Using Smartphone Sensors to Encourage Physical Activity

Paul Schimek

My Background

UCLA Urban Planning
Luskin School of Public Affairs

DUSP MIT

tBf The Boston Foundation
INNOVATION. INFORMATION. IMPACT.

TranSystems
EXPERIENCE | Transportation

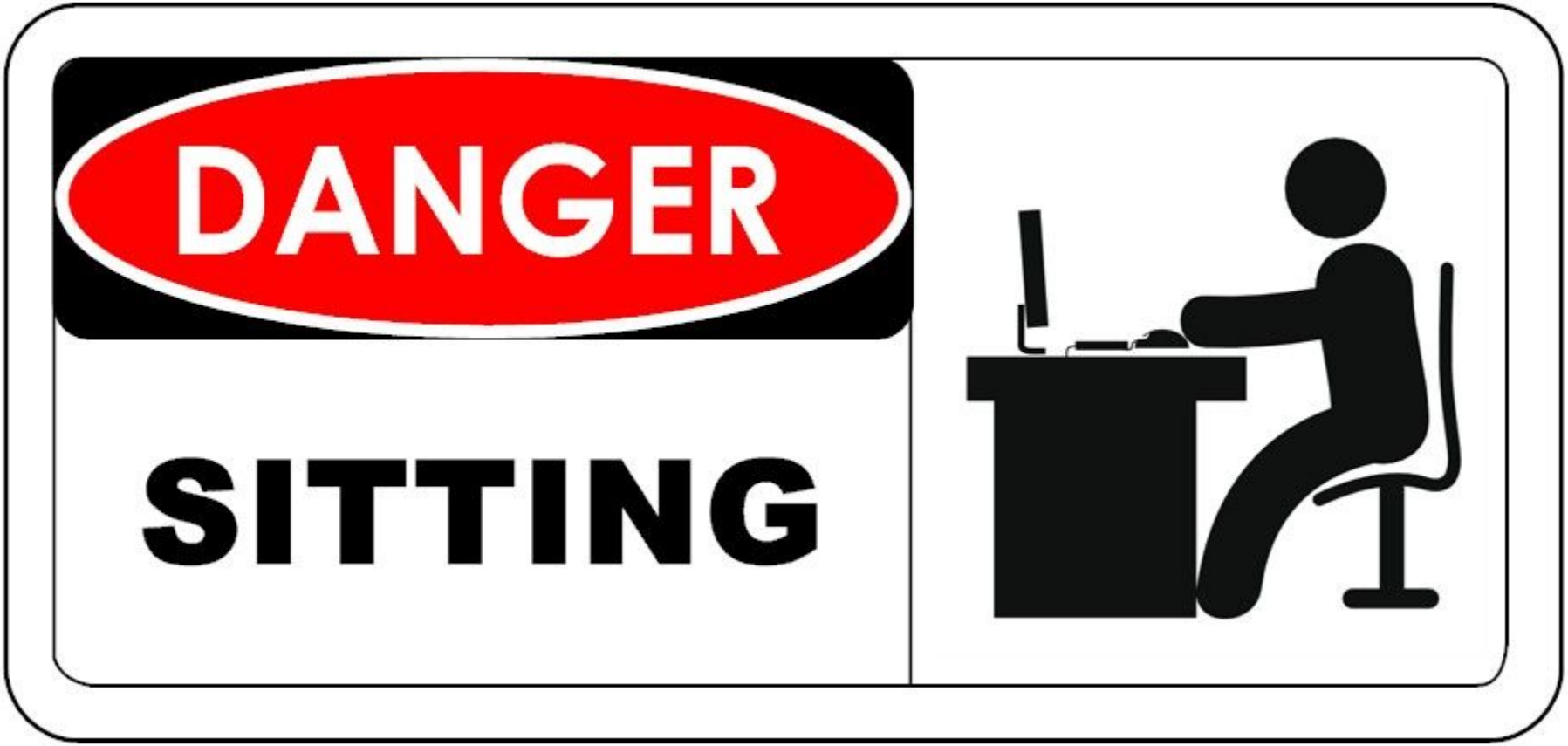
BTD
BOSTON TRANSPORTATION DEPARTMENT

Booz | Allen | Hamilton

Harvard University
Graduate School of Design

DEPARTMENT OF TRANSPORTATION
UNITED STATES OF AMERICA

One of the most dangerous activities



Annals of Internal Medicine®

ORIGINAL RESEARCH | 3 OCTOBER 2017

Patterns of Sedentary Behavior and Mortality in U.S. Middle-Aged and Older Adults: A National Cohort Study

Keith M. Diaz, PhD; Virginia J. Howard, PhD; Brent Hutto, MSPH; Natalie Colabianchi, PhD; John E. Vena, PhD; Monika M. Safford, MD; Steven N. Blair, PED; Steven P. Hooker, PhD

Article, Author, and Disclosure Information

Abstract

Background: Excessive sedentary time is ubiquitous in Western societies. Previous studies have relied on self-reporting to evaluate the total volume of sedentary time as a prognostic risk factor for mortality and have not examined whether the manner in which sedentary time is accrued (in short or long bouts) carries prognostic relevance.

SCIENCE

Americans Are Sitting More and We Have Computers to Blame

Research sets out to quantify how much time people spend sitting down; teens are more sedentary than grown-ups

By Brianna Abbott

Updated April 23, 2019 11:22 a.m. ET

Americans are sitting more than ever, and the habit starts young, according to a large study that found computer use in particular has contributed to a more sedentary lifestyle over the past two decades.

Across a range of age groups, average sitting time increased roughly an hour a day from 2007 to 2016, according to the study, which was published Tuesday in JAMA, the Journal of the American Medical Association.

Average

Goal: Physical Activity (PA) Tracker

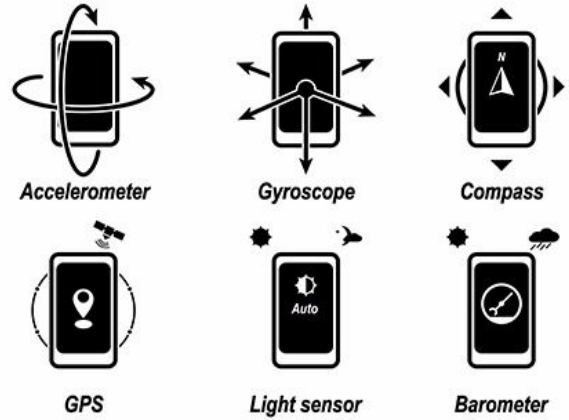
Use smartphone sensors

Sedentary vs. moving

Intensity of PA

Prompt after long period without PA

Measure and track the amount of PA



Constraints

- Recognize PA for new users
- Any phone, any position
- Preserve battery life
- Provide user feedback in real time

RealWorld Data

15 subjects

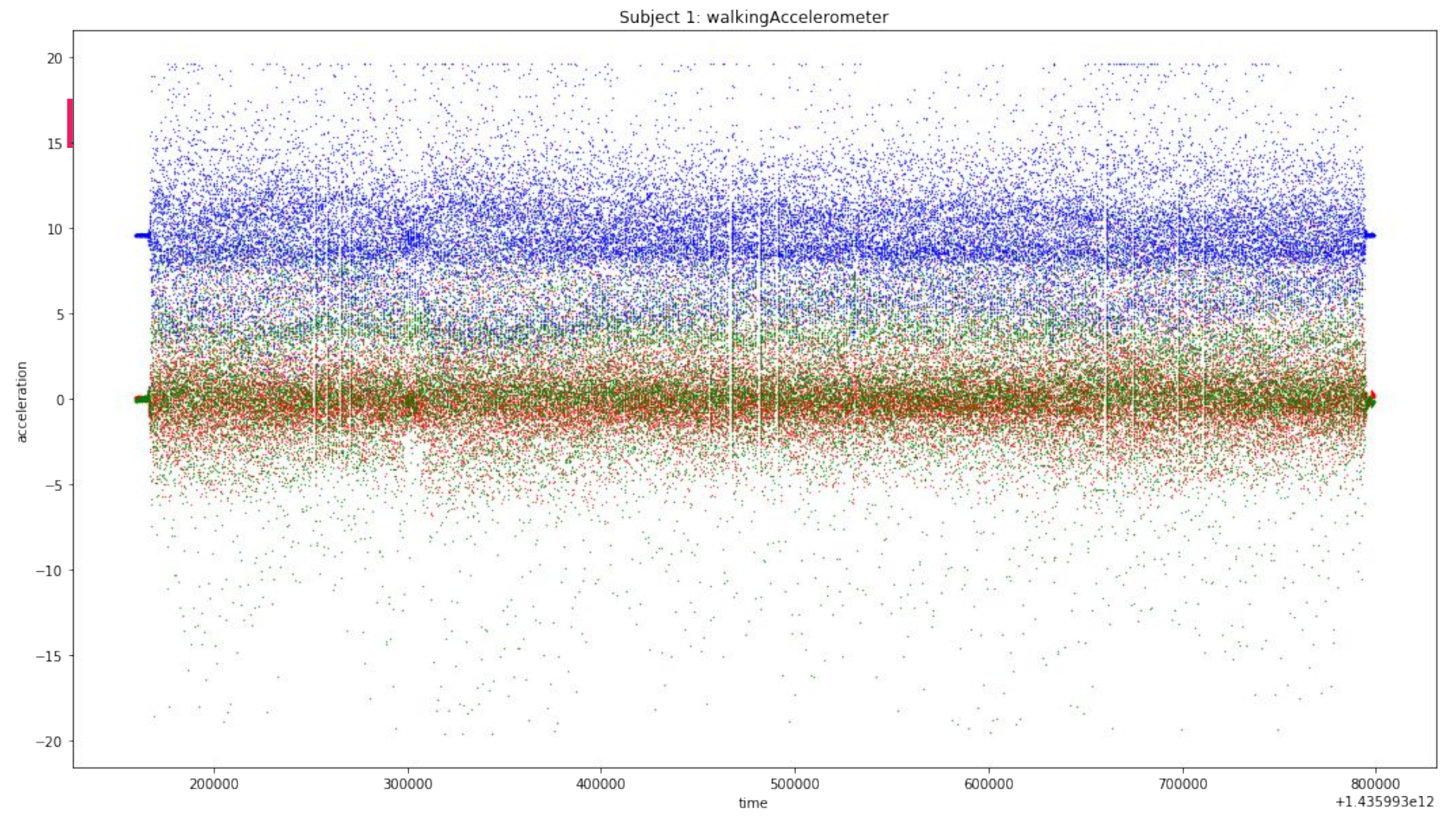
8 activities

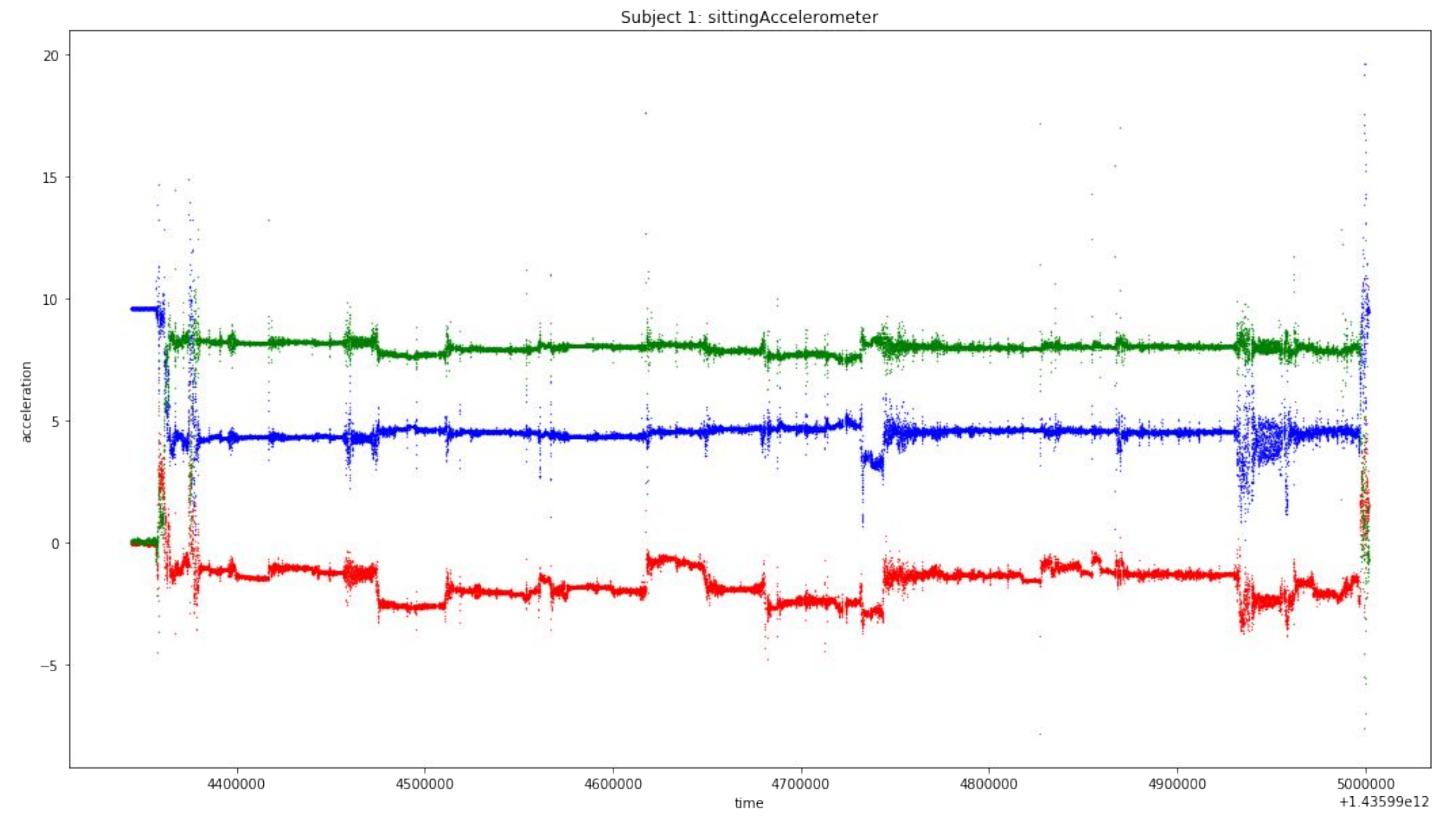
Going down stairs, going up stairs, jumping, lying, standing, sitting, running, and walking

7 devices

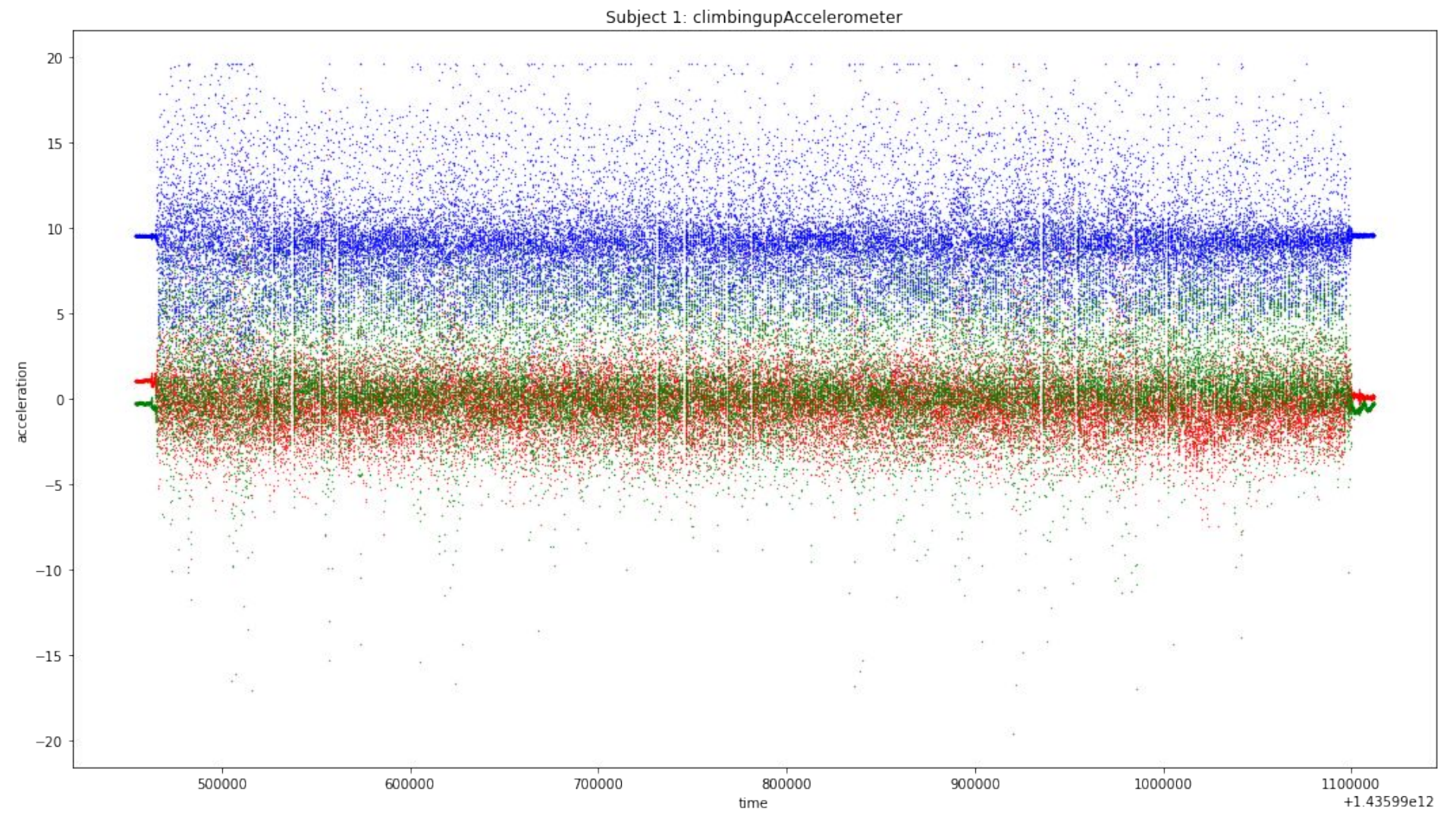
6 sensors

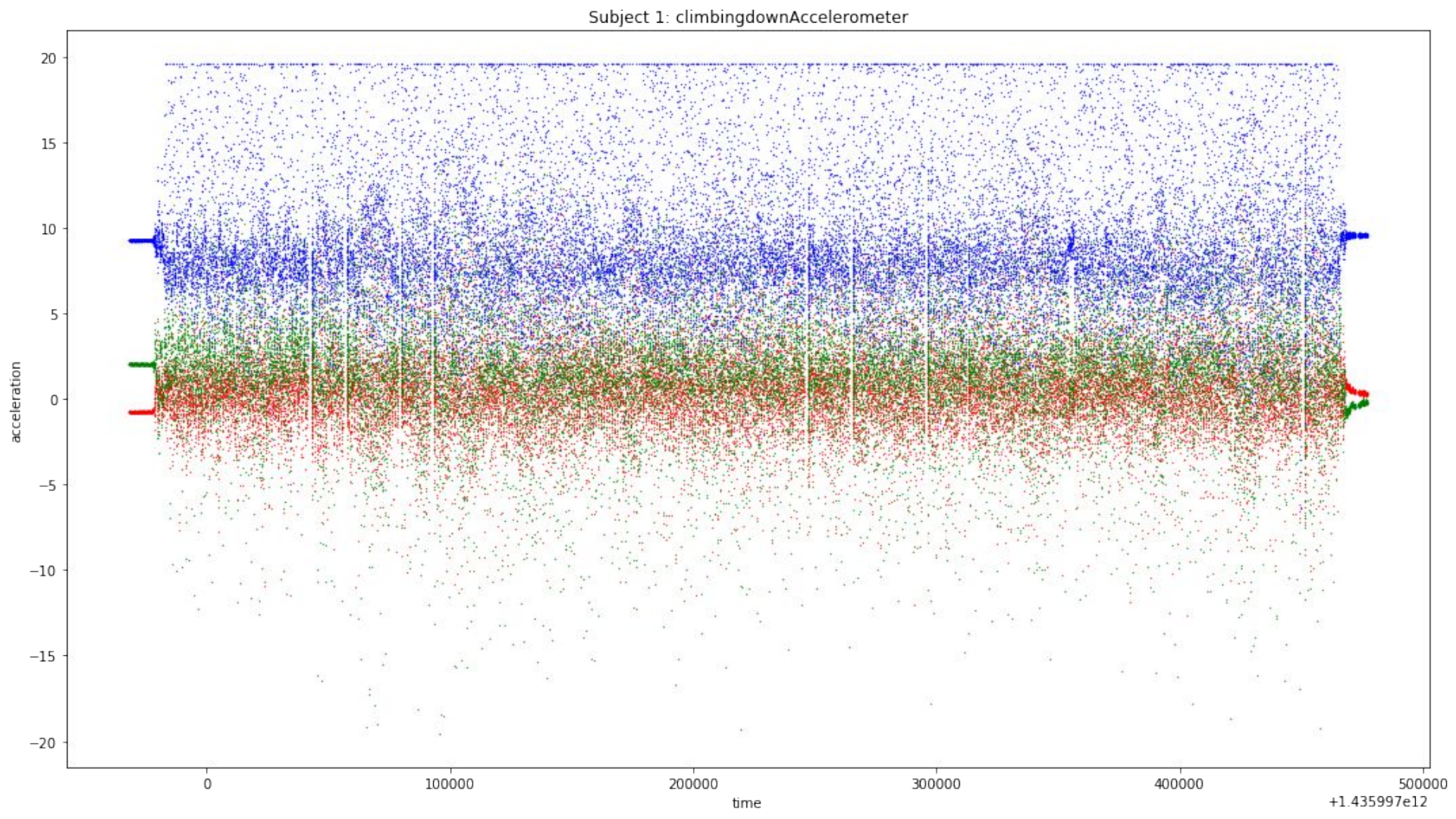
acceleration, gyroscope, GPS, light, magnetic field, and sound level

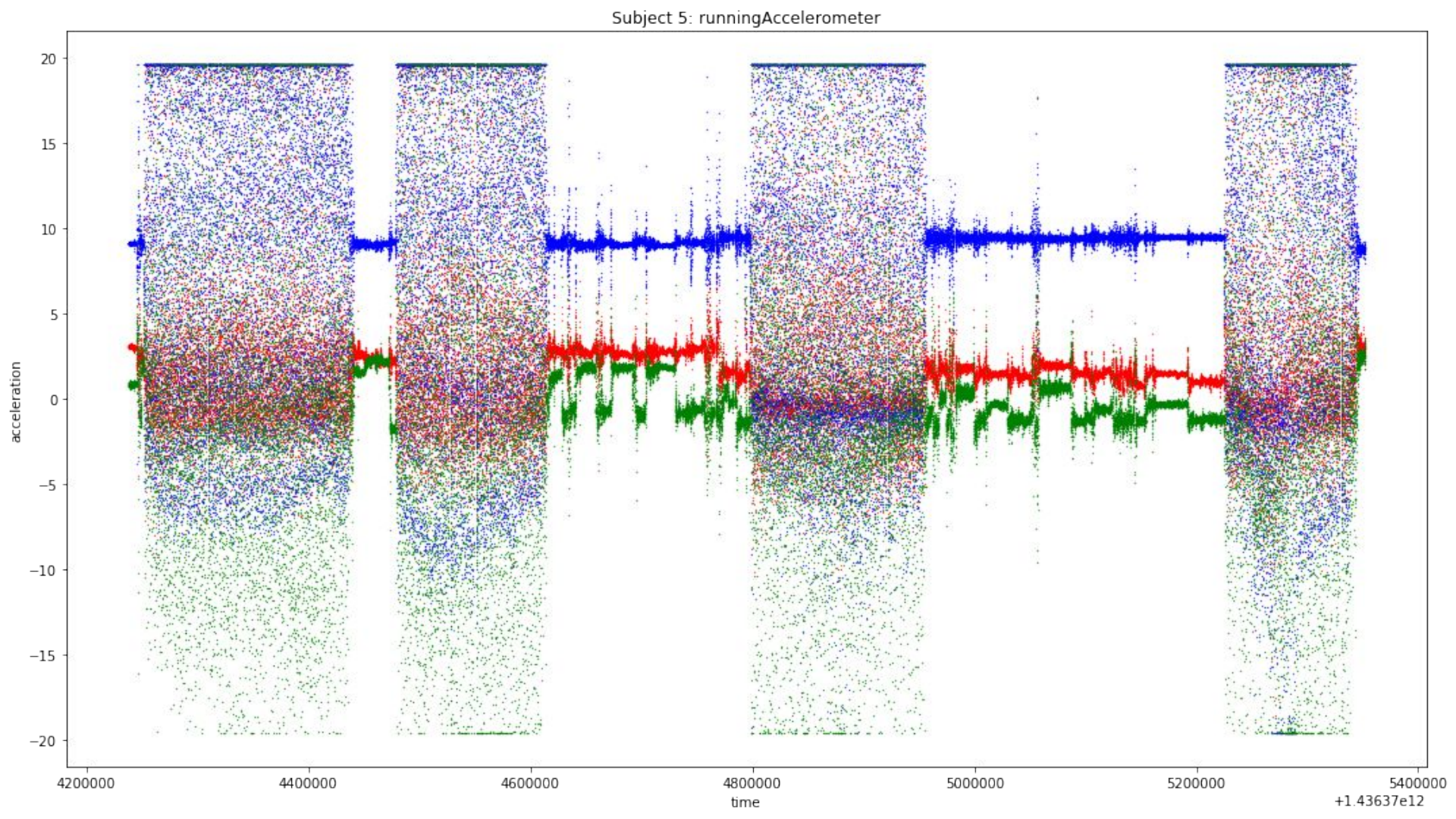

Subject 1: walkingAccelerometer
acceleration
time
+1.435993e12
20
15
10
5
0
−5
−10
−15
−20
200000
300000
400000
500000
600000
700000
800000

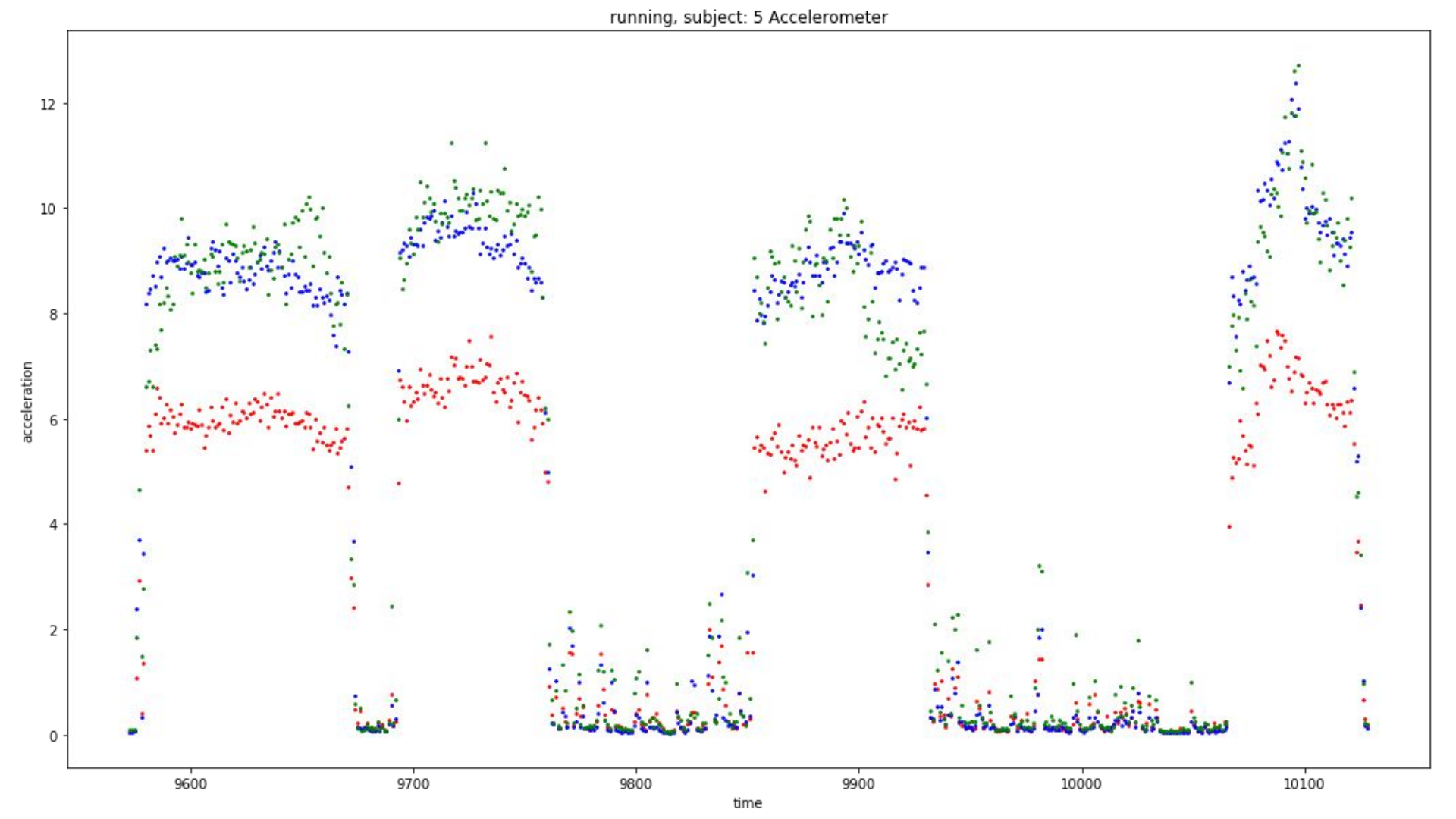

Subject 1: sittingAccelerometer
20
15
10
5
0
−5
acceleration
4400000
4500000
4600000
4700000
4800000
4900000
5000000
time
+1.43599e12


Subject 1: runningAccelerometer
acceleration
time
+1.43599e12
20
15
10
5
0
−5
−10
−15
−20
5500000
5600000
5700000
5800000
5900000
6000000
6100000


Subject 1: climbingupAccelerometer
acceleration
time
+1.43599e12
20
15
10
5
0
−5
−10
−15
−20
500000
600000
700000
800000
900000
1000000
1100000


Subject 1: climbingdownAccelerometer
acceleration
time
+1.435997e12
20
15
10
5
0
−5
−10
−15
−20
0
100000
200000
300000
400000
500000


Subject 5: runningAccelerometer
acceleration
20
15
10
5
0
−5
−10
−15
−20
4200000
4400000
4600000
4800000
5000000
5200000
5400000
time
+1.43637e12


running, subject: 5 Accelerometer
acceleration
time
0
2
4
6
8
10
12
9600
9700
9800
9900
10000
10100

Data Preparation

- Samples of 100 observations (2 sec)
  - Mean, Standard Deviation, Range
- Filter mislabeled data

Recode into 3 PA Classes

Sedentary (standing, sitting, lying)

Light-Moderate PA (walking, going up stairs, going down stairs)

Vigorous PA (running, jumping)

Train / Validate / Holdout

Modeling Data - 10 subjects

- Training - 7 subjects
- Validation - 3 subjects

Holdout Data - 5 subjects

- Not seen by model

Modeling Approach

18 features: 2 sensors (acc, gyr) x 3 dimensions (x,y,z) x 3 metrics (mean, sd, range)

Logistic regression accuracy:

- Training: 100%
- Validation: 97%

Also tried: K-Nearest Neighbors, Naive Bayes, various types of Decision Trees, Support Vector Machine, XGBoost

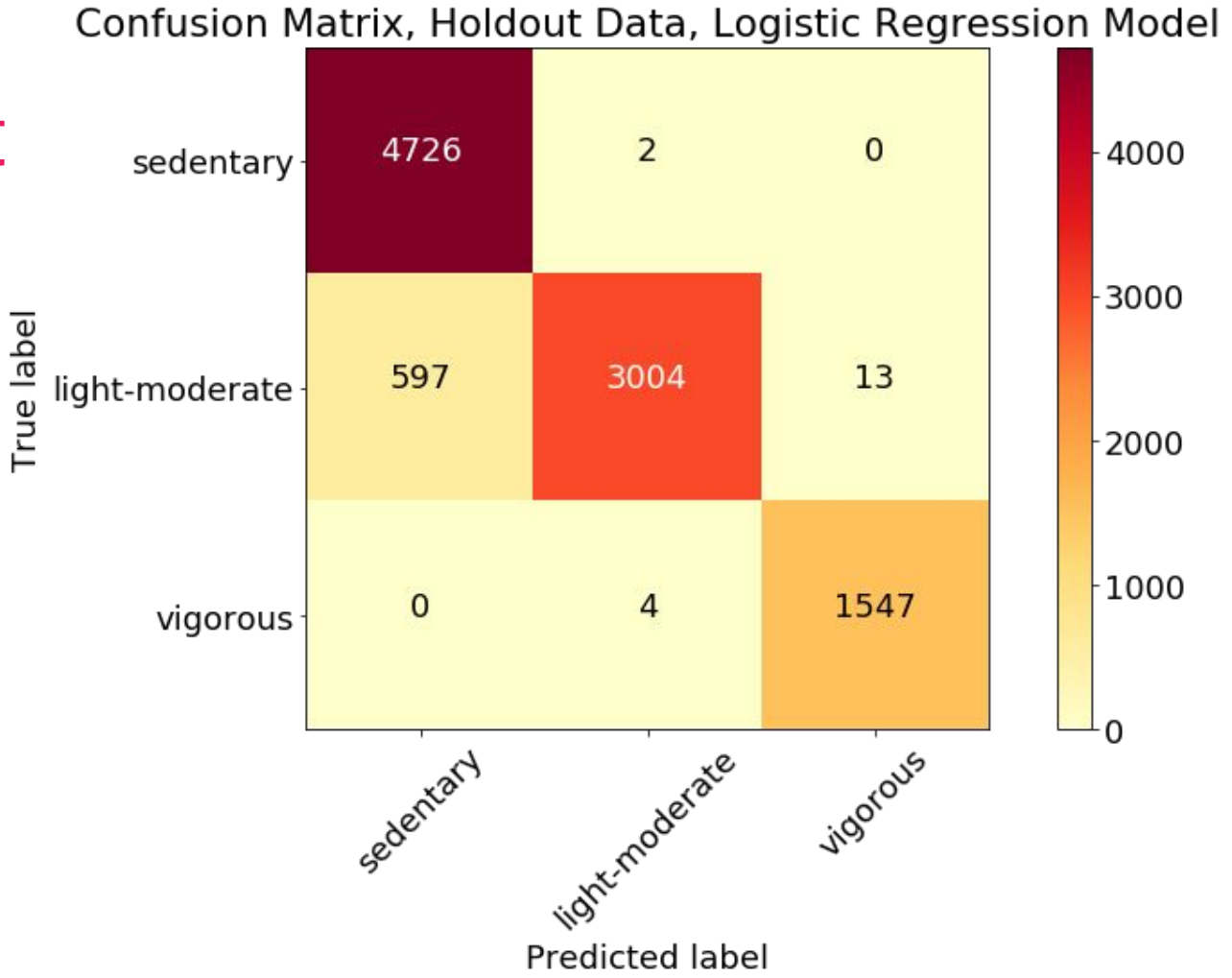
None had better results than LR

LR Test on Holdout

Tested on 5 subjects not previously used

Accelerometer only

94% accuracy



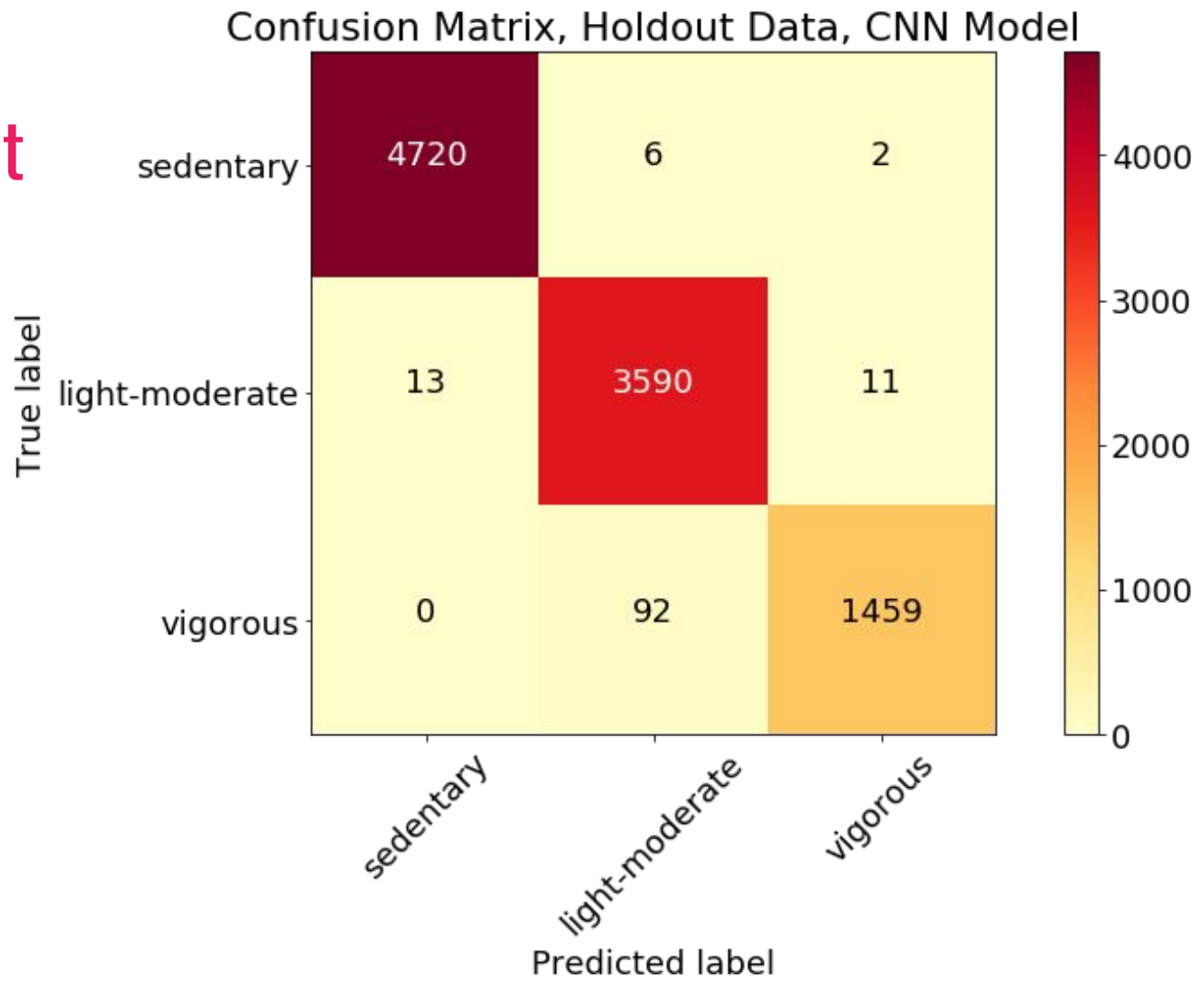
Neural Network Model

- **Raw sensor data** as array of 100 obs x 6 features (2 sensors x 3D)
- **2 Recurrent layers** to handle the time dimension using Gated Recurrent Unit (GRU)
- **2 Convolutional layers** to learn important features, prevent overfitting, and reduce number of parameters
- **Batch normalization** layer to further reduce overfitting

NN Test on Holdout

Accelerometer & gyroscope data.

99% accuracy



Takeaways and Next Steps

- **Simplify** the problem to match use case
- **Validate** on previously unseen subjects
- **Simple hand-crafted features** with LR regression produces good results
- **Neural network** has even better results
- **Test** with lower sampling rate
- **Expand** model to other data