

# The Importance of Margins of Error and Mapping

By Lisa Berry and Diana Lavery

Almost all quantifications in your life are estimates: the speed on your speedometer, the weight on your bathroom scale, and the temperature on the thermometer in your oven.

Unless you've had these instruments calibrated recently, they are off by a bit. But you can still drive, track your weight, and cook dinner without fearing the effects of these small errors.

The same concept applies to the data we map. We can still understand which communities have higher or lower overall population even though the data contains some level of error.

Many fields of science rely on samples for their studies. Physical scientists work with water and soil samples to learn about the larger ecosystems. Doctors and medical scientists work with biometric lab samples to learn about the whole body. Social scientists often work with data that is a sample of a total population.

This is particularly true of the US Census Bureau's various surveys, which give us a representation of the population, without surveying every single person in the United States. Sampling is a cost-effective way to provide insight about a population and is a common practice when creating demographic datasets of all types.

When we create maps of sampled data—commonly using demographic and socioeconomic data—it is critical to understand the reliability of our data. It is also important to effectively communicate to our map readers the built-in error that comes with sampled data in a way that avoids causing them to mistrust the data. To understand how to do this, let's explore margins of error and what they mean to our mapping projects.

## What Are Margins of Error?

Margins of error, or MOEs, are an artifact of sampled data. For example, the American Community Survey (ACS) from the US Census Bureau offers a margin of error for the data estimates it provides. A MOE tells people using the data that the estimate is not an exact figure, but rather a range of possible values. A MOE helps us figure out that range.

↓ Figure 1: Calculating MOE for a sample of 361 people.

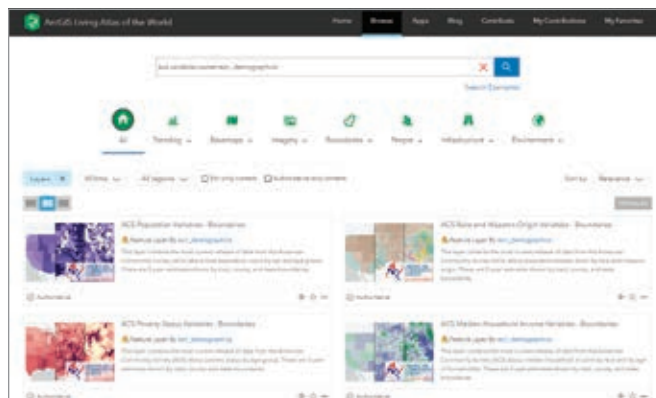
### Coefficient of Variation

Margins of error, or MOEs, are an artifact of sampled data. The American Community Survey (ACS) from the US Census Bureau provides MOEs for its data estimates that tell people using the data that the estimate is not an exact figure, but a range of possible values.

$$\left( \frac{\left( \frac{\text{MOE}}{1.645} \right)}{\text{Estimate}} \right) \times 100$$

Estimate: 361  
Margin of Error: 158

$$\left( \frac{\left( \frac{158}{1.645} \right)}{361} \right) \times 100 = 26.6\%$$



↑↑ Figure 2: The ACS layers in ArcGIS Living Atlas of the World are free to use and do not require credits or a subscription.

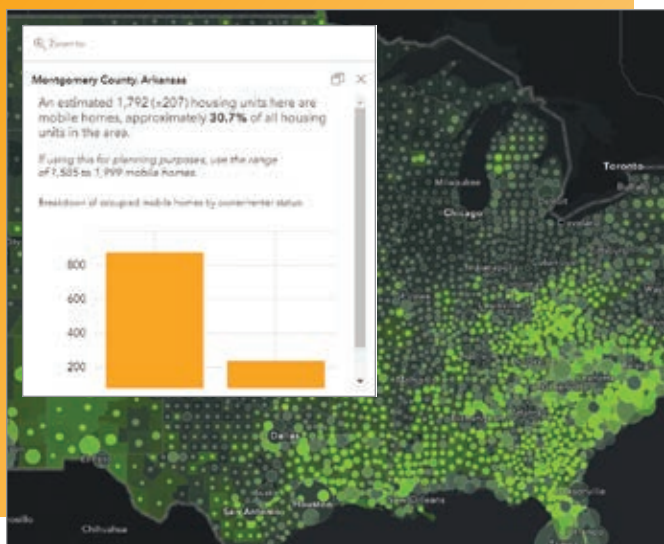
↑ Figure 3: Symbology can show MOE. An ArcGIS Arcade expression was used to calculate the coefficient of variation to extract the areas where there is lower reliability of the data.

As shown in the example in Figure 1, if the estimate for the number of people of a certain group in an area is 361, there will be an associated margin of error for that estimate. If the MOE is 158, the actual number of people in that group there falls somewhere between 203 and 519.

This range of values is known as the confidence interval, and it tells us that the Census Bureau is 90 percent confident that the count of population is between those upper and lower values.

## Why Use MOEs in Your Maps?

As stated in *Understanding and Using American Community Survey Data*, a handbook on using ACS data published by the US Census Bureau, "Estimates with smaller MOEs—relative to the value of the estimate—will have narrower confidence intervals indicating that the estimate is more precise and has less sampling error



↑ Figure 4: Communicate MOEs in a pop-up as in this map of mobile homes.

associated with it.” This tells us that not all MOEs and confidence intervals are equal.

In general, the larger the population, the smaller the MOE, and conversely, the smaller the population, the larger the MOE. Geographically, this means that states and counties typically have smaller MOEs than tracts and block groups because there are fewer respondents at smaller geography levels. Demographically, this means that estimates of a variable—as home ownership, education, health insurance status, or internet availability—has been disaggregated by age, sex, race/ethnicity, veteran status, or another characteristic, will have higher MOEs the more it is disaggregated, because the population being sampled as well as the sample is getting smaller and smaller.

Also, some population groups, such as those in low-income areas, are harder to survey due to lower response rates. For those areas, MOEs tend to be higher. When using the five-year estimates from ACS, the sample size is increased since there are 60 months of sample data pooled together, but even then, MOEs are often large.

There are two main types of survey errors: sampling and nonsampling.

- Sampling errors are caused solely by the fact that the entire population was not surveyed. Because a survey is only a sample, or subset, of the population, sampling errors are unavoidable. That is why—in part—figures that come from a sample are known as estimates. ACS MOEs only reflect sampling error.
- Any errors that are not caused by general sampling errors are nonsampling errors. An example of a systematic nonsampling error that is cited by the US Census Bureau could occur if no one from a

sampled housing unit is available during the data collection time frame. This error is known as unit nonresponse. It increases the chances that bias appears in the final survey.

These are just a few of the many factors that can impact the reliability of our data. Because these errors exist and can come from so many sources, it is important to effectively communicate the associated margins of error to our map audience. Your map reader could see a map and assume that the numbers are exact, when, in fact, they contain sampling errors. Being transparent about margins of error creates accountability for both the map’s creator and those making decisions from the data.

## Evaluating Margins of Error

One way to evaluate the reliability of an estimate is to understand the relationship between the estimate and its associated margin of error. One measure of reliability uses the coefficient of variation, which is a fancy way of describing the error as a percent of the estimate. In the previous example, which had a MOE of 158, the percent coefficient of variation would be 26.6. If an error is large in relation to the estimate, this coefficient will be large, which indicates a lower reliability. The higher the coefficient, the lower the reliability.

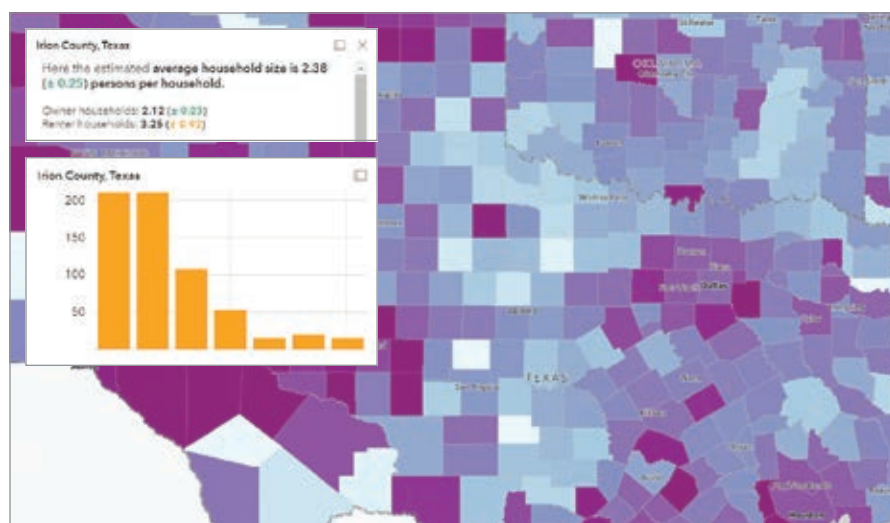
Note that 1.645 is used, since the ACS estimates are provided by Census Bureau at a 90 percent confidence level, and under a standard normal bell curve, 90 percent of the area beneath the curve is between 1.645 and -1.645. To convert to a different confidence level, use a different constant here, such as 1.96 for a 95 percent confidence level. If the MOE is 0, the estimate is likely controlled to be equal to a fixed value and has no sampling error.

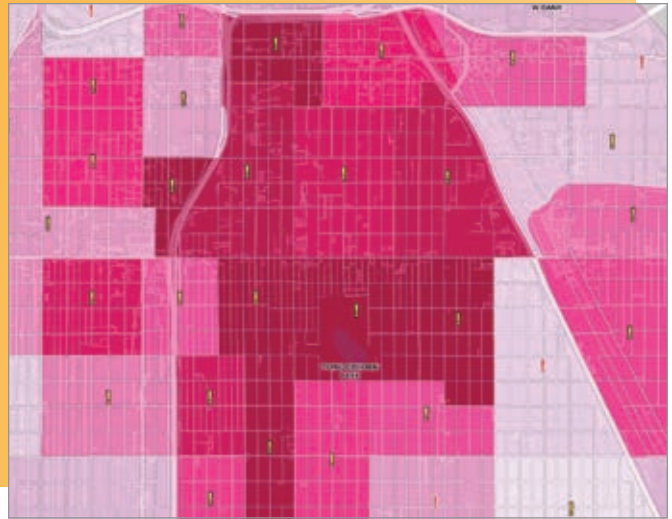
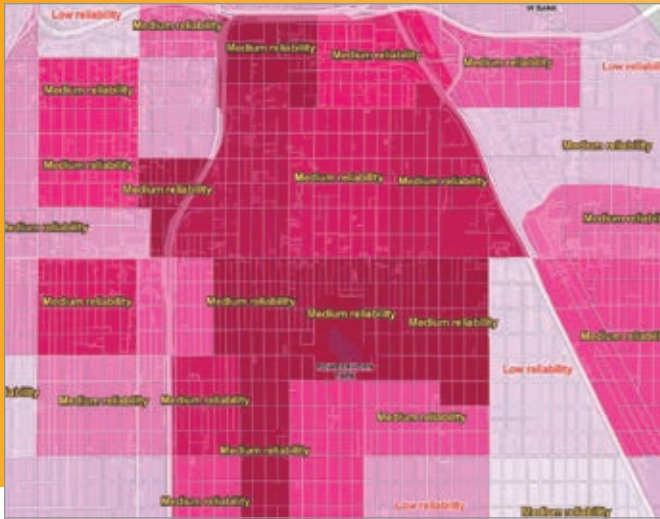
## Ways to Use MOEs and Reliability in Mapping

ACS data is available through various GIS workflows in ArcGIS. This article provides a few examples of how to locate ACS data and map MOEs so that they can be better understood.

With ready-to-use Census ACS layers from ArcGIS Living Atlas of the World (<https://bit.ly/3q5sjPm>), you can access thousands of ACS variables and their MOEs. They are free to use and do not

↓ Figure 5: This example utilizes an ArcGIS Arcade expression to create custom colors that indicate the reliability of the estimates.





↑ Figure 6: These scale dependent labels indicate data reliability only when zoomed in (left) and are not visible when the map is zoomed out.

require any credits or a subscription to use. These layers are organized by various topics. One or many ACS tables are included in each layer. Each ACS estimate or precalculated percentage comes with its associated MOE.

To see the list of fields for a layer, go to the ArcGIS Online item details page, click the Data tab, and click Fields. These layers can be easily customized in ArcGIS Online, ArcGIS Pro, or ArcGIS Enterprise. They provide data at state, county, and census tract levels of geography.

You can include MOEs in maps you make with these layers using symbology, pop-ups, custom ArcGIS Arcade expressions, and labels. Although the maps accompanying this article were created in ArcGIS Online using ArcGIS Living Atlas layers and Arcade expressions, these maps can be replicated in ArcGIS Pro using the same data and expressions.

## Symbology

Symbology is one method for showing MOE. In the example in Figure 3, an Arcade expression was used to calculate the coefficient of variation to extract the areas where data is less reliable. These areas of low reliability are overlaid on top of the map pattern to show areas with higher MOEs. This example highlights any area with a coefficient variance of more than 40 percent.

## Pop-Ups

Figure 4 shows how MOEs could be communicated in a pop-up. Although this method is less alarming to the map reader than symbology, it still effectively warns the map reader that the data being mapped is an approximation. Note the use of words such as *estimated*, *approximately*, and *range*. Pop-ups subtly highlight that the estimates contain some amount of error without scaring users and eliminating all trust in the data.

## Custom Colors Using Arcade Expressions

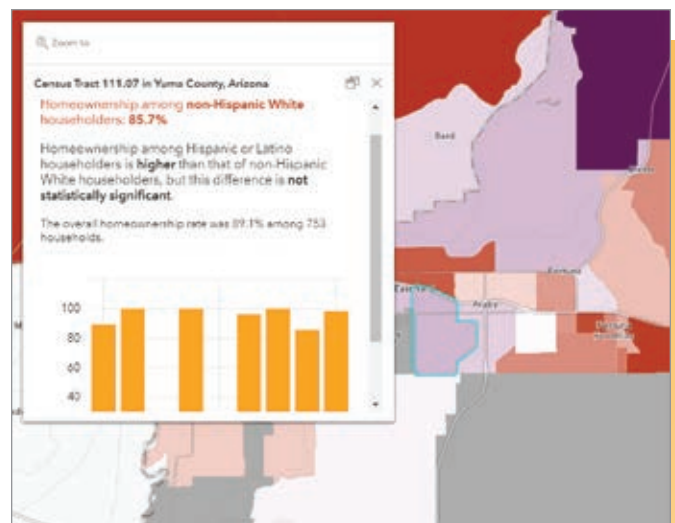
The example in Figure 5 uses some of the same techniques as the example in Figure 4 but utilizes an Arcade expression to create custom colors that show the reliability of the estimate. The color is based on

the MOE as a percent of the estimate (the coefficient of variation). Arcade helps categorize this percentage into high, medium, or low reliability. (For an explanation of high, medium, and low category thresholds, see the section Using Esri Demographics and ArcGIS Business Analyst.) Note that a disclaimer is included in the pop-up for those unfamiliar with ACS data, and a link to learn more about it.

## Labels

Figure 6 shows various ways labeling can communicate MOEs. These labels are scale dependent, so when the map is zoomed out, the map pattern is not obstructed. When the map is zoomed in to the neighborhood level, the labels become visible and display categorical estimates of the data's reliability. This map uses the same Arcade statement used as the map in Figure 5 and the label classes that are available in the new Map Viewer.

↓ Figure 7: Data reliability is conveyed statistically and visually in this map, which uses an Arcade expression to perform the statistical test on the fly and converts the results into an easy-to-read statement in the pop-up in combination with the Compare A to B mapping style.



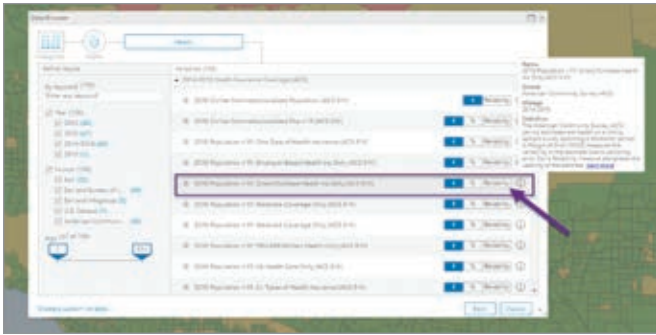
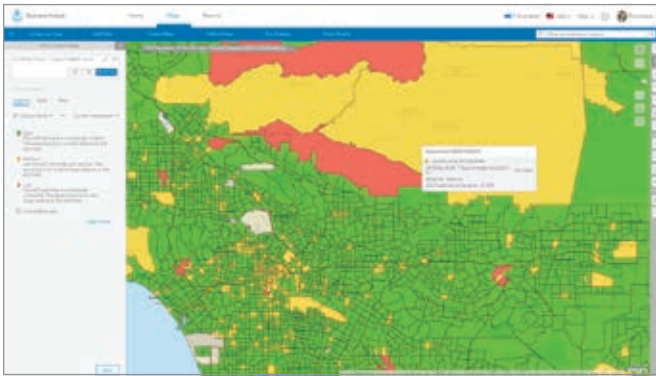


Figure 8: When choosing an ACS attribute in ArcGIS Business Analyst, the MOE is offered as a reliability estimate (REL).

Figure 9: In Business Analyst, the reliability estimate is categorized as high, medium, or low.



high (REL = 1). Between 12 and 40 percent, reliability is considered medium (REL = 2). If MOE is more than 40 percent, reliability is considered low (REL = 3).

*Special thanks to Esri staff members Helen Thompson, Jim Herries, and Steven Aviles for the map examples included in this article. Also thanks to Kevin Krivacsy; Kevin Butler; and Kyle R. Cassal, chief demographer at Esri, for providing valuable insight about margins of error and mapping.*

## About the Authors

**Lisa Berry** is a senior product engineer at Esri on the ArcGIS Living Atlas of the World team. She works to create clear and concise stories about demographic, socioeconomic, and policy topics using cartography. Berry builds data layers and tutorials to help others create their own map masterpieces.

**Diana Lavery** loves working with data. She is a senior product engineer on ArcGIS Living Atlas of the World's policy maps team and has more than a decade of experience as a practitioner of demography, sociology, economics, policy analysis, and GIS. Lavery holds a bachelor's degree in quantitative economics and a master's degree in applied demography. She enjoys strong coffee and clean datasets, usually simultaneously.

## Testing for Statistical Significance

MOEs also help calculate if two things are significantly different. By using the Statistical Testing Tool available from the US Census Bureau (<https://bit.ly/2TvmZZg>), two attributes can be compared while taking into account the MOE associated with survey estimates. The map in Figure 7 uses this z-score method to compare home ownership rates of White non-Hispanic homeowners and Hispanic or Latino homeowners. The pop-up uses an Arcade expression to perform the statistical test on the fly and convert the results into an easy-to-read statement in combination with the Compare A to B mapping style, which maps the ratio between two numbers, expressing that relationship as percentages, simple ratios, or an overall percentage and comparing two patterns both visually and statistically.

## Using Esri Demographics and ArcGIS Business Analyst

Esri Demographics can be accessed in many ways (<https://bit.ly/3gBbHuq>) from ArcGIS products, such as ArcGIS Business Analyst, and via Esri's ArcGIS GeoEnrichment Service, as well as from ArcGIS Living Atlas. There are thousands of ACS variables available at geographies for states, congressional districts, ZIP codes, census tracts, block groups, and other levels. Custom polygons can also be enriched with the attributes of your choice.

When choosing an ACS attribute in Business Analyst, MOE is offered as a reliability estimate (REL) as shown in Figure 9. The REL attribute is a reliability estimate, which categorizes the MOE by the coefficient of variation. This reliability estimate is broken into three categories: high, medium, or low reliability. If the MOE is less than 12 percent of the estimate, its reliability is considered

# Reach GIS Professionals, Managers, & Developers

Advertise in *ArcUser* and reach more than 450,000 people who are actively using GIS and Esri technology in their organizations.

ads@esri.com

Copyright © 2021 Esri. All rights reserved.