National trends in drinking water quality violations in the United States: 2009-2019

by Prajwal Seth (ps3190)

advised by Dr. Upmanu Lall

QMSS Master's Thesis, Fall 2021

Table of Contents

A. Abstract

Guaranteeing access to safe drinking water to communities across the U.S. remains a massive challenge in 2022 in the face of aging infrastructure, legacy pollution and financial struggles. Deadly outbreaks such as the 2015 Flint water crisis and the subsequent discovery of high amounts of lead in drinking water in Newark, NJ remind us of the pervasiveness of water quality violations. Very few nationwide assessments have been conducted on recent trends in drinking water quality in the United States. Studies that illuminate where violations could occur are of crucial importance, since in 2015 approximately 21 million people relied on community water systems that infringed health-based water quality standards ("Allaire, Maura, et al. 2018", pp. 2078). This paper assesses trends in drinking water quality violations in the continental U.S. from 2009-2019 by focusing on violations which are included in the Safe Drinking Water Information System (SDWIS). The dataset analyzed in this paper includes 201,397 health-based violations from 28,644 Community Water Systems (CWS's), serving approximately 94% of the population of the United States in some part. Probit regression models are created to gauge whether susceptibility factors of communities (such as median income, housing density, non-white percent, etc.) are correlated with drinking water quality violations. Concerningly, increasing time trends are found for health-based violations and total coliform violations across almost all states. Significant hot spots of health-based violations are detected in many regions, particularly in the south. Similarly, hot spots are detected for total coliform violations in the south and midwest.

B. Introduction

As COVID-19 rapidly evolves and continues its surge across the planet[1], there has never been a time before when more people stayed at home. Increases in hygiene standards such as washing hands with soap and water, cleaning floors and utensils, and regular showers have led to an estimated 21% increase[2] in household water consumption in the U.S. ("Campos, Marcus André Siqueira, et al."). As a result, it is crucial that the water that we consume is free from toxic contaminants, and that we are made aware if a drinking water quality violation occurs. Large-scale contamination of water sources is a pressing issue. For example, a $1.3 million fine was recently given to the Port of Morrow along the Columbia River in Oregon for over 1,000 nitrate violations from 2018 to 2021[3]. In light of such concerns, this paper makes an attempt to analyze drinking water quality violations in the continental U.S. for the years 2009-2019. It follows the footsteps of the 2018 paper 'National trends in drinking water quality violations' by Maura Allaire et al.[4], which achieved the same task for the time period 1982-2015.

In June 1974 the U.S. Congress passed the Safe Drinking Water Act (SDWA) and compliance with its standards came into effect in June 1977. The SDWA sets upper limits (called 'maximum contaminant levels' i.e. MCLs) for 90+ contaminants, treatment techniques, and monitoring and reporting methods for drinking water across the U.S. It also enables states to set their own standards for drinking water quality as long as these standards are at least as strict as those set by the national SDWA. Due to the SDWA, Public Water Systems (PWS's) are required to frequently monitor contaminants in source water, treated water, and distributed water. The

[1] https://covid19.who.int/
[2] https://www.phyn.com/press/residential-water-consumption-spikes-during-covid-19-pandemic/
[3] https://www.oregonlive.com/environment/2022/01/port-on-columbia-river-fined-13m-over-nitrate-violations.html
[4] https://www.pnas.org/content/pnas/115/9/2078.full.pdf

largest subcategory of Public Water Systems are Community Water Systems (CWS's). CWS's provide water to the majority of the U.S. population; approximately 94% of the U.S. population received at least some water from a CWS in 2019[5]. This paper focuses on water quality violations in CWS's only.

The Safe Drinking Water Information System (SDWIS) is a public database maintained by the U.S. Environmental Protection Agency that hosts water quality violations for all CWS's in the country. All CWS's are required to report SDWA violations to one of many primary agencies (such as states, territories, and Native American tribal agencies), which forward this information to the EPA for incorporation into the SDWIS. Some violations of the SDWA rules are termed as 'health-based' violations. Examples of health-based violations include failures in CWS operations that can jeopardize public health, such exceeding the MCL for chemical contaminants (e.g. arsenic, nitrate, lead) and microbes (e.g. total coliforms). Other examples of health-based violations include non-compliance with treatment techniques (e.g. the surface water treatment rule), and exceeding the maximum residual disinfectant level (MRDL). Health-based violations have the potential to cause immediate illness (such as E. coli bacteria leading to stomach cramps, bloody diarrhea and vomiting[6]) or long-term health complications due to prolonged exposure to toxic chemicals[7]. This paper focuses on health-based violations, and uses quantitative methods to identify variables that correlate with their occurences. It also includes hot spot maps of the continental U.S. where such violations occurred in 2009-2019.

[5] https://cfpub.epa.gov/roe/indicator.cfm?i=45
[6] https://www.mayoclinic.org/diseases-conditions/e-coli/symptoms-causes/syc-20372058
[7] https://www.pnas.org/content/pnas/suppl/2019/09/24/1905385116.DCSupplemental/pnas.1905385116.sapp.pdf pp. 13

C. Relevant Works

The 2018 paper 'National trends in drinking water quality violations' by Maura Allaire, Haowei Wu and Upmanu Lall serves as the seminal paper in this domain (and is hence referred to as 'the original paper' in some sections). In it, the authors created a panel dataset of health-related violations from 17,900 Community Water Systems (CWS's) from 1982 to 2015. Decennial Census data and the American Community Survey data was used to add county-level covariates to the dataset of water quality violations. The authors employed probit regression and probit time series analysis to determine vulnerability factors associated with health-based violations and total coliform violations. Principal-component regression was also conducted to simplify the models and reduce multicollinearity. They determined that the most important indicator in predicting whether or not a CWS would face a violation in a given year was if it had reported a violation the year before. Rates of violation occurrence were found to be the highest for small-sized CWS's in rural areas. Compliance with SDWA rules was associated with purchased water sources and private ownership of CWS's. Hot spot analysis of health-based violations from 1982-2015 was also conducted, and the hot spots of health-based violations were found to shift over time. From 1982-1992 the hot spots were located in the Northwest, California, and Pennsylvania. From 1993-2003 the hot spots had shifted to Oklahoma, Tennessee and Idaho. Lastly, from 2004-2015 hot spots were found in Oklahoma and Texas.

Two of the authors of the aforementioned paper (Maura Allaire and Upmanu Lall) published a paper in 2019 called 'Detecting community response to water quality violations using bottled water sales'. In this paper, the authors created a panel dataset of health-based water quality violations and sales of bottled water for 2,151 counties from 2004 to 2015.

Approximately 95% of the population of the continental US was included in their dataset. A fixed effects model was used by them to estimate the change in weekly sales of bottled water due to poor water quality. A 14.1% increase in bottled water sales was determined for tier 1 violations (i.e. violations that posed an immediate health risk). A 4.9% increase in bottled water sales was found for tier 2 violations (i.e. violations that have the potential for adverse health consequences following prolonged exposure). Counties which faced repeat water quality violations faced an estimated 16.8% increase in bottled water sales for tier 1 violations and 5.8% increase for tier 2 violations. For pathogen violations only, both tier 1 and tier 2 violations were associated with a significant increase in bottled water sales. Bottled water sales were also found to spike during summer months. Alarmingly, the authors observed that rural, low-income counties did not take significant averting actions against tier 1 nitrate violations. It was posited that rural, low-income counties either resorted to boiling water as an aversive measure, or, concerningly, did not take any action. A limitation of the study was that the population of the continental US which was excluded from the study was significantly more rural than the paper's study sample. This would imply that the results noted above could not be generalized to the excluded counties.

An exploratory study of health violations in the entire SDWIS database (from its inception in 1978, to 2019) was conducted by Michielssen, Senne, et al. in their paper 'Trends in microbiological drinking water quality violations across the United States'. The researchers analyzed health-based water quality violations from all types of Public Water Systems (PWS's): Community Water Systems (CWS's), non-transient Non-community Water systems (NTNCWS's), and transient Non-community Water Systems (TNCWS's). In particular, an

emphasis was placed by the authors on microbiological regulations that applied to all public water systems. These were the total coliform rule (TCR) of 1990, and the revised total coliform rule (RTCR) of 2016. The authors found that Total Coliform Rule violations were the most common form of health-based violation, and encompassed 51.5% of all health-based violations. However, after the Total Coliform Rule was revised to the RTCR (Revised Total Coliform Rule) in April 2016, the number of such violations decreased substantially. They noted that for almost all of the top nine contaminants/violation rules, the number of violations increased substantially just after a new regulation became effective and gradually decreased as PWSs caught up to the deficiencies. Very small PWSs and TNCWSs (Transient Non-Community Water Systems) were found to be disproportionately associated with total coliform and E. coli violations. The transition to the RTCR exacerbated this trend for them for total coliform violations.

The 2021 paper ‘A critical review of point-of-use drinking water treatment in the United States’ by Wu, Jishan, et al. surveyed water quality regulations and violations in the U.S. The authors noted that domestic wells (i.e. private/homeowner wells) were the dominant source of drinking water for over 43 million people living in rural areas of the United States. The authors included maps showing the distribution of domestic wells in the United States, as well as how many wells were affected by arsenic violations. Unsettlingly, approximately 2.1 million people in the continental U.S. regularly drew water from private wells with arsenic concentrations of greater than 10 micrograms per litre, in 2021. The authors also called attention to the fact that there were many Contaminants of emerging concern (CECs) which are currently not regulated by the EPA. They mentioned that a USGS study found that 80% of streams in the U.S. contained some form of emerging contaminant such as pharmaceuticals, hormones, detergents, plasticizers,

fire retardants, pesticides, etc. Though the concentrations of these contaminants were low, they are closely linked to human diseases and are currently unregulated (“Bilal, Muhammad, et al.”).

The 2005 paper “Public or Private Drinking Water? The Effects of Ownership and Benchmark Competition on U.S. Water System Regulatory Compliance and Household Water Expenditures” asked the question: should private firms own water systems, or should governments? The authors employed a panel dataset which included every CWS in the U.S. from 1997 to 2003 to test the effects of ownership and competition on regulatory compliance. They found almost no difference between private and public water systems in such compliance. Privately owned systems were found to report fewer contaminant violations than government owned ones, but reported more monitoring and reporting violations. The results were inverted for water systems that served over 100,000 people: privately owned water systems reported fewer monitoring and reporting violations but more contaminant violations. The authors noted that competition among water systems was a significant reducer of violations (a theme we shall also observe later in this paper). Another measure that was found by the authors to decrease violations was if water systems were required to disclose water quality test results to consumers.

D. Data

The Safe Drinking Water Act (SDWA) violations and water systems data was downloaded from the Environmental Protection Agency's website[8]. The file name for the violations data is 'SDWA_VIOLATIONS_ENFORCEMENT.csv' and that for water systems data is 'SDWA_PUB_WATER_SYSTEMS.csv'. These files contain information for all violations and all water systems present in the Safe Drinking Water Information System (SDWIS) database. A data dictionary containing data types and column descriptions for all columns within these files can also be found on the EPA's website[9]. The file 'SDWA_REF_CODE_VALUES.csv' in the data download contains word descriptions for contaminant code numbers, and shall be used for analyzing the top contaminants in the dataset.

The water systems file (downloaded above) was filtered according to three criteria in accordance with the criteria in the paper by Maura Allaire et al., 2018[10]. The first criterion was that the water system in question should be a community water system (CWS). Hence, Transient non-community water systems (TNCWS) and Non-transient non-community water systems (NTNCWS) were excluded from this study. The second criterion was that the water system should have experienced at least one violation in 2009 or prior. Lastly, the water system should serve a population of at least 500 people. This ensures that water systems which are too small in scale do not act as outliers in the data. The total number of Community Water Systems (CWS's) that matched the filter criterion mentioned above were 28,644. The comparable number from the study from Maura Allaire et al.'s 2018 paper[11] was 17,900 CWS's.

[8] https://echo.epa.gov/files/echodownloads/SDWA_latest_downloads.zip
[9] https://echo.epa.gov/tools/data-downloads/sdwa-download-summary
[10] https://www.pnas.org/content/pnas/115/9/2078.full.pdf pp. 2079
[11] https://www.pnas.org/content/pnas/115/9/2078.full.pdf pp. 1

Filters were also applied to the violations dataset, consistent with the filters in the original paper. First, only violations from water systems that met the criteria listed above were included in the study. Second, the violations dataset was restricted to 'health-based' violations. Hence, only maximum contaminant level (MCL), maximum residual disinfectant level (MRDL), and treatment technique (TT) violations were included in the dataset. The violations dataset was also restricted to include only those violations that occurred between 2009 and 2019. A copy of the violations dataset was then made for total coliform violations. This included only those violations whose violation codes were 3100 (Coliform TCR), 3014 (E. COLI), 3013 (Fecal Coliform), 8000 (Revised Total Coliform Rule) or 3000 (Coliform Pre-TCR). Lastly, only violations from the continental United States were included in the dataset. Violations from Puerto Rico, Alaska, and Hawaii were excluded in this analysis.

5-year estimates of the American Community Survey (ACS) from 2009 to 2019 were downloaded from Social Explorer[12]. Using these files, the columns for non-white percentage, housing density, and median income for all counties were added to the water quality violations dataset. Each county's name and state was used to merge the ACS survey data into the SDWA water quality violations dataset for every year of the study.

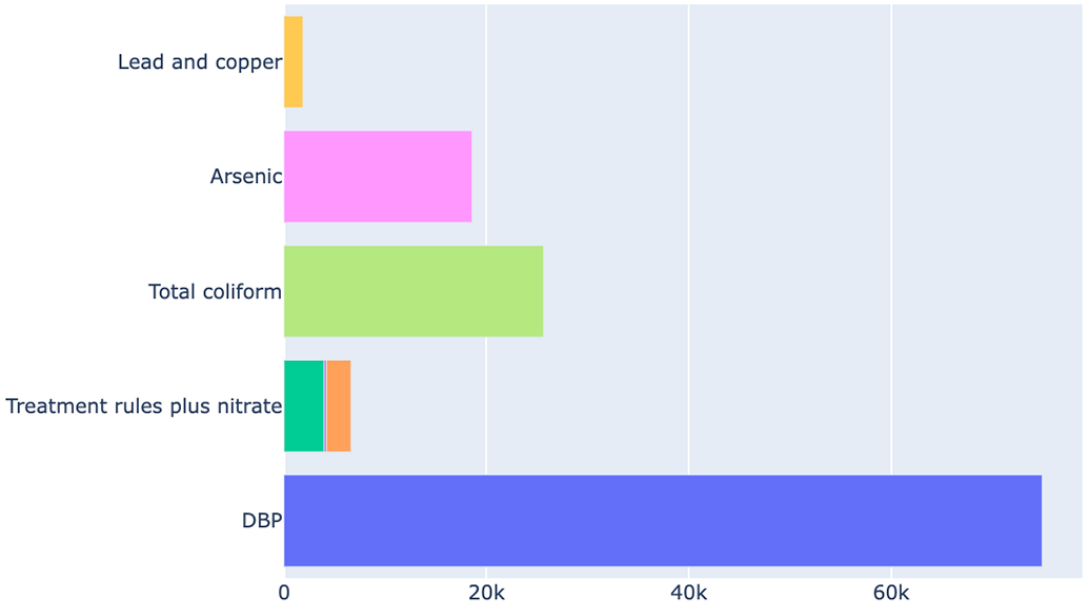
It was found that 7,842 CWS's reported at least one health-based violation from 2009-2019, and 3,962 CWS's reported at least one total coliform violation. The total number of health-based violations from 2009-2019 was 201,397 and the number of total coliform violations was 25,658. The top 10 states with water systems that had the highest number of health-based violations in the study period were: Texas (40,769 violations), Oklahoma (32,274 violations),

[12] https://www.socialexplorer.com/explore-tables

Louisiana (19,338 violations), Missouri (7,014 violations), Pennsylvania (5,797 violations), Kentucky (5,765 violations), California (5,572 violations), Arkansas (5,428 violations), Illinois (5,206 violations), and North Carolina (5,000 violations). The top 10 states with water systems that had the highest number of total coliform violations in the study period were: Texas (2,469 violations), Missouri (2,099 violations), Louisiana (1,857 violations), Nebraska (1,726 violations), Arkansas (1,263 violations), Mississippi (1,038 violations), Massachusetts (996 violations), Oklahoma (981 violations), California (903 violations) and Florida (762 violations).

The top 5 contaminants contributed to approximately 75% of all health-based violations. They were: Total trihalomethanes (TTHM) (34.8% of all health-based violations, 70,029 violations in total), Total Haloacetic Acids (HAA5) (14.5% of all health-based violations, 29,250 violations in total), Coliform (TCR) (12.1% of all health-based violations, 24,301 violations in total), Arsenic (9.23% of all health-based violations, 18,583 violations in total) and Combined Radium -226 and -228 (4.35% of all health-based violations, 8,758 violations in total). A better understanding of health-based violations by contaminant type can be gained from the following plot[13]:

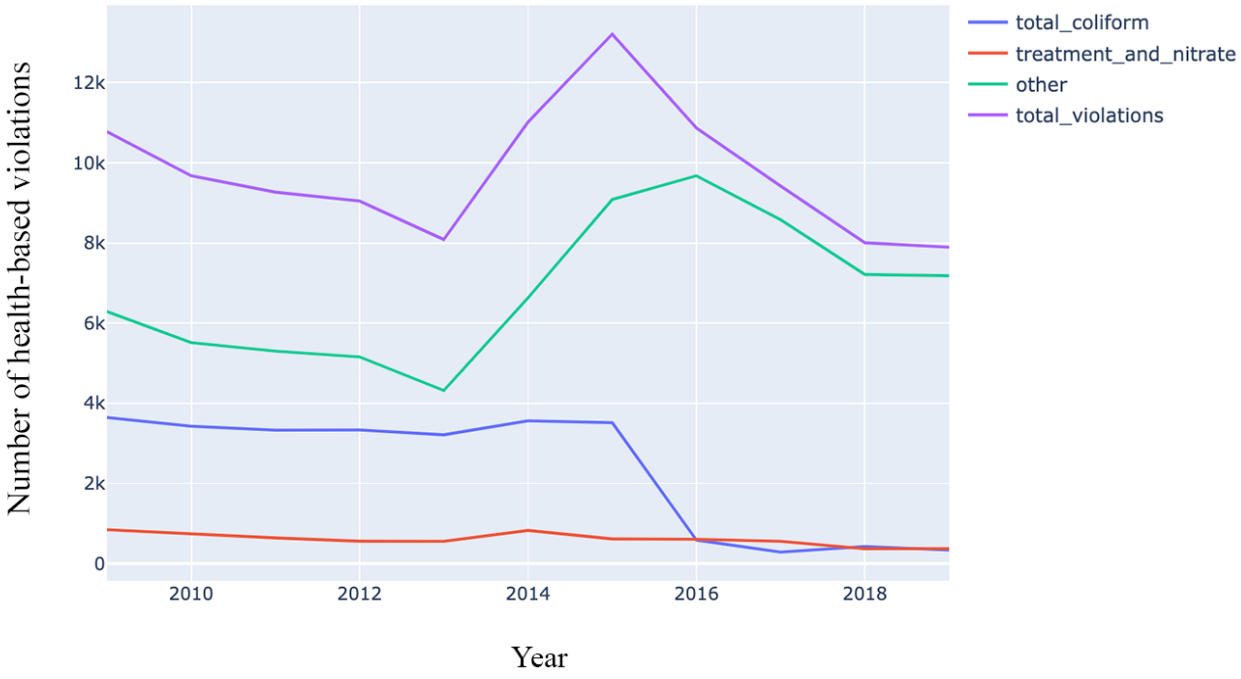
[13] The categories included in this figure are derived from Table 1 on pp. 2080 of Maura Allaire et al. 2018's paper. https://www.pnas.org/content/pnas/115/9/2078.full.pdf



Number of health-based violations from 2009-2019

We can see that among the five categories visualized, DBP violations dominated with 74,942 violations. Total coliform violations were the second most common form of health-based violation among the above categories, having 25,658 violations. Arsenic was third, with 18,583 violations. 'Treatment rules plus nitrate' was fourth, and within it the green bar represents 'Nitrate' with 3,968 violations, the orange bar represents 'Nitrate-Nitrite' with 2,483 violations, and the purple bar in the middle represents 'Nitrite' violations with 218 violations. The total number of 'Treatment rules plus nitrate' violations was 6,669. Lead and copper were the least common form of violation among these categories, with 1,870 violations in total. Note that the original paper by Maura Allaire et al., 2018 only included violations belonging to the above five categories of contaminants in their dataset of health-based violations. On the contrary, I decided to include every single health-based violation from 2009-2019 in my dataset, irrespective of whether or not the violation's contaminant code belonged to these five categories.

The time trend of violations grouped by contaminants can be seen in the following visualization:



The purple line represents total violations, the green line represents 'other' violations (explained in next sentence), the blue line represents total coliform violations, and the red line represents treatment and nitrate violations. The variable called 'other' includes DBP violations and violations belonging to Radionuclides and Inorganic and organic chemicals. We see a spike in these 'other' violations from 2013 onwards, when the stage 2 of DBP rules became enforceable.

MCL (Maximum contaminant level) violations were the most prevalent form of health-based violations, with 172,424 violations belonging to this category. The second most prevalent form of health-based violations were Treatment technique (TT) violations, with 28,749 violations. Maximum Residual Disinfectant Level (MRDL) violations were extremely infrequent, with only 224 such violations during the study period.

Variable descriptions

The water quality violations dataset was aggregated by the columns 'PWSID' (which is a unique code assigned to every CWS) and 'year'. The total number of pairs of CWS's and years in the study period was 307,219. This means that, for example, if a CWS had a unique PWSID of 'NY12345', then the number of such pairs of 'NY12345' and a year in the study sample was 307,219 in total. If a CWS did not incur a violation in a given year, it still had a row included for it for that year (the dependent variables took on a value of 0 in such a case). Hence, a CWS did not need to have a violation in a given year for it to be included in the dataset. Health-based violations were observed in about 6.03% of the 307,219 CWS-year observations, while total coliform violations were observed in about 1.90% of all CWS-year observations. Each row in the aggregated violations dataset had independent (X) and dependent (Y) variable columns. The independent variables (X) that were included in the dataset for every CWS for each year (from 2009-2019) were:

- nonwhite_percent: calculated as 1 minus the ratio of white population (column 'A03001_002' in ACS data) to total population (column 'A03001_001' in ACS data) in the county that the CWS belongs to
- housing_density: calculated as the number of households by household type in the CWS's county (column 'A10008' in ACS data) divided by the county's land area in sq. miles (column 'A00003' in ACS data)
- median_income: the median household income of the county that the CWS belongs to, adjusted for inflation (column 'A14006' in ACS data)

- prev_yr_all_viol: takes a value of 1 if the CWS incurred a health-based violation the previous year, otherwise takes a value of 0
- prev_yr_coliform_viol: takes a value of 1 if the CWS incurred a total coliform violation the previous year, otherwise takes a value of 0
- is_private: takes a value of 1 if the CWS is privately owned, otherwise takes a value of 0
- utility_small: takes a value of 1 if the CWS serves less than 3301 people, otherwise takes a value of 0
- utility_medium: takes a value of 1 if the CWS serves between 3301 and 9999 people, otherwise takes a value of 0
- utility_large: takes a value of 1 if the CWS serves between greater than 10,000 people, otherwise takes a value of 0
- purchased: takes a value of 1 if the CWS's primary water source is purchased water, otherwise takes a value of 0
- surface_water: takes a value of 1 if the CWS's primary water source is surface water, otherwise takes a value of 0
- ground_water: takes a value of 1 if the CWS's primary water source is groundwater, otherwise takes a value of 0
- hhi: the Herfindahl–Hirschman index (HHI) of the county, calculated as the sum of squared market share of each CWS in a county. For example, if a county had a total population of 300 and had two CWS's serving 100 and 200 people each, then the hhi of the county would be (100/300)^2 + (200/300)^2 = 0.555.
- violation_year: the year in which the violation occurred in that CWS

- ln_median_income: natural log of ‘median_income’
- ln_housing_density: natural log of ‘housing_density’

And the dependent variables (Y) included in each row were:

- had_violation_this_year: takes a value of 1 if the CWS incurred at least one health-based violation that year, otherwise takes a value of 0
- had_coliform_violation_this_year: takes a value of 1 if the CWS incurred at least one total coliform violation that year, otherwise takes a value of 0

Below are tables which provide summary statistics of all the aforementioned variables.

Summary statistics - Total violations

variable	*count*	*mean*	*std*	*min*	*max*
nonwhite_percent	307205	0.200253	0.154603	0	0.916724
housing_density	307205	186.77	402.663	0.139181	33514.7
median_income	307205	51905	14381.9	18869	142299
prev_yr_all_viol	307205	0.0630328	0.243022	0	1
is_private	307205	0.21888	0.413488	0	1
utility_small	307205	0.623785	0.484436	0	1
utility_medium	307205	0.204167	0.403092	0	1
utility_large	307205	0.172048	0.377423	0	1
purchased	307205	0.286769	0.452254	0	1
surface_water	307205	0.13645	0.343266	0	1
ground_water	307205	0.565108	0.495744	0	1
hhi	307205	0.280794	0.19269	0.0304018	1
violation_year	307205	2014	3.16223	2009	2019
ln_median_income	307205	10.8216	0.264204	9.84528	11.8657
ln_housing_density	307205	4.01933	1.66	-1.97198	10.4197
Y	307205	0.0603083	0.238057	0	1

The dependent variable (Y) in this case represents whether or not the CWS incurred a health-based violation in a given year.

Summary statistics - Total coliform violations

variable	*count*	*mean*	*std*	*min*	*max*
nonwhite_percent	307205	0.200253	0.154603	0	0.916724
housing_density	307205	186.77	402.663	0.139181	33514.7
median_income	307205	51905	14381.9	18869	142299
prev_yr_coliform_viol	307205	0.0215329	0.145153	0	1
is_private	307205	0.21888	0.413488	0	1
utility_small	307205	0.623785	0.484436	0	1
utility_medium	307205	0.204167	0.403092	0	1
utility_large	307205	0.172048	0.377423	0	1
purchased	307205	0.286769	0.452254	0	1
surface_water	307205	0.13645	0.343266	0	1
ground_water	307205	0.565108	0.495744	0	1
hhi	307205	0.280794	0.19269	0.0304018	1
violation_year	307205	2014	3.16223	2009	2019
ln_median_income	307205	10.8216	0.264204	9.84528	11.8657
ln_housing_density	307205	4.01933	1.66	-1.97198	10.4197
Y	307205	0.0190361	0.136652	0	1

The dependent variable (Y) in this case represents whether or not the CWS incurred a total coliform violation in a given year. Note that none of the independent variables change from the previous table except ‘prev_yr_coliform_viol’ (which replaces ‘prev_yr_all_viol’).

E. Methods

The two methods used in this paper are probit regression and hot spot analysis. The first method that I will describe is probit regression. From page 2079 of Maura Allaire et al.'s 2018 paper[14], we have the following formula for the probit model:

$$\Pr(y_{it} = 1 \mid X) = \Phi(\beta_0 + \beta_x x_i + \gamma_{jt} C_{jt} + \alpha_t T_t + \phi_k S_k)$$

- y_{it} is the binary indicator that CWS i incurred a violation in year t
- x_i represents time-invariant CWS characteristics (prev_yr_all_viol, prev_yr_coliform_viol, is_private, utility_small, utility_medium, utility_large, purchased, surface_water, ground_water, hhi, violation_year)
- C_{jt} represent county-level statistics for county j in year t (nonwhite_percent, ln_median_income, ln_housing_density)
- There are dummy variables for year (T_t) and state (S_k) to control for changes in laws and/or compliance over time, as well as state-level effects.
- A time trend regression analysis was also carried out which used the formula above but added two extra terms: a linear time trend, and an interaction term of the linear time trend and state dummy variable.

[14] https://www.pnas.org/content/pnas/115/9/2078.full.pdf

The second method of analysis used in this paper is hot spot analysis. Hot spot analysis is a method of local spatial autocorrelation which is used to determine whether clusters of counties with violations are statistically significant. Only those counties will be found to be significant which themselves have a large number of violations per CWS and are also surrounded by counties with similarly high violations. From page 3 of the supporting information[15] section of the 2018 paper written by Maura Allaire et al., the formula for the local Getis-Ord statistic is:

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{i,j}(d)\, x_j - W_i \bar{x}}{s\left(\sqrt{\frac{n * \sum_{j=1}^{n} w_{i,j}^2 - W_i^2}{n-1}}\right)}$$

- $G_i^*(d)$ is the Getis-ord statistic of a county. The greater this value is, the more significant the clustering of hot spots
- x_j is the number of health-based or total coliform violations per CWS in county j
- $w_{i,j}$ is the spatial weight between two counties, i and j. The spatial weight between counties is calculated for all pairs of counties in the dataset. It takes the value of 1 if the Euclidean distance between the centroids of two counties is less than a threshold distance 'd', and is otherwise 0. The threshold distance was set to 146.2 km, which was also its value in the 2018 paper[16]. The logic in the original

[15] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf
[16] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf pp. 3

paper for selecting this value for 'd' was that it ensured that every county had at least one neighbor.

- W_i is the sum of the spatial weights $w_{i,j}$
- $\bar{x}$ is equal to $\frac{1}{n}\sum_{j=1}^{n} x_j$ i.e. the mean of all x_j's
- s is equal to $\sqrt{\frac{1}{n} * (\sum_{j=1}^{n} x_j^2) - (\bar{x})^2}$

F. Results

The methods of probit regression and hot spot analysis were applied to the water quality violations dataset for the years 2009-2019. This section includes tables for probit regression coefficients, marginal effects, time trends, choropleth maps and hot spots for all health-based violations (referred to as 'total violations') and total coliform violations. All analysis was performed in Python, and the algorithm for conducting hot spot analysis was implemented from scratch.

Probit regression

a. Total violations

	Y: Total violations	
	No interactions	Interactions
	(1)	(2)
prev_yr_all_viol	1.507***	1.502***
	(0.010)	(0.010)
is_private	-0.114***	-0.073***
	(0.012)	(0.013)
utility_medium	0.060***	0.062***
	(0.011)	(0.012)
utility_large	0.063***	0.145***
	(0.013)	(0.015)
purchased	0.069***	0.067***
	(0.011)	(0.011)
surface_water	0.297***	0.385***
	(0.012)	(0.018)
is_private:surface_water		-0.164***
		(0.035)
is_private:utility_large		-0.133***
		(0.038)
utility_medium:surface_water		-0.035
		(0.026)
utility_large:surface_water		-0.214***

		(0.026)
ln_median_income	-0.129***	-0.135***
	(0.025)	(0.025)
nonwhite_percent	-0.082**	-0.083**
	(0.038)	(0.038)
ln_housing_density	-0.065***	-0.066***
	(0.004)	(0.004)
hhi	0.113***	0.120***
	(0.024)	(0.024)
Intercept	-0.529**	-0.471*
	(0.263)	(0.263)
Observations	307,205	307,205
Pseudo R^2	0.2307	0.2315
Residual Std. Error	1.000 (df=307136)	1.000 (df=307132)
F Statistic	(df=68; 307136)	(df=72; 307132)
Note:		*p<0.1; **p<0.05; ***p<0.01

We find that all independent variables are statistically significant except 'utility_medium:surface_water'. In the original paper, 'utility_medium:surface_water' was statistically significant, while 'ln_median_income' was the only statistically insignificant variable[17]. The two statistically significant variables whose coefficients have opposite signs from the original paper are 'purchased' (negative in the original paper) and 'nonwhite_percent'

[17] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf Table S3, pp. 11

(positive in the original paper)[18]. The independent variables which are statistically significant and have a negative correlation with total violations are: ‘is_private’, ‘ln_median_income’, ‘nonwhite_percent’, ‘ln_housing_density’, ‘is_private:surface_water’, ‘is_private:utility_large’, ‘utility_large:surface_water’. We also note that the sign of none of the independent variables changed after the interaction terms were added.

Marginal effects:

Variable	*No interactions (1) - Marginal effect*	*No interactions (1) - Std error*	*Interactions (2) - Marginal effect*	*Interactions (2) - Std error*
prev_yr_all_viol	0.1210 (***)	0.001	0.1206 (***)	0.001
is_private	-0.0092 (***)	0.001	-0.0059 (***)	0.001
utility_medium	0.0049 (***)	0.001	0.0050 (***)	0.001
utility_large	0.0051 (***)	0.001	0.0117 (***)	0.001
purchased	0.0056 (***)	0.001	0.0054 (***)	0.001
surface_water	0.0239 (***)	0.001	0.0309 (***)	0.001
ln_median_income	-0.0103 (***)	0.002	-0.0109 (***)	0.002
nonwhite_percent	-0.0066 (**)	0.003	-0.0067 (**)	0.003
ln_housing_density	-0.0052 (***)	0.000	-0.0053 (***)	0.000
hhi	0.0091 (***)	0.002	0.0096 (***)	0.002

[18] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf Table S3, pp. 11

The signs of the marginal effects of 'ln_median_income', 'nonwhite_percent' and 'purchased' are the opposite from those in the published paper[19] ('ln_median_income' was statistically insignificant in the 2018 paper). All other variables retain the same signs and orders of magnitude of their average marginal effects with the original paper. The variable with the highest average marginal effect is 'prev_yr_all_viol' i.e. whether or not the CWS had a health-based violation the year before. For 2009-19, the probability of a health-based violation for a utility purchasing water is 0.56% higher than a utility with a groundwater source, ceteris paribus (according to model 1). The probability of a health-based violation for a privately owned utility is 0.92% lower than a government-owned utility, ceteris paribus (also according to model 1). Both models agree that counties which are low-income and rural are associated with a higher likelihood of health-based violations. Surprisingly, counties with racial diversity (i.e. a higher non-white percentage) seem not to suffer more from health-based violations than those which are predominantly white.

[19] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf Table S4, pp. 11

b. Total violations, time trend regression

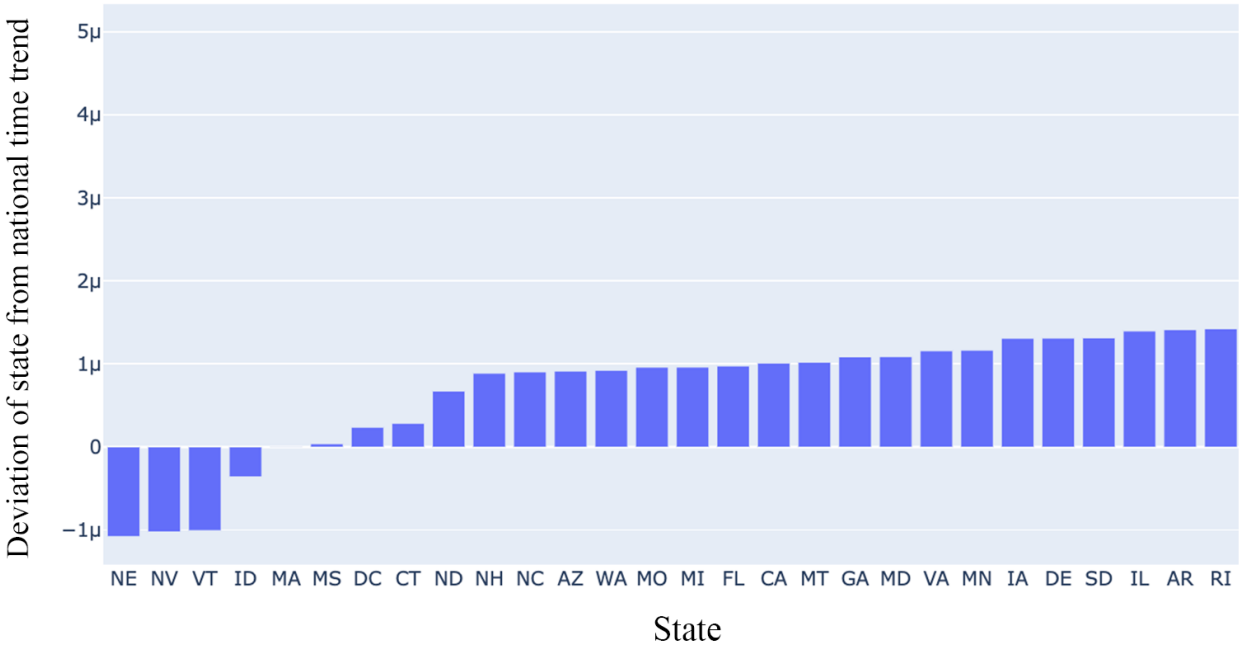
	Y: Total violations	
	State-specific time trend: not included	State-specific time trend: included
	(1)	(2)
prev_yr_all_viol	1.411***	1.401***
	(0.011)	(0.011)
is_private	-0.098***	-0.099***
	(0.014)	(0.014)
utility_medium	0.076***	0.077***
	(0.012)	(0.012)
utility_large	0.177***	0.178***
	(0.015)	(0.015)
purchased	0.033***	0.033***
	(0.011)	(0.011)
surface_water	0.322***	0.327***
	(0.018)	(0.018)
is_private:surface_water	-0.104***	-0.103***
	(0.036)	(0.036)
is_private:utility_large	-0.153***	-0.156***
	(0.038)	(0.038)
utility_medium:surface_water	-0.043*	-0.044*
	(0.026)	(0.026)

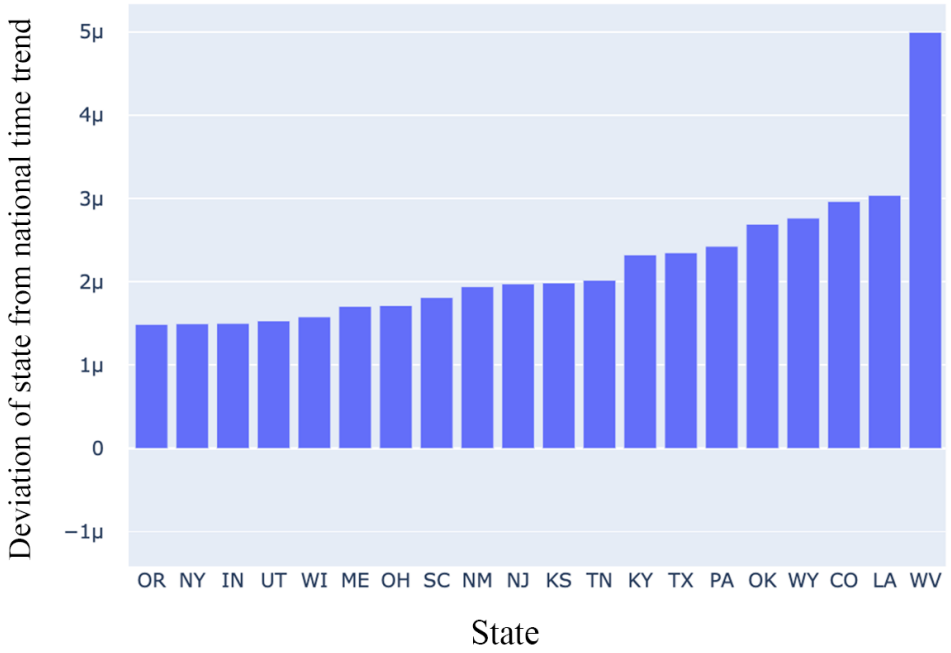
utility_large:surface_water	-0.193***	-0.195***
	(0.026)	(0.026)
ln_median_income	-0.120***	-0.122***
	(0.026)	(0.026)
nonwhite_percent	-0.188***	-0.177***
	(0.040)	(0.040)
ln_housing_density	-0.090***	-0.091***
	(0.004)	(0.004)
hhi	0.381***	0.384***
	(0.025)	(0.025)
violation_year	0.000	0.000***
	(33.570)	(nan)
Intercept	-0.183	-0.184***
	(67441.478)	(nan)
Observations	307,205	307,205
Pseudo R^2	0.2489	0.2520
Residual Std. Error	1.000 (df=307131)	1.000 (df=307083)
F Statistic	(df=73; 307131)	(df=121; 307083)
Note:		*p<0.1; **p<0.05; ***p<0.01

All independent variables are found to be statistically significant except 'utility_medium:surface_water' and 'violation_year' (significant only in model 2). In contrast, the variables which were not statistically significant in the original paper were

'ln_median_income' and 'violation_year' [20]. The two statistically significant variables whose signs differ from the original paper are 'purchased' (negative in the original paper) and 'nonwhite_percent' (positive in the original paper) [21]. The variables which are found to be statistically significant and negatively associated with total violations are: 'is_private', 'ln_median_income', 'nonwhite_percent', 'ln_housing_density', 'is_private:surface_water', 'is_private:utility_large' and 'utility_large:surface_water'. From model 2, we can plot the deviation of states' time trends from the national time trend. The deviation is the estimated coefficient of the interaction term of state dummy variables and the linear time trend. Two images for this plot have been used instead of one, for readability purposes. It is found that Nebraska, Nevada, Vermont and Idaho are only states whose time trends deviate negatively from the national average. All other states have an increasing time trend for the study period. West Virginia (a clear outlier), Louisiana and Colorado show the greatest positive deviation for health-based violations.

[20] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf Table S5, pp. 12
[21] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf Table S5, pp. 12


Deviation of state from national time trend
5μ
4μ
3μ
2μ
1μ
0
−1μ
NE NV VT ID MA MS DC CT ND NH NC AZ WA MO MI FL CA MT GA MD VA MN IA DE SD IL AR RI
State


Deviation of state from national time trend
5μ
4μ
3μ
2μ
1μ
0
−1μ
OR NY IN UT WI ME OH SC NM NJ KS TN KY TX PA OK WY CO LA WV
State

c. Total coliform violations

	Y: Total coliform violations	
	No interactions	Interactions
	(1)	(2)
prev_yr_coliform_viol	0.899***	0.895***
	(0.020)	(0.020)
is_private	-0.062***	-0.028
	(0.017)	(0.019)
utility_medium	0.200***	0.178***
	(0.015)	(0.016)
utility_large	0.279***	0.327***
	(0.017)	(0.020)
purchased	-0.089***	-0.090***
	(0.015)	(0.015)
surface_water	-0.191***	-0.184***
	(0.020)	(0.036)
is_private:surface_water		-0.189***
		(0.070)
is_private:utility_large		-0.145***
		(0.050)
utility_medium:surface_water		0.155***
		(0.047)
utility_large:surface_water		-0.113**

		(0.047)
ln_median_income	-0.040	-0.045
	(0.036)	(0.036)
nonwhite_percent	-0.034	-0.031
	(0.053)	(0.053)
ln_housing_density	-0.053***	-0.053***
	(0.006)	(0.006)
hhi	0.091***	0.095***
	(0.034)	(0.034)
Intercept	-1.633***	-1.583***
	(0.381)	(0.382)
Observations	307,205	307,205
Pseudo R^2	0.1506	0.1516
Residual Std. Error	1.000 (df=307136)	1.000 (df=307132)
F Statistic	(df=68; 307136)	(df=72; 307132)
Note:		*p<0.1; **p<0.05; ***p<0.01

From the above table, we note the statistical significance of all independent variables except 'ln_median_income', 'nonwhite_percent' and 'is_private' (significant only in model 1). In the published study, the only variables which were not statistically significant were 'ln_median_income' and 'is_private:surface_water'[22]. The variables which are statistically significant and found to be negatively correlated with total coliform violations are: 'purchased',

[22] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf Table S3, pp. 11

‘ln_housing_density’, ‘surface_water’, ‘is_private:surface_water’, ‘is_private:utility_large’, ‘utility_medium:surface_water’, ‘utility_large:surface_water’ and ‘is_private’ (only in model 1). The coefficients of all independent variables share the same sign with the coefficients in the published paper, except ‘nonwhite_percent’.

Marginal effects:

Variable	*No interactions (1) - Marginal effect*	*No interactions (1) - Std error*	*Interactions (2) - Marginal effect*	*Interactions (2) - Std error*
prev_yr_coliform_viol	0.0227 (***)	0.001	0.0224 (***)	0.001
is_private	-0.0016 (***)	0.000	-0.0007	0.000
utility_medium	0.0050 (***)	0.000	0.0045 (***)	0.000
utility_large	0.0070 (***)	0.000	0.0082 (***)	0.000
purchased	-0.0022 (***)	0.000	-0.0023 (***)	0.000
surface_water	-0.0048 (***)	0.001	-0.0046 (***)	0.001
ln_median_income	-0.0010	0.001	-0.0011	0.001
nonwhite_percent	-0.0008	0.001	-0.0008	0.001
ln_housing_density	-0.0013 (***)	0.000	-0.0013 (***)	0.000
hhi	0.0023 (***)	0.001	0.0024 (***)	0.001

We note the statistical significance of the marginal effect of all independent variables except ‘ln_median_income’, ‘nonwhite_percent’ and ‘is_private’ (statistically significant only in model 1). In the published study, only ‘ln_median_income’ did not have a statistically significant

average marginal effect[23]. The variable 'prev_yr_coliform_viol' has the largest marginal effect. The signs of the marginal effects of all independent variables are the same as in the original paper, except 'nonwhite_percent'. From model 1, the probability of a total coliform violation for a utility purchasing water is 0.22% lower than a utility with a groundwater source, ceteris paribus. Also from the same model, the probability of a total coliform violation for a privately owned utility is 0.16% lower than a government-owned utility, ceteris paribus. Populations with a high population density are found to be associated with a lower likelihood of total coliform violations.

[23] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf Table S4, pp. 11

d. Total coliform violations, time trend regression

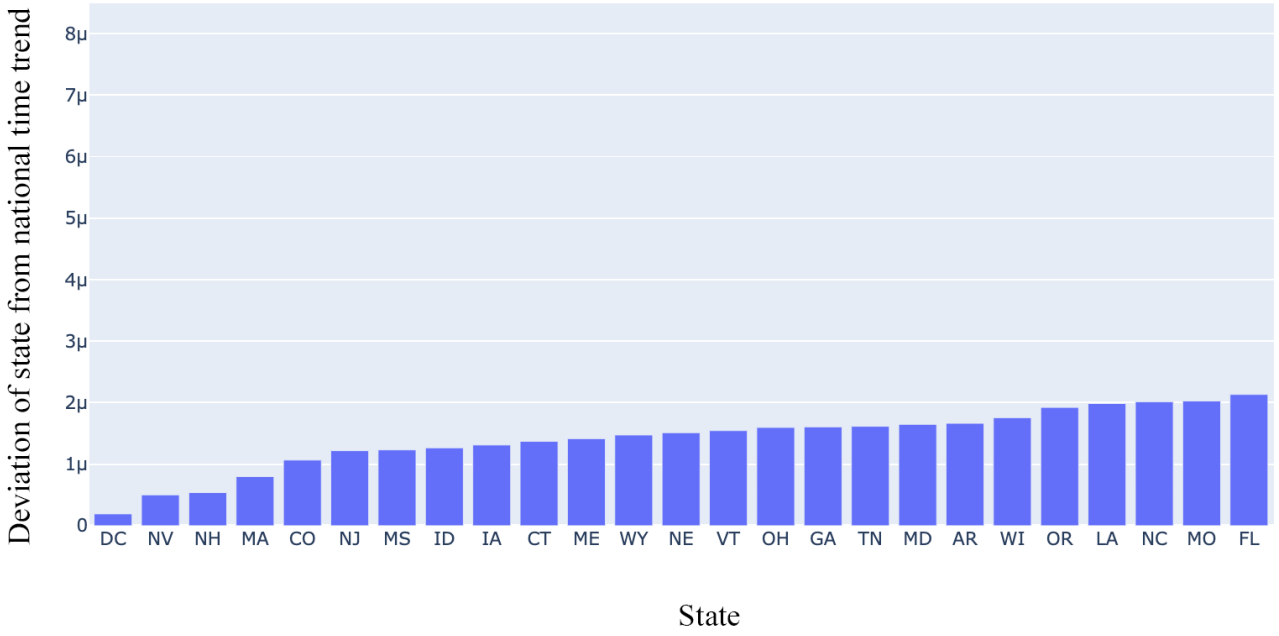
	Y: Total coliform violations	
	State-specific time trend: not included	State-specific time trend: included
	(1)	(2)
prev_yr_coliform_viol	0.879***	0.872***
	(0.020)	(0.020)
is_private	-0.040**	-0.040**
	(0.019)	(0.019)
utility_medium	0.184***	0.184***
	(0.016)	(0.016)
utility_large	0.339***	0.341***
	(0.020)	(0.020)
purchased	-0.107***	-0.108***
	(0.015)	(0.015)
surface_water	-0.218***	-0.222***
	(0.036)	(0.036)
is_private:surface_water	-0.159**	-0.158**
	(0.070)	(0.071)
is_private:utility_large	-0.151***	-0.154***
	(0.050)	(0.050)
utility_medium:surface_water	0.151***	0.156***
	(0.047)	(0.047)
utility_large:surface_water	-0.098**	-0.099**

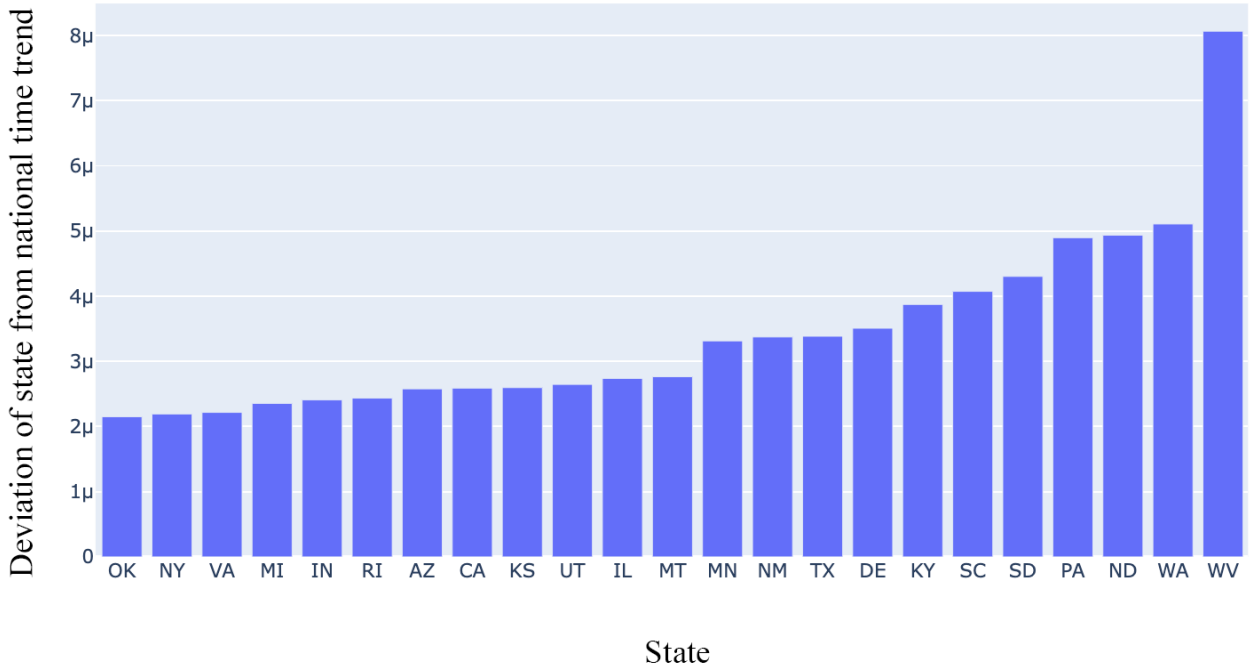
	(0.047)	(0.047)
ln_median_income	-0.035	-0.043
	(0.037)	(0.037)
nonwhite_percent	-0.079	-0.069
	(0.054)	(0.054)
ln_housing_density	-0.065***	-0.064***
	(0.006)	(0.006)
hhi	0.218***	0.216***
	(0.035)	(0.035)
violation_year	-0.001	-0.001***
	(62.404)	(nan)
Intercept	-0.069	-0.068***
	(125369.852)	(nan)
Observations	307,205	307,205
Pseudo R^2	0.1556	0.1593
Residual Std. Error	1.000 (df=307131)	1.000 (df=307083)
F Statistic	(df=73; 307131)	(df=121; 307083)
Note:		*p<0.1; **p<0.05; ***p<0.01

All independent variables are found to be statistically significant except 'ln_median_income', 'nonwhite_percent' and 'violation_year' (only significant in model 2). In comparison, the variables in the original paper which were not statistically significant were

'ln_median_income', 'is_private:surface_water', and 'violation_year' [24]. The coefficients of all independent variables share the same sign with the coefficients in the published paper, except 'nonwhite_percent'. The statistically significant variables that are found to correlate negatively with total coliform violations are: 'is_private', 'purchased', 'surface_water', 'is_private:surface_water', 'is_private:utility_large', 'utility_large:surface_water', 'ln_housing_density' and 'violation_year'. We can once again plot the deviation of states from the national time trend using the second model. All states are found to show an increasing time trend with respect to total coliform violations. West Virginia is once again noted to be an outlier in terms of its time trend possessing the greatest positive deviation from the national average. It is followed by Washington, North Dakota and Pennsylvania. The states with the lowest deviation from the national average are Washington DC, Nevada, New Hampshire and Massachusetts.

[24] https://www.pnas.org/content/pnas/suppl/2018/02/07/1719805115.DCSupplemental/pnas.201719805SI.pdf Table S4, pp. 11

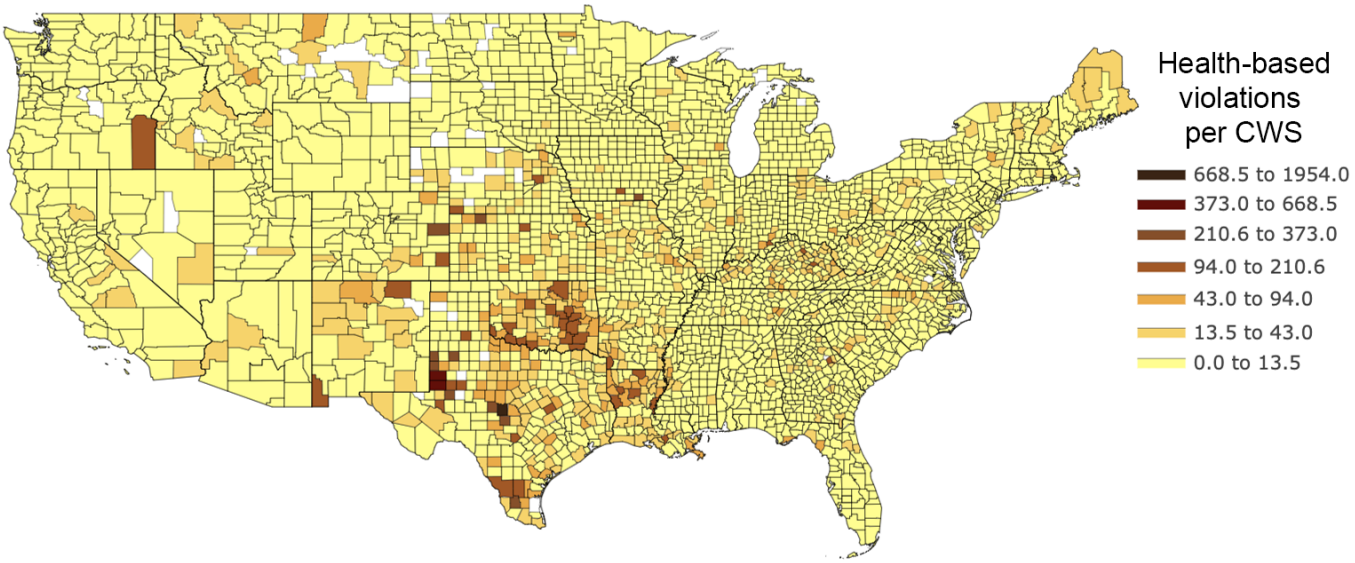

Deviation of state from national time trend
8μ
7μ
6μ
5μ
4μ
3μ
2μ
1μ
0
DC NV NH MA CO NJ MS ID IA CT ME WY NE VT OH GA TN MD AR WI OR LA NC MO FL
State


Deviation of state from national time trend
8μ
7μ
6μ
5μ
4μ
3μ
2μ
1μ
0
OK NY VA MI IN RI AZ CA KS UT IL MT MN NM TX DE KY SC SD PA ND WA WV
State

Hotspot analysis

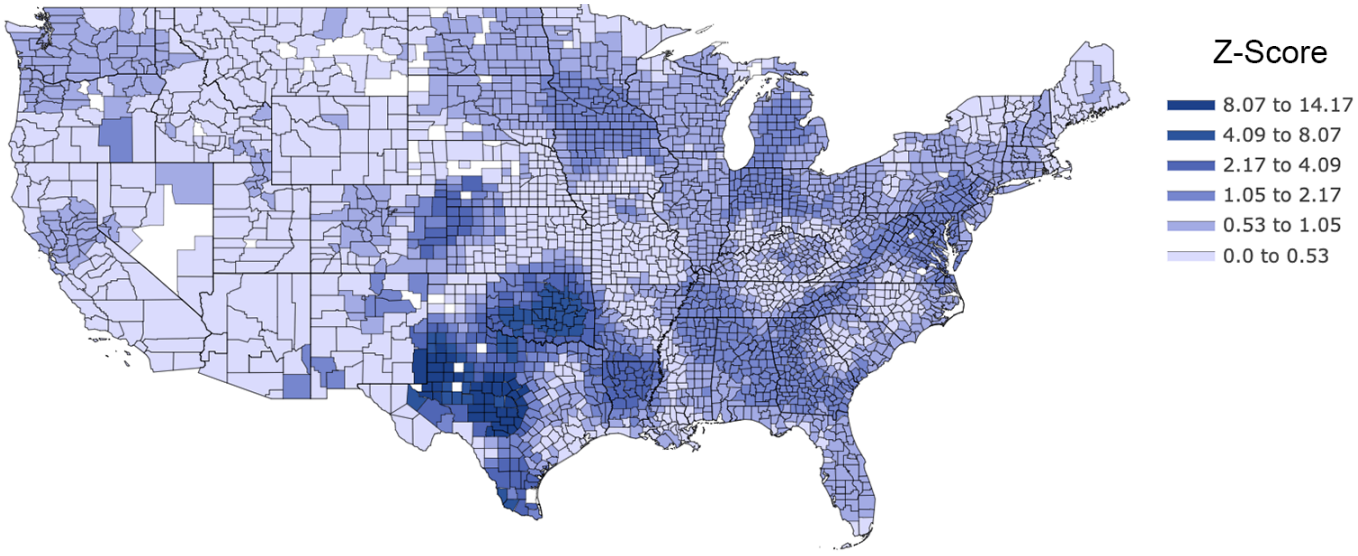
a. Total violations

We first create a choropleth map showing the number of health-based violations per CWS from 2009-2019:



The intervals in the legend were selected based on the Jenks natural breaks classification method[25]. The states of Texas, Oklahoma, Louisiana, New Mexico, Colorado, Kansas, Nebraska, Iowa and Oregon are home to counties which have a high number of health-based violations per CWS. Next, in order to ascertain whether or not clusters of counties are statistically significant, we use hot spot analysis to calculate the Z-scores of the clusters:

[25] The Jenkspy library is used to implement the 'natural breaks' algorithm (https://github.com/mthh/jenkspy)

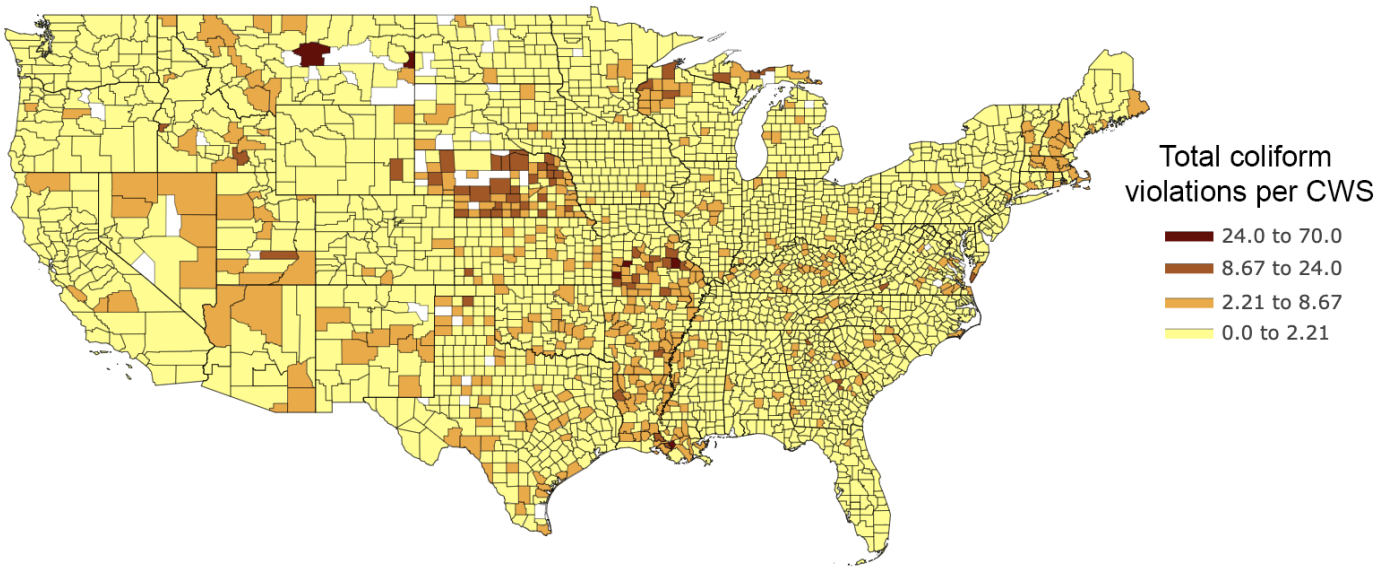


We notice intense spatial clustering of health-based violations in the states of Texas, Oklahoma, Louisiana, Kansas, Nebraska and Colorado. We also notice light hot spots in Mississippi, Alabama, Georgia, Tennessee, North Carolina, Minnesota, Iowa, Michigan, Indiana, Ohio, Virginia, Maryland, West Virginia, Pennsylvania, Arizona and New Mexico. These findings mirror those of the original study, which found hot spots of health-based violations in the Southwest US, particularly in Texas and Oklahoma[26].

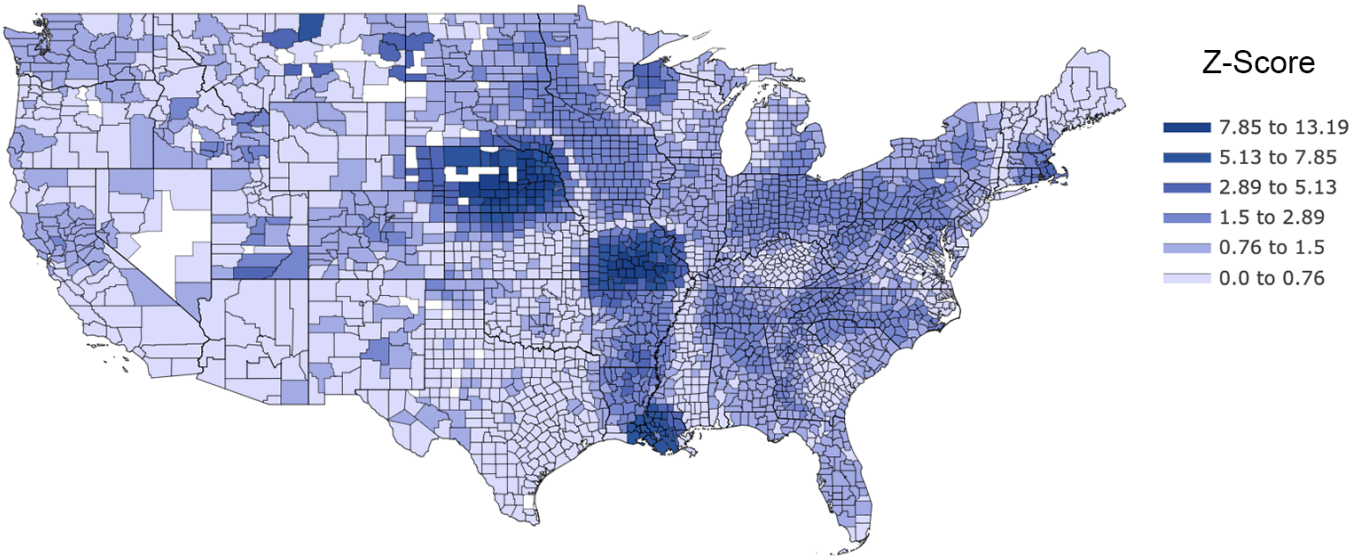
[26] https://www.pnas.org/content/pnas/115/9/2078.full.pdf pp. 2082

b. Total coliform violations

As we did in the case of total violations, we first create a choropleth map showing the number of total coliform violations per CWS from 2009-2019:



We observe that Montana, Nebraska, Missouri, Arkansas, Louisiana, and Wisconsin are home to counties which have a high number of total coliform violations per CWS. Next, in order to ascertain whether or not clusters of counties are statistically significant, we use hot spot analysis to calculate the Z-scores of the clusters:



We notice intense spatial clustering of total coliform violations in the states of Nebraska, Missouri, and Louisiana. We also notice light hot spots in Montana, Utah, North Dakota, Wisconsin, Massachusetts, Minnesota, Idaho, Iowa, South Dakota, Indiana, Ohio, Tennessee, Alabama, North Carolina, South Carolina, Georgia, Pennsylvania, West Virginia, Florida, New York, Connecticut and Rhode Island.

G. Limitations

There exist a number of limitations to the analyses conducted in this paper. Firstly, it is estimated that between 26 to 38 percent of health-based violations are not reported to the SDWIS database ("Government Accountability Office"). For this reason, total coliform violations had to be analysed, as these violations are more accurately reported than other violations ("US Environmental Protection Agency"). Another major limitation faced during the analysis was the lack of access to data on CWS's funding reports. As noted earlier, one of the biggest contributing factors to violation occurrence was whether or not a CWS had incurred a violation the year prior. It is extremely likely that utilities which are under monetary stress will find it difficult to upgrade/repair systems that lead to repeat violations. However, it is not possible to gather data about the monetary strength of CWS's as they are not all publicly traded companies. This leads to the use of variables such as median income to 'proxy' for the financial health of CWS's in that county. Moreover, the 5-year estimates of the American Community Survey (ACS) were only available from 2009 to 2019. Hence, the study period could not have been extended without using additional data sources (such as US Census data). There were 715 CWS's (out of 28,644) for which the county could not be determined. Of these, 25 CWS's had incurred at least one health-based violation in 2009-2019, and 12 had incurred at least one total coliform violation. A similar problem was faced during hot spot analysis, where the latitude and longitude of the centroid of at least 7 counties could not be determined. Lastly, 14 rows had to be excluded from the dataset before conducting probit regression because the American Community Survey data lacked housing density or median income data for these counties.

H. Conclusion

In agreement with the findings of the original paper, the most significant indicator of whether or not a CWS would incur a violation in a given year was whether it had incurred a violation in the year prior. Surface water sources were correlated with an increase in total violations in a county, but a decrease in total coliform violations. Private, urban CWS's were associated with fewer total violations and total coliform violations. Medium-sized CWS's relying on surface water sources were noted to correlate with an increase in total coliform violations. The higher the Herfindahl–Hirschman index (hhi) of a county was, the greater the chance that the CWS's in it incurred a health-based or total coliform violation. This implies that counties which have fewer CWS's and greater size differences in the populations served by its CWS's are worse off. Contrary to the findings of the original paper, counties that enjoyed a higher proportion of non-white population were not always found to be correlated with higher water quality violations. Instead, some models noted with statistical significance that an increase in the nonwhite-percent of a county led to a reduction in its health-based violations. Also contrary to the findings of the original paper, purchased water was found to be correlated with an increase in health-based violations. However, purchased water was also found to be significant in reducing total coliform violations.

Significant hot spots of health-based violations were detected in parts of Texas, Oklahoma, Louisiana, Kansas, Nebraska and Colorado. The same was the case for hot spots of total coliform violations in Nebraska, Missouri, and Louisiana. Hot spots of health-based violations did not move from their locations in 2004-15 [27]. This shows that underperforming

[27] See appendix

CWS's in these states are still in need of increased technical guidance and financial assistance to combat repeat violations. Low-income, rural communities whose CWS's did not have private water sources were at a high risk of facing health-based and total coliform violations. Underreporting of violations also remains a persistent problem in the SDWIS database at the time of writing ("Josset, Laureline, et al."). Hence, the findings of this paper largely echo those of the 2018 paper by Maura Allaire et al. by which it was inspired. One suggestion that the author of this paper has is to expand the use of sensors such as KETOS Shield[28] to municipalities to immediately identify dangerous contaminant levels and monitor water quality in real-time.

[28] https://ketos.co/shield/

I. References

Allaire, Maura, et al. “National Trends in Drinking Water Quality Violations.” *PNAS*, National Academy of Sciences, 27 Feb. 2018, www.pnas.org/content/115/9/2078.

Allaire, Maura, et al. “Detecting Community Response to Water Quality Violations Using Bottled Water Sales.” *PNAS*, National Academy of Sciences, 15 Oct. 2019, www.pnas.org/content/116/42/20917.

Bilal, Muhammad, et al. “Emerging Contaminants of High Concern and Their Enzyme-Assisted Biodegradation – A Review.” *Environment International*, Pergamon, 17 Jan. 2019, www.sciencedirect.com/science/article/pii/S0160412018323523.

Campos, Marcus André Siqueira, et al. “Impact of the COVID-19 Pandemic on Water Consumption Behaviour.” *Water Supply*, IWA Publishing, 1 Dec. 2021, iwaponline.com/ws/article/21/8/4058/82385/Impact-of-the-COVID-19-pandemic-on-water.

Getis, Arthur, and J. K. Ord. “The Analysis of Spatial Association by Use of Distance Statistics.” *Geographical Analysis*, vol. 24, no. 3, 1992, pp. 189–206., https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Government Accountability Office. “Drinking water: Compliance problems undermine EPA program as new challenges emerge”. *GAO/RCED-90-127*, 1990. https://www.gao.gov/products/rced-90-127. Accessed January 29, 2022.

Josset, Laureline, et al. “The U.S. Water Data Gap-A Survey of State-Level Water Data Platforms to Inform the Development of a National Water Portal.” *AGU Journals*, 25 Apr. 2019, agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018EF001063.

Michielssen, Senne, et al. “Trends in Microbiological Drinking Water Quality Violations across the United States.” *Environmental Science: Water Research & Technology*, The Royal Society of Chemistry, 18 Sept. 2020, pubs.rsc.org/en/content/articlelanding/2020/ew/d0ew00710b. Accessed January 29, 2022.

US Environmental Protection Agency. “Data reliability analysis of the EPA safe drinking water information system/federal version (SDWIS/FED)”. *US Environmental Protection Agency*, 2000, EPA 816-R-00-020. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=40001BK0.TXT. Accessed January 29, 2022.

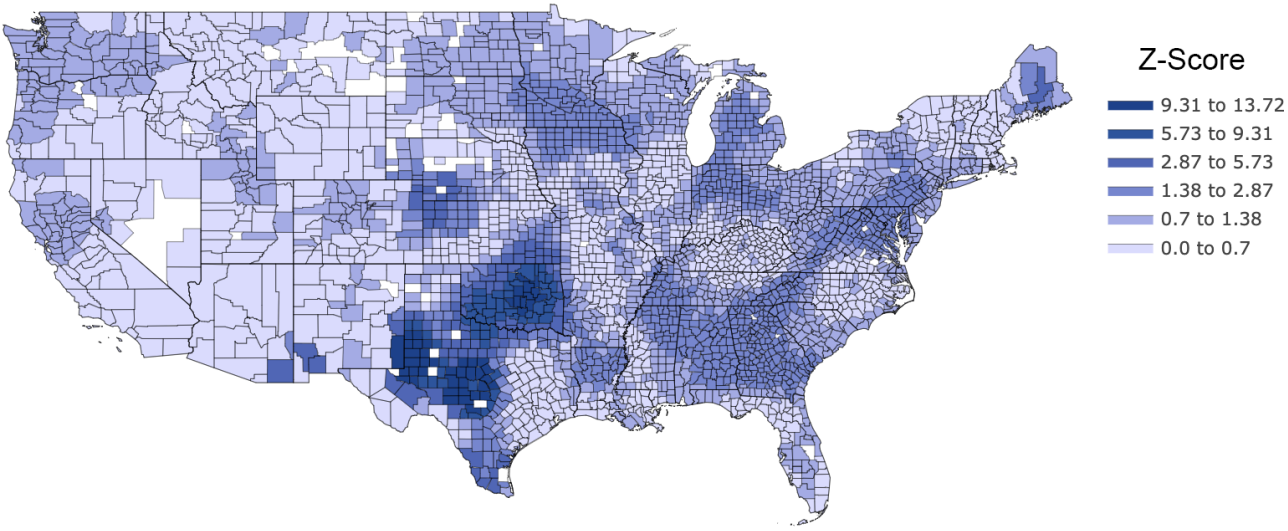
Wallsten, Scott, and Katrina Kosec. “Public or Private Drinking Water? The Effects of Ownership and Benchmark Competition on U.S. Water System Regulatory Compliance and Household Water Expenditures.” *SSRN*, 20 Apr. 2005, papers.ssrn.com/sol3/papers.cfm?abstract_id=707131.

"'We've Always Known Ours Was Contaminated': The Trouble with America's Water." *The Guardian*, 15 Sept. 2020, www.theguardian.com/us-news/2020/sep/15/america-water-crisis-contamination-pollution-infrastructure.

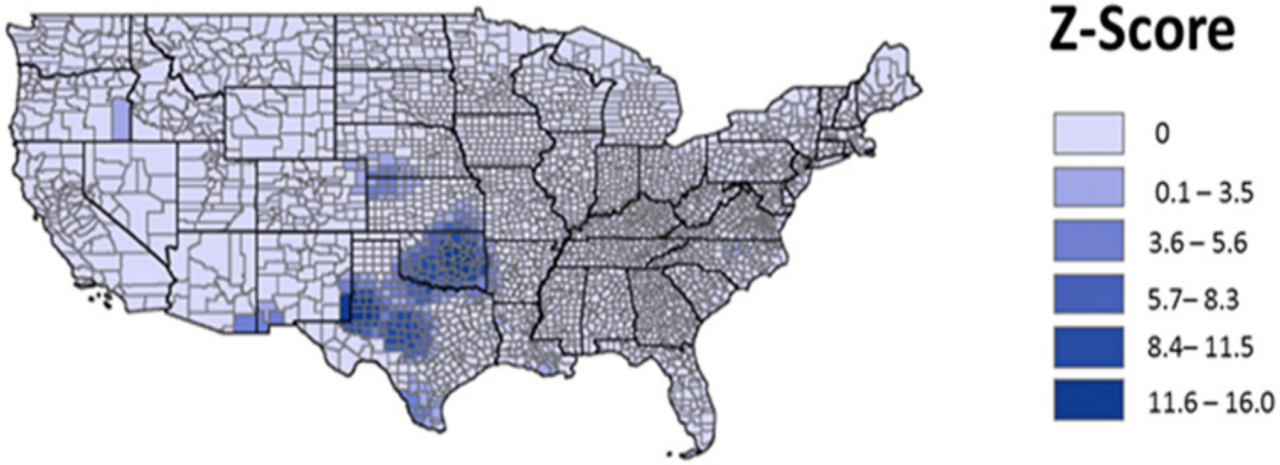
Wu, Jishan, et al. "A Critical Review of Point-of-Use Drinking Water Treatment in the United States." *Nature News*, 22 July 2021, www.nature.com/articles/s41545-021-00128-z.

J. Appendix

Hot spots of health-based violations from 2004-2015 made using the code with this paper:



This is equivalent to Fig. S2-C in the supporting information for Maura Allaire et al.'s 2018 paper[29]:



[29] https://www.pnas.org/content/suppl/2018/02/07/1719805115.DCSupplemental, pp. 8