This sample is a **cross section**, a snapshot of the population that includes people of different ages. Each person is included only once.

Another type of sample is **longitudinal**, where the cases are tracked over time, each person being included more than once in the data frame. A longitudinal data set for the runners looks like this:

DATA FILE
Cherry-Blosssom-Long.csv

Name	Year of Birth	Age	Time	Sex	Year
a. renee callahan	1966	37	103.3	F	2003
a. renee callahan	1966	42	112.4	F	2008
aaren pastor	1991	14	94.4	F	2005
aaren pastor	1991	15	74.8	F	2006
aaron alton	1974	31	86.9	M	2005
aaron alton	1974	32	92.5	M	2006
aaron alton	1974	33	92.4	M	2007
aaron alton	1974	34	89.3	M	2008

... and so on.

The individual runners have been tracked from year to year, so each individual person shows up in multiple rows.

If your concern is to understand how individual change as they age, it's best to collect data that show such change in individuals. Using cross-sectional data to study a longitudinal problem is risky. Suppose, as seems likely, that younger runners who are slow tend to drop out of racing as they age, so the older runners who do participate are those who tend to be faster. This could bias your estimate of how running speed changes with age.

2.5 Computational Technique

2.5.1 Reading Tabular Data into R

Data used in statistical modeling are usually organized into tables, often created using spreadsheet software. Most people presume that the same software used to create a table of data should be used to display and analyze it. This is part of the reason for the popularity of spreadsheet programs such as Excel and Google Spreadsheets.

For serious statistical work, it's helpful to take another approach that strictly separates the processes of data collection and of data analysis: use one program to create data files and another program to analyze the data stored in those files. By doing this, one guarantees that the original data are not modified accidentally in the process of analyzing them. This also makes it possible to perform many different analyses of the data; modelers often create and compare many different models of the same data.