MAPPING CENSUS 2000 IN THE CENSUS ATLAS OF THE UNITED STATES

Brewer, C.A., Suchan, T.A. and Tait, A.

Department of Geography, Pennsylvania State University. E-mail: cbrewer@psu.edu
Population Distribution Branch, U.S. Census Bureau. E-mail: trudy.a.suchan@census.gov
International Mapping Associates, Ellicott City, Maryland. E-mail: alex@ima-maps.com

ABSTRACT

The U.S. Census Bureau's *Census Atlas of the United States* is intended for a general-interest audience. It will comprise maps, text, and figures selected and organized to increase public understanding of population and housing characteristics across the United States. It will seek to answer the map users' natural question, "What is my place like compared with others?" We view the atlas as an icon for what the Census Bureau does—collects data that provide a continuous record of the U.S. population and captures the cultural era, as the classic atlases of U.S. Census data in the late 19th century did.

The 2003 atlas will include interpretations of population and housing data from 100-percent and sample data from Census 2000. In addition to representations of the total population, topics include characteristics of urban and rural populations, immigration, citizenship, ancestry, age groupings, sex, family composition, housing, income, and occupation. These data will primarily be presented using nationwide maps at the county scale, but the atlas will also include maps of other geographic areas, such as American Indian areas, metropolitan areas, and selected cities.

We are in the midst of preparing the atlas and will discuss the strategies we have taken for color use and data classification for choropleth mapping. We are using a wide variety of color schemes on maps and organizing them with signature colors for basic population characteristics such as age and sex. These colors will cue comparison across map series and chapters. We are classing data with the goal of emphasizing comparison of maps within series. We are using four strategies during classification: rounding aggressively, using meaningful breaks, sharing breaks, and masking unstable values.

The atlas will be published in early 2004 as a large-format bound book and on the Web.

1. INTRODUCTION

As we prepare this proceedings paper, production of the *Census Atlas of the United States* is in progress. We have been working on topic selection, page layouts, overall book design, and production of maps and graphs. Since we are midway through production, we will describe our decision making on some details of map preparation. For this paper, we present two topics: color scheme planning and choropleth map classification. First, we begin with an overview of the project background.

2. BACKGROUND

The U.S. Census Bureau's *Census Atlas of the United States* is intended for a general-interest audience. It will comprise maps, text, and figures selected and organized to increase public understanding of population and housing characteristics across the United States. It will seek to answer the map users' natural question, "What is my place like compared with others?" The *Census Atlas* will be completed in 2003 and will go to press early 2004. Until Census 2000 (the 2000 decennial census), the U.S. Census Bureau had not published a decennial statistical atlas since the 1920s. In 2001, the impact of a collection of maps, including 10-year change maps, was demonstrated in *Mapping Census 2000: The Geography of U.S. Diversity* [1]. Brewer and Suchan designed and produced that book as a prototype for the more comprehensive *Statistical Atlas* we will present at the ICA meeting.

Many maps in newspapers and online have illustrated Census 2000 stories, but these maps tend to emphasize either a local geographic area or limited topics. The U.S. Census Bureau has the full range of data for 2000, access to historical data that will provide temporal context for Census 2000 results, plus subject-matter experts. These resources have enabled us to produce an atlas that will give a lively and coherent picture of U.S. population and housing at the turn of the 20th century. We view the atlas as an icon for what the Census Bureau does—collects data that provide a continuous

record of the U.S. population and captures the cultural era, as the classic census atlases of the late 19th and early 20th century did [2].

The 2003 atlas will include interpretations of population and housing data from 100-percent and sample data from Census 2000. In addition to representations of the total population, topics include characteristics of urban and rural populations, immigration, citizenship, ancestry, age groupings, sex, family composition, housing, income, and occupation. These data will primarily be presented using nationwide maps at the county scale, but the atlas will also include maps of other geographic areas, such as American Indian areas and census tracts for metropolitan areas and selected cities. The atlas will include maps of change using historical decennial census data. As examples, the atlas may include map series showing all places ever over 10,000 population from 1790-2000 and rural population comparisons between 1900 and 2000.

Like *Mapping Census 2000,* the atlas effort will yield multiple final products. The print volume will be hard bound, with a large page size and about 300 pages. The *Census Atlas* will also be available on the Web in PDF format, and related products used to produce the atlas will be concurrently released to the public, such as data files, graphics files, and generalized geography files.

3. COLOR SCHEMES

As we planned the atlas, the group of census researchers, cartographers, and graphic artists had spirited discussions about how to structure the atlas using map color schemes and signature colors associated with sections and/or chapters. In early discussions, the group looked at 'approved' Census colors (a set of rich colors with Pantone ink definitions) and talked about using a few to distinguish atlas Parts (preliminary plans split the atlas into three Parts with about five chapters per Part). Differently colored gradient lines below titles, as used in the *Atlas of Oregon* [3], was one example use of signature colors.

We spent some time talking about the three basic types of color schemes (sequential, diverging, qualitative) and showing ColorBrewer examples <www.ColorBrewer.org> [4] to be sure the designers understood that many colors would be within the maps. We wanted them to understand that they should not design with the assumption of particular colors in all maps or that they could arbitrarily alter map colors to suit a design plan. We had been working with layout templates repeatedly shown with purple and yellow-green-blue schemes. We were concerned that this stability in the look of examples might over-influence design planning if we did not clarify the intended variety.

The main options we debated for structuring color in the atlas are described below.

3.1 Use a signature color scheme per chapter.

A disadvantage of single schemes per chapter is that fifteen different but all pleasing color schemes would be difficult to design. Thus, some chapters would be declared 'bad' by opinionated critics of particular hue choices. A variation on this idea would be to use related schemes with similar hues by atlas Part. For example, we might use a variety of schemes with blue in them for Part 1, variations on green in Part 2, and variations on red in Part 3. We were worried, however, that we would not achieve much variety in actual colors given this constraint, making the differences too subtle to look intentional.

3.2 Use the same color scheme within map series.

Series are usually limited to a few pages of maps, and chapters contain multiple series of related maps. Using the same colors within series (and different colors between series) would produce chapters with a variety of color schemes. Sets of related maps would be visually linked by shared colors. This plan would reinforce series relationships and be easy for a reader to understand. On the other hand, this structure might unnecessarily inhibit the design because links within series may be sufficiently established by map proximity and titling.

3.3 Use color to link maps between series.

Through discussion, we selected some basic 'immutable' characteristics of people that could reasonably be used to structure color choices throughout the book. This option would create attractive pages because they would have multiple schemes on a page and would be colorful.

We felt that the characteristics with potential to be linked to identifying colors were:

- age (young, working, old)
- sex (male, female)
- race/ethnicity (one scheme for all race maps; we did not pick a color for each race)

We would then use a limited set of schemes for variety in all other topics, and perhaps save some of the most attractive for this role. Altogether, this strategy would produce about ten sequential schemes used through the book.

3.4 Use a scheme for each type of data processing.
We would use one color scheme each for the types of maps listed below:

- density
- percent of total
- percent of some subset (e.g., percent of all children)
- ratio (e.g., dependency)
- number of people (raw counts)
- difference in number
- index (e.g., diversity calculation)
- qualitative (e.g., prevalence)

This strategy might aid interpretation by cueing readers to a basic characteristic of the data processing, rather than the population topic. A disproportionate number of maps would be percentage maps, but choosing some of the most attractive schemes for this category might be satisfactory.

3.5 Draft color choices
We opted for the third choice (3.3 above) for organizing color use in the atlas. We used color to link maps between series. We used ColorBrewer schemes [4] as a starting set to plan color schemes for the atlas. We used more hue transition schemes because they add more hue variety to a page of sequential maps.

We planned signature colors for the 'immutables' (basic characteristics of people) that would be consistent throughout the atlas:

- Age series followed each other nicely with the last hue for one age group beginning as the light color for the next age group: young, yellow to dark green; working age, light green to dark blue; old, light blue to dark purple (note the yellow-green-blue-purple flow through these three schemes).
- Sex schemes are basically red and blue sequential schemes with a bit of pizzazz (light orange to dark red and light to dark blue with a red-white-blue diverging scheme for sex ratio data).
- Nativity data uses two basic schemes for native (yellow-orange-dark red) and foreign (light blue to dark green).
- Race/ethnicity series will be shown in the most colorful of the sequential schemes (yellow-green-blue-purple) to allow pages to be colorful with comparable maps. We debated showing each group with a signature one-hue sequence (e.g., purples for African American, greens for Hispanics, etc.) but felt that would get too busy for seven-group sets.

If a map series involves two immutable characteristics, we will need to choose one scheme set over another given the primary story told with the data. For example, if age and sex are mapped, we will visually emphasize the age or the sex dimension in the color selection. This will be a subjective decision, partly chosen to add variety to a chapter (to differ from schemes that preceded the map series).

Beyond sequential schemes, we planned for a few diverging schemes when the data mapped has an obvious critical value midway through the data range. For example, we agreed on using the default purple-white-green diverging scheme. We nixed the green-yellow-red diverging because of the difficulty it will present to colorblind readers. We chose a qualitative scheme (Set 2 in ColorBrewer [4]; not too pastel, not too primary), but we realized we will need a much longer qualitative set for the ancestry data.

Time series data will be marked with more neutral schemes. Alex prepared sepia and warm gray sets to complement the yellow-orange-brown sequential scheme. We noted that a time series scheme to complement a diverging scheme was also needed, and we discussed using a desaturated version of the 2000 scheme as a more general strategy. For example, using grayed red-blue to go with a saturated red-blue sex ratio scheme. We expect this color work will be challenging.

Color schemes are only part of the look of choropleth maps. Data classification also affects their appearance, and we will discuss our classing decisions in the next sections.

4. CHOROPLETH CLASSIFICATION

When we first look at raw choropleth maps of new data for the atlas, we class them with natural breaks from ArcGIS [5]. Those initial maps give us a good starting view of the data distributions. The basic strategy we recommended to our cartographers (being sure to emphasize that this is partially opinion) was to work with four goals: round aggressively, use meaningful breaks, share breaks, and mask unstable values. These basic ideas grow from our work on *Mapping Census 2000* [1] and are described in Brewer [6]. How far we take the idea of comparability will be interesting. We may be able to accomplish comparability between some chapters but working toward comparability throughout the book would be too difficult.

We worked on this process by demonstrating classing for a few series and then critiquing classes prepared by Jim Miller, the cartographer taking the lead on the classification decisions map by map. He also made use of data histograms (provided by ArcMap) to augment his planning.

The goal of classification was facilitating comparison across maps in series; not optimizing individual maps but not compromising them either. Thus, our look at the default natural breaks maps first. As we prepared map series classifications, we watched that class adjustments for series comparison did not radically change map patterns, missing the within-map stories.

4.1 Round aggressively

If you are working from natural breaks classing (or quantiles if that seems a suitable strategy), do not hesitate to round class breaks to make the legend more usable, making numbers easier for readers to think with. And round aggressively. We are talking about yanking a number like 337 up to 500 (not small rounding like 30.3 rounded to 30.0; see Figure 1).

4.2 Meaningful breaks

Use meaningful breaks across series where possible (Figure 1). The obvious choices are U.S. overall rates for the data in the series. Other meaningful breaks are zero change, even ratio, 50%, twice or half the U.S. rate, and thresholds like the poverty rate or frontier density.

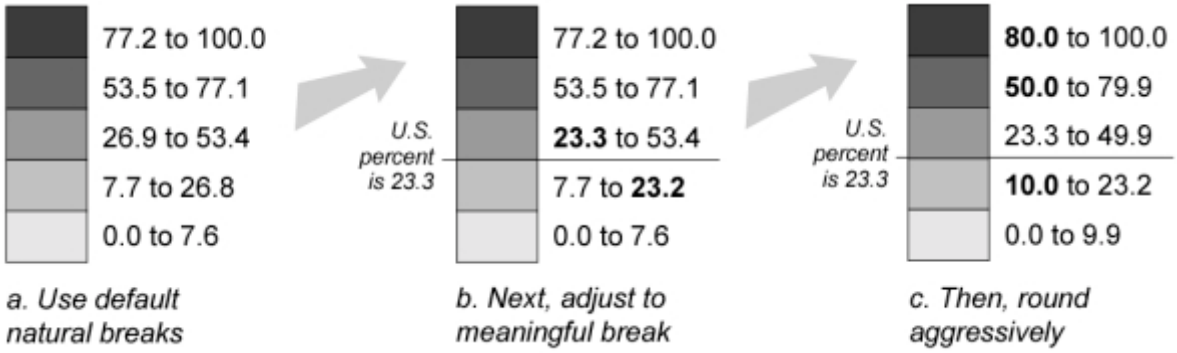

Figure 1. Adjust classes by replacing breaks with meaningful values relevant to the data (such as the U.S. overall percent, twice U.S. percent, and other meaningful breaks). Round aggressively to make the natural breaks default classes easier to use.

Construct diverging schemes that hinge on a meaningful break in the data classing. Use of a diverging scheme is an obvious choice when there is a zero or even ratio (1:1) midway through the data range that can be marked by the break between lightest classes. Data series suited to sequential schemes may also be seen as diverging. Consider using an overall U.S. percent or rate for the total population as the critical value that is the hinge in a diverging color scheme (Figure 2). We did this with the "Percent Under Age 18" maps in the *Mapping Census 2000* prototype [1].

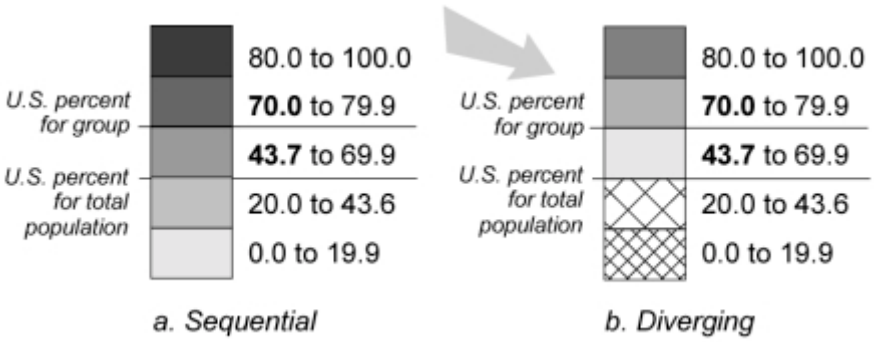

Figure 2. Consider using a total population critical value for building a diverging scheme.

4.3 Share breaks

Using a set of overall U.S. percentages to class a set of maps is easy if they are well separated, such as 20.1, 40.7, 60.5. This set of three U.S. rates, with one or two other shared breaks to complete map patterns, would rapidly build a complete set of breaks useful on all maps (Figure 3). It is not usually that easy though.

Map series with many maps and U.S. rates that are close together are more challenging (Figure 4). The strategy here is to develop a set of classes that are shared by all maps and then adjust individual maps to include a critical break, deviating from the general class set by a modest amount. Looking at potential groupings within the set critical values can help this process. For example, a map series that includes U.S. percents of 18.5, 20.1, 22.7 and 35.0, 42.3, 44.7 could have breaks at 20 and 40. On the map with a U.S. percent of 42.3, it would replace 40 and remaining breaks would include the round value of 20 (Figure 5). Alternatively, the maps could include breaks at 20, 35, and 45 if more classes seemed like a better solution for the series.

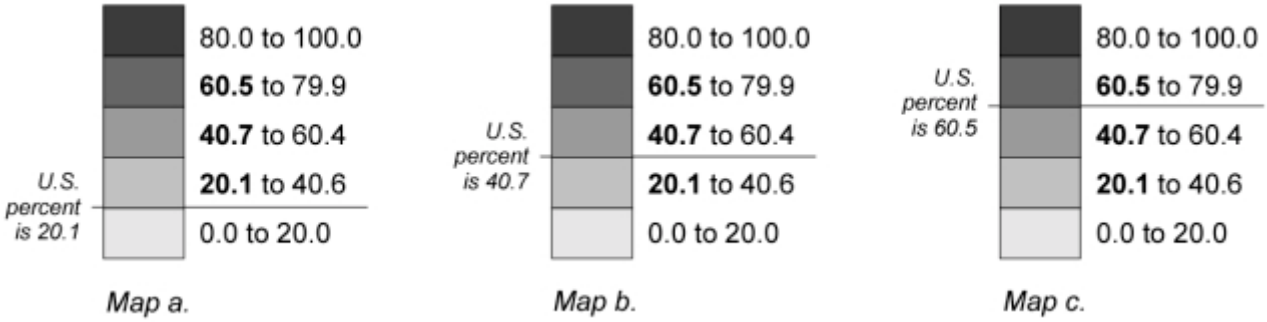

Figure 3. All breaks are shared among maps within a series of three for this example. Bold values are determined by U.S. percentages for individual maps.

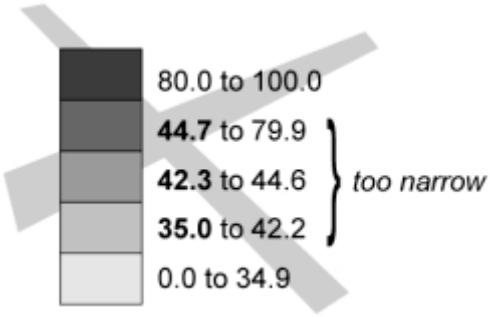

Figure 4. Shared breaks may produce classes that are too narrow. There is no fixed definition of 'too narrow.' The meaningfulness of a data range is based on the larger context of the mapped data.

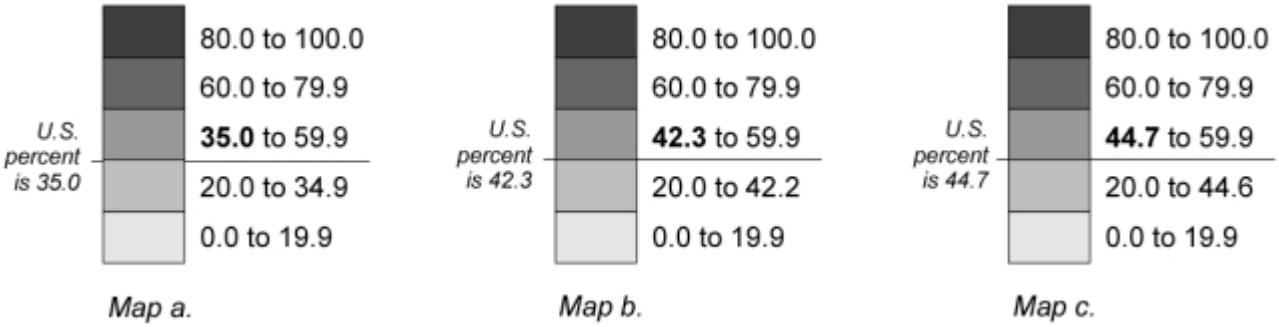

Figure 5. Shared breaks for comparison. Classes are adjusted to include a meaningful break for each map in this example series of three.

Map series may include pairs that you want to be completely comparable, with identical classing (Figure 6). The classes throughout the series may then be more generally comparable, with similar but not equal classing. This strategy is similar to that shown in Figure 5 with more constraints.

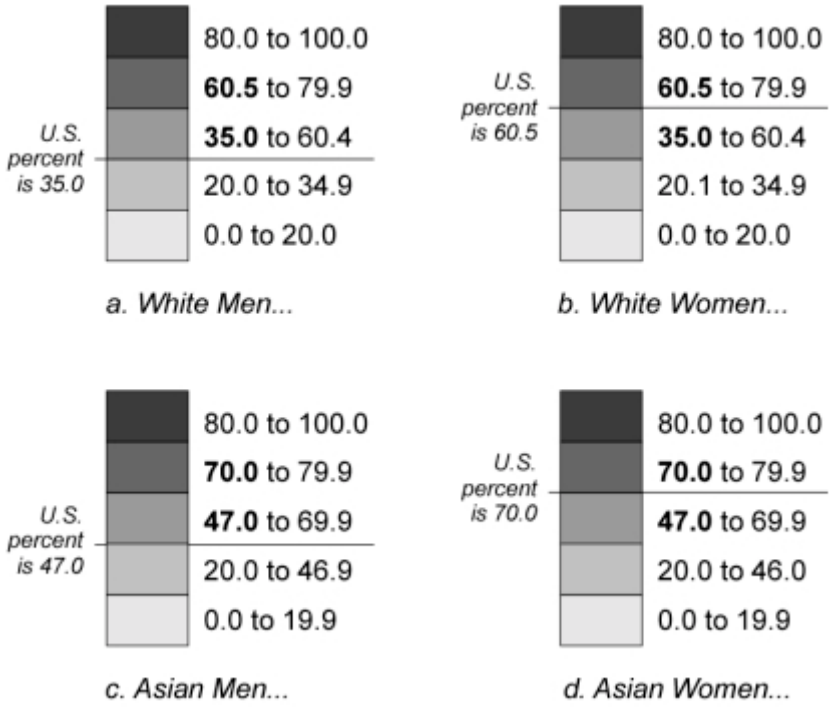

Figure 6. Attend to pairs within series. You may decide that particular pairs need to be directly comparable, with other maps more loosely comparable in the series. In the example, the same-race pairs (a-b, c-d) have the exact same breaks in the series. (These hypothetical classes are not based on race data.)

Narrow classes may be necessary for particular maps in a series, but you may not want to use them on all maps. An example situation is when a U.S. ratio for one group is close to the even ratio and you want to include both breaks on the map (Figure 7). You can insert a class for an individual map and hold room in your color sequence on all maps for this added color. Inserting a color is preferable to shifting color (or lightness) meanings up or down a class between maps. That change would interfere with comparing maps through the series (note that in Figure 6 the dark gray represents 105 to 150 on all maps).

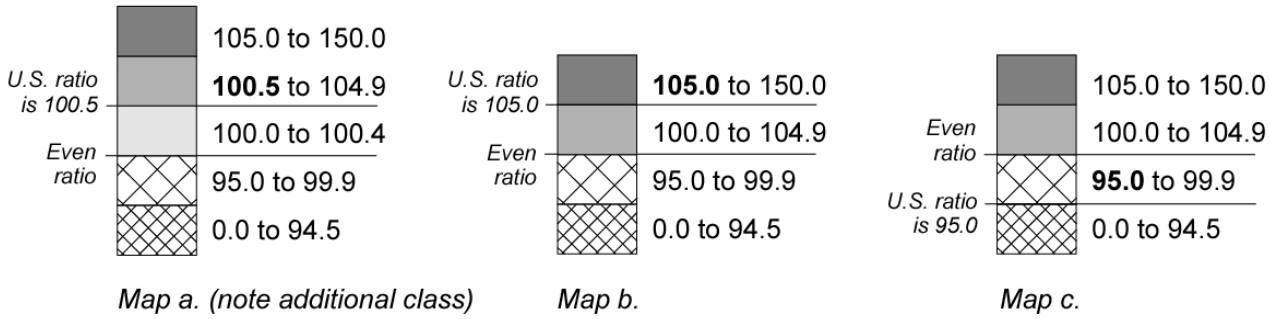

Figure 7. An individual map may be better classed with an extra narrow class that does not seem useful on other maps in the series. A narrow class is added to the first diverging legend shown (100.0 to 100.4). A new gray is added for only this class, maintaining magnitudes for the darker classes among the set of three.

Very narrow classes at the bottom of the data range, such as for mapping small populations of Native Hawaiian and Other Pacific Islander (NHOPI), can be treated the same way. Include a new color, added to the scheme, rather than bumping all colors down a class to squeeze in that added break (we added a pink below yellow for NHOPI maps in *Mapping Census 2000* [1]).

4.4 Masking

Setting criteria for masking counties with low numbers is a decision related to data classing. The goal behind this map augmentation is to mask extreme derived values (such as percentages, ratios, and rates) produced when calculations are made with small base populations (i.e., small denominator or universe). The values being masked are not incorrect, but

their extreme highs and lows interfere with seeing patterns in the distribution of more stable values for counties with larger denominator populations. In *Mapping Census 2000* [1] we used 100 people in a county as the threshold for masking low numbers, and that number was widely approved by managers critiquing draft maps. Masked counties were shown in white.

We realized that maps of sample data (from the Census 2000 'long form') may require a higher threshold because data are based on responses from approximately 1 in 6 households. Data for 100 people are extrapolated from the answers of approximately 17, which is a small number that will produce potentially unstable summary values. For household sample data, increasing the threshold to perhaps 200 or 300 households is appropriate. We have not yet decided on a threshold.

We also discussed the need to remove maps from the atlas that have too many counties masked. We suggested omitting a map that has more than ten percent of counties below the threshold (approximately 300 counties). This suggestion was tempered with the acknowledgement that maps that are part of a series may need to stay in the atlas even though they have more than ten percent of counties masked because they are needed to complete the message embodied by a series. For example, an increase in population that pulls increasing numbers of counties above the threshold, to become part of mapped data distributions, is an interesting aspect of a time-series map sequence. The unmasking of increasing numbers of counties over time is preferable to having heterogeneous patterns on early maps caused by instability of small-number calculations.

The locations and number of low-population counties shown by masking provides information about the mapped population that complements the distribution of mapped values. Maps with large masked areas should not be considered poor maps. The size of masked areas may also be larger than expected, even with less than ten-percent masking, because low-population rural counties cover more land area than small dense urban counties in the United States. Masking these large counties improves the overall map because their unstable extremes form large blocks of distracting color that provide unreliable information. It is better to know that there are small numbers in the group mapped in these large areas than to suggest unexpected or extreme characteristics over large areas.

5. SUMMARY

We have been thinking carefully about color use and data classing as we prepare the new *Census Atlas of the United States*. When we present our work at the ICA meeting in August, we will be showing examples of our final decisions. Look for the printed book and web version after publication to see how our decisions were constrained by the realities of the mapped data and atlas design. And read the maps to see the fascinating portrait of the U.S. population that the atlas presents.

6. ACKNOWLEDGEMENTS

This work is funded by the Population Division of the U.S. Census Bureau. James Fitzsimmons, John Long and Nancy Gordon at the Census Bureau are key to making this atlas a reality. For the portion of the work described in this paper, we would like to acknowledged Jim Miller at International Mapping Associates in Maryland for producing map series on which we developed and debated these ideas.

7. REFERENCES

[1] C.A. Brewer and T.S. Suchan, "Mapping Census 2000: The Geography of U.S. Diversity" (Census Special Report CENSR/01-1), pub U.S. Government Printing Office (2001). <www.census.gov/population/www/cen2000/atlas.html>

[2] H. Gannett, "Statistical Atlas of the United States" (based on 12th Census), pub United States Census Office (1903).

[3] W.G. Loy, S. Allan, A.R. Buckley, and J.E. Meacham, "Atlas of Oregon", pub University of Oregon Press (2001).

[4] C.A. Brewer, G.W. Hatchard, and M.A. Harrower, ColorBrewer in Print: A Catalog of Color Schemes for Maps, Cartography and GIS, 30 p.5 (2003).

[5] C.A. Brewer and L.W. Pickle, Comparison of Methods for Classifying Epidemiological Data on Choropleth Maps in Series, Annals of the Association of American Geographers, 92 p.662 (2002).

[6] C.A. Brewer, Reflections on Mapping Census 2000, Cartography and GIS, 28 p.213 (2001).

MAPPING CENSUS 2000 IN THE CENSUS ATLAS OF THE UNITED STATES

Brewer, C.A., Suchan, T.A. and Tait, A.

Department of Geography, Pennsylvania State University. E-mail: cbrewer@psu.edu
Population Distribution Branch, U.S. Census Bureau. E-mail: trudy.a.suchan@census.gov
International Mapping Associates, Ellicott City, Maryland. E-mail: alex@ima-maps.com

Biography

Cynthia Brewer is an Associate Professor at the Pennsylvania State University. She completed PhD and MS degrees at Michigan State University with Judy Olson. Her BA was in Fine Arts at the University of Guelph (Ontario Canada). She began teaching at Penn State in 1994. Previously, she was an Assistant Professor at San Diego State University (1991 to 1994) and a Visiting Lecturer at the University of California at Santa Barbara (1986/87).

Cindy's research and publications have focused on map design, color theory, hypothesis generation, and choropleth classification with applications in epidemiology. She has worked on contracted research for the National Center for Health Statistics, U.S. Census Bureau, BioMedware, and the National Park Service. She has presented her work at over 30 conferences on geography, cartography, and color science.

She has been a member of the Editorial Board for the Professional Geographer, Cartography and GIS, and Cartographic Perspectives. She was guest editor for the 1999 U.S. National Report to the International Cartographic Association. She was also Past President of North American Cartographic Information Society.