How to Avoid Mistakes When Using Census Data

The previous two installments of *Research Corner* introduced readers to the tools provided by Census.gov for accessing data, comparing counties, and creating visual representations of comparisons. This month, *Research Corner* will inform readers about instances in which there is a risk of misinterpreting or misusing census data. It is important to work to prevent these kinds of mistakes when using census data to describe and compare counties.

Mean versus median

Statistics describe trends in a data set. One of the most commonly referenced statistics is the *mean,* or average. The mean is found by adding together all the values of the data set and dividing by the number of values in that set. Though the mean is usually used to describe the "average" score in a data set, this number can be misleading when there are extreme numbers, referred to as "outliers" in the data. These outliers can skew data so that the average is not actually representative of the values provided.

There is another statistic that can also be used to describe what number is "typical" in a data set: the *median*. The median is the value of the data set that lies in the middle of the values when they are sorted from smallest to largest. The census and other surveys will often use the median to give an example of a "typical" value instead of using the statistical average. They use the median because it is less influenced by extreme values in the data set than the mean. For example, the median is used to describe the price of rent or income in a locality because it gives the best representation of what is usual for that area.

Figure 1 shows how the mean and median can describe the same data set differently. Because of the larger number in the set, the mean (average) salary is higher than most of the salaries of the hypothetical county. The median, the middle value of the data set, gives a more accurate description of the salaries. This shows why the census gives median numbers for characteristics like income or home value.

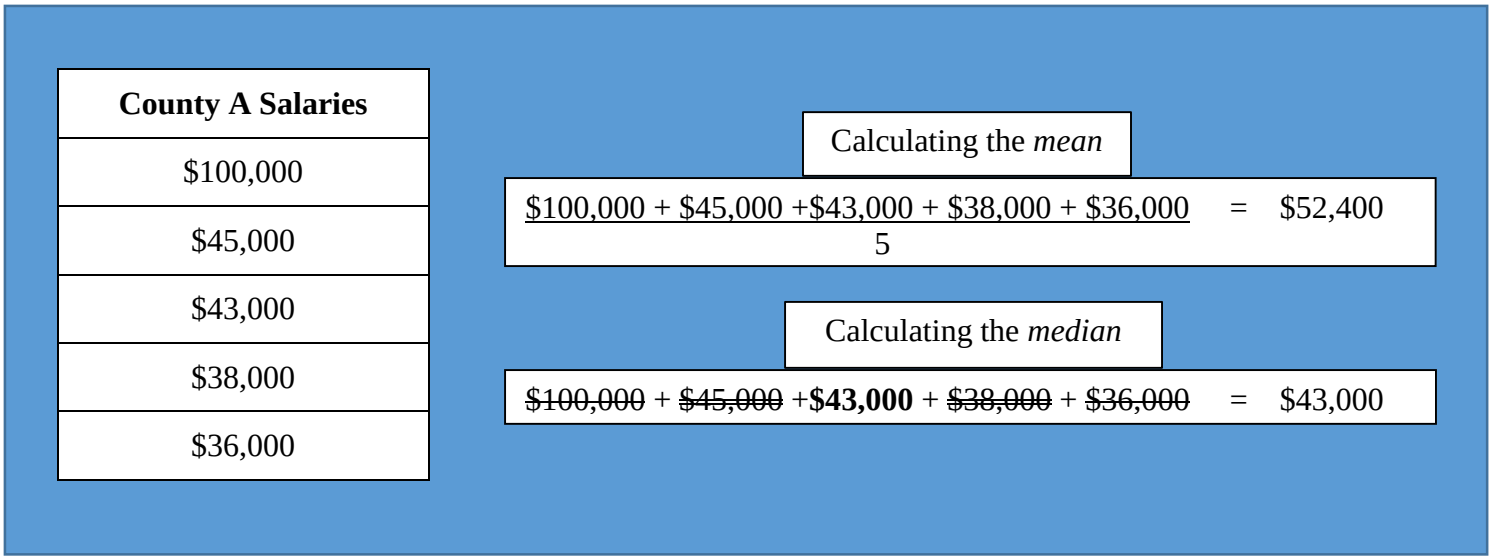

Figure 1. Calculating the mean and median of income of County A Salaries

Considering multiple characteristics

Using the QuickFacts Beta tool is a great way to draw compelling parallels between counties. However, when creating comparisons to argue a point, readers should be sure to select the counties to compare carefully. Being as counties vary in their characteristics, they may be comparable in one sense but not in another. For instance, let's say a researcher is looking to compare senior services offerings in their county. This researcher may look to another county of similar population to be a good county for comparison. However, though two counties may have similar overall populations, their population of residents over 65 years of age may be extremely different. Or perhaps two counties have very different population sizes, but are similar in land

area. It is important to be aware of the influence of various characteristics when selecting localities so that conclusions drawn from comparisons can be as strong and accurate as possible.

For example, Figure 2 shows a comparison of four counties with similar populations: Union County, Putnam County, McDuffie County, and Dodge County. These counties differ, however, for other characteristics. Union County has a higher percentage of persons over 65 years of age in its population than does Putnam County. This may imply that Union County has more senior-related needs to meet with county services than does Putnam County. Furthermore, while McDuffie County and Dodge County have similar percentages of their populations over 65 years of age, they differ greatly on population per square mile. Though the size of the population needing senior services is similar in McDuffie County and Dodge County, the delivery of these services may vary because of the difference in population density. This difference in specific population size and density should be taken into account if comparing the services offered to seniors in the two counties.

	County			
Characteristic	Union	Putnam	McDuffie	Dodge
Population of county	21,984	21,192	21,370	20,976
Persons 65 years and over	**26.60%**	**18.10%**	13.60%	13.80%
Population per square mile	66.30	61.60	**85.00**	**44.00**

Figure 2. Comparison of four counties with similar populations

Looking at long term trends over time

In addition to ensuring the counties you choose to compare are appropriate for your research needs, another important concern arises when making county comparisons over time. Within census data and data from other surveys are multiple collection years and estimates for

different time periods. If looking to draw conclusions about changes between points in time, it is important for researchers to distinguish between a simple annual fluctuation and a distinct change in the data. Looking at multiple years or points in time helps to avoid this. For instance, Figure 3 shows two graphs of the population of Liberty County. If looking at only data for 2012 and 2013, it would seem as though there was a significant decrease in population during this period. However, viewing multiple years shows that, both before and after the 2012-2013 time period, the population increased. This provides an example of the potential to misinterpret data when comparing across time. Understanding overall trends is important to accurate analysis.

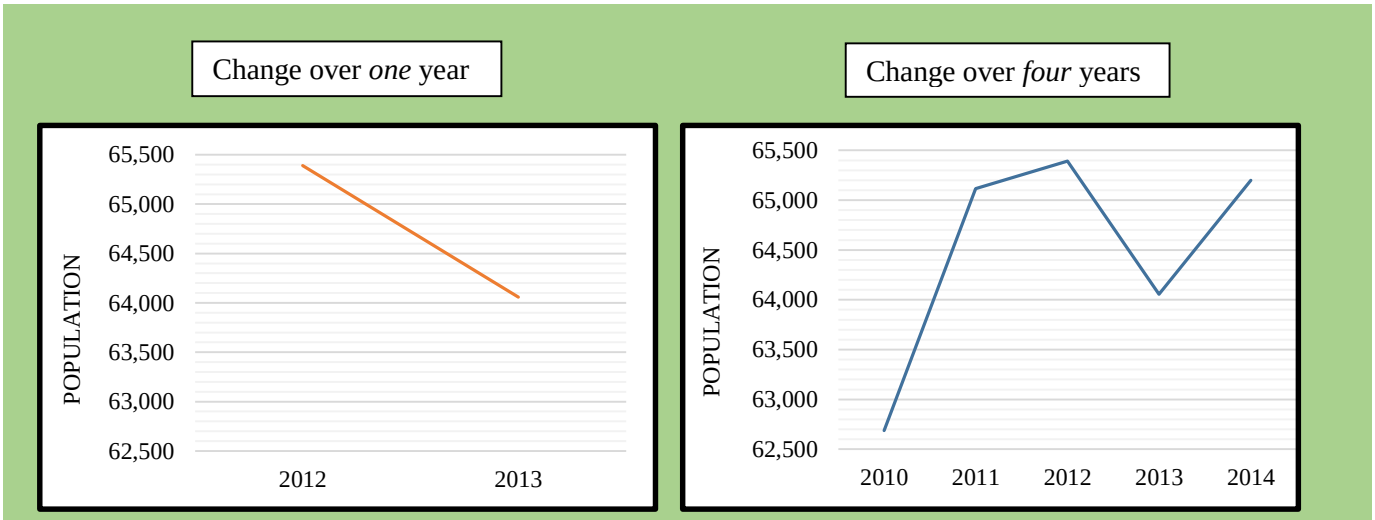

Figure 3. Population of Liberty County

Furthermore, when comparing how two counties change over time, observing percent change can help equalize variations so they are more comparable. For instance, even if two counties increase in population by the same amount of people, the percent increase would better portray the impact the increase may have on the county. Figure 4, for example, shows the census population counts for Glynn County and Long County from 2010 to 2014. While the population increase across the five years was similar for both counties, the percent change shows how the

increase of about 2,400 people is a more substantial change for Long County than for Glynn County.

County	2010	2011	2012	2013	2014	Change in population 2010-2014	Percent change 2010-2014
Glynn County	79,802	80,216	80,960	81,533	82,175	2,373	2.89%
Long County	14,679	15,243	16,161	16,641	17,113	2,434	14.22%

Figure 4. Population changes in Glynn County and Long County, 2010-2014

Conclusion

In order for researchers to be successful in conveying correct conclusions, they must be aware of the dynamic nature of the data available to use when describing and comparing counties. The past three census-focused installments of *Research Corner* have mentioned a few of the limitless possibilities for putting census data to work for counties. Whether it's when requesting funding, demonstrating reach, drawing comparisons, or dispelling myths, tactful use of census data can provide counties with compelling support. Armed with an understanding of the online tools available for accessing data and with guidance on avoiding mistakes of misinterpretation, readers should feel confident in putting census data to work for their counties.