# Documentation for Mobility Trends, Year 2 Dataset 1

The FHWA Mobility Trends dataset 1 for Year 2 is a recreation of the modeling dataset for Year 1, but collected on a county-level basis across the US. This document walks through how this dataset was collected, features of the data, and nuances of the final dataset that was put together.

The dataset features the following performance metrics and indicators:

* Vehicle Miles Traveled (VMT)
* Transportation Greenhouse Gas Emissions (GHG)
* Transit Mode Share (TMS)
* Real Gross Domestic Product (GDP)
* Population
* Unemployment Rate
* Lane Miles
* EV Charging Stations
* Unlinked Passenger Trips (UPT)
* Telework Rates

The data also includes the six GEMS model Geotypes[[1]](#footnote-2) for each county, allowing the Mobility Trends team the option to explore the data based on urban, rural, dispersed, polycentric, and monocentric characteristics.

The following portions of this document walk through each of the above listed performance metrics and indicators and how they were put together.

## VMT

Initial files provided by FHWA (Robert Rozycki):

* County VMT 2000-2008: Provides total estimates, not broken out by highway type. Seems to have significantly reduced numbers compared to 2011-2015 estimates.
* County VMT 2011-2015: Provides total estimates, not broken out by highway type. 2011 state totals almost double 2008 state totals (i.e., Sum of all FIPS code for a given state).
* County VMT 2016-2020: VMT broken out by highway type.

Upon review and SME judgement from Matthew Jacobs, the team hypothesized that prior to 2011, data is broken out to show VMT for only federal use roads, while 2011 and after includes federal and non-federal roads. We believe that 2000-2008 data represents only:

* Rural Interstate
* Rural Other Principal Arterial
* Urban Interstate
* Urban Principal Arterial
* Urban Freeway/Expressway Road types.

The following steps document how the data was compiled, paired down, and imputed to create a dataset for 2000-2020 for all counties.

1. Data for each set of FHWA provided files (2000-2008, 2011-2015, and 2016-2020) were concatenated in Excel to create one file for each year range. In cases where only Daily VMT was provided, that number was multiplied by 365 (or 366 for leap years) to create the annual VMT number.
2. Data for 2016-2020 was summed in two ways: a) the total VMT over all road categories and b) total VMT over the 2000-2008 represented roads. A calculation is then made to create the percent federal VMT is of the total VMT across each FIPS code for each year from 2016-2020. These 5 years of percentages are then averaged and multiplied by each matching FIPS code for each year of the 2011-2015 data. This produces a consistent VMT calculation across all years.
3. The data is then pivoted to create a table with columns: FIPS Code, each year from 2000-2020 with VMT as the values.
4. Counties with less than 10 data points are removed from the sample.

Note: While the District of Columbia has only one fips code (11001), VMT is captured for DC across five subdistricts. To simply the later joins, these five subdistricts are summed together to build the single D.C. fips code.

## GHG

Greenhouse Gas Emissions were pulled from the Environmental Protection Agency (EPA) National Emissions Inventory (NEI) onroad data for 2008, 2011, 2014, 2017, and 2020. For the modeling dataset, the team chose the sum of Methane and Carbon Dioxide emissions for each FIPS code and year as the GHG performance metric to be modeled on.

1. Filter each dataset down to only the Methane and Carbon Dioxide emissions. Concatenate each the region within each year’s dataset to put all the fips codes together.
2. Sum total emissions (in tons) for Methane and Carbon Dioxide.
3. Join onto County VMT using year and FIPS code.

## TMS

Transit Mode Share (TMS) is pulled from the American Community Survey 1-year PUMS estimates. The 1-year PUMS files use Public Use Microdata Areas (PUMAS) as its most granular geographic area, rather than counties. To create the TMS estimate:

1. Query the 1-year PUMS files for the JWTR variable and join a crosswalk table that connects PUMAS to FIPS codes.
2. Aggregate the sum of weighted values for all persons that indicated they used transit, and all users who responded that they commuted to work (i.e., all people who answered for the JWTR variable) for each FIPS over each year.
3. Divide the weighted value of transit users by the total weight for each FIPS code and year to create the TMS variable.

## Real Gross Domestic Product

Real GDP was pulled from the BEA at: <https://apps.bea.gov/regional/downloadzip.cfm>

BEA provides County GDP level data for 2001 – 2021. These values were easily joined to County VMT based on the year and FIPS code after being “melted” from summary table to the columns “FIPS Code”, “YEAR”, and “GDP.”

Note: Virginia has FIPS codes for both counties and cities within the state. The BEA creates its own Virginia FIPS codes which can be a combination of a county and one or two cities. Using this [link](https://www.dss.virginia.gov/family/cc_providertrain/region_lookup.pdf) from the Virginia state government website, we were able to properly relate the combined BEA territories into their individual FIPS territories. Moreover, we use employment data to distribute the BEA’s GDP estimate across the combined counties using a ratio of number of employed within a sub-territory of the BEA to the whole territory’s employment.

## Population

Population for the corresponding decades were pulled from the following sources:

* 2010s: <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html>
* 2000s: <https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-counties.html>
* 2020: <https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html>

This data did not include the FIPS code, but rather the County Name and State associated with FIPS codes. A crosswalk table from County name to FIPS code was joined onto this dataset. From there, these values were joined to County VMT.

Note: Connecticut, as of 2020, changed its FIPS designations from 8 counties to 9 ‘regional planning zones’. These new ‘regional planning zones’ do not overlap with the original counties. Thus data for Connecticut in 2020 is not added.

## Unemployment Rate

County annual averages for labor force and associated unemployment rate are found at the Bureau of Labor Statistics under the Title “Labor Force data by county, 20XX annual averages” at <https://www.bls.gov/lau/tables.htm> where XX completes the year of interest.

This data included FIPS code and year, which was easily transformed and joined to Count VMT.

## Lane Miles

Lane mile data was received from FHWA (Robert Rozycki) for the years 2000-2020. This data required some preprocessing in order to be usable on a county level. FHWA provided the following files:

* Countylnmile00-09.xlsx – this file included only the year, FIPS code, and lane miles. 2009 data is only partially included.
* County\_lane\_miles\_2010\_2019.xlsx – this file includes year, FIPS code, “Lane miles”, “Local Lane Miles”, and “Total Lane Miles”
* County\_lane\_miles\_2020.xlsx – this file includes year, FIPS code, “Lane miles”, “Local Lane Miles”, and “Total Lane Miles”

In looking through the data, the team found some inconsistent reporting from the 2010-2019 dataset, where often the ‘total lane miles’ data aggregated for a state exceeded the total lane miles reported by the state within the Highway Statistics Series (HSS). This was unique state to state, with some being equal to the ‘Total Lane Miles’ and others being equal to the ‘Lane Miles’. To compensate for this, the team:

1. Pulled the state totals reported for each year of the HSS.
2. Compared the State totals from HSS with the aggregate sum of “Lane Miles” and ‘Total Lane Miles across all counties in a state.
3. Kept the data that was closest to the reported number within HSS.

The data was then transformed and joined to County VMT.

## EV Charging Stations

Data for Electric Vehicle (EV) Charging stations was collected from the Alternative Fuels Data Center’s Developer API at <https://developer.nrel.gov/docs/transportation/alt-fuel-stations-v1/>. This API allowed the team to collect a json file of every documented electric charging station, the year it opened, and the latitude and longitude code associated with it.

Once this data was collected, the team then used the Federal Communications Commission (FCC) Geo API at [https://geo.fcc.gov/api/census/#!/block/get\_block\_find](https://geo.fcc.gov/api/census/%23!/block/get_block_find), which converted latitude and longitude codes from the charging stations to the FIPS code that it is located in. From there, based on the year and FIPS code, the data was aggregated for the sum of charging stations from 2000 to 2020.

## Transit Unlinked Passenger Trips

Unlinked Passenger Trips (UPT) was created through referencing the April 2023 [Monthly Ridership data](https://www.transit.dot.gov/sites/fta.dot.gov/files/2023-06/April%202023%20Complete%20Monthly%20Ridership%20%28with%20adjustments%20and%20estimates%29_0.xlsx) from the National Transit Database (NTD), as well as transit statistics within the American Community Survey (ACS).

First, monthly ridership data is aggregated annually and looked at on a urbanized area (UZA) level. Each UZA may encompasses multiple FIPS codes. To avoid applying the total UPT for a UZA across all FIPS (imagine giving all counties that use the New York City transportation system the same UPT value), the team instead found a measure to distribute UPT. To do this, the team pulled the number of Transit Users who commute to work from the American Community Survey, using the weighted total per PUMA, which is then crosswalked to FIPS. From there, it is aggregated across each FIPS code across each year. Then, using a UZA to FIPS crosswalk, the associated UZA is joined to the American Community Survey data, where the weighted total is aggregated per UZA. After doing this, the team divided the weighted total of number of commuters for a given FIPS code by the weighted total of number of commuters for a given UZA to determine the percent of total commuters within a UZA that are from a FIPS code. This percent is then multiplied by the Monthly Ridership Data for a UZA to determine a FIPS codes’ total annual ridership.

## Telework

Telework is pulled from the American Community Survey 1-year PUMS estimates. The 1-year PUMS files use “PUMAS” as its most granular geographic area, rather than counties. To create the TMS estimate:

1. Query the 1-year PUMS files for the JWTR variable and join a crosswalk table that connects PUMAS to FIPS codes.
2. Aggregate the sum of weighted values for all persons that indicated they teleworked, and all users who responded that they commuted to work (i.e., all people who answered for the JWTR variable) for each FIPS over each year.
3. Divide the weighted value of teleworkers by the total weight for each FIPS code and year to create the TMS variable.

1. [Geospatial Economic Multimodal Systems Modeling (GEMS) | FHWA (dot.gov)](https://highways.dot.gov/research/projects/geospatial-economic-multimodal-systems-modeling-gems) [↑](#footnote-ref-2)