Who Benefits from Surface Water Pollution Programs?

Qianping Ren[a] Jeremy West[a]*

[a]University of California, Santa Cruz

January 2023

Abstract

The Clean Water Act provides around 180 million dollars annually for nonpoint source pollution programs. We examine how state governments deploy these funds. We find that watersheds with wealthier and more white populations are less polluted but more likely to receive pollution project funding. Supporting analyses show that these disparities can be explained by spatial differences in local government capacity. Socioeconomically disadvantaged areas have fewer resources to compete for grants and constrained match funding for grant proposal requirements. Our findings suggest that a competitive application process is an inequitable way to determine environmental funding priorities and could amplify justice concerns.

JEL: D63, H72, Q53, Q56, R51

Keywords: nonpoint source pollution, public finance, environmental justice

*Ren: qren4@ucsc.edu. West (corresponding author): westj@ucsc.edu. We thank Alex Porteous for helpful discussions about the EPA's Grants Reporting and Tracking System. We thank seminar participants at UC-Santa Cruz for valuable comments. The conclusions and any errors in this work are those of the authors. This research did not receive any specific grant from funding agencies.

1 Introduction

Nonpoint source pollution is the leading cause of water quality problems in the United States (EPA, 2011). NPS pollutants—including oil, sediment, fertilizer and other agricultural and urban runoff—affect the quality of drinking water supplies and harm fisheries and wildlife. Surface water pollution also reduces the recreation and amenity value of contaminated resources (Kuwayama et al., 2022). To address water pollution from known (point) sources, policymakers usually utilize regulations and wastewater treatment facilities (Keiser and Shapiro, 2019a). In contrast, the diffuse nature of nonpoint sources makes NPS water pollution a considerably more complicated challenge (Chambers and Quiggin, 1996).

Motivated by this concern, Congress amended the Clean Water Act in 1987 to establish federal funding for nonpoint source pollution management. The Section 319(h) program allocates funds annually to the designated water quality agency in each state, as well as to tribes and U.S. territories. Since 1990, more than five billion dollars of federal grants have been provided; in 2022, the total was \$178 million. Congress determines total Section 319 funding, then provides each state a lump sum amount based on fixed proportions. For example, California always receives 5.34 percent, Texas receives 4.75 percent, and so on (Figure A1 shows the full distribution). These funds are used by state and local governments for NPS projects, often in combination with funding from other sources.[1]

States have broad discretion over Section 319 funds, and scope for potential projects exceeds available resources, such that funding decisions entail an opportunity cost. To set funding priorities, states ask local governments to submit project proposals (EPA, 2011). This competitive funding process could help states to implement projects where they are most beneficial, especially if local governments have better information about constituent preferences (Oates, 1999). However, this process might favor wealthier communities. Federal Section 319 funding requires at least 40 percent match funding, and most states pass through this requirement. Funding from local sources also helps in the proposal ranking process. Local governments vary greatly in their resources and capacity to successfully write grants (Gargan, 1981). Communities also vary in political influence (Becker, 1983). If these nonenvironmental factors play a large enough role, then socioeconomically advantaged communities may receive an outsized share of the funding.

Our study tests this hypothesis, using the spatial distribution of funded projects to

[1]Some examples of project deliverables are sediment retention basins, riparian fencing for livestock, and treatments for algae blooms. Section 319 grants are also used to develop Watershed Restoration Action Strategies and Total Maximum Daily Load planning needed to obtain larger state and federal grants.

evaluate racial and socioeconomic disparities in grant awards. Our analysis level is a "subwatershed," the smallest comprehensive hydrologic unit of the U.S. Geological Survey (there are about 103,000 subwatersheds, more than twice the number of ZIP codes). Measuring relationships at a granular level helps to avoid attenuation bias from the "ecological fallacy" (Banzhaf et al., 2019). Using EPA data, we determine whether each subwatershed is included in new Section 319 funding awards annually. We match race and socioeconomic characteristics for subwatershed populations using decennial censuses and American Community Surveys, spatially joined at the (most granular) Census Block level. After dropping unpopulated areas, our study sample contains over 88,000 subwatersheds spanning 1990-2020, with more than 74,000 project funding awards at the subwatershed-year level.

We find evidence of significant disparities in grant funding rates. For instance, the average rate of annual funding inclusion is 0.59 percent for subwatersheds with less than 10 percent white population, compared to three percent for majority-white subwatersheds. Only one percent of subwatersheds with income per capita in the bottom quartile are funded per award cycle, while areas with income in the top quartile have a funding rate of almost 4.5 percent. Using home values, the annual funding rate is 1.35 percent for subwatersheds in the bottom quartile but over four percent for areas in the top quartile.

This evidence is bolstered by regression analysis. Including statistical controls is controversial in the environmental justice literature, because conditioning on covariates may obscure the descriptive relationships of primary interest (Banzhaf et al., 2019). However, it seems sensible to control at least for state-by-year fixed effects, the statutory level of decision-making in funding allocations. Our most saturated specifications also control for subwatershed population and rural status. These linear regressions show that a ten percentage points increase in white population share is associated with 0.1 percentage points increase in the likelihood of receiving funding. At the sample average funding rate of 2.9 percent, this amounts to a 3.5 percent increase in funding probability. Similarly, we find that, at the mean, a ten percent increase in per-capita income (median home value) is associated with a 1.5 percent (1.3 percent) increase in funding probability.

Numerous economic mechanisms could explain these disparities. As discussed above, local government capacity is likely an important factor.[2] Using data from the Census of Governments, we find that local governments serving socioeconomically advantaged subwatersheds have significantly larger revenues and tax receipts. These financial resources are

[2]For instance, a specialized environmental consulting industry exists to help local governments—that can afford the services—to obtain Section 319 grants, including drafting the grant proposal itself.

both directly (via match funding requirements) and indirectly valuable in obtaining grants from state governments. We find no evidence that funding disparities result from an equity-efficiency tradeoff. Using comprehensive granular measurements of surface water pollution, we find—perhaps unsurprisingly—that socioeconomically advantaged subwatersheds are less polluted. We also rule out the importance of spillovers, such as if advantaged populations were more likely to live in upstream subwatersheds.

Our paper makes several contributions. Most directly, we show how federal nonpoint source water pollution funding is distributed in the United States. Case studies in the literature find "mixed results" for racial and socioeconomic disparities in the locations of wetlands projects (Dernoga et al., 2015). Our study provides the first comprehensive nationwide evaluation of Clean Water Act Section 319 funding. In doing so, we contribute to a growing literature on water policy, helping to address the "dearth of economic research on water pollution" (Keiser and Shapiro, 2019a). Evaluations of Clean Water Act and Safe Drinking Water Act policies focus on information disclosure, polluting facility inspections, and wastewater treatment plants (e.g. Bennear and Olmstead, 2008; Grooms, 2015; Keiser and Shapiro, 2019b). We provide evidence for the economics of nonpoint source pollution programs, a "critical area" for research (Olmstead, 2010). The evidence also sheds light on some of the "equity and distributional concerns that arise from changing clean water policies" (Keiser et al., 2022).

More generally, we demonstrate a role for public finance in environmental justice. The literature distinguishes distributive justice—the distribution of pollution—from procedural justice—the decision-making process yielding this distribution (Banzhaf et al., 2019). Our paper speaks to both concerns. We show that poorer populations and people of color live in areas with more surface water pollution. And, we show that pollution program funding decisions are an important factor underlying these differences in exposure. Empirical research on procedural justice has focused primarily on inequitable enforcement of regulation (e.g. Konisky et al., 2021; Marion and West, 2022). We provide evidence that public funding decisions can be another source of environmental disparities. The results of this study suggest that competitive Requests for Proposals and local match funding requirements may be an inequitable process for deciding where to implement environmental improvement projects.

2 Data

Our study uses an annual panel of subwatersheds during 1990-2020. We start with the U.S. Geological Survey's (USGS) Watershed Boundary Dataset. The country is divided into contiguous hydrologic units at various scales using a nested structure: region, subregion, basin, subbasin, watershed, and subwatershed.[3] There are about 103,000 subwatersheds, the smallest comprehensive unit, with an average area of 106 square kilometers.

We use Section 319 funding data from the EPA's Grants Reporting and Tracking System (GRTS). For each sponsored project, the database records details such as the state, award year, and project status. States report total grant awards by project but do not disaggregate funding per subwatershed.[4] Because multiple subwatersheds are often included in the same project, we focus on the extensive margin of whether a subwatershed is included in an award year, rather than funding amounts. Project locations are sparsely available for early years of the data, before a 2002 EPA requirement that projects be geolocated. We provide additional discussion and evidence considering this policy change in Section 3.

For characteristics of subwatershed populations, we use Census Bureau data provided by the Integrated Public Use Microdata Series (IPUMS, Manson et al., 2021). We use the complete count tabulations of population by race at the Census Block level from the 1990, 2000, 2010 and 2020 decennial censuses.[5] We use income per capita and median home values at the Census Tract level from the 1990 and 2000 census long forms and the twelve American Community Survey (ACS) 5-year waves, 2005-2009 through 2016-2020. We merge this race and socioeconomic data to subwatershed units by spatially joining the subwatershed and Census Block geodatabases, using subwatershed-block population weights to assign characteristics. To have coverage in non-census years, we impute census characteristics within-subwatershed using the last observation carried forward method. For example, we assign a subwatershed's 1990 white population share to observations for years 1991-1999.

We use the Census Bureau's Census of Governments and Annual Survey of State and Local Government Finances for data on individual county and municipal/township governments during 1990 through 2019, accessed via the Government Finance Database (Pierson et al., 2015). The Census Bureau surveys the compete count of governments every five years (for our study: 1992, 1997, 2002, 2007, 2012, 2017) and uses a sample in the intervening

[3]Technically, our unit of analysis is a 12-digit Hydrologic Unit Code. Appendix B has additional details.

[4]Not all grants relate to specific subwatersheds. Section 319 funding can also be used for hiring and other more general nonpoint source water pollution activities.

[5]The United States has over eight million unique Census Blocks with an average population of 41 people.

years. By spatially joining subwatersheds to government jurisdictions, we determine the revenue and tax receipts for local governments serving each subwatershed. We exclude federal and state government revenue, and use data fields that isolate resources sourced locally and exclude any intergovernmental transfers.

To quantify pollution, we follow related literature in selecting dissolved oxygen as the measure of water quality (Keiser and Shapiro, 2019b). Our analysis uses dissolved oxygen deficit, defined as 100 minus dissolved oxygen saturation and expressed as a percentage. Nonpoint source pollution increases dissolved oxygen deficits as microorganisms decompose pollutants. A larger value indicates more polluted water. We include data from USGS's National Water Information System, EPA's Storet, and EPA's Storet Legacy Data Center. The data provides fairly comprehensive coverage, and we impute missing dissolved oxygen values for about 28 percent of subwatershed-years. Finally, we use the National Hydrography Dataset Plus, an EPA-USGS collaboration, to determine the presence and extent of any water (downstream) outflows for each subwatershed within the national stream network.

We restrict our analysis to subwatershed-year observations with census population and nonmissing data on income, home values, and pollution.[6] This sample includes 2,554,099 observations covering 88,437 subwatersheds during 1990-2020. Appendix B provides additional discussion of the data preparation and Table A1 presents the summary statistics of key variables. On average, 2.9 percent of subwatersheds receive a grant per year. Subwatershed populations are on average 86.5 percent white, with a mean income per capita of 26,791 (2020$). The average subwatershed has a median home value of 147,924 (2020$).

3 Results

3.1 Socioeconomically advantaged areas receive more grants

As discussed in Section 1, states have broad discretion in distributing Section 319 funds. Evidence of these decisions is provided in Figure 1, a map of subwatershed-level funding frequency during 1990-2020. Note that a subwatershed being included in multiple years of funding means that the subwatershed had grant *awards* in multiple years—not that the same grant-funded project took multiple years (which is also often the case). The same project can receive multiple rounds of funding, however. Some states spread Section 319 funding broadly within their jurisdictions, while others provided funding more narrowly. Overall, the

[6] We drop 637,134 observations (about 14,500 subwatersheds) in total from the balanced panel of *all* U.S. subwatersheds during 1990-2020. The vast majority are dropped due to having no human population.

map shows broad spatial heterogeneity in funding allocations. Our analysis below explores whether these funding decisions are related to racial and socioeconomic characteristics.

First, we consider the average grant funding rate for different subsets of subwatersheds. Figure 2 presents these results, showing a clear pattern. Panel (a) groups subwatersheds by white population share. The average rate of annual funding inclusion is 0.59 percent for subwatersheds with less than 10 percent white population, 1.56 percent for subwatersheds with between 10 to 20 percent white population, 2.75 percent for areas with between 20 to 50 percent white population, and 3.0 percent for majority-white subwatersheds.[7] Panel (b) shows funding rates by quartile of income per capita (using 2020$), which are 0.96 percent, 2.44 percent, 3.77 percent, and 4.44 percent. For quartiles of subwatershed median home values (in 2020$), the respective rates in Panel (c) are 1.35, 2.79, 3.33, and 4.14 percent. For all three characteristics shown in Figure 2, the funding rate for the most socioeconomically advantaged group is three to five times that for the most disadvantaged group. Altogether, these unadjusted relationships suggest significant disparities in grant awards.

We use regression analysis to further explore these patterns. The raw averages shown in Figure 2 do not account for any intertemporal or cross-state variation, which might be important. For one, there may be aggregate time-varying factors that are correlated both with socioeconomic characteristics and with funding probability. Furthermore, the statutory level of decision making is at the state level each year. States vary substantially in the racial and socioeconomic composition of their subwatersheds (as maps in Figure A2 show). If states also differ in their funding priorities, then there may be cross-state correlations between socioeconomic characteristics and subwatershed funding probability. To address these potential confounders, the regression models include state-by-year fixed effects. Our more saturated specifications also include rural area-by-year fixed effects and log-population controls, as these factors may influence the scope for pollution projects.[8] The regression specification is:

$$\mathrm{I}\{\text{funding}\}_{isrt} = \beta \text{socioeconomic\_factor}_{it} + \theta_{st} + \phi_{rt} + \gamma \ln(\text{population})_{it} + \epsilon_{isrt} \tag{1}$$

where $\mathrm{I}\{\text{funding}\}_{isrt}$ is an indicator for whether subwatershed i in state s with rural designation r is included in the awarded grants for year t, $\text{socioeconomic\_factor}_{it}$ is the race or socioeconomic variable of interest (examined using separate regressions), θ_{st} are state-by-

[7]Panel (a) of Figure 2 groups subwatersheds by white population share using these ad hoc bins, instead of using quartiles, because even the 8th percentile subwatershed has a majority-white population.

[8]We use the Census Bureau's binary definition for rural or urban status. Because a subwatershed's rural status might be endogenous with its grant funding, we define all rural areas as of 1990.

year fixed effects, ϕ_{rt} are rural area-by-year fixed effects, $\ln(\text{population})_{it}$ is the natural log of the subwatershed's population at time t, and ϵ_{isrt} is the idiosyncratic error term.

Table 1 presents the regression results. An observation is a subwatershed-year tuple and standard errors are two-way clustered by subwatershed and year. Each panel shows results using the indicated socioeconomic characteristic. The first three columns use the full sample and include different sets of fixed effects and controls. All coefficients are positive and statistically significant at the 10 percent level. For the most saturated specification, Column (3), all estimates are significant at the one percent level. These estimates show that a ten percentage points increase in white population share is associated with 0.1 percentage points increase in a subwatershed's funding inclusion rate, a 3.5 percent increase at the mean. A ten percent increase in per-capita income (median home value) is associated with 0.043 (0.038) percentage points increase in the funding rate, a 1.5 (1.3) percent increase at the mean.

The final two columns subset the data by time period. As noted in Section 2, the EPA changed regulations starting in 2002 to require states to report geolocation information for funded projects. During the first twelve years of data, relatively few grant awards are defined at the subwatershed level. As shown in the bottom panel of Table 1, the 1990-2001 period has 124 subwatershed funding inclusions per year on average, while 2002-2020 has 3,824 per year on average. Although it is possible that states selectively reported project locations prior to 2002 in order to mask socioeconomic disparities, classical attenuation bias is much more likely. Still using Equation (1), Column (4) estimates results for the 1990-2001 period, finding small and insignificant relationships. Column (5) uses the 2002-2020 period and shows estimates that are about 60 to 90 percent larger than those in Column (3). [9] We view these subsample results as a convincing robustness check of our full sample evidence. Collectively, these regression findings reinforce the evidence shown above that a disproportionate share of grants are awarded to socioeconomically advantaged communities.[10]

[9] Figure A3 plots the coefficients from estimating regressions separately for each year of data. Other than the obvious differences between the 1990-2001 and 2002-2020 periods, there are no discernible time trends.

[10] As shown in Figure 1, almost every subwatershed is included in funding during 1990-2020 in Arkansas, Connecticut, Delaware, Louisiana, Virginia, and West Virginia. Specifically, these six states each funded more than 96 percent of subwatersheds. Our examination of state records indicates this is primarily driven by institutional details of these states, such as the Chesapeake Bay and Virginia Waters Clean-Up and Oversight Act of 2006. Because such comprehensive funding provides essentially no cross-sectional identifying variation, Table A2 replicates Table 1 after dropping these six states. Removing the attenuation bias from these states, the estimated socioeconomic disparities in Section 319 funding rates are robust and quantitatively larger.

3.2 Local government capacity can explain funding disparities

Next, we evaluate potential mechanisms for these disparities. Section 319 funding decisions are made by state water quality agencies each year. However, almost all states allocate grants to local projects using an annual application process, such as Requests for Proposals (EPA, 2011).[11] Socioeconomically advantaged communities arguably have a competitive advantage in this process. Some of the advantage might be political, if wealthier populations can use political connections to influence grant funding decisions (Becker, 1983).

More likely, the advantage is financial (Gargan, 1981). One stylized fact of funding awards is the requirement of matching funds from grant applicants. Federal Section 319 funding requires at least a 40 percent match from state and local sources, and most states pass through all or part of this requirement. Even if match funding is not explicitly required, pledging local funds helps an application in the proposal evaluation process.[12] Moreover, pollution projects are often capital-intensive, and financial revenue is a major determinant of local governments' ability to secure bond financing (Butler and Yi, 2022). Financial resources are also indirectly valuable, increasing the capacity to dedicate staff to writing grant proposals or to hire consultants to craft more competitive applications.

We empirically test the importance of a local government capacity mechanism, finding that government resources are significantly greater in socioeconomically advantaged subwatersheds. To do so, we estimate versions of Equation (1) for measures of revenue of the local governments—counties, municipalities, and townships—that serve each subwatershed. To better quantify *local* capacity, we exclude federal and state government revenue and intergovernmental transfers. This analysis uses our panel of subwatersheds matched to annual data for 1990-2019 from six Census of Governments years and the intervening Annual Surveys of State and Local Government Finances.[13]

Table 2 presents these estimates. The first two columns use measures of local government revenue as the dependent variable. Column (1) shows that the three socioeconomic terms have a strong positive relationship with log-revenue from all sources. Because total revenue includes revenue from providing specific services (such as utilities and public transit), we also estimate relationships using tax receipts, which should more closely represent a gov-

[11] The exception is Oklahoma, which coordinates funding decisions with local conservation districts.

[12] For example, Massachusetts, North Carolina, and Texas require awardees to cover at least 40 percent of the project cost. California requires a minimum local match of 25 percent, with few exceptions. Florida and Ohio do not require match funding, but local funding contributes to a higher proposal ranking.

[13] Public finance data is not available for every subwatershed-year in our primary sample, which drops by 609,660 observations (about 24 percent). In Table A3, we reproduce the estimates from Table 1 using this reduced sample. Results remain statistically significant and quantitatively show slightly larger disparities.

ernment's fungible resources. These results in Column (2) show an even greater association with the socioeconomic characteristics. Quantitatively, we estimate a semi-elasticity of 0.32 between tax receipts and subwatersheds' white population share. The elasticity between tax receipts and income per capita is 0.56. The elasticity of tax receipts to home value is 0.13. In Columns (3) and (4), respectively, we use local government log-revenue and log-tax receipts as the explanatory terms. As in Table 1, the dependent variable is an indicator for subwatershed inclusion in the grant awards for the year. These estimates are also positive and significant, both statistically and economically, showing that areas served by higher-resource governments are more likely to be awarded grant funding.[14]

It may be unsurprising that more affluent areas have better-funded local governments, and we emphasize that these are descriptive patterns—not causal estimates. Likely, other mechanisms are also factors for why socioeconomically advantaged areas receive a disproportionate share of Section 319 grant awards. Nonetheless, these findings provide compelling evidence supporting that a local government capacity mechanism could drive the disparities in grant awards documented above.

3.3 No evidence that disparities are an equity-efficiency tradeoff

A second possibility is that socioeconomic disparities are an unintended consequence of distributing funding to locations where grants are most beneficial. Despite its severity, nonpoint source water pollution remains largely unregulated in the United States (Keiser and Shapiro, 2019a). The Section 319 program is the country's primary policy addressing this concern. We find in Ren and West (2022) that funded projects significantly reduce pollution in treated subwatersheds, arguably providing substantial benefit. If the social marginal benefit of a project is larger in socioeconomically advantaged subwatersheds, it could be economically efficient to award more grants to these areas.

To assess this potential mechanism, we first evaluate the relationship between water pollution levels and socioeconomic characteristics.[15] There are diminishing marginal benefits from reducing water pollution (Olmstead, 2010). So, project funding should generally be more impactful in more polluted areas. We estimate Equation (1) for a dependent variable of dissolved oxygen deficit, where a larger DO deficit indicates more polluted water.

Table 3 presents these results, using the same specifications as Table 1. All estimates

[14]The point estimates might seem small, but government revenues vary dramatically. For example, local government revenue for the 75th percentile subwatershed is more than 950 percent that of the median.

[15]Figure A4 provides a map of subwatershed average dissolved oxygen deficit during 1990-2020.

are negative, statistically significant, and robust across the three columns. The estimates show that a ten percentage points increase in white population share is associated with about 0.32 percentage points reduction in DO deficit, or 2.6 percent of the sample mean. A direct interpretation of this magnitude is challenging because most effects of DO deficits are nonlinear, such as thresholds at which aquatic species can no longer survive. As one point of comparison, the average annual change in DO deficit for our panel is 0.1 percentage points reduction per year. Thus, in a manner of speaking, subwatersheds with more white populations have water quality that is effectively years ahead in pollution reductions.

Compared to the racial disparity, the estimates for income and home values are somewhat more muted. Regardless, we can rule out that the disparities in grant awards are because affluent communities are more polluted. A ten percent increase in subwatershed income per capita is associated with about 0.05 percentage points reduction in DO deficit. A ten percent increase in median home values corresponds to about 0.03 percentage points reduction. This evidence shows that—within state-year and controlling for population and rural status—socioeconomically advantaged populations live in relatively *less* polluted subwatersheds.

Social marginal benefits could also differ due to spillovers. Given the watershed-based nature of our study, there should be relatively little potential for spillover effects from most Section 319 projects. By definition, a watershed is an area of land where all water drains to a common location. However, surface water does flow within a network of rivers and streams. We use data mapping the flow of water within the continental United States to determine downstream outflows between subwatersheds.[16] In total, there are about 19,000 rivers and streams with inter-subwatershed flow, and about 56 percent of subwatersheds in our sample are upstream of at least one other subwatershed. We construct three measures of outflows. First, we define a binary indicator for whether a subwatershed has any downstream outflows into another subwatershed(s). Second, we determine the total number of outflows per subwatershed, i.e. the count of rivers and streams that flow into another subwatershed(s). Finally, conditional on a subwatershed having outflows, we determine the total number of downstream subwatersheds, including non-adjacent subwatersheds.

We estimate regressions using these three measures of subwatershed spillovers as dependent variables in Equation (1). There is no evidence that downstream spillovers are associated with subwatersheds' socioeconomic characteristics. The estimates (shown in Table A4) are minuscule and statistically insignificant for all three outflow measures and all three socioeconomic terms. For instance, we find that a ten percentage point increase in

[16]We verified that the funding results in Table 1 are virtually unchanged without Alaska and Hawaii.

white population share corresponds to only 0.01 percentage points increase in subwatershed outflow probability—on a mean of 56.3 percent. Evidently, wealthier and more white populations do not systematically live in more upstream subwatersheds.

The results in this section imply that socioeconomically advantaged populations do not live in areas where water pollution projects are more effective. Whether advantaged populations obtain greater social *benefit* from Section 319 funding depends also on preferences, which are unobserved. One might be tempted to interpret their disproportionate share of grant awards as evidence of a stronger revealed preference for addressing water pollution. We caution that such an interpretation conflates differences in willingness to pay with the significant differences in ability to pay that are demonstrated above. Empirically, we find no evidence that the funding disparities are an equity-efficiency tradeoff.

4 Conclusions

For a variety of reasons, both technical and political, the United States addresses nonpoint source water pollution primarily by funding local public works projects. This study has examined where state governments use these Clean Water Act Section 319 funds. We find that watersheds with socioeconomically advantaged populations are less polluted but more likely to receive pollution project funding. We demonstrate that this disparity can be explained by differences in local government capacity. The funding is usually awarded through a competitive application process, such as Requests for Proposals. Governments that serve poorer communities and people of color generally have fewer resources to craft a competitive proposal application or to meet match funding requirements.

These findings have several policy implications. Most directly, this evidence suggests that the benefits of surface water pollution programs are inequitably distributed. Although it is state government agencies that directly make Section 319 funding decisions, federal policymakers can encourage these awards to account for disparate local capacities. As one step towards a more equitable process, some states now waive match funding requirements for grants to disadvantaged communities. Recent guidance from the Environmental Protection Agency asks other states do to the same (Hall, 2021). However, if a state waives a local match requirement, the state must then contribute the statutory 40 percent match, which puts the onerous on state governments to fund environmental justice. Instead, match funding waivers for disadvantaged communities could be implemented systematically.

More broadly, our study indicates that a competitive application process may be an inequitable way to determine environmental funding priorities. Requests for Proposals and match funding requirements are ubiquitous facets of the project selection processes that governments and other organizations rely on to make funding decisions. The funding available for environmental improvement projects is often scarce, and a competitive process encourages applicants to be vocal about why projects in their communities should be prioritized. A competitive process also favors those with the resources to make their voice heard.

References

S. Banzhaf, L. Ma, and C. Timmins. Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives*, 33(1):185–208, 2019.

G. S. Becker. A theory of competition among pressure groups for political influence. *The Quarterly Journal of Economics*, 98(3):371–400, 1983.

L. S. Bennear and S. M. Olmstead. The impacts of the "right to know": Information disclosure and the violation of drinking water standards. *Journal of Environmental Economics and Management*, 56(2):117–130, 2008.

A. W. Butler and H. Yi. Aging and public financing costs: Evidence from U.S. municipal bond markets. *Journal of Public Economics*, 211:104665, 2022.

R. G. Chambers and J. Quiggin. Non-point-source pollution regulation as a multi-task principal-agent problem. *Journal of Public Economics*, 59(1):95–116, 1996.

M. A. Dernoga, S. Wilson, C. Jiang, and F. Tutman. Environmental justice disparities in Maryland's watershed restoration programs. *Environmental Science & Policy*, 45:67–78, 2015.

EPA. A national evaluation of the Clean Water Act Section 319 program. Technical report, United States Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds, 2011.

J. J. Gargan. Consideration of local government capacity. *Public Administration Review*, 41 (6):649–658, 1981.

K. K. Grooms. Enforcing the Clean Water Act: The effect of state-level corruption on compliance. *Journal of Environmental Economics and Management*, 73:50–78, 2015.

L. Hall. Near-term actions to support environmental justice in the nonpoint source program. United States Environmental Protection Agency memorandum to state and territorial Section 319 nonpoint source program managers and staff, 2021.

D. A. Keiser and J. S. Shapiro. US water pollution regulation over the past half century: Burning waters to crystal springs? *Journal of Economic Perspectives*, 33(4):51–75, 2019a.

D. A. Keiser and J. S. Shapiro. Consequences of the Clean Water Act and the demand for water quality. *The Quarterly Journal of Economics*, 134(1):349–396, 2019b.

D. A. Keiser, S. M. Olmstead, K. J. Boyle, V. B. Flatt, B. L. Keeler, D. J. Phaneuf, J. S. Shapiro, and J. P. Shimshack. The evolution of the 'waters of the United States' and the role of economics. *Review of Environmental Economics and Policy*, 16(1):146–152, 2022.

D. M. Konisky, C. Reenock, and S. Conley. Environmental injustice in Clean Water Act enforcement: Racial and income disparities in inspection time. *Environmental Research Letters*, 16(8):084020, 2021.

Y. Kuwayama, S. Olmstead, and J. Zheng. A more comprehensive estimate of the value of water quality. *Journal of Public Economics*, 207:104600, 2022.

S. Manson, J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles. IPUMS national historical geographic information system: version 16, 2021.

J. Marion and J. West. Dirty business: Principal-agent problems in hazardous waste remediation. Working paper, 2022.

W. E. Oates. An essay on fiscal federalism. *Journal of Economic Literature*, 37(3):1120–1149, 1999.

S. M. Olmstead. The economics of water quality. *Review of Environmental Economics and Policy*, 4(1):44–62, 2010.

K. Pierson, M. L. Hand, and F. Thompson. The government finance database: A common resource for quantitative research in public financial analysis. *PLoS ONE*, 10(6), 2015.

Q. Ren and J. West. Cleaner waters and urbanization. Working paper, 2022.

Figure 1: Map of subwatershed grant award frequency during 1990-2020

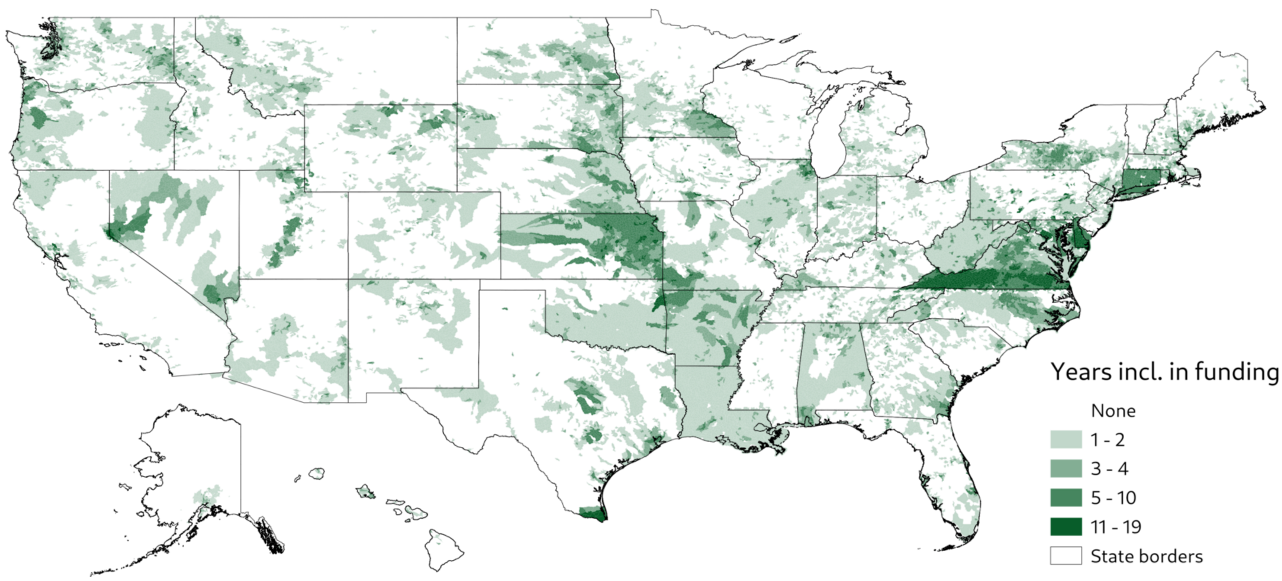
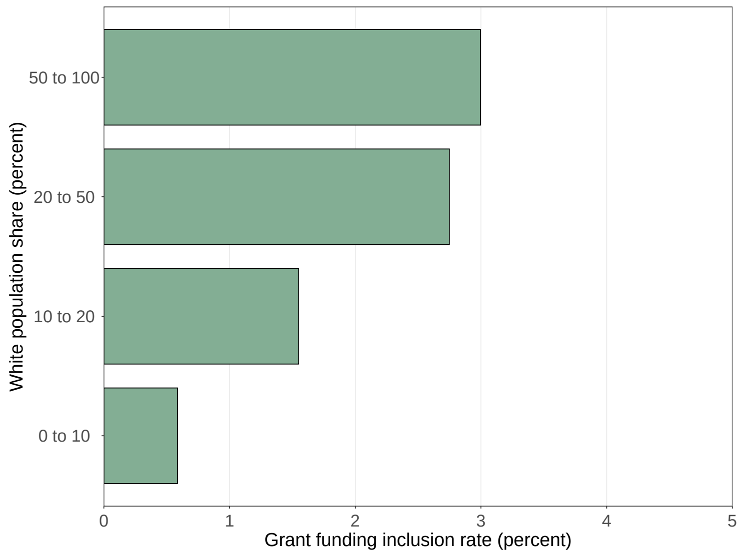
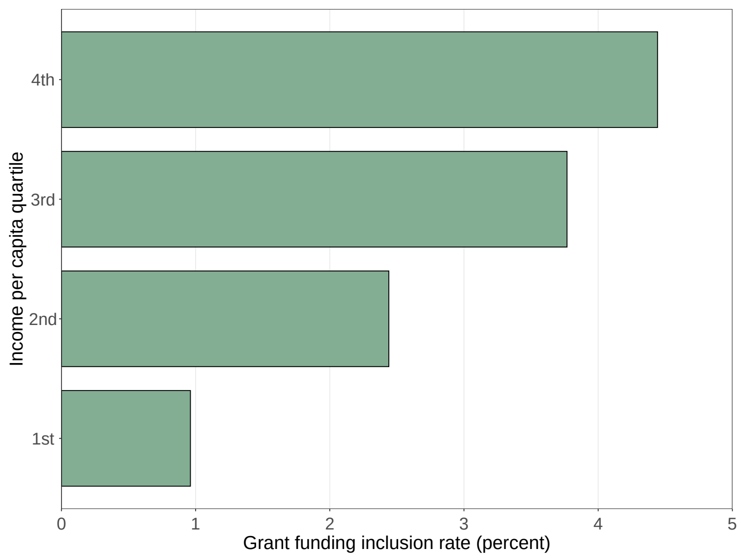


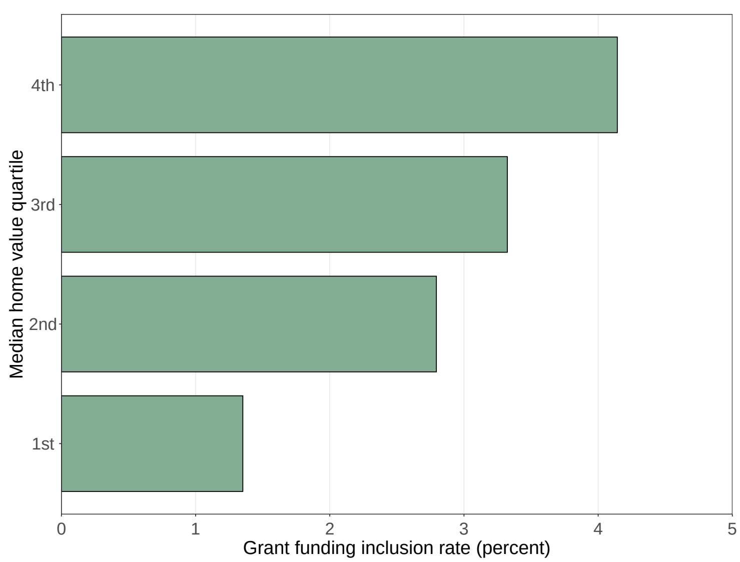
Figure 2: Relationships between socioeconomic characteristics and grant funding rates



(a) White population share



(b) Income per capita



(c) Median home value

Notes: Data covers years 1990-2020. The average rate of inclusion in annual grant awards is 2.9 percent.

Table 1: Estimated relationships between socioeconomic characteristics and grant funding

	Dependent variable: 100 × indicator for subwatershed inclusion in grant awards for the year				
	Full sample: 1990-2020			Subset by time periods	
				1990-2001	2002-2020
	(1)	(2)	(3)	(4)	(5)
Panel [A]: White population share	0.500*	0.742**	1.005***	-0.021	1.633***
	(0.274)	(0.284)	(0.308)	(0.033)	(0.460)
Panel [B]: Log(income per capita)	0.878***	0.635***	0.432***	0.043	0.719***
	(0.177)	(0.153)	(0.141)	(0.028)	(0.211)
Panel [C]: Log(median home value)	0.705***	0.556***	0.383***	0.021	0.714***
	(0.160)	(0.136)	(0.108)	(0.019)	(0.174)
State × year fixed effects	Yes	Yes	Yes	Yes	Yes
I{1990 rural area} × year FE		Yes	Yes	Yes	Yes
Population controls			Yes	Yes	Yes
Dependent variable mean	2.903	2.903	2.903	0.146	4.737
Total subwatershed funding awards	74,144	74,144	74,144	1,486	72,658
Number of subwatersheds	88,437	88,437	88,437	87,176	85,762
Observations	2,554,099	2,554,099	2,554,099	1,020,150	1,533,949

Notes: Each observation is a subwatershed-year tuple. All dollar values are in 2020$. Standard errors are two-way clustered by subwatershed and year.

*** p<0.01, ** p<0.05, * p<0.1

Table 2: Estimated relationships between socioeconomic characteristics and public finances and estimated relationships between public finances and grant funding

	Dependent variable			
	Natural log of local government revenue		100 × indicator for subwatershed inclusion in grant awards for the year	
	All revenue (1)	Tax receipts (2)	(3)	(4)
Panel [A]: White population share	0.083***	0.315***		
	(0.026)	(0.059)		
Panel [B]: Log(income per capita)	0.372***	0.556***		
	(0.025)	(0.036)		
Panel [C]: Log(median home value)	0.074***	0.129***		
	(0.009)	(0.009)		
Panel [D]: Log(all revenue)			0.064*	
			(0.035)	
Panel [E]: Log(tax receipts)				0.084**
				(0.040)
State × year fixed effects	Yes	Yes	Yes	Yes
I{1990 rural area} × year FE	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes
Dependent variable mean			3.150	3.150
Number of subwatersheds	84,262	84,262	84,262	84,262
Observations	1,944,439	1,944,439	1,944,439	1,944,439

Notes: Each observation is a subwatershed-year tuple. Data covers years 1990-2019 and includes only “own source” revenue to local governments such as counties, cities, and townships. Intergovernmental transfers from federal and state sources are excluded. All dollar values are in 2020$. Standard errors are two-way clustered by subwatershed and year.

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Estimated relationships between socioeconomic characteristics and water pollution

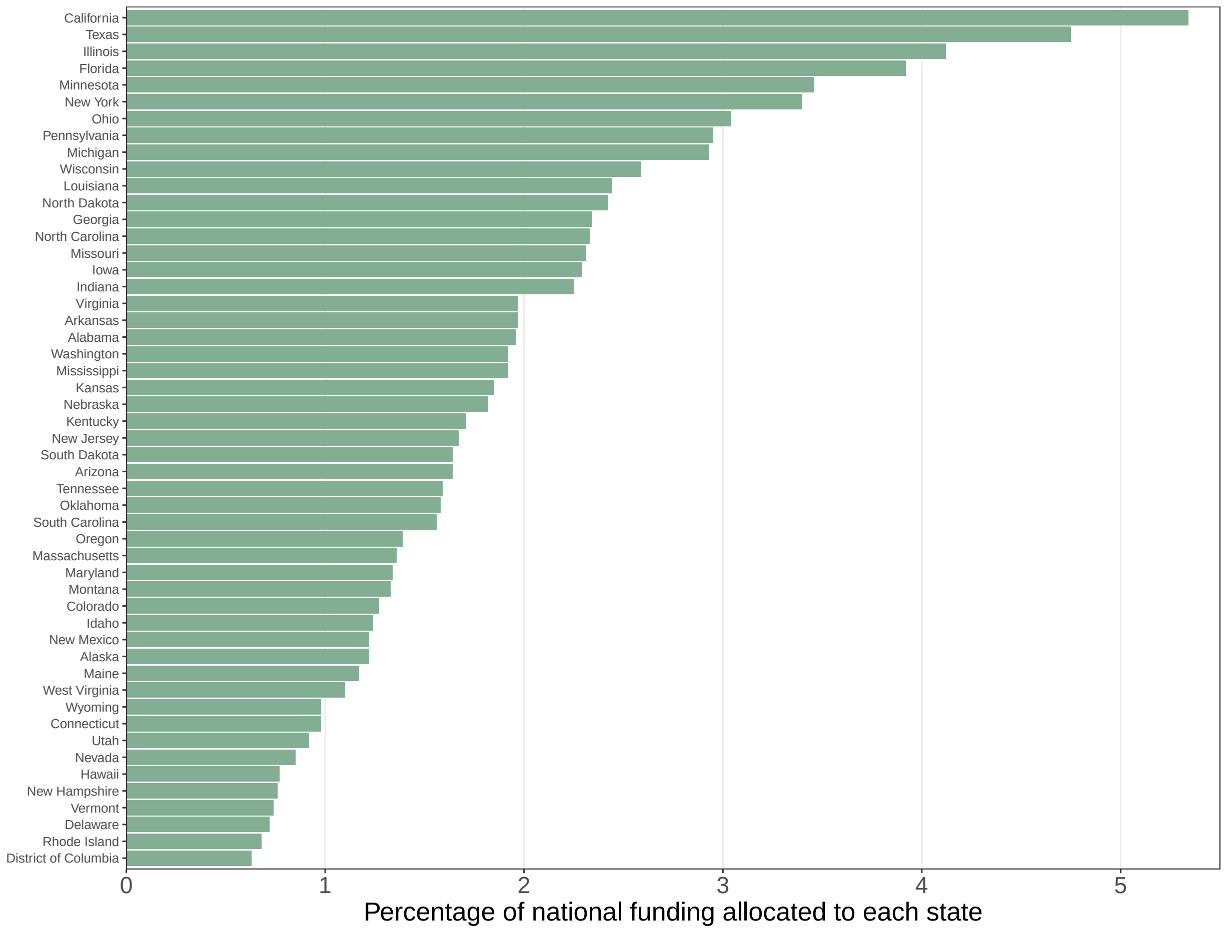
	Dependent variable: dissolved oxygen deficit		
	(1)	(2)	(3)
Panel [A]: White population share	-3.111***	-3.136***	-3.246***
	(0.434)	(0.438)	(0.432)
Panel [B]:Log(income per capita)	-0.543***	-0.518***	-0.449**
	(0.175)	(0.174)	(0.177)
Panel [C]: Log(median home value)	-0.336***	-0.325***	-0.266**
	(0.104)	(0.101)	(0.101)
State × year fixed effects	Yes	Yes	Yes
I{1990 rural area} × year FE		Yes	Yes
Population controls			Yes
Dependent variable mean	12.513	12.513	12.513
Number of subwatersheds	88,437	88,437	88,437
Observations	2,554,099	2,554,099	2,554,099

Notes: Each observation is a subwatershed-year tuple. Data covers years 1990-2020. Dissolved oxygen deficit equals 100 minus dissolved oxygen saturation, measured as a percentage. A larger value indicates more polluted water. All dollar values are in 2020$. Standard errors are two-way clustered by subwatershed and year.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

A Appendix figures and tables

Figure A1: Clean Water Act Section 319 funding allocations to states each year



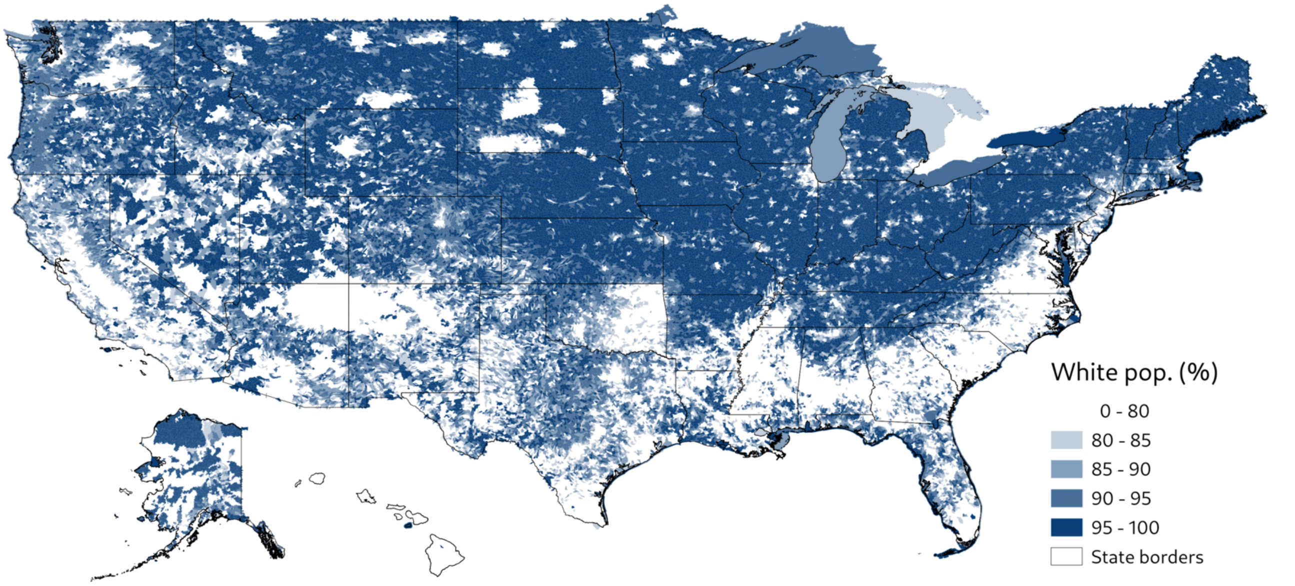
Notes: One-third percent of total funding is provided to tribes. The figure shows 98.35 percent of the remaining national allocation. The other 1.65 percent is provided to U.S. territories: Puerto Rico receives 0.56 percent and American Samoa, Guam, Marianas, and the Virgin Islands each receive 0.27 percent.

Table A1: Summary statistics

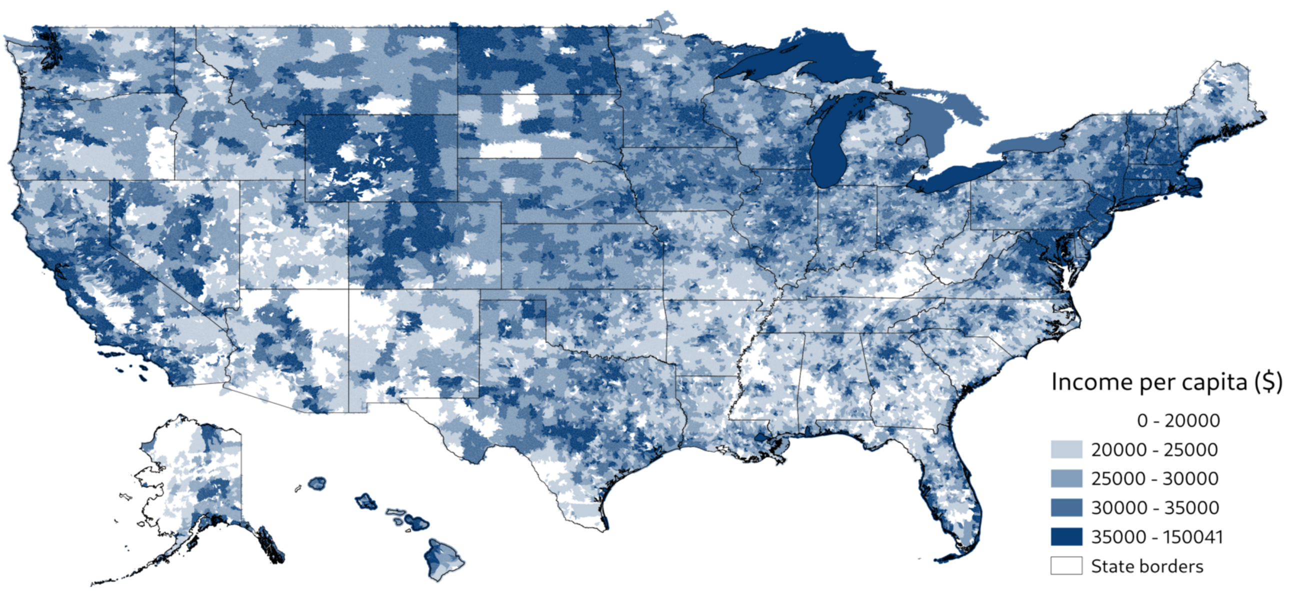
	Median	Mean	SD	Observations
100×I{grant in year}	0.000	2.903	16.789	2,554,099
White population share	0.963	0.865	0.223	2,554,099
Income per capita (2020$)	25,421	26,791	9,112	2,554,099
Median home value (2020$)	120,273	147,924	106,027	2,554,099

Notes: Each observation is a subwatershed-year tuple. Data covers years 1990-2020. I{grant in year} indicates whether a subwatershed is included in the awarded grants during the year.

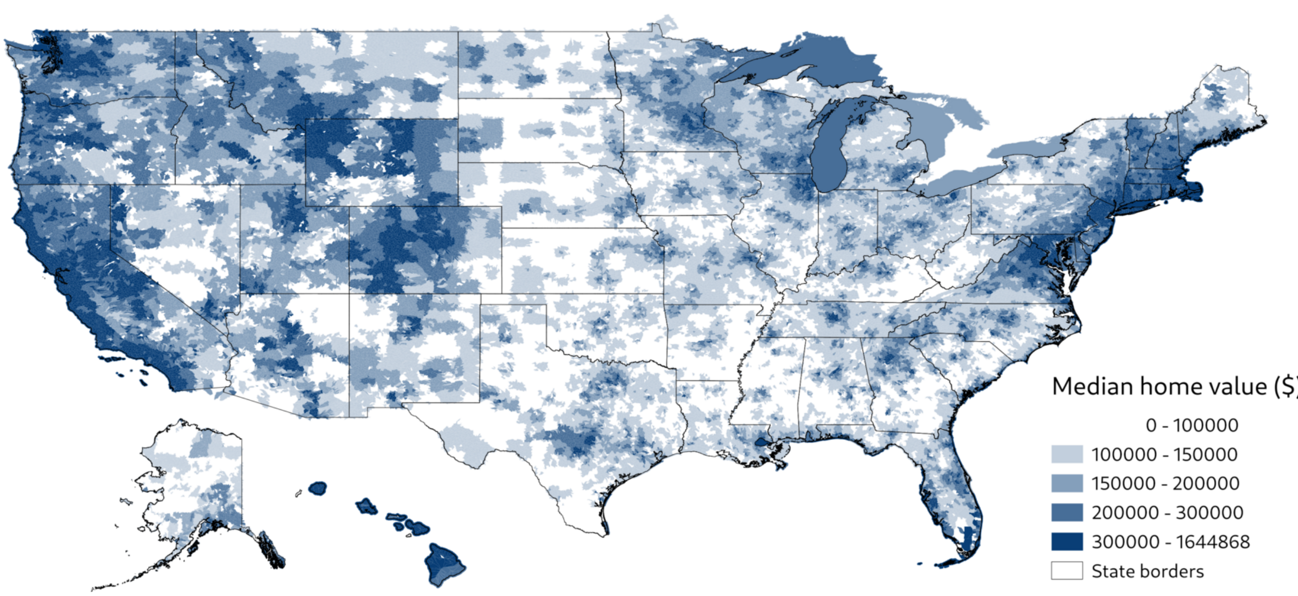
Figure A2: Maps of subwatershed racial and socioeconomic characteristics during 1990-2020



(a) White population share (percent)

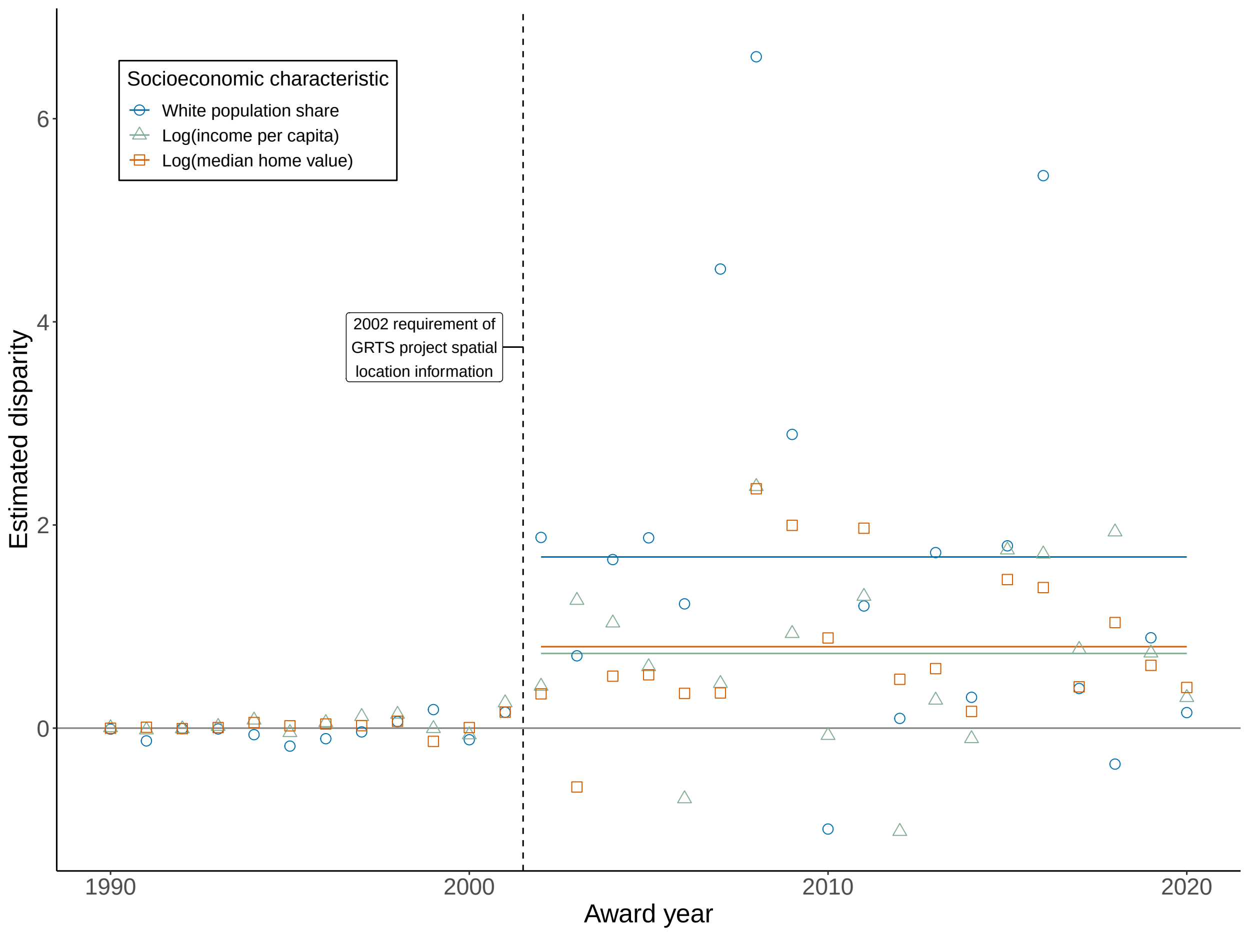


(b) Income per capita (2020$)



(c) Median home value (2020$)

Figure A3: Annual estimates relating socioeconomic characteristics and grant funding



Notes: The figure plots coefficients from estimating regressions separately for each year of data. The dependent variable is 100 × an indicator for subwatershed inclusion in grant awards for the year. The regressions control for state fixed effects, subwatershed rural area indicators, and subwatershed population. The horizontal lines show the average of estimates for each characteristic during 2002-2020. In total, 1,486 grant awards are matched to subwatersheds during 1990-2001 and 72,658 awards during 2002-2020.

Table A2: Estimated relationships between socioeconomic characteristics and grant funding: using the sub-sample of states with cross-sectional identifying variation

	Dependent variable: 100 × indicator for subwatershed inclusion in grant awards for the year				
	Full sample: 1990-2020			Subset by time periods	
				1990-2001	2002-2020
	(1)	(2)	(3)	(4)	(5)
Panel [A]: White population share	0.516**	0.746***	1.019***	-0.031	1.666***
	(0.246)	(0.258)	(0.285)	(0.033)	(0.414)
Panel [B]: Log(income per capita)	1.215***	0.974***	0.777***	0.054	1.295***
	(0.223)	(0.190)	(0.166)	(0.031)	(0.209)
Panel [C]: Log(median home value)	0.872***	0.727***	0.558***	0.022	1.057***
	(0.199)	(0.175)	(0.144)	(0.020)	(0.224)
State × year fixed effects	Yes	Yes	Yes	Yes	Yes
I{1990 rural area} × year FE		Yes	Yes	Yes	Yes
Population controls			Yes	Yes	Yes
Dependent variable mean	2.348	2.348	2.348	0.141	3.820
Total subwatershed funding awards	56,585	56,585	56,585	1,360	55,225
Number of subwatersheds	83,756	83,756	83,756	82,515	81,083
Observations	2,410,022	2,410,022	2,410,022	964,420	1,445,602

Notes: This table replicates Table 1 after dropping all subwatersheds in Arkansas, Connecticut, Delaware, Louisiana, Virginia, and West Virginia. More than 96 percent of subwatersheds in each of these states are included in grant awards during 1990 to 2020, providing essentially no cross-sectional variation. Standard errors are two-way clustered by subwatershed and year. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Estimated relationships between socioeconomic characteristics and grant funding: using the sample with available public finance data

	Dependent variable: 100 × indicator for subwatershed inclusion in grant awards for the year				
	Full sample: 1990-2019			Subset by time periods	
				1990-2001	2002-2019
	(1)	(2)	(3)	(4)	(5)
Panel [A]: White population share	0.630*	0.919**	1.270***	-0.001	1.939***
	(0.328)	(0.337)	(0.362)	(0.039)	(0.513)
Panel [B]: Log(income per capita)	1.032***	0.771***	0.592***	0.051*	0.948***
	(0.203)	(0.177)	(0.164)	(0.029)	(0.234)
Panel [C]: Log(median home value)	0.722***	0.573***	0.419***	0.018	0.769***
	(0.169)	(0.146)	(0.120)	(0.023)	(0.195)
State × year fixed effects	Yes	Yes	Yes	Yes	Yes
I{1990 rural area} × year FE		Yes	Yes	Yes	Yes
Population controls			Yes	Yes	Yes
Dependent variable mean	3.150	3.150	3.150	0.166	5.078
Total subwatershed funding awards	61,253	61,253	61,253	1,264	59,989
Number of subwatersheds	84,262	84,262	84,262	83,360	81,300
Observations	1,944,439	1,944,439	1,944,439	763,005	1,181,434

Notes: Each observation is a subwatershed-year tuple. The full sample in this table is the same as the public finance sample used in Table 2. All dollar values are in 2020$. Standard errors are two-way clustered by subwatershed and year.

*** p<0.01, ** p<0.05, * p<0.1

Figure A4: Map of subwatershed average dissolved oxygen deficit during 1990-2020

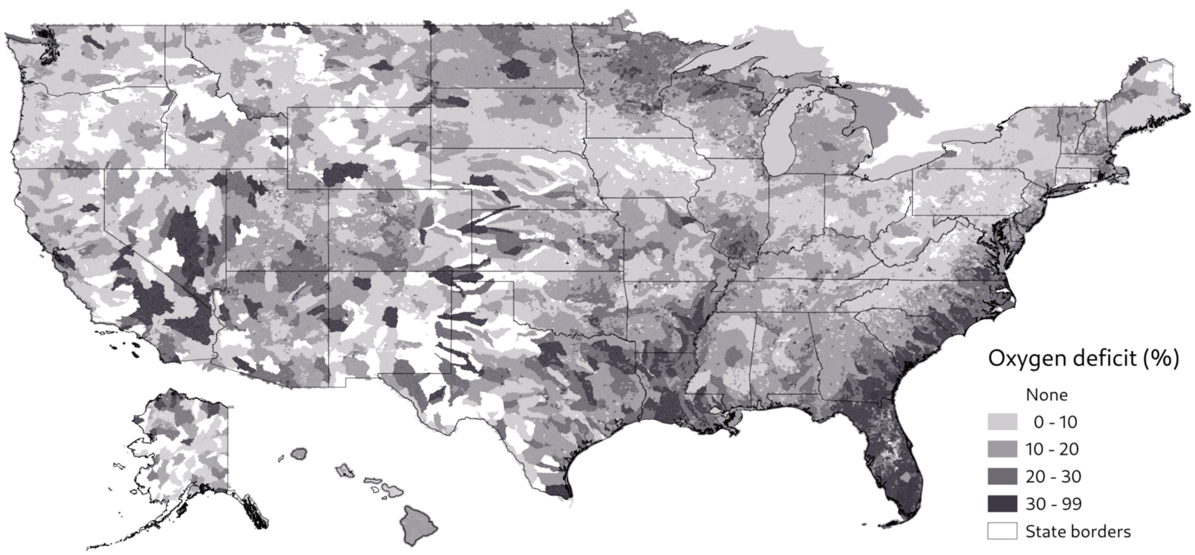


Table A4: Estimated relationships between socioeconomic characteristics and spillovers

	Indicator that subwatershed has downstream outflows (1)	Total number of downstream outflows (2)	Number of downstream subwatersheds if any (3)
Panel [A]: White population share	0.001	-0.005	0.065
	(0.009)	(0.009)	(0.581)
Panel [B]:Log(income per capita)	-0.003	-0.006	-0.477
	(0.006)	(0.006)	(0.328)
Panel [C]: Log(median home value)	-0.002	-0.003	-0.097
	(0.003)	(0.003)	(0.171)
State × year fixed effects	Yes	Yes	Yes
I{1990 rural area} × year FE	Yes	Yes	Yes
Population controls	Yes	Yes	Yes
Dependent variable mean	0.563	0.571	7.986
Number of subwatersheds	80,209	80,209	44,881
Observations	2,396,406	2,396,406	1,349,382

Notes: Alaska and Hawaii are not included in the flowline data. The outcome in Column (1) equals one if the subwatershed has any river or stream flowing out into another subwatershed(s), and is otherwise zero. Column (2) uses the total number of outflowing rivers or streams. Column (3) conditions on the subwatershed having outflow and uses the total number of downstream subwatersheds. All dollar values are in 2020$. Standard errors are two-way clustered by subwatershed and year.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

B Data appendix

This appendix provides additional details on the collection and cleaning of the data used in our study. We use administrative datasets from several sources and organize the data at the subwatershed level, the unit of analysis in this study. We then match data from different sources together to construct a panel of subwatersheds spanning from 1990 to 2020.

B.1 Subwatersheds as the unit of analysis

The spatial area of the United States is divided and sub-divided into cross sections of successively smaller "hydrologic units" in the Watershed Boundary Dataset (WBD), a seamless, national dataset developed by U.S. Geological Survey (USGS). Hydrologic units are a nested hierarchical system based on surface hydrologic features. Technically, any hydrologic unit at any aggregation is a watershed, in the general sense of the term: an area of land where all water drains to a common location. However, for more specific purposes, there are six primary levels of hydrologic units: region, subregion, basin, subbasin, watershed, and subwatershed. Each polygon area is identified by a unique hydrologic unit code (HUC) consisting of two to twelve digits: 2-digit HUC for region, 4-digit HUC for subregion, 6-digit HUC for basin, 8-digit HUC for subbasin, 10-digit HUC for watershed, and 12-digit HUC for subwatershed. In some cases, USGS provides HUC identifiers down to a 16-digit level, but the 12-digit subwatershed code (HUC12) is the most granular level with full coverage nationally. Our study data consists initially of all 102,943 HUC12-identified subwatersheds.

B.2 Clean Water Act Section 319 grants data

We use Section 319 funding data from the U.S. Environmental Protection Agency's Grants Reporting and Tracking System (GRTS). GRTS is designed to be the primary tool for monitoring and maintaining the funded nonpoint source pollution targeting programs sponsored by Section 319 grants. GRTS directly pulls grant information from EPA databases and grantees are required to report project progress annually. The system provides public "Interactive Reports" at the subwatershed level, which include details about the funded projects. We make use of these reports in our study, obtaining the report for each project. Consistently reported details include the state, project identifier, award fiscal year, grant award amount, and project status. Many projects report additional information as well.

GRTS is a dynamic database, and we use the data version as of September 2021. We keep projects with award years during 1990 to 2020. There are 9,824 projects that match to a specific subwatershed(s), which covers 33,446 subwatersheds, about one third of all subwatersheds. In earlier years of the data, relatively few projects report the specific (subwatershed) location within the state, before an EPA requirement that spatial information be reported starting in 2002. The same project can receive newly-awarded funding in multiple award cycles, and the same subwatershed can benefit from multiple distinct projects.

States report total grant awards by project each year but do not disaggregate funding per subwatershed. Also, multiple subwatersheds are often included in the same project. For these reasons, we focus on the extensive margin of whether a subwatershed is included in a funding year, rather than funding amounts. We treat the award fiscal year as the funding year. For example, we treat all subwatersheds that were included in funding awards made during fiscal year 2020 as being included in funding for year 2020. For our study, it is reasonable to treat the fiscal year as the funding year, as states usually announce the decisions of grant awards for a given fiscal year during the same year, though in different months depending on the state (e.g. March in Delaware, May in California, and June in Illinois).

The average annual funding inclusion rate is 2.4 percent for the full balanced panel of 3,191,233 observations covering all of the United States' 102,943 subwatersheds spanning years 1990-2020. In our filtered study sample of 2,554,099 (populated) subwatershed-year observations, the mean annual funding inclusion rate is 2.9 percent. Details on the sample construction are provided below in Section B.7.

B.3 Race and socioeconomic data

This section describes our process to transform the census race and socioeconomic data at block level or tract level into subwatershed-level data. The data are provided by Integrated Public Use Microdata Series (IPUMS, Manson et al., 2021). Our study focuses on three characteristics of populations: white population share, income per capita and median home value. We selected these characteristics because there is comprehensive and standardized coverage across the decennial census forms and American Community Survey waves during our study period of 1990 through 2020.

We use the most spatially granular measurement of each characteristic. Specifically, we collect the complete count tabulations of population by race at Census Block level from the four decennial censuses (1990, 2000, 2010 and 2020) and income per capita and housing values at Census Tract level from two decennial census long forms (1990 and 2000) and the twelve waves of American Community Survey (ACS) 5-year estimates, which are 2005-2009, 2006-2010, 2007-2011, 2008-2012, 2009-2013, 2010-2014, 2011-2015, 2012-2016, 2013-2017, 2014-2018, 2015-2019 and 2016-2020. ACS replaced the long form starting in 2010 and has been released annually since then.

B.3.1 Spatial crosswalks between subwatersheds and Census Blocks

We spatially join subwatersheds to Census Blocks using the related geodatabases. We use the subwatershed boundary polygons from the Watershed Boundary Dataset (WBD) discussed above. We use the Census Block polygons in geodatabases for the four decennial censuses (1990, 2000, 2010 and 2020) as provided by IPUMS. Each subwatershed, the smallest comprehensive hydrologic unit in WBD and the unit of analysis in this study, is uniquely identified by a 12-digit Hydrologic Unit Code (HUC12). Each Census Block for each census year is uniquely identified in the data by a Federal Information Processing Standards (FIPS) code and by an IPUMS GISJOIN code. The United States has over eight million unique Census Blocks and about 103,000 HUC12 subwatersheds.

We spatially join subwatersheds with Census Blocks for each census year to form spacial crosswalks between subwatershed(s) and block(s), including the size of their overlapping area. For each intersection, we calculate the intersection area ratio, which is the total intersecting area of an intersecting subwatershed-block divided by the total area of the Census Block, i.e. the fraction of the block's area located within the subwatershed. Note that, because there are far more Census Blocks than subwatersheds, the majority of blocks (about two-thirds) spatially intersect with only a single subwatershed.

B.3.2 Population and white population share at subwatershed level

We use the intersection area ratios calculated above as weights to calculate the weighted population counts for each intersection. For example, if a particular Census Block has 20 percent of its area intersecting one subwatershed and 80 percent of its area in a second subwatershed, we would assign 20 percent of the block's population to the first subwatershed and 80 percent to the second. We then add up the weighted population counts by subwatershed to determine population counts at the subwatershed level.

Importantly, our use of these intersection area ratios ensures that the total United States population for a given census is the same when measured at the Census Block level as when measured at the subwatershed level. That is, we map population from blocks to subwatersheds, holding total population constant. And, because Census Blocks are quite small—the average block has a total population of 41 people—this approach should have very little measurement error of subwatershed-level population. We apply the same process for white population counts to determine subwatershed-level white population totals. The subwatershed-level white population share is then defined as the white population divided by the total population for each subwatershed.

B.3.3 Income per capita and median home value at subwatershed level

We construct subwatershed-block population weights to assign Census Tract-level income per capita and median home values to subwatersheds. A subwatershed-block population weight is the population of the subwatershed-block intersection divided by the total subwatershed population. That is, we determine what share of a subwatershed's (calculated) population is contained within each intersecting Census Block—and by extension, what share of subwatershed population is contained within each intersecting Census Tract.

Using these subwatershed-block population weights, we calculate the weighted income per capita for each subwatershed-tract intersection. Then we use these calculated weighted income per capita values for each subwatershed-tract to generate income per capita at the subwatershed level. As a simple example, suppose that 25 percent of a subwatershed's population lives in Census Tract A and the other 75 percent lives in Census Tract B. Suppose that the income per capita is \$20,000 in Tract A and that income per capita is \$30,000 in Tract B. Our process calculates the subwatershed-level income per capita as \$27,500 (i.e. $0.25 \times 20{,}000 + 0.75 \times 30{,}000$). The same process applies to median home value. Because we are taking a weighted average of Tract-level median values, our measure of subwatershed-level median home value is technically not the *median* value of homes in the subwatershed. We use the phrase "median home value" throughout our paper to be consistent with the term as defined in the original census data source.

As with the weighted population calculations, this method maps tract-level characteristics to subwatersheds in a way that generally maintains overall data properties. For instance, the sum total of income should be the same if calculated at the tract-level or at the subwatershed-level. Moreover, by using the block level data to calculate subwatershed-block population weights, our approach minimizes the measurement error of assigning Tract-level characteristics to subwatersheds (compared to, say, taking the simple average value across all overlapping tracts and assigning that average value to the subwatershed).

B.3.4 Imputing race and socioeconomic characteristics within subwatershed

Our panel ranges from 1990 to 2020. To have coverage in non-census years, we impute the three race and socioeconomic characteristics within-subwatershed using the last observation carried forward (LOCF) method. For example, we assign a subwatershed's year 2000 white population share to observations for the same subwatershed in years 2001-2009 (and then use the 2010 census value for 2010). As another example, we assign a subwatershed's year 2000 income per capita to observations for the subwatershed in years 2001-2004 (and then use the 2005-2009 ACS value for the subwatershed in 2005). This sort of practice is extremely common in research using census data. After this imputation, we then adjust all dollar values to be in 2020$ using the CPI-U price deflator. Figure A2 in Appendix A presents the maps of white population share, income per capita (2020$), and median home value (2020$) at the subwatershed level, averaged across time during 1990-2020.

B.4 Local government finance data

We use the Census Bureau's Annual Survey of State and Local Government Finances and Census of Governments for local public finance data, conveniently organized into a single Government Finance Database (GFD, Pierson et al., 2015). GFD contains all of the Annual Survey of State and Local Government Finances and Census of Governments data from 1967-2019. The Census of Governments is conducted for the compete count of governments every five years starting from 1967. The annual survey uses a sample in the intervening years, e.g. the four years between 2012 and 2017. In this study, we use the data of six census years (1992, 1997, 2002, 2007, 2012, 2017) and survey years from 1990 through 2019, the most recent year of available GFD data. We focus on local governments by restricting the types of government to county, municipal and township. We take several steps to organize the public finance data at the subwatershed level, described just below.

B.4.1 Matching subwatersheds to counties, municipalities, and townships

As discussed in Section B.3 above, we spatially join subwatersheds and Census Blocks for the four decennial censuses (1990, 2000, 2010 and 2020). The subwatershed geodatabase comes from WBD, as discussed in Section B.1. The Census Block geodatabases are from IPUMS, as introduced in Section B.3. This provides the full set of Census Blocks that spatially intersect with each subwatershed for each of the four census years. By the structure of Federal Information Processing Standards (FIPS) codes, Census Blocks are nested within counties and Census Blocks are also nested within Census Places (municipalities and townships). So, the spatial join from subwatersheds to Census Blocks also provides a spatial join from subwatersheds to the three types of local government units. Note that, while counties do not spatially overlap with other counties, counties do overlap with Census Places, e.g. Harris County in Texas overlaps with the city of Houston, Texas. Thus, a subwatershed may be served by two "layers" of local governments: county(ies) FIPS and Census Place(s) FIPS.

B.4.2 Calculating revenue per capita measures for each local government

We focus on two types of government revenue: total revenue and total tax receipts. We isolate revenue from the impact of potentially endogenous intergovernmental transfers by using the "own source" version of the revenue measure. Excluding intergovernmental transfers also provides a better proxy for variation in *local* government capacity. We then calculate total revenue per capita (unit: $) and tax receipts per capita (unit: $), which are the total reported values divided by the population served by the specific government unit, using the population listed in the GFD data for each government unit by year. The result of this process is that for each (available) county FIPS by year and Census Place FIPS by year, we have the value for the total government revenue per capita and total tax receipts per capita. As with the Section 319 grants, we treat the fiscal year as the year of our subwatershed-year panel.

B.4.3 Local government financial data at subwatershed level

We assign a decennial census year to each county FIPS by year and Census Place FIPS by year to use in merging to the subwatershed panel. For example, we assign GFD data for years 1990-1999 a decennial census year of 1990. Using the subwatershed-FIPS crosswalk for each decennial census year, we merge the local governmental financial data to the subwatershed-year panel, separately for county FIPS and for Census Place FIPS. This links the public finance data to subwatersheds. At this point, an observation is either a subwatershed-year-county FIPS tuple or a subwatershed-year-Census Place FIPS tuple.

Because a subwatershed can overlap with the area served by multiple county governments and/or multiple municipality/township governments, we take the average of the per-capita revenue values by subwatershed-year at the county level and (separately) at the municipal/township level, which yields two panels of subwatershed-year observations. In other words, we determine the average revenue (and tax receipts) per capita for each "layer" of government for each subwatershed. We then add the county-level value and the municipal/township value for each subwatershed-year to obtain the total revenue per capita of local governments for each subwatershed-year in our panel.

There are 1,986,302 subwatershed-year tuples with government revenue data, covering 87,243 subwatersheds during years 1990-2019. We convert the two revenue per capita terms to 2020$ using the CPI-U price deflator. The average subwatershed-year total local government revenue per capita is 2,278.88 (2020$). The average for tax receipts per capita is 1,196.89 (2020$). Most of this difference is because total revenue includes revenue from specific public services, such as utilities and public transit. Finally, we multiply the per-capita values by subwatershed population to calculate the subwatershed-year-level total local government revenue and total local government tax receipts. The average subwatershed-year-level revenue is 24,394,930 (2020$) and the average tax receipts are 9,714,997 (2020$).

B.5 Water pollution data

This section describes how we collect and clean the data on water pollution used in this study. The data comes from three sources: EPA's Storet; EPA's Storet Legacy Data Center; and USGS's National Water Information System (NWIS), accessed via the Water Quality Portal (WQP). We use dissolved oxygen (DO) to quantify water pollution. As discussed in Keiser and Shapiro (2019b), DO has appealing properties as a measure of water pollution. Specifically, DO saturation is "among the most common omnibus measures of water quality in research, it responds to a wide variety of pollutants, and it is a continuous (rather than binary) measure of pollution." Following Keiser and Shapiro (2019b), we use a variation of DO saturation that they term as "dissolved oxygen deficit" (DO deficit), which is defined as 100 minus dissolved oxygen saturation and expressed as a percentage. Nonpoint source pollution increases dissolved oxygen deficits as microorganisms decompose pollutants. A larger DO deficit value indicates more polluted water.

We use several conversions to make the DO deficit measurements comparable across the three data sources, which have some variation in water pollution measurement, such as using different methods or units. For example, Storet Legacy and NWIS assign a single parameter code to each measurement, while Storet does not use a parameter code. We use a matching (crosswalk) table provided by EPA to make the measurements comparable across the databases. For Storet data, we also calculate "dissolved oxygen saturation" (unit: %) from "dissolved oxygen (DO)" (unit: mg/l), based on a formula described below.

B.5.1 Restricting to surface waters

We restrict the sample to surface waters based on the type of monitoring site, as listed in the databases. There is a minor difference between WQP and Storet Legacy in the sets of possible monitoring sites. In WQP, we restrict the site types to: Aggregate surface-water-use, Estuary, Spring, Lake, Reservoir, Impoundment, and Stream. In Storet Legacy, we restrict site types to: RESERV, ESTURY, LAKE, SPRING, STREAM, RUNOFF and IMPDMT. These restrictions exclude types of monitoring sites such as underground pools. We exclude sites at dams using regular expressions to test for the word “dam” in the station name.

B.5.2 Dissolved oxygen deficit

The EPA and USGS quantify water pollution using a variety of types of measurements. Storet Legacy and NWIS assign a single parameter code to each measurement. For example, 00301 indicates “dissolved oxygen saturation” with the unit of percent. The data also provides details about each parameter code, such as the Result Temperature Basis. The full list of parameter codes is provided by the USGS. However, no parameter codes are available in Storet. Instead, the EPA provides a matching table between Storet and NWIS. We use this crosswalk table to ensure that the DO deficit measures of water pollution we use are comparable across the three data repositories. We apply standard characteristic names “dissolved oxygen (DO)” and “dissolved oxygen saturation” to extract data from Storet.

For measurements of DO saturation, we can directly calculate DO deficit, which equals 100 minus DO saturation, measured in percentage points. We use a formula to convert DO with the unit of milligrams-per-liter to DO saturation (percent). The conversion formula is $DO_{percent} = \frac{DO_{mg/L}}{4.68/(31.5+Temp)}$, where $Temp$ indicates water temperature in Celsius. We then pool all DO deficit measurements from the three sources, along with the corresponding monitoring site locations (latitude and longitude). In a few cases, the same reading appears in both Storet and Storet Legacy, and we drop the duplicates. To limit the influence of outliers, we winsorize readings above the 99th percentile of the distribution to the 99th percentile and readings below the 1st percentile of the distribution to the 1st percentile.

B.5.3 Water pollution data at subwatershed level

In total, we have 8,365,470 unique readings for DO deficit in years 1990-2020. We link each DO deficit reading to the corresponding subwatershed (a polygon) based on the geographical location (latitude and longitude) of the monitoring site. Then we take the average by subwatershed and year to form a panel of DO deficit values at the subwatershed-year level. This panel has 313,399 nonmissing values for years 1990-2020, covering 39,143 subwatersheds. For subwatershed-year observations with missing (local) DO deficit readings, we assign DO deficit readings from the subbasin level (HUC8) to those subwatersheds that fall within the subbasin but are not covered by direct pollution readings. The DO deficit at the subbasin level is calculated by taking the average of all pollution readings that fall within the subbasin for each year. At this point, we have 1,987,095 nonmissing values for years 1990-2020, covering 91,932 subwatersheds. We then use the LOCF method to impute missing values over time within subwatershed.

After these steps, we have 2,984,711 nonmissing values of DO deficit, a nearly balanced panel for all subwatershed-years from 1990 to 2020. Whereas the full balanced panel for all of the United States' 102,943 subwatersheds spanning years 1990-2020 should have 3,191,233 observations, our nonmissing values of DO deficit account for 93.5 percent of the sample. In our main study sample (N=2,554,099), 27.57 percent of DO deficit values are determined using the LOCF imputation. We verify that the use of subbasin-level measurements and within-subwatershed imputation should not be a concern. For one, the average DO deficit for the subwatershed-year panel is 14.65 when using only the subwatershed-level readings, compared to 12.42 when also using subbasin-level readings, and 12.51 after the LOCF imputation. More broadly, we plot the density distributions for these three sets of DO deficit readings, as shown in Figure B1. The subwatershed-level readings have a somewhat thicker right tail—as expected, given these more granular measurements have a larger variance. However, the three distributions overlap closely. Additionally, we observe that DO deficit follows an approximately normal distribution, a very similar pattern as that documented in Keiser and Shapiro (2019b). Figure A4 in Appendix A shows the map of average DO deficit at the subwatershed level. There is large variation across the nation's subwatersheds.

B.6 Subwatershed downstream spillover measures

We construct three subwatershed-level measures to capture the potential for spillovers across subwatersheds. They are: 1) an indicator for whether a subwatershed has any downstream outflow into another subwatershed(s), 2) the total number of outflowing rivers or streams into another subwatershed(s), and 3) the total number of downstream subwatersheds, including those that are non-adjacent, conditional on the subwatershed having any outflows. To form these measures, we use data from the National Hydrography Dataset Plus (NHDPlus). The NHDPlus is a national geospatial surface water framework—a high-resolution map of the river and stream network—developed and maintained by the EPA in collaboration with USGS. We utilize the NHDPlus national seamless flowline geodatabase and a NHDPlus-provided crosswalk between flowline features and subwatersheds (HUC12) in the WBD.

There are 2,691,339 features in the national seamless flowline geodatabase, and each is identified by a unique "comid." These features form 1,013,033 "levelpathi." A levelpathi in the NHDPlus is what would more commonly be called a river or stream. Each levelpathi is composed of one or more comid, or river segments. For each comid, the data indicate the "pathlength" distance to the terminal feature downstream along the main path of the river. Thus, within the same levelpathi, a comid with a larger pathlength is comparatively more upstream. This also indicates the direction of water flow.

We use the NHDPlus-provided crosswalk to determine which subwatershed(s) each comid river segment intersects. After this merge, we have 1,012,930 intersecting levelpathi. We restrict the data to only rivers and streams by keeping flowline types designated as "streamriver" or "artificialpath", excluding other types such as "coastline." This restriction excludes 2.8 percent of the sample, leaving 1,056,454 unique subwatershed-levelpathi tuples covering 1,008,335 unique levelpathi. As these observation counts indicate, the vast majority of levelpathi exist in only a single subwatershed, consistent with how watersheds are defined. We drop the levelpathi that have only a single comid segment, which leaves 475,073 unique subwatershed-levelpathi tuples covering 426,954 unique levelpathi. In total, there are 19,348 rivers or streams that flow across or between multiple subwatersheds.

We determine whether a subwatershed has an outflow into another subwatershed(s) based on the sequence of river segments along the river or stream through subwatersheds, using the pathlength term discussed above. Figure B2 shows one example. San Fernando Creek in Texas flows through three subwatersheds. We highlight the Middle San Fernando Creek subwatershed as an example. Although various other flowlines are within this subwatershed, San Fernando Creek is the only outflow from the Middle San Fernando Creek into another subwatershed, i.e. the Lower San Fernando Creek subwatershed. Thus, for Middle San Fernando Creek subwatershed, the binary measure "whether a subwatershed has any downstream outflow into another subwatershed(s)" takes a value of one. The measure of the "total number of outflowing rivers or streams into another subwatershed(s)" is also one, as there is only one outflow (San Fernando Creek). The third measure is constructed by counting "the total number of downstream subwatersheds, including those that are non-adjacent." Again in the example, this value is also one, as there is only one subwatershed that is downstream of the Middle San Fernando Creek subwatershed. If the San Fernando Creek instead continued to flow beyond the Lower San Fernando Creek subwatershed and into other subwatersheds, then the first two measures would retain a value of one while the third would increase to count these additional downstream subwatersheds.

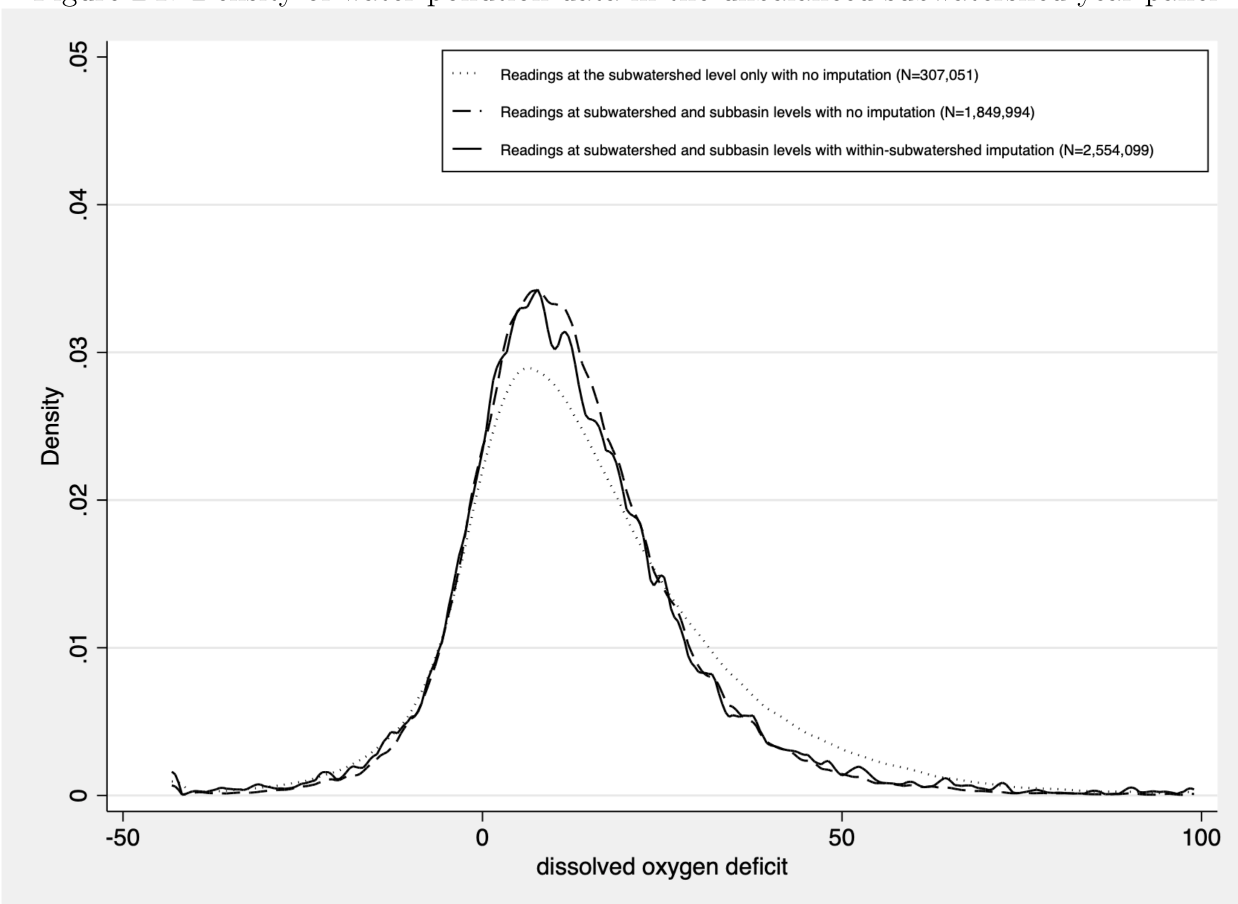
For each subwatershed, we calculate these three measures of downstream spillovers. Our final sample has 47,480 subwatersheds with outflow to another subwatershed(s). Conditional on having any outflows, the mean number of outflows downstream to another subwatershed(s) is 1.01, ranging from one to five. On average, 7.77 subwatersheds are downstream of a subwatershed with outflows, ranging from one to 267.

B.7 Forming the panel of subwatersheds as the study sample

The full balanced panel of all 102,943 subwatersheds spanning years 1990-2020 has 3,191,233 observations. Our primary (unbalanced) analysis panel has 2,554,099 subwatershed-year observations. We first restrict the sample to observations with census population. This restriction drops 537,615 observations, 84.4 percent of the total decrease. We require non-missing (imputed) pollution data, which drops another 92,752 observations, 14.6 percent of the total decrease. We also drop a handful of observations with missing income per capita or missing median home value. Finally, we drop the 31 observations for the District of Columbia, which has only one subwatershed—and hence no variation conditional on state-by-year fixed effects. Our study sample with 2,554,099 observations is around 80 percent of the original full balanced panel. Because the NHDPlus does not include flowline data for Alaska and Hawaii, we drop those two states for our spillovers analysis, keeping the 2,396,406 subwatershed-year observations located within the continental United States.

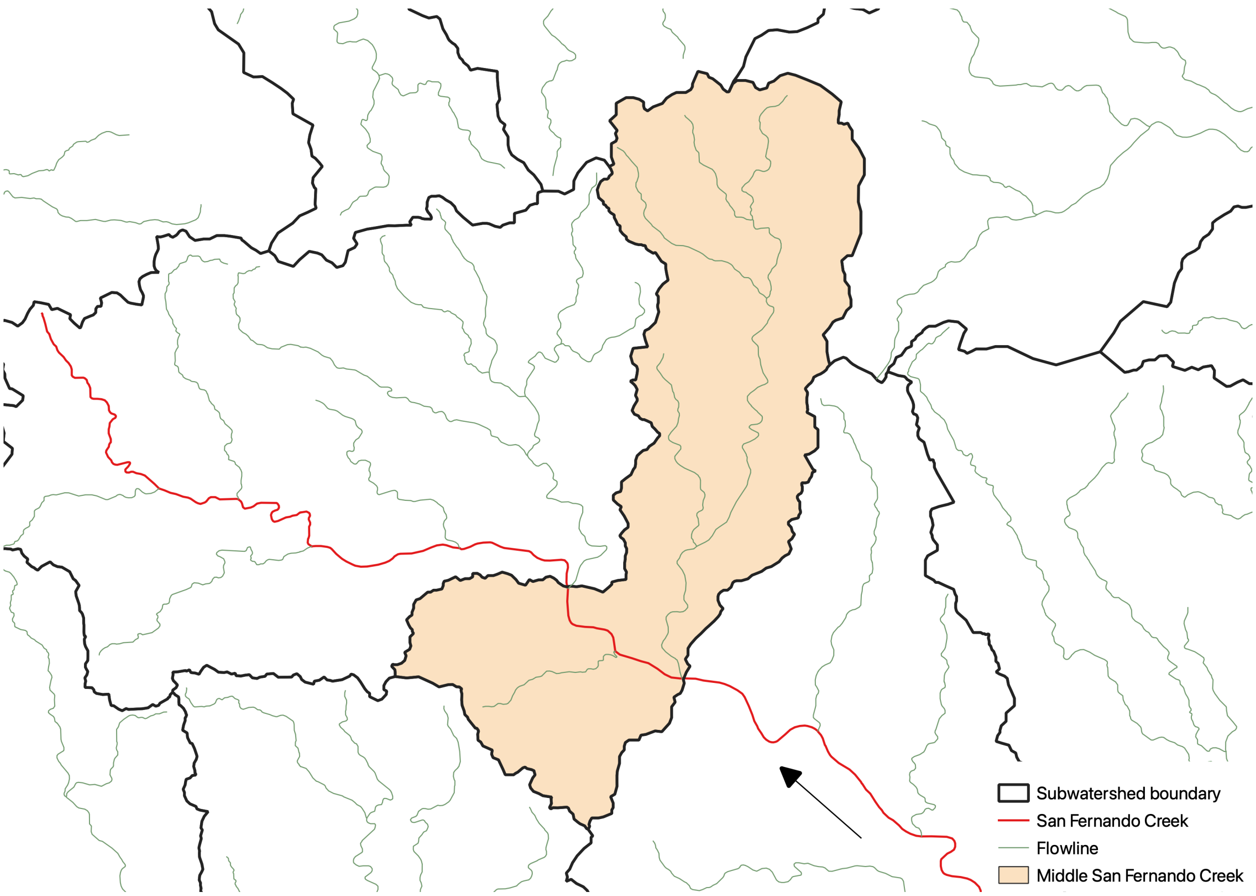
We use a subset of our primary subwatershed-year panel for the public finance analyses. As discussed in Section B.4, the Census of Governments is conducted for the compete count of governments every five years, and the Census Bureau surveys a sample of local governments in the intervening years. In total, 609,660 observations (23.9 percent) from the main study sample are not covered by the local government finance data, reducing the subwatershed-year count from 2,554,099 observations to 1,944,439 observations. The Census Bureau uses a stratified random sampling framework to include local governments in the non-census years, so subwatersheds' socioeconomic characteristics should be orthogonal to the probability of inclusion in this reduced panel. We additionally verify that our main results showing the relationship between socioeconomic characteristics and grant funding are similar when using this reduced panel that has available public finance data.

Figure B1: Density of water pollution data in the unbalanced subwatershed-year panel



Notes: Data covers years 1990-2020. Dissolved oxygen deficit equals 100 minus dissolved oxygen saturation, measured in percentage points. A negative deficit value is possible since dissolved oxygen saturation can exceed 100 percent due to photosynthetically active plants and algae.

Figure B2: Map of Middle San Fernando Creek subwatershed in Texas



Notes: The figure shows a map of Middle San Fernando Creek subwatershed (in light orange) in Texas. The red line is the San Fernando Creek flowline, and the green lines depict other flowlines (streams and river segments). The black arrow indicates the flow direction of San Fernando Creek.