Phillip Cork
Final Report on “Investigating Redlining's Impact on Safe Drinking Water Access”

Executive Summary

Through the Justice40 Initiative, the Biden Administration has made historic investments to address environmental injustices and the climate crisis while prioritizing historically marginalized communities. With $50 billion set aside for improving access to safe drinking water, it is crucial these funds go to the communities that need them most. While safe drinking water was recently added as a consideration for the Justice40 Initiative in its 1.0 release as of November 2022, there remains a lack of comprehensive data about where communities get their water and to what extent these water systems have been historically funded to relieve environmental injustices nationwide. To this end, this report considers the potential usefulness of deploying historic redlining data as a proxy for the flow of historic funding to identify ways to prioritize improving safe drinking water access in the future. It then concludes with a pair of recommendations: improving the data collection process and using currently available indicators to ensure funding reaches prioritized communities effectively and efficiently.

Data Visualization

The analysis and data visualizations discussed in this report can be found and explored at https://philcork.shinyapps.io/waterquality_redlining/.

Background

Environmental Justice

Shortly after taking office, President Biden signed Executive Order 14008, *Tackling the Climate Crisis at Home and Abroad.* Among other items, this executive order launched the Justice40 Initiative, earmarking at least 40% of federal funds going towards addressing the effects of environmental injustice and climate change in disadvantaged communities. With the administration’s commitment to addressing both systemic and lingering causes, there is an unprecedented opportunity to improve the lives of those who have unfairly bourn the burden of environmental inequity for generations.

The first step in making sure funds flow to the right communities is identifying which communities should be denoted as disadvantaged in the first place. To this goal, the Council for Environmental Quality were charged with this task and recently published version 1.0 of their open source CEJST tool, which identifies communities as disadvantaged based on meeting one of several climate and environmental vulnerabilities thresholds and a pair of economic criteria.

Safe Drinking Water

One facet of environmental justice the updated CEJST tool takes into account is access to safe drinking water. With more than $50 billion in federal funds flowing to improving water quality

and replacing lead pipes through laws earmarked for Justice40 consideration, it is crucial to identify where these funds should go to help the communities most in need. One hurdle in this process is that funding and reporting for utility infrastructure occurs at the community water system (CWS) level, often not reporting the communities who will benefit from the investments or those who are harmed by insufficient drinking water quality.

Redlining

The historic practice of redlining continues to have disproportionate negative affects on communities of color and other minority populations. The roots of this form of intentional, systemic disenfranchisement date back to the 1930s, when the Home Owners' Loan Corporation created maps of cities and towns across the country, grading portions of the city based on their "mortgage security." The neighborhoods deemed to be of the lowest risk when lenders decided who should receive loans were denoted with an "A" grade and color-coded in green. "Still Desirable" neighborhoods were depicted in blue with a "B" grade and "Definitely Declining" areas received a "C" grade and were coded in yellow. Meanwhile, the neighborhoods receiving "D" grades were labeled as "hazardous" and color-coded in red.

Lenders were considered prudent to refuse loans in these areas or to only provide investment on a conservative basis. These grades took into account not only the area's mix of residential and industrial development, but the racial and ethnic makeup of its communities as well. These discriminatory loan practices, grounded in rampant, systemic racism, directly prevented the accumulation of intergenerational wealth and stymied the flow of both private and public capital. Further, the legacy of redlining can be observed across a number of environmental vulnerabilities including flood risk, extreme heat, and others.

Data

This project combines data regarding drinking water quality, community water systems, and historic redlining practices. This report provides a brief overview of the sources, accessibility, and collection methods for each data source before considering possible bias and sampling concerns.

Sources

Redlining and census data is available through fivethirtyeight's GitHub repository. In their article, The Lasting Legacy of Redlining, the authors examine the redlining maps created in the 1930s and 1940s and compare their evaluations of each neighborhood in metropolitan areas to the current demographic distribution of these areas based on 2020 census data. Fivethirtyeight originally downloaded the redlining data from the Mapping Inequality Project. Additionally, data regarding the locations of community water systems is available from the Environmental Policy Innovation Center (EPIC). Finally, the EPA makes data regarding the condition and quality of drinking water available through the Safe Drinking Water Information System, or SDWIS.

License & Privacy
Fivethirtyeight and Mapping Inequality make their data available through the Creative Commons license system in which sharing and adapting the dataset is permissible with required attribution. The license also prevents derivative works from applying additional restrictions. The Service Area Bounds dataset provided by EPIC is available under the MIT open source license. All other data is publicly available through the Census Bureau or EPA. Given that each data source is both publicly available and conducts analysis at the neighborhood unit with population estimates from the census, there are not any pertinent privacy concerns regarding combining these datasets or potentially including any personally identifying information.

Collection Methods
For this project, each dataset is either collected directly from the source where the data was originally produced, (the EPA's "SDWIS Fed Reporting Services system", the Census Bureau's API, and Mapping Inequality Project's website), or from publicly available Github repositories, (fivethirtyeight's data and EPIC's community water system shapefiles).

Biases & Sampling
It is worth noting that analysis will be limited to locales where there is both high quality community water system data (referred to as "Tier 1" regions in the EPIC dataset) and metropolitan areas analyzed in the original fivethirtyeight article. As such, it is possible the subset dataset is biased in that it is limited to states in which data quality is higher, which may be meaningfully distinct from states in which data quality is of lesser quality or missing altogether.

Methodology
To prepare these datasets for analysis and presenting the subsequent results, a number of data cleaning, transformation, and visualization techniques were conducted.

Data Cleaning & Transformation Methods
The SDWIS dataset is recorded at the individual violation unit of analysis and thus needed to be aggregated up to the community water system unit of analysis. The data also needed to be filtered to only include health-related violations over the last five years so that comparatively minor administrative or filing errors are not prioritized in the analysis. Finally, the health violations per community water system were rolled up into a binary variable denoting whether a given water system reported a health violation lasting more than thirty days in the last two years. Other minor data transformations were used to format and simplify the dataset to be used in data visualization following the initial data exploratory analysis process.

Since the service area bounds dataset is being used primarily for its geospatial component, there was minimal data cleaning required outside of reducing the dataset to regions with high quality data (Tier 1) where redlining data also exists (metropolitan areas). Because it is released in part with an article detailing their analysis, the Fivethirtyeight dataset, which combines redlining

spatial data with 2020 census data, was already cleaned, particularly in comparison to most publicly available datasets.

Instead, the most critical data transformation involving these two geospatial datasets was the use of the custom 'partial_assignment' function. Taking in two geospatial data frames as parameters, this function returns a data frame that lists each overlap between the two sets of observations and the extent of the area they share in common. This is how the project evaluates the overlap between the SAB and redline dataset. For each water system, the function considers which neighborhoods it overlaps with, the neighborhood's grades, and the total area of the water system. These features can then be aggregated to determine how the area of the water system is divided up by various HOLC grades for neighborhoods that overlap with it.

Data Visualization Methods

The final presentation of this project includes three data visualizations, all made in R and made publicly available as a Shiny app. Of these visualizations, two are interactive bar plots, both made with the popular ggplot package, while one is an interactive map made with the leaflet package. The first bar plot highlights key statistics regarding health violations aggregated up to the state level, where sufficiently high quality state data exists. In this interface, users can choose whether to display one of four statistics. Next, an interactive map presents a geospatial interpretation of the overlap between community water systems and the historic redlining dataset for each urban area in Texas included in both datasets. This map allows users to more intuitively explore the overlap between the two geospatial datasets used in this project, observing firsthand how the water systems and HOLC grade neighbors intersect with one another. Finally, the second bar plot depicts community water systems within a given state, highlighting the breakdown of how much each system overlaps with HOLC-graded communities in a more comparative analysis. Users can compare across three states to see how historic redlining practices may or may not correlate with the presence of persistent health violations in water systems.

Talking Points & Recommendations

Considering this process and the resulting data visualizations, a few takeaways stand out for trying to answer whether comparing redlining practices with community water system data can help identify communities who should be prioritized for additional investment.

First, different trends in where health violation appear across even just a small subset of states like Texas, Utah, and Pennsylvania, suggesting that, at least nationally, there may not be a clear, definitive correlation between redlining practices and a lack of safe drinking water. Thus, this sort of analysis may be more useful in some areas than others and could potentially serve as an additional layer of consideration. Without additional contextual data, it likely cannot serve as a singular tool for guiding how to further invest in marginalized communities, even through just the lens of improving access to safe drinking water.

Second, it is worth considering the extent to which a majority of water systems serve communities of diverse HOLC grades, at least where the grading maps exist. This reality perhaps explains why some water systems that overlap with a large portion of historically redlined neighbors don't experience health violations - because they also serve communities that were more invested in and supported as well. However, it is also worth noting that this data only considers urban areas, leaving out rural communities which are likely underfunded for different reasons and often have less capacity to deal with safe drinking water issues.

Finally, these two points suggest an even more pressing issue in the decision-making process surrounding the pursuit of environmental justice - improving data collection, data access, and data quality. This analysis has been limited to a small subset of states because data regarding community water systems and whom they serve is severely lacking. The EPA's SDWIS data is considered notably incomplete and often relies on proposals and applications, rather than verified estimates about the communities they serve. And that is to say nothing of the data that isn't collected at all, but would make this analysis and other endeavors exponentially more effective.

Recommendations

With these findings in mind, there are two recommendations.

First, local, state, and federal governments should commit to enhancing data gathering capabilities and increased transparency. This analysis is decisively hindered by many states having insufficient data quality necessary for such work. It is possible a more definitive relationship would be present were the data more comprehensive or that more targeted investment recommendations could be made if the data were more granular and precise.

Second, with the data that is available, decision makers should deploy it to prioritize investing in safe drinking water where communities have been typically marginalized. In this case, there are clearly a handful of cities in Texas where historically under-invested communities lack consistent access to safe drinking water. The state can and should provide resources to those communities to ensure they apply for grants and are able to receive part of the historic investments being made in drinking water infrastructure. Thankfully, this process is underway, now that both water pollution and redlining are being folded into the Justice40 Initiative definition of whether a community will receive priority funding.

Conclusion

The Biden administration has made admirable steps in pursuing environmental justice and tackling the existential crisis of climate change. The scope of these challenges require broad-reaching initiatives and holistic solutions powered by decisions informed by the best available data. While the tools and methods for these efforts are being developed, it is crucial to continue moving forward and implementing changes with the currently available resources. This report shows that while an imperfect measure, the historic redline practices can serve as a starting point worth considering for ensuring federal funds are improving lives of those most in need.