HIGHWAY RESEARCH RECORD

Number 194

Information Systems for Land Use and Transportation Planning

8 Reports

	Subject Area
15	Transportation Economics
81	Urban Transportation Administration
83	Urban Land Use

HIGHWAY RESEARCH BOARD
DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

Washington, D.C., 1967
Publication 1525

Price: $3.40

Available from

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

Department of Urban Transportation Planning

Pyke Johnson, Chairman
Washington, D. C.

HIGHWAY RESEARCH BOARD STAFF

J. A. Scott, Urban Transportation Planning Specialist

COMMITTEE ON LAND USE EVALUATION

(As of December 31, 1966)

John R. Hamburg, Chairman
Creighton Hamburg, Incorporated
Delmar, New York

Joseph R. Stowers, Secretary
Civil Engineering Department
Northwestern University
The Technological Institute, Evanston, Illinois

Frederick T. Aschman, Executive Vice President, Barton-Aschman Associates, Incorporated, Chicago, Illinois
Robert E. Barraclough, Chief, Land Development Section, Tri-State Transportation Committee, New York City
Stuart F. Chapin, Jr., Research Director, Center for Urban and Regional Studies, University of North Carolina, Chapel Hill
Salvatore V. Ferrera, Chief Economist, Chicago Area Transportation Study, Chicago, Illinois
Aaron Fleisher, Associate Professor of Urban Studies, Department of City and Regional Planning, Massachusetts Institute of Technology, Cambridge
Walter D. Gaby, Director, Community Development Branch, Urban Renewal Administration, Department of Housing and Urban Development, Washington, D. C.
William L. Garrison, Director, The Transportation Center, Northwestern University, Evanston, Illinois
David A. Grossman, Acting Assistant Director for Program Review, Community Action Program, Washington, D. C.
Walter G. Hansen, Alan M. Voorhees and Associates, Inc., McLean, Virginia
Britton Harris, Professor of City and Regional Planning, Institute for Environmental Studies, University of Pennsylvania, Philadelphia
C. David Loeks, Director, Twin Cities Metropolitan Planning Commission, University at Fairview, St. Paul, Minnesota
David D. Longmaid, Director, Delaware Valley Regional Planning Commission, Philadelphia, Pennsylvania
Ira S. Lowry, Logistics Department, The RAND Corporation, Santa Monica, California
John K. Mladinov, Director, Puget Sound Regional Transportation Study, Seattle, Washington
Kenneth J. Schlager, Chief Systems Engineer, Southeastern Wisconsin Regional Planning Commission, Waukesha
R. J. Wheeler, College of Engineering, University of Missouri, Columbia
Lowdon Wingo, Jr., Resources for the Future, Inc., Washington, D. C.

COMMITTEE ON URBAN MEASUREMENTS
(As of December 31, 1966)

Edgar M. Horwood, Chairman
Professor of Civil Engineering
University of Washington, Seattle

Kenneth J. Dueker, Secretary
Assistant Professor of Civil Engineering
University of Wisconsin, Madison

George Belknap, Director of Research Development, Office of Programs Policy, U. S. Department of Housing and Urban Development, Washington, D. C.
William S. Cadzow, Jr., Vice President, Quinn and Associates, Inc., Wyncote, Pennsylvania
Leon Monroe Cole, Department of City and Regional Planning, Harvard University, Cambridge, Massachusetts
Roger L. Creighton, Director, Transportation Planning & Programming Subdivision, Division of Finance and Planning, New York State Department of Public Works, Albany
Robert B. Dial, Systems Analyst, Alan M. Voorhees and Associates, McLean, Virginia
Charles H. Graves, U. S. Department of Housing and Urban Development, Urban Renewal Administration, Washington, D. C.
Richard K. Guenther, Bethesda, Maryland
Edward F. R. Hearle, Assistant Director for Program Planning, Office of Regional Economic Development, U. S. Department of Commerce, Washington, D. C.
George C. Hemmens, Department of City and Regional Planning, University of North Carolina, Chapel Hill
Herbert S. Levinson, Principal Associate, Wilbur Smith and Associates, New Haven, Connecticut
T. J. Morawski, Assistant Division Engineer, U. S. Bureau of Public Roads, Region Three, Nashville, Tennessee
Robert L. Morris, Senior Transportation Planner, DOWNTOWN PROGRESS, Washington, D. C.
Clark D. Rogers, Instructor in Urban Planning, Graduate School of Public and International Affairs, University of Pittsburgh, Pittsburgh, Pennsylvania
Lawrence E. Schlesinger, Chevy Chase, Maryland

Department of Economics, Finance and Administration

Guilford P. St. Clair, Chairman
Chief, National Highway Planning Division
U. S. Bureau of Public Roads, Washington, D. C.

R. C. Blensly, Vice Chairman
Salem, Oregon

HIGHWAY RESEARCH BOARD STAFF

K. E. Cook, Highway Economist

COMMITTEE ON ECONOMIC FORECASTING
(As of December 31, 1966)

C. A. Steele, Chairman
Deputy Chief, Economic Research Division
U. S. Bureau of Public Roads, Washington, D. C.

Edmond L. Kanwit, Secretary
Supervisory Economist
U. S. Bureau of Public Roads, Washington, D. C.

V. Lewis Bassie, Director, Bureau of Economic and Business Research, University of Illinois, Urbana
James S. Burch, State Planning Engineer, North Carolina State Highway Commission, Raleigh
Robinson Newcomb, Washington, D. C.
Robert W. Paterson, Director, Research Center, School of Business and Public Administration, University of Missouri, Columbia
Marvin Tummins, Department of Economics, University of Virginia, Charlottesville

Foreword

The eight papers in this RECORD focus on varying aspects of data requirements and needs for land-use models and information systems for urban transportation planning. The papers represent a part of an increasing body of literature that is specifically addressed to questions of data requirements and needs for planning and decision-making.

Barraclough emphasizes conservativeness in selecting items of information for land-use models because of financial constraints and the relative youthfulness of the land-use model field. He suggests four different minimum sets of items of information for modeling. The fourth set, information on accessibility, vacant land, land-use areas, and floor area, appears to Barraclough to have important advantages over other minimum sets. A strong case is made by the author for floor area as having important advantages over all the minimum sets.

McMains and Fleisher describe the specifications and design of a computer system for urban studies. They emphasize various component elements of the computer system including, first, a hierarchical structure for data storage which allows for more detailed inquiries into the kinds of information contained in the system, fast retrieval of the information selected, the structure as well as the data to be edited, and automatic editing of the structure with addition of any new data; second, an equivalent structure containing programs to analyze and manipulate the data retrieved and a set of graphical procedures for displaying the results; and finally, a simple English-like language for dealing with these structures and their contents.

Hansen and Voight outline various U. S. Bureau of Census programs and activities of particular value to planners, researchers, and administrators involved in decisions which require knowledge of the characteristics of small urban areas. The authors describe plans for further research and development in applying Census data to small-area problems under live field conditions in New Haven, and the potential they hope to explore in new geographic coding capabilities in cooperation with federal, state and local agencies.

Bottiny and Goley present results of research on the initial steps in the development of an urban area transportation typology. Various equations and concepts were tested and tried for estimating certain indications of automobile availability and travel characteristics.

Worrall postulates that present modeling efforts are constrained by the characteristics of existing data systems and proposes, therefore, the development of a new form of data system for the continuous monitoring of urban information. The mechanism proposed is that of a permanent household response panel, analogous to the panels frequently employed in research and market research.

Dueker discusses the needs of transportation planning for information and postulates that the needs make specialized demands on information systems concepts and technology requiring a means of responding to unforeseeable data demands.

Essential to accommodating unforeseeable demands, according to Dueker, are a generalized data-handling capability and data organized in an intermediate form.

Horwood outlines logical bases of new urban information systems under development, discusses applications to general urban analysis and relates these application potentials to urban transportation research, analysis, and planning. Particular emphasis is given to the relationship between new information-handling capabilities and the needs of urban transportation planning.

Calkins discusses the Ottawa Street Address Conversion System being developed in the Ottawa-Hull Metropolitan Area of Ontario. It is a geocoding system and the author points out its functions and describes the current project in Ottawa.

Contents

Information for Land-Use Models

ROBERT E. BARRACLOUGH, Tri-State Transportation Commission, New York

Land-use models are critical in estimating future travel, but the models available so far in this young field of research leave much to be desired. The information needs for models are therefore unsettled. This paper attempts to determine the best approaches to the general problem of information needs for land-use models at this time, using as a criterion the advancing or retarding of the development of better land-use models.

•THE purpose of this paper is to consider some aspects of the problem of items of information for land-use models. Many different types of information have been indicated as required or desirable and each has an associated cost. The money available for information for land-use models is relatively limited and far from sufficient to buy all the items of information needed for all the metropolitan areas in the United States.

One solution is to limit the number of metropolitan areas that can obtain information for land-use models in a given year. In fact, this solution is presently in effect. On the one hand, not all metropolitan areas have looked for information for land-use models in the same year, and on the other hand, of those looking for funds in the same year, some have been refused funds, and others have been granted only a portion of the funds requested. (Almost all information funds for land-use models are included presently with funds for comprehensive and transportation planning programs administered by the U.S. Department of Housing and Urban Development and the U.S. Bureau of Public Roads.) A good long-run program must consider the information needs of all metropolitan areas and the optimum time period over which these needs should be funded. This kind of program would require a considerable increase in the amount and rate of expenditures for information over present levels. Limiting the number of metropolitan areas that can obtain information for land-use models in a given year is only a short-run solution to the general problem.

Another solution is to increase the amounts of money available for information for land-use models. Such funds should be greatly increased. The time is long overdue for tying the funds spent on planning land development (including funds for information) to the amounts spent on constructing land development. Funds for planning transportation facilities have been a fixed percentage of the moneys spent on constructing transportation facilities for several decades. It is more difficult to do the same thing for land development planning because of the multiplicity of sources for construction funds, but a way should be found to do something similar. The billions of dollars that will be spent on land development in the next 30 years will be many times greater than the billions spent in the last 30 years. Better ways for spending these dollars should be sought constantly through land development planning. Expansion of funds for land development planning (including funds for information) is a basic step in the solution of the general information problem. Even if the planning information dollars are greatly increased, they will still be insufficient to buy all the items of information that have been indicated as necessary or desirable for land-use models for all metropolitan areas in the United States.

Paper sponsored by Committee on Land Use Evaluation and presented at the 46th Annual Meeting.

Another attack on the problem is to reduce unit costs of information collection and handling so that, for a given level of reliability or accuracy, the same items of information can be obtained for a lesser per unit cost. Better data collection methods, better geo-coding systems, better data reduction systems, better data analysis systems, and better data processing hardware have reduced unit costs very substantially in the recent past and thus have helped greatly to reduce the general information problem. Cost-benefit evaluations for information handling and collection improvements, if available, would show very high benefits for the costs involved, and it is perhaps in this area that the returns on investment will be greatest.

Another aspect of the general information problem is the actual items of information that are obtained for land-use models. This part of the problem becomes more important as "data bank" proposals mushroom. It becomes more important in situations where operations concerned with collection and processing information start to have a life of their own, in spite of the fact that there is general agreement among users that these operations should be service-oriented. It also becomes more important in situations where the users are, for good reasons referred to later, not sure of their information requirements.

ULTIMATE PRODUCTS OF A LAND-USE MODEL

Land-use models are tools whose ultimate purpose is to facilitate planning, whether it is land development planning, transportation planning or comprehensive planning. Land-use models facilitate planning by producing forecasts of the future uses of land in each small area of a metropolitan region—forecasts of the types of use, the intensities of these uses and their locations. The land-use planner can use the output of the land-use model, the forecast, as a basis for preparing land-use plans. If the thinking of the planner that goes into the preparation of land-use plans can be incorporated into the model, then the model may be used to produce a land-use plan.

Several important claims are made for land-use models: (a) that they force explicit statements about the ingredients of the model so that the resulting forecast or plan should be more objective, and less subjective, than a forecast or plan not based on a model; (b) that a model permits answering policymaker's questions of the "what-if-we-did-this" variety. The policymaker's idea becomes an ingredient, the model is run, and the result is a forecast or plan based on inclusion of the policymaker's idea. This can be so time-consuming and complex as to be almost impossible without a model; (c) that models permit computerization of the preparation of a forecast or plan so that once the model is completed, computer programs written, and the basic input ready, it should be possible to produce several forecasts or plans much more rapidly than it would be possible to produce two plans or forecasts without the use of a model; and (d) that it should be easier to compare the quantities of land uses shown in the forecasts or plans with the quantities shown in later land-use surveys. Observed differences between earlier forecasts and later surveys may be used to refine and improve a model so that its forecasts become more accurate. This operation can be very time-consuming and complex without a model. No doubt, this partly explains why it has been almost entirely neglected to date.

Underlying these important claims is the fact that the information input for a model must be expressed quantitatively and the model itself must be expressed mathematically. The information must be quantitative (i.e., it must be in a form that can be manipulated mathematically), and it must be in a form that can be used in a computer (i.e., it must be in machine-readable form).

Because models are expressed mathematically, much of the discussion is carried on in mathematical terms and it is sometimes difficult, especially for the non-mathematician, to visualize the ultimate products of a land-use model. It is important to have a good understanding of model outputs in order to have a good understanding of some of the basic general requirements of information for input to the model. Broadly speaking, the computer run of the model will produce a forecast or plan expressed in quantities of land use for each small area throughout the metropolitan region.

To obtain a clear picture of the output of a model, suppose that land-use models capable of accurate forecasts did exist in 1930, and that a model were used in 1930 to produce a forecast of future land development in the Tri-State region (the metropolitan region centered on New York City that covers parts of the three states of Connecticut, New Jersey and New York, with a total area of 8,000 sq mi and a total population of 18,000,000). Suppose further that a three-dimensional diagram were built in 1930 to portray the forecast output from the 1930 land-use model. For convenience, suppose that the forecast year were 1963. The diagram would look very much like the one shown in Figure 1.

This three-dimensional stick diagram[1] shows the 1963 total floor area in each square mile throughout the Tri-State region. The highest stick in the diagram represents about 170 million sq ft of floor area on the square mile surrounding Grand Central Station; the regional total of floor area represented in this diagram is 9 billion square feet, about 500 sq ft per person.

This supposed forecast of 1963 land development from the supposed 1930 land-use model gives the quantities of land uses by small area, or intensities of land uses and their locations throughout the region. It would have been very useful to the planners of the 1930's that were concerned with planning the future of the Tri-State region. If this 1930 forecast had also reflected a forecast of quantities of travel demand, it would have been very useful to the 1930 transportation planner too.

A comparison of Figure 1 with Figure 2 demonstrates that the 1963 travel demand in fact was reflected in the 1963 floor area. Figure 2 shows the 1963 person trip

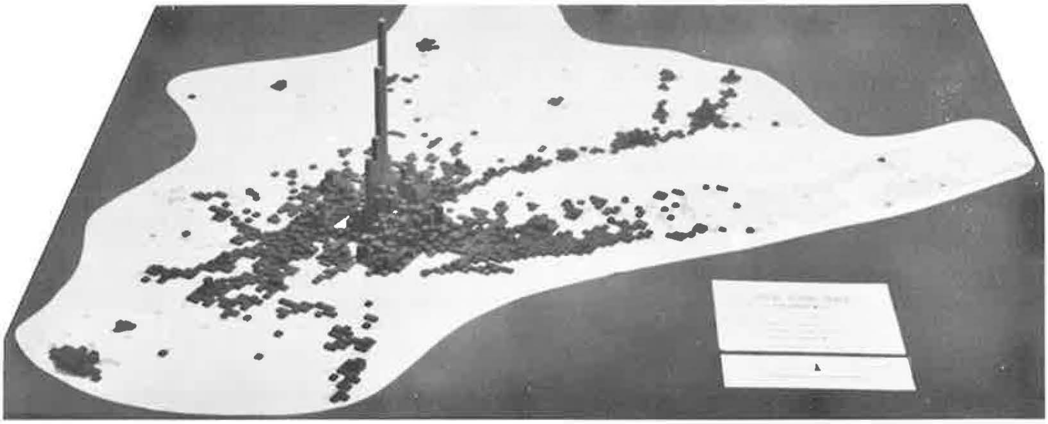


Figure 1. Total floor space in the Tri-State region, 1963. The Tri-State region contains 9 billion sq ft of floor area. Each block represents total floor area per square mile. The highest value is 173,000,000 sq ft for the square mile around Grand Central Station in Manhattan.

---

[1] It would be more correct to refer to the figures as three-dimensional models instead of diagrams, but this is purposely avoided to minimize confusion with "model" as in land-use model.

The stick diagrams of 1963 conditions in the Tri-State region are based on computerized measures of floor area and trips that resulted from large-scale field surveys, which were carried out by the Tri-State Transportation Commission in late 1963 and 1964; field work was followed by coding, land and floor area measurement, and conversion of the raw data to magnetic computer tape in 1965 and 1966. Computer maps were produced from the tape records and the numbers displayed on these computer maps for each square mile were used to determine stick lengths and thus to produce the three-dimensional stick diagrams.

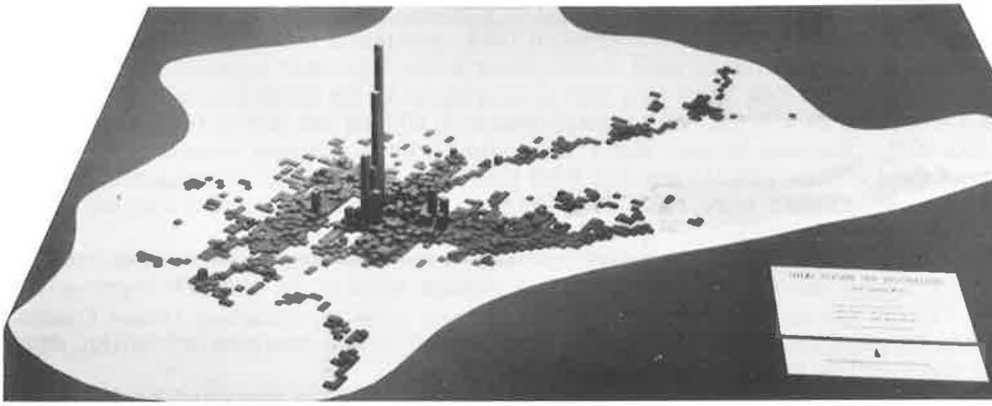


Figure 2. Total person trip destinations in the Tri-State region, 1963. The Tri-State region has a total of 34,000,000 person trip destinations daily. The highest value square mile has 800,000 person trip destinations daily.

destinations per square mile in the Tri-State region, and again the highest value appears in the same square mile that contained the highest value for floor area. This square mile had 800,000 person trip destinations daily on an average weekday, about four destinations per 1,000 sq ft of floor area. Because of the close similarities between the two figures, it is evident that the average rate of four trips per 1,000 sq ft of floor area, or a rate very close to this, is found throughout the region.

It can be supposed further that the 1930 forecast of 1963 land development was subdivided into broad types of land use—residential and nonresidential—and that again the forecasts were represented in three-dimensional diagrams. Figures 3 and 4 show the 1963 residential floor area and 1963 nonresidential floor area, respectively. Figure 5 shows the 1963 trip destinations to nonresidential land uses. Comparison of Figures 4 and 5 indicates, again, striking parallels between the quantities of floor area and the quantities of trip destinations.

These three-dimensional diagrams of 1963 conditions, supposedly forecast from 1930 land-use models, should convey a clear picture of the output of an imaginary 1930

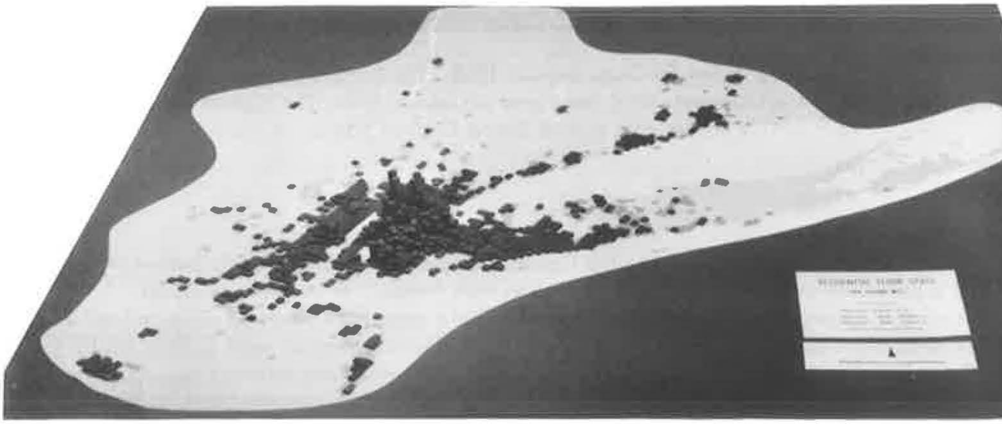


Figure 3. Residential floor space in the Tri-State region, 1963. The Tri-State region contains a total of 5.6 billion sq ft of residential floor area, about 310 sq ft per person. The highest value square mile contains 55,000,000 sq ft of residential floor area.

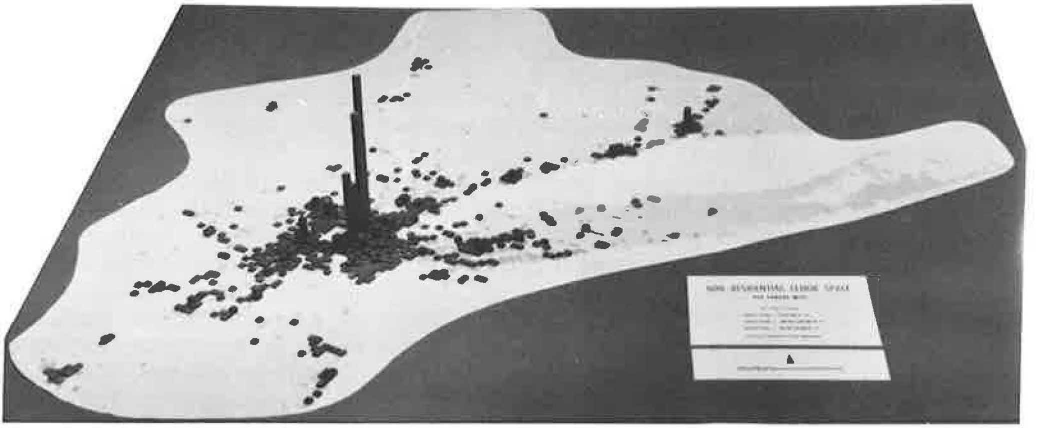


Figure 4. Nonresidential floor space in the Tri-State region, 1963. The Tri-State region contains a total of 3.4 billion sq ft of nonresidential floor space, about 190 sq ft per person. The highest value square mile contains 145,000,000 sq ft of nonresidential floor area.

land-use model. Today's land-use models should be producing equivalent statements for, say, 1990. They should give forecasts and plans of the types, intensities and locations of uses of land by small areas throughout the region.

From the viewpoint of information requirements, there are a number of other important characteristics of the output of models to be considered, whether these are forecasts or plans, or both. If the models are built to produce forecasts without plans, the forecasts should be in a form that allows them to be used easily by a land-use planner. The forecasts and plans should be cast in terms that facilitate testing and evaluation. They should be in terms that facilitate the preparation of detailed area forecasts or plans which will tell what quantities of new factories, offices, shopping, housing, recreation fields, and other facilities will be built, when and where they will be built, and at what cost.

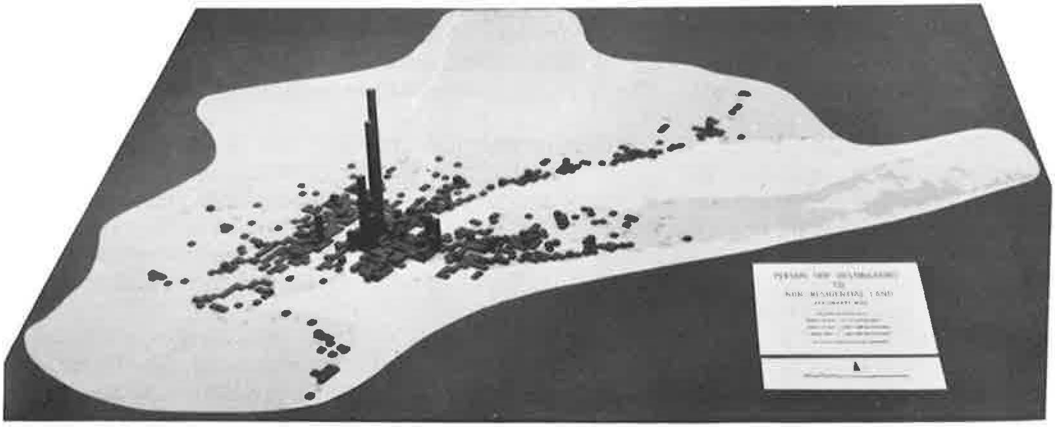


Figure 5. Person trip destinations to nonresidential land uses, 1963. The Tri-State region has a total of 18,000,000 person trip destinations daily to nonresidential land uses, about one per person. The highest value square mile has 720,000 person trip destinations daily.

They should be in a form that permits translation of future land development quantities into future travel demand so that the transportation planner can test proposed networks of transportation facilities, derive a "best" transportation plan, and indicate what facilities should be built, where, when, and at what cost. They should be easily translatable into terms that apply to cost-effectiveness and capital budget studies, or planning-programming-budgeting systems.

Three-dimensional diagrams of the region which show the planned future public capital investment for each square mile, instead of floor area or trip destinations, would be very useful to elected officials, budget directors, planners, public and private developers, as well as the general public. Proposed public non-capital expenditures, which can be pinned to locations, could also be shown in comparable three-dimensional diagrams. Adding together the diagrams for proposed capital and non-capital expenditures would give a clear picture of just where in the region the public moneys should be spent in the future. In most regions this picture is totally lacking at present; in others it is blurred. It should exist clearly in every region and should be used to tie together the work of the various planning, budgeting, and operational agencies concerned with executing governmental responsibilities in regard to building, maintaining and improving the physical, social and economic facets of our metropolitan regions.

From this discussion, the following basic general requirements may be noted in regard to information inputs for land-use models: (a) the information must be in a quantified and machine-readable form, (b) the information must refer to the intensities and locations of land uses in the region by small areas, (c) the information should be in terms that facilitate later reading-off of quantities of travel demand by small areas as a basis for transportation planning, and (d) the information should be in terms that facilitate reading-off of capital investment and other costs as well as guideline information for the preparation of detailed forecasts and plans for small areas.

DIFFICULTIES IN OBTAINING INFORMATION FOR LAND-USE MODELS

There are a number of practical difficulties that any comprehensive or transportation planning study must face today in obtaining information for land-use models, or in deciding on the items of information that will be obtained. Some of the more important of these practical difficulties are given here.

First, the land-use model field is still quite young. Although much work has been done, a great deal remains to be done. There is room for considerable improvement and a number of questions come to mind in trying to second-guess the shape and direction of these improvements. Will there be fundamental changes in the structure of existing land-use models? If so, when are such changes likely to occur? What direction will they take? What effect will they have on information requirements?

Although a wide variety of land-use models have been proposed, only a few are operational today, and some of these only cover an aspect of total land use, for example, commercial land-use models. There is no general agreement as to which of the existing operational models is the best.

Second, although it may appear wise to wait until much better land-use models are developed before spending more money on information for models, it would in fact neither be wise nor practical. It would appear, given the uncertainty about the particular merits of existing models, that perhaps the wisest course would be to delay the information-gathering until an intensive model-building effort or land development theory effort had been completed. But land-use models developed in isolation without reference to data are not likely to be better.

In the absence of a good theory that satisfactorily explains land development (why it occurs where it does, and when it does, quantitatively), there is need, in efforts to build better land-use models, to continually develop hypotheses and to test these against measurements from the real world. For efficient testing and development of hypotheses, the measurements from the real world should be readily available. This calls for carefully selected, well-organized, on-premises, computerized stocks of information and good information-handling weaponry. This should not be interpreted as indicating a need for a data bank. (Data banks are beyond the scope of this paper, but it may be said in passing that the need for a data bank should be studied very carefully and should be justified on a cost-benefit basis.)

It would appear that the requirement for this basic fund of theory-testing data could be regarded separately from the requirement of data for input to the model when it is finally developed, but because of time and cost considerations, the two sets of data requirements are usually "lumped together." This probably tends to minimize data collected for theory-testing purposes only, but it is not clear that this is a basic obstacle to the development of better land-use models.

Third, currently from 30 to 50 percent of the budget of a comprehensive or transportation planning study may go to information when the full costs are taken into account. A related, but perhaps more important, point is that it takes two to three years after specification of the information required to get the information into usable shape. Map and field work take some time, but even greater time is taken in coding, land-area measuring, checking, conversion to machine format, machine editing and file correction. Geo-coding or locational identification of the data is in itself a time-consuming task. Often data collection is thought of in terms of field work only and the time for processing subsequent to field work is grossly underestimated.

Fourth, in relation to the need, there is only a very small number of persons who are well qualified to make sound judgments on both the costs and usefulness of different types of information for land-use models. Persons who know about data collection costs seldom can evaluate the usefulness of the data, and vice versa. Staff who will be working on the land-use model may not be "on board" before data collection starts. Although this is certainly not desirable, it may be necessary in order to adhere to study schedules. Many persons on the staffs of land development transportation studies come from a variety of backgrounds and are new to the field. At present, there is no easy way for them to learn quickly all that they need to know to be able to make sound judgments on the information required for land-use models.

In the final analysis, the decision on what items of information are to be obtained for the land-use model (to be run two or three years later and to be developed partly in the interim) rests with persons in charge at the outset of the study. They must have considerable skilled technical judgment and ability to anticipate and weigh the probable future courses of action, including land-use model building action. A heavy responsibility rests with these persons because their determinations on data to be collected or not collected greatly affects the future model-building directions the study can take. If an item of information is not identified as necessary prior to the start of data collection, it is unlikely that it can be collected at a later date. If an item is deemed necessary at the outset and is not used later, valuable time and money have been wasted.

Also, these difficult decisions on data collection must be faced at the beginning of the study when a considerable amount of time is also required on basic organizational problems—recruitment, space, furniture, supplies, work program, budget, financing, accounting, and administrative procedures.

One general conclusion can be drawn from this brief discussion of some of the practical difficulties that must be faced by current studies: There is great merit in carefully examining the proposed use of each item of information before a decision is reached to collect it. In other words, there is merit in considering a minimum set of items of information for the land-use model. It would be desirable for these items also to have a broad long-term usefulness, to be relatively low in cost, and to be relatively reliable or accurate.

ITEMS OF INFORMATION

A wide variety of land-use models and sub-models have been proposed and some of these have been made operational. The information requirements for all models and sub-models, taken as a whole, are formidable. The following list is not an exact and exhaustive catalog, but rather a summary indication of the sorts of information that are regarded as necessary or desirable for land-use models:

- total population by place of residence;
- population by age-sex groups by place of residence;
- population by family size groups by place of residence;
- population by annual family income groups by place of residence;
- population by industry groups by place of residence;
- population by occupational groups by place of residence;
- total labor force by place of residence;
- labor force by industry groups by place of residence;
- total employment by place of work;
- employment by industry groups by place of work;
- employment by occupational groups by place of work;
- employment by income groups by place of work;
- total annual retail sales by place of sale;
- annual retail sales by retailing groups by place of sale;
- total value of manufactured products by place of manufacture;
- value of manufactured products by industry groups by place of manufacture;
- total government expenditures by place of agency;
- capital and operating government expenditures;
- government expenditures, capital and operating, by agency;
- total person trips by place of destination;
- total person trips by land-use groups by place of destination;
- total market value of land by small area;
- market value of land by land-use groups by small area;
- total market value of land and buildings by small area;
- market value of land and buildings by structural-type groups by small area;
- total housing units by small area;
- housing units by type of structure by small area;
- housing units by density class by small area;
- housing units by condition of structure by small area;
- housing units by age of structure by small area;
- total floor area by small area;
- floor area by land-use groups by small area;
- land area by land-use groups by small area;
- accessibility to region by small area; and
- distance (airline time or cost) to all parts of the region or to the center of the region by small area.

This list represents a massive data-collection and machine-data-handling problem. The problem is compounded by the fact that the items must be on at least one, common small-area basis suited to the particular requirements of the land-use model for which they are intended, and by the fact that flexibility in geographic or subject groupings of the data (including cross-tabulations) is a fundamental and established characteristic of a modern information system. The need for flexibility calls for retaining the basic data in a disaggregated form from which a large variety of geographic and subject summarizations (including cross-tabulations) are possible. Retention of the data in this form means a larger number of records, which increases the size of the data-collection and machine-data-handling problem.

What are the minimum items of information required for a land-use model? If a land-use model has been decided on at the outset of data collection, it is possible to restrict the items of information to just those needed for the model. It is even possible, although probably less desirable from the viewpoint of the need for geographic flexibility, to settle for just those geographic aggregates called for by the land-use model. This course does minimize the items of information required, and it has much to commend it if quick results are important. But as was mentioned earlier, there is a great deal of room for improvement in existing land-use models, so the quick results may be unsatisfactory. If this course is selected, it would be wise to have a clear understanding of land-use model limitations from the beginning.

It is unlikely that there will be a decision before data collection begins as to what land-use model to use. There will probably be a general idea of improving the best of existing models or a wish to try something new, or both. In this case, it is extremely difficult to select wisely a set of minimum items of information which are certain to be highly useful and yet reasonable in cost. Also, the precise details of items of information depend so heavily on the land-use model course being considered, that the question of minimum sets of information items can only be discussed in very general terms.

The list of items of information given can be generalized into four "orders" of information.

First order information refers to the differential qualities of the land platform itself as they affect the quantities of development that can occur on land. In its crudest form, this information would embrace all vacant land, whether flat, steep, swampy or covered by water, and it would simply indicate whether or not the vacant land could be developed.

A quality of the land platform beyond its "buildability," acknowledged to be of prime importance for land-use models, is its accessibility to all other parts of the region, and to other regions. Distance from the region's main center is the simplest measure. Other measures have been used or suggested, such as the sum of distances from several centers weighted according to the size of center; the sum of airline distances from each small area to all other small areas; gravity-type measures which take into account the "drawing-power" of each small area, the minimum travel time to it from all other areas and the "sending-power" of each of the other areas. Recently, there have been both expressions of doubt as to the efficacy of these measures and indications that their importance in land-use models has been greatly underestimated. Apparently, much more work is necessary in this area.

More sophisticated information on the land platform itself would consist of combined measures of buildability, accessibility and availability—a single measure of the probability of development. Supposedly, this is reflected in the true market value of land, which is based on costs of raw land, clearing, leveling, subdividing, drainage, services, and access roads, and such factors as (a) whether the land is held speculatively, or in reserve; (b) the greater attractiveness of some land, due to such things as proximity to beaches, lakes, rivers, natural beauty, areas of historic interest, and prestige value; and (c) the relative accessibility of the land to all other land in the region. True market value would be a measure that could be applied to developed land as well as to vacant land; it could be applied on a comparative basis from one region to another and thus provide a universal site value or probability of development. However, market value of land has not been collected and attempts to obtain it through assessed valuation have been disappointing. It is mentioned here to point to a possible direction of further development of first order information.

Information on the land platform may be obtained on a large-unit area basis, but the desirability (and cost) of ultimately recording the information in land ownership parcel units should be kept in mind.

Second order information concerns the use made of the land platform. This type of information is usually collected in a land-use inventory. A variety of land-use classification schemes has been used in the past, but in the future these are likely to adhere closely to the Standard Land Use Coding Manual recently issued jointly by the U.S. Department of Housing and Urban Development and the U.S. Bureau of Public Roads. Generally, this provides for categorizing land uses on the basis of the activity on the premises and is tied to the Standard Industrial Classification scheme (promulgated by the U.S. Bureau of the Budget) wherever possible. As with the Standard Industrial Classification, there is provision in the Standard Land Use Classification for grouping into broad categories or detailed categories within broad categories (1-, 2-, 3- or 4-digit levels of classification are provided).

A basic measure for second order information is the quantity of land area in each use. This set of information is usually best recorded on a parcel-by-parcel basis to provide maximum flexibility, to facilitate updating, and to take advantage of secondary sources.

Third order information refers to the quantity of development on the land platform within each land-use parcel. The development may be measured in terms of floor area. Numbers of housing units can provide a crude but fairly satisfactory proxy for residential floor area where this has worthwhile cost advantages, in which case the main emphasis in floor area measurement can be directed at nonresidential floor area.

Information on the quality and/or age of the floor area is a possible further dimension. Preferably, this should be in terms of a dollar value per square foot, which reflects present cost of replacement less an allowance for depreciation and obsolescence, and comes close to market value. In this form it could be combined with the market value for land. For some purposes, Census data on condition of housing units may suffice for a measure of the condition of residential floor area. Not only should floor area information be obtained on a per-structure within-land-use-parcel basis, but also it should be recorded on a per-firm or per-household (or housing unit) basis within a structure.

Fourth order information concerns the occupancy of floor area. Occupancy of land without floor area is insignificant and may be disregarded. Examples of floor area occupancy measures are numbers of residents, numbers of employees, numbers of shoppers, visitors and schoolchildren. The crude numbers of occupants can be given further dimensions, such as numbers of resident families by income groups and numbers of shoppers by retail sales groups. Also, numbers of person trip destinations may be considered as reflecting additional dimensions to crude measures of occupancy.

Occupancy information should be obtained on a basis of households (or housing units) or firms. Information on individuals should be recorded on a within-household or within-firm basis. This does not mean that a full census is necessary to obtain occupancy data. Sample surveys will give better and cheaper results in many cases.

Referring again to the question of a minimum set of items of information for land-use models, an admittedly oversimplified grouping has been made which comprises four minimum sets of information given as follows:

1. Accessibility, employment, and population data—a combination of first and fourth order information;
2. Accessibility, vacant land, land-use areas, employment, and population data—a combination of first, second and fourth order information;
3. Accessibility, vacant land, land-use areas, and trip destinations—a combination of first, second and third order information; and
4. Accessibility, vacant land, land-use areas, and floor areas—a combination of first, second and third order information.

None of these minimum sets of items of information embrace all four orders of information. No land-use models developed to date have required all four orders. The cost would be very high indeed, if the four orders of information were collected and maintained on a nesting basis—occupants, within floor area, within structure, within land-use parcel. The closest approximation to proposals for this nesting arrangement has been in data bank proposals. These types of proposals raise questions as to whether the costs and uses of the items of information have been considered carefully.

The first minimum set of items of information (accessibility, employment, and population) is probably the cheapest and perhaps the poorest. The chief objection is the great difficulty, amounting for all practical purposes to impossibility, of obtaining reliable employment data at the work site. Not only is it virtually impossible to obtain reliable data by small place-of-work areas, but also data for past time periods are seldom available. The lack of coverage of all forms of employment in the basic source, plus the reporting of "branch" or "field" employment as at the location of the head office instead of the work site, are the main causes of employment data inadequacies.

Population data are available from the U.S. Bureau of the Census decennially, and sometimes between decades from special censuses. The chief difficulty with these data is usually the geographic areas by which they are available, particularly data for past periods. Frequently, the areas are too large and thus do not permit any geographic flexibility.

Often these objections are overcome in transportation land development studies by obtaining the population and employment data in a sample home interview origin and destination survey, the employment data being "first-work-trip" data. This does give reliable existing population and employment data, but, with rare exceptions, data for past time periods are not available in this way.

The potential role of the U.S. Bureau of the Census in surmounting employment data difficulties should be mentioned. If the Bureau would collect 1970 Census place-of-work data by street address and code them to block face, most of the present difficulties with employment data would be overcome. The Bureau sees the great value of this item not only for employment data but also for journey-to-work data, and for data on the socioeconomic characteristics of workers and their families by small place-of-work areas. Also, the cost (estimated at 5 million dollars) is reasonable, but the Bureau doubts how much support they enjoy nationally among Census data users for this item; consequently, the future remains uncertain until the demand and need are established.

Another fundamental objection to the first set of minimum data is the lack of a tie of the population and employment data to land-use areas so that net density measures can be calculated. Without this tie, only gross density measures are possible.

Gross density measures (which do not take into account the net amounts of land occupied) seem to be insufficient at this stage in the development of land-use models. Few, if any, of the operational models generate densities within the model. They require that estimates of future densities be supplied.

Existing net densities appear to be a more useful basis for estimating future net densities than existing gross densities. This is because these estimates are based on the assumption that future net densities will tend to be close to existing net densities in nearby areas. ("Future" here covers a 20- to 30-year period.)

When in later land-use models densities can be generated within the model, the requirement for net density data will perhaps be eliminated or replaced by a requirement for gross density data. If so, the net density data obtained in the interim would then be able to be converted easily to gross density data. Gross densities can always be computed, given the data for computing net densities, but the opposite is not possible.

The second minimum set of items of information (accessibility, vacant land, land-use areas, employment, and population) differs from the first in that vacant land and land-use areas are added items. Most of the discussion and the objections raised in regard to the first minimum data set apply equally to this second minimum set, with one notable exception. The second set of data permits net density computations.

There remains, however, the problem of correlating the employment and population data with the land-use area data. Temporal, geographic, and definitional differences enter here, and the inaccuracy of the employment data does not make correlation easier. The land-use, employment, and population data for reasons of cost and practicality are not collected in a single inventory (assuring data correlation) but rather as three separate inventories, usually at different points in time, within different geographic areas, by different agencies, and with different definitions. An example of a definitional difference is managers and other office staff of a manufacturing company being classed as "manufacturing" in the employment data, and as "office" or "commercial" in the land-use data. In the future, use of the Standard Land Use Classification scheme will lead to closer accord between the definitions used for employment data and those used for land-use inventories, but other problems in correlation of these data are likely to remain for some time.

The third minimum set of information (accessibility, vacant land, land-use areas, and trip destinations) has been used only recently. The objections stated in connection with the first and second minimum data sets are not present in the third data set. The trip destination data are collected at the same time as the land-use area data. They are coded to block, which permits geographic flexibility; and, although collected in two separate inventories, differences in definitions are largely avoided.

One advantage of this data set (or the model for which it is collected) is that the output in trip destinations is directly usable in the assignment of trips to a network for transportation planning purposes. This is a significant advantage. On the other hand, some people find it difficult to visualize trip destinations as a measure of the intensity of the use of land.

The fourth minimum set of items of information (accessibility, vacant land, land-use areas, and floor areas) is new to the land-use model field and has not been tested, but it appears to have great promise.

This data set will be tested at the Tri-State Transportation Commission study in the near future. The data have already been collected and converted into machine-readable records.

The fourth minimum data set does not have any of the objections raised in connection with the first and second minimum data sets. It is firmly expected that the floor area data can be used directly as input to a process for direct estimation of traffic at a point. The process has been developed recently at Tri-State by Morton Schneider. This would give the fourth data set the same advantage as the third data set.

The direct traffic estimation process has not been fully tested using floor area density data as input, but it has been tested with trip end density data and the results have been good. Reports on the method and the tests are available from Tri-State. The process will be tested with floor area data shortly. The striking parallels between floor area and trip ends (Figs. 1, 2, 4 and 5) are reason for thinking that the results of these tests will be good.

There are a number of points of interest in considering the use of floor area as an input (and output) for a land-use model. Floor area is universally understood and used. The realtor, builder, financier, reporter, architect, engineer, planner, executive, administrator, housing analyst, economist, government official, politician, home owner, renter, landlord and others have a clear idea of what is meant by floor area.

Floor area is a universal density measure which reflects equally well the intensity of nonresidential, residential or total activities. Employment densities for nonresidential activities and population densities for residential activities are difficult to compare with each other. The sum of employment and population is not a good measure of total intensity, because some things are not covered, e.g., visitors or shoppers. Total floor area is a good measure of total intensity, and its components (residential, manufacturing, office, shopping, public, etc.) are good measures of specific intensities.

Floor area is also a direct measure of density of development for each site or land-use area. It is easy to use floor area, before construction and after, as a measure to control density of development. Floor area ratio (FAR), the ratio of floor area to land area, has been used by several government agencies in different parts of the world to control height and bulk of buildings. Its flexibility in details is often cited as an advantage. The same FAR permits a high building with low site coverage or a low building with high site coverage. Although flexible at the detailed level, FAR is rigid at a generalized level. This makes it useful both as a general planning measure and as a detailed planning control measure.

The department of buildings in each local jurisdiction, which is charged with the responsibility of ensuring that buildings are erected and maintained in a decent, safe and sanitary condition, is particularly concerned with structures and the floor area they contain. Items of interest on structures include: frontage, depth, number of stories, total floor area, use, age, condition and type of construction.

Many services are tied or linked to structures and the floor area they contain. For example, water, gas, electricity and telephone services, and sometimes heat and air conditioning, are delivered by pipe or cable; sewage waste is carried away in pipes; air pollutants are expelled from structures; persons and goods are picked up and delivered to structures, and the floor area they contain, through the transportation network, as is refuse. The point was made earlier that the general average rate for delivery of persons per unit of floor area is fairly stable. It is likely that the quantities of demand for other services correlate fairly well with floor area, again taking general averages; thus, floor area can be useful data for planning these services as well.

Floor area is a good basis for estimating construction costs of either individual buildings, groups of buildings, or all the buildings in a town or city. Construction costs for support services no doubt can be estimated reliably from floor area. If land costs and transportation facilities costs are added, the task of estimating the cost of a whole plan is within easy reach. Total cost estimates should be prepared as a step in plan evaluation and also plan implementation.

If small area employment and population estimates are required for other aspects of the planning program, they could be derived from floor area estimates with a fair measure of reliability. Occupancy rates vary, but these variations affect land-use model results to the same degree whether one derives space required from occupancy estimates (employment and population) or vice versa. Estimates of future employment, based on unreliable existing employment data, may be less reliable than estimates of future employment derived from estimates of future floor area. This could be because data on existing floor area are more reliable than data on existing employment.

It is not suggested that employment and population data and estimates may be entirely neglected, but that floor area can be a good substitute at the small area level. Estimates of future regional total population and total employment, and possibly some components of these totals, would be needed as a basis for estimating future regional totals of floor area. Accordingly, data on existing total population and employment would be needed, but not at a small-area level.

Floor area data are not available in an edited machine-readable form from a single national source as are the census of population and housing data. But machine-readable property tax assessor's records may be adapted to give both land and floor area data. How much adaptation is necessary is, of course, the critical question.

The Tri-State Transportation Commission obtained land and floor area data for New York City from the tax assessor's records. Outside of New York City, use of aerial-photo mosaics and maps was combined with a field-listing operation to obtain land and floor area data. Because of the multiplicity of taxing jurisdictions, each with its own peculiar system of records, it was not feasible to consider using tax assessor's records outside of New York City.

Tax assessor's records everywhere could be easily adaptable for use as land and floor space inventories, if they were standardized. Money to take care of the additional initial costs of standardization and modification, coupled with education of tax assessors as to the needs of land and floor area inventories, would probably do more to achieve this goal than anything else. But even leaving aside potential cost savings from the use of tax assessor's records, which have been standardized and modified to meet the special needs of land and floor space inventories, the cost of a land and floor space inventory is still relatively low when compared to the real costs of the censuses of population and housing, manufacturing, and business. It is estimated that the ratio of costs is approximately 1:20.

In countries where census data are poor or nonexistent, this ratio has fundamental significance. In studies within the United States, the budget is likely to reflect the reverse of this ratio (because the study must bear the full costs of a land and floor space inventory) but is closer to just tape-copying costs for census data; the big costs for census data are borne by the general taxpayer.

The 1:20 cost ratio applies even in the United States, however, when back-dating or inter-censal updating is contemplated. There are greater opportunities for constructing floor area data for past time periods than for employment or population data. Old maps and aerial photos are the prime sources for past floor area data. Old WPA property survey data also have been found satisfactory for small area aggregates.

Sets of very simple longitudinal data, such as floor area, may do more to improve land-use models than elaborate data for one point in time.

Use of aerial photographs, fire insurance maps, building permits, tax assessor's records and field surveys are all feasible methods of updating, depending on the amount of elaboration required. The potential exists for automatic or semiautomatic interpretation and measurement of floor area from aerial photographs by photogrammetric means; updating of floor area data then would boil down to a cost closer to that of "flying" an area by airplane or satellite.

SUMMARY

There are a number of reasons for considering a minimum set of items of information for land-use models. The money available for information is limited. Currently, 30 to 50 percent of the budget and two to three years of the schedule of a new study can be absorbed in building an information base. The land-use model field is young. The direction and extent of changes in the future, with respect to information needs, are not clear. There is the danger that a study may collect information that is never used.

There is great merit in conservativeness in selecting items of information. The need for each item of information should be clear at the outset. This is not easy, especially when a study is not sure at the outset what the land-use model will be like when it is finally built.

There are a number of different minimum sets of items of information. Four were suggested and considered. The fourth set included information on accessibility, vacant land, land-use areas and floor area. This minimum set of items of information has not been tested yet, but it appears to have important advantages over other minimum sets. In particular, floor area has very important advantages. These may be summarized as follows. Recording and measuring of mixed uses are facilitated. The types, intensities, and locations of all land and floor uses can be collected in a single inventory, thus doing away with the problems that appear with attempts to "marry" data from two or more separate sources. Floor area is well understood in land-use planning. The ratio of floor area to land area (FAR) has been used extensively in land development planning and in the subsequent regulation of development in accordance with the plan. Floor area is evidently closely related to measures of travel demand, which makes it acceptable as a base for transportation planning. Floor area is probably as good a base as any for estimation of gross construction costs for capital budgeting or cost-effectiveness studies. Floor area is well understood by land developers, realtors, builders and the general public. It is probably as good a base as any for preparing detailed local area plans from generalized regional plans. It is relatively easy to collect, verify, measure, update and back-date. Its collection is probably more susceptible to mechanization than other types of data. It is not subject to confidentiality strictures because it can be so easily observed. The main disadvantage of floor area data appears to be that it is not readily available from secondary sources at present. It is possible that with the help and cooperation of tax assessors, their records could be modified and standardized for this purpose so that these data would be readily available.

The outputs of land-use models can be useful for many different aspects of future planning—comprehensive planning, transportation planning, land development planning, water and sewer planning, and electricity, gas and telephone planning. Also, they could be used for planning, programming and budgeting systems. Figures 1, 3 and 4 indicate what useful land-use model outputs would look like if represented in three-dimensional diagrams. The fourth minimum data set probably would be sufficient in terms of information for these types of outputs.

The fourth minimum data set has much to commend it, for both short-term and long-term usefulness, and it deserves careful consideration by new studies. Even if this minimum data set is not chosen, the concept is still worthwhile. It can help to reduce the information problems that large studies face. Also, it can help to scale down many other problems. It need not become a negative narrowing device. Easy expandability can be, and should be, a built-in feature of the minimum data set. It should be easy to add experimental pilot data for special research testing needs when these are clarified and judged as likely to be worthwhile.

A Computer System for Urban Studies

WREN McMAINS and AARON FLEISHER, Department of City and Regional Planning, Massachusetts Institute of Technology

•SOCIAL sciences are not generally thought of as laboratory sciences, and social scientists seem content to find their experiments wherever they can. Not being of their own making, such experiments are rarely pointed, and sometimes not even to the point. This disability is so common that it is considered natural and inevitable. It is neither. We do not mean to infer that social science laboratories can attain the elegance, facility and finality of the physical sciences. We do assert, however, that the gap can be closed enough to make the enterprise well worthwhile.

Our purposes will be better understood if we describe first the context in which the system is embedded. We are working toward a laboratory of urban studies, and we intend to equip our laboratory with these facilities:

1. A data library organized for both ease and efficiency of access, search, retrieval, editing and expansion;
2. A library of mathematical programs by which these data can be analyzed and manipulated;
3. A set of graphical procedures by which results can be presented;
4. A data structure by which to describe urban problems and designs;
5. A set of orders and procedures which would help in the design and operation of experiments and simulations; and
6. A repertoire of programs and a facility for programming to help in the analysis, design and evaluation of plans and decisions.

All the facilities will be programmed for a modern digital computer operating as a time-shared machine. A computer operating in this mode allows real-time communication between the user and the system without using more computer time than that required for computation. While the computer is waiting for a request from one user, it can service other users at remote locations. The user can start and direct all operations from his control console. He will be able to examine the data before using them, to look at intermediate steps in the analysis and processing, to edit, modify, reconsider, or resume the process. It will be possible to use almost all parts of the system in a batch processing mode; however, this limits the intercourse between computer and user. This mode of operation is useful for production runs that become routine when the experiment is completely designed. In either mode of operation, these facilities can be used separately or sequentially.

SYSTEM SPECIFICATIONS

From the outset it was clear that the system must meet two prime specifications. Most of its users will know little mathematics and less programming. Moreover, they will be unaccustomed to operating in a context where any mistake, however small, is catastrophic. They are not likely to appreciate that punctuation, spelling, form and order must be precise. They are also likely to be easily discouraged. Therefore, the system must guide its user along his way. It ought not to stop sullenly when it encounters mistakes. The system must be capable of calculating the costs and other ramifications of what it is the user asks for so that the user can decide if he really

---

Paper sponsored by Committee on Land Use Evaluation and presented at the 46th Annual Meeting.

wants what he is about to get. Yet the system must also be capable of responding to the subtleties of the capable programmer.

The second is a specification on the library. We intend a data library whose acquisition policy is the same as that of a university library. The purpose is a complete rather than a critical collection. The library therefore will be large and its users heterogeneous. A computer-operated index and acquisitions system then becomes necessary.

Therefore, we have decided to accept the data in the form that they arrive. No structure is imposed a priori. That option is left to the user. At present he is able to order his data in multidimensional tables, which is a simple kind of list structure. In the future he will be able to avail himself of more elaborate list processes. The index, however, will be highly structured, which is to say that structure will be imposed on the names and characteristics of the data and not on the data themselves. The index will act somewhat as a card catalogue except that it will be capable of far finer discriminations and be more easily modified. The retrieval system is intended to be able to follow all the ramifications contained in the index.

Acquisitions and Indexing in the Data Library

The library must be able to acquire data in any machine-readable form and, if necessary, to edit and correct them. Its capabilities will rest largely in the index provided for search, and the procedures for retrieval and display.

All data files can be accepted and entered as they arrive. Only the mode in which a file is written and the location and kinds of information need be described in order to make the file immediately readable. If it is considered desirable, the file can be rewritten into a form which can be read faster.

A description of the contents of each data file will be entered into the index. The index must list the names by which the variables are commonly known, the geographic places to which they refer, their dates, sources and statistical characteristics. Each category will be listed separately and all of them cross-referenced. In the process of assimilating a set of data into the library and correcting, editing or rewriting them, the index will simultaneously be modified. All this machinery will operate as part of the procedure of acquisition.

Searching the index will proceed by way of a systematic interrogation of the computer. The searcher, for example, may start by asking, "What do you have on income by industry?" How the question is worded will not be critical. The computer may answer by asking, "Where?" The searcher might say "Boston." Then the computer would produce the names and a brief description of the sources containing such data. The searcher would pick what he wants, or ask for further details, or go back to a higher level in the hierarchically structured index.

We are not certain how best to order these questions in logical nets. Rather than work toward some figure of merit given a priori, we propose a subsystem that would observe the use of the index and make improvements a posteriori. Then it would be simple to keep count of the categories searched. This information is important where the library is used routinely. An old fashioned card catalogue of the more frequently requested categories could be a more efficient way to display them. Thus we are not committed to the computer for all purposes.

It would be more complicated, but not very difficult, to record the paths taken through the index. Efficiency would be increased by easing those paths more frequently taken. They could, for example, be replaced by direct links. The index may also become more useful if, to each variable, were attached the names of all the other variables it has been coupled with in the making of tables. Easing paths through the index and augmenting the list of cross referents can be changes that the system itself would work, in which case one could say that the index "learns" and incorporates its efficiences.

The Retrieval Procedures

For our purposes, retrieval requires programs that call information from one of several very large files of surveys, such as the individual responses from a transportation study or the one in a thousand U.S. Bureau of the Census tape. Because the files are so large, speed and efficiency are extremely important, and for this reason it may be necessary to rewrite some files into a form which can be read faster.

We have decided on table making as a basic device for data retrieval. To obtain a table one needs only to list the variables as they are named by the index and order "tabulate." The system is presently capable of producing tables of 35 dimensions. That seems large enough. New variables that are transformations of the old ones can be created and used when making tables. The tables can be printed, displayed visually, or stored for future reference and manipulation.

Tables will not be the only mode of presentation. Curves and histograms will also be possible choices. And since urban problems frequently turn upon patterns of location, we can also display data values on geographical maps.

The Manipulative Procedures

We shall have on immediate call a repertoire of programs to operate on the data files that the user assembles in the form of tables. These are the kinds of manipulations we have in mind:

1. Arithmetic;
2. Function generators: logarithmic, exponential, polynomial, trigonometric, etc;
3. Matrix algebra;
4. Statistical analysis: moments, such as means and variances, tests of hypotheses, limits of confidence, and measures of significance;
5. Regression: simple, multiple and simultaneous; principal component, factor and discriminant analysis;
6. Curve fitting;
7. Graphics: data maps, curve tracing, histograms;
8. Mathematical optimation techniques (including linear programming); and
9. Et cetera.

We mean the "et cetera" literally. The system shall be devised for the easy integration of any program, prepackaged or specially tailored, or for the immediate addition of a datum without necessarily having to go through the library routines.

The results of such additions and manipulations can be retained in the private file of the user who makes them, to be called on as he wishes, or they can become a part of the public files of the general library by routing them through the acquisitions and indexing subsystem. The library in all its parts, therefore, will grow by its use.

Simulations and Experiments

A laboratory cannot be made from a set of experiments prepared a priori. It must be able to accommodate easily at least a wide range of purposes. We meet this requirement in two ways. It will be possible to write computer programs within the system, and possible to fill prescribed data structures from the files in the library.

Simulation is a kind of experiment. We have singled it out for separate mention because it is peculiarly a product of the computer's capabilities. We will include the basic structure for many simulations and the means for linking new simulations to the existing system.

The experiment is the user's choice, but we anticipate that the following kinds may be of general interest.

1. The computational machinery for making analyses of urban and regional economies, such as sector, economic base, and input-output.
2. The analysis, allocation and assignment algorithms in transportation networks. In particular we shall further explore aggregative behavior as a function of individual utility.
3. The inverse power and exponential allocation rules for the distribution of travel and land-use assignments.
4. Optimal location and trip distribution as obtained from linear programming formulations.

Analysis, Design and Evaluation of Decisions

At present, most of the time we have allocated to this matter is occupied with the analysis and evaluation of the visual qualities of physical designs. For this purpose we are learning to program the computer to act as a hemi-demi-semi literate draftsman. To do this in a way that is most convenient for a designer, both the input and the output of the machine must be in the form of drawings. That is the technical problem. The analytical problem requires a decision on what kind of drawings to make. We have tentatively settled on sequences of perspective taken along a specified path through the design. These perspective drawings will be the raw material for studies in perception and visual quality.

We have also begun to work on the problems in the processes of design itself. The specific problem we have posed is to devise methods by which to construct sets of rules (i. e., algorithms) that would replicate the product of design. An analysis of the processes of design would come out of the analysis of those sets of rules that are successful replicators. Concentrating on replication is an indirect route to examining the process of design, but it is one that has the advantage of relatively clear verification. It has been chosen for this reason.

THE PRESENT STATUS OF THE SYSTEM

Let us now consider in more detail the system as it exists today. First, we shall give examples of some of the existing requests and their uses. Suppose a person wanted to find information on mobility in the Boston area—in particular a table showing length of residence vs the family income level. The commands would be these:

```
USE/FILE     BOSTON/HOUSEHOLD
TABULATE     MOBILITY     YEARS/AT/ADDRESS.VS.INCOME/LEVEL
EXECUTE
R/P/PRINT    MOBILITY
C/P/PRINT    MOBILITY
```

The "USE/FILE" request tells what data file is to be used ("**BOSTON/HOUSEHOLD**"). The "**TABULATE**" statement orders a table where the rows are length of residence and the columns are different income levels, and calls this table "**MOBILITY.**" The "**EXECUTE**" request says that we have specified all the tables required at this time, go ahead and make them. The last two requests say that we want the table printed twice, first as percentages of row totals and then as percentages of column totals.

If we are now interested in comparing these results for the region with a specific area in the region, we might write these commands:

```
USE/FILE     BOSTON/HOUSEHOLD
COMBINE      YEARS/AT/ADDRESS  0-1, 2-5, 6-10, 11-99
COMBINE      RESIDENCE/ZONE-INTO-DORCHESTER  61, 62, 63, 64, 65, 107
COMBINE      RESIDENCE/ZONE-INTO-NEW/RES  61-65 + 107, 1-60 + 66 - 106 + 108 - 626
TABULATE     MOBIL/I     DORCHESTER.VS.YEARS/AT/ADDRESS.VS.INCOME/LEVEL
TABULATE     MOBIL/2     NEW/RES .VS. YEARS/AT/ADDRESS .VS. INCOME/LEVEL
EXECUTE
PRINT        MOBIL/1
R/P/PRINT    MOBIL/2
PRINT        MOBIL/2
```

We are using the same data file as before, but this time we have decided to change (or combine) "**YEARS/AT/ADDRESS**" into a variable with the same name and four levels, or lengths of residence, instead of one hundred. We recombine the variable "**RESIDENCE/ZONE**" into two new variables. The first is called "**DORCHESTER**" and will have six levels, or zones in this case. The second, called "NEW/RES", will have only two levels—one containing all zones in Dorchester, the other containing all other zones in the region. We then ask for two 3-dimensional tables to be made. The first will be printed as six 2-dimensional tables with the four new levels of "YEARS/AT/ADDRESS" as rows and the income levels as columns. (There are six of these tables because there are six levels of "DORCHESTER.") The second table, which is really two 2-dimensional tables, will be printed twice, once giving percentages of row tables, and again giving absolute values.

Because it is an on-line system, the user will get almost immediate response in the form of advice, reprimands, and rewards. The system is able to inform the user of the consequences of his requests (amount of information that will be generated and the cost) before executing the requests. This will limit the amount of useless information generated by misunderstandings on the part of the user. The reprimands administered by the system are gentle and meant to show the user the error in his ways.

The critical time in a system of this type is the time required to operate on each logical record, not the time used in reading requests. The user speed is very slow relative to the time required for interpreting his requests. Some of this waiting time, therefore, is used to set up very efficient programs which will be executed as the data is read (remember that many tables can be made in one pass of the file). The marginal cost of the extra tables is very low because the system is input/output bound during this phase, unless a very large number of tables are requested. This is accomplished by double buffering, which allows the system to operate at read speed most of the time.

The following operations are performed as each logical record is read: (a) unpacking and placing the data in the variable list; (b) combining a variable; (c) arithmetic operations on variables; (d) user written subroutines; and (e) incrementing the tables. From these, programs are compiled to perform a and c as the requests are made; b and e are accomplished by pre-programmed routines which are supplied parameters and parameter lists indirectly via the requests; whereas d makes use of routines compiled by other compilers available for general use via the time-sharing system (e.g., MAD, FORTRAN, ALGOL, assembly language, etc.). We supply no general compilation features—this has been done very well by others—but we do supply several special purpose compilers indirectly.

Once tables have been made, they can be added, subtracted, multiplied and divided element by element by other tables. They can be restructured, or saved for future manipulations. They can be printed as absolute values or as percentages of totals.

Rows, columns, planes, etc. are labeled as they are printed—if labels are available—otherwise, numeric labels are used. Labels contained with the file are always available automatically. Labels for variables created by the user can be entered by the user before the table is printed, and will be used in every table in which this variable occurs.

The tables can be written on the disk for delayed printing, or saved on the disk in their internal form for later use. It is also possible for the user to save the entire system in its present state so he can return to it at another time.

The system can handle its own allocation of table storage space, or it can be assisted by the user, thereby taking advantage of his knowledge of future requests. An N-dimensional table where the i th dimension has n_i levels requires $\prod_{i=1}^{N} (n_i) + N + 1$ locations. The N + 1 locations contain pointers to names and labels, and dimensionality information. All of the tables are stored in core if room is available. If not, then room is made by shifting tables which are not required at the moment to secondary storage. Under user control, the user is informed of the overflow and he then has several options: (a) deleting tables for which he has no further use; (b) moving tables to permanent secondary storage (leaves tables in internal form, but saves them for use on another day); (c) moving tables to temporary secondary storage (same storage used in automatic mode); and (d) continuing in either manual or automatic mode.

Automatic rearrangement of storage is done only when necessary. Sometimes it will only involve primary storage. However, user initiated allocation can be done at any time. The user can also perform the same operations on labels which are intermixed with tables in storage.

There are three acquisition policies that can be followed.

1. Adding information about the contents of the data file to the index. This involves writing what we call a dictionary for the new file and supplying it to the system for the purpose of updating the index.
2. Writing the dictionary as in 1, but this time letting the system rewrite the tape into a form which can be read faster at the same time it updates the index.
3. Adding the file without a dictionary. This would require the user to write a dictionary for any variables used each time he came to the system (the individual user could save his own dictionary parts for future use). However, this would require another user of the same tape to duplicate efforts and most important would not allow for updating of the index which makes it difficult or impossible for future users to find the data. This type of action only makes sense in the case of a private file which is not to be added to the library and is to be read only once or twice.

The normal acquisitions procedure will include writing the dictionary and rewriting the file into a packed binary form. This can be done automatically by the system once the librarian has supplied the dictionary, and at very little added cost if it is done at the same time as the initial error checking. It is much faster to unpack a word into several parts than it is to read even a single additional word. The time saved by reading fewer words and not having to convert BCD to binary on each pass of the file will make up costs of the original rewriting and error checking in approximately three passes of a tape file.

Once a file has been added to the library, the user refers to the variables by name. He need never concern himself either with the location of the variable in the logical record or its packing. These are taken care of by the file dictionary, which is added to the system by the librarian in much the same way a card for a new book is added to a catalogue.

The cost of modifying the index structure is very low compared to the cost of modifying a data structure, and in fact the index is modified continually and automatically (heuristically) as determined from the present uses of the system.

The especial value of our computer system does not reside in its parts. We think that those we made work well, but they are not very novel. The system is important because it implements, not the application of the computer to the specifics of research problems, but the availability of the computer as a general instrument in a research laboratory. This difference is large and rests in the specifications of availability and instrumentality. Consider the automobile. It is immediately available for a wide variety of purposes, and one can be a good driver without understanding anything of its operating principles. The computer in a research laboratory should do as well, as easily, as often.

Availability of Census Data for Urban Areas

MORRIS H. HANSEN and ROBERT B. VOIGHT, U. S. Bureau of the Census

This paper discusses current research and development under way and planned for the future to provide census information covering the social and economic characteristics of the population and its housing to be collected in the 1970 decennial Census for a variety of small geographic units within urban areas. The potential flexibility in aggregating the data in spatial units of significance for local urban planning is outlined. The possibilities of associating a grid coordinate system with the data to permit analysis by uniform land patterns, and the significance in distance relationships between specified areas or from specified points, are explored. The paper summarizes current proposals for new types of data of significance to urban planners which may be made available from the 1970 Censuses as well as new ways of making the data available in more convenient form via computer tapes as well as through more timely publications systems.

Proposed developments in the 1967 Economic Censuses of Business and Industry, Transportation and Construction as they relate to the data needs of urban areas are discussed.

•AS our society becomes increasingly urbanized, the Bureau of the Census in its role as a producer of basic data is devoting increasing attention to the information needs of those engaged in dealing with this social and economic transformation.

A quick review will show that as the urban population increased from about 50 percent in 1920 to almost 70 percent in 1960, the Bureau of the Census identified metropolitan districts (the thickly settled territory in and around cities of 50,000 or more) and published data for them in 1930. They also expanded the data published for the standard metropolitan statistical areas as established for 1950 and 1960, and initiated the publication of data for the urbanized areas surrounding places of 50,000 or more in 1950. The Census tract program was initiated in 1910 when tracts were tabulated for 8 large cities. Publication of data for these small homogenous areas was expanded from 70 cities in 1940 to all of the central cities in the 212 standard metropolitan statistical areas in 1960. In 1960, 133 of these standard metropolitan statistical areas, representing 59 percent of the total population of the country, were completely tracted and the tract data published. Since its initiation in 1940, the Census of Housing, in addition to tract data, has provided data for individual city blocks in places of 50,000 or more.

Additional aids to urban planning and administration, transportation systems, business and industry marketing, and facilities locations have been made available by the publication of economic census data for central business districts in cities of 100,000 and over. Such districts are defined as one or more census tracts comprising an area of very high land valuation, characterized by a high concentration of retail business, offices, theaters, hotels, and services where there is high traffic flow. Central business districts were first defined for use in the 1954 Census of Business. In the 1958 Census of Business publication program, additional small area detail was provided for

---

Paper sponsored by Committee on Land Use Evaluation and presented at the 46th Annual Meeting.

109 central business districts and 472 other major retail centers in 116 standard metropolitan statistical areas. The latter areas are defined as concentrations of retail stores which include a major general merchandise store. They include not only planned shopping centers but also the older string street and community developments. Where a planned shopping center is involved, the boundaries include, in addition to the center, all adjacent blocks having at least one store of the general merchandise, apparel, or furniture-appliance category. In the latest Census of Business, data are provided for 131 central business districts and for nearly 1000 major retail trading centers in 116 standard metropolitan statistical areas.

PLANNING FOR THE 1970 CENSUS

Statistics for small areas are a unique contribution of a census, and this attribute is receiving special attention as the Census Bureau prepares for the 1970 Censuses of Population and Housing. In a series of 22 regional meetings held with groups of representative users of census data some time ago, the dominant expression of interest centered on data for small areas and reflected local interest in the many legislative programs relating to economic opportunity, economic development, transportation systems, housing, and health programs.

The Census Bureau is committed to providing the maximum informational backup to these programs within its budget limitations and legal authorizations. As the planning for the 1970 Censuses proceeds, the needs of the potential users have been solicited from advisory committees to the Census, other federal agencies, the Federal Statistics Users' Conference (composed of members from business, industry and labor), academic bodies, and state and local officials. In an effort to provide maximum benefit from the Census program, improved methods of data collection and processing are being developed, more advanced data processing equipment is being developed and acquired, and more powerful programming systems are being constructed.

The taking of a census involves the construction of a major systems design, and the 1970 design is the result of extensive and continuing research efforts. The basic new design decision from which the subsequent steps will be taken is the use of a mail-out-mail-back enumeration (with enumerative follow-up as necessary) in much of the nation.

A mail population and housing census requires a complete file of residential addresses coded to reflect the geographical detail to be observed in the census. A major advance of the 1970 Censuses will be the geographic coding system (being developed), which will permit all addresses located on the specific sides of a city block to be accounted for. To achieve this spatial accuracy, a map-making project to produce maps showing reasonably accurate detail covering the urbanized areas has been started. The gain will be especially important in suburban areas where available maps have been particularly deficient. The maps being prepared are based on copies of U.S. Geological Survey 7½-minute quadrangle maps and enlarged from a scale of 1 in. equals 2,000 ft to 1 in. equals 800 ft. The topographic detail is dropped and the street layouts are updated through cooperative efforts worked out with local planning groups. The Bureau is reciprocating for this help by providing copies of the updated maps. This effort is now about half completed. The maps, when finished, will cover about 100,000 sq mi of urban territory (1).

Geographic codes in the major urban areas will be assigned by computerized address coding guides, which will include street names, block-face identification, intersecting streets, the range of address numbers for each block face according to census tract and zip code area, and the area identification codes required for Census Bureau tabulations. An optional code field is provided on worksheets which local cooperating groups may use as a key to relate Census data to their particular local administrative or operating areas. However, the block-face coding guides will provide great flexibility in making it possible to tabulate census and other information for any desired area, whether or not they are pre-identified in the optional coding field. Both computer tapes and printouts of these address coding guides will be made available along with the Census maps, as tools for small area identification.

In addition to a coding guide that will identify an address recorded in a specified way to a small area, procedures and programs are being developed for taking uncontrolled address information (in which there may be variations in the forms of recording the address, such as in spelling and in formats of address recording) in machine readable form and identifying and associating the principal address elements with those in the address coding guide. Addresses can be recorded in many ways, and some manual intervention will be necessary for situations not adequately covered in the computer programs. The ability to code uncontrolled addresses will facilitate associating address information in various administrative record systems with the reference guide, and thus the tabulation of local records and census information for the same small geographic areas.

Through the use of address coding guides, the Census will be able, for the first time, to record information for geographic units ranging in size from one side of a city block to an entire urbanized area. Tabulations will be possible, for example, for both sides of a street or streets through a city or area. The limit to the flexibility of the information available for various areas will be disclosure rules and the cost of tabulation. The Census Bureau does not plan to provide this capability for all city delivery areas, but it does hope to accomplish this for entire urbanized areas and for cities of 25, 000 or more inhabitants (not all of which may be possible), or for the bulk of such areas. It may be necessary to cut back on some of the smaller cities and urbanized areas. A further limitation is the extent of city delivery postal service; beyond these areas the Bureau does not plan to code to the block-face level, although reporting by block is expected to be feasible to the boundaries of urbanized areas. In any event, census data will be available in far more geographic detail than ever before.

With little added effort, a copy of the "Census" address coding guide for an area can be modified locally for broader use by the addition of identification codes for areas such as police precincts, health areas and so forth. With this accomplished, local flexibility is virtually unlimited. Traffic information, for example, can be matched to the modified coding guide and the traffic data assigned not only to traffic zones, but simultaneously to health, school and other areas, as well as to census tract and block. The same can be done with other local information.

It is anticipated that this development in data flexibility, which will provide a standard set of small geographic bits as building blocks in assembling data in virtually unlimited types of areas, will be one of the major contributions of the 1970 Censuses to planners in all fields.

Coordinates

Another possibility of considerable importance in certain fields of interest stems from our proposal to identify the locations of blocks or block faces by coordinates. Although this is not a certainty, it is definitely planned. In any event, the system will be so designed that coordinates can be introduced later if resources are not available to introduce them into the system prior to the Census.

Within the areas covered by address coding guides we expect to have coordinates for block faces; for other parts of urbanized areas, coordinates for blocks. For the smaller cities outside of urbanized areas, and for rural areas, coordinates probably will be established for "standard locations" consisting primarily of census tracts or minor civil divisions in the rural areas. The coordinates will be recorded in degrees of latitude and longitude to four decimal places, i.e., to 36 ft at most, but those who wish to employ state plane or other standard coordinates, rather than latitudes and longitudes, will be able to convert them.

This program opens up a whole area of data availability and analysis heretofore not attainable. Spatial relationships of social and economic data can be examined, density and distance correlates established, and statistical aggregates established in terms of concentric circles from a given point, in equal squares of certain size, or other ways. The characteristics of people and housing for a certain distance along each side of a proposed highway or freeway right-of-way can be examined in considerable detail to determine the dislocation impact of alternative routes. Many other important uses will emerge only after considerable exposure to this facility after the Censuses are taken.

NEW HAVEN CENSUS USE STUDY

The opportunity is now available to make important advances following the 1970 Census to serve the evolving modern needs of numerous federal, state, local and private programs in urban areas that depend on Census statistics for analysis and for planning. Steps by the Census Bureau to meet the data needs of urban areas in the 1970's may also, as a by-product, make a significant contribution toward standardizing the definition and format for local data, a much desired objective emphasized by many public officials and students of the urban scene. Comparative analyses of activities and changes in one urban area as against others depend on local statistics more compatible in definition and form than are now generally available. Because the Census program is national in scope and because Census data occupy a central position in the information needs of many urban planning programs, it is reasonable to expect that a general framework developed to allow integration of Census data with other locally generated data will act to encourage greater uniformity in the coding and definition of local data in urban areas (2).

To test the feasibility of these Census spin-offs of data availability, to explore the potential uses of small area data from the Census and other sources under live conditions, and to prepackage certain programs determined to be of demonstrated value in various urban areas so that they may be available for use immediately following the 1970 Censuses, the Bureau is embarking on a small area data research and development study.

To work toward the accomplishment of these goals, a development project is being undertaken with the cooperation of interested federal, state and local agencies in the New Haven, Connecticut, Standard Metropolitan Statistical Area. In April 1967, the Census Bureau was to conduct a pretest of the procedures developed for a mail-out-mail-back population and housing census in 1970.

New Haven and Connecticut officials have expressed their willingness to cooperate in a small area research and development study based on the results of the Census pretest and subsequent surveys in the area. This project will have as its function the following:

1. Explore both the current uses and likely future needs for Census data in existing local, state and federal programs.
2. Develop systems that will allow efficient merging of Census statistical aggregates with other local and state data to meet the needs of specifically defined programs. Studies involving the matching of individual census and local records may be made also in ways that fully protect the confidentiality of the census and other data, and make information available in the form of statistical aggregates.
3. Investigate the benefits of cooperative data collection between Census and other local, state and federal programs. For example, a study will be made as to the feasibility of coordinating the regular population and housing censuses with the origin-destination studies made by or for state highway departments, and of using Census data to meet part of the needs of such surveys.
4. Investigate the level of detail and the form in which Census data should be made available to local users. Included also will be a study of the form in which Census data can best be made available on magnetic tape.
5. Develop a package of programs for use by local communities to allow rapid conversion of periodic Census tabulations into information (e.g., summary reports, charts, and possibly map-type displays) useful for local analysis.
6. Analyze the results of the demonstration for potential strategies to be incorporated in the 1970 Census plans and in local community programs to take advantage of Census and other information.

A number of federal departments are participating in the New Haven Census Use Study, including the Department of Transportation, the Department of Housing and Urban Development, and the Department of Health, Education, and Welfare. Of specific interest is the planned cooperation of the Bureau of the Census and the Bureau of Public Roads to test the feasibility of conducting by mail an origin-and-destination survey in a sample of households designated at the time of the pretest Census. The sample households will receive the travel questionnaire shortly after the Census enumeration. Thus, timely social and economic factors obtained in the Census may be readily merged with the travel data to provide statistical aggregates of social factors associated with travel characteristics in and between the traffic zones established in the standard metropolitan statistical area. Also of interest will be the attempt to compare the travel study results with the journey to work information obtained in the Census for both the households included in the travel sample and the entire Census 25 percent sample. These feasibility tests may point the way to achieving substantial economies in the travel surveys, which are called for by the Highway Transportation Act every 5 years in all standard metropolitan statistical areas. The tests may also provide the means of achieving greater comparability among all standard metropolitan statistical areas for future decisions regarding urban transportation planning on a nationwide basis.

Since the New Haven Census Use Study Office was opened in September 1966, the staff has been conducting an inventory of the data needs and interests of the various local and state agencies whose programs can be assisted by the provision of Census data for the particular geographic areas they observe in their operating programs. Whenever these interests seem to be representative of those in all urban areas, it is our intention to test the feasibility of building tabulation programs to provide the required data and evaluate their usefulness. If the evaluations demonstrate that they are effective, we hope to provide standardized programs for such tabulations to any community which calls for them following the 1970 Censuses. Lack of sufficient time and money no doubt will preclude the possibility of exploring a wide range of potential programs, but we hope to test several important possibilities.

The potential yield from these experiments will be of substantial interest to a number of federal agencies engaged in the support of various urban development programs. To the extent that we are able to demonstrate the direct effectiveness of the small area data programs, which we expect to develop and test over the next year and a half, a significant cost benefit should accrue to many urban areas.

NEW DATA DEVELOPMENTS FOR 1970

At this point it may be appropriate to outline briefly the new data developments contemplated for the 1970 Census. Whether all these can be brought into being in time for use remains to be determined. In recognition of the needs expressed by national organizations as well as local interests, an effort is being made to assure that the entire area of all standard metropolitan statistical areas will be tracted in time for the 1970 Census. In the few instances in which there is insufficient local interest and tracts are not defined by the usual methods through local initiative, the Bureau plans to develop them on its own. It is expected that in 1970 the number of tracts which will be recognized in the Census publications will be about 33,000 in approximately 230 standard metropolitan statistical areas.

We have also heard from people in many places smaller than 50,000 asking that their cities be provided with census statistics on a tract basis. The Bureau has announced that it will recognize tracts which are established in these cities through local initiative. Recognizing the tracts means that we will tabulate statistics by tract, but it does not mean that we will promise to publish the tabulations for these smaller areas in the regular Census reports. Nevertheless, the unpublished tabulations will be available for local use or publication under local auspices (3).

Housing census results have regularly been published for city blocks. Limited statistics were issued for city blocks in all cities having a population of 50,000 or over prior to the 1960 Census. In addition, the Bureau had announced that other communities wanting statistics by city blocks could arrange to have them if they would prepare the necessary block identification materials and reimburse the Bureau for the added costs. Block statistics, including a limited number of housing items and the total population, were published for nearly 750,000 city blocks. In 1970, we hope to extend the block reports to the closely built-up areas surrounding cities of 50,000 and over, i.e., we hope to include the entire urbanized area. An attempt will also be made to provide block statistics for cities with a population of 25,000 to 50,000. If these additions can be effected, the total number of blocks is likely to be on the order of 1,600,000; roughly twice the number for which reports were issued for 1960. It should be stressed at this point that these expansions of the block data are hoped for; however, it cannot yet be stated with assurance that resources will be available such that these hopes can be fully realized.

Many suggestions have been received for modifications in the content of the Census questionnaire. The Bureau, with the aid of its various advisory committees, has been evaluating these suggestions and, although final decisions have not yet been reached, the content of the New Haven Pretest will, in many respects, be an approximation to the final questionnaire content which must be frozen in early 1968.

There is a clear call for greater detail on place of work. If it is possible to secure reasonably accurate identification of places of work by street and number, as in the case of residence, the coding of work place to block faces to be aggregated by small areas as desired will be technically feasible. This, coupled with information on methods of transportation used to go to work, should provide information of considerable interest to many agencies.

Migration during the last 5 years conceivably could also be studied in terms of small areas within the city or the metropolitan area by the use of the coding of detailed addresses. It has been proposed that a social security number be added to facilitate comparison of records in successive censuses, as well as the matching of records that are available from other agencies. The result would, again, be statistical tabulations that preserve the confidentiality of the census and other information. The request for a social security number will be included in the 25 percent sample in the New Haven Pretest Census to test this possibility.

Other significant content modifications introduced for the New Haven Pretest and being considered for 1970 include the following:

1. Birthplace of parents—this has been expanded to include the specific states in which the father and mother were born.
2. Major activity 5 years ago—working in armed forces, looking for work or on layoff, keeping house, going to school, etc.
3. School attendance—a distinction is made between parochial and other private schools.
4. School attainment—in addition to the provision for designating years of schooling through 6 years of college, the highest college degree received and the major field of the highest degree is requested. The respondent is also asked to indicate if he has completed a vocational, business, or technical training program before or after 1960, if he has ever completed apprenticeship training leading to journeyman status before or after 1960, and if he is using any of these types of training in his current job or business.
5. Employment status—the street address of place of work and zip code has been added to the usual work place identification items to permit finer detail for small areas.
6. Journey to work—the mode of travel category has been expanded to indicate if the worker is a driver or passenger in a private auto.
7. Occupation—in addition to the kind of work, the person is requested to describe his most important activities and duties and give his job title to improve the occupational classification problem.
8. Type of employment—a distinction is made regarding type of government employment as to whether it is federal, state or local, and the self-employed are asked to indicate whether or not their business is incorporated.

9. Income—a separate entry is provided for showing net income from farm operations.

10. Income other than earnings—rather than a lump sum entry, the person is asked to show separate amounts received from social security payments; the total from interest, dividends, and rentals, trusts, etc., indicating whether interest on savings accounts, dividends from investments, and rental or other income is included; the amount from public assistance or welfare payments; the amount from pensions, government or private; and amounts from other sources indicating if it is unemployment insurance, veterans pension, workmen's compensation, illness or accident benefits, and any other regular payments received.

A change in the housing census will be in the method to be used for measuring housing quality. When the enumerator visited the household in the 1960 Census, he classified the housing unit as sound, deteriorating, or dilapidated on the basis of observation. Because this evaluation involved difficult judgments that could not be consistently applied, a high proportion of the units were inaccurately classified. Although the statistics were reasonably good for comparing fairly large areas, at the block level where statistics represented the work of a single enumerator the errors were so great that they impaired seriously the usefulness of the statistics. The 1960 statistics as a whole understated the number and proportion of dilapidated units. Tests conducted recently to see if householders could report faults in their living quarters by means of self-enumeration showed that a serious undercount of dilapidated units would result if this procedure were used. For 1970, the Bureau is developing a quality rating from a combination of items such as rent or value, heating equipment, kitchen facilities and number of closets. This rating would be shown with data on plumbing facilities, which are also useful in determining the quality of housing.

The Bureau expects to provide two more categories in the value of property questions to identify the $35,000 to $50,000 group and those of $50,000 or more. Additional information will be available on families living in high-rise apartments because the respondents will be asked to indicate whether they live in buildings having 13 or more floors. This should provide another dimension to the study of potential traffic and parking densities in small geographic areas.

NEW FORMS OF DATA DISSEMINATION

We should like to make a few remarks concerning new ways, new forms, and new methods of making Census data available to users in the future. As some of you may recall, the first extensive use of the computer in mass data processing and summarization took place in the 1950 Census when some of the information was processed by the Bureau on UNIVAC Number 1. All of you are aware of the advances in technology, the number of computers now operating, and the many statistical applications currently being made by computers. We have recently taken a number of steps in an effort to meet the needs that have been expressed for fuller access to information in machine-readable form (4).

In the 1970 Census, the Bureau may develop standard packages for which data will be available and for which computer programs will have been written. Mainly they will be for local areas. A local group that uses this package will get it at a considerably lower cost than if it were to request special tabulations that required new programming.

We have made available on magnetic tape several unpublished (and some published) tabulations from the 1960 Census and certain others, and we have published detailed descriptions of them. A complete inventory of machine-readable data and selected special tabulations through 1964 has been published as a supplement to the Bureau of Census Catalog.

We have begun development of a small staff to:

1. Explore uses and needs and to take the leadership in resolving problems, develop new policies and programs, and improve the Bureau's ability to supply data in the most efficient tape configurations.

2. Provide guidance and assistance to users in order to streamline procedures for responding to special requests. (We expect, however, that the consumer would still have direct communication with the subject specialists, because this new function would not take over the work of the subject divisions. It may help in facilitating the identification of sources of information and perhaps help remove road blocks that might otherwise arise.)

3. Advertise the availability of information in machine-readable form and study the market. The confidentiality of individual returns will, of course, be maintained.

4. Develop and sponsor proposals for putting machine-readable records in more accessible form including carrying of corrections into the records, and preparation of standardized programs for tabulation, summarization and analysis. This would include, to the extent that the demand exists, machine-readable equivalents of published information as well as unpublished tabulations. The Bureau will prepare for much more use of computer tapes after the 1970 Census.

5. Develop proposals for obtaining staff and developing procedures for accomplishing special services so that such work would proceed more nearly in parallel with other standard operations. Obviously, priority questions would remain, but we hope many delays and problems previously encountered by the consumer would be avoided or reduced.

Other Small Area Data Resources

The Bureau also makes significant contributions to small area data needs from the Economic Censuses and several of its current surveys. The following developments are illustrative of the Bureau's efforts in this direction.

Beginning with the 1964 data, the County Business Patterns reports for individual states are being issued annually. These publications present county-by-county statistics, as well as standard metropolitan statistical area and state totals by kind of business or employment, taxable payrolls, and the number and employment size of reporting business units. These data are shown for detailed kinds of business in agricultural services, forestry and fisheries, mining, contract construction, manufacturing, transportation and other public utilities, wholesale trade, retail trade, finance, insurance, real estate and services.

In the Census of Business taken every 5 years, the Bureau provides data on the number of establishments, sales for the year, weekly payroll and the number of employees in retail, wholesale, and service trades for states, standard metropolitan statistical areas, and counties, and for cities of 2,500 or more for retail and service and cities of 5,000 or more for wholesale statistics. Additional geographic detail, e.g., Census tract statistics, may be obtained by special tabulations made on a reimbursable basis. Limited retail sales data are also available each month for the 20 largest standard metropolitan statistical areas, and the five largest cities based on monthly sample surveys conducted by the Bureau. In addition, statistics on total sales of department stores are published each month for approximately 130 standard metropolitan statistical areas, close to 50 cities, 10 central business districts, and some 15 other local areas.

Reports are compiled from the 5-year Census of Manufactures providing data on value of shipments, value added by manufacturing, employment, payrolls, man-hours, new capital expenditures, inventories, and number of establishments for all industries in counties, standard metropolitan statistical areas, and cities of 10,000 or more. A Census of Manufactures report series, which may be of particular value to state, regional and local planning, is "Location of Manufacturing Plants." These publications provide the count of establishments in 7 categories of employment size in each of 450 industries. One series shows the data for each industry by county, and a second series presents the data for each county by industry. Limited industry data are also available each year for standard metropolitan statistical areas, from annual sample surveys of manufacturing establishments conducted by the Bureau.

The 1967 Economic Censuses

In the 1967 Economic Censuses for which intensive preparation is now under way, some new developments will probably be undertaken for retail trade activities. For example, providing that the feasibility of reporting is confirmed, there will be inquiries on under-roof floor space and on under-roof selling space.

Subject to availability of funds, tabulations will be made for a number of major retail centers located in standard metropolitan statistical areas beyond those covered in the 1963 Census, and possibly for some very large retail centers outside of standard metropolitan statistical areas.

Census of Construction

The Bureau is planning to expand the 1967 Census of Business to include the construction industry. The last comprehensive statistical effort to collect data describing the inputs, outputs, and structure of the construction field was made in the form of a construction census 26 years ago. We expect to provide a general statistical description of the structure, composition, and activities of the construction industries based on information from a sample of establishments engaged in all types of public and private construction. This will include residential and nonresidential building, highways and dams, new construction, maintenance and repairs, general contracting, special trades subcontracting, and contract construction as well as speculative or operative construction.

It is anticipated that the published results will contain tabulations and cross-tabulations relating to the following:

1. The primary and secondary kinds of business activities undertaken by construction establishments.
2. The geographic distribution of construction establishments by various characteristics.
3. Size distributions of establishments by employment, total business receipts, and construction receipts.
4. Expenditures by the industry for payroll, materials and supplies, capital equipment, and for the rental or leasing of machinery and equipment.
5. Types of construction work undertaken by establishments including the location (state) of activity, whether publicly or privately owned, and whether relating to new construction or maintenance, repairs, additions, and alterations on existing structures.
6. Distribution of establishments showing the extent of subcontracting.
7. Although the extent of geographic detail is not yet determined, we anticipate that data will be shown for the majority of the standard metropolitan statistical areas as well as states, divisions and Census regions.

Census of Transportation

Most of the data collected in the Bureau's first Census of Transportation covering 1963 have now been published in four separate report series.

The National Travel Survey produced statistics showing national and regional passenger transportation patterns for 1963, and their relationship to socioeconomic and geographic factors. The tables contain comparative data on trips and travelers, by purpose of trip, means of transportation, duration, origin, destination and distance of trip, size of party, and kind of lodgings. Travel is presented for households in such measures as income, occupation, and employment status.

The Truck Inventory and Use Survey yielded data concerning the Nation's trucking resources, such as the number of trucks classified by physical characteristics, occupational use, rough measures of the intensity of vehicle utilization, and geographic distribution of vehicles by states and regions.

The Commodity Transportation Survey provided data for 1963 concerning the physical and geographic distribution of commodities shipped by the manufacturing sector of the national economy. A series of 24 reports in the Shipper Series presents data on the flow of 24 groups of manufactured commodities from area of origin to market or redistribution area in terms of tons and ton-miles according to means of transport and distance of shipment.

Twenty-five Production Area Reports are available from the Commodity Transportation Survey. These give information on the flow of traffic from manufacturing plants located in each of 25 selected production areas composed of one or more adjacent standard metropolitan statistical areas. The reports provide statistics on total tons and ton-miles of commodities shipped from each production area classified by means of transport, distance, and destination. The designation of the production areas makes possible comparisons between traffic flow data and economic and demographic statistics available in other Census reports.

The Bus and Truck Carrier Survey supplied statistics concerning for-hire carriers that are not subject to economic regulation by the Interstate Commerce Commission. These data are comparable to the primary statistics obtained by the Commission from carriers under its jurisdiction. The information from the two sources—the Interstate Commerce Commission and the Census—represent essentially the national universe of for-hire motor carriers of persons and property.

Present plans for the 1967 Census of Transportation call for covering essentially the same subject matter area as the 1963 program with the added potential of measures of change between the two Censuses. On the basis of the 1963 experience and improved processing techniques, we expect to complete the 1967 program in about half the time.

The 1967 Census of Governments

The quinquennial Census of Governments to be conducted in 1967 will supply statistics on finances for a variety of small areas including revenue, expenditures, debt and financial assets for school districts enrolling 3,000 or more pupils, for county governments, and for municipalities and township governments having 10,000 or more inhabitants. In addition, findings will be published for standard metropolitan statistical areas and their component counties showing the numbers of local governments by type and size, local government employment, and local government finances in considerable detail. Data will be included on full-time employment of state governments in standard metropolitan statistical areas, and on selected items of direct state expenditures in these areas. Key figures on employment and finances will be shown for each of the 18,000 local governments within the standard metropolitan statistical areas.

Data Guide

The Census Bureau has recently released a Directory of Federal Statistics for Local Areas, 1966, which updates and expands the coverage of the Directory of Federal Statistics for Metropolitan Areas, ussued in 1962 by the Advisory Commission on Intergovernmental Relations.

The new publication is a comprehensive reference guide to nationwide social and economic data for geographic and political areas below the state level which are contained in publications of federal agencies. No statistics are given. The Directory presents a summary, in tabular form, of the kind of detailed information to be found in periodicals, census and survey reports, releases, and special reports issued by federal agencies. It also shows the kinds of areas for which data are available, the frequency with which the items are reported, and the specific source document in which the figures appear. The statistics referred to in this guidebook are limited to current information, i.e., since 1960, for such areas as standard metropolitan statistical areas, counties, cities, townships, city blocks, and rural areas. The subject matter is arranged under 22 major headings, including such diverse topics as business and commerce, governments, population, climate, etc. A subject index is included.

This Directory is being distributed to state and local planning agencies and cooperating federal agencies. Copies may be purchased from the Superintendent of Documents, Government Printing Office, Washington, D. C. 20402.

Population Estimates for Small Areas

Although the Bureau of the Census has a well-established state and national population estimates program, it was not until 1963 that the Bureau undertook the development of population estimates for standard metropolitan statistical areas. The initial report in the standard metropolitan statistical area series was released in April 1964. It shows estimates for the 15 largest standard metropolitan statistical areas and their 68 constituent counties. The program has expanded each year since then and currently estimates of population are published for each of the counties in the 55 largest standard metropolitan statistical areas. In effect, the program now includes all metropolitan areas that had a population of over 500,000 in 1960. These 55 metropolitan areas include 190 counties with a 1965 combined estimated population of about 90 million.

By next year it is hoped that the reports will include population estimates for the largest 75 metropolitan areas of the country, including about 230 counties. Each of these standard metropolitan statistical areas had more than 300,000 population in 1960. A major target of the program is to provide estimates for the 100 largest metropolitan areas and their constituent counties by the end of the decade.

The Census Bureau does not, as part of its regular program, prepare population projections for areas below the state level. Recently, however, negotiations with the Office of Civil Defense concerning their needs for metropolitan area projections have led to an agreement to develop over the next 1½ years population projections for each of the metropolitan areas in the country for 1975. In connection with this project, population estimates for mid-1965 for each metropolitan county will be developed (2).

SUMMARY

In the foregoing discussion we have attempted to outline the various Bureau programs and activities in terms of the statistics of particular value to planners, researchers, and administrators involved in decisions which require knowledge of the characteristics of small urban areas. We have described our plans for further research and development in the application of Census data to small area data problems under live field conditions in New Haven, Connecticut, and the potential we hope to explore in the new geographic coding capabilities in cooperation with federal, state, and local agencies. We have pointed out the expansion in data of significance for small areas, and the new methods for its use and dissemination which we anticipate in the upcoming demographic and economic censuses, as well as new items which we believe will be of use in transportation research and planning. Finally, we have listed a number of recent publications which should assist many people dealing with local problems. We welcome suggestions and ideas for additional fruitful avenues of exploration in the use of Census data.

REFERENCES

1. Fay, W. T., and Hagan, R. L. Computer-Based Geographic Coding for the 1970 Census. Paper presented at American Statistical Association, Los Angeles, Aug. 15, 1966.
2. Bureau of the Census. Small-Area-Data Activities, Vol. 1, No. 3, Nov. 1966.
3. Taeuber, Conrad. Features of the 1970 Census of Special Interest to Local Users. Paper presented at the 24th Interstate Conf. on Labor Statistics, Chicago, p. 2, June 4, 1966.
4. Hansen, M. H. Some New Development in Increasing the Availability of Census Information. Paper presented at American Statistical Association, Chicago, pp. 6-8, Dec. 10, 1964.

A Classification of Urbanized Areas for Transportation Analysis

WALTER H. BOTTINY and BEATRICE T. GOLEY, U.S. Bureau of Public Roads

As an initial step in the development of an urban area transportation typology, 213 urbanized areas (1960) were tentatively segregated into four population size groups and then subdivided, on the basis of population density, into three density classes for each group. This approach was based on the theory that automobile availability and the mode of transport used in the work trip are functions, primarily, of population size and density.

Analysis revealed the inadequacy of population density in explaining urban area variations in most of the selected indicators of transportation and socioeconomic characteristics.

An inferential statistical method, the stepwise multiple-regression procedure, was used to analyze the interrelationship between 6 transportation variables (dependent) and 12 socioeconomic variables (independent) in each of the four population size groups. Several of the equations developed have potential use for estimating certain indicators of automobile availability and travel characteristics.

A regrouping of urbanized areas, using as the criterion a ratio of the number of automobiles available to the number of persons employed, resulted in urbanized area types which in each population size group display distinguishing transportation characteristics as well as characteristics underlying socioeconomic factors and geographic location.

•THIS research was performed by division staff assigned to this task, which is one of several included in the national research project, "Underlying Factors in Urban Transportation Analysis."

One major area of concern in the development of transport requirements and the formation of new methods of forecasting demand for highway transport is the need to examine systematically the underlying factors in urban transportation. The design of this project contemplates a number of tasks at the national level which involve, primarily, a collation and correlation of significant variables and empirical values that have been associated with the demand for urban transportation.

OBJECTIVE AND SCOPE

The objective of the typology task is to develop urban area transportation types that will facilitate the analyses of urban transportation and the estimation of present and future urban travel characteristics. Segregation of urban areas by a transportation classification is to be accomplished with the aim of identifying the major factors underlying the transportation "mix" of an area. Statistical methods of analysis will be used to test the relative strength of underlying variables and the design of equations for estimating certain urban transport variables. An estimating manual or handbook will be prepared.

Characteristics of the physical elements, population, income, economic function and growth, as well as transport, are included in the scope of this task.

With its stated purpose of facilitating urban transportation analysis, the typology task is fundamental to the overall project. This task is also comprehensive in scope

---

Paper sponsored by Committee on Economic Forecasting and presented at the 46th Annual Meeting.

taking into account certain aspects of the other task areas: forecasting, population characteristics and change, urban economics, effects of rising income, and emerging urban form.

URBAN AND TRANSPORTATION DEFINED

To describe the scope of this project and to more easily focus on the major areas of concern, it is necessary to define the terms urban and transportation.

Urban Area

Urban generally connotes a geographic area which is "built up" or has well populated industrial, commercial and housing structures in close proximity. The measures of this condition vary somewhat among agencies involved in studying these areas as well as within individual studies prepared by the same agency.

The U. S. Bureau of the Census has the most complete and comprehensive data on population and its geographic distribution. Use of Census definitions, therefore, has the advantages of comparability and consistency.

Basic census classification of urban places are standard metropolitan statistical areas (SMSA's), urbanized areas, cities, and places with over 2, 500 population.

In this report, urban area and urbanized area are the same (Fig. 1). Urbanized area was selected as the study area over SMSA and city for several reasons:

1. Urban transportation analysis should concern itself with the area in which almost all of the urban transportation problems exist.
2. The SMSA, composed as it is of whole counties, takes in much rural area.
3. City (the political entity) varies widely in the population density at its limits. New York or Philadelphia, for example, have high densities extending beyond their borders; whereas Houston and Oklahoma City boundaries take in vast areas of extremely low density.
4. Urbanized areas generally represent the thickly settled core of the SMSA's with the exception of the New York-Northeastern New Jersey and the Chicago-Northwestern Indiana urbanized areas.

It is recognized, of course, that many characteristics and relationships can often be identified whether the unit of study is the urbanized area, the SMSA, or the city.

Urban Transportation

Transportation in an urban area is thought of as the movement of people and goods from place to place. Transportation of persons may be described in terms of pedestrians, vehicle passengers, or vehicle drivers. Mode of transport may also be considered in relation to purpose of trip—work, shop, social, recreation, etc. Purpose of trip and mode of travel may be considered in relation to the numbers and types of vehicles available. Other descriptors of urban transportation are trip frequencies, distances and speed, and miles of travel by various modes during a certain period of time. Therefore, a systematic examination of the underlying factors in urban transportation should begin with a comparative analysis of urban complexes and their transportation characteristics.

THEORETICAL BACKGROUND

Choice of Transport Mode

The use of a particular mode of transport by an individual in an urban area depends initially on the choices available to him. The variety of choices available, whether on rail, road or sidewalk, vary from city to city, but some cities are similar in the combination of facilities provided. Some are equipped with extensive rail facilities but not so well equipped to facilitate motor vehicle travel.

In the largest cities the transportation spectrum runs from the elaborate rail transit system in New York to insignificant rail transit facilities in Los Angeles. Between

Figure 1. Urbanized areas, 1960.

these two extremes the transport type shifts abruptly from rail to road, after taking into account Boston, Philadelphia and Chicago.

Through the remainder of this transport spectrum are variations among the cities in the types of road facilities provided—the more adequate the road transport network, the less adequate the bus service. Design standards and extensiveness of the road structures improve as the level of bus service declines, from Washington, D.C., down through Baltimore, Pittsburgh, Milwaukee, Cleveland, Buffalo, St. Louis, and San Francisco.

Beyond this range, transport availability almost limits the choice to automobile because of the emphasis on facilities provided for this type of travel. The large areas in this category, in ascending order according to adequacy of highway facilities, are Minneapolis-St. Paul, Detroit, Houston, and Los Angeles.

The varied choice of transport has gradually shifted from the mix of rapid rail, streetcar, bus, automobile, and walking to a mix which excludes rapid rail and streetcar but with significant emphasis on bus and walking, and then to a mix in which the choice is narrowed almost exclusively to automobile.

Tomorrow's Choice—A Function of Yesterday's Investment

This description was intended to underscore the fundamental importance of the type and density of physical structures erected to accommodate social and economic activities and the transportation facilities constructed to service those structures, e.g., the location and design of buildings and rail facilities and of collector and arterial streets and highways. It is these factors which explain to a great extent the differences between areas in the relative use of private and collective passenger vehicles. Historically then, use of the private vehicle increased, relative to the use of the collective vehicle, as the road structure improved, relative to rail improvements.

The old rapid rail cities, working in the direction of providing facilities for the private car, have had some success. The transformation from rail to road, however, is very slow and becomes slower as additional portions of the skeletal system of the city are dismembered. The relatively permanent remaining portions of the skeleton, including the rail system serving it, become real "sore" spots which resist the change.

Often, the road system is superimposed on the rail system, thus providing the commuter with a choice of modes. Where one supplants the other, however, opportunity for choice has not been improved.

The relative success in changing from rail to road, or from streetcar and bus to automobile is, again, dependent on the degree of permanence of investment that was made, or is made, in the skeletal structure of the city, a key element of which is its transport system.

The structure of the core and inner ring of New York, for example, is permanently in place. Office, government, educational buildings as well as apartment houses and the subways that serve them are firmly set for a long time to come (1). The same is true in varying degrees in Chicago, Philadelphia and Boston—cities which have common backgrounds of growth.

Los Angeles and Houston, on the other hand, are heavily committed to motor vehicle transport. Areas such as these can therefore be expected to rely completely on the highway facilities that make up such an important part of their multinucleated skeletal structures (2).

Hypotheses

Initial examination of indicators of the general characteristics of an urban area and that area's reliance on various modes of transportation led to the hypothesis that a classification or typing of urban areas would facilitate analyses and make more comprehensible the data used in describing how interactions of several variables differ from one type of urban area to another.

Conceivably, a transportation classification of urban areas would be useful in conjunction with urban area transportation studies. It may be possible to estimate certain characteristics of travel patterns in a particular urban area on the basis of the findings in completed studies of other areas of the same type. Conversely, it is also conceivable that some findings of completed studies of a small sample of urban areas of a particular type may have general application to other areas of that type.

Urban areas having similar transportation characteristics may have similar socioeconomic characteristics. By establishing quantitative relationships between certain aspects of the two characteristics, a measurement is provided for forecasting future travel characteristics of an urban area type (3).

Studies of travel patterns, such as those made in origin-destination surveys, may be aided by the development of such indicators as trip frequencies, trip purpose, time of travel, and mode of travel for common use in areas of a particular type.

Data

The empirical research was carried out with the use of data provided by the Bureau of the Census. These data are available in several publications prepared by the Census Bureau. By their completeness and availability, these data satisfied the need for obtaining economically measures of both the urban transport characteristics and the factors underlying them that were judged to be best suited for this experiment and for the possible potential use of the results. For a discussion of the reasoning behind choice of characteristics and their measures see the section "Selection of Variables."

APPROACH—STATIC PHASE

Interrelationships of transportation and socioeconomic variables in each urbanized area are examined as they exist at one point in time (a static or cross-section analysis). This approach was used as a first phase of the task in order to: (a) identify existing major underlying factors; (b) measure their relative influence on currently used indicators of urban travel; (c) develop mathematical equations for estimating present travel characteristics in areas where such information is not yet available; (d) develop mathematical equations for estimating future travel characteristics in areas where current data are available; and (e) develop a classification of urban areas based on transportation indicators and underlying factors.

Examinations of the dynamics involved in the growth of an urban area and changes in its transportation requirements are being made and the results of these analyses will be covered in a subsequent report. An integration of the static and dynamic analyses will be attempted to give insight into the relative importance of factors underlying transportation requirements by urban area type, the changing importance of these interrelationships, and the probable effects on the status of transportation types and individual areas within these types.

RESEARCH PLAN

The research plan contemplates analyses of urban transportation related to general characteristics of an urban area—structure, function, form and growth, and transportation characteristics, such as automobile availability and the use of various mixes of the four general characteristics. Certain structural indicators selected from those such as population size, density and age composition, personal income, and age of the area, will be combined with selected measures of economic function, geometric form and growth.

For example, a medium-sized, rapidly expanding urban area having a youthful population, new housing, a diversified industry mix, and a circular form on flat terrain would be expected to have a high rate of automobile ownership and use.

On the other hand, a large, densely populated urban area, having low rates of increase in population and income, a manufacturing-oriented economy, a relatively old population, and a ribbon shape paralleled by rough terrain would be expected to have relatively low rates of automobile ownership and use.

Accomplishments

The work accomplishments given in this report cover correlation analyses of certain structural measures with measures of aggregate transportation use for 1960. The transportation variables are also related to one measure of an area's economic base. Growth data, i. e., data on changes in the structural and functional variables have been compiled and prepared for mechanical statistical correlation analyses.

Research Not Yet Completed

The characteristic of form has not yet been related to transportation; however, considerations of urban form led to two broad categories: (a) geometric shape of the area—circular, square, rectangular, radial, linear, oval, triangular, bowl; and (b) transportation-terminal type—seaport, lakeport, riverport, airport, railroad or highway junction, or some combination thereof.

It is expected that the shape in combination with the transportation base of an area would indicate the relative accessibility of zones within an area and that area's conduciveness to automobile transport and its potential paths of expansion.

A linear-shaped seaport, for example, indicates orientation toward the waterfront where much activity is concentrated. Its linear shape indicates topographical barriers which inhibit growth in two directions when a long waterfront is paralleled by mountainous terrain a few miles distant. High population densities usually develop with these conditions, which make rail or highway mass transit efficient. Seattle appears to fit this description.

A radial-shaped railroad and/or highway junction indicates relatively high central business district (CBD) orientation and development along the railroad lines and highways emerging from it. The degree of orientation toward the automobile for the area as a whole will depend on the population density and concentration of activity in the CBD and immediate environs. In the rest of the area, conditions of this form are amenable to automobile use. From this perspective the potential growth may be seen in the spaces between the radials. Dallas would seem to fit this description.

These and other examples of the hypothetical relationships between urban form and transportation remain to be tested. Significant results of these tests when coupled with the key indicators of population and economic growth and related to transportation characteristics should reveal the variables underlying transportation use and the relationship between changes in transportation variables and urban form.

Employment concentration in certain industries has been used to measure the economic function of an urban area. Data on employment by industry for 1950 and 1960 have been compiled in order to show changes in function. Comparative analyses are being made to determine what relationship may exist, if any, between changes in economic function and automobile availability.

Trend data, where available, are being compiled for the purpose of comparing changes in social and economic factors with changes in the transportation classification of urban areas. Supplemental reports are being prepared.

Other aspects of this task in which pilot studies and experiments have been made but in which considerable research remains to be done include the following:

1. Comparisons of social status scores of urban areas. These scores, which are composed of weighted indices of educational, occupational and income levels, are intended for use in the study of interrelationships between social status, other socioeconomic factors and transportation use.
2. Bureau of the Census data on the journey to work have been tabulated for several urbanized area for the purpose of making comparative analyses of the central city orientation of urban areas as indicated by worker commuting volumes into and out of central cities.
3. Test the usefulness of the typology in estimating certain trip characteristics for urban areas of a particular type. Data from comprehensive transportation planning studies will be used as inputs into the estimating equations developed for areas of one type or classification. It is contemplated that experimentation and modification using

these data will result in equations which may be generally applied to one type of area.

4. Investigation is continuing into the potential use of several different measures of geometric form and growth patterns of urban areas. The objective is to determine the interrelationships that may exist between changes in urban transportation characteristics and patterns of development.

A STRATIFICATION OF URBANIZED AREAS

The literature covering all facets of urban complexes suggests that features be grouped under the following general categories: urban structure (4, 5, 6), function (7, 8, 9, 10, 11), form (12, 13, 14), and growth (15, 16, 17, 18). Characteristics of the human as well as the physical agglomeration come under the heading of structure. Buildings and transport structures and the people they serve are described under this broad heading. Features of composition or makeup include the size and type of buildings and transport facilities and the magnitude of the population and its distribution. Related population characteristics, such as age composition and distribution by income groups, are also included to portray the basic physical and human attributes as they might be related to transportation.

Population Size and Density Groups

Population size is a measure of the relative magnitude of an urban area as well as an indicator of relative volumes of persons and goods movement. A comparison of population totals between areas suggests that a segregation of areas into a small number of size groups would improve their comparability.

Following this reasoning, the 213 urbanized areas were divided into four groups according to total number of inhabitants. A ratio of the largest to the smallest group population totals was used as one guide in order to get some degree of uniformity in population range (Table 1).

Population density indicates the relative compactness of an area, a second important consideration related to urban transportation (5). Whereas size is indicative of overall traffic volume, density should indicate relative congestion (19).

Each of the four size groups was then divided into subgroups based on a weighted, gross population density measure.

The number of persons per square mile in the central city was weighted by the proportion of the urbanized area's population residing in the central city at that density. The same computation was made for the urban fringe and the sum of the two weighted densities gave the measures for each urbanized area as summarized in Table 2.[1]

Areas in the two largest size groups distributed themselves rather equally among high and low densities. In the medium and smallest size groups the density distributions were not even. Areas in each of these groups tended to cluster around a moderate density.

Size and density groups were thus established for experimental purposes to determine the effect of these variables on urban transportation characteristics. Effort could then be made to identify the distinguishing transportation characteristics of each size and density group.

It was anticipated that, while size and density groups may lead to transportation types, it would probably be necessary to regroup the areas on different bases after regression analyses were made.

Population size and density indicate basic structrual characteristics of an urban area. Age of an area, i.e., the period in which it grew most rapidly, indicates the type and spatial arrangement of buildings and the type of transportation service available. Other important characteristics of urban makeup are the age composition of the

---

[1] The original manuscript of this paper contained an appendix with supporting tabular data given in detail. These tables may be obtained from the Highway Research Board by special arrangement. Inquiries should refer to XS-10, Record 194.

TABLE 1

URBANIZED AREA POPULATION SIZE GROUPS[a]

Group Number	Population Ranges (thousands)			Number of Areas	Population Distribution	
	Low	High	Ratio		Total	Percent
1	1,000	14,000	1:14.0	16	51,786	54.5
2	350	999	1:2.9	36	21,154	22.7
3	150	349	1:2.3	59	13,153	12.6
4	50	149	1:3.0	102	9,755	10.2
Total	50	14,000	1:280.0	213	95,848	100.0

[a] In thousands.

Source: Adapted from U.S. Census of Population: 1960, Final Report PC(1)-1A, Table 23, U.S. Department of Commerce, Bureau of the Census.

population and the distribution of the population according to income levels. The foregoing indicators of an urban area's structure were selected in the order of their considered relative importance to analysis of urban transportation.

Size-Group Assumptions

Segregation of the 213 urbanized areas into the four population-size groups (Table 1) was based on certain assumptions and considerations.

One assumption is that the larger the area the more complicated its transportation problem; urbanized areas with populations of 1 million or more have the severest problems.

It was recognized that, although the range in size of the largest areas is broad (1 to 14 million), there are only 16 areas in the group. This number does not lend itself to being subdivided because the number of observations would then be too small to make valid correlation tests using statistical methods.

The remaining 197 areas were then divided into three groups to reduce the variance in size and the assumed related transport problems.

This experimental grouping was based on several considerations: (a) minimizing the number of groups; (b) reducing the variation in size to proportions which would

TABLE 2

URBANIZED AREA POPULATION SIZE AND WEIGHTED DENSITY GROUPS

Group Number		Weighted Population Density Range[a]		Number of Areas	Population Distribution	
Size	Density	Low	High		Total (thousands)	Percent
1	1	8,966	15,785	5	26,183	27.3
1	2	6,212	7,458	6	13,643	14.2
1	3	2,702	5,730	5	11,960	12.5
2	1	5,070	6,459	12	6,730	7.0
2	2	3,521	4,788	11	6,113	6.4
2	3	1,165	3,496	13	8,311	8.7
3	1	4,673	8,266	14	3,319	3.5
3	2	3,026	4,472	27	5,679	5.9
3	3	1,502	2,969	18	4,155	4.3
4	1	4,024	10,004	32	2,950	3.1
4	2	2,518	3,947	41	4,329	4.5
4	3	680	2,447	29	2,476	2.6
12 Group Total		680	15,785	213	95,848	100.0

[a] Persons per square mile.

Source: Adapted from U.S. Census of Population: 1960, Final Report PC(1)-1A, Table 22.

make group average characteristics more representative; and (c) making comparative analyses more meaningful.

Almost 96 million persons (53 percent of the population) lived in the 213 urbanized areas in 1960. The 16 largest areas contained nearly 52 million, almost 30 percent of the total national population. In other words, almost one-third of the population of the United States resides in 16 urban agglomerations, each having more than 1 million persons.

The 102 urbanized areas in the smallest population size group (50,000 to 150,000) contained less than 10 million persons, or 10 percent of the total urbanized area population in 1960

Table 1 gives the wide difference between the largest and smallest areas. Data on the number of areas and the population distribution among the four size groups show clearly the tendency toward population concentration.

Variations such as these, in the population size of urbanized areas, indicate the variations among the areas in the magnitudes of travel volumes. Areas such as New York, Los Angeles and Chicago in the largest size group are reasonably expected to have transportation requirements of a more complex nature than those of areas in the other size groups, because of the larger number of persons and greater expanse of land area to be served.

Density Group Assumptions

Surface movement of persons within an area is expected to be more or less congested and more or less rapid, depending largely on the density at which the people reside and work. Relative densities are given in Table 2.

Theoretically, then, population density is considered to be the major factor underlying the demand for the types and extents of urban transportation systems. Densely populated urban complexes are expected to be best served by vehicles that have a large passenger-carrying capacity, thereby providing mass movement of persons relative to vehicles. In other words, population density is closely associated in a positive way with general orientation toward persons, and in a negative way with automobile orientation.

TABLE 3

AVERAGE LAND AREA, POPULATION SIZE AND WEIGHTED DENSITY OF URBANIZED AREA GROUPS

Group Number		Number of Areas	Land Area (sq mi)	Population	
Size	Density			Total (thousands)	Weighted Density[a]
1	1	5	766	5,237	10,712
1	2	6	512	2,274	6,961
1	3	5	675	2,592	4,677
1	All	16	642	3,237	7,420
2	1	12	136	561	5,496
2	2	11	160	556	4,230
2	3	13	275	639	2,720
2	All	36	194	588	4,107
3	1	14	61	238	5,928
3	2	27	64	210	3,838
3	3	18	105	231	2,481
3	All	59	76	223	3,920
4	1	32	22	92	5,279
4	2	41	38	106	3,116
4	3	29	54	85	1,846
4	All	102	38	96	3,433

[a] Persons per square mile.

Source: Adapted from U.S. Census of Population: 1960, Final Report PC(1)-1A, Table 22.

Population density being equal, differences between urban areas in automobile orientation are often attributed to differences in income level. On the theory that relatively high income can overcome, to some extent, the congestion costs imposed by high population density (higher vehicle operating, maintenance, insurance, and parking costs), areas with relatively high income will have greater reliance on the automobile than an area of similar density but lower income.

Comparisons of Urbanized Area Size and Density Groups

The influence of New York, Los Angeles, and Chicago on the average size and density of the 16 largest areas is indicated in Table 3. An average population total of 3.24 million for the largest group is 5½ times the average size of the second largest group (588 thousand). Differences are far less between the smaller groups with the second group being over 2½ times the size of group 3, and the third group a little more than twice the size of group 4.

Averages for the density groups within each size group do not vary widely, with the exception of these in the largest size group. Population totals, for example, range from 5.24 million down to 2.59 million in the group of areas being largest in size and highest in density; whereas there is relatively little variation in size among the density groups in each of the other three size groups. Thus, similarity in population size among the density groups within each size group is necessarily reflected in the negative correlation between the square miles of land area and the population densities in each size group (again with the exception of size group 1). Population size was thereby seen as a constant within size groups 2, 3, and 4.

Selection of Variables

Automobile ownership measured by the auto-employment ratio was selected as a key dependent variable in order to overcome differences in population characteristics that often underlie urban area variations in autos per capita and autos per dwelling unit—two commonly used indicators of automobile reliance. Differences between areas in auto-employment ratios were expected to be more sensitive to factors indicating the relative personal need, desire and ability to own autos, the automobile conduciveness of areas as measured by area age, and age distribution of the population and income distribution, in addition to population size and density.

Generally, need to own an automobile is implied in the measured age of an area, its population density, and the newness of its structures. The longer an area has been established and the more time that has passed since its period of most rapid expansion, the higher will be its population density, the older will be its structures, and the less automobile oriented it will presently be. Although its newer suburbs are conducive to automobile use, its central city still relies somewhat on the older modes of mass transit and walking in many trips—automobile need is not great.

Urbanized areas of more recent development are characterized by low population densities and new buildings designed to fit into the automobile age. Little or no provision is made for the older modes of transport in these areas; therefore, need for an automobile is great.

Desire to own automobiles is indicated by the proportion of population in the age groups associated with high rates of automobile availability. Under 18, 18 to 64, and 65 and over were the three age-group distributions that are compared with rates of automobile availability. A high proportion of persons under 18 indicates a relatively large number of families in early stages of the life cycle. These families predominate in suburbia where multicar ownership is desirable, with husband and wife each wanting to own a car.

The middle-age group, 18 to 64, takes in most of the workers. People in one-person households and mature families make up this group which usually predominates in central cities where the ownership and use of an automobile is less desirable, especially where adequate mass transit is available.

Large proportions of persons in the 65-and-over group are presumably associated with both high and low auto ownership rates, depending on the size and density of an area and the availability of mass transit. In the large, dense areas, older persons in the central city would not have a high rate of auto ownership, but those in this age group who live in areas that are not very large or dense are expected to own automobiles.

In areas where there is a high proportion of persons in this age group, it was expected that the number of retirees owning cars would effect an unrealistically high rate of automobiles per 100 persons employed. Areas in which retirees owning cars would destroy the usefulness of the auto-employment criterion were expected to be few and could be readily recognized.

Where the need and desirability of owning an automobile exists, differences in rates of automobile ownership will be accounted for by differences in ability to own and operate automobiles. Ability is measured in dollars of personal income on a per capita basis and on the basis of proportions of high and low-income families ($10,000 and over and under $3,000).

It seemed logical to expect areas of similar size, density, age composition, and distribution of the population to have differences in automobile ownership rates which could be explained in almost every case by differences in the level and distribution of income.

ANALYSIS OF URBANIZED AREAS BY POPULATION SIZE AND DENSITY GROUPS

Visual Comparison

In Table 4, socioeconomic and transportation variables have been added to population size, density, and land area for each of the four size groups. The positive relationships between these three measures are shown. Income per capita is also positively related to size and age of the area. The average values were computed from tabulations of individual area data which are not included herein.

It is interesting to note the dissimilarities between the largest urban areas and those in the three remaining groups in the average values of certain characteristics. On the

TABLE 4

AVERAGE VALUES OF SOCIOECONOMIC AND TRANSPORTATION CHARACTERISTICS OF URBANIZED AREAS BY POPULATION SIZE GROUPS

Characteristic	Population Size Groups			
	1,000,000 and over	350,000 - 1,000,000	150,000 - 350,000	50,000 - 150,000
General:				
1. Total population, thousands	3,237	588	223	96
2. Population density, persons per sq mi	7,420	4,107	3,920	3,433
3. Land area, sq mi	642	194	76	38
4. Income per capita, 1959 $	2,250	2,052	1,947	1,873
5. Age of area, year	1914	1914	1919	1921
6. New housing units, %	26.6	33.2	30.9	28.5
7. Population 18-64 years, %	57.1	56.0	55.9	55.5
8. Population 65 and over, %	8.7	8.5	9.0	8.8
9. Population under 18, %	34.2	35.5	35.1	35.7
10. Family income under $3,000, %	12.5	15.9	18.1	18.1
11. Family income $10,000 plus, %	21.3	17.6	15.5	14.2
12. Persons per occupied housing unit	3.23	3.24	3.25	3.26
13. Autos per 100 employed residents	76.1	86.3	85.7	84.8
Primary work trip mode, %:				
14. Auto or car pool	61.41	70.17	71.29	72.25
15. R.R., subway, or elevated	4.48	0.16	0.34	0.25
16. Bus or streetcar	17.28	12.40	9.83	7.36
17. Walk	7.88	7.26	9.58	11.57
18. Other	1.46	1.78	2.10	2.37
19. Work at home	2.37	3.58	2.51	2.51
20. Not reported	5.13	4.66	4.35	3.70
Occupied housing units, %, with:				
21. No car available	25.0	20.1	20.0	20.3
22. One car available	56.0	56.3	57.8	57.1
23. Two cars available	16.9	20.9	19.8	20.1
24. Three or more cars	2.2	2.7	2.5	2.5

Sources: Adapted from data in the U.S. Censuses of Population and Housing: 1960; the Statistical Abstract of the United States: 1964; and the County and City Data Book: 1962.

average, population density and per capita income are higher in the largest areas than in areas making up the remaining groups.

Except for the largest areas, the average number of automobiles per 100 employed residents decreases slightly from the second largest to the smallest areas as both average population density and average per capita income decline.

Comparison of primary work trip mode shows that in the largest size group almost 22 percent of the workers rely on mass transit. In the three smaller area groups there is a gradual decline in the use of mass transit and a corresponding increase in walking. Use of the auto or car pool increases sharply from the largest to the second largest size group and then shows slight increases to the third and fourth groups. However, in the latter two groups a slight decline in the number of autos per 100 employed residents accompanies the increase in auto or car pool travel to work. Because these movements are associated with the shift from transit to walking, it may be assumed that as the size of the area becomes smaller, mass transit becomes less adequate; walking and car-pooling become more necessary. Car-pooling would, therefore, appear to be an increasingly larger proportion of the auto and car pool total.

Data on automobiles available by occupied housing units depict similarities among the three smallest size groups in all four characteristics (Table 4). The largest size group differs significantly from the other groups in the percent of households with no cars and those having two or more cars.

Density Groups

Characteristics of the urbanized areas in each of the 12 size-density groups are given in Table 5. (The areas in each of the four size groups were subdivided into high, moderate and low density groups.)

The data for each density group in the areas of the largest size group demonstrate the effect of density on the measures of availability and use of the automobile and other means of transportation. The number of automobiles available per 100 employed residents increases sharply as density drops from high to moderate to low. This negative correlation is also apparent in the data on the percentage distribution of housing units with one, two and three or more cars, but only in the areas having a population of 1 million or more.

In the second largest size group there is considerable narrowing of differences between density groups in the indicators of automobile availability. In the largest size group, as density drops from high to low the number of autos per 100 employed residents increases by 20, from 65 to 85. In the second largest group the range in the same automobile ownership indicator is reduced to approximately 12 as the number of autos per 100 employed persons increases from 82 to 94 with the reduction in population density. A similar range of 80 to 92 is given for the third largest group, whereas in the smallest size group no correlation and practically no variation is indicated between the density groups.

The effect of density on the distribution of housing units according to the number of automobiles available is reflected in the data for the largest areas, and to a much lesser extent in the two middle-sized groups, but not at all in the smallest area group.

It is significant that the urbanized areas in the lowest density group of the 350,000 to 1,000,000 population size group have the highest rate of automobile ownership (93.8), the lowest proportion (17.5 percent) of households with no car available, the highest percentages of two-car (24.5 percent) households, and households with three or more cars (3.2 percent).

Of greater significance, however, are the similarities in the percentages of workers who traveled to work by auto or car pool. Omitting the two highest density groups in the largest area class, the range in the percentages of workers using the auto in the work trip is a narrow 68 to 73 percent (rounded). When these figures are compared with those in the line of figures immediately above, the interrelationships between the two and population density are quite clear. Within each population size group, change from density group to density group in the number of autos per 100 employed residents is accompanied by a similar change in the percentage of workers using the automobile

TABLE 5

AVERAGE VALUES OF SOCIOECONOMIC AND TRANSPORTATION CHARACTERISTICS OF URBANIZED AREAS BY SIZE AND WEIGHTED POPULATION DENSITY GROUPS

Characteristic	Population Size by Weighted Population Density Groups											
	1,000,000 and over			350,000-1,000,000			150,000-350,000			50,000-150,000		
	Over 8,000	6,000 - 8,000	Under 6,000	Over 5,000	3,500 - 5,000	Under 3,500	Over 4,500	3,000 - 4,500	Under 3,000	Over 4,000	2,500 - 4,000	Under 2,500
General:												
1. Total population, thousands	5,237	2,274	2,392	561	556	639	237	210	231	92	106	85
2. Population density, persons per sq mi	10,712	6,961	4,677	5,496	4,230	2,720	5,928	3,838	2,481	5,279	3,116	1,846
3. Land area, sq mi	766	512	675	136	160	275	61	64	105	22	38	54
4. Income per capita, 1959 $	2,195	2,316	2,223	2,089	2,031	2,036	2,032	1,874	1,989	1,897	1,915	1,787
5. Age of area, year	1904	1913	1924	1905	1912	1923	1904	1921	1927	1916	1917	1932
6. New housing units, %	21.8	31.0	32.5	27.4	29.0	42.4	23.8	30.5	36.9	25.9	29.8	29.6
7. Population 18-64, %	57.7	58.6	56.3	55.5	55.9	55.9	56.5	56.3	54.8	55.5	55.8	55.3
8. Population 65 and over, %	8.8	7.1	8.6	9.3	8.4	7.2	9.2	8.0	10.4	9.0	9.0	8.2
9. Population under 18 years, %	33.4	33.9	35.3	34.3	35.6	36.3	34.2	35.7	34.8	35.6	35.2	36.5
10. Family income under $3,000. %	12.6	12.3	12.6	15.1	15.8	16.7	14.4	19.8	18.6	16.1	17.4	21.3
11. Family income $10,000 plus, %	20.3	22.9	20.2	17.5	17.9	17.5	16.3	14.9	15.7	14.7	14.5	13.3
12. Persons per occupied housing unit	3.3	3.2	3.2	3.2	3.3	3.3	3.2	3.3	3.2	3.3	3.2	3.3
13. Auto per 100 employed residents	65.1	78.1	84.5	81.5	82.6	93.8	79.8	84.4	92.4	84.0	85.6	84.7
Primary work trip mode, %:												
14. Auto or car pool	52.9	63.1	67.9	67.5	69.8	73.0	69.2	71.3	72.9	72.0	73.0	71.8
15. R.R., subway, elevated	10.8	2.8	0.2	0.1	0.2	0.2	0.2	0.1	0.8	0.3	0.2	0.4
16. Bus or streetcar	18.7	17.5	15.6	13.8	13.7	10.0	11.8	10.1	8.0	7.1	7.5	7.4
17. Walk	8.5	7.4	7.8	8.4	7.7	5.9	10.7	9.8	8.4	12.8	10.7	11.5
18. Other	1.3	1.6	1.5	1.6	1.7	2.0	1.5	2.2	2.4	2.2	2.3	2.6
19. Work at home	2.3	2.4	2.4	3.7	2.8	4.1	2.1	2.2	3.3	2.5	2.4	2.7
20. Not reported	5.4	5.2	4.7	5.0	4.2	4.8	4.5	4.3	4.3	3.4	4.0	3.6
Occupied housing units, %, with:												
21. No car available	31.4	23.9	19.8	21.6	21.5	17.5	20.7	20.6	18.5	20.0	20.3	20.6
22. One car available	54.9	56.1	57.0	58.2	56.1	54.8	59.7	56.5	58.2	58.9	56.4	56.0
23. Two cars available	12.2	17.8	20.5	18.0	19.9	24.5	17.5	20.5	20.5	18.6	20.7	20.9
24. Three or more cars	1.6	2.3	2.6	2.3	2.5	3.2	2.1	2.5	2.8	2.4	2.6	2.5

Sources: Adapted from data in the U.S. Censuses of Population and Housing: 1960; the Statistical Abstract of the United States: 1964; and its supplement, the County and City Data Book: 1962.

in the work trip. The changes are in the same direction; however, the amplitudes of change differ considerably from size group to size group.

In the largest size group, an increase of 20 in the automobile availability indicator (from 65 to 85, rounded) is accompanied by a rise of 15 percentage points in the use of the auto (53 to 68 percent, rounded). The same two series show increases of 12 autos and 5 percentage points, respectively, in the second largest group of areas.

In the third largest group, use of the auto in the work trip becomes even less responsive to change in the number of autos per 100 workers as a change of 12 in the number of autos is reflected in a change of 4 percentage points in use. There is virtually a stable relationship between the two variables in all three density groups of the smallest areas.

The relative stability in the percentage of workers using the auto in the work trip suggests the existence of a norm in this regard. In all but the largest and densest urbanized areas, where rail rapid transit plays an important role, there is wide variation among the urbanized area groups in the use of the bus or streetcar and walking to work, but there is little difference in the use of the auto. Use of the bus or streetcar shows some inverse relationship with density but a pronounced positive relationship between population size groups (see Table 4). Walking to work has a negative relationship with both density and size. The positive relationship between size of the area and the use of bus or streetcar in the work trip along with the negative relationship between size of area and walking to work explain the relative stability of the proportion of workers using the automobile in the work trip.

From the foregoing comparisons and further examination of the density group data in Table 5, it becomes apparent that within each size group most of the selected socioeconomic and transportation variables are insensitive to changes in population density.

Among the density groups within each size group, except the largest, there is little variation in the averages of such characteristics as total population, income per capita, the age group components of the population, the proportions of high and low-income families, the number of persons per occupied housing unit, the percent of workers using an automobile or car pool in the work trip, and the percent of housing units with one car available.

Characteristics which show a tendency to vary with changes in density of areas of a particular size are as follows: (a) age of area; (b) the percent of housing units built from 1950 to 1960; (c) the number of automobiles per 100 employed residents; (d) the percent of workers using bus or streetcar or walking to work; and (e) the percent of housing units with no car or two or three cars available. Areas in the small size group, however, do not exhibit this sensitivity to differences in population density.

Population density thus fails as a criterion for general use in distinguishing urbanized areas of a particular type or mix of transportation usage.

SIMPLE CORRELATION AND STEPWISE MULTIPLE-REGRESSION ANALYSES OF FOUR GROUPS OF URBANIZED AREAS

A 24-variable correlation coefficient matrix was prepared for each of the four population size groups. Intercorrelations of variables selected from the matrices are given in Table 6. Twelve of the 24 variables were indicators of demographic, social and economic characteristics. Transportation indicators comprised the other 12 variables.

Six of the 12 transportation variables were selected to be dependent variables in a 12-step (12 socioeconomic independent variables) multiple-regression analysis to be performed for each of the four population size groups (20).

Simple Correlation Analysis of Urbanized Areas in Four Population Size Groups

Results of a simple correlation analysis are summarized in Table 6. Correlation matrices of the 24 variables for each of the four size groups provided the coefficient correlations. Table 6 gives the correlation coefficients for two transportation variables related to each of 12 socioeconomic variables in each of the four population size groups.

TABLE 6

INTERCORRELATIONS OF SELECTIVE VARIABLES

(Simple Correlation Coefficients, r)

Socioeconomic Variable	Urbanized Area Population Size Groups			
	1,000,000 and over	350,000 - 1,000,000	150,000 - 350,000	50,000 - 150,000
(a) Autos Available per 100 Employed Persons				
1. Population, 1960	-0.286	0.065	0.155	-0.074
2. Population density, weighted	-0.749	-0.467	-0.320	-0.091
3. Land area, sq mi	-0.060	0.250	0.206	-0.017
4. Income per capita, 1959 $	0.127	0.218	0.099	0.138
5. Age of area	-0.493	-0.716	-0.597	-0.359
6. Housing units, built 1950-60, %	0.684	0.736	0.745	0.597
7. Population 18-64, %	-0.360	-0.234	-0.635	-0.120
8. Population 65 and over, %	-0.287	-0.187	0.235	-0.371
9. Population under 18, %	0.462	0.306	0.144	0.380
10. Families with income under $3,000, %	0.189	-0.163	0.199	-0.085
11. Families with incomes $10,000 plus, %	0.072	0.190	0.052	0.176
12. Persons per occupied housing unit	-0.148	-0.183	-0.239	0.005
Value of r at 0.05 level of significance	0.500	0.350	0.250	0.190
Number of urbanized areas	16	36	59	102
(b) Percent of Work Trips by Auto or Car Pool				
1. Population, 1960	0.576	0.180	0.131	-0.045
2. Population density, weighted	-0.808	-0.429	-0.210	-0.087
3. Land areas, sq mi	-0.412	-0.238	0.091	-0.060
4. Income per capita, 1959 $	-0.186	0.322	-0.068	0.097
5. Age of area	-0.516	-0.513	-0.354	-0.268
6. Housing units, built 1950-60, %	0.674	0.434	0.569	0.548
7. Population 18-64, %	-0.630	-0.374	-0.593	-0.148
8. Population 65 and over, %	-0.411	0.042	-0.121	-0.460
9. Population under 18, %	0.741	0.197	0.447	0.471
10. Families with incomes under $3,000, %	0.266	-0.435	0.001	-0.125
11. Families with incomes $10,000 plus, %	-0.145	0.237	-0.017	0.154
12. Persons per occupied housing unit	-0.209	-0.192	0.006	0.112
Value of r at 0.05 level of significance	0.500	0.350	0.250	0.190
Number of urbanized areas	16	36	59	102

Sources: Adapted from the U. S. Censuses of Population and Housing: 1960, and the County and City Data Book: 1962.

Number of Autos Available per 100 Employed Persons

In Table 6(a) the number of autos available per 100 employed persons has a negative correlation with population density in each size group. As the value of r drops from -0.749 in the largest group to -0.091 in the smallest group, the acceptable level of r also drops, so that only in the smallest group does r (-0.091) fall below the value at which the chances would be greater than 1 in every 20 (0.05) that the value of r could be obtained from a sample of two variables having zero correlation.

The r value of 0.745 has a high level of significance for housing units between 1950 and 1960 and autos available per 100 employed persons in the size group 150,000 to 350,000. A significant r value (0.597) for these two variables is also obtained in the smallest size group.

Age of area is significantly correlated (at the 0.05 level of significance) with automobiles per 100 employed persons in all but the largest size group.

No significant relationships are indicated between the automobile-employment ratio and population size, land area size, income per capita, the proportions of low and high-income families, or the number of persons per occupied housing unit.

In the largest population size group, the number of automobiles per 100 employed persons is significantly related in a negative way to population density and in a positive way to the percentage of housing units built between 1950 and 1960. None of the other variables is significant at the 0.05 level, although age of area and the proportion of population under 18 years of age become significant at a lower level (above 0.15 or slightly below an 85 percent confidence that the coefficient did not come from two variables having zero correlation).

Population density and age of area have significant negative correlations with the automobile-employment ratio in the 350,000 to 1,000,000 population size group. The percent of housing units built between 1950 and 1960 has a very significant correlation coefficient of 0.736, considering that an r value of 0.350 is significant.

TABLE 7

POPULATION DENSITY vs AUTOMOBILE AVAILABILITY

Automobile Availablility	Size Group			
	1,000,000 and over	350,000-1,000,000	150,000-350,000	50,000-150,000
Autos per 100 employed persons	-0.749	-0.467	-0.320	-0.091
No-car housing units, %	0.879	0.384	0.162	0.012
Two-car housing units, %	-0.684	-0.590	-0.252	-0.199

In the third largest size group, population density, age of area and the percentage of new housing units have significant r values. In addition, the proportion of the population in the 18 to 64-year age bracket assumes importance.

Population density is not an important factor underlying the number of autos per 100 employed persons in urbanized areas of the smallest size group. The proportions of the population in the young and old age groups assume importance for the first time in this group of areas.

Percent of Work Trips by Automobile or Car Pool

In all four size groups, the age of an area and the percent of its housing units that were built in the 1950-60 period have significant correlation coefficients with the percent of workers who use an auto or car pool in the work trip (Table 6(b)). The age distribution of the population becomes important, in all four groups, in relation to this work-trip mode variable. A greater proportion of persons in the 18 to 64-year age group bears a negative relationship with use of the auto in the work trip, whereas a high proportion of the young groups is positively related to auto use in the work trip. This finding is in line with the expectation that those using the auto in the work trip are largely suburbanites belonging to young families having a preponderance of children under 18. The negative correlation between use of the auto in the work trip and the percent of persons aged 18 to 64 is presumed to be due to people of this age comprising a high proportion of the central city dwellers who use mass transit or walk to work.

Another noteworthy item in Table 6 is the significance of the proportion of low-income families as a depressant on the use of the automobile in the second largest urban areas.

Summary

In summary, findings from the simple correlation experiment support the hypothetical explanations of differences in automobile orientation between urbanized areas of particular population size groups, but do not support these hypotheses for areas of other size groups.

These findings give some support to the hypothesis that high population density inhibits automobile ownership, whereas at lower densities, automobile availability is correlated with income.

Table 7 compares correlation coefficients between population density and each of three measures of automobile availability in each of the four size groups. It illustrates the declining importance of population density from the largest to the smallest size groups.

Population density is extremely high in some areas of the largest size group, and it is at this level that density becomes a strong inhibitor of highway use.

The contention that income takes over as a prime determinant of automobile availability where high population density leaves off is supported in this analysis only in a few special cases.

Income level, measured by income per capita, and income distribution, measured by the percent of high ($10,000 per year and over) and low (less than $3,000 per year) income families, bears an interesting variety of relationships to the measures of automobile availability and use.

In none of the groups does per capita income have a significant relationship with the number of automobiles available per 100 employed persons. Significant relationships were indicated, however, between income level and the proportion of housing units with no car available in areas of all but the largest size. The correlation coefficient between income and no-car households in the largest areas was a very low 0.050. Two-car households showed significant correlation with income per capita in only the smallest (50,000 to 150,000) size areas (see footnote 1).

The distribution of income assumes significance as an underlying factor in its relation to no-car households. Coefficients indicated that this is so in all but the largest size areas. In the largest areas income distribution, like income level, bears no significant relationship to any of the three measures of automobile availability.

The factor which most consistently has a significant correlation coefficient with automobile availability is the percent of housing units built between 1950 and 1960. These results are in line with expectations based on the fact that most housing construction in the 1950's was designed with the consideration that the automobile was the primary means of transportation .

Age of area is another consistently significant factor underlying differences in the rate of automobile availability, although the coefficients for this variable are always lower than those for the housing units built between 1950 and 1960. Even though there is undoubtedly collinearity between age of area and housing units built between 1950 and 1960, the fact that coefficients for recent housing construction are maintained at a similarly high level from size group to size group while age of area coefficients generally decline from the second largest to the smallest size group indicates that age of area declines in importance relative to new housing construction, as area size declines.

Numerous other comparisons could be made of the simple correlation coefficients obtained for measures of automobile availability and the several measures of urban structure. Primary mode of travel in the work trip related to the same socioeconomic variables individually produce correlation coefficients which are commonly high for housing units built between 1950 and 1960, and the age of the area—the same two variables that are important in explaining automobile availability differences.

Multiple-Regression Analysis of Urbanized Area Size Groups

Simple correlation analysis has indicated the relative importance of each socioeconomic variable in explaining differences between urbanized areas in each of six indicators of automobile availability and use. Although that analysis gives some insight into relationships between one individual transportation variable and one socioeconomic variable, it does not show the interrelations between one transportation variable and a combination of more than one socioeconomic variable. The purpose of the multiple-regression analysis is to select the combination of given socioeconomic variables that best explains variations in the transportation variable and which serves as a basis for estimating the value of the dependent transportation variable under the various socioeconomic condiditons.

All of the variables involved in the analytical phase are the same as those used in the earlier simple correlation and comparative analyses.

For each of the four population size groups, stepwise multipile-regression analysis was performed separately for each of the six dependent variables. The same 12 independent variables were used in all 24 operations (4 size groups times 6 dependent variables).

Estimating Equations

Equations for estimating the ratio of the number of automobiles available to the number of employed persons for an urbanized area in each size group are given in Table 8.

Table 9 gives equations for estimating the percent of workers using the automobile or car pool in the work trip. Each equation is the best fit for an area in the particular population size group for which the analysis was made.

TABLE 8

EQUATIONS FOR ESTIMATING THE AUTOMOBILE-EMPLOYMENT RATIO IN URBANIZED AREAS WITH 1,000,000 POPULATION OR MORE*

Step No.	Independent Variable	Estimating Equation	$S_{y.x}$	R	R^2
1	Population density (PD), 100 persons per sq mile	Y* = 0.9785 - 0.0028(PD) (0.0006)[a]	0.0800	0.728	0.529
2	Land area (LA)	Y = 0.9778 - 0.0039(PD) + 0.00011(LA) (0.0006) (0.00004)	0.0665	0.822	0.676
3	Pct. housing units built between 1950-60 (HU50-60)	Y = 0.8482 - 0.0032(PD) + 0.00009(LA) + 0.00344(HU50-60) (0.0008) (0.00004) (0.00282)	0.0653	0.827	0.684
4	Persons per occupied housing unit (POHU)	Y = 1.2252 - 0.0030(PD) + 0.00007(LA) + 0.00372(HU50-60) - 0.11908(POHU) (0.0009) (0.00005) (0.00285) (0.12699)	0.0656[b]	0.827	0.684
5	Population under 18 (P-18), %	Y = 0.8619 - 0.0025(PD) + 0.00010(LA) - 0.00245(HU50-60) - 0.44402(POHU) (0.0008) (0.00004) (0.00404) (0.20049) + 0.04457(P-18) (0.02270)	0.0585	0.866	0.750
6	Population, 1960 (P), 100,000's	Y = 0.6235 - 0.0007(PD) + 0.00035(LA) - 0.00289(HU50-60) - 0.52532(POHU) (0.0013) (0.00015) (0.00372) (0.19021) + 0.05531(P-18) - 0.0043(P) (0.02178) (0.0025)	0.0537	0.887	0.787
7[c]	Population 18 to 64 (P18-64), %	Y = 3.4923 + 0.0008(PD) + 0.00041(LA) + 0.00304(HU50-60) - 0.51888(POHU) (0.0014) (0.00014) (0.00457) (0.16836) + 0.02845(P-18) - 0.0059(P) - 0.03912(P18-64) (0.02404) (0.0024) (0.02093)	0.0475	0.912	0.831

*Number of automobiles available per employed resident: $\bar{Y}$ = 0.7606; N = 16.

[a] Standard error of the regression coefficient.

[b] The observed increase in the standard error of estimate is presumed to result from correction for the number of degrees of freedom in the estimating equation. The standard error ($S_{y.x}$) above declines from 10.5 percent to 6.2 percent of the mean ($\bar{Y}$), indicating the small variance of the dependent variable about its mean ($\bar{Y}$).

[c] Additional steps through the 11th did not improve the estimating accuracy of this equation. The 11th step resulted in the following measures: $S_{y.x}$ = 0.0548; R = 0.883; and R^2 = 0.779.

Sources: Adapted from data in the U.S. Bureau of the Census, U.S. Census of Population and Housing: 1960.

TABLE 9

EQUATIONS FOR ESTIMATING THE PERCENTAGE OF WORKERS USING PRIVATE AUTOMOBILE OR CAR POOL IN URBANIZED AREAS WITH 1,000,000 POPULATION OR MORE*

Step No.	Independent Variable	Estimating Equation	$S_{y.x}$	R	R^2
1	Population density (PD), 100 persons per sq mile	Y^* = 82.3282 − 0.2819 (PD) (0.0549)[a]	6.3419	0.793	0.628
2	Population under 18 (P-18), %	Y = 2.31316 − 0.1984(PD) + 2.15875(P-18) (0.0592) (0.91092)	5.4997	0.849	0.720
3	Persons per occupied housing unit (POHU)	Y = 16.20246 − 0.1433(PD) + 3.78984(P-18) − 22.86700(POHU) (0.0624) (1.22726) (12.55461)	5.0666	0.873	0.763
4	Population 18 to 64 (P18-64), %	Y = 164.56937 − 0.1016(PD) + 3.58818(P-18) − 31.24073(POHU) (0.0670) (1.18956) (13.47753) −2.05774(P18-64) (1.46887)	4.8749	0.883	0.780
5	Pct. housing units built between 1950-60 (HU50-60)	Y = 237.01325 − 0.0865(PD) + 2.07349(P-18) − 23.30511(POHU) (0.0710) (2.29622) (17.11353) −3.03121(P18-64) + 0.31002(HU50-60) (1.95187) (0.39923)	4.9653[b]	0.879	0.772
6	Population 1960 (P) (100,000)	Y = 299.36194 − 0.0140(PD) + 1.24838(P-18) − 24.15563(POHU) (0.1119) (2.52559) (17.39143) −3.74269(P18-64) + 0.53571(HU50-60) − 0.0656(P) (2.15149) (0.48502) (0.0776)	5.0375[b]	0.875	0.765
7[c]	Land area (LA)	Y = 346.13190 + 0.1982(PD) + 1.94225(P-18) − 32.28520(POHU) (0.1186) (1.97910) (13.85564) −4.84864(P18-64) + 0.56483(HU50-60) − 0.5670(P) + 0.03046(LA) (1.72298) (0.37684) (0.1998) (0.01157)	3.9122	0.927	0.858

*Percentage of workers using private automobile or car pool: $\overline{Y}$ = 61.4062; N = 16.

[a] Standard error of the regression coefficient.

[b] The observed increases in the standard error of estimate is presumed to result from correction for the number of degrees of freedom in the estimating equation. The standard error ($S_{y.x}$) declines from 10.3 percent to 6.4 percent of the mean ($\overline{Y}$), indicating the small variance of the dependent variable about its mean ($\overline{Y}$).

[c] Additional steps through the 11th did not improve the estimating accuracy of this equation. The 11th step resulted in the following measures: $S_{y.x}$ = 5.0624; R = 0.874; and R^2 = 0.763.

Sources: Adapted from data in the U.S. Bureau of the Census, U.S. Censuses of Population and Housing: 1960.

Tables 8 and 9 indicate the changing relative importance of each independent variable from step to step. Changes in the value of a coefficient and its individual standard error measure its changing influence. In Table 9, for example, the value of the population density coefficient declined from the first through the sixth step while its standard error became greater.[2]

The equations in Tables 8 and 9 represent a best fit for the regression line within the scatter of plotted data for the urbanized areas with a population of 1,000,000 or more. Although these are considered the best combination of independent variables to use in estimating the dependent variable, they are only the best that could be developed on the basis of the 12 independent variables chosen for the analysis.

For other size groups, the goodness of fit of the best estimating equation is sometimes not very much better than the equation derived with half the number of steps. This has been demonstrated by detailed supporting data not reproduced herein (see footnote 1).

The first three steps of the analysis of the group of smallest areas (50,000 to 150,000 population) resulted in the use of the same three independent variables in the same order as for the areas in the 150,000 to 350,000 population size group.

An important feature of the stepwise regression method is the variety of equations it produces. In some cases this offers a choice to the user. The choice depends on two major considerations—the availability of data and the degree of accuracy desired (21).

An example of possible choices may be seen by comparing the equations for estimating the automobile-employment ratio in an urbanized area with a population of 1 million or more (Table 8). In this case it may be desirable to use the third step equation. The additional improvement in the measures of reliability may not warrant the added effort required to obtain the necessary data. Also, some variables added do not meet the desired level of significance as measured by the t test.

Limitations

Of fundamental importance in considering the value or usefulness of any of these equations is the fact that none of these model equations tells the whole story. Significant proportions of the differences between urbanized areas are left unexplained. It is obvious, therefore, that this analysis omitted some important factors underlying the urbanized area differentials in the values of the dependent variables.

Tests show that at various steps one or more regression coefficients lose significance. This occurs at different steps among the four urbanized area size groups and for estimating different dependent variables. Equations in Tables 8 and 9, which were derived from analyses of data for the largest areas (1 million or more population), have variables entering beyond step three that are not significant. In the equations for the three smaller size area groups, variables of insignificant value did not enter until the fifth or sixth step (see footnote 1).

It must be pointed out, however, that this initial experiment tested the adequacy of estimating certain transport characteristics on the basis of a limited number of socioeconomic factors and excluding other transport characteristics. The intent of this approach was to develop equations constructed of a minimum, a few perhaps, independent socioeconomic variables. Estimating tools of this nature would, it seems, facilitate the problem of estimating control totals of automobile availability, travel volumes or other highway planning guides.

Summary

In summary this initial experimental computer run, using the multiple-regression technique, has produced rather crude, static, model equations for estimating certain urban transportation variables.

---

[2] Horizontal lines in the equation tables are shown after the number in the "Step No." column, which includes the last significant independent variable added, measured by comparing the regression coefficient against its standard error.

These results do, however, justify the selection of the variables chosen, and, as stated earlier, lend positive evidence to the hypothesis that the group of factors underlying urban transportation use varies among urbanized areas of different sizes.

Another finding was the lack of support for the priority, or order, in which the socioeconomic variables were selected. Income level, measured by income per capita, followed population size and density, and land area in their prejudged relative importance. Age composition of the population and distribution of families by high and low incomes were considered to be of less importance, as was the relative newness of housing units and age of the area.

As a result of this analysis, income per capita is an important factor in relation to the proportion of occupied housing units with no car available. On the other hand, the proportion of housing units built between 1950 and 1960 is closely associated with all automobile availability variables.

URBANIZED AREAS CLASSIFIED ACCORDING TO AUTOMOBILE RELIANCE

The Classification Criterion

Statistical analyses of the data in each group gave strong evidence that population density would not be useful as a basis for identifying, more closely, transportation characteristics common to a group of areas. Instead, the ratio of automobiles available to total employment was used as the criterion.

Six general types of urbanized areas result from this highway transportation viewpoint. Relative reliance on the automobile for personal mobility becomes the indicator used in segregating these classes. The number of automobiles available for every 100 employed persons, the measure of automobile reliance, is a composite of: (a) the number of automobiles available as derived from the Census data on automobiles available by occupied housing units; and (b) the total number of civilian residents employed. The number of occupied housing units reported as having one, two, and three or more cars available was multiplied by one, two, and three, respectively, to get total cars available for an urbanized area. The total number of persons employed includes all resident civilians 14 years of age and over who had a job at the time of the April 1960 Census.

Automobiles available are the only accessible data which indicate the extent of automobile ownership in all urbanized areas. These data include automobiles available for use by a member or members of an occupied housing unit, regardless of whether the automobile is owned by an occupant of the housing unit. Autos operated by persons stationed at military bases, students living in university dormitories, and other institutional persons were not included. Cars in large fleets of company-owned cars and other cars belonging to small business firms, but available to person or persons in the housing unit, are thus accounted for. Not accounted for are taxis, pickup or larger trucks, although these vehicles in many cases may serve the same purpose as a passenger car or station wagon. To the extent that this condition exists, there is a bias in the low automobile availability count for these areas. For the purpose of this study, however, it is assumed that the number of areas having these high rates of automobile substitution by other motor vehicle types would not be large enough to impair seriously the comparability of automobile availability between groups of urbanized areas.

The autos per capita ratio is hampered by variations between urban areas in the population age compositions—areas having large or small segments of persons under and over normal auto ownership age (18-65). Autos per worker narrows the population to persons who are capable of driving (with some exception, i.e., handicapped workers). It reflects ability of some workers to pay the cost of more than one automobile in a household. Effects of variance in size and composition of the household are largely overcome by using employed persons. Using employed persons as a component also avoids the necessity of estimating households with one, two, three or more, or no cars. Autos per 100 employed residents of an urban area could thus be expected to be more directly related to the underlying environmental factors, such as population size and density, age of an area, age composition of its population, and the levels and distribution of personal income.

Household members who are not employed but have automobiles available, e.g., suburban housewives and children of driving age, inflate the numerator in this ratio (autos available) without affecting the employment denominator. Inflation of the automobile figure relative to employment illustrates more clearly an urban area's relative dependence on the automobile for transportation and the relative ability of its employed residents to provide this transportation for dependents.

Unusually high numbers of automobiles relative to workers are noted in a few places, such as St. Petersburg, Fla., and Lawton, Okla. This reflects the unusually large proportion of retired persons in St. Petersburg and the large number of military personnel living off base in Lawton. These persons are not included in the total number of persons employed, although the automobiles available to them are included in the total number of automobiles available.

The abnormally high ratio of autos to workers in St. Petersburg emphasizes the degree of dependency on the automobile by both workers and retirees in that area. It also implies traffic problems of a different sort since the trip purposes and destinations of retirees would make shopping and social-recreation trips unusually high relative to work trips. Traffic should, therefore, have comparatively low peaks and be more diffused.

Auto Reliance and the Work Trip

The worker and characteristics of his trip to work best describe an urban area's degree of highway use. As a basis for estimating and predicting total highway traffic volumes for an urban area, the number of workers is preferable to the number of housing units or the total population. Highway capacity is generally designed to accommodate peak-hour traffic volumes, usually made up of vehicles being used in the work trip. Knowledge of the number of workers and their reliance on the automobile, relative to other modes of transport in their work trips, is therefore fundamental to sound decisions on highway requirements. An indication of the extent to which workers of an urban area rely on the automobile for transportation is to be found in the ratio of the number of automobiles available for the number of workers residing in the area.

Automobile availability and use of the auto in the trip to work are closely correlated. Urban areas with large numbers of automobiles available relative to the number of workers can be expected to have high proportions of workers using the automobile in the work trip. Areas with high rates of autos to workers usually have high proportions of multicar households which are outgrowths of the need for an automobile by the spouse and/or children of the worker who uses his car in the work trip.

Conversely, urban areas with lower rates of autos to workers will not be as high in the proportion of multicar households, because the need for an automobile by other members of the worker's family is not as widespread inasmuch as some of the workers leave their cars at home and use another mode of transport in the work trip.

Autos per dwelling or housing unit is not as directly connected to the work trip, since urban areas differ significantly in the number of workers per dwelling unit as well as in the number of automobiles available per dwelling unit. The correlation between the two is such that autos per dwelling unit is not a satisfactory substitute.

In areas having a similar number of workers per dwelling unit, the rates of autos per dwelling unit often vary widely. Urbanized areas in Connecticut, for example, average between 130 and 135 workers and 100 automobiles for every 100 dwelling units. In California, urban areas average approximately 115 workers and over 120 autos for every 100 dwelling units. In some urban areas, such as those in Texas, the relationship between the two data series is positive, but in many states it is a negative correlation. By relating autos directly to workers, the effect of variations in workers per dwelling unit is avoided.

Areas which have a relatively high rate of autos to workers will usually have a high proportion of workers driving the auto or riding in a car pool in the work trip. Where a relatively high rate of auto availability is associated with a relatively low rate of use in the work trip, the implication is that car pooling is more extensive than in other areas with similarly high rates of auto availability.

TABLE 10

AVERAGE VALUES OF SELECTED TRANSPORTATION AND SOCIOECONOMIC CHARACTERISTICS OF URBANIZED AREAS BY AREA TYPE

		Transportation Characteristics									Socioeconomic Characteristics						
Area		Automobiles Available per 100 Employed Residents[a]	Primary Mode of Work Trip[b]					Percent of Occupied Housing Units with		Area Age		Newness of Housing Units[d] (%)	Young and Old Age Groups (%)		Proportions of Low and High Income Families (%)	
Type	Number		Auto or Car Pool (%)	Mass Transit (%)			Walk (%)	No Car	Two Cars	Year	Years before 1960[c]		Under 18	65 Plus	Under $3,000	$10,000 Plus
				Total	R.R., Subway, Elevated	Bus, Streetcar										
A	30	106.8	78.6	4.2	0.1	4.1	7.1	13.7	27.1	1940	20	45.7	36.3	8.6	17.2	16.6
B	33	94.5	75.4	7.3	0.7	6.6	7.6	16.3	25.0	1923	37	36.4	37.0	7.7	17.2	16.7
C	65	84.7	72.2	9.0	0.1	8.9	10.1	19.5	19.4	1916	44	27.1	35.5	8.6	17.0	15.3
D	65	75.3	66.3	13.6	0.2	13.3	11.6	23.9	16.5	1912	48	24.6	34.4	9.1	17.7	15.5
E	19	65.6	62.9	15.5	2.2	13.4	13.2	29.3	13.4	1909	51	19.4	33.0	10.2	17.7	13.9
F	1	56.3	35.9	45.1	31.6	13.5	9.2	41.4	10.5	1900	60	19.5	30.6	9.6	12.4	22.7

[a] The total number of automobiles available was calculated by the BPR Office of Planning from the number of occupied housing units having one, two, or three or more cars available as reported in the Census of Housing, 1960. Total automobiles available was then divided by the total number of employed residents reported in the County and City Data Book: 1962.

[b] The principal means of transportation to work during the week prior to the 1960 Census of Population in April.

[c] The arithmetic average of the years nearest to the census year beginning with 1900.

[d] Percent of 1960 housing units in structures built between 1950 and 1960.

Sources: Adapted from data in the U.S. Census of Population and Housing: 1960; the Statistical Abstract of the United States: 1964; and its supplement, the County and City Data Book: 1962.

Transportation Types of Urbanized Areas

Selected characteristics of urbanized area types are given in Table 10 as mean unweighted averages of the data for individual areas (see footnote 1). The area types are arrayed in descending order of reliance on the automobile as measured by the number of automobiles available for every 100 employed residents. Each area type, A through F, is divided into the four population size groups.

The distribution of the numbers of areas among the types (Table 10) makes it apparent that the distribution by type is not normal, but is skewed toward the upper end of the scale where automobile reliance is greatest.

Interrelationships between the transportation characteristics are also apparent. The changes from high to low in the proportion of workers using the automobile in the work trip and in the percent of occupied housing units having two cars available show a positive relationship with the number of automobiles per 100 employed residents. The increase from Type E to Type A in the auto use data is, however, much more gradual than the increase in the auto ownership indicator. This suggests a possible ceiling or limit on use of the automobile in the work trip corresponding with what may be a floor in the proportion of workers who walk to work or use mass transit. Included in this hard core of walkers and transit users (7.1 percent walkers and 4.2 percent transit users in Type A areas) are persons who live close to their places of work and find walking the most convenient form of transport; those who for various reasons, such as age, health or desire, cannot or will not drive a car; those workers who cannot afford to own a car and for various reasons cannot or will not ride as a passenger in a car pool; those workers who own one car and prefer to leave it at home for general use by other drivers in the family rather than buy another car; those workers who have an aversion toward driving a car on congested streets or on less congested freeways where a breakdown is most troublesome and costly; and finally, those workers who cannot afford the cost of parking or for other reasons find mass transit more economical, reliable, convenient or pleasurable.

The data on social and economic characteristics reveal the close relationship between automobile availability and the proportion of housing units that were built between 1950 and 1960. In the Type A group, high rates of automobile availability are accompanied by high rates of new housing (Table 10). Change in automobile availability from one area type to another is related to area age also. The influence of the age composition of the population is less clear, although the proportion of youngsters (under 18) shows a positive relation to automobile availability while increasing old age proportions of the population are less clearly associated with declining rates of automobile availability.

TABLE 11

AUTOMOBILE OWNERSHIP, MODE OF TRAVEL TO WORK, AND RELATED SOCIOECONOMIC CHARACTERISTICS OF URBANIZED AREAS

Population Size and Area Type		Number of Areas	Transportation Characteristics								Socioeconomic Characteristics						
			Automobiles Available per 100 Employed Residents[a]	Primary Mode of Work Trip[b] (%)					Percent of Occupied Housing Units with		Area Age[c]		Newness of Housing Units[d] (%)	Young and Old Age Groups (%)		Proportions of Low and High Income Families (%)	
				Auto or Car Pool	Mass Transit			Walk									
Size	Type				Total	R.R., Subway, Elevated	Bus, Street-car		No Car	Two Cars	Year	Years before 1960		Under 18	65 Plus	Under $3,000	$10,000 Plus
1,000,000 and over-Group 1	A	1	101.6	77.3	7.8	0.0	7.8	5.0	16.0	28.0	1920	40	40.4	33.8	8.9	12.3	24.6
	B	1	89.7	74.5	11.1	0.0	11.1	4.3	17.2	26.7	1950	10	43.9	38.2	5.3	17.9	17.9
	C	4	84.6	67.4	17.3	0.5	16.8	7.1	20.1	20.5	1912	48	27.1	34.8	8.7	11.7	22.2
	D	5	73.8	61.5	21.3	0.2	21.1	8.6	24.8	14.2	1920	40	25.8	34.3	8.7	12.3	20.6
	E	4	65.5	54.4	27.1	9.2	17.8	9.0	30.0	13.0	1900	60	21.0	33.5	9.2	12.4	20.7
	F	1	56.3	35.9	45.1	31.6	13.5	9.2	41.4	10.5	1900	60	19.5	30.6	9.6	12.4	22.7
350,000 to 1,000,000-Group 2	A	5	107.9	78.1	4.9	0.4	4.4	5.7	12.4	28.7	1946	14	53.3	37.1	7.6	13.4	21.0
	B	7	94.3	72.8	10.2	0.1	10.1	5.9	18.1	24.3	1914	46	37.6	35.3	8.8	17.4	16.4
	C	11	84.5	70.6	12.7	0.1	12.6	7.0	19.9	20.3	1910	50	30.7	35.7	8.0	16.4	16.8
	D	13	75.1	65.4	16.7	0.1	16.5	8.8	24.3	16.6	1904	56	25.1	34.8	9.0	15.7	17.7
	E	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	F	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
150,000 to 350,000-Group 3	A	10	107.6	78.4	5.6	0.0	5.5	6.3	15.1	24.3	1939	21	46.4	34.6	11.5	20.1	15.2
	B	7	94.2	73.3	9.5	1.8	7.7	6.7	16.3	24.9	1923	37	37.8	38.1	7.3	18.6	18.3
	C	18	85.2	72.9	9.0	0.1	9.0	9.4	18.5	19.9	1921	39	27.8	34.9	8.0	17.0	15.4
	D	20	76.1	67.1	12.7	0.2	12.5	11.7	23.7	16.9	1910	50	24.4	34.1	8.8	18.6	14.9
	E	4	66.9	64.0	15.4	0.4	15.0	13.0	27.0	12.8	1900	60	16.8	31.0	11.8	15.2	14.2
	F	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
50,000 to 150,000-Group 4	A	14	106.2	79.1	2.7	0.0	2.7	8.3	13.1	28.5	1941	19	42.8	37.4	6.9	16.9	15.5
	B	18	94.9	77.4	5.1	0.5	4.6	8.8	15.6	25.2	1924	36	35.0	37.2	7.7	16.6	16.2
	C	32	84.6	73.1	6.8	0.1	6.6	11.9	19.9	18.7	1916	44	25.4	35.8	9.2	17.9	13.9
	D	27	75.1	67.0	11.2	0.3	11.0	13.4	23.8	16.5	1916	44	24.1	34.5	9.4	19.0	13.9
	E	11	65.1	65.6	11.4	0.2	11.1	14.8	29.9	13.8	1916	44	19.7	33.5	10.0	20.6	11.2
	F	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

---

[a] The total number of automobiles available was calculated by BPR Office of Planning from the number of occupied housing units having one, two, or three or more cars available as reported in the Census of Housing, 1960. Total automobiles available was then divided by the total number of employed residents reported in the County and City Data Book: 1962.

[b] The principal means of transportation to work during the week prior to the 1960 Census of Population in April.

[c] The arithmetic average of the years nearest to the census year beginning with 1900 in which the population of the central city of group 1 reached 500,000; group 2 reached 100,000; group 3 reached 50,000; and group 4 reached 25,000.

[d] Percent of 1960 housing units in structures built between 1950 and 1960.

Sources: Adapted from data in the U.S. Census of Population and Housing: 1960; the Statistical Abstract of the United States: 1964; and the County and City Data Book: 1962.

Urbanized Area Transportation Types in Each Population Size Group

In Table 11, average values are given for each urbanized area type within each of the four population size groups. Among the largest areas there is only one each of type A (Los Angeles), type B (Houston) and type F (New York).

In all size groups there is a clustering of areas in the Type C and D categories. Together, these two types include those areas having from 70 to 90 automobiles per 100 employed residents. Of the 213 urbanized areas, 130 are in the C and D classes.

Type A and B areas comprise approximately one-third of the total number of areas in each size group except the largest.

In size group 2, Type A consists of five California urbanized areas—San Diego, San Bernardino-Riverside, San Jose and Sacramento plus Phoenix, Arizona (Table 11). As a group, these areas have the highest number of automobiles available per 100 employed residents (107.9 percent) and, correspondingly, the lowest rate of households without cars (12.4 percent) and the highest rate of two-car households (28.7 percent). They are young areas, having recently experienced rapid growth as shown by the large proportions of new housing structures (53.3 percent) and the relatively large proportions (37.1 percent) of young persons under 18 years of age. Generally, they have below-average proportions of low-income families (13.4 percent) and above-average proportions of high-income families (21.0 percent). In addition, all five areas have low population densities. The preceding characterizes the type of large (350,000 to 1,000,000 population) urbanized areas in which more reliance is placed on the automobile for personal transportation than in any other area type in the same size range.

In contrast, the Type D area in the same size group is lowest in the number of automobiles available per 100 employed residents (75.1 percent), has the highest rate of no-car households (24.3 percent), and the lowest rate of two-car households (16.6 percent) for any area of similar size.

Typically, a Type D area is older, has a relatively small proportion (25.1 percent) of its housing units in new structures (built between 1950 and 1960), a relatively low proportion (34.8 percent) of persons under 18 and relatively large proportion of its population in the 65 and over age bracket (9.0). The proportion of low-income families (15.7 percent) in an area of this type is generally slightly higher than a Type A area (13.4 percent) and the proportion of high-income families (17.7 percent) is slightly lower than they are in an area of Type A (21.0 percent).

Eleven of the 13, somewhat less automobile-dependent urbanized areas in size Group 2, are located east of the Mississippi River. Five are in the New York-New England region.

Summary

Urbanized areas have been segrated into six types for the purpose of facilitating analyses of urban transportation and the factors underlying its demand. Degree of reliance on the automobile as the chief mode of personal transportation is the criterion used in establishing the six-area classification. The number of automobiles available for every 100 residents employed measures automobile reliance.

Only one of the 16 largest areas has more than 90 automobiles per 100 employed persons—Los Angeles with 102.

New York, with 56 automobiles for every 100 workers, is a type by itself, being the only area below the 60 to 70 rates of the next lowest group of areas. But then, New York is also unique in terms of the total number of residents (14 million), the size of the land area occupied (1,892 sq mi), the population density of its central city (23,000 persons per sq mi), and in the structure and use of its transportation systems (more freeway and subway mileage than any other urbanized area).

While all urbanized areas are generally auto-oriented, the urbanized areas in the class with the lowest degree of automobile orientation Type E, are all (with the exception of Chicago) located on the eastern seaboard (Fig. 2).

Reliance on the automobile is practically total in the urbanized areas in Type A. These areas are distributed geographically in the Far West, Southwest, and Florida.

(Adapted from data in the U.S. Censuses of Population and Housing: 1960.)

Figure 2. Geographic distribution of urbanized area types with over 150,000 population.

Nine of the 30 areas in this type are located in California, 7 in Texas, 4 in Florida, and the remaining 10 in Arizona, New Mexico, Washington, and Montana.

CONCLUSIONS AND RECOMMENDATIONS

Comparative and correlation analyses were made of the transportation and socioeconomic characteristics of four groups of urbanized areas. The four groups resulted from a stratification, by population size, of 213 urbanized areas. These analyses showed that the socioeconomic factors underlying selected transportation variables often differ between size groups. One factor, population density, was shown to be a major influencing factor only in the largest areas. Newness of housing units as measured by the percent of housing units in structures built between 1950 and 1960 was found to be significantly correlated with several of the transportation variables (Fig. 3). Measures of income (income per capita, percent of families with income under $3,000 per year and those with incomes of $10,000 per year or more) were of little or no significance.

Estimating equations were derived from a stepwise multiple-regression analysis performed on the IBM-1401 computer. The computer program was, therefore, less sophisticated than those used with later model computers. Data transformations were not made for any variables and only one computer run was made. The resulting equations have not been tested against actual data but tests of reliability indicate that the independent variables in Eq. 2 in both Table 8 and Table 9 are significantly related to the dependent variables. One or more independent variables in the remaining equations do not pass the test of significance.

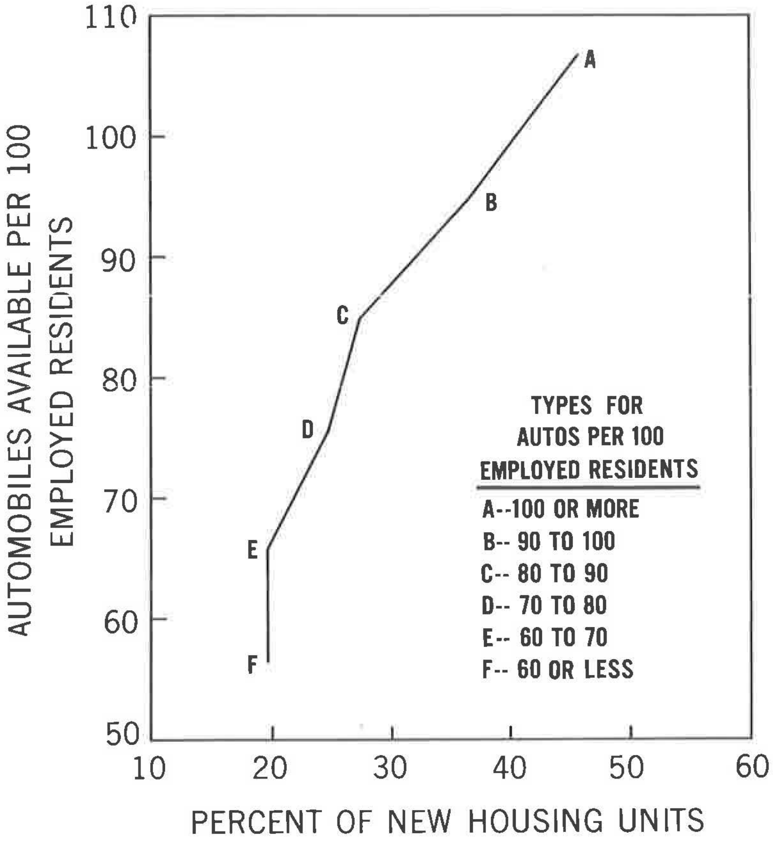


Figure 3. Age of housing units related to automobiles available by urbanized area type.

Some observations on the makeup of the equations should be of interest to persons concerned with urban road traffic estimating. Equations for estimating the automobile-employment ratio in urbanized areas are given in Table 8 (see footnote 1). For areas with a population of one million or more the multiple-regression procedure selected population density and land area for the independent variables in Step No. 2. Population density and land area did not appear as significant variables in estimating the automobile-employment ratio in areas belonging to the three smaller size groups.

Table 8 equations are for estimating the percentage of workers using private automobile or car pool in the work trip in urbanized areas with one million or more population. Other equations were derived (see footnote 1) for estimating the percentage of occupied housing units with no car available in the largest areas. In both cases, the Step No. 2 equations had the same independent variables—population density and the percentage of the population under 18—indicating the collinearity between the proportion of no car households and the proportion of workers using the automobile in the work trip.

A tentative transportation classification of urban areas has been developed based on a measure of automobile reliance (the ratio of the number of automobiles available to the number of residents employed). Grouping the urbanized areas on this basis resulted in six transportation types or classes. Generally, areas in each transportation class display similar social and economic characteristics. These similarities became stronger after the areas in each type were subdivided into the four population size groups used in the earlier analysis.

Recommendations

Research conducted in this phase of the typology task has accomplished its objective of establishing a tentative transportation classification of urban areas. This classification is tentative because the analysis has covered only structure, which is one of four general urban area characteristics—structure, function, form and growth—selected for this task. This is also a static analysis which has indicated relationships, of varying strength, between certain transportation and structural characteristics of the urban areas as they existed in 1960. It remains to be seen whether this classification, or the estimating equations developed in the process, can be of any use to persons involved in forecasting for urban highway purposes. It is recommended, therefore, that the equations given in this report and elsewhere (see footnote 1) be tested for predictability and, to improve their accuracy, adjustments or transformations be made where the need for such are indicated.

A second recommendation is that a stepwise multiple-regression analysis be made of selected variables for the urban areas in each of five area types—A, B, C, D and E.

To determine the changes taking place in the relationships between the variables studied in the static analysis, it is recommended that trend analysis be made of the areas in each tentative class and the growth dimension added to the classification criteria. Time series on economic function would be included in this phase to assess the influence of this factor on changes in urban growth and transportation requirements.

Urban form, the fourth major feature of an urban area, should be analyzed to obtain greater insight into its relationship with transportation and to determine the interrelationships that may exist between these two variables in combination with the other major variables.

In each of the recommended successive phases of research there is an implied awareness that experimentation will often be necessary to develop adequate measures of the factors involved. Also implied is the expectation that the model equations developed in each study will be tested for accuracy in estimating or forecasting certain transportation factors which would serve as controls or guides to traffic engineers and urban planners.

In the research reported here, two examples may be cited in which experimentation yielded certain measures of key variables. A weighted population density was developed to amplify central city congestion where it existed, and relative urban sprawl in areas where it existed. An experimentally developed measure of automobile availability is the automobiles-to-employment ratio which generally overcomes some of the shortcomings of the automobiles to dwelling units or the automobiles per capita ratios.

Predictability of the equations has not yet been tested against actual data, although the equations produced at the 12th step of the multiple-regression analyses had been tested. These 12th-step equations included the variables in the equations given in this report as well as others which were not significant or did not improve the standard error of estimate or the measures of correlation and determination, R and R^2. Comparisons between the actual data and the data predicted by the 12th-step equations give mixed results—close predictions for certain dependent variables in particular urban area size groups but wide differences in other cases.

ACKNOWLEDGMENTS

We wish to express our appreciation to Sidney Goldstein, former Chief of the Economics and Requirements Division, for his material aid. Special thanks go to Edmond L. Kanwit, leader of the sociology task group, for his helpful criticism and suggestions. Significant information and counsel were also received from Martin Fertal, Charles Culp, Carl Swerdloff, and Nathan Lieder of the Office of Planning. Our thanks also for the special help received from Ezio C. Cerrelli of the Automatic Data Processing Division.

REFERENCES

1. Row, Arthur T. A Consultant's Report to the Tri-State Transportation Committee on a Reconnaissance of the Tri-State Region. New Haven, Conn., 1965, pp. 3-7.
2. Chapin, F. Stuart, Jr. Foundations of Urban Planning. In Urban Life and Form, (Werner Z. Hirsch, Ed.). Holt, Rinehart and Winston, New York, p. 233, 1963.
3. Schnore, Leo F. Transportation Systems, Socio-Economic Systems, and the Individual. Transportation Design Considerations, National Academy of Sciences-National Research Council Publ. 841, Washington, pp. 201-12, 1961.
4. Berry, Brian J. L., Simmons, James W., and Tennant, Robert J. Urban Population Densities: Structure and Change. Geograph. Rev., pp. 389-405, July 1963.
5. Levinson, Herbert S., and Wynn, F. Houston. Effects of Density on Urban Transportation Requirements. Highway Research Record 2, pp. 38-64, 1963.
6. Shryock, Henry S., Jr. Population Distribution and Population Movements in the United States. Highway Research Record 2, pp. 65-78, 1963.
7. Zwick, Charles J. The Demand for Transportation Services in a Growing Economy. Highway Research Record 2, pp. 3-5, 1963.
8. Colwell, Robert C. Interactions Between Transportation and Urban Economic Growth. Highway Research Record 2, pp. 6-11, 1963.
9. Jones, Victor, and Forstall, Richard L. Classification of Metropolitan Areas. The Municipal Yearbook, 1963. The Internat. City Managers' Assoc., Chicago, pp. 31-44, 1963.
10. Harris, Chauncey D. A Functional Classification of Cities in the United States. Reprint from Geograph. Rev., Vol. 33, Jan. 1943.
11. Nelson, Howard J. A Service Classification of American Cities. Readings in Urban Geography. (Harold M. Mayer and Clyde F. Kohn, Eds.). Univ. of Chicago Press, p. 139, 1959.
12. Lynch, Kevin, and Rodwin, Lloyd. A Theory of Urban Form. Jour. of the Amer. Inst. of Planners, Vol. 24, No. 4, 1958.
13. Passoneau, Joseph Russell. The Emergence of City Form. Urban Life and Form. (Werner Z. Hirsch, Ed.). Holt, Rinehart and Winston, New York, 1963.
14. Lynch, Kevin. The Form of Cities. Sci. Amer., Vol. 190, No. 4, p. 54, April 1954.

15. Rams, Edwin M. Principles of City Land Values. Academy Publishing Co., Worthington, Ohio, 1964, (especially Chs. 3 and 14).
16. Newling, Bruce E. Urban Growth and Spatial Structure: Mathematical Models and Empirical Evidence. Cornell Univ., Ithaca, N.Y., Unpublished Manuscript, 1965.
17. Pred, Allan. Industrialization, Initial Advantage, and American Metropolitan Growth. Geograph. Rev., Vol. 55, No. 2, pp. 158-185, 1965.
18. Niedercorn, John H., and Hearle, Edward F. R. Recent Land-Use Trends in Forty-Eight Large American Cities. The Rand Corp., Santa Monica, Calif., June 1963.
19. Ogburn, William F., and Duncan, Otis D. City Size as a Sociological Variable. Contributions to Urban Sociology. (Ernest W. Burgess and Donald J. Bogue, Eds.). Chicago, p. 135, 1964.
20. Croxton, Frederick E., and Cowden, Dudley J. Applied General Statistics. New York, Chs. 22 and 24, 1939.
21. Ezekiel, Mordecai. Methods of Correlation Analysis. John Wiley and Sons, Inc., New York, 1930, (especially Ch. 12).
22. Shuldiner, Paul W. Non-Residential Trip Generation Analysis. Dept. of Civil Eng. and The Transportation Center, Northwestern Univ., Evanston, Ill., p. 67, 1965.
23. Greenwood, George W. Traffic Linkage Patterns Between a Metroplitan Area and the Communities Within Its Region of Influence. The Bureau of Community Planning, Univ. of Illinois, in cooperation with Illinois, Div. of Highways and U.S. Bureau of Public Roads, Unpublished, May 1965.

The Urban Panel as a Longitudinal Data Source

R. D. WORRALL, Department of Civil Engineering, Northwestern University

This paper examines some of the problems posed by the lack of an adequate, longitudinal data set describing the locational preferences, daily activity sets and daily travel patterns of urban households, and proposes the use of a permanent response panel—analogous to the consumer panels employed in market research—as a means of collecting such information.

Consideration is given to the need for both long- and short-run data, with particular reference to (a) model structure and calibration, (b) forecasting, and (c) before-and-after studies. A brief description is given of a case study, employing the panel technique, in which data on daily travel behavior were collected for a sample of 50 households for a period of one month. Finally, a scheme is proposed for the continuous monitoring of urban travel demand, integrated within the framework of a continuing urban transportation study.

•THE development of a satisfactory land-use model is constrained by a wide range of considerations. One of the most common of these is the availability of an adequate data set. Existing data systems, particularly those developed prior to the current quantitative focus of model building activity, tend to meet only some of the analyst's requirements. At the macro-level, for example, he is often faced with a dirth of accurate economic and time-series data. At the micro-level there is usually an equivalent lack of detailed, disaggregate information, specified at the level of precision which he requires and couched in a longitudinal rather than a cross-sectional format.

This paper considers the operational feasibility of a specialized data system—the urban panel—designed to provide detailed, time-series information on the daily activity patterns, locational and behavioral preferences and trip-making characteristics of a controlled sample of individual households. The technique is based on the concept of the continuous household response panel, analogous to the consumer panels frequently employed in public opinion and market research. Its potential application to household behavior patterns has important implications both for the structure of future land-use models and for the design and operation of future data systems.

In the present case, emphasis is placed on the application of panel techniques to the study of urban travel behavior. This emphasis is merely a matter of convenience, reflecting the particular background of the author. The paper might equally well have focused on one of the other, parallel topics of interest to the urban model builder.

MONITORING URBAN TRAVEL

It is generally acknowledged that the structure of urban travel is subject to a continuous process of change. Constant modifications in the existing transportation environment—the network of roads, transit lines, and the distribution of potential destination points—coupled with parallel changes in the behavioral preferences and activity sets of the households making up the metropolitan population all exert a continuously changing influence on the daily demand for travel. Some of these changes—or "shocks" as they may be termed—have lasting, long-term effects (e.g., the opening of a new expressway or transit line); others yield only transitory consequences (e.g., a sudden snow storm

---

Paper sponsored by Committee on Land Use Evaluation and presented at the 46th Annual Meeting.

or the breakdown of the family car). In either case their detection and evaluation, under the generic term "induced and diverted traffic," is of considerable interest to the transportation analyst.

Conventional before and after and home-interview surveys provide two possible, but rather unsatisfactory, approaches. Both techniques provide the planner with only a limited, and rather ill-defined, information set based on conditions which pertained to one or perhaps two isolated points in time. They yield no really significant information on the structure of changes in travel behavior. Similarly, continuing volume-count or ridership surveys, though exhibiting the necessary temporal characteristics, do not contain sufficiently detailed information to permit a truly penetrating analysis to be performed.

A fourth line of approach, and that which will be enlarged on here, is the use of a continuous household response panel, designed to continuously monitor changes in travel behavior in a manner analogous to the panel techniques used in public opinion sampling and market analysis. A panel consists essentially of a set of respondent units whose behavior is observed over a continuous term. Assuming knowledge of the sample and population characteristics, the results of a panel study provide the analyst with a valid basis for the detection and analysis of time-dependent behavior [this interpretation of "panel" differs from that used by Davis (1)]. In the present case, the respondent units are households (or equivalent traffic generators), and the behavior under study is the day-to-day changes which occur in the established patterns of urban travel.

In its conventional format, the panel survey may be diagrammed as:

Group	Selection Criteria	Before Measurement	Natural Exposure to Treatment	After Measurement
Test	Nil	T_B	Yes	T_A
Control	Nil	C_B	No	C_A

There is an obvious equivalence between this two-wave design and a simple, non-randomized before and after survey with test and control groups. In this context, "wave" refers to a set of test-control observations made at a single point in time. A two-wave panel, thus, involves two cross sections in time, an n-wave panel involves n successive observations. The distinction should be drawn, however, between the artificially inserted treatments of such a before and after survey, and the natural occurrence of a "treatment" in a panel study completely divorced from the control of the analyst.

The two-wave panel illustrated may be extended to the more common n-wave panel simply by extending the sets of controlled C and test T measurements horizontally, resulting in a format superficially equivalent to a multiple time-series design, but with the provision that successive natural treatments may occur between any two successive time waves.

The design outlined may be modified in a variety of different ways. It need not, in fact, take the form of a formal panel design at all. It could equally well be based on successive random sampling from within a single, well defined selection-frame. Alternatively, it might utilize one of a variety of partial-overlap designs (2). Similarly, the panel units themselves may be defined variously as elements of a purposive quota sample, a stratified probability sample or a multistage cluster sample. The period and intensity of the actual monitoring process may vary, from the extreme of a comprehensive daily record to a regular, once-a-week analysis or a randomized coverage with average frequency of 14 or 28 days.

The permutation of possible designs is almost endless. So is the list of devices which may be used to circumvent the operational problems which arise in such a study (e.g., the decay of an initial panel due to migration and maturation within the sampling frame). No attempt will be made here to treat all of these considerations comprehensively. Rather, attention will be directed toward the design of one or more feasible survey formats; feasible, that is, with respect to cost, complexity and the fundamental requirements of the travel analyst.

SAMPLE DESIGN—SELECTION OF RESPONDENT HOUSEHOLDS

Two major decisions must be made at the outset of any monitoring design: the desired intensity of the sample coverage and the desirability of sample stratification. Consider first some of the problems of sample coverage.

In a conventional home-interview study, a sample size of between 5 percent and 20 percent is aimed for, depending on the size of the urban area (the larger percent pertains to areas of less than 50,000 population, the smaller to areas of over 1,000,000). An equivalent rate would be impracticable and even undesirable in a continuing panel analysis. Two alternatives present themselves: the selection of a much smaller, wholly randomized sample, dispersed uniformly throughout the urban area, or the use of a small number of high intensity, cluster samples in which respondents are concentrated in only one or two different locations. The benefits and disadvantages of the two designs are relatively well defined; however, the choice between them is not quite so clear-cut.

A randomized sample, provided its size is adequate, will generally yield the more efficient estimate of elementary unit variance and the sounder basis for generalized induction. The cluster sample is cheaper to operate (per element) and provides for closer control over the external environment (2). The efficiency of the cluster statistics may if necessary be improved (at an increase in cost) by a process of multistage sampling.

In general, it is the author's feeling that the former, randomized design is likely to provide the more efficient basis for a continuous monitoring operation. Presumably, the analyst is concerned more with the identification of a general temporal effect, pertinent over the entire metropolitan area, than with the behavior of particular clusters within that area. Furthermore, the degree of effective environmental control attainable within a cluster sample, unless the survey is designed for a very explicit locational purpose (e.g., the evaluation of induction and diversion effects consequent on the opening of a new expressway), is likely to be minimal. A more effective general control is provided by the randomization process implicit in the non-cluster design.

Accepting the dispersed, randomized design in principle, one is then faced with the question of its desirability or stratification.

Superficially, stratification appears to be eminently desirable, preferably in terms of stratified random sampling or, if this proves too costly, in terms of purposive, quota sampling. It has been shown (3) that the intensity and, to a lesser extent, the temporal variability of daily travel to and from the home vary with the characteristics of a single household. Theoretically, therefore, in the sense that efficient stratification results in more efficient estimates of sample parameters and also permits more flexible and perhaps more appropriate subsample analysis, a stronger panel design should result from making use of this additional information.

Unfortunately, however, this is not the case. An efficient stratification scheme is predicated on the identification of a limited number of pertinent variables, all of which relate closely to the phenomena under study and all of which are based on an available and reasonable recent information set.

In the case of travel analysis, a variety of parameters act to influence the behavior of a single household—family size, structure, income level, occupation, maturity—the importance of each varying with the precise behavior under analysis. The nature and diffusion of these characteristics generate two serious problems.

Information on many of the pertinent parameters is not readily available in conventional sampling frames. (The census, for example, contains neither disaggregate measures of household structure nor of stage in the family life cycle, both of which are of significant importance in determining the frequency of daily trip-making from the home.) Similarly, frame data generally used in stratification and quota design—age, sex and race measurements—are not particularly pertinent to the study of household travel. This, therefore, raises a serious problem of frame selection and stratification control.

Even if an adequate sampling frame is identified, the diffusion of pertinent variables introduces a second and equally serious combinatorial problem. Consider, for example, six variables divided into three strata each. This yields a requirement for 3^6 or 729 separate stratification cells. Given that considerably more than one observation is required per cell, the total sample size rapidly reaches astronomical proportions. A variety of extremely complex, stratified probability designs are proposed in the literature (2) to overcome this latter problem. None of these, however, is particularly pertinent to the present discussion. Conventional, simplistic stratification is virtually meaningless.

There remain two alternative lines of approach: either a purposive, nonrandomized design, subject to all of the consequent biases and external invalidities of non-experimental research, or the use of a simple, non-stratified but wholly randomized selection procedure. The optimal choice between these two cannot be generalized usefully. For some panel analyses, the purposive design is most appropriate (e.g., the study of reaction to a rapid transit innovation where the effect is postulated to be restricted to a limited subpopulation); in others, such as the generalized monitoring scheme proposed here (where it is intended to induce general conclusions from specific survey data), a wholly randomized design is indicated.

In each case, considerations of cost dictate a sample size no larger than 2,500 households, whereas requirements of statistical efficiency, as based on the variability of behavior observed in a 4-week case study of 104 households in Skokie, Illinois (see Appendix), suggest a minimal, non-supplemented sampling ratio of not less than ¼ percent to ⅛ percent. These latter figures are theoretical estimates based on an assumption of no significant taxonomical structure. There is evidence to suggest that some such structure may well exist, based on an analysis of the case study data referred to previously. Should further research, based on the initial operation of a randomized panel sample, reveal the existence of a meaningful taxonomy of household types, then these may be used as a foundation for a more efficient stratification design and hence a reduction in the total sample size.

So far it has been tacitly assumed that the monitoring operation is to be performed only on household travel and not on a combination of residential, industrial and commercial trip-making. Some brief justification for this assumption is in order.

Residential based travel, particularly if allowance is made for the individual links in each home-home trip sequence, accounts for more than 80 percent of the total daily travel in an urban area. Wholly commercial and industrial trip-making adds up to only 20 percent or less of the total. One may logically argue, therefore, that any continuous monitoring of nonresidential traffic is likely to yield only an insignificant, and possibly rather expensive, marginal return. Furthermore, the extended study of a firm's distributory activities at the level of detail required here, makes for a considerable imposition on its traffic department. It is in fact virtually impracticable to collect such data on a continuing basis (4).

Should it be considered essential that data on nonresidential travel be made available to supplement that derived from the household sector, it is suggested that a separate subsample of industrial and manufacturing concerns be developed, which would then be subjected to a considerably less intensive and more specific surveillance. In particular, it is suggested that the onus of recording the necessary information be allowed to fall directly on the survey staff rather than be left to the firm's employees.

In view of the major logistical problems involved and the probability of only marginal returns, no further consideration will be given here to the monitoring of nonresidential travel.

CHOICE OF SAMPLING FRAME

A wide variety of alternate sampling frames are available to the urban survey designer, ranging from census records, city directory or voters' lists to utility company files and the results of recent specialized surveys. A detailed discussion of each of these and the appropriate techniques to follow in their use would be inappropriate here. Extensive discussions are given in the methodological literature, particularly by Kish (2, Ch. 8, 9, 11) and Hansen et al (5). However, a number of brief, general comments may be made at this point.

Probably the most efficient sampling frame for a monitoring study is a recent home-interview traffic survey. Such a survey is generally based on an equal probability sample selection method or systematic sample drawn from a completely updated dwelling list. Its use also has the advantage both of tying in the survey directly with a previous data collection exercise and also providing some additional base data for sample control and possibly for sample stratification (the latter comment presumes the identification of the meaningful household taxonomy discussed earlier).

If a survey is unobtainable, then refer to one or the other of the standard frames mentioned. In this event, a variety of problems are likely to be encountered—missing or duplicate listings, ambiguous and rare elements, incompatible subsample sizes and so on. Each of these is treated in detail in the literature, for example, by the use of multistage "compact-segment" techniques (2, pp. 313-315).

Perhaps the most serious frame problem which is likely to arise in a monitoring design is that of sample maturation. Approximately one in every five American families changes its place of residence each year. This means that over a period of two years migration alone may reduce the original sample base by 35 percent. As Sobol (6) has shown, response decay may increase this figure still further, while similar effects may arise due to population growth, births and deaths, and marriages.

A number of gimmicks may be employed to reduce the bias introduced by such maturation processes. Kish (2, pp. 472-474) has suggested the use of dwelling units or compact segments—i.e., clusters of adjacent dwelling units whose composition but not periphery may change over a period of time—rather than families as the basic repetition unit. Sharp has suggested the matching of neighboring household pairs (7) and several writers have discussed techniques for continuously updating existing sampling frames. The final choice is up to the individual analyst. No one technique is uniquely preferable to any other, although one may perhaps argue in favor of Kish's compact-segment technique in the case of a repetitive random sample design and his fixed dwelling unit method in the case of a formal panel study.

SURVEY FORMAT

Given that an acceptable sample has been developed, one further fundamental decision remains to be made. This centers on the selection of formal panel techniques vs the use of repetitive random samples or partial overlap designs. (In a repetitive, randomized design the same single set of households is not included at each stage of the survey. Separate random samples are drawn for each survey wave from a common sampling frame.) The primary benefits to be gained from a panel study (in which the same, randomly selected households are subjected to continuing analysis throughout the study period) are an improvement in statistical inference (2, p. 475), a reduction in overall costs and a considerable increase in the total level of information obtained. Counterbalancing these advantages is a problem due to sample and response decay, sample maturation, surveillance and possible external invalidity.

The severity of these problems is significantly reduced in both of the alternative designs cited. The use of repetitive random samples, for example, removes all question of sample maturation and response decay. It also allows for a relatively high degree of external validity, especially if the randomized sample is inserted into a continuing, quasi-experimental design, such as "separate sample, pretest-post-test control" design. A partial overlap design (in which a specified sample turnover—say, 25 percent or 33 percent—occurs on each survey wave) similarly avoids all major problems of sample maintenance and surveillance. In neither case, however, are the total survey costs as low nor the total amount of longitudinal data obtained as great as for the panel study.

As in the case of the sampling frame, there is no simple, optimal choice between the three possible formats. Each has its own peculiar advantages and disadvantages, none of which significantly outweigh those of the other two. In fact, for a continuing, monitoring operation in which a major concern is the development of a longitudinal data set defining the temporal stability of household travel, a case may be made for the simultaneous use of all three, with the panel study forming a core around which supplementary randomized or overlapping designs may be structured. This is the approach selected here and illustrated subsequently in an example panel design.

SURVEY OPERATION

Perhaps the most tempting pitfall in any continuing survey is the collection of too much information. It is extremely important, therefore, that the range and intensity of the monitoring questions, the frequency with which they are asked and particularly their pertinence to travel analysis be strictly controlled. Considerable caution should be exercised lest the monitoring operation turn into a general behavioral surveillance, with all of the obnoxious implications that such a process bears for the encroachment of privacy.

Given that the core of the proposed monitoring operation is to be formed by a continuing panel study, it is suggested that the self-administered travel diary illustrated in the Appendix and used successfully in the case study described be adopted as the primary data collection mechanism. The diary is designed for flexible use over periods varying from one day up to several weeks. The range of information requested is compact and comprehensive, and the diary format has been proven acceptable to a sample of laymen. Its use in the case study did not reveal significant response bias, either in terms of its acceptance by different types of household or of its sustained use as a travel record over a period of several weeks. Where additional information over and above that recorded in the diary is required, supplementary questionnaires could be used. It should be stressed again, however, that the temptation to ask too many supplementary questions should be vigorously avoided.

Use of the personal travel diary as the primary data collection mechanism gives rise to a potential problem of data redundancy. Although a recent empirical analysis emphasized the essential temporal instability of household travel, it also revealed a considerable degree of daily repetition, particularly from the viewpoint of the respondent who may be forced to record the details of a single, repetitive trip for each of several days for several weeks (3). Two suggestions may be made to reduce the labor of this exercise.

First, those trips which are known to be highly repetitious—i.e., the journey to work and the journey to school may be pre-coded—and, once their regular pattern has been established, full details may be requested only on the occasion of a divergence from this pattern. This technique was tested on ten of the sample families included in the case study and was found, subject to adequate surveillance, to yield acceptable results.

Second, a much more effective device in reducing the redundancy and labor of daily data recording is to vary the period and technique of the monitoring process. During the case study, data were collected on a day-to-day basis over the entire survey period. Obviously, while this is feasible and in fact desirable for a limited 4-week study, it is both undesirable and unnecessary in a continuing study.

It is suggested, therefore, that an operational format be adopted in which the sample families are requested to maintain a daily diary for an initial period of 3 to 6 weeks. During this time they would be subjected to training and surveillance at the rate of two calls per week. At the end of this period, the contact frequency would drop to an average of two days per month, with the respondents being requested to maintain a daily trip record only for those two specific days. The selection of the successive contact dates would be randomized within the selected sampling frame.

Primary reliance would be placed throughout on the self-administered questionnaire. After the initial 3- to 6- week period each successive contact would be made by mail (i.e., a self-administered diary is mailed to the family two or three days ahead of the sample day), supplemented by a telephone call on the immediately preceding day and a subsequent personal call by a member of the survey staff to collect and vet the completed form. During this personal contact, a series of brief checks would be performed on the completed diary by the staff member. Further detailed checks would take place in the survey office as part of the audit procedure (see Appendix). Supplementary contacts, as necessary, would be made primarily by mail or by telephone. Minimum reliance would be placed on extensive personal interviews. It is suggested that considerable use be made of telephone contacts to detect the consequences of a particular, short-lived environmental "shock" and possibly to evaluate public reaction to a particular experiment in system control. An example of the latter may be the efficacy of helicopter surveillance of freeway congestion and the broadcasting of alternate routes over commercial radio channels.

The panel study outlined makes up the core of the proposed monitoring operation. It provides the nucleus of a capability to detect and evaluate changes in travel demand on a continuing period, and at the same time measures public reaction to transportation innovation and change. Extended over a period of several years, it would provide a base of longitudinal data of considerable utility to the transportation analyst. Its format is entirely flexible. The duration of the initial, intensive survey period might vary from 3 to 6 weeks; the frequency of subsequent contact might range from one day/month to one day/week. (Extension or subtraction beyond these extremes is likely to result in significant over or under collection of information.)

Centered around this core would be a variety of less frequent, non-panel designs. The structure of these supplementary studies might also vary considerably, depending on the intensity of the panel coverage, the precise information requirements, and the size of the urban area. One might consider a three monthly, randomized set of two-day samples, or an annual, one-week study, again based on a wholly randomized sample selection procedure. Similarly, the supplementary studies might incorporate a degree of successive overlap—say, 25 percent or even 50 percent—and they might involve either the use of the standard diary format or a supplementary, simplified questionnaire. Contact might be made by mail, by telephone (preferable for the simpler forms of survey), or by conventional home interviews. The extensive use of the home-interview technique obviously suffers from the primary disadvantage of high relative cost. It is, however, likely to provide the best return on a given investment for the more complex forms of investigation, while the maintenance of a permanently employed, highly trained survey staff introduces considerable scale economies into a continuing transportation study with obligations to provide limited amounts of up-to-date survey data. The coordination of the proposed monitoring scheme within such a continuing transportation study is discussed in the next section.

Some form of respondent incentive was found in the case study to be essential to the efficient operation of a permanent monitoring scheme. It is suggested that using trading stamps redeemable at a local store for goods of the respondent's choice, and employed effectively in the case study, be utilized for this purpose.

A truly accurate estimate of the level of payment necessary to maintain a sustained, rather than a month-long response period cannot be deduced directly from the case-study findings. In the absence of more adequate information, it is suggested that a payment level of 50 trading stamps/day plus a terminal bonus of 500 stamps be utilized during the initial 3- to 6-week period, followed by regular monthly payments of 250 stamps/month. This is equivalent (for the stamps used in case study) to an annual expenditure per family, for the first year, of $13.81, followed by $7.50 for each succeeding year.

DATA MANAGEMENT CONSIDERATIONS AND INTEGRATION OF A MONITORING STUDY WITHIN A CONTINUING TRANSPORTATION STUDY

A continuous monitoring scheme generates a rather terrifying set of data management problems. The author's experience with a sample of only 104 households suggests strongly that any expansion of the diary technique to incorporate 500 to 1000 household units should be accompanied by a relatively sophisticated data-handling capability. A detailed discussion of the necessary data management format would take more space than is available here. However, a number of general comments may be made.

The data-management system may take one of a number of different forms. It should, however, consist of the following: (a) a data bank, in which a continuing set of basic travel information is retained on a permanent basis; (b) a complementary environmental record containing details of the occurrence of specific environmental shocks; (c) a sequential analysis set, in which running "averages" are maintained of a set of pertinent travel statistics (e.g., daily trip-making variances, visitation rates, etc.).

The nucleus of the system would be the travel data bank. This would contain details of each person- and group-trip made by each panel household, summarized in terms of the trip's origin and destination point, mode, start and end time, and trip purpose (the latter information being coded according to a two-digit classification). Permanent base data on the characteristics of the sample families would also be retained in the same source, with the environmental and sequential-analysis records aligned in concert with the structure of the main data bank.

The previous remarks are intended to be no more than the most cursory outline of an extremely complex operation. They serve merely to underline the essential need for an adequate data-management capability. This requirement in turn leads directly into a consideration of the integration of the proposed monitoring scheme within the framework of a continuing transportation study.

It may be validly assumed that any major transportation planning agency will already possess, possibly in reduced form, an ability to resolve large quantities of basic travel data. Integration of the monitoring scheme within the framework of such a study, therefore, is likely to reduce considerably the initial problems of data processing. More importantly, such an integration is both logical and also of obvious utility to the concept of a continuing planning operation. In the case of an established transportation study, the monitoring operation may be viewed as a device to measure the temporal transformation of travel demand away from that observed for the base year. Equally, the development of a longitudinal data set may provide a means for verifying, and if necessary modifying, the basic set of predictive relationships.

In the case of a newly initiated study, yet to collect its basic data set, the concept of an extended monitoring scheme is even more intriguing. Here there is a chance to modify and experiment with the established dogma of data collection. One may conceive of a variety of possible sampling formats: a spatially diffuse, randomized, cross-section sample based on a 2-, 5- or 7-day trip-record and a sampling ratio of only 3 percent to 10 percent rather than the usual 5 percent to 20 percent; a conventional one-day home-interview sample, supplemented by spatial clusters of 25 to 30 panel households; a combination of an original 7 percent one-day, base study; a regular 6-monthly or annual random 3-day sample (at a rate of, say, ¼ percent for a city of 400,000); and a continuing 500 household panel analysis. The possible permutations are virtually endless.

In each case, the initial extended base study would provide both a more accurate estimate of an individual household's (or zone's) average daily trip-production and also of the necessary spatial distribution of trip linkages required as input into the initial planning process (3). The extended monitoring and supplemental surveys provide a measure of the changing pattern of travel demand and of the stability of the predictive estimates developed from the base data. Together, the two data sets represent a considerably more complete and meaningful information package than that obtained from a conventional cross-sectional survey. The additional costs of the longitudinal surveys, though high, would be defrayed in part by the lower expenditures necessary for the reduced-sample base studies and, hopefully, by an increased forecasting accuracy.

The preceding discussion has been couched deliberately in very general terms. No mention has been made of specific sampling rates, sample sizes or design schedules. Obviously, a useful discussion in these terms may be presented only in the context of a specific design situation—a city of 400,000 population, a metropolitan area of 3,000,000—and a given set of data requirements. General statements regarding desirable sample sizes and sample combinations, separated from a specific design context, are meaningless. The final section of this paper, therefore, is devoted to an example of a possible monitoring scheme applied to the limited case of the Village of Skokie, Illinois.

TABLE 1

ESTIMATED ANNUAL COSTS FOR PROPOSED SKOKIE STUDY (FIRST YEAR)

Initial Contact Costs (assumed acceptance rate = 50%)	
Drawing and validating sample	$ 500.00
Interviewer time, at $2.00/hr (including travel cost)	1,130.00
Preparation and printing of survey forms	700.00
Mail, telephone and miscellaneous	50.00
Secretarial time	800.00
Total cost	$ 3,180.00
Avg. cost/accepting household	$ 14.42
Sustained Surveillance Costs (assumed "medium level surveillance" for first 4 weeks, 2 calls/month thereafter)	
Interviewer time, at $2.00/hr (including travel cost)	$10,740.00
Office audit, coding and punching of data	28,760.00
Total cost	$39,500.00
Avg. cost/household/day	$ 3.60
Reimbursement Costs (assumed repayment rate 50 stamps/day for 4 weeks + 500 bonus and 250/month thereafter)	
Total reimbursement cost	$ 2,552.00
Avg. cost/household	$ 11.60
Repetitive Random Sample (assumed acceptance rate = 75%)	
Drawing and validating sample	$ 500.00
Interviewer time, at $2.00/hr (including travel cost)	2,950.00
Preparation and printing of survey forms	100.00
Mail, telephone and secretarial time	500.00
Office audit, coding and punching of data	1,200.00
Reimbursement cost (at 50 stamps/day)	166.00
Total cost	$ 5,416.00
Avg. cost/household/day	$ 4.10
Total annual costs	$50,648.00
Annual cost/household/day	$ 4.11

TABLE 2

ESTIMATED ANNUAL COSTS FOR PROPOSED SKOKIE STUDY (SECOND YEAR)

Sustained Surveillance Costs (assumed 2 calls/month)	
Interviewer time	$ 8,540.00
Office audit, coding and punching data	10,560.00
Total cost	$19,100.00
Avg. cost/household/day	$ 3.62
Reimbursement Costs (assumed repayment rate of 250 stamps/month)	
Total reimbursement cost	$ 1,650.00
Avg. cost/household	$ 7.50
Repetitive Random Sample (as for first year)	
Total cost	$ 5,416.00
Avg. cost/household/day	$ 4.10
Sample Renewal Costs (estimated sample decay 25 percent per annum)	
Maintaining and updating sample frame	$ 250.00
Interviewer time	300.00
Mail, telephone and miscellaneous	25.00
Total cost	$ 75.00
Avg. cost/accepting household	$ 10.42
Miscellaneous	
Preparation and printing of survey forms	$ 200.00
Secretarial time	200.00
Total cost	$ 400.00
Total annual cost	$27,141.00
Annual cost/household/day	$ 4.11

A POSSIBLE MONITORING STUDY FOR SKOKIE, ILLINOIS

Skokie lies 14 miles northwest of the Chicago Loop. It has a current population of approximately 68,000, divided into some 22,000 households. An 11 percent cross-sectional home-interview sample, drawn by the Chicago Area Transportation Study in 1964, yielded 2,101 separate dwelling units. The cost of the data collection and coding operations associated with this conventional, one-day home-interview study was roughly $25,000.00 (3).

The following longitudinal format is proposed as a supplement to the 1964 CATS survey:

1. Continuing Panel Study

Sampling ratio	1 percent
Sample size	220 households
Operational format:	
Initial survey period	4 weeks
Continuing survey rate	2 days/month
Surveillance level:	
Medium	(biweekly housecalls, see Appendix)
Repayment level:	
Initial period	50 stamps/day + 500 bonus
Continuing period	250 stamps/month
Assumed sample decay rate	25 percent per year

2. Repetitive Random Sample

Sampling cycle	3 months
Sampling ratio	½ percent
Annual sample size	440 households
Survey period	3 consecutive weekdays
Repayment level	50 stamps/day

In each case, the survey mechanism is the self-administered travel diary, supplemented by semiweekly monitoring calls in the initial phase of the panel study, and in-

TABLE 3

ESTIMATED ANNUAL COSTS FOR A 2000 ELEMENT MONITORING SCHEME (FIRST YEAR)

Initial Contact Costs (assumed acceptance rate = 50%)	
Drawing and validating sample	$ 1,500.00
Interviewer time, at $2.00/hr (including travel cost)	8,500.00
Preparation and printing of survey forms	3,500.00
Mail, telephone and miscellaneous	250.00
Secretarial time	3,000.00
Total cost	$ 16,750.00
Avg. cost/accepting household	$ 8.38
Sustained Surveillance Costs (assumed "medium level surveillance" for first 6 weeks, 2 calls/month thereafter)	
Interviewer time, at $2.00/hr (including travel cost)	$172,000.00
Office audit, coding and punching of data	250,000.00
Total cost	$422,000.00
Avg. cost/household/day	$ 3.29
Reimbursement Costs (assumed repayment rate 50 stamps/day for 6 weeks + 500 bonus and 250/month thereafter)	
Total reimbursement cost	$ 27,620.00
Avg. cost/household	13.81
Total annual survey costs	$466,370.00
Avg. annual cost/household/day	$ 3.64

TABLE 4

ESTIMATED ANNUAL COSTS FOR A 2000 ELEMENT MONITORING SCHEME (SECOND YEAR)

Sustained Surveillance Costs (assumed 2 calls/month)	
Interviewer time	$ 46,000.00
Office audit, coding and punching data	128,000.00
Total cost	$174,000.00
Avg. cost/household/day	$ 3.60
Reimbursement Costs (assumed repayment rate of 250 stamps/month)	
Total reimbursement cost	$ 15,000.00
Avg. cost/household	$ 7.50
Sample Renewal Costs (estimated sample decay 25 percent per annum)	
Maintaining and updating sample frame	$ 750.00
Interviewer time	2,060.00
Mail, telephone and miscellaneous	75.00
Total cost	$ 2,885.00
Avg. cost/accepting household	$ 5.75
Miscellaneous	
Preparation and printing of survey forms	$ 2,500.00
Secretarial time	3,000.00
Total cost	$ 5,500.00
Total annual cost	$197,385.00
Avg. annual cost/household/day	$ 4.11

dividual house calls for each continuing panel survey contact and also for each random sample selection (for the latter, personal visits are made both at the start and end of each 3-day period). The sampling frame for both surveys is the 1964 CATS home-interview study. An annual decay rate of 25 percent is assumed within the panel sample.

The approximate annual costs of the two surveys, exclusive of data analysis and the salaries of professional personnel, are given in Tables 1 and 2. Table 1 refers to the first year of operation and Table 2 to all subsequent years. The unit survey costs in each case are based on those observed for the case study described in the Appendix.

Note the relatively high annual expenditure involved even for a small-scale study such as that illustrated here—$50,648.00 in the first year and $27,141.00 in each subsequent year. These figures are exclusive of any data analysis costs or professional salaries. Estimation of a single, lump sum of $40,000.00 to cover these two items (probably a little conservative in terms of computer costs) yields estimated total annual costs for the first and second years of slightly more than $90,000.00 and slightly more than $67,000.00 respectively.

These figures do, of course, represent a somewhat artificial situation. It is unlikely that a full-scale monitoring scheme would ever be seriously considered for an area the size of Skokie. Much more probable is the application of the method to an entire metropolitan area such as Chicago.

In this case, one may argue that a much larger sample, say, 2000 households would be required. Tables 3 and 4 summarize the estimated costs of an expanded panel study (no repetitive random sampling) applied to such a sample. In this case, allowing for an estimate of $100,000.00 for the necessary computer capability (again a rather conservative figure), the total survey expenditure during the initial 12-month period comes to $566,370.00 or an average figure of $283.14 per household. For subsequent years, allowing for an average annual sample decay rate of 25 percent and a similar computer charge, the equivalent figures are $299,450.00 and $149.73.

Appendix

THE HOUSEHOLD PANEL: A CASE STUDY OF TRAVEL BEHAVIOR

From November 1965 to March 1966, a case study was performed of the daily travel of a sample of families residing in Skokie, Illinois. Skokie is a suburban dormitory community of approximately 68,000 people (Fig. 1). The population is predominately middle class, more than two-thirds of the households living in single family dwelling units and slightly less than one-third in flats, apartments or townhouses. The study was designed both to test a set of hypotheses concerning the temporal patterns of household travel and also to evaluate the suitability of the personal travel diary as a survey mechanism.

A total of 104 families were requested to maintain a record of their daily trip-making for a period of four weeks. Information was requested on the number of people involved, the points of origin and destination, start and end times, travel mode and trip purpose for all trips made by the household during the duration of the survey.

The survey mechanism in each case was the self-administered travel diary. Three different diary formats were tested, two designed for completion by the individual and one by the entire household. Figure 2 illustrates format No. 3 (household diary), which yielded the most successful results. Note the open-ended question on trip purpose.

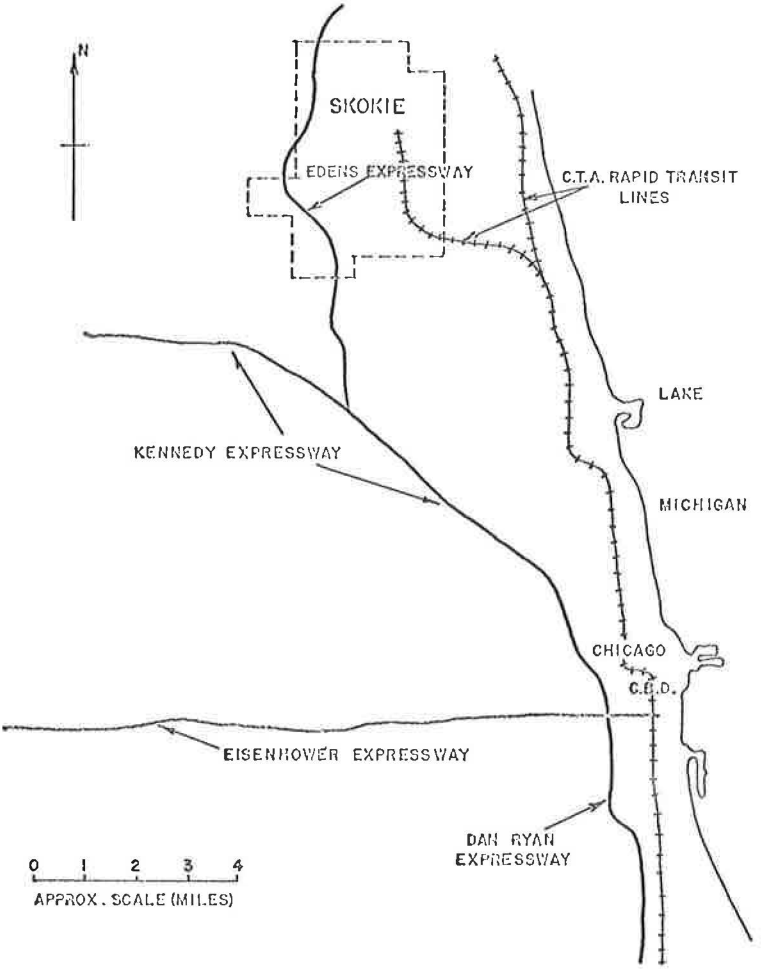
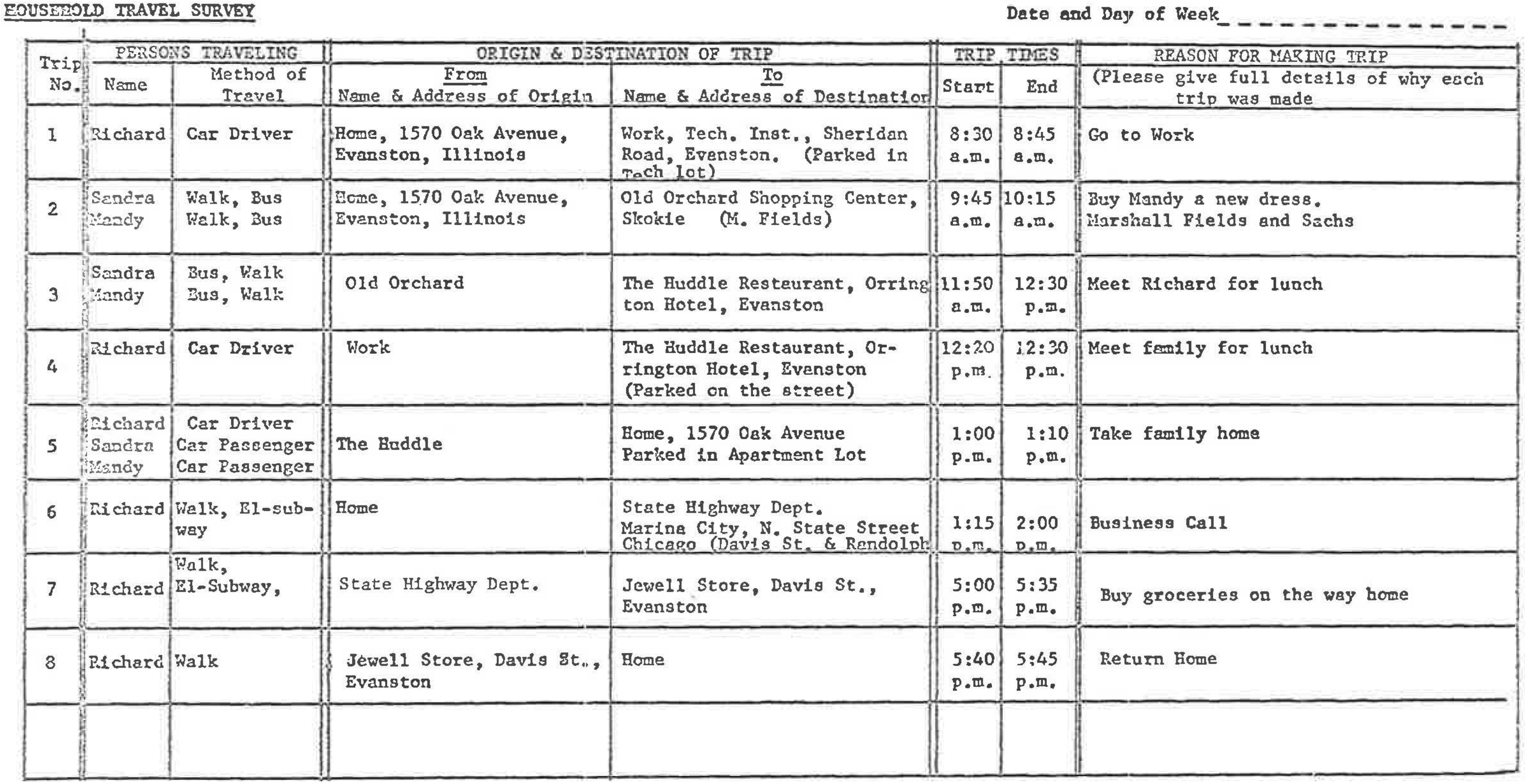


Figure 1. Location of Village of Skokie, Illinois.



HOUSEHOLD TRAVEL SURVEY

Date and Day of Week_______________

Trip No.	PERSONS TRAVELING		ORIGIN & DESTINATION OF TRIP		TRIP TIMES		REASON FOR MAKING TRIP
	Name	Method of Travel	From Name & Address of Origin	To Name & Address of Destination	Start	End	(Please give full details of why each trip was made
1	Richard	Car Driver	Home, 1570 Oak Avenue, Evanston, Illinois	Work, Tech. Inst., Sheridan Road, Evanston. (Parked in Tech lot)	8:30 a.m.	8:45 a.m.	Go to Work
2	Sandra Mandy	Walk, Bus Walk, Bus	Home, 1570 Oak Avenue, Evanston, Illinois	Old Orchard Shopping Center, Skokie (M. Fields)	9:45 a.m.	10:15 a.m.	Buy Mandy a new dress. Marshall Fields and Sachs
3	Sandra Mandy	Bus, Walk Bus, Walk	Old Orchard	The Huddle Resteurant, Orring ton Hotel, Evanston	11:50 a.m.	12:30 p.m.	Meet Richard for lunch
4	Richard	Car Driver	Work	The Huddle Restaurant, Orrington Hotel, Evanston (Parked on the street)	12:20 p.m.	12:30 p.m.	Meet family for lunch
5	Richard Sandra Mandy	Car Driver Car Passenger Car Passenger	The Huddle	Home, 1570 Oak Avenue Parked in Apartment Lot	1:00 p.m.	1:10 p.m.	Take family home
6	Richard	Walk, El-subway	Home	State Highway Dept. Marina City, N. State Street Chicago (Davis St. & Randolph	1:15 p.m.	2:00 p.m.	Business Call
7	Richard	Walk, El-Subway,	State Highway Dept.	Jewell Store, Davis St., Evanston	5:00 p.m.	5:35 p.m.	Buy groceries on the way home
8	Richard	Walk	Jewell Store, Davis St., Evanston	Home	5:40 p.m.	5:45 p.m.	Return Home

Figure 2. Typical sheet from household travel diary (diary format No. 3).

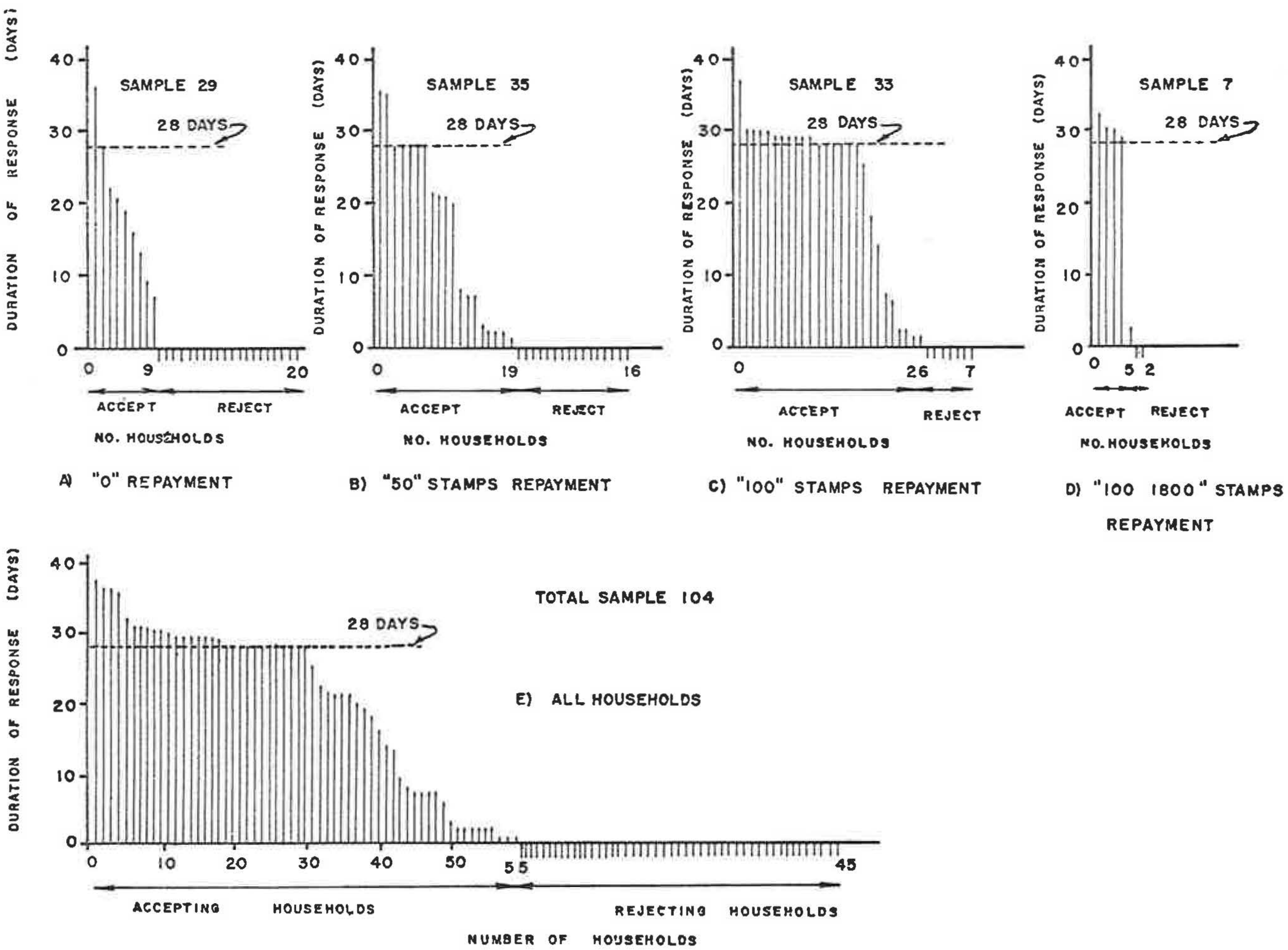


Figure 3. Distribution of sustained response rates by repayment level.

To evaluate respondent sensitivity to reimbursement and supervision, the total sample was subdivided into four repayment categories and three different surveillance groupings. Reimbursement in each case took the form of trading stamps redeemable for items of the respondent's choice at a local store. Surveillance techniques varied from regular house calls by the survey staff, to telephone contacts, logical consistency checks and the application of a simple recall questionnaire. Supplementary data on household structure, personal travel preferences and regular activity patterns were obtained through separate questionnaires.

RESULTS

Figure 3 shows the distribution of initial response to the request to maintain a travel diary. Out of the 104 households contacted, 45 rejected the request outright. Of those accepting, 30 maintained a comprehensive diary for the desired period of 28 days; the remainder kept a record for periods varying from 2 to 26 days.

The main factor influencing initial response was the prospective level of reimbursement. Once a family had agreed to cooperate, the level of information obtained was dependent primarily on the degree of surveillance provided by the survey staff. No significant bias was apparent in the socioeconomic structure of the families accepting and rejecting the initial survey contact. Simiarly, socioeconomic characteristics appeared to have little or no bearing on the maintenance of a sustained response.

It was found that as a minimum each household had to be visited by a member of the survey staff once every 3 to 4 days. At each visit the interviewer collected and checked the diaries for the previous 3 to 4 days, and also provided the family with a fresh set of blank forms. At the same time an extra series of predesigned questions were directed at the respondents to determine whether any trips had been inadvertently omitted from the daily records.

On receipt of the completed diaries in the survey office, a further, more comprehensive consistency check was performed, including both a repeat of the checks performed by the interviewer, and also an audit to determine possible errors of omission (i.e., a total failure to report certain kinds of trips), and, as the survey progressed, a check on the variance of the family's behavior from one week to the next. A detailed specification of the office monitoring procedure is given in Table 5. If any major errors or ambiguities were identified as a result of these office checks, the interviewer was instructed to clarify them during his next house call. As a supplement to the house calls, contact was also maintained daily with each family by telephone.

Almost all the cooperating families, after a period of approximately one hour's coaching and 2 or 3 days experience with the use of the diary format, showed themselves to be capable of providing an accurate and concise record of their day-to-day travel, provided that they were constantly reminded of the need to do so.

The commonest recording deficiencies (based on the result of the audit procedures) were definition of destination points (this occurred in approximately 6 percent of all ultimately recorded trips), specification of trip purpose (10 percent of all cases), and a failure to record approximately 8 percent of trips back to the home. Errors of total omission (i.e., the initial failure to record a trip in any form) varied considerably from family to family. Overall, the average rate was slightly less than 5 percent of all ultimately repeated trips.

The use of a self-administered questionnaire raises serious problems of possible reactivity (i.e., the influencing of behavior due to the fact that it is under study). This suggests that the data obtained from such a study may not represent a valid statement of "free" or unconstrained behavior.

Although accepting the validity of this comment, the writer feels that its importance should not be overplayed. People are essentially creatures of habit in their daily tripmaking. Further, they exhibit a strong predilection for the pursuit of apparent irrationalities—they do not always choose the shortest or quickest route, they do not minimize their total travel effort nor do they usually maximize their expected net benefits. Rather, they tend to pursue a pattern of behavior which satisfies their demands and does not create too high a level of inconvenience.

TABLE 5

OUTLINE OF OFFICE MONITORING PROCEDURE

General

- Check for complete diary set by day and by person

Individual Trip Records—Inclusion of Data

- Check for inclusion of traveler names
- Check for complete home-home round trips for each person
- Check for ambiguity of person trip participation
- Check for method of travel, by person and by trip
- Check origin and destination specification
- Check for successive origin and destination links
- Check trip-start times, trip-end times
- Check trip-time ambiguities
- Check for trip purpose details
- Note any ambiguities and illegible statements

Individual Trip Records—Completeness

- Construct daily travel diagram, point to point, by person and by group; check for consistency
- Construct trip-timing diagram, by purpose and by person; check for consistency
- Cross-check personal and household travel records (where applicable), for group-trip participation
- Compute running-average, daily trip-making rates, for 9 major trip-purposes and for subset of minor purposes; check for consistency and deviations from average trip-making frequency; check for work trips (one or more/worker/day), school trips (one or more/child/day), shopping trips and social-recreational trips

The maintenance of a travel diary, if it in fact has any reactive effect, probably tends to simply underline the more blatant irrationalities of behavior. The diarist, for example, may note that consolidation of several independent trips into a single "linked" trip might reduce his total travel effort. Alternatively, he may note that increased efficiencies would result from a change in the timing of his trips or perhaps from the use of another travel mode. In all events, any significant reactive consequence of diary maintenance should be reflected in a significant change in the pattern of daily travel over the period of diary keeping.

In no case was such a change observed—in terms either of an increase or decrease in total trip-making, a displacement of trip timing or a transference between travel modes—for any one of the sample households included in this study. This is not to say categorically, of course, that no reactivity occurred. The diary mechanism may simply have been too crude to detect any such effects, or else the period of the survey may have been too short for them to become apparent.

REFERENCES

1. Davis, J. A. Panel Analysis: Techniques and Concepts in the Interpretation of Repeated Measurements. Univ. of Chicago, 1963.
2. Kish, L. Survey Sampling. John Wiley and Sons, New York, 1965, pp. 148-150.
3. Garrison, W. L., and Worrall, R. D. Monitoring Urban Travel. Report submitted to NCHRP, Sept. 1966.
4. Yates, R. Sampling Methods for Censuses and Surveys, 3rd Ed. Charles Griffin and Co., London, 1960.
5. Hansen, M. H., Hurwitz, W. N., and Madow, W. G. Sample Survey Methods and Theory. Vol. 1 and 2, John Wiley and Sons, New York, 1953.
6. Sobol, M. G. Panel Mortality and Panel Bias. Jour. Amer. Statist. Assoc., Vol. 54, pp. 52-68, 1959.
7. Sharp, H., and Feldt, A. Some Factors in a Probability Sample Survey of a Metropolitan Community. Amer. Soc. Rev., Vol. 24, pp. 650-661, 1959.

Application of Information System Concepts to Transportation Planning

KENNETH J. DUEKER, Department of Civil Engineering and Department of Urban and Regional Planning, University of Wisconsin

AS defined by Bourne (1), "An information system is . . . a complex of people, equipment, and procedures working together to provide needed information to a group of users." This definition emphasizes a total system to provide information, not merely a computer system. The concern here is with man-machine systems to provide information for transportation analysis and planning. The system needs for transportation planning have many similarities and many dissimilarities with the well-defined areas of business data processing and scientific computing. The major difference relates to the concern over location, i.e., where the spatial variation of data observations is important.

In transportation planning, considerable attention is given to spatial and temporal variations of phenomena such as land use and population. Typically, these variations are observed as differences in areal aggregates, e.g., population densities or auto ownership rates for subareas of the study area. More sophisticated mathematical modeling of urban structure requires more detailed data—more detailed in locatability, frequent observations, and stratification of population.

Transportation planning requires more detailed information concerning household behavior and more accurate data depicting the urban environment. Information systems concepts applied to transportation planning analyses primarily relate to the development of methods for manipulating data and organizing data to enable efficient access in a variety of forms. On one hand, transportation planning is concerned with the generation of statistical information for parameter estimation and calibration of predictive models. For the most part, ad hoc and unrelated data sets are collected and manipulated by brute force techniques. Data, such as origin-destination data, are immediately areally aggregated. This precludes many taxonomical analyses. On the other hand, transportation planning involves the utilization of various mathematical models for testing alternative proposals. This involves the preparation and processing of large files of data, in conjunction with the predictive models, that simulate changes in the urban structure.

Both the generation of statistical information for predicting the behavior of identifiable and locatable groups of households, and the evaluation of alternative plans require considerable manipulation of data. To efficiently handle data, the transportation planner must be able to communicate directly with the computer and use it as a tool. He should not have to depend on a programmer. He must be given the software capability to handle a large number of the data-handling problems personally. Similarly, data must be organized in ways that enable accurate depiction of complex real world phenomena, and must be in forms that can be transformed to meet a variety of needs.

DATA-HANDLING CAPABILITY

A data-handling capability is a user- and problem-oriented programming language that brings the computer closer to the user. It offers users a great deal of power with a minimal requirement for specialized programming skills. A problem-oriented

---

Paper sponsored by Committee on Urban Measurements and presented at the 46th Annual Meeting.

programming language is one designed for the particular needs for a subject matter. In transportation planning, one of the needs is for the handling of statistical data relating to small areas.

Large-scale urban transportation planning studies have taken a lead in developing specialized substantive systems and rudimentary data-handling capabilities. Specialized information systems consisting of socioeconomic data and highway network data have been developed. These data, along with models for trip generation, trip distribution, and traffic assignment, enable the testing of alternative transportation system configurations.

Unfortunately, too little effort has been expended in developing generalized capabilities to handle the processing and preparation of data for modeling. However, there have been exceptions. One of these is SPAN (Statistical Processing and Analysis System), a system for management of urban data that has many generalized data-handling capabilities (2). SPAN is a large-scale data management, file processing, and statistical analysis system programmed for the IBM 7090/7094. Initiated at the Penn-Jersey Transportation Study in late 1962, it was completed by System Development Corporation with the support of the Bureau of Public Roads and is being used in the Bay Area Transportation Study.

Briefly, SPAN is a flexible system for reducing, manipulating, and displaying data. Processing functions in SPAN are "implicitly" programmed, i.e., the user communicates with the system through simple directives and parameters stated in English-like sentences. The user selects, through the parameters, a particular configuration of pre-programmed procedural options and adapts program operation to his precise requirements.

The advantage of SPAN is that it offers an entire system capable of handling the bulk of transportation planning demands for data manipulation. The disadvantage of SPAN relates to its implementation. Essentially, SPAN is technologically bound to second generation computing equipment. Conversion of SPAN to third generation computing systems will require extensive modification, both in terms of rewriting package programs written in machine language and rewriting the SPAN Systems Supervisor that operates under the FORTRAN II Monitor System. However, a more serious problem is conceptual, relating to an inability to handle data that are not organized in serial files, i.e., sequential records on magnetic tape.

Having greater data-handling capability, but less statistical capabilities, is an IBM-developed set of programs for System/360. GIS (Generalized Information System) is designed to perform data-file establishment, maintenance, retrieval, and presentation operations common to many application areas (3). GIS enables the development of a data base in a convenient form for access. It is also an adaptable user-oriented system for manipulating data to fit a wide range of special requirements.

GIS provides for interrelating data items in different physical data sets. For example, linkage may be used to relate trips of a trip file to households of a household file that make the trip. GIS also enables development of a chain of subordinate items that are associated with a master item. For travel behavior data, there may be a chain for each household master item. Trips generated by a household are subordinate items on a chain for that household. Chaining and linkages are used to represent the relationships between real world phenomena. The utility of structuring data in these forms will be explored in subsequent sections.

If procedures are incorporated into GIS to handle problems associated with spatial data, GIS would serve the data-handling needs of transportation planning very adequately. For example, data-reduction statements to determine whether a point is within a polygon, to calculate land areas, and to perform simple statistical tests are essential in handling data on spatially distributed phenomena.

SPAN and GIS are only illustrative of the kinds of capabilities that are or will become available. Generalized data-handling capabilities offer the transportation planner greater freedom to communicate with the computer. As implied in the discussion of GIS, the data-handling capability and a conceptual scheme for organizing data are highly interrelated. Organization of data, especially data on spatially distributed phenomena, is of great importance.

DATA ORGANIZATION

When phenomena from the real world are observed and recorded, data are created. Quantitative data are defined as collections of values on selected properties of entities. For example, a person weighs 150 pounds. One hundred and fifty is the data value. Weight is the property and the person is the entity. Figure 1 shows data representation. The elements of the matrix are the values, the columns represent the properties and the rows represent the entities. Normally, data are organized as serial records, where entities (in this case persons) are separate records.

Data are observations on some set of phenomena with location and time constraints. These data may be organized as a discrete observation on a phenomenon or as aggregates organized by location and/or time. Data organized and aggregated by location are commonly called areal data. Typical examples of areal data are data summarized to census tracks or traffic zones. The actual phenomena of interest may be housing units, households, persons or trips, but the individual units or entities are now areal units. In areal data, location defines the entity structure. The phenomenon is secondary to the principal argument of location.

Data organized and aggregated by time intervals are commonly called time-series data. Typically, such time intervals as day, week, month, quarter, year, and decade are used. Observations on phenomena such as employment, population, production and sales are often aggregated to time intervals for analyzing variations over time. In time-series data, the entity is defined by the principal argument of time.

Data may also be organized in list structures. A list is defined as a linked sequence of items. Each item contains a pointer or the address of the next item. Thus, the items do not need to be stored in physical sequence. Stored in addressable storage, the next item of an array is implied, whereas the successor item of a list must be specified. An item on a list contains both datum and the address of the next item. This address is called the link or pointer. A list is shown in Figure 2.

A powerful feature of list structures is that the insertion and deletion of data on lists is quite easy. Such operations merely involve changing the linkage of the list. For example, to delete the datum in item c of Figure 2, the link address of Cell b is changed from c to d. List structures for data storage are becoming increasingly

PROPERTY
(e.g., weight, age, income, origin zone)

ENTITIES
(e.g., persons, households, trips, traffic zone)

$$\begin{array}{c|ccccccc}
 & 1 & 2 & 3 & \cdot & \cdot & \cdot & n \\
\hline
1 & v_{11} & v_{12} & v_{13} & & \cdot & \cdot & v_{1n} \\
2 & v_{21} & v_{22} & v_{23} & \cdot & \cdot & \cdot & v_{2n} \\
3 & v_{31} & v_{32} & v_{33} & \cdot & \cdot & \cdot & v_{3n} \\
\vdots & \vdots & \vdots & \vdots & & v_{ij} & & \\
m & v_{m1} & v_{m2} & v_{m3} & & \cdot & \cdot & v_{mn}
\end{array}$$

Figure 1. Data representation.

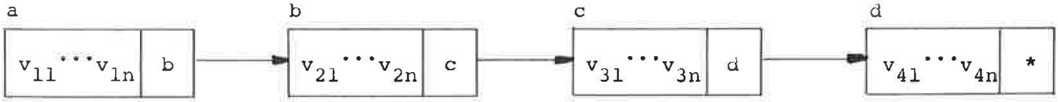


Figure 2. A list.

important as addressable storage devices are becoming cheaper. Transportation planners will find list structures convenient for storing and accessing data on spatial phenomena.

PROBLEMS WITH AREAL DATA

The inference of relationships between characteristics, such as the number of trips per household and household size, is often based on areal data. In using areal data, the areas or traffic zones must be assumed homogeneous. Such analyses focus on variations between traffic analysis zones with the homogeneity assumption setting within-area variation equal to zero. Because household size does not vary appreciably between areas, it is not usually found to be good for predicting the number of trips per household.

Less reliance should be placed on making inferences from spatially aggregated data. Yet, analysis of variations between areas or over space is necessary in transportation planning. This requires greater care and skill in organizing data. For example, it may require that separately collected data relating to the same phenomena be linked. This linkage of entities enables direct relationships to be investigated rather than relying on inferences drawn from unrelated data. Similarly, greater emphasis should be placed on stratifying disaggregated data. For example, rather than using average household size by traffic zone to predict trip production, the disaggregated data should be stratified to enable estimation of trip production for various household sizes.

The need is to organize data in forms that are amenable to linkage, stratification, and aggregation. All too often, the available forms of data restrict or control the analysis. Data are immediately aggregated to minimize the amount of data and thus limit all subsequent data analysis. Newer computers offer considerably more speed and direct access storage. These features provide greater opportunities for more extensive and detailed analyses. To take advantage of this, new concepts of data organization are needed.

INTERMEDIATE FORMS FOR DATA

An important and well recognized data organizational need is the transformation of data from some initial form to some report or product form. However, what is not recognized is the need sometimes to organize data in an intermediate form. An intermediate form is deemed desirable to enable manipulation to a variety of forms. It is often necessary to manipulate data to forms that are completely unanticipated. To provide the necessary flexibility to organize data in unanticipated forms requires: (a) a data-handling capability, and (b) data organized in a flexible form. The desire to organize data in intermediate forms eliminates the need to go back to raw source data every time a specialized request is made.

For example, assume the existence of longitudinal data on household locations. For a sample of households, assume household characteristics such as income, family structure, and cars owned. Also, assume the prior household locations of the families, going back 10 years, were collected. For the various types of analyses that can be performed from a data set such as these, it would be desirable to organize these data in some intermediate form, i. e., something more structured and accessible than the original data set.

To maximize the utility of data in an intermediate form, these data must be organized in such a manner that a generalized data-handling capability can easily transform

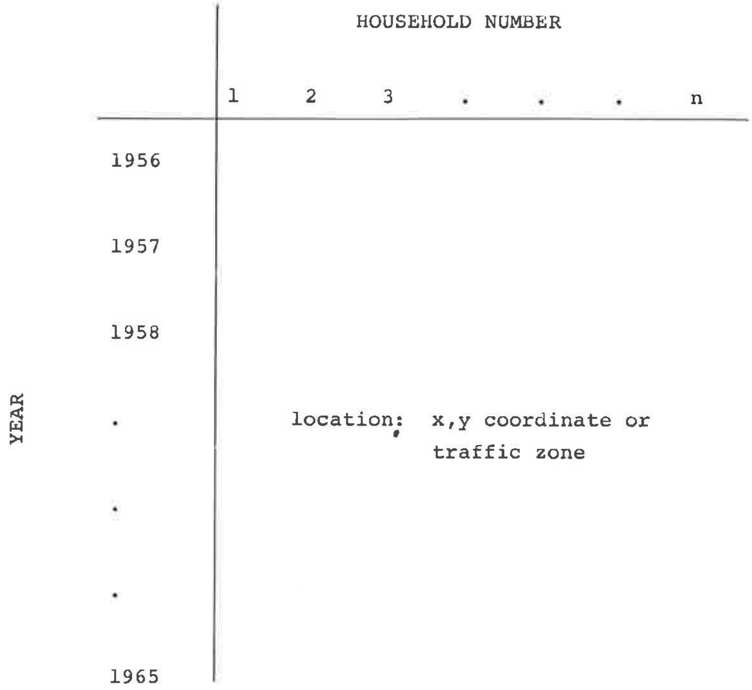


Figure 3. Household location by year.

these data from the intermediate form to a desired form for a specific analysis. To illustrate this concept, Figure 3 shows a matrix of household observations vs the year. The elements of the matrix may contain the x and y coordinate representing the person's home location during that year or a traffic zone number. Data in this form

TRIP FILE	HOUSEHOLD FILE
Origin	Family size
Destination	Number of autos
Purpose	Income
:	:
Link to HH	Link to first trip by a HH member
Link to next trip by member of same HH	

Figure 4. Intermediate form of origin-destination data.

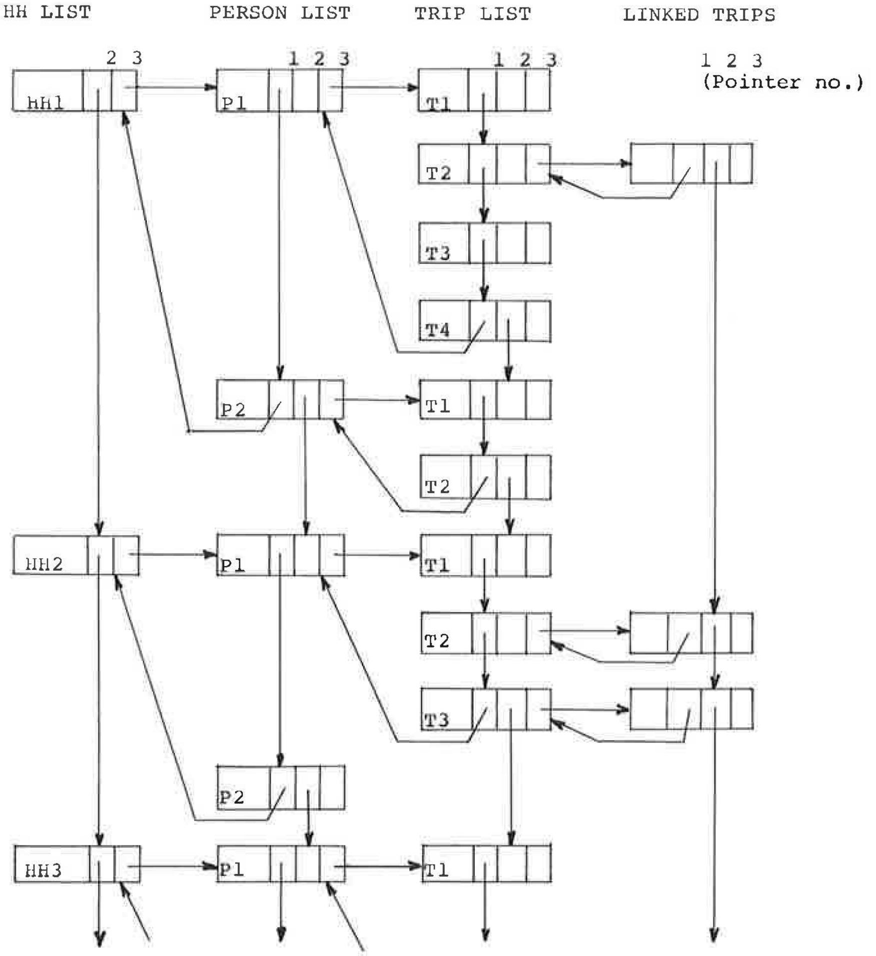


Figure 5. Intermediate form of origin-destination data showing the relationships between items.

may be accessed in a variety of ways. For example, for a given location, the matrix can be searched for households located in that position. In a similar fashion, specification of the year enables creation of a file of household locations for that year. In addition, the matrix enables retrieval of households occupying a specific location for a given year. By utilizing this data matrix, longitudinal data of the type described can be recast in a variety of forms. Recasting from this intermediate form is much simpler than returning to original data observations where often the concern must be with records of unequal length.

Origin and destination data may also be transformed to an intermediate form. To facilitate the creation of specialized data files for the analysis of trip behavior data, it may be desirable to relate trip records of a trip file to households of the household file and vice versa (Fig. 4). In Figure 4, origin-destination data in an intermediate form consist of two files or lists in addressable or direct access storage, one file for trips and another file for households. The trip file contains a record for each trip. As well as characteristics of the trip itself, each record contains a link or pointer to the household generating that particular trip. Also, each entity or record in the trip file contains a pointer to the next trip from that same household. In a similar fashion, all the trips for a particular household are chronologically linked. The household file contains household entities possessing not only characteristics of the household but a pointer or link to the first trip made by a member of that household.

Data organized and described here are of a flexible form for use in creating new files to study such phenomena as a specific trip purpose by persons from households of specific incomes, trips to a specified destination by households headed by an individual of a specified occupation, or analysis of trips linked to households with specified income and car ownership characteristics.

With origin-destination data of the form shown in Figure 4, transformations can be made to very specific forms, regardless of whether the primary argument relates to the trip or to the household. For analysis of households that make a particular type of trip, the primary argument would be trip purpose. Only households making this type of trip are selected for further analysis. For analysis of trips made by a particular class of household, the primary argument is household class. Only trips from households of the particular type are selected for further analysis.

Figure 5 shows a more complex data structure in which origin-destination data are organized as three main lists—household, person, and trip. These lists are interconnected and relate persons to characteristics of their household and trips to characteristics of the person making that trip. This is a similar but more complex representation of the data described in Figure 4. In Figure 5, the trip-making behavior of individual members of the household are distinguishable.

The items containing characteristics of households are linked and form a household list. Each household is linked to a subordinate list of items containing characteristics of persons belonging to that household. Also, each person item is linked to another subordinate chain of items containing characteristics of trips made by that person. In addition, list concepts may be used to represent linked trips, i. e., multimode trips with a single purpose.

In the schematic notation of Figure 5, the last element or pointer 3 links to a subordinate item (e.g., links household items to a person item). Pointer 1 forms the chain (e.g., persons in the same household). Pointer 2 links unrelated households, persons, or trips, and forms a continuous chain of like items. Pointers of Type 2 are desirable to enable scan or search of just one list, say, the trip list if that is all that is needed. Pointers of Type 1 that close on a master item are desirable (e.g., that pointer from the last trip made by a person back to the person item). This formation of a circular chain enables the obtaining of household characteristics for a select set of persons.

In addition to the linkages shown in Figure 5, it may be desirable to link trips originating or destined to the same traffic analysis zones. This concept envisions a chain for each traffic analysis zone. Each trip originating in a zone would be linked to form a chain. The trips on a chain may be ordered by time of origination, purpose, or other trip characteristics.

Need for organizing trip data in this way must be considered in light of difficulties in structuring data in chains and compared to brute force searches of large data files. This comparison depends on the extent to which these data are used. Often-used data warrant more extensive organization. With more efficient data-handling capabilities and more extensive and broader utilization of data, these considerations tend to favor detailed organization of data in intermediate forms.

CONCLUSIONS

Information system concepts applied to problems in transportation planning focus on the need for a generalized data-handling capability and on the need for data organized in intermediate forms. A generalized data-handling capability is needed for rapid creation and manipulation of data files.

Storage of data in complex but flexible forms enables response to a wide variety of problems and analyses and eliminates initial and arbitrary aggregation of data. To make these data-organization concepts attractive and feasible, a data-handling capability must be able to create and update these complex data structures as well as retrieve data from them. Successful data-handling capabilities will be those that can do both. Data-handling capability and data organization are interrelated, with the important concept being flexibility.

Many transportation study staffs and highway departments have adequate data-handling capabilities for present demands and equipment. However, the coming generation of computers offers considerably more speed and direct access storage. These features provide greater opportunities for more extensive and detailed analysis of travel behavior. To take advantage of greater speed and storage capacity, new analysis concepts are required. This means new concepts of data organization and the development of means to create and access these more complex data structures are needed.

REFERENCES

1. Bourne, Charles P. A Review of the Methodology of Information System Design. Information Systems Workshop: The Designer's Responsibility and His Methodology. Spartan Books, 1962.
2. Almendinger, Vladmir B. SPAN: A System for Data Management. Urban Information and Policy Decisions (C. D. Rogers, Ed.). Inst. of Local Govt., Graduate School of Public and Internat. Affairs, Univ. of Pittsburgh, 1965.
3. Generalized Information System: Application Description. International Business Machines Corp., 1965.

Urban Information Systems and Transportation Planning

EDGAR M. HORWOOD, Director of the Urban Planning Center,
University of Washington

The rapid evolution of multipurpose urban information systems along with emergent changes in census technology will provide opportunities for a new generation of transportation analysis and planning studies. This paper outlines the logical bases of new systems under development, discusses applications to general urban analysis and relates these application potentials to urban transportation research, analysis and planning. Particular emphasis is given to the relationship between new information handling capabilities and the needs of urban transportation planning.

•URBAN area transportation planning studies of the past decade have been required to develop a new technology of information-automation not shared by the urban planning field in general. The impetus for this technological development has stemmed from three major factors. First, the studies have been financed at a high enough level and have had sufficiently specialized objectives to permit the application of computer technology. Second, the teams of professionals undertaking the studies included relatively more members with data processing capabilities than found on the coexistent urban planning and renewal staffs. Finally, there has been a more sophisticated level of assistance from the federal supervising agency in relation to urban transportation planning studies than from the federal agency relating to community planning and renewal activities.

These generalizations are not presented to discredit one group of professionals as compared with another, but simply to recognize facts and circumstances that are fairly obvious to those with long-term contacts in the two fields. In all fairness to the Department of Housing and Urban Development, its antecedent agencies were too sparsely staffed to mount the kind of attack which the U. S. Bureau of Public Roads has done in developing computer programs, training a large number of state and local officials in the traffic assignment process, and maintaining some sort of an organization to service requests for information on computer programs. Although the influence of the highway engineer has been substantial in moving the urban transportation planning studies in the direction of automated systems, it is only fair to give a great deal of credit to many professionals in the transportation planning studies who have not come from this field but who have exercised strong leadership in adapting to information-automation, such as the technical leadership of the Detroit and Chicago area transportation studies of the 1950s.

In reflecting on the developments of the past decade, it is interesting to note that the information systems developed in the urban transportation planning field have been essentially the ad hoc systems needed to produce data for urban transportation planning models and to meet the predominantly single-purpose needs of these studies—traffic forecasting. The systems naturally have not addressed themselves to the across-the-board aspects of urban and regional planning. There has been very little spin-off in

---

Paper sponsored by Committee on Urban Measurements and presented at the 46th Annual Meeting.

utilizing the information gathered, not only for this reason, but also because of the organizational systems in which urban planning, urban renewal and transportation planning studies are conducted. Examples of payoffs in information systems developments for general planning purposes are scarce. Occasionally, planning and renewal agencies have found it practical to contract the processing of specific data reports from the transportation area study agencies, but in these instances the cost of data reduction has been relatively large because files have been organized in ways primarily useful for the purposes of the transportation planning groups.

Today, approximately the tenth year of man's general capability to communicate with the computer by using a general programming language, let us take stock of where we are in regard to the needs of the spectrum of users of urban information.

1. Strong pressures have emerged for the coordination of urban planning activities, transportation included, at least on the level of professional studies if not within a coordinated political framework. The transportation planning studies are no longer conceived as being ad hoc studies, but as continual ones with ongoing staffs and resources. There have been joint efforts from several sources on the federal level to pool resources to conduct urban analysis and planning.
2. Strong interest has now emerged both on the political and administrative level toward the establishment of urban and regional information systems for a broad variety of purposes which cut across the responsibility lines of a number of agencies.
3. The climate for information assembly and manipulation in the planning and renewal agencies is rapidly changing in favor of automated systems and ones which are as equally sophisticated in their demands as the foregoing transportation studies.
4. We are at the fulcrum point in the transition from one generation of computer technology to another, a time in which the value of stock in programming and systems development is relatively low because of the new capabilities in the emerging technology, and we have on hand a full decade of programming that will have a rather rapid depreciation.
5. There is a substantial increase in the sophistication of professionals in a growing number of fields in computer applications and computer operations, and we can expect a substantially different climate and professional capability in the emerging planning agencies.

All of the these facts point to a new climate, new attitudes, new capabilities, new methods of attacking the problems and new problems. They tell us in effect to take a closer look at integrated urban information systems for multipurposes as well as for a higher level of utility and capability for the various categories of users, including the transportation and planning segment of the urban planning area.

How will more sophisticated multipurpose urban information systems change the nature of the urban transportation planning process? To assist in answering this question, I have outlined in the following section what I believe are a handful of basic urban information systems needed for multi-projects.

BASIC URBAN INFORMATION SYSTEMS AND APPLICATIONS

This part of the paper describes tools and processing systems for handling data for an urban area information system and some of the more apparent applications. The material is concerned with the planning, development, testing and refinement of tools which are needed for the total spectrum of urban planning activity. Six basic components of the system are outlined as follows:

1. A geocoding system designed to convert data input by street address location identifiers to geographic coordinates, allowing information retrieval by arbitrary areas of interest as well as traditional areas of record, such as census tracts.
2. A query system designed to facilitate querying and manipulation of large data bases.
3. An automated graphic display system for map making and data display.

4. A plan test system for testing, by simulation, alternative proposals regarding employment distribution, residential densities, transportation facilities and the effect of capital expenditures and priorities on the planned growth of various sectors.

5. A planning operations system designed to assist in internal agency day-to-day routine information processing needs, such as document retrieval, report generation and the production of statistical reports for fixed-time series intervals.

6. A capital improvements and work scheduling system based on critical path methods of analysis, designed to integrate the planning and programming of all public works development in the area.

The first three components are parts of a continuum and are complementary. The first enables location of data, the second manipulation of these data, and the third data display. More will be said about their relationships as they are discussed in depth.

Before proceeding to describe each of the subsystems, some general remarks are in order concerning the choice of system components. The components recommended here stem from three requirements: (a) they are the most basic needed in terms of developing a data-handling capability; (b) they relate to the most fundamental questions which planning agencies must answer in terms of its programs; and (c) they will upgrade planning and programming capabilities. With the exception of the query element, the other systems proposed have already proved operationally feasible from test applications or experience.

Geocoding System

A geocoding system automatically relates a data observation, event or happening to a mapped location. A geocoding system of the type developed by Dial (1) and implemented by Calkins (2) is suggested as an essential tool for positioning data observations in two-dimensional space as well as retrieving data in highly flexible ways. Many data of concern in urban analysis are only locatable by street address identifiers or more easily coded by such identifiers. However, street addresses are very cumbersome to use in data retrieval. Coordinates, or X and Y spatial references, are far more powerful location identifiers from the standpoint of information retrieval capability.

To accomplish the translation of street addresses to grid coordinates, it is necessary to compile a street-address-to-coordinate directory.[1] Once compiled for a city or an urban area, the directory serves in the translation of any data having street address locational identifiers. Assistance in compiling directories is provided by a computer program that allocates entire lengths of streets into segments created by intersecting streets. Associated with street segment records are address range limits for each block, and grid coordinate values of the street segment ends or street intersections. The addresses of input data are systematically compared to the street segment records of the street-address-to-coordinate directory to accomplish the translation. Thus, any record of interest, such as a housing unit, incidence of communicable disease or the location of an economic activity, has a grid coordinate value automatically assigned to it if the input record has its street address coded.

If the conversion is to grid coordinates rather than areal unit codes, such as census tracts, the coordinated data observations may be assigned later to arbitrarily delineated areal units by a query system. This is done through a procedure that tests whether the coordinates of data observations queried are within polygon boundaries describing the areal units. This capability utilizes geocoded data. The assignment of coordinates to polygons is part of the query system that is proposed and described here.

As stated earlier, the geocoding system permits the conversion of data entities whose locational identifiers are street addresses to grid coordinate identifiers. This

---

[1] An operational test of such a system is now under way as part of a National Capital Commission—Dominion Bureau of Statistics test in the Ottawa, Ontario region, using 1966 Census data for retrieval analysis.

conversion opens the door for the use of many kinds of data that were previously unusable because they were not readily locatable. Many data exist, related to persons or properties, that are coded only by street address. Data such as building permit applications, employee address records, charge account records, pupil residence records, and hospital patient discharge records may be made amenable to spatial analysis by this system.

Monitoring subsequent population change by utilizing building permit application data is necessary to maintain current population estimates. Typical questions that building permit data are able to answer relate to time, location and magnitude of new construction. Coordination of building permit data enable their automatic allocation to any arbitrary analysis area of interest or traditional enumeration districts such as census tracts, traffic zones or school attendance areas. If building demolition records are entered into the system as well as the record of population movement from utility connection information, then a permanent system is available to monitor population change with very small error.

Besides monitoring population change as discussed, a geocoding system may assist in studying the spatial distribution of a great many urban phenomena important to planning analysis. These include, but are not limited to, the following:

1. The trade area of shopping centers;
2. The tributary area of hospital patients;
3. The location of commercial or industrial land uses;
4. The location of different land-use adjustments;
5. The tributary area of employment centers;
6. The location of federal employee residences;
7. The location of dwellings by types and values;
8. The location of communicable diseases of various types;
9. The location of traffic accidents by various classifications; and
10. The location of welfare cases of different types.

In each case, address records must be obtained from source documents such as public records, commercial directories or especially contrived sources which involve the cooperation of people and business establishments.

The applications of geocoded data have virtually no limits of utility for regional and local planning agencies, a variety of municipal departments, and both public and private welfare agencies. Once geocoded data have been developed, a project which is not to be underestimated in scope, these data are amenable to visual inspection by automated graphic display or for various types of statistical analysis or to arrive at notions of the urban ecological processes, including concentration, dispersion, clustering and time changes in the settlement patterns of people, households and economic activities. These changes should be relatable to planning policies.

Query System

A query system connotes a means or tool to access data on spatially distributed phenomena in a variety of combinations and with relative economy. The emphasis must initially be upon the system development rather than a collection of data. Without such a philosophy, each specific planning task will require its own specific data retrieval systems and no general flexible processing capability will be gained.

The query system is designed not only to handle data with street address identifiers, but also to combine the entire spectrum of data collected by different identifiers as well as for differing time spans and spatial boundaries. The spatial data query system essentially consists of a user-oriented computer programming language and schemes for organizing spatial data. This system is conceived to give planners a great deal of flexibility and power in handling and preparing data for analysis and reports.

The query system described requires that data be locatable in space by means of coordinate identifiers or areal unit codes. Geocoding, or the automatic assignment of spatial coordinates to entities, and the query system are complementary. Geocoding assigns coordinates and the query system manipulates or handles coordinated data. The existing Dial-Calkins geocoding system has some elementary, although important, query capabilities, e.g., capability to assign polygon identifiers to data observations.

To illustrate the query system better, assume the existence of elementary school student residence records in machine-processible form. Also assume these data have been geocoded by a street-address-to-grid-coordinate program. Various school enrollment areas may then be tested using the query system. This system is capable of selecting entities based on satisfaction of specified values for properties of these data. For example, students could be assigned to school service areas by testing whether their locational coordinates are within polygons describing these areas. In addition, students of a particular grade may be specified for selection. These kinds of queries may aid in the allocation of teachers to specific grades in specific schools.

Similarly, travel behavior data from interviews of households may be queried. Typical queries relating to data from household travel behavior studies are (a) select household entities with incomes exceeding $10,000 and who own more than one car; (b) select household entities owning no cars and whose family size is more than two; and (c) select household entities that make more than 10 trips per day.

Examples such as these use all three of the complementary systems—geocoding, query, and graphic display. Geocoding is used for ease of inputing spatial data, a query system to enable manipulation, and graphic display for spatially positioning the output.

Emphasis is placed here on the requirement for building a general query system, much as an assembly line must be built either to produce one or a thousand fabricated products. When attention is not given to the general development of a process to handle miscellaneous queries, common situations arise in which it may take several hours of computer time and the input of many reels of tape to answer relatively simple questions, much as it would require considerable time and cost to fabricate a car without the existence of an assembly line.

Ongoing agencies that envision using a large data base must develop a generalized data-handling facility. Such a facility is necessary to organize, store, retrieve and report all kinds of data on spatially distributed phenomena. Whether these data are used to assist in the evaluation and control of a process of operation, or whether they are to provide a basis for planning and to evaluate alternative plans, generalized processing expertise is essential, and, in fact, equally important as the data itself.

The central function of a query system is to facilitate the query and search of data and to simplify programming instruction for the desired output. It is important that a query system be designed for ease of use by persons with minimal computer experience, but who are skilled in planning analysis. A data-handling procedure designed for ease of use is the single most important element of an information system. Without a flexible system, many data are effectively locked in due to the time and cost of retrieval or manipulation.

To develop an effective automated information operation, an agency must generate an efficient means of reaching magnetically stored data, manipulating it and reporting results. This may be accomplished in several ways. One way is to develop and acquire package programs that perform all anticipated desired functions. By specifying parameters, or limits, the package programs can meet specific needs; but unfortunately, all the desired needs are not initially known. One step above acquisition of separate package programs is a system or collection of user-oriented package programs operated by English-like instructions that perform a spectrum of frequently desired tasks, such as the production of time-series statistical observations or graphic display. At a higher level, task-oriented programming languages enable compilation of more flexible and powerful instructions for highly specific tasks.

A most promising data-handling means is being developed by IBM for the System/360 (3). It is called GIS (Generalized Information System) and provides great power for storing, manipulating, retrieving, and presenting data. GIS enables storage of and retrieval from complex and linked data sets. In addition, the syntax of GIS is very flexible and in English-like language. However, GIS is not particularly oriented to handling spatial data. It is necessary to develop a subset of GIS, or an independent system to handle spatial data.

Alternatively, an agency could utilize existing data programming systems such as SPAN (4) or MARK III (5). However, each of these systems was designed and implemented on specific computer configurations and is independent of a flexible spatial coding system. Their dependence on second generation hardware also proves limiting. Second generation is the term generally applied to the computer hardware configurations in use in the first half of the 1960s, such as the IBM 7000 Series, GE 235, and CDC 3600. Third generation includes the IBM Series 360, GE 645, and CDC 6600.

Unlike commercial and scientific fields, computer programming systems for urban regional analysis needs are relatively underdeveloped. Of the systems described thus far, only the geocoding system is highly operational. Clearly, it is not specifically the purpose of an agency to become engaged in information systems research. To solve problems and assist in operations, it is nevertheless important for all organizations with large-scale information needs to contemplate advancing the state of the art of information systems operations consistent with the scope of their responsibilities.

Automated Graphic Display System

Automatic graphic display of information is essential for the presentation of a spatial pattern in the form of maps. Map imagery is presented via automated plotting hardware or cathode ray tube photographic output. In addition to displaying data through map imagery, there is a need to reproduce maps themselves at varying scales and with varying information according to the specific project at hand. This type of graphic output is now fully developed in the aerospace industries. The technology must be adopted to urban area planning and analysis needs, if manpower requirements are to be kept within reasonable limits as work demand grows.

Whereas the information system component discussed previously is concerned with spatial retrieval, it requires a graphic display subsystem to carry through in terms of a visual reporting of the information retrieved. For example, a general spatial query system could retrieve from large data files the number of people living in housing units 50 or more years of age within 5 miles of a projected employment center, assuming such data to be part of the base. The knowledge of this number itself would be only partially significant. Significance of much greater importance would be attached to a visual display of these housing locations interpreted through a map screen of the urban area. Such imagery could give some highly visible clues as to the geography of rehabilitation needs.

An automated graphic display system is essentially the application of computer programming to the ordering of data for display on output equipment. These data are ordered in such a way that the output of the system will be in the form of a map or a graph.

Output devices are of three types: (a) the high-speed impact printer, which is found at almost every large-scale computer installation; (b) the inkline plotter of either the plane table or platten type, wherein the motion of a pen is programmed to produce the image much as it would be produced by hand; and (c) the cathode ray tube (CRT) which looks like an oscillograph and for all practical purposes may be likened to a television image. Of the three graphic output modes, the CRT is the most efficient because it produces the image at electronic speed; the only production time limitation is in the photographic equipment which records the image produced on the scope. Impact printers are relatively quick and have the advantage of being very handy, but their output resembles that produced by a typewriter. The inkline plotter is probably the best tool for cartographic simulation of conventional map production. Highly efficient package programs have already been developed for graphic display via the impact printer (6).

A special adaptation of the CRT is the "light pen," which is a stylus-like instrument connected to the CRT by a cable and held in the hand. The user can apply the light pen to images displayed on the CRT either to enter data or retrieve data for ground locations which are visually evident by the display of a street line map. It is a short step from the present technological capabilities described in the query system to adapt to a system whereby the user can describe a perimeter of ground space by light pen outline in reference to a map image of the city programmed to appear on the CRT and call for data to be retrieved or summarized for that area. Practical examples of such use would be obtaining a feedback on the number of people living within an arbitrarily designated area, the value of land within some perimeter, and the number of cases of a communicable disease in a given area.

Perhaps the most immediate need for automated cartography is in the production of maps superimposed with information selected to meet the demands at hand. In this regard, a by-product of the geocoding system is a digitized record of all street segments. These segments may be displayed in any scale, either by inkline plotter or by CRT scope.

Digitized street segment records may be immediately valuable in displaying information of relevancy to municipal engineering and public works analysis as well as transportation planning. For example, if a file of traffic-volume counts is developed for street lengths, an adaptation of the display program permits the streets to be shown at varying widths corresponding directly to the traffic volumes. Similarly, a file of street lengths can be produced showing only streets of particular widths or pavement conditions. Current CRT technology even permits colored imagery and photographic reproduction, which could display roads of one type in red, those of another in green, and so forth.

Before leaving the subject of computer graphic systems, it is worthwhile to note that systems are now available to assist in the urban design process by constructing perspective drawings from very basic plan and elevation data. The conventional procedures of urban design require, at some stage of planning, a substantial expenditure of time and effort in the development of perspective drawings and the preparation of models for both design analysis and lay evaluation. Recent developments in computer output systems, such as the refinement of the inkline plotter and CRT display, now permit new tools of three-dimensional analysis which give the design team a multitude of perspective views stemming from the one-time coding of the spatial location of corresponding points of the given object (building, building groupment, or other spatial arrangement) with reference to the orthogonal distance from the picture plane and ground level.

A simple analogy would be the production of a multitude of perspective views of a building from a single set of coded information capable of production by a nonprofessional design aide. The viewing position could be made to vary in height and azimuth position, or various combinations. Views can be projected on a CRT recorder at speeds limited only by photographic requirements (60 frames/minute for practical considerations). In fact, motion pictures can be produced easily from the output which simulate moving around the building in space or walking through an urban spatial setting. This process may not preclude the need to build a model for public relations purposes, but it can substantially aid the design team in the predetermination of the spatial elements of the solution.

In summary of this section, an automated graphic display system will not only have a substantial impact in upgrading graphic study capabilities of the planning agencies, but also should ultimately give them the capability of information retrieval in a real-time sense for decision-making purposes.

Of all of the systems proposed in this section, no doubt the graphic display system could be the most utilitarian in both the savings of drafting time and the provision of services to other agencies in the region. For an agency continually involved in the production of maps, it is an economic certainty that before more than a few years pass there will be justification for the rental of in-house plotting equipment. Virtually all major transportation studies in the United States have already attained this level of in-house need. An automated graphic display system is not only necessary to upgrade map production, but to produce output reports which use a visual format to convey the meaning of the information. This type of output facilitates both technical, management and policy decisions.

Plan Test System

The effectiveness of plans, policies, and alternate proposals may be tested in a variety of ways ranging from simple visual observations to complex mathematical models. Contemporary planning places considerable hopes on a capability to monitor or test the plans and policies.

A distinction should be made between the testing of alternate proposals in the planning process as compared with the testing of the plan or policy itself in regard to events which occur after its inception. In the latter case, plans and policies can often be tested by relatively simple information feedbacks. For example, a system which reports on land-use changes or adjustments, not strictly in compliance with the planned land-use policy, may be a valuable tool in determining whether or not the policy is honored in general or breached in day-to-day adjustments. In this regard, it is common to encounter land-use adjustments that are substantially contrary to plans and which, while conceivably within the legal framework of adjustment, display the inoperation of a plan or the unwillingness of an administration body to implement it. These statements are not meant to imply that the adjustment of a plan in contradiction to its policy is necessarily illogical, but merely to indicate that feedback systems can give a fairly good indication of policy success, whether it is dealing with land use, prison paroles, welfare programs or other matters.

Time-series data of historical nature can further shed light on the effectiveness of plans and policies. An example here is the evaluation of urban areas which have been considered timely for rehabilitation through housing code enforcement. If records indicate that housing codes are not enforceable and that rehabilitation money is not being generated for an area, then it is probable that the housing stock is beyond rehabilitation status in general.

Information feedback systems can go a long way toward evaluating the impact of plans and policies in an ex post facto way. Their value should not be underestimated. On the other hand, the problem of pre-evaluating planning alternatives from the standpoint of social and economic justification, impacts and potential success is a much more sophisticated exercise, and calls for the use of models in which the input variables can be altered and the estimated results studied for different mixes of inputs. In the traffic planning and shopping center planning fields, gravity and accessibility models are typically used for preplan testing purposes and have been found to be substantially successful. It is the purpose of this section, however, to present the dimensions of at least one major plan test system as an example.

The plan test system described here is an illustrative model.[2] Its utility is in answering questions of the impact of decisions on the location of new employment centers in an urban region. This model presumes employment inputs can be anticipated 3 to 5 years in advance, both in the government and private service sectors.

Because the model-building state of the art is relatively crude, partially because of limited data-handling capabilities, only a simple model is formulated. Initially, the emphasis should be on analysis of data for calibration or estimation of parameters for the model. This slow and deliberate approach to model building is based on experience in other model-building efforts where too much was attempted or promised without sufficient data base or data-handling capability to effectuate it.

Among the most important determinants of urban patterns are employment locations and transportation facilities. Thus, the suggested model allocates increases in area employment at specific loci to residential areas. Such a model is deemed useful to test alternative proposals of large employment center locations in terms of impacts on travel facilities and residential densities. It is felt that a linear programming model, which distributes population, given inputs of employment and travel times, enables policy makers to evaluate decisions as to locations of large national government employment centers. In addition, needed transportation improvements may be shown as a result of model application, and potential pressure for development may be anticipated.

---

[2] This model rationale has been proposed by Kenneth J. Dueker in a study for the National Capital Commission of Canada, forthcoming.

A linear programming allocation process or solution method is proposed. Linear programming formulation enables optimal allocation of persons to residential areas based on minimizing the aggregate work travel cost for the new employees being allocated. This allocation is subject to constraints of transport capacity, housing supply, and employment demand.

Only incremental employment changes are to be allocated. The model is designed to allocate new employment to existing or new housing. Thus, the concern is not in replicating existing patterns, but in allocating future population growth to residential areas. This incremental approach creates some difficulties, however. For example, transportation capacity between employment zones and residential areas must be expressed as remaining available capacity, and housing supply as available housing supply. Another problem, that of generating many variables within the model, is partially circumvented by a short planning horizon of 5 years and by an iterative solution to achieve a balance between employment, available residences and transport capacities. In a long-range model, variables such as service employment must be endogenously generated. In a short-range model, service employment can be exogenously estimated and considered input to the model. If the model does not allocate sufficient population for that area, the employment inputs must be modified and the model rerun. Similarly, the model must be rerun until travel times and capacities are properly related. Subsequent refinements of the model might be to make these iterations internal to the model or automatic.

Planning Operations System

Inherent in large planning offices are high costs in accessing records and developing documents, reports and graphs by hand. Also, much repetition of office operations is required because of document updating requirements, such as the periodic production of statistical reports. Automated retrieval systems are mandatory if large manpower inputs are to be avoided. Such systems will reduce the amount of nonproductive work and provide more current and extensive information on available documents, records, reports and graphs.

Document retrieval systems, such as KWIC (Key-Work in Context), SDI (Selective Dissemination of Information), and "peekaboo" retrieval systems should be operationally investigated. These systems need to be evaluated in terms of being able to retrieve documents by subject, author, agency location, and in terms of specific agency needs.

Administrative records, such as property management data, that may be of value for planning are recommended for investigation, although organization of property data for management purposes is necessarily different than for planning. For planning, these data must be geographically ordered. Though the ordering and data-handling systems for management and planning are different, one system should rely on the other to collect and update that part of the data base relevant to its own needs. Ideally, the planning system should be designed to utilize data from the management system, but must be a separate system itself because of its differing requirements.

Critical path planning accounts for the restraining interrelationships between the various improvements. Within the restraining framework, the various improvements must be scheduled. At this point resource limitations are considered. Resources, such as finances, manpower and equipment, are allocated and the improvements scheduled to level resource fluctuations over time. The key is to make the entire process systematic and partially automated to enable rapid reevaluation for testing alternative capital improvements proposals.

Elements of the capital improvements program go through several cycles of budgeting, programming, and scheduling as they pass from the conceptual stage to realization. Within the overall budget, projects are in various stages of planning, moving from long-range planning to design investigations, to preliminary design review, to final design, and finally to construction. At each stage in planning an improvement, new information is added to the programming, scheduling, and budgeting process. Incorporation of these cyclic considerations into the Capital Improvements Program system will offer a great deal of monitoring and control capability.

RELATIONSHIP TO TRANSPORTATION PLANNING

Not all of the systems described have equal implications for urban transportation planning. Some, such as the graphic display and work scheduling elements, are part of the urban transportation planning technology that has evolved to date. Likewise, the plan test system is inherent in the transportation planning modeling work that has gone on for at least a decade, and therefore poses no uniquely new concept to the transportation planner. These systems have been described in somewhat broader terms, however, as part of the total spectrum of urban information systems needs.

Of far greater importance to the transportation planning process are the geocoding and general spatial data query systems. These are not part of the conventional technologies of urban transportation planning studies, and they are perhaps the one phase of the transportation planning process that is in most need of attention from the technological development standpoint.

The significant contribution of the geocoding system is that it eliminates reliance on preconceived notions of region, such as the traffic enumeration zone, and at the same time permits the aggregation of information from many sources for query areas that may constitute an analysis spectrum of areal configurations. For example, traffic enumeration district zones traditionally increase with distance from a central focus, and thereby mask out the true nature of trips to subfoci. A regional shopping center complex, for example, along with associated commercial and industrial land uses, may be found at the intersection of four traffic enumeration zones in such a way that a unique activity focus is not observed at the destination end of the trip. Further, the size of the traffic analysis zone at that location in the region may be such as to mask the importance of destinations to the shopping center because both origins and destinations fall within only one zone.

The importance of the specified spatial query system for transportation planning lies in the capability of matching the characteristics of specific households with their specific travel generating characteristics, rather than using the average values for enumeration districts themselves. At the existing level of technology, the averaging of characteristics for areas masks out the true characteristics of the regression of one variable upon another. At present, we also infer that the characteristics of selected samples cover their areal entities, whereas with the general spatial query systems it should be possible to link the properties of entities at the basic level of enumeration.

ACKNOWLEDGMENT

The author wishes to acknowledge substantial contribution to this paper by Kenneth J. Dueker, Assistant Professor of Civil Engineering and Urban Planning, University of Wisconsin.

REFERENCES

1. Dial, Robert B. Street Address Conversion System. Urban Data Center, Univ. of Washington, 1964.
2. Calkins, Hugh W. Operations Manual for Street Address Conversion Systems. Urban Data Center, Univ. of Washington, 1965.
3. Generalized Information Systems: Application Description. International Business Machines Corp., 1965.
4. Almendinger, Vladimer B. SPAN: A System for Data Management. Urban Information and Policy Decisions (C. D. Rogers, Ed.). Inst. of Local Govt., Graduate School of Public and Internat. Affairs, Univ. of Pittsburgh, 1965.
5. Postley, John H. A General Purpose Urban Management Data System. Proc. Third Annual Conf. on Urban Information Systems and Programs, Amer. Soc. of Planning Officials and Northwestern Univ., 1965.
6. Using Computer Graphic for Community Renewal. CRP Guide Number 1, U. S. Govt. Printing Office, Washington, 1964.

Ottawa Street Address Conversion System

HUGH W. CALKINS, Research Planner, National Capital Commission, Ottawa

•THE Ottawa Street Address Conversion System is a geocoding system being developed for the Ottawa-Hull Metropolitan Area in Ontario. This report describing the system will be presented in two parts. The first part contains some general remarks relating to the function of a geocoding system and the second part describes the current project in Ottawa, a geocoding system based on street addresses.

GEOCODING SYSTEMS

Geocoding is the process of assigning a spatially significant identifier to a data observation. A spatially significant identifier is defined as one from which the relative spatial location of a data observation is evident from the value of the identifier. For example, if two properties are identified by parcel numbers (No. 202-013-001 and No. 202-016-024) or street addresses (1091 3rd Ave. and 820 Main St.), these identifications do not describe the relative spatial locations of the parcels. However, if the two parcels are identified in terms of x-y coordinate locations (2510, 4400 and 2730, 3782), the values of the identifier represent relative spatial locations. Coordinates are the most flexible spatial identification system. Coordinates, however, do not replace nonspatial forms of data identification, but rather, the coordinates are a generalization which represent all spatial reference units. Coordinates can replace all geographic coding for areal units, i.e., traffic zones, census tracts, etc. Computer techniques can then be used to assemble the data by any set of areal units using the coordinate identifiers of the data observations.

For example, the first use of the Ottawa street address geocoding system will be to determine coordinates for each household in the Ottawa Metropolitan Area as enumerated in the 1966 census. These coordinates will be placed on the household data files maintained by the census department. Population characteristics can then be summarized and reported by any arbitrary set of areal units, i.e., traffic zones, planning districts, etc.

A geocoding system is only one part of an overall information system capability. The major steps in an information system are (a) collecting data; (b) organizing data; (c) storing data; and (d) retrieving data. A geocoding system is part of the process of organizing the data and it establishes the identifiers needed to enable maximum flexibility for retrieval on spatial criteria.

In order for data with spatial significance to have the capability of providing meaningful information for a specified problem, it must have an identifier which is spatially significant. That is, all data must be referenced to a point in space or areal unit. In the past, data have usually been referenced to a set of more or less single purpose areal units, e.g., census tracts, and traffic zones. Two major flaws to this system have long been recognized.

1. All anticipated sets of areal units must be specified before the time the data are collected, a severe restraint at the beginning of the study; and
2. Data collected for different sets of areal units are often non-comparable.

Data referenced to the larger areal units are in fact summarized to the level of that areal unit, and retrieval on spatial criteria is limited to the set of areal units as defined before the data were collected. As long as data are processed by manual methods,

---

Paper sponsored by Committee on Urban Measurements and presented at the 46th Annual Meeting.

it is virtually impossible to break away from this system due to limitations in manpower resources. However, the speed offered by computers overcomes this restriction.

Spatial units can be generalized as follows: coordinate region, query regions, and uniform data regions. The coordinate region is the area represented by a single coordinate point on a grid system and has a size and shape equal to a square whose sides are equal in length to the distance between points on the grid system. The query region is an arbitrarily delineated areal unit, e.g., traffic zone or census tract. Query regions are represented as polygons, each vertex of the polygon being represented by a coordinate region. Uniform data regions indicate the areal extent of some data characteristic, and are described in the same manner as query regions. The common point of reference to all three regions is the coordinate region.

A coordinate system is therefore used to establish and retain flexibility in the definition of query regions and in the description of uniform data regions. Individual data observations must be identified by unique coordinates. These coordinates should be based on a standard grid system such as the State Plane-Coordinate System.

Having established the need for data to be identified by coordinate values, it only remains to develop the techniques of assigning coordinates to the data observations. The magnitude of this task should not be underestimated because it involves assigning a coordinate value to every individual data observation, e.g., property, household, etc. However, automated techniques are being developed to substantially reduce the time and manpower resources needed to do this job. The street address conversion system being developed in Ottawa is such a technique.

OTTAWA GEOCODING SYSTEM

The system described here is one which derives the coordinates from street addresses. There are three coordinate sets (x, y) assigned to each address: (a) the unique proxy; (b) the midpoint of the block face; and (c) the block centroid (Fig. 1). The unique proxy coordinate value is a unique identification for each address; however, its spatial location may not always be exact. The other two coordinate values represent small areal units for data summarization—the block face and the block.

The sequence of steps in this system is shown in Figure 2. The data observations, identified by street addresses, are converted to machine records (punch cards or magnetic tape). These records are processed by a computer program which assigns any combination of the three possible coordinate sets to each data record. The program

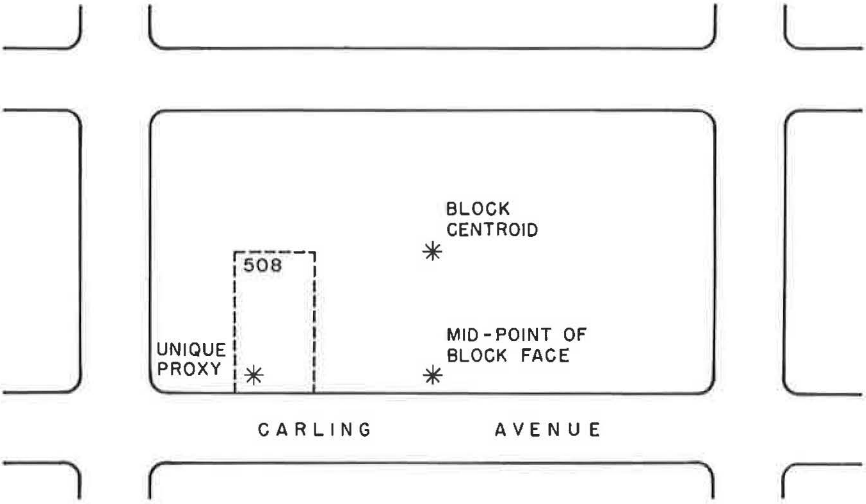


Figure 1. Coordinates.

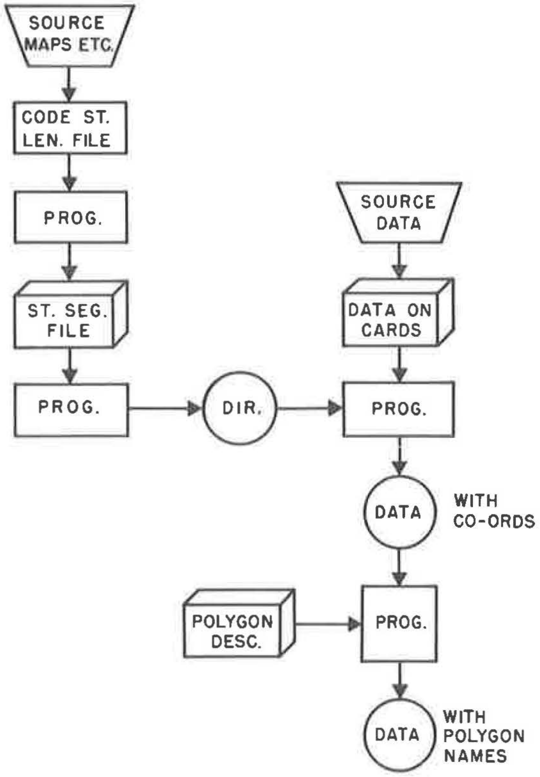


Figure 2. Flow chart.

relies on a directory for the information needed to assign the coordinates. The product of this first step, data with coordinate identifiers, is then processed by another computer program which assigns to each data record an identification code representing the appropriate areal unit (query region). The set of areal units is described as polygons.

The final result is data identified by coordinates and by specified areal units. At this point the data can be summarized, analyzed, and reported on the basis of any of the identifiers, coordinate sets or areal units. The capability to retrieve data by arbitrary polygon sets is the query capability built into the current geocoding system. Although this is far from a general query capability, it obviates the need to code geographic identifications at the time the data are assembled.

In the flow chart of the system (Fig. 2), the directory is shown as one of the major inputs to the step in which coordinate sets are assigned. The building of this directory is a computer-assisted process.

The steps required to build the directory are (a) code the street length file from the source documents; (b) run the computer program to create the street segment file from the street length file; (c) make corrections to the street segment file (manual process); and (d) run the computer program to build the directory from the street segment file.

The basic unit of the directory is the street segment record (Fig. 3). The street segment record is a straight-line segment between the centers of two consecutive intersections or abrupt changes in a street's direction. The following information makes up a street segment record:

1. Street name;
2. Range of street addresses;
3. Coordinate values (x, y) at each end;
4. Block centroids (x, y) on each side; and
5. An indicator for determining which side of the street relates to even and odd addresses.

The address range represents both sides of the street. In order for the system to function properly, the addresses on each side of the street segment must fall within the same range.

The street segment file is prepared from the street length file by a computer program. A street length is an assumed straight length of street over which addresses are always increasing. A street length represents a series of street segments by the following (Fig. 4):

1. Street name;
2. Range of street addresses;
3. Coordinate values (x, y) at each end; and
4. An indicator for determining which side of the street relates to even and odd addresses.

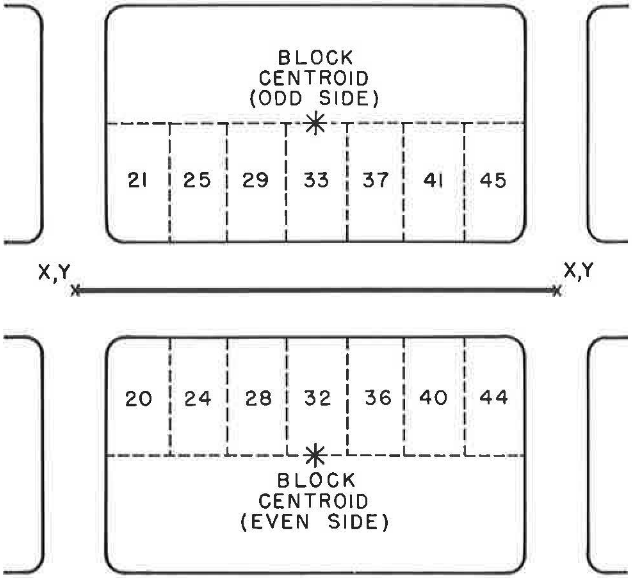


Figure 3. Street segment.

The program divides each street length into its appropriate number of segments and calculates the address range, coordinate values for the end points and the block centroids for each segment. Thus, for any street pattern the initial coding process need only record the end points of all street lengths. The hypothetical street pattern (Fig. 5) would require 19 street length records. Processing these records by the computer would yield 131 street segment records. The street segment file as produced by the computer is not complete. In areas of very regular street patterns, up to 90 percent of the street segment file is created by the computer, the remainder must be done by hand. However, as the street pattern becomes less regular the amount of work done by the computer drops, thereby increasing the manual work.

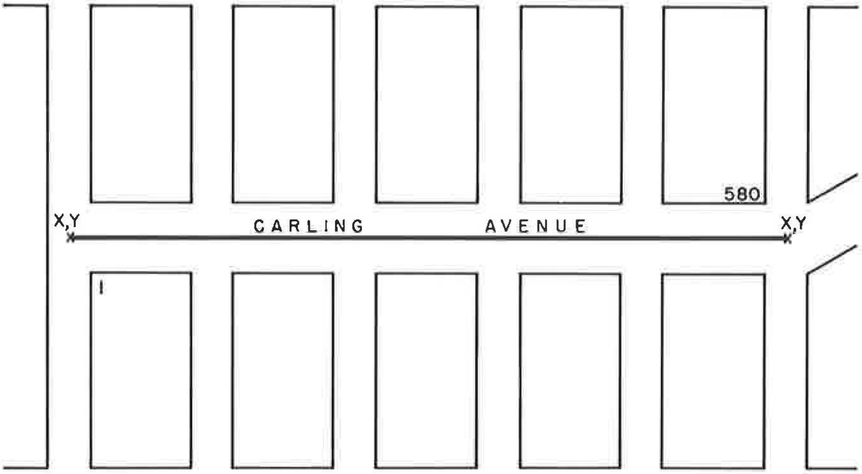


Figure 4. Street length.

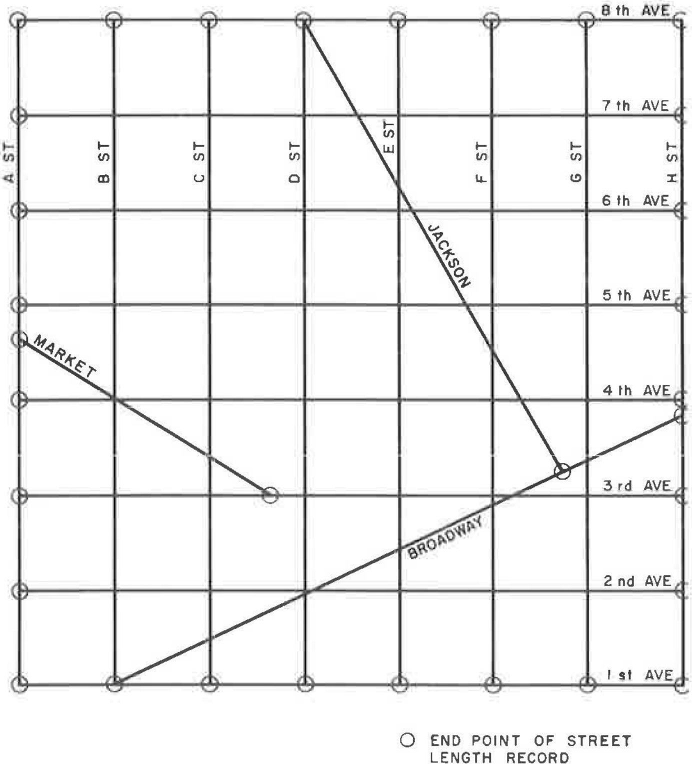


Figure 5. Hypothetical street pattern.

Operational experience in the application of the street address geocoding system to the Ottawa Metropolitan Area disclosed certain problems. The first problem related to the regularity of street addresses.

As was noted earlier, the system requires that the address range for a street segment represent both sides. Addresses must always be increasing by more or less constant increments to successfully meet this requirement. Many cities have adopted a system of assigning a range of 100 addresses to each block, thus producing a very regular address pattern. However, when any system of a less regular nature is used, then it is possible for address ranges to have characteristics which are not compatible with the system. Briefly, some of the situations which can develop are (a) addresses increasing faster for one side of a street than the other; (b) even and odd addresses on the same side of a street segment, such as on a circle or crescent; and (c) addresses out of sequence.

Another major problem which can limit the use of this geocoding system is the areal extent of urban street addresses. If the area of interest is completely urbanized, then this is not a problem. However, if rural areas are to be included, the present geocoding system cannot be used over the entire area. In the Ottawa Metropolitan Area, urban and rural areas are intermixed (Fig. 6) so that the street address geocoding system can only represent a discontinuous area.

The original system has required only slight modifications to deal with the situations described. To solve the problem of irregular patterns of street addressing, new techniques have been developed:

1. Block face records are substituted for street segment records in the directory when one address range cannot represent both sides of the street without creating an ambiguous situation with respect to other street segments of the same street.

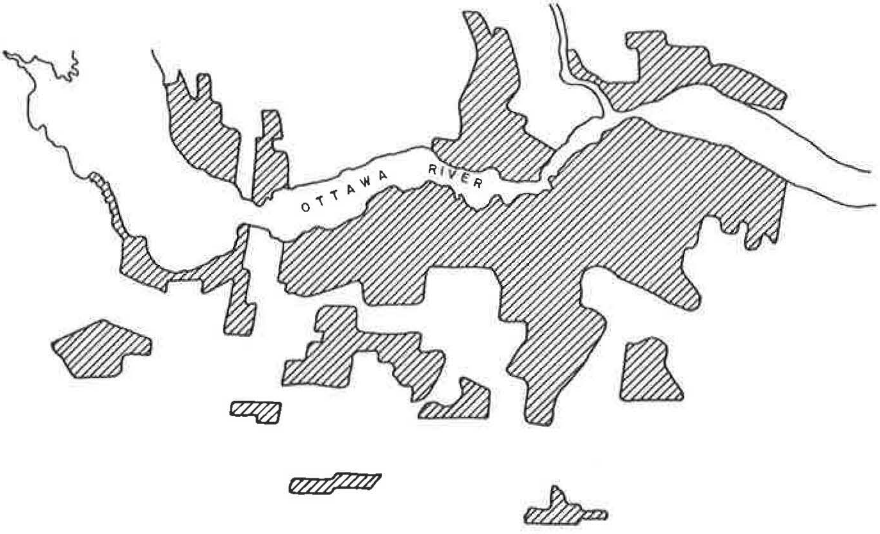


Figure 6. Extent of urban postal addresses.

2. Block face records are used to record segments where both even and odd numbers are on the same side of the street. Generally, in this case only one side of the street is developed.

3. Out-of-sequence addresses are coded as individual entries, the coordinates being derived manually.

Although it is possible to accommodate almost any situation, the amount of manual effort required to build the directory increases substantially as the street addressing pattern becomes less regular.

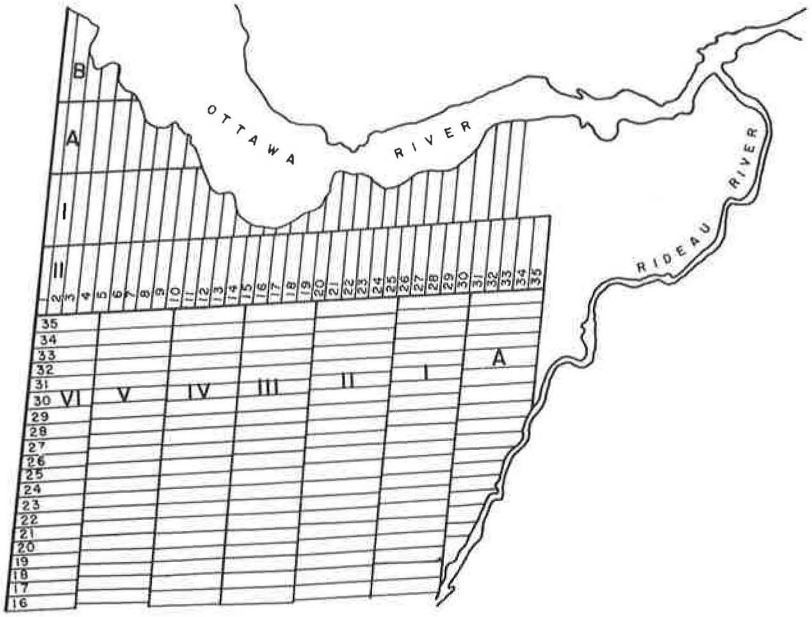


Figure 7. Lots and concessions.

In order to include the rural fringe areas in the geocoding process, the methods of rural addressing and land identification were examined. Three methods were identified: (a) rural mail delivery routes; (b) P.O. Box numbers; and (c) rural land survey (concessions and lots, Fig. 7).

The concession/lot system was found to have spatial characteristics similar to the concept of the street segment. The significant characteristic for the geocoding system is that there is a linear relationship between all lots within one concession. Therefore, if the lot numbers are used in place of the address range and the concession name is substituted for the street name, a pseudo street segment can be created to represent a whole concession. If this pseudo segment is located through the middle of the concession, each lot can be split into two pieces (north half, south half or east half, west half) for geocoding, each piece being 100 acres in size.

The changes introduced to the system have broadened its scope and increased its potential application. For example, the Canadian Dominion Bureau of Statistics is currently investigating geocoding applications through participation with the National Capital Commission in developing the Ottawa street address system. The Dominion Bureau of Statistics (DBS) plans to use geocoding techniques for census data as well as other nationally collected statistics, thereby giving DBS the capability to summarize any of their data to any set of areal units as a service to users. The ability to automatically geocode data on the basis of various identifiers substantially increases usefulness of the data in all studies relating to urban areas.

REFERENCES

1. Dial, Robert B. Street Address Conversion System. Res. Rept. No. 1, Urban Data Center, Univ. of Washington, 1964.
2. Calkins, Hugh W. Operations Manual for Street Address Conversion System. Res. Rept. No. 2, Urban Data Center, Univ. of Washington, 1965.