EPIC's Drinking Water Explorer Tool - Methodology

Version 1.0
Last application data update occurred on: February 20th, 2026
Data dictionary
Github link

Go back to the top

Community Water System Service Area Boundaries

Maintained by: Environmental Protection Agency (EPA)
Website link

- Internal update frequency:
  - Annually on Oct 1st.
  - Continuous data collection started on: Oct 1st, 2025
- Description:
  - This dataset contains the service area boundaries for Community Water Systems for the U.S. and select territories (Puerto Rico, Guam, and the Northern Mariana Islands), which describes the geographic area that receives drinking water from a water system. Since the latest update in October 2025, over half of the ~50,000 water system service area boundaries are derived from authoritative data sources.
- Caveats:
  - Boundaries for water systems may be modeled, and/or overlap with other water systems. Please note there is one duplicated system in the raw dataset "ND2801430" (we removed this duplicate for the application) and a water systems may be associated with more than one water system IDs, "PR0005086; PR0005066" for example.
- EPIC's data score: 83%
  - This score is based on the data contained within the service area boundary dataset (water system ID, water system name, etc.) and the geometries of these boundaries. To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 11 different variables. Combined with our estimate for how well these variables reflect the larger data universe, this produced the mean quality score above. For boundaries, we checked for the presence of overlaps, which produced a quality score of 82% (in other words, approximately 82% of service areas do not overlap with other service area boundaries). This produced a mean score for the dataset.
- Recommended uses:
  - Estimating the area served by a community water system
  - Estimating the population characteristics of a served by a community served by a water system
- Methods:
  - EPA SABs were downloaded from GitHub. When a PWSID contained multiple boundaries (specifically these: "FL1070685", "FL1190789", "FL6580531", "IA2573701", "MT0001923", "VA5019052"), the boundaries were merged to create a single multipolyon. There was one distinct SAB that appeared to be duplicated, "ND2801430", and the boundary that was generated by the EPA ORD was retained, as the original source appeared to just contain water districts. There are also a couple of boundaries with multiple water system IDs (PR0005086; PR0005066, for example) and these were kept as a single record. This means that we will likely not have water system information for this record (since it contains

Go back to the top

records for two separate water systems), but we will have interpolated data (income, climate vulnerability indexes, etc.) for this boundary. Invalid geometries were corrected using the st_make_valid() function, and boundaries were simplified using a st_simplify() function. All area calculations were completed using the st_area() function using an alberts equal area projection (crs = 5070).

Safe Drinking Water Information System

Maintained by: Environmental Protection Agency (EPA)
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
  - Quarterly
- Description:
  - Water system violation, enforcement, and system information from the federal Safe Drinking Water Information System (SDWIS).
- Caveats:
  - There are known issues with water systems underreporting their violations
  - The source water protection code is only available for certain states in SDWIS. This is because SDWIS tracks whether the system has implemented source water protection according to state policy.
  - The maximum age of a system is often set at ~45 years old because the "first_reported_date" (first reported date for the system, based on SDWA data dictionary) in SDWIS is "1979-03-01" or early 1980. Considering SDWA started in 1974, these systems may have been operating for longer, but this is the date they started being tracked by the EPA.
- EPIC's data score: 76.6%
  - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 40 different variables. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score.
- Recommended uses:
  - Calculating the number & types of violations & enforcement actions a water system has had over the past years
  - Gathering water system characteristics (water system ID, name, years in operation, etc.)
- Methods:
  - Downloaded data from the bulk SDWA download and pulled out the violations_enforcement dataset, the ref_code values, and the pub_water_systems dataset. Since the violations and enforcement data contain duplicated rows and IDs by violation and enforcement action, a unique ID was created by concatenating the

Go back to the top

pwsid and violation id and duplicates were removed. Data were filtered based on water system IDs within the SABs dataset.
    - Disaggregated health-based violations by rule codes were selected by where the "is_health_based_ind" field was "Y", data were filtered to where "compl_per_begin_date" (the begin period in which the violation was identified), and merged using the ref_code dataset identify specific rule violations. Violations were analyzed over the past 10 years (i.e., where the violation year > the year the worker last ran - 10 years), and 5 years using the same method. If the system has been operating for less than 5 or 10 years, relevant violation summaries were replaced with "Not Enough Data" to avoid incorrect comparisons across systems. In the tool, this shows up as “NA”.
    - Open health-based violations where instances where "is_health_based_ind" is "Y" and where the violation status was "addressed" or "unaddressed" (i.e., not "resolved" or "archived").
    - Violations over the past 5 or 10 years that required a public notification (simplified to "pn") were instances where the public_notification_tier == 1. All of these were health-based violations.
    - Phone numbers, boolean values, and gw_sw_codes were standardized for the application.

Drinking Water State Revolving Funds Awards

Maintained by: Environmental Protection Agency (EPA)
Related to service area boundaries by: EPIC through PWSIDs
Weblink

- Internal update frequency:
    - Yearly, as updated by EPA.
    - The last updated occurred on 2026-02-16. We are working to automate this process.
- Description:
    - This dataset contains downloaded data from EPA’s State Revolving Fund Public Portal for federal fiscal years 2021-2025.
- Caveats:
    - This dataset only contains water systems that received Drinking Water State Revolving Funding from 2021-2025 or amendments during this time period.
    - Principle forgiveness represents "Additional Subsidy Amount", which includes direct principle forgiveness, as well as negative interest rates and grants.
    - Some peculiarities with latest assistance agreement dates, affecting two projects with dates as 1919 & 1997 - likely due to data entry. These records are not removed for the purposes of the application.
- EPIC’s data score: 100%

Go back to the top

    - To create this score, we checked for the data completeness (how many “NA”s are there?) and how many duplicated records exist (if applicable) across the different variables. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score.
- Recommended uses:
    - Identifying which systems have or have not received assistance from the DWSRF program from 2021-2025.
- Methods:
    - Data was manually downloaded from the EPA’s State Revolving Fund Public Portal. Three variables are created. Data is grouped by PWSID to represent the total number of projects. Total assistance corresponds to the Current Agreement Amount. Note that principle forgiveness represents "Additional Subsidy Amount", which includes direct principle forgiveness, as well as negative interest rates and grants.

Census Variables

Maintained by: U.S. Census Bureau & EPA
Related to service area boundaries by: EPIC
Data link

- Internal update frequency:
    - Quarterly (requires manual update)
- Description:
    - ACS 2021 estimates for EPA's Service Area Boundaries using methods developed by EPIC. This includes over 70 different census variables.
- Caveats:
    - Please note that the EPA ORD crosswalk is missing approximately 1,296 water systems. As a result, we may be missing certain count variables for water systems but have estimated income variables (which were interpolated based on water system boundary).
- EPIC’s data score: 76%
    - To create this score, we checked for the data completeness (how many “NA”s are there?) and how many duplicated records exist (if applicable) across 30 different variables that were created after cleaning our crosswalked / interpolated dataset. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
    - Estimation of the number of people captured within a service area and their main characteristics (race, ethnicity, income, etc.)
    - Identifying water systems that may serve vulnerable communities (elderly, children, low income, etc.)
- Methods:

Go back to the top

- The census tract and service area boundary crosswalk file was downloaded from the EPA ORD github website. Essentially, this table uses Microsoft building footprints to weight census data to water system service area boundaries. Data from duplicated systems were removed (specifically these: "FL1070685", "FL1190789", "FL6580531", "IA2573701", "MT0001923", "VA5019052"), as they sometimes produced weights that were greater than one. Data from the ACS 2021 5-year estimate were collected. All non-income variables were multiplied by the specific water system and tract weight, grouped by the water system ID, and then summed to estimate totals at the water system level.
- Prior to weighted interpolation, income variables (median household income, lowest quintile of income) were related to the percentage of the total universe (essentially median household income divided by the sum of median household incomes across the entire U.S.) to generate a percentage that could be treated as spatially intensive. Weights for weighted interpolation were created using the number of households in census blocks from the 2020 decennial census. All spatial datasets used for weighting were projected to Albert's equal area projection (crs = 5070) prior to interpolating.
- After interpolation, the percentage was multiplied by the total universe (i.e., sum of median household income across the entire U.S.) to relate the variable back to income. Water systems that overlap with multiple states were handled using a mean.
- Because census count variables and income variables are handled differently (i.e., a crosswalk file versus household weighted interpolation), there may be instances where a water system does not have census count estimates (i.e., total population) but income estimates (median household income), and vice versa.

Percent Change in Census Variables

Maintained by: U.S. Census Bureau
Related to service area boundaries by: EPIC
Data link

- Internal update frequency:
  - Quarterly (requires manual update)
- Description:
  - The 10-year percent change in census variables for EPA's Service Area Boundaries using methods developed by EPIC. This includes over 70 different census variables.
- Caveats:
  - Not all census variables were available for the 2011 ACS 5-year estimate. As a result, the percent change in annual water and sewer rates, schooling information, population without a computer, and health insurance could not be estimated.
  - Water systems with duplicated boundaries could not be calculated
- EPIC's data score: 76%

Go back to the top

    - To create this score, we checked for the data completeness (how many “NA”s are there?) and how many duplicated records exist (if applicable) across 11 different variables that were created after cleaning our crosswalked / interpolated dataset. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
    - Estimation of how the population characteristics contained within a service area has changed over the past 10 years.
    - Identifying what service areas are serving communities that have rapidly grown/shrunk in people and income.
- Methods:
    - The census tract and service area boundary crosswalk file was downloaded from the EPA ORD github website. Essentially, this table uses Microsoft building footprints to weight census data to water system service area boundaries. Data from duplicated systems were removed (specifically these: "FL1070685", "FL1190789", "FL6580531", "IA2573701", "MT0001923", "VA5019052"), as they sometimes produced weights that were greater than one. Then, we downloaded NHGIS crosswalk files to relate 2020 census tracts back to 2010 census geometries.
    - To do this, we merged the EPA ORD crosswalk with the NHGIS crosswalk file and multiplied the total number of Microsoft building footprints by the 2010 tract housing weight to get an estimate of the 2010 tract building total. Here, we’re assuming that the housing weights would likely be similar/correlated to the number of buildings. After that, we then grouped by the 2010 census tract geoid and summed unique tract building totals. The unique values were summed because a single tract may overlap with multiple service areas, and therefore have the same tract building total. This avoids double counting. The new building totals for the 2010 census tracts were then added back to the main (ORD crosswalk and NHGIS file) to create a new column with the 2010 census tract building totals.
    - We then multiplied the number of buildings that overlapped in the ORD crosswalk file by the census tract housing weights from the NHGIS file to estimate how many buildings overlapped with EPA service area boundaries.
    - This final dataset is then grouped by water system ID, 2010 census tract, and 2010 census tract microsoft building totals, and the total number of buildings that overlap with service area boundaries (after the weight is applied above) is summed. This creates a new weight by dividing this number by the total number of estimated Microsoft buildings in the 2010 census tract.
    - Data from the ACS 2011 5-year estimate were collected. All non-income variables were multiplied by the specific water system and tract weight, grouped by the water system ID, and then summed to estimate totals at the water system level. Not all census variables are available in 2011 (i.e., annual water and sewer rates,

Go back to the top

schooling information, population without a computer, and health insurance), and the percent change in these variables cannot be computed.

- Prior to weighted interpolation, income variables (median household income, lowest quintile of income) were related to the percentage of the total universe (essentially median household income divided by the sum of median household incomes across the entire U.S.) to generate a percentage that could be treated as spatially intensive. Weights for weighted interpolation were created using the number of households in census blocks from the 2020 decennial census. . All spatial datasets used for weighting were projected to Albert's equal area projection (crs = 5070) prior to interpolating.
- After interpolation, the percentage was multiplied by the total universe (i.e., sum of median household income across the entire U.S.) to relate the variable back to income. Water systems that overlap with multiple states were handled using a mean.
- If a water system has been operating for < 10 years (based on the water system dataset from the EPA's Safe Drinking Water Information System), these values are replaced with "NA".
- To calculate the percent change, the following formula was used after adjusting income variables for inflation using procedures similar to the inflation adjuster from the Bureau of Labor Statistics: 100*((2021 census variable - 2011 census variable) / 2011 census variable)

EJScreen

Maintained by: Public Environmental Data Partners
Originally developed by: Environmental Protection Agency
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
  - Quarterly (requires manual update)
- Description:
  - As described on the tool website, "EJScreen, the Environmental Justice Screening and Mapping Tool, is a free and public online screening and mapping tool that contains environmental and demographic data on different areas of the United States."
- Caveats:
  - The drinking water score was developed based on an older vintage of the EPA service area boundaries, which were related to the block group level in EJScreen. Considering service area boundaries have been updated since this score was created, EPIC is working to recreate this drinking water score at the water system level.
- EPIC's data score: 74%

Go back to the top

        - To create this score, we checked for the data completeness (how many “NA”s are there?) and how many duplicated records exist (if applicable) across 3 different variables that were created after cleaning our interpolated dataset. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 74%.
- Methods:
    - Data from EJScreen, as hosted on the harvard dataverse website, were downloaded and filtered to remove data from territories (gathering census data for these locations is more difficult, but it is on our radar for the next version of the tool). This dataset was merged with ACS 2022 block group geometries by geoid. EJScreen data were estimated to EPA SABs using a mix of interpolation methods. If the variable was population-focused, the variable was interpolated using census block populations or houses as weights (based on whether the variable is population or household focused). If the variable was area-based, then areal interpolation was used. For both of these methods, variables were treated as either spatially intensive (percentiles), or extensive (counts). All spatial datasets were projected to Albert's equal area projection (crs = 5070) prior to interpolating.
    - Due to inconsistencies in the current SABs & the SABs used in the original EJScreen calculation, the drinking water non-compliance score is not currently available. We hope to include the drinking water non-compliance score in the future based on the latest information and boundaries provided by the EPA.

Climate and Economic Justice Screening Tool

Maintained by: Public Environmental Data Partners
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Quarterly (requires manual update)
- Description:
    - As described on the website, “The CEJST uses publicly-available, nationally-consistent datasets to identify disadvantaged communities.The datasets are indicators of burdens that disadvantaged communities face. These burdens are related to climate change, the environment, health, and economic opportunity. The CEJST features a userfriendly, searchable map of all 50 states, the District of Columbia, and the U.S. territories. Communities are considered disadvantaged if they are in census tracts that meet the thresholds for at least one of the tool’s categories of burden, or if they are on lands within the boundaries of Federally Recognized Tribes. "
- Caveats:
    - Data from CEJST were interpolated (see methods for details). As a result, some variables for some water systems could not be calculated.

Go back to the top

- EPIC's data score: 74%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 4 different variables that were created after cleaning our interpolated dataset. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 74%.
- Recommended uses:
    - Identifying water systems that serve communities that are identified as disadvantaged by CEJST
- Methods:
    - Data from the CEJST, as hosted on the PEDP website, were downloaded. To do this, the shapefile was downloaded to obtain the codebook and census tract geometries. The census tract geometries were combined with the community list data, which contains the variables of interest.
    - CEJST variables were estimated to EPA SABs using a mix of interpolation methods. If the variable was population-focused, the variable was interpolated using census block populations or houses as weights (based on whether the variable is population or household focused). If the variable was area-based, then areal interpolation was used. For both of these methods, variables were treated as either spatially intensive (percentiles), or extensive (counts). For this method, 2010 census geometries were used to match the vintage of CEJST. All spatial datasets were projected to Albert's equal area projection (crs = 5070) prior to interpolating.

Social Vulnerability Index

Maintained by: Center for Disease Control
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Quarterly (requires manual update)
- Description:
    - As described on the website, "Social vulnerability refers to the demographic and socioeconomic factors (such as poverty, lack of access to transportation, and crowded housing) that adversely affect communities that encounter hazards and other community-level stressors. These stressors can include natural or human-caused disasters (such as tornadoes or chemical spills) or disease outbreaks (such as COVID-19)."
- Caveats:
    - Data from the SVI were interpolated (see methods for details). As a result, some variables for some water systems could not be calculated.
- EPIC's data score: 77%

Go back to the top

    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 2 different variables that were created after cleaning our interpolated dataset. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 77%.
- Recommended uses:
    - Identifying water systems that serve communities that are identified as vulnerable by the SVI
- Methods:
    - Data from the SVI were downloaded from the CDC website. For the rpl_themes flag, values of -999 were transformed to NAs. Based on documentation, this means that the value was unavailable from the original census data and this was to avoid this value being interpolated as a true numeric value. Data from the SVI were estimated to EPA SABs using a mix of interpolation methods. If the variable was population-focused, the variable was interpolated using census block populations or houses as weights (based on whether the variable is population or household focused). If the variable was area-based, then areal interpolation was used. For both of these methods, variables were treated as either spatially intensive (percentiles, for example), or extensive (counts of people, households, etc). The documentation doesn't explicitly state the census vintage, but since many of the variables are from the ACS 2022 5yr estimate, we used 2020 census block data for weighting. All spatial datasets were projected to Albert's equal area projection (crs = 5070) prior to interpolating.

Water System Watersheds for Well & Intake Locations

Maintained by: Environmental Protection Agency (EPA)'s How's my Waterway?
Website link

- Internal update frequency:
    - Static, received from EPA Oct 7th 2024
- Description:
    - Obtained from the EPA - contains the HUC12 of well and intake locations for water systems.
- Caveats:
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 8 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains data from NPDES, underground storage tanks, facilities with risk management plans, and impaired waterways. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.

Go back to the top

- Recommended uses:
    - Identifying which watershed(s) a water system pulls from
- Methods:
    - Data were collected via email, and the dataset was filtered to only contain records from EPA's community water system service area boundary dataset.

Climate Vulnerability Index

Maintained by: Environmental Defense Fund
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Quarterly (requires manual update)
- Description:
    - As described on the website, ""The U.S. Climate Vulnerability Index helps you see which communities face the greatest challenges from the impacts of a changing climate. This tool shows what is driving the challenges, so policymakers and communities themselves can take action to build climate resilience where it is needed most."
- Caveats:
    - Data from the CVI were interpolated (see methods for details). As a result, some variables for some water systems could not be calculated.
- EPIC's data score: 77%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 5 different variables that were created after cleaning our interpolated dataset. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 77%.
- Recommended uses:
    - Identifying water systems that serve communities that are vulnerable to climate change.
- Methods:
    - Data from the CVI were downloaded using the Master CVI dataset from their website (this spreadsheet contains the overall climate vulnerability and percentile scores) and related to 2010 census geographies. Estimated to EPA SABs using a mix of interpolation methods. If the variable was population-focused, the variable was interpolated using census block populations or houses as weights (based on whether the variable is population or household focused). If the variable was area-based, then areal interpolation was used. For both of these methods, variables were treated as either spatially intensive (percentiles), or extensive (counts). For this method, 2010 census geometries were used to match the vintage of the CVI.

Go back to the top

All spatial datasets were projected to Albert's equal area projection (crs = 5070) prior to interpolating.

National Pollutant Discharge Elimination System (NPDES) Permits

Maintained by: Environmental Protection Agency (EPA)'s ECHO
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
  - Quarterly
  - Continuous data collection started on: N/A (memory issue, working on it)
- Description:
  - Outfalls data from the National Pollutant Discharge Elimination System (NPDES) from the EPA. "The NPDES permit program addresses water pollution by regulating point sources that discharge pollutants to waters of the United States. Created in 1972 by the Clean Water Act, the NPDES permit program is authorized to state governments by EPA to perform many permitting, administrative, and enforcement aspects of the program."
- Caveats:
  - Please note that if a water system is pulling from a watershed that contains a lot of permitted outfalls, that does not necessarily mean the drinking water quality is at risk or of poor quality. Water systems employ a variety of treatment techniques to filter contaminants from the source water. This variable is aimed to communicate potential hazards, rather than an actual risk to drinking water.
  - With this method, since we sorted permitted features into HUC12s (& a single permit number may apply to multiple features), a single external_permit_nmbr may match to multiple HUC12s. So, a permit in violation could show up in multiple HUCs, even if that specific feature within the HUC12 is not in violation.
- EPIC's data score: 76%
  - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 8 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains data from NPDES, underground storage tanks, facilities with risk management plans, and impaired waterways. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
  - Identifying what water systems are pulling from a watershed that contains a certain amount of permitted discharge outfalls.
- Methods:
  - NPDES permits were downloaded using the ECHO outfalls layer. Data were filtered to contain permits with "Effective", "Admin Continued", or "Pending" status

Go back to the top

(permit_status_dec), and permits with locations outside of the U.S. were removed (more specifically, we buffered the U.S. boundaries to 50 miles offshore, and then ran the st_within function to capture permits that were within 50 miles of the shoreline - we recognize this is not the same as the definition of WOTUS, but since we're relating this information to HUC12s, which are on land, they would've been removed in the following step). Point data (containing all permitted features) were then sorted into HUC12s (specifically, we looped through each state, grabbed all of their HUC12 using the get_huc function, and made the CRS the same as the NPDES data before running an st_intersection with spherical geometry turned off), which further removed any offshore permits or permits with geometry errors.

- Data were simplified to permit-level data (i.e., feature codes and lat/longs were removed, since they contained duplicated permit data, and there data were simplified to unique entries), grouped by huc12, and the total number of unique external permit numbers (external_permit_nmbr) were counted, those with violations (where cwa_current_status is "violation identified" or "significant/category I noncompliance") were counted, those with significant violations (where cwa_current_stats is "significant/category I noncompliance") were counted, as well as the number of permits with effluent violations (where cwp_current_snc_status is "Effluent - Monthly Average Limit" or "Effluent - Non-monthly Average Limit"). It should be noted that violation data are from the "most current data available", including data that have been reported after the last quarterly noncompliance report was issued.
- With this method, since we sorted permitted features into HUC12s (& a single permit number may apply to multiple features), a single external_permit_nmbr may match to multiple HUC12s. So, a permit in violation could show up in multiple HUCs, even if that specific feature within the HUC12 is not in violation. We could fix this by sorting through the discharge monitoring reports and looking at data by permitted feature, but this is computationally intense (grabbing violation history can take ~5 mins, and we have ~400,000 of them) and could be considered in a future version. Also, if a permitted feature exists at the nexus of multiple HUC12s (MTR108729, for example), the data for this permit is duplicated across those overlapping HUC12s.
- All spatial datasets were projected to Albert's equal area projection (crs = 5070) prior to running intersections.

303(d) Impaired Waterways

Maintained by: Environmental Protection Agency (EPA)'s ATTAINS Dataset
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Quarterly

Go back to the top

    - Continuous data collection started on: N/A (memory issue, working on it)
- Description:
    - Data from assessed streams - "State water quality assessment decisions reported to EPA under the Integrated Report (IR), and Clean Water Act Sections 303(d) and 305(b). This service provides summary information for each Assessment Unit. For more detailed data, please reference the Assessment Total Maximum Daily Load (TMDL) Tracking and Implementation System (ATTAINS) web services. Information on those web services is provided on the public ATTAINS website: https://www.epa.gov/waterdata/attains"
- Caveats:
    - This information is reported by states and the amount of assessed waters and streams varies across states.
    - Please note that if a water system is pulling from a watershed that contains a lot of impaired waterways, that does not necessarily mean the drinking water quality is at risk or of poor quality. Water systems employ a variety of treatment techniques to filter contaminants from the source water. This variable is aimed to communicate potential hazards, rather than an actual risk to drinking water.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 8 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains data from NPDES, underground storage tanks, facilities with risk management plans, and impaired waterways. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
    - Identifying what water systems are pulling from a watershed that contains a certain amount of impaired waterways
- Methods:
    - Data downloaded from ArcGIS REST services, grouped by HUC12 and summarized to identify the number of assessed streams, how many of those are impaired, threatened, and how many of them are on a 303d list (impaired stream without a TMDL). These sums are based on the presence of a "Y" in the assessed, impaired, threatened, and on303dlists. The not_assessed column was based on the presence of "N" in the isassessed column. Please note all of these data are state reported from the ATTAINS web services, and there is a bunch of variation between states.
    - Based on the metadata:
        - isassessed = If the state has monitored a water and made an Assessment decision about the Assessment Unit, it is considered Assessed. If the state has defined the Assessment Unit but has not yet monitored and assessed it, then it is Not Assessed.

Go back to the top

        - impaired = If any part of the Assessment Unit fails to meet its water quality standards, it is calculated as impaired.
        - threatened = Threatened means that one or more Uses is Fully Supporting but experiencing a declining trend and likely to become impaired in the next reporting cycle. Waters that are Threatened are part of the Clean Water Act Section 303(d) list, unless a TMDL has been created for them. A null value is the same as isThreatened = 'N'.
        - 303d list = If the Assessment Unit is impaired by a pollutant and still needs to be addressed by a TMDL or other pollution control measure, it falls on the Clean Water Act (CWA) Section 303(d) List (which is also known as EPA IR Category 5). Note: This does not include all impaired waters. For example, Assessment Units that are impaired but already have a TMDL would fall into EPA IR Category 4a, and would not be on the CWA Section 303(d) list.
    - All spatial datasets were projected to Albert's equal area projection (crs = 5070) prior to running intersections.

Underground Storage Tanks

Maintained by: Environmental Protection Agency (EPA)'s Office of Underground Storage Tanks
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Static
- Description:
    - Point data containing the locations of facilities with active underground storage tanks. This data is from the UST Finder, which "is a flexible web application containing a comprehensive, national map of underground storage tank (UST) and UST release data. UST Finder provides the attributes and locations of active and closed USTs, UST facilities, and UST releases in states, tribal lands, and U.S. territories. The U. S. Environmental Protection Agency (EPA) developed UST Finder in partnership with the Association of State and Territorial Solid Waste Management Officials (ASTSWMO)." This information collected from September 2018 through December 2019 and represents a snapshot of UST facilities at time of data collection.
    - Information on UST releases are not summarized, as these events are typically resolved in 1-2 years and the dataset was deemed too old to include in the application.
- Caveats:
    - Please note that if a water system is pulling from a watershed that contains a lot of open underground storage tanks, that does not necessarily mean the drinking water quality is at risk or of poor quality. Water systems employ a variety of

Go back to the top

treatment techniques to filter contaminants from the source water. This variable is aimed to communicate potential hazards, rather than an actual risk to drinking water.
    - Please note this dataset is static, and only contains a snapshot of information collected from Sept 2018 - December 2019.
    - Because the exact location of underground storage tanks are not provided, it is assumed that facilities with underground storage tanks are close in proximity to the storage tanks they manage.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 8 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains data from NPDES, underground storage tanks, facilities with risk management plans, and impaired waterways. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
    - Identifying what water systems are pulling from a watershed that contains a certain amount of open underground storage tanks.
- Methods:
    - Point data was downloaded using ArcGIS REST services and data were filtered to only contain records whose facility status is active. Point data were sorted into HUC12s (points without geometry or those outside of the U.S. were removed, and we looped through each state, grabbed all of their HUC12 using the get_huc function, and made the CRS the same as the UST sites before running an st_intersection with spherical geometry turned off), grouped by HUC12s, and the total number of open underground storage tanks was summed using the open_us_ts field and temporarily out of service USTs, as these are considered "open" by EPA standards.
    - All spatial datasets were projected to Albert's equal area projection (crs = 5070) prior to running intersections.

Facilities with Risk Management Plans

Maintained by: Environmental Protection Agency (EPA)
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Quarterly
    - Continuous data collection started on: N/A (memory issue, working on it)
- Description:

Go back to the top

        - Facilities with active risk management plans. "EPA's Facility Registry Service (FRS) subset of facilities that link to the Risk Management Plan (RMP) database. This web feature service contains location and facility identification information from EPA's Facility Registry Service (FRS) for the subset of facilities that link to the Risk Management Plan (RMP) System. The Risk Management Plan (RMP) database stores the risk management plans reported by companies that handle, manufacture, use, or store certain flammable or toxic substances, as required under section 112(r) of the Clean Air Act (CAA)."
- Caveats:
    - Please note that if a water system is pulling from a watershed that contains a lot of facilities with risk management plans, that does not necessarily mean the drinking water quality is at risk or of poor quality. Water systems employ a variety of treatment techniques to filter contaminants from the source water. This variable is aimed to communicate potential hazards, rather than an actual risk to drinking water.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 8 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains data from NPDES, underground storage tanks, facilities with risk management plans, and impaired waterways. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
    - Identifying what water systems are pulling from a watershed that contains a certain amount of facilities with risk management plans.
- Methods:
    - Point data was downloaded using ArcGIS REST services and data were filtered to only contain records whose facility status is "active". Point data were sorted into HUC12s using the nhdtoolsPlus R package (specifically, we looped through each state, grabbed all of their HUC12 using the get_huc function, and made the CRS the same as the RMP sites before running an st_intersection with spherical geometry turned off), grouped by HUC12s, and the total number of facilities with Risk Management Plans were summed using the number of unique registry ids within each huc12.
    - All spatial datasets were projected to Albert's equal area projection (crs = 5070) prior to running intersections.

Oregon Drinking Water Advisories

Maintained by: Oregon Public Health Authority - Drinking Water Data Online
Related to service area boundaries by: EPIC

Go back to the top

Website link

- Internal update frequency:
  - Quarterly
  - Continuous data collection started on: Oct 2nd, 2025
- Description:
  - Publicly available drinking water advisories for water systems located in Oregon. Provided by the Oregon Public Health Authority.
- Caveats:
  - Continuous data collection started on Oct 2nd, 2025, and updates are completed quarterly. Please refer to the source link for the most current information.
  - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
  - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
  - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
  - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
  - Screening for water systems that have had disruptions in service.
- Methods:
  - Oregon boil water notice data were scraped from the webpage in R. Water system IDs were fixed by appending "OR41" to the PWS column, and boil water notices that appeared prior to 2017 were removed. This is because "Water advisory tracking began in May, 2017; most advisories from before this time will not be displayed here." Data are updated by pulling the latest webpage data in R, and checking that data against data from the last pull in S3. Records are updated (i.e., if they have closed since the last data pull), and new records are added. Data were filtered to only records for CWS.

West Virginia Drinking Water Advisories

Maintained by: WV Department of Health and Human Services
Related to service area boundaries by: EPIC

Go back to the top

Website link

- Internal update frequency:
  - Quarterly
  - Continuous data collection started on: Oct 2nd, 2025
- Description:
  - Publicly available drinking water advisories for water systems located in West Virginia. Provided by the West Virginia Department of Health and Human Resources.
- Caveats:
  - Please note that a water system in this dataset may report multiple & potentially separate advisories on the same day. For example, a system may report multiple advisories for each area that is affected by the same main break. As a result, the number of advisories for this state may appear very large in comparison to other states.
  - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
  - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
  - Continuous data collection started on Oct 2nd, 2025, and updates are completed quarterly. Please refer to the source link for the most current information.
  - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
- EPIC's data score: 76%
  - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
  - Screening for water systems that have had disruptions in service.
- Methods:
  - The webpage was inspected and the API endpoint containing these data were found. Data were read into R and filtered to only contain data for community water systems. Data are updated by pulling the latest API data in R, and checking that data against data from the last pull in S3. Records are updated (i.e., if they have closed since the last data pull), and new records are added. Because our method creates a unique ID using the date the advisory was issued to detect dataset updates, records that have been edited may appear as a separate advisory.

Go back to the top

New Mexico Drinking Water Advisories

Maintained by: New Mexico Environment Department
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
  - Quarterly
  - Continuous data collection started on: Oct 2nd, 2025
- Description:
  - Publicly available drinking water advisories for water systems located in New Mexico. Provided by the New Mexico Environment Department.
- Caveats:
  - Please note that if an advisory listed multiple water system IDs, the record was duplicated for each water system.
  - Continuous data collection started on Oct 2nd, 2025, and updates are completed quarterly. This is not a comprehensive record, as water system IDs are not always provided. Please refer to the source link for the most current information.
  - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
  - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
  - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
  - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
  - Screening for water systems that have had disruptions in service.
- Methods:
  - Data were scraped from this website. Issued and advisory lifted dates were converted into standard format, and notices without a water system ID were partially matched using the water system names from the EPA's SABs datasets. Due to incomplete PWSIDs, the data were not filtered to CWS. If a single notice was related to multiple water systems, the advisory row was duplicated with the

Go back to the top

separate PWSIDs. Data are updated by pulling the latest webpage data in R, and checking that data against data from the last pull in S3. Records are updated (i.e., if they have closed since the last data pull), and new records are added.

Ohio Drinking Water Advisories

Maintained by: Ohio Environmental Protection Agency
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
  - Static, this website has been retired by the state, last update was 7/30/2025
- Description:
  - Publicly available drinking water advisories for water systems located in Ohio. Previously provided by the Ohio Environmental Protection Agency.
- Caveats:
  - This website has been retired by the state and therefore this dataset is no longer continuously updated.
  - The last update of this dataset occurred on July 30th, 2025. This dataset is no longer maintained by the state of Ohio, and is not a comprehensive record. Please refer to your water system for the most current information.
  - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
  - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
  - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
  - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
  - Screening for water systems that have had disruptions in service.
- Methods:
  - Open advisories were collected from this website. Data were pulled using ArcGIS REST services into R, and dates were formatted appropriately. Data were filtered to only contain data from CWS. For updates, the latest list of active advisories were

Go back to the top

pulled and a unique ID was created to investigate differences between the active list and previous record. Data were split into: new advisories, still active advisories, and advisories that are assumed closed (no longer on active list), and binded together.

Rhode Island Drinking Water Advisories

Maintained by: Rhode Island Department of Health
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
  - Static, due to changes in data as a record moves from open to lifted, this dataset is more complicated to update automatically over time. This is on our radar for the next version of the tool. Last update was 7/30/2025
- Description:
  - Publicly available drinking water advisories for water systems located in Rhode Island. Provided by the Rhode Island Department of Health.
- Caveats:
  - The last update of this dataset occurred on July 30th, 2025. Some records may be missing the date the advisory was issued because that information is not provided for recently lifted advisories. Please refer to the source link for the most current information.
  - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
  - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
  - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
  - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
  - Screening for water systems that have had disruptions in service.
- Methods:

Go back to the top

- After inspecting this website, we found this spreadsheet that the webpage is pulling from. Data from all four tabs were pulled, binded together (with a flag to note ongoing or lifted general or PFAS notices), and data were filtered to CWS. Do not drink orders were inferred where the column "water_systems_under_do_not_drink" was not NA, and boil water notices where the column water_systems_under_boil_water_order was not NA. Otherwise, the type of notice or advisory is NA.

Washington Drinking Water Advisories

Maintained by: Washington Department of Health
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
  - Daily
  - Continuous data collection started on: Sept 12th, 2025
- Description:
  - Publicly available drinking water advisories for water systems located in Washington. Provided by the Washington Department of Health.
- Caveats:
  - Please note this is not a comprehensive record, as this state does not report PWSIDs and is therefore difficult to match with other datasets.
  - Continuous data collection started on Sept 12th, 2025, and updates are completed quarterly. This is not a comprehensive record, as water system IDs are not always provided. Please refer to the source link for the most current information.
  - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
  - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
  - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
  - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:

Go back to the top

    - Screening for water systems that have had disruptions in service.
- Methods:
    - Active water alerts were scraped from this website in R. Data were merged by water system name to match water system IDs (only 6 were successful). Data could not be filtered to only contain data from CWS because the PWSID column was incomplete. For updates, the latest list of active advisories were pulled and a unique ID was created to investigate differences between the active list and previous record. Data were split into: new advisories, still active advisories, and advisories that are assumed closed (no longer on active list), and binded together.

Missouri Drinking Water Advisories

Maintained by: State of Missouri Data Portal
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Daily
    - Continuous data collection started on: Sept 25th, 2025
- Description:
    - Publicly available drinking water advisories for water systems located in Missouri. Provided by the State of Missouri Data Portal.
- Caveats:
    - Continuous data collection started on Sept 25th, 2025, and updates are completed quarterly. This is not a comprehensive record, as this dataset only contains advisories from long-term contaminant issues. Please refer to the source link for the most current information.
    - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
    - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
    - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.

Go back to the top

- Recommended uses:
    - Screening for water systems that have had disruptions in service.
- Methods:
    - Active water advisories were downloaded using the API from this website in R. Data were filtered to only contain data from CWS. As of May 21st 2025, we have 0 records for CWS. For updates, the latest list of active advisories were pulled and a unique ID was created to investigate differences between the active list and previous record. Data were split into: new advisories, still active advisories, and advisories that are assumed closed (no longer on active list), and binded together.

Maine Drinking Water Advisories

Maintained by: Maine Department of Health and Human Services
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Daily
    - Continuous data collection started on: Sept 12th, 2025
- Description:
    - Publicly available drinking water advisories for water systems located in Maine. Provided by the Maine Department of Health and Human Services.
- Caveats:
    - Continuous data collection started on Sept 12th, 2025, and updates are completed quarterly. This is not a comprehensive record, as water system IDs are not always provided. Please refer to the source link for the most current information.
    - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
    - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
    - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:

Go back to the top

    - Screening for water systems that have had disruptions in service.
- Methods:
    - Active boil water notices and do not drink orders were downloaded from this website in R. Data were filtered to contain data from CWS, as of May 21st 2025, we have 1 record for a CWS. There were quite a few systems noted at CWS on the state page that did not match to a service area boundary, and were assumed to be errors. For updates, the latest list of active advisories were pulled and a unique ID was created to investigate differences between the active list and previous record. Data were split into: new advisories, still active advisories, and advisories that are assumed closed (no longer on active list), and binded together.

Alaska Drinking Water Advisories

Maintained by: AK Department of Environmental Conservation
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Daily
    - Continuous data collection started on: Sept 10th, 2025
- Description:
    - Publicly available drinking water advisories for water systems located in Alaska. Provided by the AK Department of Environmental Conservation.
- Caveats:
    - Continuous data collection started on Sept 10th, 2025, and updates are completed quarterly. Please refer to the source link for the most current information.
    - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
    - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
    - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:

Go back to the top

    - Screening for water systems that have had disruptions in service.
- Methods:
    - Active boil water notices and do not drink orders were downloaded from this web application in R. Data were pulled using ArcGIS REST services into R, and dates were formatted appropriately. Data were filtered to contain data from CWS. For updates, the latest list of active advisories were pulled and a unique ID was created to investigate differences between the active list and previous record. Data were split into: new advisories, still active advisories, and advisories that are assumed closed (no longer on active list), and binded together.

Arkansas Drinking Water Advisories

Maintained by: AR Department of Health
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Daily
    - Continuous data collection started on Feb 12th, 2026
- Description:
    - Publicly available drinking water advisories for water systems located in Arkansas. Provided by the AR Department of Health.
- Caveats:
    - Please note there was a gap in our data collection from May 21st, 2025 to Feb 3rd, 2026, as there was a temporary block on web scraping.
    - The last update of this dataset occurred on Feb 3rd, 2026. Please refer to the source link for the most current information.
    - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
    - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
    - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.

Go back to the top

- Recommended uses:
    - Screening for water systems that have had disruptions in service.
- Methods:
    - Active boil water orders were scraped from this website in R. Data were merged with water system information from SDWIS to match system names to PWSID, but since PWSID was not supplied, data were not filtered to CWS. However, as of May 21st 2025, all water system names matched to a PWSID in the SABs dataset. Data are updated by pulling the latest webpage data in R, and checking that data against data from the last pull in S3. Records are updated (i.e., if they have closed since the last data pull), and new records are added. Data were assumed to be boil water orders, based on the title of the table.

Florida Drinking Water Advisories

Maintained by: Florida Health
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Quarterly
    - Continuous data collection started on: Oct 2nd, 2025
- Description:
    - Publicly available drinking water advisories for water systems located in Florida. Provided by Florida Health.
- Caveats:
    - Continuous data collection started on Oct 2nd, 2025, and updates are completed quarterly. This is not a comprehensive record - Florida only lists advisories declared during a natural disaster, such as a hurricane or tropical weather event. Please refer to the source link for the most current information.
    - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
    - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:

Go back to the top

        - Screening for water systems that have had disruptions in service due to a large storm event.
- Methods:
    - Boil water advisories were scraped from this website in R. Data were merged with water system information from SDWIS to match system names to PWSID, but since PWSID was not supplied, data were not filtered to CWS. As of May 21st 2025, there is only 1 record for a system, and we don't know if this system is a CWS. Data are updated by pulling the latest webpage data in R, and checking that data against data from the last pull in S3. Records are updated (i.e., if they have closed since the last data pull), and new records are added.

Massachusetts Drinking Water Advisories

Maintained by: Massachusetts Department of Environmental Protection
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Currently static, since building a web driver on an AWS instance is a bit tricky.
- Description:
    - Publicly available drinking water advisories for water systems located in Massachusetts. Provided by Massachusetts Department of Environmental Protection.
- Caveats:
    - The last update of this dataset occurred on Oct 30th, 2025. Please refer to the source link for the most current information.
    - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
    - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
    - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:

Go back to the top

    - Screening for water systems that have had disruptions in service.
- Methods:
    - Active and terminated boil water advisories and do not drink orders were scraped from this website in R. Since these data are not downloadable, easily scrape-able, and there is no public API endpoint, data were scraped using RSelenium webdrivers. This code presses the "search" button to gather the full list of active and terminated advisories, and loops through the contents of each page to grab the "pop-up" that denotes the PWSID for the water system. The code then loops through each page (max is set to 200 pages, and on May 27th there are only 85) until a final page is not detected. Since the returned PWSID does not contain the state code (i.e., "4350001" instead of "MA4350001"), "MA" was appended to the start of each ID. On May 27th, 2025, there doesn't appear to be any Tribal CWS (which often have the EPA region code appended to the start of the PWSID), but we should check for pwsid longer than 9 characters before doing joins, in the case that a Tribal CWS was under a boil water notice. Because of this, data are not filtered to CWS.

Louisiana Drinking Water Advisories

Maintained by: Louisiana Department of Health
Related to service area boundaries by: EPIC
Website link

- Internal update frequency:
    - Quarterly, requires manual update
- Description:
    - Publicly available drinking water advisories for water systems located in Louisiana. Provided by Louisiana Department of Health. Last update was July 9th 2025.
- Caveats:
    - The last update of this dataset was completed on July 9th, 2025. Due to data download limits, this dataset includes system-issued boil water notices from 2024-2025, and state-issued boil water advisories from 2015-2025. This is not a comprehensive record. Please refer to the source link for the most current information.
    - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
    - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
    - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%

Go back to the top

        - To create this score, we checked for the data completeness (how many “NA”s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
    - Screening for water systems that have had disruptions in service.
- Methods:
    - The webpage was developed using Angular, which means it's super dynamic and there's a lot of javascript that makes scraping these data using web drivers really complicated. Names, IDs, and classes are duplicated across various HTML elements, which makes it difficult for the webdriver to identify what element I'm attempting to click. I tested various methods using RSelenium, and after connecting to the webpage, was getting a blank page with very basic html code and no actual elements to interact with. After some more digging, I think this is because the webpage is protected by reCAPTCHA, which is blocking non-human traffic from accessing the website content.
    - I did locate a very basic API endpoint from looking at network traffic after navigating to violations/boil water notices, but was getting a status code of 500 after attempting to access it using a GET() request in R. So, this data need to be exported manually.
    - Do to this, navigate to the drinking water watch page (https://sdw.ldh.la.gov/), click on "violations" go to the violation type, and select "BN - STATE ISSUED BOIL NOTICE" and select a 5yr window (mine was 7/9/2020 - 7/9/2025). Click on the "water system type" and select "community", click "search" and export the data to excel. Do the same for "BA - SYSTEM ISSUED BOIL ADVISORY" but filter for a 1year window (mine was 7/9/2024 - 7/9/2025) because the "export all" function can only email 5,000 records and selecting anything beyond one year often exceeds 5,000 records. Email the records using the "export all" function and check to make sure the records don't get capped to 5,000. Ours on 7/9 had 2,216 records.

Texas Drinking Water Advisories

Maintained by: Texas Commission on Environmental Quality (TCEQ)
Related to service area boundaries by: EPIC
Website link (N/A, FOIA’d)

- Internal update frequency:
    - Static, FOIA’d data from TCEQ. The last update of this dataset was completed on April 17th, 2024.
- Description:

Go back to the top

    - FOIA'd drinking water advisories for water systems located in Texas. Provided by the Texas Commission on Environmental Quality.
- Caveats:
    - Please note this dataset only contains records from 2018 - 2024.
    - The last update of this dataset was completed on April 17th, 2024. This is not a comprehensive record, and we are working to automatically update this dataset in the future. Please refer to the source link for the most current information, as it is provided on the state SDWIS website.
    - Please note some drinking water advisories may be planned maintenance! We hope to distinguish these from contaminant / infrastructure events in the future, if possible with the data available.
    - To account for differences in the scale of reporting between states, it is assumed a water system would report maximum one advisory on a given day.
    - If states edit the date an advisory was reported or edit the date an advisory was lifted after it has been filled out, this is captured as new advisory. It is difficult to discern edited advisories from completely new advisories.
- EPIC's data score: 76%
    - To create this score, we checked for the data completeness (how many "NA"s are there?) and how many duplicated records exist (if applicable) across 10 different variables that were created after cleaning our dataset. Please note that this score checks our summarized dataset, which contains summarized data for all states where we have boil water notice and advisory data. Combined with our estimate for how well these variables reflect the larger data universe, this produced a mean quality score of 76%.
- Recommended uses:
    - Screening for water systems that have had disruptions in service.
- Methods:
    - Data were FOIA'd from TCEQ on April 17th, 2024. The issued and rescinded dates were converted to proper date formats, and an issued date in 2027 was removed and assumed to be an error. These dates were then converted to our standardized "date_issued" and "date_lifted" fields, because these are dates likely entered by the utility (there is a reported_date and achieved_date, which have similar/identical definitions to the fields used above, but we are assuming these are the dates the advisory was reported in SDWIS). There are four events where the date_issued is missing, and 10 events with a value >= 1 in the "reissued count" field. Since these events are folded up into the same "date_issued", this is considered a single boil water notice event for the utility (i.e., we didn't duplicate these events, and kept them as a single row). We then filtered for water systems with service area boundaries from the EPA.

Go back to the top

Go back to the top