

## AUTOMATING STATISTICAL GRAPHICS: A TOOL FOR COMMUNICATION

Vincent P. Barabba, Bureau of the Census

Today I'm going to discuss with you some efforts, both inside and outside the Census Bureau, to develop a fully automated and standardized graphic presentation system.

This is a very important endeavor. People who deal with statistics today are trying to cope with an information and data explosion that is increasing so fast that by the end of the next decade, as Cal Schmid pointed out, we may be contending with perhaps six times the present volume of statistical information. Given these realities it is imperative that we find immediately more uniform and more effective ways to communicate statistics quickly, accurately and more understandably.

What I have to say ties in directly with the presentations of my colleagues in their session. Standards and perception of graphics are inseparable from a useful, fully automated standardized system. We are all concerned with what we might call graphic literacy.

As we go through this presentation, I'm asking you to consider factors that will be useful to researchers in graphical statistics in developing a better sense of direction, not only regarding hardware and software but also in answering such questions as what statistics lend themselves to the system and how should they be presented. Although the Bureau has had little dialogue with other groups on this subject in the past, if the system is to develop effectively, we must have much dialogue in the future. We need your input because, in my judgment, the system we are developing will have a strong impact on the future use of statistics generated by the Federal statistical system.

There is at the present time little empirical evidence that graphics have a greater impact than textual or tabular material. We can say that intuitively graphics appear to be more effective; they certainly attract attention to a greater extent. Although users of data have been skeptical of the so-called "graphics method of statistics" for more than 100 years, I believe that it is time we begin, in a systematic way, to gather empirical evidence on the impact of graphics. We need to demonstrate the strengths of the various forms of graphic presentation in order to justify the investment in the system. Dr. Wainer's efforts represent a first approach to that need.

To be most useful, the systems must be standardized. Factors which may be considered as part of a standardized system include portability, adaptability to specific needs, flexibility to provide for simple as well as sophisticated usage at varying levels of complexity and

financial investment, and ability to be used by people with a variety of skills and training.

What I have just said about graphics, skepticism about their use, and difficulty in arriving at suitable standards is not new. For instance, 700 years ago the Mayan Indians had a highly developed written language based almost entirely on symbols of people and faces. An interesting treatise on the development of statistical graphics is given by James R. Beniger in his unpublished paper entitled "Science's Unwritten History: The Development of Quantitative and Statistical Graphics." As Cal Schmid pointed out, the International Statistical Institute which was organized in 1885 gave serious consideration to standards of graphic presentation but with little success. That same year, the noted economist, Professor Alfred Marshall, wrote the following in the Jubilee Volume of the Statistical Society of London:

"The graphic method of statistics, though inferior to the numerical in accuracy of representation, has the advantage of enabling the eye to take in at once a long series of facts. . . . Its defects are such that many statisticians seldom use it except for the purpose of popular exposition, and for this purpose I must confess it has great dangers. I would however venture to suggest the inquiry whether the method has had a fair chance. It seems to me that so long as it is used in a desultory and unsystematic manner its faults produce their full effect, but its virtues do not."

Although many of the concerns expressed by Professor Marshall almost 100 years ago are still being echoed today, it appears that the climate in the statistical community is such that serious efforts to present data more effectively through the medium of graphics must be considered.

For many reasons, the Bureau of the Census is interested in promoting this activity. First, it is a major contributor to (some might say a culprit in) the information explosion. Second, we have a historical as well as current statistical base on which to build. Third, our users have indicated that our methods of display are cumbersome, require time to digest and have the potential for inaccurate interpretation. Fourth, the Bureau has much of the basic software and hardware that is required to develop an automated standardized system and last but far from least, we have innovative, well trained personnel to develop the procedures and demonstrate their usefulness.

Our ultimate goal at the Bureau, which may

take a few years to realize, is to put together a system that will produce statistical displays in black and white or in color, in a variety of formats. The displays include those produced electronically on a television screen and on videotape, in the form of color printouts, color slides, and microfilm, or whatever format demonstrates a potential for effective communication quickly and at reasonable cost.

We have, in fact, already made a good start in the development of this automated system for presenting statistics. One product that has resulted from our efforts thus far is a new chartbook of social and economic trends. The issue for August contains 86 pages, including both text and more than 165 charts and 4 maps -- all created by computer at the Census Bureau.

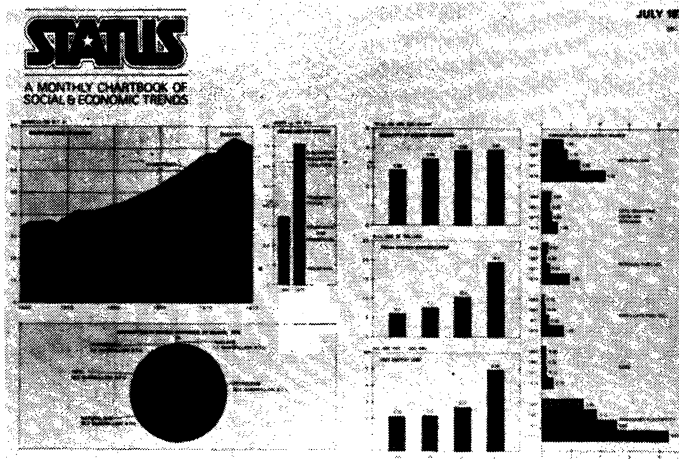
This publication grew from a series of briefing notes on domestic developments prepared by the Bureau from data compiled from the entire Federal statistical system for the President and Vice President, starting in April of last year. The graphic approach that the Bureau took impressed them to the extent that they decided that a more extensive publication, stressing the use of 4-color graphics, would be of great value not only to the Government, but also to the public.

The result is our new magazine, called STATUS, which stands for Statistics, United States. On the first of July, STATUS made its formal debut in ceremonies at the Bureau commemorating the 25th birthday of UNIVAC I. Vice President Rockefeller attended, and in his remarks he pointed out that in 25 years we have progressed from the ability to simply produce more information via computer, to an important new dimension-- the actual dissemination of information in graphic form generated by the current family of computers.

Here are some line charts from STATUS. They are actually in four colors but are reproduced black and white here. It took the computerized system less than five minutes to produce each of these high-quality charts, many times less than it would take an artist to do the work. And,

of course, the more complicated the chart, the more time an artist would have to spend on them, whereas the system do can them all relatively quickly.

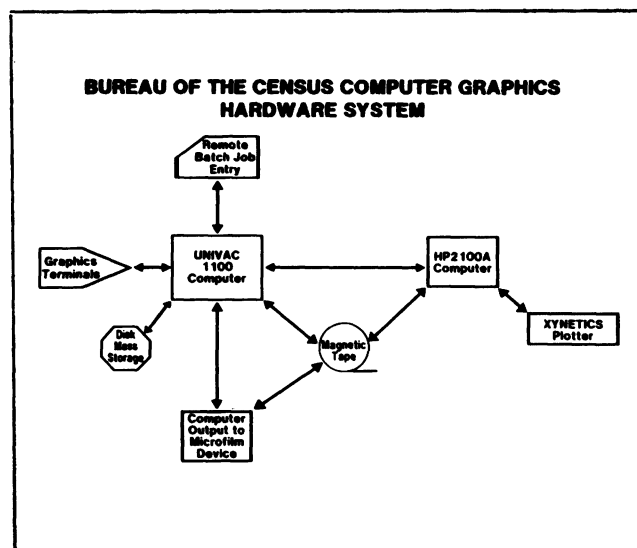
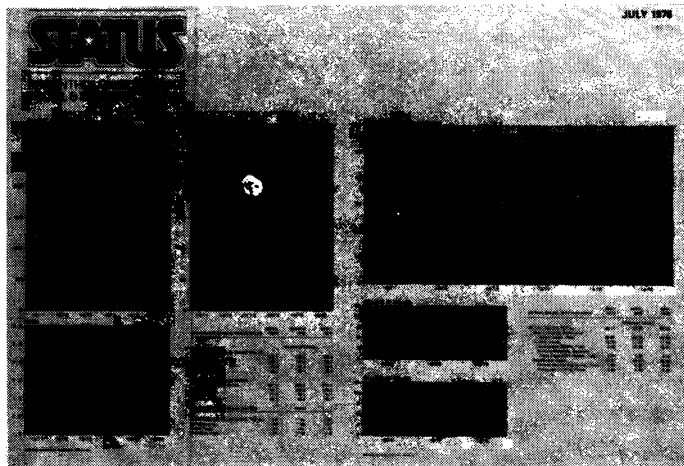
Here are some more graphics from STATUS, illustrating the system's ability to produce bar charts and pie charts as well as line charts. I might add that the comments that we already have



received on STATUS -- both in and out of Government -- have expressed a great deal of amazement at the speed at which we have been able to produce the statistics in graphic form. Data less than two weeks old, in some cases, appear graphically in published form.

Let's take a look at the computer hardware system that we use to produce these graphics.

In this schematic, the boxes in the middle represent our central processors. On the left is the graphics terminal, which can feed graphical descriptions into the computer through a keyboard and display the resulting charts on a screen. The other two main pieces of hardware receive the instructions from the computer and then actually



produce the hard copy. They are the Xynetics flatbed plotter, shown on the right in the schematic, and the computer output to microfilm equipment, shown at the bottom of the schematic, which we call the COM system. The arrows simply show the data communications lines. These normally are telephone lines from a remote graphics terminal location to the computers, and from the computers to the hardware that makes the graphics.

These three main units -- the graphics terminal, the plotter and the COM system -- cost the Census Bureau about \$570,000, but less sophisticated counterparts would be much less expensive. For about \$50,000 you can obtain basic graphics.

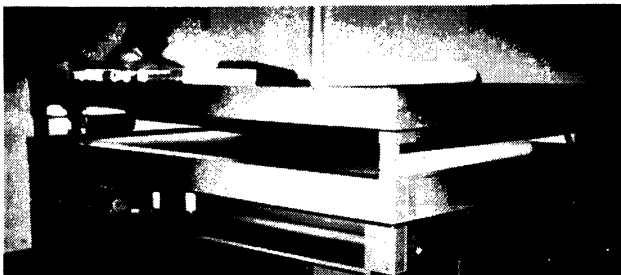
The graphics terminal consists of a cathode ray tube, or CRT as we call it, which is like a television tube, and the keyboard for issuing instructions. The operator is able to see on the screen of the CRT what is actually being cor



structed from information fed to the computer. If you don't like what you see, you can issue further instructions to the computer and change any one element in the picture.

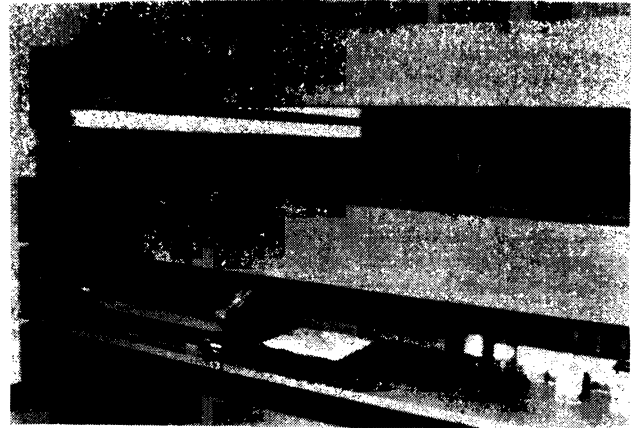
When you are satisfied, you just instruct the computer to have either of the hard copy devices produce the chart. You also can get a paper copy or copies in seconds through a copying machine that is hooked up to the CRT. This CRT runs about \$10,000, and this is down from \$12,000 five years ago for a terminal that had much less capability.

This is the Xynetics flatbed plotter that draws the charts from instructions received from the



computer. We acquired this 2 years ago for about \$110,000. Today they run from \$20,000 for a 1-pen plotter to as high as \$200,000 for the most for the most sophisticated version.

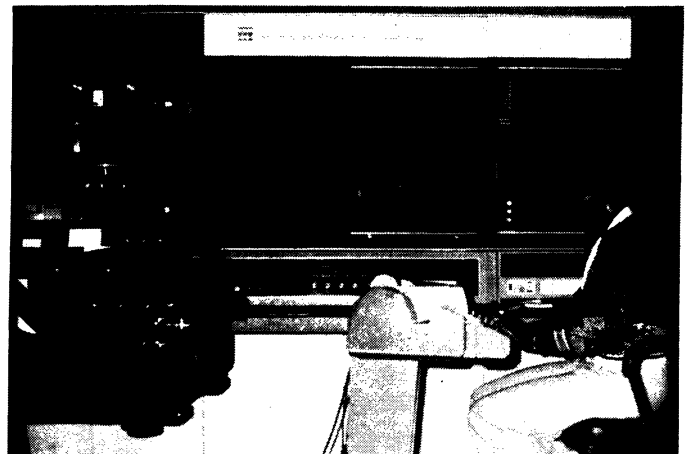
You can see here the drawing head that can produce graphics. It is program controlled and and is capable of plotting charts in different sizes,



in black, or in black plus three additional colors made bypens that can be soft tipped, ball point, or liquid ink, and that can draw on mylar, plain plain bond paper, or tracing paper.

The weekly White House briefings, which incidentally we still prepare, use charts produced by this plotter. For STATUS, however, we still have to go through color selection and separation, have the printing plates made, and then do the printing. A little later I will describe equipment that will eliminate this and directly produce full-color hard copy -- and our ability to use this new capability may not be more than a year or so away.

This is the COM system, once again the letters C-O-M standing for computer output to microfilm. On the right is the tape drive, which has a magnetic tape on it that has the instructions from the CRT via the computer. In the center is a screen that allows you to see the image of the graphic. If you like it,

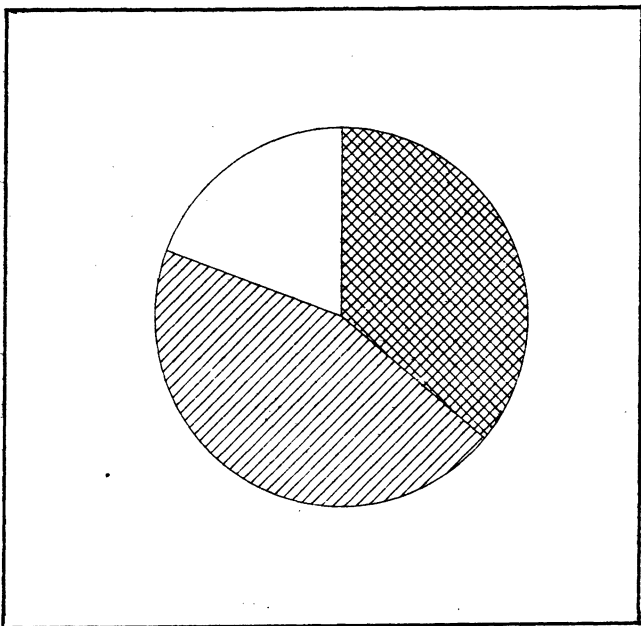


you can project the image on the screen at the left, where it is photographed by any of four cameras -- a 16 millimeter, a 35, a 105 millimeter microfiche camera, or a large 310 millimeter camera that will directly produce 8 1/2 by 11 inch pages on either photo typesetting paper, or on film.

It's important to note that the printshop still has to add colors, if they are needed, but the COM system eliminates the outside photographic process. This particular system cost \$450,000, but a basic COM machine could be purchased for \$150,000.

So these are the three types of graphic devices that we have now in our system. The big breakthrough is that we have added to the basic ability of the computer and have entered the custom stage where we can ask the computer to give us clear and accurate graphics made to order. Here are some examples.

For instance, here is a simple pie chart, created by the programmed computer. It has three slices with three different shadings.

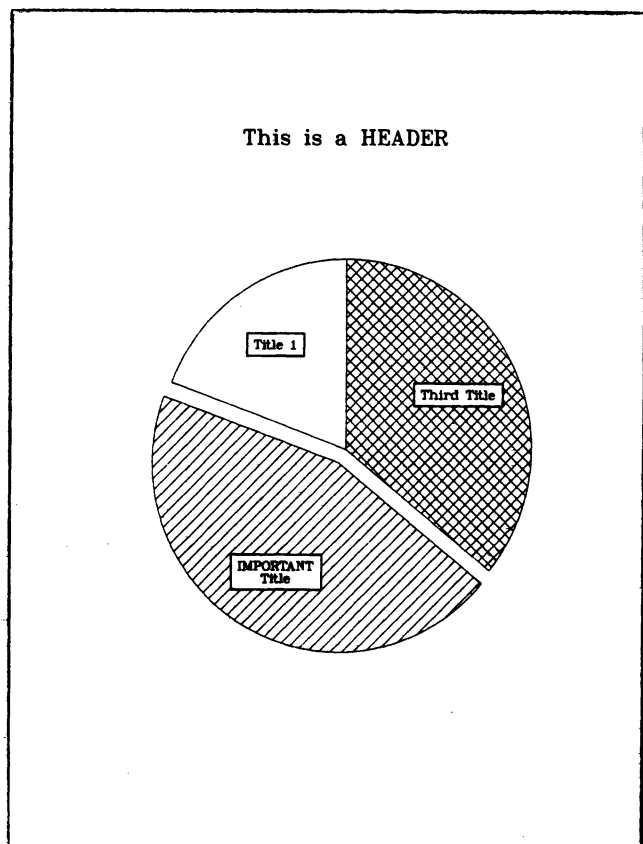


It was created through these simple, English-language instructions. The first was to lay out three slices. The word "end" means that there are no more graphic instructions. And then we gave the data values -- in this

```
# LAYOUT SLICES 3
# END
14 32.3 26
```

case 14, 32.3 and 26 -- and we generate our pie chart.

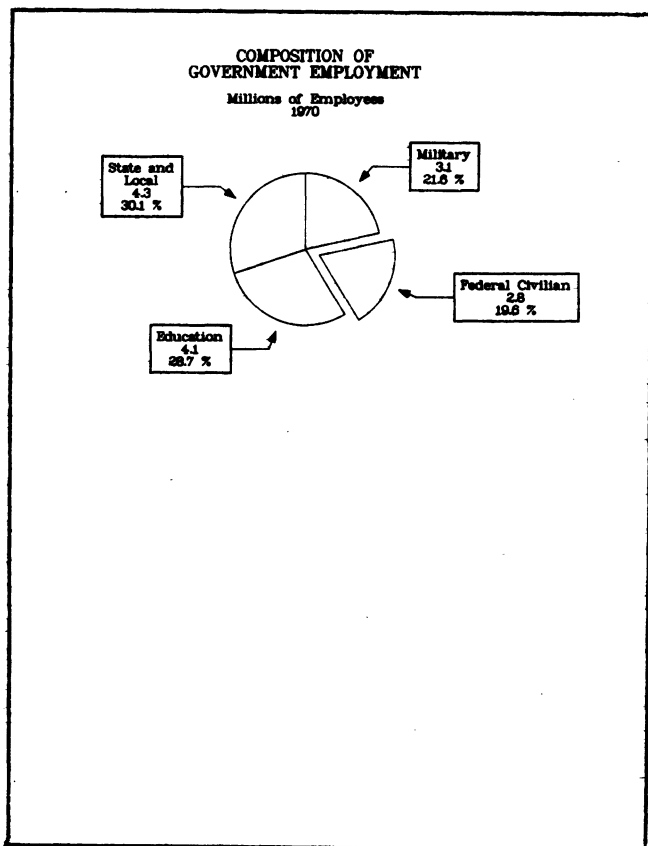
But suppose you want more sophistication. You want a separated slice; a title, or header, at the top; and titles on the slices, including one that is more important than the other two as we have here.



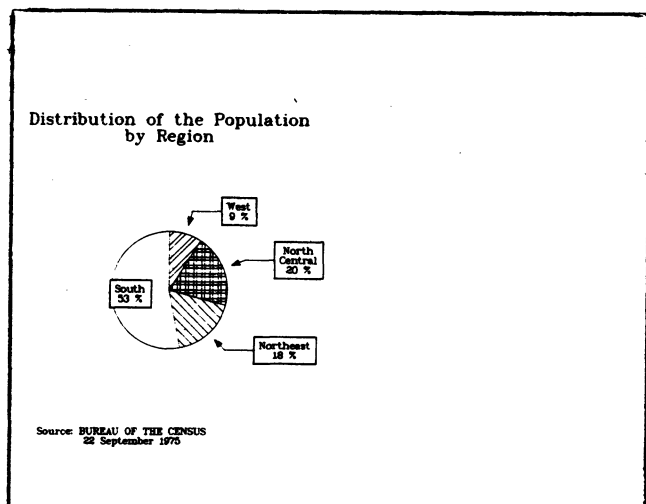
Using the same data values, these are the additional instructions that were needed to produce the chart. The computer took care of all the housekeeping. It determined the size of the pie, the size of the page, the textures, and where to put the titles -- all through previous programming.

```
# LAYOUT SLICES 3
# BROKEN 2
# HEADING/ <T> HIS IS A <HEADER >/
# SLICE TITLES/ <T> ITLE 1/
#                   <IMPORTANT> & <T> ITLE/
#                   <T> HIRD <T> ITLE/
# END
14 32.3 26
```

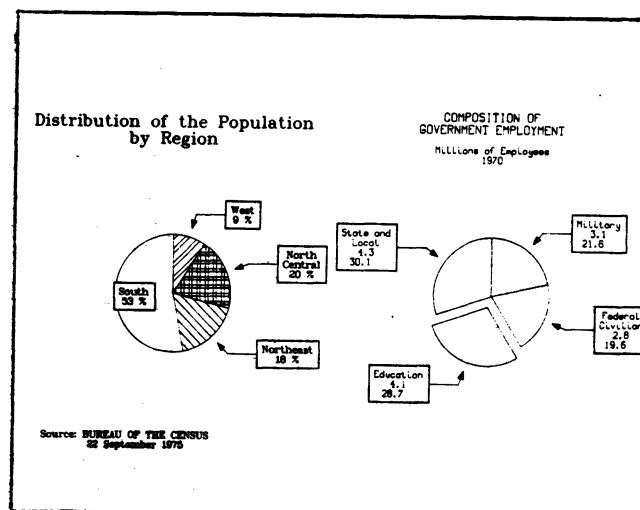
You can have the computer put the chart at the top of the page, and in this case put titles in boxes and indicate the slices with arrows.



Or you can put the chart on one side of a horizontal page. You could create text material by the computer, and strip it into an appropriate place as part of the graphic -- on the right side in this particular situation.

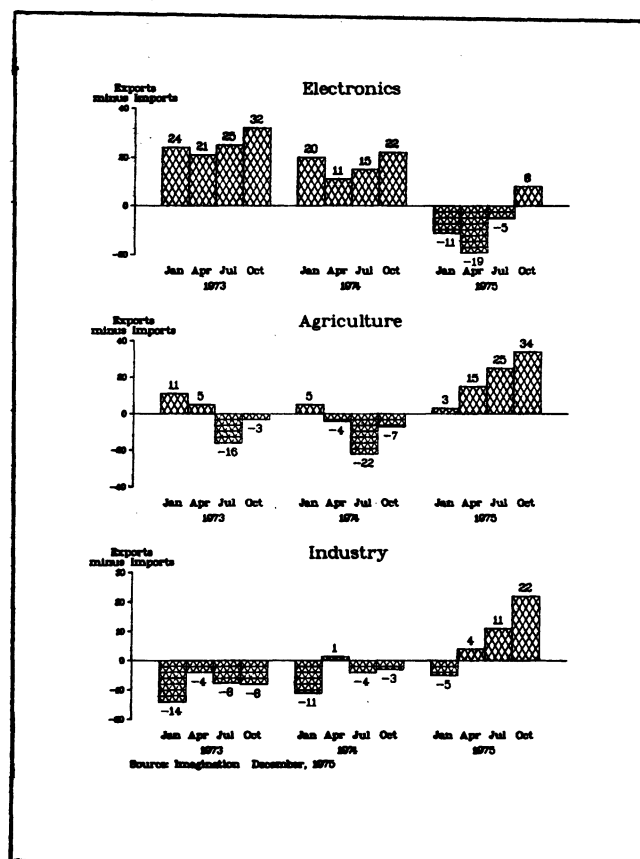


You can put charts side by side, if you want, and the computer will place all these things for you. And again, you can change what you don't like. For instance, you could have asked the com-

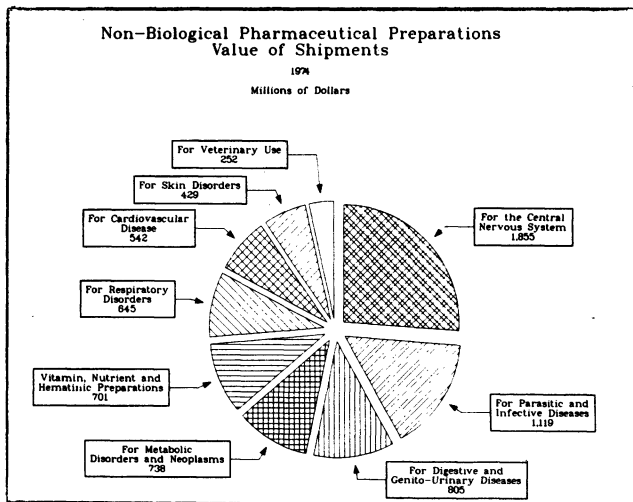


puter to move that left pie chart a little more to the left, if you wanted to, after looking at it on the CRT. Also note here that we can put the titles in a variety of type styles.

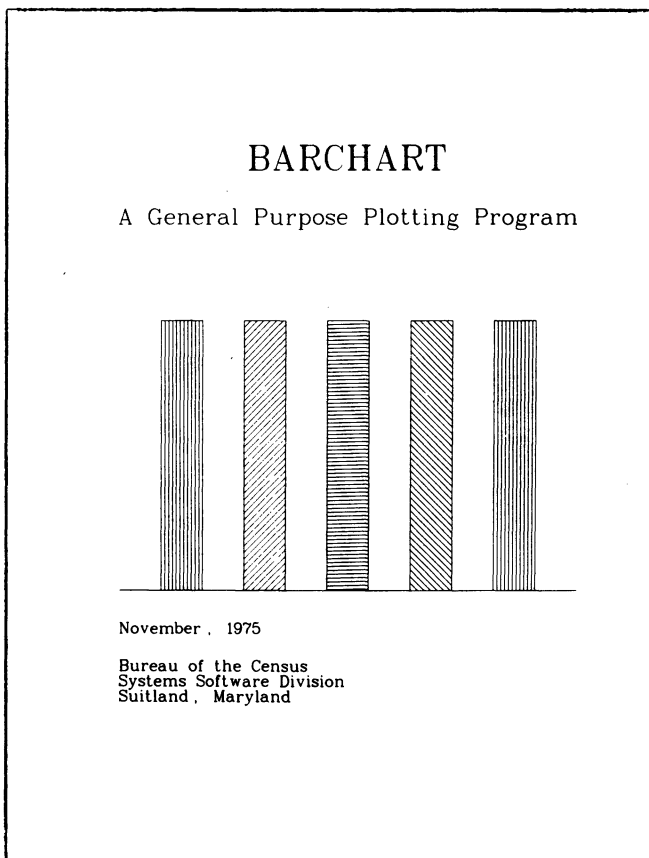
And here is a case where you have three bar charts on the same page, with scales on the left side.



Here is a chart (see next page) that was produced for a Census Bureau publication, which contains quite a bit of graphic information. You can imagine the amount of time that it would take an artist to do this particular chart, considering all the titles, boxes, arrows and hash marks.

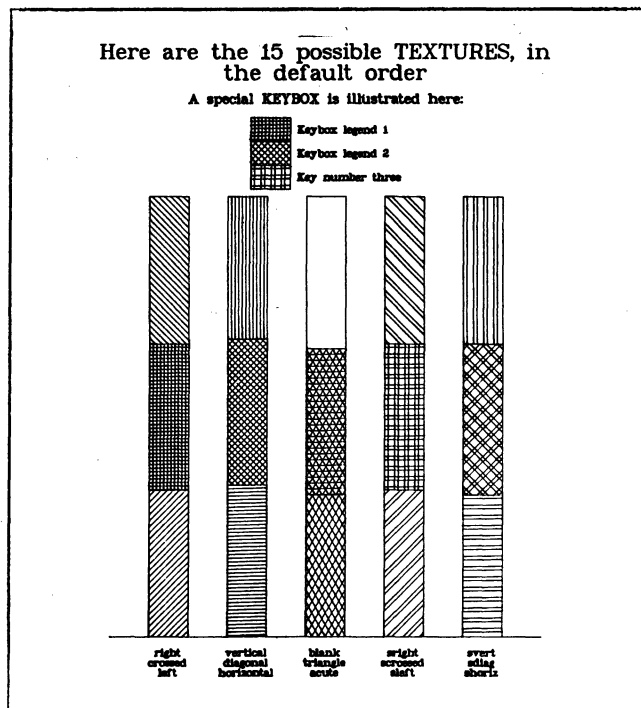


And you can get a variety of hash markings in bar charts.

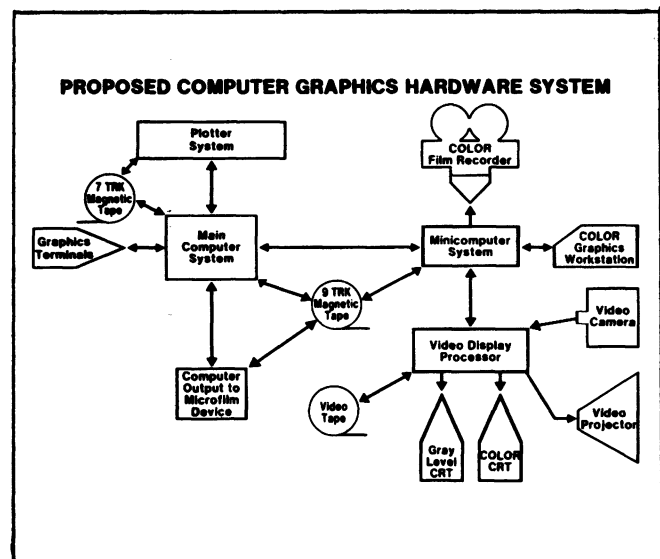


This illustration (see top of next column) shows 15 possible textures that have been programmed into our computer. Actually, an infinite variety of textures could be selected and programmed.

So this is the type of thing we can do today with the system, but we can only do it in black and white, except for the limited set of four colorpens in the plotter. So what we are aiming for is to develop the ability to do these same things in full color.



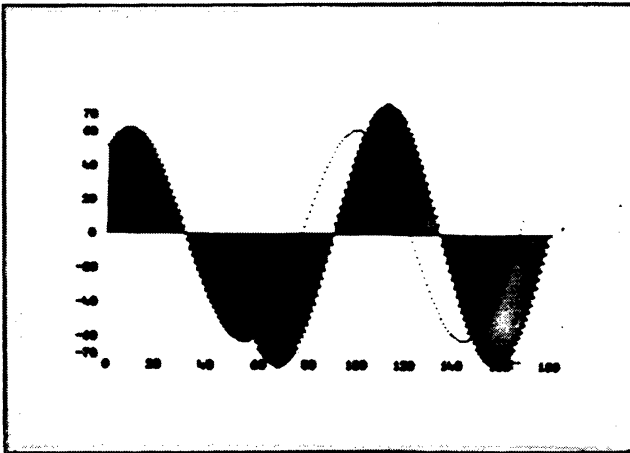
This is a schematic of our present system plus the equipment that would be needed to provide color capability. The present system is on the left, with the graphics terminal, the plotter system and the COM device. On the right is what we envision from the color standpoint. And let me point out that these color components already have been developed by companies outside the Census Bureau.



The photograph at the top of the next page shows the color CRT, although this illustration is of necessity in black and white. In addition to being able to change the composition of the graphics at will, this CRT will allow you to make color choices and quality judgments on the spot, until you are satisfied. It has, literally, a dial-a-color capability. These units would cost from \$20,000 to \$70,000.



Here is a chart that illustrates the display capability of a color CRT, although again it is in black and white here. This was actually produced by a color CRT developed by one of the companies. In fact this is a photograph of the CRT face. Again, the computer has done all of the scaling and the rest of the design.



There are several devices being developed at the present time that will be able to produce hard copies in color very quickly, once you are happy with the colors on the CRT. One of the devices would be interfaced with a color copying machine. Also, of course, you could

store your approved color graphic and call it up for reproduction at any time.

Another valuable addition to a full-color system will be a video projector that will allow you to project a graphic image onto a movie-type of screen, electronically. Video equipment can cost anywhere from \$5,000 to \$90,000.

Equipment also has been developed that will permit you to put images created on the color CRT screen onto color film. Equipment of this type would be in the \$50,000 to \$125,000 range.

One point to remember is that the graphics that will be available in color also will continue to be available in black and white. And they also will be available in tones of gray, which is not the case at present with our system.

One of the things that private industry is working on now is the development of information in a computerized data base that can tie into the rest of an automated graphics presentation system. This would permit the decision-maker to have direct interaction with the data file through the CRT or some similar approach. And this would be either in color or in black and white or gray, and a hard copy of what he wants would be produced by issuing a simple instruction.

So these are some of the things that we have today, either at the Census Bureau or in the research laboratories of private industry. As is illustrated by STATUS magazine, we already have the automated tools to produce effective and accurate graphics, and we hope within a year to have acquired the equipment to create color as well as black and white, both electronically and in hard copies of different kinds. Then we will have to put the entire system together, which should take a little more time.

Ultimately, we would like to see this system, or one of a modified nature, made available to users in any location, at reasonable cost and in a standardized form that will permit a maximum amount of flexibility. In the meantime, we will encourage and participate in research to determine the extent of the impact that graphics have on people, and if necessary take account of this impact as we continue to develop the system.