

# FitBit Aggregation Script README

by Alex Ramsdell

Last Updated: Tuesday, Jan 31st, 2017

## Script Name / Location

`~/fbtrack/scripts/fb_aggregate/fb_aggregate.py`

## Important File locations and format information:

- JSON FitBit data source directory: `~/fbtrack/data/raw` .
- Aggregated and converted data files: `~/fbtrack/data/csv` .
- Aggregation output filename format: `<aggregation_type>_<timestamp>.csv` , e.g. `hourly_2017-01-31.csv` .
- Month Day Subject Calculator File: `~/fbtrack/scripts/fb_aggregate/lib/month_day_subject_calculator.xlsx` .

## How To Use The Script

At a command-line prompt, change directory to the `fb_aggregate` directory inside the fbtrack scripts directory, e.g.:

```
fbtrack@uw ~ $ cd ~/fbtrack/scripts/fb_aggregate
```

Run the `fb_aggregate.py` script. It should take a few minutes to complete. There is a progress meter which tells you the percent complete. When the aggregation script completes, it will tell you the location of the newly created aggregation files.

```
fbtrack@uw ~/fbtrack/scripts/fb_aggregate $ ./fb_aggregate.py
0% of source JSON files have been processed.
5% of source JSON files have been processed.
10% of source JSON files have been processed.
...
output files:
/Users/fbtrack/fbtrack/data/csv/2017-01-31/hourly_2017-01-31.csv
/Users/fbtrack/fbtrack/data/csv/2017-01-31/sleep_2017-01-31.csv
/Users/fbtrack/fbtrack/data/csv/2017-01-31/daily_2017-01-31.csv
```

**Note:** Everything on the command-line from the `$` to the left is part of the terminal prompt and is there for informational purposes, so you shouldn't type that if you are running this yourself.

## How The convert.py Script Works

1) It goes through each JSON data file in `~/fbtrack/data/raw` and creates a list of unique subject ids by parsing the filenames for the embedded id. The filenames are in the format `<subject_id>_<capture_date>.json` .

2) A time-stamped directory is created in the `~/fbtrack/data/csv` directory for these aggregated files, e.g. `~/fbtrack/data/csv/2017-01-31` .

3) It then goes through each subject's files, parses them into Python objects and aggregates the data into hourly, daily and sleep formats.

4) Those aggregations are converted to CSV and appended to a time-stamped file for that aggregation type, which is inside a time-stamped directory for that day, e.g. the directory/file might be: `~/fbtrack/data/csv/2017-01-31/sleep_2017-01-31.csv`.

## Caveats

1) Re-running the Script: if you for some reason need to run the `fb_aggregate.py` script again on the same day, the script will append the new results to the pre-existing files, so it's best to remove them before you run it again.

2) **The Month Day Subject Calculator spreadsheet has been renamed to exclude spaces in the title.** I've also reconverted it to .xlsx because the file's metadata appeared corrupted in the form I received via email (an .xlsx file is really a zip file of XML documents). I checked to make sure the content matched after I did this.

## Output Notes

- CSV output fields containing a `.` indicates that the data isn't available.

## FitBit Field Mapping to SEA Variables

### Note:

- Any SEA variable below in brackets indicates that the field isn't properly available through Fitbit's data. I've made my best guess as to how to calculate the value using other provided data, but it should be reviewed.

## General

subjectid: The subject's assigned SEA id.

msid: N/A. Used for the band acquisition but not for the FitBit acquisition.

date: Date of captured FitBit data.

day: Subject's canonical day of SEA study.

month: Subject's canonical month of SEA study for a given subject.

## Sleep-Specific

dur:

- Meaning: Duration, total duration of sleep reading
- FitBit field used: 'duration'
- Unit: hours, converted from FitBit's millisecond reading

sdur:

- Meaning: Duration, total duration of time asleep
- FitBit field used: 'minutesAsleep'
- Unit: hours, converted from FitBit's minute-level reading

sleff:

- Meaning: Sleep efficiency
- FitBit field used: 'efficiency'
- Unit: percentage value of time asleep / total time in bed

sdurrl:

- Meaning: Duration of restless sleep

- FitBit field used: 'restless Duration'
- Unit: hours, converted from FitBit reading in minutes

[ sdurrf ]:

- Meaning: Duration of restful sleep in hours
- FitBit field used: 'minutesAsleep' - 'restlessDuration'
- Unit: Converted from calculation using minutes Asleep - sdurrl. See note in 'build\_sleep.py'.

numwu: Number of Awakenings. Uses FitBit 'awakeCount' field.

- Meaning:
- FitBit field used:
- Unit:

awdur: Duration of Time awake during main sleep in hours. Converted from FitBit readings in minutes.

- Meaning:
- FitBit field used:
- Unit:

fadur: Fall Asleep Duration. Uses FitBit's minutesToFallAsleep Reading.

- Meaning:
- FitBit field used:
- Unit:

fat:

- Meaning: Fall Asleep Time.
- FitBit field used: 'minutesToFallAsleep'
- Unit: Measured in hours, converted from FitBit minute-level readings.

[ wutime ]:

- Meaning: Wake-up Time.

Not explicitly recorded by FitBit, but see note in 'build\_sleep.py'.

- FitBit field used: Not available. Could be calculated by other available fields?
- Unit: datetime

[ hrrest ]:

- Meaning: Heart Rate at Rest
- FitBit field used: N/A. FitBit does not provide this data for sleep readings.
- Unit: N/A.

## Activity

avghr: Average heart rate. Calculated on the hourly and daily level from minute-level FitBit readings.

- Meaning:
- FitBit field used:
- Unit:

pkhr: Peak heart rate. Calculated on the hourly and daily level from minute-level FitBit readings.

- Meaning:
- FitBit field used:
- Unit:

lwahr: Lowest heart rate. Calculated on the hourly and daily level from minute-level FitBit readings.

- Meaning:
- FitBit field used:
- Unit:

step:

- Meaning: Steps
- FitBit field used: 'activities-steps' and 'activities-steps-intraday'
- Unit: Calculated on the hourly and daily level from minute-level FitBit readings.

cal:

- Meaning: Calories
- FitBit field used: 'activities-calories' and 'activities-calories-intraday'
- Unit: Integer count for daily reading, Float count for intraday readings on the minute-level.

dtof:

- Meaning: Total distance on foot.

- FitBitField used: 'activities-distance' and 'activities-distance-intraday'
- Unit: miles calculated on the hourly and daily level from minute-level and day-level readings.

dt:

- Meaning: Total distance on foot.
- FitBit field used: 'activities-distance' and 'activities-distance-intraday'
- Unit: Miles, calculated on the hourly and daily level from minute-level and day-level readings.

hrrest:

- Meaning: Resting Heart Rate
- FitBit Field used 'activities-heart.restingHeartRate'.
- Unit: Integer count of beats per minute.

hract:

- Meaning: Hours Active
- FitBit field used: 'duration' in general 'activities'
- Unit: hours.