

CHAPTER 4. DESCRIPTION OF DATA FILES

4-A. STRUCTURE OF THE DATA FILES

BASIC STRUCTURE

The 1995 NPTS Public Use Data are organized into six different data files, which are available to users in SAS, ASCII, or DBF formats. Exhibit 4.1 illustrates the structure of the six files, with a description of which data are included in each file, the applicable questionnaire sections, the record level, and the variables which are needed to uniquely identify a record (ID variables).

The file variables are identified by variable name in the SAS versions. For each file variable, the code book contains:

- the variable type & length
- whether it was a variable on the 1990 NPTS dataset
- the label, which is a brief description of the variable
- the section and item number of the questionnaire or other source of the data
- value ranges and special codes
- the frequency of responses for each value or code shown
- comments, as necessary

The variables in the ASCII files are on the file in the following order:

- ID and weight variables, followed by
- question response variables in questionnaire order, and ending with
- variables used to describe the geography, stratification variables, date of interview variables, and derived variables.

See **Appendix I** for the lists of ASCII variables, including the starting position and length of each variable and their order on the NPTS files.

Users should be aware that the ID variables TRPNUM (for travel day trips) and TRIPNUM (for travel period trips) are two different variables. Also, data for all of the travel day trips, including segmented trips, are included in the travel day trip file. More detail about the segmented trips is included in the segmented

travel day trip file.

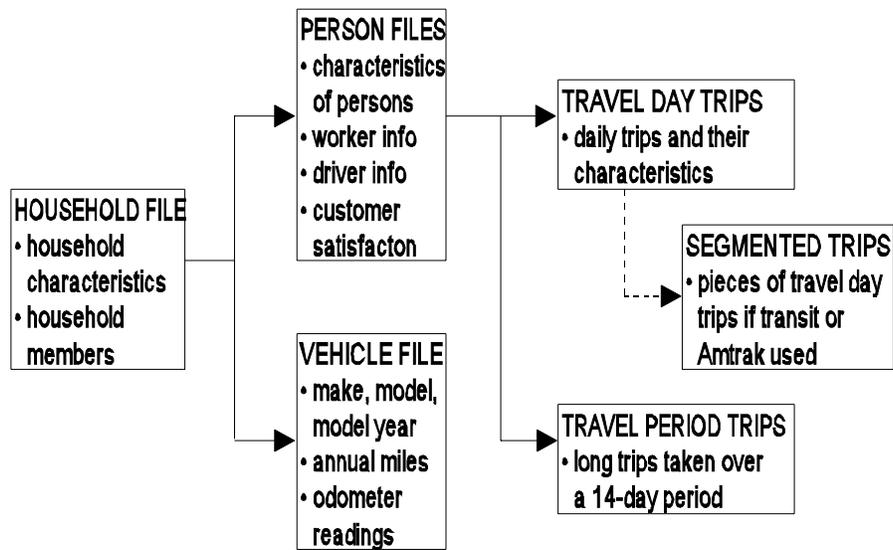
Exhibit 4.1 - Structure of 1995 NPTS Data Files

Data Files	Information Included	Record Level	ID Variables
Household file	Data unique to a household, or questions asked once for each sample household. Questions from interview sections: C -Home and Neighborhood D -Person Data J - Household Location, and K -Household Income	One record per household	HOUSEID
Person file	Data determined once for each completed person interview. Questions from interview sections: E - Driver Info. & Customer Eval. F - Education & Travel to Work I - Income of Persons not included in Household Income	One record per person	HOUSEID and PERSONID
Vehicle file	Data relating to each of the household's vehicles. Questions from interview section: B - Vehicle Data	One record per vehicle	HOUSEID and VEHID
Travel day trip file	Data about each trip the person made on the household's randomly-assigned travel day. Questions from interview section: G - Travel Day	One record per travel day trip	HOUSEID, PERSONID, and TRPNUM
Segmented travel day trip file	Data for up to 4 segments of each segmented travel day trip the person made on travel day. Based on responses to questions 28-30 and other questions of interview section G - Travel Day	One record per segmented travel day trip.	HOUSEID, PERSONID, and TRPNUM

Travel period file	Data that is asked once for every trip of at least 75 miles one way that the person took during a 14-day period ending on travel day. Questions from interview section: H - Travel Period	One record per travel period (14 days) trip.	HOUSEID, PERSONID, and TRIPNUM
---------------------------	---	--	--------------------------------

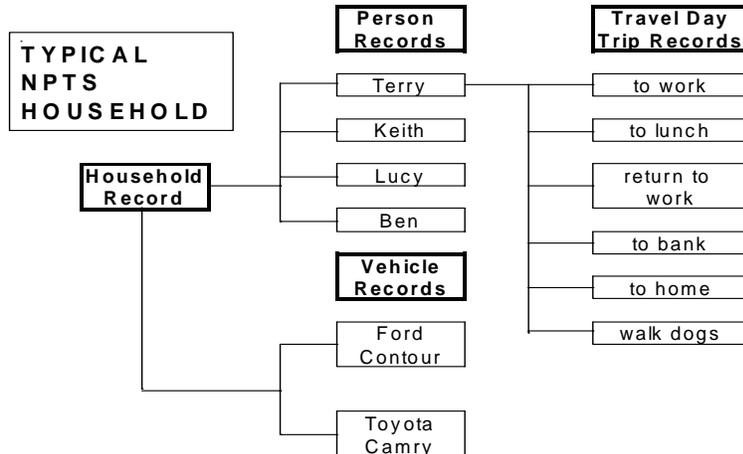
4-B. RELATIONSHIP BETWEEN THE SIX NPTS DATA FILES

The chart below depicts the six NPTS data files and their relationship.



**TYPICAL
NPTS
HOUSEHOLD**

The next chart shows how the records would appear for the data reported by the Typical NPTS Household example introduced in Chapters 1 and 2. Remember that this household reported only a portion of what would have been reported in an actual NPTS interview.



- NOTES:**
- This follows the Typical NPTS Household material in Chapter 2. In a real household, there would probably be trips by each household member.
 - Terry had no long trips, so there is no Travel Period file for her
 - Terry had no segmented trips, so there are no Segmented trip files for her

**TRAVEL DAY
AND TRAVEL
PERIOD TRIPS**

These two sections of the questionnaire are designed to complement each other. When the data from the two sections are combined, a more complete picture of personal travel is obtained.

In the travel day section, the respondent is asked to report all trips of any length during the 24-hour period designated as their "travel day." The travel day is designed to collect the types of trips typically made on a daily basis, such as trips to work, to the store, running errands, and visiting friends.

Because people make longer trips less frequently, respondents are asked to report any long trips, defined as 75 miles or more one-way, taken over a two week period. This is known as the travel period.

Once the travel day is designated for a household, the travel period becomes the thirteen days preceding the travel day plus the travel day. Because the travel day is also included in the travel period, if the respondent took a long trip on travel day, this would be reported in both trip sections of the questionnaire. These trips, which are called "overlap" trips, must be subtracted from travel day data when the user combines travel day and travel period. To do this, omit the trips on travel day that have the a "yes" response (code=01) in the OVERLAP variable.

SEGMENTED TRIPS

In the 1995 NPTS, as in the 1990, certain trips were "segmented", that is, they were broken into their component parts. A trip was segmented if both of these conditions were met:

- there was a change of vehicle or a change of mode on the trips, AND
- one of the modes used was a public transit mode or Amtrak. Public transit modes include bus, subway, elevated rail, commuter train, streetcar or trolley car.

This was done to collect more complete data on multi-modal trips, with particular emphasis on the use of public transit.

There was a limit of four segments per trip, and the typical travel day trip information was collected, along with the mode, start time and duration of each of the segments.

TRIPS NOT SEGMENTED

Trips in which the respondent went from one private vehicle (POV) to another were not segmented. For example, the respondent drives his sport utility vehicle to the pickup point for his carpool, then rides to work in the car of another carpool member. These POV-to-POV transfers were not segmented because they would have added to the respondent burden in reporting travel day trips, without an offsetting value in improving

our understanding of travel behavior. When more than one POV was used for a trip, the travel mode was assigned to the vehicle type used for the longest distance. In the example above, the car was probably the mode used for the longest distance.

RELATIONSHIP OF TRAVEL DAY TRIPS AND SEGMENTED TRIPS

Segmented trips are a small subset of the universe of NPTS travel day trips. Of the 409,025 travel day trips collected in the 1995 NPTS, 3,779 or less than one percent, are segmented. When a travel day trip is segmented, most of the trip information is on the travel day trip record. In addition, a segmented trip record is established on the segmented trip file. This segmented record, which has the same ID variables as the travel day trip record, contains the unique information on each of the segments of the trip, such as the mode, start time and duration in minutes. Even though a trip can have up to four segments, there is only one segmented trip record established. The data for each segment is listed in variables with names like SEG1_MIN, SEG2_MIN, SEG3_MIN, etc. to accommodate the characteristics of up to four segments in one record.

WHEN IS A RECORD ON THE FILE

The purpose of this subsection is to present information on the NPTS file structure that has confused data users in the past.

Household Record - There is one record for each household in the dataset, also called a "useable" household.

Vehicle Record - There is a vehicle record for each vehicle owned by or available to a useable household. If the household has no vehicles, there will not be any vehicle records. The number of household vehicles, including zero vehicles, is available on the household record in the variable, HHVEHCNT.

Person Record - There is a person record for each **interviewed** person in the household. For example, a household consists of three people, Tom, Dick & Harry. Tom and Dick were interviewed for the NPTS, but Harry was never available, despite repeated attempts. There will be a person record for Tom and one for Dick. No person record will exist for Harry, but his characteristics will be available to the analyst on the household file (see **HOUSEHOLD**

MEMBER VARIABLES below.)

Travel Day Trip Record - There is a trip record for each trip taken by an interviewed person in a useable household. So, in our example above, if Tom makes six trips, there will be six travel day trip records on the file. Suppose Dick was ill and stayed home all day. There are no travel day trips records for Dick, however, there is a person record for him, since he was interviewed. The person file variable, SAMEPLC, i.e. "stayed in the same place all day?", will confirm that Dick was interviewed for travel day and reported no trips. No travel day trip records will exist for Harry, since he was not interviewed. Likewise, there will be no person file record for Harry.

In earlier NPTSs, before "stayed in same place all day?" was asked, data users assumed that the lack of a travel day trip record for Dick meant that he was not interviewed for his travel day travel. This is not true. If there is a person record for that person, they were interviewed for travel day. Note that about 12 percent of the 1995 NPTS respondents reported no travel day trips. While some of these non-travelling people may be "soft refusals" who did not want to bother reporting their trips, many of them are legitimate non-travellers. Remember that the NPTS travel days encompass all 365 days of the year, including holidays and weekends.

Segmented Trip Record - A segmented trip record will be present only when a travel day trip meets the two conditions for segmented treatment (see **SEGMENTED TRIPS** discussion above). In our example, assuming that none of Tom's six trips met those conditions, there would be no segmented trip records for him.

Travel Period Trip Record - A record is present only when a qualifying trip was made by the respondent. Thus, if Dick, who was sick on travel day, had made a 250-mile trip the week before travel day, there would a travel period trip record for him. Because little detail is collected on the long trips in the travel period section, there is only one record for each roundtrip.

HOUSEHOLD MEMBER VARIABLES

For the 1995 NPTS, the characteristics of all household members, whether interviewed or not, are available on the Household File. These characteristics were included to allow the user to address a number of travel behavior and survey method research issues. The characteristics are contained in the variables starting with P1 through P10 and, for each household member, the information includes:

- age (P1_AGE, P2_AGE, etc.)
- sex (P1_SEX, etc.)
- relationship to household respondent (P1_RELAT, etc.)
- driver status (P1_DRVR, etc.)
- worker status (P1_WKR, etc.) and
- response status, i.e., eligible or not eligible, interviewed or not interviewed, self interview or proxy (P1_STAT, etc.).

In earlier NPTSs it was difficult to obtain a complete picture of the household members, because a person record is only on the datafile when a household member is interviewed for the survey. The characteristics of all household members were gathered in the household interview, but in the 1990 NPTS the characteristics of those household members not interviewed were discarded. This caused some limitations on the analyses that could be performed, so it was decided to retain characteristics of all household members in the 1995 NPTS.

4-C. CODEBOOK

CODEBOOK FORMAT

The documentation includes a codebook, with sections for each of the data files. The codebook contains critical user information about each variable in each of the files. The codebook is arranged in a two-page format, with the variables in Exhibit 4.2 beginning on the left-hand side and continuing across the two facing pages. Exhibit 4.2 lists the items that correspond to the codebook columns, along with a brief description of the contents of each column.

CODEBOOK EXAMPLE

As an example, the third column of Exhibit 4.2 shows the codebook information for the variable named BUS_DIST.

- It is a numeric variable of width 5 including the decimal point (up to 3 digits before the decimal and one after).
- This question was not asked in the 1990 NPTS.
- This variable contains the distance in miles from the home to the nearest bus stop, reported in response to item 2.1 of questionnaire section C.
- The value range and the frequencies show that the file contains 26,160 reports ranging from 0 to 100 miles; that 1,245 household respondents said they could not ascertain the distance, and 15 refused to answer the question. It also shows that the question was legitimately skipped in the 14,613 households in which the household respondent answered no or don't know to question C-1 "Is local bus service available in your town or city?"
- The comment for this variable tells the user that the responses in blocks have been converted to miles using a factor of 9 blocks per mile.

Exhibit 4.2 - Contents of the 1995 NPTS Code Books

Column Heading	Description of Contents	Example Variable (from Household File)
Target Variable	The variable name	BUS_DIST
Variable Type	C = character; N = numeric	N
Width	Maximum variable length	5.1
1990 Variable Name	S = same name in 1990 NPTS N = new variable in 1995 NPTS * = variable values external to the survey	N
Variable Label	Short description of the variable	Distance to bus (miles)
Section	Source section(s) of the questionnaire	C
Item ID	Source item(s) in the questionnaire section	2.1
Value Range & Codes	Either lists all possible values of the variable, a range of the values, or a combination of the two	(0 - 100) 994 = Legitimate skip 998 = Not ascertained 999 = Refused
Frequencies	Shows the number of records in the file for each listed value	0-100 = 26,160 994 = 14,613 998 = 1,245 999 = 15
Comments	Gives additional information to users, or refers to relevant discussion in other sections of the documentation	Miles as reported, blocks converted to miles (9/mile)

**COMPAR-
ABILITY
WITH 1990
NPTS**

Emphasis was placed on making the 1995 NPTS data files comparable with the 1990 NPTS data files.

- To the extent possible, the same variable names as in 1990 were used for variables based upon the same information. In cases where the information is basically the same as 1990, but it was asked in a slightly different way or context, the similar 1990 variable name will be listed in the codebook column labeled "1990 variable".
- The same general scheme was used again for legitimate skip, not ascertained and refusal codes.
- The documentation in this volume is intended to cover at least the same content as the "1990 NPTS User's Guide for the Public Use Tapes", (Publication FHWA-PL-92-007).

4-D. VARIABLES REPEATED

**REPEATED
VARIABLES**

In addition to the information specific to its file (e.g., the travel day file contains data on the individual travel day trips), each of the six files includes variables from other files to be used along with its own variables. This is done for the convenience of the data user, to minimize the need to merge data from multiple files. Although this format is less desirable from a data storage standpoint, it significantly simplifies subsequent data manipulation.

**HOUSEHOLD
LEVEL
REPEATED
VARIABLES**

The following commonly used variables are included in all six data files:

VARIABLE DESCRIPTION	VARIABLE NAME
Census Division	CENSUS_D
Census Region	CENSUS_R
Number of household drivers	DRVRCNT
CMSA of household (Consolidated Metropolitan Statistical Area)	HHCMSA

Household family income category	HHFAMINC
MSA of household (Metropolitan Statistical Area)	HHMSA
Number of household members	HHSIZE
Number of household vehicles	HHVEHCNT
Hispanic status of household reference person	HH_HISP
Race of household reference person	HH_RACE
Household life cycle	LIF_CYC
Population size of MSA	MSASIZE
Presence or absence of rapid rail (i.e., subway, elevated rail)	RAIL
Substratum within major stratum for low-density or high-density residential phone numbers	SUBSTRAT
Travel day month (May 1995 through June 1996)	TDAY_MON
Travel day year	TDAY_YR
Major sample stratum	VARSTRAT
Number of household workers	WRKCOUNT.

BLOCK GROUP REPEATED VARIABLES -

Four of the variables that describe the block group of the interviewed household are also repeated on the other files (except the segmented trip file). These four variables are:

VARIABLE DESCRIPTION	VARIABLE NAME
Median household income, block group	HBHINMED
Housing unit density, block group	HBHRES DN
Urban/rural code, block group	HBHUR
Population density, block group	HBPPOPDN

PERSON LEVEL REPEATED VARIABLES

There are a few person-level variables that are repeated on the three trip files (travel day, travel period, and segmented trips.). These are:

VARIABLE DESCRIPTION	VARIABLE NAME
Whether respondent is a driver	DRIVER
Was this a proxy interview	PROXY
Respondent's age	R_AGE

Respondent's sex	R_SEX
Whether respondent is a worker	WORKER.

4- E. VARIABLES ADDED

ADDED VARIABLES

An added variable is an item on the dataset that is not a response to a question in the interview. Numerous variables were developed and added to the data base, including:

- summary variables to aid data analysis,
- external variables to describe the geographic area surrounding the respondents' household and work locations, and
- flag variables to identify data records that have been edited.

HOUSEHOLD LEVEL

Common-required variables were calculated and included on the data files so they would not need to be constructed each time they were needed. The variables that are repeated on all six files are indicated with an asterisk.

VARIABLE DESCRIPTION	VARIABLE NAME
* Number of drivers in the household	DRVRCNT
Number of eligible household members	HHELGCNT
* Total number of persons in the household	HHSIZE
* Number of household vehicles	HHVEHCNT
Number of household members under 5 years of age	HH_0TO4
Number of household members not eligible for NPTS (e.g., under 5 years of age, determined not to reside in the household, or incapable of being interviewed)	INELGCNT
* Life cycle of the household	LIF_CYC
Variable indicating non-family income reported in the person file	NONFMFLG
Number of person interviews completed for the household	RESP_CNT

Day of week for the household's travel day	TRAVDAY
* Number of workers in the household	WRKCNT.

**TRACT &
BLOCK
GROUP
CHARACTER-
ISTICS**

A number of geographically-based variables obtained from Claritas, Inc. were added to the database. These variables are based on Census tract or block group level projections of 1990 Census data to 1995. They provide the data user with characteristics of the respondent's neighborhood, which can supplement to the data collected on the respondent's household.

The tract and block group were identified by geocoding the reported home and work addresses from the survey. The addresses used to geocode the home and workplace locations were removed from the dataset for confidentiality reasons.

All of the household level variables are on the Household file, and the workplace variables are on the Person file. Four of these variables were repeated on all files except the Segmented Trip file (see **BLOCK GROUP REPEATED VARIABLES** above).

Appendix L contains more information on the tract and block group variables.

TRAVEL DAY

The derived variables added to the travel day file are:

VARIABLE DESCRIPTION	VARIABLE NAME
Whether the trip began during AM or PM hours	DAYNIGHT
Difference in days between the household travel day and the person interview date	DIFFDATE
The number of minutes spent at destination of previous trip	DWELTIME
Total number of persons on the travel day trip	NUMONTRP
Variable identifying travel day	

POV trips for which the respondent was the driver VTR_FLG

TRAVEL PERIOD

There are two derived variables added to this file:

VARIABLE DESCRIPTION	VARIABLE NAME
Straight line distance of the travel period trip, based on household location and reported trip destination	CALCDIST
Imputed variable identifying the driver of a travel period trip	DRVR_TRP.

1990 TRIP PURPOSES

The trip purpose definitions for the 1995 NPTS differed from those used in the 1990 NPTS. In addition to the 1995 trip purpose, each trip was recoded into the variable WHYTRP90 to mimic the 1990 NPTS trip purpose definitions.

The 1995 trip purposes use a "from-to" format, while the 1990 purposes were based on coding a "main reason" for the trip. As a result, the trip purpose codes used in 1995 differed from the 1990 trip purposes in the following ways:

- Returning home is a 1995 trip purpose but was not a 1990 NPTS trip purpose. In 1990, the trip purpose was assigned to the activity that was the main reason the person was away from home.
- In 1990, if one of the reasons was work, the return trip home was assigned a work purpose, even if there were incidental trips made on the way home.
- In 1990, if there were multiple purposes for being away from home and work was not one of them, the respondent was asked main reason for the trips. Because this "main reason" format was not used in the 1995 survey, when the 1995 purposes were recoded to the 1990 scheme, the activity the person spent the most time at while away from home was assigned as the main purpose for the return trip home. The variable, DWELTIME, was created to determine this.

The recoded 1990 trip purposes will be particularly useful for analyses comparing the 1990 and 1995 data by purpose. See **Appendix M** for more detail on trip purposes and trip purpose variables on the 1995 dataset.

TRIP CHAINING

Part of the recoding of trip purposes to the 1990 purpose involved creating trip chains. For this purpose, the chains were defined by trips ending at home, work or someplace else.

There are several derived variables on the Travel Day file developed to define trip chains. The variable CHAIN file indexes the trip chains defined for each a person's travel day. Each trip reported for a respondent was assigned to a "chain", after ordering the person's travel day trips by STRTTIME from 4:00 am to 3:59 am. Trips with missing STRTTIME values were sorted to the beginning of the list. All trips within a chain are sequentially numbered in the variable CHAINTRP. Variables TRPNUM_A and TRPNUM_B identify the first and last trips in each chain. The variables FROM_A and TO_B identify the origin and destination of the chains in terms of home, work or someplace else (H, W, or S).

Some of these chains do not begin or end at either home or work, as some respondents did not take such trips. Also, some persons reported only a single trip on the travel day, such as returning home from vacation. It is possible to select a subset of chains that are anchored by home and work using FROM_A and TO_B. Note that some trip chains involve only one or two trips, which might exclude them from other types of trip chaining analyses.