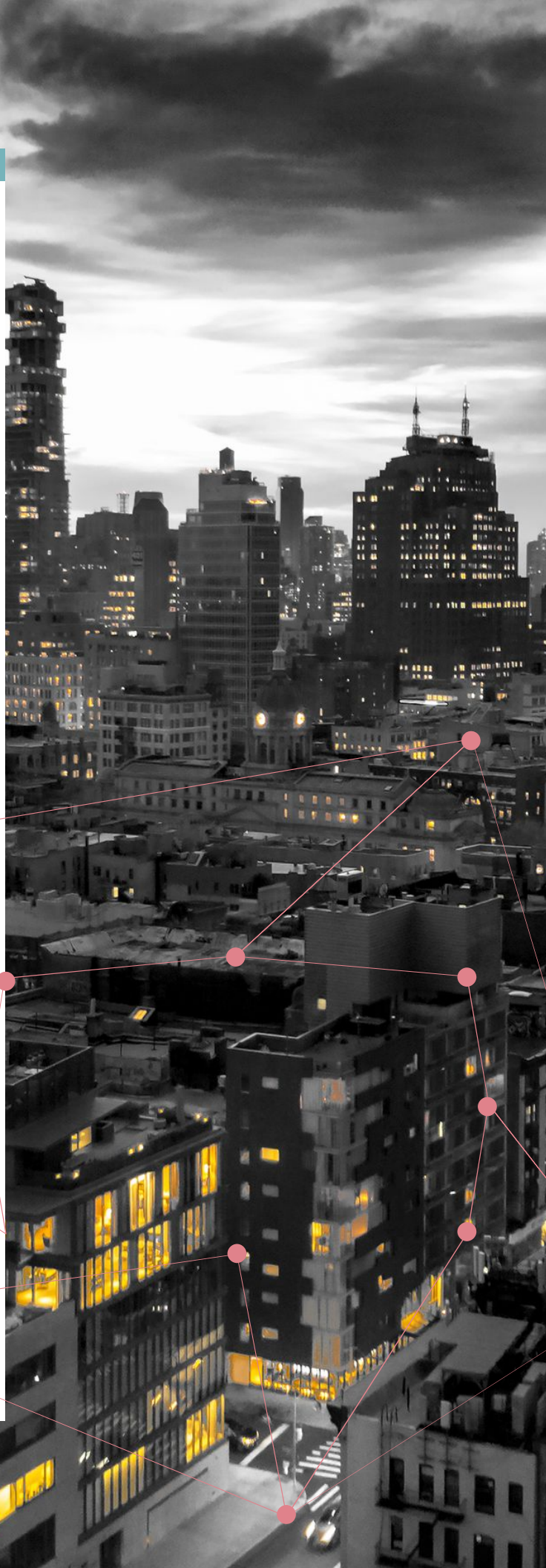
Community Stress Index Users' Guide

Prepared for NAF
by the NYU Marron Institute

September 24, 2021



NEW YORK UNIVERSITY

Contents

Contents continued

Contents continued

Figures

Introduction

The Community Stress Index aims to consolidate data on the communities (or "neighborhoods") around the ~23,000 public (including charter) high schools in the United States.

The Index reflects the level of "stress," in five domains, that school communities experience, to provide context when reviewing student learning outcomes and as a tool to help resource allocation for decisionmakers. The index comprises five *domains*: Economic, Education, Health, Housing, and Crime. A larger value of the measure (or "score") on a domain indicates more stressful community conditions.

The scores for the five community stress domains are weighted equally and are then aggregated into a composite index using an arithmetic mean.

Community stress= ⅕(Economic stress) + ⅕(Education stress) + ⅕(Health stress) + ⅕(Housing stress) + ⅕(Crime stress)

Each domain comprises several *indicators*, each with an associated *measure*. Measures are available at different geographic levels, obtained from various sources with different updating frequencies (and different most recent year, as many sources have several years lag in reporting) and with different geographic coverage (see How measures were selected, page 35; special attention needs to be paid to Puerto Rico, with limited coverage). See Table 1 (page 5) for characteristics of index measures.

Table 1. Characteristics of Index Measures.

Domain	#	Indicator	Measure	Source	Frequency	Level	Puerto Rico	Limited extent
Economic	1	Unemployment	% 16+ unemployed but seeking work	Bureau of Labor Statistics	Annually (last 2018)	County	Yes	
	2	Poverty	% 0–17 in poverty	U.S. Census Bureau	Annually	Census tract	Yes	
	3	Access to broadband internet	% households with subscription to any broadband service	U.S. Census Bureau	Annually	Census tract	Yes	
	4	Single-parent household	% 0–17 living with one parent	County Health Rankings	Annually	County	No	
Education	5	High-school graduates	% 25+ high-school grad or equivalent	U.S. Census Bureau	Annually	Census tract	Yes	
	6	Linguistic isolation	% limited-English-speaking households	U.S. Census Bureau	Annually	Census Tract	Yes	
Health	7	Access to healthcare	% 6–18 with health-insurance coverage	U.S. Census Bureau	Annually	Census tract	Yes	
	8	Infant mortality rate	Infant deaths / 1,000 live births	County Health Rankings	Annually	County	No	
	9	SNAP recipients	% households with children under 18 receiving SNAP	U.S. Census Bureau	Annually	Census tract	Yes	
	10	Low birth weight	% live births <2.5 kg	County Health Rankings	Annually	County	No	
	11	Lead exposure risk	Lead-exposure risk index	City Health Dashboard	Annually	Census tract	No	Yes*
Housing	12	Housing-vacancy rate	% housing units vacant	U.S. Census Bureau	Annually	Census tract	Yes	
	13	Housing affordability	% households spend ≥30% income on housing	U.S. Census Bureau	Annually	Census tract	Yes	
	14	Park access	% within 10-minute walk of green space	City Health Dashboard	Static	Census tract	No	Yes*
Crime	15	Violent crime	Violent crimes / 100,000	County Health Rankings	Annually	County	No	
	16	Incarceration rate	Jail incarceration / 100,000, 15–64	Vera Institute	Annually	County	No	

*In 2020, the City Health Dashboard expanded from covering the 500 largest cities to over 750 cities (all cities over 50,000). We currently include the 500 cities and are working to include the 750+. More details in cityhealthdashboard.com/technical-documentation.

Domains, Indicators, and Measures

Economic

The economic domain is assessed by unemployment, poverty, access to broadband internet, and single-parent household. Values of these four indicators are weighted equally in deriving the domain score

Unemployment

Description

% 16+ unemployed but seeking work

Unemployment is associated with low income, poor healthcare access, and psychological stress. It has many negative effects on physical and mental health. In general, those who are unemployed report more days of poor mental or physical health than those who are employed. Unemployment can lead to reduced self-worth manifesting as depression, anxiety, poor self-esteem, and unhealthy coping behaviors, such as alcohol and substance use. A measure, percent unemployment rate for adults in the labor force, is available from the Bureau of Labor Statistics. Data are updated annually and are available nationally at the county level; most recent data are for 2019.[1]

Frequency Annually
Level County
Data source
U.S. Bureau of Labor Statistics:
bls.gov/lau/#tables
Details on how unemployment is measured at:
bls.gov/cps/eetech_methods.pdf
Additional calculations
Raw data are as percentages and are converted to rates.
Missingness geographic extent
No counties (and no schools within our sample) in the U.S. are missing values of this measure. Includes data on Puerto Rico.

[1] In some instances, data are updated annually but there are reporting lags of one or more years. County Health Rankings and City Health Dashboard use tertiary data sources and, in some cases, querying for the latest year available does not guarantee that measures will be for those years. For example, County Health Rankings published 2021 data on violent crime using FBI data from 2014 and 2016.

Poverty

Description

% children age 0–17 in poverty

Childhood poverty is associated with poor access to services, psychological stress, and poor school performance. A measure, percent children 0–17 in poverty, is available from USDA. Data are updated annually and are available nationally, at the tract, ZCTA, and school-district levels (among other geographies); most recent data are for 2018.

Frequency Annually

Level Census tract

Data source

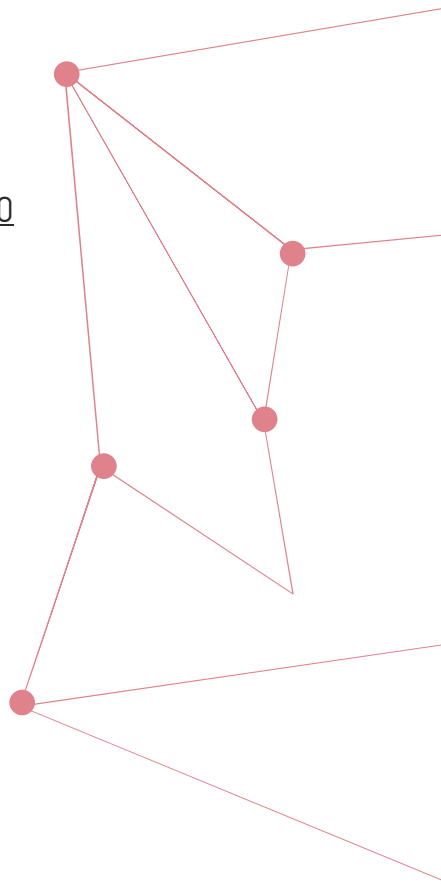
American Community Survey 5-year estimates (ACS-5):
data.census.gov/cedsci/table?q=B17020_004E,B17020_005E,B17020_002E&tid=ACSDT1Y2019.B17020
B17020_004E: 6 to 11 years
B17020_005E: 12 to 17 years
B17020_002E: Total

Additional calculations

Poverty = (B17020_004E + B17020_005E) / B17020_002E

Missingness geographic extent

1.4% of census tracts are missing this measure (2019). No schools are missing this measure.[2] Includes data on Puerto Rico.



[2] See Notes on missingness, page 28.

Access to broadband internet

Description

% households with subscription to any broadband service

Access to broadband internet is increasingly important to taking advantage of education technology, for social connection, and for employment in growth industries. A measure, percent households with subscription to any broadband service, is available from the U.S. Census Bureau, American Community Survey. Data are updated annually and are available nationally, at the tract, ZCTA, and school-district levels, (among other geographies); most recent data are for 2019.

Frequency Annually

Level Census tract

Data source

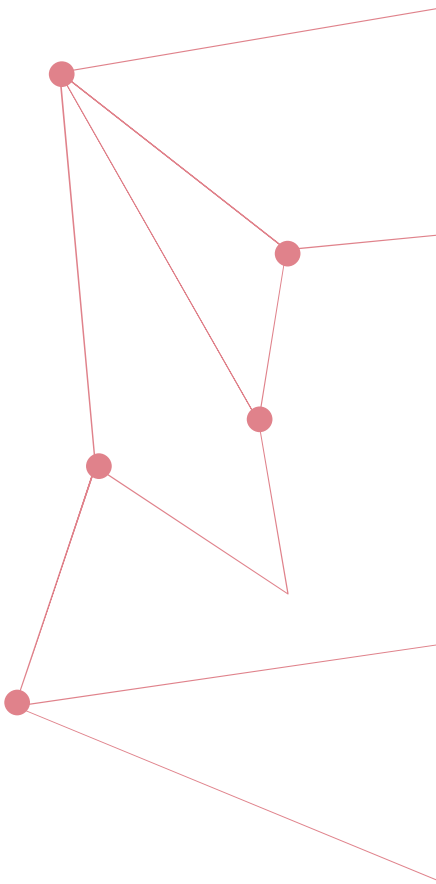
American Community Survey 5-year estimates (ACS-5):
api.census.gov/data/2019/acs/acs5/groups/B28002.html
B28002_004E: Total households with an internet subscription (broadband of any kind)
B28002_001E: Total number of households

Additional calculations

Access to broadband internet = B28002_004E / B28002_001E

Missingness geographic extent

1.2% of census tracts are missing this measure (2019). No schools are missing this measure. Includes data on Puerto Rico.



Single-parent household

Description

% 0–17 living with one parent

Raising a child or children as a single parent can be stress-inducing, without another parent to balance work and home life. Single-parent households may also be associated with other causes or indicators of stress. Data are updated annually and are available nationally at the county level; most recent data are for 2019.

Frequency Annually

Level County

Data source

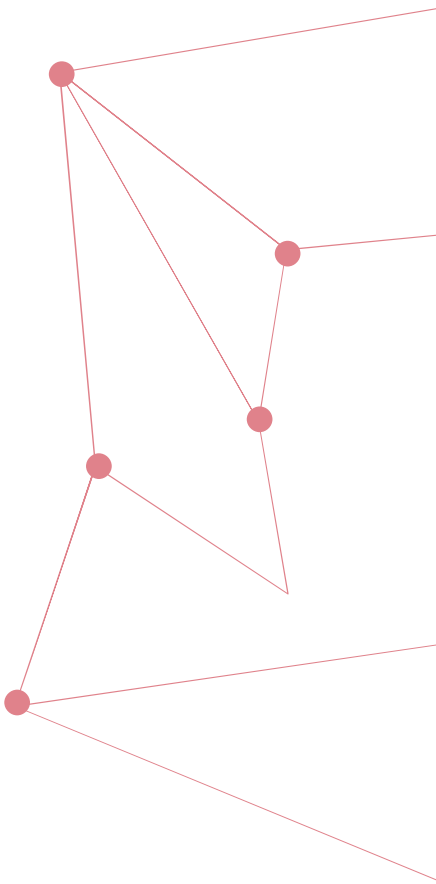
County Health Rankings: countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/social-and-economic-factors/family-social-support/children-in-single-parent-households

Additional calculations

Raw data are as percentages and are converted to rates.

Missingness geographic extent

2.5% of counties are missing this measure (2019). 0.8% of schools are missing this measure. Does not include data on Puerto Rico.



Education

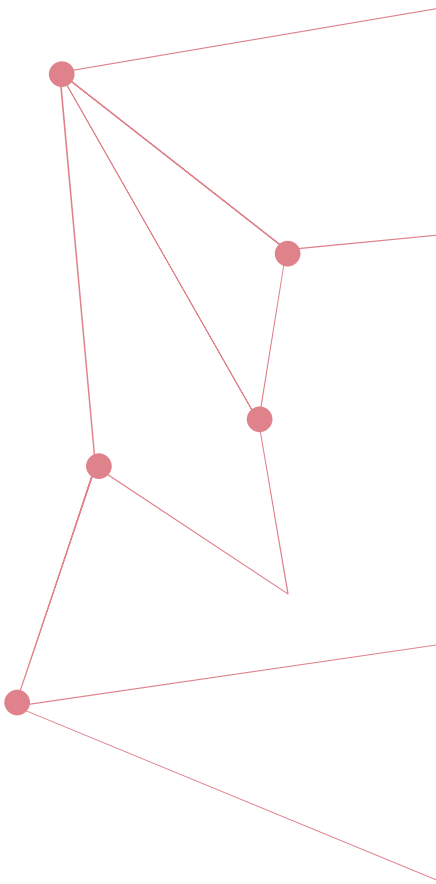
The education domain is assessed by high school graduation and linguistic isolation. Values of these two indicators are weighted equally in deriving the domain score.

High-school graduates

Description

% 25+ high-school graduate or equivalent

Adults having a high-school diploma or equivalent are likely better able to provide for children and serve as education role models. A measure, percent population 25+ high-school graduate (or equivalent), is available from the U.S. Census Bureau, American Community Survey. Data are updated annually and are available nationally, at the tract, ZCTA, and school-district levels (among other geographies); most recent data are for 2019.



Frequency Annually

Level Census tract

Data source

American Community Survey 5-year estimates (ACS-5):
api.census.gov/data/2019/acs/acs5/groups/B15002.html
B15002_011E: Male high-school graduates (includes equivalency) 25 and over
B15002_028E: Female high-school graduates (includes equivalency) 25 and over
B15002_001E: Total population 25 and over

Additional calculations

High-school graduates = (B15002_011E + B15002_028E)/(B15002_001E)

Missingness geographic extent

1.0% of census tracts are missing this measure (2019). No schools are missing this measure. Includes data on Puerto Rico.

Linguistic isolation

Description

% limited English-speaking households

Difficulty communicating in or understanding English limits opportunities in most communities in the U.S. A measure, percent limited English-speaking households, is available from the U.S. Census Bureau, American Community Survey. Data are updated annually and are available nationally, at the tract, ZCTA, and school-district levels, (among other geopgraphies); most recent data are for 2019.

Frequency Annually

Level Census tract

Data source

American Community Survey 5-year estimates (ACS-5):
api.census.gov/data/2019/acs/acs5/subject/groups/S1602.html
S1602_C03_001E: Limited English-speaking households
S1602_C01_001E: All households

Additional calculations

Linguistic isolation = S1602_C03_001E / S1602_C01_001E

Missingness geographic extent

1.2% of census tracts are missing this measure (2019). No schools are missing this measure. Includes data on Puerto Rico.

Notes

A "limited English-speaking household" is one in which no member 14 years old and over (1) speaks only English or (2) speaks a non-English language and speaks English "very well." In other words, all members 14 years old and over have at least some difficulty with English. By definition, English-only households cannot belong to this group. Previous Census Bureau data products referred to these households as "linguistically isolated" and "Household where no one age 14 and over speaks English only or speaks English 'very well.'"

See more in: census.gov/topics/population/language-use/about/faqs.html

Health

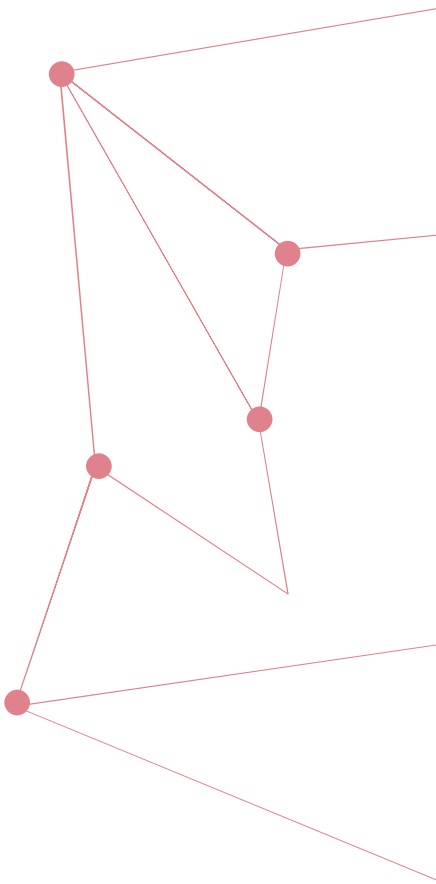
The health domain is assessed by access to healthcare, infant mortality, SNAP recipients, low birthweight, and lead exposure risk. Values of these five indicators are weighted equally in deriving the domain score.

Access to healthcare

Description

% 6–18 with health-insurance coverage

Access to healthcare is important to maintaining good health, and poor access is stressful to a community. Measuring access is difficult, as low healthcare usage could indicate good health and so low demand, or poor health and low supply or inability to pay. A measure, percent children 6–18 with health-insurance coverage, is available from the U.S. Census Bureau, American Community Survey. Data are updated annually and are available nationally, at the tract, ZCTA, and school-district levels, (among other geographies); most recent data are for 2019. And a similar measure, percent children 6–18 with public health-insurance coverage, is also available.



Frequency Annually

Level Census tract

Data source

American Community Survey 5-year estimates (ACS-5):
api.census.gov/data/2019/acs/acs1/subject/groups/S2701.html
S2701_C03_003E: Percent of civil noninstitutionalized population 6–18 that has health insurance.

Missingness geographic extent

1.4% of census tracts are missing this measure (2019). No schools are missing this measure, Includes data on Puerto Rico.

Infant mortality rate

Description

Number of all infant deaths (within 1 year) per 1,000 live births

Morbidity and early mortality are strong indicators of stress. An especially telling indicator is infant mortality, which is rare but notably high in especially stressed communities. A measure, infant deaths (within 1 year) per 1,000 live births, is available from County Health Rankings, derived from data from the National Center for Health Statistics, National Vital Statistics System. Data are updated annually and are available nationally, at the county level; most recent data are for 2019.

From County Health Rankings: "Infant mortality represents the health of the most vulnerable age group and can help with interpreting the years of potential life lost (YPLL) rate in a county. Infant mortality is also commonly used to examine global health differences, as well as to understand historic racial inequities in the U.S."

Frequency Annually

Level County

Data source

County Health Rankings
countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-outcomes/length-of-life/infant-mortality

Missingness geographic extent

60.9% of counties are missing this measure (2019). 22.9% of schools are missing this measure. Most of the missingness is likely attributable to low infant-death rates. See note below. No data on Puerto Rico.

Notes

"Infant death is a relatively rare event in most counties."

"A missing value is reported for counties with fewer than 20 infant deaths in the time frame."

SNAP recipients

Description

% households with children under 18 receiving SNAP

SNAP is an important program for supporting children in low-income households, to ensure that they receive adequate nutrition, critical to good health; a high incidence of SNAP recipients indicates challenges to nutrition in the community. A measure, percent households with children under 18 receiving SNAP, is available from the U.S. Census Bureau, American Community Survey. Data are updated annually and are available nationally, at the tract, ZCTA, and school-district levels, (among other geographies); most recent data are for 2019.

Frequency Annually

Level Census tract

Data source

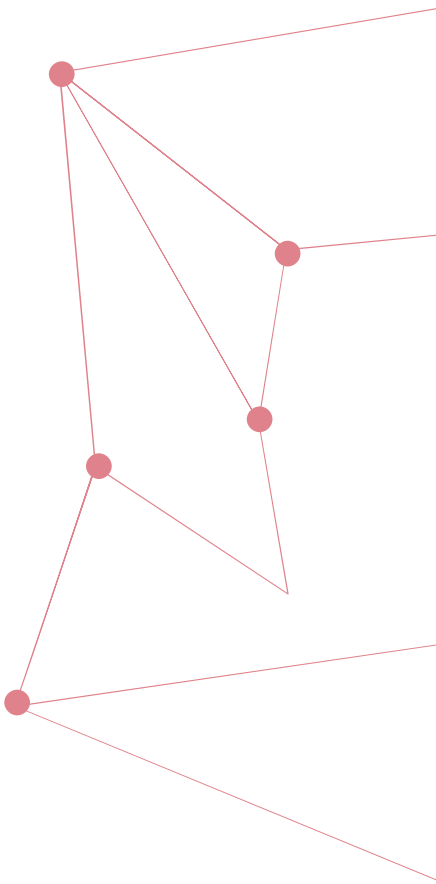
American Community Survey 5-year estimates (ACS-5): api.census.gov/data/2019/acs/acs5/subject/groups/S2201.html S2201_C04_009E: Estimate. Percent households receiving food stamps/SNAP. Households with children under 18 years.

Additional calculations

Raw data are as percentages and are converted to rates.

Missingness geographic extent

3.6% of census tracts are missing this measure (2019). 0.1% of schools are missing this measure. Includes data on Puerto Rico.



Low birth weight

Description

% live births <2.5 kg

Babies born with low birth weight are more likely to experience long-term physical-health problems and cognitive-development challenges later in life. Low birth weight is associated with poor healthcare access and maternal behavioral-health problems. There are significant demographic differences in low birth-weight prevalence, with babies born to Black mothers more likely to be low birth weight compared to babies born to White, non-Hispanic mothers.

A measure, percent of live births with birth weight < 2.5 kg, is available from the National Center for Health Statistics, and we acquire it via County Health Rankings. Data are available at the county level and are updated annually. Most recent data are for 2019.

Frequency Annually

Level County

Data source

countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-outcomes/quality-of-life/low-birthweight

Additional calculations

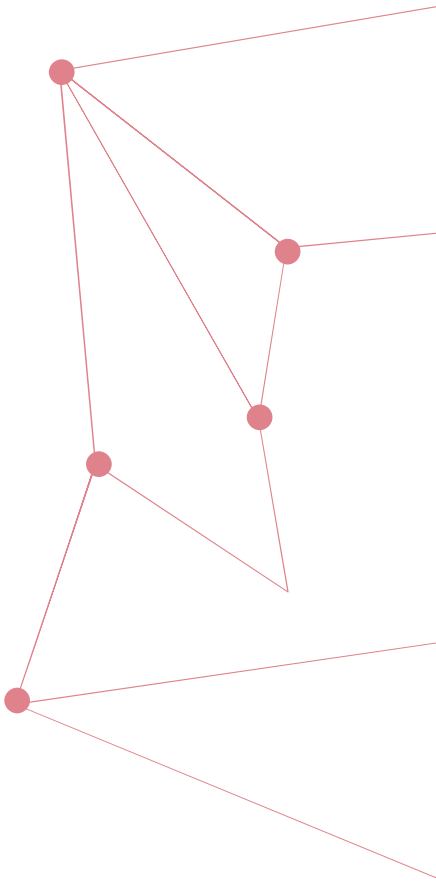
Raw data are as percentages and are converted to rates.

Missingness geographic extent

5.7% of census tracts are missing this measure (2019). 1.3% of schools are missing this measure. No data on Puerto Rico.

Notes

From County Health Rankings: "Births are counted in the county corresponding to the mother's address on the child's birth certificate, not the county the child was born in."

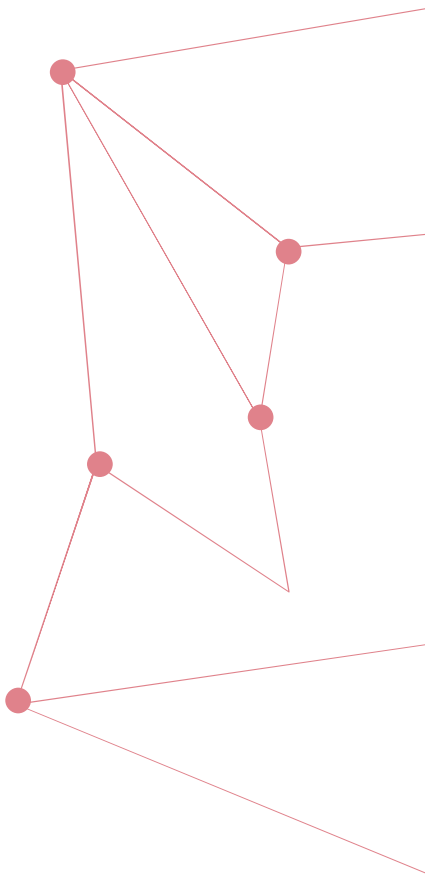


Lead exposure risk

Description

Lead-exposure-risk index (see notes below)

Exposure to lead can seriously damage a child's brain and nervous system, impair growth and development, and yield learning or behavioral problems, and hearing/speech damage. Living in high-lead-risk census tracts is associated with greater harms for children from lower- versus higher-income families—lower cognitive test scores, smaller cortical surface area, and smaller cortical volume. Data are updated annually and are available nationally at the census-tract and county levels; most recent data are for 2019.



Frequency Annually

Level Census tract

Data source

City Health Dashboard:
cityhealthdashboard.com/measure/48
More at cityhealthdashboard.com/technical-documentation

Missingness geographic extent

This measure is available only for the five hundred most populous cities in the United States. 51.5% of census tracts are missing this measure (2018). 53.8% of schools are missing this measure. No data on Puerto Rico.

Notes

From the data source:
Housing with potential lead risk
We count the number of housing units in each of five time periods: pre-1939, 1940–59, 1960–79, 1980–99, and 2000 or newer. The count of housing units in each time period is weighted by the likelihood of lead exposure in housing of that era, which results in an overall percent of area housing likely to have some risk of lead exposure.
Lead exposure-risk index
We took the housing with potential lead risk and factored in information about the percentage of households living at or below 125% of the poverty level. We standardized, weighted, summed, and ranked these values from 1, or lowest risk, to 10, or highest risk, to create a scale of overall lead exposure risk.

Housing

The housing domain is assessed by vacancy rate, housing affordability, and park access. Values of these three indicators are weighted equally in deriving the domain score.

Housing–vacancy rate

Description

% housing units vacant

Vacant housing may be an indicator of neighborhood stress—declining population, high foreclosures, or overbuilding preceding an economic crash. A measure, percent housing units vacant, is available from the U.S. Census Bureau, American Community Survey. Data are updated annually and are available nationally, at the tract, ZCTA, and school-district levels, (among other geographies); most recent data are for 2019.

Frequency Annually

Level Census tract

Data source

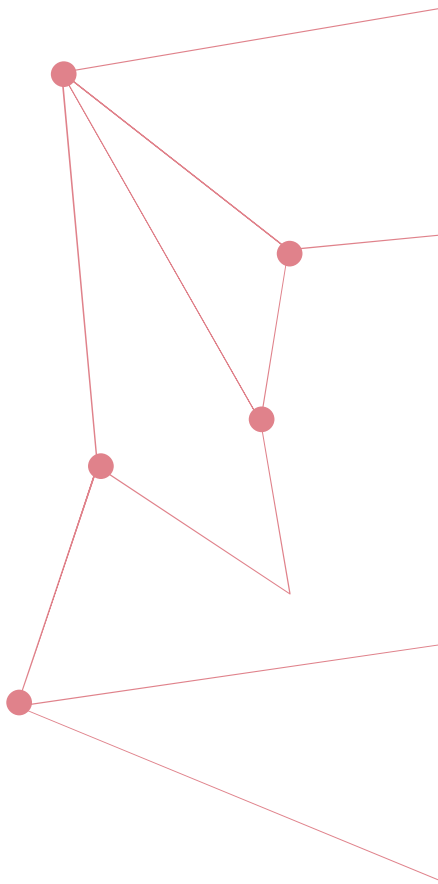
American Community Survey 5-year estimates (ACS-5):
api.census.gov/data/2019/acs/acs1/profile/groups/DP04.html
DP04_0003E: vacant housing units
DP04_0001E: total housing units

Additional calculations

DP04_0003E / DP04_0001E

Missingness geographic extent

1.2% of census tracts are missing this measure (2019). No schools are missing this measure. Includes data on Puerto Rico.



Housing affordability

Description

Share of households spending more than 30% of income on housing.

While high absolute housing costs often reflect wealthy, low-stress neighborhoods, lower-income neighborhoods in high-cost regions can be beset with high housing-cost burdens. A measure, percent households that spend 30% or more of their income on housing, can be derived from data from the U.S. Census Bureau, American Community Survey. Data are updated annually and are available nationally, at the tract, ZCTA, and school-district levels, (among other geographies); most recent data are for 2019.

Frequency Annually

Level Census tract

Data source

American Community Survey 5-year estimates (ACS-5):
api.census.gov/data/2019/acs/acs5/subject/groups/S2503.html

The number of housing units by household income bracket is available. The percentage of housing units that spend 30% or more of household income in housing is also available by income bracket. Using these variables, it is possible to calculate the share of households spending 30% or more of income on housing.

$$\frac{\sum_{i\ income\ brackets} x_i \cdot y_i}{n}$$

Where x_i is the number of occupied households in bracket i, y_i is the percentage of occupied households in bracket i that spend 30 percent or more of their income on housing, and n is the total number of occupied units.

Since the predefined brackets in x and y are not equal (see variables below), we grouped them so they are comparable.

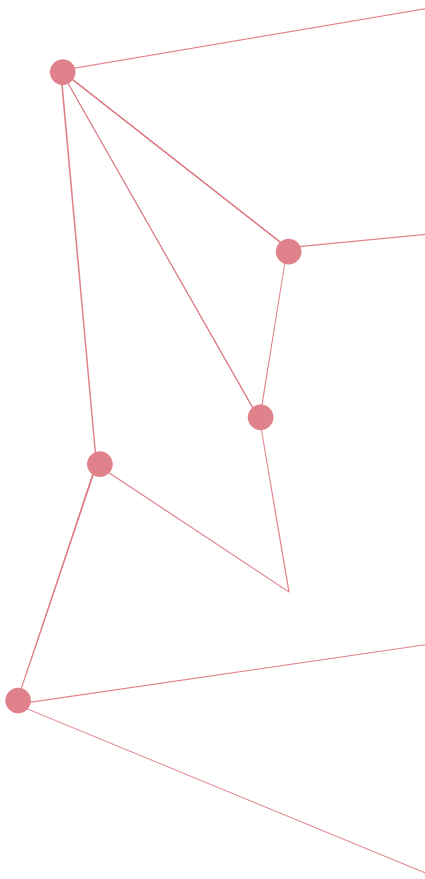
Additional calculations

Owner-occupied housing units[3]

((002E+003E+004E+005E).028E+007E.032E+(008E.036E)+009E.040E+(010E+011E+012E)*044E)/(001E.001E)

[3] Variable names are shortened—all have the S2503_C01_ prefix.

Where:
S2503_C01_001E: Occupied housing units (estimated).
S2503_C02_002E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) less than $5,000.
S2503_C02_003E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $5,000 to $9,999.
S2503_C02_004E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $10,000 to $14,999.
S2503_C02_005E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $15,000 to $19,999.
S2503_C02_028E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income of less than $20,000.
S2503_C02_006E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $20,000 to $24,999.
S2503_C02_007E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $25,000 to $34,999.
S2503_C02_032E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income $20,000 to $34,999.
S2503_C02_008E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $35,000 to $49,999.
S2503_C02_036E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income $20,000 to $49,999.
S2503_C02_009E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $50,000 to $74,999.
S2503_C02_040E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income $50,000 to $74,999.
S2503_C02_010E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) 75,000 to $99,999.
S2503_C02_011E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) 100,000 to $149,999.
S2503_C02_012E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) 150,000 or more.
S2503_C02_044E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income $75,000 or more.



Missingness geographic extent

1.2% of census tracts are missing this measure (2019). No schools are missing this measure. Includes data on Puerto Rico.

Park access

Description

% population living within a 10-minute walk of green space

Public parks are critical resources for physical activity in underserved communities. Even mental health is strongly associated with residential distance from parks. Parks can also have a positive impact on environmental outcomes like air pollution. A measure, percent population living within a 10-minute walk of a green space, from ParkServe, was published in 2018, available nationally at the tract level. These data are also available to us through a data repository maintained by NYU Population Health.

From City Health Dashboard: "Access to parks has been associated with better mental and physical health, lower death rates, and improved social connection. While a growing body of research is making the evidence base linking park access and health stronger, the reasons for this link are not yet fully clear. Research has shown that green space may be associated with increased physical activity, reduced stress, improved quality of life, and better mental health across different groups of people. Also, parks and green space may be associated with increased neighborhood social ties by serving as meeting places for local residents and spaces for community events. Parks may also have positive impacts on environmental outcomes such as reducing air pollution and extreme heat, especially in urban areas."

Estimates for this measure are from 2017 ParkServe data. The City Health Dashboard sources ParkServe data from 2018, ParkServe is using demographic data from 2020 and geographic data from 2010, likely because that information changes less frequently. We do not find the update frequency for demographic and geographic data on the ParkServe website.

Frequency unknown
Level Census tract
Data source
City Health Dashboard cityhealthdashboard.com/measure/51
More on cityhealthdashboard.com/technical-documentation
and tpl.org/parkserve/about
Missingness geographic extent
This measure is available only for the five hundred most populous cities in the United States. 51.2% of census tracts are missing this measure. 53.8% of schools are missing this measure. No data on Puerto Rico.

Crime

The crime domain is assessed by violent crime and the incarceration rate. Values of these two indicators are weighted equally in deriving the domain score.

Violent crime

Description

Number of reported violent-crime offenses per 100,000 population

Violent crime exacts a toll on victims and the community.[4] Crime data are notoriously fickle, as jurisdictions vary in their definitions of crime and reporting by local law-enforcement agencies to the FBI and BJS is unreliable. Murder is the most difficult crime to sweep under the rug; murder, while rare, is generally correlated with other violent crime in larger jurisdictions and so is a good proxy. In smaller jurisdictions, which may have zero or one murders a year, it is a very noisy indicator. Also, while actual violent crime rates and public perceptions can change very quickly, reporting and data collection are slow so that "uniform," national data are typically reported with a lag.

A measure, number of reported violent crime offenses per 100,000 population, is available from County Health Rankings, sourced from the FBI Uniform Crime Reporting Program. Data are updated annually and are available nationally at the county level.

Frequency Annually (lagged)
Level County
Data source
American Community Survey 5-year estimates (ACS-5):
County Health Rankings:
countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/social-and-economic-factors/community-safety/violent-crime-rate
Additional calculations
All observations with a violent crime rate value higher than the 95th percentile of the variable were assigned that value instead.
Missingness geographic extent
8.4% of counties are missing this measure (2019). 3.0% of schools are missing this measure. No data on Puerto Rico.

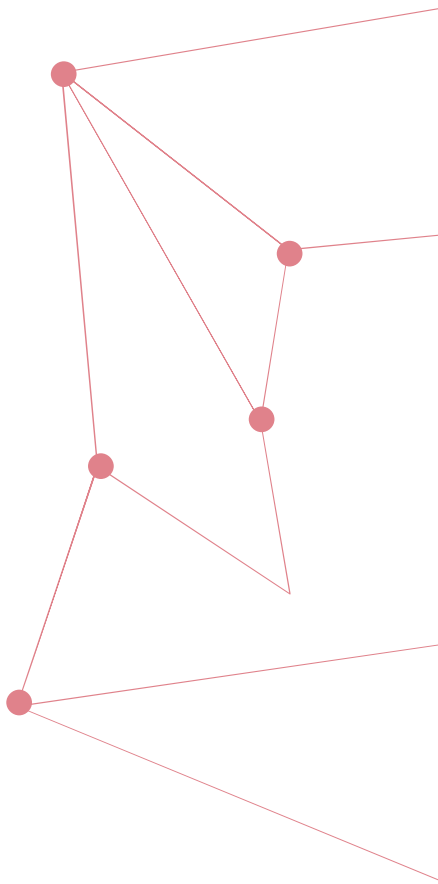
[4] Violent crimes are defined as offenses that involve face-to-face confrontation between a victim and a perpetrator, including homicide, rape, robbery, and aggravated assault.

Incarceration rate

Description

Jail incarceration / 100,000, 15–64

Incarceration imposes stress on the community that incarcerated people leave behind. Prison population figures are generally complete and timely, but there is no uniform collection of data on where prison inmates are from. Jails mostly house people who are from the jurisdiction that they are in (typically counties), but jail-population data are slow to be collected and reported—the most recent census was in 2013. A measure, jail incarceration rate per 100,000 residents 15–64, is available from the Vera Institute of Justice. This is a new mechanism; data will be updated periodically and are available nationally at the county level. Most recent data are for 2018.



Frequency Annually

Level County

Data source

Vera Institute of Justice:
github.com/vera-institute/incarceration-trends

Missingness geographic extent

8.0% of counties are missing this measure (2018). 5.4% of schools are missing this measure. No data on Puerto Rico.

Index Calculation

Notes on pre-processing

To construct a composite index, all measures must be expressed in similar fashions ("commensurate"). As the raw data associated with the measures for this index are expressed in a variety of fashions (percentages, rates, etc.), they must be transformed ("pre-processed") to be expressed similarly —as numbers from 0 to 100.

The raw data associated with the measures selected for the index do not all have the same sign (or "directionality"); i.e., "more" of one measure may indicate better conditions in the community, whereas more of another may indicate worse conditions. For example, more unemployment in a given census tract indicates higher community stress, whereas more access to health-insurance coverage indicates lower community stress. Measures for which the raw data have larger values associated with higher stress are transformed by reversing the sign.

Different measures have different ranges of actual values (on the 0–100 scale). To capture the full extent of variation by community, for each measure, some measures are truncated, so that all values above or below selected "cut points" are treated as, respectively, being the minimum or maximum value of that measure. For example, when the poverty measure is greater than 40%, the raw data are truncated to 40%, which becomes a score of 100.

Table 2 summarizes the pre-processing for each measure.

Table 2. Measures Pre-Processing.

Domain	#	Indicator	Measure	Original Scale	Scaled range	Updated cuts
Economic	1	Unemployment	% 16+ unemployed but seeking work	%	–	
	2	Poverty	% 0–17 in poverty	%	–	>0.4 → 0.4
	3	Access to broadband internet	% households with subscription to any broadband service	%	+	<0.4 → 0.4
	4	Single-parent household	% under 6 or 6–17 living with one parent	%	–	
Education	5	High-school graduates	% 25+ high-school grad or equivalent	%	+	
	6	Linguistic isolation	% limited English-speaking households	%	–	>0.25 → 0.25
Health	7	Access to healthcare	% 6–18 with health-insurance coverage	%	+	
	8	Infant mortality rate	Infant deaths / 1,000 live births	>=0	–	
	9	SNAP recipients	% households with children under 18 receiving SNAP	%	–	
	10	Low birth weight	% live births <2.5 kg	%	–	
	11	Lead exposure risk	Lead-exposure risk index	1-10	–	
Housing	12	Housing-vacancy rate	% housing units vacant	%	–	>0.4 → 0.4
	13	Housing affordability	% households spend ≥50% income on housing	%	–	
	14	Park access	% within 10-minute walk of green space	%	+	
Crime	15	Violent crime	Violent crimes / 100,000	>=0	–	95th percentile
	16	Incarceration rate	Jail incarceration / 100,000, 15–64	>=0	–	95th percentile

Data assembly: School Attendance Boundary (SAB)

The intent of the index is to characterize the challenges faced by public high schools and their student bodies, in order to leverage resources and put education outcomes data into context. The natural unit of analysis for the index is a public high school. For this purpose, ideally, we would have data reflecting each measure on the well-defined geographic area from which the school's students are drawn. Many schools have such an area, which the National Center for Education Statistics, NCES, defines as a School Attendance Boundary (SAB);[5] some schools, such as magnet schools or districts with open enrollment (as in New York City), draw from an entire school district. And some schools do not have an NCES SAB, for reasons unspecified.

Unfortunately, none of the measures for this index are collected at the SAB level. So, we synthesize measures at the SAB level, from data collected at other levels. The smallest geographic units for which there are publicly available data on the measures are census tract or county, depending on the measure (see Table 1, page 5). We conduct an "assembly process," where data at the census-tract or county level are assigned to schools associated with the geographic units and using SABs, when possible.

[5] An SAB has an associated shapefile, which defines the geographic boundaries.

The first step in the assembly process is the creation of two data files (in .csv format) that contain all measures collected at the county level and at the census-tract level, separately. This is done by downloading or collecting information from the data sources detailed above, using API calls or https requests and attaching the information to reference tables on the universe of counties and census tracts (see Data sources, below).

Once these data are available, measure values are assigned to each school depending on the following logic, which applies for measures that are available at the census-tract level:

1. If an SAB is available (n~12,1970[6]), then all intersecting census tracts[7] are selected, and their measure values population-weighted and averaged.
2. If an SAB is not available (n~9,928), the school's location (geographic coordinates) is used to determine the census tract where it is located. ZCTAs are reasonable proxies for school "neighborhoods," and correspondence tables between ZCTAs and census tracts are available.[8] Measures from the census tracts that intersect the school's zip code are population weighted and averaged.
3. 323 out of 461 schools in New York City have the same SAB, the geographic extent of the city, due to the school district's open-enrollment policy. The strategy in #2 is applied in this case, as well as for four schools with open enrollment in the Los Angeles Unified School District, with the purpose of making the index values more localized and thus relevant. This strategy was applied to 327 high schools in total.

For measures available at the county level, all schools in the county have the same value.

[6] We started with 23,225 high schools from a 2018–2019 NCES data set (A), along with a 2015–2016 data set consisting of 14,172 high schools with a column specifying their SABs and corresponding geographies (B). When (A) and (B) were merged, 1,202 schools were lost, so we were left with 12,970 high schools with SAB information that we can use. This mismatch may be due to the difference in publication dates for the data sets, but there may be other reasons for this discrepancy that are unknown to us.
[7] Census tracts and counties also have shapefiles, so it can be determined which census tracts and counties intersect with which SABs. Almost all SABs lie entirely within one county and intersect with multiple census tracts.
[8] See census.gov/geographies/reference-files/2010/geo/relationship-files.html#par_textimage_674173622.

With all schools having values assigned for each measure and each domain, the composite index values can now be calculated:

$$\text{Index} = \sum_{i \text{ in domains}} \sum_{j \text{ in metrics in domain } i} d_i m_{ij} x_{ij},$$

where d_i is the weight of domain i relative to the value of the index; m_{ij} is the weight of measure j in domain i, relative to the domain; and x_{ij} is the standardized value of measure i in domain j.

In simpler terms, for each school, we create a score for each domain, which is a weighted average of the measures in that domain. If a domain has, say, three indicators, x, y, and z, each with a possible range of 0–100, then the score for that domain is $a*x + b*y + c*z$, where the weights a, b, and c add up to one. The weights indicate how much relative importance we put on each of the measures—they are subjective choices. For simplicity, we make each of the weights in a domain equal (in the example above, the weights a, b, and c are all 1/3).[9]

The index is then a weighted sum of the five domain scores. Again, we make the weights equal, so each is 1/5.

[9] The weights are adjustable in the developer tool.

Notes on missingness

If a measure in a domain is missing for a school, the weights of all other measures in the domain are re-calculated so that their sum is 1.0, proportionally to the original intended weights. For instance, if four measures in a domain had weights [0.5, 0.15, 0.2, 0.15] and the first measure is missing, then the remaining measures, which sum to 0.5, are adjusted to [0.0, 0.3, 04, 0.3].

If all measures within a domain are missing, then the entire domain is marked as N/A. Similarly, if one of the domains is missing, then the weights of the remaining domains are adjusted proportionally to their original weights.

For measures that are available at the census tract and schools with known SAB, a school will be marked as missing such measure if and only if all of the census tracts intersecting its SAB are missing it. For measures available at the county level, if a measure is missing in a county, all schools within such county will be marked as missing it. For measures available at the census tract level and schools with no SAB, if a measure is missing in a census tract, all schools within such census tract will be marked as missing it.

Data Sources

Three kinds of information were needed to build the index:

- Information about schools, since schools are the unit of analysis of the index.
  - School-attendance boundaries (SABs) for 14,172 high schools. See nces.ed.gov/programs/edge/sabs. Data are from the 2015–2016 NCES School Attendance Boundary Survey (SABS).
  - List of high schools for the 2018–2019 school year from the NCES Common Core of Data, CCD. See nces.ed.gov/ccd/files.asp. Includes information on each school, such as name, geographic coordinates, zip code, county, city, and state, for 23,225 high schools.[10] This is our universe of schools.
- Information about geographic units.
  - Census-tract populations were gathered to implement the weighting scheme in strategy #2 when needed. This information comes from the U.S. Census Bureau.
  - Geographic boundaries of census tracts. They are used to determine which schools belong to which census tracts, which is important for strategy #2.
  - 2010 ZCTA to Census Tract Relationship File Layout. This list matches ZCTAs with all census tracts where there is a non-null geographic intersection. This is needed for when strategy #2 in the assembly process is implemented.
  - Total county populations were added to the data set for consistency.
- Source data for measures: See below

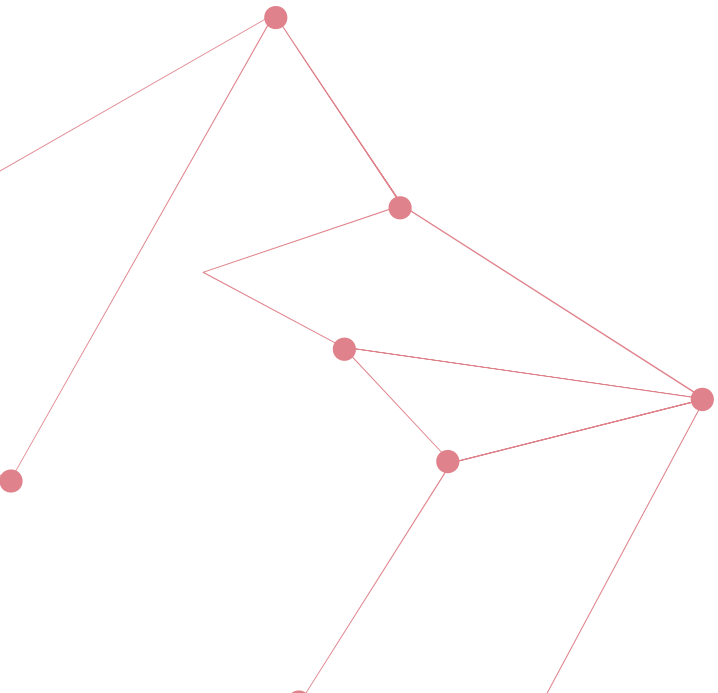
[10] Excluding schools from Guam, American Samoa, and the Virgin Islands, and excluding all schools that are not high schools. We count 23,243 schools before excluding these schools.

Additionally, NAF provided the list of schools they partner with. It was added to the database and is an optional column in the online index display.

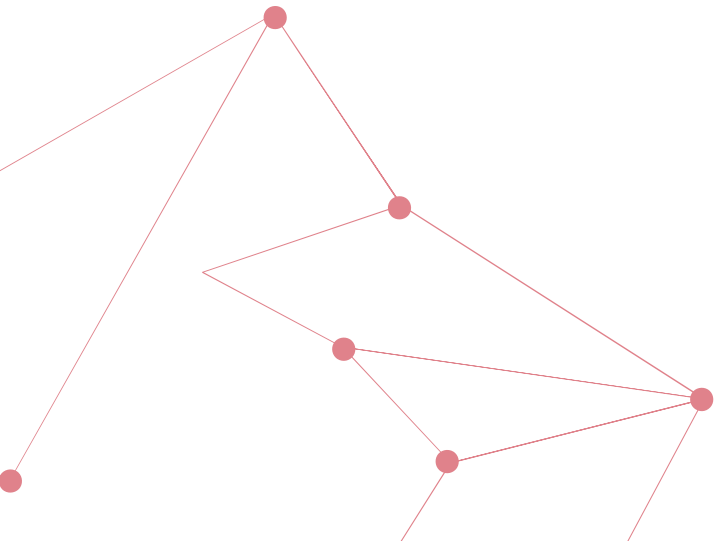
For measures:

From census.gov/data/developers/data-sets/acs-5year.html. There are different API call structures for each kind of table:

- **Detailed Tables** contain the most detailed cross-tabulations, many of which are published down to block groups. The data are population counts. There are over 20,000 variables in this dataset.
  - B28002_004E: Households with an internet subscription of any kind
  - B28002_001E: Presence and types of internet subscriptions in household
  - B15002_001E: Total population—sex by educational attainment for the population 25 years and over
  - B15002_011E: Male—sex by educational attainment for the population 25 years and over
  - B15002_028E: Female—sex by educational attainment for the population 25 years and over
  - B17020_002E: Total population—income in the past 12 months below poverty level
  - B17020_004E: Ages 6 to 11—income in the past 12 months below poverty level
  - B17020_005E: Ages 12 to 17—income in the past 12 months below poverty level



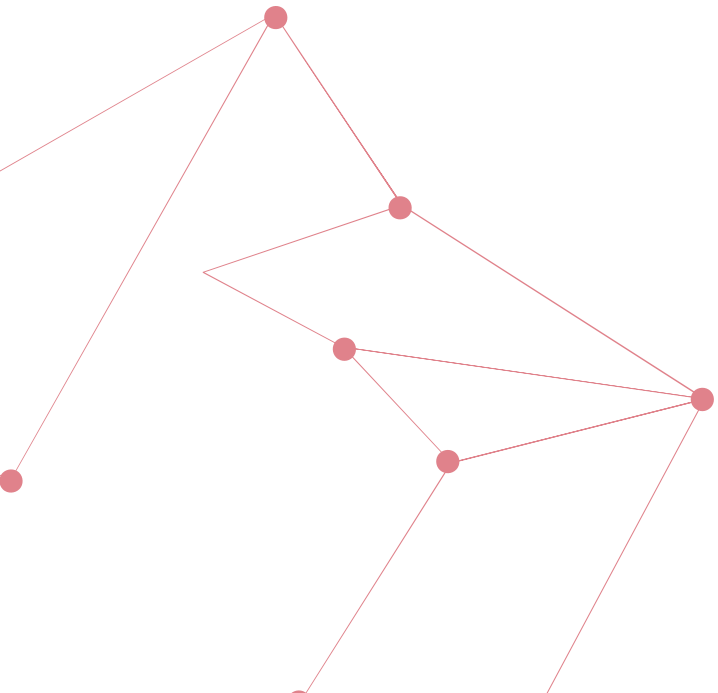
- **Subject Tables** provide an overview of the estimates available on a particular topic. The data are presented as population counts and percentages. There are over 18,000 variables in this dataset.
  - ACS-5 subject tables that were used for the index:
  - S1602_C03_001E: Limited English-speaking households
  - S1602_C01_001E: Total—limited English-speaking households
  - S2701_C03_003E: Percent insured—civilian noninstitutionalized population—6 to 18 years of age
  - S2201_C04_009E: Percent households with children under 18 years receiving food stamps
  - S2503_C01_001E: Occupied housing units (estimated)
  - S2503_C02_002E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) less than $5,000
  - S2503_C02_003E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $5,000 to $9,999
  - S2503_C02_004E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $10,000 to $14,999
  - S2503_C02_005E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $15,000 to $19,999
  - S2503_C02_028E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income of less than $20,000
  - S2503_C02_006E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $20,000 to $24,999
  - S2503_C02_007E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $25,000 to $34,999



- S2503_C02_032E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income $20,000 to $34,999
- S2503_C02_008E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $35,000 to $49,999
- S2503_C02_036E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income $20,000 to $49,999
- S2503_C02_009E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) $50,000 to $74,999
- S2503_C02_040E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income $50,000 to $74,999
- S2503_C02_010E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) 75,000 to $99,999
- S2503_C02_011E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) 100,000 to $149,999
- S2503_C02_012E: Percent occupied housing units with household income in the past 12 months (in 2019 inflation-adjusted dollars) 150,000 or more
- S2503_C02_044E: Percent occupied housing units with monthly housing costs of 30 percent of more of household income in the last 12 months—for households with annual income $75,000 or more

- **Data Profiles** contain broad social, economic, housing, and demographic information. The data are presented as population counts and percentages. There are over 1,000 variables in this dataset.
  ACS-5 profile tables that were used in the index:
  - DP04_0003E: Vacant housing units
  - DP04_0001E: Total housing units

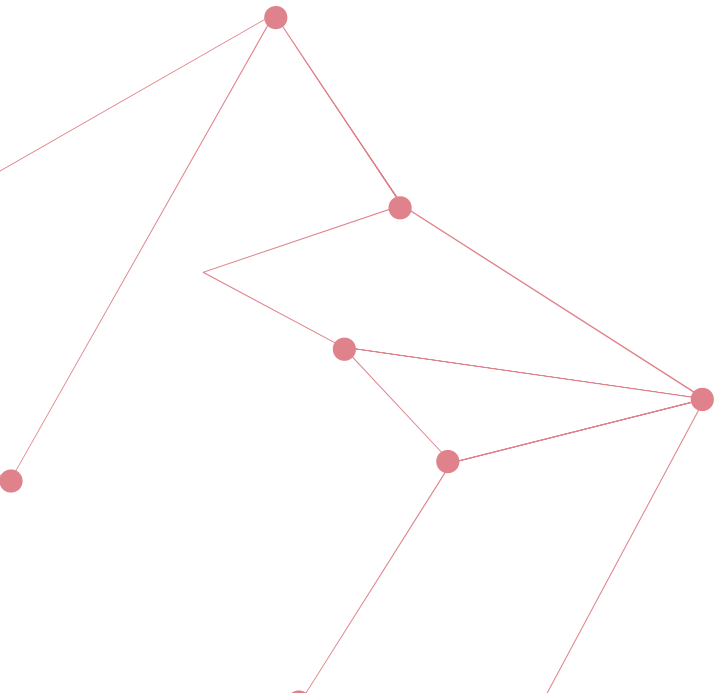
In some instances, data are updated annually but there are reporting lags of one or more years. County Health Rankings and City Health Dashboard use tertiary data sources and, in some cases, querying for the latest year available does not guarantee that measures will be for those years. For example, County Health Rankings published 2021 data on violent crime using FBI data from 2014 and 2016.



Notes on inclusion

Data from Puerto Rico are collected for nine of the sixteen measures. Those measures are unemployment, poverty, access to broadband internet, high-school graduates, linguistic isolation, access to healthcare, SNAP recipients, housing vacancy rate, and housing affordability.

Minorities (i.e., other than "non-Hispanic Whites") are systematically undercounted in the American Community Survey data collected by the U.S. Census Bureau. While undercounted communities are disserved when associated census data are used for population-based apportionment of resources or representation, they are of less concern for our present purposes. That is, census tracts are, by design, small and relatively homogeneous. So, even if smaller shares of the population in high-minority tracts are surveyed than in low-minority tracts, the people surveyed nonetheless accurately reflect the population in their tract, so there is no systematic bias in the tract-level characteristics.

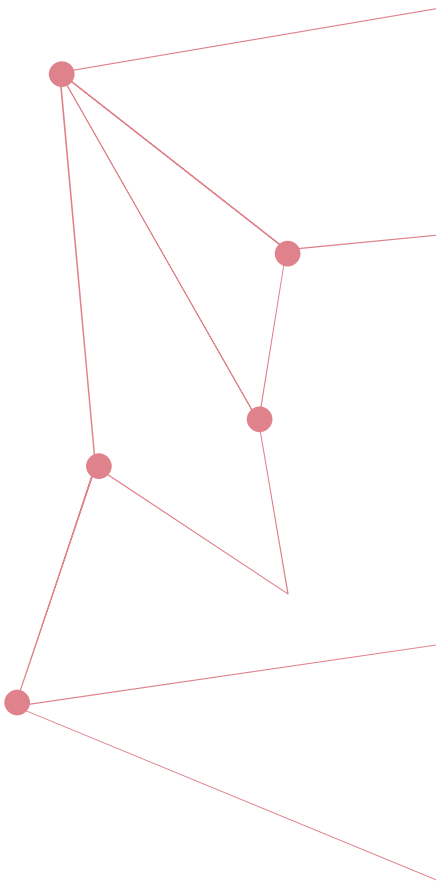


How measures were selected

Measures were selected using the following criteria:

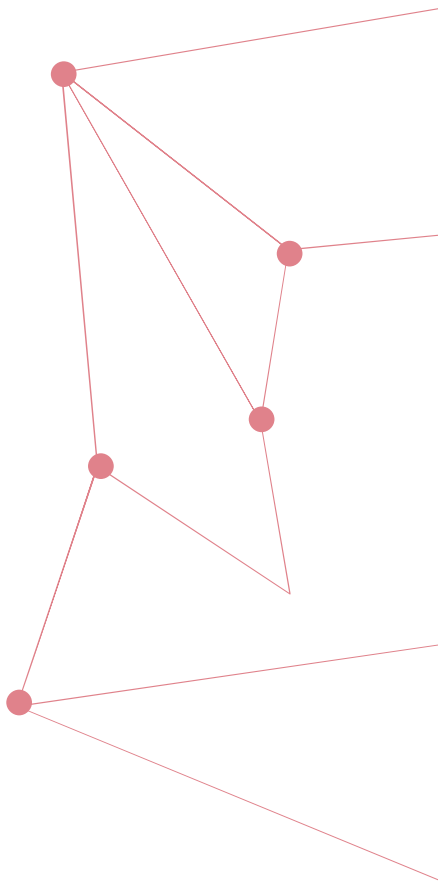
Indicator/measure attributes

- **Relevance**: Does the indicator/measure capture some important feature of the domain, so that variation in the indicator/measure affects the value of the index? More relevant is better.
- **Distinctiveness**: Does the indicator/measure pick up aspects that others in the index do not? In other words, is it not highly correlated with another indicator/measure? (It is, then, an attribute of an indicator/measure only in the context of others in the index.) More distinctive is better.
- **Monotonicity**: Is it clear that (all else equal) an increasing value of an indicator/measure is either good (e.g., access to dental care) or bad (e.g., lead level in tap water)? Some indicators/measures are probably best at some intermediate level (e.g., average number of physician visits per year). More clearly monotonic is better.
- **Potentiality**: Ostensible "access" indicator/measures can be complex, as they may reflect both supply and demand. The share of the population using a good or service may reflect both its availability/affordability to those who want it and the share that want it. For example, consider measuring "access to public transportation" by "percentage of adults using public transit to commute to work." A low figure could reflect low demand, which is fully met by supply, or high demand that is unmet by supply. Even good data that shed light on "does everyone who wants X have access to X?" can founder on whether wanting X itself is an indicator of stress—is it better to have a large share wanting X and getting it ("good access") or a small share wanting X and not getting it ("poor access")?
- **Variation**: Does the indicator/measure vary much across the units of analysis? In the U.S., for instance, there is near-universal access to electricity, so the measure "residential electrification rate" is not especially useful in distinguishing low-stress from high-stress neighborhoods. More variation is better.



Data Attributes

- **Accessibility**: Are the data available from a source that is easily accessible electronically, at no cost, without restrictions on use (e.g., U.S. Census), or are there technical/administrative/financial barriers (e.g., proprietary market-research companies)? Are the data in an easy-to-use format (e.g., .csv files) with good documentation, or in a difficult-to-use format (e.g., .pdf files)? Are the data readily downloaded in one file transfer, or does a script need to be written to automate (e.g., state-by-state download)? More accessible is better.
- **Coverage**: Are the data available over the entire area of interest (e.g., all U.S.), or only some areas (e.g., large cities)? Broader coverage is better.
- **Granularity**: Are the data available at a fine geographic level (e.g., census tract), or only at coarser levels (e.g., state)? More granular is better. Combining data at different, overlapping levels of granularity (e.g., ZCTAs and tracts) complicates index construction.
- **Frequency**: Are the data updated frequently (e.g., annually), infrequently (e.g., decennially), or never (i.e., from a one-time study)? More frequent is better.
- **Recency**: Were the data updated recently (e.g., 2019), or not recently (e.g., 2013)? More recent is better.



Usage

How to use the tool

The website hosts a user interface that includes a customizable table with visualizations, as well as search and filter options.

Data visualizations

Some visual encodings display information in the tool:

Index table

This is the main component of the site. Each row represents a high school in the NCES data set. Data on each school (columns) include: school name, school district, state, county, and zip code, and the values of the index and each of the domains. Toggle switches allow for selecting which columns to display.

Flags

These flags (either or both) appear next to some school names:

* Gini index for county household income higher than 0.5. *Click* on  to display exact value.
† More than 3 measures missing.

The first flag applies to 9.0% of the schools. The Gini index is a measure of income inequality at the county level.[11] Since seven of the index measures are collected at the county level, and some counties have high inequality, interpretation of index and domain values for these schools needs to consider this. For example, consider a high-inequality county, with some very wealthy and some very poor neighborhoods. Schools in those respective neighborhoods will have the same values of the county-level measures, even if the local values are considerably different.

[11] Notionally, if every household had the same income, the Gini index would be 0, and if one household had all the income (and all the others had no income), the Gini index would be 1.

The second flag applies to 3.4% of the schools. Index values are displayed for these schools, but a measure of caution should be applied in interpreting index values when data are missing.

Medians and hover-over distribution descriptives

When hovering over a school's index or domain values, each cell in the corresponding column (domain) displays the horizontal-length-encoded median value for each domain or for the index value, as well as details such as the value, the percentile rank (the percentage of schools with domain or index values lower than those of each school) and the median value of such column across all schools. See Figure 1.

Figure 1. Median Values Displayed as Horizontal Lengths and Hover-over Window with Descriptive Values.

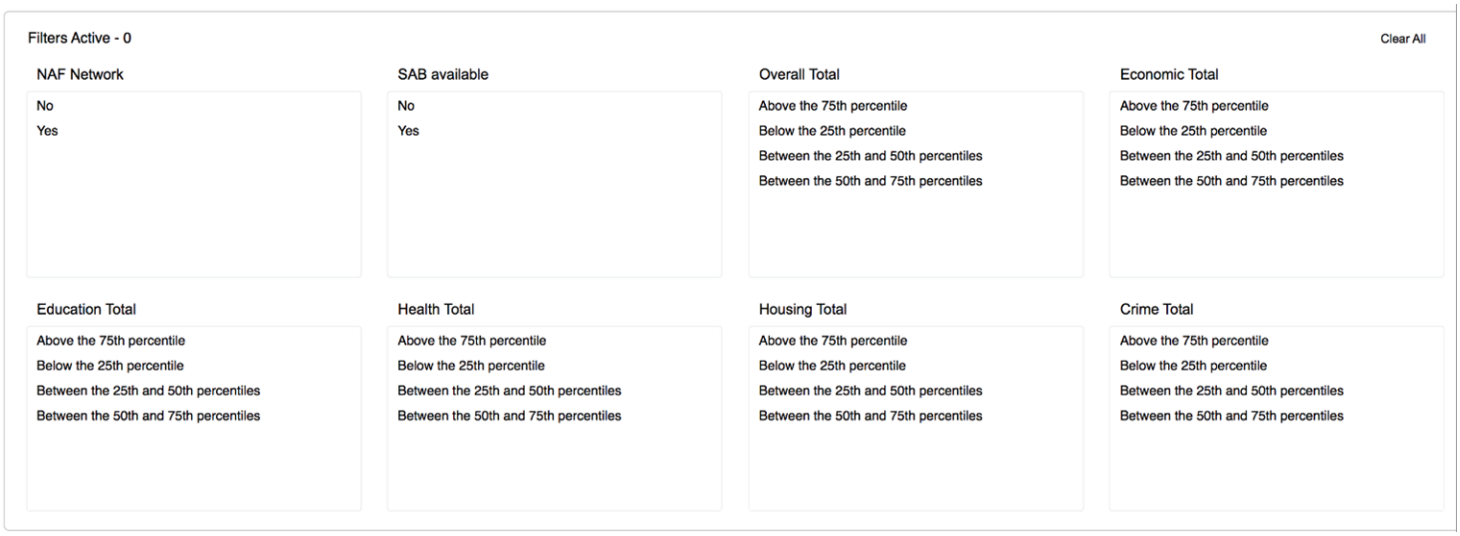
Domain Totals						
Economic	Education	Health	Housing	Crime	Composite Score	
56	82	41	47	64	58	
57	83	41	44	64	58	
57	82	41	45	64	58	
59	87	36	30	64	55	
51	68	42	47	64	54	
51	62	41	45	64	53	
51	65	40	44	64	53	
58		34	20	64	52	
58		34	20	64	52	

Value: 51
Percentile Rank: 96th
Median: 34

Filters

When clicking the "Filter Results" button, the following options appear (see Figure 2). Selecting one or more groups reduces the set of schools that appear in the table visualization (and in exported files) to those (if any) that match the selected criteria. For examples, selecting "Yes" for "NAF Network" displays only high schools that belong to the NAF network. "SAB available" are those schools for which there is an SAB shapefile, rather than our proxies for communities using zip codes.

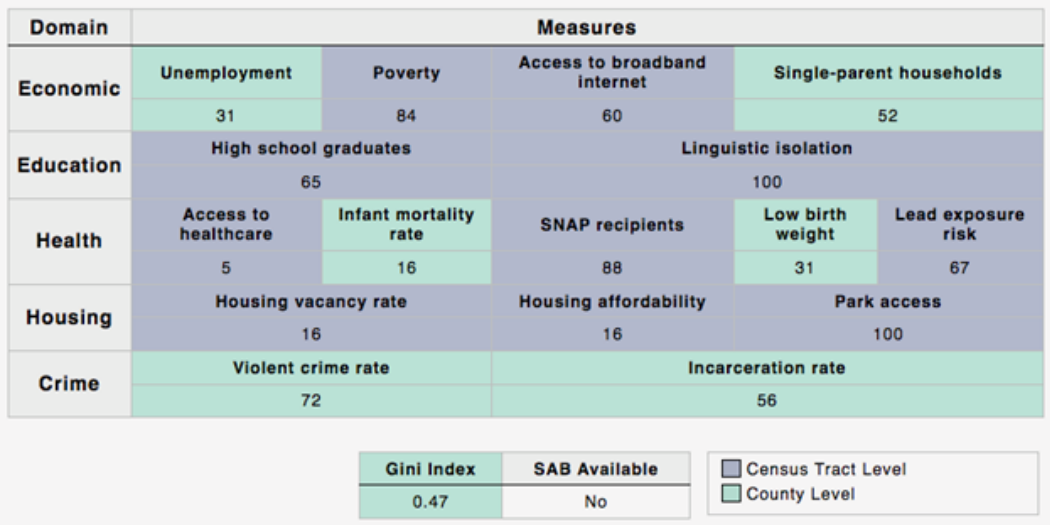
Figure 2. Filter Settings.



Details on each school

Clicking the ⊕ symbol on the right expands details about all measures in each domain (see Figure 3, page 40).

Figure 3. Domain and Measure Details.

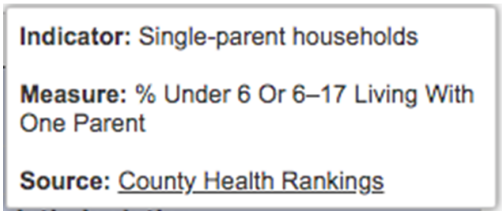


Domain	Measures				
Economic	Unemployment	Poverty	Access to broadband internet	Single-parent households	
	31	84	60	52	
Education	High school graduates		Linguistic isolation		
	65		100		
Health	Access to healthcare	Infant mortality rate	SNAP recipients	Low birth weight	Lead exposure risk
	5	16	88	31	67
Housing	Housing vacancy rate		Housing affordability	Park access	
	16		16	100	
Crime	Violent crime rate		Incarceration rate		
	72		56		

Gini Index	SAB Available
0.47	No

Census Tract Level
County Level

Figure 4. Measure Description



Hovering over variable names results in the display of a variable description (see Figure 4, left).

Search options

Searches can be performed by state (with a dropdown of predefined options including all 50 states and Puerto Rico), county, city, school zip code,[12] school district, and school name, using complete words or parts of words (or numbers, in the case of zip codes).

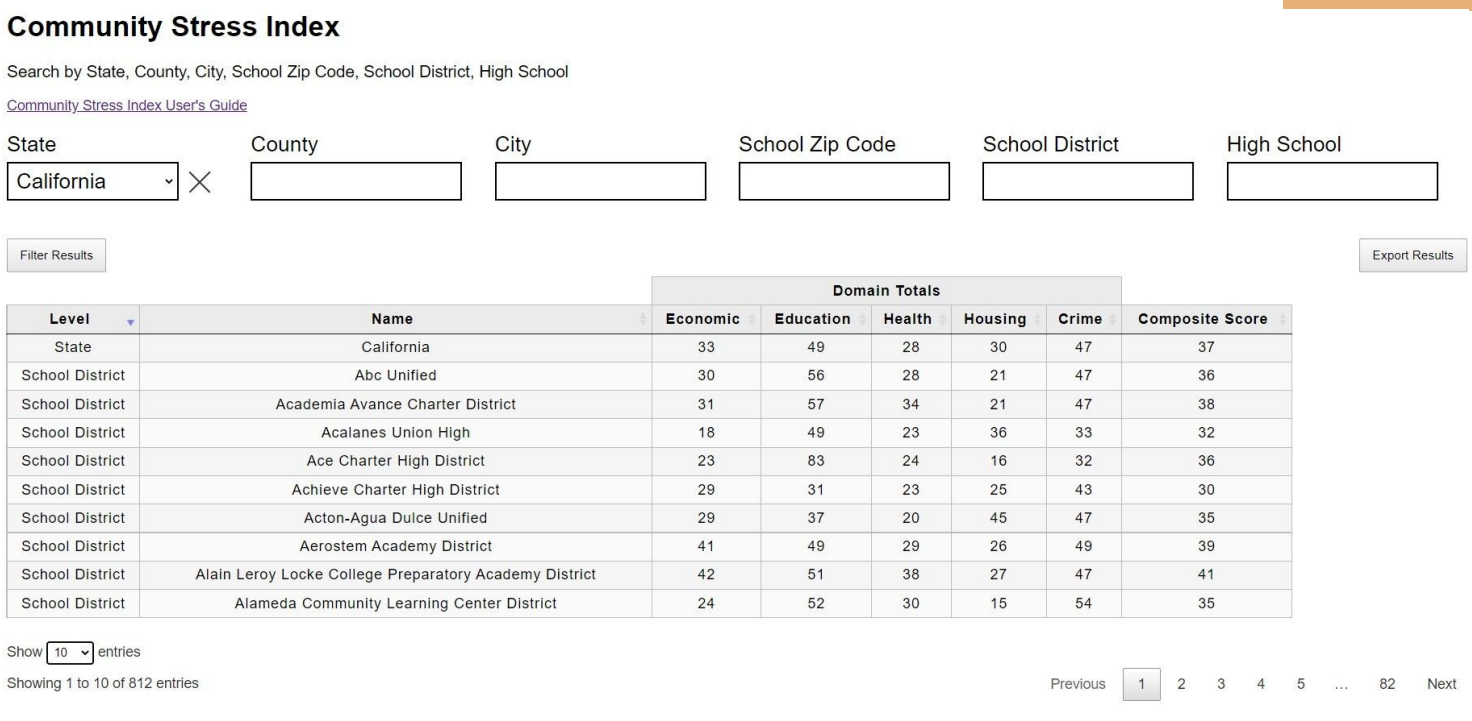
When executing a search and hitting enter, the main visualization table displays all schools that match the search criterion, and a second table appears above it. This table displays aggregated values at the state, county, zip code, or school-district levels, depending on the query terms.

[12] This search returns any schools located in the selected zip code, not schools outside the zip code serving students in it.

Using multiple criteria displays results for schools (if any) but, due to potential ambiguities in aggregation levels (for example, when geographies are not disjoint), the aggregated-results table does not appear.

In general, searching by a specific geographic unit yields aggregation. Searching by state yields a table displaying aggregated values at the state level and all school districts within the state (see Figure 5, below). Searching by county yields aggregated results at the county and school-district levels. Searching by city returns aggregated results by city (all cities matching the name in the query) and school district (all school districts within such cities). Searching by school does not generate an aggregated results table; doing this displays only those schools in the main table that match the name in the query.

Figure 5. Aggregated Values by State Search.



Community Stress Index

Search by State, County, City, School Zip Code, School District, High School

Community Stress Index User's Guide

State: California | County | City | School Zip Code | School District | High School

Filter Results | Export Results

		Domain Totals					
Level	Name	Economic	Education	Health	Housing	Crime	Composite Score
State	California	33	49	28	30	47	37
School District	Abc Unified	30	56	28	21	47	36
School District	Academia Avance Charter District	31	57	34	21	47	38
School District	Acalanes Union High	18	49	23	36	33	32
School District	Ace Charter High District	23	83	24	16	32	36
School District	Achieve Charter High District	29	31	23	25	43	30
School District	Acton-Agua Dulce Unified	29	37	20	45	47	35
School District	Aerostem Academy District	41	49	29	26	49	39
School District	Alain Leroy Locke College Preparatory Academy District	42	51	38	27	47	41
School District	Alameda Community Learning Center District	24	52	30	15	54	35

Show 10 entries

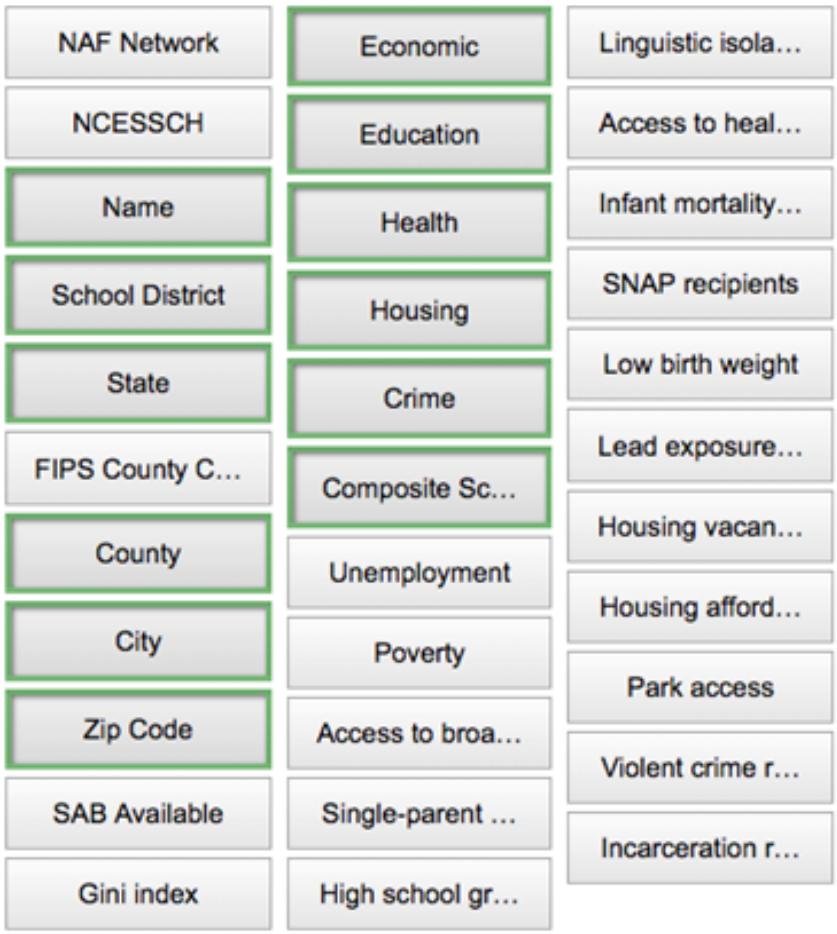
Showing 1 to 10 of 812 entries

Previous 1 2 3 4 5 … 82 Next

Customization

Clicking on the "Column visibility" button reveals options for which columns are to be displayed (see Figure 6, page 42).

Figure 6. Column Visibility Toggle.



These options can be toggled on and off, to make columns with matching names visible or invisible in the main table visualization. The columns of the default view of the table are highlighted in Figure 6.

Data export

Filters and search results (or the default view of the index) at the school level can be exported into .csv or .xlsx format (see Figure 7, page 43). These files contain all measure values regardless of the column-visibility settings.

"Excel" is a built-in export for the current view (visible columns, current ordering, and active filters) with rounded values. "Raw data" exports filtered high schools with raw metrics values. "Raw totals" exports filtered high-school data after processing them, with three decimal places (this is the only difference from the "Excel" option). "Raw" options are in .csv format.

Figure 7. Data Export Options