

Census Data as a Source for Urban Transportation Planning

RONALD J. FISHER and ARTHUR B. SOSSLAU, Tri-State Transportation Committee

•POPULATION, housing and numerous other socioeconomic data collected and published by the United States Bureau of the Census decennially have been a valuable source to urban transportation planning studies for analyses and forecasts. In 1960, information was collected for the first time on the journey-to-work and automobile ownership. These data have greatly enhanced the value of the census for urban transportation planning studies. In addition to the printed reports, the 1960 Census data are also available on computer magnetic tapes for use by other agencies.

The Tri-State Transportation Committee, financed by Connecticut, New Jersey, and New York, the Federal Government through the U.S. Bureau of Public Roads, and the U.S. Department of Housing and Urban Development, has made considerable use of census data both in printed form and on magnetic tapes. In addition to analyses made, a number of reports have been prepared, covering the journey-to-work in the Tri-State Region, as well as tape files that can be readily used by other agencies in the region. The experience gained in the use of this source may be of value to others contemplating its use.

The purpose of this paper is five-fold:

1. To present the uses made of the census data by the Tri-State Transportation Committee.
2. To indicate possibilities for the use of future censuses.
3. To describe the census data sources available on magnetic tape.
4. To discuss the limitations of the data for transportation planning.
5. To help those interested in the census as a data source for transportation planning to form suggestions for improving future censuses.

USES OF CENSUS DATA

A considerable number of uses and analyses of the census data have already been undertaken by the Tri-State Transportation Committee, including:

1. Selection of the Tri-State cordon line.
2. Selection and verification of home interview sampling accomplished from utility company records.
3. Study of trends in population and housing units from 1940 to 1960.
4. Examination of travel-oriented characteristics, such as mode choice and trip length.
5. Comparison of data obtained from the Tri-State 1 percent home interview survey with comparable data from the census to check the validity of the survey.
6. Examination of residential mobility characteristics.
7. Preparation of displays and reports that have provided insights into work travel and related characteristics, which will be further analyzed from the home interview survey.

Selection of Cordon Line

The choice of the Tri-State cordon line was predicated on the following basic factor: the area thus enclosed would include all continuous urban development as well as most of the expected population increase estimated for the future (1980). The two items

considered in determining the extent of urbanization in the Tri-State Region were population density and car ownership. These data were gathered from printed census sources. Figure 1 is a plot of population density per acre by municipality for the Tri-State Region. It is a good indication of population dispersion from highly intense urbanization centering on Manhattan to a suburban and then rural density as the distance from the center increases. It provides a visual idea of the area that should be encompassed by a cordon line.

The census also provided a more usable indication of car ownership by geographic area than was readily available from automobile registration sources. Figure 2 shows the per-acre distribution of automobiles by county and again provides a visualization of a cordon line location.

Population and auto availability data as well as criteria on roadside interview station locations permitted a number of tentative lines to be established, each with an encompassed population determined from the census. The tentative cordon lines were used to determine the extent and shape of the study area. Between the tentative cordons, segmented population figures were calculated to evaluate the addition of certain land areas to each preceding cordon line to determine if each particular population increase was warranted. The line finally chosen is shown in Figure 3.

Selection and Verification of Home Interview Sample

A 1 percent home interview survey was collected from households within the cordon line shown in Figure 3. A number of sampling frames were used as sources for selection of a 1 percent probability sample of living places in the cordon area.

Within New York City, the census provided the basis for the sample selection. In cooperation with another public agency, which had already completed the necessary preparatory work, a 1 percent clustered area probability sample was selected. The sampling frames were defined by two strata:

1. The civilian, non-institutional population living in housing units and other special dwelling places in existence according to the 1960 Census of Population and Housing. Census block data were used for this purpose.
2. Housing units built during the period between the 1960 Census and February 28, 1963, as represented by occupancy certificates obtained from the New York City Department of Buildings.

Outside of New York City, the records of the various electric utility companies were used as a sampling frame. However, some towns were not covered this way. Instead, these enclaves were sampled by means of a block field listing procedure. Immediately after the selection of the sample from the frame, various checks were applied to insure its reasonableness when compared to published sources. Again, census data on housing units by municipality were used after updating by building permit data for this comparison. Any large discrepancies were checked in detail.

Study of Population, Housing and Employment Trends 1940-1960

As a preliminary step to understanding past growth, and as an aid to the forecasting of future population, housing and employment characteristics of the residents of the region, a "county level" minimum comparability file has been developed and made operational. This magnetic tape file contains 18 data items including: "population"—all persons by 5-year age groups; "housing units"—total by race and tenure; and "labor force"—occupation by sex and industry. The file was prepared from published and unpublished U.S. Census sources for 1940 and 1950 and from census tapes for 1960. It covers all 25 counties in the Tri-State Region. The data stratifications used by the census for the three periods have been compressed and regrouped to provide definitional consistency and comparability over the period 1940-1960.

Selected items can be retrieved as needed, in phase with the requirements of the Tri-State Transportation Committee's analytical progress. The file has been used in preparation of a first projection of total regional and county populations to 1985, preparation of a series of county population density maps over the 1940-1985 period, and

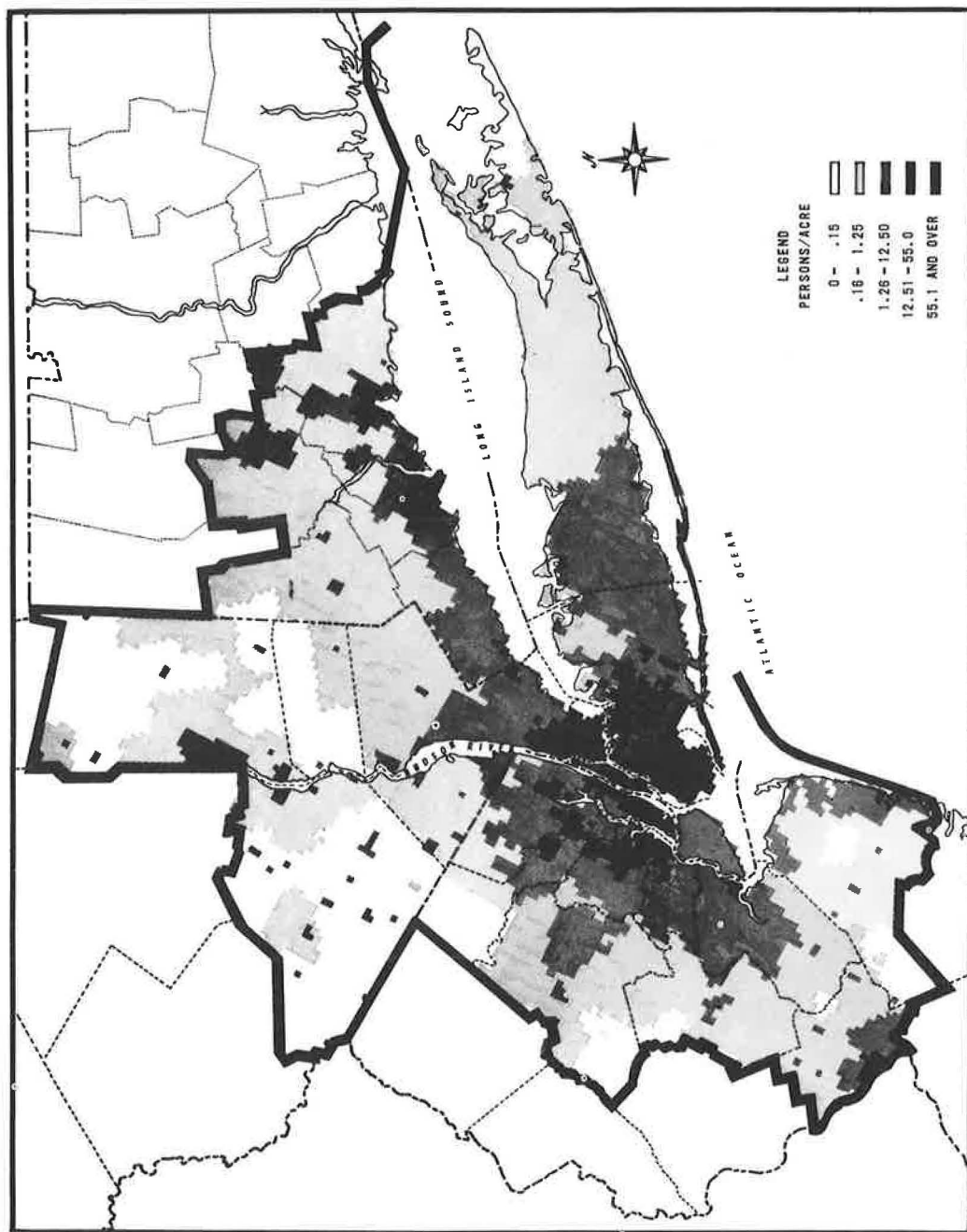


Figure 1. Population density by municipality.

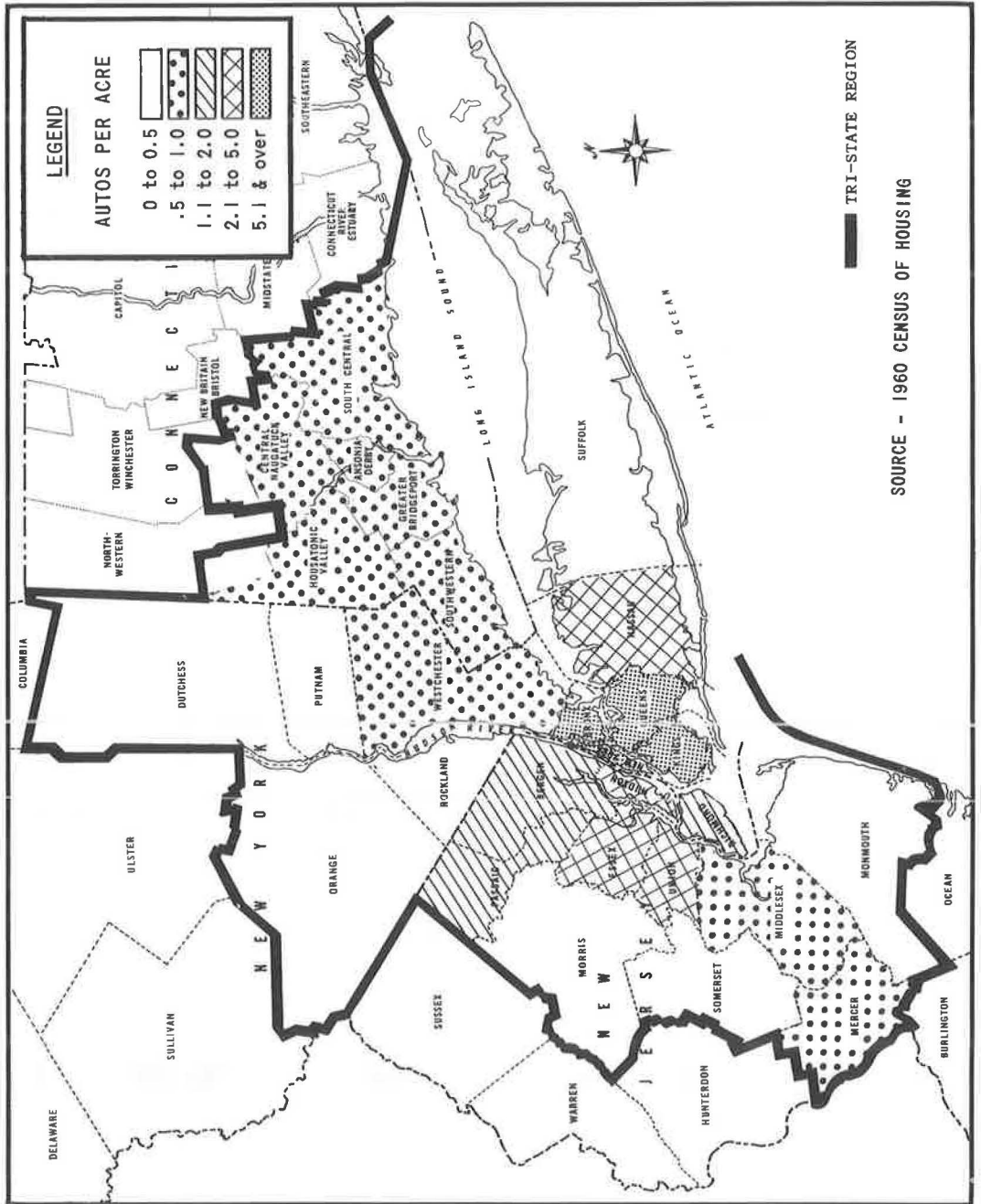


Figure 2. Autos per acre by county.

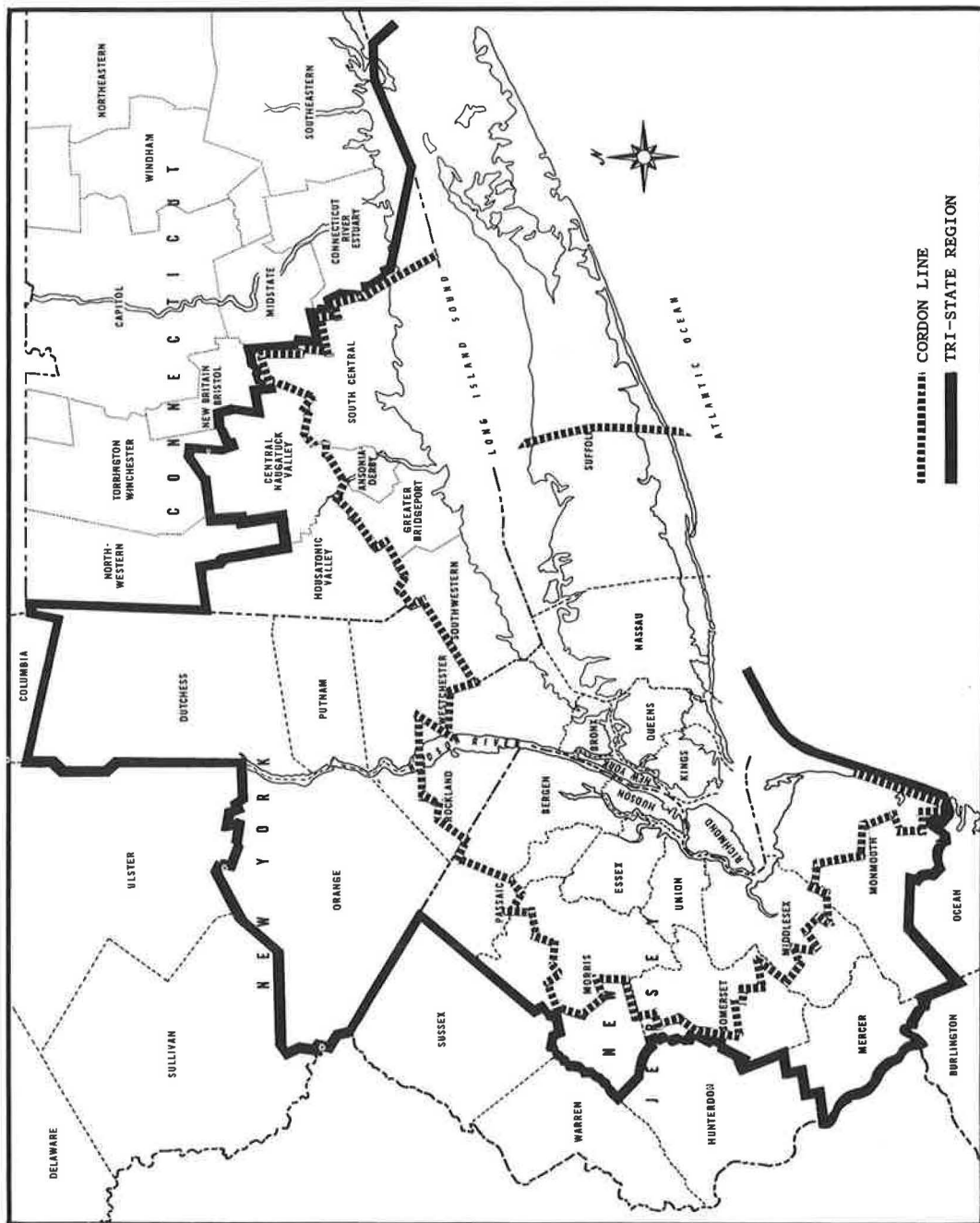


Figure 3. Tri-State cordon area.

preparation of preliminary estimates of the number of occupied housing units by structure type by county as an input to the interim traffic estimating procedures developed by Tri-State. The file also assisted in estimating the 1985 distribution of families and unrelated individuals by income class and in preparing preliminary sketches of land development alternatives.

More detailed analysis to which this file will be put include such efforts as age-cohort survival techniques of population projections, and intensive investigation of relationships between population distribution, composition (age, sex, race, etc.) and developments in income, employment, occupation and levels of education.

Travel Oriented Characteristics Such as Mode Choice and Trip Length

Although the level of areal detail available in the census journey-to-work material is at present gross (being composed of counties and major cities within the Tri-State Region), analysis of trip lengths by workers using different major modes of travel and of varying socioeconomic characteristics was made.

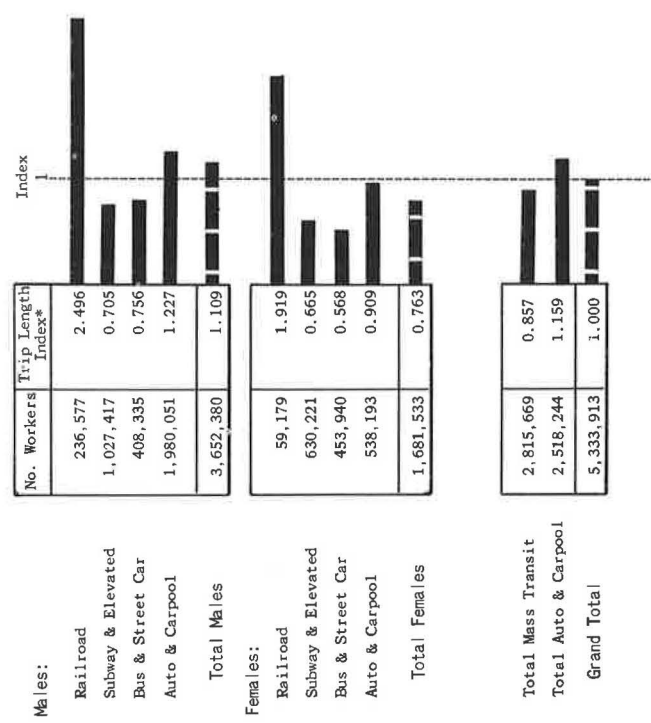
For the Tri-State Region, the 29BB census file was consolidated to 3332 records. Each record contained a residence code, a work place code, and information on the number of journey-to-work trips by mode, by sex, by age, by occupation and by income. The geographic areas were counties and major cities (groups of towns in Connecticut) and amounted to 67 zones in the Tri-State Region, 47 of which had complete employment coverage. The approximate geographic centroids were determined for each zone, and their coordinates were entered in each zone-to-zone record. From these data, the distances between the residence and work places were calculated and entered in each record, making possible the calculation of average trip lengths for each mode and socio-economic characteristic in the record.

On the basis of these data, a few limitations must be placed on trip lengths. First, people reported the place worked the longest during the week prior to the interview, if they had more than one job. Data were recorded at the usual place of residence even though the respondent may have been interviewed elsewhere. This tends to increase the trip lengths. Second, the gross areal detail tends to increase all average trip lengths. Third, only the primary mode is recorded from the respondent's interpretation of the mode involving the longest travel distance. Average trip length for a mode would be different if each leg of a multi-mode trip could be given weighted consideration. Finally, the data do not completely cover all employment in the Region. Only data for Standard Metropolitan Statistical Areas (SMSA's) with populations of 250,000 or more (there are seven SMSA's in the Tri-State Region) were processed by the Bureau of the Census. Approximately 90 percent of the Region's workers are represented in the 29BB tapes.

Consequently, the trip lengths developed from present census sources should best be viewed on their relative significance rather than their absolute value. An index, derived by dividing the average trip length for all workers using mass transit or auto and carpool into each subpopulation defined by mode or socioeconomic status, was calculated for this purpose. The data proved useful, since they offered information on trip lengths before the Tri-State travel surveys were processed.

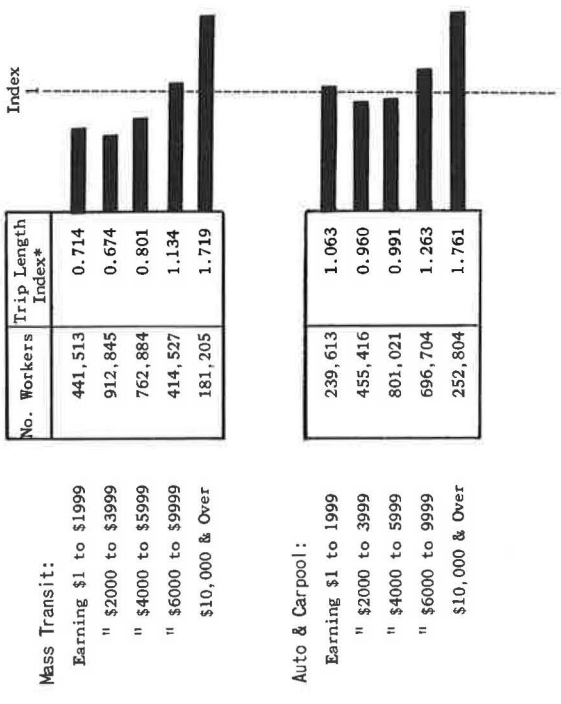
Figures 4 through 8 show the results of the trip length analysis. A recapitulation of the basic findings from these trip length data follows:

1. People using some form of mass transit for work trips travel 30 percent less on the average than those using auto or carpool.
2. Considering just the railroad commuters in the portion using mass transit, one finds longer average trip lengths than in any other grouping of workers analyzed.
3. Workers grouped by income or mode and income indicate longer trips are made by higher incomes.
4. Trip lengths vary slightly with age.
5. Males are likely to make a longer trip than females of the same age or occupation, or traveling via the same mode.



*THE BASE FOR COMPUTING THESE INDICES IS THE AVERAGE TRIP LENGTH OF ALL WORKERS TRAVELING VIA MASS TRANSIT OR AUTO AND CARPOOL.

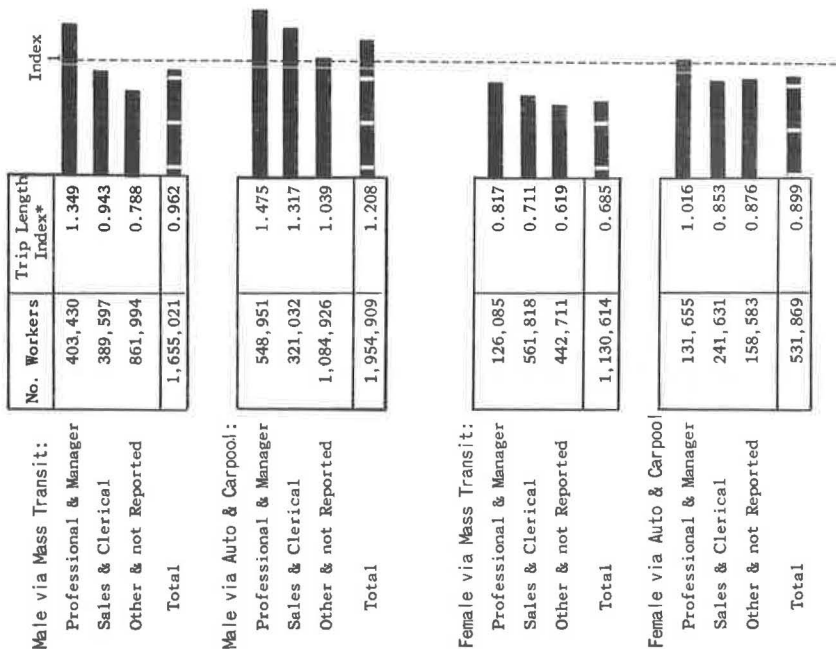
Figure 4. Average trip lengths for the journey-to-work in the Tri-State Region, by sex, by mode.



* EXCLUDED FROM THIS DISTRIBUTION ARE WORKERS NOT REPORTING MODE AND WORKERS WITH NO EARNINGS.

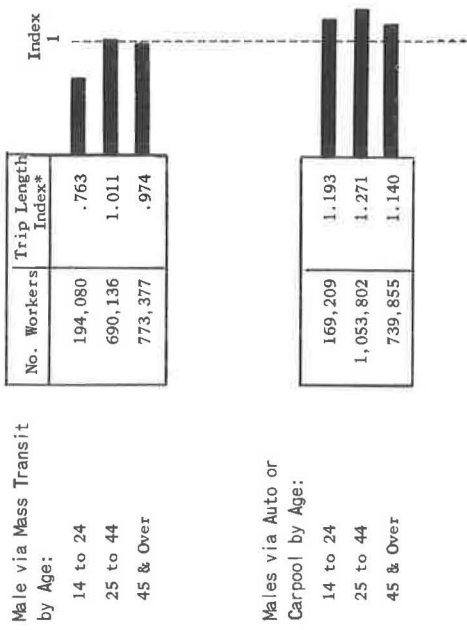
** THE BASE FOR COMPUTING THESE INDICES IS THE AVERAGE TRIP LENGTH OF ALL WORKERS TRAVELING VIA MASS TRANSIT OR AUTO AND CARPOOL.

Figure 5. Average trip lengths for the journey-to-work in the Tri-State Region, by earnings, by mode.



* EXCLUDED FROM THIS DISTRIBUTION ARE WORKERS NOT REPORTING MODE AND WORKERS IN THE ARMED FORCES.
 ** THE BASE FOR COMPUTING THESE INDICES IS THE AVERAGE TRIP LENGTH OF ALL WORKERS TRAVELING VIA MASS TRANSIT OR AUTO AND CARPOOL.

Figure 6. Average trip lengths for the journey-to-work in the Tri-State Region, by sex, by mode, by occupation.



* THE BASE FOR COMPUTING THESE INDICES IS THE AVERAGE TRIP LENGTH OF ALL WORKERS TRAVELING VIA MASS TRANSIT OR AUTO AND CARPOOL.

Figure 7. Average trip lengths for the journey-to-work in the Tri-State Region, by mode, by sex, by age.

6. Trip lengths for people who moved to a different house between 1955 and 1960 are slightly longer than those who did not move.

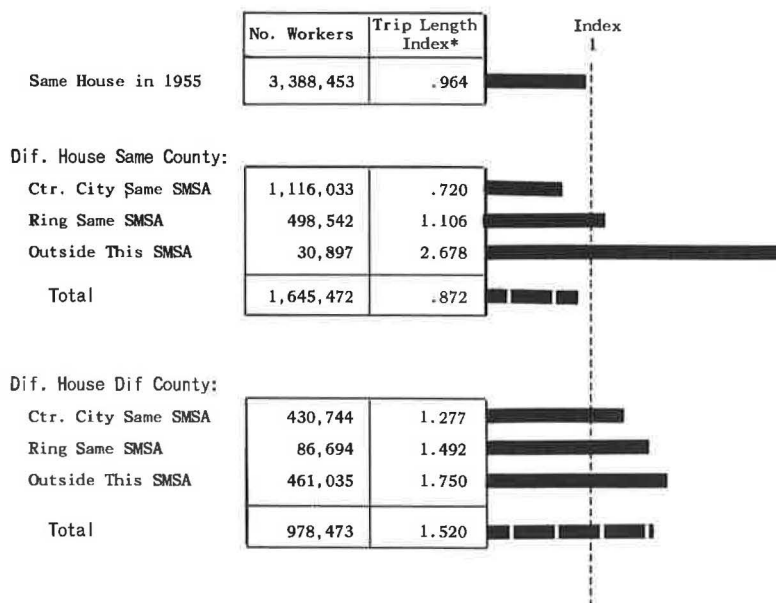
7. People who moved into the Tri-State Region since 1955 or lived in a different county in a different SMSA in the Tri-State Region in 1955, traveled 80 percent farther to work than people in the Region who had not moved.

In addition to the trip length analysis, tabulations were produced from 29BB journey-to-work tapes for each of the 47 zones having complete data coverage in the Tri-State Region that show the cross correlations between the percentage of workers using mass transit and the following socioeconomic characteristics: age, sex, earnings and occupation.

Due to the rather large data areas, only a limited overview was possible. However, this overview provided groundwork for more detailed work trip analysis to be undertaken with the extensive travel surveys made by the Tri-State Transportation Committee. A summary of the findings will be presented here.

The propensity for males and/or females in particular age groups to use mass transit for work trips is shown in Figure 9. These data are further stratified as to population density at place of residence and place of work for internal trips, at place of residence for export trips, and place of work for import trips. Figure 9 also shows the relationship of the captive rider market to mass transit usage. The females as a group are more apt to use mass transit than males. Mass transit usage dips for both sexes in the 25 to 44-year old age group.

Figure 10 shows the relation between worker earnings and mass transit usage. Internal and import worker streams at this rather gross level of detail show an inverse relationship between earnings and mass transit usage. Figure 11 shows that this is mostly caused by an unusually large portion of workers with high earnings going to Manhattan. In general, a larger portion of the export workers with high earnings are



*THE BASE FOR COMPUTING THESE INDICES IS THE AVERAGE TRIP LENGTH OF ALL WORKERS TRAVELING VIA MASS TRANSIT OR AUTO AND CARPOOL.

Figure 8. Average trip lengths for the journey-to-work in the Tri-State Region, by residence in 1955.

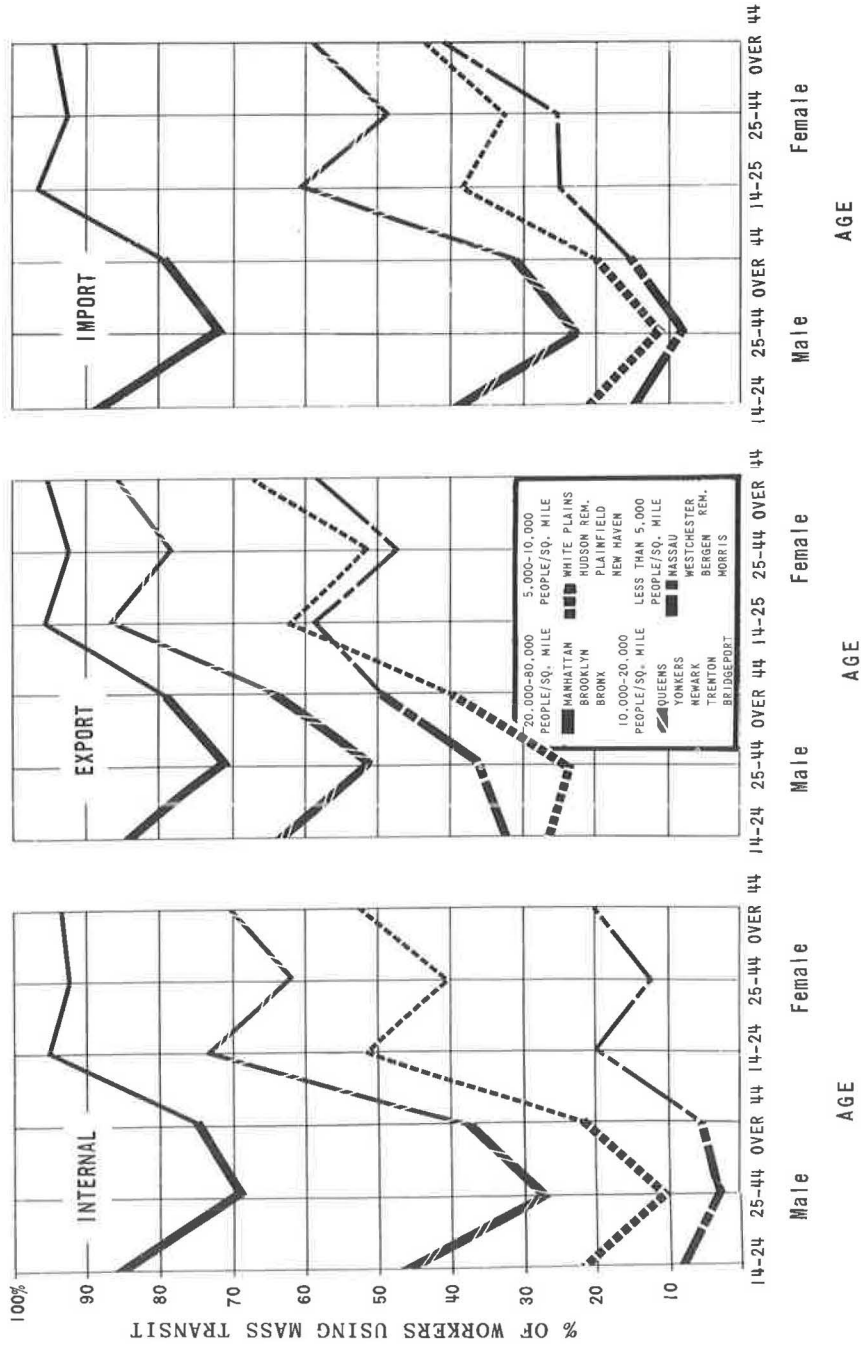


Figure 9. Mass transit usage by age, sex, population density.

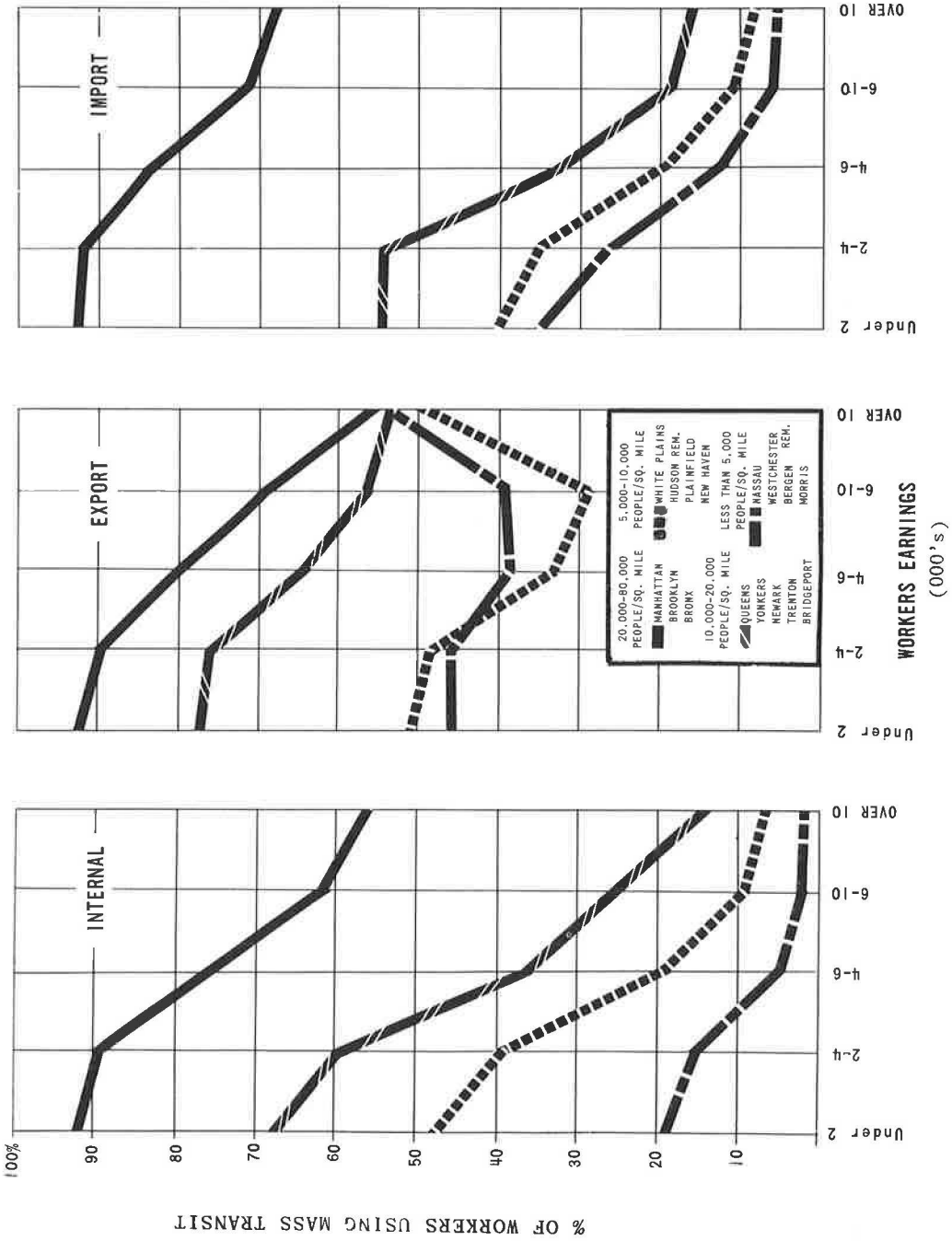


Figure 10. Mass transit usage in different density and income strata.

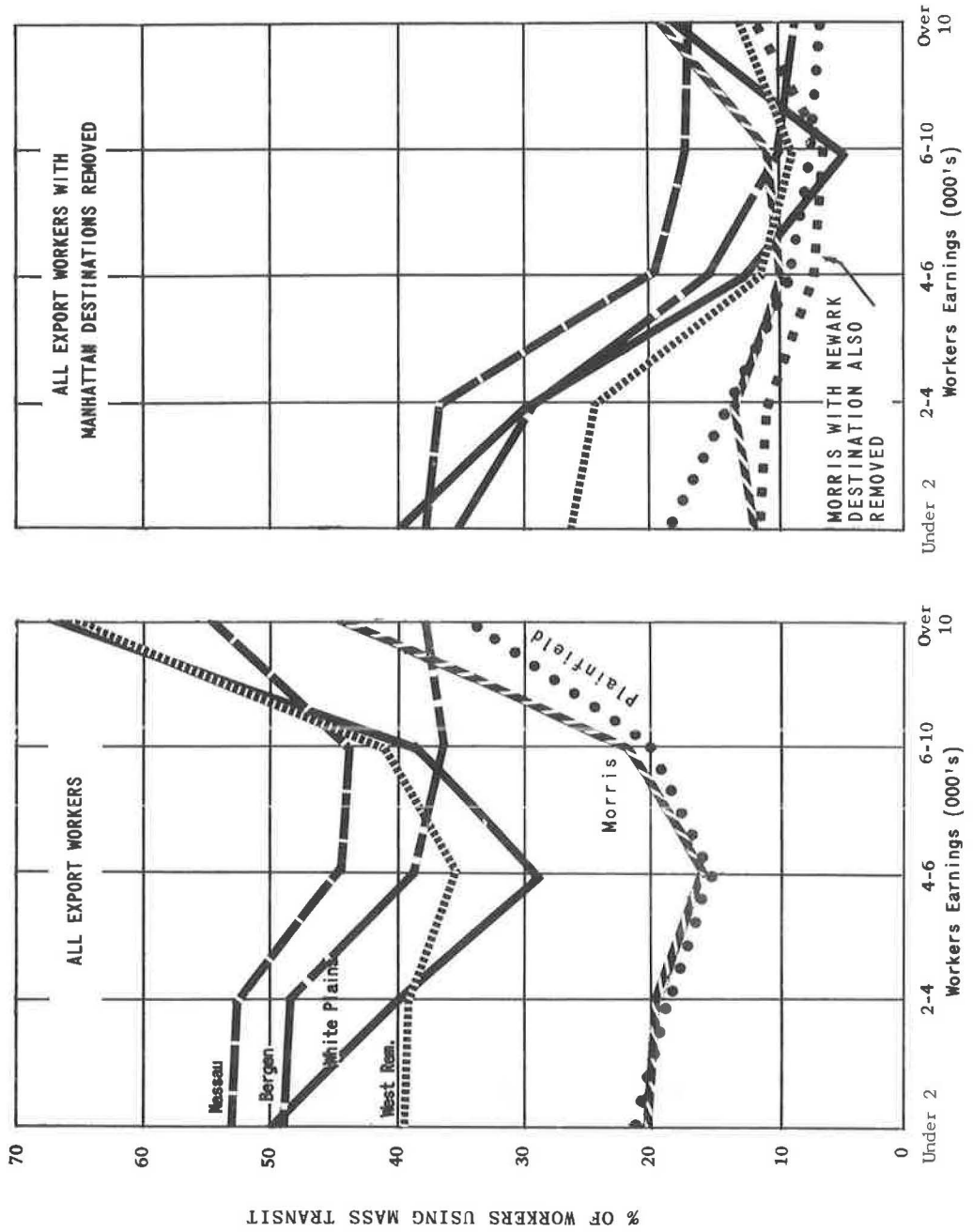


Figure 11. Manhattan influence on mass transit usage.

destined for highly dense areas. It appears that at this gross level of detail mass transit usage decreases with rising income when the density of the destination does not increase greatly over the origin.

Figure 12 shows the mass transit usage to Manhattan from each of the three states in the Tri-State Region. It gives an interesting hint of how the cost, relative comfort and convenience of transit service influence the market. At low levels of income, the rate of mass transit usage to Manhattan is less in New Jersey and Connecticut than in New York. In New York State, the rate of mass transit usage is inversely related to income and the graph is very similar to the one plotted for Manhattan's import workers. Rising mass transit usage with rising income is very evident in Connecticut.

Again, looking for the relationship at a lower level of detail, Figures 13 and 14 show mass transit usage between particular origins and destinations. Unfortunately, this is the lowest level of detail available on the census computer tapes.

Comparison of Survey Data With Census Data

A traditional use of the census data is currently being undertaken by the Tri-State Transportation Committee. That is, the comparison of home interview survey results with that of the census. The Tri-State home interview sample contains 1 percent of the households in the region. The census contains some comparable items on a 25 percent and 100 percent basis. Reasonableness checks of the survey data with the census provide an indication of the completeness of the survey and the validity of the sample. The types of items which will be compared by geographic area include:

1. Population by age and sex;
2. Number of dwelling units;
3. Number of dwelling units by the number of units in the structure;
4. Income;
5. Occupation and industry of resident workers;
6. Number of vehicles available;
7. Number of families with 0, 1, 2, and 3 or more vehicles;
8. The distribution of the first work trip on a gross area basis, such as county;
9. Various ratios obtained from the above items such as persons per household and cars per person.

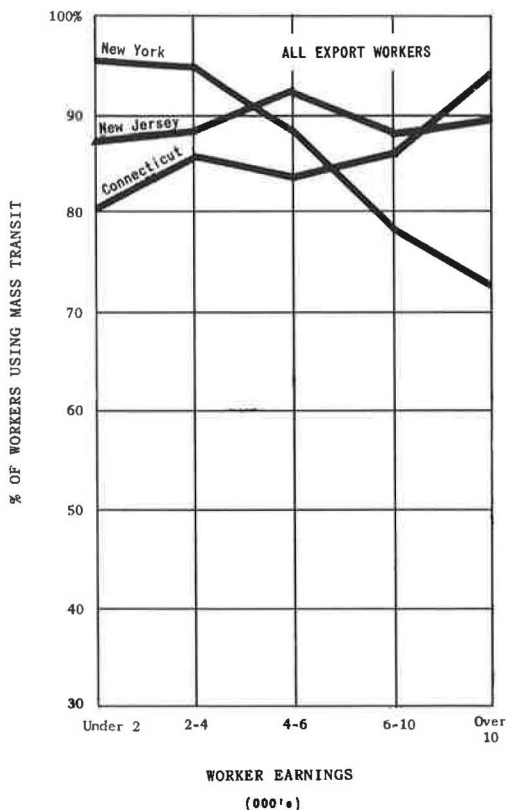


Figure 12. Mass transit usage to Manhattan from portion of each state in Tri-State Region.

The Census data are for 1960; the home interview survey is for 1963. Care must be taken that apparent differences are not due to changes occurring during the three years and that adjustments are provided where necessary prior to comparison. With this consideration in mind, many data items will be compared on a proportion basis as well as absolute values. For example, the proportion of workers to Manhattan from a residence area such as Brooklyn will be compared as well as the total number of work trips from Brooklyn to Manhattan.

Residential Mobility

In the 1960 Census of Population, mobility data were collected for all persons five years and older living in a standard

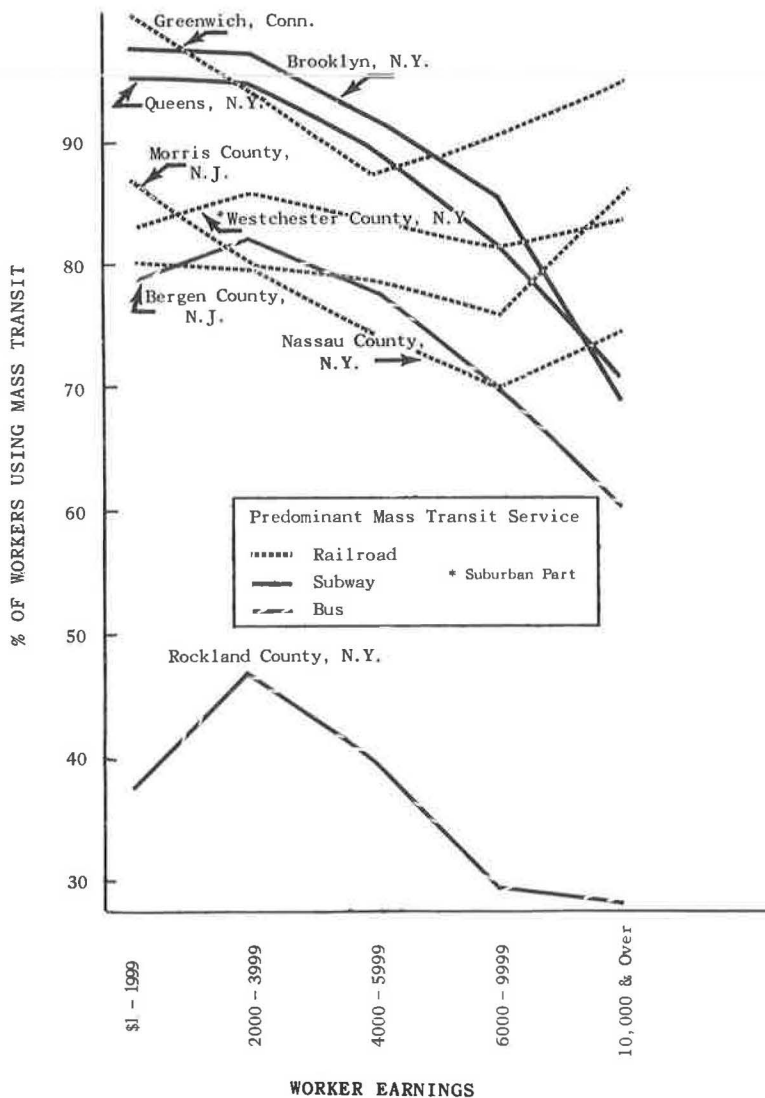


Figure 13. Mass transit usage to Manhattan from selected origins.

metropolitan area. These data describe the residence of these persons on April 1, 1955, according to the following classifications:

Residence in 1955 for Persons 5 years old and over, 1960

- Same House as in 1960
- Different House in U.S.
 - Central City of this SMSA
 - Other Part of this SMSA
 - Outside this SMSA
 - North and West
 - South
- Abroad
- Moved, Residence in 1955 not reported

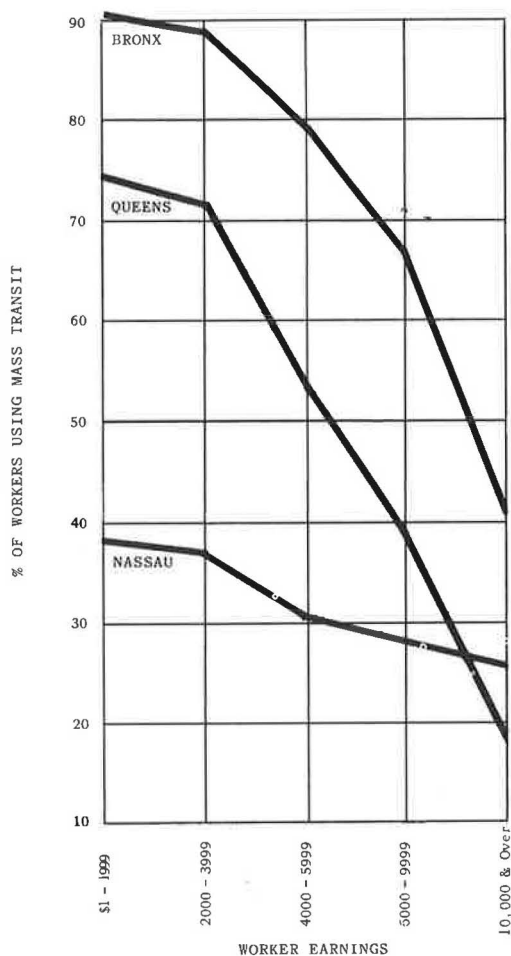


Figure 14. Mass transit usage to Brooklyn from selected origins.

These mobility data viewed at the county or city level enable the analyst to take an overall view of the residential movements within the region. An analysis of the relative stability of the population by geographic area is useful for forecasting population and employment related characteristics.

Data for the New York SMSA are used to illustrate some of the findings from a preliminary probe that can be derived from the census on mobility patterns. For New York City, movement (1955 to 1960) has been within, with 88 percent of the moving population (persons who reside in a specific house or apartment in New York City in 1960 but resided elsewhere in the U.S. in 1955) contained in the city. In fact, more people migrated to New York City from abroad (foreign country, Commonwealth of Puerto Rico, or a possession of the United States) than entered this area from the United States outside the New York SMSA. The counties outside New York City may also be compared as to trends in the five-year period 1955-60 (see Table below).

This mobility may also be viewed at a lower level using the census tracts as basic units. At this level, Table H-2 of U.S. Census of Population and Housing (PHC 1) reports the year moved into the present housing unit by the following stratifications: (a) 1958 to March 1960, (b) 1954 to 1957, (c) 1940 to 1953, and (d) 1939 or earlier.

Use of the census 29BB computer tapes yields additional data in explaining the effect of mobility patterns. The tape output describes the universe of workers who

have moved since 1955 into the following two categories: (a) moved within the county, and (b) moved into the county. In addition, the above groupings are also stratified into: (1) central city, and (2) ring of the SMSA under study.

County	Persons Residing in County 1960 but in Diff. House in U.S. in 1955		All Persons (% Persons that have not changed residence 1955-60)
	% from N.Y.C.	% from Total N.Y. SMSA	
Nassau	50	89	58
Suffolk	38	89	48
Rockland	32	79	47
Westchester	23	83	54

Preparation of Displays and Reports

Summarization and analysis of the census material was undertaken by the Tri-State Transportation Committee shortly after interviewing for the travel surveys was started in the field. During the period in which the Tri-State survey data were collected, coded, edited, factored and summarized, the census data offered a source for obtaining insights into work travel and related characteristics which would be further analyzed from the home interview survey. The material also proved useful for some immediate action, or short-range planning work, undertaken by the Committee and for answering specific questions concerning transit usage in a few locations in the Tri-State Region.

The data, in magnetic tape form, also lend themselves readily to the preparation of automatic data displays on the Tri-State modified EAI model 3500 data plotter. The basic source for the automatic plots was the 29I census file. This tape contains detailed population characteristics for the 4103 census areas (generally tracts) in the Tri-State Region. For a number of data items, inputs were prepared for the plotter that would allow plotting a different symbol for each of 20 ranges established for each item. These symbols are plotted at the geographic centroid of each census area. A final display is prepared manually after analyses of the plot.

A later advancement of this technique provided for the assignment of grid squares to each census tract with the subsequent automatic plotting of a completed color display.

The tract data from the 29I census tapes also allowed XY data plots to be automatically prepared by a 1401 computer on a 1403 printer. For example, a plot was prepared relating percent auto usage to automobiles per household as shown in Figure 15.

A number of descriptive reports prepared from the census material presented a good first picture of the journey-to-work that would be subsequently brightened by the Tri-State travel surveys. Their titles are listed as follows along with a brief annotation.

1. Journey-to-Work in the Tri-State Region, May 1964: Describes work travel in the Tri-State Region from three viewpoints: (a) those workers leaving each county to work (export), (b) those workers coming into each county for work (import), and (c) those living and working in the same county (internal). A square trip table containing 24 counties is contained for mass transit trips, for automobile trips, and for total trips.

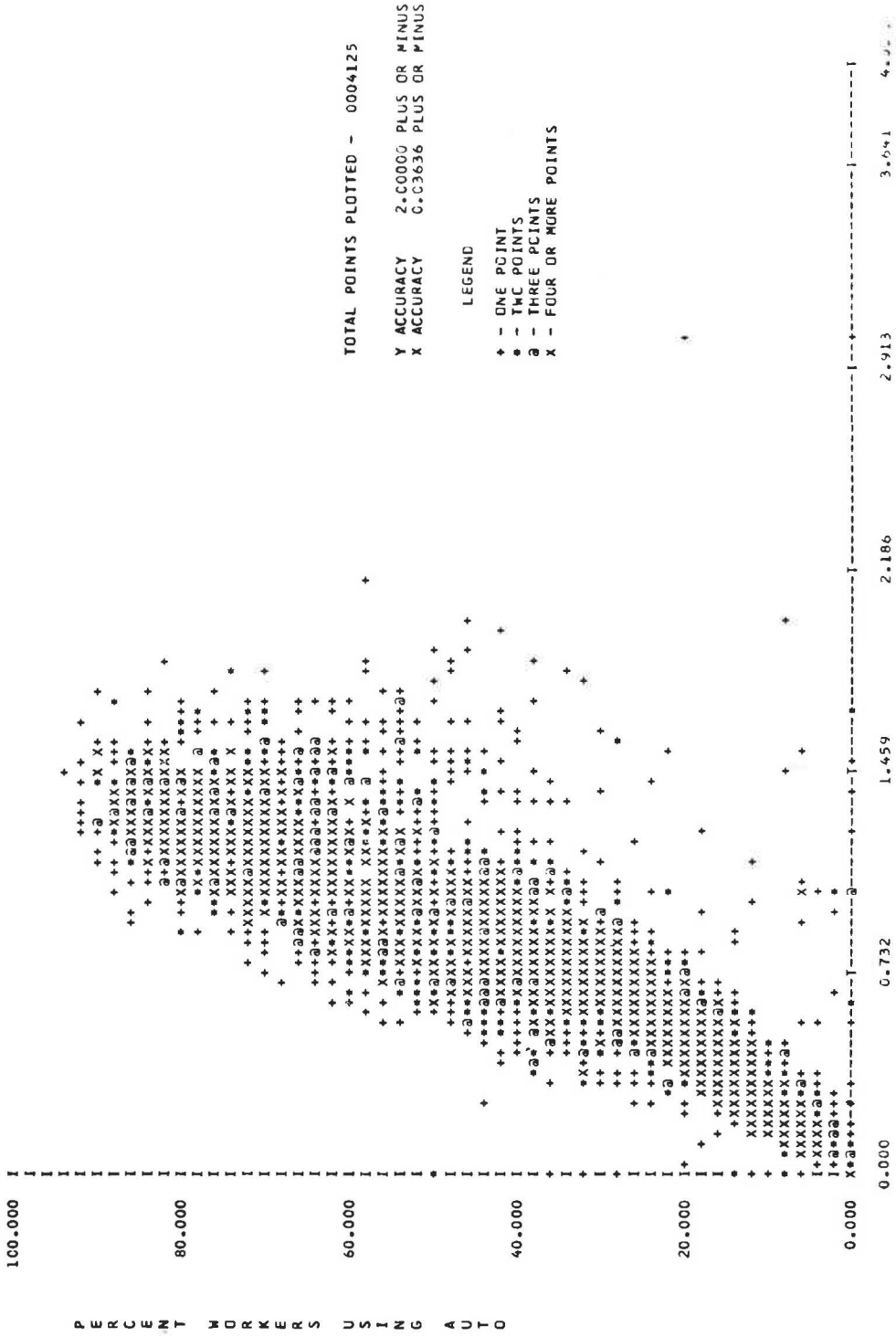
2. Journey-to-Work in the New York City SMSA: Describes the characteristics of workers using public and private transportation by a number of socio-economic characteristics, including income, occupation, sex, and age. The subgroups described are those who live and work in New York City, those who live and work in a suburban county, those who live in the suburbs and work in New York City, those who live in New York City and work in the suburbs, etc. Figure 16 is a sample display for those who live and work in New York City.

3. Characteristics of Workers by Place of Residence—Interim Technical Report 4014-3442: Set of tables for each of 47 counties and major cities in the Tri-State Region (for which complete coverage in the journey-to-work survey is available) containing the number and percent of workers by mode, occupation, income, hours worked, age and sex, housing, schooling, mobility, and class of worker. The data are listed by place of residence for three worker groups: internal, export and total workers.

4. Characteristics of Workers by Place of Employment—Interim Technical Report 4014-3442: Same as Item 3 except data are summarized at the employment place rather than residence and the three worker groups are internal, import, and total employees.

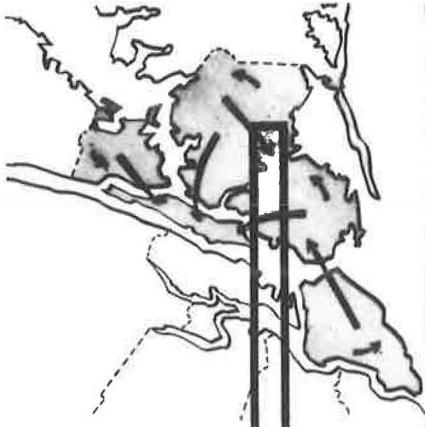
5. Characteristics of Mass Transit Users by Place of Residence—Interim Technical Report 4014-3442: Set of tables for each of 47 counties and major cities containing information on the users of mass transit. The number of transit users are presented by age and sex, occupation and sex, earnings, age and occupation. The percent of each group using mass transit is also included. The workers are further subdivided into internal workers, export workers and total workers.

6. Characteristics of Mass Transit Users by Place of Employment—Interim Technical Report 4014-3442: Same as Item 5 except data are summarized at the employment



AUTOMOBILES PER HOUSEHOLD

Figure 15. Percent of workers using auto vs autos per household.



selected characteristics of workers classified according to whether they usually went to work by public transportation or by private automobile

2,040,000 Public transportation users	66.5%
540,000 private automobile users	17.9%
480,000 all other	15.6%
3,060,000 Total workers	100.0%

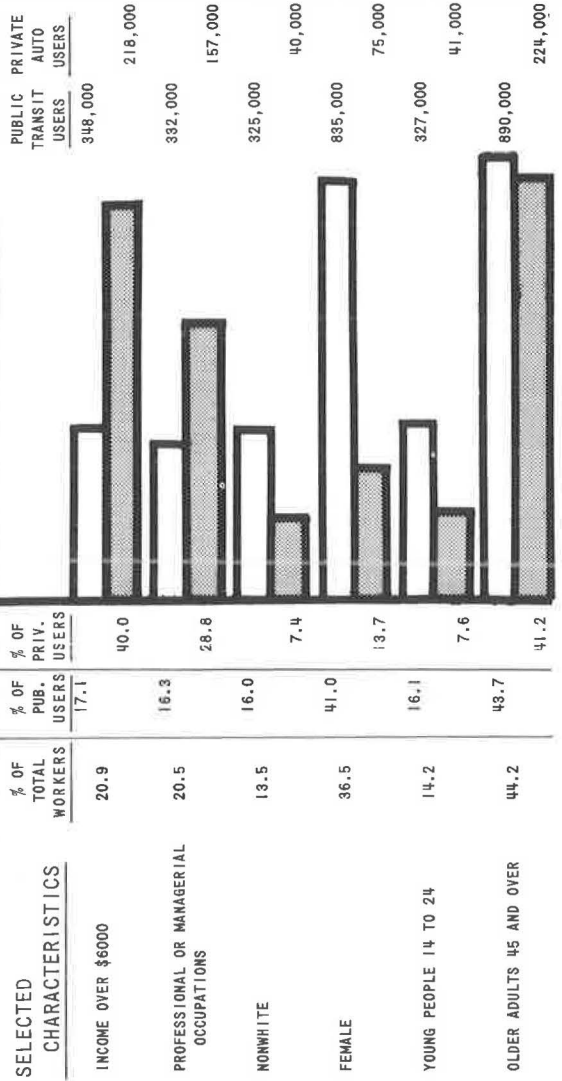


Figure 16. Employed persons living and working in New York City.

place rather than residence. The three worker categories are internal, import, and total employees.

7. Employment by Industry in the Tri-State Region—Ronald J. Fisher, Tri-State Transportation Committee Technical Bulletin, May 1965: Summarizes the workers by industry reported in the census journey-to-work survey for each of the seven SMSA's in the Tri-State Region and compares these with the Chicago, Philadelphia and Pittsburgh SMSA's. Also shows the relative number of workers in each industry category by geographic area within the Tri-State Region.

In addition to the foregoing reports, the tapes obtained from the Bureau of the Census and processed by Tri-State have been used by several agencies in the Tri-State Region. One such use was for determining the impact of New Haven Railroad passenger service discontinuance on the highway system. The data have also been used in the study of possible inconveniences that might result from the elimination of certain stops on a New Jersey railroad.

The data from these same sources are also being analyzed in connection with an extensive mass transportation demonstration project in the Queens-Long Island sector of New York. The Port of New York Authority is using the data as a source for predicting Hudson River crossings. The data have also provided useful material for inclusion in talks before local civic groups.

USE OF FUTURE CENSUS MATERIAL

The work of a number of transportation planning studies has determined those characteristics that appear to be the most reliable indicators of trip generation in an urban area. These can usually be broken into two categories: (a) resident end characteristics, (b) nonresident end characteristics.

At the resident end, the total trips generated by the residents of an urban area appear to be strongly related to automobile availability, population density, housing type, income, family size, distance from the central business district, and accessibility. Residential destinations also have been found to be related to these same variables. Other indicators have been used but are generally correlated with one or more of the variables listed above. In fact, use of all seven variables as shown above would probably be a folly since they are also very much correlated with one another.

At the nonresidential end, the variables which have been found useful in estimating the number of trips generated are: employment by industry or occupation, retail sales, the floor area or gross area of various land use types, and accessibility. Again, other variables can be added to the list, but generally with little gain, and inclusion of all the above would again probably be a mistake.

The variables mentioned for residential and nonresidential trip estimation fall into two general categories: (a) those related to the characteristics of the population, and (b) those related to the transportation system and the land development of an area. The Bureau of the Census has provided much of the former data in the past and can provide even more useful data for future planning work.

The purpose here is to outline a hypothetical procedure for estimating trips in a region (one that is not far different from those used by a number of transportation studies) and show how census data can be useful to such a procedure.

Trip Model

1. The total trips generated by the residents of an urban area (and trip destinations to residential land) are equal to some function of car availability and net residential density (structure type, such as single-family, two-family, three- and four-family, and multi-family structures may be a substitute variable).

$$Gr_i = f(AA_i, d_i)$$

where

Gr_i = trips generated by the residents of zone i ;
 AA_i = autos owned by the residents of zone i ; and
 d_i = net residential density of zone i .

2. Automobile availability is equal to some function of family income and net residential density.

$$AA_i = f(I_i, d_i)$$

where

I_i = median income of residents of area i .

3. The trip destinations at nonresidential land are equal to some function of the employment density and number of employees by industry type.

$$Gnr_i = f\left(\frac{E_i}{A_i}, e1_i, e2_i, \dots, en_i\right)$$

where

Gnr_i = nonresidential trips to zone i ;
 E_i = total employees in zone i ;
 $e1_i$ = employees in industry 1 in zone i ; and
 A_i = area of nonresidential land in zone i .

For sake of discussion, it is assumed that the mode of travel used by the residents of an area may be forecasted by the above variables and system characteristics such as cost and speed, and that the estimation procedures developed are based on data collected from travel inventories. These procedures are then used to forecast travel to some future year based upon estimates of population, employment, automobile ownership (based on income and density), and the land area to be allocated to residential and nonresidential uses by area. Intermediate year forecasts are also made, perhaps on a five-year basis. Future censuses will allow a critical review of the forecast by providing, at least once every ten years, those variables upon which the trip estimates rest. That is, every ten years rather complete information on population, income, automobiles available, and, hopefully, employment data by small areas such as census tracts will be available. All of these data were available by census tract in the 1960 census, except for employment data at the work place in sufficient areal detail.

Employment data were available from the journey-to-work survey, for what is similar to the first work trip, by occupation and industry, but at a gross level of detail (generally county and major city). Of course, the number of workers by occupation and industry was available at the residence by tract, but this is of no value to the specific purpose of estimating nonresidential trip generation. What is desired is the number of employees by work place to as small an areal definition as possible. Such data are now difficult to bring together from other sources.

The trip estimating process relates travel to certain variables, including population, employment, automobile availability, and income. These variables must be forecast as the foundation for the travel estimates. They are also key elements to land-use forecasting. One of the best uses to which data from future censuses can be put is the evaluation of long-range forecasts on an incremental basis. At least once every ten years the incremental forecasts of population, employment, income and the journey-to-work can be compared to what is actually happening as reflected in the census data. Studies can be made of any discrepancies between the estimates and the actual values

and a means developed for adjusting the long-range forecasts. Models for estimating future travel and land development can be adjusted and reused during the census years to sharpen estimates and estimating procedures.

Also, in addition to the possible widespread application of census data in model work, there is a more limited use possible in metropolitan areas that are evaluating rather sizable expenditures for public transportation. Mass transit is a rather specialized service mainly encouraged by the congestion at peak hours. At the present time, a model approach has not been very successful in representing work travel, which is the greatest portion of peak-hour travel in most urban areas. Obtaining what is nearly equivalent to the first work trip coded to detailed residence and work place locations will provide the major portion of the traffic data for cost-benefit studies of public transportation facilities now being proposed in many areas to relieve congestion. However, the array of origins and destinations actually occurring may be too widely scattered to be attracted to public transportation, which can only serve a limited number of origin and destination points. Detailed work trip data could provide the precise information to analyze the worker transportation market and the capital expenditures that are justified to service this demand for transportation. A developing transportation and data communications technology is bound to have profound effects on this market, and future censuses could provide invaluable evidence of the ensuing changes.

LIMITATIONS ON USE OF DATA FOR TRANSPORTATION PLANNING

It is natural that certain problems will arise in the use of a data source intended for such universal use as the U.S. Census. The complexity of such a massive data handling operation is probably beyond the imagination of most.

The limitations found by the authors in their use of census data, although narrowed in scope to their particular analyses, may help others in their use of this data source. Also, consistent with the Bureau of the Census policy to improve each succeeding census, certain present limitations may be rectified in future censuses.

Definitive Documentation

The major source of documentation is at the beginning of each published tabulation. General data collection procedures are explained, and definitions are given for certain populations included in the tabulations. This documentation is helpful when working with the magnetic tapes used to produce these tabulations. Also, the tape layout is described by the Bureau of the Census Decennial Operations Division and Demographic Operations Division Technical Memoranda. However, there are data on these magnetic tapes that are not defined in either of these documents. Certain definitions have been obtained through recontact with the Bureau of the Census and research through census procedural manuals. There is no single source of documentation—such as a user's manual—for data on magnetic tape.

Tape Format

The data from the 1960 Census were available at the Bureau of the Census on UNIVAC tape. Other tape formats must be specially requested and conversion paid for by the user. For example, the authors obtained a conversion to magnetic tape for use on IBM equipment. The first conversion was done at the Bureau of the Census. These tapes then required additional handling and programming for conversion from the XS3 and binary languages to the BCD language for use on the IBM 1401. This process involved many transmissions with the Bureau of the Census. Tapes had to be replaced, because they would not read into the computer, or because they had "garbage" instead of valid records. The 1960 data cannot be obtained for any computer system in a "Go" status that would allow the user to make a minimum of summary checks before using.

Comparability with Earlier Censuses

A limited number of data items are carried consistently in published tabulations from earlier censuses. It was not until 1960 that data were available on computer tape.

Consequently, these earlier data must be transferred to a computer media from printed sources by the user, if he wishes to do any trend analysis work.

Record File

The procedure to identify a record on Census Tape Series 29I is rather complex. The file was normally found in sort: by state, by county, by minor civil division, by place, by tract prefix, by tract basic and by tract suffix. This identification requires 20 characters, some alpha and some numeric. Certain areas do not have officially designated census tracts and a pseudo-tract was created. Pseudo-tracts comprise wards in untraced cities of 25,000 or more, separate urban places of 2,500 or more, and the remainder of minor civil divisions or census county divisions in untraced areas. In some instances, data records have been found for places that have no defined boundaries. Also, there are 30 tracts in the area studied by the authors that are for crews of vessels and do not represent data for a physical portion of land in the study area. Those types of data are possible in any area with port facilities.

Geographic Identification

The record identification for each of the census tape series links the data to a particular geographic location through the use of a coding manual, general map, and in the case of the Census Tape Series 29I, a census tract map. This procedure allows only a very limited display of data on a map for a particular area, because of the laborious task of manually determining the geographic location for the display of data. Map coordinates are available for census tracts on the Universal Transverse Mercator (UTM) grid system, but now must be inserted in each record by the user. These coordinates are a key part of the Damage Assessment System for the Office of Civil Defense, Department of Defense, and are published in the National Location Code Manual.

Another ingredient that is important in this visual display of data is the area of the census tract involved. This information must be obtained from sources outside the Bureau of the Census.

Areal Detail

Population and housing data and auto availability are available by census tract. A limited amount of journey-to-work data, occupation, industry, mode, and 13 work destinations, are given at the place of residence of the worker by census tract. These 13 work destinations are gross areas, such as a whole county or major city. Only the total number of workers going to each of these work destinations is given.

Additional journey-to-work data are available on Census Tape Series 29BB at the place of employment and place of residence. The level of areal detail is major city, county, or remainder of county, except in the New England States where towns are used. These places of employment and residence are identified with a Universal Area Code on the Census Tape Series 29BB. They range in size, for the area studied by the authors, from one square mile to 922 square miles. The population of people living in these areas ranged from less than 1000 people to over 2.5 million. Individual records for some of the very small populations could be aggregated; however, there was no way to obtain a lower level of areal detail in the large areas. Significant variations in the choice of mode for work trips from such large areas as Queens County, N. Y., with a population of 1.8 million people, are lost in the data record, which is for all of Queens. Trip length analyses have obvious limitations when using such large summaries.

The table on page 69 gives a summary of the areal detail for the census tape data.

Areal Coverage

It was mentioned in the discussion of areal detail that 13 places of work were given for a particular census tract by place of residence on Census Tape Series 29I. The same 13 places of work were usually used for all the census tracts in a particular county. These places of work differed between counties and, in the area studied by the authors, there were 85 different employment areas. There were overlapping definitions

	Enumeration District	Census Block	Census Tract	Universal Area Code
100% Data	29A	34	29B	
25% Sample			291 ^a	29BB

^aThis tape contains some 100 percent data and place of work to 13 places of work about equal in size to UAC's.

for an employment area. For example, in New York State, Manhattan was carried as a separate employment destination, but in Northern New Jersey, New York City (the five boroughs) was one destination. In other words, an employment area may be uniquely defined in census tract records by place of residence for one county, aggregated with other employment areas for census tract records in another county, or not included as an employment area.

More universal and complete coverage of employment is possible from the Census Tape Series 29BB. The limitations of areal detail have been mentioned. In addition, the data were only processed to this gross level of detail for workers who either live and/or work in a Standard Metropolitan Statistical Area (SMSA) of 250,000 or more people. For example, the people who lived in Pike County, Pa. (which is not even in a SMSA) but worked in Manhattan, are contained in a record for the New York SMSA. The reverse commute would also be contained in a record for the New York SMSA. Those who commute between Pike County, Pa., and Somerset County, N.J., are not available in a record, because neither place is in a SMSA of over 250,000 people. In the area studied by the authors, sizable portions were missing employment data such as that for Somerset, Middlesex, and Monmouth counties, N.J., a 1092-sq mi area with a 1960 population of 874,000 people. These data on Census Tape Series 29BB were prepared by the Census Bureau for 101 SMSA's in the United States with 250,000 or more population in 1960.

General Data Limitations

The population of workers described is defined as anyone 14 years or older who worked at least once in the week prior to being interviewed in the 1960 census or was then a member of the Armed Forces. Distributions of these total workers do not include members of the Armed Forces by occupation or by industry. This population of workers is a cross between the average daily employment traditionally studied in a travel survey and total employment statistics compiled by certain state agencies. It does not include the location of second jobs; just the one place of work where the most hours were spent is recorded. This is roughly equivalent to the first work trip from a travel survey.

The mode data are for the primary mode. If more than one mode was used in getting to work, the mode involving the greatest travel distance as judged by the respondent is recorded.

The data are carried at the person's usual place of residence, even though he may have been working in another area at the time of the census. For example, a person who has an apartment in Manhattan for ease of commuting during the week, but actually lives in Boston or Florida, would be recorded as commuting from Boston or Florida.

The cross tabulation between socioeconomic characteristics of the worker and the mode used is limited to auto or carpool and public transportation modes. The workers using each of these two modes are distributed by: white or non-white, sex, age, by sex and occupation (three occupation categories), and by earnings.

There are a limited number of cross tabulations for workers irrespective of mode. Occupations are divided into 13 categories for each sex. Workers are divided into family heads and other relatives and then by income category for each. The primary means of transportation to work by eight mode categories is given for each sex.

Included in most of these cross tabulations and in straight distributions of the workers are data in an unreported category. The magnitude of these data have been found to range from 1 to 10 percent of the total population involved. In some cases, these may be distributed in proportion to the reported information by the user. Of course, this is impossible where the unreported information is in two or more sub-populations in a cross tabulation.

Usually summations of each distribution in a universe should be made to determine the correct base for ratio computations. For example, the total number of workers in a record on the Census Tape Series 29BB may be obtained from the summation of just three fields in the record. However, this total will not allow 100 percent coverage for the mode data, unless the unreported mode is carried as a mode category. On the Census Tape Series 29I, the total number of housing units derived by summing over the distribution by number of units in the structure (25% sample) does not always agree with published totals (100% data) or the totals derived by summing over the distribution by condition and plumbing (100% data).

The Census Tape Series 29I does not include a population distribution by age. Consequently, it is necessary to process Census Tape Series 29B to obtain these data.

SUGGESTED IMPROVEMENTS FOR 1970 U. S. CENSUS

Although the Bureau of the Census has collected data that have been of significant value to transportation studies in the past, the adoption of questions on automobile availability and the journey-to-work (collected for the first time in the 1960 census) has enhanced the value of the source and indicates the desire of the Bureau of the Census to provide data for special purposes, such as transportation planning. Professionals have had a chance to use the newly collected data and should be in a position to analyze its value. Additional improvements would further enhance the value of the data to users and make their analyses easier. The authors have found the data to be a valuable source. Certain limitations have been observed and suggestions formed, which may be an aid in developing criteria for improving the value of the data collected in the 1970 census.

The authors present these suggestions with full knowledge that other factors must be considered. The authors have no information as to the cost and logistical problems involved.

Suggestions for making the next census more readily usable are based on the use made of the census data by the authors and the limitations presented in the previous section.

1. Employment data, which are used as a basic variable by many transportation studies, are difficult to obtain since coverage in various sources is usually not complete. The census journey-to-work question in the 1960 census obtains a large portion of an area's employment. However, the data obtained contain information on only a single work trip for each employee, if made at least once during the week prior to the census. Missed are second jobs and workers who are ill or on vacation. It may be possible that the journey-to-work question in the 1970 census be framed similar to the following:

If you are employed: (a) Where did you work yesterday, and what mode of travel did you use? Both of these would be for the primary job. (b) If you hold more than one job, where are the other jobs located? (c) If you did not work yesterday, where is your regular place of employment?

2. It is further suggested that the employment places obtained from the preceding questions be coded to some smaller geographic area than Universal Area Code zones. Since the population data are coded to the census tract as a major aggregation level, perhaps it would be possible to use tracts or combinations of them for coding employment data.

3. The mode of travel currently includes the category, "auto or carpool." To be consistent with the usual modes collected in home interview travel surveys, it is suggested that "auto driver" and "auto passenger" be considered as separate modes.

4. At present the census publications and tape files are limited to journey-to-work information for Standard Metropolitan Statistical Areas of 250,000 and over population. In the Tri-State Region, for example, this covers only 90 percent of the total workers journeying to work. It is proposed that complete coverage be provided at least for those areas that are included in the study areas of the urban transportation studies, established in conformance with the requirements of the 1962 Highway Act.

5. Automobile availability is provided at the census tract level in the 1960 census. However, the rate of sampling was variable, from 5 to 20 percent, although the journey-to-work data were collected on a uniform 25 percent sample basis. Since auto availability is of considerable usefulness, it is suggested that it be collected on a uniform 25 percent sample basis.

6. Cross tabulations are not available to any great degree on the 1960 census tapes. Even though a minimum of cross tabulations results in a great increase in the size of the data records, consideration should be given to increasing the number of cross tabulations. The form of the cross tabulations will not be described here, since the possibilities are so great that a consensus from users must provide the combinations desired.

7. It is recognized that only from the past census has the Bureau of the Census been in a position to supply data in magnetic tape form to other agencies. Although layouts have been provided for each file, the uninitiated have experienced much difficulty in determining exactly what is available in each tape file, the definitions of various terms, and the coverage provided. Since much use may be made of tape files by agencies other than the Bureau of the Census, it is hoped that considerable additional effort will be expended in the 1970 census to prepare detailed descriptions of the tape files, including data coverage and definitions used. A user's manual for the tape files would be very helpful.

8. It is suggested that magnetic tapes be processed to be handled on all manufacturers' computers to eliminate the time-consuming process of conversion from the census tape to other tape forms at the time of request.

9. The Bureau of the Census might consider establishing a service unit within the organization to provide users with assistance in both the use and possible correcting of discrepancies found on the tapes. Such help was readily supplied to the Tri-State Transportation Committee by the Bureau of the Census; however, it was felt that this help was provided by people who were pulled away from their usual responsibilities.

10. Control totals for the fields contained on the census tapes should be provided along with the tapes supplied to allow the user to insure that they have been processed correctly in his subsequent uses. Machine-read errors will be more readily apparent.

11. It would make it easier for the user if the exceptions to obtaining control totals were eliminated. The data should be adjusted for, not reported. Where 100 percent sample totals are available, a distribution determined from a 25 percent sample should be adjusted to this control total.

12. Consideration should be given to establishing, as soon as possible, officially designated census tract boundaries in all transportation study areas.

13. The establishment of geographic identification for each tract in the form of coordinates in the data records should be considered.

14. One computation of the gross areas of census tracts and placement in the respective records should be considered.

15. It is suggested that two data files be prepared for use in transportation planning. One file should contain pertinent population and housing data; the other should contain the journey-to-work data.

16. Finally, for the convenience of the user, a distribution of the total population by age groupings should be included in the population and housing data file.

The aforementioned suggestions, which have been formulated by the authors' use of the census material, do not necessarily represent those of the Tri-State Transportation Committee. Others in the organization are using the data and may also form suggestions. The implementation of improvements to the data in the census for transportation planning uses is an evolutionary process, which must be based on past uses and

evaluations, as well as foreseeable uses. In no way should the limitations and suggestions discussed be considered as a criticism of the Bureau of the Census. It is hoped that this discussion will help users of census material to form suggestions for the 1970 census and interest others in becoming census data users.

ACKNOWLEDGMENTS

The authors wish to acknowledge the following staff members of the Tri-State Transportation Committee who contributed material for the paper: Harold Deutschman, Transportation Engineer I; David Ornstein, Chief, Regional Plan Analysis Section; Paul Rackow, Sr. Statistician III; Richard Warchol, Travel Analyst; and Herbert Mills, Systems Analyst II, who prepared programs for conversion and summary of the census tapes obtained from the Bureau of the Census.

Special thanks go to Edward F. Sullivan, Chief, Travel Analysis Section, Tri-State Transportation Committee, under whose supervision this work was carried out.

The preparation of this report was financed in part through federal funds made available by the U. S. Bureau of Public Roads; the federal grant from the Urban Renewal Administration of the U. S. Department of Housing and Urban Development, under the planning assistance program authorized by Section 701 of the Housing Act of 1954, as amended; and in cooperation with the three states of Connecticut, New Jersey and New York.