# Analyzing and Interpreting Surveys

**ENGL 7360: Justice-Oriented Writing Assessment**
Professor Mya Poe
Fall 2022

**Taught by:** Juniper Johnson and Ana Abraham

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# Workshop Agenda

- Review: Social science methods + Survey design
- Introduction to NCES & DataLab
- Navigating DataLab
  - Statistical Analysis with PowerStats
  - Table and chart analysis with Tables Library
- Activity: DataLab exploration and discussion

Slides and presentation materials available at:
https://bit.ly/fa22-poe-surveyanalysis



*Feel free to ask questions at any point during the presentation!*

# Review: Data Ethics + Survey Design

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# Big Data & Algorithmic Bias

**Big Data:** vast amounts of data collected by companies, governments, and other groups about users, often analyzed and used for advertising, marketing, surveillance, etc. with the goal of ***predicting individual user behavior based on patterns from the user***, turning our information into a ***product***.

**Algorithmic bias**: perpetuation of biases in algorithmic processes, either through use of data that misrepresents populations or technology design.

> **"What is counted—like being a man or a woman—often becomes the basis for policymaking and resource allocation. By contrast, what is not counted—like being nonbinary—becomes invisible..."**
>
> Catherine D'Ignazio & Lauren Klein, Data Feminism, 2020



*Feel free to ask questions at any point during the presentation!*

# Critical Questions for Researchers:

- What **information** is being collected and from where? To whom does this data **belong**?
- How is it being **collected**? Do **participants** know that it is collected, how it will be collected, and how will it be used?
- **How** will the data be analyzed? What **biases** and **ideologies** may be implicit in this analysis?
- Who will this research impact? Who will it **benefit**? Who will it potentially **harm**?



*Feel free to ask questions at any point during the presentation!*

# Fundamentals of Social Science Research

**Variables**: sets of attributes or characteristics that can take on different values depending on what is being measured.

- **Ordinal Variables**: variables that can be measured and ordered, either quantitatively or qualitatively.
- **Nominal Variables**: variables without order, often where there is no expected or measured value collected.

**Common Inquiry Errors:** these are examples of common errors when creating or measuring variables.

- **Inaccurate Observations**
- **Overgeneralization**
- **Selective Observation**
- **Illogical Reasoning**



*Feel free to ask questions at any point during the presentation!*

# Survey Design: Types of Questions

**Multiple choice:** questions that only have specific answers and the user can only click one. For example, “What is your favorite physical exercise?”

**Likert-scale:** a type of response scale in which respondents can specify their level of agreement, importance or satisfaction typically in 5 points, (1) strongly agree to (5) strongly disagree.

**Checkboxes**: questions that only have specific answers and the user can click multiple. For example, “select all the physical exercises you did last week”.

**Linear Scale**: questions that invite users to choose from a lower to higher number that match their experience. Typically on a scale from 1 to 10; allows a more granular measure of affect and participants are able to express their degrees of response.

**Paragraph/Short Answer**: open-ended questions that usually invite longer prose. For example, “Describe your favorite gym exercise and how it makes you feel”.



*Feel free to ask questions at any point during the presentation!*

# Statistical Analysis

**Descriptive Statistics**: summary statistics used to quantitatively describe or summarize features of a collection of information or dataset, including frequency, central tendency (mean), variation, etc.

**Association:** a form of inferential statistics used to make inferences of a sample set of data for a larger populations; includes measures of association like Chi square to show how strongly two variables are related.

**Correlation:** a form of inferential statistics that is used to show an empirical relationship between two variables like if a change in one variable is associated with change in another. **Remember, correlation does not mean causation!**



*Feel free to ask questions at any point during the presentation!*

# Introduction: National Center for Education Statistics

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# National Center for Education Statistics (NCES)


National Center for Education Statistics (NCES) is a federal entity that collects and analyzes data on education in the US and internationally, from early childhood up through post-secondary educational levels.

Examples of some studies and assessments include:

- National Assessment of Educational Progress (NAEP)
- Education Longitudinal Study (ELS)
- National Household Education Survey (NHES)
- Baccalaureate and Beyond (B&B)



*Feel free to ask questions at any point during the presentation!*

# Closer Look: NCES Fast Facts

**Fast Facts**: a series of short, consumable fact sheets composed of "concise information on a range of educational issues" that is drawn from published studies from the NCES.

- Degree conferred by race/ethnicity and sex
- Undergraduate graduation rates
- Student Debt
- Students with Disabilities
- Race/ethnicity of college faculty

**Questions for Consideration:**

- What do you notice about how these fast facts are formatted?
- What sort of patterns do you see in the questions that they are asking?
- Is there anything missing?
- Are there other things you wish you could know based on this information?



*Feel free to ask questions at any point during the presentation!*

# NCES DataLab

**DataLab** is a web-based platform providing access, navigation, analysis, and presentation of data collected by the NCES, including the **PowerStats** and **Tables Library Tools**

## Tables Library

- Repository for data tables from all NCES publications
- Allows users to search by keyword, filter by source and topic, and download for Excel or CSV

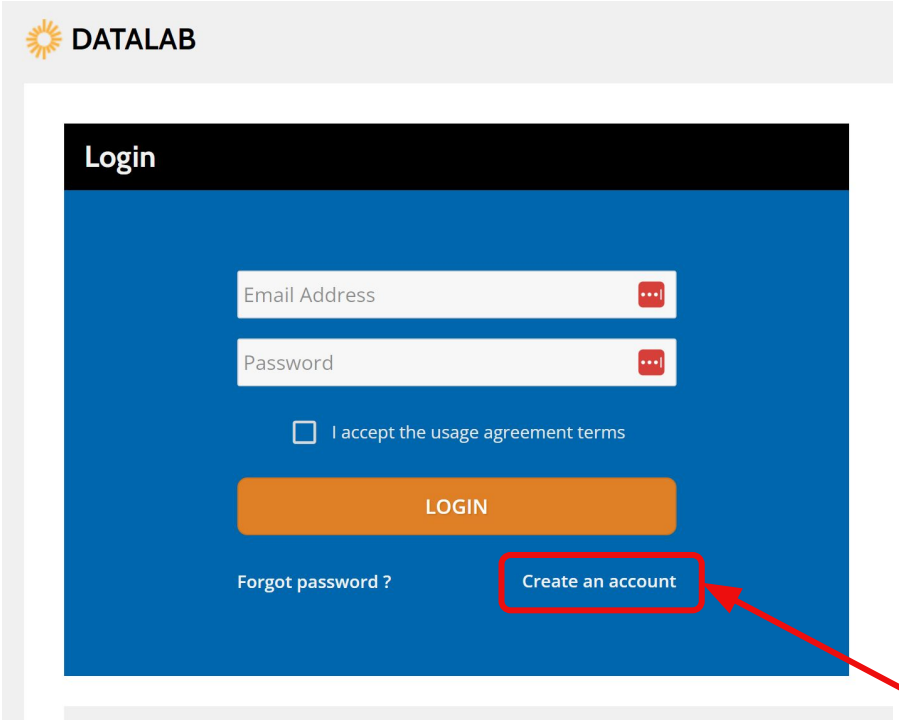
## PowerStats

- Generate statistical analyses and data visualizations
- Create and run percentage distributions, averages, medians, linear and logistic regressions, correlation matrices, etc.



*Feel free to ask questions at any point during the presentation!*

# Create a DataLab Account




**NCES DATA USAGE AGREEMENT**

Under law, public use data collected and distributed by the National Center for Education Statistics (NCES) may be used only for statistical purposes. Any effort to determine the identity of any reported case by public-use data users is prohibited by law. Violations are subject to Class E felony charges of a fine up to $250,000 and/or a prison term up to 5 years.

NCES does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted or modified in the dataset to protect the true characteristics of individual cases. Any intentional identification or disclosure of a person or institution violates the assurances of confidentiality given to the providers of the information. Therefore, users shall:

- Use the data in any dataset for statistical purposes only.
- Make no use of the identity of any person or institution discovered inadvertently, and advise NCES of any such discovery.
- Not link any dataset with individually identifiable data from other NCES or non-NCES datasets.

To proceed you must signify your agreement to comply with the above-stated statutorily based requirements.

**Note**: NCES data policies for survey results.



*Feel free to ask questions at any point during the presentation!*

# Your Turn!

Create your own DataLab account by going to:
https://nces.ed.gov/datalab/membership/login

1) Enter in your preferred email address
2) Check account for DataLab account verification
3) Create your password.

After creating an account, go to DataLab homepage.



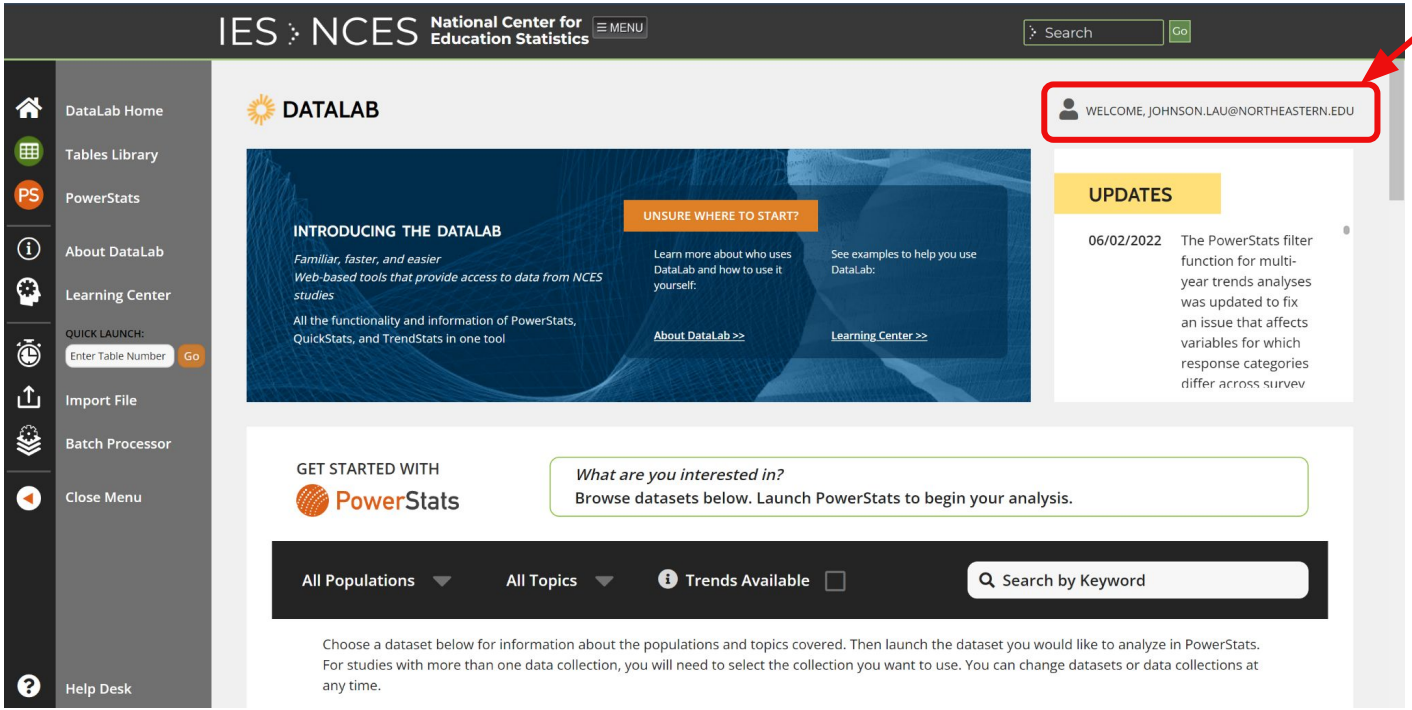
*Feel free to ask questions at any point during the presentation!*

# Navigating NCES DataLab

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# DataLab Interface

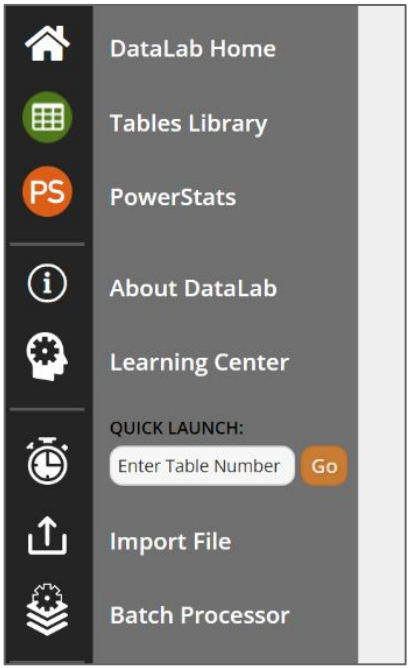



**User Profile:**
You can change certain settings for your user profile, including preferences for data (confidence and statistical significance) and access any tables that you have saved in the interface.



*Feel free to ask questions at any point during the presentation!*

# DataLab Menu

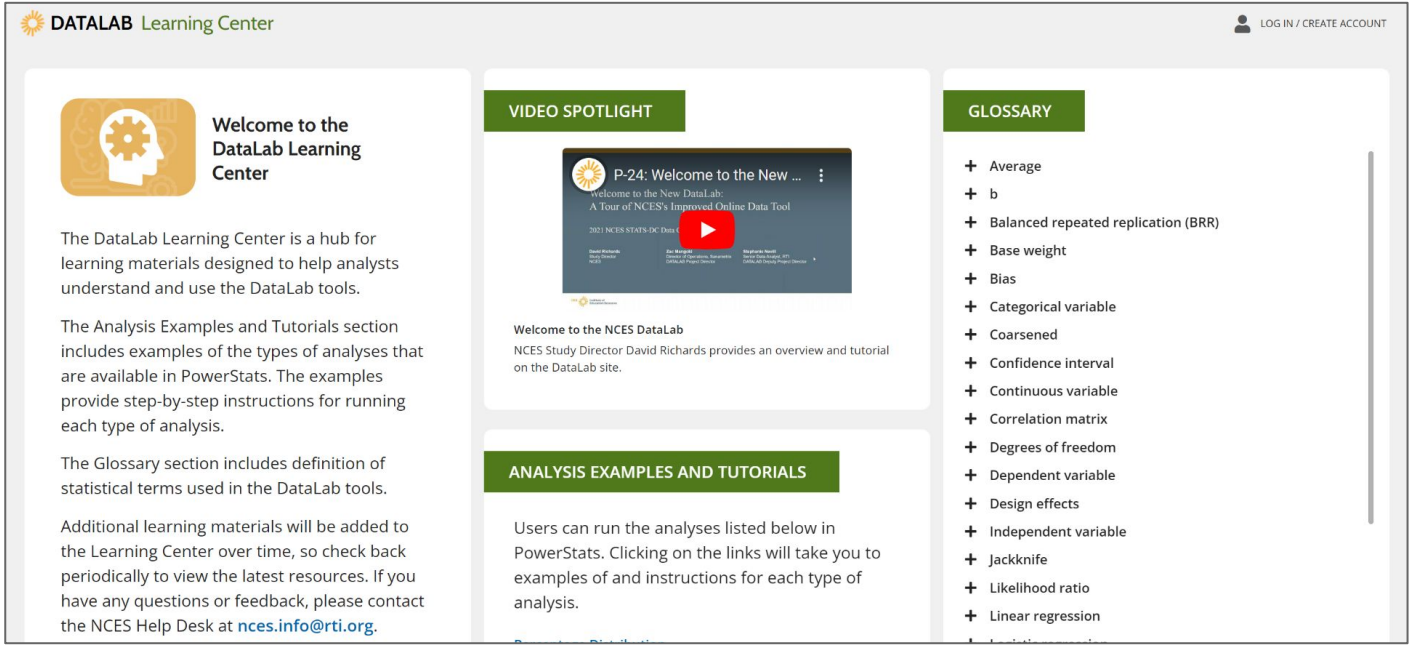



- **Home**: choose different studies or surveys to explore.
- **Tables Library:** explore tables from NCES publications.
- **PowerStats**: interface to explore data through different statistical analyses.
- **About**: access for FAQs, resources, and tutorials.
- **Learning Center:** tutorials for navigating DataLab interface and examples of analyses.
- **Import**: upload XML or JSON files for tables to edit in DataLab.
- **Batch Processor:** process multiple queries from specific files/surveys.



*Feel free to ask questions at any point during the presentation!*

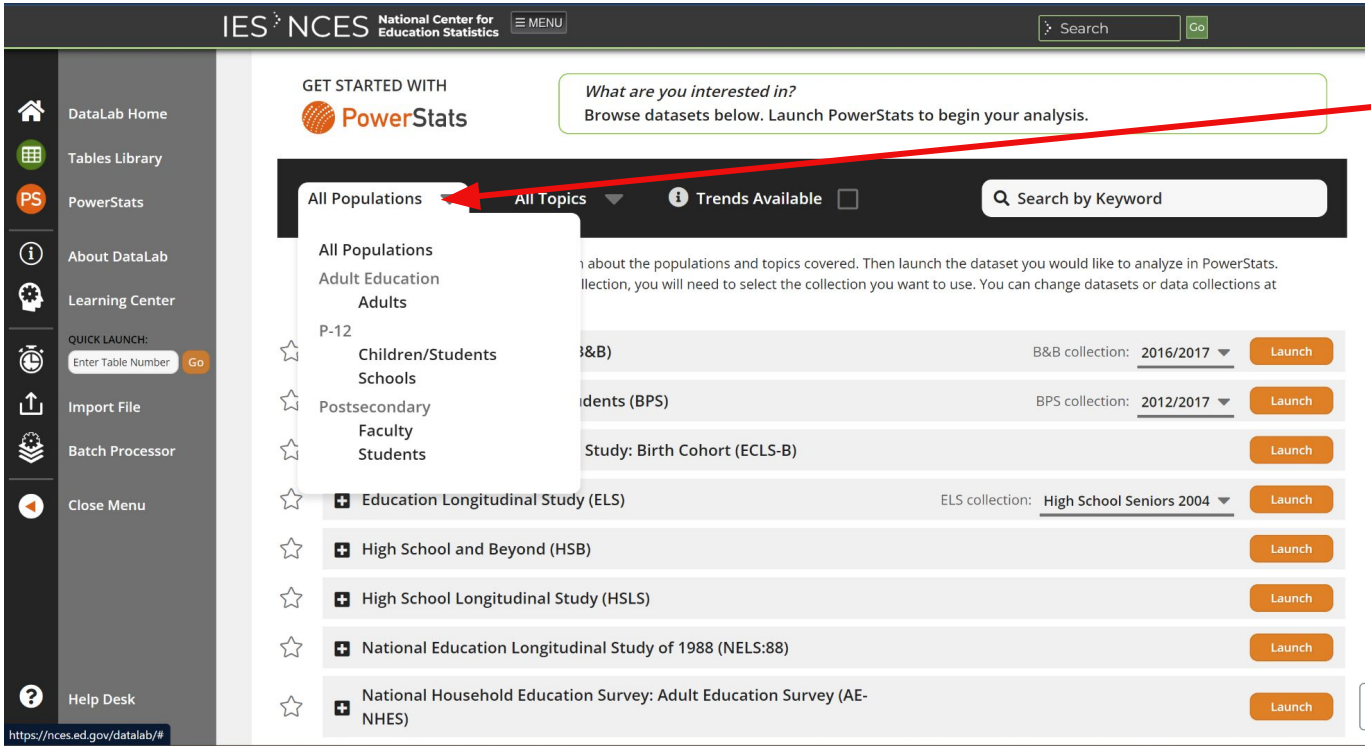
# DataLab Learning Center





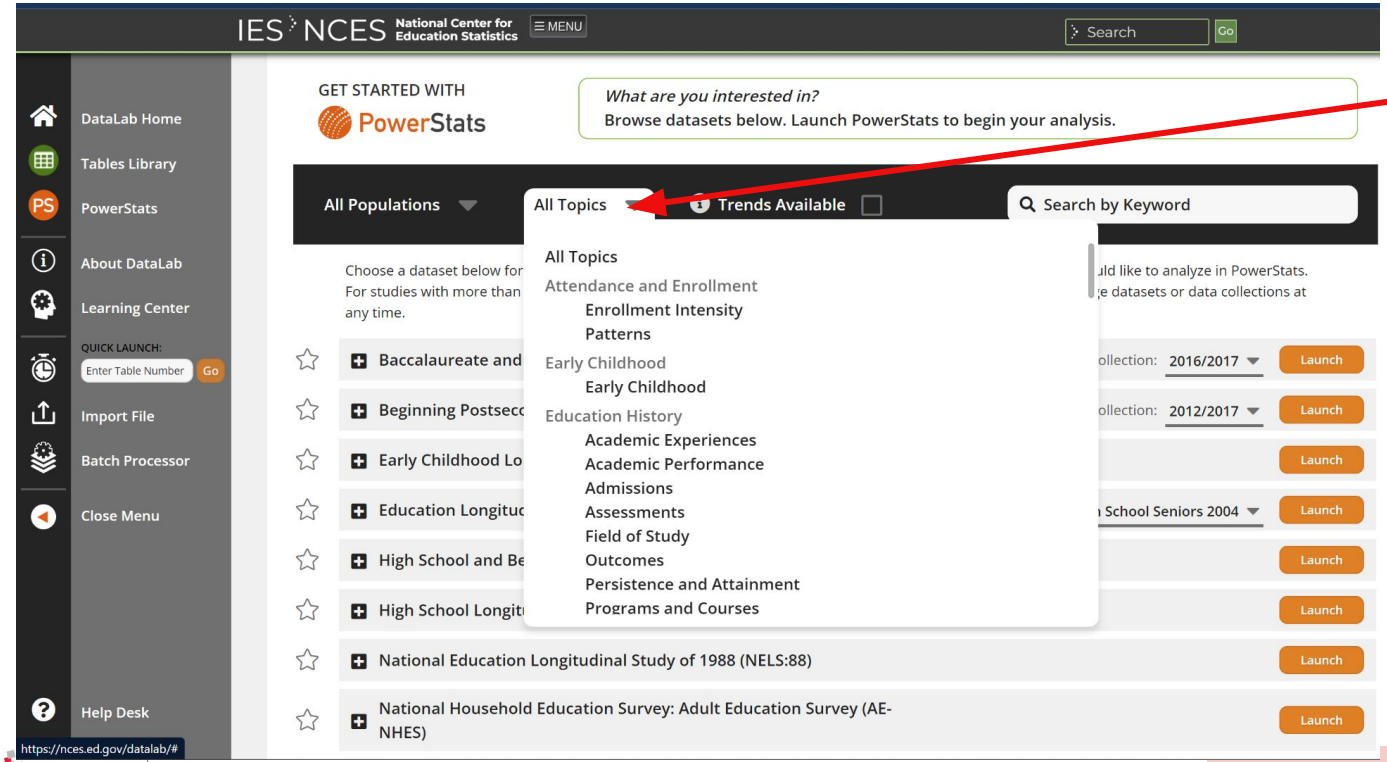

*Feel free to ask questions at any point during the presentation!*

# DataLab PowerStats: Populations




**Populations:**
You can choose which assessment, study, or survey data to investigate based on different populations using this drop down menu.

# DataLab PowerStats: Topics

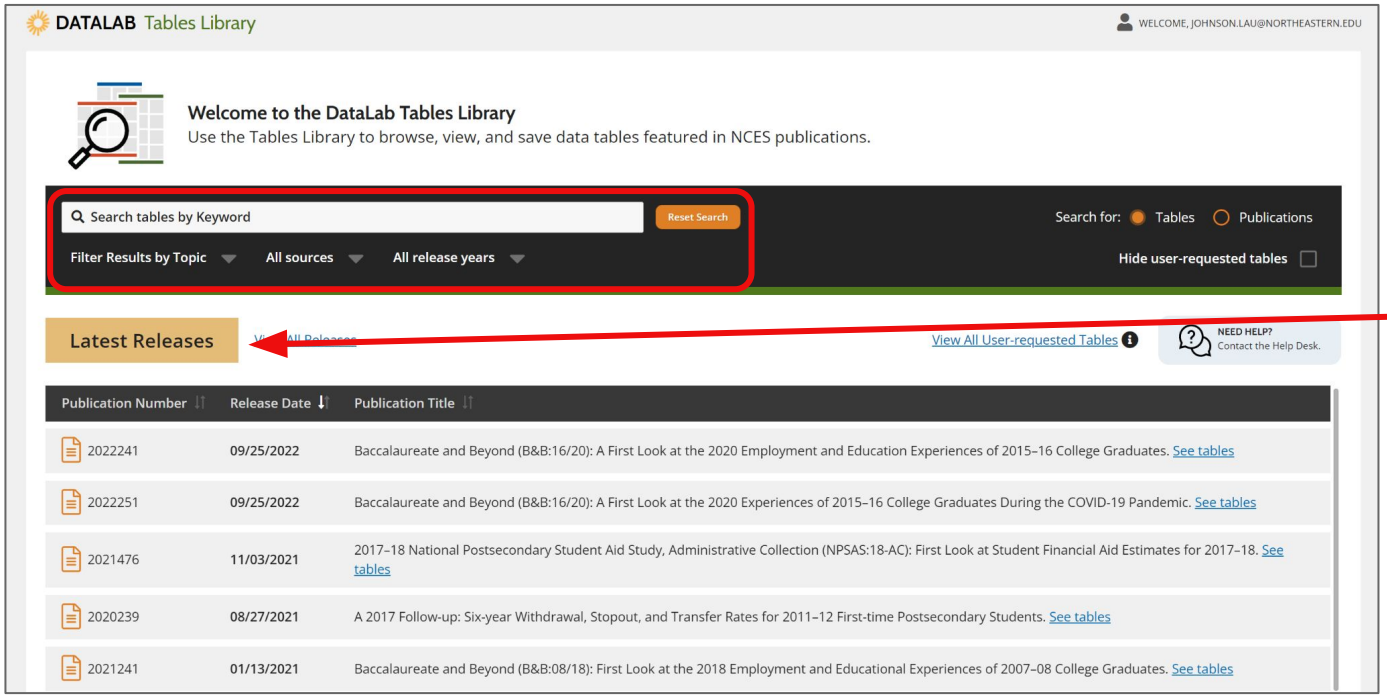



**Major Topics:**
You can also choose which assessment, study, or survey data to investigate based on different study topics using this drop down menu.



*Feel free to ask questions at any point during the presentation!*

# DataLab Tables Library

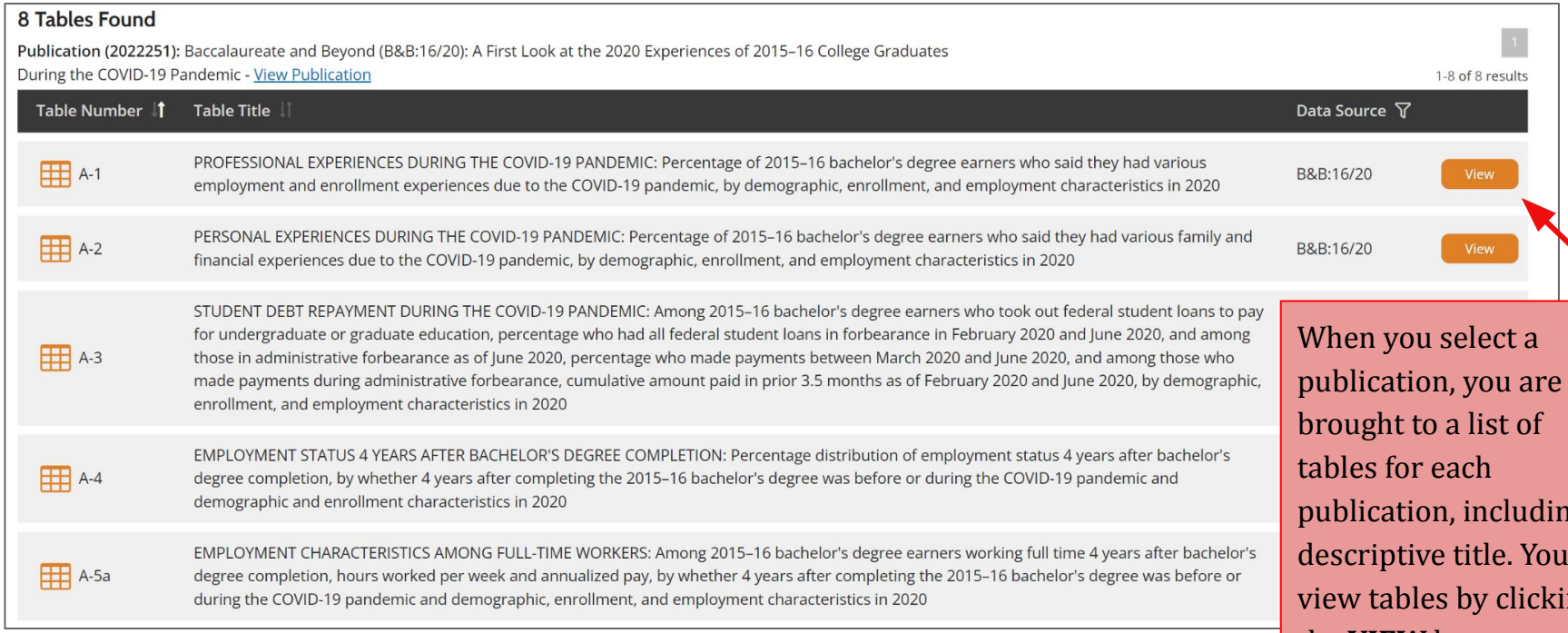



**Tables Library:**
You can explore data tables featured in NCES publications in the Tables Library, looking at latest releases by publication title or all releases. You can also search the database by keywords or by topic.



*Feel free to ask questions at any point during the presentation!*

# Example Tables



**8 Tables Found**

**Publication (2022251):** Baccalaureate and Beyond (B&B:16/20): A First Look at the 2020 Experiences of 2015–16 College Graduates During the COVID-19 Pandemic - View Publication

1-8 of 8 results

| Table Number | Table Title | Data Source | |
|---|---|---|---|
| A-1 | PROFESSIONAL EXPERIENCES DURING THE COVID-19 PANDEMIC: Percentage of 2015–16 bachelor's degree earners who said they had various employment and enrollment experiences due to the COVID-19 pandemic, by demographic, enrollment, and employment characteristics in 2020 | B&B:16/20 | View |
| A-2 | PERSONAL EXPERIENCES DURING THE COVID-19 PANDEMIC: Percentage of 2015–16 bachelor's degree earners who said they had various family and financial experiences due to the COVID-19 pandemic, by demographic, enrollment, and employment characteristics in 2020 | B&B:16/20 | View |
| A-3 | STUDENT DEBT REPAYMENT DURING THE COVID-19 PANDEMIC: Among 2015–16 bachelor's degree earners who took out federal student loans to pay for undergraduate or graduate education, percentage who had all federal student loans in forbearance in February 2020 and June 2020, and among those in administrative forbearance as of June 2020, percentage who made payments between March 2020 and June 2020, and among those who made payments during administrative forbearance, cumulative amount paid in prior 3.5 months as of February 2020 and June 2020, by demographic, enrollment, and employment characteristics in 2020 | | |
| A-4 | EMPLOYMENT STATUS 4 YEARS AFTER BACHELOR'S DEGREE COMPLETION: Percentage distribution of employment status 4 years after bachelor's degree completion, by whether 4 years after completing the 2015–16 bachelor's degree was before or during the COVID-19 pandemic and demographic and enrollment characteristics in 2020 | | |
| A-5a | EMPLOYMENT CHARACTERISTICS AMONG FULL-TIME WORKERS: Among 2015–16 bachelor's degree earners working full time 4 years after bachelor's degree completion, hours worked per week and annualized pay, by whether 4 years after completing the 2015–16 bachelor's degree was before or during the COVID-19 pandemic and demographic, enrollment, and employment characteristics in 2020 | | |

When you select a publication, you are brought to a list of tables for each publication, including a descriptive title. You can view tables by clicking the **VIEW** button at the right.



*Feel free to ask questions at any point during the presentation!*

# Using PowerStats on DataLab

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# DataLab Statistical References

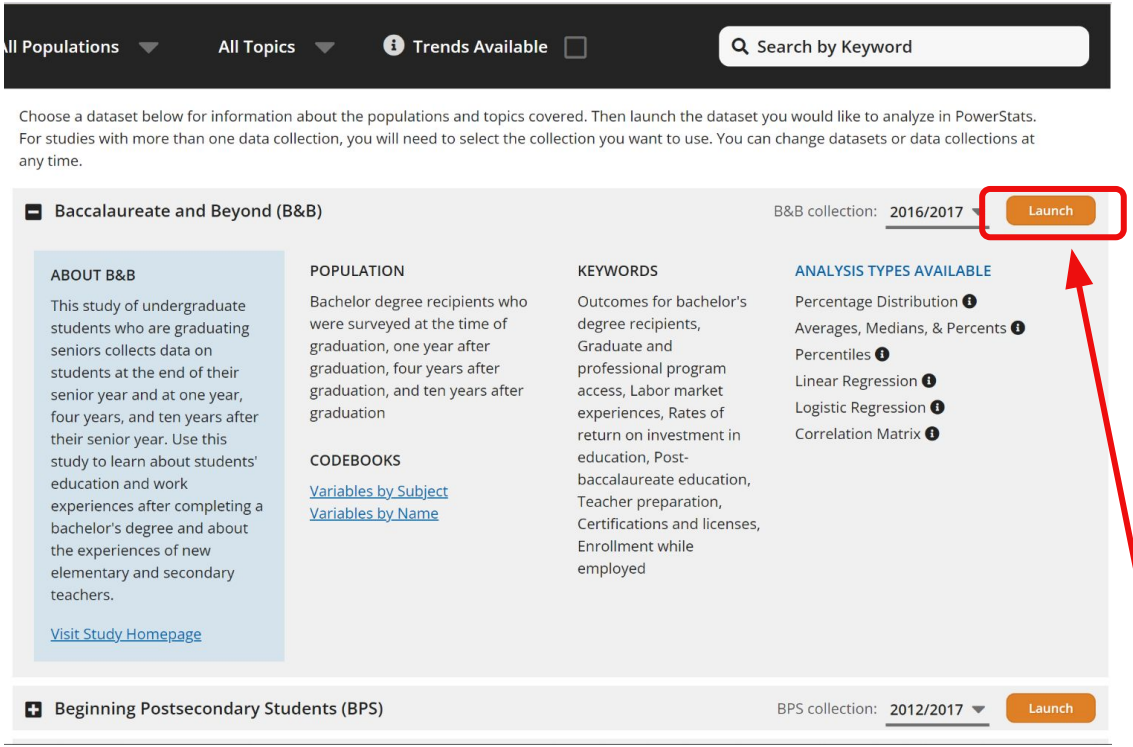
There are a number of different statistical analyses available with the PowerStats tool on the DataLab tool. Below is a brief explanation of three of the tools we recommend exploring today with links to step-by-step instructions for setting up each type of analysis:

- Percentage Distribution: a distribution of frequency in which the total frequency equals 100 and individual variable or class frequencies is expressed as proportion to the total frequency (x/100 or x%).
- Averages, Medians, & Percents: descriptive statistics that mark central tendency and frequency distribution of a dataset.
- Percentiles: a number or value at which a specified s percentage of data falls below, expressed as the “xth” percentile.



*Feel free to ask questions at any point during the presentation!*

# Case Study: Baccalaureate & Beyond




**B&B Study:**
Assessment of undergraduate students during their senior year, then 4 and 10 years after graduation to learn about education and work experiences.

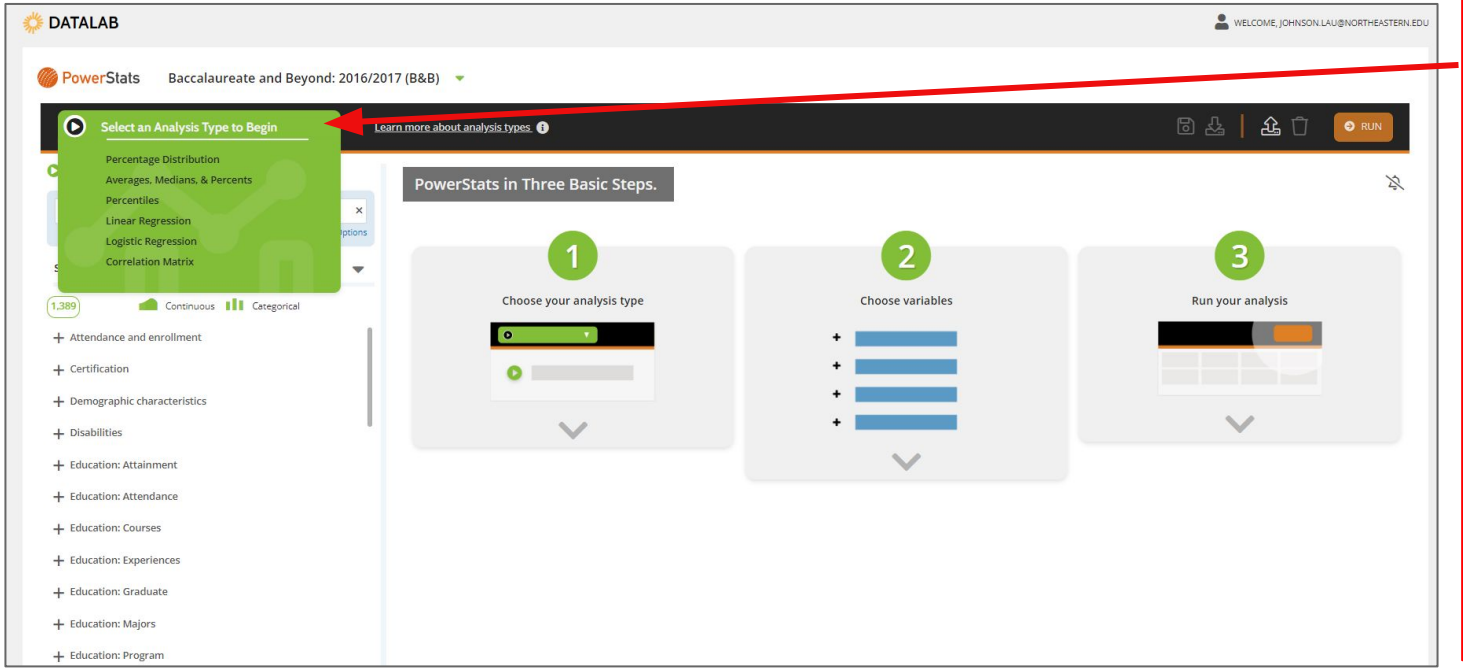
The DataLab interface shows you information on each study's **population, keywords,** and **types of analysis available.**

To get started, click on the **LAUNCH** button at the right.



*Feel free to ask questions at any point during the presentation!*

# Step 1: Choose your analysis type



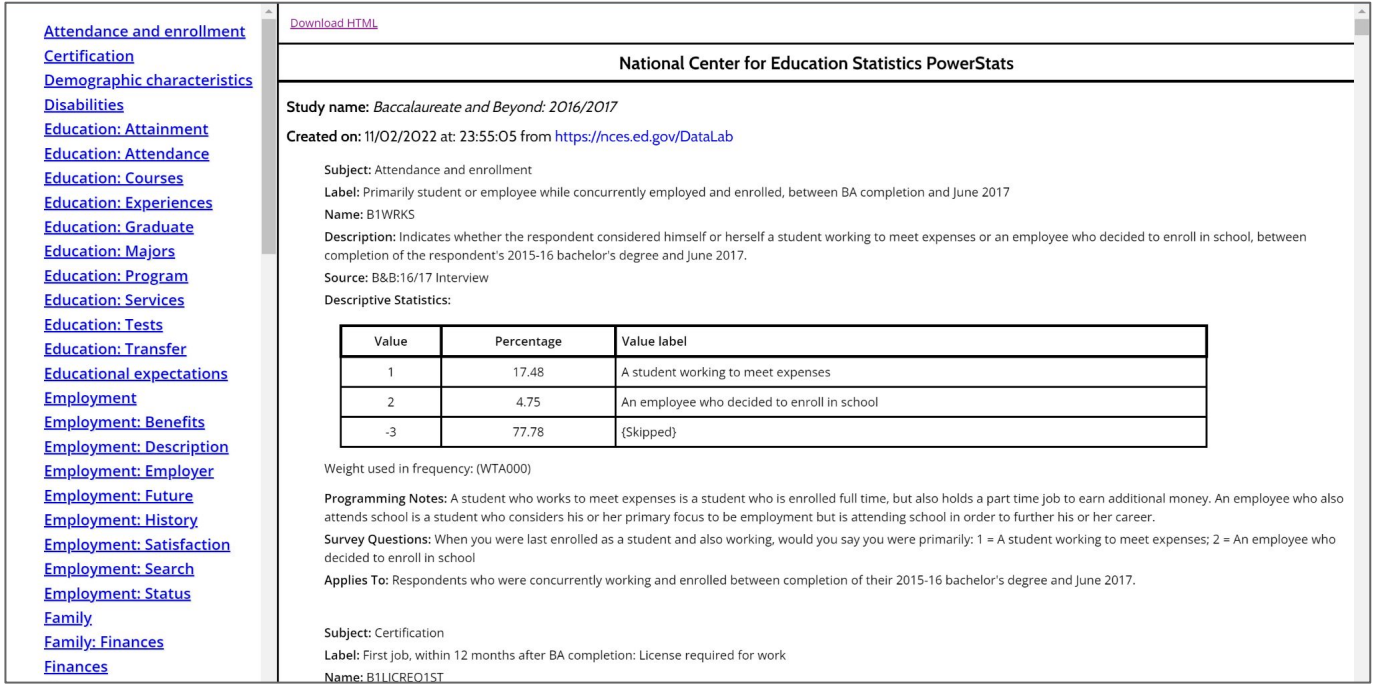

**Step 1: Choose your analysis type.**

There a number of built-in statistical analyses available in the DataLab interface, including **percentile distribution**, **averages**, **linear regressions**, and **correlation matrices**.



*Feel free to ask questions at any point during the presentation!*

# Study Variables (by Subject)



Attendance and enrollment
Certification
Demographic characteristics
Disabilities
Education: Attainment
Education: Attendance
Education: Courses
Education: Experiences
Education: Graduate
Education: Majors
Education: Program
Education: Services
Education: Tests
Education: Transfer
Educational expectations
Employment
Employment: Benefits
Employment: Description
Employment: Employer
Employment: Future
Employment: History
Employment: Satisfaction
Employment: Search
Employment: Status
Family
Family: Finances
Finances

Download HTML

**National Center for Education Statistics PowerStats**

**Study name:** *Baccalaureate and Beyond: 2016/2017*

**Created on:** 11/02/2022 at: 23:55:05 from https://nces.ed.gov/DataLab

**Subject:** Attendance and enrollment

**Label:** Primarily student or employee while concurrently employed and enrolled, between BA completion and June 2017

**Name:** B1WRKS

**Description:** Indicates whether the respondent considered himself or herself a student working to meet expenses or an employee who decided to enroll in school, between completion of the respondent's 2015-16 bachelor's degree and June 2017.

**Source:** B&B:16/17 Interview

**Descriptive Statistics:**

| Value | Percentage | Value label |
|---|---|---|
| 1 | 17.48 | A student working to meet expenses |
| 2 | 4.75 | An employee who decided to enroll in school |
| -3 | 77.78 | {Skipped} |

Weight used in frequency: (WTA000)

**Programming Notes:** A student who works to meet expenses is a student who is enrolled full time, but also holds a part time job to earn additional money. An employee who also attends school is a student who considers his or her primary focus to be employment but is attending school in order to further his or her career.

**Survey Questions:** When you were last enrolled as a student and also working, would you say you were primarily: 1 = A student working to meet expenses; 2 = An employee who decided to enroll in school

**Applies To:** Respondents who were concurrently working and enrolled between completion of their 2015-16 bachelor's degree and June 2017.

**Subject:** Certification

**Label:** First job, within 12 months after BA completion: License required for work

**Name:** B1LICREQ1ST

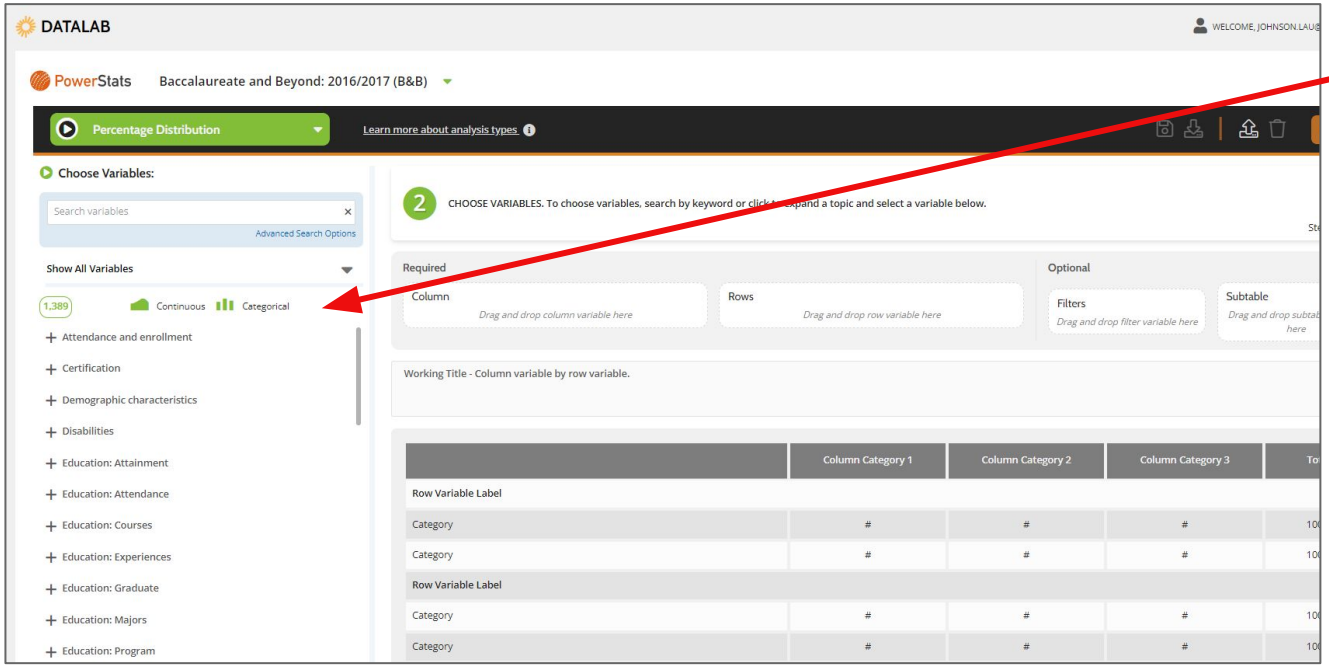
**Study Variables:**
Each of the codebooks for the NCES studies include detailed information for the different variables including:

- **Subject title**
- **Label + Name**
- **Description**
- **Source**
- **Descriptive Statistics:** percentage of values for this single variable
- **Programming notes:** information about when participants were surveyed about this topic.
- **Associated Survey Questions** (including possible answers)



*Feel free to ask questions at any point during the presentation!*

# Step 2: Choose your variables.




**Step 2: Choose your variables.**

Each study has many different variables, more than we can look at!

**Continuous variables:** Variables with data that is categorized or grouped.

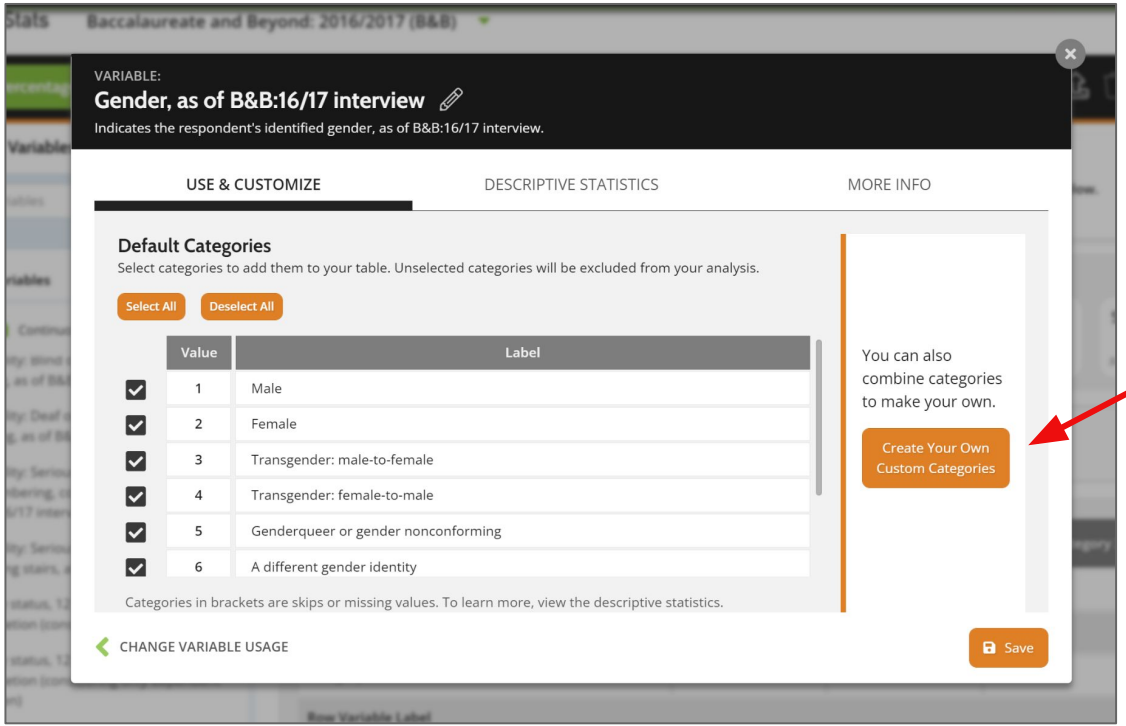
**Categorical variables**: Variables that can be measured, but with a theoretically infinite number of values.

**Drag your variable to the column and row sections.**



*Feel free to ask questions at any point during the presentation!*

# Variable Customization: Gender




**Variable Customization Prompt**
After choosing a variable, a variable usage and customization box will open up on the screen.

In this window, you can customize your variable. Each variable has different options, but this allows you to change or ***customize*** the variables beyond the **default values.**

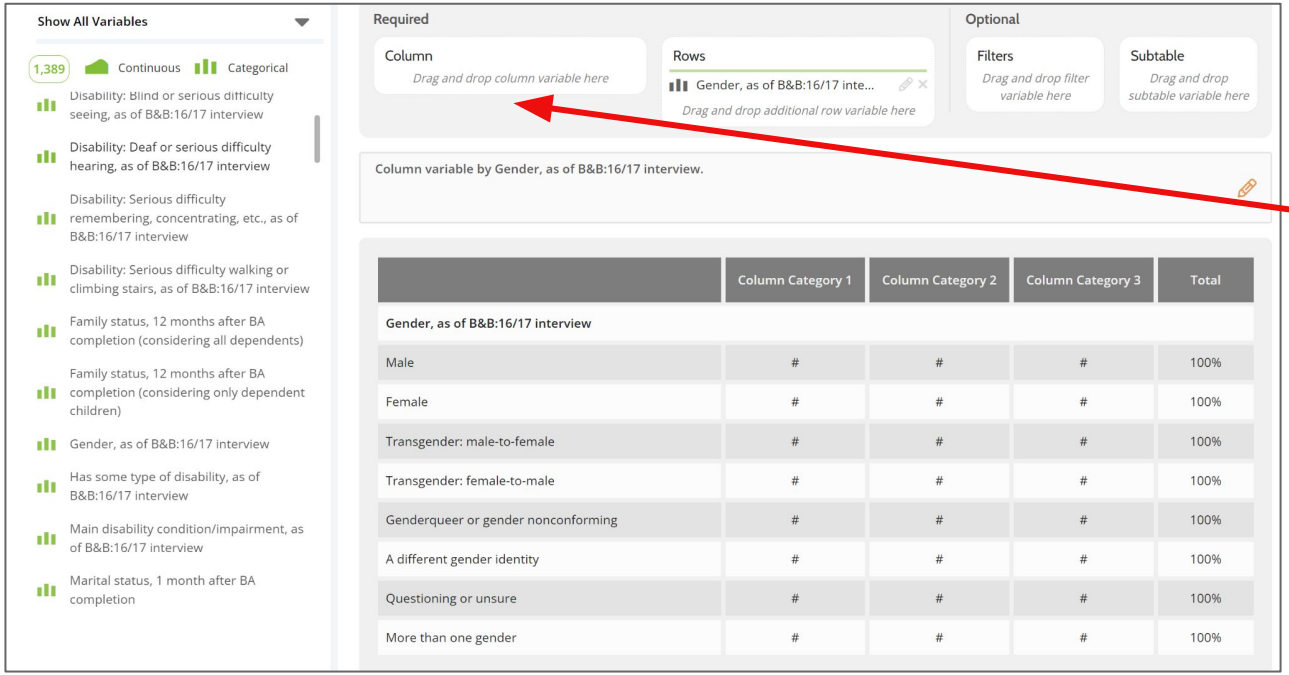
For example, you can choose to leave out certain variable values from your analysis and table, or create new categories of your own to group variable values.

**Question: what are some possible benefits or drawbacks of this feature?**



*Feel free to ask questions at any point during the presentation!*

# Step 2: Choose your variables (cont.)



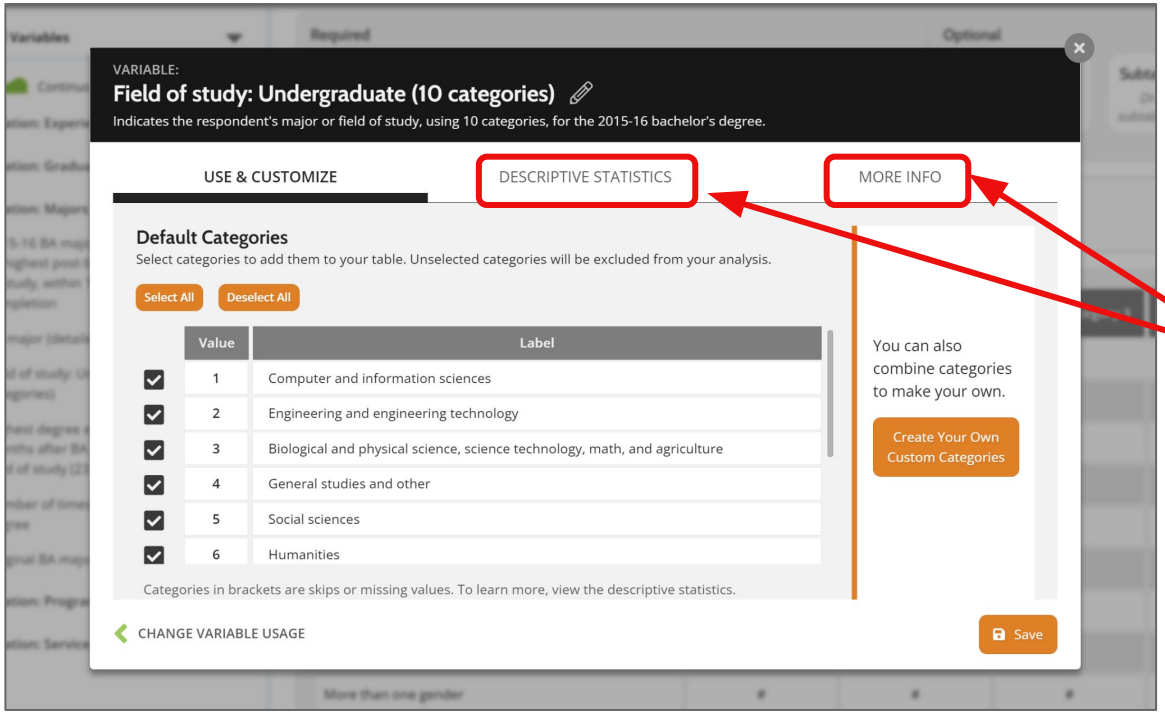

**Step 2: Choose your variables (cont).**

After choosing one of your variables, you will need to choose the second. You can swap the variables between the column or row categories to change how the data will be displayed in the final table. After choosing one variable, it will show you what the format for the table will look like before you run it.



*Feel free to ask questions at any point during the presentation!*

# Variable Customization: Field of Study



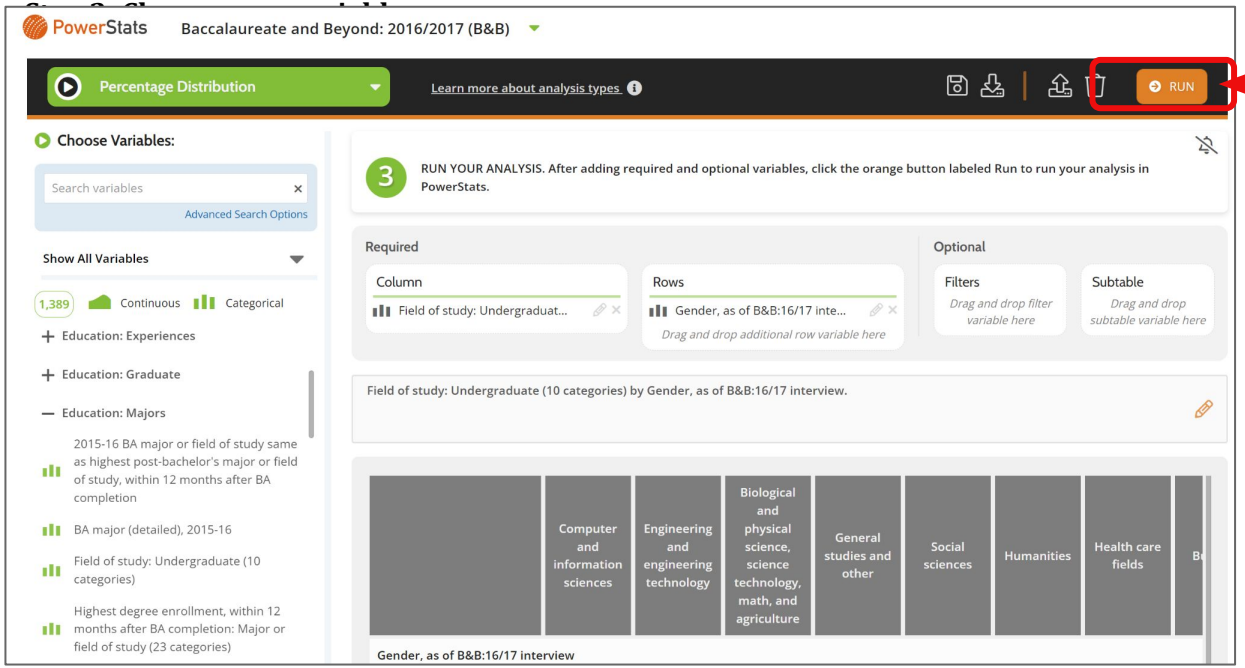

**Variable Customization Prompt**
Here is an example of another customization prompt for a second variable, this time for field of undergraduate study.

While having access to the codebook for the study variables is helpful, you can access this information about the name, subject, descriptive statistics, and survey questions for each variables on the other tabs **DESCRIPTIVE STATISTICS** and **MORE INFO.**



*Feel free to ask questions at any point during the presentation!*

# Step 3: Run your analysis




**Step 3: Run your Analysis.**

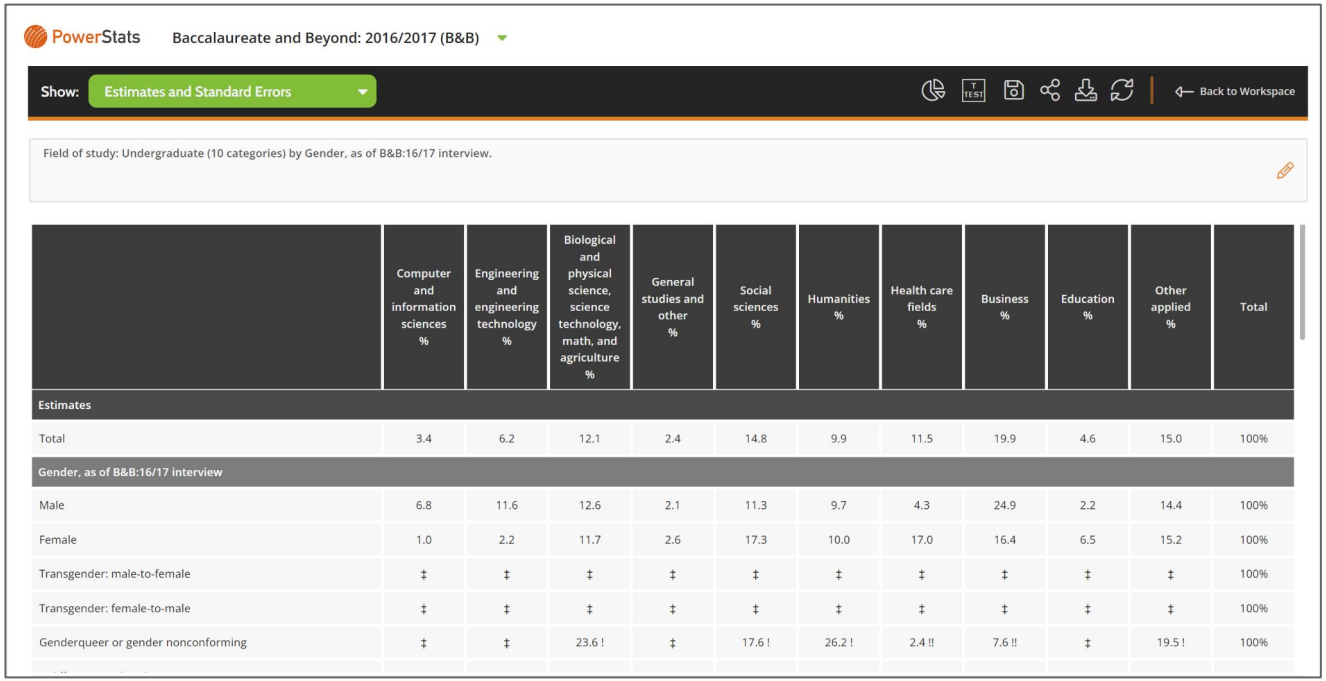
Once you have selected your variables from the list and formatted them (choose which value options you want or not, if applicable), hit the **RUN** button.

**NOTE**: you can change the title of the table in the box with the orange edit arrow. Below that is a "draft" preview of what the table will look like, including the units of the numbers expected in each of the cells (% here).



*Feel free to ask questions at any point during the presentation!*

# Percentile Distribution: Gender + field of study



PowerStats Baccalaureate and Beyond: 2016/2017 (B&B)

Show: Estimates and Standard Errors

Back to Workspace

Field of study: Undergraduate (10 categories) by Gender, as of B&B:16/17 interview.

| | Computer and information sciences % | Engineering and engineering technology % | Biological and physical science, science technology, math, and agriculture % | General studies and other % | Social sciences % | Humanities % | Health care fields % | Business % | Education % | Other applied % | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Estimates** | | | | | | | | | | | |
| Total | 3.4 | 6.2 | 12.1 | 2.4 | 14.8 | 9.9 | 11.5 | 19.9 | 4.6 | 15.0 | 100% |
| **Gender, as of B&B:16/17 interview** | | | | | | | | | | | |
| Male | 6.8 | 11.6 | 12.6 | 2.1 | 11.3 | 9.7 | 4.3 | 24.9 | 2.2 | 14.4 | 100% |
| Female | 1.0 | 2.2 | 11.7 | 2.6 | 17.3 | 10.0 | 17.0 | 16.4 | 6.5 | 15.2 | 100% |
| Transgender: male-to-female | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | 100% |
| Transgender: female-to-male | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | ‡ | 100% |
| Genderqueer or gender nonconforming | ‡ | ‡ | 23.6 ! | ‡ | 17.6 ! | 26.2 ! | 2.4 !! | 7.6 !! | ‡ | 19.5 ! | 100% |

**Step 4: Explore your resulting statistical table!**

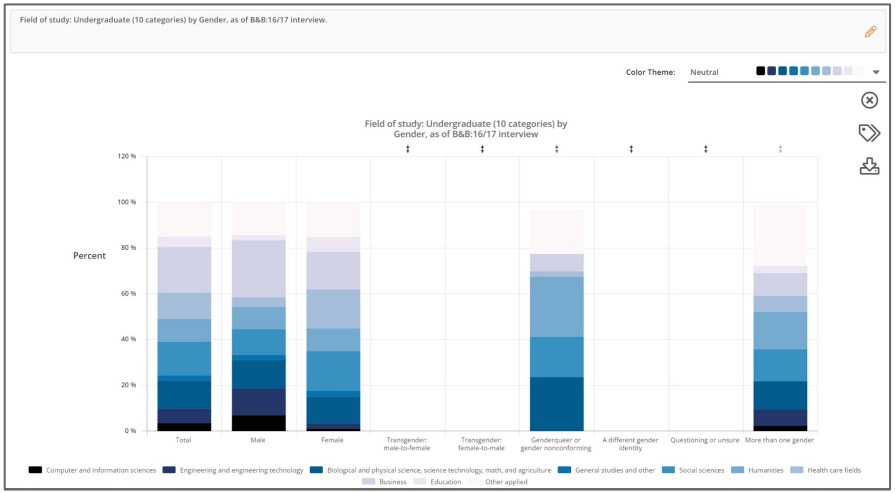
Once you have run your statistical analysis, you will see a large table that has run the calculation for the variables you selected. Feel free to scroll down and explore the information it yielded.



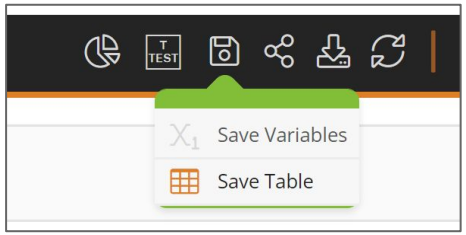
*Feel free to ask questions at any point during the presentation!*

# Next Steps: what can I do with this analysis?

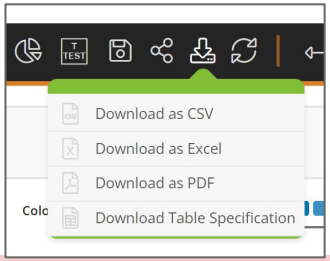
1. **Create charts:** after running an analyses, you can create bar charts, pie charts, or a stacked bar chart from the data.




2. **Save your table to your DataLab library.**




2. **Download your table and analyses:** choose from CSV, Excel, PDF, or other specifications. This includes the confidence levels, relative standard error, etc.






*Feel free to ask questions at any point during the presentation!*

# Using DataLab: Exploration and Discussion

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# Your Turn!

Use one of the PowerStats options to explore different ways to analyze the Baccalaureate & Beyond NCES study. If you have questions about how to set up these different analyses, see the step-by-step tutorials in the DataLab Learning Center.

Consider the following questions:

1. What variables interest you? How might you find variables in this (or other studies) that connect to your research interests?
2. What kind of questions are you interested in asking about these variables? What comes to mind?
3. Are there any variables that are missing?



*Feel free to ask questions at any point during the presentation!*

# Your Turn!

Consider the following additional questions:

1. What does it mean to have caution around statistical analysis?
2. Did you try any of the variable customization features? Why or why not?
3. How might customizing the values for variables help or hinder your analysis or presentation of statistical results?
4. Did you use the table/chart tool? Did being able to present the data in a non-table format help you look at it in a new way?
5. What connections do you see between this tool and the discussions you have had about surveys/assessments so far?
6. How does thinking or looking at data in a database level change what kind of questions or observations you have?



*Feel free to ask questions at any point during the presentation!*

# Thank you!

If you have any questions, contact us at: nulab.info@gmail.com

**Sign up for office hours at:** http://calendly.com/diti-nu/

We'd love your feedback! Please fill out a short survey here: https://bit.ly/diti-feedback

Slides & handouts available at: https://bit.ly/fa22-poe-surveyanalysis



*Feel free to ask questions at any point during the presentation!*