
APPENDIX D: DESCRIPTION OF DATA

APPENDIX D

DESCRIPTION OF DATA

STRUCTURE OF THE DATA FILES

BASIC STRUCTURE

The EIA public-use data are organized into two different data files, which are available to users in two formats: Microsoft Excel and ASCII (comma-separated values) on http://www.eia.doe.gov/emeu/rtecs/nhts_survey/2001/index.html#pubuse. Table D1 illustrates the structure of the two files, with a description of which data are included in each file, the record level, and the variables that are needed to uniquely identify a record (i.e., Identifier or ID variable).

For each data variable, the codebook contains:

- a variable name,
- a code(s) associated with the variable and code descriptions,
- a variable label.

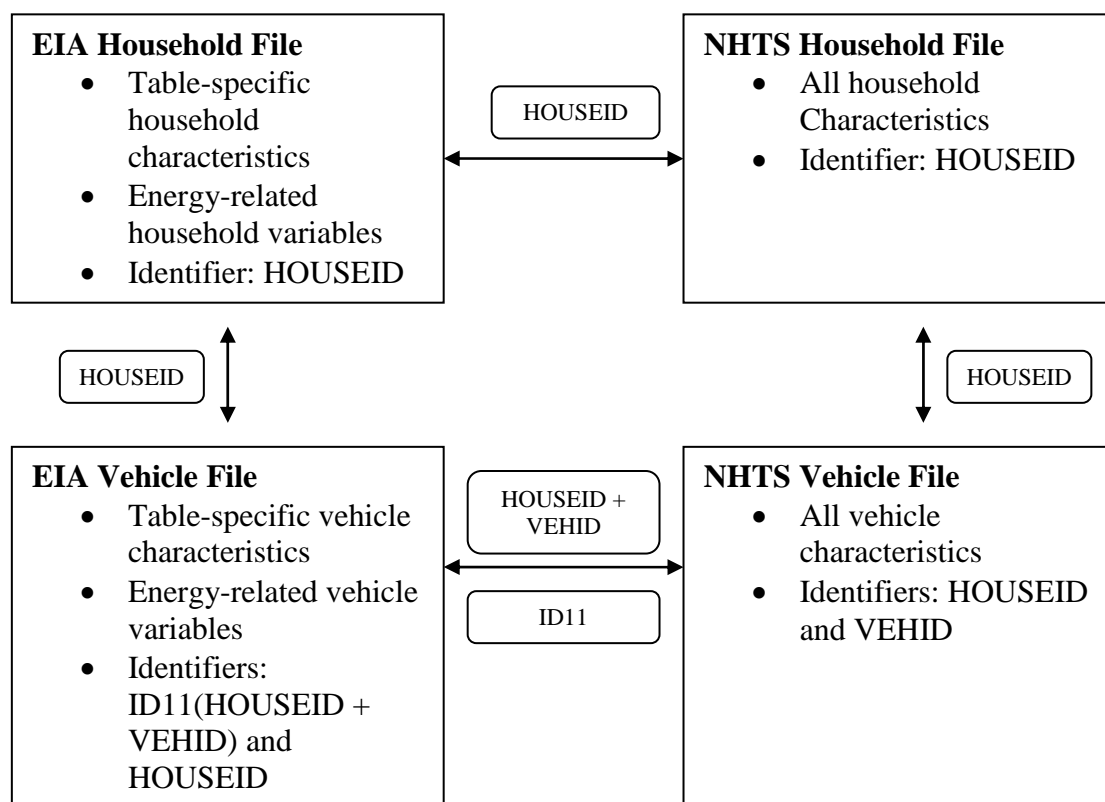
Table D1. Data Files

Data File	Information Included	Observation Level	ID Variable	National Sample Weight
EIA Household File	Energy-related data for households	One record per household	HOUSEID HOUSEID VEHID ID11	EXPFLHHN
EIA Vehicle File	Energy-related data for vehicles	One record per vehicle	(HOUSEID + VEHID)	EXPFLHHN

RELATIONSHIP AMONG U.S. DOT AND EIA PUBLIC-USE DATA FILES

The figure below depicts the EIA and the January 2004 release of the 2001 NHTS data, as well as their inter-relationships.

Figure D1. Relationships among Data Files Released by EIA and NHTS



UNDERSTANDING THE DATA FILES

The purpose of this section is to present information about the EIA household and vehicle file to data users. Based on a recommendation from U.S. DOT, only “100-percent households” and their light-duty vehicles are contained in EIA’s public-use files; hence, care should be taken when combining data from NHTS with EIA’s augmented energy and energy-related data to avoid any misinterpretations.

The following files are available as public-use data:

- **EIA Household File** contains table-specific household characteristics and energy-related household variables for 21,178 households. **Important:** To calculate results comparable to those found in this report, data users must exclude in their tabulations those households that do not have light-duty vehicles. EIA has included a variable, HHVEHLC, for that purpose.
- **EIA Vehicle File** contains table-specific vehicle characteristics and energy-related vehicle variables for 42,736 light-duty vehicles that EIA used to calculate the number and type of light-duty passenger vehicles, annual mileage, type of fuel used, and price paid for fuel, all of which are required to measure the consumption and expenditure associated with the nation’s residential travel.

- **NHTS Household File** contains all the households data collected during by the NHTS. To identify the households in the national sample of “100-percent household” having at least one light-duty passenger vehicle (passenger cars, vans, sport utility vehicles, pickup trucks and recreational vehicles), use the VEHTYPE and SMPLSRCE variables found on the **NHTS Vehicle File**.
- **NHTS Vehicle File** contains all vehicles and vehicle data collected from NHTS survey respondents.

CODEBOOK FOR EIA AUGMENTATIONS

This section contains information on the variables in each of the two EIA data files: household and vehicle data files. The codebook has been included first for household variables, followed by the codebook for the vehicle variables.

For each variable the codebook provides:

- variable name, this is the variable name used in the 2001 NHTS;
- data level, either household- or vehicle-level data file;
- type, numeric or character (alphanumeric) data
- codes and descriptions, a description of every possible code identified with a variable; and,
- label, a brief description of a variable.

Table D2. Codebook for Public-Use Data Produced by EIA

	Variable Name	Data Level	Type	Codes and Descriptions	Label
1	AVGVEHS	Household	N	Numeric values	Number of vehicles in a household, expressed as vehicle-years, AVGVEHS
2	CENSUS_D	Household	C	"1" - New England "2" - Middle Atlantic "3" - East North Central "4" - West North Central "5" - South Atlantic "6" - East South Central "7" - West South Central "8" - Mountain "9" - Pacific	Household Census Division, CENSUS_D

Table D2. Codebook for Public-Use Data Produced by EIA

Variable Name	Data Level	Type	Codes and Descriptions	Label
3 CENSUS_R	Household	C	"1" - Northeast "2" - North Central "3" - South "4" - West	Household Census Region, CENSUS_R
4 DRVRCNTC	Household	N	1 - 1 driver 2 - 2 drivers 3 - 3 drivers 4 - 4 or more drivers	EIA Number of drivers, DRVRCNTC
5 EHHWT#	Household	N	Replicate weight values	Replicate weight_#, EHHWT#
6 EXPFLHHN	Household	N	National sample weight values	HH Weight-100% completed - NATL, EXPFLHHN
7 HHCHILD	Household	N	0 - Households without children 1 - Households with children	EIA Household has children, HHCHILD
8 HHCOMP	Household	N	1 - Under 7 year, age of oldest child, household with children 2 - 7 to 15 years, age of oldest child, household with children 3 - 16 to 17 years, age of oldest child, household with children 4 - Under 35 years, age of householder, one adult with no children 5 - 35 to 59 years, age of householder, one adult with no children 6 - 60 years or more, age of householder, one adult with no children 7 - Under 35 years, age of householder, two or more adults with no children 8 - 35 to 59 years, age of householder, two or more adults with no children 9 - 60 years or more, age of householder, two or more adults with no children	EIA Household composition, HHCOMP

Table D2. Codebook for Public-Use Data Produced by EIA

Variable Name	Data Level	Type	Codes and Descriptions	Label	
9	HHINCOME	Household	N	1 - Less than \$5,000 2 - \$5,000 to \$9,999 3 - \$10,000 to \$14,999 4 - \$15,000 to \$19,999 5 - \$20,000 to \$24,999 6 - \$25,000 to \$34,999 7 - \$35,000 to \$49,999 8 - \$50,000 to \$74,999 9 - \$75,000 or more -9 - Don't know	EIA Family income, HHINCOME
10	HHR_HISP	Household	C	"1" - Yes "2" - No	Hispanic status of HH respondent, HHR_HISP
11	HHR_RACEC	Household	N	1 - White 2 - Black 3 - Other	EIA Race of householder, HHR_RACEC
12	HHR_SEX	Household	C	"1" - Male "2" - Female	Gender of HH respondent, HHR_SEX
13	HHSIZEC	Household	N	1 - One person 2 - Two persons 3 - Three persons 4 - Four persons 5 - Five persons 6 - Six or more persons	EIA Household size, HHSIZEC
14	HHVEHLC	Household	N	Numeric Values	Filter variable for calculating estimates and std errors for household statistics, HHVEHLC
15	HOUSEID	Household	C	Character Values	Household Identification Number, HOUSEID
16	LIF_CYC	Household	C	"01" - One adult, no children "02" - Two plus adults, no children "03" - One adult, youngest child 0-5 "04" - Two plus adults, youngest child 0-5 "05" - one adult, youngest child 6-15 "06" - Two plus adults, youngest child 6-15 "07" - One adult, youngest child 16-21	HH Life Cycle, LIF_CYC

Table D2. Codebook for Public-Use Data Produced by EIA

Variable Name	Data Level	Type	Codes and Descriptions	Label
			"08" - Two plus adults, youngest child 16-21 "09" - One adult, retired, no children "10" - Two plus adults, retired, no children	
17 PDRVAGEC	Household	N	1 - 16 to 17 years 2 - 18 to 22 years 3 - 23 to 29 years 4 - 30 to 39 years 5 - 40 to 49 years 6 - 50 to 59 years 7 - 60 to 69 years 8 - 70 to 79 years 9 - 80 years and over -9 - Don't know	EIA Age of primary driver, PDRVAGEC
18 PERCOSTC	Household	N	1 - Less than 2 2 - 2 to 3.9 3 - 4 to 5.9 4 - 6 to 7.9 5 - 8 or more	EIA Percent of income for motor fuel, PERCOSTC
19 POOR100	Household	N	Numeric values	Households with incomes below 100 percent of the poverty level, POOR100
20 POOR125	Household	N	Numeric values	Households with incomes between 100 to 150 percent of the poverty level, POOR125
21 POOR150	Household	N	Numeric values	Households with incomes above 150 percent of the poverty level, POOR150
22 POORC	Household	N	1 - Below 100 percent 2 - 100 to 150 percent 3 - Above 150 percent -9 - Don't know	EIA Income relative to poverty line, POORC

Table D2. Codebook for Public-Use Data Produced by EIA

Variable Name	Data Level	Type	Codes and Descriptions	Label	
23	SMPLSRCE	Household	C	"01" - National Sample "02" - Baltimore Add-on "03" - Des Moines Add-on "04" - Hawaii Add-on "05" - Kentucky Add-on "06" - Lancaster PA Add-on "07" - New York Add-on "08" - Oahu Add-on "09" - Texas Add-on "10" - Wisconsin Add-on	Sample where the case originated, SMPLSRCE
24	TOTBTUSINK	Household	N	Numeric values	Total household fuel consumed in units of 1000 British Thermal Units (BTUs), TOTBTUSINK
25	TOTCOST	Household	N	Numeric values	Household Expenditures, TOTCOST
26	TOTCOSTC	Household	N	1 - \$500 or less 2 - \$501 to \$1,000 3 - \$1001 to \$1,500 4 - \$1,501 to \$2,000 5 - \$2,001 or more	EIA Motor fuel expenditures, TOTCOSTC
27	TOTGALS	Household	N	Numeric values	Household gallon of gasoline equivalent, TOTGALS
28	TOTMILES	Household	N	Numeric values	Household vehicle miles traveled, TOTMILES
29	URBRUR	Household	C	"1" - Urban "2" - Rural	Household in urban/rural area, URBRUR
1	BTUSINK	Vehicle	N	Numeric values	Fuel consumed in units of 1000 British Thermal Units (BTUs), BTUSINK
2	COST	Vehicle	N	Numeric values	Motor fuel expenditures, COST
3	EHHWT#	Vehicle	N	Replicates sample weight values	Replicate weight_#, EHHWT#
4	EIADMPG	Vehicle	N	Numeric values	EIA derived miles per equivalent-gallon, EIADMPG

Table D2. Codebook for Public-Use Data Produced by EIA

Variable Name	Data Level	Type	Codes and Descriptions	Label	
5	EIADMPGC	Vehicle	N	1 - 10.9 or less 2 - 11 to 12.9 3 - 13 to 15.9 4 - 16 to 18.9 5 - 19 to 21.9 6 - 22 to 24.9 7 - 25 to 29.9 8 - 30 or More	EIA Fuel economy, EIADMPGC
6	EPATMPG	Vehicle	N	Numeric values	Unadjusted 55/45 combined fuel economy, EPATMPG
7	EPATMPGF	Vehicle	C	"10#" - represents a NHTS sample vehicle that had a single model name "matching" to eligible NHTSA file records using four linking variables: vehicle manufacturer, vehicle model, vehicle model year, and vehicle type. "20#" - represents a NHTS sample vehicle that had multiple model names "matching" to eligible NHTSA file records using four linking variables: vehicle manufacturer, vehicle model, vehicle model year, and vehicle type. "30#" - represents a NHTS sample vehicle that had a single model name "matching" to eligible NHTSA file records using three linking variables: vehicle manufacturer, vehicle model, and vehicle model year. "40#" - represents a NHTS sample vehicle that had multiple model names "matching" to eligible NHTSA file records using three linking variables: vehicle manufacturer, vehicle model, and vehicle model year.	Imputation flag for EPATMPG variable, EPATMPGF

Table D2. Codebook for Public-Use Data Produced by EIA

Variable Name	Data Level	Type	Codes and Descriptions	Label
			<p>"50#" - represents a NHTS sample vehicle that "matched" to eligible NHTSA file records using three linking variables: vehicle manufacturer, vehicle type, and vehicle model year.</p>	
			<p>"600" - represents a "match" based on EIA expert analysis using subject matter experience, in conjunction with past RTECS. Additionally, this imputation flag value represents recreation vehicles (VEHTYPE = "06"), where MPG(EPA 55/45) has been fixed at a fuel economy rates based on the U.S. Department of Transportation, National Transportation Statistics 2000.</p>	
			<p>"001" - denotes a NHTS sample vehicle that was internally hot-decked to match with a median value of the Composite fuel economy defined by categories of one or more vehicle characteristics, such as make, model, model year, and vehicle type. None of these imputed values were subsequently re-used in the actual hot-decking procedure of median values.</p>	
			<p>"999" - represents an imputation flag where no eligible NHTSA file records were found to "match" a NHTS sample vehicle.</p>	
<p>* = For EPATMPGF codes, # is a number between 0 and 5. This number, #, represents a year increment. Due to the errors in respondents reporting accurate model year or, to a lesser extent, due to deficiencies in the NHTSA files, it was necessary to incrementally increase or decrease (not simultaneously increase and decrease) the model year for "matching" to successively larger range of years. If, for example, an eligible match was not found for a NHTS sample vehicle having the following attributes: Volkswagen, Sirocco, 1990, Automobile. Toggling of model years, by a single year increase followed by a single year decrease of the reported model year, resulted in a match with a Volkswagen, Sirocco, 1988, Automobile. In this example, the Volkswagen, Sirocco, 1990, Automobile, while seemingly a respondent reporting error, would receive an imputation flag of</p>				

Table D2. Codebook for Public-Use Data Produced by EIA

Variable Name	Data Level	Type	Codes and Descriptions	Label	
"102" due to the "match" with the NHTSA file record corresponding to a Volkswagen, Sirocco, 1988, Automobile.					
8	EXPFLHHN	Vehicle	N	Numeric national weight values	HH Weight-100% completed - NATL, EXPFLHHN
9	FUELTYPE	Vehicle	N	1 - Diesel 2 - Natural gas 3 - Electricity 4 - Gasoline	Type of transportation fuel, FUELTYPE
10	GALS	Vehicle	N	Numeric values	Motor fuel consumption, GALS
11	ID11	Vehicle	C	Character Values	HOUSEID plus VEHID, ID11
12	MILES	Vehicle	N	Numeric Values	Vehicle miles traveled, MILES
13	SMPLSRCE	Vehicle	C	"01" - National Sample "02" - Baltimore Add-on "03" - Des Moines Add-on "04" - Hawaii Add-on "05" - Kentucky Add-on "06" - Lancaster PA Add-on "07" - New York Add-on "08" - Oahu Add-on "09" - Texas Add-on "10" - Wisconsin Add-on	Sample where the case originated, SMPLSRCE
14	TVEHSHARE	Vehicle	N	Numeric values (between 0 and 1)	Prorated fraction of possession of vehicle, in vehicle-year, TVEHSHARE
15	VEHID	Vehicle	C	Character values	Vehicle ID number, VEHID
16	VEHTYPE	Vehicle	C	"01" - Passenger car "02" - Vans (Large and Minivans) "03" - Sport Utility Vehicle "04" - Pickup truck "06" - Recreational vehicle	Type of vehicle, VEHTYPE

Table D2. Codebook for Public-Use Data Produced by EIA

Variable Name	Data Level	Type	Codes and Descriptions	Label
17 VEHYEARC	Vehicle	N	1 - 2001 to 2002 2 - 2000 3 - 1999 4 - 1998 5 - 1997 6 - 1996 7 - 1995 8 - 1992 to 1994 9 - 1989 to 1991 10 - 1986 to 1988 11 - 1983 to 1985 12 - 1980 to 1982 13 - 1977 to 1979 14 - 1976 or earlier -9 - Don't know	EIA Model year, VEHYEARC

Figure D2. Map of U.S. Census and Division Areas

