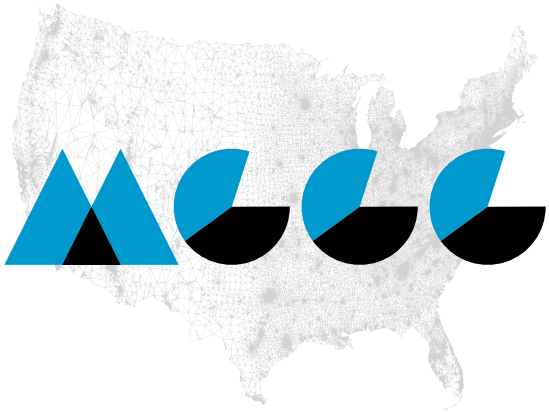
Conventions for Constructing Demographic Categories



Data and Democracy Lab

Abstract

In this note we explain the construction of VAP and CVAP from Census variables, using both the ACS and the Decennial (PL 94-171) as sources.

Contributors

Chris Donnay, Moon Duchin, and Peter Rock compiled this information.

1 Race and ethnicity categories

We take the following convention for dividing up combinations of race and ethnicity.

We choose an order for simple group names as follows: Black, Hispanic, Asian, AMIN, Other, and White.

This means that we include in BPOP or BVAP anyone who answered “yes" to the checkbox question of whether they are Black. In other words, this is what experts sometimes call “Any-Part-Black," or “Black Alone or In Combination" – not what is known as “Black Alone" or “Single-Race Black."

Then, of the people who remain, we take HPOP or HVAP to include as anyone who indicated that they were Hispanic (but were not already counted as Black). Then, of the people who remain, we compute AVAP as anyone who indicated that they were Asian or NHPI (Native Hawaiian/Pacific Islander), so as to correspond to the category usually abbreviated “AAPI." And similarly with AMIN, Other, and White. This has the effect that White encompasses only non-Hispanic, single-race White respondents.

2 Voting age population

Next, we’ll spell out very precisely how voting age population is created as sums of basic Census variables. We draw from the P3 table, which contains the counts of people over the age of 18, i.e., the VAP data. Total population is constructed in an exactly analogous manner using the P1 table in place of P3.

To compute HVAP, we use the P4 table, which contains a breakdown of voting age people that are NOT Hispanic. So to compute HVAP, we subtract the P4 columns from their corresponding P3 columns to get the count of people who are Hispanic. We exclude any columns used in the computation of BVAP, so the P3/P4 columns involved are given in Table 2.

2020 P3 Column	Variable name
p3_004n	black
p3_011n	white_black
p3_016n	black_amin
p3_017n	black_asian
p3_018n	black_nhpi
p3_019n	black_other
p3_027n	white_black_amin
p3_028n	white_black_asian
p3_029n	white_black_nhpi
p3_030n	white_black_other
p3_037n	black_amin_asian
p3_038n	black_amin_nhpi
p3_039n	black_amin_other
p3_040n	black_asian_nhpi
p3_041n	black_asian_other
p3_042n	black_nhpi_other
p3_048n	white_black_amin_asian
p3_049n	white_black_amin_nhpi
p3_050n	white_black_amin_other
p3_051n	white_black_asian_nhpi
p3_052n	white_black_asian_other
p3_053n	white_black_nhpi_other
p3_058n	black_amin_asian_nhpi
p3_059n	black_amin_asian_other
p3_060n	black_amin_nhpi_other
p3_061n	black_asian_nhpi_other
p3_064n	white_black_amin_asian_nhpi
p3_065n	white_black_amin_asian_other
p3_066n	white_black_amin_nhpi_other
p3_067n	white_black_asian_nhpi_other
p3_069n	black_amin_asian_nhpi_other
p3_071n	white_black_amin_asian_nhpi_other

Table 1. 32 variables that are summed to compute BVAP.

2020 P3 and P4 Column Calculation	Variable name
p3_003n – p4_005n	white
p3_005n – p4_007n	amin
p3_006n – p4_008n	asian
p3_007n – p4_009n	nhpi
p3_008n – p4_010n	other
p3_012n – p4_014n	white_amin
p3_013n – p4_015n	white_asian
p3_014n – p4_016n	white_nhpi
p3_015n – p4_017n	white_other
p3_020n – p4_022n	amin_asian
p3_021n – p4_023n	amin_nhpi
p3_022n – p4_024n	amin_other
p3_023n – p4_025n	asian_nhpi
p3_024n – p4_026n	asian_other
p3_025n – p4_027n	nhpi_other
p3_031n – p4_033n	white_amin_asian
p3_032n – p4_034n	white_amin_nhpi
p3_033n – p4_035n	white_amin_other
p3_034n – p4_036n	white_asian_nhpi
p3_035n – p4_037n	white_asian_other
p3_036n – p4_038n	white_nhpi_other
p3_043n – p4_045n	amin_asian_nhpi
p3_044n – p4_046n	amin_asian_other
p3_045n – p4_047n	amin_nhpi_other
p3_046n – p4_048n	asian_nhpi_other
p3_054n – p4_056n	white_amin_asian_nhpi
p3_055n – p4_057n	white_amin_asian_other
p3_056n – p4_058n	white_amin_nhpi_other
p3_057n – p4_059n	white_asian_nhpi_other
p3_062n – p4_064n	amin_asian_nhpi_other
p3_068n – p4_070n	white_amin_asian_nhpi_other

Table 2. 31 compound variables that are summed to compute HVAP. For each variable, we take the P3 value and subtract the P4 value, since P4 counts people who are NOT Hispanic.

	2020 P4 Column	Variable name
AVAP	p4_008n	asian
	p4_009n	nhpi
	p4_015n	white_asian
	p4_016n	white_nhpi
	p4_022n	amin_asian
	p4_023n	amin_nhpi
	p4_025n	asian_nhpi
	p4_026n	asian_other
	p4_027n	nhpi_other
	p4_033n	white_amin_asian
	p4_034n	white_amin_nhpi
	p4_036n	white_asian_nhpi
	p4_037n	white_asian_other
	p4_038n	white_nhpi_other
	p4_045n	amin_asian_nhpi
	p4_046n	amin_asian_other
	p4_047n	amin_nhpi_other
	p4_048n	asian_nhpi_other
	p4_056n	white_amin_asian_nhpi
	p4_057n	white_amin_asian_other
	p4_058n	white_amin_nhpi_other
	p4_059n	white_asian_nhpi_other
	p4_064n	amin_asian_nhpi_other
	p4_070n	white_amin_asian_nhpi_other
AMINVAP	p4_007n	amin
	p4_014n	white_amin
	p4_024n	amin_other
	p4_035n	white_amin_other
OVAP	p4_010n	other
	p4_017n	white_other
WVAP	p4_005n	white

Table 3. AVAP (24 variables), AMINVAP (4 variables), OVAP (2 variables), and WVAP (1 variable). We use the P4 variables so that we do not count anyone who was labeled Hispanic.

3 Citizen voting age population via “discounting"

The ACS reports CVAP at the tract level in the primary tables, which is a much higher level of aggregation than the block data in the Decennial release. There are various special tabulations that let you put 5-year ACS data with certain fairly coarse racial categories onto block groups, but the categories do not match up with the Decennial data and the race schema outlined above is not possible with ACS data.

We often need to conduct analysis of CVAP at the block level. For many of our data projects, we use the following computation, which uses ACS at the tract level to estimate the citizenship rate of various broad racial/ethnic groups, then applies that “discount" (percentage reducing the total) to VAP, in the categories described and calculated above.

Suppose block X belongs to tract Y in state Z, and we want to estimate BCVAP. Take the citizenship rate for the group in the tract and apply it to the voting age population in the block:

$$\frac{\text{BCVAP}_{\text{ACS}}(\textit{tract } Y)}{\text{BVAP}_{\text{ACS}}(\textit{tract } Y)} \cdot \text{BVAP}_{\text{PL}}(\textit{block } X) = \text{BCVAP}(\textit{block } X).$$

If the denominator in a calculation like this is too low (we have chosen the admittedly arbitrary threshold of **20**), then we swap in the citizenship rate in the state since the tract did not allow for a meaningful estimate:

$$\frac{\text{BCVAP}_{\text{ACS}}(\textit{state } Z)}{\text{BVAP}_{\text{ACS}}(\textit{state } Z)} \cdot \text{BVAP}_{\text{PL}}(\textit{block } X) = \text{BCVAP}(\textit{block } X).$$

Note to readers: the last part is clearly a crude fix when the local signal is weak. A more sophisticated Bayesian approach that borrows statistical power from neighboring tracts is coming soon, via work of Duchin–Schein–Wolfram.

It’s important to note that the race categories available in the ACS data are different—not as granular—as the categories in the Decennial. So where the PL categories of Black, Hispanic, Asian, AMIN, and White are constructed as described above, these are instead treated as simple (not summed) categories in the ACS.[1]

Finally, the total CVAP in a block is computed as the sum of the estimated CVAP in the five categories. This is reasonable because each of the five kinds of CVAP is computed as a sum of VAP, over mutually disjoint groups.

[1]Non-Hispanic members of the “Some Other Race" category in the Decennial data are rare, and their citizenship rate is likely not well represented by the Other rate in the ACS, because that includes many Hispanic respondents. Therefore, in our ACS calculation, we fold Other in with White to make five categories rather than six.