Volume Title: Regional Income

Volume Author/Editor: Conference in Research in Income and Wealth

Volume Publisher: NBER

Volume ISBN: 0-870-14177-5

Volume URL: http://www.nber.org/books/unkn57-3

Publication Date: 1957

Chapter Title: The Geographic Area in Regional Economic Research

Chapter Author: Morris B. Ullman, Robert C. Klove

Chapter URL: http://www.nber.org/chapters/c7602

Chapter pages in book: (p. 87 - 112)

The Geographic Area in Regional Economic Research

MORRIS B. ULLMAN AND ROBERT C. KLOVE,
BUREAU OF THE CENSUS

Because it is practical, quantitative data are usually grouped in classes and expressed as totals or averages. But if the classes are too inclusive, the totals or averages may confuse rather than clarify the outlines of a phenomenon. Economic analysis may need the data presented in small classes or subunits. Often these are geographic or spatial.

Solutions to similar problems tend to fall into a pattern. The pattern may be considerably influenced by the form in which the basic data are presented, especially where recompilation of the data is expensive and time-consuming. Yet, to the best of our knowledge, existing regional or areal patterns have never been systematically examined to determine their aptness for the problems of economic research. A complete examination of this type is beyond the scope of this paper. But we describe some of the geographic areas now available, indicating their usefulness for economic studies, and summarize some of the factors that influenced their formation.[1] In the last section, as an example, we describe the procedure followed in establishing a set of regions consisting of states.

Geographic Areas Available for Statistical Use

The standard sources of area data are the statistical reports of government agencies and, in particular, the reports of major censuses. In a major census, the Bureau of the Census usually collects some information from all units in the universe and, therefore, can prepare significant subtotals for any detail desired, if they will be generally useful and funds are available. In fact, a major contribution to regional research by the Census Bureau is the determination

[1] Two important factors are not discussed. Administrative requirements have been responsible for developing the use of geographic areas and for the availability of a significant amount of data for particular regions. Examples of studies of administrative factors are *Area and Administration* by James W. Fesler (University of Alabama Press, 1949), and the comprehensive study of the National Resources Committee, *Regional Factors in National Planning* (1935). Historical developments also have influenced the formation of many of the regions now in use. An excellent series of chapters, some of which cover this aspect of regional formation, is presented in *Regionalism in America,* Merrill Jensen, editor (University of Wisconsin Press, 1951).

of which areas are of most widespread use and which detail should be published for them.

The maximum amount of geographic detail is published for the censuses of population and housing, and the areas used include not only states, counties, cities, and minor civil divisions but also such aggregates as standard metropolitan areas, urbanized areas, state economic areas, census tracts, unincorporated places, city blocks, and others.[2] Other censuses, such as those of agriculture, manufactures, and business, have selected those areas required by the data.

POLITICAL AREAS VERSUS STATISTICAL AREAS

Until a half century ago, federally collected statistics were almost exclusively for political areas, all with definite legal boundaries. Recognition of the need for special statistical areas developed slowly and for a long time almost all were combinations of political units. During the nineteenth century, states were combined into large geographic regions which by 1910 had crystallized into the set of census regions and divisions observed to this day. In recent decades, counties were combined into types of farming areas and, in 1950, into the new state economic areas and economic subregions. Selected counties or combinations of counties were recognized as early as 1900 as industrial districts by the census of manufactures, and, in 1950, the standard metropolitan areas were defined in terms of counties by a federal interagency committee.

Other statistical areas with boundaries not necessarily observing the limits of political units have been developed in recent years. This type of area represents an attempt to define statistical areas precisely in terms of their inherent characteristics or relationships and recognizes that many of the political units have lost much of their original structural and functional meaning. This development includes such areas as census tracts, urbanized areas, unincorporated places, and census county divisions. Census tracts were first used as early as 1910; the others are of recent origin.

Each of these areas has value for certain purposes. Urbanized areas present the extent, the pattern, and the statistical characteristics of the truly urban portions of the metropolitan areas. Census tracts provide detailed information about the internal structure of the metropolitan areas, or about the similarities and differences of their component segments. Unincorporated places are communities

[2] Definitions of these areas are included in the reports of the 1950 censuses of population and housing and also in a four-page geographic report, *Census Areas of 1950*, Bureau of the Census, Series Geo. 1, 1951.

not recognized in earlier statistical enumerations, but communities which, size for size, are fully as significant in most respects as their formally organized, legal municipal counterparts. Census county divisions are presently being delimited in some states to replace frequently changing civil divisions or those with boundaries unfamiliar to the local inhabitants. These areas facilitate enumeration and improve the significance of local statistics.

All of the statistical areas mentioned above are census areas, but other areas with nonpolitical boundaries are also used by business and by local community organizations. The use of special areas for statistical purposes may be limited by the lack of census or other general data for comparable analysis. The development of census areas, as well as the use of existing political areas, offers economic research ready-made areas of considerable potential statistical value.

AREA HIERARCHIES

In regional analysis, areas of all sizes may be needed either separately or in sequence. For a broad-gauged regional study of the country as a whole, large areas are desired. Where a regional study seeks an understanding of major sections of the country, areas of intermediate size are useful, but for a detailed understanding, small areas are required. To serve these uses, various area hierarchies are available which, for this descriptive purpose, may be divided into four principal types: (1) all political units, (2) groupings of states, (3) groupings of counties, and (4) metropolitan or urban areas.

1. Political units form the most commonly used area system, ranging in size from the nation through states and counties to the smallest civil divisions of counties and cities. The large number of these areas and the fact that their boundaries generally fail to observe precise physical and cultural differences reduce their usefulness.

2. State groupings form a limited system in terms of size range, because the state component is in most cases a rather large area. In census use, these consist of either three or four census geographic regions (Northeast, North Central, South, and West) and nine census geographic divisions. Others have used different combinations of states.

3. County groupings offer the greatest opportunity to develop a broad-ranged area system that will make for more precision in regional analysis, although up to the present their use has been restricted because of the limited acceptance of area groupings based on counties. The system of 501 state economic areas devised for

the 1950 census by the Census Bureau in cooperation with the Department of Agriculture appears to be a beginning in the development of widely accepted county groupings.[3] The 501 areas were combined into 119 economic subregions for census use.[4] Bogue has gone further in combining them into thirteen economic regions and five economic provinces.[5] Only the state economic areas observe state boundaries; for the others, these lines are ignored.

Statistical area hierarchies for the whole country based on political units smaller than counties are impractical because of the difficulty and cost of collecting statistics for minor civil divisions or other small areas, the large number of units involved, the absence of accepted techniques for achieving general purpose areas at this level of precision, and for other reasons.

4. The urban area hierarchy is made up of a number of areas defined in various ways. These include metropolitan areas, urbanized areas, the political cities, communities, census tracts, city blocks, and others.

This pattern of developing regions in terms of a collapsible system or a hierarchy has the advantage of presenting data so that the analyst has available either large or small areas or can analyze the results of one level of grouping, in terms of another level, with a minimum of work.

AREAS FOR CURRENT DATA

Apart from census data, there is little statistical information for small areas available on a comparable basis nationally, largely because of the cost. In the increased development of broad constructs, such as national income, and the increased use of sampling surveys, the emphasis is placed on speed and on national totals. For constructs composed of a large number of elements, the precision of the results is related to the size of the aggregate. In sampling, the precision depends, in general, on the number of cases rather than on their proportion to the total. Thus, a sample for a region requires about as many cases as a sample for the United States to obtain the same degree of precision. In planning programs and designing surveys, cost factors are fundamental in the recent trend of limiting current data to national totals or, at best, broad regions.

[3] *State Economic Areas*, Bureau of the Census, 1951.

[4] *Economic Subregions of the United States*, Bureau of the Census and Bureau of Agricultural Economics, Census—BAE 19, 1953.

[5] Donald J. Bogue, "An Outline of the Complete System of Economic Areas," *American Journal of Sociology*, September 1954.

Conceptual Basis for Delineation of Statistical Areas

Statistical areas, as defined in the preceding section and as used by statistics-collecting agencies, are almost all general purpose areas. They are adapted to the presentation of data on particular subjects; some for agriculture, others for industry, and still others for population use or for distinguishing between the urban and rural aspects of our economy. Yet each is designed for many uses not all of which can be anticipated. Indeed, general purpose statistical areas are always used for a special purpose, because analysis, regional or otherwise, has no meaning except in terms of special purposes.

In view of this, why not have a set of special purpose areas for each regional problem? If areas were developed for each problem, statistical chaos would result. Regional problems are related, but if the analytical results are not comparable in terms of area, the relationships are largely lost. This sort of regional statistical confusion exists now to some degree, but steps are being taken by federal statistical agencies to correct the situation.

Illustrating successful action in achieving widely adopted general purpose areas for general statistical use are the standard metropolitan areas defined under the direction of the Bureau of the Budget. Several federal agencies had used metropolitan areas defined in various ways before, but most now use a single, uniformly defined set of areas. These areas have also gained wide acceptance outside government both locally and nationally. This trend toward standard areas is only in its beginning, but there is evidence that it is satisfying a need for regional statistical order in the geographic presentation of statistics.

Comparability of results is an important reason for using general purpose statistical areas, but not the only one. The problem of retabulating data for small areas into a set of special purpose statistical areas is frequently not feasible because of the time and cost required. An additional difficulty is the defining of meaningful special purpose areas. The collecting agency has other reasons for using general purpose areas. Many types of statistical data cannot be made available for small area "building blocks" but can be presented for larger statistical areas. For example, industrial and business statistics cannot be presented for small areas because of the disclosure law, which prohibits revealing statistics that apply to a single establishment or a small group of establishments. Statistics collected on a sample basis have low reliability for small areas and are therefore presented only for larger areas. Survey errors in the collection and

tabulation of statistical data are sometimes significant in the presentation of small area data and the problem of removing them is most costly.

The foregoing discussion does not in any way deny the need for small area, statistical building blocks. They are a necessary and essential part of any broadly conceived and useful statistical program and presentation. General purpose statistical areas cannot satisfy all the needs of regional statistical analysis. But they can satisfy many special purpose needs and they can help to point up the kind of special purpose areas required for those problems whose needs they do not precisely meet. Any large statistics-gathering agency, such as the Census Bureau, must keep a proper balance in its area publication program between data for general purpose statistical areas and the building block statistics.

The technical procedures as well as the principles or criteria for the delineation of statistical areas have been slow in developing and rarely have been stated or presented in detail. Little record exists of how the earlier census statistical areas were chosen, but what evidence there is indicates that the methods were heavily subjective. The set of geographic regions and divisions still used by the Census Bureau were developed before the turn of the century by Henry Gannett, Geographer of the Census, but the considerations which led to these areas are largely lost to history. Gradually a number of the factors that should govern the delineation of statistical areas have been formulated. Many are implicit in discussions about statistical areas and in their definitions, but not all are given the consideration due them. Persons using statistical areas for regional analysis need to understand the basis of delineation so as to select the proper areas for the problem at hand. Where the researcher must devise his own special-purpose areas, knowledge of the conceptual basis of the statistical areas is even more important.

HOMOGENEOUS OR FUNCTIONAL EMPHASIS

The delineation of regions involves one of two fundamental criteria. One may define a region by the likeness of its component characteristics—the homogeneity principle—or by the presence of a nucleus and an area of influence—the functional integration principle.

An area may be alike in a single characteristic or in a group of characteristics. It may be alike in some but differ in others, but the similarities in the characteristics and trends must outweigh the differences. The characteristic may be population, economy, land

use, physical environment, or another basic characteristic or combination of characteristics. The measurement techniques involve the use of statistical indexes, maps, aerial photos, and field work. Such areas are useful for analyzing differentials, since the assumption can be made that all units within an area are similar.

The census geographic regions and divisions, which are combinations of states, recognize certain broad homogeneous characteristics of the different parts of the country. Having been in use for over a half century during which the United States has experienced many changes, they may need to be changed too. The last section of this paper discusses the work of a Department of Commerce committee that has been studying a proposed revision.

Other examples of homogeneous statistical areas are census tracts and urbanized areas. Census tracts, the small areas into which larger cities and their metropolitan areas are divided, are more or less homogeneous when first devised but tend to lose this homogeneity with the passage of time. Fundamental changes in their layout are inadvisable because comparability with the past would be lost. On the other hand, urbanized areas which are designed mainly to mark urban fringe settlement at the time of the census, must be changed with each population and housing census because, at least in recent years, urban fringe growth has been so rapid in this country.

The functional integration principle requires that a region comprise all the area that is organized around a centralizing node, or nucleus, and that operates as a unit. Functional areas include metropolitan areas, community and neighborhood areas, trading areas, traffic flow areas, and all other areas that are integrated through communication or movement of one kind or another. Integration, like homogeneity, may be based on a single factor or on a group of related factors, but may be more difficult to measure. Few statistics are collected on movements, trading currents, traffic flow, or communications, and fewer yet are published. What is available is often difficult to interpret. Again a good deal of information on the limits of these areas may be obtained through field observation and mapping, but this technique is costly and often not feasible.

Among the statistical areas used by the Census Bureau, standard metropolitan areas and census county divisions are examples of functionally integrated areas. The standard metropolitan areas try to show the area of close economic and social integration around the larger cities of the country; census county divisions, the trading or community areas within counties. The census county divisions are new statistical areas. Most of them will be used for the first time

in the 1960 censuses and will replace the minor civil divisions in those states where the latter have been inadequate for one reason or another.

Actually, in the delineation of most statistical areas, both principles are used to some extent. The difference is mainly in emphasis. For example, census tract criteria emphasize homogeneity but also recognize the importance as boundaries of such major barriers as railways, freeways, and waterways which often separate one community from another. Indeed, some census tracts are neighborhoods, and in many cities census tracts in combination are recognized as communities, retail trade areas, or central business districts. In the case of standard metropolitan areas, the criteria may be divided into two kinds: those concerned with metropolitan character including population concentration and place-of-work concentration and those concerned with economic and social integration between the central city and outlying areas. Furthermore, standard metropolitan areas fit into the country-wide system of state economic areas. The nonmetropolitan state economic areas were defined almost exclusively on the principle of grouping together those counties with similar economic and social statistical indexes. In this broader system of areas, the standard metropolitan areas clearly represent homogeneous areas distinct from their nonmetropolitan counterparts. Considerations of scale or level in the hierarchy of areas are always relative, and what may be homogeneous among large areas may be grossly heterogeneous among small areas.

While homogeneity and functional integration are basic criteria in the delineation of areas, a number of other factors require careful consideration. These we shall discuss mainly from the point of view of the analyst or the user of statistics.

NUMBER, SIZE, AND GRID

Number of Areas. Other things being equal, the number of areas used for analysis of a region should be the minimum necessary for the detail of analysis required. For example, in studying agricultural potentials of the Midwest, one might use states or state economic areas but ordinarily not counties. In studying a single state, counties are appropriate. If the region of study is a state economic area or a county, data for minor civil divisions or enumeration districts are preferable. The amount of time needed for posting and compiling data will be in direct proportion to the number of areas used.

Size of Areas. The size of statistical areas, assuming all other factors are nonrestricting, depends on the degree of detail desired.

If only broad regional generalizations are needed, larger areas, such as geographic divisions or states, can be used without hesitation. If great detail is required for rural areas, minor civil divisions or smaller areas can be used; for urban areas, census tracts or even blocks, where available. The areas, however, should be large enough to furnish significant data in the detail needed.

Uniform size, either in terms of area or of population, does not seem to be required for most problems. Too great disparity, however, tends to make interpretation difficult. Per capita figures, density, or other ratios are often used to overcome the size differences between areas. Another technique often used is analysis in terms of size groupings.

Completeness of Grid. Another choice to be made in selecting the geographic area is the degree of completeness with which the areas include the universe. Most political areas, as well as the larger statistical areas, cover the entire country. Regions, divisions, states, state economic areas, and counties, include all parts of the United States.

Small statistical areas, such as census tracts or blocks, usually cover only a limited area. These are quite useful for special purposes, but caution must be exercised in using such areas for studies that may require analysis beyond the area covered. However, if a complete grid is needed and the coverage is not complete, it is sometimes possible to substitute other small areas for territory not covered by the grid.

BOUNDARY AND RECOGNITION PROBLEMS

Definiteness of Boundaries. Boundaries should be definite, well known, and easily identifiable through observation or inquiry in the field. Otherwise, accurate allocation of data is impossible and its interpretation is difficult. Incorrect boundary information may result in the omission or duplication of certain areas and may affect the analysis adversely.

When statistical areas are defined in terms of political units, the boundaries are generally assumed to be definite and accurate, but a word of caution is in order. In the United States, the Bureau of the Census regularly seeks to improve the accurate recording of the political boundaries observed in its surveys. Many boundaries change frequently, particularly those of incorporated cities, and many are still not accurately shown on maps, partly because the land area to which they apply has never been surveyed and mapped in accordance with acceptable standards. Some boundaries are not precisely known by local officials or inhabitants. Nevertheless, for all

practical purposes the census coverage for political areas in the United States is highly reliable and ordinarily is accepted by the user of the statistics. Maps showing the political composition of these statistical areas are a requirement for interpretation of the data.

When statistical areas are defined in terms of unique nonpolitical limits, the boundaries must follow easily discernible features, such as highways, railroads, streams, or similar objective lines that are subject to ready identification by enumerators. Many users of the data for these areas require maps for analysis.

Precision of Definition. It is important to recognize how closely the boundaries follow the geographic entity defined. The definition of standard metropolitan areas in terms of entire counties is an approximation adopted for the convenience of using county data, since a certain amount of nonintegrated area is normally included and a small amount of integrated area may be excluded. By contrast, the urbanized area limits itself more precisely to the central city and the built-up area around it. Precision is obtained by the use of highly irregular boundaries for an area for which limited general data are available.

Precision of definition is a feature of the areas established specifically for statistical purposes and is usually related to the size of the units or building blocks. Geographic divisions and regions of the United States certainly are not precision areas. Economic subregions and state economic areas are more finely drawn, but the most precise areas would not follow political lines. Precision of definition is an objective in the design of census tracts.

In some economic problems, a precise definition of areas, either homogeneous or functional, is necessary. If precisely defined areas do not exist, they must be defined, usually by combining areas for which statistics are already available, such as city blocks, census tracts, minor civil divisions, or counties.

Contiguity of Area. Area contiguity is desirable for practical purposes, but exceptions are sometimes made. For example, many urbanized areas have urban exclaves that in terms of all other characteristics are a part of the main urban fringe. A few census tracts and economic subregions also have exclaves. The presence of enclaves in a few statistical areas is a related feature.

Recognition of Area. To all users of statistical areas, the ability to recognize each area under analysis, that is, to know in general its meanings, its location, and its limits, is important. At the national level, users are familiar with New England, the Middle Atlantic states, the North Central region, etc., which are defined in terms of

states. Recognition of areas defined along county boundaries, is not quite so easy. Bogue's economic regions contain areas, such as the Great Lakes and Central Appalachian regions, that most users would recognize fairly easily but would require a map for adequate understanding. When the United States is divided into smaller areas, such as the state economic areas or the retail or wholesale trading areas of individual cities, recognition of the area definition requires maps or lists of the component units, because the boundaries between areas may be drawn in so many different ways, and also because the areas may be identified only by number or letters in the presentation of the data.

Some regional economic descriptions may not require knowledge of the exact boundaries. State, county, and city statistics are often used by persons who know only the general locations of the areas. It is often more important to understand the concept of the area and know its general location than to have precise knowledge of its boundaries. Every student is familiar with the major political area concepts—state, county, and city. The minor civil divisions are less well known, but the concepts are familiar to most persons, especially in their local areas.

For the special statistical areas, knowledge of the precise boundaries can only be known through careful examination of maps or lists of political unit inclusions. Understanding of the concept of these areas make many of them useful without graphic definition. We refer to such areas as urbanized areas, unincorporated places, and census county divisions. For census tracts that are small divisions of large cities and their metropolitan areas, most analysis requires maps.

DATA AVAILABILITY AND COMPARABILITY

Availability. The state and the county are the most practical units for area construction, since most census data are available by states and counties. Many areas have considerable local information, such as utilities data, local market data, local school data, etc. However, most such data are unique to a particular area, and when several units are to be studied together, very frequently the data are not comparable because of differences in concept, techniques of collection, or method of estimation. This leaves the major censuses as the principal source for comparable data over the country as a whole.

In producing current data, early availability is emphasized. Grouping by small areas is usually avoided, since it delays the release of the information and is more expensive. Among the exceptions where

current data are available for counties on a national scale are the social security data published in *County Business Patterns*[6] and bank deposit data. These exceptions, however, are very few and most of them are represented in the *County and City Data Book* compiled by the Bureau of the Census.

Comparability. For any trend or historical study, the researcher must be aware of changes in both the type of area and its boundaries as well as any conceptual changes in its definition. Many areas are of very recent origin, e.g. standard metropolitan areas, urbanized areas, unincorporated places, state economic areas, etc. Where these areas are defined in terms of political units, it is possible to assemble data for earlier years. For areas with unique boundaries, earlier data are not available.

Even data for political areas must be used with care, because the smaller areas change. Many counties have been subdivided since the early censuses; minor civil divisions in some states change substantially during every decade; and hundreds of annexations and detachments are made each year for the incorporated places. These changes from one census to the next are recorded in the footnotes in Volume I of the census of population. For the future, it should be kept in mind that many of the new statistical areas, such as urbanized areas and unincorporated places, will change their limits from one census to another. These areas will have conceptual comparability but not area comparability.

Delineation of a General Purpose Grouping of States

The defining of a set of regions usually starts with the selection of criteria appropriate to the use to be made of the regions. When the regions are to be used for many purposes, some of which cannot be anticipated, one can select a single broad criteria on the basis that it is related to many other criteria and assume that such a regionalization will have general utility. The criteria might be income, population, or the type of economy.

Another approach would be to select a number of criteria and develop a technique for either combining them into a single measure or reconciling differences between regionalizations that would have resulted if the different criteria had been used separately. Such technicians as Odum, Elliott, Mangus, Lively, Baker, Hagood, Bogue, and others have used variations of this approach.

A description of the actions of a committee that undertook such

[6] *County Business Patterns,* Bureau of the Census and Bureau of Old-Age and Survivors Insurance, 1955.

a regionalization for presenting general purpose statistics illustrates the problems of area selection and delimitation. This committee, consisting of representatives of the Bureau of the Census, the Office of Business Economics, and the Area Development Division of the Office of Technical Services, considered the feasibility of standardizing the regional presentation of general purpose data within the Department of Commerce. The Office of Statistical Standards of the Bureau of the Budget was also interested in the work of this group and is now considering their report for possible government-wide application. In view of potential use of the new regionalization, it was decided to use states as units. States have definite boundaries, completely cover the United States, are readily recognized, and have all types of data available which can readily be made comparable with the past by simple compilation. On the other hand, the state is a large unit and many states are quite heterogeneous, making precise definition difficult, if not impossible. The first questions to be faced were: What should be the basis for the grouping? How many groups should be made?

After preliminary consideration, the committee agreed on the following ground rules:

1. Homogeneity with regard to economic and social factors was to be the principal criterion for grouping states into regions.
2. Each region was to consist of two or more geographically contiguous states.
3. Insofar as feasible, the classification was to be based on objective methods, and the influence of personal judgment was to be minimized.
4. The number of regions to be delineated was not decided at the beginning of the work, but six and twelve were set as the outer limits.

PROCEDURE

The committee examined the approach in which the classification is based on the homogeneity of a large number of individual measures which are combined to depict the economic and social structure of a state. Such an approach—whether carried out by visual inspection and analysis or by a formal statistical method—places a heavy premium on personal judgment in selecting and weighting the factors. No general framework was found to exist for describing statistically the over-all economic and social structure of an area.

In the absence of such a general framework, the committee turned to two bodies of data to describe the economic composition of the

various states: (1) the series of income payments, which is the most comprehensive measure of economic activity available on a state basis, and (2) the data on industrial distribution of the employed labor force. These data, the committee agreed, would furnish a framework for drawing a regional classification based on economic factors.

Another regional classification based on noneconomic factors [7] was then discussed. Six subject fields were defined as covering the pertinent factors: (1) population size, distribution, and growth; (2) racial and ethnic composition; (3) socio-economic status or level of living; (4) transportation and communication; (5) health; and (6) history and tradition. Statistical series were to be selected to describe the first five fields. The sixth, history and tradition, could not be quantified, but it entered into consideration later in preparing the groupings.

The regional classification based on economic factors and the one based on noneconomic factors were to be merged into a single regional grouping. Finally, the new regional system was to be tested statistically for homogeneity against the three systems currently in use within the Department of Commerce.

GROUPING ACCORDING TO INCOME PAYMENTS

The committee considered the following broad aspects in drawing the map based on income payments: (1) composition of income in 1950, (2) level of income in 1950, and (3) trend of income over the period 1929–1950. Of these three, composition of income was considered most important. Both the level and trend are, in part, a reflection of it.

The composition of income payments in each state was measured by expressing the income derived from each industrial source as a percentage of total income in the state. Accordingly, distributions were made among twenty-three income sources from data available on the worksheets of the Office of Business Economics. A finer classification of income sources than this appeared to be of limited additional value. More detailed data were used for reference, however, in determining the classification of some states.

In the United States, manufacturing and agriculture account for much of the interstate variation in composition of income. The first step, therefore, was to classify the states into groups by the relative importance of manufacturing and agriculture. These groups were modified to satisfy the criterion of geographic contiguity. The states

[7] In the work of the committee and in this description the term "noneconomic" is used to represent data other than national income and labor force data.

within the resulting groups were then analyzed for homogeneity in regard to all twenty-three components. At this stage, the trend and level of income were also considered, and if the position of a state was doubtful, a finer classification of income payments was used.

GROUPING ACCORDING TO LABOR FORCE

The states were grouped according to industrial employment as reported in the 1950 census of population to check the grouping based on income payments. For this purpose, the number of employed in each of eighteen industrial groupings was expressed as a percentage of the total number employed in the state. To determine a breakdown for agriculture, data from the 1950 census of agriculture was used to subdivide employed persons in agriculture into nine types of farming groups, making a total of twenty-seven groups. The procedure of visual inspection and analysis, which was used for the income grouping, was also used for the labor force grouping. To minimize the effect of personal judgment and other bias, the regionalization on the basis of the labor force was prepared by persons other than those who worked on income payments.

GROUPING ACCORDING TO NONECONOMIC FACTORS

The committee assembled available statistical series for each of the noneconomic factors and used them to group the states into regions. This, too, was done by visual inspection. For each of the five groups listed above, the committee selected a few series that seemed to be best suited to show differences among the states.

Frequently, the committee encountered a state that seemed to fall between two regions and to have some of the characteristics of each of them. In fact, the regional assignment of fourteen of the forty-eight states was not definite. It was evident that, while the core of each region was homogeneous, the peripheral areas took on the characteristics of adjoining regions. The committee subjected each of the problem states to intensive examination to determine the grouping which it most resembled.

To illustrate the procedure used, Maryland was assigned to the Middle Atlantic states on the basis of the following decisions for the different series:

1. *Population size, distribution, and growth*	*Middle Atlantic*
Population per square mile, 1950	Middle Atlantic
Percentage change in population, 1940 to 1950	marginal

Percentage of population in urban areas, 1950	Middle Atlantic
Percentage of population in open country (outside places of 1,000 or more), 1950	Middle Atlantic
Net migration, 1940 to 1950, as percentage of 1950 population	Middle Atlantic
Percentage of population living in same house in 1949 and 1950	Upper South
Births in 1950 per 1,000 population	marginal
Percentage of nonfarm dwelling units in one-dwelling-unit detached structures, 1950	Middle Atlantic
2. *Racial and ethnic composition*	*marginal*
Percentage of population who are Negro, 1950	Upper South
Percentage of population who are foreign-born White, 1950	Middle Atlantic
Percentage of population who are Negro, Indian, or White with Spanish surname	not applicable here
3. *Socio-economic status and level of living*	*Middle Atlantic*
Median income of families, 1949	Middle Atlantic
Median income of families and unrelated individuals, 1949	Middle Atlantic
Median years of school completed by persons twenty-five years old and over, 1950	marginal
Farm-operator family level of living index	Middle Atlantic
Ratio of children plus aged persons to persons in labor force, 1950	Middle Atlantic
4. *Transportation and communication*	*Middle Atlantic*
Registered borrowers in public libraries, 1945, per 1,000 population, 1950	Upper South
Telephones per 1,000 population, 1950	Middle Atlantic
Tons of newsprint consumed for newspapers, 1947, per 1,000 population, 1950	Middle Atlantic

Percentage of farm operators who travel 0.0 to 0.2 miles over dirt or unimproved road to get to trading center	Middle Atlantic
Ratio of persons eighteen years old and over to total automobiles registered	Middle Atlantic
5. *Health*	*Middle Atlantic*
Infant deaths per 1,000 live births, 1949	Middle Atlantic
Death rate from tuberculosis, all forms, 1949	Upper South
Deaths per 1,000 population, 1950	marginal
Percentage of births attended by physician in hospital, 1949	Middle Atlantic

INTEGRATION INTO A SINGLE GROUPING

The committee examined and merged the regional classifications developed from the economic and the noneconomic data. States that were marginal were reviewed individually. Data for economic and noneconomic factors were examined for the marginal states and an agreement on the most appropriate classification was reached. The result put the forty-eight states into nine groups as shown on Map 1. The combination of these divisions into a grouping of four regions was also reviewed. The results of this grouping is shown on Map 2.

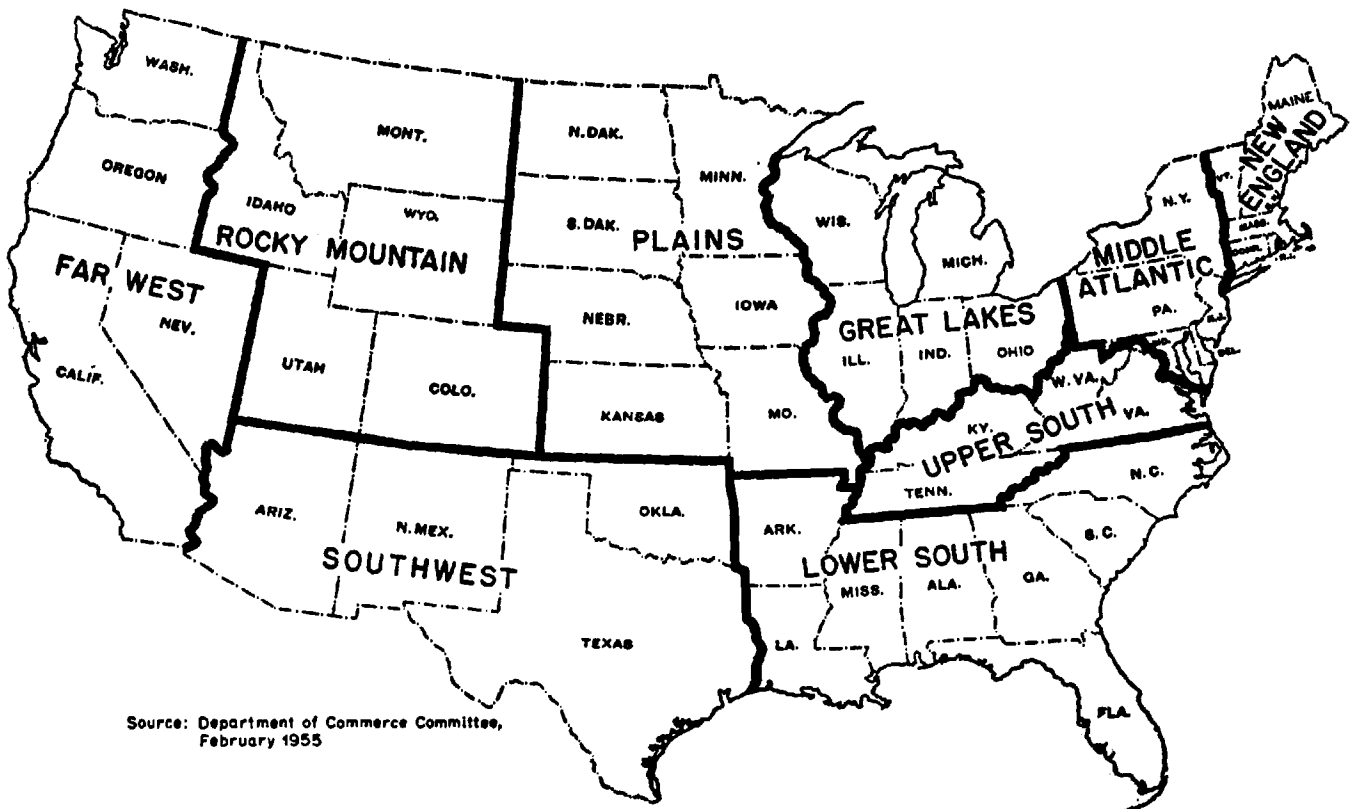
SOME PROBLEM AREAS

Assignment of Louisiana and Missouri. The economics-based system placed Louisiana in the Southwest and Missouri in the Plains region. On the basis of social characteristics, on the other hand, Louisiana was classified in the Lower South and Missouri in the Great Lakes region. In each case the assignment was marginal. Further analysis of the income payments and social data resulted in the classification of Louisiana in the Lower South and Missouri in the Plains region.

Division of the Southeastern Area. The Office of Business Economics did not regard any subdivision of the southeastern areas as desirable for the final classification. The Census Bureau, however, recommended a two-way breakdown into Upper South and Lower South. While this subdivision was an outgrowth of the noneconomic analysis, it was also based on certain pragmatic considerations. One was the geographic size of this large twelve-state area. This alone, according to the Census Bureau, made some subdivision desirable. In addition, the Census Bureau noted that a region of more than

MAP 1

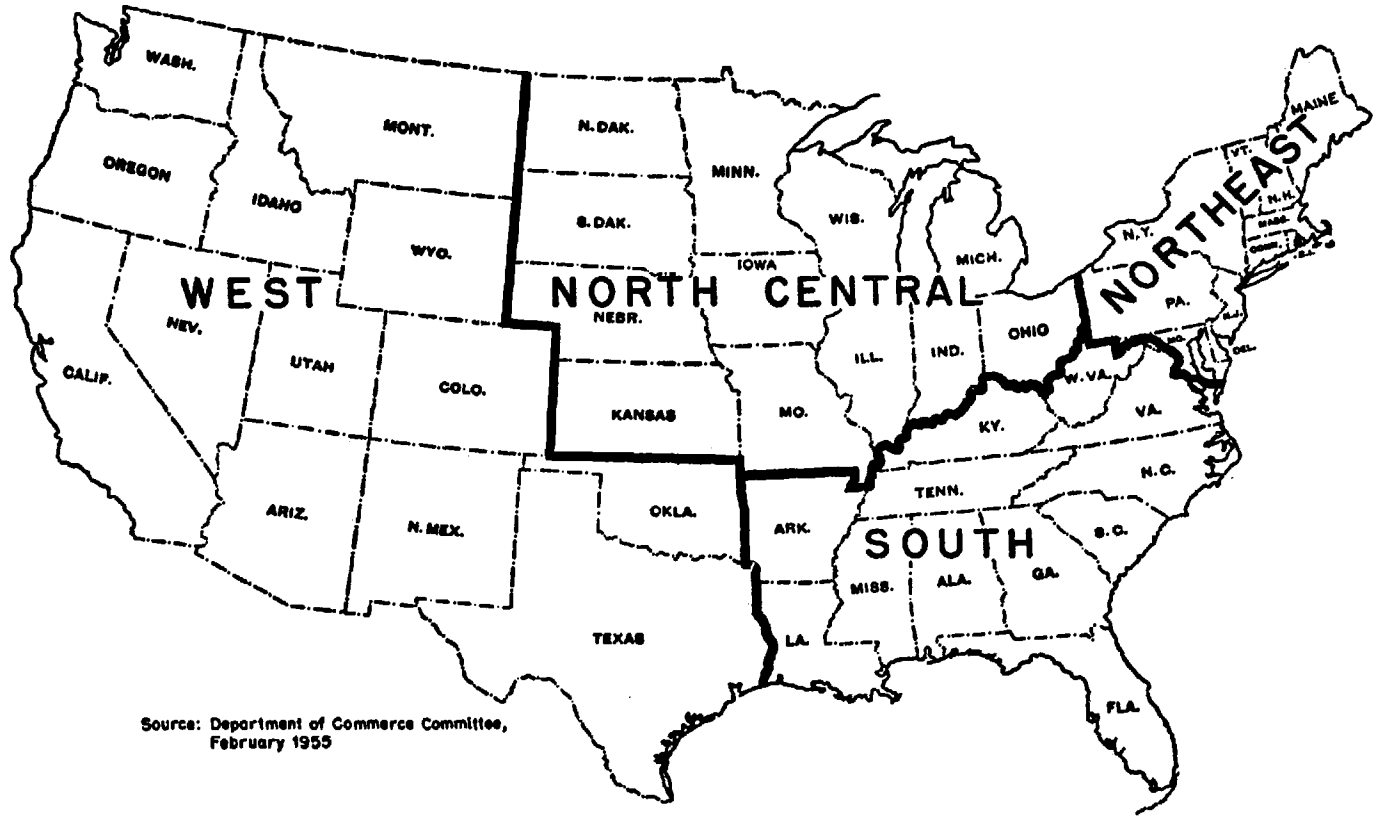
Proposed Grouping of States into Nine Geographic Divisions



Source: Department of Commerce Committee, February 1955

MAP 2

Proposed Grouping of States into Four Geographic Divisions



Source: Department of Commerce Committee, February 1955

nine states would present practical difficulties in coding operations.

On the basis of the income payments analysis, separation of West Virginia, Virginia, and Kentucky from the rest of the Southeast could be justified. The subgroup classification of Tennessee was somewhat marginal, although income factors alone would place it in the Lower South. On the basis of noneconomic factors, West Virginia, Kentucky, and Tennessee appeared to constitute a relatively homogeneous grouping, with the assignments of Virginia and North Carolina marginal as between inclusion with these states or in the Lower South.

Discussion of this problem by the Commerce Department's regional classification committee did not result in a definitive, majority opinion on either the desirability of subdividing the southeastern area or the nature of the split if one should be made. The decision to adopt the split shown on Map 1 was thus somewhat tenuous, especially with regard to the assignment of North Carolina as between the Upper and Lower South.

Recognition of the New England Region. The classification procedures outlined in the preceding sections resulted in a delineation of the Northeast which combined the New England and Middle Atlantic divisions. Such groupings, though in accord with both the social and economic factors examined, were considered impractical on several counts. In terms of economic aggregates, a single northeastern region appeared too large. The splitting off of Maine, New Hampshire, and Vermont as a unit—the course indicated by the data—would worsen matters by creating an extremely small region and at the same time reducing only slightly the size of the original group. Finally, and most important, it was felt that the culture, historical, and traditional grounds for considering the six New England states as a region should be given precedence over the statistical approach. Accordingly, the classification was modified to show New England as a regional division.

MEASUREMENT OF HOMOGENEITY

A statistical measure of homogeneity was developed for testing the results agreed upon. A group of statistical series were selected to which the test was applied.

The variance was selected as the best measure of the extent to which one classification was more homogeneous than another with respect to a particular series. In this measure, the differences were obtained between a statistic for the region and the comparable statistics for each state within the region. These differences were squared and the results added. Each individual value was weighted by the relative importance of that characteristic for the state measured by

the aggregate upon which the ratio was based. For example, the weight would be the population if the characteristic were a per capita item.

The variance would be zero for any characteristic if all states within a region had the same value. It would be relatively large for a characteristic that varied widely between states within a region, and it would be relatively small for a characteristic that did not vary widely between the states. A good set of regions by this criteria would have a low variability between states within a region, and a larger variability between regions.

This measure of regional homogeneity is affected by differences in the number of regions in the different classifications compared. For example, a classification with nine groupings should appear to be more homogeneous than one with seven. This factor was considered when the results of the test were evaluated.

To limit the amount of calculation needed, nine series were selected as quite diverse and indicators of a large number of other series. They were selected independently of the series used for making the original groupings. The items were:

1. Per capita income, 1951
2. Percentage urban, 1950
3. Percentage of income from manufacturing, 1951
4. Percentage of income from agriculture, 1951
5. Percentage Negro, Indian, and White with Spanish surname, 1950
6. Telephones per 1,000 population, 1950
7. Infant deaths per 1,000 live births, 1949
8. Net migration, 1940–1950, as percentage of 1950 population
9. Percentage change in total income, 1929–1951

The first two series are broad general measures, the third and fourth measure parts of the economy; the fifth, sixth, and seventh are social measures, and the last two series show historical trends.

The sole purpose of the test was to compare groupings, so it was necessary to calculate only the ratios of the measures for the groupings. Calculations were simplified by omitting factors in the formula that were constant for any grouping. The only element needed for the comparison was the variance factor, which was proportional to

$$\Sigma_r \Sigma_s y_{rs} (x_{rs} - x_r)^2$$

where x was the characteristic used for measurement and y was the number of elements in the series. The subscripts r and s indicate region and state.

The resulting measure of homogeneity has meaning only when it

is used for comparison with similar calculations for other groupings of states. This measure was calculated for each item separately. No over-all group index was devised, because the committee believed that comparisons on a series-by-series basis were better. The results of the computations are shown in Table 1 in terms of ratios of the values to the present census groupings.

TABLE 1

Comparison of Homogeneity Measures for Alternative Groupings of States into Divisions

(*ratio to value for present census grouping*)

	Present Census Divisions, 9 Groups	*OBE (Modified Odum) Regions, 7 Groups*	*Area Development Divisions, 9 Groups*	*Proposed Divisions, 9 Groups*
Per capita income, 1951	1.00	1.16	0.94	0.82
Percentage urban, 1950	1.00	1.29	0.88	0.89
Percentage of income from manufacturing, 1951	1.00	1.72	1.03	1.04
Percentage of income from agriculture, 1951	1.00	1.36	1.03	0.91
Percentage Negro, Indian, and White with Spanish surname, 1950	1.00	0.76	0.95	0.53
Telephones per 1,000 population, 1950	1.00	1.01	0.83	0.71
Infant deaths per 1,000 live births, 1949	1.00	0.94	1.15	0.71
Net migration, 1940–1950, as percentage of 1950 population	1.00	1.25	1.03	1.08
Percentage change in total income, 1929–1951	1.00	1.16	1.17	1.14

Source: Department of Commerce, Regional Classification Committee.

On the basis of the nine series, the proposed classification provided more homogeneous regions than any of the other three classifications now used in the Department of Commerce. The committee concluded that the new classification could be proposed by the Department of Commerce as a standard classification for the regional presentation of social and economic data.

To provide experience on a broader base, a report was prepared and transmitted to the Office of Statistical Standards of the Bureau of the Budget. Comments of other government agencies and of groups and persons outside the government are to be requested. When these comments are received, the proposed set of regions can be further evaluated and a final recommendation made, with possible government-wide application. It is hoped that if such a system

is adopted, broad regional descriptions, the availability of basic data, and other general features pertinent for economic research will be improved by the use of these standard regions. Adding these to the political divisions (the states and counties) and to the state economic area grid will make standard regions available for most medium and large area needs.

COMMENT

MARGARET JARMAN HAGOOD, Agricultural Marketing Service, Department of Agriculture

My comments on the paper by Morris Ullman and Robert Klove will be limited, since I agree with its major premises and with the general approach used in the specific problem of regional delineation now being considered by the Department of Commerce.

I believe that the authors should have given more emphasis to comparability with the past. Since a number of statistical series in many statistical agencies (including the Department of Agriculture) have been developed and issued over long periods for the nine major geographic regions, no new grouping of states should replace the old one, unless it can be definitely proven to be better. Documentation of any proposal for change must be detailed and convincing to gain acceptance.

The authors raise the problem of what to call classes of areas. There seems to be little question about the use of "political areas" for areas with recognized political boundaries. For other types of areas, the choice of a name is not simple. The proposal to call such areas "artificial" seems ill advised, since they are usually constructed to approximate more closely areas that could be considered "natural." The authors suggest an alternative: "statistical" areas. The areas may be used for other purposes; however, the Census Bureau and other agencies concerned with the mass production of statistical data use the areas most often for statistical purposes, so I consider "statistical areas" preferable to "artificial areas."

In work done earlier at the Institute for Research in Social Science of the University of North Carolina, I attempted to delineate a set of group-of-state regions using the same general approach as Ullman and Klove, but with somewhat more elaborate statistical methods.[1] Data on fifty-one population-census-derived variables and fifty-one

[1] For a brief report of this project, see Margaret Jarmon Hagood "Statistical Methods for Delineation of Regions Applied to Data on Agriculture and Population," *Social Forces*, March 1943, pp. 287–297. A typescript of the full report

agricultural-census-derived variables were combined in two ways to provide a set of regions with maximum internal homogeneity with respect to these variables. First, I used correlation and factor or component analysis to provide summarizing indexes for each state based on a linear equation involving the 102 variables. Grouping together states with similar scores on this index, I constructed regions internally alike in the average levels of the variables.

Secondly, I computed inverse correlation coefficients between pairs of adjacent or nearby states for the series of the states' standard scores on the 102 variables listed in the same order. High correlation coefficients between pairs of states indicated a similarity of patterns between the two states in ranking on the set of 102 variables. A borderline or "problem" state was then assigned to the region to which it was most similar as indicated by the values of the composite indexes and the sizes of the relevant correlation coefficients.

Using the results of these two methods for the problem of the regional allocation of Maryland and Delaware, I reached a conclusion similar to Ullman and Klove that these states should be grouped with the states to the north rather than with those to the south.

FRANK A. HANNA, Duke University

A system of general purpose regions for state data presentation can help to preserve the maximum detail, by industry and by area, in segments where avoiding the disclosure of the operation of individual concerns is a problem. Manufacturing, mining, and wholesale trade are affected most. Where disclosure is not a serious problem, state data may be used as building blocks to construct a variety of regions.

Data for a detailed manufacturing industry that cannot be published for a state are now treated in two ways. First, they are grouped with data for other industries in the same major group and an "all other industries in the same major group" category is published for the state. Secondly, they are grouped with data for same detailed industry to get a total that can be published at a regional level. In the second case, the structure of the regions has an important bearing on the quantity and utility of the data that can be published. The existing structure requires the suppression of otherwise publishable detail to prevent indirect disclosure through the subtraction of published detail from published group totals. Any proposed change in

---

containing all the basic data is on file at the Institute for Research in Social Science, University of North Carolina; the statistical methods and computation techniques used are presented in *Statistics for Sociologists*, rev. ed., Holt, 1952, pp. 523–547.

the regions should be examined to see that it would reduce the need to combine state data for important detailed industries.

Beyond an increase in the utility of industrial data, about the most that can be expected from a general purpose grouping of states is that in tables organized by regions, the states with roughly similar characteristics will appear close together. This feature is facilitated when there is no insistence on an alphabetic arrangement of the states within a region, and borderline states can be placed at the beginning or end of the listing.