Codebook

# BostonWalks study

Date: 26.09.2024

Contact MIT

- Andres Sevtsuk (a.sevtsuk@mit.edu)

Contact Northeastern

- Ryan Wang (q.wang@northeastern.edu)

# Contents

# 1. Overview

The data constitutes of two main parts. The GPS tracking comes from Motiontag (part 1), the socio-demographic and attitudinal indicators come from the Qualtrics survey (part 2).

Motiontag provides three different tables. The first "user_stats" keep a record of all users that registered with the Catch-my-Day app. It includes mostly technical indicators, but importantly, it links the central Motiontag identifier (the identifier used across the Motiontag tables) with the central Qualtrics identifier (the identifier used in the Qualtrics data. Both identifiers have the following form:

**Central Motiontag user identifier: ab332347-6616-4c0c-a646-6794434d8649**

**Central Qualtrics user identifier: BW18098**

The second table "storylines" provides the detected events for each user. Every line represents one event (track or stay) from one user. The events can be sorted chronologically to result in a continuous timeline of events. The geometries need to be decoded to obtain the sequence of lat/lon pairs. Note that these points are not fully equivalent to the raw GPS measurements (third table), as the Motiontag pipeline includes a high-level filtering of outliers.

The third table includes all waypoints of all users, and can be queried using the Motiontag identifier and the start & end datetime of a certain storyline event.

# 2. Motiontag (part 1)

## 2.1. User statistics

Relevant file: user_stats.csv

Generation: as provided through Motiontag backend.

Columns:

- **id: central Motiontag user identifier**
- email: email address used by the user to register in the Catch-my-Day app
- created_at: datetime of user registrations [UTC]
- activated_at: datetime of tracking activation [UTC]
- operating_system: operating system of user phone
- os_version: version of operating system of user phone
- handset: type identifier of user phone
- first_tracked_at: first recorded track [UTC]
- latest_tracked_at: last recorded track [UTC]
- friendly_user: internal Motiontag indicator / not relevant
- latest_confirmed_finished_at: last confirmation (validation of event) [UTC]
- confirmed_tracks_count: number of confirmed/validated tracks
- unconfirmed_tracks_count: number of not confirmed/validated tracks
- tracks_count: total number of recorded tracks
- unconfirmed_days_count: number of days with full confirmation/validation
- active_days_count: number of days where tracking was active

3

- days_in_range_count: total number of days between first and last recorded track
- **registration_code: central Qualtrics user identifier**
- confirmed_days_count: number of days without full confirmation/validation
- inactive_days_count: number of days where tracking was not active
- latest_data_transmission_at: latest data transmission of user phone [UTC]
- latest_storyline_finished_at: latest recorded storyline item (track/stay) [UTC]

## 2.2.  Storylines

Relevant file: storylines_MAPC.csv

Generation: as provided through Motiontag backend. Filtered for MAPC region and enriched with plausibility metrics.

Plausible speed & distance: based on following (debatable) threshold values:

```
"""
plausibility levels average speed
Car:          150kph > avg_speed              including Carsharingmobility, Taxi/Uber, Motorbike, Scooter, Escooter, Boat
Bike:          35kph > avg_speed              including Bikesharing
Ebike:         50kph > avg_speed
Walk:          12kph > avg_speed
Etrottinett:   25kph > avg_speed
Train:        200kph > avg_speed              including RegionalTrain
Subway:       100kph > avg_speed              including Tram, LightRail
Bus:          120kph > avg_speed
Cablecar:      50kph > avg_speed              including Ferry
Ski:          100kph > avg_speed
Airplane:            avg_speed > 200kph

plausibility levels distance
Bike:          50km > distance               including Bikesharing, Ebike
Walk:          20km > distance               including Etrottinett, Ski
Cablecar:      10km > distance
Train:              distance > 0.25km         including LightRail, RegionalTrain, Tram, Subway
Airplane:           distance > 25km
"""
```

Avg. walking speed threshold of 12kph. Measurements get rather noisy for very short trips due to in-accuracies of GPS signal (e.g. urban canyon effect), see plot below.
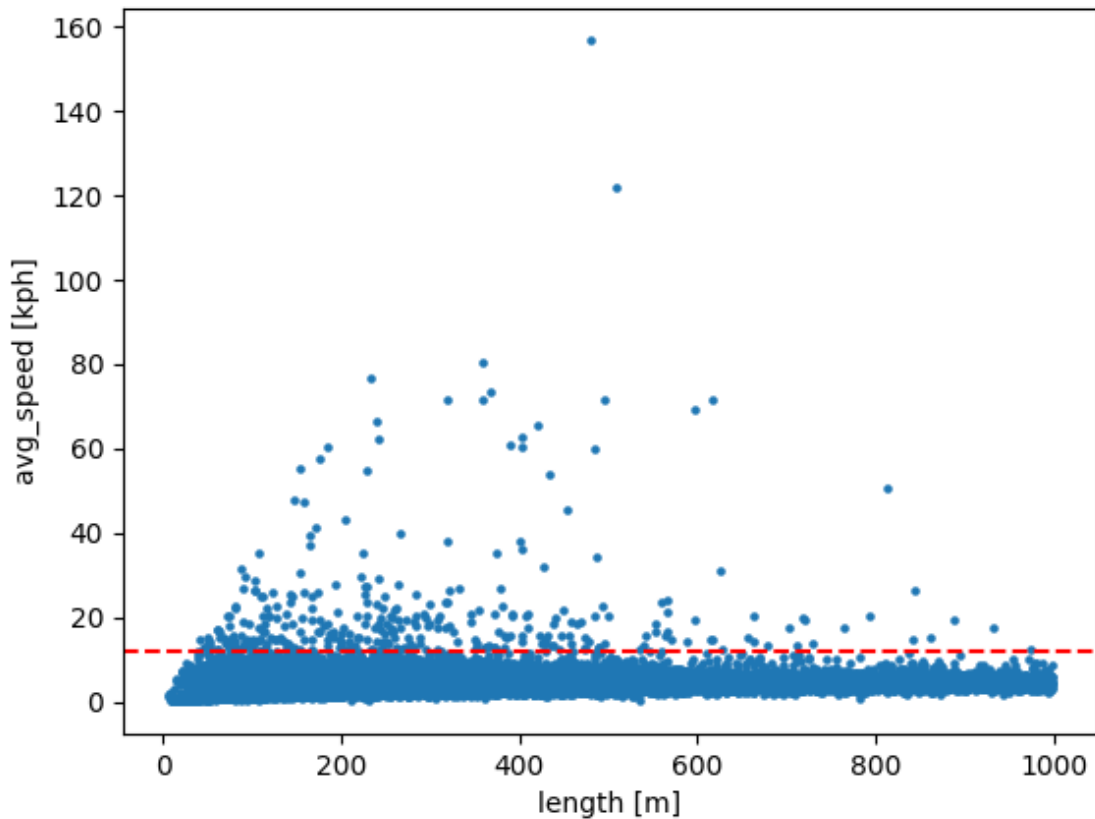
Columns (Motiontag backend):

- id: event (track/stay) identifier
- **user_id: central Motiontag user identifier**
- type: type of event (track = in motion, stay = stationary)
- started_at: datetime of event start [UTC] (to get local time convert started_at using started_at_timezone)
- started_at_timezone: timezone of datetime of event start
- finished_at: datetime of event end [UTC]
- finished_at_timezone: timezone of datetime of event end
- length: lengths in meters
- detected_mode: initially detected mode (see "detectable" in appendix)
- mode: final mode (after potential correction by user)
- purpose: final purpose (automated detection of home & work location after first labeling by user)
- geometry: WKB HEX encoded geometry
- confirmed_at: datetime of event confirmation [UTC]
- started_on: date of event start [UTC]
- misdetected_completely: indicator from Motiontag pipeline, not relevant
- merged: indicator if event was merged with previous event
- created_at: datetime of event object creation in backend [UTC]
- updated_at: datetime of event update (e.g. change of mode/purpose) [UTC]

Columns (post-processing, only for tracks):

- duration_min: event duration in minutes (min)
- avg_speed_kph: length/duration in kilometers per hours (kph)
- plausible_speed: 1 if plausible speed, 0 otherwise

- plausible_distance: 1 if plausible distance, 0 otherwise
- on_water: 1 if any part of trip within waterbodies, 0 otherwise
- circle_route: 1 if straight distance between origin and destination is <100m, 0 otherwise
- detour_ratio: division between length and straight distance between origin and destination
- seg_XXX: number of segments longer than XXX m
- plausible_segments: 1 if no segments longer than 500m, 0 otherwise
- within_rail_buffer_100m: 1 if full trip within 100m buffer to MBTA commuter or rapid transit rail lines, 0 otherwise
- walkrun: 2 if 12 > avg_speed_kmh >= 5, 1 if avg_speed_kmh < 5, 0 otherwise

## 2.3.  Waypoints

Relevant file: waypoints.csv

Generation: as provided through Motiontag backend.

Columns:

- **user_id: central Motiontag user identifier**
- tracked_at: datetime of signal [UTC]
- latitude: WSG84
- longitude: WSG84
- accuracy: if available, phone-specific accuracy metric
- speed: if available, speed in meter per second
- altitude: if available, altitude in meters
- course: if available, heading in degrees

Note: to get raw gps waypoints of a certain storyline event, query using user_id, started_at, finished_at from storyline items.

# 3. Qualtrics (part 2)

Relevant file: qualtrics.csv

Generation: As provided through Qualtrics survey responses.

Columns:

- start_date: datetime of survey start [MTZ]
- end_date: datetime of survey end [MTZ]
- progress: survey progress (0% - 100%)
- duration: time to complete the survey (seconds)
- finished: is the survey completed ('True' or 'False')
- recorded_date: datetime of survey record [MTZ]
- response_id: Qualtrics entry identifier
- user_language: language used in survey (EN: English, ES: Spanish, PT-BR: Portuguese)

- age: age in years
- gender: gender (Male, Female or Other)
- ethnicity: ethnicity (Hispanic or Not Hispanic)
- race: race
- zip: residential zip code
- indiv_income: pre-tax annual household income in US dollars
- occup: occupation
- educ: highest level of education
- hhsize: adults (18 years or older) living in household
- nchild_hh: children (under 18 years) living in household
- height: height (ft + in)
- weight: weight (lbs)
- chronic_dis: user's chronic diseases
- freq_phys_act: frequency of engaging in physical activities

Following questions with the "walking_" tag are answered in the format of: Agreed, Somewhat agree, Indifferent, Somewhat disagree, disagree.

- walking_leisure: I like to walk for leisure
- walking_short: I like walking, but only for short distances
- walking_over_transport: When possible, I prefer walking over other transport modes
- walking_tired: Walking makes me tired
- walking_relaxed: Walking makes me relaxed/comfortable
- walking_unsafe: Walking is dangerous/unsafe
- walking_healthy: Walking has health benefits
- walking_env: Walking is good for the environment
- walking_walkable: I would like to live in a walkable neighborhood with easy access to transport, shops, amenities, etc.

- pollution_sources: What do you think are the main sources of air pollution in your daily life, select all those that apply: Traffic and transport, Industrial activities, Household heating and cooking, Agricultural activities, Natural sources (e.g. wildfires, dust storm)
- pollution_sources_other: What do you think are the main sources of air pollution in your daily life, other (user provided)
- pollution_measures: Do you take measures to reduce air pollution, select all those that apply: Using sustainable modes of transport (e.g. public transport, cycling and walking), Uisng energy efficient applicanes, Recycling and reducing waste, Save energy at home (e.g. lower heating), Plant trees or maintain green spaces.
- pollution_measures_other: Do you take measures to reduce air pollution, other (user provided)
- aq_rate: how do you rate the air quality in your neighborhood? (Good to Very Poor)
- aq_inform_imp: how important is it for you to be informed about air quality?

- participation: are you interested to participate in the tracking part of the study? ('Yes or 'No')
- walk_quarter_miles: are you able to walk a 1/4 mile (400 meters) without pain? ('Yes or 'No')
- smartphone: do you use a smartphone? ('Yes or 'No')
- professional_driver: are you a professional driver? ('Yes or 'No')

- uber: do you drive for Uber/Lyft or similar services? ('Yes' or 'No')
- uber_driving_time: time spent for driving Uber/Lyft or similar services (in percentage)
- uber_driving_distance: miles driven per week for Uber/Lyft or similar services
- **id: central Qualtrics user identifier**

Note: All answers are given in the string format, integer/categorical conversions should be made depending on the usage purpose.

# 4. Appendix (Modes & Purposes)

| Name | Key | Id | SVG | PNG | PNG white 3x | Color | Carbon emission per km | Public transport | detectable? |
|---|---|---|---|---|---|---|---|---|---|
| Airplane | airplane | Mode::Airplane | | | | #e9b100 | 196 | | ✓ |
| Bicycle | bicycle | Mode::Bicycle | | | | #90a72f | 0 | | ✓ |
| Bikesharing | bikesharing | Mode::Bikesharing | | | | #90a72f | 0 | | |
| Boat | boat | Mode::Boat | | | | #6e6cad | 139 | | |
| Bus | bus | Mode::Bus | | | | #5175ae | 74 | ✓ | ✓ |
| Cable car | cablecar | Mode::Cablecar | | | | #2ea8c5 | 44 | ✓ | ✓ |
| Car | car | Mode::Car | | | | #d57f0e | 139 | | ✓ |
| E-Bike | ebicycle | Mode::Ebicycle | | | | #90a72f | 16 | | |
| Electric kick scooter | etrottinett | Mode::Etrottinett | | | | #ee6767 | 16 | | |
| Electric scooter | escooter | Mode::Escooter | | | | #4a90e2 | 16 | | |
| Ferry | ferry | Mode::Ferry | | | | #6e6cad | 139 | ✓ | ✓ |
| Mobility Car-Sharing | carsharing_mobility | Mode::CarsharingMobility | | | | #be1924 | 139 | | |
| Motorbike | motorbike | Mode::Motorbike | | | | #46926c | 139 | | |
| Other | other | Mode::Other | | | | #35ad9c | 0 | | ✓ |
| Pikmi | ridepooling_pikmi | Mode::RidepoolingPikmi | | | | #0096cc | 104 | | |

| Name | Key | Id | SVG | PNG | PNG white 3x | Color | Carbon emission per km | Public transport | detectable? |
|---|---|---|---|---|---|---|---|---|---|
| Rapid transit railway | light_rail | Mode::LightRail | | | | #35ad9c | 74 | ✓ | ✓ |
| Regional train | regional_train | Mode::RegionalTrain | | | | #cb4288 | 72 | ✓ | ✓ |
| Scooter | scooter | Mode::Scooter | | | | #6b7fe5 | 72 | | |
| Ski | ski | Mode::Ski | | | | #f3a536 | 0 | | |
| Subway | subway | Mode::Subway | | | | #3e8eb6 | 74 | ✓ | ✓ |
| Taxi/Uber | taxi_uber | Mode::TaxiUber | | | | #f0d722 | 139 | | |
| Train | train | Mode::Train | | | | #d6393a | 43 | ✓ | ✓ |
| Tramway | tram | Mode::Tram | | | | #2ea8c5 | 74 | ✓ | ✓ |
| Walking | walk | Mode::Walk | | | | #debb00 | 0 | | ✓ |

# Purposes

| name | key | id | svg | png_white_3x | marker_png | color |
|------|-----|-----|-----|--------------|------------|-------|
| At home | home | home | | | | #4f4f4f |
| Drop off / pick up someone | assistance | assistance | | | | #4f4f4f |
| Eat out | eat | eat | | | | #4f4f4f |
| Education | study | study | | | | #4f4f4f |
| Errand | errand | errand | | | | #4f4f4f |
| Family & Friends | family_friends | family_friends | | | | #4f4f4f |
| Leisure | leisure | leisure | | | | #4f4f4f |
| Medical visit | medical_visit | medical_visit | | | | #4f4f4f |
| Other | other | other | | | | #4f4f4f |
| Shopping | shopping | shopping | | | | #4f4f4f |
| Sports | sport | sport | | | | #4f4f4f |
| Unknown | unknown | unknown | | | | #4f4f4f |
| Waiting | wait | wait | | | | #4f4f4f |
| Working | work | work | | | | #4f4f4f |