Expanding the Exposure Assessment Landscape Using New Information Tools

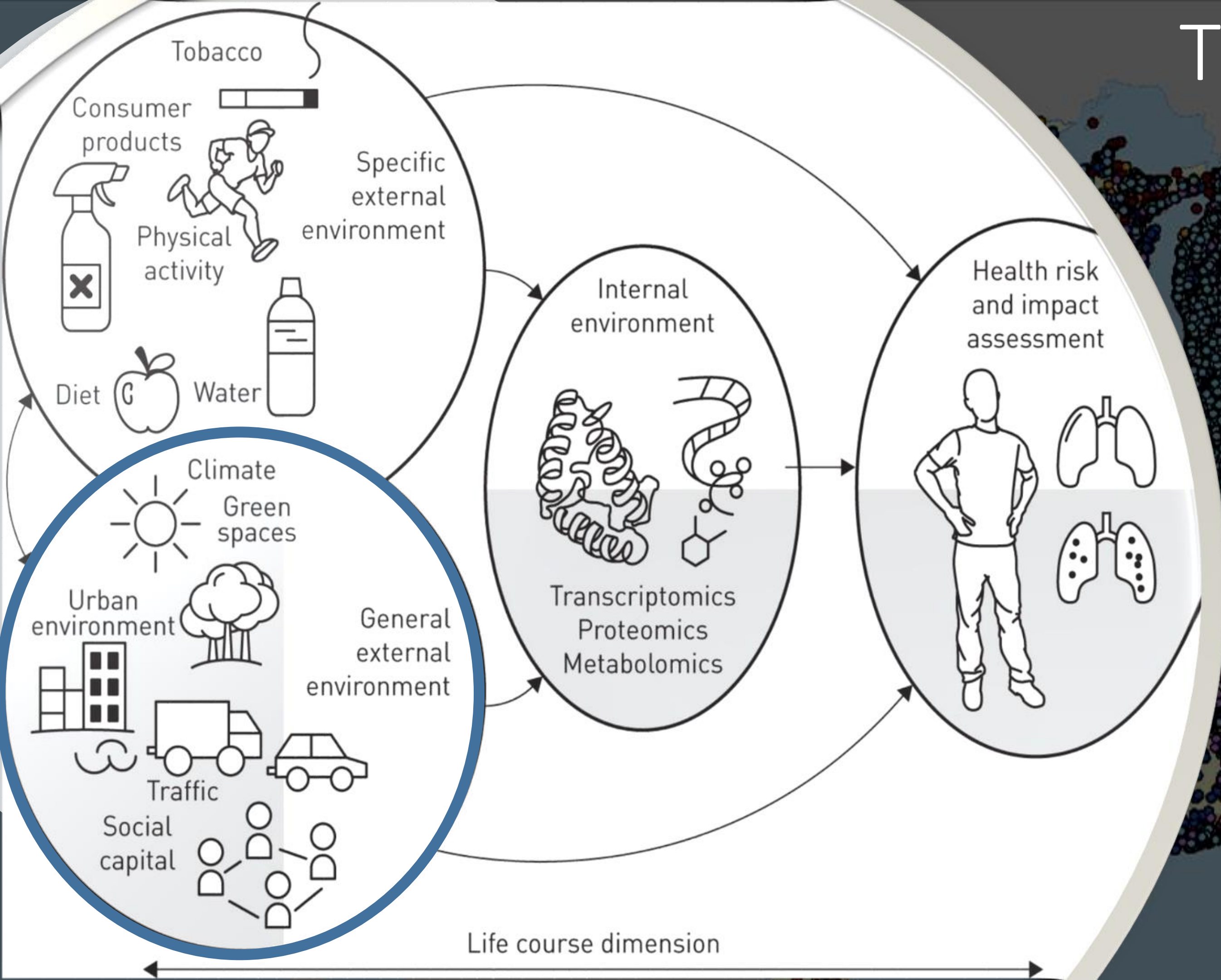
Peter James, ScD
Associate Professor of Population Medicine
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute

Associate Professor
Department of Environmental Health
Harvard TH Chan School of Public Health

HEI Annual Conference
May 1, 2023

The Places in Which We Live, Work, and Play Influence Our Health

The Exposome



Siroux et al. 2016

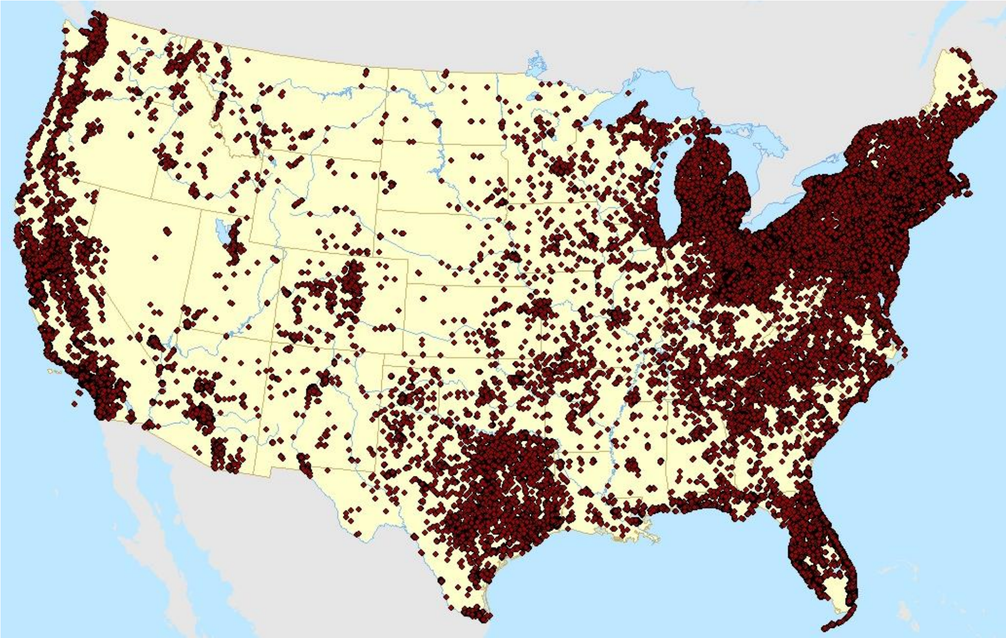
Nurses' Health Study Cohort

Began in 1976 with 121,701 female nurses aged 30-55

Originally from 11 States, have moved throughout the United States

Biennial questionnaires with disease and mortality follow-up

Residential mailing addresses geocoded from 1986-2018



Location of NHS residential addresses over follow-up

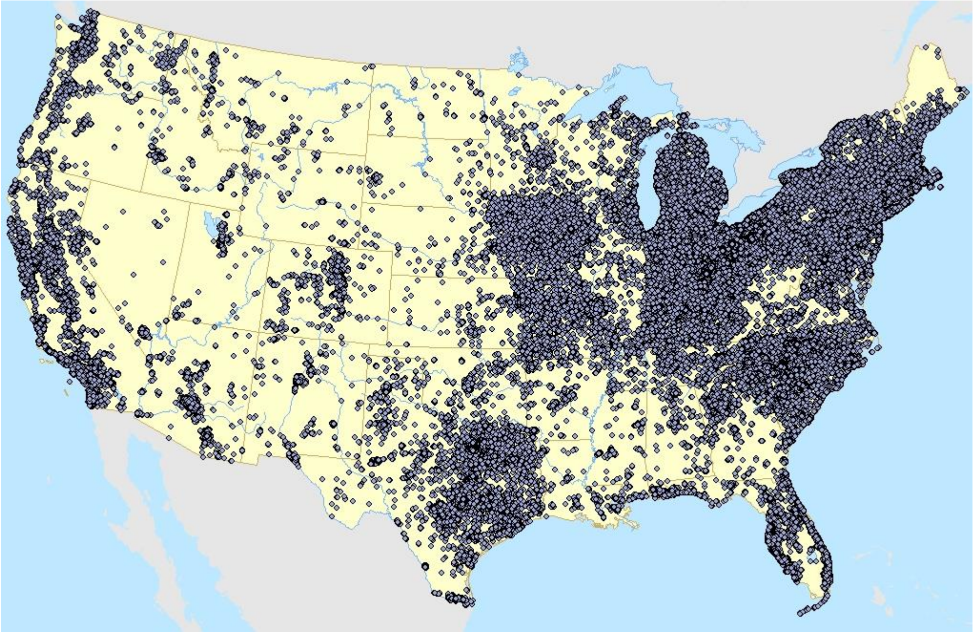
Nurses’ Health Study II Cohort

Began in 1989 with 116,000 female nurses aged 25-42

Originally from 14 States, have moved throughout the United States

Biennial questionnaires with disease and mortality follow-up

Biennial addresses geocoded from 1989-2017



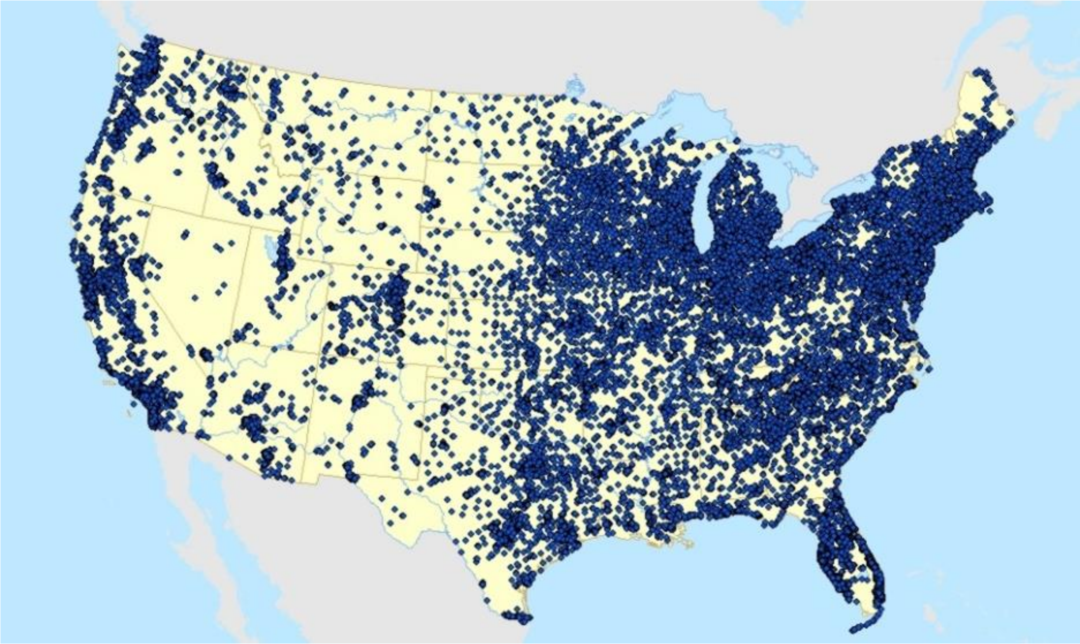
Location of NHS II residential addresses over follow-up

Health Professionals Follow-Up Study

Began in 1986 with 51,529 male health professionals (e.g., dentists, optometrists)

Biennial questionnaires with disease and mortality follow-up

Biennial addresses geocoded from 1986-2018



Location of HPFS mailing addresses over follow-up

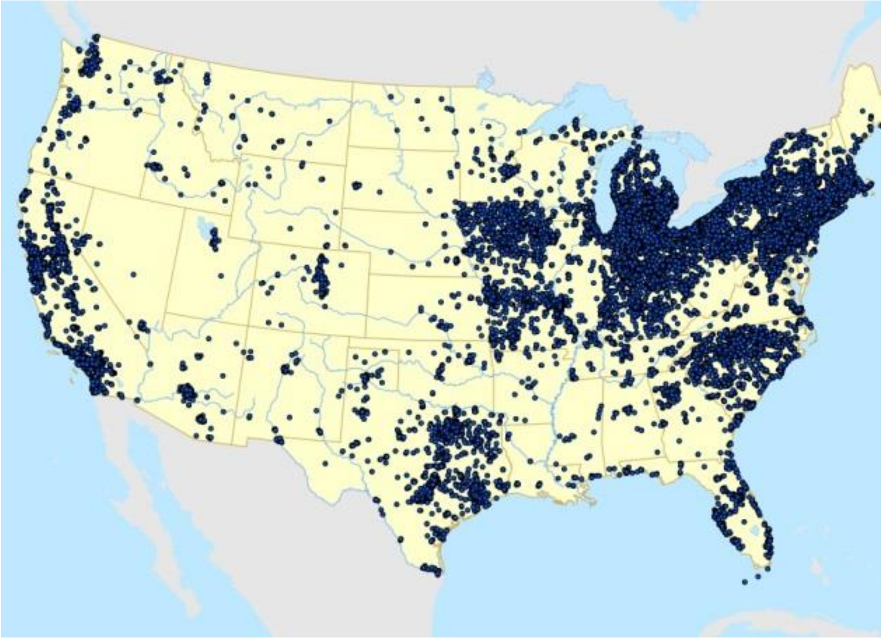
Growing Up Today Study

Children of NHSII participants

GUTS1 (N=16,882) enrolled in 1996, ages 9-14

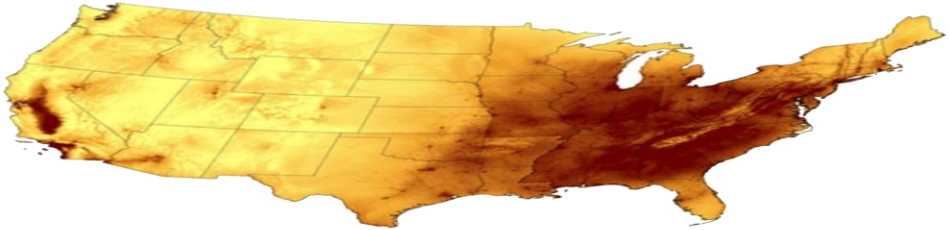
GUTS2 (N=10,923) enrolled in 2004, ages 10-17

Complete follow-up questionnaires each year through age 18, every other year after age 18



Location of GUTS residential addresses over follow-up

Environmental Exposures

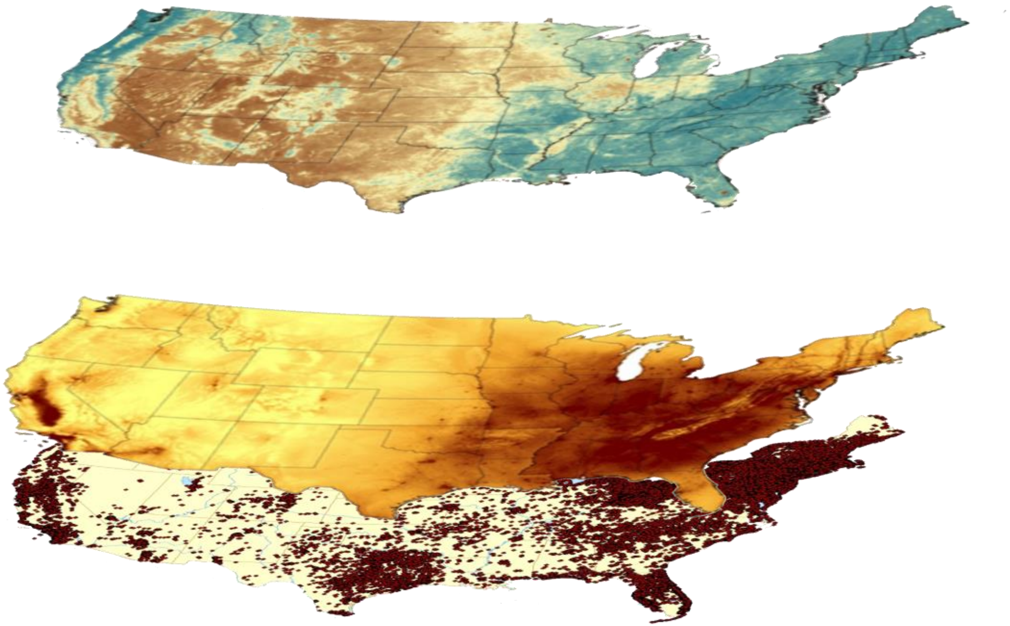


$PM_{2.5}$

- Spatiotemporal model to predict monthly $PM_{2.5}$ levels at any point (address) in the coterminous US
    - Generalized additive mixed models
    - EPA Air Quality Monitoring System data
    - Geospatial and meteorological predictors

Graphics by Cindy Hu

Environmental Exposures

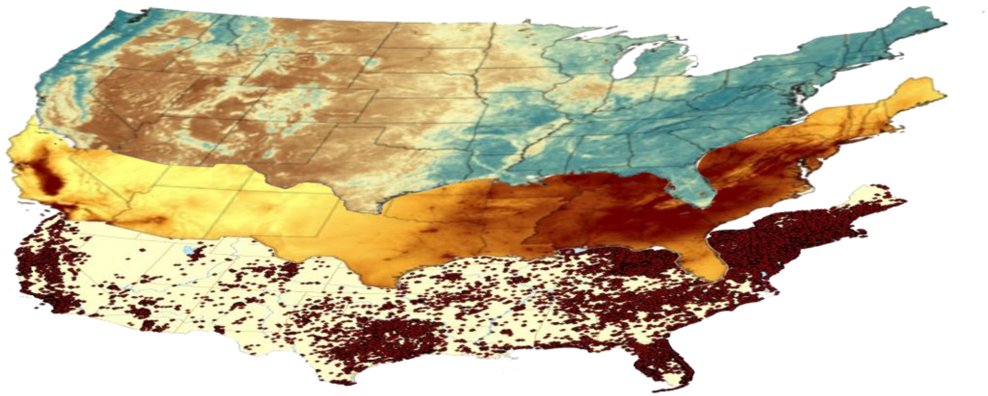


Greenness

- Normalized Difference Vegetation Index (NDVI)
  - From Landsat satellite imagery
- Satellite measure of vegetation based on the wavelengths absorbed by photosynthetically active plants
- Available at 30m resolution every 16 days across the globe from 1984-present

Graphics by Cindy Hu

Environmental Exposures

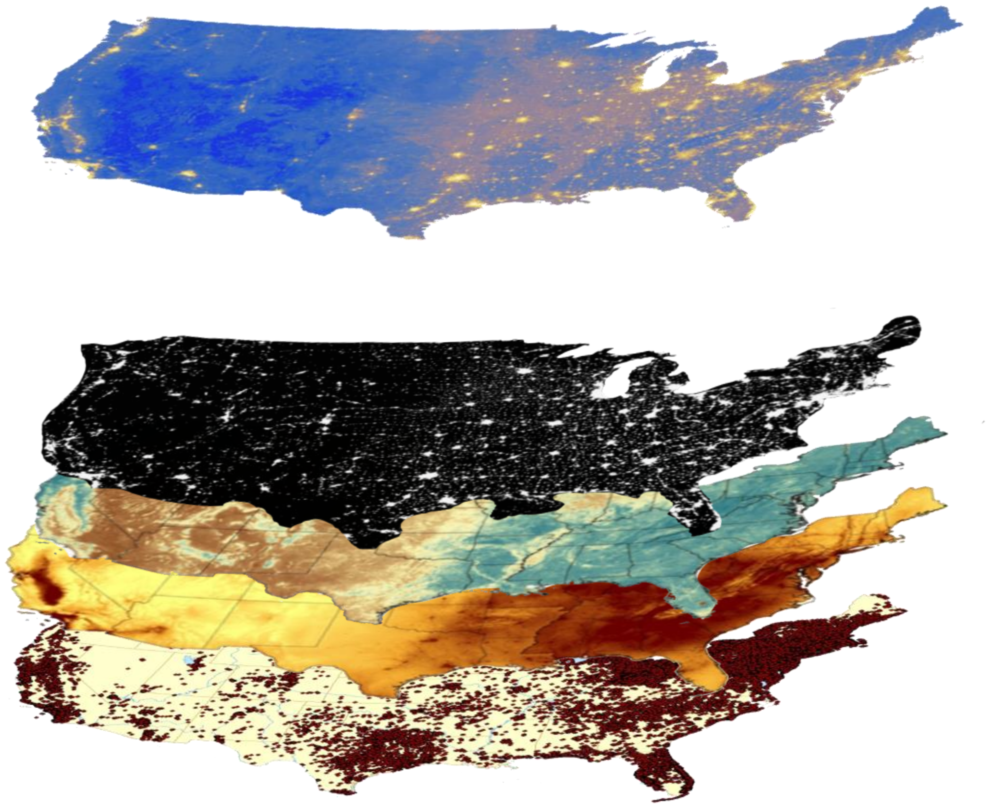


Light at Night

- From the U.S Defense Meteorological Satellite Program's Operational Linescan System
- Satellite imagery to detect light at night in nighttime radiance units
- Annual averages at 1 km^2 resolution starting in 1990s

Graphics by Cindy Hu

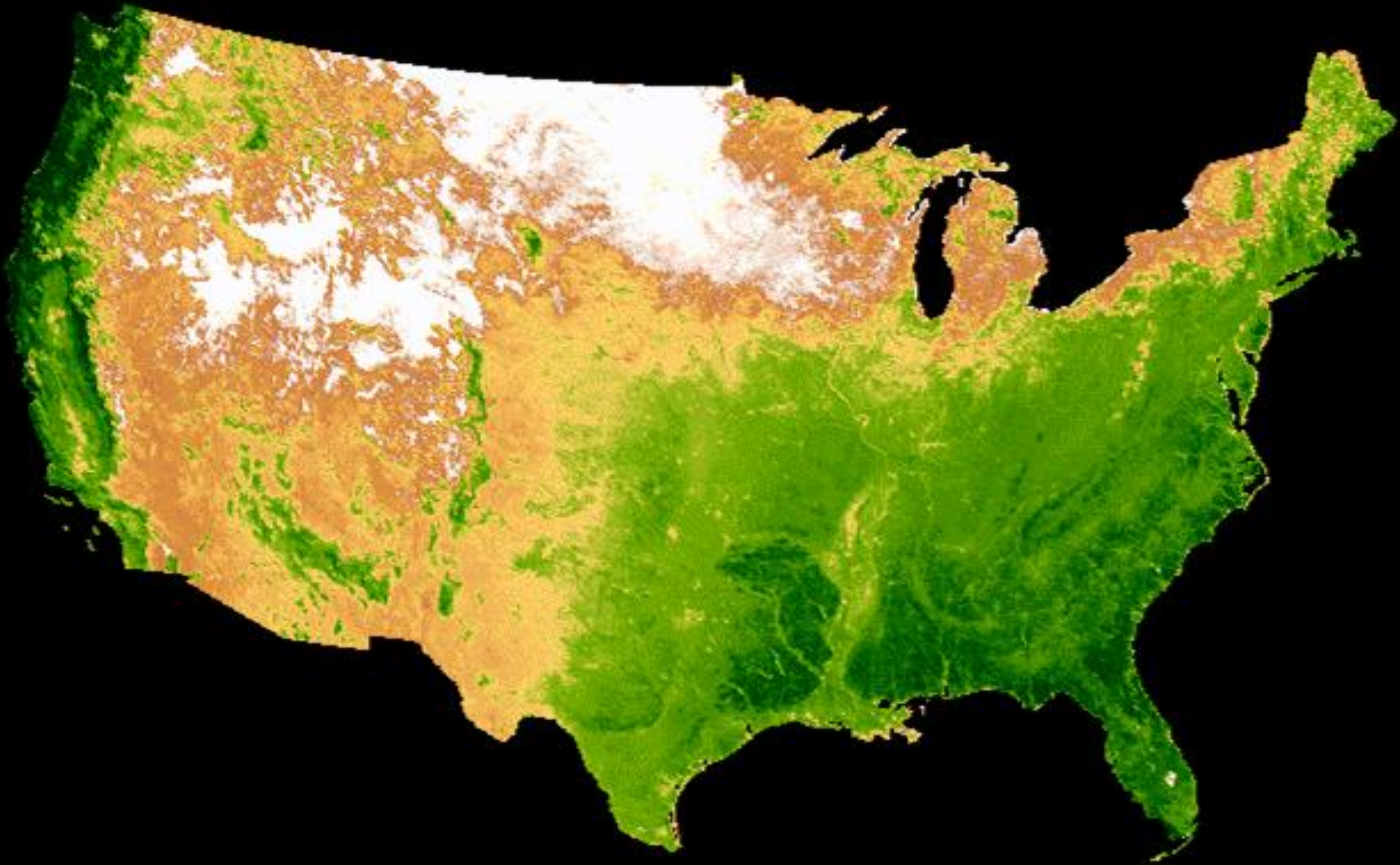
Environmental Exposures



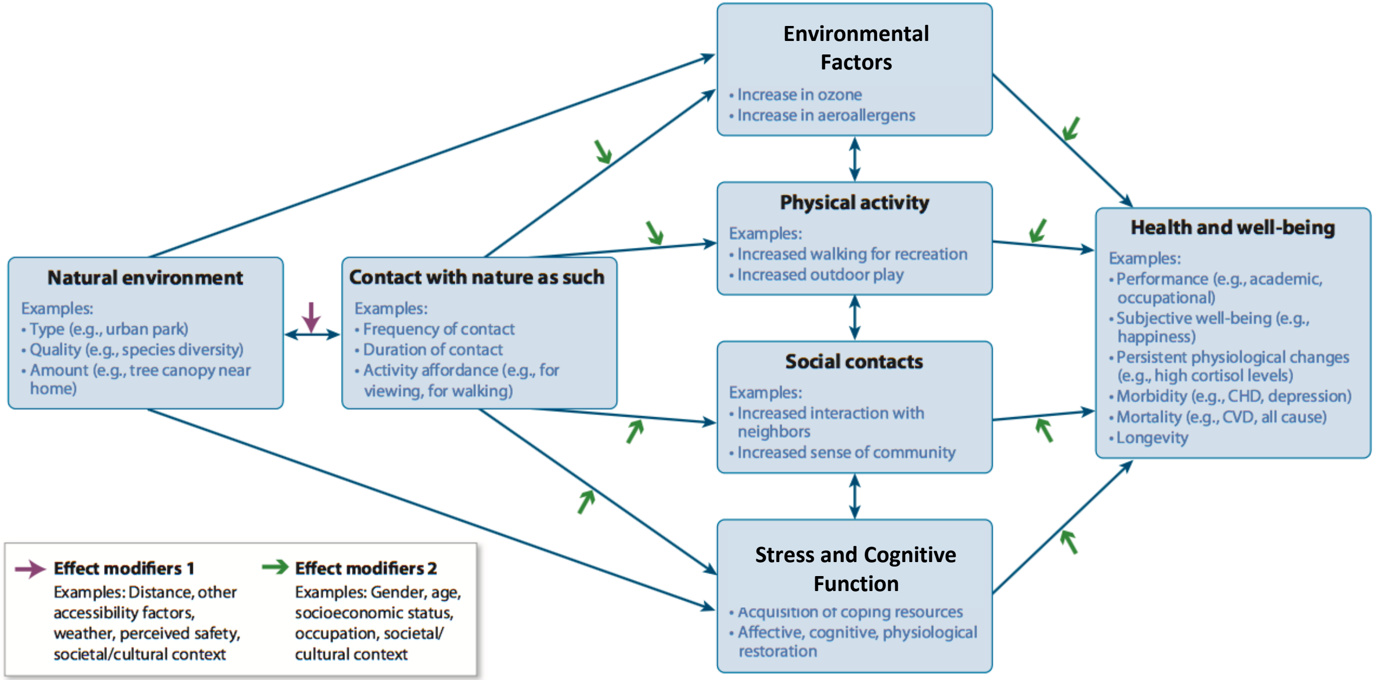
Noise

- From the US National Park Service model of outdoor noise, using noise monitor data from 2000 – 2014 and land use regression
    - 270 m^2 grid
    - Anthropogenic nighttime noise median, Ldn, Leq, and other metrics available

Graphics by Cindy Hu

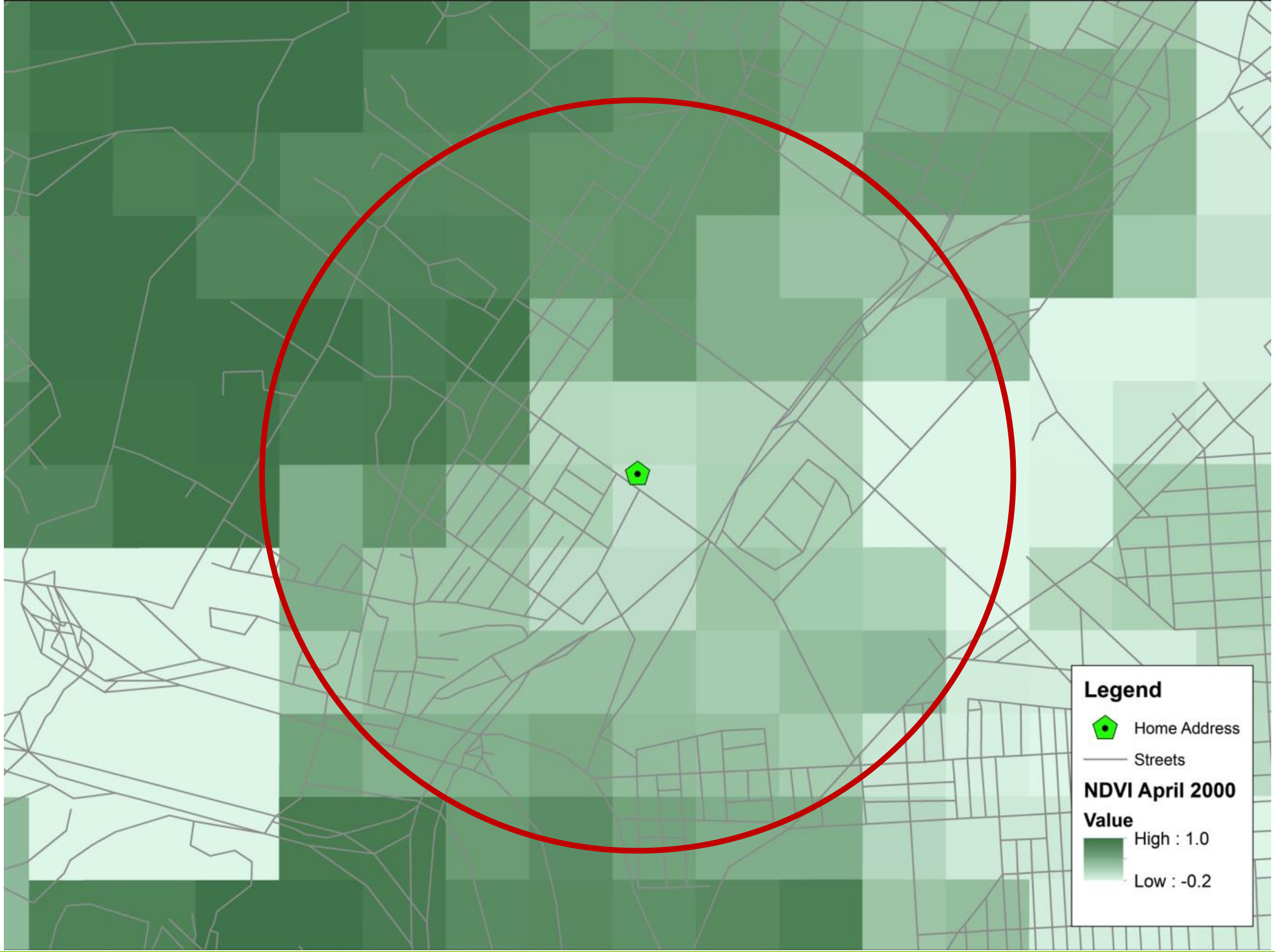


Mechanisms for Nature and Long-Term Health Outcomes



Hartig et al. 2014

James et al. *Environmental Health Perspectives*. 2016.



James et al. *Environmental Health Perspectives*. 2016.

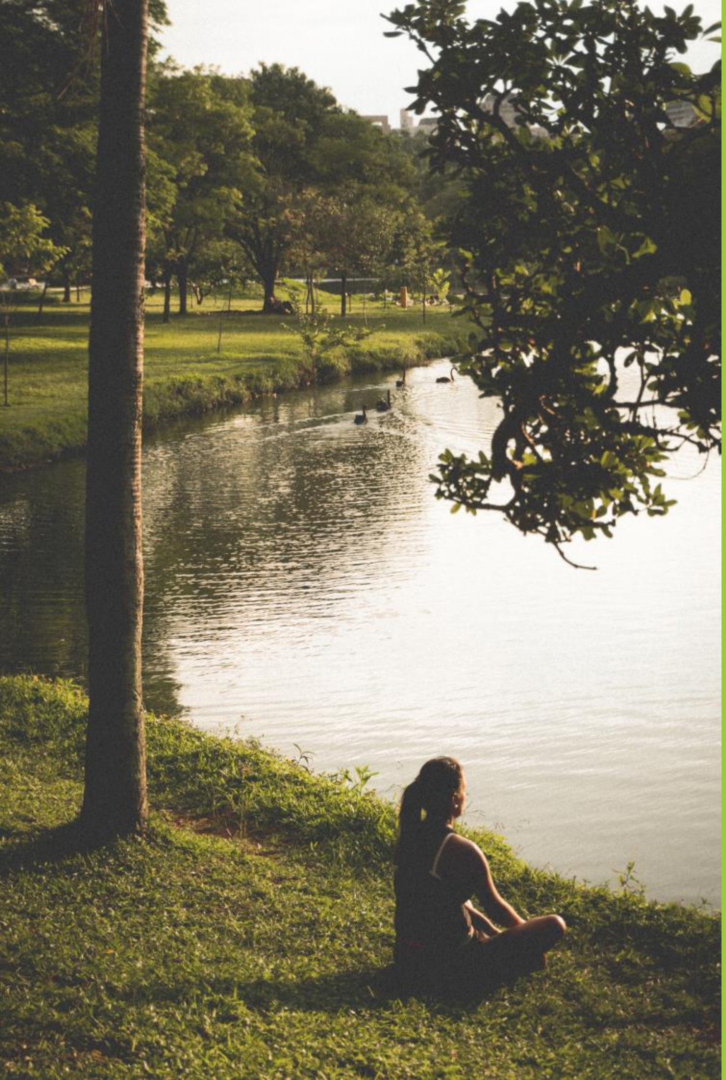
Greenness and All-Cause Mortality in the Nurses' Health Study (N=108,630 from 2000-2008)

Cumulative Average Greenness in 250m	Fully Adjusted HR (95% CI)
Greenness Quintile 1	Ref
Greenness Quintile 2	0.92 (0.86, 0.98)
Greenness Quintile 3	0.90 (0.84, 0.96)
Greenness Quintile 4	0.94 (0.88, 1.00)
Greenness Quintile 5	0.88 (0.82, 0.94)
P for Trend	0.002

Hazard ratios are adjusted for calendar time, age, race, smoking, individual SES, Census tract median home value and income

James et al. *Environmental Health Perspectives*. 2016.

Growing Scientific Evidence for Health Benefits of Nature in an Exposomic Framework

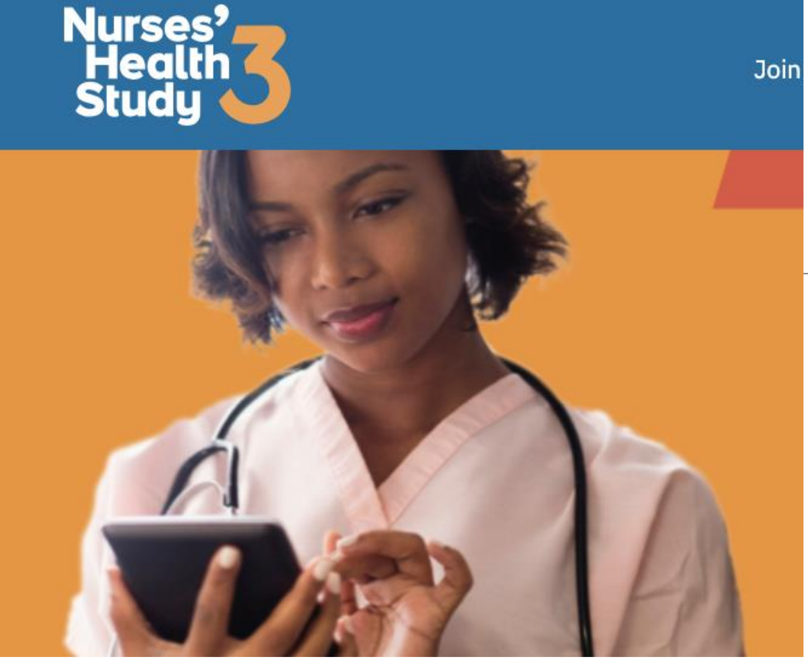


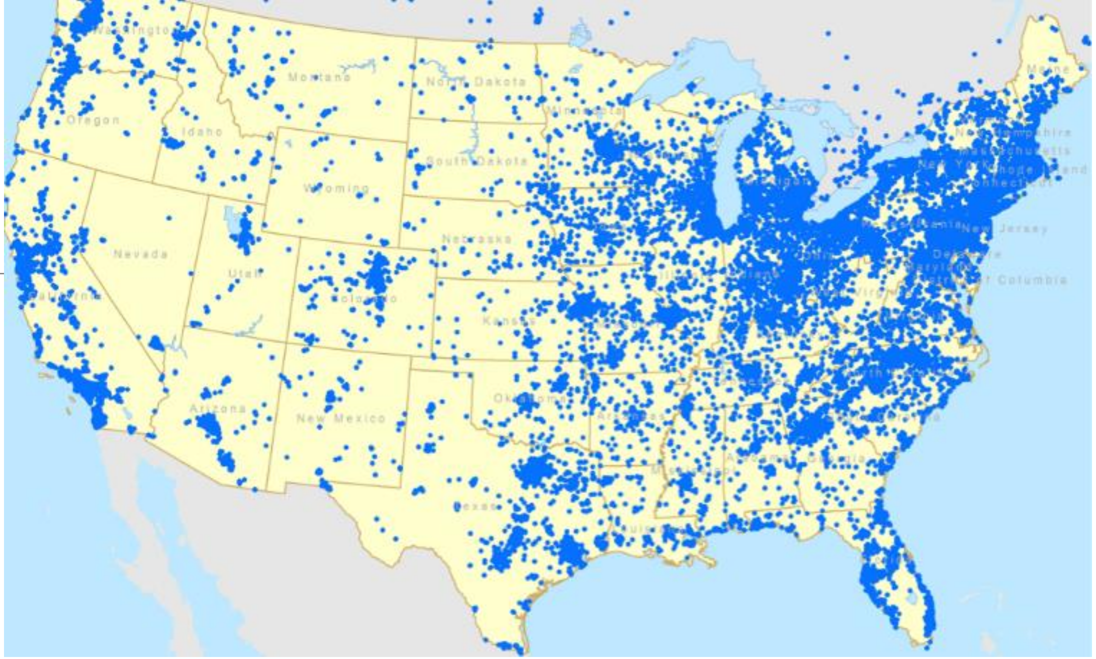
Limitations of Epidemiologic Research on Nature and Health

Is "greenness" around the home the right measure?

- What area around the home? Are we capturing contact with or exposure to nature?
- We focus on vegetation, but what are the specific "active ingredients" in nature?
- Without specificity, we are limited in causal inference and in policy relevance

What tools can we use to address these limitations?





Nurses' Health Study 3

nhs3.org

Open cohort of 49,000+ nurses and nursing students ages 19-46 across the entire US

Web-based questionnaires every six months

Participants open emails using predominantly smartphones

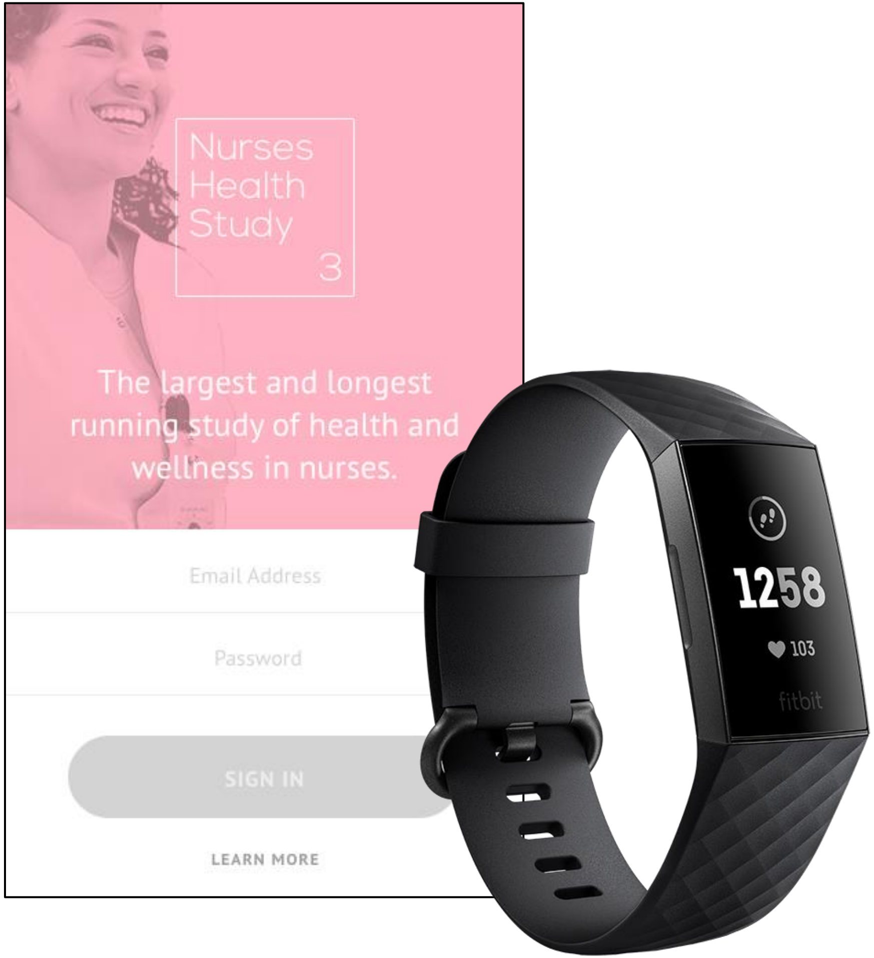
Nurses' Health Study 3 Mobile Health Substudy

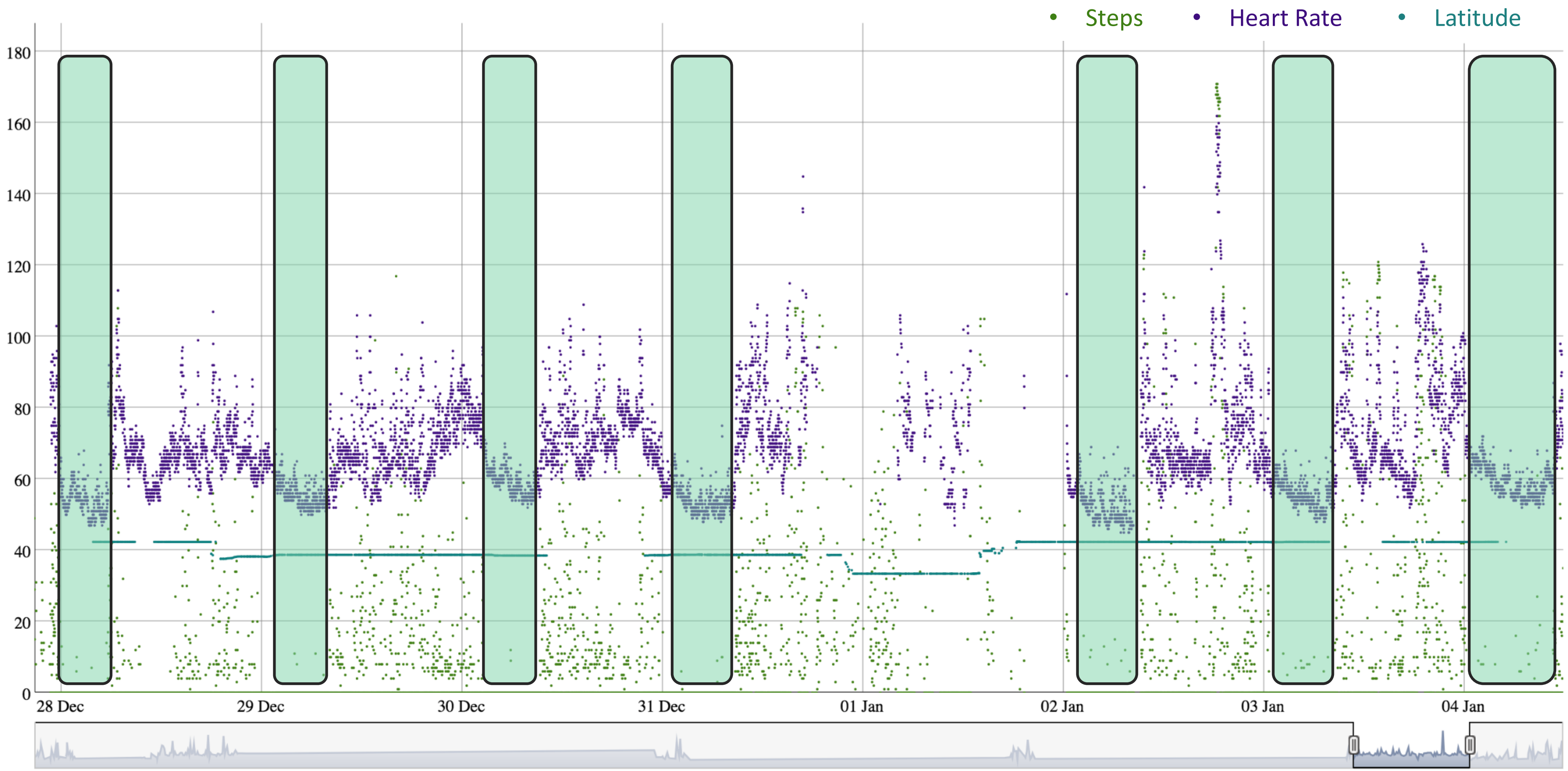
450 participants underwent seven-day sampling periods four times over a year

Fitbit device to measure **steps, heart rate, and sleep at the minute-level**

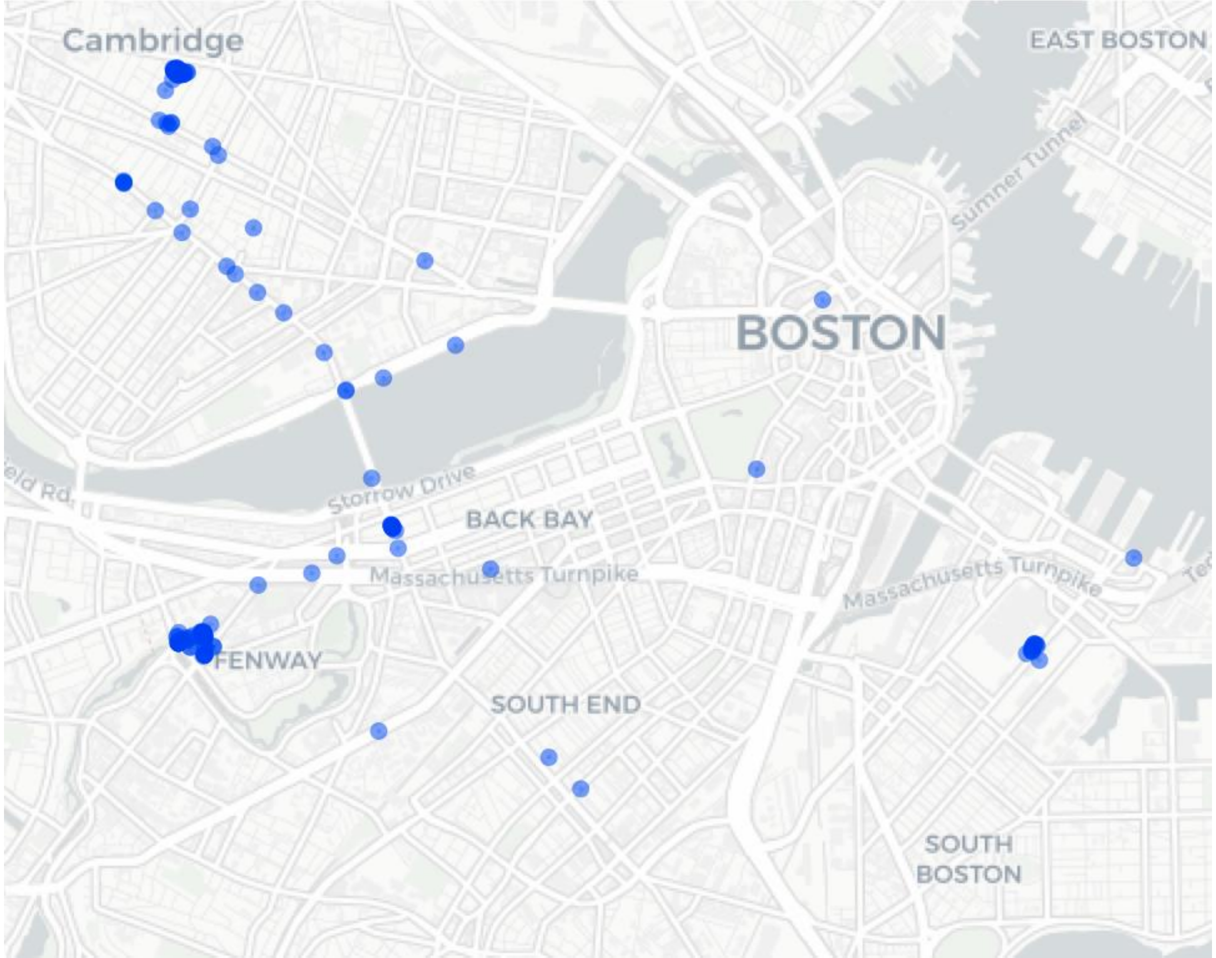
Custom smartphone app to measure **location and administer questionnaires**

Efficient, passive, objective, low-cost measurement of high spatial- and temporal-resolution data on health behaviors

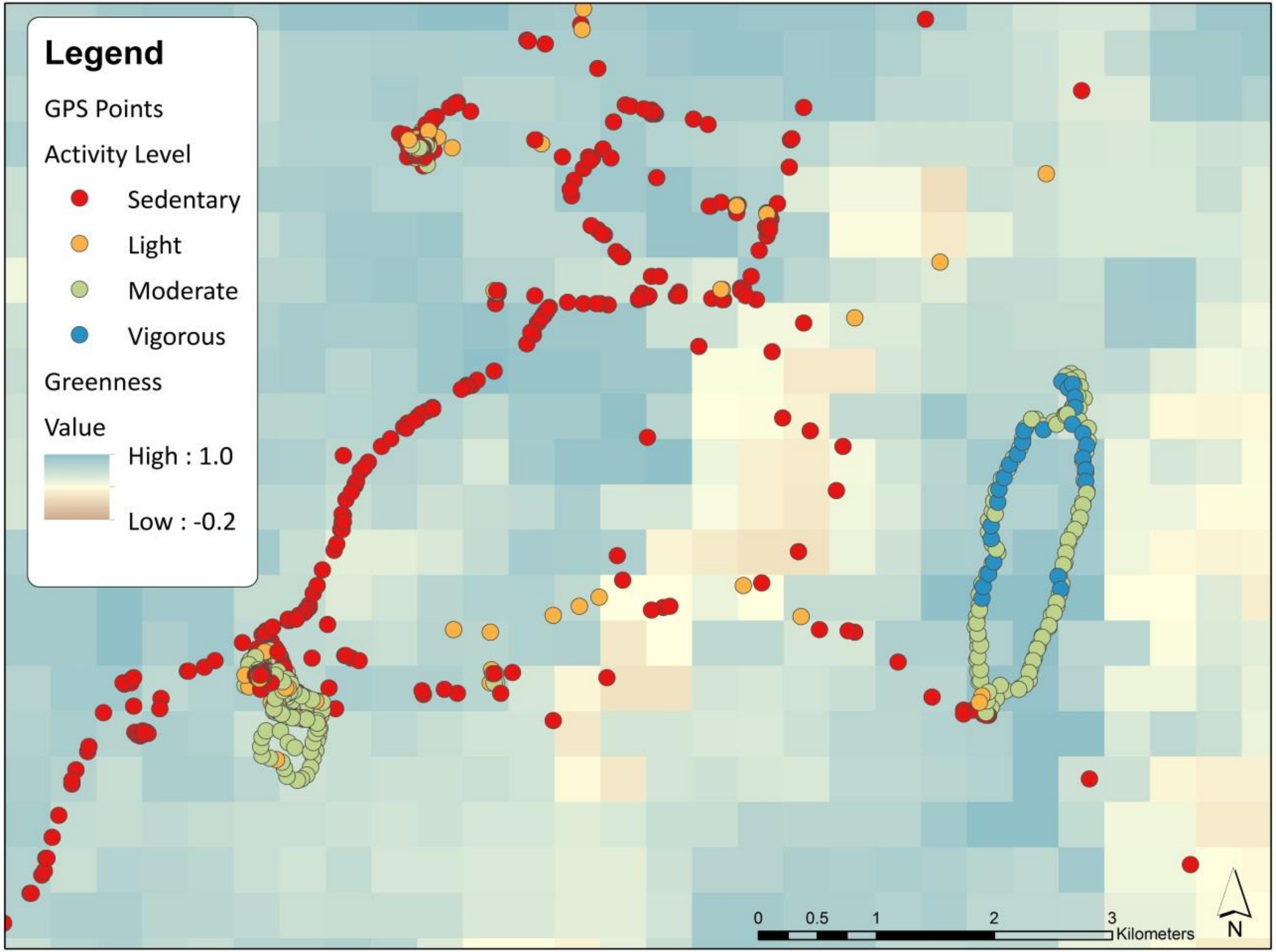



Steps
Heart Rate
Latitude
180
160
140
120
100
80
60
40
20
0
28 Dec
29 Dec
30 Dec
31 Dec
01 Jan
02 Jan
03 Jan
04 Jan

Cambridge
EAST BOSTON
Soldiers Field Rd
Sumner Tunnel
BOSTON
Massachusetts Turnpike
ALLSTON
Soldiers Field Rd
Storrow Drive
BACK BAY
Massachusetts Turnpike
Massachusetts Turnpike
FENWAY
SOUTH END



Cambridge
EAST BOSTON
Sumner Tunnel
BOSTON
Storrow Drive
BACK BAY
Massachusetts Turnpike
Massachusetts Turnpike
FENWAY
SOUTH END
SOUTH BOSTON

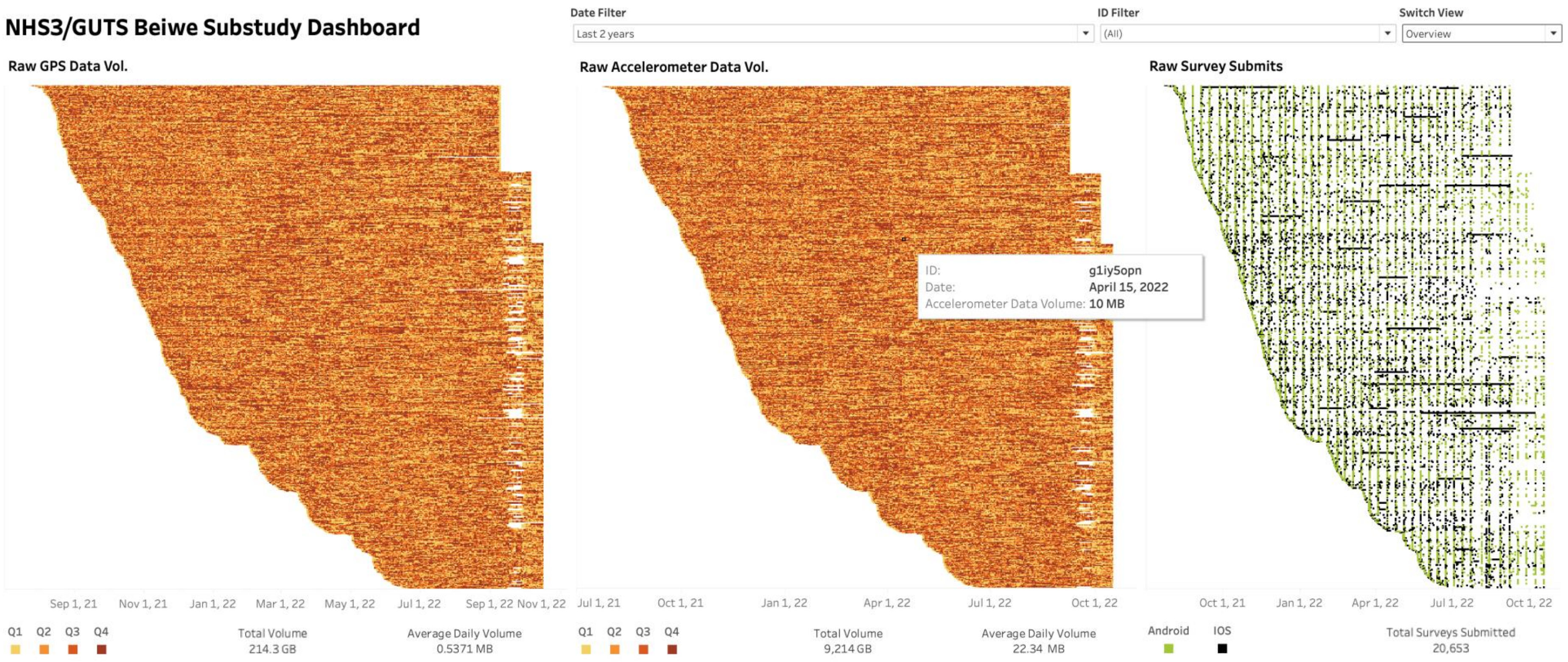

Legend
GPS Points
Activity Level
Sedentary
Light
Moderate
Vigorous
Greenness
Value
High : 1.0
Low : -0.2
0
0.5
1
2
3
Kilometers
N

Digital Phenotyping in Epidemiological Cohorts

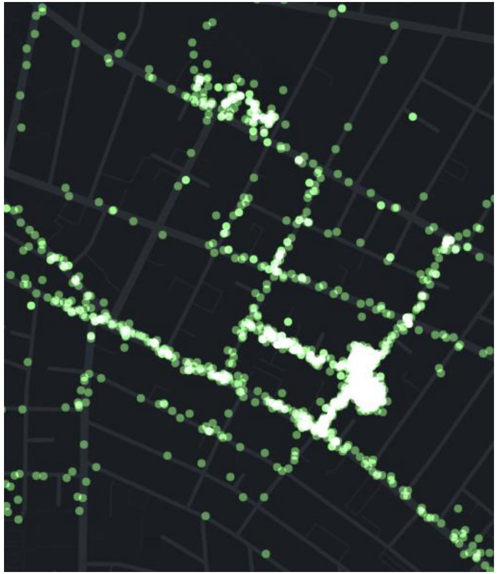
Stored and processed by **cloud computing** infrastructure

The collection of **smartphone-based** GPS, accelerometer, sleep and social network data

Beiwe Smartphone Study (N=2,394)



Massive Data Volumes



An illustration of GPS trajectories of a test subject over a year

Average follow-up of **195 days** per participant

Averaged 12.7 valid hours of accelerometer data and 14.3 valid hours of GPS data (**430,000+ participant-days**)

23,000+ questionnaires completed with response rates range between 40%~50%

Collected data are currently under processing in **cloud-based servers** to derive environmental exposures and physical activity

Notes. Data summary as of November 1st, 2022. The data collection remains ongoing for ~50% of participants.

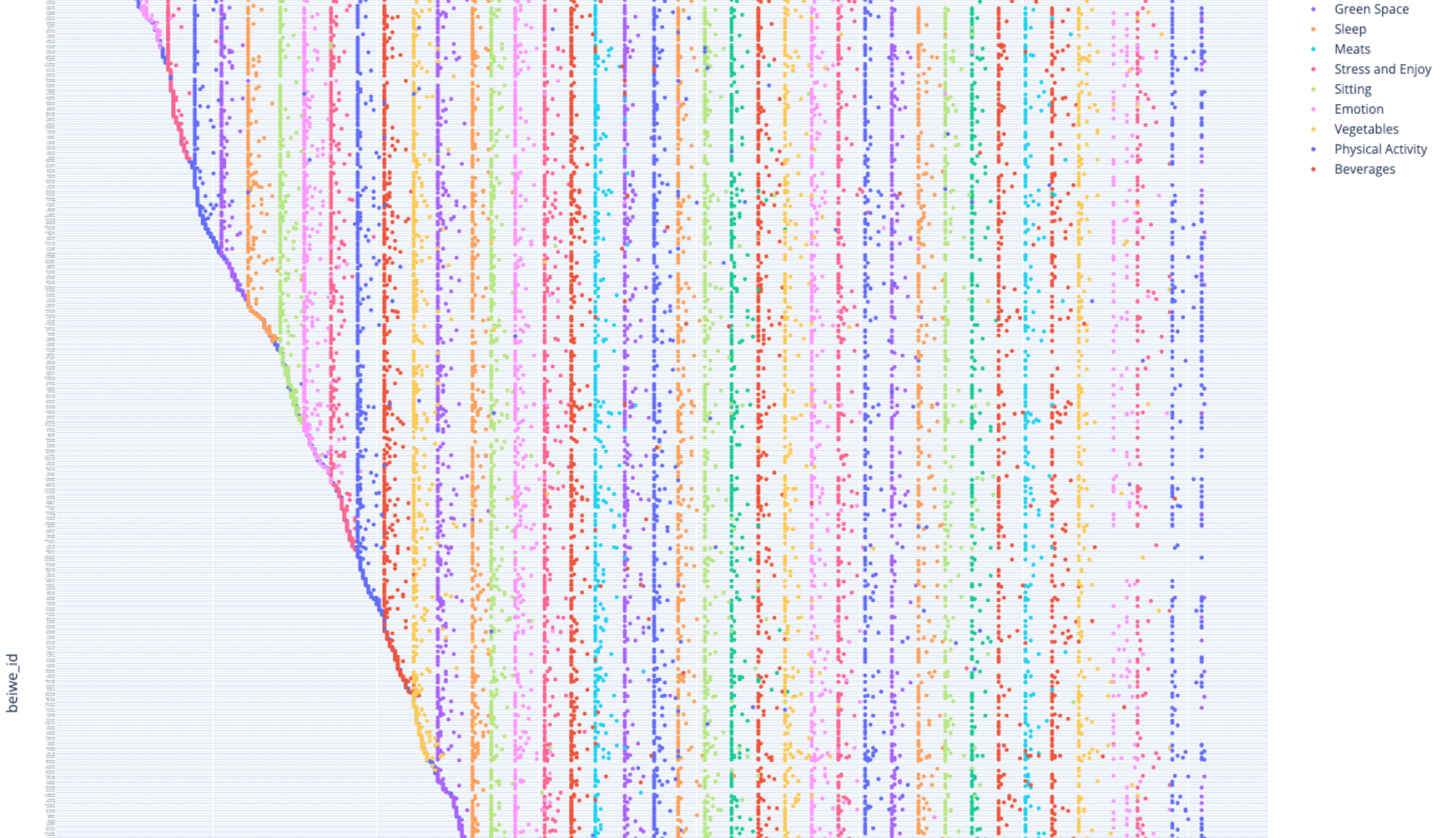
Potential Applications

14 months of GPS trajectories by day based on data from a test subject.

Health & Place
Volume 51, May 2018, Pages 52-60

How many days of global positioning system (GPS) monitoring do you need to measure activity space environments in health research?

Shannon N. Zenk [a], Stephen A. Matthews [b], Amber N. Kraft [c], Kelly K. Jones [a]



beiwe_id
Green Space
Sleep
Meats
Stress and Enjoy
Sitting
Emotion
Vegetables
Physical Activity
Beverages

Remote Assessment of Cognition through Smartphone Applications

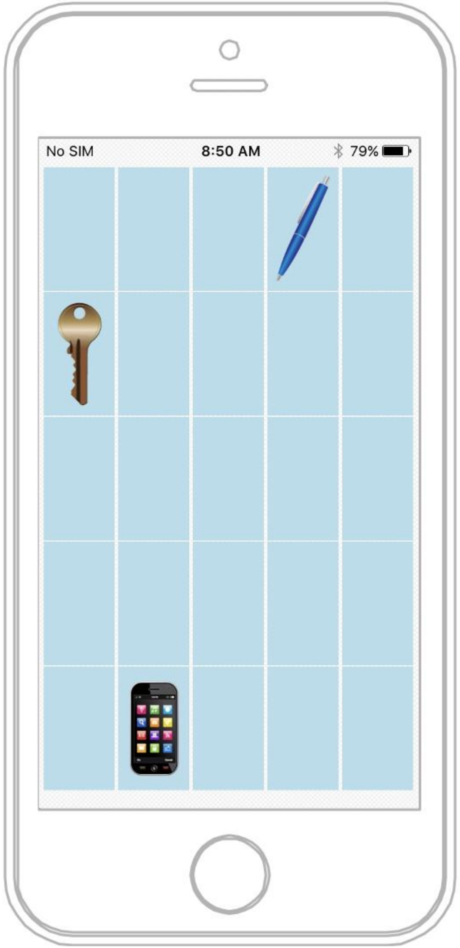
Grants under review to incorporate the Ambulatory Research in Cognition (ARC) platform into Beiwe

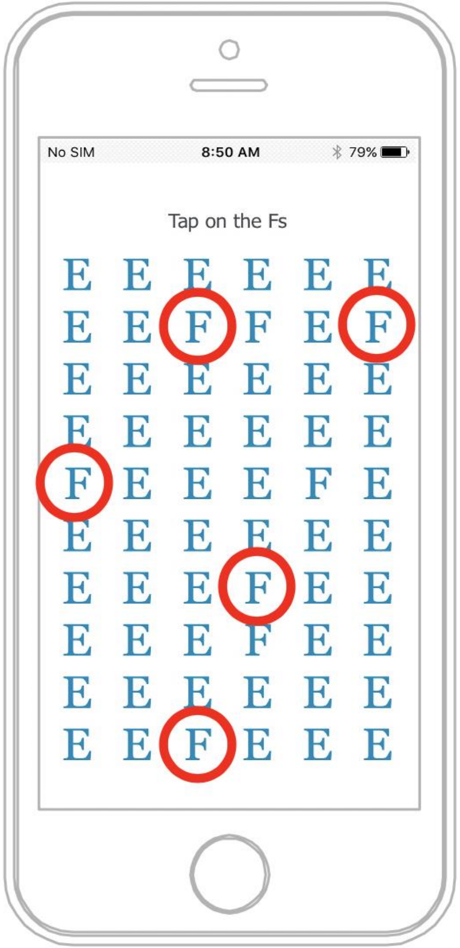
Repeated testing in free-living environments

- Reduces the impact of within-person variability
- Improves measurement precision
- Enhances assessment of ecologically valid cognitive function

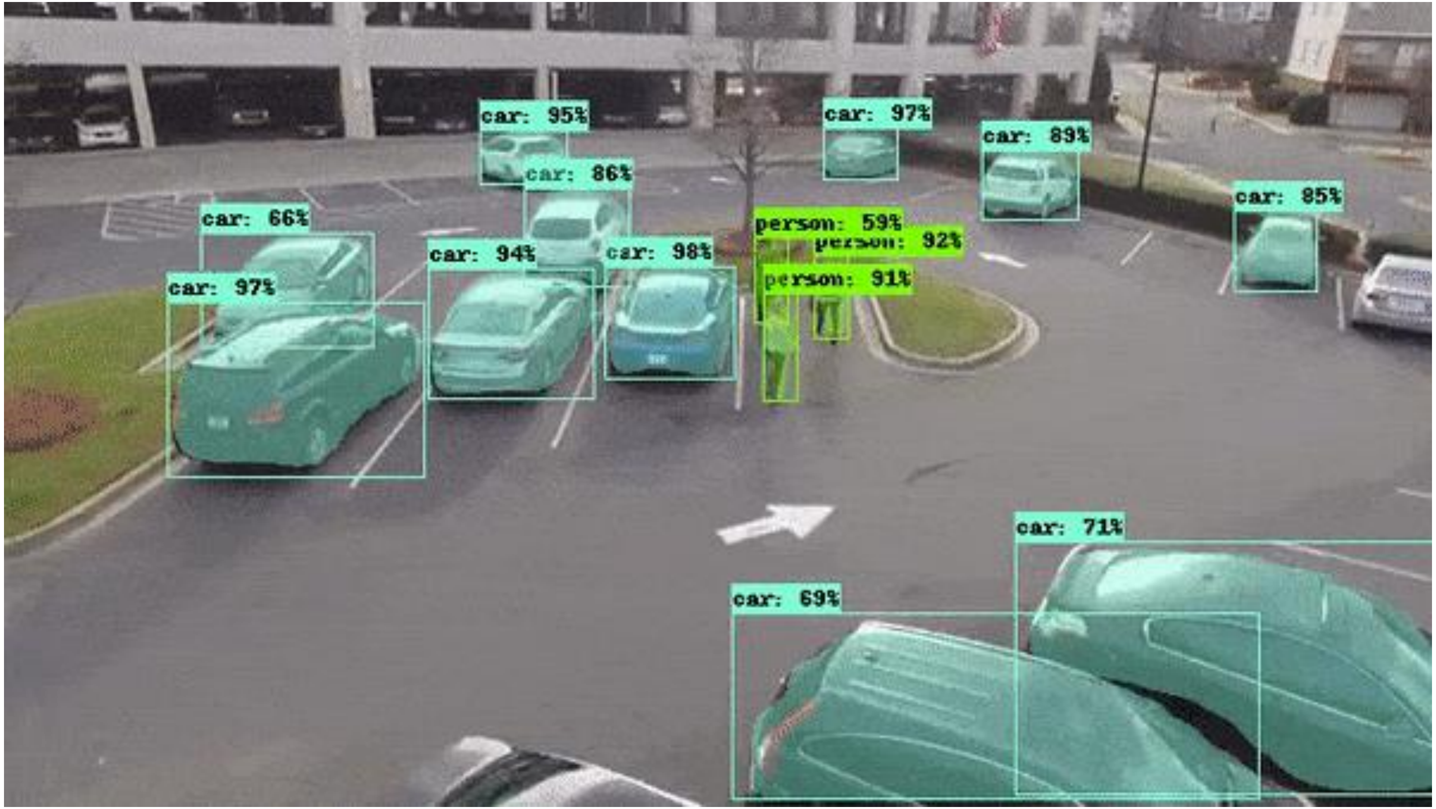
Completed in 2 minutes or less

Timestamped and geotagged





Exposure Assessment through Deep Learning



From: https://towardsdatascience.com/using-tensorflow-object-detection-to-do-pixel-wise-classification-702bf2605182

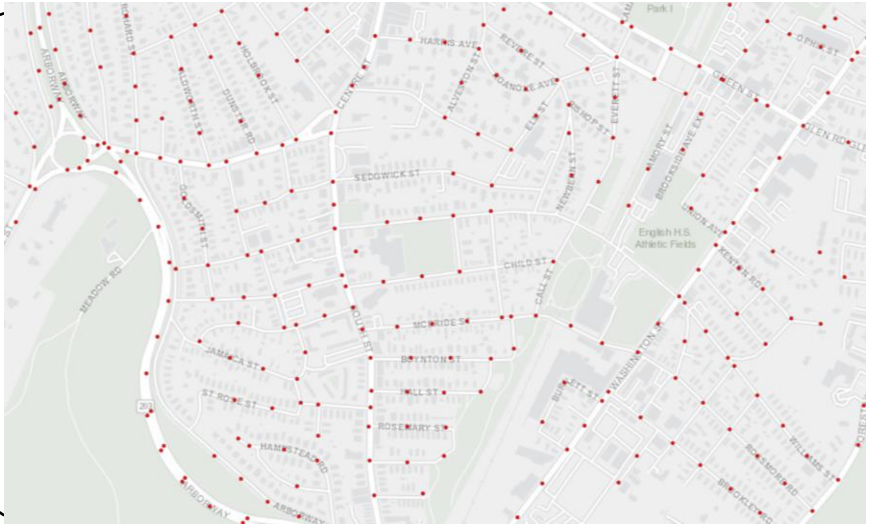
Image-based Approaches to Spatial Factors and Health

1. Satellite-based data tell us little about the quality of natural and built environments
2. Geocoded street-level images provide insight into specific environmental features from an on-the-ground perspective
3. Machine learning approaches can segment specific components of an image for analysis

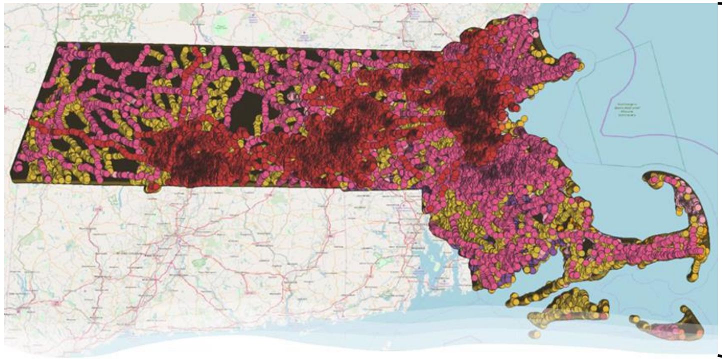
Legend
Home Address
Streets
NDVI April 2000
Value
High : 1.0
Low : -0.2


road
sidewalk
building
wall
fence
pole
traffic light
traffic sign
vegetation
terrain
sky
person
rider
car
truck
bus
train
motorcycle
bicycle

Create Grids



Obtain Near-Grid GSV Images from 2007-2022



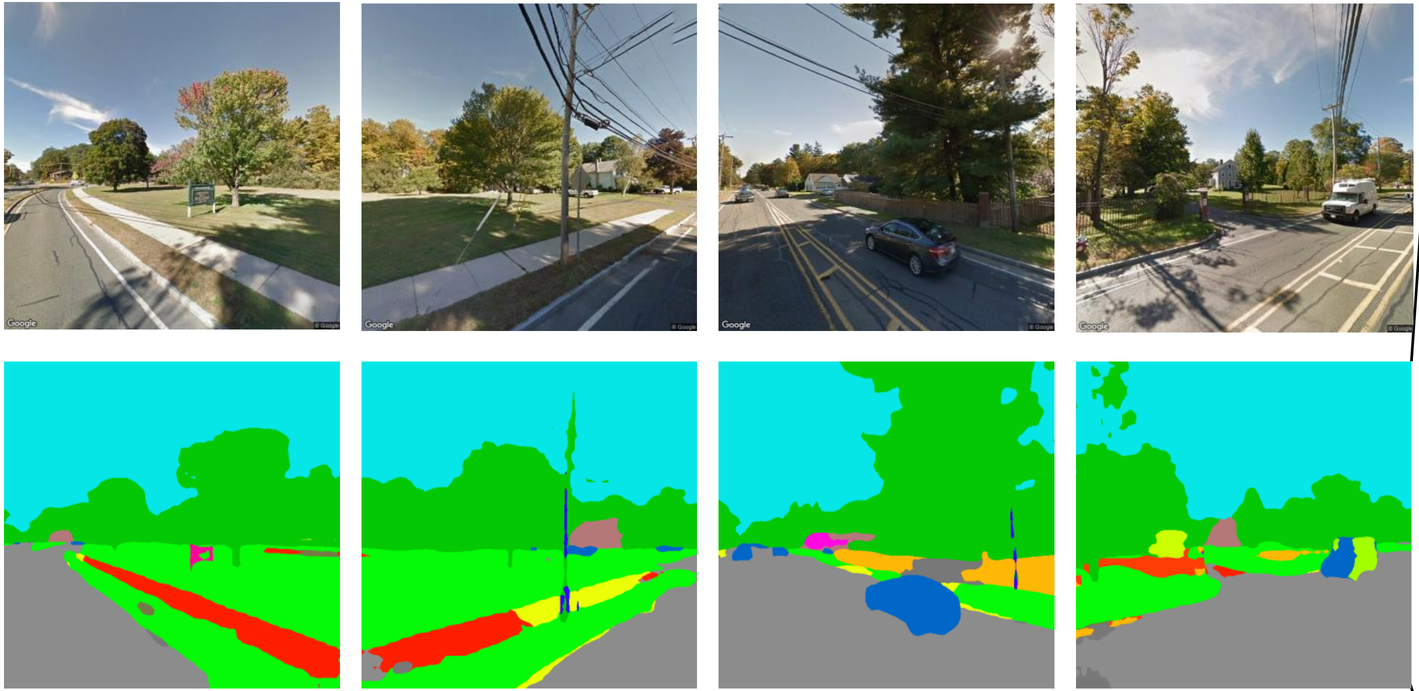
ALL

2018

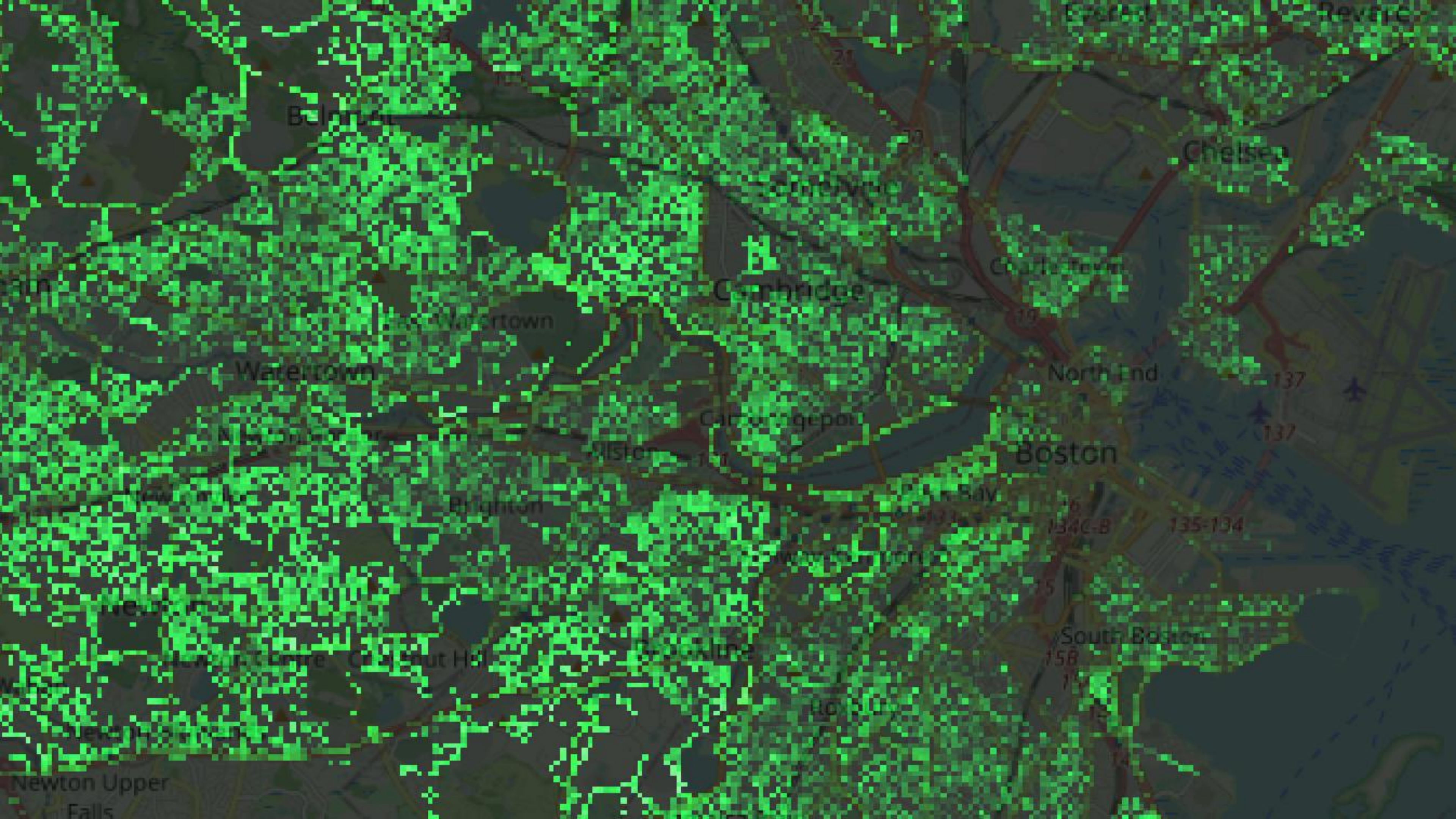
2017

Deep Learning of Google Street View Images to Develop Novel Spatial Exposures

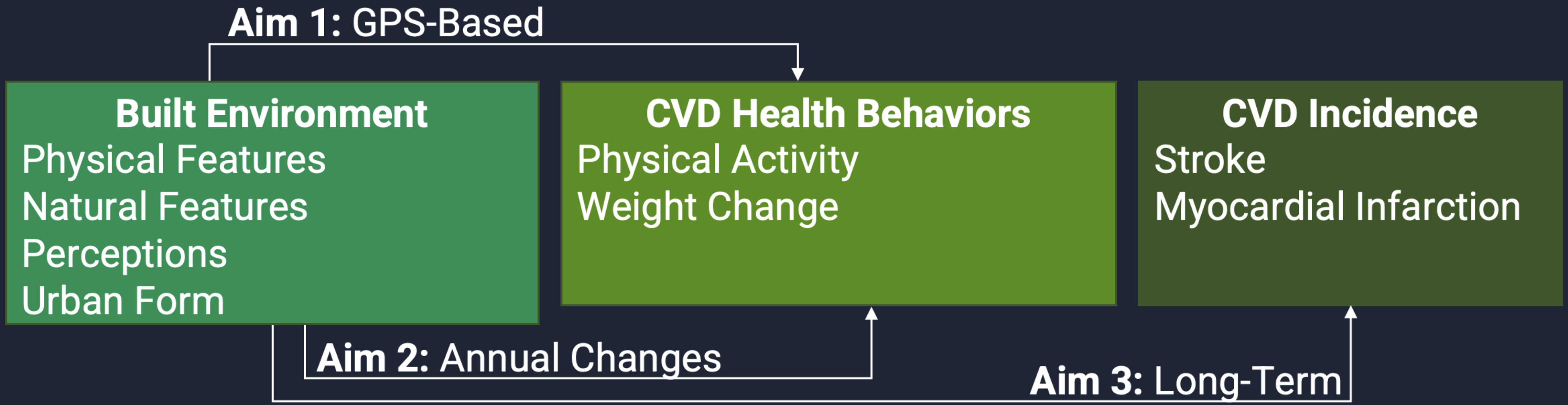
Process Google Street View Images



Idx	Name	Percentages
1	wall	1.0%
2	building;edifice	0.6%
3	sky	32.6%
4	floor;flooring	0.0%
5	tree	23.6%
6	ceiling	0.0%
7	road;route	22.3%
8	bed	0.0%
9	windowpane;window	0.0%
10	grass	13.3%
143	plate	0.0%
144	monitor;monitoring;device	0.0%
145	bulletin;board;notice;board	0.0%
146	shower	0.0%
147	radiator	0.0%
148	glass;drinking;glass	0.0%
149	clock	0.0%
150	flag	0.0%


Everett
Revere
Belmont
Chelsea
Cambridge
Charlestown
East Watertown
Watertown
North End
Boston
Allston
Brighton
Back Bay
South Boston
Brookline
Newton Centre
Chestnut Hill
Newton Upper Falls
137
135-134
134C-B
158

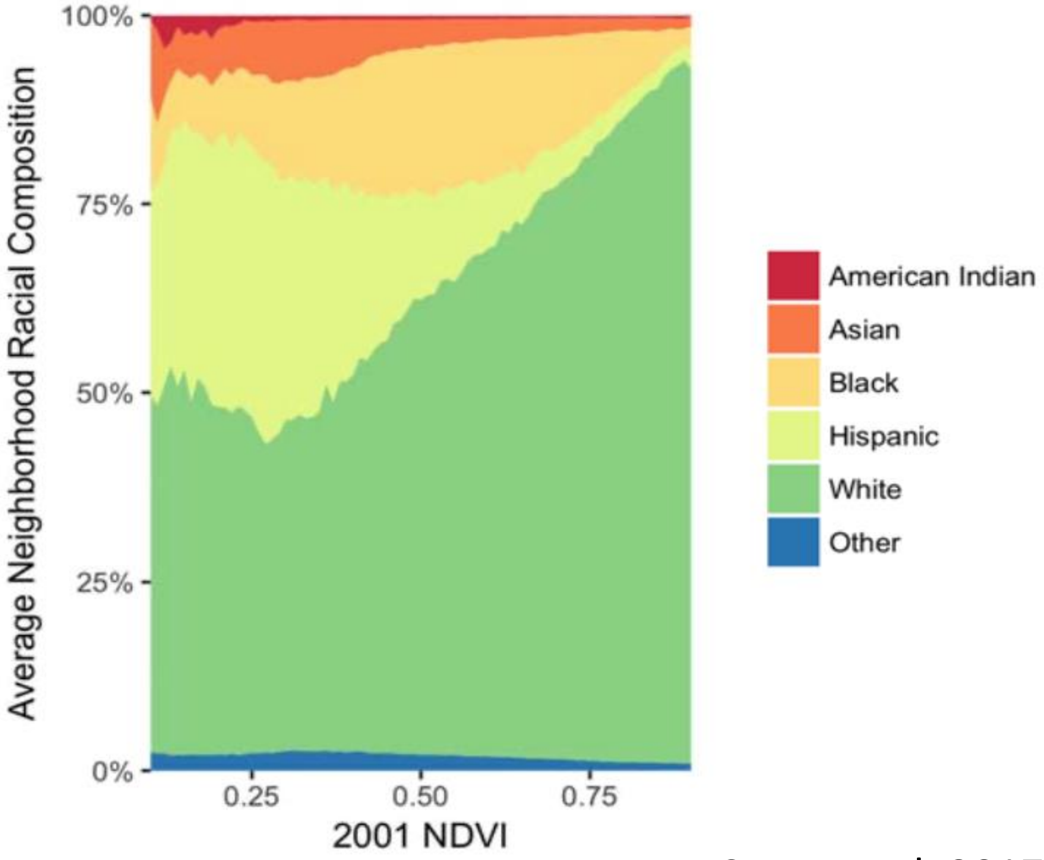
NHLBI-funded R01: Built Environment Assessment through Computer visiON (BEACON): Applying Deep Learning to Street-Level and Satellite Images to Estimate Built Environment Effects on Cardiovascular Health



Peter James (Harvard Medical School), Perry Hystad, Andrew Larkin, Lizhong Chen (Oregon State), Steve Hankey (VA Tech), Wenwen Zhang (Rutgers), Esra Suel (ETH Zurich), Jaime Hart (Brigham and Women's Hospital), and Eric Rimm (Harvard TH Chan School of Public Health)

Disparities in Access to Green Space

- White neighborhoods have greater access to green space
- Exposure disparities persist over generations, possibly contributing to health disparities by race and socioeconomic status
- Greenspace may reduce health disparities



Casey et al. 2017

The Climate Crisis

Green space has massive co-benefits for human health and the health of natural systems

- Major carbon sink
  - Planting a half trillion trees could **reduce atmospheric carbon by about 25 percent**
  - **Enough to negate about 20 years of human-produced carbon emissions** at the current rate, or about half of all carbon emitted by humans since 1960
- Prevents soil erosion
- Provides substantial cooling in urban environments
- Counterpoint of wildfires

In the context of equity, low-income blocks have 15.2% less tree cover and are 1.5^{o}C hotter than high-income blocks

Conclusions

Exposomic analyses account for multiple correlated exposures

Mobile health approaches add granularity on personalized exposures and health behaviors

Deep learning algorithms provide specific, time-varying, objective data on environmental exposures from unprecedented perspectives

These data combined may provide relevant, actionable information for planners, policy makers, and community members to design places that are optimal for health

Acknowledgements

Thanks to Francine Laden, Jaime Hart, Marcia Pescador Jimenez, Esra Suel, Debby Pi-i Lin, Grete Wilt, Cindy Hu, Charlotte Roscoe, Perry Hystad, Steve Hankey, Wenwen Zhang, Li Yi, Eric Rimm, Brent Coull, Jorge Chavarro, and Emily Oken

Special thanks to Grete Wilt and Li Yi for some of the slides in this presentation

Funding: R00 CA201542, R01 HL150119, CRESSH P50 MD010428, U2 COD023375, U24 OD023382, U24 OD023319, U2 COD023375, R01 HD034568, UH3 OD023286, U01 HL145386

Questions: pjames@hsph.harvard.edu or visit **spaceelab.org**