

Normalizing Census Data Using ArcMap

By George Dailey, ESRI K–12 Education Program Manager

Editor’s note: A common mistake made by people new to mapping is comparing areas based on a count statistic such as the number of people who fall into a particular category (e.g., marital status, age, ethnicity). This is not a very meaningful analysis because areas are usually arbitrary in size, and larger areas typically will have more people. Normalizing data factors out the size of areas by transforming counts (measures of magnitude) into ratios (measures of intensity).

Ratio maps can be quickly designed in ArcMap. To normalize data in ArcMap, select a field to map (numerator) and a field to standardize against (denominator). ArcMap creates a proportion by performing simple division and maps that proportion.

An attribute can be normalized in ArcMap using two methods. In the first method, the attribute value for one feature is divided by the sum of that attribute value for all features, turning the resulting ratio values into a percent of the total. See Figure 1 for the formula and an example of this method.

$$\frac{\text{Attribute value for feature } x}{\text{Sum of attribute values in all features}} = \text{Percent of total contained in feature } x$$

$$\frac{15 \text{ persons of Hispanic origin (in } x\text{)}}{450 \text{ persons of Hispanic origin (in total)}} = 0.333 = 3.33 \text{ percent of total}$$

Figure 1: Normalize an attribute as a percent of the total.

In the second method, the values for an attribute are normalized by the values in another attribute that is the universe upon which the first attribute is based or is a member. See Figure 2 for the formula and an example.

$$\frac{15 \text{ persons of Hispanic origin (in } x\text{)}}{300 \text{ total persons (in } x\text{)}} = 0.05 = 5 \text{ percent of the population in } x \text{ is Hispanic}$$

$$\frac{\text{Attribute value for feature } x}{\text{Universe value for feature } x} = \text{Proportion (percentage) of universe that is the attribute}$$

Figure 2: One attribute normalized by another attribute

Data can also be normalized using an attribute that associates with other attributes. The demographic concept of sex ratio (i.e., number of males per 100 females) is an example of this method and illustrated in Figure 3.

$$\frac{135 \text{ males (in } x\text{)}}{165 \text{ females (in } x\text{)}} = .818 \text{ or about } 82 \text{ men for every } 100 \text{ women}$$

Figure 3: Normalizing an attribute that associates with other attributes

Adding a Time Element

Attributes can combine to show influence or change over time, such as population at one time versus population at another time. When the numerator and denominator are the same, the result is one, which indicates there is no change over time. Values above one indicate positive change, and values below one indicate negative change. The example in Figure 4 shows both types of change.

$$\frac{300 \text{ persons in } x \text{ (in } 1990\text{)}}{250 \text{ persons in } x \text{ (in } 1980\text{)}} = 1.2 \text{ 1990 population is } 1.2 \text{ times that of } 1980, \text{ or a } 20 \text{ percent change}$$

$$\frac{250 \text{ persons in } x \text{ (in } 1980\text{)}}{400 \text{ persons in } x \text{ (in } 1970\text{)}} = 0.625 \text{ 1980 population is } 0.625 \text{ that of } 1970, \text{ or a } -37.5 \text{ percent change}$$

Figure 4: Positive and negative change over time

Normalizing Data in ArcMap

Data can be more normalized directly from the ArcMap standard interface. In an ArcMap document, right-click on the layer with the attribute data that will be normalized and choose Properties from the context menu.

1. In the Layer Properties dialog box, click on the Symbology tab and select Quantities.
2. In the Field section, choose the field that will be used as the numerator from the Value drop down. Using the previous example, this would be the Hispanic population.
3. From the Normalization drop down, choose the field that will be used as the denominator—in this case, Total Population.

Apples and Oranges

Know data before normalizing it. Normalizing unrelated data is like mixing apples and oranges. It makes fruit salad, not good analysis. Normalizing data is a powerful tool for map display. However, normalizing can be inappropriately applied if the data being mapped isn’t well understood. It is easy to build compellingly accurate but completely false ratio maps using the sociodemographic data compiled from the United States decennial census of population and housing.

While the construction of a ratio map based on the percent of total is fairly difficult to misapply, creating a map around a median, average, or statistical value would not be appropriate. The application of one attribute value against another has many possible erroneous associations. See Figure 5 for an example using data on people of Hispanic origin that illustrates an appropriate ratio that uses an attribute and its universe.

$$\frac{\text{Hispanic pop.}}{\text{Total pop.}} = \frac{\text{Attribute}}{\text{Universe}} = \frac{\text{Numerator}}{\text{Denominator}} = \frac{\text{Classification field}}{\text{Normalize by}}$$

Figure 5: An appropriate ratio using an attribute and its universe

Universes and Units

It is easy to look at an attribute table and see all kinds of potential associations, but these associations may not be valid. For instance, looking at the single-family houses (one unit) field in data on Hispanic populations might lead to the creation of a map intended to show the ratio of Hispanics living in single-family homes. Unfortunately, because the wrong universe was used, the resulting set of values is meaningless.

$$\frac{\text{Hispanic population}}{\text{Single-family homes}} = \frac{\text{Attribute}}{\text{Wrong universe}} = \text{Nonsense}$$

Figure 6: Inappropriate normalizing field

While normalizing data is conceptually easy, it can prove challenging to apply correctly and ArcMap will not stop a user from creating inappropriate equations. It is important to become familiar with a data item's universe. The universe is the value or population that forms the base from which the data item in question is a subset. For example, when creating the proportion of persons aged five to nine years old across various geographic entities, the universe is the total population.

Know data before normalizing it. Normalizing unrelated data is like mixing apples and oranges. It makes fruit salad, not good analysis.

Intelligent Analysis

While it may seem possible, even reasonable, to create a ratio of anything simply because the data is present in an attribute table, without considering what is being mapped, the result will likely be garbage if the data does not warrant the association.

Mapping Hispanic Americans who live in single-family homes, the example previously cited, illustrates the concepts of unit of analysis and levels of summarization associated with the data. Generally, data available from the Census Bureau has been summarized or blended to some level of geography—census block, census tract, city, county, state, or other geographic unit—removing access to individual responses from individual households or persons. With aggregated data, the unit of analysis is the county, state, or other geographic unit. This protects the confidentiality of individual responses. When attempting to create a cross tabulation of aggregated data using the ArcMap normalizing function, the math will work but the answer will not be what was intended.

A proper cross tabulation requires data at the atomic (i.e., non-aggregated) level. With census data, this means working with data for individual households and persons. The only Census Bureau source for data at this level is the Public Use Microdata Samples (PUMS). PUMS does not include name and specific location information but does allow for state, county, and other higher-order geographic and cross-tabulation analyses. Only bureau employees have access to specific data on individuals for small geographic areas such as census blocks. PUMS data would allow analysis of Hispanic Americans who live in single-family houses but not at very low levels of geography.

What Fields Are Reasonable?

What census data fields are reasonable to use in normalization? Using summarized data from the Census Bureau, there are numerous normalizing associations that can be made. The accompanying table lists classification field-normalizing field pairs for a range of census data items. It also represents many standard data items found in data products available directly from the Census Bureau and those available from ESRI and ESRI Business Information Solutions (ESRI BIS). The normalizing pairs are presented with common names. These may or may not match up with field names in individual data tables. It is important to become familiar with the composition of the data table being used by investigating its data documentation (also known as metadata).

Classification Field	Normalize by
Age, gender, national origin	Total population (persons)
Marital status	Population over age 15
Household composition (e.g., one-person households)	Total households
Contract rent	Renter-occupied housing units
Housing value	Owner-occupied housing units
Household income	Total households
Labor force	Population over age 16

Figure 7: Field-normalizing field pairs for census data

Sources	URL
Gateway to Census 2000	www.census.gov/main/www/cen2000.html
Introduction to Census 2000 Data	www.census.gov/dmd/www/products.html
Population and Housing Definitions	factfinder.census.gov/home/en/eps/glossary_a.html
PowerPoint Presentations	www.census.gov/mso/www/pres_lib/index2.html
Census 2000 Information	www.esribis.com/data/census2000.html
<i>Unlocking the Census with GIS</i>	www.esri.com/esripress

Figure 8: Resources for learning more information about using census data.