DISPARITIES IN COMMUNITY WATER SYSTEMS' COMPLIANCE WITH THE SAFE DRINKING WATER ACT

A Thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
the requirements for
the Degree

Master of Science

In

Geographic Information Science

by

Zoe Larissa Statman-Weil

San Francisco, California

January 2019

CERTIFICATION OF APPROVAL

I certify that I have read Disparities in Community Water Systems' Compliance with the Safe Drinking Water Act by Zoe Larissa Statman-Weil, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree Master of Science in Geographic Information Science at San Francisco State University.

Leora Nanus, Ph.D.
Assistant Professor

Nancy Wilkinson, Ph.D.
Professor

DISPARITIES IN COMMUNITY WATER SYSTEMS' COMPLIANCE WITH THE SAFE DRINKING WATER ACT

Zoe Larissa Statman-Weil
San Francisco, California
2019

Understanding potential disparities in communities' compliance with the Safe Drinking Water Act (SDWA) will help managers effectively and fairly allocate funding for improving drinking water systems. This study investigated the relationship between socioeconomic status and race, and violations to the SDWA by community water system (CWS) in the state of Pennsylvania. The sociodemographic characteristics of the water systems were estimated using three different CWS-level spatial analysis methods (areal weighting, dasymetric mapping, areal interpolation) and a county-level spatial analysis method. Negative binomial regression was then applied to determine if these sociodemographic characteristics, and other water system variables, are associated with the number of total and/or health-based SDWA violations. Conclusive evidence of environmental injustice in SDWA violations by race or class was not found; however, this study did determine that small, rural CWSs are likely to have more violations than other CWSs. This research demonstrates that the spatial analysis method selected for an environmental justice study can affect the results and conclusions of the research.

I certify that the Abstract is a correct representation of the content of this thesis.

Lian Narius
Chair, Thesis Committee

12/06/18
Date

ACKNOWLEDGEMENTS

I would like to thank Dr. Leora Nanus for meeting with me regularly for over a year, helping me think through my project, and encouraging me to produce my best research. She made the challenge of a thesis very enjoyable and rewarding and for that I am truly grateful. I would also like to thank Dr. Nancy Wilkinson for listening to my first thesis idea and encouraging me to work with her and Dr. Nanus. Her consistent positive support has given me great confidence in myself as a researcher. I am very lucky to have worked with such great advisors who asked challenging questions and provided thoughtful feedback. Lastly, I have deep gratitude for my partner, Jake Hanft, who fed me many meals and listened to countless hours of me talking about my classes and thesis as I worked through ideas. I could not have done it without his support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figures Page

1. Introduction

While there has been limited research on the relationship between access to safe drinking water and the sociodemographic and socioeconomic makeup of communities, interest in drinking water equity in the United States has increased due to the recent and ongoing crisis in Flint, Michigan. In this low-income, majority-black city, lead levels well above the Safe Drinking Water Act (SDWA) legal threshold were found. The heightened public attention provides researchers and advocates the opportunity to explore and discuss access to safe drinking water in the context of environmental justice (Greenberg, 2016; Olson & Fedinick, 2016). This crisis highlights the need to conduct environmental justice analyses of public utility water systems' compliance with the SDWA. As explored in more detail below, compliance with the SDWA does not equate safe drinking water quality. However, compliance does *suggest* safer drinking water, and SDWA data is the best nationally available public data to use to predict the drinking water quality of public water systems.

A. Safe Drinking Water Act

The primary federal law that protects drinking water in the United States is the SDWA, enacted in 1974, and amended and reauthorized in 1986 and 1996 (U.S. EPA, 2017a). Under the SDWA, the United States Environmental Protection Agency (EPA) is required to identify and develop rules related to harmful contaminants in drinking water

Public water systems are defined by the EPA as serving at least 25 people or having at least 15 service connections for at least 60 days a year. There are three types of public water systems:

- **Community water systems** serve the same population year-round (e.g. homes).
- **Non-transient non-community water systems** serve the same population at least six months a year, but not the entire year (e.g. schools, churches, factories)
- **Transient non-community water systems** serve a changing population (e.g., campgrounds, gas stations)

U.S. EPA 2017a

distributed by public water systems (PWSs) (see the callout box for the definition of a PWS and the three subset water system types) (U.S. EPA, 2017a).[1] These rules establish maximum contaminant levels (MCLs) or treatment protocols for around 100 contaminants (Fedinick, Wu, Pandithartne, & Olson, 2017). The initial law was passed with the intention of providing communities with safe drinking water through treatment, but the 1996 amendments expanded the scope of the law and included protections for source water, operator training, infrastructure funding, and public education and information. (U.S. EPA, 2004).

Although the SDWA was enacted to protect drinking water in the United States, it does not guarantee Americans clean water (Balazs, 2011). The EPA has only decided to regulate certain contaminants, including nitrate, arsenic, and lead, for which it has determined a maximum contaminant level goal (MCLG) where there would be no expected health risk over a lifetime of exposure on a daily basis (U.S. EPA, 2004, 2017b). The EPA then sets the MCL, the determined enforceable limit of the contaminant allowed to be distributed by PWSs, as close to the MCLG as possible, while taking cost of treatment into consideration (U.S. EPA, 2004, 2017b). If there is no reliable method of detecting contaminants, or the EPA does not consider it economically or technically feasible to set an MCL, a required treatment technique is then used by the PWS to remove the contaminants from the drinking water (U.S. EPA, 2004, 2017b). The cost-benefit analysis that factors into the MCL and treatment requirements is considered controversial (Cory & Rahman, 2009).

The EPA is also slow to regulate new contaminants. It is required to reassess its regulations of all contaminants under the SDWA every six years; however, it has not established a regulatory standard for a new contaminant in over 20 years (Environmental

[1] Although this project looks at community water systems in its analysis, PWSs are referenced throughout the paper in situations where the cited source or study references PWSs. PWS is the umbrella term of which a community water system is a subset.

Working Group, 2017; Fedinick et al., 2017). The result of this hiatus in regulatory action is over 250 contaminants, including emerging contaminants, that are detected in drinking water in the United States at concentrations greater than scientists have determined healthy but that remain within the current legal limits defined by the SDWA or are unregulated (Environmental Working Group, 2017).

If a PWS does not comply with an existing SDWA standard, it is cited for a violation. Not all SDWA violations indicate a contaminant was found above its MCL; a violation could mean an administrative SDWA rule was violated (U.S. EPA, 2017c). Health-based violations consist of exceedances of MCLs, exceedances of disinfectant concentrations thresholds, and improper water treatment. Non-health-based violations include failure to monitor regularly or report results on-time, failure to notify the public per the requirements of the SDWA (e.g., if there is a serious health problem with drinking water), and failure to publish annual consumer confidence reports (U.S. EPA, 2017c). Even seemingly non-health-based violations, such as failure to monitor and report, can indicate an underlying health-based issue (Fedinick et al., 2017). Previous work showed that fewer MCL health violations in smaller privately owned PWSs, but more non-health-based violations, may indicate that fewer MCL violations are a result of insufficient monitoring and reporting (Allaire, Wu, & Lall, 2018; Wallsten & Kosec, 2008). Thus, it is important to look at disparities across all SDWA violations.

Although the EPA sets the national MCLs and treatment standards, most states have applied for "primacy," which means the state oversees SDWA implementation, ensuring compliance, conducting inspections, offering trainings, and enforcing the standards at the state's PWSs (U.S. EPA, 2004). If PWSs are not meeting the requirements of the SDWA, both the EPA and states can take legal action (U.S. EPA, 2004). Enforcement actions by either states or the EPA are uncommon. The NRDC found in its analysis of SDWA data that enforcement actions were taken for only 13.1 percent of all SDWA violations in 2015 across the U.S. (Fedinick et al., 2017).

In its 2013 National Public Water Systems Compliance Report, the EPA conducted an audit of primacy agencies' water programs (U.S. EPA's Office of Enforcement and Compliance Assurance, 2013). States with primacy are required to input all violation and enforcement data in the EPA's Safe Drinking Water Information System/Federal Version (SDWIS) database (U.S. EPA's Office of Enforcement and Compliance Assurance, 2013). The assessment of states' water programs, which included comparing the states' SDWA files and data with the data contained in the SDWIS database, led to the conclusion that the "violation data are substantially incomplete" (U.S. EPA's Office of Enforcement and Compliance Assurance, 2013). Incomplete data entry could mean that even monitoring and reporting violations are not entered into the SDWIS database. In sum, record of PWSs' violations are incomplete and enforcement of penalties is weak. However, the data stored in the EPA's SDWIS database is the most comprehensive dataset available for information about PWSs and their drinking water quality in the United States, and is commonly used in national assessments of the water systems in the U.S. (Allaire et al., 2018; Fedinick et al., 2017).

B. **Environmental Justice Analysis Related to Drinking Water Quality**

The EPA defines environmental justice as "the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income, with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies" (U.S. EPA, 2018). Environmental justice encompasses the concepts of both distributional equity, meaning environmental harms and benefits are distributed equally, and procedural justice, which is the fair enforcement of and compliance with environmental laws and policies (Balazs, 2011). Environmental justice studies typically investigate potential disparities based on race and/or socioeconomic status (SES).

Environmental justice research on inequalities in access to clean drinking water has increased since 2014 as a result of the drinking water crisis in Flint, Michigan that occurred after the city switched its water source to the Flint River. Due to the corrosive quality of this new water source, the lead levels in Flint's drinking water spiked, resulting in an increase in blood lead levels of the community's children (Campbell, Greenberg, Mankikar, & Ross, 2016). Several case studies have concluded that environmental injustice and environmental racism played a role in the slow response by authorities to address this crisis (Butler, Scammell, & Benson, 2016; Campbell et al., 2016). This has resulted in a national interest and discussion regarding fair access to clean and safe drinking water.

A handful of quantitative studies have examined the relationship between sociodemographic factors and SES of communities and drinking water quality, some as a direct response to the Flint crisis. These studies have primarily occurred in the western United States and have focused on exposure to contaminants and/or compliance with drinking water standards (Balazs, Morello-Frosch, Hubbard, & Ray, 2011, 2012; Cory & Rahman, 2009; The Environmental Justice Coalition for Water, 2005). Balazs (2011) coined the term "joint burden analysis" to refer to the research approach of investigating both aspects of drinking water quality. While exposure and compliance burdens are related, acknowledging they cause different hardships within a community helps elucidate the complexities of drinking water research.

Investigating exposure levels provides insight into the health risks of a community, but lack of resources to comply with regulations may render low-income communities unable to mitigate health risks. Communities with a large percentage of residents with a lower SES may have a harder time complying with EPA standards if they lack the technical, managerial and financial capacity (TMF) to decrease contaminants in drinking water, such as arsenic and nitrate, and/or to implement proper treatment. A lack of resources in non-compliant systems can shift the financial burden of compliance from the CWS to the

customers with a lower SES through increased water costs. If the cost of mitigation is too high, the community may oppose efforts to reduce exposure and/or rely heavily on bottled water, resulting in continued exposure and/or higher costs (Balazs & Ray, 2014; National Research Council, 1997). The lack of TMF and related consequences are even more marked in rural communities where there is a smaller customer base to generate revenue (National Research Council, 1997).

C. **Environmental Justice Quantitative Assessment**

This section provides an overview of quantitative findings of previous environmental justice studies analyzing compliance with drinking water regulations and exposure to contamination in drinking water. The spatial analysis methods and techniques of some of these previous studies are presented in *Section 1-D.*

The Environmental Justice Coalition (EJC) assessed disparities in access to safe drinking water at the county level in California between the years 1995 and 2000, and found that counties with the greatest number of drinking water violations[2] had the highest percentage of people of color, people living below the poverty line, and Latinos (all calculated separately) (The Environmental Justice Coalition for Water, 2005). The EJC's results were included in its report "Thirsty for Justice: A People's Blueprint for California Water" (The Environmental Justice Coalition for Water, 2005). Since the report did not explain the methods used to make this assessment, the validity of the analysis could not be assessed (The Environmental Justice Coalition for Water, 2005). EJC divided the counties into three categories (most violations, least violations, and others), but did not define the parameters for each county group. Simple bar graphs present the sociodemographic and socioeconomic characteristics of the counties with the most violations, least violations, and all others, but no statistical analysis appears to have

[2] The report does not state clearly that it is assessing violations to the SDWA, but it is likely that is what "violations" refers to in this report (The Environmental Justice Coalition for Water, 2005).

been conducted (The Environmental Justice Coalition for Water, 2005). EJC found the most prominent difference in violations correlated with the percentage of Latinos: counties with the most violations had 41% Latinos, counties with the least violations had 16% Latinos, and all other counties had 25% Latinos (The Environmental Justice Coalition for Water, 2005).

In another California study, Balazs et al. (2011) compared the average nitrate concentrations at points of entry into each community water system (CWS) in the San Joaquin Valley to the sociodemographic characteristics of the population potentially exposed. The researchers found that CWSs with less than 200 service connections had greater nitrate levels in communities with a higher Latino population (Balazs et al., 2011). No significant relationships between nitrate contamination and race/ethnicity or home ownership were found for larger CWSs (Balazs et al., 2011).

A few studies have calculated drinking water disparities related to arsenic contamination. Balazs et al. (2012) found that CWSs in California's San Joaquin Valley serving residents with higher home ownership rates, as an indication of economic security, had lower arsenic levels, and that this relationship was even stronger in the smaller systems. CWSs serving residents with higher home ownership rates received fewer arsenic MCL violations under the Safe Drinking Water Act (SDWA), and the inverse was true for CWSs serving more people of color (Balazs et al., 2012). Balazs et al. (2012) concluded that their findings showed a joint burden of both exposure and compliance for poorer communities of color.

Cory and Rahman (2009) also looked at the relationship of arsenic compliance to sociodemographic factors and SES of residents in Arizona. Arsenic concentrations were analyzed by zip code and it was determined that there were no social disparities in the enforcement of the revised arsenic standard, which the EPA was required to update under the 1996 SDWA amendments. The revised arsenic standard was effective February 22,

2002, and PWSs were required to comply by 2006, making PWSs' transition to compliance a unique research opportunity (Cory & Rahman, 2009). This Arizona study found no differences in compliance with the revised arsenic standard by race, income (per capita and per household) or average value of a house (Cory & Rahman, 2009).

Switzer and Teodoro (2017) investigated the impact of race, ethnicity, and SES on compliance with the SDWA on a national scale, with a focus on health-based regulations. They concluded that the racial and ethnic composition of the community served by the PWS "predicts drinking water quality," and that "black and Hispanic populations most strongly predict SDWA violations in low-SES communities" (Switzer & Teodoro, 2017). However, since the researchers were looking at drinking water system *compliance*, it is unfair to make a conclusion regarding the quality of the drinking water. Additionally, the authors do not provide an explanation of how the demographic characteristics of the communities served by the PWSs were estimated, which may be problematic as this method could influence the results. See *Section 1-D* for more details on estimating demographic characteristics within a given area.

Recent studies have analyzed national trends in drinking water violations. Allaire et al. (2018) analyzed spatial and temporal trends in SDWIS data by county and by CWS between 1982 and 2015 at the national level. They assigned each CWS the census information of the county in which its address listed in the SDWIS database was located. They found rural CWSs were more likely to have violations, and that CWSs serving low-income, communities of color were more likely to receive total coliform violations, specifically (Allaire et al., 2018).[3] As the research spanned decades, responses to new regulatory rules were also analyzed (Allaire et al., 2018). After a new rule related to disinfection byproducts was put into effect in the early 2000s, low-income rural counties were found to be much slower to comply than urban counties (Allaire et al., 2018). Their

[3] The paper notes that total coliform violations are more accurately reported than other violations (Allaire et al., 2018).

findings may indicate environmental justice concerns but more research is needed (Allaire et al., 2018).

In order to assess inequalities in compliance with the SDWA, a comparison between the number of violations of the SDWA and the characteristics of the population served by a CWS is needed, primarily the socioeconomic status (SES) and the percentage people of color. SES is represented by a number of different variables in environmental justice studies. Balazs et al. (2011, 2012) used home ownership rate, which has been accepted as a proxy for affluence and political power (Krieger, Williams, & Moss, 1997; Morello-Frosch, Pastor, & Sadd, 2001). Other environmental justice studies use mean income (both per capita and per household) or percent below poverty level (Cory & Rahman, 2009; Ogneva-Himmelberger & Huang, 2015; The Environmental Justice Coalition for Water, 2005).

D. **Environmental Justice Spatial Analysis**

Analyzing the sociodemographic characteristics and SES of populations potentially exposed to environmental harms is critical to environmental justice and health equity research. GIS makes quantitative environmental justice analysis feasible and accessible (Holifield, Chakraborty, & Walker, 2017). Typical geospatial environmental justice analysis identifies an environmental "bad" (e.g., a polluting facility) or an environmental "good" (e.g., park) and then evaluates the sociodemographic and socioeconomic characteristics of the community affected by the target feature, often with the use of a defined buffer of impact (Chakraborty, Maantay, & Brender, 2011). Since these discrete or continuous areas do not usually line up spatially with available geospatial population data, researchers face the task of identifying a method that will best estimate the characteristics of a population within the exposure area (Holifield et al., 2017). Commonly used methods in the literature include areal weighting, dasymetric mapping, and areal interpolation defined below:

- **Areal weighting.** This method assigns a proportion of the population to the affected area relative to the percent of the geographic unit within the discrete affected area boundary (e.g., a buffer) (Holifield et al., 2017). For example, if 10% of the population within a given geographic unit is below the poverty line and 40% of the unit is within a buffer representing exposure to an industrial facility, then 40% of the unit's total population would be considered within the area of exposure, and 10% of that population would be estimated to be below the poverty line. This method assumes equal distribution of the population within the census tracts.
- **Dasymetric Mapping.** This advanced analysis technique utilizes additional ancillary data, such as land use or zoning data, to better estimate the population distribution within a given geographic unit. For example, if a census tract includes a large city park, it can be assumed that no individuals reside in that area. This method may not be as used as commonly compared to the other simpler methods as it is fairly complex and requires a higher degree of knowledge of GIS (Maantay & Maroko, 2009).
- **Areal Interpolation.** This geostatistical kriging-based method creates a continuous prediction surface from polygon data that can be reaggregated to new polygons (Hallisey et al., 2017; Krivoruchko, Gribov, & Krause, 2011)

Previous research on drinking water disparities studies did not include complete datasets of the digital boundaries of the water systems they studied so these standard methods of estimating populations within a given buffer could not be applied. Cory and Rahman (2009) averaged the contaminant levels of arsenic for every PWS within a given zip code, and then compared that average concentration of arsenic to the demographic characteristics of the zip code. The researchers used zip codes as the geographical unit of

analysis because a significant percentage of PWSs in Arizona serve more than one census tract, and the larger zip code area would reduce the problem of area overlaps. However, this study does not address the possibility that PWSs could, and often do, serve more than one zip code (Cory & Rahman, 2009).

Balazs et al. (2011) compared two techniques of estimating the demographics of the population served by each CWS for their analysis of nitrate contamination in drinking water in San Joaquin Valley, California. The first method used areal weighting with the use of digitized CWS boundaries, the second included averaging the demographic characteristics of every census block that contained a CWS source (well field, surface water intake, and treatment plants). The second approach was found to be adequate and utilized for their analysis. Balazs et al. (2012) applied the same approach to their analysis of arsenic contamination in drinking water in the same area. The SDWIS database does not contain water system source locations so this method could not be applied in different states.

E. **Research Questions**

This thesis is an environmental justice analysis of water systems' compliance with the SDWA and focuses on the following research questions:

> 1) Are there social disparities in CWS compliance with the SDWA? Are there more violations (total and health-based) in low-income communities, communities with a higher proportion of people of color, and/or rural communities?
>
> 2) How do the results differ depending on the spatial analysis method used to estimate the demographic characteristics of the population served by the CWS?

The specific spatial analysis methods are discussed in more detail in *Section 2-C*. Three of these methods use GIS techniques to estimate the demographic characteristics of the

specific CWS's population (referred to as the "CWS-level analyses") and one assigns the CWS county's demographic characteristics to that CWS (referred to as the "county-level analysis"). Percent below poverty line will be used as the proxy variable for SES in this study. Several other covariates related to the characteristics of the CWS itself (e.g., private vs. public) will also be factored into the statistical model based on the literature. See *Section 2-D* for more information on what other variables were included in the model and why they were chosen.

2. Methods

A. Study Area

Very few states have publicly available PWS or CWS boundary data. Due to the availability and quality of the data available in Pennsylvania, it was selected as the study area for this thesis.

Pennsylvania has the sixth largest population in the United States (12.78 million people) and ranks ninth in population density (286 people/square mile) (Cedar Lake Ventures Inc., 2018). The two largest cities by population are Philadelphia (1.6 million people), located in the southeast corner, and Pittsburgh (305 thousand people), located in the southwest corner of the state **(Figure 1)** (Cedar Lake Ventures Inc., 2018). The Allegheny mountains extend southwest to northeast across the state between these two cities **(Figure 1)**. Marcellus Shale, a rock formation from which natural gas can be extracted with the use of hydraulic fracturing technology (or "fracking"), underlies approximately 64% of Pennsylvania, excepting the southeast corner (Clough & Bell, 2016; Ogneva-Himmelberger & Huang, 2015). Fracking wells, referred to by the Pennsylvania Department of Environmental Protection (PA DEP) as unconventional wells, are scattered across the Marcellus Shale, with the highest density south of Pittsburgh and in the northeast corner of the state (Amico, DeBelius, Detrow, & Stiles, 2015).

Pennsylvania on average has a higher percentage white population than the rest of the United States, with 77.7% non-Hispanic white, approximately 15% higher than the United States as a whole. The largest minority populations are black (11.0%) and Hispanic (6.1%), both smaller than their respective proportions in the United States as a whole. Rural Pennsylvania is heavily white non-Hispanic (Cedar Lake Ventures Inc., 2018). The mean income in Pennsylvania is similar to the mean in the United States, at $54.9 thousand annually (Cedar Lake Ventures Inc., 2018). The percentage of households on foods stamps in Pennsylvania (13.0%) ranks it 26th in the United States in this poverty metric (Cedar Lake Ventures Inc., 2018).

Figure 1: Map of study area, state of Pennsylvania.

B. **GIS Data**

The data used for the spatial analysis in this study is presented in this section. The source of each dataset is identified below, followed by a description of any data processing that was conducted, including the criteria used for selection or categorization.

The PA DEP has produced a near-complete GIS shapefile of the active state CWS boundaries (http://data-padep-1.opendata.arcgis.com/datasets; accessed 8/20/2017).[4] The dataset contains 1,853 CWS boundaries, which according to the metadata represent over 90% of the states' CWS boundaries, a higher percentage than most, if not all, other states. Review of the dataset determined that 20 of the water systems were either not CWSs or were inactive. These were removed and 1,833 active CWS boundaries remained. The metadata for this shapefile describes the accuracy as follows:

> *"Since the boundaries polygons are based on maps submitted by each PWS, a method to eliminate the overlapping service areas and boundary gaps between neighboring service areas was needed. When overlaps were identified a map was created and sent back to the PWS to correct the overlapping areas. In cases when no verification was available the overlaps were merged with a service area based on the best available information"* (Pennsylvania Department of Environmental Protection, 2017).

This PWS data was used to identify the area served by each CWS. As discussed in more detail under *Section 2-C*, several different spatial analysis techniques were evaluated to estimate the demographics of the community served by each water system.

SDWA violation data was queried and downloaded from the online SDWIS database (https://ofmpub.epa.gov/apex/sfdw/f?p=108:200; accessed 10/15/2017). The time period ranges from January 1, 2012 to December 31, 2016, a five-year time period was selected to provide more violation data to analyze, to better assess the operation of the CWSs.

---

[4] Although the data is titled "Public Water Supplier's (PWS) Service Areas," the metadata states that all non-transient noncommunity water systems and transient noncommunity water systems are excluded (Pennsylvania Department of Environmental Protection, 2017), leaving only community water systems.

Following the methods of Allaire et al. (2018), the time period in which a violation occurred was defined by the compliance period begin date. It is assumed that a violation did not occur within the time period of interest if a CWS did not have a violation entry in the SDWIS database for the years selected. Since the PA DEP geospatial public water supply boundary data only includes active CWS data, "CWS" was selected as the PWS Type and "Active" as the activity status. To summarize, the SDWIS database was queried as follows:

- PWS Activity Status = "Active"
- PWS Type = "Community Water System"
- Compliance Period Begin Date < 1/1/2017
- Compliance Period Begin Date >= 1/1/2012

These data were then processed to produce a dataset containing the PWS ID, number of total violations within the time period of interest, and the number of health-based violations within the time period of interest. The data were then joined to the PA DEP CWS boundary shapefile, and all CWSs contained in the shapefile that did not have any corresponding violation data in the SDWIS spreadsheet were identified as having zero violations, health-based or otherwise.

Populations served by the CWSs were estimated based on census tract-level data using a 2017 geodatabase containing census tract and county boundaries from the U.S. Census Bureau (https://www.census.gov/geo/maps-data/data/tiger-geodatabases.html; accessed 11/10/2017). Census tracts were selected as the unit of analysis instead of the census block group, a smaller geographical unit, as data associated with census tracts reflect a larger number of survey respondents and a smaller margin of error, and are thus considered a more reliable estimate of the population (Ogneva-Himmelberger & Huang, 2015).

Population and demographic characteristics of the Pennsylvania census tracts and counties from the U.S. Census Bureau were also included in the analysis (https://factfinder.census.gov/faces/nav/jsf/pages/download_center.xhtml; accessed 11/10/2017). Specifically, total tract population, percent population below the poverty level and percent people of color (including the Hispanic population) were compiled from the 2011-2015 American Community Survey (ACS) 5-Year Estimates. The census tract data contained 3,218 tracts, and the county data contained 67 counties (https://www.census.gov/geo/maps-data/data/tiger-geodatabases.html). Processing of the data included removing all census tracts missing data for the total census tract population or the demographic data, which resulted in a dataset containing 3,202 census tracts. A threshold of >0.24 km^2 was used for census tract size to allow for application of the areal interpolation method, resulting in 3,166 census tracts.

Delineated urban areas in Pennsylvania were also obtained from the U.S. Census Bureau (https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2016&layergroup=Urban+Areas; accessed 2/20/2018). For the CWS-level analysis, a CWS was classified as urban when 50% or more of the area of the CWS was located within an urban area, and otherwise it was classified as rural. Classifying by the rural and urban population count within each CWS was not an option as there is not an accurate way of determining what percentage of the rural or urban population within a census tract lives within the CWS. The percentage of the CWS within an urban area was determined to be the best available method to estimate whether the water system was primarily urban or rural. 754 of the CWSs were classified as urban and 1,079 were classified as rural.

Alternatively, the counties were classified as rural or urban based on the percentage of the population living within a rural or urban area in a given county according the U.S. Census. The classification method was different for the county-level analysis compared to the CWS-level analysis because the county rural/urban demographics are readily

available from the U.S. Census Bureau (https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_SF1_P2&prodType=table; accessed 2/20/2018). If 50% or more of the county's population was defined as urban according to the U.S Census data, then the county was classified as urban. Of the 67 counties, 37 were classified as urban and 30 were classified as rural.

Land cover data obtained from the U.S. Geological Survey helped estimate the population of the community served within each water system. The 2011 National Land Cover Database (NLCD) (https://www.mrlc.gov/nlcd11_leg.php; accessed 2/11/2018) was clipped to the area surrounding the Pennsylvania state boundary. A shapefile identifying residential land was created by selecting the land classified in the NLCD as developed at low, medium, and high intensity, which consists of areas with 20% to 100% impervious surface land cover according to the NLCD metadata. These three NLCD classes were selected as they are defined as residential in the NLCD; they include single-family housing units, and areas where people live and work at a high density, such as apartments, row houses, and commercial and industrial areas. Some studies have used the same classification techniques to identify residential land cover (Ogneva-Himmelberger & Huang, 2015), while others also include the "Developed, Open Space" classification, which accounts for impervious surfaces of less than 20% of the total land cover and primarily includes parks, golf courses, recreation areas and some large-lot single-family housing units (Zandbergen & Ignizio, 2010).

C. **Spatial Analysis**

Four spatial analysis methods were used to estimate the demographic and socioeconomic characteristics of the population served by each CWS. These estimates were compared to the total violations and health-based violations of each CWS. The first three methods were areal weighting, dasymetric mapping, and areal interpolation (the "CWS-level

analyses"). The last method used the demographic characteristics of the county in which the CWS is located, which is the technique that was used by Allaire et al. (2018) on a national scale and is referred to as the "county-level analysis." The descriptions of these four methods below refer to estimating the population served by each CWS, which means estimating both the total population but also the population with a characteristic of interest (e.g., below the poverty line).

Method 1: Areal Weighting

An areal weighting model was developed in ArcGIS to estimate the population served by a given CWS based on the proportion of the intersecting census tracts that are within the CWS boundary. This method assumes equal distribution of the population across each census tract.

After the model was used to estimate the population served by each CWS, several of the smaller CWSs had estimates of zero people served. These CWSs were excluded from the analysis, leaving a total of 1,751 CWSs included in the areal weighting analysis.

Method 2: Dasymetric Mapping

The areal weighting model was modified such that the population served by each CWS was estimated based on the proportion of the intersecting census tracts' residential land within the CWS boundary, rather than the proportion of the tract's entire area. The residential land was identified using the NLCD data. Dasymetric mapping is considered the most detailed and accurate of all methods used in this study because it is based on a more realistic distribution of the population (Holifield et al., 2017). Thus, dasymetric mapping is used as the standard for comparison in this paper.

CWSs without any residential land and CWSs where the model resulted in an estimate of zero people served were excluded from the analysis. A total of 1,749 CWSs were included in the dasymetric mapping analysis.

Method 3: Areal Interpolation

Using ArcMap's 10.5.1's Geostatistical Wizard, the areal interpolation tool was implemented to estimate the population of a given characteristic within each CWS. The geostatistical interpolation model selected was the binomial areal kriging model, which is designed to be used for rate data. Using the visual variography tools available in the Geostatistical Wizard, a kriging interpolation model was fit to a plot of covariance versus distance. A Stable model was used for the percent below the poverty line data and a Spherical model was used for the percent people of color data. The model parameters, such as lattice spacing, lag size, and number of lags, were adjusted with the goal of achieving a standardized root mean square as close to 1.0 as possible. The result was a continuous prediction surface of the population characteristics, such as percentage of the population below the poverty line. This data was then reaggregated to the CWS boundaries. See **Figure 2** for a visual display of the steps of this method.

Method 4: County-Level Analysis

The last method assigned the percent people of color and the percent below the poverty line of the CWS's county to that CWS. Although some CWSs serve more than one county, the primary one was determined based on the county listed in the SDWIS database. The rural or urban classification of the county was also assigned to the CWS. Essentially this is the method that would need to be used if no CWS boundary data was available.

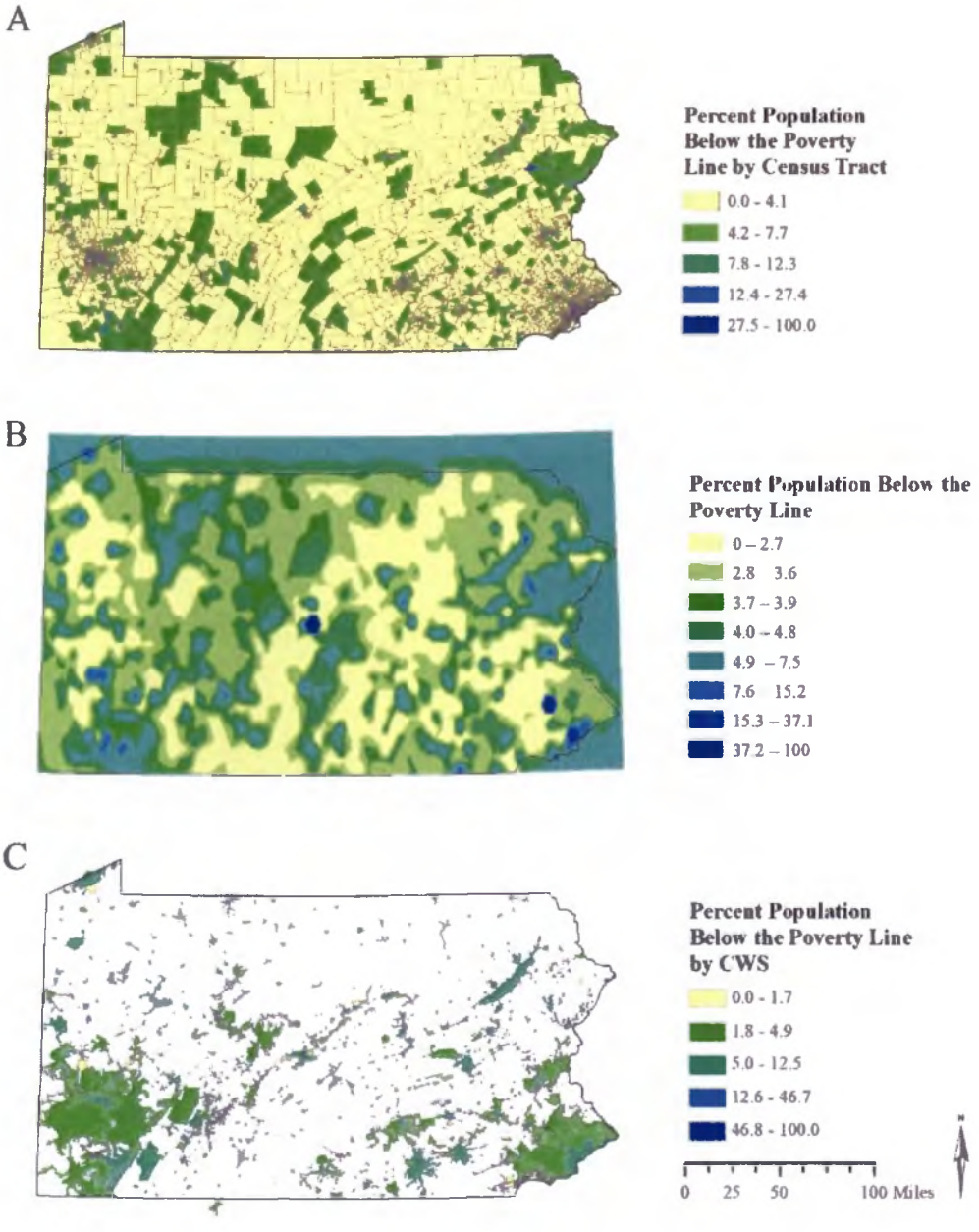

Figure 2: Visual display of the steps of areal interpolation: A) percent below the poverty line by census tract; B) continuous prediction surface of the percent below the poverty line; and C) estimates of percent below the poverty line reaggregated to the CWS boundaries. Intervals in the legend are selected based on the Jenks natural breaks classification method.

D. **Statistical Analysis**

Model

A regression model was used to determine which characteristics of the water system and population served by the water system best predict the number of violations. Since the dependent variable is a count of SDWA violations, a Poisson model, which is often used to model counts, was determined to be more appropriate than a linear regression model (Switzer & Teodoro, 2017; Wallsten & Kosec, 2008). However, a Poisson model requires the variance and the mean of the dependent variable to be equal, and the variance of the SDWA violations is much greater than the mean of the violations, a situation referred to as overdispersion ("NCSS Documentation," 2018; Switzer & Teodoro, 2017; Wallsten & Kosec, 2008). Thus, negative binomial regression, a model similar to Poisson which allows for overdispersion, was selected ("NCSS Documentation," 2018; Switzer & Teodoro, 2017; Wallsten & Kosec, 2008). Negative binomial regression was applied to all four spatial analysis results for both health-based violations and total violations.

Covariates

The covariates used in the model were selected based on the related literature. The percentage of the population below the poverty line was chosen as a proxy for SES. The percentage of non-Latino whites within the population was the second key variable chosen, one used frequently in environmental justice studies (Balazs et al., 2011; Cory & Rahman, 2009).

Characteristics of the water systems themselves are often included in environmental justice studies on drinking water quality as they are considered potentially confounding variables, but also can provide their own insight into the operations of CWSs. The type of ownership (i.e., private vs. public) is thought to potentially have an effect on SDWA compliance and has been included in several studies (Allaire et al., 2018; Balazs et al., 2011, 2012; Konisky & Teodoro, 2016; Wallsten & Kosec, 2008). Public water systems

have been found to have more violations than private water systems (Konisky & Teodoro, 2016). However, one study that focused primarily on the effects of ownership on the number of SDWA violations found ownership type did not affect compliance (Wallsten & Kosec, 2008). Smaller systems generally have less TMF for proper regulation and enforcement of the SDWA, and thus size of the system is a key variable to include (National Research Council, 1997). As noted above, Balazs et al. (2011) found evidence of environmental injustice in small systems with less than 200 connections and did not find the same trends in larger systems. Other studies have included size of the water system as well (Konisky & Teodoro, 2016). SDWA violations or drinking water contamination have also been found to be higher in rural areas compared to urban areas (Allaire et al., 2018).

Water source, primarily groundwater compared with surface water, is also included as a variable in similar drinking water quality studies (Balazs, 2011; Balazs et al., 2012; Switzer & Teodoro, 2017). Different water sources are susceptible to different types of contaminants and problems. Systems that rely on groundwater have fewer regulatory requirements compared to systems that utilize surface water (Allaire et al., 2018). Groundwater supply has been found to positively predict SDWA violations (Konisky & Teodoro, 2016). In sum, this study included the following variables in its regression model: 1) percent below the poverty line; 2) percent non-Latino white; 3) rural/urban; 4) private/public; 4) size of the system (simplified to systems with less than 200 and more than 200 connections); and 5) groundwater/surface water source. **Table 1** summarizes the covariates used in the negative binomial regression and the criteria used for each variable. A test for multicollinearity confirmed none of the selected variables were strongly correlated, and all could be included in regression model.

Table 1: Covariates used in the negative binomial regression analysis

Covariate	Criteria
Percent Below the Poverty Line	Percentage of the population below the poverty line according to the U.S. Census data.
Percent People of Color	Percentage population that is not classified as white in the U.S. Census data, including Hispanic population.
Rural	CWS-level analysis: CWSs with less than 50% of their area within an urban area as defined by the U.S. Census Bureau. County-level analysis: CWSs within counties with more than 50% of their population classified as rural population.
Public	CWS where the ownership type is classified in the SDWIS database as 'local government', 'federal government', or 'state government'.
Small size	CWS with less than 200 connections.
Groundwater Source	CWS where the primary source is classified in the SDWIS database as 'ground water', 'ground water purchased', 'groundwater under influence of surface water', and 'purchased ground water under influence of surface water source'.

3. Results

A. Spatial Analysis

The spatial distribution of the number of total and health-based SDWA violations by CWS is shown in **Figure 3** and the summary statistics of this data are shown in **Table 2**. Pennsylvania CWSs received between 0 and 804 total violations, and between 0 and 41 health-based violations during the five-year period of interest. The averages of total and health-based violations per CWS, no matter how many CWSs were included in the specific analysis, were 25 and 1, respectively (**Table 2**). There is no clear pattern in the spatial distribution of the total and health-based SDWA violations (**Figure 3**).

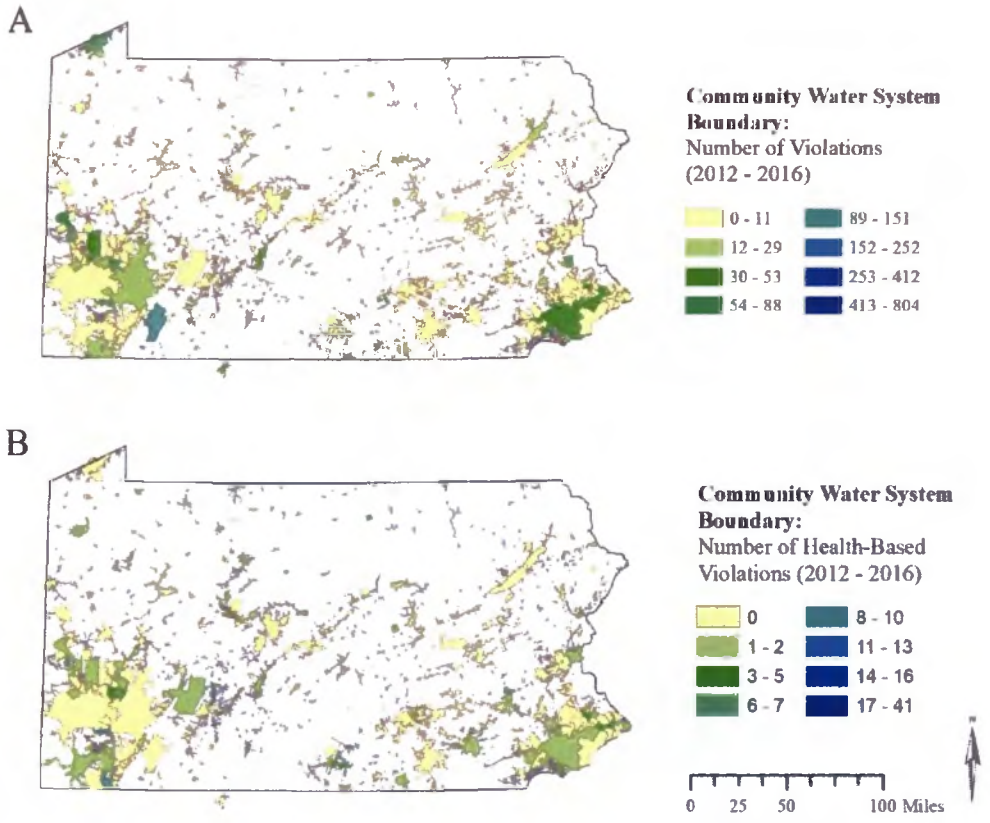

Figure 3: Number of SDWA violations by CWS in Pennsylvania (2012 - 2016): A) total violations; and B) health-based violations. Intervals in legend are selected based on the Jenks natural breaks classification method.

Table 2: Violations per CWSs in Pennsylvania (2012 - 2016)

Statistics	Method (# of CWSs included in analysis)							
	Areal Weighting (1751)		Dasymetric Mapping (1749)		Areal Interpolation (1833)		County-level (1833)	
	All	Health-based	All	Health-based	All	Health-based	All	Health-based
Min	0	0	0	0	0	0	0	0
Max	804	41	804	41	804	41	804	41
Mean	25	1	25	1	25	1	25	1
Standard Deviation (SD)	56.3	2.1	55.8	2.1	56.5	2.1	56.5	2.1

CWS-Level Analyses

The summary statistics of the relevant SES and sociodemographic data by census tract in Pennsylvania are shown in **Table 3**, and their spatial distribution is shown in **Figure 4**. The percentage of the population below the poverty line ranged between 0.4% and 100%, with a mean of 5.95% (**Table 3**). The percentage of people of color ranged between 0% and 100%, with a mean of 22.5% (**Table 3**). There is no clear pattern in the spatial distribution of the population below the poverty line by census tract (**Figure 4A**). However, census tracts with a higher percentage people of color appear to be located in the bigger urban centers, specifically Philadelphia and Pittsburgh (**Figure 4B**).

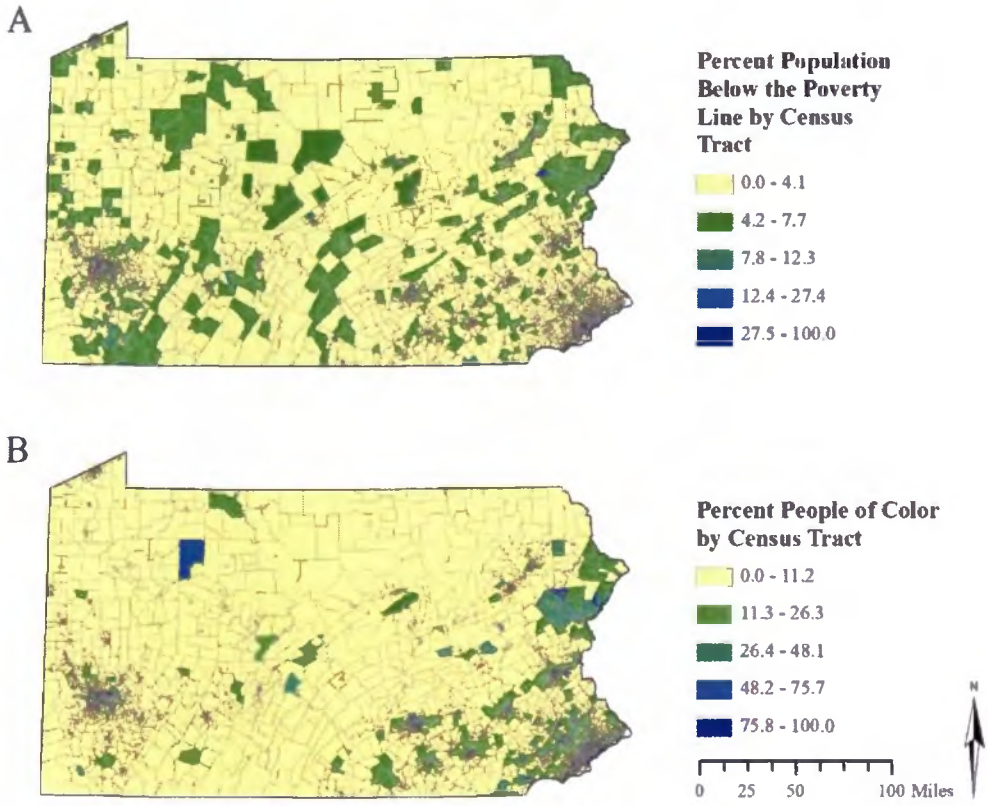

Figure 4: Characteristics of Pennsylvania population by census tract: A) percent below the poverty line; and B) percent people of color. Intervals in legend are selected based on the Jenks natural breaks classification method.

Table 3: Socioeconomic and sociodemographic data for census tracts in Pennsylvania

Statistics	Percent Population Below Poverty Line	Percent People of Color
Min	0.4	0
Max	100	100
Mean	5.95	22.5
SD	6.09	26.6

The results of the three CWS-level spatial analysis methods for assessing percent below the poverty line by CWS are shown in **Figure 5**, and the results of these three methods for assessing percent people of color by CWS are shown in **Figure 6.** At the state scale, the spatial distribution of the percent below the poverty line by CWS does not appear to vary by spatial analysis method (**Figures 5A-C**). However, the percent people of color by CWS estimated using areal interpolation (**Figure 6C**) appears to vary spatially compared to the results of the areal weighting and dasymetric mapping methods (**Figures 6A** and **6B**, respectively). The CWSs in the Philadelphia area in **Figures 6A** and **6B** have a higher percentage people of color than that shown in **Figure 6C**. There is also slight variation in the CWSs surrounding Pittsburgh.

The summary statistics of these two variables by spatial analysis method are also outlined in **Table 5**. The minimum percentage of the population below the poverty line by CWS is 0% for all three methods. The maximum percentage below the poverty line is 100% for both the areal weighing and dasymetric mapping methods, and 99.3% for the areal interpolation method. The mean values of the percentage below the poverty line are relatively similar across methods and range between 2.66% and 2.82%. The minimum percentage of people of color by CWS is also 0%. The maximum is 100% in the areal weighting method, but much less in the other two methods. The maximum people of color percentages estimated from the dasymetric mapping and areal interpolation methods are 71.6% and 76.3%, respectively. The mean values of the percentage of people

of color by CWS are also relatively similar for all three methods and range between 6.93% and 8.60%.

Although at the state scale the results appear similar, especially for **Figures 5A-C,** a closer look at the maps show how the spatial methods produce different results. **Figure 7** demonstrates an example of the variation in the estimation of the percent below the poverty line between the three different CWS-level spatial analysis methods. Two CWSs have been labeled as CWS #1 and CWS #2 in **Figure 7** for the purpose of discussing these variations. CWS #1 is estimated to have 1.96, 1.70, and 0.2 percent of the population below the poverty line through the areal weighting, dasymetric mapping and areal interpolation spatial analysis methods, respectively (**Figures 7A, 7B and 7C,** respectively). Similarly, CWS #2 is estimated to have 2.32, 2.34, and 0 percent of the population below the poverty line through the areal weighting, dasymetric mapping and areal interpolation spatial analysis methods, respectively (**Figures 7A, 7B and 7C,** respectively).

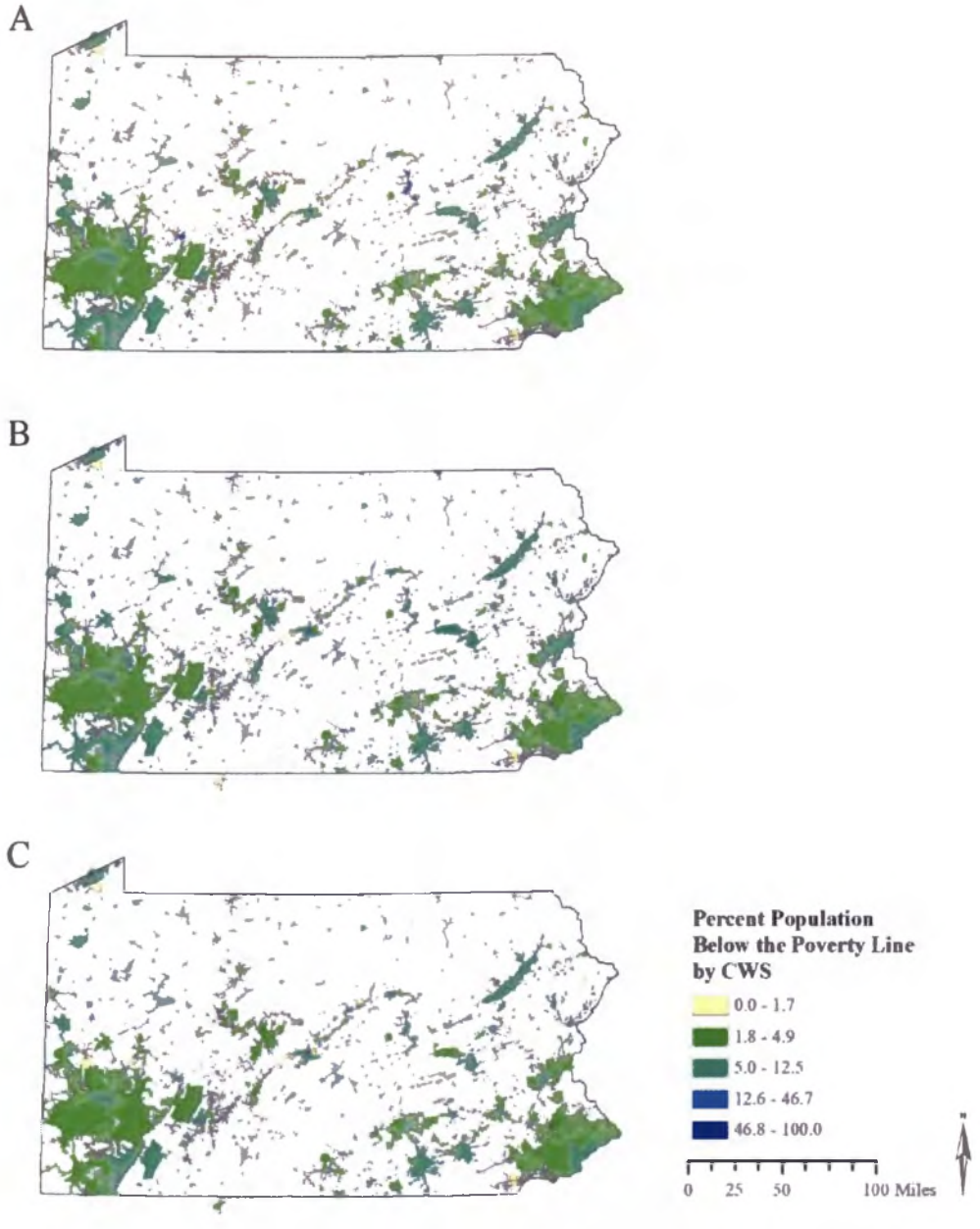

Figure 5: Percent below the poverty line by CWS calculated using different spatial analysis methods: A) areal weighting; B) dasymetric mapping; and C) areal interpolation. Intervals in legend are selected based on the Jenks natural breaks classification method.

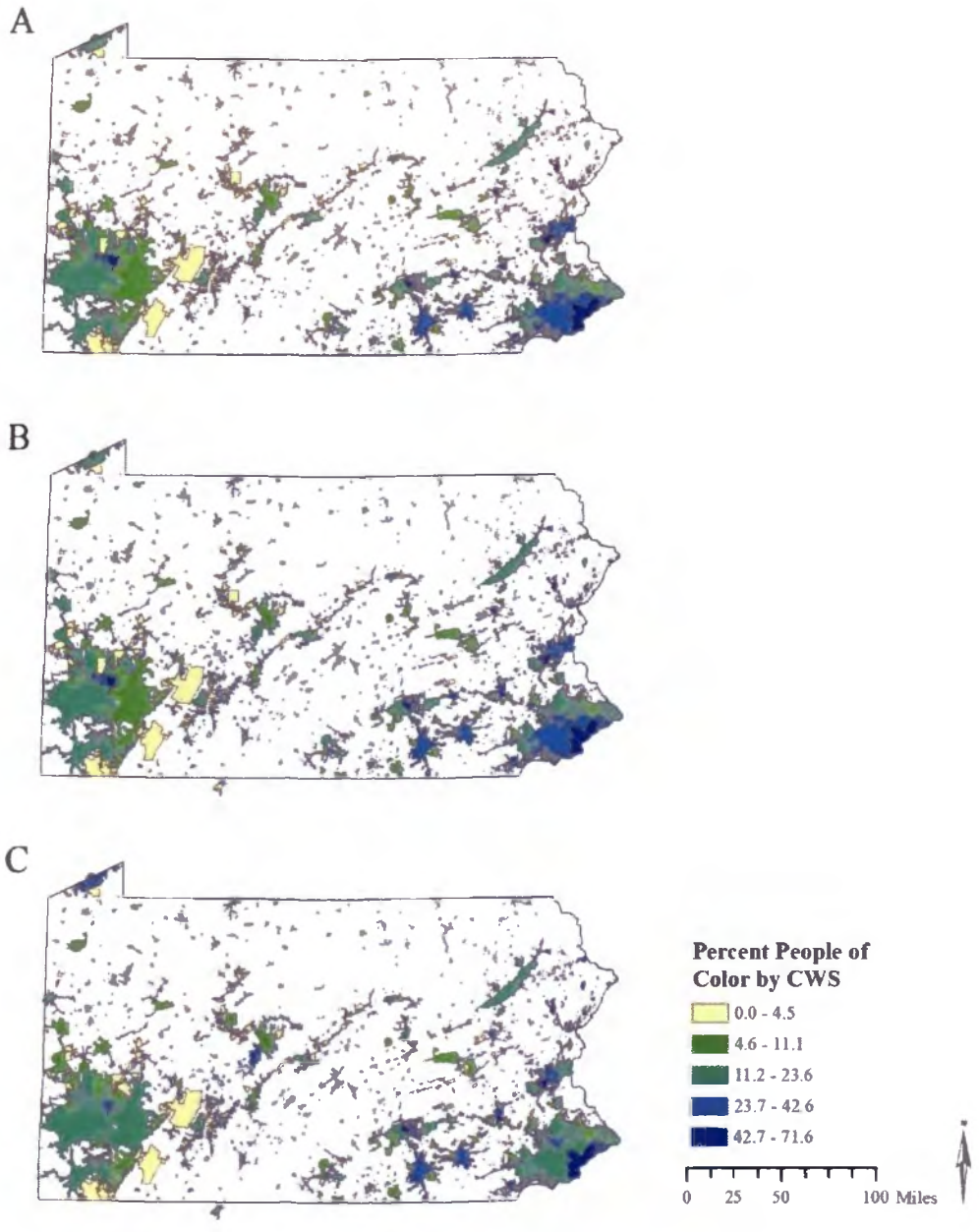

Figure 6: Percent people of color by CWS calculated using different spatial analysis methods: A) areal weighting; B) dasymetric mapping; and C) areal interpolation. Intervals in legend are selected based on the Jenks natural breaks classification method.

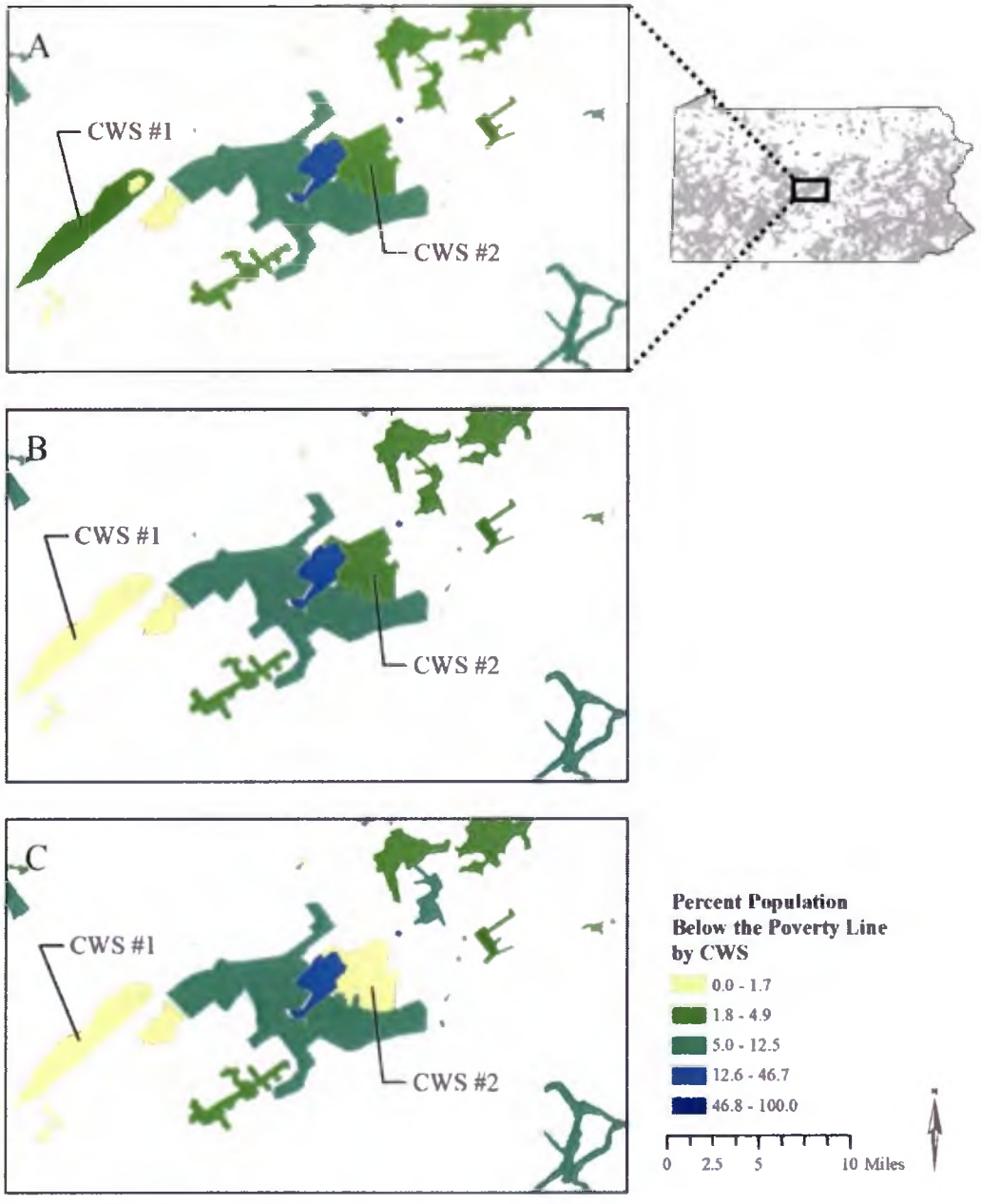

Figure 7: Percent population below the poverty line by CWS calculated using different spatial analysis methods: A) areal weighting; B) dasymetric mapping; and C) areal interpolation. Intervals in legend are selected based on the Jenks natural breaks classification method.

County-Level Analysis

The summary statistics of the SES and sociodemographic data by county in Pennsylvania are shown in **Table 4**, and their spatial distribution is shown in **Figure 8**. The percentage below the poverty line by county ranges between 6.0% and 26.4%, with a mean of 13.1% (**Figure 8A**), and the percentage people of color by county ranges between 2.4% and 64.2%, with an average of 11.2% (**Figure 8B**). There is not a clear spatial pattern in the percent below the poverty line by county, but **Figure 8B** shows the highest percentages of people of color reside in the counties surrounding Philadelphia.

Table 4: Socioeconomic and sociodemographic data for counties in Pennsylvania

Statistics	Percent Population Below Poverty Line	Percent People of Color
Min	6.00	2.40
Max	26.4	64.2
Mean	13.1	11.2
SD	3.27	10.6

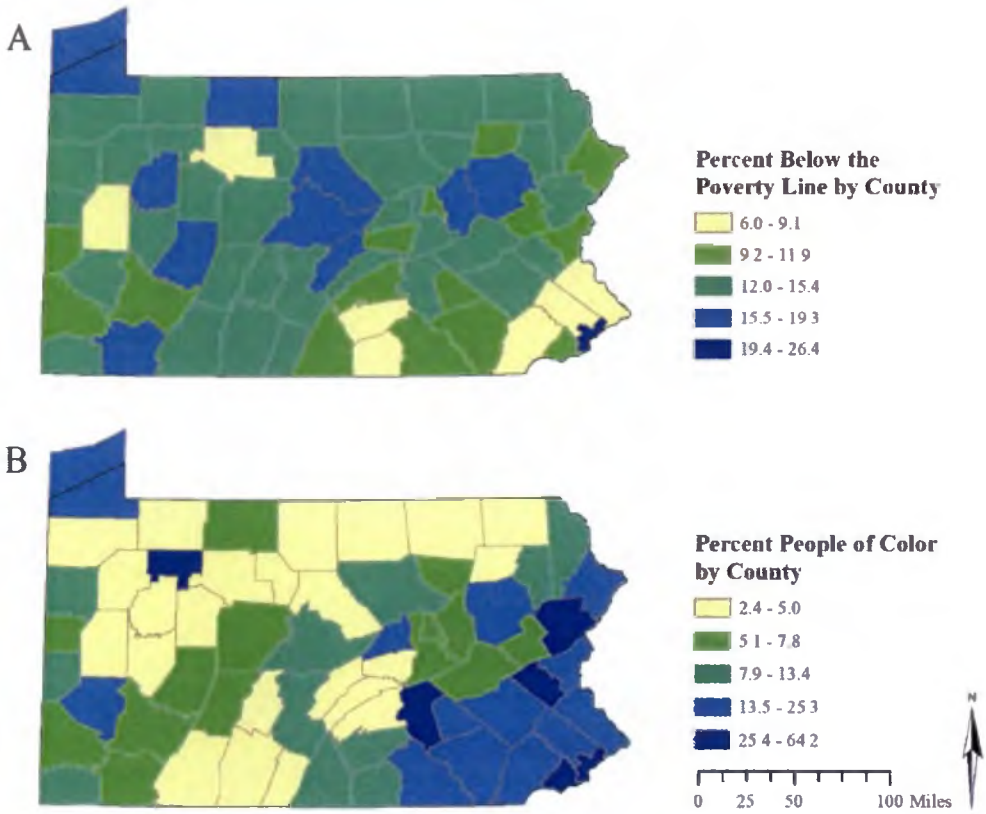

Figure 8: Sociodemographic data by county: A) percent below the poverty line; and B) percent people of color. Intervals in legend are selected based on the Jenks natural breaks classification method.

It is important to visualize how the county-level analysis can vary by county to understand the limitations of this method. **Figure 9** demonstrates how CWSs are spatially distributed in three Pennsylvania counties compared to census data. The entirety of the first, Philadelphia County, is served by only one CWS so the analysis at the county level is the same as the analysis at the CWS level (**Figures 9A1** and **9A2**). In the second, Fayette County, several of the CWSs partially within the county's boundaries, but located in another county according to the SDWIS database, had Fayette County's demographic data excluded from their analysis (shown in grey) (**Figures 9B1** and **9B2**). These grey

CWSs are associated with other counties in the SDWIS database, and thus those counties' demographic data were assigned to them for the statistical analysis. Additionally, although the CWSs within the county only cover a portion of the county's area, the census data of the entire county was assigned to each of those CWSs. The last county, Cameron County, has only two small CWSs and two census tracts. One of the CWSs lines up closely with a census tract, while the other census tract covers the rest of the county, including the second small CWS. This suggests that for the CWS that lines up well with a census tract, the sociodemographic characteristics estimated from the CWS-level analysis would be more accurate compared to the county-level analysis (**Figures 9C1** and **9C2**).

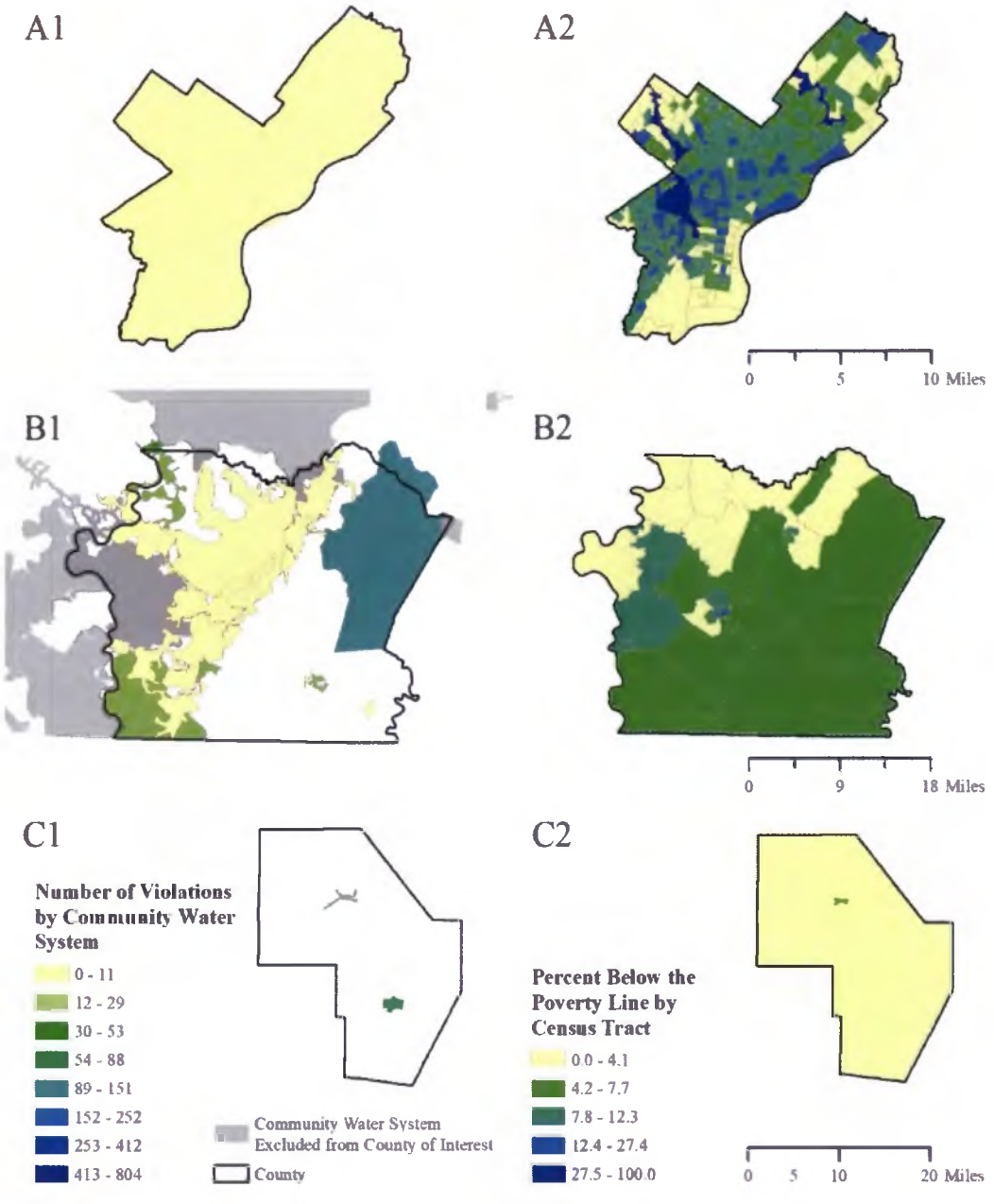

Figure 9: Examples of analysis at the county level: A1) the number of total violations by CWS in Philadelphia County; B1) the number of total violations by CWS in Fayette County; C1) the number of total violations by CWS in Cameron County; A2) the percent below the poverty line by census tract in Philadelphia County; B2) the percent below the poverty line by census tract in Fayette County; and C2) the percent below the poverty line by census tract in Cameron County. Intervals in legend are selected based on the Jenks natural breaks classification method.

Summary Statistics by CWSs

The summary statistics for the sociodemographic and SES variables by CWS and the categorical variables by CWS for all four spatial analysis methods are presented in **Table 5** and **Table 6**, respectively. The summary statistics of the continuous variables (percent below the poverty line and percent people of color) are relatively similar across the three CWS-level analyses but there are some stark differences between these three analyses and the county-level analysis. The maximum value for the percent below the poverty line by county (26.4%) is significantly less than that estimated by the other three methods (ranging between 99.3% and 100%). Similarly, the maximum value of the percent people of color by county (64.2%) is less than the other three CWS-level methods (ranging between 71.6% and 100%). The mean value of the percent below the poverty line by county (12.5%) is much higher than the values estimated by the other three CWS-level methods (ranging between 2.66% and 3.83%). The mean value of the percent people of color by county (12.8%) is also higher than the other three methods (ranging between 6.93% and 8.60%), but the difference is not as drastic.

Table 5: Summary statistics by CWS for continuous variables

Statistics	Method (# of CWSs included in analysis)			
	Areal Weighting (1751)	Dasymetric Mapping (1749)	Areal Interpolation (1833)	County-level (1833)
Percent Below the Poverty Line				
Min	0.00	0.00	0.00	6.00
Max	100	100	99.3	26.4
Mean	2.66	3.56	3.83	12.5
SD	4.52	4.21	3.35	3.13
Percent People of Color				
Min	0.00	0.00	0.00	2.40
Max	100	71.6	76.3	64.2
Mean	6.93	7.79	8.60	12.8
SD	10.6	10.0	10.1	8.24

The categorical variable summary statistics for the public, small size, and groundwater sources variables are very similar across all four methods since the data for each CWS comes directly from the SDWIS database and is not dependent on the spatial analysis method. The numbers vary slightly because of the difference in the number of CWSs included in each analysis. The county-level analysis estimated significantly fewer rural CWSs (559) and more urban CWSs (1,275) compared to the other three methods (ranging between 1,005 and 1,079 rural systems and 740 and 754 urban systems).

Table 6: Count of categorical variables

Variable	Method (# of CWSs included in analysis)							
	Areal Weighting (1751)		Dasymetric Mapping (1749)		Areal Interpolation (1833)		County-level (1833)	
	Yes	No	Yes	No	Yes	No	Yes	No
Rural	1,011	740	1,005	744	1,079	754	559	1,274
Public	726	1,025	725	1,024	729	1,104	729	1,104
Small size	1,009	742	1,007	742	1,091	742	1,091	742
Groundwater Source	1,353	398	1,350	399	1,434	399	1,434	399

B. **Statistical Analysis**

The coefficient estimates for the variables included in the regression models are shown in **Table 7** for the total SDWA violations, and **Table 8** for the health-based SDWA violations. The significance levels shown are $p<0.01$, $p<0.05$, and $p<0.10$, similar to other statistical analyses conducted with SDWA data (Allaire et al., 2018; Wallsten & Kosec, 2008). After the models were run, the raw residuals (actual minus predicted) were mapped to determine if there were any spatial trends that indicated a key spatial variable may have been excluded from the analysis. No clear trend was found so additional variables were not included.

Table 7 shows that in the three CWS-level regression models, an increase in the percent of the population below the poverty line resulted in a decrease in the number of total SDWA violations ($p<0.10$), while no significant effect was found in the county-level regression model. The percent people of color had a significant negative effect in the county-level model but was not found to be significant in the other three models. The type of ownership (public vs private) did not have a significant effect ($p>0.10$) on the outputs of any of the models. The effects of whether the CWS was in a rural area and whether it had fewer than 200 connections were both positive and highly significant ($p<0.01$) across all four models, suggesting that a rural and/or smaller CWS (less than 200 connections) is likely to have more violations than an urban and/or large CWS. Lastly, all models but the areal interpolation model showed more total violations in CWSs that had groundwater as a drinking water source ($p<0.10$). In sum, the effects of the rural, public, size, and water source on the total violations were comparable across all four total violation models (**Table 7**). In contrast, the effects of the percent below the poverty line and percent people of color on the total violations were comparable for the CWS-level analysis models, but not the county-level model (**Table 7**).

Table 7: Determinants of total SDWA violations

Variable	Method (# of CWSs included in analysis)							
	Areal Weighting (1751)		Dasymetric Mapping (1749)		Areal Interpolation (1833)		County-level (1833)	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Percent Below the Poverty Line	-0.016***	0.009	-0.016***	0.009	-0.018***	0.011	0.010	0.011
Percent People of Color	0.004	0.004	0.003	0.004	0.001	0.004	-0.009***	0.005
Rural	0.332*	0.073	0.380*	0.073	0.363*	0.715	0.455*	0.085
Public	0.128	0.90	0.124	0.090	0.092	0.089	-0.049	0.088
Small size	0.400*	0.094	0.419*	0.093	0.418*	0.092	0.405*	0.090
Groundwater Source	0.153***	0.099	0.169***	0.099	0.150	0.098	0.162***	0.096
Constant	2.613*	0.124	2.560*	0.126	2.630*	0.125	2.662*	0.193
AIC	14,111.65		14,101.23		14,855.20		14,821.50	
Log-Likelihood	-7048.8		-7,043.6		-7,420.6		-7403.7	
Pearson chi2	3,956.21		3,796.18		4,025.54		3,792.29	

Coeff. = Coeffecients
SE = Standard Error
Significance levels: *p<0.01, **p<0.05, ***p<0.10

The results of the models predicting health-based SDWA violations were similar to the results of the models predicting total SDWA violations for the rural, public, and size variables. Refer to **Table 8** for the results of the four regression models predicting health-based violations. There was a significant ($p<0.05$) positive correlation between the rural and small size CWS variables and the number of health-based violations for all four models, except for the size of the CWS was not a significant variable under the areal weighting model ($p>0.10$). Whether the CWS was public or private had no significant effect ($p>0.10$) for any of the models predicting health-based violations, similar to the

findings of the models predicting total violations. All models showed fewer health-based violations in CWSs that had groundwater as a drinking water source ($p<0.01$), which is the opposite of what was found in the total violation models. The percent below the poverty line did not have a significant effect on the number of health-based violations in the CWS-level models ($p>0.10$) but did have a significant negative effect in the county-level model ($p<0.10$). The percent people of color had a positive effect for the areal weighting and dasymetric mapping models ($p<0.10$) but had no significant effect in the other two models. Similarly to the total SDWA violation model results, the findings were analogous across all four models for the CWS characteristic variables (rural, public, small size, water source), but were not comparable for the SES and sociodemographic variables (**Table 8**).

Table 8: Determinants of health-based SDWA violations

Variable	Method (# of CWSs included in analysis)							
	Areal Weighting (1751)		Dasymetric Mapping (1749)		Areal Interpolation (1833)		County-level (1833)	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Percent Below the Poverty Line	-0.015	0.014	-0.019	0.015	-0.004	0.016	-0.034***	0.018
Percent People of Color	0.010***	0.006	0.010***	0.006	0.008	0.006	0.008	0.008
Rural	0.335*	0.119	0.319*	0.119	0.311*	0.116	0.264**	0.139
Public	0.007	0.144	-0.083	0.145	-0.120	0.143	-0.062	0.145
Small size	0.237	0.152	0.333**	0.152	0.311**	0.150	0.392*	0.150
Groundwater Source	-0.8067*	0.156	-0.9705*	0.157	-0.969*	0.156	-1.023*	0.155
Constant	-0.091	0.195	0.063	0.198	0.044	0.952	0.487	0.310
AIC	3,792.12		3,809.11		3,993.61		3,994.34	
Log-Likelihood	-1,889.10		-1,897.6		-1,989.8		-1990.2	
Pearson chi2	2,397.81		2,133.53		2,243.75		2,235.21	

Coeff. = Coeffecients
SE = Standard Error
Significance levels: *p<0.01, **p<0.05, ***p<0.10

As noted above, the effect of the rural variable was significant across all models ($p<0.05$). Rural CWSs are likely to have more total and health-based SDWA violations compared to urban CWSs. **Figure 10** shows the mean total and health-based SDWA violations by spatial analysis method in rural systems compared to urban systems. The results are relatively similar across methods, although the county-level method estimated an average that is about 5 violations higher than the other methods (~35 violations/system compared to ~30 violations/system). The difference between rural and urban CWSs is

much more pronounced for the total SDWA violations compared to the health-based violations (**Figure 10**). On average, a rural CWS experienced approximately 10 to 15 more total SDWA violations in the time period of interest than the urban CWSs.

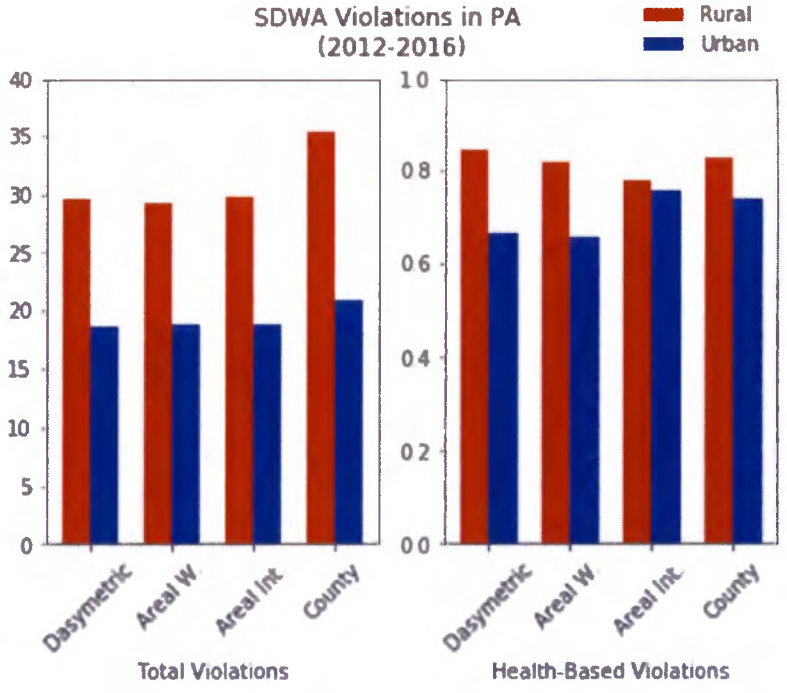

Figure 10: Mean total and health-based SDWA violations in rural and urban CWSs as estimated by the four different analysis methods: dasymetric mapping, areal weighting, areal interpolation, and by county.

4. Discussion

A. Regression Model Comparison

As discussed above, the continuous and categorical variable summary statistics were relatively similar across the four spatial analysis methods. However, there were a few key differences. The maximum values of the percent below the poverty line and percent people of color were significantly less in the county-level analysis. This is likely due to the fact that the county maximum value is an average of the population in a large area, while the other three methods capture smaller areas of the county, and thus include more

outliers. Additionally, the county-level analysis identified substantially fewer rural CWSs than the other three methods. This is likely because the method of categorizing a county as rural or urban was based on total number of people in the county living in rural areas compared to urban areas, whereas the CWS-level analysis was based on the area of a CWS system within a U.S. Census Bureau-designated urban area. Since fewer people live in rural areas by definition, the county had to be significantly rural to have a greater rural population and thus be classified as rural. Despite this discrepancy, **Figure 10** and the regression results (**Table 7** and **Table 8**) demonstrate that rural CWSs tend to have more violations than urban CWSs across all four spatial analysis methods.

The results of the regression models for both total and health-based SDWA violations were relatively similar across all four spatial analysis methods with regard to the CWS characteristics (rural vs urban, public vs private, size, water source), but not the demographic data. The similarity across regression models in the results of the CWS characteristic variables is likely due to the fact that all of the data other than the rural/urban variable came directly from the SDWIS database and the only difference between regression models was how many CWSs were included in the analysis. However, it is clear from looking at the results of the SES and sociodemographic data coefficients that the spatial analysis method selected can directly affect the conclusions of an environmental justice study, particularly when factoring in the county-level analysis. These findings are similar to recent environmental justice research, which has also shown that the selected spatial analysis method can have a direct impact on the outcome (Maantay & Maroko, 2009; Ogneva-Himmelberger & Huang, 2015).

The results of the regression models with regard to the percent below the poverty line and percent people of color were different for both sets of regression models predicting total and health-based SDWA violations. In the total SDWA violations analysis, the county-level method was the only one that did not find the percent below the poverty line to have a significant effect ($p>0.10$), and the only method to find that the percent people of color

did have a significant effect ($p<0.10$). The differences were similar in the health-based SDWA violations models. The county-level regression model was the only one that found percent below the poverty line had a significant effect ($p<0.10$), and both the county-level and areal interpolation methods did not find the percent people of color had a significant effect, while the other two methods did ($p<0.10$).

The differences between the results of the county-level analysis compared to the CWS-level analyses were not surprising. As discussed above, **Figure 9** shows how the total violation data is spatially distributed in three Pennsylvania counties. **Figure 9B,** which highlights Fayette county, demonstrates how CWSs may cover more than one county, and how large portions of a county's area, and thus some of its population, may be factored into the analysis of that CWS even when far outside the boundaries of the system. Pennsylvania has many counties like Fayette County, which would mean census data of areas without CWSs is inappropriately being included in the county-level analysis. This situation is not unique to Pennsylvania and should be evaluated for other states across the U.S. as CWSs can cover more than one county, and many counties have areas not covered by CWSs.

These variations in the regression model results can have a large effect on research conclusions and resulting actions. This is demonstrated by the findings of the effect of the percentage of people of color within a CWS on the number of health-based SDWA violations. If one was only analyzing the county-level or areal interpolation regression models, one could conclude that the percentage people of color has no effect on the number of health-based violations. However, the other two regression models (using the results of the areal weighting and dasymetric mapping spatial analysis methods) show a positive effect, meaning that the higher the percentage of residents of color, the more likely the CWS will receive a health-based violation.

B. **Environmental Justice Analysis**

This study's finding that poorer CWSs had fewer total SDWA violations across three of the four models is surprising, but not unusual. A national study that investigated violations by every CWS between 1997 and 2003 found that wealthier counties had more SDWA violations. That national study used income as a proxy for wealth (Wallsten & Kosec, 2008). Another state-specific study found per capita income and average house value had no statistically significant effect on Arizona's implementation of a 2002 arsenic standard (Cory & Rahman, 2009). Similarly, in this Pennsylvania-specific study, the percent below the poverty line within a CWS was not found to have an effect on the number of health-based SDWA violations.

This study showed a positive relationship in two of the regression models between the number of people of color within a CWS and the number of health-based SDWA violations, and a negative relationship in the county-level regression model between the number of people of color within a CWS and the number of total violations. The other the regression models showed no significant relationship between the percent people of color and the total or health-based SDWA violations. Previous studies also did not find that a greater percentage of people of color necessarily resulted in more violations or an increase in concentrations of contaminants. In the study of Arizona's implementation of the new arsenic standard, Cory and Rahman (2009) did not find evidence that communities of color experienced inequitable implementation of the new standard. While Balazs et al. (2012) found that communities of color had greater odds of having an MCL violation, they did not find that these CWSs had higher arsenic concentrations. Balazs et al. (2011) also did not find a statistically significant relationship between race/ethnicity and nitrate levels in CWSs with more than 200 connections in San Joaquin Valley, California, although they did within communities of less than 200 connections. In sum, this study did not find conclusive evidence of environmental injustice in CWS violations by race or class, as was the case in several related studies.

The most significant finding of this study is that rural and small CWSs are likely to have more violations than other CWSs. Allaire et al. (2018) found that rural areas had more violations than urban areas, and concluded that small, rural CWSs relying on surface water sources had the highest predicted probability of a SDWA violation. As discussed in the introduction, rural communities have a smaller customer base to generate the necessary revenue for proper treatment technology and maintenance and to hire experienced utility managers. One proposed solution to the struggles of small CWSs is to combine smaller systems into a larger PWS (National Research Council, 1997). However, this is not an easy solution in practice. Balazs and Ray (2014) describes an example in San Joquin Valley in which a small CWS with poor water quality, including high nitrate concentrations, hoped to consolidate with a larger system with more wells and lower nitrate levels, but the larger city resisted the consolidation.

In 1997, the National Research Council's Committee on Small Water Supply Systems wrote a book titled, "Safe Water From Every Tap: Improving Water Service to Small Communities," outlining the challenges faced by small rural water systems, in addition to proposed actions and solutions. It is clear from the results of the Allaire et al. (2018) national study and this Pennsylvania-specific study that rural communities are still facing the same problems in accessing quality drinking water two decades later.

The distribution of population characteristics within Pennsylvania is not reflective of all states, which could help explain the lack of a finding of environmental injustice. Rural areas in Pennsylvania tend to have a higher percentage of white individuals compared to states such as California or Arizona (Cedar Lake Ventures Inc., 2018). Studies in California did find higher rates of violations in communities of color (Balazs et al., 2012; The Environmental Justice Coalition for Water, 2005), but this may be related to the distribution of diversity in California, where many of the communities of color are rural (Cedar Lake Ventures Inc., 2018). The PA DEP defines its own Environmental Justice areas as census tracts with more than 30 percent people of color and over 20% of the

population is "in poverty." These areas tend to be clustered in urban areas in Pennsylvania, such as Philadelphia and Pittsburgh (Fractracker Alliance, 2018). Thus, it is not entirely surprising that in this Pennsylvania-specific study, strong evidence of environmental injustice was not found, as at a national scale, it is the rural areas that are found to have a much higher rate of violations (Allaire et al., 2018). It is important that future state or region-specific studies keep the distribution of the relevant population characteristics in mind when analyzing the results.

C. **Limitations**

There were several limitations in this research associated with the data and spatial analysis methods. There are likely inaccuracies in the U.S. Census Bureau data that was used for the percent people of color and percent below the poverty line. Additionally, this study used the 2011-2015 American Community Survey (ACS) 5-Year Estimates, but the SDWA data was for the years 2012-2016. The CWS boundaries may contain errors as each water system reported the boundaries themselves and the PA DEP only assessed the boundaries if there was overlap (Pennsylvania Department of Environmental Protection, 2017). The undereporting of violations identified as a problem within the SDWIS database could also have an impact on the count of total and health-based violations (Balazs, 2011; U.S. EPA's Office of Enforcement and Compliance Assurance, 2013). Similarly, it is possible that the SDWIS database is not entirely up-to-date on the descriptions of its CWSs, meaning that some of the categorical data (e.g., public vs private) may be inaccurate.

Research suggests there may be problems with the SDWIS dataset specific to Pennsylvania. Within the last couple of years, the PA DEP has become in danger of losing primacy as "it lacks the necessary staffing and resources to enforce safe drinking water standards." The number of unaddressed SDWA violations in Pennsylvania nearly doubled in the five years prior to December 2016, essentially the time period of this study

(Cusick, 2017). While an increase in violations would not impact this study, a lack of funding could mean the PA DEP has not been properly reporting violations to the federal government via the SDWIS database. If the state is not entering all the violation data into the SDWIS database, monitoring and reporting violations may not appear in the dataset (U.S. EPA's Office of Enforcement and Compliance Assurance, 2013). In addition, the PA DEP has not recently been meeting the minimum number of required inspections of PWSs, which could mean PWSs are committing violations without the state's knowledge and thus these violations would not be recorded in the SDWIS database (Cusick, 2017). The potential gaps in the SDWA violation dataset used in this study may have had an impact on the conclusion regarding environmental injustice. Since undereporting of violations has been identified as a problem in the entire SDWIS database, other states may have similar issues (U.S. EPA's Office of Enforcement and Compliance Assurance, 2013).

Compliance with the SDWA does not equate safe drinking water quality, and thus this study cannot be considered an environmental justice analysis of access to safe drinking water. This is true for the many reasons outlined in the introduction, including that the EPA is slow to regulate new contaminants (Fedinick et al., 2017). Pennsylvania has experienced a sharp increase in fracking in the last decade, an activity acknowledged to cause water pollution and yet its wastes and wastewater are still inadequately monitored and regulated (Clough & Bell, 2016). With Marcellus Shale underlying the majority of Pennsylvania, fracking has grown significantly since the first well was drilled in the state in 2005; as of October 2018, there were 11,437 active unconventional wells (Clough & Bell, 2016; Fractracker Alliance, 2018). As the complete chemical makeup of the hydraulic fracturing fluid is considered a trade secret, relatively few specifics are known about the potential harmful health and environmental effects on water (Ogneva-Himmelberger & Huang, 2015). However, since 2008, there have been 13,473 violations of state environmental regulations given to unconventional wells (Fractracker Alliance,

2018). As this study is only assessing compliance with the SDWA, which has not been updated in 20 years, it would likely not capture any disparities found in exposure to fracking contaminants, even though the industry has grown so rapidly in Pennsylvania. Environmental justice studies of the location of fracking wells have been completed though; Ogneva-Himmelberger and Huang (2015) found that communities surrounding unconventional gas wells in Pennsylvania tended to be low-income, but they did not find evidence of any disparities by race.

Transforming the scale of spatial data in order to compare datasets of different scales is a common challenge in geographic analysis (Hallisey et al., 2017). This study utilized several different geospatial techniques to convert the census data to the scale of the CWS in order to understand how that transformation potentially affected the results. However, many similar studies only utilized one method of spatial data transformation (Allaire et al., 2018; Cory & Rahman, 2009). Allaire et al. (2018) determined that "[a]ssigning census information of one county to each CWS is reasonable, given that over 97% of systems in our study serve only a single county." While that statement addresses the circumstances of CWSs straddling more than one county, it does not address the fact that the county's census data do not necessarily reflect that of the CWS. Assessing demographics at the county scale will mask the variations in population characteristics at a smaller community level. Allaire et al. (2018) made no effort to compare methods to support their assertion that their method was "reasonable." Other studies did include a comparison of methods to provide support for their technique of estimating the demographics of the population served by each CWS (Balazs et al., 2011). As CWSs in Pennsylvania do not track the demographics of their communities served, there is not an effective way of assessing the accuracy of the spatial analysis methods used in this study to estimate the SES and racial makeup of the communities served. There is uncertainty with all four spatial analysis methods used, and thus any applications of these methods

needs to be done with an understanding of the challenges of transforming the scale of spatial data.

D. **Future Research**

There are two primary findings of this study that could help inform future research: 1) the results of environmental justice studies can vary depending on the spatial analysis methods used to estimate the sociodemographic characteristics of the communities of interest; and 2) small, rural CWSs tend to experience more violations of the SDWA. The finding that results can vary depending on the spatial analysis methods used has important implications for similar environmental justice analyses conducted in the future. When the CWS boundary data exists, it is best to use the most comprehensive method, which in this case would be dasymetric mapping. However, as so few states produce CWS boundary data, the county-level analysis method is sometimes the only option. For future national-scale county-level studies, such as the one recently completed by Allaire et al. (2018), it would be prudent to use the same statistical methods used in the study on data produced from a CWS-level spatial analysis, such as dasymetric mapping, in a state with available CWS boundary GIS data. This would allow the researchers to compare the county-level results of their methods to a more robust method of spatial analysis. Understanding the potential bias in a study would help shape the discussion of the results and the conclusions drawn from them.

Balazs and Ray (2014) developed the Drinking Water Disparities Framework that identifies actors and systems that perpetuate social inequalities in drinking water and addresses the challenges faced by small (i.e., rural) water systems. They argue that funding at the regional level is needed to support TMF capacity and help CWSs develop engineering and financial strategies for proper infrastructure and SDWA compliance (Balazs & Lubell, 2014). Future studies should focus on how regional agencies or organizations can best and most efficiently support CWSs with obtaining compliance,

and policy makers need to take seriously the need for funding, particularly in small rural communities and/or low-income communities of color, as is demonstrated by Balazs and Ray (2014) and the crisis in Flint. State-wide studies like this one could be beneficial in identifying any potential disparities in the state, and what areas are in most need of regional support to reach compliance. This could assist in providing the most efficient allocation of funding to the areas in most need. Case studies of certain regions successfully reducing their violations could be a beneficial blueprint for similar CWSs.

5. Conclusion

This study demonstrates that the spatial analysis method used in an environmental justice study can significantly affect the outcome and interpretations. The findings of the regression models varied significantly regarding the SES and sociodemographic variables across the four spatial analysis methods for both total and health-based SDWA violations. As noted above, the dasymetric mapping method is the most robust and most reflective of the population's distribution (Holifield et al., 2017), and thus was used as the basis of comparison for the other methods used in this study. While the county-level model was comparable to the dasymetric mapping model across many variables, it differed in the key environmental justice-related categories. For example, if the spatial analysis had only included county-level data, we might have concluded that health-based violations were not affected by the percentage of people of color, while the dasymetric mapping results showed that percentage of people of color corresponded to an increase in health-based violations. It is important to be aware of the potential effects of the spatial analysis method used and community-level variations may be masked if county-level data is used.

Conclusive evidence of environmental injustice in SDWA violations by race or class was not found in this study; however, the research did identify a disparity in SDWA compliance in urban vs small, rural communities. Rural CWSs in Pennsylvania are likely to have more total and health-based SDWA violations than their urban counterparts, a

discrepancy which was also found to be true on a national scale for health-based violations (Allaire et al., 2018). It is possible that the fact that the poorer communities of color are centered in the urban areas in Pennsylvania is the reason evidence of environmental injustice was not found.

Disparities in many states go beyond the rural vs urban divide, however, as other studies have found evidence of environmental injustice by race and class with regard to drinking water quality. (Balazs et al., 2011, 2012). Qualitative assessments of the recent events in Flint, Michigan also demonstrate the inequalities faced by poorer communities of color in drinking water quality, compliance and enforcement (Butler et al., 2016). Prior studies analyzed in conjunction with this one show that disparities in drinking water quality and compliance need to be addressed through research and increased funding at the local, state and national level.

6. References

Allaire, M., Wu, H., & Lall, U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences*, *115*(9), 2078–2083. https://doi.org/10.1073/pnas.1719805115

Amico, C., DeBelius, D., Detrow, Sc., & Stiles, M. (2015). Natural Gas Drilling in Pennsylvania. Retrieved from http://stateimpact.npr.org/pennsylvania/drilling/

Balazs, C. L. (2011). *Just Water? Social Disparities and Drinking Water Quality in California's San Joaquin Valley.* University of California, Berkeley.

Balazs, C. L., & Lubell, M. (2014). Social learning in an environmental justice context: a case study of integrated regional water management. *Water Policy*, *16*, 97–120. https://doi.org/10.2166/wp.2014.101

Balazs, C. L., Morello-Frosch, R., Hubbard, A. E., & Ray, I. (2011). Social disparities in nitrate-contaminated drinking water in California's San Joaquin Valley. *Environmental Health Perspectives*, *119*(9), 1272–1278. https://doi.org/10.1016/j.socscimed.2013.05.029

Balazs, C. L., Morello-Frosch, R., Hubbard, A. E., & Ray, I. (2012). Environmental justice implications of arsenic contamination in California's San Joaquin Valley: a cross-sectional, cluster-design examining exposure and compliance in community drinking water systems. *Environmental Health*, *11*(84), 1–12. https://doi.org/10.1186/1476-069X-11-84

Balazs, C. L., & Ray, I. (2014). The drinking water disparities framework: On the origins and persistence of inequities in exposure. *American Journal of Public Health*, *104*(4), 603–611. https://doi.org/10.2105/AJPH.2013.301664

Butler, L. J., Scammell, M. K., & Benson, E. B. (2016). The Flint, Michigan, Water Crisis: A Case Study in Regulatory Failure and Environmental Injustice. *Environmental Justice*, *9*(4), 93–97. https://doi.org/10.1089/env.2016.0014

Campbell, C., Greenberg, R., Mankikar, D., & Ross, R. D. (2016). A case study of environmental injustice: The failure in Flint. *International Journal of Environmental Research and Public Health*, *13*(10), 951. https://doi.org/10.3390/ijerph13100951

Cedar Lake Ventures Inc. (2018). The Demographic Stastical Atlas of the United States. Retrieved from https://statisticalatlas.com

Chakraborty, J., Maantay, J. A., & Brender, J. D. (2011). Disproportionate proximity to environmental health hazards: Methods, models, and measurement. *American Journal of Public Health*, *101*(S1), 27–36. https://doi.org/10.2105/AJPH.2010.300109

Clough, E., & Bell, D. (2016). Just fracking: a distributive environmental justice analysis of unconventional gas development in Pennsylvania, USA. *Environmental Research Letters*, *11*(2), 025001. https://doi.org/10.1088/1748-9326/11/2/025001

Cory, D. C., & Rahman, T. (2009). Environmental justice and enforcement of the safe drinking water act: The Arizona arsenic experience. *Ecological Economics*, *68*(6), 1825–1837. https://doi.org/10.1016/j.ecolecon.2008.12.010

Cusick, M. (2017, February 1). Pennsylvania lacks resources to enforce safe water standards, says EPA. *StateImpact Pennsylvania*. Retrieved from https://stateimpact.npr.org/pennsylvania/2017/02/01/pennsylvania-lacks-resources-to-enforce-safe-water-standards-says-epa/

Environmental Working Group. (2017). State of American Drinking Water. Retrieved from https://www.ewg.org/tapwater/state-of-american-drinking-water.php#.WZoWZyiGNPY

Fedinick, K. P., Wu, M., Pandithartne, M., & Olson, E. (2017). *Threats on Tap: Widespread Violations Highlight Need for Investment in Water Infrastructure and Protections*. Retrieved from https://www.nrdc.org/resources/threats-tap-widespread-violations-water-infrastructure

Fractracker Alliance. (2018). Pennsylvania Shale Viewer. Retrieved from https://www.fractracker.org/map/us/pennsylvania/pa-shale-viewer/

Greenberg, M. R. (2016). Delivering fresh water: Critical infrastructure, environmental justice, and Flint, Michigan. *American Journal of Public Health*, *106*(8), 1358–1360. https://doi.org/10.2105/AJPH.2016.303235

Hallisey, E., Tai, E., Berens, A., Wilt, G., Peipins, L., Lewis, B., … Lunsford, N. B. (2017). Transforming geographic scale: A comparison of combined population and areal weighting to other interpolation methods. *International Journal of Health Geographics*, *16*(1), 1–16. https://doi.org/10.1186/s12942-017-0102-z

Holifield, R., Chakraborty, J., & Walker, G. (Eds.). (2017). *The Routledge Handbook of Environmental Justice* (1st editio). New York, New York: Routledge.

Konisky, D. M., & Teodoro, M. P. (2016). When Governments Regulate Governments, *60*(3), 559–574. https://doi.org/10.1111/ajps.12221

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring Social Class in US Public Health Research: Concepts, Methodologies, and Guidelines. *Annual Review of Public Health*, *18*, 341–378. https://doi.org/10.1146/annurev.publhealth.18.1.341

Krivoruchko, K., Gribov, A., & Krause, E. (2011). Multivariate areal interpolation for continuous and count data. *Procedia Environmental Sciences*, *3*, 14–19. https://doi.org/10.1016/j.proenv.2011.02.004

Maantay, J., & Maroko, A. (2009). Mapping urban risk: Flood hazards, race, & environmental justice in New York. *Applied Geography*, *29*, 111–124. https://doi.org/10.1016/j.apgeog.2008.08.002

Morello-Frosch, R., Pastor, M., & Sadd, J. (2001). Environmental Justice and Southern California's "Riskscape": The Distribution of Air Toxics Exposures and Health Risks among Diverse Communities. *Urban Affairs Review*, *36*(4), 551–578. https://doi.org/10.1177/10780870122184993

National Research Council. (1997). *Safe Water From Every Tap: Improving Water Service to Small Communities*. Washington DC: National Academy Press. https://doi.org/10.17226/5291

NCSS Documentation. (2018). NCSS LLC. Retrieved from https://www.ncss.com/software/ncss/ncss-documentation/

Ogneva-Himmelberger, Y., & Huang, L. (2015). Spatial distribution of unconventional gas wells and human populations in the Marcellus Shale in the United States: Vulnerability analysis. *Applied Geography*, *60*, 165–174. https://doi.org/10.1016/j.apgeog.2015.03.011

Olson, E., & Fedinick, K. P. (2016). *What's in Your Water? Flint and Beyond: Analysis of EPA Data Reveals Widespread Lead Crisis Potentially Affecting Millions of Americans*. Retrieved from https://www.nrdc.org/resources/whats-your-water-flint-and-beyond

Pennsylvania Department of Environmental Protection. (2017). *Public Water Supplier's (PWS) Service Areas* [Shapefile]. Retrieved from http://data-padep-1.opendata.arcgis.com/datasets/fdf53cdeb2ec42b8a9422569f2e9531b_225?geometry=-90.618%2C38.409%2C-77.566%2C43.387

Pennsylvania Department of Environmental Protection. (2017). Public Water Supplier's PWS Service Areas, Full Metadata. Retrieved from http://www.pasda.psu.edu/uci/FullMetadataDisplay.aspx?file=PublicWaterSupply2017_07.xml

Switzer, D., & Teodoro, M. P. (2017). Class, Race, Ethnicity, and Justice in Safe Drinking Water Compliance. *Social Science Quarterly*, 1–12. https://doi.org/10.1111/ssqu.12397

The Environmental Justice Coalition for Water. (2005). *Thirsty for Justice: A People's Blueprint for California Water*.

U.S. Census Bureau. (2010). *Urban and Rural 2010 Census Summary File* [Population count spreadsheet]. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_SF1_P2&prodType=table.

U.S. Census Bureau (2016). *2011-2015 American Community Survey 5-Year Estimates* [Population count spreadsheets]. Retrieved from https://factfinder.census.gov/faces/nav/jsf/pages/download_center.xhtml.

U.S. Census Bureau. (2016). *Urban Areas* [Shapefile]. Retrieved from https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2016&layergroup=Urban+Areas.

U.S. Census Bureau. (2017). *TIGER State Level Geodatabase (Pennsylvania)* [Geodatabase]. Retrieved from https://www.census.gov/geo/maps-data/data/tiger-geodatabases.html .

U.S. Geological Survey (2011). *National Land Cover Database* [Raster]. Retrieved from https://www.mrlc.gov/nlcd11_leg.php.

U.S. EPA's Office of Enforcement and Compliance Assurance. (2013). *Providing Safe Drinking Water in America: 2013 National Public Water Systems*. Retrieved from https://www.epa.gov/sites/production/files/2015-06/documents/sdwacom2013.pdf

U.S. EPA. (2004). *Understanding the Safe Drinking Water Act.*

U.S. EPA. (2017a). Background on Drinking Water Standards in the Safe Drinking Water Act (SDWA). Retrieved February 8, 2017, from https://www.epa.gov/dwstandardsregulations/background-drinking-water-standards-safe-drinking-water-act-sdwa

U.S. EPA. (2017b). How EPA Regulates Contaminants. Retrieved August 20, 2017, from https://www.epa.gov/dwregdev/how-epa-regulates-drinking-water-contaminants

U.S. EPA. (2017c). Safe Drinking Water Act (SDWA) Resources and FAQ. Retrieved from https://echo.epa.gov/help/sdwa-faqs#Q5

U.S. EPA. (2018). Environmental Justice. Retrieved June 1, 2018, from https://www.epa.gov/environmentaljustice

U.S. EPA. (2018). *Safe Drinking Water Information System* [Violation data]. Retrieved from https://ofmpub.epa.gov/apex/sfdw/f?p=108:200.

Wallsten, S., & Kosec, K. (2008). The effects of ownership and benchmark competition: An empirical analysis of U.S. water systems. *International Journal of Industrial Organization*, *26*, 186–205. https://doi.org/10.1016/j.ijindorg.2006.11.001

Zandbergen, P. A., & Ignizio, D. A. (2010). Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. *Cartography and Geographic Information Science*, *37*(3), 199–214. https://doi.org/10.1559/152304010792194985