Supporting Planners' Work with Uncertain Demographic Data

Amy L. Griffin,*[1] Seth E. Spielman,[2] Nicholas T. Nagle,[3] Jason Jurjevich,[4] Meg Merrick,[4] David C. Folch[2]

[1] UNSW Canberra

[2] Colorado University at Boulder

[3] University of Tennessee-Knoxville

[4] Portland State University

ABSTRACT

Maps are often made without the explicit inclusion of the uncertainty in the mapped data. Map readers, in turn, often assume that mapped information is accurate and certain. Developments in the methodology used to generate population estimates in the United States have led to an increase in the attribute uncertainty associated with many small area estimates. Therefore, adequately communicating this uncertainty is of increasing importance to making quality decisions based upon population estimates. We report here on a context-of-use study that informed the design of a visualization that planners can use to understand the uncertainty in mapped estimates. We are currently undertaking a user study to assess the extent to which planners a) understand the mapped uncertainties from the visualizations and b) draw upon the uncertainty information when making decisions with the uncertain data.

Keywords: uncertainty, geovisualization, urban planning.

Index Terms:

1 INTRODUCTION

Sample surveys such as population censuses produce estimates of demographic characteristics of populations. Each of these estimates has an uncertainty associated with its value and responsible use of the data requires an understanding of its uncertainty. Therefore, it is common for statistical agencies to publish margins of error alongside demographic estimates. A key question lies in whether or not users of population estimates pay any attention to these margins of error when determining whether the data are fit for the use to which they would like to put them.

An answer to this question is needed more urgently than ever because the US Census Bureau has recently made major changes in the methods they use to generate US population statistics that have led to an increase in the attribute uncertainty of population estimates. Prior to 2010, estimates for most demographic characteristics were generated through the decennial census' long form, which sampled approximately one in six households across the country every ten years. Since 2005, the US Census Bureau has collected data annually through the American Community Survey (ACS), with approximately one in forty households sampled each year. These yearly estimates are then aggregated into multi-year estimates to produce more certain data. The problem of attribute data uncertainty is most acute for one-year estimates for the smaller spatial units (Fig 1). Therefore it has become more important than ever for users of census statistics to examine margins of error to determine fitness of use for their application.

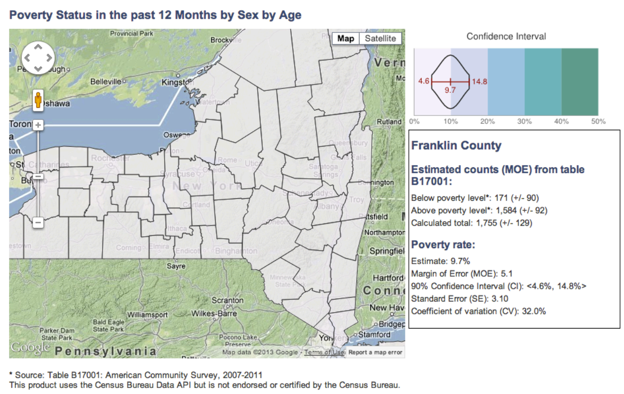

Figure 1: An example of data from the American Community Survey with a very large margin of error, making it impossible to tell which class the estimate actually falls within. Figure generated from a web-mapping application from the Cornell Center on Applied Demographics. http://pad.human.cornell.edu/Unlisted/uncertaintymap_fullinfo_api_B17001.cfm.

While a fair amount of attention has been directed to the development of representations of uncertainty in geospatial information, relatively less research has worked towards understanding what users take away from looking at these uncertainty representations [1]. Even fewer studies have been contacted 'in vivo': in the context of real-world tasks, situated in real work environments among end users whose work demands consideration of the uncertainty in the information they work with [2]. The overall aim of our broader research project is to understand how to best support the ability of urban planners, who are 'power users' of census population estimates, to understand uncertainty represented in maps of ACS data so that they can establish fitness for use. Here, we report on results of our context of use study and the design of our prototype visualizations of uncertainty in mapped population estimates.

2 CONTEXT OF USE STUDY

To understand context of use, we took a multi-method approach, conducting semi-structured interviews (n = 7), analysis of artifacts generated from ACS data (e.g., reports, presentations, fact sheets and other publications), and a survey questionnaire (n

* a.griffin@adfa.edu.au

= 214) to produce more generalizable results from a broader range of participants.

The semi-structured interviews have allowed us to identify the types of tasks for which the planners in our sample use ACS data. Typical uses included [rank in frequency of response by the subsequent survey questionnaire]:

- Creating community profiles [3]
- Comparisons of the community over time [1]
- Comparisons to other communities, especially when competing for grants [2]
- Developing a transportation master plan [5]
- Identifying transport dependent populations or choice commuters [4]
- Describing county to county commuter flows [6]

Our artifact analysis turned up very few examples of inclusion of uncertainty estimates in maps made from ACS data and distributed to the public or decision makers. We sampled a selection of publicly available reports, community profiles, atlases, and other documents produced by planners from a wide variety of communities.

Planners' failure to use or communicate uncertainty information alongside ACS estimates is driven largely by their perceptions of what their clients want or need. In some cases, they believe or have been explicitly told that uncertainty information should not be included in grant application submissions:

"*They just want the number*. That's all they care about….So if there's uncertainty in the data, people may not have confidence in it….*And the grant apps, they're not asking for the margins of error. They're not asking about uncertainty*. They're just comparing that application versus somebody else's study area, and they're going well…we're going to award this one because it's got better numbers." (I2)

This problem is compounded by the fact that at least some planners do not have a good understanding of ACS uncertainty information to make valid comparisons between places. This task requires explicit attention to the uncertainty in the population estimates for each location being compared. For example, one respondent noted:

"Any good statistics class, software, person who just does statistics will show… you have to include a margin of error when you do the type of sampling. *However, we just don't use it* [the margins of error]. Nobody….unless you're a statistics type person presenting to statistics professors where you have to have your footnotes in there…for the actual real world studies, what I said is the case. *If you're comparing ACS to ACS, it really doesn't matter. They're going to have the same margin of error, more or less*." (I6)

This was confirmed in our follow-up survey when *less than half* of survey respondents agreed with the statement: "Demographic and economic estimates from the American Community Survey are only suitable for making comparisons between places if margins of error are considered."

3 Conclusions and Future Work

Some of the most important results from the context of use study include:

- There are specific tasks for which planners could benefit from effective methods for representing attribute uncertainty information, especially forms of benchmarking, such as *comparisons between places* (tract to tract, city to city, city to state, city to nation) or *comparisons over time*.
- Planners currently use ACS attribute uncertainty information in very limited ways.
- At least some planners have important misconceptions about attribute uncertainty in American Community Survey data.

Based upon what we learned in the context of use study, we designed a series of controlled experiments that required our user group to make comparisons between places using a task that planners commonly have to undertake: defining an economic improvement zone based upon a set of eligibility criteria. These currently in-progress user studies with planners compare (infrequently) but currently used visual representations of attribute uncertainty (i.e., current 'best practice' among the planning community) with our own representation designs (Fig 2).

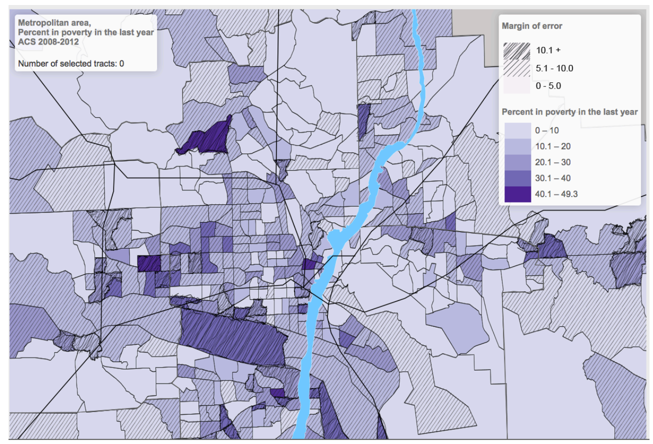

Figure 2: Example experimental stimulus used in our controlled experiments. The uncertainty representation draws upon the 'sketchy' library developed by [3].

References

[1] A. M. MacEachren, A. C. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan. Visualizing geospatial uncertainty: What we know and what we need to know. *Cartography & Geographic Information Science*. 32: 139-60, 2005.

[2] D. Lloyd, J. Dykes. Human-Centered Approaches in Geovisualization Design: Investigating Multiple Methods Through a Long-Term Case Study. *IEEE Transactions in Visualization and Computer Graphics*, 17: 2498-2507, 2011.

[3] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, A. Slingsby. Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics,* 18: 2749-2758, 2012.