Video 3: The U.S. Census Bureau and https://data.census.gov

Jylisa Doney
University of Idaho, Social Sciences Librarian
jylisadoney@uidaho.edu

Transcript

00:00 Hello again, welcome to Video 3.
00:02 In this video, we're going to practice
00:05 using data.census.gov
00:07 as the U.S. Census Bureau is phasing out
00:09 American FactFinder after June 2019.
00:12 Data.census.gov is a powerful tool
00:15 that allows visitors to search for statistical information
00:18 and sometimes find aggregate data
00:20 on a variety of topics.
00:22 As I mentioned in Video 2,
00:24 the Census Bureau conducts over 100 surveys,
00:27 and if you are new to Census data,
00:29 you may not know which survey collects the data
00:31 that would be relevant for your topic.
00:33 That's where data.census.gov comes in.
00:36 All statistical information included in this resource
00:39 is compiled from raw survey data,
00:41 and the actual surveys are cited within each source.
00:45 So we can use a tool like data.census.gov
00:48 to find statistical information about our topic,
00:51 identify the survey that collected this information,
00:54 visit the actual survey page,
00:56 and find and download the raw data.
00:59 I'm going to demonstrate each of these steps
01:01 in this video.
01:02 We're also going to discuss how to find
01:04 related documentation that can help us understand
01:07 the data we download.
01:08 To begin, let's navigate to data.census.gov.

[Data.census.gov: https://data.census.gov]

01:13 Once the page loads,
01:14 we can either type details about our topic
01:16 into the search box or click on "advanced search."
01:19 Let's use the "advanced search" for this demonstration.
01:22 I find the "advanced search" more useful
01:24 because it provides filters
01:25 that help us construct our search.
01:27 First, let's click on the "Topics" filter.
01:30 From here, we can click on the
01:32 broader topic categories that interest us,
01:34 such as "educational attainment"

01:36 as well as "income and poverty."
01:39 Filter options that are grayed-out
01:41 are unavailable in data.census.gov
01:44 based on your previous selections.
01:46 You can also select other filter options
01:49 such as "Geography."
01:50 After selecting filters, we can see that each selection
01:54 has been added to the search query.
01:56 Now we can click "View All Results" to run our search.
02:00 The results page on data.census.gov
02:03 will show you tables, maps, and pages
02:05 that relate to your search.
02:07 For our purposes, the tables are the most relevant result
02:10 as they include the statistical information
02:13 that can help us identify the survey
02:15 that collected data related to our topic.
02:18 Click on "View All Tables"
02:20 to see the results that match our search query.
02:23 You can scroll through the related tables
02:25 and click on the titles
02:26 to see the associated information.
02:29 After clicking into a table, you can see the
02:31 survey or program where it came from
02:33 as well as the universe, date, and estimate information.
02:37 We can also see that the results are organized
02:39 based on a geographic filter selection.
02:42 If a table seems relevant to your data need,
02:45 you can then navigate directly
02:46 to the survey or program page on the Census website
02:49 to find and download the microdata.
02:51 In this case, the statistical information reported
02:54 came from the American Community Survey.
02:57 So let's navigate to census.gov,

[United States Census Bureau: https://www.census.gov]

03:00 and from here we can hover over "Surveys/Programs"
03:03 and click on "American Community Survey."
03:06 From this page we can hover over "Data"
03:08 and see two options to get the microdata or PUMS,
03:11 which stands for Public Use Microdata Sample.
03:15 The first is "Data via FTP" and the second
03:18 is "PUMS Data," and although it's a good option,
03:21 it currently directs you to American FactFinder
03:23 to download data.
03:25 Since the Census is phasing out American FactFinder,
03:28 let's click on the "Data via FTP" option.
03:31 We can then click on the hyperlink under "Link"
03:33 for "Public Use Microdata Sample (PUMS) Files."
03:37 This new page arranges
03:39 American Community Survey data based on year

03:41 and has data from 1996 to present.
03:44 So let's click on "2017,"
03:46 and select the "5-year estimate" option.
03:49 This page looks a bit daunting
03:50 but is actually very well-organized.
03:53 The organizational structure and naming convention
03:55 is detailed in the PDF included on this page,
03:58 but I'll give a brief explanation.
04:00 The first part of the file name identifies the file format,
04:03 either CSV or UNIX.
04:06 CSVs can be opened in Excel and imported into SPSS,
04:10 while the UNIX files can be opened in the SAS tool.
04:13 The second part of the final name
04:15 represents the record level,
04:16 either the household, "h," or the person, "p."
04:20 Lastly, the two letters before the .zip
04:23 are the state abbreviation.
04:25 So let's scroll down and click on the file name,
04:27 "csv_pid.zip."
04:31 Once we download and save,
04:32 as well as open this file,
04:34 we can see that it includes two separate files.
04:37 The ReadMe file shares
04:38 introductory information about the data
04:40 and details any changes to the survey
04:42 since it was last conducted.
04:44 While the raw data file includes
04:46 the data that we can analyze
04:47 to investigate our research question or topic.
04:50 Since this is a raw data file and hasn't undergone
04:53 any simplification or cleaning,
04:55 we can't really understand
04:56 any of the information that's included.
04:58 To facilitate our data analysis,
05:01 we now need to find the associated
05:02 codebooks or data dictionaries,
05:04 which include the information we need
05:06 to understand each of the variables
05:08 and potential answer options.
05:10 To access the documentation,
05:12 we need to navigate back
05:13 to the specific survey page
05:14 on the Census Bureau website.

[American Community Survey: https://www.census.gov/programs-surveys/acs]

05:16 From here, we can hover over
05:18 "Technical Documentation,"
05:20 and then click on "PUMS Documentation."
05:23 Once we click on "PUMS Documentation,"
05:26 we are directed to a page that provides

05:28 additional information about the PUMS data.
05:30 We can then click on the link for
05:32 "PUMS Technical Documentation."
05:35 The information and documentation on this new page
05:38 is organized based on the year of the survey
05:41 and defaults to the most recent year.
05:43 We can also see the ReadMe file,
05:45 which was included in our initial data download.
05:48 One of the most important pieces of documentation
05:51 is the "PUMS Data Dictionary,"
05:53 which lists and describes the associated variables
05:55 for the particular year and survey type.
05:58 Be sure to select the documentation or data dictionary
06:01 that matches the data you downloaded.
06:03 Let's download the PDF version for the
06:05 "2013 to 2017 ACS 5-year PUMS Data Dictionary."
06:11 When using ACS data, the data dictionary
06:14 will include the variables used in both the
06:15 housing records and the person records.
06:18 So for this data dictionary,
06:20 we will need to scroll to page 29 in this document
06:23 to see the list of person record variables.
06:26 Now that we have the data dictionary,
06:28 we can begin the process of analyzing our data.
06:31 Coming up in Video 4, I'm going to demonstrate
06:34 how to use an IPUMS repository
06:36 to find aggregate geographic data.