CSCE 590-1: From Data to Decisions with Open Data: A Practical Introduction to AI

Lecture 4: A Systematic AI Approach – AI For Water

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

21ST JAN 2021

Carolinian Creed: "I will practice personal and academic integrity."

Organization of Lecture 4

- Introduction Segment
  - Recap of Lecture 3
- Main Segment
  - Real-World Problems – Water
  - A Systematic Approach
    - Value of decision: before and after
    - Data-needed
    - Method
    - Evaluation
    - Integrating with overall process
- Concluding Segment
  - About Next Lecture – Lecture 5
  - Ask me anything

Introduction Segment

Recap of Lecture 3

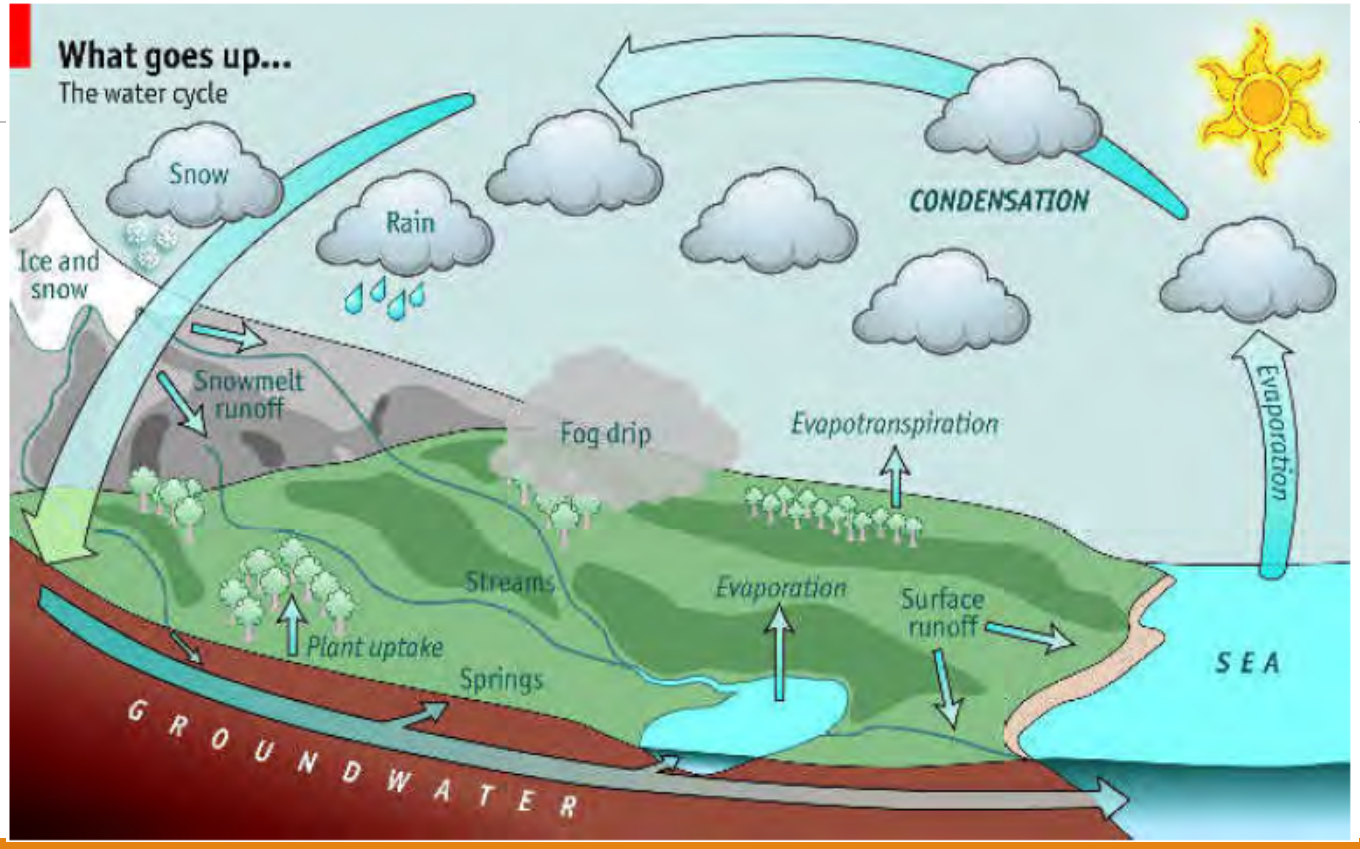
- We looked at problems of the real world
- Understood what is meant by “Smart City”. Also referred to as “Cyber-Physical System”.
- Introduced a systematic framework to introduce data-driven innovations
- Looked at examples in Health
  - Saw examples of AI methods: decision making, randomized control trial (synthetic control)

Main Segment

Blue Water project: https://sites.google.com/site/biplavsrivastava/research-1/bluewater
AAAI 2017 Tutorial: https://sites.google.com/site/aiwatertutorial/

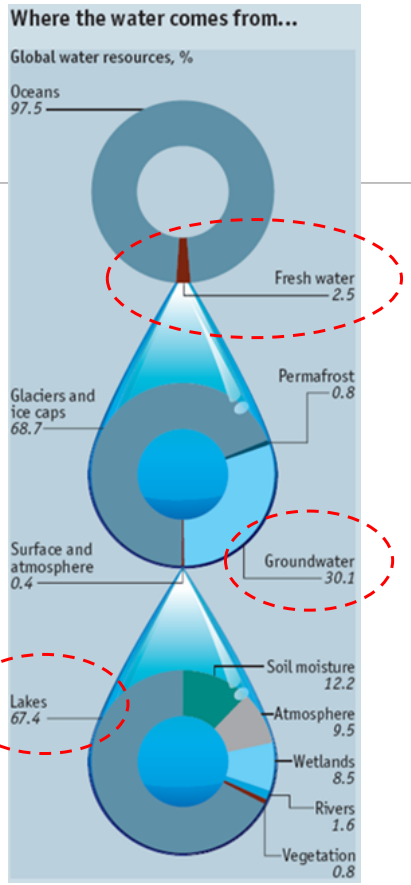
Water Cycle (aka Hydrological Cycle)

Source: Economist, May 20, 2010

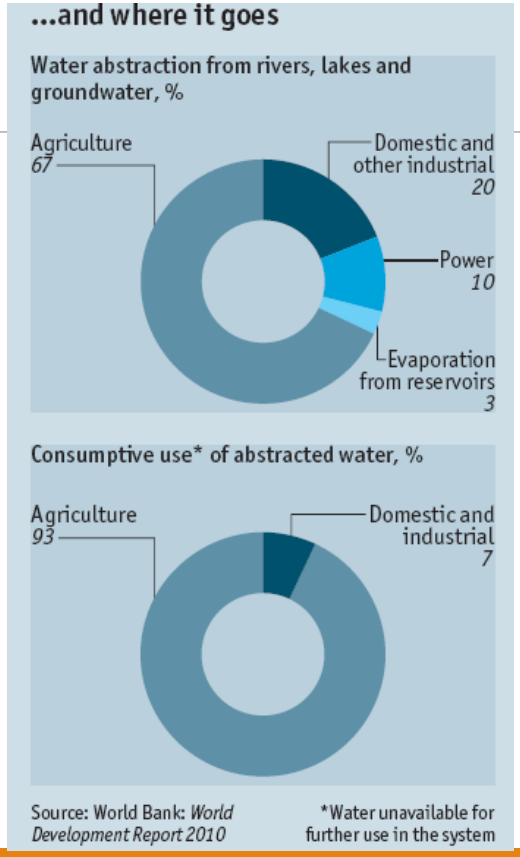


Fresh Water: Supply and Demand

Supply

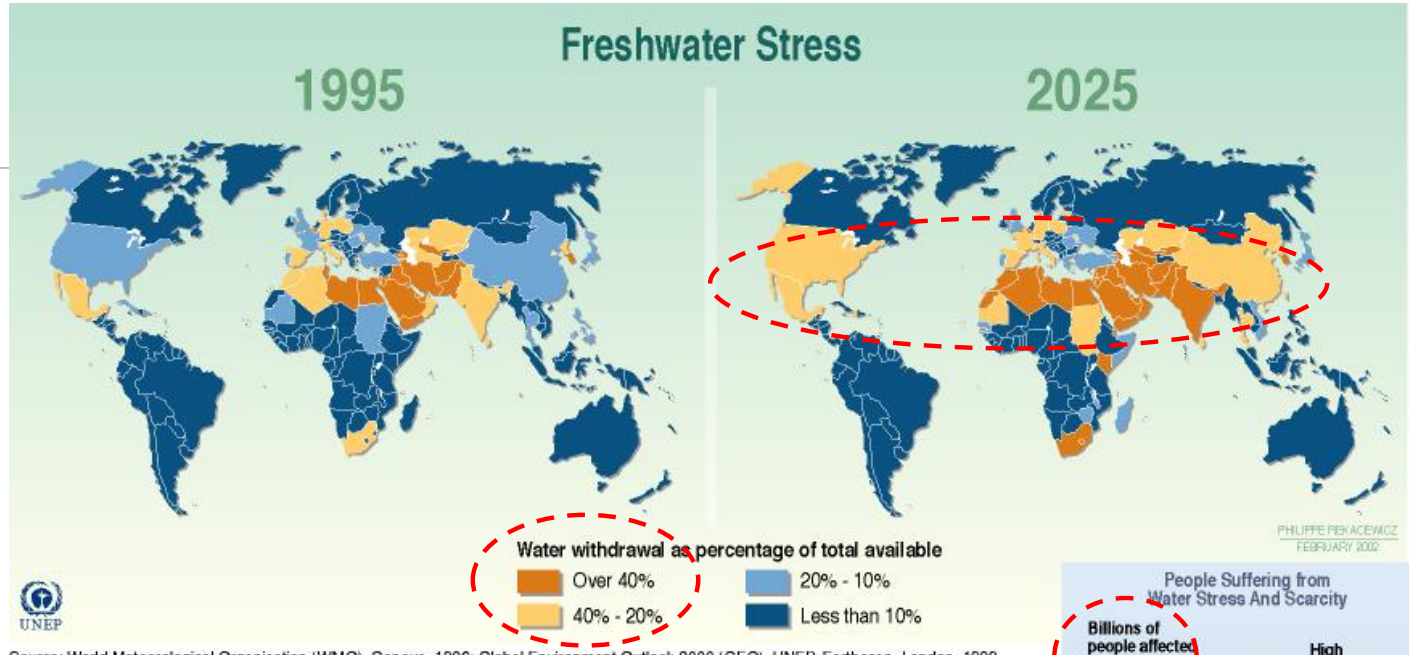


Demand

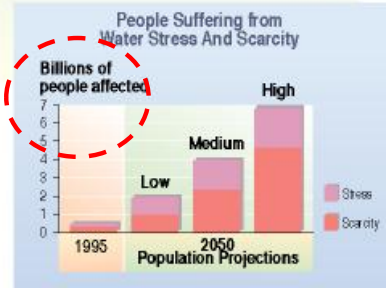


Source: Economist, May 20, 2010

Fresh Water Stress: Spatial Distribution



Source: World Meteorological Organisation (WMO), Geneva, 1996; Global Environment Outlook 2000 (GEO), UNEP, Earthscan, London, 1999.



Freshwater stress will affect both developed and developing nations. **Billions** will be affected.

Better information flow is critical for better water flow

The nature of water management must rapidly evolve

From	*To*
Manual Data Collection	Automated Sensing
Managing in Isolation	Managing Collaboratively
Intermittent Measurement	Real-Time Measurement
Multiple Data Sets	Data Integration
"Guesstimation" Tools	Modeled Decision Support
Commodity Pricing	Value Pricing
Tactical Problem Solving	Strategic Risk Management

"One barrier to better management of water resources is simply lack of data — where the water is, where it's going, how much is being used and for what purposes, how much might be saved by doing things differently. In this way, the water problem is largely an information problem. The information we can assemble has a huge bearing on how we cope with a world at peak water."

Source: **Wired Magazine, "Peak Water: Aquifers and Rivers Are Running Dry. How Three Regions Are Coping", Matthew Power, April 21st, 2008**

Challenges and Need for AI

Challenges

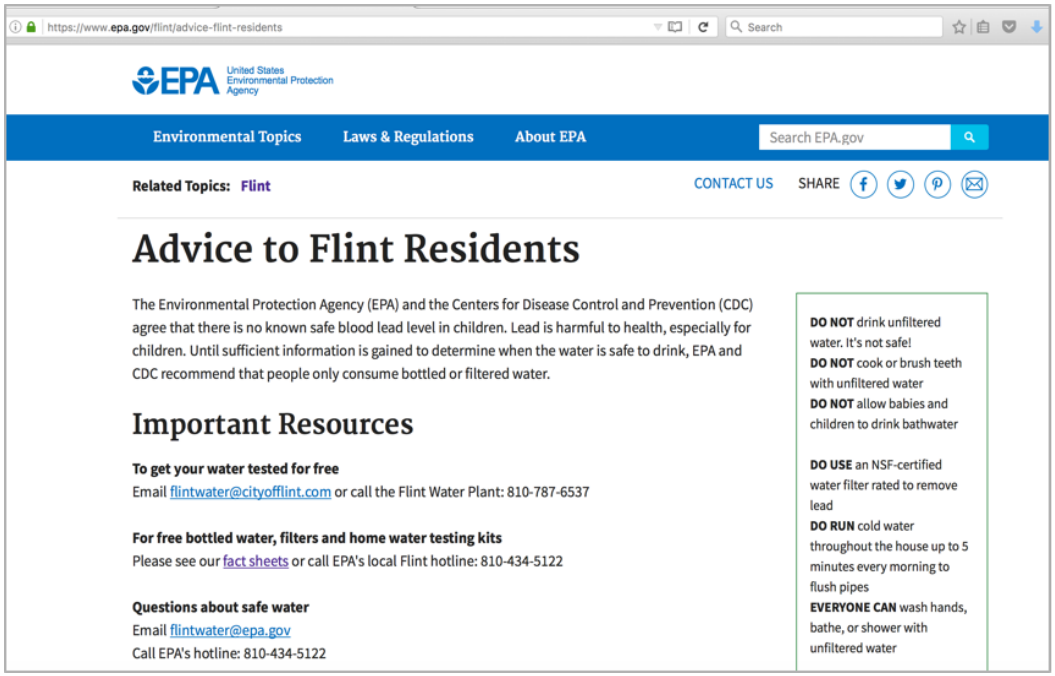
- Increasing demand due to
  - Population
  - Changing water-intensive lifestyle
  - Industrial growth
- Shrinking supplies
  - Erratic rains due to climate change
  - Sewage / effluent increase
- Poor management
  - Below cost, unsustainable, pricing
  - Delayed or neglected maintenance

Water is the next flash point for wars

Value of Water Pollution Data

- Government for business decisions
  - Source attribution
  - Sewage treatment
  - Public Health
- Individuals for personal decisions
  - Bathing (Religious, Lifestyle)
  - Recreation
  - Community practices
- Businesses
  - Inland transportation
  - Water usage and treatment options

Water Information to People: Static and Post-Facto Advisories (US)



Advisories to public for Flint Residents, MI, USA

Physical signage at a lake in Washington, USA

Problem and Objective

Guide every day people, who may be non-experts, with a multi-modal assistant to take data-based decisions specific to their needs, leveraging complex water quality data.

Audience

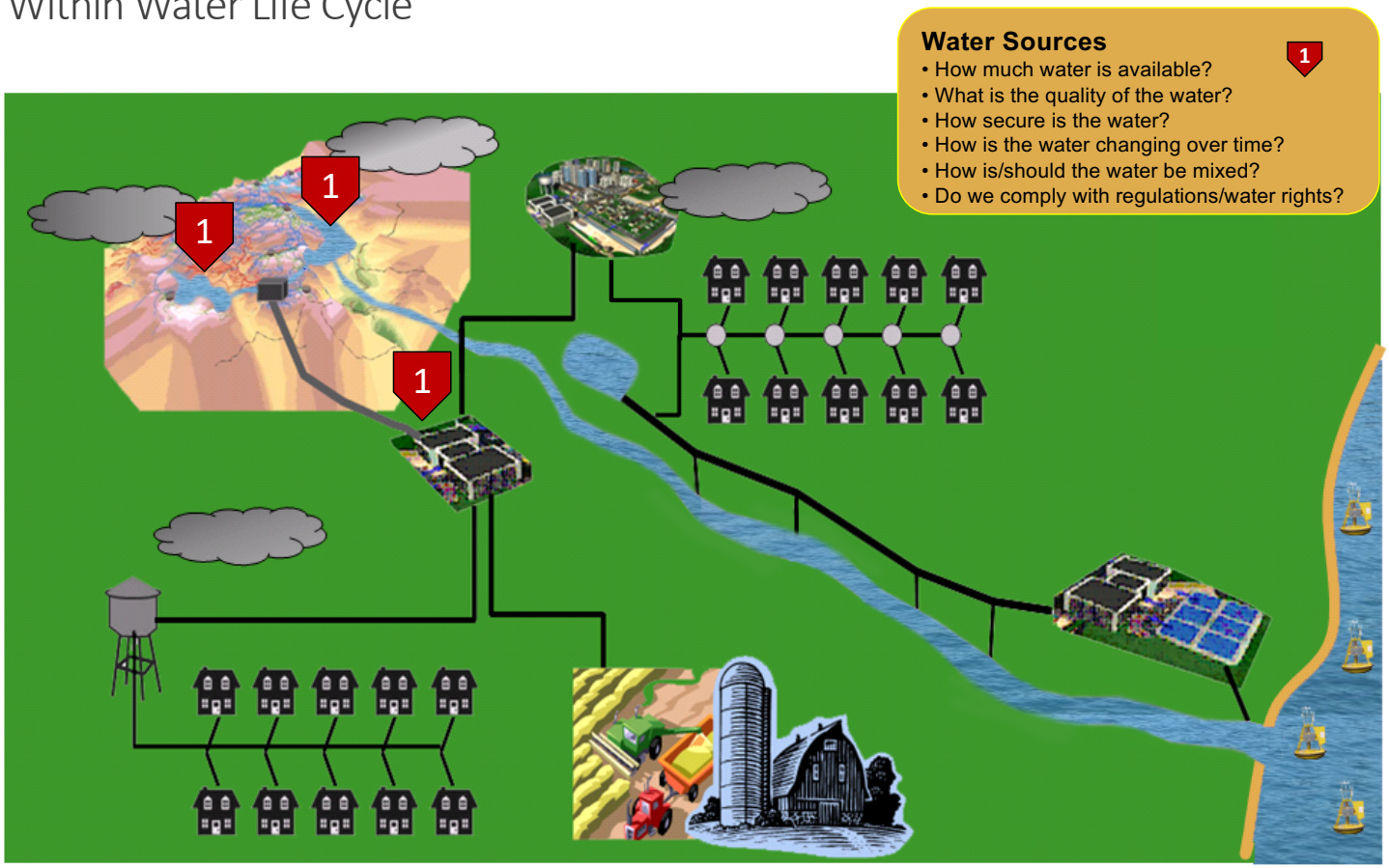
- General Public that wants to understand water quality at a specific location (e.g., swimming)
- Professionals with responsibility for regions (e.g., public health)

Before and After

Now: Static, non-interactive, non-contextual, lacks data details

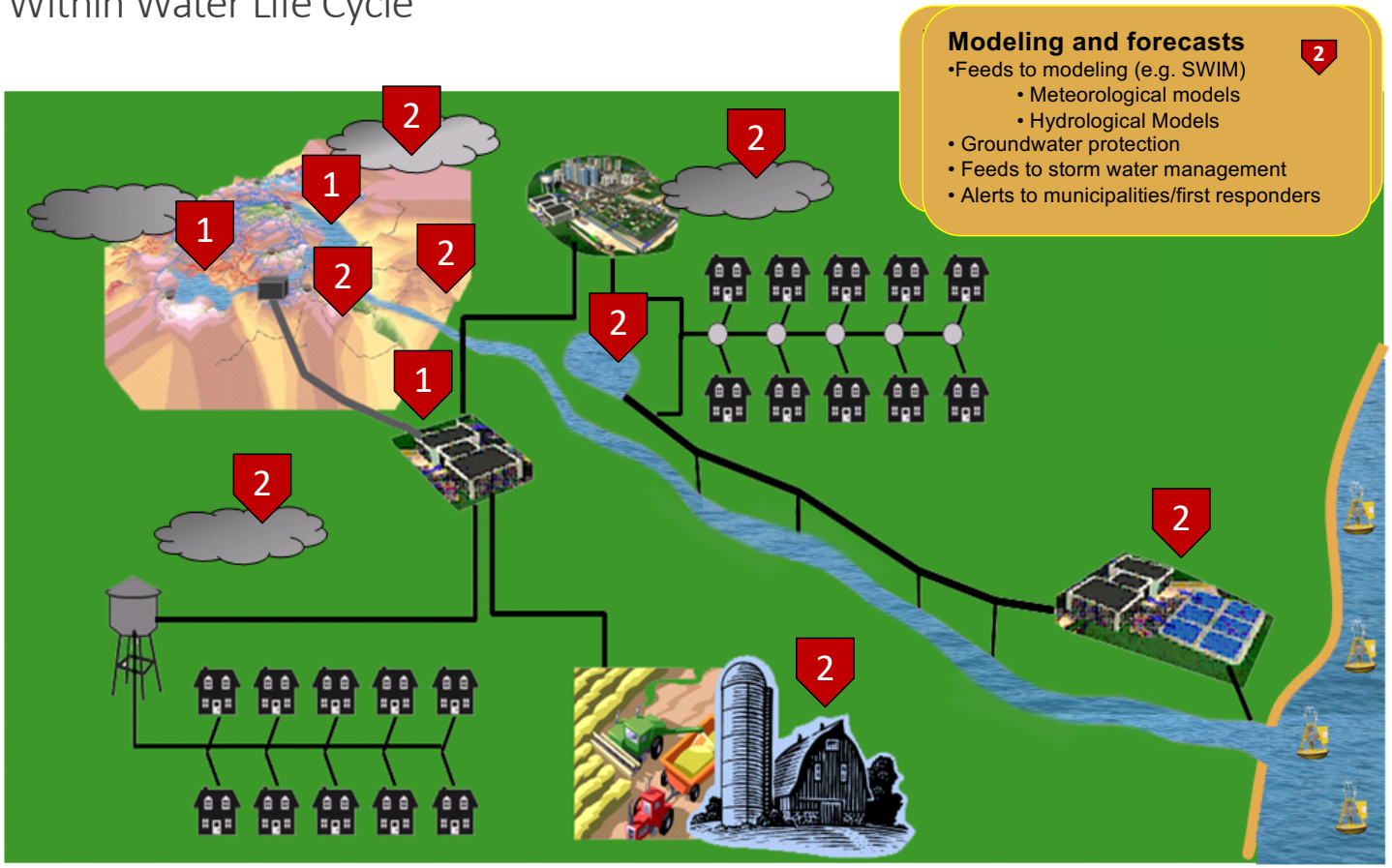
Future: Anywhere, interactive, explain with data, contextual

Analytics Within Water Life Cycle



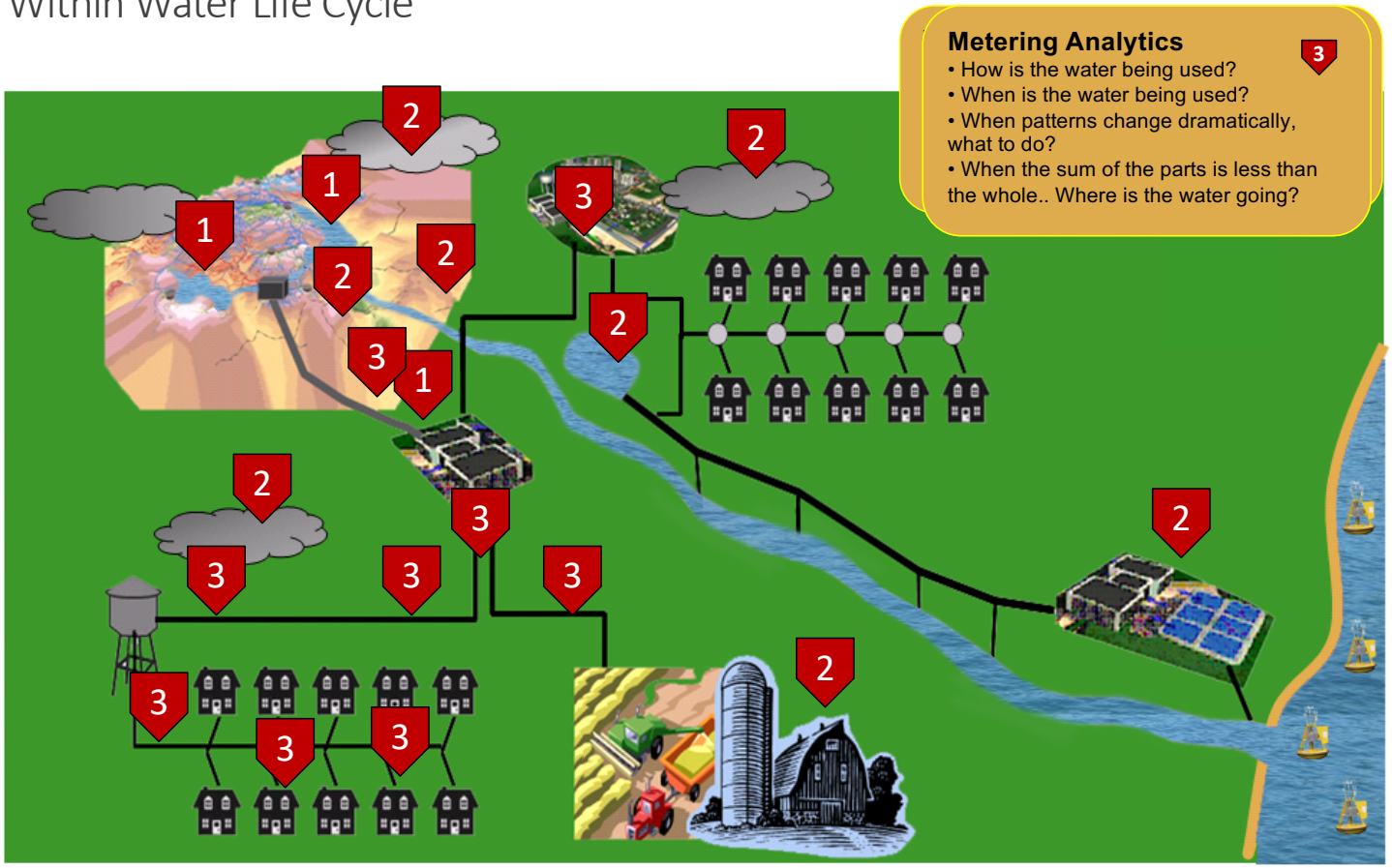
Advanced Water Management Capabilities

Analytics Within Water Life Cycle



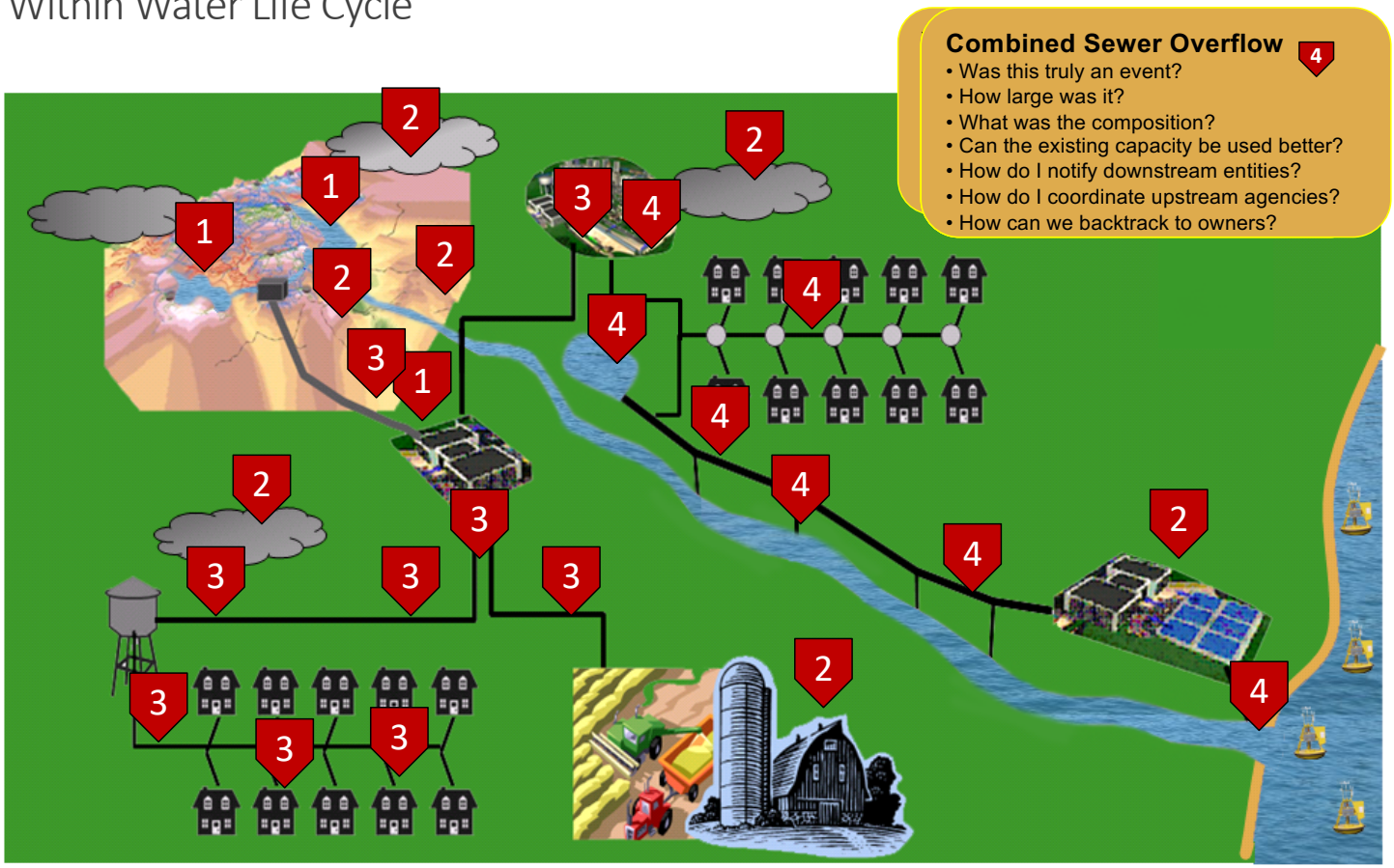
Advanced Water Management Capabilities

Analytics Within Water Life Cycle



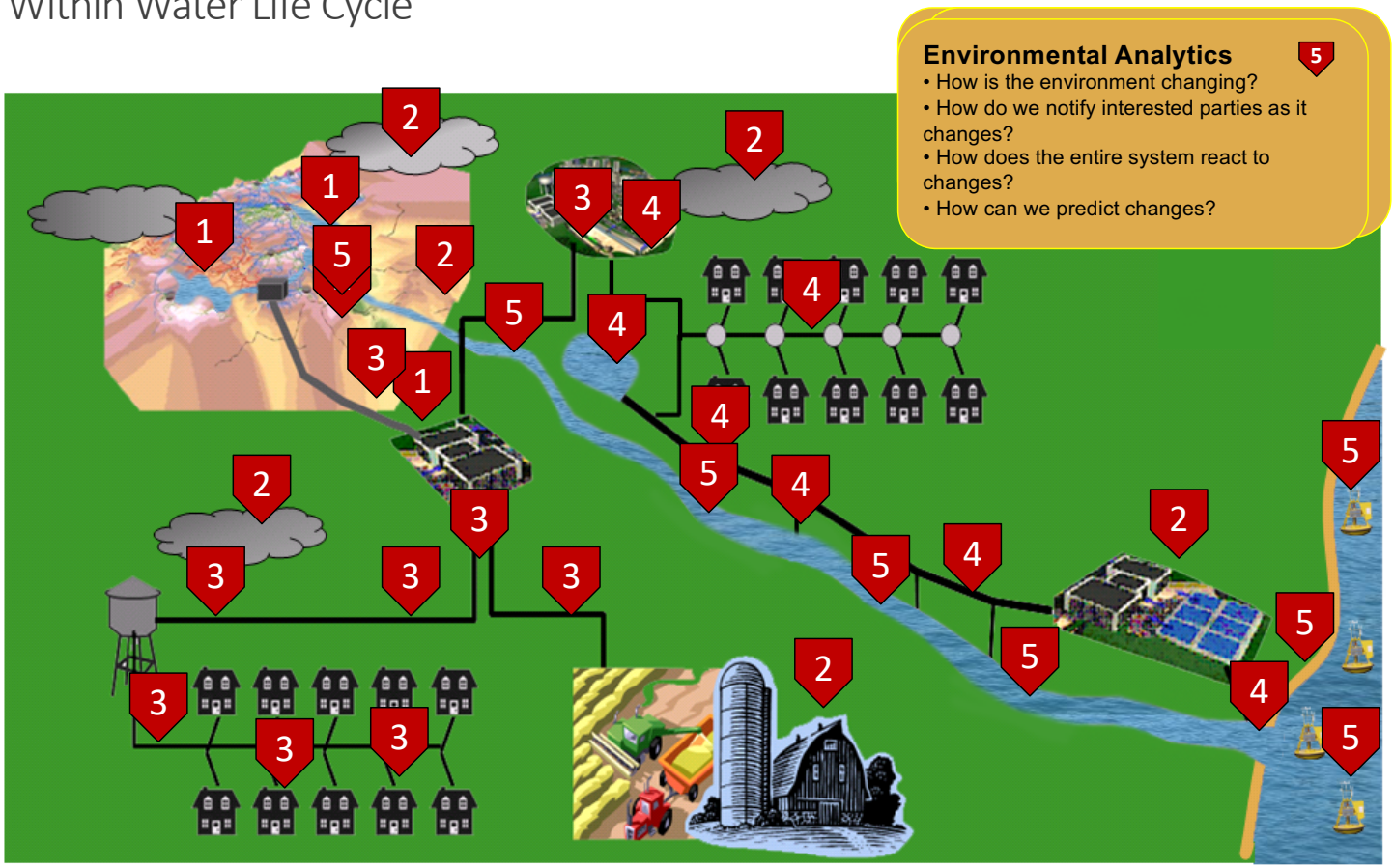
Advanced Water Management Capabilities

Analytics Within Water Life Cycle



Advanced Water Management Capabilities

Analytics Within Water Life Cycle



Advanced Water Management Capabilities

Exercise: Water Problems in South Carolina

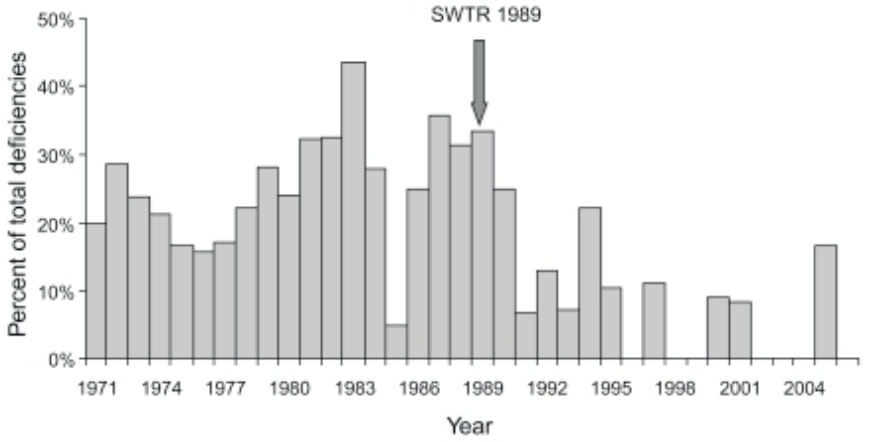
- Problems
  - Whether SC specific?
- What people are doing today?
- Data

Case Study: Advice to People in US

USA: Water Safety Rules and Monitoring Reduce Water-Borne Diseases

Number of reported waterborne-disease outbreaks in public drinking water systems—United States, 1971–2006 (N = 680).

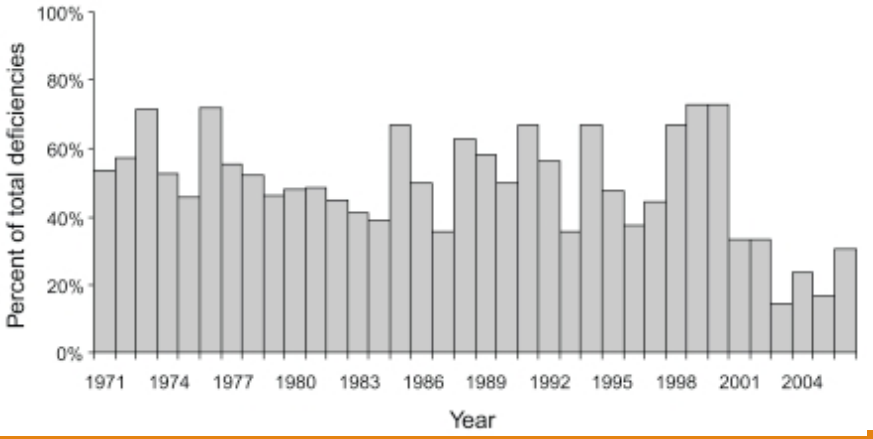
SOURCE: CDC, unpublished WBDOSS data.



Proportion of deficiencies in public drinking water systems associated with untreated or improperly treated ground water—United States, 1971–2006.

Deficiency = antecedent event or situation that results in exposure of persons to a disease-causing agent or agents. May be single or multiple deficiencies associated with each outbreak.

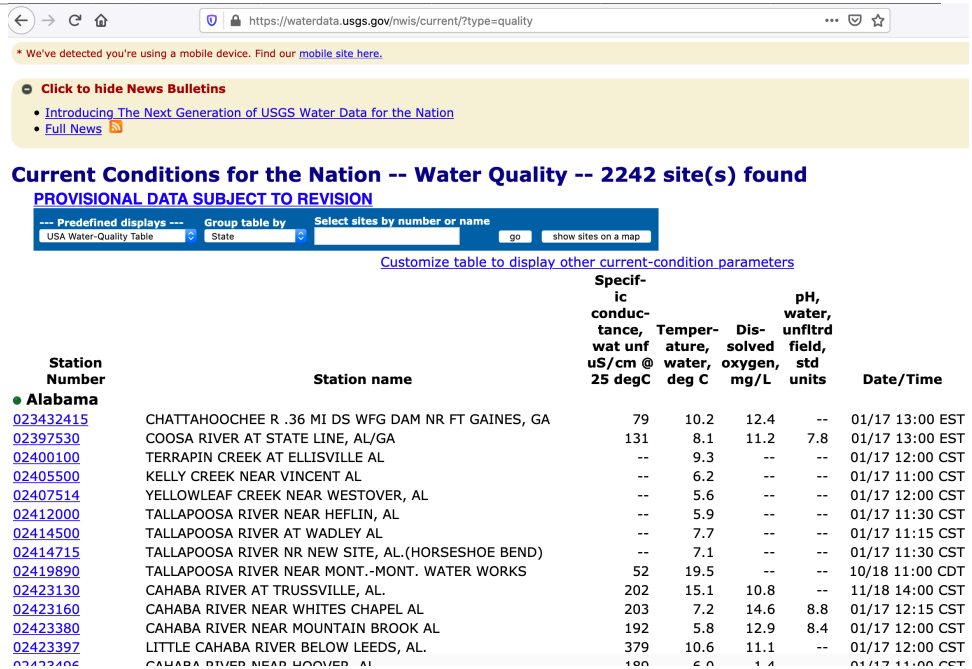
SOURCE: CDC, unpublished WBDOSS data.



SWTR: Surface Water Treatment rule of 1989

US Geological Survey's Water Data

- Web interface : https://waterdata.usgs.gov/nwis
- REST interface: https://waterservices.usgs.gov/rest/DV-Service.html
- Snapshot: https://waterdata.usgs.gov/nwis/current/?type=quality



https://waterdata.usgs.gov/nwis/current/?type=quality

* We've detected you're using a mobile device. Find our mobile site here.

Click to hide News Bulletins

- Introducing The Next Generation of USGS Water Data for the Nation
- Full News

Current Conditions for the Nation -- Water Quality -- 2242 site(s) found

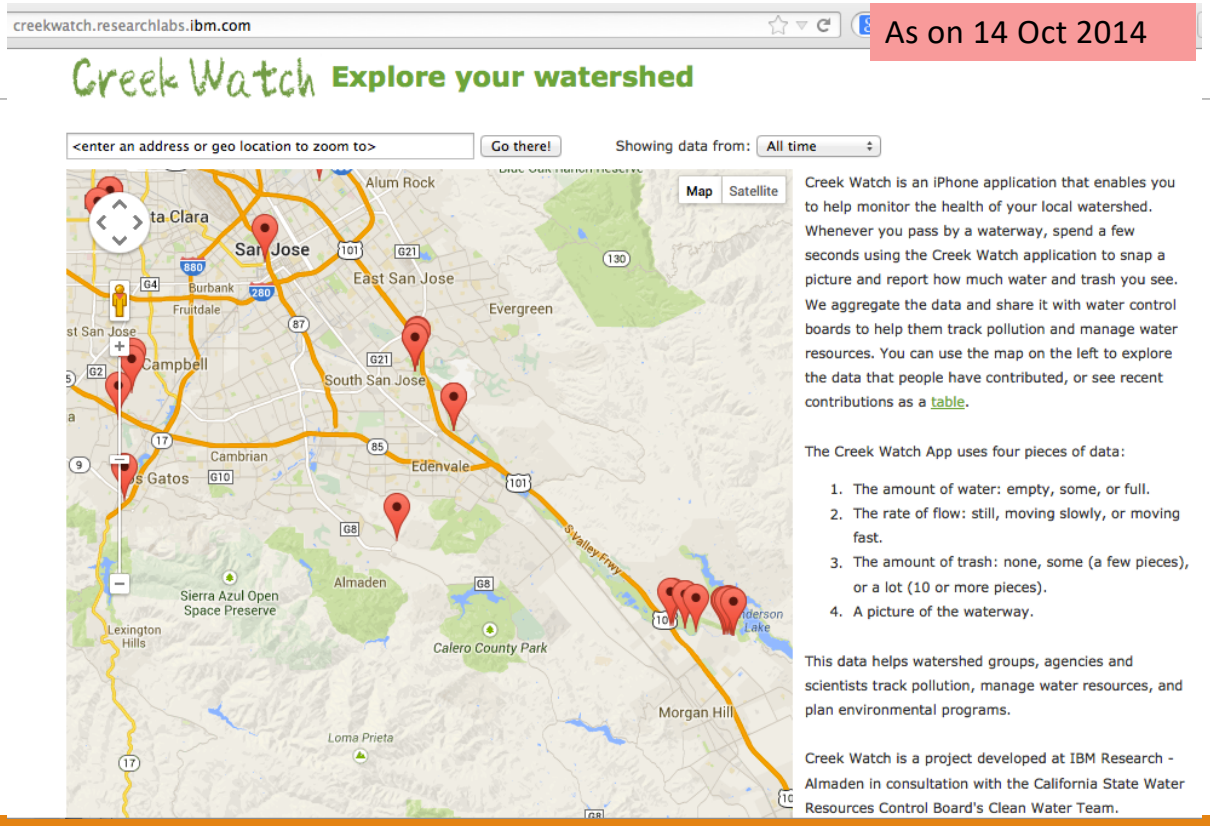
PROVISIONAL DATA SUBJECT TO REVISION

--- Predefined displays --- | USA Water-Quality Table | Group table by | State | Select sites by number or name | go | show sites on a map

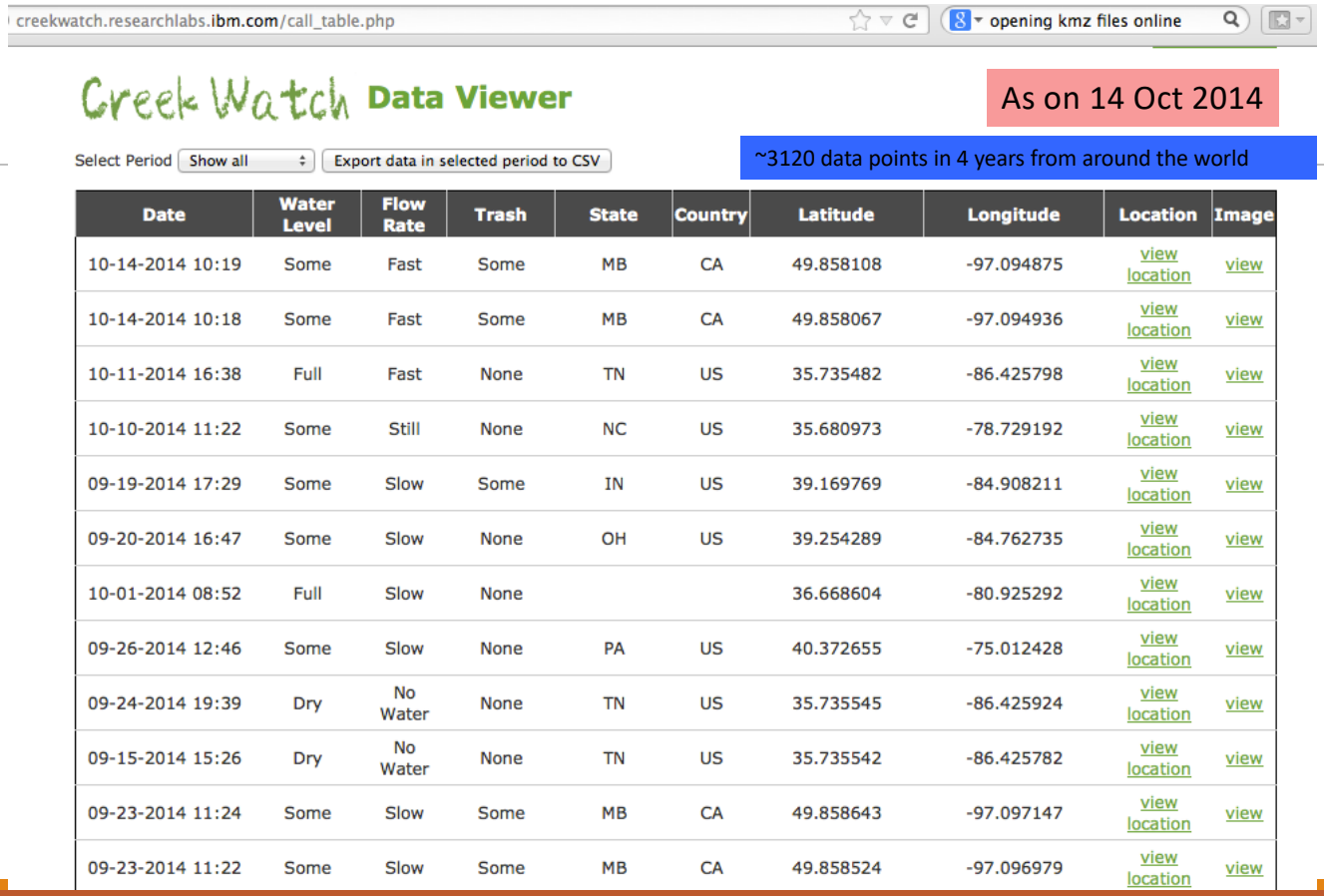
Customize table to display other current-condition parameters

Station Number	Station name	Specif-ic conduc-tance, wat unf uS/cm @ 25 degC	Temper-ature, water, deg C	Dis-solved oxygen, mg/L	pH, water, unfltrd field, std units	Date/Time
Alabama						
023432415	CHATTAHOOCHEE R .36 MI DS WFG DAM NR FT GAINES, GA	79	10.2	12.4	--	01/17 13:00 EST
02397530	COOSA RIVER AT STATE LINE, AL/GA	131	8.1	11.2	7.8	01/17 13:00 EST
02400100	TERRAPIN CREEK AT ELLISVILLE AL	--	9.3	--	--	01/17 12:00 CST
02405500	KELLY CREEK NEAR VINCENT AL	--	6.2	--	--	01/17 11:00 CST
02407514	YELLOWLEAF CREEK NEAR WESTOVER, AL	--	5.6	--	--	01/17 12:00 CST
02412000	TALLAPOOSA RIVER NEAR HEFLIN, AL	--	5.9	--	--	01/17 11:30 CST
02414500	TALLAPOOSA RIVER AT WADLEY AL	--	7.7	--	--	01/17 11:15 CST
02414715	TALLAPOOSA RIVER NR NEW SITE, AL.(HORSESHOE BEND)	--	7.1	--	--	01/17 11:30 CST
02419890	TALLAPOOSA RIVER NEAR MONT.-MONT. WATER WORKS	52	19.5	--	--	10/18 11:00 CDT
02423130	CAHABA RIVER AT TRUSSVILLE, AL.	202	15.1	10.8	--	11/18 14:00 CST
02423160	CAHABA RIVER NEAR WHITES CHAPEL AL	203	7.2	14.6	8.8	01/17 12:15 CST
02423380	CAHABA RIVER NEAR MOUNTAIN BROOK AL	192	5.8	12.9	8.4	01/17 12:00 CST
02423397	LITTLE CAHABA RIVER BELOW LEEDS, AL.	379	10.6	11.1	--	01/17 12:00 CST

Creek Watch – Crowd Sourced Water Information Collection

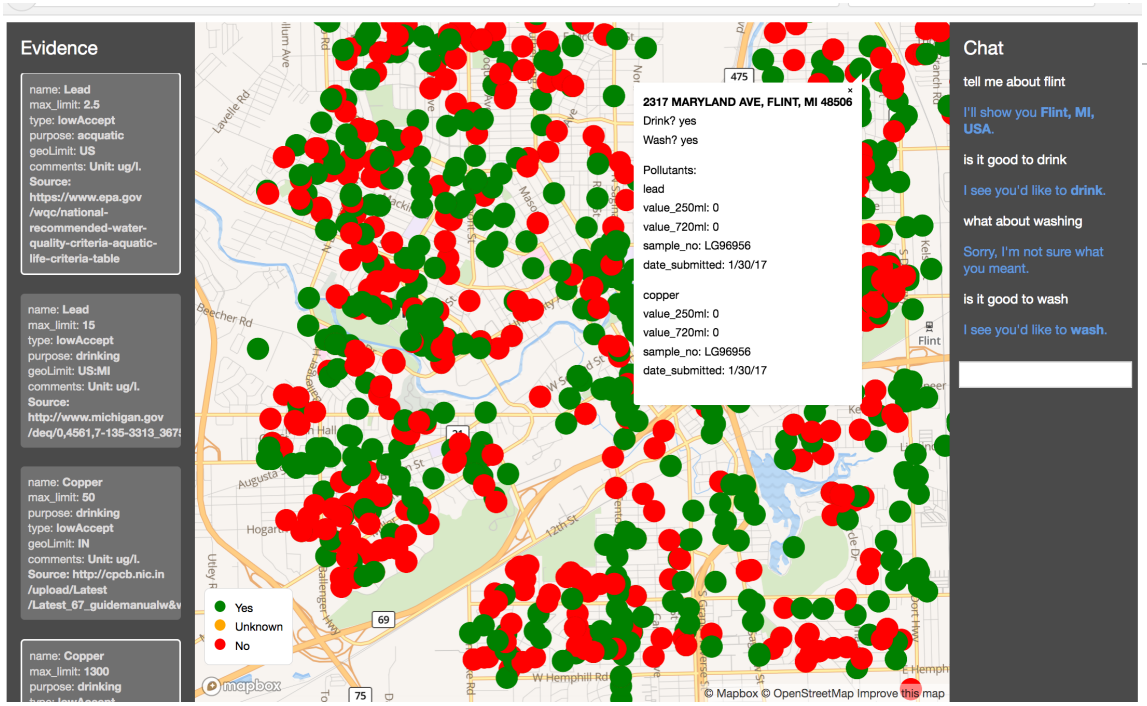


Location: http://creekwatch.researchlabs.ibm.com/call_table.php



Date	Water Level	Flow Rate	Trash	State	Country	Latitude	Longitude	Location	Image
10-14-2014 10:19	Some	Fast	Some	MB	CA	49.858108	-97.094875	view location	view
10-14-2014 10:18	Some	Fast	Some	MB	CA	49.858067	-97.094936	view location	view
10-11-2014 16:38	Full	Fast	None	TN	US	35.735482	-86.425798	view location	view
10-10-2014 11:22	Some	Still	None	NC	US	35.680973	-78.729192	view location	view
09-19-2014 17:29	Some	Slow	Some	IN	US	39.169769	-84.908211	view location	view
09-20-2014 16:47	Some	Slow	None	OH	US	39.254289	-84.762735	view location	view
10-01-2014 08:52	Full	Slow	None			36.668604	-80.925292	view location	view
09-26-2014 12:46	Some	Slow	None	PA	US	40.372655	-75.012428	view location	view
09-24-2014 19:39	Dry	No Water	None	TN	US	35.735545	-86.425924	view location	view
09-15-2014 15:26	Dry	No Water	None	TN	US	35.735542	-86.425782	view location	view
09-23-2014 11:24	Some	Slow	Some	MB	CA	49.858643	-97.097147	view location	view
09-23-2014 11:22	Some	Slow	Some	MB	CA	49.858524	-97.096979	view location	view

Water Advice Via Conversation Interfaces



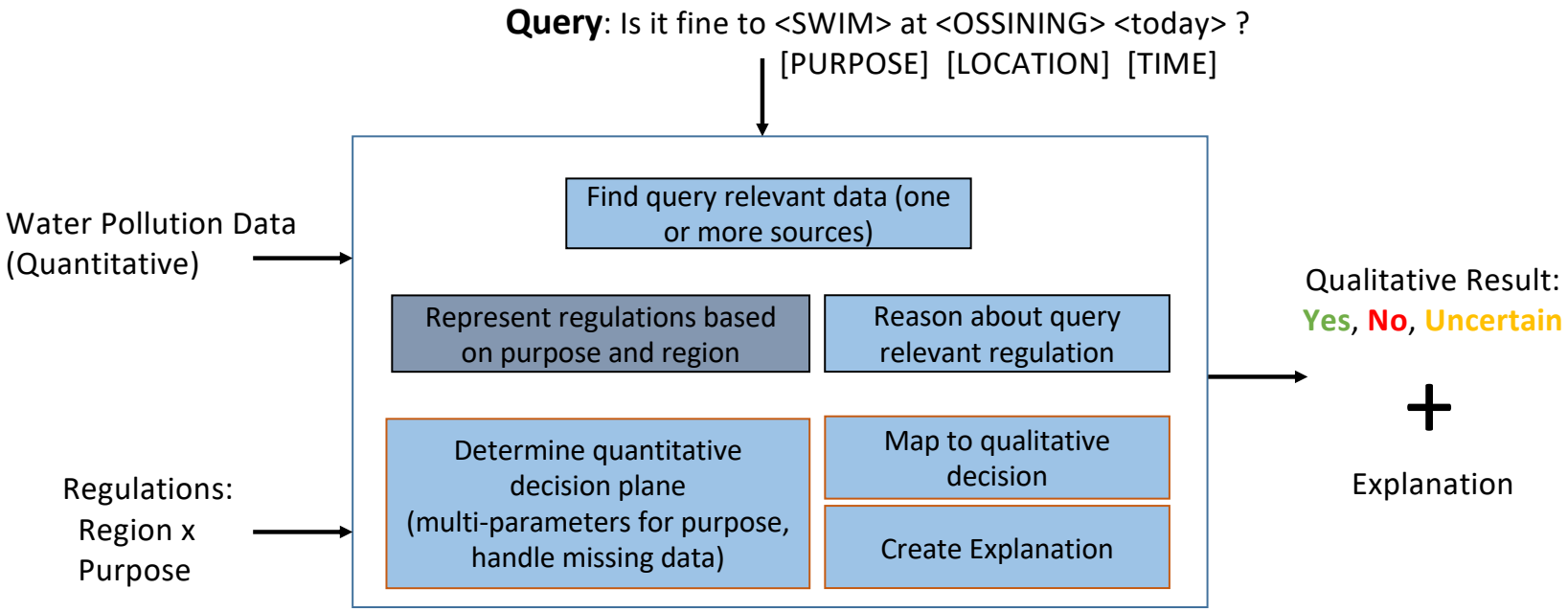
•Video: https://www.youtube.com/watch?v=z4x44sxC3zA

•Paper: Jason Ellis, Biplav Srivastava, Rachel Bellamy and Andy Aaron, Water Advisor - A Data-Driven, Multi-Modal, Contextual Assistant to Help with Water Usage Decisions, AAAI 2018. [Demonstration paper].

Modes (R to L):

1. Conversation
2. Map
3. Documents

Architecture



Data, Analysis and (AI) Challenges

Data Needed

- Mapping of activities (e.g., bathing, drinking, fishing) to water quality parameters
  - Created and made open-source at: https://github.com/biplav-s/water-info
  - Novelty: None existed before
- Regulations applicable to geography and activity, plus water quality parameters
  - Demo: US EPA, Michigan, New York, India
  - Challenge: under-specified, conflicting
- Actual water quality data
  - Demo: US Geological Survey (USGS)

AI Methods

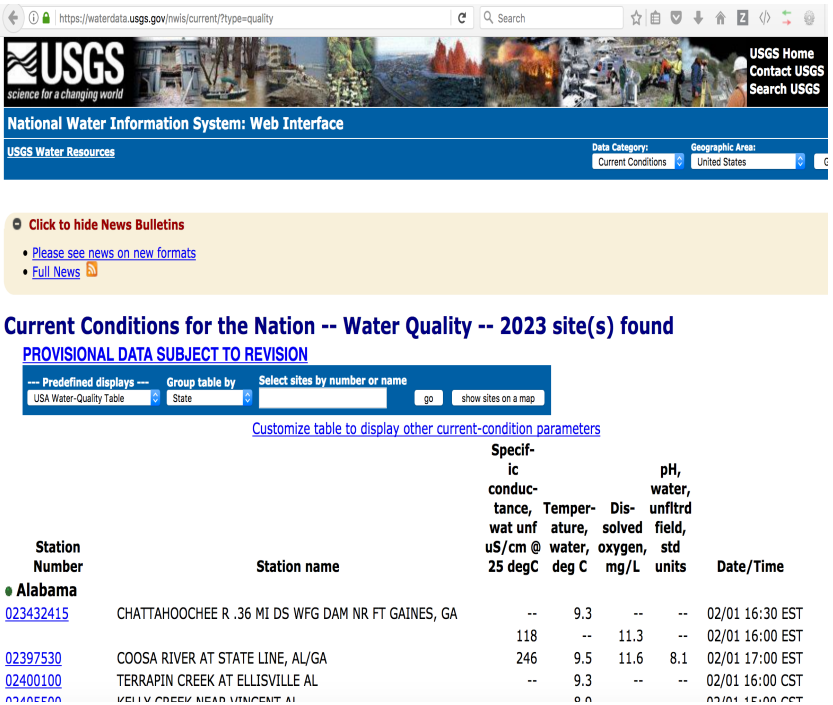
- Constraints, Rule-based processing, Preferences, Knowledge Graph

Sample Standards: Just in Drinking Water

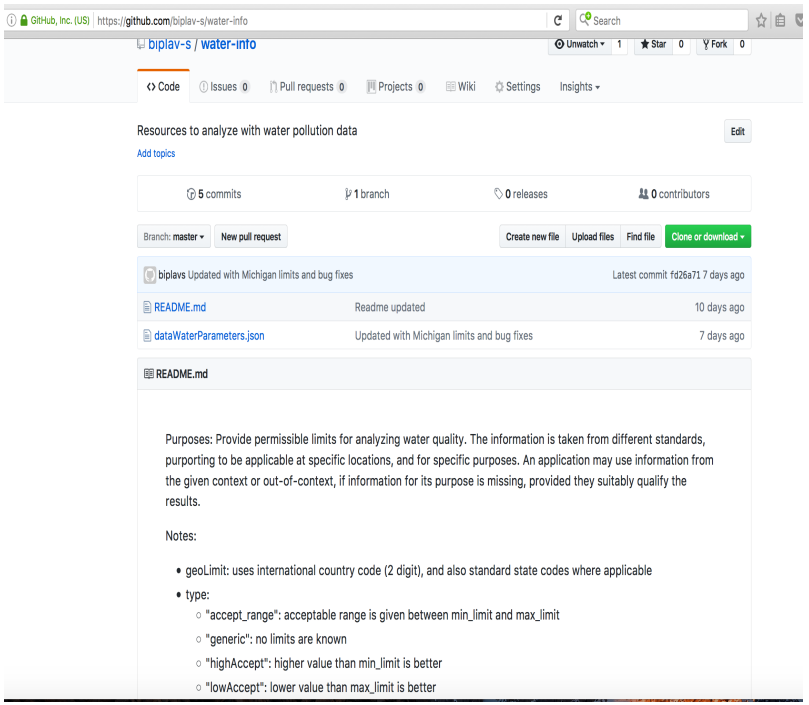
Source: https://en.wikipedia.org/wiki/Drinking_water_quality_standards

Parameter	World Health Organization	European Union	United States	China
Acrylamide	"	0.10 µg/l	"	"
Arsenic	10µg/l	10 µg/l	10µg/l	50µg/l
Antimony	ns	5.0 µg/l	6.0 µg/l	"
Barium	700µg/l	ns	2 mg/L	"
Benzene	10µg/l	1.0 µg/l	5 µg/l	"
Benzo(a)pyrene	"	0.010 µg/l	0.2 µg/l	0.0028 µg/l
Boron	2.4mg/l	1.0 mg/L	"	"
Bromate	"	10 µg/l	10 µg/l	"
Cadmium	3 µg/l	5 µg/l	5 µg/l	5 µg/l
Chromium	50µg/l	50 µg/l	0.1 mg/L	50 µg/l (Cr6)
Copper	"	2.0 mg/l	TT	1 mg/l
Cyanide	"	50 µg/l	0.2 mg/L	50 µg/l
1,2-dichloroethane	"	3.0 µg/l	5 µg/l	"
Epichlorohydrin	"	0.10 µg/l	"	"
Fluoride	1.5 mg/l	1.5 mg/l	4 mg/l	1 mg/l
Lead	"	10 µg/l	15 µg/l	10 µg/l
Mercury	6 µg/l	1 µg/l	2 µg/l	0.05 µg/l
Nickel	"	20 µg/l	"	"
Nitrate	50 mg/l	50 mg/l	10 mg/L (as N)	10 mg/L (as N)
Nitrite	"	0.50 mg/l	1 mg/L (as N)	"

Water Data from USGS



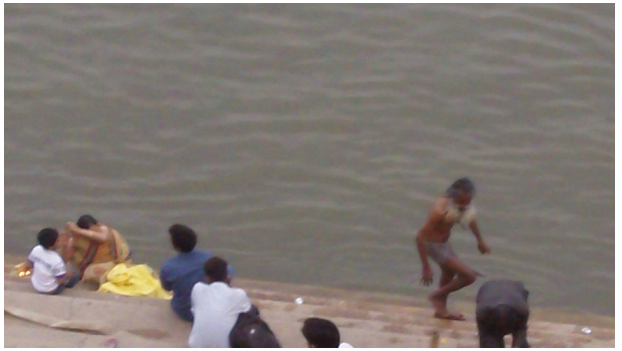
Mapping Made Available on Github



Case Study: Advice to People in India

[India] Ganga – Local Ground Situation @ Varanasi (Assi/ Tulsi Ghats) + Patna

Assi Ghat post recent cleanup



Bathing on Tulsi Ghat

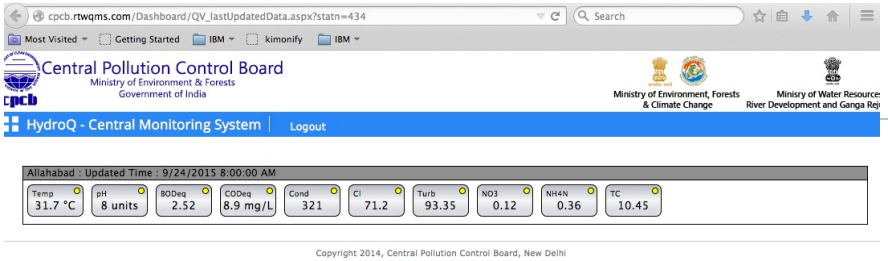
A nullah draining into Ganga

A manual powered boat

Common scene around Indian water bodies

Photos at Gandhi Ghat, Patna on 18 March 2015 during 1700-1800 Hrs
Photos of/ at Assi/ Tulsi Ghat, Varanasi on 25 March 2015 during 1700-1800 Hrs

Decision Example –River Water Pollution



Value – To individuals, businesses, government institutions

- **Individual Use Example** – Can I take a bath? Will it cause me dysentery?
- **Business Use Examples** – How should govt spend money on sewage treatment for maximum disease reduction? If an industry is being set-up, which site is best among feasible alternatives with least impact and latest technology?

Data – Quantitative as well as qualitative

- Dissolved oxygen,
- pH,
- … 30+ measurable quantities of interest

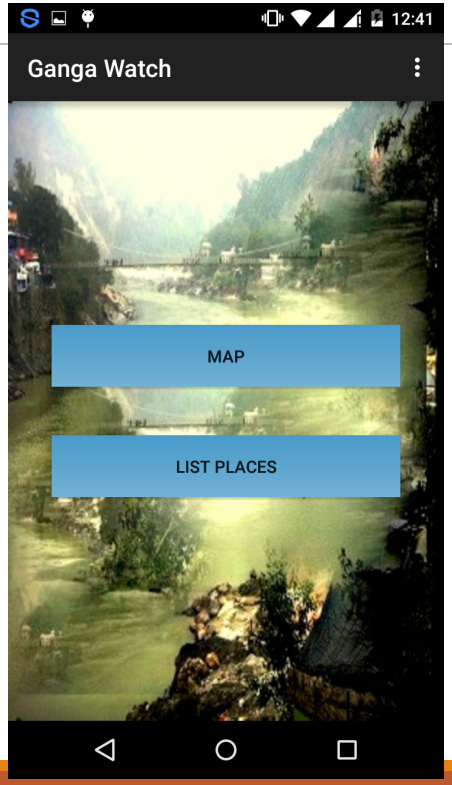
Access –

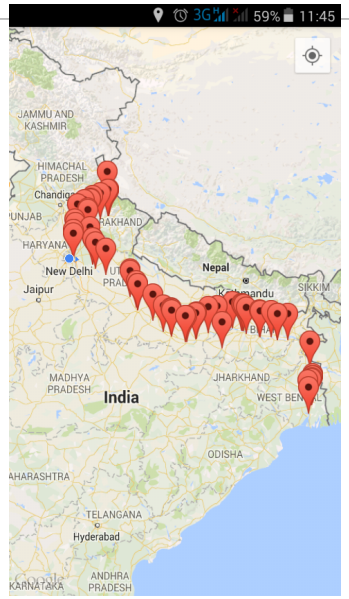
- Today, little, and that too in water technical jargon
- In pdf documents, website

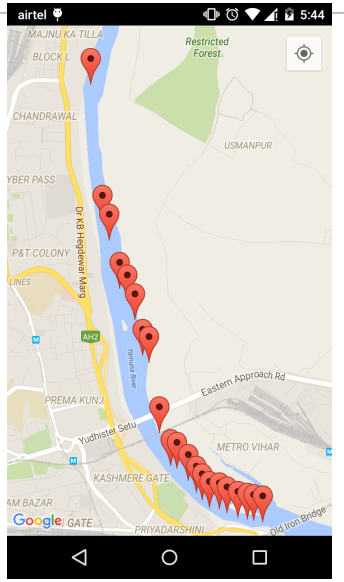
Key Idea: Can we make insights available when needed and help people make better decisions?

Download by searching for “GangaWatch” on Android Playstore in India.

Demo: GangaWatch (1/2)

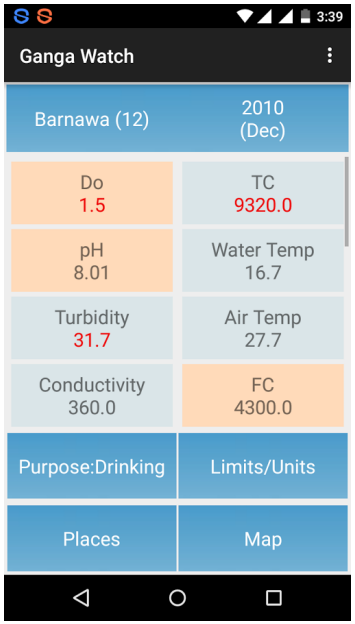


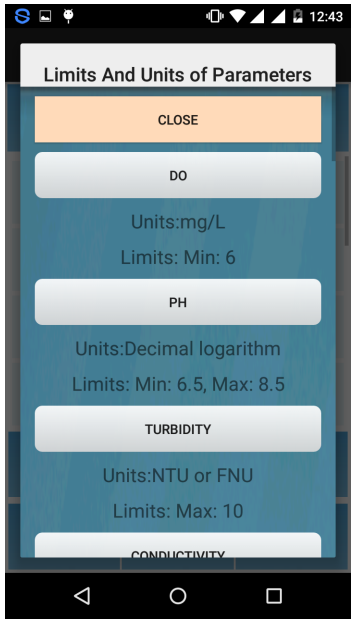


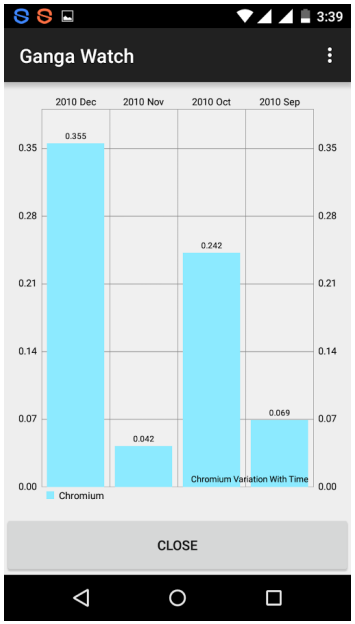


IBM’s Collected Fine-grained,Geo-tagged Data from a Real Time Run on Yamuna

Demo: GangaWatch (2/2)

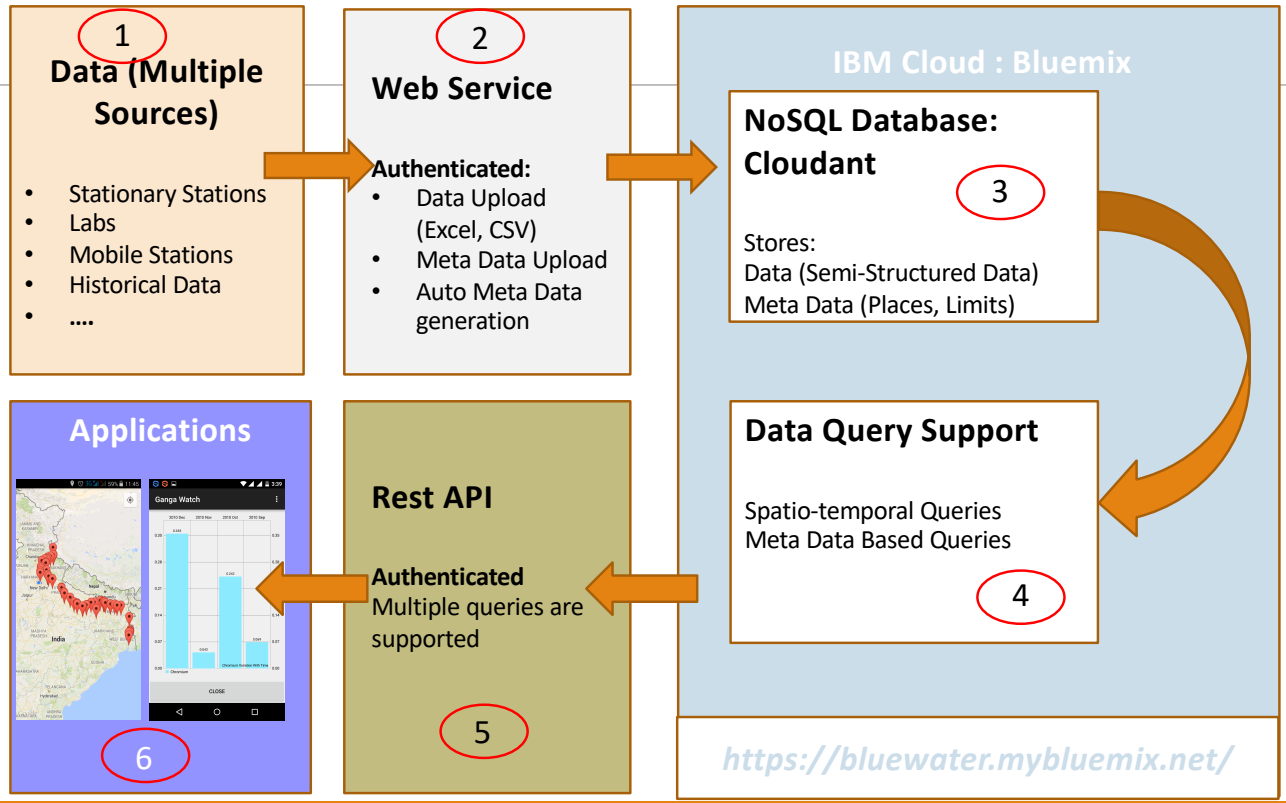






Paper at: https://arxiv.org/abs/1701.08212

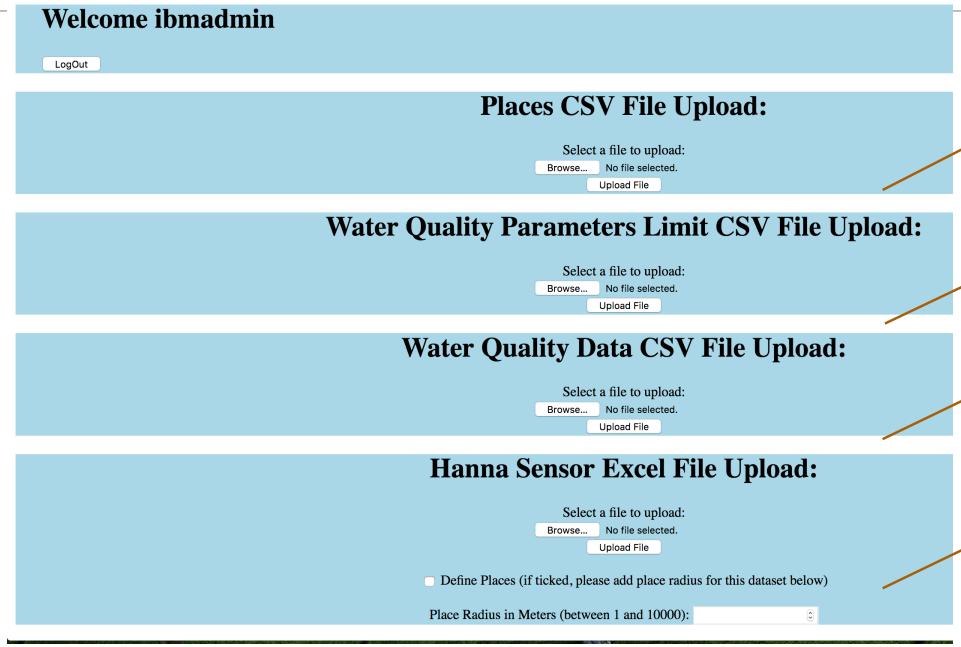
Blue Water Architecture: Water Data



BlueWater In-Depth: Web Service

Authenticated Data Upload

https://bluewater.mybluemix.net



Manual Places Definitions

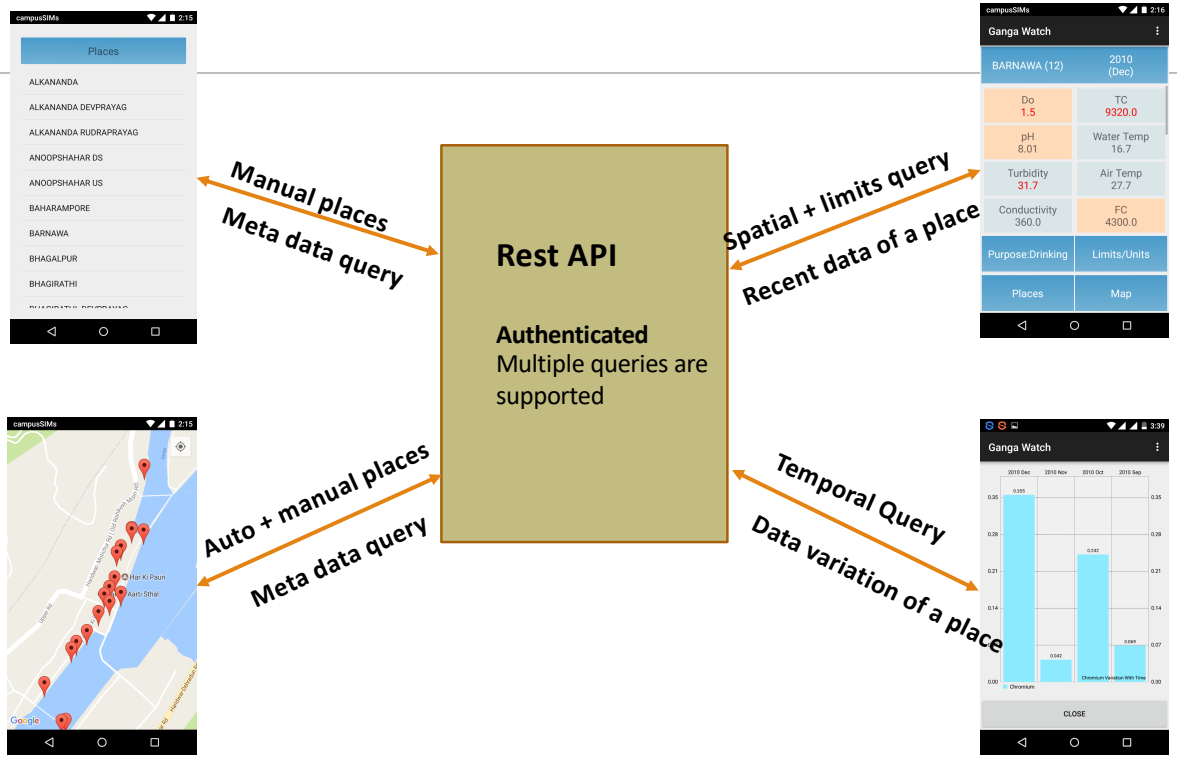
Limits on Water Quality

Water Quality Data of Parameters in CSV

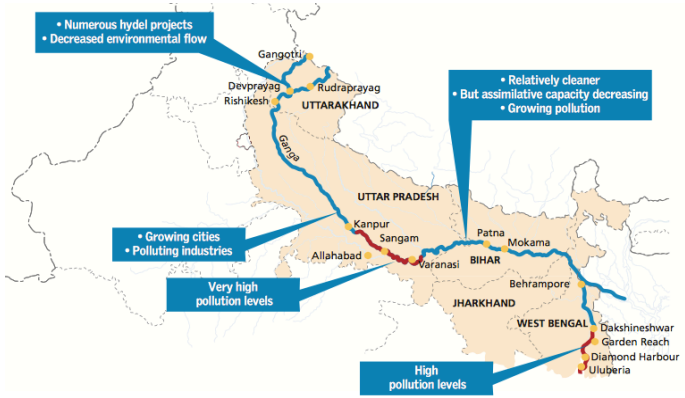
Water Quality Data of Parameters in Excel

Resources for water quality: https://github.com/biplav-s/water-info/

BlueWater In-Depth: GangaWatch Using REST API



Experience Working in the Field



UTTARAKHAND	UTTAR PRADESH	BIHAR	JHARKHAND	WEST BENGAL
450 km	1,000 km	405 km	40 km	520 km
14 drains	43 drains	25 drains		54 drains
440 MLD	3,270 MLD	580 MLD		1,780 MLD

Note: MLD: million litre per day (the figures refer to the collective discharge from the drains into the river)
Source: CPCB 2013, *Pollution Assessment: River Ganga*, Central Pollution Control Board, MoEF, July

Water-bodies of focus:
(1) Hindon, sub-tributary Yamuna, tributary of Ganga
(2) Yamuna, tributary of Ganga
(3) Ganga

Details in *AAAI 2017 Tutorial on AI and Water*:
https://sites.google.com/site/aiwatertutorial/

Usage Scenario – Tourism Impact

Reference:

An Open, Multi-Sensor, Dataset of Water Pollution of Ganga Basin and its Application to Understand Impact of Large Religious Gathering,
B Srivastava, S Sandha, V Raychoudhury, S Randhawa, V Kapoor, A Agrawal
arXiv preprint arXiv:1612.05626

Use-Case: Understand Impact of a Large-Scale Religious cum Tourism Event

Haridwar Ardh Khumbh Mela 2016

- January 1, 2016 to April 30, 2016
- Over 100 millions attended; Many took dip in river at select spots
- Major bath sub-events during the period have high burst of visitors

Question

- How much does human activity impact river?
- Where is the impact highest? Of what kind?

Date (2016)	Day	Main Bathing Event (Snan)
14th January	Thursday	Makar Sankranti
12th February	Friday	Vasant Panchami
22nd February	Monday	Magh Purnima
7th March	Monday	Mahashivratri
7th April	Thursday	Chaitra Amavasya
8th April	Friday	Chaitra Shukla Pratipada
14th April	Thursday	Mesha Sankranti
15th April	Friday	Ram Navami
22nd April	Friday	Chaitra shukla Purnima

Sources:

1. https://en.wikipedia.org/wiki/Kumbh_Mela
2. http://www.kumbhamela.net/kumbha-mela-haridwar.html
2. http://www.thegreatananda.com/ardh-kumbh-mela-2016-haridwar/

Ardh Kumbh 2016, Haridwar

TABLE I WATER STANDARDS FOR DIFFERENT ACTIVITIES

Activity	pH	Conductivity (µS/cm)	DO (mg/L)	Turbidity (FNU)
Drinking	6.5-8.5	≤ 2250	≥ 6	1-5
Bathing	6.5-8.5	--	≥ 5	--
Irrigation	6.0-8.5	≤ 2250	≥ 6	1-5

Territorial Bird's Eye View: ~76 KM (Road Distance)

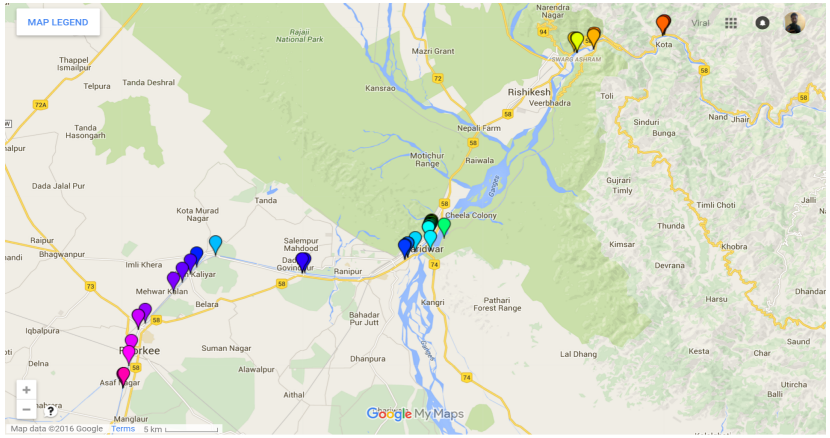
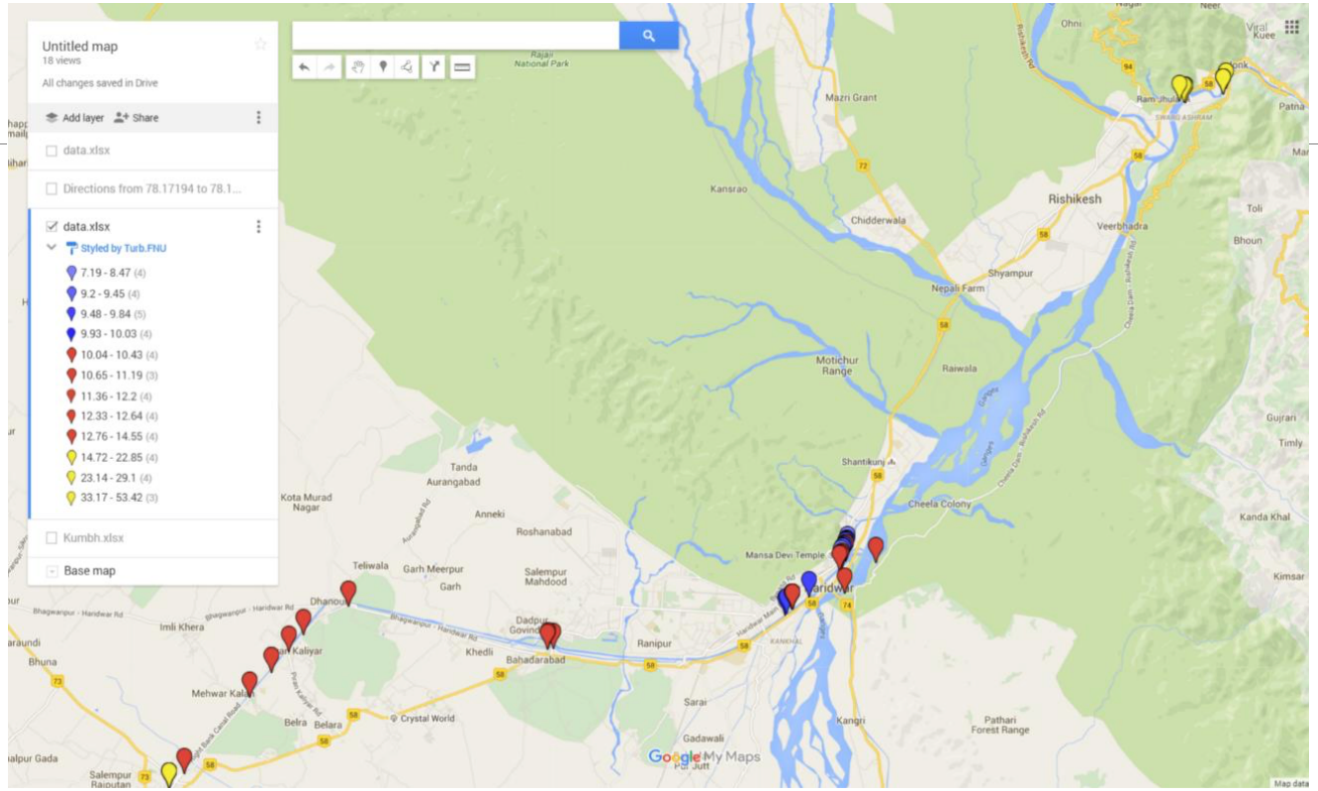


TABLE III DATA COLLECTION DATES

Date	Kumbh Bathing Day/ Other day	Place	Time
Feb. 27	Other day	Haridwar and Roorkee	3-7 PM
Feb. 28	Other day	Shivpuri, Rishikesh, Roorkee	11 AM - 6 PM
Mar. 07	Bathing Day	Haridwar	6-10 AM
Apr. 07	Bathing Day	Haridwar	1-4 PM
Apr. 15	Bathing Day	Haridwar	7-10 AM

Turbidity Variations

Feb 27-28, 2016

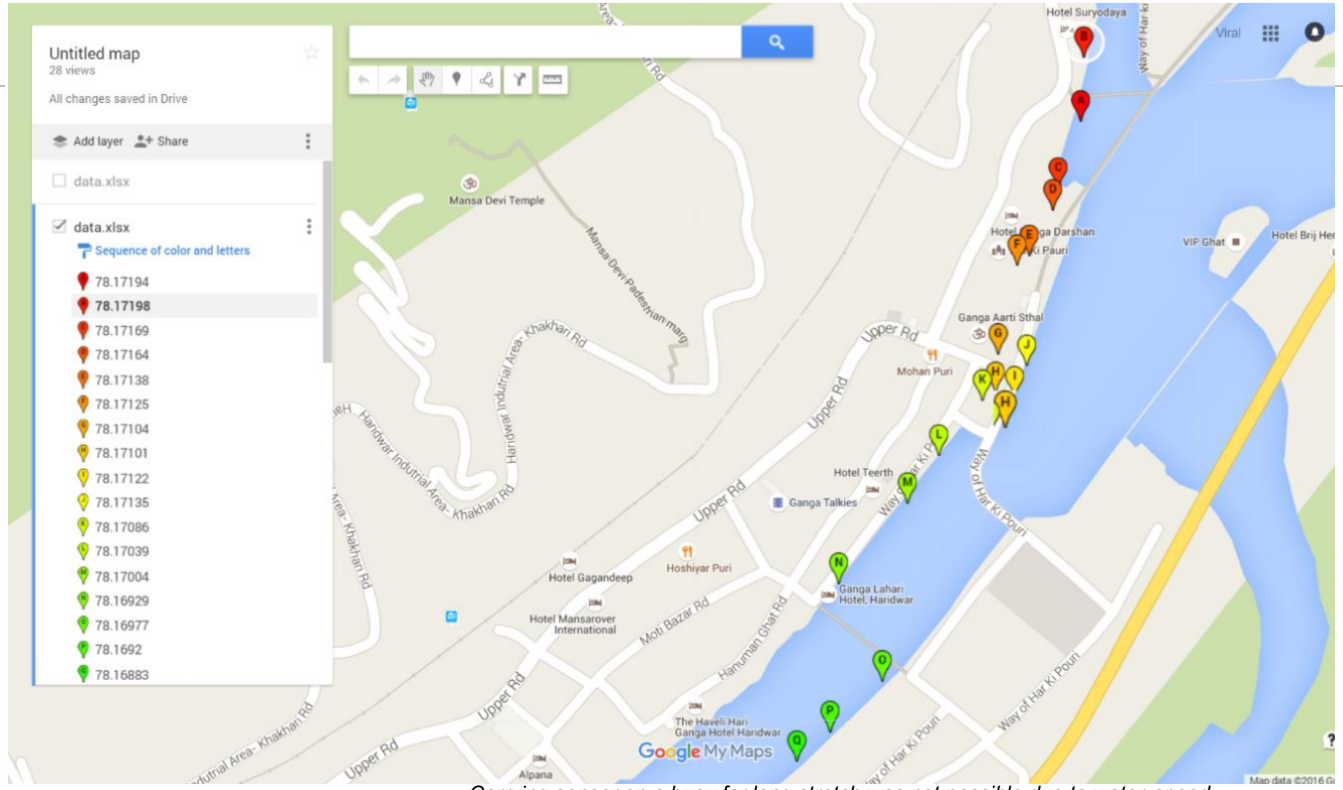


Turbidity values at different places (places marked red have turbidity value above the drinking range, places marked blues ha turbidity value in range of drinking water)

Data Collection Points around Har-ki-pauri, Haridwar

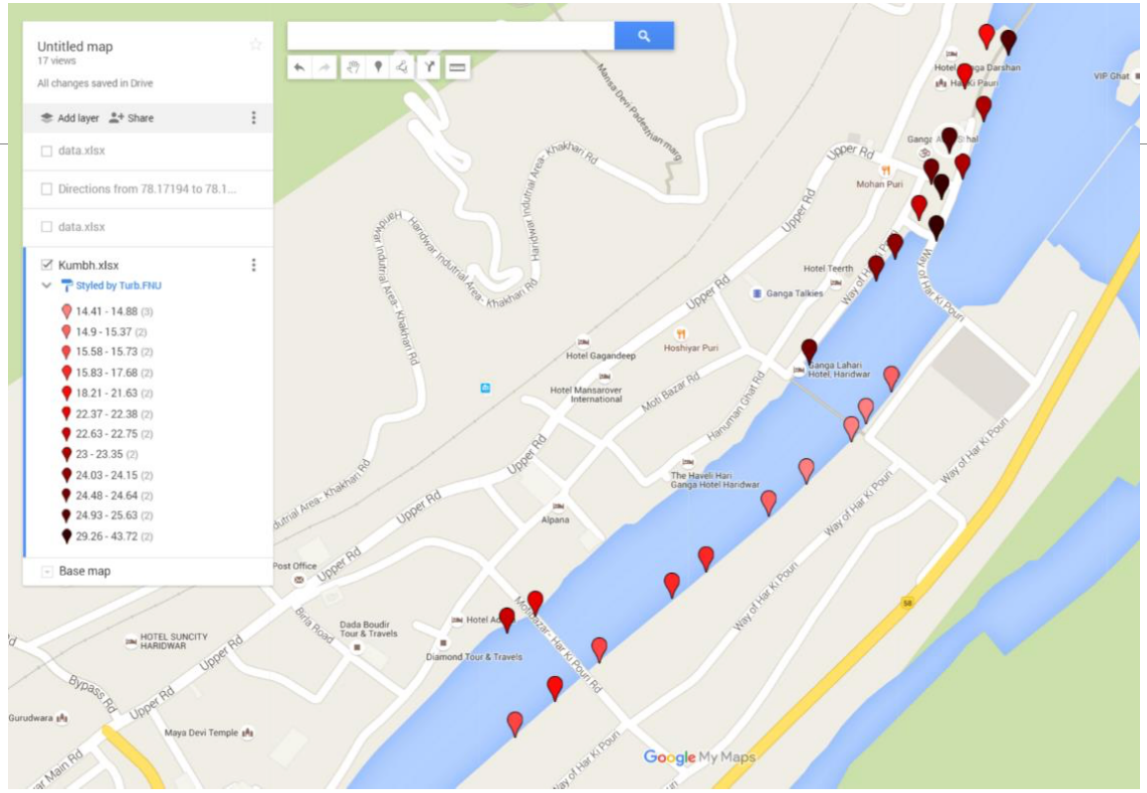
Feb 27-28, 2016

45+ places from Rishikesh to Ganga Canal (Roorkee) (75+ KM)



Carrying sensor on a buoy for long stretch was not possible due to water speed.

Pollution on Major Bath Day around Har-ki-pauri, Haridwar
March 7, 2016



Turbidity values at different places (places marked red have turbidity value above the drinking range, places marked blues ha turbidity value in range of drinking water)

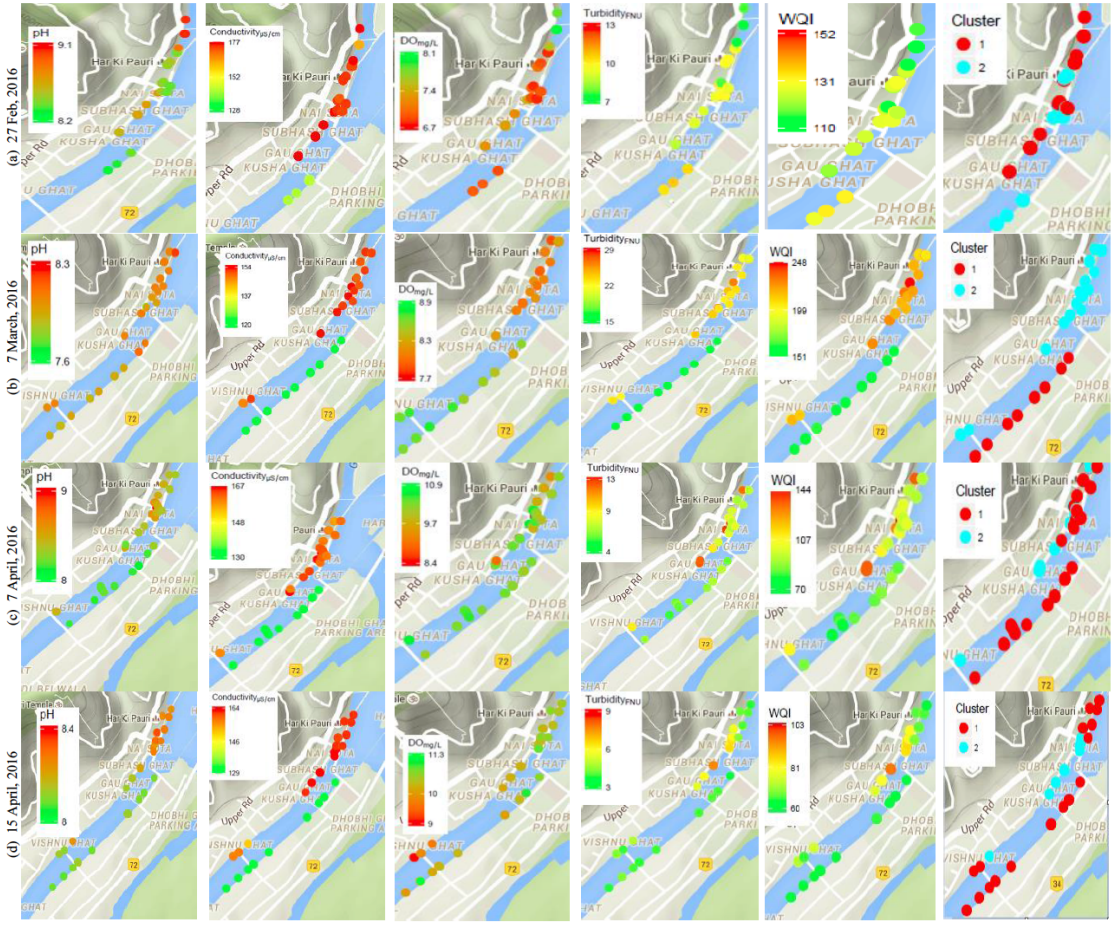


Figure 5 Heat Maps to Show Day-wise (a-d) variation of pH, Conductivity, DO, Turbidity, WQI, and Cluters during Ardh Kumbh 2016

AI in the Scenario

Data cleaning, normalization, missing values

- Quantitative to qualitative data conversation
  - Water Quality Index
- Aggregate Binary Clustering using parameters with opposite polarities (e.g., Dissolved Oxygen, Turbidity), interval functions (e.g., pH)

Advanced

- Generating consolidated qualitative assessment across multiple parameters
- Explaining and validating assessment

Analytics: Potential Use Cases

S. No.	Stakeholder	Use case	Data	Analytical techniques
1	IT	Identifying and removing outliers, data validation	Sensor data	Data mining (outlier detection)
2	Individual	Which bathing site to use?	Sensor data, ghat data	Rule-based decision support
3	Individual/ Economy	What crops can I grow that will flourish in available water?	Sensor data, crop data	Distributed data integration, co-relation
4	Institution	Determine trends/anomalies in pollution levels	Sensor data, weather data	Time series analysis, anomaly detection
5	Institution	Attribute source of pollution at a location	Sensor data, demographics, industry data	Physical modeling, inversion
6	Institution	Sewage treatment strategy and operational planning	Sensor data, demographics data, STP data	Multi-objective optimization
7	Institution	Promoting wildlife/ dolphins	Sensor data, wildlife data	Rule-based decision support

Exercise: Discussion AI in Water Problems Identified in South Carolina

- Data
- AI techniques
  - Representation
  - Reasoning
  - Learning
  - Interaction with People

Lecture 4: Concluding Comments

- We covered a background in water
  - Looked at common problems
- Looked at decision-support situations from US and a few solutions
- Looked at decision-support situations from India and a few solutions
- We explored the scope of AI

Concluding Segment

Think of Project and Act

- Look at problems of the real world around
- Identify
  - Value of decision: before and after
  - Data-needed
  - Method
  - Evaluation
  - Integrating with overall process
- Enter in student spreadsheet

About Next Lecture – Lecture 4

Lecture 4: Structured Data

- Data Preparation
- Knowledge Graph
  - Standardizing data
  - Annotating for discovery