Details About the Variables, Their Coding, and Uses

SIMPLEDENSITY

SIMPLEDENSITY is the population of a given PUMA divided by its land area, both of which are standard variables in the Census Bureau's databases aggregated by PUMA. It represents the average population density across the entire PUMA.

GEODENSITY

GEODENSITY is an index for expressing the concentration of population within a PUMA developed by Jonathan P. Schroeder and José D. Pacas of the University of Minnesota's Institute for Social Research and Data Innovation. It is needed because of the wide variation in population densities that can exist within a single PUMA. Rather than use the average density across the entire PUMA, it can be more revealing to use the average of the densities of smaller units, weighted by population. GEODENSITY operationalizes this insight. It represents the population-weighted geometric mean of the PUMA's census tract population densities—that is, it measures the mean density of census tracts on a logarithmic scale. The use of logged values is prompted by the severely skewed distribution of census tract densities. The algorithm for computing GEODENSITY is

$$d_{GM} = \exp\left(\frac{\sum P_i \log d_i}{\sum P_i}\right)$$

where d_{GM} is the population-weighted geometric mean and P_i and d_i are respectively the population and density of the census tracts in a given PUMA.

The difference between the interpretation of the two density variables is best described with a concrete example. The Arizona PUMA numbered 700 (AZ-700) consists of Yuma County. AZ-700 is huge—5,519 square miles, larger than Connecticut—and has an unweighted population density of just 38 people per square mile. SIMPLEDENSITY gives you an accurate image of a PUMA that consists overwhelmingly of sparsely populated land. But 93 percent of the 209,658 people who live in AZ-700 live in the city of Yuma and its suburbs, which occupy just 3.9 percent of the PUMA's land area. The geometric mean of 1,358 gives you an accurate indicator of the population density that is experienced by almost all the residents of AZ-700, and that's usually the indicator of interest when you are using the individuals in the ACS as your unit of analysis.

Almost all the large differences between the two measures are found in rural PUMAs. For the 728 PUMAs in which GEODENSITY was more than twice as large as SIMPLEDENSITY, the mean

SIMPLEDENSITY was just 449 people per square mile. For practical purposes, SIMPLEDENSITY and GEODENSITY can be used interchangeably for PUMAs that are located entirely within large cities.

METPOP10

METPOP10 is an index for expressing the size of the commuting population where PUMA households are located based on 2010 decennial census data, also developed by Schroeder and Pacas. The "10" in the label distinguishes it from METPOP00, a parallel variable in the IPUMS-USA ACS based on the 2000 decennial census data.

METPOP10 is operationalized as the population-weighted mean of the MSAs associated with each PUMA. The algorithm is

$$P_{GMj} = \exp\left(\frac{\sum P_{ij} \log P_i}{\sum P_{ij}}\right)$$

where P_{GMi} is the populated-weighted geometric mean of the populations of CBSAs and noncore counties in PUMA *j*, P_i is the population of the CBSA or noncore county *i*, and P_{ij} is the population in the area of intersection between *i* and *j*.

For the 78 percent of PUMAs that lie within a single CBSA, the value of METPOP10 is the same as the total population of the CBSA's population (e.g., the value of METPOP10 for a PUMA containing a suburb of Dallas is the same as the entire population of the Dallas-Fort Worth-Arlington CBSA). Elsewhere, METPOP10 summarizes the sizes of all CBSAs associated with the PUMA. For counties that are part of a PUMA but not part of any CBSA, METPOP10 uses the county's population as an approximation of the commuting system size.

Schroeder and Pacas use the geometric mean because commuting system populations, like census tract densities, have a roughly log-normal distribution. Using the geometric mean in effect represents relative differences in populations rather than absolute differences. To see why this is appropriate, think about the population difference between two cities of 100,000 and 200,000 people. The latter city is twice the size of the former. Now think about the same absolute difference in population—100,000—if the total populations of two cities are 1,000,000 and 1,100,000. The latter city is just 10 percent larger than the former. Increments of a given size should not have the same weight across all city populations.

For a complete description of GEODENSITY and METPOP10 and their many potential uses, see Jonathan P. Schroeder and José D. Pacas (2021) "Across the Rural–Urban Universe: Two Continuous Indices of Urbanization for U.S. Census Microdata," *Spatial Demography*, 9(1), 131-154. doi:10.1007/s40980-021-00081-y.

PUMACITY

The sources for the PUMACITY codes are the Census Bureau's PUMA maps, Google Maps, and the variable labeled PUMA.

The Census Bureau provides a description of each PUMA, given in the variable PUMANAME. These descriptions fall into three categories: ones that includes the name of a town or city, ones that list the counties in the PUMA, and ones that describe a region or parts of counties.

When the description includes the name of a town or city. These PUMAs present the easiest coding decisions. An example is PUMA CA-6103: "Placer County (East/High Country Region)—Auburn & Colfax Cities." The first-named place in PUMANAME is almost always the largest town or city in the PUMA. There are exceptions, but the default coding decision was to use the first-named place as PUMACITY—in this case, Auburn CA.

When the city is a major metropolis, PUMANAME includes a description of neighborhoods in the PUMA, a directional descriptor of the PUMA, or both. An example of a neighborhood description without a directional descriptor is NY-3708: "NYC-Bronx Community District 4—Concourse, Highbridge & Mount Eden." In a case like this, PUMACITY consists of the first two named neighborhoods and a directional descriptor based on the map of the city in Google Maps. Thus PUMACITY for NY-3708 is coded "SW Bronx (Concourse, Highbridge)."

An example of a PUMA label with a directional descriptor but no neighborhood descriptor is TX-2311, "Dallas City (North)—South of I-635 & West of US-75." The Census Bureau's map of TX-2311 was used to determine the borders of the PUMA and Google Maps was used to look for a meaningful neighborhood descriptor. It turns out that TX-2311 contains Preston Hollow, a well-known affluent Dallas neighborhood. The coding of PUMACITY is therefore "N Dallas (Preston Hollow)."

Most of the directional abbreviations used in coding PUMACITY—N, S, E, W, NW, NE, SW, SE—are self-explanatory. The descriptors "Central" and "Downtown" in PUMANAME are abbreviated as "C" in PUMACITY. When an abbreviation has two descriptors, the first one indicates the section of the city; the second indicates the part of that section. Thus "EC" refers to the central part of the eastern section of the city, while "CE" refers to eastern part of the central section of the city. An abbreviation such as "E & N" denotes "east and north." A direction that is spelled out (e.g., "West Dallas") is an accepted name for a district or neighborhood, not a directional descriptor.

PUMAs that were identified in the official description as "Central" or "Downtown" without other qualification in a city of at least 250,000 people are coded as "Central <cityname>" to make it easy to select a set of PUMAs containing the center of a major city.

When the description lists the counties that make up the PUMA. For PUMAs outside major metropolitan areas, the state agencies that define the PUMAs commonly use a set of entire counties. An example is Indian PUMA IN-3000, "Jefferson, Jennings, Decatur & Scott Counties." These PUMAs are also easy to code for PUMACITY using the Census Bureau datafile that gives the county (or counties, for cities that overlap county borders) for every town, city, or other Census Designated Place (CDP) in the nation. When the datafile is sorted by state, county, and population of the town or city, identifying PUMACITY is a straightforward matter of determining the largest town or city in each county and choosing the largest among them. The largest town in the four counties that make up IN-3000, and hence PUMACITY, turns out to be Greensburg IN, population 12,312.

Note that the datafile used to identify the largest place in a PUMA when coding PUMACITY during the fall of 2021 was based on estimated 2020 populations. The actual 2020 population figures from the 2020 Decennial Census became available toward the end of this process, and they have been substituted for all values of PUMAPOP, PUMACITYPOP, PRINCITYPOP, and CBSACITYPOP. It is possible that in a few cases the place that was the largest in a PUMA using the estimated 2020 populations is not the same as the place that was the largest in the actual 2020 figures.

When the description contains a region and/or parts of counties. Several hundred PUMAs are in this category. An example of a regional descriptor is GA-402, "Coastal Regional Commission (East)—Chatham County (East & Outside Savannah City)." If whole counties are involved, the identification of PUMACITY is as easy as a PUMANAME that list the counties. When parts of counties are involved, coding PUMACITY became more demanding. For GA-402, only the eastern part of the county outside Savannah is part of the PUMA. In such cases, it is necessary to look up the population of each of the towns in the specific part of the county or counties included in the PUMA to determine PUMACITY.

Coding PUMACITYPOP in states that recognize townships. The objective of PUMACITYPOP is to reflect the urban population of towns and cities. This is complicated by the 20 states that recognize the township as a geographic unit. In most of those states, the township is a much larger entity than the town located within it. The distortions this produces can be important, especially in New York and Illinois. In New York's Suffolk County, for example, the village called Babylon has an actual population of 12,188. Babylon township has a population of 209,785. PUMACITYPOP reports the town population instead of the township population. Approximations were necessary for a few places in Connecticut for which no separate population figure for the town within a township could be identified. It is also possible that some errors remain—the "places" defined by the Census Bureau are sometimes ambiguous. For a description of the differences among the different communities that the Census Bureau includes under the label "place," See Michael Ratcliffe,

Understanding "Place" in Census Bureau Data Products, available for download at www.census.gov/content/dam/Census/data/developers/understandingplace.pdf.

PRINCITY

The codes for PRINCITY are the officially designated Principal Cities for the MSAs in the Census Bureau inventory. A Principal City was associated with a PUMA under one of three conditions.

1. The Principal City is also PUMACITY or immediately adjacent to PUMACITY.
2. The PUMA does not contain a Principal City, but a Principal City borders on PUMACITY in a contiguous urban area. "Urban area" is operationally defined as within the legal borders of a Census Designated Place or an adjacent residential or commercial area that is colored gray on Google Maps. If the Principal City within the PUMA is a micropolitan area and is contiguous with the urban area of the Principal City of a metropolitan area in an adjacent PUMA, PRINCITY is coded as the Principal City for the metropolitan area.
3. The PUMA is in a large contiguous urban area that includes more than one Principal City but is not adjacent to any of them. In these cases, PRINCITY is the closest of the Principal Cities.

If the PUMA is not associated with a Principal City under one or more of these three conditions, PRINCITY is coded "No Principal City."

One exception to the coding rules for PRINCITY involves towns that still had a few miles of open country separating them from the contiguous urban area of a large city in Google Maps as of January 2022 but showed rapid growth since 2010 in the newly released 2020 population numbers. All PUMAs that had originally been coded as "No Principal City" that experienced 2010–2020 population growth of 10 percent or more were examined on Google Maps. If a Principal City was within a few miles of the contiguous urban area in such cases, it was assigned as PRINCITY to the PUMA in question under the assumption that the green space separating PUMACITY from the Principal City will have disappeared when Google Maps is updated with the 2020 numbers. The most conspicuous example of this rapid change is the corridor from San Antonio to Austin in Texas, which appears to be in the process of becoming a contiguous urban area along the entire 60 miles from the northern suburbs of San Antonio to the southern suburbs of Austin.

The other exception to the coding rules for PRINCITY involves the CBSA labeled "New York-Newark-Jersey City." It is unlike any of the others in its lack of differentiation over an extremely broad geographic area that encompasses the New York City's five boroughs, all of Long Island, Westchester and Rockland counties, and all of northern New Jersey. It has just six MSAs and hence six codes for PRINCITY to subdivide its 19.6 million people. By way of comparison, The CBSA

labeled "Los Angeles-Long Beach-Anaheim" with 12.8 million people, has 18 Metro Areas supplemented by another 8 for the 4.2 million in the adjacent "Riverside-San Bernardino-Ontario" CBSA.

To make PRINCITY for the New York CBSA more informative, PRINCITY is coded with "NY" plus the name of the borough if it is located within New York City. PRINCITYPOP is the population of the borough. For PUMAs in the New York CBSA outside of the five boroughs, PRINCITY is the name of the county and PRINCITYPOP IS THE SUM OF THE POPULATIONS OF PUMAS IN THAT COUNTY.

CBSACITY

CBSACITY is the name of the most populous city in the CBSA associated with a PUMA. For example, CBSACITY for the CBSA officially labeled "Los Angeles-Long-Beach-Anaheim" is Los Angeles. CBSACITYPOP is the population of CBSACITY.

Three exceptions were made for CBSAs that cover such a long linear distance that PUMAs could be much closer to a second-largest major city in the CBSA than to the largest city. The three CBSAs that fit this description were San Francisco-Oakland-Berkeley, Miami-Fort Lauderdale-Pompano Beach, and Dallas-Fort Worth-Arlington. Oakland was coded as CBSACITY for PUMAs in the contiguous urban areas on the east side of San Francisco Bay. Fort Lauderdale was coded as CBSACITY for PUMAs in the contiguous urban areas of Broward and Palm Counties. Fort Worth was coded as CBSACITY for PUMAs in the contiguous urban areas of Tarrant County.

If examination of the PUMA map revealed no Principal City, the map of CBSAs in Social Explorer was used to determine if PUMACITY fell within a CBSA. If yes, the largest city in that CBSA and its population were coded for CBSACITY and CBSACITYPOP. When PUMACITY fell outside any CBSA, Social Explorer's map of Metro Areas was superimposed on its map of PUMAs to determine whether a significant portion of the PUMA other than PUMACITY was associated with a CBSA. If yes, the largest city in that CBSA and its population were coded for CBSACITY and CBSACITYPOP. PUMAs that failed to show a link to a CBSA by either test are coded "Not in a CBSA."

ALTMETRO

A longstanding problem with CBSAs defined by the Census Bureau is that they can include so many different environments, from rural to densely urban. For example, the Washington DC Metro Area stretches from downtown Washington more than 80 miles northward and 70 miles westward. For the last 50 miles in both directions, the suburbs have been left behind, replaced by farm country with an occasional town and no big cities.

Knowing PUMACITY in combination with GEODENSITY can reliably identify where a PUMA stands at the extremes. If the population of PUMACITY is 15,000 and GEODENSITY is less than 100, you can

be sure that the PUMA is overwhelmingly rural; if PUMACITYPOP is 700,000 and GEODENSITY is more than 10,000, you can be sure that the PUMA is in the heart of a metropolis. But between those extremes, the combinations of PUMACITYPOP and GEODENSITY can represent very different rural/urban environments.

The ACS files available at IPUMS-USA include a variable called METRO that at first glance would seem to be useful in distinguishing different types of PUMAS, but almost half of the individuals are coded as "Metropolitan status indeterminable (mixed)" or "In metropolitan area: Central/principal city status indeterminable (mixed)," effectively uninterpretable. But even a code of "In metropolitan area: In central/principal city" doesn't tell you whether a person is at the urban center of the city or in its residential periphery. A code of "In metropolitan area: Not in central/principal city" doesn't tell you whether a person lives in a nearby suburb of the Metro Area's Principal City, a distant stand-alone town, or in a rural part of the Metro Area. ALTMETRO is intended to be a more useful version of METRO.

The initial step in assigning codes to ALTMETRO was to identify candidate PUMAs for each category, then examine ambiguous cases individually. The specific procedures used for each of the five codes are as follows.

1. *"Rural."* An agricultural or otherwise sparsely populated PUMA with a largest place of fewer than 20,000 people that is not contiguous with another city.

   The pool of candidate PUMAs was identified by selecting all PUMAs with PUMACITYPOP<20,000 & (PRINCITY=PUMACITY OR PRINCITY="NO PRINCIPAL CITY" OR CBSACITY="NOT IN A CBSA"). The algorithm identified 421 candidate PUMAs. Upon examination of those with the highest population densities, the candidate PUMAs were restricted to those with both GEODENSITY and SIMPLEDENSITY less than 1000. This left 395 candidate PUMAs. These were examined case by case to determine the proximity of PUMACITY to other towns or cities, recoding the candidate PUMAs as "Rural," "Town," or "Satellite" based on the results of the case-by-case examination. This left 304 PUMAs/

   The objective of the "Rural" category was to identify PUMAs in which people in the largest town are typical for small towns in a rural area. The names of the PUMACITYs coded "Rural" were examined, looking for ones that are unusually well-known. This led to recodes of the PUMAs containing Augusta ME (Maine's capital city) and Williamsburg VA (home of William & Mary University and a well-known tourist attraction) as "Town" despite meeting the other criteria for "Rural." Other tourist destinations recoded as "Town" were Steamboat Springs CO, Branson MO, Kill Devil

Hills NC, and Pinehurst NC. Los Alamos NM, a uniquely cosmopolitan small town, was also coded "Town" instead of "Rural." The net number of PUMAs classified as "Rural" was 298.

2. *"Town."* PUMAs with a largest place of fewer than 50,000 people that is not contiguous with another town or city."

3. *"Small City."* PUMAs with a largest place with 50,000–149,999 people that is not contiguous with another city.

   The categories "Town" and "Small City" have a common purpose for places with different population ranges: to identify PUMAs that did not qualify as "Rural" but had a largest place—PUMACITY—that is a stand-alone community, with not only an independent identity as a Census Bureau "place" but also an independent identity socially and economically.

   The candidate PUMAs were identified by this logic: If PUMACITY is also the Principal City of the PUMA, or if there is no Principal City associated with the PUMA, then by definition PUMACITY cannot be the satellite of a larger city unless it is in the contiguous urban area of a larger CBSACITY. Thus the candidates with the appropriate population ranges for "Town" and "Small Cities" had to meet the criterion:

   (PUMACITY=PRINCITY or PRINCITY="No Principal City") & ALTMETRO≠"Rural"

   Applying these criteria identified 481 PUMAs as candidates for "Town" and 339 candidates for "Small City."

4. *"Satellite."* PUMAs with a largest town or city that is a suburb or satellite to a larger city.

   A PUMA in which the largest town or city is associated with another and larger place that is a Principal City of a CBSA must be a satellite of that Principal City. Candidates for classification as "Satellite" were identified using the criterion PUMACITY ≠ PRINCITY, with the qualifier that the entire PUMA was not embedded in a large city. This produced a candidate pool of 701 PUMAs.

5. *"Core City."* PUMAS with a city of 150,000 or more that is either the largest city in the CBSA or larger than any other contiguous town or city.

   All PUMAs with a stand-alone city of at least 150,000 population are potentially Core Cities. The problem is to determine which cities of that size are actually satellites to an even larger contiguous city.

   Applying the criterion PUMACITYPOP>=150,000 identified 577 PUMAs as candidates for "Core City."

These successive steps left 45 PUMAs that that qualified both as a candidate for both "Satellite" and as "Core City" because the largest city in the PUMA had a population of 150,000 or more but it was in a contiguous urban space with a larger Principal City in another PUMA. All of these were coded as "Satellite."

Two anomalous cases were PUMAs with a city larger than 150,000 people that is not associated with a Principal City. They are Lancaster and Palmdale CA, adjacent cities in adjacent PUMAs that fall within the Los Angeles CBSA but are unambiguously separated from Los Angeles's contiguous urban area, lying on the other side of the mountains to the north of Los Angeles. Curiously, neither has been designated as a Principal City of a Metro Area. They have very similar populations—155,822 and 150,498. Together, they form a substantial city of more than 300,000. For lack of a better alternative, Palmdale was coded as "Core City" and Lancaster as "Satellite."

Preparing the Merged Database

Preparing a database that combines the variables in *PUMA Descriptors* with an ACS datafile of individuals has four steps.

1. Download an ACS dataset from IPUMS USA with the variables about individuals that you want to use. Include the variables for the state FIP code and the PUMA number. In the IPUMS USA ACS files, these variables are named STATEFIP AND PUMA RESPECTIVELY. RENAME PUMA TO PUMANUM.
2. Merge the ACS dataset with *PUMA Descriptors* using its variables labeled STATEFIP and PUMANUM.
3. Download an ACS dataset that has summary statistics for PUMAs on the demographic and socioeconomic variables you want to use (e.g., number of people with college degrees, median family income, or racial composition of the PUMA). Either Social Explorer or IPUMS NHGIS can be used for this purpose. Include variables with the state FIP code and the PUMA number. In Social Explorer, these variables are included automatically included the download. Their names in Social Explorer files are STATE and PUMA5 respectively. Rename them to STATEFIP and PUMANUM.
4. Merge the datafile created in step #3 with the ACS file of individuals created in step #2 using STATEFIP and PUMANUM.