



Community Indicators: Census Variables

Data Methodology and Statistical Choices

How the community indicators are sourced, derived, and tested for reliability

Cameron Caputi and Miranda Santillo

Updated June 2026

1. Overview

The census community indicators dashboards present American Community Survey (ACS) measures for the eight-county region and for the towns and cities within it. This document records the data source, how each measure is produced, how statistical reliability is handled, and how the published values are validated. It is the reference for anyone reviewing the numbers or maintaining the pipeline. The companion document, Annual Update Procedure, covers the operational steps.

Scope. This document covers the census-sourced dashboards only. The dashboard system also includes environmental and transportation dashboards, which are built by separate pipelines and do not draw on the ACS. Everything in this document, the margin-of-error treatment, suppression, regrouping, and validation, applies to the census dashboards. The non-census dashboards and their sources are listed in Section 2.1 for reference but are not otherwise covered here.

2. Data source and vintage

Source. All measures come from the U.S. Census Bureau American Community Survey (ACS) 5-Year estimates. The 5-year product is used, rather than the 1-year, because the dashboards report at the county subdivision level and for individual race and ethnicity groups. At those small populations the 1-year estimates are either not published or too unreliable to use; the 5-year estimates pool sixty months of sample to produce usable figures.

Vintage. A 5-year estimate is labeled by its end year. The current vintage, labeled 2024, covers 2020 through 2024. Every variable in a release uses the same vintage.

2.1 Companion dashboards not covered here (non-census sources)

The environmental and transportation dashboards are produced by separate pipelines (environmental.py and transportation.py) and use the agency sources below rather than the ACS. They carry their own reliability conventions and are checked against their respective source releases, not against data.census.gov. They are listed here only so the published system is documented in one place. This table lists the indicators that are currently published.

Area	Indicator	Source
Environmental	Asthma deaths (age-adjusted) and hospitalizations by age group	NYS Dept. of Health Asthma Dashboard (SPARCS data)
Environmental	Clean Energy Communities status and designation	NYSERDA, Clean Energy Communities Program, CEC Map (as of November 2024)
Environmental	Clean energy jobs	NYSERDA, New York Clean Energy Industry Report
Environmental	Food access	USDA Economic Research Service, Food Access Research Atlas (2019)
Transportation	Transit on-time performance	RGRTA via NY Open Data
Transportation	Transit ridership	RGRTA via NY Open Data

3. Geographies

Two levels are published. Counties, the fixed set of eight, are always shown. County subdivisions, the towns and cities, are shown when they have at least one reliable estimate; subdivisions with no reliable estimate in any year are removed from the selector (see Section 8).

4. Variables and source tables

The table below maps each published census chart to its ACS source table. It lists the twenty-five census charts that are currently published. All have been independently reconstructed from raw Census data and reconciled against the dashboard output cell by cell in the most recent validation.

Variable	ACS table	Audited
Race and ethnicity distribution	B03002 (regrouped)	Yes
Age distribution	S0101 (regrouped)	Yes
Commute time distribution	B08303 (regrouped)	Yes
Transportation mode distribution	B08301 (regrouped)	Yes
Median household income	B19013	Yes
Median owner-occupied value	DP04	Yes
Median rent	DP04	Yes
Percent below poverty (past 12 months)	S1701	Yes
Percent under 18 below poverty	S1701	Yes
Percent housing units vacant	DP04	Yes
Percent owner-occupied	DP04	Yes
Percent households receiving SNAP	S2201	Yes
Percent with an internet subscription	S2801	Yes
Percent uninsured (civilian noninstitutionalized)	S2701	Yes
Percent with no vehicle available	S2504	Yes
Total population	B01003 / S0101	Yes
Homeownership rate by race	B25003 plus iterations	Yes
Median income by race	B19013 plus iterations	Yes
Poverty by race, all ages	B17001 plus iterations	Yes
Poverty by age	S1701	Yes
Female poverty, total	B17001	Yes
Senior poverty, total	S1701 (age 65+ line)	Yes
Employment-to-population ratio	S2301	Yes
Rent burden, combined rate	B25070 (summed)	Yes
Owner cost burden, combined rate	B25091 (summed)	Yes

Race iteration suffixes. Across the iteration tables the suffixes used are H (White alone, not Hispanic or Latino), B (Black or African American alone), I (Hispanic or Latino), D (Asian alone), and G (Two or More Races).

5. Published versus derived estimates

Some variables arrive from Census already expressed as percentages and are used directly. The subject tables (the S series) and the data profile percent columns (DP04 ...PE) fall in this group; age distribution, for example, uses the published S0101 percentages.

Other variables are derived as a share of a total from a count table (the B series). A derived share is computed as the estimate divided by its denominator, times one hundred. Race composition and commute time are derived this way, and so is the homeownership rate, which is owner-occupied households divided by total occupied households for each group. For these count-derived charts, the underlying numerator count is retained alongside the rate and its margin, and appears in the chart hover (for example, the number of rent-burdened households behind a rent-burden percentage). Retaining the count does not change any published rate or margin; it adds context only. Whenever a value is derived, its margin of error must also be derived, which is the subject of the next section.

Combined burden rates. The dashboard publishes a single combined cost-burden rate for renters and for homeowners: the share of households that are burdened at all. It is formed by summing the two burdened brackets (cost burdened, 30 to 49.9 percent, and severely burdened, 50 percent or more) into one rate. The summed count is the sum of the two bracket counts, and the combined margin is the root of the sum of the squared bracket margins, per Section 6.2. This is the same summing method used for the transportation mode catch-all in Section 7.4.

6. Margins of error

What a margin of error is. Every ACS estimate carries a margin of error (MOE), the half-width of a 90 percent confidence interval. An estimate of 20 percent with an MOE of 3 means the true value is, with 90 percent confidence, roughly between 17 and 23 percent.

6.1 Relative MOE and the reliability cutoff

Reliability is judged by the relative MOE, the margin expressed as a share of the estimate:

$$\text{relative MOE (\%)} = \text{MOE} / \text{estimate} \times 100$$

An estimate is suppressed when its relative MOE exceeds 50 percent (MOE_CUTOFF = 50). Equivalently, an estimate is kept only when it is at least twice its own margin of error. Suppressed values are not plotted; they are labeled so that the reader sees that data exists but is too uncertain to show.

Why 50 percent. The cutoff is a deliberate balance, not an arbitrary round number. A relative MOE of 50 percent is equivalent to a coefficient of variation of about 30 percent, which is close to the level commonly used to flag survey estimates as unreliable. The choice trades two failure modes against each other. A stricter cutoff would blank out a large share of the small-town and small-group cells, leaving the dashboards mostly empty at the subdivision level, which is exactly where the audience most wants local detail. A looser cutoff would display estimates so noisy that a reader could draw conclusions from sampling noise. Fifty percent keeps the maps and charts populated while removing figures that cannot be read with any confidence, and it is applied uniformly to every variable so that the standard is consistent and easy to defend.

6.2 MOE of a combined category (a sum)

When fine categories are summed into coarse groups (Section 7), the combined margin is the root of the sum of the squared component margins:

$$\text{MOE}_{\text{sum}} = \text{sqrt}(\text{MOE}_1^2 + \text{MOE}_2^2 + \dots + \text{MOE}_k^2)$$

Census zero-estimate rule. Among the components whose estimate is zero, only the single largest margin is kept; the others are dropped from the sum. This follows Census guidance and prevents a string of zero cells from inflating the combined margin.

Because the summed shares share a common denominator (the area total), this root-sum-of-squares is slightly conservative: it omits a small reduction that the shared denominator would allow. That is a deliberate choice to err toward suppression, and it is consistent with the rest of the pipeline (see Section 6.5).

6.3 MOE of a derived proportion

A derived percentage is a proportion, $p = \text{numerator} / \text{denominator}$. The correct margin depends on whether the numerator is part of the denominator.

Numerator is a subset of the denominator (for example, one race group within the total population, or owner-occupied households within total occupied households). The recommended formula, the proportion MOE or `moe_prop`, subtracts a term:

$$\text{MOE}_{\text{prop}} = \text{sqrt}(\text{MOE}_{\text{num}}^2 - p^2 \times \text{MOE}_{\text{den}}^2) / \text{denominator}$$

The minus sign reflects that some of the numerator's uncertainty is already contained in the denominator's, so the two are not independent. This makes the proportion MOE smaller than a naive combination would suggest.

6.4 Three kinds of zero on a rate

Why a separate rule is needed. The relative-MOE test divides by the estimate, so it is undefined when the estimate is zero. Without a dedicated rule a zero with a very large margin would pass through unchallenged and display as a confident 0. A zero is also genuinely ambiguous: it can mean three different things, and the margin of error is what tells them apart. The rate is resolved into these three cases:

Case	Signature	Shown as
No denominator	estimate 0, MOE 0 or missing (for example, a race group with zero households in a place)	No data
A real zero rate	estimate 0, with a small valid MOE at or below the zero limit	0
An unreliable zero	estimate 0, with a large MOE above the zero limit	Suppressed

The zero limit (MOE_ZERO_LIMIT = 10) is in the estimate's own units, which for shares and rates is percentage points. It is set so that a confident near-zero (a margin under about one point) reads as a genuine 0, while a zero whose interval could plausibly reach well above ten points is suppressed. The no-denominator case is important on the race charts: showing 0 percent for a group that has no households in a place would read as a real disparity, when the truth is simply that the group is not present, so it is shown as No data.

Why ten points. In ACS data a genuine zero, a rare group truly absent from a place, almost always carries a small derived margin, well under one percentage point, while a zero produced by thin sample carries a much larger one. A threshold of ten points sits comfortably between the two, so confident zeros are still displayed as 0 rather than being over-suppressed, and only zeros whose confidence interval could plausibly reach a meaningful share are withheld. The value is a documented judgment call and can be tuned; ten was chosen to match the permissive spirit of the 50 percent cutoff rather than to suppress aggressively.

6.5 The homeownership MOE: an explicit choice

The homeownership rate by race is a proportion whose numerator (owner-occupied households of a race) is a subset of its denominator (total occupied households of that race). By the rule in Section 6.3, the proportion formula (moe_prop) is the textbook choice.

The pipeline instead uses the more conservative ratio-style margin. This is intentional, for two reasons. First, consistency: the only place the dashboard computes its own margin (the combined-category sum in Section 6.2) deliberately errs conservative and toward suppression, so a conservative homeownership margin matches that lean, whereas the proportion formula would make homeownership the single looser exception. Second, boundary safety: at 100 percent (or 0 percent) homeownership on a handful of households, the proportion formula collapses the margin toward zero, which would publish a falsely precise rate; the conservative margin suppresses those cells instead.

The effect is narrow. The underlying rates are identical under either formula. Only the display threshold differs, so a small number of small race-group cells are suppressed that the proportion formula would have shown. This is a reliability-display decision, not a difference in the computed values.

7. Category regrouping

Four distributions are collapsed from many fine ACS categories into a few readable groups: age, commute time, race and ethnicity, and transportation mode. Regrouping improves readability and reliability, since the combined groups have larger counts and smaller relative margins than the individual fine categories. Shares are summed and the margins are combined per Section 6.2. The rent burden and owner cost burden combined rates (Section 5) are formed by the same summing method.

7.1 Age (S0101): 18 buckets to 6 groups

Display group	Source age buckets
Under 5	Under 5 years
5 to 19	5 to 9, 10 to 14, 15 to 19
20 to 24	20 to 24
25 to 49	25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49
50 to 64	50 to 54, 55 to 59, 60 to 64
65 and over	65 to 69, 70 to 74, 75 to 79, 80 to 84, 85 and over

7.2 Commute time (B08303): 12 bins to 4 groups

Display group	Source bins
Less than 10 min	Less than 5, 5 to 9
10 to 29 min	10 to 14, 15 to 19, 20 to 24, 25 to 29
30 to 59 min	30 to 34, 35 to 39, 40 to 44, 45 to 59
60+ min	60 to 89, 90 or more

7.3 Race and ethnicity composition (B03002): 8 groups to 6

Display group	Source groups
White alone, not Hispanic or Latino	White alone, not Hispanic or Latino
Black or African American alone	Black or African American alone
Hispanic or Latino	Hispanic or Latino (any race)
Asian alone	Asian alone
Two or More Races	Two or More Races
Other races and ethnicities	American Indian and Alaska Native alone, Native Hawaiian and Pacific Islander alone, Some Other Race alone

Composition shares sum to approximately 100 percent within each geography, which the validation confirms.

7.4 Transportation mode (B08301): worker commute modes

Display group	Source lines
Drove Alone	Car, truck, or van, drove alone
Carpooled	Car, truck, or van, carpooled (all sizes)
Public Transportation	Public transportation (excludes taxicab)
Bicycle	Bicycle
Walked	Walked
Taxi, Motorcycle, or Other	Taxicab, Motorcycle, Other means
Worked from Home	Worked from home

Combined catch-all. The Taxi, Motorcycle, or Other group sums three separate source lines (taxicab, motorcycle, and other means) into a single bar. It is built with the category-distribution method, so the three counts are added before dividing by the worker total, and their margins are combined per Section 6.2. Computing the catch-all as a single mapped line instead would understate it, because two of the three source lines would be dropped.

7.5 Rent and owner cost burden brackets (B25070, B25091)

The combined rent burden rate sums the two burdened brackets below (30 to 49.9 percent and 50 percent or more) over renter-occupied units. The owner cost burden combined rate is formed the same way from the analogous B25091 brackets over owner-occupied units. The bracket structure for rent burden is:

Bracket	Definition
Not Burdened	Less than 30 percent of household income (all brackets below 30 percent)
Rent Burdened	30.0 to 49.9 percent
Severely Burdened	50.0 percent or more
Not Computed	Not computed

8. Suppression and display rules

- **Point suppression.** Any observation that fails the relative MOE test or the zero rules has its value removed and is labeled Suppressed or No data, on bars, trend lines, and maps alike. These labels are rendered vertically so they do not overlap adjacent bars.
- **Subdivision drop.** A town or city with no reliable estimate in any year is removed from the subdivision selector entirely, so the list is not cluttered with places that can never display.
- **Counties are never dropped.** The eight-county set is fixed and always present, even where a particular cell is suppressed.

9. Race rate charts: groups shown

The by-race rate charts (homeownership, median income, and poverty by race, all ages) display five groups: White alone not Hispanic or Latino, Black or African American alone, Hispanic or Latino, Asian alone, and Two or More Races. The smaller iterations (American Indian and Alaska Native, Native Hawaiian and Pacific Islander, and Some Other Race) are not charted as rates, because at this geography their counts are too small to produce reliable rates. They remain part of the composition chart inside the combined Other group.

10. Validation

The published values are checked on two layers before release.

- **Layer A, dashboard against the data pull.** Confirms the transform code (regrouping, derivation, suppression) reproduces the values in the R pull CSVs.
- **Layer B, data pull against raw Census.** An independent tool downloads the source tables from data.census.gov and rebuilds each displayed value from scratch using the formulas in this document, then a diff compares the independent reconstruction to the dashboard output cell by cell.

The dashboard publishes twenty-five census charts. In the most recent validation, all twenty-five were independently reconstructed from data.census.gov and compared cell by cell against the dashboard output, across the eight counties and a sample of towns chosen to span the population range where suppression takes effect. Every comparable value matched. Every label matched except the homeownership by race cells that differ only because of the deliberate conservative margin choice in Section 6.5. The composition, age, commute time, and transportation mode shares all summed to approximately 100 percent.

11. Constants and tolerances

Setting	Value	Meaning
MOE_CUTOFF	50	Suppress when relative MOE exceeds this percent
MOE_ZERO_LIMIT	10	For a zero estimate, the margin (in the estimate's units) above which the zero is suppressed
Value tolerance (audit)	0.4	Allowed difference for shares and percentages, in points
Value tolerance (audit)	~2	Allowed difference for dollar and count variables

12. Known limitations

- ACS estimates carry sampling error by design. Margins of error are reported and drive suppression, but uncertainty remains in every displayed estimate.
- The homeownership by race chart, by the conservative margin choice, displays fewer small-group cells than the standard proportion formula would. The rates shown are correct; some reliable-by-the-looser-standard cells are withheld.
- Validation samples a set of subdivisions rather than every town. The transform is uniform across subdivisions, so a systematic error would surface in the sample, but cell-by-cell coverage of every place is not claimed.

13. Formula quick reference

Quantity	Formula
Derived share or rate	$\text{estimate} = \text{numerator} / \text{denominator} \times 100$
Relative MOE	$\text{MOE} / \text{estimate} \times 100$
MOE of a sum	$\sqrt{\text{sum of component MOE}^2}$, largest zero only
MOE of a proportion (subset)	$\sqrt{\text{MOE}_{\text{num}}^2 - p^2 \times \text{MOE}_{\text{den}}^2} / \text{denominator}$
MOE of a ratio (non-subset)	$\sqrt{\text{MOE}_{\text{num}}^2 + p^2 \times \text{MOE}_{\text{den}}^2} / \text{denominator}$
Reliability test	keep when relative MOE \leq 50 percent