CENSUS STATISTICAL MAPPING AND THE USERS

Robert C. Klove
U.S. Bureau of the Census

INTRODUCTION

Cartography at the Census Bureau is concerned with three types of map work, mapping for enumeration, the preparation of statistical area identification maps, and what is generally known as statistical mapping. It is with the third type of mapping that this paper deals and the focus is on our experience in planning and designing census statistical maps rather than on methods and techniques of production. These planning and design problems are closely related to the map reading and perception problems of the users.

My presentation is organized under three general headings: (1) Statistical mapping by the Census Bureau, (2) Who are the users and what kind of maps do they need? and (3) Statistical map design problems and the users, or do the maps tell the story they should?

STATISTICAL MAPPING BY THE CENSUS BUREAU

Census statistical cartography goes back over one hundred years to the first Statistical Atlas produced by Francis Walker after the 1870 census. It reached a peak in 1890 under Henry Gannett and then gradually declined to a low level in the twenties and thirties. It began to recover in the forties and fifties, burst into color again in the sixties, and in the seventies is becoming computer generated and user oriented. This capsule history of statistical mapping at the Census Bureau is covered in more detail in an earlier paper by the author.*

The revival of statistical mapping at the Census Bureau was probably due to many factors. One of these undoubtedly was an increasing awareness of user needs. Another was the strengthening of the professional staff and particularly the geographic staff. This was accompanied by a revival of interest in area measurement and resulted in the publication of a large two-sheet color map showing population density by minor civil divisions in the United States after the 1940 census. With the 1945

*R.C. Klove, "Statistical Cartography at the Bureau of the Census", International Yearbook of Cartography, VII, 1967, pp. 191-199 (Gutersloh, C. Bertelsmann Verlag).

Census of Agriculture, the Census Bureau assumed responsibility for preparing and publishing agricultural statistical maps for census reports, a task earlier undertaken by the Department of Agriculture. For the 1950 Census of Population a number of black and white statistical maps were prepared for the summary volume and a large single-sheet population distribution map of the United States in color was published. Some statistical maps were introduced into the reports of the economic censuses in the fifties, particularly the 1958 Census of Manufactures. With 1960, statistical maps in color were introduced in the summary volume of the Census of Population. Then in 1963 came the initiation of the large separate United States maps in color, familiarly known as the GE-50 series. These maps were developed to illustrate some of the more important statistical distributions found in each of the various censuses, both demographic and economic. We, who originated this map series, felt that there was a public need for maps of this type, and we sought to make them as useful and attractive as possible. Much of the remainder of this paper is concerned with our experiences in developing this particular set of maps while I was associated with the Geography Division.

WHO ARE THE USERS?

We know something about our statistical map users, but we need to know a great deal more. Some we have talked with; others have written to us. Most of the separate maps as well as those in publications have been sold by the Government Printing Office. We do know that in the early years of issuance of the GE-50 map series several thousand copies of most of the maps were sold. Many have also been distributed to depository libraries.

The users are distributed in about five groups: (1) the academic world, including teachers and students, (2) Federal, State, and local governments (3) economic activities, especially marketing but also production, (4) nonprofit institutions for research, and (5) others. We also know that the maps are used for at least three purposes: (1) as a reference, or to understand quickly the geographic distribution of the statistics, (2) for research, or to relate other distributions to the one in hand, and (3) as a tool in teaching.

During the last decade a concerted and greatly increased effort has been made by the Bureau of the Census to find out more about the users of census statistics and their data needs. Since the mid-sixties the Bureau has had a Data User Services Division which acts to assist users and to determine their needs. This division has publicized the maps and exhibited them but has not been in a position to determine how statistical maps are used and how they could be improved. The Geography Division has provided and manned exhibits of its maps at annual meetings of academic associations with increasing frequency in recent years. Through such activities only small segments of the users and potential users are reached. It is frequently discovered that there are conflicting uses and that it is difficult to determine how statistical mapping can be improved to satisfy most users.

The Geography Division conducted a small survey in 1973 primarily to find out how users of the GE-50 United States map series might react to a reduction in the scale of these maps from 1:5,000,000 to 1:7,500,000 with comparable reduction in

sheet size. Printed samples of maps at both scales were submitted and the letter indicated that the Geography Division preferred the smaller size. Letters were sent to 53 persons. Included were editors of publications which have given publicity to these maps in the past, selected Federal government officials, and officials of other organizations with map interests, but the major group consisted of university professors of geography with a known interest in cartography. Of the 31 who replied, 18 were for reduction, 11 favored the present scale, and 2 gave no preference. Those for reduction noted that the smaller scale still provided the same information and that the smaller sheet sizes were easier to examine and work with and also easier to file. Several mentioned as an advantage that they were the same scale as the National Atlas maps of the U. S. Geological Survey. Those favoring holding the present scale of 1:5,000,000 generally expressed themselves in a strong manner, perhaps to make their views clear in opposition to the preference expressed for the reduction by the Geography Division letter. Their argument was clear and unanimous that the larger map was needed for classroom instruction. There was indication that some would prefer even larger scale maps comparable in size with the average wall maps. There was also criticism of the 1:7,500,000 scale map - doubt that it could ever show all county names legibly and also, that without county names it would be better for most purposes at page size.

A further mini-survey was made at the annual meeting of the National Council for Geographic Education in Washington, D.C. in 1974. Visitors to the Census Bureau exhibit were asked to register their opinions concerning the two differently scaled maps exhibited. Of 66 polled, 56 were for the 1:5,000,000 scale, 6 for the 1:7,500,000 scale, and 4 for both scales. Many of these had used the maps in teaching geography at the elementary, junior high, and senior high school levels.

As yet, the Geography Division has made no change in the publication scale of this map series. The samples were very small and certainly not unbiased, but more important, the two results were contradictory. In that case it is usually better to stay with what you have until more definite information is available. There could be another solution, as some have suggested, and that is publication of the maps also in 2 x 2 inch color slides, but that has not been investigated.

Size of map sheet and map scale for a particular map series are only two characteristics out of countless numbers that could be investigated in order to learn more about the effectiveness of these statistical maps. I believe the Bureau of the Census should develop controlled surveys of map users to gain information for statistical map improvement, although I know this approach will be difficult.

The Bureau of the Census has a great interest in getting more people to use its statistics and to use them in the most intelligent way. If good statistical maps and also graphs help people to understand statistics better, and, therefore, a greater demand for them is developed, more will be produced.

STATISTICAL MAP DESIGN PROBLEMS AND THE USERS

Better solutions to many design problems in statistical mapping could be achieved if we, the cartographers, knew more about how the users draw information from the

maps. Research cartographers have made progress in recent years in learning some of the answers, but the results are still fragmentary and the production cartographers have to rely on their best empirical judgment for a great many decisions made in preparing statistical maps. Therefore, I thought it would be useful to present some of the experience gained in four of the more critical planning and design phases of statistical mapping at the Census Bureau.

SUBJECT MATTER SELECTION

In selecting subjects for statistical mapping for a census one must examine the statistics that were collected to determine what data are suitable for mapping. Absolute figures should only be presented in distribution maps using dots or graduated circles, or the like. For choropleth maps we need density figures, averages, percentages, or ratios of one statistic to another. In both cases we need a geographic breakdown by States, counties, SMSA's, cities, or smaller areas. For most censuses we know the conventional subjects that have been mapped in the past and the more important of these distributions bear repeating with every new census, but trying to improve the format. We must seek out types of data which have never been mapped and also the data elements which are being collected for the first time. This is the kind of cartographic planning that should be done for each census. Usually it has been done at the Census Bureau by Geography Division.

Far more data could be presented in statistical map form than are. Some of these maps would be useful, but others would present patterns very similar to related subjects already mapped. For example, we could make dot and circle maps of male and female population separately, but their patterns probably could not be distinguished from each other and would be very similar to total population. On the other hand, a sex ratio map could be useful.

Some census subjects lend themselves to statistical mapping much better than others. Agriculture is one. It has many subjects such as different crops, livestock, equipment, and practices, and these are spread widely as every region has some form of agriculture. Population is also widely spread, but its depiction on small-scale statistical maps is hampered because so much of the population is urban and concentrated on a very small proportion of the land (1.5 percent in the United States in 1970). Economic activities also are concentrated in small areas, but the uses of their statistics are subject to a further difficulty, that of disclosure of confidential information. The census laws prevent release of information which might reveal the magnitude of activity of individual establishments or companies. The same law applies to population and agricultural data, but because the numbers in these fields are greater, the effect is minimal.

TYPES OF STATISTICAL MAPS

Users have their preferences and prejudices about statistical maps. Some express a dislike for maps with graduated circles, because they say it is difficult to

measure the value of any given circle. While there is some truth to this, it misses the purpose of a statistical distribution map which, in my mind, is to give an overall correct view of distribution in terms of geographic concentrations and patterns. If you want precise figures, you go to the statistical table, but a table won't tell you much about the distribution of small areas even though you know the areas and try to form a mental map of them. Normally, reading of the map gives the answer quickly, but in general rather than precise terms, because the data are expressed in number of points (dots) or in interval shadings (choropleths or isopleths). Statistical maps do not replace statistical tables. They are complimentary rather than competitive. Also, the comparison of statistical maps of the same subject made for different years enables one to understand where changes in distribution have occurred over time.

The selection of the type of statistical map to use for a particular distribution is not too difficult, because there are basically only four types of statistical maps: (1) maps with point symbols (usually dots and graduated circles) to show the distribution of absolute numbers; (2) choropleth maps (with data areas shaded) to show densities, averages, and rations; (3) isopleth maps (with contour lines) also to show densities, averages, and ratios; and (4) flow maps (with proportioned lines and arrows) to show movements from one area to another. Of course, there are countless variations in these maps and many combinations and these are the ways in which cartographic ingenuity can contribute much to making more informative maps. The Census Bureau's statistical mapping programs have used all of these types except the isopleth map. It would seem that the isopleth map has been neglected for at least four reasons: (1) its use for economic and social statistics is less familiar to the public; (2) its use would make the value identification of individual geographic areas such as States or counties, more difficult to obtain from the map; (3) its construction has been somewhat more difficult; and (4) in the mapping of census data the interest is not so much in the slope of the statistical surface which isopleth maps show so well, but rather in how the statistical areas are grouped in regions and their relative values.

Before leaving the subject of types of statistical maps let me illustrate what happens when the choropleth type of map is selected to show absolute numbers. On Map No. 47, "Number of Negro Persons: 1970" (Figure 9, p. 260) a large county like Los Angeles is much more prominent than the five counties of New York or Chicago area which are relatively small in area but have much larger Black population. Also, compare Map No. 49, (Figure 10, p. 261) "American Indians: 1970," with Map No. 14, "American Indians: 1960," (Figure 11, p. 262) which used graduated circles and see how the latter gives a more accurate picture not emphasizing some large counties with low densities of population as in Nevada and Oregon.

CLASS INTERVALS

A statistical map uses symbolization to show data distribution, because ordinarily it is impossible to provide details for each statistical area. Accordingly, the data must be generalized into classes. It is important that these classes or class intervals be well selected so as to reveal an accurate picture of the statistical surface of the map.

In developing class intervals for choropleth maps at the Census Bureau we have found that five classes are about ideal and rarely ever have used more than eight.

Five shades in black and white or color are easily distinguished, but there is increasing difficulty with more shades.

Different kinds of class intervals may be developed for different sets of data. The sizes of the intervals may be equal, or vary in some regular or irregular manner. The purpose is to divide the areas into only five or six classes, which will show the regional variations accurately. While recognizing that there are various ways to select intervals, our usual method involved developing a frequency distribution of the data or of a random sample of the data if we were dealing for instance, with over 3,000 counties. Then, we determined the mean average or the median, the highest and lowest values, and where there were breaks in the distribution with few areas or none. Within this range and with this information we set up tentative class intervals. We also considered it desirable for general public use and easier understanding that the interval breaks come at whole numbers or preferably numbers ending in 0 or 5. The mean and medians ordinarily should be in one of the middle classes and each class should have enough representatives so that it would show up clearly on the final map. If the distribution were skewed or otherwise abnormal, the class intervals reflected this. Also, if a map was repeated from one census to another, we considered it desirable for comparative purposes to maintain the same class intervals.

USE OF COLOR

We have had two problems with color on statistical maps. One is selecting the right colors to make the maps most effective for the users and that usually is the Geography Division's responsibility. The other is getting the printers always to use the colors as instructed, and sometimes this has been difficult for the separate U.S. map series (GE-50) where each printing order is separately contracted out by the Government Printing Office. Almost every order has been done by a different printing plant and they have been scattered across the United States. It is a tribute to the quality of the American color printing industry that specifications are followed as well as they are. Nevertheless, once in awhile, the work is not as good as one would like to see it.

What I have to say about the selection of color for statistical maps concerns our experience with the United States maps, GE-50 series, during the time that I was directing their development and ends not long before map numbers 40 though 45 were printed. I feel that I can freely criticize these maps and tell you what I like and don't like about them and what I believe we learned from making them.

Cartographers as well as others recognize that color and its use are a subjective matter and that people see colors differently. When we started to use color, we had no definite color plan. We knew we had to distinguish high values from low values and that the intervals next to each other on the graded scale should be enough different to be distinguished easily. We experimented and while we certainly never achieved the perfect color scheme, I believe we learned by trial and error some rules or principles which may be worthwhile to pass on for the considerations of others who face similar problems with color on choropleth maps. Some if not all of these conclusions have been reached by others, I will concede.

COLOR GRADATION RULES

1. Reject a full color spectrum or even a modified full spectrum, because the spectrum colors (red-orange-yellow-green-blue-violet) range in value from dark to light and back to dark and therefore are misleading for showing statistical value gradients from high to low.

With the first two choropleth maps in the GE-50 series we experimented with a color gradation that might be called a modified full spectrum. Map No. 6, "Families with Income Under $3,000 in 1959," has six percentage classes from high to low with the colors of red-brown-orange-yellow-green-blue. The color values go from dark to light to dark. The reader has to go to the legend to learn that red is high and blue is low and he might be bothered by the considerable jump in color value between yellow and green. Map No. 7, "Older Americans: 1960," with only five classes drops the red and brown and substitutes a pink that is rather weak for the highest value. After these two maps we gave up on using the full spectrum.

2. Ordinarily reject a single primary color in different values from dark to light, because it fails to show sufficient differentation between classes if there are five or more.

We found it works well with three or four, but five is doubtful. Map No. 39, "Percent of Population Urban: 1970," uses five shades of violet or purple, but it also has some pink. I think it is fairly effective but some people don't like purple.

3. Favor a shortened modified spectrum with two basic schemes ranging from dark to light values:

a. From violet through blue and green to yellow

b. From red through orange to yellow

These may be modified in countless ways by various mixing of adjacent colors and by varying the amounts of gray or white. They should always have value shadings from dark to light. See Map No. 8, "Per Capita Money Income for 1959," and Nos. 9 and 10 on education. Each has seven class intervals from purple through blue, green, and yellow to a cream. On Map Nos. 12 and 13 on employment, purple was dropped and a lighter color added in the green-yellow level. Map No. 16, "Negro Population as Percent of Total Population," uses seven color values from dark brown through yellow to cream. On Map Nos. 29, 31, and 33 on agriculture, color values from dark green through yellow to cream were used. Green and yellow are appropriate for suggesting the subject of agriculture.

4. For percent change maps which show gains and losses we have favored blue values for increases and red values for decreases.

We selected this scheme for use in 1960 for maps in the summary volume of the Census of Population. There was precedent for this in part. Red was being used for decrease dots on increase-decrease dot maps for showing crop and livestock changes for the censuses of agriculture. We were only allowed two colors other than black on the population maps so we chose blue as the other color. Once in Amsterdam when I was showing these maps, someone asked why we didn't use red for gain and blue for loss. They were probably thinking red is the dominant color and is used for high mountains and blue for the seas. We could have made that selection, but I suspect that whoever first used red for loss on maps was thinking of being in debt or losing or "in the red" as the American expression goes. In any event, this practice has become a convention on census maps and to change it now could introduce confusion. See Map No. 42, "Population Trends" which illustrates this color scheme rather well. Nevertheless, we did depart from this rigid convention for three agriculture maps of percent change where we used shades of green and yellow for increases and a light brown for all decreases. See Map Nos. 30, 32, and 34.

Selecting colors for choropleth maps is difficult and takes a great deal of time and patience. Also, what looks good in small patches looks quite different sometimes when spread all over the map. Someone has also asked why we haven't taken one color scheme and stayed with it. My answer to that is that I don't believe we have found a perfect color gradation scheme and I have some doubts that we ever will, because I believe people see the same colors differently. Also, I believe varied color schemes heighten people's interest in such maps as these. However, some of the recent maps in this series have failed to follow closely the rules of color gradation design which we used earlier, but I shall leave their effectiveness to your judgement. Note particularly some of the ethnic and income maps.

Before ending this discussion I wish to make two comments about possible improvements in color on dot and graduated circle maps as used in this same series of maps. The first is to suggest the use of even brighter colors of high intensity or chroma for dots and circles, because they occupy relatively little space on the maps and need to be emphasized. The second comment concerns the development of some means of making the land area with the dot and circle symbols stand out, perhaps by a contrasting tint on the land area or a stronger blue on the oceans. I make these comments because in studying the maps and particularly the slides while preparing this paper, I found that choropleth maps with greater expanses of color catch one's attention much better than some of the dot and circle maps.