

## State-by-Year ACS Data Set Documentation

### Overview

The state-by-year ACS data set includes state-level data for all 50 states and the District of Columbia for the years 2018 to 2022. All variables in this data set are constructed using individual-level survey data from the American Community Survey (ACS) 1-year files. You can read more about the ACS at the [IPUMS website](#).

- Data set is .csv formatted – to load into R, use the `read_csv()` function from `tidyverse`.
- Each row of the data set corresponds to a state and year, so rows can be uniquely identified using a combination of `statefip` or `state_name` and `year`.
- Documentation includes lists all variables and includes definitions and coding notes.
  - Pay attention to variable definitions – in some cases, the manner in which variables are constructed should inform how you interpret a given variable. For example, commuting variables below are all conditional on employment.

Variable Name	Definition	Coding and Notes
year	Year (covers 2018-2022)	
statefip	State FIPS code – numeric ID for states	See <a href="#">link here</a> for list of states and codes (you can also use the <code>state_name</code> variable to ID states)
state_name	State name (string format)	
census_region_name	Census region name	Provides a way of grouping states across regions of the US, defined by the Census Bureau – includes West, Midwest, Northeast, and Southeast. See <a href="#">link here for map</a> of regions.
labor_force_participation_rate	Labor force participation rate	Denominator includes all individuals ages 16 and older.
employment_rate	Employment rate	Fraction of the labor force that is currently employed.
hhincome	Average household income	Average household income – includes income from all sources from all individuals living with respondent.
incwage	Average individual wage and salary income	Average individual earned income from wages and salary – calculated using personal earnings of survey respondents.
poverty_rate	Fraction of population with incomes under poverty line	Family-level poverty rate equal to fraction of people with family incomes less than the federal poverty line (FPL). FPL is calculated based on a number of factors including family size.

<b>Variable Name</b>	<b>Definition</b>	<b>Coding and Notes</b>
food_stamp_reciprocity	Food stamp / SNAP reciprocity rate	Includes anyone who participated in SNAP program in past year.
homeownership_rate	Fraction of the population who owns their own home	Fraction of the population that owns their own home (including both homes owned free and clear and homes under mortgages).
rental_rate	Fraction the population who lives in a rental housing unit	
rent_average_total	Average total rent (includes utilities)	Average value of total rent paid by renters. Total rent here includes both the contracted lease price of the housing unit as well as utilities. Using total rent allows for better comparisons across housing units (because some apartments include utilities as part of rent while others don't, using just the lease price can be misleading).
attending_school	Fraction of population currently attending school	Fraction of all individuals that are currently enrolled in school (includes pre-K through post-graduate education).
educ_hs_grad	High school graduation rate	Fraction of the population that has graduated high school. Note that this calculated across the entire population, i.e., the denominator includes children who are not old enough to have graduated yet.
educ_college_grad	College graduation rate	Fraction of the population that has graduated college – as with high school graduation rate above, calculated across entire population.
pct_hispanic	Fraction of population reporting Hispanic ethnicity	Fraction of the population that reports Hispanic ethnicity – note that in the ACS, ethnicity and race are separate questions. Thus, adding across the Hispanic variable and race variables below may yield values greater than 100%.
pct_white	Fraction of population that is white	Fraction of the population that reports their race as white – note that the race categories in the ACS are not mutually exclusive (survey respondents can select multiple options).
pct_black	Fraction of population that is black	Fraction of the population that reports their race as black – note that the race categories in the ACS are not mutually exclusive (survey respondents can select multiple options).

<b>Variable Name</b>	<b>Definition</b>	<b>Coding and Notes</b>
pct_asian	Fraction of population that is Asian	Fraction of the population that reports their race as Asian – note that the race categories in the ACS are not mutually exclusive (survey respondents can select multiple options).
pct_female	Fraction of population that is female	
pct_married	Fraction of population that is married	
commute_car	Fraction of population that commutes to work by car	Denominator includes everyone who is employed (i.e., commute variables are calculated conditional on working).
commute_public	Fraction of population that commutes via public transportation	Public transportation includes bus, light rail, trolley, subway, and trains. Denominator includes everyone who is employed (i.e., commute variables are calculated conditional on working).
commute_time	Average amount of daily time spent commuting to work	Number of minutes employed survey respondents report spending commuting to work on an average day.
work_from_home	Fraction of population that works from home	As with commute variables, conditional on employment.
migrate_no_move	Fraction of the population that has not moved in past year	
migrate_same_state	Fraction of the population that has moved within the same state at some point in past year	
migrate_new_state	Fraction of the population that has moved across states at some point in past year	