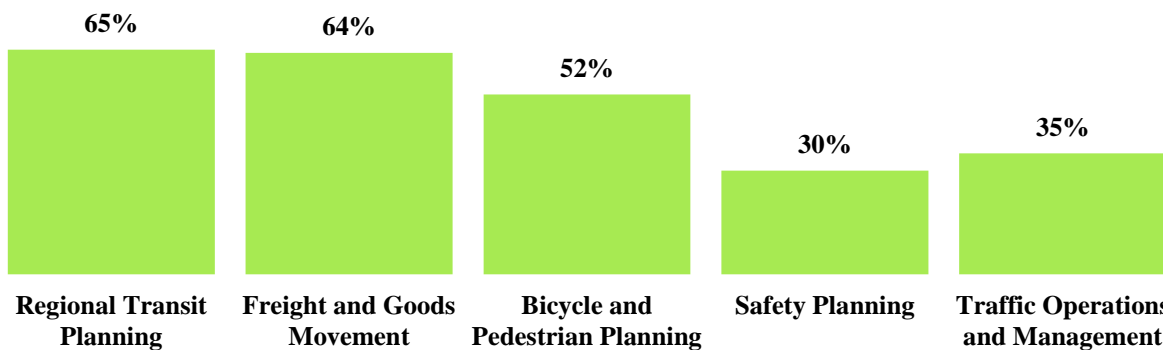


Figure A-5. Question 1D: What types of projects do you use employment data for in modal planning?

A follow-up, open-ended question collected other uses of employment data, including instances where employment was essential to the planning process. Employment data has been used widely in analyses: from Title VI analyses to hurricane evacuation planning, legislative responses and metro newsletters, transit indexes to corridor analyses.

Question 1E: Please tell us a little about how you've used employment data in other project applications or purposes?

- We use MPO HHTS data for O/D analysis; Commute Trip Reduction Program for employee Zip Code origins; State ESD and Census OTM data for jobs in corridors, etc.; ACS for Vehicle Available, mode of travel to work, general demographics; transit boardings and alightings for station area planning and analysis.
- Developed a transit corridor typology index using block level Census Population and LEHD employment.
- Use ACS demographic data to construct a Residential Transit Orientation Index at the block group level and use LEHD data to identify employment concentrations.
- Developing performance measures for the region's long-range plan. Corridor and station level analysis including economic characteristics analysis was done around the region as part of a Sustainable Communities grant from HUD.
- Metrotrends newsletters, annual economic forecast/report for city, planning studies for local cities and counties. Answering data requests from member jurisdictions and general public.
- We create an employment inventory every 3-4 years beginning with InfoUSA point level data and then fold in other sources of data for sectors identified as weak in the InfoUSA data. We assess other aggregate sources of larger area employment data to identify holes in our inventory.
- Imputed cellular data to develop models of regular workplace location choice, time of day, and mode choice (including auto occupancy level).
- Increased employment for analysis of TOD areas.
- Hurricane Evacuation. State of Florida updates its hurricane evacuation data through the Regional Planning Councils.
- Parking studies.
- We used employment data from both private and State resources to locate every distribution center and warehouse in the State.
- Mapping transit routes, greenways and sidewalks in relation to poverty status, income status and/or habitations of foreign born residents.
- Use InfoUSA, BEA, CES, QCEW, ACS, LEHD, and/or CTPP for local government data requests. Also using a business establishment database from Michigan Department of Transportation for our input into our socioeconomic and land use forecast.
- We use it in Title VI analyses of transit service changes. We use it as a source to identify potential corridors for new transit service. Through the Remix product, we use it to analyze the changes in accessibility to jobs from changes in the transit network.
- Used in understanding employment type, numbers, and densities by specific address location when siting a multi-modal transit facility.
- InfoGroup employer information was used in the development of corridor importance factors. An upcoming analysis of carpooling will include commutation information from LODES.
- I have used LEHD and ACS CTPP data for multiple types of projects. We have used LEHD to identify what percent of workers could be served by commuter rail service.

- LEHD and CTPP was used to develop ridership and parking forecasts, and to determine the impact of new fare policies to serve riders.
- I have used this data to provide maps for our members.
- Market research for potential downtown retailers.
- Crucial to the development of the MPO's Land Use and Travel Demand Models as well as our Regional Transportation Plans.
- Legislative responses, statewide analysis model, high speed rail analysis, and international good movement
- TDM Estimation Trip Characteristics Market Segmentation, Stop Locations, Bike/Ped Planning Development of Performance Metrics.
- Data used in research for measuring the employment impacts of transportation improvements. Data used in TIGER applications.
- Analysis of housing and transportation affordability in region - building upon the idea of the Center for Neighborhood Technologies model, and before HUD released their Location Affordability model. Used the data from LEHD/ LODES to approximate driving distances for work on a town-by-town level.
- Ridership projection for regional transit route study.
- We've used employment data in annexation studies, but we primarily use it for MTP development.

Employment Data Best Practices

This set of questions explored best practices for problem-solving, data visualization, and interesting applications of employment data. Multiple agencies indicated they cross-check employment data against different sources or across different years using a variety of techniques and databases. Nearly all agencies responding to the survey indicated interest in data visualization and many shared best practices for mapping.

Question 2: What solutions have you developed for addressing data weaknesses or issues?

- | | |
|-------------------|---|
| Satellite imagery | <ul style="list-style-type: none">• Google earth to verify locations and size• To confirm employment data, we usually take two approaches at the same time. First we use satellite imagery to confirm the employment facilities.• Based on this information, we contact local governments to verify its employment data including size, location, personal information, etc.• Using high resolution aerial imagery (identifying parking lots capacity) for individual businesses (sometimes the geographic coordinates of the employment center, if provided is not the actual location, but rather the human resources office). |
| “Ground truthing” | <ul style="list-style-type: none">• We verify the locations of firms from QCEW via phone and field review.• Verify top employers by phone. |

- | | |
|-------------------------------------|---|
| Rely on local knowledge | <ul style="list-style-type: none"> • Local knowledge and field work. • Have local planners perform consistency checks of the largest employers in their jurisdiction. |
| Cross-check with other data sources | <ul style="list-style-type: none"> • Using multiple sources (e.g., VEC and military). • We track development activity. • Population demand for commercial space, retail and services. • For the mapping we've done, I have used asymmetric mapping using our Open Space geodatabase to exclude jobs from locations jobs are highly unlikely to be located. • Utilized cellular data, Bluetooth, GPS, conduct own surveys. Participate in NHTS, use multiple sources to cross check other data, build relationships with data providers to enhance accuracy. • One trick I've used to be able to compare different sources and possibly data from different years is to convert raw numbers of employees to percentages of workers resident in a particular county (our model area includes nine whole and two partial counties) who work outside their home. Estimating margins of error (if not already provided, as with the ACS) is also helpful in order to compare ranges of percentages (overlapping or disjoint?). |

Question 3: How have you utilized data visualization, mapping, display, or analytical tools to better communicate and access employment data?

Agencies create density maps—either dot density or choropleth (heat map)—to aid in project planning. Mapping allows planners to locate future commercial centers, as well as map commuting flows. Other data mapping software used includes Tableau, Google Earth, Geomedia, LEHD OnTheMap and TransCad. When mapped, employment data is often summarized at the census block or TAZ level. Some agencies publish interactive maps online, and much of the visualization software is migrating to the cloud. Programs or applications used to visualize data include the following:

- | | |
|-------------------|------------------|
| • ArcGIS/ArcView. | • Bing Maps. |
| • ESRI online. | • Postgre SWL. |
| • Tableau. | • Quantum GIS. |
| • Google Earth. | • Transcad. |
| • Geomedia. | • Delphi Studio. |
| • LEHD OnTheMap. | • D3. |
| • OpenGeo Suite. | • Leaflet. |

The following are examples of maps available online:

- **Denver Visual Resources:** <https://drcog.org/services-and-resources/denver-regional-visual-resources>.
- **Chicago Metropolitan Agency for Planning:** <http://www.cmap.illinois.gov/economy/regional-economic-indicators>.

- **Mid-Region Council of Governments:** <http://www.mrcog-nm.gov/transportation/technical-services>.

Question 4: What are some best practices for effectively using employment data?

- We were able to use the employment point data to support a small starts application for BRT by drawing radii around proposed stops and quantifying the number of jobs that would have improved access. We also performed an accessibility analysis to show how many more people could reach these destinations in a shorter time.
- Using employment data to look at commuter patterns is really important. ACS data only tells us limited information, without the spatial component. We care less about how many minutes someone travels, and instead are interested in exactly where they are going from and to - which is available through LODES. We work with small towns - between 250 and 10,000 people - so a lot of employment data isn't available to us through other methods.
- Caltrans should have trained staff to provide detailed employment data at zip code detail. Could do within EAB if we were given sufficient time to self train and complete an analysis or requests.
- Compare figures to other sources such as State's employment agency, US Census American Community Survey and local studies.
- We create regional data for local governments and regional partners to use. When it came to our attention that some of our member governments were buying and cleaning QCEW as well, we intervened to stop the duplication of effort. We continued to buy and clean (e.g., geocode, distribute government jobs, add sole proprietors) QCEW data, but now we distribute the cleaned product to local governments that have signed a license agreement with the DLE. Not only does this save time and reduce duplication of effort, but it also helps standardize the employment data that we are using for local and regional analysis.
- Geocoding and have local planners review the data since they are most familiar with their areas.
- FDOT provides training for Comprehensive Demand Modeling. This is an introductory level class which includes employment data and data structure. Our employment data structure is governed through the Florida Standard Urban Modeling Structure for consistency throughout the State.
- Iowa DOT does a great job of keeping their Infogroup data quality-checked by transmitting known updates to Infogroup so known errors aren't replicated. We are considering following their model.
- I use NAICS codes cross referenced with the ITE trip generation guides to identify number of trips produced by certain types of employment.
- Calculations of change of access to jobs within 45 minutes from a point in the region.
- Other than in using for trip production in traffic modeling, it is a great way to use in cluster analysis. Cluster analysis tells us from a regional stand point what industries make up the landscape so our planners can better match transportation planning to the economic landscape.
- We created zonal level employment estimates by 2-digit NAICS code for the Tennessee DOT's statewide model using data fusion with LEHD block level employment estimates and a proprietary address level database together with county level control totals.
- Carefully paying attention to the definitions used for each data series. Conversion of the disparate units of measure to a single concept is helpful (e.g., number of persons working on the basis of a full-time equivalent man-hours conversion). Use caution and be aware of the pitfalls of the data sources. Creation of a one-stop for the information would be beneficial, especially if it were able to be integrated into graphic or geospatial software easily. Creating a department-wide best practice information sheet would ensure the data are used correctly.

- Scrutinize available data for accuracy -- not really interesting but essential to producing quality results.
- We have been trying to utilize LEHD data more and more. It does not have confidentiality issues and has proved to be very useful for a number of different statewide efforts. There is no specific project I would want to highlight over others; the approach has just been to get researchers and practitioners easy access to consistent statewide employment data from LEHD and doing that has worked that dataset into a number of our tools and processes.
- Continuing to build relationships to enhance data and utilizing Delphi panels at the local level.
- TIGER applications are the most recent and useful examples I have seen lately. FHWA discontinued forecasting highway construction employment.
- The CTR tools are the paradigm of SDOT's BPs. Large volumes of transportation facility data stored as map layers and citywide GIS provides excellent other data, including census, general infrastructure, cartographic. SDOT must have dozens of staff trained and using a broad range of data, including economic data and travel data, but they are not organized.

Employment Data Topics of Interest

Survey respondents were not overly concerned with technical methods or details of employment data sources, and were interested finding and using employment data in day-to-day planning activities.

Question 5: When asked what employment topics they would like to learn more about, agencies ranked topics in the following order:

1. Availability accuracy, and applications of public datasets.
2. Use of origin-destination and location data effectively and appropriately.
3. Availability, accuracy, and applications of commercial datasets.
4. Methodology and data collection of key data sources.
5. Accuracy of origin-destination and location data.
6. Detailed information on employment categorization and counts.
7. Availability and release schedule for key public datasets.

Question 6: When asked what Frequently Asked Questions they would like to see included in the resources guide, respondents suggested:

- GPS methods.
- How employment data can affect choices in transportation.
- Methods of combining or reconciling disparate employment estimates.
- Examples of visualizations of Origin/Destination data.
- Metadata for all employment data.
- Different error rates for different size regions, so that different places may better understand which products are most appropriate for specific locations and purposes.
- Ways to quickly explain to the public how data is collected and some of its shortfalls.
- Where to access public datasets.

- Table comparing the public datasets available. Similar to the table published in CTPP Status Report, October 2013. Create a similar comparison table for commercial datasets.
- The pitfalls or “traps” to be aware of when comparing different data sources, specifically the different data collection methods and data purpose. For example, QCEW data originates with workers’ comp or Unemployment Insurance (UI) claims, and therefore doesn’t cover small firms, independent contractors, or self-employed individuals.
- Availability of multiple sources for the employment data and if any are free of charge how often is the information updated.
- The public is frequently confused by the difference between household-survey and the place of work nonfarm payroll jobs (and similar) sources.
- Lists of best data sources/links for information categories.
- A simple how-to-use guide would be very helpful. I do not use this data every day, and have to think about the best way of obtaining it each time.
- Detailed information about differences between public and commercial datasets.
- What is the most reliable way to secure and enter data for self-employed businesses that are currently missed?
- Frequency of updates, number of samples (metadata) who uses the data, how they use the data, examples of visualization, programs used for visualization.
- Large corporations often list all employees at Administrative Headquarters. Would like to see more localized breakdowns. Also the “work from home” option by number of days, etc. is becoming a larger impact on traffic/transit ridership.
- What is the most efficient method to review large datasets?
- Common methodology for assignment of employment data. What employment vendor has the most accurate data? How often should the employment data be updated?
- How reliable is the data? When was the data collected? What was the methodology for collecting?
- How can we get accurate employment data? How do we account for hard-to-reach employment such as home-based businesses, migrant farmworkers, etc.? What are typical employment per building size (sq. ft.) estimates that can be used for sketch-level planning? What is an effective way to determine this based on local data?
- How is employment data obtained and how frequently is it updated?
- Examples of how to properly use the data, techniques on checking accuracy of mapped data, shortcuts to mapping data, suggestions on ways to pair data to create a map that displays relevant info for planning.
- What are the best data sources and methodologies for analysis for different types of transit planning questions?
- How the military population plays into employment. We know that active military is excluded from the data. However, year to year changes in military personnel will help us better tailor a transportation network to a substantial segment of the population.
- How do costs measure up against accuracy of third party databases? How much do agencies invest in manually checking and correcting these databases and is it worth it? How are agencies using employment data to perform cost/benefit analysis at project and regional plan levels?

- Where do I get the information and how do I transform the raw information into what I need? Which one will provide what I need? What is it lacking? Are these available in spatial datasets? Why don't these match? Do I need to pay for 'good' information?
- How much error there is among LEHD compared to CTPP data? Which datasets to use when?
- Ways or methods to assess the error in these datasets for a person's specific region or use area.
- Concise information regarding the differences between various datasets in terms of types of employment that are included or not included. For example the county-level totals of employees are much higher in the Bureau of Economic Analysis (BEA) counts than in the Bureau of Labor Statistics (BLS) counts.
- NAICS codes for employment projections.
- Best practices and/or innovative approaches to post-processing geocoded employment data.
- What is the smallest unit of data available? County subdivisions? County? Can you self-define your area of interest?
- How to verify employment? What has worked, and which methods are not useful.
- What are best sources of (fill in the blank) data? Is local or regional employment data readily available for my city, county, region? Are there clear instructions for extraction data and referrals to forums or agency assistance?
- Where do I find the available publicly available employment datasets?

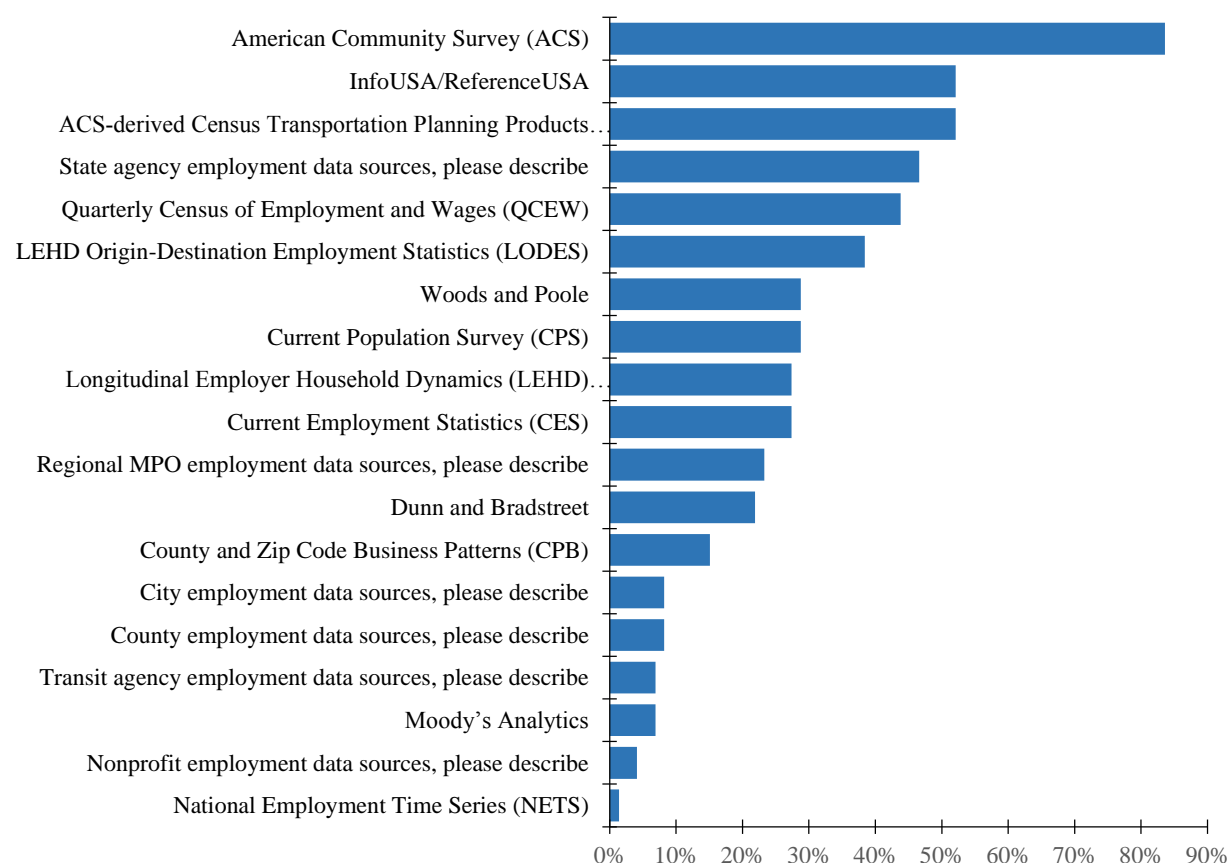


Figure A-6. Question 7: What data sources do you use most frequently?

Employment Data Strengths and Weaknesses

The following summary of perceived strengths and weaknesses of data sources is based on survey data and does not imply endorsement by NCHRP.

Current Population Survey

Strengths

- Methodology, trends.

Weaknesses

- Geographic resolution, origin/destination, workplace granularity, accuracy.
- Lack of availability except for large metro areas.
- Difficult to search through voluminous documentation.
- Requires knowledge and understanding of information collection and purpose.

Current Employment Statistics

Strengths

- Overall view of growing or shrinking industry sectors and seasonality.
- Cost.

Weaknesses

- Geographic resolution.
- Not available for middle-sized and small metro areas; not meant for local level information.

Quarterly Census of Employment and Wages

Strengths

- Geographic resolution.
- Attribute detail.
- Trends.
- Cost.
- Establishment counts.
- Helpful to combine with other sources for a better picture (i.e., Transearch).

Weaknesses

- Micro data has confidentiality restrictions.

American Community Survey—CTPP

Strengths

- Geographic resolution, origin/destination, GIS compatibility, attribute detail, cost.
- Includes all employment including farm workers, part time, seasonal.
- Use for proprietary info that is missing from other sources.
- Prepared specifically for transportation planning community using input from that community.

Weaknesses

- Accuracy, timeliness.
- Better search option needed.
- Poor resolution for large rural tracks.
- Suppression of records and higher margin of error at small geography.

LEHD Quarterly Workforce Indicators

Strengths

- Geographic resolution, origin/destination, GIS compatibility, attribute detail, cost.
- Visualizations.
- No need for contract and can share freely with others.

Weaknesses

- Accuracy.
- Changes to North American Industry Classification System (NAICS) make time-series comparisons problematic.
- The “fuzziness” added to protect confidentiality.
- Inability to see who is being geocoded incorrectly.

LEHD – Origin-Destination Employment Statistics

Strengths

- Geographic resolution, origin/destination, GIS compatibility, cost.
- Source of block group-level commutation information and worker demographics.

Weaknesses

- Workplace granularity.
- Not all jobs included in series.
- Could be offered at multiple geography levels (i.e., BG, Tract, county, so users do not have to make calculations).
- The “fuzziness” added to protect confidentiality.
- Large datasets require knowledge on how to create and utilize. Know that it is older information (lags at least a year) and takes some time getting used to.
- Some States opting out causes issues with timeliness of data releases, as well as skewed data if working with a bi-State region.

Dun and Bradstreet

Strengths

- Geographic resolution, GIS compatibility, workplace granularity.
- Data is geocoded.

Weaknesses

- Cost.
- Lack of compatibility in employment-by-industry with other sets.
- Substantial data cleaning is needed (employment levels, location, industry codes, duplicate records at one address for the same business).

InfoUSA

Strengths

- Geographic resolution, GIS compatibility, attribute detail.
- Attribute availability.

Weaknesses

- Opportunity cost, cost.
- Lack of compatibility in employment-by-industry with over sets (especially Dun and Bradstreet).
- Purchasing through department procedures.
- Substantial data cleaning is needed (employment levels, location, industry codes, duplicate records at one address for the same business).
- Post-processing needed.
- Expensive source of specific employer information that is easily geocoded, but rife with errors.
- Cannot be collected as a “snapshot in time” due to means of collecting data.

Moody’s Analytics

Strengths

- Accuracy, timeliness.

Weaknesses

- Cost.

Woods & Poole

Strengths

- Attribute detail, trends.
- What we used for forecasting but not useful for current/existing datasets.
- Recognized source.

Weaknesses

- Geographic resolution, cost.
- Difficulty in sharing data.

State Sources

Strengths

- Geographic resolution, accuracy, timeliness, trends, cost.
- State Web sites easier to use than Bureau of Labor Statistics (BLS).
- Can work with other State departments to improve data collection and reporting.
- Local, trusted source.
- Great data source and worthwhile link with the pulse of local economic situations due to contact with Department of Labor.
- Some States can only specify a boundary (i.e., center city) and receive a total count of UI covered workers.

Weaknesses

- Geographic resolution, origin/destination, workplace granularity.
- Not familiar with all the uses.
- Some States provide little detail on age, sex, incomes, and other demographics of workers/jobs.
- Not on TAZ level.
- Data is sometimes aggregated by similar codes.
- Process to secure such data has numerous steps, is time-consuming, and confidentiality requirements decreases the utility of the data. Significant data cleaning required.
- Variable due to benchmarking.
- Requires confidentiality training and a memorandum of understanding (MOU) with the State Department of Labor; possible costs involved.
- Data needs a lot of cleaning.
- Data does not have enough attributes.

Regional Sources

Strengths

- Geographic resolution, GIS compatibility, attribute detail, workplace granularity.

Weaknesses

- Timeliness, opportunity cost.
- Still working out issues, such as meta data and version control.
- Difficult to obtain from all sources required.
- Not all MPOs have MPO-based forecasts.
- Historically only surveyed every five to seven years for model updates.

County Sources

Strengths

- Geographic resolution, GIS compatibility, attribute detail, workplace granularity.

Weaknesses

- Workplace granularity, methodology, timeliness, trends, opportunity cost.

City Sources

Strengths

- Geographic resolution, GIS compatibility, attribute detail, workplace granularity, accuracy.

Weaknesses

- Attribute detail, methodology, timeliness, opportunity cost.

Transit Sources

Strengths

- Geographic resolution, origin/destination, GIS compatibility, attribute detail, accuracy.

Weaknesses

- Methodology, timeliness, trends, opportunity cost, cost.

Nonprofit Sources

Strengths

- Geographic resolution, origin/destination, GIS compatibility, attribute detail, accuracy.

Weaknesses

- Methodology, timeliness, trends, opportunity cost, cost.