

GIS: Normalizing Census Data in ArcView

Summary

Normalizing data using ArcView's Legend Editor is a quick way to design ratio maps. Census data for a range of geographic areas (tracts, counties, states) provides a rich source of these mapping opportunities. This document is a guide to the concept of normalizing with ArcView using census data for examples. It offers background on some key data concepts, possible pitfalls, and suggestions on many "normalizing data pairs."

Materials

ArcView or ArcVoyager software

Instructional Procedures

Normalizing Census Data in ArcView: Concepts and Roadmap

Normalizing data using ArcView's Legend Editor is a quick way to design ratio maps. By normalizing data (an attribute, a field) using this Legend Editor function, you are saying that you want to create a proportion or ratio and have it display in a map. At its essence, you are simply selecting a field to map (or establish as the numerator) and then a field to standardize against (or a denominator). You are creating a proportion (percentage) by having ArcView perform simple division.

In ArcView you can normalize an attribute in one of two ways:

By the sum total of the attribute's values, turning the resulting ratio values into a percent of the total.

By the values in another attribute, where generally that other attribute is the universe upon which the first attribute is based or is a member.

Some examples will help.

As a percent of total, consider the following

$$\frac{\text{Attribute value for feature x}}{\text{Sum of attribute values in all features}} = \text{Proportion (\% of total contained in feature x)}$$
$$\frac{15 \text{ persons of Hispanic origin in x}}{450 \text{ persons of Hispanic origin (in total)}} = 0.333 = 3.33\% \text{ of total}$$

As one attribute normalized by another, consider this:

$$\frac{\text{Attribute value for feature x}}{\text{Universe value for feature x}} = \text{Proportion (5) of universe that is the attribute}$$
$$\frac{15 \text{ persons of Hispanic origin (in x)}}{300 \text{ total persons (in x)}} = 0.05 = 5.0\% \text{ of the population in x is Hispanic}$$

Also consider normalizing using attributes that associate with other attributes, for instance the demographic concept of sex ratio (number of males per 100 females)

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

Or, attributes that can combine to show influence or change over time, such as population at one time versus population at another time. (NOTE: A resultant value of one [1] means that the numerator and denominator are the same. In the case of analyzing population change such a value would mean that population has not changed across the time period. Therefore, values

above one [1] indicate positive change and those below one [1] negative change.)

300 persons in x (in 1990)			
250 persons in x (in 1980)			= 1990 pop is 1.2 times that of 1980, or a 20% change
250 persons in x (in 1980)			
400 persons in x (in 1970)	=		1980 pop. is 0.625 that of 1970, or a -37.5% change

Know your data before normalizing: Mixing apples and bananas equals fruit cocktail. The examples help point out the power of normalizing data for map display. However, they also imply the extent to which this function can be inappropriately applied because of a misunderstanding about the data being mapped. One of the richest arenas for the construction of compellingly accurate and completely false ratio maps is the range of sociodemographic data available from the U.S. decennial census of population and housing.

While the construction of a ratio map based on the percent of total is fairly difficult to misapply (however creating one around a median, average, or statistical value would not be appropriate), the application of one attribute's values against another's is ripe with possibilities for erroneous associations.

Going back to the Hispanic origin example, an appropriate ratio makes use of an attribute and its universe.

Hispanic pop	=	Attribute	=	Numerator	=	Classification field
Total pop		Universe		Denominator		Normalize by

It is just as easy to look through the data in the table you are mapping and begin to see potential associations to create. For instance, you see Hispanic population in the table and you see single-family houses (one unit). A quick reaction might be to create what you think is a map of the ratio of Hispanics living in single-family homes. Unfortunately, the reaction is wrong. You will end up with a set of values but they will be meaningless.

Hispanic pop	=	Attribute	=	Nonsense
Single-family homes		Wrong universe		

The concept of normalizing may be easy but the use can prove challenging since ArcView will not stop you from creating inappropriate equations. It is up to you to make that determination. Understanding data to normalize means understanding universes and units of analysis. In working with Census Bureau data for example, it is important to become familiar with a data item's universe. In other words, the value or population that forms the base from which the data item in question is a subset. Therefore, if you want to create:

The proportion of persons age 5-9 across various geographic entities
Then the universe is total persons or total population.

While it may seem reasonable, or possible, to create a ratio of anything simply because the data are present in a table, chances are without considering what is being mapped the result can be garbage. This is the case with the example of Hispanics living in single family homes. The data themselves do not warrant the association.

Your inability to create the example of "Hispanic Americans who live in single-family homes" is a function of the concepts of unit of analysis and levels of summarization associated with the data. For instance, all the Census Bureau TIGER 95 data available from [ESRI's ArcData Online](http://www.esri.com/data/arcdata) and generally data available from the Census Bureau have been summarized to some level of geography: a census block, tract, city, county, state, or other geographic unit. This means that the

uniqueness of and access to individual responses from individual households or persons have disappeared. Data about individual persons and households have been blended (or summarized) with those of others. In this aggregated data, the unit of analysis is the county, state, or other geographic unit (which by the way protects the confidentiality of individual responses). In attempting to create a "cross-tabulation" of data by using the ArcView Legend Editor normalizing function, the math works but the answer is not what was intended.

In order to obtain a proper cross-tabulation, you need to work with the data at the "atomic" level. In the census, this means working with the data for individual households and persons. The only Census Bureau source for these data is the Public Use Microdata Samples (PUMS) which do not include name and specific location information but which allow for state, county, and other higherorder geographic and cross-tabulation analyses. In other words, this means you could explore data about Hispanic Americans who live in single-family houses but you cannot do it at very low levels of geography. No one except the Bureau has access to specific individual data especially for small geographic areas, such as census blocks.

So 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. Below is a listing of classification field-normalizing field pairs for a range of Bureau data items. The list was fashioned generally around the census data available from ArcData Online. It also represents many standard data items found in data products available directly from the Census Bureau. 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 you are using by investigating its data documentation also known as metadata.

Classification field

Normalize by

Age, gender, race, Hispanic origin (NOTE: Hispanic origin is not considered a race, but ethnicity)

Total population (persons)

Marital status

Population age 15+

Family composition (e.g., married families with children)

Total families

Household composition (e.g., one-person households)

Total housing units

Group quarters (e.g., persons in institutions)

Total population

Ability to speak English, migration (place of residence)

Population age 5+

Citizenship, place of birth

Total population

Labor force, employment, industry, occupation, travel to work

Population age 16+

Educational attainment (years of school completed)

Population age 25+

Persons enrolled in school (student population)

Population age 3+

Household Income

Total households

Occupancy status (occupied, vacant, seasonal units), rooms in unit

Total housing units

Housing tenure (owner vs. renter), persons in units, vehicles per household

Total households

Housing value

Owner occupied housing units

Contract rent

Renter occupied housing units

Units in structure, source of water and heat, sewer, age of housing units

Total housing units

Need more on normalizing data with ArcView?

To learn more about normalizing data in ArcView as well as other aspects of working with the Legend Editor, go to Chapter 6 "*Symbolizing your data*" in *Using ArcView GIS*, the user manual that comes with ArcView.

Need more on Census Bureau data concepts and documentation?

To learn more about the kinds of Census Bureau data available from [ESRI's ArcData Online site](#) as well as to begin to get a broader understanding of Census Bureau data and geographic concepts, click on this [link](#).

You can also research these topics and learn about the data that will become available from the 2000 Census by going to the links below.

Bibliography

Copyright © 1999 ESRI Canada

Authors

[Utah LessonPlans](#)