The University of Chicago

Opt In, Left Out:

Environmental Justice Impacts of Voluntary Lead Testing in Chicago

By Hailey Sobel

-

A thesis submitted for partial fulfillment of the requirements for a Bachelor of Arts degree in

Public Policy Studies

Paper presented to:

Public Policy Studies Preceptor, Alexis Pearson

Faculty Advisor, Karlyn Gorski

Department of Public Policy Studies

April 14, 2025

Acknowledgments

The guidance and encouragement I have received throughout this project has been exceptional. I first want to recognize my incredible preceptor, Alexis Pearson, who went above and beyond for each of her students. Alexis' care, thoughtful edits, guiding questions, analytical mind, and support have improved this paper far beyond my initial expectations. I also want to express my gratitude to my advisor, Professor Karlyn Gorski, who pushed me to pursue surprising findings and interrogate the meaning of my results. Thank you to the UChicago faculty within and outside of the Public Policy department who have truly shaped the way I have thought about government, data, research, and interpretation over the last four years. Lastly, I'd like to thank my friends who have read countless pages, offered insightful critique, and fueled late night writing sessions with laughter and energy.

Abstract

With approximately 400,000 lead service lines, Chicago has the most municipal lead water infrastructure in the United States. The city, which once mandated installation of lead pipes, is now confronting the problem with voluntary water testing and service line replacement programs. Requiring citizens to take an active role in requesting and performing government services is controversial: it engages citizens in political participation and provides information to city officials, but also exacerbates inequalities because residents with fewer resources tend to be excluded. In this paper, I use Chicago's 311 testing request database alongside census data to investigate whether barriers limit program utilization. I find that both objective risk and socioeconomic factors determine free lead testing engagement, but citizens adapt to real and perceived barriers with significant success. I recommend that Chicago implements a free, integrated, and mandatory testing and replacement system to eliminate all lead service lines by the federal 2047 deadline.

Table of Contents

Acknowledgments	**1**
Abstract	**2**
Table of Contents	**3**
Introduction	**4**
Background	**7**
History of lead use	7
Health damage from lead	8
Disproportionate harm	10
Legislative history	11
Chicago's water distribution system	13
Progress	15
Literature review	**16**
Theoretical Frameworks	16
Responsiveness to need and demand	20
Propensity to complain	22
Insight from Chicago's lead request system	27
Methods	**28**
Results	**33**
Key features of request database	33
Spatial distribution	38
Regression analysis	42
Time progression	60
Analysis	**62**
Limitations	**68**
Policy Implications	**69**
Conclusion	**75**
References	**78**
Appendix A: Descriptive Statistics	**90**
Appendix B: Gender of head of household	**91**
Appendix C: Code	**92**

Introduction

Overnight in 1986, the Safe Drinking Water Act Amendments transformed lead drinking water lines in Chicago from mandated to prohibited (Eng, 2020). Despite the discovery and publication of the dangers of lead water pipes since the mid-19th century, strong lobbying from the lead and plumbing industries succeeded in downplaying the risks and advocating for continued use (Rabin, 2008). In the November 1923 issue of National Geographic, the National Lead Company published a full page ad entitled “Lead helps to guard your health” (National Geographic Society, 1923). Throughout the next several decades, a coalition of lead companies pursued a highly organized action plan to counter the spreading awareness of the poisoning caused by their products (McCormick and Uteuova, 2022).

The Lead Industries Association lobbied municipalities like Chicago to not just allow, but require, the use of lead water pipes in building codes (McCormick and Uteuova, 2022). In tactics reminiscent of the tobacco industry, lead companies funded medical research to cast doubt on the drastic health effects of lead exposure (McCormick and Uteuova, 2022). The Association also activated the political capital of skilled plumbers unions that benefited from these requirements because they prevented a shift to less technically complicated processes that unskilled and unlicensed workers could perform (Rabin, 2008; Eng, 2018). This dynamic was especially pervasive in Chicago, where the Plumbers Union was particularly strong. Its president, Stephen M. Bailey, was childhood friends with Mayor Richard J. Daley, and he served as an important political supporter throughout Daley’s two decades in office from 1955 to 1976 (Eng, 2018). Bailey helped marshal labor votes in municipal elections, and Chicago steadfastly maintained its edict for lead service lines, even as the vast majority of other major cities banned their use in the face of mounting medical evidence.

For decades, every new building or renovation was connected to Chicago's water distribution system with lead. However, in the 1960s and 1970s, as polluting industries became more visible and criticized, the federal government increasingly took responsibility for regulating environmental and safety standards (Lewis, 1988). Yet, even after the Safe Drinking Water Act Amendments in 1986 prevented any future additions of lead service lines, there was little serious political will to remove and replace the existing infrastructure in Chicago. When Mayor Rahm Emanuel replaced the city's water mains in 2012, the lead pipes were left in place (Eng, 2018). This legacy of underappreciated risk and required use has left Chicago with over 380,000 lead pipes, the most of any city in the United States (EDF, 2024).

Chicago is now trying to comprehensively replace all the lead infrastructure lying under the city streets. Federal rulemaking from the Biden administration has created an accelerated timeline and made resources available for that pursuit (EPA, 2025). Complementary lead testing provided through the Chicago 311 municipal service system[1] aims to help prioritize which service lines are removed first and protect citizens with temporary measures in the meantime (Lead-Safe Chicago, n.d.). According to the mayor's office, this program aims to "[lead] with equity so that residents with the greatest need will be provided the highest levels of supports first, and ultimately any resident who wants to replace their lead service line will be able to do so" (Eng, 2020). However, determining where the greatest need is through a voluntary and citizen-initiated testing system is problematic. Structural barriers like lack of awareness, trust in government, available time, literacy, and capacity can limit overall engagement in bureaucratic systems, even when they provide critical programs (e.g., Herd and Moynihan, 2019). Moreover, barriers tend to affect different groups of people to varying extents, exacerbating inequalities

[1] 311 is a non-emergency version of the 911 operating system that connects residents with municipal departments and agencies to access government services and resources.

along wealth and racial lines (e.g., Christensen et al., 2020; Herd and Moynihan, 2019; Masood and Nisar, 2020). Despite the theoretical grounding, existing empirical literature is conflicted on the strength of these socioeconomic effects, and the Chicago lead testing system remains understudied.

Throughout the following research, I investigated the impacts of these institutional factors in Chicago's voluntary lead testing program by leveraging the public database of 311 requests on the city data portal (City of Chicago, n.d.). I aimed to elucidate whether significant barriers impede use of the testing program, and explored the existence of social inequities in program implementation. Specifically, I examined whether patterns of who contacts and successfully engages with the system can be better predicted by objective lead risk or demographic and socioeconomic factors that indicate resource access. In this pursuit, I mapped demographic and request patterns across the city, built correlational models between explanatory variables and request details, and tracked changes over time.

I find dramatically low utilization and high rates of attrition, as well as substantial relationships between program utilization and both socioeconomic and risk factors. These results indicate the existence of prohibitive barriers that unequally affect groups with lower socioeconomic status. Surprisingly, I find that high minority populations are positively correlated to uptake after controlling for both need and income, but are negatively related to successful completion of a request. I suggest that this contrast reveals the existence of barriers at multiple steps in the testing process that have varying impacts. Additionally, the evidence suggests that residents adapt to real and perceived barriers through their request type and contact method, emphasizing that citizens are active participants even in burdensome systems. Based on these insights, I recommend that Chicago replaces the current patchwork of service line identification, voluntary water testing, and means–tested replacement programs with an integrated, universal, and mandatory effort following the example of Newark, New Jersey. I also suggest that municipalities across the country reconsider providing services and distributing resources based on voluntary or citizen-initiated contacts in order to prevent exacerbating existing inequalities in the urban landscape.

Background

History of lead use

Humans have used lead for over 6 millenia in household products like jewelry, cosmetics, paints, pottery, and glassware (Needleman, 1999). The metal has key physical properties that encourage its use in a variety of industries: longevity, malleability, low melting point, and a remarkable ability for holding pigment. During the Roman empire, lead transformed into a mass produced and extremely prevalent commodity, used for everything from sweeteners to urns, and featured in the historic water distribution system through aqueducts and plumbing (Needleman, 1999). In modern times, major sources of lead exposure to humans have included leaded gasoline, lead paint, emissions from industrial processes, and public water systems (Levin, 2008). In the United States, significant installation of lead water lines began at the end of the 19th century due to durability and flexibility advantages over the iron pipes they replaced (Rabin, 2008). By 1900, nearly three-quarters of cities with over 30,000 residents used pipes that contained lead (Rabin, 2008). Remnants from many of these uses continue to contaminate the soils, water systems, and houses of Americans (Rabin, 2008). Lead from paint and dust now cause an estimated 70% of elevated blood lead blood levels in the United States, while drinking water is responsible for 10-20% of total lead exposure (Levin et al., 2008).

Health damage from lead

Societies have discovered and rediscovered the deleterious health effects of lead exposure throughout history. At least three times in human development—Antiquity, the Middle Ages, and the Modern Era—people have recognized and understood lead poisoning (Hernberg, 2000). Yet, each time, these insights were largely forgotten and lead continued to be widely used. Even as historians theorized in the mid-1900s that lead exposure contributed to severe health issues in the past and potentially even the decline of the Roman Empire, the United States continued to allow the use of lead pipes, paint, and gasoline (Gilfillan, 1965; Moore, 2019).

Rabin's foundational work on the lead industry, "A Modest Campaign," documents the history of political activism by lead companies to influence regulation (2008). As lead pipes came into wide use across America in the 1800s, Rabin describes how officials, industry leaders, scientists, doctors, and the public debated the risk of drinking water delivered through lead lines. Though it was generally known that lead could have toxic effects in acute dosages, chronic exposure effects and the possibility of leaching into the water supply were less clear. As early as the 1850s, health, medical, and engineering professionals, as well as the general public, raised concerns about lead service lines. Reports continued to accumulate, and by 1890 the Massachusetts State Board of Health issued an advisory against using lead pipes. The public health community reached near-consensus about the dangers by the first decades of the 20th century, and increased public awareness placed pressure on local governments to reduce toxic exposure. As cities began to outlaw lead plumbing, however, active pushes from production, manufacturing industries, and unions sought to diminish medical outcomes and enshrine lead's role as a crucial part of infrastructure. The Lead Industries Association even funded the defense of landlords accused in court of poisoning tenants through contamination from the building's pipes.

In the 1960s, in response to increasing scientific reports, the Centers for Disease Control and Prevention (CDC) first established blood lead level thresholds for children (Rabin, 2008). Since then, health institutions have become increasingly stringent, substantially lowering the threshold levels at which lead is considered dangerous. In fact, the World Health Organization declares that there is no safe level of lead exposure (WHO, 2024). Scientific and medical studies now illuminate the pathways and ultimate consequences of contamination. These studies show that lead is primarily a neurotoxin that impairs cognition and is especially dangerous during childhood development (Froehlich et al. 2009; Lanphear et al. 2005; Lidsky and Schneider, 2003; Schnaas et al. 2006). Even extremely low concentrations cause persistent behavioral and intellectual deficiencies in individuals, including increased rates of crime (Aizer et al. 2018; Canfield et al., 2003; Reyes, 2015; Wright et al. 2008). These neurological effects deteriorate academic performance and future expected earnings for elementary school kids (Hollingsworth et al., 2020). In Chicago specifically, low levels of lead exposure are associated with significantly decreased academic performance on statewide standardized tests by 3rd grade, with lead toxicity responsible for nearly 15% of failure on these exams (Evens et al., 2015).

At the population level, the magnitude of lead emissions is directly linked to the rate of aggravated assault across American cities over time, and experts have modeled the substantial decline of IQ in generations exposed to lead in childhood (Mielke and Zahran, 2012; McFarland et al., 2021). Reduction in lead exposure increases population reading and math scores and narrows the racial academic achievement gap (Sorensen et al., 2019). Beyond cognitive effects, chronic lead poisoning in both adults and children results in higher instances of cardiovascular, cancer, and overall mortality (Menke et al., 2006; Schober et al., 2006). In fact, 18% of all-cause mortality in the United States is attributable to lead exposure, corresponding to hundreds of thousands of deaths a year (Lanphear et al., 2018). Extensive evidence proves that even small doses of lead pose a dramatic health risk with catastrophic individual and societal effects.[2]

Disproportionate harm

The harm caused by lead exposure is not equally distributed across Americans. Like many other health and environmental crises (e.g., air pollution in Liu et al., 2021, highway traffic in Rowangould, 2013, and toxic waste in Reichel, 2018), racial minorities and low income families disproportionately bear the brunt of lead's devastating consequences. Childhood blood tests show that lead toxicity has declined across sociodemographic groups due to regulations banning leaded gasoline and lead paint (Egan et al., 2021; Jones et al., 2009; Levin, 2008). Still, elevated blood lead levels remain highest among children that are Black or Hispanic, have low socioeconomic status, and live in older housing (Egan et al., 2021; Jones et al., 2009; Levin, 2008). Hundreds of thousands of children between 1-11 years old still exceed the reference value for blood concentrations set by the CDC (Egan et al., 2021). These patterns are strong in Chicago, where high blood lead levels among children are concentrated in predominantly minority communities on the South and West sides of the city (Chicago Department of Public Health, 2021; Sampson & Winter, 2016; Tang & Carrel, 2022; White & Gala, 2022). Racial inequalities in lead exposure persist even after controlling for socioeconomic factors, though differences have declined in recent years (Sampson & Winter, 2016).

[2] For a comprehensive overview of the health effects of lead, explore the EPA's Integrated Science Assessment for Lead (U.S. EPA, 2024).

Lead poisoning disparities among racial and socioeconomically disadvantaged neighborhoods are reflected in inequities in exposure pathways. For example, lead paint is more prevalent in households under the poverty line, and soil-bound lead exposure is higher in urban areas with higher density and closer proximity to vehicle traffic (Levin, 2008). In Chicago, areas of the city with higher Black and Hispanic populations are at increased risk of lead-contamination through the municipal water system (Huynh et al., 2024; Sarfarpour et al., 2024). Recent inequities during system failures such as the water contamination crisis in Flint further demonstrate the vulnerability of disadvantaged populations to chemical exposure (Campbell et al., 2016; Hanna-Attisha, 2016).

Legislative history

Throughout the 20th century, municipalities had vastly disparate responses to reports of water contamination and pressure from the plumbing and lead production industries. Rabin describes how various states and localities prohibited, restricted, encouraged, or required lead service lines (2008). He notes that until the 1960s, the federal government did not see its role as legislating to prevent environmental health hazards, and so did not generally enact national regulations. Instead, the prevailing belief held that industries were responsible for regulating themselves. Where legislative protection on drinking water did exist, it prioritized eliminating infectious diseases from the water supply (Lewis, 1988).

This pattern of delayed regulatory coverage persisted in other lead industries as well. For example, lead was a valuable additive to gasoline because it helped the fuel to burn consistently and prevented “knocking,” which contributes to engine damage and performance issues (Jiffy Lube, n.d.). For decades, the gasoline and automobile industries operated under voluntary standards for lead (EIA, 2022). The Clean Air Act in 1970 created the first regulatory standards for lead and emissions of other automotive pollutants. In response, manufacturers implemented catalytic converters, which transformed vehicle emissions into less harmful compounds. Leaded gasoline was prohibited for vehicles after 1975 because the lead content damaged catalytic converters. Prior to this regulation, half of living Americans had elevated blood lead levels as children, when exposure is most dangerous—nearly all children in the late 1960s and early 1970s had elevated blood lead levels (McFarland et al., 2021).

In 1986, amendments to the Safe Drinking Water Act prohibited the construction of new lead pipes for drinking water across the country. In 1991, the Environmental Protection Agency (EPA) published the Lead and Copper Rule (LCR), which required water management systems to conduct quality monitoring at the faucet. It established the "action level," the concentration at which further steps to mitigate lead leaching must be taken, at 15 parts per billion (ppb) in at least 10% of tested taps. These required steps include educating residents and replacing lead service lines. The EPA revised the LCR in 2021 to target schools and childcare facilities, mandate inventories of service line materials, and change education and replacement requirements. The most recent update to the EPA rule is the 2024 Lead and Copper Rule Improvements. Under this regulation, the Biden Administration lowered the lead action level to 10 ppb and directed systems to replace all lead service lines in the next decade (starting in November 2027). However, some cities with considerable lead infrastructure, including Chicago, received a delayed deadline that allows for 20 years for total replacement instead of 10 (Circle Blue, 2024).

Chicago's water distribution system

The Chicago Department of Water Management (DWM) sources tap water from Lake Michigan and treats it through a series of purification and microbial processing steps in one of two centralized plants (Chicago, n.d.). This water is then pumped throughout the city and suburbs from the main lines of the plumbing grid. The water leaving the water treatment centers is constantly monitored for contaminants and safety standards, and is generally not considered a risk for lead exposure. Similarly, the water mains that distribute water across the city are not made of lead (Corley, 2016). In fact, lead generally enters the water supply in the last 100 feet of its journey, through the service lines that connect the mains to buildings (Reyes & Hawthorne, 2024). In Chicago, roughly 80% of these lines are lead (McCormick et al., 2022), though there is not a full, definitive record of the location and material of service lines. Instead, the DWM publishes a water service line inventory that predicts the material of a building's supply pipes based on field verification and visual inspection, construction records, historical data, building age, and self reports.

In order to prevent leaching from pipes that do contain lead, the DWM adds corrosion control chemicals that maintains a protective solid coating on the inside of the line (Chicago Department of Water Management, n.d.). However, depending on the chemical characteristics of the water flowing through the pipe, the condition of the metal, the frequency of flushing, and how long water is left stagnant, residences can still experience leaching and high levels of lead exposure. In accordance with federal regulations, Chicago also monitors lead concentrations in samples from at-risk taps, but even neighboring houses can have starkly different contaminant levels due to the number of variables influencing corrosion.

Since 2016, Chicago's 311 municipal service request system has allowed any resident to order a free self-administered lead test kit or an evaluation visit from an expert to ensure that their tap water is safe. Residents can file orders through call, text, online website, mobile app, or through aldermen and city departments. If residents request a visit, they coordinate a day and time for a technician to come to the home and conduct the test. Self-administered tests are mailed to the provided address, along with an instruction pamphlet available in English, Spanish, Polish, or Simplified Chinese. Residents then collect three samples at specific time intervals, arrange a collection time, and leave the kit outside to be collected by the city. For both the test kit and visit, residents must leave the water stagnant for a minimum of six hours before beginning to collect samples. This means that no water can be used for drinking, showering, cooking, washing, or flushing toilets.

The free testing program is vital to preventing lead poisoning because elevated results help people qualify for and engage with lead service line replacement (LSLR) programs and encourage households to use bottled water or filters in the meantime. Chicago offers five replacement programs. In the Block-Long Replacement Program, the DWM performs complimentary service line replacement when planned maintenance on the plumbing system disrupts the pipes. Under the Breaks and Leaks Replacement Program, residents are required to report service line issues to the city, which will then replace the entire line for free.[3] In the absence of public construction, breaks, or leaks, Chicago offers three voluntary programs. The Daycare Lead Service Line Replacement Program qualifies buildings that host services for young children in low-income communities for free service line replacement. Meanwhile, the

[3] Historically, the city is responsible for replacing the public side of a service line and the property owner is responsible for replacing the private side. Cost-sharing in this manner caused environmental justice issues because partial service line replacements increased exposure to lead, and lower income households were less likely to be able to complete the private side replacement (Baehler et al., 2021). However, a 2023 Illinois law mandates that lead service lines must be replaced in their entirety. The Block-Long Replacement Program and Breaks and Leaks Program are the result of this legislation (Lead-Safe Chicago, n.d.).

Homeowner-Initiated Replacement Program waives permitting fees for any resident that pursues lead service line replacement on their own. Finally, the Equity Lead Service Line Replacement Program provides complete service line replacement for owner-occupied households below 80% of the area median income, but funds are limited and the program prioritizes households that have used the testing program and received elevated lead results (Lead-Safe Chicago, n.d.).

From 2016 to 2018, nearly 70% of the returned tests detected lead (Reyes & Hawthorne, 2024). Over a longer period spanning from 2016 to 2021, 5% of the tests exceeded the federal government action level, with one remarkable test revealing 1,100 ppb of lead (McCormick et al., 2022). Around one-third of the tests surpassed the limit for lead in bottled water (McCormick et al., 2022). In many community areas, the average lead exposure exceeded that allowed for bottled water even after flushing the pipes for three minutes, despite Chicago advocating for flushing as one of the most effective ways to prevent lead poisoning (Reyes & Hawthorne, 2024). Based on tests collected between 2016 and 2023, Huynh et al. (2024) estimate that 68% of Chicago's children receive drinking water with elevated levels of lead. Still, on the city's webpage about testing and filters, the Department of Water Management declares: "To be clear, Chicago's water is safe to drink and we are taking this step out of an abundance of caution" (Chicago Department of Water Management, n.d.). The discrepancy between the available evidence and the city's cavalier attitude further emphasize the need for comprehensive and citizen-informed testing efforts.

Progress

The Lead and Copper Rule Improvements represent a significantly accelerated timeline for total lead service line replacement. Prior to the federal rule, Illinois Law gave Chicago until

2077 to eliminate the lead infrastructure (Londcor & UIS PAR, 2024). As the latest in a series of federal grants and loans, the Biden administration also provided hundreds of millions of dollars to Illinois' Drinking Water State Revolving Fund for water system improvements like removing lead pipes (Lippert, 2023). Despite increasing political attention and funding, progress has been slow. Between 2020 and 2022, municipal programs replaced only 280 of Chicago's nearly 400,000 service lines (Chase, 2022). By July 2024, Chicago had replaced 6,820 total pipes (Anastasakos, 2024). This substantial increase reflected in part the creation of new programs in response to implementation of a state law prohibiting partial service line replacements (Chase, 2020). However, even the increased rate of replacement would be insufficient to remove all lead service lines within the century, much less the federal timeline of twenty years. Chicago lags far behind other cities like Newark, New Jersey, which replaced all of its 23,000 service lines in under three years (Cunningham, n.d.). At an estimated $15,000 to $26,000 per service line, total service line replacement in Chicago could cost up to $10 billion (Circle Blue, 2024).

Literature review

Theoretical Frameworks

Handling resident needs is complex. As Borins (2008) explains:

> Wide sociological gulfs can separate officials from their publics… exacerbated by social inequalities and group discrimination. When they face these social and cultural walls, even well-intentioned officials can have a difficult time discerning a community's needs and devising effective means to satisfy them (p. 57).

Due to informational gaps and deficits of understanding, direct citizen involvement in the policy making process appears beneficial for guiding government decisions. Unfortunately, participation is not equally accessible to all.

In their book *Administrative Burden: Policymaking by Other Means*, Pamela Herd and Donald Moynihan (2018) theorize that citizen engagement with government is mediated by barriers to resource or service provision. The authors outline three major types of costs that typically construct this "administrative burden." *Learning costs* involve the process of gaining awareness and understanding of a program, *compliance costs* concern the regulations and instructions to successfully participate, and *psychological costs* include the mental toll that the process exacts. While costs are inevitable to an extent, in many cases they are intentionally imposed as the result of policy choices to restrict access or discourage participation. Even when an administrative burden is unintentional or the result of insufficient capacity, it can significantly undermine the efficacy of program implementation

Moreover, Herd and Moynihan explain that administrative burden is "distributive" (2018, p. 3). Bureaucratic costs exacerbate societal inequalities because the people with the fewest resources have the most difficult time overcoming burdens and accessing resources, despite often being the people that need those services the most. These resources are not simply monetary, but also include "forms of human capital, such as education, cognitive and noncognitive skills, or a social network" (Herd and Moynihan, 2018, p. 7). In addition to social, cultural, and economic capital, access to administrative capital (bureaucratic expertise) shapes how people respond to and their success in overcoming administrative burden (Christensen et al., 2020; Masood and Nisar, 2020). Low uptake rates for many programs reflect the existence of prohibitory barriers.

The 311 testing program requires cooperation between the government and citizens to produce government services in a process called "coproduction" (Brudney & England, 1983). On the surface, the 311 program for requesting tap water lead tests aims to minimize burdens for citizens. There is no fee or financial barrier associated with the program. Participants can request tests through multiple platforms like a website, app, or phone call. The city provides the option to either send an expert to conduct a test in-person or a self-test kit with step-by-step instructions. In fact, coproduction through 311 service systems in general improves communication between citizens and their government, boosting responsiveness and satisfaction with local governance (Clark and Shurik, 2014; Clark and Shurik, 2015).

Yet, the voluntary testing program is a form of engagement triggered by the resident, and citizen-initiated contact theory suggests that socioeconomic status determines which people will originate communication with their government in order to receive services (Jones et al., 1977; Thomas, 1980). This "socioeconomic model" posits that higher levels of awareness and financial, human capital, and psychological resources enable privileged individuals to have higher levels of political participation (Jones et al., 1977; Thomas, 1980; Verba and Nie, 1972, Verba et al., 1993). Though specific models vary, empirical studies suggest that both level of need (or self-perceived need) and socioeconomic status influence which citizens initiate contact (Jones et al., 1977; Sharp, 1982; Thomas, 1980). In this case, socioeconomic status mediates the relationship between needs and action, and people with certain characteristics (e.g., low levels of education, income, and social-cultural capital) are less likely to be willing and able to engage with the government in order to have their needs met.

However, quantifying the determinants of citizen participation has been complicated empirically because socioeconomic status is used as both a measure of need and as a measure of awareness or other resources that make initiating contact more feasible (Jones et al., 1977). The presumptive inverse relationship with need and direct relationship with resources makes teasing out the impacts of socioeconomic status difficult. Therefore, researchers have theorized and tested positive, negative, and parabolic relationships based on the balance between the two factors (Jones et al., 1977; Sharp, 1982). The voluntary lead testing program in Chicago presents a valuable opportunity to differentiate these forces because contamination risk (objective need for services), requests (demand for services), status of requests (service provision), and socioeconomic status can be determined separately; this represents a unique opportunity because such data is usually limited (Cook et al., 2024).

In addition to providing valuable data for research, 311 systems also help inform city governments on resident needs, so that services can be appropriately distributed (Cook et al., 2024). Cook et al. provide a simple model for how municipalities provide services (2024). Under their conceptualization, there are two ways that need translates into provision of services. First, city monitoring systems may pick up directly on the status of city needs and respond proactively by addressing the issue. Second, need can be translated through citizen involvement when residents demand services based on their experiences, such as through 311 reporting systems. However, the existence of administrative burdens and citizen-initiated inequalities mean this process may not reflect true levels of need. When cities rely on citizen complaints to decide who receives services, differing likelihoods of participation perpetuate and reproduce systems of inequality (Slough, 2024). Frequently, variable engagement occurs along income and racial lines (Cook et al., 2024). Once demand is expressed, cities respond by fulfilling (or not fulfilling) the desired service.

Under this framing, inequality in the realm of municipal services can come from one of four paths. First, historical and modern patterns of social, economic, and political developments create heterogeneous distributions of service needs (Chicago Department of Public Health, 2021; Sampson & Winter, 2016; Tang & Carrel, 2022; White & Gala, 2022). Second, informational gaps on patterns of need and prioritization of specific groups or areas of the city bias where the city pursues action (Mladenka, 1980). Third, citizens have varying propensities to demand government response based on the resources they have access to and the barriers they face (e.g., Minkoff, 2015). Finally, cities respond differently to demand coming from different people or geographic regions (Jones et al., 1977). Underlying distributional patterns of lead poisoning has already been addressed above. In the following sections, I address the existing literature on each of the other routes to environmental injustice as they apply to lead contamination in Chicago.

Responsiveness to need and demand

Municipal responsiveness matters in two ways: proactive management of lead risk measurement and reactive follow-up to resident demand expressed through 311 requests for lead tests. Federal legislation through the Lead and Copper Rule standardizes how objective measures of lead risk are determined through sampling (40 C.F.R. §141.86 (1991-2025)). These regulations are supposed to guarantee that municipal water systems accurately capture the extent of contamination. The testing is compared to the federal action level (currently 10 ppb) to determine whether the system needs to take additional steps like educating residents or replacing service lines.

However, to meet the requirements for compliance, Chicago only tests 50 homes once every three years, most of which “were owned by water department employees or retirees living on the Far Northwest and Far Southwest Sides” (Reyes & Hawthorne, 2024). This sample is exceedingly small, infrequent, and biased towards areas of the city with higher socioeconomic status. Evidence from the Chicago Lead in Drinking Water Study in 2011-2012 found that the federal sampling protocol underestimates lead contamination in drinking water (Toral et al., 2013). Results from the 311 citizen testing program that show widespread and severe lead contamination further emphasize that the regulatory-based testing results are extremely unrepresentative of the true risk Chicago residents face from lead exposure (Huynh et al., 2024; Reyes & Hawthorne, 2024). As a Virginia Tech researcher that serves as an EPA advisory panelist explains: "Chicago's testing blows out of the water one of the foundations of (federal regulations), namely that current lead-in-water monitoring requirements yield reliable information about the extent and severity of contamination across a service area" (Reyes & Hawthorne, 2024).[4] The sampling that the city provides outside of the 311 system is therefore insufficient to assure universal clean water access, leaving citizen-initiated testing as an important health pathway.

Historically, studies have shown that cities responded quicker to citizen-initiated contacts requesting services when those contacts originated from wealthier neighborhoods (Jones et al., 1977). However, as technology advances, bureaucratic decision making becomes more automatic, standardized, and algorithm based (Clark et al., 2018). In these modern systems, there is less discretion where bias and discrimination can penetrate and effect government responsiveness (Maynard-Moody and Musheno, 2012; Nisar and Masood, 2019). Indeed, across 311 systems in 15 large US cities, governments tend to respond equitably regardless of where the request is coming from (Clark et al., 2020). Though results vary significantly by request type and city, the authors conclude that there is no substantively significant or systematic bias in response times. When differences in responsiveness do occur along racial and wealth lines, they are largely attributable to different types of requests (Hamel & Holliday, 2024; Wichowsky, 2022). Among requests of a given type, there is limited and contradictory evidence of discrimination in meeting citizen demand (Hamel & Holliday, 2024; Wichowsky, 2022).

[4] Note that updates to the Lead and Copper Rule attempt to address some of these testing issues.

In the case of Chicago's lead test request system, the opportunity for bias is particularly scarce. Kits are sent out centrally through the United States Postal service, and residents themselves establish pickup times for collection by the city (Souders et al., 2024). Once kits are picked up by staff members, they are tested by the city and results are sent back in 6-8 weeks (Souders et al., 2024). This means that it is reasonable to expect Chicago to send lead test kits to citizens that request them, and analyze those that are correctly completed and returned, without bias. Requests for in-person visits may have more potential for bias, but the existing literature suggests it is likely minimal (Clark et al., 2020; Hamel & Holliday, 2024). Overall, responsiveness from the city is generally inadequate as a direct response to need but equitable as a response to demand. This dichotomy highlights the critical importance of high levels of representative citizen participation to ensure provision of government resources.

Propensity to complain

The likelihood for citizens to initiate a government contact depends on financial, social, cultural, psychological, and administrative resources that leave significant possibility for inequalities at the intersection of administrative burden and citizen-initiated contact theory. However, the empirical literature reports conflicting results on the relationship between "propensity to complain" and socioeconomic and demographic factors (Kontokosta et al., 2017). For example, researchers have found that citizen reports of heating and hot water violations in New York City are significantly lower for communities with lower income, less education, higher proportions of English-limited citizens, and greater percentages of minority residents (Kontokosta et al., 2017). More generally, income is positively associated with propensity to contact the New York City 311 system for government goods, while minority makeup is negatively associated (Minkoff, 2015). Similarly, in Houston, high prevalence of Hispanic and

Black residents as well as low socioeconomic status correspond to a lower propensity to file a 311 report for potholes, despite relatively low costs of using the system (Cook et al., 2024). This has led to the perverse result that objective need (pothole number) is higher in the areas with fewer requests, emphasizing the systematic discrepancy in engagement from minorities and low income residents. Impediments to navigating bureaucratic processes for people without certain social resources often filter out those with the most need, even when there is no financial cost.

In contrast, San Francisco's 311 system shows no evidence of biases in engagement based on socioeconomic status (Clark and Brudney, 2018). Additionally, inclusion of disadvantaged groups has increased over the time the program has been in operation, suggesting a learning effect and increased awareness (Clark and Brudney, 2018). In Tallahassee, 311 requests relating to power outages in the wake of a hurricane were disproportionately created by minority groups, potentially suggesting that modern technological resident engagement systems can narrow distributional inequities (Xu and Tang, 2020). Meanwhile, in Boston, lower family income reduced the likelihood of utilizing 311 services, but there was little evidence of a racial impact, and higher levels of education actually decreased the propensity to contact (Clark et al., 2013). It is important to note that these seemingly conflicting studies may be the result of inadequately accounting for heterogeneous need distribution between groups. If the need for government services is elevated among disadvantaged groups, but willingness and ability to pursue those services is low, it can be hard to distinguish between the competing effects.

When it comes to Chicago's lead testing program, administrative burden dynamics have been theorized to apply, but this proposition has not been thoroughly empirically tested (Souders et al., 2024). Preliminary evidence suggests that there is a negative correlation between census blocks with high Black and Hispanic populations and the likelihood of self-conducted testing

(Huynh et al., 2024). While informative, Huynh and colleagues' study is primarily focused on estimating childhood lead exposure, and the investigation into testing behavior is limited to the probability of a census tract having at least one completed test of lead concentration.

This body of literature is vital in establishing a quantitative relationship between sociodemographic features and political participation, yet it does little to elucidate which barriers are influencing the macroscopic trends. In order to develop a comprehensive framework of administrative burden in the free lead request system in Chicago, it is essential to catalog the potential barriers and assess support for their impact.

The 311 approach offers key benefits that may ameliorate several sources of burden. It eliminates financial burden by providing testing kits or visits for free. The process and outcomes are highly routinized, leaving limited room for discretion in implementation that can result in discrimination and discourages portions of the population from participating (Maynard-Moody and Musheno, 2012; Nisar and Masood, 2019). At the same time, the defined procedure avoids the feeling of a special appeal, a perception that limits willingness to engage (Thomas, 1980). This also dampens the role feelings of political efficacy have on willingness to engage (Bovaird et al., 2014; Prats and Meunier, 2021), and since political efficacy is associated with socioeconomic measures and race (Brady et al., 1995; Leighley and Vedlitz, 1999), reduces inequality. Finally, the program is universally accessible, rather than means-tested, which eliminates paperwork requirements and sources of shame or stigma (Herd and Moynihan, 2018).

Despite these benefits, significant barriers remain. First, residents must have risk knowledge, or awareness of the possible existence of lead contamination and the associated health impacts. Additionally, they require "action knowledge," or awareness that the program exists and an understanding of how to use the 311 system to request a test (Flanagan et al., 2015).

These barriers represent *learning costs* in Herd & Moynihan's framework. Third, residents need to have the capacity to go through the request system and successfully complete the test, and moreover they must have sufficient confidence in their ability to do so to be willing to commit time and energy to the process. This capacity includes literacy in one of the languages offered, available time to order, conduct and return the test, ability to leave water stagnant for the required time, ability to follow the instructions to successfully collect water samples,[5] means to set up a collection time, and capability to overcome frequent glitches and problems with the system.[6] These barriers represent *compliance costs* in Herd & Moynihan's framework. The option to request an in-person visit from a professional may help alleviate some of these burdens. Finally, the level of trust and confidence residents have in governmental public health systems regulates access to the program (Teodoro, 2022). Indeed, some Chicago residents express suspicion of the lead service line replacement programs (Chase, 2022).

These forms of personal, social, cultural, and administrative burden disproportionately affect historically disadvantaged populations (Moynihan and Herd, 2018). For example, in a study on superfund sites, researchers showed that certain measures of socioeconomic status influence awareness of environmental hazards (Rhubart and Robertson, 2020). Additionally, phone and internet-based services offer accessibility advantages, but technology-mediated services can undermine participation from under-resourced groups, reinforcing existing inequalities in access (Das and Das, 2021; Linos et al., 2021; Pew, 2010).

[5] This is less straightforward than one might assume. In focus groups, the instruction manual sent along with lead test kits (before redesign in 2022) "was described as intimidating, confusing, and overwhelming" (Souders et al., 2024). The faculty advisor on my thesis, a public policy professor at the University of Chicago, realized she had in fact made a mistake in collection after returning the water samples.

[6] For example, in October, 2024, the confirmation email for requesting a kit came with the following message: "We apologize as the city's 311 system is experiencing some technical IT issues with service request creation. If there are any changes or corrections in the above information please contact Chicago Department of Water Management."

The closest analogies to the burdens operating in Chicago's 311-based lead test request system come from other public health testing environments, and these cases reinforce the importance of barriers and their association with sociodemographic factors. Education doubles consent among opioid users to take free, at-home HIV tests, and the overall rate of completion and follow up is low (Ballard et al., 2021). The researchers suggest that the at-home option solves several existing barriers to HIV testing such as limited capacity, stigma, concerns about confidentiality, and transportation access, but other barriers such as difficulty with returning the tests and trust of the system continue to limit successful uptake of the program. Flanagan et al. (2015) surveyed communities in Central Maine with high potential for arsenic contamination that rely on private wells for drinking water, finding that less than half of well owners have ever tested for arsenic. Moreover, the probability of testing increases with education and income, and the study points to four factors that increase testing for arsenic: knowledge of the health risk, knowledge of what can be done to address the risk, belief that testing is not onerous, and neighborhood norms for testing.

During the water contamination disaster in Flint, the local government provided free, voluntary lead tests for tap water (Chojnacki et al., 2017). Similar to Chicago's system, residents had to conduct the tests themselves and submit them for analysis. They also had to pick up the tests from a distribution center, which may provide obstacles to lower income residents who have long work days and lack action knowledge (Chojnacki et al., 2017). In addition, the authors note that the self-testing procedure is finicky, relying on citizens to leave their water stagnant for an extended period of time and sampling the first liter–mirroring Chicago's system–which may not be feasible for many households. Indeed, more valuable homes were associated with a higher likelihood and frequency of lead tests, suggesting that self-selection perpetuates disproportionately high exposure to lead for less affluent residents (Chojnacki et al., 2017). After the Flint water crisis, Black residents were more likely to perceive the contamination as intentional, racially motivated, and discriminatory on the basis of class (Ezell et al., 2022). Patterns like this suggest that certain groups of residents distrust municipal water management programs, which could limit their participation in government-run testing services. However, the relevance of case studies on Flint may be limited in their applications to Chicago because of the structural differences between an urgent, salient environmental disaster, and the more diffuse, continuous effects of infrastructure.

Insight from Chicago's lead request system

Despite ample scholarship on the creation of inequalities in political participation and government service provision through the intersection of sociodemographic factors with administrative burdens, no study has capitalized on Chicago's 311 request database to investigate lead testing behavior through this lens. Though administrative burden theory has been applied to this system (Souders et al., 2024), empirical treatment has focused on which census blocks have been tested at all (Huynh et al., 2024) or how interventions decrease compliance costs once requests are made (Souders et al., 2024). These studies leave out the rich data available on the request side of the database. By accessing the 311 request log through the Chicago Data Portal, this thesis will assess mechanisms for the observed patterns in testing disparities. Information such as how a request is filed, the type of test requested, the status of the request, and change over time can be exploited to test theories about the barriers facing self-initiated testing.

Moreover, this approach diverges from previous administrative burden literature by focusing on a status quo context rather than a crisis (Chojnacki et al., 2017; Ezell et al., 2022; Xu and Tang, 2020), assessing more than one year of data (Clark et al., 2013), and isolating the impacts of burden versus objective need on the observed demographic participation discrepancies (Cook et al., 2024). In light of the 2024 revision to the EPA's Lead and Copper Rule, which mandates the identification and replacement of all lead pipes within a decade, elucidating the obstacles and pitfalls of voluntary citizen-led testing is vital. Understanding what makes voluntary testing ineffective will help policy makers design more equitable programs that target those at greatest risk of both lead exposure and being left out of government services.

Methods

Analyzing the dynamics and barriers in Chicago's free lead test request system required two types of data. The first is the 311 Service Request Dataset, which I accessed through the Chicago Data Portal. Since this database contains requests for all municipal services and information, I queried and downloaded the entries coded with the SR Type WCA2 or WCA3, which correspond to a Water Lead Test Kit Request and Water Lead Test Visit Request respectively. I downloaded this data on October 24, 2024. The columns of this dataframe include the following relevant information: request type (self-testing kit vs professional testing visit), department that created the the request (including aldermen, the Department of Water Management, or 311), status (open, closed, canceled), manner of origin (eg internet, phone call, email), date of creation and last modification, and geographic location. The Department of Water Management webpage claims that "the results [of the lead tests] will be posted on DWM's website at the 'block level' - this means that the last two digits of the street number of the address will be removed and replaced with 'XX.'" However, this anonymization is not present on the request side. Each request includes the full, uncensored address, which can ostensibly be matched to the results when combined with the date and other information, eliminating the privacy the city claims to offer.

The second set of information is demographic details for the city. I obtained estimates for educational attainment, English fluency, affluence, poverty, internet access, and race by census tract through the American Community Survey conducted by the US Census Bureau. In particular, I used the 2022 ACS 5-year estimates of census tract level characteristics. The key variables I extracted to represent those characteristics are summarized in Table 1. These data points are separately provided in social, economic, and demographic characteristic datasets, which I downloaded for all the census tracts in Cook County, Illinois and then merged into one dataframe through their Geo IDs.

Demographic characteristic	**Representative ACS variable**	**Variable interpretation**
Education	DP02_0067PE	Percent of population 25 years and over that are high school graduates or have higher educational attainment
English fluency	DP02_0115PE	Percent of population 5 years and over that speak a language other than English at home and speak English less than very well

Affluence	DP03_0062E	Median household income in 2022 inflation-adjusted dollars
Poverty	DP03_0119PE	Percent of families whose income in the past 12 months is below the poverty level
Internet access	DP02_0154PE	Percent of households with a broadband Internet subscription
Race	DP05_0067PE DP05_0073PE	Percent of population that is Black or African American Percent of population that is Hispanic
Households	DP02_0001E	Total number of households
Housing age	DP04_0021PE – DP04_0026PE	Proportion of housing stock of different ages

Table 1: Definitions of demographic variables from the American Community Survey.

Due to the political circumstances of lead pipe prescription and prohibition, housing age provides a strong proxy for risk of exposure. I produced this variable by calculating the percent of houses in the geographic unit built before 1989, and used this as an explanatory variable for engagement behavior. There is a slight mismatch between the data threshold on housing age (1989) and the legislative boundary (1986), but the two year transition period for enforcement brings the dates nearly into alignment.

I conducted all statistical analysis and mapping in the R programming language. In order to merge the request and demographic data, I had to assign each request to the census tract in which it belongs because the geographic designations from the request database and the census information are misaligned. Using the tidycensus, tigris, and sf packages, I retrieved shapefiles for the Illinois census tracts and the Chicago boundary shapefile, and removed the census tracts located outside of the city. I then converted longitude and latitude coordinates from the request database into a geospatial data type, and used this information to link each request to the tract in which the coordinates fall. Finally, I created a summary count of the number of requests, proportion of each type of request, percent of completed requests, and origin of request for each census tract and merged it with the ACS data by matching Geo IDs.

During this process, I also cleaned the data by ensuring data were of the correct type and removing unreadable or missing entries (e.g. income above 250,000 is marked as 250,000+, which cannot be read as an integer). After removing erroneous request entities, I was left with 97,278 requests.[7] Finally, I divided the total number of requests by the total number of households in each tract to create a requests per household metric that accounts for differential population between geographic units. To better understand the structure of the assembled data, I computed descriptive statistics and maps of the explanatory and outcome variables. I plotted

[7] On April 2, 2024 there appears to have been a system error. The Department of Water Management filed several hundred requests for each of nine addresses on this day, all of which were canceled a day later on April 3, 2024. For one additional address, 74 requests were filed on May 2, 2023 by phone call, and all but one were canceled a day later on May 3, 2023. It seems as though the first request, which was eventually completed, was real, but the remainder were in error. I removed these entries (excluding the one completed request, n = 3,897) from the dataset on the assumption of a glitch, leaving 97,332 requests. I also removed requests with addresses corresponding to the Department of Water Management and 311 Operations center (n = 54), leaving 97,278 requests.

each variable from the American Community Survey in a box plot showing the distribution of that variable across census tracts. I also visualized the spatial distribution of the indicators within the municipal boundary. I repeated this procedure for the number of requests, number of requests per household, type of request, and percent of completed requests. I also created bar graphs for the origin and department that created the requests.

Once I joined the datasets together and had a basic understanding of the traits, I began statistical analysis. I created scatter plots and ran linear regressions between the average number requests per household as the outcome variable and the set of demographic features as explanatory variables. Initially, I conducted single variable linear regression models with each explanatory variable independently. Next, I included all of the explanatory variables in one multivariable linear regression model to differentiate the relative strengths of the correlative relationships and separate possible confounders. I tested the assumptions of this model by plotting the residuals in a Q-Q plot and conducting a multicollinearity correlation matrix.

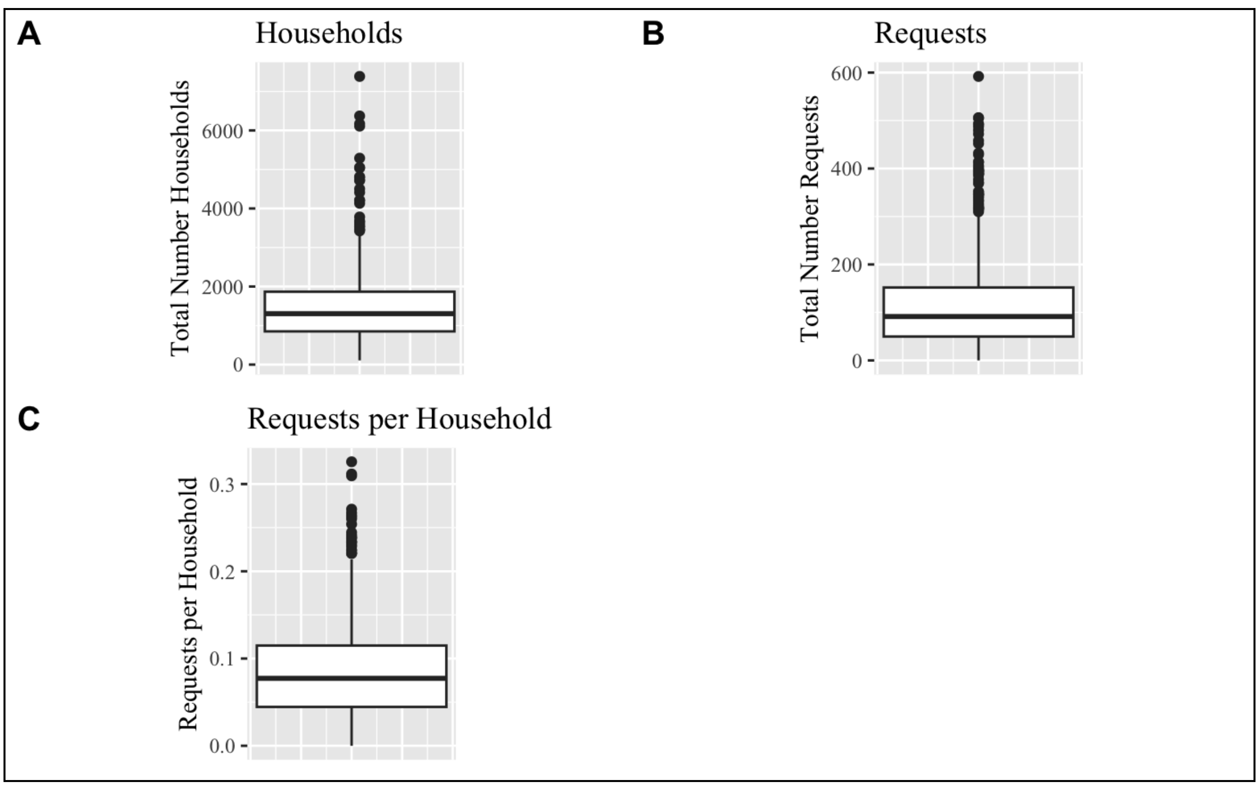
In addition to investigating the connection between demographic characteristics and likelihood of engaging with the testing program, I explored potential underlying mechanisms for the observed patterns. I replicated the single and multivariable analyses with the same explanatory variables and used the proportion of requests for visits, percent of completed requests (older than one year), and the fraction of requests filed through the internet as outcome variables.

Finally, I investigated changes over time. I plotted the request volume for each day within the study period by grouping the request database by the created date. I also determined the change in regression coefficients over time by splitting the request database into year-long segments and repeating the regression of requests per household against explanatory variables in each year group.

Results

Key features of request database

The City of Chicago's database on lead requests reveals drastically low utilization of the program. The median number of requests per household among census tracts is 0.077, meaning that on average only around 8% of households in a given census tract have capitalized on the free provision of tests (Figure 1).[8] While several census tracts exhibit significantly higher rates of program engagement, no census tract has had more than one third of households use the 311 request system.



[8] This assumes that multiple tests are not being requested within a single household.

Figure 1: Summary statistics on the number of households (A), total number of requests for lead tests (B), and the requests per household (C) in Chicago. Each data point represents one census tract and requests per household is calculated as the number of requests divided by the number of households in each census tract (n = 836).

Furthermore, the majority of requests were never successfully completed. The database has a column titled "Status" which records the current condition of each entry as "Completed," "Open," or "Cancelled." An "open" status indicates that the water sample has neither been successfully returned to the city and analyzed nor has the file been explicitly cancelled. I removed the most recent year of data under the assumption that these open entries may still be in progress, and that any requests not completed within a year would remain unfinished. I chose this cutoff because the average length of time between the created and closed date for completed requests is 255 days, and 75% of completed requests took fewer than 283 days. The median proportion of completed requests in a census tract is 34.8%, with the middle half of data falling in the narrow band between 30.1% and 39.9% (Figure 2). Few census tracts have a completion rate above 50%, and only one has a 100% completion rate.

While citizens can choose between a self-administered kit and an examination conducted by a city-approved expert, the vast majority (87%) of requests are for self-test kits. Across census tracts, the median proportion of requests for an in-person visit is 10.4%, indicating strong preference for performing the sampling oneself (Figure 2). However, the range of data has a fair amount of spread, with one census tract having 60% of its requests made for the city official option.

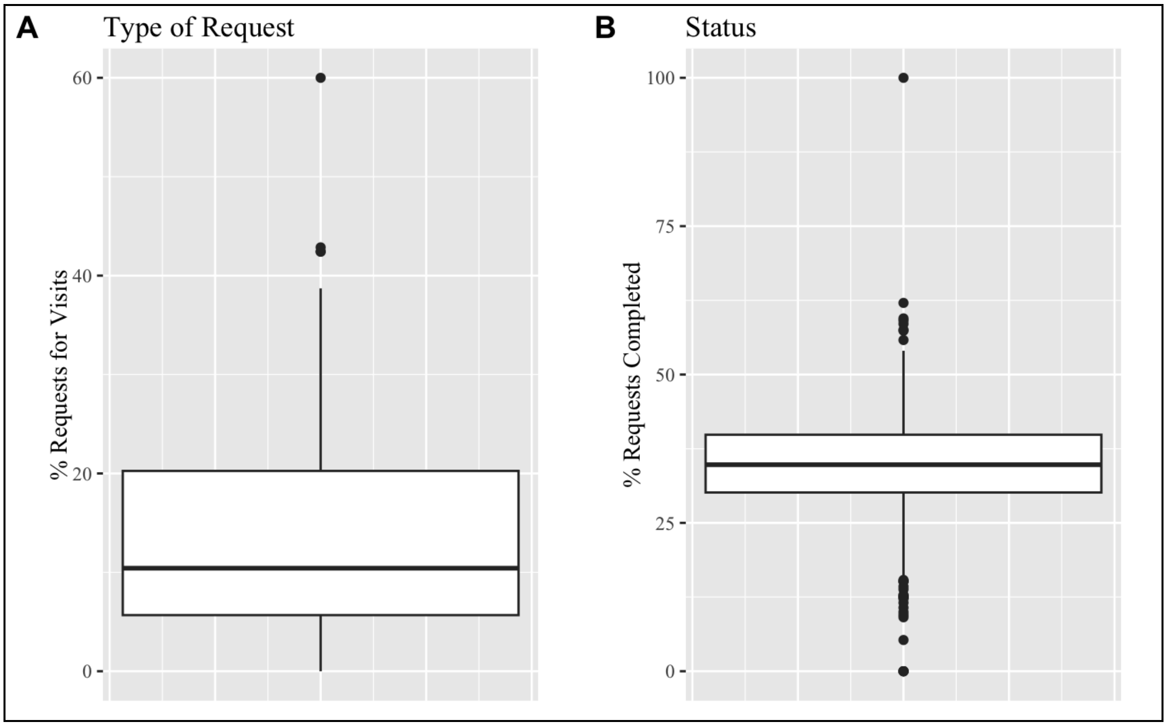
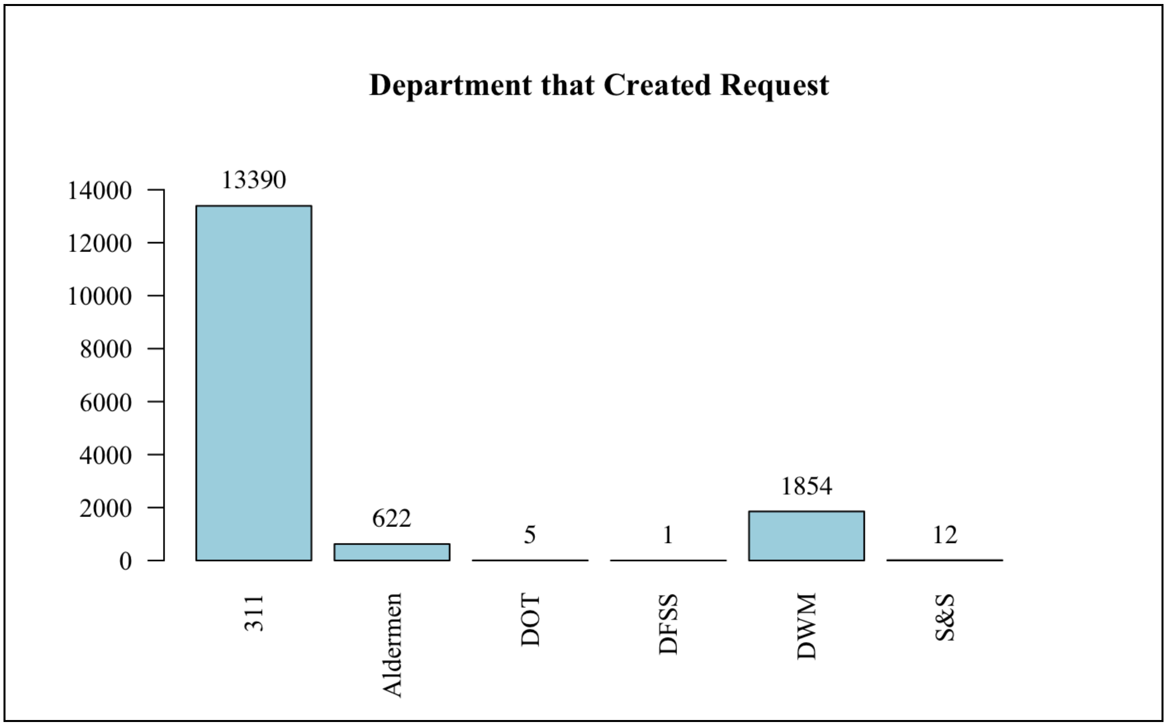


Figure 2: Summary statistics on the type of lead tests requested (A) and the current status of requests (B). The distribution of the type of request is expressed through the percent of requests for visits from professionals and the distribution of status is expressed through the percent of completion. Each data point represents one census tract, with all census tracts with at least one request included (n = 783).

Two other notable aspects of the request data are the department that initiated the request and the request's origin. While providing similar and overlapping information, the former indicates whether the service request was opened by a city department and the latter represents how the request was filed, whether by City operations or individual usage (Figure 3). Nearly all of the requests opened by a department were created by 311, the non-emergency municipal service system, or the Department of Water Management (DWM). Aldermen made up the largest share of the remaining requests, with the Department of Transportation (DOT), Department of Family and Support Services (DFSS), and Department of Streets and Sanitation (S&S) accounting for the rest.



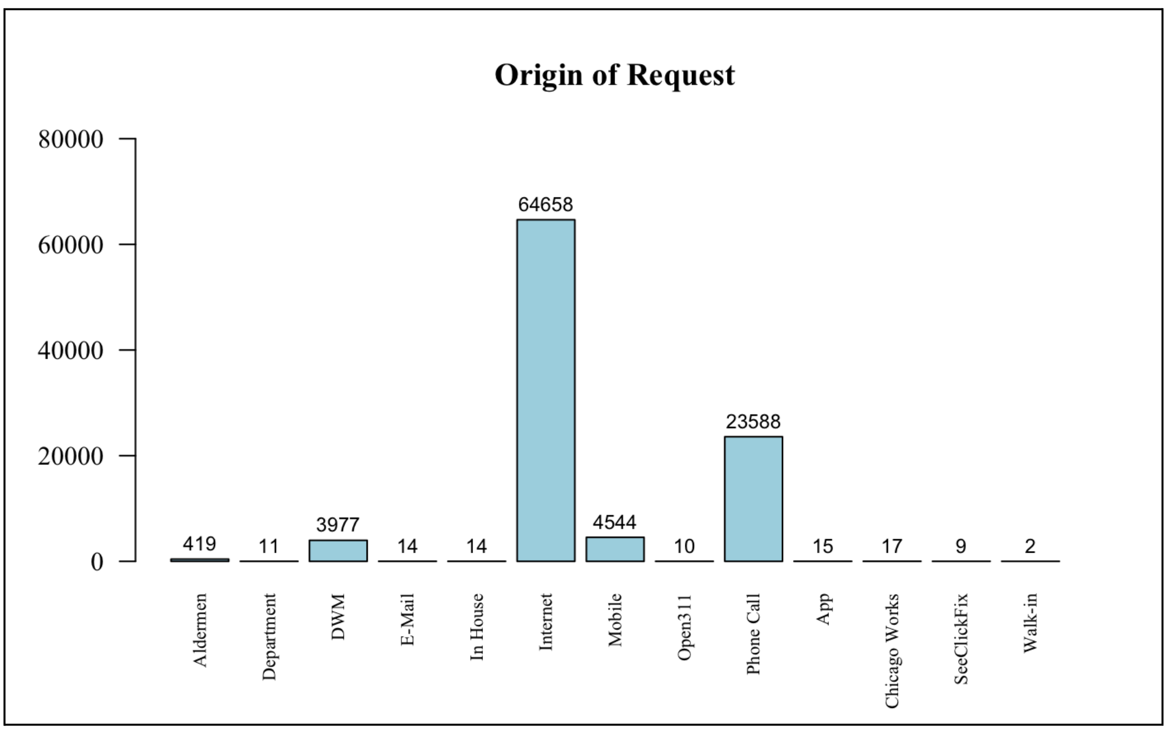


Figure 3: Frequency distribution of the creating department (top) and origin (bottom) of lead tests requested. Frequency is the number of individual requests filed from 2018-2024.

The origin data shows that most requests were inputted through the internet, with phone calls and mobile devices activity the second and third most frequent. Figure 4 summarizes the origin, type of request, and current status of the 97,278 requests. Crucially, requests for visits are significantly more likely to be completed (48%) versus requests for self-test kits (31%).

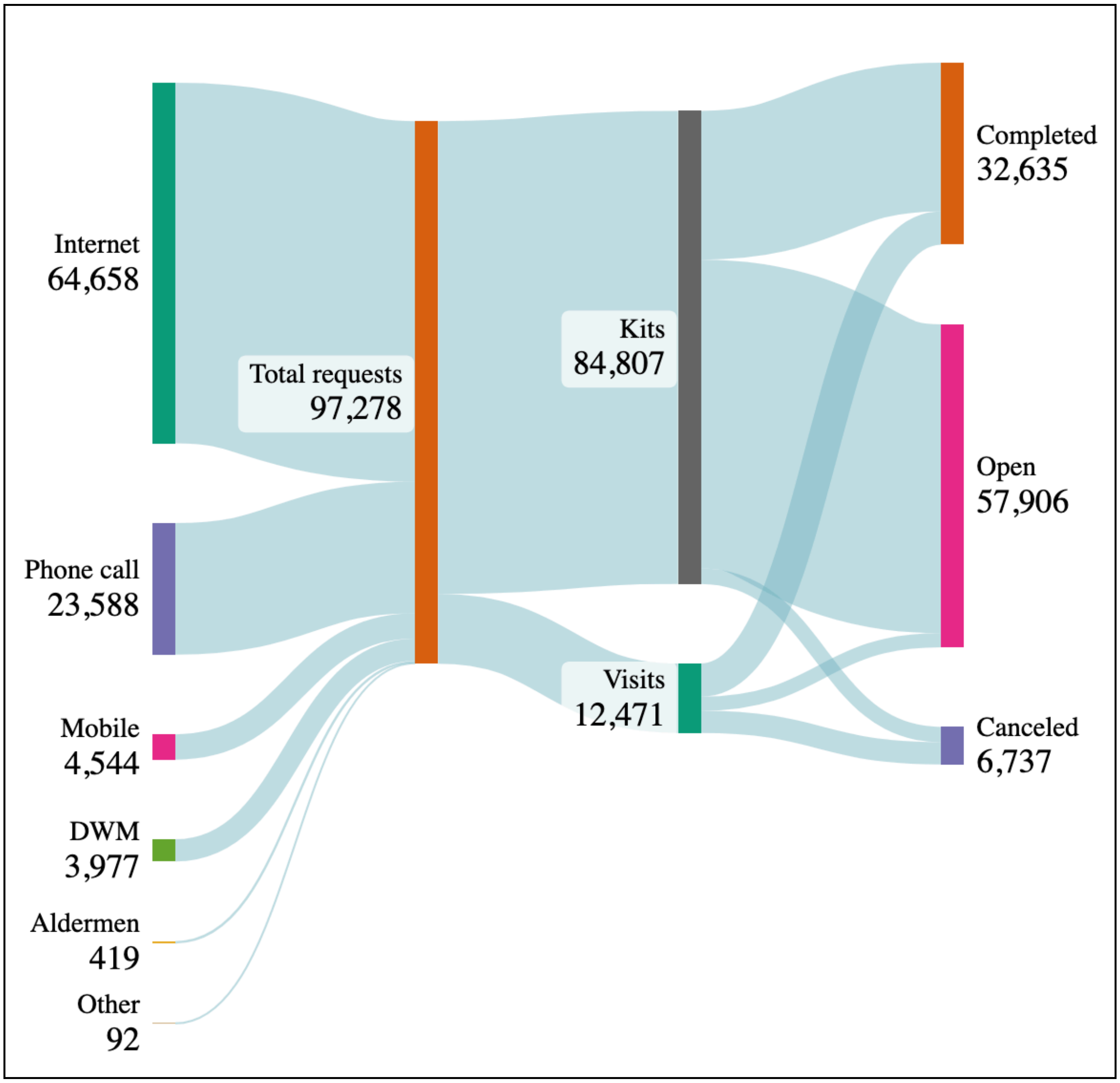


Figure 4: Sankey diagram tracking the flow of 311 lead test service requests from origin, to type of test, to current status (total n = 97,278).

Spatial distribution

Mapping requests by census tract across the city reveal characteristic spatial patterns in the use of the free lead test program (Figure 5). The highest numbers of requests are from census tracts on the Northwest and South portions of the city. After accounting for variability in the population and household sizes of census tracts by dividing requests by number of households, this distribution persists. High and low testing behavior clusters across the city.

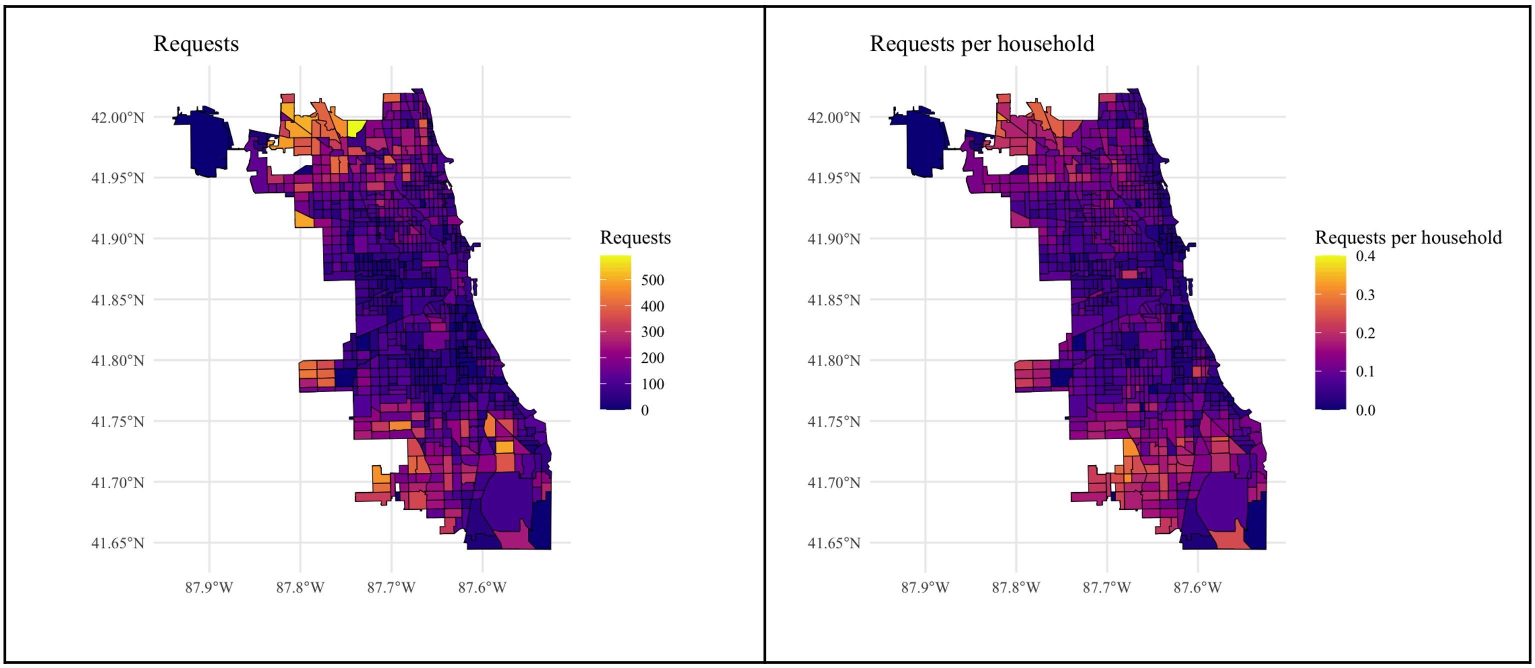
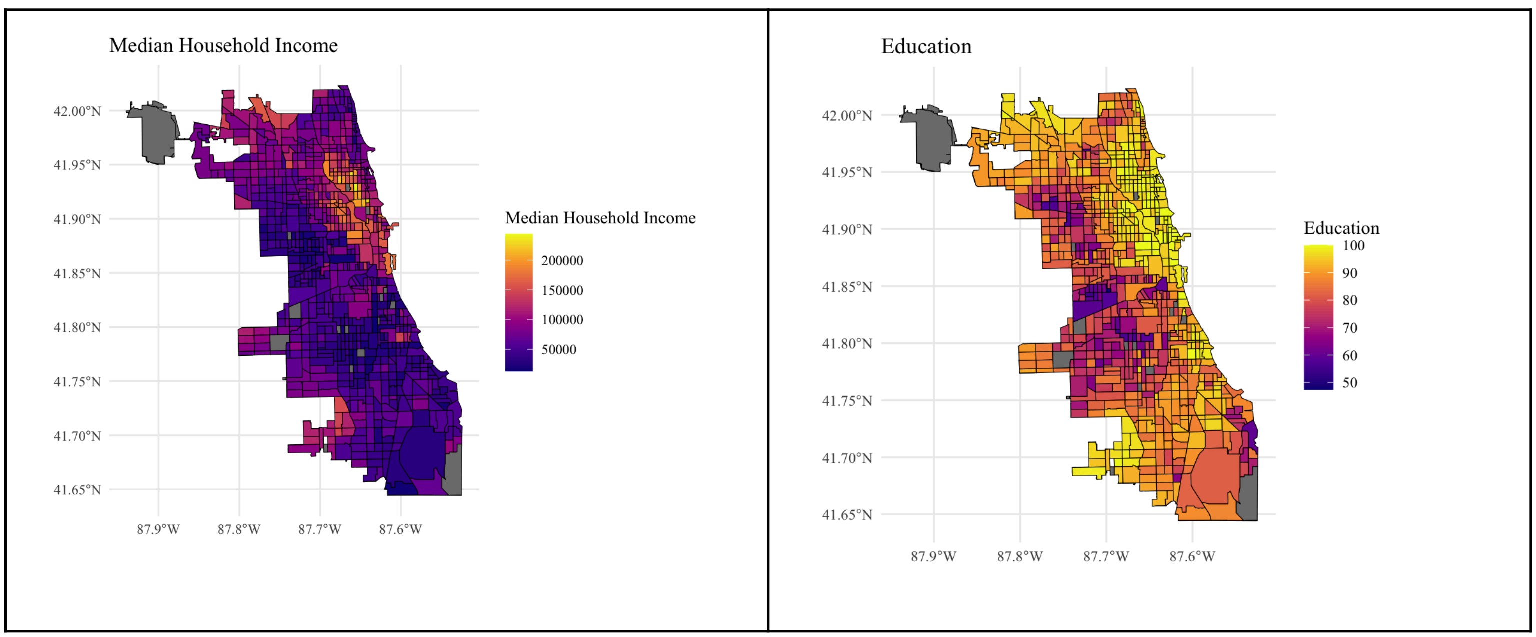
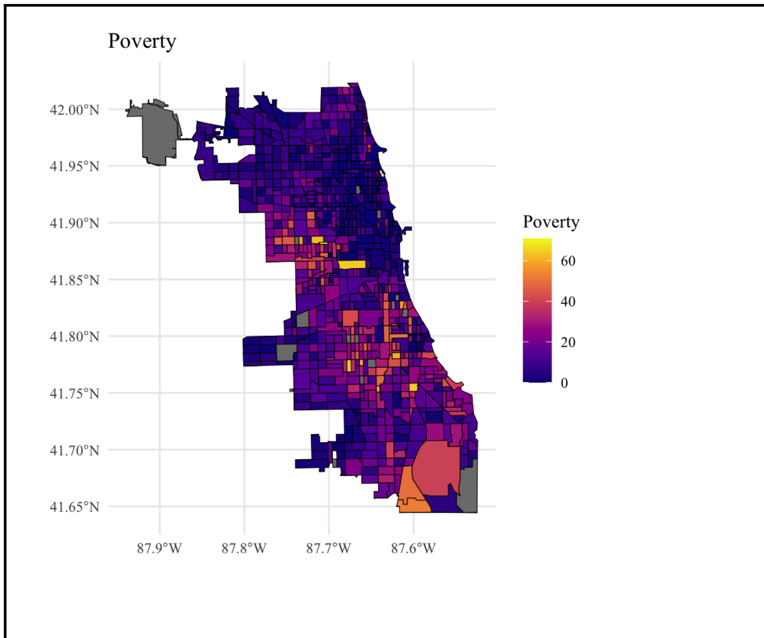
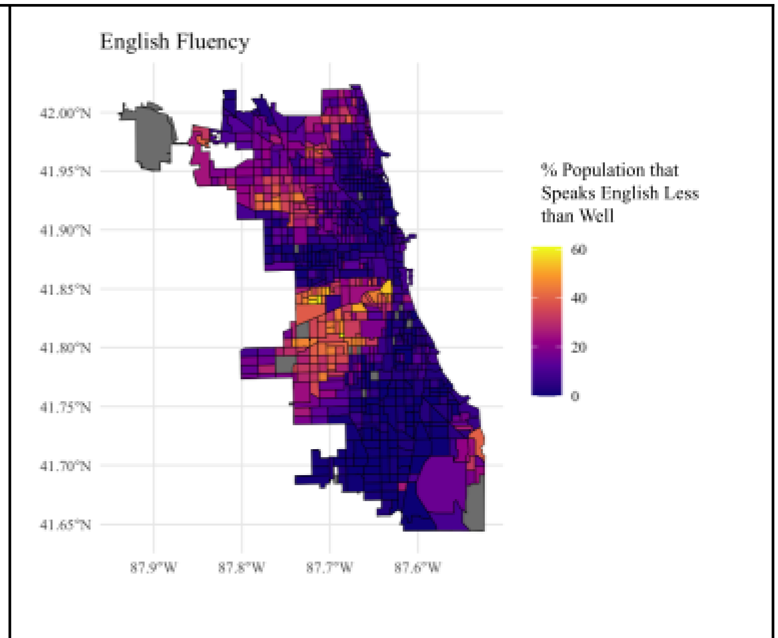


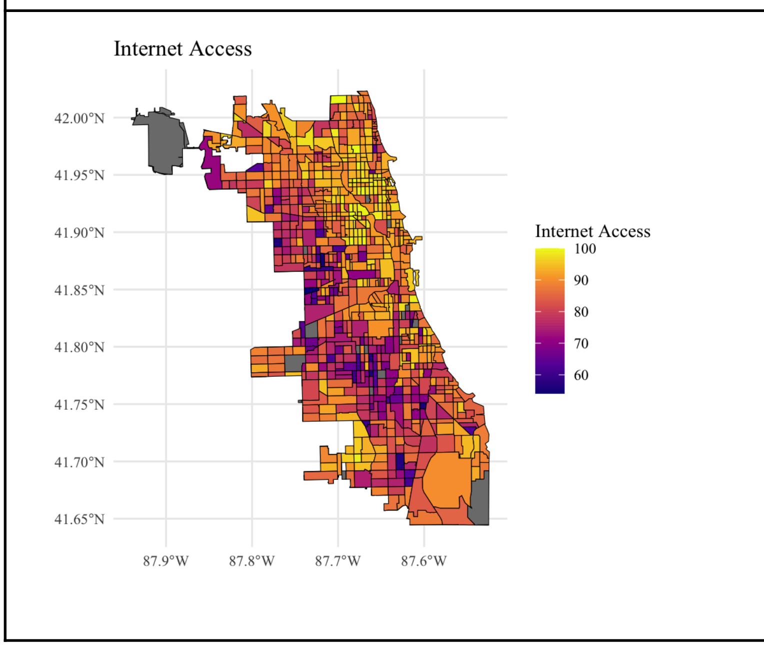
Figure 5: Heat maps of the total number of requests and the average number of requests per household by census tract in Chicago. Lighter colors (yellow end of the spectrum) represents higher values.

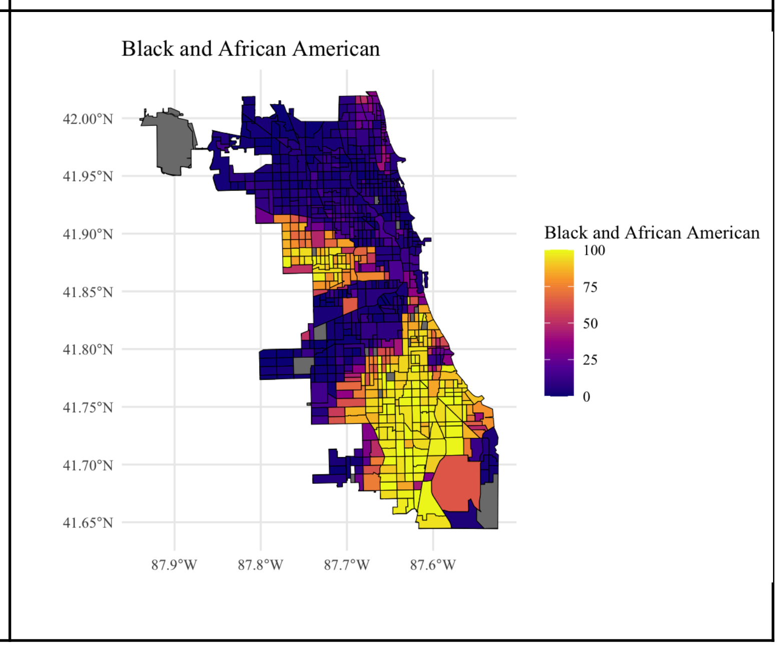
These patterns can be visually compared to demographic distribution maps (Figure 6) as an initial investigation into the determinants of testing behavior. For example, the census tracts with high median household incomes and education levels (excluding the downtown area) are on the far Northwest and Southwest sides of the city, where the highest testing behavior occurs. The downtown area has the lowest proportion of houses built before 1989, which is a strong proxy for lead water contamination risk given the legislative history of Chicago. Thus, mapping suggests that income and education are correlated with higher engagement with the lead testing program, except in those areas where risk of lead is low. Similarly, the band of high levels of poverty stretching from the West to the far South side of the city is associated with low testing behavior. The relationship between levels of Black and African American residents and program engagement is less clear. The Black and African American population is heavily concentrated in two bands on the West and South sides of the city. The western cluster is associated with very low rates of test requests, while the southern cluster overlaps with one of the regions of highest testing.




Poverty
42.00°N
41.95°N
41.90°N
41.85°N
41.80°N
41.75°N
41.70°N
41.65°N
87.9°W
87.8°W
87.7°W
87.6°W
Poverty
60
40
20
0


English Fluency
42.00°N
41.95°N
41.90°N
41.85°N
41.80°N
41.75°N
41.70°N
41.65°N
87.9°W
87.8°W
87.7°W
87.6°W
% Population that Speaks English Less than Well
60
40
20
0


Internet Access
42.00°N
41.95°N
41.90°N
41.85°N
41.80°N
41.75°N
41.70°N
41.65°N
87.9°W
87.8°W
87.7°W
87.6°W
Internet Access
100
90
80
70
60


Black and African American
42.00°N
41.95°N
41.90°N
41.85°N
41.80°N
41.75°N
41.70°N
41.65°N
87.9°W
87.8°W
87.7°W
87.6°W
Black and African American
100
75
50
25
0

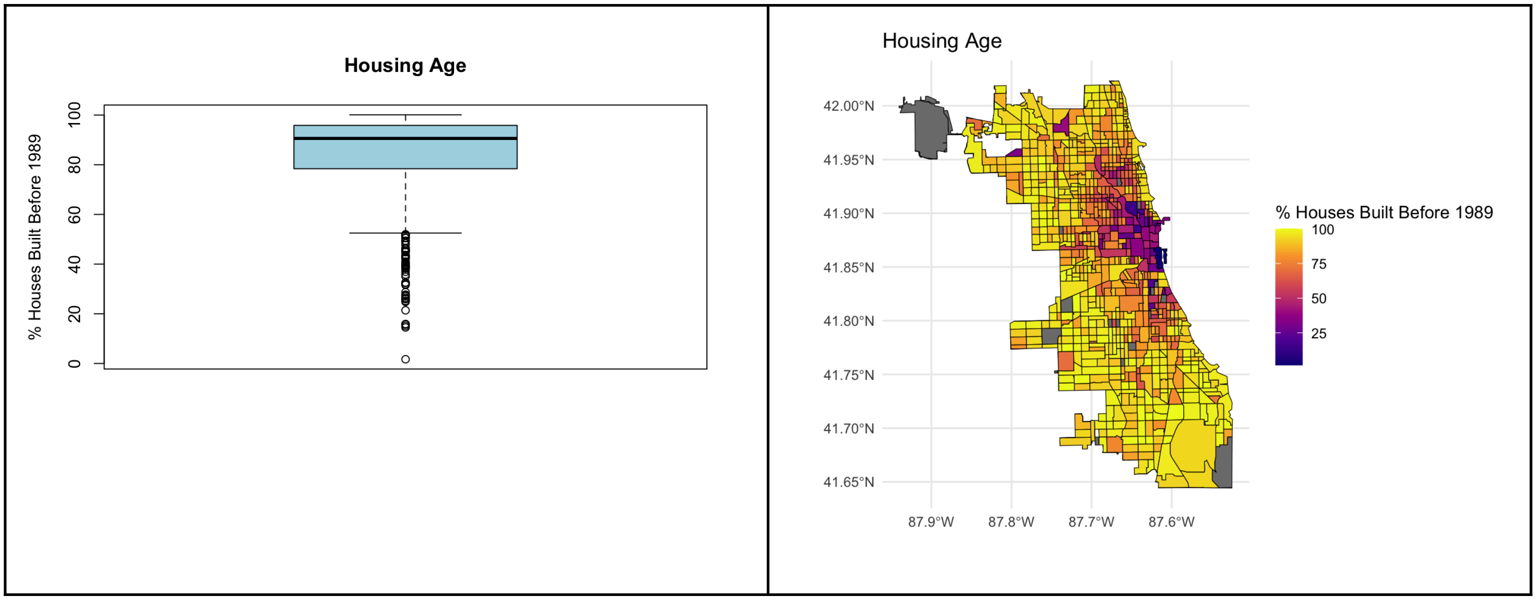


Figure 6: Heat maps of median household income (in USD), education, poverty, English fluency, internet access, Black and African American makeup (in percent), and housing age by census tract in Chicago. Education is measured by the percent of the population over 25 years old with high school diplomas, poverty by the percent of families in poverty, and internet access by the percent of households with a broadband subscription. Note: English fluency is measured by the percentage of the population that speaks English less than very well, so higher values represent less fluency. Lighter colors (yellow end of the spectrum) represents higher values. Also displayed are summary statistics for housing age in Chicago and a heat map for housing age as measured by the percent of houses in each census tract built before 1989.

Regression analysis

Regression analysis helps formalize and quantify these relationships, teasing out the nuances and dependencies. The first set of single variable regressions I conducted are of the form $y_i = \beta_0 + \beta_1 x_i + \varepsilon$, where y_i represents the predicted requests per household of census tract i, β_0 is the intercept, β_1 is the regression coefficient on a demographic variable, x_i is the value of the demographic variable for census tract i, and ε is an error term. The set of demographic variables includes median household income, percentage of families in poverty, education as measured by the percent of people with a high school education or above, percent of households with internet access, English fluency as measured by the share of the population that speaks English less than very well, and Black/African American and Hispanic populations.

Figure 7 displays the results for median household income. There is a strong positive relationship between median household income and requests per household. Over the entire range of incomes, the estimated coefficient is $3.322 * 10^{-7}$ which is statistically significant at the $p < 0.001$ level. Substantively, this means that an increase of $10,000 in the median household income of a census tract is associated on average with a 0.003322 increase in the requests per household. This is approximately equivalent to a 0.3 percentage point increase in the likelihood of a given household having requested a test. The R^2 value is 0.06, which means that, by itself, household income can explain 6% of the variation in testing behavior. Despite the high level of statistical significance, the substantive predictive power is thus modest. However, the scatter plot demonstrates a stronger relationship between the variables at incomes below $150,000. Above this threshold, requests per household declines, which may represent census tracts that are located in areas without buildings old enough to pose a lead hazard in tap water. Capping census tracts below this income and repeating the statistical analysis increases the estimated coefficient to 6.126* 10^{-7} and the R^2 to 0.11. While both analyses are statistically significant, the relationship between income and testing behavior is substantively stronger at lower incomes.

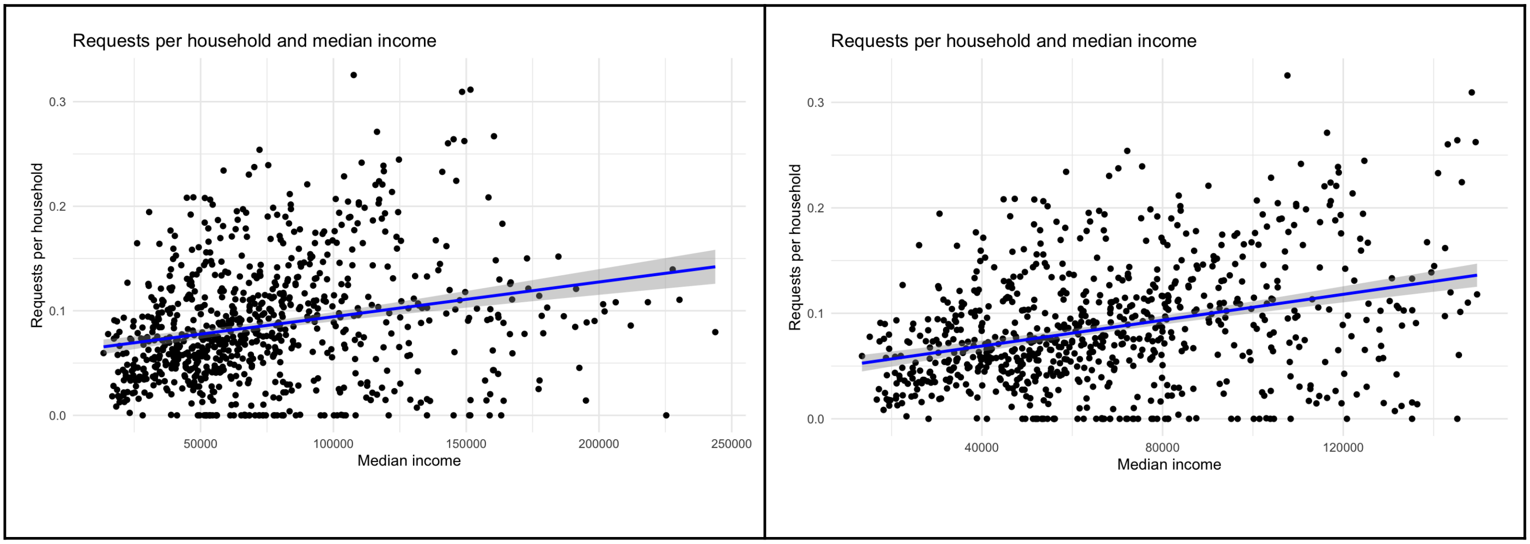


Figure 7: Scatterplots and regression lines for the number of requests per household against median household income for all income levels (left) and median income capped at $150,000 (right). The unit of analysis is the census tract.

I ran analogous regressions for the other demographic factors (Figure 8). Poverty, education, internet access, and English proficiency were all statistically significant predictors of testing behavior at the $p < 0.001$ level. The estimated coefficient for poverty is -0.0011 and the R^2 value is 0.06. This means that an increase in the proportion of families in poverty by ten percentage points is associated with a 0.011 decline in the average requests per household, or approximately a 1.1 percentage point decrease in the likelihood of a given household having requested a lead test. This variation can account for 6% of the variation in testing behavior. Education was positively correlated with requests per household with an estimated coefficient of 0.0011 and an R^2 of 0.04. Internet access was positively correlated with an estimated coefficient of 0.001 and an R^2 of 0.02. English proficiency was positively correlated with an estimated coefficient of 0.0007 and an R^2 of 0.02. Again, these variables are statistically significant but substantively do not explain much of the observed variance. Note that the regression line for English proficiency is negatively sloped because the ACS variable measuring this trait is percent of the population that speaks English less than very well.

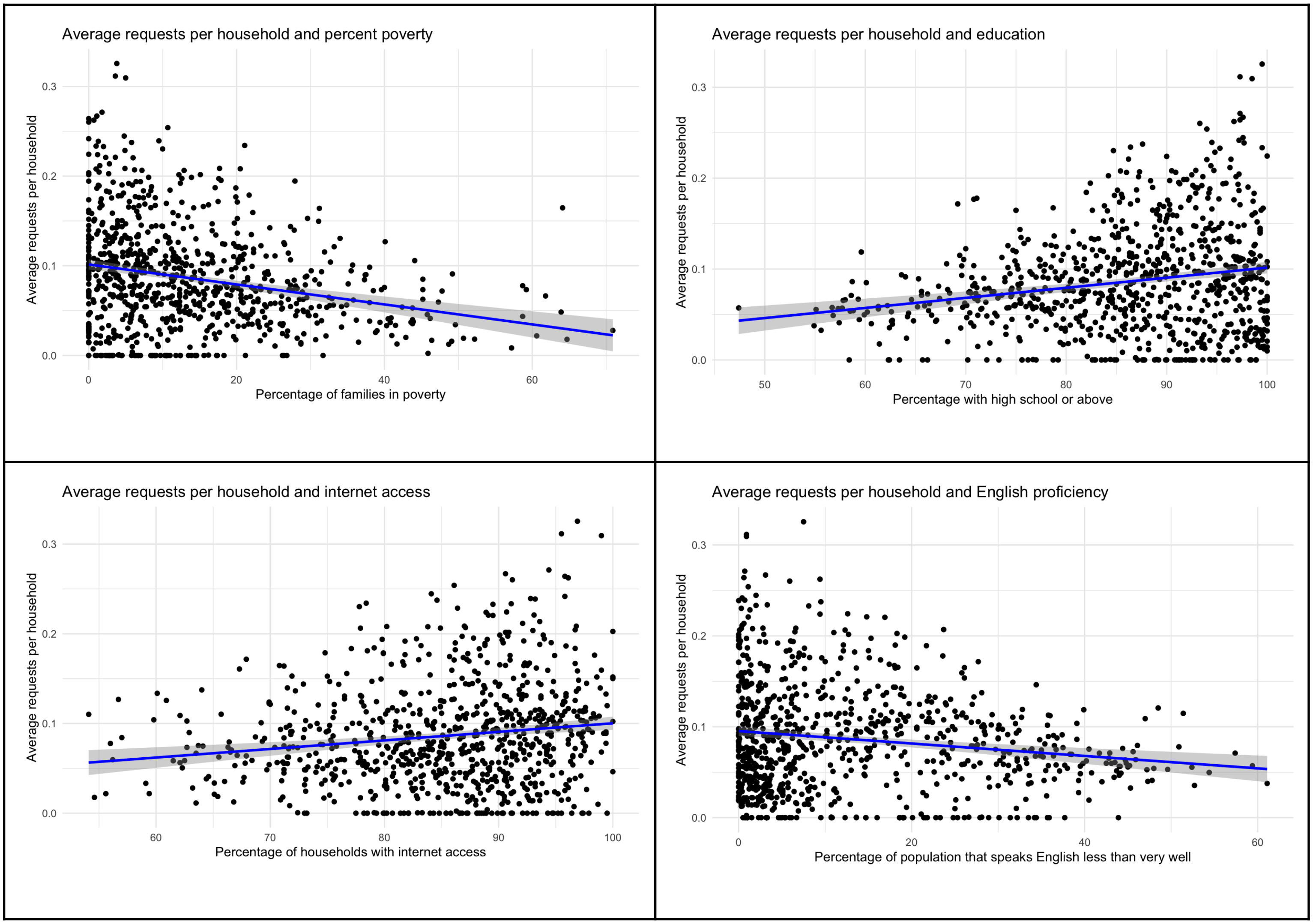


Figure 8: Scatterplots and regression lines for the number of requests per household against poverty, education, internet access, and English proficiency. The unit of analysis is the census tract.

Interestingly, while there were slight positive and negative coefficients for the Black and African American and Hispanic variables respectively, these relationships were not statistically significant at a $p < 0.01$ level. This result is unexpected due to the correlation between these variables and variables tested above, as well as prior literature.

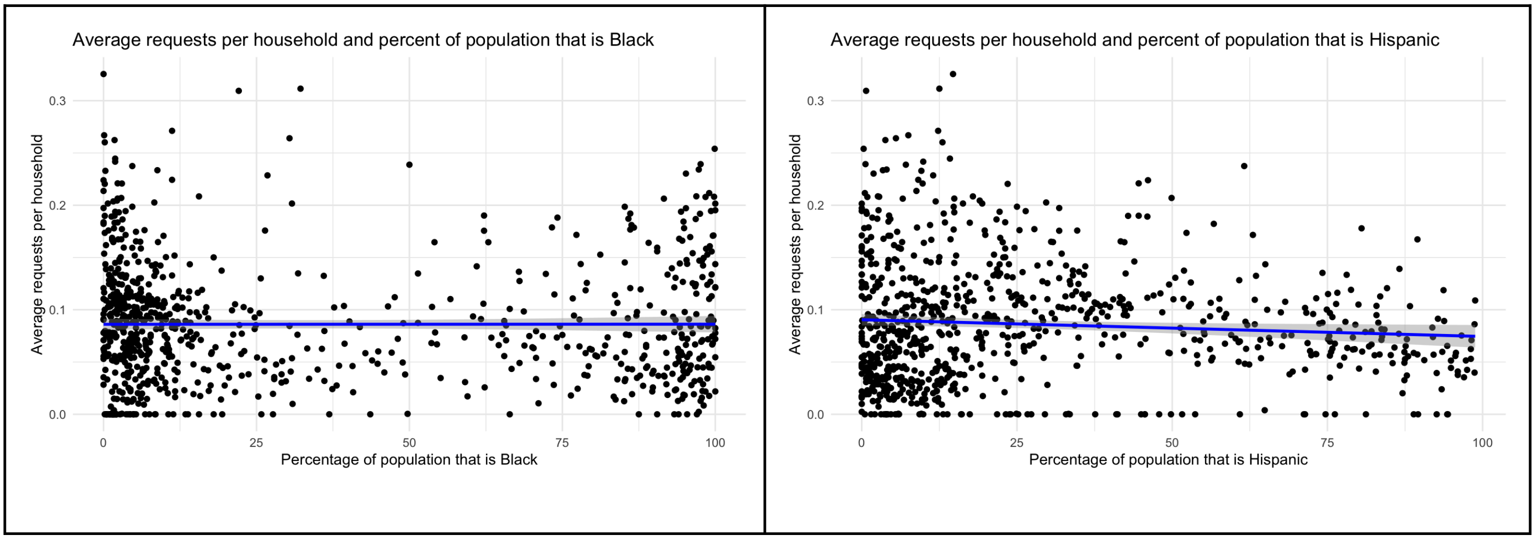


Figure 9: Scatterplots and regression lines for the number of requests per household against Black and Hispanic populations. The unit of analysis is the census tract.

Another potential determinant of who tests for lead in their tap water is objective risk. Regardless of other personal characteristics, presumably a higher possibility of lead contamination increases the likelihood of requesting a test. While lead hazard cannot be determined explicitly with the given data, housing age provides a good proxy, as explained above. As expected, there is a positive correlation between the proportion of houses in a census tract built before 1989 and the requests per household, with an estimated coefficient of 0.00086 with a significance threshold $p < 0.001$ (Figure 10). Therefore, for a ten percentage point increase in the proportion of houses built before 1989, there is an associated increase in the requests per household of approximately 0.01. The R^2 of 0.06 shows that 6% of the variation can be explained by housing age.

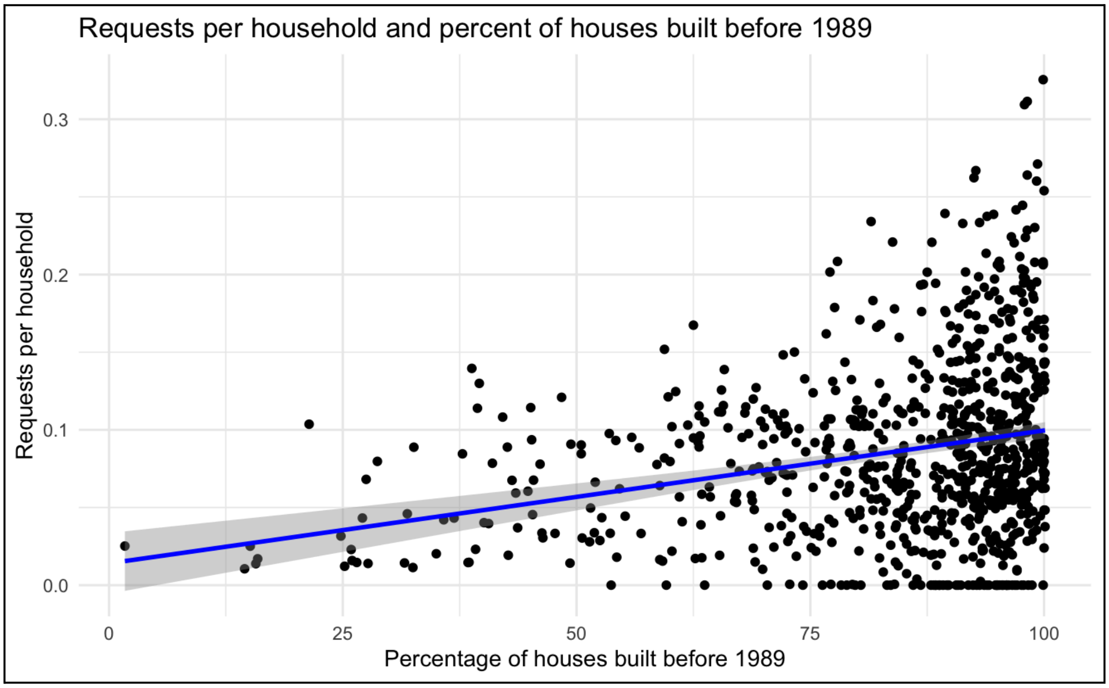


Figure 10: Scatterplot and regression line for the number of requests per household against housing age represented by the percentage of houses built before 1989. The unit of analysis is the census tract.

Explanatory variable	**Estimated coefficient**	**R^2**
Median household income	3.322×10^{-7}***	0.06
Percent of families in poverty	-0.0011***	0.06
Education (high school or above)	0.0011***	0.04
Internet access	0.001***	0.02
English proficiency	0.0007***	0.02
Black or African American	1.052×10^{-6}	0
Hispanic	-1.601×10^{-4}*	0.01

Housing age	0.00086	0.06

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 2: Regression table for the requests per household model.

Due to the possible associations between the independent variables, I proceeded to conduct a multivariate analysis of the form $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + \beta_5 x_{i,5} + \beta_6 x_{i,6} + \beta_7 x_{i,7} + \beta_8 x_{i,8} + \varepsilon$, where y_i represents the predicted requests per household of census tract i, β_0 is the intercept, β_1–β_8 are the regression coefficients for the demographic variables, $x_{i,1}$–$x_{i,8}$ are the value of the demographic variables for census tract i, and ε is an error term. The demographic variables are the same as included in the prior single variable regressions.

When the variables are all included together in the same regression, education, internet access, and English fluency drop out of significance at the $p < 0.01$ level. This change occurs because there is collinearity between these variables and the other explanatory variables such as income. For example, on average, higher levels of education increase income potential, and higher levels of income increase educational achievement. The single variable regression between request frequency and education captures both the effects of education and part of the effects of income because income and education are associated. Due to the overlap in variation patterns, the effect originally calculated between each independent variable and the dependent variable overestimated the real relationship. The multivariable regression serves to quantify the impact of each explanatory variable independently by holding the other variables constant. While it is difficult to discern whether education, internet access, and English fluency are confounders or mechanisms through which the other variables impact testing behavior, the impact can be statistically attributed better to the remaining factors.

More interestingly, by controlling for the full set of variables, Black and African American population becomes significant at the $p < 0.001$ level, and Hispanic population becomes significant at the $p < 0.01$ level. Furthermore, the estimated coefficients are positive, implying that as the proportion of these minority populations increase, the incidence of testing increases. Therefore, once variables like income, poverty, and risk are held constant, Black and Hispanic populations are positively associated with engaging with the free lead testing program. For a parsimonious model, I dropped the independent variables that did not yield significant coefficients at $p < 0.001$ and reran the multivariate regression. This final model yielded a coefficient of 6.29×10^{-7} for median household income, -7.95×10^{-4} for poverty, 4.70×10^{-4} for Black and African American makeup, and 1.38×10^{-3} for housing age (Figure 11). Together the variables can account for 24% of the variation in testing behavior (Adjusted $R^2 = 0.24$). This is a significant improvement over the R^2 values for the individual explanatory variables. While the multivariate regression still explains a minority of the variation present, R^2 values above 0.1 are generally acceptable in empirical social science research as long as the predictors are statistically significant (Ozili, 2023).

Explanatory variable	**Estimated coefficient**
Median household income	6.29×10^{-7}***
Percent of families in poverty	-7.95×10^{-4}***
Black and African American	4.70×10^{-4}***
Housing age	1.38×10^{-3}***

*** $p < 0.001$

Table 3: Reduced regression table for the requests per household model.

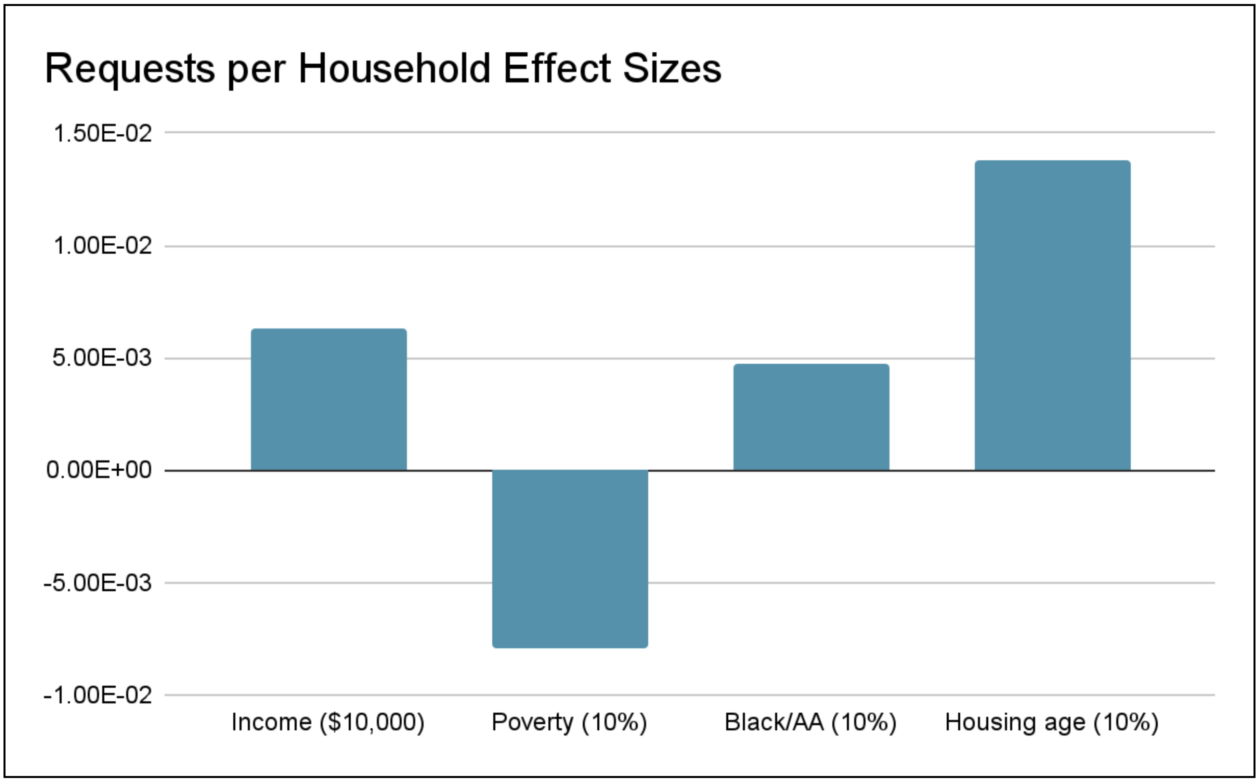


Figure 11: Effect size comparison of the independent variables in the simplified model of requests per household. The y-axis represents the change in average requests per household associated with an increase of $10,000 median household income, 10 percentage point increase of households in poverty, 10 percentage point increase of Black and African American population, and 10 percentage point increase in the proportion of houses built before 1989 respectively for each bar.

To test the assumptions of the multivariate model, I graphed the model residuals, normal Q-Q plot, and collinearity matrix (Figure 12). The residuals are nearly normal and the collinearity between independent variables is generally below the threshold of ~0.6.

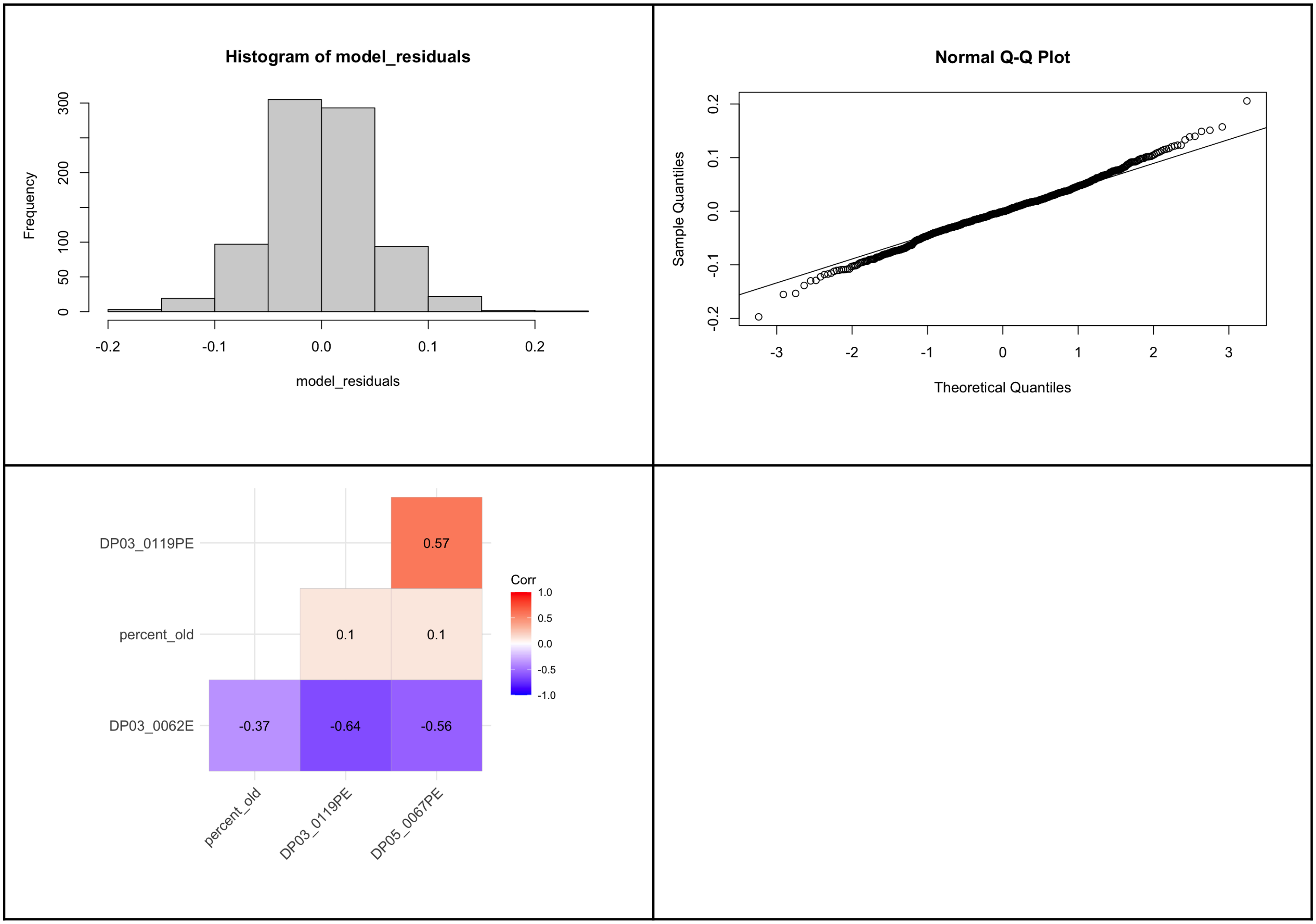


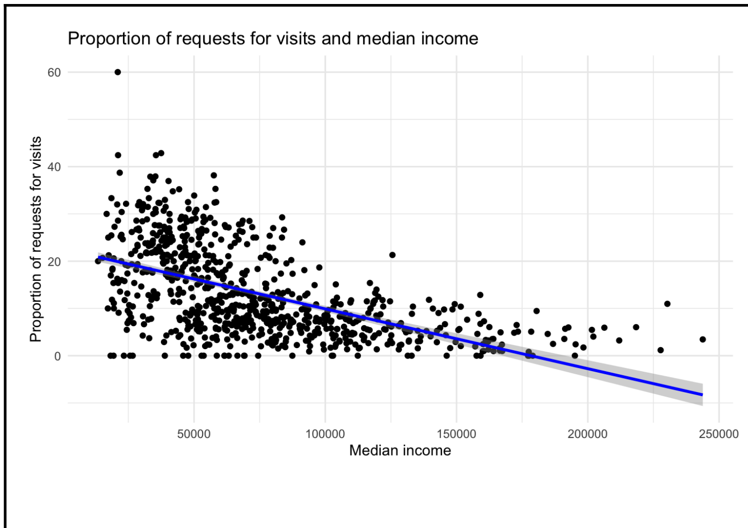
Figure 12: Histogram distribution of model residuals, normal Q-Q plot, and collinearity to matrix to check assumptions for multivariate regression.

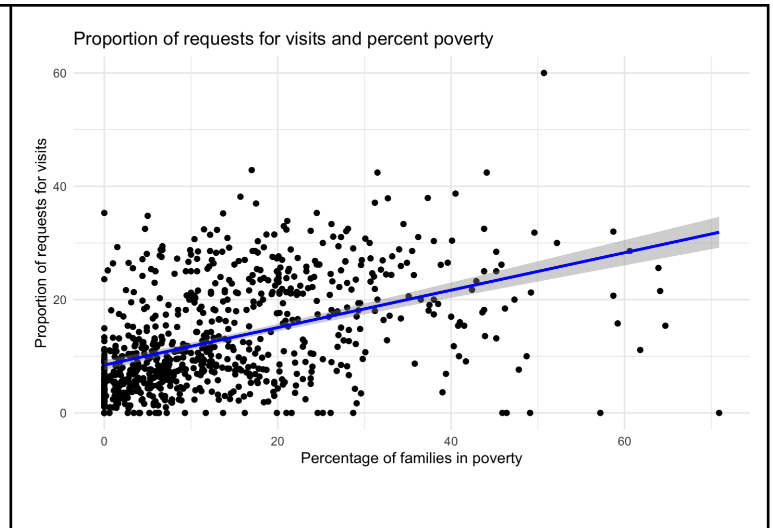
Next, I conducted a similar set of regression analyses with the proportion of requests for visits (in contrast to self-test kits) as the dependent variable. Rather than simple quantity of testing as explored in the previous section, this investigation delves into how requests are filed in order to elucidate potential barriers. Figure 13 displays the scatterplots and regression lines between proportion of requests for visits and the set of demographic variables. The following table displays the estimated coefficients, significance levels, and R^2 values for each single-variable regression.

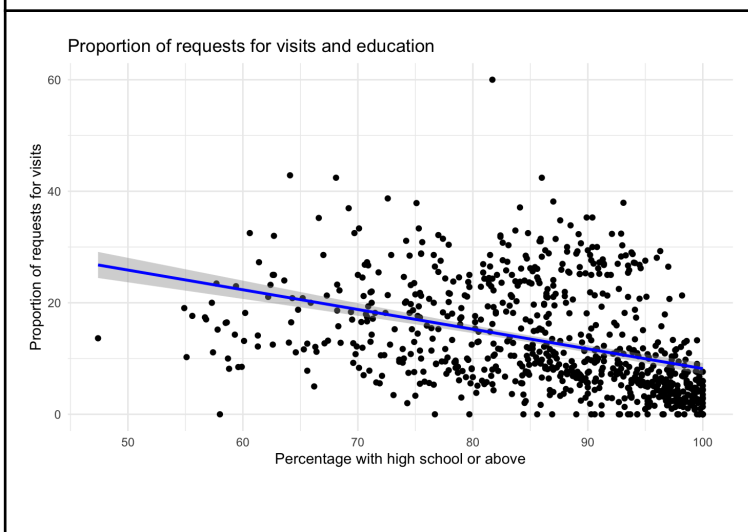
Explanatory variable	**Estimated coefficient**	**R^2**
Median household income	-1.265×10^{-4}***	0.30
Percent of families in poverty	0.3306***	0.20
Education (high school or above)	-0.3534***	0.15
Internet access	-0.4895***	0.23
English proficiency	-0.03433	0.002
Black or African American	0.1594***	0.41
Hispanic	0.0023	0

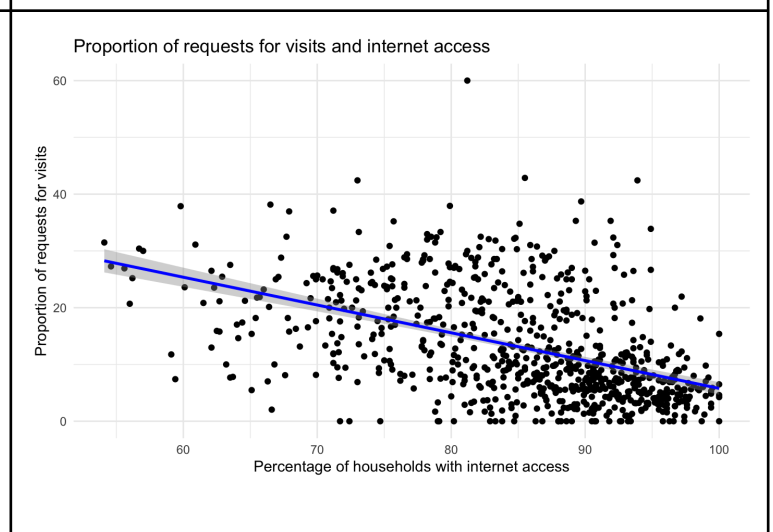
*** $p < 0.001$

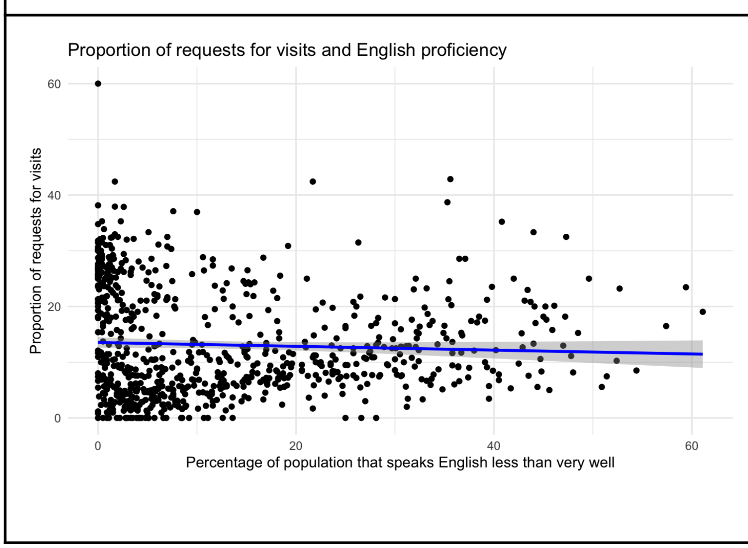
Table 4: Regression table for the proportion of requests for visits model.

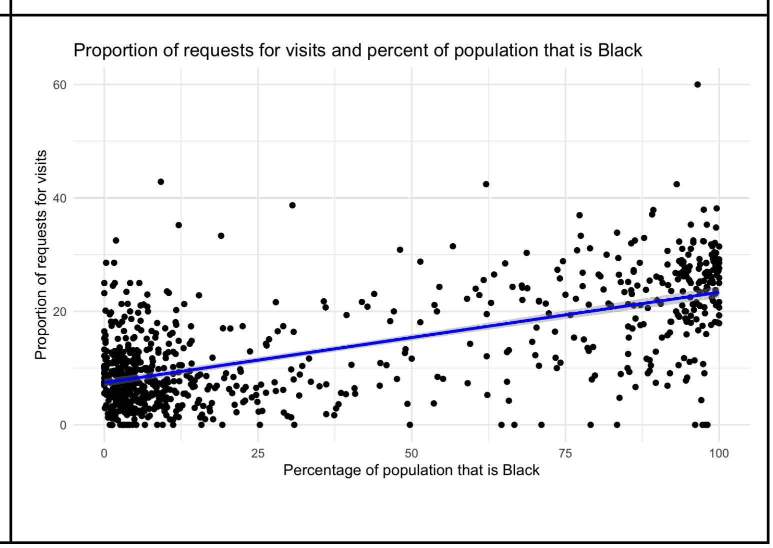

Proportion of requests for visits and median income
Proportion of requests for visits
60
40
20
0
50000
100000
150000
200000
250000
Median income


Proportion of requests for visits and percent poverty
Proportion of requests for visits
60
40
20
0
0
20
40
60
Percentage of families in poverty


Proportion of requests for visits and education
Proportion of requests for visits
60
40
20
0
50
60
70
80
90
100
Percentage with high school or above


Proportion of requests for visits and internet access
Proportion of requests for visits
60
40
20
0
60
70
80
90
100
Percentage of households with internet access


Proportion of requests for visits and English proficiency
Proportion of requests for visits
60
40
20
0
0
20
40
60
Percentage of population that speaks English less than very well


Proportion of requests for visits and percent of population that is Black
Proportion of requests for visits
60
40
20
0
0
25
50
75
100
Percentage of population that is Black

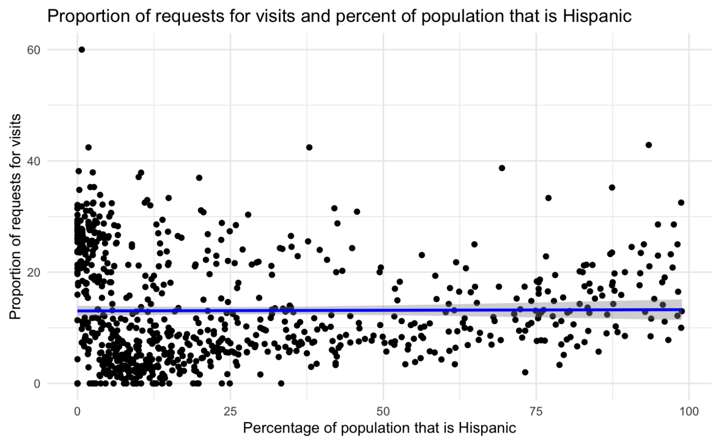


Figure 13: Scatterplots and regression lines for the proportion of requests for in-person visits against median income, poverty, education, internet access, English proficiency, Black, and Hispanic populations. The unit of analysis is the census tract.

All explanatory variables are statistically significant besides English proficiency and Hispanic population. Among the significant variables, poverty and Black population are positively correlated with the likelihood of requesting an in-person visit. To contextualize the estimated coefficient of 0.1594 on Black population, a ten percentage point increase in the Black population of a census tract is associated with a 1.594 percentage point increase in the proportion of requests for visits. Meanwhile, household income, education, and internet access are negatively correlated. Black population, income, and internet access have the highest explanatory power for determining whether an in-person or self-test kit is requested. As before, I ran a multivariate regression with all independent variables and then simplified the model by removing variables based on the significance levels. The updated statistical values from this reduced analysis are in the table below. Together, internet access, Black population, and Hispanic population can account for 57% of variation in request type (adjusted $R^2 = 0.57$).

Explanatory variable	Estimated coefficient
Internet access	-0.1074***
Black or African American	0.1988***
Hispanic	0.1349***

*** $p < 0.001$

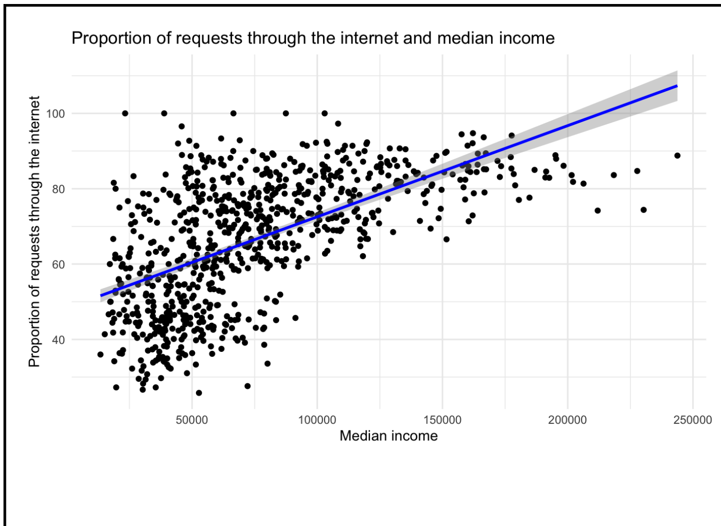
Table 5: Reduced regression table for the proportion of requests for visits model.

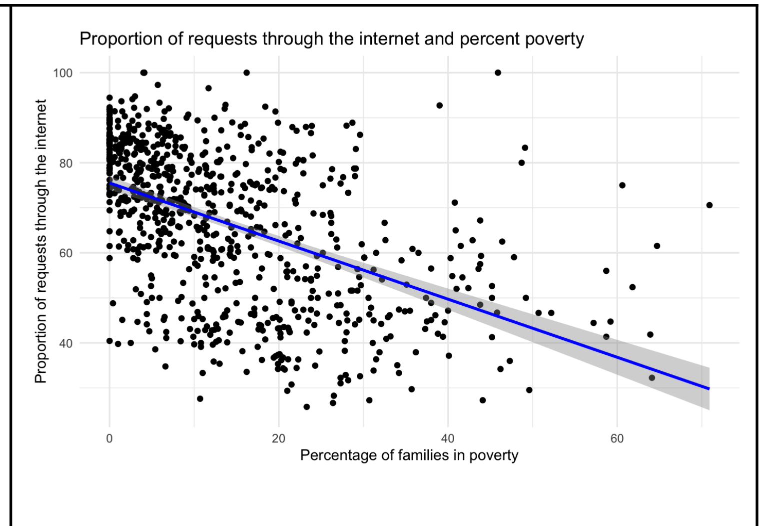
Another way to explore usage of the 311 request program is the manner in which citizens interact with the platform. While people primarily access the system through the internet, the proportion is highly variable by census tract. Figure 14 displays the relationships between the proportion of requests filed through the internet and the set of demographic variables through scatterplots and associated regression lines. The following table displays the estimated coefficients, significance levels, and R^2 values for each single variable regression.

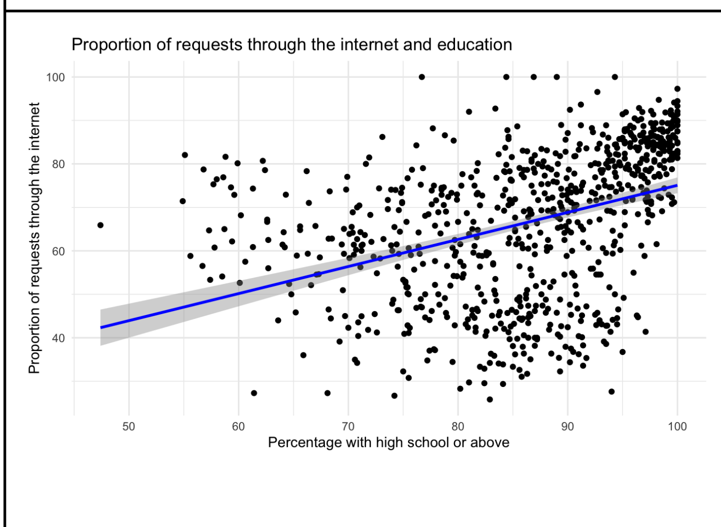
Explanatory variable	Estimated coefficient	R^2
Median household income	2.420×10^{-4}***	0.35
Percent of families in poverty	-0.6451***	0.24
Education (high school or above)	0.6227***	0.15
Internet access	0.9549***	0.27
English proficiency	0.1527***	0.01
Black or African American	-0.3202***	0.53
Hispanic	0.0225	0

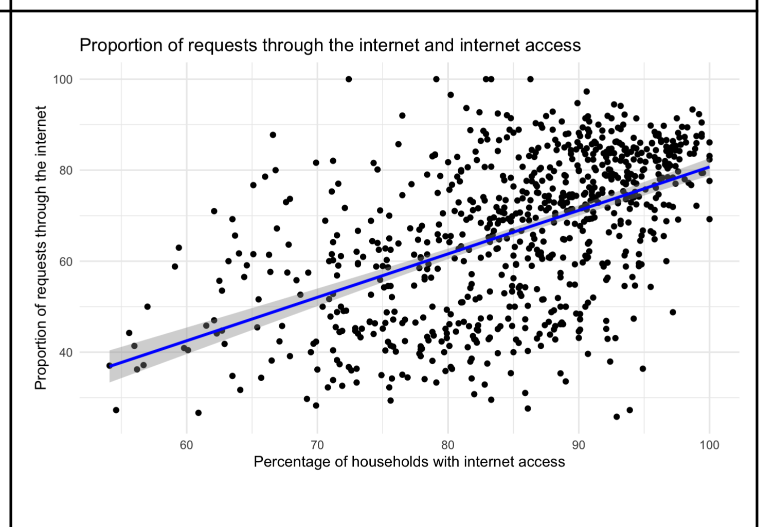
*** $p < 0.001$

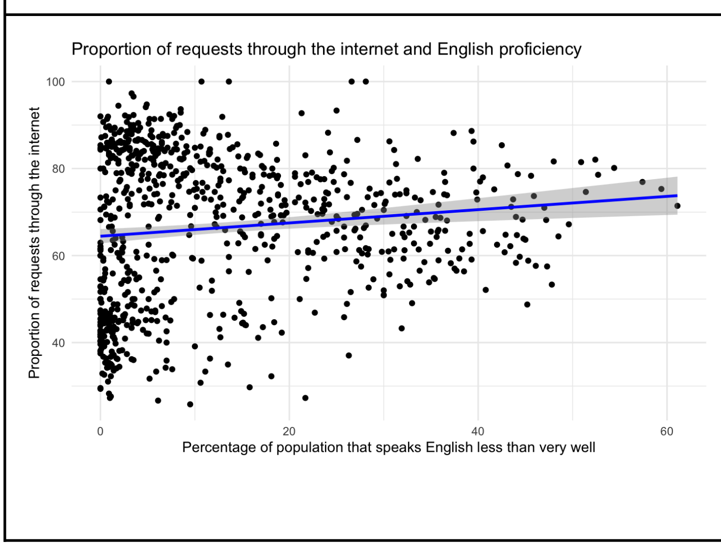
Table 6: Regression table for the proportion of requests through the internet model.

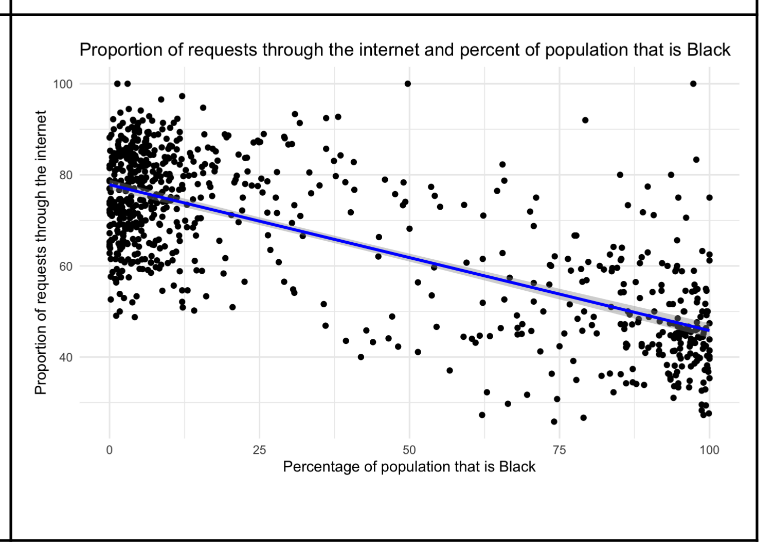

Proportion of requests through the internet and median income
Proportion of requests through the internet
100
80
60
40
50000
100000
150000
200000
250000
Median income


Proportion of requests through the internet and percent poverty
Proportion of requests through the internet
100
80
60
40
0
20
40
60
Percentage of families in poverty


Proportion of requests through the internet and education
Proportion of requests through the internet
100
80
60
40
50
60
70
80
90
100
Percentage with high school or above


Proportion of requests through the internet and internet access
Proportion of requests through the internet
100
80
60
40
60
70
80
90
100
Percentage of households with internet access


Proportion of requests through the internet and English proficiency
Proportion of requests through the internet
100
80
60
40
0
20
40
60
Percentage of population that speaks English less than very well


Proportion of requests through the internet and percent of population that is Black
Proportion of requests through the internet
100
80
60
40
0
25
50
75
100
Percentage of population that is Black

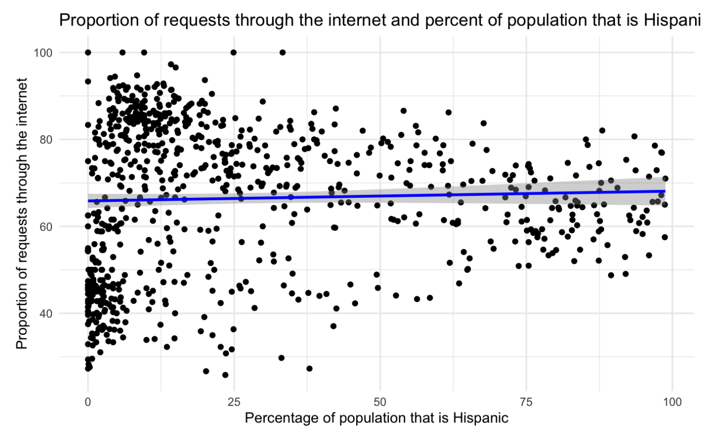


Figure 14: Scatterplots and regression lines for the proportion of requests through the internet against median income, poverty, education, internet access, English proficiency, Black, and Hispanic populations. The unit of analysis is the census tract.

Besides Hispanic population, all explanatory variables are statistically significant. Increases in income, education, and internet access correspond to increases in the likelihood of filing a request through the internet. In contrast, high poverty rates, English proficiency, and Black population are associated with increased likelihood of requesting a test through a non-internet method. Black population, median household income, and internet access can explain the highest amounts of variation in request origin. Again, I combined the explanatory variables into a single regression, followed by removing insignificant variables for parsimony. The table below displays the statistical values from the simplified model. Together, internet access, Black population, and Hispanic population variables can account for 68% of variation in request type (adjusted $R^2 = 0.68$).

Explanatory variable	Estimated coefficient
Internet access	0.2291***
Black or African American	-0.3838***
Hispanic	-0.2324***

*** $p < 0.001$

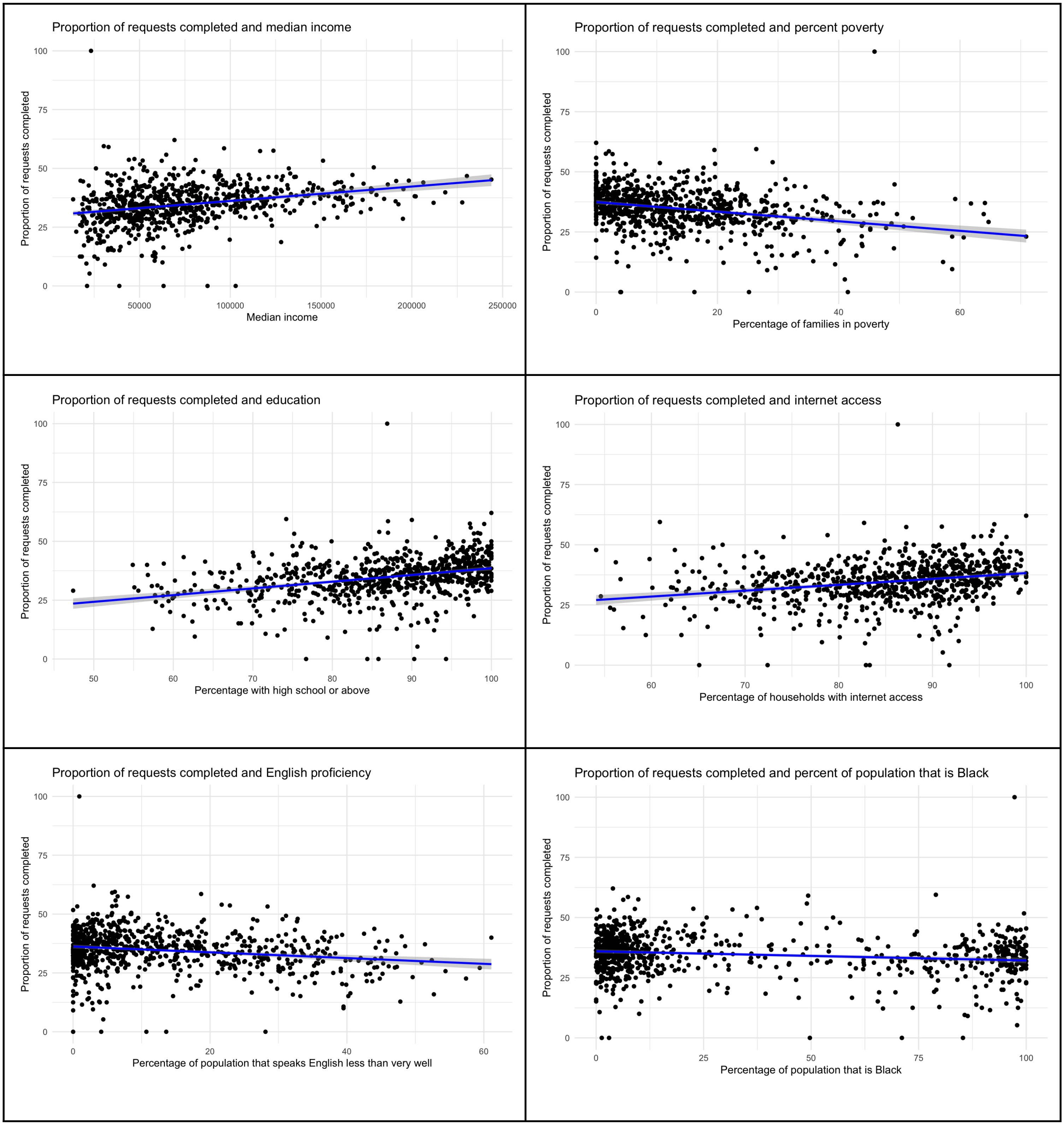
Table 7: Reduced regression table for the proportion of requests through the internet model.

Finally, in addition to disparities in the rates of requesting tests, the proportion of requested tests that are successfully completed is variable and correlates with the demographic features of the census tracts. Summary statistics from single variable regressions are available in the following table, and corresponding scatterplots and regression lines are in Figure 15.

Explanatory variable	Estimated coefficient	R^2
Median household income	6.133×10^{-5}***	0.08
Percent of families in poverty	-0.1991***	0.08
Education (high school or above)	0.2853***	0.12
Internet access	0.2441***	0.07
English proficiency	-0.1222***	0.03
Black or African American	-0.0381***	0.03
Hispanic	-0.0663***	0.04

*** $p < 0.001$

Table 8: Regression table for the completed requests model.


Proportion of requests completed and median income
Proportion of requests completed
Median income
0
25
50
75
100
50000
100000
150000
200000
250000
Proportion of requests completed and percent poverty
Proportion of requests completed
Percentage of families in poverty
0
20
40
60
Proportion of requests completed and education
Proportion of requests completed
Percentage with high school or above
50
60
70
80
90
100
Proportion of requests completed and internet access
Proportion of requests completed
Percentage of households with internet access
60
70
80
90
100
Proportion of requests completed and English proficiency
Proportion of requests completed
Percentage of population that speaks English less than very well
0
20
40
60
Proportion of requests completed and percent of population that is Black
Proportion of requests completed
Percentage of population that is Black
0
25
50
75
100

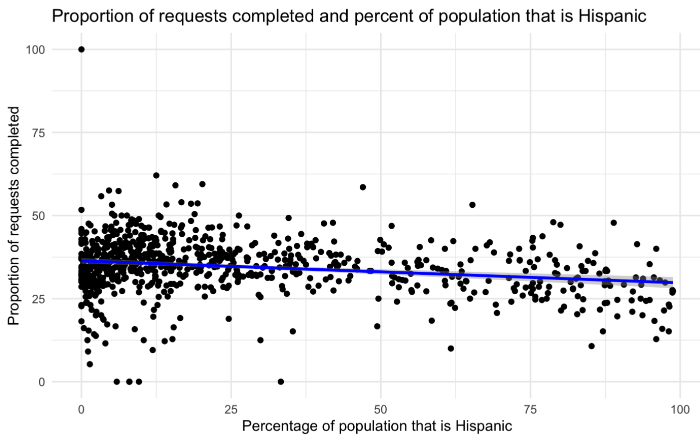


Figure 15: Scatterplots and regression lines for the proportion of requests completed against median income, poverty, education, internet access, English proficiency, Black, and Hispanic populations. The unit of analysis is the census tract.

All explanatory variables are statistically significant, though the substantive magnitude of the correlative relationship is generally small. Income, education, English proficiency, and internet access are positively correlated with successful completion of a test once a request is filed, while poverty and Black and Hispanic populations are negatively correlated. None of the variables individually can explain more than 12% of the variance in completion. The results from the simplified multivariate regression are shown in the following table (adjusted $R^2 = 0.16$).

Explanatory variable	**Estimated coefficient**
Percent of families in poverty	-0.1168***
Black or African American	-0.0553***
Hispanic	-0.1075***

*** $p < 0.001$

Table 9: Reduced regression table for the completed requests model.

Time progression

As program implementation continues over time, the number of tests requested does not uniformly increase nor decrease (Figure 16). Instead, testing rates generally stay low but are punctuated periodically with spikes of higher activity. Two peaks in the first year of implementation persist over significant segments of time, but most of these peaks are very short term, consisting sometimes of a single day. The largest spike occurred in July of 2019: 1,632 requests were filed on July 9th, 2,027 on July 10th, and 885 on July 11th. These are the three largest request volumes for a single day over the entire time of the program.

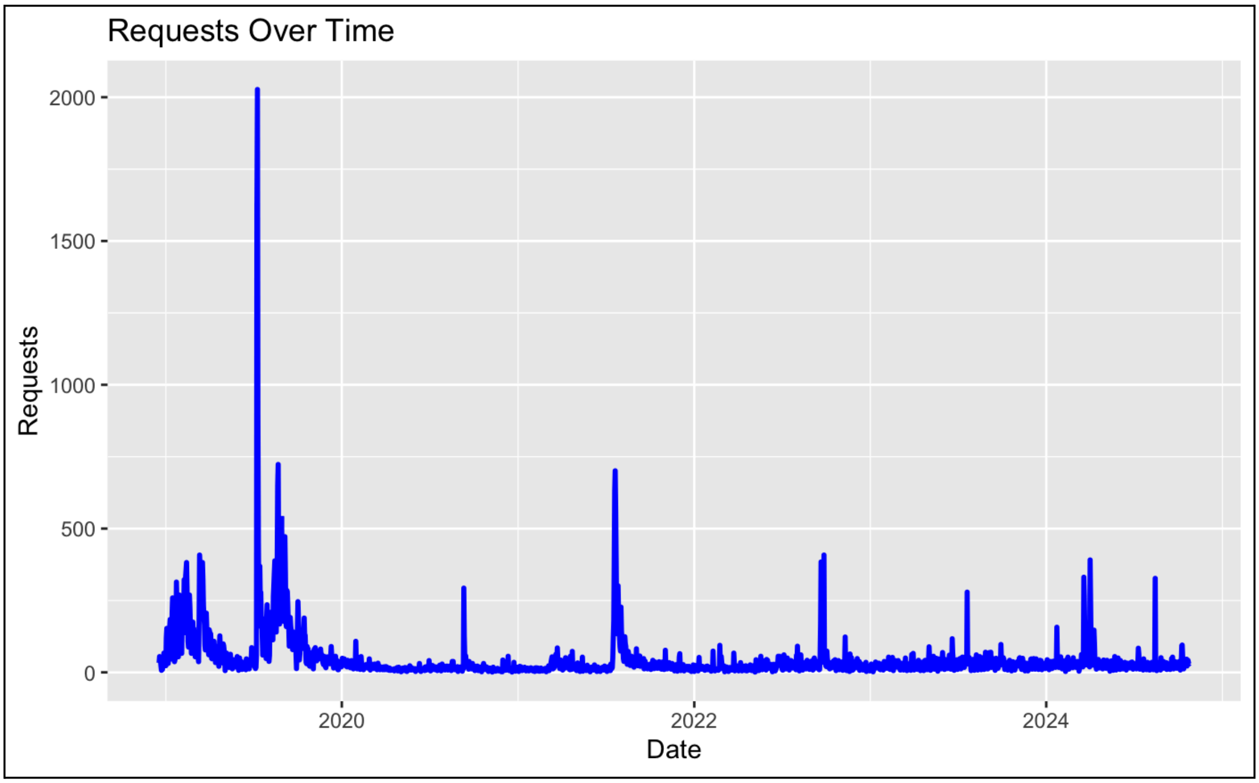


Figure 16: Total number of requests for lead test kits or visits through the 311 program over time (2018-2024).

In addition to volume of tests, the regression coefficients relating demographic variables to the number of requests per household change over time (Figure 17). This suggests that the strength of the relationship between the potential explanatory variables and frequency of testing is not constant. For all independent variables that had statistically significant coefficients in the multivariate regression, the highest magnitude coefficient occurs in 2019, which also marked the year with the most requests total. In general, the coefficient magnitudes were also higher in 2021 and 2023 when compared to 2020, 2022, and 2024, leading to an oscillating biannual pattern.

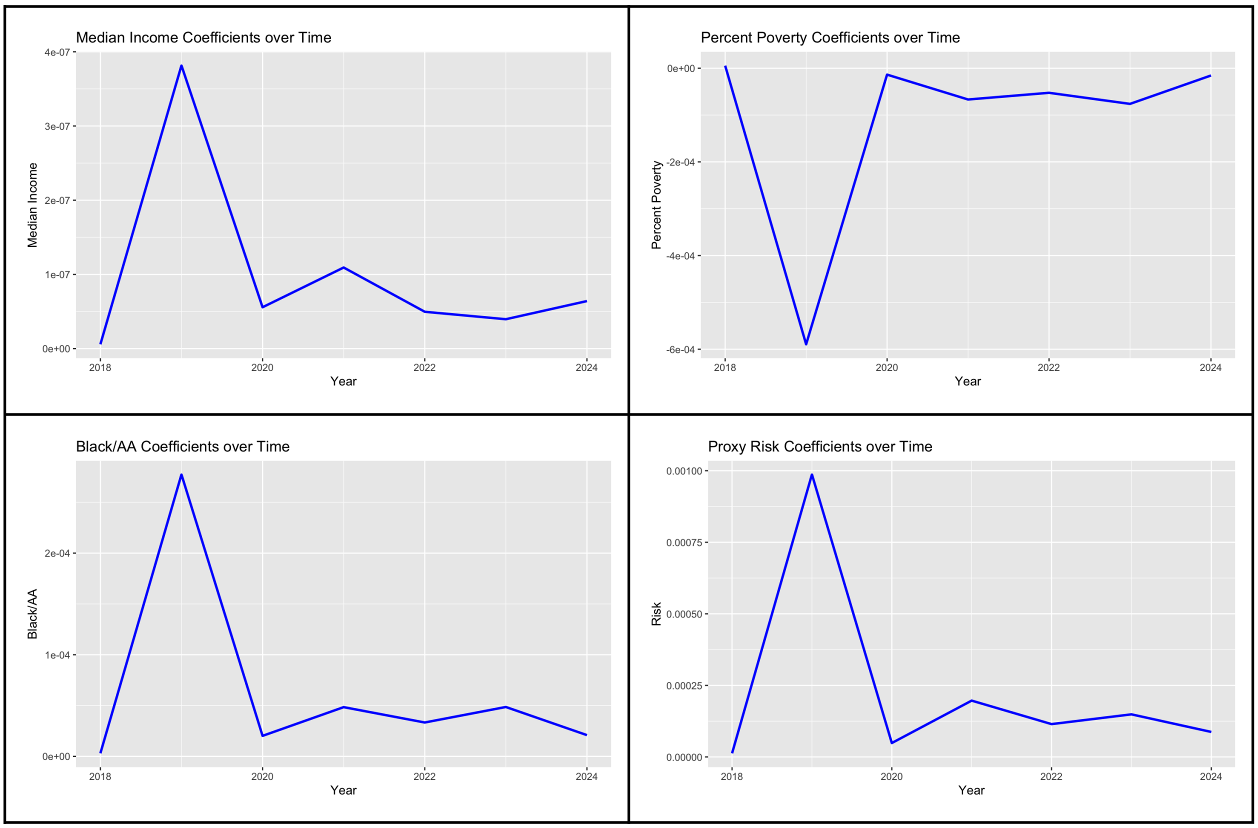


Figure 17: Coefficients from the multivariable linear regression relating requests per household to demographic variables (median income, households in poverty, Black and African American population, and proxy risk measured by housing age) over time (2018-2024).

Analysis

Analysis reveals that significant barriers exist throughout the multi-stage process of resident based tap water lead contamination testing. While low rates of requests per household indicate that residents experience burden at the initial step of engagement, the low rates of completion among those that do request tests suggests another layer of obstacles. Together, these two features are indicative of the administrative burden framework at play (Herd and Moynihan, 2019). As residents must become aware of their lead risk, the existence of the 311 program, and how to participate, they experience *learning costs*. These costs are reflected in the request volume. Similarly, as residents navigate the bureaucratic requirements to request a test, leave their water stagnant, follow the instructions for water sample collection, coordinate a pick-up time, and leave the completed test for collection, they experience *compliance costs*, which are reproduced in the completion rates.

Not only do both of these forms of cost appear to be strongly prohibitive, but they are also not equally borne by all Chicagoans. The evidence supports a modified needs-based hypothesis: engagement (as measured by frequency of tests) increases with need (proxy lead risk measured by the age of the housing stock), but is conditioned by other forces based on burden. In particular, income, poverty, and Black and African American population provide the best predictors for the number of requests made per household, while poverty, Black and African American population, and Hispanic population provide the best predictors for completion rates.

Poverty and income perform in the expected manner as measures of socioeconomic status and resource access; declines in poverty and increases in income are associated with both an increased likelihood of requesting a test at all and a greater chance of successfully completing the test once received. This result provides evidence for the distributive effects of administrative burden. Inequalities in participation are also organized spatially across the city along historical, social, cultural, and economic lines ingrained into the built environment. However, the impact of race is more complex. While increases in both minority populations decrease the proportion of completed requests, Black and African American population is positively correlated with request volume once controlling for other forms of socioeconomic status and objective risk of lead. This result contradicts prior research that suggests political participation (including demand expression through 311 systems) is negatively associated with minority status (Cook et al., 2024; Grumbach & Sahn, 2020; Kontokosta et al., 2017; Minkoff, 2015; Morris & Grange, 2024), though there is some evidence for higher participation among Black citizens due to group consciousness during the Civil Rights Movement (Verba & Nie, 1987). More specifically, my data provides opposing evidence to the finding from Huynh et al. (2024) that Black and Hispanic populations are negatively associated with a census block's odds of being tested for lead.

I have three hypotheses for the observed positive relationship between Black/African American population and request number. The first is an elevated level of distrust of water delivery service. According to the Healthy Chicago Survey 2021-2022 conducted by the Chicago Department of Public Health, Black and Hispanic households are less likely than white households to use tap water as their primary drinking water source, instead tending to rely on bottled water (Huynh et al., 2024). This behavior likely indicates, in part, a belief that tap water is not safe (Gorelick et al., 2011). Thus, minority populations may be more likely to request lead tests, even once objective risk is controlled for, because they disproportionately fear contamination in their water. Rather than decrease government interaction as I predicted, distrust of publicly provided tap water could increase participation in the testing program.

My second hypothesis is that race may be acting through a mechanism of gender or family structure to affect testing rates. Black families are over twice as likely to be female-headed as white families due to social factors like mass incarceration (Craigie et al., 2018), and women have higher levels of health awareness (Bidmon & Terlutter, 2015) and are more proactive about their health (Rudoy & Leis, 2018). In a preliminary examination of the relationship between gender and testing behavior, I found that both female-headed and male-headed households were negatively associated with requesting lead tests (versus cohabiting or married households), but that this negative relationship was stronger for male-headed households (Appendix B). Therefore, a high Black and African American population in a census tract may increase testing behavior because there are more female-headed households and these households are more likely to request a test than male-headed households. Additionally, Black and Hispanic women tend to have kids younger (Schweizer & Guzzo, 2020) and have higher general fertility (PRB, 2003) than white women. Since lead poses the most danger to children, this demographic pattern could also lead to higher testing rates among Black and African American households because of heightened concern about child development.

The third hypothesis is that there are social or cultural networks in census tracts with high minority populations that advocate and help provide the resources for requesting a lead test. Higher levels of social organization stimulate governmental engagement (Minkoff, 2015), and could potentially overcome barrier-based determinants of program participation. Future research should further pursue these hypotheses to help explain the unusual pattern of involvement observed here. In any case, my results suggest that the relationship Huynh et al. (2024) found between race and screening likelihood is probably due to either confounding by other socioeconomic variables or that the positive relationship between minority population and testing requests is outweighed by the negative relationship between Black and Hispanic population and successfully completing a test once received.

Overall, the 311 dataset focusing on lead testing requests provides a valuable, need-specific data source for examining the dynamics of systematic inequality in the demand, and therefore provision, of city services. Furthermore, the focus on lead testing provides a subject specific investigation, which is crucial because the salience of different types of governmental service/resource provision is variable for different groups (Thomas, 1982). When looking just at median household income as a predictor of resource-based engagement, the results mimic the parabolic socioeconomic model derived by Jones et al. (1977). The highest levels of requests per household occur in the middle of the income range, where both need and resources are presumed to be moderately high. However, a more nuanced analysis reveals an interplay between objective risk (i.e., need), socioeconomic status, and race. Program participation increases with need, resources, and minority population. Moreover, burden appears to operate differently at the learning cost versus compliance cost stage, where minority population switches from a positive to a negative indicator of participation. The research presented here suggests that even when disadvantaged populations are able to overcome initial barriers to program engagement, administrative burden in the compliance stages prevents the successful reaping of government benefits.

Despite this discouraging result, I find evidence that citizens are adapting to the barriers they face in the lead testing program. There is high variability in the proportion or requests for visits from a professional (in contrast to self-test kits) between census tracts, and this proportion is sharply related to socioeconomic and demographic factors. The likelihood of requesting a visit increases with poverty and the population of racial minorities, while it decreases with income, education, and internet access. Thus, it appears that when households have less resources, they prefer to have their tap water tested by a city official rather than conduct the test themselves. There are several possible explanations for this behavior. It could indicate awareness of compliance barriers, a lack of confidence in successfully completing testing, or time barriers. Given the especially strong influence of internet access in predicting the type of request made, this pattern may also be a result of features of the request system itself. If people with lower socioeconomic status preferentially interact with the program through alternative routes to the internet (i.e., phone call, direct contact with municipal department, alderperson), they may be encouraged to request a visit. Regardless of the cause, this behavioral adaptation is successful; visits are much more likely to resolve in a completed request than self-test kits are, which suggests they allow residents to bypass compliance costs.[9] The substantive effect sizes and explanatory power are also considerably higher here than for overall request volume or completion rates, so choice between visits and kits may be offsetting the visibility of administrative burden.

Another way citizen behavior diverges in use of the 311 system is through request origin, or how people access the program. Around two thirds of requests come through the internet, but this proportion is highly variable by census tract. According to existing literature, providing government services online can reduce barriers to access (Das and Das, 2021), but there is a digital divide in who interacts with the government. Higher levels of income and education dramatically increase the likelihood of online participation (Pew, 2010). Indeed, the relationship between requesting tests through the internet and the socioeconomic and demographic explanatory variables reinforces the existence of a digital divide. Nearly 70% of the variation in

[9] The groups that request kits and visits are not comparable. However, the evidence suggests that the groups that are more likely to request visits would actually be less likely to successfully complete the test, so this effect should be, if anything, underestimated.

internet usage can be explained by just three variables: internet access (positive relationship), Black and African American population (negative relationship), and Hispanic population (negative relationship). The strong impact of internet access is intuitive and emphasizes the importance of getting people online for equal opportunity in governance, but Black and Hispanic populations are negatively associated with using the internet even once controlling for access. Exploring potential cultural or social explanations is essential for designing political participation to be obtainable for all.

Finally, it is worth exploring the change in request volume and the strength of the regression relationships over time. As mentioned above, the number of requests stays low over time with discrete peaks, the strongest occurring in July, 2019. The drastic nature of this elevated period of requests suggests an exogenous event that stimulated knowledge or awareness of the program. For example, a neighborhood campaign promoting lead testing or a news report detailing lead contamination from tap water could provide a sharp, time-constrained increase in testing. One candidate is the ongoing water crisis that was occurring in Flint, Michigan, though I found no immediate connection to the specific time period between July 9th and 11th. Additionally, the test requests issued during this period are not dominated by one ward or region of the city (Figure 18), suggesting that it was not the targeted actions of one alderperson that caused the marked increase of request volume. More research is needed to connect the July 2019 peak in requests for lead tests, and to a lesser extent later spikes, to specific historical events.

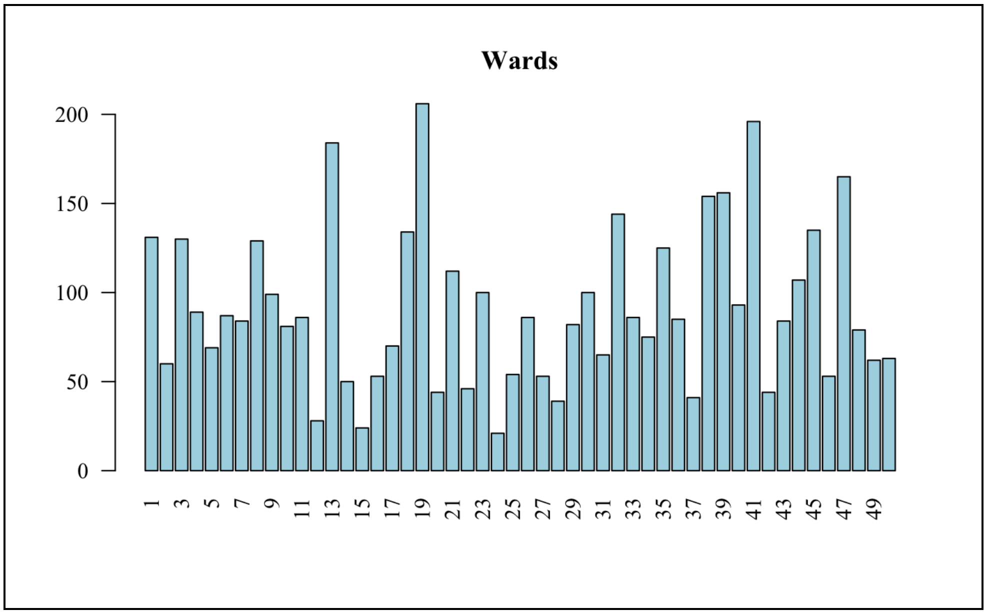


Figure 18: Distribution of requests for lead test kits and visits by ward between July 9th and July 11th, 2019.

The time series of regression coefficients reveals that the explanatory factors examined in this analysis are strongest in the first full year of program implementation. This builds on previous research that finds inequities in resident-expressed demand disappear over time (Clark and Brudney, 2018). Increasing rates of participation among historically disadvantaged groups suggests that the greatest danger of unequal demand occurs early on in rollout, when levels of awareness are lowest. The later oscillating pattern observed in the magnitude of regression coefficients could potentially be reflective of electoral cycles, but there is not enough information in the current analysis to pursue this theory, and the substantive changes are very small.

Limitations

There are several limitations in this study. First, while cleaning the city data (311 request database), I found several clear instances of technical error in request filing (see Methods section). I consciously removed these faulty entries, however their existence raises concerns about the validity of the database itself. Additionally, as with previous studies in this field (e.g., Clark et al., 2013; Clark and Brudney, 2018; Cook et al., 2024; Minkoff, 2015), I relied on aggregated census data to determine demographic and socioeconomic characteristics because these attributes were not available at the household level. Household variation that is not captured by the community level data could be introducing noise, and I may be missing barriers that occur at a household level. Though census data erases the ability to connect the explanatory variables to individual requests, it succeeds in capturing the macroscopic spatial and community based dynamics. Finally, quantitative data is limited in explaining complex behavior. In the case of lead testing in Chicago, additional variables in the dataset capturing features of origin, type, and status provide especially rich insight into behavioral patterns. Nevertheless, future research should pursue a qualitative branch that explores potential explanations for the data trends and includes interviews with users/non-users to determine resident experiences with barriers.

Policy Implications

Chicago's current approach to the city's extensive lead pipe infrastructure and high risk of tap water contamination relies on active citizen participation. While the city provides financial and technical support in the form of free lead tests, filters, and service line replacements, residents are expected to initiate contact with the government and carry out multi-step processes rife with both learning and compliance costs in order to access resources. This program constitutes a form of "coproduction," where both the government and citizens contribute to public service provision (Brudney & England, 1983). Coproduction is controversial. Researchers and policy experts often laud coproduction for fostering good citizenship, encouraging active participation in government, improving governance, and strengthening the relationship between citizen and government (Brudney & England, 1983; Clark et al., 2013; Levine, 1984). In contrast, other authors criticize reliance on coproduction as an abdication of government responsibility in response to fiscal or capacity restraints that perpetuates inequalities (Brudney & England, 1983; Hwang et al., 2024).

By combining administrative burden and citizen initiated contact theory, I propose that voluntary coproduction programs like Chicago's lead testing initiative inherently produce barriers that inequitably restrict access to essential services. The evidence I have presented of low uptake rates and the association between resources and successful participation supports this assertion. Herd & Moynihan (2019) argue that to eliminate the harmful effects of administrative burden, we should increase government capacity and shift burden from the individual to government. Coproduction does precisely the opposite, relieving the public system of cost by requiring more of the private individual (Brudney & England, 1983). I suggest that this tradeoff only has the potential to be beneficial when citizens have more insight than the government into a particular issue or its solution. In the case of lead testing and remediation, this situation is unlikely to be true.

Access to lead testing and exposure knowledge is critical not just because it empowers residents to take preventative measures like using filters or buying bottled water, but also because it provides the basis for city-sponsored lead service line replacement programs. The replacement programs introduce an additional level of learning, compliance, and psychological costs as they are structured as other voluntary, citizen–initiated processes with the added barriers posed by an application and qualification process. Based on the evidence presented here, there are likely high levels of bias in which service lines are replaced, despite a superficially equal system and programs geared directly towards those with lower socioeconomic resources (i.e., the Equity Lead Service Line Replacement program).

The dynamics of administrative burden are especially disturbing considering the 2024 Lead and Copper Rule Improvements that mandate an accelerated timeline for total lead service line removal (two decades for Chicago). As the City prepares its plan to complete this monumental infrastructural feat, it will need to identify where lead service lines are, prioritize the order in which to remove them, and determine the logistics of removal and funding. If Chicago continues to pursue these objectives through their established reliance on voluntary, citizen-led coproduction, it is destined to exacerbate the lead contamination disparities plaguing the city today.

Therefore, I recommend that the City of Chicago implements a universal, mandatory testing and replacement program in pursuit of the federal replacement deadline in 2047. This move would not be unprecedented. In fact, it is already mandatory in Chicago to report a broken or leaking service line to 311, after which the City coordinates replacement of the entire service line for free (Lead-Safe Chicago, n.d.). Other municipalities have applied this principle more generally. Between 2019 and 2022, Newark, New Jersey replaced all 23,000 operating lead service lines (Cunningham, n.d.). This remarkable success followed the passage of a city ordinance that declared simply "it is hereby established that the existence of lead service lines is prohibited in the City of Newark" and mandated building owners replace the plumbing or provide certification that their service lines were lead–free (City of Newark Code § 16:23).

To enable and enforce the requirement, the edict set a 90 day deadline with available extensions and created a Lead Service Line Replacement Program that helps owners coordinate with contractors and funds the renovation (City of Newark Code § 16:23). If the owner did not sign up for the program nor pursue replacement on their own, the ordinance detailed how city officials and contractors should secure right of entry to the property, including legal action to overcome access denial (City of Newark Code § 16:23). In the seven months preceding this ordinance, Newark replaced 650 service lines under a voluntary and only partially subsidized framework similar to Chicago (Cunningham, n.d.). After the mandate passed, the replacement rate accelerated to 80-100 service lines per day, and all 23,000 pipes were replaced in under three years (Cunningham, n.d.). In comparison, by July 2024, Chicago had only replaced 6,820 service lines total (Anastasakos, 2024).[10]

A mandatory lead replacement program would face significant challenges in Chicago: Chicago has nearly nine times the population of Newark and lacks a definitive inventory of the quantity and location of its service lines (Largest Cities, n.d.). The Department of Water Management has concerns about funding, labor availability, and logistics (Moilanen, 2024). In addition, the DWM worries that residents are reluctant to participate in replacement (Moilanen, 2024). A compulsory edict may garner resentment and resistance from residents that do not want to undergo the hassle of coordination and construction. The city would also need to be careful how it enforces compliance without penalizing people that struggle to meet the requirements, as these are likely the same people left behind by the current voluntary based programs.

However, there are a series of steps and considerations that the City can pursue to ensure equitable and efficient remediation. Prior to replacement, Chicago should enact a mandatory service line material identification and water quality testing program. This program could combine the current 311 system with principles of the Newark ordinance. Chicago would continue to provide the option of either free lead test kits or expert visits, but with an established

[10] It is possible that additional service lines were replaced by homeowners without city involvement.

mandatory compliance deadline. Past this date, city officials would attempt to contact residents, overcoming the bias involved in the citizen-initiated contact stage, and eventually follow the procedure to attain right of entry access. In addition to water testing, residents would also need to provide a self-assessed[11] or certified inspection of the material of the service line. Unless the available evidence allows for confident rejection of a lead provenance, all buildings built before 1986 should be assumed to have lead. The combination of these two branches of investigation will empower the city to create a geospatial catalog of lead service lines and prioritize removal where contamination levels are currently the most elevated.

Throughout both the preparatory and replacement phases, the city must universally cover homeowner and resident costs, rather than relying on the current set of programs that include rigorous qualification requirements and excludes renters, who make up over half of Chicago's households (Institute for Housing Studies, 2023). The required money can be collected through Drinking Water State Revolving Funds, Water Infrastructure Finance and Innovation Act loans, and funds made available through the Bipartisan Infrastructure Law and Water Infrastructure Improvements for the Nation Act. If there is a funding gap, another option involves leaving part of the cost to the water utility, which could then distribute the cost through service fees beyond the already vulnerable households that need service line replacements (Perry, 2024).

Under a mandatory program, Chicago could also capitalize on the efficiency of replacing lines by block rather than individually (Gonzalez, 2023). The city already attempts to benefit from block-by-block replacement through the Block-Long Replacement Program, but this initiative is currently limited to scheduled maintenance by the Department of Water Management (Lead-Safe Chicago, n.d.). To deal with the higher workforce demand (Illinois Lead Service Line

[11] Evidence of service line replacement or renovation records, or following the steps outlined here: https://www.chicagowaterquality.org/LSLIdentification.

Replacement Advisory Board, 2023), Chicago should pursue an apprenticeship program like the one Newark pioneered to simultaneously provide jobs and increase capacity (Wells of Opportunity, 2020).

In the absence of universal testing followed by replacement, there are still options that minimize the administrative burden faced by residents. As low uptake rates, low completion rates, and disparities across groups exemplifies, interventions need to take place in both the initial stage of engagement and in the subsequent compliance and return steps. The Department of Water Management has already pursued these goals, reducing compliance costs through a text message reminder and a simplified redesign of the instruction manual (Souders et al., 2024). The improvements were largely successful, substantially increasing completion rates across income bands. Additionally, I have shown that offering multiple ways to use the 311 system and providing both a self-test and visit option help increase accessibility among historically disadvantaged groups.

Yet, the indicators of obstructive barriers persist. Future interventions should focus on both learning and compliance costs. In particular, the City could raise awareness of lead risk and increase advocacy for the free testing program by sending pamphlets warning residents of single households and two flat homes built before 1986 that they likely have their water delivered by lead pipes and instructing them how to order a test. The Department of Water Management should also reach out to residents that have requested test kits two weeks after delivery to schedule pick up. This action would have the dual benefit of reminding people to complete the tests and removing the requirement that citizens initiate contact with the government a second time to receive the testing service. Most importantly, Chicago should extend the innovations of the testing system to the lead replacement programs. Rather than require an application and be conditioned on income requirements, daycare location, emergency breaks and leaks, or planned maintenance, free service line replacement should be universally accessible. Finally, the City should proactively contact citizens to engage with the program, as it is currently doing for daycare centers in low-income areas, rather than rely on citizen-initiated contact (Lead-Safe Chicago, n.d.).

As these reforms are implemented, Chicago should remain vigilant that its efforts are not exacerbating existing inequities. In the context of domestic well water, research has shown that even when testing was promoted and provided for free, successfully increasing testing rates, the connection between testing and socioeconomic status or education was exacerbated (Flanagan et al., 2016). Rather than equalize testing access, these programs continued to have greatest utilization by educated, wealthier homes.

Beyond lead service line testing and replacement, this research has broad implications for the wider world of public service provision. I caution local governments to reconsider their reliance on voluntary programs and citizen-initiated contact to gather information on city needs and distribute resources. Cities should be careful to identify when public burden is illegitimate, and analyze the value of universal programs as an alternative (Nisar & Masood, 2022). In a country where public trust in government is historically scarce (Pew Research Center, 2024), delivering services in an effective and fair manner is critical.

Conclusion

Ingesting lead carries devastating health consequences, especially for children. In addition to increasing cardiovascular, cancer, and overall mortality, lead is a potent neurotoxin. Exposure to lead impairs cognitive development, disrupts behavioral function, and significantly contributes to societal levels of crime and violence. Like other avenues of lead exposure, water contamination disproportionately affects low income groups and racial minorities. Historical political clout from lead industry and plumbers unions has left Chicago with extensive lead infrastructure, with lead pipes making up 80% of service lines in the city. Despite federal regulations requiring monitoring standards, municipal surveillance is insufficient, leaving the extent of lead exposure under-tested and under-recognized by the government. While city based testing reveals little lead risk, 70% of tests conducted through Chicago's voluntary, resident-initiated lead testing program detect elevated lead levels.

Unfortunately, existing literature suggests that offering lead testing through an opt-in system that requires citizens to reach out to the government creates intrinsic barriers to participation. Not only does this burden lower overall access, but theory predicts that those with the least resources in society will be underrepresented the most. Inequitable access is particularly problematic when such a high proportion of the tests show elevated lead levels. Additionally, the results of tests empower residents to order filters or qualify them for free service line replacement. Thus, when people are left out of the testing process, it leaves the most vulnerable citizens without support.

By analyzing patterns in the 311 request database and census data, I find evidence of barriers stemming from both learning and compliance costs. This means that program implementation is restricting the effectiveness of the policy. Additionally, those with low socioeconomic status are more likely to be excluded from participation, reinforcing evidence for the distributive effects of administrative burden. Interestingly, once controlled for socioeconomic status and objective risk, high populations of Black and Hispanic residents increased the likelihood of engaging with the program by requesting a lead test. However, these populations are also associated with lower rates of completing a test once requested. This discrepancy indicates that compliance costs are the major factor affecting minority participation. Finally, distinct patterns in the type of test requested reveals adaptive behavior by residents that helps minimize the impact of barriers.

This study provides a significant improvement over previous research on administrative burden in service request systems. Lead testing behavior provides unique insight because demand and need for services are distinct and quantifiable. Additional features included in the request database also enable deeper investigations into mechanistic explanations. Nevertheless, the aggregated nature of the census data and the quantitative focus of this thesis provide limitations to interpreting behavior on an individual scale. Future research should attempt to fill this gap through survey instruments and qualitative interviews that focus further on resident experiences.

Despite previously requiring lead service lines by law, Chicago today does not conduct universal lead testing nor provide free replacement to every resident. This failure perpetuates structural inequalities by preventing the people already most likely to have lead exposure from accessing the resources they need to address it. Since lead is a neurotoxin that interrupts cognitive development, worsens academic achievement, lowers impulse control, reduces future earnings, and increases the likelihood of crime, the city's neglect to solve the issue preserves a historical legacy that prevents social mobility. Lead exposure is therefore a viable mechanism for the intergenerational continuation of poverty. However, this problem is not intractable. By investing in a universal and mandatory testing and replacement effort, Chicago can prevent the past from poisoning the future.

References

Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. "Do Low Levels of Blood Lead Reduce Children's Future Test Scores?" *American Economic Journal. Applied Economics* 10, no. 1 (January 2018): 307–41. https://doi.org/10.1257/app.20160404.

Baehler, Karen J., Marquise McGraw, Michele J. Aquino, Ryan Heslin, Lindsay McCormick, and Tom Neltner. "Full Lead Service Line Replacement: A Case Study of Equity in Environmental Remediation." *Sustainability* 14, no. 1 (January 2022): 352. https://doi.org/10.3390/su14010352.

Ballard, April M., Regine Haardöerfer, Nadya Prood, Chukwudi Mbagwu, Hannah L. F. Cooper, and April M. Young. "Willingness to Participate in At-Home HIV Testing Among Young Adults Who Use Opioids in Rural Appalachia." *AIDS and Behavior* 25, no. 3 (March 2021): 699–708. https://doi.org/10.1007/s10461-020-03034-6.

Ballotpedia. "Largest Cities in the United States by Population." Ballotpedia. Accessed April 9, 2025. https://ballotpedia.org/Largest_cities_in_the_United_States_by_population.

Bell, Peter. "Public Trust in Government: 1958-2024." *Pew Research Center* (blog), June 24, 2024. https://www.pewresearch.org/politics/2024/06/24/public-trust-in-government-1958-2024/.

Bidmon, Sonja, and Ralf Terlutter. "Gender Differences in Searching for Health Information on the Internet and the Virtual Patient-Physician Relationship in Germany: Exploratory Results on How Men and Women Differ and Why." *Journal of Medical Internet Research* 17, no. 6 (June 22, 2015): e156. https://doi.org/10.2196/jmir.4127.

Binns, Helen J., Susan A. LeBailly, Ann R. Fingar, and Stephen Saunders. "Evaluation of Risk Assessment Questions Used to Target Blood Lead Screening in Illinois." *Pediatrics* 103, no. 1 (January 1, 1999): 100–106. https://doi.org/10.1542/peds.103.1.100.

Borins, Sandford F. *Innovations in Government: Research, Recognition, and Replication*. Brookings Institution Press, 2009.

Bovaird, T., G. G. Van Ryzin, E. Loeffler, and S. Parrado. "Activating Citizens to Participate in Collective Co-Production of Public Services." *Journal of Social Policy* 44, no. 1 (January 2015): 1–23. https://doi.org/10.1017/S0047279414000567.

Brady, Henry E., Sidney Verba, and Kay Lehman Schlozman. "Beyond SES: A Resource Model of Political Participation." *American Political Science Review* 89, no. 2 (June 1995): 271–94. https://doi.org/10.2307/2082425.

Brudney, Jeffrey L., and Robert E. England. "Toward a Definition of the Coproduction Concept." *Public Administration Review* 43, no. 1 (1983): 59–65. https://doi.org/10.2307/975300.

Burlingame, Gary A., Cathy Bailey, James Nelson, Verna J. Arnette, Scott Bradway, Andrea R. Holthouse Putz, Alan Stark, et al. "Lessons Learned From Helping Schools Manage Lead in Drinking Water to Protect Children's Health." *Journal AWWA* 110, no. 10 (2018): 44–53.

Campbell, Carla, Rachael Greenberg, Deepa Mankikar, and Ronald D. Ross. "A Case Study of Environmental Injustice: The Failure in Flint." *International Journal of Environmental Research and Public Health* 13, no. 10 (September 27, 2016): 951. https://doi.org/10.3390/ijerph13100951.

Canfield, Richard L., Charles R. Henderson, Deborah A. Cory-Slechta, Christopher Cox, Todd A. Jusko, and Bruce P. Lanphear. "Intellectual Impairment in Children with Blood Lead Concentrations below 10 Microg per Deciliter." *The New England Journal of Medicine* 348, no. 16 (April 17, 2003): 1517–26. https://doi.org/10.1056/NEJMoa022848.

Chicago. “Water Supply.” City of Chicago. Accessed November 28, 2024. https://www.chicago.gov/content/city/en/depts/water/provdrs/supply.html.
Chicago Department of Public Health. “CHILDHOOD LEAD POISONING PREVENTION & DATA REPORT, 2019-2021.” Chicago, IL: City of Chicago, 2023.
Chicago Department of Water Management. “Know Your Water.” Chicago Department of Water Management. Accessed April 12, 2025. https://www.chicagowaterquality.org/home.
———. “Water Service Line Inventory.” Chicago Department of Water Management. Accessed November 28, 2024. https://sli.chicagowaterquality.org/.
Chojnacki, Alex, Chengyu Dai, Arya Farahi, Guangsha Shi, Jared Webb, Daniel T. Zhang, Jacob Abernethy, and Eric Schwartz. “A Data Science Approach to Understanding Residential Water Contamination in Flint.” In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1407–16. KDD ’17. New York, NY, USA: Association for Computing Machinery, 2017. https://doi.org/10.1145/3097983.3098078.
Christensen, Julian, Lene Aarøe, Martin Baekgaard, Pamela Herd, and Donald P. Moynihan. “Human Capital and Administrative Burden: The Role of Cognitive Resources in Citizen-State Interactions.” *Public Administration Review* 80, no. 1 (2020): 127–36. https://doi.org/10.1111/puar.13134.
Cilliers, Louise, and Francois Retief. “Chapter 14 - Lead Poisoning and the Downfall of Rome: Reality or Myth?” In *Toxicology in Antiquity (Second Edition)*, edited by Philip Wexler, 221–29. History of Toxicology and Environmental Health. Academic Press, 2019. https://doi.org/10.1016/B978-0-12-815339-0.00014-7.
Circle Blue. “New EPA Federal Water Rules Put Chicago on Track to Replace All Lead Pipes Within 20 Years.” Circle of Blue, December 18, 2024. http://www.circleofblue.org/2024/supply/water-management/infrastructure/new-epa-federal-water-rules-put-chicago-on-track-to-replace-all-lead-pipes-within-20-years/.
City of Chicago. “311 Service Requests | City of Chicago | Data Portal.” Accessed April 11, 2025. https://data.cityofchicago.org/Service-Requests/311-Service-Requests/v6vf-nfxy/about_data.
Clark, Benjamin Y., and Jeffrey L. Brudney. “Citizen Representation in City Government-Driven Crowdsourcing.” *Comput. Supported Coop. Work* 27, no. 3–6 (December 1, 2018): 1153–80. https://doi.org/10.1007/s10606-018-9308-2.
Clark, Benjamin Y., Brudney, Jeffrey L., and Jang, Sung-Gheel. “Coproduction of Government Services and the New Information Technology: Investigating the Distributional Biases.” *Public Administration Review* 73, no. 5 (August 12, 2013): 687–701. https://doi.org/10.1111/puar.12092.
Clark, Benjamin Y., Jeffrey L. Brudney, Sung-Gheel Jang, and Bradford Davy. “Do Advanced Information Technologies Produce Equitable Government Responses in Coproduction: An Examination of 311 Systems in 15 U.S. Cities.” *The American Review of Public Administration* 50, no. 3 (April 2020): 315–27. https://doi.org/10.1177/0275074019894564.
Clark, Benjamin Y., and Maria Shurik. “Improving Citizen Satisfaction with Local Government Using 311 Systems: The Case of San Francisco, California.” In *Innovation in the Public and Nonprofit Sectors*. Routledge, 2015.

Clark, Benjamin Y., and Maria Shurik (Rokakis). "Do 311 Systems Shape Citizen Satisfaction with Local Government?" SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, September 1, 2014. https://doi.org/10.2139/ssrn.2491034.

Combined Sewer Overflow Committee. "Wells of Opportunity: Training Residents and Prioritizing Local Hiring for Water Infrastructure Projects in Newark." Jersey Water Works, November 2020. https://www.jerseywaterworks.org/resources/wells-of-opportunity-training-residents-and-prioritizing-local-hiring-for-water-infrastructure-projects-in-newark/.

Cook, Scott J., Samantha Zuhlke, and Robin Saywitz. "Potholes, 311 Reports, and a Theory of Heterogeneous Resident Demand for City Services." *Policy Studies Journal* 52, no. 3 (August 2024): 647–69. https://doi.org/10.1111/psj.12540.

Corley, Cheryl. "Chicago's Upgrades To Aging Water Lines May Disturb Lead Pipes." *NPR*, April 14, 2016, sec. Environment. https://www.npr.org/2016/04/14/474130954/chicagos-upgrades-to-aging-water-lines-may-disturb-lead-pipes.

Courtney, Joseph G. "Decreases in Young Children Who Received Blood Lead Level Testing During COVID-19 — 34 Jurisdictions, January–May 2020." *MMWR. Morbidity and Mortality Weekly Report* 70 (2021). https://doi.org/10.15585/mmwr.mm7005a2.

Craigie, Terry-Ann, Samuel L. Myers, and William A. Darity. "Racial Differences in the Effect of Marriageable Males on Female Family Headship." *Journal of Demographic Economics* 84, no. 3 (September 2018): 231–56. https://doi.org/10.1017/dem.2018.3.

Cunningham, Maureen. "Echoing Newark: How American Cities Can Replicate Newark's Success in Replacing Over 23,000 Lead Pipes in Under Three Years." Environmental Policy Innovation Center. Accessed April 9, 2025. https://www.policyinnovation.org/insights/echoing-newark-how-american-cities-can-replicate-newarks-success-in-replacing-over-23000-lead-pipes-in-under-three-years.

Das, Amit, and Shobha S. Das. "E-Government and Entrepreneurship: Online Government Services and the Ease of Starting Business." *Information Systems Frontiers* 24, no. 3 (June 1, 2022): 1027–39. https://doi.org/10.1007/s10796-021-10121-z.

Del Rio, Michelle, and Jacqueline Gibson. "Factors Influencing Inequities in Lead Exposure in United States Children: A Systematic Review." *International Public Health Journal* 15, no. 3 (2023): 263–81.

Del Rio, Michelle and Gibson, Jacqueline MacDonald. "Factors Influencing Inequities in Lead Exposure in United States Children: A Systematic Review." *International Public Health Journal* 15, no. 3 (2023): 263–81.

Del Toral, Miguel A., Andrea Porter, and Michael R. Schock. "Detection and Evaluation of Elevated Lead Release from Service Lines: A Field Study." *Environmental Science & Technology* 47, no. 16 (August 20, 2013): 9300–9307. https://doi.org/10.1021/es4003636.

Dignam, Timothy, Rachel B. Kaufmann, Lauren LeStourgeon, and Mary Jean Brown. "Control of Lead Sources in the United States, 1970-2017: Public Health Progress and Current Challenges to Eliminating Lead Exposure." *Journal of Public Health Management and Practice : JPHMP* 25, no. Suppl 1 LEAD POISONING PREVENTION (February 2019): S13. https://doi.org/10.1097/PHH.0000000000000889.

Egan, Kathryn B., Cheryl R. Cornwell, Joseph G. Courtney, and Adrienne S. Ettinger. "Blood Lead Levels in U.S. Children Ages 1–11 Years, 1976–2016." *Environmental Health Perspectives* 129, no. 3 (March 2021): 037003. https://doi.org/10.1289/EHP7932.

EIA. “Gasoline Explained: Gasoline and the Environment.” EIA, December 29, 2022. https://www.eia.gov/energyexplained/gasoline/gasoline-and-the-environment-leaded-gasoline.php.

Eng, Monica. “How Political Clout And Union Power Created Chicago’s Lead Water Problem.” WBEZChicago, September 15, 2018. https://www.wbez.org/curious-city/2018/09/15/how-political-clout-and-union-power-created-chicagos-lead-water-problem.

———. “What We Do — And Don’t — Know About Chicago’s Lead Water Problem.” WBEZChicago, September 19, 2020. https://www.wbez.org/curious-city/2020/09/19/what-we-do-and-dont-know-about-chicagos-lead-water-problem.

Environmental Defense Fund. “Top 10 Cities with the Most Lead Pipes.” Environmental Defense Fund, April 23, 2024. https://www.edf.org/top-10-cities-most-lead-pipes.

Evens, Anne, Daniel Hryhorczuk, Bruce P. Lanphear, Kristin M. Rankin, Dan A. Lewis, Linda Forst, and Deborah Rosenberg. “The Impact of Low-Level Lead Toxicity on School Performance among Children in the Chicago Public Schools: A Population-Based Retrospective Cohort Study.” *Environmental Health: A Global Access Science Source* 14 (April 7, 2015): 21. https://doi.org/10.1186/s12940-015-0008-9.

Ezell, Jerel M., and Elizabeth C. Chase. “Forming a Critical Race Theory of Environmental Disaster: Understanding Social Meanings and Health Threat Perception in the Flint Water Crisis.” *Journal of Environmental Management* 320 (October 2022): 115886. https://doi.org/10.1016/j.jenvman.2022.115886.

Feigenbaum, James J, and Andrew B. Hall. “How High-Income Neighborhoods Receive More Service from Municipal Government: Evidence from City Administrative Data.” *SSRN Electronic Journal*, 2015. https://doi.org/10.2139/ssrn.2631106.

Flanagan, Sara V., Robert G. Marvinney, and Yan Zheng. “Influences on Domestic Well Water Testing Behavior in a Central Maine Area with Frequent Groundwater Arsenic Occurrence.” *Science of The Total Environment* 505 (February 1, 2015): 1274–81. https://doi.org/10.1016/j.scitotenv.2014.05.017.

Flanagan, Sara V., Steven E. Spayd, Nicholas A. Procopio, Steven N. Chillrud, James Ross, Stuart Braman, and Yan Zheng. “Arsenic in Private Well Water Part 2 of 3: Who Benefits the Most from Traditional Testing Promotion?” *Science of The Total Environment* 562 (August 15, 2016): 1010–18. https://doi.org/10.1016/j.scitotenv.2016.03.199.

Froehlich, Tanya E., Bruce P. Lanphear, Peggy Auinger, Richard Hornung, Jeffery N. Epstein, Joe Braun, and Robert S. Kahn. “Association of Tobacco and Lead Exposures with Attention-Deficit/Hyperactivity Disorder.” *Pediatrics* 124, no. 6 (December 2009): e1054-1063. https://doi.org/10.1542/peds.2009-0738.

Gazze, Ludovica. “Hassles and Environmental Health Screenings: Evidence from Lead Tests in Illinois.” *Journal of Human Resources* 59, no. 4 (July 2024): 1090–1121. https://doi.org/10.3368/jhr.0221-11478R2.

Gilfillan, S. C. “Lead Poisoning and the Fall of Rome.” *Journal of Occupational Medicine* 7, no. 2 (1965): 53–60.

Gonzalez, Brianna. “Chicago’s Leaded Legacy: Past, Present, and Future,” n.d.

Gorelick, Marc H., Lindsay Gould, Mark Nimmer, Duke Wagner, Mary Heath, Hiba Bashir, and David C. Brousseau. “Perceptions About Water and Increased Use of Bottled Water in

Minority Children." *Archives of Pediatrics & Adolescent Medicine* 165, no. 10 (October 1, 2011): 928–32. https://doi.org/10.1001/archpediatrics.2011.83.

Grumbach, Jacob M., and Alexander Sahn. "Race and Representation in Campaign Finance." *American Political Science Review* 114, no. 1 (February 2020): 206–21. https://doi.org/10.1017/S0003055419000637.

Hamel, Brian T., and Derek E. Holliday. "Unequal Responsiveness in City Service Delivery: Evidence from 42 Million 311 Calls." *Quarterly Journal of Political Science* 19, no. 3 (2024): 243–74. https://doi.org/10.1561/100.00022089.

Hanna-Attisha, Mona, Jenny LaChance, Richard Casey Sadler, and Allison Champney Schnepp. "Elevated Blood Lead Levels in Children Associated With the Flint Drinking Water Crisis: A Spatial Analysis of Risk and Public Health Response." *American Journal of Public Health* 106, no. 2 (February 2016): 283–90. https://doi.org/10.2105/AJPH.2015.303003.

Heinrich, Carolyn J. "The Bite of Administrative Burden: A Theoretical and Empirical Investigation." *Journal of Public Administration Research and Theory* 26, no. 3 (July 2016): 403–20.

Herd, Pamela, and Donald P. Moynihan. *Administrative Burden: Policymaking by Other Means*. Russell Sage Foundation, 2019.

Hernberg, Sven. "Lead Poisoning in a Historical Perspective." *American Journal of Industrial Medicine* 38, no. 3 (2000): 244–54. https://doi.org/10.1002/1097-0274(200009)38:3<244::AID-AJIM3>3.0.CO;2-F.

Hollingsworth, Alex, Mike Huang, Ivan J Rudik, and Nicholas J Sanders. "A Thousand Cuts: Cumulative Lead Exposure Reduces Academic Achievement," n.d.

Huynh, Benjamin Q., Elizabeth T. Chin, and Mathew V. Kiang. "Estimated Childhood Lead Exposure From Drinking Water in Chicago." *JAMA Pediatrics* 178, no. 5 (May 2024): 473–79. https://doi.org/10.1001/jamapediatrics.2024.0133.

Hwang, Hyunseok, Vedlitz ,Arnold, Mostafavi ,Ali, and Sungdae and Lim. "Socially Equitable Public Service Provision: Towards Inclusive Coproduction of Public Infrastructure." *International Journal of Public Administration* 0, no. 0 (n.d.): 1–15. https://doi.org/10.1080/01900692.2024.2427652.

Jiffy Lube, "What Is Engine Knock?"

Jones, Bryan D., Greenberg, Saadia R., Kaufman, Clifford, and Drew, Joseph. "Bureaucratic Response to Citizen-Initiated Contacts: Environmental Enforcement in Detroit." *American Political Science Review*, 1977. https://www.cambridge.org/core/journals/american-political-science-review/article/abs/bureaucratic-response-to-citizeninitiated-contacts-environmental-enforcement-in-detroit/9613ABC2B516CAAA2D27D76439E2BDB7.

Jones, Robert L., David M. Homa, Pamela A. Meyer, Debra J. Brody, Kathleen L. Caldwell, James L. Pirkle, and Mary Jean Brown. "Trends in Blood Lead Levels and Blood Lead Testing Among US Children Aged 1 to 5 Years, 1988–2004." *Pediatrics* 123, no. 3 (March 1, 2009): e376–85. https://doi.org/10.1542/peds.2007-3608.

Katner, Adrienne, Kelsey J. Pieper, Yanna Lambrinidou, Komal Brown, Chih-Yang Hu, Howard W. Mielke, and Marc A. Edwards. "Weaknesses in Federal Drinking Water Regulations and Public Health Policies That Impede Lead Poisoning Prevention and Environmental Justice." *Environmental Justice* 9, no. 4 (August 2016): 109–17. https://doi.org/10.1089/env.2016.0012.

Kemper, Alex R., and Sarah J. Clark. "Physician Barriers to Lead Testing of Medicaid-Enrolled Children." *Ambulatory Pediatrics* 5, no. 5 (September 1, 2005): 290–93. https://doi.org/10.1367/A05-008R.1.

Kirkwood, James. "Collection of Reports (Condensed) and Opinions of Chemists in Regard to the Use of Lead Pipe for Service Pipe in the Distribution of Water for the Supply of Cities." Wellcome Collection. Accessed November 12, 2024. https://wellcomecollection.org/works/qrkg9u95/items.

Knowles, Jason, Pistone, Ann, and Green, Maggie. "Chicago Lead Pipes: City Working to Replace over 400K Service Lines; How You Can Test Your Water." ABC7 Chicago, July 20, 2023. https://abc7chicago.com/lead-in-water-pipes-chicago-testing/13526059/.

Kontokosta, Constantine, Boyeong Hong, and Kristi Korsberg. "Equity in 311 Reporting: Understanding Socio-Spatial Differentials in the Propensity to Complain." arXiv, 2017. https://doi.org/10.48550/ARXIV.1710.02452.

Lanphear, Bruce P., Richard Hornung, Jane Khoury, Kimberly Yolton, Peter Baghurst, David C. Bellinger, Richard L. Canfield, et al. "Low-Level Environmental Lead Exposure and Children's Intellectual Function: An International Pooled Analysis." *Environmental Health Perspectives* 113, no. 7 (July 2005): 894–99. https://doi.org/10.1289/ehp.7688.

Lanphear, Bruce P., Stephen Rauch, Peggy Auinger, Ryan W. Allen, and Richard W. Hornung. "Low-Level Lead Exposure and Mortality in US Adults: A Population-Based Cohort Study." *The Lancet. Public Health* 3, no. 4 (April 2018): e177–84. https://doi.org/10.1016/S2468-2667(18)30025-2.

Lead Service Line Replacement Advisory Board. "Report of Recommendations," June 29, 2023.

LeBrón, Alana M. W., Ivy R. Torres, Enrique Valencia, Miriam López Dominguez, Deyaneira Guadalupe Garcia-Sanchez, Michael D. Logue, and Jun Wu. "The State of Public Health Lead Policies: Implications for Urban Health Inequities and Recommendations for Health Equity." *International Journal of Environmental Research and Public Health* 16, no. 6 (January 2019): 1064. https://doi.org/10.3390/ijerph16061064.

Leighley, Jan E., and Arnold Vedlitz. "Race, Ethnicity, and Political Participation: Competing Models and Contrasting Explanations." *The Journal of Politics* 61, no. 4 (November 1999): 1092–1114. https://doi.org/10.2307/2647555.

Levin, Ronnie, Mary Jean Brown, Michael E. Kashtock, David E. Jacobs, Elizabeth A. Whelan, Joanne Rodman, Michael R. Schock, Alma Padilla, and Thomas Sinks. "Lead Exposures in U.S. Children, 2008: Implications for Prevention." *Environmental Health Perspectives* 116, no. 10 (May 19, 2008): 1285. https://doi.org/10.1289/ehp.11241.

Levine, Charles H., and Glenn Fisher. "Citizenship and Service Delivery: The Promise of Coproduction." *Public Administration Review* 44 (1984): 178–89. https://doi.org/10.2307/975559.

Lewis, Jack. "Looking Backward: A Historical Perspective on Environmental Regulations." Reports and Assessments, March 1988. https://www.epa.gov/archive/epa/aboutepa/looking-backward-historical-perspective-environmental-regulations.html.

Lidsky, Theodore I., and Jay S. Schneider. "Lead Neurotoxicity in Children: Basic Mechanisms and Clinical Correlates." *Brain: A Journal of Neurology* 126, no. Pt 1 (January 2003): 5–19. https://doi.org/10.1093/brain/awg014.

Linos, Katerina, Melissa Carlson, Laura Jakli, Nadia Dalma, Isabelle Cohen, Afroditi Veloudaki, and Stavros Nikiforos Spyrellis. "How Do Disadvantaged Groups Seek Information About

Public Services? A Randomized Controlled Trial of Communication Technologies." *Public Administration Review* 82, no. 4 (2022): 708–20. https://doi.org/10.1111/puar.13437.
Lippert, Julia, James Montgomery, and Camille DeMarco. "Lead Health Fairs: A Community-Based Approach to Addressing Lead Exposure in Chicago." *Health Education & Behavior* 48, no. 6 (December 1, 2021): 758–68. https://doi.org/10.1177/1090198120954359.
Liu, Jiawen, Lara P. Clark, Matthew J. Bechle, Anjum Hajat, Sun-Young Kim, Allen L. Robinson, Lianne Sheppard, Adam A. Szpiro, and Julian D. Marshall. "Disparities in Air Pollution Exposure in the United States by Race/Ethnicity and Income, 1990–2010." *Environmental Health Perspectives* 129, no. 12 (December 2021): 127005. https://doi.org/10.1289/EHP8584.
MacDonald, Kristi, and Mara Tippett. "Reducing Public Exposure to Common, Harmful Well Water Contaminants through Targeted Outreach to Highly Susceptible Neighborhoods as a Method of Increasing the Likelihood of Testing and Treatment of Water from Private Wells." *Journal of Water and Health* 18, no. 4 (July 8, 2020): 522–32. https://doi.org/10.2166/wh.2020.059.
Mandatory Replacement of Lead Service Line, 16:23 City of Newark Code § (2019). https://ecode360.com/36709572.
Masood, Ayesha, and Muhammad Azfar Nisar. "Administrative Capital and Citizens' Responses to Administrative Burden." *Journal of Public Administration Research and Theory* 31, no. 1 (January 1, 2021): 56–72. https://doi.org/10.1093/jopart/muaa031.
Mathews, Allison, Donaldson Conserve, Hailey Mason, Lé Marus Alston, Stuart Rennie, and Joseph Tucker. "'Informed and Empowered': A Mixed-Methods Study of Crowdsourcing Contests to Promote Uptake of HIV Self-Testing Kits among African Americans." *Journal of Virus Eradication* 6, no. 2 (April 2020): 74–80. https://doi.org/10.1016/S2055-6640(20)30020-0.
Maynard-Moody, Steven, and Michael Musheno. "Social Equities and Inequities in Practice: Street-Level Workers as Agents and Pragmatists." *Public Administration Review* 72, no. s1 (2012): S16–23. https://doi.org/10.1111/j.1540-6210.2012.02633.x.
McCormick, Erin, and Aliya Uteuova. "Profiting from Poison: How the US Lead Industry Knowingly Created a Water Crisis." *The Guardian*, September 22, 2022, sec. US news. https://www.theguardian.com/us-news/2022/sep/22/us-lead-industry-history-water-crisis.
McCormick, Erin, Aliya Uteuova, Taylor Moore, and Taylor Moore with photographs by Jamie Kelter Davis. "Revealed: The 'Shocking' Levels of Toxic Lead in Chicago Tap Water." *The Guardian*, September 21, 2022, sec. US news. https://www.theguardian.com/us-news/2022/sep/21/lead-contamination-chicago-tap-water-revealed.
McFarland, Michael J., Matt E. Hauer, and Aaron Reuben. "Half of US Population Exposed to Adverse Lead Levels in Early Childhood." *Proceedings of the National Academy of Sciences* 119, no. 11 (March 15, 2022): e2118631119. https://doi.org/10.1073/pnas.2118631119.
Menke, Andy, Paul Muntner, Vecihi Batuman, Ellen K. Silbergeld, and Eliseo Guallar. "Blood Lead Below 0.48 Mmol/L (10 Mg/dL) and Mortality Among US Adults." *Circulation* 114, no. 13 (September 26, 2006): 1388–94. https://doi.org/10.1161/CIRCULATIONAHA.106.628321.

Michael Phillis and Alexa St. John. “In Policy Reversal, Trump Eliminates Help for Black and Latino Communities Hit Harder by Pollution.” AP News, January 26, 2025. https://apnews.com/article/climate-health-trump-environmental-justice-pollution-59baf164c81049d3995d2dfe1eceda8c.

Mielke, Howard W., and Sammy Zahran. “The Urban Rise and Fall of Air Lead (Pb) and the Latent Surge and Retreat of Societal Violence.” *Environment International* 43 (August 1, 2012): 48–55. https://doi.org/10.1016/j.envint.2012.03.005.

Minkoff, Scott L. “NYC 311: A Tract-Level Analysis of Citizen–Government Contacting in New York City.” *Urban Affairs Review* 52, no. 2 (March 1, 2016): 211–46. https://doi.org/10.1177/1078087415577796.

Mladenka, Kenneth R. “The Distribution of an Urban Public Service: The Changing Role of Race and Politics.” *Urban Affairs Quarterly* 24, no. 4 (June 1, 1989): 556–83. https://doi.org/10.1177/004208168902400405.

Moore, Joanna Faye. “DEATH METAL: Characterising the Effects of Environmental Lead Pollution on Mobility and Childhood Health within the Roman Empire.,” n.d.

Moore, Joanna, Kori Filipek, Vana Kalenderian, Rebecca Gowland, Elliott Hamilton, Jane Evans, and Janet Montgomery. “Death Metal: Evidence for the Impact of Lead Poisoning on Childhood Health within the Roman Empire.” *International Journal of Osteoarchaeology* 31, no. 5 (2021): 846–56. https://doi.org/10.1002/oa.3001.

Morris, Kevin, and Coryn Grange. “Growing Racial Disparities in Voter Turnout, 2008–2022 | Brennan Center for Justice.” Brennan Center for Justice, March 2, 2024. https://www.brennancenter.org/our-work/research-reports/growing-racial-disparities-voter-turnout-2008-2022.

National Geographic Society. *National Geographic November 1923*, 1923. http://archive.org/details/nationalgeographic19231101.

Needleman, Herbert. “History of Lead Poisoning in the World.” Biological Diversity, 1999.

Nilsson, J Peter. “The Long Term Effects of Early Childhood Lead Exposure: Evidence from the Phase-out of Leaded Gasoline,” n.d.

Nisar, Muhammad Azfar, and Ayesha Masood. “Dealing with Disgust: Street-Level Bureaucrats as Agents of Kafkaesque Bureaucracy.” *Organization* 27, no. 6 (November 1, 2020): 882–99. https://doi.org/10.1177/1350508419883382.

Nisar, Muhammad Azfar, and Ayesha and Masood. “Are All Burdens Bad? Disentangling Illegitimate Administrative Burdens through Public Value Accounting.” *Asia Pacific Journal of Public Administration* 45, no. 4 (October 2, 2023): 385–403. https://doi.org/10.1080/23276665.2022.2088581.

Olympian Water Testing. “The Connection Between Lead Testing and Economic Development.” Olympian Water Testing, March 5, 2023. https://olympianwatertesting.com/aquawiki/lead/the-connection-between-lead-testing-and-economic-development/.

Ozili, Peterson K. “The Acceptable R-Square in Empirical Modelling for Social Science Research,” n.d.

Pamukcu, Duygu, and Christopher Zobel. “Characterizing 311 System Reactions to a Global Health Emergency,” 2021. https://doi.org/10.24251/HICSS.2021.272.

Paul, Michael P., Pierce Rigrod, Steve Wingate, and Mark E. Borsuk. “A Community-Driven Intervention in Tuftonboro, New Hampshire, Succeeds in Altering Water Testing Behavior.” *Journal of Environmental Health* 78, no. 5 (December 2015): 30–39.

Perry, Chakena D. "What the EPA's New Federal Water Rule Means for Chicagoans," October 10, 2024. https://www.nrdc.org/bio/chakena-d-perry/what-epas-new-federal-water-rule-means-chicagoans.

Pew Research Center. "How Americans Interact with Government Online." Pew Research Center, April 27, 2010. https://www.pewresearch.org/2010/04/27/how-americans-interact-with-government-online/.

Prats, Mariana and Meunier, Axel. "Political Efficacy and Participation: An Empirical Analysis in European Countries." OECD Working Papers on Public Governance. Vol. 46. OECD Working Papers on Public Governance, June 16, 2021. https://doi.org/10.1787/4548cad8-en.

PRB. "U.S. Fertility Rates Higher Among Minorities." PRB, January 2003. https://www.prb.org/resources/u-s-fertility-rates-higher-among-minorities/.

Rabin, Richard. "The Lead Industry and Lead Water Pipes 'A MODEST CAMPAIGN.'" *American Journal of Public Health* 98, no. 9 (September 2008): 1584. https://doi.org/10.2105/AJPH.2007.113555.

Reichel, Chloe. "Toxic Waste Sites and Environmental Justice: Research Roundup," n.d.

Reporting (PAR), Cole Longcor, UIS Public Affairs. "As State Continues to Inventory Lead Pipes, Full Replacement Deadlines Are Decades Away." Capitol News Illinois, April 13, 2024. https://capitolnewsillinois.com/news/as-state-continues-to-inventory-lead-pipes-full-replacement-deadlines-are-decades-away/.

Reyes, Cecilia and Hawthorne, Michael. "Brain-Damaging Lead Found in Tap Water in Hundreds of Homes Tested across Chicago, Results Show – Chicago Tribune." Chicago Tribune, June 14, 2024. https://www.chicagotribune.com/2018/04/12/brain-damaging-lead-found-in-tap-water-in-hundreds-of-homes-tested-across-chicago-results-show-2/.

Reyes, Jessica Wolpaw. "Lead Exposure and Behavior: Effects on Antisocial and Risky Behavior Among Children and Adolescents." *Economic Inquiry* 53, no. 3 (2015): 1580–1605. https://doi.org/10.1111/ecin.12202.

Rhubart, Danielle Christine, and Anya M. Galli Robertson. "The Right to Knowledge and the Superfund Program: A Fundamental Cause Approach to Disparities in Resident Awareness of Hazardous Waste Sites." *Environmental Justice* 13, no. 5 (October 2020): 181–88. https://doi.org/10.1089/env.2020.0020.

Rowangould, Gregory M. "A Census of the US Near-Roadway Population: Public Health and Environmental Justice Considerations." *Transportation Research Part D: Transport and Environment* 25 (December 2013): 59–67. https://doi.org/10.1016/j.trd.2013.08.003.

Safarpour, Haniye, Deborah A. Carroll, and Lauryn A. Spearing. "Spatial and Demographic Trends in Lead Levels: The Case of Chicago's Water System," March 18, 2024, 641–50. https://doi.org/10.1061/9780784485279.064.

Sampson, Robert J., and Alix S. Winter. "THE RACIAL ECOLOGY OF LEAD POISONING: Toxic Inequality in Chicago Neighborhoods, 1995-2013." *Du Bois Review: Social Science Research on Race* 13, no. 2 (October 2016): 261–83. https://doi.org/10.1017/S1742058X16000151.

Schnaas, Lourdes, Stephen J. Rothenberg, Maria-Fernanda Flores, Sandra Martinez, Carmen Hernandez, Erica Osorio, Silvia Ruiz Velasco, and Estela Perroni. "Reduced Intellectual

Development in Children with Prenatal Lead Exposure." *Environmental Health Perspectives* 114, no. 5 (May 2006): 791–97. https://doi.org/10.1289/ehp.8552.
Schober, Susan E., Lisa B. Mirel, Barry I. Graubard, Debra J. Brody, and Katherine M. Flegal. "Blood Lead Levels and Death from All Causes, Cardiovascular Disease, and Cancer: Results from the NHANES III Mortality Study." *Environmental Health Perspectives* 114, no. 10 (October 2006): 1538–41. https://doi.org/10.1289/ehp.9123.
Schweizer, Valerie and Guzzo, Karen Benjamin. "Distributions of Age at First Birth, 1960-2018." Family Profiles. Bowling Green State University National Center for Family & Marriage Research, 2020. https://www.bgsu.edu/ncfmr/resources/data/family-profiles/schweizer-guzzo-distribution-age-first-birth-fp-20-11.html.
Sharp, Elaine B. "Citizen-Initiated Contacting of Government Officials and Socioeconomic Status: Determining the Relationship and Accounting for It." *The American Political Science Review* 76, no. 1 (1982): 109–15. https://doi.org/10.2307/1960446.
Shaw, Souradet Y., Stella Leung, Shajy Isac, Helgar Musyoki, Mary Mugambi, Japheth Kioko, Janet Musimbi, et al. "Assessing Awareness and Use of HIV Self-Testing Kits after the Introduction of a Community-Based HIV Self-Testing Programme among Men Who Have Sex with Men in Kenya." Edited by Kevin Escandón. *PLOS Global Public Health* 3, no. 8 (August 18, 2023): e0001547. https://doi.org/10.1371/journal.pgph.0001547.
Slough, Tara. "Oversight, Capacity, and Inequality," 2024.
Sobin, Christina, Marisela GUTIéRREZ-VEGA, Gisel Flores-Montoya, Michelle Del Rio, Juan M. Alvarez, Alexander Obeng, Jaleen Avila, and Ganga Hettiarachchi. "Improving Equitability and Inclusion for Testing and Detection of Lead Poisoning in US Children." *The Milbank Quarterly* 101, no. 1 (2023): 48–73. https://doi.org/10.1111/1468-0009.12596.
Sorensen, Lucy C., Ashley M. Fox, Heyjie Jung, and Erika G. Martin. "Lead Exposure and Academic Achievement: Evidence from Childhood Lead Poisoning Prevention Efforts." *Journal of Population Economics* 32, no. 1 (2019): 179–218.
Souders, Sarah, Leonard Lopoo, Sheena Mirpuri, Patrick Schwer, and Emily Cardon. "Decreasing Compliance Costs Using Behavioral Interventions: An Evaluation of Lead Testing in Chicago." *Journal of Behavioral Public Administration* 7 (July 12, 2024). https://doi.org/10.30636/jbpa.71.349.
Southtown, SAMANTHA MOILANEN | Daily. "Chicago Needs to Accelerate Lead-Pipe Replacement Plan, Increase Outreach, Experts Say." *Chicago Tribune* (blog), May 27, 2024. https://www.chicagotribune.com/2024/05/27/chicago-lead-pipes/.
Steven Manson, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. "IPUMS National Historical Geographic Information System," 2024. IPUMS. https://doi.org/10.18128/D050.V19.0.
Stillo III, Frank. "Racial Disparities in Access to Municipal Drinking Water: Lead Exposure Risk and a Risk Communication Intervention." University of North Carolina at Chapel Hill, 2019. https://cdr.lib.unc.edu/concern/dissertations/sj139685h.
Tang, Zhuo, and Margaret Carrel. "The More Things Change the More They Stay the Same: Persistent Spatial Inequity in Pediatric Lead Poisoning in Chicago." *Applied Geography* 142 (May 1, 2022): 102686. https://doi.org/10.1016/j.apgeog.2022.102686.
Taylor, Ashley, Alexis Garretson, Karen H. Bieluch, Kate L. Buckman, Hannah Lust, Cait Bailey, Anna E. Farrell, et al. "A Mixed Methods Approach to Understanding the Public Health Impact of a School-Based Citizen Science Program to Reduce Arsenic in Private

Well Water." *Environmental Health Perspectives* 132, no. 8 (August 2024): 087006. https://doi.org/10.1289/EHP13421.
Teodoro, Manuel P., Samantha Zuhlke, and David Switzer. *The Profits of Distrust: Citizen-Consumers, Drinking Water, and the Crisis of Confidence in American Government*. Business and Public Policy. Cambridge: Cambridge University Press, 2022. https://doi.org/10.1017/9781009244893.
Thomas, John Clayton. "Citizen-Initiated Contacts with Government Agencies: A Test of Three Theories." *American Journal of Political Science* 26, no. 3 (1982): 504–22. https://doi.org/10.2307/2110940.
Tong, Michelle, Samantha Artiga, and Robin Rudowitz Published. "Mitigating Childhood Lead Exposure and Disparities: Medicaid and Other Federal Initiatives." *KFF* (blog), May 20, 2022. https://www.kff.org/racial-equity-and-health-policy/issue-brief/mitigating-childhood-lead-exposure-and-disparities-medicaid-and-other-federal-initiatives/.
University, Institute for Housing Studies-DePaul. "2023 State of Rental Housing in the City of Chicago." Institute for Housing Studies - DePaul University. Accessed April 9, 2025. https://www.housingstudies.org/releases/2023-state-rental-housing-city-chicago/.
US EPA, CENTER FOR PUBLIC HEALTH & ENVIRONMENTAL ASSESSMENT. "Integrated Science Assessment (ISA) for Lead (Final Report)." Reports and Assessments. US EPA, November 20, 2023. https://assessments.epa.gov/isa/document/&deid%3D359536.
US EPA, OA. "Biden-Harris Administration Announces $332 Million for Drinking Water Infrastructure Upgrades in Illinois." News Release, April 4, 2023. Illinois. https://www.epa.gov/newsreleases/biden-harris-administration-announces-332-million-drinking-water-infrastructure.
US EPA, OW. "Lead and Copper Rule." Other Policies and Guidance, October 13, 2015. https://www.epa.gov/dwreginfo/lead-and-copper-rule.
———. "Lead and Copper Rule Improvements." Overviews and Factsheets, May 4, 2022. https://www.epa.gov/ground-water-and-drinking-water/lead-and-copper-rule-improvements.
US EPA, REG 05. "Chicago Lead in Drinking Water Study." Reports and Assessments, October 28, 2015. Illinois. https://www.epa.gov/il/chicago-lead-drinking-water-study.
Verba, Sidney, and Norman H. Nie. *Participation in America: Political Democracy and Social Equality*. Chicago, IL: University of Chicago Press, 1987. https://press.uchicago.edu/ucp/books/book/chicago/P/bo3637096.html.
Verba, Sidney, Kay Lehman Schlozman, Henry Brady, and Norman H. Nie. "Race, Ethnicity and Political Resources: Participation in the United States." *British Journal of Political Science* 23, no. 4 (1993): 453–97.
Wait, Kory, Adrienne Katner, Daniel Gallagher, Marc Edwards, Wilson Mize, Crystal Lee Pow Jackson, and Kelsey J. Pieper. "Disparities in Well Water Outreach and Assistance Offered by Local Health Departments: A North Carolina Case Study." *Science of The Total Environment* 747 (December 10, 2020): 141173. https://doi.org/10.1016/j.scitotenv.2020.141173.
Wang, Lingjing, Cheng Qian, Philipp Kats, Constantine Kontokosta, and Stanislav Sobolevsky. "Structure of 311 Service Requests as a Signature of Urban Location." Edited by Yanguang Chen. *PLOS ONE* 12, no. 10 (October 17, 2017): e0186314. https://doi.org/10.1371/journal.pone.0186314.

White, Deidre, and Tekleab Gala. “Environmental Injustice? Disparities in the Exposure to Environmental Lead Poisoning and Risks among Children in the Chicago Neighborhoods.” *American Journal of Public Health Research* 10, no. 3 (June 15, 2022): 124–33. https://doi.org/10.12691/ajphr-10-3-5.

Wichowsky, Amber, Paru Shah, and Amanda Heideman. “Call and Response? Neighborhood Inequality and Political Voice.” *Urban Affairs Review* 58, no. 4 (July 1, 2022): 1182–97. https://doi.org/10.1177/10780874211021327.

World Health Organization. “Lead Poisoning.” World Health Organization, September 27, 2024. https://www.who.int/news-room/fact-sheets/detail/lead-poisoning-and-health.

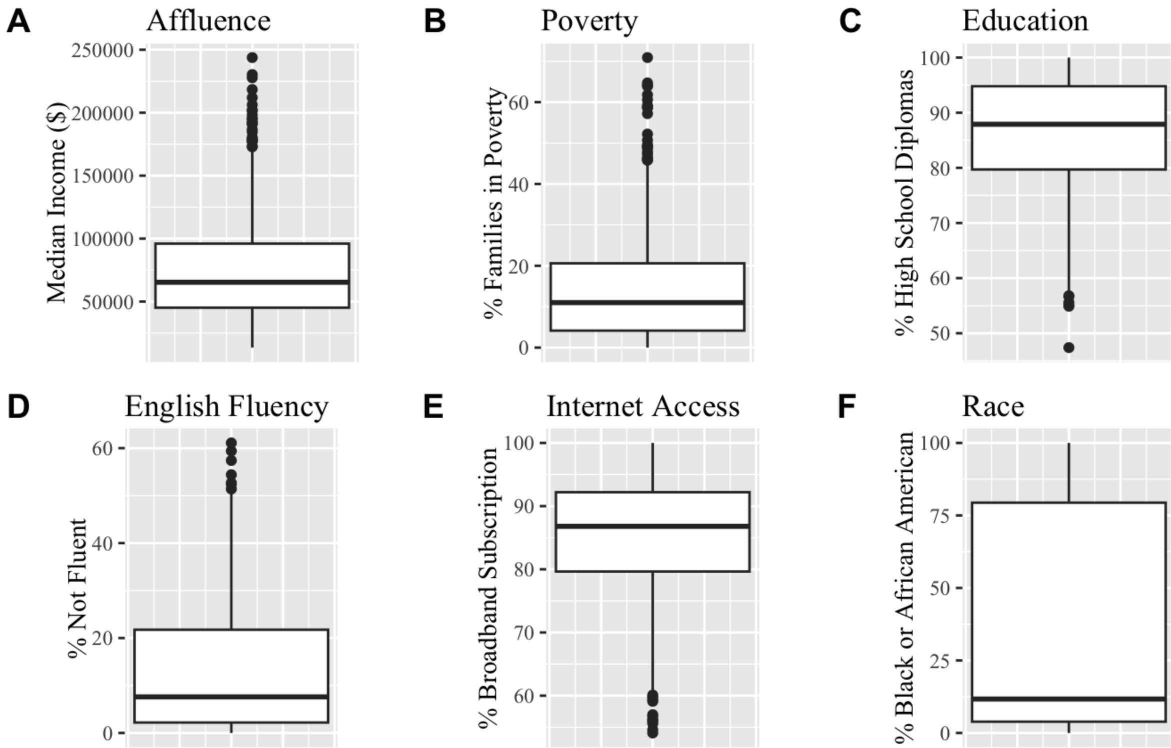
Wright, John Paul, Kim N. Dietrich, M. Douglas Ris, Richard W. Hornung, Stephanie D. Wessel, Bruce P. Lanphear, Mona Ho, and Mary N. Rae. “Association of Prenatal and Childhood Blood Lead Concentrations with Criminal Arrests in Early Adulthood.” *PLoS Medicine* 5, no. 5 (May 27, 2008): e101. https://doi.org/10.1371/journal.pmed.0050101.

Wyman, Oliver. “Females Are Discontent but Darn Proactive About Their Health and Healthcare.” Accessed April 7, 2025. https://www.oliverwyman.com/our-expertise/perspectives/health/2019/mar/females-are-unhappy-but-darn-proactive-about-their-health-and-he.html.

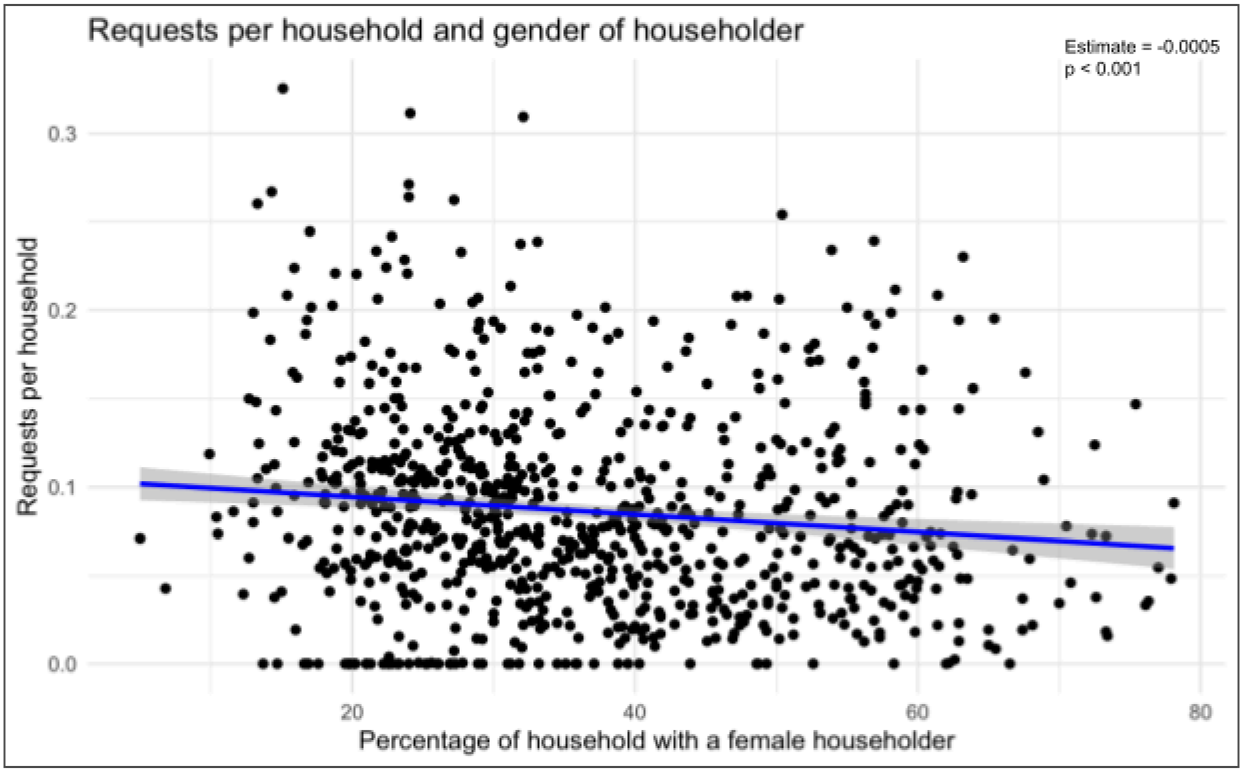
Xu, Corey Kewei, and Tian Tang. “Closing the Gap or Widening the Divide: The Impacts of Technology-Enabled Coproduction on Equity in Public Service Delivery.” *Public Administration Review* 80, no. 6 (2020): 962–75. https://doi.org/10.1111/puar.13222.

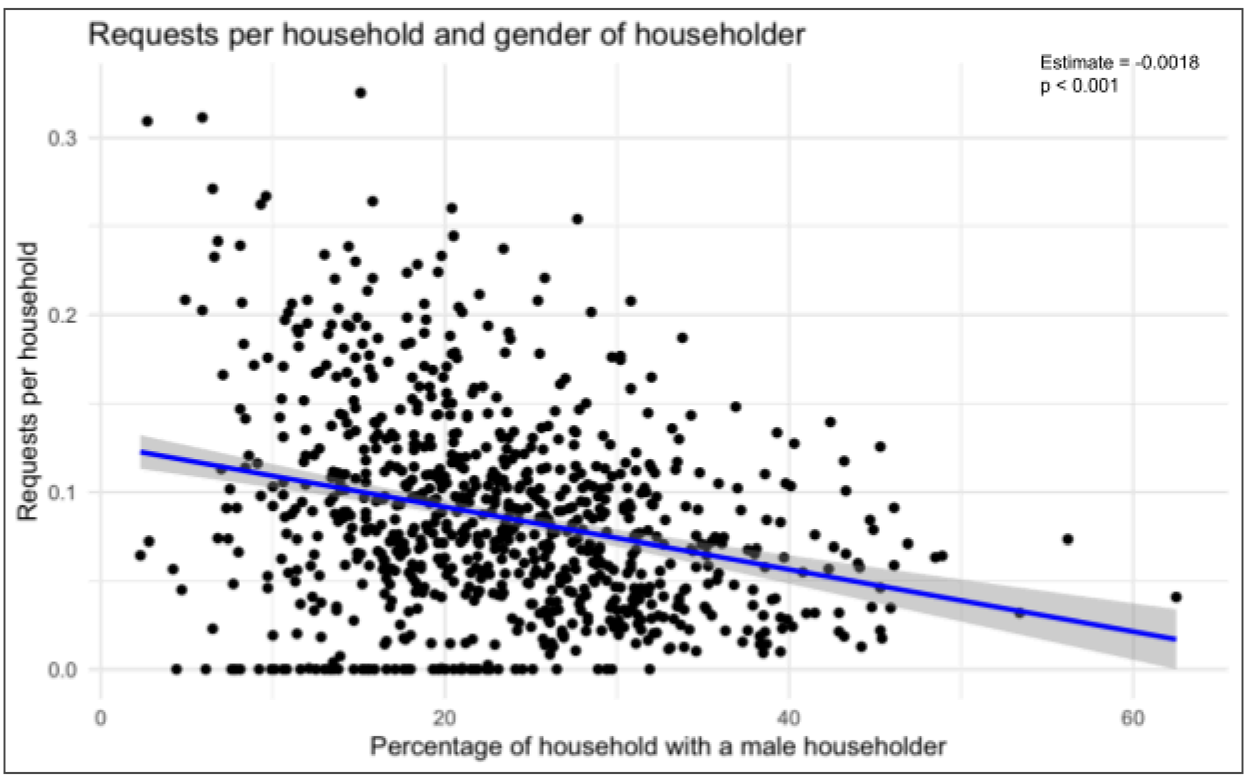
Zhang, Ci, Deborah Koniak-Griffin, Han-Zhu Qian, Lloyd A. Goldsamt, Honghong Wang, Mary-Lynn Brecht, and Xianhong Li. “Impact of Providing Free HIV Self-Testing Kits on Frequency of Testing among Men Who Have Sex with Men and Their Sexual Partners in China: A Randomized Controlled Trial.” Edited by T Charles Witzel. *PLOS Medicine* 17, no. 10 (October 9, 2020): e1003365. https://doi.org/10.1371/journal.pmed.1003365.

Appendix A: Descriptive Statistics



__Appendix B: Gender of head of household__





Appendix C: Code

```r
# Pivot tables for request dataset

# Pivot table for SR_TYPE
requesttypes <- PivotTable$new()
requesttypes$addData(requests)
requesttypes$addColumnDataGroups("SR_TYPE")
requesttypes$defineCalculation(calculationName = "total", summariseExpression
= "n()")
requesttypes$renderPivot()

# Pivot table for CREATED_DEPARTMENT (department that created service
request)
department <- PivotTable$new()
department$addData(requests)
department$addColumnDataGroups("CREATED_DEPARTMENT")
department$defineCalculation(calculationName = "total", summariseExpression =
"n()")
department$renderPivot()

# Pivot table for STATUS
status <- PivotTable$new()
status$addData(requests)
status$addColumnDataGroups("STATUS")
status$defineCalculation(calculationName = "total", summariseExpression =
"n()")
status$renderPivot()

# Pivot table for ORIGIN
origin <- PivotTable$new()
origin$addData(requests)
origin$addColumnDataGroups("ORIGIN")
origin$defineCalculation(calculationName = "total", summariseExpression =
"n()")
origin$renderPivot()

# Pivot table for zipcode (ZIP_CODE)
zipcode <- PivotTable$new()
zipcode$addData(requests)
zipcode$addColumnDataGroups("ZIP_CODE")
zipcode$defineCalculation(calculationName = "total", summariseExpression =
"n()")
zipcode$renderPivot()

# Pivot table for COMMUNITY_AREA
community <- PivotTable$new()
community$addData(requests)
community$addColumnDataGroups("COMMUNITY_AREA")
```

```r
community$defineCalculation(calculationName = "total", summariseExpression =
"n()")
community$renderPivot()

# Merging requests dataset with census data

# Retrieve Illinois shapefiles and Chicago boundary shapefile, and remove IL
census tracts not in Chicago
tracts_US <- st_read("~/Desktop/PBPL
Thesis/nhgis0001_shape/nhgis0001_shapefile_tl2022_us_tract_2022/US_tract_2022
.shx")
tract_IL <- tracts_US[tracts_US$STATEFP == 17,]
Chicago_boundary <- st_read("chipoli/chipoli.shp")
tract_IL <- st_transform(tract_IL, st_crs(Chicago_boundary))
tracts_Chicago <- tract_IL %>%
  st_intersection(Chicago_boundary)

# Examples for requesting census data and mapping
dc_income <- get_acs(
  geography = "tract",
  variables = "B19013_001",
  state = "DC",
  year = 2020,
  geometry = TRUE
)
# plot(dc_income["estimate"])

chicago_income <- get_acs(
  geography = "tract",
  variables = "DP03_0062E",
  state = "Illinois",
  county = "Cook County",
  year = 2022,
  geometry = TRUE
)
chicago_income_filtered <- st_intersection(chicago_income, Chicago_boundary)

# Convert coordinates into geospatial data type
request_points <- requests %>%
  select(SR_NUMBER, LONGITUDE, LATITUDE) %>%
  filter(!is.na(LATITUDE)) %>%
  st_as_sf(coords = c("LONGITUDE", "LATITUDE"), crs = st_crs(tracts_Chicago))

# Assign coordinates to census tracts
request_points_tract<- st_join(request_points, tracts_Chicago)

# Merge census tract information with requests
requests <- merge(requests, request_points_tract, by = "SR_NUMBER")
```

```r
# Clean data (ie look at duplicates)
duplicate_rows <- requests[duplicated(requests$STREET_ADDRESS) |
duplicated(requests$STREET_ADDRESS, fromLast = TRUE), ]
duplicate_counts <- table(requests$STREET_ADDRESS)
frequent_duplicates <- names(duplicate_counts[duplicate_counts >= 10])
duplicate_rows <- requests[requests$STREET_ADDRESS %in% frequent_duplicates,
]
# 1000 E Ohio Street is Chicago Water Quality DWM and 2111 W Lexington St is
311 Operations
# possible system mistakes: 1226 W 97th Pl, 1328 N Claremont Ave, 2214 N
California Ave, 2537 N Bernard St, 2814 N Drake Ave, 5348 S Keeler Ave, 5425
S Bishop St (two seem real), 8015 S State St, 816 W 51st Pl, 856 N Waller Ave
# most others with 10 or more duplicates seem to contain daycare centers,
condos, or doctors offices
requests <- requests[!requests$STREET_ADDRESS %in% c("1226 W 97TH PL", "1328
N CLAREMONT AVE", "2214 N CALIFORNIA AVE", "2537 N BERNARD ST", "2814 N DRAKE
AVE", "5348 S KEELER AVE", "8015 S STATE ST", "816 W 51ST PL", "856 N WALLER
AVE"), ]
glitch <- requests$STREET_ADDRESS == "5425 S BISHOP ST" & requests$STATUS ==
"Canceled"
requests <- requests[!glitch, ] # keep seemingly real ones
requests <- requests[!requests$STREET_ADDRESS %in% c("1000 E OHIO ST", "2111
W Lexington ST"), ]


# New dataframe with counts of unique entries for geo_id
geo_id_request_counts <- requests %>%
  group_by(GEOID) %>%
  summarise(count = n(), .groups = 'drop')  # Count occurrences and drop
grouping
colnames(geo_id_request_counts) <- c("GEO_ID", "Requests")

# Change ACS geo_id to match with requests and limit ACS data to Chicago
tracts
ACS <- ACS %>%
  mutate(GEO_ID = substr(GEO_ID, 10, 20))
Chicago_tract_numbers <- tracts_Chicago$GEOID
ACS <- ACS %>%
  filter(GEO_ID %in% Chicago_tract_numbers)

# Merge requests with ACS data
# ADD ZEROS
df <- merge(geo_id_request_counts, ACS, by = "GEO_ID", all = TRUE)
df[is.na(df)] <- 0
df <- df %>%
  mutate(DP02_0001E = as.integer(DP02_0001E))
df <- df %>%
  filter(DP03_0062E != "-")
df <- df %>%
```

```r
  mutate(DP03_0062E = case_when(
    DP03_0062E == "250,000+" ~ "250,000",
    TRUE ~ DP03_0062E              # Keep original value for all others
  ))
df <- df %>%
  mutate(DP03_0062E = as.double(DP03_0062E))
df <- df %>%
  mutate(DP03_0119PE = as.double(DP03_0119PE))
df <- df %>%
  filter(!is.na(DP03_0062E))
df <- df %>%
  mutate(DP02_0067PE = as.double(DP02_0067PE))
df <- df %>%
  mutate(DP02_0068PE = as.double(DP02_0068PE))
df <- df %>%
  mutate(DP02_0154PE = as.double(DP02_0154PE))
df <- df %>%
  mutate(DP02_0115PE = as.double(DP02_0115PE))
df <- df %>%
  mutate(DP05_0067PE = as.double(DP05_0067PE))
df <- df %>%
  mutate(DP04_0026PE = as.double(DP04_0026PE))
df <- df %>%
  mutate(DP04_0025PE = as.double(DP04_0025PE))
df <- df %>%
  mutate(DP04_0024PE = as.double(DP04_0024PE))
df <- df %>%
  mutate(DP04_0023PE = as.double(DP04_0023PE))
df <- df %>%
  mutate(DP04_0022PE = as.double(DP04_0022PE))
df <- df %>%
  mutate(DP04_0021PE = as.double(DP04_0021PE))
df <- df %>%
  mutate(DP02_0006PE = as.double(DP02_0006PE))
df <- df %>%
  mutate(DP02_0007PE = as.double(DP02_0007PE))
df <- df %>%
  mutate(DP02_0010PE = as.double(DP02_0010PE))
df <- df %>%
  mutate(DP02_0011PE = as.double(DP02_0011PE))
df <- df %>%
  mutate(Requests_per_household = Requests / DP02_0001E)
df <- df %>%
  mutate(DP05_0073PE = as.double(DP05_0073PE))
df$percent_old <- df$DP04_0026PE + df$DP04_0025PE + df$DP04_0024PE +
df$DP04_0023PE + df$DP04_0022PE + df$DP04_0021PE
df <- subset(df, df$GEO_ID != 0)

# Descriptive statistics
# Calculate percentage of each category by location
```

```r
requests_percentage <- requests %>%
  group_by(GEOID, SR_TYPE) %>%
  summarise(count = n(), .groups = "drop") %>%
  group_by(GEOID) %>%
  mutate(total_count = sum(count),
         percentage = (count / total_count) * 100) %>%
  select(-count, -total_count)  # Remove intermediate columns

# Pivot the data to wide format
request_type <- requests_percentage %>%
  pivot_wider(names_from = SR_TYPE, values_from = percentage, values_fill =
list(percentage = 0))
request_type <- request_type %>%
  rename(GEO_ID = GEOID)
request_type <- request_type %>%
  rename(Visit_request = "Water Lead Test Visit Request")

# Combine with census data
request_type <- merge(request_type, df, by = "GEO_ID")

# Remove most recent year of data
requests$date <- substr(requests$CREATED_DATE, 1, 10)
requests$date <- as.Date(requests$date, format = "%m/%d/%Y")
requests_old <- requests %>%
  filter(date < as.Date("2023-10-29"))

# Calculate percentage of each category by location
status_percentage <- requests_old %>%
  group_by(GEOID, STATUS) %>%
  summarise(count = n(), .groups = "drop") %>%
  group_by(GEOID) %>%
  mutate(total_count = sum(count),
         percentage = (count / total_count) * 100) %>%
  select(-count, -total_count)  # Remove intermediate columns

# Pivot the data to wide format
status <- status_percentage %>%
  pivot_wider(names_from = STATUS, values_from = percentage, values_fill =
list(percentage = 0))
status <- status %>%
  rename(GEO_ID = GEOID)

# Combine with census data
status <- merge(status, df, by = "GEO_ID")

# Box plots
boxa <- ggplot(df, aes(y = DP03_0062E)) +
  geom_boxplot() +
  ggtitle("Affluence") +
```

```
  ylab("Median Income ($)") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
boxb <- ggplot(df, aes(y = DP03_0119PE)) +
  geom_boxplot() +
  ggtitle("Poverty") +
  ylab("% Families in Poverty") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
boxc <- ggplot(df, aes(y = DP02_0067PE)) +
  geom_boxplot() +
  ggtitle("Education") +
  ylab("% High School Diplomas") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
boxd <- ggplot(df, aes(y = DP02_0115PE)) +
  geom_boxplot() +
  ggtitle("English Fluency") +
  ylab("% Not Fluent") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
boxe <- ggplot(df, aes(y = DP02_0154PE)) +
  geom_boxplot() +
  ggtitle("Internet Access") +
  ylab("% Broadband Subscription") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
boxf <- ggplot(df, aes(y = DP05_0067PE)) +
  geom_boxplot() +
  ggtitle("Race") +
  ylab("% Black or African American") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
```

```
    text = element_text(family = "Times New Roman"))
# font_import()
plot_grid(boxa, boxb, boxc, boxd, boxe, boxf, labels="AUTO")

boxg <- ggplot(df, aes(y = DP02_0001E)) +
  geom_boxplot() +
  ggtitle("Households") +
  ylab("Total Number Households") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
boxh <- ggplot(df, aes(y = Requests)) +
  geom_boxplot() +
  ggtitle("Requests") +
  ylab("Total Number Requests") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
boxi <- ggplot(df, aes(y = Requests_per_household)) +
  geom_boxplot() +
  ggtitle("Requests per Household") +
  ylab("Requests per Household") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
plot_grid(boxg, boxh, boxi, labels="AUTO")

boxj <- ggplot(request_type, aes(y = Visit_request)) +
  geom_boxplot() +
  ggtitle("Type of Request") +
  ylab("% Requests for Visits") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    text = element_text(family = "Times New Roman"))
boxk <- ggplot(status, aes(y = Completed)) +
  geom_boxplot() +
  ggtitle("Status") +
  ylab("% Requests Completed") +
  theme(aspect.ratio = 1.5,
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
```

```r
    text = element_text(family = "Times New Roman"))

par(family = "Times")
par(mar = c(5, 5, 5, 4))
category_count_department <- table(requests$CREATED_DEPARTMENT)
bara <- barplot(category_count_department, main = "Department that Created
Request", col = "lightblue", las = 2, names.arg = c("311", "Aldermen", "DOT",
"DFSS", "DWM", "S&S"), family = "Times New Roman", ylim = c(0, 15000))
text(bara, category_count_department + 1000 ,
paste(category_count_department, sep="") ,cex=1)



plot_grid(boxj, boxk, labels="AUTO")



requests$CREATED_DATE <- as.POSIXct(requests$CREATED_DATE, format = "%m/%d/%Y
%I:%M:%S %p")
requests$CLOSED_DATE <- as.POSIXct(requests$CLOSED_DATE, format = "%m/%d/%Y
%I:%M:%S %p")
requests$CREATED_DATE <- as.Date(requests$CREATED_DATE)
requests$CLOSED_DATE <- as.Date(requests$CLOSED_DATE)

# Filter data for 'Completed' status and calculate the time difference
completed <- requests %>%
  filter(STATUS == "Completed") %>%
  mutate(time_diff = as.numeric(difftime(CLOSED_DATE, CREATED_DATE, units =
"days")))

# Calculate the average length of time (in days)
average_time <- mean(completed$time_diff, na.rm = TRUE)

# Print the average time
print(average_time)
summary(completed$time_diff)

# Maps
df <- df %>%
  mutate(GEOID = GEO_ID)
merged_geometry_data <- merge(chicago_income_filtered, df, by = "GEOID",
all.x = TRUE)
merged_geometry_data$Requests[is.na(merged_geometry_data$Requests)] <- 0
merged_geometry_data$Requests_per_household[is.na(merged_geometry_data$Reques
ts_per_household)] <- 0
ggplot(merged_geometry_data) +
  geom_sf(aes(fill = DP03_0062E), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", name = "Median Household Income") +
```

```
# Choose a color palette
  labs(title = "Median Household Income") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(family = "Times New Roman"))

ggplot(merged_geometry_data) +
  geom_sf(aes(fill = DP02_0067PE), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", name = "Education") +  # Choose a
color palette
  labs(title = "Education") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(family = "Times New Roman"))

ggplot(merged_geometry_data) +
  geom_sf(aes(fill = DP03_0119PE), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", name = "Poverty") +  # Choose a
color palette
  labs(title = "Poverty") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(family = "Times New Roman"))

ggplot(merged_geometry_data) +
  geom_sf(aes(fill = DP02_0115PE), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", name = "% of population that speaks
English less than well") +  # Choose a color palette
  labs(title = "English Fluency") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(family = "Times New Roman"))

ggplot(merged_geometry_data) +
  geom_sf(aes(fill = DP02_0154PE), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", name = "Internet Access") +  #
Choose a color palette
  labs(title = "Internet Access") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(family = "Times New Roman"))

ggplot(merged_geometry_data) +
  geom_sf(aes(fill = DP05_0067PE), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
```

```
  scale_fill_viridis_c(option = "plasma", name = "Black and African
American") +  # Choose a color palette
  labs(title = "Black and African American") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(family = "Times New Roman"))

ggplot(merged_geometry_data) +
  geom_sf(aes(fill = Requests), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", name = "Requests") +  # Choose a
color palette
  labs(title = "Requests") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(family = "Times New Roman"))

ggplot(merged_geometry_data) +
  geom_sf(aes(fill = Requests_per_household), color = "black", size = 0.001)
+  # Color tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", limits = c(0,0.4), oob =
scales::squish, name = "Requests per household") +  # Crop outliers
  labs(title = "Requests per household") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(family = "Times New Roman"))

# Scatter plot of total number of requests per census tract against median
household income
ggplot(df, aes(x = DP03_0062E, y = Requests)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Total number of requests and median income",
       x = "Median income",
       y = "Total number of requests") +
  theme_minimal()
model <- lm(Requests ~ DP03_0062E, data = df)
summary(model)

# Scatter plot of total number of requests per census tract against median
household income less than 150,000
df_filtered <- df %>%
  filter(DP03_0062E < 150000)
ggplot(df_filtered, aes(x = DP03_0062E, y = Requests)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Total number of requests and median income",
       x = "Median income",
       y = "Total number of requests") +
  theme_minimal()
```

```r
model <- lm(Requests ~ DP03_0062E, data = df_filtered)
summary(model)

# Scatter plot of average requests per household against median household
income
ggplot(df, aes(x = DP03_0062E, y = Requests_per_household)) +
  geom_point() +
   geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Requests per household and median income",
       x = "Median income",
       y = "Requests per household") +
  theme_minimal()
model <- lm(Requests_per_household ~ DP03_0062E, data = df)
summary(model)

# Scatter plot of average requests per household against median household
income less than 150,000
ggplot(df_filtered, aes(x = DP03_0062E, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Requests per household and median income",
       x = "Median income",
       y = "Requests per household") +
  theme_minimal()
model <- lm(Requests_per_household ~ DP03_0062E, data = df_filtered)
summary(model)

# Pull out max outlier
df[which.max(df$Requests_per_household),]
outlier <- requests[requests$GEOID == 17031612100,]
# remove?
outliertest <- PivotTable$new()
outliertest$addData(outlier)
outliertest$addColumnDataGroups("STREET_ADDRESS")
outliertest$defineCalculation(calculationName = "total", summariseExpression
= "n()")
outliertest$renderPivot()

# Scatter plot of total number of requests per census tract against percent
of families in poverty
ggplot(df, aes(x = DP03_0119PE, y = Requests)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Total number of requests and percent poverty",
       x = "Percentage of families in poverty",
       y = "Total number of requests") +
  theme_minimal()
model <- lm(Requests ~ DP03_0119PE, data = df)
summary(model)
```

```r
# Scatter plot of total number of requests per household against percent of
families in poverty
ggplot(df, aes(x = DP03_0119PE, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Average requests per household and percent poverty",
       x = "Percentage of families in poverty",
       y = "Average requests per household") +
  theme_minimal()
model <- lm(Requests_per_household ~ DP03_0119PE, data = df)
summary(model)

# Scatter plot of total number of requests per census tract against percent
of people above 25 with a high school degree or higher
ggplot(df, aes(x = DP02_0067PE, y = Requests)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Total number of requests and education",
       x = "Percentage with high school or above",
       y = "Total number of requests") +
  theme_minimal()
model <- lm(Requests ~ DP02_0067PE, data = df)
summary(model)

# Scatter plot of total number of requests per household against percent of
people above 25 with a high school degree or higher
ggplot(df, aes(x = DP02_0067PE, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Average requests per household and education",
       x = "Percentage with high school or above",
       y = "Average requests per household") +
  theme_minimal()
model <- lm(Requests_per_household ~ DP02_0067PE, data = df)
summary(model)

# Scatter plot of total number of requests per census tract against percent
of households with internet
ggplot(df, aes(x = DP02_0154PE, y = Requests)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Total number of requests and internet access",
       x = "Percentage of households with internet access",
       y = "Total number of requests") +
  theme_minimal()
model <- lm(Requests ~ DP02_0154PE, data = df)
summary(model)

# Scatter plot of total number of requests per household against percent of
households with internet
```

```
ggplot(df, aes(x = DP02_0154PE, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Average requests per household and internet access",
       x = "Percentage of households with internet access",
       y = "Average requests per household") +
  theme_minimal()
model <- lm(Requests_per_household ~ DP02_0154PE, data = df)
summary(model)

# Scatter plot of total number of requests per census tract against percent
of population that speaks English less than very well
ggplot(df, aes(x = DP02_0115PE, y = Requests)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Total number of requests and English proficiency",
       x = "Percentage of population that speaks English less than very
well",
       y = "Total number of requests") +
  theme_minimal()
model <- lm(Requests ~ DP02_0115PE, data = df)
summary(model)

# Scatter plot of total number of requests per household against percent of
population that speaks English less than very well
ggplot(df, aes(x = DP02_0115PE, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Average requests per household and English proficiency",
       x = "Percentage of population that speaks English less than very
well",
       y = "Average requests per household") +
  theme_minimal()
model <- lm(Requests_per_household ~ DP02_0115PE, data = df)
summary(model)

# Scatter plot of total number of requests per census tract against percent
of population that is Black or African American
ggplot(df, aes(x = DP05_0067PE, y = Requests)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Total number of requests and percent of population that is
Black",
       x = "Percentage of population that is Black",
       y = "Total number of requests") +
  theme_minimal()
model <- lm(Requests ~ DP05_0067PE, data = df)
summary(model)

# Scatter plot of total number of requests per household against percent of
```

```
population that is Black or African American
ggplot(df, aes(x = DP05_0067PE, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Average requests per household and percent of population that
is Black",
       x = "Percentage of population that is Black",
       y = "Average requests per household") +
  theme_minimal()
model <- lm(Requests_per_household ~ DP05_0067PE, data = df)
summary(model)

# Scatter plot of total number of requests per census tract against percent
of population that is Hispanic
ggplot(df, aes(x = DP05_0073PE, y = Requests)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Total number of requests and percent of population that is
Hispanic",
       x = "Percentage of population that is Hispanic",
       y = "Total number of requests") +
  theme_minimal()
model <- lm(Requests ~ DP05_0073PE, data = df)
summary(model)

# Scatter plot of total number of requests per household against percent of
population that is Hispanic
ggplot(df, aes(x = DP05_0073PE, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Average requests per household and percent of population that
is Hispanic",
       x = "Percentage of population that is Hispanic",
       y = "Average requests per household") +
  theme_minimal()
model <- lm(Requests_per_household ~ DP05_0073PE, data = df)
summary(model)

# Controlling for variables, ie including multiple variables in model and
analyzing individual coefficients


# Add concentration of lead/risk as explanatory variable

# Housing age as a proxy
boxplot(df$percent_old, main="Housing Age", ylab="% Houses Built Before
1989", col="lightblue")
ggplot(merged_geometry_data) +
  geom_sf(aes(fill = percent_old), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
```

```
  scale_fill_viridis_c(option = "plasma", name = "% Houses Built Before 
1989") +  # Choose a color palette
  labs(title = "Housing Age") +
  theme_minimal() +
  theme(legend.position = "right")

risk <- lm(formula = Requests_per_household ~ percent_old, data = df)
summary(risk)
ggplot(df, aes(x = percent_old, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Requests per household and percent of houses built before 
1989",
       x = "Percentage of houses built before 1989",
       y = "Requests per household") +
  theme_minimal()

model <- lm(formula = Requests_per_household ~ DP03_0062E + DP03_0119PE + 
DP02_0067PE + DP02_0154PE + DP02_0115PE + DP05_0067PE + DP05_0073PE + 
percent_old, data = df)
summary(model)

# Only include the significant explanatory variables

model_simple <- lm(formula = Requests_per_household ~ DP03_0062E + 
DP03_0119PE + DP05_0067PE + percent_old, data = df)
summary(model_simple)



# Model residuals
model_residuals = model_simple$residuals
hist(model_residuals)
qqnorm(model_residuals)
qqline(model_residuals) # right skewed

# Multicollinearity assumption check
reduced <- df[c("DP03_0062E", "DP03_0119PE", "DP05_0067PE", "percent_old")]
corr_matrix = round(cor(reduced), 2)
ggcorrplot(corr_matrix, hc.order = TRUE, type = "lower", lab = TRUE)

# Gini importance index
gini_df <- df[c("Requests_per_household", "DP03_0062E", "DP03_0119PE", 
"DP02_0067PE", "DP02_0154PE", "DP02_0115PE", "DP05_0067PE", "DP05_0073PE")]
rf_model <- randomForest(Requests_per_household ~ ., data = gini_df, 
importance = TRUE)
print(rf_model)
importance(rf_model)
varImpPlot(rf_model, type = 2)
```

```r
# Investigation into differences in requests for test kits and test visits

# Calculate percentage of each category by location
requests_percentage <- requests %>%
  group_by(GEOID, SR_TYPE) %>%
  summarise(count = n(), .groups = "drop") %>%
  group_by(GEOID) %>%
  mutate(total_count = sum(count),
         percentage = (count / total_count) * 100) %>%
  select(-count, -total_count)  # Remove intermediate columns

# Pivot the data to wide format
request_type <- requests_percentage %>%
  pivot_wider(names_from = SR_TYPE, values_from = percentage, values_fill =
list(percentage = 0))
request_type <- request_type %>%
  rename(GEO_ID = GEOID)
request_type <- request_type %>%
  rename(Visit_request = "Water Lead Test Visit Request")

# Combine with census data
request_type <- merge(request_type, df, by = "GEO_ID")

# Comparisons with demographic data
ggplot(request_type, aes(x = DP03_0062E, y = Visit_request)) +
  geom_point() +
   geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests for visits and median income",
       x = "Median income",
       y = "Proportion of requests for visits") +
  theme_minimal()
model <- lm(Visit_request ~ DP03_0062E, data = request_type)
summary(model)

ggplot(request_type, aes(x = DP03_0119PE, y = Visit_request)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests for visits and percent poverty",
       x = "Percentage of families in poverty",
       y = "Proportion of requests for visits") +
  theme_minimal()
model <- lm(Visit_request ~ DP03_0119PE, data = request_type)
summary(model)

ggplot(request_type, aes(x = DP02_0067PE, y = Visit_request)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests for visits and education",
       x = "Percentage with high school or above",
```

```
       y = "Proportion of requests for visits") +
  theme_minimal()
model <- lm(Visit_request ~ DP02_0067PE, data = request_type)
summary(model)

ggplot(request_type, aes(x = DP02_0154PE, y = Visit_request)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests for visits and internet access",
       x = "Percentage of households with internet access",
       y = "Proportion of requests for visits") +
  theme_minimal()
model <- lm(Visit_request ~ DP02_0154PE, data = request_type)
summary(model)

ggplot(request_type, aes(x = DP02_0115PE, y = Visit_request)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests for visits and English proficiency",
       x = "Percentage of population that speaks English less than very
well",
       y = "Proportion of requests for visits") +
  theme_minimal()
model <- lm(Visit_request ~ DP02_0115PE, data = request_type)
summary(model)

ggplot(request_type, aes(x = DP05_0067PE, y = Visit_request)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests for visits and percent of population
that is Black",
       x = "Percentage of population that is Black",
       y = "Proportion of requests for visits") +
  theme_minimal()
model <- lm(Visit_request ~ DP05_0067PE, data = request_type)
summary(model)

ggplot(request_type, aes(x = DP05_0073PE, y = Visit_request)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests for visits and percent of population
that is Hispanic",
       x = "Percentage of population that is Hispanic",
       y = "Proportion of requests for visits") +
  theme_minimal()
model <- lm(Visit_request ~ DP05_0073PE, data = request_type)
summary(model)

# plot_grid(scattera, scatterb, scatterc, scatterd, scattere, scatterf,
```

```r
labels="AUTO")

model <- lm(formula = Visit_request ~ DP03_0062E + DP03_0119PE + DP02_0067PE
+ DP02_0154PE + DP02_0115PE + DP05_0067PE + DP05_0073PE, data = request_type)
summary(model)

simplified_model <- lm(formula = Visit_request ~ DP02_0154PE + DP05_0067PE +
DP05_0073PE, data = request_type)
summary(simplified_model)

gini_df <- request_type[c("Visit_request", "DP02_0154PE", "DP05_0067PE",
"DP05_0073PE")]
rf_model <- randomForest(Visit_request ~ ., data = gini_df, importance =
TRUE)
print(rf_model)
importance(rf_model)
varImpPlot(rf_model, type = 2)

# Investigation into status of requests, are the requests that are cancelled
or left open different from those that are completed

# Remove most recent year of data
#requests$date <- substr(requests$CREATED_DATE, 1, 10)
#requests$date <- as.Date(requests$date, format = "%Y/%m/%d")
requests_old <- requests %>%
  filter(CREATED_DATE < as.Date("2023-10-29"))

# Calculate percentage of each category by location
status_percentage <- requests_old %>%
  group_by(GEOID, STATUS) %>%
  summarise(count = n(), .groups = "drop") %>%
  group_by(GEOID) %>%
  mutate(total_count = sum(count),
         percentage = (count / total_count) * 100) %>%
  select(-count, -total_count)  # Remove intermediate columns

# Pivot the data to wide format
status <- status_percentage %>%
  pivot_wider(names_from = STATUS, values_from = percentage, values_fill =
list(percentage = 0))
status <- status %>%
  rename(GEO_ID = GEOID)

# Combine with census data
status <- merge(status, df, by = "GEO_ID")

# Comparisons with demographic data
ggplot(status, aes(x = DP03_0062E, y = Completed)) +
  geom_point() +
```

```r
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests completed and median income",
       x = "Median income",
       y = "Proportion of requests completed") +
  theme_minimal()
model <- lm(Completed ~ DP03_0062E, data = status)
summary(model)

ggplot(status, aes(x = DP03_0119PE, y = Completed)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests completed and percent poverty",
       x = "Percentage of families in poverty",
       y = "Proportion of requests completed") +
  theme_minimal()
model <- lm(Completed ~ DP03_0119PE, data = status)
summary(model)

ggplot(status, aes(x = DP02_0067PE, y = Completed)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests completed and education",
       x = "Percentage with high school or above",
       y = "Proportion of requests completed") +
  theme_minimal()
model <- lm(Completed ~ DP02_0067PE, data = status)
summary(model)

ggplot(status, aes(x = DP02_0154PE, y = Completed)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests completed and internet access",
       x = "Percentage of households with internet access",
       y = "Proportion of requests completed") +
  theme_minimal()
model <- lm(Completed ~ DP02_0154PE, data = status)
summary(model)

ggplot(status, aes(x = DP02_0115PE, y = Completed)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests completed and English proficiency",
       x = "Percentage of population that speaks English less than very
well",
       y = "Proportion of requests completed") +
  theme_minimal()
model <- lm(Completed ~ DP02_0115PE, data = status)
summary(model)
```

```r
ggplot(status, aes(x = DP05_0067PE, y = Completed)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests completed and percent of population
that is Black",
       x = "Percentage of population that is Black",
       y = "Proportion of requests completed") +
  theme_minimal()
model <- lm(Completed ~ DP05_0067PE, data = status)
summary(model)

ggplot(status, aes(x = DP05_0073PE, y = Completed)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests completed and percent of population
that is Hispanic",
       x = "Percentage of population that is Hispanic",
       y = "Proportion of requests completed") +
  theme_minimal()
model <- lm(Completed ~ DP05_0073PE, data = status)
summary(model)

model <- lm(formula = Completed ~ DP03_0062E + DP03_0119PE + DP02_0067PE +
DP02_0154PE + DP02_0115PE + DP05_0067PE + DP05_0073PE, data = status)
summary(model)

model_simple <- lm(formula = Completed ~ DP03_0119PE + DP05_0067PE +
DP05_0073PE, data = status)
summary(model_simple)

gini_df <- status[c("Completed", "DP03_0062E", "DP03_0119PE", "DP02_0067PE",
"DP02_0154PE", "DP02_0115PE", "DP05_0067PE", "DP05_0073PD")]
rf_model <- randomForest(Completed ~ ., data = gini_df, importance = TRUE)
print(rf_model)
importance(rf_model)
varImpPlot(rf_model, type = 2)

# Investigation into origin, how are people filing requests

# Calculate percentage of each category by location
origins_percentage <- requests %>%
  group_by(GEOID, ORIGIN) %>%
  summarise(count = n(), .groups = "drop") %>%
  group_by(GEOID) %>%
  mutate(total_count = sum(count),
         percentage = (count / total_count) * 100) %>%
  select(-count, -total_count)  # Remove intermediate columns

# Pivot the data to wide format
```

```
request_origin <- origins_percentage %>%
  pivot_wider(names_from = ORIGIN, values_from = percentage, values_fill =
list(percentage = 0))
request_origin <- request_origin %>%
  rename(GEO_ID = GEOID,
         Alderman = "Alderman's Office",
         Mobile = "Mobile Device",
         Call = "Phone Call",
         Salesforce = "Salesforce Mobile App",
         Department = "City Department",
         In_house = "Generated In House")

# Combine with census data
request_origin <- merge(request_origin, df, by = "GEO_ID")

# Comparisons with demographic data
ggplot(request_origin, aes(x = DP03_0062E, y = Internet)) +
  geom_point() +
   geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests through the internet and median
income",
       x = "Median income",
       y = "Proportion of requests through the internet") +
  theme_minimal()
model <- lm(Internet ~ DP03_0062E, data = request_origin)
summary(model)

ggplot(request_origin, aes(x = DP03_0119PE, y = Internet)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests through the internet and percent
poverty",
       x = "Percentage of families in poverty",
       y = "Proportion of requests through the internet") +
  theme_minimal()
model <- lm(Internet ~ DP03_0119PE, data = request_origin)
summary(model)

ggplot(request_origin, aes(x = DP02_0067PE, y = Internet)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests through the internet and education",
       x = "Percentage with high school or above",
       y = "Proportion of requests through the internet") +
  theme_minimal()
model <- lm(Internet ~ DP02_0067PE, data = request_origin)
summary(model)

ggplot(request_origin, aes(x = DP02_0154PE, y = Internet)) +
```

```
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests through the internet and internet
access",
       x = "Percentage of households with internet access",
       y = "Proportion of requests through the internet") +
  theme_minimal()
model <- lm(Internet ~ DP02_0154PE, data = request_origin)
summary(model)

ggplot(request_origin, aes(x = DP02_0115PE, y = Internet)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests through the internet and English
proficiency",
       x = "Percentage of population that speaks English less than very
well",
       y = "Proportion of requests through the internet") +
  theme_minimal()
model <- lm(Internet ~ DP02_0115PE, data = request_origin)
summary(model)

ggplot(request_origin, aes(x = DP05_0067PE, y = Internet)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests through the internet and percent of
population that is Black",
       x = "Percentage of population that is Black",
       y = "Proportion of requests through the internet") +
  theme_minimal()
model <- lm(Internet ~ DP05_0067PE, data = request_origin)
summary(model)

ggplot(request_origin, aes(x = DP05_0073PE, y = Internet)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Proportion of requests through the internet and percent of
population that is Hispanic",
       x = "Percentage of population that is Hispanic",
       y = "Proportion of requests through the internet") +
  theme_minimal()
model <- lm(Internet ~ DP05_0073PE, data = request_origin)
summary(model)

model <- lm(formula = Internet ~ DP03_0062E + DP03_0119PE + DP02_0067PE +
DP02_0154PE + DP02_0115PE + DP05_0067PE + DP05_0073PE, data = request_origin)
summary(model)

simplified_model <- lm(formula = Internet ~ DP02_0154PE + DP05_0067PE +
```

```
DP05_0073PE, data = request_origin)
summary(simplified_model)

gini_df <- request_origin[c("Internet", "DP03_0062E", "DP03_0119PE",
"DP02_0067PE", "DP02_0154PE", "DP02_0115PE", "DP05_0067PE", "DP05_0073PE")]
rf_model <- randomForest(Internet ~ ., data = gini_df, importance = TRUE)
print(rf_model)
importance(rf_model)
varImpPlot(rf_model, type = 2)

# Investigation into time, has the pattern of who is requesting tests changed
over time (could indicate becoming more well known)

# Create a year column in the original request dataframe
requests$year <- year(requests$CREATED_DATE)

# New dataframe with counts of unique entries for geo_id
year_request_counts <- requests %>%
  group_by(GEOID, year) %>%
  summarise(count = n(), .groups = 'drop')  # Count occurrences and drop
grouping
colnames(year_request_counts) <- c("GEO_ID", "Year", "Requests")

# Merge requests with ACS data
df_years <- merge(year_request_counts, ACS, by = "GEO_ID")
df_years <- df_years %>%
  mutate(DP02_0001E = as.integer(DP02_0001E))
df_years <- df_years %>%
  filter(DP03_0062E != "-")
df_years <- df_years %>%
  mutate(DP03_0062E = case_when(
    DP03_0062E == "250,000+" ~ "250,000",
    TRUE ~ DP03_0062E              # Keep original value for all others
  ))
df_years <- df_years %>%
  mutate(DP03_0062E = as.double(DP03_0062E))
df_years <- df_years %>%
  mutate(DP03_0119PE = as.double(DP03_0119PE))
df_years <- df_years %>%
  filter(!is.na(DP03_0062E))
df_years <- df_years %>%
  mutate(DP02_0067PE = as.double(DP02_0067PE))
df_years <- df_years %>%
  mutate(DP02_0068PE = as.double(DP02_0068PE))
df_years <- df_years %>%
  mutate(DP02_0154PE = as.double(DP02_0154PE))
df_years <- df_years %>%
  mutate(DP02_0115PE = as.double(DP02_0115PE))
df_years <- df_years %>%
  mutate(DP05_0067PE = as.double(DP05_0067PE))
```

```
df_years <- df_years %>%
  mutate(DP04_0026PE = as.double(DP04_0026PE))
df_years <- df_years %>%
  mutate(DP04_0025PE = as.double(DP04_0025PE))
df_years <- df_years %>%
  mutate(DP04_0024PE = as.double(DP04_0024PE))
df_years <- df_years %>%
  mutate(DP04_0023PE = as.double(DP04_0023PE))
df_years <- df_years %>%
  mutate(DP04_0022PE = as.double(DP04_0022PE))
df_years <- df_years %>%
  mutate(DP04_0021PE = as.double(DP04_0021PE))
df_years <- df_years %>%
  mutate(DP02_0006PE = as.double(DP02_0006PE))
df_years <- df_years %>%
  mutate(DP02_0007PE = as.double(DP02_0007PE))
df_years <- df_years %>%
  mutate(DP02_0010PE = as.double(DP02_0010PE))
df_years <- df_years %>%
  mutate(DP02_0011PE = as.double(DP02_0011PE))
df_years <- df_years %>%
  mutate(DP05_0073PE = as.double(DP05_0073PE))
df_years <- df_years %>%
  mutate(Requests_per_household = Requests / DP02_0001E)
df_years$percent_old <- df_years$DP04_0026PE + df_years$DP04_0025PE +
df_years$DP04_0024PE + df_years$DP04_0023PE + df_years$DP04_0022PE +
df_years$DP04_0021PE

coefficients_list <- list()
for (i in unique(df_years$Year)) {
  subset_data <- subset(df_years, Year == i)
  model <- lm(Requests_per_household ~ DP03_0062E + DP03_0119PE + DP05_0067PE
+ percent_old, data = subset_data)
  coefficients_list[[as.character(i)]] <- coef(model)
}
coefficients_df <- as.data.frame(coefficients_list)
coefficients_df <- t(coefficients_df)
coefficients_df <- as.data.frame(coefficients_df)
colnames(coefficients_df) <- c("Intercept", "DP03_0062E", "DP03_0119PE",
"DP05_0067PE", "percent_old")
coefficients_df$Year <- c("2019", "2020", "2023", "2024", "2021", "2022",
"2018")
coefficients_df$Year <- as.numeric(coefficients_df$Year)

ggplot(coefficients_df, aes(x = Year, y = Intercept)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Intercept over Time",
       x = "Year", y = "Intercept")
ggplot(coefficients_df, aes(x = Year, y = DP03_0062E)) +
  geom_line(color = "blue", size = 1) +
```

```
  labs(title = "Median Income Coefficients over Time",
       x = "Year", y = "Median Income")
ggplot(coefficients_df, aes(x = Year, y = DP03_0119PE)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Percent Poverty Coefficients over Time",
       x = "Year", y = "Percent Poverty")
ggplot(coefficients_df, aes(x = Year, y = DP05_0067PE)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Black/AA Coefficients over Time",
       x = "Year", y = "Black/AA")
ggplot(coefficients_df, aes(x = Year, y = percent_old)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Proxy Risk Coefficients over Time",
       x = "Year", y = "Risk")


# New dataframe with counts of unique entries for geo_id
day_request_counts <- requests %>%
  group_by(GEOID, CREATED_DATE) %>%
  summarise(count = n(), .groups = 'drop')  # Count occurrences and drop
grouping
colnames(day_request_counts) <- c("GEO_ID", "Day", "Requests")

# Merge requests with ACS data
df_days <- merge(day_request_counts, ACS, by = "GEO_ID", all = TRUE)
df_days <- df_days %>%
  mutate(DP02_0001E = as.integer(DP02_0001E))
df_days <- df_days %>%
  filter(DP03_0062E != "-")
df_days <- df_days %>%
  mutate(DP03_0062E = case_when(
    DP03_0062E == "250,000+" ~ "250,000",
    TRUE ~ DP03_0062E                # Keep original value for all others
  ))
df_days <- df_days %>%
  mutate(DP03_0062E = as.double(DP03_0062E))
df_days <- df_days %>%
  mutate(DP03_0119PE = as.double(DP03_0119PE))
df_days <- df_days %>%
  filter(!is.na(DP03_0062E))
df_days <- df_days %>%
  mutate(DP02_0067PE = as.double(DP02_0067PE))
df_days <- df_days %>%
  mutate(DP02_0068PE = as.double(DP02_0068PE))
df_days <- df_days %>%
  mutate(DP02_0154PE = as.double(DP02_0154PE))
df_days <- df_days %>%
  mutate(DP02_0115PE = as.double(DP02_0115PE))
df_days <- df_days %>%
```

```
  mutate(DP05_0067PE = as.double(DP05_0067PE))
df_days <- df_days %>%
  mutate(DP04_0026PE = as.double(DP04_0026PE))
df_days <- df_days %>%
  mutate(DP04_0025PE = as.double(DP04_0025PE))
df_days <- df_days %>%
  mutate(DP04_0024PE = as.double(DP04_0024PE))
df_days <- df_days %>%
  mutate(DP04_0023PE = as.double(DP04_0023PE))
df_days <- df_days %>%
  mutate(DP04_0022PE = as.double(DP04_0022PE))
df_days <- df_days %>%
  mutate(DP04_0021PE = as.double(DP04_0021PE))
df_days <- df_days %>%
  mutate(DP02_0006PE = as.double(DP02_0006PE))
df_days <- df_days %>%
  mutate(DP02_0007PE = as.double(DP02_0007PE))
df_days <- df_days %>%
  mutate(DP02_0010PE = as.double(DP02_0010PE))
df_days <- df_days %>%
  mutate(DP02_0011PE = as.double(DP02_0011PE))
df_days <- df_days %>%
  mutate(DP05_0073PE = as.double(DP05_0073PE))
df_days <- df_days %>%
  mutate(Requests_per_household = Requests / DP02_0001E)
df_days$percent_old <- df_days$DP04_0026PE + df_days$DP04_0025PE +
df_days$DP04_0024PE + df_days$DP04_0023PE + df_days$DP04_0022PE +
df_days$DP04_0021PE

summarized_df <- day_request_counts %>%
  group_by(Day) %>%
  summarise(total_requests = sum(Requests), .groups = 'drop')

ggplot(summarized_df, aes(x = Day, y = total_requests)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Requests Over Time", x = "Date", y = "Requests")

# Spikes in July 2019
requests_July <- requests %>%
  filter(year(CREATED_DATE) == 2019, month(CREATED_DATE) == 7)

requests_spike <- requests[requests$CREATED_DATE == '2019-07-09' |
requests$CREATED_DATE == '2019-07-10' | requests$CREATED_DATE ==
'2019-07-11',]
category_count_ward <- table(requests_spike$WARD)
barplot(category_count_ward, main = "Wards", col = "lightblue", las = 2,
family = "Times New Roman", ylim = c(0, 200))

# Add gender of head of household as a variable
boxplot(df$DP02_0010PE, main="Female householders", ylab="% Households with a
```

```
female householder", col="lightblue")
ggplot(merged_geometry_data) +
  geom_sf(aes(fill = DP02_0010PE), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", name = "% Households with a female
householder") +  # Choose a color palette
  labs(title = "Female householders") +
  theme_minimal() +
  theme(legend.position = "right")

boxplot(df$DP02_0006PE, main="Male householders", ylab="% Households with a
male householder", col="lightblue")
ggplot(merged_geometry_data) +
  geom_sf(aes(fill = DP02_0006PE), color = "black", size = 0.001) +  # Color
tracts by the 'value' column
  scale_fill_viridis_c(option = "plasma", name = "% Households with a male
householder") +  # Choose a color palette
  labs(title = "Male householders") +
  theme_minimal() +
  theme(legend.position = "right")

# Female householder, no spouse/partner present
gender <- lm(formula = Requests_per_household ~ DP02_0010PE, data = df)
summary(gender)
ggplot(df, aes(x = DP02_0010PE, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Requests per household and gender of householder",
       x = "Percentage of household with a female householder",
       y = "Requests per household") +
  theme_minimal()

model <- lm(formula = Requests_per_household ~ DP03_0062E + DP03_0119PE +
DP02_0067PE + DP02_0154PE + DP02_0115PE + DP05_0067PE + DP02_0010PE +
DP05_0073PE + percent_old, data = df)
summary(model)

model_simple <- lm(formula = Requests_per_household ~ DP03_0062E +
DP03_0119PE + DP05_0067PE + DP02_0010PE + percent_old, data = df)
summary(model_simple)

# Male householder, no spouse/partner present

gender <- lm(formula = Requests_per_household ~ DP02_0006PE, data = df)
summary(gender)
ggplot(df, aes(x = DP02_0006PE, y = Requests_per_household)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(title = "Requests per household and gender of householder",
```

```
      x = "Percentage of household with a male householder",
      y = "Requests per household") +
  theme_minimal()

model <- lm(formula = Requests_per_household ~ DP03_0062E + DP03_0119PE +
DP02_0067PE + DP02_0154PE + DP02_0115PE + DP05_0067PE + DP02_0006PE +
DP05_0073PE + percent_old, data = df)
summary(model)

model_simple <- lm(formula = Requests_per_household ~ DP03_0062E +
DP03_0119PE + DP05_0067PE + DP02_0006PE + percent_old, data = df)
summary(model_simple)
```