

NEIGHBORHOOD INFORMATION SYSTEM
USER'S MANUAL AND DATA DESCRIPTION
FINAL DRAFT

Donald Musa
Gambhir Bhatta
Steven D. Manners

University Center For Social and Urban Research
University of Pittsburgh

October, 1989

TABLE OF CONTENTS

1. Introduction	1
2. Operation of the Program	2
2.1 Overview	2
2.2 Use of Menus and Keys	5
2.3 RETRIEVE DATA - Selecting Numeric Variables	6
2.4 GEOGRAPHIC SELECTION - Selecting and Modifying Geography	8
2.5 MODIFY SELECTION SET - Modifying Numeric Variables	16
2.6 OUTPUT DATA - Obtaining Output	21
2.7 RETRIEVE TEXT	24
2.8 RUN REPORTS	25
2.9 Help System	26
2.10 QUIT/EXIT TO DBASE - Exiting from the Program	27
3. Description of the Data	28
3.1 Overview	28
3.2 Census Data	31
3.3 Police Crime Statistics	35
3.4 Building Construction Permits	36
APPENDIX A. Listing of Census Data Tables and Variables	A1
APPENDIX B. Listing of Crime Data Tables and Variables	B1
APPENDIX C. Listing of Building Permit Data Tables and Variables	C1
APPENDIX D. Pittsburgh Census Tracts and Neighborhood Definitions	D1
APPENDIX E. Allegheny County Census Tracts and Municipality Definitions	E1

1. INTRODUCTION

The Neighborhood Information System (NIS) is a computerized data retrieval and processing system that allows users to gain access to various sets of data and operate on them to generate hybrid data sets for analysis and policy-making purposes. The system makes it possible to assemble sets of data containing information that community planners utilize to analyze neighborhood change for small geographic areas based on census tracts. The NIS was developed at the University Center for Social and Urban Research (UCSUR) at the University of Pittsburgh for the City of Pittsburgh Department of City Planning (DCP) and includes data for both the City of Pittsburgh and Allegheny County, Pennsylvania. Specifically, the data currently available in the system includes census tract data for the City of Pittsburgh and Allegheny County for each of the censuses from 1930 through 1980, Pittsburgh Police crime data for 1977 through 1987, and building construction permit data for the City of Pittsburgh for 1981 through 1989 (January-May).

The system has the potential to efficiently manage a great deal more data including both additional numeric information and textual information such as neighborhood descriptions, lists of neighborhood leaders, bibliographies, and any other text which might require quick and easy retrieval. Additional tract level numeric data types may be included in the system relatively easily as they are made available.

In its current form, the NIS is not intended to provide a complete and finished data retrieval system, but rather to make easier and more efficient the process of retrieval and combination of data from different sources at the tract or neighborhood level. The emphasis in the development of the system thus far has been on the procurement and preparation of the data rather than in providing a totally polished program. Therefore, the program is viewed as a prototype needing further development and has some limitations, missing elements and "rough spots" which can be improved with additional effort.

This manual is intended as a reference guide for users of the NIS system. Part 2 describes the operation of the program on a function by function basis, and provides a manual for use of the system. Part 3 describes the data contained in the system in greater detail. A complete list of data files and the variables contained in each file as well as the census tract geography for each census year will be found in the Appendix. Additional information concerning use of the NIS or regarding problems encountered should be reported to Donald Musa or Steven D. Manners at (412) 624-5442.

2. OPERATION OF THE PROGRAM

2.1 Overview

The NIS is a user friendly, menu driven data retrieval and processing system which accesses a set of data files containing a variety of tract level data. The NIS data is contained in dBase IV database files and may be accessed directly through dBase IV as well as through the NIS system.¹ The NIS program is a stand-alone, executable program written in the dBase programming language and compiled using the Clipper compiler from Nantucket Corporation. The program does not require dBase for operation, and can be run directly on any microcomputer that supports PC/MS-DOS version 2.0 or greater, preferably on an AT class computer.

The program resides in a directory called \NIS on the computer and must be run from there (this can be modified if it presents a problem). The data resides in subdirectories of the \NIS directory, called \NIS\CENSUS, \NIS\CRIME and \NIS\CONSTRUC. Data files which are used in the system include database files (extension .DBF) containing both data and descriptive information utilized in the system, and database index files (extension .NTX)² which speed up access to the data. The data files are described in more detail in Part 3 below and in the Appendix. The program is started by typing NIS followed by the ENTER key when in the \NIS directory. This executes a batch file called NIS.BAT which in turn calls the program (called NISV1.EXE).

The NIS program currently processes only numeric information of three types: Pittsburgh and Allegheny County census data, Pittsburgh crime statistics, and Pittsburgh building construction permits. While the main menu of the program shows selections for

¹ Most of the data files can be accessed by dBase III as well; however, some have more fields than dBase III can handle (over 128 fields) and are not recognized as dBase files by dBase III.

² The index files are not dBase compatible index files (extension .NDX) but are Clipper index files (extension .NTX), so they cannot be created using dBase. Clipper index files were chosen because they are both faster and smaller than dBase index files. Since the data itself doesn't change, the index files don't change and therefore don't need to be modified. However, three utility programs are supplied with the NIS system to recreate the index files should this be necessary for any reason: INDXC.EXE to create indices for the census database files, INDXR.EXE to create indices for the crime database files and INDXB.EXE to create indices for the building permit database files. These programs are located in the data subdirectories and each should be invoked (with the appropriate subdirectory as default) by typing in its name followed by the ENTER key (e.g., "INDXC" followed by ENTER).

textual types of information processing, the capability for dealing with this type of information has not yet been included in the program. Hopefully, applications will arise which will allow the further development of the system to include entry, editing and retrieval of textual information. Until then, however, the program suggests the capabilities which are possible. The following functional description of the program deals only with choosing, manipulating and retrieving numeric information.

The heart of the NIS program is its ability to easily construct a data set for examination or output from the selection of data items contained in the different NIS data sources. This is accomplished by the specification of what is called a variable "selection set" in the program. The selection set consists of a set of data items identified by variable name which is maintained as an internal data set within the NIS. This set is selected a variable at a time by the user (the selection set starts out empty) through the process of choosing individual data items from their source tables. When any operation is carried out which concerns variables (such as computing a new variable from existing ones or outputting data) the selection set defines the variables potentially operated on.

The variables chosen for the selection set may be displayed or printed at any time by selecting DISPLAY SELECTION SET under the main menu choice MODIFY SELECTION SET. The selection set is only available while the program is running - when the program is terminated, the selection set is lost. However, the user may elect to save the selection set by choosing SAVE SELECTION SET under MODIFY SELECTION SET on the main menu and supplying a file name. The saved selection set may then be recalled later or during a different program session.

The operation of the program depends on the separate selection and modification of variables via the choice of a selection set as outlined above, and the specification of a census tract geography to determine the units for reporting the data. The data is stored in the system for each census year at the census tract level, and it may be recalled at this level for either the City of Pittsburgh tracts or for the tracts in the remainder of Allegheny County. However, the program also allows the user to select subsets of the entire set of tract data (e.g., to examine data for the tracts in the Central Business District alone) or to group tracts together in any aggregation (e.g., to output data at the neighborhood level with neighborhoods defined as aggregations of entire tracts). A predefined set of neighborhood definitions for the City and a predefined set of municipality definitions for the remainder of the County are included in the system. As with the variable selection set, the census tract geography which has been selected by the user may be saved to a file and recalled during a later program session.

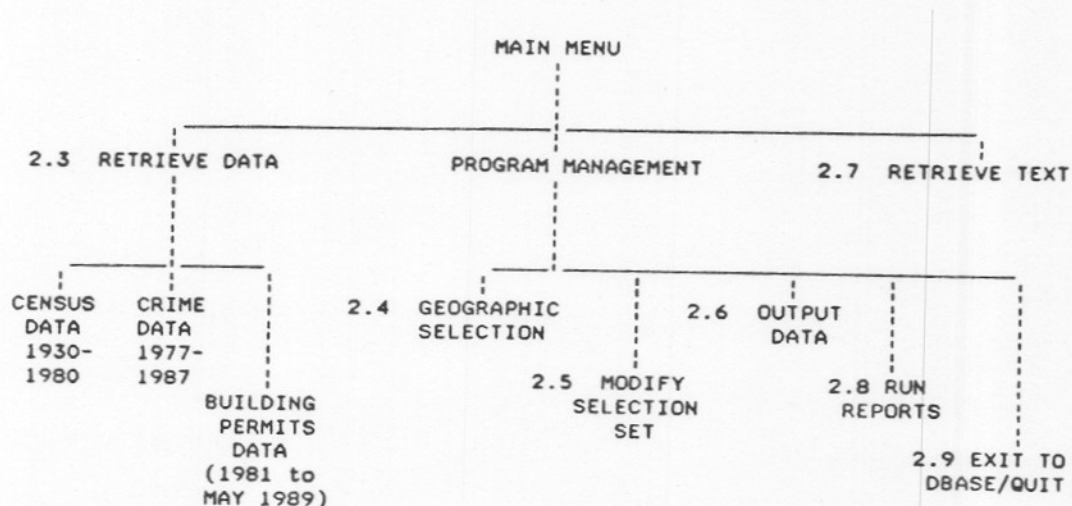
After variables and geography have been specified, the

selected data can then be output in various forms. The basic output is a data matrix made up of variables as columns and geographic units (tracts, neighborhoods, etc.) as rows. This data matrix may be output to the screen, the printer or a DOS file for further processing in other programs, either as a standard matrix or as a transposed matrix (variables as rows and geographic units as columns). In addition to data matrix output, a dBase database file may be created containing the entire data matrix (including all computed variables) and having the specified geography. This file may be used in the dBase program or transferred to other programs such as SPSS/PC+ for statistical analysis. DBase data files are widely used, and many other programs convert dBase files directly for their own use, making this a valuable feature for data transfer between programs. The last type of output supplied by the program is a frequency listing of a single variable in the selection set. This is most useful for variables having relatively few values, perhaps as a result of a computation procedure.

The program description which follows is structured by the various menu selections which can be made in the program. Figure 1 presents a schematic diagram of the main menu selections, which represent first level program functions, with the section of the manual describing them. This figure will be helpful in understanding the structure of both the program and the manual.

FIGURE 1

SCHEMATIC DIAGRAM OF THE MAIN MENU SELECTIONS
(FIRST LEVEL PROGRAM FUNCTIONS)



2.2 Use of Menus and Keys

The program is menu oriented with a limited number of choices at any given point rather than command line oriented with an unlimited number of choices. This means that the user does not have to know ahead of time what he or she wants to do, but may instead browse through the menus for the selections to make. In general, selection occurs on all menus by moving the highlight bar to the desired entry (with the arrow keys or other movement keys) and pressing ENTER or the SPACE bar to make a selection. On the main menu only, a selection may be made by entering the first letter of the selection. The selection will be immediately executed. ESC and F10 terminate the current menu and backup one menu or to the main menu, respectively.

Since the program is menu driven, a limited number of key strokes are allowed throughout the program and most of these keys have the same function everywhere. The most important of these keys are shown in reverse video on the bottom line of the screen during most operations. They include:

ENTER key or SPACE key - Selects the current choice on the menu.

ESC key - Terminates the current menu and moves back to the previous menu; does not make a selection.

F1 key - Supplies help. The help system is not operative in the first version of the program.

F9 key - Prints the variable selection set when it is displayed on the screen.

F10 key - Terminates the current menu and returns to the main menu; does not make a selection.

NUMBER PAD keys - Control movement on the current menu:

ARROW keys move one selection up, down, right or left.

HOME key moves to the first selection.

END key moves to the last selection.

PGUP key moves one page up.

PGDN key moves one page down.

2.3 RETRIEVE DATA - Selecting Numeric Variables

The selection of numeric data items (or variables) begins on the main menu where the data source (or type) is selected. Currently three data types are available, census data, police crime statistics and building construction permit data. The selection procedure is the same for all three. Selection occurs here and in other menus by moving the highlight bar to the desired entry and pressing ENTER or the SPACE bar to make a selection. When one of the three data types has been selected, another menu is presented on the screen which asks for the selection of a census tract year. This is a census year from 1930 to 1980 specifying the census tract definitions which will be utilized for the selection set - the tract definitions used by the Census bureau for that year.

The selection of a census tract year is important. Since census tract definitions are not compatible from census to census, data stored for tracts defined by one census are incompatible with data stored for tracts from a different census. The current version of the program will allow only one census tract definition to be used at a time. Only data which is stored for that set of census tracts will be available until the census tract year is changed. The census tract year may be changed in the GEOGRAPHIC SELECTION module (Section 2.4.1 below).

Thus, time series data sets at the census tract level cannot be constructed at all, while time series at the neighborhood level (or at any aggregate level based on tracts) from multiple censuses can only be constructed indirectly. For aggregated data, the only way to create this sort of data set at present is to output a set of data for each census year and combine the sets outside the NIS program, for example in dBase or SPSS/PC+. Future versions of the program can incorporate the capability to construct time series for aggregated data if it is desired.

After a census tract year is selected, the user is prompted to pick a data table from a menu showing the list of tables available for that census year. The census data table structure for the most part follows that of the Census Tract Statistics Books published by the Census Bureau and available for the Pittsburgh SMSA for the more recent censuses and for the City and County previously. Therefore, these books may usually be utilized as documentation for the data structure contained in the program at both the table and individual variable level.³

When a table has been selected, the individual variables or

³ For a more comprehensive description of the data structure and the similarities and dissimilarities with the Census Tract Statistics Books, see Part III and the Appendix where a complete list of tables and variables is presented.

data items in that table are presented as descriptive labels on the screen for selection by the user. At this point in the selection process, the user is selecting vectors of data (by variable name) to be included in the output. As elsewhere, a choice is made by pressing ENTER or SPACE when the highlight bar is placed on an entry. An asterisk (*) appears to the left of the variable label on the screen indicating that the variable has been selected. During the selection of variables, the ENTER and SPACE keys act as toggle switches; that is, they toggle on or off a particular variable to be included in the select set. Thus, if an incorrect selection is made, it can be unselected by placing the highlight bar on it and pressing ENTER or SPACE again. Note that the variables are included in the select set in the order in which they are chosen. Therefore, variables should, in general, be chosen in the order desired for the final output. However, variables may be reordered and/or deleted at any time after selection under MODIFY SELECT SET on the main menu (see below).

Data may be selected from any number of tables in this way. In order to select variables from a different table, pressing the ESC key will back up to the list of tables available (within the same data type) for the census tract year and a new choice can be made. Selection of variables within the new table is the same as above. Any tables which contain data for the particular census tract year definitions chosen may be accessed in this way. Pressing the F10 key returns the program to the main menu and allows the choice of another data type. At present, in addition to census data, Pittsburgh Police crime statistics are available for 1970 census tracts (1977 through 1984) and for 1980 census tracts (1985 through 1987), and Pittsburgh building construction permits are available for 1970 census tracts (1981 through May, 1989).

The NIS program maintains the selection set as long as it is not intentionally cleared by the user and the program is not exited, so any table can be reentered at any time for more selections or to unselect any choices.

2.4 GEOGRAPHIC SELECTION - Selecting and Modifying Geography

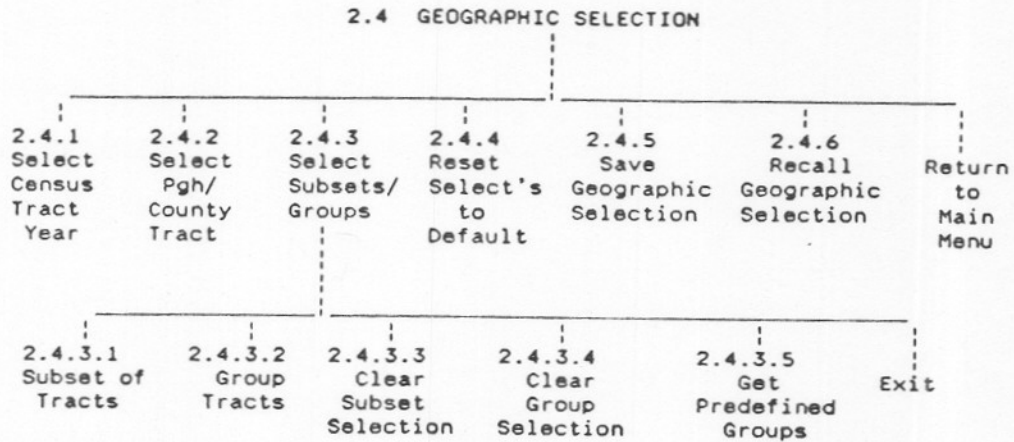
The NIS data system and program is based on census tract geography; that is, the data is stored in the data files by census tract. In order to select a set of tracts for data output or to group tracts to the neighborhood or higher aggregation level, the census tract geography must be manipulated. These manipulations may be of several types: a) The program will permit the selection of a subset of census tracts of any size and will permit the grouping together of tracts to form aggregates of larger size. The only requirement is that the aggregates be composites of whole tracts. b) Since each census year from 1930 through 1980 has differing census tract definitions, the data from each census year at the tract level is incompatible with the data from any other year and the program can allow only one census tract definition to be utilized at a time. Therefore, a census tract year (and only one) must be selected for each set of data to be output. c) Two separate sets of data are available, for census tracts in the City of Pittsburgh and for census tracts in the remainder of Allegheny County. In the current version of the program, these data sets are mutually exclusive. A choice must be made between them. d) Finally, to avoid reentering geographic selections, the program has the capability to save these selections to a DOS file and recall them later.

The GEOGRAPHIC SELECTION option on the main menu provides a number of procedures for accomplishing all of these tasks: selecting, grouping and saving/recalling a set of geographic definitions for data output. When this option is chosen, a submenu appears on the screen listing the geographic selection functions and allowing the user to select among them. The highlight bar should be moved to the desired function and ENTER or SPACE pressed to select it. A description of each of the geographic functions follows.

If no selections are made, a set of defaults will govern the geography at data output time: The census tract year defaults to 1980, thus 1980 census data and any other data, such as police crime statistics for 1985, 1986, 1987, which are defined for 1980 tracts are available. The data set defaults to City of Pittsburgh data, consequently only Pittsburgh census tracts are available. Finally, the default with respect to selection of a subset or a grouping of tracts is that none is in effect; the entire set of tracts with no grouping will be output. In order to change any of these, the options described below must be utilized.

Figure 2 provides a schematic diagram of the GEOGRAPHIC SELECTION module and its components, including the section number of the manual for each selection. This is provided to make it easier to locate the component description in the manual.

FIGURE 2
SCHEMATIC DIAGRAM OF THE GEOGRAPHIC SELECTION MODULE
AND ITS COMPONENTS



2.4.1 SELECT CENSUS TRACT YEAR - To define the census tract geography in use, a census tract year must be selected. Only one can be in use at a time. The selection of the census tract year can occur in one of two ways: either as the first step of the process of data selection and retrieval described above, or by utilizing SELECT CENSUS TRACT YEAR under GEOGRAPHIC SELECTION to select or change the year. The two options present the same screen and work the same way. However, the year can only be selected or changed during the data selection phase until variables have been added to the selection set. After variables have been added to the selection set the tract year menu will not be presented on the screen during data selection anymore. The only way to change the tract year in this case is to do it explicitly using SELECT CENSUS TRACT YEAR. Note that if the tract year is changed after variables have been selected, the selection set will be cleared and the variable selections lost. A warning message will be displayed before this happens giving the user a choice of whether to continue or not.

The actual operation of the tract year selection menu is straightforward. A menu showing the census years from 1930 to 1980 is presented on the screen and the user must move the highlight bar onto the year to be chosen and press ENTER or SPACE. As mentioned above, if no selection is made, the census tract year defaults to 1980.

2.4.2 SELECT PITTSBURGH/COUNTY TRACT SET - Two separate sets of tract data having the same variables are contained in the NIS data system. These are data for the census tracts in the City of Pittsburgh and data for the remainder of the census tracts in Allegheny County. In the current version of the program, these two data sets are alternatives and they cannot be accessed at the same time.⁴ The program normally accesses the Pittsburgh tract set as a default. This option provides a way to change the tract set to the County tracts or back again. The screen menu for this selection shows the two choices, and the user must move the highlight bar to the desired selection and press ENTER or SPACE.

Data for Allegheny County is available only for the census years from 1940 through 1980. For 1930, only the Pittsburgh tract data is available. Therefore, if 1930 variables are selected from the Pittsburgh data set and then an attempt is made to change to the County data set for the same variables using this option, an anomalous situation is created. After a warning message indicating that the selection set will be cleared, the user has an opportunity to stop the process. If the choice is made to continue, the selection set is cleared (all variables are lost) and the year is reset to 1930.

Two tables from the 1980 data set are available only for the Pittsburgh tracts as well, Table P-1 and Table H-1.⁵ The same anomalous situation may arise for variables selected from these two tables, however they are handled differently by the program since they may be intermixed with other 1980 variables which are available for the County. Basically, these variables are ignored at output time when data for the County is being processed. A zero is output for them (or any computed variable which contains them) in place of a data value in the output data set.

⁴ Construction of a data set for the municipalities of Allegheny County, including Pittsburgh, can be done indirectly. Grouping of the County tract set to the municipal level (see below) and the outputting of this data to a file provides municipal data for all municipalities other than the City. For the Pittsburgh data set, all tracts can be grouped together sending one output record to a file for the City. The two files can then be combined outside the NIS program to create the full municipal data set for the County.

⁵ Tables X-1 and X-2, derived from the Census Bureau STF3 data tape, contain most of the same information as these two tables and are available for Allegheny County. See Part 3 and the Appendix for more information on the contents and sources of the data.

2.4.3 SELECT SUBSETS/GROUPS - The ability to present data for a subset of the entire set of census tracts and the ability to aggregate the tract level data by grouping together census tracts are two of the most useful features of the NIS. Grouping tracts in particular allows the user to present the data at the geographic level of his or her choosing (e.g., neighborhoods, regions, wards, etc. for the City, and municipalities, regions, etc. for Allegheny County). This selection on the main menu provides access to these features. When the SELECT SUBSETS/GROUPS option is chosen, a submenu is presented to list the subset and grouping options. These are described below.

2.4.3.1 SUBSET OF TRACTS - This option provides an opportunity to select a subset of the census tracts for which data will be output. When the user elects to choose a subset of the whole set of tracts with this option, the subset is chosen from the data set that is currently in use. (At any point in the program, either the census tracts for the City of Pittsburgh or the census tracts for Allegheny County are the data set in use.)

When this option is chosen from the SELECT SUBSETS/GROUPS menu, a list of the census tract numbers for the data set currently in use and for the census tract year selected (e.g., 1930 through 1980) are presented on the screen. Along with the tract numbers, the neighborhood name (Pittsburgh tracts) or municipality name (Allegheny County tracts) assigned to the tract are shown. These names reflect the predefined groups described in Section 2.4.3.5 below. Also shown on the screen is the county abbreviation, which in all cases is 'AL' for Allegheny. This information was included since future versions of the program may include data from other counties in the Pittsburgh SMSA.

The user may move through this menu (which takes up multiple screens) the same way as through other menus - using the number pad keys. As elsewhere, a choice of census tract is made by pressing ENTER or SPACE when the highlight bar is placed on an entry. An asterisk (*) appears to the left of the tract number on the screen indicating that the tract has been selected. On this menu, the ENTER and SPACE keys act as toggle switches; that is, they toggle on or off a particular tract to be included in the set. Thus, if an incorrect selection is made, it can be unselected by placing the highlight bar on it and pressing ENTER or SPACE again. The order of selection for the census tract set is immaterial - data will always be presented in tract order.

The selection of subsets of tracts and grouping tracts (see next Section) occur independently from each other. If a grouping of tracts has been carried out and a subselection of tracts is also made, output will be generated for only those tracts which have been selected, whether grouped or not. For instance, if a group of tracts 1, 2, and 3 has been defined, but only tracts 1 and 3 have been selected as part of a subset, the group output will only

contain data from tracts 1 and 3 grouped together. Thus the user must be aware of the interaction of the two procedures if both are in effect. Both the tract selection set (via asterisks) and the group definitions (via group numbers) are shown on the screen during both procedures.

2.4.3.2 GROUP TRACTS - This procedure is the main mechanism to create aggregations of tracts for output. The groups are created by assigning a group number to a tract. Groups may be assigned to some or all tracts, and grouped and ungrouped tracts may be mixed in the output.

When this option is chosen from the SELECT SUBSETS/GROUPS menu, a list of the census tract numbers for the data set currently in use and for the census tract year selected (e.g., 1930 through 1980) are presented on the screen. Along with the tract numbers, the neighborhood name (Pittsburgh tracts) or municipality name (Allegheny County tracts) assigned to the tract are shown. These names reflect the predefined groups described in Section 2.4.3.5 below. Also shown on the screen is the county abbreviation, which in all cases is 'AL' for Allegheny. This information was included since future versions of the program may include data from other counties in the Pittsburgh SMSA.

The user may move through this menu (which takes up multiple screens) the same way as through other menus - using the number pad keys. The procedure for grouping tracts is to enter the group number (up to three digits long) for each tract which is to be grouped when the highlight bar is on the line containing the tract information. These numbers must be assigned by the user and are arbitrary. The group number appears on the screen to the left of the tract number and is registered by pressing the ENTER or SPACE key or by using the number pad keys to move the highlight bar off the entry to be grouped. The group number may be changed by placing the highlight bar back on the same line and reentering the new group number. The new number supercedes the old number. To remove any group membership from a tract, enter a zero followed by ENTER, SPACE or a number pad movement key.

The grouping of tracts specified by the user is maintained by the program in a variable with an element for each tract in the data set. This variable contains the group number (a number from 0 to 999) assigned by the user. A zero in the grouping variable indicates that the tract is not grouped with any other tracts while any number from 1 to 999 indicates that the tract is grouped with all other tracts having the same assigned grouping number.

When output is created, the data from all tracts which are grouped together are combined for each variable to create the group value which is then output. In the current version of the program, the only procedure for combining values is to add them together to create the combined group value. Thus, the grouping procedure is

only meaningful for counts and is not meaningful for percentages, means, medians, etc. This is very important to remember when selecting variables for grouping. If additional combination procedures are desired, a future version of the program can incorporate them.

The selection of subsets of tracts and grouping tracts (see previous Section) occur independently from each other. If a grouping of tracts has been carried out and a subselection of tracts is also made, output will be generated for only those tracts which have been selected, whether grouped or not. For instance, if a group of tracts 1, 2, and 3 has been defined, but only tracts 1 and 3 have been selected as part of a subset, the group output will only contain data from tracts 1 and 3 grouped together. Thus the user must be aware of the interaction of the two procedures if both are in effect. Both the tract selection set (via asterisks) and the group definitions (via group numbers) are shown on the screen during both procedures.

2.4.3.3 CLEAR SUBSET SELECTION - The purpose of this option is to remove a previously selected subset of tracts and reset the census tract selection set to all the tracts of the current data set (either City of Pittsburgh or Allegheny County). In other words, this procedure puts the subset selection set back to its condition when the program was first started. The tract selection set is maintained by the program in a variable with an element for each tract in the data set. This variable contains a zero or a one for each tract with one indicating membership in the selection set and zero indicating non-membership. This procedure resets all tract selection variable elements to one, thus indicating no subselection. This procedure operates independently from the grouping selection and the other geographic selections (census year and data set) and has no effect on them. No additional input is required from the user when this option is chosen.

2.4.3.4 CLEAR GROUPS SELECTION - The purpose of this option is to remove a previously defined grouping of tracts and reset the census tract group indicator to no grouping. This procedure puts the group indicator variable back to its condition when the program was first started. The grouping of tracts specified by the user is maintained by the program in a variable with an element for each tract in the data set. This variable contains the group number (a number from 0 to 999) assigned by the user in the grouping procedure (see Section 2.4.3.2 above) or in the selection of predefined groupings (see Section 2.4.3.5 below). A zero in the grouping variable indicates that the tract is not grouped with any other tracts while any number from 1 to 999 indicates that the tract is grouped with all other tracts having the same assigned grouping number. This procedure resets all the grouping indicator numbers to zero, thus indicating no grouping. This procedure operates independently from the subset of tracts selection indicator and the other geographic selections (census year and data

set) and has no effect on them. No additional input is required from the user when this option is chosen.

2.4.3.5 GET PREDEFINED GROUPS - This procedure provides a quick way to group the census tracts of the current data set (either City of Pittsburgh or Allegheny County) by selecting the predefined set of groups already contained in the system. These groups establish a set of neighborhoods (by a grouping of whole tracts) for each census year for the City of Pittsburgh tract set and similarly establish a set of municipalities for each census year for the Allegheny County tracts. The effect of choosing this option is to assign the neighborhood or municipal grouping numbers already in the NIS to the grouping indicator variable, producing the same result as if the user had used the grouping procedure (see Section 2.4.3.2 above) to assign the same group numbers. Once assigned, these group numbers may be changed or otherwise edited by the user using the grouping procedure. No additional input is required from the user when this option is chosen.

For the City, the predefined neighborhood groupings are based on Department of City Planning (DCP) neighborhood definitions for 1980 and 1970 and are carried backward through the 1930 census by utilizing the 1970 neighborhood boundaries and assigning the tracts for previous years to neighborhoods based on DCP maps. For Allegheny County, the predefined municipality groupings are based on the Census Bureau municipality codes and Census Bureau maps and are involve less judgement. The complete list of census tracts and their group assignments are shown in Appendix D for the Pittsburgh tracts and in Appendix E for the Allegheny County tracts.

These definitions of neighborhoods and municipalities as groupings of whole census tracts are UCSUR's attempt to provide useful groups, but they may be changed if they are not accurate or do not represent the most useful grouping set. User input in this regard will be useful in changing the definitions if necessary. Also, additional predefined tract groupings for either alternate neighborhood or municipality groupings or aggregations at other levels (e.g., ward, region, etc) are quite easy to add to the NIS program in the future if they are desired. Note that alternate groupings may always be created by the user utilizing the grouping procedure and then saved to a DOS file for future use. While somewhat time consuming initially, the effect of this procedure, once completed, is the same as having additional predefined groupings built into the system.

2.4.4 RESET SELECTIONS TO DEFAULT - This option provides a quick way to reset all the geographic selections to their default values at the same time. This includes resetting the census tract year to 1980, the census tract data set to the City of Pittsburgh data set, and removing any subselection or grouping of tracts which is in effect. If a selection set of variables has been made from any census year other than 1980, the selection set must be cleared

as well. A warning message indicating that the selection set will be cleared is shown on the screen and the user has an option to stop the process. If the choice is made to continue, the selection set is cleared (all variables selections are lost) and the year is reset to 1980. No additional action is required on the part of the user other than the selection of this option from the GEOGRAPHIC SELECTION MENU.

2.4.5 SAVE GEOGRAPHIC SELECTIONS - The NIS program has the capability to save to a file and later recall the geographic configuration which is in effect at any time during program use. This capability operates independently from the saving of a selection set of variables. Thus, saved variable selection sets and saved geographic configurations may be mixed in any allowable combination for which data is available for the creation of output. This capability allows the user to set up standard geographic configurations (in addition to the predefined groupings described above) for later use, such as ward or regional groupings or the neighborhoods for a particular section of the City (a grouping of a subset of tracts), save them, and then recall them for use whenever desired.

When this option is selected, a filename is requested for the file to which to store the configuration. Only the filename itself without a period or an extension should be supplied. The geographic configuration is then stored to a dBase database file called "filename".GEO. Note that the extension automatically provided to this file is different from those provided to the two variable selection set dBase files. If both a selection set and a geographic configuration are to be saved, the same root filename can be used for both with no conflict occurring.

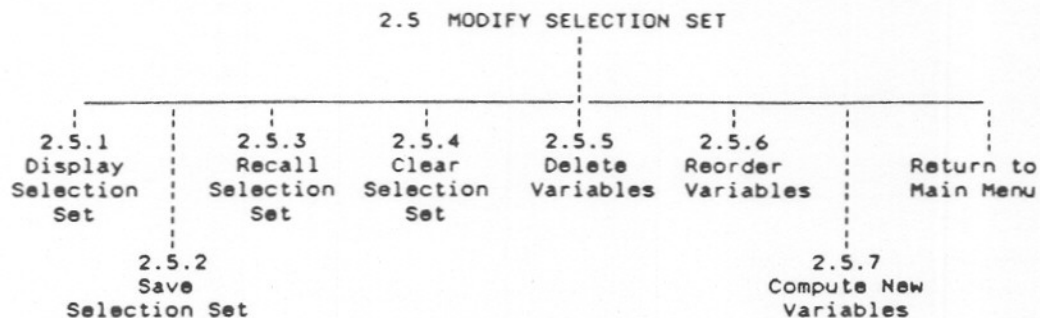
2.4.6 RECALL GEOGRAPHIC SELECTIONS - This option reads a previously saved file containing a geographic configuration. The program requests the filename of the file to recall. Only the filename without a period or an extension should be supplied. The program expects the file to have the extension ".GEO" so the extension should not be changed on a previously saved file. The recalled geographic configuration replaces any geographic selections that were previously in effect.

2.5 MODIFY SELECTION SET - Modifying Numeric Variables

The selection on the main menu labelled MODIFY SELECTION SET provides a sub menu with a number of functions for manipulating the variables of the selection set. These include displaying the variable labels for each variable in the selection set, saving the selection set to a file and recalling it from the file, clearing the selection set in order to start a new set, deleting or reordering variables currently in the set, and computing new variables based on those already chosen. These procedures may be carried out at any time after variables have been added to the selection set. Figure 3 shows a schematic diagram of the MODIFY SELECTION SET module and its components with the manual section number for each component.

FIGURE 3

SCHEMATIC DIAGRAM OF THE MODIFY SELECTION SET MODULE AND ITS COMPONENTS



2.5.1 DISPLAY SELECTION SET - When this choice from the modification menu is made, the program presents on the screen (for viewing only) the labels for all the variables in the selection set. A letter from the sequence of letters (A-Z,AA-AZ,BA-BZ,etc.) is placed at the beginning of each label for identification purposes. This letter is used in the other modification functions to select variables to be operated on, and is provided here for reference purposes. If there are more variables in the set than will fit on one screen, the number pad keys can be used to scroll through the set. The UP ARROW, DOWN ARROW, HOME, END, PGUP and

PGDN keys all act as they do elsewhere in the program to allow movement through the list. The selection set list can be printed by pressing the F9 key while the selection set is displayed on the screen. Make sure the printer is on and ready for output before printing.

2.5.2 SAVE SELECTION SET - Saving the selection set to a DOS file is a straightforward operation which allows for preserving and recovering work done during an earlier program session. When this option is chosen, the user is asked for a filename to save the selection set to. The filename should not include a period or filename extension. The selection set is saved in two DOS files, both having the same filename but with different extensions. All normal variables in the selection set are saved in a file called "filename".DBF (DBF is the dBase database default extension). This is a dBase database file. The text containing an expansion of the formula for a computed variable, if any exist, is also stored in a dBase database file which is called "filename".CMP. If there are no computed variables in the selection set, this file is not created.

2.5.3 RECALL SELECTION SET - When the saved selection set is recalled by choosing this option, the filename is again asked for. Again, only the filename itself without a period or an extension should be entered since the program assumes the extensions described above. The recalled selection set replaces the existing selection set if there is one. Thus the recalled set cannot be appended to a current set.

2.5.4 CLEAR SELECTION SET - This option provides a quick way to delete all the variables from the selection set and begin again. While it is always possible to delete each variable individually, this procedure simplifies the operation if all the variables are to be discarded. When the set is cleared, all record of the previously selected set is lost. No further input from the user is required after this option is chosen.

2.5.5 DELETE VARIABLES - This option provides a way of deleting individual variables from the selection set. The program presents the list of variable labels on the screen in the same form as the display selection set output, with a sequence letter at the left of the label. Multi-page lists of variables may be scrolled through using the number pad keys. The user is asked to enter the letter of the variable to be deleted from the set. Enter the letter and an ENTER or SPACE; the variable is removed and the revised selection set displayed on the screen.

2.5.6 REORDER VARIABLES - After variables have been added to the selection set and output is desired, it may be necessary to reorder the variables, for instance for a particular analysis. This procedure allows the user to accomplish this. The order of the variables shown on the display is the same order in which they

will be output to the screen, printer or file from the output module.

When this option is selected, the selection set labels appear on the screen in the same form as the display selection set output, with a sequence letter at the left of the label. Multi-page lists of variables may be scrolled through using the number pad keys. The program asks for the variable to be moved (enter the letter and an ENTER or SPACE) and the location to which it will be moved (again enter a letter and an ENTER or SPACE). The original variable (first letter entered) will be moved so that it is now at the spot denoted by the second letter entered, and the variable which was there originally (and those before or after it) will be moved up or down a spot depending on its relationship to the original variable (before it or after it). The first variable's slot will be filled by an adjacent variable moving either up or down. Thus the letters used and the number of them will remain the same while the order of the variables will change.

2.5.7 COMPUTE NEW VARIABLES - Most census numbers are counts, and this is true of the other numeric data sources in the NIS as well. However, the most useful figures for analysis are often percentages or other statistics derived from the raw counts. The NIS program has incorporated the capability of calculating many of these statistics. This option makes this possible by allowing the computation of a new variable from those variables already in the selection set.

When this option is selected, the user is first asked to supply a variable label for the variable to be computed. This label will be used in all displays of the selection set after the variable computation process is completed. Enter the label in the highlight box shown on the screen (up to 70 characters) and then press ENTER.

After the label is entered, the selection set variable labels appear on the screen in the same form as the display selection set output, with a sequence letter at the left of the label. Multi-page lists of variables may be scrolled through using the number pad keys. Near the bottom of the screen, the phrase "New Variable=" appears with a 60 character space to the right of it. The mathematical formula for the new variable should be entered into this space.

Rules for the formula are as follows: Each select set variable in the formula should be represented by the sequence letter before it in the variable list. In general, the formula should be in the form allowed by dBase for operations on fields. The following characters are allowed in the formula: alphabet letters denoting existing variables, numbers, operators, parentheses and finally, the period and the space characters. Characters which are not allowed will either produce a beep indicating an error or will be ignored (not entered into the

formula). The following operations (operators) are allowed in the formula: addition (+), subtraction (-), multiplication (*), division (/), and exponentiation (**). In general, spaces between variable letters, operators, etc. should not be used except where absolutely necessary for clarity. Parentheses should be used liberally to delimit operations. Normal left to right parsing of the formula occurs following the rules for expressions used by dBase.

As an example, if the percentage of females in a set of tracts is to be calculated, the number of females (suppose this is variable B) can be divided by the total number of people (suppose this is variable C) for each tract and the result multiplied by 100. The formula for this computed variable could be expressed as $((B/C)*100)$.

The computed variable is placed in the selection set as is any other variable. The computed variable is always given a length of 12 with 2 decimal places. While this often wastes space on output, determining the actual length required for the data values and whether decimal places are necessary is beyond the scope of this version of the program.

The result for each computation formula is actually calculated at the time of output of the data. The formula is first expanded by supplying the database names of the variables denoted by the letters and then Clipper does the actual computations, referring to the database file(s) for the values to substitute into the formula.

When the selection set is saved to a DOS file (see SAVE SELECTION SET above), the text of the formula is stored in expanded form (i.e., containing actual field names rather than letters) in a secondary dBase file with the extension ".CMP". This file is only created if there are computed variables in the selection set when it is saved. The computed variables in the saved selection set and in this file are given names beginning with "COMP" followed by a two digit computed variable sequence number (e.g., the second computed variable in a selection set would be called "COMP02"). At present, the only way to review or edit a computed variable once it is entered is to save the selection set to a file and examine and modify the expanded formula in the ".CMP" file. However, modifying this file is not recommended unless the user understands dBase syntax and the data structure of the NIS system fairly well.

Note that although the length of the formula may be up to 60 characters, the limiting factor more likely to be encountered is the expanded length of the formula since this must be stored in a character field accessible by dBase IV. This type of field has a length limit of 254 characters and therefore, so must the expanded formula. While Clipper has the capability to create and store longer character fields, these fields would not be accessible from

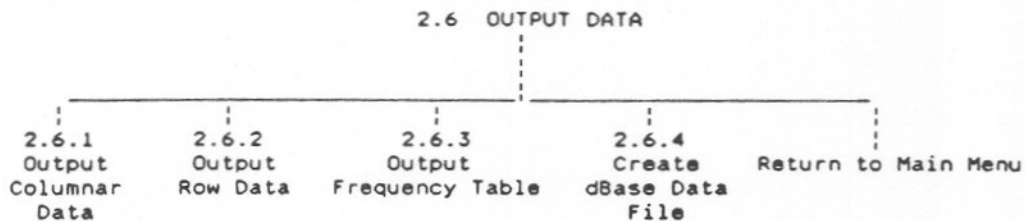
dBase. In order to allow access to the computed variable, the 254 limit has been used. No warning is currently given if this limit is exceeded, and the expanded formula in the save file must be examined for error (the rightmost part of the formula will be cutoff) if the results do not appear to be accurate. In practical terms though, unless there are more than 12 or 13 references to variables (by sequence letter) in the formula, this type of error condition is very unlikely. If this limit is found to be too restrictive, another version of the program can easily be supplied which will allow a much longer expanded formula (but no dBase access to the formula).

2.6 OUTPUT DATA - Obtaining Output

The procedures found on this menu are the primary means of outputting data from the NIS program. Four type of output are currently available: 1) outputting a data matrix in columnar form (tracts or other geography as rows and variables as columns) to the screen, a printer or a file, 2) outputting a data matrix in row form (variables as rows and tracts or other geography as columns) to the screen, a printer or a file, 3) a one variable frequency table showing the number of occurrences of each value of a specified variable, and 4) the creation of a dBase database file containing the variables in the select set (as fields) for the geographical configuration in effect (as rows). Figure 4 shows a schematic diagram of the OUTPUT DATA module and its components with the manual section numbers for each component.

FIGURE 4

SCHEMATIC DIAGRAM OF THE OUTPUT DATA MODULE AND ITS COMPONENTS



2.6.1 OUTPUT COLUMNAR DATA - This option provides output of the selection set of variables as columns in a data matrix with the geographic selections, tracts or aggregate groupings of tracts, as rows - the standard output. This type of output may be sent to either the screen, a printer or a file for possible transfer to other programs.

The variables are output in exactly the order they are found in the variable selection set. The data is presented without headings. The list of selection set variables labels can be

printed out (using DISPLAY SELECTION SET under MODIFY SELECTION SET on the main menu) for help in identification of variables.

When output is sent to the screen, a screen-full of data is presented at a time with a pause - the user strikes a key when more output is desired. The output process can be terminated at any time by pressing the ESC (escape) key.

When output is to be sent to the printer, make sure the printer is on and ready for printing before starting or an error condition will be created. No output will be sent to the screen while the printing is going on.

When output is sent to a file, the program will ask for a file name. The full file name including the period and extension desired should be entered. No further input from the user is required.

2.6.2 OUTPUT ROW DATA - This option provides output of the selection set of variables as rows in a data matrix with the geographic selections, tracts or aggregate groupings of tracts, as columns. This type of output may also be sent to either the screen, a printer or a file for possible transfer to other programs. This form of output is identical to columnar output with the data matrix transposed. Output in this form will be very difficult to read for a standard type of data matrix, where the rows (tracts or other geography) far outnumber the columns (variables). Where the "row" form of output is more useful is for the situation where data is output for a small number of tracts or other geographic units for a set of variables.

The variables are output in exactly the order they are found in the variable selection set. The data is presented without headings. The list of selection set variables labels can be printed out (using DISPLAY SELECTION SET under MODIFY SELECTION SET on the main menu) for help in identification of variables.

When output is sent to the screen, a screen-full of data is presented at a time with a pause - the user strikes a key when more output is desired. The output process can be terminated at any time by pressing the ESC (escape) key.

When output is to be sent to the printer, make sure the printer is on and ready for printing before starting or an error condition will be created. No output will be sent to the screen while the printing is going on.

When output is sent to a file, the program will ask for a file name. The full file name including the period and extension desired should be entered. No further input from the user is required.

2.6.3 OUTPUT FREQUENCY TABLE - The frequency table output is a frequency count of the occurrence of each value in a column of data. This information is currently presented only on the screen with one screen-full appearing at a time with a pause for the user to strike a key when more output is desired.⁶ This is most useful when the number of values contained in a column of data is relatively small. Frequency distributions may be created for computed variables as well as for the variables contained in the data set. The frequency distribution will take into account both subsets of tracts which have been selected and groupings of tracts which have been defined. That is, the frequency distribution will be carried out only on the subset of tracts selected and will include one value for each group rather than a value for each group element (tract).

When this option is chosen from the output menu, the variable labels for the select set are presented on the screen in the form used for the display of the selection set, with a letter indicating the sequence of variables. For multi-page variable selection sets, the number pad keys allow movement through the list. The user is asked to specify the variable for the frequency table by letter. When a letter is entered, followed by an ENTER or SPACE key, the frequency distribution is presented on the screen.

2.6.4 CREATE DATABASE FROM SELECT SET - This procedure provides a quick and easy way to create a dBase database file containing data for the variable select set for the geographic configuration in effect. The data in this database file is identical to the data which is output in the "column" and "row" output procedures, however it is output to a dBase file in column form. This file is accessible from dBase III+ or dBase IV. Note that dBase database files are translatable to the formats used by many other programs either by dBase itself (e.g., Lotus 123) or by the target program (e.g., SPSS/PC+). Therefore, this option provides the easiest way to transfer data to other programs.

The column names used for the dBase file are the names used in the NIS data system and are listed in full in the Appendix. Computed variables are given the name "COMP" followed by a sequence number. The column lengths and decimal spaces are also those used in the NIS system (see Appendix). Computed variables are assigned length 12 with two decimal places to allow for a very large computed number. This procedure may take a long time to complete with no screen activity and little evident computer activity (i.e., disk read/writes), so be patient.

⁶ Note that a printed copy of the output can be created by printing each screen full individually.

2.7 RETRIEVE TEXT

The program currently processes only numeric information. While the main menu shows selections for textual types of information processing under the heading RETRIEVE TEXT, the capability for dealing with this type of information has not yet been built into the program. The intent in including this selection is to show that the NIS program is not limited to numeric information and can provide written information at the tract, neighborhood, or any geographic aggregation level. The examples shown on the main menu suggest some possible applications of the text processing system: annotated bibliographies, report or other literature excerpts, documentation of sources, and various neighborhood characteristics such as lists of groups and leaders and textual descriptions of neighborhoods which could contain the latest developments. If desired, the text processing system can incorporate text entry and editing capabilities to allow for maintaining the information base. Hopefully, applications will arise which will allow the further development of the system to allow entry, editing and retrieval of textual information. Until then, however, the program merely suggests capabilities which are possible.

2.8 RUN REPORTS

The main menu has a selection labelled RUN REPORTS. This is also a currently inoperative procedure which suggests possible future expansion of the program. The function of this option is to provide a menu of preprogrammed standard reports which can be automatically executed upon selection. These reports can be flexible and user designed and may include any data contained in the system for any year. An example might be a standard neighborhood profile which can be produced for output whenever desired.

In addition to the capability of producing output from a standard report, it is also possible to provide a "report generator" module which will allow the user to create (and perhaps modify) reports as desired and add them to the existing library of standard reports. If there is sufficient interest in this function, it will be added to the NIS program.

2.9 Help System

Another capability which is only suggested in the program but which may be developed further in the future is the help system. As mentioned previously, the emphasis in this phase of system development was on data procurement and providing the capability of data retrieval. Due to a lack of development time, a help system was not included in the current version of the program. A full help system which includes not only instructions on how to operate the program but also the source for each data item and notes on special information necessary to use the item (e.g., how it was derived) can be included in future versions of the program.

2.10 QUIT/EXIT TO DBASE - Exiting from the program

Exiting from the program is straightforward. Selecting QUIT on the main menu terminates the program, deletes all temporary files and returns to DOS. All selections made during the run are lost unless they have been explicitly saved. Selecting EXIT TO DBASE on the main menu has the same effect except that the dBase program is run. In order to successfully start dBase, it has to be accessible from the default directory (i.e., on the path). If this is not the case, this form of exit should not be used or the batch file which controls the program will go into a continuous loop. This is not a serious situation, and a Control-C combination will terminate the batch job with no damage. However, make sure dBase will run when you start it from the NIS directory before exiting from the NIS in this way.

3. DESCRIPTION OF THE DATA

3.1 Overview

The NIS, in its current form, is based on census tract geography. That is, the data is stored for each variable in the system at the census tract level. The data may be retrieved from the program at the census tract level or at any aggregate level composed of groupings of whole census tracts. However, the raw data for the system is limited to data available for census tracts.

Data is available in the NIS for the City of Pittsburgh census tracts and for the census tracts in the remainder of Allegheny County. Since the principal goal of the system is to provide data for the City alone, the data for Allegheny County has been stored in separate data files and is accessed separately in the program. Data need not be available for both the City and the County to be included in the system.

There are three primary sets of data in the NIS: (1) census data; (2) police crime statistics data; and (3) building construction permit data (including the value of new constructions). The census data is available from 1930 to 1980 (on a decennial basis) and includes most of the important tables in the Census Tract Statistics Books⁷ published by the Census Bureau for the City of Pittsburgh, Allegheny County or the Pittsburgh SMSA, depending on the census year. The crime data is available for the City of Pittsburgh census tracts from 1977 to 1987 on an annual basis. The source of the data is the Police Statistical Reports published annually by the Pittsburgh Department of Police. Finally, the data on building construction permits and the value of new constructions is available on an annual basis from 1981 to 1989 (January to May 1989) again for the City of Pittsburgh tracts only. This data was gathered directly from the ledgers maintained by the Bureau of Building Inspections of the City of Pittsburgh. The description of each of the data sets which follows includes their sources, coverage, limitations and other relevant information. The remainder of this section provides some general information relevant to all data sources.

The data is structured by tables each of which is stored in one dBase database file. Each table contains data values for a set of census tracts (either Pittsburgh or Allegheny County) from a specific census year on a set of related variables. The census tract definitions are taken directly from the tract delineations used by the Census Bureau for the particular census year. The set

⁷. Not all the data in the Census Tract Statistics Books is included in the NIS. The tables for inclusion were chosen jointly by UCSUR and the DCP for some census years, and were based on available computerized data for the other years.

of variables either is derived from one table in the Census documentation or provide logically related information for a given time period (e.g., annually). For the census data, these tables closely parallel the tables in the Census Tract Statistics Books which may therefore be used as partial documentation for the system data. The NIS tables have been given the same names and titles as the tables from the Tract books to facilitate their use in this way. For the crime and building permit data, all the available information is presented in annual tables.

The complete list of tables (one per data file) and the variables in each table for the three data categories are listed in Appendix A through C. The Appendix tables show, for every data table, the variable name used for each variable in the NIS system, the number of columns the variable is stored in (its length), the number of decimal places it has if any, and the variable label used for the variable in the system. The variable label is the only information the user will normally see in the NIS since it is this label that is presented on the screen in the data retrieval module. While every effort was made to make these labels as descriptive as possible, in some cases they had to be abbreviated or shortened because of lack of space. These labels may be traced back to the original documentation for the data source if any questions arise.

The variable names are the dBase field names used for the variables in the databases and are transparent to the user in normal usage. However, if a data set is saved as a dBase database from the program, or if the user wishes to use the database files directly, these names must be used. The variable names consist of a letter indicating the data type (the letter "C" for the census data, the letter "R" for the crime data, and the letter "B" for the building permit data) followed by either a one digit year for the census data (e.g., "8" for 1980) or a two digit year for the crime and building permit data (e.g., "80" for 1980), the letter "V" and finally a sequentially assigned variable number. Thus, the first census variable in the system from the 1980 census has the name "C8V1", and the 10th crime variable from 1980 has the name "R80V10". The variables from all tables in a census year (within data type) are numbered sequentially across tables. The census tract number is always the first variable in each table and is called "TRACT" followed by the two digit census year (e.g., "TRACT80").

Appendix D and Appendix E show the census tract geography used in the NIS system for the City of Pittsburgh tracts and the Allegheny County tracts, respectively. These tables show the census tract year, the tract number and the neighborhood (Pittsburgh) or municipality (Allegheny County) the tract is located in by name and code number. These code numbers are the grouping numbers assigned when the GET PREDEFINED GROUPS option is used in the GEOGRAPHIC SELECTION part of the program (see Section 2.4.3.5).

For 1970 and 1980, when Pittsburgh neighborhoods were recognized officially, the assignment of tracts to neighborhood was determined from maps provided by the Department of City Planning. For the earlier census years, neighborhood boundaries were based on those of 1970 and tract to neighborhood assignments were made by comparing the census tract maps of 1960, 1950, 1940, and 1930, with those of 1980 and 1970. The neighborhood code numbers for the City were derived from assigning numbers to the neighborhoods listed alphabetically. There were 88 neighborhoods in 1980, 78 in 1960 and 1970, and 76 in 1950, 1940, and 1930.

The municipalities in Allegheny County were also assigned codes. For the 1980 and 1970 censuses, which have FIPS codes assigned to the municipalities, census coding was used. For the other census years, the municipalities were listed alphabetically and numbers were assigned to them in ascending order. The determination of tract to municipality assignment was straightforward from Census maps and publications. Note that the census tracts are listed in sorted order in this Appendix (and when output from the program), whereas in the tract book they are listed by municipality.

3.2 Census Data

Census data is available for each of the census years from 1930 through 1980 - those censuses for which tracted data is available for Pittsburgh. In general, the census data in the NIS comes from the Census Tract Statistics Publication for Pittsburgh for each census, much of it key punched directly from these books. The principal exception to this is the 1980 data which is primarily derived from the Summary Tape File 3 (STF3) data tape from the Census Bureau. The data for 1930 and 1970 was entered from the tract books by UCSUR. The data for 1940, 1950 and 1960 was provided by the Social Development Center at the University of Chicago. As part of a long term study of urban conditions, they have key entered most of the data in the tract books for the entire U.S. for these censuses. They provided this data plus Census tape file data for other census years to UCSUR on computer tape. A summary description of the data for each of the census years is given in the sections below. Unless otherwise noted, the census data tables have data available for both the City of Pittsburgh and Allegheny County.

The census data is contained in a total of 84 tables - one per dBase database file. Of these tables, 47 contain data for the City of Pittsburgh tracts (8 in 1930; 6 in 1940; 3 in 1950; 5 in 1960; 6 in 1970; and 19 in 1980), and 37 contain data for Allegheny County tracts (6 in 1940; 3 in 1950; 5 in 1960; 6 in 1970; and 17 in 1980). A complete listing of all census data in the system is shown in Appendix A.

Some general points which apply to all data items and tables are: (1) The value of zero in the data, in addition to being a true zero, may also denote unavailability or inapplicability of data. In general, the NIS replaces all unavailable data (whether suppressed by the Census Bureau for confidentiality, missing, or otherwise not available) with zero. (2) Not all tables from a census year are included in the program; and not all the variables from a table are always included.⁸ (3) In some cases, additional variables are generated from existing ones.⁹

Finally, it should be noted that it is very important that the

⁸. The decision to exclude certain whole tables and certain variables from tables was arrived at in collaboration with the City Planning Department.

⁹. For instance, the Census documentation often provides the total number of persons and the number of females in a tract. In cases like this the NIS usually provides the number of males as well. This is derived by subtracting the number of females from the total number of persons. Similar derivations were carried out whenever they were straightforward.

user become familiar with the Census Bureau publications describing the sources and limitations of the census data. The NIS system has simply provided the census data to the user as is, but the characteristics of the data must be fully understood to ensure correct usage. For example, the comparability of variables across census years may be questionable even though their labels may sound very similar. This can only be determined by examination of census variable definitions.

3.2.1 1980 Tables - There are 19 data tables for 1980 in the NIS system - eleven for the Population Characteristics series, and eight for the Housing Characteristics series. All but two of these tables are derived from the STF3 data tape. Only Table P-1 (population series) and Table H-1 (housing series) are derived from the Census Tract book. These two tables were entered by electronic scanning of the tract book¹⁰ and include data for only the City of Pittsburgh tracts. In order to have comparable data for both the City and County tracts, additional tables from the STF3 tape with similar information to these two tables were included. Table X-1 is a supplement to Table P-1 and Table X-2 is a supplement of Table H-1, and both of these tables have both City and County data.

Note that the STF3 tape data is primarily sample data inflated by the Census Bureau to represent the total population, whereas many variables (but not all) from the Census Tract books are 100% counts. These two different methods of arriving at the population counts are likely to result in differing values, and this may be evident in a comparison of the 1980 data in the NIS (largely STF3 data) with the 1980 Census Tract book. For the most part, these differences are minor. However, the user should be aware of these differences to ensure proper usage of the data. See the Census Bureau documentation and hardcopy of the data for a complete discussion of the characteristics of the data.

Also, see the Correction Note in the Introduction to the 1980 Census Tract Book for Pittsburgh for late corrections to the total population and housing counts in certain tracts. These corrected figures were not incorporated into the NIS since corrected counts are not available for the remainder of the variables; therefore, inconsistent counts would occur (e.g., the sum of total males and total females would not be equal to the corrected total population).

3.2.2 1970 Tables - There are 6 tables taken directly from the 1970 Census Tract Statistics book included in the NIS system. Four of these are for population series data, and two are for housing

¹⁰ The scanning of the census data was done on a trial basis and proved not to be cost effective due to the excessive time spent in checking the data for accuracy. These two tables have been thoroughly checked and verified.

series data. These data were key entered by UCSUR from the tract book for the Pittsburgh SMSA. Since the source of the tables is the tract book, this book can be used to document the system data, and variables for all the tables are exactly the same as those in the book. The only difference is that the County tracts are listed in sorted order when output from the program, whereas in the book they are listed in municipal order.

3.2.3 1960 Tables - There are five tables included from the 1960 census in the system. Three are from the population series and two are from the housing series. The source of the data for these tables is the computer tape from the Social Development Center at the University of Chicago. The 1960 data they provided is derived from a combination of tract book and computer file. Some of the data is from a special census done for the Atomic Energy Commission in 1960. The balance of the data was entered directly from the book (such as anything to do with race, data on unrelated individuals in Table P-1, and also extensive data from Table H-2). For the most part, however, the tract book may be used as documentation for this data as well.

Note that for the 1960 City of Pittsburgh tracts, all tract numbers contain alpha codes. These alpha codes are changed to numeric codes for use in the NIS. Thus, Tract 1A becomes Tract 101, and Tract 7E becomes Tract 705. However, for municipalities other than Pittsburgh, the alpha codes in front of the tract number are left unchanged. Thus, Tract WIB26900 in Wilkinsburg Borough is left as it is in the book.

The data received from the University of Chicago had missing entries for Tract MKC 209. Since some of the data was collected for the Atomic Energy Commission special census and, therefore, was not listed in the tract book, some values could not be recovered. Zeroes appear for this tract for those variables for which no data was available in the tract book. Finally, there are some split tracts in the SMSA data for the 1960 census. However, the values for the partial tracts are not listed in the NIS. Rather, the combined totals of the partial tracts are provided.

3.2.4 1950 Tables - There are three tables included for the 1950 census in the program. See Appendix A for a listing of these tables. The source of the 1950 data is the computer tape from the Social Development Center at the University of Chicago. This data was key entered by them directly from the 1950 Census Tract book, so the book may be used as documentation. Note that in the NIS, as for the 1960 census, Pittsburgh tracts with alpha codes have been changed to numeric codes. Thus, for example, Tract 26A becomes Tract 2601.

3.2.5 1940 Tables - For the 1940 census, there are 6 tables included in the system. See Appendix A for a listing of these tables. The source of the 1940 data is the computer tape from the

Social Development Center at the University of Chicago. This data was key entered by them directly from the 1940 Census Tract book. Note that, as for 1960 and 1950, Pittsburgh census tracts with alpha codes in them are changed to numeric codes. Also note that: (1) in Table 2 the variables for the male nonwhite population are arrived at by adding the figures for male negroes and males of other races; the same is true for the female nonwhite population; (2) in Table 3, the variable "Median School Years Completed" is available for males and females but not for total persons; (3) in Table 4, for the variable "Percent of all Occupied Units that are Owner-Occupied" some tracts have values in the system and not in the book because a new figure was possible to reconstruct from other data; (4) finally, in Table 5, the variable "Number of all Owner-Occupied Buildings" was also reconstructed from partial data. The data for some of the tracts in Table 5 may be slightly different from those shown in the census book due to reconstruction based on the explanatory note at the bottom of the table.

3.2.6 1930 Tables - The system includes eight tables from the 1930 census. The data for these tables is available only for the City of Pittsburgh since tracted data was not provided for the balance of the County in this census. See the Appendix for a listing of these tables and the variables included. The source for these tables is the 1930 census book and the data was key entered at UCSUR. As in 1940 through 1960, Pittsburgh census tracts containing alpha codes have been changed to numeric codes.

Note that in Table 1, the variables for whites are the sum of native whites (native parentage), native whites (foreign or mixed parentage), and foreign-born whites. Finally, in Table 7, note that the variables in the system are for the marital status of males and females of all classes.

3.3 Crime Data

The Crime statistics for the Pittsburgh tracts are taken from the Pittsburgh Department of Police Statistical Report, published annually from 1977-1987 by the Department of Police, City of Pittsburgh. One table per year is included in the NIS system. Unfortunately, there is an incompatibility between the earlier years and the later years in the tract definitions used -- the data for 1977 to 1984 were presented by the police statistical reports based on 1970 census tracts, while the data for 1985 to 1987 were based on 1980 tracts. Thus, a time series cannot be constructed at the tract level for the entire series because of differences in the tract definitions. It is possible, however, to construct this time series at the neighborhood level (or at other aggregations).

The crime statistics for 1986 and 1987 include data for all the census tracts in the city; however, they also include a tract entry marked "other". This column refers to those cases where multiple tracts (or even jurisdictions) were involved in a crime investigation. This entry has been dropped from the data set due to its small size and to its presence in only the last two years.

The police statistical reports from each year should be examined for possible changes in the definitions of certain crimes or other changes in the reporting conventions utilized in order to ensure proper usage of the data.

3.3 Building Permits

The building permit data was gathered directly by UCSUR from the ledgers kept by the Bureau of Building Inspections (BBI) of the City of Pittsburgh. This data is available in the NIS annually for the period from January 1, 1981 to May 31, 1989.¹¹

The data in the ledgers maintained by the BBI is recorded by date (year, month, and day), street address, and the ward the construction is located in. The ledgers also list a dollar value and a description of the construction. The description reveals whether it is a new construction, renovation, or an addition to an existing structure. The NIS includes data on only the new constructions.

The following pieces of information were copied from the BBI ledgers: (1) ward, (2) year, (3) month, (4) street address, (5) item of construction (e.g. garage in dwelling, office building, etc.), and (6) dollar value of construction. While the data is presented on a daily and monthly basis, for the purposes of the NIS data system, it has been aggregated to an annual basis. The street addresses were matched with census tract boundaries (with the help of maps and a 1970 census tract index) and 1970 tracts were determined for each address. Each new construction was classified as either residential or non-residential.¹² Data was then aggregated for each tract for each type of construction for each year.¹³

The primary limitation of the data set is that the 1970 census tract index was used to determine the tracts for the street addresses. As a few street addresses had changed, or been added since 1970, tracts had to be arbitrarily assigned to these addresses with the help of a map. While the first two digits of a tract denote the ward number, given the 1970/1980 differences in tracts, the ward numbers did not always offer a clue as to what tract the data actually belonged to. An error rate has been determined for the data based on the possible errors in tract classification. The error rate is calculated thus:

¹¹. The BBI did not have any files prior to 1981.

¹². Occasionally, arbitrary criteria were used to determine the residential or nonresidential nature of the construction. For example, construction of a garage in a dwelling is considered to be residential construction, yet when it is not stated as such (and the entry merely lists "Garage" with a construction value of \$30,000), then subjective judgments have to be used.

¹³. The file with the aggregated data is also available in a LOTUS data file with a format that enables further additions.

Total number of entries:	2156
Number of possible classification errors:	137
Error rate: $\{(100/2156)*137\}$	6.35 %