

# Getting education data from common sources

## EDH7916

Benjamin Skinner

There are many places you can find higher education data. Below are a few publicly-available sources with instructions on how to find and download data sets to use in your research. This list is by no means complete, but should give you a general idea of what's available.

### NCES Surveys

The National Center for Education Statistics (NCES)<sup>1</sup> is part of the Department of Education's Institute of Education Sciences (IES). NCES offers a large number of resources for researchers interested in higher education.



The screenshot shows the homepage of the National Center for Education Statistics (NCES) website. At the top, there is a navigation bar with links for "About Us", "Surveys and Programs", "Publications and Products", and "Data and Tools". Below the navigation bar, there is a section titled "WHAT'S NEW AT NCES" featuring three recent publications: "A Look at Postsecondary Certificates and Associate's Degrees in Occupational Fields of Study" (May 13), "Highlights from the PISA 2018 Financial Literacy Assessment" (May 7), and "School Nurses in U.S. Public Schools" (April 30). To the right of these news items is a section titled "THE CONDITION OF EDUCATION" which features a circular graphic with various icons representing different fields of study and labor force outcomes. The main content area of the page shows a classroom scene with students raising their hands.

Among these are longitudinal surveys that follow different cohorts of students from high school through college and beyond.

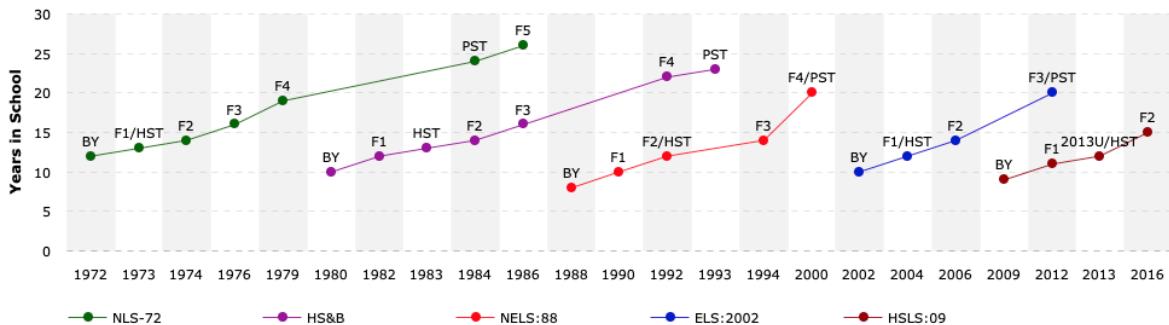
<sup>1</sup><https://nces.ed.gov>

## Historical Background

HSLS:09    **ELS:2002**    NELS:88    HS&B    NLS-72

**ELS:2002** must be seen in the context of the prior NCES high school studies. ELS:2002 looks back to these past studies, upon which it builds and to which its findings will be compared. At the same time, it uniquely enhances the accomplishments of its predecessors by updating the content of the survey and extending the time line to a new decade.

Research Design for the NCES High School Cohorts



NLS-72: National Longitudinal Study of the High School Class of 1972  
 HS&B: High School and Beyond  
 NELS:88: National Education Longitudinal Study of 1988  
 ELS:2002: Education Longitudinal Study of 2002  
 HSLS:09: High School Longitudinal Study of 2009

BY: Base Year  
 F1: First follow-up data collection  
 F2: Second follow-up data collection  
 F3: Third follow-up data collection  
 F4: Fourth follow-up data collection  
 F5: Fifth follow-up data collection

2013 U: 2013 update  
 HST: High School Transcript  
 PST: Post-secondary Transcript

Specifically, we'll go through the steps to download the Education Longitudinal Study of 2002 (ELS)<sup>2</sup>. The good news is that the process is the same for the other surveys.

To get to the online code book you'll use to download the raw data files, head to the NCES homepage ([nces.ed.gov](http://nces.ed.gov)<sup>3</sup>) and click in the following order: 1. *Menu* 1. *Data & Tools* 1. *Downloads Microdata/Raw Data* 1. *EDAT*

<sup>2</sup><https://nces.ed.gov/surveys/els2002/>

<sup>3</sup><https://nces.ed.gov>

The screenshot shows the NCES homepage with a dark header. The header includes the IES > NCES logo, the text "National Center for Education Statistics", a "MENU" button, a search bar, and a "Co" link. A red box labeled "(1)" highlights the "MENU" button. A red box labeled "(2)" highlights the "Data & Tools" link under the Surveys & Programs section. A red box labeled "(3)" highlights the "Downloads Microdata/Raw Data" link under the Data & Tools section. A red box labeled "(4)" highlights the "EDAT" link under the Downloads Microdata/Raw Data section. The main content area features a sidebar with links like "About Us", "Surveys and Programs", "Publications and Products", and "Data and Tools".

#### WHAT'S NEW AT NCES

- MAY 13** [A Look at Postsecondary Certificates and Associate's Degrees in Occupational Fields of Study](#)  
This new set of web tables examines trends in subbaccalaureate occupational education from 2005-06 to 2016-17. [» More info](#)
- MAY 7** [Highlights from the PISA 2018 Financial Literacy Assessment](#)  
This web-based report presents the 2018 results of the PISA financial literacy assessment. [» More info](#)
- APR 30** [School Nurses in U.S. Public Schools](#)  
This Data Point describes the presence of school nurses in public schools for the 2007-08, 2011-12, and 2015-16 school years, and by selected school characteristics for the 2015-16 school year. [» More info](#)
- APR 23** [A look at the postsecondary enrollment of students who took career and](#)

#### THE CONDITION OF EDUCATION



Browse key indicators on the condition of education in the United States at all levels, from prekindergarten through postsecondary, as well as labor force outcomes and international comparisons. The indicators summarize important developments and trends using the latest statistics, which are updated throughout the year as new data become available.

You may see a couple of popups — just agree. Once you've clicked through those, you should see the code book home screen. From here, you can access a number of data sets. We'll focus on ELS, but as I said above, the process is the same: just choose another data set from the code book homepage if you'd rather use that data.

The screenshot shows the NCES Online Codebook homepage. At the top, there's a navigation bar with the IES > NCES logo, the National Center for Education Statistics, a menu icon, a search bar, and links for 'Need Help?', 'LOG IN', and 'Feedback'. Below the header, a teal banner reads 'Public-Use' and 'Explore the Codebooks, Download Datasets, Create Syntax'. It includes a sub-instruction: 'Online Codebook guides you through selecting a survey, population, and variables relevant to your analysis.' To the right of the banner is a 'Take a Tour' button. The main content area displays eight study cards: ECLS-K (Kindergarten Class of 1998-99), ELS (Education Longitudinal Study of 2002), HSLS (High School Longitudinal Study of 2009), NELS (National Education Longitudinal Study of 1988), NHES (National Household Education Survey), SASS (Schools and Staffing and Teacher Follow-up Surveys), and SSOCS (School Survey on Crime and Safety). The ELS card is highlighted with a red box around its title and subtitle, and a red arrow points from the 'Click here for ELS' button at the top left towards it.

When you choose ELS, you'll see the online code book. You can (and should) use this to learn about variables — their definitions, how they're constructed, missing values, *etc.* You can also use this tool to only select a few variables for download. Don't do that! You should plan to download the full data set and do any filtering or subsetting in your analytic code.

Click the *Downloads* button to get the data.

Use at least one search selection above to explore this codebook.

Select variables of interest to create a syntax file for your preferred software. Download the dataset formatted for your preferred software.

**Data Files**

**DOWNLOAD**

**CREATE SYNTAX**

**Click here to download the data files**

You'll be presented with a number of file types. Because you are using R, you could read in all these data types — either with standard functions or functions from the tidyverse [haven](#)<sup>4</sup> package.

My recommendation:

- **If you don't care about labels:** Download the CSV version for maximum portability
- **If you'd like labeled data:** Download the STATA version and use `haven::read_dta()` to input the data

I generally like labels, so we'll choose the STATA version

---

<sup>4</sup><https://haven.tidyverse.org>

The screenshot shows a web browser window for nces.ed.gov. At the top, the NCES logo and menu are visible. Below the header, the title "ONLINE CODEBOOK" is displayed next to a blue and green icon. On the right side of the header, there are links for "Switch Codebook", "Need Help?", "LOG IN", and "Feedback". A search bar with a magnifying glass icon and a "Go" button are also present. The main content area has a teal header bar with the text "ELS (Education Longitudinal Study of 2002)". Below this, the heading "Select Format" is centered. Under "Statistical Software Formats", there are five buttons: SPSS, SAS, S-Plus, Stata, and R. Under "Generic Formats", there are two buttons: ASCII and CSV. At the bottom left, a teal button labeled "BACK" with a left arrow icon is visible. The footer of the page contains the NCES logo and navigation links for IES, NCEC, NCEER, NCES, and various policies and standards.

After choosing your file version, you can finally download the files. Go ahead and click each box to download all the files.

The screenshot shows a web browser window for the NCES Online Codebook. The URL is nces.ed.gov. The page title is "IES > NCES National Center for Education Statistics". The main content area is titled "ELS (Education Longitudinal Study of 2002)" and features a section for "Download Data Files". It lists four download options:

- DOWNLOAD ELS 2002-I2 STUDENT FILE V1.0 STATA (zip file contains 1 file (16.0 MB))
- DOWNLOAD ELS 2002-I2 STUDENT BRR FILE V1.0 STATA (zip file contains 1 file (163.7 MB))
- DOWNLOAD ELS 2002-I2 OTHER FILES V1.0 STATA (zip file contains 3 files (932 KB))
- DOWNLOAD ELS 2002-I2 V1.0 CODEBOOK AND LAYOUT (zip file contains 6 files (1.2 MB))

At the bottom left is a "Back" button, and at the bottom right are links for "IES > NCES National Center for Education Statistics", "Explore the Institute of Education Sciences", "IES Centers", "IES Policies and Standards", "Public Access Policy", "Privacy and Security Policies", "NCES Statistical Standards", "Additional Resources", "ERIC", "Sitemap", and "Organizational Chart".

## NLS

The Bureau of Labor Statistics (aside from a lot of other useful information) has a number of National Longitudinal Surveys<sup>5</sup>. These are similar to those from the NCES, but much more expansive. They include:

- National Longitudinal Survey of Youth 1997 (NLSY97)
- National Longitudinal Survey of Youth 1979 (NLSY79)
- NLSY79 Children and Young Adults (NLSCYA)
- National Longitudinal Surveys of Young Women and Mature Women (NLSW)
- National Longitudinal Surveys of Young Men and Older Men (NLSM)

If you decide to use one of these surveys in your work, it will probably be the NLSY97, which began following a cohort of high schools students in 1997.

<sup>5</sup><https://www.bls.gov/nls/>

An official website of the United States government [Here is how you know](#)

bls.gov

United States Department of Labor

Follow Us [Twitter](#) | [Release Calendar](#) | [Blog](#)

Search BLS.gov

**U.S. BUREAU OF LABOR STATISTICS**

HOME ▾ SUBJECTS ▾ DATA TOOLS ▾ PUBLICATIONS ▾ ECONOMIC RELEASES ▾ STUDENTS ▾ BETA ▾

## National Longitudinal Surveys

NLS HOME

BROWSE NLS

NLS GENERAL OVERVIEWS

NLS NEWS RELEASES

NLS TABLES

NLS PUBLICATIONS

NLS FAQS

CONTACT NLS

SEARCH NLS

NLS TOPICS

NLSY97

NLSY79

NLSY79 CHILD & YOUNG ADULT

NLS ORIGINAL COHORTS

OBTAIN DATA

The National Longitudinal Surveys (NLS) are a set of surveys designed to gather information at multiple points in time on the labor market activities and other significant life events of several groups of men and women. NLS data have served as an important tool for economists, sociologists, and other researchers for more than 50 years. [Learn about the different NLS cohorts.](#)

**Announcement**

» [Redesign of various National Longitudinal Surveys webpages](#)

**On This Page**

» [NLS General Overviews](#)

» [NLS News Releases](#)

» [NLS Tables](#)

» [NLS Data](#)

» [NLS Publications](#)

» [NLS FAQs](#)

» [NLS Related Links](#)

» [Contact NLS](#)

**NLS General Overviews**

- [Introduction to the NLS Cohorts](#)– Information about the Bureau of Labor Statistics NLS Programs.
- Topics covered in the surveys:
  - [National Longitudinal Survey of Youth 1997 \(NLSY97\)](#) 
  - National Longitudinal Survey of Youth 1979 (NLSY79)
  - NLSY79 Children and Young Adults (NLSCYA)
  - National Longitudinal Survey of Young Women and Mature Women (NLSW)
  - National Longitudinal Survey of Young Men and Older Men (NLSM)
- [Accessing NLS Data](#)– Information on how to access public use and confidential data.
- [NLS Glossary](#)– This section provides glossary of terms that are important to the NLS cohorts.

National Longitudinal Survey of Youth Analysis Team. The NLS program celebrated its 50th anniversary with a conference on September 17, 2017.

## Investigator

Scrolling down on the NLS97 page, you'll see a section for **Accessing** the data via the *Investigator*. Because the NLSY is so large, you may choose to go this route.

**Accessing Data**

View [Accessing Data](#) for detailed information on accessing the NLSY97 public-use and the confidential files. The public-use NLS data are available on the [Investigator](#). The links below provide answers to frequently asked questions about requesting the NLSY97 Geocode, Zip-Code and Census Tract files, and the School Survey data.

- [How to Request Access to the NLSY79, NLSY79 Young Adult and NLSY97 Geocode data?](#)
- [How to Request Access to the NLSY Zip Code and Census Tract Files and NLSY97 School Survey Data?](#)

**Latest News Release**

- **May 05, 2020:** Labor Market Activity, Education, and Partner Status Among America's Young Adults at 33: Results from a Longitudinal Survey [NLSY97 Round 18] ([HTML](#)) ([PDF](#)) ([Supplemental Tables](#))
- [View Archived NLSY97 News Releases](#)
- [View News Releases Tables](#)

**Cross-Cohort NLSY79/97 Overview**

An NLSY79/NLSY97 cross-cohort beta release is now available on Investigator. This dataset harmonizes NLS data across two cohorts, allowing users added opportunities to perform cross-cohort comparisons. Initially the dataset includes interview date, reason for noninterview, age, marital status, highest grade attended, highest grade completed, and employment status, with possible expansion to over 100 variables. Fixed background variables are also included, such as sex, race, and AFQT score. See [Cross Cohort Harmonization Overview](#) document for more information. To access the data, go to <https://www.nlsinfo.org/investigator> and select "NLS Cross Cohort Beta."

**Last Modified Date:** May 05, 2020

RECOMMEND THIS PAGE USING:  Facebook  Twitter  LinkedIn

[Home](#) | [Subjects](#) | [Data Tools](#) | [Publications](#) | [Economic Releases](#) | [Students](#) | [Beta](#)



**U.S. BUREAU OF LABOR STATISTICS**

<b>INFO</b>	<b>RESOURCES</b>	<b>ABOUT THIS SITE</b>
<a href="#">What's New</a>	<a href="#">Inspector General (OIG)</a>	<a href="#">Sitemap</a>
<a href="#">FAQs</a>	<a href="#">Budget and Performance</a>	<a href="#">Freedom of Information Act</a>
<a href="#">A-Z</a>	<a href="#">No Fear Act</a>	<a href="#">Privacy &amp; Security Statement</a>
<a href="#">Glossary</a>	<a href="#">USA.gov</a>	<a href="#">Disclaimers</a>
<a href="#">About BLS</a>		<a href="#">Linking &amp; Copyright Info</a>
<a href="#">Careers @ BLS</a>		<a href="#">Important Website Notices</a>

You'll be shown a new external link. Click it to go to the data investigator.

NLSY97 Data Overview

Thanks for visiting us. You are now exiting the U.S. Bureau of Labor Statistics website : U.S. Bureau of Labor Statistics

An official website of the United States government [Here is how you know](#)

United States Department of Labor

Follow Us [Twitter](#) | [Release Calendar](#) | [Blog](#)

[Search BLS.gov](#)

[HOME](#) | [SUBJECTS](#) | [DATA TOOLS](#) | [PUBLICATIONS](#) | [ECONOMIC RELEASES](#) | [STUDENTS](#) | [BETA](#)

## BLS Information

[PRINT:](#)

Thanks for visiting us. You are now exiting the U.S. Bureau of Labor Statistics website

You are exiting the U.S. Bureau of Labor Statistics website. Thank you for visiting. Please click on the link below to continue.

<https://www.nlsinfo.org/investigator/pages/login>

RECOMMEND THIS PAGE USING:  Facebook  Twitter  LinkedIn

[Home](#) | [Subjects](#) | [Data Tools](#) | [Publications](#) | [Economic Releases](#) | [Students](#) | [Beta](#)



**U.S. BUREAU OF LABOR STATISTICS**

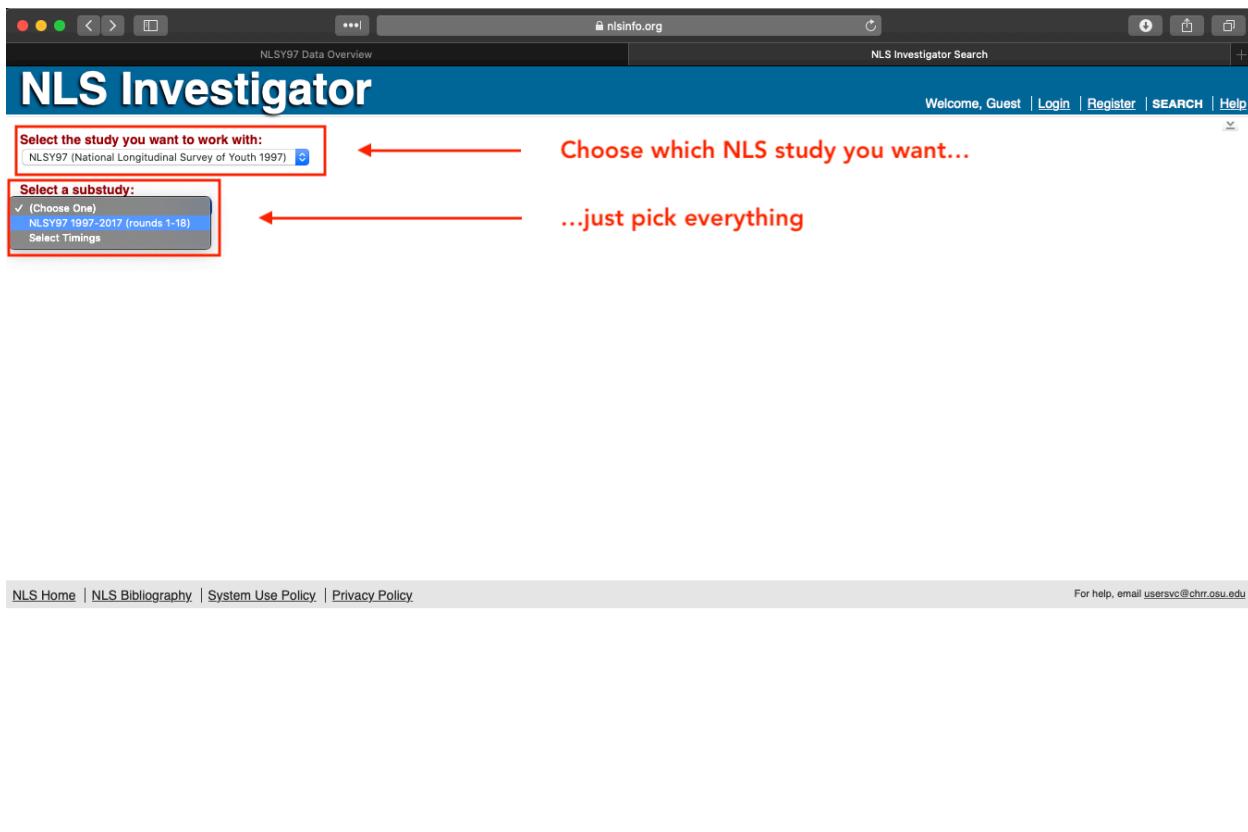
Postal Square Building  
2 Massachusetts Avenue NE  
Washington, DC 20212-0001

<b>INFO</b>	<b>RESOURCES</b>	<b>ABOUT THIS SITE</b>
<a href="#">What's New</a>	<a href="#">Inspector General (OIG)</a>	<a href="#">Sitemap</a>
<a href="#">FAQs</a>	<a href="#">Budget and Performance</a>	<a href="#">Freedom of Information Act</a>
<a href="#">A-Z</a>	<a href="#">No Fear Act</a>	<a href="#">Privacy &amp; Security Statement</a>
<a href="#">Glossary</a>	<a href="#">USA.gov</a>	<a href="#">Disclaimers</a>
<a href="#">About BLS</a>		<a href="#">Linking &amp; Copyright Info</a>
<a href="#">Careers @ BLS</a>		<a href="#">Important Website Notices</a>
<a href="#">Find It! DOL</a>		<a href="#">Help &amp; Tutorials</a>
<a href="#">Join our Mailing Lists</a>		

You can create a log in, which is nice if you come back often since you can save tag sets (variable groups you want to download), or just log in as a guest.

The screenshot shows the NLS Investigator login page. On the left, there is a 'Log In' form with fields for 'Username' and 'Password', and buttons for 'Forgot Password' and 'Login'. On the right, the main content area has a heading 'Welcome to NLS Investigator' and a paragraph about the National Longitudinal Surveys (NLS). It lists 'The seven NLS cohorts are:' followed by a bulleted list: National Longitudinal Survey of Youth 1997 (NLSY97), National Longitudinal Survey of Youth 1979 (NLSY79), NLSY79 Child and Young Adult, Older Men, Mature Women, Young Men, and Young Women. Below this, a note says 'To access data for any of the seven NLS cohorts use the login box to the left or begin searching as guest.' A red arrow points from the text 'You can log in, or just search as a guest' to the 'begin searching' link. At the bottom, there are links for 'NLS Home', 'NLS Bibliography', 'System Use Policy', 'Privacy Policy', and 'For help, email usersvc@chrr.osu.edu'.

Once inside, the investigator will allow you to choose which NLS data you want to access. So even though we got here via the NLSY97 page, we can still look at NLSY79 data if we want. Whichever you choose, go ahead and choose to look at all data rounds.



Now you can explore the data via the menu tree on the left side of the page. When you find a variable or set of variables you want to download, be sure to click the box next to the variable name. When you are finished, click the *Save/Download* tab.

Screenshot of the NLS Investigator interface showing the 'Review Selected Variables' screen.

**Select the study you want to work with:** NLSY97 (National Longitudinal Survey of Youth 1997)

**Select a substudy:** NLSY97 1997-2017 (rounds 1-18) Released December 05, 2019

**Additional Resources:** Errata, Documentation (user's guide, questionnaires and other materials), Custom Weights

To start a new search [click here](#)

**Click here when ready to download** (red arrow points to the 'Save / Download' button)

**Choose Tagsets Variable Search Review Selected Variables (6) Codebook Save / Download** (the 'Save / Download' button is highlighted with a red box)

**Browse Index Browse Index with Search Search**

**Index of Selected Variables**

- Education, Training & Achievement Scores (23402)
  - Education (18222)
    - Summary measures (518)
      - All schools (279)
        - Cumulative (15)
        - + By year (264)
      - + Primary & secondary school (124)
      - + College (115)
    - + Schooling history (1640)
    - + School-specific characteristics (roster) (16064)
  - Training (810)
    - + Cognitive tests (19)
    - + Transcript data (4458)
  - Employment (27430)
  - Household, Geography & Contextual Variables (9200)
  - Dating, Marriage & Cohabitation (1718)
    - + Sexual Activity, Pregnancy & Fertility (1749)
    - + Children (2387)
    - + Parents, Family Process & Childhood (187)

**QUESTION NAME VARIABLE TITLE YEAR**

RNUM	QUESTION NAME	VARIABLE TITLE	YEAR
1	Z9032200 CVC_GRADES_REPEAT_EVER_2007 # GRADES R HAS EVER REPEATED 2007		XRND
2	Z9032300 CVC_GRADE_SKIPPED_EVER_2007 # GRADES R EVER SKIPPED 2007		XRND
3	Z9083800 CVC_HGC_EVER	RS HIGHEST GRADE COMPLETED	XRND
4	Z9083900 CVC_HIGHEST_DEGREE_EVER	HIGHEST DEGREE RECEIVED	XRND
5	Z9084000 CVC_SCH_ATTEND_EVER	# SCHOOLS R ATTEND EVER	XRND
6	Z9084100 CVC_GED	MONTH R RECEIVED GED	XRND
7	Z9084200 CVC_HS_DIPLOMA	DATE RECEIVED HIGH SCHOOL DIPLOMA	XRND
8	Z9084300 CVC_AA_DEGREE	DATE RECEIVED ASSOCIATES DEGREE	XRND
9	Z9084400 CVC_BA_DEGREE	DATE RECEIVED BACHELORS DEGREE	XRND
10	Z9084500 CVC_PROF_DEGREE	DATE R RECEIVED PROFESSIONAL DEGREE	XRND
11	Z9084600 CVC_PHD_DEGREE	DATE R RECEIVED DOCTORAL DEGREE	XRND
12	Z9084700 CVC_MA_DEGREE	DATE R RECEIVED MASTERS DEGREE	XRND
13	Z9084800 CVC_HS_LEFT_DATE	DATE LEFT PRIMARY/SECONDARY SCHOOL	XRND
14	Z9084900 CVC_HS_LEFT_HGC	HIGHEST PRIMARY/SECONDARY GRADE COMPLETED	XRND
15	Z9085000 CVC_HS_LEFT_HGA	HIGHEST PRIMARY/SECONDARY GRADE ATTENDED	XRND

**Click in box to select variable** (red arrow points to the first column of the table)

On the next screen, you'll be able to save your tag set (meaning, not download the data, but name and keep the variable list you've chosen for a later date) or download.

Choose the *Advanced Download* tab. Within that tab, choose which file type you want to download (I've chosen just plain CSV here) and what you want to call the download. When you're ready, click the *download* button.

Screenshot of the NLS Investigator website showing the Advanced Download process.

**Select the study you want to work with:** NLSY97 (National Longitudinal Survey of Youth 1997)

**Select a substudy:** NLSY97 1997-2017 (rounds 1-18) Released December 05, 2019

**Additional Resources:** Errata, Documentation (user's guide, questionnaires and other materials), Custom Weights

**Choose Tagsets | Variable Search | Review Selected Variables (6) | Codebook | Save / Download**

**Save Tagset | Basic Download | Advanced Download (highlighted with a red box) | Manage Downloads**

**Customize your advanced download:**

**(1) Advanced Download:** Create Download of Data (radio button selected) →

- Tagset (list of selected variables)
- SAS® control file (includes the datafile of selected variables)
- SPSS® control file (includes the datafile of selected variables)
- STATA® dictionary file of selected variables
- R® Source code (includes the datafile of selected variables)
- Codebook of selected variables
- Short Description File
- Comma-delimited datafile of selected variables (to be read in Excel, etc.)

Column headers -- Use Reference Number Question Name (does not guarantee uniqueness)

**(2) Choose file type(s):** (arrow pointing to the file type checkboxes)

**(3) Name and download:** Filename: nlsy97\_subset → (arrow pointing to the filename input field)

Download status appears under 'Manage Downloads' tab.  
Downloads may be deleted after 10 days of inactivity.

NLS Home | NLS Bibliography | System Use Policy | Privacy Policy      For help, email usersvc@chrr.osu.edu

After your data set is prepared, you can download it to your computer.

Screenshot of the NLS Investigator website showing the Manage Downloads tab.

**Select the study you want to work with:** NLSY97 (National Longitudinal Survey of Youth 1997)

**Select a substudy:** NLSY97 1997-2017 (rounds 1-18) Released December 05, 2019

**Additional Resources:** Errata, Documentation (user's guide, questionnaires and other materials), Custom Weights

**Choose Tagsets | Variable Search | Review Selected Variables (6) | Codebook | Save / Download**

**Save Tagset | Basic Download | Advanced Download | Manage Downloads (highlighted with a red box)**

**Download Status:** All downloads are available. Please click a download link below to begin downloading.

**All Available Downloads:**

Date	Study	Name	Size	Download
2020-05-16 15:38:15	NLSY97	nlsy97_subset	37.1K	<a href="#">download</a> (highlighted with a red box)

[Delete Selected Files](#)

NLS Home | NLS Bibliography | System Use Policy | Privacy Policy      For help, email usersvc@chrr.osu.edu

## Direct download of full NLS data sets

Alternately, you can directly download the full NLS data files at [www.nlsinfo.org/accessing-data-cohorts<sup>6</sup>](https://www.nlsinfo.org/accessing-data-cohorts). I would recommend this approach if you think you're going to want a large number of variables. Also, you'll eventually find it easier to do your variable selection in R rather than via the Investigator.

The screenshot shows a web browser displaying the NLS website. The URL in the address bar is [nlsinfo.org](https://nlsinfo.org). The page title is "National Longitudinal Surveys". The main navigation menu includes "Getting Started", "Cohorts", and "Access Data / Investigator". The "Access Data / Investigator" tab is currently selected. On the left, there is a sidebar titled "HELPFUL LINKS" with links to "Subscribe for Updates", "Tutorials", "Investigator User Guide", "Investigator Video Tutorials", and "Glossary". The main content area is titled "Accessing Data / Cohorts". It contains text about building custom data sets in Investigator instead of downloading entire sets due to the large amount of NLS data. It lists three available cohorts: "NLS Youth 1997 (NLSY97) - updated February 14, 2020", "NLS Youth 1979 (NLSY79) - updated February 14, 2020", and "NLSY79 Child and Young Adult (NLSCYA) - updated February 14, 2020". Below this, a section titled "Currently unavailable:" lists four cohorts: "NLS Mature Women", "NLS Young Women", "NLS Older Men", and "NLS Young Men". At the bottom of the page, there is footer information for the National Longitudinal Survey and user services.

## IPEDS

For institution-level information, the Integrated Postsecondary Education Data System (IPEDS)<sup>7</sup> will likely be your first stop. While the IPEDS site ([nces.ed.gov/ipeds<sup>8</sup>](https://nces.ed.gov/ipeds)) will let you explore individual institutions or use a portal to select particular variables (like the BLS investigator), you'll want to just download the raw files. Begin by selecting *Use the Data*.

<sup>6</sup>[https://www.nlsinfo.org/accessing-data-cohrots](https://www.nlsinfo.org/accessing-data-cohorts)

<sup>7</sup><https://nces.ed.gov/ipeds/>

<sup>8</sup><https://nces.ed.gov/ipeds/>

IES > NCES National Center for Education Statistics ☰ MENU

IPEDS Integrated Postsecondary Education Data System

Search IPEDS 🔍 ⌂ ⓘ ⓘ ⓘ ⌂

Your primary source for information on U.S. colleges, universities, and technical and vocational institutions. Learn More

Find your college Use The Data Report Your Data Join In

IES > NCES National Center for Education Statistics

Explore the Institute of Education Sciences

IES Centers

IES Policies and Standards

Additional Resources

Home

NCEE

NCER

Public Access Policy

Privacy and Security Policies

NCEES Statistical Standards

ERIC

Sitemap

Organizational Chart

On the next page, look in the right column for the section *Survey Data*. From the drop down menu, choose *Complete data files*.

**NB:** If you know how to work with databases, the *Access databases* may be useful for you. But to use these, you either need a Microsoft Access license or a program to convert them to another format (like SQLite).

IES > NCES National Center for Education Statistics ≡ MENU

**IPEDS** Integrated Postsecondary Education Data System

Search IPEDS 🔍 Home Help User Profile FAQ ≡

IPEDS Data Use Help Desk (866) 558-0658 or [ipedstools@rti.org](mailto:ipedstools@rti.org)

## Use the Data

Access IPEDS data submitted to NCES through our data tools or download the data to conduct your research

**Data Explorer** 🔍  
Search for tables, charts, publications, or other products related to postsecondary education by keywords and filters.

**IPEDS Survey Components** 📊  
Learn more about the individual IPEDS survey components: view training videos, infographics, answers to frequently asked questions, survey forms, and more.

**Data Trends** 📈  
Use the Trend Generator to view trends on most frequently asked subject areas including: Enrollment, Completions, Graduation Rates, Employees and Staff, Institutional Revenues, and Financial Aid.

**Look Up an Institution** 🔍  
Look up information for one institution at a time. Data can be viewed in two forms: institution profile (similar to College Navigator) and reported data (institution's response to each survey question).

**Data Feedback Report** 📊  
Download, print, or customize an institution's Data Feedback Report, a report that graphically summarizes selected institutional data and compares the data with peer institutions.

**Statistical Tables** 📊  
Create simple descriptive statistics (e.g., total, count, average, median, standard deviation, percentiles) on selected IPEDS institutions and variables.

**Compare Institutions** 📊  
Download IPEDS data files for more than 7,000 institutions and up to 250 variables. Data files are provided in comma separated value (\*.csv) format.

**Survey Data** ⬇️ **(1) Choose Survey Data**  
Data are available starting with the 1980-81 collection year for the *Complete data files* and *Custom data files* functions, which zip the data into comma separated value (\*.csv). Beginning with the 2004-05 collection year, data for each collection year are compiled into an Access database.  
Select download option  
Access database  
**Complete data files** ? (2) Choose Complete data files  
Custom data files  
Select your shortcut

**Publications and Products** 📄  
Review publications using IPEDS data including First Looks, Web Tables, methodology reports, and Digest Tables.

**Resources** 📚  
[Overview of IPEDS Data](#)  
[IPEDS Glossary](#)  
[Archived Survey Forms](#)  
Delta Cost Project

You may see a popup window — if so, just agree. You'll now see a pretty lonely page. If you have a specific file or year you know you want, use the two drop down menus to filter your search. Otherwise, just click the *Continue* button to see your options.

The screenshot shows a web browser window for the IPEDS Data Center at [nces.ed.gov](http://nces.ed.gov). The main header includes the IES > NCES logo, the National Center for Education Statistics, and a menu icon. Below the header, the IPEDS logo and the Integrated Postsecondary Education Data System are displayed. The page title is "Complete Data Files". A navigation bar at the top right includes "Data Tools | Help Desk 1 866-558-0658", "Start over", "Save session", "Help", and a "MAIN MENU" button. A search bar with "Search" and "Go" buttons is also present. The main content area features a blue header bar with "Years & Surveys" and dropdown menus for "All years" and "All surveys". A red arrow points to a blue "Continue" button. Below this, a note states "Data files are available in ZIP format.". At the bottom of the page is a footer with links to various IES and NCES centers, policies, and additional resources.

Click the links in the *Data File* column to get zipped versions of the CSV files. If you want a Stata data file instead, choose the link from the *Stata Data File* column. You will probably want to grab the *Dictionary* file while you're at it.

*How do I know which file I need?*, you might be asking. If you are unsure, you may want to download the dictionary file first and check for the data element(s) you think you need. After a while, you'll get better at knowing (or reasonably guessing) which file is the one you need based on the names.

The screenshot shows the IPEDS Data Center interface. At the top, there's a navigation bar with links for 'Data Tools', 'Help Desk 1 866-558-0658', 'Start over', 'Save session', 'Help', and 'MAIN MENU'. Below this is a search bar with a 'Search' button and a 'Go' button.

The main content area is titled 'Complete Data Files' and has a 'Data Release Info' link. It features a blue header bar with 'Years & Surveys' and dropdown menus for 'All years' and 'All surveys', along with a 'Continue' button.

Below this is a table listing data files available in ZIP format for the year 2018. The columns include Year, Survey, Title, Data File, Stata File, Programs, and Dictionary. Several links in the table are highlighted with red boxes and arrows pointing to them:

- An arrow points to the 'Data File' column header with the text 'Choose CSV or STATA file'.
- An arrow points to the 'Dictionary' link in the last column with the text 'Get data dictionary'.
- Red boxes highlight the 'HD2018' and 'HD2018\_STATA' links under the first row.
- Red boxes highlight the 'IC2018' and 'IC2018\_STATA' links under the second row.
- Red boxes highlight the 'IC2018\_AY' and 'IC2018\_AY\_STATA' links under the third row.
- Red boxes highlight the 'IC2018\_PY' and 'IC2018\_PY\_STATA' links under the fourth row.
- Red boxes highlight the 'FLAGS2018' and 'FLAGS2018\_STATA' links under the fifth row.
- Red boxes highlight the 'EFFY2018' and 'EFFY2018\_STATA' links under the sixth row.
- Red boxes highlight the 'EFIA2018' and 'EFIA2018\_STATA' links under the seventh row.
- Red boxes highlight the 'FLAGS2018' and 'FLAGS2018\_STATA' links under the eighth row.
- Red boxes highlight the 'ADM2018' and 'ADM2018\_STATA' links under the ninth row.
- Red boxes highlight the 'FLAGS2018' and 'FLAGS2018\_STATA' links under the tenth row.
- Red boxes highlight the 'EF2018A' and 'EF2018A\_STATA' links under the eleventh row.
- Red boxes highlight the 'EF2018CP' and 'EF2018CP\_STATA' links under the twelfth row.

Year	Survey	Title	Data File	Stata File	Programs	Dictionary
2018	Institutional Characteristics	Directory information	<a href="#">HD2018</a>	<a href="#">HD2018_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	Institutional Characteristics	Educational offerings, organization, services and athletic associations	<a href="#">IC2018</a>	<a href="#">IC2018_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	Institutional Characteristics	Student charges for academic year programs	<a href="#">IC2018_AY</a>	<a href="#">IC2018_AY_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	Institutional Characteristics	Student charges by program (vocational programs)	<a href="#">IC2018_PY</a>	<a href="#">IC2018_PY_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	Institutional Characteristics	Response status for all survey components	<a href="#">FLAGS2018</a>	<a href="#">FLAGS2018_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	12-Month Enrollment	12-month unduplicated headcount: 2017-18	<a href="#">EFFY2018</a>	<a href="#">EFFY2018_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	12-Month Enrollment	12-month instructional activity: 2017-18	<a href="#">EFIA2018</a>	<a href="#">EFIA2018_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	12-Month Enrollment	Response status for all survey components	<a href="#">FLAGS2018</a>	<a href="#">FLAGS2018_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	Admissions and Test Scores	Admission considerations, applications, admissions, employees and test scores, fall 2018	<a href="#">ADM2018</a>	<a href="#">ADM2018_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	Admissions and Test Scores	Response status for all survey components	<a href="#">FLAGS2018</a>	<a href="#">FLAGS2018_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	Fall Enrollment	Race/ethnicity, gender, attendance status, and level of student: Fall 2018	<a href="#">EF2018A</a>	<a href="#">EF2018A_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>
2018	Fall Enrollment	Major field of study, race/ethnicity, gender, attendance status, and	<a href="#">EF2018CP</a>	<a href="#">EF2018CP_STATA</a>	SPSS, SAS, STATA	<a href="#">Dictionary</a>

## Download all of IPEDS via R

If you don't want to bother with the portal, I've written an R script that will download the entirety of IPEDS to your computer (a little over 1 GB if you only want one type of data file). See [github.com/btskinner/downloadipeds<sup>9</sup>](https://github.com/btskinner/downloadipeds) for the script and information on how to use it.

## College Scorecard

Though it's intended to give students and their families better information about their college options, the College Scorecard<sup>10</sup> offers data that's useful for research. In particular, you can find earnings data linked to schools and programs that you can't find anywhere else.

## Direct

If you go to the College Scorecard homepage ([collegescorecard.ed.gov<sup>11</sup>](https://collegescorecard.ed.gov)), you'll see the portal that students use. Scroll to the bottom of the page.

<sup>9</sup><https://github.com/btskinner/downloadipeds>

<sup>10</sup><https://collegescorecard.ed.gov>

<sup>11</sup><https://collegescorecard.ed.gov>

The screenshot shows the homepage of the College Scorecard. At the top, there's a navigation bar with icons for back, forward, and search. The URL is college scorecard.ed.gov. Below the bar, the U.S. Department of Education logo and the text "College Scorecard" are displayed. A large blue banner features the slogan "FIND THE RIGHT FIT." and the subtext "Find out about colleges: their programs, costs, admissions, results, and more." Below the banner are three search options: "CUSTOM SEARCH", "NAME SEARCH" (which is highlighted in blue), and "SHOW ME OPTIONS". A search input field with a magnifying glass icon and the placeholder "Type to search" is centered. The main content area has two columns. The left column, titled "Alternative Pathways to a Career", contains text about apprenticeships and a button labeled "FIND ONE THAT'S RIGHT FOR YOU!". The right column, titled "Learn More About Paying for College", contains text about financial aid and a button labeled "START MY FAFSA® FORM".

At the bottom of the page, you'll see a link to download the data files that power the Scorecard.

The screenshot shows the same homepage as above, but with a red arrow pointing from the text "Click to access full data files" on the left side towards the "Powered by" link at the bottom of the page. The "Powered by" link is enclosed in a red box. The rest of the page content is identical to the first screenshot.

On the data page, you can download the full set of files or just the latest data. Unless you have a good reason to do otherwise, I would recommend getting all the data. You may also want to follow the *Documentation* tab to get the data documentation.

U.S. DEPARTMENT OF EDUCATION  
College Scorecard

DATA HOME DOCUMENTATION CHANGE LOG GLOSSARY

**Download the Data** You may want to get documentation, too

This data was last updated March 30, 2020.

**All Data Files** • Click to download files

289 MB ZIP

Most Recent Institution-Level Data

148 MB CSV

Most Recent Data by Field of Study

51 MB CSV

Looking for help? Visit [StackExchange](#) or email the help desk at [scorecarddata@rti.org](mailto:scorecarddata@rti.org) For inquiries by members of the press, please contact [press@ed.gov](mailto:press@ed.gov).

Powered by [College Scorecard Data](#) | v2.2.1 | [Contact Us](#) | [Notices](#) | [Glossary](#) | [U.S. Department of Education](#)  
ZIP Code latitude and longitude provided by [GeoNames](#) under a Creative Commons Attribution 3.0 License.

## rscorecard

You can also download College Scorecard data directly from R using the **rscorecard** package, which accesses Scorecard data via an API. See [btskinner.io/rscorecard](http://btskinner.io/rscorecard)<sup>12</sup> for more information and examples.

## American Community Survey (ACS)

The American Community Survey (ACS)<sup>13</sup> is part of the U.S. Census that, unlike the decennial census, collects data each year. While it has information on education that you may want to use directly, the ACS is also a great source for place-based data that you can merge with other data sets (student-level data, for example, if you know where they live). The ACS homepage is here: [census.gov/programs-surveys/acs](http://census.gov/programs-surveys/acs)<sup>14</sup>.

There are a few ways to access ACS data. I will show you how to get the public use micro sample (PUMS) data. From the ACS home page, click on the *Data* link on the left.

<sup>12</sup><https://www.btskinner.io/rscorecard/>

<sup>13</sup><https://www.census.gov/programs-surveys/acs>

<sup>14</sup><https://www.census.gov/programs-surveys/acs>

United States Census Bureau

BROWSE BY TOPIC EXPLORE DATA LIBRARY SURVEYS/ PROGRAMS INFORMATION FOR... FIND A CODE ABOUT US

// Census.gov > Our Surveys & Programs > American Community Survey (ACS)

## American Community Survey (ACS)

The American Community Survey (ACS) helps local officials, community leaders, and businesses understand the changes taking place in their communities. It is the premier source for detailed population and housing information about our nation.

**Data**

Guidance for Data Users

Geography & ACS

Technical Documentation

Methodology

Library

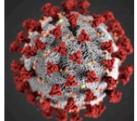
Operations and Administration

Contact Us

< Back to Our Surveys & Programs

 Respond to the ACS Learn how









How do I respond to the ACS?

ACS data help with COVID-19 response efforts

Should I respond to both the ACS and 2020 Census?

**Data**

**2018 Data Release**

Learn more about 2018 ACS data releases, including the latest release of new 2014-2018 ACS 5-year PUMS files and Variance Replicate Estimates, as well as the new 2018 ACS 1-year Supplemental Estimates.

Is this page helpful?

On the next screen, click on the *PUMS* link which again is on the left.

United States Census Bureau

BROWSE BY TOPIC EXPLORE DATA LIBRARY SURVEYS/ PROGRAMS INFORMATION FOR... FIND A CODE ABOUT US

// Census.gov > Our Surveys & Programs > American Community Survey (ACS) > American Community Survey Data

## American Community Survey Data

**PUMS Data**

New Data Every Year!

The American Community Survey collects data on an ongoing basis, January through December, to provide every community with the information they need to make important decisions

We release new data every year, in the form of estimates, in a variety of tables, tools, and analytical reports.

**Data Tables and Tools**

The Data Tables and Tools pages provide explanations and links to our most popular data tables on data.census.gov.

**Data Profiles**

**Narrative Profiles**

**Subject Tables**

Related Information

Federal Statistical Research Data Centers

Available APIs

Contact Us

You May Be Interested In

RELATED TOPICS

Contact Us

AROUND THE BUREAU

Email Updates

MOST POPULAR

Share Your ACS Data Story

Is this page helpful?

There are a couple of ways to get PUMS data: from the old FTP site or the newer data.census.gov site. Though it's less pretty, I'll show you the FTP version (if you have used FTP applications before to access data, you can use those here).

The screenshot shows a web browser window for census.gov. The left sidebar has a tree view under 'AMERICAN COMMUNITY SURVEY (ACS)' with nodes like 'Data Tables & Tools', 'Data via FTP', 'Summary File Data', and 'PUMS Data' (which is selected). Below that are links for 'Variance Replicate Tables', 'Race/Ethnicity and American Indian & Alaska Native Data', 'Custom Tables', and a 'Back to American Community Survey Data' link. There's also a 'Respond to the ACS' section with a 'Learn how' link. The main content area has a search bar and navigation links for 'EXPLORE DATA', 'LIBRARY', 'SURVEYS/ PROGRAMS', 'INFORMATION FOR...', 'FIND A CODE', and 'ABOUT US'. The breadcrumb trail shows: Census.gov > Our Surveys & Programs > American Community Survey (ACS) > American Community Survey Data > PUMS Data. The main title is 'PUMS Data'. Below it, a note says 'Supporting documentation for the data below is available on the [PUMS Documentation](#) page.' A sidebar on the right is titled 'Related Information' with links for 'Share Your ACS Data Story', 'Contact Us', 'Introduction to the Public Use Microdata Sample (PUMS) File', and 'You May Be Interested In' sections. The central part shows two columns of file names: 'Available on the FTP site' (2018, 2017, 2016, 2015, 2014, 2013) and 'PUMS on data.census.gov' (2014-2018, 2018, 2013-2017, 2017, 2012-2016, 2016). A red box highlights the 2018 ACS PUMS link. At the bottom right is a feedback box: 'Is this page helpful? X' with 'Yes' and 'No' buttons.

Available on the FTP site	PUMS on data.census.gov
2018 ACS PUMS	2014-2018 ACS 5-year PUMS
2017 ACS PUMS	2018 ACS 1-year PUMS
2016 ACS PUMS	2013-2017 ACS 5-year PUMS
2015 ACS PUMS	2017 ACS 1-year PUMS
2014 ACS PUMS	2012-2016 ACS 5-year PUMS
2013 ACS PUMS	2016 ACS 1-year PUMS

You'll notice the FTP page looks like a file system. That's basically what it is. Click on the file name you want. For more information on whether you want 1-, 3-, or 5-year estimates, check out this page: [census.gov/programs-surveys/acs/guidance/estimates.html<sup>15</sup>](https://www.census.gov/programs-surveys/acs/guidance/estimates.html).

<sup>15</sup><https://www.census.gov/programs-surveys/acs/guidance/estimates.html>

The screenshot shows a Mac OS X desktop environment with a browser window open to the [www2.census.gov](http://www2.census.gov) website. The page displays a file directory under 'Parent Directory' with two entries: '1-Year/' (modified 12-Nov-2019 08:34) and '5-Year/' (modified 29-Jan-2020 08:26). Below the directory is a navigation menu with several sections: ABOUT US, FIND DATA, BUSINESS & ECONOMY, PEOPLE & HOUSEHOLDS, SPECIAL TOPICS, and NEWSROOM. Each section contains a list of links related to its category. At the bottom of the page, there is a 'CONNECT WITH US' section featuring social media icons for Facebook, Twitter, LinkedIn, YouTube, and other platforms, along with links to Accessibility, Information Quality, FOIA, Data Protection and Privacy Policy, and the U.S. Department of Commerce.

---

On the final page, you can choose data at the state level. There are two basic types of files, each pertaining to sections of the survey:

- **Housing:** These files start with *h* after the underscore
- **Person:** These files start with *p* after the underscore

For more information about which file to use or PUMS more generally, visit [census.gov/programs-surveys/acs/technical-documentation/pums/about.html<sup>16</sup>](https://www.census.gov/programs-surveys/acs/technical-documentation/pums/about.html).

<sup>16</sup><https://www.census.gov/programs-surveys/acs/technical-documentation/pums/about.html>

Choose your state file(s) ("h" for housing; "p" for person")

- "csv\_hak.zip" = Housing file for Alaska
- "csv\_pak.zip" = Person file for Alaska

Name	Last modified	Size	Description
Parent Directory		-	
PUMS_file_naming_convention.pdf	12-Nov-2019 08:34	36K	
csv_hak.zip	31-Oct-2019 15:19	709K	
csv_hal.zip	31-Oct-2019 15:19	3.8M	
csv_har.zip	31-Oct-2019 15:19	2.4M	
csv_haz.zip	31-Oct-2019 15:19	5.0M	
csv_hca.zip	31-Oct-2019 15:19	23M	
csv_hco.zip	31-Oct-2019 15:19	4.1M	
csv_hct.zip	31-Oct-2019 15:18	2.7M	
csv_hdc.zip	31-Oct-2019 15:18	727K	
csv_hde.zip	31-Oct-2019 15:18	921K	
csv_hfl.zip	31-Oct-2019 15:18	15M	
csv_hga.zip	31-Oct-2019 15:18	7.2M	
csv_hhi.zip	31-Oct-2019 15:18	1.1M	
csv_hia.zip	31-Oct-2019 15:18	2.5M	
csv_hid.zip	31-Oct-2019 15:18	1.4M	
csv_hil.zip	31-Oct-2019 15:18	8.9M	
csv_hin.zip	31-Oct-2019 15:18	4.9M	
csv_hks.zip	31-Oct-2019 15:18	2.3M	
csv_hky.zip	31-Oct-2019 15:17	3.4M	
csv_hla.zip	31-Oct-2019 15:17	3.5M	
csv_hma.zip	31-Oct-2019 15:17	4.9M	
csv_hmd.zip	31-Oct-2019 15:17	4.2M	
csv_hme.zip	31-Oct-2019 15:17	1.4M	
csv_hmi.zip	31-Oct-2019 15:17	7.6M	
csv_hmn.zip	31-Oct-2019 15:17	4.1M	
csv_hmo.zip	31-Oct-2019 15:17	4.7M	
csv_hms.zip	31-Oct-2019 15:17	2.3M	
csv_hmt.zip	31-Oct-2019 15:17	1.0M	

## Other

Below are some other data sources you may find useful, either on their own or joined with the data sets above.

- Current Population Survey ([www.census.gov/programs-surveys/cps.html](https://www.census.gov/programs-surveys/cps.html))<sup>17</sup>
- TIGER/Line Shapefiles ([www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html](https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html))<sup>18</sup>
- Bureau of Labor Statistics ([www.bls.gov/data/](https://www.bls.gov/data/))<sup>19</sup>
- Delta Cost Project ([deltacostproject.org/delta-cost-data](https://deltacostproject.org/delta-cost-data))<sup>20</sup>
- Urban Institute Data Explorer ([educationdata.urban.org/data-explorer/](https://educationdata.urban.org/data-explorer/))<sup>21</sup>
- PISA ([www.oecd.org/pisa/data/](https://www.oecd.org/pisa/data/))<sup>22</sup>

<sup>17</sup><https://www.census.gov/programs-surveys/cps.html>

<sup>18</sup><https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>

<sup>19</sup><https://www.bls.gov/data/>

<sup>20</sup><https://deltacostproject.org/delta-cost-data>

<sup>21</sup><https://educationdata.urban.org/data-explorer/>

<sup>22</sup><https://www.oecd.org/pisa/data/>