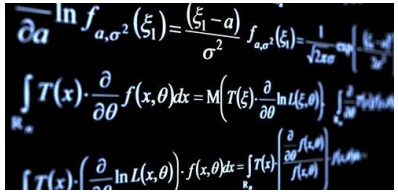
Big Data Approaches to Psychological Processes

Simon Dennis

Complex Human Data Hub
University of Melbourne
Unforgettable Research Services Pty Ltd




764.4
2.96
25:38


WHERE?
WHAT?
WHO?
HOW?
WHY?
WHEN?

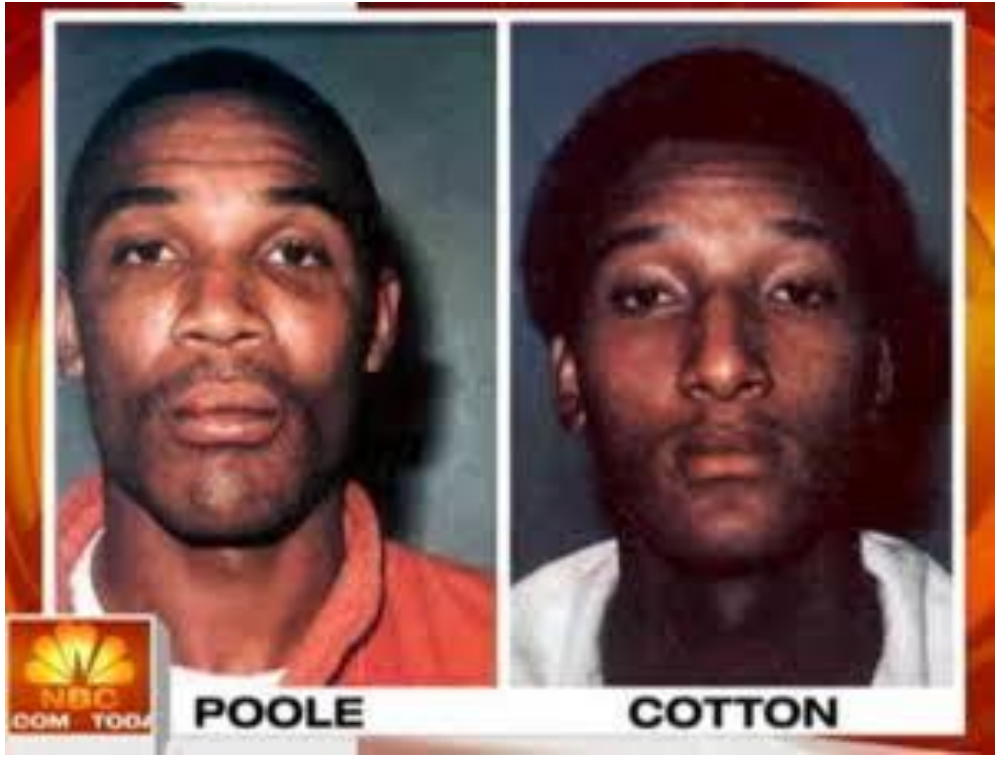
In 1984, Ronald Cotton was convicted for two counts of rape and two counts of burglary.

He was sentenced to life + 50 years.

Cotton was exonerated in 1995, after spending over 10 years in prison when DNA evidence demonstrated it had been Bobby Poole who had committed the crimes.

His convictions were based on an eyewitness misidentification made by one of the victims, Jennifer Thompson-Cannino.

Misidentification occurred in a photo lineup and then in a physical line up

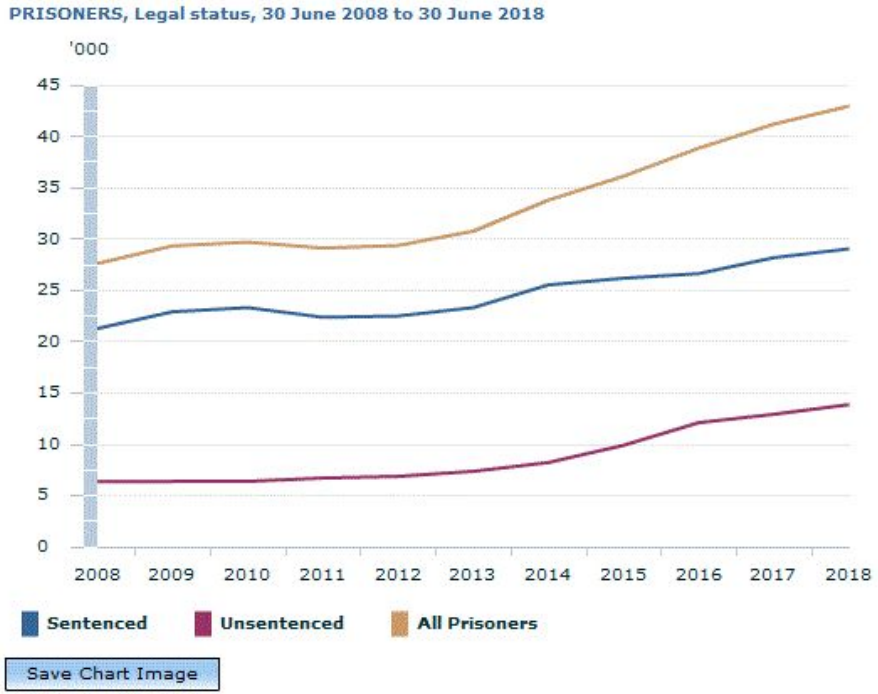


In a retrial in 1987, when presented with the real rapist (Bobby Poole) and Ronald Cotton, Thompson-Cannino chose Cotton again.

Captain Mike Gauldon, the detective in the case

Cotton also misremembered where he had been at the time of the crimes

His false alibi led jurors to the conclusion that he was lying



The Innocence Project estimates that between 2% and 5% of prisoners have been falsely convicted.

In the United States, 2,298,300 people were incarcerated in 2016 (Prison Policy Initiative) suggesting over 100,000 people might be falsely imprisoned.

Australia had a prison population of 42,974 at 30 June 2018 (ABS), suggesting that over 2000 individuals may be serving time for crimes they did not commit.

How reliably can people remember where they have been?

What factors affect this memory?

What questions might the detective have asked to uncovered the false memory?

Can’t use laboratory paradigms to answer these questions.

For 2 months collect

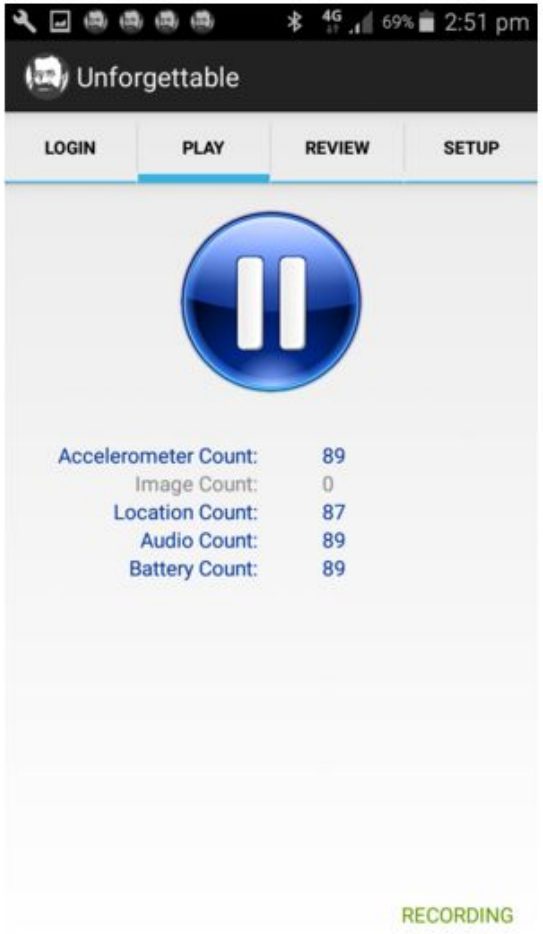
GPS coords every 10 minutes, about 8640 pairs per subject

Accelerometry vectors 10 times a second continuously, about 51M data points per subject (then FFT)

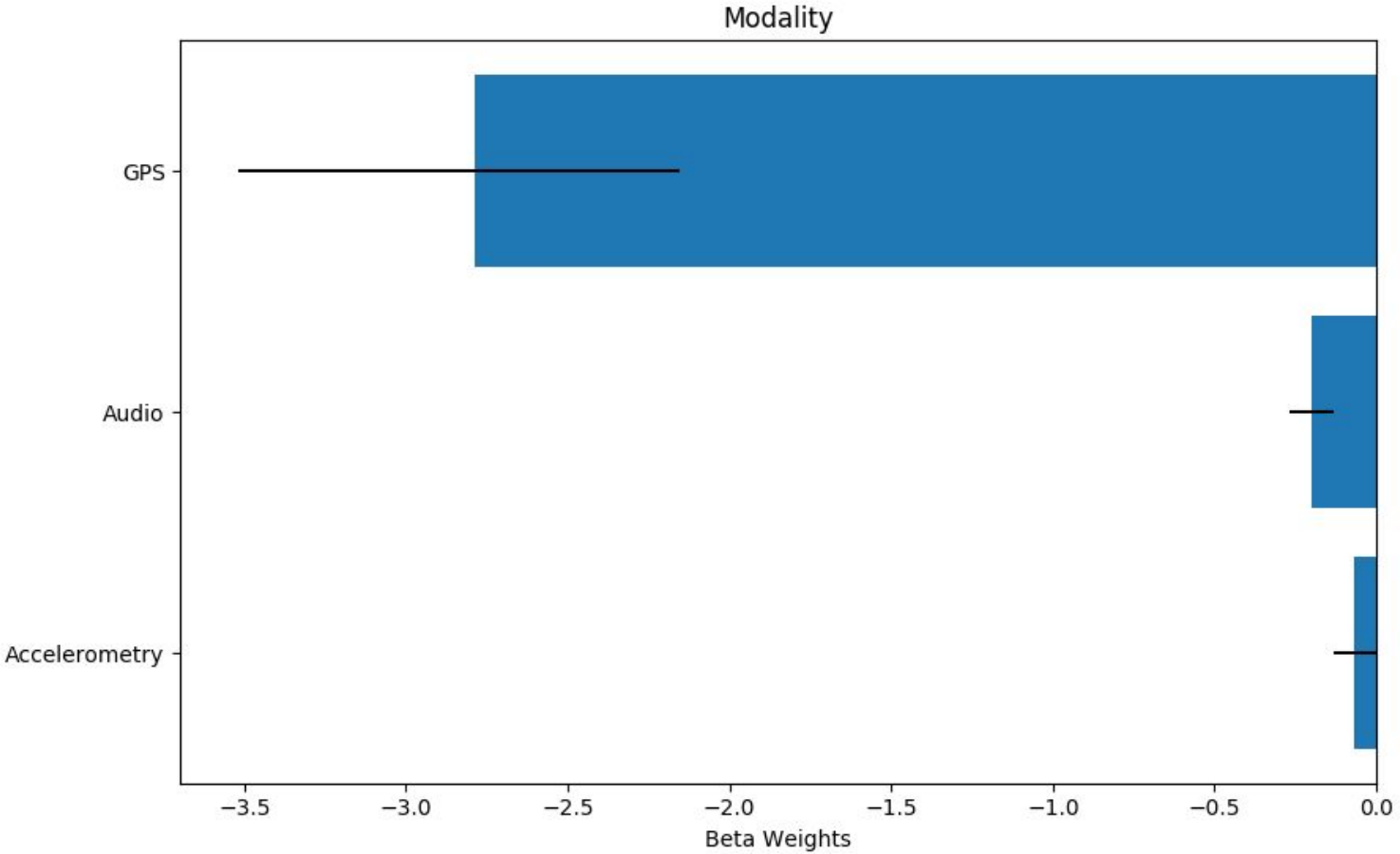
Audio as 500 milliseconds of MFCCs every 10 minutes

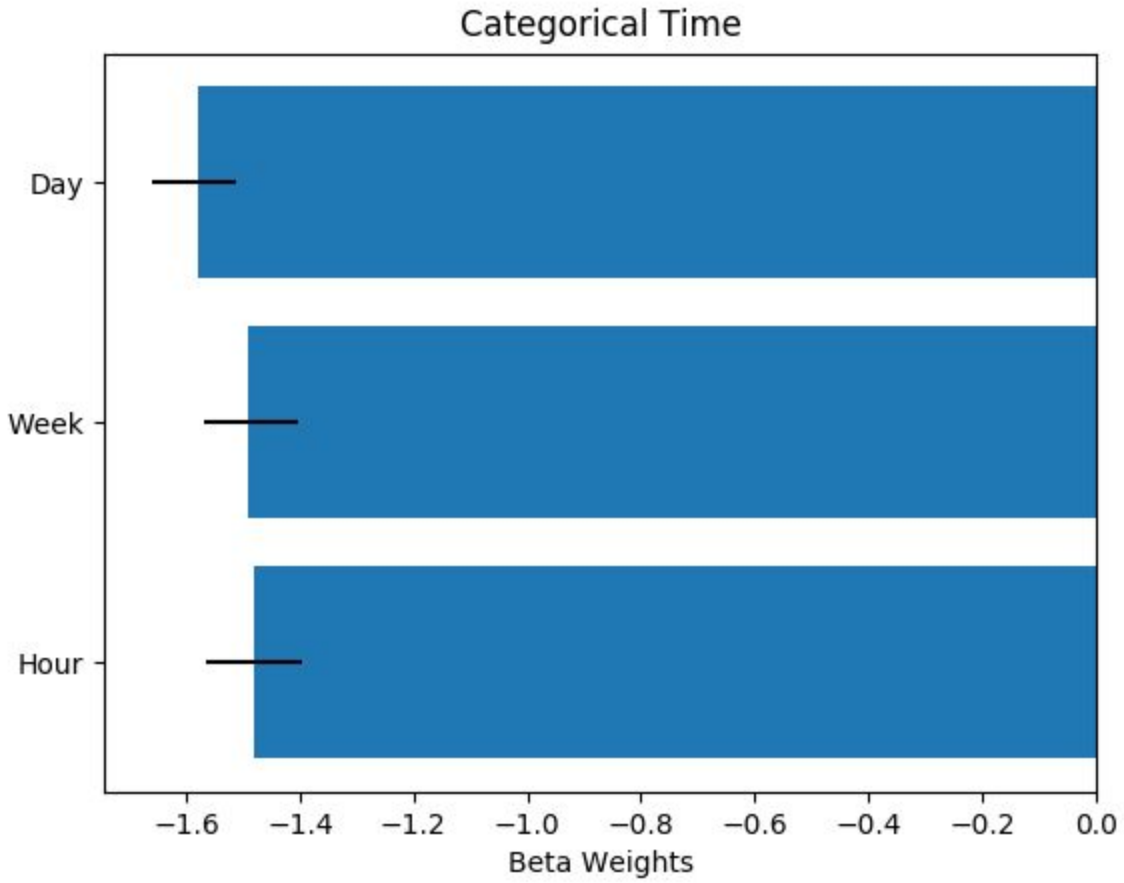
One week retention interval

Then receive a memory test



Click marker (📍) where you were on Tuesday 2018-02-13 at 20:00


Modality
GPS
Audio
Accelerometry
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
Beta Weights


Categorical Time
Day
Week
Hour
−1.6
−1.4
−1.2
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
Beta Weights

Investigators should ask:

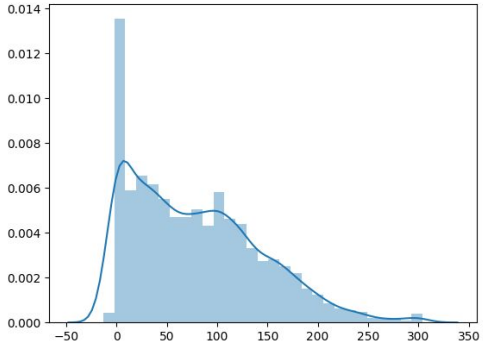
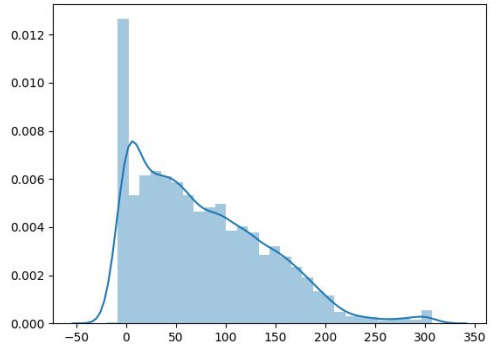
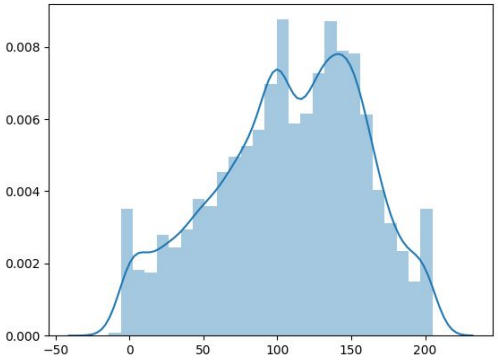
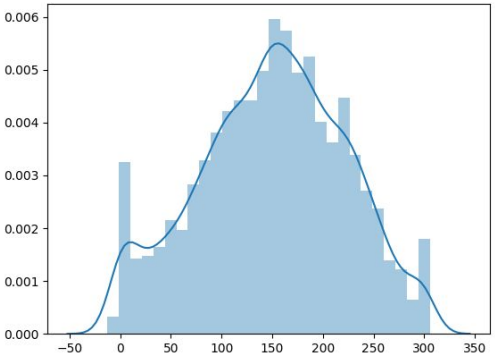
At what other times were you at this place?

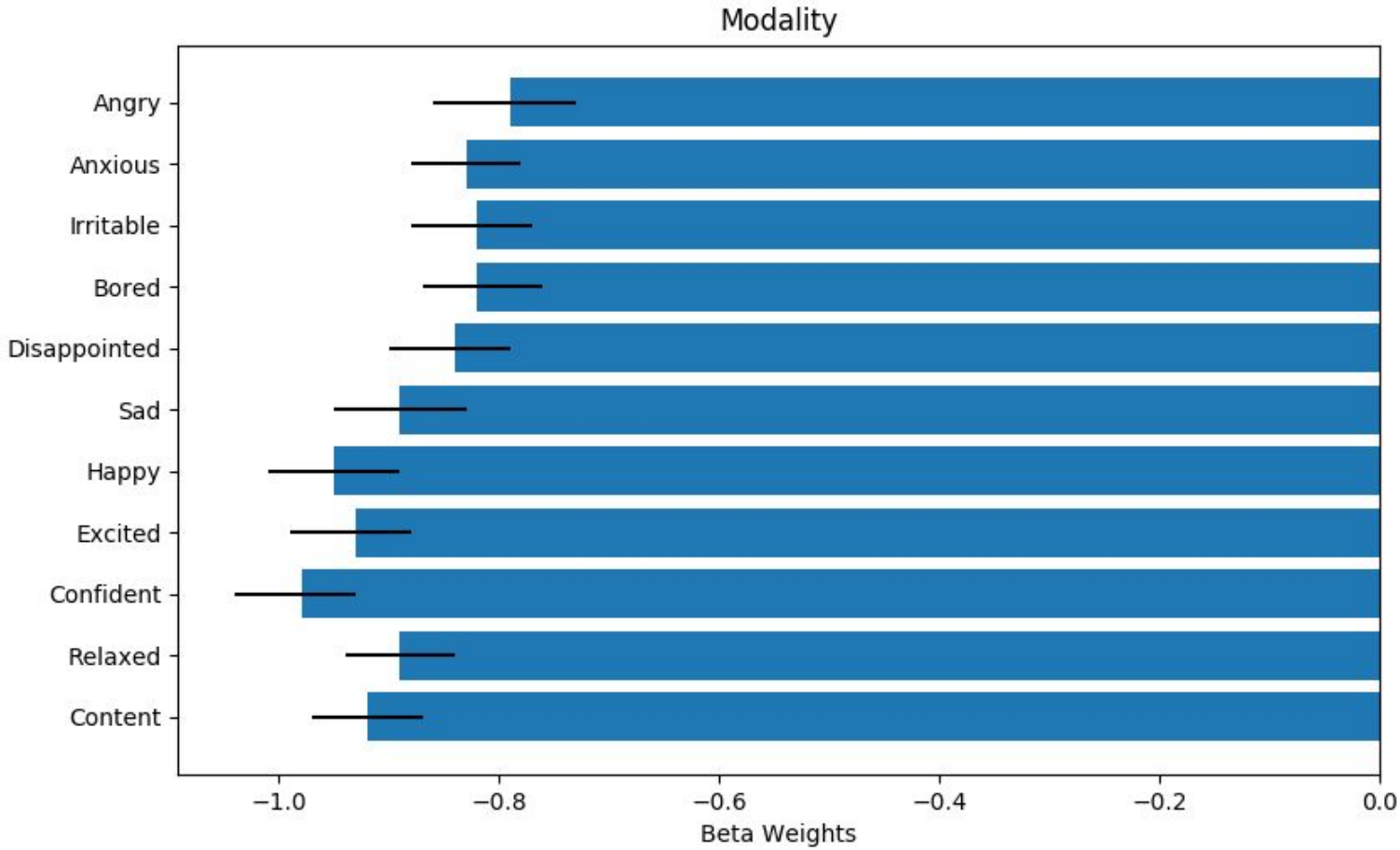
Where were you the week before and after on this day?

Where were you the day before and after at this time?

But still need to determine if we have the right representations for the other modalities

The impact of emotion on memory for WHERE

	Low Arousal	High Arousal
Negative	.007 	0.017 
Positive	0.14* 	-0.10* 


Modality
Angry
Anxious
Irritable
Bored
Disappointed
Sad
Happy
Excited
Confident
Relaxed
Content
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
Beta Weights

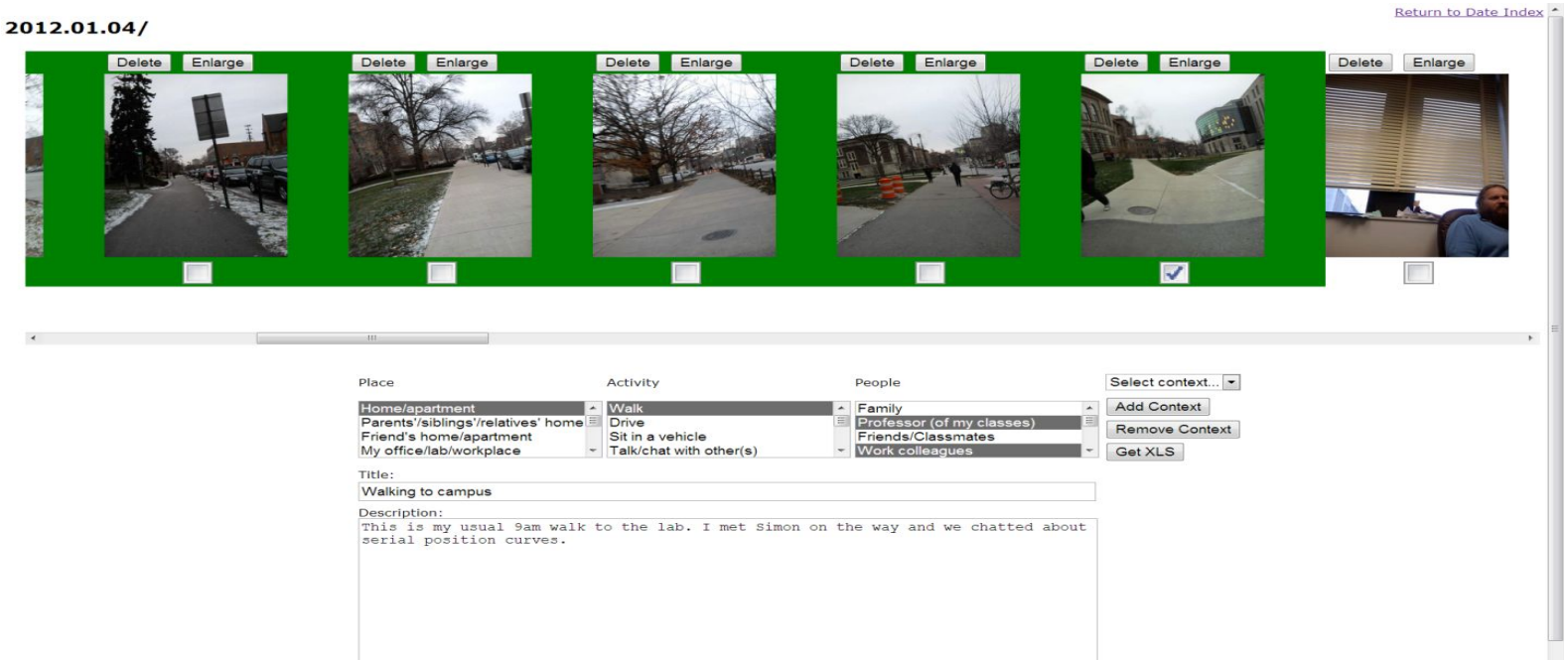

WHERE?
WHAT?
WHO?
HOW?
WHY?
WHEN?

*

The Data

Each night participants:

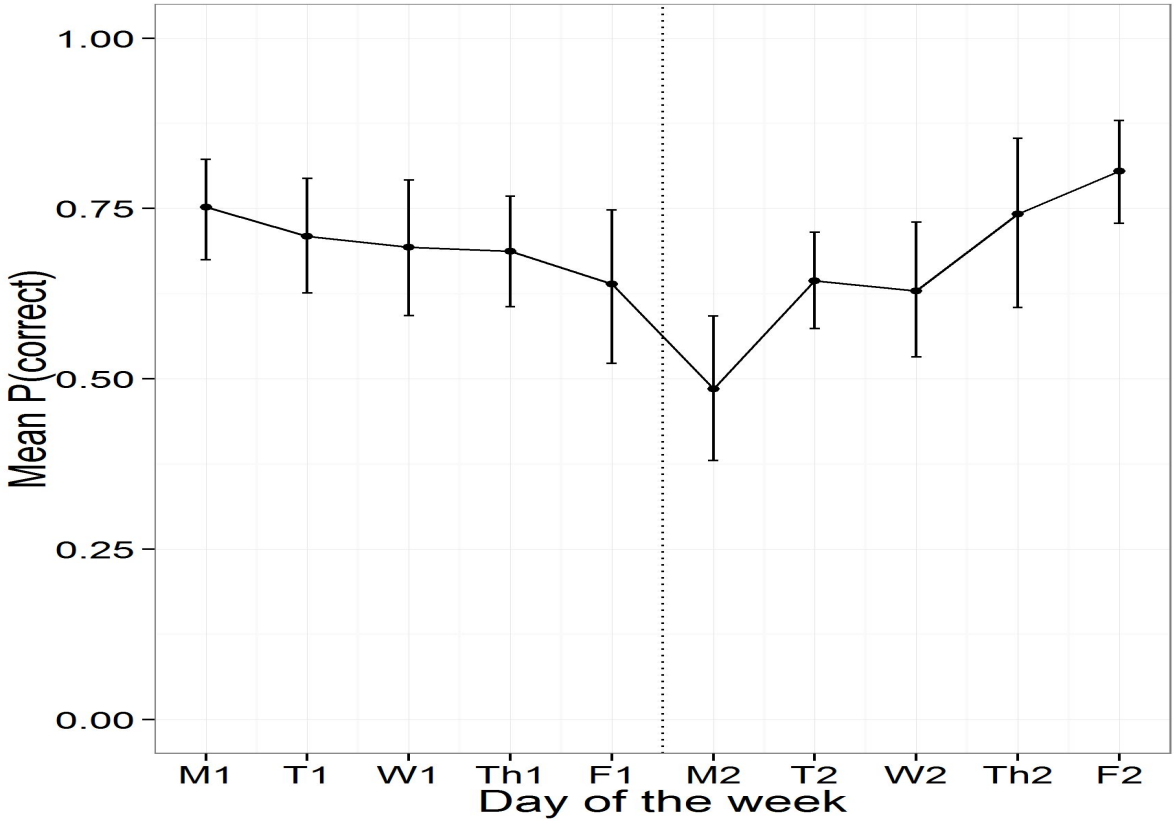
Segment images into contexts (episodes)
Provide a description
Provide activity, place and people tags



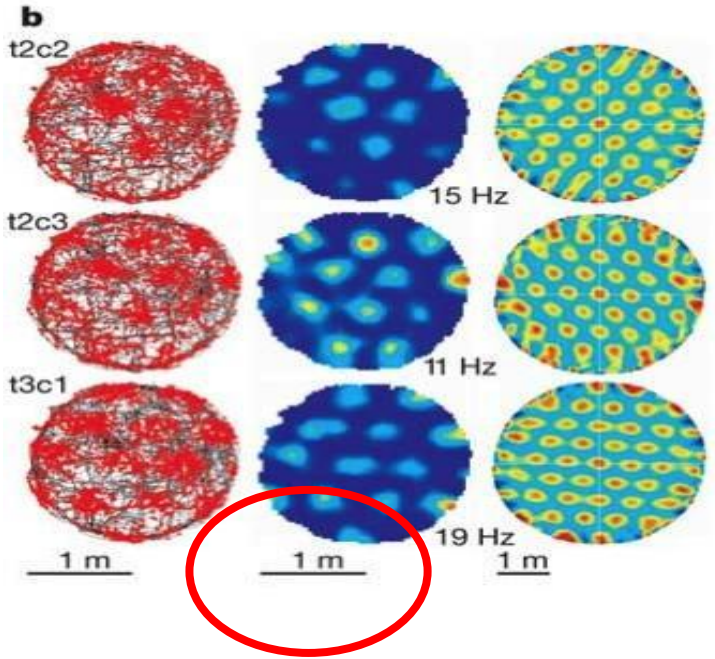
Did this happen last week or the week before?

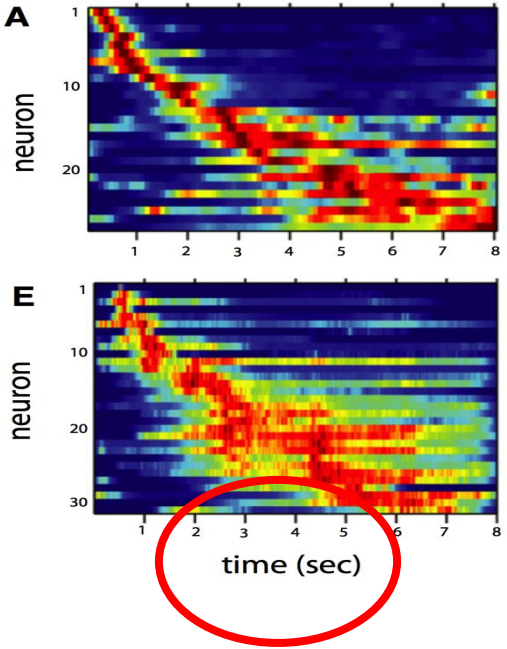
P(correct)
0.0
0.2
0.4
0.6
0.8
1.0
45
9
59
57
21
48
194
107
97
69
75
194
4
62
407
118
54
24
44
448
'Party'
'Airport/Bus~station'
'Kids'
'Care for/play with child/baby'
'Stranger'
'Park/Museum/Zoo'
'Spouse/Partner'
'Shopping'
'Salesperson/Clerk/Cashier'
'Library'
'Boyfriend/Girlfriend'
'Thinking'
'Other office'
'Hobbies'
'Work'
'Restaurant/Cafe/Bar'
'Other person's office/workplace'
'Pet'
'Giving a lecture/presentation'
'My office/lab/workplace'

Week Discrimination



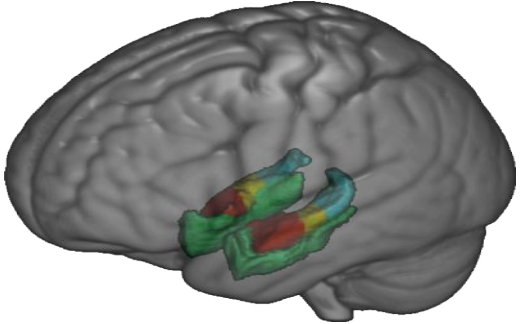

WHERE?
WHAT?
WHO?
HOW?
WHY?
WHEN?

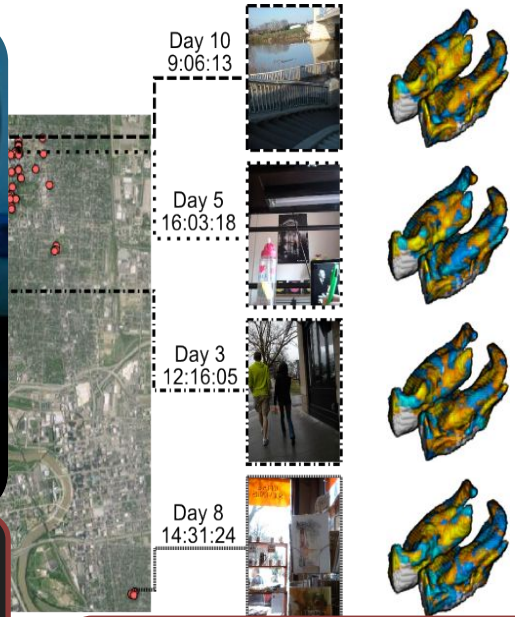

b
t2c2
15 Hz
t2c3
11 Hz
t3c1
19 Hz
1 m
1 m
1 m


A
neuron
1
10
20
1
2
3
4
5
6
7
8
E
neuron
1
10
20
30
1
2
3
4
5
7
8
time (sec)

Experience Sampling
(2-4 weeks)

Reminiscence Task
in fMRI





Compare Patterns of Brain Activity to Structure of the World

8 s

Were you able to remember the event?

Yes No

3s

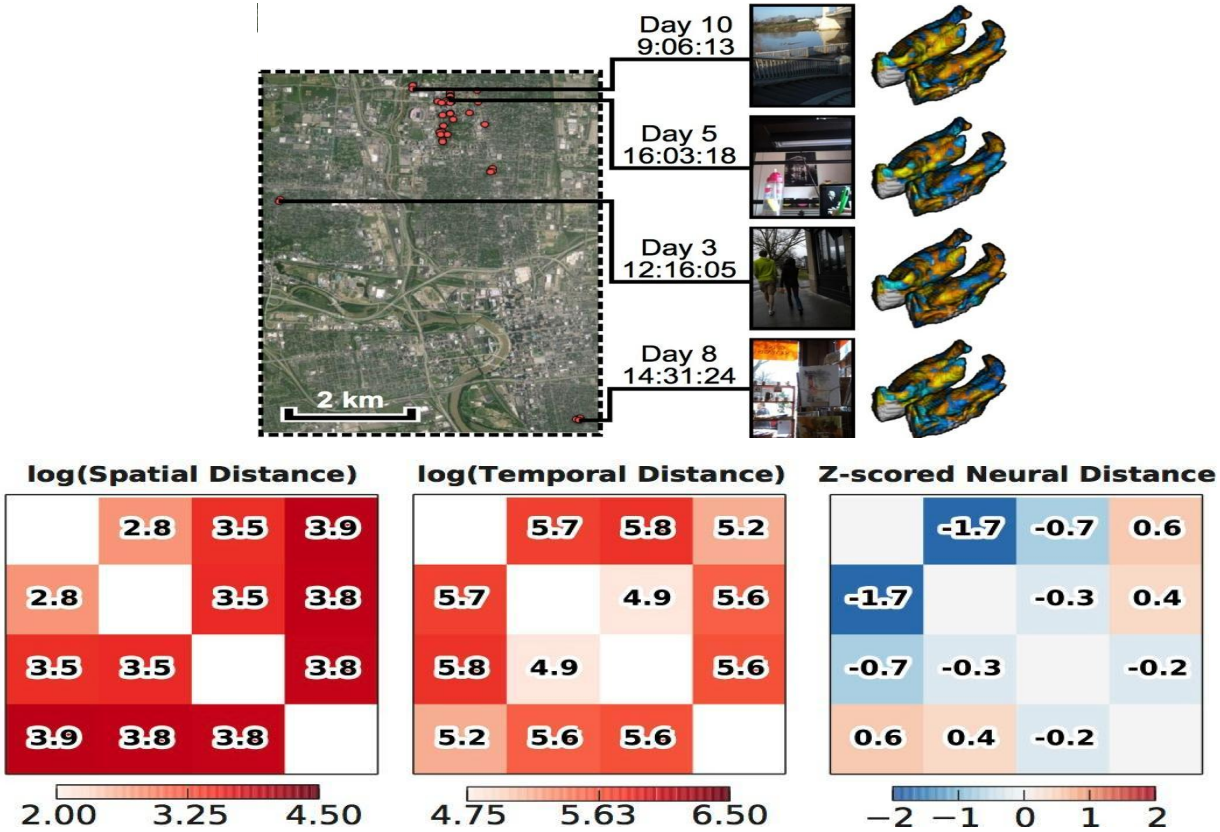
How vivid was your memory?

Lots of detail Very little detail

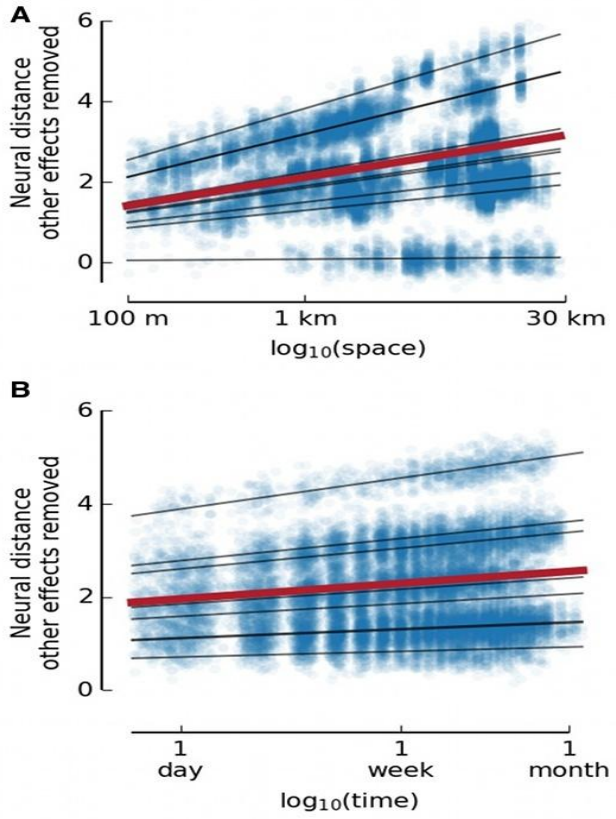
3s

4-10 s ITI

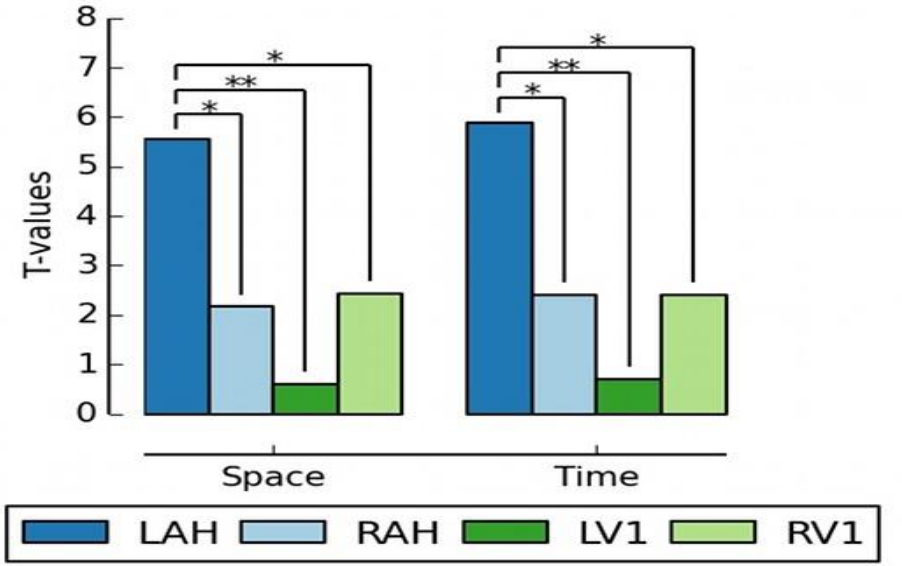
Representational Similarity Analysis (RSA)

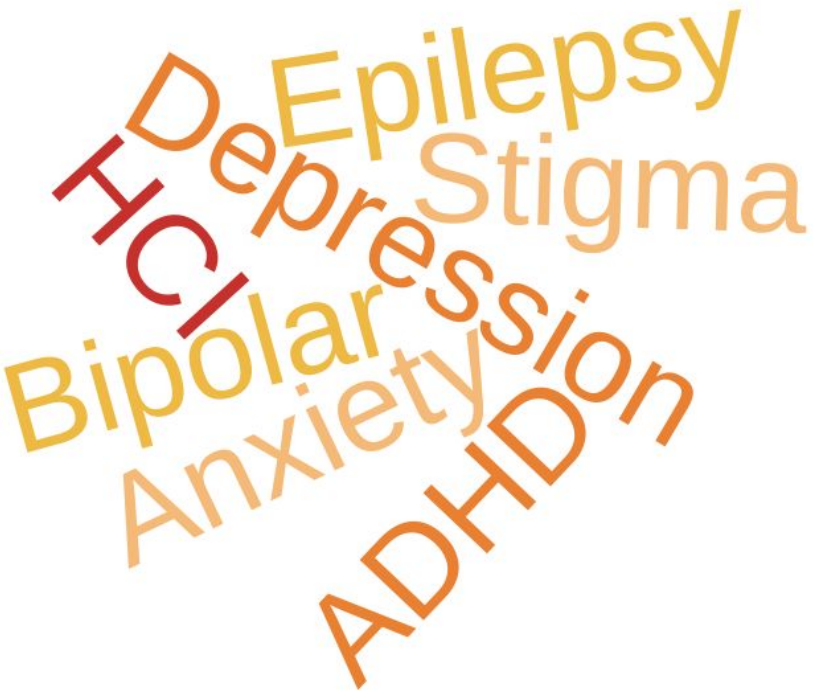


Significant in Left Ant. Hippocampus



More than in other control regions

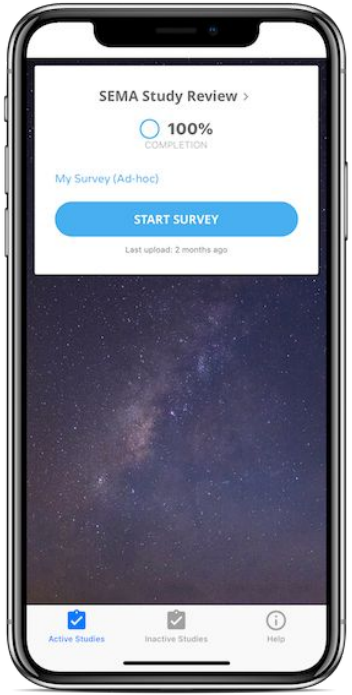


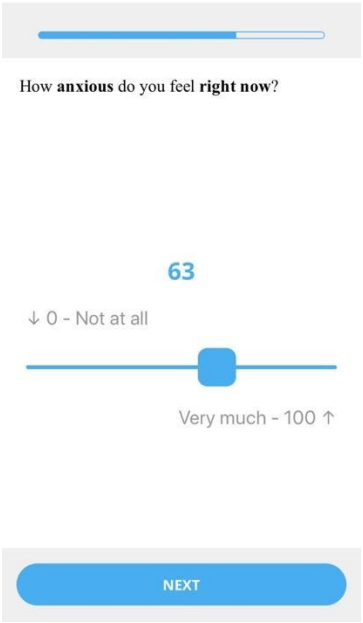

Epilepsy
Depression
Stigma
HCI
Bipolar
Anxiety
ADHD

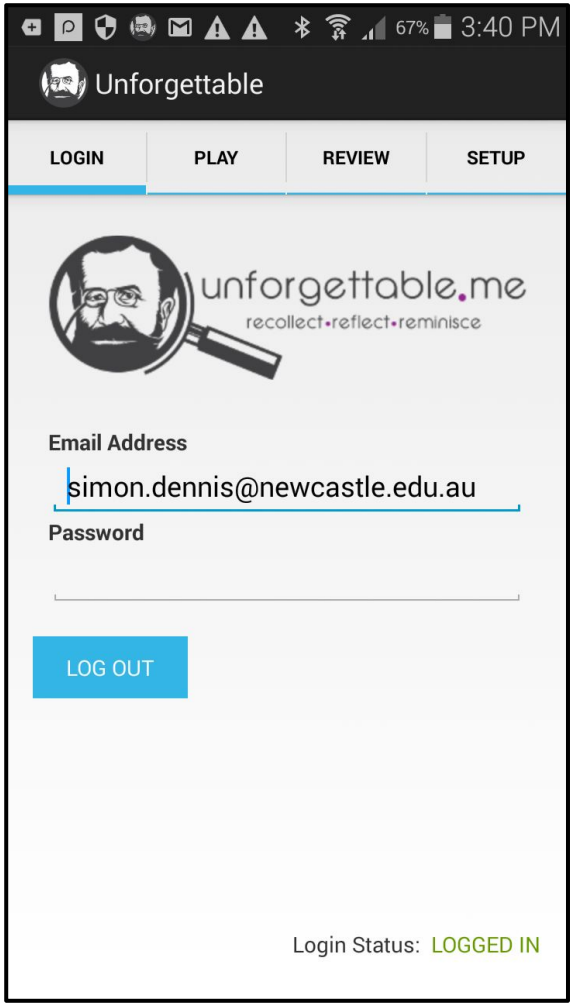

LET'S GET
STARTED
SEARCH
20%
15%
35%
50%

IFTTT provides access to over 600 data streams

SEMA3


SEMA Study Review ›
100%
COMPLETION
My Survey (Ad-hoc)
START SURVEY
Last upload: 2 months ago
Active Studies
Inactive Studies
Help


How **anxious** do you feel **right now**?
63
↓ 0 - Not at all
Very much - 100 ↑
NEXT

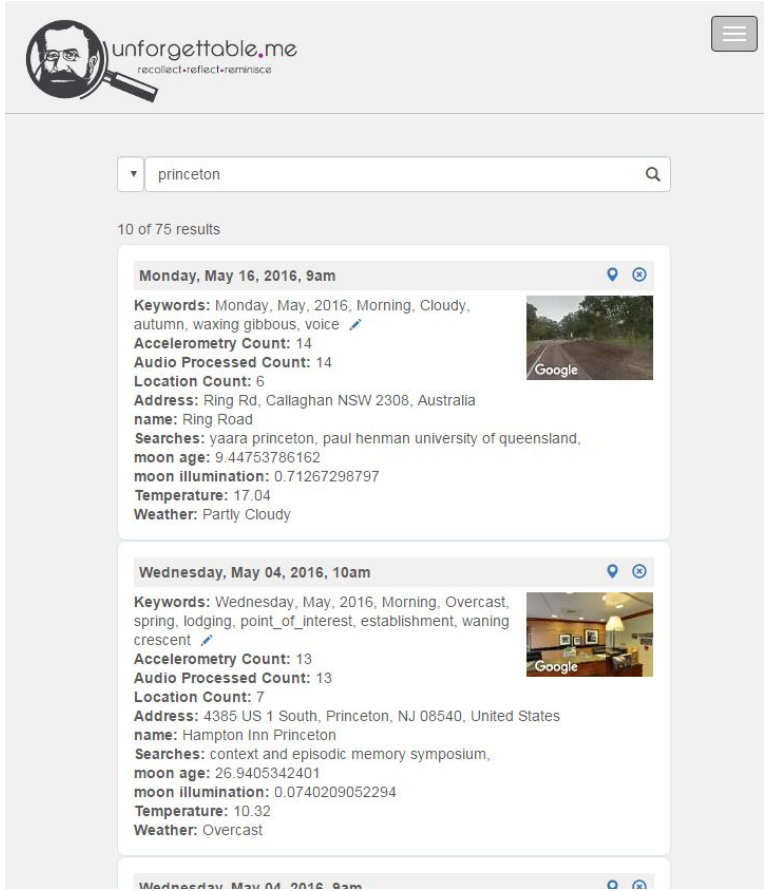


Unforgettable App can collect GPS, accelerometry, obfuscated audio and images

Allows control over which streams are collected

Data is stored on phone to start with and can be deleted before it reaches our servers

The rate of data collection can be adjusted to optimize battery life

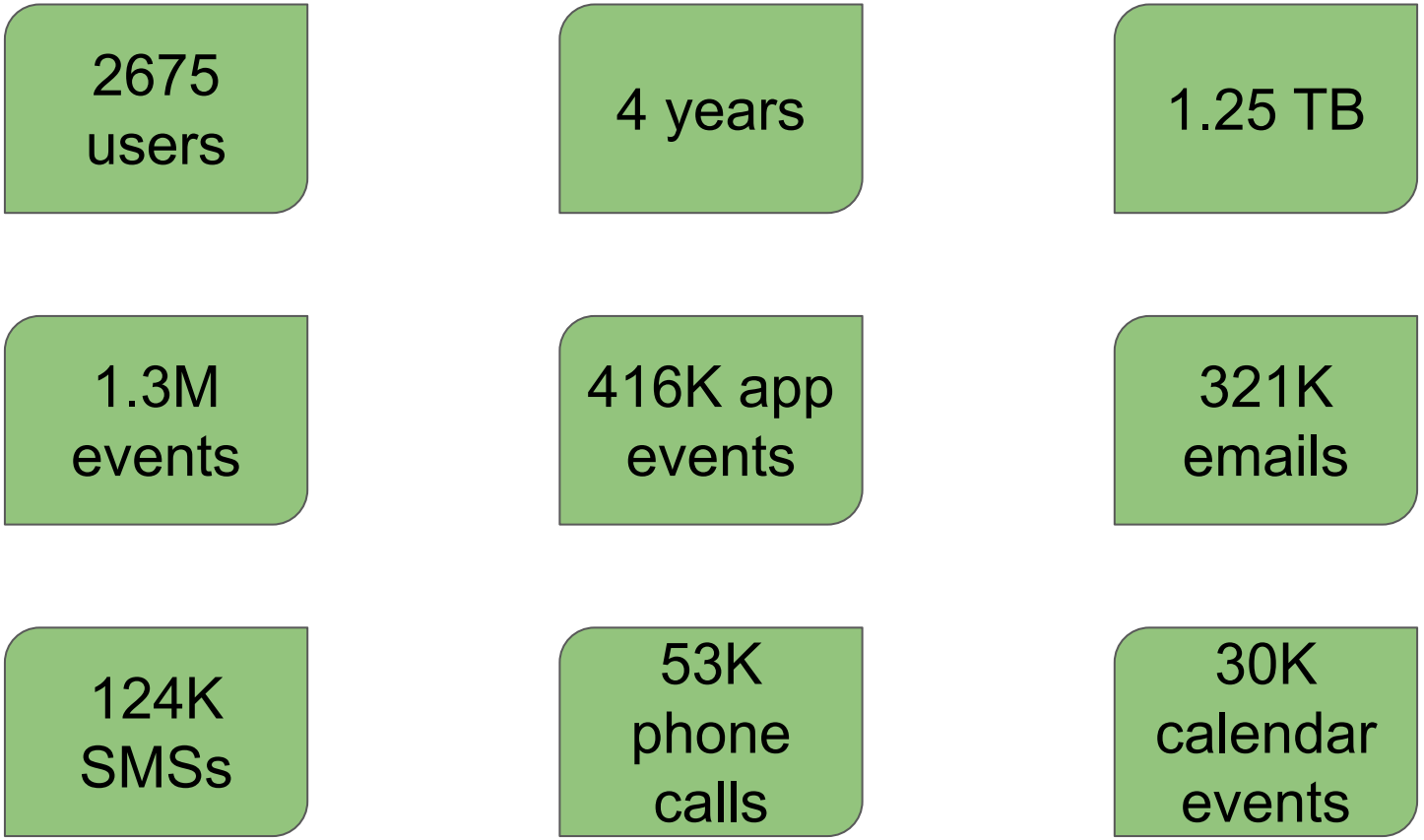


Participants receive:

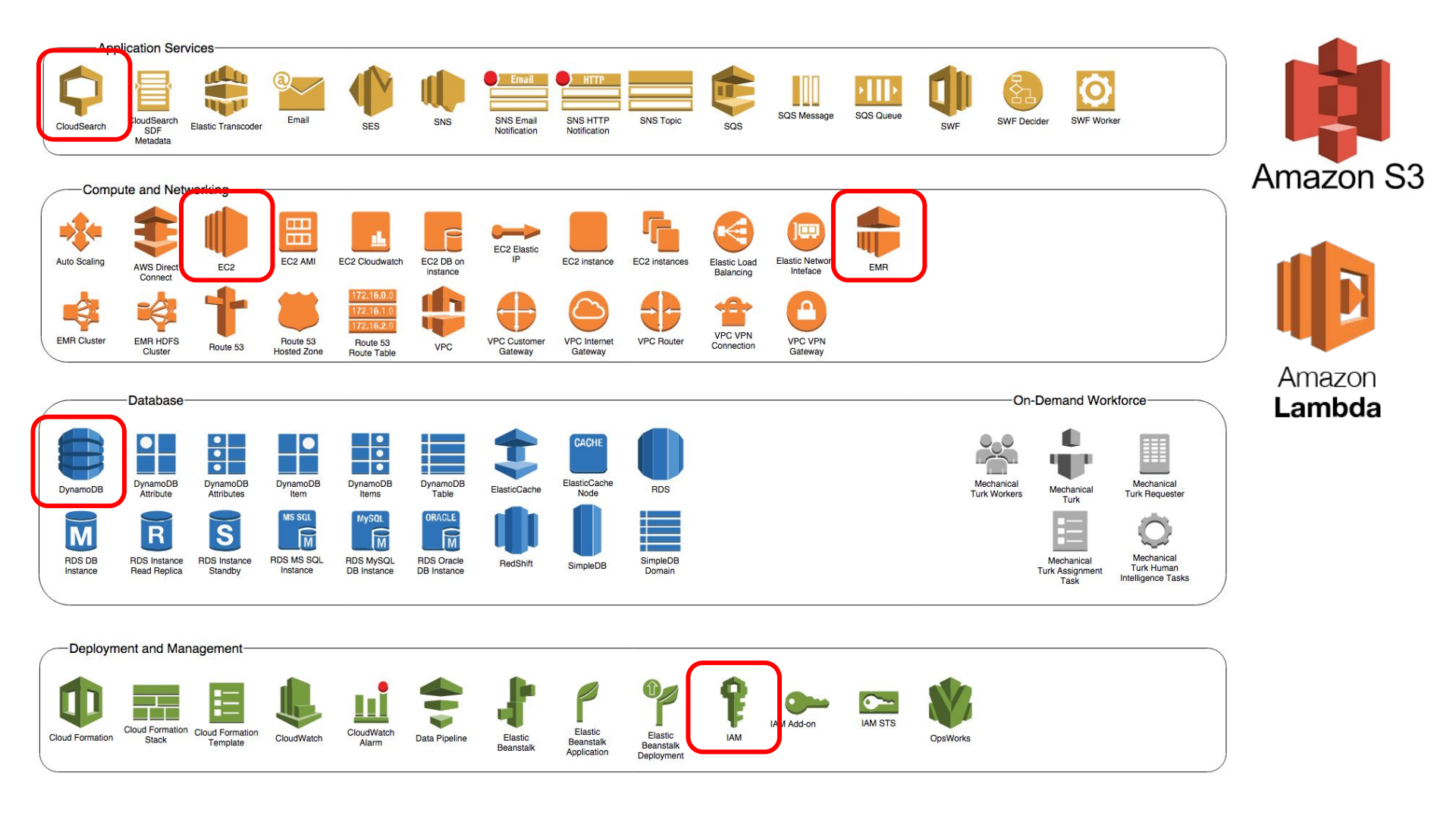
Search tools that allow them to recollect events from their lives

Visualization tools that allow them to reflect on patterns in their lives

Compensation for leasing their data to researchers



Data Marketplace (November 2019)


Application Services
CloudSearch
CloudSearch SDF Metadata
Elastic Transcoder
Email
SES
SNS
SNS Email Notification
SNS HTTP Notification
SNS Topic
SQS
SQS Message
SQS Queue
SWF
SWF Decider
SWF Worker
Compute and Networking
Auto Scaling
AWS Direct Connect
EC2
EC2 AMI
EC2 Cloudwatch
EC2 DB on instance
EC2 Elastic IP
EC2 instance
EC2 instances
Elastic Load Balancing
Elastic Networ Inteface
EMR
EMR Cluster
EMR HDFS Cluster
Route 53
Route 53 Hosted Zone
Route 53 Route Table
VPC
VPC Customer Gateway
VPC Internet Gateway
VPC Router
VPC VPN Connection
VPC VPN Gateway
Database
DynamoDB
DynamoDB Attribute
DynamoDB Attributes
DynamoDB Item
DynamoDB Items
DynamoDB Table
ElasticCache
ElasticCache Node
RDS
RDS DB Instance
RDS Instance Read Replica
RDS Instance Standby
RDS MS SQL Instance
RDS MySQL DB Instance
RDS Oracle DB Instance
RedShift
SimpleDB
SimpleDB Domain
On-Demand Workforce
Mechanical Turk Workers
Mechanical Turk
Mechanical Turk Requester
Mechanical Turk Assignment Task
Mechanical Turk Human Intelligence Tasks
Deployment and Management
Cloud Formation
Cloud Formation Stack
Cloud Formation Template
CloudWatch
CloudWatch Alarm
Data Pipeline
Elastic Beanstalk
Elastic Beanstalk Application
Elastic Beanstalk Deployment
IAM
IAM Add-on
IAM STS
OpsWorks
Amazon S3
Amazon Lambda

87 million Facebook profiles taken

Used to influence the US presidential election??

Alexandr Kogan is a social psychologist who studies personality
Research associate at Cambridge University
“This is your digital life” app tested the big five
About 270,000 people used the app

In mid 2018, Public Transport Victoria (PTV) released 15 million de-identified myki records

Re-identification

Within three months researchers at the University of Melbourne had been able to identify themselves in the data, a co-traveller and a member of parliament.

https://pursuit.unimelb.edu.au/articles/two-data-points-enough-to-spot-you-in-open-transport-records

Most people in the dataset are identifiable from just a handful of touch on or touch off events.

The Private Language

Current data analysis languages such as R or python have not been designed to be privacy preserving.

We have developed a new open source language called Private that allows researchers to analyse data without having access to it. They can't see it and they can't copy it.

But they can write programs to analyse the data.

All results are checked computationally to determine the privacy risk before release.

General Data Protection Regulation

Breach Notification

Right to Access

Right to be Forgotten

Data Portability

Privacy by Design

Data Officers

Tanya Preminger

Thanks to

Grant number FA9550-09-1-0614

Ben Stone
Jihun Hamm
Mikhail Belkin
Vishnu Sreekumar
Per Sederberg
Troy Smith
Dylan Neilson
Hyungwook Yim
Paul Garrett
Nathan Evans
Michael Diamond

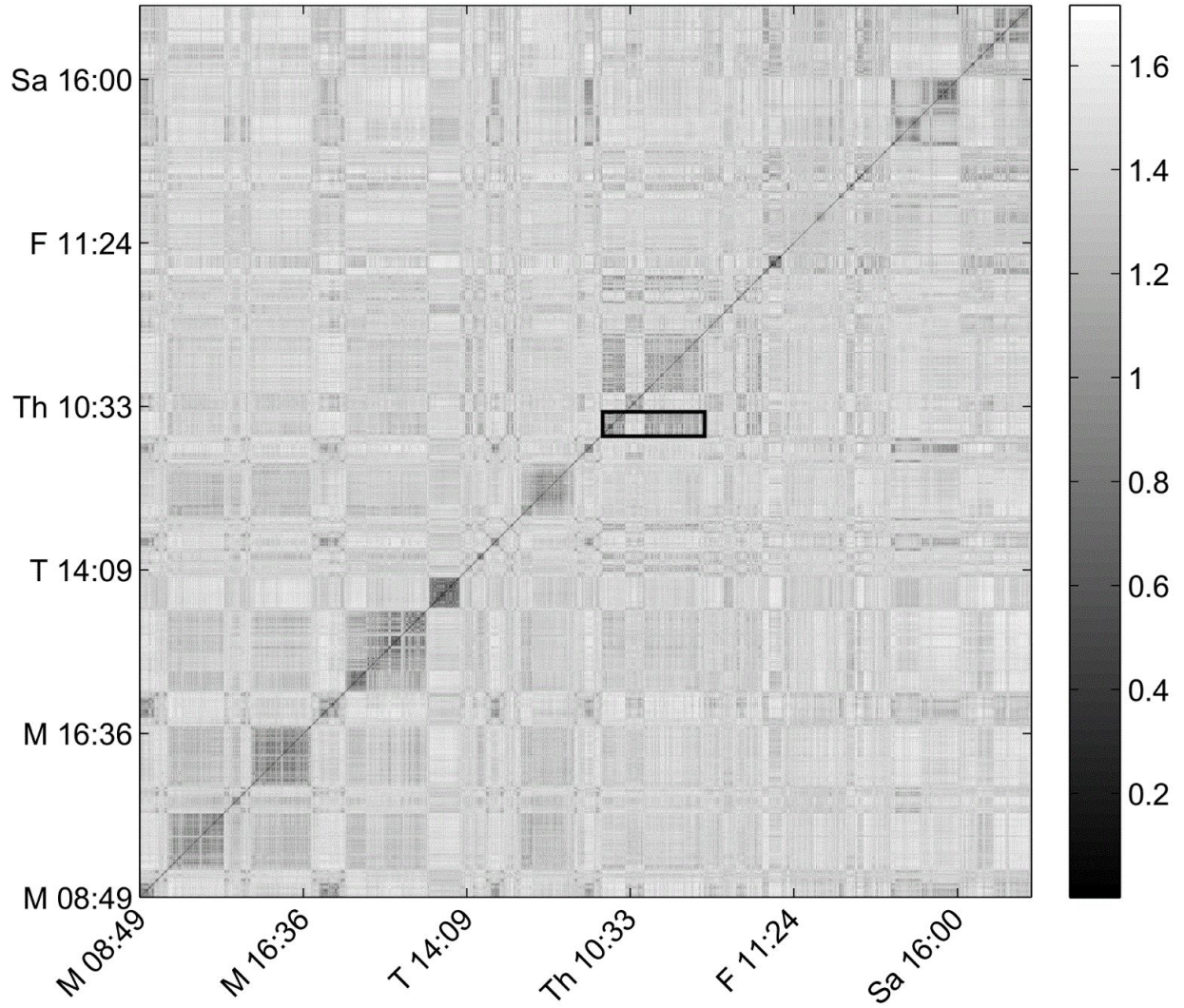
Students & Volunteers

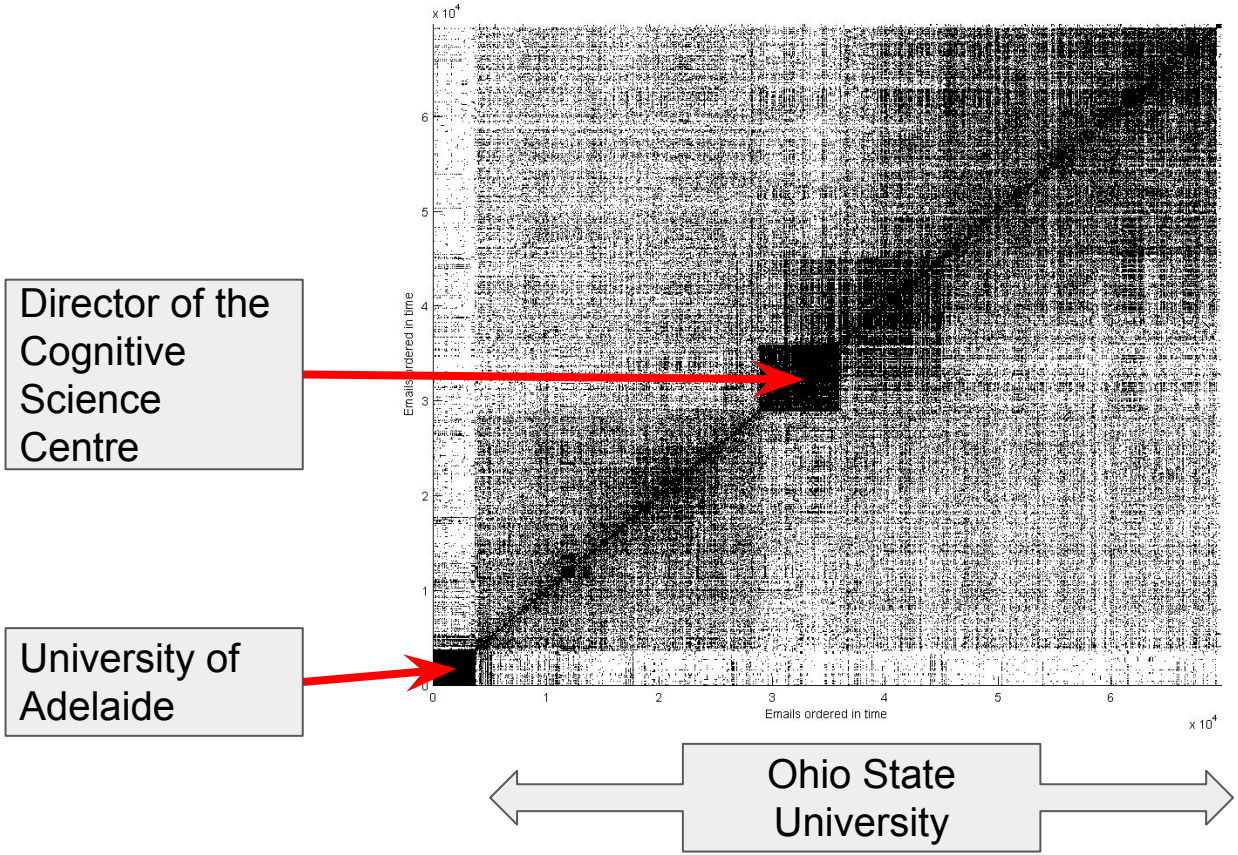
Corrine McGann
Gabrielle Matthews
Kay Chan
Samuel Collison
Mhairi Gray
Anita Gray
Megan Baker
Viviana Pinzon Morales
Rachael Vince
Alison Rasmussen
Ruby Walker
Adelaide McKenzie
Elizabeth Laliberte
Courtney O'Brien

The nature of context

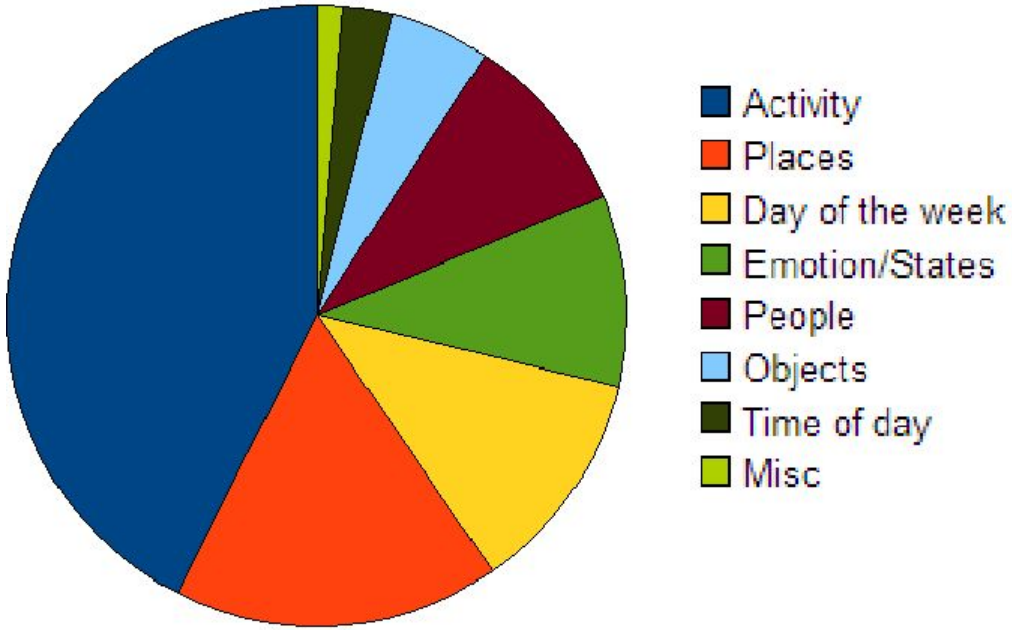
Sreekumar, V., Dennis, S., Doxas, I., Zhuang, Y., & Belkin, M. (2014). The geometry and dynamics of lifelogs: discovering the organizational principles of human experience. PloS one, 9(5), e97166.

Sreekumar, V., Dennis, S., & Doxas, I. (2017). The episodic nature of experience: a dynamical systems analysis. Cognitive Science, 41(5), 1377-1393.

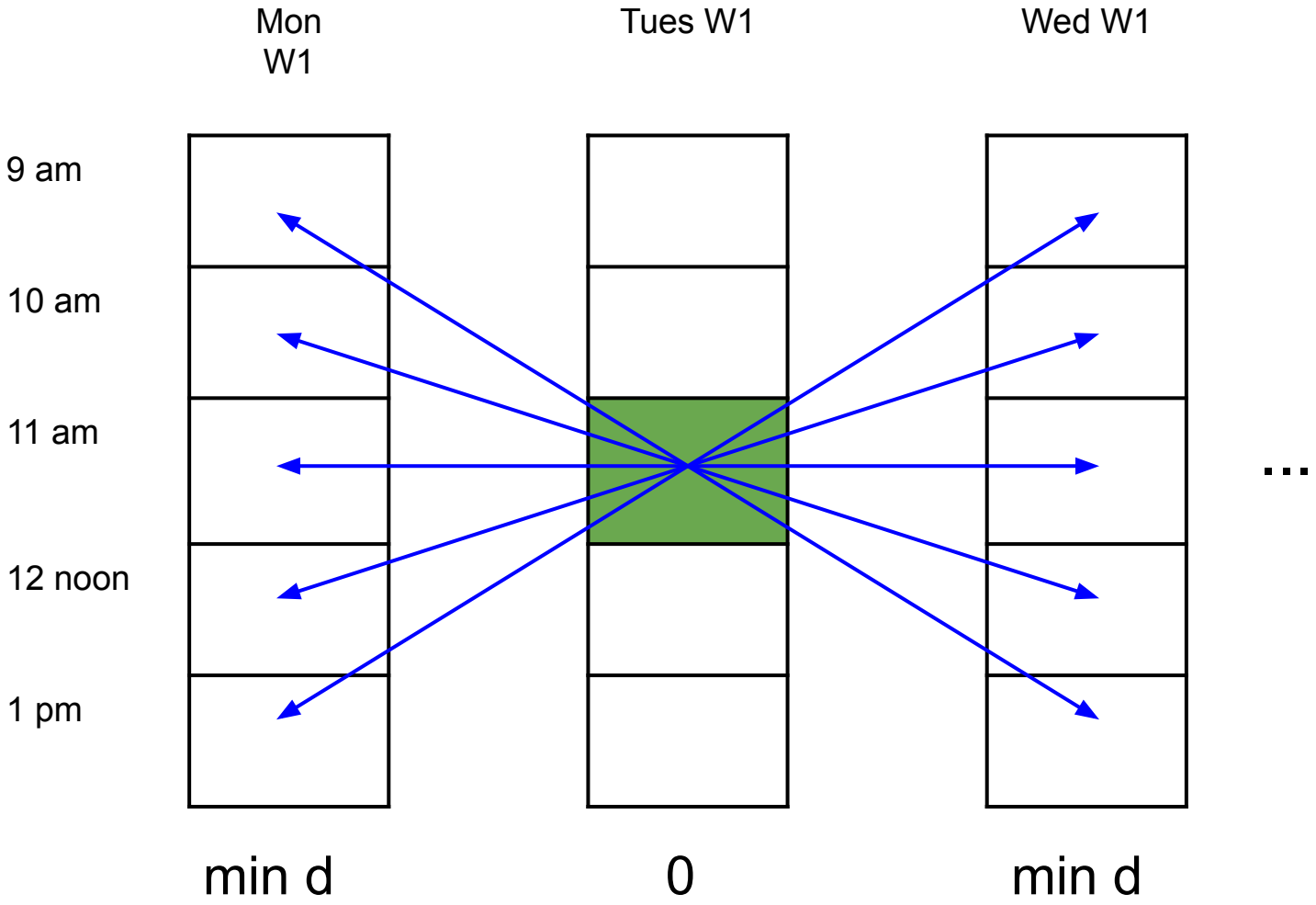

Sa 16:00
F 11:24
Th 10:33
T 14:09
M 16:36
M 08:49
M 08:49
M 16:36
T 14:09
Th 10:33
F 11:24
Sa 16:00
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2

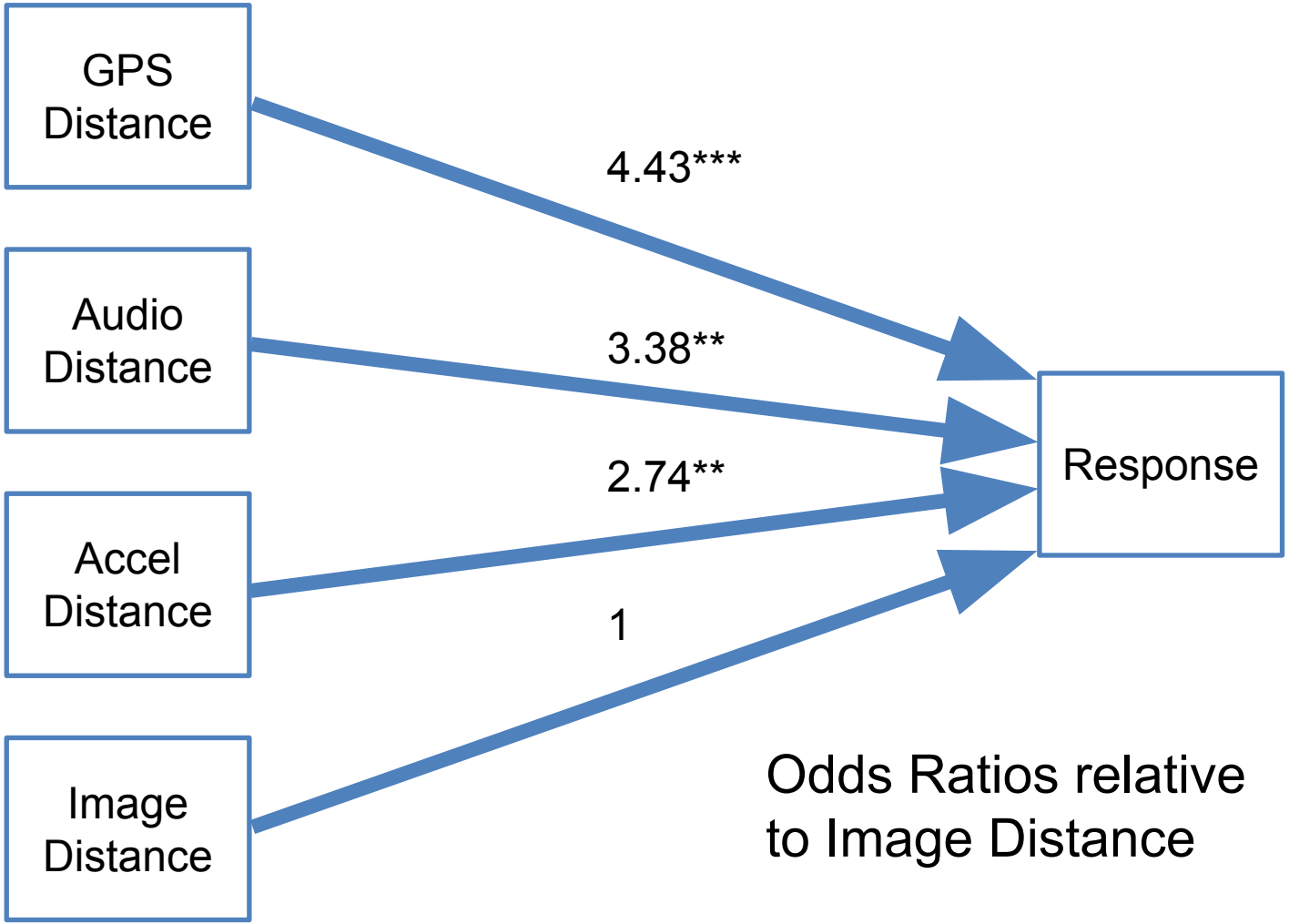


Recurrence plot of 30823 SD emails from 2007 to 2012



In lab, we manipulate place and time
But activity seems to dominate


Mon
W1
Tues W1
Wed W1
9 am
10 am
11 am
12 noon
1 pm
...
min d
0
min d


GPS Distance
Audio Distance
Accel Distance
Image Distance
4.43***
3.38**
2.74**
1
Response
Odds Ratios relative to Image Distance