



evidation

New ways to measure health in everyday life

MAY 2019

Benjamin Bradshaw
Senior Health Data Scientist
bbradshaw@evidation.com

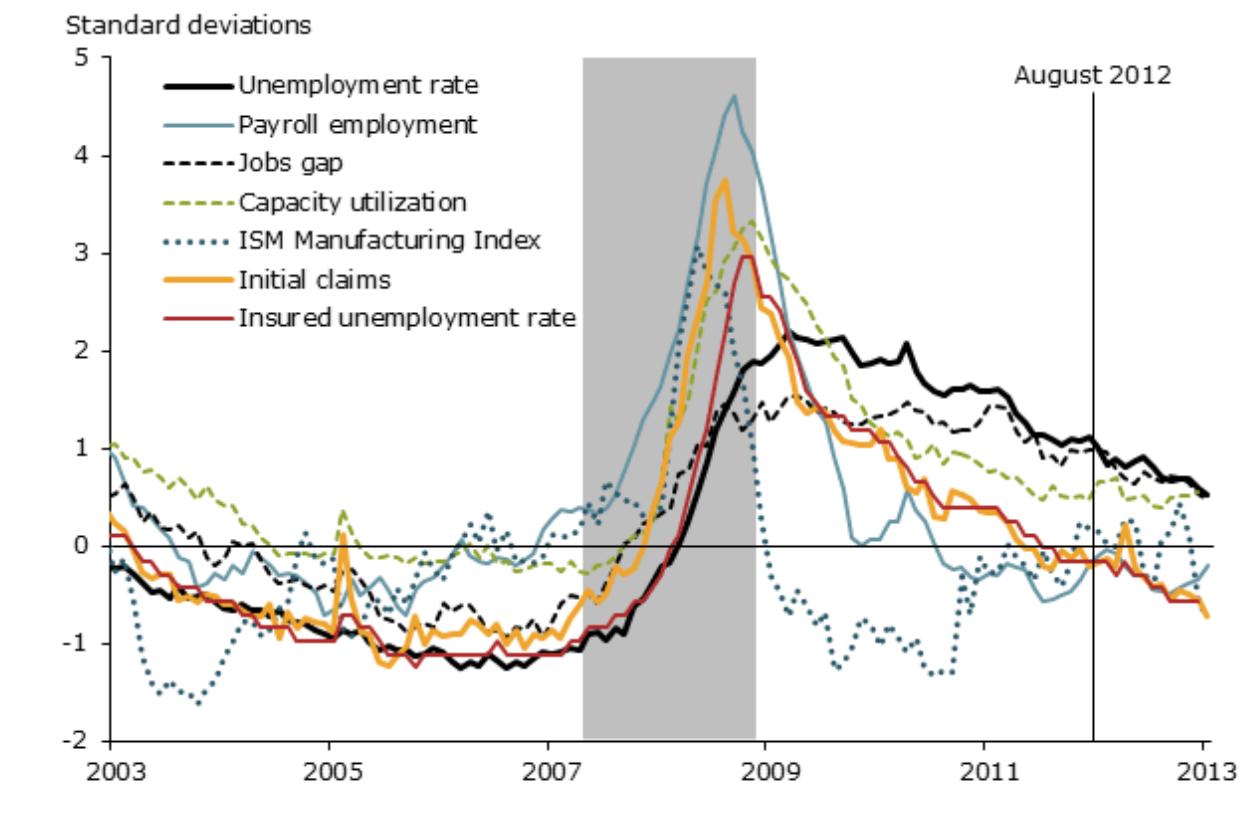
Goals of this talk

1. Give you an overview of Evidation Health: including we started, how we are structured, and what we do
2. Give an overview of data science at Evidation Health: How we use IoT data + EHR data to provide insights
3. Machine learning @ Evidation: Give a quick sampler of past projects at Evidation
4. Give you a quick notebook hands on case study that walks through the types of analysis we do at Evidation
5. Provide some practical tips for aspiring data scientists + Q & A (If time allows)
6. Data Science team background (if interesting & we have time)

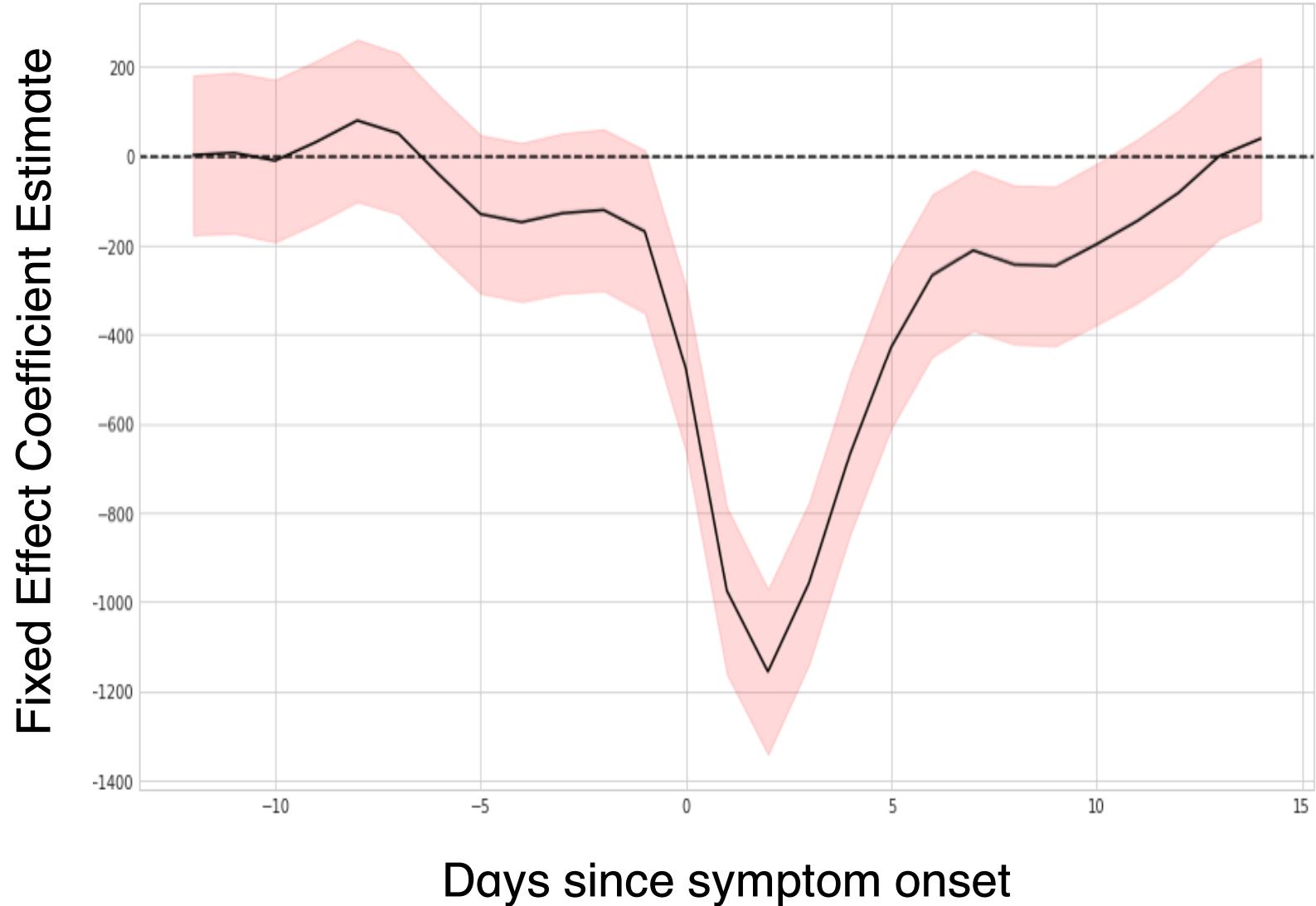
My background - Many paths to data science

- Currently: Senior Health Data Scientist @evidation
- Former Research Associate at the Federal Reserve Bank of San Francisco (applied microeconomics)
- Teach for America Alumni (Colorado Core 2013 - teaching high school statistics)

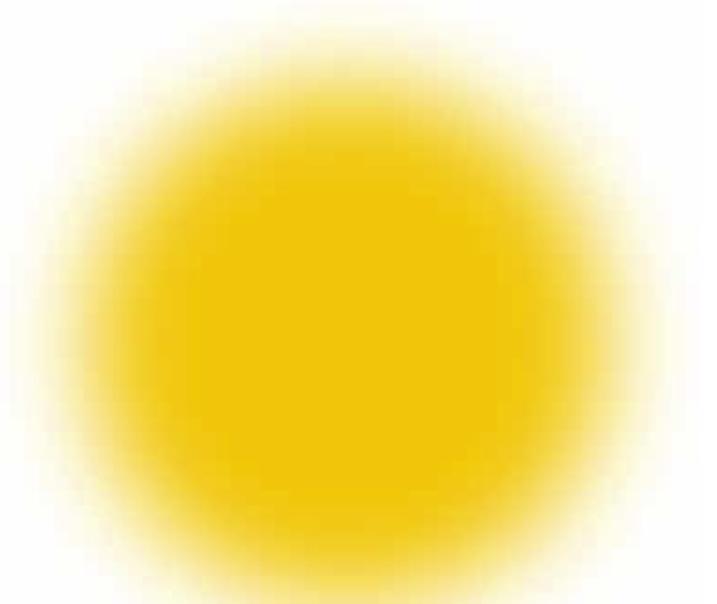
Figure 1
Paths of leading indicators for unemployment rate



Estimated mean daily decrease in step count



Company Overview



Evidation Health is a technology company
that helps the most innovative healthcare
companies understand everyday behaviors
to create better health outcomes.

Our Mission:
**We are enabling and empowering everyone to
participate in better health outcomes.**



evidation

Started on the hills of Santa Barbara, CA 6 years ago
3 offices, 120+ employees, \$30M+ raised
Work with Hospitals, Clinics, Universities, Insurance,
Pharma, Governmental Organization

Achievement

HOME

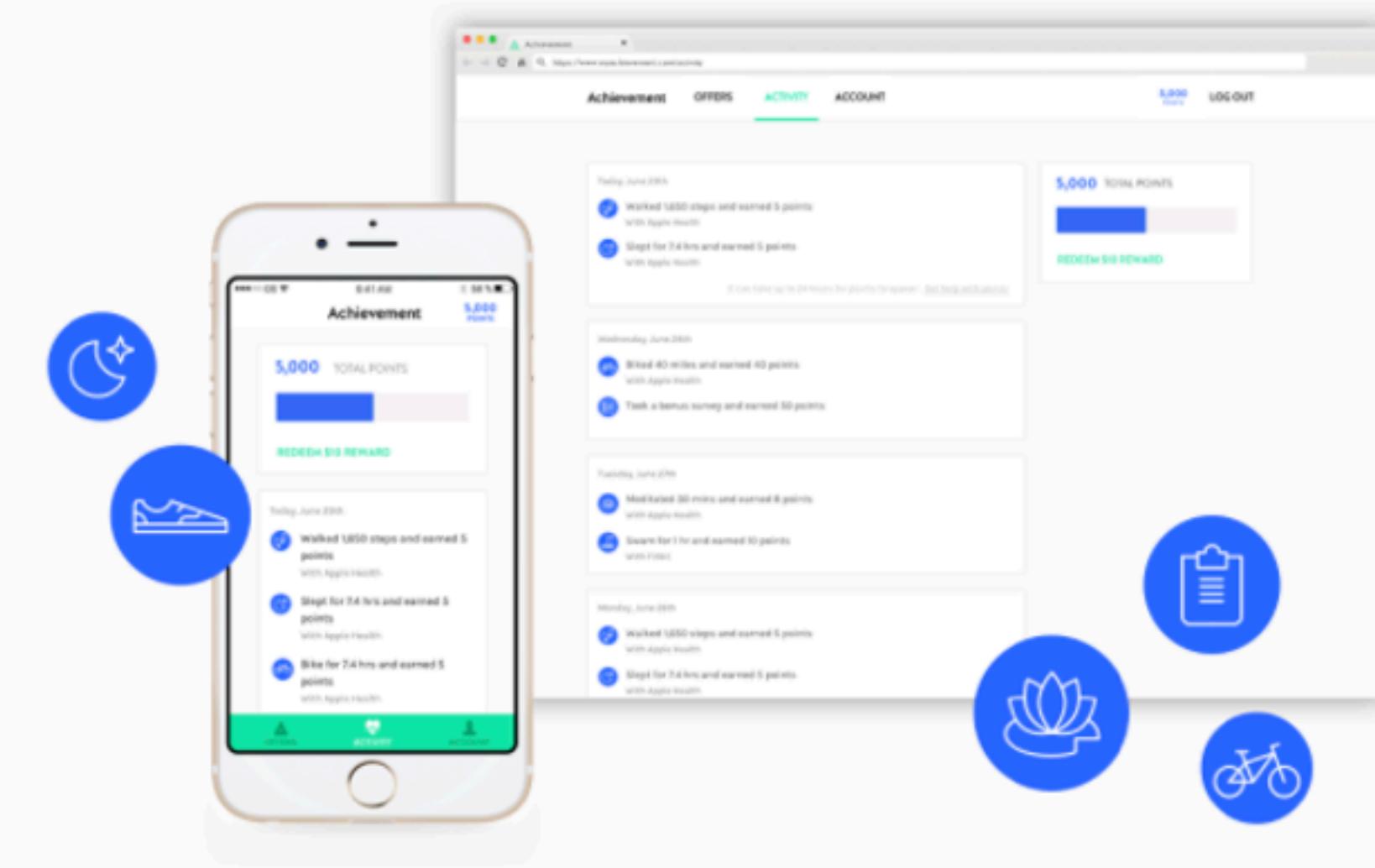
ABOUT

LOG IN

WE SERIOUSLY PAY YOU TO TRACK YOUR STEPS.

We pay you (via PayPal or direct deposits) for doing healthy things! Earn for tracking steps, sleep, meals, tweets & more! It's free to sign up for Achievement and start earning.

SIGN UP

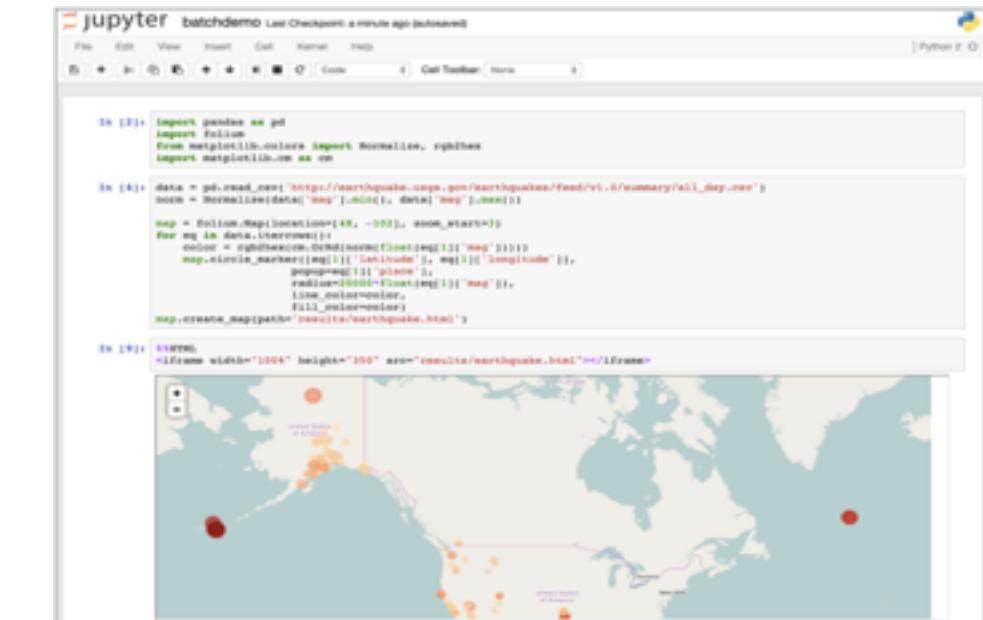
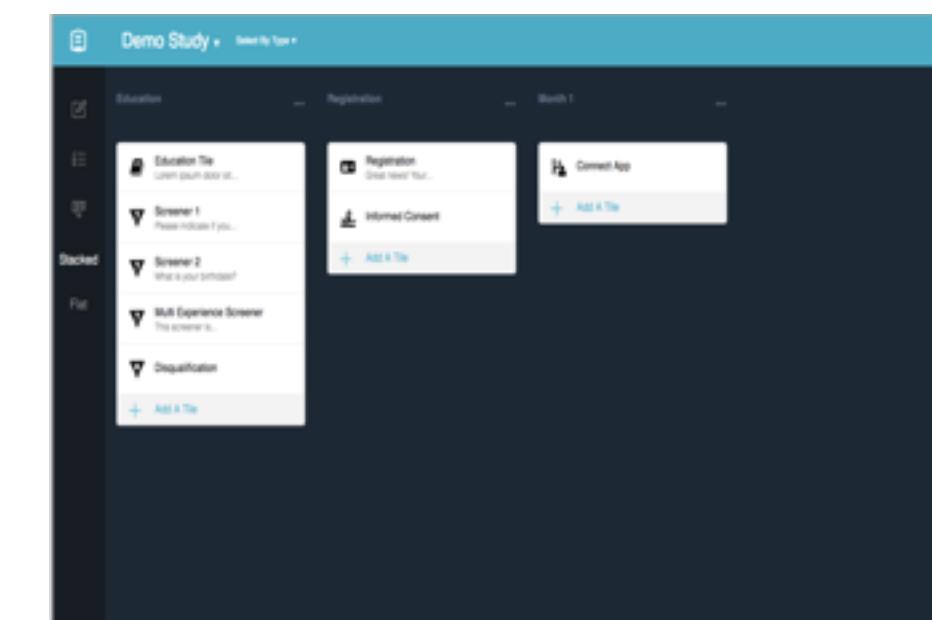
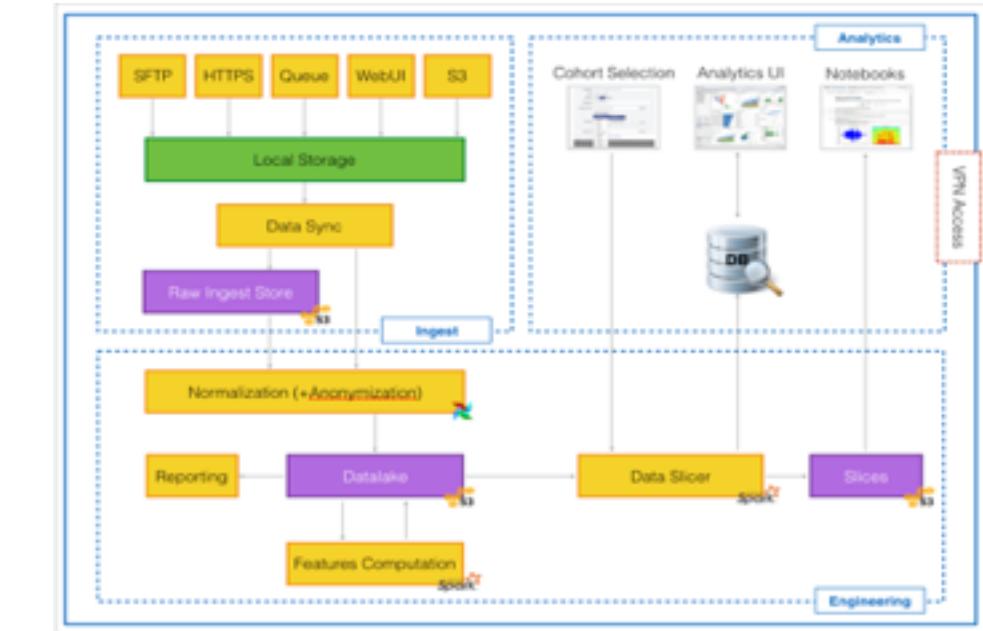
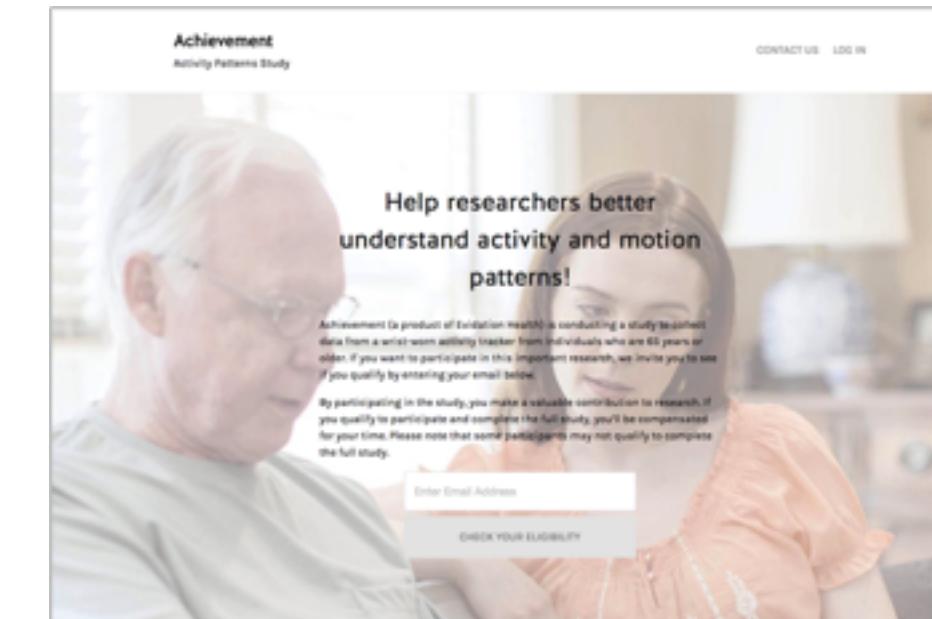


"My sister told me about it, and it seemed too good to be true at first. But when she said that she had already received \$20, I signed up. It was an easy and a monetary motivation really."

Achievement

3.2M+ Members | iOS/Android/Web

Evidation has three distinctive analytical bodies that drive insights



Data Connectivity

100+ Devices/Apps

Highly Scalable

Weather, Census, EMR

Study Management

Design, Management Multi-Site, Physical/Virtual Protocol + IRB Approval

Analysis Platform

HIPAA Compliant

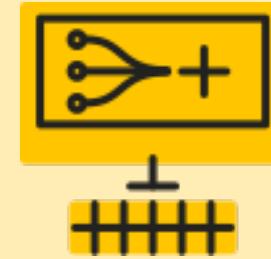
Exploration & Research

Python, R, MATLAB

Harnessing such massive data allows us to ask different questions about human health.

Can a conversation be used to detect the hidden beginnings of Alzheimer's?

DIGITAL BIOMARKER



Is passively measured accelerometry more sensitive to changes in ALS than the gold standard questionnaire?

NOVEL ENDPOINT



Can characterizing everyday life improve treatment for patients in chronic pain?

LIVING LAB



Projects (2017/2019, sample)

OBJECTIVE	POPULATION	SIZE
Measure correlation of hypoglycemic events and cognition in T1 diabetes	Type 1 Diabetes	25
Test impact of digital health intervention on lung function and medication adherence	Cystic fibrosis	40
Develop digital measure to identify and quantify migraine events	Migraine	50, 300
Validate a virtual, multi domain lifestyle intervention	Cognitive decline	85
Determine signs of cognitive decline using digital measures	Cognitive decline	140
Measure Multiple Sclerosis status and progression with passive trackers	Multiple sclerosis	200
Detect digital signals of early onset of Alzheimer's disease and progression	MCI / AD	375
Measure consumer preferences in allergy population to inform patient profiles	Allergy	500
Study relationship between symptoms, resource utilization, and behavior traits in individuals reporting severe anxiety	Anxiety	1,100
Identify digital and behavioral phenotypes for lower gut health	Gastrointestinal	1,500
Detect digital signals in chronic pain	Chronic pain	10,000
Create phenotypes based on behaviors and understand relationship between flu vaccinations and type 2 diabetes management	Type 2 diabetes	54,000
Evaluate cardiovascular outcomes related to atrial fibrillation interventions	Cardiovascular disease	150,000
Examine association between activity patterns and adherence	Hypertension and diabetes	300,000

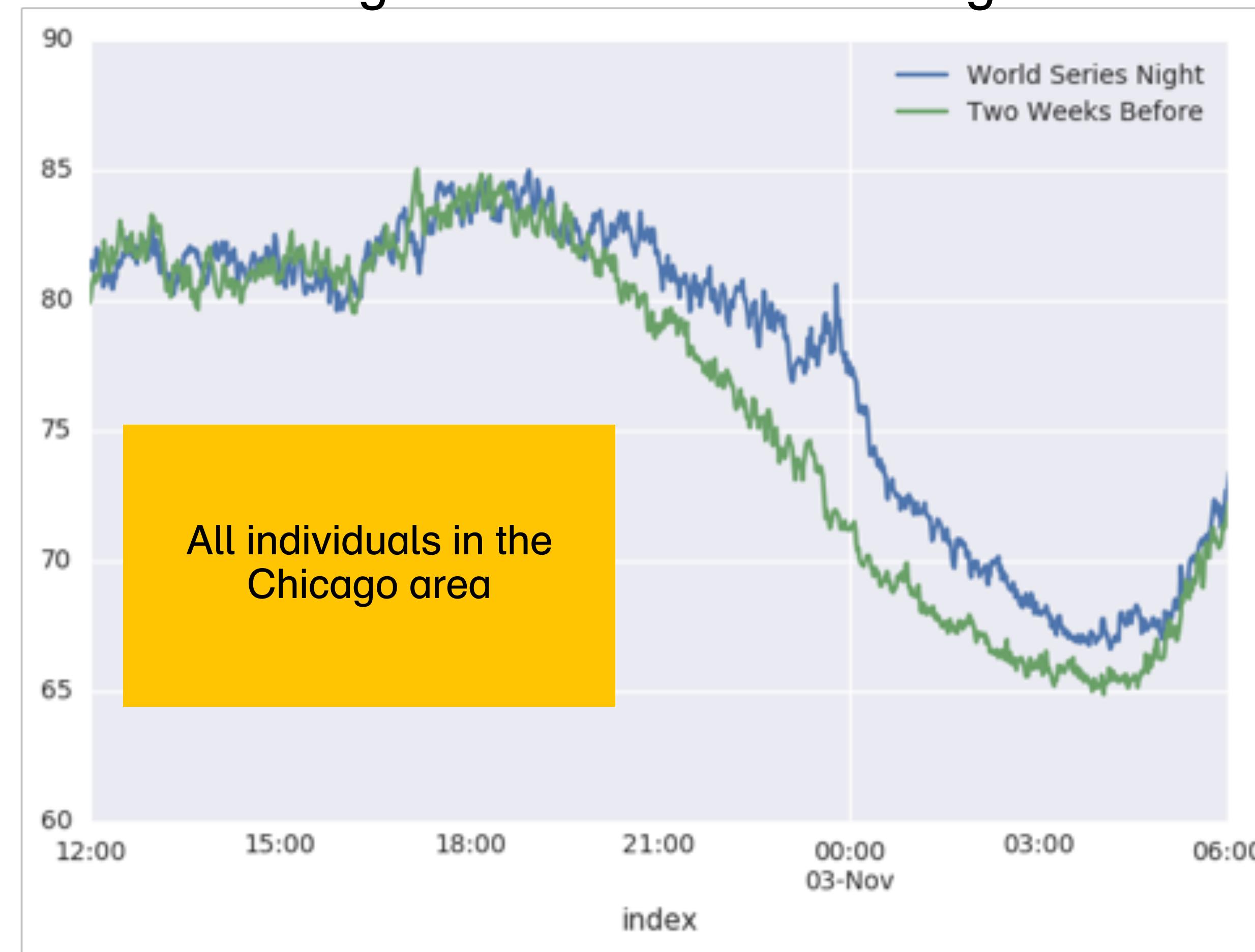
Collaborators (sample)

- Dr. Rich Milani, MD, Ochsner Health (co-author)
- Dr. John Torous, MD, Harvard (co-author)
- Prof. Mitesh Patel, MD, UPenn Med (co-author)
- Dr. David Kerr, MD, Sansum Diabetes (co-author)
- Prof. Nigam Shah, MD, Stanford (collaborator)
- Prof. Niteesh Choudry, MD PhD, Harvard/HMS (co-author)
- Prof. David Sontag, PhD, MIT (co-author)
- Prof. Susan Murphy, PhD Harvard (grant proposal co-authors)
- Prof. Francesca Gino, PhD, Prof. Leslie John, PhD, Harvard/HBS (co-authors)
- Prof. Katy Milkman, PhD, UPenn/Wharton (co-author)
- Dr. Sue Zbikowski, PhD, Humana, Inc (co-author)
- Dr. Sandrine Sansom, PhD, Sanofi (co-author)
- Dr. Diego Silva, MD, Novartis (co-author)

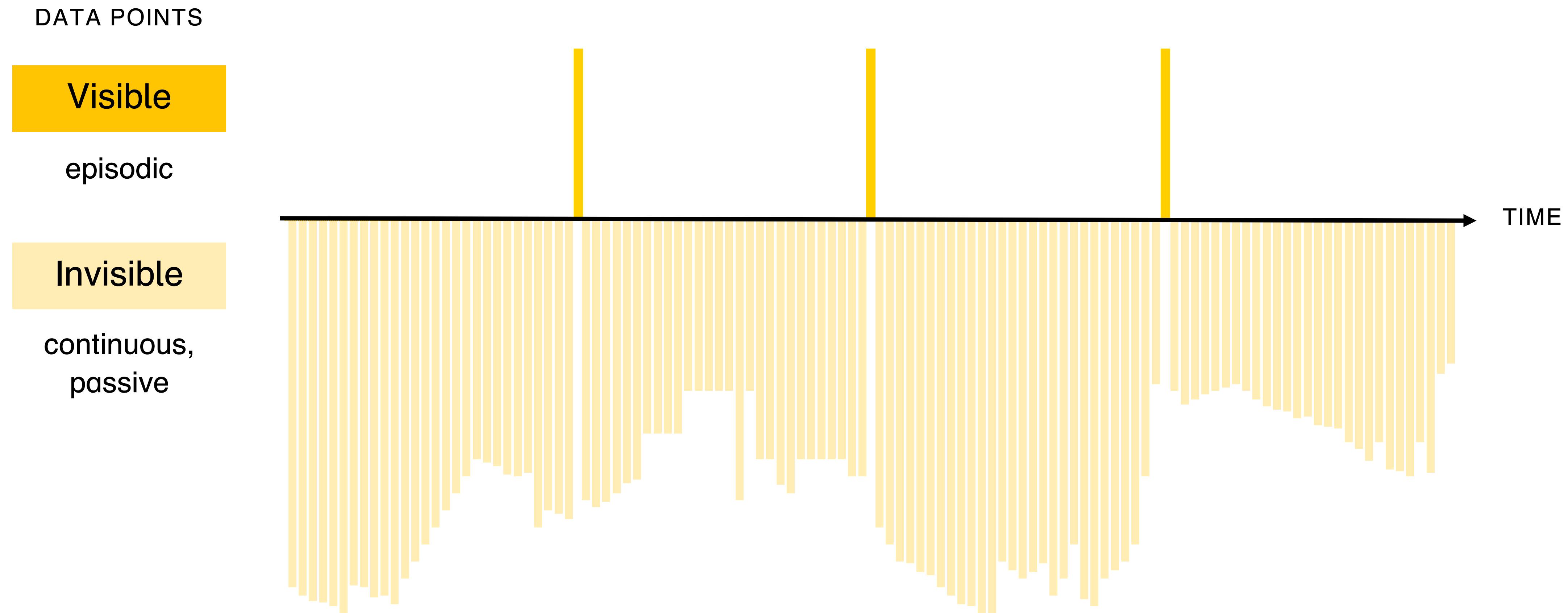
Data Science @
Evidation Health

Lots of Data + Simple Algorithms = Great Insights.

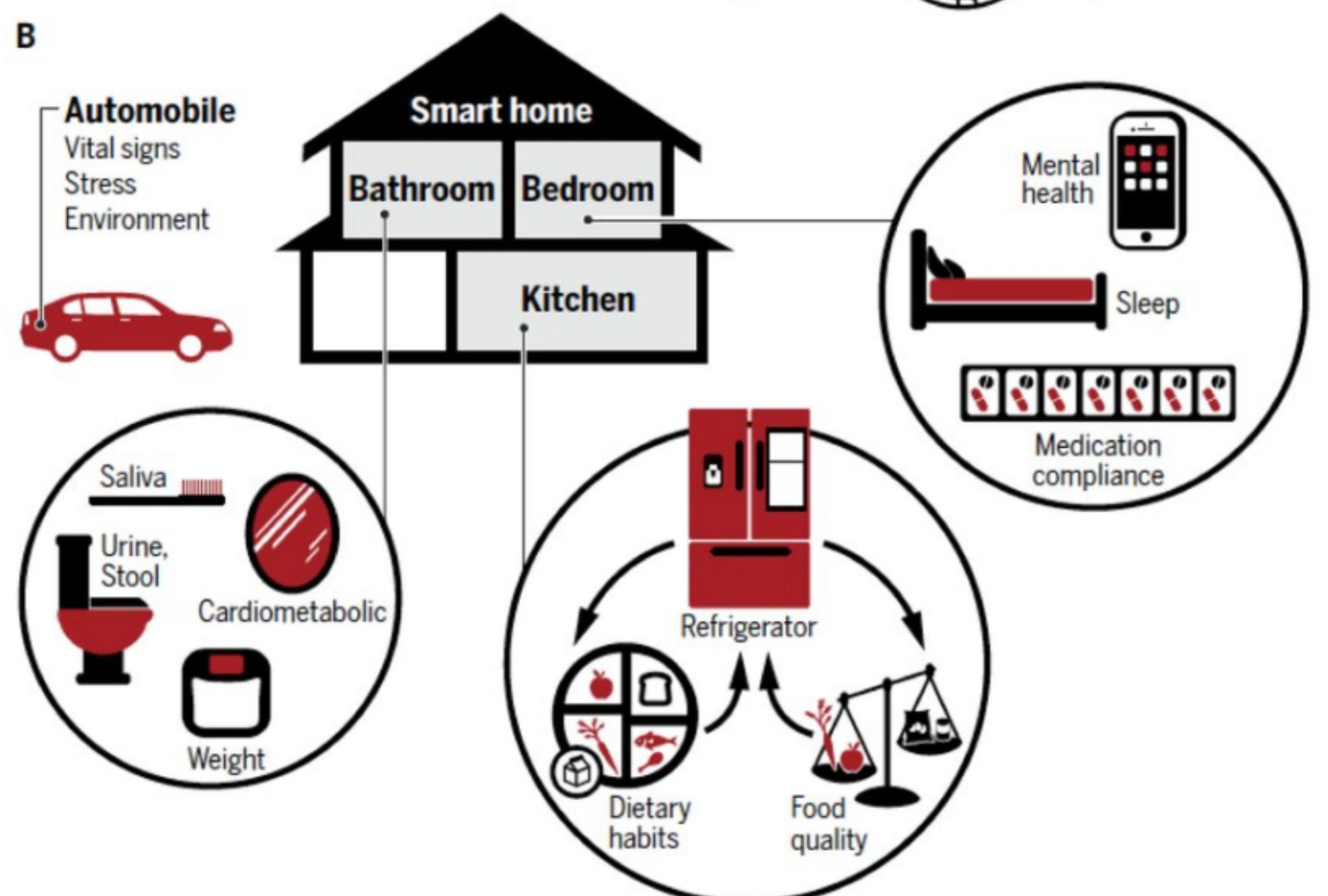
Population average minute level HR
during the World Series 2016 night



Patients and their outcomes have historically been characterized using limited, visible-to-the-system data sets.



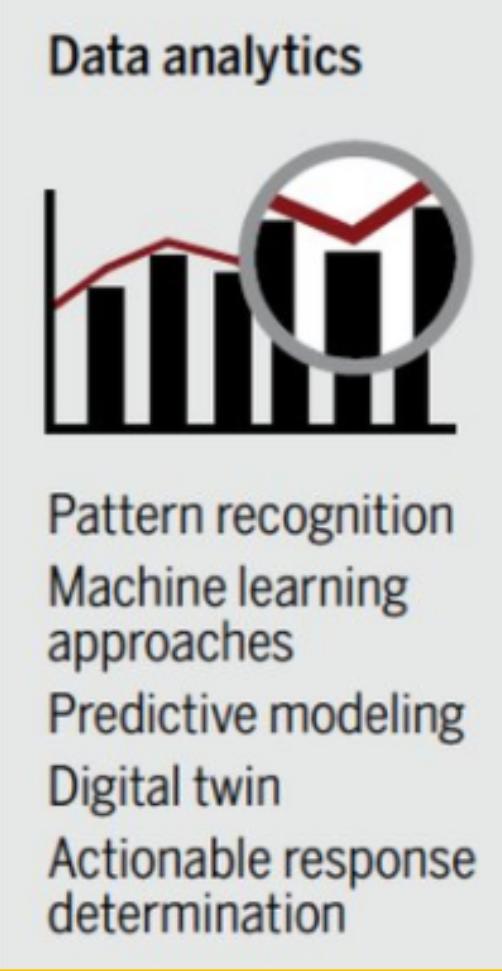
PRECISION HEALTH

Toward achieving precision healthSanjiv Sam Gambhir,^{1,2,3,4*} T. Jessie Ge,^{1,5} Ophir Vermesh,¹ Ryan Spitzer^{1,4}**Risk assessment at all stages of life**

Family history
Genetic screening
Socioeconomic factors

Customized monitoring

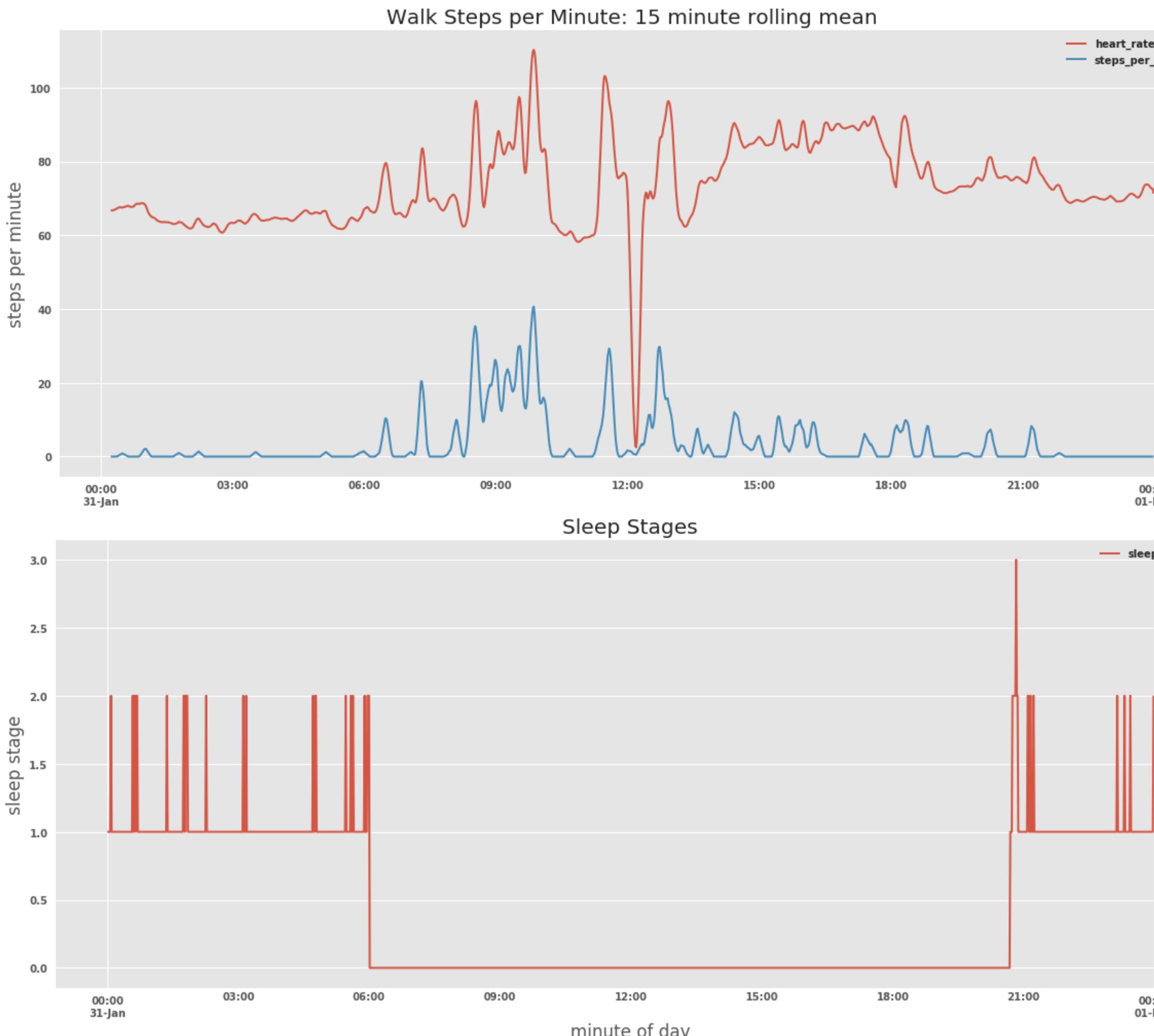
Implantables
Wearables
Smart home
Biomarkers
Imaging

**Learning and adaptation**

Evidation Health uses a variety of tools to ingest, process, store, and analyze data from hundreds of different data sources



Behavioral features give fine grained individual level context along a variety of informative dimensions



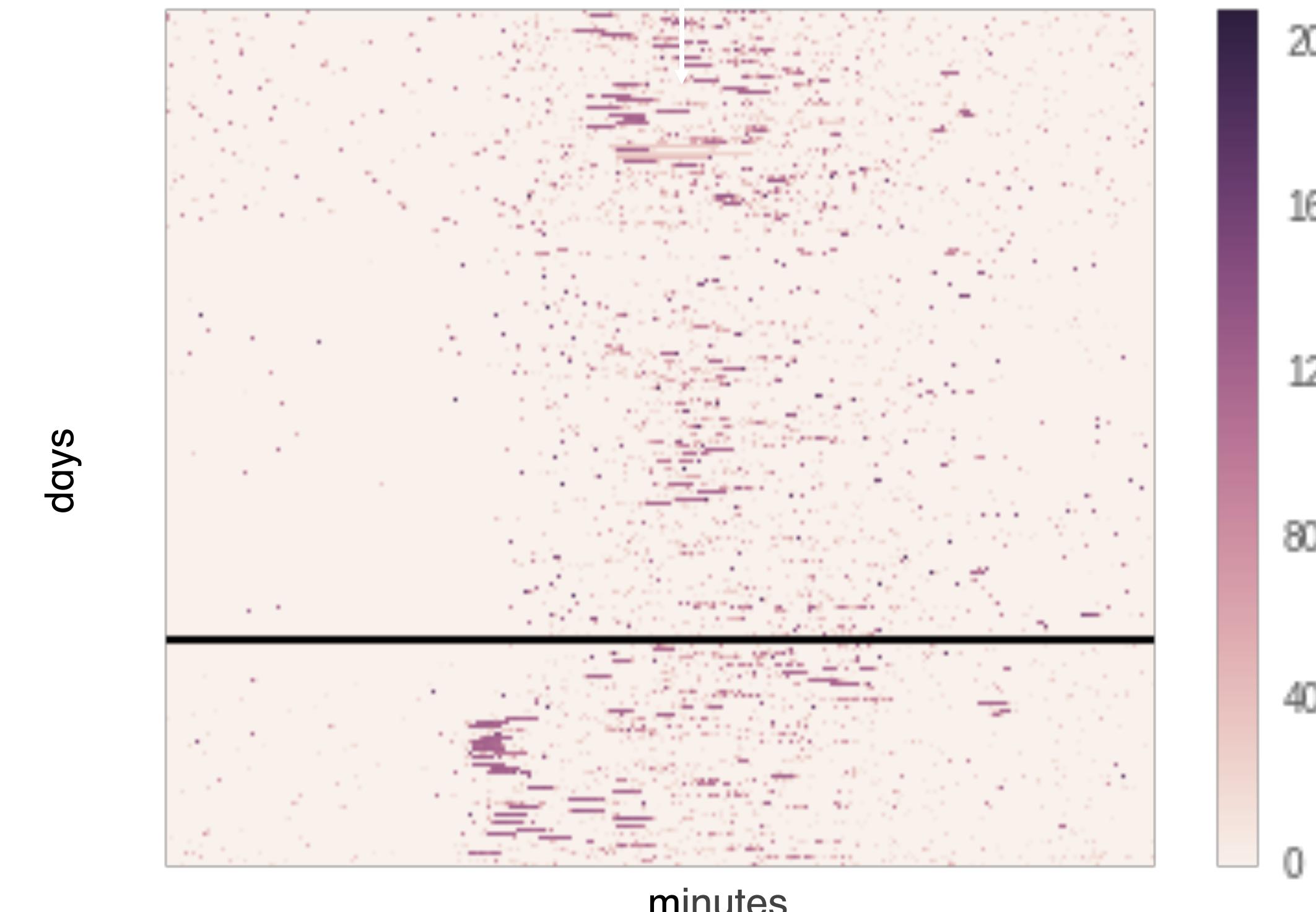
Contextual-Behavioral features

Based on prototype model and assumptions about Landmark's operations

VARIABLE	VALUE
TIME OF FIRST STEP	613 AM
MAXIMAL 10 MINUTE ROLLING STEPS	41.6 SPM
TOTAL STEPS WALKED DURING DAY	5230 STEPS
Steps	SLEEP REGULARITY INDEX
	NAP COUNT
	NUMBER OF "AWAKE" REGIONS
	SLEEP EFFICIENCY
Sleep	RESTING HEART RATE
	(05 TH , 95 TH) PERCENTILE OF HEART RATE
Heart Rate	62 BPM
	(61, 96) BPM

Over all managed populations, Evidation has collected >1 billion medical records & >10 trillion IoT data points

Provider Name	From Date	Thru Date	Paid Date	Total Charge	Total Paid	CPT CPT Code	Moc	CPT Description	Prime ICD9	ICD9Desc
WALGREEN	12/27/09	12/27/2009	12/27/2009	0.00	0.00	RX		TEMODAR 100 MG CAPSU		
WALGREEN	12/27/09	12/27/2009	12/27/2009	0.00	0.00	RX		TEMODAR 20 MG CAPSUL		
WALGREEN	12/27/09	12/27/2009	12/27/2009	0.00	0.00	RX		TEMODAR 5 MG CAPSULE		
WALGREEN	12/22/09	12/22/2009	12/22/2009	1.60	1.60	RX		PROMETHAZINE 25 MG T		
DAVID BRACH MD	11/24/09	11/24/2009	12/10/2009	588.00	449.00	77301	26	NTSTY MODUL RADTHX P	191.4	MALIGNANT NEO OCCIP
DAVID BRACH MD	11/24/09	11/24/2009	12/10/2009	588.00	449.00	77301	26	NTSTY MODUL RADTHX P	191.4	MALIGNANT NEO OCCIP



Machine Learning



CASE STUDY

Flu Vaccination Predictive Model

(Research@Evidation et al., in submission to JBI)

OBJECTIVE

Conduct a retrospective, observational cohort study on trends in influenza vaccination to inform a prospective study aimed at increasing vaccination rates through mobile messaging promotion.

We cast the problem as a predictive modeling exercise: from all the available data of the individual, predict whether they're going to vaccinate next year

**Get the
FLU  SHOT
not the flu!**

STUDY SETUP

- The behavioral characteristics that drive propensity to vaccination against flu on 521,978 members of a fully insured commercial population.
- Data from 2014-2015
- Goal: predict whether an individual is going to get a flu shot in 2016
- Data included utilize:
 - Claims, pharma benefits, eligibility data
 - Contextual information (census, weather)
 - Minute level data on sleep, steps, heart rate collected from activity trackers, linked at the patient level



CASE STUDY

Flu Vaccination Predictive Model: Results

(Research@Evidation et al., in submission to JBI)

Data 2014-2015, .5M people

679 Features

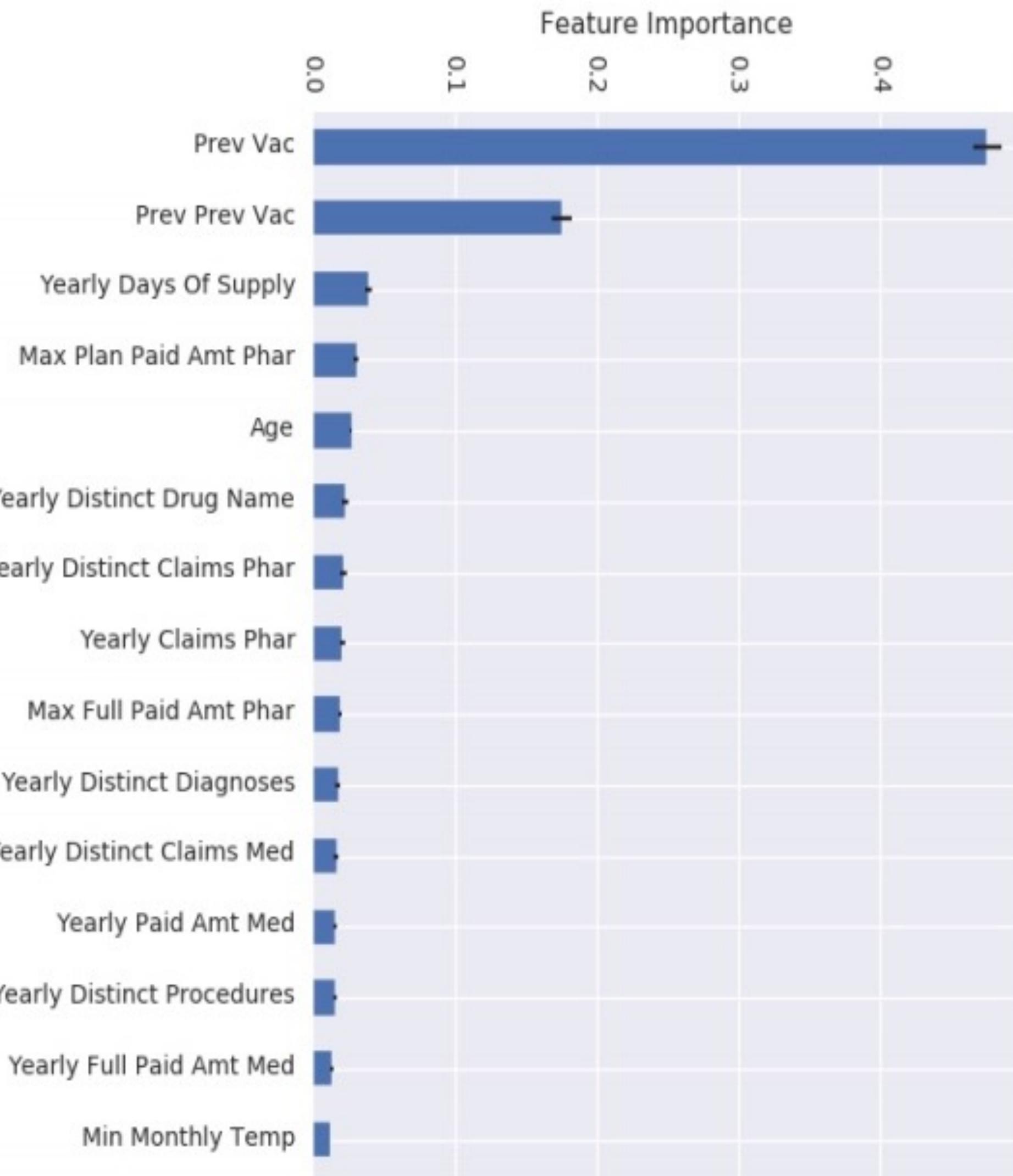
Lasso Feature Selection

236 Features

Random Forest Classifier

Who will get a flu shot in 2016?

0.83 AUROC



SUB-COHORT OF ACTIVITY TRACKERS
ABLATION ANALYSIS

PV	MC	PA	PV+PA	PV+MC
AUC	0.77	0.64	0.58	0.78

PV = Previous Year Vaccination

MC = Medical Claims

PA = Physical Activity



CASE STUDY

Voice & Speech Biomarker for Mild Cognitive Decline

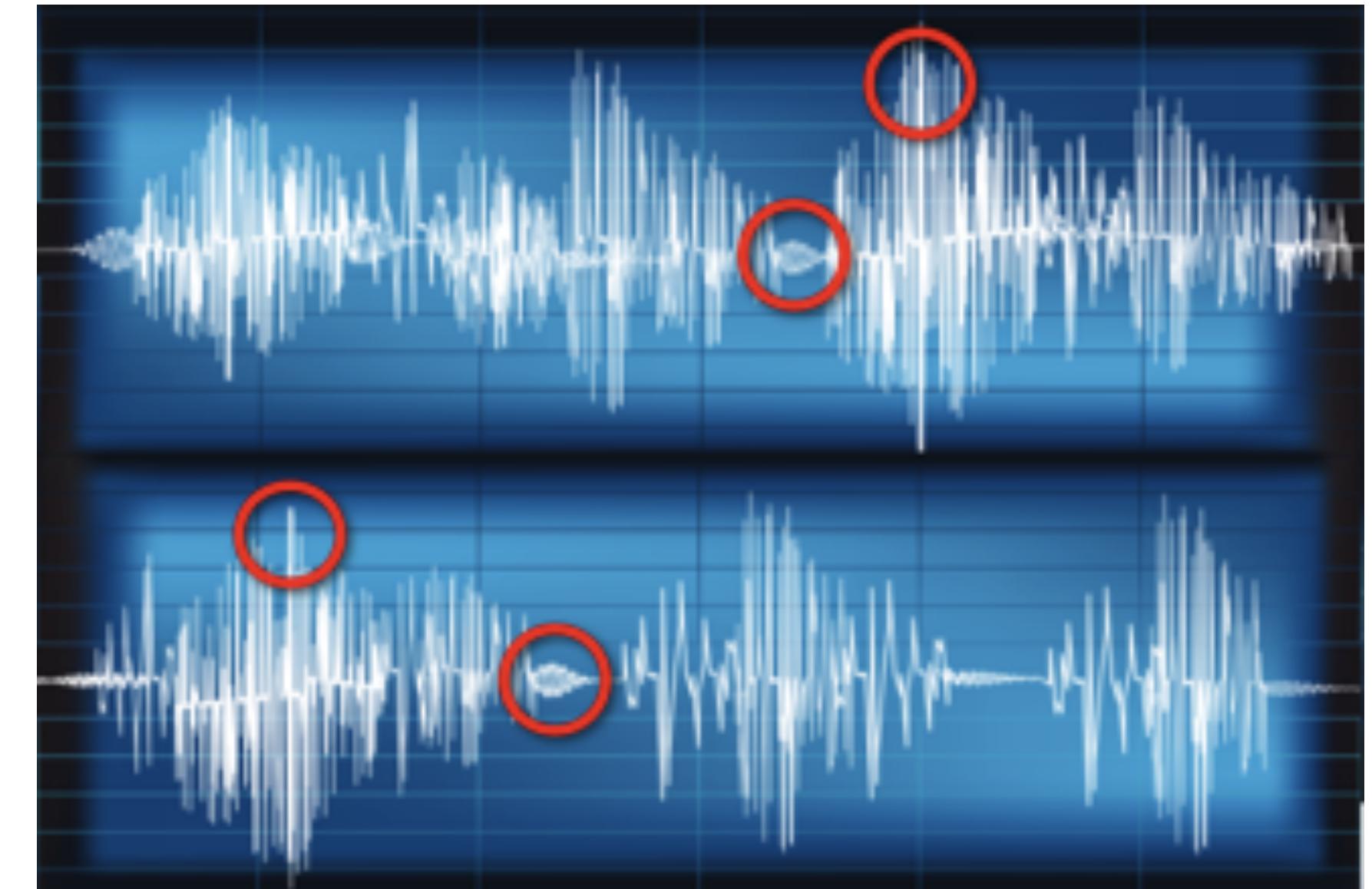
(Research@Evidation et al., Submitted)

DATA SOURCES

- Large National Cohort Study, 200+ neuro psych exams
- 50+ years of longitudinal medical + voice files

MODELING SETUP

Use speech transcription & audio biomarkers to detect cognitive decline over time.



RESULTS DETECTING DEMENTIA

- Demographics: 0.52 AUC
- Simple acoustic features: 0.81 AUC
- Acoustic features + POS features: 0.91 AUC

IMPORTANT FEATURES

- Time spent answering specific questions
- Uncertainty in the automatic transcription
- Use of stop-words ('the', 'and', 'is',...)
- Variability of vocal tones

Case Study: Influenza Surveillance Using Wearable Mobile Health Devices

Case Study: Can commercial wearables be used to improve flu surveillance and estimate the burden of influenza?

Why do we care about influenza?

- The CDC estimates that during the 2017/18 Flu season
48.8 million people contracted the flu

- 22.7 went to a health care provider for the flu

- 959,000 people were hospitalized due to influenza

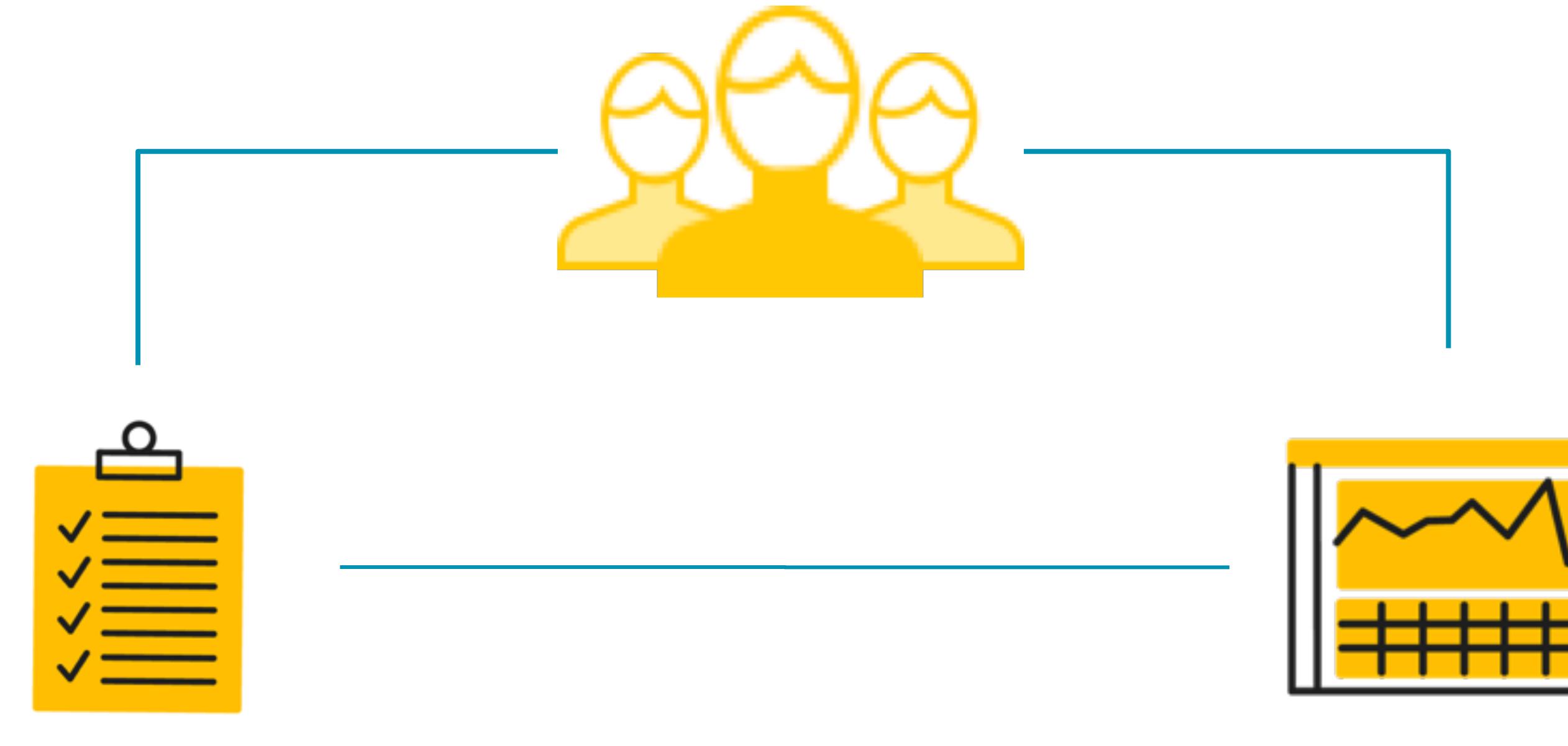
- 79,400 people died from influenza

Influenza nowcasting and forecasting

What is currently used to estimate and forecast disease burden?

- Currently the CDC uses ILINet and FluSurvNet, a large network of outpatient clinics and hospitals that send weekly reports of the number of patients with ILI (influenza like illness) symptoms, by age group to the agency
- These reports are aggregated to estimate flu burden as well as forecast flu in the short term, but there is typically a 2 week lag in reporting, and doesn't take into account non-medically attended ILI
- Google Flu trends attempted to use aggregated search queries related to ILI symptoms in order to “nowcast” influenza, the project failed spectacularly due to the feedback loop present in the system

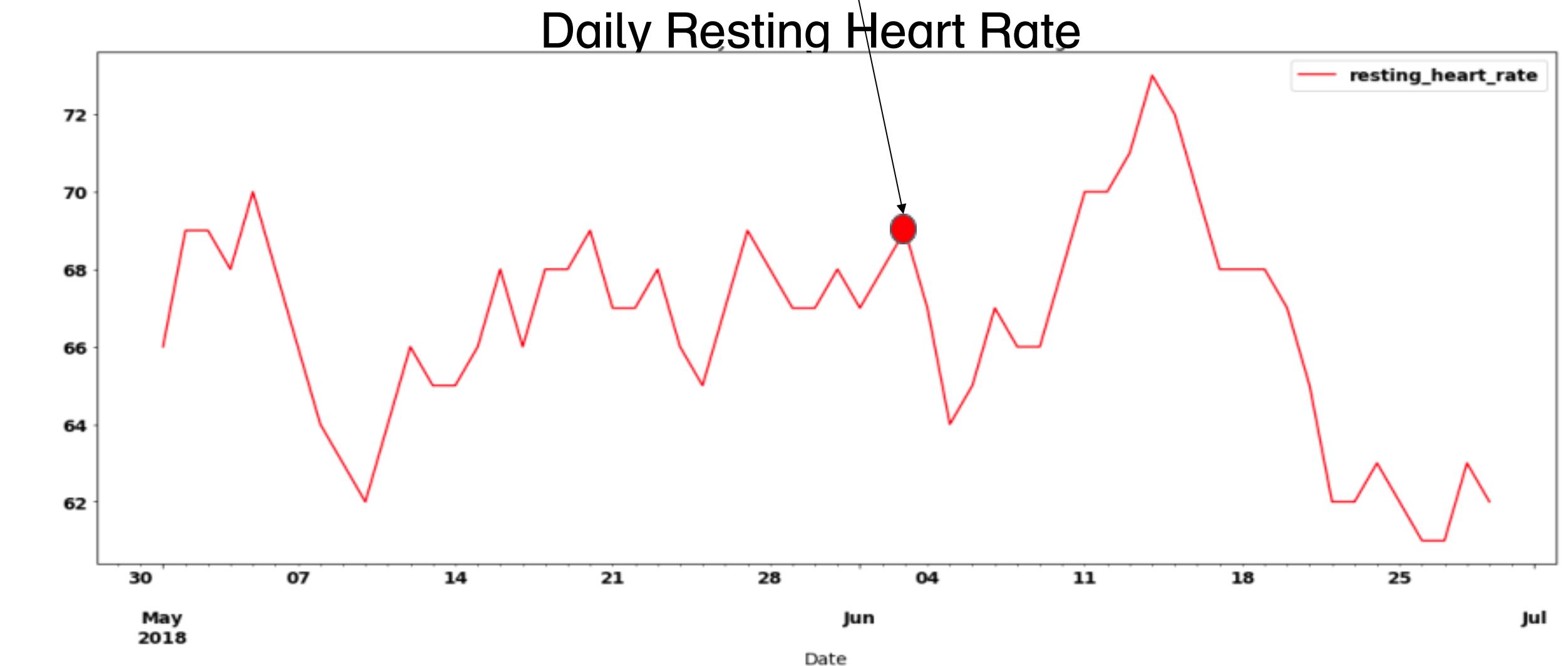
Designing an analysis: Combining self reported ILL events with passively measured behavioral and physiological data



DEMOGRAPHICS	MEDICAL	BEHAVIORAL
Age	Diagnoses	Step counts
Gender	Procedures	Average sleep length
ZIP Code	Utilization	HEART RATE
Employment status	Patient Cost	Weigh-in frequency
		Health apps installed

High resolution data gives powerful contextual information about participant physiology and behavior

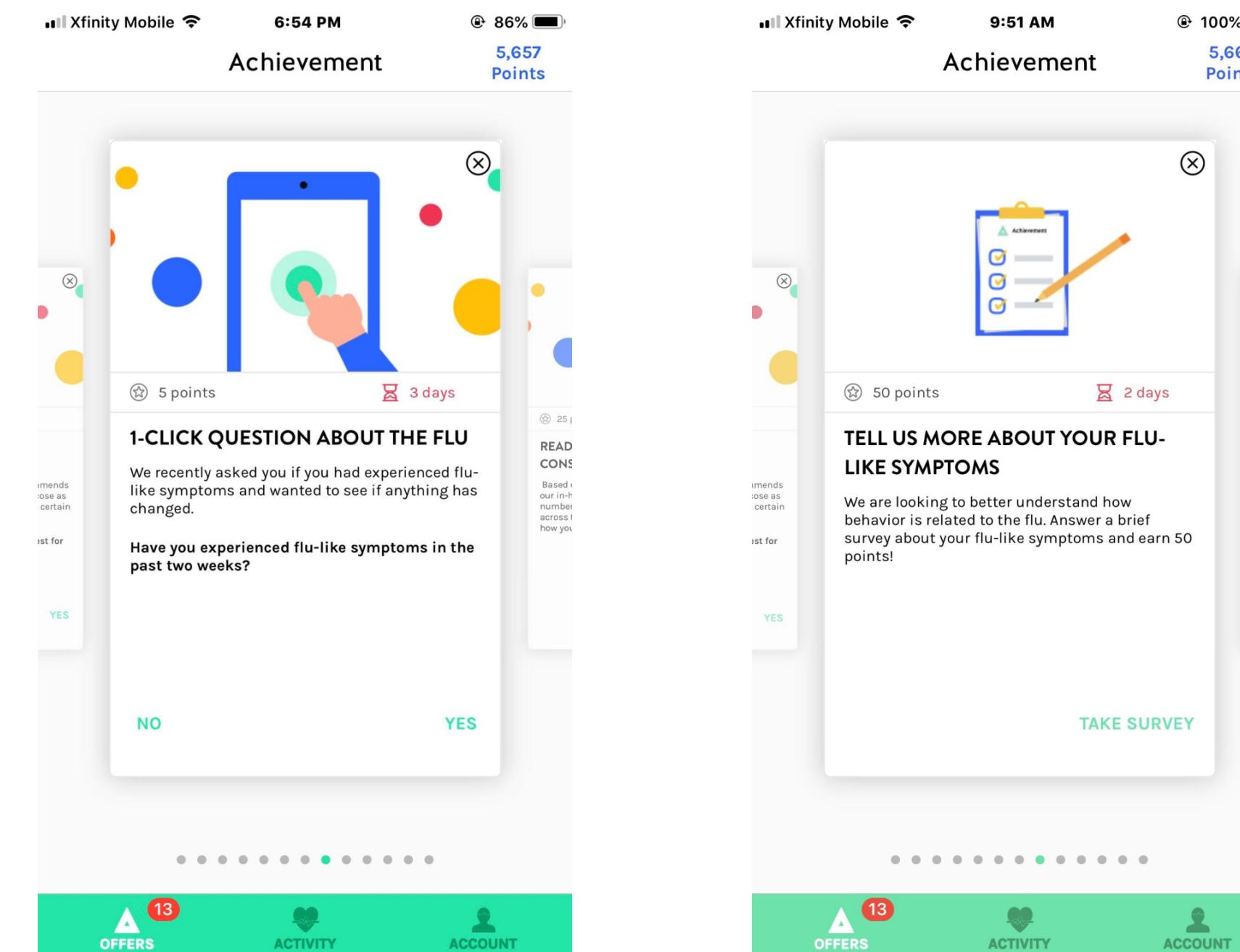
- We collect multiple streams of raw “intraday” data
- These raw time series are then aggregated into “features” using various statistical transformations in order to capture various aspects of behavior and physiology
- We combine activity features with other collected data, such as self-reported demographics and survey data



The Achievement Flu Study was based on a 12-week period in 2018 during which our population was surveyed

- 12 weeks of the 2017/18 flu season
 - Initiated Feb 2018
- Users queried as to flu symptoms in past two weeks
 - “yes” = full flu survey, “no” = repeat two weeks later

Flu Symptom 1-Click



Mobile / Web Flu Survey

Mobile / Web Flu Survey

* 5. What other symptoms did you have? Please select all that apply.

Cough
 Sore throat
 Shortness of breath
 Chills
 Nausea/Vomiting
 Diarrhea
 Body aches
 Headache
 Other (please specify)

* 6. Did you see a physician or seek medical attention for this illness?

Yes
 No

The Achievement Flu Survey included 13 items to assess the impact of ILI

- Survey Items
 - Symptoms
 - Symptoms dates
 - Vaccination status
 - Health-seeking behavior
 - Physician Diagnosis / Method
 - Hospitalization
 - Prescription
 - Missed time at work
- Responses to Achievement platform
 - Demographics
 - General health
 - Wearable data

Moving over to the notebooks!

For the rest of the class we'll be running through a notebook example:

https://github.com/evidation-health/EvidationDataScienceModule/tree/master/2019_flu_analysis

Or just google: “EvidationDataScienceModule”

Practical Tips for aspiring data scientists

Learn the lingo, and deeply understand the implications and limitations of your methods.

Talk like a Data Scientist



data scientist

"If you would be a real ~~seeker after truth~~, it is necessary that at least once in your life you doubt, as far as possible, all things"

~René Descartes

- "AUC" (AUROC)
- "Data is bad" (There's no bad data, only bad data science)
- "Signal-bound" (Fitbit can't predict your salary)
- "Predict" (Only sometimes it's the future)
- "P-hacking" (If you torture the data long enough, it will confess)
- "Significant" (Doesn't necessarily mean significant)

Learn to write good code, it's an upfront investment that pays for itself many times over

When you don't create things, you become defined by your tastes rather than ability. Your tastes only narrow and exclude people. So create.

- Jonathan Gillette

Everyone you work with can teach you something, look for opportunities to learn from people, and ask for good teachers and mentors!

CEO @ FRBSF



Teaching Coach
@ Teach for America



Software Engineer
@ Guild Engineer

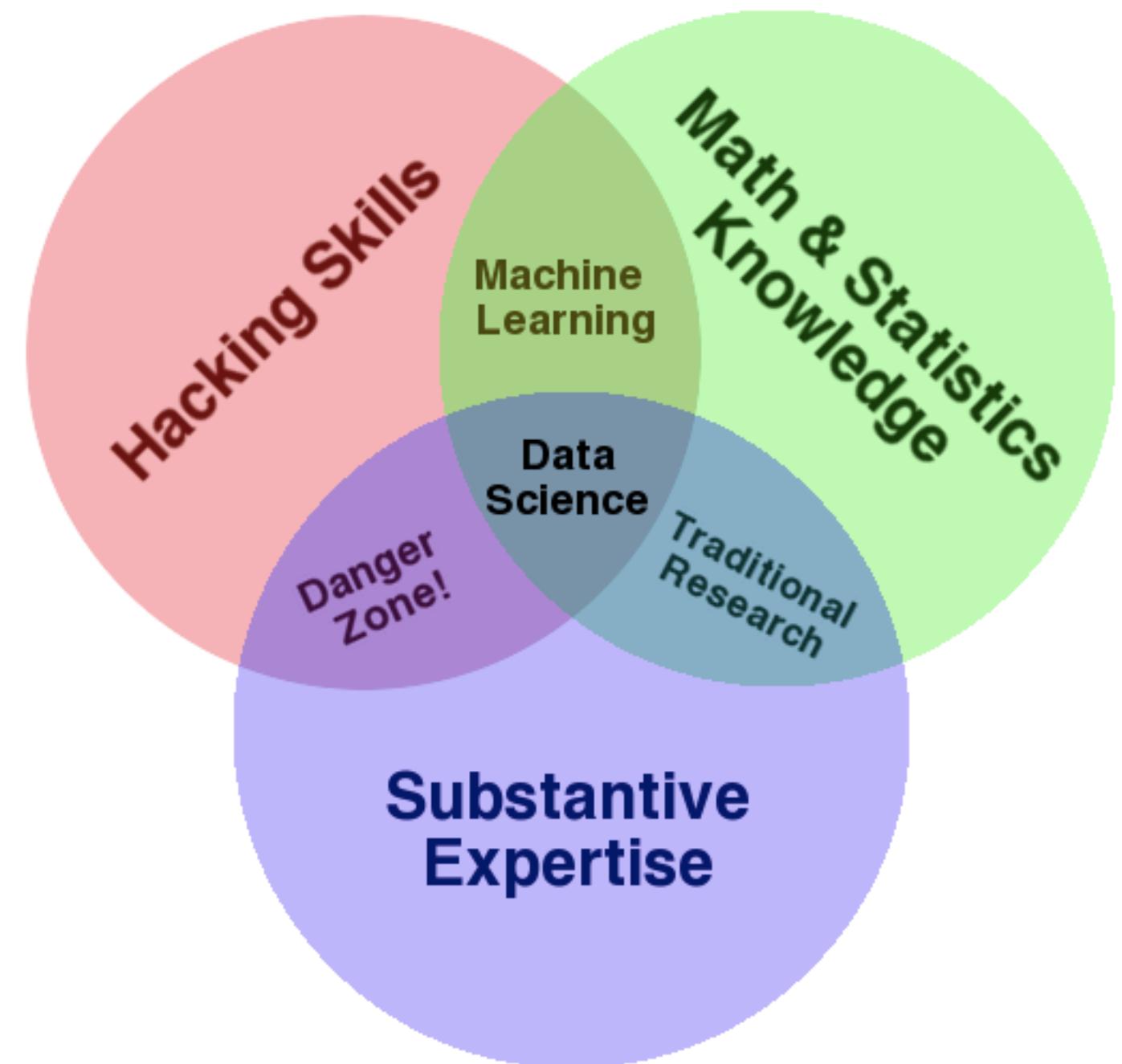


Data Scientist
@ Evidation Health



Focus on developing strong fundamentals as a generalist before become a domain expert. They are robust to changes in fashion.

Traditional Data Science Skills Venn Diagram:



How we think about talent at Evidation:

Need to have (necessary) conditions:

- Writes good code
- Communicates clearly: both verbally and in written form
- Intellectual curiosity
- Work ethic

Nice to have (but NOT sufficient):

- Statistical expert
- Machine learning expert
- Deep learning expert
- PhD
- Medical expertise

Take a problem centered approach to solving problems, not a tools centered approach.

Don't let urgent work take the place of important work.

Let me summarize. You've got to work on important problems. I deny that it is all luck, but I admit there is a fair element of luck. I subscribe to Pasteur's "Luck favors the prepared mind." I favor heavily what I did. Friday afternoons for years - great thoughts only - means that I committed 10% of my time trying to understand the bigger problems in the field, i.e. what was and what was not important. I found in the early - days I had believed `this' and yet had spent all week marching in `that' direction. It was kind of foolish. If I really believe the action is over there, why do I march in this direction? I either had to change my goal or change what I did. So I changed something I did and I marched in the direction I thought was important. It's that easy.

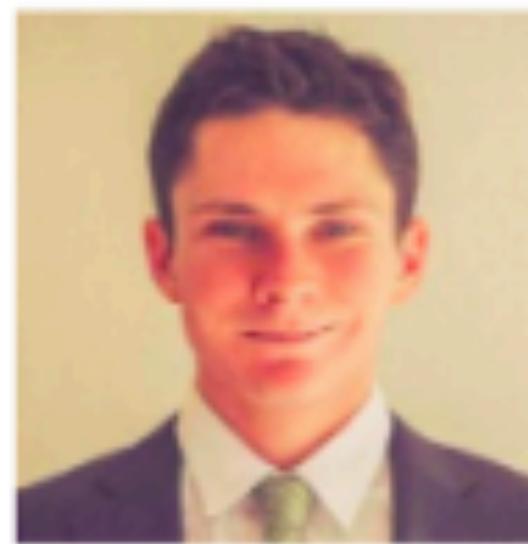
- Richard Hamming

Many roads to data science

Data Science Team, May 2019



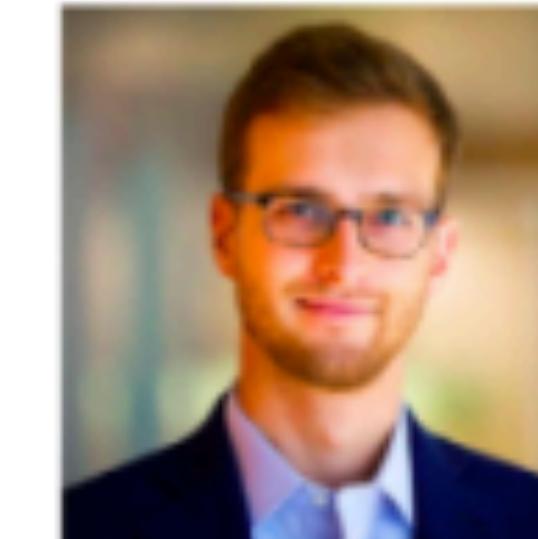
Wei-Nchih (Weech)
MD, PhD, MPH, Stanford
Kaiser Permanente,
HP labs



Ben Bradshaw
BS, Math+Econ, UCSD
Federal Reserve Bank



Mackenzie Wildman
PhD, Fin. Math, Lehigh
UCSB A. Professor



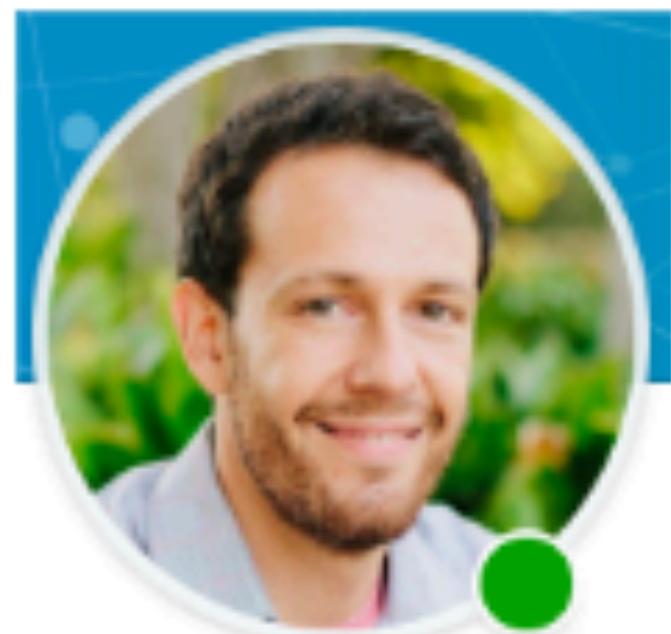
Filip Jankovic
ME, UofM
Coursera



Nikki Marinsek
PhD Candidate
Psych UCSB



Ernesto Ramirez
PhD UCSD



Luca Foschini
PhD, UCSB
Evidation Co-founder,
Chief Data Scientist
Google, ask.com



Joanne Shin
MS Electrical Engineering
Formerly: Kaiser



Mitchell Maier
MS Materials Engineering
Formerly



Raul Eulogio
BS Probability and
Statistics, UCSB