Analysis of drinking water water contaminant occurrence in the Northeast and Southeast United States

https://github.com/rachel-gonsenhauser/Final_Project_Environmental_Data_Analytics

Rachel Gonsenhauser

Abstract

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

While the EPA establishes standards for 90 drinking water contaminants by means of the federal Safe Drinking Water Act (SDWA) and its regulations, public water systems still often struggle to remain in compliance with such policies (USEPA, 2020). This issue of compliance with the SDWA can stem from myriad causes, for instance financial capacity of the water system. This is especially concerning in areas where geologic conditions and/or anthropogenic activities frequently introduce contaminants into drinking water supplies. Additionally, some known contaminants still have yet to be regulated by the EPA, such as poly- and perfluoroaklyl substances (PFAS), which introduces even more complexity to the issue of water quality monitoring of drinking water sources.

This analysis seeks to investigate the co-occurrence of water quality indicators including arsenic, trihalomethane, uranium, and PFAS, which originate from both geogenic and anthropogenic sources. Additionally, given pervasive questions related to environmental justice and how socioeconomic factors may be related to water quality indicators, this analysis seeks to examine relationships between water quality indicators and county-level median household income (MHI) and size of the population served by a given community water system (CWS), which is often a proxy for how rural an area is and the financial capacity of a CWS. Additionally, questions regarding how contaminant occurrence differs across time and between states are explored.

To narrow the scope of this project, most analyses are targeted to southeastern region states and northeastern region states. These regions were chosen given their differences in geology and socioeconmic makeup. Additionally, individual case studies of Massachusetts and North Carolina are explored in further depth. As arsenic is present in much of the underlying geology in New England and other northeastern states, arsenic data is used is many of the analyses performed. Due to issues of PFAS data limitations, discussed in more detail in the subsequent section, analyses using PFAS data are limited. Specifically, the following questions are explored in this analysis: 1. Do arsenic concentrations vary significantly across time and from state to state in northeastern and southeastern states? 2. Do socioeconomic factors or the presence of other conntaminants predict arsenic concentrations in northeastern and southeastern states? Do these same relationships hold in the case studies explored (Massachusetts and North Carolina)? 3. Do MHI and the presence of drinking water contaminants predict the population served by a CWS in northeastern and southeastern states? Do these same relationships hold in the case studies explored (Massachusetts and North Carolina)? 4. Do PFAS concentrations vary signinficantly across time and from state to state in the United States? Are socioeconomic factors significant predictors of PFAS concentrations?

2 Dataset Information

Data used for this analysis was downloaded from the Centers for Disease Control and Prevention (CDC)'s National Environmental Public Health Tracking Network at Centers for Disease Control and Prevention (CDC)'s National Environmental Public Health Tracking Network https://ephtracking.cdc.gov/DataExplorer/#/. Output from this online tool containing geographic and CWS data associated with individual variables was combined into the final processed dataset used for this analysis.

Parameter	**Summary**
Number of states	28
Number of CWSs	25,583
Water quality indicators	Arsenic, trihalomethane, uranium, PFAS
Socioeconomic variables	Population served by CWS, MHI
Data collection time span	1999-2018

Column heading	**Variable description**	**Data range**
stateFIPS	Federal Information Processing Standard state code	N/A
State	state measurement was taken in	N/A
countyFIPS	Federal Information Processing Standard county code	N/A
County	county measurement was taken in	N/A
Year	year measurement was taken in	N/A
Arsenic_ugL	mean arsenic concentration (micrograms per liter)	1-2,422 micrograms/liter
PWS.ID	Public Water System Identification Number	N/A
CWS.Name	Community Water System Name	N/A
Population.Served	number of people served by CWS	0-8,271,000 people
MHI	median household income ($)	$16,435-$113,336
PFAS_ppt	PFAS conncentration (parts per trillion)	1-60 ppt
TTHM_ugl	mean trihalomethane concentration (micrograms per liter)	0-219.20 micrograms/liter
Uranium_ugL	mean uranium concentration (micrograms per liter)	0-379.10 micrograms/liter
MCL_TTM	whether MCL for trihalomethanes is exceeded	N/A
MCL_Uranium	whether MCL for uranium is exceeded	N/A
MCL_Arsenic	whether MCL for arsenic is exceeded	N/A

It should be noted that PFAS data was only available for 2013-2015. For ease of analysis, this date range was changed to 2014 during the raw dataset wrangling process to create a common annual unit of analysis for all variables.

3 Exploratory Analysis

4 Analysis

4.1 Question 1: <insert specific question here and add additional subsections for additional questions below, if needed>

4.2 Question 2:

5 Summary and Conclusions

6 References

United States Environmental Protection Agency (USEPA). 2020. Safe Drinking Water Act (SDWA). Retrieved from: https://www.epa.gov/sdwa.