

Useful Datasets for the Ec980x “Personal Economics” Research Paper

These are some of the most commonly used resources for labor economists and also for students in Ec980x. But there are many more that are tailored to an issue or question. The secondary sources that you will consult may guide you to sources in addition to these.

National Survey Data

The national surveys listed below have been administered over many years and cover large samples of the US population. They are well-suited for analyses that examine trends over time or differences across geographies.

American Community Survey (ACS): <https://usa.ipums.org/usa/>

Annual survey data covering approximately 1 percent of the population. Available for years 2005-present and, among a smaller sample size, for 2001-2004. Contains information about household and family composition. Provides individual level data on marital status, cohabitation, fertility over past year, demographics, education, work, income, and migration for each member of the household. From 2008 onward contains detailed information about the number of times married, year of last marriage, and changes in marital status over the past year. Geographic information is provided at the Public Use Microdata Area level (geographic areas with 100K people or more, which do not cross state lines). Can be used in conjunction with the decadal census. Note: The ACS began in 2001 to replace the long-form on the decennial census. Unlike the decennial census, the ACS isn't a full count each year. Rather, it is a different slice of the US each year.

Decadal Census: <https://usa.ipums.org/usa/>

All decadal censuses conducted from 1850 onward. Sample size ranges from 1 to 5 percent of the population (the default sample is likely appropriate for your purposes), due to different weighting schemes across years. Data fields evolve over time. Marital status is available for 1880 onward. Cohabitation is available for recent decades. Number of times married, age at first marriage, duration of current marital status, and number of children ever born are available for ad hoc years between 1900 and 1980. Also contains information about household and family composition and provides individual level data on marital status, demographics, education, work, income, and migration for each member of the household. Geographic information is available at the Public Use Microdata Area level (areas with 100K people or more, which do not cross state lines). Can be used in conjunction with the ACS.

Current Population Survey (CPS): <https://cps.ipums.org/cps/>

Monthly survey data from the 1960s onward covering approximately 50,000 households per month. Basic monthly data contain information about household and family composition and provide individual level data on marital status, demographics, education, and work for each member of the household. Information about cohabiting partners is available from 2008 onward. The fertility supplement provides detailed data on marriage and childbearing and has been administered approximately every other June since the 1970s. Although data fields in the fertility supplement change somewhat over time, information about the number of live births ever had and birth year of most recent child is consistently provided over the duration of the survey.

Geographic information is provided at the metropolitan statistical area level. Note: The CPS began in 1944 but the micro-data exist only since 1962.

Topic Specific Survey Data

The surveys listed below are national surveys that take a deep dive into the time use activities, health-related behaviors, and education of respondents. They are well-suited for analyses that seek to explore these topics.

American Time Use Survey (ATUS): <https://www.atusdata.org/atus-action/variables/group>
Nationally representative time diary survey sampling approximately 26,000 individuals each year since 2003. Drawn from the CPS, it contains information about individual and household characteristics typically provided in the CPS (see above) along with time use information. Respondents are interviewed one time about how they spent their time on the previous day, where they were, and whom they were with. Activity codes are detailed and cover childcare, housework, paid work, and leisure, among many other time uses.

National Health Interview Survey (NHIS): <https://nhis.ipums.org/nhis/>
Nationally representative survey covering a wide range of health status and utilization measures administered annually since around 1960. Questionnaires undergo major revisions every 10-15 years, so many topics are only surveyed on an ad hoc basis. Dataset covers women's health (including fertility and family planning), general health, mental health, health behaviors, and health insurance, among other health topics, and also contains information about family and household structure, marital status, cohabitation, demographics, education, work, and income.

National Survey of College Graduates (NSCG): <https://www.census.gov/programs-surveys/nscg/data.html>
Biennial survey of college graduates that has been conducted since the 1970s (data available online for 1993 onward). Sample size is approximately 100,000 individuals. Contains information about marital status, number of children, detailed educational history (including university characteristics and degrees obtained), work, income, and job satisfaction for college graduates.

Longitudinal Survey Data

The longitudinal surveys listed below track individuals over time ask them very detailed questions. These surveys are well-suited for analyses that examine specific behaviors and attitudes as well as those that require information about earlier experiences and beliefs.

National Longitudinal Survey of Youth (NLSY79): <https://www.nlsinfo.org/investigator/>
Nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979. Individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis. Data are broad, containing detailed information about marriage, cohabitation, sexual activity, pregnancy, and fertility along with details about childhood, education, labor market activity, health, crime, and attitudes. Children who were born to women in the cohort are also tracked as part of the survey (NLSY79 Child & Young Adult). Can be combined with the NLSY97 (NLS Cross Cohort Beta).

National Longitudinal Survey of Youth (NLSY97): <https://www.nlsinfo.org/investigator/>

Nationally representative sample of approximately 9,000 youths who were 12 to 16 years old as of December 31, 1996. Youths continue to be interviewed on an annual basis. Data are broad, containing detailed information about marriage, cohabitation, sexual activity, pregnancy, and fertility along with details about childhood, education, labor market activity, health, crime, and attitudes. Can be combined with the NLSY79 (NLS Cross Cohort Beta).

National Longitudinal Study of Adolescent to Adult Health (Add Health):

<https://www.cpc.unc.edu/projects/addhealth/documentation/publicdata>

Longitudinal study of adolescents in grades 7-12 during the 1994-95 school year followed into young adulthood with four in-home interviews. Most recent wave was conducted in 2008, when sample was 24-32, and includes marriage and cohabitation histories, pregnancy and fertility histories, educational histories, and employment histories for roughly 5,000 respondents. Earlier waves provide contextual data on family, neighborhood, community, school, friendships, peer groups, and romantic relationships.

Panel Study of Income Dynamics (PSID): <https://simba.isr.umich.edu/VS/i.aspx>

Long running longitudinal household survey which began in 1968 with a nationally representative sample of over 18,000 individuals living in 5,000 families. Information on these individuals and their descendants has been collected continuously, including data covering marriage, childbearing, child development, education, employment, income, wealth, expenditures, and health topics. Available variables evolve over time but are detailed and can cover topics such as birth control use, desired family size, and childcare expenditures, in addition to standard information about marital status and fertility.