

# 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 NCES website homepage. At the top, there is a navigation bar with the IES and NCES logos, a search bar, and a menu icon. Below this is a large banner for the 'Digest of Education Statistics', which includes a photo of students in a classroom and text stating it is the 51st edition containing statistics from prekindergarten to graduate school. Underneath the banner are five navigation links: 'The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education. Learn more about NCES', 'About Us', 'Surveys and Programs', 'Publications and Products', and 'Data and Tools'. The main content area is divided into two sections. The first, 'WHAT'S NEW AT NCES', lists four recent updates with dates and titles: 'MAY 13 A Look at Postsecondary Certificates and Associate's Degrees in Occupational Fields of Study', 'MAY 7 Highlights from the PISA 2018 Financial Literacy Assessment', 'APR 30 School Nurses in U.S. Public Schools', and 'APR A look at the postsecondary enrollment of students who took career and'. The second section, 'THE CONDITION OF EDUCATION', features a circular graphic with various educational icons and a brief description of the indicators.

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

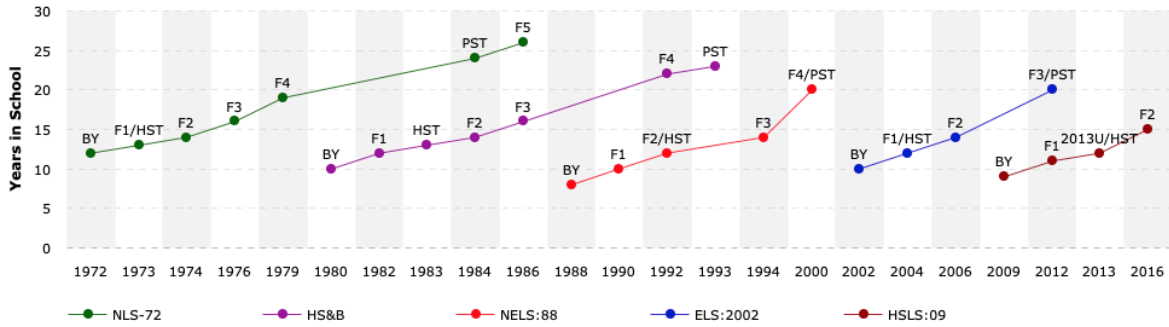
<sup>1</sup><https://nces.ed.gov>

## Historical Background

HLSL: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  
 HLSL: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](https://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 website interface. At the top, the navigation bar includes the NCES logo, a search bar, and a menu icon labeled (1). A left sidebar contains a vertical menu with items: Surveys & Programs (3), Data & Tools (2), Fast Facts, News & Events, Publications & Products, and About Us. The main content area is titled "Downloads Microdata/Raw Data" and features a sub-menu with "EDAT" (4), "Delta Cost Project", "IPEDS Data Center", and "How to apply for Restricted Use License". A right sidebar contains contact information, social media links for Twitter and Facebook, and a "KIDS' ZONE" logo. Below the navigation bar, a footer section provides a brief description of NCES and links to "About Us", "Surveys and Programs", "Publications and Products", and "Data and Tools".

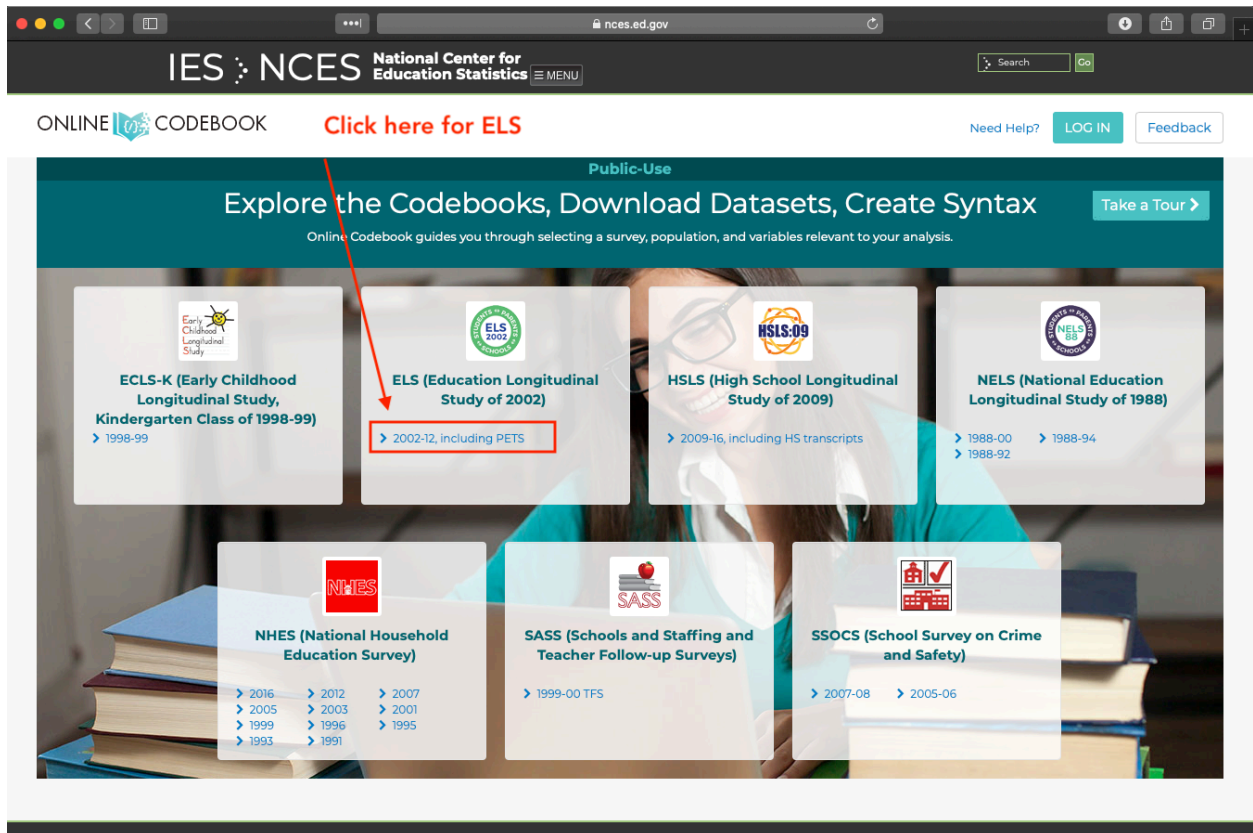
**WHAT'S NEW AT NCES**

- 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** [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.



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.

IES NCES National Center for Education Statistics

ONLINE CODEBOOK

Switch Codebook Need Help? LOG IN Feedback

ELS (Education Longitudinal Study of 2002)

Files Variable Types Keywords Options

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

IES NCES National Center for Education Statistics

Explore the Institute of Education Sciences

IES Policies and Standards

Additional Resources

IES Home About Publications Data

IES Centers NCEE NCER NCES Home

Public Access Policy Privacy and Security Policies NCES Statistical Standards Peer Review Process ED Data Inventory

ERIC Sitemap Organizational Chart

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 with the URL [nces.ed.gov](https://nces.ed.gov). The page header includes the IES NCES logo and the text "National Center for Education Statistics". Below the header, there is a navigation bar with "ONLINE CODEBOOK" and links for "Switch Codebook", "Need Help?", "LOG IN", and "Feedback". A teal banner at the top of the main content area reads "ELS (Education Longitudinal Study of 2002)".

The main content area is titled "Select Format" and is divided into two sections:

- Statistical Software Formats:** This section contains five buttons: "SPSS", "SAS", "S-Plus", "Stata", and "R".
- Generic Formats:** This section contains two buttons: "ASCII" and "CSV".

At the bottom of the main content area, there is a teal button labeled "BACK".

The footer of the page is dark and contains the IES NCES logo and the text "National Center for Education Statistics". It is organized into four columns of links:

- Explore the Institute of Education Sciences:** IES, Home, About, Publications, Data.
- IES Centers:** NCEE, NCER, NCES, Home.
- IES Policies and Standards:** Public Access Policy, Privacy and Security Policies, NCES Statistical Standards, Peer Review Process, ED Data Inventory.
- Additional Resources:** ERIC, Sitemap, Organizational Chart.

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

IES NCES National Center for Education Statistics





ONLINE CODEBOOK

Switch Codebook Need Help? LOG IN Feedback

### ELS (Education Longitudinal Study of 2002)

#### Download Data Files

Stata Formatted Data Files ▾

- DOWNLOAD ELS 2002-12 STUDENT FILE V1.0 STATA   
zip file contains 1 file (16.0 MB)
- DOWNLOAD ELS 2002-12 STUDENT BRR FILE V1.0 STATA   
zip file contains 1 file (163.7 MB)
- DOWNLOAD ELS 2002-12 OTHER FILES V1.0 STATA   
zip file contains 3 files (932 KB)
- DOWNLOAD ELS 2002-12 V1.0 CODEBOOK AND LAYOUT   
zip file contains 6 files (1.2 MB)

[← Back](#)

IES NCES National Center for Education Statistics

Explore the Institute of Education Sciences

- IES
- Home
- About

IES Centers

- NCEE
- NCER

IES Policies and Standards

- Public Access Policy
- Privacy and Security Policies
- NCES Statistical Standards

Additional Resources

- ERIC
- Sitemap
- 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/>

The screenshot shows the U.S. Bureau of Labor Statistics website. The main heading is "National Longitudinal Surveys". On the left, there is a sidebar with "BROWSE NLS" and "SEARCH NLS" sections. The main content area includes an "Announcement" about the redesign of various NLS web pages, an "On This Page" section with links to "NLS General Overviews", "NLS News Releases", "NLS Tables", "NLS Data", "NLS Publications", "NLS FAQs", "NLS Related Links", and "Contact NLS". Below this is the "NLS General Overviews" section, which contains a list of topics. One item, "National Longitudinal Survey of Youth 1997 (NLSY97)", is highlighted with a red box and a red arrow pointing to it.

## 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**

**INFO**  
[What's New](#)  
[FAQs](#)  
[A-Z](#)  
[Glossary](#)  
[About BLS](#)  
[Careers @ BLS](#)

**RESOURCES**  
[Inspector General \(OIG\)](#)  
[Budget and Performance](#)  
[No Fear Act](#)  
[USA.gov](#)

**ABOUT THIS SITE**  
[Sitemap](#)  
[Freedom of Information Act](#)  
[Privacy & Security Statement](#)  
[Disclaimers](#)  
[Linking & Copyright Info](#)  
[Important Website Notices](#)


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

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

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

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

**INFO**  
[What's New](#)  
[FAQs](#)  
[A-Z](#)  
[Glossary](#)  
[About BLS](#)  
[Careers @ BLS](#)  
[Find It! DOL](#)  
[Join our Mailing Lists](#)

**RESOURCES**  
[Inspector General \(OIG\)](#)  
[Budget and Performance](#)  
[No Fear Act](#)  
[USA.gov](#)

**ABOUT THIS SITE**  
[Sitemap](#)  
[Freedom of Information Act](#)  
[Privacy & Security Statement](#)  
[Disclaimers](#)  
[Linking & Copyright Info](#)  
[Important Website Notices](#)  
[Help & Tutorials](#)

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.

**NLS Investigator** Welcome, Guest | [LOGIN](#) | [Register](#) | [Search](#) | [Help](#)

**Log In**  
Username:   
Password:   
[Forgot Password](#)

**Welcome to NLS Investigator**

Sponsored by the Bureau of Labor Statistics, the National Longitudinal Surveys (NLS) are a family of surveys dedicated to tracking the labor market and other life experiences of American men and women.

The seven NLS cohorts are:

- 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
- Young Women

To access data for any of the seven NLS cohorts use the login box to the left or [begin searching](#) as guest.

**NLSY User-Initiated Questions:** We're soliciting suggestions for new questions to add to the NLSY97, NLSY79, and child/young adult surveys. Please visit the [NLSY User-Initiated Questions](#) page to learn how to make an informal suggestion or submit a formal proposal.

[NLS Home](#) | [NLS Bibliography](#) | [System Use Policy](#) | [Privacy Policy](#) For help, email [usersvc@chr.osu.edu](mailto:usersvc@chr.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.

NLS Investigator

Welcome, Guest | [Login](#) | [Register](#) | [SEARCH](#) | [Help](#)

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

Select a substudy:  
✓ (Choose One)  
NLSY97 1997-2017 (rounds 1-18)  
Select Timings

Choose which NLS study you want...

...just pick everything

[NLS Home](#) | [NLS Bibliography](#) | [System Use Policy](#) | [Privacy Policy](#) | For help, email [usersvc@chr.osu.edu](mailto:usersvc@chr.osu.edu)

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.

NLS Investigator

Welcome, Guest | Login | Register | SEARCH | Help

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:  
[Errors, Documentation](#) (user's guide, questionnaires and other materials)  
[Custom Weights](#)

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

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

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)

Options | Showing 15 of 15 filtered by Primary Variables | 500 | 1

|                          | RNUM     | QUESTION NAME               | VARIABLE TITLE                            | YEAR |
|--------------------------|----------|-----------------------------|-------------------------------------------|------|
| <input type="checkbox"/> | Z9032200 | CVC_GRADES_REPEAT_EVER_2007 | # GRADES R HAS EVER REPEATED 2007         | XRND |
| <input type="checkbox"/> | Z9032300 | CVC_GRADE_SKIPPED_EVER_2007 | # GRADES R EVER SKIPPED 2007              | XRND |
| <input type="checkbox"/> | Z9083800 | CVC_HGC_EVER                | RS HIGHEST GRADE COMPLETED                | XRND |
| <input type="checkbox"/> | Z9083900 | CVC_HIGHEST_DEGREE_EVER     | HIGHEST DEGREE RECEIVED                   | XRND |
| <input type="checkbox"/> | Z9084000 | CVC_SCH_ATTEND_EVER         | # SCHOOLS R ATTEND EVER                   | XRND |
| <input type="checkbox"/> | Z9084100 | CVC_GED                     | MONTH R RECEIVED GED                      | XRND |
| <input type="checkbox"/> | Z9084200 | CVC_HS_DIPLOMA              | DATE RECEIVED HIGH SCHOOL DIPLOMA         | XRND |
| <input type="checkbox"/> | Z9084300 | CVC_AA_DEGREE               | DATE RECEIVED ASSOCIATES DEGREE           | XRND |
| <input type="checkbox"/> | Z9084400 | CVC_BA_DEGREE               | DATE RECEIVED BACHELORS DEGREE            | XRND |
| <input type="checkbox"/> | Z9084500 | CVC_PROF_DEGREE             | DATE R RECEIVED PROFESSIONAL DEGREE       | XRND |
| <input type="checkbox"/> | Z9084600 | CVC_PHD_DEGREE              | DATE R RECEIVED DOCTORAL DEGREE           | XRND |
| <input type="checkbox"/> | Z9084700 | CVC_MA_DEGREE               | DATE R RECEIVED MASTERS DEGREE            | XRND |
| <input type="checkbox"/> | Z9084800 | CVC_HS_LEFT_DATE            | DATE LEFT PRIMARY/SECONDARY SCHOOL        | XRND |
| <input type="checkbox"/> | Z9084900 | CVC_HS_LEFT_HGC             | HIGHEST PRIMARY/SECONDARY GRADE COMPLETED | XRND |
| <input type="checkbox"/> | Z9085000 | CVC_HS_LEFT_HGA             | HIGHEST PRIMARY/SECONDARY GRADE ATTENDED  | XRND |

500 | 1

Click here when ready to download

Click in box to select variable

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.

The screenshot shows the 'Advanced Download' section of the NLS Investigator interface. The page title is 'NLS Investigator' and the user is logged in as 'Guest'. The main navigation bar includes 'Choose Tagsets', 'Variable Search', 'Review Selected Variables (6)', 'Codebook', and 'Save / Download'. The 'Advanced Download' tab is selected, and the 'Advanced Download' button is highlighted with a red box. The page contains several sections: 'Customize your advanced download:', 'Create Download of Data' (with sub-options for Tagset, SAS, SPSS, STATA, R, Codebook, Short Description File, and Comma-delimited datafile), 'Create Frequency / Table', and 'Apply Universe Restrictors'. A 'Filename' field contains 'nlsy97\_subset' and has a 'Download' button next to it. Red arrows point to the 'Advanced Download' button, the 'Codebook of selected variables' checkbox, and the 'Download' button. The filename field also has a red box around it. The footer includes 'NLS Home', 'NLS Bibliography', 'System Use Policy', 'Privacy Policy', and contact information for usersvc@chr.osu.edu.

**(1) Advanced Download**

**(2) Choose file type(s)**

**(3) Name and download**

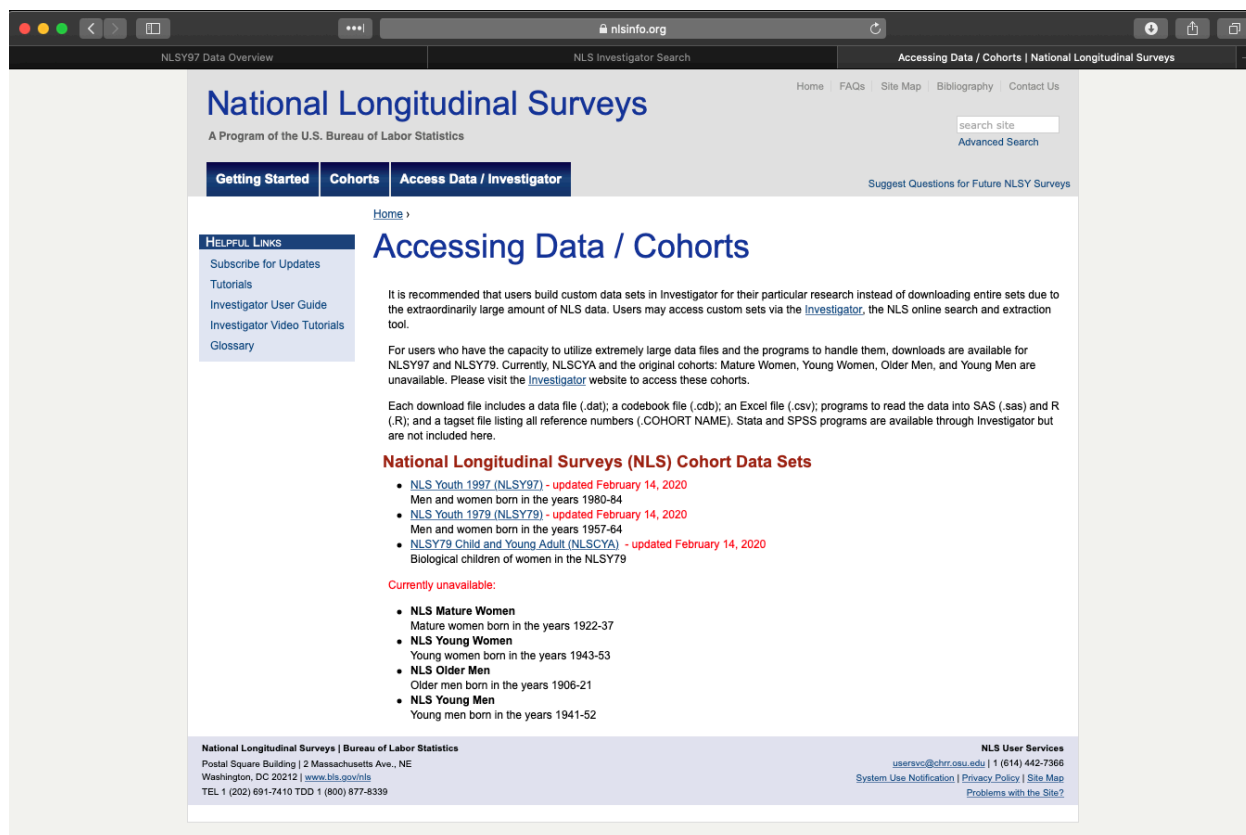
After your data set is prepared, you can download it to your computer.

The screenshot shows the 'Download Status' section of the NLS Investigator interface. The page title is 'NLS Investigator' and the user is logged in as 'Guest'. The main navigation bar includes 'Choose Tagsets', 'Variable Search', 'Review Selected Variables (6)', 'Codebook', and 'Save / Download'. The 'Advanced Download' tab is selected, and the 'Download Status' button is highlighted with a red box. The page contains a 'Download Status' section with the text 'All downloads are available. Please click a download link below to begin downloading.' and an 'All Available Downloads' section with a table. The table has columns for 'Date', 'Study', 'Name', 'Size', and 'Download'. The first row shows a download for 'NLSY97' with a size of '37.1K' and a 'download' link. A red arrow points to the 'download' link. The footer includes 'NLS Home', 'NLS Bibliography', 'System Use Policy', 'Privacy Policy', and contact information for usersvc@chr.osu.edu.

|   | Date                | Study  | Name          | Size  | Download |
|---|---------------------|--------|---------------|-------|----------|
| 1 | 2020-05-16 15:38:15 | NLSY97 | nlsy97_subset | 37.1K | download |

## Direct download of full NLS data sets

Alternately, you can directly download the full NLS data files at [www.nlsinfo.org/accessing-data-cohorts](http://www.nlsinfo.org/accessing-data-cohorts)<sup>6</sup>. 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 the website for the National Longitudinal Surveys (NLS). The page title is "Accessing Data / Cohorts". It features a navigation menu with "Getting Started", "Cohorts", and "Access Data / Investigator". A sidebar on the left lists "HELPFUL LINKS" including "Subscribe for Updates", "Tutorials", "Investigator User Guide", "Investigator Video Tutorials", and "Glossary". The main content area includes a "Home" link, a search bar, and a section titled "National Longitudinal Surveys (NLS) Cohort Data Sets". This section lists several cohorts with their update dates and descriptions:

- **NLS Youth 1997 (NLSY97)** - updated February 14, 2020  
Men and women born in the years 1980-84
- **NLS Youth 1979 (NLSY79)** - updated February 14, 2020  
Men and women born in the years 1957-64
- **NLSY79 Child and Young Adult (NLSCYA)** - updated February 14, 2020  
Biological children of women in the NLSY79

Below this, a section titled "Currently unavailable:" lists:

- **NLS Mature Women**  
Mature women born in the years 1922-37
- **NLS Young Women**  
Young women born in the years 1943-53
- **NLS Older Men**  
Older men born in the years 1906-21
- **NLS Young Men**  
Young men born in the years 1941-52

At the bottom, there is contact information for the Bureau of Labor Statistics and NLS 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/](http://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-cohorts>

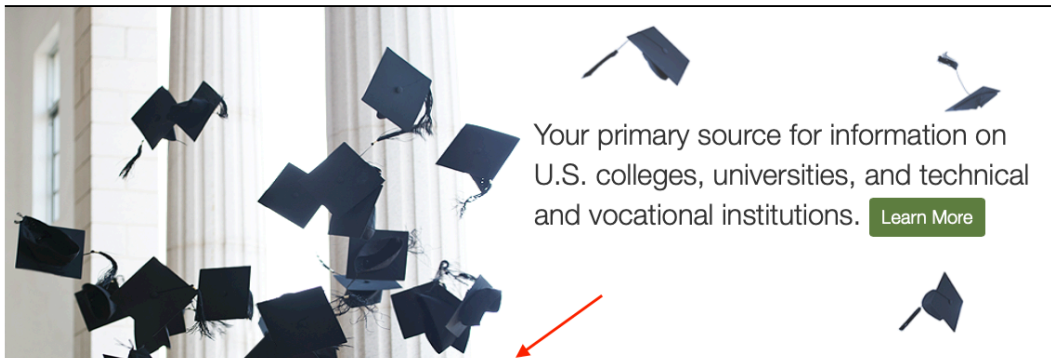
<sup>7</sup><https://nces.ed.gov/ipeds/>

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

# IPEDS

Integrated Postsecondary Education Data System

Search IPEDS



Find your college Use The Data Report Your Data Join In

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).

nces.ed.gov

IES : NCES National Center for Education Statistics MENU

Search IPEDS

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

## IPEDS Integrated Postsecondary Education Data System

### Use the Data

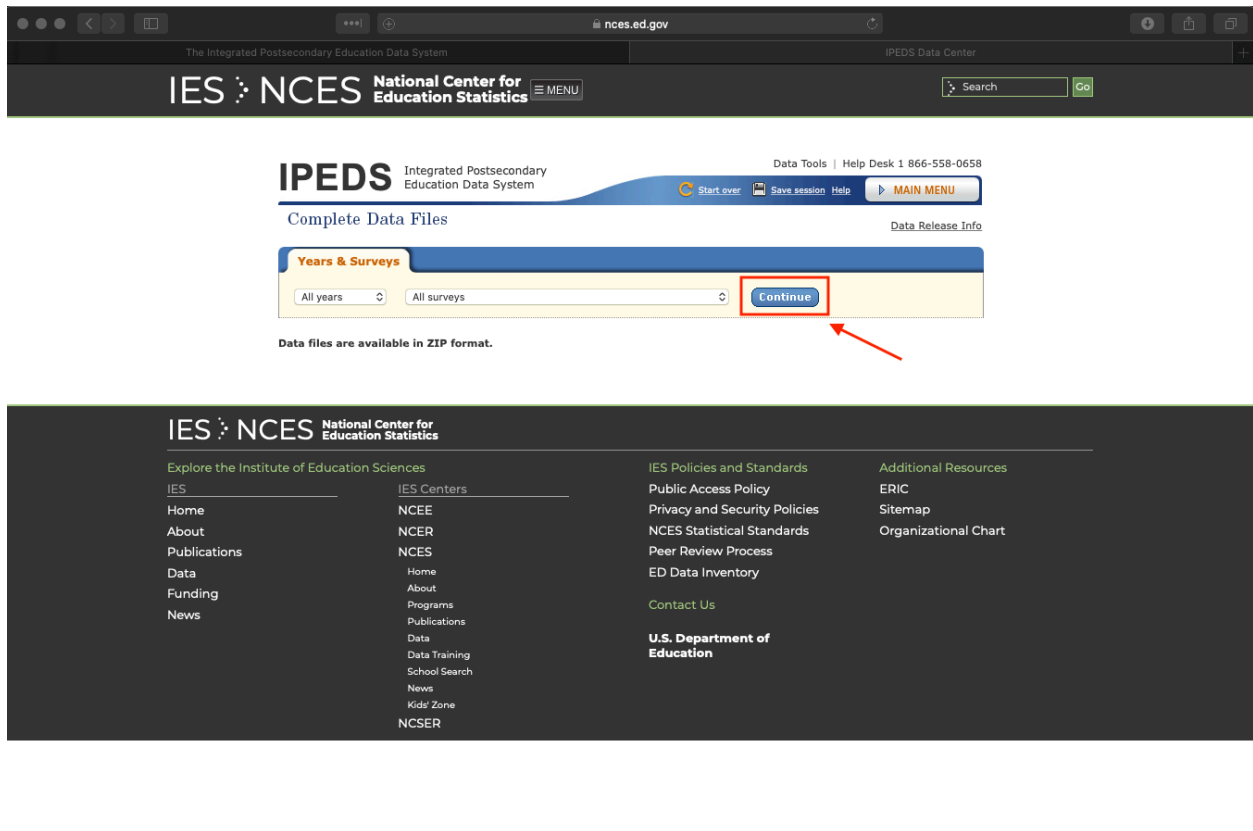
Access IPEDS data submitted to NCES through our data tools or download the data to conduct your research

- 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            |
| <b>Complete data files</b> |
| Custom data files          |
| Select your shortcut       |
- Publications and Products (2) Choose Complete data files**  
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
  - Data Cost Project

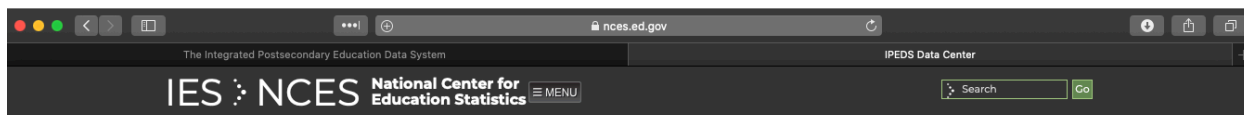
You may see a popup window — if so, just agree. You'll now see a pretty lonely page. If you have a specific file or year you know you want, use the two drop down menus to filter your search. Otherwise, just click the *Continue* button to see your options.





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.



IPEDS Integrated Postsecondary Education Data System

Data Tools | Help Desk 1 866-558-0658

Start over Save session Help MAIN MENU

Complete Data Files Data Release Info

Years & Surveys

All years All surveys Continue

Data files are available in ZIP format.

Choose CSV or STATA file

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

Get data dictionary

## 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/downloadipedes](https://github.com/btskinner/downloadipedes)<sup>9</sup> 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.

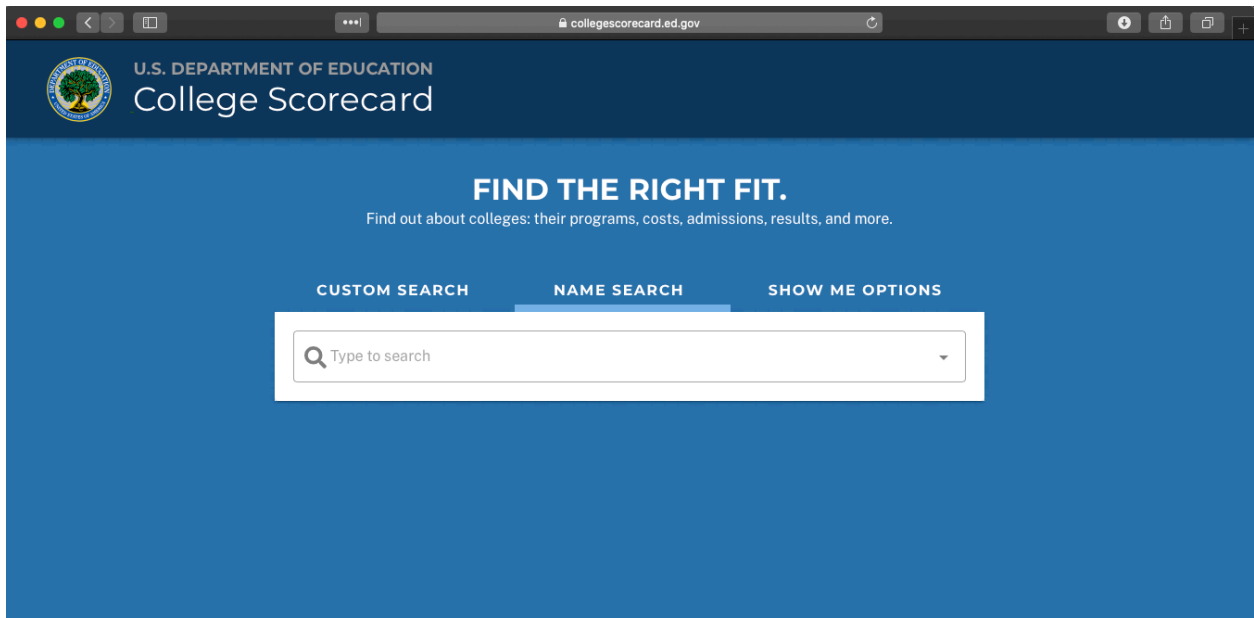
### Direct

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

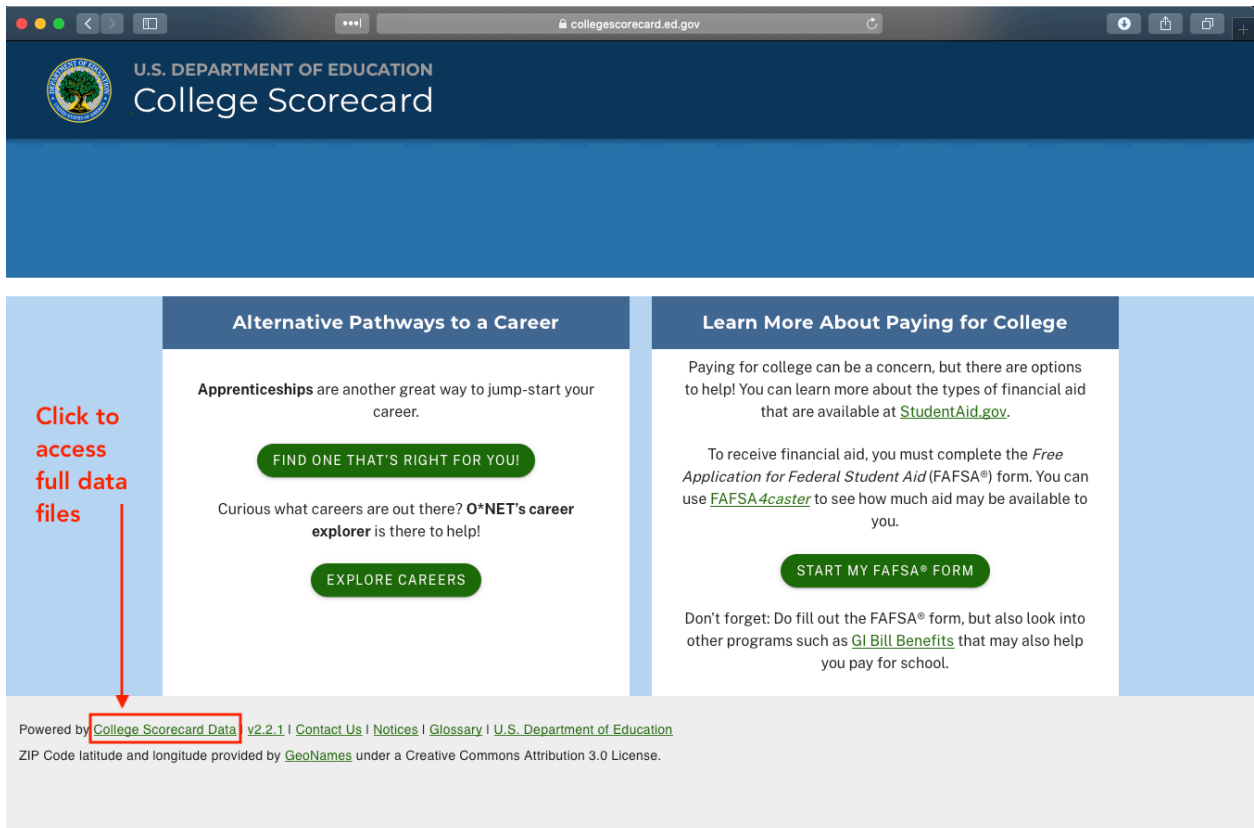
<sup>9</sup><https://github.com/btskinner/downloadipedes>

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

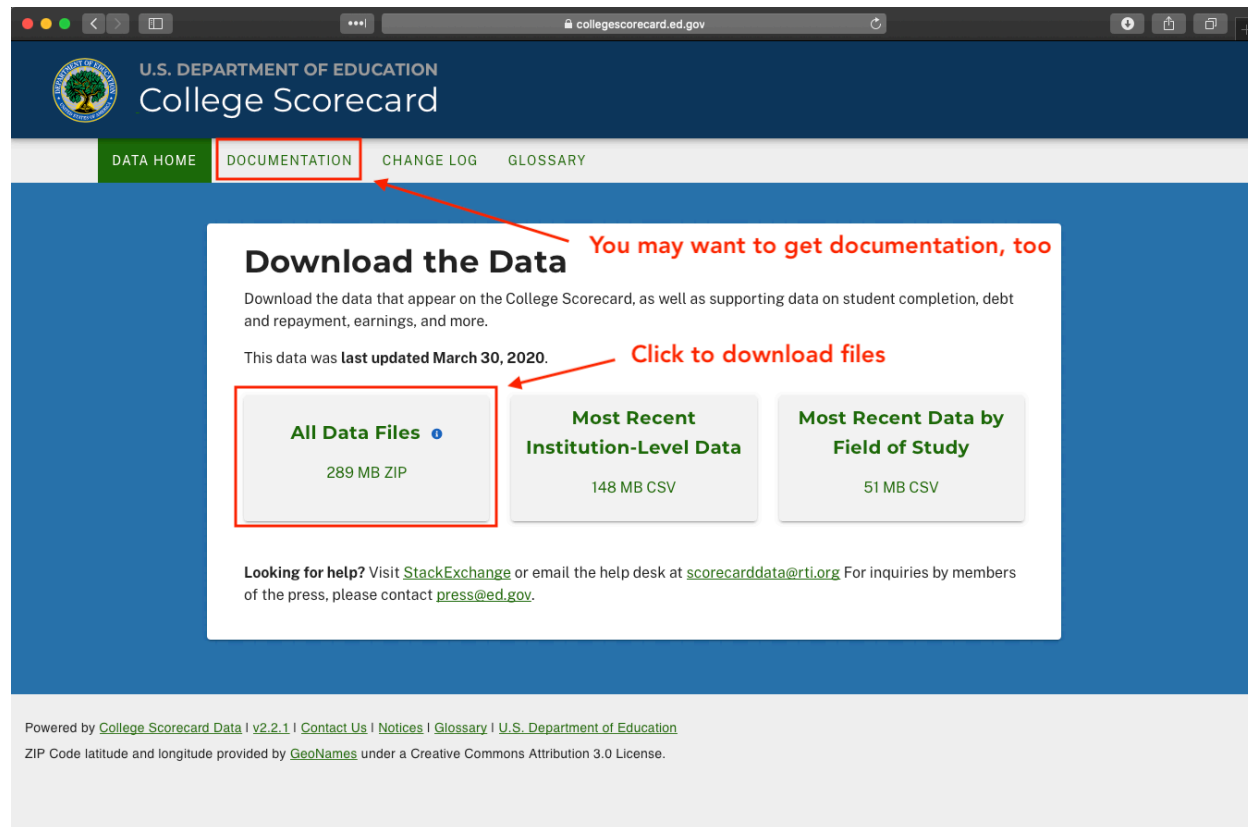
<sup>11</sup><https://collegescorecard.ed.gov>



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



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.



## 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](https://www.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](https://www.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

Search

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)

About the Survey

Respond to the Survey

News & Updates

**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

## 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.

The 2020 Census American Community Survey The importance of the A...

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? Yes No

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

United States Census Bureau

Search

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 (ACS)

Data Tables & Tools

Data via FTP

Summary File Data

**PUMS Data**

Variance Replicate Tables

Race/Ethnicity and American Indian & Alaska Native Data

Custom Tables

< Back to American Community Survey (ACS)

Respond to the ACS Learn how

## American Community Survey 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 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

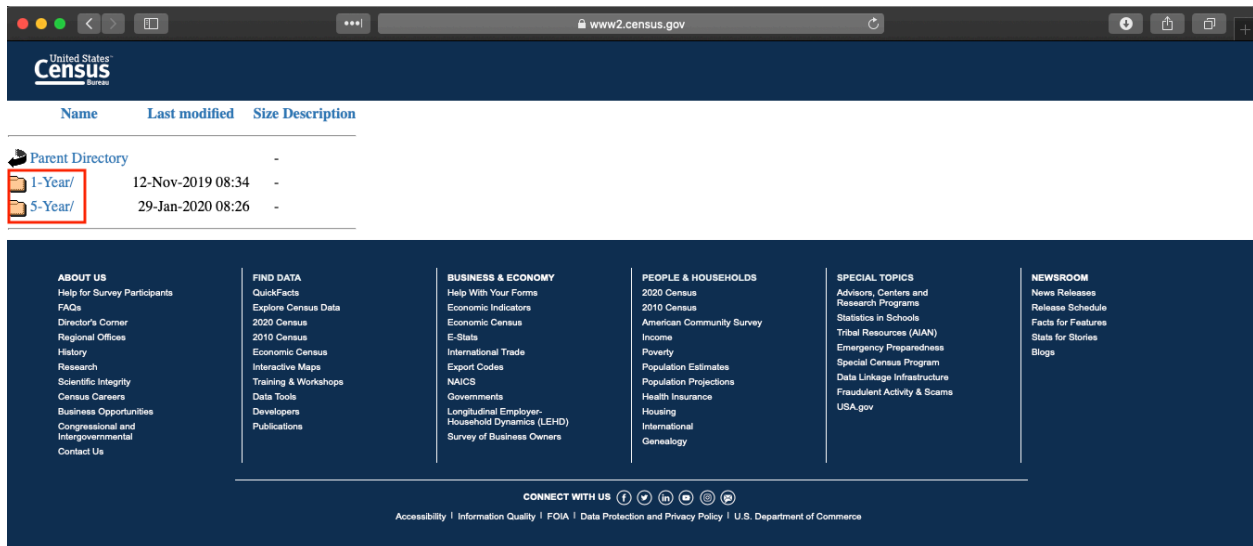
Is this page helpful? Yes No

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 the Census.gov website interface. At the top, there is a navigation bar with the United States Census Bureau logo and a search bar. Below the navigation bar, the breadcrumb trail reads: // Census.gov > Our Surveys & Programs > American Community Survey (ACS) > American Community Survey Data > PUMS Data. The main content area is titled 'PUMS Data' and includes social media icons for Facebook, Twitter, and LinkedIn. A sidebar on the left lists various data options under 'AMERICAN COMMUNITY SURVEY (ACS)', with 'PUMS Data' highlighted. The main content area is divided into two columns: 'Available on the FTP site' and 'PUMS on data.census.gov'. The 'Available on the FTP site' column lists links for 2018, 2017, 2016, 2015, 2014, and 2013 ACS PUMS, with the 2018 link highlighted by a red box. The 'PUMS on data.census.gov' column lists links for 2014-2018 ACS 5-year PUMS, 2018 ACS 1-year PUMS, 2013-2017 ACS 5-year PUMS, 2017 ACS 1-year PUMS, 2012-2016 ACS 5-year PUMS, and 2016 ACS 1-year PUMS. A right sidebar contains 'Related Information' and 'You May Be Interested In' sections. At the bottom right, there is a 'Is this page helpful?' survey with 'Yes' and 'No' buttons.

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

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

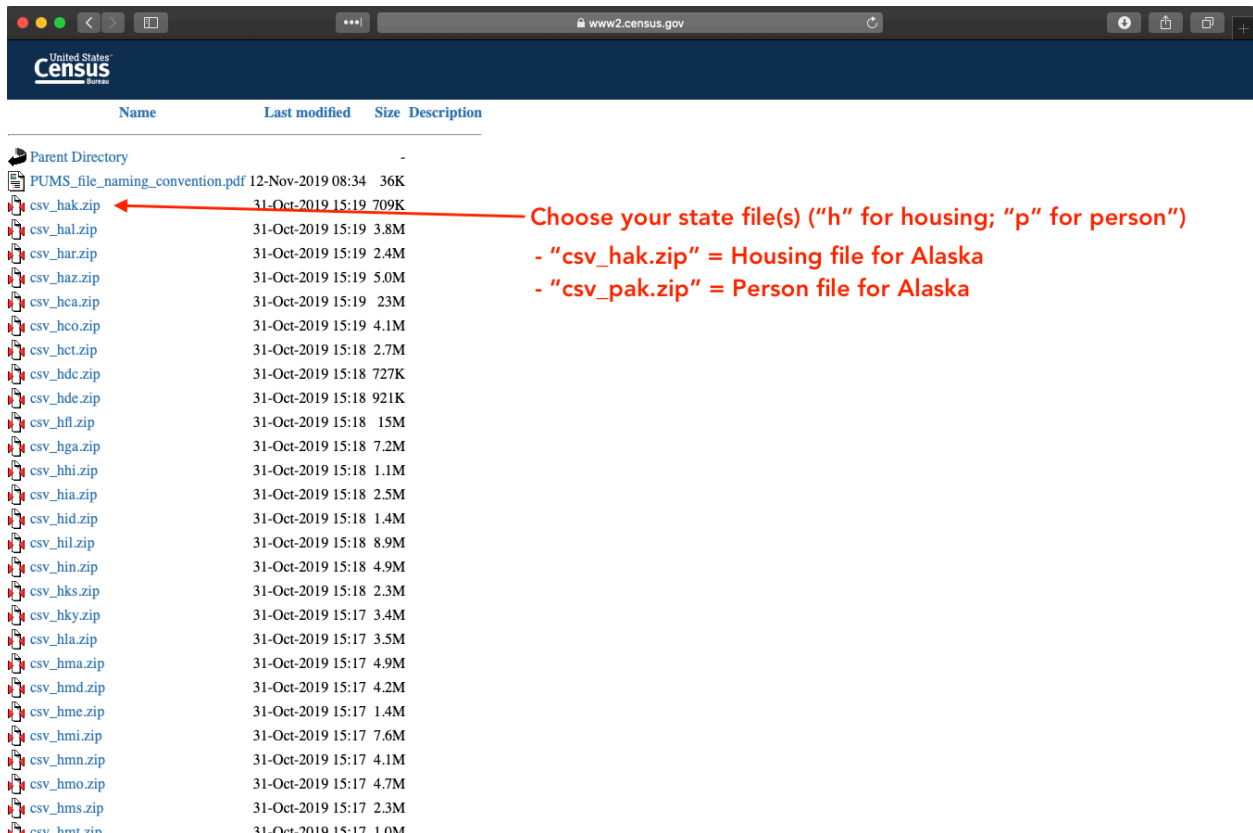


On the final page, you can choose data at the state level. There are two basic types of files, each pertaining to sections of the survey:

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

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



## 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](http://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](http://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html))<sup>18</sup>
- Bureau of Labor Statistics ([www.bls.gov/data/](http://www.bls.gov/data/))<sup>19</sup>
- Delta Cost Project ([deltacostproject.org/delta-cost-data](http://deltacostproject.org/delta-cost-data))<sup>20</sup>
- Urban Institute Data Explorer ([educationdata.urban.org/data-explorer/](http://educationdata.urban.org/data-explorer/))<sup>21</sup>
- PISA ([www.oecd.org/pisa/data/](http://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/>