

SUPPLEMENTAL GUIDE FOR REVIEWING DATASETS

1.0 IN THE MOMENT TRAVEL STUDY 2015

This document provides additional explanation of the In the Moment (ITM) Travel Study datasets beyond the variable labels in the data dictionaries.

Full documentation is provided in the In the Moment Travel Study technical memo.

1.1 | OVERVIEW

191 households are included in the 2015 In the Moment Travel Study datasets. These are households wherein at least one member of each household installed and activated rMove (the GPS travel data collection smartphone app). These datasets therefore include data from any household in which one or more members sent any form of data through the app, including GPS data and user-input survey data. This includes partially complete households in which some eligible members did not participate or some participating members have incomplete data (e.g., some surveys were not completed or some members started late or stopped before the assigned travel period was over). Recruited households in which no one was eligible or where no one downloaded or activated rMove are excluded from these datasets. Household inclusion criteria were broadly defined to include partially complete households in order to provide a larger sample of GPS data for analysis.

Six datasets are included:

1. Household-level data
2. Person-level data
3. Vehicle-level data
4. Trip-level data
5. Daily summary data
6. Location data

All user-input survey variables are included in the datasets, as well as passively collected GPS data. In addition, a limited set of derived variables have been computed to facilitate high-level analysis and use of the data.

1.2 | MISSING DATA

Data can be missing for two reasons:

- A user did not answer all or part of a survey – coded as -9998
- A technological error occurred with the data – coded as -9999

Data for missing user input are coded as -9998. This includes any time a question was asked and not answered. In cases where the user did not complete any part of the survey, the purpose, mode, and non-household travelers variables are coded as missing. When mode is answered but the follow-up questions are not answered, these follow-up questions are coded as missing.

Data for technological errors in the passive GPS collection software, such as missing timestamps, are coded as -9999. There are two cases of missing survey completion timestamps due to a technological error, and four cases of missing auto details (parking details, driver details) due to an error.

Continuous variables such as distance were not coded with this value, and were instead left as NULL when missing, in order to avoid interfering with statistical calculations of continuous variables. There is one case where the trip distance and speed could not be derived due to an error in the timestamps of the location points. In order to preserve the ordering sequence of locations, data for the locations on this trip were not edited.

A NULL (or blank) value exists in the dataset when a response was not required for a given question. For example, respondents who live alone were not asked if they traveled with any other household members. These NULL values are not errors, but are “intentionally missing” data.

1.3 | HOUSEHOLD-LEVEL DATASET

The household-level dataset has 191 rows, one row per household.

Unique identifier: Password

Example = gtxxgqhe

Households were given the identical password in 2015 as the password they had in the 2014 study.

Household info survey duration (hhinfodur)

The household information/recruit survey duration is calculated as the difference between the timestamp recorded for the last survey page and the timestamp for the first survey page. Use care when interpreting survey duration as it is possible the respondent left their web browser open for period(s) of time or may have paused during the survey and returned to finish at a later time.

First and last travel dates (traveldate_start, traveldate_end)

The travel period lasted seven sequential days. The assigned start and end date of this period are indicated by these variables. Some participants may have downloaded the smartphone app after the first travel date, and some participants may have stopped using or uninstalled the app before the last travel date.

Number of trips by HH during travel week and number of trips by HH during travel day 1 (hhnumtrips, hhnumtrips_day1)

The number of trips passively collected from or reported by each household over the 7-day travel period is indicated by “hhnumtrips”. This includes trips with GPS data (passively collected), trips derived for non-participating household members, trips derived when there was a distance gap in a person’s trip record, and trips reported missing in the daily summary surveys. The number of trips passively collected from or reported by each household on the first travel date is indicated by “hhnumtrips_day1”. It is important to note that this does not necessarily represent a complete picture of each household’s travel, as it does not include trips made separately by non-participating household members and some participating members may only be partially complete. Therefore, true household trip rates cannot be estimated.

1.4 | PERSON-LEVEL DATASET

These are the person-level variables, for which there is one row per person for a total of 505 persons, including adults and children, both GPS participating and non-participating.

Unique identifier: PersonID

Example = gtxxgqhe.01, where each PersonID is the household's password (gtxxgqhe) with a sequential person number appended for each household member (01, 02, etc.) Take to note, that although the household's password is both 2014 and 2015, it is not necessarily true that household members have the same ID in 2014 and 2015. This is due to household changes that required a new recruitment survey in 2015.

GPS participant (participant)

If a household member participated using rMove, they are designated with a 1 in the "participant" variable. Members were eligible and invited to participate if they were over age 16 and owned a smartphone compatible with rMove, but members are not considered GPS participants unless they installed and activated rMove and provided some data through the app (passively collected GPS data and/or survey responses).

Person completion status (person_status)

Person completion status is derived based on the percentage of trip surveys each person answered, the number of days with GPS data, and whether all daily surveys are answered or not. Status level 4 is considered a "fully complete" person who has answered all trip and daily surveys and recorded GPS data for four or more days. Status level 3 includes persons who have completed at least two-thirds (66%) of surveys and collected GPS data for four or more days. Status level 2 includes persons with less than two-thirds of surveys completed or fewer than four days of data. Status 1 level includes persons who downloaded rMove but provided no GPS data.

Type of household respondent (resptype)

The type of household respondent identifies a "primary" person (the person who completed the HH Info Survey), "other" household adults (any household members aged 18 and up), and children (any household members under the age of 18).

Number of trips for travel week, number of trips for travel day 1 (numtrips, numtrips_day1)

The number of trips passively collected or reported by the respondent is the count of total trip records (reported missing and recorded) associated with each person's personID for all seven travel days, while the number of trips for travel day 1 accounts for trips (reported missing and recorded) on the first travel day. Again, it should be noted that this does not necessarily represent a complete picture of an individual's travel for a day or a week if that person did not complete the study every day.

Number of daily surveys completed (num_dailies)

GPS participants were prompted to complete a "daily summary" survey at the end of each travel day to report any missed trips or to briefly describe why they did not make any trips (if that was the case). This variable indicates the count of daily surveys submitted per GPS participant.

Number of missing trips reported (numtrips_missed)

The number of missing trips reported is the sum of all missing trips reported in the daily surveys per GPS participant.

1.5 | VEHICLE-LEVEL DATASET

The vehicle-level dataset has 389 rows – one row per vehicle reported in the household info survey.

Unique identifier: vehicleID

Example = gtxxgqhe.v01 where each vehicleID is the household's password (gtxxgqhe) with a unique ID appended for each household vehicle (v01, v02, etc.)

1.6 | DAILY SUMMARY DATASET

The daily dataset has 1,981 rows – one row per GPS participant per day of the travel study. For days when no GPS data was collected and no daily survey submitted, no data is included in that row. For days when GPS data was collected but no daily survey submitted, there is a count of number of trips collected, but no other data for that row.

Day number (daynum)

This is the number of the assigned travel date (where the first assigned travel date is 1 and last assigned travel date is 7). Unique identifiers for each person-day can be created by combining personID with the day number.

Daily survey was submitted (daily_survey_complete)

If the row contains information from a daily survey that was submitted, this variable contains a “1”. In some cases, users took trips but did not take a daily survey. For these rows, this variable contains a “0”, and other daily survey information is left blank or set to 0.

Number of total trips, Number of trips passively collected on travel date (numtrips, numtrips_recorded)

The number of recorded trips is the derived number of trips passively collected by rMove on the travel date. This number plus the number of missed trips reported equals the total number of trips on the travel day (numtrips). Days on which no GPS data was passively recorded have a “0” for this variable. Data from days where a daily survey was not submitted (daily_survey_complete = 0) and 0 trips were collected should not be used when calculating an average trip count per day, as dates with 0 recorded trips and no daily survey submitted could indicate says that a person did not have rMove installed.

1.7 | TRIP-LEVEL DATASET

These are the trip-level variables, from the trips collected by the smartphone app during the travel period. The trip-level dataset has 13,689 trips – one row per passively collected trip. This includes trips collected directly from GPS participants' devices (whether the trip survey was completed or not) as well as trips “copied” for non-participants who were identified in the travel party for participants' trips. Trips that started on the last travel date (May 11, 2015) but ended on the day after the last travel date (May 12, 2015) are included in the dataset, but trips that started before the first travel date are not included.

Unique identifier: tripID

Example = gtxxgqhe.01.001 where each tripID is the household's password (gtxxgqhe) with a sequential person number (01, 02, etc.) appended for each household member and a sequential trip number (001, 002, etc.) appended for each unique trip made by an individual participant. Trip numbers are sequential for an individual across all days in the travel period.

Origin and destination latitude/longitude (olat, olon, dlat, dlon)

The origin and destination coordinates are the coordinates of the first and last locations for which data was collected on each trip.

Origin trip purpose and destination trip purpose (o_purpose, d_purpose)

Respondents report the destination trip purpose in the trip surveys. The origin purpose is derived from the destination purpose of the previous trip, except for first trip in in the travel period, or where a trip falls after a trip with no survey data. For surveys where the destination was not reported, that trip's destination purpose and the origin purpose of the subsequent trip are coded -9998. Origin purpose is coded as NULL when the trip is the first trip of the travel period or occurs after a derived "teleport" trip (see below.)

Start and end time (time_start, time_end)

Timestamp on the first point and last point collected on the trip. These timestamps correspond to the timestamps in the locations table for a given personID.

Trip duration in minutes (trip_duration)

Travel time is derived as the difference between the start and end timestamp of the trip.

Trip distance in miles (trip_distance)

The trip distance is derived from the sum of all distances between points included in the trip.

Implied speed in miles per hour (implied_speed_mph)

Implied speed is equal to the derived trip distance over the trip duration, in units of miles per hour. When the duration rounded to 0 minutes, the speed could not be derived and is instead NULL in these cases.

Travel mode (mode)

Respondents were allowed to select more than one mode. For trips where more than one mode was selected, the "mode" variable shows a value of 99, and the variables "mode1", "mode2", and "mode3" are populated with the modes chosen. No trips were reported to have more than three chosen modes. Follow-up questions were asked depending on the type of mode selected. Because the mode "Vanpool" was listed under both "vehicle" and "transit" modes, some Vanpool mode trips include auto details and some include transit details. In future versions of rMove, these will be coded as two different modes.

Error reported (error)

Users reported when a trip had an error, such as the wrong start or end time, or where more than one stop was made on a trip. Some reported errors were corrected during post-processing, but in some cases, users reported an error incorrectly, or an error could not be fixed in processing. For example, some trips were

reported as “not moving” although it is clear from the GPS trace that a trip was made, and these were retained in the dataset. Error flags were removed when an error was corrected during data processing.

Member 1-8 was on trip (member1 through member8)

Users could report whether household members traveled with them on a trip. When another household member is reported on a trip, their personID is listed in the “member” variables. When a person travels alone but has other members in their household, the “member” field is populated for the person who took the trip. When a person is the only member of their household, all of the member variables are NULL.

Survey revision flags (revised_at, revised_count)

After a survey was completed, users were able to change information for that survey within 24 hours of completing the trip survey. Revised_count indicates the number of times a survey was revised, and revised_at shows the timestamp at which the survey was last revised.

Trip was split from another trip (flag_split)

During data processing, when a trip was determined to have more than one stop (such as a drop-off trip or an exercise “loop” trip that started and ended in the same place), the trip was typically “split” by the analyst at the point where an obvious stop was made or at the furthest point of a loop trip. Because of the nature of the data processing, there are some instances where an extra stop was reported (error = 1), or the path is such that a stop might have been made (for example, a driving trip that appeared to take a detour before arriving at the destination), but the trip was not split due to uncertainty of where the stop occurred or possible analyst oversight. The resulting trips from a split trip are denoted in the “flag_split” variable.

Trip was merged from multiple trips (flag_merge)

Similarly, when it was either reported by the respondent (error = 2) or made obvious by the trip path that a trip was actually part of the previous or next trip (for example, someone waited at a stop light and rMove determined that the stop light was the end of the trip), the two adjacent trips were joined together in data processing. When it was not obvious whether a trip should be joined, typically because of a long dwell time between trips (time that elapsed between the end of one trip and beginning of the next), no action was taken and the trips were left as two separate trips, even if reported as a “part of another trip” error. Additionally, in cases where a phone lost service or battery and stopped data collection mid-trip, there was no path for the remainder of the trip, and the trip was not joined to another trip. Merged trips are denoted in the “flag_merge” variable.

Trip was derived when ≥ 250 m gap existed in person’s trip record (flag_teleport)

When the end of one trip and the start of the next trip are greater than or equal to 250 meters apart, this was categorized as a “teleport” case, meaning an unrecorded trip happened between the two recorded trips and caused the person to appear as though they “teleported” from one location to another. When this happened, a trip was derived using the starting and ending coordinates matching those of the trip before and after. These “teleport” trips have only two coordinates, the origin and destination. The start and end timestamps match the time the previous trip ended and the time the next trip started. This can lead to very long derived trip durations, because a “teleport” sometimes occurs in between two trips that occurred several hours apart.

Trip was derived for non-GPS participant (nonresp_trip)

Trips were derived for non-GPS participants when reported by other GPS-participating household members to be on a trip. In instances where more than one household member reported another member on a trip with overlapping travel times, the trip with the earlier start time is chosen, so that overlapping trips are eliminated. The corresponding original tripID is stored in “nonresp_originaltripid.

Person ID of original trip (trip_personID)

This is the ID of the participant who originally took the trip, and is included so that derived trips can be joined to the locations on “personID” in the location dataset. Joining the trip and location tables is explained in more detail below.

1.8 | LOCATION-LEVEL DATASET

The location-level dataset has 843,920 rows – one row per location collected from 276 GPS participants for whom data was passively collected.

Unique person-location identifiers

Unique identifiers can be created by combining personID with locnum. The unique locations are not linked to specific tripIDs because there is not necessarily a one-to-one relationship between points and trips. This is because in cases where a trip was split, the resulting two trips share the point (location and timestamp) at which the trip was split. This means that the ending point of one trip is sometimes the exact same point as the beginning point of the next trip.

Location number (locnum)

The order in which the point was collected for each person is denoted by “locnum”. Locnum is unique for each person and is sequential per personID, acting as an index of location rows.

Time collected (collected_at)

The timestamp at which the point was collected by rMove. Collection timestamps can be the same between two points for the same person record, and in this case, the points should be ordered by “locnum”.

Location accuracy in meters (accuracy)

For points collected on an Android device: Accuracy is the measurement in meters of the radius in which there is 68% confidence that the point lies. In other words, if you draw a circle centered at this location's latitude and longitude, and with a radius equal to the accuracy, then there is a 68% probability that the true location is inside the circle.

For points collected on an iOS device: Accuracy is the measurement in meters of the radius in which the location likely lies, though Apple currently provides no documentation on the level of confidence of this radius.

Heading in degrees (heading)

Heading is the direction in degrees (due north being 0) collected by rMove. This is the direction in which the smartphone was traveling when the location point was collected.

1.9 | USING THE TRIP AND LOCATION DATASETS

Trip traces (or routes/paths) are made up of several sequential collected location “points”, which were collected at varying intervals (typically every 5 seconds during a trip.) This section describes how to interpret the trip and location dataset together in order to view trip traces.

Selecting location points for trips

In order to find all the points within a given trip, users should first identify the personID for the participant who made the trip (trip_personID). Then search for the rows in the location data with this personID where the “collected_at” timestamp field falls between the trip’s “trip_start” and “trip_end” times, including the start and end times themselves. To list the locations in sequential order, sort the location rows by “locnum” (the location index by personID).

Order of locations within a trip

Some points will share a timestamp, due to technological limitations of rMove that are currently being assessed and upgraded. When the “collected_at” timestamp is the same for more than one row within a person’s locations, the location index “locnum” should be used to determine the ordering of the points, as this indicates which point was collected first.