MERCATUS CENTER
George Mason University

DATA RESOURCE • COMMUTER MARKET ACCESS

Mercatus Commuter Market Access Dataset for Congested Auto Travel

Technical Documentation

Prepared for the Mercatus Center by

klumpentown

Contents

1. Overview

This document describes the Employment and Labor Market Access Dataset for Congested Auto Travel, which measures job and market accessibility from US Census tracts in the 50 states and Washington, D.C. under congested automobile travel conditions. Two accessibility approaches are available:

1. Raw counts binned by travel time interval, and
2. Distance-weighted aggregate accessibility scores

Each of these approaches are computed in both directions: job access (from a worker's tract to reachable jobs, via Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) Workplace Area Characteristics) and market access (from an employer's tract to reachable workers or population, via LODES Residence Area Characteristics and separately via American Community Survey (ACS) tract population characteristics). All products are produced at the Census Tract level, national in scope, and cover auto travel only. Most products draw on 2023-vintage data (see §2.1.2 for exceptions); ACS-derived market access products instead reflect the 2019-2023 5-Year ACS estimate window (§2.1.3).

Each product is available as a single national file, with identical per-tract extracts provided at the state level for convenience. The two ACS-derived market access products are provided as independent files, joined to their LODES-derived counterparts by tract and travel-time band rather than merged into a single file (§2.1.3). Methodology for all products is described in Section 2; column-level detail is in Section 3.

1.1. Data Products

- Raw workplace area characteristic job counts binned by five-minute travel time intervals.
- Tract-level distance-weighted workplace area characteristic job counts.
- Raw residence area characteristic worker counts binned by five-minute travel time intervals.
- Tract-level distance-weighted residence area characteristic worker counts (market access).
- Raw ACS tract population characteristics binned by five-minute travel time intervals.
- Tract-level distance-weighted ACS population characteristics (market access).

1.1.1. Employment Counts by Five-Minute Travel Time Intervals

This dataset contains employment (job) counts accessible from a given US Census tract in five-minute congested travel time intervals. Job categories and counts are derived from the LODES dataset provided by the United States Census. More information on the naming convention and structure of this dataset can be found in §3.1 on page 13.

1.1.2. Distance-Weighted Job Accessibility Scores

This dataset contains tract-level job accessibility scores, in which jobs reachable at longer travel times are weighted lower than those reachable at shorter travel times. The weighting is derived from a decay function estimated separately (see §2.2.2). More information on the naming convention and structure of this dataset can be found in §3.2 on page 15.

1.1.3. Worker Counts by Five-Minute Travel Time Interval

This dataset contains worker counts reachable from a given US Census tract in five-minute congested travel time intervals, measured in the reverse direction of the employment counts above: for a given origin tract, this reports the labor pool of residents reachable within each travel-time band. Worker categories and counts are derived from the LODES Residence Area Characteristics (RAC) dataset provided by the United States Census. More information on the naming convention and structure of this dataset can be found in §3.3 on page 17.

1.1.4. Distance-Weighted Market Access Scores (LODES)

This dataset contains tract-level market access scores, in which reachable workers at longer travel times are weighted lower than those reachable at shorter travel times, using the same decay function applied to job accessibility (see §2.2.2). More information on the naming convention and structure of this dataset can be found in §3.4 on page 19.

1.1.5. ACS Population Characteristics by Five-Minute Travel Time Interval

This dataset contains American Community Survey (ACS) population characteristics – total population, adult population, unemployment, and household income bands – reachable from a given US Census tract in five-minute congested travel time intervals, measured in the same reverse direction as the RAC-derived worker counts above. This product is provided as an independent, ACS-sourced complement to the LODES-derived worker counts, and is joined to the same travel-time bands by a shared key rather than merged into a single file (see §2.1.3). More information on the naming convention and structure of this dataset can be found in §3.5 on page 21.

1.1.6. Distance-Weighted Market Access Scores (ACS)

This dataset contains tract-level market access scores computed from ACS population characteristics rather than LODES worker counts, using the same decay function applied elsewhere in this dataset (see §2.2.2). More information on the naming convention and structure of this dataset can be found in §3.6 on page 23.

2. Methods

2.1. Data Sources and Vintages

2.1.1. Streetlight Congested Travel Time Matrices

Congested auto travel times.

Origin-destination points. Streetlight travel time matrices are computed directly at the Census Tract level, using each tract's population-weighted centroid as its representative origin and destination point. This differs from the block-to-tract crosswalk approach used for LODES data (§2.1.2): rather than aggregating finer-grained records up to tract, Streetlight represents each tract as a single point from the outset. Population weighting places this point closer to where residents are actually concentrated within the tract, rather than at the tract's geometric center, which may fall in an unpopulated area (e.g., parkland, industrial land, or a large lot at the tract's edge). Travel times in this dataset should therefore be interpreted as point-to-point estimates between representative locations, not as an aggregate or average of travel times across all possible origin-destination pairs within the two tracts.

Coverage. Travel time computations were attempted for census tract centroids in the 50 United States and Washington D.C. Puerto Rico, Guam, US Virgin Islands, Northern Mariana Islands, and American Samoa were not included in this dataset. There are a handful of census tracts in remote areas whose centroids do not meaningfully snap to the road network, and as a result travel times are not computed. **Table A.1 lists census tract coverage by state with notes on missing census tracts.** 99.93% of census tracts are included in this dataset.

In 2022 the Census Bureau reclassified Connecticut's county-equivalents from eight counties to nine planning regions, changing the county segment of every tract GEOID (e.g., `09001010101` became `09190010101`). The population-weighted centroid file used to seed the routed origin-destination matrix was never updated to reflect this change, so the matrix continued to reference tracts by their pre-2022 county codes while the labor-market (LODES) and current geographic (TIGER2025) data had already moved to the new planning-region codes. Because the underlying tract boundaries were largely unchanged, the fix did not require re-routing the state: a spatial crosswalk was built by matching each legacy centroid point to its corresponding tract in the current TIGER2025 geography, then applied directly to the existing matrix. This resulted in 98.52% of census tracts being covered in the sate (866 of 879 tracts); the small remaining gap (13 tracts) is most plausibly explained by a handful of genuine tract boundary splits introduced during the 2022 reclassification itself.

2.1.2. LODES8 Origin-Destination Employment Statistics

Job and worker counts are derived from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES), Format Version 8.4. LODES is released as state-level CSV files in three groups: Origin-Destination (OD, jobs tied to both a home and work Census Block), Residence Area Characteristics (RAC, jobs totaled by home block), and Workplace Area Characteristics (WAC, jobs totaled by work block). This dataset draws on WAC for job-access computations and RAC for labor-access computations; OD is not used directly, as travel time between tracts is instead supplied by the congested travel time matrices (§2.1.1).

Vintage. LODES data years span 2002-2023 for most states. This dataset uses the most recent available year for each state: 2023 for all states except Michigan (2021) and Alaska (2016), where 2023 data was not available at time of production. Job counts for these two states should be interpreted with this lag in mind, particularly for analyses sensitive to post-pandemic labor market shifts.

Geography. LODES8 is enumerated at the 2020 Census Tabulation Block level and aggregated to Census Tract for this dataset, consistent with the tract-level geography used throughout the analysis.

Geography crosswalk. Tract identifiers for aggregation are taken from the LODES geography crosswalk file (`[ST]_xwalk.csv.gz`) rather than inferred by truncating the first 11 characters of the Census block geocode. While the block geocode's leading digits do correspond to state, county, and tract FIPS codes by construction, using the crosswalk avoids relying on that convention holding for every vintage and edge case, and correctly captures blocks whose tract assignment has been adjusted independent of the raw geocode string (e.g., where geographic boundary corrections postdate the block code itself). The crosswalk's primary key, `tabblk2020`, is stable across the 2020 decennial cycle and is the same key used to join block-level LODES data to its corresponding tract; all tract aggregations in this dataset use this join rather than a string-parsed geocode.

Job type. This dataset uses Job Type JT00 (All Jobs), which includes both private and federal employment. Firm-specific columns (`CF*`, covering firm age and firm size) are only populated for private jobs (JT02) and are not included in this dataset's derived products.

Coverage caveats. A small number of state-years lack OD/WAC data entirely (residents employed elsewhere are still captured in RAC); none of the affected years fall within the 2023/2021/2016 vintages used here, so this does not affect the current release. Demographic segmentation (age, education, earnings, industry) has been available for all included states and vintages since 2009 and is populated throughout this dataset's LODES-derived columns.

Disclosure avoidance. LODES data is reviewed by the U.S. Census Bureau under approved disclosure avoidance practices prior to release, consistent with standard Census confidentiality protections for block-level tabulations.

2.1.3. American Community Survey (ACS) 5-Year Estimates

Two distinct uses of ACS data appear in this dataset: national-level commute-duration data used to calibrate the decay function (§2.2.2), and tract-level population characteristics used in the labor-access product. Both draw from the 2023 5-Year American Community Survey estimates, accessed via the Census Bureau API.

Vintage and reliability. Unlike LODES8, which reflects a single reference year, the ACS 5-Year product is a rolling average of survey responses collected continuously over 2019-2023. Estimates should be interpreted as characteristic of that five-year window rather than as a point-in-time snapshot of any single year, including 2023 itself. ACS estimates also carry sampling error, reported by the Bureau as margins of error (MOE); this dataset does not currently propagate MOE values into its derived products, and users requiring uncertainty bounds on ACS-derived figures should consult the source tables directly.

Decay curve calibration (Table B08134). The decay function described in §2.2.2 is calibrated from Table B08134 (Means of Transportation to Work by Travel Time), restricted to workers commuting by car, truck, or van, and queried at the national level (`geo_unit=us:*`) rather than by tract. This is a single national estimate applied uniformly across all tracts; it does not vary by geography, and does not capture regional differences in commute-time distributions.

Tract population characteristics (Tables B01001, B01003, B19001, B23025). The labor-access product draws four tables at the Census Tract level: total population (B01003), population by age and sex (B01001, used to derive total adult population), household income bands (B19001), and unemployment status for the population 16 years and over (B23025). These are joined to tract geography using the same 2020 Census Tract boundaries used throughout this dataset.

Comparability with LODES. ACS-derived income and employment figures are not directly comparable to their LODES counterparts and should not be summed or reconciled without accounting for their differing definitions: B19001 reports *household* income, collected annually, at the household's *residence*; LODES CE01-CE03 reports *individual job* earnings, collected monthly, at the worker's *workplace* (or, in the RAC-derived labor-access product, at the worker's residence but describing their current job). Similarly, B23025's unemployed count is defined relative to the civilian labor force aged 16 and over, not relative to total population, and should not be interpreted as a share of all residents. Comparisons between ACS and LODES tables in this dataset are intended as a rough cross-source reasonableness check (e.g., whether total population and total employed workers in a reachable area are of a plausible relative magnitude), not as a reconciled or blended statistic.

2.2. Methodology

2.2.1. Binned Access Approach

The binned access products (§1.1.1/3.1, §1.1.3/3.3) report raw counts of jobs or workers reachable within each five-minute travel-time interval, without any distance weighting applied. For origin tract i and travel-time band $[t_k, t_{k+1})$:

$$B_i^{[t_k, t_{k+1})} = \sum_{j \,:\, t_{ij} \in [t_k, t_{k+1})} O_j \tag{2.1}$$

where O_j is the relevant count at destination tract j (jobs, for the WAC-derived product; workers, for the RAC-derived product) and t_{ij} is the travel time between i and j from the travel time matrix (§2.1.1). Bands are left-inclusive and right-exclusive, as described in the data dictionary entries for each product. This formulation makes no assumption about the relative desirability of destinations at different travel times within a band, or between bands – a job reachable in 5 minutes and one reachable in 55 minutes are both counted, undifferentiated except by which band they fall into. The weighted access approach (§2.2.2) addresses this by applying a continuous decay function in place of discrete banding.

2.2.2. Weighted Access Approach

Census tract level job access scores and market access scores were computed using a weighted decay function, where destinations requiring more travel time were weighted lower than the average. In place of the discrete travel-time bands used in §2.2.1, a continuous decay function $w(t)$ is applied to each destination's contribution, so that closer destinations count more heavily than farther ones within an origin tract's score, rather than being grouped into equally-weighted bands. For origin tract i:

$$A_i = \sum_j O_j \cdot w(t_{ij}) \tag{2.2}$$

where O_j is the relevant count at destination tract j (jobs for the job-access product, or workers for the market/labor access product), and $w(t_{ij})$ is the decay function described below, evaluated at the travel time between origin i and destination j. The same decay function is applied in both directions; the following section describes its estimation and calibration.

Estimating a Decay Curve

To appropriately weight destinations at different travel times, a decay function $w(t)$ was estimated from national-level binned commute-duration data from the 2023 American Community Survey (ACS) 5-Year estimates, Table B08134 (travel time to work, workers commuting by car, truck, or van).

Domain. The congested travel time matrices underlying this dataset are computed to a maximum of 60 minutes. The decay function's domain is matched to this same 60-minute ceiling: the ACS "60 or more minutes" bin is excluded from calibration, so that $w(t)$ is estimated only over the travel-time range the underlying accessibility calculations actually span. Excluding this bin means the calibrated

survival function does not reach zero at $t = 60$; the residual weight represents the share of commuters whose travel time exceeds the modeled domain and is not assigned to any tract-pair in this dataset. This is a consequence of the 60-minute matrix cutoff, not an artifact of the decay function itself, and is discussed further in §2.3.1. For the data in this domain, the mean one-way travel time was 23.32 minutes, and the median was 22 minutes.

Functional form. The decay function is specified as lognormal in travel time:

$$w(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \tag{2.3}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. This form was selected on maximum-entropy grounds: given the empirical mean and variance of *log*-travel-time as the informative moments, the lognormal is the least-informative distribution consistent with them, rather than a form selected by best-fit search among arbitrary candidates. This corresponds to a standard assumption in travel-time modeling that variation in commute duration arises from multiplicative (compounding) delay factors rather than additive ones. A comparison between three common candidate fits (lognormal, normal, and exponential) are provided in Figure 2.1

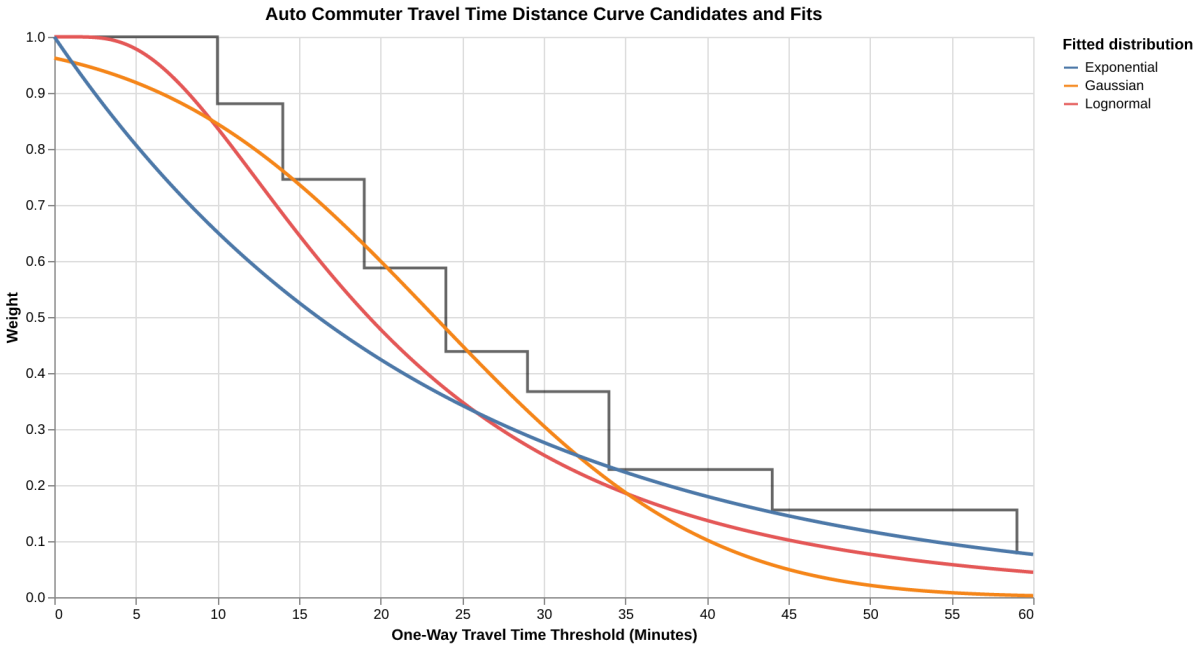

Figure 2.1.: Auto commuter travel time distance curves considered, compared with stepwise empirical data. A log-normal decay curve was selected.

Calibration. The parameters μ and σ were calibrated to the empirical mean and variance of log-commute-time from the ACS B08134 bins (excluding the 60+ bin, per above), equivalent to maximum-likelihood estimation for the lognormal family. The calibrated function is:

$$w(t) = 1 - \Phi\left(\frac{\ln t - 2.9584}{0.6676}\right) \tag{2.4}$$

Application. $w(t)$ as calibrated in Equation 2.4 is substituted into the general weighted access formula (Equation 2.2) to produce both the job-access and market-access scores.

2.3. Known Limitations

2.3.1. The 60-Minute Cutoff and Supercommuting

The congested travel time matrices underlying every product in this dataset are computed to a maximum of 60 minutes (§2.1.1). Workers whose actual commute exceeds 60 minutes (so-called "supercommuters") are therefore entirely absent from this dataset's job access, market access, and binned interval products; they are not down-weighted or approximated, they are simply not represented in any tract-pair. The decay curve calibration (§2.2.2) reflects this same domain restriction: the ACS "60 or more minutes" bin was excluded from calibration rather than assigned an arbitrary upper bound, and the resulting survival function does not reach zero at $t = 60$. The excluded population of workers reporting a 60-minute-or-longer commute in ACS B08134 represented 7.65% of national commuters in the calibration data. This share is not uniform across geography; tracts in large or sparsely-connected metropolitan areas are likely to have a higher share of unrepresented supercommuters than this dataset can reflect.

2.3.2. A Single National Decay Curve

The decay function $w(t)$ (§2.2.2) is calibrated once, from a single national ACS estimate, and applied uniformly to every tract in every product. It does not vary by region, urban form, or local commute-time distribution. A dense urban tract and a rural tract are weighted by the identical function, even though the true relationship between travel time and destination desirability plausibly differs between the two. Users interested in sub-national or metro-specific decay behavior should treat this dataset's weighted products as a national default rather than a locally-calibrated estimate.

2.3.3. Symmetric Decay Assumption Between Job and Market Access

The same decay function $w(t)$, calibrated from worker commute-time data, is applied to both job access (§1.1.2) and market access (§1.1.4, §1.1.6) products. This assumes that travel-time impedance is symmetric between a worker's tolerance for commuting to a job and an employer's effective reach into a labor pool. This is a modeling convenience, not a measured or independently justified equivalence; no attempt has been made in this dataset to estimate a separate decay function for the employer-side direction.

2.3.4. Realized Travel Behavior as a Proxy for Impedance

The decay function's functional form is selected on maximum-entropy grounds (§2.2.2), which provides a principled justification for the lognormal family given the empirical moments of log-commute-time. This argument is silent on mechanism: it does not establish that the observed decline in commute frequency with travel time reflects genuine travel-time disutility, as opposed to a confounded outcome of residential sorting, housing costs, and land-use patterns that happen to correlate with

distance. A confounded process and a genuine impedance-driven process can produce the same realized commute-time distribution. This dataset's accessibility scores should be interpreted as descriptive of realized commuting patterns weighted by a theoretically-motivated decay assumption, not as a validated behavioral model of destination choice.

2.3.5. Point-to-Point Travel Times

Travel times are computed between each tract's population-weighted centroid (§2.1.1), not as an average across all possible origin-destination pairs within the two tracts. A resident located near a tract's edge may experience a meaningfully different travel time than the centroid-to-centroid value used throughout this dataset, particularly for large or irregularly-shaped tracts. This understates within-tract heterogeneity in accessibility.

3. Data Dictionaries

3.1. Workplace Area Characteristic Jobs by Travel Time Interval

This data product is available at the national and state level. All sub-national files include destinations accessible outside of the provided geography and are identical to the national dataset; these files are provided for convenience only.

3.1.1. Files

- **National:** `wac_S000_JT00_binned_5min.csv.gz`
- **State:** `[ST]_wac_S000_JT00_binned_5min.csv.gz`, where:
  - `[ST]` is the two-letter capitalized postal code identifier for the sate (e.g. AL for Alabama).

3.1.2. Data Notes

This data generally reproduces job values in the LEHD/LODES provided by the United States Census. Note the following that may impact analyses with this data:

- For most states, LODES data from 2023 was available and used, with the exception of Michigan (2021) and Alaska (2016). Job growth over time, and the effect of the Covid-19 pandemic on specific job sectors may impact the usefulness of the data in the future.
- Care should be taken when normalizing or performing any kind of correlative analysis with this data, as certain columns are sums of others. For example, columns `CD01` through `CD04` sum to `CA02 + CA03`, not the total jobs columns `C000`, while other families do sum to `C000`. Refer to the technical documentation for LODES data for more details.
- Columns containing firm-specific information (`CF*`) provided in the LODES data are only nonzero for job types specific to private firms. These columns were removed in this data product.
- See §2.1.1 for information on census tract coverage.

3.1.3. Data Dictionary

With the exception of the census tract identifier `CT20`, each column in the dataset takes the form `[prefix]_[suffix]`. The prefix reproduces the corresponding column from the LODES dataset; these are described in the table below. The suffix indicates a travel time interval, in minutes, by automobile. Intervals are *incremental* rather than cumulative: each column reports the jobs reachable within that band alone, so a cumulative measure is obtained by summing across successive intervals. Bands are left-inclusive and right-exclusive. For example, the column `CE01_40-45` contains the number of jobs with earnings of $1,250/month or less reachable from tract `CT20` in 40 minutes or more and less than 45 minutes of travel time.

Prefix	Type	Description	Units
CT20	string	2020 Census tract GEOID (origin)	—
C000	integer	Total jobs	count
CA01	integer	Jobs, workers age 29 or younger	count
CA02	integer	Jobs, workers age 30 to 54	count
CA03	integer	Jobs, workers age 55 or older	count
CE01	integer	Jobs, $1,250/mo or less	count
CE02	integer	Jobs, $1,251–$3,333/mo	count
CE03	integer	Jobs, greater than $3,333/mo	count
CNS01	integer	Jobs, NAICS 11 (Agriculture, Forestry, Fishing and Hunting)	count
CNS02	integer	Jobs, NAICS 21 (Mining, Quarrying, and Oil and Gas Extraction)	count
CNS03	integer	Jobs, NAICS 22 (Utilities)	count
CNS04	integer	Jobs, NAICS 23 (Construction)	count
CNS05	integer	Jobs, NAICS 31–33 (Manufacturing)	count
CNS06	integer	Jobs, NAICS 42 (Wholesale Trade)	count
CNS07	integer	Jobs, NAICS 44–45 (Retail Trade)	count
CNS08	integer	Jobs, NAICS 48–49 (Transportation and Warehousing)	count
CNS09	integer	Jobs, NAICS 51 (Information)	count
CNS10	integer	Jobs, NAICS 52 (Finance and Insurance)	count
CNS11	integer	Jobs, NAICS 53 (Real Estate and Rental and Leasing)	count
CNS12	integer	Jobs, NAICS 54 (Professional, Scientific, and Technical Services)	count
CNS13	integer	Jobs, NAICS 55 (Management of Companies and Enterprises)	count
CNS14	integer	Jobs, NAICS 56 (Administrative and Support and Waste Management and Remediation Services)	count
CNS15	integer	Jobs, NAICS 61 (Educational Services)	count
CNS16	integer	Jobs, NAICS 62 (Health Care and Social Assistance)	count
CNS17	integer	Jobs, NAICS 71 (Arts, Entertainment, and Recreation)	count
CNS18	integer	Jobs, NAICS 72 (Accommodation and Food Services)	count
CNS19	integer	Jobs, NAICS 81 (Other Services, except Public Administration)	count
CNS20	integer	Jobs, NAICS 92 (Public Administration)	count
CR01	integer	Jobs, workers of race White Alone	count
CR02	integer	Jobs, workers of race Black or African American Alone	count
CR03	integer	Jobs, workers of race American Indian or Alaska Native Alone	count
CR04	integer	Jobs, workers of race Asian Alone	count
CR05	integer	Jobs, workers of race Native Hawaiian or Other Pacific Islander Alone	count
CR07	integer	Jobs, workers of two or more race groups	count
CT01	integer	Jobs, workers not Hispanic or Latino	count
CT02	integer	Jobs, workers Hispanic or Latino	count
CD01	integer	Jobs, workers with less than high school	count
CD02	integer	Jobs, workers with high school or equivalent, no college	count
CD03	integer	Jobs, workers with some college or Associate degree	count

Prefix	Type	Description	Units
CD04	integer	Jobs, workers with Bachelor's degree or higher	count
CS01	integer	Jobs, male workers	count
CS02	integer	Jobs, female workers	count

3.2. Distance-Weighted Job Accessibility Scores

This data product is available at the national and state level. All sub-national files include destinations accessible outside of the provided geography and are identical to the national dataset; these files are provided for convenience only.

3.2.1. Files

- **National:** `wac_S000_JT00_weighted.csv.gz`
- **State:** `[ST]_wac_S000_JT00_weighted.csv.gz`, where:
  - `[ST]` is the two-letter capitalized postal code identifier for the state (e.g. AL for Alabama).

3.2.2. Data Notes

This data applies the decay function described in §2.2.2 to the job counts described in §3.1. Note the following that may impact analyses with this data:

- Unlike the binned interval product (§3.1), this product reports a single weighted score per attribute per origin tract, rather than a value per travel-time band. Scores are not integer job counts; they are the sum of job counts across all reachable destination tracts, each discounted by $w(t_{ij})$ as described in §2.2.2.
- Scores are not directly comparable across attributes with different totals (e.g., `CE03_score` cannot be meaningfully compared to `CD04_score`) without normalizing by each attribute's total reachable count.
- Columns containing firm-specific information (`CF*`) provided in the LODES data are only nonzero for job types specific to private firms. These columns were removed in this data product.
- For most states, LODES data from 2023 was available and used, with the exception of Michigan (2021) and Alaska (2016).
- See §2.1.1 for information on census tract coverage.

3.2.3. Data Dictionary

Each column in the dataset takes the form `[prefix]_score`. The prefix reproduces the corresponding column from the LODES WAC dataset, described in the table below; `score` indicates the value is a travel-time-weighted sum rather than a raw count. For example, the column `CE01_score` contains the decay-weighted sum of jobs with earnings of $1,250/month or less, reachable from tract `CT20` across all travel time bands.

Prefix	Type	Description	Units
CT20	string	2020 Census tract GEOID (origin)	—
C000	float	Total jobs, weighted	score
CA01	float	Jobs, workers age 29 or younger, weighted	score
CA02	float	Jobs, workers age 30 to 54, weighted	score
CA03	float	Jobs, workers age 55 or older, weighted	score
CE01	float	Jobs, $1,250/mo or less, weighted	score
CE02	float	Jobs, $1,251–$3,333/mo, weighted	score
CE03	float	Jobs, greater than $3,333/mo, weighted	score
CNS01	float	Jobs, NAICS 11 (Agriculture, Forestry, Fishing and Hunting), weighted	score
CNS02	float	Jobs, NAICS 21 (Mining, Quarrying, and Oil and Gas Extraction), weighted	score
CNS03	float	Jobs, NAICS 22 (Utilities), weighted	score
CNS04	float	Jobs, NAICS 23 (Construction), weighted	score
CNS05	float	Jobs, NAICS 31–33 (Manufacturing), weighted	score
CNS06	float	Jobs, NAICS 42 (Wholesale Trade), weighted	score
CNS07	float	Jobs, NAICS 44–45 (Retail Trade), weighted	score
CNS08	float	Jobs, NAICS 48–49 (Transportation and Warehousing), weighted	score
CNS09	float	Jobs, NAICS 51 (Information), weighted	score
CNS10	float	Jobs, NAICS 52 (Finance and Insurance), weighted	score
CNS11	float	Jobs, NAICS 53 (Real Estate and Rental and Leasing), weighted	score
CNS12	float	Jobs, NAICS 54 (Professional, Scientific, and Technical Services), weighted	score
CNS13	float	Jobs, NAICS 55 (Management of Companies and Enterprises), weighted	score
CNS14	float	Jobs, NAICS 56 (Administrative and Support and Waste Management and Remediation Services), weighted	score
CNS15	float	Jobs, NAICS 61 (Educational Services), weighted	score
CNS16	float	Jobs, NAICS 62 (Health Care and Social Assistance), weighted	score
CNS17	float	Jobs, NAICS 71 (Arts, Entertainment, and Recreation), weighted	score
CNS18	float	Jobs, NAICS 72 (Accommodation and Food Services), weighted	score
CNS19	float	Jobs, NAICS 81 (Other Services, except Public Administration), weighted	score
CNS20	float	Jobs, NAICS 92 (Public Administration), weighted	score
CR01	float	Jobs, workers of race White Alone, weighted	score
CR02	float	Jobs, workers of race Black or African American Alone, weighted	score
CR03	float	Jobs, workers of race American Indian or Alaska Native Alone, weighted	score
CR04	float	Jobs, workers of race Asian Alone, weighted	score

Prefix	Type	Description	Units
CR05	float	Jobs, workers of race Native Hawaiian or Other Pacific Islander Alone, weighted	score
CR07	float	Jobs, workers of two or more race groups, weighted	score
CT01	float	Jobs, workers not Hispanic or Latino, weighted	score
CT02	float	Jobs, workers Hispanic or Latino, weighted	score
CD01	float	Jobs, workers with less than high school, weighted	score
CD02	float	Jobs, workers with high school or equivalent, no college, weighted	score
CD03	float	Jobs, workers with some college or Associate degree, weighted	score
CD04	float	Jobs, workers with Bachelor's degree or higher, weighted	score
CS01	float	Jobs, male workers, weighted	score
CS02	float	Jobs, female workers, weighted	score

3.3. Residence Area Characteristic Workers by Travel Time Interval

This data product is available at the national and state level. All sub-national files include destinations accessible outside of the provided geography and are identical to the national dataset; these files are provided for convenience only.

3.3.1. Files

- **National:** `rac_S000_JT00_binned_5min.csv.gz`
- **State:** `[ST]_rac_S000_JT00_binned_5min.csv.gz`, where:
  - `[ST]` is the two-letter capitalized postal code identifier for the state (e.g. AL for Alabama).

3.3.2. Data Notes

This data generally reproduces worker values in the LEHD/LODES Residence Area Characteristics (RAC) file provided by the United States Census. Note the following that may impact analyses with this data:

- For most states, LODES data from 2023 was available and used, with the exception of Michigan (2021) and Alaska (2016). Job growth over time, and the effect of the Covid-19 pandemic on specific job sectors may impact the usefulness of the data in the future.
- Care should be taken when normalizing or performing any kind of correlative analysis with this data, as certain columns are sums of others. For example, columns `CD01` through `CD04` sum to `CA02 + CA03`, not the total workers column `C000`, while other families do sum to `C000`. Refer to the technical documentation for LODES data for more details.
- Columns containing firm-specific information (`CF*`) provided in the LODES data are only nonzero for job types specific to private firms. These columns were removed in this data product.
- Race, ethnicity, and sex columns (`CR*`, `CT*`, `CS*`) provided in the LODES data were removed in this data product.

- Earnings (`CE*`) and industry (`CNS*`) columns describe the resident's current job, not their labor-market availability more broadly. A worker's current earnings band or industry reflects the job they presently hold, not necessarily the wage they would accept or the sector they would be willing to work in if reachable by a different employer. These columns should be interpreted as *characteristics of the resident workforce currently reachable within a given travel-time band*, not as a direct measure of labor supply at a given price point or for a given industry.
- See §2.1.1 for information on census tract coverage.

3.3.3. Data Dictionary

With the exception of the census tract identifier `CT20`, each column in the dataset takes the form `[prefix]_[suffix]`. The prefix reproduces the corresponding column from the LODES RAC dataset; these are described in the table below. The suffix indicates a travel time interval, in minutes, by automobile, measured in the reverse direction of the Workplace Area Characteristic product: bands represent travel time from workers' residence tracts to the origin tract `CT20`, i.e. the labor pool reachable by an employer located at `CT20`. Intervals are *incremental* rather than cumulative: each column reports the workers reachable within that band alone, so a cumulative measure is obtained by summing across successive intervals. Bands are left-inclusive and right-exclusive. For example, the column `CE01_40-45` contains the number of workers with earnings of $1,250/month or less reachable from tract `CT20` in 40 minutes or more and less than 45 minutes of travel time.

Prefix	Type	Description	Units
CT20	string	2020 Census tract GEOID (origin)	—
C000	integer	Total workers	count
CA01	integer	Workers age 29 or younger	count
CA02	integer	Workers age 30 to 54	count
CA03	integer	Workers age 55 or older	count
CD01	integer	Workers with less than high school	count
CD02	integer	Workers with high school or equivalent, no college	count
CD03	integer	Workers with some college or Associate degree	count
CD04	integer	Workers with Bachelor's degree or higher	count
CE01	integer	Workers with current job earnings $1,250/mo or less	count
CE02	integer	Workers with current job earnings $1,251–$3,333/mo	count
CE03	integer	Workers with current job earnings greater than $3,333/mo	count
CNS01	integer	Workers currently employed in NAICS 11 (Agriculture, Forestry, Fishing and Hunting)	count
CNS02	integer	Workers currently employed in NAICS 21 (Mining, Quarrying, and Oil and Gas Extraction)	count
CNS03	integer	Workers currently employed in NAICS 22 (Utilities)	count
CNS04	integer	Workers currently employed in NAICS 23 (Construction)	count
CNS05	integer	Workers currently employed in NAICS 31–33 (Manufacturing)	count
CNS06	integer	Workers currently employed in NAICS 42 (Wholesale Trade)	count
CNS07	integer	Workers currently employed in NAICS 44–45 (Retail Trade)	count

Prefix	Type	Description	Units
CNS08	integer	Workers currently employed in NAICS 48–49 (Transportation and Warehousing)	count
CNS09	integer	Workers currently employed in NAICS 51 (Information)	count
CNS10	integer	Workers currently employed in NAICS 52 (Finance and Insurance)	count
CNS11	integer	Workers currently employed in NAICS 53 (Real Estate and Rental and Leasing)	count
CNS12	integer	Workers currently employed in NAICS 54 (Professional, Scientific, and Technical Services)	count
CNS13	integer	Workers currently employed in NAICS 55 (Management of Companies and Enterprises)	count
CNS14	integer	Workers currently employed in NAICS 56 (Administrative and Support and Waste Management and Remediation Services)	count
CNS15	integer	Workers currently employed in NAICS 61 (Educational Services)	count
CNS16	integer	Workers currently employed in NAICS 62 (Health Care and Social Assistance)	count
CNS17	integer	Workers currently employed in NAICS 71 (Arts, Entertainment, and Recreation)	count
CNS18	integer	Workers currently employed in NAICS 72 (Accommodation and Food Services)	count
CNS19	integer	Workers currently employed in NAICS 81 (Other Services, except Public Administration)	count
CNS20	integer	Workers currently employed in NAICS 92 (Public Administration)	count

3.4. Distance-Weighted Market Access Scores (LODES)

This data product is available at the national and state level. All sub-national files include destinations accessible outside of the provided geography and are identical to the national dataset; these files are provided for convenience only.

3.4.1. Files

- **National:** `rac_S000_JT00_weighted.csv.gz`
- **State:** `[ST]_rac_S000_JT00_weighted.csv.gz`, where:
  - `[ST]` is the two-letter capitalized postal code identifier for the state (e.g. AL for Alabama).

3.4.2. Data Notes

This data applies the decay function described in §2.2.2 to the worker counts described in §3.3. Note the following that may impact analyses with this data:

- Unlike the binned interval product (§3.3), this product reports a single weighted score per attribute per origin tract, rather than a value per travel-time band. Scores are not integer worker counts; they are the sum of worker counts across all reachable destination tracts, each discounted by $w(t_{ij})$ as described in §2.2.2.
- Scores are not directly comparable across attributes with different totals (e.g., `CE03_score` cannot be meaningfully compared to `CD04_score`) without normalizing by each attribute's total reachable count.
- The same current-job interpretive caveat that applies to `CE*` and `CNS*` in the binned product (§3.3) applies here: these scores describe reachable workers by their current job's earnings or industry, not labor supply at a given wage or for a given industry.
- For most states, LODES data from 2023 was available and used, with the exception of Michigan (2021) and Alaska (2016).
- See §2.1.1 for information on census tract coverage.

3.4.3. Data Dictionary

Each column in the dataset takes the form `[prefix]_score`. The prefix reproduces the corresponding column from the LODES RAC dataset, described in the table below; `score` indicates the value is a travel-time-weighted sum rather than a raw count. For example, the column `CE01_score` contains the decay-weighted sum of workers earning $1,250/month or less, reachable from tract `CT20` across all travel time bands.

Prefix	Type	Description	Units
CT20	string	2020 Census tract GEOID (origin)	—
C000	float	Total workers, weighted	score
CA01	float	Workers age 29 or younger, weighted	score
CA02	float	Workers age 30 to 54, weighted	score
CA03	float	Workers age 55 or older, weighted	score
CD01	float	Workers with less than high school, weighted	score
CD02	float	Workers with high school or equivalent, no college, weighted	score
CD03	float	Workers with some college or Associate degree, weighted	score
CD04	float	Workers with Bachelor's degree or higher, weighted	score
CE01	float	Workers with current job earnings $1,250/mo or less, weighted	score
CE02	float	Workers with current job earnings $1,251–$3,333/mo, weighted	score
CE03	float	Workers with current job earnings greater than $3,333/mo, weighted	score
CNS01	float	Workers currently employed in NAICS 11 (Agriculture, Forestry, Fishing and Hunting), weighted	score
CNS02	float	Workers currently employed in NAICS 21 (Mining, Quarrying, and Oil and Gas Extraction), weighted	score
CNS03	float	Workers currently employed in NAICS 22 (Utilities), weighted	score
CNS04	float	Workers currently employed in NAICS 23 (Construction), weighted	score

Prefix	Type	Description	Units
CNS05	float	Workers currently employed in NAICS 31–33 (Manufacturing), weighted	score
CNS06	float	Workers currently employed in NAICS 42 (Wholesale Trade), weighted	score
CNS07	float	Workers currently employed in NAICS 44–45 (Retail Trade), weighted	score
CNS08	float	Workers currently employed in NAICS 48–49 (Transportation and Warehousing), weighted	score
CNS09	float	Workers currently employed in NAICS 51 (Information), weighted	score
CNS10	float	Workers currently employed in NAICS 52 (Finance and Insurance), weighted	score
CNS11	float	Workers currently employed in NAICS 53 (Real Estate and Rental and Leasing), weighted	score
CNS12	float	Workers currently employed in NAICS 54 (Professional, Scientific, and Technical Services), weighted	score
CNS13	float	Workers currently employed in NAICS 55 (Management of Companies and Enterprises), weighted	score
CNS14	float	Workers currently employed in NAICS 56 (Administrative and Support and Waste Management and Remediation Services), weighted	score
CNS15	float	Workers currently employed in NAICS 61 (Educational Services), weighted	score
CNS16	float	Workers currently employed in NAICS 62 (Health Care and Social Assistance), weighted	score
CNS17	float	Workers currently employed in NAICS 71 (Arts, Entertainment, and Recreation), weighted	score
CNS18	float	Workers currently employed in NAICS 72 (Accommodation and Food Services), weighted	score
CNS19	float	Workers currently employed in NAICS 81 (Other Services, except Public Administration), weighted	score
CNS20	float	Workers currently employed in NAICS 92 (Public Administration), weighted	score

3.5. ACS Population Characteristics by Five-Minute Travel Time Interval

This data product is available at the national and state level. All sub-national files include destinations accessible outside of the provided geography and are identical to the national dataset; these files are provided for convenience only.

3.5.1. Files

- **National:** `acs_population_binned_5min.csv.gz`
- **State:** `[ST]_acs_population_binned_5min.csv.gz`, where:
  - `[ST]` is the two-letter capitalized postal code identifier for the state (e.g. AL for Alabama).

3.5.2. Data Notes

This data is derived from 2023 5-Year American Community Survey (ACS) estimates, described in §2.1.3. Note the following that may impact analyses with this data:

- ACS 5-Year estimates reflect a rolling average over 2019-2023, not a single-year snapshot; they are not directly contemporaneous with the 2023 LODES vintage used elsewhere in this dataset.
- This data is provided as an independent complement to the LODES-derived worker counts (§3.3), joined by tract and travel-time band rather than merged into a single file. See §2.1.3 for guidance on comparing the two.
- Margins of error (MOE) reported by the Census Bureau for these estimates are not included in this product. Users requiring uncertainty bounds should consult the source ACS tables directly.
- `adult_pop` is derived by summing the 18-and-over age bins of Table B01001 across both sexes; it is not a single native ACS cell.
- `unemployed_pop` (Table B23025) is defined relative to the civilian labor force aged 16 and over, not relative to total population.
- Household income bands (Table B19001) are not directly comparable to LODES job-earnings bands (`CE01`–`CE03`); see §2.1.3 for the comparability caveat.
- See §2.1.1 for information on census tract coverage.

3.5.3. Data Dictionary

With the exception of the census tract identifier `CT20`, each column in the dataset takes the form `[prefix]_[suffix]`. The suffix indicates a travel time interval, in minutes, by automobile, measured in the same reverse direction as the RAC-derived worker counts (§3.3): bands represent travel time from a reachable tract to the origin tract `CT20`. Intervals are *incremental* rather than cumulative, and bands are left-inclusive and right-exclusive. For example, the column `unemployed_pop_40-45` contains the unemployed population reachable from tract `CT20` in 40 minutes or more and less than 45 minutes of travel time.

Prefix	Type	Description	Units
CT20	string	2020 Census tract GEOID (origin)	—
total_population	integer	Total population	count
adult_population	integer	Population age 18 and over	count
unemployed_pop	integer	Unemployed population, civilian labor force age 16+	count
total_households	integer	Total households (income band denominator)	count
hh_income_lt_10k	integer	Households with income less than $10,000	count
hh_income_10k_15k	integer	Households with income $10,000–$14,999	count

Prefix	Type	Description	Units
hh_income_15k_20k	integer	Households with income $15,000–$19,999	count
hh_income_20k_25k	integer	Households with income $20,000–$24,999	count
hh_income_25k_30k	integer	Households with income $25,000–$29,999	count
hh_income_30k_35k	integer	Households with income $30,000–$34,999	count
hh_income_35k_40k	integer	Households with income $35,000–$39,999	count
hh_income_40k_45k	integer	Households with income $40,000–$44,999	count
hh_income_45k_50k	integer	Households with income $45,000–$49,999	count
hh_income_50k_60k	integer	Households with income $50,000–$59,999	count
hh_income_60k_75k	integer	Households with income $60,000–$74,999	count
hh_income_75k_100k	integer	Households with income $75,000–$99,999	count
hh_income_100k_125k	integer	Households with income $100,000–$124,999	count
hh_income_125k_150k	integer	Households with income $125,000–$149,999	count
hh_income_150k_200k	integer	Households with income $150,000–$199,999	count
hh_income_200k_plus	integer	Households with income $200,000 or more	count

3.6. Distance-Weighted Market Access Scores (ACS)

This data product is available at the national and state level. All sub-national files include destinations accessible outside of the provided geography and are identical to the national dataset; these files are provided for convenience only.

3.6.1. Files

- **National:** `acs_population_weighted.csv.gz`
- **State:** `[ST]_acs_population_weighted.csv.gz`, where:
  - `[ST]` is the two-letter capitalized postal code identifier for the state (e.g. AL for Alabama).

3.6.2. Data Notes

This data applies the decay function described in §2.2.2 to the ACS population characteristics described in §3.5. Note the following that may impact analyses with this data:

- As with the LODES-based weighted product (§3.4), this product reports a single weighted score per attribute per origin tract, rather than a value per travel-time band. Scores are the sum of population counts across all reachable destination tracts, each discounted by $w(t_{ij})$.
- Scores are not directly comparable across attributes with different totals without normalizing by each attribute's total reachable count.
- ACS 5-Year estimates reflect a rolling average over 2019-2023; see §2.1.3 for vintage and reliability considerations, including the absence of propagated margins of error in this product.

3.6.3. Data Dictionary

Each column in the dataset takes the form `[prefix]_score`. The prefix reproduces the corresponding column from §3.5; `score` indicates the value is a travel-time-weighted sum rather than a raw count.

Prefix	Type	Description	Units
CT20	string	2020 Census tract GEOID (origin)	—
total_population	float	Total population, weighted	score
adult_population	float	Population age 18 and over, weighted	score
unemployed_pop	float	Unemployed population, civilian labor force age 16+, weighted	score
total_households	float	Total households, weighted	score
hh_income_lt_10k	float	Households with income less than $10,000, weighted	score
hh_income_10k_15k	float	Households with income $10,000–$14,999, weighted	score
hh_income_15k_20k	float	Households with income $15,000–$19,999, weighted	score
hh_income_20k_25k	float	Households with income $20,000–$24,999, weighted	score
hh_income_25k_30k	float	Households with income $25,000–$29,999, weighted	score
hh_income_30k_35k	float	Households with income $30,000–$34,999, weighted	score
hh_income_35k_40k	float	Households with income $35,000–$39,999, weighted	score
hh_income_40k_45k	float	Households with income $40,000–$44,999, weighted	score
hh_income_45k_50k	float	Households with income $45,000–$49,999, weighted	score
hh_income_50k_60k	float	Households with income $50,000–$59,999, weighted	score
hh_income_60k_75k	float	Households with income $60,000–$74,999, weighted	score
hh_income_75k_100k	float	Households with income $75,000–$99,999, weighted	score
hh_income_100k_125k	float	Households with income $100,000–$124,999, weighted	score
hh_income_125k_150k	float	Households with income $125,000–$149,999, weighted	score
hh_income_150k_200k	float	Households with income $150,000–$199,999, weighted	score
hh_income_200k_plus	float	Households with income $200,000 or more, weighted	score

A. Census Tract Coverage

Table A.1.: Census tract coverage in the travel-time matrix, by state, relative to the official census tract list, using only tracts that contain some amount of land (`ALAND > 0`). "Missing" = official land tracts with no OD-matrix origin. "Extra" = OD-matrix origins that are technically water tracts (`ALAND = 0`), but still received a route.

State	Official	Present	Missing	Extra	Coverage	Notes
Alabama	1,435	1,435	0	0	100.00%	
Alaska	177	157	20	0	88.70%	Minimal road network in remote areas.
Arizona	1,765	1,764	1	0	99.94%	
Arkansas	823	823	0	0	100.00%	
California	9,107	9,111	6	10	99.93%	Missing Farallon and Channel Islands, remote Kern/Imperial Co. desert.
Colorado	1,447	1,446	1	0	99.93%	
Connecticut	879	866	13	0	98.52%	2022 GEOID reclassification (fixed via crosswalk); remainder is boundary splits.
Delaware	259	260	0	1	100.00%	Water-tract extras.
Florida	5,114	5,116	2	4	99.96%	Keys/barrier-island tracts; mixed routability.
Georgia	2,791	2,791	0	0	100.00%	
Hawaii	436	451	3	18	99.31%	Extensive inter-island water tracts; most route via a ferry/bridge-adjacent point, a few small islands do not.
Idaho	456	456	0	0	100.00%	
Illinois	3,263	3,263	0	0	100.00%	
Indiana	1,693	1,693	0	0	100.00%	
Iowa	896	896	0	0	100.00%	
Kansas	829	829	0	0	100.00%	
Kentucky	1,306	1,306	0	0	100.00%	
Louisiana	1,376	1,376	0	0	100.00%	
Maine	400	400	0	0	100.00%	

Continued on next page

Table A.1 – continued from previous page

State	Official	Present	Missing	Extra	Coverage	Notes
Maryland	1,463	1,465	0	2	100.00%	Water-tract extras.
Massachusetts	1,613	1,614	0	1	100.00%	Water-tract extras.
Michigan	2,971	2,972	1	2	99.97%	Isolated Great Lakes island tracts; mixed routability.
Minnesota	1,502	1,502	0	0	100.00%	
Mississippi	875	875	0	0	100.00%	
Missouri	1,654	1,654	0	0	100.00%	
Montana	319	316	3	0	99.06%	Remote/rural tracts with no road-network route.
Nebraska	553	553	0	0	100.00%	
Nevada	775	774	1	0	99.87%	
New Hampshire	349	349	0	0	100.00%	
New Jersey	2,175	2,176	0	1	100.00%	Water-tract extras.
New Mexico	612	612	0	0	100.00%	
New York	5,393	5,400	0	7	100.00%	Water-tract extras.
North Carolina	2,660	2,665	2	7	99.92%	Outer Banks barrier-island tracts; mixed routability.
North Dakota	228	228	0	0	100.00%	
Ohio	3,162	3,162	0	0	100.00%	
Oklahoma	1,205	1,205	0	0	100.00%	
Oregon	994	996	0	2	100.00%	Water-tract extras.
Pennsylvania	3,445	3,445	0	0	100.00%	
Rhode Island	247	249	0	2	100.00%	Water-tract extras.
South Carolina	1,317	1,319	0	2	100.00%	Water-tract extras.
South Dakota	242	242	0	0	100.00%	
Tennessee	1,701	1,701	0	0	100.00%	
Texas	6,883	6,883	0	0	100.00%	
Utah	716	715	1	0	99.86%	Non-routable West Desert tract.
Vermont	193	193	0	0	100.00%	
Virginia	2,186	2,188	0	2	100.00%	Water-tract extras.
Washington	1,772	1,776	2	6	99.89%	Puget Sound / San Juan Islands tracts; mixed routability.
Washington DC	206	206	0	0	100.00%	
West Virginia	546	546	0	0	100.00%	
Wisconsin	1,526	1,528	0	2	100.00%	Water-tract extras.
Wyoming	160	159	1	0	99.38%	
Total	**84,095**	**84,107**	**57**	**69**	**99.93%**	