Environmental Justice Chemical Water Contamination in Arizona

ENVIRONMENTAL JUSTICE CHEMICAL WATER CONTAMINATION IN ARIZONA

by

Lina Benitez

A Thesis Submitted to the Faculty of the

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2023

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by: Lina Benitez
titled: Environmental Justice: Chemical Water Contamination in Arizona

and recommend that it be accepted as fulfilling the thesis requirement for the Master's Degree.

George Frisvold
George Frisvold (May 10, 2023 20:11 PDT)
George Frisvold

Date: May 10, 2023

Bonnie G Colby
Bonnie G Colby (May 11, 2023 06:41 PDT)
Bonnie G Colby

Date: May 11, 2023

Laura A. Bakkensen
Laura Bakkensen

Date: May 11, 2023

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement.

George Frisvold
George Frisvold (May 10, 2023 20:11 PDT)
George Frisvold
Thesis Committee Chair
Agricultural & Resource Economics

Date: May 10, 2023

Signature: Lina Benitez Segura (May 10, 2023 11:38 PDT)

Email: lmbs90@arizona.edu

Contents

Abstract

Environmental justice (EJ) concerns the disproportionate impact of environmental hazards and pollution on vulnerable and marginalized communities. Previous studies have identified that these marginalized communities include low-income populations, people of color, and other relegated groups who may have limited access to resources, political power, and decision-making processes. In this study, I focus on the association between socio-economic variables (race, poverty rate, age segments of the population such as children and elderly people, gender) and concentrations or violations of arsenic, lead, and copper in drinking water in Arizona. The results show that the principal chemical contaminant in drinking water in the state is arsenic, with an average concentration in drinking water samples collected between 2009 to 2022 of 5 ug/L, which is significantly higher than the U.S. average level of 2 ug/L. For arsenic, there is a disproportionate exposure among low-income groups, Hispanic communities, and tracts with a higher proportion of children. The study also supports the disproportionate lead exposure for black communities and copper exposure for American Indian communities.

Chapter 1: Introduction

Over the last decade and a half, the southwest area of the U.S has experienced significant drought conditions. These conditions have reduced water in the Colorado River, which was already overallocated to the seven states in its basin. Cuts in the water were imposed in 2022 and 2023 and Arizona will lose about one-fifth of its share (ASU News, 2022). According to the Arizona U.S Drought Monitor the conditions of drought are more extreme during the last two years compared to any other period since 2000 when the measurement started (Arizona Department of Water Resources, 2023). Although, with scarcity, chemical contaminants in water sources will be more concentrated. Poorer water quality may lead to higher labor or other treatment costs for municipalities, higher concentrations of contaminants in drinking water, or both.

Arizona relies on groundwater for 40 percent of its water supply where the concentration of contaminants such as arsenic and copper is higher compared to surface water (James, Evans, & Evans, 2020). The most common contaminant found in Arizona groundwater in a concentration above health-based drinking water standards is arsenic (Uhlman, Rock, & Artiola, 2009). Also, large copper mining deposits and extraction activities have implications for groundwater quality. Besides lead pipelines which are common in old households are the principal source of lead in drinking water (EPA, 2023). Long exposure to these contaminants in the population may cause diseases such as cancer, brain damage, and kidney chronic disease, among others (R, FA, DK, I, & Clarke E, 2015)

This study aims to explore the effect of socio-economic characteristics on the occurrence of violations of the standard for arsenic, copper, and lead in drinking water in Arizona and identify relevant socio-economic features of the communities with higher exposure to drinking water contaminants. It characterizes contaminant concentration levels and safe drinking water violations for Arizona community water systems. It then matches this information with socio-demographic characteristics of the population affected by the contaminants. Using conceptual frameworks from the environmental justice literature, it assesses characteristics of communities that have been overburdened by water pollution during the last 13 years.

In terms of analyzing water pollution for arsenic, lead, and copper, different challenges arise for each contaminant that has distinct implications for environmental justice studies. Generally, the purpose of this type of study is to bridge the gap in water pollution between population segments with different characteristics. However, an important challenge is the cost associated with the solution. In the case of arsenic contamination, the treatment is very expensive compared to lead and copper in drinking water. For Arizona, the presence of higher levels of arsenic in groundwater imposes a significant challenge for water systems. In rural areas, the water systems are relying heavily on groundwater. Unlike urban areas that typically have access to surface water from rivers, lakes, and reservoirs, rural areas in Arizona often have limited options for water supply and depend largely on underground aquifers. Groundwater in rural Arizona is often pumped from wells and aquifers, with some areas relying solely on groundwater for their water needs. However, excessive groundwater pumping can cause depletion of aquifers and lead to land subsidence, as well as water quality issues such as increased arsenic levels. In rural areas, where mostly small water systems use groundwater, the absence of water treatment could result in the subsequent closure of these systems due to excessive cost.

Environmental justice is 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 (EPA, 2023). Previous studies of environmental justice for air and water pollution have shown that low-income households and people of color have greater exposure to environmental hazards (Banzhaf, Ma, & Timmins, Environmental Justice: The Economics of Race, Place and Pollution, 2019). Environmental justice studies use variables associated with race, ethnicity, sex, age, income, and property rights linked to various, negative environmental exposures (Chakraborty, Rus, Henstra, & Minano, 2022) For example, one analysis of arsenic water contamination on public water systems in California showed that variables such as homeownership rate are associated with a lower level of contamination (Balazs, Morello, & Ray, 2012). Another, in Arizona, found low-income communities were at a disproportionate risk for arsenic contamination in drinking water (Rahman, 2009).

This study starts with a description of the study area, the data used in the analysis, and changes across space and time in measures of water contamination. Then it will review the literature on the relevant policies, water quality regulations, and the current contamination situation. Later, it explores conceptual models used in environmental justice studies. A discussion of the data and methods employed in this study follows. The next chapter details econometric model specifications and results of regression analyses. It concludes by presenting conclusions and policy implications.

1.1 Study Area

To assess the relationship between local household characteristics served by community water systems and violations of standards for arsenic, copper, and lead in drinking water, it is first necessary to identify sources of water quality information for at a local level. The Arizona Department of Environmental Quality (AZDWIS) has a publicly available safe drinking water database, with information on the concentration of chemical contaminants and violations of standards by water systems from 1990 to 2022. The socio-economic variables of that characterize residents served by these water systems come from U.S. Census Data or from the American Community Survey database.

1.2 Data and Methods

Water quality sample reports of public water supply systems (PWSS) are available on the AZDWIS website. Data include violations of the standards by individual PWSS and contaminant concentration levels. The PWSS has three different categories: Community Water systems (CWS), which are systems that supply water to the same population year-round; Non-Transient Non-Community Water Systems (NTNCWS), which are systems that regularly supplies water to at least 25 of the same people at least six months per year, and Transient Non-Community Water Systems (TNCWS) that provide water in a place where people do not remain for long periods

(EPA U. S., Information about Public Water Systems, 2022). Considering these categories, it is possible to identify the area of the population served by each CWS using information on samples that is collected on a yearly basis. The available public information of served population area is only accessible to CWS. Considering this constraint, the study includes CWS sample results.

Furthermore, the socio-demographic information is available at a state, county, census tract, and census block level in the U.S. Census Data and the American Community Survey (ACS). In this analysis, the data is evaluated at the census tract level. The publicly available socio-demographic data of interest for Arizona are accessible from 2009 to 2022.

The data from the CWS are panel data with information collected across multiple systems and years. Two principal approaches were used to measure negative environmental exposures: concentration levels of contaminants (continuous variables) and drinking water standard violations. The latter occurs if contaminant concentration levels exceed a threshold (a binary variable). Determinants of violations are examined using standard binary dependent variable models (e.g. logit models). For concentration levels data are lower-bound censored because of the detection limits of the devices used to measure those concentration levels. The study experimented with different approaches to address this censoring issue.

Chapter 2: Objective

The objective of this study is to estimate a regression model to assess which factors explain differences in the concentrations of arsenic, lead, and copper (and drinking water safety violations) in Arizona Community Water Systems (CWS) between 2009 to 2022. The research question is whether there are associations between the socio-economic characteristics of the populations served by these CWS and drinking water safety. Specifically, from an environmental justice perspective, the study seeks to examine whether historically overburdened and underserved communities face greater exposure to drinking contaminants.

Chapter 3: Literature Review

3.1 Federal Policy: Justice 40

In January 2021, President Biden made a historic commitment and signed the Executive Order 14008 to establish the Justice 40 initiative. This program sets a goal that 40 percent of the overall benefits of certain Federal investments flow to disadvantaged communities that are marginalized or underserved and overburdened by pollution. One of the categories of this program looks for the reduction of legacy pollution and the development of critical clean water and wastewater infrastructure (The White House, 2022).

Based on the Justice 40 initiative, the Environmental Protection Agency (EPA) includes a new strategic goal in its Strategic Plan focused on the advancement of environmental justice to protect human health and the environment for all people, with an emphasis on historically overburdened and underserved communities. In addition, the EPA developed programs such as Clean Water

State Revolving Fund, Reducing Lead in Drinking Water and Superfund to reduce chemical contamination in water source and drinking water (EPA, 2022).

3.2 Water quality regulations

3.2.1 Arsenic

In 2001, EPA stringent the standard for the concentration of arsenic in drinking water that applies to CWS and NTNCWS. The new standard was 10 parts per billion (ppb), which replaced the old standard of 50 ppb. This value protects consumers from the effects of long-term, chronic exposure to arsenic. PWSS must comply with the lower standard by January 2006 (EPA, 2022).

3.2.2 Lead and Copper

In 1991, EPA published the Lead and Copper Rule, a regulation to control lead and copper in drinking water. The regulation requires systems to monitor drinking water at consumer taps. If copper concentrations exceed an action level of 1.3 ppm or if lead concentrations exceed an action level of 15 ppb on more than 10% of consumer taps sampled, it is considered a violation (EPA, 2022).

In January 2021 the EPA established the Lead and Copper Rule revisions (LCRR) with a compliance date of October 16, 2024. Under this rule, the CWS and NTNCWS must identify all service lines from the main to the building, regardless of the usage of the water or the materials of these lines. PWSS serving more than 50,000 people must post information about lead service lines on a publicly available website (ADEQ, 2021).

In December 2021, Biden administration announced a plan to replace 100% of lead pipes in US homes. As a result, in 2024 a revised Lead and Copper rule is set to go into effect, which aims to strengthen the testing and treatment requirement for lead and copper in drinking water (EPA, 2023). The revised rule requires water systems to identify and replace all lead service lines in their distribution systems, requires corrosion treatment to be optimized, and lowers the action level for lead from 15 ppb to 12 ppb. The new rule also requires public disclosure of lead service line information and testing results (EPA, 2021).

3.3 Chemical Water Contamination in Arizona

Arizona relies on groundwater for 40 percent of its water supply (James, Evans, & Evans, 2020) where the concentration of contaminants such as arsenic and copper is higher compared to surface water.

3.3.1 Arsenic

The most common contaminant found in Arizona groundwater in a concentration above health-based drinking water standards is arsenic. There are three significant geologic sources of arsenic in Arizona: First, granite bedrock with valuable mineral ore often contains elevated concentrations of arsenic. Second, aquifers consisting of alluvium eroded from granite bedrock may also contain arsenic, and third layers of ancient sedimentary rock with high concentrations of arsenic and uranium in the Colorado Plateau, a northern region of Arizona (Kristine Uhlman, 2009).

A study of arsenic and nitrate concentrations developed by the U.S Geological Survey in the basin-fill aquifers of the Southwestern United States concluded that 43% of the area was predicted to equal or exceed the standard for arsenic. Much of the area is within a belt of basins along the western portion of the Basin and Range Physiographic Province that includes almost all of Nevada and parts of California and Arizona (U.S Geological Survey, 2012).

Contaminants produced by natural processes or for old pipelines such as arsenic, copper, and lead are relevant in the analysis of chemical contamination because the average concentration levels of these contaminants in public water systems in the southwest region are higher than the medium level in the United States. (Nigra A. E., et al., 2020). Long exposure to these contaminants in the population may result in diseases such as cancer, brain damage, and kidney chronic disease among others (EPA, 2022)

In Arizona, there are disparities across communities in access clean water. In the northeastern region of the state, the Hopi tribe has issues accessing clean water. The principal issue for this tribe is arsenic contamination. During the decade of 1960, U.S Government developed the infrastructure to guarantee water access for this community. However, in 2006, after the adjustment of the Arsenic standard from 50 ppb to 10 ppb, EPA identified that in the public water systems of this community, the levels of arsenic were higher than the limit. Afterward, the tribe leaders and EPA started to develop a project to find alternatives to water sources. However, the only solution was changing the water source to an area located on the north side of the tribal area (James I. , 2020). Recently, the Hopi community is building the infrastructure to guarantee access to clean water with appropriate arsenic concentrations through the Arsenic Mitigation Project, with the support of the federal government (Navajo Hopi Observer, 2022).

3.3.2 Lead and Copper

The sources of water contamination for copper and lead are copper and lead pipes, mining activity contamination, and farming or industrial pollution. In the case of Arizona, there are two principal sources: mining activities and pipelines.

First, consider mining activity. Arizona has a long history of metal mining and smelting. Starting around 1850, frontier explorers mined for copper, gold, and lead throughout Arizona. The establishment of these mines before modern mining practices and environmental regulations has impacted the environment and public health. These sites are no longer in operation, but they are affecting the water sources around the site (ADEQ, Surface Water Quality Remediation Sites, 2023).

ADEQ has developed a remediation process in these sites that are located near Prescott, Miami, and Nogales, Arizona. The most harmful effects of these old mining sites are contamination in water sources by copper, lead, and zinc. The risk for copper and lead contamination in drinking water from these sources could be limited because the PWS uses treatment methods to reduce the concentration in residential use (ADEQ, 2023). Nevertheless, the relationship between the proximity of CWS to mining sites and lead and copper contamination is examined in later regression models.

Second, consider contamination by old pipes from copper and lead. Lead can enter drinking water when plumbing materials that contain lead corrode. In homes with lead pipes that connect the home to the water main, these pipes are typically the most common source of lead in drinking water. Lead pipes are more likely to be found in older cities and homes built before 1986 (EPA, 2023). Arizona has an estimated 12,000 lead pipes currently in service (Knappenberger, 2022). Comparing the number of identified pipelines of these materials with other states, Arizona is at the lowest level of risk with less than 16,000 pipelines. Nevertheless, Phoenix, Tucson, Prescott, and Tempe have instituted projects to replace old pipelines with new ones in response to potential health risks. Finally, with the new regulation of LCRR, an updated inventory of lead and copper pipelines will appear in December of this year (2023), to help programs replacing pipelines.

3.4 Interrelations between ethnicity, income, poverty, home ownership and age of home.

Research has shown that there are interrelationships between ethnicity, income, poverty, home ownership, and age of home. First, some reports have shown that ethnic minorities are more likely to have lower rates of home ownership compared to white households. A report of the U.S Department of the Treasury using information of the Census Bureau that for the homeownership rate in U.S for white communities was 75% compared to 45% for black households, 48% for Hispanic households and 57% for non-Hispanic households of any other race. In general, these disparities have changed little over the last three decades (U.S Department of the Treasury, 2022). These disparities in home ownership are often linked to differences in income, as well as historical discrimination and exclusion from the housing market (Bayer & Wilcox, 2019).

Additionally, research has also shown that the age of a home is closely linked to its value and the socio-economic characteristics of its occupants. Older homes are often located in low-income and minority neighborhoods and are associated with a higher prevalence of health hazards, such as lead paint and poor air quality. These homes may also require more maintenance and repairs, which can be a financial burden for low-income homeowners (Office of Policy Development and Research , 2005).

Furthermore, home ownership has been found to be a key factor in reducing poverty and increasing wealth for households. However, low-income households are often excluded from the home ownership market due to credit constraints and lack of financial resources. This exclusion can perpetuate poverty and contribute to a widening wealth gap between ethnic groups. (Federal Reserve Bank, 2020)

Overall, the interrelationships between ethnicity, income, poverty, home ownership, and age of home are complex and multifaceted. Understanding these relationships is critical for developing policies and interventions aimed at reducing poverty and promoting equitable access to housing for all individuals, regardless of their socio-economic status or ethnic background.

3.5 Economics of Environmental Justice

The roots of the environmental justice studies established correlations between exposure to pollution, race, and poverty in the 1980's. Banzhaf, Ma, and Timmins (2019) developed a review of the environmental justice literature where it intersects with work by economists. They identify four principal potential causal mechanisms of environmental justice: disproportionate sitting, coming to the Nuisance, Coasean bargaining and political economy and government.

First, disproportionate sitting at the beginning was the firm taste-based discrimination by incorporating into their decision-making a preference for protecting whites from pollution (Becker, 1957). The approach is that firms might site their polluting activity based on local economic conditions which are correlated in space with residential demographic patterns and the government agencies make decisions that affect the location of such facilities. Wolverton (2009) modeling firm location as a decision variable found that a disproportionate share of people of color exposed to pollutants seems to arise more from economic factors such as land, labor, and access to transportation than directly from local demographics.

An alternate approach suggests that the households chose where to live based on their willingness (and ability) to pay for a clean environment. Housing location choice depends both on their preferences and budget. As a result, household sorts by income across levels of services, a process known as "stratification" (Banzhaf and Walsh, 2013). The categorization process may result in additional consequences on neighborhood characteristics as well as impacts that could reinforce the initial sorting patterns. The relations between environmental quality and demographics, environmental quality with housing prices, high-income households with a clean environment, neighborhood amenities, and environmental quality, and high-income households and amenities explain not only the sorting process but also create a self-reinforcing multiplier effect.

Firms have preferences over the willingness to pay to locate at a certain place and where to locate their industrial facilities. In addition, households have a tolerance for pollution and a willingness to accept it as compensation for industrial activity nearby. The Coase theorem holds that there is a process of negotiation and market transactions between firms and households that ensures the efficient use of resources. However, the distribution of payments in this structure depends on the initial allocation of property. In this approach, the amount of the payment the firm needs to make will be determined between the two parties according to their relative bargaining power (Hamilton, 1995).

The final consideration is the effect of government and political economy on the distribution of pollution across the population. Governments can affect the distribution of pollution through enforcement patterns, legislation, and bureaucratic monitoring. Regulators used technical and polluting factors to allocate policy tools prioritizing remediation and regulation depending on resource and time constraints. However, households with the highest willingness to pay to avoid pollution may influence the government and regulators to guarantee better conditions for them. There is evidence that regulatory actions are at least correlated with the political power of local communities and it highlights the relevant connection between Coasean processes and political economy, both are tied up in property rights and the enforcement of those rights. (Banzhaf, Ma, & Timmins, 2019)

In addition, the review of environmental justice (Banzhaf, Ma, & Timmins, 2019) described the Intergenerational Effects of Pollution as a relevant aspect of environmental justice. Exposure to pollutants has deleterious long-term effects on human capital and earning potential, meaning that pollution exposure extends poverty to the next generation . Economists study these effects in two steps. First, pollution affects birth outcomes. Second, birth outcomes impact later-life outcomes. Studies have found a connection between pollution exposure affecting birth outcomes such as birth weight, gestation length, and infant mortality, and others demonstrate that neonatal health can impact later socioeconomic status (fetal origins hypothesis) (Almond & Currie 2011, Almond et al. 2017). Other studies connect childhood exposure to later life outcomes (Currie 2011, Voorheis 2017).

3.6 Environmental Justice Studies in drinking water pollution in the U.S

In the last section, there is a description of the relationship between socio-economic and racial/ethnic disparities associated with pollution. In this section, there is a review of the principal studies of environmental justice that accounted for disparities in drinking water pollution concentrations in vulnerable populations in the United States.

Drinking water contaminant exposure reflects structural national and regional inequality in the U.S. Previous studies have identified inequalities in public drinking water exposure at a national level using the information on the concentration level of the contaminants. For arsenic and uranium Morata, et al. (2022) associated CWS concentration data between 2000 and 2011 at a county level across the US. with a higher exposure among Hispanic and American Indian populations. Also, Nigra, Cazacu-Deluca, & Navas-Acien, 2022 for arsenic identified using a database of concentrations in CWS at a county level across US between 2006 and 2011, greater exposure in communities with a lower percentage of adults with high school education and Nigra A. E., et al., 2020 found elevated arsenic exposure in CWS at a county level that served small populations, reliant on groundwater and served Hispanic communities. For nitrate Schaider, Swetschinski, Campbell, & Rudel, 2019 identified that the percentage of Hispanic residents served by each CWS was significantly associated with a higher nitrate concentrations, this study was developed at a county level across US. Some evidence indicates inequalities at a regional level. In California, a census block-level study concluded that higher CWS arsenic concentrations were associated with a higher proportion of Hispanic residents (Pace, et al., 2021).

In Arizona, a zip-code-level study identify a negative significant associations of the percentage of black people with arsenic violation, while for hispanics it identified a positive coefficients associated with the exposure to arsenic in drinking water. However the analysis did not identified any relationship between arsenic contamination and low-income groups (Cory & Rahman, 2009).

In addition, preliminary studies analyzed the inequality in drinking water pollution using the data of violations of drinking water quality standards. McDonald & Jones, 2018 found in a study of violations for all the accounted contaminants in the U.S. using data between 2011 to 2015 that the population served by CWS with a higher number of violations and repeated violations were associated with communities of minorities, low-income households, and uninsured households. For arsenic, Foster, Pennino, Compton, Leibowitz, & Kile, 2019 identified for CWS data between 2008 and 2017 a decreasing tendency in the percentage of CWS with violations of arsenic standards starting in 2008 with 1.3% of active CWS and finishing in 2017 with 0.55%. A summary of the previous studies in water pollution on environmental justice in the US is available in Table 1.

Table 1. Summary of the principals studies

Authors	Geographic specificity	Geographic Scope	Years Covered	Dependent Variable	Significant Associations
McDonald, Yolanda J, PhD; Jones, Nicole E	County	National	2011-2015	Drinking water violations	Minorities, low-income households, and uninsured households
Morata, et al. (2022)	County	National	2000-2011	Arsenic and Uranium concentration	Hispanics and American Indians
Nigra A. E., et al., 2020	County	National	2006-2011	Arsenic concentration	Percentage of adults with highschool education
(Cory & Rahman, 2009)	Zip codes	State	2002-2004	Arsenic concentration	
(Schaider, Swetschinski, Campbell, & Rudel, 2019)	City and County	National	2010-2014	Nitrate concentration	Percentage of hispanics
(Pace, et al., 2021)	CWS	State	2011-2019	Contaminants concentrations	People of color

This study gives another perspective to the EJ concerns about arsenic exposure in the population in Arizona. In next sections we describe the identified associations between low income communities and racial/ethnic disparities with arsenic exposure. In addition, in this analysis we included additional variables related to characteristics of sex and age to identify additional disparities not only related with race and income in vulnerable segments of the population.

In addition, this analysis includes lead and copper as contaminants of interest given the national interest that Biden administration is giving to eliminate any source of lead in drinking water, and the revised version of the lead and coppe rule that is set to go into effect in 2024. This study, is the first that analyze lead and copper exposure in Arizona to analyze if there are any EJ concerns associated to disparities of low income or any racial disparities in the population.

3.7 Empirical Models

Previous studies identified significant relationships between concentrations of contaminants and socioeconomic characteristics using linear regression models, such as panel data models (Nigra A. E., et al., 2020), mixed effect regression models (Schaider, Swetschinski, Campbell, & Rudel, 2019), and quantile regression model (Nigra A. E., et al., 2020). In addition, other studies applied spatial econometric models such as spatial lag regression (Nigra, Cazacu-Deluca, & Navas-Acien, 2022) and geographically weighted regression (Morata, et al., 2022). In the analysis of violations of standards, prior studies used logistic regression models as the empirical model (McDonald & Jones, 2018) (Cory & Rahman, 2009).

The following sections describe the principal econometric models used in the development of this study: the Tobit model, binary dependent variable model, and panel data model.

3.7.1 Tobit Models

Tobit models refer to regression models in which the range of the dependent variable is constrained in some way. This model was first suggested by Tobin (1958) which analyzed the relationship between income and expenditure and limited the dependent variable (expenditure) to only positive values. These models are known as censored or truncated, censored is when the limit observations are in the data and truncation is when the observations are not in the sample (Amemiya, 1984).

The censored sample is representative of the population because all observations are included. The model can be censored from below or above. In this study, the data is censored from below, the dependent variable (y) is observed if the latent variable (y^*) is above the limit (L) and the limit is observed for the censored observations.

$$y = \begin{cases} y^* \ if \ y^* > L \\ L \ if \ y^* \leq L \end{cases}$$

The Tobit model assumes that the latent dependent variable is normally distributed, and the limit is equal to L. The model is defined as follows:

$$y_i{}^* = x\beta + \varepsilon$$

$$y_i{}^* \sim N(\mu^*, \sigma^2)$$

$$\alpha = (L - \mu^*)/\sigma$$

Using the maximum likelihood method, Tobin (1958) established the probabilities for censored normal data, the following version adjusted the standard model where the limit was zero to a below-limit model where the threshold is L:

$$Prob(y = L|x) = Prob(y^* \leq L|x) = Prob[(y^* - \mu^*)/\sigma \leq (L - \mu^*)/\sigma|x]$$
$$= Prob[z \leq (L - \mu^*)/\sigma|x] = \phi(\alpha)$$

$$Prob(y > L|x) = Prob(y^* > 0|x) = 1 - \phi(\alpha)$$

Now, to find the expected values of this model in a censored normal distribution the equation when the dependent variable is higher than the limit (L) is the following:

$$E[y|y^* > L] = \mu^* + \sigma\lambda(\alpha)$$

The ratio between the probability density function (PDF) and the Cumulative distribution function (CDF) is called the inverse mill's ratio $\lambda(.)$. The equation is defined as follows:

$$\lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)}$$

Considering that the dependent variable is censored, the marginal effects are different than the OLS model that is applied over the latent variable (y^*). The marginal effect equation in the Tobit model is defined as follows:

$$\frac{\partial E(y|x)}{\partial x_k} = \beta_k \Phi\left(\frac{x'\beta}{\sigma}\right)$$

3.7.2 Binary-dependent Variable Model

In the analysis of violations of Arsenic and Lead and Copper Rules, which are binary random variables that take only values zero and one (zero: non-violation of the standard, one: violation of the standard), there are two principal econometric models which use a binary random dependent variable: probit and logit. The probit model assumption is normal distribution in the error, on the other side in the logit model the assumption is that the error is logistically distributed.

Binary regression techniques allow us to estimate the effects of the explanatory variables (Xs) on the underlying dependent variable (Y*) (Williams & Jorgensen, 2023). They can also be used to see how the independent variables affect the probability of being in one category of the observed Y as opposed to another. The latent variable in binary regressions can be written as follows:

$$y_i = \begin{cases} 1 & y_i^* > 0 \\ 0 & y_i^* \leq 0 \end{cases}$$

$$y_i{}^* = x\beta + \varepsilon$$

The probability of the outcome occurrence for the two possible responses changed depending on the model used. The equations for these probabilities for logit and probit models are the following:

Probit

$$Prob(y_i = 0) = \Phi(-\beta' x_i)$$

$$Prob(y_i = 1) = \Phi(\beta' x_i)$$

Logit

$$Prob(y_i = 0) = \frac{e^{-\beta x_i}}{1 + e^{-\beta x_i}}$$

$$Prob(y_i = 1) = \frac{1}{1 + e^{-\beta x_i}}$$

To estimate the effect of changes in one explanatory variable in the occurrence of the dependent variable $(y_i = 1)$ it is relevant to estimate the marginal effects in probit and logit model. The estimation is calculated according to the following equations:

Probit

$$\frac{\partial prob(y_i = 1)}{\partial x_i} = \phi(\beta' x_i) * \beta_i$$

Logit

$$\frac{\partial prob(y_i = 1)}{\partial x_i} = \frac{e^{-\beta\prime x_i}}{1 + e^{-\beta\prime x_i}{}^2} * \beta_i$$

3.7.3 Panel Data Models

Panel data models considered a regression which varies across individuals and time. However, under this scenario it is not possible to use a regular OLS model because there is a presence of heterogeneity. One approach to solve this problem is to use dummy variable per individual (Least Squares Dummy Variable Regression). However, if the model has any variable which does not vary in time, then there will be a problem of multicollinearity. The best approach to control individual heterogeneity is use panel data models.

A panel data regression differs from a regular cross sectional or time series regression because it has double subscripts on its variables with i denoting individuals and t denoting time (Baltagi, 2005)

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}$$

Most of the panel data application utilize a one-way error component variable for the disturbances with the following equation, where μ_i denotes the unobservable individual-specific effect and v_{it} denotes the remainder disturbance. There are two principal approaches of panel data model: Fixed Effects Model (FE) and Random Effects Model (RE).

$$u_{it} = \mu_i + v_{it}$$

In the fixed effects model μ_i is assumed to be fixed parameters to be estimated and reminder disturbances stochastic with v_{it} for all i and t. Since α_i is not observable, the Fixed Effects Model eliminates α_i by de-meaning the variables using the within transformation:

$$(y_{it} - \bar{y}_i) = (\alpha_i - \bar{\alpha}_i) + (x'_{it} - \bar{x}_i)\beta + (u_{it} - \bar{u}_i) ==> \qquad \widetilde{y_{it}} = \widetilde{x_{it}}\beta + \widetilde{u_{it}}$$

Another alternative is the first difference transformation, which use the first difference as the subtracted value.

The random effects model is a model to analyze data that takes into account the fact that some factors affecting the outcome may vary randomly across individuals or groups. This model implies a homoscedastic variance for all individuals i in time t (σ_μ^2) and an equi-correlated block-diagonal covariance matrix which exhibits serial correlation over time only between the disturbances of the same individual (σ_v^2) and zero otherwise. Under this approximation, it is essential to adjust the OLS model because the standard errors are correlated over time. The following transformation is proposed to have uncorrelated standard errors:

$$(y_{it} - \theta\bar{y}_i) = (\alpha_i - \bar{\alpha_\imath}) + (x'_{it} - \theta\bar{x_\imath})\beta + (\mu_i - \theta\bar{\mu_\imath}) + (v_{it} - \theta\bar{v_\imath})$$

The equation to estimate the parameter θ is the following calculation. If the value is equal to zero, the model corresponds to pooled OLS, and if the value is equal to 1 it corresponds to the Fixed Effects within estimator.

$$\hat{\theta} = 1 - \sqrt{\frac{\sigma_V^2}{\sigma_V^2 + T\sigma_\mu^2}}$$

To check the best approximation in an analysis between fixed and random effects panel data model, the Hausman test is used to decide whether to use the fixed effects (FE) or random effects (RE) estimators. This test used the difference in RE and FE coefficients and their covariances. The null hypothesis of this test assumes no correlation between individual specific effects and independent variables, in this case FE and RE coefficients are not significantly different other and in case the null hypothesis is not rejected RE is a better approximation. The equation to calculate the test statistic is the following:

$$W = (\beta_{RE} - \beta_{FE})'(var(\beta_{RE}) - var(\beta_{FE})^{-1}(\beta_{RE} - \beta_{FE}) \sim x_2$$

Chapter 4: Data

4.1 Census Data

The Census Bureau's mission is to serve as the United States provider of quality data about the US people and economy. Every ten years the US Census attempts to provide a count of every resident in the United States, the last two Censuses were implemented for 2010 and 2020. In addition, the US Census Bureau conducts the American Community Survey (ACS), which is a nationwide survey designed to provide communities with reliable and timely social, economic, housing, and demographic data every year. The ACS has 1-year estimates, which are data that have been collected over 12 months and are available for geographic areas with at least 65,000 people and 5-year estimates represent data collected over 60 months (U.S Census Bureau, 2020). This study includes information from 2009 to 2022 of the US Census (2010 and 2020), ACS 5-years (2015), and ACS 1-year (2009, 2011, 2012, 2013, 2014, 2016, 2018, 2019, 2021.

Another significant source of information is the unit level of analysis. The Census Bureau has a standard hierarchy of Census Geographic entities starting with the national level, later regions, divisions, states, counties, tracts, and blocks. In this study, the subdivision used is the Census Tract, because it is the lowest population unit with available data for the analyzed period. Census Tracts generally have a population size between 1,200 and 8,000 people with an average size of 4,000 people. (U.S Census Bureau, 2022)

The variables of interest in a study of environmental justice are the socio-economic characteristics of the population in the relevant area. For instance, as explained in the literature review section, this study analyzes variables such as race, ethnicity, age, and gender of the population to identify any gaps in the contaminant exposure, and on the economic side, there are variables such as income per household and poverty rate. Besides, considering the government and political economy connection in environmental justice issues, a proxy of the economic status and economic power of a particular region is measured using the home ownership rate per unit of analysis.

The variables used for race and ethnicity are the percent of Hispanic, non-Hispanic White, non-Hispanic Native American, and non-Hispanic Black people. The Hispanic ratio is estimated using all the population of Hispanics over the total population. White, native American, and Black ratios are estimated using the population of the race that is not Hispanic over the total population. Figure 1 includes information on the Hispanic, White, Native American, and Black mean percentage of the population in Arizona during the relevant period.

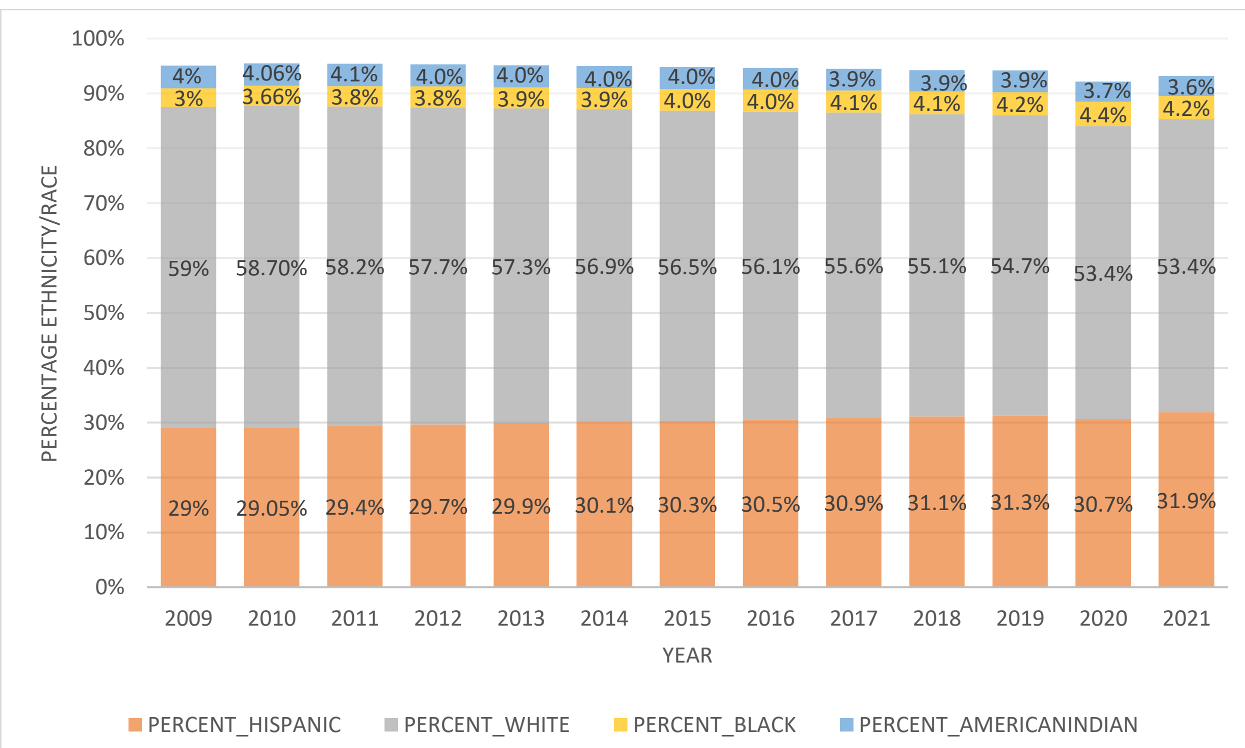


Figure 1. Percentage of population by Ethnicity/Race 2009-2021

The highest percentage of the Arizona population is White with a decreasing trend in the analyzed period changing from 58.70% in 2009 to 53.4% in 2021. In contrast, the Hispanic percentage has been increasing from 29% in 2009 to 31.9% in 2021. For the Black and American Indian communities, the percentages are stable with an average percentage of 4%.

The ethnicity and race changed significantly depending on the region. Table 2 contains the distribution of the population in the Decennial Census 2020 per county. In the counties of Cochise, Coconino, Gila, Graham, La Paz, Maricopa, Mohave, Pima, Pinal, and Yavapai the greatest percentage of the population is White. In Yuma, Greenlee, and Santa Cruz the highest proportion of the populace is Hispanic. In Apache County, 70% of the population is American Indian, while in Navajo County, American Indians are 44% of the population.

Table 2. Percentage of population by ethnicity/race 2020

County	PERCENT HISPANIC	PERCENT WHITE	PERCENT BLACK	PERCENT AMERICAN INDIAN
APACHE	6%	21%	0%	70%
COCHISE	34%	54%	3%	1%
COCONINO	15%	53%	1%	24%
GILA	17%	61%	0%	16%
GRAHAM	30%	53%	1%	13%
GREENLEE	46%	46%	1%	3%
LA PAZ	25%	55%	1%	14%
MARICOPA	31%	53%	6%	2%
MOHAVE	16%	75%	1%	2%
NAVAJO	10%	42%	1%	44%
PIMA	36%	51%	3%	2%
PINAL	29%	56%	5%	4%
SANTA CRUZ	83%	15%	0%	0%
YAVAPAI	15%	78%	1%	1%
YUMA	64%	30%	2%	1%

The relevant variables of age and sex account the population with historically marginal conditions. In this study, we consider the percentage of elderly people (age>64 years) and children younger than five years old are relevant variables because contaminants such as arsenic and copper in the long-term can cause chronic diseases, and these segments of the population are at higher risk (WHO, 2023). On the other side, variables such as the percentage of female-led households are included to analyze any gap associated with gender in exposure to the contaminants. The mean values of these variables for the analyzed period are comprised in Figure 2. In previous studies, Alvarez, C & Evans,C (2021), used percent of female headed-households as a marker of deprivation (Mosher 2001 & Smith 2007) and as an independent predictor of risk of exposure to environmental toxins (Lievanos 2019).

Regarding the age of individuals in Arizona, the population is aging. The proportion of individuals more than 64 years old was 12.9% in 2009, this proportion increased in Arizona during the analyzed period until 19.7% in 2021, while the share of children decreased from 5.9% to 5.6%. In addition, female-headed households increased by 3.2 percentage points over the study period from 9.2% to 12.5% of the population.

The homeownership rate is calculated as the ratio between owner occupied households over occupied households. Previous studies identified this variable as a good proxy of economic/political power (Balazs, Morello, & Ray, 2012). Figure 3 shows the trend of this variable in Arizona, the rate declined between 2009 to 2016 from 68.3% to 62.4%, then it rises until 2021 with 66.7%.

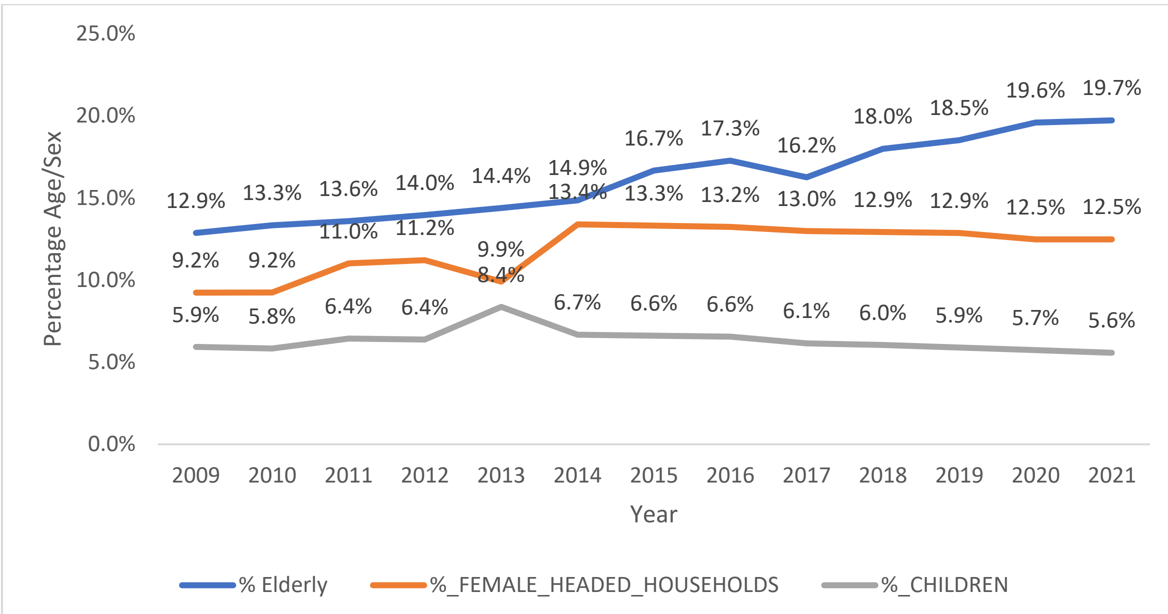


Figure 2. Percentage of population by Sex/Age 2009-2022

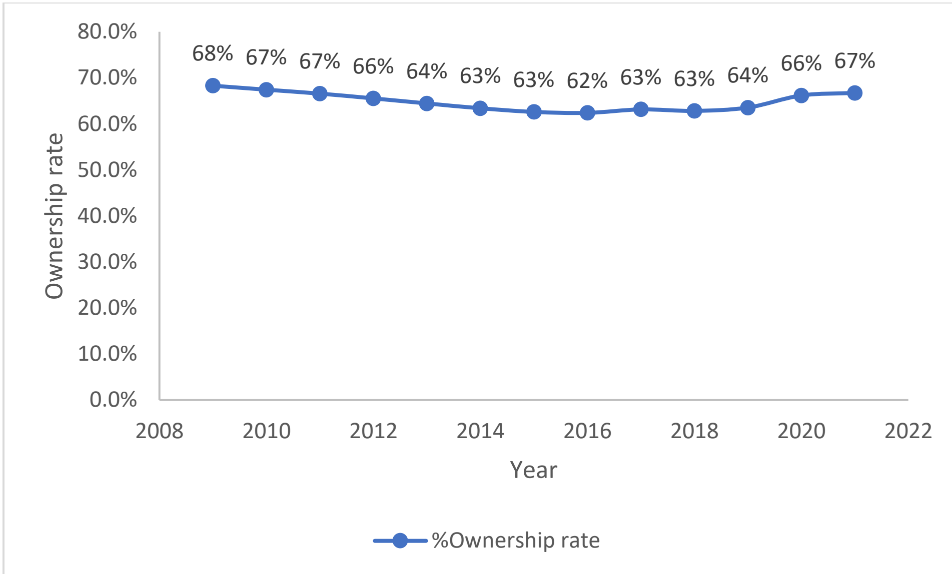
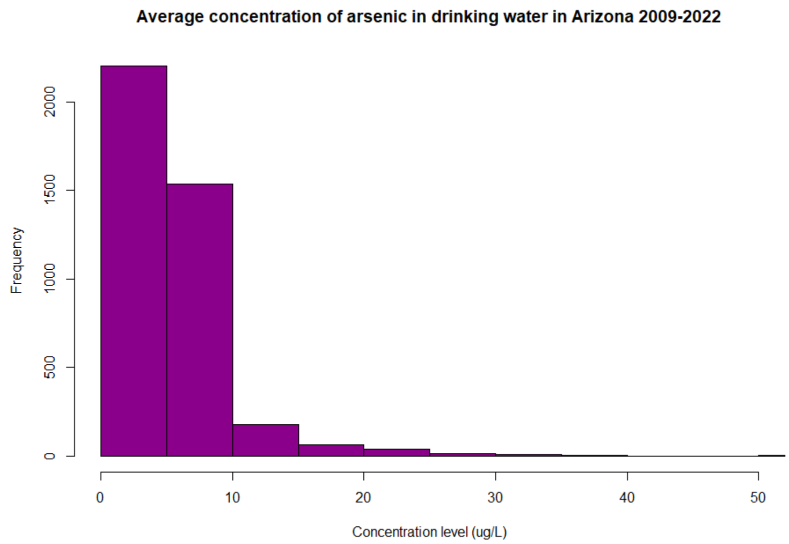


Figure 3. Home Ownership Rate 2009-2022

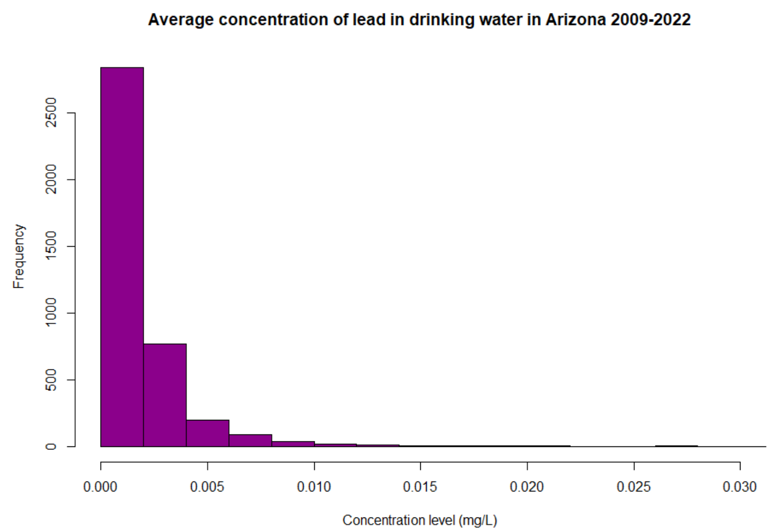
4.2 Water quality variables

More than 6 million people in Arizona get drinking water from a controlled public water system regulated by the Arizona Department of Environmental Quality (ADEQ). To guarantee the quality of drinking water and meet the state and federal Safe Drinking Water Act standards. ADEQ, with the help of local counties, assesses the drinking water source and supervises regulations that govern water system design, operation, and construction. ADEQ helps measure drinking water quality through required scheduled tests of all public water systems for a wide variety of potential contaminants. ADEQ has a public database that provides information related to regulated public water systems in Arizona, the Safe Drinking Information System (SDWIS). (ADEQ, 2023)

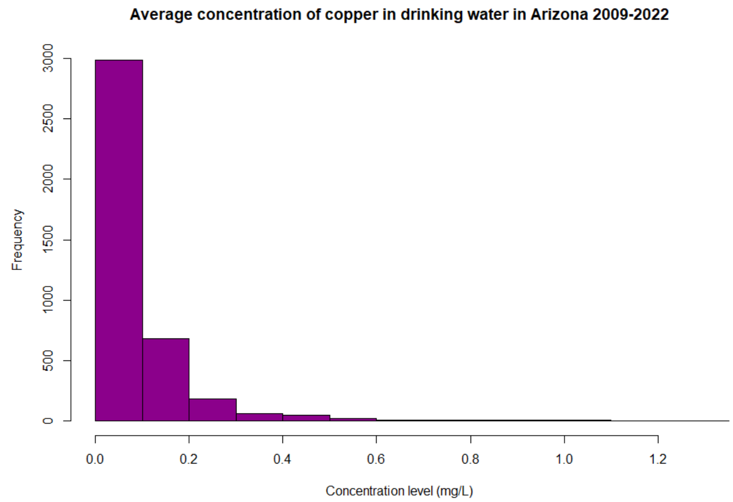
The SDWIS includes information on tap water samples from public water systems in the state of Arizona since 1983 in the sample results section. Besides, it contains the records of the violations of the drinking water standards for the PWS. For this study, the information of interest is the sample results for Arsenic, Copper, and Lead and the violations of the Lead and Copper Rule and Arsenic Standard. The distributions of the concentrations in the analyzed database which includes average water levels per year per CWS from 2009 and 2022, the histograms are in Figure 4. For Arsenic almost 95% of the samples are below 10 ug/L (limit of arsenic standard), in the case of lead most of the observations are below 0.015 mg/L (limit of lead rule), and for copper, there are just 3 observations in the whole sample above the limit for copper standard (1.3 mg/L)


Average concentration of arsenic in drinking water in Arizona 2009-2022
Frequency
0
500
1000
1500
2000
0
10
20
30
40
50
Concentration level (ug/L)

(a)


Average concentration of lead in drinking water in Arizona 2009-2022
Frequency
0
500
1000
1500
2000
2500
0.000
0.005
0.010
0.015
0.020
0.025
0.030
Concentration level (mg/L)

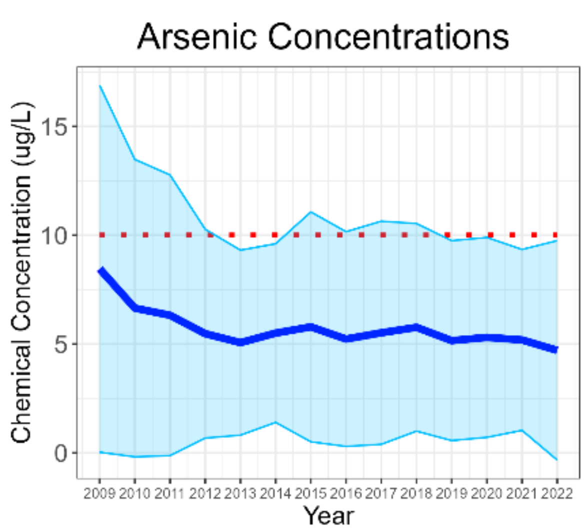
(b)



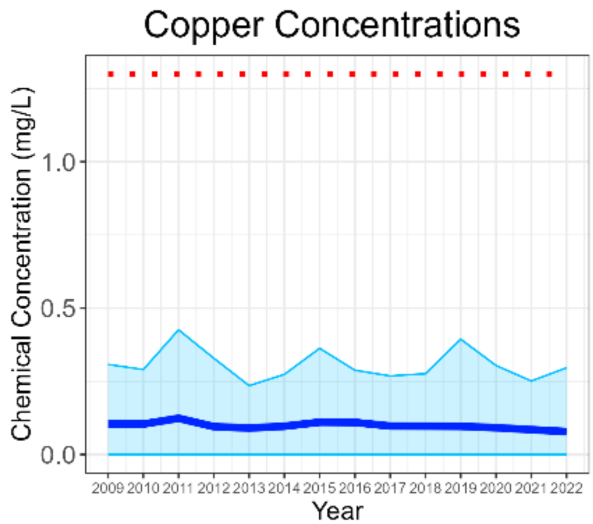
(c)

Figure 4. Histograms of contaminants concentration levels

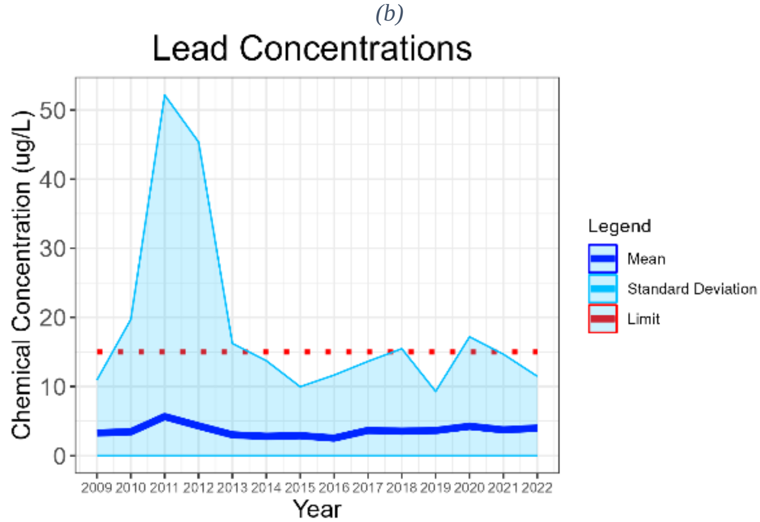
The concentration of contaminants is measured in micrograms per liter (UG/L) for Arsenic and Lead and milligrams per liter (MG/L) for copper. For measurement purposes the minimum level of concentration is the level of detection, this level varies depending on the contaminant and the CWS. Figure 3 contains the average concentration and the interval across one standard deviation of a contaminant in the analyzed period for the CWS linked with this study and the limit of the standard. In Figure 5 (a) the tendency of arsenic concentration is declining with a maximum mean value adjoining the limit of the Arsenic standard in 2009 and the lowest mean level last year at 4.70 UG/L. In Figure 5 (b) the concentrations of copper are steady between 2009 and 2022 with an average rate of 0.1 mg/L, some years have higher deviation for specific extreme measurements, these outcomes are distant from the Copper Rule standard of 1.3 mg/L, this chemical does not represent an imminent risk of exposure for residents in Arizona. Besides, in Figure 5(c) contamination for lead in drinking water samples has a stable trend with an average value of around 4 ug/L and some extreme observations in 2011 and 2012.



(a)



(b)



(c)

Figure 5. Home Ownership Rate 2009-2022 a) Arsenic, b) Copper and c) Lead.

The analysis of violations includes the entries that contain measurements for concentrations per each CWS. In Figure 6 there is a summary of the percentage of CWS with violations per rule for the analyzed period. The violations are grouped into two standards: The arsenic standard and the Lead and Copper Rule. There is a decreasing trend in arsenic violations which supports the good results of initiatives such as Source Water Protection developed by the ADEQ to reduce exposure to this pollutant (ADEQ, 2020). In contrast, the number of violations for Lead and Copper increased from 2009 to 2017 and then decreased in the period between 2017 and 2022.

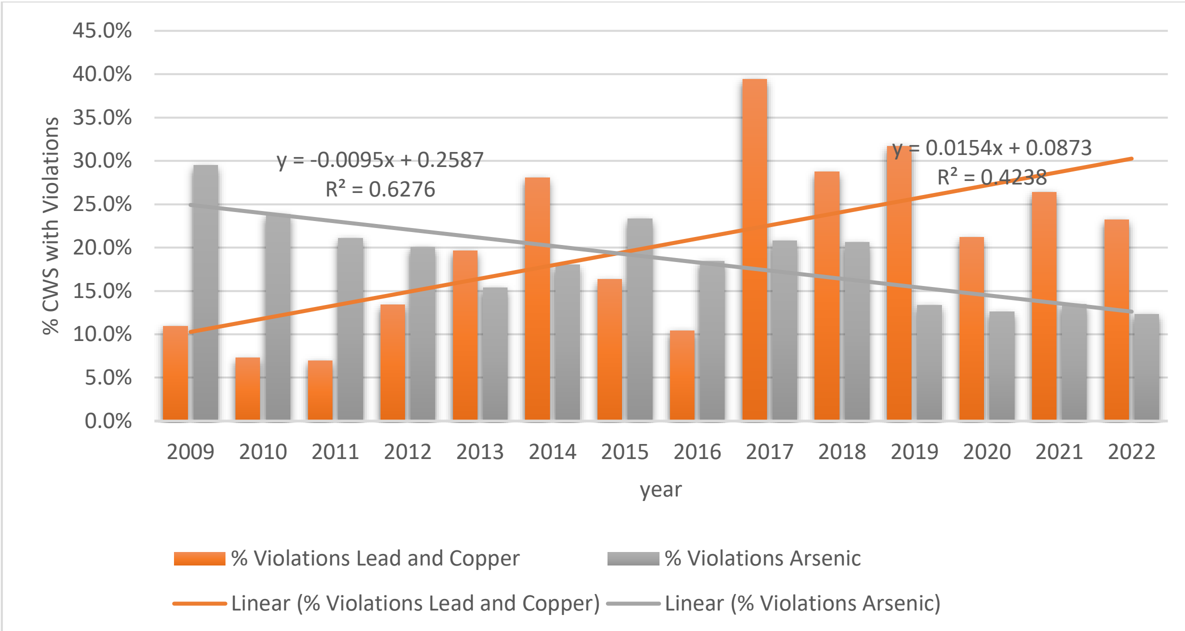


Figure 6. Percentage of CWS with violations per rule

4.3 Economic variables

Income and poverty level are relevant economic characteristics in environmental justice studies to evaluate the relationship between income level and water pollution. The poverty level accounts for individuals or families whose collective income in the preceding 12 months was below the national poverty level of the United States. In 2021, about 12.8% of Arizona's population lived below the poverty level, however, this variable has been improving in the last decade considering that the level of poverty in 2012 was 18.7%.

According to the 2021 American Community Survey, analyzing the level of poverty at a county level, Maricopa, Pina, and Pinal have the lowest percentage of poverty (12%), which are the most populated counties with a total population of 5,822,874, accounting for 82% of the total state population. On the other hand, Graham, Santa Cruz, and Yuma are the counties with the highest level of poverty of around 18%. In general, the level of poverty has been decreasing. Figure 7 shows the percentage of poverty per county for 2012, 2016, and 2021.

In 2021, on average the individuals in Arizona experienced a higher poverty rate than the national proportion (11.4%). The poverty rate for the American Indian communities of Arizona exceeds the state average by a significant amount with 29.6% in 2021. The white non-Hispanic population posted the lowest poverty rate in Arizona with 8.8% while for Hispanic households it was 17.2%, and for black households it was 16.7%. These disparities in poverty rates are linked to historical and systemic factors, such as discrimination, lack of access to education and job opportunities and housing segregation. This intersection of poverty and race/ethnicity can also contribute to environmental injustice, as low-income communities of color are often disproportionately affected by pollution.

The trend in the poverty rate was decreasing across race and ethnicity, greater gaps were presented in the Hispanic community changing from 28.8% in 2012 to 17.2% in 2021 and the American Indian population starting at 38.5% to 29.6%. Figure 8 displays poverty rate levels across races and ethnicities in 2012 and 2021.

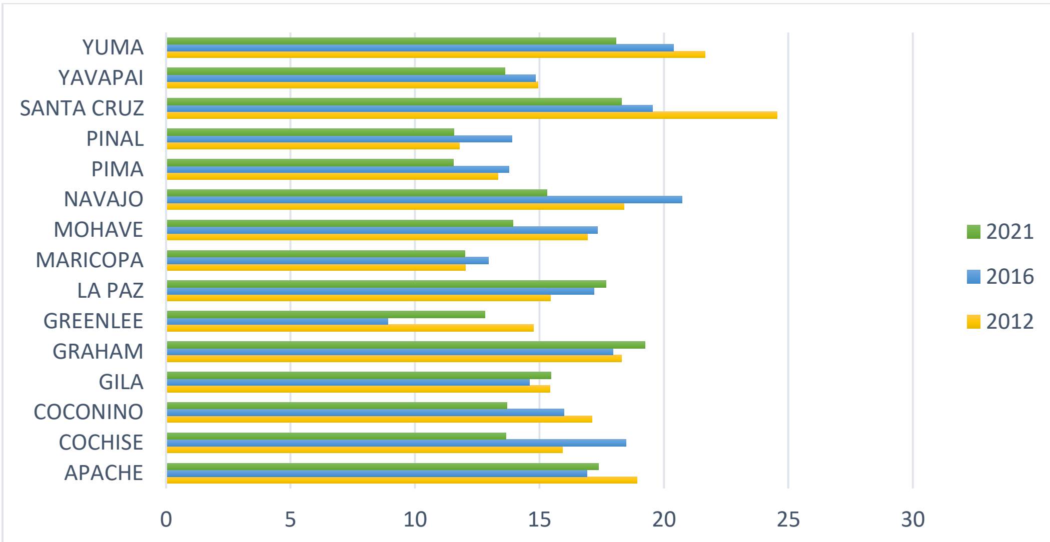


Figure 7. Poverty Rate per county 2012, 2016 and 2021

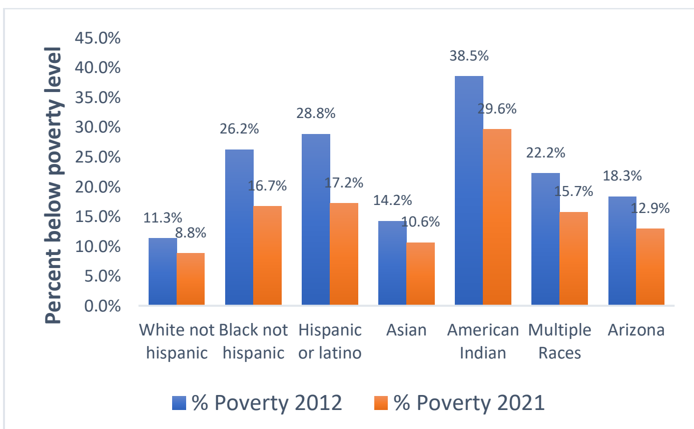


Figure 8. Percent of population below poverty per race/ethnicity 2012 and 2021

Evaluating the relation between age and poverty, the distribution of poverty rate across different age ranges including children, individuals under 18 years, adults, and elderly people. There is an indirect relation between age and poverty rate in Arizona. Considering the change between 2012 and 2021, the households with children enhanced their economic situation changing from a poverty rate of 27% to 19%, a similar trend occurred for adults. However, for elderly people, the rate of poverty increased from 8% to 10%.

Another substantial variable in the relation of economic status and contamination exposure is household income. The American Community Survey of 5 years collected data from 2009 to 2021, containing information on annual median income per household for all the census tracts in Arizona, there is a significant variation over time and counties. Figure 10.a shows the tendency of the average income yearly. Between 2009 and 2015 the median income was stable, and after 2016 the income has been increasing to a median value of approximately 65.000 USD per household. Figure 10.b presents the distribution of income across counties. Maricopa, Greenlee, Pinal, and Coconino are the counties with the highest levels, whereas Santa Cruz, La Paz, and Apache are the counties with the lowest values. The comparison between these results and the poverty rates shows that for most of the counties, income has been increasing while the poverty rate has been decreasing.

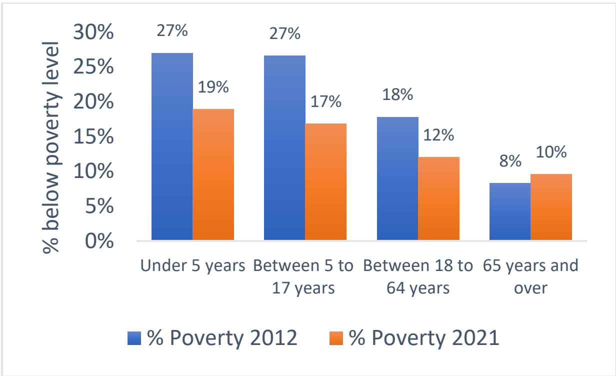
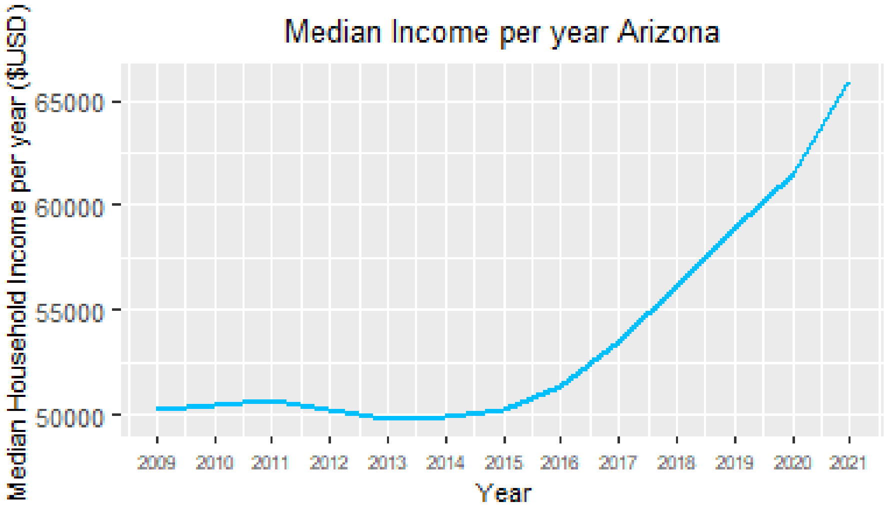
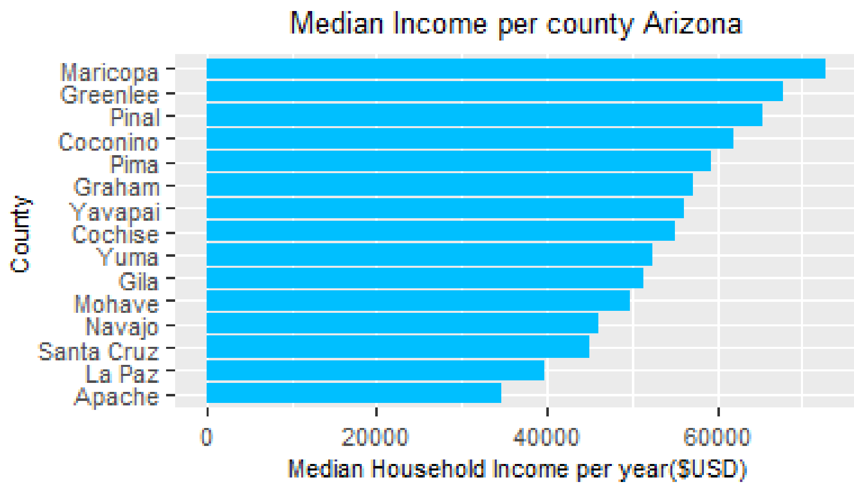


Figure 9. Poverty Rate by Age (2012 and 2021)



(a)



(b)

Figure 10. Median household income per year a)Period: 2009-2022 b) Income per county 2021

4.4 Community Water Systems

The distribution of CWS across Arizona is included in the following map (Figure 11). Most of the systems are in the central and southern areas of the state where the highest populations are located. The database of the residents served by the CWS of Arizona is a shapefile with area per system published in the Arizona Department of Water Sources in 2022 (Arizona Department of Water Resources, 2022). The information includes the area served for 718 CWS that provides water for 6.65 million people in the state. After debugging the data, the active CWS with socio-economic variables is 708 CWS which served 6.57 million people in Arizona, 90% of the population in the state.

There are different categories of size of the CWS, there are large number of CWS that served very small communities (<500 inhabitants). Only a few served large populations (>10,000 inhabitants) and very large populations (>100.000 inhabitants). The distribution of services per size is included in Figure 12a. In addition, another relevant variable related to water pollution is the water source, there is a higher correlation between water chemical pollution and groundwater sources for the natural phenomenon that affects groundwater. In Arizona 60% of the population is served by surface water sources and 40% by groundwater, the distribution per source is included in Figure 12b. In general, the characteristics of the CWS have not changed over time.

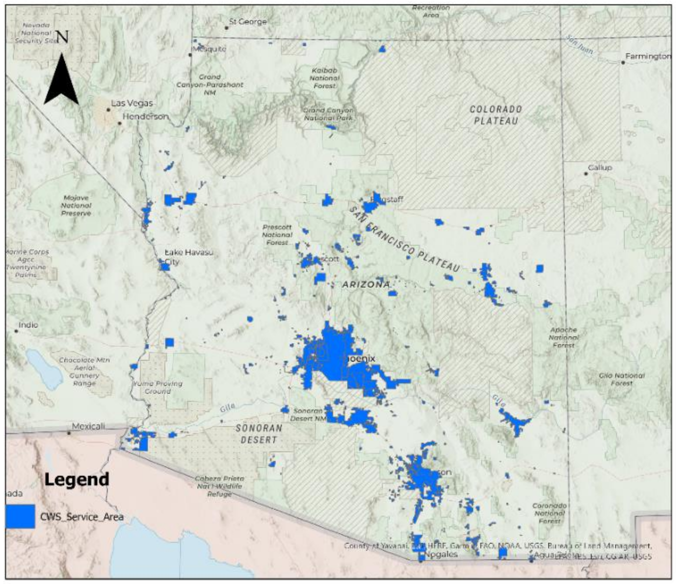
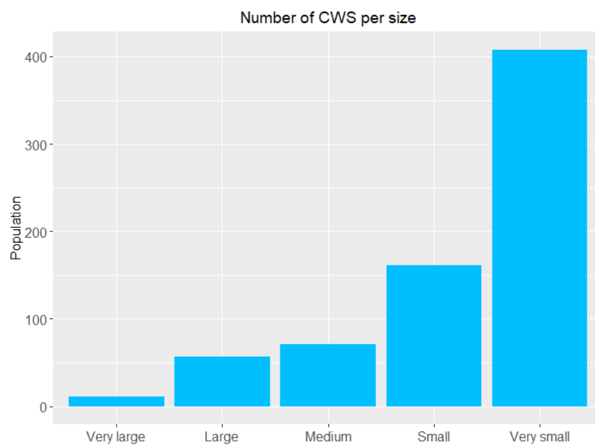
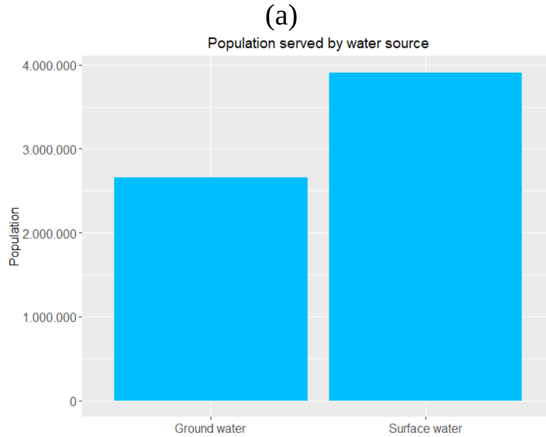


Figure 11. Area of served population by CWS in Arizona 2022



(a)



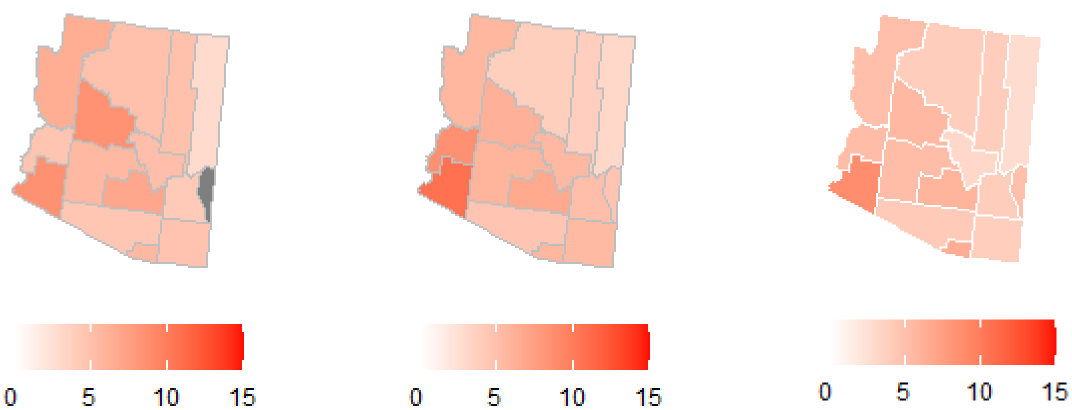
(b)

Figure 12. Distribution per size and water source of CWS in Arizona

4.5 Spatial characteristics

According to the review of environmental justice studies of water pollution in the U.S., there is a significant variance in the contamination events associated with the location of the public water systems. Given this concern, in this study one of the control variables is the county per CWS, the purpose is to estimate the effect at a county level and avoid the noise of spatial correlation associated with the estimation of the model. There are 15 counties in Arizona, in previous sections, the variance of socioeconomic variables has been described. Now, the changes in the concentration of contaminants for 2010, 2015, and 2021 per county are shown in Figures 13, 14, and 15.

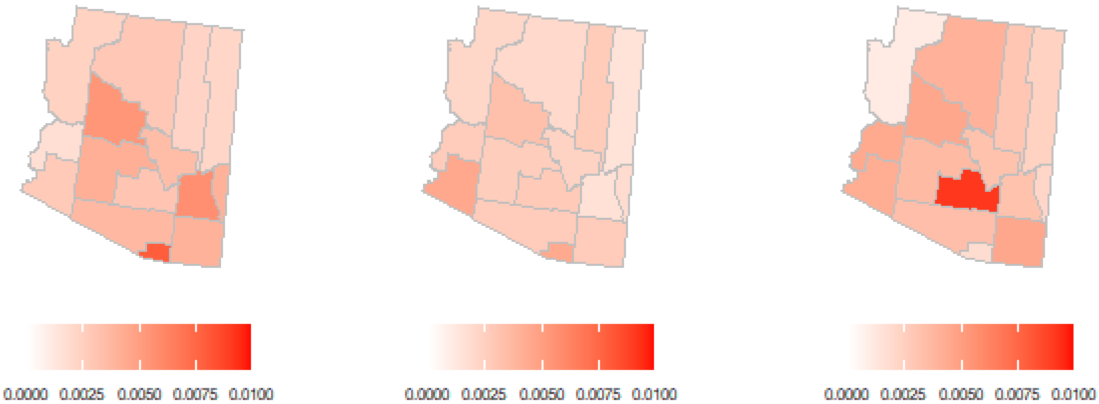
For Arsenic, the highest concentrations in CWS samples were found in Yuma, Yavapai, and Santa Cruz. However, there is a significant decrease in the levels over time. For counties in the northeast area of the state (Apache, Navajo, and Coconino), the concentrations were lower, though, the results have been calculated with only a few samples per county. The average concentrations of arsenic in water per county are shown in Figure 13.



Arsenic Mean Concentration 2010 *Arsenic Mean Concentration 2015* *Arsenic Mean Concentration 2021*

Figure 13. Arsenic average concentrations per county (ug/L)

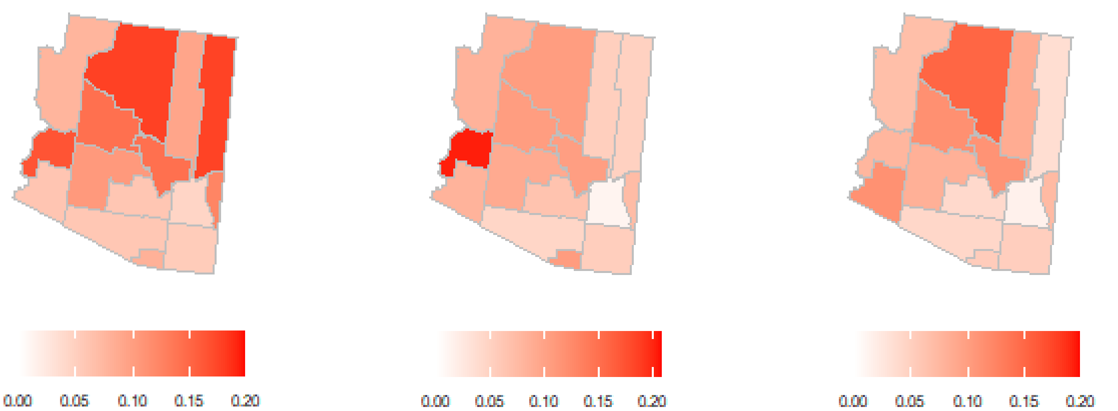
The Lead and Copper Rule established a standard with a maximum level in drinking water of 1.3 ppm (1.3 mg/L) for copper and 15 ppb (0.015 mg/L) for lead. Figure 14 shows that the concentrations of lead have been increasing from 2010 to 2021. Pinal is the county with the highest level in 2021. In contrast, the levels of copper have been decreasing over time, nonetheless, Coconino maintains a higher level compared to other counties. Figure 15 shows concentrations of copper in drinking water per county.



Lead Mean Concentration 2010 *Lead Mean Concentration 2015* *Lead Mean Concentration 2021*

Figure 14. Lead average concentrations per county (mg/L)

Section 2.3 includes the principal sources of contamination for the analyzed contaminants. One of the principal sources is mining activity in the area, to control for this variable, a shapefile of active mining areas published by U.S Geological Survey is included in the analysis (USGS, 2022). In addition, regardless of the open sites, several abandoned mining places are located in the state which are another pollution source, to account for these contaminants a shape file of superfund sites is incorporated in the study. Figure 16 displays the distribution of mining sites and superfund places in the state, most of which are in the central and southern regions.



Copper Mean Concentration 2010 *Copper Mean Concentration 2015* *Copper Mean Concentration 2021*

Figure 15. Copper average concentrations per county (mg/L)

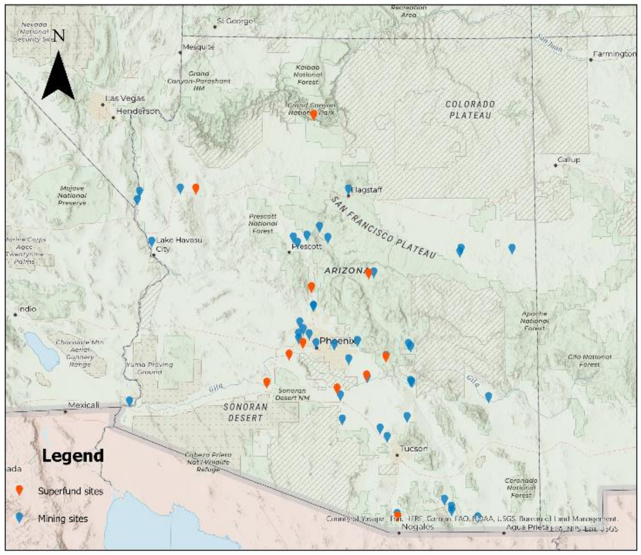


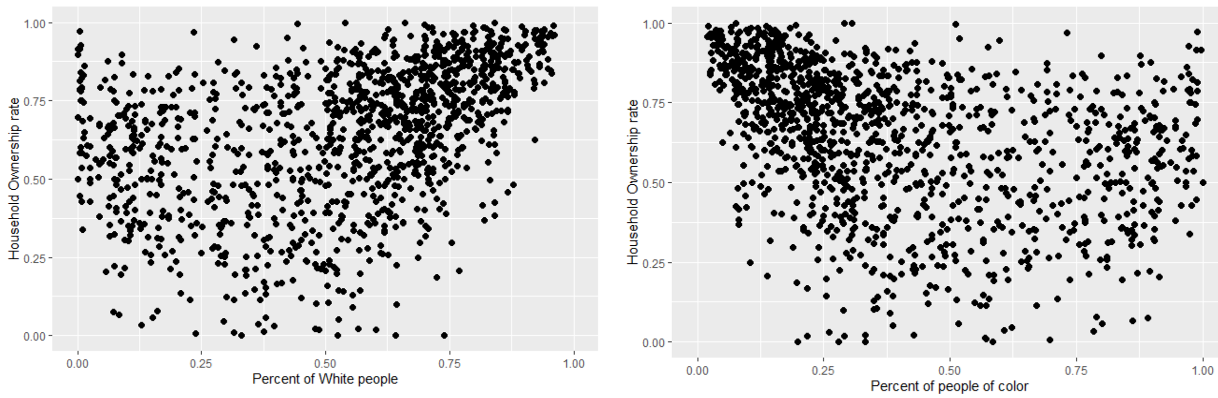
Figure 16. Mining sites and superfund sites in Arizona

The contaminants concentrations have decreased in the analyzed period for arsenic in most of the counties. (Foster, Pennino, Compton, Leibowitz, & Kile, 2019) identified a trend of reduction in the violations of for arsenic at a national level between 2009 to 2016 which, it may be inferred that the enforcement the arsenic standard may be contributed to the reduction of arsenic exposure. However, there is no evidence of this trend at a state level.

4.6 Interrelations between ethnicity/race, poverty rate, home ownership and age of homes in Arizona

As we discussed in previous sections there are significant differences in poverty rate according to the race or ethnicity of the community in Arizona. In general, communities of people of color have a higher poverty rate than white communities with the higher disparities in Hispanics and American Indians. Besides, in this section the identification of other interrelations is included such as homeownership across different races or ethnicities and different poverty rate levels, and age of households in segments of the population with different races and income.

First, in 2020, homeownership rates in Arizona were 69.55% for tracts with a higher proportion of white people, compared to an average rate of 54.13% for tracts with a higher percentage of people of color. For specific ethnic or racial groups, the average rate was 54.34% for tracts with mostly Hispanics, and 70.70% for tracts with a greater proportion of American Indians. Figure 17 represents the relationship between the percentage of white people and people of color for the homeownership rate. In general, most inhabitants of white communities are homeowners, while in tracts with mostly people of color, the relationship is inverse, reaffirming that communities with low incomes have lower access to homeownership.



(a) (b)

Figure 17. Homeownership rate for different ethnic/racial groups in 2020.

Second, looking the relationship of the household ownership rate with the poverty rate when the poverty rate is lower the rate of homeownership is higher, supporting the previous relation about a lower ownership rate in communities mostly composed of people of color. In summary, in the census tracts with low income there is a higher proportion of people of color and a lower household ownership rate. Figure 18 shows the distribution of household ownership rate for each poverty rate per census tract in 2020.

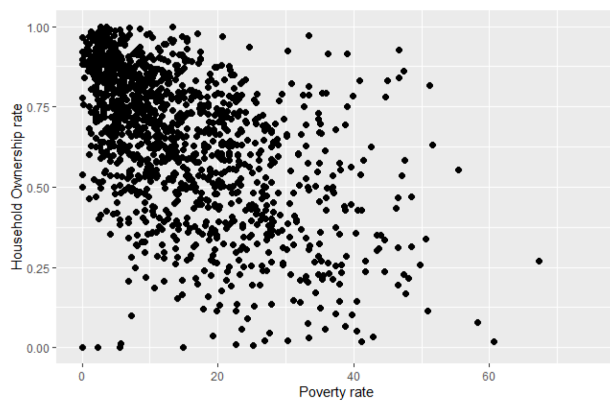


Figure 18. Homeownership rate vs poverty rate in 2020

There is a potential relationship between household age and the presence of lead pipelines. Older houses are more likely to have lead pipes or plumbing fixtures that contain lead which can contaminate drinking water. In many cases, lead pipes were installed in homes built before 1986, the year when the Safe Drinking Water Act prohibited the use of lead pipes in new constructions (EPA, 2017). As a proxy to the households built before 1980's, the EPA in the EJ Screen database included the housing units built before 1960 as an environmental factor (EPA, 2023).

Using the variable of houses built before 1960 and the ethnic/ racial groups proportions in 2020 at a census tract level, in average the tracts with mostly people of color had 17% of houses built before 1960 while the mean value for tracts with mostly white people had only 6.8% of old houses. This is relation that may result in a higher lead exposure in drinking water in people of color communities because there is a higher percentage of old households in those communities. Figure

19 shows the distribution of the percentage of old houses in white and people of color tracts. Finally, in Appendix 3.2.c a graph shows the relationship between poverty rate and the percentage of houses built before 1960. In general, if the poverty rate in a tract is lower the percentage of old houses is also lower.

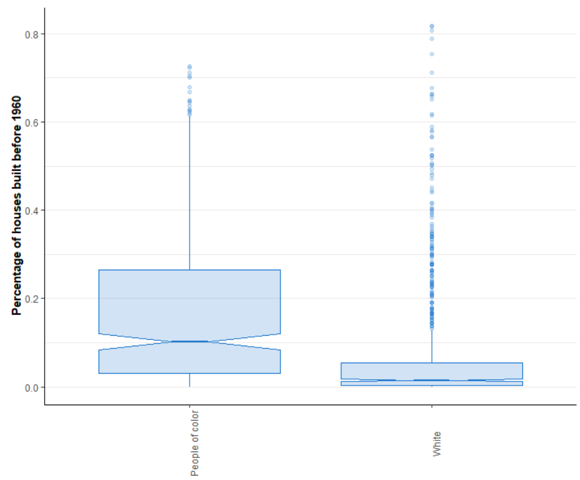


Figure 19. Percentage of houses built before 1960 in tract with a higher proportion of people of color and whites.

Chapter 5: Empirical Models

The data discussed in Chapter 3 are included in statistical analyzes using the models described in the last sections of the literature review in Chapter 2. Considering the dependent variables included in this study there are two different approaches: a continuous variable of the concentrations of chemical contaminants in the drinking water samples and a binary dependent variable of the violations in the rules for arsenic and lead and copper.

For the first model, initial rounds of statistical analysis employ Ordinary Least Squares (OLS), then panel data models for contaminants concentration levels. This chapter describes the preliminary independent variables tested for statistical significance to determine which are meaningful in predicting the concentration of chemical contaminants in drinking water. Then, robustness checks are conducted on the panel data regressions and find the variances of error terms from these models, and test for heteroscedasticity and autocorrelation.

In addition, potential alternative functional forms for the concentration levels with a limited dependent variable and their strengths and weaknesses are described. Besides, a discussion about the marginal errors in the limited dependent variable is evaluated, and their interpretation in the significant independent variables.

In the second approach, the evaluation of the binary outcome variable models is developed to establish the prediction accuracy and estimate the best approach with violations for chemical contaminants included in this analysis as the dependent variable.

5.1 Empirical OLS Regressions

In environmental justice studies a correlation has been identified between race, poverty, age, and pollution. In this study, the dependent variable of pollution is the concentration level of arsenic, lead, and copper in drinking water. Potential measurement errors in the samples could be reported to the Arizona Department of Environmental Quality depending on the differences in the minimum level of detection of the equipment used to measure the concentration level per CWS across the state. Besides, to enhance the internal validity and limit the influence of confounding in this study additional control variables are included, such as characteristics of the CWS and spatial variables.

The variables of interest are race or ethnicity characteristics discovered the differences between Hispanic, black, and American Indian communities to white communities. Economic status characteristics such as poverty and income are measured using the mean household income per year and the percentage of people below the poverty level. Age and sex are characteristics that are correlated with pollution levels, to determine a possible connection between age and pollution, the analyzed variables are percentages of the elderly and children population. To check any relationship between sex and pollution the percentage of female-led households is used.

Considering that the CWS are the units of this analysis, the empirical model includes as control variables characteristics of CWS. The determinants of interest are size and principal water source (groundwater or superficial).

In Chapter 2, the description of chemical water contamination in Arizona is included, a link was identified between mining activity and water contamination levels in Arsenic and Copper. This study applies two spatial variables that potentially correlate with pollution, first census tracts where a mining site is active and superfund sites where mining activity was developed. Finally, a dummy variable per county is included in the model to account for spatial differences.

Initially, the socioeconomic characteristics and spatial variables were available at a census tract level. To link the census tract level data with the CWS, the analysis used the polygons of the served area per CWS to determine census tracts and shared area per CWS to determine a weighted average per variable. As a result, a database with the list of the active CWS and their socioeconomic, spatial and system variables were found for the period of interest. This information was the basis for the initial OLS model.

The significant variables in the OLS models are described in Table 3. For the Arsenic empirical OLS model, the percent of Hispanics, the rate of homeownership, and the dummy variable for groundwater had significant positive coefficients. For Hispanics, in Chapter 3, the distribution of race per county was described, there is a significant difference between counties with a higher proportion of white people (base case) and Hispanics. Yuma and Santa Cruz are the counties with the highest levels of Arsenic with a higher proportion of Hispanics living in these areas. In Chapter 2, the relationship between groundwater and Arsenic contamination was established. However, for home ownership, the sign of this coefficient is not negative as expected. On the other hand, the proportion of black people, percent of the elderly population, and the location of superfund sites have significant coefficients with a negative sign. For the black group, the reason is that this community is located mostly in Maricopa County which is one of the counties with the lowest concentration of arsenic. Elderly people have the lowest level of poverty, the negative coefficient confirms the link between poverty and pollution. Finally, most of the superfund sites in Arizona are abandoned uranium and copper mines, which are not related to arsenic.

For the copper empirical OLS model, there are significant positive coefficients for the ratio of American Indians, poverty rate, and size of the CWS. First, as we described in Chapter 2 American Indian communities historically have suffered from water pollution. Besides, the positive sign of the poverty rate is the confirmation of the connection between poverty and race for the copper model. Finally, in this case, there is a positive relation between the size of the CWS a copper concentration. However, considering that the levels of copper in drinking water in most of the cases are below the limit, it means that large CWS are probably not treated the water for this contaminant. On the other side, the percentage of Hispanics, the percentage of children, and the dummy of groundwater have negative significant coefficients. For Hispanic communities, the concentrations of copper in drinking water are lower in the counties of the southern region where a higher proportion of Hispanics lived.

For the lead empirical OLS model, there are only 2 significant coefficients a negative sign with the percentage of children and a positive sign with the size of CWS. However, the variables included in this model do not explain the concentration of the contaminant in a good proportion, the R2 for this model is only 2%.

Table 3. Significant variables OLS Model

Variable	Arsenic	Copper	Lead
PERCENT_HISPANIC	1.39	-0.076	
PERCENT_BLACK	-9.208		
PERCENT_AMERICAN_INDIAN		0.199	
POVERTY_RATE		0.001	
PERCENT_ELDERLY	-0.946		
PERCENT_CHILDREN		-0.11	-0.007
PERCENT_HOMEOWNERSHIP_RATE	2.098		
D_GROUNDWATER	1.359	-0.048	
SIZE_CWS		0.007	0.0002
SUPERFUND_SITE	-0.979		

The spatial characteristics of the CWS are relevant to the analysis. Table 4 shows the significant coefficients of the dummies per county associated with concentration levels, using as base case scenario Maricopa County. For Arsenic, there are significant positive coefficients for Pinal, Yavapai, and Yuma, while negative coefficients with Apache, Cochise, Coconino, Gila, Graham, Navajo, Pima, and Santa Cruz. These results confirm the distribution of arsenic per county shown in Figure 13, accounting for counties with higher or lower concentrations of arsenic compared to Maricopa County.

For copper, one of the counties with the lowest concentrations is Maricopa, compared to this county most of the other counties have significant positive coefficients. Finally, for lead levels, Apache and Mohave have significant negative coefficients, and Pinal and Yuma have significant positive coefficients. The results of the complete OLS model are in Appendix 1.

Table 4. Significant county dummy variables OLS Model

County	Arsenic	Copper	Lead
d_APACHE	-3.312***		-0.001**
d_COCHISE	-1.252***		
d_COCONINO	-1.291***	0.022**	
d_GILA	-1.895***	0.025**	
d_GRAHAM	-1.354*		
d_GREENLEE		0.072***	
d_LAPAZ		0.056***	
d_MOHAVE			-0.001**
d_NAVAJO	-1.626***		
d_PIMA	-1.755***		
d_PINAL	0.647**		0.001***
d_SANTACRUZ	-0.884*	0.040***	
d_YAVAPAI	0.763***	0.045***	
d_YUMA	2.725***	0.046***	0.001**

Note: *p<0.1; **p<0.05; ***p<0.01

Specified that the analyzed data varies across CWS and time, the empirical OLS model described earlier is a cross-sectional analysis that accounts only for the variability of the individuals, consequently, a relation between the errors in different periods for the same individuals appeared generating a problem of autocorrelation. The panel data model is a better approach, this regression captures the effect of variation between individuals and time. The next section describes the empirical model and significant coefficients.

5.2 Empirical panel data regression

The empirical panel data regression is derived from the OLS model, the only difference is the variation across time. The model is described in Equation 1, the dependent variable is the concentration of contaminants per year t per CWS i. The explanatory variables are race, economic status and power, sex, age, characteristics of the CWS, variables related to contaminant source, and dummies at a county level. The base case comparison is the communities of white people in Maricopa County.

$$C_{it} = \beta_0 + \sum_{1}^{3} \beta_i RACE_{it} + \sum_{4}^{6} \beta_i ECONOMIC_{it} + \sum_{7}^{9} \beta_i SEX\&AGE_{it} + \sum_{10}^{11} \beta_i CWS_i + \sum_{12}^{13} \beta_i Csource_i + \sum_{14}^{28} \beta_i D\_county_i + \varepsilon_{it}$$

To estimate the best approach between the fixed effects model and random effects model for panel data Hausman test is used, the description of this test was described in section 2.6.3. The results of the test (p-value) for the analyzed contaminants are shown in Table 5, given that the null hypothesis of this test uses random effects. The best option considering an alpha value of 0.05 is to use random effects for arsenic and lead and fixed effects for copper.

Table 5. Hausman Test p values

CONTAMINANT	HAUSMAN TEST (P VALUE)
ARSENIC	0.1267
LEAD	0.1972
COPPER	0.0370

Potential issues with the assumptions of panel fixed or random effect models are the presence of heteroscedasticity or autocorrelation. To account for these problems, robust standard errors are applied in the estimation, and the Durbin-Watson test was used to assess autocorrelation. The test for arsenic, lead, and copper show problems of autocorrelation.

The results for the significant coefficients of time change variables in the random effects models of arsenic and lead and the fixed effects model for copper controlling for heteroscedasticity and autocorrelation are shown in Table 6. In the arsenic case, percent of Hispanics, the rate of poverty, and the homeownership proportion have positive significant coefficients and percent of elderly, percent of black have negative significant coefficients. For lead, only the percentage of children has a negative significant coefficient and for copper percent Hispanic, the percentage elderly, and the percentage of children have negative significant coefficients, while the percentage of families below the poverty level and homeownership rate have negative significant coefficients.

Likewise, for time-invariant characteristics in the random effect panel data models for arsenic and lead the results are included in Table 7, for arsenic a significant positive relation is identified for groundwater sources and a negative coefficient for superfund sites. Considering the dummy variables at a county level the sign and significance are equivalent for the OLS model. However, the positive significant coefficients on the arsenic model for Pinal, Yavapai, and Yuma are not significant in the random effect approach.

Table 6. Results of significant variables panel data models.

	Dependent variable:		
	Arsenicmean (Random)	LeadMean (Random)	copperMean (Fixed)
PERCENT_HISPANIC	1.518*		-0.079***
PERCENT_BLACK	-8.720*		
PERCENTAGE_POVERTY	0.008*		0.0004*
PERCENT_Elderly	-1.972***		-0.027*
PERCENT_CHILD		-0.007**	-0.095*
PERCENT_Ownership_rate	1.954***		0.032**
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7. Significant time-invariant variables coefficients

	Dependent variable:	
	Arsenicmean (Random)	LeadMean (Random)
d_groundwater	1.313***	0.0001
Superfund	-1.064*	-0.0001
Miningsite	0.183	-0.001**

A robustness check of the results was conducted. In this case, the estimation of the significant coefficients was done using not only the average concentrations per year per CWS but also the maximum concentration levels per period for all the units. The results are shown in Appendix 1. In most cases, the significance level and sign of the coefficients are the same. There are slight differences in some variables such as the significance of the size per CWS in the models that used max concentrations as the dependent variable.

Another potential issue in the model is multicollinearity. First a calculation of the correlation between the dependent variables was found, the correlation table is included in Appendix 5. In general, the only variable with a high correlation is income and poverty rate with a correlation of less than -0.5. Then, using the Vif command a final checking was done, the results of the Vif values for the explanatory variables used in the panel data models are in Table 8. In conclusion, all the vif values are between 1 to 2 meaning that there is a moderate correlation between a given predictor variable and other variables in the model, and multicollinearity is not a problem in the explanatory variables.

Table 8. Vif of explanatory variables

Variable	VIF	Variable	VIF
PERCENT_HISPANIC	1.55	d_COCHISE	1.42
PERCENT_BLACK	1.11	d_COCONINO	1.39
PERCENT_AMERICANINDIAN	1.25	d_GILA	1.41
PERCENTAGE_POVERTY	1.15	d_GRAHAM	1.06
PERCENT_Elderly	1.30	d_GREENLEE	1.06
PERCENT_CHILD	1.11	d_LAPAZ	1.26
PERCENT_FEMALELED	1.12	d_MOHAVE	1.54
PERCENT_Ownership_rate	1.20	d_NAVAJO	1.39
d_groundwater	1.37	d_PIMA	1.91
Sizecat	1.49	d_PINAL	1.51
Superfund	1.08	d_SANTACRUZ	1.22
Miningsite	1.31	d_YAVAPAI	1.96
d_APACHE	1.21	d_YUMA	1.35

5.3 Empirical Tobit Regression

To avoid false-positive quantification of a constituent, very low concentrations are censored and reported as a “less than” value by the laboratories. Censoring levels changed over time as methods change and differ depending on the accuracy of the equipment used to measure the concentration. The limit of detection (MDL) is defined as the minimum concentration that can be measured with 99% confidence.

For this study, the database of SDWIS has censoring levels. Before the determination of the mean values, the total samples with concentration levels were 24,390 for arsenic, 47,776 for copper, and 47,799 for lead. These observations have 20%, 11%, and 67% of censoring values for arsenic, copper, and lead respectively. For OLS and panel data models, the concentration levels were established at the MDL.

Furthermore, MDL varies depending on the contaminant and laboratory that processes the information, the proportion of censoring observations per limit of detection for each contaminant is described in Table 9. Given this characteristic of the information, a censored panel data model is used as a better approximation to estimate the impact of interest variables in concentration levels.

In section 2.6.1 censored data regression model (Tobit model) was described. One of the constraints of this method is the usage of only one limit per side (right or left). Given that the information has different MDLs, only the most common limit was selected and the observations with concentrations below the limit were adjusted to it. The MDL was 1 ug/L, 0.01 mg/L, and 0.005 mg/L for arsenic, copper, and lead. However, for the lead given that the MDL of 0.001 mg/L is especially frequent and lower than the limit of 0.005 mg/L, the study uses the limit of 0.001 mg/L to avoid the adjustment of more data.

Table 9. Censoring observations per limit of detection.

Arsenic		*Copper*		*Lead*	
MDL	**ug/L**	**MDL**	**mg/L**	**MDL**	**mg/L**
0.5	122	0.001	81	0.0005	4506
1	**3160**	0.002	167	**0.001**	**11706**
2	255	0.004	61	0.002	1973
2.5	49	0.005	1178	0.0022	56
3	589	**0.01**	**2097**	0.0025	24
5	506	0.02	1234	0.005	13586
10	66	0.05	166	Other	126
Others	75	Other	74		
Total	**4822**	**Total**	**5058**	**Total**	**31977**

Using the average levels of contaminants after the adjustment of the observations to the minimum level, a Tobit panel data model was used to estimate the significant coefficients. The results of the major variables are included in Table 10. Considering that income is not a significant variable in the previous models and the variable is correlated with poverty level, the result of the Tobit model does not include income as an explanatory variable.

Table 10. Significant variables Tobit Model

	Dependent variable:		
	Arsenic_mean (1)	Lead_mean (2)	Copper_mean (3)
PERCENT_HISPANIC	0.468*	-0.00192*	-0.1554**
PERCENT_BLACK	-9.660**	0.00860*	
PERCENTAGE_POVERTY	0.0134*	0.00000599***	
PERCENT_Elderly	-2.311***		
PERCENT_AMERICAN_INDIAN			0.2000*
PERCENT_Ownership_rate	2.123***	0.00172*	
d_groundwater	1.352**		
Sizecat		0.0010***	0.02305***
d_APACHE	-3.773***		
d_COCHISE	-2.909***		
d_COCONINO	-2.089***		
d_GILA	-3.377***	0.00146*	
d_LAPAZ		0.00216*	
d_GRAHAM	-3.541***		
d_NAVAJO	-2.973***		
d_PIMA	-2.394***		
d_PINAL		0.00149**	
d_SANTACRUZ			0.2512**
d_YAVAPAI		0.0009*	
Note:	*p<0.1; **p<0.05; ***p<0.01		

For time-variant characteristics, in the arsenic model, the percentage of Hispanics, the rate of poverty, and the homeownership proportion have positive significant coefficients and percent of elderly, percent of black have negative significant coefficients. For lead, percent of black people, rate of poverty, and homeownership rate have positive significant coefficients and for copper percent Hispanic has a negative significant coefficient, while American Indians has a positive significant coefficient.

5.4 Empirical model binary dependent variable

Chapter 3 includes the description of the two principal binary models: Probit and Logit. In general, there are not significant differences in the estimation between these models. However, as a first step in this study, a verification of the differences between the real violation values and estimated violations for probit and logit was conducted.

The results for arsenic, copper, and lead are shown in Table 11. In general, the average of the predicted values is similar comparing both options for all the contaminants. As a result, the analysis estimates the accuracy of both options using 0.5 in the prediction as the threshold, when the prediction was lower than 0.5 it was considered a non-violation, while higher than 0.5 was a violation. The percentage of correct estimation in the arsenic case is the same for logit and probit, for lead and copper, the probit model estimates the violations correctly in 79.25% of the observations, while the logit model estimates it correctly in 79.20%.

Table 11. Predicted values for probit and logit

Arsenic					
Variable	Obs	Mean	Std.dev	Min	Max
Violation	4,054	0.1835224	0.387142	0	1
pprobit	4,054	0.1266536	0.0727487	0.0050747	0.3872446
plogit	4,054	0.1272856	0.0729838	0.0052445	0.3888368
Lead and Copper					
Variable	Obs	Mean	Std.dev	Min	Max
Violation	4,005	0.2062422	0.4046569	0	1
pprobit	4,005	0.2005294	0.117713	0.0137002	0.7163303
plogit	4,005	0.2009151	0.1182206	0.017601	0.71684

The significant coefficients for the binary dependent variable models for Arsenic (logit) and Lead & Copper (probit) are described in Table 12. For arsenic, the rate of poverty and percent of homeownership are positive coefficients, while the percent of black and percent of elderly are negative coefficients. For lead and copper, the percentage of Hispanics and elderly have positive coefficients, while the size of the CWS is a negative coefficient. For the first regression, the results are consistent with the Tobit model of concentrations for arsenic. However, the second model has dissimilar results.

For the lead and copper rule, a clarification is that a violation might be registered if one of the contaminants exceeds the standard, analyzing both contaminants under the same rule produced a discrepancy in the model. However, in this case, 20% of the observations registered a violation but only 13% of these violations exceed the limits for lead or copper. Besides, a violation might be reported for other reasons such as if the CWS does not perform certain actions of public education or lead service line replacement or if the samples were collected improperly or were not reported, or if the treatment was done incorrectly (EPA, 2019).

Table 12. Results Binary Dependent variable models

	Arsenic Logit (1) Violation	Lead&Copper Probit (2) Violation
Sizecat		-0.253***
PERCENT_HISPANIC		0.526*
PERCENT_BLACK	-10.98**	
PERCENT POVERTY	0.0315**	
PERCENT_ElDERLY	-1.476*	0.523*
PERCENT_Home Ownership	1.337*	
d_APACHE		1.150***
d_COCONINO		0.430*
d_GILA	-2.998**	
d_GREENLEE		1.121***
d_LAPAZ		0.795***
d_MOHAVE		0.877***
d_NAVAJO	-2.660**	0.884***
d_PIMA	-2.402***	
d_PINAL		0.378*
d_YUMA		1.021***

Chapter 6: Econometric Results

Chapter 4 described the significant variables of the censored data model for concentrations of the contaminants and binary-dependent regressions for violations of the arsenic standard and lead and copper rule. This section has a description of the econometric results and the potential effect of race, economic status, or age on pollution. Besides, checking the spatial correlation presented in the models and the significant effect of spatial variables. The goal of this section is to identify any gap associated with environmental justice.

6.1 Tobit model for concentrations

The identification of the significance in Chapter 4 pushes on the relation between relevant socioeconomic variables and pollution. Considering that the coefficients are not the same as the marginal effects in the Tobit model, Table 13 includes the marginal effects for the significant coefficients per contaminant.

Table 13. Econometric results Tobit Model

	Marginal Effects:		
	Arsenic_mean (1)	Lead_mean (2)	Copper_mean (3)
PERCENT_HISPANIC	0.4064*	-0.000972*	-0.088543**
PERCENT_BLACK	-7.4925**	0.004352*	
PERCENTAGE_POVERTY	0.0104*	0.0000303***	
PERCENT_Elderly	-1.7980***		
PERCENT_AMERICAN_INDIAN			0.113943*
PERCENT_Ownership_rate	1.6582***	0.0008705*	
d_groundwater	1.0559**		
Sizecat		0.0005244***	0.013135***
d_APACHE	-2.9664***		
d_COCHISE	-2.2909***		
d_COCONINO	-1.6210***		
d_GILA	-2.8654***	0.0007408*	
d_LAPAZ		0.0010946*	
d_GRAHAM	-2.8155***		
d_NAVAJO	-2.2566***		
d_PIMA	-1.8807***		
d_PINAL		0.000757**	
d_SANTACRUZ			0.143118**
d_YAVAPAI		0.0004962*	
Note:	*p<0.1; **p<0.05; ***p<0.01		

6.1.1 Arsenic

The Tobit model of arsenic presents a causal significant relation between race and pollution. For Hispanics, the relation is positive and for the percentage of Black, the relation is negative. The unit of the percent of Hispanics is in decimals, the interpretation of the coefficient is that if the percentage of Hispanics increases by 1% then the concentration of arsenic increases by 0.004 ug/L. This finding is consistent with the result developed at the county level by Morata, et al., 2022 meaning that the relationship in the U.S. is similar to the Arizona level where a disproportionate exposure of Hispanic communities was identified.

On the other hand, the relationship between the black percentage population and arsenic concentration is shown using white communities as the baseline. Appendix 2 (1. a) includes the plots showing the relations between race/ethnicity and concentration of contaminants. The relation between the proportion of white people and arsenic concentration has a positive correlation, while the figure between black and arsenic concentration shows a negative correlation. Given that the percentage of black people is in decimal units, if the percentage of blacks increases by 1% then the concentration of arsenic will be lowered by 0.075 ug/L.

For arsenic, there is a significant effect on the rate of poverty, confirming the hypothesis of a causal relationship between income in pollution as Banzhaf, Ma, & Timmins, 2019 established in their study. The level of poverty is measured as a percentage, meaning that if the rate of poverty increases by 1%, it produces an increase of 0.01 ug/L of arsenic. In general, this effect may be produced by the Coasean bargaining effect. Likewise, in Appendix 2 1. b, the relation between income and arsenic concentration is included with a clear pattern of a higher concentration level in low-income zip codes which reaffirms the relation with poverty rate. One of the possible explanations for this effect is a low level of access to education as the reason for a higher rate of poverty and a potential effect of intergenerational poverty. To confirm this hypothesis the variable of the Percentage of people with less than a high school diploma is taken from the EPA Environmental Justice Database at a tract level, the relation between this variable and the poverty rate is included in Appendix 3.4. there is a clear trend where communities with less educated people have a higher poverty rate.

Zip codes where a higher percentage of the population is over 64 years old have a lower exposure to arsenic, thus, if the percentage of elderly people increases by 1% it produces a drop of 0.017 ug/L in arsenic. This association is connected to a lower poverty rate in this segment of the population. Appendix 3 shows the negative relation between the elderly and the poverty rate.

Home-ownership rate in Arizona shows a positive correlation with arsenic concentration (Appendix 3.1). In 2017, 63% of households were owners. The homeownership in nonmetro Arizona was 72%, compared to 62% in Maricopa and Pima Counties (State of Arizona, 2020). Given this association, the plot of arsenic concentration of metropolitan and non-metropolitan areas (Appendix2. 1e) shows a difference in exposure to arsenic, in average the concentration in Pima and Maricopa was 4.94 ug/L and 5.59 ug/L for the other counties. The T-test of the difference between arsenic concentrations means of metropolitan and non-metropolitan areas confirms that the difference is significant. As a result, there is a positive relation between home-ownership rate and arsenic exposure associated with a higher rate in non-metropolitan counties where the concentration of arsenic is significantly higher than in metropolitan counties.

The most common contaminant found in Arizona groundwater in a concentration above health-based drinking water standards is arsenic (Uhlman, Rock, & Artiola, 2009). The model confirms that CWS which used groundwater as its primary source has a significantly higher level of arsenic than CWS which used surface water. Groundwater as a source is associated with an increase in 1.05 ug/L in arsenic concentration.

Considering the significant dummy spatial variables, in the counties of the northeast area of Arizona the concentration of arsenic in the CWS is lowered than for the CWS of Maricopa. This area does not have too many public services given the restricted access to water sources, making the comparison not representative of the source and more associated with a lack of information. Likewise, for Pima County, the concentration is significantly lower than in Maricopa.

6.1.2 Lead

There is one important consideration in lead concentrations effects, almost 60% of the sample results are censored and not identified at the minimum level of detection. Black people were exposed to a higher concentration of lead, this effect is associated with the places where the African American community lived in Arizona, (Appendix 2 (3e)).

One of the principal causes of elevated lead and copper in water is the age of the pipelines, an older house is more likely to have lead pipelines, as a proxy to estimate this relation the EPA EJScreen database has information on the percentage of houses built before 1960 per tract in the

U.S (EPA, 2023). A way to establish the connection between this proxy and any racial disparities is to compare the distribution of the variable and the percentage of the racial/ethnic group. The relation between the percentage of Hispanic and houses built before 1960 (Appendix 3.2) shows a higher concentration in tracts where the percentage of Hispanics is lower, while, for white communities, it is concentrated in tracts where the percentage of white is higher. Thus, the relationship is negative and exposure to lead is lower in Hispanic communities.

The relation between pollution and poverty rate in lead is positive, meaning that communities with lower income have higher concentrations of lead in drinking water. If the poverty rate increases by 1% the concentration of lead are higher at 0.00003 mg/L. Using the proxy of houses built before 1960, the plot with the relation between the poverty rate and the proxy (Appendix 3.2.c) shows that for tracts with a poverty rate above the mean the percentage of old households is higher increasing the potential exposure to lead in drinking water. A sample t-test between the means of the groups above and below the average rate of poverty confirms that the difference is significant.

As in arsenic, there is a significant positive effect of home ownership on lead exposure, this relation is not clear given the distribution of home-ownership rate in the state. In Pinal and Yuma where the exposure is higher, the home-ownership rate is lower than in other counties in the state (Appendix 3.3).

The Tobit model results estimate a positive effect of size on lead and copper concentrations of 0.0005 mg/L for lead and 0.013 mg/L. Appendix 2 (2.d) shows the distribution of copper concentrations per size of CWS and Appendix 2 (3. e) displays the spreading of lead per size of CWS. For copper and lead, the distribution of samples is more concentrated in CWS of very small and small sizes given the number of services in the database per category. Copper concentrations are higher in the medium, large, and very large sizes explaining the positive coefficient associated with the outcome variable. For lead. the highest concentrations were registered in medium-size CWS, and it is representative of more samples than CWS with large or very large size, as a result given that this variable was defined as a categorical variable (1: very small to 5: very large) a higher percentage in the medium category explains the positive coefficient.

The significant positive dummy variables for counties are Gila, Pinal, La Paz, and Yavapai. These counties have higher concentration levels for lead than Maricopa. (Appendix 2.3.d)

6.1.3 Copper

The model includes a database with the location of the mining site as a possible indicator of copper pollution in drinking water. However, the variable was not significant in the Tobit Model. Given this result, the other important source of copper in drinking water is the corrosion of plumbing materials. As in lead, the concentration is connected to the household age and the variable of the percentage of households built before 1960 is a good indicator, a higher concentration of old households was identified in tracts where less population is Hispanic compared with whites, explaining the negative significant coefficient of percent of Hispanics in the copper model. Besides, a significant positive coefficient was estimated for the percent of American Indians in the copper model, Appendix 2.2.f has the concentration of copper in tracts where most of the population is American Indian and in tracts where that is not the case, comparing the means of both groups there is a significant difference between the average concentration of copper confirming the positive coefficient for Hispanics in the copper model. Finally, Santa Cruz has a positive significant coefficient, the average concentrations per county (Appendix 2) show that the exposure in Santa Cruz is higher than in Maricopa.

6.2 Logit panel data model for Violations

From the significant variables identified in Chapter 4 for the logit model for violations of the arsenic standard and lead and copper rule, in this section, the results of the marginal effects of the binary dependent models are included. Table 14 includes the marginal effects for the significant coefficients per standard.

Table 14. Marginal Effects binary dependent variable models

	Delta-method.Arsenic		Delta-method. Lead &Copper	
	dy/dx	se	dy/dx	Se
Sizecat			-0.0565235***	0.0097431
PERCENT_CHILD	0.2004275*	0.1066505		
PERCENT_HISPANIC			0.1174819*	0.055801
PERCENT_BLACK	-0.6639663***	0.2367916		
PERCENTAGE_POVERTY	0.0019039***	0.0006509		
PERCENT_ELDERLY	-0.0892388*	0.0407833	0.1167223**	0.0486929
d_APACHE			0.2567041***	0.0536108
d_COCONINO			0.0959757*	0.0442403
d_GILA	-0.1813183***	0.065212		
d_GREENLEE			0.2501837***	0.0748385
d_LAPAZ			0.1775139***	0.0518857
d_MOHAVE			0.1956585***	0.037337
d_NAVAJO	-0.160842***	0.0605974	0.1972176***	0.0404322
d_PIMA	-0.1452901***	0.037735		
d_PINAL			0.0843948**	0.0356475
d_YUMA			0.2279227***	0.0436664

Note:
*p<0.05; **p<0.01;
***p<0.001

6.2.1 Arsenic Standard

The marginal effects of the model for violations of arsenic standard show a negative association between age and pollution, for tracts where the higher percentage of the population is elderly people the arsenic exposure is lower and for tracts with a higher fraction of children the concentration is higher. Appendix 4.1.a shows the relationship between variables that accounts for the age of the population and violations of arsenic standard, in the group of violations the percentage of children is higher, and the percentage of elderly people is lower compared with the group of non-violations. Besides, there is a significant difference in the means between both groups. In summary, if the proportion of elderly people increases by 1%, the probability that an arsenic violation occurs decreases by 0.0009 on a [0,1] scale while if that happens for the percent of children the probability increases by 0.002. For public health policies, this finding is a concern because the association of the children population with arsenic violations may be associated with a higher exposure of this segment of the population. Early childhood exposure has been linked to negative impacts on cognitive development (WHO, 2023).

The percentage of black people is the only significant variable of race/ethnicity in the logit model, there is a negative coefficient for this variable. This is associated with lower exposure to arsenic in the tracts where a higher proportion is black. Appendix 4.1.b shows the connection between violations and percent of black people, there is a significant difference in the means in the group of violations the percentage of black people is 1.5%, while in the group of non-violation has an average of 2.2%.

As same as in the Tobit arsenic model there is a positive association between poverty and pollution if the percentage of poverty increases by 1% the effect in violation of arsenic occurrence increases by 0.002.

There is a significant difference in the occurrence of violations in arsenic standards between counties, Appendix 4.1.c shows the percentage of samples with violations per county. Gila, Navajo, and Pima have samples with a significantly lower number of violations than Maricopa. In Gila according to the model, when a sample is taken the probability that an arsenic violation is registered is lower by 18%, in Navajo and Pima the probability of a violation drops by 16% and 14% respectively.

6.2.2 Lead & Copper Rule

Given that according to the Lead and Copper Rule, exceeding the action level is not a violation and in Chapter 4 the identification of the differences between violations and high concentration levels in the database was identified. One of the issues in the interpretation of the probit model for the Lead and Copper Rule is the identification of the cause of the violation. For example, if the cause is that the services are not taking samples or the samples are collected improperly, or the system is not replacing lead service lines, these assumptions may expand the exposure to lead, and copper identified in the Tobit model. On the other side, if the violation is registered for the absence of public education, the interpretation may be different.

A negative significant coefficient is associated with the size of CWS, this may be associated with the absence of treatment or samples in the system with a small size that can conclude in a violation of the standard. The percentage of violations of the arsenic standard per size differs between CWS of large (13.55%) or very large size (9.46%) and CWS of small (19.47%) or very small size (22.08%).

The percentage of Hispanics is a significant positive variable of race/ethnicity in the arsenic violations model. This relation is associated with the higher percentage of Hispanics in the group of violations compared with the group of non-violations and the significant difference between them. Comparing these results with the distribution in white communities where the group of violations has a lower percentage of white people than the group of non-violations. In Appendix 4.2.a the distribution of percentage of whites and Hispanics across the group of violation and non-violation of lead and copper rule is included. If the percentage of Hispanics increases by 1% the effect in violation of lead and copper occurrence increases by 0.12.

Another significant variable in the probit model for lead and copper is the proportion of elderly people. Finally, most of the counties' dummy variables have significant positive coefficients as a consequence of the low percentage of violations in Maricopa (baseline). The graph with the percent of violations per county (Appendix 4.2.c) shows that Greenlee, Yuma, and Apache are the counties with the highest presence of violations with more than 40% of violation occurrence in the analyzed data and have also a significant difference with Maricopa county. As a result, if the CWS is in Greenlee, Yuma, and Apache the effects in violation of lead and copper occurrence increase by 0.25, 0.26, and 0.22 respectively.

Chapter 7: Conclusion and next steps

This study addressed an original mix in the analysis of drinking water chemical contamination. First, it included the analysis of violations of arsenic standard, and lead and copper rule; on the other side, it shows the analysis of the concentrations of the relevant contaminants using an environmental justice approach at a state level.

The study found that there were disparities in the exposure to lead and copper in drinking water across the population of Arizona. The violations were more common in community water systems serving small populations of less than 3,300 inhabitants. These findings were consistent with those of McDonald & Jones (2018) in the U.S. The use of census tracts to analyze socio-economic characteristics coupled with the average population served by the CWS showed a significant association between contaminant exposure or violations and socio-economic characteristics. Additionally, the study found that Apache county had the highest violations, which is consistent with the high proportion of CWS with small populations in the county. In terms of lead concentrations, Pinal and La Paz counties had higher exposure, while Santa Cruz was the most affected county in terms of copper. These results further highlight the spatial differences in contaminant concentrations in Arizona, as reported by Morata et al. (2022).

Furthermore, for lead concentrations, one of the main causes for the presence of this contaminant in drinking water is the existence of lead pipelines, specifically in old constructions. Using the percentage of households built before 1960 as a proxy, a positive association between this variable and communities with high rates of poverty was identified. Likewise, for the lead concentration model, the coefficient of black people was positive given that the highest concentration level of lead were identified in Pinal County, which has a superior proportion of black people compared to other counties in Arizona. For the model of violations of lead and copper, a significant positive association was established with the proportion of Hispanics; however, the cause of these violations was not associated with higher concentration levels given that the model for concentrations in lead and copper exposed negative significant coefficients. In addition, in communities with an elevated proportion of American Indians, the concentrations of copper were higher.

On the other hand, for arsenic, this study shows that there is a significant association between CWS using groundwater as their primary source and higher concentration levels of arsenic, which confirms the finding that there is a higher concentration of arsenic in groundwater at a level above health-based drinking water standard in Arizona. Besides, this research explored the differences in the concentrations and violations across the counties in Arizona. The violations of the rules appeared principally in Greenlee and Yuma for arsenic. On the other hand, for the models of concentrations, the metropolitan areas (Pima and Maricopa counties) have lower concentrations compared to non-metropolitan areas.

Previous environmental justice studies have identified a significant association between poverty and pollution (Banzhaf, Ma, & Timmins, 2019). This research confirms the relationship between poverty and water pollution, as evidenced by the violations model for arsenic and the concentrations model for arsenic and lead. This association may be attributed to the intergenerational poverty effect, where communities with higher poverty rates have lower access to education. We used the percentage of the population with less than a high school diploma as a relevant variable to assess education access. With this identification in the arsenic models, we found a significant association which is a huge concern in EJ studies, this relationship was not identified in the study of Cory & Rahman, 2009, meaning that probably the increase in the income inequality in Arizona between 2002 and 2020 may produce a higher disproportionate exposure in arsenic on low income communities.

Overall, our results highlight significant racial and ethnic disparities in Arizona. In the model of violations and concentrations in arsenic, the proportion of black communities had a negative coefficient, suggesting that areas with a higher proportion of black people had lower levels of arsenic. This is likely due to the fact that black communities are predominantly located in urban areas where the presence of arsenic is lower. In contrast, the percentage of Hispanics in a community emerged as a driver for higher concentrations of arsenic, linked with the elevated rate of poverty in locations where the greatest percent of the population is Hispanic compared to white communities. Though, for ethnic/racial disparities we ratified the previous findings of Cory & Rahman, 2009 about the positive association of higher concentration in hispanics communities and negative associations with black communities meaning that the effect identified with information between 2002 to 2004 is still significant using information between 2009 to 2022.

The analysis revealed that the proportion of elderly people in a community had a statistically negative coefficient for arsenic models given that this segment of the population has a lower rate of poverty. On the other side, the percentage of children in the tract had a significative positive coefficient. This finding is relevant for public health implications, such as in the case of arsenic, the exposure to the contaminant in early childhood may cause impacts on their cognitive development.

However, the lower exposure in elderly people reduced health outcomes in this segment of the population such as diabetes, pulmonary disease, cardiovascular disease, and skin cancer.

These findings may help to understand how environmental justice and drinking water quality are linked, and that arsenic is the biggest problem for chemical contamination in drinking water in Arizona. The latter is supported by the elevated concentration levels which were on average 5.35 ug/L, being higher than 2 ug/L - the average level for the U.S.- for the analyzed database between 2009 and 2022. Also, this exposure is more concentrated in non-urban areas of the state presenting a challenge of investment and execution for policymakers because these areas have a higher concentration of small and very small CWS and the population is more spread in these territories.

This study reafirms implications about the racial disparities in the exposure to contaminants such as arsenic in drinking water. In general, the groups with the higher exposure are hispanics for arsenic, black people for lead and american indians for copper, policymakers can target resources and interventions to these communities, ensuring that they receive the assistance they need to adress the problem such as water treatment, pipelines change or educational programs.

In addition, in the future, environmental justice studies focused on contamination of drinking water can enhance the development of equitable policies. Communities with a higher proportion of people of color and/or low income, which are typically the populations most exposed to pollution, can be targeted by policies aimed at alleviating these disparities and achieving equity for the population of interest. One way to start would be to develop a state-level database where communities can report if their areas are not included, such as community water systems that are not part of the active database of the Arizona Department of Environmental Quality (AZDEQ).

There are two principal limitations in this study. First, there is a discrepancy between the results for the lead and copper rule violations model and the Tobit model for concentrations of these contaminants. The reason is that in the database of ADEQ, there is an absence of an explanation for the violation, and there were various reasons for a violation in this rule that were not associated with a concentration higher than the standard limit. For further analysis, it is relevant to access a database to establish the cause of the violation and to link it with the exposure. Second, the appropriate model for the analysis of the concentrations of the contaminants was a Tobit Model. Although, this approach uses a unique censored level. In this study, various limits of detection impose that part of the censored data was not limited to this threshold. Besides, for the model of lead concentrations, almost 60% of the samples were censored which was significantly higher than copper and arsenic cases.

In interpreting the results of this study, there are some potential issues to consider. The sorting of people within community water systems could have implications for the interpretation of the results and conclusions drawn from the aggregated data. People with similar characteristics tend to cluster together in certain areas, which can create biases in the results. For example, in a CWS serving a particular area, the proportion of people from a certain race and income may be similar, which could lead to an overestimation of the impact of race on water chemical contamination. Conversely, in very large CWSs, there may be huge disparities in the characteristics of the served population, and at the aggregated level, some differences may not be clear, resulting in an underestimation of the impacts.

Finally, the initiative Justice 40 invested 40% of the federal resources to reduce the disparities of marginal communities across the U.S. This program has been helping to create more available data related to environmental justice and facilitate the identification of disproportionate exposures to pollutants. As a next step of this research, spatial analysis can be added to account for spatial dependency using spatial lag and spatial weighted regression models, while a reduction to a lower unit basis can be done given that the census data is available at a block level since 2020.

References

ADEQ. (2020). *Source Water Protection.* Retrieved from https://www.azdeq.gov/node/735

ADEQ. (2023, February 10). *Safe Drinking Water*. Retrieved from http://www.azdeq.gov/node/242

ADEQ, A. D. (2021, July). *Lead and Copper Rule Revisions (LCRR) Resources, Updates & Focus Group Information*. Retrieved from https://azdeq.gov/node/8705

ADEQ, A. D. (2023, February 10). *Surface Water Quality Remedation Sites*. Retrieved from https://azdeq.gov/LegacyMines#:~:text=Their%20historic%20mine%20workings%20can,and%20the%20State%20are%20exceeded.

Almond, D., & Curie, J. (2011). Killing me softly: the fetal origins hypothesis. *Economics Perspectives*, 153-172.

Almond, D., Curie, J., & Duque, V. (2017). Childhood circumstances and adult outcomes. *NBER Work*.

Alvarez, C., & Evans, C. (2021). Intersectional environmental justice and population health inequalities: A novel approach. *Social Science and Medicine*.

Amemiya, T. (1984). Tobit Models: A Survey. *Journal of Econometrics*, 3-61.

Arizona Department of Water Resources. (2022, March 28). *CWS Service Area.* Retrieved from Arizona Deapartament Water Resources GIS Data: https://gisdata2016-11-18t150447874z-azwater.opendata.arcgis.com/datasets/cws-service-area-1/explore?location=34.114259%2C-111.993688%2C7.18

Arizona Department of Water Resources. (2023). *Drought Status*. Retrieved from https://new.azwater.gov/drought/drought-status

ASU News. (2022, November 15). The Future of Water in Arizona.

Balazs, C., Morello, R., & Ray, A. H. (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*.

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data. Third Edition.* John Wiley & Sons, Ltd.

Banzhaf, S., & Randall, W. (2013). Segregation and Tiebout Sorting: The Link Between Place-based Investments and Neighborhood Tipping. *Journal of Urban Economics*, 83-98.

Banzhaf, S., Ma, L., & Timmins, a. C. (2019). Environmental Justice: The Economics of Race, Place and Pollution. *Journal of Economic Perspectives*, 185-208.

Banzhaf, S., Ma, L., & Timmins, C. (2019). Environmental Justice: Establishing Causal Relationships. *Annual Review of Resource Economics*, 377-398.

Banzhaf, S., Ma, L., & Timmins, C. (2019). Environmental Justice: The Economics of Race, Place and Pollution. *Journal of Economic Perspectives*, 185-208.

Bayer, A., & Wilcox, D. W. (2019). The unequal distribution of economic education: A report on the race, ethnicity, and gender of economics majors at U.S. colleges and universities. *The Journal of Economic Education*.

Becker, G. (1957). The economics of discrimination. *University of Chicago Press*.

Chakraborty, L., Rus, H., Henstra, D., & Minano, J. T. (2022). Exploring spatial heterogeneity and environmental injustices in exposure to flood hazards using geographically weighted regression. *Environmental Research*.

Cory, D., & Rahman, T. (2009). Environmental justice and enforcement of the safe drinking water act: The Arizona arsenic experience. *Ecological Economics*, 1825-1837.

Curie, J. (2011). Inequality at birth: some causes and consequences. *American Economics Reviews*, 1-22.

EPA. (2019). *Understanding the Lead and Copper Rule.*

EPA. (2022, November). Retrieved from Chemical Contaminant Rules: https://www.epa.gov/dwreginfo/chemical-contaminant-rules

EPA. (2022). *National Primary Drinking Water Regulations*. Retrieved from https://www.epa.gov/sites/default/files/2016-06/documents/npwdr_complete_table.pdf

EPA. (2023, March 3). *EJScreen: Environmental Justice Screening and Mapping Tool.* Retrieved from https://www.epa.gov/ejscreen/ejscreen-map-descriptions

EPA. (2023, January 27). *General Information about Lead in Drinking Water.* Retrieved from https://www.epa.gov/ground-water-and-drinking-water/basic-information-about-lead-drinking-water#getinto

EPA, U. S. (2022, November 15). *Information about Public Water Systems*. Retrieved from https://www.epa.gov/dwreginfo/information-about-public-water-systems

EPA, U. S. (2022, July 11). *Justice 40 at EPA*. Retrieved from https://www.epa.gov/environmentaljustice/justice40-epa

Federal Reserve Bank. (2020). *Credit and Income Inequality.* New York.

Foster, S., Pennino, M., Compton, J., Leibowitz, S., & Kile, M. (2019). Arsenic Drinking Water Violations Decreased across the United States following Revision of the Maximum Contaminant Level. *Environmental Science and Technology*, 11478 - 11485.

Frees, E. W. (2004). *Longitudinal and Panel Data. Analysis and Applications in the social sciences.* Cambridge University Press.

Hamilton, J. (1995). Testing for Environmental Racism: Prejudice, Profits, Political Power? *Journal of Policy Analysis and Management*, 107-132.

James, I. (2020, December 14). *'We need water to survive': Hopi Tribe pushes for solutions in long struggle for water.* Retrieved from AZ Central: https://www.azcentral.com/in-depth/news/local/arizona-environment/2020/12/14/hopi-tribe-pushes-solutions-many-without-clean-drinking-water/3731341001/

James, T., Evans, A., & Evans, E. M. (2020). The Economic Importance of Groundwater in Arizona, 2010-2018. *Arizona State University. Seidman Research Institute*.

Knappenberger, R. (2022, September 6). *Cronkite News.* Retrieved from Federal funds for water projects a fraction of what Arizona says it needs.

Kristine Uhlman, C. R. (2009). *Arizona Drinking Water Well Contaminants.* Arizona Cooperative Extension.

Lievanos, R. (2019). Air toxic clusters revisited : intersectional environmental inequalities and indigenous deprivation in the U.S. Environmental Protection agency regions. *Race soc*, 161-184.

McDonald, Y., & Jones, N. (2018). Drinking water violations and environmental justice in the United States, 2011-2015. *American Journal of Public Health*, 1401-1407.

Morata, I. M., Bostick, B., Conroy-Ben, O., Duncan, D., Jones, M., & Spaur, M. (2022). Nationwide geospatial analysis of county racial and ethnic composition and public drinking water arsenic and uranium. *Nature Communications*.

Musher, C. (2001). Predicting drug arrest rates: con ict and social dis-organization perspectives. *Crime Delinquency*.

Navajo Hopi Observer. (2022, July 19). *Hopi Tribe celebrates opening of arsenic mitigation project.* Retrieved from NHO: https://www.nhonews.com/news/2022/jul/19/hopi-tribe-celebrates-opening-arsenic-mitigation-p/

Nigra, A. E., Chen, Q., Chillrud, S. N., Wang, L., Harvey, D., Mailloux, B., & Navas-Acien, P. F.-L. (2020). Inequalities in Public Water Arsenic Concentrations in Counties and Community Water Systems across the United States, 2006–2011. *Environmental Health Perspectives*.

Nigra, A. E., Chen, Q., Chillrud, S. N., Wang, L., Harvey, D., Mailloux, B., . . . Navas-Acien, A. (2020). Inequalities in Public Water Arsenic Concentrations in Counties and Community Water Systems across the United States, 2006-2011. *Environmental Health Perspectives*.

Nigra, A., Cazacu-Deluca, A., & Navas-Acien, A. (2022). Socioeconomic vulnerability and public water arsenic concentrations across the US. *Environmental Pollution*.

Office of Policy Development and Research . (2005). *Homeownership Gaps Among Low-Income and Minority Borrower and Neighborhood.* Cambridge, MA: Abt Associates Inc.

Pace, C., Balazs, C., Bangia, K., Depsky, N., Renteria, A., Morello-Frosch, R., & Cushing, L. J. (2021). Inequities in Drinking Water Quality Among Domestic Well Communities and Community Water Systems, California, 2011–2019. *American Journal of Public Health*.

R, Q., FA, A., DK, E., I, L., & Clarke E, M. K. (2015). Association of arsenic with adverse pregnancy outcomes/infant mortality: a systematic review and meta-analysis. *Environmental Health Perspectives*, 412-421.

Rahman, D. C. (2009). Environmental justice and enforcement of the safe drinking water act: The Arizona arsenic experience. *Ecological Economics*, 1825-1837.

Schaider, L., Swetschinski, L., Campbell, C., & Rudel, R. (2019). Environmental justice and drinking water quality: are there socioeconomic disparities in nitrate levels in U.S. drinking water? *Environmental Health*.

Smith, C. (2007). Economic deprivation and environmental inequality in Postindustrial Detroit: a comparison of land 11 and superfund locations. *Organ Environ*, 25-43.

State of Arizona. (2020). *Analysis of Impediments to fair Housing Choice*. Retrieved from https://housing.az.gov/sites/default/files/documents/files/AZ-2020-Analysis-of-Impediments-to-Fair-Housing-Choice_0.pdf

The White House. (2022). *Justice 40*. Retrieved from https://www.whitehouse.gov/environmentaljustice/justice40/

U.S Census Bureau. (2020). *Understanding and Using American Community Survey Data.*

U.S Census Bureau. (2022, April 11). *Glossary*. Retrieved from https://www.census.gov/programs-surveys/geography/about/glossary.html#par_textimage_13

U.S Department of the Treasury. (2022, November 4). *Racial Differences in Economic Security: Housing.* Retrieved from https://home.treasury.gov/news/featured-stories/racial-differences-in-economic-security-housing

U.S Geological Survey. (2012). *Maps of Estimated Nitrate and Arsenic Concentrations for Basin-Fill Aquifers of the Southwestern United States.* National Water-Quality Assessment Program.

Uhlman, K., Rock, C., & Artiola, J. (2009). *Arizona Drinking Water Well Contaminants.* Arizona Cooperative Extension.

USGS. (2022). *Active Mines in Arizona.* Retrieved from https://uagis.maps.arcgis.com/apps/webappviewer/index.html?id=9eceb192cd86497e8eed04113302db8b

Voorheis, J. (2017). Air quality, human capital formation and the long-term effects of environmental inequality at birth. *US Census Bureau*.

WHO, W. H. (2023). *Arsenic. Key Facts.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/arsenic#:~:text=Long%2Dterm%20exposure%20to%20arsenic,increased%20deaths%20in%20young%20adults.

Williams, R., & Jorgensen, A. (2023). Comparing logit & probit coefficients between nested models. *Social Science Research*.

Appendix 1. Results regression models.

Table 1. OLS Results

	Dependent variable:		
	Arsenicmean	copperMean	LeadMean
	(1)	(2)	(3)
PERCENT_HISPANIC	1.390**	-0.076***	-0.0002
	(0.584)	(0.015)	(0.001)
PERCENT_BLACK	-9.208***	-0.057	0.006*
	(2.704)	(0.077)	(0.004)
PERCENT_AMERICANINDIAN	-0.019	0.199***	0.0001
	(1.302)	(0.032)	(0.002)
PERCENTAGE_POVERTY	-0.005	0.001***	0.00001
	(0.011)	(0.0003)	(0.00001)
Income	0.00001**	0.00000***	-0.000
	(0.00000)	(0.00000)	(0.000)
PERCENT_Elderly	-0.946*	0.008	-0.0003
	(0.573)	(0.015)	(0.001)
PERCENT_CHILD	0.454	-0.110**	-0.007**
	(2.007)	(0.054)	(0.003)
PERCENT_FEMALELED	-0.875	0.042	-0.003
	(1.425)	(0.040)	(0.002)
PERCENT_Ownership_rate	2.098***	0.012	0.001
	(0.469)	(0.013)	(0.001)
d_groundwater	1.359***	-0.048***	0.0001
	(0.273)	(0.009)	(0.0004)
Sizecat	-0.095	0.007***	0.0002*
	(0.072)	(0.002)	(0.0001)
d_APACHE	-3.312***	-0.015	-0.001**
	(0.692)	(0.014)	(0.001)
d_COCHISE	-1.252***	-0.014	-0.0005
	(0.349)	(0.009)	(0.0004)
d_COCONINO	-1.291***	0.022**	0.0001
	(0.395)	(0.011)	(0.001)
d_GILA	-1.895***	0.025**	0.0002
	(0.406)	(0.010)	(0.0005)
d_GRAHAM	-1.354*	-0.007	-0.0001
	(0.776)	(0.019)	(0.001)
d_GREENLEE	-0.343	0.072***	-0.0001
	(0.876)	(0.018)	(0.001)
d_LAPAZ	0.063	0.056***	0.0004
	(0.509)	(0.013)	(0.001)

	Dependent variable:		
	Arsenicmean	copperMean	LeadMean
	(1)	(2)	(3)
d_MOHAVE	0.321	-0.0001	-0.001**
	(0.352)	(0.010)	(0.0005)
d_NAVAJO	-1.626***	0.003	-0.001
	(0.430)	(0.010)	(0.001)
d_PIMA	-1.755***	-0.009	-0.0004
	(0.246)	(0.007)	(0.0003)
d_PINAL	0.647**	-0.003	0.001***
	(0.276)	(0.009)	(0.0004)
d_SANTACRUZ	-0.884*	0.040***	0.0001
	(0.487)	(0.014)	(0.001)
d_YAVAPAI	0.763***	0.045***	0.001
	(0.270)	(0.008)	(0.0004)
d_YUMA	2.725***	0.046***	0.001**
	(0.363)	(0.011)	(0.001)
Superfund	-0.979**	0.017	-0.0001
	(0.460)	(0.016)	(0.001)
Miningsite	-0.006	-0.001	-0.001*
	(0.277)	(0.009)	(0.0004)
Constant	2.671***	0.080***	0.003***
	(0.609)	(0.018)	(0.001)
Observations	4,054	4,005	4,005
R2	0.092	0.100	0.020
Adjusted R2	0.086	0.094	0.013
Residual Std. Error	4.232 (df = 4026)	0.114 (df = 3977)	0.006 (df = 3977)
F Statistic	15.160*** (df = 27; 4026)	16.445*** (df = 27; 3977)	2.935*** (df = 27; 3977)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2. Panel Data Model Results

	Dependent variable:					
	Arsenicmean	ArsenicMax	LeadMean	LeadMax	copperMean	copperMax
	(1)	(2)	(3)	(4)	(5)	(6)
PERCENT_HISPANIC	0.152*	1.223*	-0.0001	0.002	-0.079***	-0.117***
	(0.731)	(1.264)	(0.001)	(0.007)	(0.029)	(0.154)
PERCENT_BLACK	-8.720***	-16.474***	0.005	0.037	-0.103	-0.127
	(3.108)	(5.566)	(0.004)	(0.038)	(0.112)	(0.586)
PERCENT_AMERICANINDIAN	-2.204	0.832	0.0002	-0.001	-0.092	-0.242
	(1.887)	(3.151)	(0.002)	(0.016)	(0.091)	(0.476)
PERCENTAGE_POVERTY	0.008	0.016	0	0.0002	0.0004	-0.001
	(0.01)	(0.018)	(0.00001)	(0.0001)	(0.0003)	(0.002)
Income	0	0	0	0	-0.00000*	0
	(0.00001)	(0.00001)	(0)	(0)	(0)	(0)
PERCENT_Elderly	-1.972***	-2.308**	-0.0004	-0.002	-0.027*	-0.018*
	(0.545)	(1.006)	(0.001)	(0.007)	(0.016)	(0.082)
PERCENT_CHILD	0.662	1.458	-0.007**	-0.01*	-0.095*	-0.036*
	(1.657)	(3.119)	(0.003)	(0.026)	(0.049)	(0.255)
PERCENT_FEMALELED	-0.207	0.186	-0.003	-0.012	-0.028	-0.169
	(1.254)	(2.343)	(0.002)	(0.019)	(0.036)	(0.19)
PERCENT_Ownership_rate	1.954***	1.830**	0.001	0.006	0.032**	0.128*
	(0.473)	(0.867)	(0.001)	(0.006)		
d_groundwater	1.313**	1.653*	0.0001	-0.00004		
	(0.548)	(0.847)	(0.0004)	(0.004)		
Sizecat	-0.004	0.364*	0.0002	0.006***		
	(0.138)	(0.215)	(0.0001)	(0.001)		
d_APACHE	-3.151***	-4.828***	-0.001**	-0.01		
	(0.998)	(1.611)	(0.001)	(0.007)		
d_COCHISE	-2.181***	-3.006***	-0.001	-0.008*		
	(0.579)	(0.917)	(0.0005)	(0.004)		
d_COCONINO	-1.819***	-2.617**	0.0001	-0.003		
	(0.683)	(1.081)	(0.001)	(0.005)		
d_GILA	-2.884***	-3.151***	0.0002	0.001		
	(0.665)	(1.062)	(0.001)	(0.005)		
d_GRAHAM	-2.262*	-3.454	-0.0001	-0.01		
	(1.351)	(2.124)	(0.001)	(0.009)		
d_GREENLEE	-1.581	-1.292	-0.0002	-0.006		
	(1.435)	(2.295)	(0.001)	(0.009)		
d_LAPAZ	-0.884	-1.216	0.0003	-0.004		
	(0.876)	(1.386)	(0.001)	(0.007)		
d_MOHAVE	-0.551	-1.161	-0.001**	-0.005		
	(0.608)	(0.968)	(0.001)	(0.005)		

	Dependent variable:					
	Arsenicmean	ArsenicMax	LeadMean	LeadMax	copperMean	copperMax
	(1)	(2)	(3)	(4)	(5)	(6)
d_NAVAJO	-2.247***	-3.333***	-0.001	-0.008		
	(0.665)	(1.066)	(0.001)	(0.005)		
d_PIMA	-1.971***	-2.831***	-0.0004	-0.008**		
	(0.449)	(0.703)	(0.0004)	(0.004)		
d_PINAL	0.201	-0.276	0.001**	0.005		
	(0.543)	(0.845)	(0.0004)	(0.004)		
d_SANTACRUZ	-0.819	-1.652	-0.00004	-0.006		
	(0.882)	(1.382)	(0.001)	(0.007)		
d_YAVAPAI	0.189	1.993**	0.001	-0.001		
	(0.507)	(0.799)	(0.0004)	(0.004)		
d_YUMA	0.92	1.171	0.001*	-0.005		
	(0.698)	(1.092)	(0.001)	(0.005)		
Superfund	-1.064	-1.046	-0.0001	0.005		
	(0.975)	(1.501)	(0.001)	(0.008)		
Miningsite	0.183	0.258	-0.001*	-0.010**		
	(0.576)	(0.888)	(0.0005)	(0.005)		
Constant	3.553***	3.771**	0.004***	0.0002		
	(0.893)	(1.474)	(0.001)	(0.009)		
Observations	4054	4054	4005	4005	4005	4005
R2	0.013	0.003	0.016	0.018	0.007	0.002
F Statistic	123.628***	115.312***	70.394***	73.430***	2.569***	0.615***
Note:				*p<0.1;	**p<0.05;	***p<0.01

Table 3. Tobit Panel Data Model Results

	Dependent variable:		
	ArsenicCensored	CopperCensored	LeadCensored
	(1)	(2)	(3)
PERCENT_HISPANIC	0.468*	-0.155**	-0.00192*
	(0.61)	(-3.15)	(-2.04)
PERCENT_BLACK	-9.661**	-0.0607	0.0086
	(-2.88)	(-0.24)	(1.79)
PERCENT_AMERICAN INDIAN	-3.384	0.2	-0.00138
	(-1.61)	(1.85)	(-0.66)
PERCENTAGE POVERTY	0.0134*	-0.0000602	0.0000599***
	(1.31)	(-0.07)	(3.75)
PERCENT_ELDERLY	-2.311***	0.0192	-0.000889
	(-3.95)	(0.4)	(-1.00)
PERCENT_CHILDREN	0.753	0.00792	-0.00359
	(0.42)	(0.05)	(-1.14)
PERCENT_FEMALELED	-0.57	0.128	-0.00335
	(-0.43)	(1.01)	(-1.44)
PERCENT_Homeownership	2.123***	-0.0084	0.00172*
	(4.26)	(-0.21)	(2.26)
d_groundwa~r	1.352*	-0.0375	0.000271
	(2.25)	(-1.31)	(0.5)
Sizecat	0.0733	0.0231**	0.00104***
	(0.53)	(3.17)	(7.35)
d_APACHE	-3.774***	-0.0423	-0.00104
	(-3.44)	(-0.89)	(-1.11)
d_COCHISE	-2.910***	-0.00755	0.000196
	(-4.66)	(-0.26)	(0.34)
d_COCONINO	-2.089**	0.0159	0.000857
	(-2.79)	(0.45)	(1.24)
d_GILA	-3.677***	0.0291	0.00146*
	(-5.13)	(0.89)	(2.3)
d_GRAHAM	-3.541*	-0.0166	0.00016
	(-2.33)	(-0.27)	(0.13)
d_GREENLEE	-1.727	0.0808	0.00151
	(-1.10)	(1.29)	(1.22)
d_LAPAZ	-0.799	0.0476	0.00216**
	(-0.86)	(1.09)	(2.58)
d_MOHAVE	-0.576	-0.0211	0.000586
	(-0.90)	(-0.67)	(0.97)

	Dependent variable:		
	ArsenicCensored	CopperCensored	LeadCensored
	(1)	(2)	(3)
d_NAVAJO	-2.874***	0.000582	0.00031
	(-3.97)	(0.02)	(0.47)
d_PIMA	-2.395***	-0.0122	-0.0000677
	(-4.92)	(-0.52)	(-0.15)
d_PINAL	0.244	-0.00361	0.00150**
	(0.42)	(-0.13)	(2.72)
d_SANTACRUZ	-0.915	0.251***	0.00134
	(-0.95)	(5.51)	(1.49)
d_YAVAPAI	0.381	0.0401	0.000981
	(0.72)	(1.5)	(1.87)
d_YUMA	0.734	0.0543	0.000856
	(0.97)	(1.5)	(1.2)
Superfund	0.0521	-0.000197	
	(0.99)	(-0.20)	
Miningsite	-0.00517	-0.00124*	
	(-0.17)	(-2.13)	
_cons	2.835***	0.0795	-0.00260**
	(3.35)	(1.59)	(-2.73)
Observations	4,054	4,005	4,005
sigma_u	3.027***	0.0541***	0.00160***
	-27.36	-6.46	-10.12
sigma_u	3.348***	0.357***	0.00611***

Note: *p<0.05; **p<0.01; ***p<0.001

Table 4. Binary Dependent Panel Data Model Results

	Dependent variable:	
	Violation Arsenic	Violation L&C
	(1)	(2)
PERCENT_HISPANIC	-1.078	0.526*
	(-1.28)	(2.1)
PERCENT_BLACK	-10.98**	1.898
	(-2.82)	(1.49)
PERCENT_AMERICAN INDIAN	-3.641	0.466
	(-1.39)	(0.83)
PERCENTAGE POVERTY	0.0315**	-0.00736
	(2.92)	(-1.82)
PERCENT_ElDERLY	-1.476*	0.523*
	(-2.20)	(2.39)
PERCENT_CHILDREN	3.314	0.873
	(1.88)	(1.17)
PERCENT_FEMALELED	-0.692	0.489
	(-0.51)	(0.86)
PERCENT_Homeownership	1.337*	0.0416
	(2.42)	(0.21)
d_groundwa~r	0.497	0.0684
	(0.74)	(0.4)
Sizecat	0.0991	-0.253***
	(0.59)	(-5.76)
d_APACHE	0	1.150***
	(.)	(4.74)
d_COCHISE	-1.858*	0.285
	(-2.28)	(1.73)
d_COCONINO	-0.949	0.430*
	(-1.16)	(2.17)
d_GILA	-2.998**	-0.0483
	(-2.76)	(-0.25)
d_GRAHAM	-0.679	0.5
	(-0.41)	(1.47)
d_GREENLEE	0.95	1.121***
	(0.59)	(3.33)
d_LAPAZ	-2.443	0.795***
	(-1.76)	(3.41)
d_MOHAVE	-0.345	0.877***
	(-0.51)	(5.19)

	Dependent variable:	
	Violation Arsenic	Violation L&C
	(1)	(2)
d_NAVAJO	-2.660**	0.884***
	(-2.65)	(4.83)
d_PIMA	-2.402***	0.000723
	(-3.84)	(0.01)
d_PINAL	0.469	0.378*
	(0.8)	(2.37)
d_SANTACRUZ	0.377	0.337
	(0.4)	(1.33)
d_YAVAPAI	0.417	-0.0407
	(0.79)	(-0.26)
d_YUMA	1.394	1.021***
	(1.87)	(5.16)
Superfund	0.4	-0.102
	(0.36)	(-0.29)
Miningsite	-0.919	-0.158
	(-1.26)	(-0.81)
_cons	-4.650***	-1.273***
	(-4.91)	(-4.76)
Observations	4,054	4,005
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 5. Marginal Effects Tobit Model

	Delta-method.Arsenic		Delta-method. Lead		Delta-method. Copper	
	dy/dx	se	dy/dx	se	dy/dx	se
d_groundwater	1.055929**	0.47	0.0001	0.00	-0.0214	0.02
Sizecat	0.0549	0.12	0.0005244***	0.00	0.0131351***	0.00
PERCENT_FEMALELED	-0.4635	1.04	-0.0017	0.00	0.0731	0.07
PERCENT_CHILD	0.5831	1.38	-0.0018	0.00	0.0045	0.10
PERCENT_HISPANIC	0.4064237*	0.60	-0.000972*	0.00	-0.0885426**	0.03
PERCENT_BLACK	-7.492571**	2.61	0.0043522*	0.00	-0.0346	0.15
PERCENT_AMERICAN_INDIAN	-2.5706	1.63	-0.0007	0.00	0.1139	0.06
PERCENTAGE_POVERTY	0.0105559*	0.01	0.0000303***	0.00	0.0000	0.00
PERCENT_ELDERLY	-1.798036***	0.46	-0.0004	0.00	0.0109	0.03
PERCENT_HomeOwnership	1.658266***	0.39	0.0008705*	0.00	-0.0048	0.02
d_APACHE	-2.966467***	0.86	-0.0005	0.00	-0.0241	0.03
d_COCHISE	-2.290961***	0.49	0.0001	0.00	-0.0043	0.02
d_COCONINO	-1.621089***	0.58	0.0004	0.00	0.0091	0.02
d_GILA	-2.865379***	0.57	0.0007408*	0.00	0.0166	0.02
d_GRAHAM	-2.815577***	1.19	0.0001	0.00	-0.0094	0.04
d_GREENLEE	-1.3621	1.23	0.0008	0.00	0.0460	0.04
d_LAPAZ	-0.6275	0.73	0.0010946*	0.00	0.0271	0.02
d_MOHAVE	-0.4488	0.50	0.0003	0.00	-0.0120	0.02
d_NAVAJO	-2.256671***	0.57	0.0002	0.00	0.0003	0.02
d_PIMA	-1.880667***	0.38	0.0000	0.00	-0.0070	0.01
d_PINAL	0.2030	0.45	0.000757**	0.00	-0.0021	0.02
d_SANTACRUZ	-0.6857	0.75	0.0007	0.00	0.143118**	0.03
d_YAVAPAI	0.2946	0.41	0.0004962*	0.00	0.0229	0.02
d_YUMA	0.5432	0.59	0.0004	0.00	0.0309	0.02
Superfund	-0.9370	0.83	-0.0001	0.00	0.0297	0.03
Miningsite	0.2179	0.49	-0.0006	0.00	-0.0029	0.02

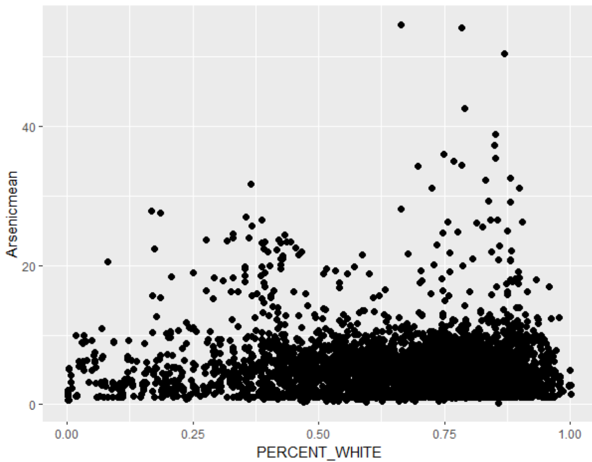
Table 6. Marginal effects binary dependent variable models

	Delta-method.Arsenic		Delta-method. Lead &Copper	
	dy/dx	se	dy/dx	se
d_groundwater	0.030062	0.040409	0.0152667	0.0380079
Sizecat	0.0059962	0.01009	-0.0565235***	0.0097431
PERCENT_FEMALELED	-0.0418745	0.082308	0.1091273	0.1270793
PERCENT_CHILD	0.2004275*	0.106651	0.19488	0.166057
PERCENT_HISPANIC	-0.0651959	0.050423	0.1174819*	0.055801
PERCENT_BLACK	-0.6639663***	0.236792	0.4235945	0.2837087
PERCENT_AMERICAN_INDIAN	-0.2201719	0.1575	0.1041205	0.1246573
PERCENTAGE_POVERTY	0.0019039***	0.000651	-0.001643	0.0009025
PERCENT_ELDERLY	-0.0892388*	0.040783	0.1167223**	0.0486929
PERCENT_HomeOwnership	0.0808359	0.03365	0.0092916	0.0433516
d_APACHE	0	(omitted)	0.2567041***	0.0536108
d_COCHISE	-0.1123629*	0.048901	0.0635791	0.0367601
d_COCONINO	-0.0574075	0.049685	0.0959757*	0.0442403
d_GILA	-0.1813183***	0.065212	-0.010786	0.0431977
d_GRAHAM	-0.0410561	0.099197	0.1116041	0.0757405
d_GREENLEE	0.0574621	0.097241	0.2501837***	0.0748385
d_LAPAZ	-0.147777	0.083256	0.1775139***	0.0518857
d_MOHAVE	-0.0208384	0.040591	0.1956585***	0.037337
d_NAVAJO	-0.160842***	0.060597	0.1972176***	0.0404322
d_PIMA	-0.1452901***	0.037735	0.0001614	0.0306819
d_PINAL	0.028347	0.03546	0.0843948**	0.0356475
d_SANTACRUZ	0.0227878	0.057527	0.0752379	0.0565725
d_YAVAPAI	0.025221	0.031987	-0.0090853	0.034893
d_YUMA	0.0843077	0.045206	0.2279227***	0.0436664
Superfund	0.0241679	0.066486	-0.022695	0.077015
Miningsite	-0.0556088	0.044056	-0.0351748	0.0433166

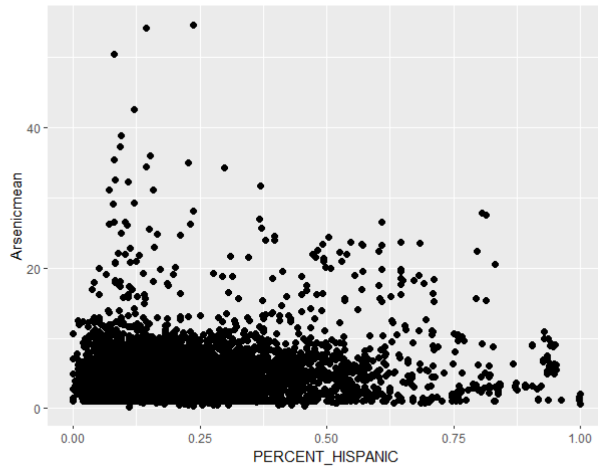
Note: *p<0.05; **p<0.01; ***p<0.001

Appendix 2. Graphics relations between race, income, age, and concentration of contaminants

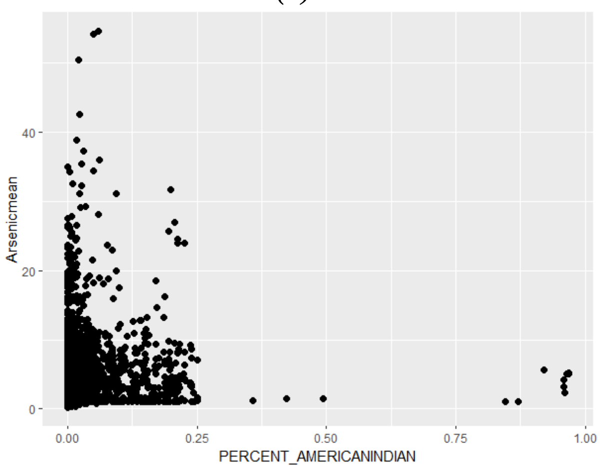
1. Arsenic
    a. Relation between ethnicity and arsenic concentration (ug/L)



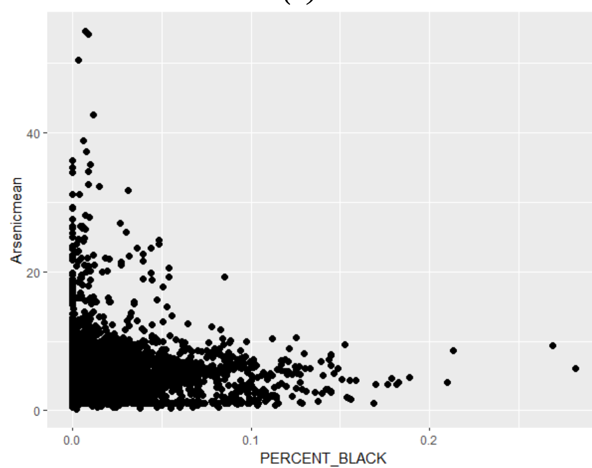
(a)



(b)

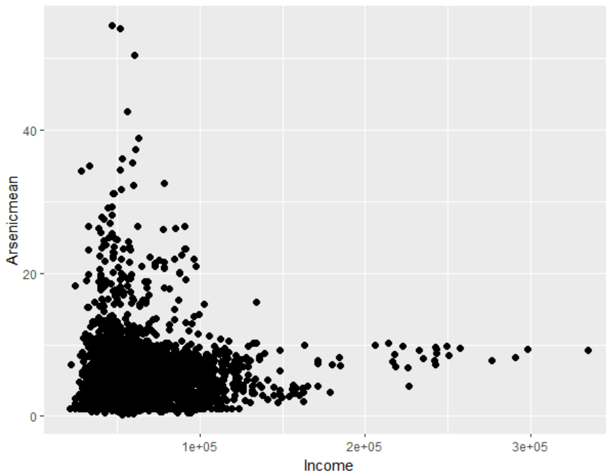


(c)

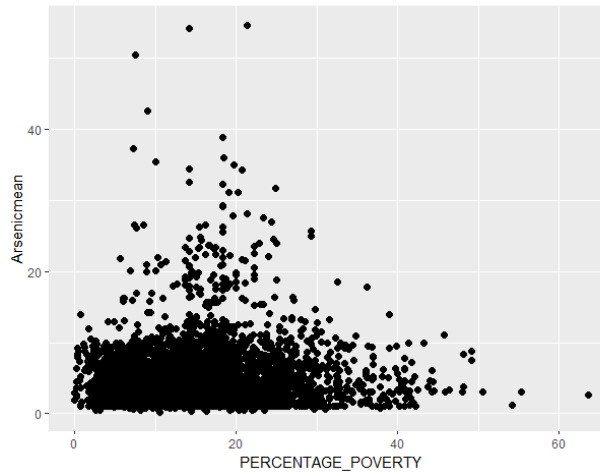


(d)

b. Relation between income and arsenic concentration (ug/L)

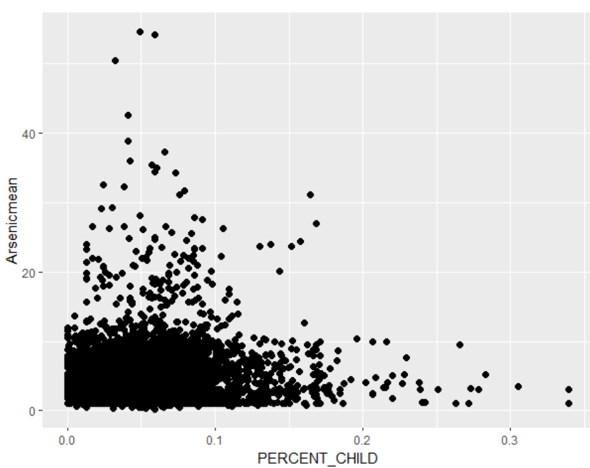


(a)

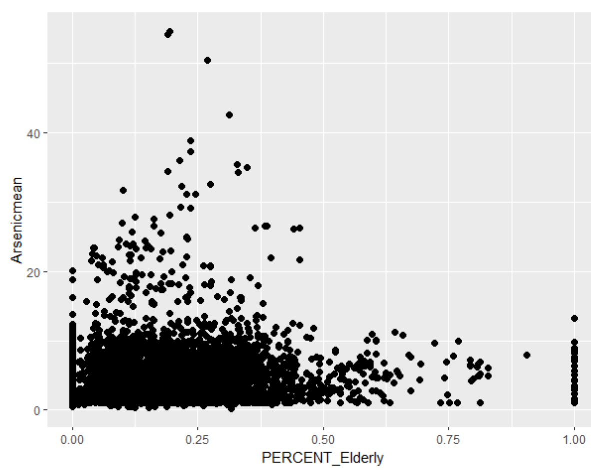


(b)

c. Relation between age and pollution

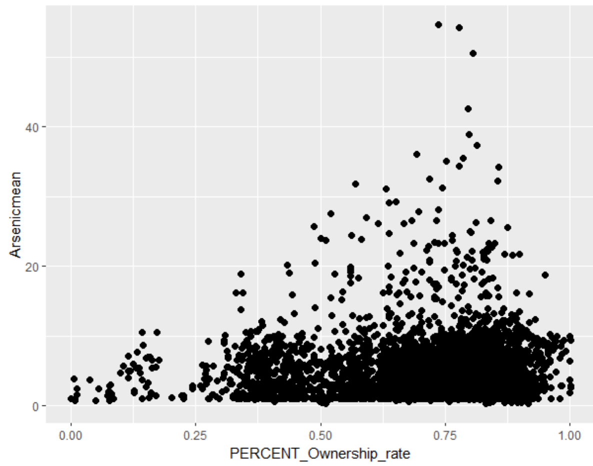


(a)

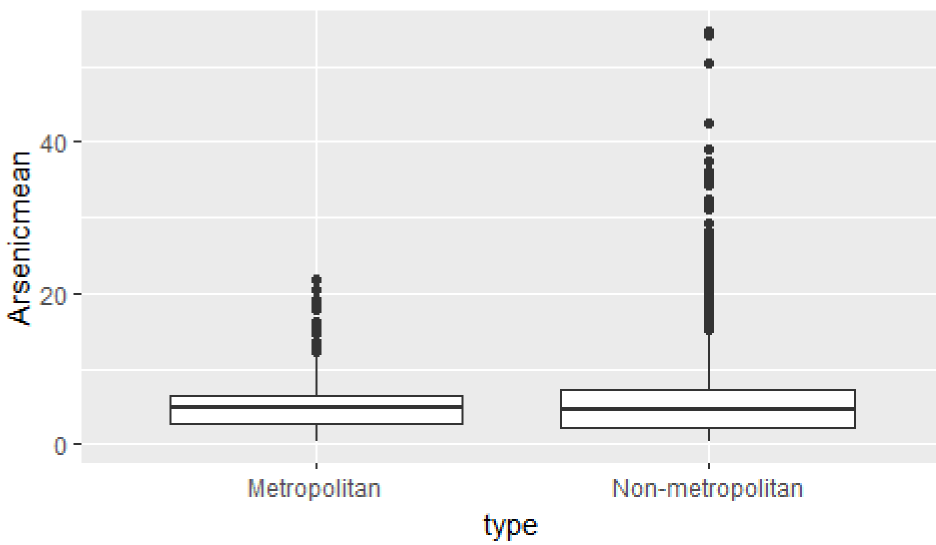


(b)

d. Relation between Home-ownership and pollution

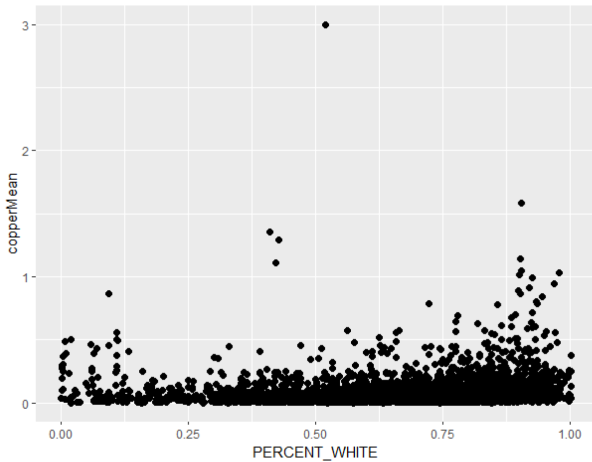


e. Concentration of Arsenic for metropolitan and non-metropolitan areas

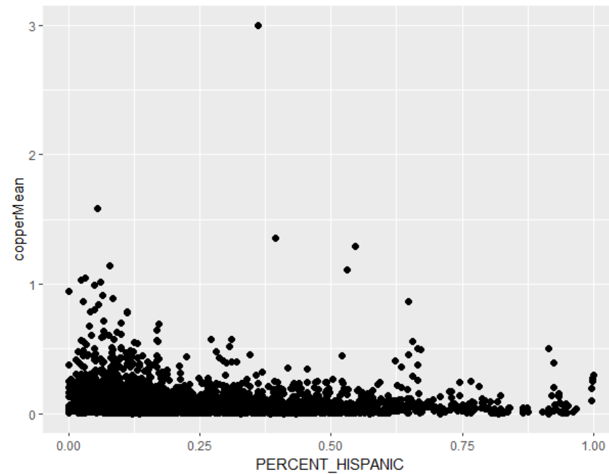


2. Copper

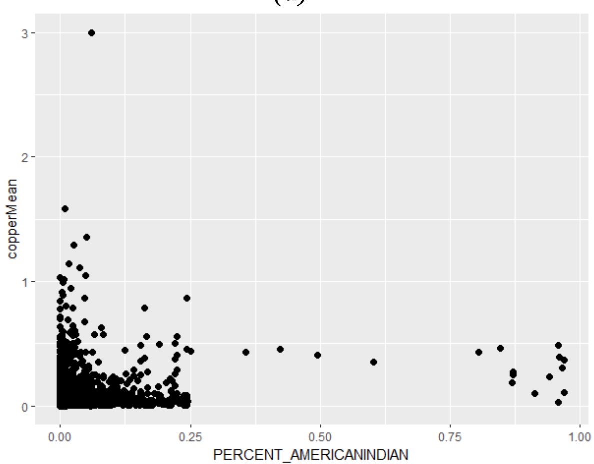
a. Relation between ethnicity and copper concentration (ug/L)



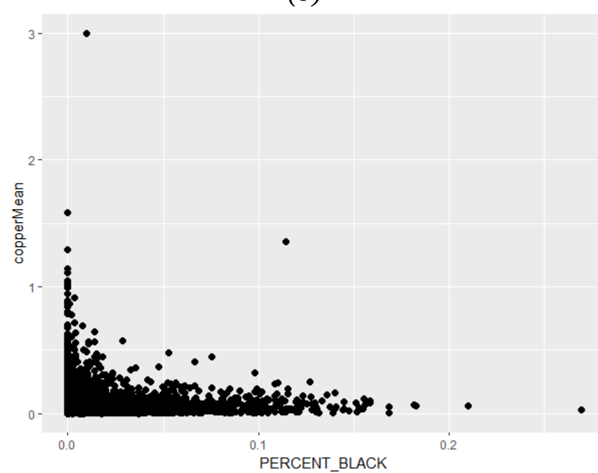
(a)



(b)

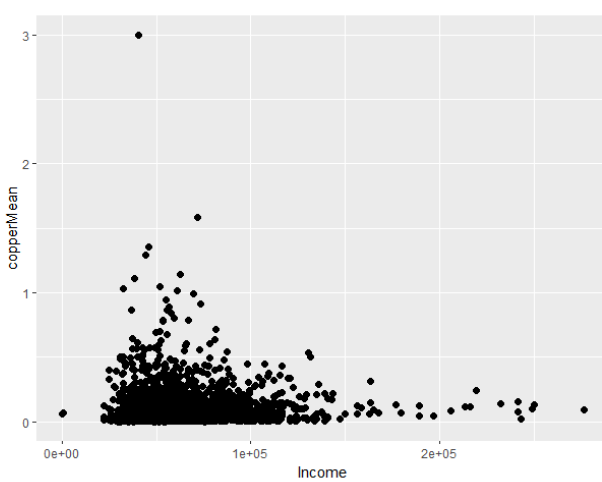


(c)

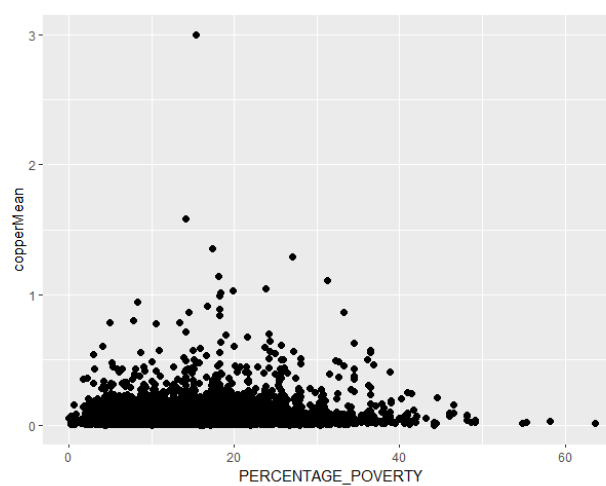


(d)

b. Relation between income and arsenic concentration (ug/L)

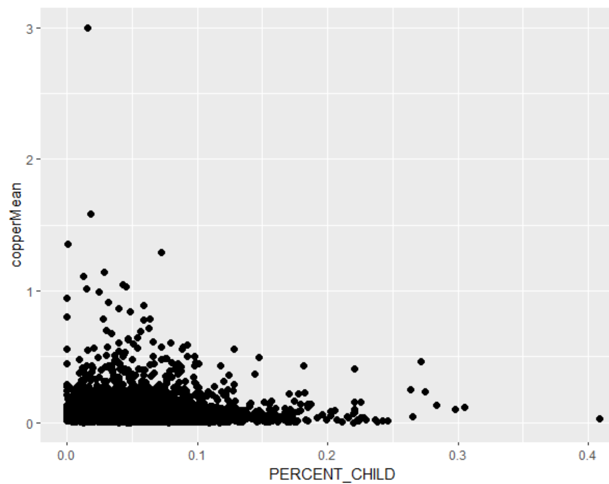


(a)

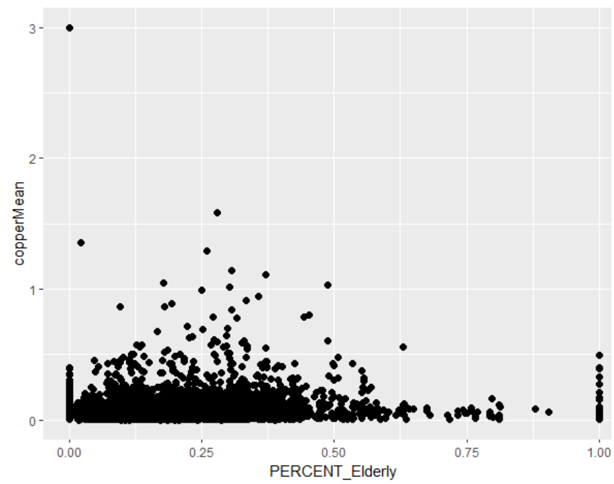


(b)

c. Relation between age and pollution

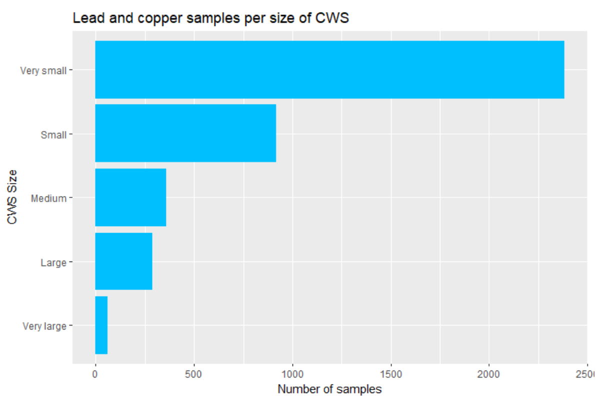


(a)

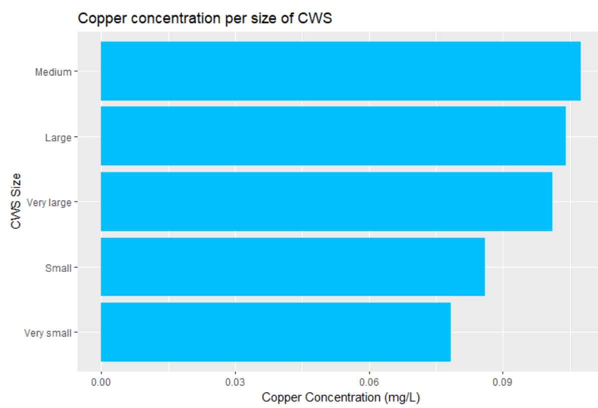


(b)

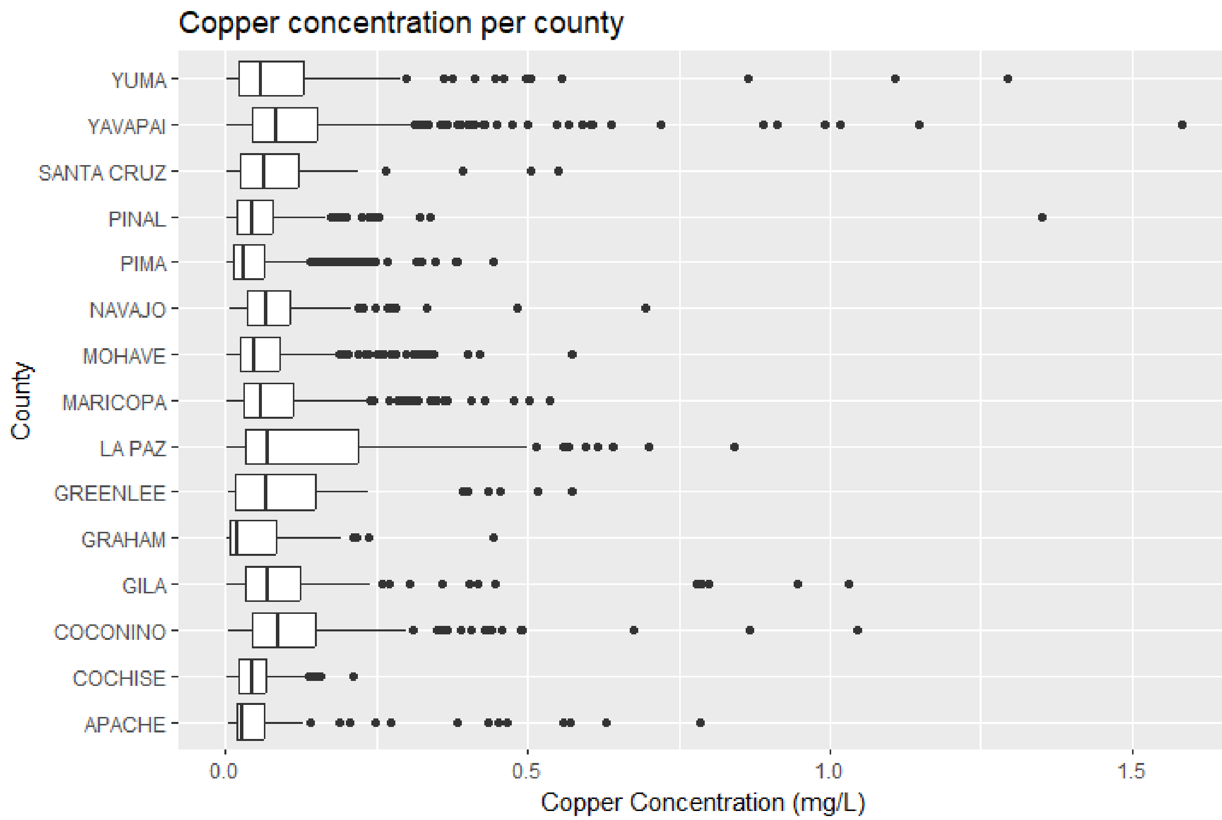
d. Relation between size and copper



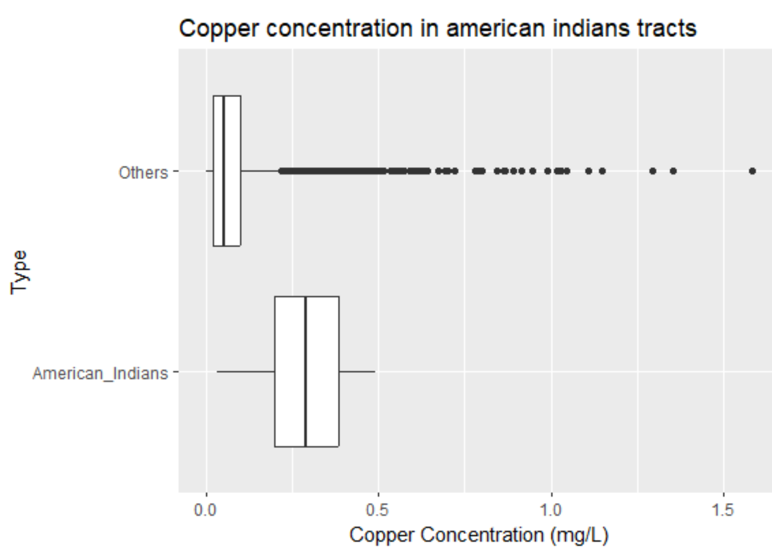
(a)



e. Concentrations of copper per county

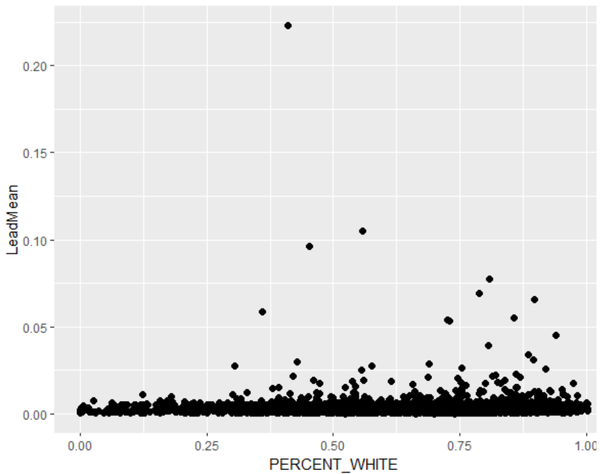


f. Concentrations of copper in tracts where the higher percent of the population is American Indians

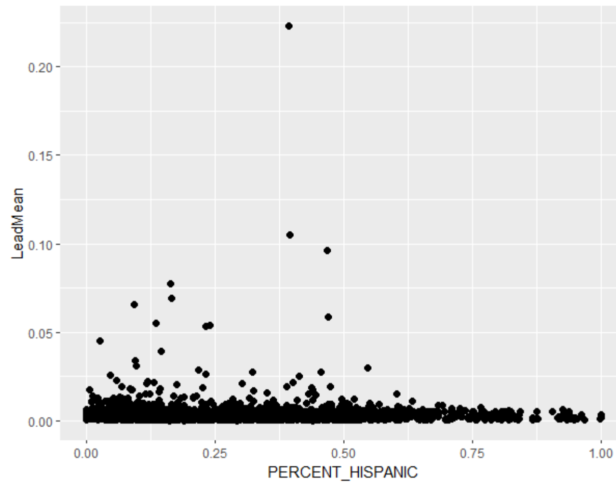


3. Lead

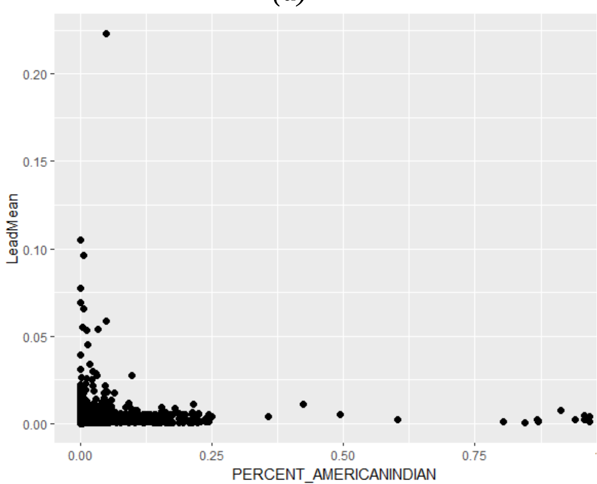
a. Relation between ethnicity and lead concentration (ug/L)



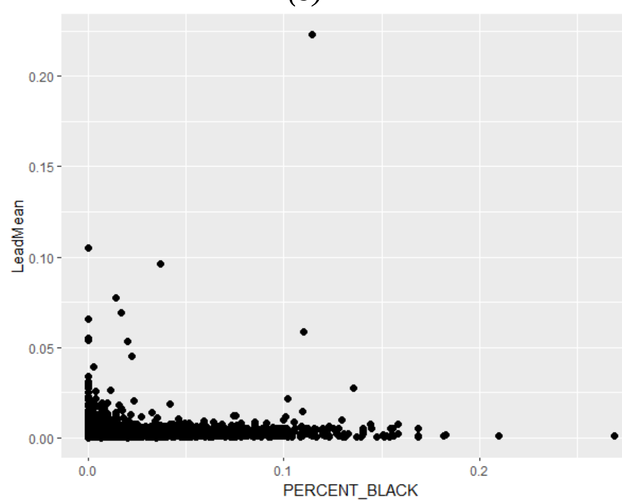
(a)



(b)

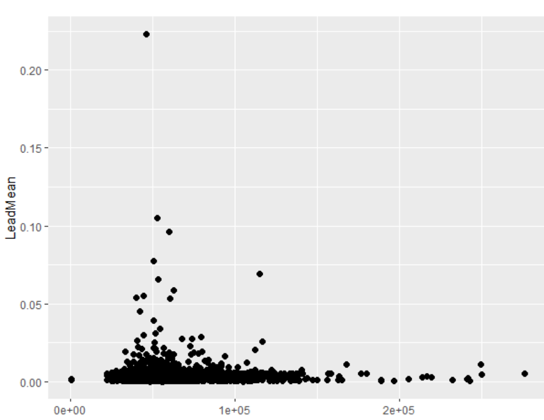


(c)

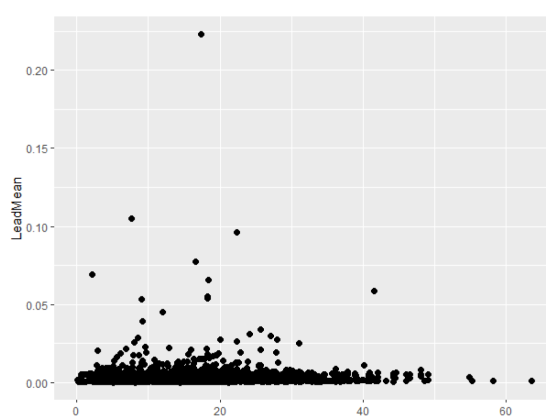


(d)

b. Relation between income and lead concentration (ug/L)

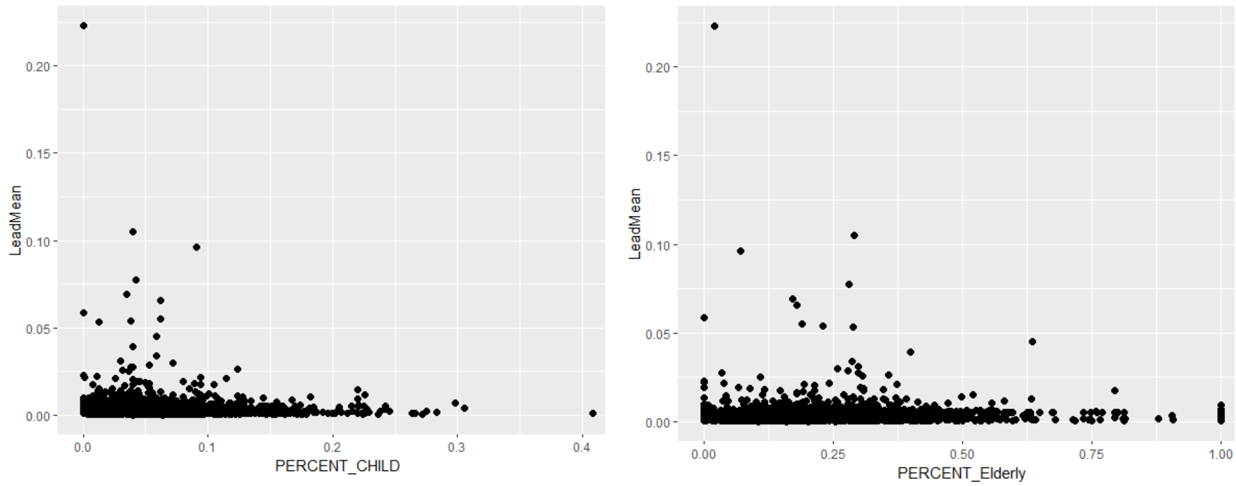


(a)



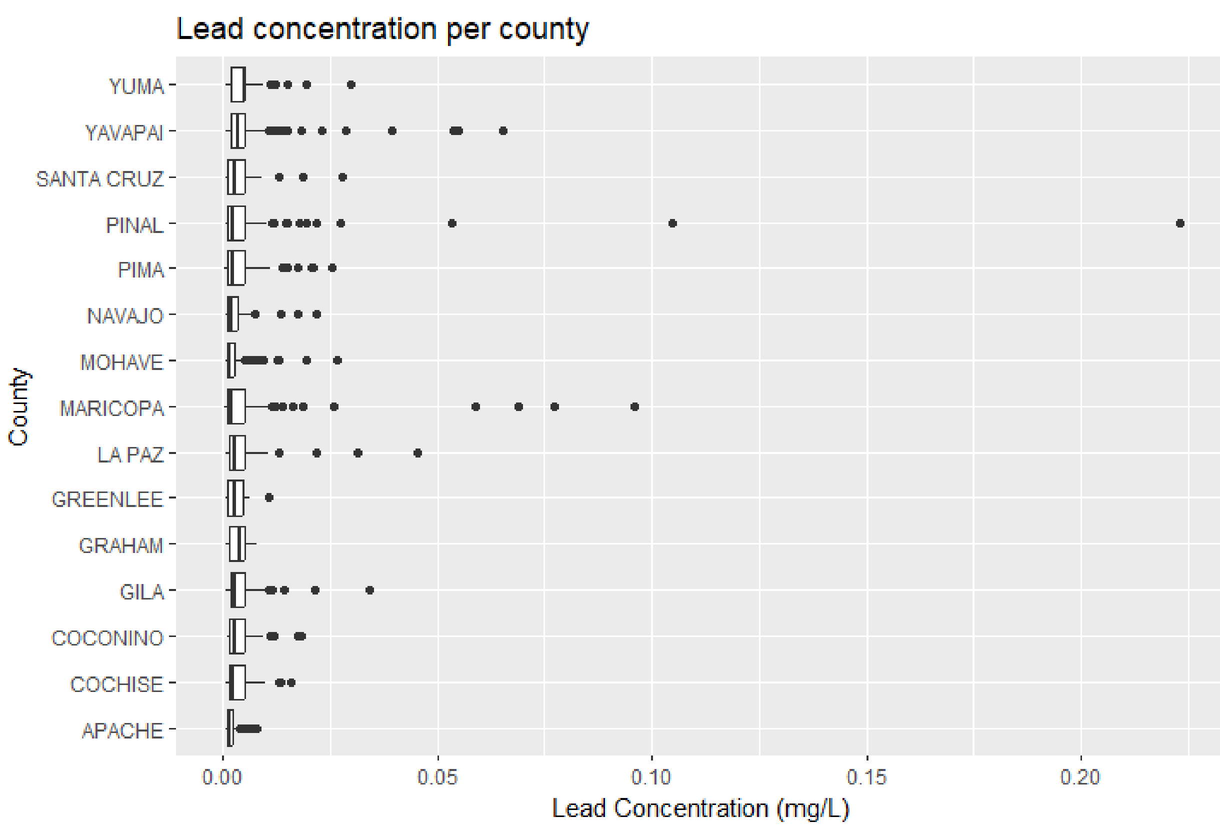
(b)

c. Relation between age and lead concentration

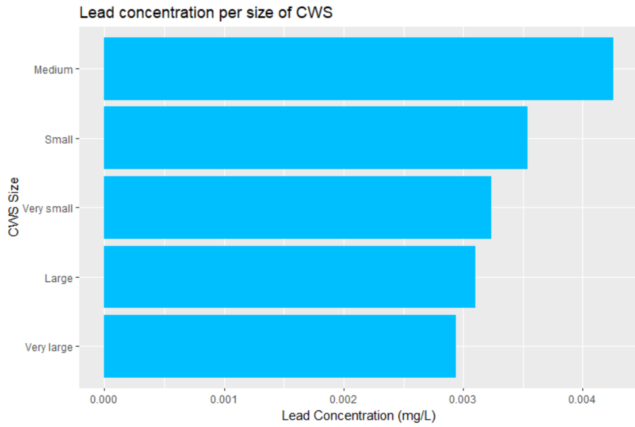


(a) *(b)*

d. Lead concentration per county

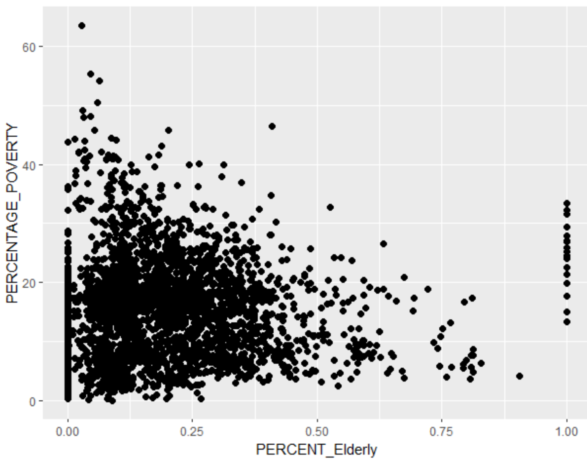


e. Relation between CWS size and lead.

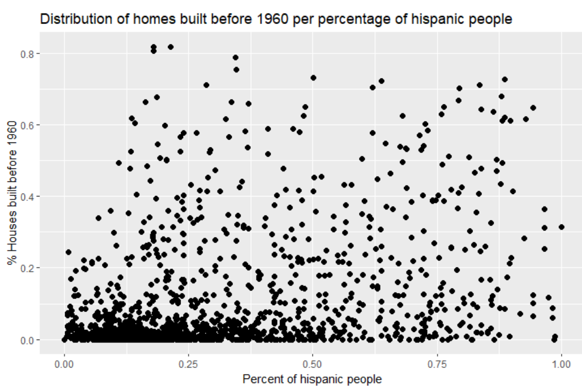


Appendix 3. Plots with the relation between explanatory variables

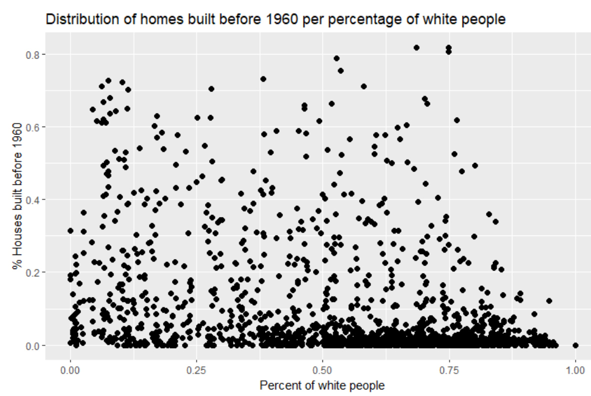
1. Percent of elderly vs rate of poverty



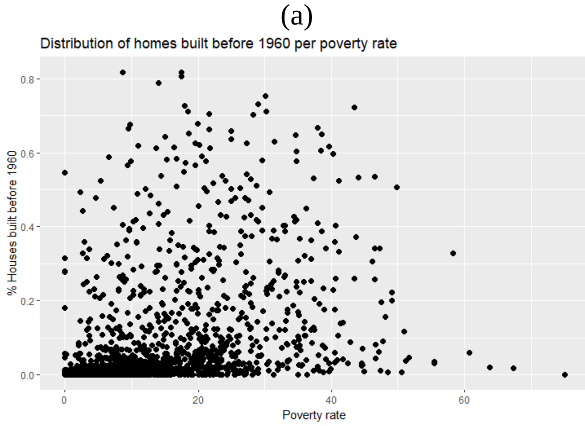
2. Distribution of homes built before 1960 and race.



(a)

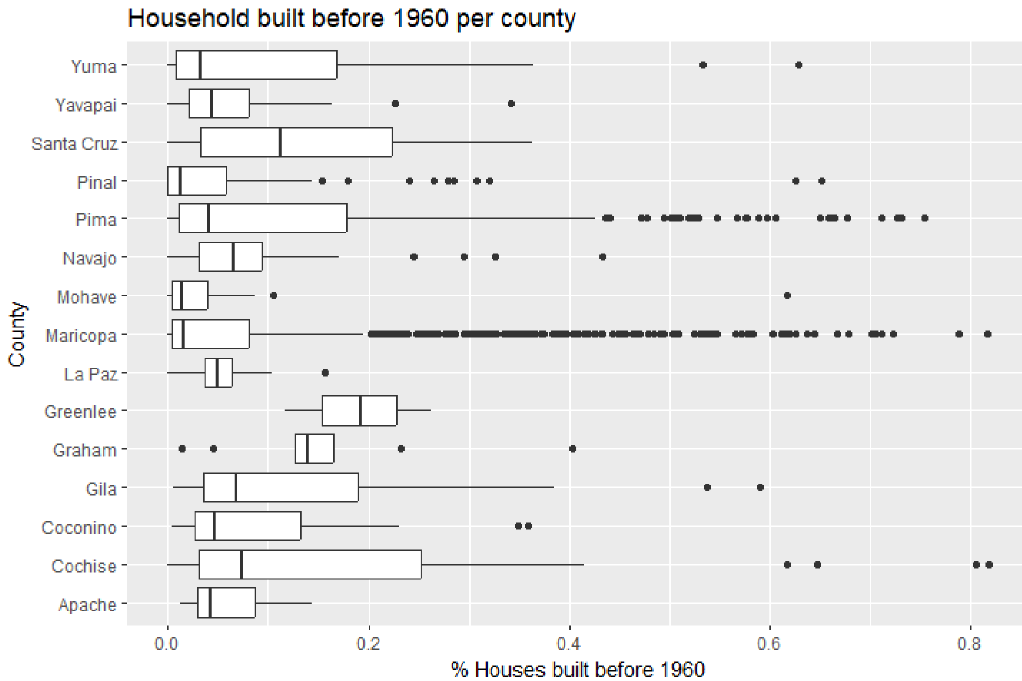


(b)

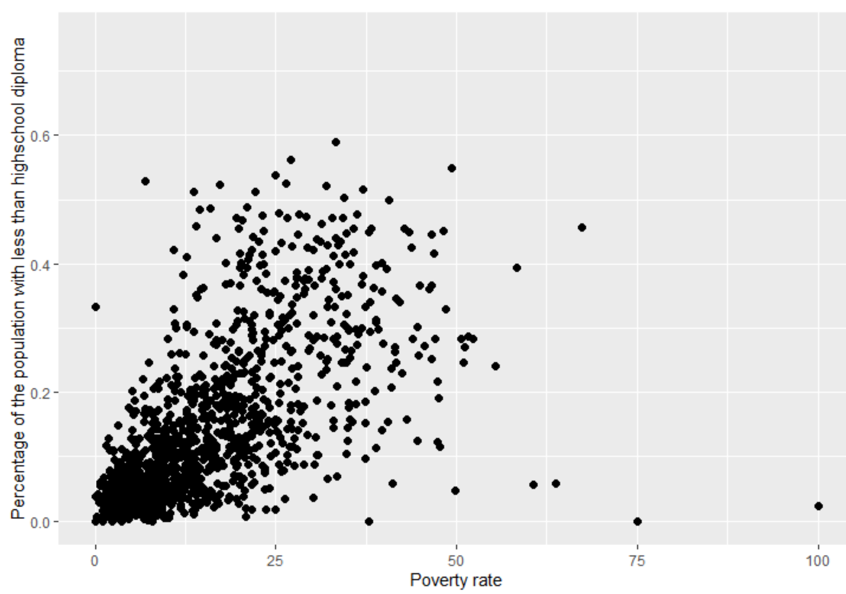


(c)

3. Home-ownership rate per county

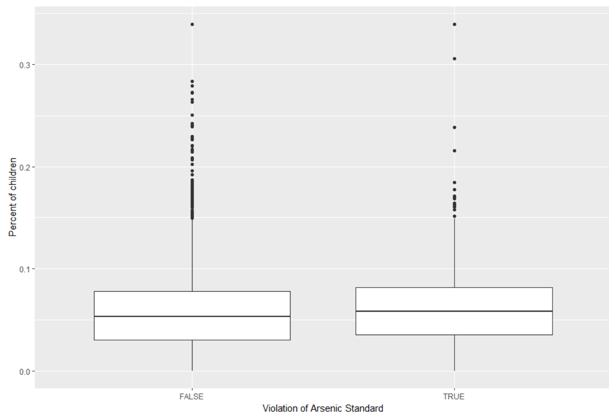


4. Relation between population with less than high school diploma and poverty rate

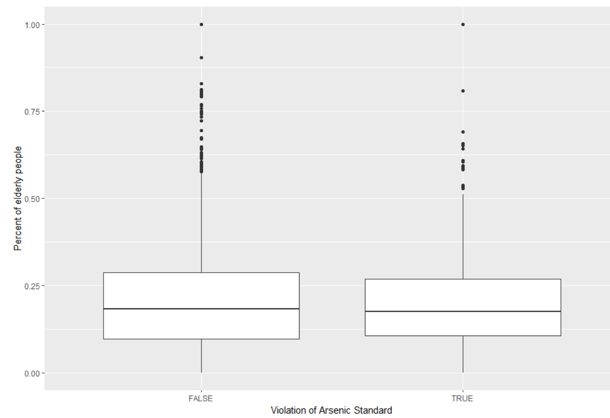


Appendix 4. Graphics relations between race, income, age and violations of standards

1. Arsenic
a. Relation between violations of arsenic standard and age

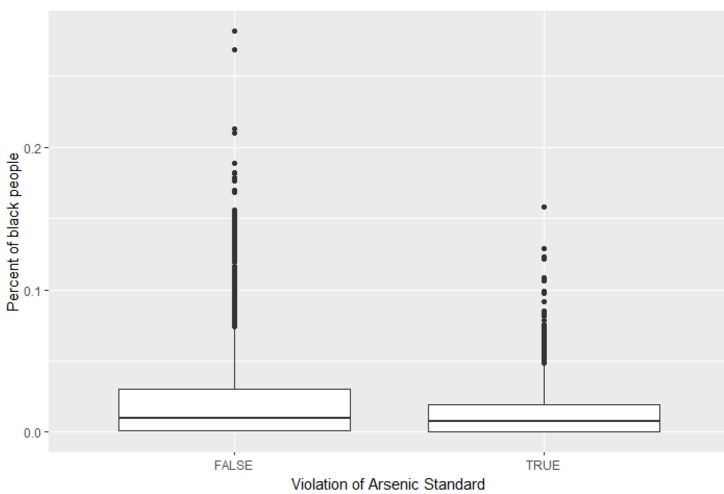


(a)

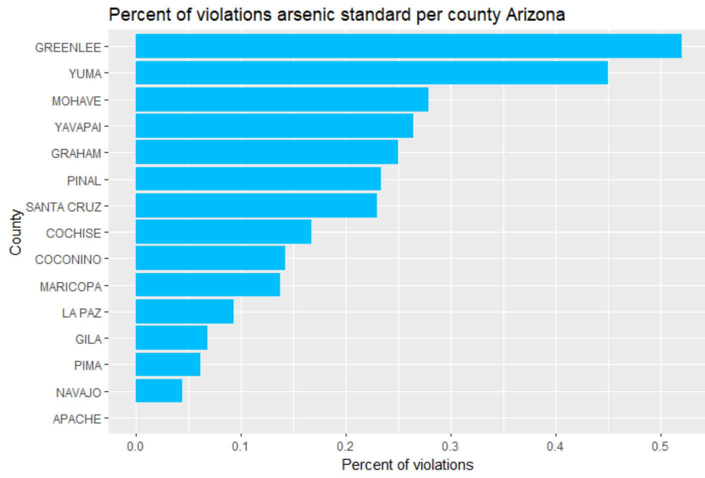


(b)

b. Relation between violations of arsenic standard and race

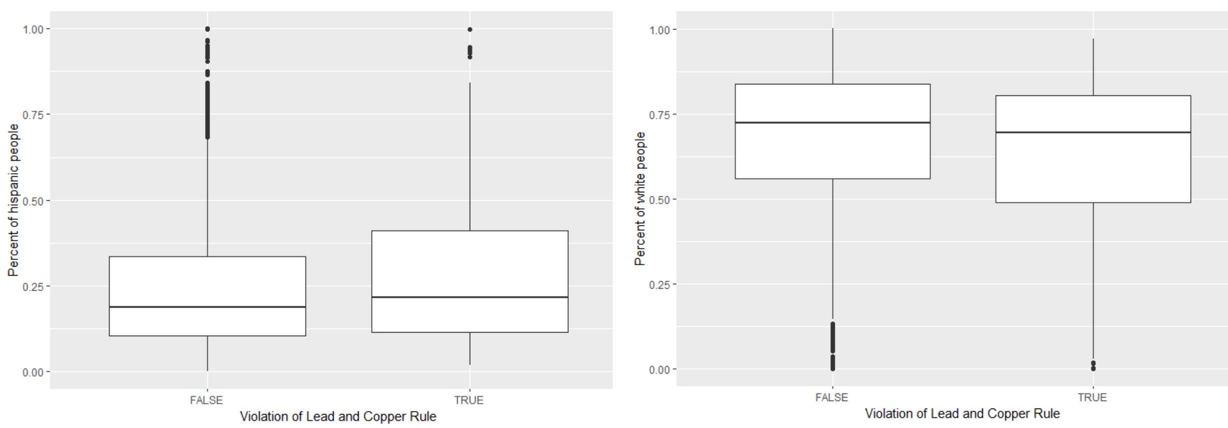


c. Distribution of percentage of violations per county.



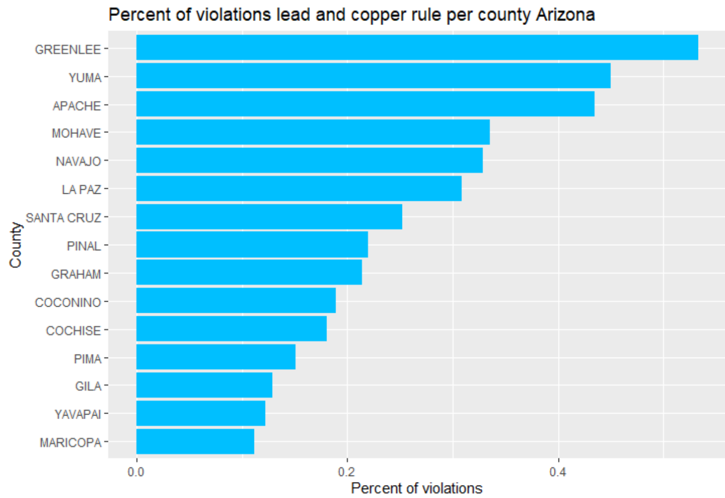
2. Lead and Copper

a. Relation between violations of lead and copper rule and race/ethnicity



(a) (b)

b. Distribution of violations for lead and copper rule per county



Appendix 5. Correlation Table

	d_groundwater	Sizecat	PERCENT_FEMALELED	PERCENT_CHILD	PERCENT_HISPANIC	PERCENT_BLACK	PERCENT_AMERICANINDIAN	Income	PERCENT_Elderly	PERCENT_Ownership_rate
d_groundwater	1.00	-0.45	0.05	0.08	0.11	-0.10	0.05	-0.22	0.05	0.10
Sizecat	-0.45	1.00	0.00	-0.12	-0.06	0.17	-0.07	0.23	0.01	-0.04
RCENT_FEMALELED	0.05	0.00	1.00	0.35	0.41	0.12	0.25	-0.25	-0.22	-0.21
PERCENT_CHILD	0.08	-0.12	0.35	1.00	0.33	0.14	0.14	-0.02	-0.36	-0.13
'ERCENT_HISPANIC	0.11	-0.06	0.41	0.33	1.00	0.18	0.03	-0.23	-0.33	-0.20
PERCENT_BLACK	-0.10	0.17	0.12	0.14	0.18	1.00	-0.05	0.17	-0.34	-0.12
AMERICANINDIAN	0.05	-0.07	0.25	0.14	0.03	-0.05	1.00	-0.16	-0.08	-0.13
Income	-0.22	0.23	-0.25	-0.02	-0.23	0.17	-0.16	1.00	-0.14	0.22
PERCENT_Elderly	0.05	0.01	-0.22	-0.36	-0.33	-0.34	-0.08	-0.14	1.00	0.35
NT_Ownership_rate	0.10	-0.04	-0.21	-0.13	-0.20	-0.12	-0.13	0.22	0.35	1.00
d_APACHE	0.03	-0.01	0.01	0.04	-0.05	-0.05	0.19	-0.02	0.00	0.05
d_COCHISE	0.02	-0.04	-0.03	-0.01	0.04	0.06	-0.07	-0.10	-0.01	-0.01
d_COCONINO	-0.05	0.00	0.00	-0.04	-0.13	-0.09	0.29	0.03	-0.05	-0.09
d_GILA	-0.02	-0.08	0.00	-0.04	-0.09	-0.12	-0.03	-0.12	0.13	0.04
d_GRAHAM	0.03	0.06	0.02	0.03	0.04	0.03	0.00	-0.03	-0.05	0.01
d_GREENLEE	-0.01	-0.05	0.04	0.05	0.12	-0.03	0.00	-0.03	-0.02	-0.03
d_LAPAZ	-0.08	-0.09	-0.09	-0.10	-0.04	-0.08	-0.04	-0.14	0.24	0.05
d_MOHAVE	0.04	0.00	-0.05	-0.06	-0.17	-0.07	-0.03	-0.16	0.16	-0.01
d_NAVAJO	0.06	0.01	0.01	0.03	-0.12	-0.09	0.08	-0.07	0.02	0.01
d_PIMA	0.13	-0.02	0.03	0.01	0.15	0.03	-0.08	0.05	-0.03	0.06
d_PINAL	0.07	0.05	0.10	0.04	0.15	0.15	0.18	-0.07	-0.09	0.00
d_SANTACRUZ	0.05	-0.01	0.02	0.01	0.25	-0.10	-0.04	-0.03	0.07	-0.01
d_YAVAPAI	0.16	-0.13	-0.03	-0.13	-0.30	-0.24	-0.07	-0.14	0.15	-0.02
d_YUMA	-0.10	-0.03	0.06	0.07	0.32	-0.02	-0.03	-0.06	-0.11	-0.13
Superfund	-0.04	0.13	0.02	-0.04	0.06	0.02	0.03	-0.04	0.00	0.01
Miningsite	-0.23	0.47	0.06	-0.06	0.01	0.05	-0.04	0.02	-0.01	-0.05
:ENTAGE_POVERTY	0.11	-0.15	0.29	0.22	0.31	-0.14	0.27	-0.51	-0.06	-0.17

	d_APACHE	d_COCHISE	d_COCONINO	d_GILA	d_GRAHAM	d_GREENLEE	d_LAPAZ	d_MOHAVE	d_NAVAJO	d_PIMA	d_PINAL	d_SANTACRUZ	d_YAVAPAI	d_YUMA	Superfund	Miningsite	PERCENTAGE_POVERTY
d_groundwater	0.03	0.02	-0.05	-0.02	0.03	-0.01	-0.08	0.04	0.06	0.13	0.07	0.05	0.16	-0.10	-0.04	-0.23	0.11
Sizecat	-0.01	-0.04	0.00	-0.08	0.06	-0.05	-0.09	0.00	0.01	-0.02	0.05	-0.01	-0.13	-0.03	0.13	0.47	-0.15
RCENT_FEMALELED	0.01	-0.03	0.00	0.00	0.02	0.04	-0.09	-0.05	0.01	0.03	0.10	0.02	-0.03	0.06	0.02	0.06	0.29
PERCENT_CHILD	0.04	-0.01	-0.04	-0.04	0.03	0.05	-0.10	-0.06	0.03	0.01	0.04	0.01	-0.13	0.07	-0.04	-0.06	0.22
'ERCENT_HISPANIC	-0.05	0.04	-0.13	-0.09	0.04	0.12	-0.04	-0.17	-0.12	0.15	0.15	0.25	-0.30	0.32	0.06	0.01	0.31
PERCENT_BLACK	-0.05	0.06	-0.09	-0.12	0.03	-0.03	-0.08	-0.07	-0.09	0.03	0.15	-0.10	-0.24	-0.02	0.02	0.05	-0.14
AMERICANINDIAN	0.19	-0.07	0.29	-0.03	0.00	0.00	-0.04	-0.03	0.08	-0.08	0.18	-0.04	-0.07	-0.03	0.03	-0.04	0.27
Income	-0.02	-0.10	0.03	-0.12	-0.03	-0.03	-0.14	-0.16	-0.07	0.05	-0.07	-0.03	-0.14	-0.06	-0.04	0.02	-0.51
PERCENT_Elderly	0.00	-0.01	-0.05	0.13	-0.05	-0.02	0.24	0.16	0.02	-0.03	-0.09	0.07	0.15	-0.11	0.00	-0.01	-0.06
NT_Ownership_rate	0.05	-0.01	-0.09	0.04	0.01	-0.03	0.05	-0.01	0.01	0.06	0.00	-0.01	-0.02	-0.13	0.01	-0.05	-0.17
d_APACHE	1.00	-0.02	-0.02	-0.02	-0.01	-0.01	-0.02	-0.03	-0.02	-0.04	-0.03	-0.02	-0.05	-0.02	-0.02	-0.01	0.07
d_COCHISE	-0.02	1.00	-0.05	-0.05	-0.02	-0.02	-0.04	-0.06	-0.04	-0.09	-0.08	-0.04	-0.11	-0.05	-0.04	-0.01	0.02
d_COCONINO	-0.02	-0.05	1.00	-0.04	-0.02	-0.02	-0.03	-0.05	-0.04	-0.08	-0.07	-0.03	-0.10	-0.05	-0.02	-0.01	0.03
d_GILA	-0.02	-0.05	-0.04	1.00	-0.02	-0.02	-0.03	-0.05	-0.04	-0.08	-0.07	-0.03	-0.10	-0.05	0.02	0.07	0.05
d_GRAHAM	-0.01	-0.02	-0.02	-0.02	1.00	-0.01	-0.01	-0.02	-0.02	-0.04	-0.03	-0.01	-0.04	-0.02	-0.01	0.11	0.03
d_GREENLEE	-0.01	-0.02	-0.02	-0.02	-0.01	1.00	-0.01	-0.02	-0.01	-0.03	-0.03	-0.01	-0.04	-0.02	-0.01	-0.02	-0.02
d_LAPAZ	-0.02	-0.04	-0.03	-0.03	-0.01	-0.01	1.00	-0.04	-0.03	-0.06	-0.05	-0.02	-0.07	-0.04	-0.02	-0.05	0.07
d_MOHAVE	-0.03	-0.06	-0.05	-0.05	-0.02	-0.02	-0.04	1.00	-0.05	-0.11	-0.09	-0.04	-0.13	-0.06	0.05	0.03	0.15
d_NAVAJO	-0.02	-0.04	-0.04	-0.04	-0.02	-0.01	-0.03	-0.05	1.00	-0.07	-0.06	-0.03	-0.09	-0.04	-0.03	-0.02	0.08
d_PIMA	-0.04	-0.09	-0.08	-0.08	-0.04	-0.03	-0.06	-0.11	-0.07	1.00	-0.13	-0.06	-0.19	-0.09	-0.06	-0.09	-0.04
d_PINAL	-0.03	-0.08	-0.07	-0.07	-0.03	-0.03	-0.05	-0.09	-0.06	-0.13	1.00	-0.05	-0.16	-0.08	0.12	0.08	0.02
d_SANTACRUZ	-0.02	-0.04	-0.03	-0.03	-0.01	-0.01	-0.02	-0.04	-0.03	-0.06	-0.05	1.00	-0.08	-0.04	0.03	-0.02	0.01
d_YAVAPAI	-0.05	-0.11	-0.10	-0.10	-0.04	-0.04	-0.07	-0.13	-0.09	-0.19	-0.16	-0.08	1.00	-0.11	-0.01	-0.06	-0.01
d_YUMA	-0.02	-0.05	-0.05	-0.05	-0.02	-0.02	-0.04	-0.06	-0.04	-0.09	-0.08	-0.04	-0.11	1.00	-0.03	-0.01	0.10
Superfund	-0.02	-0.04	-0.02	0.02	-0.01	-0.01	-0.02	0.05	-0.03	-0.06	0.12	0.03	-0.01	-0.03	1.00	0.19	0.04
Miningsite	-0.01	-0.01	-0.01	0.07	0.11	-0.02	-0.05	0.03	-0.02	-0.09	0.08	-0.02	-0.06	-0.01	0.19	1.00	-0.04
:ENTAGE_POVERTY	0.07	0.02	0.03	0.05	0.03	-0.02	0.07	0.15	0.08	-0.04	0.02	0.01	-0.01	0.10	0.04	-0.04	1.00