Assessing Human Uncertainty to Inform Smarter Infrastructure Decisions by Leveraging Big Data

Newsha Ajami
Stanford Woods Institute for the Environment
Bill Lane Center for the American West
Stanford University

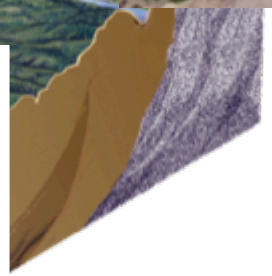
GEWEX Scientific Steering Committee
Pasadena, California
January 29, 2020

We live in a world defined by 20th century infrastructure, and 21st century water needs and challenges.

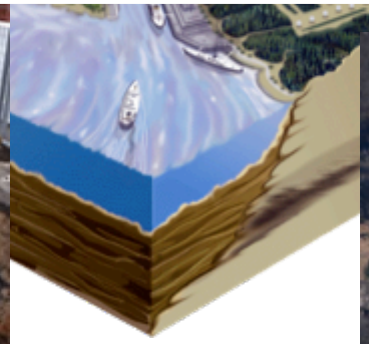
20th Century Centralized Water Infrastructure Model





- *Once through systems*
- Based on ***abundance*** and hydrologic ***stationarity***
- ***Top-down governance structure***


The Washington Post
Democracy Dies in Darkness

World • Analysis

After years of drought, Cape Town is set to run out of water

ronmental

Climate change increasing probability of extreme events (e.g. floods, droughts)

Middle East

In Iran, parched lands hollowed by water pumping now sinking

ue to lack of water

Search jobs | Sign in | Search | US edition

The Guardian

Opinion
Drought

The only thing as certain as drought in Australia is the stupid call to build new dams

THE WALL STREET JOURNAL.

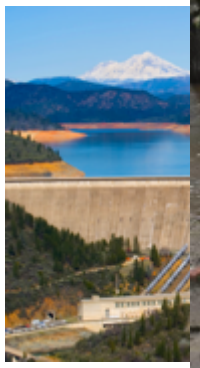
California Orders Unprecedented, Mandatory

ence

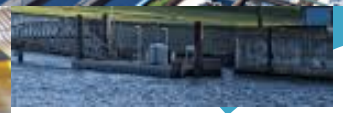
Los Angeles Times

How bad is the drought? Here are some sobering answers.

Transitioning to the 21st Century Infrastructure Model





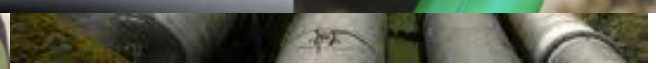


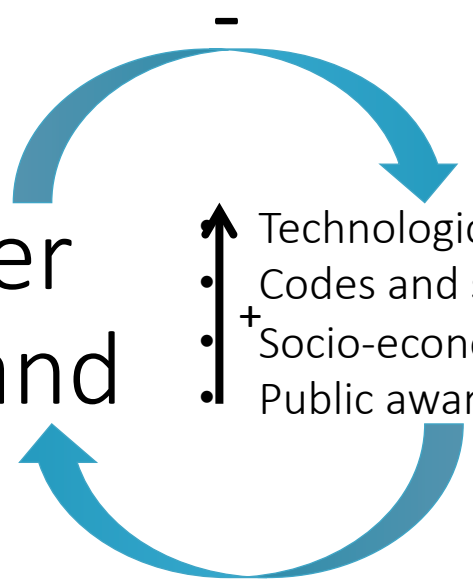
Water

emand

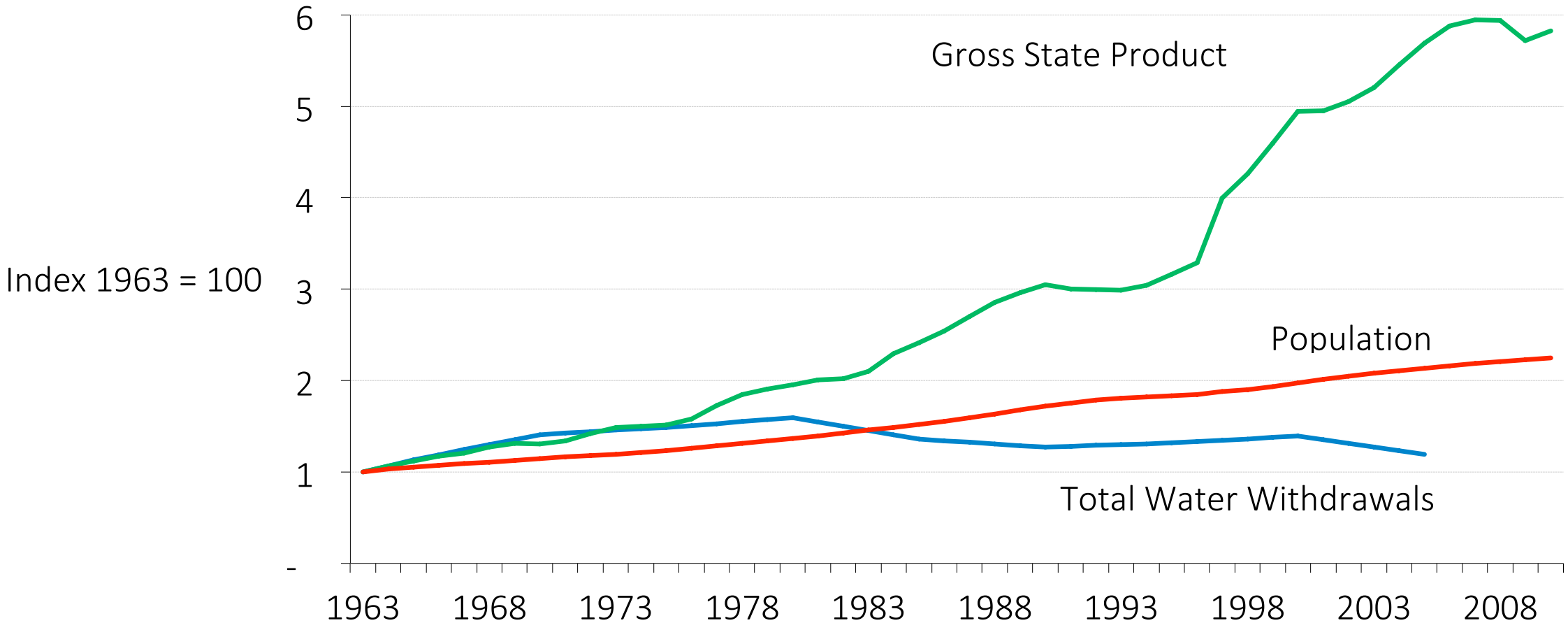
Water





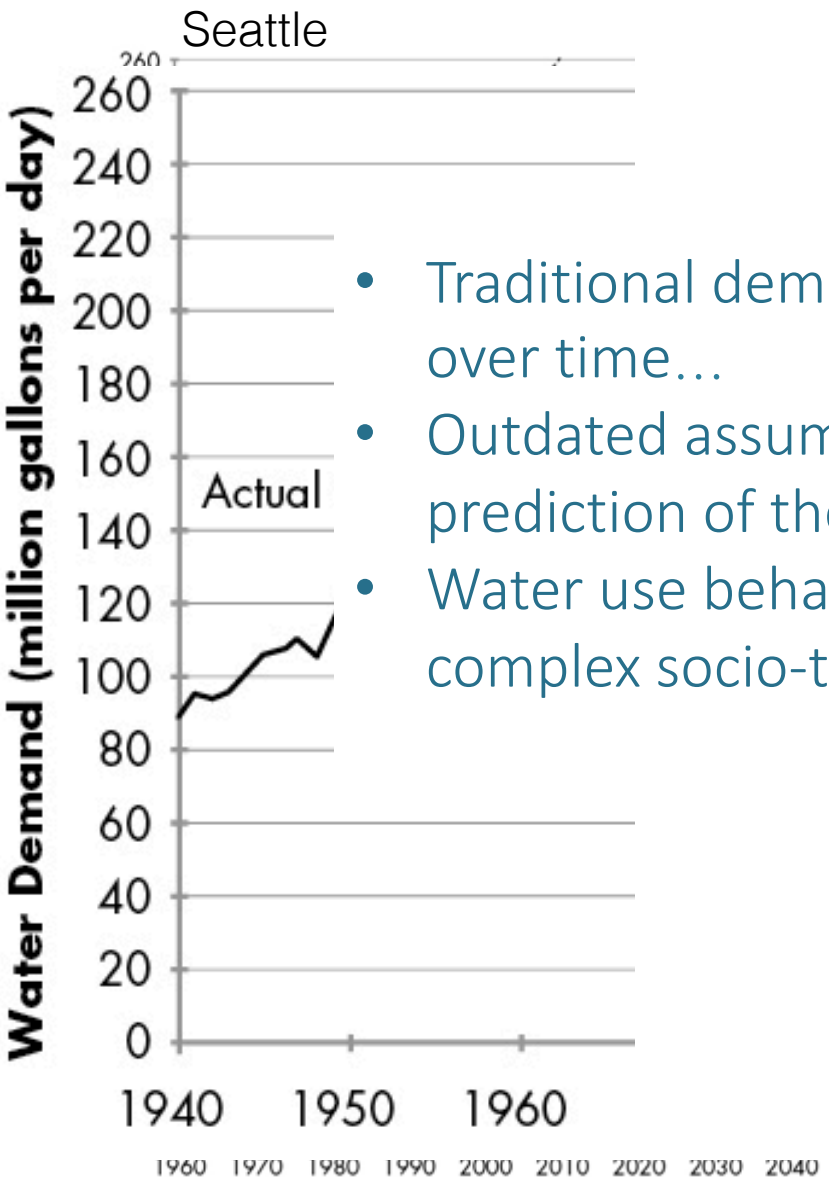
Water Demand and Population growth have decoupled in recent decades: California



Source: Hanak et al. 2012

Figure 1: California Economy, Population, and Water Use

Demand forecasting is the foundation of planning and decision-making

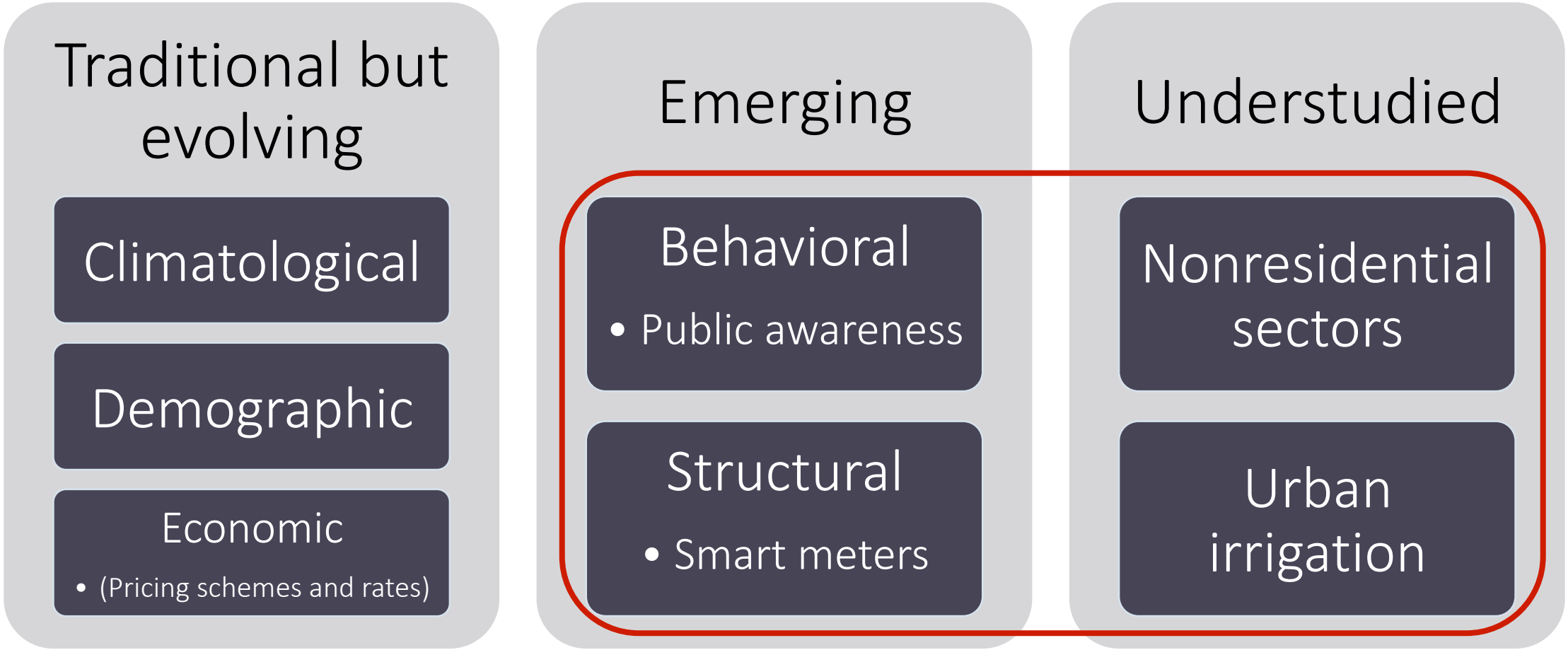


- Traditional demand models assume static conditions over time…
- Outdated assumption that the past is an accurate prediction of the future
- Water use behaviors are dynamic, and the result of complex socio-technical interactions.

1992 1996 2000 2004 2008 2012 2016 2020

Pacific Institute, 2016

Water use is complex



How can we harness
big data
to assess
evolving water demand
dynamics?

What is big data?

A stunning 90% of the data created by humanity has been generated in just the past two years.

- Volume (lots of data)
- Resolution (spatial and temporal)
- Variety (many different sources)
- Computationally intensive

Two Applications of Big Data in Water

Generating new data

Assessing water demand uncertainty and social memory

Combining high resolution data

Large landscape irrigation conservation behavior

Applications #1: Web scraping and media data

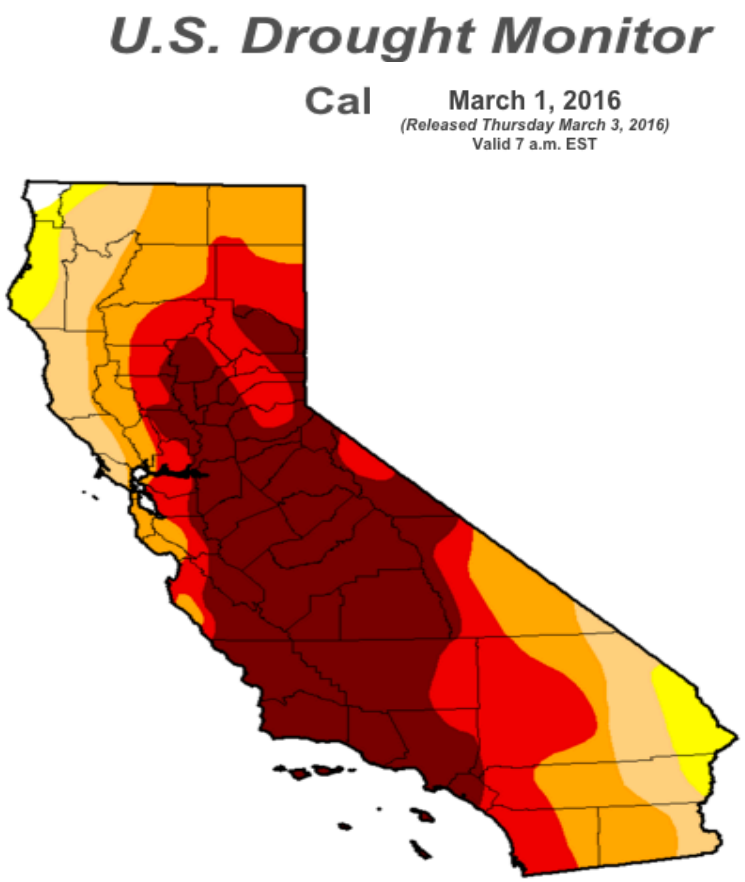
Data sources

Assessing water demand uncertainty and social memory

Single Family Residential Sector

The 2012-2016 California drought was unprecedented,

Hydrologically



Media Coverage

THE WALL STREET JOURNAL.

California Orders Unprecedented, Mandatory Water Cuts

THE SACRAMENTO BEE

Jerry Brown declares California drought emergency, urges 20 percent cut in water use

Los Angeles Times

How bad is the drought? Here are some sobering answers.

Articulate: a new search algorithm to measure news media coverage

Existing tools

Proprietary, inflexible, data-limited

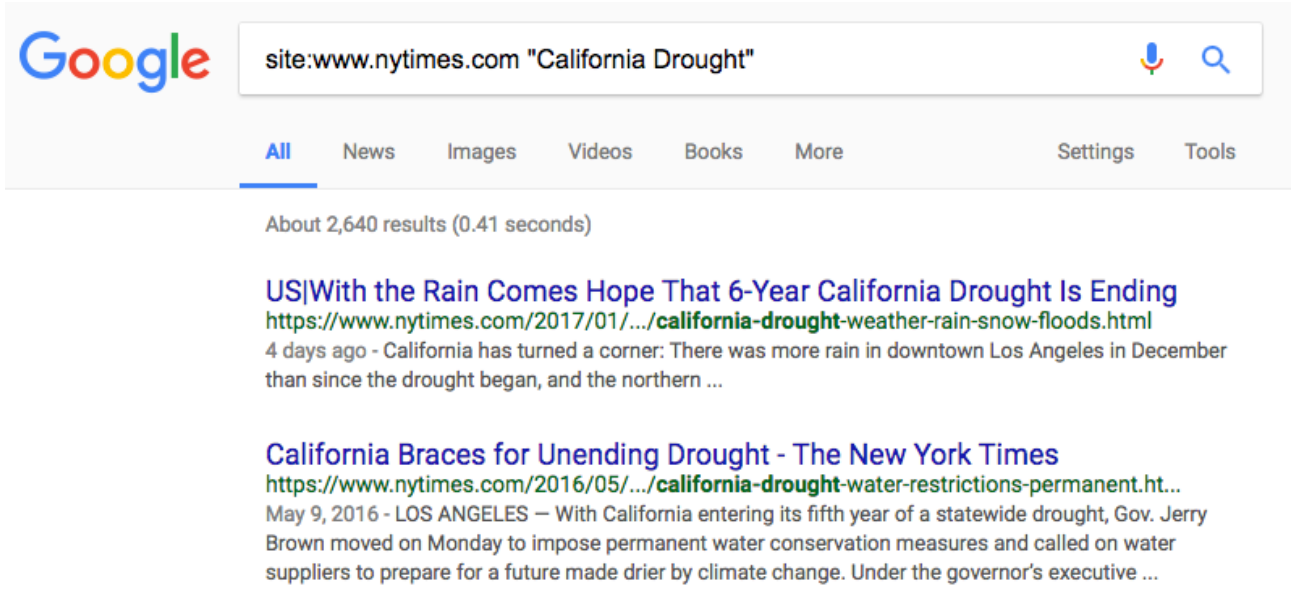
ProQuest

NewsBank inc.

LexisNexis®

Articulate

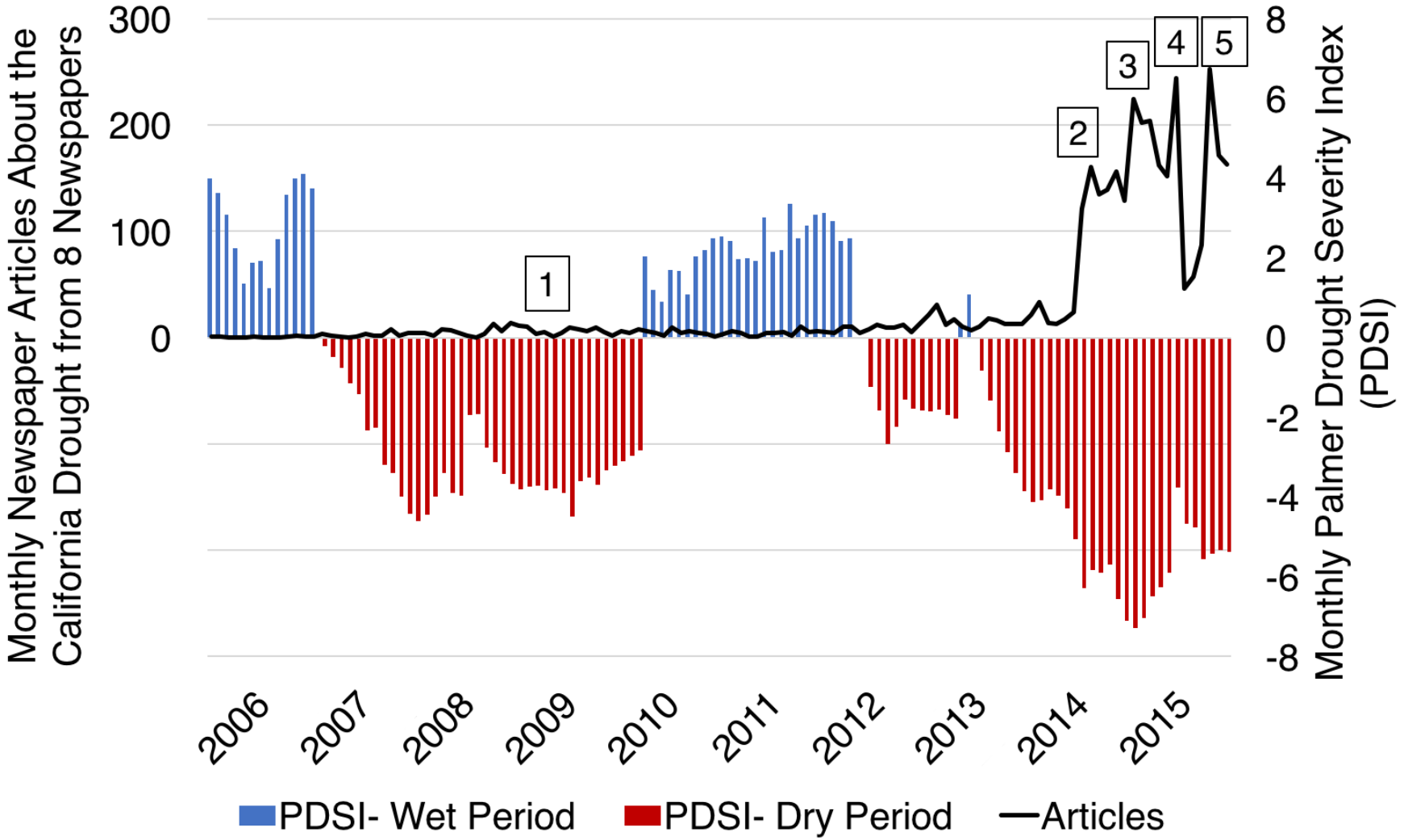
Flexible and open-source, written in Python, interacts with Google Custom Search Engine API with user specified inputs and outputs



(Roby, Gonzales, Quesnel, and Ajami , *Environmental Modeling and Software,* 2018)

A Tale of Two Droughts

Drought-related coverage was only during recent drought and spiked during political/climatic events:



1 Feb. 2009
Governor declares drought state of emergency

2 Jan. 2014
Governor declares drought state of emergency

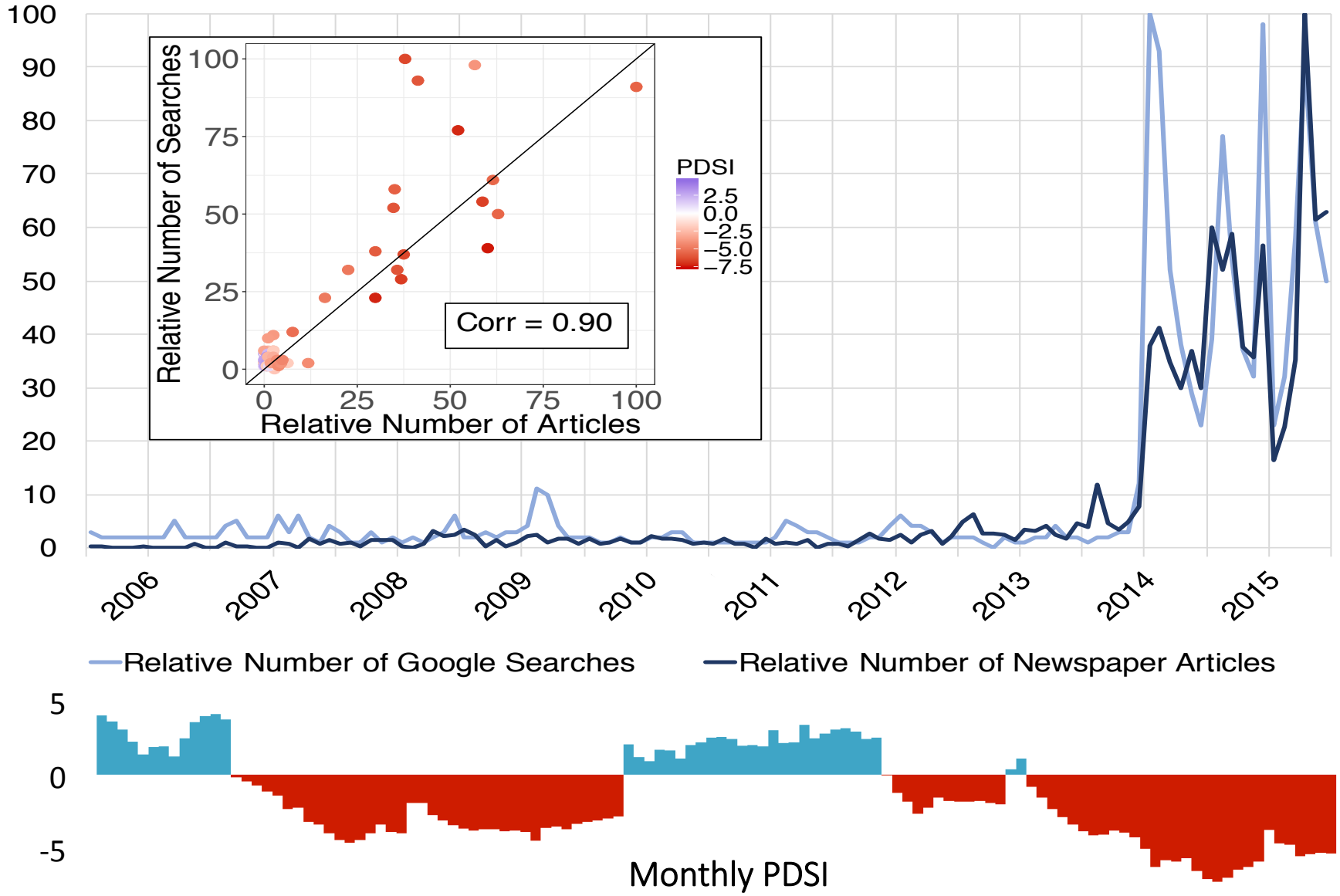
3 July/Aug. 2014
Mandatory outdoor water conservation regulation

4 December 2014
Rain Event

5 April 2015
Mandatory statewide water use restrictions

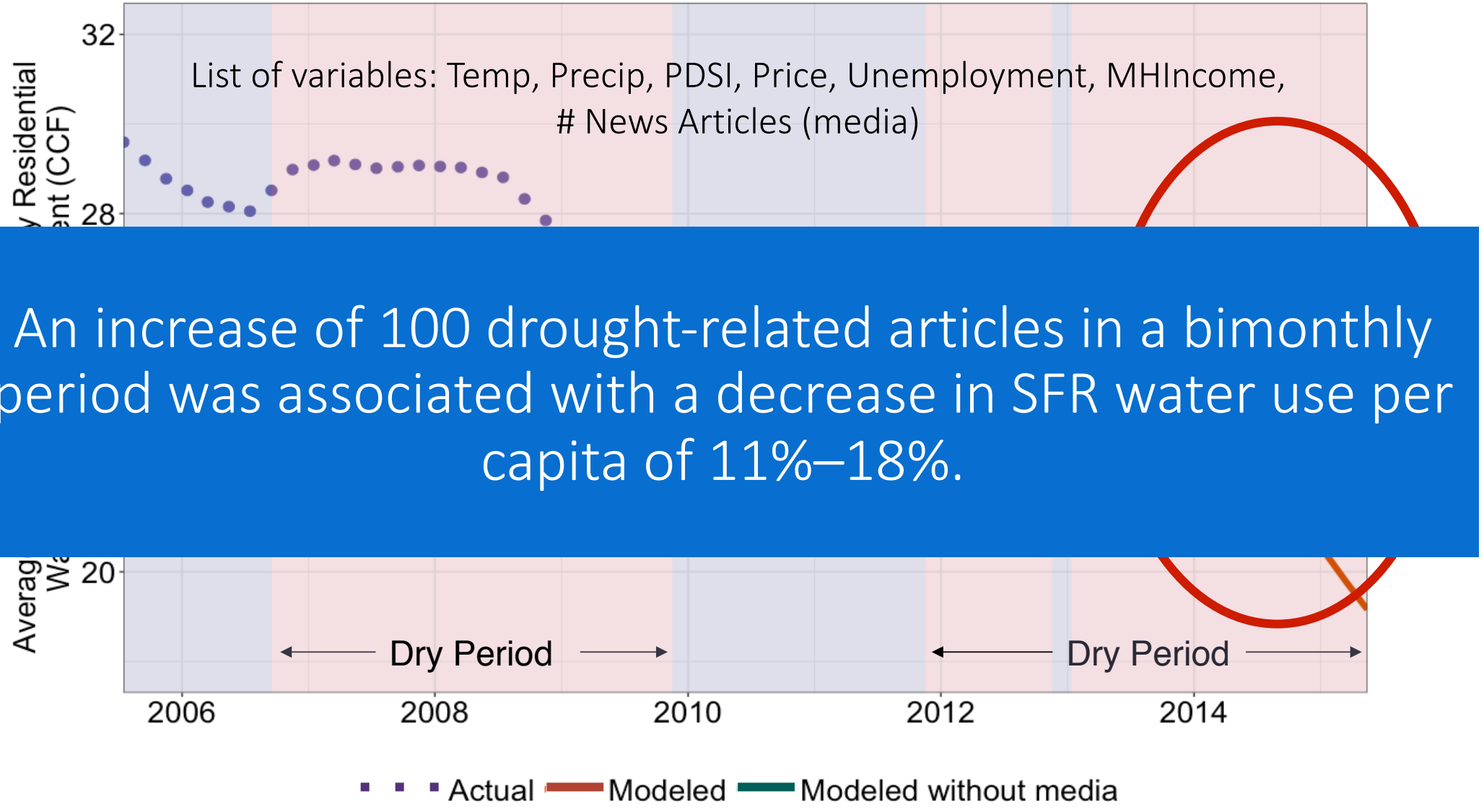
(Roby, Gonzales, Quesnel, and Ajami , *Environmental Modeling and Software,* 2018)

Media coverage and internet search trends are highly correlated:



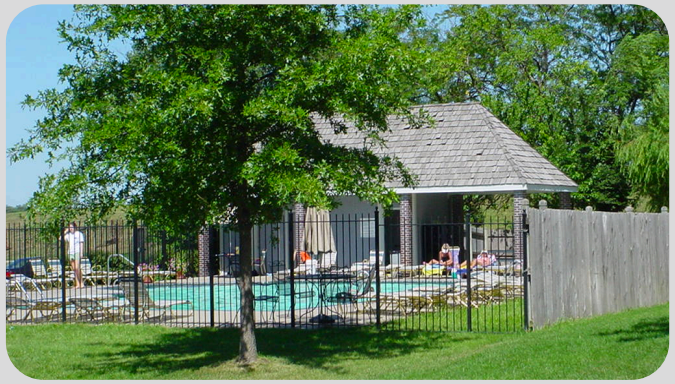
(Quesnel and Ajami, *Science Advances,* 2017)

A counterfactual scenario shows that media has a significant signal:



(Quesnel and Ajami, *Science Advances,* 2017)

Applications #2: Matching various data sources



Large landscape irrigation conservation behavior

The American Obsession with Lawns

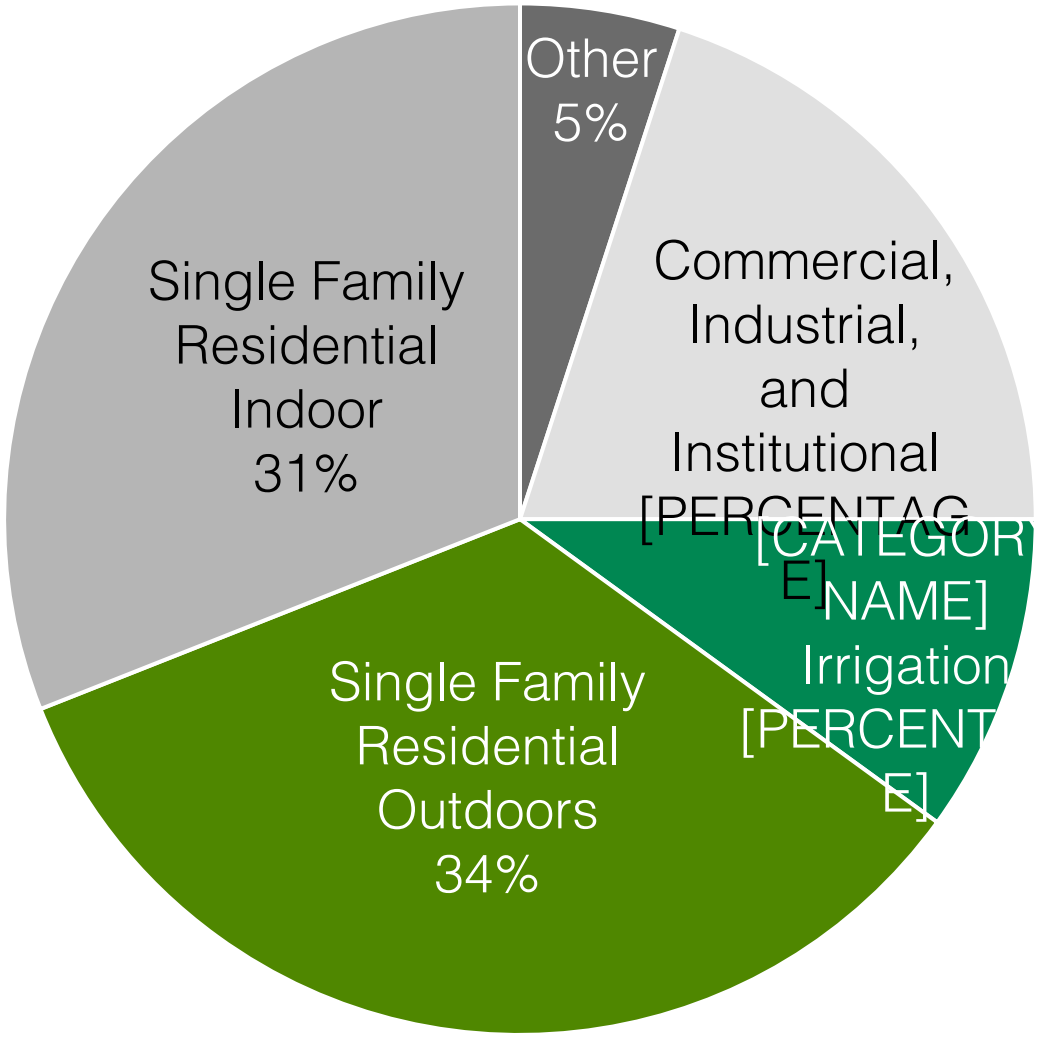
Lawns are the most grown crop in the U.S.—and they're not one that anyone can eat; their primary purpose is to make us look and feel good about ourselves

By Krystal D'Costa on May 3, 2017

SCIENTIFIC AMERICAN.

Source: Milesi et al (2005) *Env. Mgmt*

California's urban water use



Almost half of California's urban wat demand is for outdoo use

Source: 2016 Urban Water Efficiency, California Water Plan, CA DWR

Measurement technologies

Large landscape irrigation conservation behavior

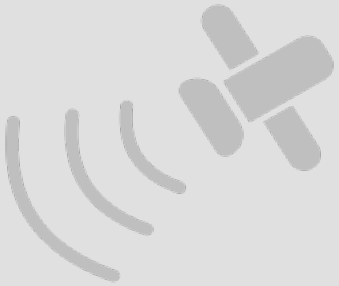
Measurement technologies

Irrigation and vegetation health connections during drought

Nonresidential Irrigation Sector

Measurement technologies

Large landscape irrigation conservation behavior

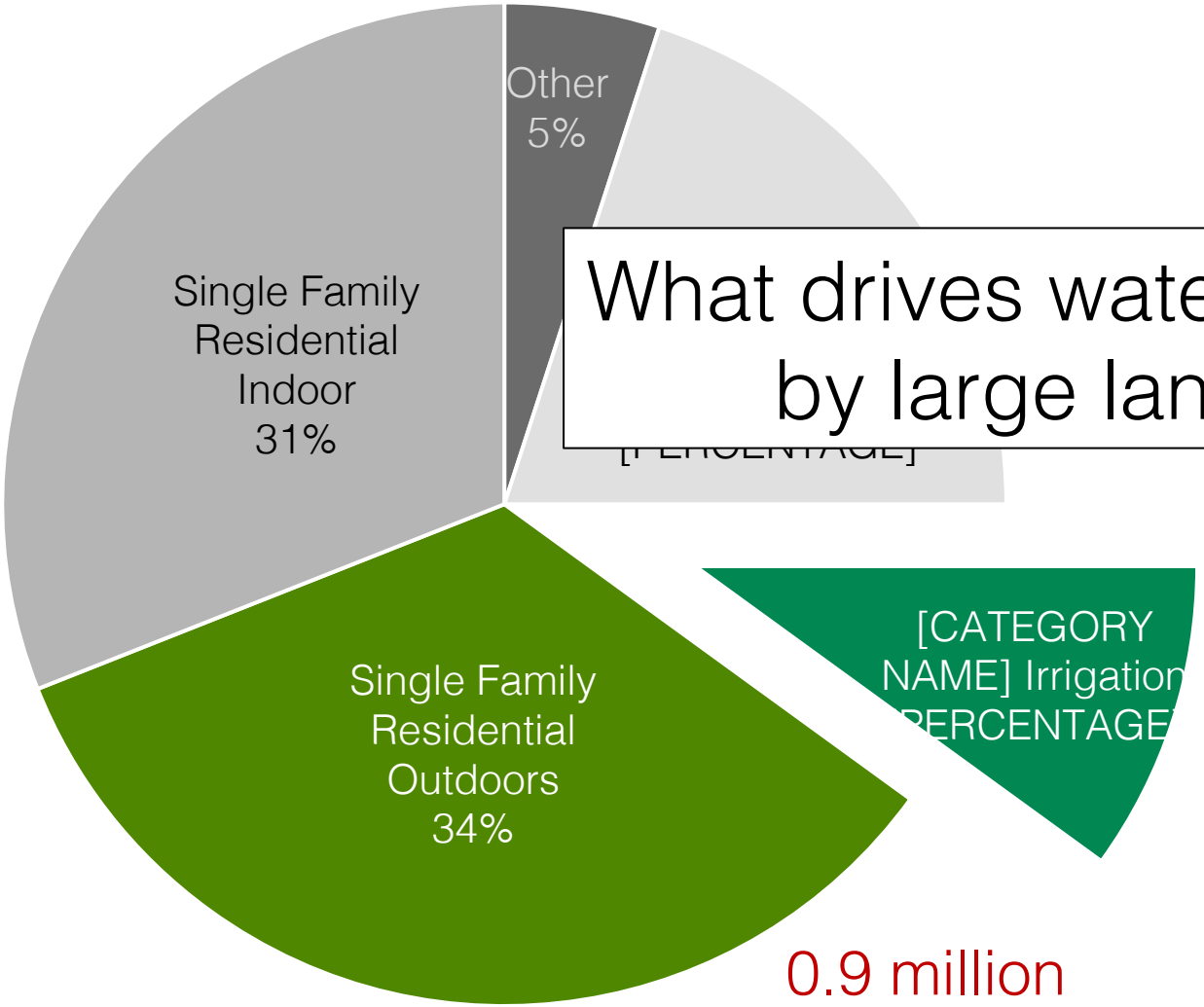


Measurement technologies

Irrigation and vegetation health connections during drought

Nonresidential Irrigation Sector

California's urban water use



0.9 million
acre-feet per year

What drives water use and conservation by large landscape irrigators?

Commercial, Industrial, Institutional

Multifamily Residential and Homeowners' Associations

Source: 2016 Urban Water Efficiency, California Water Plan, CA DWR

City of Redwood City

- 629 large landscape irrigators
- Potable or recycled water connections
- Commercial, industrial, institutional or multifamily residential

<u>Approach</u>

Evaluate heterogeneous water use and conservation behavior

Utilizing water data from dedicated outdoor AMI

Daily water use from smart meters

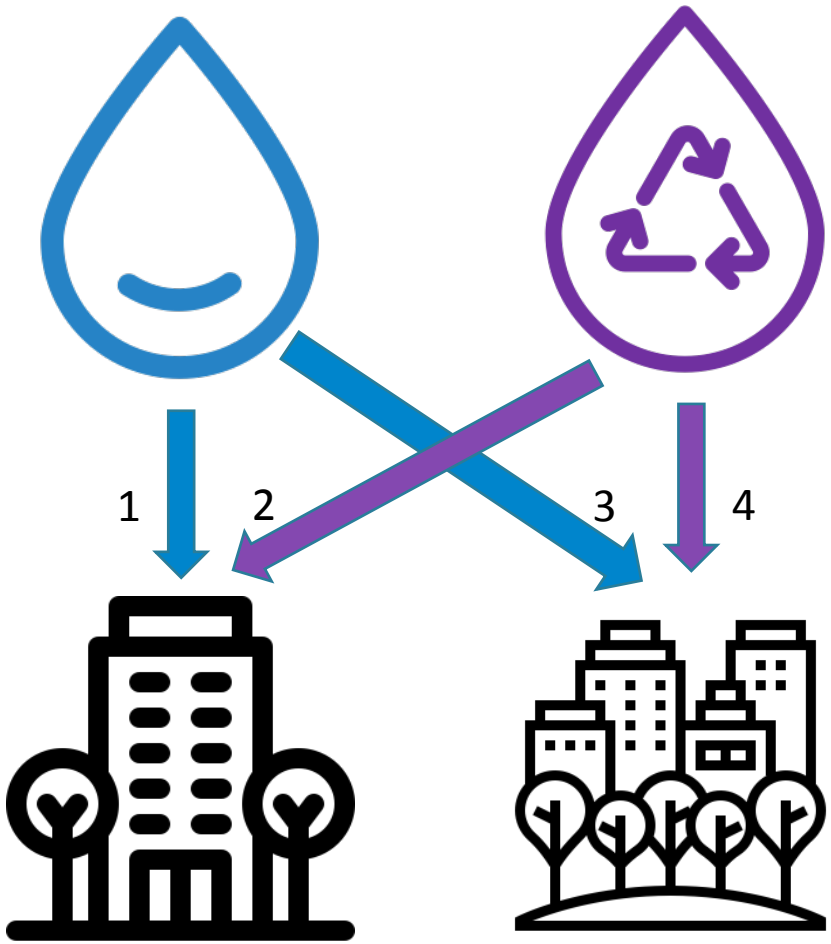
(1.7 million observations)

Data cleaning, processing, integration, aggregation

Data-driven analyses of **conservation** and weekly water use behavior

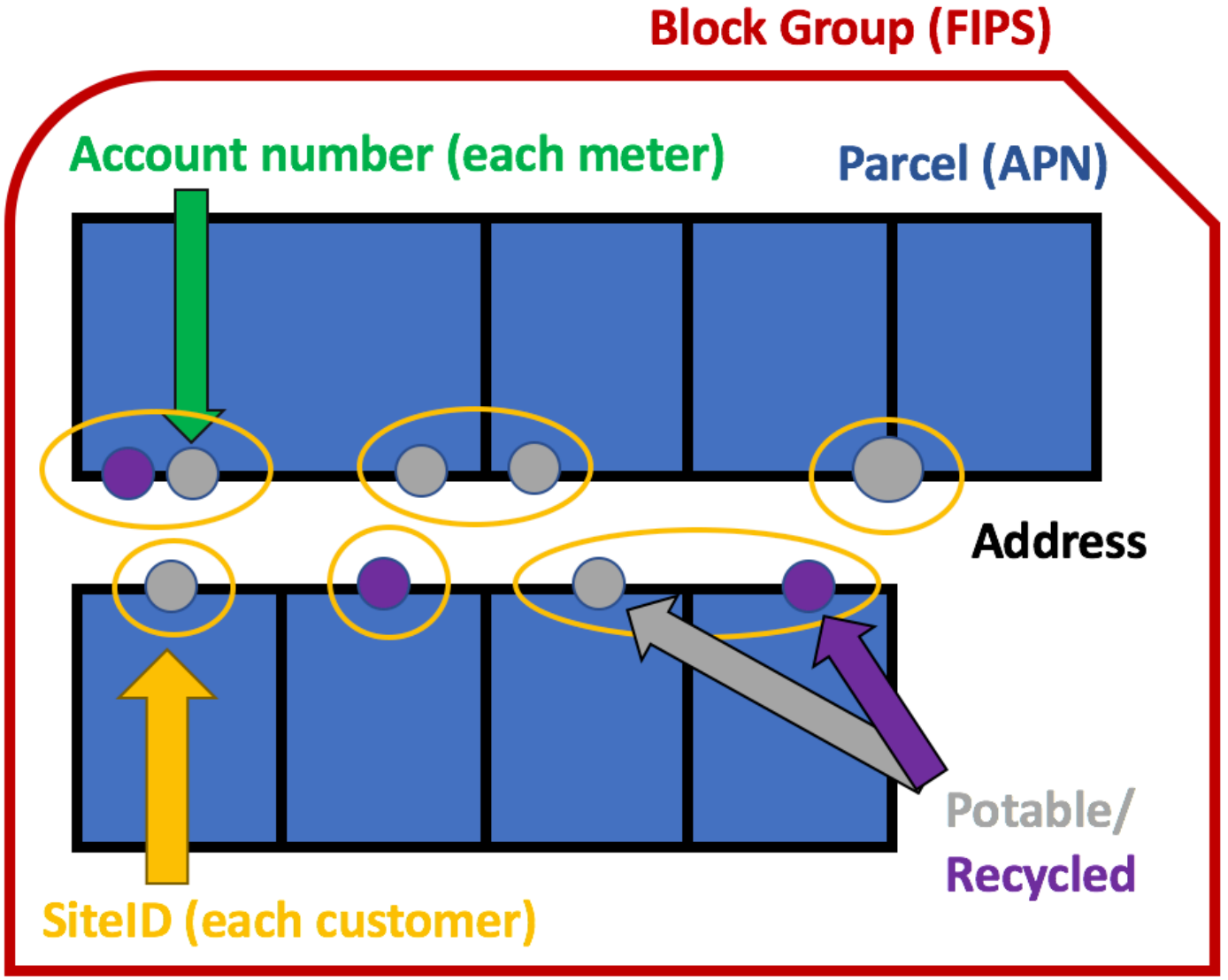
Customer heterogeneity

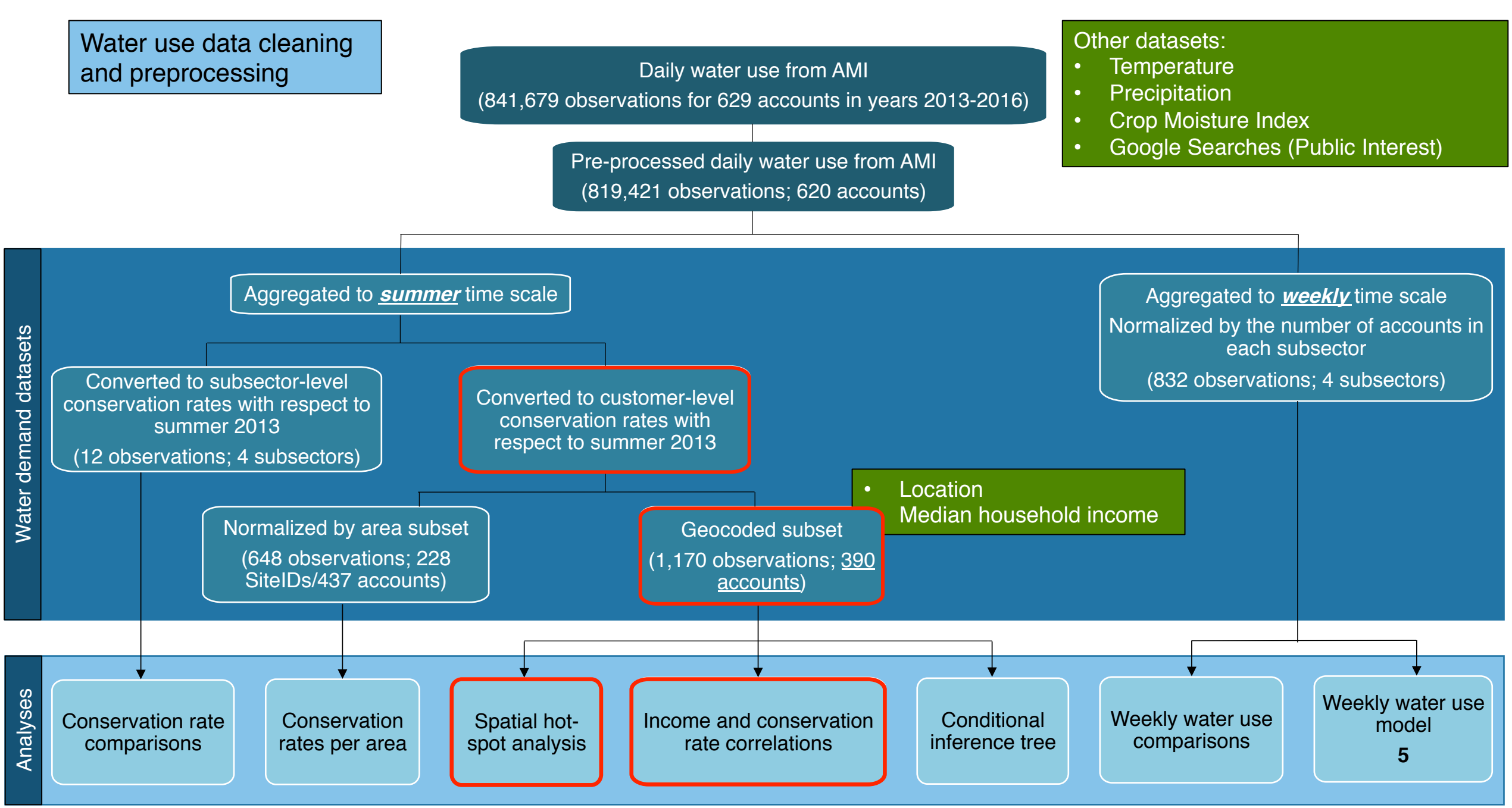
(4 subsectors)



Different identifiers are used in different datasets

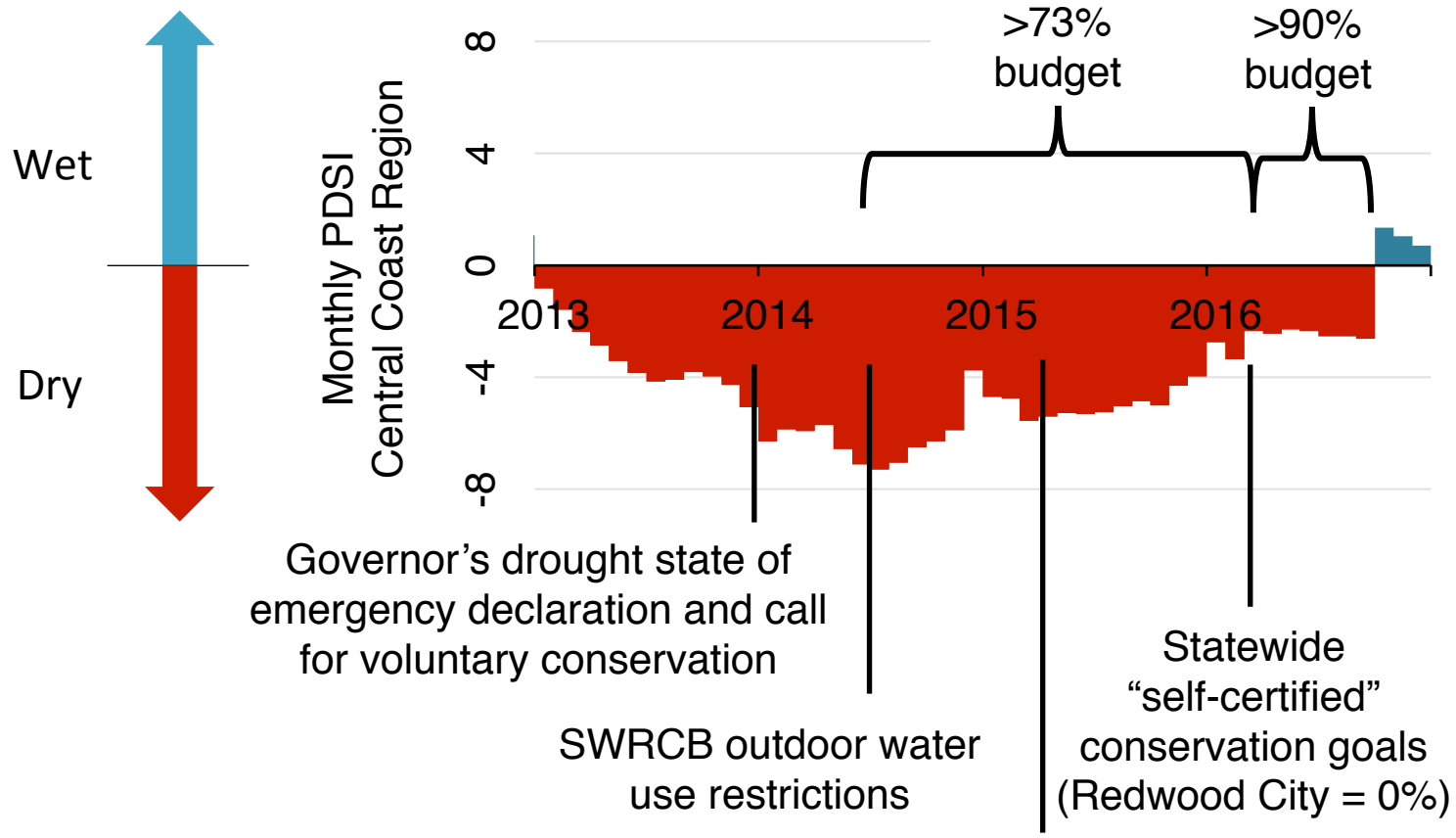
- Water use is by account number
- RWC Area and Budgets are by SiteID
- Location (RWC GIS File) is by account number
- Remote sensing is by parcel/account number combinations




Water use data cleaning and preprocessing
Daily water use from AMI
(841,679 observations for 629 accounts in years 2013-2016)
Other datasets:
• Temperature
• Precipitation
• Crop Moisture Index
• Google Searches (Public Interest)
Pre-processed daily water use from AMI
(819,421 observations; 620 accounts)
Water demand datasets
Aggregated to summer time scale
Aggregated to weekly time scale
Normalized by the number of accounts in each subsector
(832 observations; 4 subsectors)
Converted to subsector-level conservation rates with respect to summer 2013
(12 observations; 4 subsectors)
Converted to customer-level conservation rates with respect to summer 2013
• Location
Median household income
Normalized by area subset
(648 observations; 228 SiteIDs/437 accounts)
Geocoded subset
(1,170 observations; 390 accounts)
Analyses
Conservation rate comparisons
Conservation rates per area
Spatial hot-spot analysis
Income and conservation rate correlations
Conditional inference tree
Weekly water use comparisons
Weekly water use model
5

Evaluate 2013–2016 water use to capture changes during different policy periods

Redwood City policy context



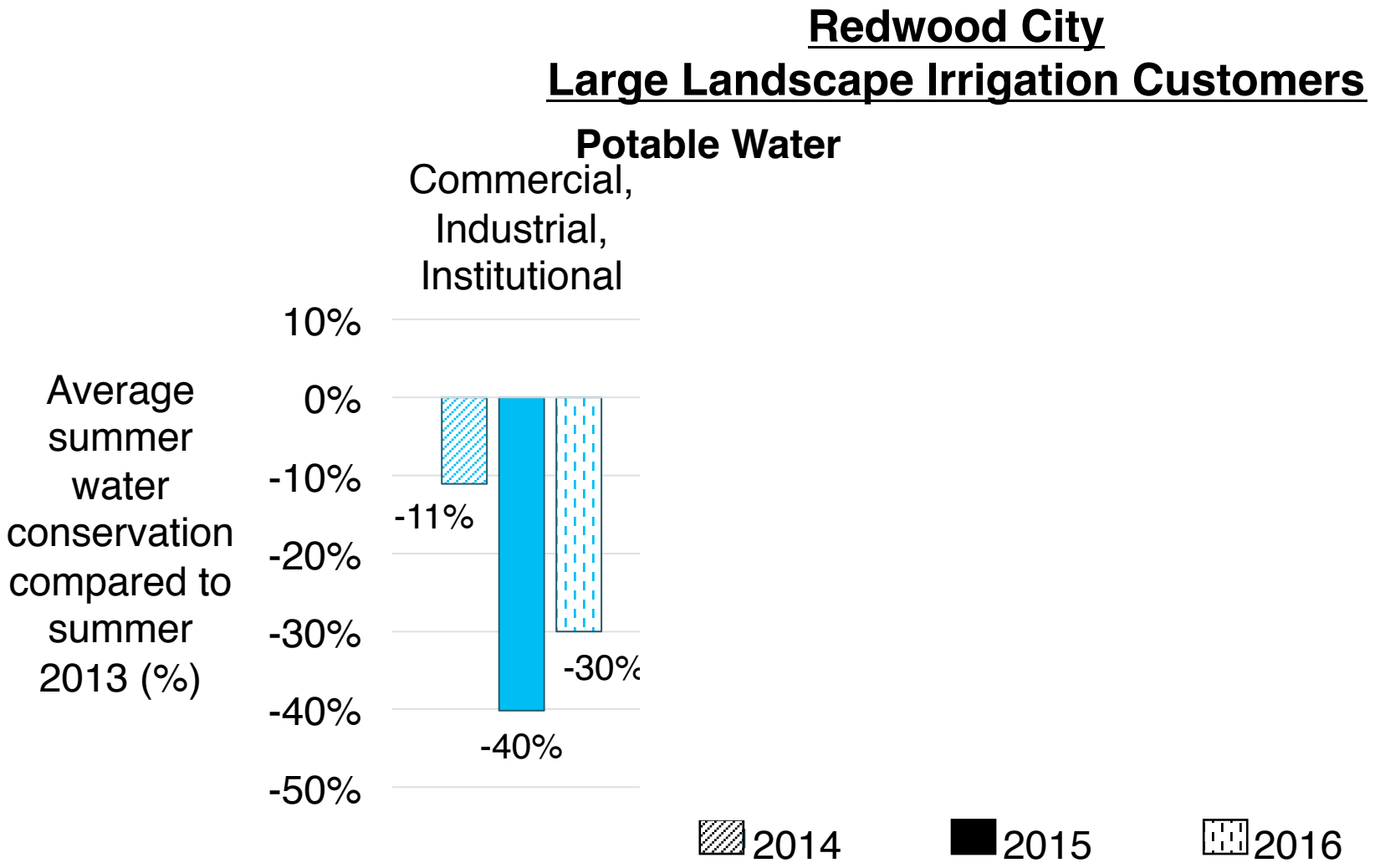
Potable water irrigators
subject to fines for going over their water-use budgets

Recycled water irrigators
did not face any conservation restrictions during the drought

California policy context

Redwood City
Non-residential irrigation customers conserved in parallel to California residents despite receiving different or no mandates

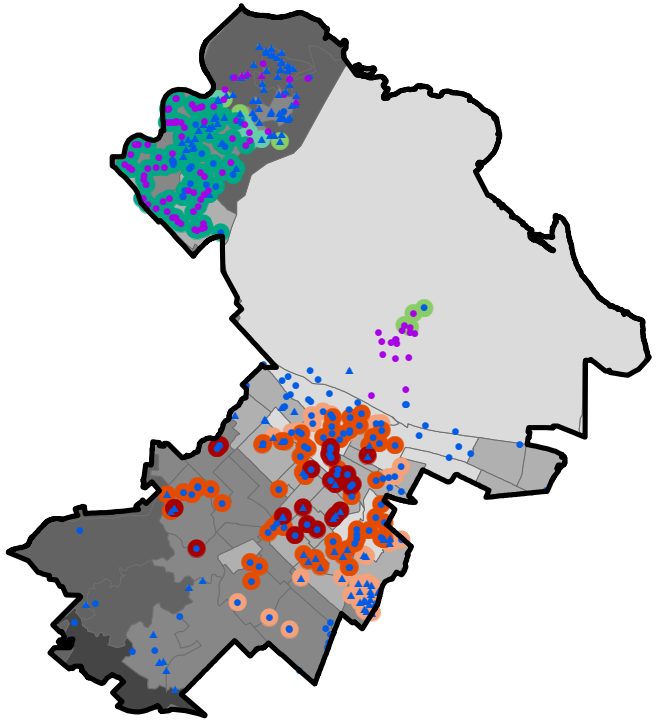
Potable and recycled irrigation conservation patterns



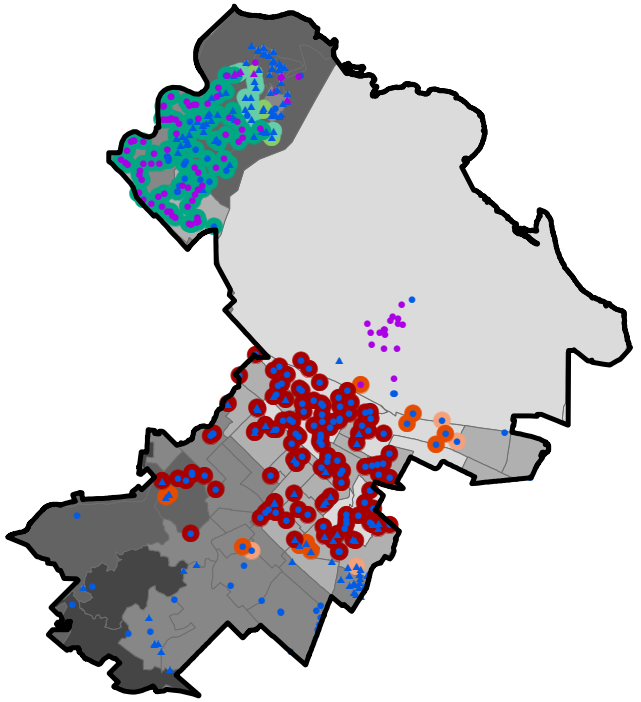
Quesnel and Ajami (2019) *Water Resources Research*

Conservation hotspots show importance of neighborhood norms

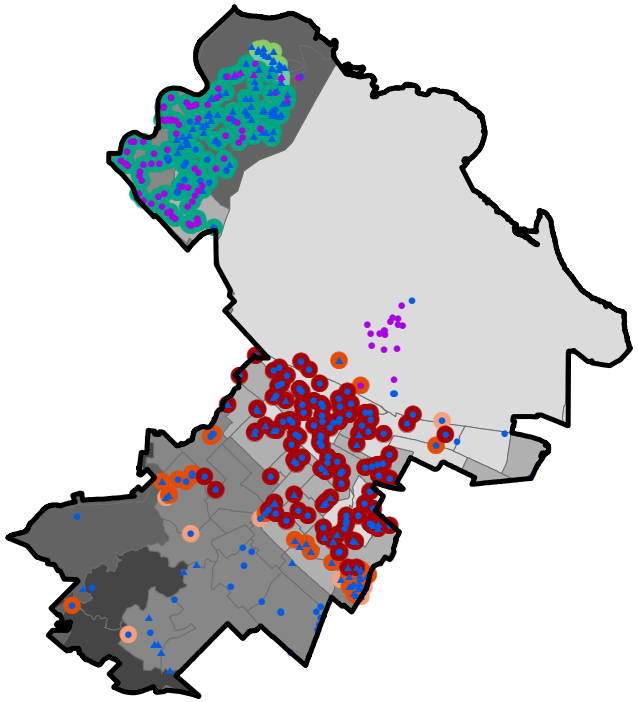
Summer 2014



Summer 2015



Summer 2016



Getis-Ord Gi* statistic

$$G_i^* = \sum_{j=1}^{n} w_{i,j} x_j - X \sum_{j=1}^{n} w_{i,j} \Big/ \sqrt{\Big[\sum_{j=1}^{n} x_j^2 / n - \Big(\sum_{j=1}^{n} x_j / n\Big)^2\Big] \sqrt{\Big[n \sum_{j=1}^{n} w$$

Conservation hot-spot

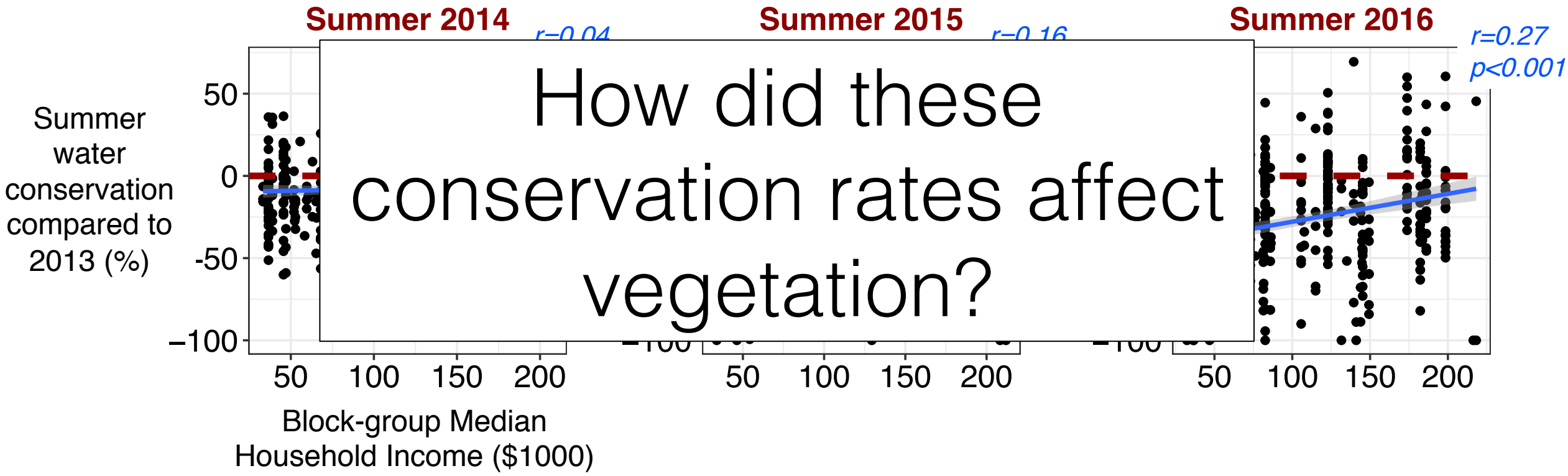
Customer with potable water

Conservation cold-spot

Customer with recycled water

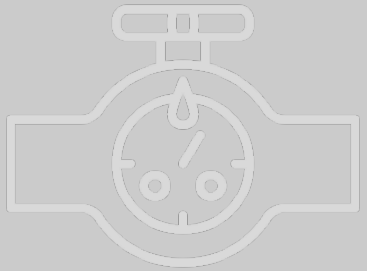
Quesnel and Ajami (2019) *Water Resources Research*

Correlation between nonresidential irrigation conservation and neighborhood affluence



How did these conservation rates affect vegetation?

Quesnel and Ajami (2019) *Water Resources Research*



Measurement technologies

Large landscape irrigation conservation behavior

Measurement technologies

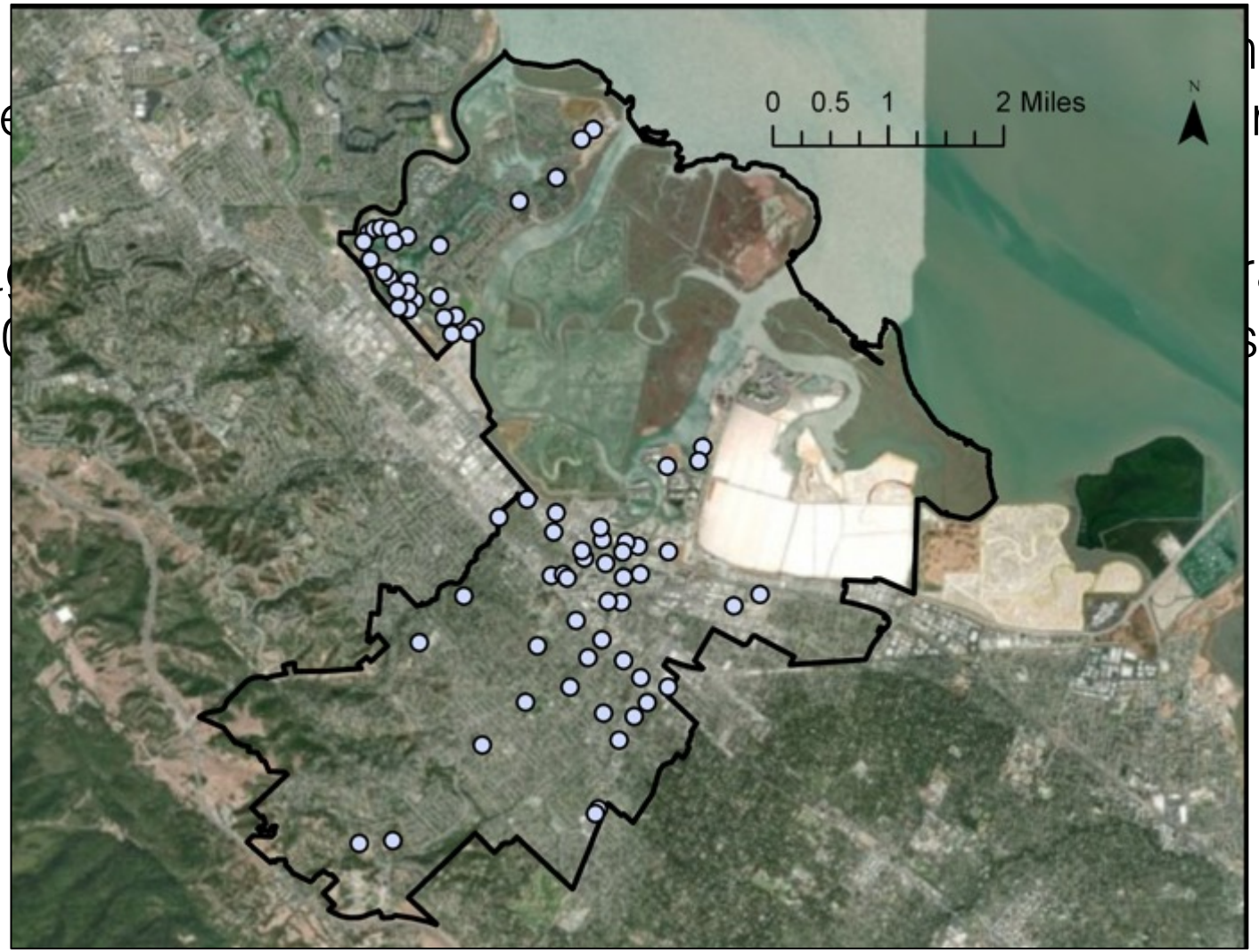
Irrigation and vegetation health connections during drought

Nonresidential Irrigation Sector

Linking water use and remote sensing data

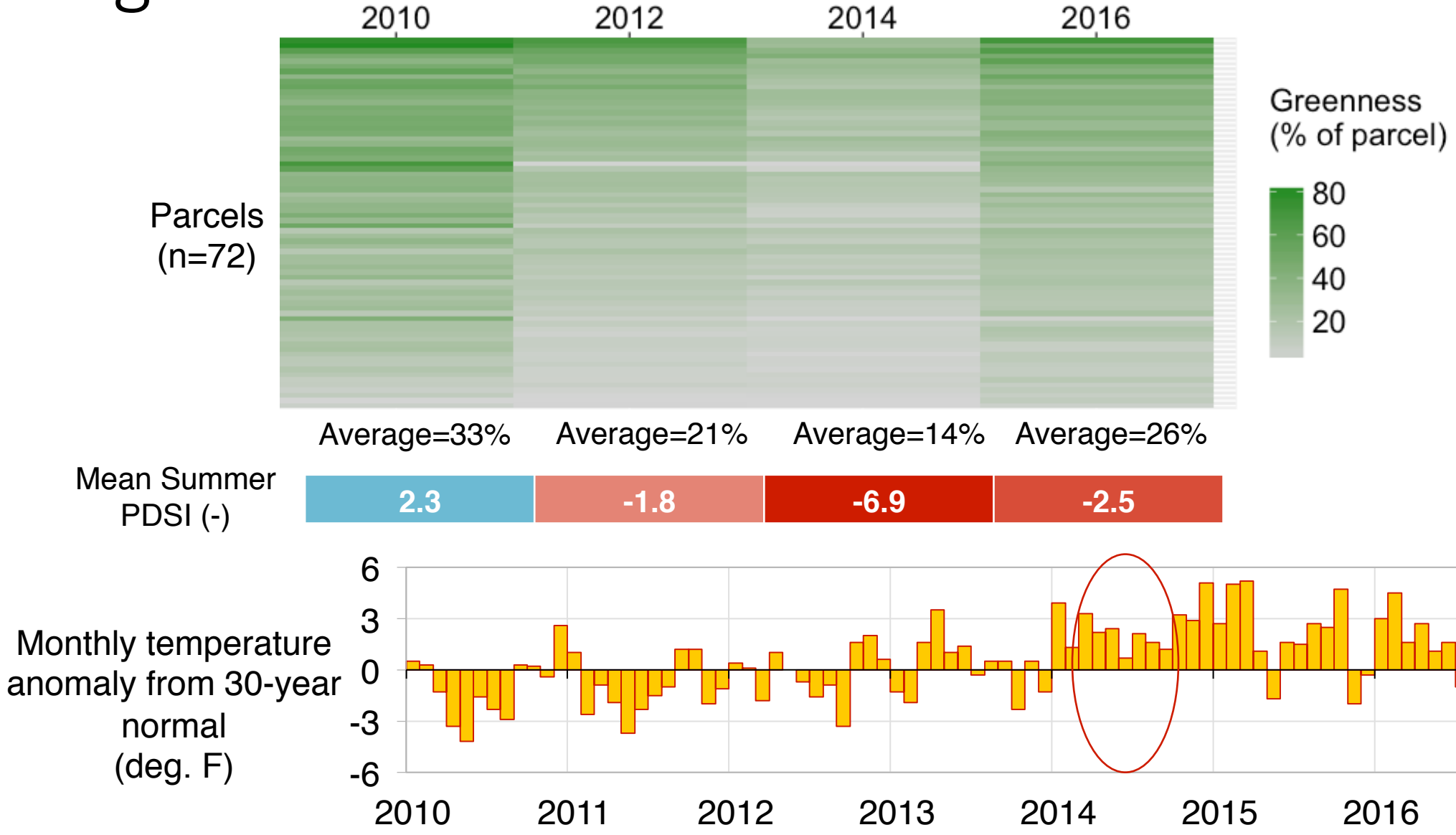
de re

Aerial imag ram (NAIP)
for 201(solution)



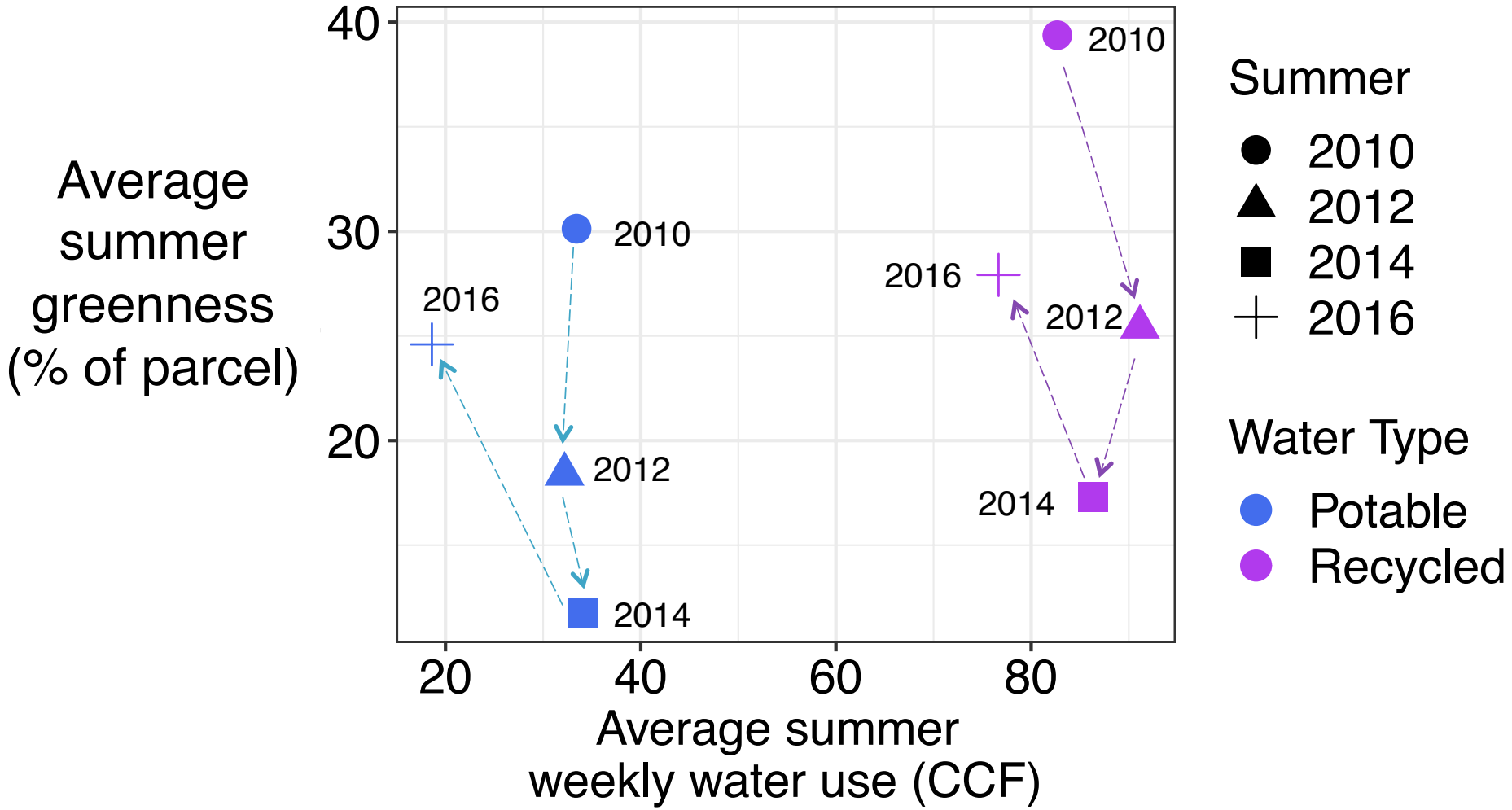
Quesnel, Ajami, and Marx (2019) *Environmental Research Letters*

Parcel greenness followed climatic conditions



Quesnel, Ajami, and Marx (2019) *Environmental Research Letters*

Greenness is not directly tied to water use



Quesnel, Ajami, and Marx (2019) *Environmental Research Letters*

Conclusions and Broader Impacts

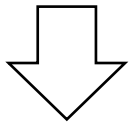
Data revolution, increased computational power, and interdisciplinary methods can help us to better understand human-water dynamics

New proxies for evolving social realities

Emerging data sources and new data aggregators

Modern water infrastructure systems

Evolving water use drivers, patterns, and trends



More informed and optimal decision-making for infrastructure development and demand-side management efforts

Thank you!

"We can't solve problems by using the same kind of thinking we used when we created them." Albert Einstein

email: newsha@stanford.edu

Questions?