Commuting Distances and Times under Different Transportation Modes for U.S. Households Dataset
Dataset available at: https://github.com/peteryz/employment-od

This U.S. Department of Transportation-funded dataset is preserved in the GitHub Repository (https://github.com/), and is available at https://github.com/peteryz/employment-od

Metadata from the GitHub Repository record:
Below is the attached README, that can be found in the GitHub Repository:

Commuting Distances and Times under Different Transportation Modes for U.S. Households

Citation

U.S. Household Commuting Dataset and Transportation Fairness (under review), Hao Hao, Hai Wang, Peter Zhang.

Summary

Efficient and fair transportation planning creates opportunities and equity for jobs, health care, and education. Therefore, data consolidation for transportation systems provides basis for evidence based policies. In this work, we construct a dataset that documents home-to-job commuting time and distance information for the 100 most populated U.S. urban areas. Our dataset builds on the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Dataset, which provides origin (home) and destination (job) location information for households. For these origin-destination (OD) pairs, we derive commuting time and distance information under different travel modes, each a combination of walking, public transit, and ridesourcing. We construct data under different modes so policymakers and researchers have information about alternatives and can perform what-if analysis. Towards the end of this data sheet, we document a sample use case to illustrate this goal.

Data Description

Each instance in our dataset contains commuting time and distance information for one OD pair. We randomly sampled 100,000 OD pairs from the U.S. Census Longitudinal Employer-Household Dynamics Dataset (1,000 data points each for the 100 most populated urban areas, the list or urban areas with the corresponding indices can be found under `metadata/Urban_Area_ID.csv`. In particular, each instance contains travel time and distance information for five different travel modes, as documented in Table 1.

Table 1: Description of five travel modes. FM = first-mile, LM = last-mile. For the instances that do not have public transit available between OD pairs, only Mode Five (ridesourcing or driving throughout) travel time/distance are provided.

Travel Mode	Origin to FM Stop	FM Stop to LM Stop	LM Stop to Destination
Mode One	Walking	Public Transit	Walking
Mode Two	Walking	Public Transit	Ridesourcing
Mode Three	Ridesourcing	Public Transit	Walking
Mode Four	Ridesourcing	Public Transit	Ridesourcing
Mode Five	Ridesourcing from Origin to Destination		

The definition of columns of the dataset is provided in Table 2, with sample instances of the dataset shown in Table 3. The data files are organized as follows: `data/duration/Duration_M1M5_{Urban_Area_Index=i}.csv` and `data/distance/Distance_M1M5_{Urban_Area_Index=i}.csv` each contains the travel time and travel duration information of $1,000$ OD pairs within the ith urban area.

Table 2: Definitions of the columns. For each urban area, there are two data tables, one for travel time, and one for travel distance.

Column	Definition
OD Pair	Index of the OD pair
FM Walk	Duration (seconds)/distance (meters) of first-mile by walking
LM Walk	Duration (seconds)/distance (meters) of last-mile by walking
FM Drive	Duration (seconds)/distance (meters) of first-mile by ridesourcing
LM Drive	Duration (seconds)/distance (meters) of last-mile by ridesourcing
M1	Total duration (seconds)/distance (meters) by Mode One
M2	Total duration (seconds)/distance (meters) by Mode Two
M3	Total duration (seconds)/distance (meters) by Mode Three
M4	Total duration (seconds)/distance (meters) by Mode Four
M5	Total duration (seconds)/distance (meters) by Mode Five

Table 3: Sample instances of the travel duration data table.

OD Pair	FM Walk	LM Walk	FM Drive	LM Drive	Transit
0	810	367	190	177	4,643
1	714	134	125	46	3,639
2	643	144	172	118	480
. . .	. . .	. . .	. . .	. . .	. . .

	M1	M2	M3	M4	M5
	5,820	5,630	5,200	5,010	1,433
(continued)	4,487	4,399	3,898	3,810	1,086
	1,267	1,241	796	770	362
. . .	. . .	. . .	. . .	. . .	. . .

Data Construction

Please refer to the data sheet PDF.

Summary Statistics

The summary statistics of travel duration/distance across the five travel modes are shownin Tables 4 and 5. The summary statistics for the percentage of OD-pairs with access to public transit within each of the 100 urban areas is documented in Table 6.

Table 4: Summary statistics (min, max, mean, and standard deviation) of travel duration (in seconds) for five travel modes of 100,000 OD pairs.

Travel Mode	Min.	Max.	Mean	Stdev.
Mode One	174	21,570	4,421	2,570
Mode Two	169	21,284	4,047	2,504
Mode Three	184	21,356	3,935	2,489
Mode Four	122	21,070	3,561	2,422
Mode Five	30	6,632	1,098	581

Table 5: Summary statistics (min, max, mean, and standard deviation) of travel distance (in meters) for the five travel modes of 100,000 OD pairs.

Travel Mode	Min.	Max.	Mean	Stdev.
Mode One	428	239,520	21,519	16,683
Mode Two	481	239,549	21,769	16,720
Mode Three	448	239,520	21,691	16,708
Mode Four	526	239,549	21,941	16,745
Mode Five	137	149,516	17,770	13,527

Table 6: Summary statistics (min, max, mean, and standard deviation) of the percentage of OD pairs with access to public transit over 100 urban areas.

Min.	Max.	Mean	Stdev.
0.4%	92.8%	59.9%	20.1%

Figure 2 shows the histograms of OD travel duration and distance under five different travel modes.
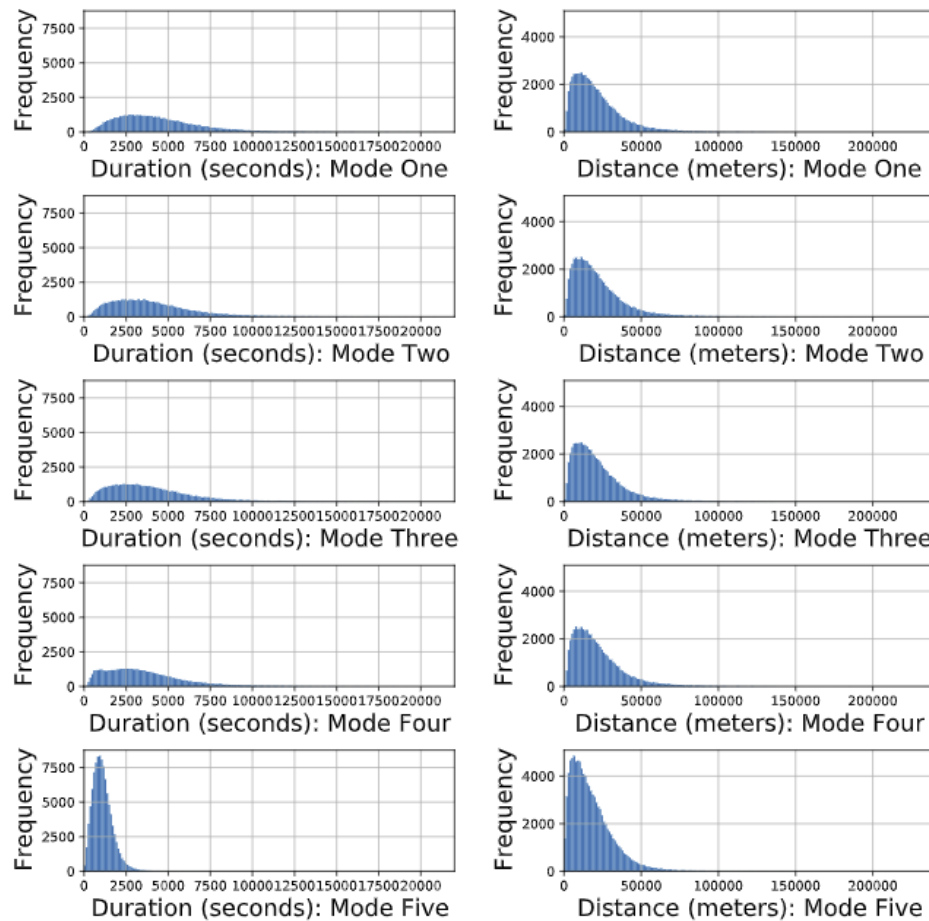
Figure 2: Histograms of OD travel duration and distance under five travel modes.

Dataset description:
This dataset contains 1 file collection, described below.

Employment-od-main.zip
- employment-od-main Folder
  - README.md
  - LICENSE
  - Commuting_Datasheet.pdf
  - metadata Folder
    - Urban_Area_ID.csv
    - .gitkeep
  - data Folder
    - .gitkeep
    - duration Folder
      - .gitkeep

- 100 .csv files with the naming structure Duration_M1M5_(0-99).csv
  - i.e. Duration_M1M5_65.csv
- distance Folder
  - .gitkeep
  - 100 .csv files with the naming structure Distance_M1M5_(0-99).csv
    - i.e. Distance_M1M5_33.csv

Recommended citation:
Hao Hao, Hai Wang, Peter Zhang. 2022. "Commuting Distances and Times under Different Transportation Modes for U.S. Households," https://github.com/peteryz/employment-od

National Transportation Library (NTL) Curation Note:
As this dataset is preserved in a repository outside U.S. DOT control, as allowed by the U.S. DOT's Public Access Plan (https://ntl.bts.gov/public-access) Section 7.4.2 Data, the NTL staff has performed *NO* additional curation actions on this dataset. NTL staff last accessed this dataset at https://github.com/peteryz/employment-od on 2022-04-26. If, in the future, you have trouble accessing this dataset at the host repository, please email NTLDataCurator@dot.gov describing your problem. NTL staff will do its best to assist you at that time.