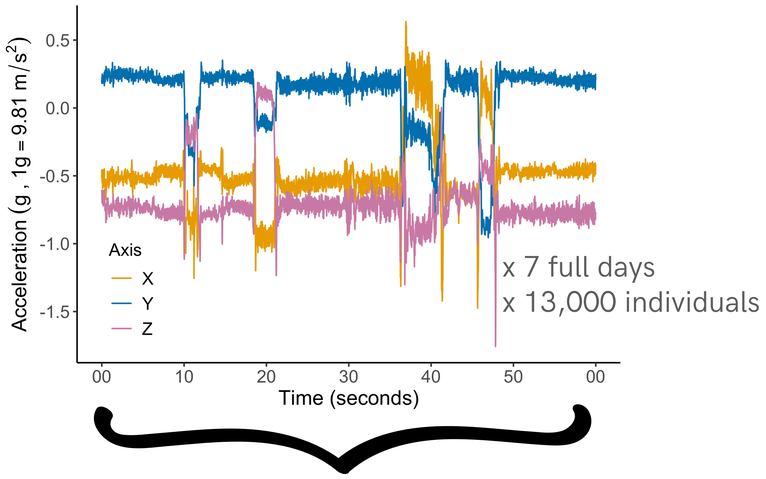
FINGERPRINTING WALKING in a large epidemiological study

Lily Koffman
John Muschelli
Ciprian Crainiceanu

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

DATA

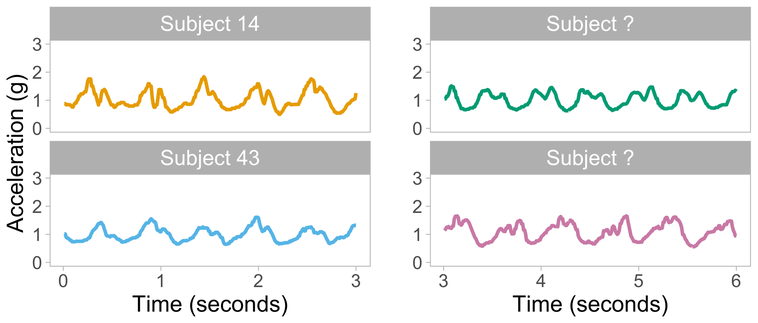
- Ntl Health & Nutrition Examination Survey (NHANES)
- Questionnaires and laboratory tests on large, **nationally representative** sample of Americans
- 2011-2014: survey included accelerometry sub-study



1 minute of data
80 observations per second per dimension

OBJECTIVE

Leverage accelerometry-derived walking patterns for biometric identification

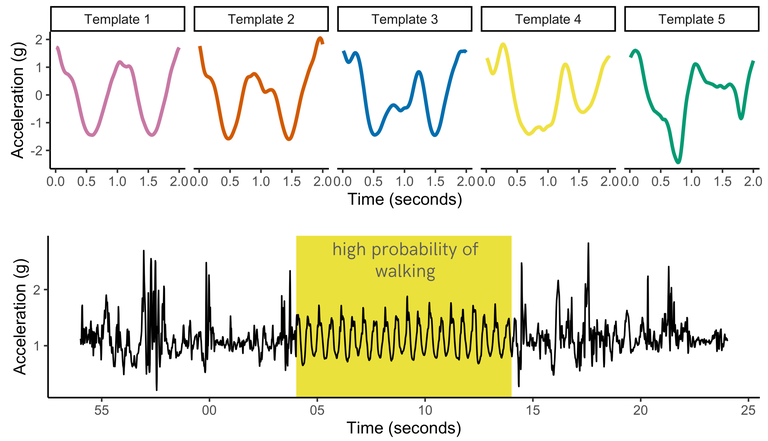


METHODS

1. Identify time periods with high probability of being walking
2. Derive **grid-cell predictors** for identification models
3. Implement various identification models under different train/test scenarios

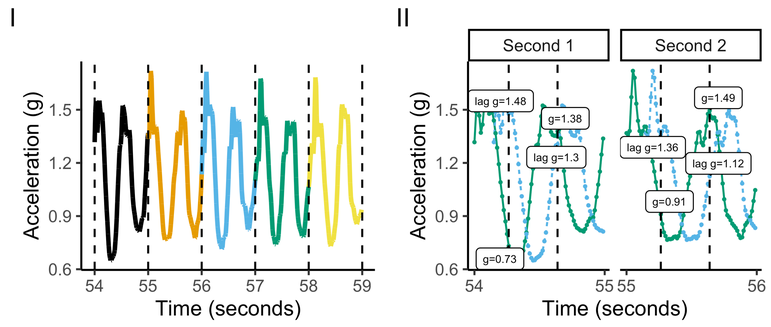
1. *IDENTIFY WALKING*

Use Adaptive Empirical Pattern Matching (ADEPT)[1] to identify walking periods from raw accelerometry data

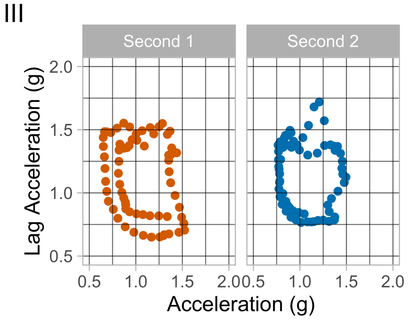


2. *DERIVE PREDICTORS*

I. Break walking periods into 1-second segments
II. Compute empirical joint distribution of acceleration and lag acceleration at each 1/80th of a second for three lags
III. Tabulate number of acceleration, lag acceleration pairs in each 0.25x0.25g square on the [0,3]gx[0,3]g grid
IV. Create predictor matrix



III



IV

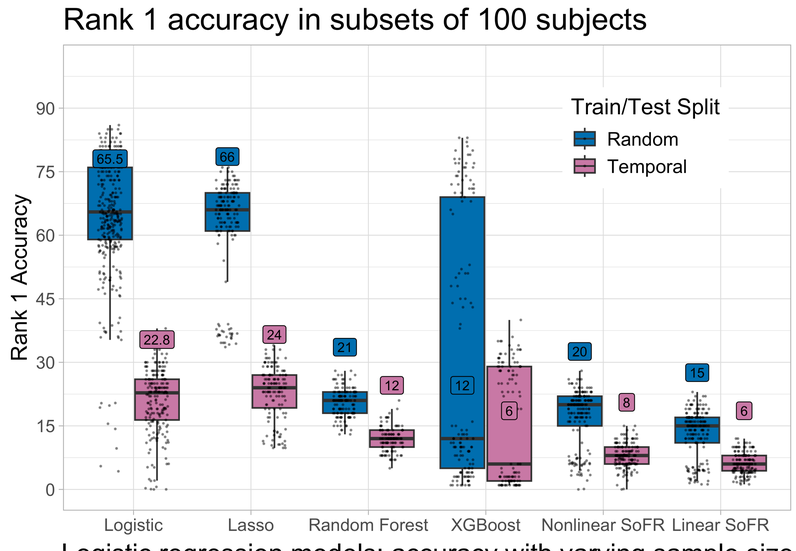
	$cell_1$	$cell_2$	$cell_3$	...
$second_1$	0	4	1	
$second_2$	0	0	2	
$second_3$	1	0	3	
⋮				

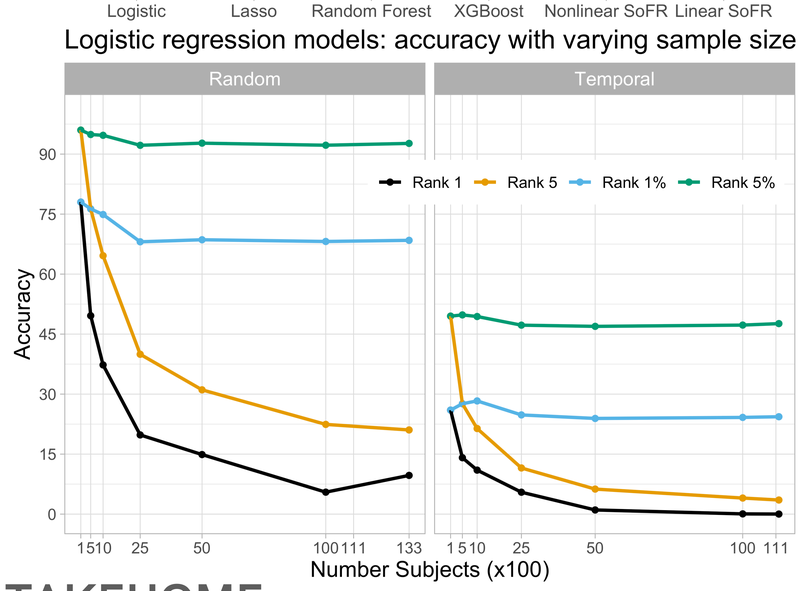
x 144 cells
x 3 lags
= 432 predictors
x 180 sec (rows) per person

3. *FIT MODELS*

- One vs. rest classification paradigm with varying size subsets of the entire sample
- Logistic regression, machine learning (random forest, XGBoost), scalar on function regression
- Training and testing randomly sampled
- Training from days 1-3, testing from days 4-7 (temporal)

RESULTS





TAKEHOME

- Accelerometry data can identify individuals from walking
- Implications for epidemiology, health, privacy/security
- Scalable to large datasets

RELATED LITERATURE
1. Karas, M., Straczkiewicz, M., Fadel, W., Harezlak, J., Crainiceanu, C. M., & Urbanek, J. K. (2021). Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation. *Biostatistics (Oxford, England)*, 22(2), 331-347
2. Koffman, L., Crainiceanu, C., & Leroux, A. (2024). Walking fingerprinting. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 73(5), 1221-1241.

github.com/lilykoff/nhanes_fingerprinting