POSTAL ZIP CODE AREA STATISTICS FROM INTERNAL REVENUE RECORDS

Robert A. Wilson, Internal Revenue Service

This paper deals primarily with an analysis of three-digit postal ZIP Code area data for 1979. These data were recently published in Statistics of Income (SOI) by the U.S. Internal Revenue Service (IRS) based on individual income tax return information. Much research has already been done on differences between tax return and Census data so that some of what follows, especially about State and national trends, is simply a repackaging of what is already known. However, this may be the first time that an analysis has been reported upon for ZIP Code areas [1].

The first section of the paper describes the use of the tax return as an indicator of population and economic shifts by State. Next, the significance of the postal ZIP Code area as a geographic reporting unit is discussed. The third section compares Census and IRS data using the State of New Jersey and its 20 three-digit ZIP Code areas as examples. The concluding comments summarize possible directions for future use of postal ZIP Code statistics.

TAX RETURNS AS GEOGRAPHIC INDICATORS

The individual income tax return makes sense as a leading indicator of population shifts, especially at the State or regional level. This is because there is a strong connection between the number of return filings and the number of individuals in the labor force, so that geographic changes in returns filed show the direction of migration of the working population. Figure A shows the increase in returns filed for Income Years 1969 and 1979 [2], years that roughly coincide with the Census. The trends shown, north to south and east to west, are clearly evidenced. The Census population data, also included in Figure A, show the same trends, but not the dramatic shifts shown by the tax returns.

In this context, the tax return filing patterns may anticipate what the next Census will show [3]. To the extent that most migration reflects the movement of younger workers to areas of real or perceived economic opportunity, then the next Census can be expected to reflect the results of this movement, namely sizable population increases as these new workers settle down to raise families [4, 5].

Economic activity has a varying effect on changes in average income by State which are more difficult to analyze. Thus, a similar chart showing income changes presents a mixed and more complicated pattern.

At first glance one might assume that any geographic patterns that emerge from a time series analysis would be difficult to interpret when based on tax return data--one reason being the changes in filing requirements that occur over the years. For 1969, for example, the basic filing requirement was $600 or more in "gross income." However, by 1979, a new, more varied, filing requirement set forth gross income cutoffs that ranged from $1,000 to $7,400, depending on whether or not taxpayers were age 65 years or over; their marital status; and, if married, whether a joint return was to be filed [6]. Therefore, all other things being equal, fewer individuals at the lower end of the income scale might have been expected to file for 1979 than for 1969.

However, a number of factors together have served to maintain comparability over the years. Obvious factors are increases in real income and the effects of inflation, both of which served to increase income enough for many individuals to meet the revised filing requirement. In addition though, voluntary filing below the income cutoffs and the filing of returns solely to obtain a refund of taxes withheld on wages or of the amount based on the earned income credit which was claimed by qualifying low-income workers, have all contributed to maintain stability in the filing population and to facilitate time series comparisons as a result.

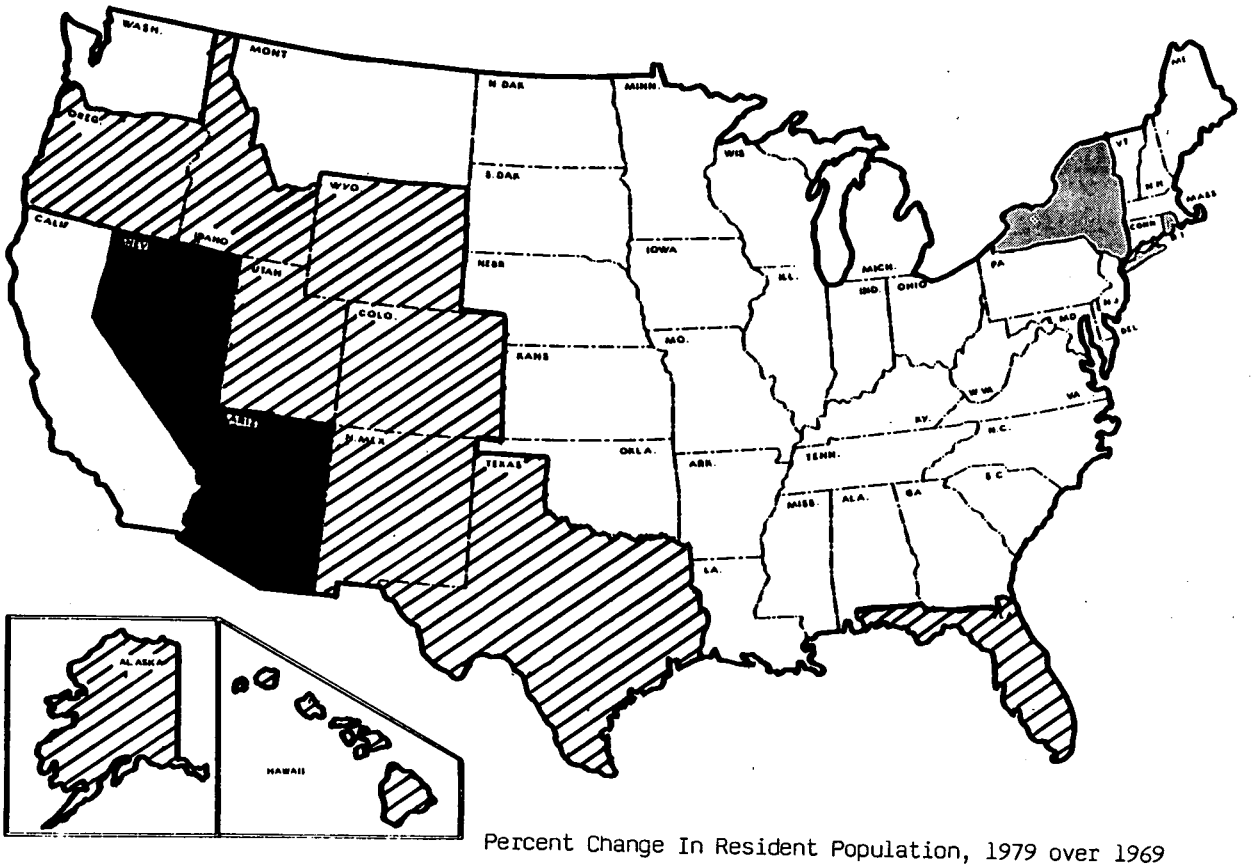
SIGNIFICANCE OF THE ZIP CODE AREA

Within States, the ZIP Code area is a relatively new geographic reporting unit overlapping political boundaries, such as those of counties, that normally serve as the basis for most local area statistics. As of 1979, there were 837 three-digit ZIP Code areas. Each represented a postal "sectional center" or "zoned city". Sectional centers serve as "hub" cities (which includes surrounding smaller town and rural areas as well) in which mail is sorted for distribution to peripheral postal outlets which are, in turn, identified by the fourth and fifth digits of the ZIP Code. While zoned city boundaries generally coincide with those of the larger cities, they may also include close-in suburbs as well.

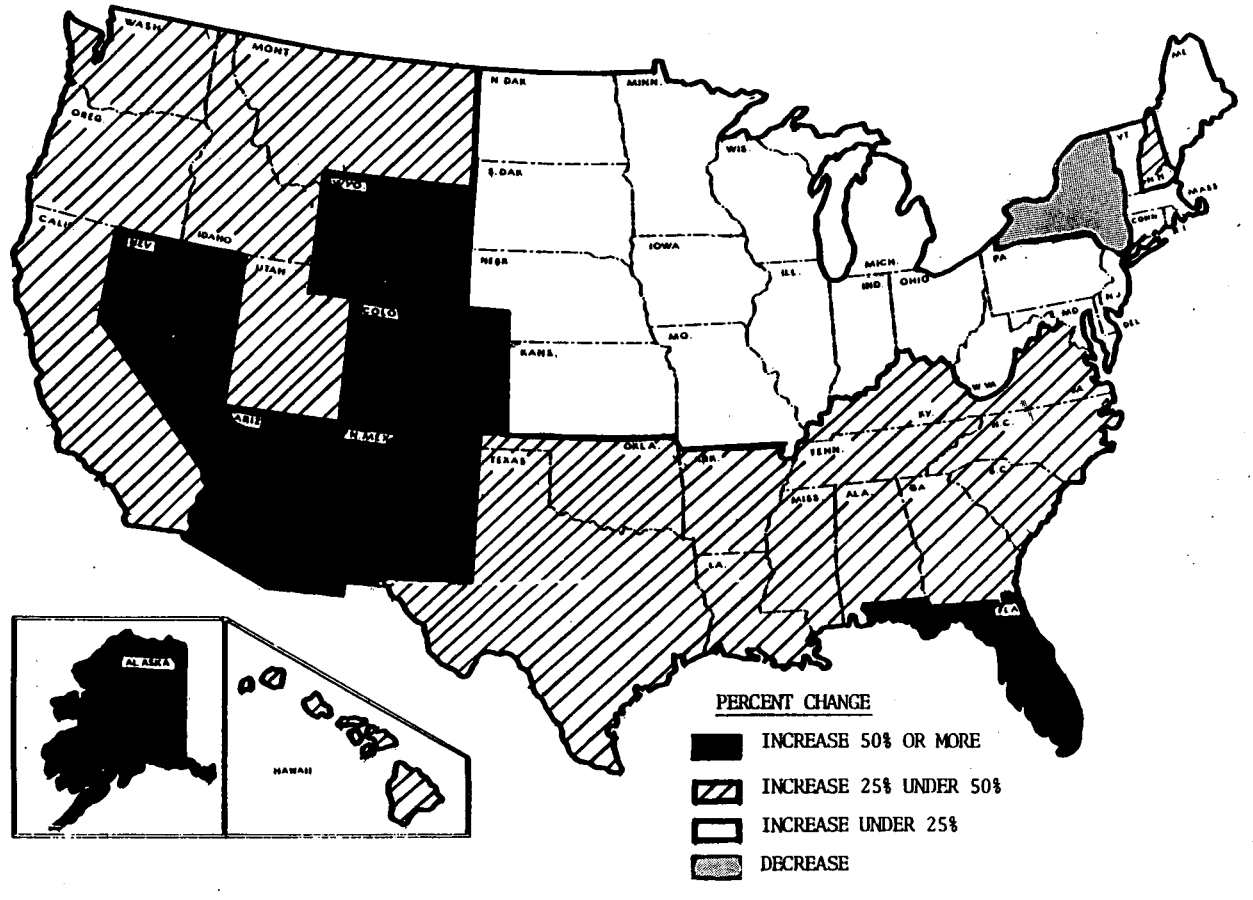
Although its primary purpose is to meet the specific needs of the U.S. Postal Service, the ZIP Code area is not without economic significance. The three-digit area, the next largest unit below the State, is set up to facilitate the movement of mail and this means that it tends to reflect transportation patterns and thus, patterns of commerce. Therefore, as economic units, three-digit areas often may be equated with marketing or commuting areas. In a sense, they may be thought of as an alternative to the county as a reporting unit for economic analysis [7].

ZIP Code areas obviously have limitations when used for this purpose. In contrast to counties, only limited data are currently available for them [8]. Moreover, because their boundaries can change, what historical data there are may be difficult to use [9]. Also, ZIP Code boundaries, particularly at the three-digit level, while significant for measuring some trends, can obscure others. In a ranking of the highest income areas based on adjusted gross income for 1979, prosperous Maryland suburbs of Washington, DC, for instance, do not even appear because, as of

Figure A.--Percent Change in Number of Returns, 1979 over 1969



Percent Change In Resident Population, 1979 over 1969



that year, they were included in the same ZIP Code areas as Washington itself, where average income was much less. (These boundaries have since changed so that this example would no longer apply).

A major area of concern in any statistical use of the ZIP Code is the accuracy of the reporting. Probably at the three-digit level, many of the discrepancies that exist may be masked. At least in the case of the income tax return though, IRS conducts an analysis of the ZIP Code at the time it validates the taxpayer address during the normal course of processing for tax administration purposes. IRS has incentives to improve on the quality of the reporting because of its need to communicate with taxpayers on a timely basis and, when taxpayers file with IRS, they are urged to use the preaddressed mailing label. IRS also has an additional incentive to assure the accuracy of the reporting: postal delivery rates are less when mail to taxpayers is presorted by ZIP Code. In any case, as a byproduct, this "perfecting" of ZIP Codes definitely results in better statistics.

Related to the question of accuracy is the significance of the taxpayer address taken as a whole, especially as it applies to its use for statistics. Some taxpayers report as their address a business address, the address of a lawyer or accountant who often times is a tax advisor or preparer, or a post office lockbox in a community (and ZIP Code area) other than the one in which they reside. In the context of these limitations, the three-digit ZIP Code area that witnessed the highest rate of increase in returns filed between 1969 and 1979 was not really an "area" at all. Rather, it represented U.S. Government facilities in Washington, DC, for which there were special ZIP Codes. One can only conclude from this seeming anomaly that, for whatever reason, many more Federal employees were reporting their business address on their income tax returns for 1979 than for 1969.

The effect of such variations in reporting addresses on ZIP Code area statistics is not known, although it is likely to be more prevalent in statistics for large urban areas than for other areas of the country. Richard Irwin of the U.S. Census Bureau estimates that from 10 to 15 percent of all individual income tax returns show a mailing address that is different from the residential address, including more than 6 million returns filed from lockbox addresses [10].

All tax return ZIP Code data are, of course, based on the tax-defined concept of income, i.e., adjusted gross income (AGI), and this means that certain kinds of income are excluded either wholly or in part. The effect of these exclusions may be compounded by definitional differences based on tax law changes.

For the majority of taxpayers, there may well be little or no difference between total economic income and AGI, but the two tend to pull apart as income increases. Thus, toward the upper end of the income scale, nontaxable interest on State and local Government obligations and capital gains become factors. In addition, certain amounts are deductible in arriving at AGI, thereby also understating economic income. Examples are certain expenses of employees, alimony paid, a portion of dividend income, and contributions to self-employed or individual retirement plans or arrangements. Moreover, for tax purposes, certain deductions have special meanings that differ from their normal accounting counterparts, with a varying effect on economic income. Depreciation is just one example. Economic income at the lower end of the income scale is also understated, by the exclusion of social security, cash public assistance and in-kind transfer payments of various sorts [11].

In a similar vein, the tax return itself is a unique reporting unit. Just as the ZIP Code area is designed to meet postal needs, the IRS reporting unit is designed for tax administration purposes. It is therefore synonymous neither with the "individual" nor with the "family" and is actually a hybrid of the two [12].

NEW JERSEY COMPARISONS

A comparison of IRS and Census data for New Jersey reveals some of the strengths and weaknesses of the tax return as the starting point for a future census based on administrative records [13]. Relevant counts from the two sources are shown below:

Table 1.--New Jersey: Census vs. IRS

Data Source		Number (thousands)		Percent
		1979	1969	Increase
Census				
	Population	7,338*	7,170	2.3
	Households.....	2,549	2,218	14.9
IRS				
	Total returns .	3,153	2,820	11.8
	Exemptions**..	7,239	7,194	0.6

*The official population for New Jersey was 7,365,000. The number shown represents the number included on the Census ZIP Code area public use file.

** Other than age or blindness.

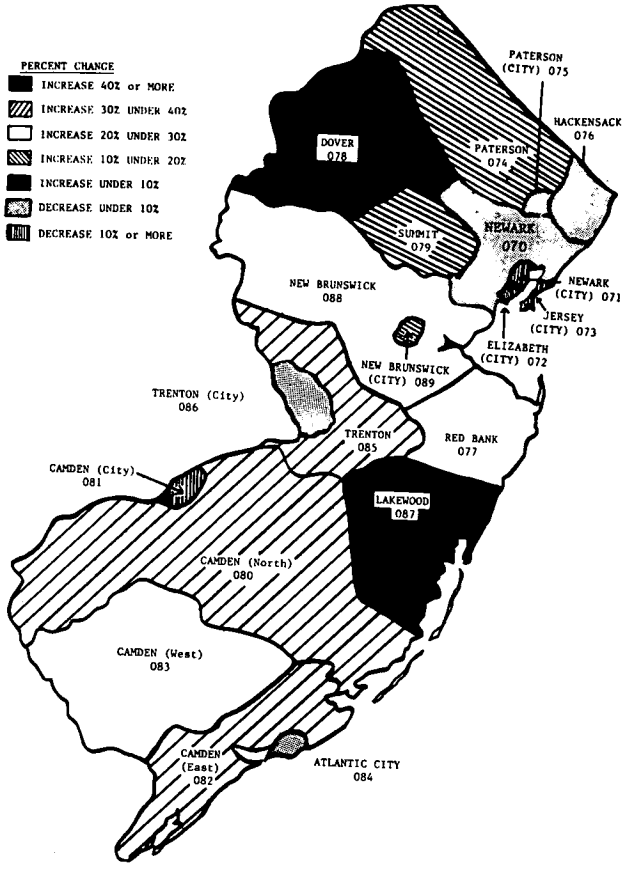
Personal exemptions other than age and blindness are the key to a reconciliation of Census and IRS counts. In general, exemptions are allowed for each taxpayer (on joint returns, husband and wife are each regarded as a taxpayer) and for "qualified" dependents. There is some double counting in exemptions claimed chiefly because some taxpayers who had to file their own return could also be counted as dependents on the return of another.

A possible warning note on the return and exemption counts: tax returns for New Jersey may be especially affected by the taxpayers who file their returns using addresses outside the State. This situation may be especially applicable to New Jersey because many of its residents commute to jobs either in Philadelphia or New York City.

Figure B shows the increase in returns filed by ZIP Code area between 1969 and 1979 [14]. The increase in the northern part of the State reflects the growth in the areas that are, or at least include, suburbs of New York City,

especially the new "outer" suburbs, plus the fallout from some of the older cities within the State, such as from Newark and Jersey City.

Figure B.--New Jersey: Percent Change In Number Of Returns, By 3-Digit ZIP Code Area, 1979 and 1969



The growth in return filings for the southern part of the State also reflects the move to suburban ZIP Code areas, with the areas that include the outer suburbs again leading the way. Cities losing population to the suburbs in addition to Philadelphia include Trenton and Camden, both older cities which are located within New Jersey. The drop in population for Trenton is not evidenced by the tax return data principally because its three-digit ZIP Code includes wealthier inner suburbs. (At the five-digit level, however, it would be clearly visible).

The growth of the suburbs in South Jersey is complicated by the companion growth of the coastal areas whose increase reflects the increasing popularity of using resort areas for permanent residences. Another factor contributing to growth is that the shore areas have become especially attractive as homes for retired persons. More definitive comparisons with Census data would undoubtedly shed more light on these trends.

When a comparison was made between AGI and Census household median income by ZIP Code, it showed that the trends in both data sets are highly correlated; in fact, the rank level correlation is .941. Using either income concept it is likely that the movement of younger, lower income, individuals into suburban areas vacated by often older, higher income, elements of the population, applies within the State, as the wealthier in turn move to "newer" suburban areas more distant from the cities. At the same time, other lower income individuals are moving to the cities to replace those who have moved to the "older" suburbs.

FUTURE DIRECTIONS

As has been established by others, data on tax returns bear a definite relationship to population size. Most reconciliations have been confined to the U.S. and State levels and, when attempted at the local area level, have been in terms of administrative subdivisions, such as the county. A conclusion drawn from the research for this paper is that when reconciliations are attempted at the local area level, by three-digit ZIP Code in this case, data peculiarities are more complex and difficult to resolve inasmuch as they bring to light problems that are not evident at more global levels of aggregation. Research needs to continue in order to resolve these discrepancies before we can begin to say we are ready for a population census based on administrative records.

On their own, ZIP Code area data can become increasingly useful as tools for economic analysis. On the one hand, users will need to become more conscious of their use. In this regard, it is unfortunate that adequate maps showing ZIP Code boundaries are so scarce [15]. Certainly market researchers are aware of the value of ZIP Code statistics, although understandably their preference is for a finer level of detail than was used for this paper, namely at the five-digit level (or at the nine-digit level if these extra digits are added to the postal system). The needs of marketers for detail raises questions about confidentiality and about the propriety of Government agencies developing data unfairly labelled (at least in the past) as being primarily for use by "junk mailers", a charge that was originally made when IRS produced its first five-digit ZIP Code tabulations, for 1966.

On the other hand, Government agencies, too, must consider whether utilizing the ZIP Code for more than an aid to assist in classifying data for the older geographic series is a worthwhile endeavor. At a time when Government resources for statistics are at a premium and are likely to continue to be, the development of new statistical series based on the postal ZIP Code may require private financial assistance, or at least a new kind of partnership between the public and private sectors, possibly along the lines recently advocated by James Bonnen [16].

ACKNOWLEDGEMENTS

The author wishes to extend his special thanks to Edward Spar of Market Statistics, a division of Bill Communications, for providing him with a copy of ZIP Code Demographics for the State of New Jersey. This publication, based on public use files of the U.S. Census

Bureau, enabled the comparisions to be made of tax return and Census data by ZIP Code area. Appreciation is also extended to Douglas Sater and John Kavaliunas of the Census Bureau for the additional data they provided, and to Richard Irwin, also of the Census Bureau, for information on the quality of the tax return addresses that serve as the basis for IRS geographic distributions. Finally, thanks must be given to Denise Herbert, Michelle Cobb and Roselind Vinson for their indispensible typing assistance.

NOTES AND REFERENCES

[1] For the underlying Internal Revenue Service data for 1979 that were used in this paper, see Wilson, Robert A. and Oh, H. Lock, "Individual Income by ZIP Code Area, 1969 and 1979", U.S. Department of the Treasury, Internal Revenue Service, Statistics of Income Bulletin, Vol. 2, No. 4, Spring 1983, U.S. Government Printing Office.

[2] See U.S. Department of the Treasury, Internal Revenue Service, Statistics of Income--1969, Individual Income Tax Returns, Publication 79, U.S. Government Printing Office, 1971, and Statistics of Income--1979, Individual Income Tax Returns, Publication 79, U.S. Government Printing Office, 1982.

[3] For a brief review of geographic patterns of filing individual income tax returns, see Blacksin, Jack and Plowden, Ray, "Statistics of Income for Individuals: A Historical Perspective", 1981 Proceedings, American Statistical Association, Section on Statistical Uses of Administrative Records.

[4] IRS State data for California, New York and Washington include overseas military personnel because army and fleet post office addresses are located in these States. Data for Maryland also include overseas military and governmental personnel to an unknown extent. Population and migration estimates by county based on individual income tax returns are available from the Internal Revenue Service. The migration data show the number of taxpayers whose county of residence changed between 1980 and 1982. AGI of migrants and nonmigrants is also available by county. For additional information about purchasing the results of this study, write to the Director, Statistics of Income Division, Internal Revenue Service, 1111 Constitution Avenue, N.W, Washington, DC 20224.

[5] For more on migration patterns, see Garnick, Daniel H., "Shifting Patterns in the Growth of Metropolitan and Nonmetropolitan Areas", U.S. Department of Commerce, Bureau of Economic Analysis, Survey of Current Business, May 1983, Vol. 63, No. 5, U.S. Government Printing Office.

[6] See Internal Revenue Code, section 6012, for relevant years for a complete description of the individual income tax return filing requirements.

[7] For years prior to 1979, Internal Revenue Service data were published in the Supplemental Report, Statistics of Income--1969, ZIP Code Area Data from Individual Income Tax Returns, Publication 649, U.S. Government Printing Office, 1972, and in the Supplemental Report, Statistics of Income--1966, ZIP Code Area Data from Individual Income Tax Returns, Publication 649, U.S. Government Printing Office, 1972. These reports present income and tax data by three-digit area; data for five-digit areas were released through the National Technical Information Service.

[8] Ibid.

[9] Between 1969 and 1979, there were 56 three-digit ZIP Code areas that were increased in size, 54 that were decreased in size and 38 areas that were otherwise redefined. The total number of three-digit areas remained at 837.

[10] Irwin, Richard; Knott, Joseph J.; and Thompson, John, "Feasibility of An Administrative Records Census" (unpublished report), U.S. Census Bureau, 1983.

[11] Income in kind, such a food produced for home comsumption is another factor that might be included as "economic" in contrast to "taxable" income. Such income is likely to have a more noticeable effect at the lower end of the income scale.

[12] In this context, see Irwin, Richard and Herriot, Roger, "An Initial Look at Preparing Local Estimates of Household Size from Income Tax Returns", 1982 Proceedings, American Statistical Association, Section on Applications of Administrative Records Data.

[13] See Alvey, Wendy, and Scheuren, Fritz, "Background for an Administrative Record Census", 1982 Proceedings, American Statistical Association, Social Statistics Section.

[14] ZIP Code boundaries shown for New Jersey in Figure B were approximated in some instances based on municipality names and the five-digit ZIP Codes associated with them. See also footnotes 1 and 14.

[15] Maps showing most of the three-digit ZIP Code areas by State may be found in the Statistics of Income supplemental reports described in footnote 7. These maps, in turn, were obtained from the National ZIP Code and Post Office Directory, for 1969 and 1979, U.S. Postal Service. Maps of five-digit ZIP Code are included annually in the various Bell System Yellow Pages directories.

[16] See Bonnen, James T., "Official Statistics in Troubled Times: The Changing Environment of Producers and Users", Michigan State University, Agricultural Economics Staff Paper 83-33, prepared for the 44th Biennial Session of the International Statistical Institute, Madrid, Spain, Sept. 1983.