PLOS WATER

RESEARCH ARTICLE

The Municipal Drinking Water Database

Sara Hughes[1]*, Christine J. Kirchhoff[2¤], Katelynn Conedera[3], Mirit Friedman[3]

1 School for Environment and Sustainability, University of Michigan, Ann Arbor, Michigan, United States of America, **2** Civil and Environmental Engineering, University of Connecticut, Mansfield, Connecticut, United States of America, **3** Department of Urban and Regional Planning, University of Michigan, Ann Arbor, Michigan, United States of America

¤ Current address: School of Engineering Design and Innovation, Law, Policy and Engineering Program, and Department of Civil and Environmental Engineering, Pennsylvania State University, State College, Pennsylvania, United States of America
* hughessm@umich.edu

Citation: Hughes S, Kirchhoff CJ, Conedera K, Friedman M (2023) The Municipal Drinking Water Database. PLOS Water 2(4): e0000081. https://doi.org/10.1371/journal.pwat.0000081

Editor: Manuel Herrera, University of Cambridge, UNITED KINGDOM

Received: November 7, 2022

Accepted: March 9, 2023

Published: April 10, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pwat.0000081

Data Availability Statement: Data are available at https://doi.org/10.7910/DVN/DFB6NG.

Funding: This work was supported by NSF Award # 2048505 (S.H., C.K.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Drinking water services in the U.S. are critical for public health and economic development but face technical, political, and administrative challenges. Understanding the root cause of these challenges and how to overcome them is hindered by the lack of integrative, comprehensive data about drinking water systems and the communities they serve. The Municipal Drinking Water Database (MDWD) combines financial, institutional, demographic, and environmental conditions of U.S. municipalities and their community water systems (CWS). Municipally owned and operated CWS are ubiquitous and play a critical role in ensuring safe, affordable drinking water services for most Americans; they also offer important opportunities for understanding municipal government behavior and decision making. The MDWD is publicly available and will enable researchers and practitioners interested in viewing or tracking drinking water spending, the financial condition of municipal governments, or myriad demographic, political, institutional, and physical characteristics of municipal drinking water systems to access the data quickly and easily. This paper describes the database and its creation, details examples of how the data can be used and discusses illustrative analyses of trends and insights that can be gleaned from the database. Building and sharing more integrated datasets provides new opportunities for asking novel questions about the drivers and consequences of local decision making about drinking water.

Introduction

Drinking water services and infrastructure in the U.S. are unequal [1–3], and many communities are struggling with high water bills, water quality concerns, and aging infrastructure [4]. Moreover, drinking water systems designed for the past climate are increasingly vulnerable to disruption in today's climate, and often unprepared for future climate and other changes [5–7]. Understanding and addressing these challenges is hindered by a lack of basic information about drinking water systems and the communities they serve. Water scholars, practitioners and policy makers currently lack the ability to understand holistically drinking water decision making, including how socio-economic, political, institutional, and environmental contexts influence

Competing interests: The authors have declared that no competing interests exist.

decision making and performance outcomes (e.g., violations of the Safe Drinking Water Act). This paper describes the Municipal Drinking Water Database (MDWD), a new resource that helps to fill this critical data gap by enabling researchers and practitioners to develop and test hypotheses about the relationships between environmental conditions, community characteristics, governance structures, and drinking water decision making and performance. Building and sharing more integrated datasets provides new opportunities for asking novel questions about the drivers and consequences of local decision making about drinking water.

General-purpose local governments in the U.S. often own and operate the community water systems (CWS) responsible for providing residents with safe, affordable drinking water [8, 9]. The United State Environmental Protection Agency (USEPA) defines a CWS as a public water system "that serves at least 15 service connections used by year-round residents or regularly services at least 25 year-round residents (e.g., homes, apartments and condominiums that are occupied year-round as primary residences)" [10]. Throughout the paper we use the term CWS to refer to the municipal drinking water utilities we are interested in identifying. However, when other datasets use other terms, such as Public Water System or Water Utility we also use these terms in our descriptions. For the purpose of this paper, all of these terms are interchangeable. Municipally owned CWS serve the majority of the U.S. population [11]. Given their importance and ties to local governments, municipal CWS also offer unique and valuable opportunities for understanding the drivers and consequences of local decisions about drinking water [8]. Unfortunately, data challenges stymie researchers' and practitioners' ability to develop an integrated understanding of municipal governments and their CWS.

Data that provide comprehensive understanding of local governments and their CWS are difficult to find in readily available, comparable, and complete formats. Information about local governments, communities and households, and CWS are collected and aggregated in a range of formats, and at a range of spatial and temporal scales, without common identifiers. For example, the U.S. Census of Governments collects information about local government revenues, spending, and employees reported voluntarily by local governments every five years. Demographic data from the U.S. Census, which provide important socio-economic information about households and communities, are reported at a range of scales that vary in format depending on the state and the year. Information about CWS, both technical (e.g., water source) and regulatory (e.g., Safe Drinking Water Act (SDWA) Violations), is aggregated by the USEPA and stored in the Safe Drinking Water Information System (SDWIS). Data provided in SDWIS are reported by CWS to U.S. states and provide no simple way to connect a given CWS to the community it serves or to the local government which owns and operates it. These data limitations shape and constrain research on, and understanding of, drinking water services and infrastructure governance in the U.S. often leaving scholars to ask and answer questions using a single data source [1, 12–15], or case studies of a small number of cities [16, 17].

The MDWD helps to fill this critical gap by first identifying municipal CWS (CWS owned and operated by general-purpose local governments) and then providing information about the demographic, environmental, fiscal, institutional, political, and physical conditions of these systems over time. The database enables viewing or tracking quickly and easily drinking water spending by municipalities, the financial condition of city governments operating drinking water utilities, or the relationships among the myriad demographic, political, institutional, and physical characteristics of municipal CWS. The database also provides new opportunities to develop and test hypotheses about the relationships between community characteristics, environmental conditions, institutional form, and drinking water decision making and CWS performance (e.g., SDWA violations). The MDWD is freely available and can be downloaded at https://doi.org/10.7910/DVN/DFB6NG.

This article describes the MDWD and the steps we took to identify and integrate information about CWS, local governments and governance, environmental conditions and the communities served by municipal drinking water utilities. The unit of analysis in the MDWD is the municipality, and it includes 2,219 municipalities with 10,000 or more residents in the year 2000. We chose this population threshold to help ensure data availability and a base level of governance capacity and public service responsibilities (e.g., drinking water, parks, police). The database extends from 1997 to 2018 and includes municipalities in the 48 contiguous U.S. states. While we made every effort to include the full set of municipal drinking water systems in the U.S. associated with their respective municipality, the methods and scale of our approach mean there may be some that are not included.

To assemble the MDWD we followed a multi-step process of pre-filtering, matching, and match confirmation to generate a list of municipalities and their CWS. We then brought together information from a variety of sources including the U.S. Census, USEPA SDWIS, U. S. Census of Governments, International City/County Management Association, Voting and Election Science Team (VEST), Annual Survey of Public Employment and Payroll, and a Climate Moisture Index [18]. As a curated assemblage of existing data sources, the MDWD brings the limits and strengths associated with these sources and our methods for integration. In describing the database, we aim to be transparent about where data come from and our use of those data, and to guide users to further information about the original data sources. Despite these limitations, the MDWD provides new opportunities for exploring the social, political, environmental, and institutional underpinnings of drinking water provision and performance in the U.S. at a time when communities around the country are grappling with the challenges of aging infrastructure, rising costs, and climate change.

A primary innovation provided by the MDWD is that it links municipalities and their CWS. This linking process involved first using U.S. Census data to generate a list of municipalities that were consistently present and identifiable during the study period and had a population size of at least 10,000 residents in the 2000 census. We then used SDWIS data to generate a list of all CWS. We undertook a multi-step process of pre-filtering, matching, and match confirmation to identify the subset of municipalities that also own and manage a CWS. We took a conservative approach to matching and confirming, excluding, for example, public CWS governed by independently elected boards as they lack a direct connection to the municipal government. The output from this matching process is a final list of 2,219 municipalities and their relevant identifier codes: Federal Information Processing System (FIPS) codes for U. S. Census and elections data, GOVSid for U.S. Census of Governments and Annual Survey of Public Employment and Payroll data, and Public Water System Identification Number (PWSID) for SDWIS data. We used these codes to integrate financial, city government, demographic, and environmental data, including climatic information using Census-provided GIS shape files and a Climate Moisture Index [19] (Fig 1).

The Methods section provides more in-depth discussion and description of the creation of the MDWD. We then provide some examples of how the data can be used and discuss illustrative analyses of trends and insights that can be gleaned from the database. In the Discussion and Conclusion section we reflect on the uses and limits of the database.

Methods

The first step in creating the MDWD was identifying municipalities and then matching a CWS to that municipality in a way that was both comprehensive and accurate. This required navigating the different ways states characterize municipalities. We developed a reference guide [20] for identifying municipalities in Census data that documents these differences and how to

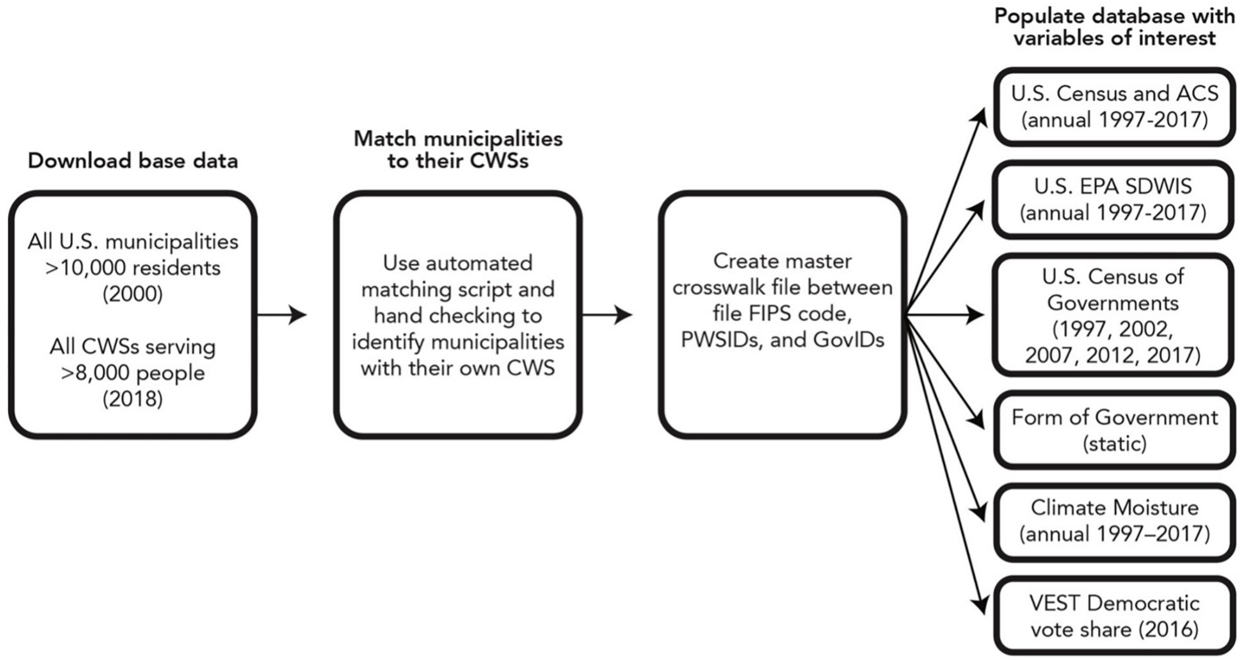

Fig 1. Conceptual diagram of the MDWD creation process.

https://doi.org/10.1371/journal.pwat.0000081.g001

navigate them. Because municipalities may change over time through consolidation and annexation, creating the dataset required identifying municipalities with CWS and ensuring we were tracking relevant metrics for the correct municipal units over space and time. We narrowed our list to municipalities that had a consistent municipal government and geographic size during our study period (1997–2018) by using the Census Bureau comparability files provided alongside every Decennial Census. These comparability files allowed us to identify municipalities over time regardless of FIPS code or name changes. We constructed a comparability key that linked 2000, 2010, and 2020 FIPS codes using the 2000–2010 and 2010–2020 comparability files. For 2000–2010 we included those municipalities with a continuing governmental status, meaning that the vast majority of the file contained only one-to-one relationships and records linking old and new FIPS codes for a particular municipality. The 2010–2020 comparability files were constructed by the Census Bureau based on geographic overlap between 2010 and 2020 shapefiles, meaning that any boundary changes over the prior ten years would result in a relationship being drawn between otherwise unrelated municipalities. Therefore, for data between 2010 and 2018 we selected only the relationship between old and new FIPS codes with the largest geographic overlap. Alaska, Hawaii, and Puerto Rico have unique and noncomparable municipal governance arrangements and were therefore excluded from the dataset. This step produced an initial list of 35,836 consistent municipalities in the 48 contiguous states. We then identified the subset of municipalities with 10,000 or more residents as of the 2000 Census, which narrowed the list to 3,930 municipalities.

After identifying municipalities and their Census designation (FIPS code), we sought to identify the subset of these municipalities that own and operate a CWS. We developed a multistep process of pre-filtering, matching, and match confirmation to identify this subset. Information about CWS location, population served, and ownership is collected by the USEPA and organized by Public Water System IDs (PWSIDs). We selected CWS serving at least 8,000 people to help ensure a true match between municipal jurisdictions (of 10,000 or more residents)

Table 1. Identifier variables included in the MDWD.

Variable Name	Variable Description	Data Source
FIPS_2000	FIPS code from 2000 census	U.S. Census
FIPS_2010	FIPS code from 2010 census	
FIPS	FIPS code from 2020 census	
state	State	
PWSID	Drinking Water System ID	U.S. EPA SDWIS
PWS.Name	Drinking Water System Name	
Name	City Name	
GOVSid	Unique ID for local governments from the Census of Governments	Government Finance Database; Census of Governments

https://doi.org/10.1371/journal.pwat.0000081.t001

and CWS service areas (e.g., they are of similar size). This produced a list of 5,489 CWS. Next, we developed an R script to match municipality names with CWS names, differentiating between confirmed matches, possible matches, and matches that otherwise did not fit our study criteria. Matching by name produced a list 1,988 municipalities with confirmed matches between the municipality name and the CWS name. Between September, 2021 and April, 2022 we checked by hand a set of 710 possible matches, relying on websites to verify the governance structure and municipal ownership of the utility. This produced an additional 233 matched municipalities and CWS, for a total of 2,219 municipal drinking water systems.

We used this matched list to create a crosswalk file that includes FIPS codes, PWSIDs, and GOVSid (used by U.S. Census of Governments in place of FIPS codes) for all the municipalities, including FIPS codes updates made by the Census between 2000 and 2020 (Table 1). We used these identifier codes to then collect and compile a range of variables.

We include a range of financial information available from the U.S. Census of Governments and compiled by the Government Finance Database (Table 2). These include measures of city government revenue, expenditures (by function and character), debt levels, and assets. We also include Consumer Price Index (CPI) values to allow for the calculation of inflation-adjusted

Table 2. Financial variables included in the MDWD.

Variable Name	Variable Description	Data Source; Manipulations
Total_Revenue	Total Revenue	Government Finance Database
Total_Expenditure	Total Expenditure	--
Total_Taxes	Total Taxes	Multiplied by $1000;
Water_Utility_Revenue	Water Utility Revenue	Adjusted for inflation
Water_Util_Total_Exp	Water Utility Total Expenditure	
Water_Util_Inter_Exp	Water Utility Intergovernmental Expenditure	
Water_Util_Cap_Outlay	Water Utility Capital Outlay	
Water_Util_Current_Exp	Water Utility Current Expenditure = Water utility total expenditure—water utility capital outlay	
Water_Util_Construct	Water Utility Construction	
Health_Total_Expend	Health Total Expenditure	
Parks___Rec_Total_Exp	Parks and Recreation Total Expenditure	
Police_Prot_Total_Exp	Police Protection Total Expenditure	
Total_Debt_Outstanding	Total Debt Outstanding	
Total_IG_Revenue	Total Intergovernmental Revenue	
Total_Rev_Own_Sources	Total Revenue Own Sources	
CPI	Consumer Price Index	US Bureau of Labor Statistics

https://doi.org/10.1371/journal.pwat.0000081.t002

Table 3. Variables from SDWIS included in the MDWD.

Variable Name	Variable Description	Data Source; Manipulations
GW_SW_CODE	Water Source; GW = Groundwater; SW = Surface Water	SDWIS
PRIMARY_SOURCE_CODE	Primary Water Source; GW = Groundwater; GWP = Ground water purchased; SW = Surface water SWP = Surface water purchased; GU = Groundwater under influence of surface water; GUP = Purchased ground water under influence of surface water source	
vio_count	Total number of health related violations in a year	SDWIS; Added all individual violations in a year

https://doi.org/10.1371/journal.pwat.0000081.t003

values. For our database, we multiplied all values by $1000 to represent true dollar values and used the corresponding annual Consumer Price Index (CPI) to adjust for inflation. Alternatively, users of the database could elect to use the Implicit Price Deflator to adjust for inflation, which measures changes in the prices of goods and services. The data included four discrete zero values for a municipality's total revenue and total expenditures and four discrete zero values for a municipality's number of full-time equivalent employees; we assumed these were reporting errors (in the self-reported data) and adjusted these zeros to be missing values. Additional financial variables, such as debt, water utility spending, and police spending, also included zero values. These were retained in the MDWD; by keeping these zeros, users of the database may make their own decisions about how to treat these zeros in their work.

We used the U.S. EPA's SDWIS system to download information about SDWA violations and water source for each municipal drinking water system (Table 3). Water source variables indicate whether the CWS relies on groundwater or surface water, and whether the water is purchased or self-produced. We collected data on all health-based violations reported between 1997 and 2018. These can take one of three forms: (1) exceedance of the maximum contaminant levels (MCLs) which specify the highest allowable contaminant concentrations in drinking water, (2) exceedance of the maximum residual disinfectant levels (MRDLs), and (3) failure to meet treatment technique requirements. All violations from a single year for each CWS were summed to create a value for total annual drinking water system violations.

We then used FIPS codes to populate the database with socio-economic and political information about the municipal residents (Table 4). These variables were drawn from the American Community Survey (ACS) and Decennial Census using Social Explorer and include information about race, education, and income levels. Demographic data are included for 2000 and 2009–2017, as ACS 5-year estimates are not available before 2009. Most variables are included in the dataset without manipulations. Median house age was calculated by subtracting the median year built from 2017 (the latest year this information is available in our study period) consistent with [21]. The percentage of individuals at or below the poverty line was calculated by adding individuals that fall at or below the poverty line across age groups and dividing it by the total population. We also include precinct-scale voting data aggregated to municipal boundaries generated and provided by VEST [22, 23]. This variable indicates the percent of a municipality's population that voted for the Democratic presidential candidate in 2016.

Finally, we include several variables that provide additional information about the broader context of these municipalities and communities (Table 5). For example, we include information about the number of full-time equivalent employees (FTE) from the Annual Survey of Public Employment and Payroll conducted by the U.S. Census. We also include a measure of the city's form of government, specifically whether or not it is a mayor-led form of government. This required compiling information from a number of sources: 2018 and 2011 surveys by the International City/County Management Association (ICMA), Tausanovitch and

Table 4. Socio-economic and political variables included in the MDWD.

Variable Name	Variable Description	Data Source; Manipulations
TOT_POP	Total Population	U.S Decennial Census/Census American Communities Survey
PCT_White	Percent of the population that's White	
PCT_Black	Percent of the population that's Black	
PCT_Hispanic	Percent of the population that's Hispanic	
PCT_25Plus_LessThanDiploma	Percent of the population over the age of 25 that has less than a high school diploma	
PCT_25Plus_Bachelors	Percent of the population over the age of 25 that has a bachelor's degree	
Median_Income	Median Income	
Below18_Poverty	Number of people under the age of 18 who are below the poverty line	
X18_64_Poverty	Number of people between the age of 18 and 64 who are below the poverty line	
X65Plus_Poverty	Number of people who are 65 years old or older who are below the poverty line	
Median_Housing_Age	Median Housing Age	U.S Decennial Census/Census American Communities Survey; 2017 minus Median Year Built
POV_POP	Total number of people below the poverty line	U.S Decennial Census/Census American Communities Survey; Added the total number of people from the three poverty categories (Below 18; 18–64; and 65+)
POV_PCT	Total number of people below the poverty line as a percent of the total population	U.S Decennial Census/Census American Communities Survey; Divided the total population below the poverty line by the total population from the ACS data
demshare_pres_2016	Democratic vote share in 2016 presidential election	Voting and Elections Science Team

https://doi.org/10.1371/journal.pwat.0000081.t004

Warshaw [24], the 1992 U.S. Census of Governments, and some hand collection and verification from city websites.

We include a measure of the climatic conditions using annual values of Willmott and Feddema's Climate Moisture Index [19]. The Moisture Index data archive provided a gridded time series of monthly moisture index values for the United States for our study period. Since these data come in a spatial grid, we assigned the closest gridded moisture index value to the local governments included in our study and aggregate monthly values to an annual average.

Results: Key features and uses of the MDWD

The final set of municipalities in the MDWD exhibit a wide range of financial, institutional, demographic, and environmental conditions (Table 6). For example, municipal populations range from 7,952 to 8.5 million residents and these populations have median household incomes that range from $13,149 to $250,001. The median age of a city's housing stock–a useful proxy for infrastructure age more broadly [21]–ranges from 12 to 78 years. Likewise, municipalities are operating within a wide range of institutional and financial capacities, such

Table 5. Additional variables included in the MDWD.

Variable Name	Variable Description	Data Source; Manipulations
Full_Time_Equivalent_Employees	Number of full-time equivalent employees for the municipality	Annual Survey of Public Employment and Payroll
month_moisture	Annual average moisture index; Higher value indicates more moisture; Lower value indicates dryer conditions	Willmott and Feddema's Moisture Index Archive
merged_FOG	Form of government; 1 = mayor; 0 = council	International City/County Management Association; Voting Sciences and Elections Team; Census of Governments

https://doi.org/10.1371/journal.pwat.0000081.t005

Table 6. Descriptive statistics for select MDWD variables using 2017 values.

	Variable	Mean	Median	Minimum	Maximum
Financial Measures	**Water Utility Spending**	$5.1 million	$1.7 million	$0	$747 million
	Health Spending	$1.1 million	$24,922	$0	$687 million
	Police Spending	$7.2 million	$2.3 million	$0	$2.3 billion
	Parks Spending	$2.7 million	$737,388	$0	$525 million
	Total Revenue	$74 million	$19 million	$183,800	$46 billion
	Total Debt	$90 million	$15 million	$0	$57 billion
City Government	**FTE in City Government**	878	242	0	445,385
	Form of Government (1 = mayor)	0.36	0	0	1
Demographics	**Population Size**	60,626	25,754	7,951	8,560,072
	Median Household Income	$55,842	$49,599	$13,149	$250,001
	Poverty Rate	14.4%	13.4%	0.4%	50.9%
	Proportion White Residents	77.0%	82.3%	1.8%	99.6%
	Median Housing Age	45 years	44 years	12 years	78 years
	Democratic Vote Share (2016)	0.53	0.51	0.092	0.99
Environment	**SDWA Violations**	0.15	0	0	28
	Climate Moisture Index Value	0.11	0.22	-0.99	0.88

https://doi.org/10.1371/journal.pwat.0000081.t006

as staff, revenue, and poverty rates. Finally, the municipalities in the MDWD are operating under varied climatic conditions: from very dry (climate moisture index of less than zero) to very wet (climate moisture index approaching 1). The municipal drinking water utilities represented in the MDWD collectively serve around 139 million people, or 40 percent of all U.S. residents. The dataset includes municipalities in all 48 contiguous U.S. states, and there is significant geographic variation in where these drinking water systems are located (Fig 2).

The MDWD is an unbalanced panel dataset and annual data availability for a given municipality varies (Fig 3). When using the MDWD, traditional approaches to dealing with an unbalanced panel dataset should be used [25]. Data availability is determined by the data source, such as the frequency and completeness of self-reported financial data, Census records (demographic data), and external datasets (Democratic vote share). The availability of financial and demographic data exhibits the greatest annual variability. Some variables, such as the climate moisture index, are available in the MDWD on an annual basis, whereas other variables are recorded as fixed values across years (e.g., form of government) or as a single annual value (e.g., 2016 Democratic vote share).

The MDWD includes information about the financial condition of the municipal governments, drawing from the Government Finance Database [26]. The most complete municipal financial data are available every five years: 1997, 2002, 2007, 2012, and 2017. We include a municipality's total revenue, own source revenue, total taxes, and total debt as key measures of financial position [27, 28]. We also include measures of city spending, such as the amount dedicated to drinking water and other core service areas such as health, police, and parks.

An initial descriptive analysis shows that the median level of spending on drinking water systems in 2017 was $1.7 million and ranged from $0 to $747 million. This level of spending is similar to municipal spending on police, which had a median of $2.3 million in 2017, compared to parks or health (median values of $24,922 and $737,388, respectively). The proportion of a city's budget dedicated to a particular critical service area is often used as a proxy for the priority given to a particular service domain [29] and can be calculated with the MDWD. For example, in 2017 municipalities dedicated an average of 13.8 percent of their annual spending to police, 5 percent to parks, and 0.8 percent to health. Per capita spending in all areas has

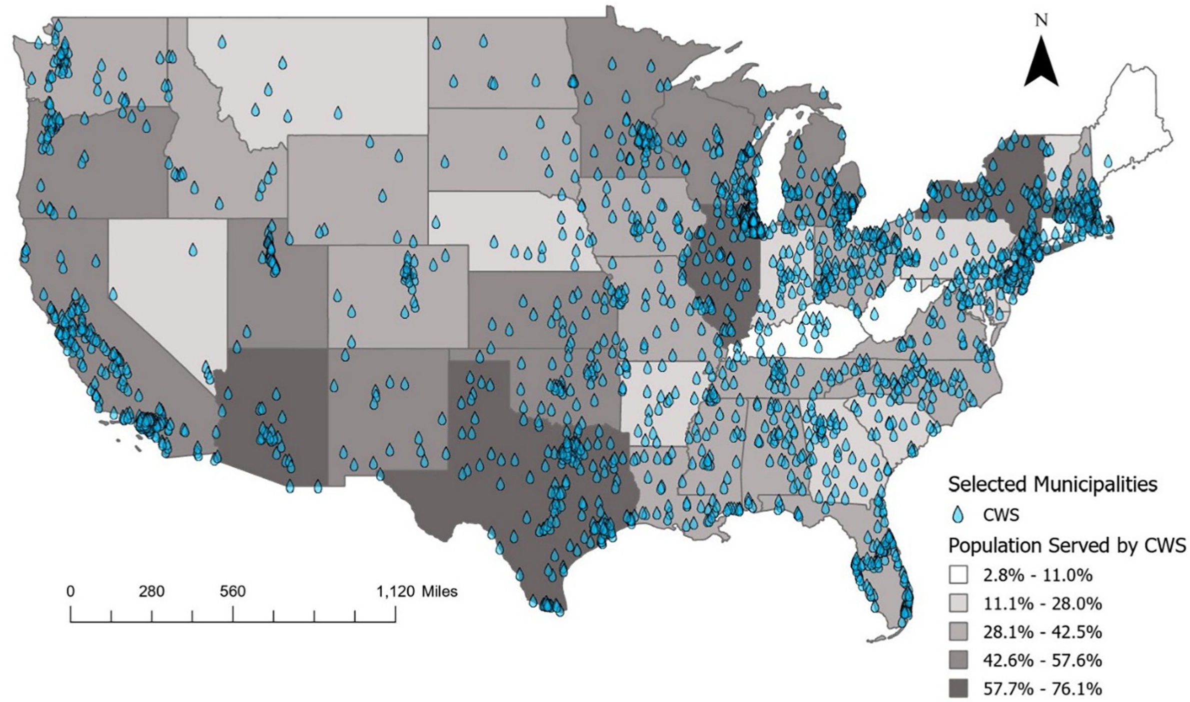

Fig 2. Location of municipal CWS in the MDWD and the proportion of each state's population they collectively serve. Map base layer and technical documentation available from U.S. Census 2020 TIGER shapefiles for states.

https://doi.org/10.1371/journal.pwat.0000081.g002

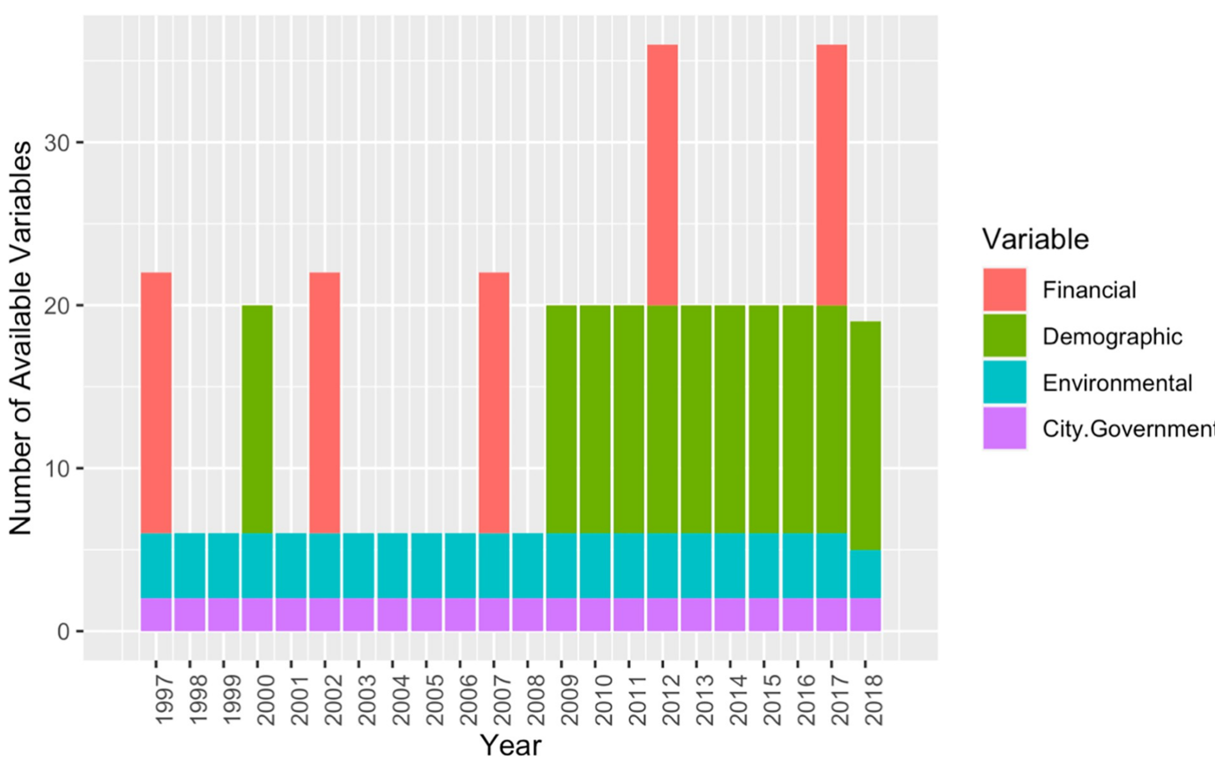

Fig 3. Annual availability in the MDWD of financial, city government, demographic, and environmental variables.

https://doi.org/10.1371/journal.pwat.0000081.g003

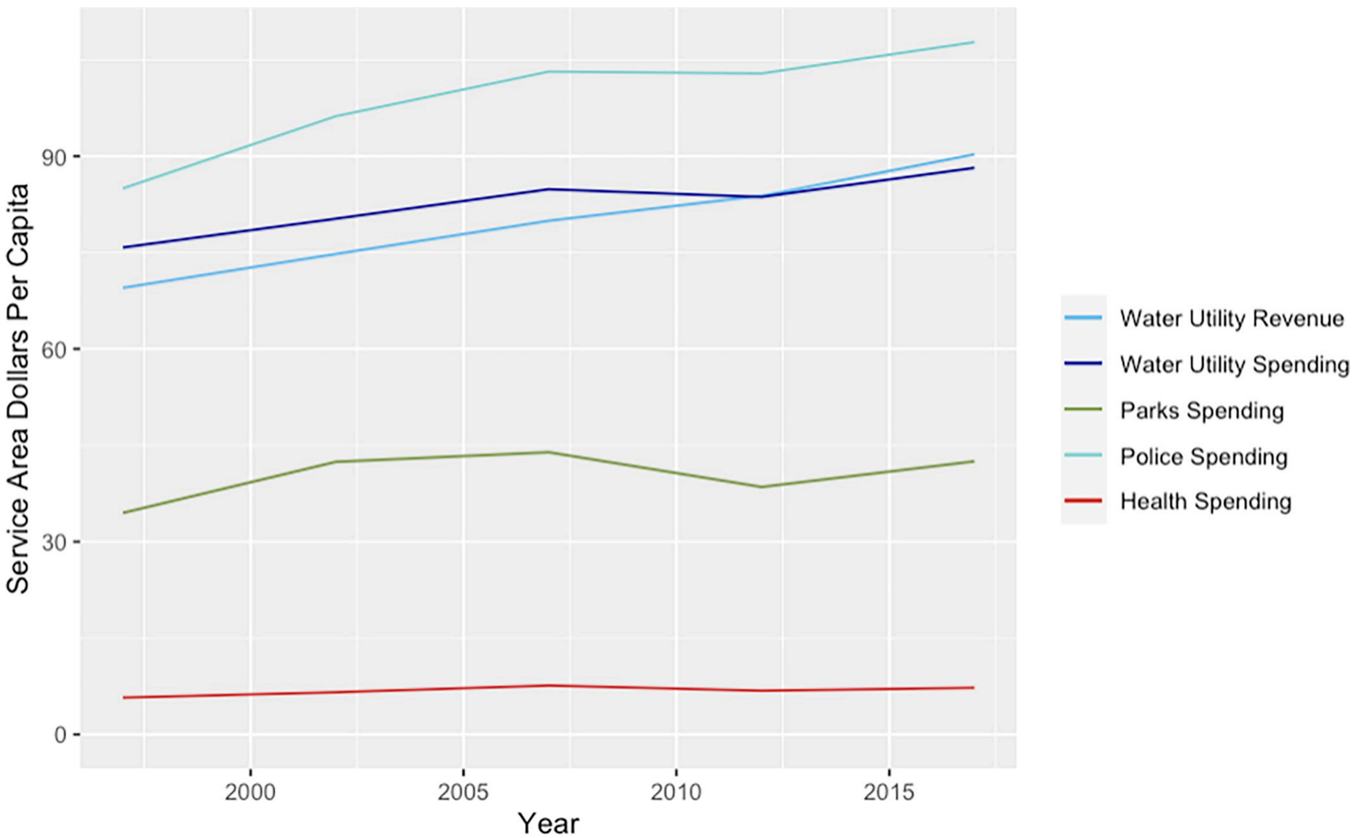

Fig 4. Inflation-adjusted per capita revenue and spending over time.

https://doi.org/10.1371/journal.pwat.0000081.g004

generally increased over time, and police spending has seen the largest increase (Fig 4). We can also examine patterns of water utility revenue and expenditures: the MDWD shows water utility revenue is currently outpacing water utility spending, but this has historically not been the case (Fig 4).

The MDWD also includes two variables that describe each city government: the city's form of government and number of full-time equivalent (FTE) employees. For form of government we include a binary measure of the institutional structure of the city government: whether or not it is a mayor-led city government. A large body of research has demonstrated a relationship between this institutional measure and a range of local policy outcomes [30–32]. We assume a municipality's form of government is consistent throughout the time-period captured by the MDWD (1997–2018). We also include information about the number of FTEs in each city government over time, drawing on data from the U.S. Census of Governments. Fig 5 illustrates the distribution of FTEs per capita in MDWD cities in 2017; the median number of total FTE is 242.

Drawing on data provided by the U.S. Census and 5-year American Community Survey estimates, we include in the MDWD a range of demographic variables. For example, we include municipal-scale information about race (percent white and percent Black); education levels; population size; and median household income. If desired, additional U.S. Census variables could be added to the database to meet the needs of individual research projects using FIPS codes to link new data to the MDWD. We also include information about the political characteristics of each municipality's residents, drawing on municipal-scale election data

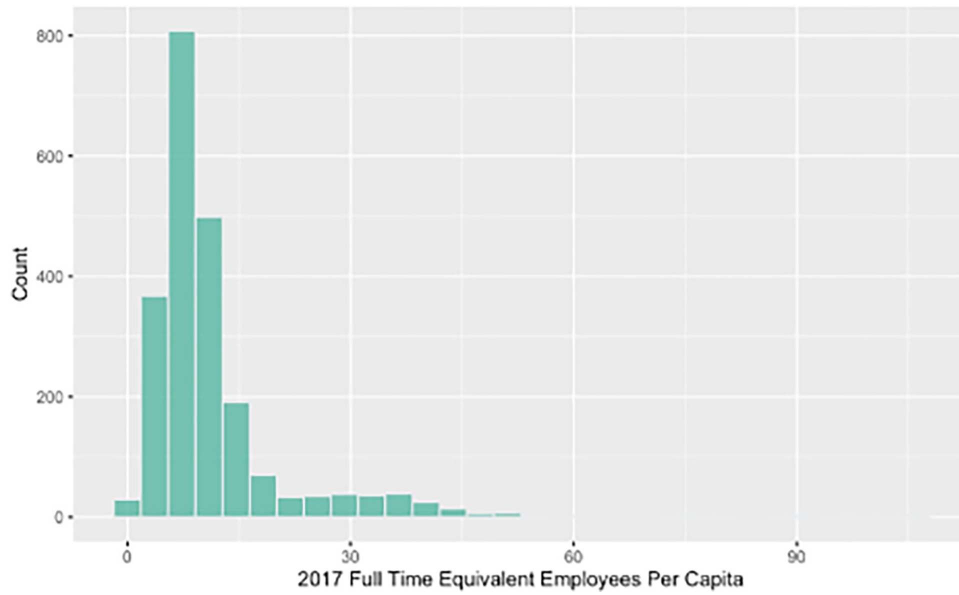

Fig 5. Distribution of Full Time Equivalent (FTE) employees in MDWD city governments in 2017.

https://doi.org/10.1371/journal.pwat.0000081.g005

compiled by the Voting and Election Science Team and provided for the MDWD [22, 23]. Aggregating precinct and county level election data to the municipal scale is time and labor intensive which make these data very difficult to access. We include the percent of each municipality that voted Democratic in the 2016 presidential election as a measure of the political characteristics or leanings of the population (Fig 6).

Finally, the MDWD includes environmental and performance information about municipal CWS. We use the Climate Moisture Index developed by Willmott and Feddema [19] to develop aggregate annual values for "the relationships between climate and the availability of

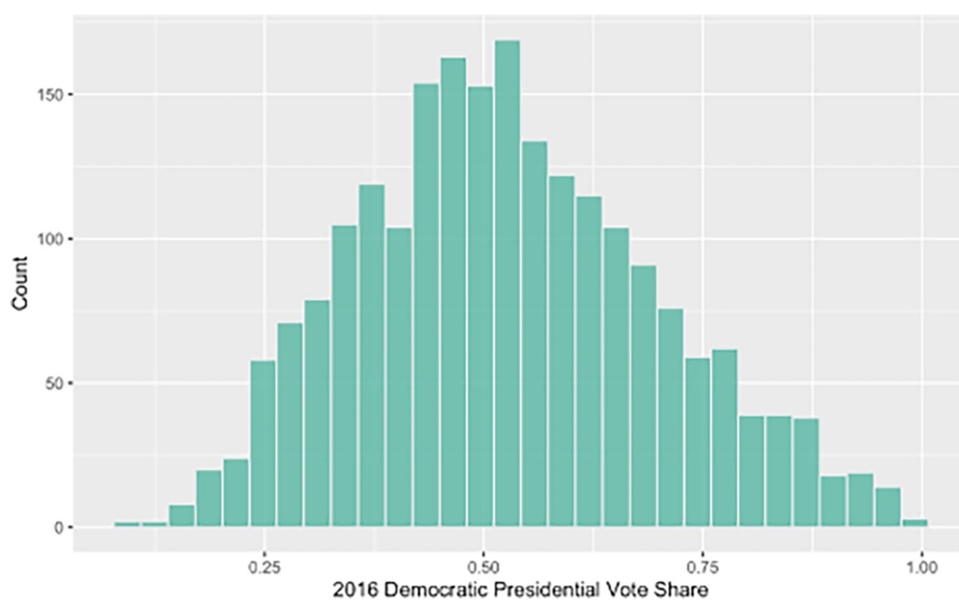

Fig 6. Distribution of the 2016 democratic vote share in MDWD cities.

https://doi.org/10.1371/journal.pwat.0000081.g006

moisture at the earth's surface" at the geographic location of each municipal CWS between 1997 and 2017. The values range from -1 to 1, indicating a more (-1) to less (1) arid climate. This environmental variable has been shown to correlate with decisions about drinking water management [33]. Climate moisture index values are available annually in the database. We also include information about the source water and incidence of SDWA violations for each CWS, both gathered from the USEPA's Safe Drinking Water Information System (SDWIS). Integrating SDWIS data with information from the Census and other datasets provides particularly new and innovative opportunities to understand municipal drinking water systems. For example, comparing drinking water spending per capita across regions of the country shows a potential relationship between regional climate and spending on drinking water systems: in 2017, regions with lower climate moisture scores (e.g., EPA regions 8 and 9) tended to spend more per capita on their drinking water systems (Fig 7).

Discussion and conclusion

Safe, affordable, accessible drinking water is a critical human need and core service area of municipal governments. Understanding the broader social, political, economic, and environmental conditions shaping the performance of drinking water systems has been hindered by data availability and compatibility (21). The MDWD is a publicly available resource to enable better understanding of how financial, environmental, demographic and governance conditions and characteristics affect the decisions made about drinking water services and the performance of these drinking water systems over time. Our descriptive analyses of a subset of these data present some initial insights and point toward opportunities for further investigation. Users of the MDWD can also use the various identifier variables to expand the number of variables and years included in the database.

While the MDWD provides new possibilities for understanding U.S. drinking water systems, there are data limitations. As an assimilation of existing datasets, the MDWD shares the intermittent availability and uncertainties associated with those datasets. Using the MDWD in longitudinal analysis will therefore require careful attention to the availability of key variables and the use of appropriate statistical techniques to account for this availability. The MDWD also focuses exclusively on municipally owned and operated drinking water systems, and takes a conservative approach to identifying these (see Methods below). The questions the database can answer are therefore limited to the operation and performance of municipal drinking water systems rather than privately owned or special purpose water systems. Finally, the MDWD provides a truncated set of variables that are available in the "feeder" datasets such as the U.S. Census, American Community Survey, and Government Finance Database. However, because the MDWD provides a range of identifier codes for each municipality (e.g., FIPS code, PWSID, and GOVSid), it is straightforward to add variables to the database as needed for specific research projects.

Our aim with the MDWD is to provide a novel and publicly available resource that can be used to generate new insights into the dynamics and performance of municipal drinking water systems. Local governments in the U.S. play an outsized role in ensuring the delivery of safe and affordable drinking water to residents, particularly as state and federal resources for drinking water infrastructure have declined [34]. Lack of data availability and compatibility has challenged our ability to better understand decision making by municipal governments. This database is a first step toward better understanding the dynamics of municipal drinking water systems and service provision, and how these are shaped by local context and decision making.

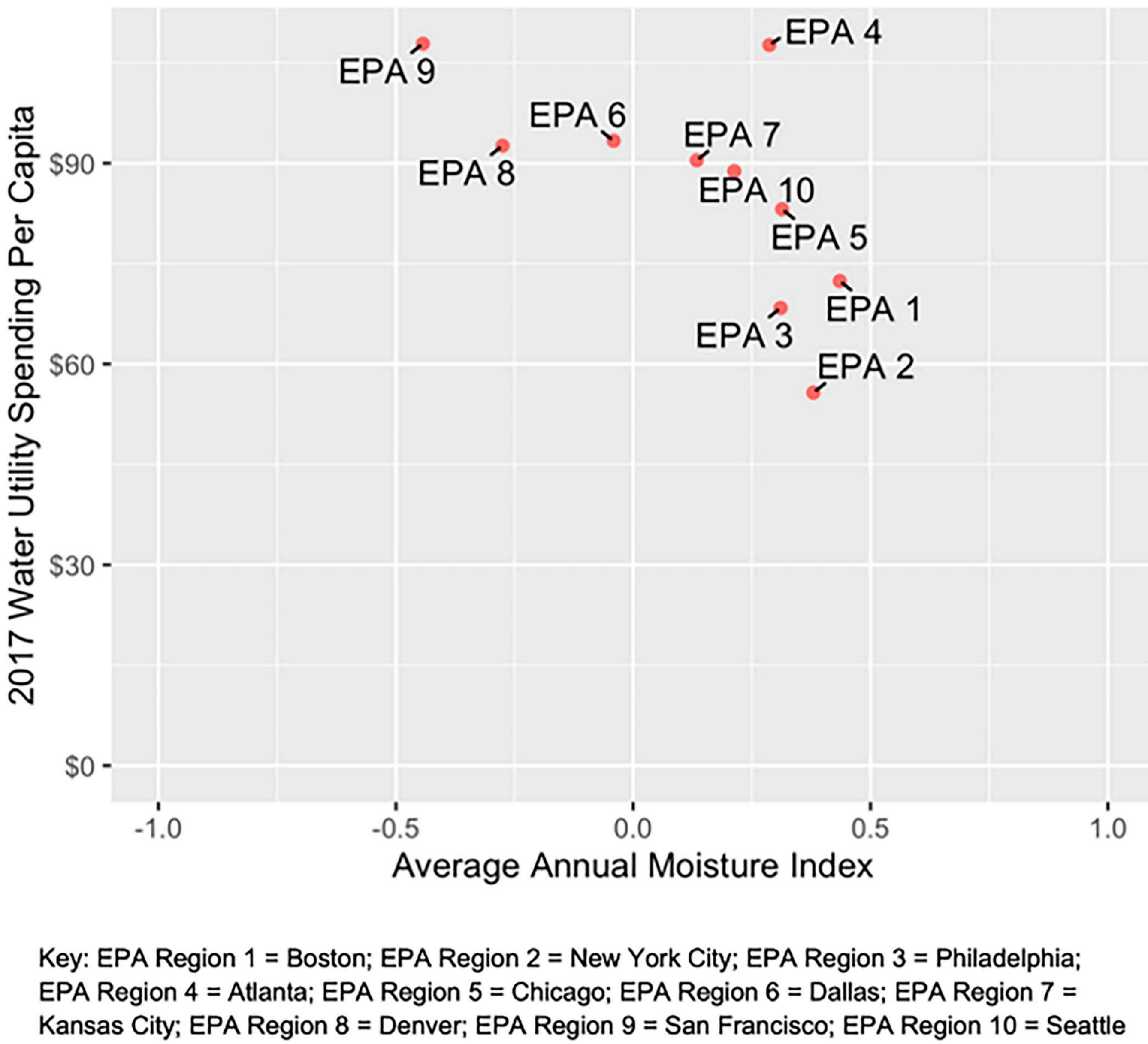

Fig 7. Water utility spending and climatic conditions across EPA regions.

https://doi.org/10.1371/journal.pwat.0000081.g007

Author Contributions

Conceptualization: Sara Hughes, Christine J. Kirchhoff.

Data curation: Katelynn Conedera, Mirit Friedman.

Formal analysis: Mirit Friedman.

Funding acquisition: Sara Hughes, Christine J. Kirchhoff.

Investigation: Christine J. Kirchhoff, Katelynn Conedera.

Methodology: Sara Hughes.

Project administration: Sara Hughes, Christine J. Kirchhoff.

Resources: Sara Hughes, Christine J. Kirchhoff.

Supervision: Sara Hughes, Christine J. Kirchhoff.

Writing – original draft: Sara Hughes, Christine J. Kirchhoff.

Writing – review & editing: Katelynn Conedera, Mirit Friedman.

References

1. Allaire M, Wu H, Lall U. National trends in drinking water quality violations. Proc Natl Acad Sci. 2018 Feb 27; 115(9):2078–83. https://doi.org/10.1073/pnas.1719805115 PMID: 29440421
2. Meehan K, Jurjevich JR, Chun NMJW, Sherrill J. Geographies of insecure water access and the housing–water nexus in US cities. Proc Natl Acad Sci. 2020 Oct 28; 117(46):28100–707.
3. Hughes S. Flint, Michigan and the Politics of Safe Drinking Water in the United States. Perspect Polit. 2020.
4. Mueller JT, Gasteyer S. The widespread and unjust drinking water and clean water crisis in the United States. Nat Commun. 2021; 12(1):1–8.
5. Kirchhoff CJ. Understanding and enhancing climate information use in water management. Clim Change. 2013 Jul 1; 119(2):495–509.
6. Dilling L, Daly ME, Travis WR, Wilhelmi OV, Klein RA. The dynamics of vulnerability: why adapting to climate variability will not always prepare us for climate change. Wiley Interdiscip Rev Clim Change. 2015; 6(4):413–25.
7. Page R, Dilling L. How experiences of climate extremes motivate adaptation among water managers. Clim Change. 2020 Aug 1; 161(3):499–516.
8. Mullin M. Governing the Tap: Special District Governance and the New Local Politics of Water. Cambridge, Massachusetts: MIT Press; 2009. 266 p.
9. Hughes S, Mullin M. Local Water Politics. In: Oxford Handbook on Water Politics and Policy. Oxford and Cambridge, MA: Oxford University Press; 2018. p. 284–309.
10. USEPA. Information about public water systems [Internet]. Washington, D.C.: United States Environmental Protection Agency; 2017. Available from: https://www.epa.gov/dwreginfo/information-about-public-water-systems
11. U.S. EPA. FACTOIDS: Drinking Water and Ground Water Statistics for 2007 [Internet]. U.S. Environmental Protection Agency; 2008. Available from: http://www.epa.gov/safewater/data/pdfs/data_factoids_2007.pdf
12. Switzer D, Teodoro MP. The Color of Drinking Water: Class, Race, Ethnicity, and Safe Drinking Water Act Compliance. J—Am Water Works Assoc. 2017 Dec 1; 109(9):40–5.
13. McDonald YJ, Anderson KM, Caballero MD, Ding KJ, Fisher DH, Morkel CP, et al. A systematic review of geospatial representation of United States community water systems. AWWA Water Sci. 2022; 4(1): e1266.
14. Chernick H, Reschovsky A. The fiscal condition of US cities: Revenues, expenditures, and the "Great Recession." J Urban Aff. 2017; 39(4):488–505.
15. Sapotichne J, Johnson M, Park YS. Stability and change in US city policymaking: evidence and a path forward. Urban Res Pract. 2013; 6(3):255–75.
16. Scott TA, Moldogaziev T, Greer RA. Drink what you can pay for: Financing infrastructure in a fragmented water system. Urban Stud. 2018; 55(13):2821–37.
17. Teodoro MP. Water and sewer affordability in the United States. AWWA Water Sci. 2019; 1(2):e1129.
18. Willmott CJ, Matsuura K. Moisture Index Archive: Monthly Time Series (1900–2017) [Internet]. 2018. Available from: http://climate.geog.udel.edu/~climate/html_pages/Global2017/README.GlobalImTs2017.html
19. Willmott CJ, Feddema JJ. A More Rational Climatic Moisture Index*. Prof Geogr. 1992; 44(1):84–8.
20. Conedera K. Census Geography Reference [Internet]. 2021. Available from: https://github.com/wcplab/censusgeog/tree/v0.1
21. Hughes S. A multidimensional approach to evaluating the vulnerability of drinking water systems. J Environ Policy Plan [Internet]. 2021 Nov 11 [cited 2021 Nov 12]; Available from: https://www.tandfonline.com/doi/abs/10.1080/1523908X.2021.2000377

22. Warshaw C, de Benedictis-Kessner J, Velez Y. Local Representation in the United States: A New Comprehensive Dataset of Elections. Work Pap [Internet]. 2022; Available from: http://www.chriswarshaw.com/papers/LocalElections_Data_combined.pdf
23. Voting and Election Science Team. 2016 Precinct-Level Election Results. Harvard Dataverse, V86. 2018.
24. Tausanovitch C, Warshaw C. Measuring constituent policy preferences in congress, state legislatures, and cities. J Polit. 2013; 75(2):330–42.
25. Biørn E. Econometrics of Panel Data: Methods and Applications. Oxford: Oxford University Press; 2016.
26. Pierson K, Hand ML, Thompson F. The Government Finance Database: A Common Resource for Quantitative Research in Public Financial Analysis. PLOS ONE. 2015 Jun 24; 10(6):e0130119. https://doi.org/10.1371/journal.pone.0130119 PMID: 26107821
27. Gordon T. Predicting Municipal Fiscal Distress: Aspiration or Reality. Lincoln Institue of Land Policy; 2018.
28. Kim Y, Warner ME. Pragmatic municipalism or austerity urbanism? Understanding local government responses to fiscal stress. Local Gov Stud. 2020;1–19.
29. Park AY, Sapotichne J. Punctuated Equilibrium and Bureaucratic Autonomy in American City Governments. Policy Stud J. 2020; 48(4):896–925.
30. Lubell M, Feiock R, Handy S. City Adoption of Environmentally Sustainable Policies in California's Central Valley. J Am Plann Assoc. 2009; 75(3):293–308.
31. Krause RM. Symbolic or substantive policy? Measuring the extent of local commitment to climate protection. Environ Plan C Gov Policy. 2011; 29(1):46–62.
32. Hughes S, Runfola DM, Cormier B. Issue Proximity and Policy Response in Local Governments. Rev Policy Res. 2018; 35(2):192–212.
33. Teodoro MP. The institutional politics of water conservation. J-Am Water Works Assoc. 2010; 102 (2):98–111.
34. ASDWA. Insufficient Resources for State Drinking Water Programs Threaten Public Health. An Analysis of State Drinking Water Programs' Resources and Needs. Associatoin of State Drinking Water Administrators; 2013.