

# scientific data



**OPEN**  
DATA DESCRIPTOR

## MapAgora, civic opportunity datasets for the study of American local politics and public policy

Jae Yeon Kim , Milan de Vries & Hahrie Han

Civil society plays a vital role in cultivating civic skills and democratic engagement. Yet studying it at scale in the United States remains challenging due to the decentralized nature of the nonprofit organizations that constitute civil society. We introduce MapAgora, a set of five new datasets that provide a comprehensive view of civic opportunity across America. Drawing on IRS tax records and the websites of over 1.7 million nonprofit organizations, we examine whether organizations offer civic opportunities, such as membership, volunteering, public events, or civic and political action, and classify them by type (e.g., political, professional, issue-based). The de-identified organization-level dataset includes indicators of civic opportunity, organizational type, and financial status. Four additional datasets aggregate this information at the ZIP code and county levels. We validate that civic opportunity is conceptually distinct from financial resources at the organizational level, while aggregate patterns align with traditional measures of socioeconomic inequality. These datasets offer new tools for understanding civic infrastructure, local democracy, and public policy.

### Background & Summary

Democracy requires people to develop capabilities for participation, such as expressing preferences and negotiating with each other in collective settings<sup>1,2</sup>. A healthy democracy represents the public and enables people to act together in ways that can build social bonds, hold public officials accountable, and realize shared interests<sup>3–6</sup>. Achieving this vision requires not only a capable state, but also a vibrant civil sector that can shape, channel, and when necessary, check government authority<sup>7</sup>. Civil society also provides the institutional and relational infrastructure for civic engagement, including collective action, community-building, service delivery, and leadership development.

In practice, this infrastructure is built and maintained by organizations, both formal and informal, that enable people to connect, participate, and lead. We refer to this as *civic infrastructure*—the tangible networks and institutions that support everyday democratic life. Within this infrastructure, *civic opportunity* refers to the availability of concrete ways for individuals to engage in public life, such as volunteering, attending events, joining organizations, or participating in civic or political action. Civic opportunity is not guaranteed by the size, visibility, or financial resources of any individual organization. Larger or wealthier organizations, such as foundations, hospitals, or universities, may not offer civic opportunities. By contrast, smaller, less-resourced organizations, such as local religious groups or fraternal societies, often serve as the most active providers of civic opportunity<sup>8</sup>. What matters for democratic life is the widespread availability of inclusive opportunities for engagement. Such opportunities are necessary because the civic skills necessary for democratic citizenship are learned, not innate. People don't naturally reach across social divides. Human behavior tends to be self-interested or parochial, especially when people encounter differences in race, language, or belief<sup>9–11</sup>. Fostering pro-democratic civic engagement in a diverse society requires the intentional cultivation of civic skills and pro-social norms<sup>12</sup>.

Much like building physical strength requires repeated training, developing civic capacity requires practice, especially across lines of difference. Just as access to gyms enables physical fitness, civic opportunity supports civic development. Alexis de Tocqueville called such settings “schools of democracy”<sup>13</sup>. Civic opportunity thus captures the *supply side* of social capital, the institutional and organizational conditions that make civic participation possible. Unlike the *demand-side* measures of social capital, such as interpersonal trust, civic opportunity emphasizes the structural availability of opportunities to engage, organize, and act collectively.

SNF Agora Institute, Johns Hopkins University, Baltimore, 21218, USA. e-mail: [jkim638@jhu.edu](mailto:jkim638@jhu.edu)

Conducting a large-scale investigation of the distribution of civic opportunity is, however, challenging, even in the United States, where Tocquevillian practices are deeply rooted in its local, democratic tradition. Associations are highly decentralized and often emerge organically. No single dataset tracks them comprehensively. With over a million such organizations nationwide, manual data collection is not feasible. This gap is notable, especially compared to the wealth of data available on U.S. elections or public meetings<sup>14,15</sup>. If similar datasets existed for local civic infrastructure, they could help scholars and practitioners better understand how these civic infrastructure influence elections, shape policy, and build political power at the local level.

Due to these limitations, most researchers face trade-offs. One approach prioritizes breadth, using one or two measurable forms of civic engagement, such as volunteering, to study civic life nationally<sup>16–18</sup>. This offers scale, but risks overlooking meaningful forms of civic or political action beyond what is easily measurable<sup>19</sup>. The other approach prioritizes depth, focusing on specific types of organizations within a city or region. This yields rich insights but makes it hard to compare across places or organization types<sup>20–23</sup>.

We used a computational approach to combine both breadth and depth. Our goal was to capture as many organizations as possible while extracting detailed information about the opportunities they provide.

To construct our datasets on civic infrastructure, we began with the Internal Revenue Service (IRS) Exempt Organizations Business Master File, a comprehensive registry of all tax-exempt organizations in the United States. This registry serves as our sampling frame and includes a broad range of nonprofit entities, regardless of whether they file annual tax returns. From this list, we linked IRS administrative records with data scraped from organizational websites. We used machine learning and rule-based classification techniques to identify the type of each organization and whether it offers one or more forms of civic opportunity, including volunteering, membership, public events, and civic or political engagement. The resulting data were aggregated at the ZIP code and county levels to reflect the geographic distribution of civic infrastructure nationwide.

While the IRS no longer updates the Form 990 series on Amazon Web Services (AWS) as of December 31, 2021 (<https://registry.opendata.aws/irs990/>), we scraped and restructured these administrative records when they were publicly accessible. These filings, which date back to tax year 2011, provide standardized information on more than 1.7 million nonprofit organizations. As summarized in Table S1 of the Online Appendix, the dataset includes a wide range of civic organizations: public charities such as food banks and universities; private philanthropic foundations such as the Gates Foundation; religious institutions including churches, mosques, and synagogues; advocacy organizations such as the American Civil Liberties Union and the National Rifle Association; labor unions such as the American Federation of Labor and Congress of Industrial Organizations; and professional associations such as the U.S. Chamber of Commerce. These organizations are registered under various nonprofit tax codes in U.S. law, including those for charitable, advocacy, labor, and professional entities. To supplement this information, we scraped “About” pages from 1,062,554 organizational websites, enabling us to classify them into 15 organizational types and identify four key forms of civic opportunity: volunteering, membership, public events, and civic or political engagement.

We acknowledge, however, that our dataset does not capture the full spectrum of civic life. Informal neighborhood groups, mutual aid networks, online communities, and fiscally sponsored projects, many of which play critical roles in civil society, are excluded if they are not formally registered or do not file with the IRS. Some civic actors intentionally forgo incorporation to engage more freely in political or faith-based work. These groups are no less civic, but they are structurally difficult to observe using standardized administrative data. Despite these limitations, IRS records remain the most cost-effective and scalable source of structured information on U.S. civil society. While imperfect, they offer a reliable foundation for measuring civic infrastructure across contexts, and our integration of digital trace data helps mitigate some of the blind spots in administrative filings.

The MapAgora datasets are constructed by linking administrative records (IRS tax filings) with digital trace data (organization websites), providing both breadth and depth in our measurement of civic infrastructure. These data illuminate where civic opportunities are available and which types of organizations provide them. Our approach enhances the utility of IRS records by extending and restructuring them for scholarly and practical use through large-scale data integration and computational classification.

We released five datasets, each designed to support a different level of analysis:

- Organization-level dataset** (de-identified, Dataset 1): This dataset includes over 1.7 million nonprofit organizations and captures whether each provides civic opportunities, the types of opportunities offered, the predicted organizational category, and financial indicators drawn from tax filings. All identifying information is removed to protect privacy and reduce the risk of misinterpretation.
- ZIP code-level datasets (Datasets 3, 5)**: These aggregated datasets summarize the number and types of civic opportunity providing organizations, as well as demographic and socioeconomic characteristics, for nearly 31,000 ZIP codes across the U.S.
- County-level dataset (Datasets 2, 4)**: Similarly, these datasets provide aggregated counts and organizational type breakdowns for over 3,200 counties, enabling regional comparison and spatial analysis of civic infrastructure.

Together, these datasets offer a comprehensive and scalable resource for studying the micro- and macro-dynamics of civic infrastructure in the United States. In the following sections of this article, we first describe how this data was collected and created using web scraping, machine learning, and rule-based text classification methods. Next, we validate that civic opportunity is conceptually distinct from financial resources at the organizational level, while aggregate patterns align with traditional measures of socioeconomic inequality. Finally, we outline how these datasets can be used to research the role of grassroots organizations in American local democracy and public policy.

Text Source	Mean Word Count	SD Word Count
IRS Mission Text	25.7	26.6
IRS Program Text	110.3	304.5
Organization Website Text	1077.0	6637.5

**Table 1.** Text length by source.

## Methods

To construct a comprehensive and interpretable dataset of civic associations in the United States, we implemented a multi-step process that integrated administrative records with large-scale digital trace data. This entire pipeline is reproducible and relies on the two open-source R packages we developed: **MapAgora** (v0.08 <https://github.com/snafagora/MapAgora>) for parsing and integrating administrative data, and **autotextclassifier** (v0.05 <https://github.com/snafagora/autotextclassifier/tree/main>) for machine-learning-based text classification.

Since December 31, 2021, the IRS no longer updates the Form 990 series data on AWS, where machine-readable filings had previously been made available. The **MapAgora** package nonetheless provides a transparent and reproducible blueprint for parsing, cleaning, and harmonizing digitized filings. Researchers and practitioners can use the package to structure extraction workflows, without navigating the complexity of each form's schema manually.

Our goal was to build a dataset that reflects not only the presence of nonprofit organizations but also the specific civic opportunities they provide. This section outlines each stage in the pipeline, from defining the sampling frame to classifying organizational types and civic opportunity categories.

**Step 1: Defining the sampling frame.** We began by constructing our sampling frame using the IRS Exempt Organizations Business Master File, a public registry of approximately 1.8 million organizations that have incorporated with the IRS to obtain and maintain tax-exempt status. Unlike datasets that rely solely on Form 990 filings, the Master File includes all registered tax-exempt organizations, including those not required to file annual returns, such as churches and certain faith-based groups. This allowed us to capture a broader and more inclusive view of the nonprofit sector.

By starting with the Master File and layering in Form 990 data, we ensured that our dataset captured both organizations with detailed financial and programmatic disclosures and those with limited or no reporting requirements. Of the approximately 1.8 million organizations listed in the IRS Master File, about 460,000 are required to file a full Form 990—either the standard 990, 990-EZ, or 990-PF—depending on the organization's size and type. Another 913,000 file only the abbreviated 990-N “postcard” form, while roughly 400,000 organizations are exempt from filing altogether. We begin by extracting filing data from those that do file and then supplement with website data from all available organizations.

**Step 2: Standardizing Form 990 variation.** Most tax-exempt organizations are required to file some version of IRS Form 990 each year, but the structure of these forms varies depending on the organization's size and type. Large nonprofit organizations submit the full Form 990, mid-sized organizations file the abbreviated 990-EZ, and small nonprofit organizations (with gross receipts under \$50,000) typically file the simplified 990-N. Private foundations must file Form 990-PF regardless of size. (See Table S1 in the Online Appendix for further detail).

To extract consistent information from structurally diverse Form 990 filings, we developed a set of Extensible Markup Language (XML) parsing functions. These functions first determine which version of the form an organization has filed (e.g., 990, 990-EZ, or 990-PF), and then extract key information such as mission statements, program descriptions, mailing addresses, and financial details. By accounting for the differing layouts and field structures of each form type, the parser creates a standardized dataset that supports large-scale classification and analysis across the entire 990 series.

**Step 3: Supplementing IRS records with organizational website data.** While IRS filings are useful for establishing organizational identity, they offer limited insight into what organizations actually do. Mission statements and program descriptions in Form 990s are typically brief, often no more than one or two sentences. Moreover, many organizations are only required to file abbreviated forms that contain little to no descriptive content. To supplement these sparse administrative records, we collected digital trace data from organizational websites, including homepage and “About” page content, which provided substantially richer text. On average, website content (1,077 words) was 10 times longer than IRS program descriptions (110 words) and 40 times longer than mission statements (26 words) (see Table 1).

**Step 4: Collecting and verifying organizational website URLs.** To obtain website URLs, we used a two-pronged approach. For organizations that filed Form 990, we extracted and verified the web address listed in the filing. For all organizations, including those with no IRS-reported URL, we conducted automated searches using the Bing Search API. These queries incorporated the organization's name and geolocation, often returning deep search results with verified URLs. When deep search data were unavailable, we selected the most plausible result from the general search output.

**Step 5: Web scraping and text extraction.** After identifying URLs for approximately 1.1 million organizations, we scraped text from both their homepage and “About” pages. We used rule-based logic to detect

Civic opportunity dimension	Regular expression
Membership	"(?<! [a-zA-Z])join"; "member"; "sign up"
Volunteering	"volunteer"; "get involved"; "getinvolved"
Holding events	"events"; "calendar"; "meeting"
Taking civic or political actions	"take action"; "takeaction"; "justice"; "social(.?)action"

**Table 2.** Civic opportunity dimensions and corresponding regular expressions for detection via webpage links. **Note:** The regular expressions listed for each civic opportunity dimension are used to identify relevant terms in organizational webpage links. Each expression includes keywords or patterns commonly associated with that form of engagement and is matched in a case-insensitive manner. For example, the expression `(?<! [a-zA-Z])join; member; sign up` detects links like `/join-us`, `/become-a-member`, or `/sign-up`, while avoiding partial matches (e.g., “enjoy” or “membership”). Similarly, the expression for volunteering captures URLs such as `/volunteer`, `/get-involved`, or `/getinvolved`. The expression for civic or political action includes variations like `/take-action`, or `/socialaction`, with `(.?)` allowing for flexible matches. These regular expressions help classify organizations based on the types of civic opportunities they publicly advertise.

common URL patterns (e.g., “about” or “about-us”) and performed content extraction while excluding boilerplate components such as navigation headers and footers.

**Step 6: Identifying civic opportunities and the types of organizations that provide them.** We then applied a two-stage data transformation process to identify civic opportunities and the types of organizations that provide them.

First, we used regular expressions to identify textual signals indicating whether an organization offered opportunities for membership, volunteering, public events, or civic/political engagement (see Table 2).

Second, we used a supervised machine learning approach to classify organizational types based on combined text from IRS records and website content. We manually labeled 9,112 organizations and trained an ensemble model that combines LASSO and XGBoost. The final model achieved approximately 90% accuracy in predicting organizational categories. The full codebook for human-labeled classifications is provided in the supplementary materials from<sup>8</sup>: [https://static-content.springer.com/esm/art%3A10.1038%2Fs41562-023-01743-1/MediaObjects/41562\\_2023\\_1743\\_MOESM1\\_ESM.pdf](https://static-content.springer.com/esm/art%3A10.1038%2Fs41562-023-01743-1/MediaObjects/41562_2023_1743_MOESM1_ESM.pdf).

This approach, illustrated in Fig. 1, builds on the methodology introduced by<sup>8</sup> and expands on it in additional detail. Beginning with the extraction of administrative records, we proceeded through geocoding, URL identification, large-scale web scraping, and classification using natural language processing and machine learning.

By integrating structured IRS records with unstructured web content, we produced a high-resolution dataset that captures both the breadth and depth of civic infrastructure across the United States. Every step in this pipeline is reproducible through public code and documentation, enabling other researchers to audit, replicate, or extend our workflow using the MapAgora and autotextclassifier R packages.

## Data Records

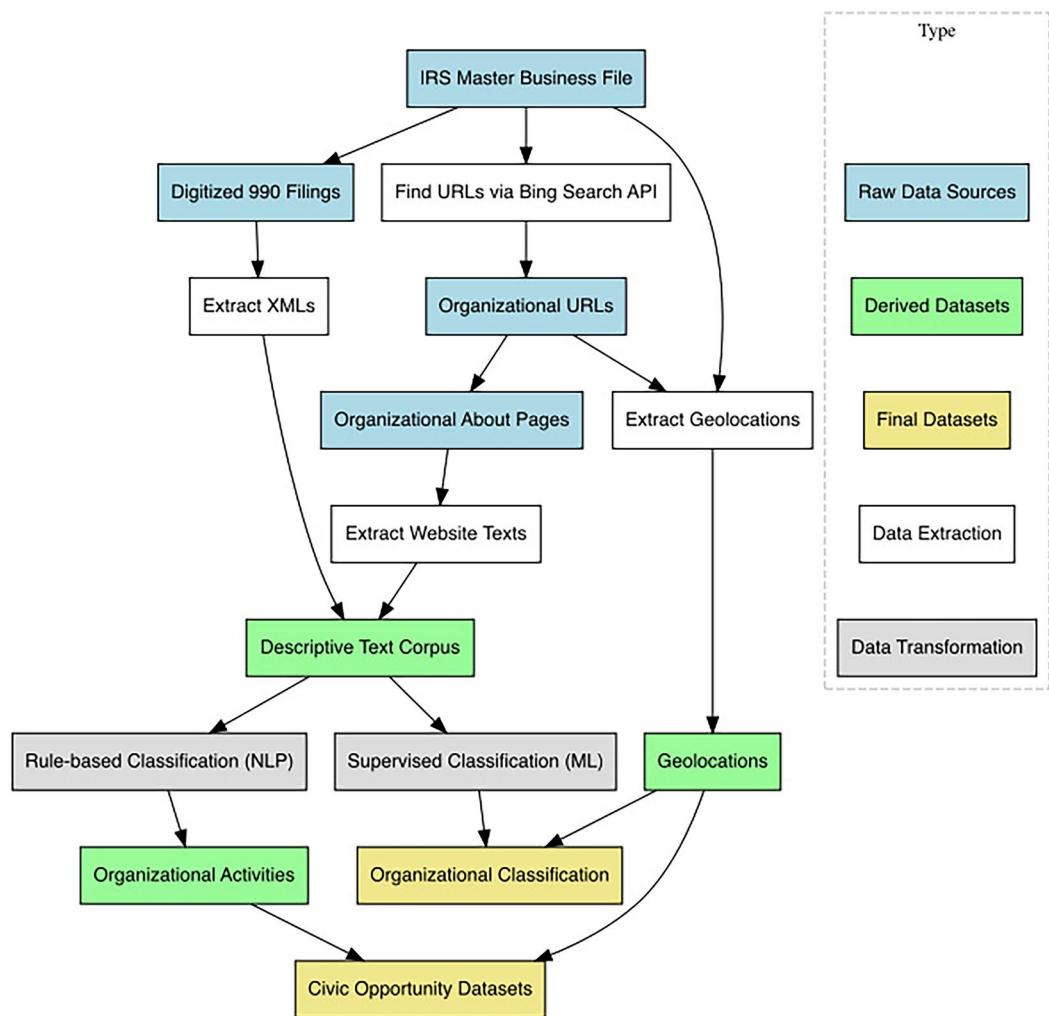
The MapAgora datasets are available via the GitHub repository ([https://github.com/snfagora/american\\_civic\\_opportunity\\_datasets](https://github.com/snfagora/american_civic_opportunity_datasets)) and the Harvard Dataverse (<https://doi.org/10.7910/DVN/IRCA7C>)<sup>24</sup>. This section provides detailed information on the availability, structure, and contents of the datasets being described.

This paper introduces five core datasets that capture the availability, intensity, and organizational composition of civic opportunities across U.S. ZIP codes and counties. Each dataset is available in .csv and .rds formats; the organization-level dataset (Dataset 1) is also available in .parquet format to support efficient processing of large-scale data. Designed to support research on place-based civic inequality and community resilience, these datasets support detailed analysis of where civic infrastructure is available and what types of organizations provide it, across ZIP codes and counties (Datasets 2–5).

Across all datasets, we define civic opportunity along four dimensions: offering membership, hosting public events, providing volunteer opportunities, and enabling civic or political action. Each organization is also classified into one of 15 types, such as political, professional, religious, or issue-based, allowing for rich analysis of the structure and diversity of civic infrastructure nationwide.

The organization-level dataset includes 1,774,798 de-identified nonprofit organizations (Dataset 1). Each observation includes binary indicators for civic opportunity provision (membership, volunteer, events, take\_action), a composite indicator (opp\_binary), an average indicator (opp\_mean), organizational type, and financial data drawn from IRS filings.

To protect privacy, we fully anonymize the organization-level dataset by removing identifiable information such as organization names and Employer Identification Numbers (EINs). The dataset includes only a row index (ID column) for reference. This de-identification step is especially important given that (1) some organizations may be politically sensitive or vulnerable to surveillance, and (2) the classifications used in this dataset are derived from predictive models applied to IRS and website text, which may contain labeling errors. While we are confident in the validity of the aggregated civic opportunity measures at the ZIP code and county levels, individual predictions may be noisy. Publishing organization identities alongside model-generated labels could risk misinterpretation by end users.



**Fig. 1** Graphical description of dataset creation.

Tables 3–5 summarize the key variables in the organizational-level dataset, along with their descriptions. Each row represents a single de-identified nonprofit organization. Key variables include:

- **Unique identifier:** `id`, a row index used as a reference; contains no identifying information.
- **Geographic identifiers:**
  - `state`: two-letter state abbreviation.
  - `city`: city name listed in the IRS record.
  - `FIPS`: county FIPS code.
  - `ZCTA`: ZIP Code Tabulation Area.
  - `is_po`: indicates whether the organization lists a P.O. Box as its mailing address (1 = yes, 0 = no).
- **Civic opportunity indicators:**
  - `membership`, `volunteer`, `events`, `take_action`: binary variables indicating whether the organization provides each type of civic opportunity (1 = provides the opportunity; 0 = does not provide it or information is unavailable).
  - `opp_binary`: equals 1 if the organization provides at least one civic opportunity; 0 otherwise.
  - `opp_mean`: the mean of the four civic opportunity binary indicators.
- **Organizational type:**
  - `predicted`: machine-learned classification of the organization (e.g., religious, political, professional).

Name	Description	Data source
id	Unique row index (used as a reference; not personally identifying)	Constructed (internal)
state	Two-letter US state abbreviation	IRS
city	City name listed in the IRS record	IRS
FIPS	County FIPS code, constructed from mailing address	Constructed (internal)
ZCTA	ZIP Code Tabulation Area, constructed from mailing address	Constructed (internal)
is_po	Indicator: uses a P.O. Box as mailing address (1 = yes, 0 = no)	IRS

**Table 3.** Identifiers in the organizational-level dataset (Dataset 1). **Note:** Variables labeled as “constructed (internal)” were created by us using information derived from available IRS records or organizational websites. These variables are not directly reported in the original data sources but were inferred or generated through data processing steps. For example, FIPS and ZCTA were constructed by geocoding mailing addresses. The id variable is a unique row index used for internal reference and contains no personally identifying information.

Name	Description	Data source
membership	Provides membership opportunities (1 = yes, 0 = no)	Website / IRS
volunteer	Provides volunteering opportunities (1 = yes, 0 = no)	Website / IRS
events	Hosts public events (1 = yes, 0 = no)	Website / IRS
take_action	Supports civic/political action (1 = yes, 0 = no)	Website / IRS
opp_binary	Equals 1 if the organization provides at least one civic opportunity; 0 otherwise	Website / IRS
opp_mean	Mean of the four civic opportunity indicators	Website / IRS
predicted	Machine-learned classification of organization type (e.g., religious)	Constructed (internal)

**Table 4.** Civic opportunity indicators and classifications in the organizational-level dataset (Dataset 1). **Note:** Variables labeled as “constructed (internal)” were created by us using information derived from available IRS records or organizational websites. These variables are not directly reported in the original data sources but were inferred or generated through data processing steps. For example, the predicted indicates the machine-learning classification of organizational type based on textual data.

Name	Description	Unit	Data source
grouping_value	Internal identifier for organizations that file joint tax returns, often part of federated networks (e.g., national organizations with chapters)	Organization group	IRS
asset_amt	Total assets	Organization	IRS
income_amt	Total income	Organization	IRS
revenue_amt	Total revenue	Organization	IRS

**Table 5.** Attributes and financials in the organizational-level dataset (Dataset 1).

- **Federated indicator:**
  - grouping\_value: an anonymized internal identifier used to track organizations that file joint returns, often within federated organizational structures (e.g., national networks with local chapters).
- **Financial attributes:**
  - asset\_amt: total assets.
  - income\_amt: total income.
  - revenue\_amt: total revenue.

The ZIP code- and county-level datasets aggregate civic opportunity indicators alongside community-level characteristics. Datasets 2, 3 include civic opportunity counts and scores for 3,281 counties and 30,988 ZIP codes. Datasets 4, 5 summarize the organizational types (e.g., religious, political, professional) providing those opportunities across geographies. Datasets 4, 5 cover 3,124 counties and 25,635 ZIP codes, reflecting 4.8% fewer counties and 17.3% fewer ZIP codes than Datasets 2, 3 due to insufficient text data for organizational classification.

Tables 6, 7 present the key variables in Datasets 2, 3 (county- and ZIP code-level civic opportunity counts), while Table 8 covers those in Datasets 4, 5 (county- and ZIP code-level civic opportunity organizational types). Table 9 lists the demographic and socioeconomic variables derived from the American Community Survey (ACS) and included in Datasets 2–5.

In Datasets 2, 3, each row represents a single county or ZIP code. Key variables include:

Name	Description
state	Two-letter US state abbreviations
FIPS	County FIPS code
n	Number of nonprofit organizations
civic_org_sum	Count of civic opportunity organizations
civic_org_sum_normalized	Civic opportunity organizations per capita
civic_opp_sum	Count of civic opportunity scores
civic_opp_sum_normalized	Civic opportunity scores per capita
civic_opp_index	Quantile-based index of normalized civic opportunity scores
volunteer_sum	Count of volunteering organizations
volunteer_sum_normalized	Volunteering organizations per capita
membership_sum	Count of membership organizations
membership_sum_normalized	Membership organizations per capita
take_action_sum	Count of civic or political action organizations
take_action_sum_normalized	Civic or political action organizations per capita
events_sum	Count of public event organizations
events_sum_normalized	Public event organizations per capita

**Table 6.** Civic opportunity indicators in the county-level dataset (Dataset 2). **Note:** state, FIPS, n are derived from IRS data. All other variables are sourced from either organizations' websites or their IRS tax filings.

Name	Description
state	Two-letter US state abbreviations
ZCTA	ZIP Code Tabulation Area
n	Number of nonprofit organizations
civic_org_sum	Count of civic opportunity organizations
civic_org_sum_normalized	Civic opportunity organizations per capita
civic_opp_sum	Count of civic opportunity scores
civic_opp_sum_normalized	Civic opportunity scores per capita
civic_opp_index	Quantile-based index of normalized civic opportunity scores
volunteer_sum	Count of volunteering organizations
volunteer_sum_normalized	Volunteering organizations per capita
membership_sum	Count of membership organizations
membership_sum_normalized	Membership organizations per capita
take_action_sum	Count of civic or political action organizations
take_action_sum_normalized	Civic or political action organizations per capita
events_sum	Count of public event organizations
events_sum_normalized	Public event organizations per capita

**Table 7.** Civic opportunity indicators in the ZIP code-level dataset (Dataset 3). **Note:** state, ZCTA, n are derived from IRS data. All other variables are sourced from either organizations' websites or their IRS tax filings.

- **Geographic identifiers:**

- state: two-letter state abbreviation.
- FIPS: county FIPS code.
- ZCTA: ZIP Code Tabulation Area.

- **Organizational counts:**

- n: total nonprofit organizations.
- civic\_org\_sum: total civic opportunity organizations.
- membership\_sum: total organizations providing membership opportunities.
- volunteer\_sum: total organizations providing volunteer opportunities.
- events\_sum: total organizations providing public event opportunities.
- take\_action\_sum: total organizations providing political or civic action opportunities.

- **Composite civic opportunity scores:**

- civic\_opp\_sum: total opportunity score.

Name	Description	Unit	Dataset
FIPS	County FIPS code	County	Dataset 4
ZCTA	ZIP Code Tabulation Area	ZIP Code	Dataset 5
n	Count of civic opportunity organizations	County, ZIP Code	Datasets 4-5
class	Predicted category of organizations	County, ZIP Code	Datasets 4-5
freq	Frequency of predicted categories	County, ZIP Code	Datasets 4-5
primary_org_cat	Primary civic opportunity provider category	County, ZIP Code	Datasets 4-5

**Table 8.** Civic opportunity organization classifications in the county- and ZIP code-level datasets (Datasets 4-5). **Note:** FIPS is derived from IRS data. All other variables are sourced from either organizations' websites or their IRS tax filings.

Name	Description	Unit
TotalPopulation	Total population	County, ZIP Code
POV150	Persons living below 150% of the poverty level	County, ZIP Code
SNGPNT	Single-parent households	County, ZIP Code
BROAD	Households without broadband subscription	County, ZIP Code
NOHSDP	Adults (25+) without high school diploma	County, ZIP Code
UNEMP	Unemployment among people 16+ in the labor force	County, ZIP Code
REMNRT	Persons of racial or ethnic minority status	County, ZIP Code

**Table 9.** Demographic and socioeconomic variables in the county and ZIP code-level datasets (Datasets 2–5). **Note:** All variables are derived from the 2017–2021 American Community Survey (ACS).

- **Normalized civic opportunity indicators (per capita):**
  - civic\_org\_sum\_normalized: total number of civic organizations per capita.
  - civic\_opp\_sum\_normalized: total civic opportunities per capita.
  - civic\_opp\_index: quintile-based civic opportunity index, derived by dividing civic\_opp\_sum\_normalized into five equal-sized bins.
  - membership\_sum\_normalized: total number of organizations providing membership opportunities per capita.
  - volunteer\_sum\_normalized: total number of organizations providing volunteer opportunities per capita.
  - events\_sum\_normalized: total number of organizations providing public event opportunities per capita.
  - take\_action\_sum\_normalized: total number of organizations providing political or civic action opportunities per capita.

- **Demographic and sociodemographic indicators:**

- TotalPopulation: total population.
- POV150: poverty rate.
- SNGPNT: single-parent households.
- BROAD: households without broadband access.
- NOHSDP: adults without a high school diploma.
- UNEMP: unemployment rate.
- REMNRT: share of racial or ethnic minority residents.

In Datasets 4, 5, each row represents a unique geography–organization type pair. Key variables includes:

- **Geographic identifiers:**

- FIPS: county FIPS code.
- ZCTA: ZIP Code Tabulation Area.

- **Organizational counts:**

- n: number of civic opportunity providing organizations of a given type in the geography.

- **Organization type classification:**

- class: predicted organizational type (e.g., religious, political, professional).

Variable	Mean [95% CI: lower, upper]
Civic opportunity organizations per 100,000	125.22 [122.61, 127.84]
Civic opportunity scores per 100,000	59.64 [58.35, 60.94]
Membership opportunity organizations per 100,000	101.07 [98.82, 103.32]
Public event opportunity organizations per 100,000	82.28 [80.36, 84.21]
Volunteer opportunity organizations per 100,000	51.77 [50.55, 52.99]
Take action opportunity organizations per 100,000	4.13 [3.91, 4.35]

**Table 10.** Civic opportunity summaries at the county level.

Variable	Mean [95% CI: lower, upper]
Civic opportunity organizations per 100,000	178.18 [172.01, 184.34]
Civic opportunity scores per 100,000	84.33 [81.46, 87.21]
Membership opportunity organizations per 100,000	145.90 [140.44, 151.37]
Public event opportunity organizations per 100,000	109.84 [106.57, 113.10]
Volunteer opportunity organizations per 100,000	73.53 [70.96, 76.09]
Take action opportunity organizations per 100,000	8.45 [7.71, 9.19]

**Table 11.** Civic opportunity summaries at the ZIP code level.

- **Relative frequency:**
  - `freq`: the proportion of civic opportunity organizations in the geography that belong to the given class.
- **Primary provider type:**
  - `primary_org_cat`: the most common civic opportunity organization type in the geography; appears once per geographic unit.

Together, these datasets offer a comprehensive and scalable resource for analyzing spatial inequality in civic infrastructure, enabling researchers to identify where civic opportunities are concentrated or lacking, and how access patterns vary across communities and regions in the United States. The following tables and figures illustrate how these datasets reveal variation in civic opportunity across counties and ZIP codes, and identify the types of organizations most responsible for providing civic access.

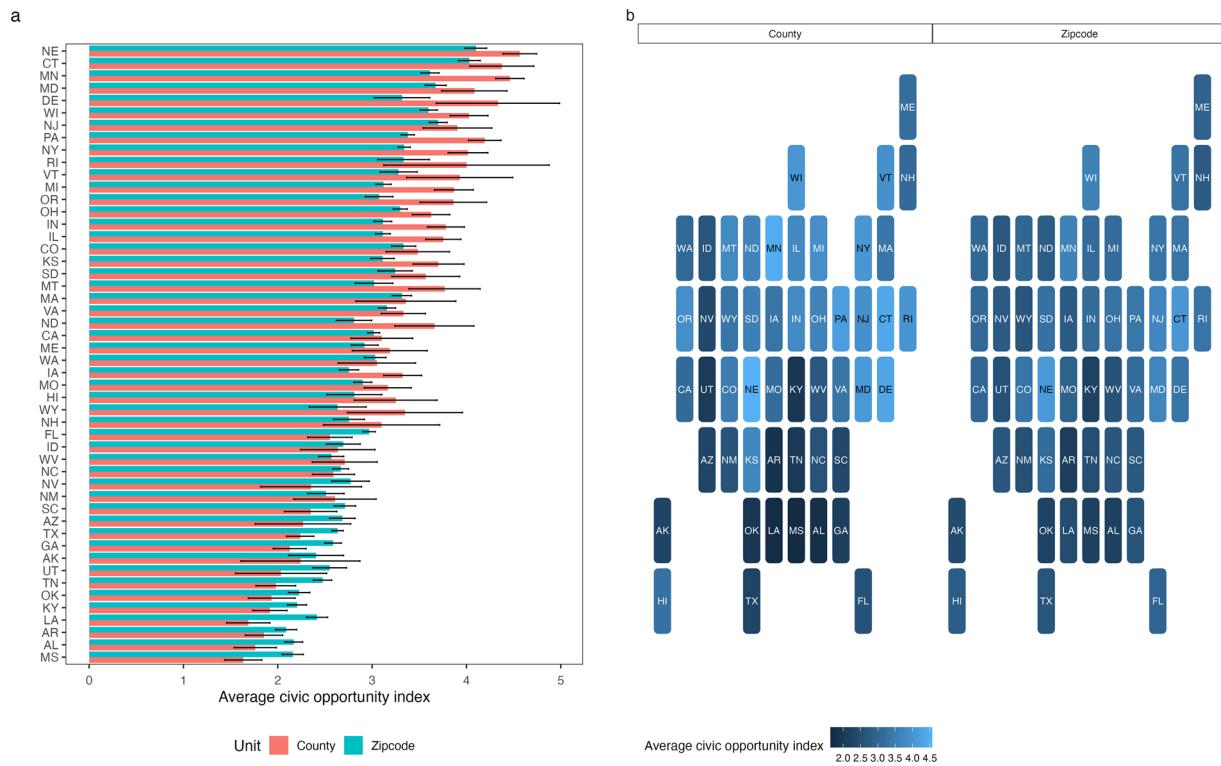
Tables 10, 11 summarize civic opportunity indicators at the county and ZIP code levels, respectively. On average, counties have 125.2 civic opportunity providing organizations (95% CI: 122.61–127.84) and a civic opportunity score of 59.6 per 100,000 residents (95% CI: 58.35–60.94). ZIP codes show higher rates due to smaller population denominators, with an average of 178.2 organizations (95% CI: 172.01–184.34) and a civic opportunity score of 84.3 (95% CI: 81.46–87.21). Across both geographic levels, membership is the most common form of civic opportunity, followed by public events, volunteering, and civic or political action, which remains the least common.

Figure 2 illustrates geographic variation in civic opportunity using a five-point index: Panel A shows group-level averages, while Panel B maps state-level differences using uniform cartograms. The results highlight substantial disparities across counties and ZIP codes. Figure 3 identifies the dominant type of civic opportunity provider in each area, revealing that fraternal and religious organizations account for the largest share nationwide at both the county and ZIP code levels.

To ensure data quality and consistency across geographic units, we applied a series of preprocessing steps and exclusion criteria when constructing the aggregated datasets. These steps were designed to improve the accuracy of spatial analysis, minimize measurement error, and address known limitations in the underlying data. The following discussion outlines our decisions regarding geographic boundaries, organizational address information, the treatment of federated structures, and the identification of civic deserts—areas where nonprofit organizations exist but do not offer any of the four defined forms of civic opportunity (membership, volunteering, public events, or civic/political action).

**Geographic boundaries.** We use ZIP codes rather than townships or municipal boundaries to ensure geographic consistency, as township definitions vary widely across the U.S., and demographic data are often unavailable at that level.

**Organizational address information.** To identify locally embedded organizations, we use the mailing address listed in IRS filings as a proxy for constituency location. Following prior work<sup>8</sup>, we exclude organizations whose primary address is a P.O. Box. This affects approximately 23% of organizations in our initial sampling frame. While some P.O. Box-using organizations may be active and locally rooted, particularly those serving



**Fig. 2** Distribution of civic opportunity across U.S. States. Notes: Panel A displays the average civic opportunity index by group, with error bars indicating 95% confidence intervals. Panel B illustrates state-level variation using uniform cartogram heatmaps. To facilitate meaningful comparison across states, we converted raw civic opportunity scores into a standardized index ranging from 1 to 5, where 1 indicates the lowest level of civic opportunity and 5 the highest. Each index bucket contains an equal number of observations.

vulnerable or rural populations, we exclude them to reduce false positives and ensure a consistent standard of local presence.

Excluding organizations with P.O. Box addresses has minimal impact on the overall distribution of organization types. On average, the absolute difference in the percentage of organizations classified into each type (e.g., political, professional, issue-based) between the full dataset and the version that excludes P.O. Boxes is 0.095 percentage points ( $SD = 0.101$ ). For more information, see Table S2 and Figure S1 in the Online Appendix.

**Federated organizations.** We also identify and retain 4,168 federated organizations (e.g., Girl Scouts, alumni associations) with multiple local chapters. When chapters are separately incorporated, we treat them as distinct organizations and assign them to separate ZIP codes or counties. However, if multiple chapters from the same parent organization are located within a single geographic unit, we count them only once to avoid duplication.

**Civic deserts.** Finally, we note that some places can be considered civic deserts, areas where nonprofit organizations exist but do not offer any civic opportunities as defined by our framework. This includes 3% of counties and 17% of ZIP codes in our dataset.

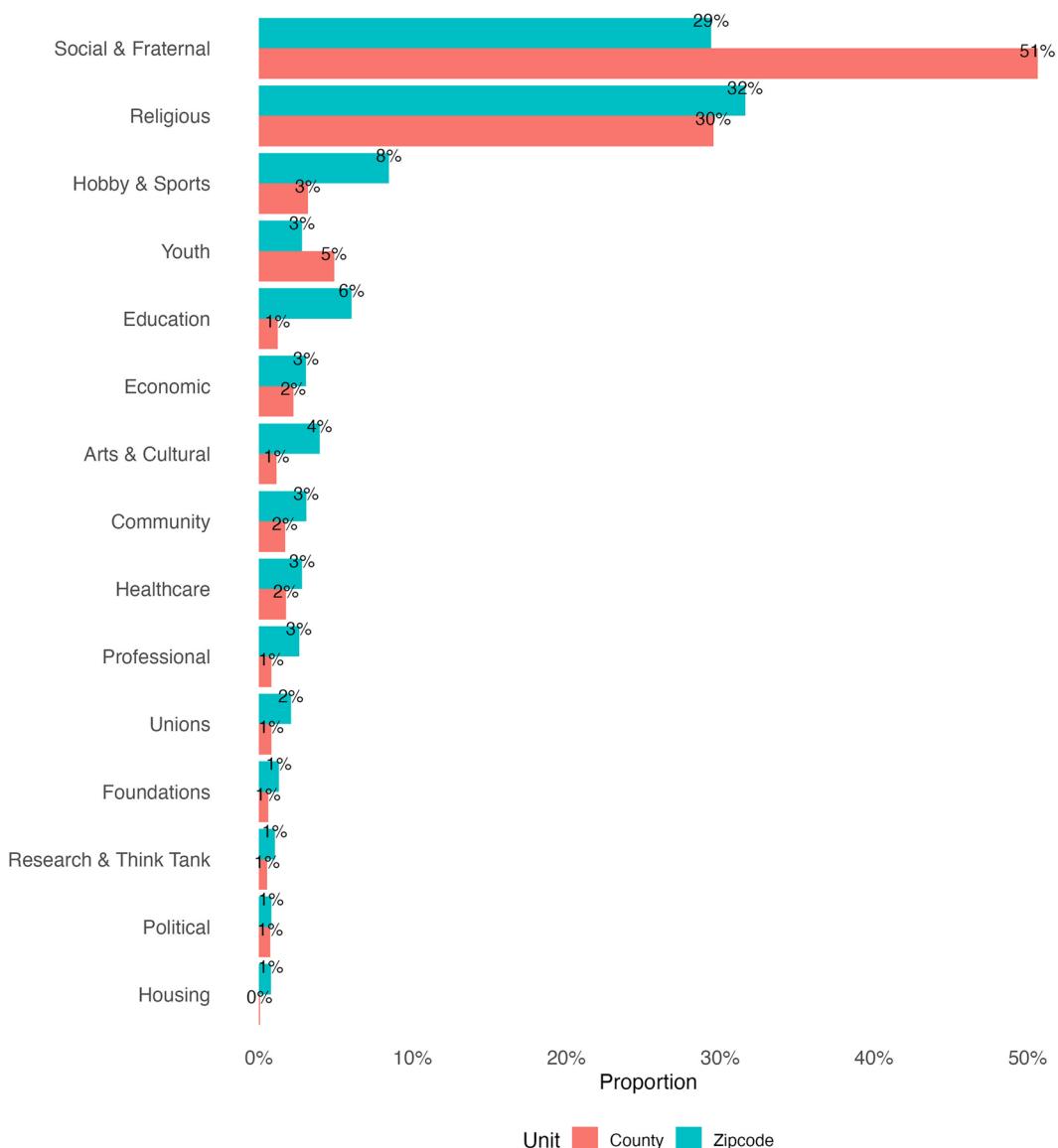
Table 12 summarizes all files included in the MapAgora dataset release, listing their file name, formats, dimensions, availability on GitHub and Harvard Dataverse, and exact directory paths within the GitHub repository. This comprehensive reference is intended to help users quickly locate and access the full set of data files.

### Technical Validation

We take several steps to ensure that the MapAgora datasets are high quality and support accurate interpretation and responsible use. Our validation strategy includes tests at the organization and aggregated levels.

We begin by validating the organization-level dataset, which includes more than 1.7 million de-identified nonprofit organizations. In Fig. 4, we examine the relationship between an organization's provision of civic opportunities and its financial resources. The y-axis displays the median values of assets, income, and revenue, and 95% confidence intervals are computed using a nonparametric bootstrap with 1,000 resamples.

We find that organizations offering more civic opportunities tend to have fewer financial resources. Specifically, those offering only one form of civic opportunity (e.g., membership, volunteering, events, or political action) tend to have the highest median levels of asset, income, and revenue, while organizations offering multiple forms of civic opportunity typically have lower financial resources. This seemingly counterintuitive result reflects the fact that many of the most civically active organizations are religious and social fraternal



**Fig. 3** Primary type of civic opportunity providing organizations in U.S. Counties and ZIP Codes. Notes: The primary organization type refers to the category comprising the largest share of civic opportunity providers in each area, shown in red at the county level and blue at the ZIP code level.

groups, which often operate with modest financial means. For example, foundations comprise 15.9% of organizations that do not provide civic opportunities, but only 3.3% of those offering low or single-type civic opportunities and 2.4% of those providing multiple forms. These patterns support the conceptual distinction between civic opportunity and organizational wealth and underscore the civic contributions of financially limited yet socially embedded organizations.

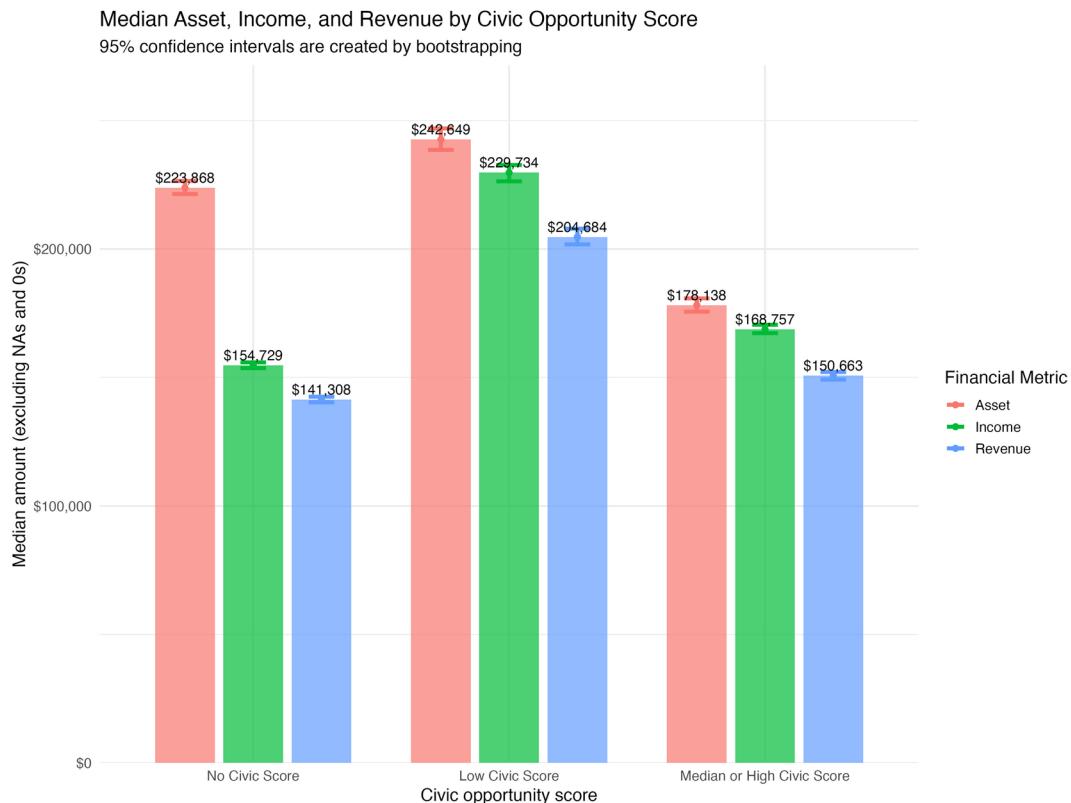
We now turn to the second and third datasets, validating our aggregate measures of civic opportunity at the county and ZIP code levels. As part of this validation, we expect civic opportunity scores to be higher in resource-rich areas and lower in resource-poor ones. To evaluate this expectation, we estimate a series of ordinary least squares (OLS) models examining the bivariate relationship between civic opportunity scores and several indicators of socioeconomic inequality.

The inequality indicators include: (1) the percentage of racial or ethnic minorities, (2) the proportion of individuals living below 150 percent of the federal poverty level, (3) the percentage of adults without a high school diploma, (4) the prevalence of single-parent households, (5) the unemployment rate among individuals aged 16 and older, and (6) the percentage of households without broadband internet access.

In these models, the civic opportunity score is as the dependent variable and is normalized per capita using the estimated population of the corresponding county or ZIP code. All inequality indicators are scaled from 0 to 100 to facilitate comparison and interpretation. The estimation model is specified as follows:

Dataset	File name	Available at	GitHub Path
Dataset 1	ind_org_df.parquet	GitHub, Dataverse	/raw_data/ind_org_df.parquet
Dataset 1	ind_org_df.csv	Dataverse only	—
Dataset 1	ind_org_df.rds	Dataverse only	—
Dataset 2	cnty_counts_cov.csv	GitHub, Dataverse	/data_outputs/cnty_counts_cov.csv
Dataset 2	cnty_counts_cov.rds	GitHub, Dataverse	/data_outputs/cnty_counts_cov.rds
Dataset 3	zcta_counts_cov.csv	GitHub, Dataverse	/data_outputs/zcta_counts_cov.csv
Dataset 3	zcta_counts_cov.rds	GitHub, Dataverse	/data_outputs/zcta_counts_cov.rds
Dataset 4	cnty_civic_org_type.csv	GitHub, Dataverse	/data_outputs/cnty_civic_org_type.csv
Dataset 4	cnty_civic_org_type.rds	GitHub, Dataverse	/data_outputs/cnty_civic_org_type.rds
Dataset 5	zcta_civic_org_type.csv	GitHub, Dataverse	/data_outputs/zcta_civic_org_type.csv
Dataset 5	zcta_civic_org_type.rds	GitHub, Dataverse	/data_outputs/zcta_civic_org_type.rds

**Table 12.** MapAgora dataset files: formats, availability, and file paths. Note: For file naming conventions, `ind_org_*` indicates individual organization-level datasets, `zcta_*` indicates ZIP code-level datasets, and `cnty_*` indicates county-level datasets.

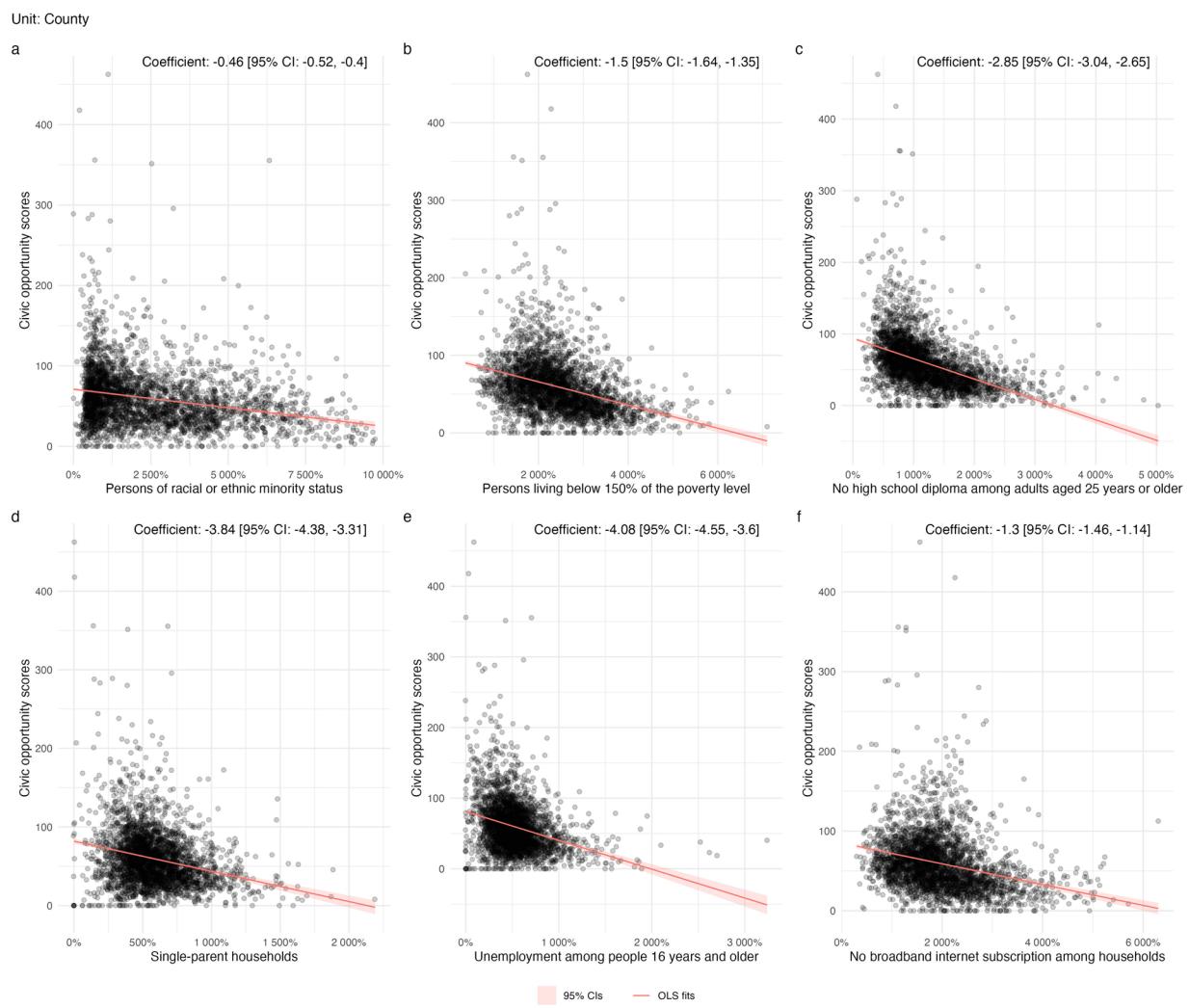


**Fig. 4** Median assets, income, and revenue of organizations by civic opportunity score. Notes: No civic score refers to organizations that do not provide any civic opportunities. Low civic score includes organizations that offer only one type of civic opportunity, as defined in the dataset: membership, volunteering, hosting events, or engaging in civic or political actions. Median or high civic score represents organizations that provide more than two of these civic opportunities to their communities. 95% confidence intervals are calculated using a nonparametric bootstrapping method with 1,000 resamples.

$$\text{Civic opportunity score per capita} = \beta_0 + \beta_1 \times \text{Inequality indicator} + \epsilon$$

where the unit of observation is a county or ZIP code.

The results show a consistent negative association between civic opportunity scores and inequality indicators at the county level (see Fig. 5). At the ZIP code level, most inequality indicators are also negatively associated with civic opportunity scores, except for the percentage of racial or ethnic minorities and single-parent households, which show no clear relationship (see Fig. 6).



**Fig. 5** Race, poverty, inequality indicators, and civic opportunity score at the county level. Notes: Race (Panel A), poverty (Panel B), and other inequality indicators (Panels C–F) plotted against the civic opportunity score at the county level. Solid lines represent OLS fits; shaded ribbons indicate 95% confidence intervals; and scattered dots show individual counties in the underlying data.

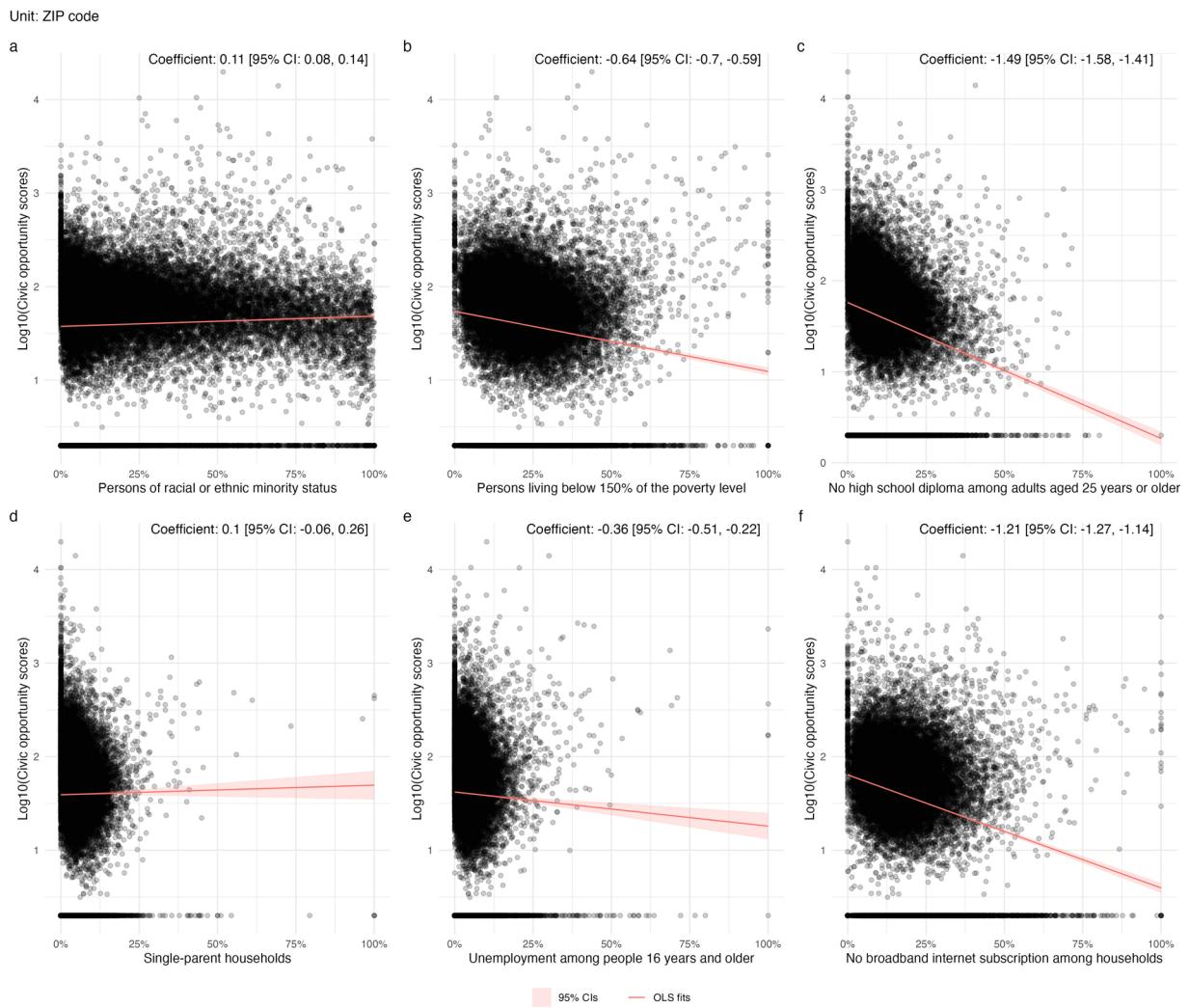
We apply a related validation approach to the fourth and fifth datasets, shifting the focus from the overall volume of civic opportunities to the types of organizations providing them. While the previous analysis examined how civic opportunity scores relate to socioeconomic inequality, this step evaluates whether communities with higher levels of inequality rely more heavily on religious and social fraternal organizations. To assess this, we construct a dependency measure defined as the proportion of civic opportunity providing organizations in each county or ZIP code that are religious or fraternal. For example, if a ZIP code has 100 civic opportunity organizations, 50 of which are religious or fraternal, the dependency score is 0.5.

This measure serves as a proxy indicator of civic infrastructure diversity. Religious and fraternal organizations often require shared beliefs or social ties for participation, which can limit access for those who do not share those characteristics. Many of these organizations are traditional in nature and may not offer inclusive spaces for newer or more diverse populations, including immigrants and LGBTQ individuals. In this sense, a heavy reliance on religious or fraternal organizations may signal constrained civic opportunity for some members of the community.

We therefore treat the dependency measure as a supplementary indicator of inequality and expect it to exhibit similar trends. To test this expectation, we estimated additional OLS models examining the bivariate relationship between the dependency measure and the same set of inequality indicators. The estimation model is specified as:

$$\text{Civic opportunity dependence} = \beta_0 + \beta_1 \times \text{Inequality indicator} + \epsilon$$

As shown in Fig. 7, the dependency measure at the county level exhibits more consistent positive associations with inequality indicators than at the ZIP code level. Five out of the six indicators are positively associated with the county-level dependency measure, while only three show a positive association at the ZIP code level.



**Fig. 6** Race, poverty, inequality indicators, and logged civic opportunity score at the ZIP code level. Notes: Race (Panel A), poverty (Panel B), and other inequality indicators (Panels C-F) plotted against the logged civic opportunity score at the ZIP code level. Solid lines represent OLS fits; shaded ribbons denote 95% confidence intervals; and scattered dots represent individual ZIP codes in the underlying data.

Across both aggregated datasets, we find that ZIP code-level measures are generally less reliable than their county-level counterparts when assessing relationships with common inequality indicators.

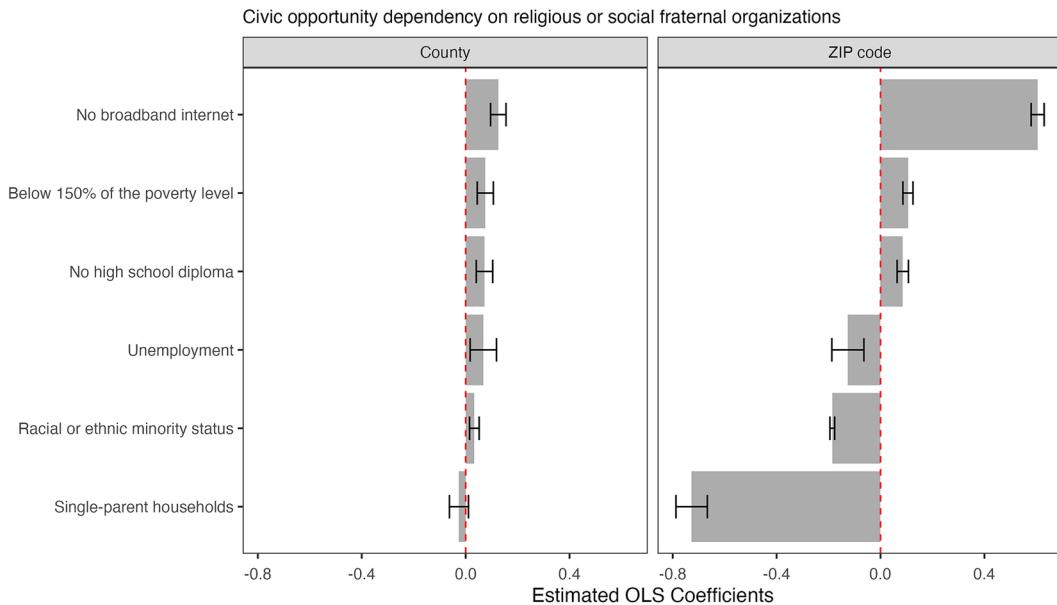
We suggest two possible reasons for this discrepancy. First, the locations of organizations (i.e., mailing addresses) may not accurately reflect where their constituents reside. While people move frequently, the physical location of an organization tends to remain fixed. Because ZIP codes are more fine-grained geographic units, they are more sensitive to this mismatch, making ZIP code-level data more prone to misrepresenting the actual distribution of civic opportunities.

Second, although ZIP codes offer greater geographic granularity than counties, they are conceptually limited as units for analyzing access to civic infrastructure. Most organizations do not define their constituencies by ZIP code boundaries, nor do they aim to serve only residents within a single ZIP code. This mismatch weakens the construct validity of ZIP codes for measuring the spatial distribution of civic opportunities.

### Usage Notes

As the most comprehensive dataset on civic opportunity in the United States, MapAgora enables researchers to examine not just what people think or do but also the opportunities in civic life in which they participate. Existing datasets on social capital often focus on individual attitudes, preferences, or behaviors. In contrast, MapAgora asks a more structural question: Which communities have access to meaningful pathways for engagement? Building on the framework developed by<sup>8</sup>, MapAgora centers on civic infrastructure, the organizational environments and resources that make participation possible.

The dataset supports research at multiple geographic and analytical scales. This flexibility enables both a fine-grained analysis of local civic ecosystems and broader investigations of regional or national patterns. For example, researchers might assess how civic infrastructure varies across rural and urban areas, or how



**Fig. 7** Inequality indicators and civic opportunity dependency on religious or social fraternal organizations. Notes: The association between race, poverty, and other inequality indicators and civic opportunity dependency on religious or social fraternal organizations at the county (left) and ZIP code (right) levels. The bars represent estimated OLS regression coefficients, and the error bars indicate 95% confidence intervals.

aggregated civic opportunity scores relate to policy outcomes or political behavior at the community or state level.

Understanding the distribution of civic opportunities is crucial for diagnosing civic decline in the United States<sup>25–27</sup>. Without comprehensive data, researchers and practitioners often rely on the most visible or well-resourced organizations, overlooking the smaller, locally embedded groups that make up the backbone of civic life in many communities<sup>1,28</sup>. In fact, religious and fraternal organizations account for approximately 80% of civic opportunity providers at the county level and 61% at the ZIP code level. Though often modest in scale and resources, these groups are central to civic life, especially in historically underserved or rural communities. They also play a critical role in ethnoracial minority communities, where religious institutions in particular have long served as venues for civic learning, collective identity, and political engagement<sup>29–32</sup>.

MapAgora enables a wide range of empirical research. Scholars can examine how the presence or types of civic opportunities, such as volunteering, public events, or civic and political action, relate to outcomes such as voter registration, turnout, or other forms of political behavior<sup>14,33–36</sup>. The dataset also supports investigations into organizational resilience and crisis response. During the early stages of the COVID-19 pandemic, for instance, many communities organized mutual aid networks that depended on preexisting civic infrastructure and local organizational capacity. Communities with stronger civic networks were better equipped to mobilize quickly and effectively<sup>8</sup>, a pattern echoed in research on disaster recovery and resilience<sup>37,38</sup>.

Second, civic opportunity data offer a framework for understanding why policy implementation varies across communities. Many public policies, especially in areas like public health, education, housing, and environmental action, require not just top-down directives but also active cooperation from civil society<sup>39–41</sup>. For example, while over 70% of residents in Marin County, California were fully vaccinated during the COVID-19 pandemic, fewer than 30% were vaccinated in Sweetwater County, Wyoming<sup>42</sup>. These disparities cannot be explained by demographics or partisanship alone. As shown by<sup>8</sup>, local civic infrastructure significantly shaped the effectiveness of the pandemic response. Similar patterns are likely to hold across a range of policy domains, underscoring the importance of civic opportunity data in advancing research on democratic capacity and policy equity.

Finally, for researchers conducting case studies or working in specific local contexts, MapAgora can also serve as a foundation for targeted discovery or supplemental data collection. Users may choose to manually audit a small number of counties or ZIP codes or cross-reference MapAgora with external sources, such as Google Maps, Yelp, or local directories, to identify additional civic organizations, such as informal neighborhood groups, mutual aid networks, and online communities. Comparing these locally identified organizations to those included in MapAgora can help researchers assess coverage and understand what kinds of organizations may be missing. Just as importantly, such comparisons can reveal why discrepancies arise, whether due to differences in formal registration status, online presence, or patterns of organizational disclosure. To aid this kind of exploration and support broader access, we also provide an interactive dashboard featuring both a county-level map and searchable table of civic opportunity indicators and their organizational types ([https://snfagora.github.io/agora\\_dashboard/](https://snfagora.github.io/agora_dashboard/)). These kinds of user-driven extensions are especially valuable for refining place-based research, mapping local civic ecosystems, or evaluating implementation contexts in public policy studies<sup>43</sup>.

## Code availability

The replication code for data generation, description, and validation is available at: [https://github.com/snffagora/american\\_civic\\_opportunity\\_datasets/tree/main](https://github.com/snffagora/american_civic_opportunity_datasets/tree/main).

Received: 30 December 2024; Accepted: 5 June 2025;

Published online: 08 July 2025

## References

1. Han, H. & Kim, J. Y. Civil society, realized: Equipping the mass public to express choice and negotiate power. *Ann. Am. Acad. Polit. Soc. Sci.* **699**, 175–185 (2022).
2. Fung, A. Associations and democracy: Between theories, hopes, and realities. *Annu. Rev. Sociol.* **29**, 515–539 (2003).
3. Hersh, E. *Politics is for Power: How to Move Beyond Political Hobbyism, Take Action, and Make Real Change*. Simon & Schuster (2020).
4. Han, H., McKenna, E. & Oyakawa, M. *Prisms of the People: Power & Organizing in Twenty-First-Century America*. University of Chicago Press (2021).
5. Gause, L. *The Advantage of Disadvantage*. Cambridge University Press (2022).
6. Ganz, M. *People, Power, Change: Organizing for Democratic Renewal*. Oxford University Press (2024).
7. Acemoglu, D. & Robinson, J. A. *The Narrow Corridor: States, Societies, and the Fate of Liberty*. Penguin UK (2019).
8. de Vries, M., Kim, J. Y. & Han, H. The unequal landscape of civic opportunity in America. *Nat. Hum. Behav.* **8**, 256–263 (2024).
9. Finkel, E. J. *et al.* Political sectarianism in America. *Science* **370**, 533–536 (2020).
10. Kinder, D. R. & Kam, C. D. *Us Against Them: Ethnocentric Foundations of American Opinion*. University of Chicago Press (2010).
11. van Bavel, J. J. *et al.* Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).
12. Han, H. & Arora, M. Igniting change: An evangelical megachurch's racial justice program. *Perspect. Polit.* **20**, 1260–1274 (2022).
13. de Tocqueville, A. *Democracy in America*. Trans. Mansfield, H. C. & Winthrop, D. University of Chicago Press (2000 [1835]).
14. Baltz, S. *et al.* American election results at the precinct level. *Sci. Data* **9**, 651 (2022).
15. Barari, S. & Simko, T. LocalView, a database of public meetings for the study of local politics and policy-making in the United States. *Sci. Data* **10**, 135 (2023).
16. Kyne, D. & Aldrich, D. P. Capturing bonding, bridging, and linking social capital through publicly available data. *Risk Hazards Crisis Public Policy* **11**, 61–86 (2020).
17. Chetty, R. *et al.* Social capital I: Measurement and associations with economic mobility. *Nature* **608**, 108–121 (2022).
18. Rupasingha, A., Goetz, S. J. & Freshwater, D. The production of social capital in US counties. *J. Socio-Econ.* **35**, 83–101 (2006).
19. Han, H. *How Organizations Develop Activists: Civic Associations and Leadership in the 21st Century*. Oxford University Press (2014).
20. Sampson, R. J. *Great American City: Chicago and the Enduring Neighborhood Effect*. University of Chicago Press (2012).
21. Ren, C. & Bloemraad, I. New methods and the study of vulnerable groups: Using machine learning to identify immigrant-oriented nonprofit organizations. *Socius* **8**, 10–21 (2022).
22. Pattillo, M. *Black on the Block: The Politics of Race and Class in the City*. University of Chicago Press (2010).
23. Levine, J. R. *Constructing Community: Urban Governance, Development, and Inequality in Boston*. Princeton University Press (2021).
24. Kim, J. Y. *et al.* MapAgora, civic opportunity datasets for the study of American local politics and public policy. Harvard Dataverse, V6. <https://doi.org/10.7910/DVN/IRCA7C> (2024).
25. Skocpol, T. & Fiorina, M. P. *Civic Engagement in American Democracy*. Brookings Institution Press (1999).
26. Skocpol, T. *Diminished Democracy: From Membership to Management in American Civic Life*. University of Oklahoma Press (2003).
27. Schlozman, K. L. *et al.* Organizations and the democratic representation of interests: What does it mean when those organizations have no members? *Perspect. Polit.* **13**, 1017–1029 (2015).
28. Hojnacki, M. & Kimball, D. C. The who and how of organizations' lobbying strategies in committee. *J. Polit.* **61**, 999–1024 (1999).
29. Harris, F. C. Something within: Religion as a mobilizer of African-American political activism. *J. Polit.* **56**, 42–68 (1994).
30. Owens, M. L. *God and Government in the Ghetto: The Politics of Church-State Collaboration in Black America*. University of Chicago Press (2008).
31. Cohen, C. J. *The Boundaries of Blackness: AIDS and the Breakdown of Black Politics*. University of Chicago Press (2009).
32. Harris-Perry, M. *Barbershops, Bibles, and BET: Everyday Talk and Black Political Thought*. Princeton University Press (2010).
33. LeRoux, K. Examining implementation of the National Voter Registration Act by nonprofit organizations: An institutional explanation. *Policy Stud. J.* **39**, 565–589 (2011).
34. LeRoux, K. & Krawczyk, K. Can nonprofit organizations increase voter turnout? Findings from an agency-based voter mobilization experiment. *Nonprofit Volunt. Sect. Q.* **43**, 272–292 (2014).
35. Owens, M. L. & Walker, H. L. The civic voluntarism of "custodial citizens": Involuntary criminal justice contact, associational life, and political participation. *Perspect. Polit.* **16**, 990–1013 (2018).
36. Walker, H. L. *Mobilized by Injustice: Criminal Justice Contact, Political Participation, and Race*. Oxford University Press (2020).
37. Aldrich, D. P. *Building Resilience: Social Capital in Post-Disaster Recovery*. University of Chicago Press (2012).
38. Aldrich, D. P. & Meyer, M. A. Social capital and community resilience. *Am. Behav. Sci.* **59**, 254–269 (2015).
39. Ostrom, E. Crossing the great divide: Coproduction, synergy, and development. *World Dev.* **24**, 1073–1087 (1996).
40. Joshi, A. & Moore, M. Institutionalised co-production: Unorthodox public service delivery in challenging environments. *J. Dev. Stud.* **40**, 31–49 (2004).
41. Pressman, J. L. & Wildavsky, A. *Implementation: How Great Expectations in Washington Are Dashed in Oakland*. University of California Press (1973).
42. Lovelace, B. COVID cases rise in U.S. counties with low vaccination rates as Delta variant spreads. *CNBC*. <https://www.cnbc.com/2021/07/13/covid-cases-rise-in-us-counties-with-low-vaccination-rates-as-delta-variant-spreads.html> Accessed 6 Dec 2024 (2021).
43. Williams, S. *Data Action: Using Data for Public Good*. MIT Press (2022).

## Acknowledgements

We acknowledge funding from the SNF Agora Institute at Johns Hopkins University and the Carnegie Corporation of New York.

## Author contributions

Conceptualization: J.Y.K. and M.d.V.; Methodology: J.Y.K. and M.d.V.; Investigation: J.Y.K. and M.d.V.; Visualization: J.Y.K. and M.d.V.; Funding acquisition: H.H.; Project administration: M.d.V. and H.H.; Supervision: M.d.V. and H.H.; Writing—original draft: J.Y.K. and M.d.V.; Writing—review and editing: J.Y.K., M.d.V. and H.H.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-05353-6>.

**Correspondence** and requests for materials should be addressed to J.Y.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025