I've had a few questions about the census lately, so I thought I'd provide a little summary. "The Census" isn't one, monolithic chunk of data. Rather, it is a collection of different data sets collected or created for different purposes. Here, we'll discuss three of the main sources for population and housing data: The Decennial Census, The American Community Survey, and the Population Estimates Program. At the end, we'll summarize where to obtain the data and how to decide which datasets to use.

**The Decennial Census**

When people refer to "The Census", they are usually referring to this data set. It has been collected every ten years since 1790, and provides a snapshot of what the country looked like on April 1st. The Decennial Census (DC) consists of two parts. Every household in the country fills out the census short form, which consists of basic demographic questions: age, gender, race, family relations, housing units, occupancy, tenure, and a few others. This 100% count of the population is stored in a data file called Summary File 1 (SF1). A second form, the long form, is sent to 1 of every 6 households and contains additional questions that include: language, ethnicity, employment status, income, poverty, education, home value, home construction, and more. The sample is then extrapolated for the general population and is stored in a data file called Summary File 3 (SF3).

The main advantage of the DC (over other data sets) is that it is extremely comprehensive, in terms of the topics it covers and the levels of geography at which it is available (from the entire nation to individual census blocks). Compared to other datasets, it is also the most "accurate" as it is based on a 100% count and on a large sample survey. The primary disadvantage is frequency, as it is only collected once every ten years.

When the Decennial Census is taken again in 2010, it will consist solely of the 100% count data collected from the short form (SF1). The long form data (SF3) will no longer be collected. It will be replaced by the American Community Survey.

**The American Community Survey**

As the name implies, the American Community Survey (ACS) is an annual survey, not an actual count. Testing for the ACS began in the late 1990s and data is available from 2003 to the present. Every year, the census surveys three million households across the country, insuring that some households from every county are collected. The ACS includes all of the questions from SF1, many of the questions from SF3, and several questions that are not covered at all in the decennial census (go here for a list of 2006 ACS topics). However, due to the small sample size, numbers from the ACS must be considered as estimates with a margin of error. For example, according to the decennial census the population of Newark, NJ was 273,546 in 2000.
According to the 2006 ACS, we can say that we are 90% confident that the population of Newark is 266,736, plus or minus 10,113 people.

The disadvantage of the ACS is obvious, in that we are getting estimates with margins of error and not counts for a fixed period of time. Another disadvantage, albeit a temporary one, is frequency. As of 2007, data was only available for areas of geography that have at least 65,000 people. Data for smaller areas was not available, because the small sample size would make the estimates unreliable. However, at the end of 2008 the ACS will release two datasets: one is an annual estimate for areas with more than 65,000 people, and the other will be a three year estimate of areas with more than 20,000 people. Taking a three year estimate for the smaller areas will make the statistics reliable enough to publish. In subsequent years, the estimate will be revised by adding data from the newest survey year and dropping data from the oldest survey year. So, from 2008 forward, data for areas with more than 20,000 people will always be available each year. By the year 2010, the bureau will have enough data to begin publishing a third dataset, which will be a five year average for ALL geographic areas. At that point, the big advantage of the ACS will be that we will have comprehensive data published on an annual basis for all geographic areas.

Population Estimates

While the decennial census is a count, and the ACS is a survey, population estimates are the result of calculations based on the cohort component method. This method counts the number of births, deaths, and migrants for the country as a whole and for each individual county within five-year age groups called cohorts. At the beginning of the decade, the bureau begins with the decennial census count (Census 2000 for example), factors in births, deaths, and migration for these areas, and creates an estimate for July 1st, 2001. In 2002 they perform the same calculations based on the estimate from 2001, and then create estimates each year based on the new demographic data and the previous years estimates. Data is available for total population, race, Hispanic origin, age (broken down by gender for each year at the national level and for five year groups below that) and housing units for states and counties. Some data is also available for metropolitan areas (which are county based) and county subdivisions (for total population only). Compared to our example for Newark, NJ above, the population estimate for 2006 was 281,402 (which is much higher than the ACS number).

The advantage of population estimates is that they are available every year and are more precise, consisting of a single value rather than a range of values. The disadvantages are that the estimates are limited to basic population figures, and they become less reliable as you move further away from a decennial census year. Once a new decennial census is taken, the bureau goes back and revises the estimates it made for each year. So the estimates from 2001 to 2009 will be revised after the 2010 census in order to accurately reflect what happened over the course of the decade. The bureau has been creating population estimates since the beginning of the 20th century.
Summary

All of this data can be accessed through the American Factfinder, which is the best place to go if you want to download actual data tables. In addition, the Population Estimates Program has a separate page where they provide summary tables. This is the best place to go for basic population rankings of states, counties, metropolitan areas, and cities. If you are looking for a summary overview of a particular place, you can use the Population Finder or Quickfacts, but you need to be careful in evaluating the data, since these overviews are a mishmash of many datasets. Here are a few clues that can help you identify the data source:

- Years ending in zero are usually from the decennial census (until the year 2010, with simultaneous publishing of the DC and ACS)
- Most numbers from the American Community Survey are accompanied by a margin or error (i.e. 266,736 +/- 10,113).
- Population estimates usually have the word "estimate" as a descriptor. Estimates are only collected for total population, age, gender, race, Hispanic origin, and number of housing units.
- Data about businesses or industries are only collected in the Economic Census. It is published every five years, in years ending with 2 and 7.

Which dataset should you use?

- Use the Decennial Census if you need comprehensive data for small areas (census tracts, block groups, blocks, zip codes) or if you need complete, contiguous coverage for a specific geographic area (all of the counties for a specific state). Also use the decennial census for making historical comparisons.
- Use Population Estimates for basic demographic values (total population, age, gender, race, Hispanic origin, total housing units) for recent years. Also use estimates when examining change between the last decennial census and today, for states, counties, metro areas, and places.
- Use the American Community Survey for current, comprehensive data that is not calculated as part of the population estimates program, for areas that have at least 65,000 people (and soon, areas with more than 20,000 people). Appropriate for states and for large metro areas, counties, and cities.

If you are affiliated with Baruch college and would like more information about Census data, contact Frank Donnelly, the Geospatial Librarian at Baruch College CUNY.