SPATIAL ANALYSIS AND THE POPULATION CENSUS

Griffith Feeney

August 1993

By *spatial analysis* we understand studies that examine data for nations or large regions, national or international, disaggregated into large numbers of geographical units, so that detailed variations in space over large areas may be comprehended. Spatial analysis is distinguished both from national analysis, which tells us about the whole but not about its geographical parts, and from local analysis, which tells us about parts but not about the whole. Spatial analysis tells us simultaneously about the whole and its geographical parts.

Spatial analysis is likely to play an increasingly important role in population studies in coming decades. This will be due in part to the wide range of substantive issues that can be more successfully addressed when spatial variation is taken into account, and in part to technological and methodological developments that will make spatial analysis easier and more effective.

The primary purpose of this paper is to suggest the importance of national population censuses for spatial analysis and to indicate some of the opportunities, demands, and problems that national census offices are likely to face. First, however, it is necessary to say something about the theoretical, substantive, and practical aspects of spatial analysis.

POPULATION CENSUSES AND SPATIAL ANALYSIS

A population census typically involves the canvassing of tens or hundreds of millions of persons. Being vastly expensive and cumbersome, censuses are rarely taken more than once every ten years. The information they provide is consequently out of date by a minimum of five years, averaged over time, and delays in producing census results typically extend this to six or seven years. The complete enumeration of population makes it imperative, moreover, that the number of questionnaire items be strictly limited.

Population surveys involve relatively minuscule numbers of persons, typically thousands or tens of thousands. They are therefore comparatively inexpensive and can be mounted relatively quickly and often. Experience has shown that they can successfully collect massive quantities of information.

By these considerations, the sample survey enjoys all the advantages, the population census all the limitations. Is it any wonder that surveys have been used more and more in population research over the past half century, censuses (relatively) less and less?

For spatial analysis of population data, however, the universal coverage of the population census gives a critical advantage. Every individual in the population is enumerated, and implicit in this enumeration is the location of the residence of the individual. Within the limits set by the questions included on the census schedule, the census records provide complete information on the spatial distribution of population.

CHOICE OF SPATIAL UNITS

The states and provinces into which countries are divided, and for which census tabulations below the national level are most often given, are unsatisfactory for spatial analysis. They are far too large and diverse to provide an adequate representation of spatial variation, and their boundaries are drawn and change in ways that confound systematic analysis.

There are two broad approaches to overcoming the disadvantage of large politico-administrative units. The first consists of disaggregating national data into a large number of small units and then re-aggregating these units into larger units suited to a particular analytical purpose. G. William Skinner's macroregional analysis of China during 1982-85, for example, is based on data for approximately two thousand counties aggregated into ten macroregions. Though the macroregions are larger than China's provinces, they cannot be obtained by aggregating provinces because macroregional boundaries cross-cut provincial boundaries and conversely.

The second approach to spatial disaggregation is more general and elegant, but also more demanding. When census records are geocoded, meaning essentially that individual records contain coordinates of latitude and longitude of the relevant enumeration district, tabulations of census records may be made for arbitrarily defined spatial units. The researcher is therefore free to chose units appropriate to the particular investigation. We may for example ask for the numbers and composition of persons living within 50 miles of a coastline or border, the distribution of persons according to the altitudes at which they live, or for information on populations living in forested areas or deserts.

REGIONAL SYSTEMS THEORY

The richness of the spatial information contained in a population census confounds ready exploitation. The states or provinces into which counties are divided, though far too large and diverse for spatial analysis, are still sufficiently numerous to discourage unit by unit repetition of national level analyses. To secure reasonable control over spatial variation requires hundreds or thousands of units. Clearly we cannot simply repeat for each unit the sorts of analyses we are accustomed to carrying out at the national level. Spatial analysis requires new approaches.

The most fully developed and promising approach at present is the regional systems theory of G. William Skinner. Regional systems theory draws on the central place theory of Christaller (1933; see also Marshall 1989, chapters 5-7) and the land use model of Von Thünen (1826). Systematic development of these ideas is beyond the scope of this paper, but some brief indication of the essential ideas is essential to the discussion, even at the risk of caricature.

Von Thünen's model addresses the tendency of land uses surrounding a city to form a series of concentric zones, with character and intensity of use decreasing as one moves out from the center of the city.

Christaller's central place theory focuses on cities and towns as providers of economic goods and services to a surrounding hinterland. The postulates of central place theory lead to a hierarchy of central places based on a parallel hierarchy of the goods and services provided by these central places. The smallest towns provide a limited set of goods and services to a small surrounding hinterland. The largest cities provide nearly every kind of good and service to a very large surrounding hinterland.

Variation between larger and smaller central places tends to be discrete rather than continuous. Central places may be classified into a limited number of levels, with all places at each level providing a similar set of goods and services. The goods and services provided by places at any given level are also provided by places at each higher level, but not by any city at a lower level. Central places above the lowest level thus have a family of hinterlands, one for the set of goods and services defining this level, and one for the sets of goods and services defining each lower level.

Because higher order central places serve larger hinterlands, there are fewer of them. The hierarchy thus forms a pyramid, with smaller numbers of larger central places toward the top and larger numbers of smaller central places at the bottom. Central places at any given level in the hierarchy tend to be uniformly distributed in space.

Regional systems theory adapts these theoretical approaches to the analysis of social structure in peasant agrarian societies. Developed initially in historical studies of China, regional systems theory has been applied to historical studies of France and Japan and to contemporary China. Given the continuing peasant agrarian character of much of Asia, the theory is directly relevant to contemporary Asian studies of many kinds, including population studies. Certain methodological innovations generated by the theory are likely to be more generally applicable.

Painstaking empirical work on Chinese cities in the late 19th century, involving the classification of some 4,000 central places in a central place hierarchy, showed systematic departures from the pattern expected on the basis of pure central place theory (Skinner 1977a-b). Rather than the uniform distribution of pure central place theory, Skinner found that the cities of traditional agrarian China cluster into nine major regions, each region corresponding approximately to the drainage basin (or portion thereof) of a major river system. Each macroregion has one (sometimes two) central metropolises, large cities at the apex of the central place hierarchy that serve the region as a whole. The regional metropolises are located in the lowland, riverine cores of the macroregion, where soil fertility is highest and transportation costs are lowest. As we move from the core to the periphery of the macroregion, the incidence of higher order cities decreases, a progression which reflects declining soil fertility and increasing transport costs. Broadly speaking, the same social and physiographic influences that define the boundaries of the macroregions give them an internal structure reminiscent of Von Thünen's concentric circles.

These regions, designated *macroregions*, are shown in Map 1, this based on an unpublished analysis by Skinner of Chinese data for 1982-85. All aspects of macroregional structure, including regional boundaries, core-periphery structure, and the position of cities in the central place hierarchy, change over time. The overall pattern in China, however, has been remarkably stable despite the tremendous change experienced over the past century, as may be indicated by a comparison of the structure for 1982-85 shown in Map 1 and the structure for 100 years earlier shown in Skinner (1977a: Maps 1-2).

Every point in the geographical space of a macroregion occupies a certain position in the hinterland of a central place at each level of the central place hierarchy, from the regional metropolis down to the lowest level central places. Position with respect to the highest level central place(s), the regional metropolis(es), defines what Skinner designates the *core-periphery structure* of the region.

There is a strong tendency, modulated by the presence of lower order central places at various points in the region, for social, economic, and demographic parameters of all kinds to vary monotonically as we move from the regional metropolis at the center of the core to the far periphery of the region. Population density, education, and income, for example, all vary systematically from high in the core to low in the periphery. The range of variation is often extreme, sometimes reaching ten to one between core and periphery.

GEOCODING AND GEOGRAPHIC INFORMATION SYSTEMS (GIS)

When spatial analysis is restricted to data from a population census, analysis using small politico-administrative units will suffice for many purposes. If we wish to analyze population census data in conjunction with other spatially referenced data, however, it is essential to have some means of identifying position in space between the census and other data sources. Geocoding provides a such a means.

The significance of spatial analysis in population research depends very much on analysis in conjunction with other data sources. In the study of population and environment interactions, for example, we want to consider population distribution in relation to land use patterns, proximity to coastlines or waterways, and other physiographic and ecological characteristics.

Geocoding of census data opens the way to spatial disaggregation in accordance with any analytical plan whatever. We may study population density in relation to rainfall, numbers and composition of population vulnerable to floods, or distribution of population in relation to altitude and background radiation.

Geographic information systems (GIS) is the generic term for computer systems for processing spatially referenced data. For a review of this very rapidly developing field see Burrough (1986). Effective exploitation of geocoded census data requires the utilization of GIS.

SOME CONSIDERATIONS FOR NATIONAL STATISTICAL OFFICES

The population census is probably the single most extensive and powerful source of information on the distribution of population phenomena in space. This information is vital for the study of many pressing human problems, including the pressure of rapidly growing populations on resources and the impacts of human society on the global environment. Geographic information systems are maturing rapidly and will extend the possibilities for the utilization of this information in the same way as other advances in computer technology have extended the more traditional uses of census information.

These developments raise important opportunities and challenges for national statistical offices, the principal producers of population census data. These new uses of population census data have the potential to create large new constituencies of users of census information, constituencies that can provide important support for the population census. To realize this support, however, national statistical offices must have or develop the capability to respond to the needs of these new groups of users.

What measures can be taken to meet these challenges? Obviously there will be many and important differences between countries, and assessments are likely to become obsolete rapidly. Obviously, also, countries differ widely in the extent to which these challenges are already being met, so that any listing of issues risks looking at once dated for some counties and hopelessly visionary for others. One must begin somewhere, however, and the following list is meant to provide a beginning.

(1) Disseminate basic population data for the lowest feasible level of politico-administrative units, in computer media, and together with suitably constructed administrative boundary maps

Published reports of small area data serve the interests of local government, but they are next to useless for spatial analysis at the national or regional level. Spatial analysis will be carried out by computer, whence the data must be in computer medium. The preferred medium at this point is probably CD-ROM. Published reports are useful primarily as documentation of the computer files.

Since nearly all published data are generated on computers, dissemination in computer medium is easy in principle, but it must be organized and implemented. During the transitional period to systematic computer media dissemination, computer files will all too often have to be recreated from printed material by scanning.

Published data invariably show names of the geographical units, and it is desirable that these be included in the computer files. Processing names by computer is substantially more complex than processing numbers, however, whence the provision of numerical codes for units in computer files is highly desirable.

It is naturally essential to have a codebook showing the correspondence between names and area codes. Because this codebook will contain thousands or tens of thousands of entries, it also should be provided in computer medium. This requires files that accommodate the relevant writing systems, e.g., non-roman alphabets and Chinese characters.

For the purpose of spatial analysis, data for small administrative units must be keyed to corresponding administrative boundary maps. Since these maps will be traced, scanned, and digitized, it is desirable that they show boundary lines only and identify units by brief numerical codes rather than by names. Because names frequently do not fit within unit boundaries, their use is likely to obscure critical boundary information.

Making boundary maps available in a digitized form usable with standard GIS systems obviates the need for scanning or digitizing, but this does not obviate the need for good hard copy boundary maps.

When working with information from a single census, one requires only a list of names of units, usually subsumed in the tabulations themselves, and a corresponding boundary map. The set of spatial units into which a country changes constantly over time, however, and comparison of data from successive censuses, or from a census and another data source with a different reference date, requires systematic control for changes in the population of units.

More generally, the analysis of population census data in conjunction with information from other sources requires the ability to generate unit name lists and boundary maps for any given point in time.

(2) Use population census data as well as other sources to develop spatially referenced data bases on central places

The importance of megacities in the developing world is widely recognized, but these cities are only the apex of the central place hierarchy. Published information on cities and towns should be supplemented by information in computer media, with numerical codes as well as place names. Central places should be geocoded for analyses involving other spatially referenced data.

The population census provides an important source of data, but there are many other sources, including censuses of establishments and of industry, directories of government and commercial services, and compilations of data on cities, towns, and villages.

Central place files must address the divergence between administrative and `real' boundaries, which varies with context. In the west, cities tend to overflow their administrative boundaries, leading to the creation of such units as the Standard Metropolitan Statistical Area (SMSA) in the United States. In China, we face the opposite problem, with geographical areas administratively designated as cities and towns containing substantial rural populations.

(3) Develop GIS capability to analyze census data in relation to data from other sources

Geocoding of census data is not required to put GIS technology to good use, but it is necessary to exploit it fully. Most national statistical offices that have not already done so are likely to introduce geocoding and develop at least limited GIS capability in time for the year 2000 round of censuses.

GIS technology is rapidly maturing, with low-cost systems for microcomputers are increasingly available and user-friendly. Formats, standards, and procedures for dealing with digitized spatial data are developing rapidly. The spread of microcomputers, word processing, and spreadsheets over the past decade provides a useful indication of what we are likely to see in the years ahead.

The creation of digitized maps to a high standard of accuracy is a highly specialized function, however, and one that may in some cases be best carried out in government departments other than the national statistical office. Questions of the division of labor and the sharing of information between the national statistical office and other agencies will surely arise. National statistical offices need to know what capabilities exist and what is being done in other units of government.

CONCLUSION

Every population census contains an unexplored and unexploited wealth of information on spatial patterns of human characteristics and behavior relevant to a wide variety of social problems. This paper has attempted a very cursory sketch of the possibilities and importance of making better use of this data. The essential argument is that these possibilities pose a significant opportunity for census takers to increase the utilization of census data among traditional users and to create important new classes of users.

ACKNOWLEDGEMENTS

I am grateful to G. William Skinner for numerous extended discussions on the subject of regional systems theory, for comments on an earlier draft of this paper, and for permission to use Maps 2

and 3. It should hardly be necessary to add that any shortcomings are my own.

REFERENCES

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford: Clarendon Press.

Christaller, Walter. 1933. Die zentralen Orte in Suddeutschland. Jena. (Central places in Southern Germany, translated by Carlisle W. Baskin. Englewood Cliffs, N.J.: Prentice-Hall, 1966)

Marshall, John U. 1989. The Structure of Urban Systems. Toronto: University of Toronto Press.

Skinner, G. William. 1964/65. Marketing and Social Structure, Parts I, II and III. Reprinted from The Journal of Asian Studies 24(1-2-3). AAS Reprint Series No. 1. Association for Asian Studies.

--. 1977a. Regional urbanization in nineteenth-century China. In G. William Skinner, ed., The City in Late Imperial China. Stanford, California: Stanford University Press.

--. 1977b. Cities and the hierarchy of local systems. In G. William Skinner, ed., The City in Late Imperial China. Stanford, California: Stanford University Press.

Von Thünen, J.H. von. 1826. Die isolierte Staat .... Berlin: (Isolated state, translated by Carla M. Wartenberg. Oxford: Permagon, 1966.)

SPATIAL ANALYSIS AND THE POPULATION CENSUS

August 1993

Griffith Feeney
Program on Population
The East-West Center
Honolulu, Hawaii 96848 USA
August 1993

Paper presented to the 15th Population Census Conference, co-sponsored by the National Statistical Office of the Republic of Korea and the East-West Center Program on Population, Seoul, Korea, 17-20 August 1993.