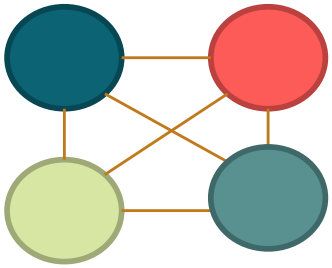
Data Science to Improve Population Health

Elaine Nsoesie, PhD
Assistant Professor of Global Health
Email: onelaine@bu.edu / Twitter: @ensoesie

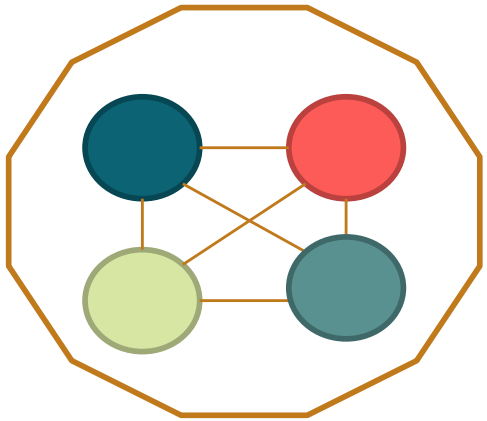
BOSTON UNIVERSITY

Network Individual-based Model

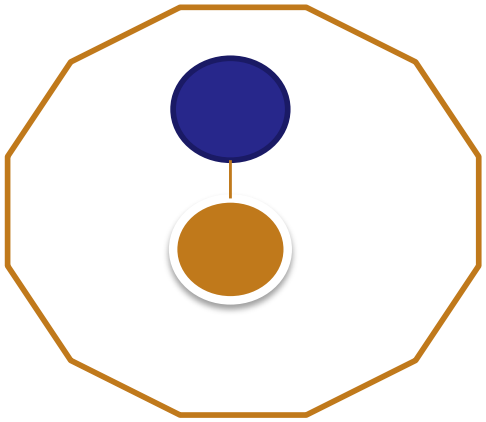


Household

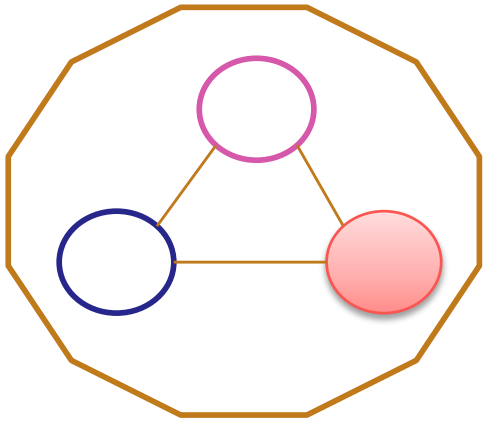
Network Individual-based Model



Household 1

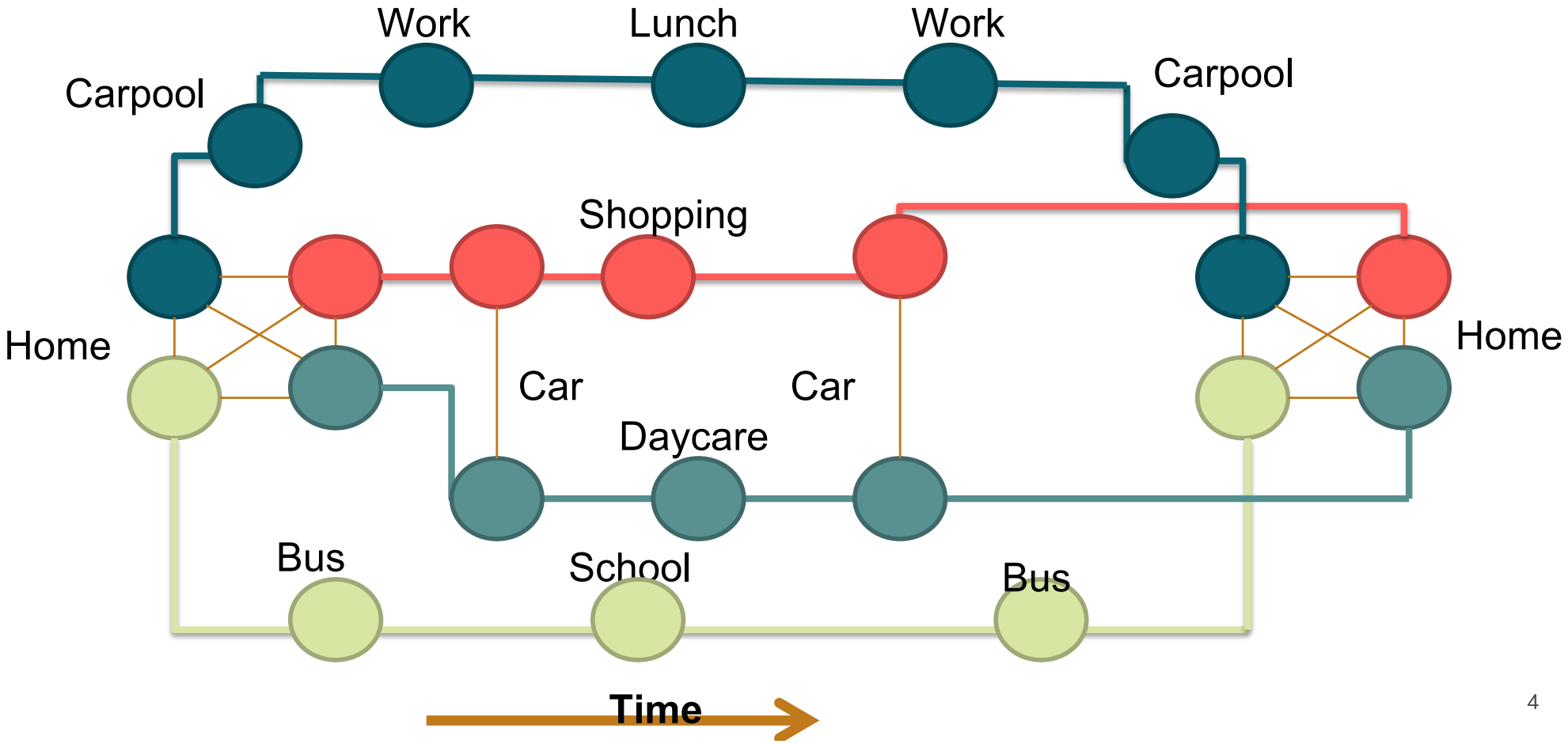


Household 2

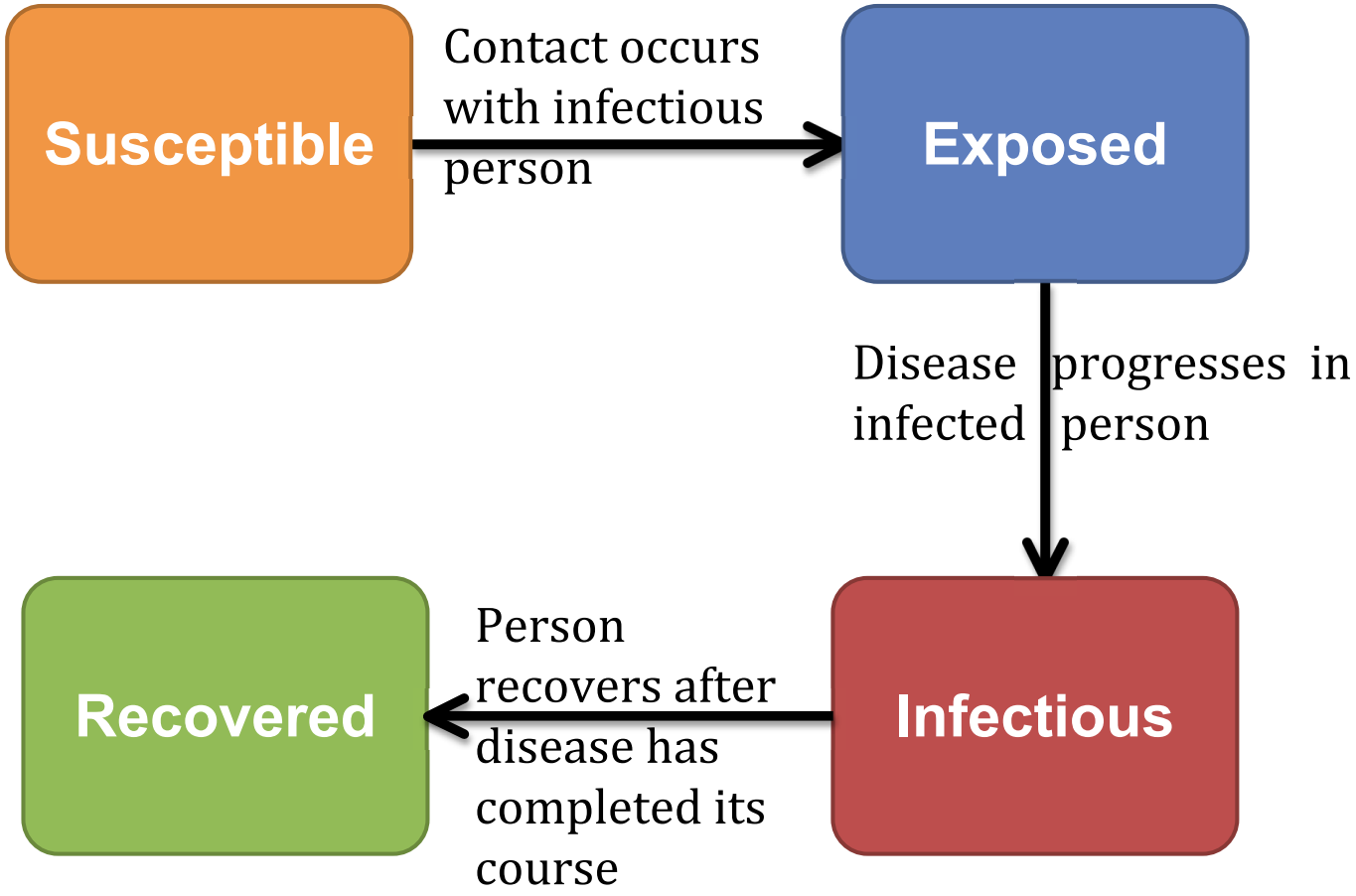


Household 3

Network Individual-based Model

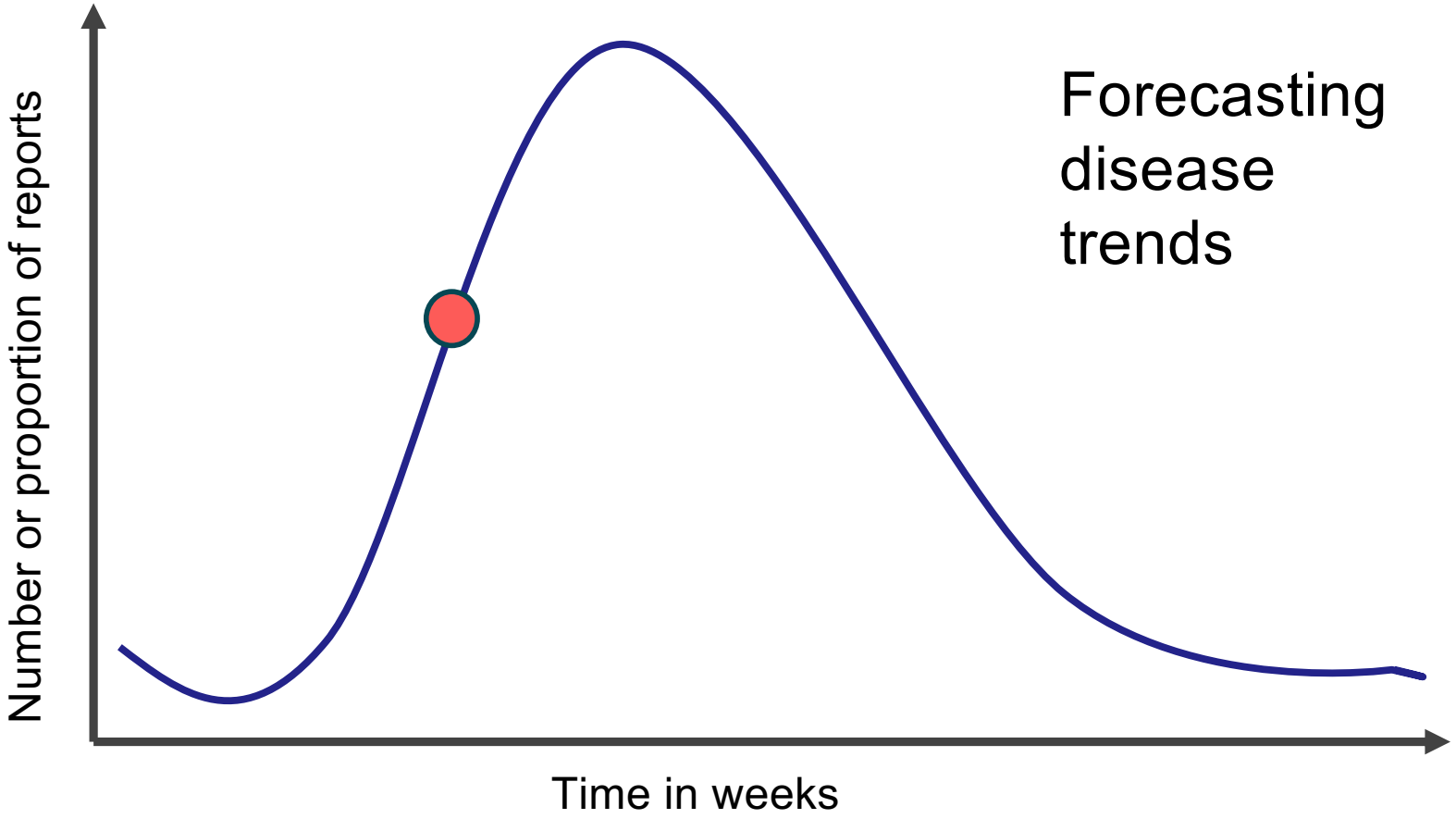


Disease model



Day: 1

Number of components: 0
Sizes: []


Number or proportion of reports
Time in weeks
Forecasting
disease
trends

Digital Epidemiology: new data streams

"Got my flu shot this morning and now my throat is sore."

"Stomach flu & normal flu in the same month. I'm officially a germaphobe."

"Such an upset stomach today. I hope it's just a bug and not the Truvada."

"My weight: 170.1 lb. 10.1 lb to go. #raceweight @Withings scale auto-tweets my weight once a week http://withings.com"

Text
Images
Videos
Sounds
Location
Biological data
etc.

Slide by Marcel Salathé

Studies have explored a variety of health questions

How can we monitor the spread of influenza using Internet search data?

Can Instagram photos provide insights into changes in mental health?

What do postings on social media tell us about the stigma associated with women's health issues?

How does misinformation about vaccines spread through social networks?

How can discussions on radio stations lead to early identification of a Cholera outbreak?

Is there seasonality in online searches for chicken pox and can we use that to predict outbreaks?

Can postings on discussion forums reveal new insights into adverse drug events?

Are there geographical and racial differences in hospital care experiences?

Do filtered images on Instagram distort the self-image of teenagers?

Can news reports mined from online sources help us detect Ebola outbreaks faster?

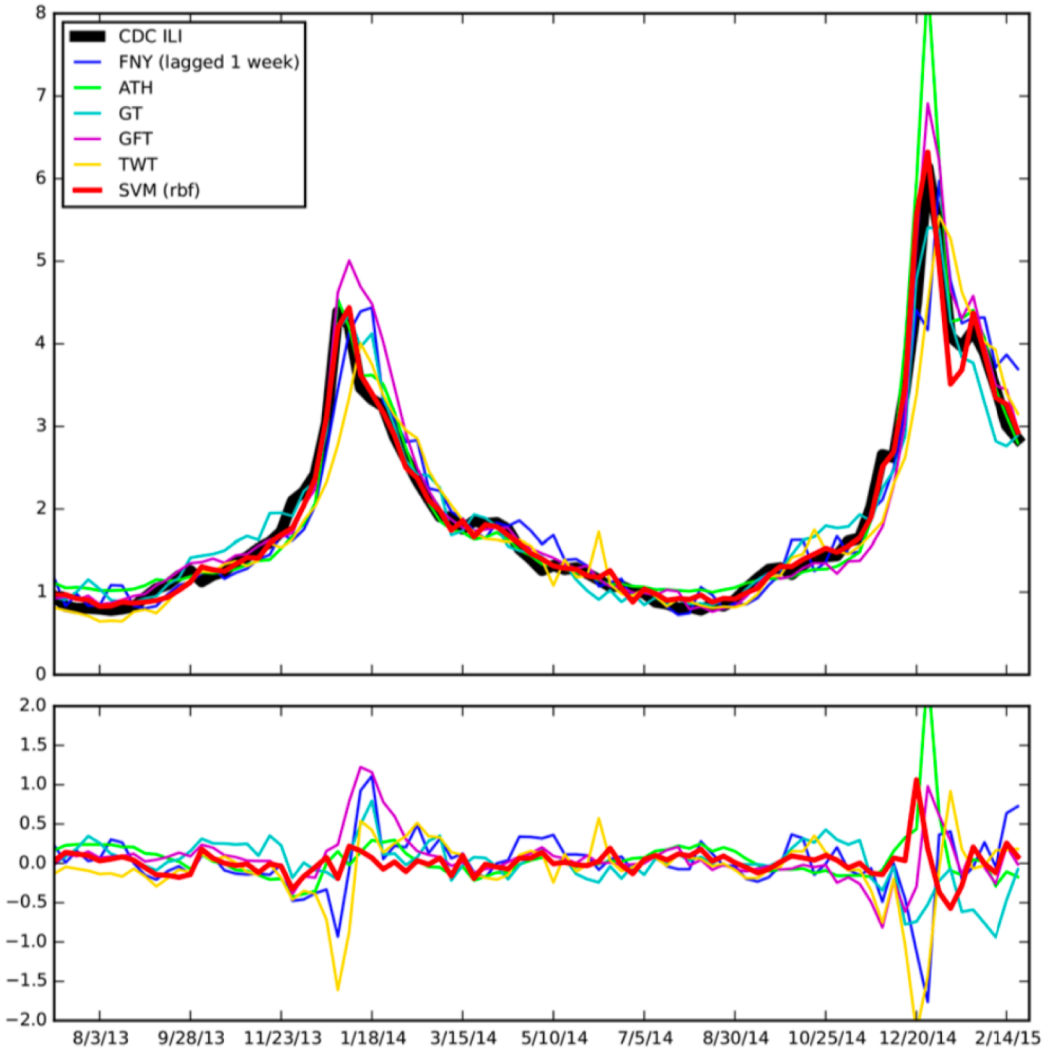


Fig 1. The CDC's %ILI (Influenza like illnesses), the performance of the 5 available predictors, the baseline predictions, and the performance of the best ensemble method for last week's predictions are displayed as a function of time (top). The errors associated with each weak predictor and the ensemble approach are shown (bottom).

CDC – Centers for Disease Control and Prevention
ILI – Influenza-like illness
FNY – FluNearYou
ATH – Athena Health
GT – Google Trends
GFT – Google Flu Trends
TWT – Twitter


Bai du 百度


yelp


G


OpenTable®

iwaspoisoned.com
SAFER DINING BY
CROWDSOURCING


DigitalGlobe


Google


HealthMap
Global Health, Local Information


flu
near you
do you have it in you?


amazon


Google Earth


UpToDate®


International Society
for Infectious Diseases


ProMED
mail

Research and Implementation

How can we use data from the Internet and digital devices to improve individual and population health?

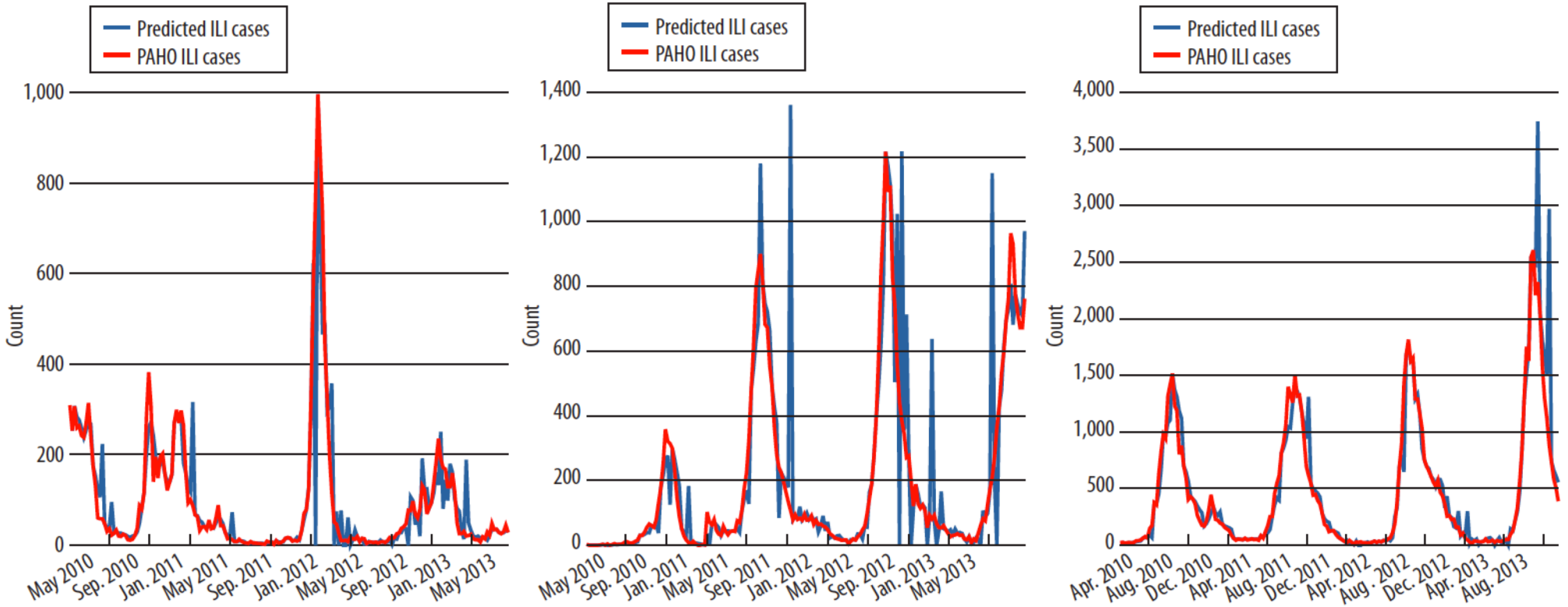
Process data and develop tools for health departments, global health organizations, and foundations for decision making.

Satellite Images

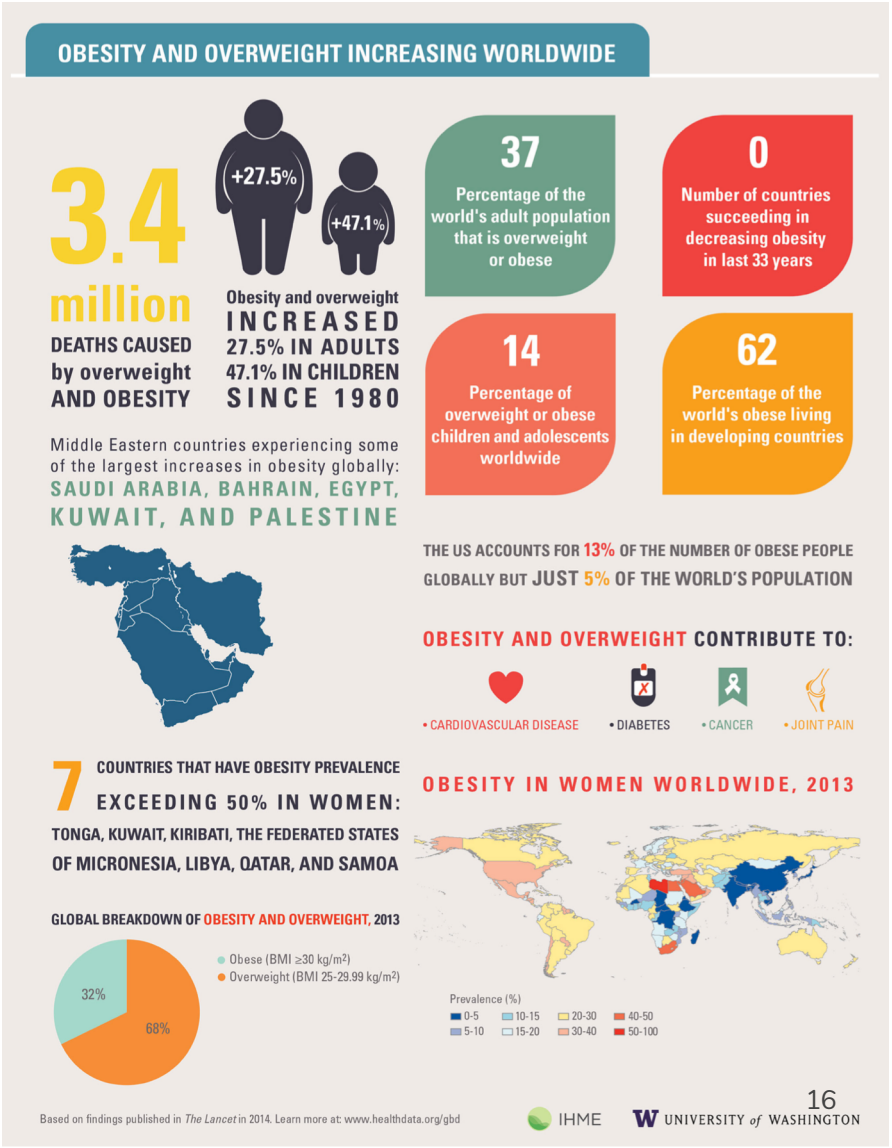
Monitoring Social Disruption in South America with Satellite Parking Data

Butler et al. IEEE. 2014

Monitoring Social Disruption in South America with Satellite Parking Data

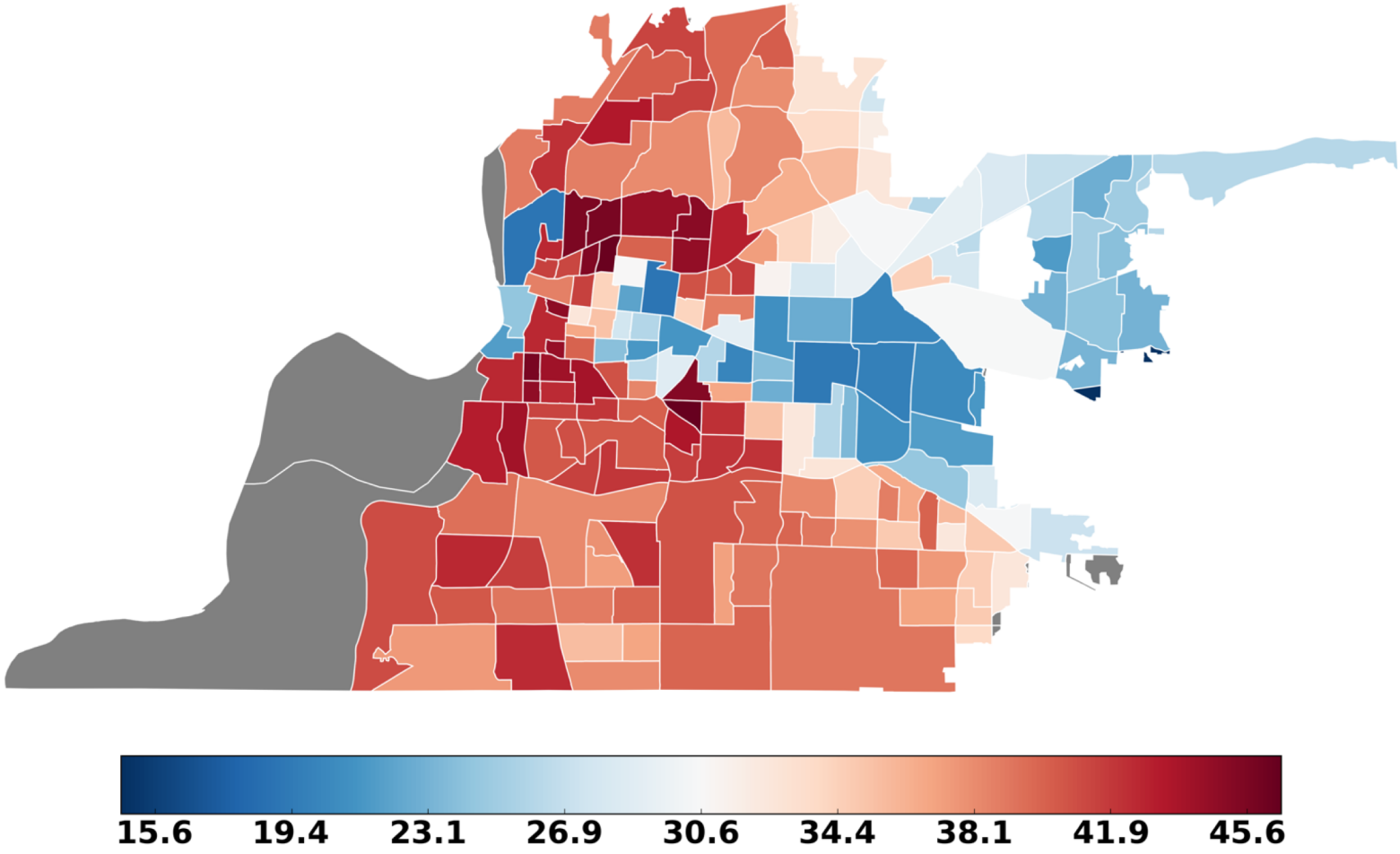


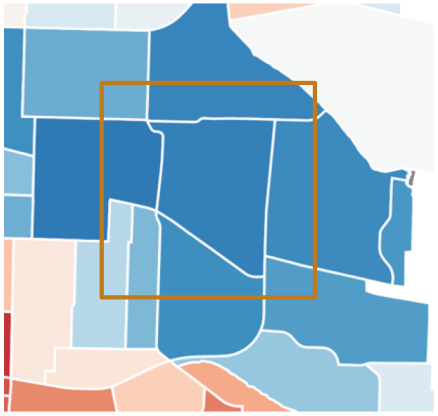
- Obesity is a complex health issue
- Multiple factors have been linked to obesity
- In the US, obesity affects about one-third of the adult population



Many environmental factors have been linked to obesity

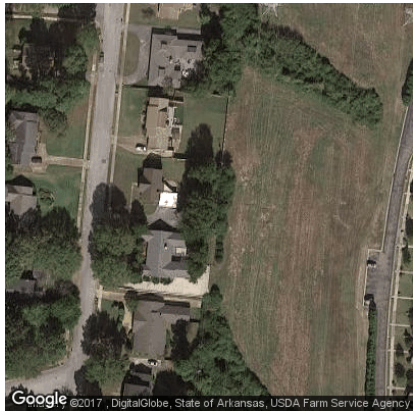
But there are differences in measures and measurements making it difficult to compare findings across cities

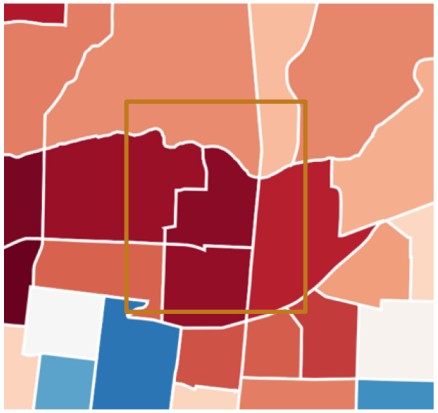

15.6
19.4
23.1
26.9
30.6
34.4
38.1
41.9
45.6




Google ©2017 , DigitalGlobe, State of Arkansas, USDA Farm Service Agency


Google ©2017 , DigitalGlobe, State of Arkansas, USDA Farm Service Agency


Google ©2017 , DigitalGlobe, State of Arkansas, USDA Farm Service Agency

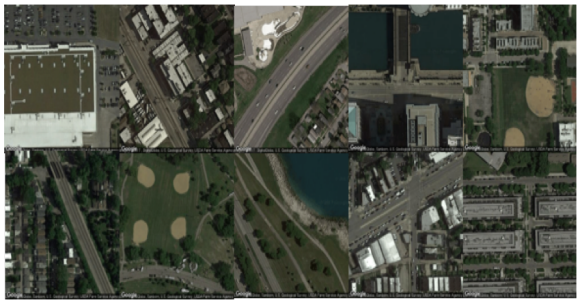


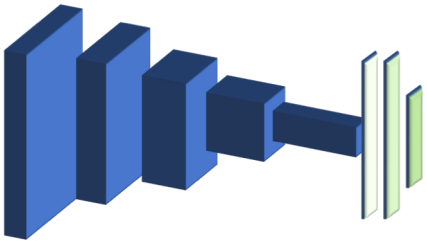

Google ©2017 , DigitalGlobe, State of Arkansas, USDA Farm Service Agency


Google ©2017 , DigitalGlobe, State of Arkansas, USDA Farm Service Agency


Google ©2017 , DigitalGlobe, State of Arkansas, USDA Farm Service Agency

Transfer Learning with VGG-CNN-F



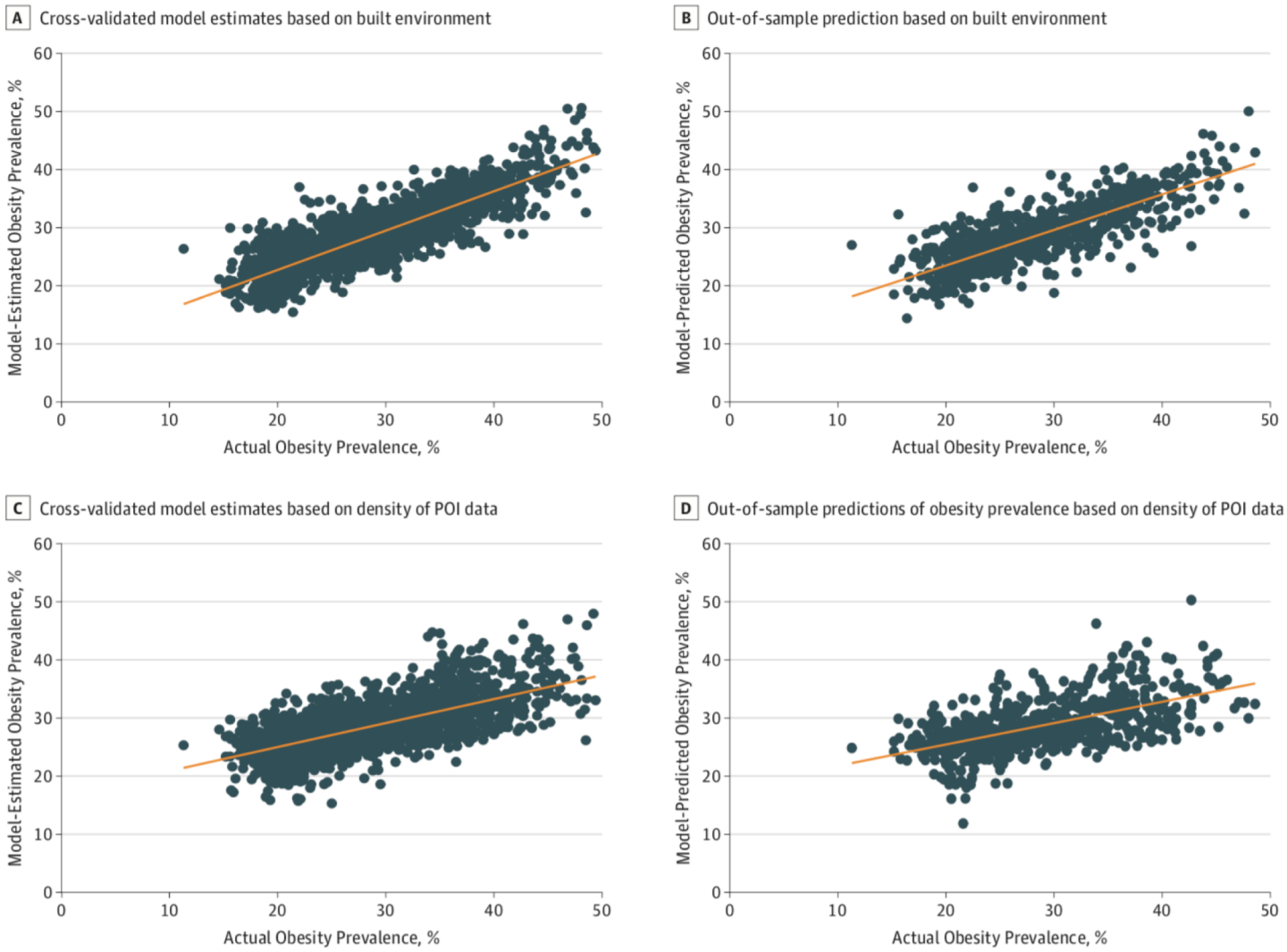


Model trained for Domain A

Feature Maps for Domain B

Predictive Model

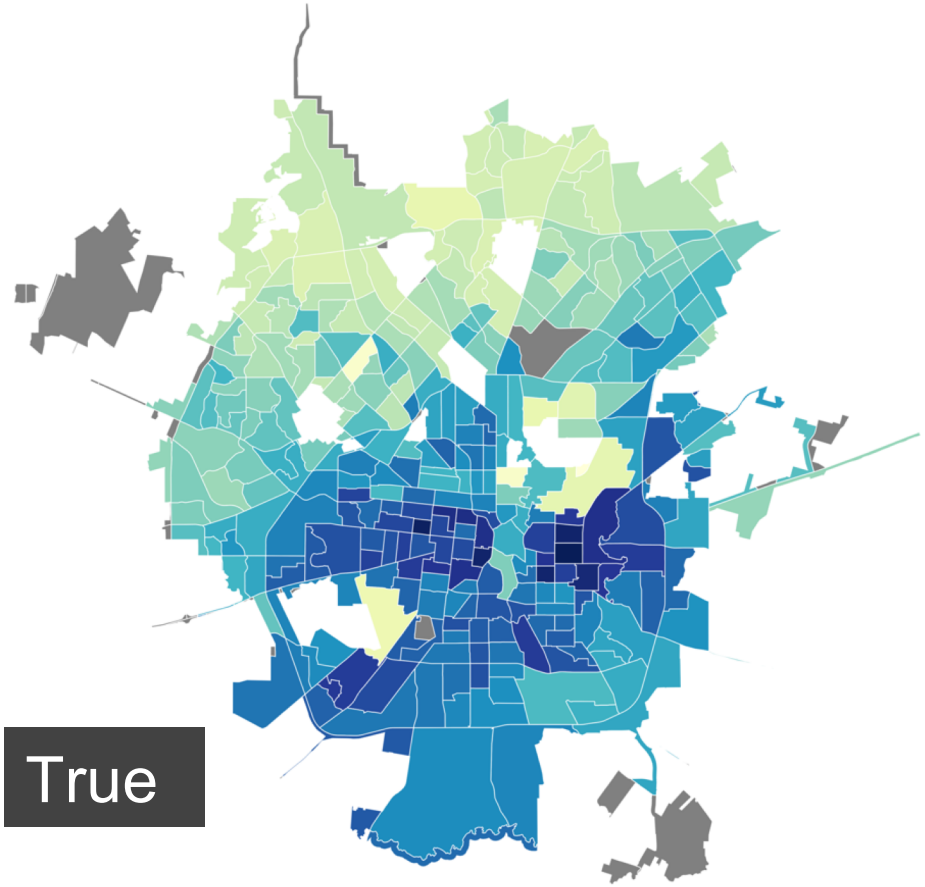
Transfer of Knowledge



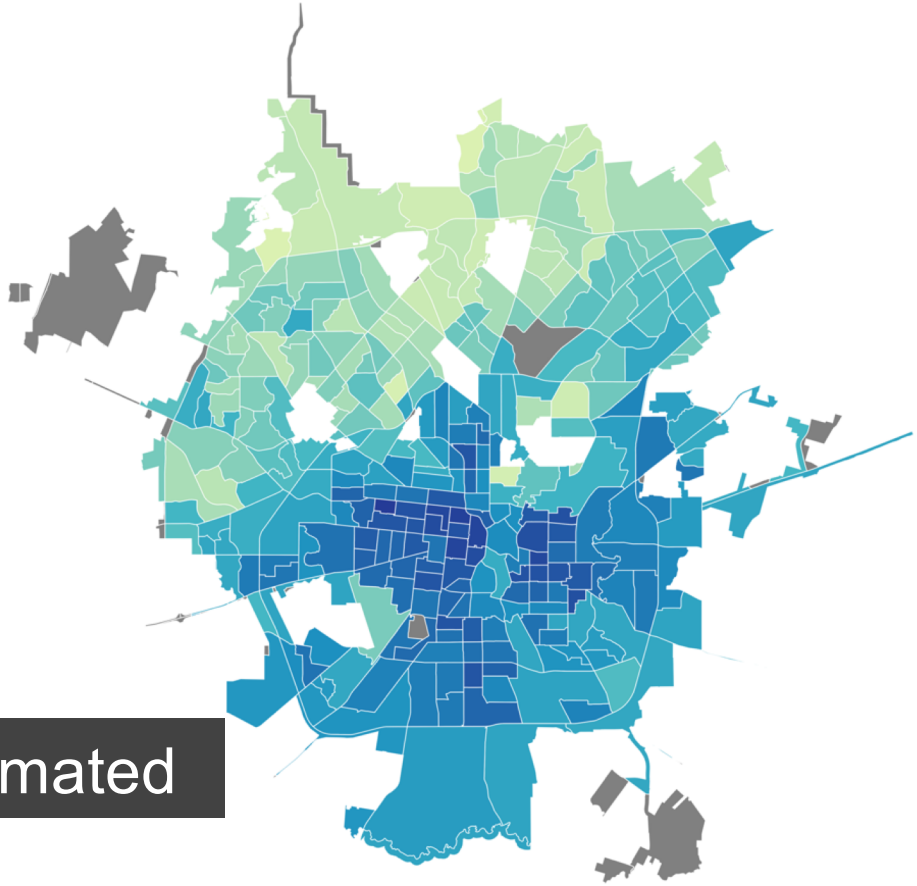
Built environment information is extracted from satellite images. POI indicates point of interest.

San Antonio, Texas

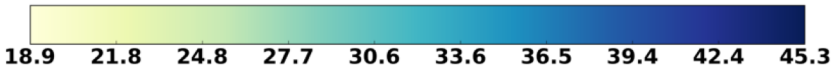
n =311

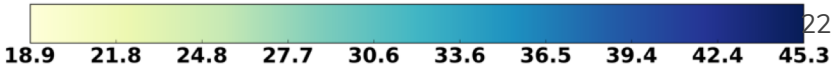


True



Estimated

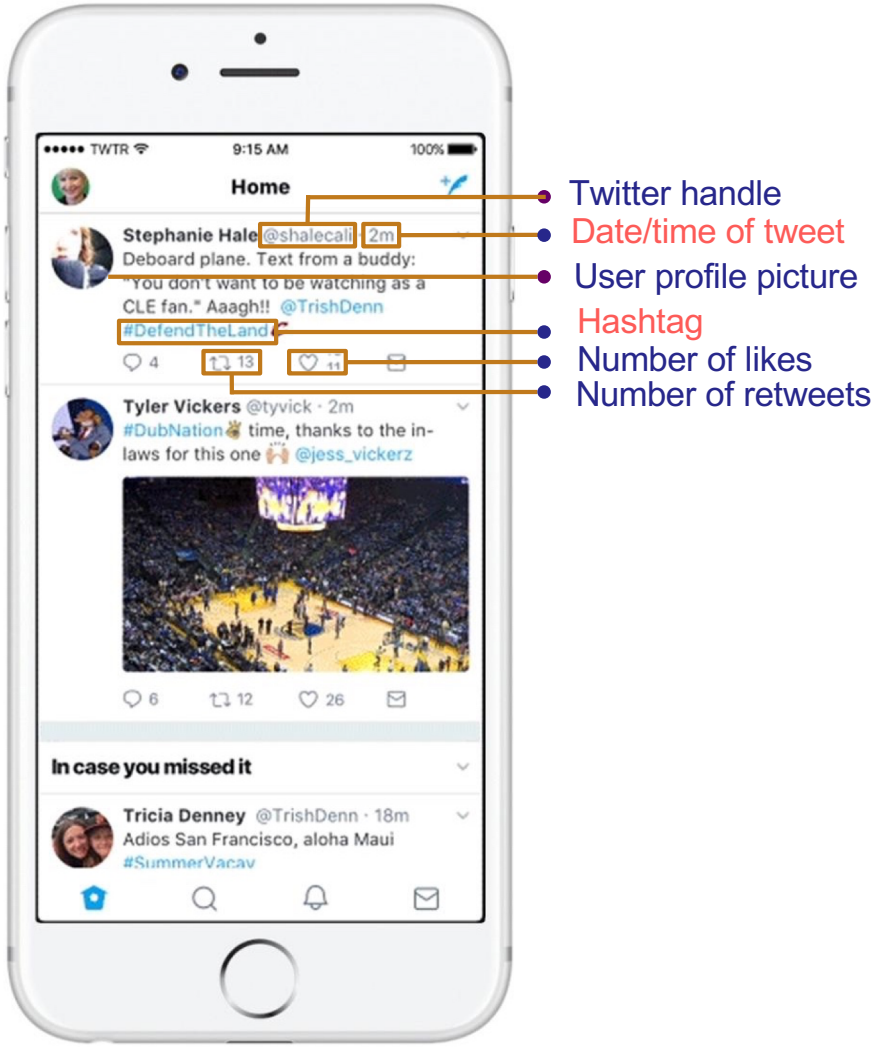




Social Media

Social media provides a window into the **built environment, interactions, access, likes/dislikes, lifestyle, behaviors, choices**

EXPLICIT DATA ABOUT USER	EXPLICIT DATA ABOUT TWEET	EXTRACTABLE DATA ABOUT USER	EXTRACTABLE DATA ABOUT TWEET
Twitter handle	Time/date of tweet	Disease state	Content
Language	Hashtag	Gender	Sentiment
Time zone	URL	Marital status	Image Analysis
Location	Geotag	Political party	Language
Date account created	App used to send tweet	Race/ethnicity	
User profile	Number of retweets	Occupation	
User profile photo	Number of favorites	Interests	
Number of followers	Total number of tweets	Religion	
Number following	Linked images	Income	



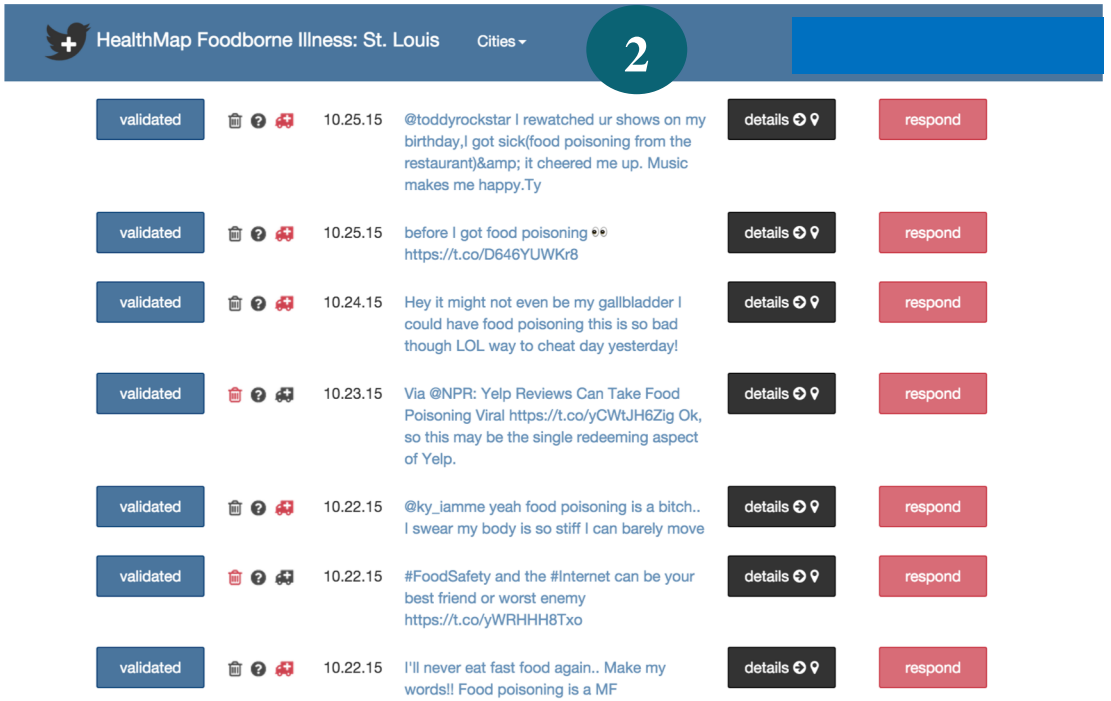
It’s complicated.
Social media data raises **important ethical and regulatory questions.**

The Impact Health of Misinformation on Public Health

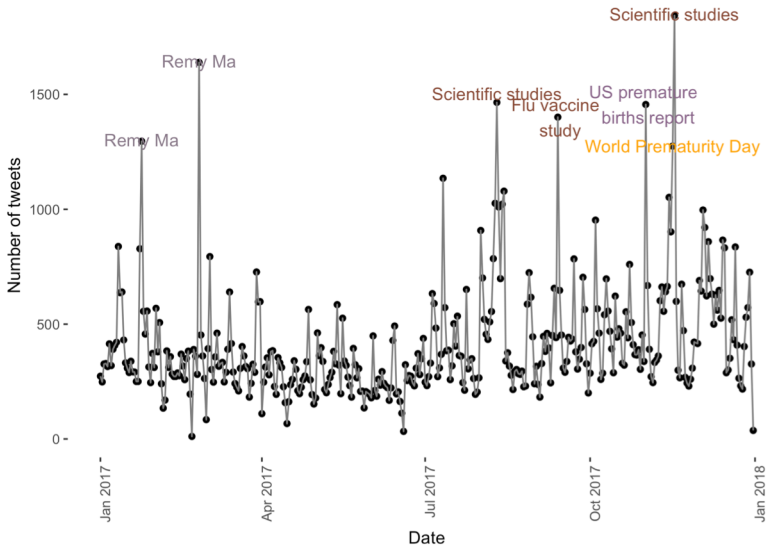
- Fake audio warning against the dangers of the yellow fever vaccine contributed to vaccine hesitancy in Brazil
- Misinformation about the spread of 2014-2016 Ebola epidemic in West Africa
- Increased incidence of measles outbreaks in the US and Europe linked to misinformation on vaccines

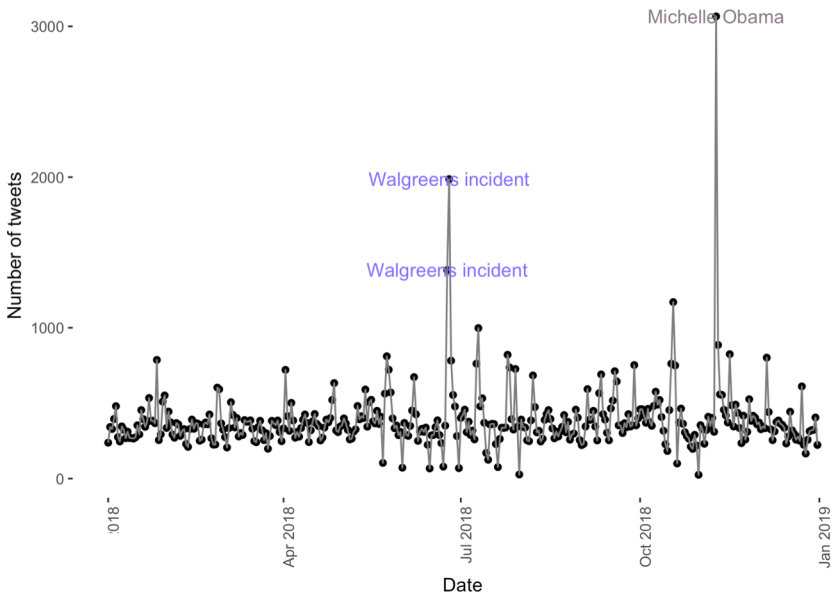
https://www.wired.com/story/when-whatsapps-fake-news-problem-threatens-public-health/

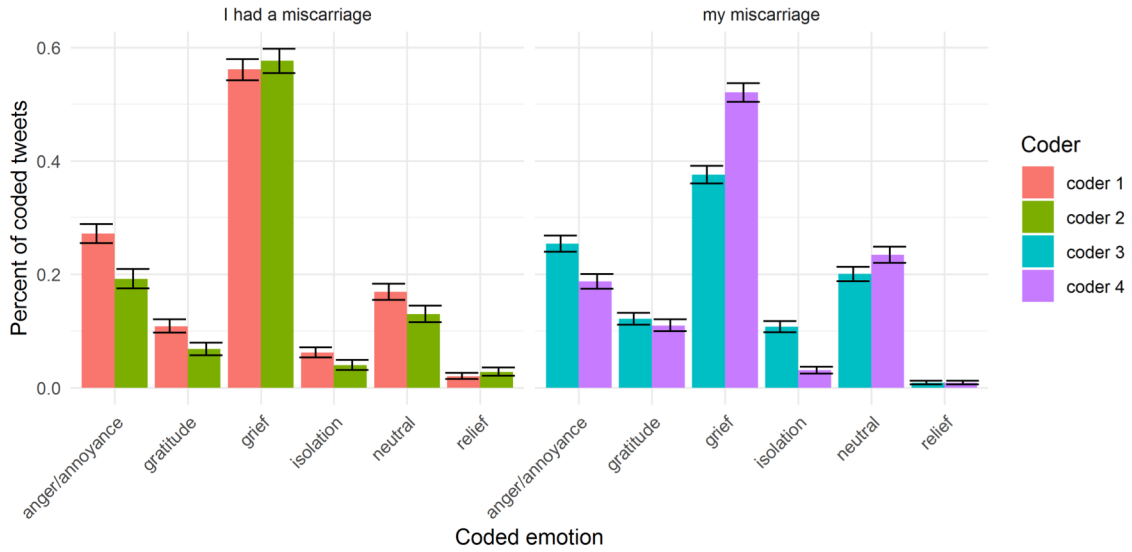
Using Twitter Data to Monitor Reports of Foodborne Illness in US Cities



Targeted restaurant inspections and surveillance







How do people talk about women's health and infertility on social media? And why does it matter?

Using digital data for health applications in Africa requires a rethinking of who, what and how.

Health in African Cities

Dar es Salaam
Travel Advisor

About 55% of the world's population lives in urban areas. This number is expected to increase to 68% by 2050.

United Nations

The field of urban health started around the turn of the 21st century is focused on the study of how and why cities influence health.

The foods we have access to

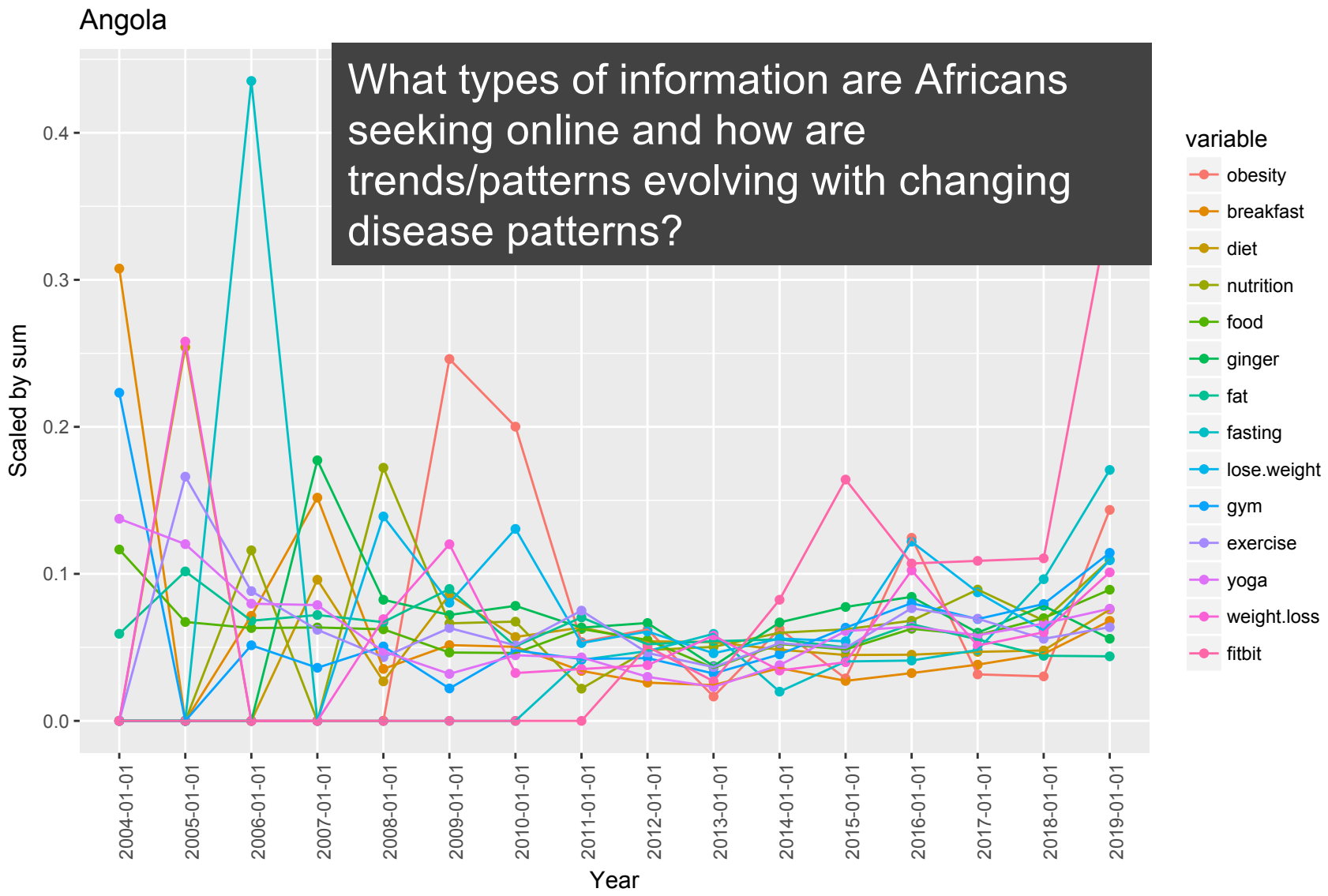
The foods we eat

Images from unsplash.com

All neighborhoods are not created equal

Drivers of health in cities

The physical environment
The social environment
Access to health and social services


Angola
What types of information are Africans seeking online and how are trends/patterns evolving with changing disease patterns?
Scaled by sum
0.0
0.1
0.2
0.3
0.4
2004-01-01
2005-01-01
2006-01-01
2007-01-01
2008-01-01
2009-01-01
2010-01-01
2011-01-01
2012-01-01
2013-01-01
2014-01-01
2015-01-01
2016-01-01
2017-01-01
2018-01-01
2019-01-01
Year
variable
obesity
breakfast
diet
nutrition
food
ginger
fat
fasting
lose.weight
gym
exercise
yoga
weight.loss
fitbit

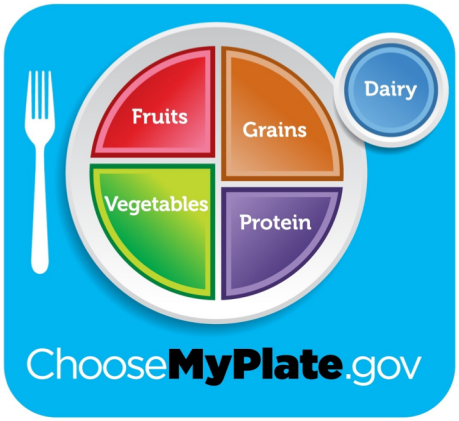
Food & Beverage on Instagram

Nutritional Composition

- 34% bread, cereal, rice, pasta, potatoes
- 26% fruits, vegetables
- 22% desserts, candy, cakes, ice cream
- 20% dairy, milk, cheese
- 18% chicken, eggs, poultry
- 13% lamb, beef, pork
- 6% fish, seafood, shellfish
- 4% snacks, chips, pretzels
- 6% soda
- 4% juice
- 12% alcohol

Nutritional Composition

- 21% homemade/fresh
- 23% heavily processed
- 24% fried, heavy oil content, butter
- 23% heavy salt content



Slide by Raina Merchant, MD

Call for collaborators

Contributors

Nina Cesare, Boston University
Pallavi Dwivedi, University of Maryland
Quynh Nguyen, University of Maryland
Adyasha Maharana, SCIOME
Bryan Lewis, University of Virginia
Boston University Spark!
John Brownstein & Jared Hawkins, Harvard Medical School
Olubusola Oladeji, Boston University
Kadija Ferryman, Data & Society
Raina Merchant, University of Pennsylvania

Thank you!

Contact
Email: onelaine@bu.edu
Twitter: @ensoesie