

Integrating big geospatial data into studies of health disparities

Marta Jankowska, PhD

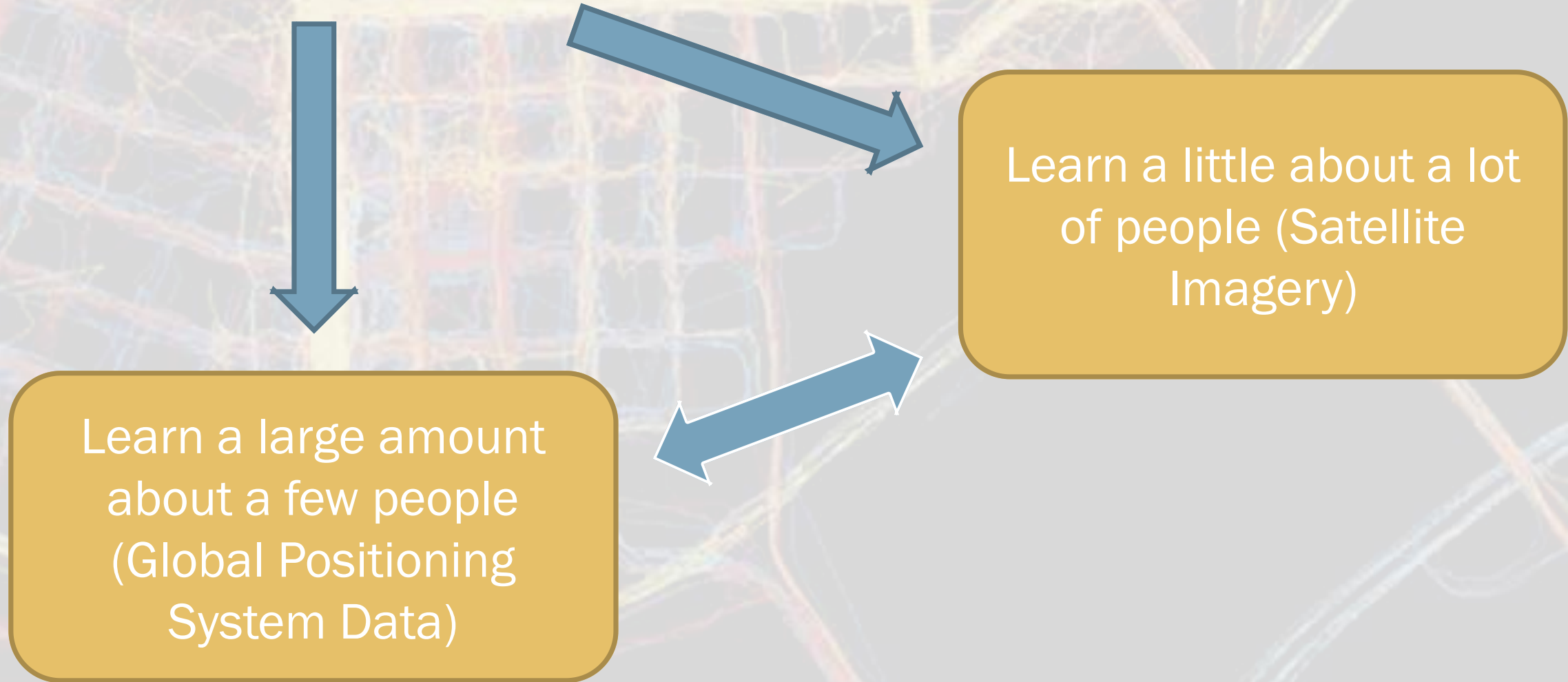
UCSD, Calit2/Qualcomm Institute

Health Data at Scale Collaboratory

[@hdscalecollab](https://twitter.com/hdscalecollab)

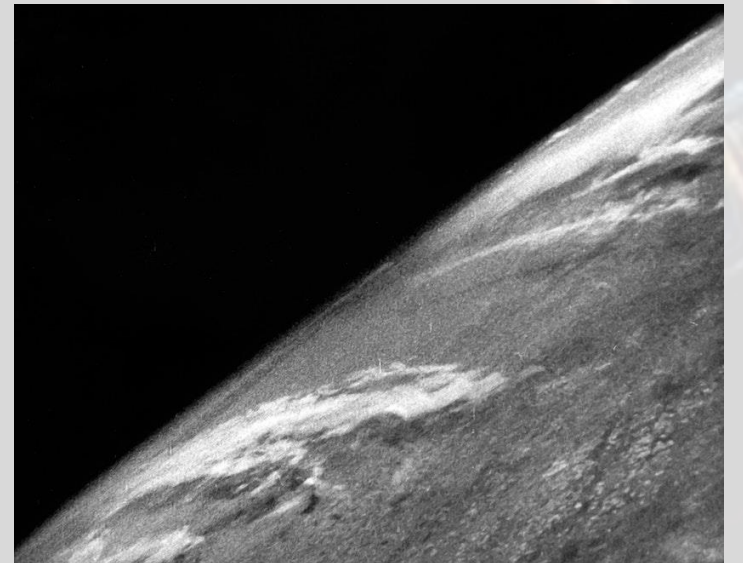
Disclosures: None

Cancer risk from environmental exposures



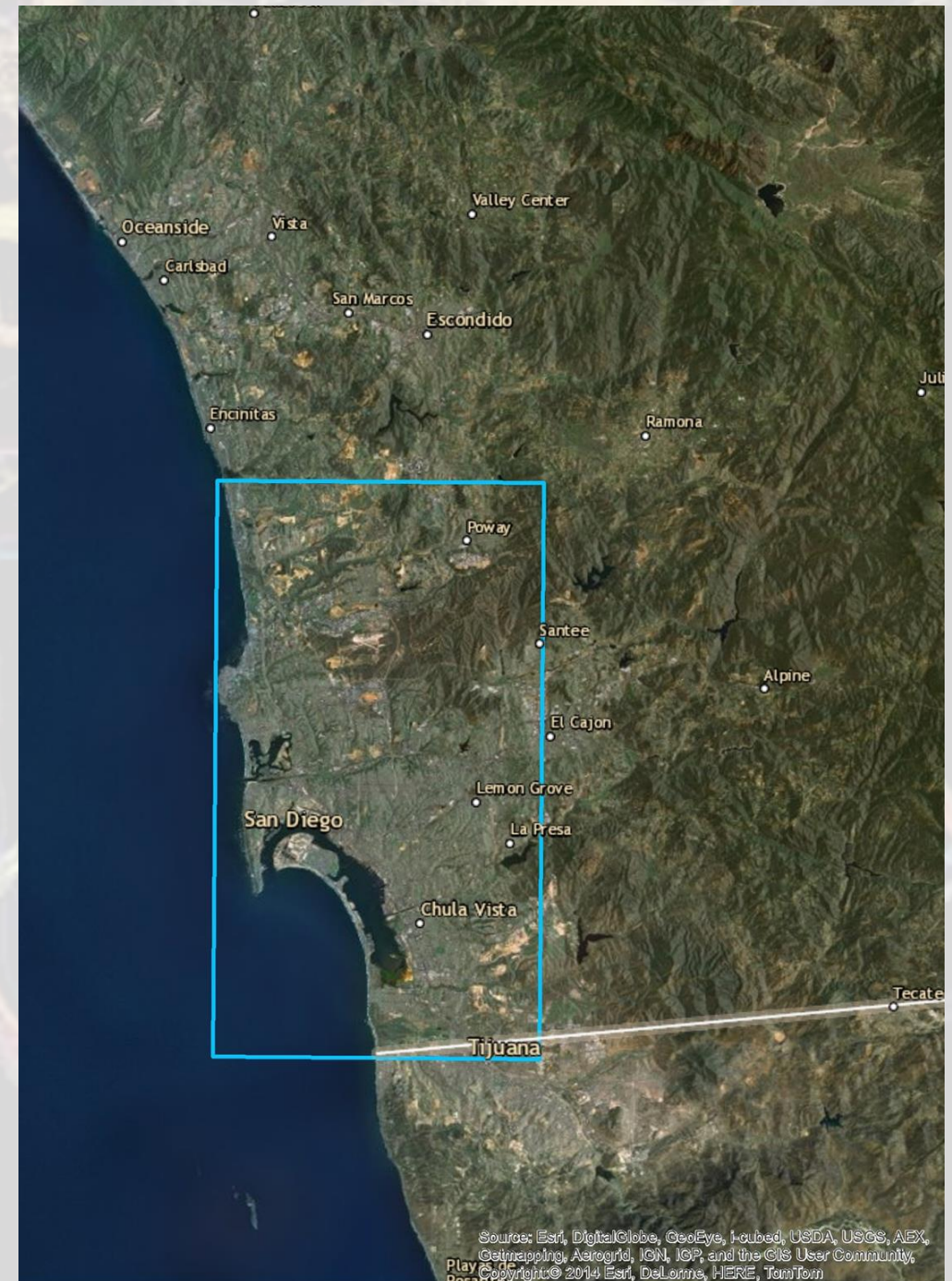
Learning a little about a lot of people with satellite imagery

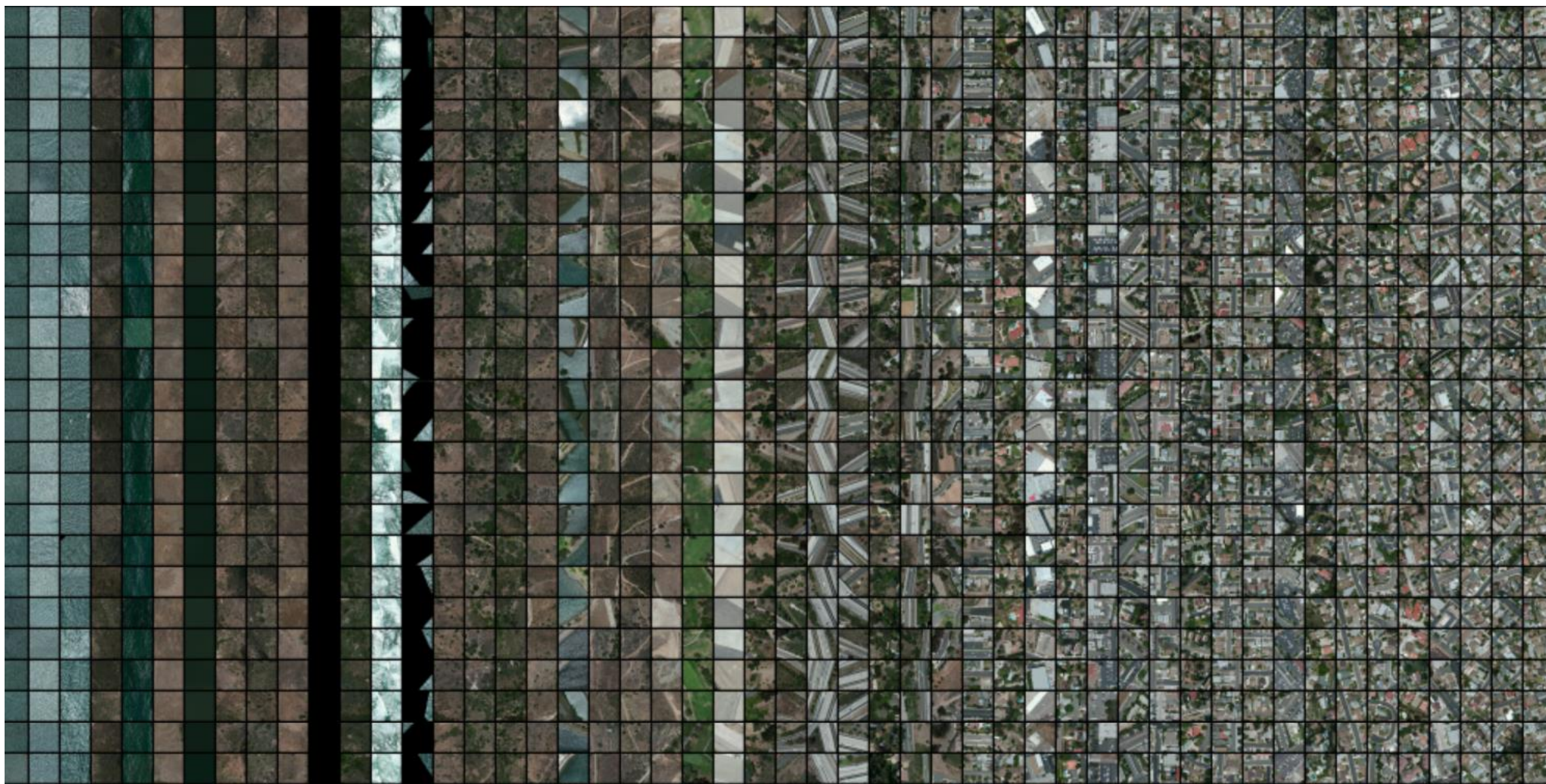
- Satellite imagery has been around since 1940
- Revolution in how to process this type of data
- Opportunity for health disparities and cancer:
 - *Ubiquitous*
 - *Repetitive in time*
 - *Flexible in spatial scale*
 - *Often free or low cost*
 - *Reliability*
- What can we learn?



Imagery - San Diego

- San Diego Dataset comes from Digital Globe's Global Basemap.
- Imagery is a composite from 2012 from several satellite passes
- 0.5 meter imagery, RGB, ~418,000 tiles, 200x200 pixels
- Applied Convolutional Neural Net algorithm and then k-means clustering on features

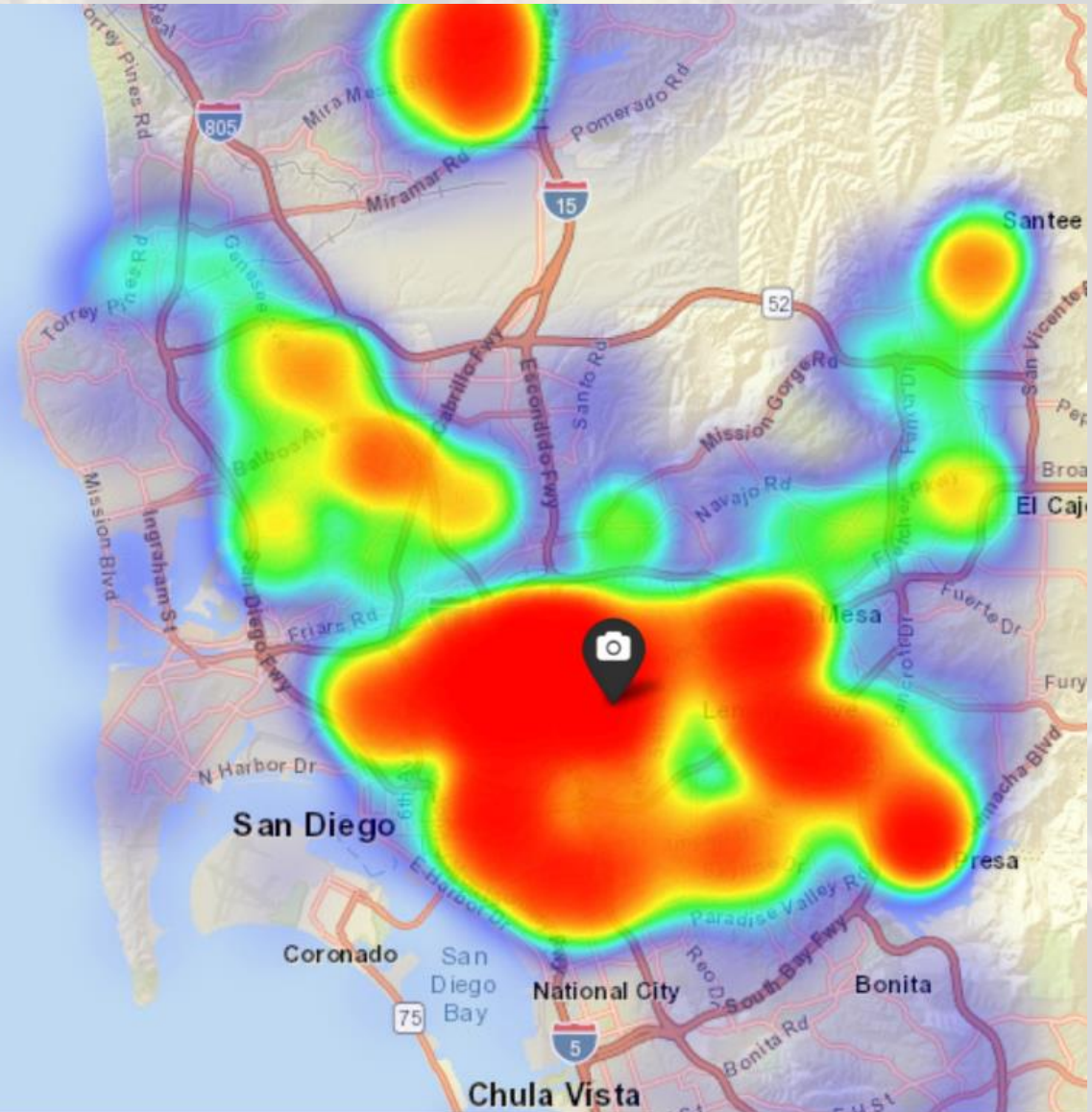




10

An aerial photograph of a residential neighborhood. The image shows several houses with dark roofs, some with light-colored siding. There are trees and greenery scattered throughout the area. A street runs along the right side of the image, with a few cars visible. The overall scene is a typical suburban residential area.

Age Group	Number of People
0-4	75
5-9	90
10-14	65
15-19	40
20-24	70
25-29	60
30-34	65
35-39	85
40-44	50
45-49	55
50-54	85
55-59	60
60-64	75
65-69	35
70-74	100
75+	55
80+	80
85+	45



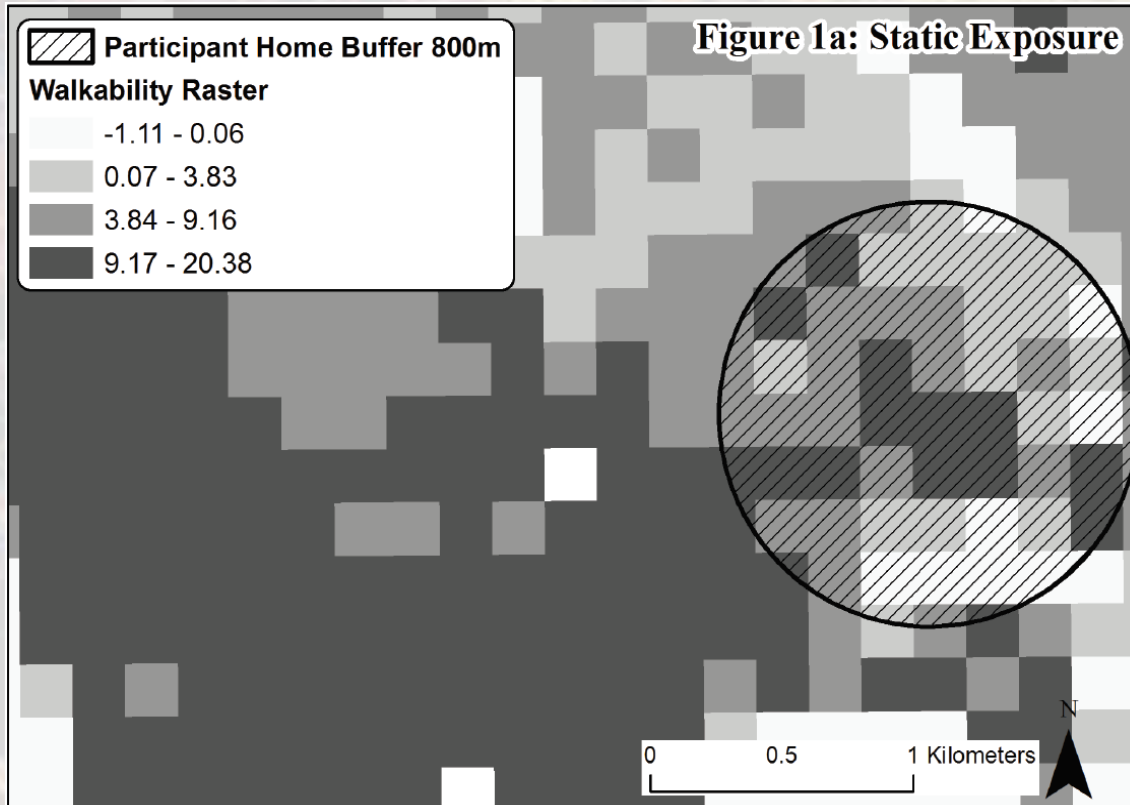
Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong K



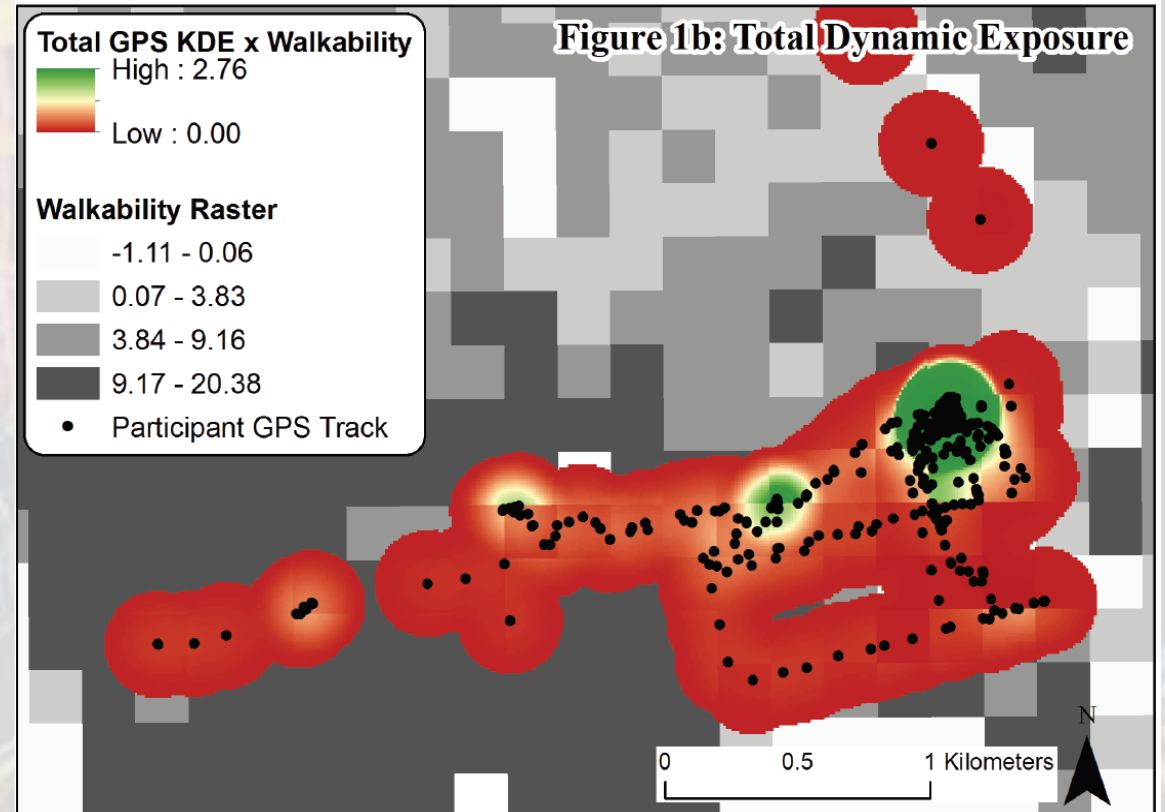
Learning a lot about a few people with Global Positioning System data

- Exposures not just where we live but where we go
- Linking with other sensors to understand behaviors or health outcomes with exposures in real time
- Can start to unravel assumptions about cancer-risk related behaviors and where they occur

Static Assessment of Environment Exposure



Dynamic Assessment of Environment Exposure

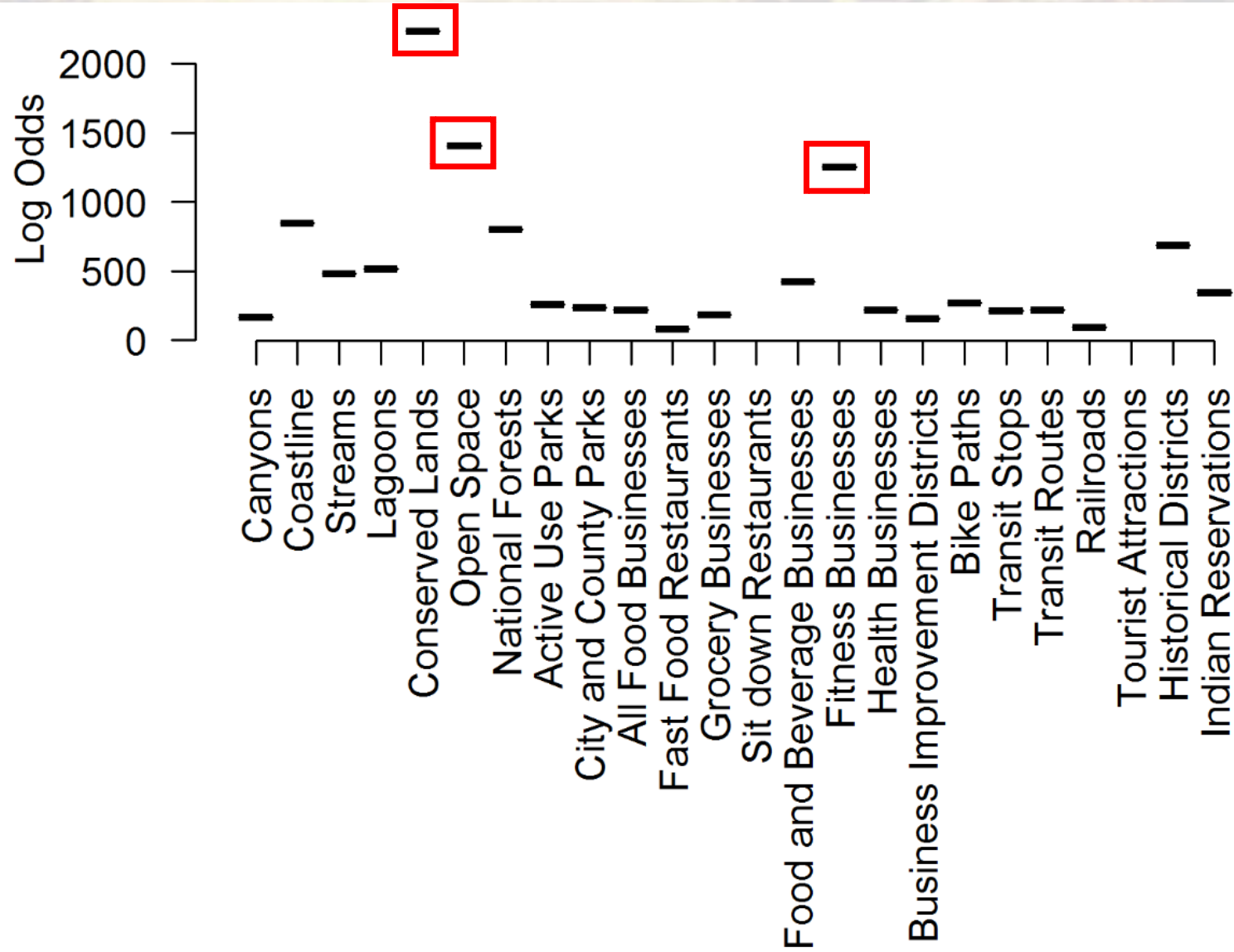


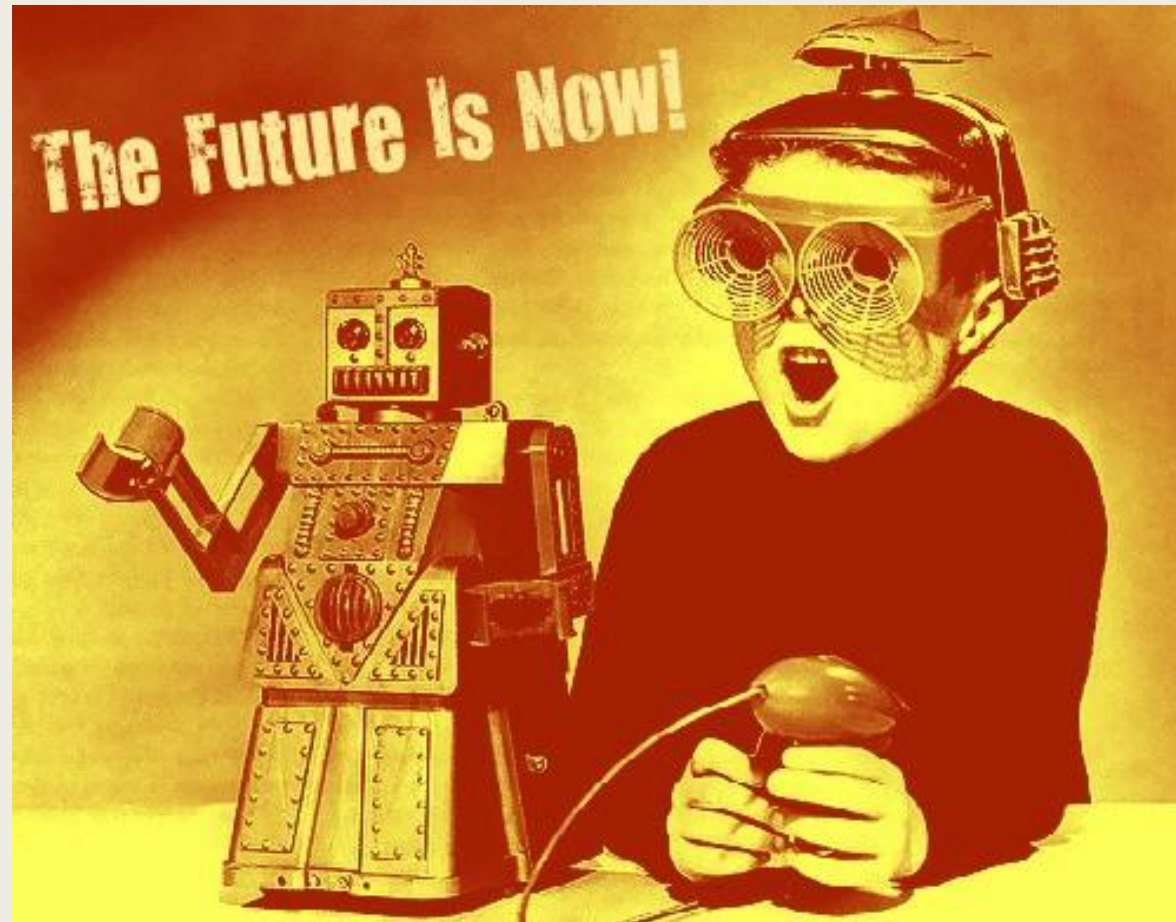
Jankowska, M. M., Natarajan, L., Godbole, S., Meseck, K., Sears, D. D., Patterson, R. E., & Kerr, J. (2017). Kernel density estimation as a measure of environmental exposure related to insulin resistance in breast Cancer survivors. *Cancer Epidemiology and Prevention Biomarkers*, 26(7), 1078-1084.

GIS Variables	Name			
Population Density	Adult Day Care	CMS Clinics + Council of Community Clinics + Family Resource Centers	Sit-down Food, Retail Slope	WIC retailers
	Agricultural Commodity	Indian Reservations	Census Block Groups	Bankruptcy filings (non-business)
	Agricultural Preserve	Lakes	Residential Density	Tax-related info: new returns, credits, dependent, etc.
Land-use Mix	Agricultural Preserve Contracts	National Forests	Walkability	Number of unemployed workers and unemployment rate
	Business Improvement Districts	Canyons	Census Tracts	Average annual wage by industry
	Business Regional Enterprise Zones	Conserved Lands	Zip Codes	home values, housing stock, rental units
Rec Facility Access	Business type: Community	All Parks	Total Population By Age	Count and percent of jobs located in a certain place
	Business type: Fitness	Railroad	Hispanic Population	total fruits and vegetables consumed per day
	Business type: Retail	Transit Routes	Household Income	count and percent obese, diabetic, physically inactive
Street Pattern	Business type: Health	Transit Stops	U.S. Tapestry	Prenatal care received, infants born with low birth weight
	Business type: Food Grocery	Bike Routes	Housing prices	Level of Influenza-like flu activity
	Business type: Food Beverage	Bike Master Plan	Crime Incidents from 2007-2011	Level of Influenza-like flu activity
Sidewalk Coverage	Business type: Food	All Roads, categorized	County Crime	Level of Influenza-like flu activity
	Business type: Community, Fitness, Health, All Food	County Roads, Curb, Sidewalks	County Alcohol Permits	Level of Influenza-like flu activity
	Community Villages: Rural villages	State Licensed Child Care	County Crime	Level of Influenza-like flu activity
Vehicular traffic	Community Villages: village	State Licensed Group Homes	County Alcohol Permits	Level of Influenza-like flu activity
	Cemetery	Streams	Street Lamp Locations	Level of Influenza-like flu activity
	Coastline	Lagoons	Deaths	Level of Influenza-like flu activity
Public Transit	Neighborhood delineations	Places/POIs	California Academic Performance Tests	Level of Influenza-like flu activity
	General Plan Housing Element Inventory	Parks Active Use	Species Occurrence	Level of Influenza-like flu activity
	County Fire Stations	Land Use	Critical Habitat	Level of Influenza-like flu activity
Slope	County General Plan Trails	Redevelopment Infill	Alcohol Retailers	Level of Influenza-like flu activity
	Congressional (US) Districts	Zoning Base	Community Gardens	Level of Influenza-like flu activity
	Council Districts	Non-residential footprints	Farm Markets/Stands	Level of Influenza-like flu activity
Greenness/vegetation	County Health and Human Services	Open Space	Food Deserts	Level of Influenza-like flu activity
	Historic Districts	Casinos	Tobacco retailers	Level of Influenza-like flu activity
	School Districts Elementary	Colleges	Retail food environment	Level of Influenza-like flu activity
Building design	School Districts Unified	Government Centers	Bureau of labor statistics	Level of Influenza-like flu activity
	Hospital Points	Law Enforcement Facilities	Fitgram (physical fitness tests)	Level of Influenza-like flu activity
		Library	Health professional Shortage Areas	Level of Influenza-like flu activity
Residential Density		County and City Rec Centers	Pharmacies	Level of Influenza-like flu activity
		Schools, area	WIC clinics	Level of Influenza-like flu activity
		Tourist Attractions	SWITRS (motor vehicle collisions)	Level of Influenza-like flu activity
		Subdivision	Earthquake faults	Level of Influenza-like flu activity
		Fast Food Food Retail	Tsunami areas	Level of Influenza-like flu activity
		Taqueria Food Retail	Climate Zones	Level of Influenza-like flu activity
			EBT retailers	Level of Influenza-like flu activity

	Layer Name	760+	Layer Name	760+
1	Canyons	602	Streams All	5627
2	National Forests	472	TRANSIT STOPS	867
3	Open space	203	TRANSIT ROUTE	15707
4	Ag preserve contracted	278	State Licensed Group Homes	3
5	Agriculture commodity	75	Library	36
6	Business improve district	4803	Railroads	973
7	Parks active use	3220	Tourist Attraction	171
8	Parks	3289	Conserved lands	2586
9	Coast	3022	Lagoons	292
10	Fitness Business	440	Bike	14829
11	Fast Food	196	Law Enforcement Facilities	21
12	Food Beverage	186	State licensed child care	72
13	Food Grocery	154	Historical districts	3619
14	Food Business	1582	Indian reservations	243
15	Sit Down	934	Health Business	2155

People who get at least 30 minutes of activity per day (Non-Hispanic vs. Hispanic/Latino people)





MAJANKOWSKA@UCSD.EDU

@hdscalecollab