

## Demographics and employment in the United States

In the wake of the Great Recession of 2009, there has been a good deal of focus on employment statistics, one of the most important metrics policymakers use to gauge the overall strength of the economy. In the United States, the government measures unemployment using the Current Population Survey (CPS), which collects demographic and employment information from a wide range of Americans each month. In this exercise, we will employ the topics reviewed in the lectures as well as a few new techniques using the September 2013 version of this rich, nationally representative dataset (available [online](#)).

The observations in the dataset represent people surveyed in the September 2013 CPS who actually completed a survey. While the full dataset has 385 variables, in this exercise we will use a more compact version of the dataset, [CPSData.csv](#), which has the following variables:

**PeopleInHousehold:** The number of people in the interviewee's household.

**Region:** The census region where the interviewee lives.

**State:** The state where the interviewee lives.

**MetroAreaCode:** A code that identifies the metropolitan area in which the interviewee lives (missing if the interviewee does not live in a metropolitan area). The mapping from codes to names of metropolitan areas is provided in the file [MetroAreaCodes.csv](#).

**Age:** The age, in years, of the interviewee. 80 represents people aged 80-84, and 85 represents people aged 85 and higher.

**Married:** The marriage status of the interviewee.

**Sex:** The sex of the interviewee.

**Education:** The maximum level of education obtained by the interviewee.

**Race:** The race of the interviewee.

**Hispanic:** Whether the interviewee is of Hispanic ethnicity.

**CountryOfBirthCode:** A code identifying the country of birth of the interviewee. The mapping from codes to names of countries is provided in the file [CountryCodes.csv](#).

**Citizenship:** The United States citizenship status of the interviewee.

**EmploymentStatus:** The status of employment of the interviewee.

**Industry:** The industry of employment of the interviewee (only available if they are employed).