

Massachusetts Vehicle Census
v.3, 2009 - 2014
Technical Documentation
October 10, 2016

Summary

The Massachusetts Vehicle Census is a catalog of information about nearly every vehicle registered in the Commonwealth of Massachusetts from 2009 to 2014. It is a valuable resource for anyone seeking to understand the factors that influence auto ownership patterns and miles driven, or to track the impact of public policies to curb driving and promote sustainable transportation. The Vehicle Census combines information from vehicle registrations, inspection records, mileage ratings, and other sources to document the ownership and mileage history of each vehicle, which is presented in two different data formats designed to protect the privacy of individual owners. Two types of data products are available: microdata (vehicle level records) and summary tables (aggregate statistics for census tracts, municipalities, etc.), each of which is accompanied by metadata tables with detailed definitions for each attribute.

All users are encouraged to carefully read this documentation for information about data definitions, the data sources and methods, and known issues and limitations.

Published by the Metropolitan Area Planning Council Data Services Department

MAPC staff: Tim Reardon, Susan Brunton, Elizabeth Irvin, Meghna Hari

In collaboration with: Paul Reim, Central Transportation Planning Staff; Ken Gillingham, PhD, Yale University School of Forestry and Environmental Studies.

In partnership with the Massachusetts Department of Transportation, Registry of Motor Vehicles

Produced with the generous support of the Barr Foundation.

On the web: vehiclecensus.mapc.org

Introduction

The Massachusetts Vehicle Census is a catalog of information about vehicle registered in the Commonwealth from 2009 to 2014. It is a valuable resource for everyone seeking to understand the key factors that influence auto ownership patterns and miles driven, and will help public agencies and communities in their efforts to build a more efficient and sustainable transportation system in the Commonwealth. The Vehicle Census combines information from vehicle registrations, inspection records, mileage ratings, and other sources to document the ownership and mileage history of each vehicle, which is presented in two different data formats designed to protect the privacy of individual owners.

The Vehicle Census is presented in two types of data products:

- **Microdata tables** include information about each individual vehicle registered in Massachusetts. Each record covers a defined period of time when the specified vehicle had a unique combination of owner, garaging address, and estimated daily mileage. Therefore, each vehicle is represented in multiple records which comprise the history of that vehicle's ownership, garaging, and mileage over time. Additional attributes include make and model of the vehicle; model year; MSRP, estimated fuel efficiency; odometer reading at the last inspection; the municipality where the vehicle was garaged; and characteristics of the vehicle's garaging address. Two approaches were used to protect the privacy of vehicle owners: records were *anonymized* by removing address information and by replacing the license plate number, owners' license number, and vehicle identification number with numeric IDs; and (depending on the level of access) records were *suppressed* by removing the geographic information from records where the vehicle's make, model, and year were unique within a given municipality at the time. These tables include approximately 34 million records for the 2009 – 2014 version of the Vehicle Census.
- **Summary Tables** compile the information about individual vehicles into summary statistics for each municipality, census tract, block group, or 250 meter grid cell. A variety of statistics are calculated for each level of geography, including total registered vehicles, average mileage per vehicle, estimated fuel consumption and emissions, and number of households. Since the microdata constitute a continuous longitudinal dataset, the summary tables are based on all the vehicle records valid on a specified day, namely the median day of each calendar quarter. As a result, there are 24+ records for each geographic area—one for each calendar quarter in the study period. To protect the privacy of vehicle owners, data were *suppressed* where there were very few households or vehicles garaged in a particular grid cell. The record count for the 2009 – 2014 Census tables is as follows:
 - Grid cell- 9.96 million records
 - Census block group- 139,580 records
 - Census tract- 41,384 records
 - Municipality- 9,828 records

In addition to the two basic types of data products, MAPC has also defined three distinct “levels of access” designed to protect the privacy of individual owners, while also making more detailed information available to researchers with a legitimate purpose and ability to ensure the security of the data.

General Public Access Files include microdata and summary tables that have been both anonymized and suppressed, making them appropriate for distribution to the general public under a general use agreement.

Complete Public Access Files include microdata and summary tables which have been thoroughly anonymized, but without data suppression, making them useful for more detailed analysis. Users will be required to specify their intended use of the data, must agree to terms of use, and will be prohibited from redistribution of bulk data.

Researcher Files include microdata tables which contain the actual VIN and characteristics of the specific parcel where the vehicle was garaged. Users interested in accessing these files must submit a research proposal and security plan; MAPC will review these materials and will grant access at its sole discretion. Approved users must sign a detailed data use agreement.

The Terms of Use for the General Public Access files are included at the end of this document. Persons interested in accessing the Complete Public Access files or Researcher files should contact MAPC.

Core Data Sources

The Vehicle Census is based on two administrative datasets maintained by the Massachusetts Registry of Motor Vehicles, a division of the Massachusetts Department of Transportation. The Automated License and Registration System (ALARS) contains the owner, license plate, and address information for each registered vehicle and is updated each time there is a 'transaction,' including transfer of title (purchase or sale), cancellation (junked vehicle), registration renewal, or change of address. A separate database contains records from safety and emissions inspections conducted as part of the state's [Vehicle Check](#) program. Vehicles are required to be inspected annually and within seven days of sale. The date of inspection, VIN, and odometer reading are all recorded on the inspection record. Comparison of odometer readings recorded during successive inspections can be used to estimate average daily mileage driven by the vehicle during the intervening period.

Data Processing

MAPC and its partners conducted extensive processing and modification of the raw RMV data to create the vehicle census in a form that could be publicly shared. The sections below describe this process.

Creating the "Registrations" Table

Registration transaction records for 2008 through June 2015 were processed and geocoded by the [Central Transportation Planning Staff](#) to create a registration history for each vehicle, which contains a distinct record for each combination of VIN, registration ID (license plate), and address. Each transaction record includes the date and nature of the transaction, the beginning date of the 2-year registration period (same as previous if renewal or new registration), and the end date of the 2-year registration period, as well as the effective address. Successive transactions for each vehicle were evaluated and records were modified to create a set of exhaustive and non-overlapping records documenting the vehicle's history. If the vehicle moved or changed hands during a 2-year registration period, the end date for that record is set to match the beginning date of the new registration with a new address or owner. Records created as a result of registration renewal, with no change in address or plate ID, were merged to create a continuous record that spans the registration renewal date.

There are a number of challenges to creating a complete registration history record from registration transaction data. Transaction data possessed by the RMV is occasionally purged, and there is a pronounced seasonality to transactions (new vehicle purchases are highest in the summer, and a wave of registrations cancellations occurs in the fall as motorcycles and other recreational vehicles are taken off the road for the winter.) As a result, frequent (at least biannual) data collection from the RMV is required for a

complete enumeration. However, MAPC and CTPS did not begin regularly scheduled and maintained acquisition of this data until late 2011. Transactions for 2008 and 2009 were available in data previously received from the RMV by CTPS. Transaction data for 2010 were not available. Since 2009 registrations span the 2010 calendar year, a partial record could be created by stitching together records from the prior and succeeding years. However, the record is missing vehicles first registered in 2010 with no subsequent record that could be used to infer a 2010 start date or garaging address.

The table below compares the vehicle census to the count of registered vehicles published by the Massachusetts Department of Revenue¹, the only other official public source of vehicle registration counts. (That source may have its own challenges, since vehicles registered and paying excise tax in multiple municipalities may be double-counted; plus it is only available through 2012.) This table shows that the Vehicle Census was 19% to 32% lower than the DOR estimate for 2009 – 2012. In 2012, the Vehicle Census was within 10% of DOR estimates, due primarily to a sharp reduction in the DOR estimates.

Year	MAVC Total (2 nd quarter)	MA DOR Total	MAVC:DOR ratio
2009	4,411,677	5,824,484	0.75
2010	4,051,664	5,884,200	0.68
2011	4,761,861	5,822,247	0.81
2012	4,848,204	5,362,872	0.90

The improved regularity of data collection from the RMV since 2012 and the narrowing gap with the RMV data allow us to state with reasonable confidence that the Vehicle Census provides a complete enumeration of registered vehicles since 2011. **However, the registration data for 2009 and 2010 are incomplete, and should be used with caution for population-level statistics such as vehicles per household or total VMT.**

Geocoding

Registration addresses were geocoded to X-Y locations using a combination of MassGIS Master Address building points² and NAVTEQ address information. Of approximately 15 million unique addresses in the registration data, 87% were geocoded. Approximately 2.7% of the records contain no street address (P.O. Box only or data entry error) and an additional 10.5% are registered to out-of-state addresses. Geocoded address points were matched to assessors parcels from the Massachusetts Land Parcel Database³, as well as census block groups and 250 meter grid cells. Neither addresses, X-Y coordinates, parcel ID, nor grid cell ID is included in either of the “Public” versions of the microdata—only municipality. Non-geocoded vehicles were inspected to determine whether they contained a valid garaging municipal ID, and this information was used for creating summary statistics. If the vehicle had a “Garage Town” which differed from the registration address (which could be the case for leased vehicles that are registered at the lessor’s address but garaged at the lessee’s home), then the registration address was not used for assigning vehicles to tracts, blocks, grid cells; these vehicles were grouped with the ungeocoded vehicles in their garaging municipality and assigned to smaller areas as described below.

Creating Mileage Estimates

¹ <http://www.mass.gov/dor/local-officials/municipal-databank-and-local-aid-unit/databank-reports-new.html>

² <http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/master-address-data.html>

³ <http://data.mapc.org/datasets>

Odometer readings from vehicle inspection records (from 2005 to the end of June 2015) and the dates of the inspections were compared to calculate the mileage driven in the intervening period and the average daily mileage. In many cases, input mistakes or other errors result in lower odometer readings being recorded in later inspections. To increase the number of valid records, MAPC used a ‘triplets’ method to extract the best estimate from three successive inspection records, with a bias toward discarding the odometer readings that would overestimate mileage (e.g., if reading at Time A was 5,000 miles, Time B was 15,000 miles, and Time C was 10,000 miles, the method would use readings A and B, and would throw out the reading at Time C.) The result was a series of mileage estimates for each VIN with an inspection date at the beginning and end, the odometer reading at the beginning inspection, and estimated daily mileage during the intervening period. Due to data entry errors or other erroneous information in the inspections data, some mileage estimates are unrealistically high (thousands of miles per day.) MAPC has chosen to leave these outliers in the data, in the event that some users may find them useful. However, this requires users to set their own rules for the removal of outliers as called for in any given application. (The summary tables, described below, do exclude any estimate >200 miles per day.)

Registrations and Estimates Table

The mileage estimates were intersected with the registration data to create a new “Registrations and Estimates” table that includes a distinct record for each combination of VIN, registration ID, address, and mileage estimate. Registration records were split where a mileage estimate begins or ends. Registration periods without a corresponding mileage estimate are retained and assigned a “false” value for the `insp_match` field. The temporal overlap between the mileage estimate and the registration record is compared to the length of the mileage estimate period as a measure of data reliability. High values for this “Percent Overlap” field mean that the vehicle had the same owner and was garaged in the same location for a large portion of the mileage estimate period; a low value means that a substantial portion of the estimated mileage may have been driven while the vehicle was owned by another person or garaged in a different location. To support temporal analysis, MAPC created Boolean fields that indicate whether a given record was valid at a specified point in time; in this case, we used the median day of each calendar quarter from January 2009 to December 2014. MAPC commonly uses the second quarter of 2014 as a reference point for summary statistics.

Vehicle Characteristics, Fuel Efficiency, and GHG Emissions

With assistance from researchers at [Yale University](#), MAPC assigned vehicle characteristics using information from a commercially-available vehicle database substantially augmented with additional values researched by Yale staff. Available vehicle attributes include make, model, model year, MSRP, curb weight, vehicle type (a composite field developed by Yale researchers), fuel type, flags for hybrid vehicles and motorcycles, and estimated fuel efficiency (miles per gallon, or MPG) based on U.S. Environmental Protection Agency economy ratings issued in 2008. Because fuel efficiency declines with vehicle age and mileage, MAPC calculated an adjusted MPG rating based on the odometer reading at the start of the mileage estimate, using efficiency decay factors from scientific and engineering literature (approximately 0.3%⁴ per year.) Adjusted MPG and estimated daily mileage were combined to generate estimated daily fuel consumption and associated greenhouse gas emissions (CO₂ equivalents) based on the GHG density of the associated fuel type. Per gallon CO₂ emissions were calculated (8.89 kgs/gallon) and factored up (CO₂ emissions/0.997) to account for other GHG gases.

⁴ Source: Modelling private car energy demand using a technological car stock model, Hannah Daly, Brian P. Ó Gallachóir; Transportation Research Part D: www.elsevier.com/locate/trd

Microdata Tables

After assigning vehicle characteristics, MAPC created a publicly-accessible version of the vehicle level microdata. The registration ID, X-Y coordinate, grid cell ID, block group and tract IDs, and zip code fields were all removed from the table, leaving the municipality as the most specific level of geographic detail. For the “General Public” version of the microdata, MAPC took additional steps to prevent individual vehicles from being identified on the basis of their make, model, and year. At any given point in time, approximately 34% of registered passenger vehicles have a make, model, and model year that is unique within the zip code where they are garaged (e.g., only one 2010 Toyota Camry in zip code 02111 during Q2 2014.) To prevent such records from being linked to a known vehicle only on the basis of their make, model, and year, MAPC suppressed the municipal ID of these passenger vehicles for the period when they would be considered unique. The vehicle information is retained, but there is no location information for the suppressed vehicle record (`muni_id=0`). MAPC also found a number of one-of-a-kind vehicles that remained unique, even after suppressing the municipality; the make, model, and year of these vehicles was suppressed to prevent individual vehicles from being identified in the dataset. This suppression was not applied to the “Complete Public” version of the microdata.

As a result of data suppression, the General Public microdata records within a given municipality cannot be considered to be a complete inventory of all registered vehicles. This table can be used to assess vehicle characteristics and mileage patterns within a municipality, but aggregate statistics must be calculated using sample weights based on the vehicle counts available in municipal summary tables, described below.

Summary Tables

To enable mapping, small area analysis, and performance metrics, the geocoded “Registrations and Estimates” records were aggregated to five geographies: 250 meter grid cells, census block group, census tracts, municipalities, and Massachusetts. The Registrations and Estimates table was sampled on the median day of each quarter and the results were aggregated for each geography. As a result, the full summary tables include multiple records for each town, tract, etc.: one for each calendar quarter in the study period.

For passenger vehicles, MAPC summarized the total number of geocoded vehicles, the number of geocoded vehicles with valid mileage estimates (between zero and 200 miles per day), and the number of geocoded vehicles with the “best” mileage estimates (more than 95% of the mileage estimate period is associated with the same owner and address.) We also calculated two different average daily mileage values: one which used all valid estimates, weighted by the percent inspection days so that the least reliable estimates have relatively less impact on the resulting average; and another which used a straight average of the “best” estimates. For each geography, MAPC also counted the number of geocoded “best” mileage estimate passenger vehicles that fell into the four quartiles of mileage estimates for all vehicles: 15 or fewer miles per day, 15-26 miles per day, 26-39 miles per day, and greater than 39 miles per day. MAPC also identified those vehicles in the top decile of all mileage estimates, which have an average daily mileage of more than 53 miles a day. (Please note: estimates greater than 53 miles per day are counted twice—in the top quartile “High Mileage vehicles” field *and* in the top decile “Very High Mileage Vehicles” field.)

As described above, 13% of vehicles could not be geocoded and some vehicles are registered to an address outside of the municipality where they are garaged. As a result, the “garage town” is the most specific geographic detail about these vehicles. MAPC allocated these vehicles to each grid cell, tract, or block group *pro rata* based on that geography’s share of the total geocoded passenger vehicles in the

municipality. The number of geocoded passenger vehicles as well as those assigned based on zip code or municipality were summed to estimate the total number of passenger vehicles garaged in that grid cell, tract, or block group. The number of commercial vehicles assigned to each grid cell, tract, or block group was also summed up, and un-geocoded commercial vehicles were allocated to grid cells, tracts, or block groups *pro rata* based on each area's share of the total geocoded commercial vehicles in the municipality. Commercial VMT is summarized in a single field which is the average of all valid mileage estimates for geocoded commercial vehicles, weighted by the percent inspection days. An estimate of total registered vehicles includes both geocoded and assigned passenger and commercial vehicles.

MAPC estimated per-household statistics by combining the count of vehicles (both geocoded and assigned) with the average mileage per vehicle and the estimated number of households (based on 2010 census or MAPC estimates described below.) Average vehicle ownership was estimated by dividing the total passenger vehicles by the estimated households. This figure was multiplied by average miles per day for passenger vehicles to produce estimates of daily per-household vehicle mileage.

MAPC estimated an "effective MPG" for each summary area based on the aggregate mileage and fuel consumption of all passenger vehicles with valid mileage estimates, and used that figure to estimate daily fuel consumption and GHG emissions associated with vehicles in the area. This "effective MPG" estimate differs from a straight average of fuel efficiency for the vehicle fleet in a given area because it accounts for different mileage patterns based on fuel efficiency (high-efficiency vehicles tend to be driven more miles than less-efficient vehicles.)

To prevent users of the General Public data from extracting information about specific vehicles in grid cells or block groups with a very small number of households and vehicles, MAPC suppressed the passenger mileage estimates and fuel consumption estimates for all geographies with fewer than two households and fewer than five geocoded passenger vehicles.

At the municipal and state level an additional attribute for estimating the change in households since 2010 was added. MAPC used building permit data collected from all Massachusetts municipalities as a proxy for change in the number of households. The number of new permitted housing units in a two-year period was summed and then divided by 8 quarters to create smooth trend lines over the two-year period. As the Vehicle Census is updated with more registration and inspection data over time, the longitudinal aspect of the dataset will allow researchers to investigate the impact of changes to the physical environment and policy change on vehicle ownership and VMT.

Smoothed Grid

At the grid cell level, vehicle per household and VMT per household statistics have some noise caused by fine-grained mismatch between household allocation and vehicle address geocoding. MAPC produced a version of the second quarter 2014 (May 15, 2014) grid cell data (smoothgrid) that was smoothed by replacing the count of households and vehicles in a cell with a 9-cell neighborhood average, and computing an average of passenger and commercial VMT per vehicle weighted by the number of vehicles in each of the 9 neighborhood cells. While the raw data are most appropriate for statistical analysis, the smoothed version may be more useful for visualization and public-facing planning efforts.

Terms of use for Massachusetts Vehicle Census - General Public Access Files

Summary Tables: Grid, Smoothed Grid, Block Group, Tract, and municipal-level (Anonymized & Suppressed)

Microdata tables (Anonymized & Suppressed)

Persons requesting access to the Massachusetts Vehicle Census General Public Access Files must provide a valid email address and must agree to the following terms of use:

Terms of Use:

Users may analyze, display, summarize, transform, or build upon Massachusetts Vehicle Census General Public Access Files (“the data”) for any legal purpose (including commercial purposes), subject to the restrictions stated below. Users are authorized to copy and redistribute the Massachusetts Vehicle Census Public Access Files in any medium or format. Any redistribution of bulk data should make reasonable efforts to include all components of the data package, including documentation, metadata, and these terms of use. All published use of the data must give appropriate credit (Massachusetts Vehicle Census, Metropolitan Area Planning Council 2015) and indicate if changes were made. Credits may be provided in any reasonable manner, but not in any way that suggests that MAPC endorses you or your use.

Users agree not to misuse the data in any way which is unlawful, illegal, fraudulent or harmful, or in connection with any unlawful, illegal, fraudulent or harmful purpose or activity. Users agree to use these datasets solely for research, statistical, commercial, or policy purposes and not investigate or identify any specific vehicle or combine the data with other information to derive or disclose personal information of any individual. User agrees to make no use of the identity of any individual discovered inadvertently, and to advise MAPC of any such discovery. Unauthorized use of the Massachusetts Vehicle Census may give rise to a legal claim or suit.

MAPC has made reasonable efforts to provide accurate data but makes no claims, no representations, and no warranties, expressed or implied, concerning the validity (expressed or implied), the reliability, or the accuracy of the Massachusetts Vehicle Census, including the implied validity of any uses of such data. Each user is responsible for determining the suitability of the data for their intended use or purpose.

Neither MAPC, MassDOT, nor their affiliates, employees, or agents shall be liable for any loss or injury caused in whole or in part by use or misuse of any information from the Massachusetts Vehicle Census.