Title

Underrepresented, understudied, underserved: Gaps and opportunities for advancing justice in disadvantaged communities

Authors

Fernandez-Bou, Angel Santiago
Ortiz-Partida, J Pablo
Dobbin, Kristin B
et al.

Publication Date

2021-08-01

DOI

10.1016/j.envsci.2021.04.014

Underrepresented, underserved, understudied: gaps and opportunities for advancing environmental justice in disadvantaged communities

Preprint · December 2020

DOI: 10.13140/RG.2.2.19036.41606

CITATION
1

READS
629

6 authors, including:

Angel Santiago Fernandez-Bou
Union of Concerned Scientists
31 PUBLICATIONS **208** CITATIONS
SEE PROFILE

Jose Pablo Ortiz Partida
Union of Concerned Scientists
39 PUBLICATIONS **469** CITATIONS
SEE PROFILE

Humberto Flores-Landeros
University of California, Merced
13 PUBLICATIONS **38** CITATIONS
SEE PROFILE

Kristin B. Dobbin
University of California, Los Angeles
15 PUBLICATIONS **144** CITATIONS
SEE PROFILE

Underrepresented, underserved, understudied: gaps and opportunities for advancing justice in disadvantaged communities

A.S. Fernandez-Bou[1]†*, J.P. Ortiz-Partida[2]†*, K.B. Dobbin[3], H. Flores-Landeros[1], L.A. Bernacchi[4], J. Medellín-Azuara[1]

Affiliations:

[1]Civil and Environmental Engineering and Environmental Systems Graduate Program, University of California, Merced, CA 95343, USA.

[2]Climate and Energy Program, Union of Concerned Scientists, Oakland, CA 94607, USA.

[3]Department of Environmental Science and Policy, University of California, Davis, CA 95616, USA

[4]Center for Information Technology Research in the Interest of Society (CITRIS) and the Banatao Institute, University of California, Merced, CA 95343, USA

† These authors contributed equally to this work

*Corresponding author. Email: afernandezbou@ucmerced.edu (A.S.F.B.); jportiz@ucsusa.org (J.P.O.P.)

Highlights

- Media, legislation, and science poorly represent the concerns of disadvantaged communities.
- Developing effective policies requires addressing nuances and issue co-occurrence.
- Community-specific knowledge is necessary to advance sustainable, effective solutions.

Abstract

A common approach in scientific research and policy is a commitment to develop projects or legislation trying to improve problems experienced by rural communities; however, lack of interaction with community members during the process tends to produce unsatisfactory results. We visited disadvantaged communities in the San Joaquin Valley of California and interviewed local stakeholders (community members and leaders, policy advocates, attorneys, and educators). Then we analyzed a corpus related to disadvantaged communities from a pool of California-related publications containing 154,000 scientific papers, 2.6 million newspaper articles, and 11,000 state legislation bills from 2017 to 2020 to estimate the frequency and quality of disadvantaged community representation. Here we present our findings describing the biases and gaps of knowledge by scientific papers, California newspaper articles, and legislation bills with respect to disadvantaged communities in California, and we suggest opportunities for scientists, media communicators, and policymakers to amplify the voices of these stakeholders. In all corpus categories, disadvantaged communities are underrepresented: about one in four Californians live in disadvantaged communities, but only one in 2000 news articles and scientific papers cover them. The concerns and priorities of disadvantaged communities do not match the public perspective of them depicted by the corpus. Developing effective policies requires addressing place-specific nuances and co-occurrence of structural inequities in partnership with local stakeholders. Holistic coverage in newspapers and community-based approaches may increase awareness and understanding of disadvantaged communities, helping tailor policy solutions and building the political leverage needed to implement them.

1. Introduction

Rural disadvantaged communities in California experience a disproportionate share of the most pressing social, environmental, and economic challenges. These challenges co-occur creating a compounding effect that leads to structural conditions of extreme inequity that are more complex than just the sum of their parts. The origin of the disadvantage often contains elements of racism, discrimination, and segregation that resulted in inequitable opportunities and interfere with such essential issues as their health, education, and overall well-being (Almaguer, 1994; Anderson, 2009; Eissinger, 2017, 2008; OEHHA, 2018; Pannu, 2012).

In California, disadvantaged communities are formally defined by their performance (worse 25 % percentile) of a score (CalEnviroScreen score) that considers several indicators of pollution burdens and population characteristics (De León, 2012; OEHHA, 2017). This term has been formally defined and is widely adopted to facilitate discussion of these overburdened and underresourced communities with widespread use across science, media, and policy (usage often mirrors "environmental justice community" or "vulnerable community" in other parts of the world). We focus on the San Joaquin Valley, the region with the largest concentration of rural disadvantaged communities in California, many of which are surrounded by the region's dominant industry, agriculture. The San Joaquin Valley has been described as a region where "Flint is everywhere" (Real, 2019) after the case of lead poisoning in tap water that disproportionally affected low-income communities and racial minorities in Flint, Michigan (Butler et al., 2016). The region contains 413 census tracts (2.2 million people) under the disadvantaged community designation (OEHHA, 2017). San Joaquin Valley disadvantaged communities have endured well-documented social, economic, environmental, and public health crises: lethal air and water quality (Balazs Carolina et al., 2011; Balazs and Ray, 2014), entrenched poverty, lack of educational opportunities (De Vore, 2008), low life expectancy (Tejada-Vera et al., 2020), health disparities (Kissam, 2020), and linguistic and social isolation (Gifford and Valdés, 2006). Despite ongoing work, efforts to address persistent inequalities in the region have consistently fallen short. For example, a 2013 report by PolicyLink (a national research and action institute advancing racial and economic equity) detailed the lack of fundamental features and infrastructure, such as safe and affordable drinking water, sewer systems, safe housing, public transportation, parks, sidewalks, and streetlights (Flegel et al., 2013). Such conditions have been documented in the valley for decades, if not a century (Eissinger, 2008), and remain in 2020.

Rural disadvantaged communities of California are neglected in essential issues such as water resources management and infrastructure (Allaire et al., 2018; Bernacchi et al., 2020; Scott et al., 2020; Ulibarri et al., 2017). They also lack political leverage since many are unincorporated and unable to vote for local politicians, and they are outnumbered by other entities at the county level (Anderson, 2009). Some disadvantaged communities have low population sizes but relatively large capital investments in water infrastructure, leading to very high water bills for some of the lowest income communities to pay for water deemed unsafe to drink (Bland, 2018). Then, some infrastructure investments are abandoned when maintenance and operation costs become too expensive for the communities to sustain. Consider the case of Lanare (Fresno County), where the community received a $1.3 million water treatment plant to remove arsenic from their drinking water that went offline after a few months because the community could not afford to operate it (Ezra David and Klain, 2017).

California needs more effective and sustainable policy solutions, and such solutions must be supported by a robust understanding of these communities. The sources of power and influence with text-based records are (1) scientific papers, (2) newspaper articles, and (3) legislation bills (Likens, 2010; Shanahan et al., 2008). Investments to solve fundamental inequities occurring in disadvantaged communities are the responsibility of policymakers. Policymakers need science-based information to make correct decisions, and lack of scientific work in these communities directly limits their capacity to act. While policymakers can offer more funding for science, scientists may not be well informed about the reality of disadvantaged communities if the coverage of news media is not enough (Likens, 2010; Shanahan et al., 2008). Inadequate information leads to ignorance by the general public who, in turn, do not exert leverage over policymakers to solve those issues faster and more effectively. That may make policymakers consider disadvantaged communities a less urgent topic than others that are more often demanded by their voters. Adequate representation in textual power can guide scientific knowledge to inform policy and investments to serve disadvantaged communities.

In this study, we developed quantitative and qualitative metrics based on semi-structured interviews with disadvantaged community stakeholders. We analyzed a corpus of California scientific papers, newspaper articles, and legislation bills with two objectives: (a) quantifying the frequency of representation of disadvantaged communities of California across the three platforms, and (b) assessing the quality of those representations. We ask the following research questions: (1) How often is the term "disadvantaged community" represented across the three platforms? (2) And to what extent does the coverage represented by these platforms align with the concerns from disadvantaged community stakeholders themselves?

2. Methods

2.1. Study location and focus

We focused our interviews on rural disadvantaged communities of the San Joaquin of California. The valley is enclosed by the Coastal and the Sierra Nevada Mountain Ranges, with a length of 430 km from Bakersfield in the south to the Delta (east of the San Francisco Bay). This region is one of the most productive farmlands in the world, with more than 20 000 km^2 of irrigated farmland (Hanak et al., 2019). The San Joaquin Valley has a population of 3.97 million people (U.S. Census Bureau, 2019), with about 2.2 million people (55 % of the total valley's population) living in 413 communities classified as disadvantaged (OEHHA, 2017). Besides, the region has a large amount of "hard-to-count" residents, including Latinxs, immigrants, low-income families, and other vulnerable individuals who live in disadvantaged communities and make the actual population higher than the official count (Latino Community Foundation, 2018).

Legislation by De León in Senate Bill 535, 2012 (De León, 2012), and Gomez in Assembly Bill 1550, 2016 (Gomez, 2016), identify disadvantaged communities in California and requires a certain amount of funds to benefit them. Senate Bill 535, 2012, requires the California Environmental Protection Agency to measure geographic, socioeconomic, public health, and environmental hazard criteria to identify disadvantaged communities in California. Assembly Bill 1550, 2016, establishes that a minimum of 25 % of the moneys available in the Greenhouse Gas Reduction Fund are invested to benefit disadvantaged communities. This legislation and the subsequent creation of the "California Communities Environmental Health Screening" tool (CalEnviroScreen) to identify disadvantaged communities by the California Environmental

Protection Agency are important landmarks. For the purpose of this study, we adopt the CalEnviroScreen 3.0 definition of disadvantaged communities as census tracts that perform in the 75 percentile or higher (worse) of the CalEnviroScreen score (before CalEnviroScreen, the definition of disadvantaged communities in California was based only on income). This score considers two broad groups: (1) pollution burden, subdivided in exposures (ozone, particulate matter 2.5 µm, diesel emissions, contaminants in drinking water, pesticides, toxic releases, traffic density; this component represents 33.3 % of the final score) and environmental effects (cleanup sites, groundwater threats, hazardous waste, impaired water bodies, and solid waste sites; this component represents 16.7 % of the final score), and (2) population characteristics, subdivided in sensitive populations (asthma, cardiovascular disease, and low weight at birth; this component represents 25 % of the final score) and socioeconomic factors (education, housing burden, linguistic isolation, poverty, and unemployment; this component represents 25 % of the final score). Each of the indicators is given as percentile of the studied census track compared with the rest of the state, and the indicators of each component are averaged to generate the components value. The weighted components result in the CalEnviroScreen score for each studied census tract, and a census track receives the disadvantaged status when its score is between the 75^{th} percentile and the 100^{th} percentile.

2.2. Interviews

We conducted 18 interviews (9 in English and 9 in Spanish) with community leaders and residents, local politicians, public servants, and specialists affiliated with nonprofit organizations and NGOs that work directly with multiple communities. We employed a snowball sampling approach, and we recruited participants at community outreach events, in committees, and by personal recommendations. During each interview, we asked broad questions about environmental risk and socioeconomic problems perceived by the interviewees (Table S3). Our semi-structured interview protocol was designed to collect spontaneous responses regarding broad perceptions of environmental problems (Adams, 2015). Socioeconomic questions included the topics of the interviewee's relationship to the community and its perceived demography, climate change perceptions, employment in the community, food access and security, and representation in policy-making decisions. Environmental justice questions covered topics of water quality, water quantity, drought vulnerability, floods, and air quality.

Interviews lasted between 20 minutes and 90 minutes, with an average of 50 minutes, and we recorded via handheld audio recording device. The audio was transcribed using Sonix.ai, an online transcription service. We reviewed each transcript to remove transcription errors and removed personal identifiers for each interviewee. We used the resulting transcripts in our qualitative analysis of the interviews.

The interviews were manually coded using the NVivo 12 Plus software (QSR International, Doncaster, Australia), and multiple topics were considered (Table S4). The codebook was developed based on the questions (Table S3) and complemented by emerging topics. By using an *a priori* codebook, the interviews were organized from different topics into clear categories: agriculture, air quality, climate change, impacts, non-environmental and environmental issues, policy, social characteristics, and water concerns. The categories were not mutually exclusive, and each sentence could be coded into multiple categories. Each interview was read and coded at the sentence level to identify what we call high-resolution categories (see section 2.4.2). High-

resolution categories illustrate specific challenges (for example, dependence on bottled water) that cannot be addressed with general categories (water issues).

2.3. Corpus selection

The corpus included publications regarding disadvantaged communities in California in general or in the Central Valley of California in Scientific papers, Newspaper articles, and Legislation bills (Table S1). The time frame studied was from January 1st, 2017, to May 31st, 2020. Each of these platforms could be independently analyzed due to their intrinsic unique characteristics, but here we elected to study them together because of the influence they exert on each other (Likens, 2010; Shanahan et al., 2008).

2.3.1. Scientific papers

We preselected all the scientific papers (research, review, and short communications) published between 2017 and May 31st, 2020 in Elsevier (sciencedirect.com) and Springer (link.springer.com) containing the word "*California*" to analyze the representation of the state of California in those databases. The preselection of articles included work conducted by California-funded researchers (for example, from the University of California), and equipment and software manufactured in California because these represent the intellectual wealth that California exports. Within that preselection, we searched for all the articles containing the expression "*disadvantaged communities*" and variations (for example, in singular and plural, or adding "unincorporated") in any part of the document (n = 198). Then, we filtered that subset of articles utilizing keywords to analyze if the article was addressing issues related to disadvantaged communities in the abstract (such as disadvantage, vulnerable, poverty, and low-income; see Table S5 for more details). We read approximately half of the articles to validate the accuracy of the keywords identifying relevant articles. Of the relevant articles (n = 68), we utilized all the titles, keywords, and abstracts of each for category-based analysis (see section 2.4).

2.3.2. Newspaper articles

Newspapers represent public access to information. As a textual body, newspapers have an outsized influence on public perception of environmental issues, risk, and health (Carvalho, 2010; Killingsworth and Palmer, 2012). We queried the Newsbank database of California Newspapers (University of California Library) for the term "*disadvantaged communities*" and variations. Of the 240 newspapers in the database, only 149 newspapers were included in the study, since 91 newspapers did not mention disadvantaged communities during the studied time frame. We found 1,440 articles that we reviewed to remove duplicates (same article in different journals), and to exclude those articles that were clearly referring to only specific urban neighborhoods. We conducted analysis on the full-text and titles of 511 newspaper articles.

2.3.3. California Assembly and Senate legislation bills

We preselected all the bills published on the California Legislative Information portal (http://leginfo.legislature.ca.gov) containing the expression "*disadvantaged communities*" in the keywords field of "Bill Search" for the periods 2017-2018 and 2019-2020 (n = 240). We removed bills with more than 100,000 words since they were likely budget bills and the analysis would not improve the results based on this methodology. Then, we located the reference to

"*disadvantaged communities*" (and variations) in the bill and excluded bills in which the term "*disadvantaged communities*" was only a definition but without a context of action towards them; if the reference was in the title or in the legislative counsel's digest of the bill, we utilized the whole bill for in-depth analysis; if the mention was in the body and not in the digest, we selected the section or sections in which it appeared, unless the mention was in the title of a law chapter or article, in which case we selected the whole chapter or article, even if it had several sections. The final number of bills selected (in whole or sections of them) was 210.

2.4. Analysis of publications

2.4.1 Theoretical low-resolution categories

We developed a theoretical framework based on the Regional Opportunity Index (Benner et al., 2014) and CalEnviroScreen (OEHHA, 2017) that contained six low-resolution categories: *Health*, *Economy*, *Education*, *Housing*, *Infrastructure*, and *Civic Life*. The low-resolution analysis serves to classify publications in broad categories that describe the opportunities in disadvantaged communities (Table S2). The classification into each of the categories was conducted by identifying the presence or not of specific keywords associated to each category (Table S2). The keywords were obtained from the metadata associated to the Regional Opportunity Index and from CalEnviroScreen 3.0. The six categories are not mutually exclusive.

2.4.2 Interview-based high-resolution categories

The high-resolution analysis aims to code the corpus documents using knowledge learned from the interviews with disadvantaged community stakeholders. This means that to develop this framework, we analyzed the documents from stakeholders' perspectives using their first-hand experience.

We performed a qualitative and quantitative analysis of the interviews that yielded 20 high-resolution categories covering specific priorities from communities' members. Community-specific issues would have remained unidentified without visiting communities and talking with their members and other stakeholders. High-resolution analysis cover categories such as flooding problems, dry wells, and dependence in bottled water, among others (Table S5). We selected keywords (Table S5) based on these categories to quantify their frequency within the three platforms (scientific papers, newspaper articles, and legislation bills). When appropriate, the keywords included their variations (for example, education, educational, educated) to better capture the representations. We assumed that the higher the frequency in each platform, the better their understanding of specific local concerns.

We conducted a cluster analysis to depict how the publications from the three studied platforms (scientific papers, newspaper articles, and legislation bills) represented the high-resolution categories. We utilized the Silhouette method to optimize the number of clusters. We aimed to find patterns in the representation of the high-resolution categories across the different platforms.

3. Results and discussion

3.1. Disadvantaged communities are overwhelmingly underrepresented

In California, 9.4 million people live in disadvantaged communities, representing 25 % of the State's population. In the San Joaquin Valley, 2.2 million people live in disadvantaged communities, representing 55% of the San Joaquin Valley's population (OEHHA, 2017).

Yet, in a sample of 154,000 scientific papers regarding California from 2017 to 2020, only 68 referred to disadvantaged communities. In a sample of 2.6 million news articles from 240 newspapers, only 1,440 articles mentioned disadvantaged communities, and 91 newspapers did not ever mention disadvantaged communities in the study period. Roughly speaking, the ratio of Californians to Californians living in disadvantaged communities is 4:1 (25 %); the ratio of the science and news writing about it is 2000:1 (0.05 %). The ratio of Californians to Californians without safe access to water is 40:1 (2.5 %); the ratio of the science and news writing about water issues in disadvantaged communities is 10,000:1 (0.01 %).

Representation in policy is more expansive. Of the 11,000 bills analyzed, 211 mentioned disadvantaged communities (ratio of 50:1), although often the mentions did not require any action that benefited disadvantaged communities. Looking more closely at representation in scientific papers (from Elsevier and Springer), none of the articles mentioned interviews to residents from disadvantaged communities in the San Joaquin Valley (or the greater Central Valley) regarding their socioeconomic and environmental concerns. Thus, besides the limited coverage of disadvantaged communities, first-hand information from residents is not surfaced.

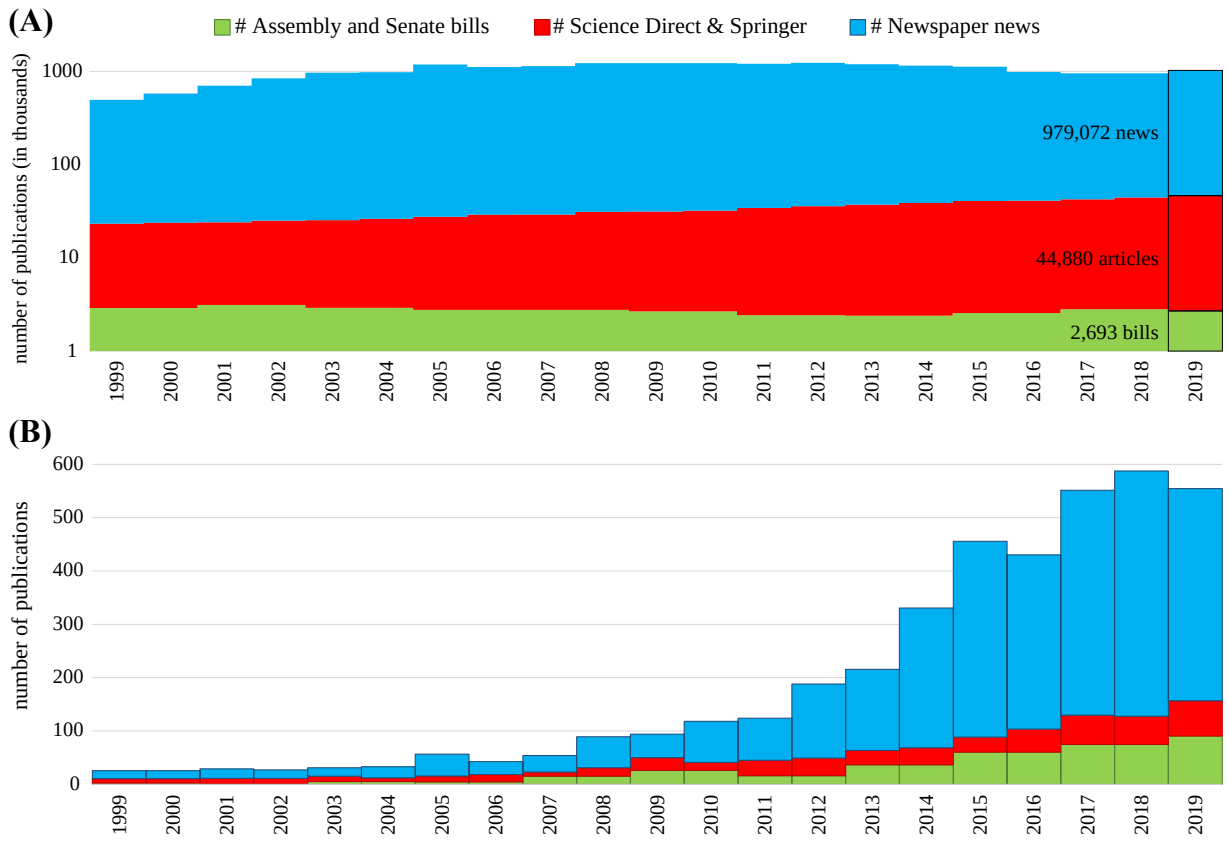

Fig. 1. Number of publications in the three studied platforms. (A) Total publications and (B) publications containing the expression "disadvantaged communities."

While the mention of disadvantaged communities in the three studied publication platforms has increased over the last two decades (particularly in the news media), disadvantaged communities remain largely underrepresented in all three platforms (Fig. 1). Despite increasing inclusivity, disadvantaged communities of California are not yet properly served by policies. Discussion about them is far dwarfed by their actual prevalence, and policies implemented to serve them cause sometimes negative effects (for example, Balazs and Lubell, 2014; Bernacchi et al., 2020; Cushing et al., 2018; Dobbin, 2020; Dobbin and Lubell, 2019; Goddard et al., 2021; Shonkoff et al., 2011). Still, that representation in legislation bills far outpaces representation in scientific papers and newspaper articles, pointing to a significant gap between the need for solutions and the knowledge and attention that can be leveraged to achieve them. In our analysis, about 2 % (211 bills) mentioned disadvantaged communities. If 2 % is not enough representation for disadvantaged communities to be properly served by policy (underserved), then 0.05 % of representation in the news and in scientific papers is subsequently not enough (underrepresented and understudied).

However, it is not possible to set a fixed threshold about how much representation is enough; such a threshold varies depending on the necessity. Disadvantaged communities often experience the burden of oppression and injustice, and therefore they have greater attention needs than non-disadvantaged communities (the distinction of equality versus equity). Taking the feminist notion of "centering the margins", 2 % of legislation bills, and 0.05 % in scientific papers and in newspaper articles is a clear obstacle to progress. The negligence shown by these three independent platforms towards disadvantaged communities is the result of their structural biases and the reciprocal influence that they exert on each other.

3.2. There is a gap between most concerns of disadvantaged communities and their representation in media, legislation, and research

Co-occurrence of problems exacerbates their consequences, and it is essential to understand the holistic context in which disadvantaged communities live. Our second objective was to study how well those platforms understand the problems, needs, and concerns of disadvantaged community residents. Our interviews with stakeholders provided valuable first-hand knowledge about environmental threats and socioeconomic challenges in their communities and possible solutions. The qualitative and quantitative analysis of the interviews resulted in 20 high-resolution categories that we grouped thematically using low-resolution categories drawn from the Regional Opportunities Index and CalEnviroScreen: *Health*, *Economy*, *Education*, *Infrastructure*, *Housing*, *Civic life*. Then we utilized keywords to measure the frequency of such categories in the three platforms. We found that the three platforms have a higher frequency in more generic categories (generic topics brought up by the interviewees, such as *Air quality*, *Water quality*, or *Education*) and lower frequency about specific issues that were raised multiple times during interviews (Fig. 2, Table 1).

An example of low-resolution and high-resolution problems is the distinction between "water problems" and specific issues that emerge from breaking down 'problems' into its components such as "dependence on bottled water", "cost of water", or "wells getting dry". The distinction is

important because high-resolution issues vary across communities, and solutions may need fundamentally different approaches. Consider the dependence on bottled water (mentioned by 94 % of the interviews): for some communities is because of water scarcity while in others is because the available water is toxic. Effective solutions to stop the reliance on bottled water may be very different.

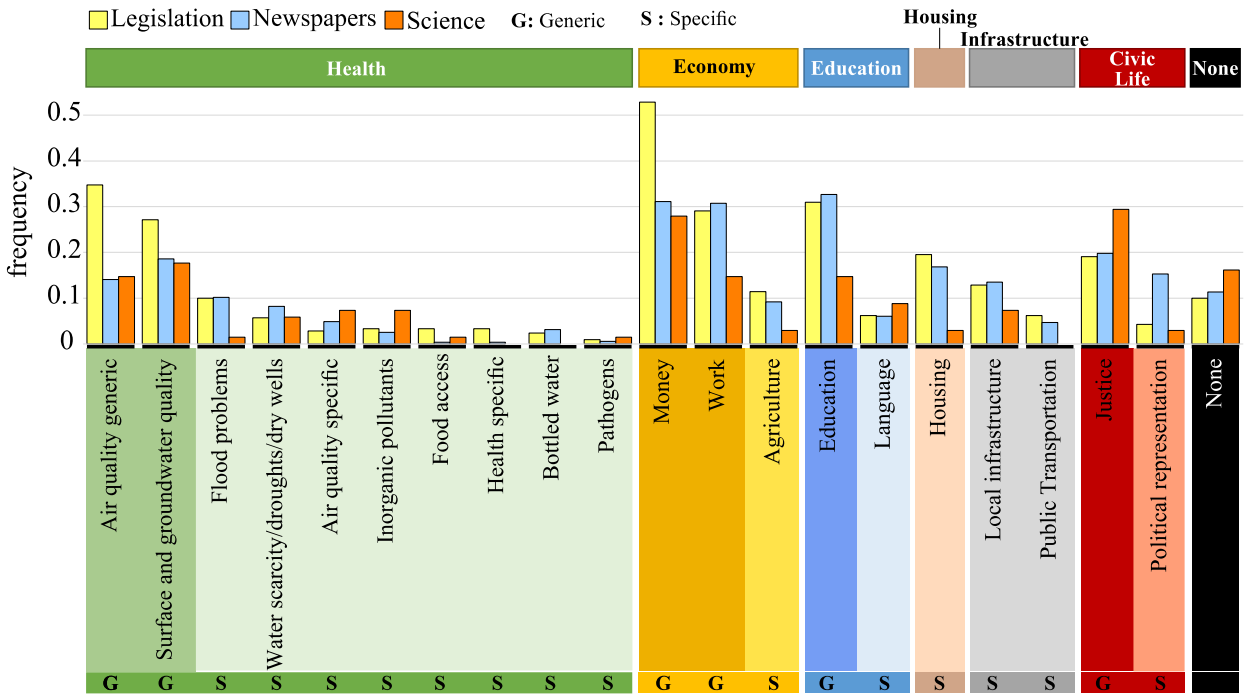

Fig. 2. Frequency of high-resolution categories by platform, classified as generic categories (G) and specific (S), and compared to the low-resolution categories (top ribbon). Generic categories are more frequent than specific categories.

We performed a cluster analysis on frequency of high-resolution categories to look for similarities in the way that the three platforms represent the communities (Fig. S2). Then we compared it with the frequency of issues raised during the interviews. The optimal number of clusters was two, and they coincided with the possible classification of "generic" and "specific" topics (Table 1). Generic problems are broad, such as water or air quality problems, fewer economic opportunities, or education, and had the highest frequency of representation across all three platforms. Specific problems are particular burdens disproportionally experienced by disadvantaged communities, such as problems with specific drinking water contaminants (arsenic, nitrates, 1,2,3,-trichloropropane), the burden of purchasing bottled water, or pesticide drift and dense dust near schools. Specific topics presented lower values in the cluster analysis, meaning that the representation across platforms was consistently lower, despite being important issues for the communities as demonstrated by the interview frequencies.

Table 1. Frequency of generic and specific issues, and corresponding low-resolution categories. The interviews frequency represents the number of interviewees that mentioned each issue.

	High-resolution category	Corresponding Low-resolution category	Frequency			
			Science	News	Bills	Interviews*
Cluster 1: Generic	Money	Economy	27.9%	31.1%	52.9%	83%
	Job opportunities	Economy	14.7%	30.7%	29.0%	67%
	Surface and groundwater quality	Health	17.6%	18.6%	27.1%	100%
	Education	Education	14.7%	32.7%	31.0%	50%
	Air quality generic	Health	14.7%	14.1%	34.8%	89%
	Justice	Civic Life	29.4%	19.8%	19.0%	39%
Cluster 2: Specific	Reliance on bottled water, interim water tanks, or filling and hauling water from neighbors	Health	0.0%	3.1%	2.4%	94%
	Inorganic pollutants (pesticides, heavy metals)	Health	7.4%	2.5%	3.3%	100%
	Pathogens (E. coli, Salmonella, Giardia)	Health	1.5%	0.6%	1.0%	56%
	Water scarcity/droughts/dry wells	Health	5.9%	8.2%	5.7%	61%
	Flooding problems	Infrastructure	1.5%	10.2%	10.0%	61%
	Air quality specific (pesticide drift and spraying, dust, smells, asthma)	Health	7.4%	4.9%	2.9%	83%
	Health specific (extreme heat, valley fever)	Health	0.0%	0.4%	3.3%	78%
	Food access	Health	1.5%	0.4%	3.3%	78%
	Agriculture	Economy	2.9%	9.2%	11.4%	100%
	Language isolation	Education	8.8%	6.1%	6.2%	50%
	Local infrastructure	Infrastructure	7.4%	13.5%	12.9%	44%
	Public Transportation	Infrastructure	0.0%	4.7%	6.2%	56%
	Housing	Housing	2.9%	16.8%	19.5%	39%
	Political representation	Civic Life	2.9%	15.3%	4.3%	78%

*Interviews were not subjected to a cluster analysis.

Comparing these high-resolution categories to the low-resolution ones, we find that *Health* appears more frequently and it is associated with two generic and seven specific high-resolution categories (Table 1). This is likely a cause of it being a prescriptive rather than preventive solution. While "water quality" was a keyword associated with *Health* in the low-resolution analysis, in the high-resolution analysis it became a keyword to one of the nine categories (*Surface and groundwater quality*) associated with *Health*. Thus, our results distinguish between "water problems" and specific issues that emerge from breaking down those broader categories. Consider the dependence on bottled water (a topic that concerns 94 % of the interviewees but with a frequency between 0 % and 3.1 % in the platforms): for some communities, it occurs because the available water is contaminated, while in others it is due to lack of water. Then, in some places, effective solutions may require filtration systems or groundwater blending (Mayzelle et al., 2015); while others may need to limit groundwater extraction near the communities to avoid that cones of depression from deeper wells take the groundwater from shallow wells serving communities (Pauloo et al., 2020); and in other locations, both approaches may be needed. Newspapers and legislation seem synchronized in the representation frequency of generic topics *Work* and *Education*, and specific categories *Local infrastructure*, *Housing*, *Justice*, *Agriculture*, and *Language*. Scientific papers and newspaper articles are in sync for generic topics *Air quality*, *Water quality*, and *Money*, but not for specific categories. In general,

there is a gap between the main concerns revealed by the interviews and the level of attention that those concerns get in the three platforms.

3.3. Developing effective policies requires addressing nuances and issue co-occurrence

Disadvantaged communities undergo disparities that must be addressed with specificity and not broadly. Specific needs tend to disappear in broad categories and large-scale classifications that do not capture the nuances of their lived reality. This generalization of topics may be one of the reasons why problems in disadvantaged communities are seldom addressed. A generic topic such as "air quality problems" may not call the attention of the public, since many locations in California have air quality problems related to traffic and wildfires; however, the air quality problems in disadvantaged communities are much more specific, such as pesticide drift entering homes through the windows, particularly at night during the summer in homes without air conditioning, or residents whose noses bleed when their communities are sprayed. Then, a policy that for example regulates emission standards to improve air quality will not address these types of community-specific concerns where vehicle emissions are not the culprit. Capturing specific problems helps to develop effective solutions, which for pesticide drift may be regulation to prohibit aircraft application of pesticides within a wider buffer from the community, or the creation of vegetation barriers to prevent particles from the surrounding farmlands to enter the communities. While there is value in broad classifications to identify problems at a large scale, they are too generic to represent accurately and to address co-occurring problems related to disadvantaged communities. Issue co-occurrence and its compounding effect may lead to negative effects if only one problem is tried to be addressed without a holistic understanding of the community. For example, the air quality in Kern is the worst in the United States, in part because of fracking activities by oil companies. Stopping the pollution source can solve air quality problems and prevent the exacerbation of climate change, but doing so without planning for socioeconomic impacts of the job loss can create new issues that will continue affecting disadvantaged communities in new ways. In this sense, disadvantaged community priorities can be wrongly perceived as paradoxical when the solution is addressing co-occurrent issues rather than one issue at a time.

Co-occurring issues can also lead to perception bias by external observers. For example, drinking water in the Central Valley city of San Joaquin (west of Fresno) often has high concentrations of sediments and pollutants (Fig. 3). The person who provided these pictures said that not everyone in the city had money to purchase bottled water, and some tried to boil it to remove its toxicity. However, this resident reported that air quality was their greatest concern because "most people can purchase bottled water, but none can buy clean air." The interviewee showed us how their car was covered with microdroplets from pesticide drift that arrived virtually everywhere in the city. That person became infected with coronavirus, a respiratory-related disease that disproportionally affects locations with poor air quality (Wu et al., 2020), while helping the most vulnerable in their community and passed away a few weeks before the present study was concluded.

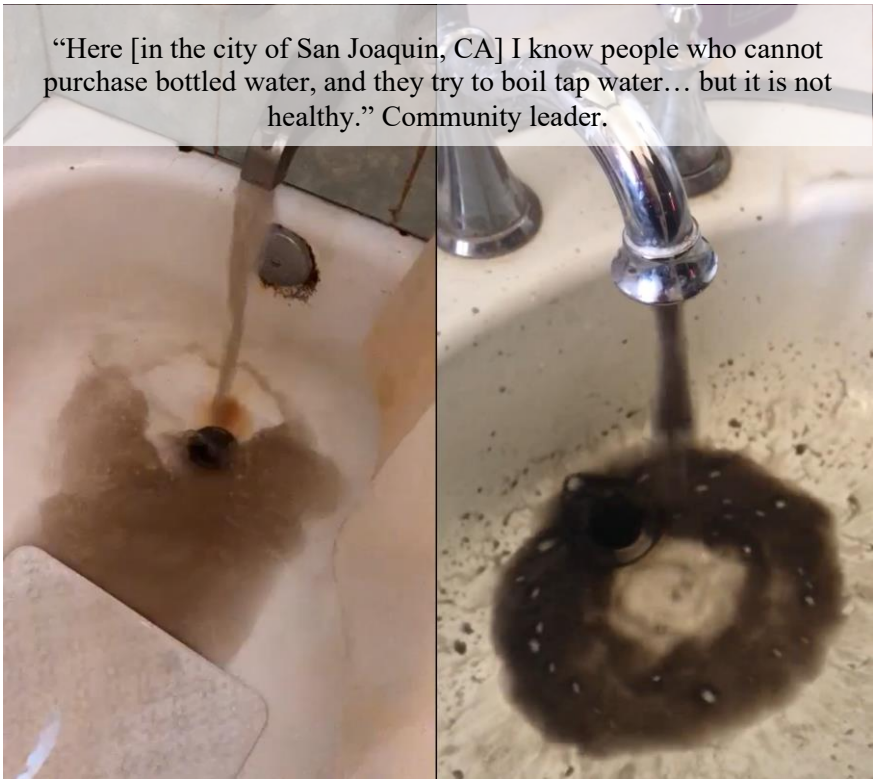

Fig. 3. Water in two homes in the city of San Joaquín, Fresno County, CA, in November 2019. Water contains high concentration of sediments and pollutants. Original quote in Spanish: "*hay personas que yo conozco que no [pueden comprar agua embotellada]. Ellos tratan de hervir [el agua de la llave] o lo que sea, pero de todos modos pues no es algo saludable.*"

Geographical scale can also be an issue. CalEnviroScreen uses census tracts to identify disadvantaged communities. While aggregating the population this way may work well in larger cities such as Los Angeles, San Jose, and Fresno, it is often too large for small rural unincorporated communities that are quite smaller than a census tract (more details in the Supplementary text). Policymakers can benefit from adding to the current pool of disadvantaged communities identified at the census tract those smaller ones for which the overall score is not "bad enough", but they have some key indicators performing very poorly. This way, the classification would be more inclusive and would avoid some extreme inequities. In addition, instead of using census tracts only, CalEnviroScreen could consider using higher resolution or additional definitions of locations to more appropriately represent small rural communities.

3.4. Community-specific knowledge is necessary to advance sustainable, effective solutions

First-hand knowledge and community perspective are critical for shaping solutions. Residents have consistently organized to demand adequate funding to address environmental injustice issues. Instances where their expertise and ideas have been embraced are among the most promising recent examples of progress. One significant success was the bill passed by the California Senate in 2019, SB-200, to devote up to $130 million annually until 2030 "to help

water systems provide an adequate and affordable supply of safe drinking water in both the near and long terms." The funding comes from a percentage (5 %) of cap-and-trade auctions for greenhouse gas emitters in the State. The California cap-and-trade program itself is aimed at improving air quality issues, but it has struggled at increasing environmental equity in disadvantaged communities (Cushing et al., 2018). The funding source is normally guaranteed, but the global pandemic dramatically decreased the auction profits (from $739 million in the last quarter of 2019 to $24 million in the May 2020 quarterly auction; data available on ww2.arb.ca.gov). However, starting in 2023, if the funding is less than $130 million, the amount will be supplemented by the General State Fund, making it a more robust funding source. Some interviewees mentioned how that the amount is less than what is needed; still, SB-200 is an important victory in the fight for environmental justice in disadvantaged communities.

Science and media have the opportunity to build on this momentum. Those impacted by social and environmental injustice have specific knowledge that is critical to the effective development of solutions rather than just addressing the symptoms (Cammarota and Fine, 2010; Morello-Frosch et al., 2005). However, valuing expert community knowledge above more traditional, hierarchical approaches to science is hardly the usual. Our results suggest that researchers assume what is better for the communities, and this ultimately renders unsatisfactory results for the communities. Scientific research benefits from bottom-up approaches to leverage local knowledge, including visits to the communities or interviews with individuals familiar with them. This allows scientists to understand the challenges firsthand, develop solutions in collaboration with local stakeholders, and increase the connection with the communities, which leads to a higher level of commitment on both sides. Community-based participatory research can help to understand the link between environmental justice and socioeconomic development in disadvantaged communities (Minkler and Wallerstein, 2011), and it is a tool to improve the rigor, relevance, and reach (the "3 Rs") of scientific studies (Balazs and Morello-Frosch, 2013). For example, decreasing carbon emissions from economic activity (for example, fossil fuel extraction and fracking) without harming the livelihood of workers from vulnerable communities who depend on it is complex and controversial; however, by addressing their needs (for example, through interviews with local experts and stakeholders), it is possible to bring social justice as well as environmental justice (Cha et al., 2020). Project evaluation and continuity require stable funding from agencies that, in turn, should hold accountable researchers for the benefits and positive impacts that their work claims to be doing for disadvantaged communities that claim they are developing research for disadvantaged communities but with little or no direct engagement with them, or resulting in no net benefit or a decimated positive impact to the communities.

News media inform the public opinion about their perception of disadvantaged communities, and this way they influence science and policy. While newspaper publications mention disadvantaged communities more frequently than science, their coverage tends to be too broad, missing the co-occurring inequities and the urgency of solutions. For example, the Sustainable Groundwater Management Act of California was passed to prevent future undesirable impacts associated with groundwater overdraft including household water outages like those that occurred in the Central Valley between 2012 and 2016. The implementation of this law requires decision-making by local stakeholders. Newspapers, however, tend to overrepresent the more powerful stakeholders while only describing disadvantaged communities a handful of times despite their legal standing in the law (Bernacchi et al., 2020). In this way, news media representation disengages disadvantaged communities from water resources management,

decreasing the law's capacity to serve the most vulnerable stakeholders that this legislation was meant to protect (Dobbin, 2020; MacLeod and Méndez-Barrientos, 2019; Méndez-Barrientos et al., 2020).

Disconnection between how society perceives disadvantaged communities and the actual roots of their problems often masks consequences of systemic inequities as deficiencies. Consider food access and education in the Central Valley. From a health perspective, physicians may encourage community members to eat healthier food, such as fruits and vegetables. However, this is sometimes difficult for residents who cannot afford the costs of healthy food or lack supermarkets and stores that sell quality food (often the closest option is an expensive convenience store in a gas station). Similarly, some of the worst-performing school districts in the country are in the Central Valley, but education may not improve by bringing the best teachers or by building new schools with state-of-the-art teaching technology. The root of educational problems is often everything but the delivery of education: children who are hungry or do not feel safe may have concentration difficulties; they may not have air conditioning when the Central Valley reaches 40 °C in the summer and fall afternoons; they may spend the whole day with their socks wet when they step in puddles in the winter while they walk for kilometers to their schools because of insufficient transportation, sidewalks, or drainage; they may be hungry because their parents have low salaries and they may have to prioritize paying rent and bottled water; they may be thirsty at schools where there is no available clean drinking water. Then, poor educational performance is not the problem, but a consequence of the co-occurrence of a plethora of systemic inequities.

Connecting community needs with public awareness and legislation to advance more tailored legislation requires mutual empowerment among policymakers, stakeholders, and the public. A recent example is a collaboration between Kamala Harris, US Senator and vice-presidential candidate, and Dolores Huerta, an iconic civil rights activist who founded the United Farm Workers with Cesar Chavez to defend farmworkers' rights. Together, they wrote an opinion letter about disparities that Black, Indigenous, and Communities of Color experience, focusing on safe and affordable drinking water (Harris and Huerta, 2020). The letter promotes the adoption of the proposed Water Justice Act (Harris, 2019), which seeks to enact $230 billion for water affordability programs and investments in clean and safe drinking water initiatives in the United States. This portrays the beginning of a path that can lead to success. However, it may not be enough if it is not pursued along with local stakeholders and integrated with other management actions to addressing systemic oppression of disadvantaged communities. Legislative fixes must focus on preventing and solving root problems rather than just focusing on the consequences of the problems. For example, it may mean first stopping pollution and then cleaning the water rather than investing only in cleaning the water and allowing pollution to continue.

Conclusions

Disadvantaged communities are underrepresented in news media, underserved by their government representatives, and understudied by science. Considering that millions of people live in disadvantaged communities, approaches to improve their living conditions need to fundamentally engage all three platforms and the communities. To untangle the systemic injustices that disadvantaged communities experience, we need to understand how multiple oppressions are intertwined and target solutions at multiple problem roots. While significant

recent efforts have made important decisive steps towards these ends, continuing to move in that direction will require the ongoing integration of local knowledge and perspectives. To succeed in such an endeavor, news media have the opportunity to increase the representation of disadvantaged communities, amplifying their voices to bring up their concerns and recommendations. If society becomes more aware thanks to proper media representation, they will be able to incentivize policymakers to create the institutional infrastructure to implement solutions. Legislators can adequately serve disadvantaged communities by partnering with them to craft sustainable solutions and allocating sufficient resources that include funding for community-based research, grassroots organizations, children and adult education, and technical assistance. And scientists must approach disadvantaged communities using more community-based research and fewer assumptions, with holistic and multidisciplinary frameworks in partnership with those most impacted, and they must share findings widely with the general public and policymakers.

References and Notes:

Adams, W., 2015. Conducting Semi-Structured Interviews, in: Handbook of Practical Program Evaluation. Wiley. https://doi.org/10.1002/9781119171386.ch19

Allaire, M., Wu, H., Lall, U., 2018. National trends in drinking water quality violations. Proc. Natl. Acad. Sci. 115, 2078–2083. https://doi.org/10.1073/pnas.1719805115

Almaguer, T., 1994. Racial Fault Lines: The Historical Origins of White Supremacy in California, Racial Fault Lines: The Historical Origins of White Supremacy in California. University of California Press, California.

Anderson, M.W., 2009. Mapped out of Local Democracy. Stanford Law Rev. 62, 931.

Balazs Carolina, Morello-Frosch Rachel, Hubbard Alan, Ray Isha, 2011. Social Disparities in Nitrate-Contaminated Drinking Water in California's San Joaquin Valley. Environ. Health Perspect. 119, 1272–1278. https://doi.org/10.1289/ehp.1002878

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

Balazs, C.L., Morello-Frosch, R., 2013. The Three Rs: How Community-Based Participatory Research Strengthens the Rigor, Relevance, and Reach of Science. Environ. Justice 6, 9–16. https://doi.org/10.1089/env.2012.0017

Balazs, C.L., Ray, I., 2014. The drinking water disparities framework: on the origins and persistence of inequities in exposure. Am. J. Public Health 104, 603–611. https://doi.org/10.2105/AJPH.2013.301664

Benner, C., Erbstein, N., London, J., Greenfield, T., Hartzog, C., Watterson, S., Godard, D., Lampinen, G., Visser, A., 2014. Regional Opportunity Index (ROI) [WWW Document]. UC Davis Cent. Reg. Change. URL https://interact.regionalchange.ucdavis.edu/roi (accessed 7.17.20).

Bernacchi, L.A., Fernandez-Bou, A.S., Viers, J.H., Valero-Fandino, J., Medellín-Azuara, J., 2020. A glass half empty: Limited voices, limited groundwater security for California. Sci. Total Environ. 738, 139529. https://doi.org/10.1016/j.scitotenv.2020.139529

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

Cammarota, J., Fine, M., 2010. Revolutionizing Education: Youth Participatory Action Research in Motion. Routledge.

Carvalho, A., 2010. Media(ted)discourses and climate change: a focus on political subjectivity and (dis)engagement. WIREs Clim. Change 1, 172–179. https://doi.org/10.1002/wcc.13

Cha, J.M., Wander, M., Pastor, M., 2020. Environmental Justice, Just Transition, and a Low-Carbon Future for California. Environ. Law Report. 50, 10216.

Cushing, L., Blaustein-Rejto, D., Wander, M., Pastor, M., Sadd, J., Zhu, A., Morello-Frosch, R., 2018. Carbon trading, co-pollutants, and environmental equity: Evidence from

California's cap-and-trade program (2011–2015). PLOS Med. 15, e1002604. https://doi.org/10.1371/journal.pmed.1002604

De León, K., 2012. SB-535, Health and Safety Code, relating to climate change.

De Vore, M.B., 2008. Attitudes and perceptions of high school career and technical education in California's Central Valley (Dissertation). Pepperdine University, Malibu, CA.

Dobbin, K.B., 2020. "Good Luck Fixing the Problem": Small Low-Income Community Participation in Collaborative Groundwater Governance and Implications for Drinking Water Source Protection. Soc. Nat. Resour. 1–18. https://doi.org/10.1080/08941920.2020.1772925

Dobbin, K.B., Lubell, M., 2019. Collaborative Governance and Environmental Justice: Disadvantaged Community Representation in California Sustainable Groundwater Management. Policy Stud. J. https://doi.org/10.1111/psj.12375

Eissinger, M.A., 2017. Re-Collecting the Past: An Examination of Rural Historically African American Settlements across the San Joaquin Valley. CreateSpace Independent Publishing Platform.

Eissinger, M.A., 2008. African Americans in the Rural San Joaquin Valley, California: Colonization Efforts and Townships. California State University, Fresno.

Ezra David, R., Klain, K., 2017. They Built It, But Couldn't Afford To Run It—Clean Drinking Water Fight Focuses On Gaps In Funding [WWW Document]. Val. Public Radio. URL https://www.kvpr.org/post/they-built-it-couldn-t-afford-run-it-clean-drinking-water-fight-focuses-gaps-funding (accessed 7.17.20).

Flegel, C., Rice, S., Mann, J., Tran, J., 2013. California unincorporated: mapping disadvantaged communities in the San Joaquin valley. Policylink Oakland, California, United States.

Gifford, B.R., Valdés, G., 2006. Chapter 7 The Linguistic Isolation of Hispanic Students in California's Public Schools: The Challenge of Reintegration. Yearb. Natl. Soc. Study Educ. 105, 125–154. https://doi.org/10.1111/j.1744-7984.2006.00079.x

Goddard, J.J., Ray, I., Balazs, C., 2021. Water affordability and human right to water implications in California. PLOS ONE 16, e0245237. https://doi.org/10.1371/journal.pone.0245237

Gomez, 2016. AB-1550, Health and Safety Code.

Hanak, E., Escriva-Bou, A., Gray, B., Green, S., Harter, T., Jezdimirovic, J., Lund, J., Medellín-Azuara, J., Moyle, P., Seavy, N., 2019. Water and the Future of the San Joaquin Valley: Overview.

Harris, K.D., 2019. Text - S.2466 - 116th Congress (2019-2020): Water Justice Act [WWW Document]. URL https://www.congress.gov/bill/116th-congress/senate-bill/2466/text (accessed 7.28.20).

Harris, K.D., Huerta, D., 2020. Opinion: Racism is fueling disparities in access to safe water [WWW Document]. Mercury News. URL https://www.mercurynews.com/2020/07/16/7301162/ (accessed 7.28.20).

Killingsworth, M.J., Palmer, J.S., 2012. Ecospeak: Rhetoric and Environmental Politics in America. Southern Illinois University Press.

Kissam, E., 2020. The impact of the COVID-19 pandemic on California farmworkers: Better local data collection and reporting will improve strategic response. Stat. J. IAOS 36, 867–898. https://doi.org/10.3233/SJI-200763

Latino Community Foundation, 2018. A roadmap for census 2020: securing a fair and accurate count of Latinos in California.

Likens, G.E., 2010. The role of science in decision making: does evidence-based science drive environmental policy? Front. Ecol. Environ. 8, e1–e9. https://doi.org/10.1890/090132

MacLeod, C., Méndez-Barrientos, L.E., 2019. Groundwater Management in California's Central Valley: A Focus on Disadvantaged Communities. Case Stud. Environ. 3, 1–13. https://doi.org/10.1525/cse.2018.001883

Mayzelle, M.M., Viers, J.H., Medellín-Azuara, J., Harter, T., 2015. Economic Feasibility of Irrigated Agricultural Land Use Buffers to Reduce Groundwater Nitrate in Rural Drinking Water Sources. Water 7, 12–37. https://doi.org/10.3390/w7010012

Méndez-Barrientos, L.E., DeVincentis, A., Rudnick, J., Dahlquist-Willard, R., Lowry, B., Gould, K., 2020. Farmer Participation and Institutional Capture in Common-Pool Resource Governance Reforms. The Case of Groundwater Management in California. Soc. Nat. Resour. 0, 1–22. https://doi.org/10.1080/08941920.2020.1756548

Minkler, M., Wallerstein, N., 2011. Community-Based Participatory Research for Health: From Process to Outcomes, Wiley Desktop Editions. Wiley.

Morello-Frosch, R., Pastor, M., Sadd, J., Porras, C., Prichard, M., 2005. Citizens, Science, and Data Judo. Methods Community-Based Particip. Res. Health 22.

OEHHA, 2018. Analysis of race/ethnicity, age, and CalEnviroScreen 3.0 Scores. Office of Environmental Health Hazard Assessment, Sacramento, CA.

OEHHA, 2017. CalEnviroScreen 3.0: Update to the California Communities Environmental Health and Screening Tool. Office of Environmental Health Hazard Assessment.

Pannu, C., 2012. Drinking Water and Exclusion: A Case Study from California's Central Valley. Calif. Law Rev. 100, 223–268.

Pauloo, R.A., Escriva-Bou, A., Dahlke, H., Fencl, A., Guillon, H., Fogg, G.E., 2020. Domestic well vulnerability to drought duration and unsustainable groundwater management in California's Central Valley. Environ. Res. Lett. 15, 044010. https://doi.org/10.1088/1748-9326/ab6f10

Real, J.A.D., 2019. 'Brown Water for Brown People': Making Sense of California's Drinking Water Crisis. N. Y. Times.

Scott, T.A., Ulibarri, N., Figueroa, O.P., 2020. NEPA and National Trends in Federal Infrastructure Siting in the United States. Rev. Policy Res. 37, 605–633. https://doi.org/10.1111/ropr.12399

Shanahan, E.A., McBeth, M.K., Hathaway, P.L., Arnell, R.J., 2008. Conduit or contributor? The role of media in policy change theory. Policy Sci. 41, 115–138. https://doi.org/10.1007/s11077-008-9058-y

Shonkoff, S.B., Morello-Frosch, R., Pastor, M., Sadd, J., 2011. The climate gap: environmental health and equity implications of climate change and mitigation policies in California—a review of the literature. Clim. Change 109, 485–503. https://doi.org/10.1007/s10584-011-0310-7

Tejada-Vera, B., Arias, E., Escobedo, L., Salant, B., 2020. Life Expectancy Estimates by U.S. Census Tract, 2010-2015 [WWW Document]. Natl. Cent. Health Stat. URL https://www.cdc.gov/nchs/data-visualization/life-expectancy/

Ulibarri, N., Cain, B.E., Ajami, N.K., 2017. A Framework for Building Efficient Environmental Permitting Processes. Sustainability 9, 180. https://doi.org/10.3390/su9020180

U.S. Census Bureau, 2019. U.S. Census Bureau QuickFacts: California; United States [WWW Document]. URL https://www.census.gov/quickfacts/fact/map/CA,US/PST045219 (accessed 3.2.21).

Wu, X., Nethery, R.C., Sabath, B.M., Braun, D., Dominici, F., 2020. Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. MedRxiv Prepr. Serv. Health Sci. 2020.04.05.20054502. https://doi.org/10.1101/2020.04.05.20054502

Acknowledgments: We are grateful to all the respondents who donated their time and provided extremely useful insights of the communities they represent. Thanks to those who helped in this project, including Miriam Campos Martinez, Robin DeLugan, Alexander Guzman, public servants, and personnel from nonprofits (including Water Education for Latino Leaders, Leadership Counsel for Justice & Accountability, Self-Help Enterprises, Community Water Center, and 4 Venir). Our semi-structured interview protocol was approved by the Institutional Review Board at the University of California, Merced with ID UCM2019118. Funding: Ortiz-Partida was partially supported by funding from the Water Foundation and Environment Now; Fernandez-Bou, Flores-Landeros, Bernacchi, Medellín-Azuara were partially funded through several grants under co-PI Medellin-Azuara including NSF INFEWS program grant number 1639268 (P.I. Characklis, UNC-Chapel Hill), USDA INFEWS grant number 2018-67004-27405 (P.I. Conklin, UC Merced), California Strategic Growth Council number CCRP0013 (P.I. McCullough, CalPoly San Luis Obispo), University of California Merced School of Engineering (P.I. Medellin-Azuara); Dobbin was supported by NSF Graduate Research Fellowship number 1650042.

Author contributions: A.S.F.B. and J.P.O.P. contributed with Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft; H.F.L. contributed with Data curation, Investigation, Writing – original draft; K.B.D. contributed with Conceptualization, Writing – review & editing. L.B. contributed with Conceptualization, Methodology, Writing – review & editing; J.M.A. contributed with Funding acquisition, Project administration, Writing – revision & editing.

Supplementary Materials

Underrepresented, underserved, understudied: gaps and opportunities for advancing justice in disadvantaged communities

A.S. Fernandez-Bou*, J.P. Ortiz-Partida*, K.B. Dobbin, H. Flores-Landeros, L.A. Bernacchi, J. Medellín-Azuara

Correspondence to: afernandezbou@ucmerced.edu (A.S.F.B.); jportiz@ucsusa.org (J.P.O.P.)

* These authors contributed equally to this work

Index

Supplementary Text

Representation results

Looking more closely at these representations, few scientific publications interviewed community residents or trusted representatives such as technical assistance providers, local politicians or nonprofit staff. Fifteen articles (in Elsevier and Springer) included some kind of interview, but often related to a very specific, predetermined concept, such as drug-related syringe use or the popularity of bicycles; and in all cases disadvantaged communities were only mentioned as part of the scope, and not as their focus. Nine scientific articles in the 154,000 California-related corpus concerned disadvantaged communities in the Central Valley, but zero interviewed residents about their socioeconomic and environmental concerns. Thus, besides the limited coverage of disadvantaged communities, first-hand information from residents is not surfaced. A limitation of the study is that the corpus of scientific articles only considered publications on the databases from Elsevier and Springer. There are however other publishers that include articles that did interviews about socioeconomic and environmental concerns in disadvantaged communities.

Low-resolution analysis results

Publications represented more frequently the categories *Health* and *Economy*, and less frequently *Housing* and *Civic Life*. Here, *Health* is broadly defined and includes water and air issues, among others, and *Economy* is sometimes the only factor to classify a community as disadvantaged; then, such high frequency of those categories was expected. Legislation presented the highest frequency and co-occurrence of low-resolution categories, indicating that policymakers have broad understanding of the issues in disadvantaged communities and bills may have a correct large-scale context (Fig. S1A). Newspapers presented lower frequency and co-occurrence of low-resolution categories than did legislation, but still in relatively high numbers, suggesting that they tend to be concentrated on narrow topics yet usually providing context to understand the situation (Fig. S1B). Science had the lowest frequency and the least holistic perspective (also with more publications on disadvantaged communities unrelated to any category under study), suggesting that scientific research may be more focused on individual issues predefined by researchers rather than based on community-based input or research (Fig. S1C).

It is worth noting that the Regional Opportunity Index also contains many subcategories that provide important details of the communities they represent. However, sometimes the data utilized by this index is broader and less specific than our interview-based approach, and it is limited by available datasets, such as the census. For example, in the Regional Opportunity Index, the *Health* category contains *Infant health* and *Births to teens*, which are relevant categories to analyze large-scale problems but do not illustrate the root of such problems, while our analysis focuses on specific health issues and their sources. Other subcategories of the index, however, are more detailed, such as those related to employment (*Job availability*, *Job growth*, *Job quality*), but they are less specific to disadvantaged communities, since the Regional Opportunity Index was developed for the whole State of California. We do not intend to provide a substitute of the Regional Opportunity Index, as its capacity at identifying large-scale issues is extremely valuable.

High-resolution analysis results. Cluster analysis

The cluster analysis revealed an optimal distribution of the high-resolution categories in two clusters, one of them with more representation than the other. The cluster with more representation (that we identify as "generic") had general categories such as *Money*, *Education*, *Work*, *Justice*, *Surface and groundwater quality*, and *Air quality*. The less represented cluster (that we call "specific") contained more specific issues related to most or all of the communities, and its categories were *Housing*, *Political representation*, *Local infrastructure*, *Agriculture*, *Language*, *Flood problems*, *Water scarcity/droughts/dry wells*, *Political representation*, *Air quality specific*, *Inorganic pollutants*, *Public Transportation*, *Food access*, *Health specific*, *Bottled water*, and *Pathogens* (Table S5).

The relevance of geographical scale

CalEnviroScreen is a pioneer and essential tool to identify disadvantaged communities and to guide policymakers to serve them adequately. However, the methodology has at least two issues that may result in inequities for some communities. One is the use of census tracts. While aggregating the population this way may work well in larger cities such as Los Angeles, San Jose, and Fresno, it is often too large for small rural unincorporated communities. For example, Tooleville is a small community (less than 500 habitants) near the Sierra Nevada foothills in Tulare county. The community has many issues, and we visited there advised by nonprofit organizations who work with them providing legal counsel to bring back their rights to water security and clean air. Tooleville is clearly a vulnerable community that suffers disproportionally from most of the indicators measured by CalEnviroScreen. However, it is not considered a disadvantaged community, since it is located inside a larger census tract with more than 5,000 habitants (census tract 6107001400), and the better performance of the rest of the locations compensates their poor living conditions.

Another issue with CalEnviroScreen is the averaging methodology of the indicators. While some of the components are weighted, utilizing the mean of the indicators can lead to good performances compensate for bad performances. For example, the west census tract of the city of Arvin (census tract 6029006303; 6,784 habitants) performs in the percentile 70 to 75 (CalEnviroScreen 3), hence out of the "disadvantaged community" classification. However, that census tract has some of the worst (highest) percentiles in the state for Ozone (98), PM 2.5 (97), Pesticides (98), Drinking Water (87), Education (100), Linguistic Isolation (92), Poverty (99), and Unemployment (95). The reason why that census tract is not considered disadvantaged is because they perform really well in Toxic releases (12), Traffic (8), Cleanups (20), Groundwater threats (9), Hazardous waste (26), Impaired Water (0), Solid Waste (0), and moderately well in Asthma (40), Low Birth Weight (48), Cardiovascular Rate (67), and Housing Burden (50).

Policymakers can benefit from adding to the pool of current disadvantaged communities identified by CalEnviroScreen those other communities for which the overall score is not "bad enough", but they have some key indicators performing very poorly. This way, the classification would be more inclusive and would avoid some extreme inequities. In addition, instead of using census tracts only, CalEnviroScreen could consider using "places" as well, as they are more appropriate for small rural communities.

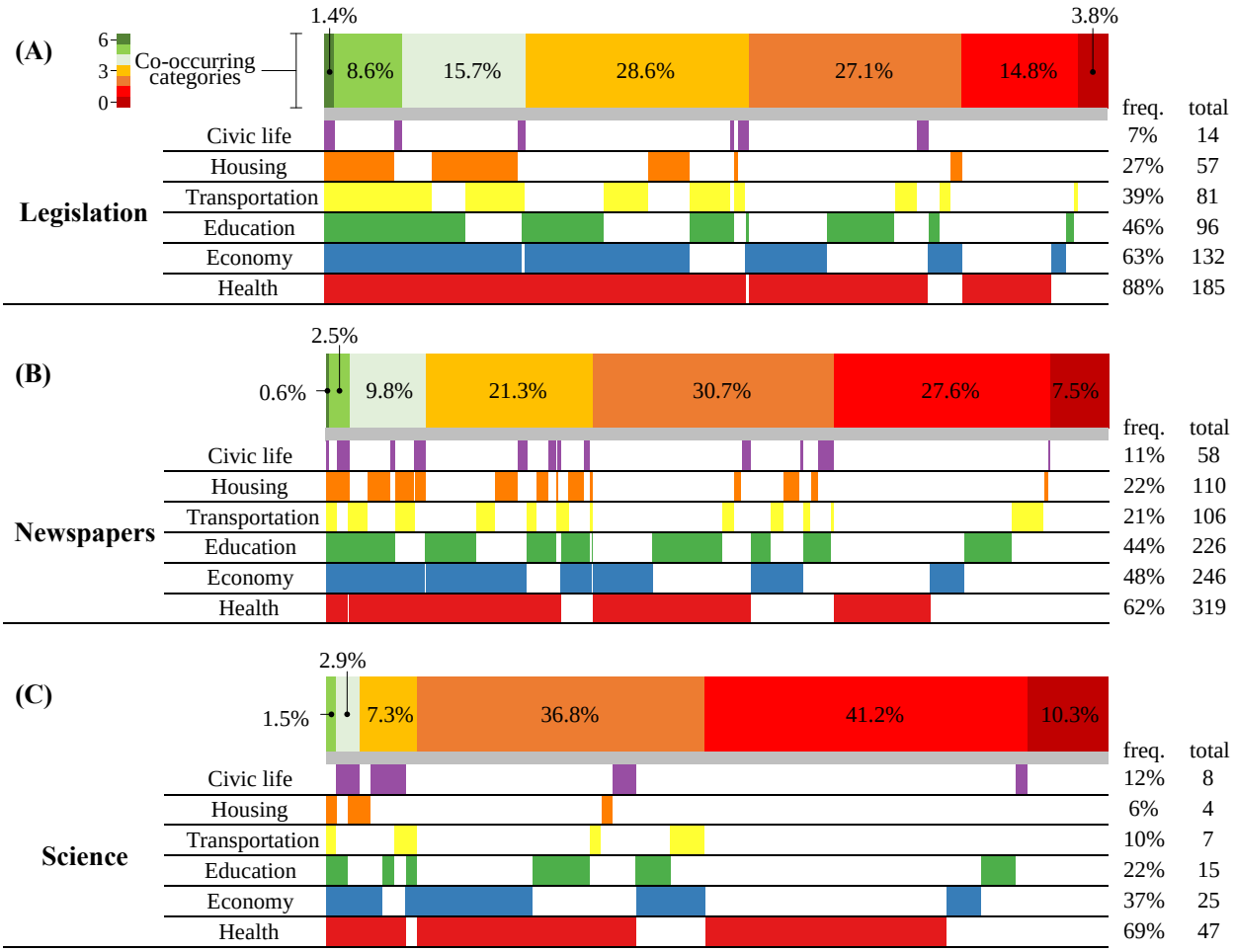

Fig. S1.

Frequency and co-occurrence of low-resolution categories. Upper ribbon represents the number of co-occurring categories. The bars representing the six studied categories are organized so that the co-occurring categories are sorted. The total number of columns is equal to the total publications analyzed per platform, and each vertical column represent a publication.

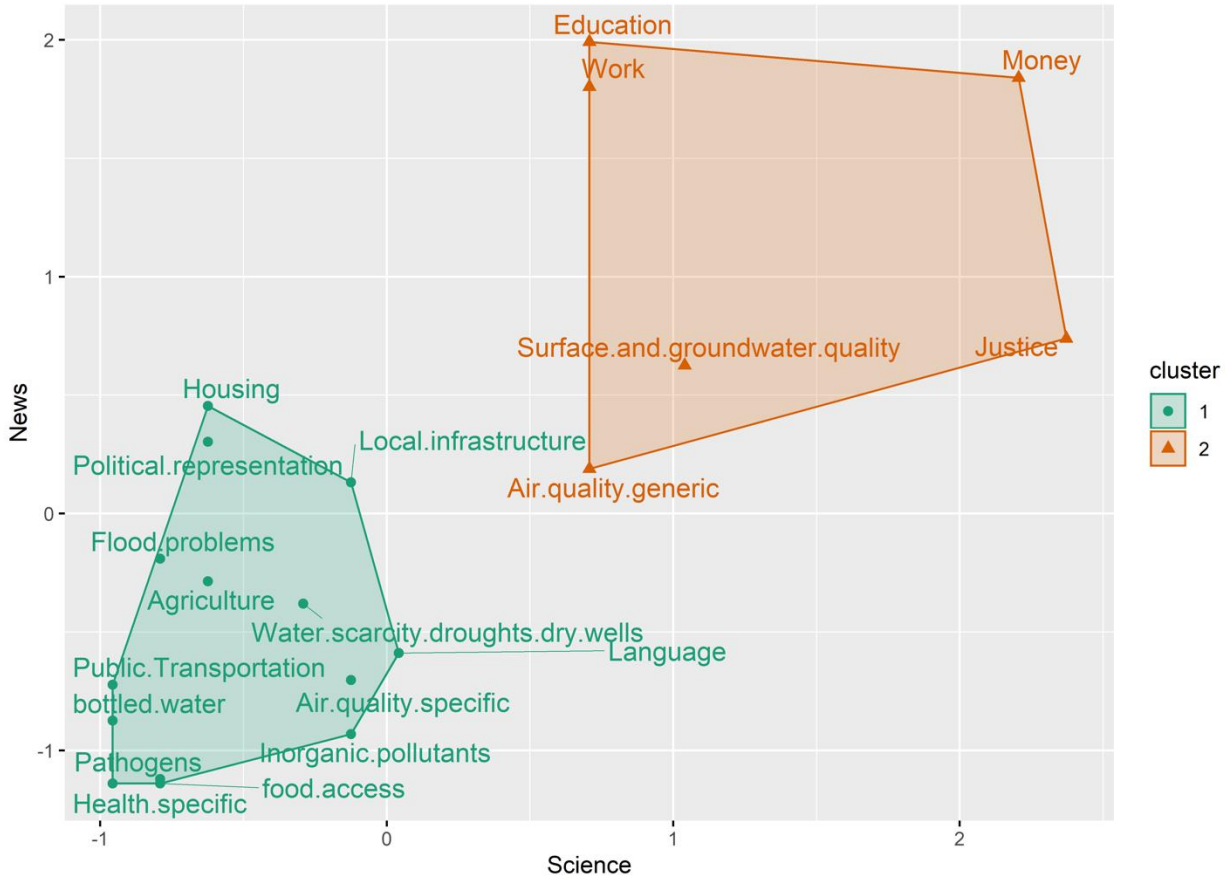

Fig. S2.

Clusters. Data points closer to the bottom, left occurred less often than those on the top, right corner. Lower values mean less occurrence in that dimension.

Table S1.

Study design

	Science	Newspapers	Legislation
Databases	**Elsevier** (sciencedirect.com) and **Springer** (link.springer.com); articles (research, reviews, and short communications) in English language	Newspaper news available online from 240 newspapers in the state of California, USA.	California Legislative Information (leginfo.legislature.ca.gov)
Time frame	between January 2017 and June 2020		
Sample search equation	All scientific articles containing the word "California"	All news available in the database with 240 Californian newspapers	All the bills available (not a sample of them, but all of them)
Sample size	**154,000 articles**	**2,612,324 articles**	**10,626 bills**
Search equation	"disadvantaged communities" AND California	"disadvantaged communities"	"disadvantaged communities"
Preselection	**198**	**1,440**	**240**
Selection	We filtered the articles against a set of keywords to verify that the article was addressing issues related to disadvantaged communities, and we read approximately half of the articles to validate the keywords.	We removed obvious duplicates (same article in different journals), and we excluded those articles that clearly were not referring to the Central Valley. Before removing duplicated, the number of items found was 1,440.	We removed those bills with more than 100,000 words, and bills in which the term "disadvantaged communities" was only defined but not used in a context of action towards them
Items analyzed	**68**	**511**	**210**
Preparation for in-depth analysis utilizing:	All the titles, keywords, and abstracts of the selected articles.	All the article bodies and their titles.	a) If search term was in title of bill, law chapter, legislative counsel's digest, then full text analyzed. b) If search term was in the body only, then only the section(s) analyzed.

5

Table S2.

Theoretical categories based on the Regional Opportunity Index and CalEnviroScreen and keywords used to classify the publications

Category	Keywords
Health	asthma, air quality, cardiovascular, disease, low birth weight, health, water quality, food security, food access, pesticide, cancer, dust, smog, particulate matter, contamination
Economy	salary, wages, employ, job, income, job availability, job quality, job growth, bank access
Education	education, college, UC eligible, CSU eligible, university, school, Math, Instruction, teacher, English proficiency, English knowledge, linguistic.
Infrastructure	commute, vehicle available, transport, internet access, vehicle own
Housing	housing, cost burden, homeowner, household income
Civic Life	voting rate, citizenship, English, neighborhood stability, us citizen, economic dependent, farming

In addition, we searched for words derived from those (such as contaminated, contaminating, and contaminate, or ownership, owned, and owning), and synonyms (for example, for smog, we searched for ozone and O_3).

Table S3.

Interviews questions.

#	Question
1	What is your relationship with this community?
2	Do you identify any “Environmental risk” for the livelihood of your community?
3	[If not addressed in the previous question] Does your community have access to clean water? [Meaning if they can drink, bathe, or cook with water from the tap]
4	[If not addressed previously] What are the major challenges that the drought brought to your community that you did not expect?
5	What are your thoughts about climate change and the potential effects on your community?
6	Can you speak about food access and security in your community?
7	What limitations does your community have compared with other places that you know?
8	How is the employment in your community?
9	Do you think the community is represented at the different levels of government?
10	What are the measures already taken to improve living standards in your community?
11	What else can be done to improve the living conditions in your community?
12	Is there anything you want to add?

5

Table S4.

Nodes in interview analysis

#	Topics
1	Agriculture: considers mentions of farms, farmers, farmworkers, agricultural activities, crop types, industry employment opportunities, wages.
2	Air quality: considers mentions of sources of air pollution, health issues related to air quality, smells.
3	Climate change: considers mentions of perceived changes in temperatures, rainfall events, drought, floods, and others.
4	Impacts: umbrella category for mention of impacts related to diseases, hydroclimatic events, smells, and lack of opportunities.
5	Non-environmental problems: considers mention of issues like corruption, affordable housing, job opportunities, education, access to food, insufficient or lack of infrastructure, language barriers.
6	Other environmental issues: considers mentions environmental issues not directly related to air and water quality, such as extreme heat, pests, soils pollution
7	Policy: mentions of elected officials, policies, regulations, representation (or lack of it), elections.
8	Social characteristics: considers mentions of the economic situation,, education, health, housing, and transportation
9	Water problems: considers mentions of water quality, water quantity, groundwater, droughts and flood impacts and experiences, water affordability and access.

5

Table S5.

Theoretical categories based on the qualitative and quantitative analysis of the interviews and keywords used to classify the publications

Category	Keywords
Money	ownership, income, salary, wages
Education	educat*, college, high school, middle school, elementary school
Work	employ*, job, job growth, job availab*, job qual*
Justice	political repre*, racism, justice, equity, oppress*
Air quality generic	air qual*, air contam*, air pollut*, particulate matter
Surface and groundwater quality	water qual*, water access, water issues, water problems, groundwater, water pollu*, water contamin*, safe water, affordable water, water affordability, clean water, clean drinking water, water bill
Housing	housing
Local infrastructure	sewage, green area, park, sidewalk, street light, ageing, infrastr*, bad street, pothole, streetlight, water treat*, lack of emergen*
Agriculture	farming, agriculture, farmwork, fieldwork
Language	speak English, English, linguistic, learn English, learning English, improve English, English skill, language
Flood problems	flood
Water scarcity/ droughts/dry wells	drought, dry*wells, wells*dry, water scarc*
Political representation	representation, vote, constituent, voting
Air quality specific	asthma, smell, dust, pesticide drift, application of pesticides, downwind, upwind, pesticide spray
Inorganic pollutants	1,2,3,-T, trichloro*, nitrate, arsenic, heavy met*
Public Transportation	public transpo*, car own, own car
food access	junk food, food desert, food acc*, quality food, healthy food, healthy meal, health*food*supermarket, supermarket*health*food, health*supermarket*food, food*supermarket*health
Health specific	valley fever, heat wave, heat stroke, extreme heat
Bottled water	bottled water, bottle of water, water bottle
Pathogens	giardia, septic tank, e. coli, salmonel*

In addition, we searched for words derived from those (such as educated, education, educator, and educational) and the use of "*" means that the search returned all the results that contain the strings on the left and right of the "*" regardless of the words that the "*" represent, such as "wells that went dry" or "healthy food options in the supermarket".