Data in DataFerrett

Table of Contents

 Administrative note: Much of this directory text is taken directly from the descriptions provided by the studies, which is why the "More Information" link is readily included. These sites also contain more data and possibly more current data than that from DataFerrett.

Terms

DataFerrett: A program that allows the public to search publicly available government data and download data sets.

Public Use Microdata Sample (PUMS): Computerized files containing a small sample of individual records, with identifying information removed, from the census long form and from the American Community Survey showing the population and housing characteristics of the people included on those forms. Provides more detailed data than other census files, but only for a small sample of the population.

Public Use Microdata Area (PUMA): An area that defines territory used for the PUMS data.

Full glossary of census terms available at http://factfinder.census.gov/home/en/epss/glossary_p.html

American Community Survey

General Description: On-going survey that shows what the American population looks like and how it lives, influencing how communities locate services and allocate resources.

Kinds of variables available: Age, sex, tenure, income, education, language spoken at home, journey to work, occupation, condominium status, shelter costs, vehicles available, etc.

Organization of Data: PUMS = **P**ublic **U**se **M**icrodata **S**ample
3-year estimates, Puerto Rico 3-year estimates, PUMS 1-year, Puerto Rico PUMS 1-year.

More information about data: http://www.census.gov/acs/www/

American Housing Survey

General Description: Data on housing for national and metropolitan areas.

Data Description: National data is collected in odd years and data for metropolitan areas is conducted biennially in even-numbered years, during the period March through November. The areas are divided into three groups, with a group rotating into sample every six years, with the schedule adjusting due to budget constraints. The national sample covers an average 55,000 housing units. Each metropolitan area sample covers 4,100 or more housing units. The AHS returns to the same housing units year after year to gather data; therefore, this survey is ideal for analyzing the flow of households through housing.

Kinds of variables available: Income, neighborhood and housing quality, housing costs, utilities cost, mortgage data, etc.

Organization of Data: Metropolitan Sample, National Sample, National Records for 1993 and 1995.

More information about data: http://www.census.gov/hhes/www/housing/ahs/overview.html

Common Core of Data (Education)

General Description: Data on education from NCES, National Center for Education Statistics.

Data Description: Annual national data on public elementary and secondary schools and local education agencies. The data is collected by survey.

Kinds of variables available: Information collected by schools such as name, address, and phone number; descriptive information about students and staff, including demographics; fiscal data, including revenues and current expenditures.

Organization of Data: One category: Public elementary/secondary universe data, variable categories include universe data and geographies.

More information about data: http://nces.ed.gov/ccd/

Consumer Expenditure Survey

General Description: Information on the spending activities of Americans.

Data Description: Buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics. The survey data is collected for the Bureau of Labor Statistics by the U.S. Census Bureau.

Kinds of variables available: Demographic data, composition of earners, resources used, rent, investment holdings (securities, savings accounts, bonds), various utilities used, etc.

Organization of Data: 1993 data, first, second, third, and fourth quarters.

More information about data: http://www.bls.gov/cex/, includes resources to get more data from this survey, up to 2007.

County Business Patterns

General Description: An annual series that provides subnational economic data by industry.

Data Description: Useful for studying the economic activity of small areas; analyzing economic changes over time; and as a benchmark for statistical series, surveys, and databases between economic censuses. Businesses use the data for analyzing market potential, measuring the effectiveness of sales and advertising programs, setting sales quotas, and developing budgets. Government agencies use the data for administration and planning. County Business Patterns covers most of the country`s economic activity. The series excludes data on self-employed individuals, employees of private households, railroad employees, agricultural production employees, and most government employees.

Kinds of variables available: Number of establishments (by size), payroll, number of employees, industry codes.

Organization of Data: County level, State level, and National level, annual data from 1998 to 2003.

More information about data: http://www.census.gov/econ/cbp/index.html

Current Population Survey

General Description: Large survey with a great deal of information on the population.

Data Description: The Current Population Survey (CPS) is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The survey has been conducted for more than 50 years. The survey covers many topics of interest about the population of the United States.

Kinds of variables available: Estimates obtained from the CPS include employment, unemployment, earnings, hours of work, and other indicators. Demographics including age, sex, race, marital status, and educational attainment, other variables include occupation, industry, and class of worker. Other information includes school enrollment, income, previous work experience, health, employee benefits, and work schedules.

Organization of data: Many categories as described in each title, from Basic to Fertility, to Veterans.

More information about data: http://www.census.gov/cps/

Decennial Census of Population and Housing

General Description: The commonly known "census" conducted every 10 years to decide the apportionment of seats in the House and to gather information on the U.S. population.

Data Description: There are 100% questions asked of every person and housing unit, generally demographic information. The census sample asks more detailed questions about population and housing including ancestry, occupation, farm residence, vehicles available, among other questions.

Kinds of variables available: Many variables on population and housing are available in the PUMS including place of work, poverty status, children in household, household income, mortgage status, and many others. There is a choice of the 1% sample and the 5% sample. Historical data includes variables like air conditioning, sewage disposal, times married, and other variables. Summary files sort the data for you into various relationships, like median age by sex or average household size.

Organization of data: Historical data, PUMS, Summary File 1, Summary File 3

More information about data:
http://factfinder.census.gov/jsp/saff/SAFFInfo.jsp?_pageId=sp4_decennial ,
http://www.census.gov/prod/www/abs/decennial/index.htm

Decennial PUMS

This category includes the same kind of data as detailed in the PUMS above, but for the 1990 census rather than the 2000 census.

Mortality

General Description: Collection of data on cause of death annually from 1991 to 1998.

Data Description: Collected for National Vital Statistics System. Data can be used to present the characteristics of those dying in the United States, to determine life expectancy, and to compare mortality trends with other countries.

Kinds of variables available: Month of death, occupation, location, cause of death.

Organization of data: Underlying Cause of Death.

More information about data: http://www.cdc.gov/nchs/deaths.htm

National Ambulatory Care Survey

General Description: Health data collected annually from office-based physicians on ambulatory patients.

Data Description: The National Center for Health Statistics (NCHS) selects an annual, national sample of approximately 3,000 office-based physicians whose offices are located in a subset of the 1980 design National Health Interview Survey primary sampling units. The sample is divided into weekly subsamples of approximately 67 physicians. Sampled physicians are screened for eligibility by telephone and are interviewed by personal visit. During the initial interview, the physician provides information on the type of practice, health plans of patients, and other items. During the reporting week, the physician is asked to maintain a log of all patient visits and to record information on a sample of patients. The sample is designed to produce annual data.

Kinds of variables available: Patient information (age, date of visit…), diagnosis information, test results, medication, procedure, provider seen (doctor, nurse, etc.), and other similar data.

Organization of data: "Core" with yearly data available.

More information about data: http://www.cdc.gov/nchs/ahcd.htm

National Health and Nutrition Examination Survey

General Description: This data studies the health and nutrition of adults and children in the U.S. It combines interviews and physical examinations.

Data Description: The survey is conducted annually and surveys the prevalent conditions of the population like diabetes, hearing loss, and other conditions. This data has been used to create informational charts used by doctors, inform FDA regulations, and regulate national programs. This survey is also presently attempting to bring more attention to the conditions of the aging population. **Note**: The data in DataFerrett is from a past version of the survey that covered 6-years. The survey as described is now a continuous survey. The 1994 survey that is available on DataFerrett is NHANES III,

covering 1988 to 1994. Since 1999, the survey is a continuous annual survey. It is now released to the public in two-year increments (i.e. 1999-2000, etc.).

Kinds of variables available: A wide range including weight, number of days skipped meals, exam data like finger swelling (age 60+), demographic data, dental exam results, meal contents (carbohydrates, protein, vitamins, etc.) for combination foods and individual foods, laboratory exam results (insulin, urinary iodine), subdivisions into adult and youth.

Organization of data: Adult, Combination foods, Examination, Individual foods, Laboratory, Variable ingredients, Youth.

More information about data: http://www.cdc.gov/nchs/nhanes.htm

National Health Interview Survey

General Description: Provides information about the prevalence and distribution of illness, its effects in terms of disability and chronic impairments, and the kind of health services people receive.

Data Description: This is a cross-sectional household interview survey. Sampling plans are redesigned after every decennial census, making datasets difficult to compare. The current sample draws 428 primary sampling units (PSU's).

Kinds of variables available: Any member of family receive income from TANF?, Cost of med/dent care past 12mo, Any overnight hospital visits in past 12mo, and other questions involving health and information about the respondent's demographics, income, etc.

Organization of data: 1993 and 1997 to current, where 1993 has many subfolders such as AIDS Knowledge (from PSAs, brochures, etc.), and the latter folder includes datasets by year.

More information about data: http://www.cdc.gov/nchs/nhis.htm

National Hospital Ambulatory Care Survey

General Description: This survey collects data on the utilization and provision of ambulatory care services in hospital emergency and outpatient departments.

Data Description: Findings are based on a national sample of visits to the emergency departments and outpatient departments of non-institutional general and short-stay hospitals. Sample size is around 600 hospitals (varying each survey). There is a 4-week reporting period in which an administrator samples and logs information on patients. This survey supplements the National Ambulatory Care Survey.

Kinds of variables available: Cause of injury, Diagnoses, What diagnostics were used, Counseling, Diagnostic – Blood Pressure, and similar variables related to diagnosis and treatment.

Organization of data: Emergency Department and Outpatient Department folders.

National Survey of Fishing, Hunting, and Wildlife

General Description: This data examines persons involved in fishing, hunting and wildlife activities, such as photography. It looks at trips taken by these individuals.

Data Description: The data was collected in three parts, the first as the screening that determined if a member of the household was involved in hunting/fishing or wildlife watching and placed the interviewees in the appropriate group. The second part identified more subjects that has not responded to the first part but were likely to respond. The third part was a detailed interview from those sampled in parts 1 & 2.

Kinds of variables available: States hunted in, kind of fishing participated in, took wildlife-related trips, how likely to take trips, etc. Some variables are confusing because there are multiples, like for states hunted in (HUNTSTD#), but this is because respondents could list up to 8 states.

Organization of data:

More information about data: http://www.census.gov/prod/www/abs/fishing.html

New York City Housing and Vacancy Survey

General Description: This study is conducted every three years to analyze NYC's compliance with rent regulation laws. The study principally collects housing data but also collects demographic data.

Data Description: The study draws from the decennial census plus new occupancy sites since the census. Approximately 18,000 units are sampled, representing the five boroughs of NYC. File structure description: There are three files: household, house, and person. There is a household record for each occupied housing unit, a house record for each vacant housing unit, and a person record for each person in occupied housing units (households). Person records can be linked to their appropriate household record by means of a sequence number. This number is unique to each housing unit and is the same on all person records within a household. There are three "weights" that appear on the various files. To generate estimates pertaining to occupied housing units or households, use the "household weight". Similarly, to generate estimates pertaining to vacant units, apply the "housing unit weight". Use the "person weight" to generate estimates of population in the housing units.

Kinds of variables available: Asked condition of building unit, observed condition, ratings of neighborhood by respondents, year moved in, down payment, mortgage status, and generally a wide range of variables available.

Organization of data: Household and Persons dataset, Vacant Units dataset. 2005 and 2002 are available in DataFerrett.

More information about data: http://www.census.gov/hhes/www/housing/nychvs/nychvs.html

Small Area Health Insurance Estimates

General Description: These are model-based estimates of health insurance coverage for counties and states. It builds on the work of the Small Area Income and Poverty Estimates set.

Data Description: The data provided are indirect estimates produced by statistical model-based methods using sample survey, decennial census, and administrative data sources. To maintain confidentiality, the Census Bureau uses procedures to assure that the estimates and related information that are released cannot be used to disclose individual data or violate other confidentiality restrictions applicable to the source data.

Kinds of variables available: Demographics, percent insured/percent uninsured, uninsured – number of people, etc.

Organization of data: State and County by Income and Demographic, years included in DataFerrett are 2005 and 2006.

More information about data: http://www.census.gov/did/www/sahie/

Small Area Income and Poverty Estimates

General Description: Data from 1995 to 2002 is available on DataFerrett. This data is meant to supplement the decennial census in estimating the number of people living in poverty in small areas.

Data Description: This is a complex data set, as the relation between income or poverty and tax and program data for the states were modeled and a subset of counties using estimates of income or poverty from the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC). The study then used the modeled relations to obtain estimates for all states and counties. For school districts, they used the model-based county estimates and the decennial census distribution of the population of poor of each county over its constituent school districts.

Kinds of variables available: poverty-# people-estimate for different age groups, Income estimate, and confidence intervals lower/upper bounds for variables, etc.

Organization of data: County and state folders with each year's data as a dataset.

More information about data: http://www.census.gov/did/www/saipe/

Social Security Administration

General Description: Data from 2001 on benefits received and other program eligibility.

Data Description: The data available in this program for Old Age, Survivors, and Disability Insurance (OASDI) contains an extract of data fields from the main database and consists of a 1 percent random, representative sample of persons entitled to receive an OASDI benefit payment for December 2001 and the data for Supplemental Security Income (SSI) is a 5 percent random, representative sample for that

program for December 2001. Descriptions of data fields are provided in the Data Dictionary for the file. The descriptions cover the content of each field, the method of presentation, and the disclosure avoidance steps taken to provide confidentiality. SSA asserts that the data is of high quality.

Kinds of variables available: Date of eligibility, residence, income, program eligibility, type of benefit, among others.

Organization of data: Split into two major programs: Old Age, Survivors, and Disability Insurance OASDI and Supplemental Security Income (SSI)

More information about data: http://www.ssa.gov/policy/

Survey of Income and Program Participation

General Description: Information on amount of income, labor force information, program participation and eligibility data, and general demographic characteristics. For more data like this, see the Survey of Program Dynamics.

Data Description: The survey design is a continuous series of national panels, with sample size ranging from approximately 14,000 to 36,700 interviewed households. The duration of each panel ranges from 2 1/2 years to 4 years. The SIPP sample is a multistage-stratified sample of the U.S. civilian non-institutionalized population. Cross-year data differs significantly: For the 1984-1993 panels, a new panel of households was introduced each year in February. A 4-year 1996 panel was introduced in April 1996; a 3-year panel was started in February 2000 but cancelled after 8 months for budget reasons; and a 3-year panel was introduced in February 2001.

Kinds of variables available: Employment status, amount of income from social security, Medicare coverage, family demographics, amount received from other programs (food stamps, etc.), among others.

Organization of data: Panel and Panel Longitudinal data. Longitudinal data includes a core file and a full panel extract while the Panel folders allow the user to access data from each "Wave"

More information about data: http://www.census.gov/sipp/

Survey of Program Dynamics

General Description: Data on demographic, social, and economic characteristics of a nationally representative sample of the U.S. population that permits the evaluation of the welfare reform legislation. For more data like this, see the Survey of Income and Program Participation.

Data Description: Data collected in these panels included family/household composition, program eligibility, access and participation, transfer income, and in-kind benefits; detailed economic and demographic data on employment and job transitions; and income.

Kinds of variables available: Education, employment, income, health insurance, poverty, demographic and other variables.

Organization of data: Cross-sectional includes data sets from each year from 1999-2002 while the Longitudinal data folder holds Third File, which pulls in data from each of these sets and sometimes limits the data, ex. Universe = Only respondents age 15+.

More information about data: http://www.census.gov/spd/