

STEAP Report Methodology Technical Documentation

What data sources are used to generate the STEAP report?

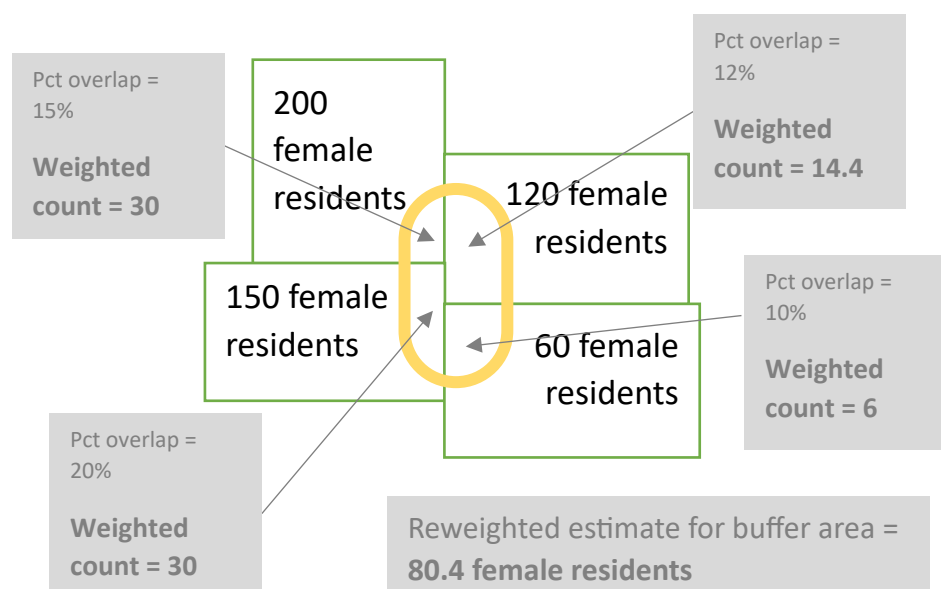
The [U.S. Census' 2021 American Community Survey \(ACS\) 5-Year Estimates summary file dataset](#) is used to generate the socioeconomic data summary tables in the STEAP report. The [U.S. Census' TIGER/Line shapefiles](#) are used to determine block group boundaries. The [DOT Disadvantage Census Tracts National Results Feature Service](#) is used to generate statistics on disadvantaged areas within the user-selected buffer.

How does STEAP calculate the statistics for a buffered area when that buffered area contains only parts of Census block groups?

Most buffers you generate as a STEAP user will not align exactly to the geographic units used by the Census to collect and report data. To generate a report tailored to your specific buffer of interest, STEAP needs to custom-aggregate the Census data to generate custom reports. When you generate a buffer, STEAP does the following:

1. Identify all of the Census block groups that your buffer overlaps.
2. Calculate the percent of overlap between your buffer and each block group that it overlaps.
3. Weight the data for each block group based on what percent of the block group is overlapped by your selected buffer. So, for example, if half of a block group is covered by your buffer, then half of its data will be used in the tabulations for the report.
4. Sum up these weighted values across all block groups that your buffer overlaps.

In the figure below, if the buffer (**yellow oval**) overlaps four block groups (**green rectangles**), the calculation for the count of female residents in the buffer would proceed as follows:



This approach relies on the assumption that the distribution of each reported phenomenon is uniform within a block group. This may not always be an accurate assumption. For example, household vehicle availability may vary within a block group if there is a transit station in one corner of the block group that causes residents to forego car ownership, while residents farther from the transit station prefer to own cars. Or, for example, the monthly housing cost may vary within a block group if the block group happens to contain a plot of expensive luxury apartments, while the rest of the block group constitutes more affordable housing options. Unfortunately, it is not possible to adjust for this kind of microgeographic non-uniformity in a screening tool like STEAP; the user should therefore keep this uniformity assumption in mind when interpreting the data.

Some of the data available in the STEAP report is only reported by the Census at the Census tract level, yet the STEAP tool is reporting at a more granular level. How does the STEAP tool handle this data?

Some data elements used in the STEAP report are only furnished by the Census at the Census tract level of aggregation. In these cases, we need to be able to say something about the particular phenomenon at the block group level of aggregation, because this is the level of aggregation required for STEAP to generate a report. So for these data elements, STEAP employs a disaggregation method that takes data available at the Census tract level and apportions it across the block groups within each given Census tract based on some basic assumptions.

For most of these data elements, STEAP makes the assumption that the phenomenon is distributed across the block groups within a Census tract proportionally to how the overall population is distributed. The following phenomena are assumed to be distributed across the block groups according to the overall population distribution:

- Nativity
- Household Type For Children Under 18 Years In Households (Excluding Householders, Spouses, And Unmarried Partners)
- Poverty Status In The Past 12 Months By Sex
- Poverty Status In The Past 12 Months By Age
- Age By Disability Status
- Veteran Status For The Civilian Population 18 Years And Over
- Monthly Housing Costs

For some tract-level data elements, however, the Census ACS furnishes related data at the block group level that can provide more detailed context for how tract-level phenomena are likely to be distributed across block groups. For example, we know that the Hispanic and Latino population within a Census tract is *not* evenly distributed across the block groups that fall within the tract. Hispanic and Latino populations may tend to cluster in particular block groups within the tract, and we want to take this into account during our disaggregation. And fortunately, the ACS 5-Year Estimates *do* furnish block group-level information about how the overall Hispanic and Latino population is distributed across block

groups. This information can be taken into account when disaggregating other variables related to the Hispanic and Latino population that are only available at the Census tract level.

Let’s dig into this example further: the “Hispanic or Latino Origin by Specific Origin” table contains variables that are only available at the Census tract level, but this is a data element that we wish to disaggregate to the block group level. Let’s say we want to make a reasonable estimate of how many residents of Mexican heritage are likely located within a specific block group. Here’s the data that is available at the Census tract level of aggregation:

B03001 HISPANIC OR LATINO ORIGIN BY SPECIFIC ORIGIN	
American Community Survey Universe: Total population 2021: ACS 5-Year Estimates Detailed Tables	
Census Tract 1259, Hennepin C...	
Label	Estimate
▼ Total:	5,055
Not Hispanic or Latino	3,063
▼ Hispanic or Latino:	1,992
Mexican	992
Puerto Rican	0
Cuban	78
Dominican (Dominican Republic)	0
▶ Central American:	1
▶ South American:	691
▶ Other Hispanic or Latino:	230

(Data example: Minnesota > Hennepin County > Census Tract 1259; screenshot from [Census Data Explorer](#))

We would not simply want to distribute the count of residents with Mexican heritage according to how the overall population is distributed. Instead, we find a more relevant **control variable** that is available at the block group level to help distribute the tract-level data more accurately. In this case, we use the following control variable, the “Hispanic or Latino” count from the “Hispanic or Latino Origin by Race” table, available for all block groups within the Census tract:

B03002 HISPANIC OR LATINO ORIGIN BY RACE				
American Community Survey Universe: Total population 2021: ACS 5-Year Estimates Detailed Tables				
	Census Tract 1259...	Block Group 1...	Block Group 2...	Block Group 3...
Label	Estimate	Estimate	Estimate	Estimate
▼ Total:	5,055	836	2,026	2,193
▶ Not Hispanic or Latino:	3,063	491	1,608	964
▶ Hispanic or Latino:	1,992	345	418	1,229

Then, to distribute the count of residents with Mexican heritage across the block groups, we perform the following calculations:

- Block Group 1 contains $345 / 1992 = 17.3\%$ of the overall “Hispanic or Latino” population within the Census tract. So, we estimate the count of Mexican-heritage residents within this block group to be: 992 (i.e. the total count of Mexican-heritage residents at the Census Tract level) * $17.3\% = 171.8$ residents.
- Block Group 2 contains $418 / 1992 = 20.9\%$ of the overall “Hispanic or Latino” population within the Census tract. So, we estimate the count of Mexican-heritage residents within this block group to be: $992 * 20.9\% = 208.2$ residents.
- Block Group 3 contains $1229 / 1992 = 61.7\%$ of the overall “Hispanic or Latino” population within the Census tract. So, we estimate the count of Mexican-heritage residents within this block group to be: $992 * 61.7\% = 612.0$ residents.

Several other variables are disaggregated in a similar manner. Here is a list of these variables, along with the control variables used to assist in their disaggregation:

Table	Variable	Control Variable
American Indian And Alaska Native (Aian) Alone Or In Any Combination By Selected Tribal Groupings (B02017)	Total Groups Tallied (B02017_001)	AMERICAN INDIAN AND ALASKA NATIVE ALONE OR IN COMBINATION WITH ONE OR MORE OTHER RACES: Total (B02010_001)
	American Indian tribes, specified (B02017_002)	B02010_001
	Cherokee (B02017_008)	B02010_001
	Chippewa (B02017_011)	B02010_001
	Navajo (B02017_026)	B02010_001
	Sioux (B02017_036)	B02010_001
Hispanic Or Latino Origin By Specific Origin (B03001)	Total (B03001_001)	HISPANIC OR LATINO ORIGIN BY RACE: Total (B03002_001)
	Not Hispanic or Latino (B03001_002)	HISPANIC OR LATINO ORIGIN BY RACE: Not Hispanic or Latino (B03002_002)
	Hispanic or Latino (B03001_003)	HISPANIC OR LATINO ORIGIN BY RACE: Hispanic or Latino (B03002_012)
	Mexican (B03001_004)	B03002_012
	Puerto Rican (B03001_005)	B03002_012
	Cuban (B03001_006)	B03002_012
	Dominican (Dominican Republic) (B03001_007)	B03002_012
	Central American (B03001_008)	B03002_012
	South American (B03001_016)	B03002_012
Other Hispanic or Latino (B03001_027)	B03002_012	
Vehicle Availability (B08201)	Total Households (B08201_001)	HOUSEHOLD TYPE (INCLUDING LIVING ALONE): Total Households (B11001_001)
	No vehicles available (B08201_002)	B11001_001
	1 vehicle available (B08201_003)	B11001_001
	2 vehicles available (B08201_004)	B11001_001
	3 vehicles available (B08201_005)	B11001_001
	4 vehicles available (B08201_006)	B11001_001

Population in Poverty by Race (B17001A-G)	People in Poverty – White Alone (B17001A_002)	RACE: White Alone (B02001_002)
	People in Poverty – Black or African American Alone (B17001B_002)	RACE: Black or African American Alone (B02001_003)
	People in Poverty – American Indian & Alaska Native Alone (B17001C_002)	RACE: American Indian & Alaska Native Alone (B02001_004)
	People in Poverty – Asian Alone (B17001D_002)	RACE: Asian Alone (B02001_005)
	People in Poverty – Native Hawaiian & other Pacific Islander Alone (B17001E_002)	RACE: Native Hawaiian & other Pacific Islander Alone (B02001_006)
	People in Poverty – Some Other Race Alone (B17001F_002)	RACE: Some Other Race Alone (B02001_007)
	People in Poverty – Two or More Races (B17001G_002)	RACE: Two or More Races (B02001_008)

How is the “Population in Poverty by Race” table calculated?

The Census ACS reports the total number of people in each Census tract for whom it is possible to determine their poverty status (B17001_001). This is the “Total” value included in the “Population in Poverty by Race” table in the STEAP report.

The Census ACS data also reports counts of people in poverty in a series of seven separate tables (B17001A-G) that each focus on a specific racial category. This is what one of these raw tables looks like in the Census data explorer, with the relevant data field highlighted below:

B17001B | POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE) ▼

American Community Survey | Universe: Black or African American alone population for whom poverty s... +1

Label	Census Tract 1011.01, Su...	Estimate
▼ Total:		3,253
▼ Income in the past 12 months below poverty level:		220
▼ Male:		104
Under 5 years		0
5 years		0

The count and percent values included in the “Population in Poverty by Race” table in the STEAP report are sourced from these Census tables and computed as follows:

Data Element	Count	Percent
Total (Population for whom poverty status is determined)	B17001_001	N/A
People in Poverty – White Alone	B17001A_002	B17001A_002 / B17001_001 * 100

	POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (WHITE ALONE): Income in the past 12 months below poverty level)	
People in Poverty – Black or African American Alone	B17001B_002	$B17001B_002 / B17001_001 * 100$
People in Poverty – American Indian & Alaska Native Alone	B17001C_002	$B17001C_002 / B17001_001 * 100$
People in Poverty – Asian Alone	B17001D_002	$B17001D_002 / B17001_001 * 100$
People in Poverty – Native Hawaiian & other Pacific Islander Alone	B17001E_002	$B17001E_002 / B17001_001 * 100$
People in Poverty – Some Other Race Alone	B17001F_002	$B17001F_002 / B17001_001 * 100$
People in Poverty – Two or More Races	B17001G_002	$B17001G_002 / B17001_001 * 100$

[How does STEAP deal with margins of error reported by the Census?](#)

The STEAP tool does not address margins of error directly, and margins of error are not included in the STEAP report output. STEAP is intended as a screening tool, and is not intended to be consumed as a statistical or research product. Users should be aware that the counts included in the STEAP report are estimates, and should be interpreted as such. This is particularly important to keep in mind when working with very small buffer areas.