ABOUT STATE AND REGIONAL DATA NEEDS*

Werner Z. Hirsch, University of California, Los Angeles

Introduction

Not long ago Charles Schultze, Director of the Bureau of the Budget, pointed to the newly emerging importance of subnational data, stating "New programs initiated to carry out the objectives of the Great Society must be solidly grounded in factual information. The national effort to raise educational levels, to increase employment, to wage war against poverty and crime, to improve transportation and housing facilities—naming only some of our objectives—requires data not now available. It also requires to a much greater extent than ever before, data on a state or local area basis."1/

In this paper I will discuss some issues involved in furnishing certain data to aid public officials in making state and regional decisions. No attempt is made to identify state and regional data needs for all users. Instead, the main concern is with governmental decisions. I was asked to confine myself to the Census of Population. To a limited extent, some liberties will be taken to go beyond this mandate in discussing ways of collecting new data, including nonpopulation data.

What then is the scope of this paper? At the outset I will explore one way in which state and regional data needs can be identified in a systematic manner. Then the framework will be applied to an example. An examination of ways to supply needed information reveals that the Census of Population already provides much of what is needed; but certain data not now collected would be more readily acquired by integrated recurring surveys than by the current, decennial enumerations. The desirability, scope and feasibility of such survey efforts are then explored.

A Framework for Identifying State and Regional Data Needs

The Bureau of the Census has long shown a deep concern for serving its clients well. Again and again, and especially in preparation for a new decennial census, the Bureau has made an all-out effort to solicit advice on what new and additional information might be needed. A particularly concerted effort was undertaken in the late fifties when the Bureau joined forces with Resources for the Future and sent out questionnaires asking potential users to indicate what new and additional regional information they would want.

It is fair to say that most, if not all, of these efforts have not produced many helpful guides. The main reason is that a shotgun approach was used, and even under the best circumstances, only expressions of highly personalized needs at any one particular moment of time could be obtained. Thus, even if researchers could carefully identify their regional data needs, the Bureau would still face the problem of verifying their validity and ranking their importance and priority. But apparently, in recent years, the Bureau has not been placed in a position where it had to take this second step too often, and there is little evidence that it has a systematic way to cope with the problem. The question therefore arises whether we cannot devise an alternative approach that is more promising.

The approach I suggest is, in a sense, intended to work backwards.2/ It starts by asking, "What are the important decisions to be made with regard to regions and states?" Once these decisions are identified, the next question should be, "What decision rules will be applied to making these decisions?" In a very real sense, the decisions, together with the decision rules, should provide the "demand" for, and establish the character of, the regional and state information needed.

Let us consider this issue in relation to regional public decisions. To decide on relevant data, we must determine in advance what issues they are designed to elucidate and what rules will be used to arrive at a decision. Public bodies need regional information to plan for the future, and to select and operate preferred programs. In planning for the future, governments look for large discrepancies or imbalances that are likely to occur. If, for example, costs are likely to exceed receipts by a wide margin, governments face a problem. If on the other hand, benefits of a program are likely to greatly exceed costs, they face an opportunity. They also look for major changes in magnitudes of variables over time, secular or cyclical, as an indication of major stresses and opportunities.

In formulating programs, i.e., selecting preferred solutions, governments have a set of at least two rules. They are concerned about economic efficiency and therefore want to examine alternative programs in terms of their relative net social benefits. But a government official, particularly an elected official, will not necessarily pick a program because it promises a very large net social benefit. While this might be an essential condition, he insists on another, namely, that the solution be consistent with his political survival. Specifically this means that

---

* Paper prepared for a presentation to the 126th Annual Meeting of the American Statistical Association held at Los Angeles, California on August 15, 1966. The paper benefited from help and advice from Sidney Sonenblum, John M. Kavanagh, Morton J. Marcus and Lawrence H. Revzan.

1. As cited in "Decision-making and Statistical Standards," by James E. Webb, The American Statistician, December 1965.

2. These issues will be discussed in much greater detail in a forthcoming volume, Design of Regional Information Systems.

he will want to know how the losses and gains are likely to be distributed among different interest groups. Detailed information is therefore needed on both social benefits and costs to determine their distribution.

Within such a screening framework, it becomes possible to identify significant data needs and hopefully order them according to their importance. To the extent that this approach succeeds, it should prevent the Census from becoming a data dump when it should be a data bank.

Identifying Some State and Regional Data Needs Within a Decision Framework

I will now apply the framework sketched out in rudimentary form in the previous section to an example, keeping in mind that the Bureau of the Census is mainly concerned with providing raw data and my main task is to reflect on the Census of Population. Therefore I will explore in some detail the nature of population data needed for state decisions on education. Data needed for estimating indirect benefits is not included, and since space is too limited to reproduce the steps that were taken, only the end results are given.

Some of the most important population information I was able to identify as being needed for education decisions, is presented in Table 1, where it is grouped under the following headings: population, households, income, consumer expenditures, leisure and special services. Admittedly this grouping is arbitrary and at one extreme could have been reduced to two categories—population and households, and at the other extreme to a dozen or more categories. In any case, as can be seen particularly from Table 1, there are many cross-classifications.

Most of the population data for education decisions should be available on a city block, census tract, municipal, county, school district and state level.

We also could have produced a table of select population data needed for state and regional water decisions. It would be quite different. For example those who make decisions about water need much less detailed population information than do those facing education problems. Clearly the demand for education is more closely related to age, sex, race, income, etc. than is the demand for water. Furthermore the population data for water decisions should be available on a municipal, county, utility district and state level.

As we compare the information requirements stated in Table 1 with the information in the 1960 Census of Population we find the Census could serve the regional decision maker reasonably well. There are exceptions, however, among them, consumer expenditures, leisure, and specialized services information. It might be possible for the Census to provide some of the additional data needs, but it appears that much of the additional information will require special surveys. For example, I cannot readily imagine the decennial census questionnaire burdened with questions designed to gain information about time spent in recreation participation per week, by age, income group, race and type of recreation. Yet this is crucial information for state and regional recreation decisions. More about the need for special studies will be set forth below.

Another, all too obvious, point should be made. We need up-to-date and recurrent information, often in greater detail, and especially on income, age and education grouping. A further important data gap pertains to detailed migration data as well as intra-area circulation data. Finally, we are often in need of longitudinal data which the Census neglects, except in very rare cases.

I would like to make sure that there is no misunderstanding about the fact that state and regional public decisions require more than population data. Other data needs might be grouped under the headings of manufacturing, commerce and service; policies and attitudes; public service; revenues and expenditures; and property, assets or wealth data.

My assignment excludes a review of the Census of Manufacturers and the Census of Governments. However, I am sure that my earlier discussion about government decisions on education and to a lesser extent on water will prompt readers to ask whether the data identified there are not available in the Census of Governments. First of all let me reply by reminding you that the structure of the Bureau of the Census is such that a given division is concerned with data pertaining to a given phenomenon, i.e., housing, governments, manufacturers, etc. These divisions are not designed mainly to produce information needed by the particular group that is surveyed; and this is not necessarily a bad arrangement. However, to come back to the initial question of whether the present Census of Governments furnishes much of the information needed by public bodies to make state and regional decisions, the answer is an emphatic no. A careful review of the Census of Governments produces the distinct conclusion that much of the information that it presently collects is only in a minor way useful to government decision making. For example, the Census of Governments is mainly looking at administrative budgets which, for decision making, often hide more than they reveal. It would be very useful for the Census of Governments to examine, for example, the desirability of helping to generate programmatic information on governments.

Finally, let me turn to wealth data, not often collected, but found so important by decision makers. The Bureau of the Census could perform a most significant service if it would collect information on the location, physical characteristics, value, utilization, etc. of water, recreation, transportation, education, housing and other physical facilities every 5 or 10 years. For example, in relation to water systems, it would be useful to distinguish between public and private facilities in the following areas: dams, reservoirs, catch and debris basins, aqueducts, drainage facilities, canals, pumping facilities, reclamation plants, desalinization plants, filtration plants, etc.

There is virtue in collecting many of the enumeration data every 5 instead of 10 years since much of the information needed for decision making tends to be of relatively little use if it is old. Decisions have to be based primarily on

Table 1

Select Population Data Needs for State Education Decisions

1. Population
   - (a) total numbers
     - i. age
       - (1) 0-4
       - (2) 5-12
       - (3) 13-17
       - (4) 18-24
       - (5) 25-65
       - (6) over 65
     - ii. sex
     - iii. race
       - (1) white
       - (2) Negro
       - (3) other
     - iv. income
       - (1) 0 - $4,999
       - (2) $5,000 - $8,999
       - (3) $9,000 - $11,999
       - (4) $12,000 - $14,999
       - (5) over $15,000
   - (b) density rates
     - i. central city areas
       - (1) age
       - (2) race
       - (3) income
     - ii. suburban areas
       - (1) age
       - (2) race
       - (3) income
     - iii. rural areas
       - (1) age
       - (2) race
       - (3) income
   - (c) migration rates
     - i. age
       - (1) 0-4
       - (2) 5-12
       - (3) 13-17
       - (4) 18-24
       - (5) 25-65
       - (6) over 65
     - ii. sex
     - iii. race
       - (1) white
       - (2) Negro
       - (3) other
     - iv. income
       - (1) 0 - $4,999
       - (2) $5,000 - $8,999
       - (3) $9,000 - $14,999
       - (4) over $15,000
2. Households
   - (a) total number
     - i. families with children ages 1-18
     - ii. income class
       - (1) 0 - $4,999
       - (2) $5,000 - $8,999
       - (3) $9,000 - $11,999
       - (4) $12,000 - $14,999
       - (5) over $15,000
   - (b) migration rates
     - i. families with children ages 1-18
     - ii. income class
       - (1) 0 - $4,999
       - (2) $5,000 - $8,999
       - (3) $9,000 - $11,999
       - (4) $12,000 - $14,999
       - (5) over $15,000
3. Income
   - (a) per capita income
     - i. total
     - ii. race
       - (1) white
       - (2) Negro
       - (3) other
   - (b) household income
     - i. total
     - ii. race
       - (1) white
       - (2) Negro
       - (3) other
4. Expenditures
   - (a) food
   - (b) clothing
   - (c) total recreation
   - (d) total transportation
   - (e) total housing
   - (f) medical
   - (g) utilities
   - (h) education
     - i. primary
     - ii. secondary
     - iii. higher
     - iv. other
   - (i) other expenditures
5. Leisure
   - (a) average hours worked per week
   - (b) time spent on non-credit educational coursework
   - (c) length of school day
     - i. primary
     - ii. secondary
     - iii. higher
     - iv. other
   - (d) time spent in recreation participation per week
     - i. primary school facilities
     - ii. secondary school facilities
     - iii. higher education facilities
     - iv. other
6. Specialized Services
   - (a) individuals requiring specialized education

current and projected information.

This leads to my next point: while the Census of Population is doing a good job, subject to certain qualifications, we need a massive new effort to produce key annual or biannual survey information. The collection of certain information, especially if it is to be provided on the basis of longitudinal studies, will involve large-scale efforts. They should not be piecemeal and ad hoc. Instead, they should be carried out by a central unit that integrates and coordinates all efforts so they are comparable and cumulative in terms of the information they provide.

Robert C. Wood, Undersecretary of the Department of Housing and Urban Development, in 1963 called for the establishment of urban observatories. They would make "a common series of investigations under a single research plan which for the first time would provide us with professionally reliable findings simultaneously in a number of areas."[3] Such urban observatories could become instrumental in the development of integrated and coordinated surveys and could provide crucial information to improve decision making.

Is the Bureau of the Census the best agency to carry out this function? The Bureau's long tradition and high competence in collecting data clearly gives it a major claim to this assignment. I would assume that it would want to seriously consider assuming such a responsibility, particularly since the need for new and additional data collection by enumeration appears to be declining. However, the Bureau of the Census would have to acquire competence in an area in which it has relatively little experience. This area concerns the identification of important subnational data needs and the evaluation of their priority. Most likely, for this purpose, the Bureau would have to establish new competence within its own organization. But in addition, it might seriously consider joint ventures in cooperation with research centers in universities and private research institutes engaged in urban and regional research.

Summary

State and regional decision makers have been in need of better and more relevant data for a long time, but this has never been more true than it is today when the nation is so greatly concerned with raising educational and employment levels, fighting crime and poverty, improving transportation and housing facilities, developing and cleaning river basins and recreational facilities, etc. Consequently the demand for data is virtually unlimited. Yet data on a state and regional basis are extremely expensive and before they are collected careful thought must be given to their usefulness. The Bureau of the Census, therefore, might have to concern itself increasingly with carefully identifying the most relevant state and regional data. Much of the information will have to be collected frequently by specialized surveys, which should be well coordinated and integrated.

Since most of the Census information is raw data needed for decisions about the future, it is most important to develop reasonably long internally consistent time series. Therefore whenever possible old definitions should be retained and if they have to be changed, the old series should be continued side by side with the new. In a recent article, Elliott R. Morss complained "Whether the objective has been understanding or prediction, it is extremely difficult to understand the recent emphasis on cross-section analyses. For predictive purposes, time series analyses of individual government units would unquestionably be superior to the cross-section approach."[4] Unfortunately, cross-section data are often all the analyst has to deal with.

In order to identify new and additional regional and state data for which there exists a high priority, a framework is needed within which such a rational identification can take place and the costs and benefits of having or not having the information can be evaluated. With this in mind we have proposed such a framework which helps identify key decisions about regions and states together with appropriate decision rules. An understanding of key decisions together with their decision rules can provide the "demand" for, and establish the character of, the regional and state information needs. This is likely to be a relatively new venture for the Bureau of the Census and it will have to develop a new capability in this respect. In doing so, it might be helped by joining forces with established regional research efforts in universities and private research institutes.

Two other important data sources on the federal level are the Internal Revenue Service, whose personal income tax returns provide a gold mine of statistical information, and the Social Security Administration records. But federal data efforts are not well integrated. As we all know a government-wide data service has been proposed recently by a committee of the Social Science Research Council. This proposal has been examined during the last year by the Office of Statistical Standards of the Bureau of the Budget and a White House appointed committee under the chairmanship of Karl Kaysen. Even if these efforts are brought to fruition, however, they will only go part of the way. Much regional and state information is available in state and local governments and private industry, especially in utility companies. No doubt, state and regional data users would greatly welcome a forceful and well conceived effort to integrate state and regional data collection, projection, and dissemination. The present system is inefficient and expensive. Efforts to bring about coordination in this field, if successful, could greatly enhance the chances of providing more appropriate and better quality state and regional data, and thus improve decision making.

---

3. Robert C. Wood, "Contributions of Political Science to Urban Form," Urban Life and Form, Werner Z. Hirsch, editor, (New York: Holt, Rinehart and Winston, Inc., 1963), p. 123.

4. Elliott R. Morss, "Some Thoughts on the Determinants of State and Local Expenditures," National Tax Journal, Vol. 19, No. 1, March 1966, p. 101.