



- [Home](#)
- [Add Your Blog](#)
- [About](#)
- [Talk Stats](#)

## A log transformation of positive and negative values

July 14, 2014

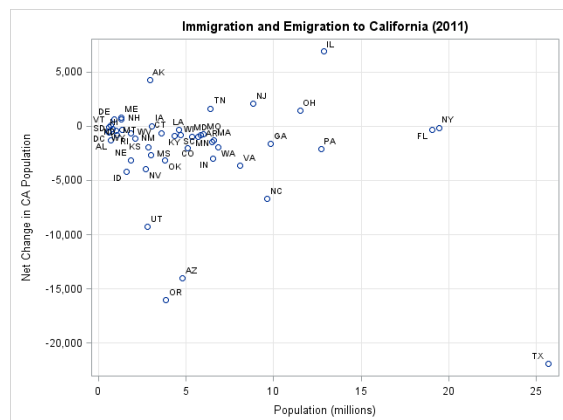
By [Rick Wicklin](#)

(This article was originally published at [The DO Loop](#), and syndicated at [StatsBlogs](#).)

In my four years of blogging, the post that has generated the most comments is ["How to handle negative values in log transformations."](#) Many people have written to describe data that contain negative values and to ask for advice about how to log-transform the data.

Today I describe a transformation that is useful when a variable ranges over several orders of magnitude in both the positive and negative directions. This situation comes up when you measure the net change in some quantity over a period of time. Examples include the net profit at companies, net change in the price of stocks, net change in jobs, and net change in population.

### A state-by-state look at net population change in California

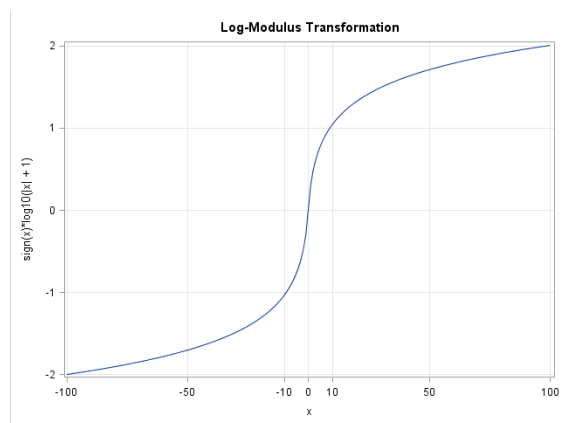


Here's a typical example. The US Census Bureau tracks state-to-state migration flows. (For a great interactive visualization of internal migration, [see the Governing Data Web site](#).) The adjacent scatter plot shows the net number of people who moved to California from some other US state in 2011, plotted against the population of the state. (Click to enlarge.) For example, 35,650 people moved to California from Arizona, whereas 49,635 moved to Arizona from California, so Arizona was responsible for a net change of  $-13,985$  to the California population. The population of Arizona is just under 5 million, so the marker for Arizona appears in the lower left portion of the graph.

Most states account for a net change in the range  $[-5000, 5000]$  and most states have populations less than 5 million. When plotted on the scale of the data, these markers are jammed into a tiny portion of the graph. Most of the graph is empty because of states such as Texas and Illinois that have large populations and are responsible for large changes to California's population.

As discussed in my blog post about [using log-scale axes to visualize variables](#), when a variable ranges over several orders of magnitudes, it is often effective to use a log transformation to see large and small values on the same graph. The population variable is a classic example of a variable that can be log-transformed. However, 80% of the values for the net change in population are negative, which rules out the standard log transformation for that variable.

### The log-modulus transformation

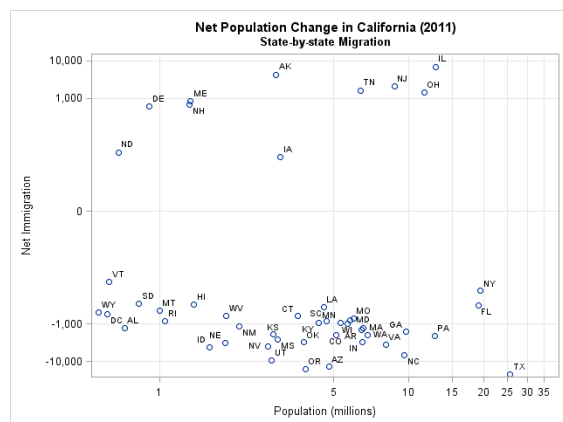


A modification of the log transformation can help spread out the magnitude of the data while preserving the sign of data. It is called the log-modulus transformation ([John and Draper, 1980](#)). The transformation takes the logarithm of the absolute value of the variable plus 1. If the original value was negative, "put back" the sign of the data by multiplying by  $-1$ . In symbols,

$$L(x) = \text{sign}(x) * \log(|x| + 1)$$

The graph of the log-modulus transformation is shown to the left. The transformation preserves zero: a value that is 0 in the original scale is also 0 in the transformed scale. The function acts like the log (base 10) function when  $x > 0$ . Notice that  $L(10) \approx 1$ ,  $L(100) \approx 2$ , and  $L(1000) \approx 3$ . This property makes it easy to interpret values of the transformed data in terms of the scale of the original data. Negative values are transformed similarly.

### Applying the log-modulus transformation



Let's see how the log-modulus transformation helps to visualize the state-by-state net change in California's population in 2011. You can [download the data for this example and the SAS program that creates the graphs](#). A previous article showed [how to use PROC SGPLOT to display a log axis on a scatter plot](#), and I have also discussed [how to create custom tick marks for log axes](#).

The scatter plot to the left shows the data after using the log-modulus transformation on the net values. The state populations have been transformed by using a standard log (base 10) transformation. The log-modulus transformation divides the data into two groups: those states that contributed to a net influx of people to California, and those states that reduced the California population. It is now easy to determine which states are in which group: The states that fed California's population were states in New England, the Rust Belt, and Alaska. It is also evident that size matters: among states that lure Californians away, the bigger states tend to attract more.

The main effect of the log-modulus transformation is to spread apart markers that are near the origin and to pull in markers that are relatively far from the origin. By using the transformation, you can visualize variables that span several orders of magnitudes in both the positive and negative directions. The resulting graph is easy to interpret if you are familiar with logarithms and powers of 10.

tags: [Data Analysis](#), [Statistical Programming](#)

◆ [Email this](#) ◆ [Share on Facebook](#) ◆ [Stumble It!](#) ◆ [Digg This!](#) ◆ [Save to del.icio.us](#) ◆ [Discuss on Newsvine](#)  
◆ [Add to Mixx!](#)

**Please comment on the article here:** [The DO Loop](#)

Tags: [Data Analysis](#), [Statistical Programming](#), [Uncategorized](#)