Why do we need longitudinal data?

- A huge number of interesting questions can be analyzed with longitudinal information:
  - Schooling decisions
  - Effect of previous labor market experience on wages.
  - Fertility decisions
  - School performance $\Rightarrow$ labor market outcomes
  - Smoking during pregnancy $\Rightarrow$ baby's health $\Rightarrow$ schooling achievement
  - and many many more
Always ask yourself:

- Is attrition a serious problem in the data?
- What is the sampling scheme of the data? are there weights?
- Does the question change over time?
- Who answer the question? (head of household, individual, teacher, etc)
- Does the data contain retrospective information? Is it reliable?
- Does the survey contain non-public information? Is it possible to get?
- Errata?
- Does my software provide enough flexibility and capability? What if I need to output things to C, Fortran, or R?
Two Important Sources of Information

- The National Longitudinal Surveys (NLS)
- Panel Study of Income Dynamics (PSID)
National Longitudinal Survey of Youth 1997 (NLSY97): Survey of young men and women born in the years 1980-84; respondents were ages 12-17 when first interviewed in 1997.

National Longitudinal Survey of Youth 1979 (NLSY79): Survey of men and women born in the years 1957-64; respondents were ages 14-22 when first interviewed in 1979.

NLSY79 Children and Young Adults: Survey of the biological children of women in the NLSY79.
National Longitudinal Surveys of Young Women and Mature Women (NLSW): The Young Women’s survey includes women who were ages 14-24 when first interviewed in 1968. The Mature Women’s survey includes women who were ages 30-44 when first interviewed in 1967. These surveys are now conducted simultaneously in odd-numbered years.

National Longitudinal Surveys of Young Men and Older Men: The Young Men’s survey, which was discontinued in 1981, includes men who were ages 14-24 when first interviewed in 1966. The Older Men’s survey, which was discontinued in 1990, includes men who were ages 45-59 when first interviewed in 1966.
The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979.

These individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis.
Labor force activity: start and stop dates for each job held since the last interview, periods in which individuals are not working but still with an employer, and labor market activities during gaps between jobs.

Additional information: marital status, fertility, and participation in government assistance programs such as unemployment insurance.
Although a primary focus of the NLSY79 survey is labor force behavior, the content of the survey is considerably broader.

The survey contains questions ranging from child-care costs to welfare receipt. For example: educational attainment, training investments, income and assets, health conditions, workplace injuries, insurance coverage, alcohol and substance abuse, sexual activity, and marital and fertility histories.

It includes an aptitude measure (AFQT), a school survey, and high school transcript information.
NLSY79

Getting the data and the manual

- NLS Database Investigator: http://www.nlsinfo.org/dbgator/index.php3
- This software provides everything you need!
- Example....
The Panel Study of Income Dynamics (PSID), begun in 1968, is a longitudinal study of a representative sample of U.S. individuals (men, women, and children) and the family units in which they reside.

It emphasizes the dynamic aspects of economic and demographic behavior.

The sample size has grown from 4,800 families in 1968 to more than 7,000 families in 2001.

At the conclusion of 2003 data collection, the PSID will have collected information about more than 65,000 individuals spanning as much as 36 years of their lives.
Income sources and amounts, Poverty status, Public assistance in the form of food or housing, Other financial matters (e.g., taxes, inter-household transfers), Family structure and demographic measures (e.g., marital events; birth and adoptions; children forming households), Labor market work (e.g., employment status, work/unemployment/vacation/sick time; occupation, industry; work experience), Housework time, Housing (e.g., own/rent, house value/rent payment, size), Geographic mobility (e.g., when and why moved; where Head grew up; all states Head has lived in), Socio-economic background (e.g., education, ethnicity, religion, military service; parents’ education, occupation, poverty status), Health (e.g., general health status; disability; 30-day emotional distress).

The PSID also includes a child supplement.
The Data Center (http://simba.isr.umich.edu/) is the most popular means for obtaining PSID data.

The Data Center also provides access to customized HTML and PDF codebooks.

In addition to the data center, data and documentation can also be obtained from the PSID Web site in the form of prepackaged files.

The packaged data files are in ASCII format.

The data center automatically merges your selected files, but example SAS and SPSS programs have been prepared to assist users with creating cross-year analysis files from packaged downloads.
Other data sets

ABC, Perry, NLS of Adolescent health, early childhood longitudinal study (ECLS), high school & beyond (HSB), National Education Longitudinal Study of 1988 (NELS:88), National Survey of Families and Households, Chicago Longitudinal Study (CPC).
Additional sources of information

- World Wide Web Resources in Economics: (WebEc) http://www.helsinki.fi/WebEc/webecc8d.html
- The University of Michigan Library - Data Center: http://www.lib.umich.edu/govdocs/stecon.html
- You can also get longitudinal data from World Bank, AIDB and local governments.