Title	Mapping American community survey data
Author(s)	Torrieri, NK; Ratcliffe, M; Wong, WSD
Citation	2011 Joint Statistical Meetings, Miami, Florida, 30 July-4 August 2011, p. 1089-1100
Issued Date	2011
URL	http://hdl.handle.net/10722/192447
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Mapping American Community Survey Data

Nancy Torrieri[1], David Wong[2], Michael Ratcliffe[3]
[1]U.S. Census Bureau, Washington, DC 20233
[2]George Mason University, Fairfax, VA 22030
[3]U.S. Census Bureau, Washington, DC 20233

Abstract

Maps are a frequently used tool to portray the Census Bureau's data and highlight spatial patterns that provide context and significance for the characteristics displayed. Maps provide visually what tables and other graphics cannot: a picture of the data, their distribution over geographic areas, and a means for interpreting the data shown by color, symbology, or explanation provided as annotations or as part of the map legend. The value of maps in enhancing an understanding of census data is well established as demonstrated by their frequent use in the media following the release of census data products. Mapping census data is common throughout government, academia, and the private sector.

Casual users of maps of statistical data may not look past what is interesting visually to analyze the underlying data that a map depicts. However, that does not absolve the mapmaker of the responsibility for informing users of the statistical qualities associated with the mapped values. The Census Bureau set new standards for communicating the statistical qualities of estimates from the American Community Survey (ACS) by including information on the level of sampling error (specifically, margins of error) associated with every ACS estimate. Now, efforts are underway to develop an operational tool that will make it possible for geographic information systems (GIS) users to communicate this information through map products as well.[1]

Key Words: American Community Survey, Mapping, Sampling Error

1. Background

Including data quality information in maps has long been of interest to cartographers.[2] A number of methods have been proposed to design maps that communicate the statistical qualities of the data portrayed. At the Census Bureau, an informal working group established in the 1990s considered the issue of statistical mapping and raised an awareness of the issues that needed to be considered.[3] However, this group did not achieve a consensus on any particular guidance that should apply. The Census Bureau's methodological standards for the presentation of statistical data apply to many kinds of data products, but no detailed guidance has been established to suggest how such standards should be implemented for maps that feature survey data.[4]

---

[1] The views expressed in this paper are those of the authors and not necessarily those of the U.S. Census Bureau.
[2] Beard and Buttenfield.
[3] U.S. Census Bureau, Statistical Mapping Workgroup Notes.
[4] U.S. Census Bureau, Statistical Quality Standards, 87.

The difficulty in reaching consensus on methods to use to incorporate statistical data into maps may reflect in part an inconsistent application of specific standards throughout the federal government. For example, although the Federal Geographic Data Committee has proposed attribute accuracy as one of six measures of spatial data quality, federal agencies have not demonstrated consistent implementation of this standard.[5] The National Academy of Sciences panel tasked to consider the usability of ACS data urged that margins of error be examined before drawing conclusions from a set of estimates, but did not advise a specific approach to implement this guidance for maps.[6]

Examples of maps of ACS estimates appear regularly in the media following the release of new ACS data products. These maps illustrate characteristics of populations and areas on topics of interest to the public. They usually cite the source of data as being the U.S. Census Bureau, and sometimes cite the specific source of the data, for example, the ACS. But examples of maps that include specific information on how the data portrayed should be interpreted are rare. More typically, an interesting statistic is displayed on a map, and a reference to the source appears in a brief note somewhere on the map or in an accompanying article. Percentages shown on the map are presented as values without errors and no information on their quality or reliability is included on the map. A typical example of a map with some of these limitations was published by the *Washington Post* in September 2010.[7]

Other examples of maps of ACS estimates are emerging from non-media sources. However, these maps also include limited information relating to the interpretation of these estimates. Figure 1 is one of a series of maps from a rural atlas developed by the U.S. Department of Agriculture's Economic Research Service.[8] This specific map depicts information on characteristics of income and employment at the county level. Information for Lake County, Montana is highlighted in a pop-up table. The ACS, as well as other surveys, are the source of the data shown on this and other maps in this atlas. While interactive links provide the ability to navigate among the maps and access information on their sources, access is not straightforward. A tab at the top of the display leads us to learn that one of the sources of the data is the ACS. Additional clicks are required to reach the Census Bureau's ACS web page, www.census.gov/acs/www/, and eventually, specific information that could be helpful to interpret the data. It is probably the case that only an ambitious map user will follow the trail of links to find information on the statistical uncertainty of ACS estimates and take the time to learn more about this topic. Users who do not concern themselves with such details could be left with the impression that the classes of estimates portrayed have statistical significance in terms of the data, when in fact the estimates representing various classes of data might not be statistically different.

[5] Federal Geographic Data Committee, 1998.
[6] Citro and Kalton, 2007,130.
[7] *Washington Post,* September 29, 2010.
[8] USDA, *Rural Atlas*.

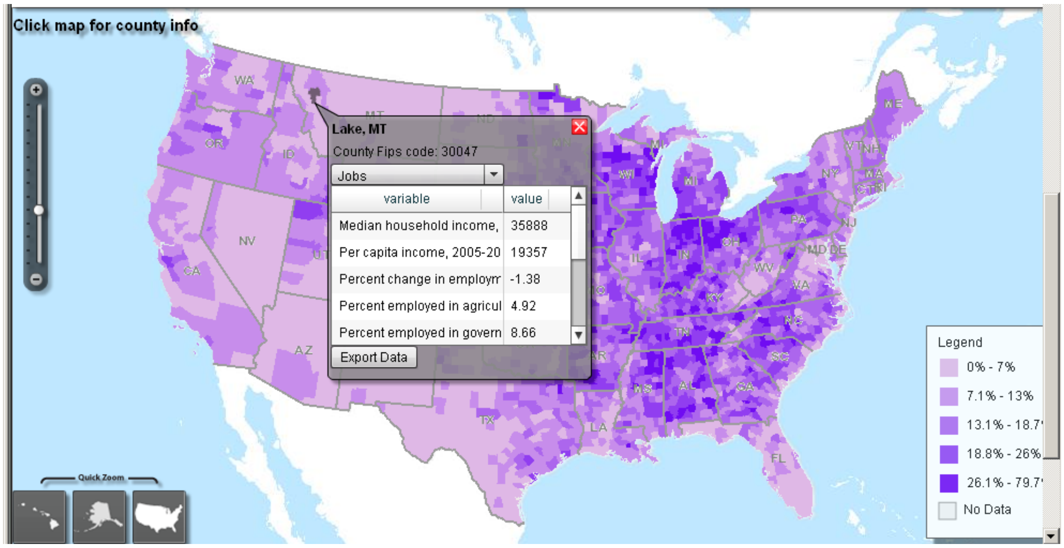

Figure 1: USDA Economic Research Service: Rural Atlas, Employment Characteristics, Lake County, Montana

It is clear from these examples that designers of maps of ACS data could benefit from standards or guidelines relating to the appropriate presentation of survey data. In the next section of this paper, we review Census Bureau efforts to improve maps of ACS data, and discuss efforts underway to develop operational tools to make it easier to map these data appropriately.

2. ACS Data in Census Bureau Maps

Since the first release of data during the testing and development phase of the ACS program, ACS data tables featured in American FactFinder (AFF), the Census Bureau's chief data dissemination vehicle, and as CD-ROM products developed before the ACS was fully implemented, have included information on the margins of error (MOEs) of ACS estimates. But the production of a map incorporating the information on the statistical uncertainty of ACS estimates did not take place until after the ACS program was fully launched. In 2007, the Census Bureau released an ACS report containing a map that distinguished statistically significant differences in real median household income by state from 2005 to 2006 (see figure 2).[9]

[9] U.S. Census Bureau, "Income, Earnings, and Poverty," 6.

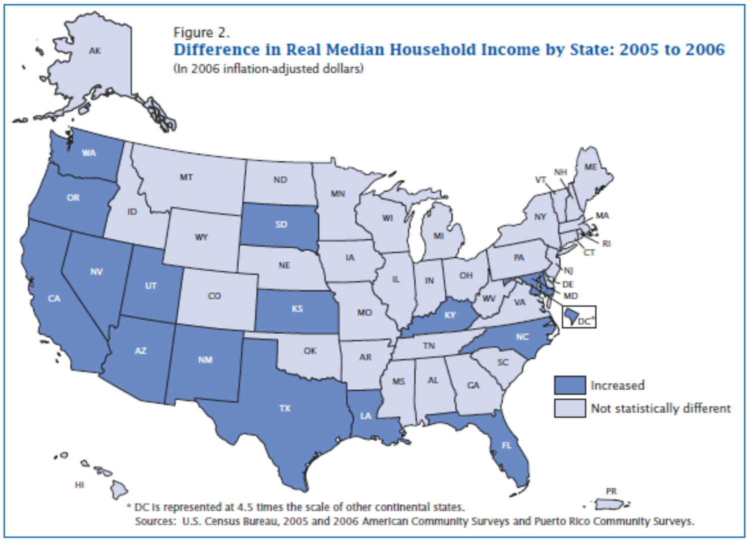

Figure 2: Income Differences by State, 2005-2006, Based On the 2005 and 2006 American Community Surveys and Puerto Rico Community Surveys

With the release of 2006 ACS data in AFF, users of AFF's thematic mapping function were provided with an interactive display option to distinguish areas for which estimates were statistically different. That option allows users to toggle between two views of mapped ACS estimates, one that portrays the distribution of a characteristic by geographic area based on the given estimate, without consideration of the margins of error, and the other that identifies areas that are not statistically different from an area of reference for a particular estimate. For example, in Figure 3, the estimate for South Dakota (the geographic area of reference identified by a round map pin) is not statistically different from the estimates for Montana, Vermont, and Arkansas. However, the estimate for South Dakota is statistically different from the estimates for Iowa, North Dakota, Hawaii, and Puerto Rico, which represent the same data class, 14.1-17.3 percent, but for which a hatched pattern is not present.

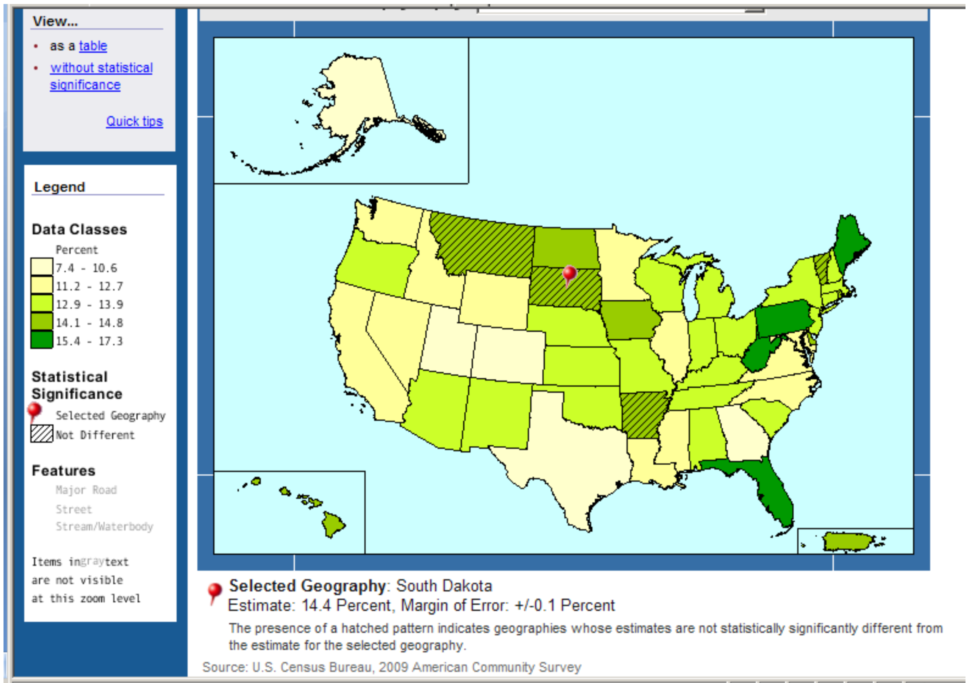

Figure 3: American FactFinder Thematic Map Showing Areas With 2009 ACS Estimates That Are Not Different From the Selected Geography, South Dakota

In further recognition of the need to reflect the statistical uncertainty of survey estimates on maps developed for public release the Census Bureau's Public Information Office developed a map that included an advisory note urging that statistical testing should be used as a basis of comparisons of estimates between areas. This map was included in the mid-December 2010 release of ACS data as part of a special series of maps portraying ACS estimates of interest.[10] A portion of a map from this series is shown in Figure 4.

[10] U.S. Census Bureau, "2009 ACS Maps."

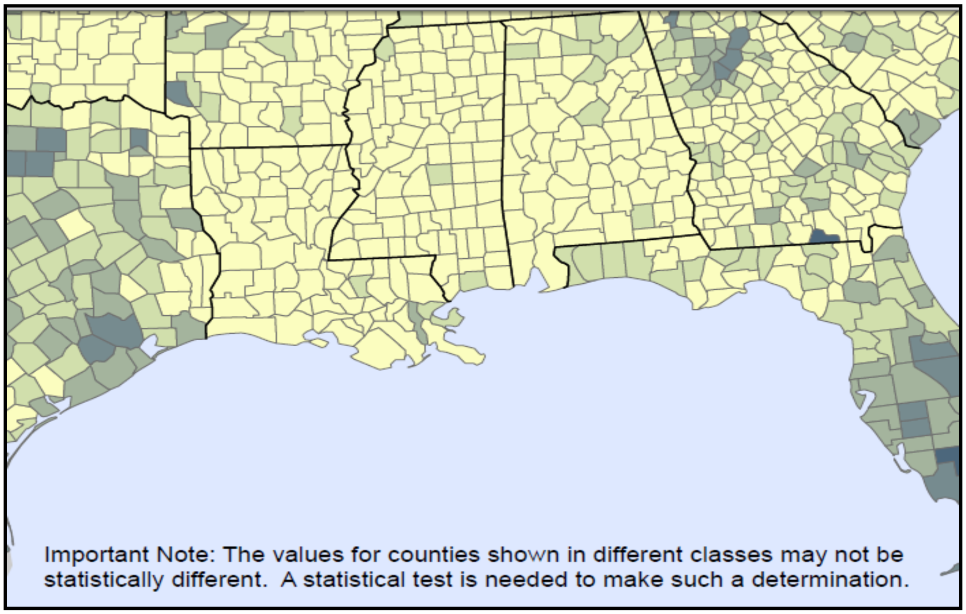

Figure 4: Portion of a Map of ACS 5-Year Estimates Displaying Advisory Message Relating to the Need for Statistical Testing

These developments reflected well on the Census Bureau's efforts to promote good mapping practices but probably did little to improve professional approaches for mapping ACS data. To address head-on the need for specific guidance or tools to map survey estimates including those produced by the ACS, the Census Bureau decided to launch an exploratory research and outreach initiative. The broad goals of this initiative were (1) to create one or more operational tools to facilitate the mapping of ACS data to include information on the statistical uncertainty of ACS estimates; and, (2) to promote the use of these tools by posting them for downloading directly from a web site, and by participating in professional meetings with stakeholders to explain why these tools are important. The remainder of the paper summarizes progress as of August 2011 on the project that serves as a vehicle for this initiative.

3. Approaches to Displaying Statistical Uncertainty in Mapping ACS Estimates

Many proposed methods can incorporate information on the statistical uncertainty of estimates on a map.[11] In the ongoing effort, two approaches were evaluated in depth and implemented for mapping ACS data using a popular desktop geographic information system (GIS) software package, ArcGIS, developed by the Environmental Systems Research Institute (ESRI). Collectively, these approaches will display the coefficients of variation (CVs) for ACS estimates together with the estimates and allow users to determine which (if any) estimates for the geographic areas on a map are significantly different from the estimate for a selected area. Work on the development of operational tools to perform these functions is complete. In both cases, the tools were designed to create maps that communicate clear messages, through map legends, colors, and patterns, that explain statistical qualities of ACS data while at the same time reflecting the best

[11] Sun and Wong, "Incorporating Data Quality Information."

practices of map design. Because the tools automate processes that would normally be required to map ACS data with quality information, they vastly decrease the time required to develop a map over that required without the aid of the tools.

The tools are bundled as an extension that functionally enhances the capabilities of ArcGIS beyond what its standard tools can provide to handle and map ACS data. The "ACS Mapping Extension" can be installed in ArcGIS version 9.3 or 10.0, which are supported by most Windows operating systems. The extension provides navigation aids through a series of pull-down menus, such as the one shown in Figure 5.

Figure 5: A Pull-Down Menu and Menu Items in the ACS Mapping Extension for ArcGIS

Besides the two mapping functions, the extension also includes step-by-step instructions to download ACS data and census boundary data from the U.S. Census Bureau website. ACS data are available in large volume, such as the 5-year estimates for census tracts for an entire state, from the download center accessible from AFF. As of February 2011, two versions of AFF were online, a legacy version that provides access to ACS data, and a second version, AFF II, designed to be the source for online access to ACS data starting in Fall 2011. The boundary data, which are based upon the Topologically Integrated Geographic Encoding and Referencing (TIGER) files from the U.S. Census Bureau, are available in shapefile format.[12] TIGER shapefiles provide the boundaries of geographic areas, such as census tracts, for which tabulations are released from the U.S. Census Bureau. Shapefiles represent a popular geospatial vector data format for GIS software. They store geometrical data in points, lines, and polygons that represent features such as landmarks, roads, lakes, or parks.

Figure 6 displays a map created by a mapping function of the extension designed to map ACS estimates together with the CVs. The map shows ACS estimates of median household income for Iowa counties with patterns representing ranges of CV values superimposed on the ACS estimates for counties. The data are 2005-2009 ACS 5-year estimates. The CV provides a measure of the relative amount of sampling error that is associated with a sample estimate. A small CV indicates that the sampling error is small relative to the estimate, and, thus, the user can be more confident that the estimate is close to the actual value.[13] It should be noted there may be little value in knowing differences among the CVs in this case, since they are so small (less than 8 percent), indicating that all of the data are of excellent quality. The cross-hatch patterns are emphasized to make them more recognizable for illustrative purposes.

---

[12] U.S. Census Bureau, "TIGER Products."
[13] U.S. Census Bureau, *A Compass for Understanding and Using American Community Survey Data*, A-13.

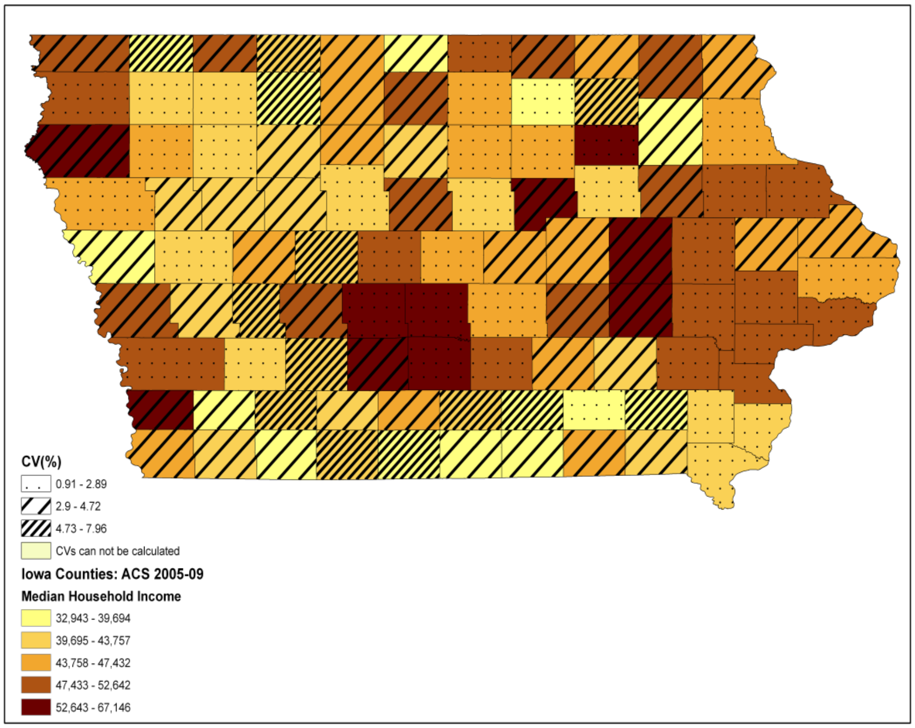

Figure 6: Overlay of the Coefficients of Variation with Median Household Income Estimates by County, Iowa based on American Community Survey 2005-2009 5-Year Estimates

A second approach to mapping ACS data is to reflect statistical differences between estimates. Geographic areas on a map are colored differently to reflect that they have different values, and it would be reasonable to expect casual users of maps to assume that units with the same color have values similar to each other. However, mapping software does not automatically assign colors to units for which values are statistically different. Areas with different colors could have estimates that are not significantly different from each other, and areas with the same colors could have significantly different estimates. The ACS mapping extension allows the user to select a geographic area of reference (in this case, Harrison County, Iowa, outlined in green), determine whether the estimates for other counties are statistically different from the area of reference, and assign a pattern to identify such areas based on a 90 percent level of confidence. Figure 7 shows a map produced using this mapping function.

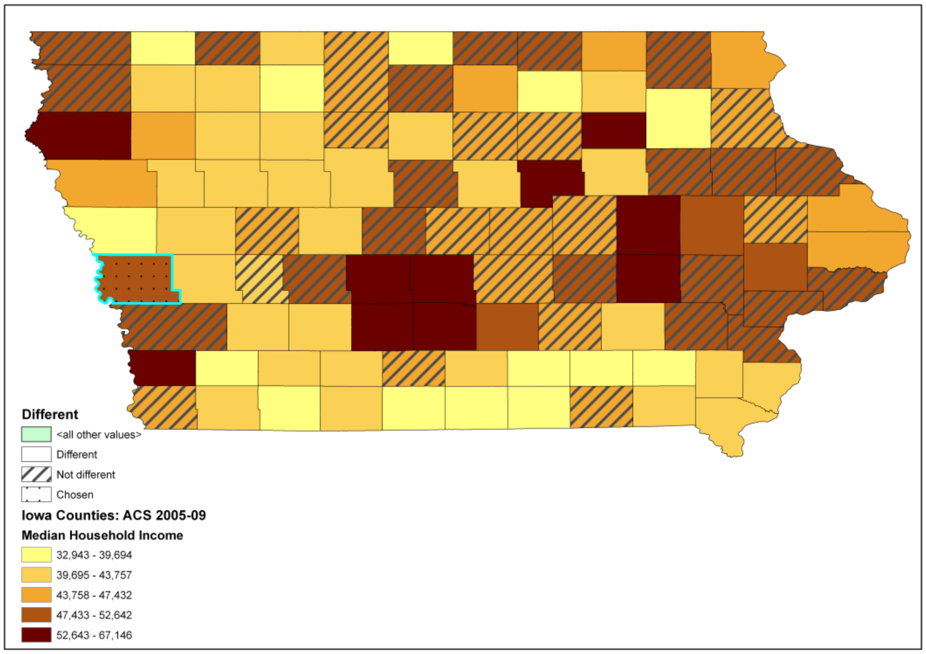

Figure 7: Counties with Median Household Income Estimates not Significantly Different from the Selected County, Harrison County, Iowa based on American Community Survey 2005-2009 5-Year Estimates

4. Issues in Mapping ACS Data

The approaches described in this paper address only some of the complexities involved in mapping ACS data. For example, the CV levels for the first mapping function are grouped into three classes ─ the default setting. These categories may not have a practical distinction. Currently, no guidelines are available to determine what CV level is "not acceptable," because such assessments are application-dependent.[14] To refine the precision of CV classes on the map, more classes need to be used, but texture overlay may not be able to accommodate more classes effectively. Other bivariate legend designs may have to be adopted to accommodate more classes for CV levels. The second mapping function allows the 1 x (N-1) comparisons, where N is the number of geographic areas in the study area. If more pair-wise comparisons are needed, the same procedure has to be repeated for multiple selected units. Such repetition is quite tedious and inefficient.

Several other issues affect the usability of these mapping approaches. First, they are most useful with maps for which the total number of geographic areas displayed is limited and simplicity in map design is important. In the case of Figure 7, the geographic areas displayed are counties, and there are only 99 in the geographic area of reference, the state

[14] "Esri Reliability Symbols Improve ACS Reports" identifies three levels of reliability based on specific coefficient of variation thresholds. However, those levels have not been evaluated and are not endorsed by the U.S. Census Bureau.

of Iowa, making the application of the approaches more feasible than would be the case had the number of geographic areas been far greater.

Another issue concerns the complexity of the display, and the ability of a user to comprehend the information conveyed. The interpretation of the maps shown in Figures 6 and 7 may be clear to an audience of statisticians, and technically preferable to such an audience. However, general audiences might find it more challenging to interpret these maps. In designing maps for these audiences, the benefit of simplicity in presentation probably outweighs the benefit of technical correctness, although a short statement supporting the appropriate interpretation of the data portrayed can usually be included without adding undue complexity or clutter to the map design. Such a statement can serve to demonstrate the map designer's responsibility for notifying the map user of the importance of interpreting the data displayed appropriately. Digital maps that include options for concealing or displaying information relating to the quality of the data displayed offer greater flexibility to the map designer.

A map of all 3,143 counties in the U.S. using either of these techniques is probably incomprehensible to most audiences. To address this issue, one could frame the statistical comparisons such that they are constrained within a state. Or, interactive visualization techniques could be used to supplement this tool. For example, interactively selecting a group of counties with reference to a preselected reference county could generate a pop-up table to reflect the results of statistical testing (the USDA map portrayed in Figure 1 includes an example of a pop-up table). The table could include the results of statistical testing that would compare the estimates for these counties with the estimate for the reference county, or compare the estimates for all counties (see Figure 8). This has a major disadvantage in that the use of such a table, as opposed to an overlay pattern, would not make it possible to discern areal patterns. Also, if too many comparisons are shown, the pop-up table becomes too complex and detailed to appreciate or understand easily.

County / State	Mercer County, NJ	Howard County, MD	Broward County, FL	Fairfax County, VA	Waldo County, ME	Chilton County, AL	Pima County, AZ	Kent County, DE
Mercer County, NJ	–	No	No	Yes	Yes	Yes	Yes	Yes
Howard County, MD	No	–	No	Yes	Yes	Yes	Yes	No
Broward County, FL	Yes	No	–	No	No	No	Yes	No
Fairfax County, VA	No	Yes	No	–	No	No	No	No
Waldo County, ME	No	Yes	Yes	No	–	No	No	Yes
Chilton County, AL	No	No	No	Yes	Yes	–	Yes	Yes
Pima County, AZ	No	No	No	No	Yes	Yes	–	No
Kent County, DE	Yes	No	Yes	No	No	Yes	No	–

Figure 8: Hypothetical Statistical Testing Results for Comparisons of Selected Counties: Statistically Significant Differences Between Estimates Noted by Affirmative ("Yes") Entries

5. Conclusion

Work on the development of the mapping tools described here has proceeded concurrently with participation in GIS-related conferences across the U.S. to explain the work underway to support this initiative. Presentations at the Applied Geography Conference (October 2010), the Association of American Geographers Conference (April 2011), the GIS in Public Health Conference (June 2011) and the ESRI Annual Users Conference (July 2011) were designed to explain the need for improved mapping in light of the issue of the statistical uncertainty of survey data in general, and ACS estimates in particular. Various technical issues associated with mapping ACS data have been discussed in a peer-reviewed published paper by one of the authors. A formal period to solicit input on the approaches and methods adopted in the ArcGIS extension began in early summer 2011, when the project website, http://gesg.gmu.edu/census/, was opened to the public to download the ArcGIS extension and to provide comments. A final version of the extension will be available in late Spring 2012.

Universities and other learning institutions where geography is offered as a major field of study, as well as federal agencies and the private sector, use GIS software to teach and practice mapping. By embracing and promoting the need for new approaches to map ACS data, the Census Bureau hopes that geographers in these institutions ─ and others who use GIS software to map survey estimates ─ will benefit. We hope this project advances the broad goal of statistical literacy that the Census Bureau promotes, and helps the Census Bureau’s data users understand the importance of cartographically representing ACS estimates appropriately.

References

Beard, M.S. and B.P. Buttenfield. 1991. Geographic Information and Analysis NCGIA Research Initiative 7: “Visualization of Spatial Data Quality.” NCGIA, Technical Paper: 91-96. Accessed August 31, 2011 from http://www.ncgia.ucsb.edu/Publications/Tech_Reports/91/91-26.pdf.

Citro, C.F. and G. Kalton (eds). *Using the American Community Survey: Benefits and Challenges*. (Washington, DC: The National Academies Press, 2007).

Environmental Systems Research Institute. “Esri Reliability Symbols Improve ACS Reports.” Esri Business Analyst Online. Accessed August 31, 2011 from http://www.esri.com/software/bao/reliability-flags.html.

Federal Geographic Data Committee. 1998. *Content Standard for Digital Geospatial Metadata (version 2.0)*. Accessed August 31, 2011 from http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf.

Sun, M. and D.W.S. Wong. 2010. “Incorporating Data Quality Information in Mapping American Community Survey Data.” *Cartography and Geographic Information Science*, 37(4): 285-300.

The Washington Post. "Percentage of Children Living Below the Poverty Line in 2009." September 29, 2010. Accessed September 26, 2011 from http://www.washingtonpost.com/wpdyn/content/graphic/2010/09/29/GR2010092901799.html?sid=ST2010092901802.

U.S. Census Bureau. *A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know*. Accessed August 31, 2011 from http://www.census.gov/acs/www/Downloads/handbooks/ACSGeneralHandbook.pdf.

U.S. Census Bureau. *American Community Survey Reports, ACS-08, Income, Earnings, and Poverty Data from the 2006 American Community Survey*: 6. Accessed August 31, 2011 from http://www.census.gov/prod/2007pubs/acs-08.pdf.

U.S. Census Bureau. "TIGER Products." Accessed August 31, 2011 from http://www.census.gov/geo/www/tiger.

U.S. Census Bureau, *U.S. Census Bureau Section 515 Information Quality Guidelines, 2002, reissued 2011*. Statistical Quality Standard E-2, "Reporting Results." Accessed August 31, 2011 from http://www.census.gov/quality/standards/index.html.

U.S. Census Bureau. "Unpublished Notes of the U.S. Census Bureau's Statistical Mapping Workgroup." 1994.

U.S. Department of Agriculture. Economic Research Service. *An Atlas of Rural and Small Town America*. Accessed August 31, 2011 from http://www.ers.usda.gov/Data/RuralAtlas/atlas.htm#map.