AGENCY FOR HEALTHCARE RESEARCH AND QUALITY
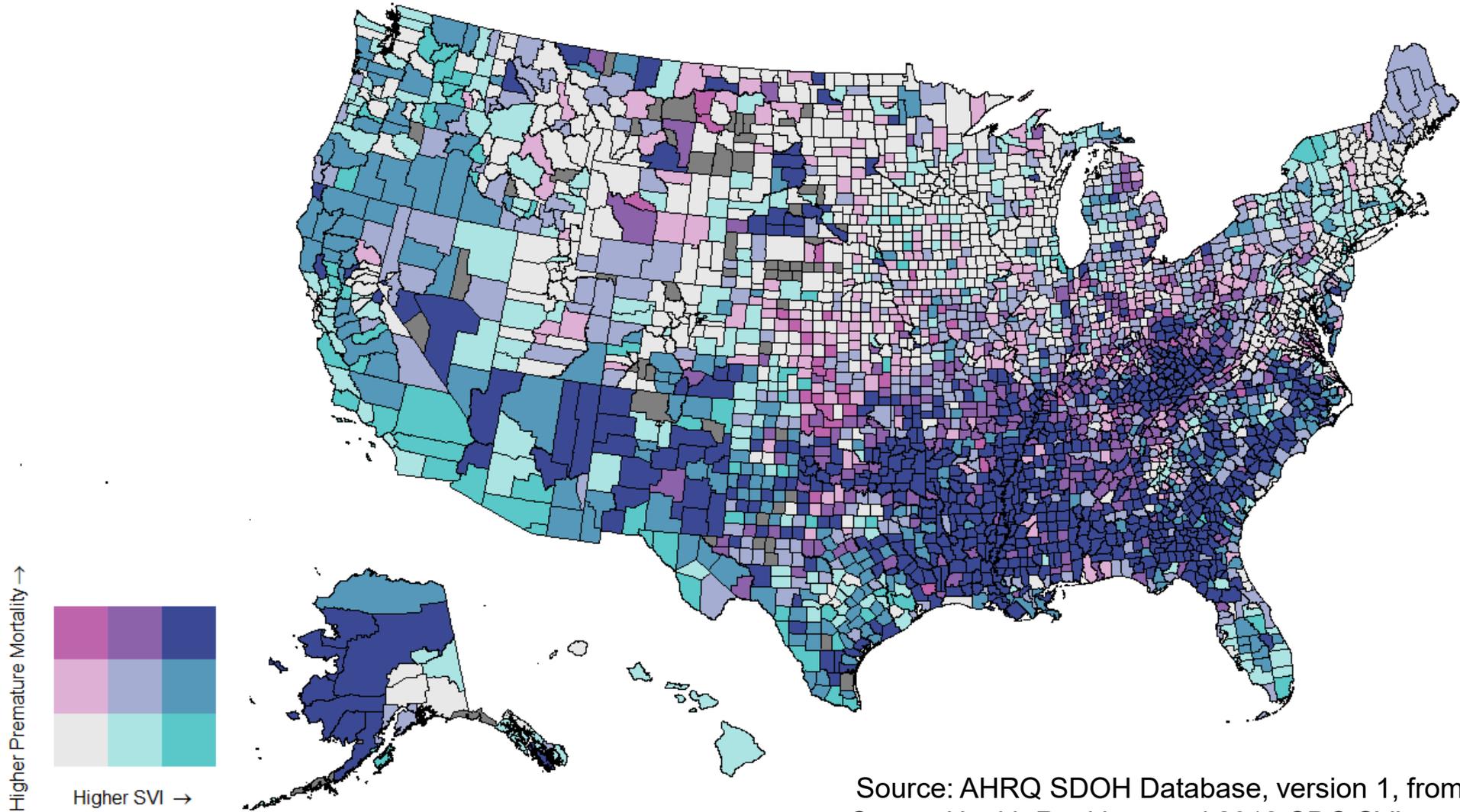
# AHRQ SDOH Database Overview

## July 2022

# Overview

- Context

- Updated Database Contents

  - ► Topics Covered in the Database

  - ► Geographic and Temporal Coverage

  - ► Documentation

  - ► Methodological Contributions

- Data Uses

# Premature Mortality and Social Vulnerability Index Across U.S. Counties in 2018



Source: AHRQ SDOH Database, version 1, from 2018 County Health Rankings and 2018 CDC SVI.
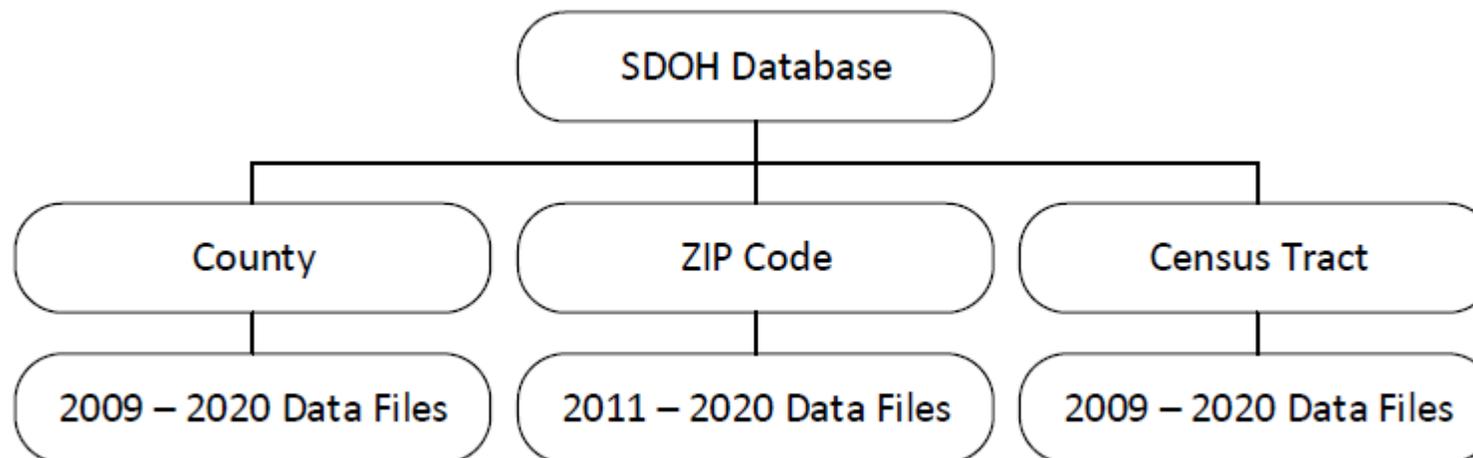
3

# SDOH Database Purpose

- Make community-level SDOH data easier to use in analyses to inform decisions to improve health outcomes
- Account for health differences across areas and identify effective interventions tailored to populations served
  - ► Patient-centered outcomes research
  - ► Healthcare delivery systems
  - ► Local, state, tribal, federal programs
- Analyses: data linkages, direct analysis
- Funded by HHS OS-PCORTF, AHRQ
  - ► This work was supported by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund Under Interagency Agreement 750119PE0K0036.

# Approach

- Develop, make publicly available, and iteratively improve database

- Environmental scan of public SDOH data
  - ▶ Identifies sources of SDOH variables and compiles a comprehensive inventory

- Initial files: public use "beta" data files
  - ▶ County: 2009-2018; ZIP Code tabulation area (ZCTA): 2011-2018
  - ▶ Use external public data
  - ▶ Covers all SDOH domains
  - ▶ Designed to link to other data by geography
  - ▶ Standardized data documentation, variable names, etc.

- Revise based on database input

# Overview of Updated Database

- Spans multiple years and three geographic levels
  - ► County
  - ► ZIP Code instead of ZCTA
  - ► Tract - new

- Draws from 44 different data sources, including over 17,000 variables across all geographic levels and years

```
                        ┌──────────────────┐
                        │  SDOH Database   │
                        └──────────────────┘
              ┌──────────────────┼──────────────────┐
      ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
      │    County    │   │   ZIP Code   │   │ Census Tract │
      └──────────────┘   └──────────────┘   └──────────────┘
             │                  │                  │
┌───────────────────┐ ┌───────────────────┐ ┌───────────────────┐
│ 2009 – 2020 Data  │ │ 2011 – 2020 Data  │ │ 2009 – 2020 Data  │
│      Files        │ │      Files        │ │      Files        │
└───────────────────┘ └───────────────────┘ └───────────────────┘
```
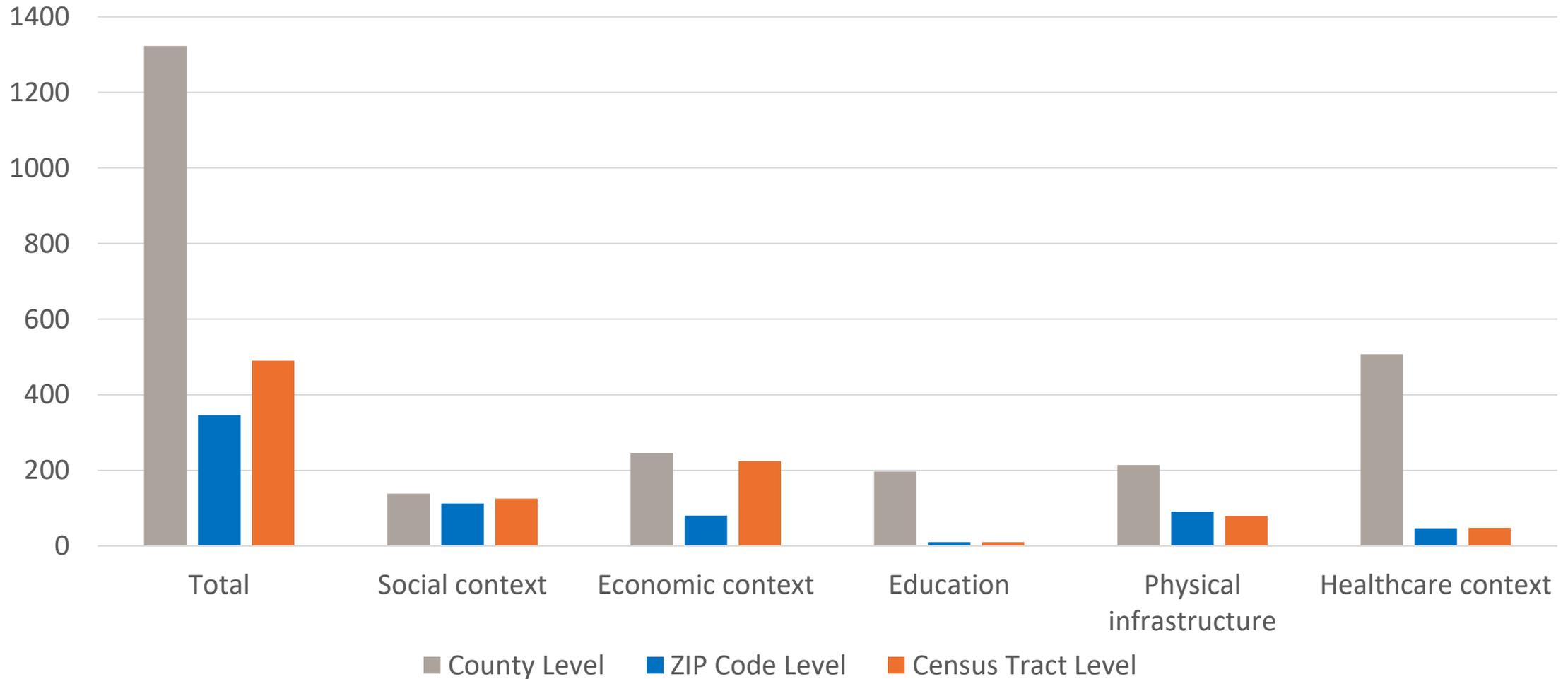
# Examples of Data Sources

# Community-Level SDOH Variables Organized by Domains and Topics

| Social Context | Economic Context | Education | Physical Infrastructure | Healthcare Context |
|---|---|---|---|---|
| • Demographics<br>• Disability<br>• Immigration<br>• Living conditions<br>• Segregation<br>• Socioeconomic disadvantage indices | • Employment<br>• Income<br>• Poverty | • Attainment<br>• Education funding<br>• Literacy<br>• Numeracy<br>• School system | • Access to exercise<br>• Crime<br>• Environment<br>• Food access<br>• Housing<br>• Industry composition<br>• Internet connectivity<br>• Migration<br>• Social services<br>• Transportation | • Characteristics of healthcare<br>  − Facilities<br>  − Providers<br>• Distance to providers<br>• Health behaviors<br>• Healthcare quality<br>• Health insurance status<br>• Utilization and cost<br>• Health outcomes |

Source: AHRQ SDOH Database, version 1.

# Variables in the SDOH Database by Domain and Geographic Level

Source: AHRQ SDOH Database, version 1.

# Examples of Variables Added from New Data Sources

| Topic | Data Source | Variables | Level |
|---|---|---|---|
| Socioeconomic Disadvantage Indices | • Census CRE<br>• HHS OMH, CDC | • Community Resilience Estimates<br>• Minority Health SVI | • County, Census tract<br>• County |
| Income | • Opportunity Atlas | • Incarceration rates by race/ethnicity<br>• Income mobility by race/ethnicity | • County, Census tract |
| Housing | • ACS | • High-cost rentals, # people in unit<br>• Units without fuel or plumbing | • County, ZIP, Census tract<br>• County, ZIP, Census tract |
| Environment | • EPA<br>• NOAA<br>• WUSTL Atmospheric Composition Analysis Group | • Climate and air quality<br>• Storms, floods, wildfires, tornadoes<br>• Particulate matter concentration | • County<br>• County<br>• County, Census tract |
| Food Access | • USDA Food Access Research Atlas<br><br>• Food Environment Atlas | • Low income and low access<br>• Distance from grocery by race/ethnicity<br>• Food banks, stores w/ WIC, farmers markets w/ SNAP | • County, Census tract<br>• County, Census tract<br><br>• County |
| Transportation | • EPA | • National Walkability Index | • Census tract |

Source: AHRQ SDOH Database, version 1.

# Examples of New Variables and Data Sources on Healthcare Context

| Subtopic | Data Source | Variables | Level |
|---|---|---|---|
| Distance to providers | • CMS Provider of Services<br><br>• Homeland Infrastructure Foundation-Level Data | • Distance to nearest clinic, ED, ICU, trauma center, obstetrics department<br>• Distance to nearest urgent care center | • ZIP Code, Census tract<br><br>• ZIP Code, Census tract |
| Provider competition | • American Hospital Association | • Herfindahl-Hirschman Index for short-term acute care hospitals | • County |
| Health behaviors and outcomes | • CDC Places | • Age-adjusted preventive services use<br>• Age-adjusted chronic conditions | • County, ZIP Code, Census tract |
| Utilization and costs | • Medicare Geographic Variation PUF<br>• LTC Focus | • Per capita Medicare spending<br>• Nursing home residents by race/ethnicity | • County<br>• County |
| Disparities in healthcare quality | • Mapping Medicare Disparities Tool | • Rates of avoidable hospitalizations, readmissions, ED visits by race/ethnicity | • County |
| Disparities in health outcomes | • CDC Atlas | • Avoidable heart disease and stroke deaths, by race/ethnicity | • County |

Source: AHRQ SDOH Database, version 1.

# Documentation

- Data source documentation
  - ► Overview of SDOH files and specifics by data source
  - ► Tables with data sources and variables by year
  - ► Variable information such as source table, original variable name, numerator/denominator
  - ► Information on
    - – Geography variables (e.g., rural-urban)
    - – Indices (e.g., Social Vulnerability Index)
  - ► Data notes

- Codebook
  - ► Excel files for each year
  - ► Can filter by domain, domain subtopics, data source

# User Notes

How do I find…?

- For each data source, the years and geographic levels available?
  - ▶ Documentation file, Table 3.1
- Variables available on a specific topic, and years and geographic levels?
  - ▶ Every codebook file, "All SDOH variable" tab, filter by domain and topic. Scroll right for geography and years available.
- A crosswalk of variable names from the beta to the updated database?
  - ▶ Documentation file, Appendix A
- Information on a specific variable?
  - ▶ Search the variable name in the documentation (variable construction, notes) and codebook (descriptive statistics in the county, ZIP, tract summary tabs)

# Documentation File Excerpts, American Community Survey

and ZIP Code SDOH files because they are not included in the ACS 5-year estimate
access nativity and citizenship data for Puerto Rico through the Puerto Rico Comn
The ACS does not contain 5-year estimates for other U.S. territories such as Ameri
Commonwealth of the Northern Mariana Islands, Guam, and the U.S. Virgin Islands
ZIP Code, or census tract files. Decennial census data are available for these territc
https://www.census.gov/data/datasets.html.

## Suppression Rules

To be reported in ACS 5-year data, estimates must represent a geographic area wi
size of at least 7,000 people. The Census Bureau Disclosure Review Board also set
rules to ensure confidentiality and protect respondent privacy. These rules pertain
number o
restriction
https://ww
suppressi

## Variable

We identif
Detailed T
collapsed
as the bas
availability

**Table 4.4**     Categories of ACS available for a limited time range

| Variable Category | Years Available | Number of Variables in SDOH files | Reason for Missing Data |
|---|---|---|---|
| Internet Connectivity | 2017-2020 | 19 | Internet and computer variables are provided in the ACS 5-year estimates for all geographies in 2017 and onwards. |
| Health insurance | 2013-2020 | 28 | Health insurance status and health insurance related federal poverty variables were first produced in the 5-year estimates for all geographies in 2013. |
| Disability Status | 2012-2020 | 5 | The first 5-year estimates on the disability status of individuals were produced for all geographies in 2012. |

## Suppression Rules

To be reported in ACS 5-year data, estimates must represent a geographic area with population size of at least 7,000 people. The Census Bureau Disclosure Review Board also sets additional rules to ensure confidentiality and protect respondent privacy. These rules pertain to a minimum number of cases required to publish a cell in a table, table cell restrictions, and table topic restrictions. For more information review the details provided here: https://www.census.gov/programs-surveys/acs/technical-documentation/data-suppression.html.

Source: AHRQ SDOH Database, version 1 Documentation.

# Documentation File Excerpt, Census County Business Patterns

## Data Set Notes

### Reliability

CCBP does not include margins of error.

More generally, CCBP covers most of the country's economic activity. The series excludes data on self-employed individuals, employees of private households, railroad employees, agricultural production employees, and most government employees. Businesses operating without an EIN, and businesses with an EIN but without employees, are excluded from the County Business Patterns universe.

The comparability of data over time may be affected by changes in industry classifications, definitions of establishments, establishment active status, and/or changes to geographic boundaries (actual or statistically defined areas).

### Geographic Levels

CCBP data include the numbers of establishments for each industry operating in a geographic area. Statistics are available on establishments at the U.S. level and by state, county, Metropolitan/Micropolitan Statistical Area, Combined Statistical Area (CSA), ZIP Code, and congressional district levels. If a county or ZIP Code does not have any establishments reported for a given NAICS code, there is no record for that geography-NAICS combination in the data file.

### Suppression Rules

In 2017, Census began suppressing data for cases that had fewer than three establishments in a geographic area. For comparison, in 2012, 59 percent of ZIP Codes with any grocery stores had two or fewer establishments. These ZIP Codes would be excluded from the 2017 – 2019 data files. CCBP does not include a suppression flag that would allow users to distinguish between fewer than three and zero, and therefore these observations are represented as missing in the SDOH.

Source: AHRQ SDOH Database, version 1.
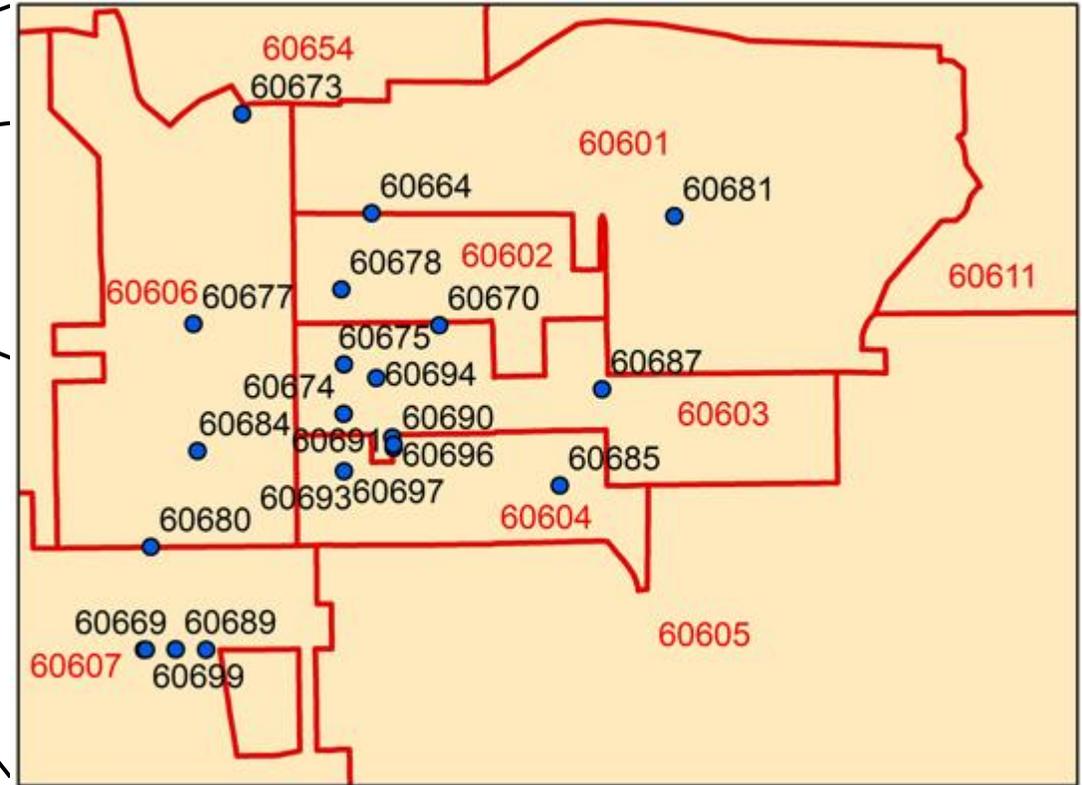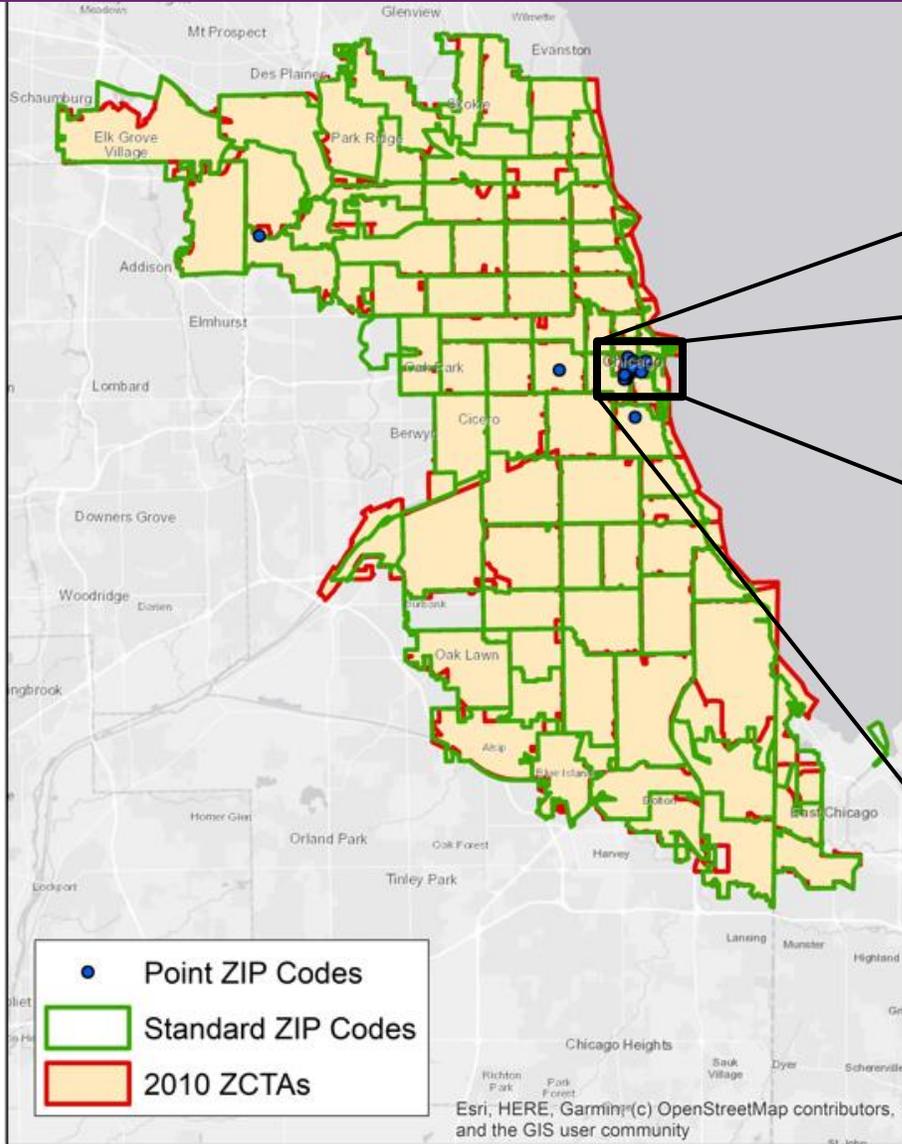
# Methodological Notes

- Naming conventions standardized between the datasets
  - ► All naming conventions include standardized abbreviations for the source data set
  - ► Names of variables and concepts are standardized across sources (e.g., similar abbreviations and names for racial/ethnic subgroups)
- In addition, variables were standardized to consistent geographic levels to facilitate ease of use
  - ► Included the creation of a ZIP Code-level file
  - ► Some source data elements needed transformation to align with traditional geographic concepts

Source: AHRQ SDOH Database, version 1.

# Creating a ZIP Code File

- Two specific issues:
  - ► Different geographic boundaries for ZIP Codes and ZCTAs
  - ► "Point ZIP Codes": special ZIP Codes representing mail delivery points
- For the SDOH database, we link ZIP Codes to ZCTAs using a 1:1 match when available, and when not relying on a spatial join
- Overall, roughly 76.4% of ZIP Codes match directly to a ZCTA, and 23.4% were matched using a spatial join
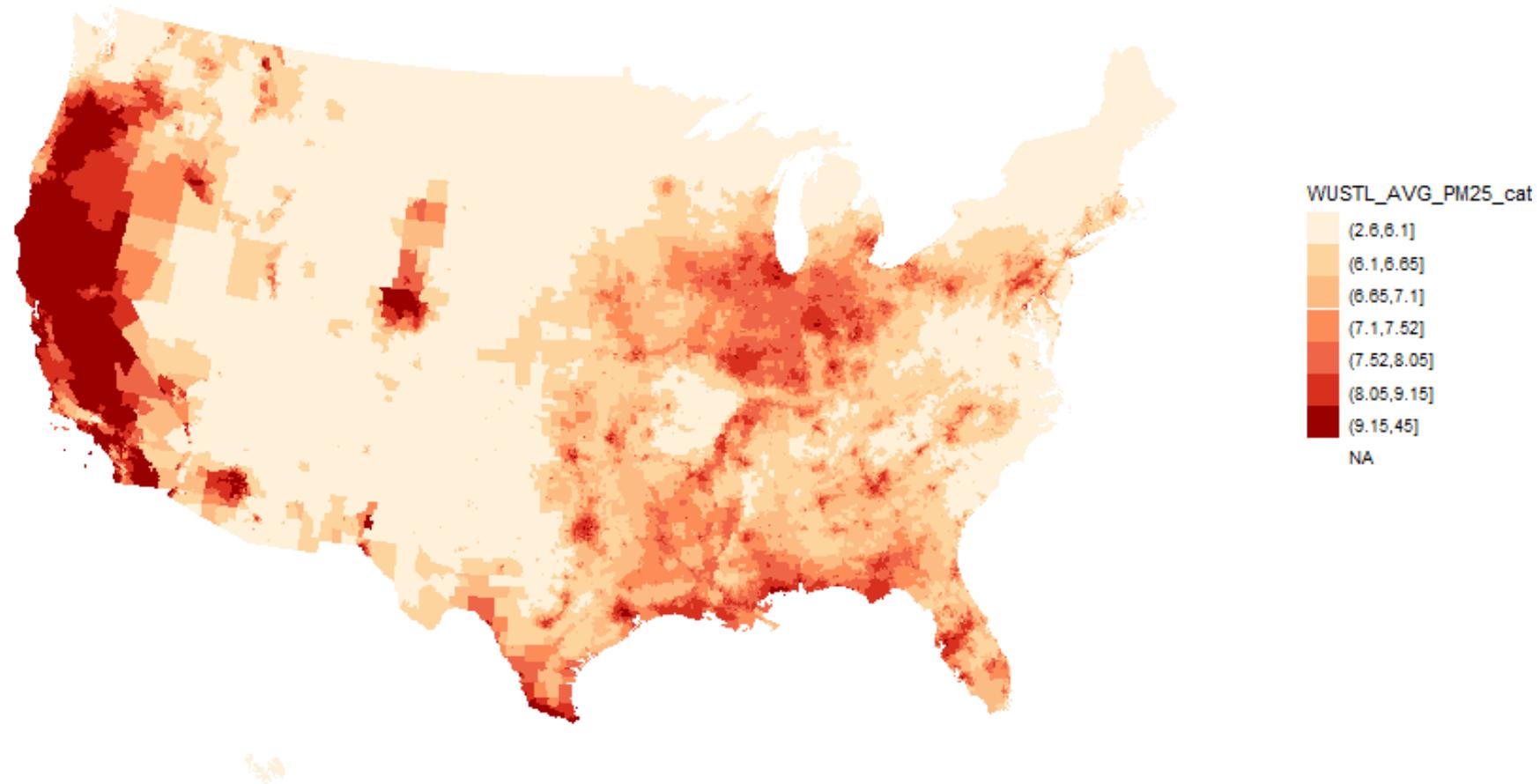
Source: AHRQ SDOH Database, version 1.

Source: AHRQ SDOH Database, version 1.
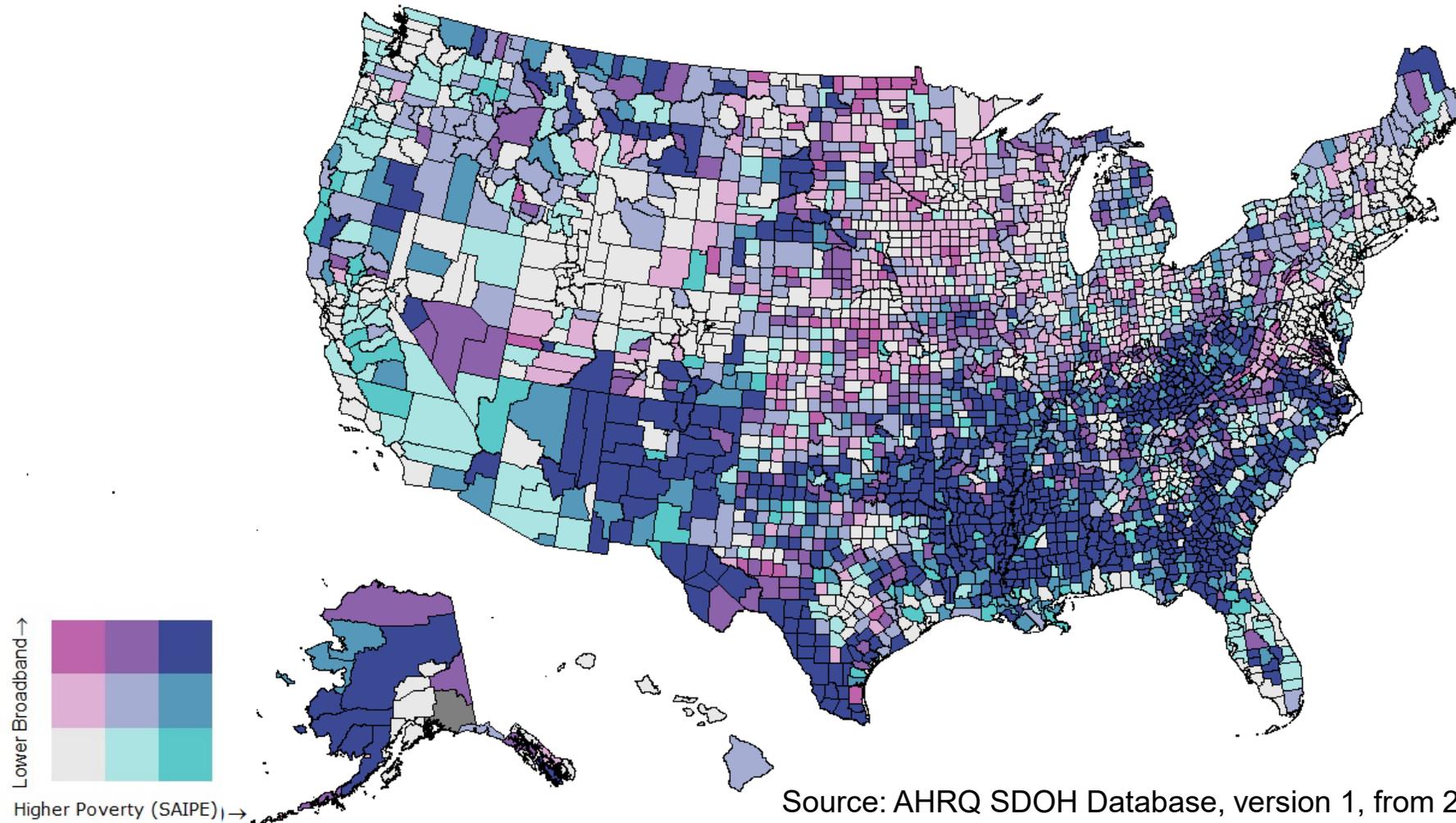
18

# Examples of Data Use

- The AHRQ SDOH Database provides a one-stop source for data to analyze characteristics of communities across the United States across multiple domains

- Can be linked with other data sources to conduct more detailed analyses

- Resource for patient-centered outcomes research

# Annual Tract-Level Mean of Air Particulate Matter Concentration, 2020



WUSTL_AVG_PM25_cat
- (2.6,6.1]
- (6.1,6.65]
- (6.65,7.1]
- (7.1,7.52]
- (7.52,8.05]
- (8.05,9.15]
- (9.15,45]
- NA

Source: AHRQ SDOH Database, version 1, from 2020 Washington University Saint Louis - Atmospheric Composition Analysis Group.
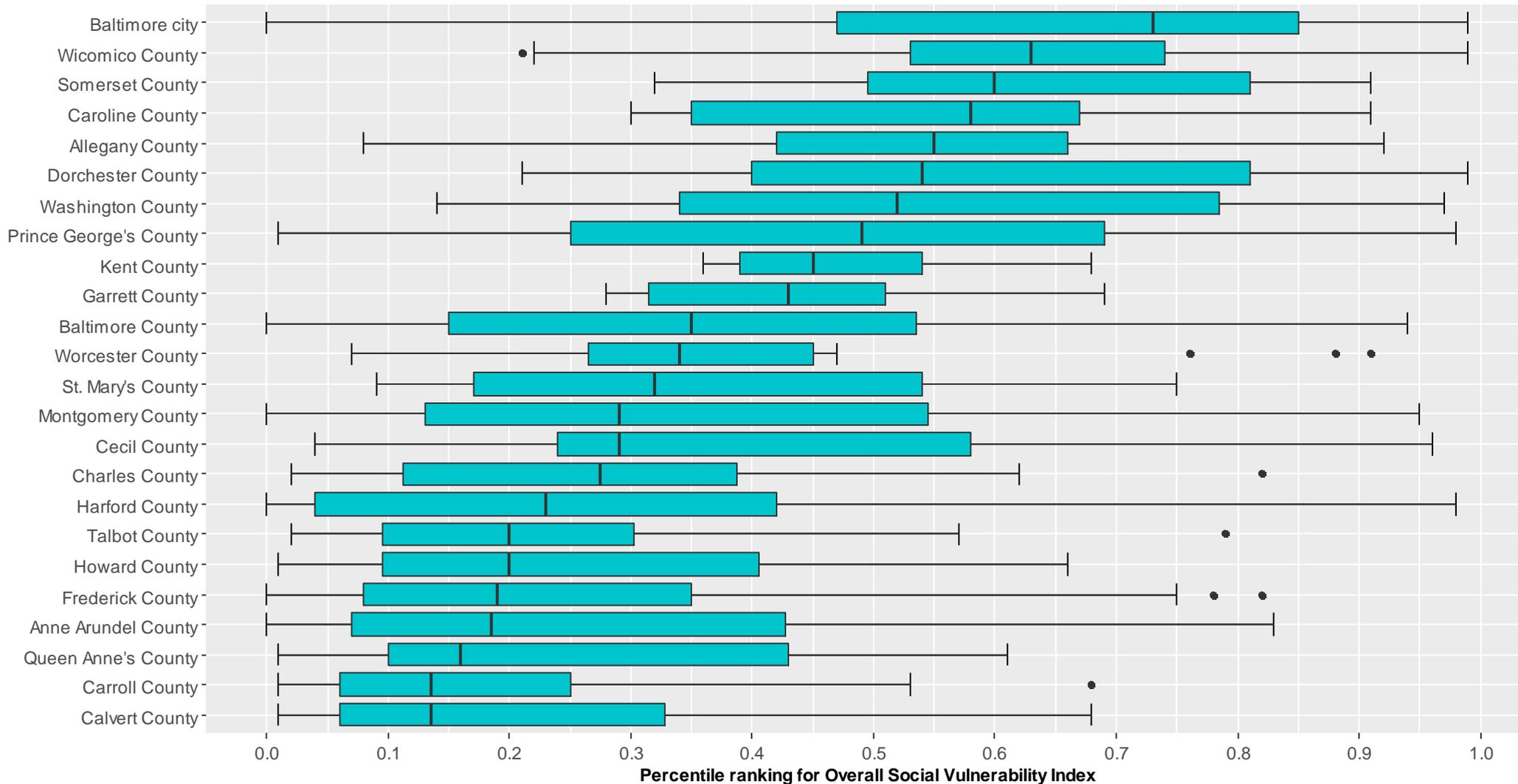
# SAIPE Poverty Rate and Broadband Use Across U.S. Counties in 2020



Source: AHRQ SDOH Database, version 1, from 2020 SAIPE and 2016-2020 American Community Survey 5-year files.

21

# Variation in Social Vulnerability Index Across and Within Maryland Counties, 2018



Distributions of tract-level Overall SVIs by Maryland county, 2018

Percentile ranking for Overall Social Vulnerability Index

# Linking AHRQ SDOH Database With Other Data Sources

| Data Type | Examples of Data Sources | Potential Linking Variables |
|---|---|---|
| Claims/Billing Data | Research Identifiable Files<br>• Medicare, Medicaid<br>Limited Data Sets<br>• All-Payer Claims Databases, Commercial (e.g., Blue Health Intelligence, MarketScan, Optum)<br>• AHRQ's HCUP | **ZIP Code** of patient or provider<br>**County** of patient or provider |
| Administrative Data | • MDS for Nursing Home and Swing Bed Providers<br>• OASIS for Home Health Providers | **ZIP Code** of provider<br>**County** of provider |
| Survey Data | • AHRQ's MEPS-Household Component Person-level Restricted Data<br>• Health Retirement Study Restricted Data<br>• National Health Interview Survey Restricted Data | **Census tract**<br>**ZIP Code**<br>**County** |
| Aggregated Data | Publicly Available Data<br>• CDC: Compressed Mortality File<br>• County-level data from registries/administrative data | **County** |
| Practice or Health System Data | • EHR data<br>• Population health management data | **Census tract**<br>**ZIP Code**<br>**County** |

# SDOH Database: Examples of Linked Data Questions

- Claims data *(Medicare claims)*
  - ▶ To what extent do changes in healthcare utilization for chronic illnesses during the COVID-19 pandemic differ by SDOH characteristics of patients' county/ ZIP Code of residence?

- Hospital billing data *(HCUP)*
  - ▶ Are differences in county and/or ZIP Code characteristics associated with differences in COVID-19 hospitalizations and outcomes?

- Individual-level survey data *(MEPS-HC)*
  - ▶ Do individuals living in areas with a shortage of primary care providers have difficulty finding a usual source of care?

- Healthcare system/EHR data *(individual practices or systems)*
  - ▶ Does knowing information on the community in which a patient lives help providers improve care?  If so, how?

HCUP is the Healthcare Cost and Utilization Project.
MEPS-HC is the Medical Expenditure Panel Survey-Household Component.

# Thank You

Feedback is welcome!

[sdoh@ahrq.hhs.gov](mailto:sdoh@ahrq.hhs.gov)