

SWarren_Enviro_Data_Proj

Repository for Simon Warren's final project in Environmental Data Analytics, Duke University.

Summary

This project is being completed as part of a Duke University Environmental Data Analytics class. The goal of the project is to create a visual representation of states, cities, counties, other administrative units, and water authorities that have announced moratoria on water shutoffs and reconnections during the COVID-19 pandemic. Currently most of the information comes from Food and Water Watch's (FWW) reporting on which governments have announced moratoria, and "live" data is kept on a spreadsheet, which does not visually depict the situation. The spreadsheet was created in March, 2020. Making a live version of this map would allow users to tell which governments have and have not put a moratorium on shutoffs quickly, and get the message out more quickly. This spreadsheet will be combined with US Census TIGER Lines for various levels of administrative areas, as well as SDWIS files for those water authorities with boundaries that do not match city or county boundaries. The map also shows which regions have announced moratoria on shutoffs but not reconnections and which have announced both, and depicts patterns in the size of populations served by the various bodies.

Investigators

Simon Warren, Duke University, simon.b.warren@duke.edu

Keywords

Water shutoffs; Water Reconnections; COVID 19; Food and Water Watch; Moratorium

Database Information

Data was collected from a spreadsheet published and updated by Food and Water Watch. (<https://docs.google.com/spreadsheets/d/e/2PACX-1vSU9gYc6bZ73BJFK1IpxTSSq5IRA0FmUqLWRvVbOw2kVXUzGrZiA5roEDVAjAP510Uk56EmhnypEc5T/pubhtml>) Data in this spreadsheet includes size of service population, moratorium information, and reconnection policies. Data is updated every 5 minutes, and is informed by official local/state government statements and media coverage. The data was first accessed for this project on April 4, 2020, but is accessed every time the script is run. A capture of the data made from a web scraping will be saved as FWW_Raw.csv

More information about FWW's research and reporting, along with a preliminary map of moratoria locations, can be found here <https://www.foodandwaterwatch.org/news/stopping-water-shutoffs-locally-not-enough-we-need-national-ban-and-service-restoration-plan> and here <https://www.foodandwaterwatch.org/news/we-need-country-wide-moratorium-water-shutoffs-amid-coronavirus>

State, county, and metropolitan area boundaries, as well as American Indian, Alaska Native, and Native Hawaiian Areas are created using US Census TIGER Lines, loaded using the "tigris" package developed by Kyle Walker of Texas Christian University. As a result, they do not appear in the "Raw" subfolder of the repository, but do appear in the "Processed" folder.

Water authority boundaries are created using EPA's Safe Drinking Water Information System (SDWIS) from the Envirofacts Data Service API for WATER_SYSTEM (https://data.epa.gov/efservice/WATER_SYSTEM/EXCEL). This table contains the name of the system as well as the name of a city or county served, and a zipcode, for geographic referencing. This data was saved as SDWIS_Raw.csv Data was accessed on April 13, 2020.

Folder structure, file formats, and naming conventions

The repository contains three folders, one of which has subfolders:

1. Data– both raw and intermediate data needed for the analysis. Included are three subfolders:

*Raw- Raw data, downloaded directly. This includes EPA SDWIS Water System data as an .xls and .csv and a .csv version of the Food and Water Watch spreadsheet for static analysis created from webscraping.

*Processed- Intermediate data and final tables. Intermediate data includes refined spreadsheets with certain columns removed and any necessary joins, in .csv format.

*Metadata- Metadata, in both .rmd and .pdf formats

2. Code– scripts needed to produce the output, including webscraping of the FWW spreadsheet, in .rmd and .pdf format.

3. Output- finalized maps, in .pdf format, although ideally eventually as .html for use with Shiny.

Data files will be named with the acronym of their source (TIGER, SDWIS, FWW), followed by `_raw` or `_processed`. When combined, they will be named `Combined_processed`.

Metadata

FWW

Status: A description of the particular action taken. There are three possible entries– “Moratorium,” “Pre-existing Moratorium,” or “Does not do them.” Most actions are “Moratorium” on shutoffs. “Pre-existing Moratorium” are in places that typically suspend water shutoffs during winter months, and extended the suspension during the pandemic. “Does not do them” refers to cities that did not use water shutoffs prior to the pandemic. City: City, county, water district/authority, state, or tribe. State: State Service Population: Number of people served by entity taking the action Restorations: Whether the order includes restoring service to residents who had been disconnected before the order. Source: Source for information on the order.

SDWIS

Not all columns from SDWIS tables are relevant, and so will not be discussed here. Those that follow are relevant: `WATER_SYSTEM.PWS_NAME`: The name of the water system. This is what matches the FWW “City” column when discussing water authorities `WATER_SYSTEM.POPULATION_SERVED_COUNT`: Population served, used to verify the match between Water System and City. `WATER_SYSTEM.ZIP_CODE`: Will be used as the location of the point for locations that do not match directly to a county or city.

TIGER Lines

Places: Census-Designated Places, standing in for towns and cities States: States Counties: Counties Native Areas: From the Census Bureau: “Federally recognized American Indian reservations and off-reservation trust land areas, state-recognized American Indian reservations, and Hawaiian home lands (HHLs).”

Scripts and code

The `Shutoff_Moratorium_Maps.rmd` contains two parts. Exploratory scripts: Three polygon layers of states, counties, and towns, are created, and merged into one layer. A point layer of zip codes will also be created. The most recent FWW spreadsheet will be datascraped and turned into a dataframe, and saved as a .csv.

Analysis scripts: The FWW tabular data is joined to the spatial data based on FWW’s “City” column, and polygons are filled with a color aesthetic based on whether a moratorium has been declared as well as whether reconnections have been ordered (green for both, blue for moratorium-only). Filling is based on a hierarchy– if a statewide moratorium is ordered, the entire state will be filled regardless of which individual cities or counties have ordered moratoria. An alpha aesthetic displays the service population.

For water districts/authorities that do not correspond to a county or city, SDWIS PWS_NAMES are used to join to the FFW “City” column, and then joined to the layer of zipcodes. A size and alpha aesthetic will be used for said point to represent service population.

Quality assurance/quality control

Shutoff orders are progressing quickly, and so some data is likely to be misreported. At the present, FWW’s reporting will be assumed to be accurate, but this map should not be interpreted as a legal or governmental document.

Since water district boundaries are not always coterminus with city or county limits, and shapefiles are not always publicly available for water district boundaries, the exact location of those authorities may be inaccurate. Water authorities are represented as points connected to the city or county they serve, with the size of the point reflecting the population they serve.