A CHALLENGE TO CARTOGRAPHERS

Vincent P. Barabba
U.S. Bureau of the Census

INTRODUCTION

I am especially pleased to speak here today -- not only because it gives me a chance to meet once again with some of the leaders of the international statistical and cartographic community, but because the subject is of great interest to me.

All of us here today are concerned with finding meaningful methods by which to communicate information. In many ways we are working against time. It has been predicted that by the end of the next decade new information will be generated and circulated at six times the present rate and 20 to 25 times the volume of a mere 15 years ago.

Even if such a forecast is exaggerated, there is little doubt the amount of data is increasing rapidly, and that new and effective ways of communicating the information these data contain must be developed.

It is my strong belief that the answer lies in developing graphic methods which complement existing data formats by summarizing data accurately and efficiently. Because of the wide variety of users, the methods of display must be standardized in a manner which allows decision-makers in various fields to easily understand what they are viewing. And finally, the methods should be as fully automated as existing technology permits.

What we are talking about is a fully automated and standardized graphic presentation system.

I think we are at a point where the scope of the problem has been adequately defined -- and now we can focus on the challenge it presents.

Unfortunately, one of the major challenges is the machinery of adoption. Developing such a system is not as difficult as getting it known and into use.

EDITOR'S NOTE: Vincent P. Barabba has been Director of the Bureau of the Census since 1973. In this capacity he has encouraged Bureau responsiveness to the needs of data users. He has taught at several universities and has made presentations at numerous national and international conferences, as well as professional society meetings. Prior to coming to the Census Bureau, he served with several data consulting firms, where he pioneered the advanced technical application of computers to analyze population data in marketing and other types of surveys. Mr. Barabba served as honorary co-chairman of Auto-Carto II.

To underline the situation, let me offer the following quotation:

> 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.

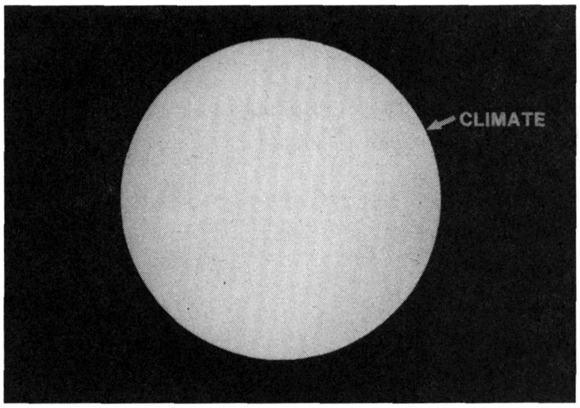
Does anyone care to guess when that was written? The style of the language may give you a clue. It comes from a paper entitled "On the Graphic Method of Statistics," written by the famous economist Professor Alfred Marshall, and appeared in the Jubilee Volume of the Statistical Society of London -- in 1885.

In the 91 years since that was written, graphic presentation still has not had its "fair chance." However, there is activity on a number of fronts. Let's look at one example of a way to enable "the eye to take in at once a long series of facts."

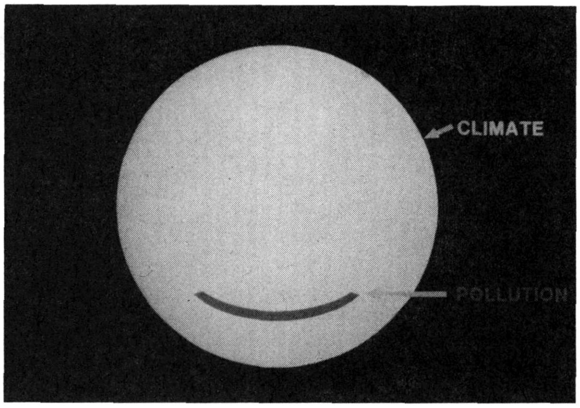
In the June 1973 <u>Journal of the American Statistical Association</u>, Herman Chernoff presented the concept of using the components of cartoon faces to map multivariate data. His purpose was to find a helpful tool to communicate information to the analyst in a form that was easier to use than the many complex tables found in the traditional computer output. A year later, McDonald and Ayers applied the Chernoff faces in the analysis of a community mortality and pollution study. In this study sixty faces represent portraits of sixty different standard metropolitan statistical areas. They are drawn so that each face represents the relationship in a metropolitan area of sixteen different variables organized into four classes of data.

CLASS	VARIABLE(S)	FEATURES CONTROLLED
CLIMATE	PRECIPITATION JANUARY TEMPERATURE JULY TEMPERATURE RELATIVE HUMIDITY	CIRCUMFERENTIAL SHAPE OF FACE FACE
POLLUTION	HC POTENTIAL NO_x POTENTIAL SO_2 POTENTIAL	THE POSITION AND SHAPE OF MOUTH
MORTALITY	TOTAL MORTALITY RATE	NOSE LENGTH
SOCIOECONOMIC	% ≥ 65 YEARS POP./HOUSEHOLD EDUCATION % SOUND HOUSING POP./MILE^2 % NON-WHITE % WHITE COLLAR % WITH INCOME <$3000	THE POSITION AND SHAPE OF THE EYES, PUPILS AND BROWS

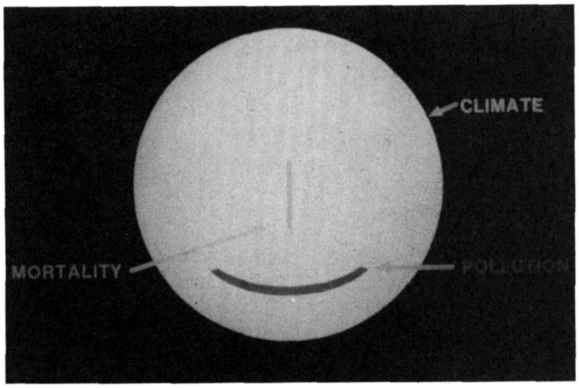
The authors have attempted to present their data graphically in the following manner:



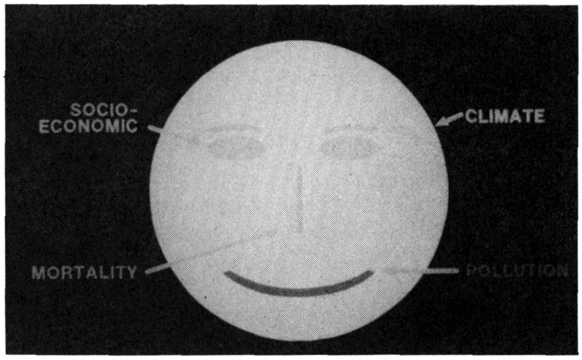
The four climate variables control the circumferential shape of the face.



The three pollution variables control the position and shape of the mouth.



The one mortality statistic controls the nose length.

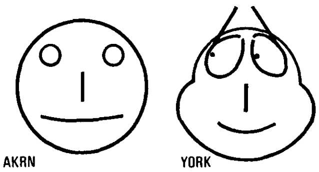


And finally, the eight socioeconomic variables control the position and shape of the eyes, pupils and brows.

This is a face representing the Akron, Ohio, SMSA. The 16 variables appear across the top. Of the 60 metropolitan areas in the study, Akron is a middle-of-the-road example. The facial features appear neutral, undistorted.

SMSA	PRECIPITATION	JAN. TEMP.	JULY TEMP.	% ≥ 65 YRS.	POP./HOUSEHOLD	EDUCATION	% SOUND HOUSING	POP./MILE²	% NON-WHITE	% WHITE COLLAR	% < $3000	HC POTENTIAL	NO_x POTENTIAL	SO_2 POTENTIAL	REL. HUMIDITY	TOTAL MORTALITY RATE
AKRON,OHIO	36	27	71	8.1	3.34	11.4	81.5	3,243	8.8	42.6	11.7	21	15	59	59	921.870

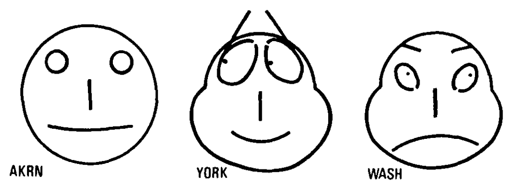
SMSA	PRECIPITATION	JAN TEMP	JULY TEMP	% ≥ 65 YRS	POP / HOUSEHOLD	EDUCATION	% SOUND HOUSING	POP / MILE²	% NON-WHITE	% WHITE COLLAR	% < $ 3000	HC POTENTIAL	NOx POTENTIAL	SO₂ POTENTIAL	REL HUMIDITY	TOTAL MORTALITY RATE
AKRON,OHIO	36	27	71	8.1	3.34	11.4	81.5	3,243	8.8	42.6	11.7	21	15	59	59	921.870
YORK,PA.	42	33	76	9.7	3.22	9.0	76.2	9,699	4.8	42.2	14.5	8	8	49	54	911.817



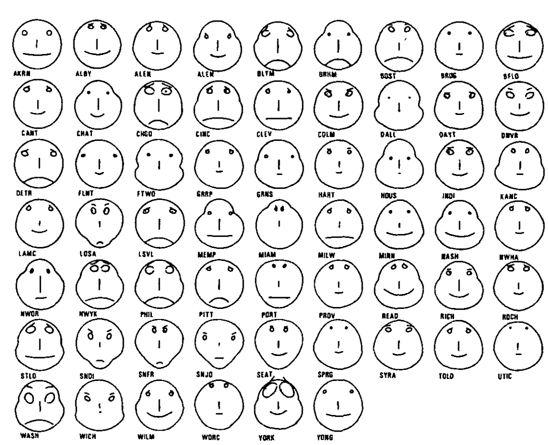
Next is York, Pennsylvania. A considerable difference is evident--especially in the larger population density, as reflected in the size of the eyes, the lower pollution potential, reflected in the shape of the mouth, and the climate, which affects the general shape of the face.

Finally, Washington, D.C. This is not an editorial comment, by the way. Immediately apparent is the high potential for all forms of pollution, and a fairly high population density. You can see also that the climate is not too different in York and Washington.

SMSA	PRECIPITATION	JAN TEMP	JULY TEMP	% ≥ 65 YRS	POP / HOUSEHOLD	EDUCATION	% SOUND HOUSING	POP / MILE²	% NON-WHITE	% WHITE COLLAR	% < $ 3000	HC POTENTIAL	NOx POTENTIAL	SO₂ POTENTIAL	REL HUMIDITY	TOTAL MORTALITY RATE
AKRON,OHIO	36	27	71	81	334	114	815	3243	8.8	426	117	21	15	59	59	921870
YORK,PA	42	33	76	97	322	90	76.2	9,699	48	422	14.5	8	8	49	54	911817
WASHINGTON, DC.	41	37	78	62	325	123	895	5308	259	597	103	65	28	102	52	967803



With a little training, an analyst can glance at the sixty faces and pick out metropolitan areas with characteristics and relationships in which he is interested. Instead of facing a page of 960 data items, he sees this presentation of faces -- a visual catalog of 16 variables for 60 SMSA's.



But just how new is the concept of presenting information through the medium of facial components?

The ancient Mayan Indians had a highly developed written language based almost entirely on glyphs of either whole persons or faces. These symbols were also intended to convey given facts -- information independent of other glyphs.

The Mayan symbols were not an alphabet in the sense we use the word. Nor were they pictography as in the Egyptian hieroglyphics. What the Mayan written language has in common with Chernoff's faces is that both employ abstract representation of known forms to transmit information expressed by the size, length or curvature of the lines making up the given symbol.

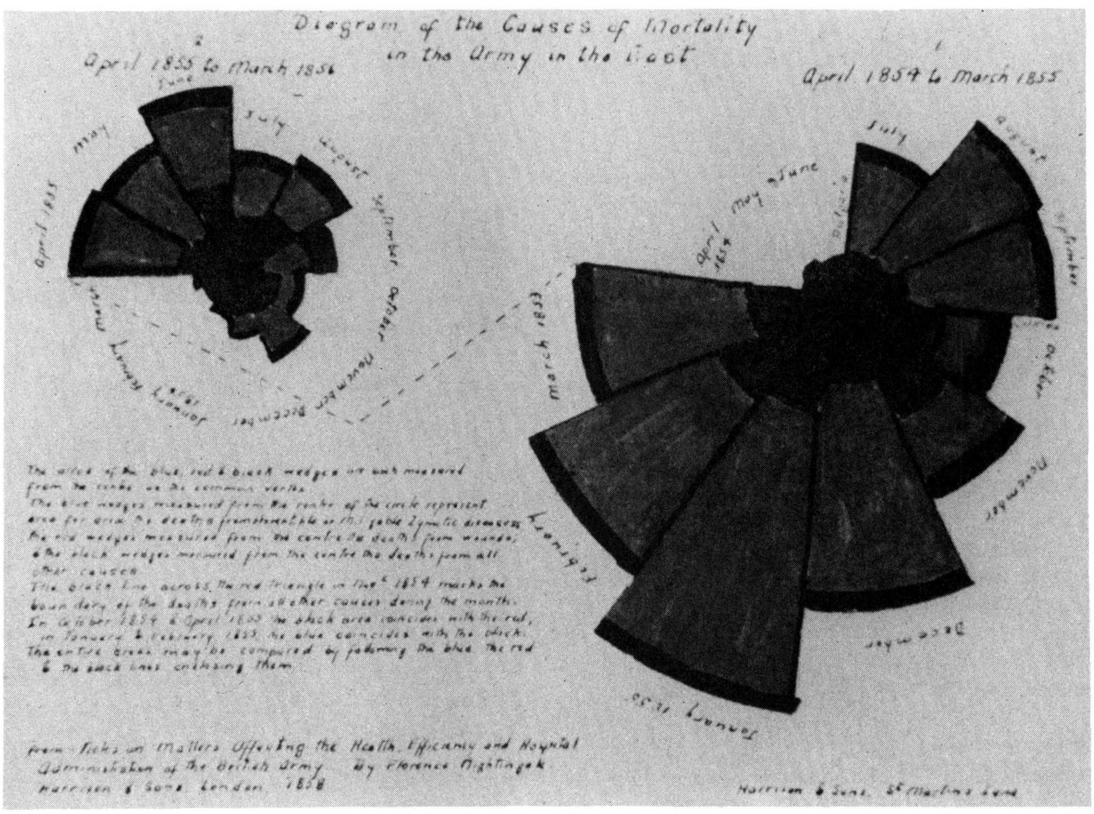
Although Chernoff's general concept may appear new to many of us, there certainly is an historical parallel in the Mayan faces of some 700 years ago.

In a recent Census Bureau seminar, Dr. Roberto Bachi of Hebrew University presented an alternative graphic alphabet. He stressed the need to understand such a new alphabet <u>before</u> attempting to evaluate the graphic system's effectiveness.

Let's look at another example.

In a recent discussion the Bureau had with another U.S. Government agency, the development of polar aerial graphs -- such as the one shown here -- was discussed.

We found that many people considered this form of graphic presentation to date from the late 1960's. However, this graph represents casualties during the Crimean War as illustrated in the 1850's by Florence Nightingale, the noted British nurse.



Clearly the ability and the desire to present information in a graphic way is not new. The phrase "one picture is worth a thousand words" has been with us a long time. Using poetic license, I would say "one statistical map is worth a thousand pictures." In this case the map on page 251 presents two variables and imparts a large amount of information. Figures 2 and 3 (pp. 252, 253) show the individual maps that went into preparing the cross-map.

STANDARDS

Another component of our goal is standardization.

Again, we have yet to break new ground. Not only is the concept of graphical standards not new, but difficulty in establishing them has been with us for a long time as well.

In 1872, over 100 years ago, the Eighth Congress of the ISI had a lively debate on the subject culminating in the following declaration: "As for uniformity of diagrams, properly called, the Congress declares that the time has not yet come to propose uniform rules." In fact, the first tangible efforts related to standards did not come about in the United States until 1936, in a report prepared by the Committee on Standards of Graphic Presentation, entitled Suggested Symbols for Plans, Maps and Charts.

One of the key reasons for having standards was pointed up in a recent article in The Cartographic Journal by Ronald Carswell and Haze Wescott of the University of Calgary. They noted that illustrative material is not automatically more informative than straight text--that a person must be taught how to use pictures and graphic illustrations in order to extract the information they contain.

My point here is that such teaching is not being accomplished because there are no standards on which to base teaching.

Before leaving the subject of standards, I want to make sure I am not creating a wrong impression. The standards I am calling for are not a set of rigid rules and regulations, but guidelines--perhaps conventions would be a better word.

AUTOMATION

The final component of our goal is fully automated.

Once more we find the idea is not new. However, in this regard most development has occurred within the lifetime of the people in this room. In fact, it was not until the late 1960's that computer-driven pen plotters came to be widely employed. The interactive graphics systems such as Ivan Sutherland's "Sketchpad" and the "Magic System" of the U.S. National Bureau of Standards are as recent as 1963 and 1964.

What then about the development of a fully automated and standardized graphical presentation system. Is it in fact an innovation? The answer is both yes and no.

Here I would like to refer to the work of Everett Rogers, a Professor of Communications at Stanford University in California. Rogers defines an innovation this way:

> An idea, practice, or object perceived as new by an individual. It matters little, so far as human behavior is concerned, whether or not an idea is "objectively" new as measured by the lapse of time since its first use or discovery. It is the perceived or subjective newness of the idea for the individual that determines his reaction to it. If the idea seems new to the individual, it is an innovation.

The important point here is that an innovation need not be truly new, in the sense that it has never existed before, in order for it to be perceived as new.

Keeping Rogers' definition in mind, let's ask the question once again--what is innovative about a fully automated and standardized graphic presentation system? The answer is the attempt to put all three elements together. Graphic Presentation--Standards--Automation.

For discussion purposes, let us accept the concept of such a system as not only an innovation but one whose time has come. Where then do we start? We must convince the decisionmakers of our society that not only do they need this system, but, that in the initial stages of development they are going to have to divert scarce resources from other projects so that we can produce the system in a form that will effectively and efficiently meet their requirements. How can we systematically go about developing and presenting an innovation that will be adopted and used by those responsible for making decisions based on information?

I find a sense of direction in the work of Rogers as he outlines the complexity of the innovation-decision process. In this model Rogers depicts four stages.

First is the KNOWLEDGE stage where the individual is exposed to the innovation's existence and gains some understanding of how it functions.

I think most people accept the statement that we generally tend to come into contact with ideas or concepts which are in general accord with our interests, needs or existing attitudes. Rogers identifies the central point of the knowledge stage as follows:

> Consideration of a new idea does not pass beyond the knowledge function if the individual does not define the information as relevant to him or if he does not seek sufficient knowledge to become adequately informed so that persuasion can take place.

Which leads us to the second stage: PERSUASION.

During this stage the individual forms a favorable or unfavorable opinion about the concept. He becomes more psychologically involved as his knowledge of the innovation increases. Rogers stresses the key point that both the knowledge and persuasion stages move only as rapidly and effectively as the channels of communication allow. If an innovation has possible meaning to many individuals the only way this may be realized is by bringing the potential user into contact with it.

The third step is the DECISION stage, and here the individual engages in activities which lead to a choice to adopt or reject the innovation.

Although the two earlier stages implied that choices were being called for, in the decision phase the type of choice is different, because it reflects a commitment to adopt or reject a new idea. An important finding Rogers has made in studying this area is that innovations which are amenable for a trial or test are generally adopted more rapidly.

Following a decision to adopt an innovation, even on a limited basis, comes the CONFIRMATION stage.

Here the individual seeks reinforcement for the decision he has made--although the possibility of a reversal remains. There seems little doubt that meaningful standardization would assist in this process.

What is it about an innovation that determines its rate of adoption? What kinds of innovations are fully adopted in months or years as distinguished from our subject today, portions of which were being discussed over 100 years ago and are not yet fully adopted? What can we do to "package" the innovation to hasten its adoption?

Rogers points to five attributes of an innovation which tend to determine its adoption rate. It is important to remember that it is the perception of the decisionmakers we are dealing with at this point, not the attributes of the innovation as seen by those who are seeking to have it adopted. The five attributes are relative advantage, compatibility, complexity, trialability, and observability.

Relative advantage is the degree to which an innovation is perceived as being better than the idea it supersedes. The degree of relative advantage is often expressed in economic profitability, but may be measured in other ways as well.

Compatibility is the degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of the receivers. An idea that is not compatible with the salient characteristics of a social system will not be adopted as rapidly as an idea that is compatible. Compatibility ensures greater security and less risk to the receiver and makes the new idea more meaningful to him.

Complexity is the degree to which an innovation is perceived as relatively difficult to understand and use. Any new idea may be classified on the complexity-simplicity continuum. Some innovations are clear in their meaning to potential adopters, others are not.

Trialability is the degree to which an innovation may be experimented with on a limited basis. New ideas that can be tried on the installment plan will generally be adopted more rapidly than innovations that are not divisible. An innovation that may be used on a trial basis is less risky for the adopter.

The impact of a fully automated system should be noted here, since it would allow many more trials of new ideas before adoption, and at greatly reduced cost.

Observability is the degree to which the results of an innovation are visible to others. The results of some ideas are easily observed and communicated to others, while some innovations are difficult to describe to others.

After more fully describing these attributes, Rogers goes on to observe how they relate to the rate of adoption: In the left side of the chart we identify the five attributes perceived by potential adopters, on the right hand side we identify whether the attribute increases or slows down the rate of adoption.

PERCEIVED	Effects the rate of adoption by increasing it	by slowing it
RELATIVE ADVANTAGE	X	
COMPATIBILITY	X	
COMPLEXITY		X
TRIALABILITY	X	
OBSERVABILITY	X	

I think that most of us would agree with the observations presented by Rogers. In fact, we might very well ask: Why was it necessary for him to state the obvious? But he goes a step further by listing the number of empirical studies that support and do not support each of the five generalizations he makes regarding attributes of the innovation and its rate of adoption.

PERCEIVED	Effects		Empirical Evidence		
	Increases	Slows	Agree	Disagree	Total
RELATIVE ADVANTAGE	X		29	14	43
COMPATIBILITY	X		18	9	27
COMPLEXITY		X	9	7	16
TRIALABILITY	X		9	4	13
OBSERVABILITY	X		7	2	9

Not all of Rogers' findings and particularly those related to complexity have been fully substantiated by empirical evidence. That, of course does not mean they are incorrect. It simply means they are hypotheses not yet fully tested.

This review of Rogers' work brings two benefits. First, his findings provide us a sense of direction in finding the most effective ways of understanding the diffusion process so that we can see that our ideas are not only developed but adopted as well. Secondly, we can see from reviewing his approach the need and the importance of empirical evidence to demonstrate the utility of our ideas as we attempt to have them adopted.

Applying Rogers' work in a specific sense, I think one of the key obstacles to be overcome lies within the statistical graphics community itself. Professor Arthur Robinson referred to it in his Presidential Address to the International Cartographic Association in Ottawa in 1972. He said:

> When one looks at the history of cartography, one cannot but be impressed by the persistence of techniques, and the strength of the urge to maintain the status quo. This is true in many areas, perhaps most obviously in the graphic.

He went on to call for the creation of a climate which fosters change and which demonstrates that rapid change is normal.

Keeping in mind the stagnation of the last century regarding graphic presentation, along with the work of Everett Rogers and the words of Arthur Robinson, let me ask a series of questions to the statistical graphics community at large. These are questions we each must deal with if we are to overcome inertia and see the rapid adoption and effective use of a fully automated standardized graphic presentation system.

-- What, if anything, will the system replace?

-- What tools and techniques are likely to be modified as a result of introducing the system?

-- Which would have to be modified if the system is accepted?

-- Who will benefit immediately and who will suffer immediately from a fully automated standardized graphic presentation system?

-- Who in the statistical community will have to abandon or change their occupations as a result of the system?

-- Is there a possibility that the system will open up new forms of cooperation, or perhaps of conflict? If so, will this be within the statistical community, or between that community and the users of data?

-- Do individuals and group leaders -- statisticians, cartographers, demographers, economists, and others who will be affected really understand the nature and purpose of the system?

-- Who will take part in planning a fully automated standardized graphic presentation system, and who will not?

-- What is the history of introducing new ideas within the statistical graphics community?

-- Other than technology, what traditional way of doing things are likely to be affected by the system? Relationships between program specialists and computer programmers? Between the statistical graphics community and data users?

These are tough questions -- and they are only a sample of the many that must be answered before a graphic presentation system such as we have been discussing will be a reality. However, I am confident satisfactory answers can be found, and that the immense pool of talent within the statistical graphics community can be brought to bear, not only on finding the answers, but doing so in a relatively short time.

The Bureau of the Census would like to begin a dialogue with both producers and users of graphic presentations. We hope this dialogue would lead to the development of a graphic presentation center which would assist in developing prototype techniques that can be systematically evaluated "in the field." Additionally the Bureau can serve as a clearing house or depository of empirical evidence which demonstrates the utility of different forms of automated graphic presentations.

In essence, I foresee a situation where the Bureau of the Census, with its available resources of people and machines, could help implement ideas of others, through the use of our existing automated technology and the availability of extensive and timely data sets. This does not mean, I hasten to add, that the Bureau has unlimited resources to contribute to this endeavor. It does mean, however, that we do have some resources and we are willing to open discussions to find how we can most effectively use these limited resources to accomplish the most good.

That's the heart of the challenge I bring with me today. Let us do more in the next ten years toward developing a graphic presentation system than was accomplished in the last 100 years. Let us approach the task not as we would an unknown -- but in the manner of a medical researcher who has isolated the cause of a disease, and is now perfecting a vaccine. Let us systematically collect empirical evidence as we proceed, so that when the time comes to convince others outside the statistical graphics community, we can speak from a basis of fact rather than our own personal desires.

This is a challenge which must be met if we are to help society as a whole meet the larger challenges we have alluded to this morning:

- The forecast of increased information flow, coming at us at an accelerated rate; and
- The dilemma of a society facing difficulty in making good decisions as well as being more realistic about the consequences of making bad ones.
- We need to develop and insure the adoption of a fully automated and standardized graphic presentation system. If we do, we will have contributed greatly to reducing the number of casualties which will occur in a world which is already beginning to face "future shock."