EPA-600/5-74-027
October 1974

Socioeconomic Environmental Studies Series

Modal Cities

Office of Research and Development
U.S. Environmental Protection Agency
Washington, D.C. 20460

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, Environmental Protection Agency, have been grouped into five series. These five broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The five series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies

This report has been assigned to the SOCIOECONOMIC ENVIRONMENTAL STUDIES series. This series includes research on environmental management, economic analysis, ecological impacts, comprehensive planning and forecasting and analysis methodologies. Included are tools for determining varying impacts of alternative policies, analyses of environmental planning techniques at the regional, state and local levels, and approaches to measuring environmental quality perceptions, as well as analysis of ecological and economic impacts of environmental protection measures. Such topics as urban form, industrial mix, growth policies, control and organizational structure are discussed in terms of optimal environmental performance. These interdisciplinary studies and systems analyses are presented in forms varying from quantitative relational analyses to management and policy-oriented reports.

EPA REVIEW NOTICE

EPA-600/5-74-027
October 1974

MODAL CITIES

by

George B. Pidot, Jr.
John W. Sommer

Grant No. 801226
Program Element 1HA096
ROAP 21ALV, Task 01

Project Officer

Dr. Philip D. Patterson
Washington Environmental Research Center
Washington, D.C. 20460

Prepared for
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 - Price $1.25

ABSTRACT

Modal cities are representative cities based on a specific set of criteria. Using principal components analysis, 224 U.S. SMSA's were examined in terms of 48 selected variables. This analysis yielded 14 dimensions, of which 7 explained 67% of the variance.

The 224 cities were then grouped using a method that minimizes the differences among cities within a group and maximizes the differences across groups. This procedure allowed for a confident selection of 9 modalities of the U.S. metropolitan system. Each city fell into a modality and was ranked relative to its distance from the mean. The two cities closest to the mean were taken as representative of that group. One unforeseen result of this research was the distinct regional character of the different groupings.

CONTENTS

TABLES

ACKNOWLEDGMENTS

This study was conducted by the Department of Geography, Dartmouth College, Hanover, New Hampshire as a result of a grant awarded by the Washington Environmental Research Center, Office of Research and Development, EPA. The principal authors of this report are John W. Sommer and George B. Pidot, Jr.; Walter Shoup is responsible for the maps shown in Appendix A. In addition, the support of Dr. Philip D. Patterson as EPA Project Officer for this grant is acknowledged.

I. INTRODUCTION

A set of modal cities has three dimensions: space, time, and set characteristics, for example, population density, value - added by manufacturing, or retail sales. The research reported on here represents an attempt to clarify two of these dimensions, space and set characteristics. This report documents the success of the identification of set characteristics of modal cities and the steps needed to develop modal spatial frameworks Some examples of spatial framework have been created which are specific to a single model.

The modal cities project goal was to establish groups of cities, each group with a combination of characteristics which defined it as different from another group. In terms of forty-eight (48) carefully selected variables we established an example of a set of modal cities groups (9) which could be used to form data bases for a simulation model. Most important, an approach to city classification is exposed in this research which will allow one to create (within the constraints of data and computer capacity) modal city sets appropriate to any simulation model. Appropriateness is judged by the demands of any particular model to account for certain real world variables. The city groupings constitute the first, and most important part of this research.

In terms of the spatial dimension of intra-urban activity location we failed to derive much which could enhance the previously defined modal city sets. This part of the research foundered on the lack of uniform data upon which to make generalizations. Advice is offered as to when and how this data might become available.

Despite its fascinating possibilities and puzzles no systematic investigation of the temporal context of cities was undertaken The time dimension (evolution, change, development, etc.) was modestly included among the modal cities set characteristics through the employment of change variables such as "% change non-white 1950-1950". Temporality is important because therein lies the dynamics of a situation but general theoretical work on Time is not as advanced as it is on Space.

The report that follows presents a detailed explanation of the construction of modal cities, what their utility might be, and an indication on how they might be improved through the incorporation of the spatial framework on intra-urban activity locations. Obviously, a successful combination of these with a consideration of the dynamic, temporal context would yield an important study which might be useful to policy makers but would require very substantial funding.

Background to City Classification

The work undertaken for Project SUPERB is directly descended from the classical efforts of urban geographers and urban economists to derive city classification schemes.[1] Such classification schemes, which usually emphasize economic data, always beg the question, "what for?", "to what end?" and it has been the opinion of several critics that most urban scheme-makers have not answered those questions satisfactorily.[2]

Smith has reviewed the purposes and techniques of town classification, and while he is dubious about the purposes he is clear about the techniques which he has divided into three main categories based on how the threshold values for group discrimination were chosen.[3] These categories are summarized below:

1) the occupational structure of a well-defined city type was chosen and its employment figures (e.g., 30% manufacturing) were chosen as a guideline

for classifying other cities. This was used by one of the earliest and most famous of the typologists, Chauncy Harris.[4] It suffers from the obvious subjectivity of the initial city selection.

2) the calculation of an arithmetic mean with associated standard deviations, so that, for example, one might discover that the average employment in services for all cities is 25% and the standard deviation is 15% thereby allowing one to identify a discrete city with an employment structure more than 40% in services as definitely a service-oriented city Nelson's classification is representative of this kind of typology.[5] This technique is arbitrary but it is more easily replicated than is Harris'.

3) by choosing some arbitrary majority quantity of employment in a category, such as 50% or more in manufacturing, as a yardstick, is common of a number of European studies.[6] Clearly, this is a crude measure.

To these one may add a growing number of factor analytic studies.[7] The three typology techniques Smith mentions are based almost entirely upon economic data and are, therefore, unable to capture other dimensions of an urban system. Employment characteristics of cities are certainly important -- that is why the traditional typologies are built on these kinds of data, but other systemic features are also important -- city size, education levels, ethnic composition, growth performance, etc. The addition of these and other features into the creation of typologies, as one finds in those employing factor analysis, is important in giving them dynamic qualities. Several major studies such as Moser and Scott's and Yamaguchi's have used a multiplicity of variables and have reduced these to synthetic dimensions by factor analysis. Characteristically, these kinds of studies have employed many more variables than the classic urban typologies, and Alford argues that the more variables included the better since there is no rational basis for excluding any variables.[8] Berry acknowledges this reasoning and asserts that "when a research worker is confronted with a mass of data and needs to reduce these data to the most parsimonious descriptive model while gaining understanding about complex patterns of association between observations and variables, the methods discussed here (principal components analysis and grouping procedures) will be of use."[9] In fact, Berry's point is highly appropriate because it has imbedded in it the rationale for judicious selection of parameters from which a city, or system of cities will be characterized and typologized. That rationale rests with the tractibility of large scale data manipulation, and the need to select from a host of possible variables. Choice is made on *a priori* grounds to be sure, but such choice is based on one's knowledge of important elements of a city's constituent parts.

For SUPERB's Modal Cities we carefully examined each variable and debated its inclusion in our final analysis. Our selection was limited in part by computational considerations, but it is not clear that a significant addition of variables would have improved our effort.

With what success have town classifications met? Smith suggests a point of reference for measurement:

> ...to be justified on other than pedagogic grounds, any classification should be relevant to a well-defined problem or class of problems. Thus when towns are classified according to function (the differentiating characteristic), we not only want to be able to say something about the function or combination of functions typical of that

> group; knowledge of membership in any one group should automatically carry with it knowledge of membership in any one group should automatically carry with it knowledge of additional characteristics of the towns in that group.[10]

Smith adds that two justifications arise for these urban classifications: the first relates to the distributional characteristics of towns of the same class and the second to the relationship of these towns to their hinterland. In either case, through an examination of the distributional characteristics light should be shed on the underlying social and economic structure of the landscape which supports the towns.[11]

The central problem for the SUPERB classification is simply to discover a technique for creating "average cities" which are typical of the American urban system. This has been done, and the results are sound. It is not clear to us that any one Modal City typology will necessarily lead directly to greater insights into fundamental socio-economic structures of our cities but apart from the creation of modal cities data bases, it does seem that some intriguing speculation may be achieved when the maps of each class of Modal City distribution are examined. The following comments of Arnold on city classification come close to the aims of the Modal City classification effort:

> Classification serves as a framework, rather than as a developer of alternatives or a predictor for management decision-making... Classification is no more nor less than an attempt to group items (physical objects, biological characteristics, economic and social data, words, etc.) on the basis of similarities or differences as measured by data. It begins with the assembly of information in the form of data.[12]

II. AIMS

The purpose of this study is to show how to classify urban areas of the United States into a relatively small set of types based on their economic, social and demographic characteristics.

The ultimate application of any of these classifications is to define types of urban areas to use for loading a simulation model. Both for diversified gaming use and intellectual interest, it is desirable to have a relatively small set of scenarios which typify the wide variety of conditions found in different urban areas of the United States. On the one hand, massive information requirements dictate the use of a small number of areas; on the other, it is attractive to represent as broad as possible a spectrum of places. The task has been to arrive at a rational selection procedure.

The approach used in this paper is to derive modal groups and then to select actual areas which most nearly represent the range of conditions encountered in that group at a particular point in time. While the simplifications inherent in any model require some abstraction and simplification from the real world data, the use of actual areas allows a fineness of calibration and testing which entirely synthetic cities would not permit. It should be emphasized that the cities selected as Modal Cities are truly representative of their class.

Fundamentally, the test model chosen to illustrate our technique (EPA's River Basin Simulation) is designed to represent an urban region with a limited portion of supporting hinterland. The most readily available statistical construct which genuinely conforms to such a region is the Census defined SMSA (Standard Metropolitan Statistical Area). Its use of the county as a building block (outside New England) results in poor delineation of the areas for some parts of the country where counties are large and urban areas compact. However, SMSA's do represent reasonably well-defined socioeconomic functional entities, which are widely accepted for analytic purposes.

We begin with all 224 of the SMSA's defined by the Census for the time our data were collected. Three had to be deleted due to lack of data availability, but we judged their omission to have minimal effect on the succeeding analysis. We prefer not to delete any areas on _a priori_ grounds of 'distorting' the results as we feel this approach introduces and narrows the base of the resulting typology. Our selection includes roughly two-thirds of the entire United States population as of 1960.

III. VARIABLE SELECTION

Our choice of variables to describe the SMSA's for purposes of classification was guided by a combination of *a priori* reasoning applied to the needs of the chosen test model and the logic of urban structure and a pragmatic appreciation of data availability. We ought to include measures of the major demographic, labor force, housing, income and business characteristics of the SMSA's with particular detail for manufacturing because of the emphasis of the model. We utilized principal components analysis to reduce the carefully selected original set of 48 variables for 221 SMSA's to 7 indices. On the basis of these summary measures, 9 classes are delineated using a grouping algorithum. Finally, representative areas are chosen for each class. Other test models would demand different variables and derive different sets of modal cities.

More detailed variables might have been interesting, particularly for services, in understanding the internal structure of urban areas, but it is doubtful that the ultimate classification would have been altered substantially. (See Tables 1 and 2). Similarly, more contemporary data would be desirable, but we believe that our typology is sufficiently generalized to withstand developments over time. It is possible that particular areas may have sufficiently altered characteristics that they would now fall into a different class, but we feel that the broad groupings would be maintained.

Studies of almost every facet of urban life include population, size, density, and growth as the major variables describing the extent and nature of urban development. They reflect the scale economies, critical mass, proximity, stage of growth, and dynamics of the economy -- public and private. Given the SMSA as an analytic unit, the degree of urbanization adds further useful information about the extent of development in the particular area. Race and age

TABLE 1

Variables Used for SMSA Classification

Number	Variable
1	Population, 1960
2	Population per square mile, 1960
3	Population increase, 1950-1960
4	Percent urban population, 1960
5	Percent Negro population, 1960
6	Percent population aged over 65, 1960
7	Median year of education of population aged over 24, 1960
8	Percent population aged over 24 with less than 5 years of school, 1960
9	Percent population aged over 24 with high school or more, 1960
10	Percent employment in manufacturing, 1960
11	Percent white collar employment, 1960
12	Percent families with income under \$3000, 1960
13	Percent families with income \$10,000 and up, 1960
14	Percent single family housing units, 1960
15	Percent housing units sound with all plumbing, 1960
16	Percent owner occupied housing units, 1960
17	Percent population aged 5 to 34 in school, 1960
18	Income per capita, 1960
19	Unemployment rate, 1960
20	Percent employment in local government, 1962
21	Value added by manufacturing per capita 1963

TABLE 1 (Continued)

Number	Variable
22	Capital expenditures percent of value added, 1963
23	Value added increase, 1958-1963
24	Retail sales per establishment, 1963
25	Percent employment in retailing, 1963
26	Other retail sales per capita, 1963
27	General merchandise retail sales per capita, 1963
28	Retail food sales per capita, 1963
29	Retail auto sales per capita, 1963
30	Retail sales increase, 1958
31	Wholesale sales per establishment, 1963
32	Wholsesale sales per capita, 1963
33	Percent employment in wholesaling, 1963
34	Increase in wholesale sales, 1958-1963
35	Selected service receipts per establishment, 1963
36	Selected service receipts per capita, 1963
37	Percent employment in selected services, 1963
38	Increase in selected service receipts, 1958-1963
	Estimated value added per capita, 1963 in:
39	Food and tobacco products
40	Textile, apparel, and leather products
41	Paper and printing
42	Chemicals, petroleum, rubber and plastic products

TABLE 1 (Continued)

Number	Variable
43	Lumber, wood products, and furniture
44	Stone, clay, and glass products
45	Primary and intermediate metal products
46	Electrical and nonelectrical machinery
47	Transportation and ordinance
48	Instruments and miscellaneous products

TABLE 2

Means and Standard Deviations of Original Variables for 221 SMSA's

Variable	Mean	Standard Deviation	Coefficient of Variation
1	526,852	1,041,653	2.00
2	494	1,003	2.00
3	32.3	34.3	1.06
4	78.6	11.9	.15
5	10.0	10.4	1.04
6	8.5	2.2	.26
7	10.9	1.0	.09
8	7.8	5.0	.64
9	43.4	7.5	.17
10	26.9	12.4	.46
11	42.7	5.4	.12
12	18.8	7.8	.41
13	14.7	5.0	.34
14	78.1	12.7	.16
15	78.8	7.8	.10
16	63.6	7.9	.12
17	25.0	3.9	.16
18	1,857	318	.17
19	5.1	1.5	.29
20	6.7	1.3	.19
21	1,166	686	.59
22	6.3	4.3	.68
23	44.3	43.9	.99

TABLE 2 (Continued)

Variable	Mean	Standard Deviation	Coefficient of Variation
24	158,578	27,810	.18
25	14.1	2.0	.14
26	631	130	.21
27	189	44	.23
28	324	52	.16
29	276	60	.22
30	24.6	14.3	.58
31	986,391	412,860	.42
32	1,786	1,203	.67
33	4.9	1.9	.39
34	27.6	25.3	.92
35	36,513	17,289	.47
36	224	194	.87
37	5.3	3.9	.74
38	36.8	21.5	.58
39	176	123	.70
40	62	100	1.61
41	106	86	.81
42	164	163	.99
43	33	49	1.48
44	54	56	1.04
45	187	181	.97
46	205	191	.93
47	133	145	1.09
48	46	57	1.24

variables are important both in describing the political dynamics as well as potential demands on the public sector. Educational achievement reflect the general quality of the labor force and may also be related to attitudes and tastes. The broad employment variables outline the distribution among types of economic activity. Income and its distribution are both the outcome of economic activity and determinants of its future direction. Housing type and quality are important physical characteristics as well as reflecting age and affluence of the area. Value added is the most comprehensive measure of manufacturing activity, and we have attempted to estimate it at the 2 digit SIC level for SMSA's. Capital expenditure rate indicates rate of expansion for this activity. The retailing, wholesaling and service variables similarly indicate the scale, extent, development and composition of the other major private economic activities.

Our selection of variables while limited reflects, we believe, the panoply of conditions observed in urban areas. The number is already such as to make classification an almost impossible task without reducing the dimension of the problem. Further, it may be argued that the variables are not independent measures but reflect closely related aspects of the urban complex. Underlying them is an enormously complicated set of economic, political, and demographic relationships which we cannot specify explicitly. We may hope to capture one view of these interactions while reducing the dimensionality of our analysis through application of the principal components technique. This will also reduce the problem of overweighting aspects of the urban setting in our subsequent classification, which happen to be reflected in a large number of our variables.

IV. PRINCIPAL COMPONENTS TECHNIQUE

Briefly, the technique creates a smaller set of artificial measures from the original collection of variables. The new indices explain as large a portion of the original variance as possible, but are uncorrelated with each other. The principal components may be analyzed *per se* to gain insights into the urban structures we are working with as well as being used for classification purposes.

Formally, we wish to specify our variables Vij (i=1,...n SMSA's; j=1,...m variables) in terms of a set of underlying components Fik (k=1,...p components) and residuals e_i

$$Vij = \sum_{k=1}^{p} W_{jk} F_{ik} + e_i$$

where W_{jk} are the weights used in combining the components to determine the original variables. If the residual terms reflect errors of measurement and sampling, then, under the usual assumptions, they 'disappear' from the covariance matrix. We assume that the components account for all the variance of the variables, and we are trying to attribute a portion of the variance to each of the components. If there are unique elements of variance in some of the original variables or we omit some of the components then the error terms do not vanish. Ideally we should know <u>a priori</u> these specific variances, or alternatively the communalities, and perform our analysis only on the latter. Here we assume that all of the variance is to be analyzed. Since we standardize our original variables, i.e., they have zero mean and unit variance, we are in effect examining the correlation matrix with unity on the diagonal. It should be noted that this standardization affects the resulting analysis in a complex fashion. The resulting weights can not be readily converted into those which would arise from nonstandardized data.

In principal components, we know the resulting variables and wish to estimate both the underlying components F_{ik} and the weights W_{jk}. This introduces a degree of indeterminacy in the results which we eliminate by constraining the components to have zero mean and unit variance. We wish to choose a set of coefficients a_{jk} for

$$F_{ik} = \sum_{j=1}^{m} a_{jk} V_{ij}$$

which minimize the residual variance, i.e., the sum of squared residuals between the original variables and their estimates based on the first component. This is equivalent to explaining as much of the original variance as possible with the first component. Having done so, we might then eliminate the effects of the first component from the original variables and estimate a second component such that it explained as much of the remaining variance as possible. This interactive procedure can be followed up to a limit of m components. We hope to find a set of r components, $r \ll m$, which will account for most of the observed variance (see Table 3).

It turns out that our problem is equivalent to finding the successive roots of the correlation matrix by solving its characteristic equation $|R - \lambda I| = \emptyset$. The solution for the largest root corresponds to that set of weights explaining the greater portion of the variance. The eigen-value is the portion of the variance explained and the accompanying eigenvector contains the weights. Successively smaller roots and their vectors correspond to subsequent components. It can also be shown that these vectors are orthogonal, or uncorrelated with each other.

We may also view the analysis in geometric terms as a rotation of the axes on which the variables are measured. The weights are in fact the direction casines used to transform the variable into the components of the new metric.

In interpreting the components we examine the correlation of each with the original set of variates (see Table 4). We also compute the component scores (F_{ik}) for each SMSA.

We attempt to verify our understanding of the components by examining areas which rank very high or low in the metric of the new variables (see Table 5). Recall that the component variables were standardized with zero mean and unit variance so we may view an area's score directly in terms of a distribution.

TABLE 3

Proportion of Total Variance
Accounted for by Principal Components

Principal Component	Eigenvalue	Percent of Pooled Variance	Cumulative Percent
1	10.52	21.9	21.9
2	7.49	15.6	37.5
3	4.11	8.6	46.1
4	3.26	6.8	52.9
5	2.67	5.6	58.5
6	2.12	4.4	62.9
7	1.73	3.6	66.5

The above refers to an analysis of 48 variables for 221 SMSA's.

TABLE 4

Zero Order Correlation Coefficients Between Principal Components and Original Variables for 221 SMSA's

Variable	Principal Components 1	2	3	4	5	6	7
1	0.28	0.25	-0.50	-0.13	0.11	-0.20	-0.17
2	0.05	0.40	-0.50	-0.35	-0.09	-0.23	0.06
3	0.56	-0.30	0.35	-0.16	-0.06	-0.24	0.05
4	0.45	0.20	-0.45	-0.00	-0.10	-0.29	-0.07
5	-0.19	-0.39	-0.14	-0.19	0.57	-0.31	0.18
6	-0.12	0.37	-0.15	-0.01	-0.36	0.62	-0.07
7	0.77	-0.02	0.13	0.45	-0.12	-0.12	0.12
8	-0.49	-0.48	-0.17	-0.38	0.24	-0.23	0.06
9	0.78	-0.09	0.14	0.42	-0.15	-0.12	0.11
10	-0.35	0.83	0.10	-0.16	0.01	0.08	0.03
11	0.70	-0.19	-0.24	0.39	-0.04	-0.18	0.14
12	-0.51	-0.72	-0.05	-0.18	0.23	0.01	0.07
13	0.72	0.48	-0.00	0.01	-0.01	-0.29	0.03
14	-0.16	-0.50	0.59	0.33	0.26	0.05	-0.13
15	0.68	0.49	0.06	0.10	-0.26	-0.08	-0.04
16	-0.05	0.15	0.60	0.50	-0.02	0.20	-0.21
17	0.03	0.01	0.18	0.14	-0.08	-0.36	0.19
18	0.71	0.56	-0.03	0.04	-0.11	-0.07	-0.05
19	-0.32	-0.16	0.08	-0.20	-0.32	-0.11	-0.53
20	0.27	-0.48	0.10	-0.09	-0.06	-0.39	-0.19
21	-0.16	0.86	0.28	-0.08	0.33	-0.06	-0.02
22	-0.10	-0.22	0.16	0.03	0.14	-0.02	-0.36
23	0.19	-0.12	0.40	-0.31	0.13	0.02	0.26

TABLE 4 (Continued)

Variable	Principal Components 1	2	3	4	5	6	7
24	0.82	0.04	0.05	0.16	0.23	-0.10	0.09
25	0.63	-0.48	0.15	-0.02	-0.03	0.17	-0.03
26	0.77	-0.01	0.12	-0.23	-0.12	0.28	-0.03
27	0.65	0.01	-0.03	0.13	0.27	0.31	-0.01
28	0.67	0.31	0.16	-0.35	-0.11	0.18	-0.12
29	0.71	-0.25	0.33	0.07	0.11	0.15	-0.03
30	0.51	-0.25	0.46	-0.45	0.05	0.02	0.29
31	0.39	0.17	-0.55	0.23	0.46	0.07	-0.04
32	0.41	0.02	-0.57	0.30	0.42	0.20	-0.04
33	0.36	-0.21	-0.52	0.33	0.41	0.22	0.01
34	0.55	-0.12	-0.17	-0.34	0.02	0.02	0.25
35	0.73	-0.04	-0.21	-0.42	0.21	0.01	-0.28
36	0.69	-0.11	-0.02	-0.47	0.08	0.15	-0.37
37	0.64	-0.28	0.03	-0.46	0.09	0.11	-0.36
38	0.38	-0.08	0.30	-0.39	0.12	0.02	0.38
39	-0.20	0.28	0.24	0.08	0.50	0.27	0.03
40	-0.27	0.35	-0.32	-0.43	-0.12	0.33	0.23
41	-0.02	0.61	0.08	-0.03	0.33	0.10	0.08
42	-0.13	0.43	0.04	-0.10	0.38	-0.30	-0.19
43	-0.14	0.05	0.15	-0.04	0.29	0.24	0.17
44	-0.18	0.25	0.39	-0.08	0.40	-0.09	-0.32
45	-0.15	0.73	0.30	0.02	0.12	-0.21	-0.17
46	0.07	0.77	0.24	-0.00	0.09	0.02	0.13
47	-0.03	0.46	0.38	0.08	0.15	-0.22	-0.09
48	0.04	0.65	-0.11	-0.19	-0.12	-0.14	0.17

$r \geq .18$ is significant at 1 percent with 200 degrees of freedom

TABLE 5

SMSA's With Extreme Principal Component Scores

1		2		3		4	
Anaheim-Santa Ana-Garden Grove, Cal.	3.07	Jersey City, N.J.	2.49	Anaheim-Santa Ana-Garden Grove, Cal.	3.32		
Las Vegas, Nev.	5.51	Kenosha, Wisc.	2.57	Anderson, Ind.	2.42		
Reno, Nev.	3.58	New Britain, Conn.	2.07	Ann Arbor, Mich.	2.63		
San Jose, Cal.	2.24	Waterbury, Conn.	2.20	Flint, Mich.	2.20		
Santa Barbara, Cal.	2.02			Kenosha, Wisc.	2.34		
Stamford, Conn.	2.11			Las Vegas, Nev.	2.14		
Washington, D.C.	2.37						
Brownsville-Harlingen-San Benito, Tex.	-2.31	Brownsville-Harlingen-San Benito, Tex.	-2.11	Boston, Mass.	-2.00	Anaheim-Santa Ana-Garden Grove, Cal.	-2.36
Gadsden, Al.	-2.07	Fayetteville,N.C.	-2.24	Chicago, Ill.	-2.47	Atlantic City, N.J.	-2.18
Johnston, Pa.	-2.10	Laredo, Tex.	-2.81	Jersey City, N.J.	-4.08	Fall River, Mass.	-2.25
				New York, N.Y.	-5.05	Huntsville, Ala.	-2.78
						Jersey City, N.J.	-3.66
						Las Vegas, Nev.	-6.44
						New Bedford, Mass.	-2.13
						New York, New York	-2.64
						Reno, Nev.	-3.12

TABLE 5 (Continued)

5		6		7	
Atlanta, Ga.	2.40	Las Vegas, Nev.	2.06	Anaheim-Santa Ana-Garden Grove, Cal.	3.11
Charleston, W. Va.	2.70	St. Joseph, Mo.	2.47	Huntsville, Ala.	3.13
Charlotte, N.C.	3.00				
Durham, N.C.	2.36				
Memphis, Tenn.	2.65				
Richmond, Va.	2.06				
Winston-Salem, N.C.	3.24				
Colorado Springs, Col.	-2.05	Ann Arbor, Mich.	-2.21	Huntington-Ashland, W. Va.	-2.73
Meriden, Conn.	-2.05	Beaumont-Pt. Arthur, Tex.	-2.05	Las Vegas, Nev.	-4.98
		Galveston-Texas C., Tex.	-2.39	Reno, Nev.	-3.78
		Jersey City, N.J.	-2.00	Steubenville-Weirton, Ohio	-3.15
		Lake Charles, La.	-2.21	Wheeling, W. Va.	-2.22
		Midland, Tex.	-2.11		
		Provo-Orem, Utah	-2.06		
		Waterbury, Conn.	-2.52		

V. RESULTS OF PRINCIPAL COMPONENTS ANALYSIS

Our analysis retains seven components for examination, based on their explanation of the pooled variance. They account for two-thirds of the original variation. A substantial portion of the variance is included from each of the original variables although less so for school enrollment, capital expenditure in manufacturing, value added, growth and lumber and wood products.

It should be noted that the arithmetic signs on components are not unique, i.e., multiplying all the coefficients for a component by a minus one does not affect the statistical properties. Thus for interpretation one may think of component with many large negative weights in terms of its inverse. Since the component scores are standardized about a zero mean, one might view an area with a large negative score as ranking high on those variables with large negative weights.

Component I

The first component is linked with high levels of income and growth, and their associated phenomena. The growth includes retailing, wholesaling and services as well as population. The labor force is highly educated and concentrated in white collar jobs. Housing quality is high. All measures of retailing and selected services are strong. Wholesaling is important to a somewhat lesser degree.

Component II

The second component reflects a dominance of manufacturing in employment and value added. The linkage is strong with paper and printing, metals, machinery, instruments and miscellaneous and less so with chemicals, petroleum, rubber and plastics, and transportation and ordinance. The people are moderately well-to-do with a notable absence of poor and those with poor education. They live in generally good quality multifamily dwellings.

Component III

The third component is the antithesis of metropolitanism. It is negatively linked to size, density and urbanization. Growth is fairly important, of retailing and manufacturing as well as of population. People live in their own, single-family homes. Wholesaling is notably absent, being a function of the heavily urbanized area. There is some concentration of stone, clay, and glass industry, also of metals, transportation - ordinance and an absence of textile and apparel manufacturing.

Component IV

The fourth component is negatively associated with all measures of services and the textile and apparel industry. On the other hand, it is linked to high educational attainment. Economic growth is poor -- for manufacturing, retailing, services and wholesaling. People tend to live in owner-occupied, single-family homes.

Component V

The fifth component stresses the presence of a black population and the absence of older people. There is emphasis on various measures of wholesaling. Manufacturing is growing and is important in food and tobacco, less so for chemical-petroleum-rubber and plastics, stone-clay and glass, and lumber and wood products.

Component VI

The sixth component is strongly representative of the aged population and consequent lower school enrollments. There is also an absence of local government employment.

Component VII

The seventh component is clearly linked to full employment, and modestly so to growth in retailing, services and wholesaling, although the levels of service employment and receipts are low. It is, however, related to a low level of manufacturing capital outlays and an absence of the stone-clay-glass industry.

VI. GROUPING PROCEDURE

Having derived a set of measures describing the multiplicity of conditions existing in urban areas, we must now categorize the SMSA's on these bases into a workable set of classes. Given our goal of a small group of representative types, we want to create these classes in such a way that the members of a class are as nearly like each other as possible.

Formally we want to define a small number of groups (g) such that the intragroup variance of the principal component measures F_{ik} is minimized.

$$\min V = \sum_{g=1}^{t} \left(\sum_{k=1}^{p} \left(\sum_{i=1}^{ng} (F_{ik} - \bar{F}_{gk})^2 \right) \right)$$

where ng is the number of members in group g and $\bar{F}_{gk}$ is the mean of factor k in group g (see Table 6). Equivalently, the intragroup differences are maximized. Optimal solutions to such grouping problems with more than trivial dimensions are intractible from a practical viewpoint. The solution we choose here is Ward's grouping algorithm, which builds up groups in a nonrecursive stepwise procedure that minimizes the increased error at each stage.[13]

TABLE 6

Mean Component Value by Type of SMSA

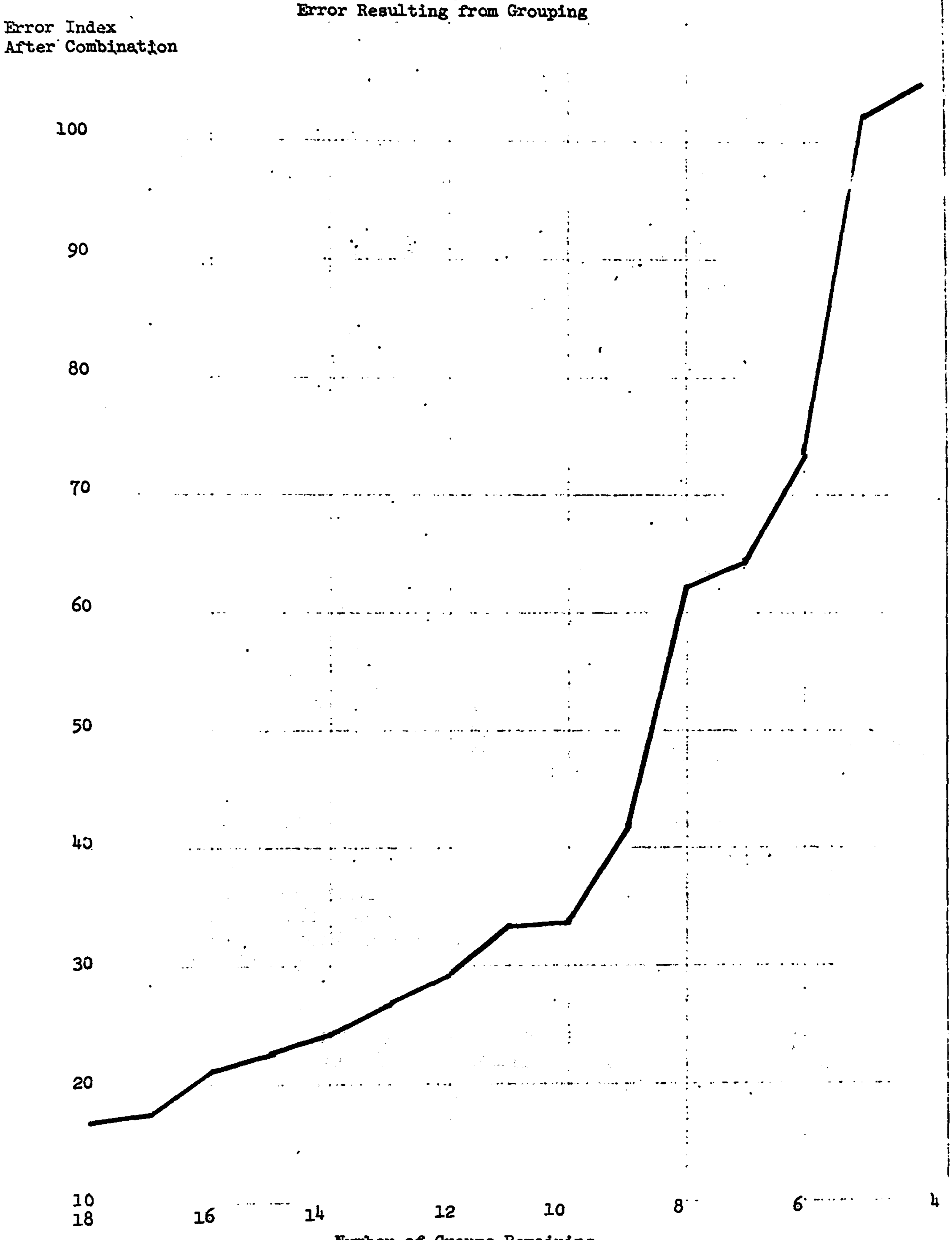
SMSA Type	Number of SMSA's in Type	Mean Value of Component #1	#2	#3	#4	#5	#6	#7
A	20	0.72	1.07	-1.45	-0.45	0.06	-0.77	-0.23
B	2	4.54	-1.35	1.32	-4.78	0.41	1.72	-4.38
C	20	0.61	-0.02	-0.92	0.92	1.14	0.67	-0.05
D	12	1.28	-0.70	1.52	-0.93	-0.25	0.05	1.00
E	33	-0.63	-0.24	0.20	-0.19	0.40	1.06	0.45
F	22	-0.69	0.68	-0.54	-0.85	-1.28	0.63	0.12
G	30	-0.84	-1.09	-0.24	-0.25	0.53	-0.96	-0.14
H	43	-0.22	0.96	0.92	0.37	0.10	-0.19	-0.24
I	39	0.51	-0.67	-0.02	0.73	-0.72	-0.31	0.02

This approach begins with each observation placed in a separate group. That pair of groups is combined which will cause the smallest increase in the error function. This function is simply the pooled intragroup variance for the measures we are using. At each subsequent step, the potential error resulting from any further combination of the remaining groups is computed and then a new error minimizing combination is selected. The procedure does not backtrack nor select groups simultaneously so that it does not result in a true optimum combination. However, if the associations among types of items being grouped are fairly strong, the resulting groupings are likely to be near optimal in terms of the error variance.

There is no statistical test to determine how many classes should be defined. The selection is based on the rough number of types one wishes to have. However, examination of the error function does indicate the cost of a particular choice; the increased cost due to a further reduction in the number of classes helps to delineate the appropriate stopping point. Because the grouping algorthm gives equal weight to the measures used as a basis for selection, one needs to consider the number of indices related to particular facets of the items and their variance. By definition, our principal components are orthogonal and maximally efficient in describing the underlying variables. Further, they are standardized to zero mean and unit variance, so no further manipulation is necessary.

Ward's algorithm was applied to our seven principal component measures for 221 SMSA's. The accompanying table indicates the behavior of the error function over the range of classes we were concerned with. However, it might be noted that increases in the error function were very small over the entire range up to this point. The very large jump in the cumulative error reducing the number of classes from nine to eight (approximately a fifty percent increase) led us to select that as the desirable level of aggregation (see Table 7). Subsequent examination of class membership confirmed the feeling that further

Table 7
Error Resulting from Grouping



combinations would submerge distinctive types. It might also be added that the class which contains only Las Vegas and Reno remains distinct with further combination until the last step.

The complete listing of SMSA's by type is given in Table 8. The geographic clustering in the resulting typology is clear, although there is no bias in our procedures to produce it. Aside from the obvious, Type B being solely Nevada, D is California and Florida, E is South and Central U.S., F is Northeast, especially New England, G is Deep South, H is Midwest and I is South Central U.S.

Type A clearly consists of very large, highly-developed urban areas across the country with important manufacturing sectors. B is highly specialized in recreation, with rapid growth and high income. Category C contains the medium size areas with a relatively smaller service sector, emphasizing distribution and some manufacturing. Class D areas are affluent and growing, but less highly urbanized. Class E represents less well-to-do areas with elderly populations. F types are traditional New England with relative stagnation, lack of wholesaling and an absence of Blacks. G areas are nonmanufacturing with rather high levels of poverty and many Blacks. The H class areas are archetypal Midwestern, stressing manufacturing, somewhat smaller but growing. Finally, the I group are reasonably affluent, medium-size regional centers, individually specializing in a variety of functions.

VII. MODAL CITIES SELECTION

The final stage of our analysis was to rank the areas within their types and select representative SMSA's for each class. This was done on the basis of the sum of square deviations of each SMSA from its class means for the seven principal components.

$$D_i = \sum_{k=1}^{7} (F_{ik} - \overline{F}_{gk})^2$$

TABLE 8

Ranking of SMSA's by Type

Census Number	Name	Deviation Score*
	Type A	
136	Newark, N.J.	0.42
149	Philadelphia, Pa.	0.94
122	Milwaukee, Wisc.	0.97
41	Cleveland, Ohio	1.37
27	Boston, Mass.	1.65
40	Cincinnati, Ohio	1.83
177	San Francisco-Oakland, Calif.	1.86
110	Los Angeles-Long Beach, Calif.	1.92
18	Baltimore, Md.	2.18
31	Buffalo, N.Y.	2.19
170	St. Louis, Mo.	2.22
39	Chicago, Ill.	3.08
146	Paterson-Clifton-Passaic, N.J.	3.08
53	Detroit, Mich.	3.65
152	Pittsburgh, Pa.	4.29
132	New Haven, Conn.	4.95
211	Washington, D.C.	7.18
192	Stamford, Conn.	8.35
135	New York, N.Y.	19.59
90	Jersey City, N.J.	26.02

* Sum of squared deviations from type means for seven grouping components.

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type B	
101	Las Vegas, Nev.	5.52
162	Reno, Nev.	5.52
	Type C	
93	Kansas City, Mo.	0.42
143	Omaha, Neb.	0.53
47	Dallas, Tex.	0.69
163	Richmond, Va.	0.99
52	Des Moines, Iowa	1.18
86	Indianapolis, Ind.	1.22
155	Portland, Ore.	1.37
123	Minneapolis-St. Paul, Minn.	1.59
89	Jacksonville, Fla.	1.79
185	Sioux Falls, S. Dak.	2.11
219	Wilmington, Del.	2.54
83	Houston, Tex.	2.89
154	Portland, Me.	2.97
184	Sioux City, Iowa	2.98
13	Atlanta, Ga.	3.14
62	Fargo-Moorhead, N. Dak.	3.45
111	Louisville, Ky.	3.75
37	Charlotte, N.C.	5.25
118	Memphis, Tenn.	5.38
169	St. Joseph, Mo.	5.53

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type D	
145	Oxnard-Ventura, Calif.	1.13
150	Phoenix, Ariz.	1.37
144	Orlando, Fla.	1.74
179	Santa Barbara, Calif.	2.09
214	W. Palm Beach, Fla.	2.23
167	Sacramento, Calif.	2.83
66	Ft. Lauderdale-Hollywood, Fla.	3.76
175	San Bernadino-Riverside-Ontario, Calif.	3.76
178	San Jose, Calif	4.00
59	Eugene, Ore.	5.18
85	Huntsville, Ala.	11.10
9	Anaheim-Santa Ana-Garden Grove, Calif.	14.66
	Type E	
95	Knoxville, Tenn.	0.91
12	Ashville, N.C.	1.29
164	Roanoke, Va.	1.32
207	Tyler, Tex.	1.37
129	Nashville, Tenn.	1.64
210	Waco, Tex.	1.70
189	Springfield, Mo.	1.72
38	Chatanooga, Tenn.	1.73
60	Evansville, Ind.	1.84
54	Dubuque, Iowa	1.88
80	Harrisburg, Pa.	2.02
200	Texarkana, Tex.	2.04

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type E (Continued)	
223	York, Pa.	2.17
105	Lexington, Ky.	2.34
7	Altoona, Pa.	2.36
67	Ft. Smith, Ark.	2.37
108	Little Rock, Ark.	2.45
114	Lynchburg, Va.	2.61
78	Greensville, S.C.	2.67
98	Lancaster, Pa.	2.97
6	Allentown-Bethlehem-Easton,Pa.	3.17
188	Springfield, Ill.	3.34
25	Bloomington-Normal, Ind.	3.50
161	Reading, Pa.	3.54
171	Salem, Ore.	3.62
77	Greensboro-High Point, N.C.	3.82
198	Tampa-St. Petersburg, Fla.	3.99
160	Raleigh, N.C.	4.25
199	Terre Haute, Ind.	4.46
15	Augusta, Ga.	5.13
56	Durham, N.C.	5.34
151	Pine Bluffs, Ark.	8.42
221	Winston-Salem, N.C.	13.82

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type F	
208	Utica-Rome, N.Y.	0.68
222	Worcester, Mass.	0.93
112	Lowell, Mass.	1.11
191	Springfield-Chicopee-Holyoke, Mass.	1.24
4	Albany-Schenectady-Troy, N.Y.	1.38
23	Binghamton, N.Y.	1.44
156	Pawtucket-Providence-Warwick, R.I.	1.55
104	Lewiston-Auburn, Me.	1.70
29	Brockton, Mass.	1.99
181	Scranton, Pa.	2.11
218	Wilkes Barre-Hazelton, Pa.	2.81
64	Fitchburg-Leominster, Mass.	2.98
119	Meriden, Conn.	3.07
117	Manchester, N.H.	3.08
130	New Bedford, Mass.	3.40
102	Lawrence-Haverhill, Mass.	4.44
28	Bridgeport, Conn.	4.93
61	Fall River, Mass.	5.39
131	New Britain, Conn.	5.77
215	Wheeling, W. Va.	8.02
91	Johnston, Pa.	8.27
14	Atlantic City, N.J.	8.32

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type G	
124	Mobile, Ala.	0.19
180	Savannah, Ga.	1.13
43	Columbia, S.C.	1.23
125	Monroe, La.	1.29
88	Jackson, Miss.	1.29
115	Macon, Ga.	1.43
44	Columbus, Ga.	1.43
174	San Antonio, Tex.	1.43
138	Norfolk-Portsmouth, Va.	1.45
206	Tuscaloosa, Ala.	1.49
96	Lafayette, La.	1.66
35	Charleston, S.C.	1.92
147	Pensacola, Fla.	2.08
126	Montgomery, Ala.	2.29
57	El Paso, Tex.	2.34
24	Birmingham, Ala.	2.56
183	Shreveport, La.	2.64
3	Albany, Ga.	2.69
220	Wilmington, N.C.	3.15
134	New Orleans, La.	3.59
63	Fayetteville, N.C.	4.23
30	Brownsville-Harlingen-San Benito, Tex.	4.78
19	Baton Rouge, La.	4.91
97	Lake Charles, La.	5.38

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type G (Continued)	
71	Gadsden, Ala.	5.46
21	Beaumont-Port Arthur, Tex.	6.74
100	Laredo, Tex.	8.22
72	Galveston-Texas City, Tex.	8.24
141	Ogden, Utah	9.28
84	Huntington-Ashland, W. Va.	9.69
	Type H	
79	Hamilton - Middletown, Ohio	0.58
168	Saginaw, Mich.	0.61
92	Kalamazoo, Mich.	0.62
127	Muncie, Ind.	0.66
166	Rockford, Ill.	0.74
49	Dayton, Ohio	0.79
213	Waterloo, Iowa	0.85
74	Grand Rapids, Mich.	0.87
32	Canton, Ohio	0.93
201	Toledo, Ohio	0.99
128	Muskegon, Mich.	1.04
2	Akron, Ohio	1.07
106	Lima, Ohio	1.25
203	Trenton, N.J.	1.43
186	South Bend, Ind.	1.51
159	Racine, Wisc.	1.58

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type H (Continued)	
99	Lansing, Mich.	1.59
224	Youngstown-Warren, Ohio	1.65
195	Syracuse, N.Y.	1.73
165	Rochester, N.Y.	1.74
87	Jackson, Mich.	1.86
50	Decatur, Ill.	2.04
48	Davenport-Rock Island-Moline, Ill.	2.10
190	Springfield, Ohio	2.29
68	Fort Wayne, Ind.	2.33
20	Bay City, Mich.	2.46
58	Erie, Pa.	2.51
148	Peoria, Ill.	2.68
158	Pueblo, Colo.	3.06
81	Hartford, Conn.	3.58
33	Cedar Rapids, Iowa	3.67
65	Flint, Mich.	3.89
10	Anderson, Ind.	4.34
109	Lorain-Elyria, Ohio	4.43
76	Green Bay, Wisc.	5.08
73	Gary-Hammond-E. Chicago, Ind.	6.12
157	Provo-Orem, Utah	6.86
153	Pittsfield, Mass.	7.14
94	Kenosha, Wisc.	7.20
36	Charleston, W. Va.	10.62
11	Ann Arbor, Mich.	11.18

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type H (Continued)	
212	Waterbury, Conn.	13.83
193	Steubenville-Weirton, Ohio	15.01
	Type I	
205	Tulsa, Okla.	0.57
196	Tacoma, Wash.	0.62
75	Great Falls, Mont.	0.69
202	Topeka, Kans.	0.91
217	Wichita Falls, Tex.	0.98
1	Abilene, Tex.	1.21
69	Fort Worth, Tex.	1.23
141	Ogden, Utah	1.23
5	Albuquerque, N. Mex.	1.49
107	Lincoln, Nebr.	1.54
173	San Angelo, Tex.	1.62
204	Tucson, Ariz.	1.73
172	Salt Lake City, Utah	1.76
16	Austin, Tex.	1.86
8	Amarillo, Tex.	1.89
116	Madison, Wisc.	1.91
45	Columbus, Ohio	1.94
176	San Diego, Calif.	2.03
142	Oklahoma City	2.10
216	Wichita, Kan.	2.14
26	Boise City, Idaho	2.21

Table 8 (Continued)

Census Number	Name	Deviation Score*
	Type I (Continued)	
113	Lubbock, Texas	2.29
34	Champaign-Urbana, Ill.	2.33
209	Vallejo-Napa, Calif.	2.37
51	Denver, Calif.	2.44
22	Billings, Mont.	2.80
182	Seattle-Everett, Wash.	2.85
187	Spokane, Wash.	2.94
42	Colorado Springs, Colo.	3.04
103	Lawton, Okla.	3.09
17	Bakersfield, Calif.	3.16
194	Stockton, Calif.	3.26
70	Fresno, Calif.	3.27
82	Honolulu, Hawaii	4.59
120	Miami, Fla.	4.71
121	Midland, Tex.	4.88
55	Duluth-Superior, Minn.	5.38
137	Newport News-Hampton, Va.	5.80
197	Tallahassee, Fla.	6.31

where F_{gk} is the mean value of component k for group g.

An area with zero deviation would have precisely the mean characteristics for its type. In examining the deviations for specific types or areas, it should be remembered that the components from which the deviations are computed were standardized with zero mean and unit variance.

While the choice of representative cities for each modal group is determined by the grouping algorithm in an absolute sense it is worthwhile considering some other factors not included in the statistical analysis which can lead one to alternative selections. As shown in Table 8, Newark, New Jersey, Las Vegas or Reno, Nevada, Kansas City, Missouri, Oxnard-Ventura, California, Knoxville, Tennessee, Utica-Rome, New York, Mobile, Alabama, Hamilton-Middletown, Ohio, and Tulsa, Oklahoma are the least deviant from the mean characteristics of their respective groups in a statistical sense, but there are some spatial considerations that temper the actual choice of the "typical" city of several of the classes.

The chief consideration that arises is that of "independence" of the city as a unit. Notwithstanding the obvious fact that the whole urban system is intensely interrelated, particularly within the megalopolitan concentrations, it does appear that Newark (Type A) and Oxnard-Ventura (Type D) are heavily influenced by their relationships to New York City and Los Angeles respectively. Therefore, we must submit that Philadelphia and Phoenix are "more typical" representatives of their categories: both are spatially separated units next on the list of deviancy from their class means. For the purposes of the SUPERB project these are also good substitutions from the point of view of water-related issues.

Similarly, the substitution of Lowell, Massachusetts (Type F) for Utica-Rome is attractive because of spatial discretness and classic New England manufacturing city water pollution problems. Worcester, Massachusetts, ranking directly behind Utica-Rome and above Lowell, would have been our choice if Lowell's position were not in the Merrimack River Basin, where pollution issues are nearly two centuries old.

Whether one chooses Las Vegas or Reno (Type B) is a literal toss-up, and while there is no question about this being a distinct class, its spareness of representative cities, and its lack of clear-cut pollution issues makes it a candidate for exclusion.

Type H, the small northern manufacturing centers, present a luxuriant set of choices for a data base. The first ten cities in this modal group are less than a standard deviation from the mean, and less than half a standard deviation separates them. For that matter, the next ten are barely more than a standard deviation away from Hamilton-Middletown, the leader. On inspection it seemed to us that Saginaw, Michigan or Rockford, Illinois might be the best choices on the basis of "independence" and water quality kinds of questions. The point is that convenience for the user of the typology should play a role in the choice here (as of course it should for each modal type).

For the other modalities (Type E, Kansas City, Type F, Knoxville, Type G, Mobile, and Type I, Tulsa) there appeared to us to be no compelling reason to seek alternative representatives.

Our summary suggestion for a list of modal cities is shown in Table 9. We have chosen the cities in pairs by class, listing our primary selection first.

In all of the cases where we have suggested alternatives we believe that the suggestions are in accord with the classification principal enunciated by Smith and quoted on page 4 of this report.

VIII. CONCLUSIONS

As we have proceeded with this classification effort we have become aware of the richness and vitality of the existing literature and current research and we are pleased to discover that others have found the kind of effort pursued in this study to be rewarding.[14] We are also happy to see that our effort is unique in the sense of employing data from, and ultimately classifying, virtually all of the SMSA's in the United States using

TABLE 9

Modal Cities Suggestions

Type A

Philadelphia
Cleveland

Type B

Las Vegas
Reno

Type C

Kansas City
Dallas

Type D

Phoenix
Orlando

Type E

Knoxville
Ashville

Type F

Lowell
Worcester

Type G

Mobile
Savannah

Type H

Saginaw
Rockford

Type I

Tulsa
Tacoma

a large number of carefully selected variables. Other studies have used more variables on fewer urban areas, but none, to our knowledge, have spanned the whole United States urban system in the same way as is presented in this report. Furthermore, our research has been set up in such a way that this study may be replicated for any period when new (or old) data is available. There are also possibilities for extending empirical research of this kind in accord with the sound assertions of Johnston concerning theory-building from regionalization techniques.[15] He points out that it is only after a classification procedure has been undertaken that the question of spatial contiguity should be considered and hypotheses formed and tested.[16] While the purpose of SUPERB has been entirely in the realm of empirical methodology -- devising a modal city typology -- there may be some attractive realms of theorizing which result from our analysis. Some suggestive commentary is included below.

Remarks on Future Research

The most intriguing area for further research appears to us to be in the reasonably well-defined regional groupings which have "fallen out" of our analysis. We have already emphasized that we proceeded with no intention to select variables which could specifically generate clustered patterns in space. Following the grouping and mapping parts of our research we rechecked the variables and concluded that there were no specific regional variables included, yet there are obvious clusters in the resultant patterns shown on the maps in Appendix A.

Type A, with its major metropolitan character is, of course, spread around the United States, but there is a firm concentration in the northeastern megalopolitan corridor.

Type B, is outstandingly concentrated in a spatial sense but what one can make of this is not obvious other than to note that these are two highly specialized recreational cities in a state with unusual laws.

Type C, cities are focused on middle America with a few on the southern piedmont. These are manufacturing and distribution centers smaller than those of Type A.

Type D, are closely related to the amenity environment. The socio-economic composition of these cities, which seem to reflect climate orientation, would be a good point of departure for examining a whole class of cities.

Type E, have a strong regional grouping in the mid-South and Border states. Some of this may be due to industry age and type, but the reasons for this grouping are not yet clear.

Type F, is a distinct spatial grouping and it is clear that the New England manufacturing city is more than just an image. Curiously, the pattern of this city-type is even tighter if one eliminates the three most deviant cases from the bottom of the list in Table 8.

Type G, southern cities, are almost as distinct as the preceding grouping. Again, if one drops the five most deviant cases from the bottom of the list there is much greater spatial packing in this group.

Type H, with its low deviancy in a statistical sense, is also localized in the north-central section of the U.S. Several of the most deviant of this class are also the most removed in real distance.

Type I, with its apparent relationship to extractive industry and military installations is lacking in regional clustering. If these two reasons are indeed heavily influential in the classification, then the derived pattern is fully plausible.

It is obvious that there are regional factors at work here, and this is interesting. If we do seek in a typology the ability to suggest something about a class of individuals it is useful to be able to point to a clustering in spatial terms as well as in aspatial, statistical terms. We believe that the analysis of these clusters could be a basis for further research.

In effect, our examination has yielded the potential for hypothesis formulation concerning the macro-scale spatial context. Why are certain kinds of cities located where they are? What do their common attributes and regional clustering mean in terms of environmental impact? What would monitoring of their migration (in a statistical sense) from one group to another over time indicate? Do these groups of cities have modal spatial frameworks, (i.e. arrangements of land use), to which the individual cities tend?

It was to this last question that the second phase of our research was directed. Data limitations defied our attempts to discover if modal arrangements do exist, but we were able to make practical suggestions concerning how one could generate data for such an analysis. We have also demonstrated a technique for loading a test simulation model from real world data and provided three data bases.

IX. REFERENCES

[1]The literature on city classifications is extensive but much of it has been reported on in B.J.L. Berry, CITY CLASSIFICATION HANDBOOK: METHODS AND APPLICATIONS. Wiley-Interscience, New York, 1972.

[2]B. J. L. Berry and F. E. Horton, GEOGRAPHIC PERSPECTIVES ON URBAN SYSTEMS. Prentice-Hall, Englewood Cliffs, New Jersey, 1970, p. 107.

[3]R.H.T. Smith, "Method and Purpose in Functional Town Classification," ANNALS OF THE ASSOCIATION OF AMERICAN GEOGRAPHERS, IV, No. 3 (September, 1965), 539-548, cited in Berry and Horton, op. cit., p. 108.

[4]C.D. Harris, "A Functional Classification of Cities in the United States," GEOGRAPHICAL REVIEW, XXXIII, No. 1 (January, 1943), 86-99.

[5]H.J. Nelson, "A Service Classification of American Cities," ECONOMIC GEOGRAPHY, XXXI, No. 3 (July, 1955), 189-210.

[6]For example, A. Aagesen, "The Population," in ATLAS OF DENMARK, ed. N. Nielsen, C. A. Reitzels Forlag, Copenhagen, 1961, 89-92. Cited in Berry and Horton, op. cit. p. 108.

[7]C. A. Moser and W. Scott, BRITISH TOWNS: A STATISTICAL STUDY OF THEIR SOCIAL AND ECONOMIC DIFFERENCES. Oliver and Boyd, Edinburgh, 1961; T. Yamaguchi, "Japanese Cities: Their Functions and Characteristics," PAPERS AND PROCEEDINGS OF THE THIRD FAR EAST CONFERENCE OF THE REGIONAL SCIENCE ASSOCIATION, Vol. 3, 1969, 141-156.

[8]Robert R. Alford, "Critical Evaluation of the Principles of City Classification," in Berry, op. cit., p. 337.

[9] B.J.L. Berry, "Grouping and Regionalizing: An Approach to the Problem using Multivariate Analysis," in W.L. Garrison and D. F. Marble, editors, QUANTITATIVE GEOGRAPHY, PART I: ECONOMIC AND CULTURAL TOPICS. Northwestern University, Evanston, Ill., 1967, p. 245. The remarks in brackets are ours.

[10] R.H.T. Smith, op. cit., p. 111.

[11] Ibid., 112.

[12] David S. Arnold, "Classification as Part of Urban Management," in B.J.L. Berry, op. cit., p. 362.

[13] D. J. Veldman, FORTRAN PROGRAMMING FOR THE BEHAVIORAL SCIENCES. Holt, Rinehart and Winston, New York, 1967.

[14] Thomas F. Golob, Eugene T. Canty, and Richard L. Gustafson, "Classification of Metropolitan Areas for the Study of Arterial Transportation." Paper presented at the 1972 Annual Meeting of the Transportation Research Forum in Denver, November 8-10, 1972. Research publication GMR-1225, Research Laboratories General Motors Corporation, Warren, Michigan.

[15] R. J. Johnston, "Grouping and Regionalizing: Some Methodological and Technical Observations," ECONOMIC GEOGRAPHY, Vol. 46, No. 2 (June, 1970), pp. 293-305.

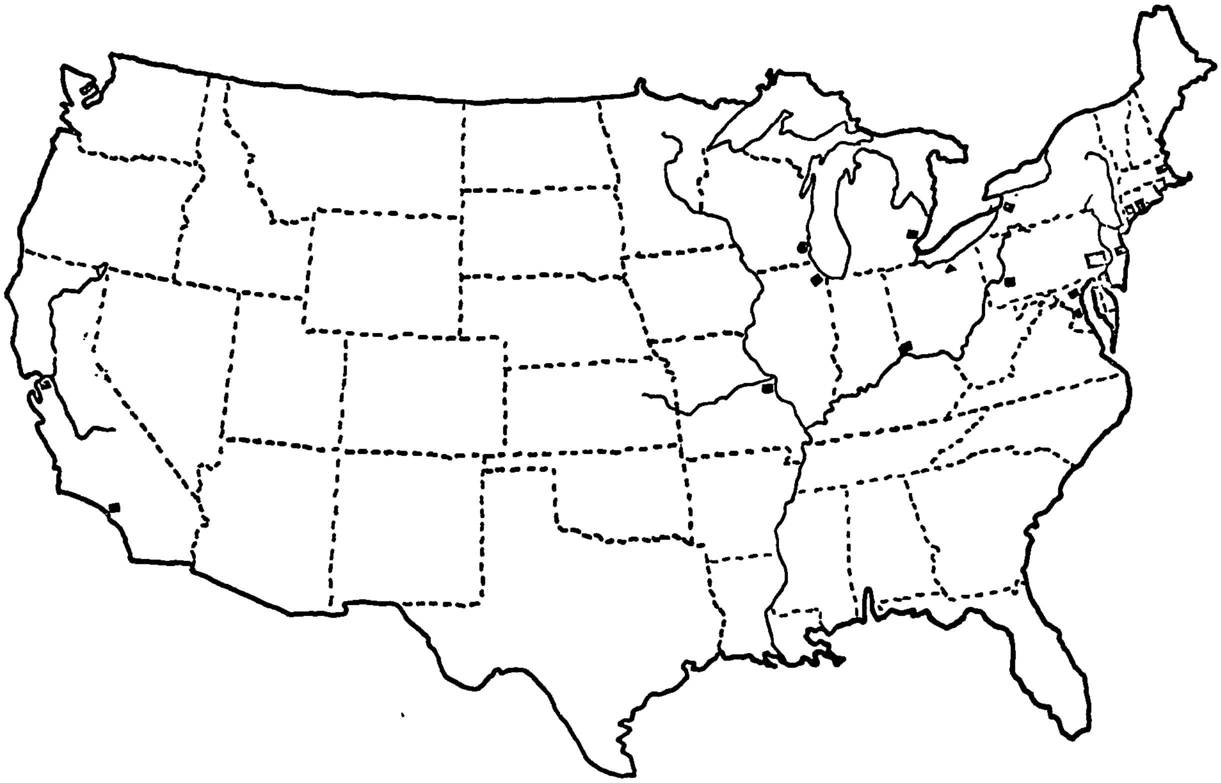
[16] Ibid., p. 302.

X. APPENDICES

APPENDIX A

MAPS

MODAL CITIES DISTRIBUTION
TYPE A



KEY

■ PHILADELPHIA, PA. (PRIMARY)

▲ CLEVELAND, OHIO (SECONDARY)

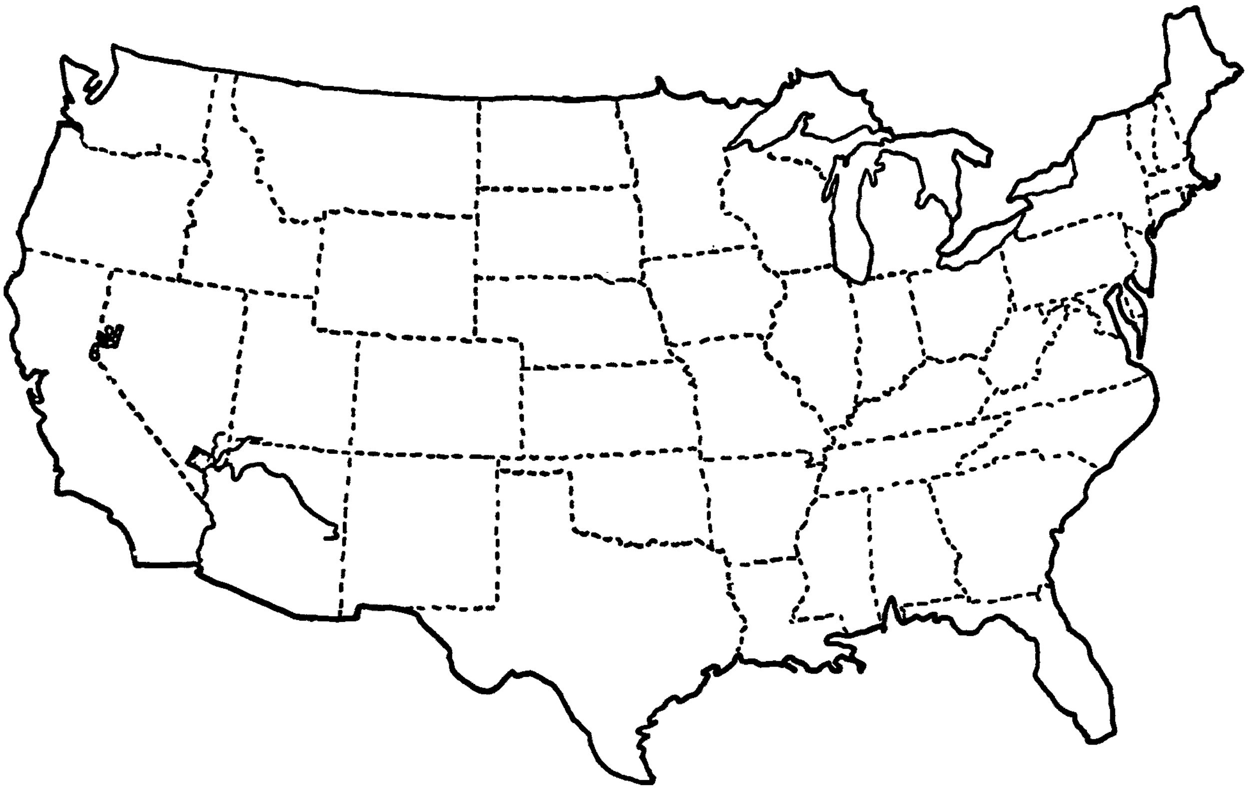
OTHER CITIES

SCALE

0 200 400 600 800 1000 MILES

0 200 400 600 800 1000 1200 1400 KILOMETERS

MODAL CITIES DISTRIBUTION
TYPE B



KEY

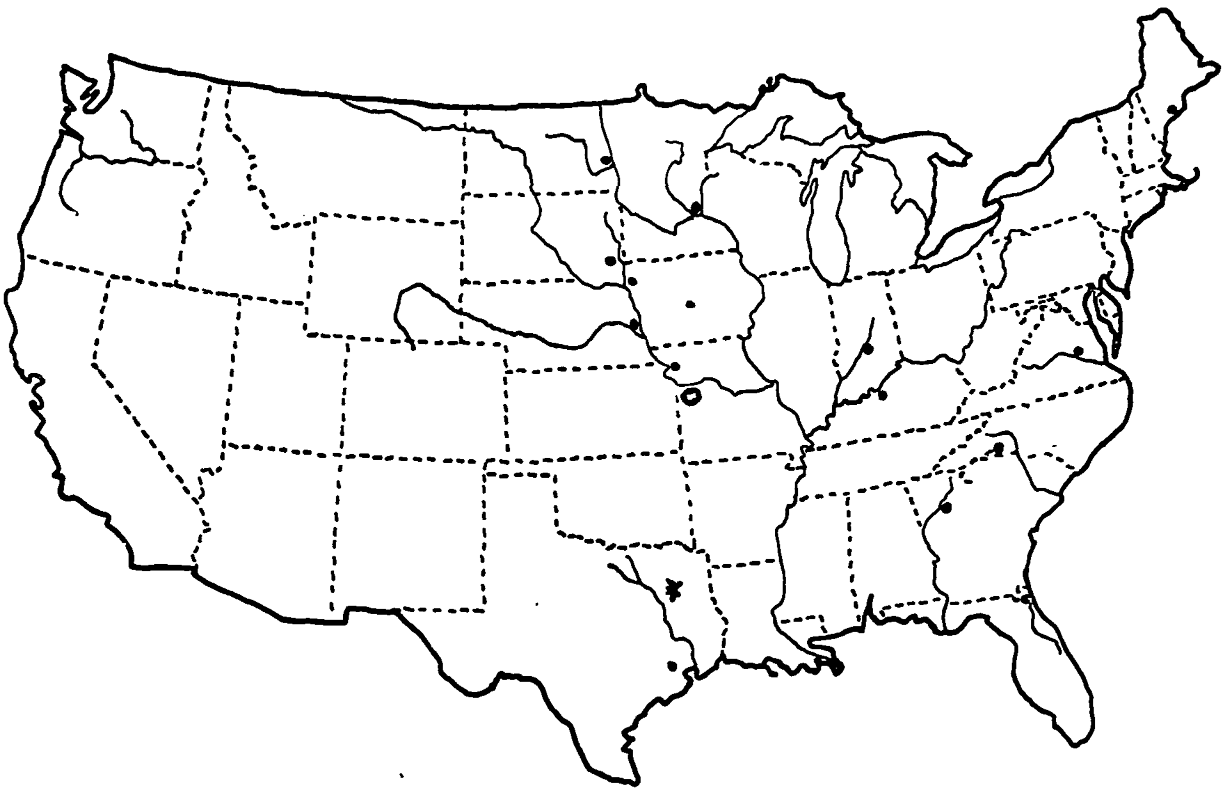
■ LAS VEGAS, NEV. (PRIMARY)

▪ RENO, NEV. (SECONDARY)

SCALE

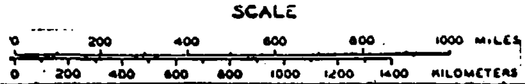
0	200	400	600	800	1000 MILES		
0	200	400	600	800	1000	1200	1400 KILOMETERS

MODAL CITIES DISTRIBUTION
TYPE C

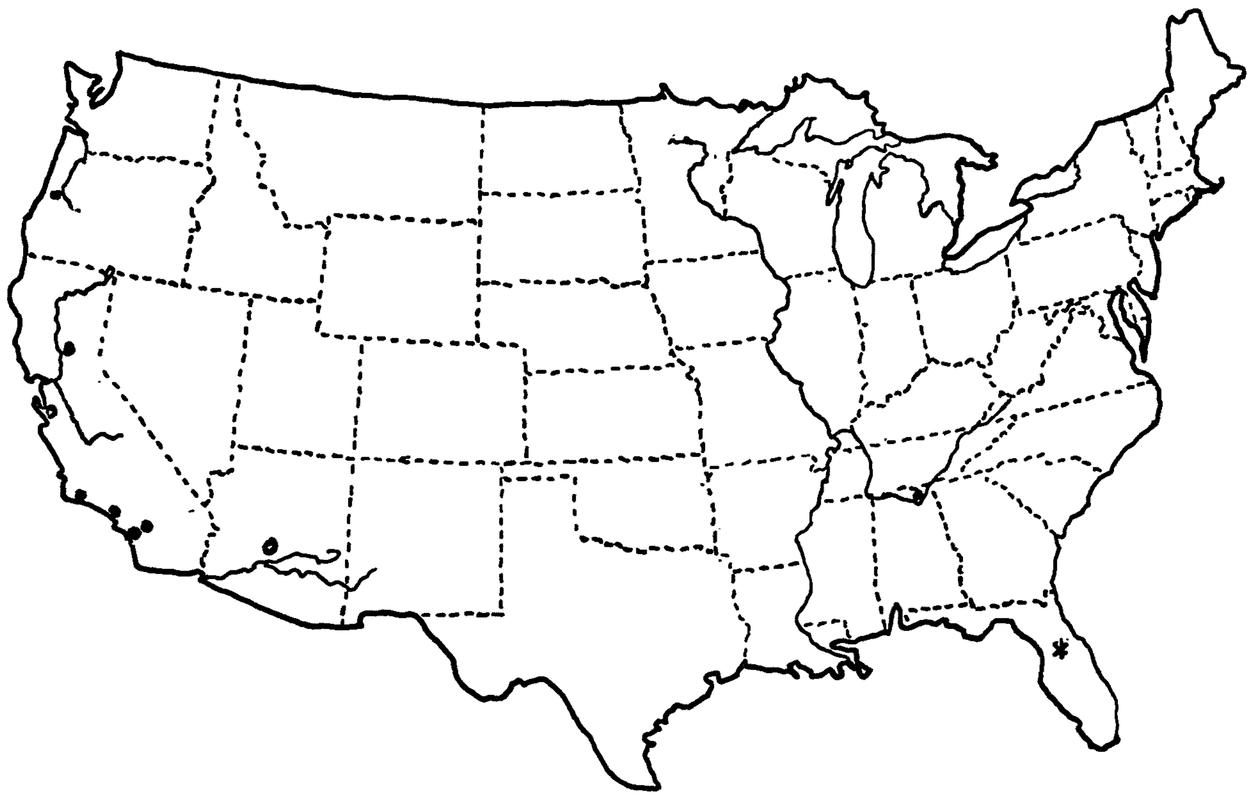


KEY

- KANSAS CITY, MO. (PRIMARY)
- DALLAS, TEXAS (SECONDARY)
- OTHER CITIES



MODAL CITIES DISTRIBUTION
TYPE D

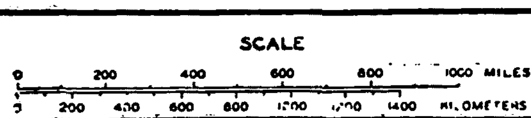


KEY

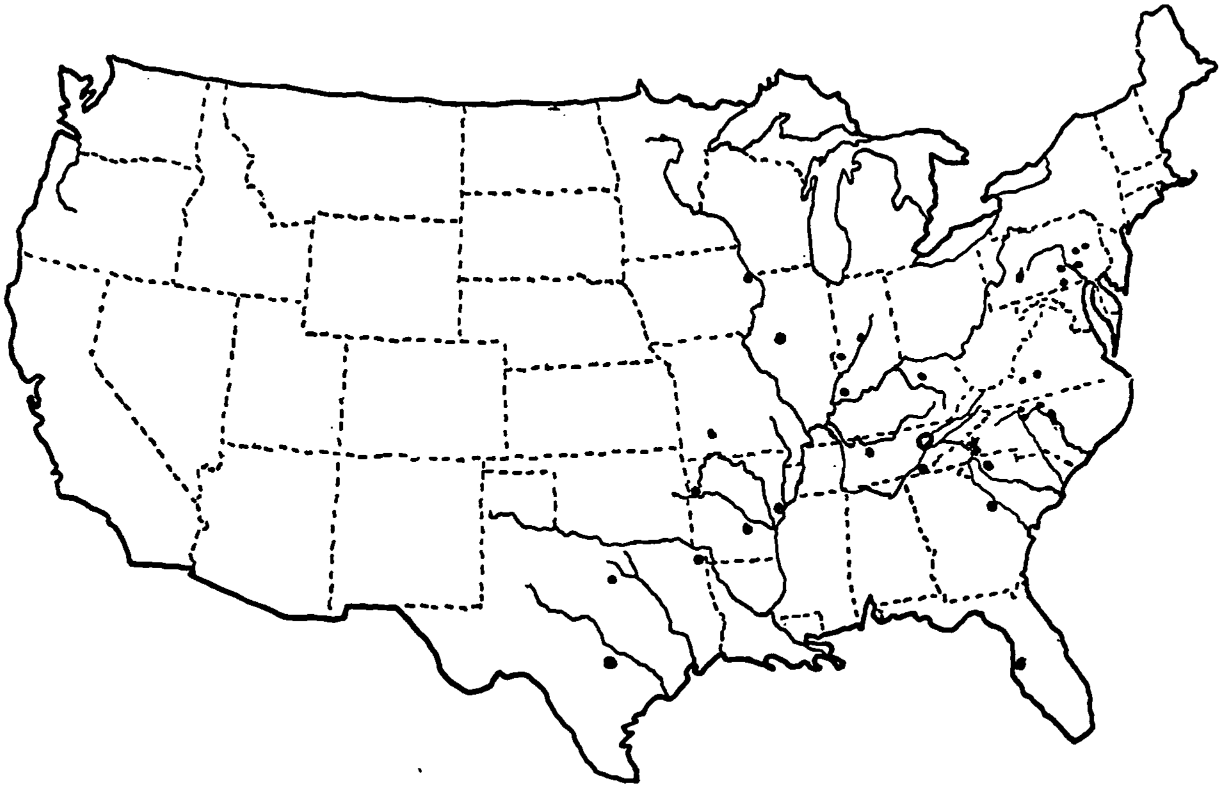
⊖ PHOENIX, AZ. (PRIMARY)

* ORLANDO, FLORIDA (SECONDARY)

• OTHER CITIES



MODAL CITIES DISTRIBUTION
TYPE E



KEY

O KNOXVILLE, TE. (PRIMARY)

✱ ASHVILLE, N.C. (SECONDARY)

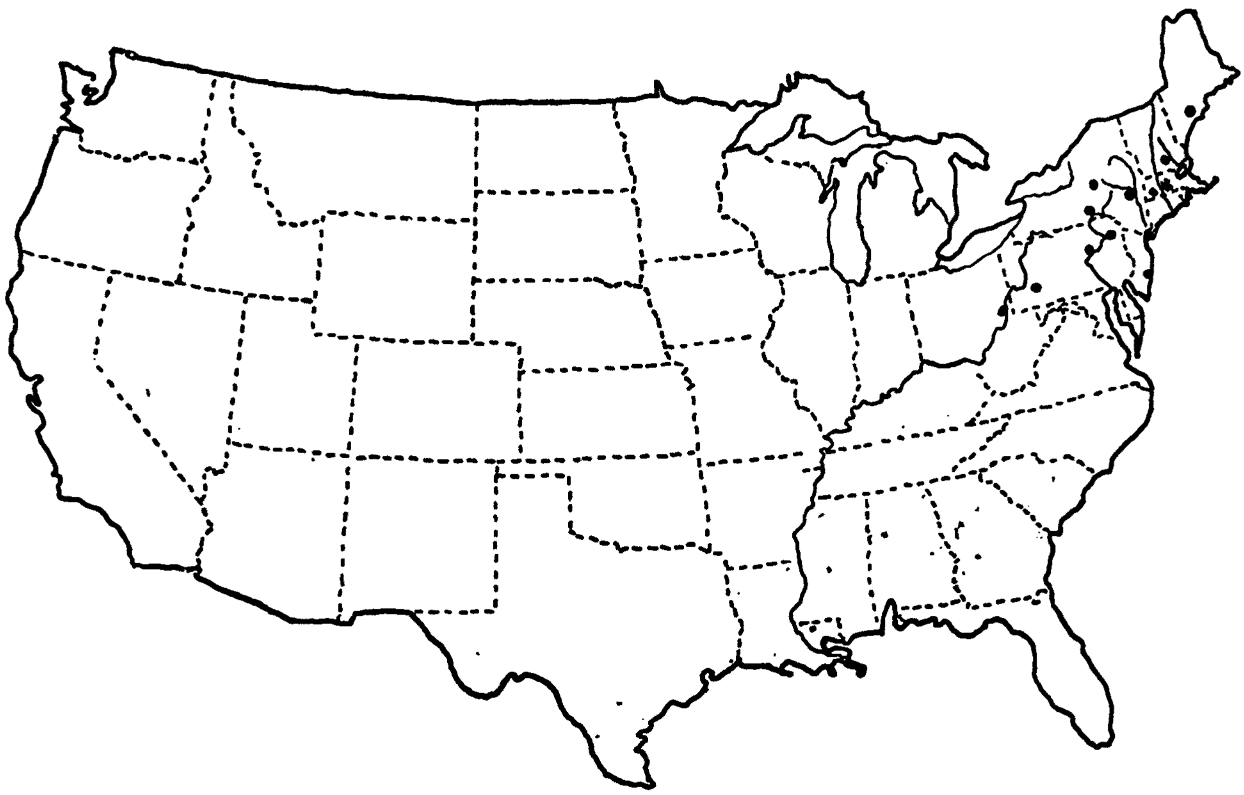
• OTHER CITIES

SCALE,

0 200 400 600 800 1000 MILES

0 200 400 600 800 1000 1200 1400 KILOMETERS

MODAL CITIES DISTRIBUTION
TYPE F



KEY

LOWELL, MASS. (PRIMARY)

WORCESTER, MASS. (SECONDARY)

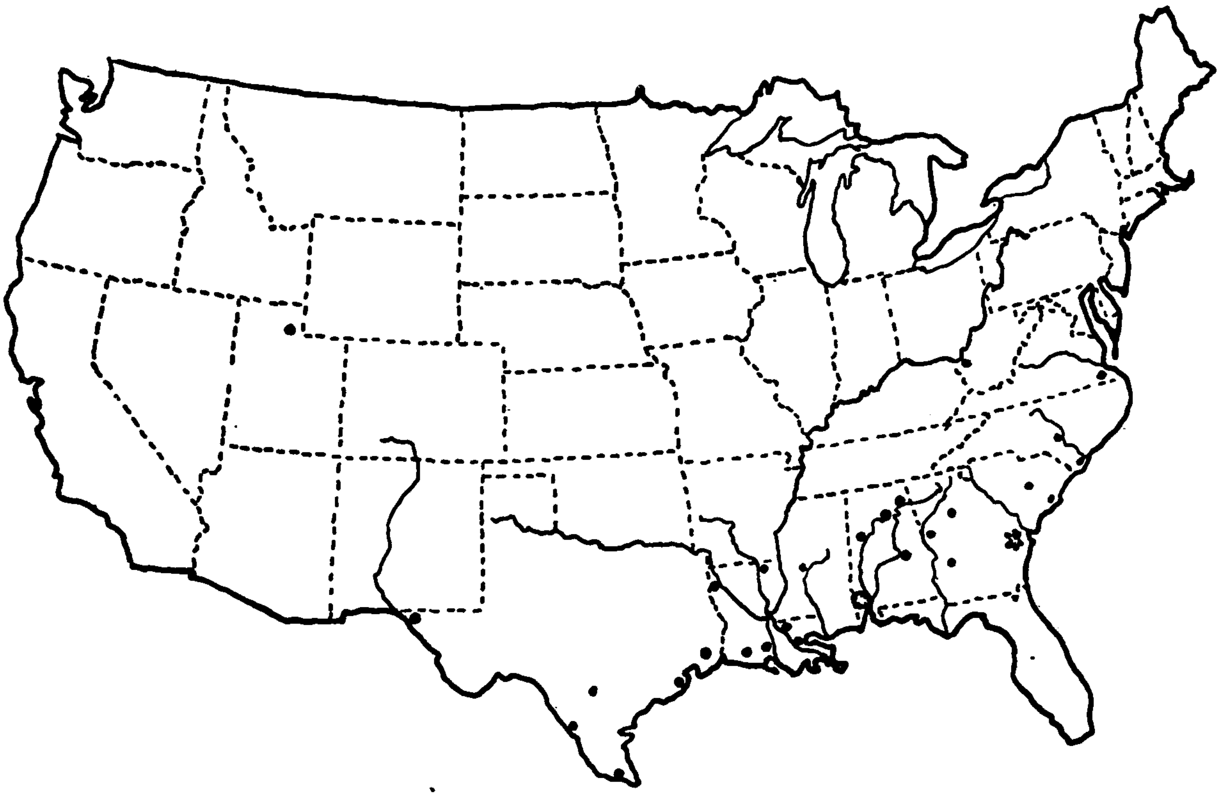
OTHER CITIES

SCALE

0 200 400 600 800 1000 MILES

0 200 400 600 800 1000 1200 1400 KILOMETERS

MODAL CITIES DISTRIBUTION
TYPE G



KEY

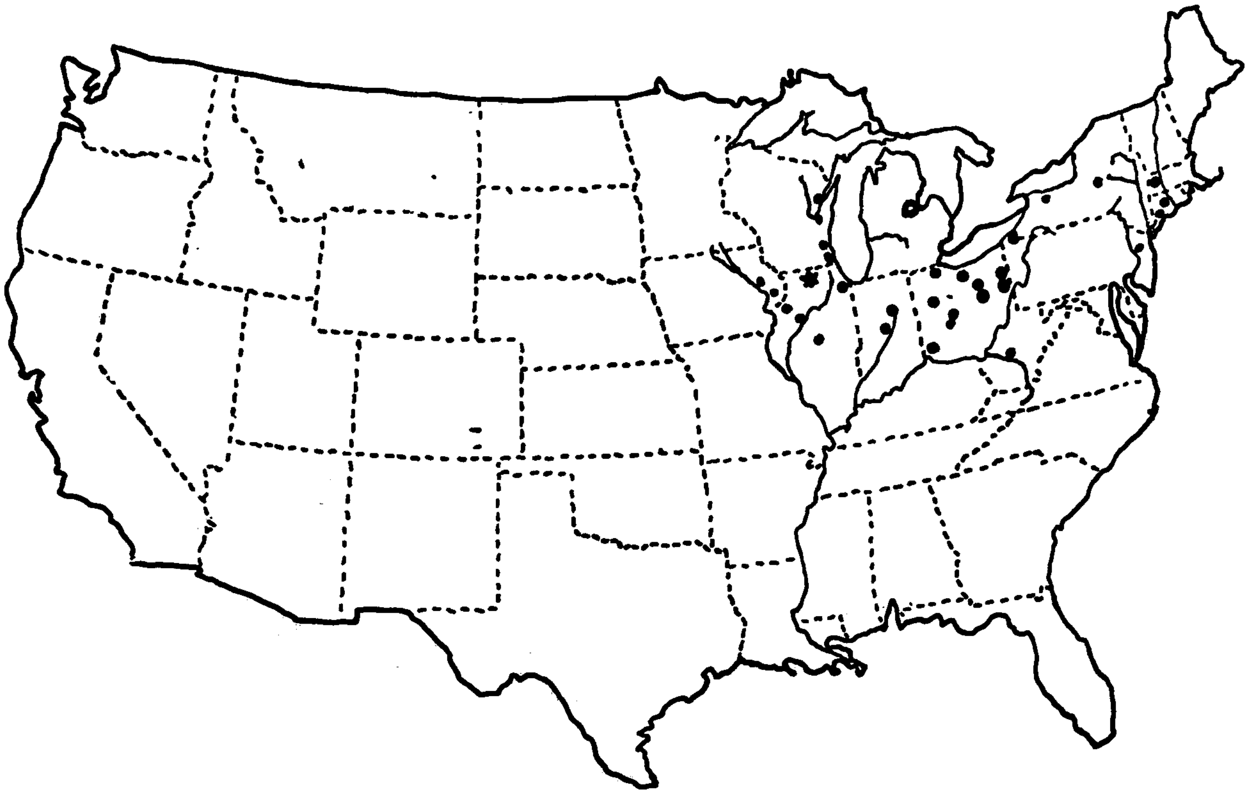
o MOBILE, ALA. (PRIMARY)

* SAVANNAH, GA. (SECONDARY)

• OTHER CITIES

SCALE
0 200 400 600 800 1000 MILES
0 200 400 600 800 1000 1200 1400 KILOMETERS

MODAL CITIES DISTRIBUTION
TYPE H



KEY

SAGINAW, MICH. (PRIMARY)

ROCKFORD, ILL. (SECONDARY)

OTHER CITIES

SCALE

0 200 400 600 800 1000 MILES

0 200 400 600 800 1000 1200 1400 KILOMETERS

MODAL CITIES DISTRIBUTION
TYPE I

KEY

TULSA, OKLA.	(PRIMARY)
TACOMA, WASH.	(SECONDARY)
OTHER CITIES	

SCALE.

0 200 400 600 800 1000 MILES

0 200 400 600 800 1000 1200 1400 KILOMETERS

SELECTED WATER RESOURCES ABSTRACTS INPUT TRANSACTION FORM	1. Report No.	2	3. Accession No. W

4. Title

Modal Cities

5. Report Date

6.

8. Performing Organization Report No.

7. Author(s)

John W. Sommer and George B. Pidot, Jr.

Department of Geography
Dartmouth College
Hanover, New Hampshire

10. Project No.

1HA096

11. Contract/Grant No.

801226

12. Sponsoring Organization Environmental Protection Agency

13. Type of Report and Period Covered

Final

15. Supplementary Notes Environmental Protection Agency Report No. EPA-600/5-74-027, dated October 1974

16. Abstract

Model cities are representative cities based on a specific set of criteria. Using principal components analysis, 224 U.S. SMSA's were examined in terms of 48 selected variables. This analysis yielded 14 dimensions, of which 7 explained 67% of the variance. The 224 cities were then grouped using a method that minimizes the differences among cities within a group and maximizes the differences across groups. This procedure allowed for a confident selection of 9 modalities of the U.S. metropolitan system. Each city fell into a modality and was ranked relative to its distance from the mean. The two cities closest to the mean were taken as representative of that group. One unforeseen result of this research was the distinct regional character of the different groupings.

17a. Descriptors

Principal Components Techniques; Principal Components Analysis; Modal Cities

17b. Identifiers

17c. COWRR Field & Group

18. Availability	19. Security Class. (Report) 20. Security Class. (Page)	21. No. of Pages 22. Price	Send To: WATER RESOURCES SCIENTIFIC INFORMATION CENTER U.S. DEPARTMENT OF THE INTERIOR WASHINGTON, D.C. 20240

Abstractor John W. Sommer | Institution Dartmouth College