Northwestern Medicine®
Feinberg School of Medicine

Wear it Well:
Data Analysis for Wearable Devices

Statistically Speaking Lecture Series

February 13, 2024

Margaret Banker, PhD
Assistant Professor
Division of Biostatistics
Department of Preventive Medicine
: margaret.banker@northwestern.edu

Northwestern Medicine®
Feinberg School of Medicine

Biostatistics Collaboration Center (BCC)

Who We Are

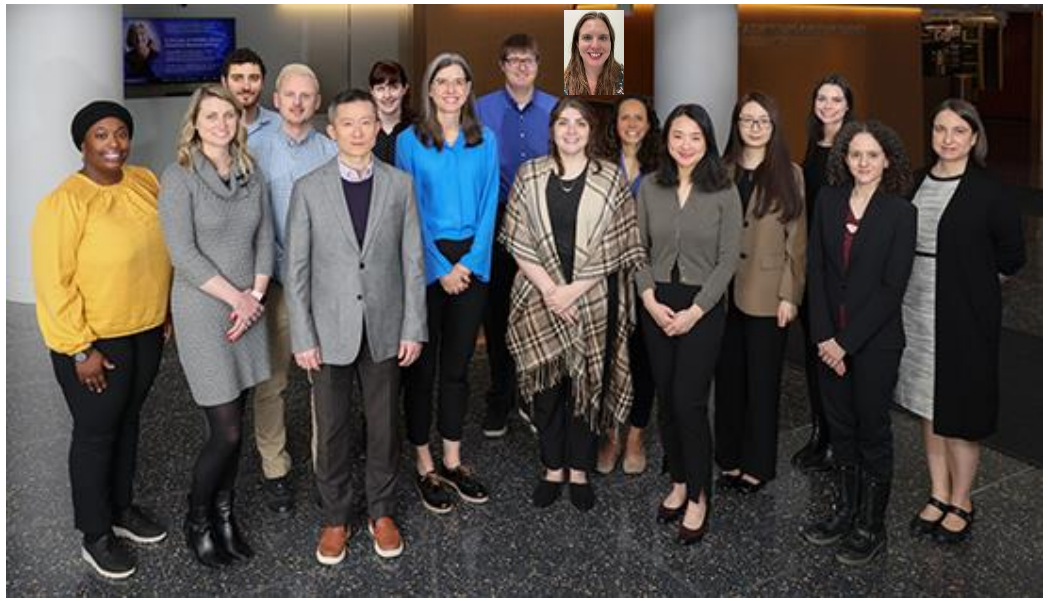
Biostatistics **C**ollaboration **C**enter

Mission: to support investigators at FSM in the conduct of high-quality, innovative health-related research by providing expertise in biostatistics, statistical programming, and data management.

Cancer-related biostatistics needs → RHLCCC's Quantitative Data Sciences Core (QDSC)

Multi-site data coordinating → NUDACC (NU Data Analysis & Coordinating Center)

Biostatistics Collaboration Center



13	PhD Faculty Statisticians
9	MS Statistical Analysts
> 300	Research Projects (Annually)
> 75	Grant Proposals (Annually)

What can we do?

Many areas of expertise, including:

- Bayesian Methods
- Big Data
- Bioinformatics
- Causal Inference
- Clinical Trials
- Database Design
- Genomics
- Longitudinal Data
- Machine Learning
- Missing Data
- Reproducibility
- Statistical Genetics
- Survival Analysis
- Wearable Device Data

Many types of software, including:

Biostatistics Collaboration Center

REQUEST AN APPOINTMENT

Come talk to us for an hour – we can learn about your work and how we might be able to help. **It's covered** by the FSM Office for Research, and the **wait is not long**!

Writing a grant?

- ✓ Power/sample size
- ✓ Statistical analysis plan
- ✓ Co-Investigator

Help with analysis? Constructing a database?

- ✓ Develop statistical analysis plan (no cost)
- ✓ Hourly charge for analysis ($145 or $175/hour)

Request an Appointment here!
https://www.feinberg.northwestern.edu/sites/bcc/index.html

Northwestern Medicine®
Feinberg School of Medicine

Northwestern Medicine®
Feinberg School of Medicine

Wearable Devices

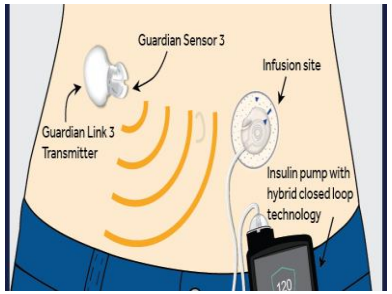
What Are They?

Wearing What?
Personal Wearable Devices

- Automatic data collection in high frequency
- Track physiological variables and clinical symptoms outside of clinical environment
- Non-invasive and convenient
- Help MD and patients deliver better decisions and health care
- Increasingly used in health research
- Inherent data challenges

Basal Temperature Tracking

Physical Activity / Sleep

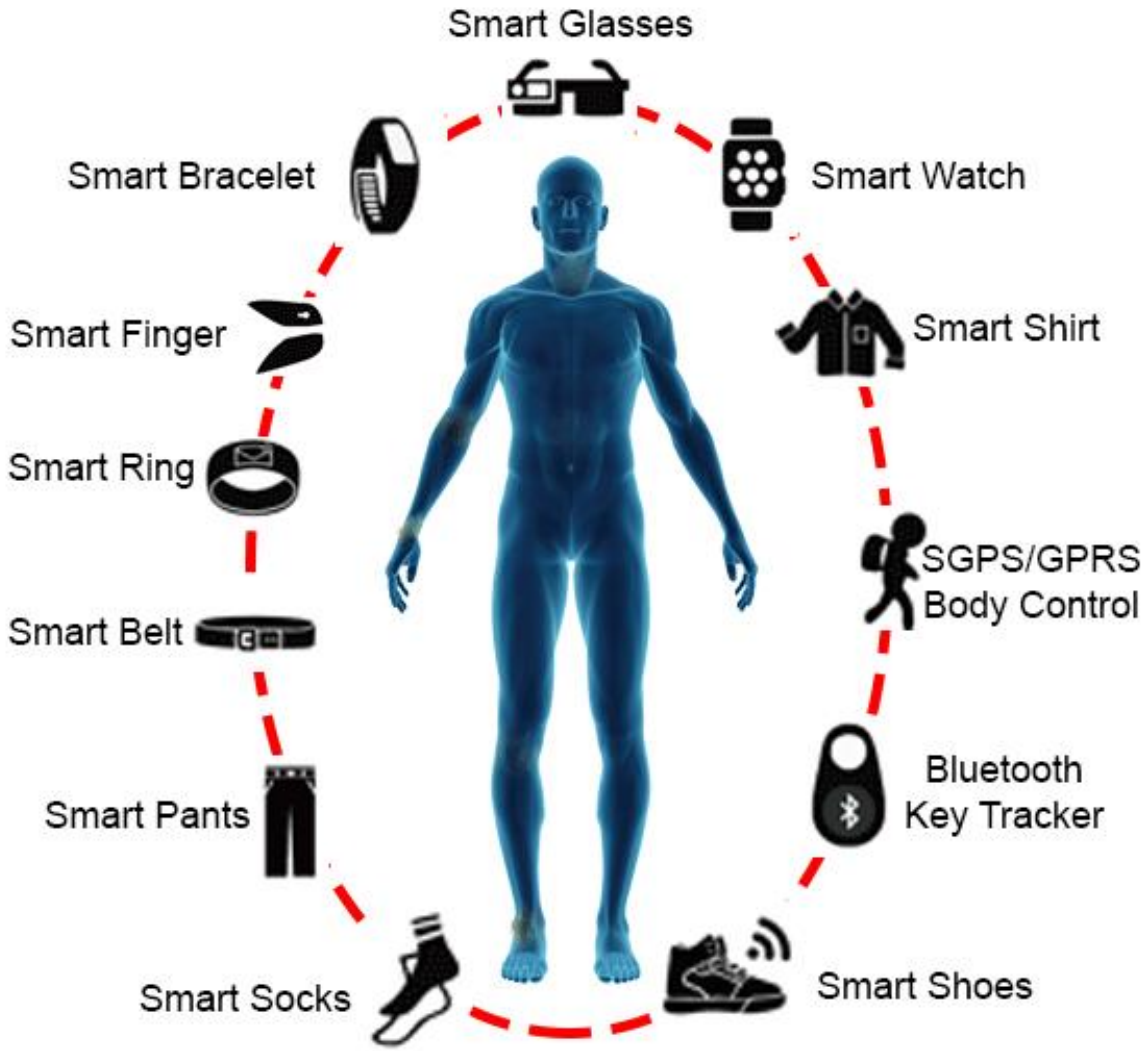


Glucose Monitoring

Phone Sensors

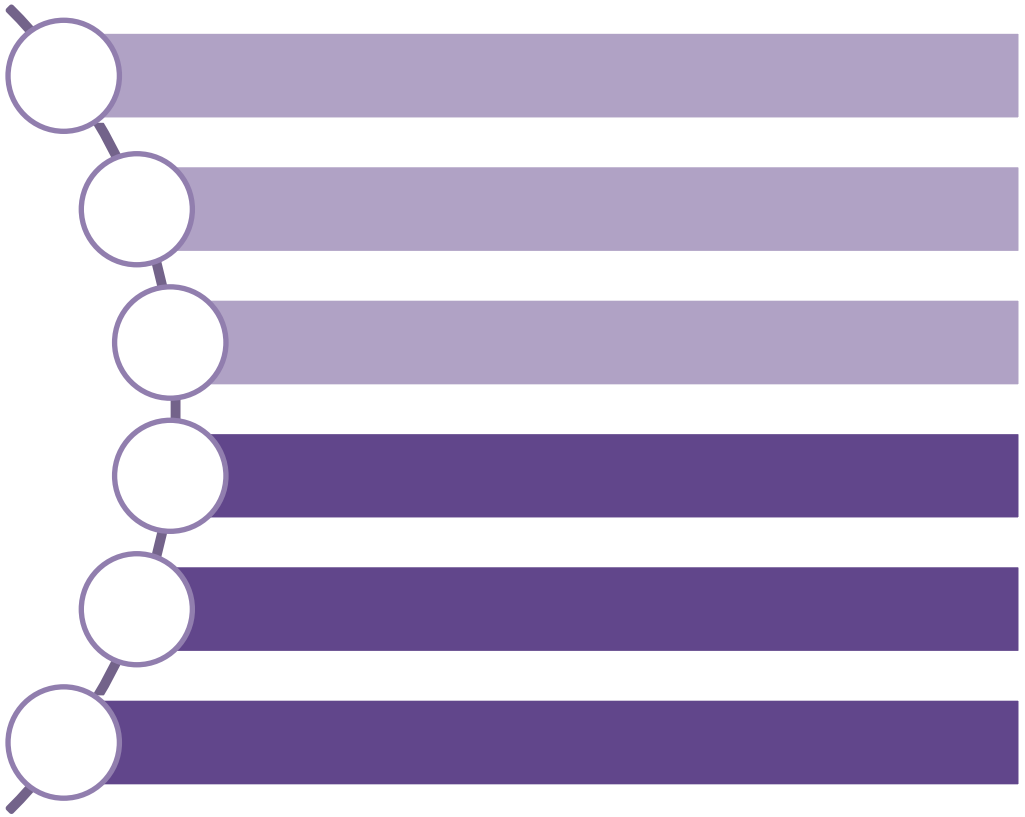
Source: Yono Lab Earbud
Source: Glucose Monitoring Image

Where to Wear?



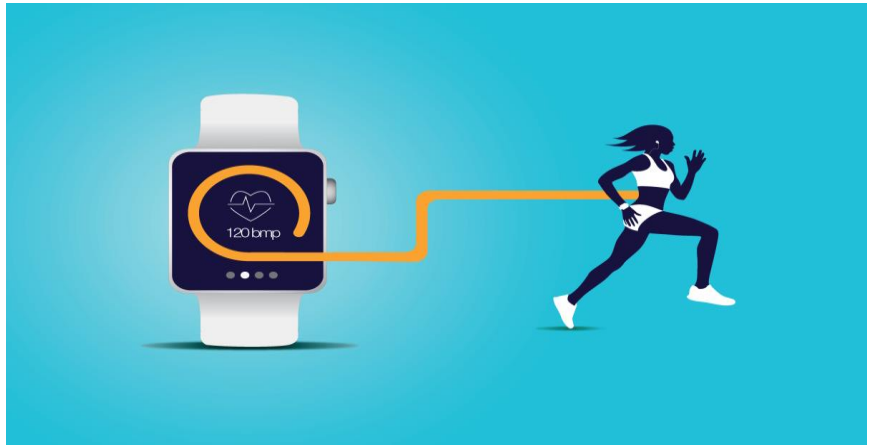
Source: Wearable Device Image

Wear for what?



Wear for what?

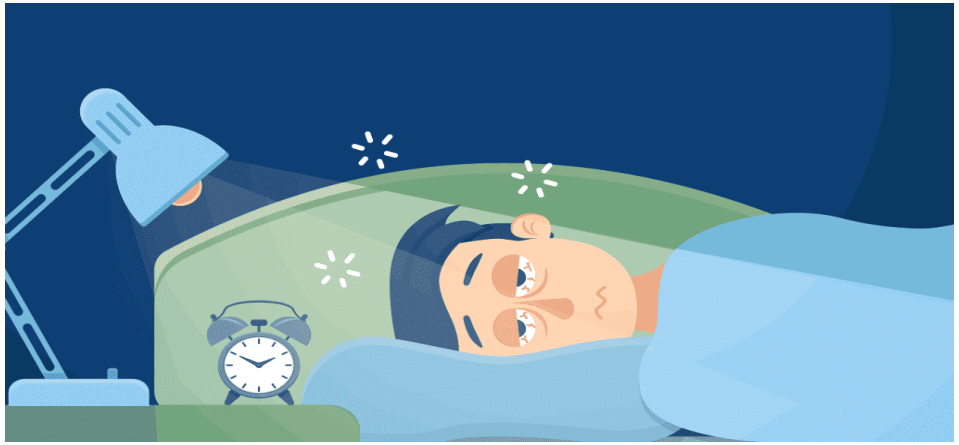
- Activity Monitoring
-
-
-
-
-



Calories Burn

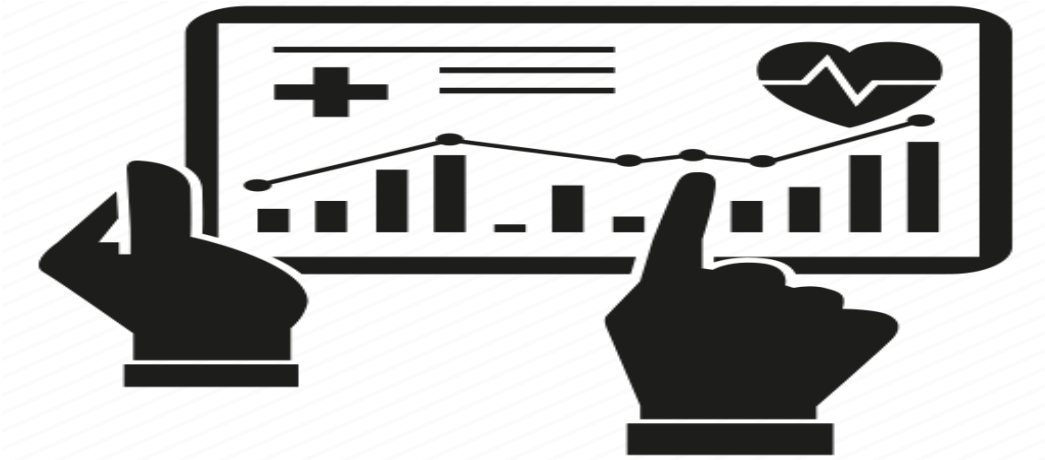
Wear for what?

- Activity Monitoring
- Sleep Tracking
-
-
-
-



Wear for what?

- Activity Monitoring
- Sleep Tracking
- Physical Health Monitoring



Blood Oxygen Level

Wear for what?

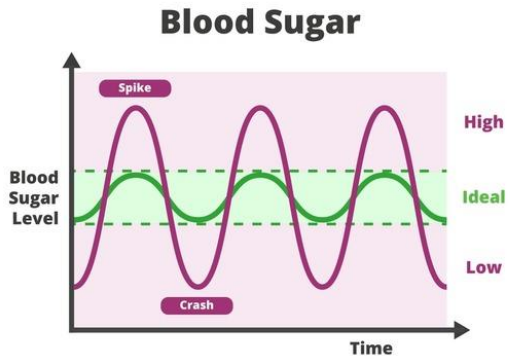
- Activity Monitoring
- Sleep Tracking
- Physical Health Monitoring
- Behavioral Monitoring

Source: Medication Adherence Cartoon

Source: Food Tracking Image

Wear for what?

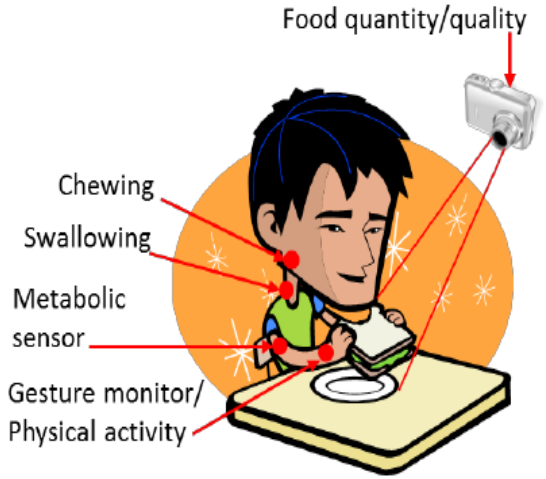
- Activity Monitoring
- Sleep Tracking
- Physical Health Monitoring
- Behavioral Monitoring
- Clinical Trials/Interventions

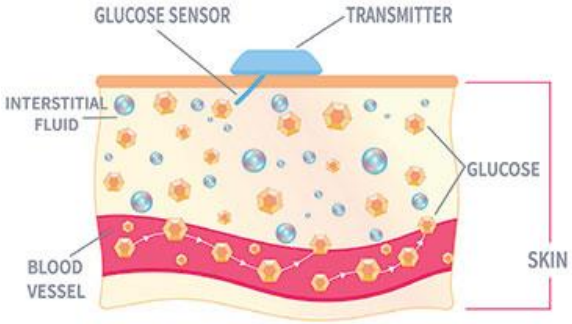


Source: Blood Sugar Image Source: Mood App Image Source: Intervention Prompt Image

Wear for what?

- Activity Monitoring
- Sleep Tracking
- Physical Health Monitoring
- Behavioral Monitoring
- Clinical Trials/Interventions
- Disease Management



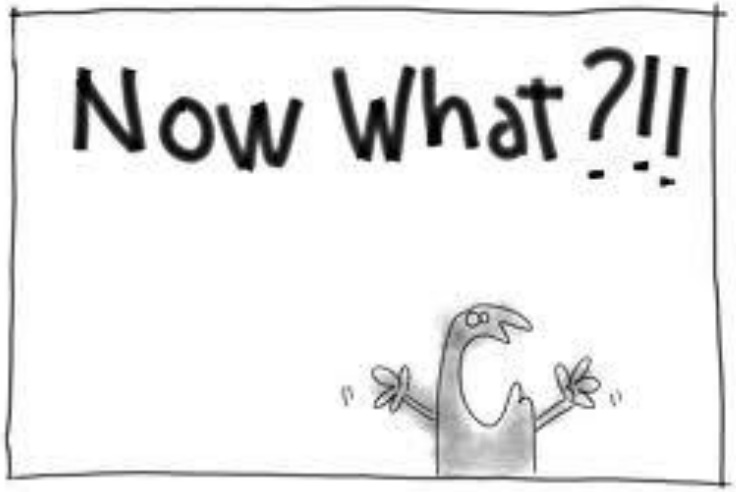


Source: Food Tracking Image

Source: CGM Image

Wear for what?

- Activity Monitoring
- Sleep Tracking
- Physical Health Monitoring
- Behavioral Monitoring
- Clinical Trials/Interventions
- Disease Management



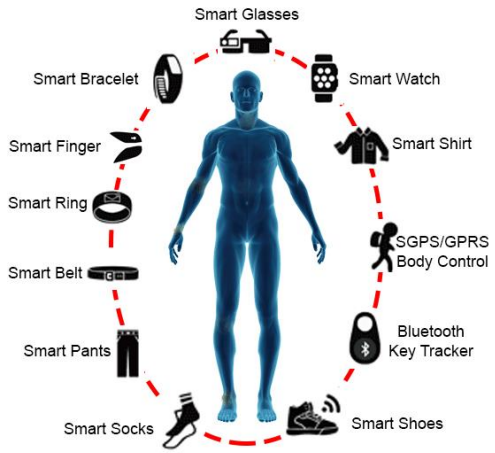
Source: Cartoon

Link Accreditation: Universe Zoom Out Gif

Northwestern Medicine®
Feinberg School of Medicine

Wearable Device Data

Cleaning/Processing/Summary Metrics

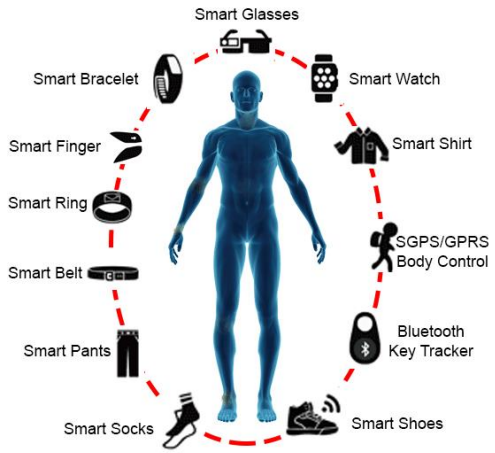


Raw Data

Processed Data

Data Analysis

Source: Wearable Device Image

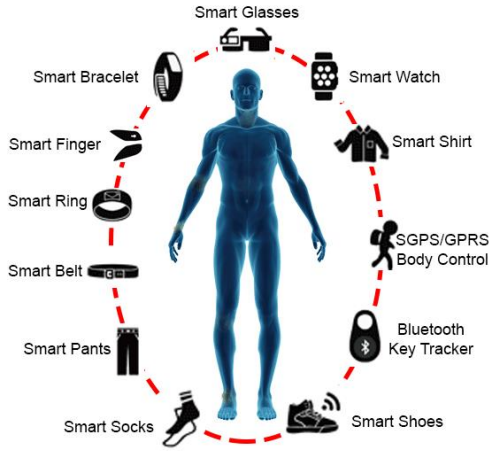


Raw Data

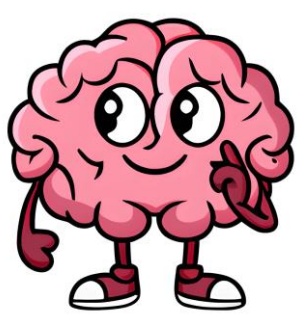
Processed Data

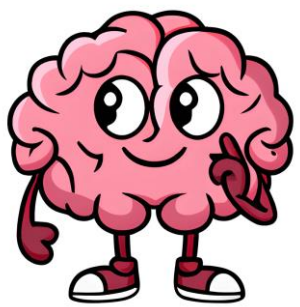
Data Analysis

Source: Wearable Device Image



Raw Data





Processed Data

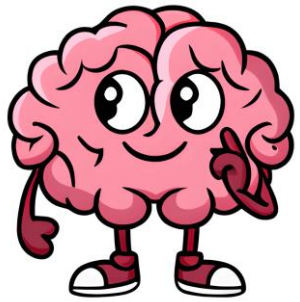
Data Analysis

Source: Wearable Device Image

Wearable Device Example: Accelerometers

- Capture continuous measures of **objective** activity
- Detects acceleration in 3 orthogonal planes
  - Vertical (axis 1, *x*)
  - Horizontal (axis 2, *y*)
  - Perpendicular(axis 3, *z*)
- Considerations
  - Raw to Processed Data?
  - Sleep? Physical Activity?
  - Single summary statistics? Patterns of activity?

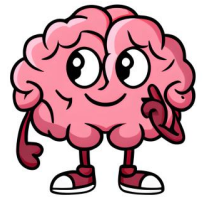
Raw Accelerometer Data



Accelerometer X	Accelerometer Y	Accelerometer Z
0.078	-1.008	0.051
0.078	-0.934	0.125
0.176	-0.961	-0.262
0.172	-0.957	-0.293
0.203	-0.961	-0.301
0.219	-0.949	-0.289
0.164	-0.953	-0.273
0.148	-0.957	-0.281
0.148	-0.973	-0.270
0.188	-0.953	-0.270
0.184	-0.957	-0.266
0.207	-0.973	-0.203
0.223	-0.965	-0.258
0.223	-0.961	-0.266
0.035	-0.953	-0.227
0.211	-0.992	-0.234
0.289	-0.945	-0.219
0.199	-0.973	-0.152

Note: Many data points per second

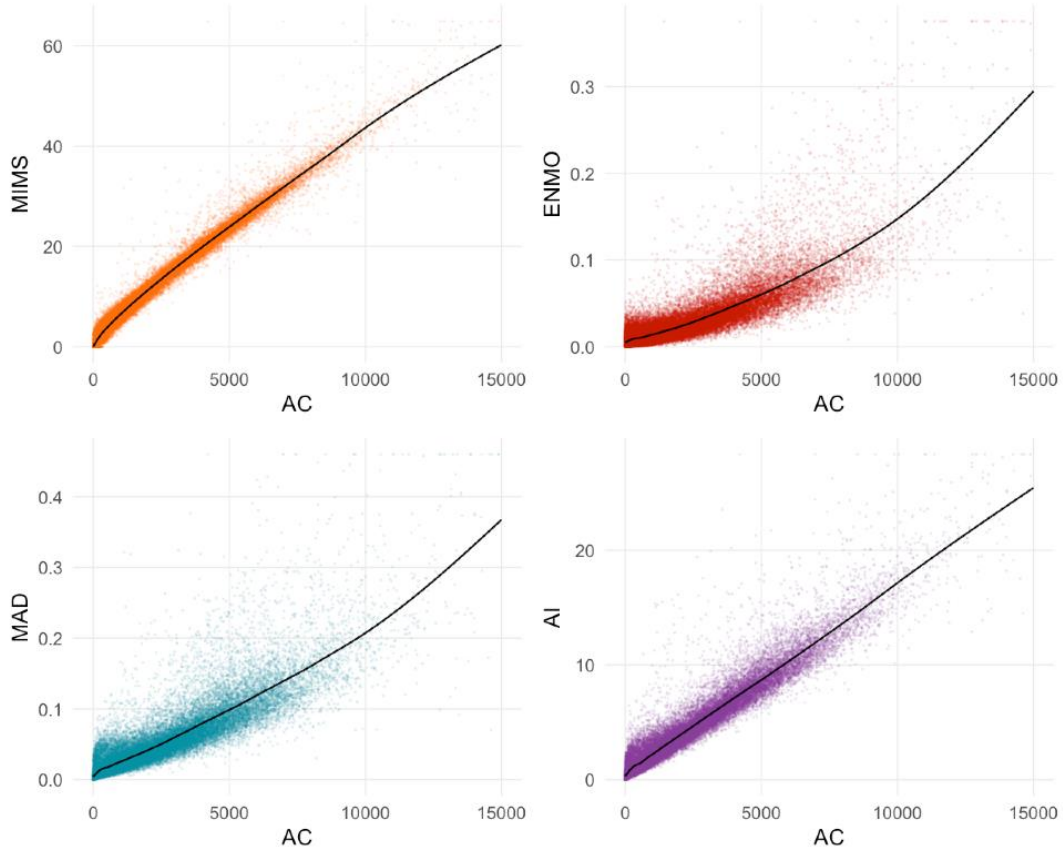
Example: 60Hz devices capture 60 readings per second



Processed Accelerometer Data

Processed Intensity Measurements

- **AC**: Activity Count (device specific)
- **MIMS:** Monitor-Independent Movement Summary
- **ENMO:** Euclidean Norm Minus One
- **MAD:** Mean Amplitude Deviation
- **AI:** Activity Intensity



Kartas, Marta et al. "Comparison of Accelerometry-Based Measures of Physical Activity: Retrospective Observational Data Analysis Study", *JMIR Mhealth and uHealth, 2022.* DOI: 10.2196/38077

Processed Accelerometer Data – from Actigraph

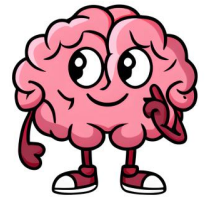
Granularity

Step Count

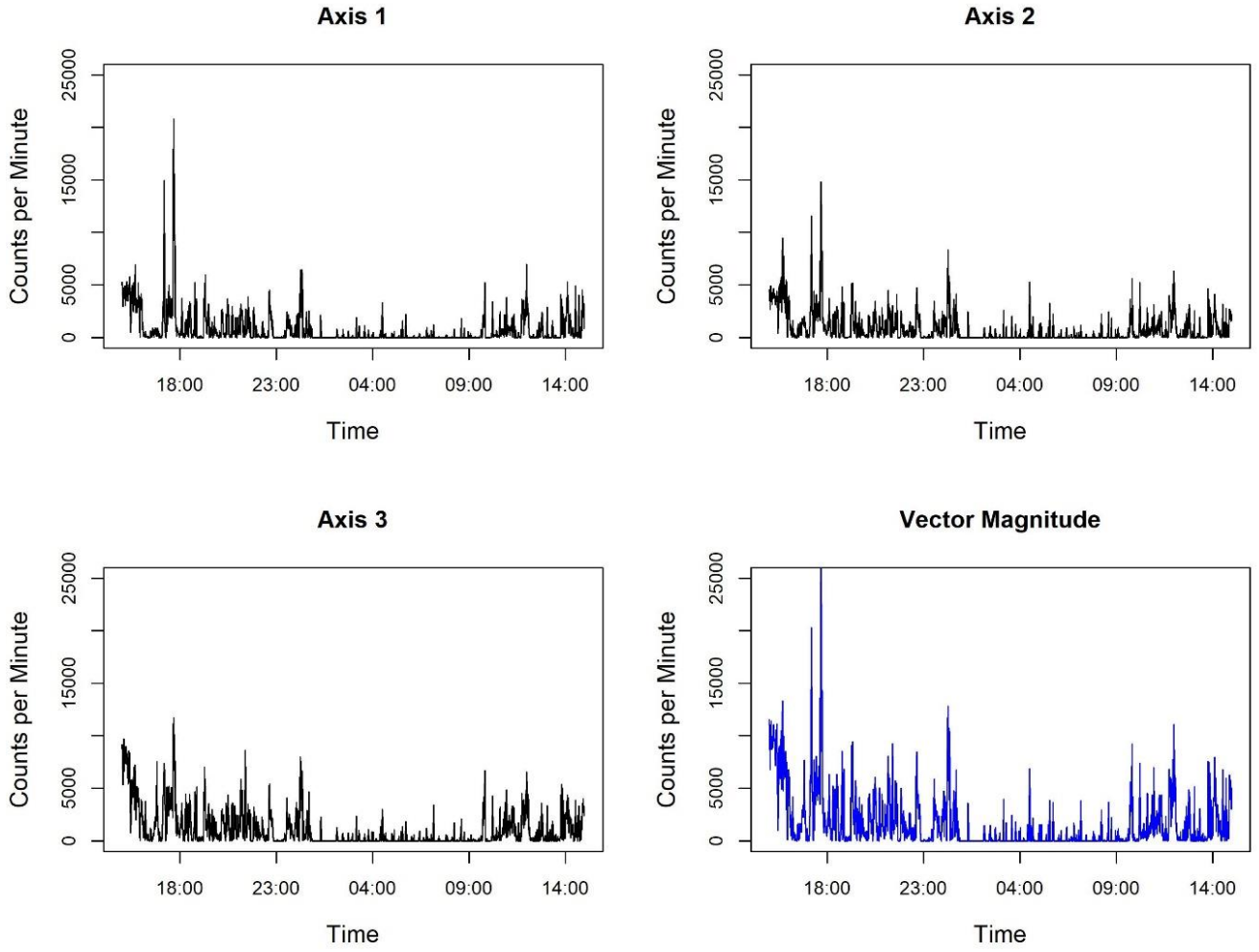
Direction and AC Intensity

Orientation Variables

Epoch	axis1	axis2	axis3	steps	inclineStanding	inclineSitting	inclineLying
6/9/2022 15:00	0	0	0	0	0	0	60
6/9/2022 15:01				3			
6/9/2022 15:02		0	0	0		60	0
6/9/2022 15:03	0	0	0	0	0	60	0
6/9/2022 15:04	572	901	1391	6	30	30	0
6/9/2022 15:05	27	311	722	1	0	57	3
6/9/2022 15:06	45	93	467	1	0	26	34
6/9/2022 15:07	845	1293	1600	8	40	2	18
6/9/2022 15:08	443	234	543	3	20	0	40
6/9/2022 15:09	1008	433	1666	8	20	28	12
6/9/2022 15:10	512	326	522	6	10	10	40
6/9/2022 15:11	1172	1001	2226	13	50	10	0
6/9/2022 15:12	228	212	889	1	10	18	32
6/9/2022 15:13	219	283	552	2	10	18	32
6/9/2022 15:14	374	164	322	5	20	17	23
6/9/2022 15:15	689	495	1721	10	50	4	6
6/9/2022 15:16	85	123	499	1	0	41	19
6/9/2022 15:17	963	850	2947	7	30	0	30



Activity Count (AC) Data

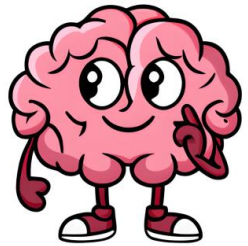


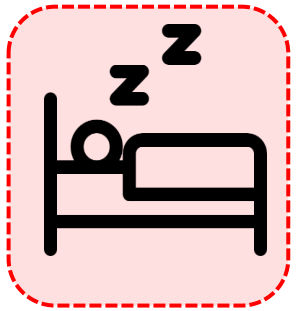
- Data expressed as number of **"Counts"** (movements) per **"Epoch"** (unit of time)
  - Epochs **can vary**… e.g. 10 second, one-minute, five-minute, etc.
- PA Intensity measures at each of 3 axes
  - Vertical (axis 1, *x*)
  - Horizontal (axis 2, *y*)
  - Perpendicular(axis 3, *z*)
  - **Vector Magnitude (VM):**

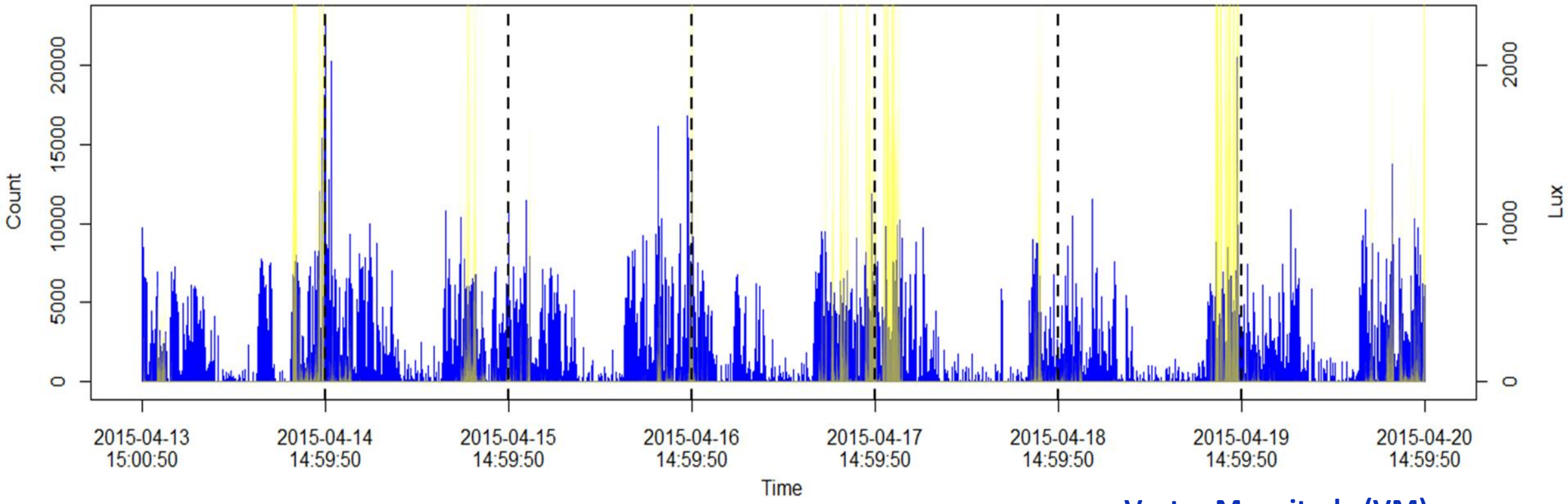
$$\sqrt{x^2 + y^2 + z^2}$$

We Processed Our Data! Now what?

Determining Metrics

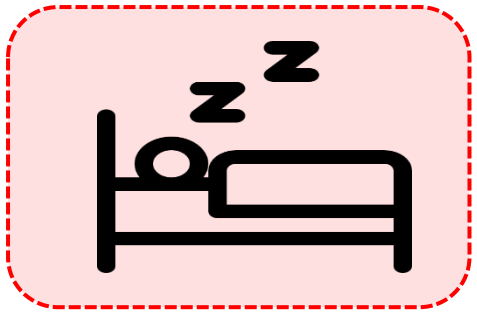


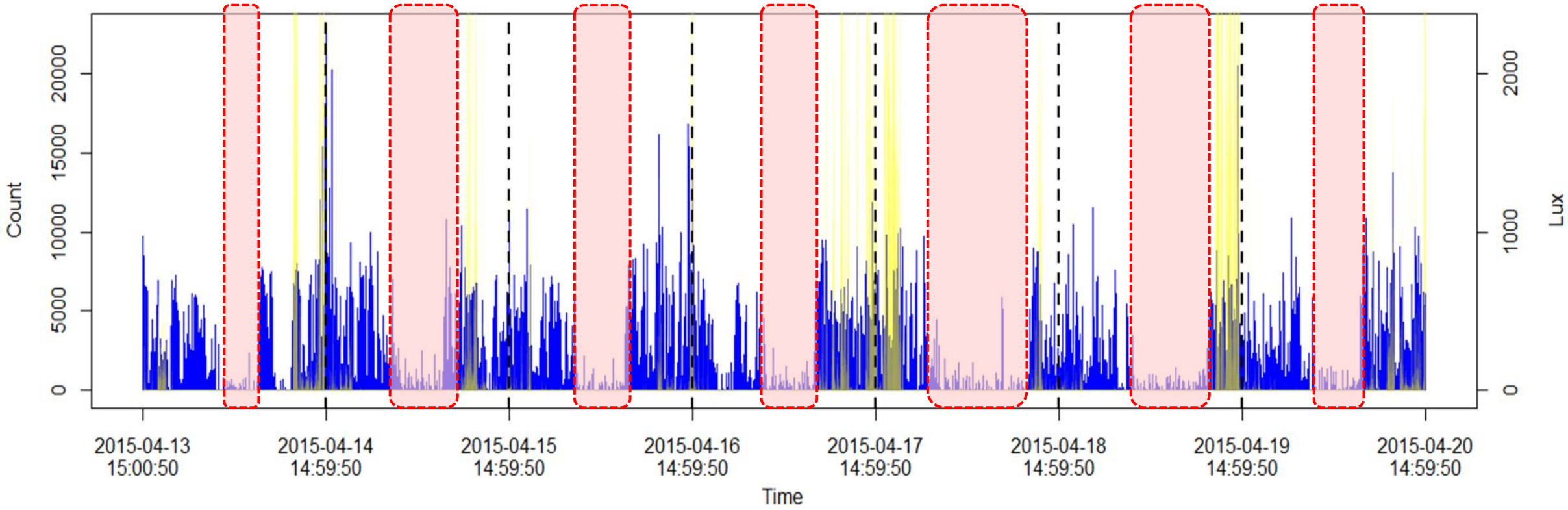


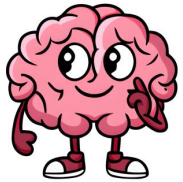


- **Vector Magnitude (VM)**
- **"Lux", or light intensity**

Analyzing Sleep Data







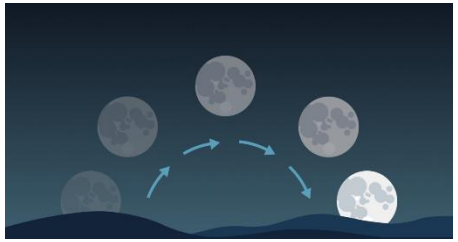
How Do We Estimate Sleep?

Analyzing Sleep Data

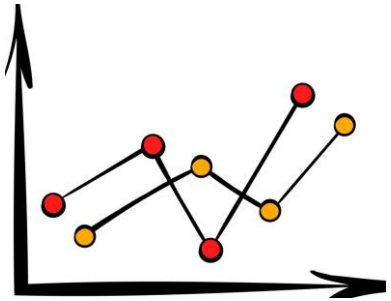
Bed Time

Wake Time

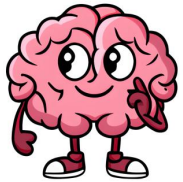
Sleep Duration



Sleep Timing

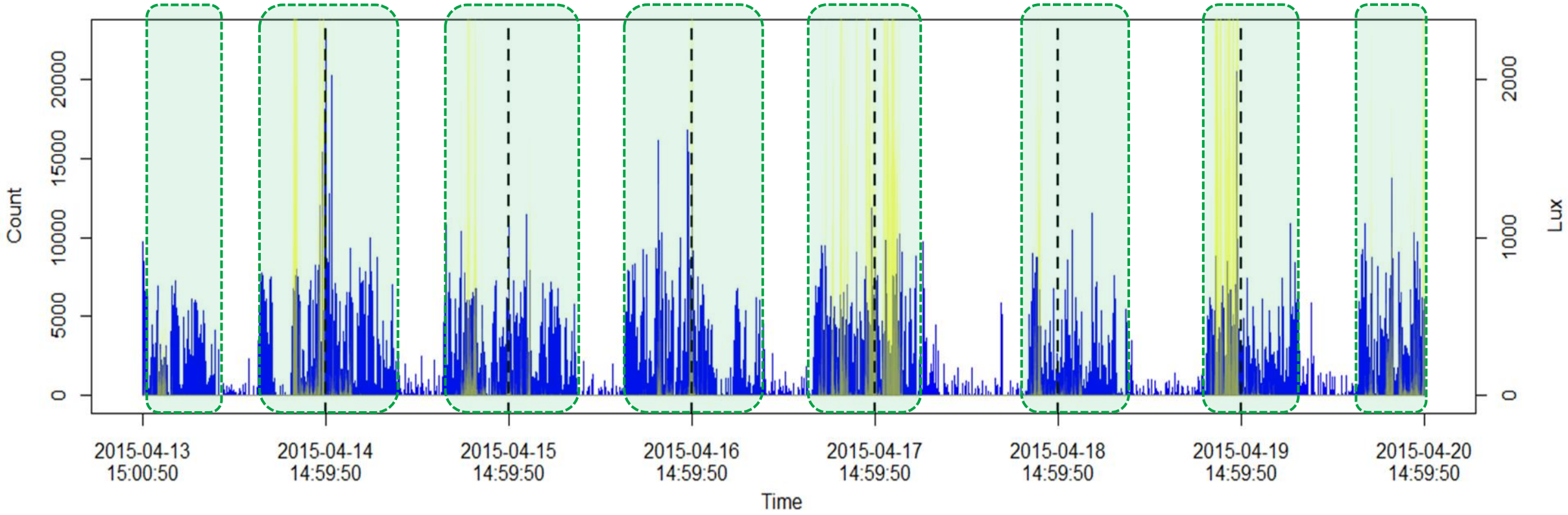


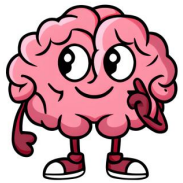
Sleep Variability



What Sleep Measurements do we Care About?

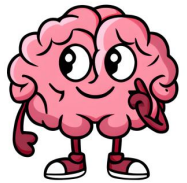
Analyzing Physical Activity (PA) Data



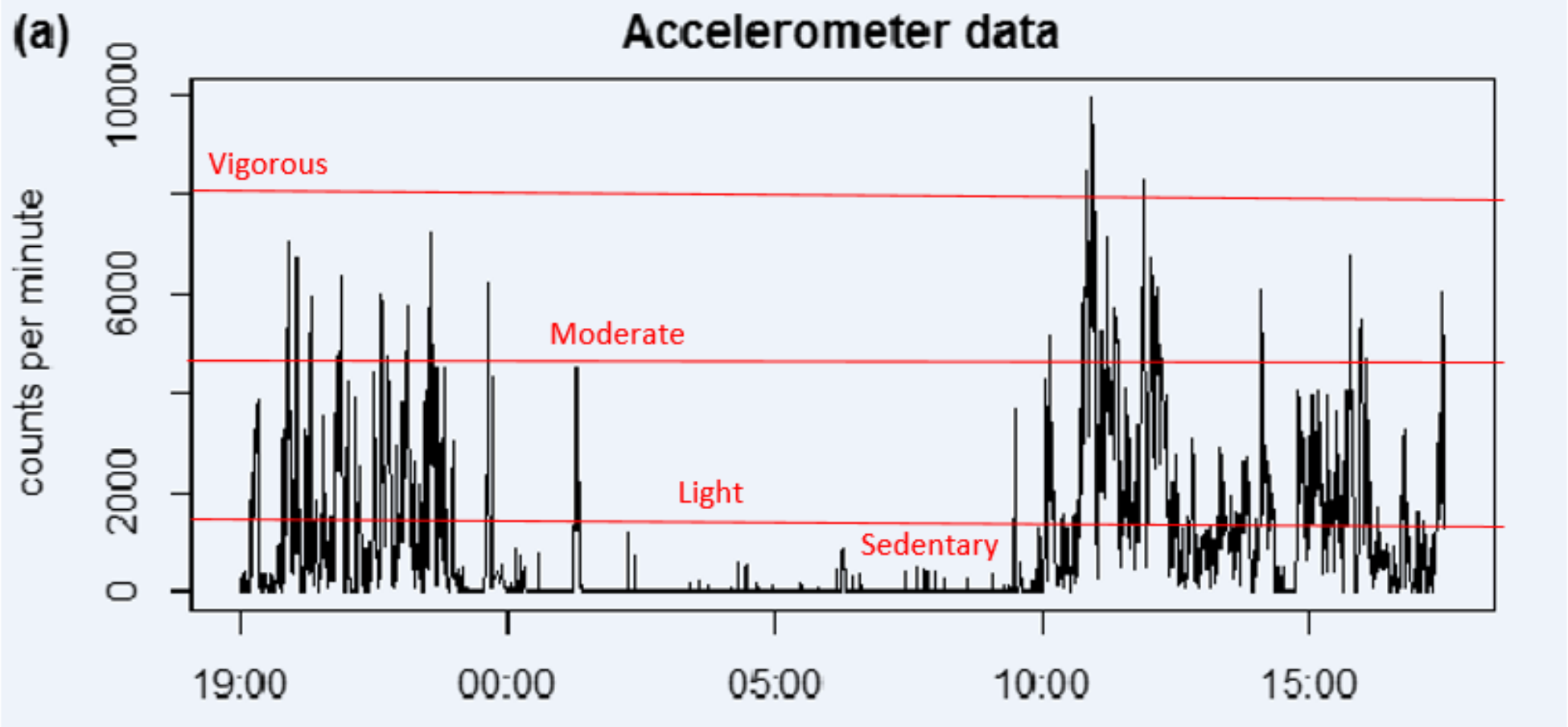


How Do We Estimate PA Intensities?

Analyzing Physical Activity (PA) Data



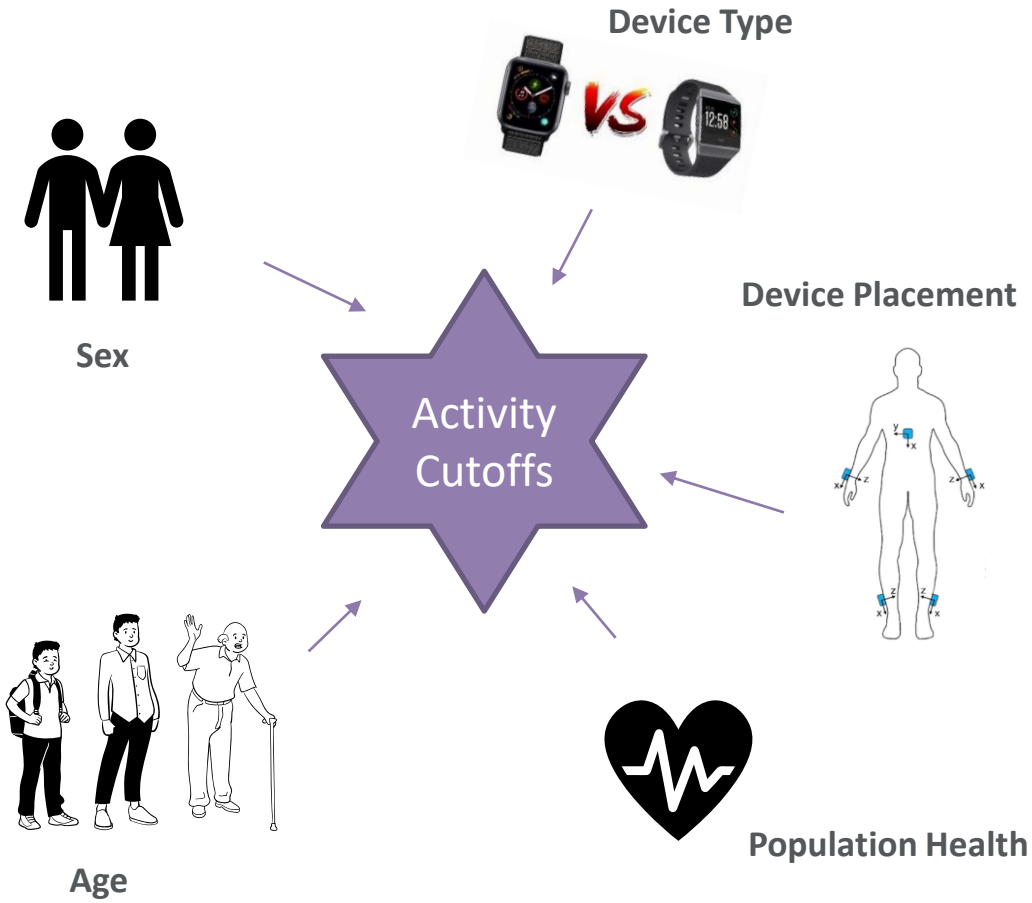
Where do the cutoffs come from?



Analyzing Physical Activity (PA) Data

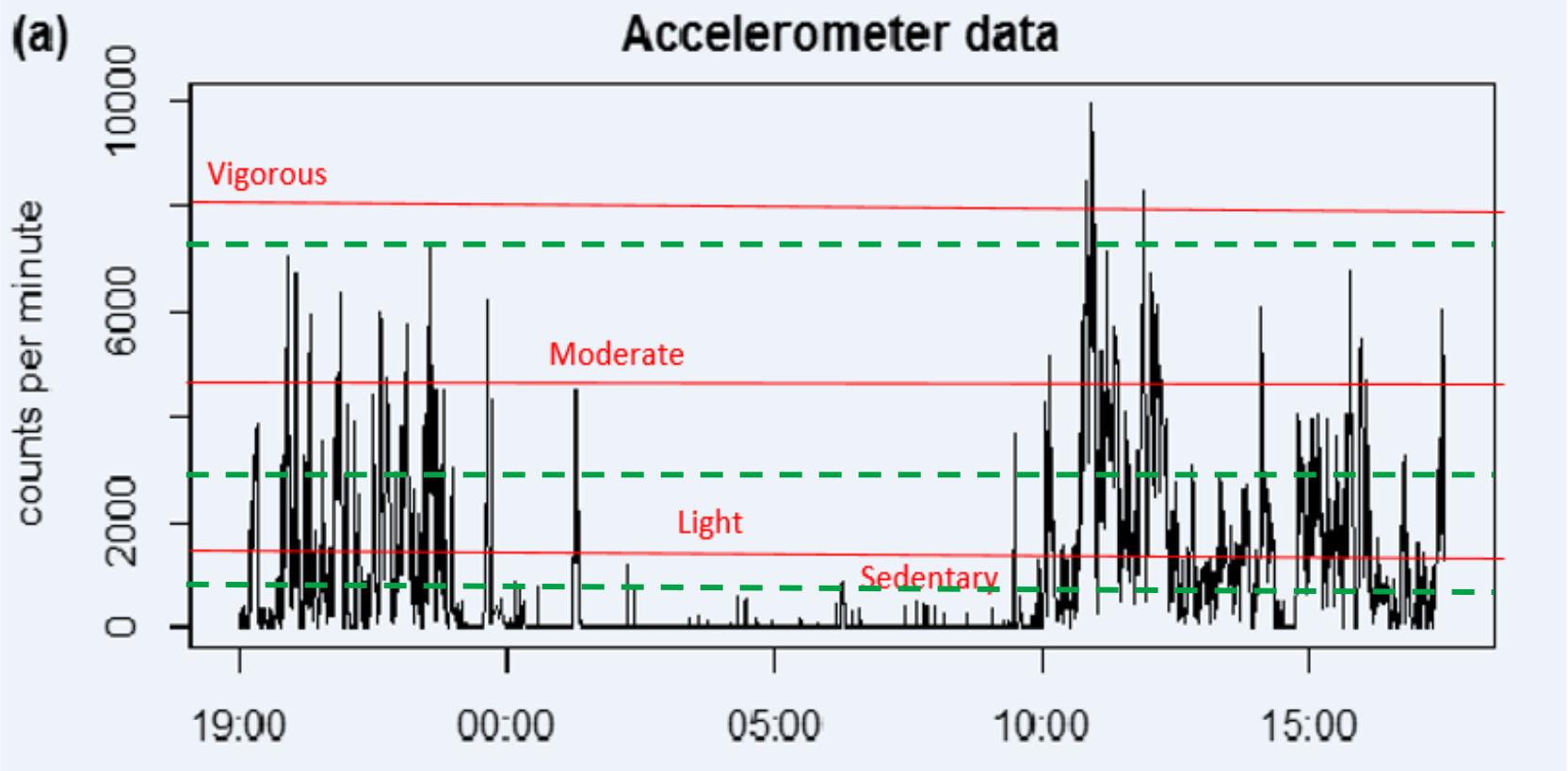
Intensity Cutpoint Classifications

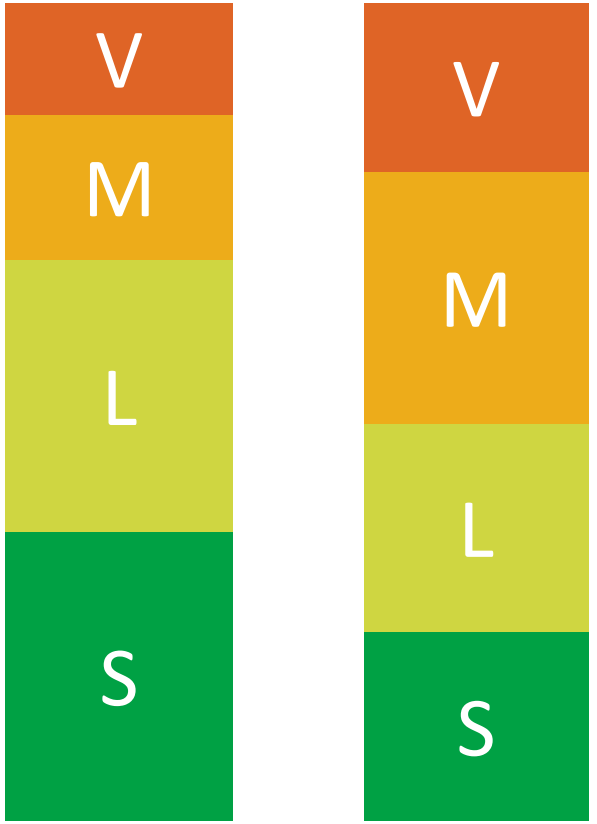
- There are **many** potential published cutoff classifications
  - Validated against different subgroup populations and studies
  - Many variables affect cutoff options
- Use of cutoffs derived from population that does not align with your own study
  - Lead to incorrect or biased activity classifications



Analyzing Physical Activity (PA) Data

Intensity Classification Cutpoints

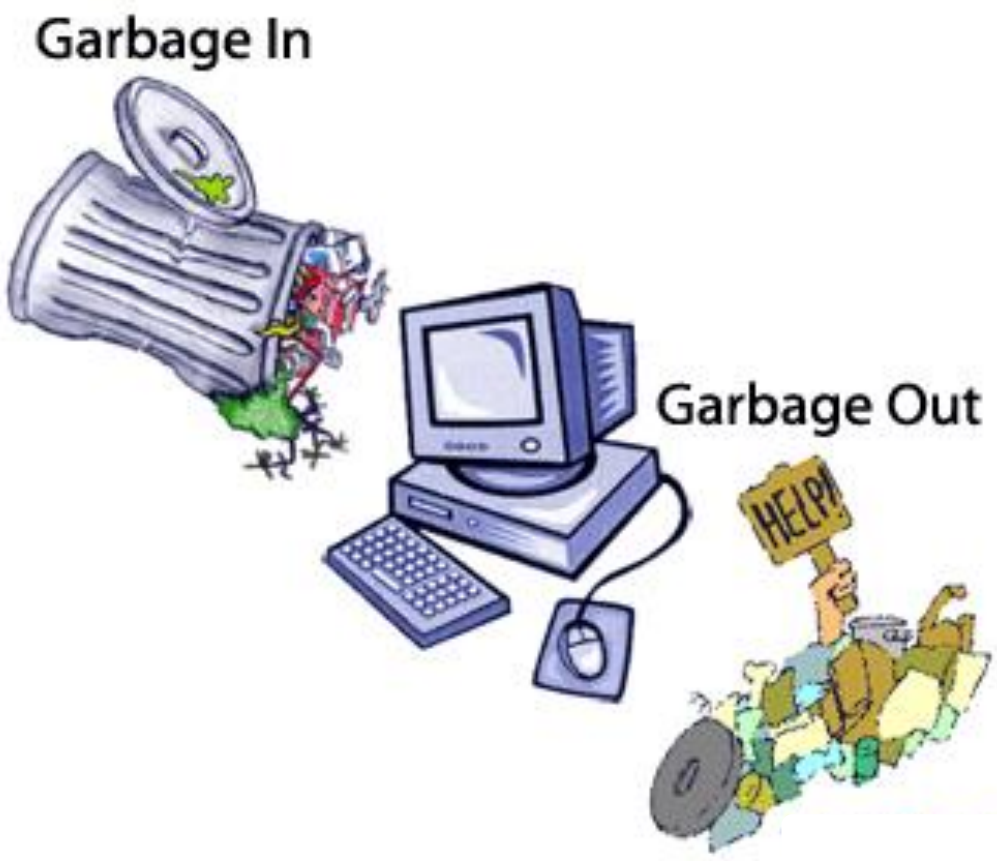


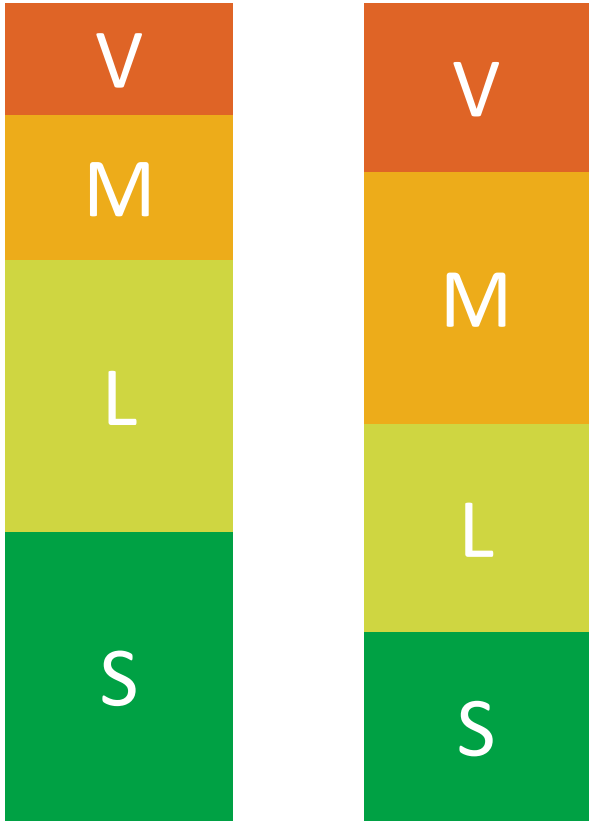


Minutes per Day at each Activity Level

Analyzing Physical Activity (PA) Data

Intensity Classification Cutpoints



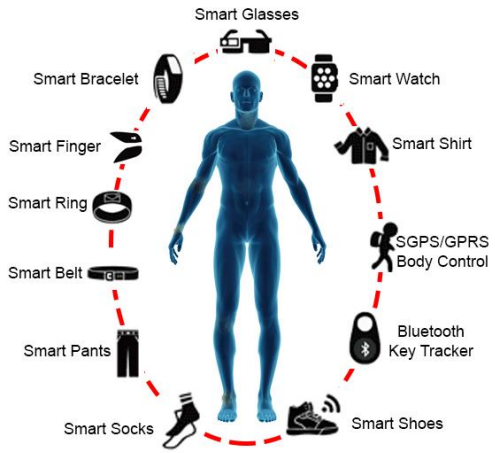


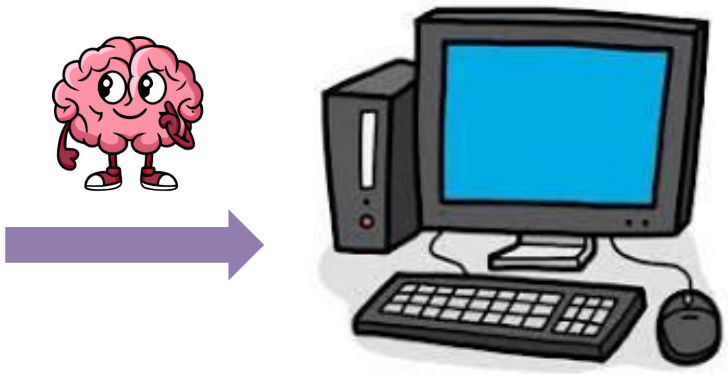
Minutes per Day at each Activity Level

Northwestern Medicine®
Feinberg School of Medicine

Analyzing Your Data

A few approaches

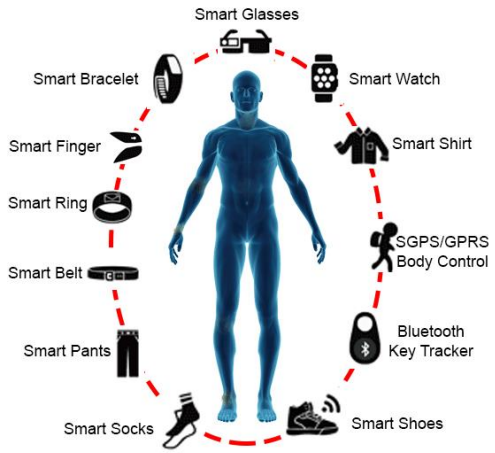


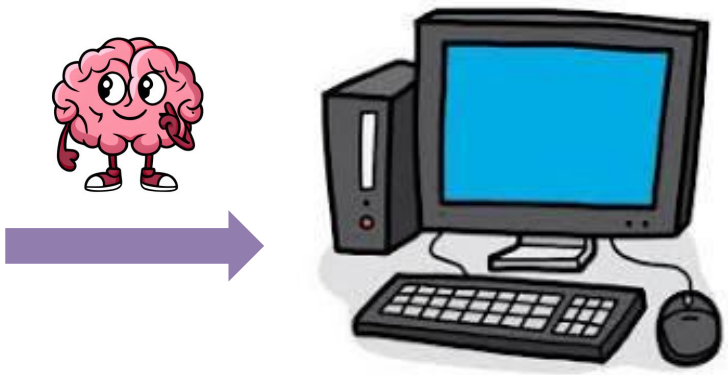


Raw Data

Processed Data

Source: Wearable Device Image





Raw Data

Processed Data

Data Analysis

Source: Wearable Device Image

Analysis of Wearable Device Data

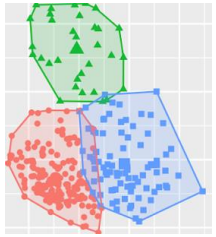
No “one-size-fits-all” approach

- Statistical Analysis depends on your research question
- Examples:
  - Time from taking blood pressure medication until BP drops below a pre-specified threshold? **Survival Analysis**
  - Association of sleep duration with (continuous) cognitive outcome? **Linear Regression**
  - Effect of daily MVPA on mood over a one-week period? **Longitudinal Modelling**

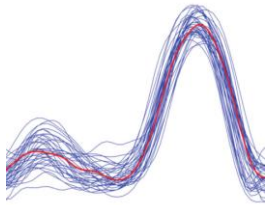
Analysis of Wearable Device Data

Highlighted Other Approaches

Substitution Analysis



Cluster Analysis

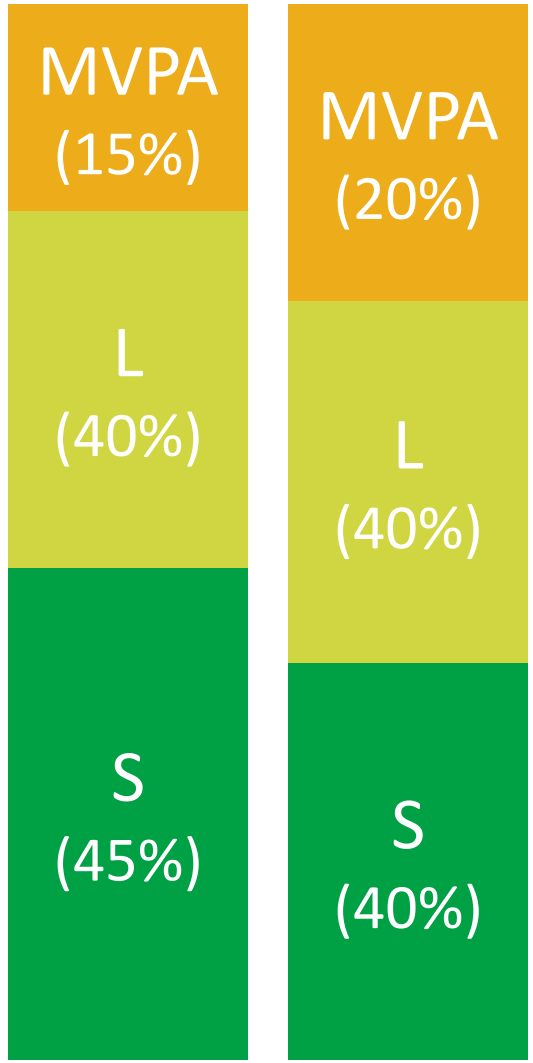


Functional Data Analysis (FDA)

Substitution Modeling

Analysis of Physical Activity

- Research Question
  - What if an individual **replaces 5%** of Sedentary Behavior with Moderate-to-Vigorous Activity (MVPA)
    - **Note:** Different question than can be answered with standard linear regression
  - How would this substitution affect a health outcome (waist circumference)
- Results
  - Substitution associated with a reduction in waist circumference by 1.35 cm with 95% CI = (– 1.91, – 0.79)



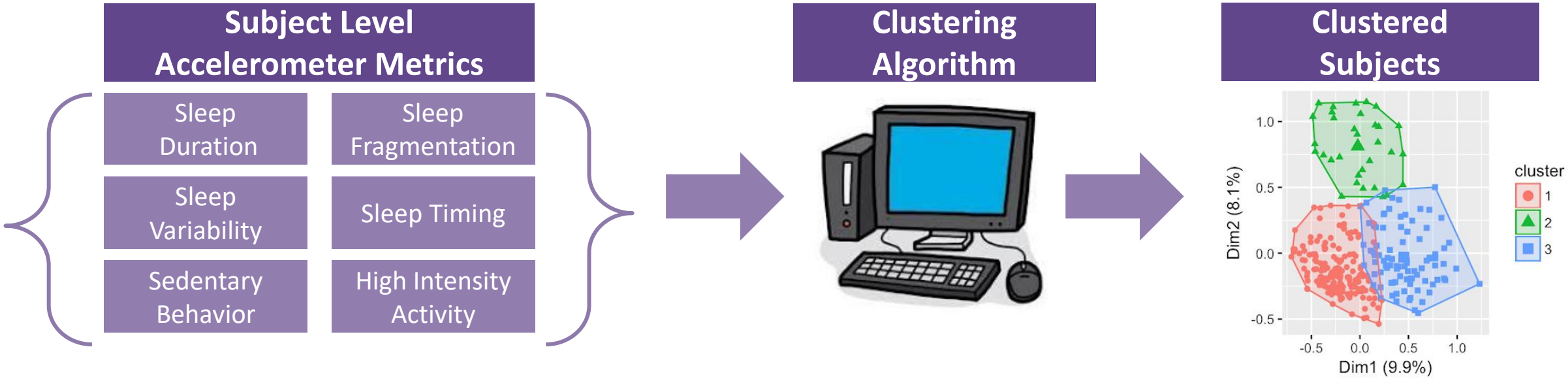
% of Wake Time at Each Activity Level

Aljahdali, A et al. "Sedentary patterns and cardiometabolic risk factors in Mexican children and adolescents: analysis of longitudinal data", *International Journal of Behavioral Nutrition and Physical Activity*, 2022. DOI:10.1186/s12966-022-01375-0

Phenotype Clustering

Analysis of multiple accelerometer metrics

- Research question
  - Can individuals be clustered into “phenotypes” based on variables of interest from accelerometers

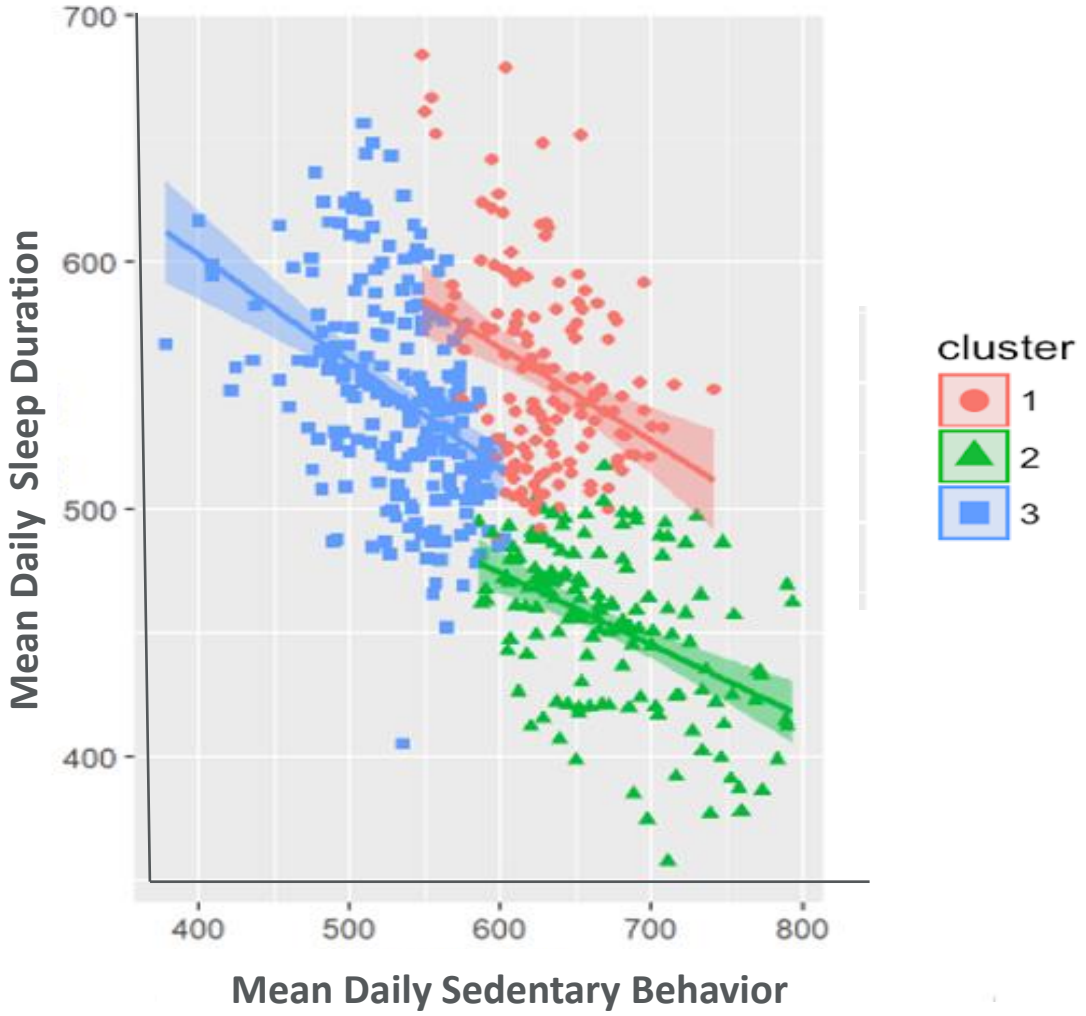


Source: Cluster Image

Phenotype Clustering

Analysis of multiple accelerometer metrics

- Assess cluster-specific subject characteristics, or “phenotypes”
  - Cluster 1:
    - ↑ Sedentary Behavior
    - ↑ Sleep Duration
  - Cluster 2:
    - ↑ Sedentary Behavior
    - ↓ Sleep Duration
  - Cluster 3:
    - ↓ Sedentary Behavior
    - ↑ Sleep Duration



Banker, M and Jansen, E et al. “Associations between sleep and physical activity behavior clusters and epigenetic age acceleration in Mexican adolescents”, Under Review at *Medicine & Science in Sports & Exercise.* 2024.

Functional Data Analysis

Data points as a Function rather than single value

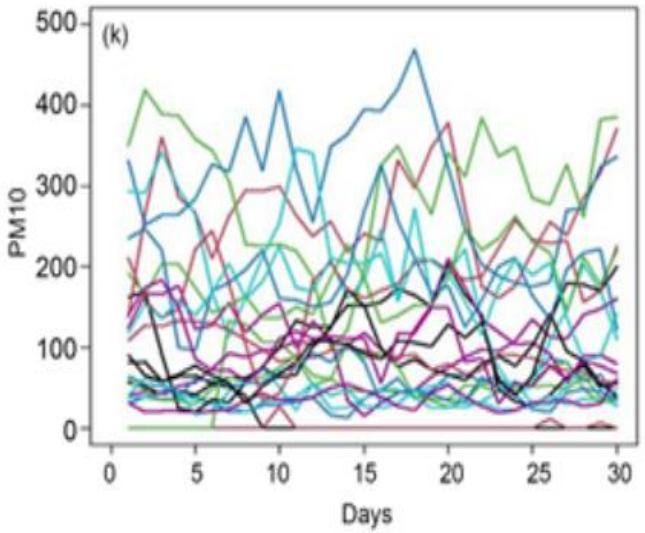
What is Functional Data?

- Observations on subjects that you can imagine as $X_i(t_i)$ where t_i is continuous
- i.e. Data in form of functions, images, shapes

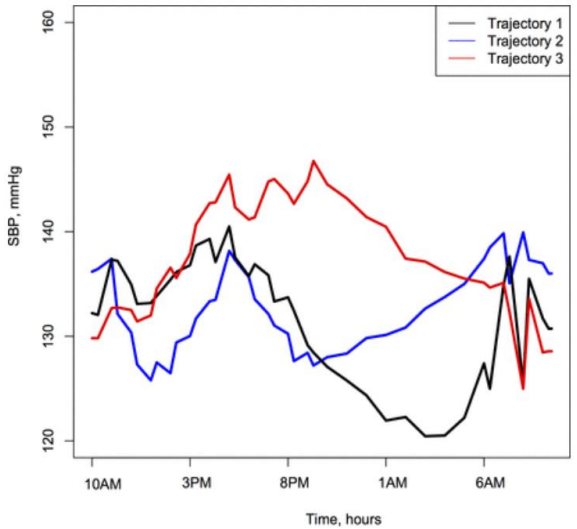
Characteristics of Functional Data

- High dimensional
- Ordered over time and space
- Recorded over a continuous domain (e.g. time)

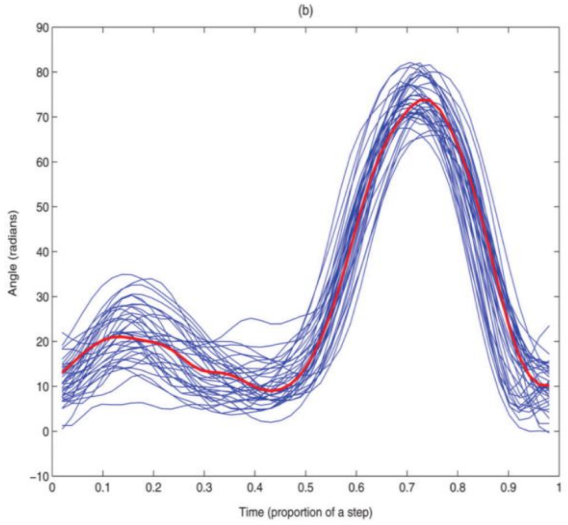
PM10 Exposure Curves



Blood Pressure



Knee Angle in Walking Gait

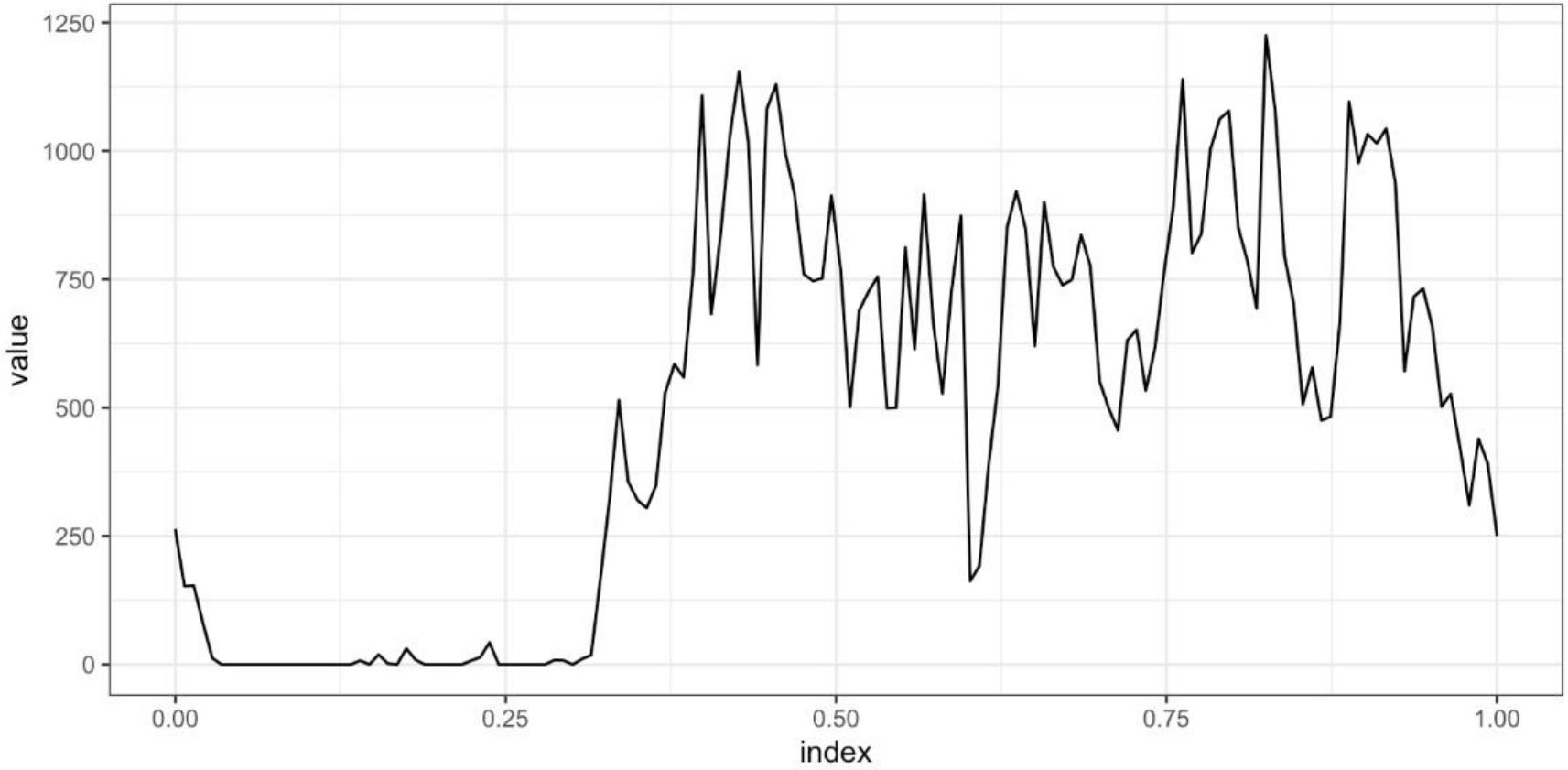


Source: PM10 Exposure Curve Source: Blood Pressure Image Source: Knee Angle Image

Functional Data Analysis

Getting Started

Functional Data from a single subject



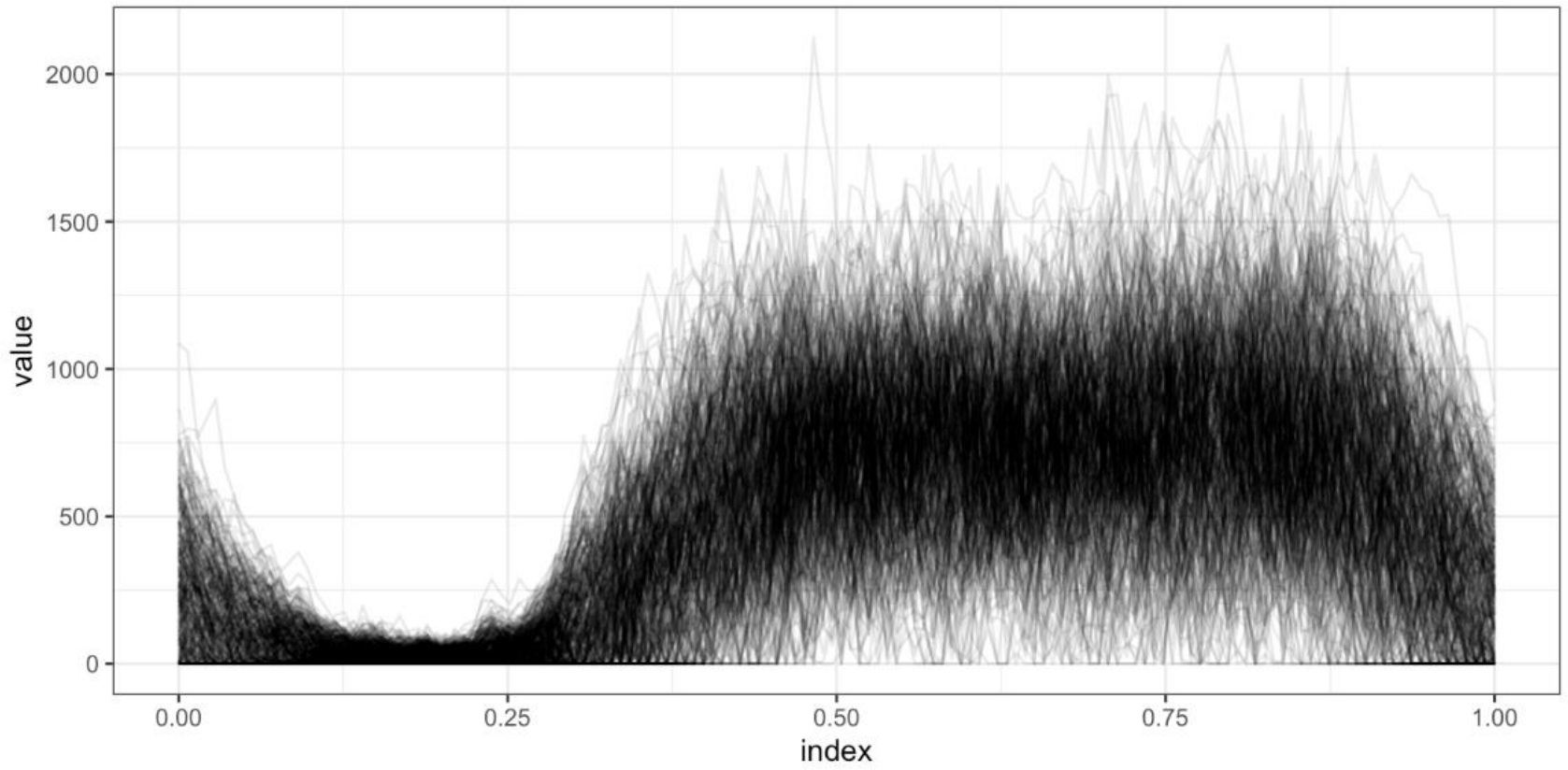
Note: Here, the full curve is the unit of analysis

Example from Jeff Goldsmith short course:
https://jeffgoldsmith.com/ICAMPAM/shortcourse_notation.html

Functional Data Analysis

Getting Started

"Spaghetti Plot": Functional Data from many subjects



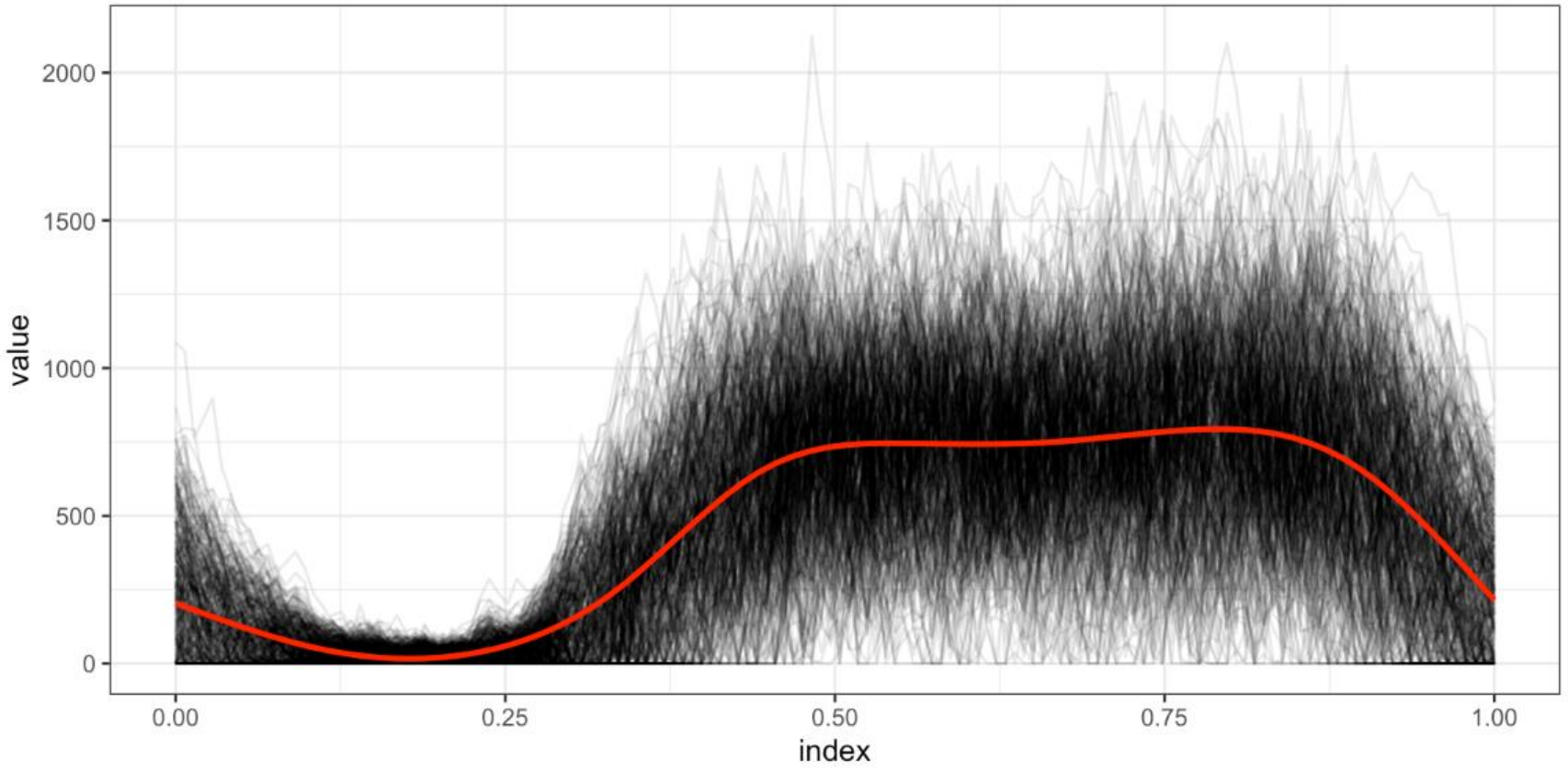
Note: Here, the full curve is the unit of analysis

Example from Jeff Goldsmith short course:
https://jeffgoldsmith.com/ICAMPAM/shortcourse_notation.html

Functional Data Analysis

Getting Started

"Spaghetti Plot": Functional Data from a many subjects



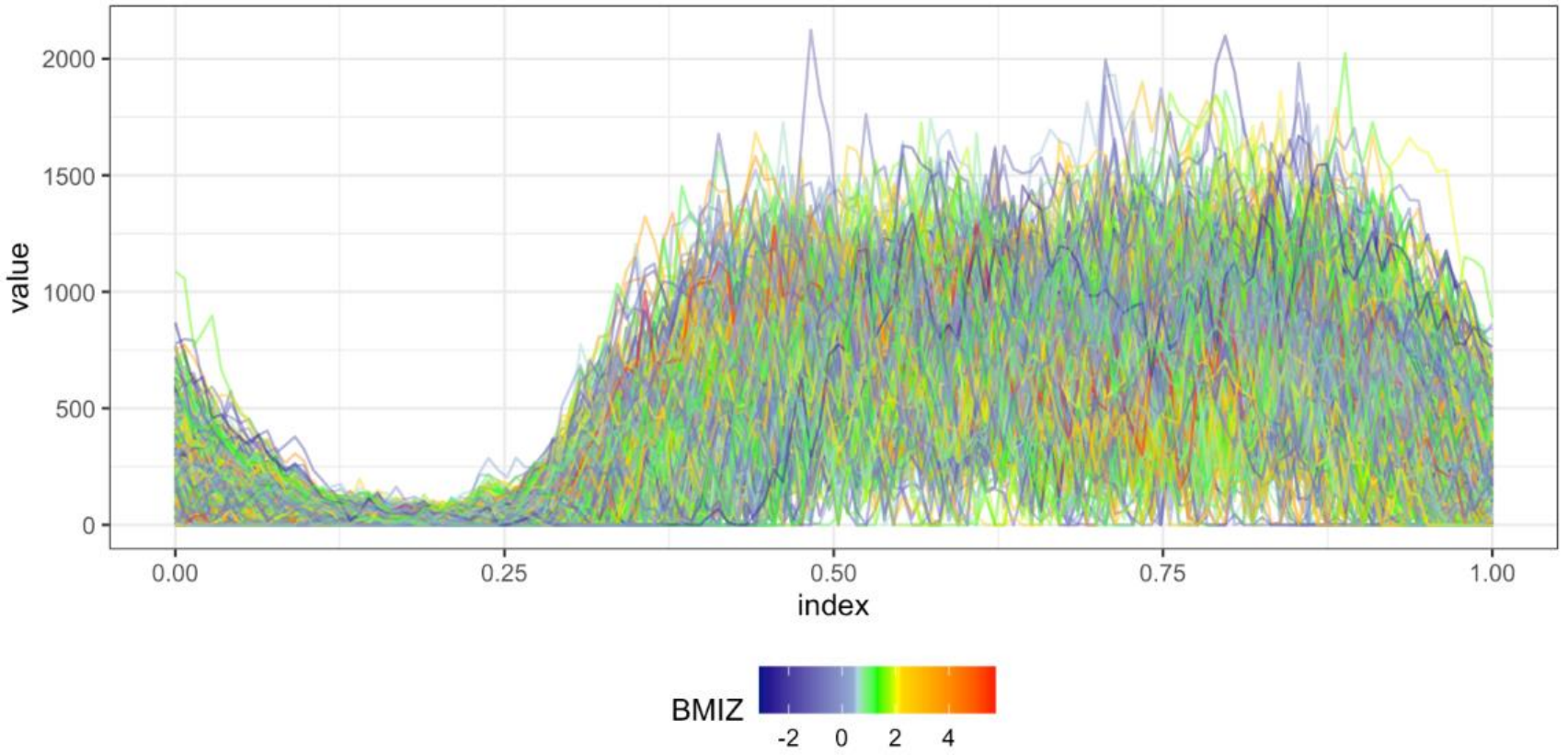
Note: Here, the full curve is the unit of analysis

Example from Jeff Goldsmith short course:
https://jeffgoldsmith.com/ICAMPAM/shortcourse_notation.html

Functional Data Analysis

Getting Started

"Rainbow Plot": Functional Data from a many subjects



Note: Here, the full curve is the unit of analysis

Example from Jeff Goldsmith short course:
https://jeffgoldsmith.com/ICAMPAM/shortcourse_notation.html

Scalar-on-Function Regression

Notation

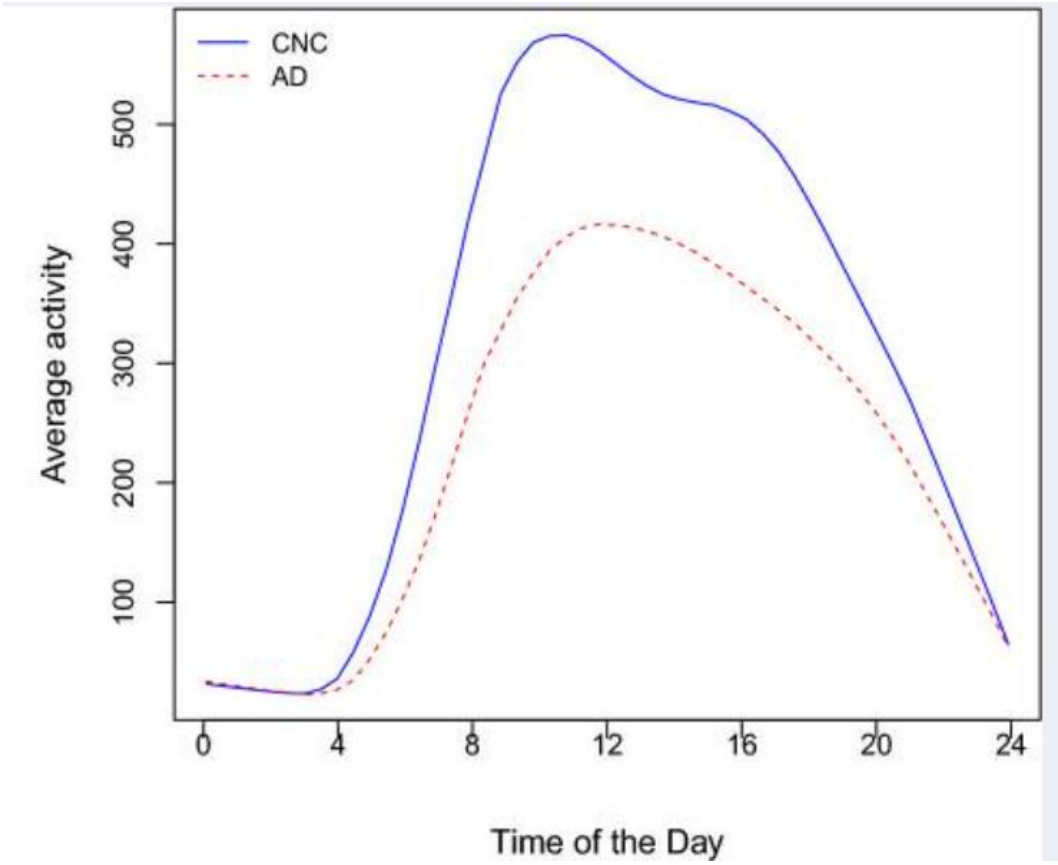
$$Y_i = \beta_0 + \int_T X_i(t)\beta(t)dt + \epsilon_i$$

- Y is scalar outcome of interest (single measurement)
  example: Health outcome of interest
- X is functional predictor/covariate
  example: PA intensity from accelerometer
- β is functional parameter we want to estimate
  example: functional effect of PA intensity on health outcome
- T is the continuous domain
  example: time

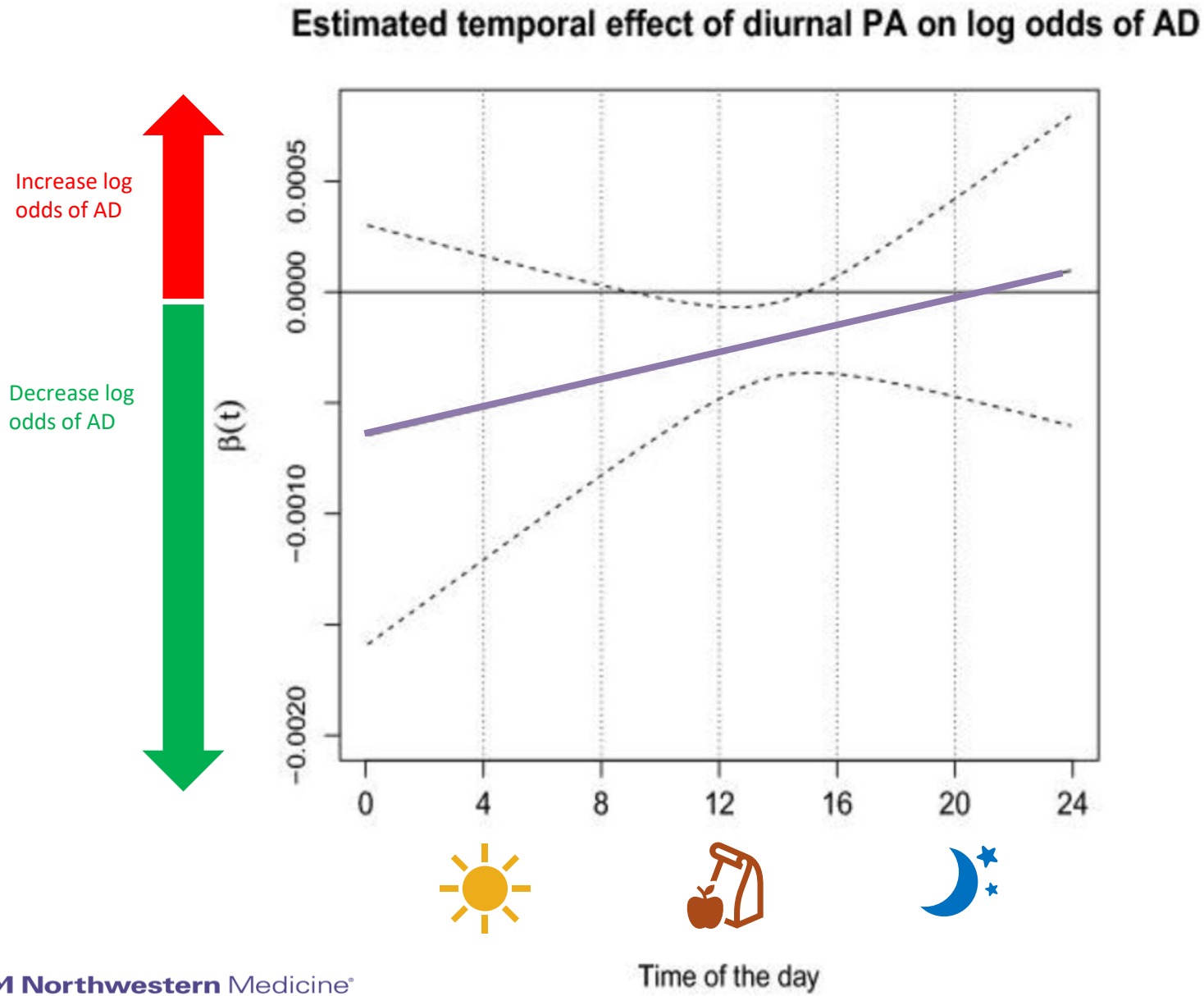
Scalar-on-Function Regression

Research Example

- Research Setup:
  - PA was monitored continuously for 7 days in older adults with mild Alzheimer's Disease (AD) and cognitively normal controls (CNC)
  - Can temporal PA profiles be used to differentiate between CNC and mild AD?
- Scalar-on-Function Regression
  - Model temporal aspects in wearable observations $X_i(t)$ over 24 hours in a day
- Results
  - Smoothed PA profiles for CNC and AD participants suggest different activity patterns



Ghosal, Rahul et al. "Scalar on time-by-distribution regression and its application for modelling associations between daily-living physical activity and cognitive functions in Alzheimer's Disease. "*Scientific Reports.* 2022.



- Functional regression coefficient β(t)

Result:

- Higher PA during morning hours (~ 10 a.m.–3 p.m.) is significantly associated with a lower odds of AD

Ghosal, Rahul et al. "Scalar on time-by-distribution regression and its application for modelling associations between daily-living physical activity and cognitive functions in Alzheimer's Disease." *Scientific Reports.* 2022.

Northwestern Medicine®
Feinberg School of Medicine

Wear it Well

Ending Thoughts

Ending Thoughts

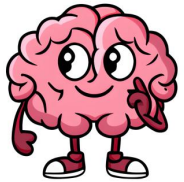
wow, so much data

I have so much **Data**!!

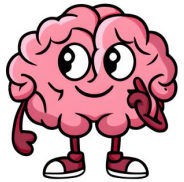
I have **SO MUCH** data!!

Ending Thoughts

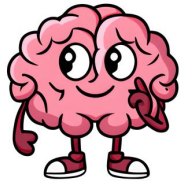
Start Planning Early



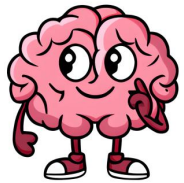
Processing raw data is **not trivial** … but there are some standard practices



There are **many metrics** you can glean from the continuous data … let your research question guide your variable creation



You have **options** in data analysis … from more standard to novel methodologies



The BCC is here to help! The earlier in the process, the better ☺

Ending Thoughts

Into The Unknown ….

Thank you!

… and Questions ☺