NASA Blue Marble 2.0

Spatial Data Science: Mapping for Impact

Maggi Kelly

Professor & CE Specialist

UC Division of Agriculture & Natural Resources

University of California at Berkeley

GEOSPATIAL INNOVATION FACILITY

Berkeley UNIVERSITY OF CALIFORNIA | College of Natural Resources

INTERNATIONAL AND EXECUTIVE PROGRAMS

IGIS | University of California Agriculture and Natural Resources | Informatics and GIS Statewide Program

Our 21st Century Mapping Toolkit

Many of the challenges we face today around food, water, equity, energy, invasive species, fire, climate change, conservation – are ***complex***, require a ***spatial approach*** and impact ***diverse public groups***.

Addressing these challenges requires innovative & resourceful ***data collection***, ***data synthesis, novel analytical tools***, and ***increased communication and cooperation*** between scientists and citizens.

IGIS

University *of* California
Agriculture and Natural Resources

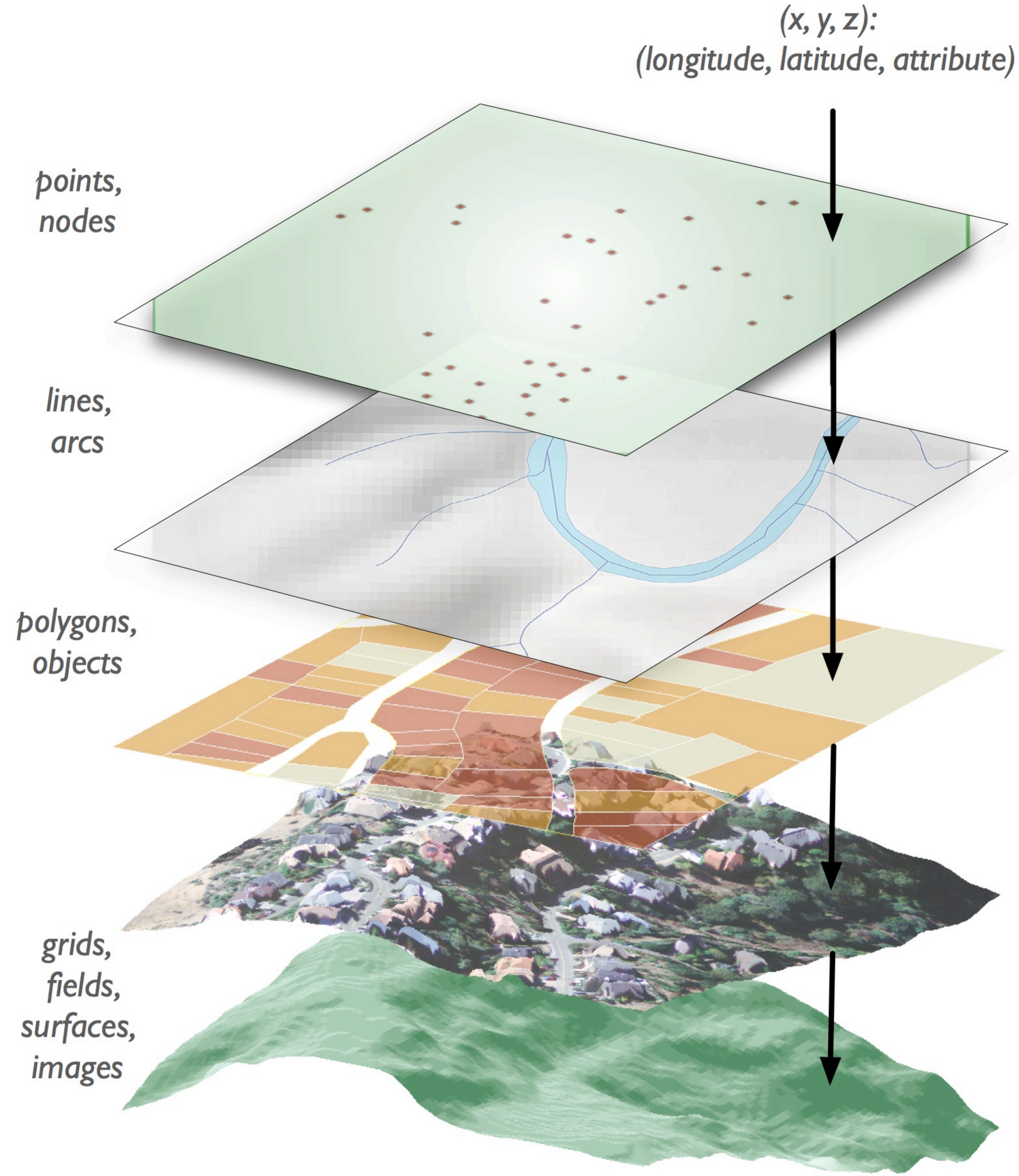
Informatics and GIS Statewide Program

Geospatial Data & Analysis

GIS: a system for the: entry and editing, storage, query and retrieval, transformation, analysis, and display (soft copy) and printing (hard copy) of spatial data

Made possible (and powerful) from the coincidence of spatial data features: *georeferenced data*

Spatial data are a *representation* of reality



Maps are Analytical Tools

Spatial analysis the process by which we turn raw data in to useful information to make decisions.

Spatial analysis is a set of methods *whose results change when the locations of the objects being analyzed change*

This is the heart of the transition from the concept of a *map as a display tool* to an *analytical tool*

Spatial analysis helps us assess:

-Patterns

-Relationships

-Trends

The GIS Landscape Has Been Changing:

- New data streams – from phones, from APIs, from new satellites and sensors…
- Technological advances in computing- advances in database integration, cluster computing, big data, more choices in coding…
- More public focus on geography…
- Global and granular scales: more personal and more broad…
- The spatial technological landscape is more relevant and more omnipresent than ever…
- More chances and more need to collaborate…

…But fundamentally, spatial data and maps encode and reveal content-rich stories about our environment

Key Questions in My Work

1. ***Understanding the current landscape***
   - Feature extraction & object-based
   - Comparing datasets and data products
   - Pattern and classification
   - Public engagement and interaction
2. ***Understanding change, histories, past legacies***
   - Data digitization & accuracy
   - Data integration
   - Reproducibility
   - Data access, open data
3. ***Predicting futures***
   - Ensemble models
   - Data accuracy and uncertainty
   - Visualization, collaboration, story telling

data: On the spatial side of things...

citizen science

drones/UAVs

high resolution imagery

dark data

lidar and radar

downscaled climate data

APIs

Landsat

nano satellites

sensor

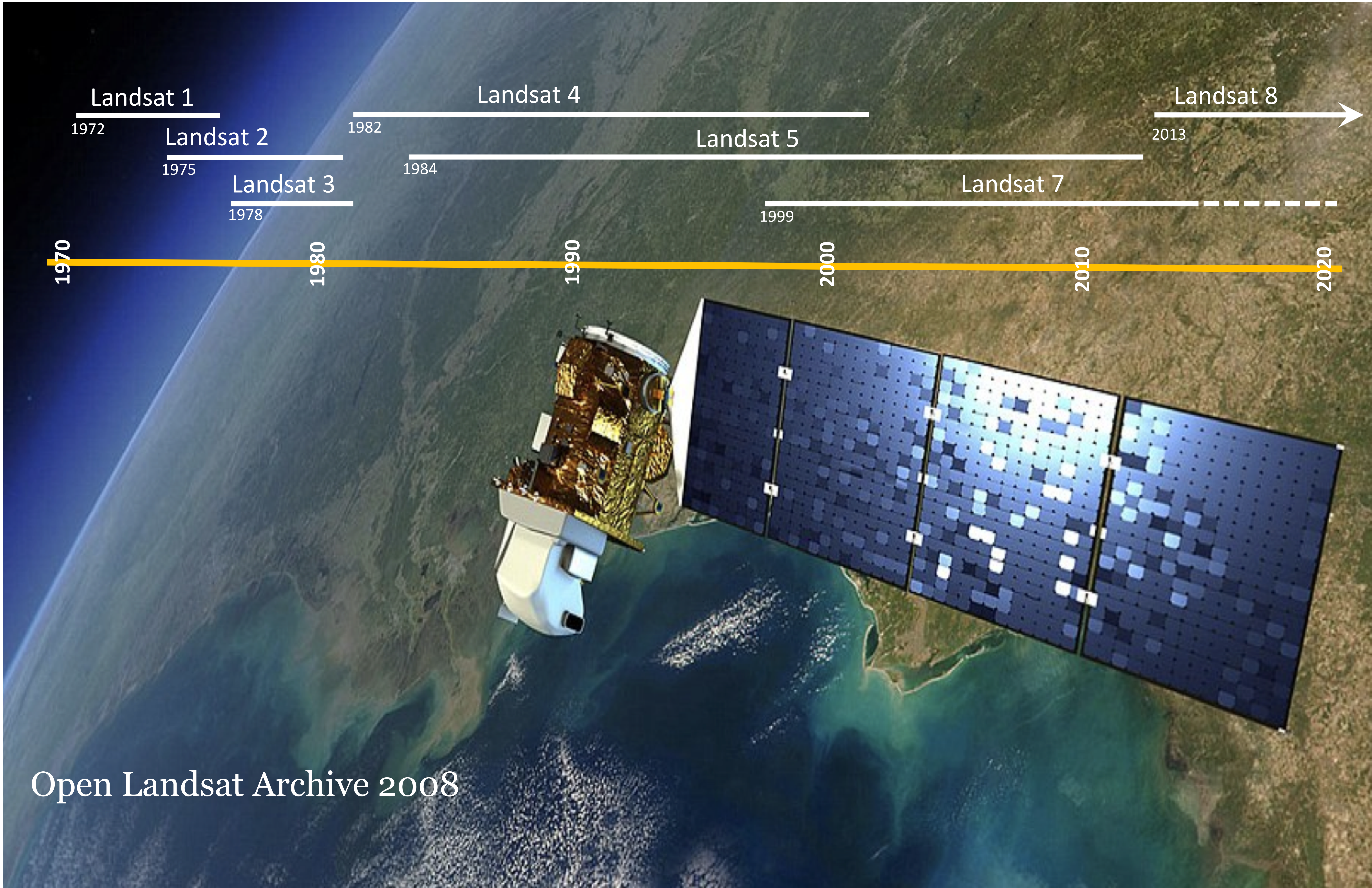
VGI

open data

networks

tracking

high spatial-temporal resolution


Landsat 1
1972
Landsat 2
1975
Landsat 3
1978
Landsat 4
1982
Landsat 5
1984
Landsat 7
1999
Landsat 8
2013
1970
1980
1990
2000
2010
2020
Open Landsat Archive 2008

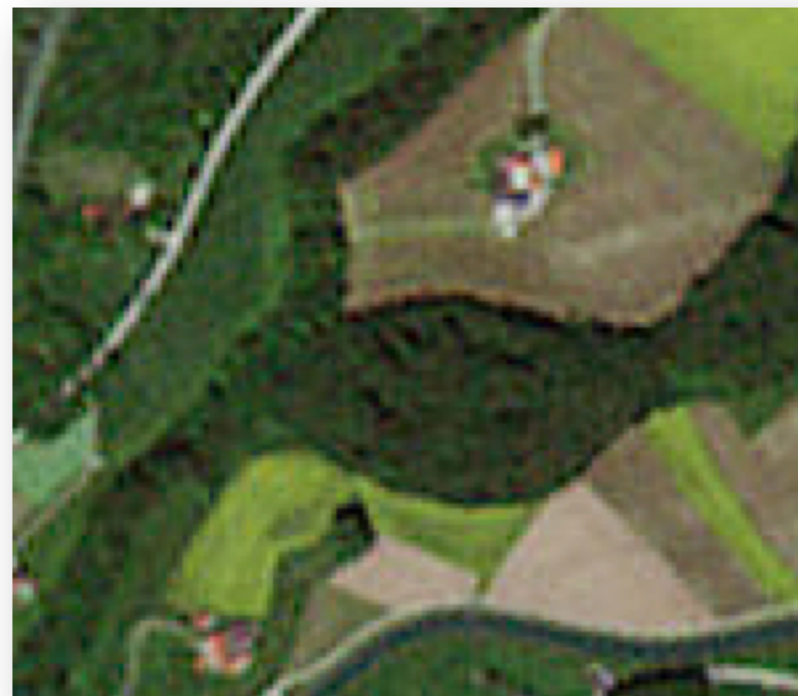
Sensor Trends: High Spatial Resolution

WorldView2	**WorldView3**	**GeoEye-2 / WorldView4**
8 bands (+Pan)	8 bands (+Pan)	4 bands (+Pan)
Daily	Daily	3 days
0.5-1.8 m	0.3-1.2 m	0.3-1.2m
Fee	Fee	Fee

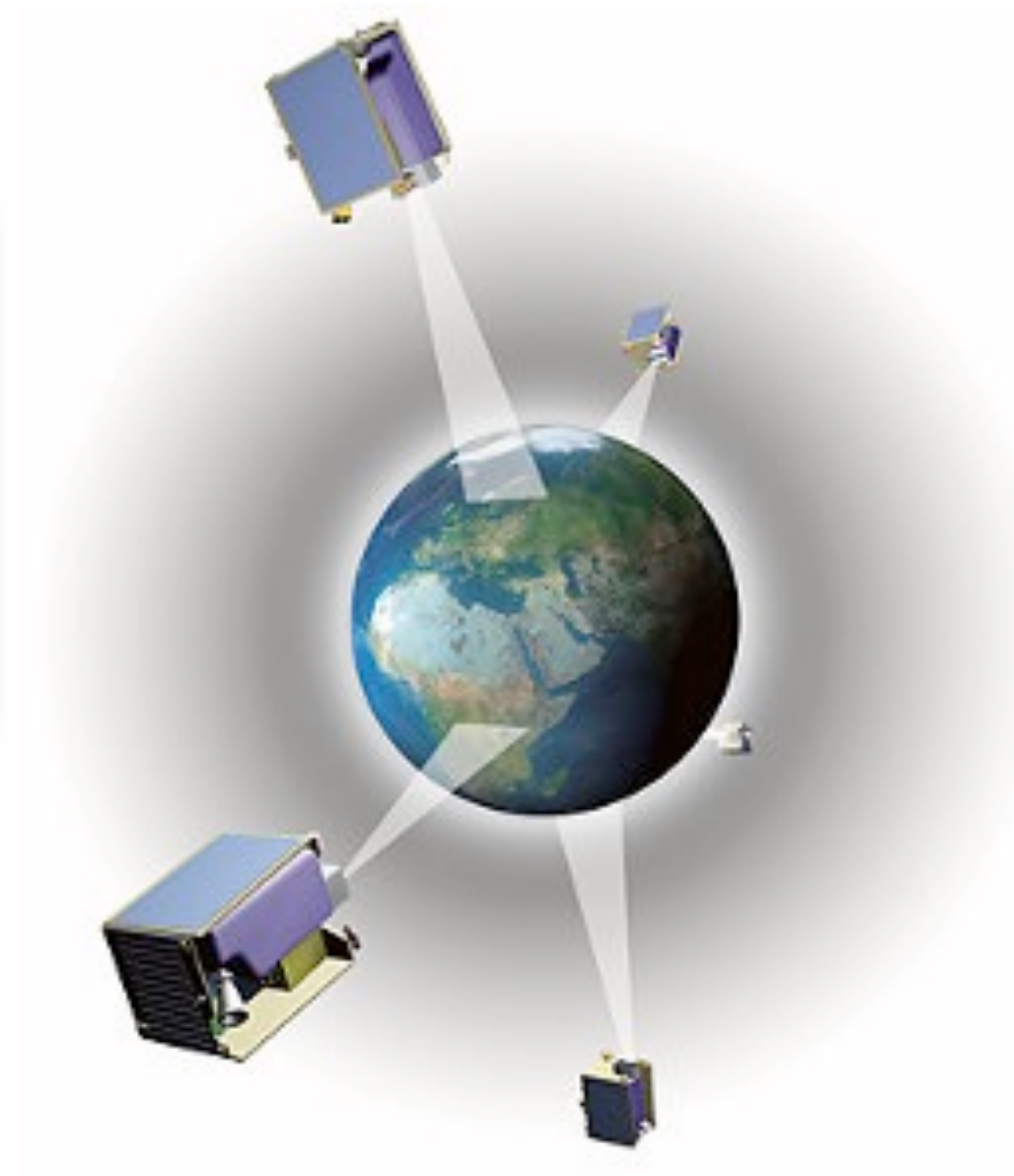
DigitalGlobe

Sensor Trends: Constellations

Sentinel2
13 bands
5-days
10, 20, 60m
2 satellites
Cost



RapidEye
5 bands
Daily
5 m
5 satellites
Cost

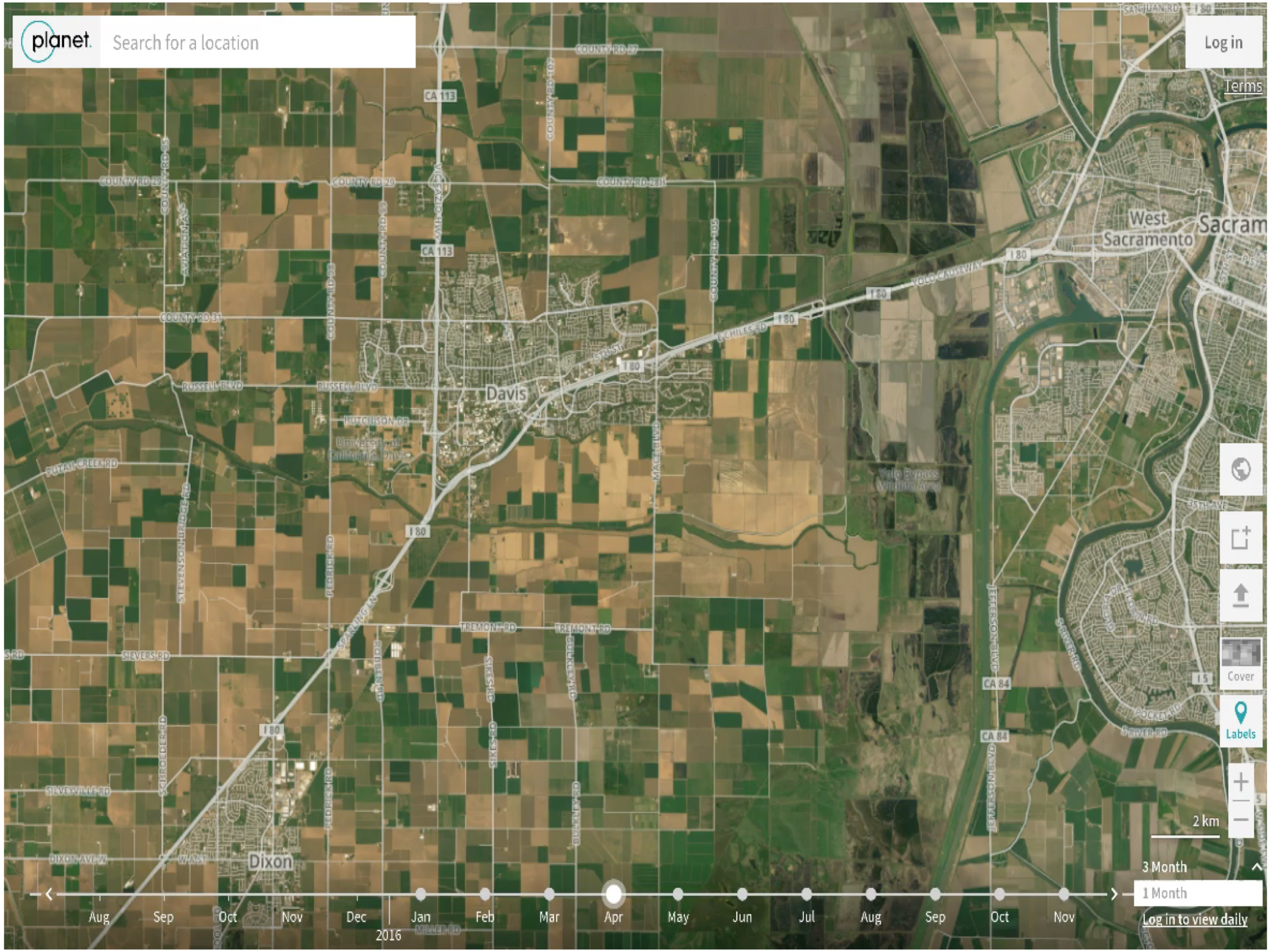


Sensor Trends: Constellations

Planet Launches 88 micro satellites ("cubesats"), Feb 2017

Planet will image the entire earth, daily, at ~4m resolution

Sensor Trends: High Spatial-Temporal Resolution



Sensor Trends: High Spatial-Temporal Resolution

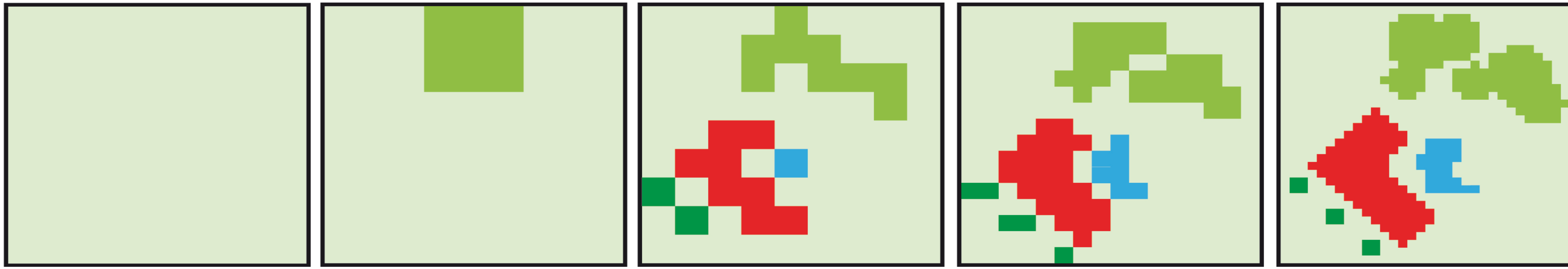
John Cherbini; Blue Oak Reserve; Feb 10; NDVI from 08:30 through 14:30

Sensor Trends: Imagery on Demand

IGIS at the Desert REC

Data Trends: Decrease in Spatial Resolution

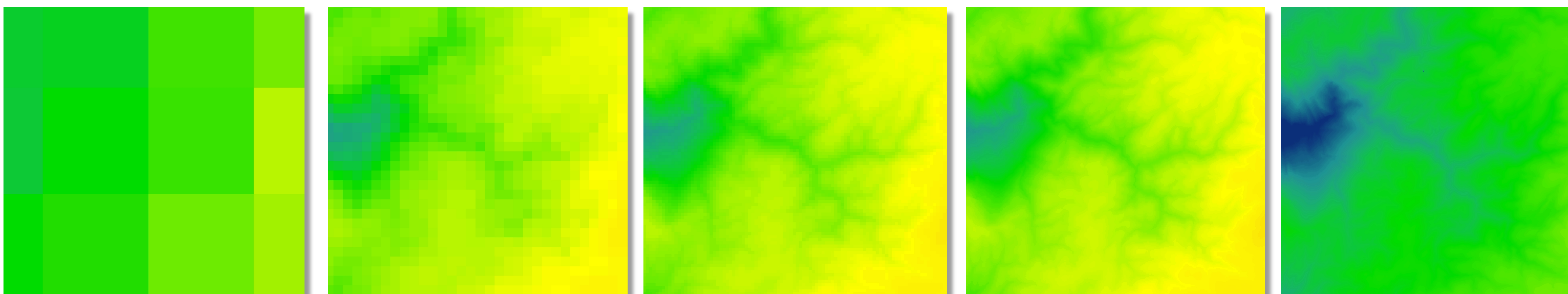
Remote Sensing



Landsat 30m | SPOT5 10m | IKONOS2 3.3m | WV-2 1.8m | GeoEye PS 0.5m

Blaschke et al. 2015

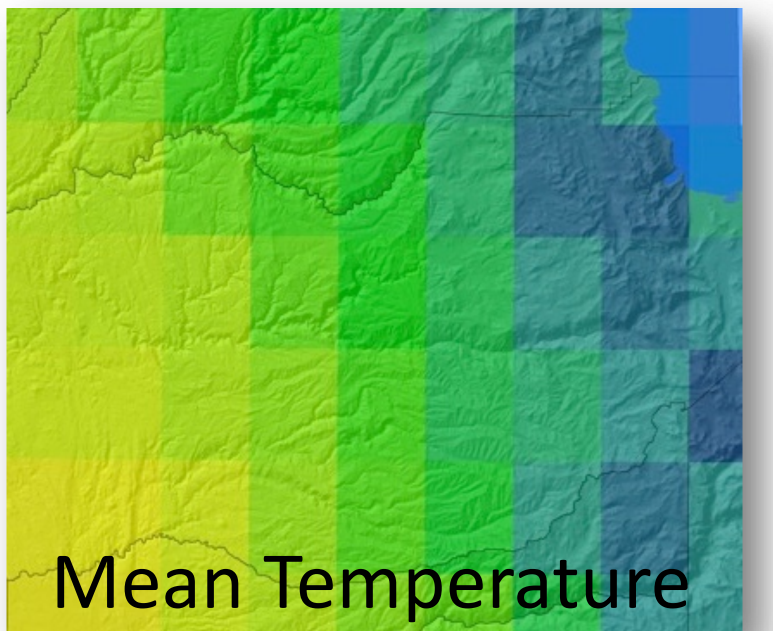
DEMs



GTOPO30 1km | SRTM 90m | USGS NED 30m | USGS NED 10m | USGS Lidar 1m

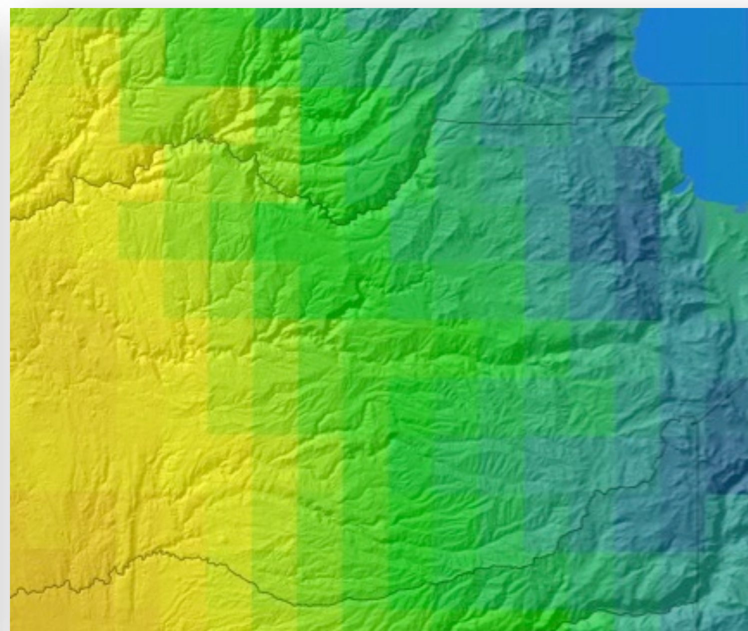
Data Trends: Climate Model Downscaling

11km resolution



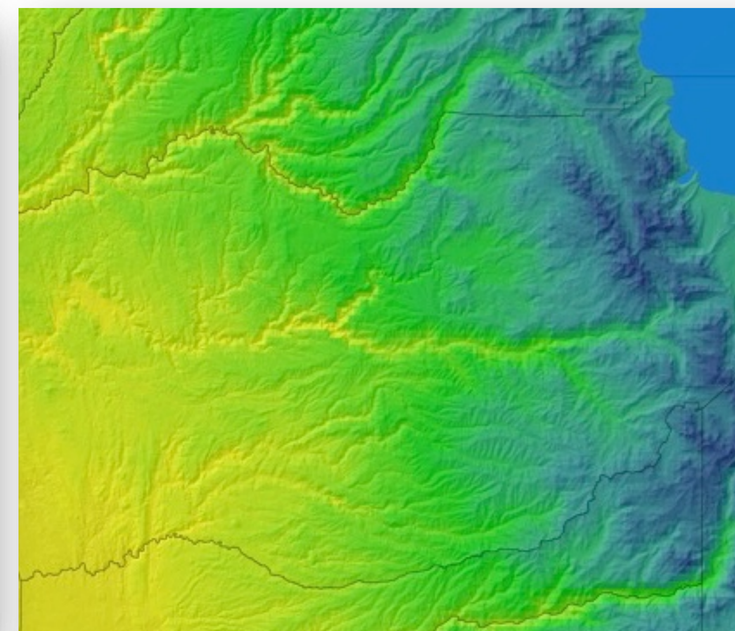
1/8-degree
Cal-Adapt.org 1.0

6km resolution



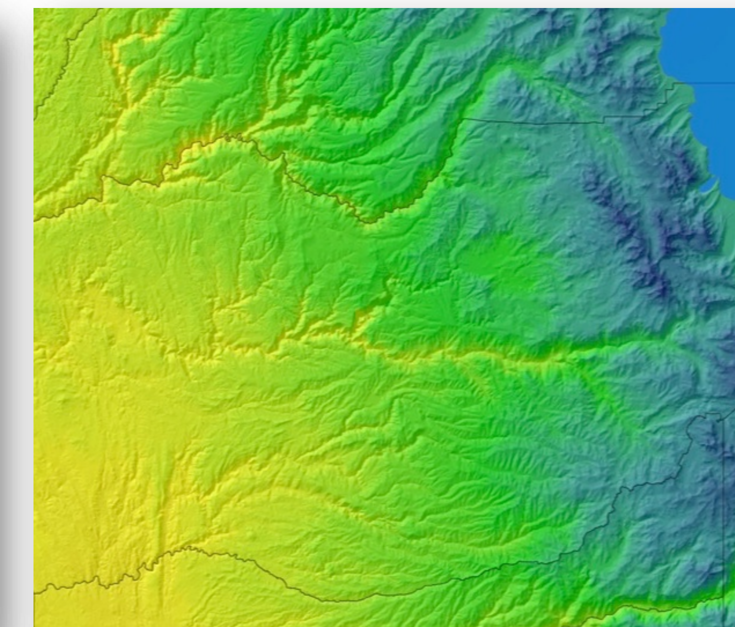
1/16-degree
Cal-Adapt.org 2.0

800m resolution



30-arcsecond
California Climate Action

270 m resolution



USGS Flints'
BASIN model

Data Trends: Detailed, Accurate Location

2005 Onwards: Mobile Acquisition

In 2015, there were 98 cellular phone subscriptions per 100 people

- World Bank World Development Indicators
– Birenboim & Shoval, 2016

www.statistica.com

Number of Mobile Phone Users (M) Globally

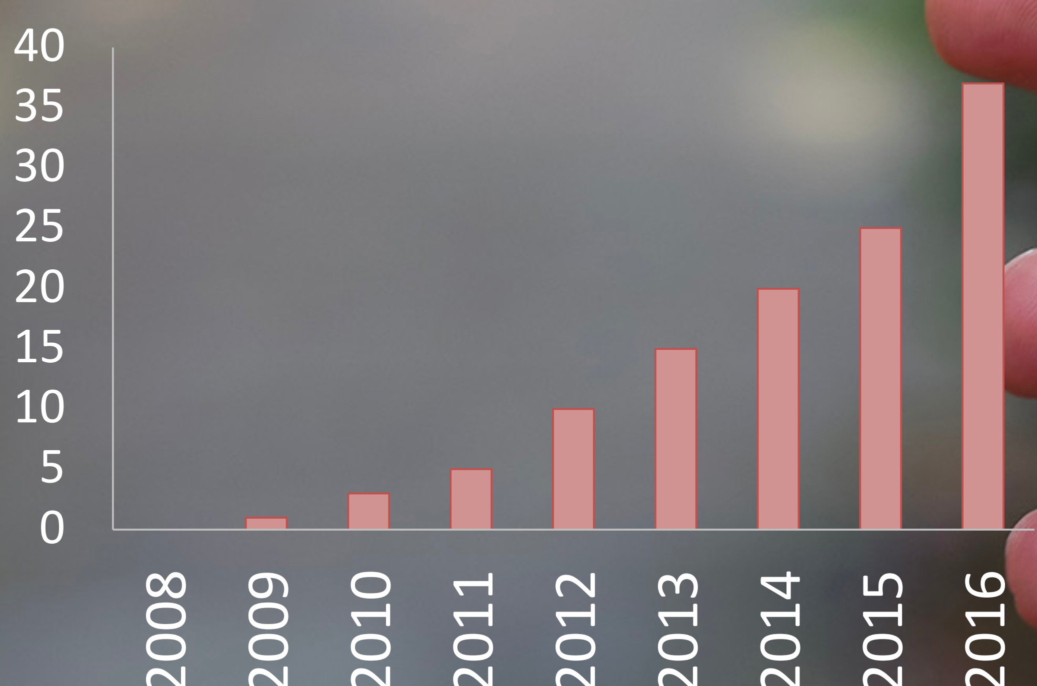
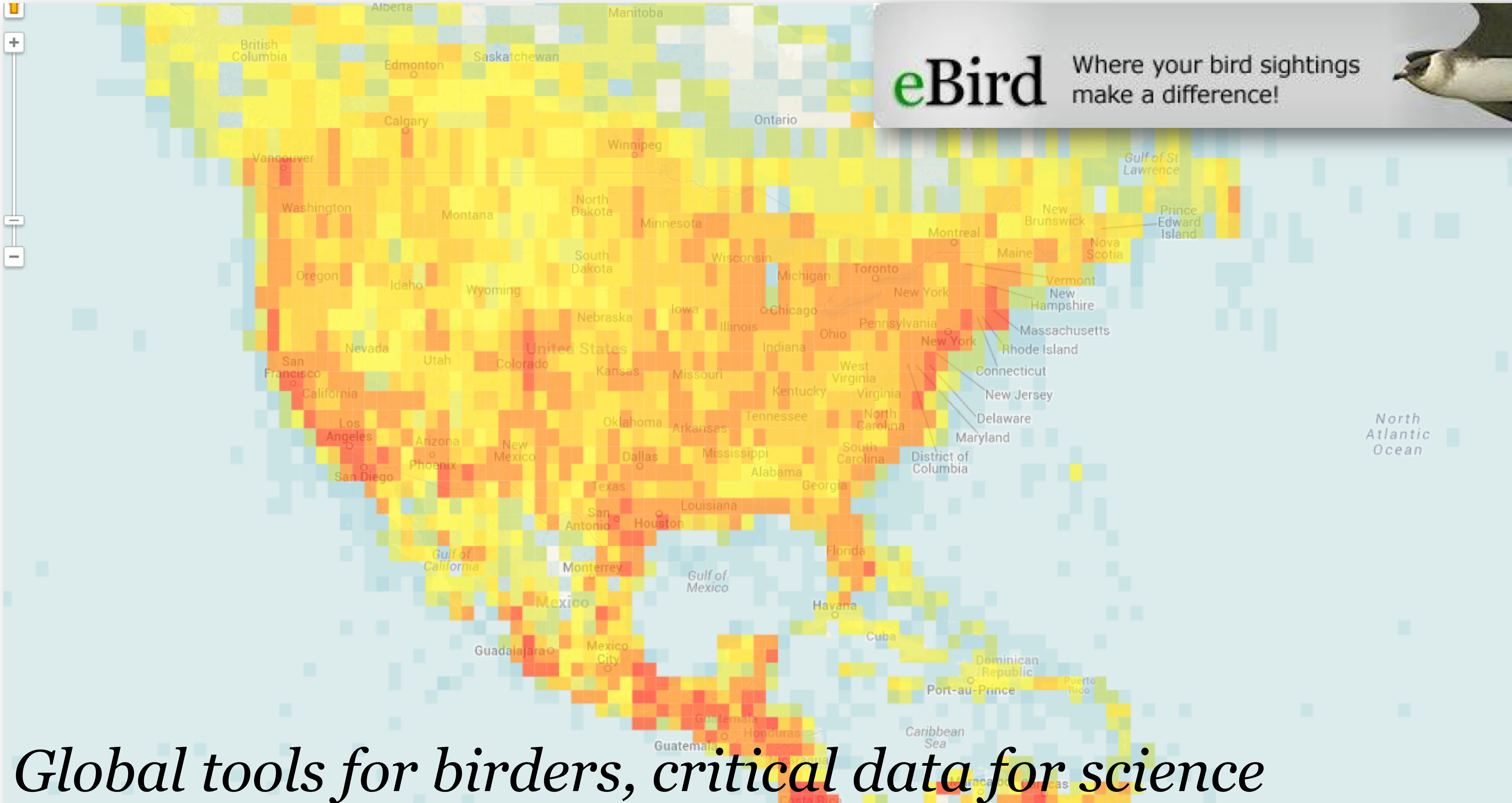


Photo credit: opentreemap

Data Trends: Citizen Science and VGI

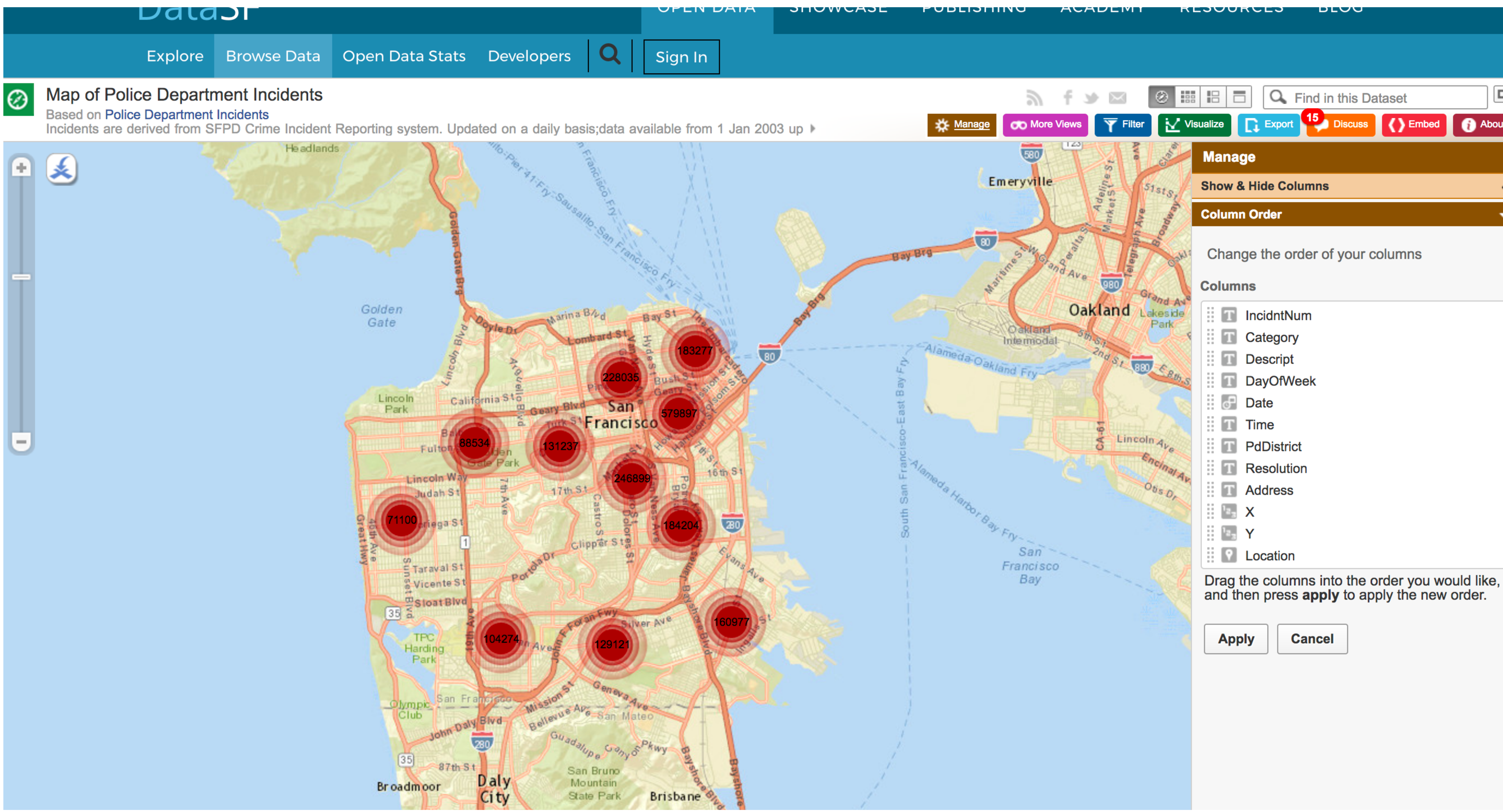
Photo credit: nerds for nature

Trends: Citizen Science & VGI



Global tools for birders, critical data for science

Trends: Open Data



https://data.sfgov.org/

data: Spatial Data is ***Big Data***

detail

breadth

continuity

focus

resilience

tools: Analysis & Web Mapping

open tools

web mapping

shared working environments

distributed computing

visualization

space-time

3D + 4D

analysis

cloud-based

Trends: Open Tools

QGIS is a user friendly Open Source GIS licensed under the GNU General Public License.

QGIS is an official project of the Open Source Geospatial Foundation (OSGeo).

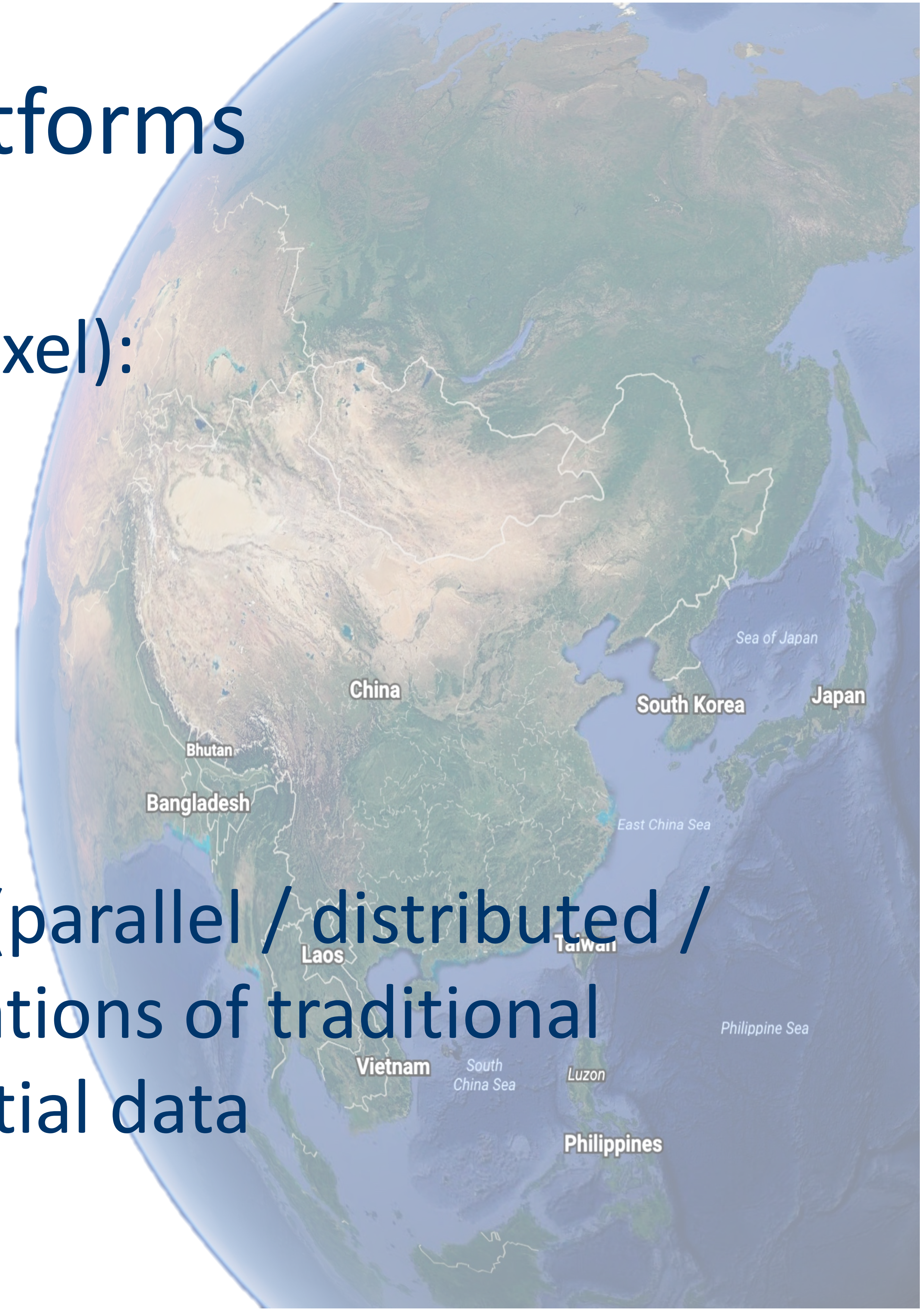
It runs on Linux, Unix, Mac OSX, Windows and Android and supports numerous vector, raster, and database formats and functionalities.

Trends: Cloud & HPC Platforms

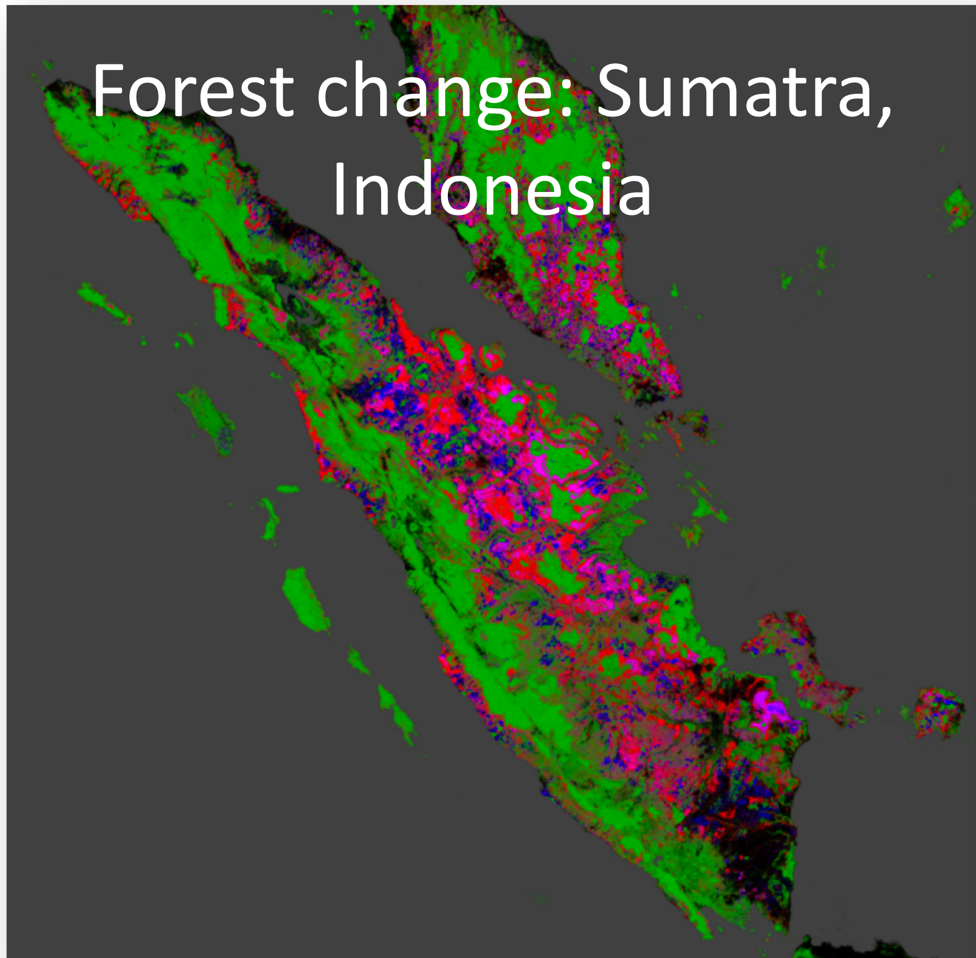
At the scale of Landsat (30m pixel):

- California → 500M pixels
- China → 10B pixels
- Globe → 800B pixels

High Performance Computing (parallel / distributed / clustered) supports the applications of traditional geospatial methods on big spatial data

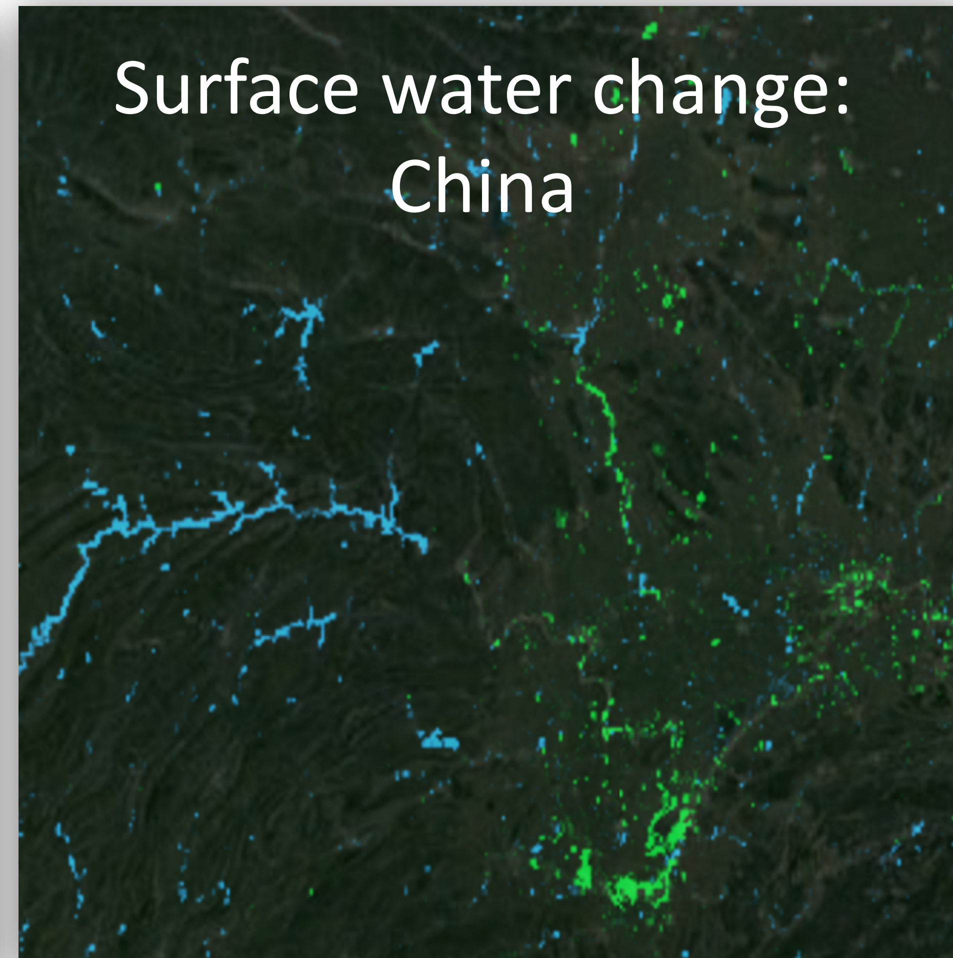


Large Scale Cloud Computation: Google Earth Engine



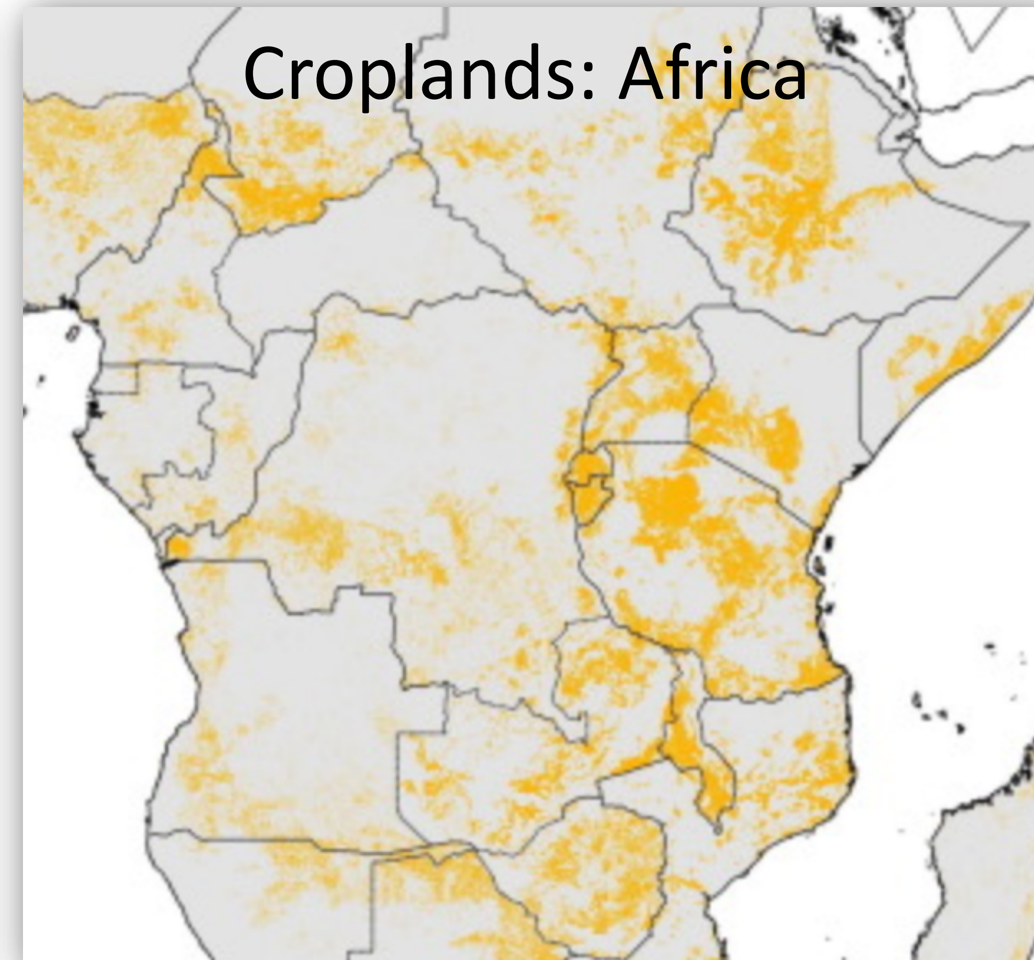
Forest Loss 2000–2013
Forest Gain 2000–2013
Both Loss and Gain
Forest Extent

Hansen et al. 2013, *Science*



Surface Water Loss
Surface Water Gain

Pekel et al. 2016, *Nature*

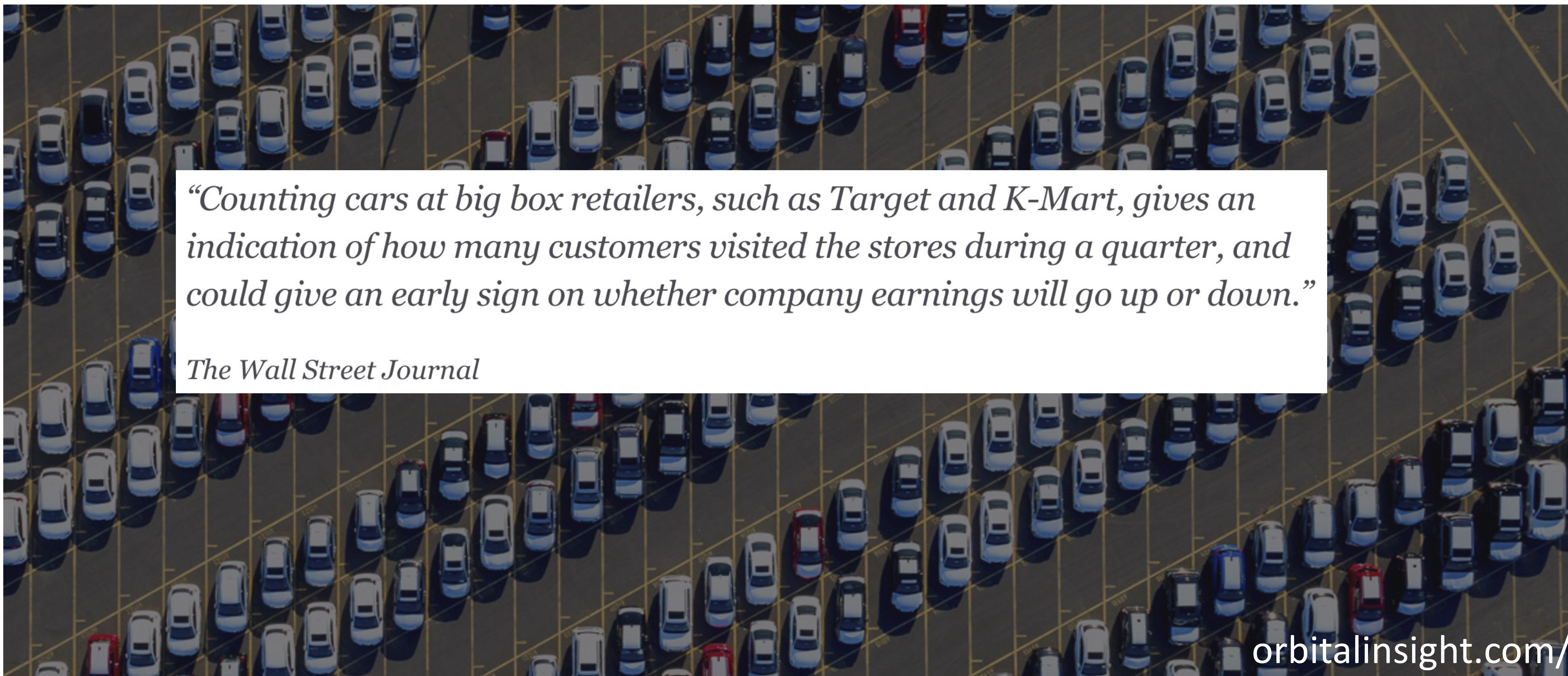


Cropland with fallow
Non-cropland

Xiong et al. 2017, *ISPRS Journal of Photogrammetry and Remote Sensing*

Trends: AI, Machine Learning, Deep Learning

Machine Learning is the practice of using algorithms to parse data, learn from it, and then make a prediction. The machine is "trained" using large amounts of data and algorithms that give it the ability to learn how to perform the task. Algorithms like support vector machines, decision trees, Random Forests, boosted DTs, artificial neural networks are all involved.



Trends: Web Mapping & Visualization

Climate Tools

Explore charts, maps and data of observed and projected climate variables for California. The tools show projections for two possible climate futures, one in which emissions peak around 2040 and then decline (RCP 4.5) and another in which emissions continue to rise throughout the 21st century (RCP 8.5). Both futures are considered possible depending on how much action we decide to take. Learn more about working with climate projections on our Resources page.

Annual Averages

Explore projected annual averages of maximum temperature, minimum temperature and precipitation for your location.

Extreme Heat

Explore projected frequency and duration of extreme heat events for your location.

Sea Level Rise - CalFloD-3D

Explore maps of inundation location and depths for San Francisco Bay Area, Sacramento - San Joaquin Delta and the California coast during near 100 year storm events coupled with projected Sea Level Rise scenarios.

Snowpack

View timelapse animation and monthly averages of projected Snow Water Equivalent, a common measurement for snowpack.

Wildfire

Annual averages of area burned for combination of 4 GCMs, 2 RCPs and 3 population growth scenarios.

Cooling Degree Days and Heating Degree Days

Explore projected changes in Heating Degree Days and Cooling Degree Days, which are a common proxy for energy needed to heat and cool buildings

Stream Flow

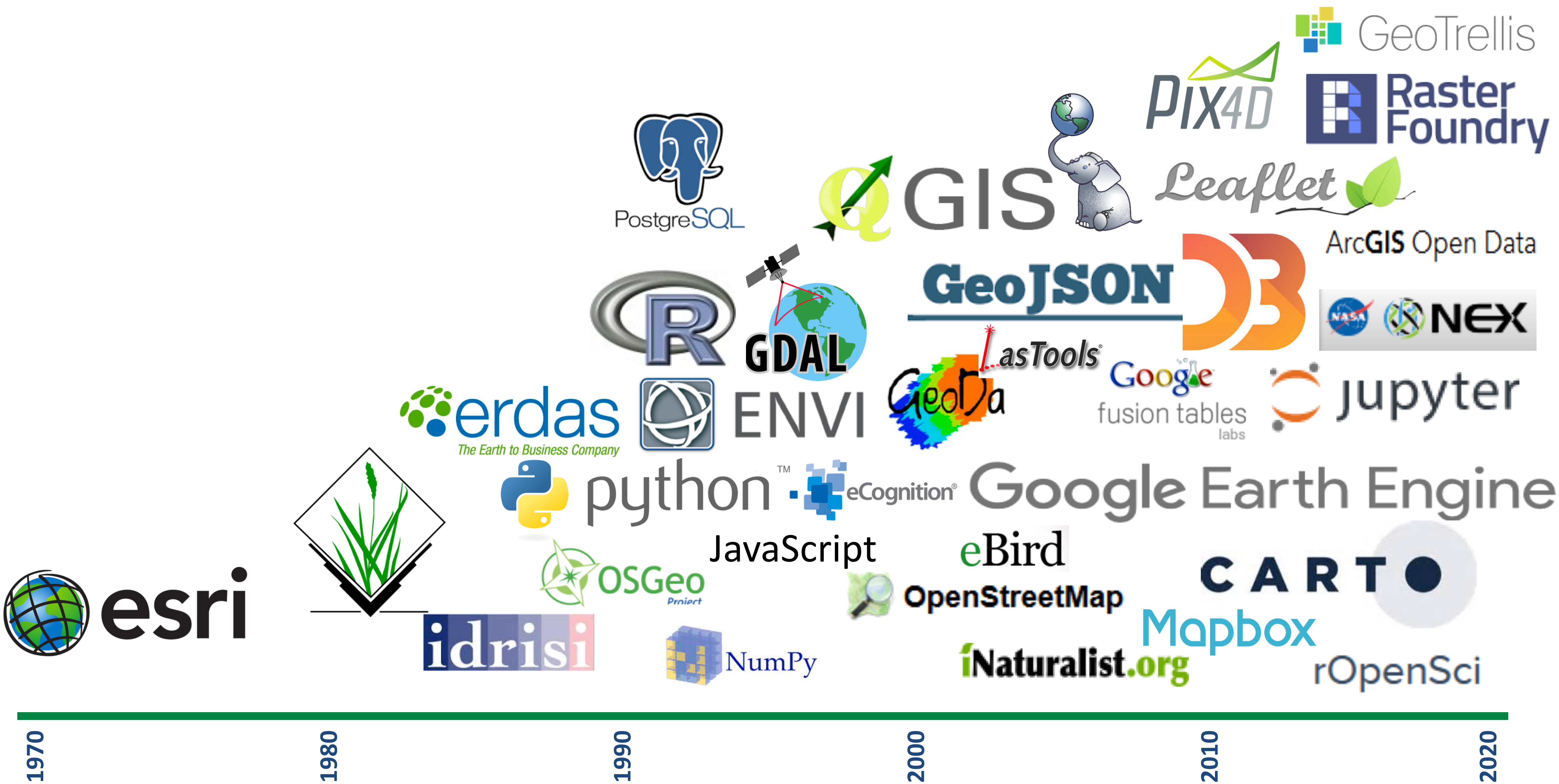
Charts of VIC routed and bias corrected streamflows driven by LOCA downscaled temperature and precipitation.

Long Drought Scenarios (LOCA)

Charts of temperature, precipitation plus a set of VIC variables from two 30 year drought periods.

Coming Soon

Trends: Choices in Analysis & Web Mapping



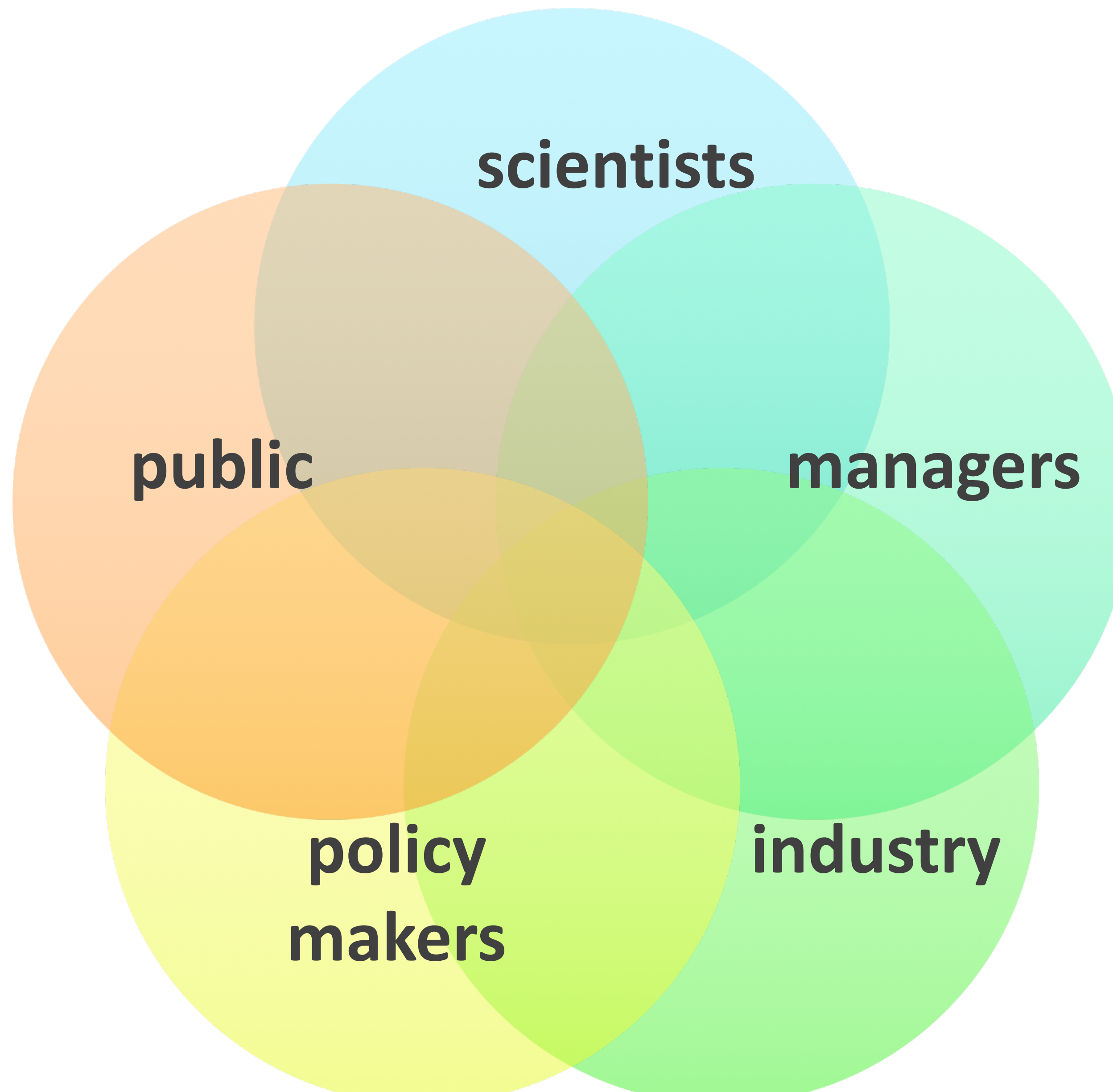
tools: The Spatial Tool Landscape is Changing Fast

bespoke workflows

reproducibility & transparency

Thanks to Lowndes et al. 2017. Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*

people: The Citizen-Science-Practitioner Nexus



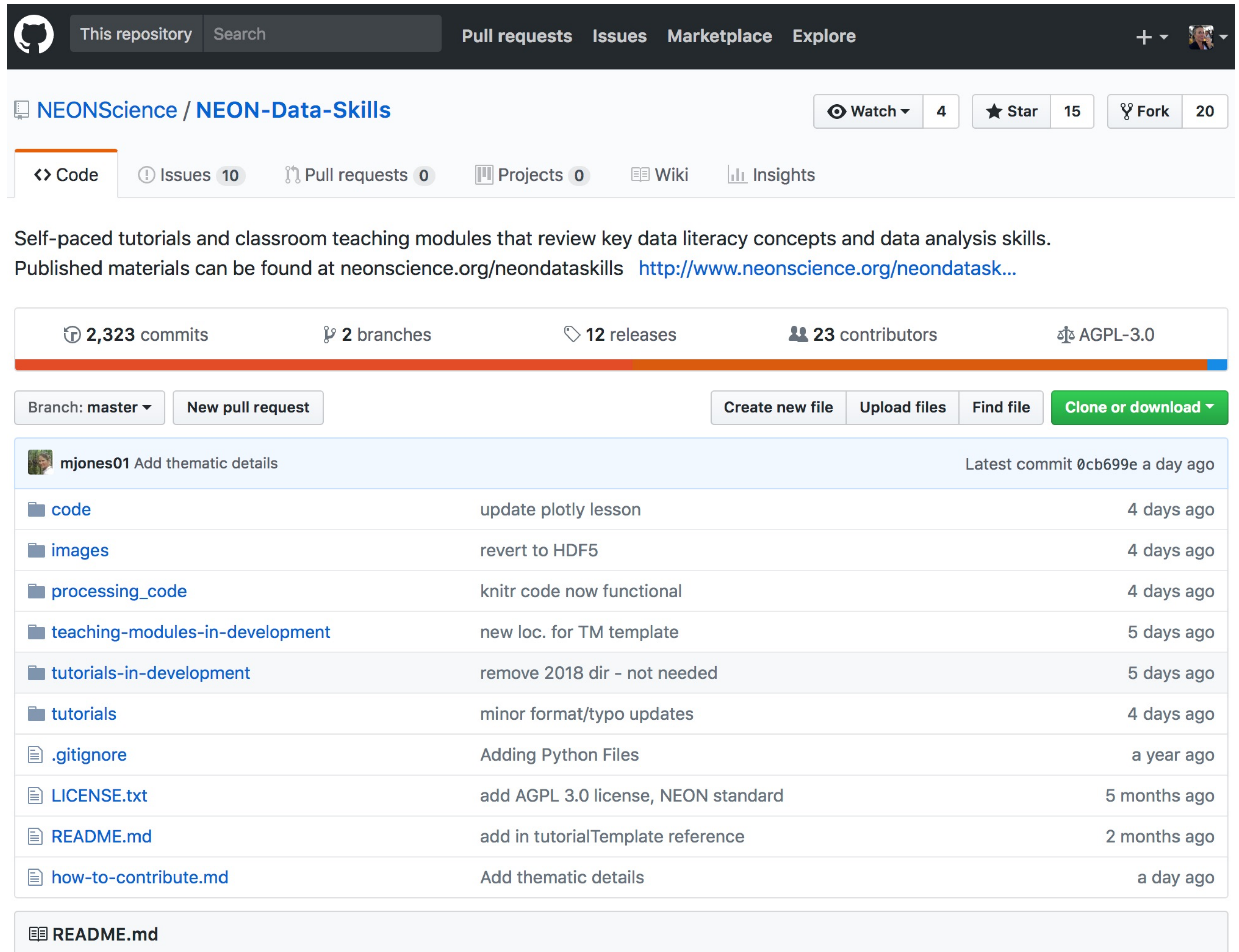
There are social dimensions to collaboration success, yet successful collaboration can also be supported by technology

Mapping makes exchanges between groups easier and more powerful

Trends: Collaborative Workflows

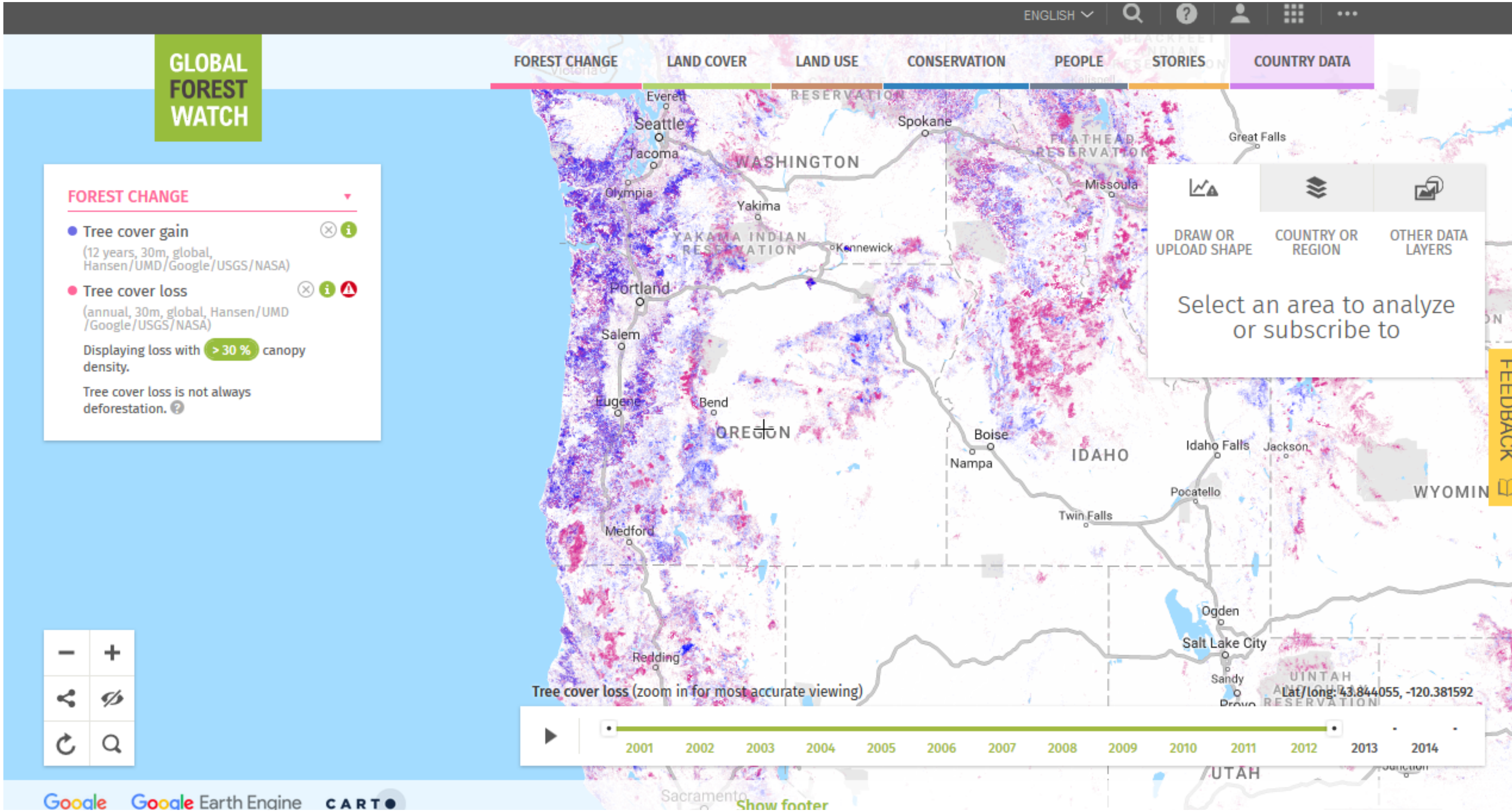
jupyter

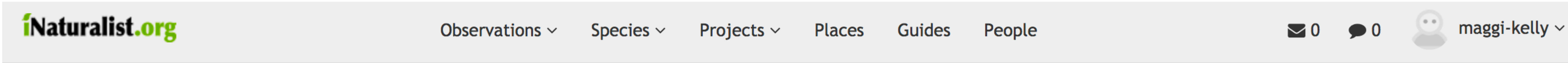
python™



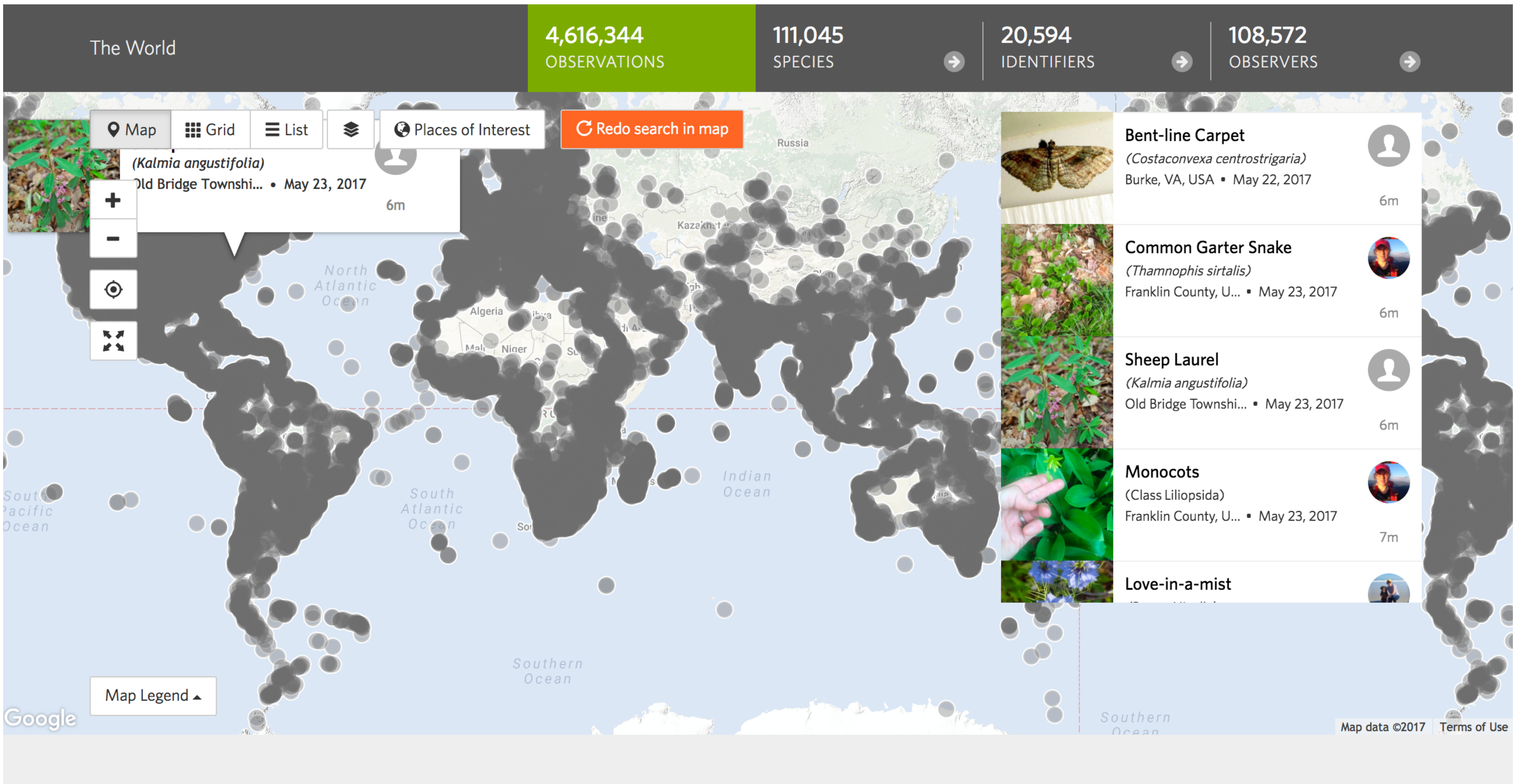
Standardized team workflow conventions that support reproducibility, collaboration & communication

Trends: Geo-collaboratories

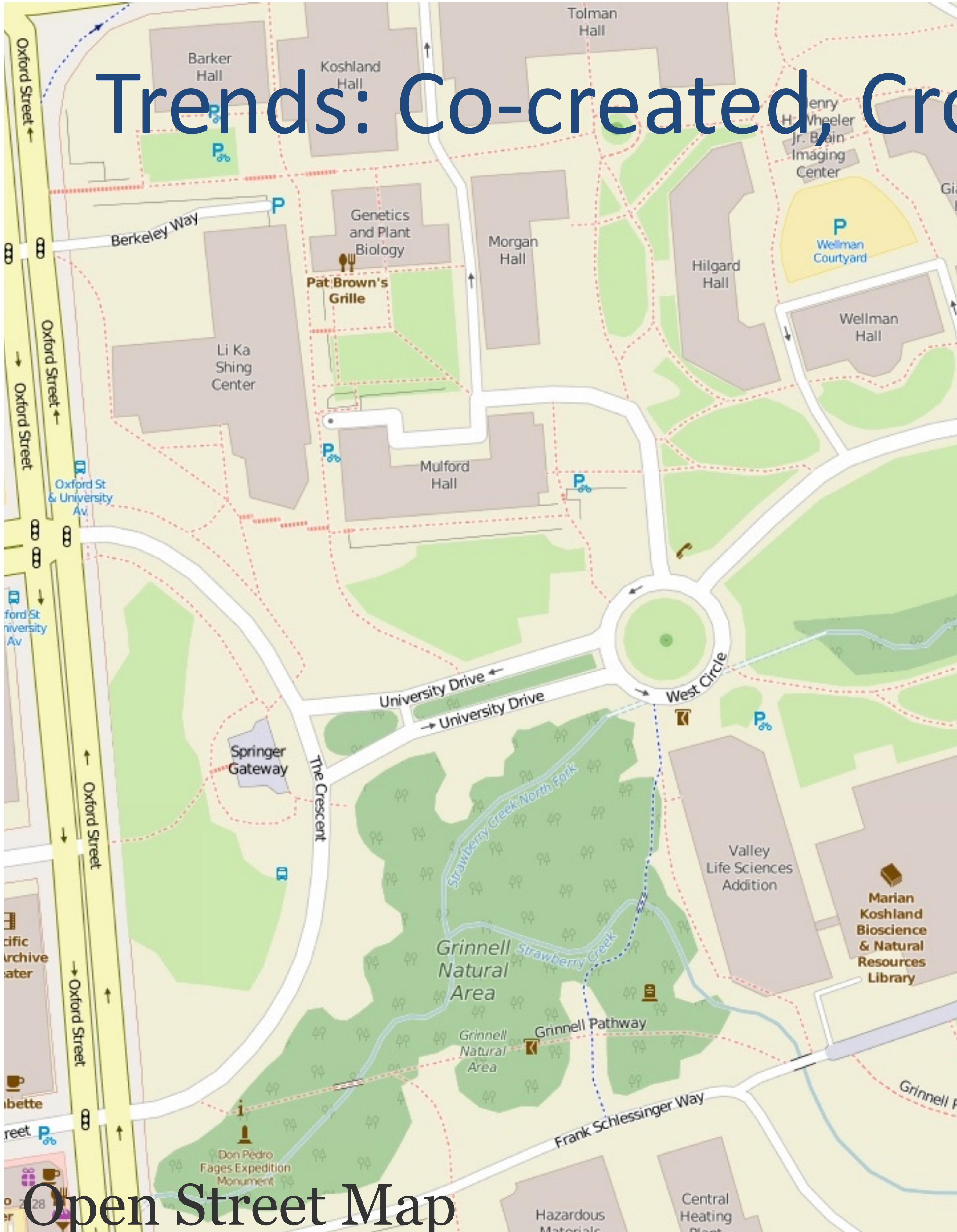




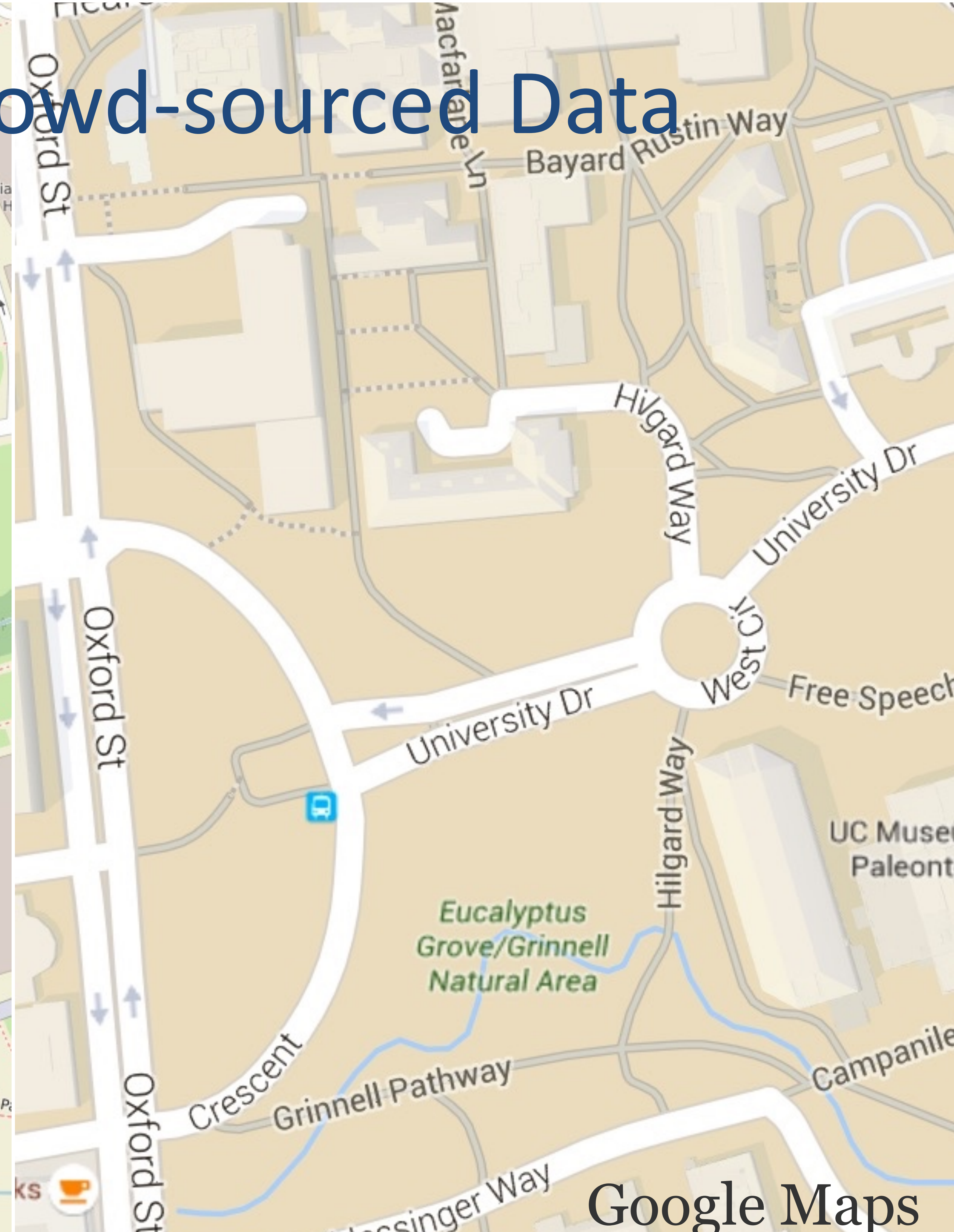
Trends: Geo-collaboratories



Trends: Co-created, Crowd-sourced Data



Open Street Map



Google Maps

people: Collaboration is Critical and Useful

collaboration

engagement

impact

Some Case Studies From My Work

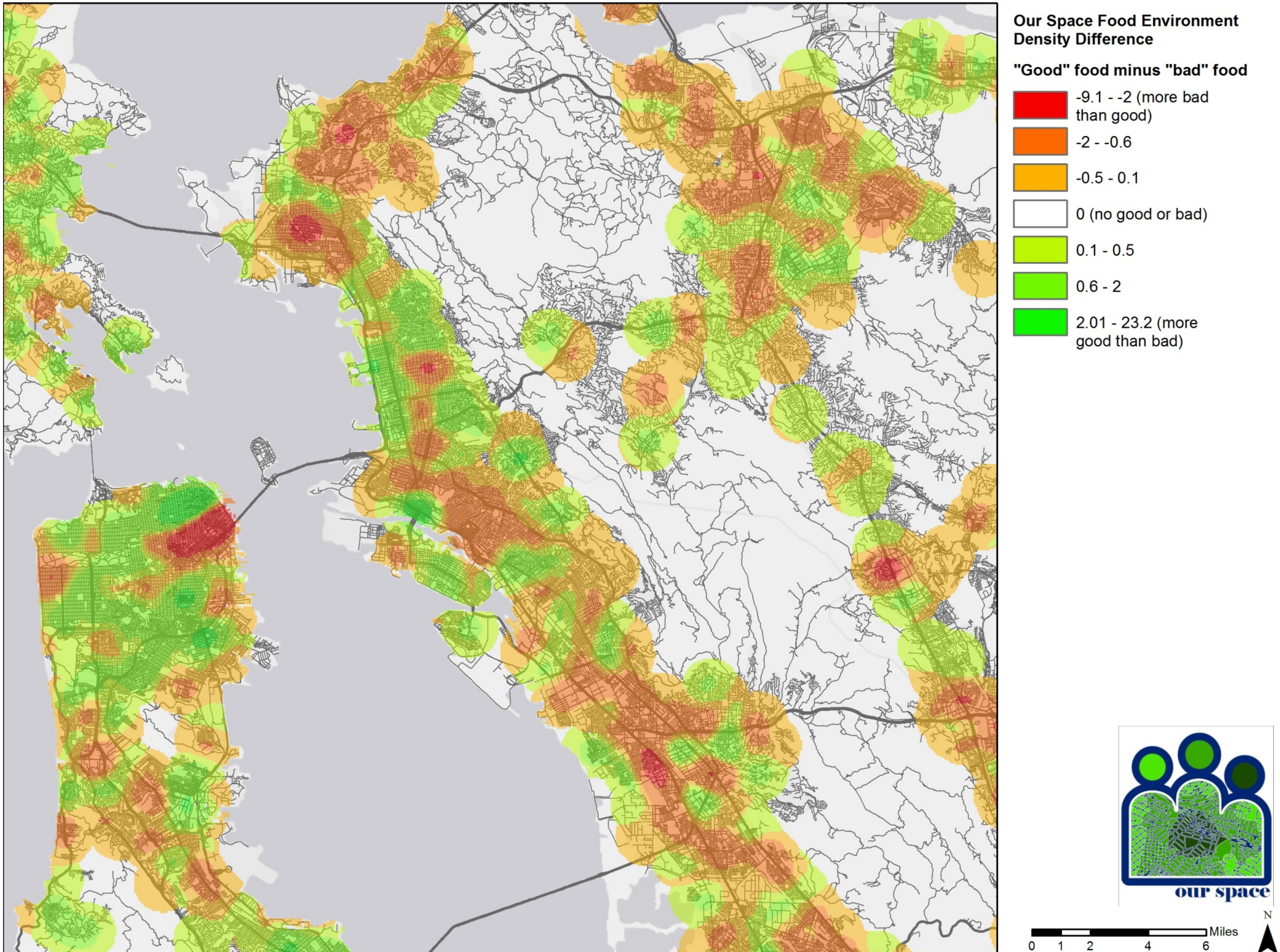
1. ***Understanding the current landscape***
   - Feature extraction & object-based
   - Comparing datasets and data products
   - Pattern and classification
   - Public engagement and interaction
2. ***Understanding change, histories, past legacies***
   - Data digitization & accuracy
   - Data integration
   - Reproducibility
   - Data access, open data
3. ***Predicting futures***
   - Ensemble models
   - Data accuracy and uncertainty
   - Visualization, collaboration, story telling

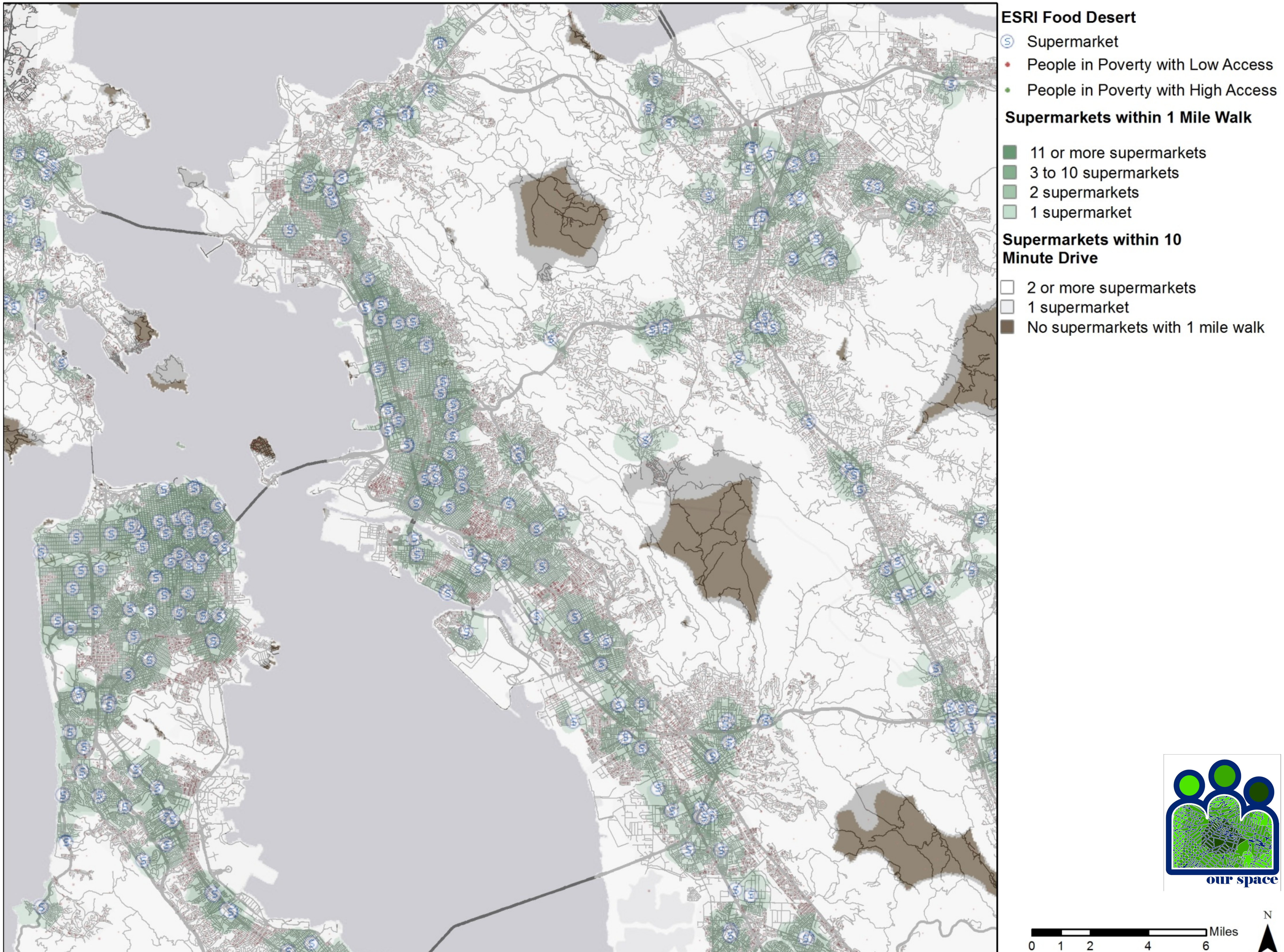
data +

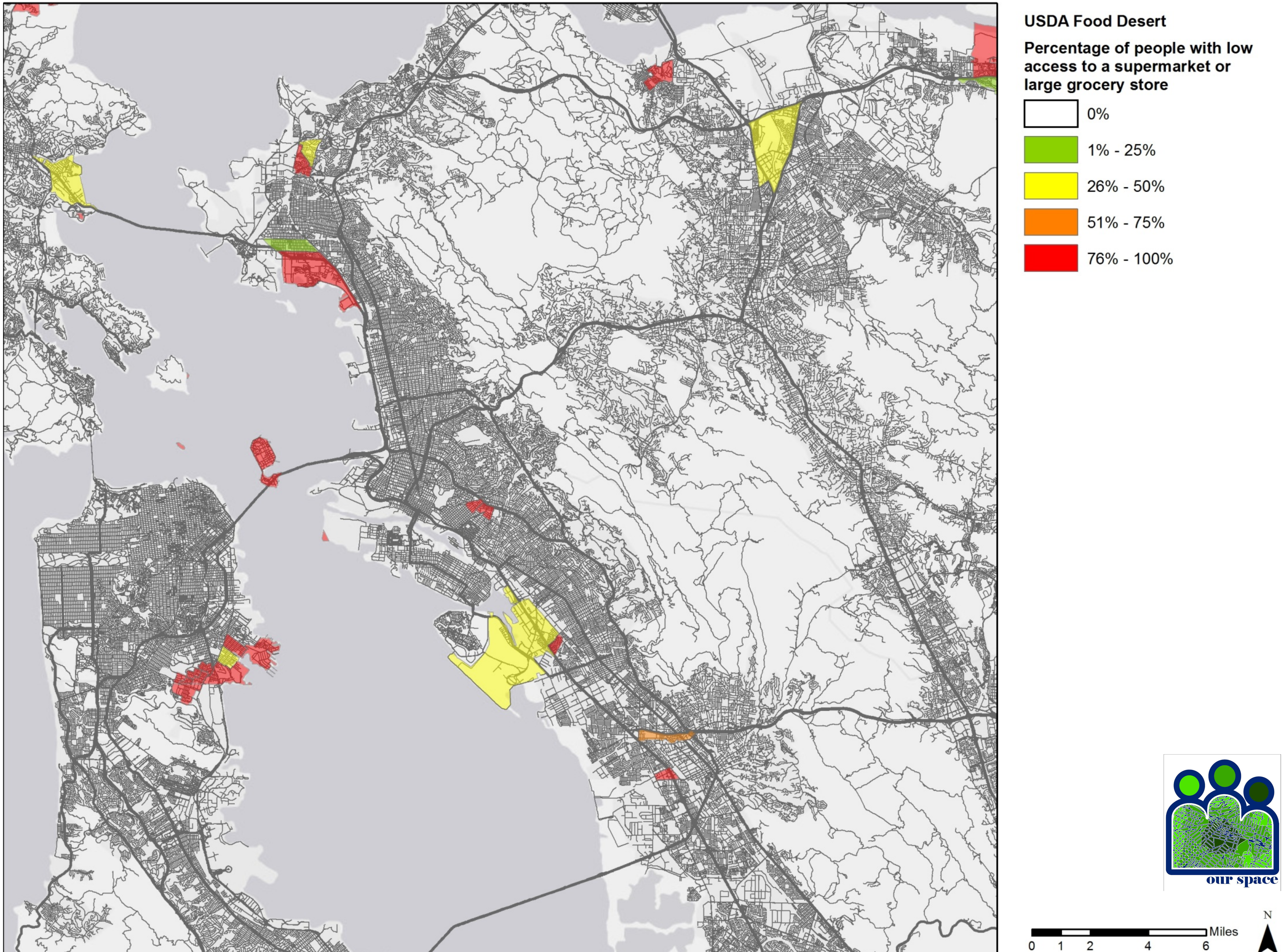
tools +

people

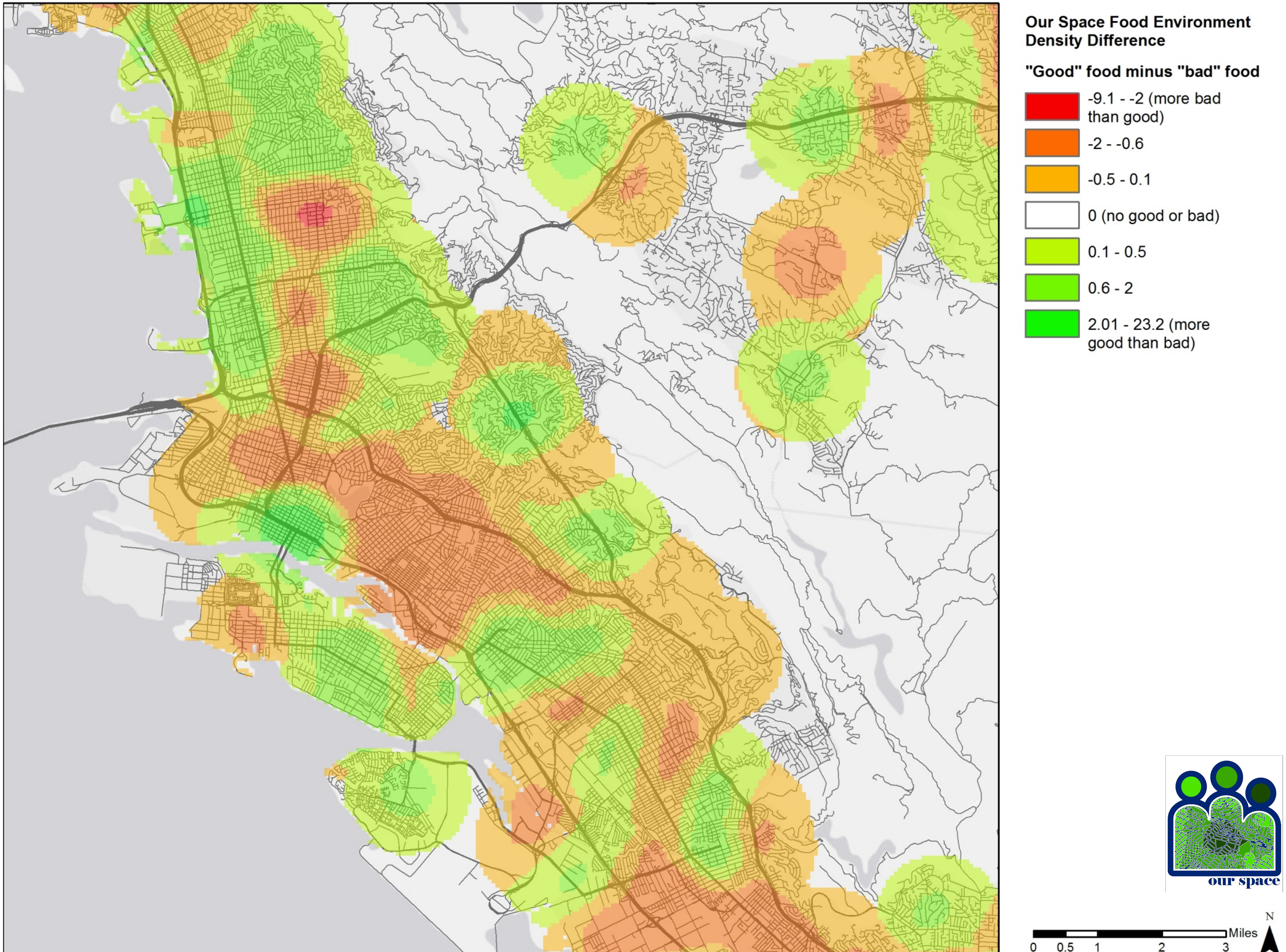
NASA Blue Marble 2.0

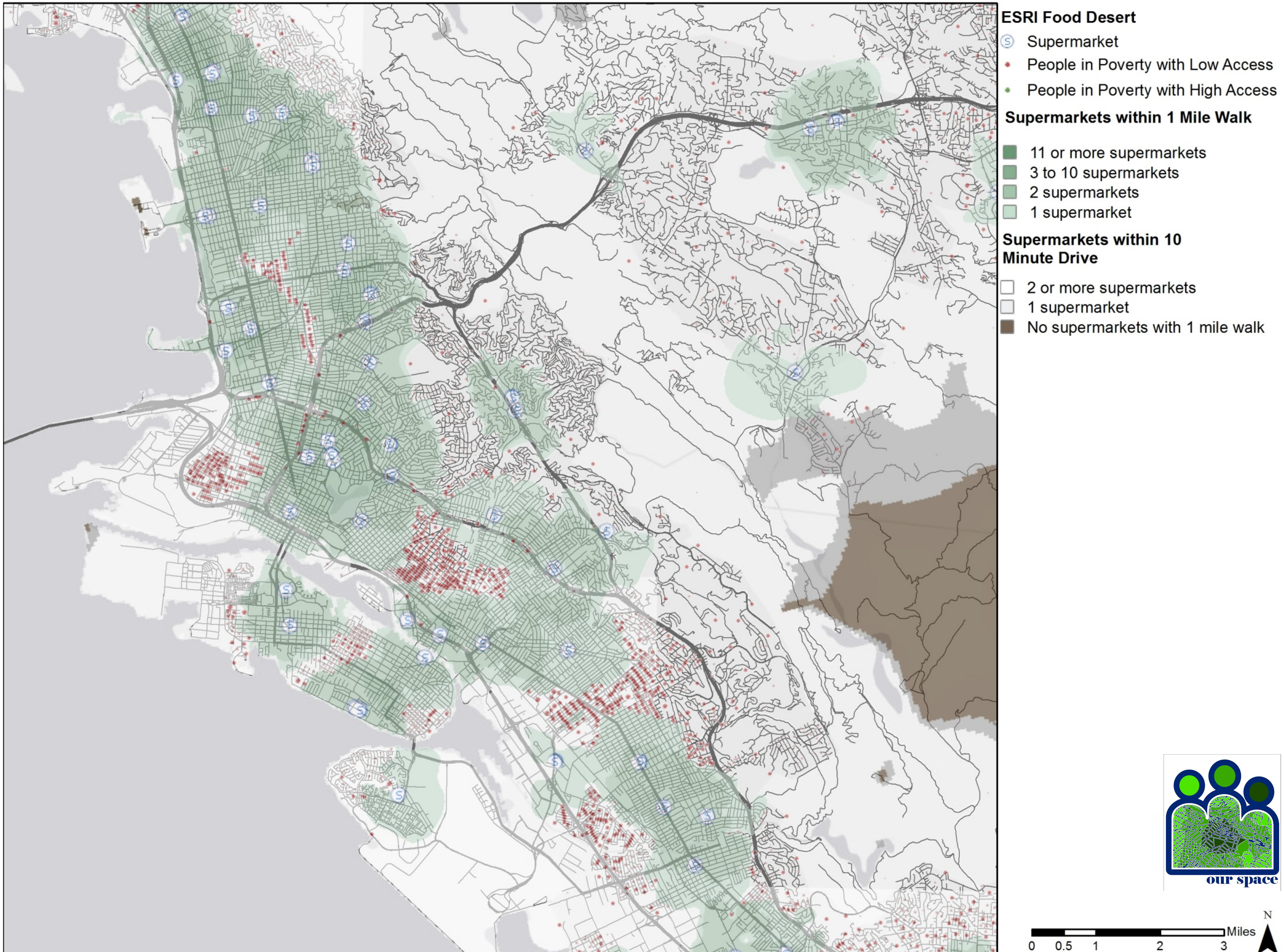

Our Space Food Environment Density Difference
"Good" food minus "bad" food
-9.1 - -2 (more bad than good)
-2 - -0.6
-0.5 - 0.1
0 (no good or bad)
0.1 - 0.5
0.6 - 2
2.01 - 23.2 (more good than bad)
our space
Miles
0 1 2 4 6
N

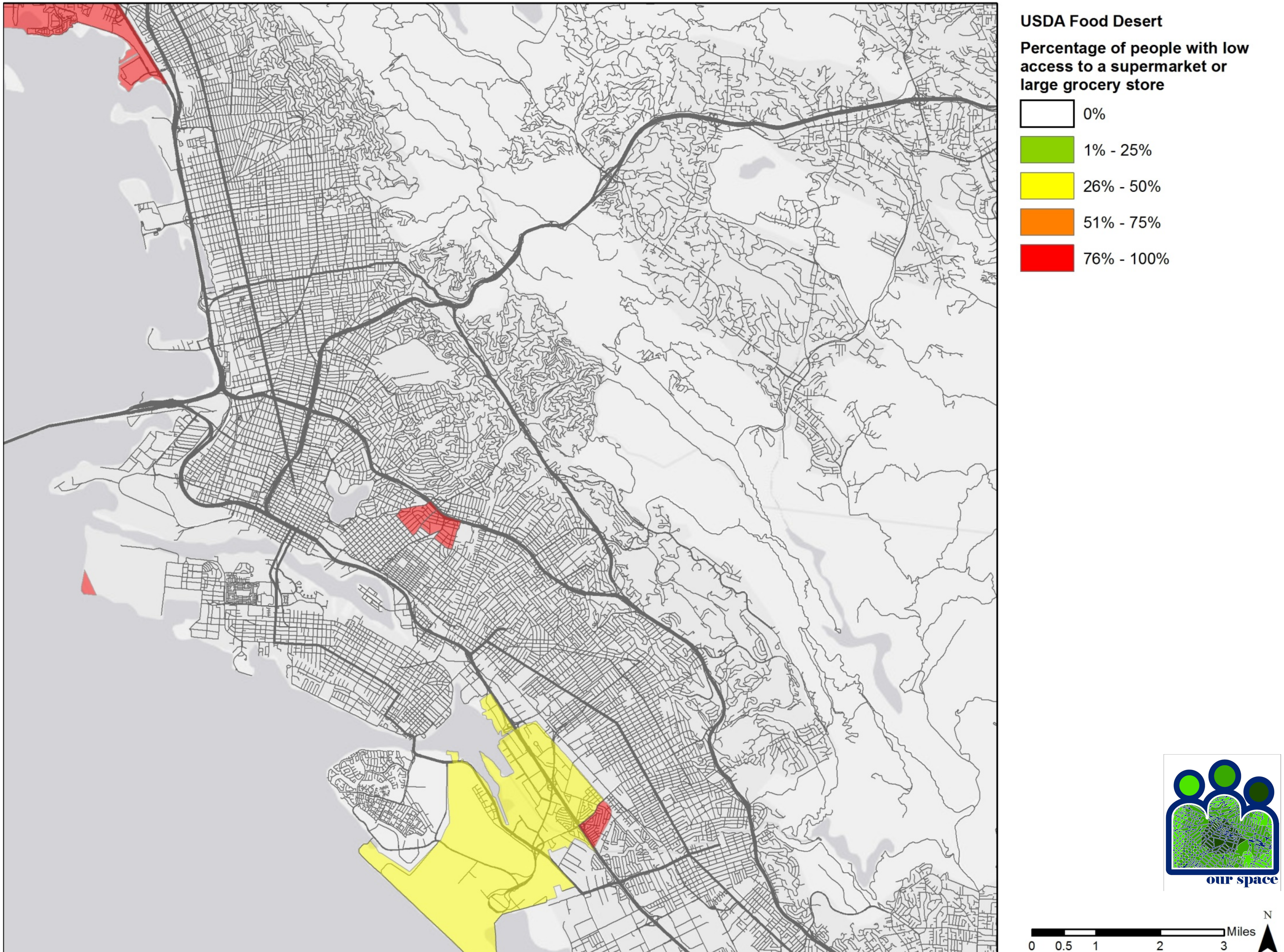

ESRI Food Desert
Supermarket
People in Poverty with Low Access
People in Poverty with High Access
Supermarkets within 1 Mile Walk
11 or more supermarkets
3 to 10 supermarkets
2 supermarkets
1 supermarket
Supermarkets within 10 Minute Drive
2 or more supermarkets
1 supermarket
No supermarkets with 1 mile walk
our space
Miles
0 1 2 4 6
N



USDA Food Desert
Percentage of people with low access to a supermarket or large grocery store
0%
1% - 25%
26% - 50%
51% - 75%
76% - 100%
our space
Miles
0
1
2
4
6
N


Our Space Food Environment
Density Difference
"Good" food minus "bad" food
-9.1 - -2 (more bad than good)
-2 - -0.6
-0.5 - 0.1
0 (no good or bad)
0.1 - 0.5
0.6 - 2
2.01 - 23.2 (more good than bad)
our space
Miles
0 0.5 1 2 3
N


ESRI Food Desert
Supermarket
People in Poverty with Low Access
People in Poverty with High Access
Supermarkets within 1 Mile Walk
11 or more supermarkets
3 to 10 supermarkets
2 supermarkets
1 supermarket
Supermarkets within 10 Minute Drive
2 or more supermarkets
1 supermarket
No supermarkets with 1 mile walk
0 0.5 1 2 3 Miles
N
our space



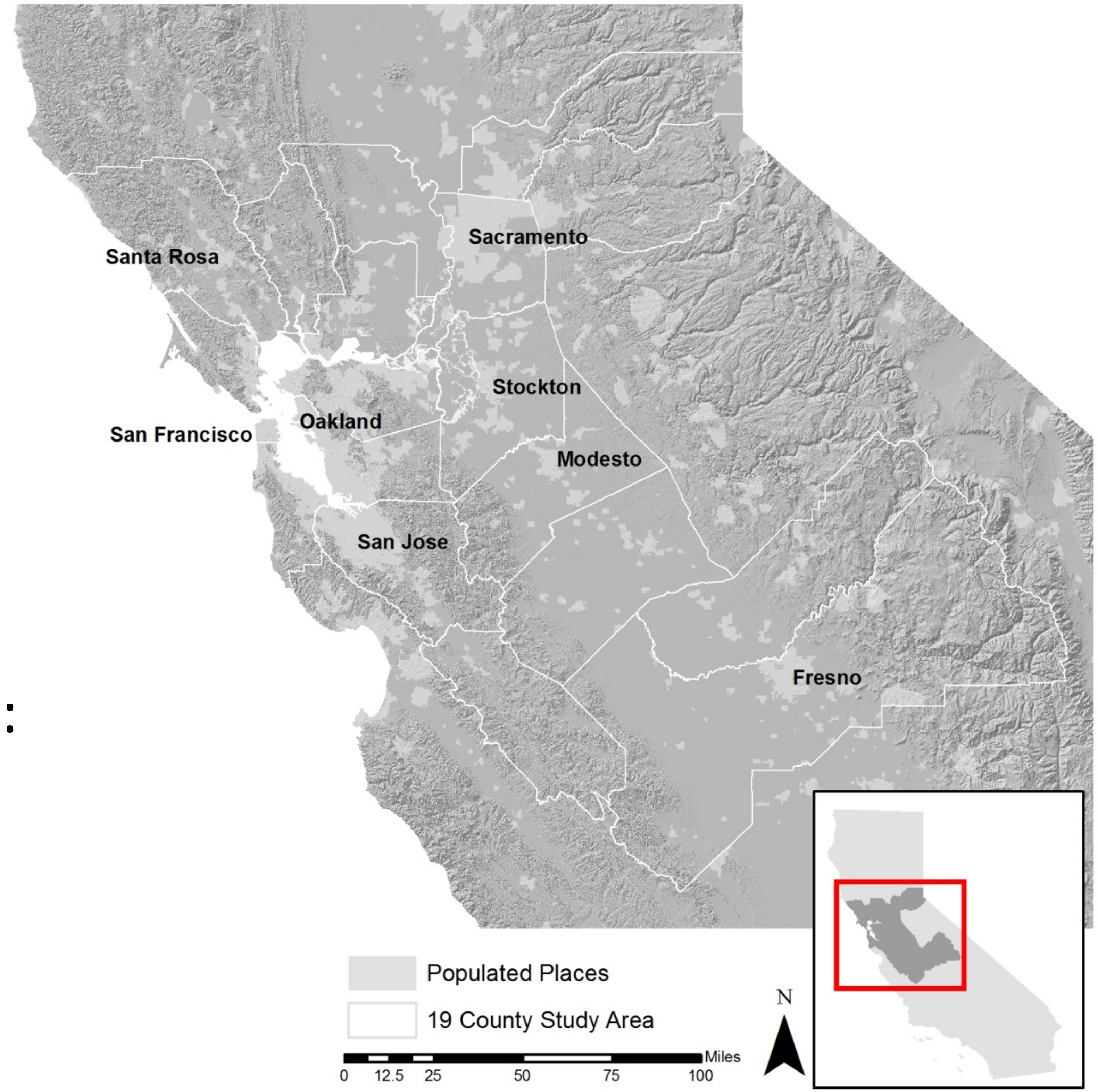
USDA Food Desert
Percentage of people with low access to a supermarket or large grocery store
0%
1% - 25%
26% - 50%
51% - 75%
76% - 100%
our space
0
0.5
1
2
3
Miles
N

Questions that can be asked using these data

- Are there patterns in my data? Are people, phenomena clustered?
- Are there local influences on health outcomes?
- How might mobility & changes in environment through time impact health outcomes?
- How can we best visualize the health landscape?

Study Area

- 19 county Kaiser Permanente Northern California service area
- Encompasses a diversity of environment types ranging from:
  - Urban to suburban to rural
  - Coastal to inland
- 2010 population: ~12 million
- Major metropolitan statistical areas:
  - San Jose-San Francisco-Oakland
  - Sacramento-Arden Arcade-Yuba City
  - Fresno-Madera



School Neighborhoods

IGIS has been working with NPI to look at social amenities in the areas surrounding schools.

Local Ground helps you tell the story of Your World

Story Maps
Create story maps that take your visitor on a journey through your data

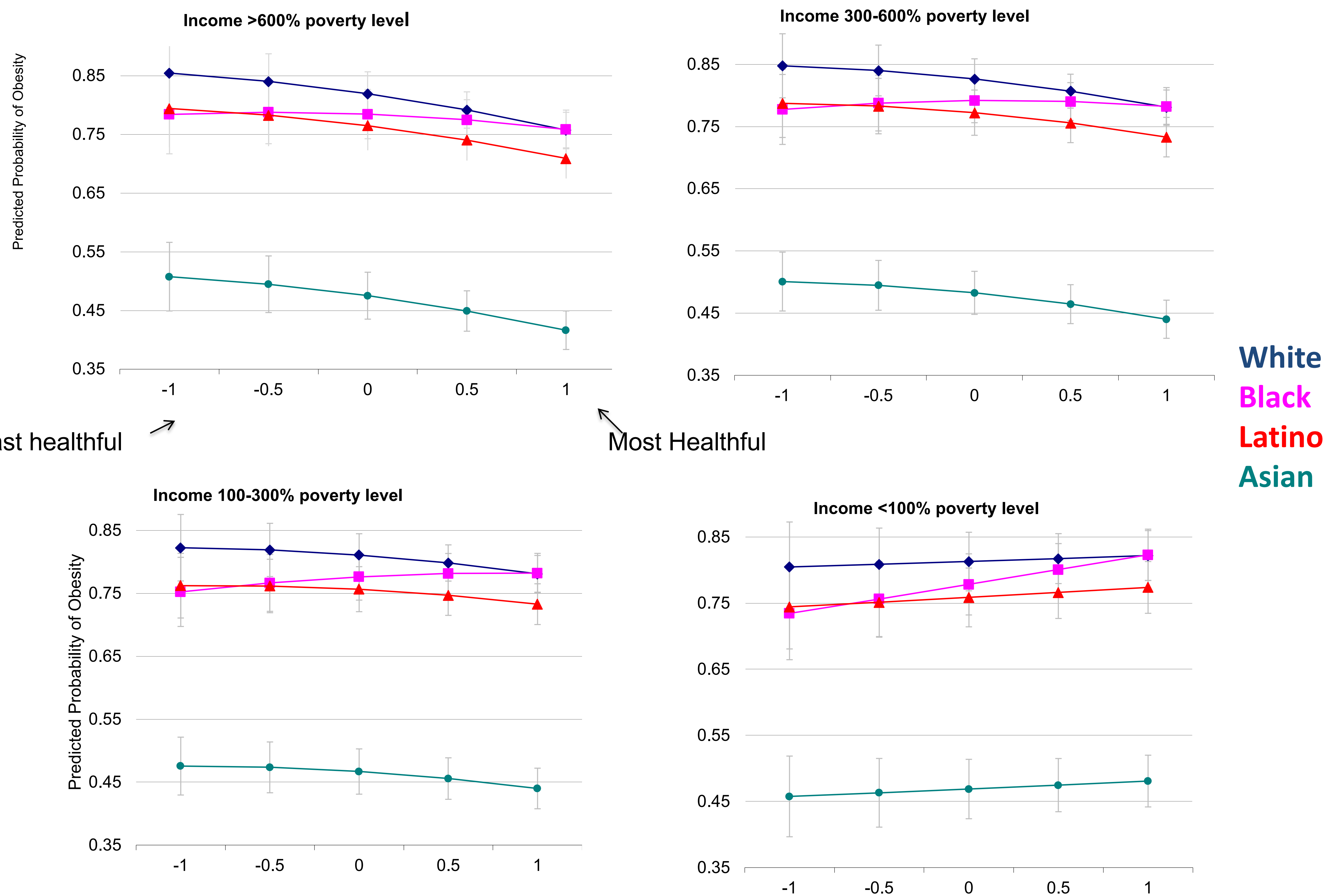
Data Visualization
Visualize your spatial data

Hand Drawn Maps
Print your map, draw on it with a colored pen, and upload it to automatically overlay your hand drawings on your map

http://localground.org/

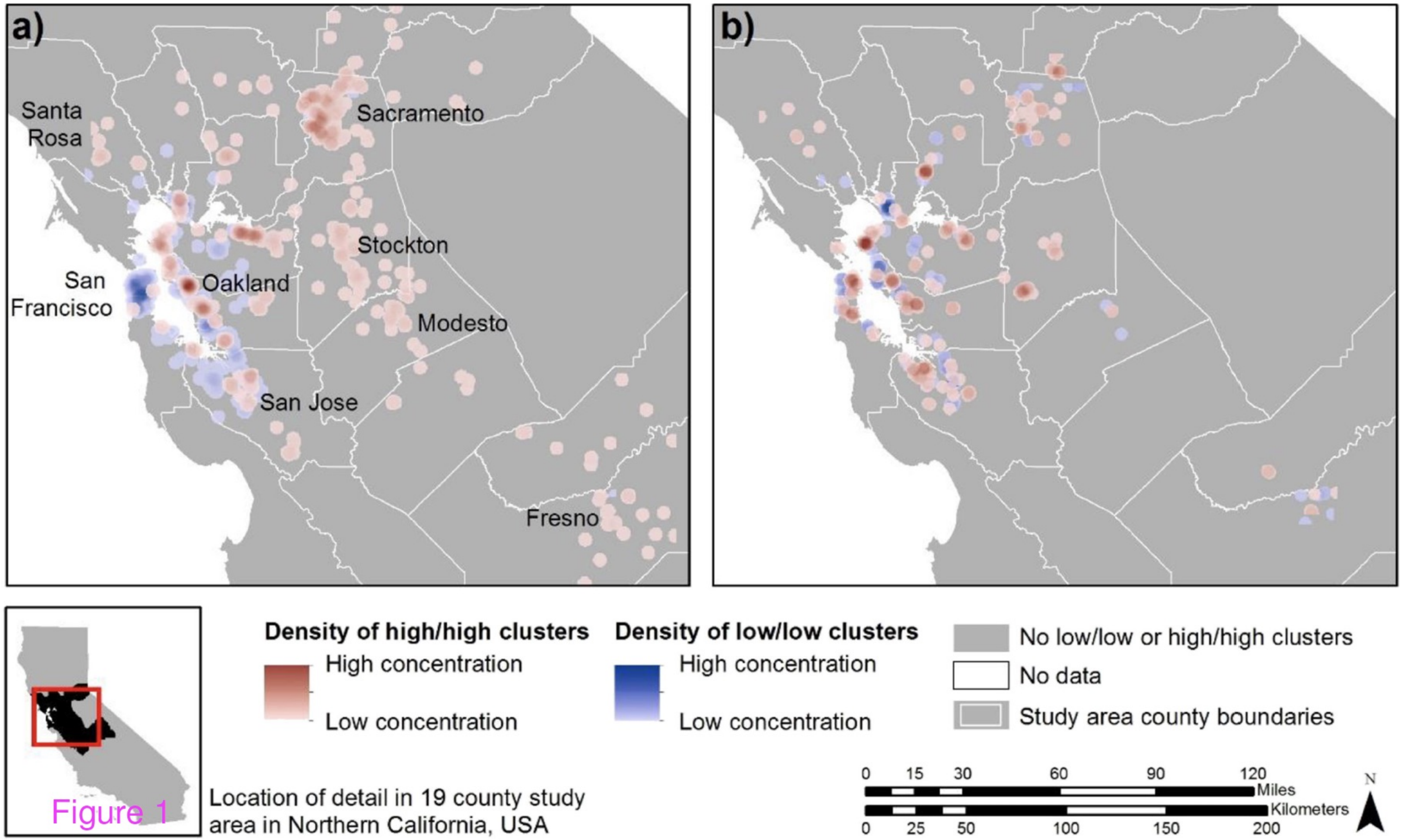
Difference in Kernel Density of BMI-Healthful vs. BMI-Unhealthful Food Venues

Examining associations between food environments and adult obesity, individual income & race/ethnicity.



Jones-Smith, J. C., M. Wharton, M. Kelly, E. Kersten, A. Karter, N. Adler, D. Schillinger, H. Moffett, and B. A. Laraia. 2013. Obesity and the food environment: income and ethnicity differences among people with diabetes, the Diabetes Study of Northern California (DISTANCE). *Diabetes Care* 36: 2697-2705

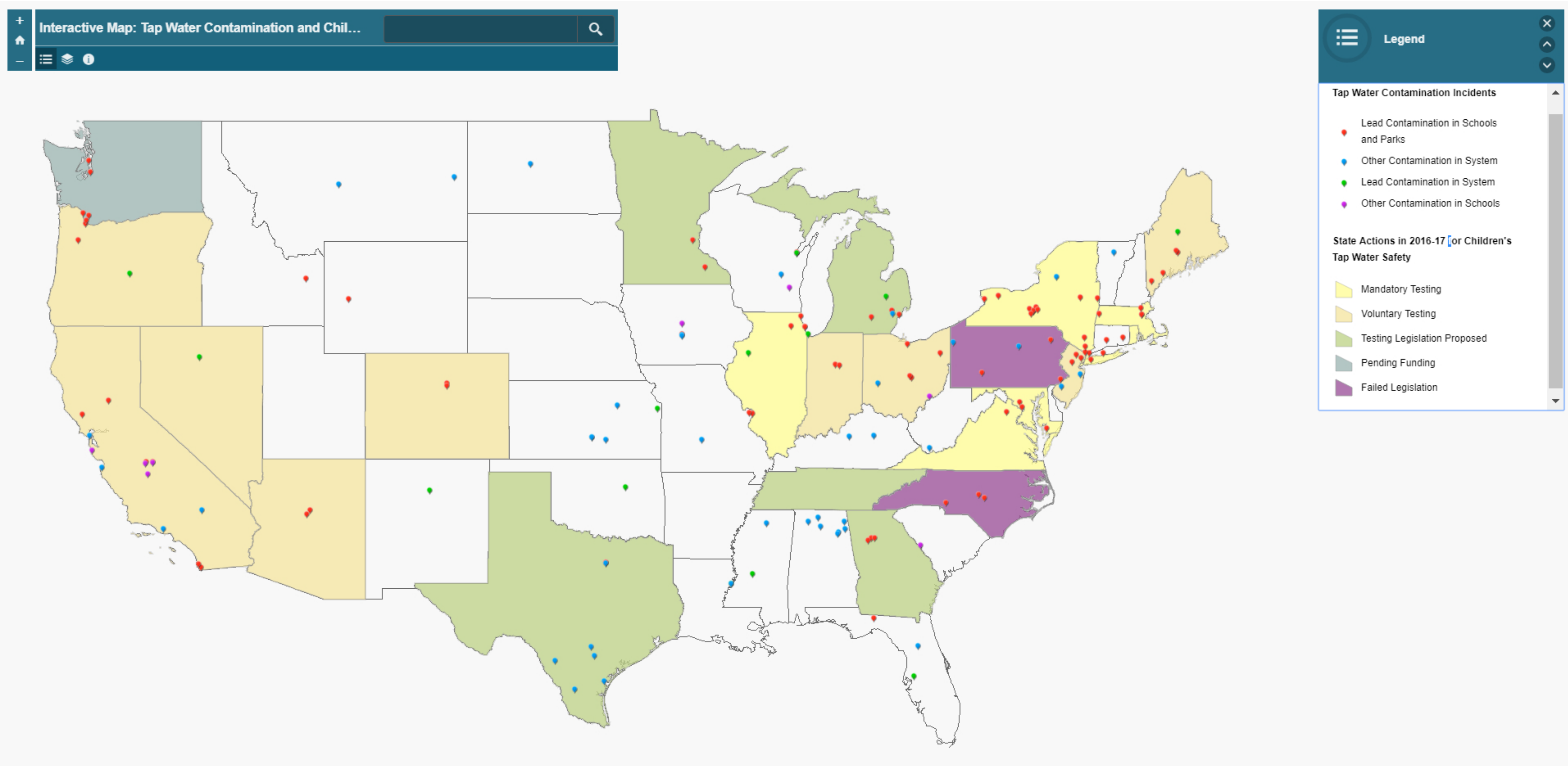
Significant level of clustering of BMI values among adults with diabetes



Examined individual demographic, socioeconomic factors, neighborhood & contextual factors. Found individual choices, conditions, and preferences may play a strong role in how individuals select the neighborhoods in which they live.

Laraia, B.A., S.D. Blanchard, A.J. Karter, J.C. Jones-Smith, E.M. Warton, E. Kersten, M. Jerrett, H.H. Moffett, N. Adler, D. Schillinger and M. Kelly. 2014. Spatial pattern of Body Mass Index among adults in the Diabetes Study of Northern California (DISTANCE). *International Journal of Health Geographics* 13:48

NPI Tap Water Contamination Project



http://www.drinkingwateralliance.org/new-map

Resources for You

http://igis.ucanr.edu/

IGIS is ANR's leader for geospatial knowledge and innovation, meeting the growing demand for spatial tools, data, training, and support across the ANR continuum

workshops **software**

data **office hours**

tools & technology

...+dronecamp! **June 2018**

The 21st Century Mapping Toolkit: Spatial Data Science

people **data** **tools**

The Spatial Data Science framework is *geographically-based*, *computing-intensive, data-rich, reproducible*, and *collaboration-focused*.

We need to:

- compile disparate **data from multiple sources;**
- use **easily available and open technology** for reproducible data analysis, sharing, and publication;
- apply **core spatial analysis concepts and methods;**
- use **machine learning, deep learning, data mining** to deal with giant spatial-temporal datasets;
- and utilize **visualization & interactive tools** to communicate with project managers, the public, and other stakeholders;

All in order to create **impactful** and **meaningful** solutions to our 21st century environmental challenges

What Do We Need?

- ***Training***: in technology, communication, and spatial concepts
- ***Funding***: for bridging technology, for linking groups, for critical natural resource challenges
- ***Coordination*** between academy, industry, and government for practical, impactful, solution-driven research to solve our natural-human challenges

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

GEOSPATIAL INNOVATION FACILITY
Cutting-Edge Mapping Technology at UC Berkeley

IGIS | University of California
Agriculture and Natural Resources | Informatics and GIS Statewide Program

Thank You

Maggi Kelly, Professor and CE Specialist, UC Berkeley

Faculty Director, Geospatial Innovation Facility (GIF)

Director ANR Statewide Program in Informatics and GIS (IGIS)

Email: maggi@berkeley.edu

Twitter: @nmaggikelly

http://kellylab.berkeley.edu

http://gif.berkeley.edu

http://igis.ucanr.edu/