Environmental Justice Analysis of Public Water Systems

Delineating Water System Service Boundaries for the U.S. Using Machine Learning Techniques

Andrew Murray | Alex Hall

US EPA Office of Research and Development

Center for Environmental Solutions & Emergency Response

WE DON'T KNOW WHO DRINKS WHAT WATER IN THE UNITED STATES

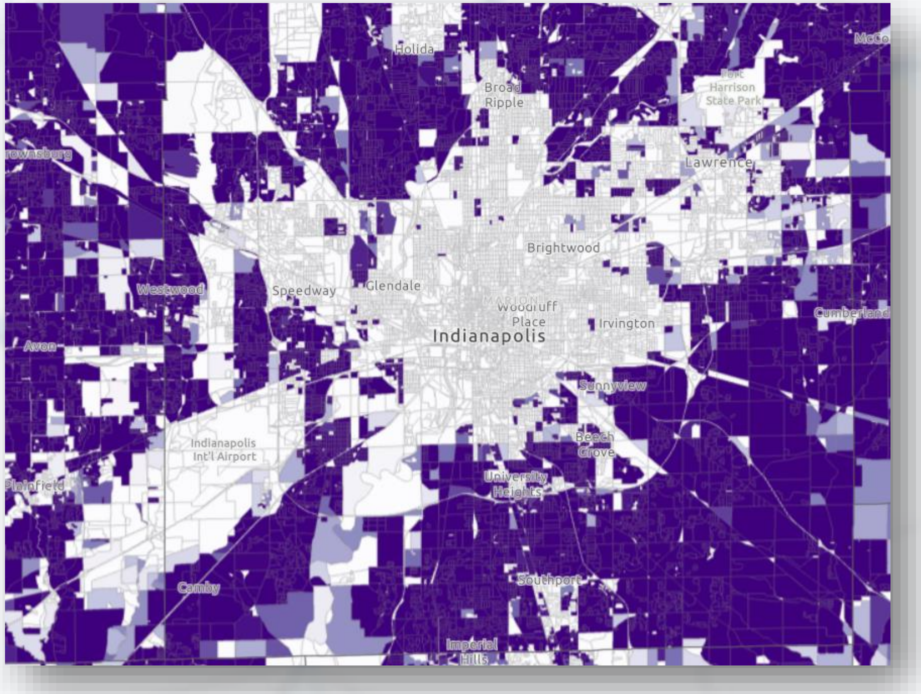
In 2020 ORD team published a dataset and paper solving half of that problem: the locations and density of **unregulated** water sources

- private domestic wells in the US—by Census block (right graphic)

We are currently working on where our **regulated** drinking water is distributed

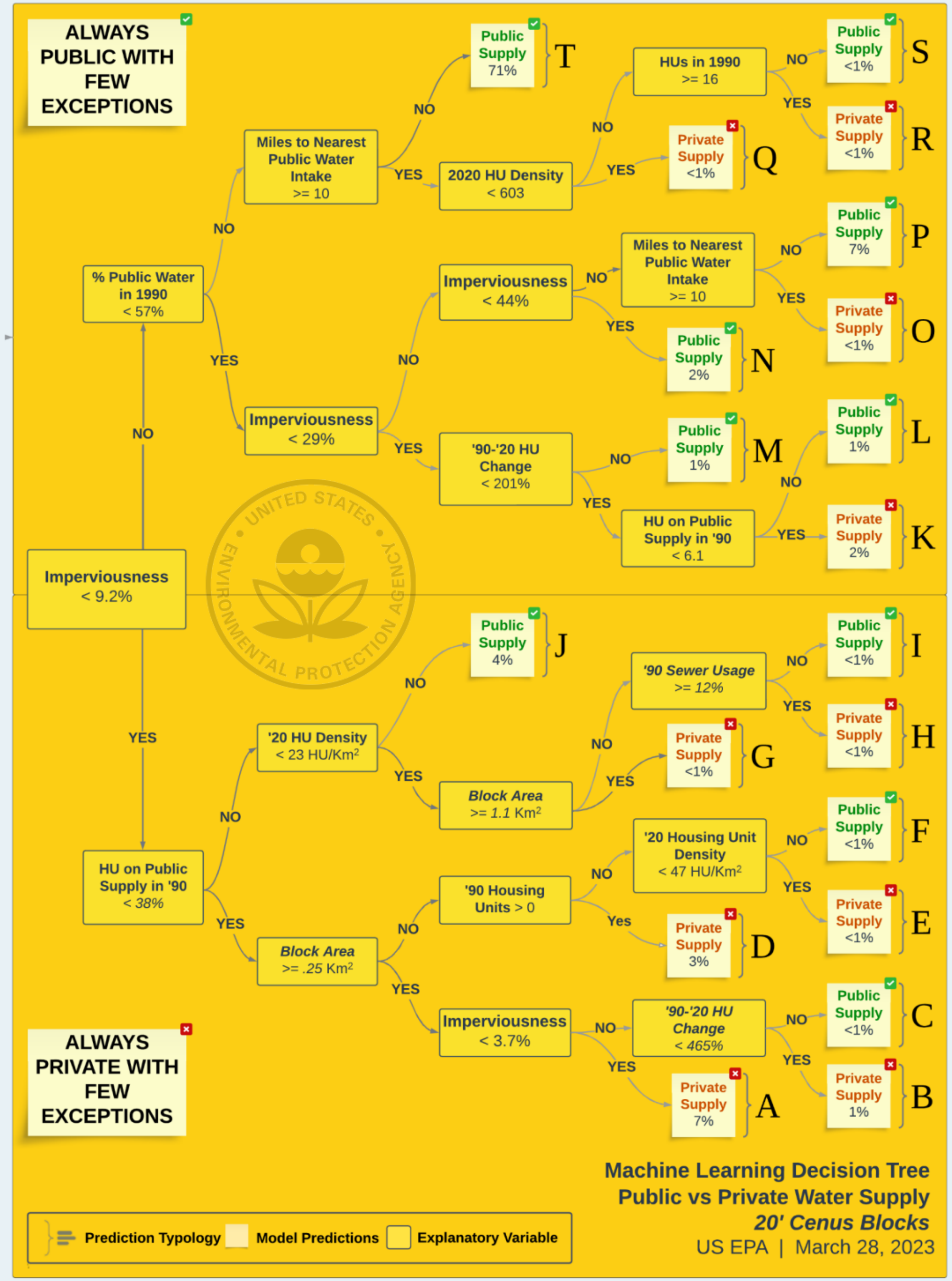
Utility of a nationally consistent, high resolution dataset of public water system boundaries

- Connect public water system violations to people/consumption on a national scale
- Identifying lead service lines to public water system boundaries on a national scale
- Understanding the environmental justice implications of impaired public water on a national scale
- Aid in identifying the potential need for public water infrastructure expansion (i.e., better allocation of resources to vulnerable populations)



PRIVATE WELLS/PUBLIC WATER

www.gispub.epa.gov/wellmap



Machine Learning Decision Tree
Public vs Private Water Supply
20' Cenus Blocks
US EPA | March 28, 2023

Model

Predicting Public Service Boundaries by Census Block

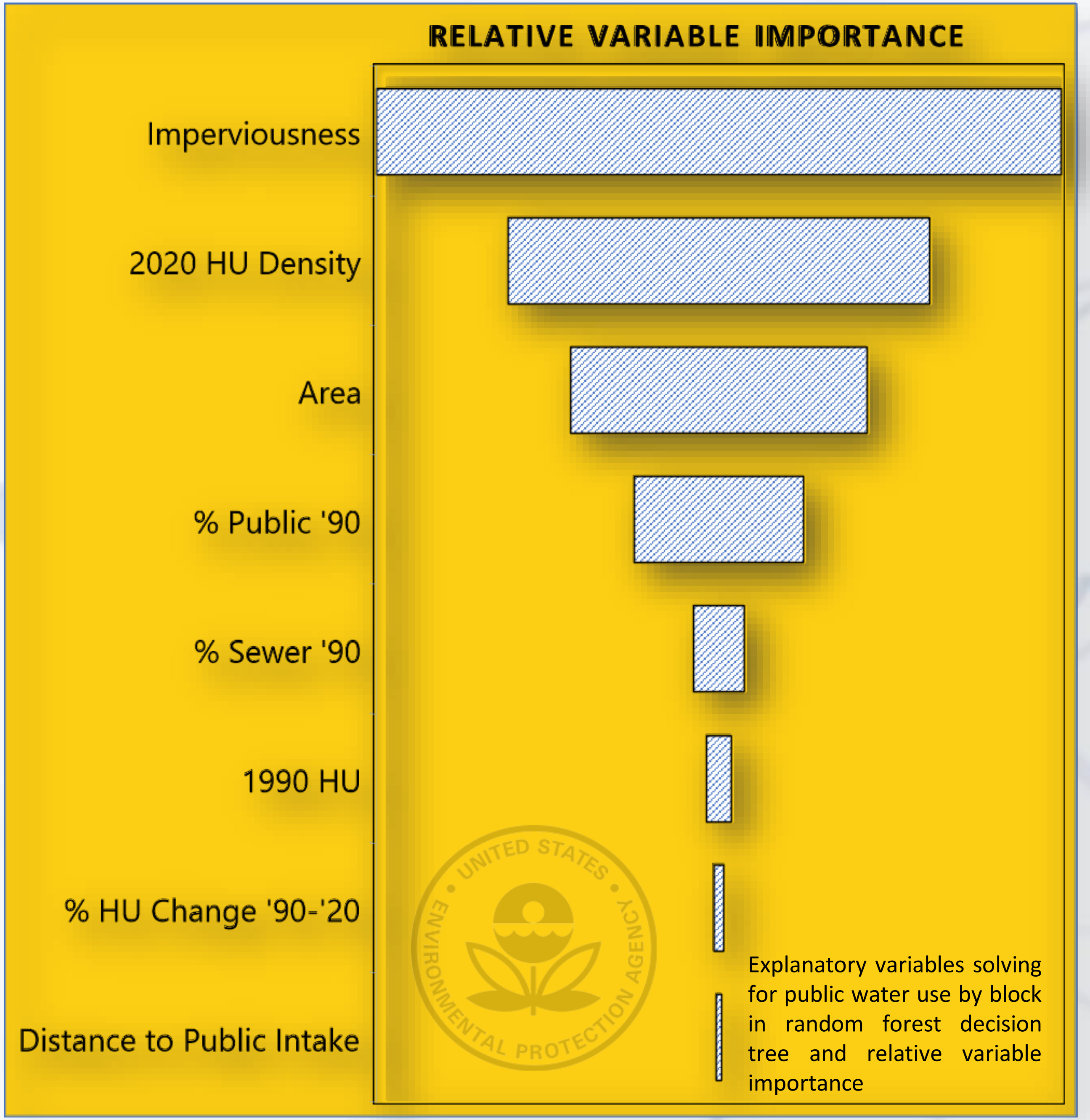
We have all the information we need to determine if a house is on public or private water

- However the confluence of factors and the partitioning of variables on certain hinge points on millions of data points is too complex for humans to define
- A machine learning decision tree solves for these almost infinite permutations and finds the correct fit and sequence of explanatory information to solve the problem

- Our model is trained and validated on 10% of all US blocks using 3 states
- 700,000 Census blocks in California, Connecticut, and New Jersey were used
- Using CA, CT, and NJ public service boundaries we identified:
  1. Blocks served by public water supply (between 1%-100% public)
  2. Blocks served by private water supply (100% private)
- Validated the model using every permutation of training and validation states (i.e., developed model using CT and CA blocks and trained on NJ; Developed model using CA and NJ and trained on CT)

Performance

Model Training State(s)	Validating State	Accuracy (R2)	Public Supply Accuracy (R2)
CT CA	NJ	94	96
CA	NJ	94	96
CT NJ	CA	93	96
CT	NJ	93	96
NJ	CA	93	96
CT	CA	92	96
NJ	CT	89	91
CA	CT	89	90
CA NJ	CT	89	96



Explanatory variables solving for public water use by block in random forest decision tree and relative variable importance

- 8 out of 10 variables were used in the model (excluded '20 HU & '90 HU Density)
- Created 20 unique community "*typologies*" based on 8 variables and 19 variable splits

4 Predominate *Typologies*

A (7%)—Homes on private wells: Very low imperviousness; not likely to be on public supply in 1990; large block area

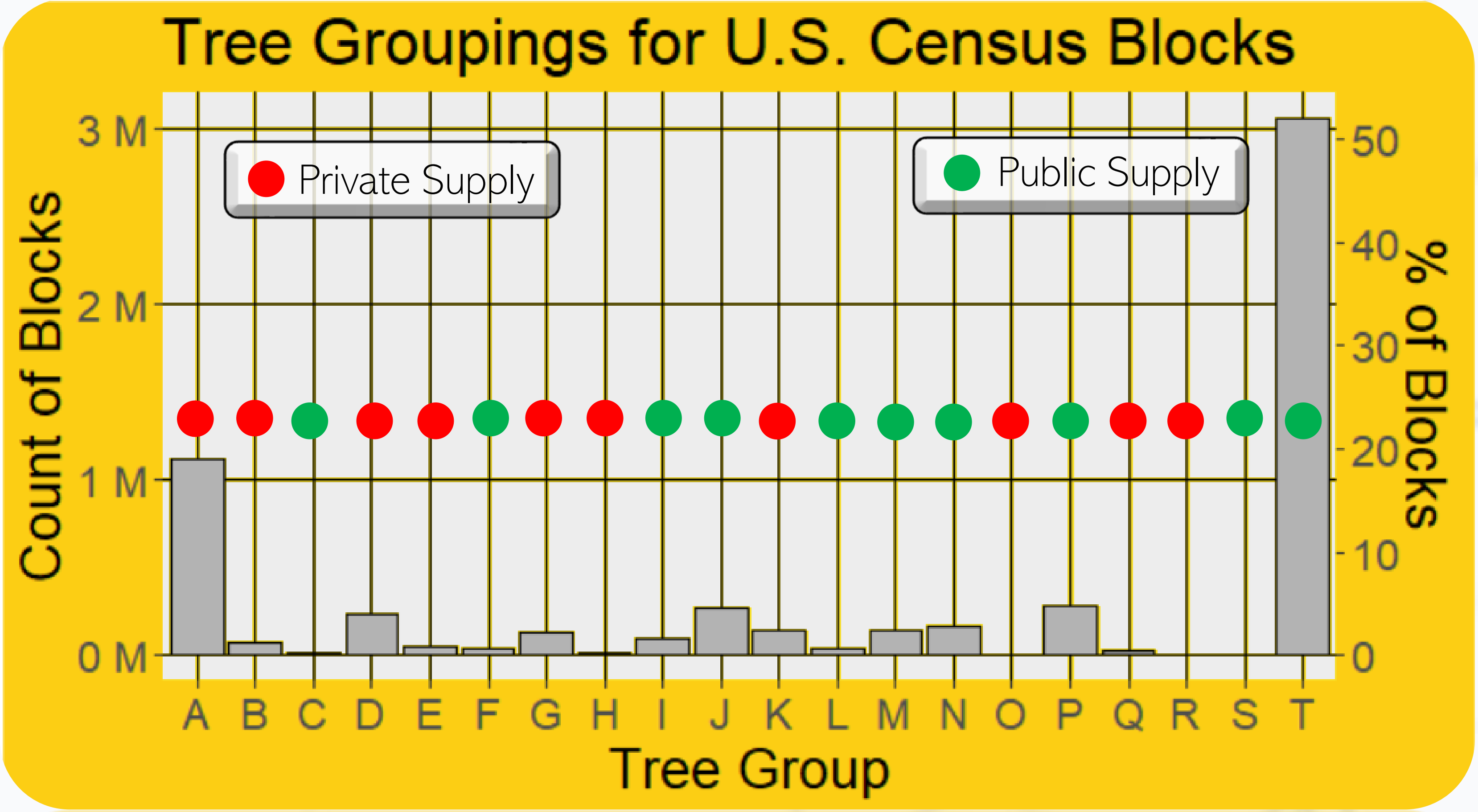
J (4%)—Residential, Low Development Urban Areas: Fairly low imperviousness; higher HU density

P (7%)—Suburban expansion: High imperviousness; minimal public supply in 1990; close to a public drinking water intake

T (71%)—Typical Urban Environment: High imperviousness; predominately public supply in 1990; close to a public drinking water intake

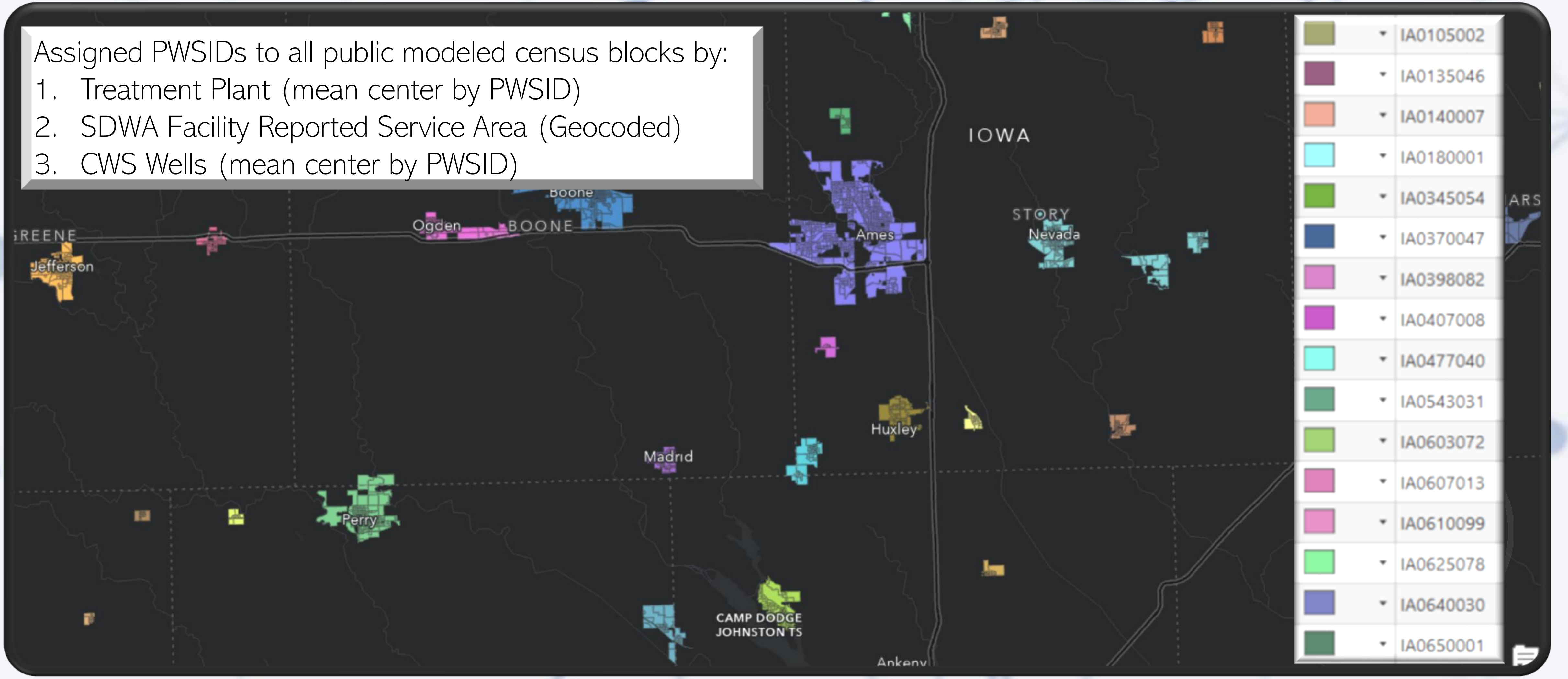
- These 4 types make up 89% of all community types
- The other 16 typologies are rare exceptions to these 4 rules

Model Applied to All States



- Group A went from 7% in the training set to over 20% when applied to the U.S.
- Concomitantly, Group T went from 70% in the training set to 50% when applied to the U.S.
- This is a result of the model better reflecting the U.S. demographics as a whole

Public Water System ID Assignment: 585 Unique Community Water Systems Delineated in Iowa



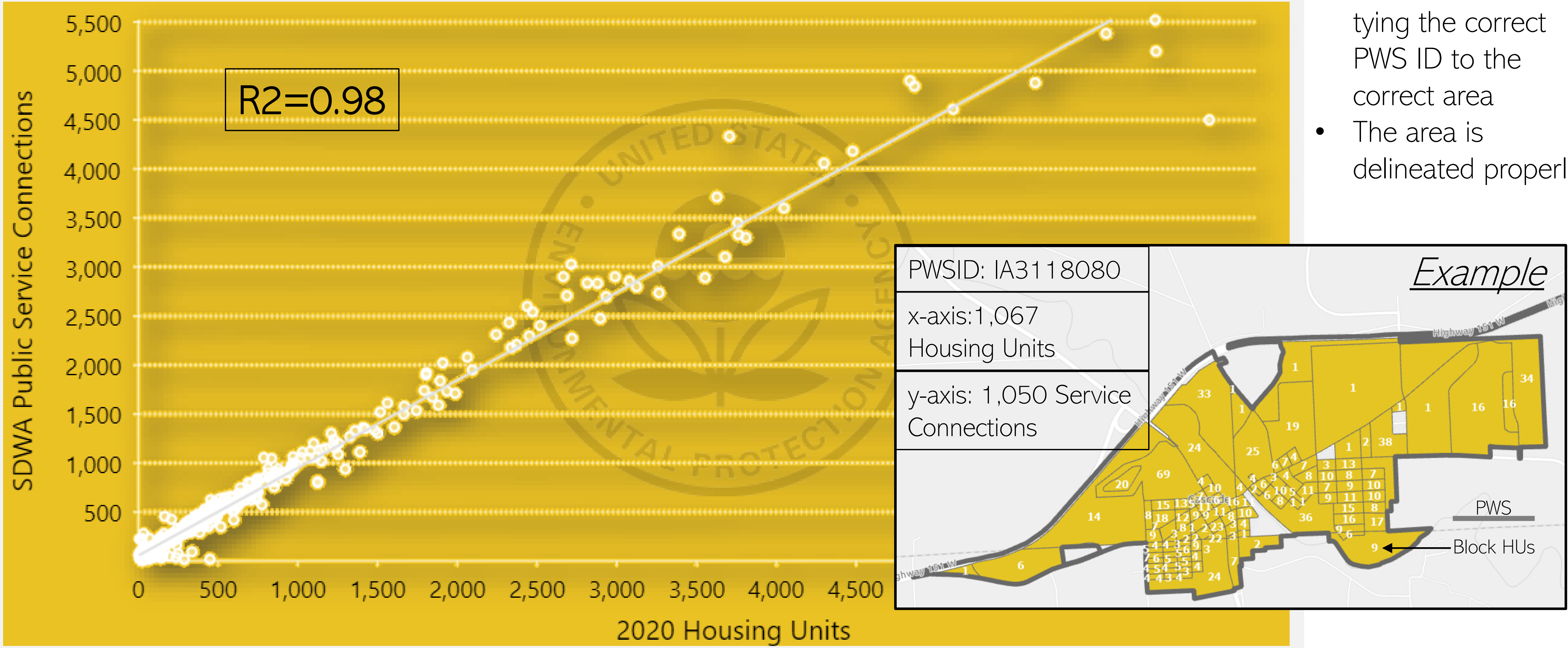
Assigned PWSIDs to all public modeled census blocks by:

1. Treatment Plant (mean center by PWSID)
2. SDWA Facility Reported Service Area (Geocoded)
3. CWS Wells (mean center by PWSID)

■	IA0105002
■	IA0135046
■	IA0140007
■	IA0180001
■	IA0345054
■	IA0370047
■	IA0398082
■	IA0407008
■	IA0477040
■	IA0543031
■	IA0603072
■	IA0607013
■	IA0610099
■	IA0625078
■	IA0640030
■	IA0650001

PWSID Allocation Validation

SDWA Community Water System Service Connections vs. 2020 Census Block Housing Units | Iowa CWS: n=585



Example

PWSID: IA3118080

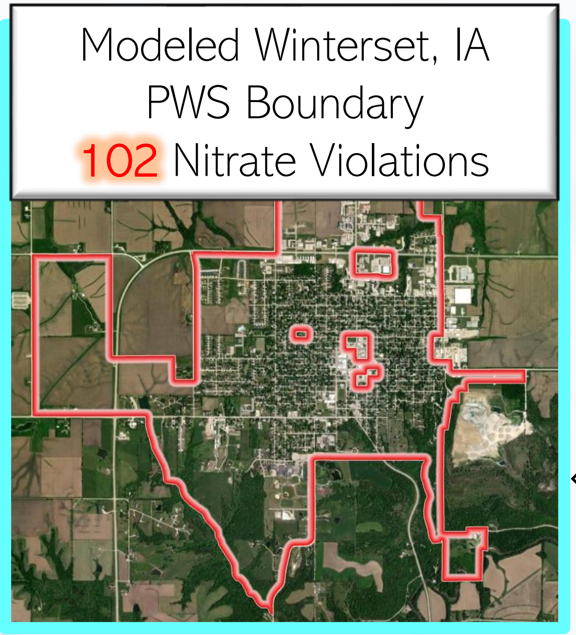
x-axis:1,067 Housing Units

y-axis: 1,050 Service Connections

Validates that:

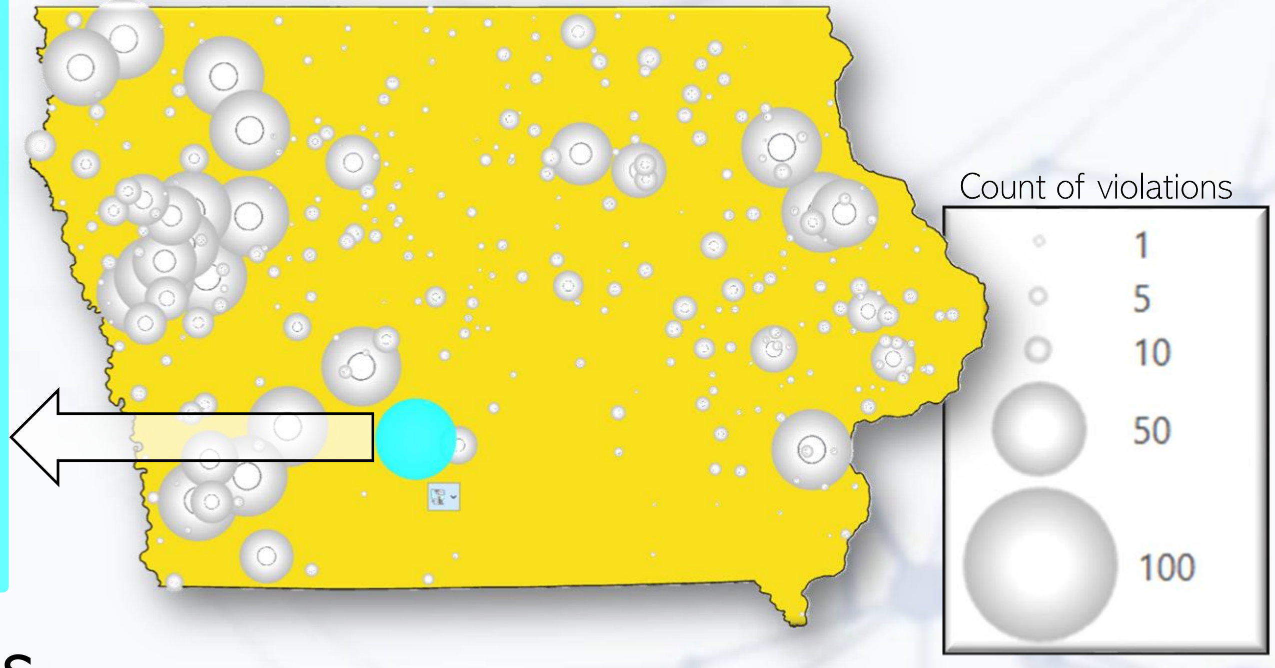
- We know we are tying the correct PWS ID to the correct area
- The area is delineated properly

Implications

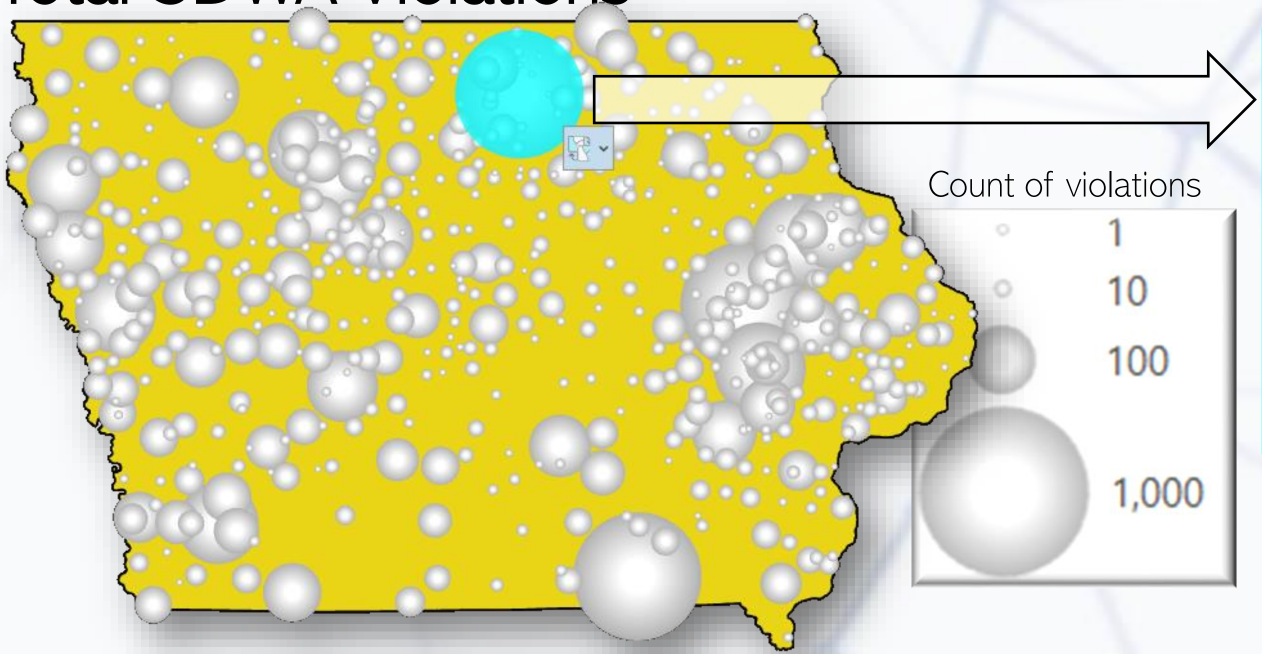


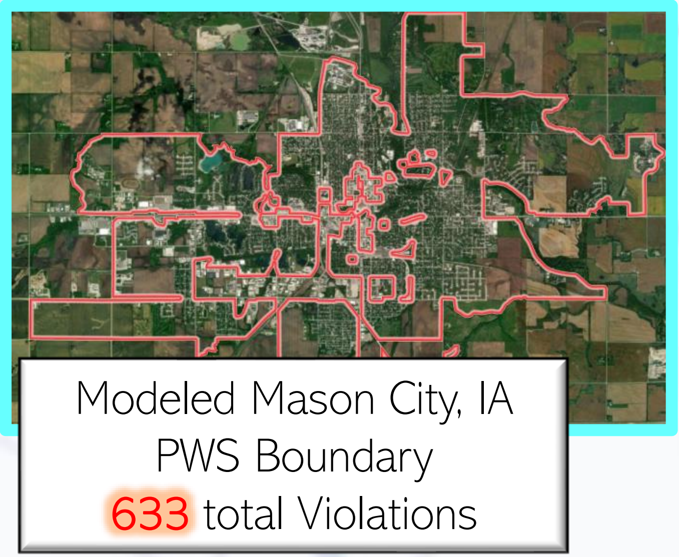
Modeled Winterset, IA
PWS Boundary
102 Nitrate Violations

SDWA Nitrate Violations



Total SDWA Violations





Modeled Mason City, IA
PWS Boundary
633 total Violations

- We can say with 90%+ confidence which house has been drinking what water with what SDWA violation in Iowa
- Because these boundaries are by Census blocks they can be integrated with EJScreen (block groups) to look at EJ issues