SOLUTIONS JOURNALISM NETWORK

Story Recipe: Student Retention Rates

Nearly one in five first-year students at four-year colleges don't return after summer break. That translates to hundreds of thousands of students who may never complete their degree, with potentially grave financial consequences – both immediate (losing the cost of attendance and missing time in the workforce) and long-term (lifelong diminished earnings capacity).

Data from the U.S. Department of Education's Integrated Postsecondary Education Data System, known as IPEDs, offers a window into the widely varying retention rates at U.S. colleges and universities. Some of the insights are not surprising: For-profit schools are over-represented among those with very low retention rates; and the most-selective schools in the nation have little trouble keeping nearly 100 percent of their students from year to year.

But there are deeper insights to be found, particularly when looking at data across multiple years. Those figures can point to schools that have lost ground in the struggle to keep students enrolled, and that can be an important story. But the data can also point to institutions that have shown great improvement, with a growing percentage of students who stay in school and advance toward their degrees.

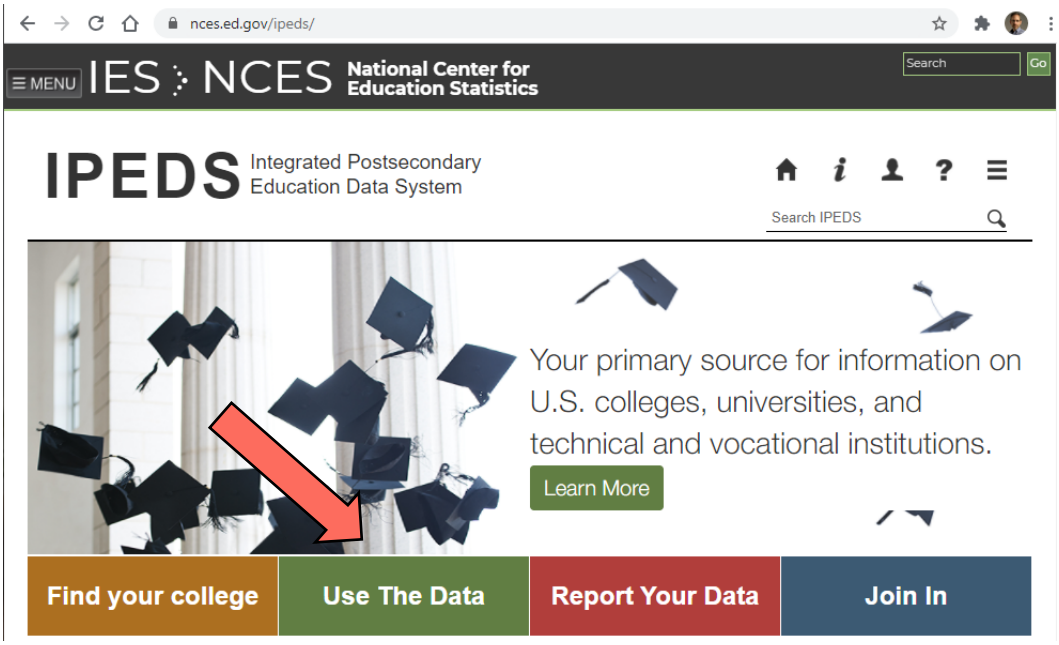
Those better-performers are "positive deviants" – outliers in data that might lead you to institutions that have found a better way, and might offer an opportunity for solutions journalism. If it turns out those positive deviants deployed specific policies to support students and increase retention, then reporting on those institutions can make for great accountability journalism by deflating the excuses of poor performers. And if those policies are replicable from school to school, taking a solutions lens can lead to more engaging and useful public-service journalism – just the sort of reporting studies show our audiences crave.

Solutions journalism isn't advocacy or fluff. It isn't silver-bullet journalism. It's rigorous, evidence-based reporting that explores the limitations of a response as aggressively as its benefits. And it is the natural evolution beyond reporting that focuses solely on problems, with no examination of responses that have a time-tested record of better outcomes.

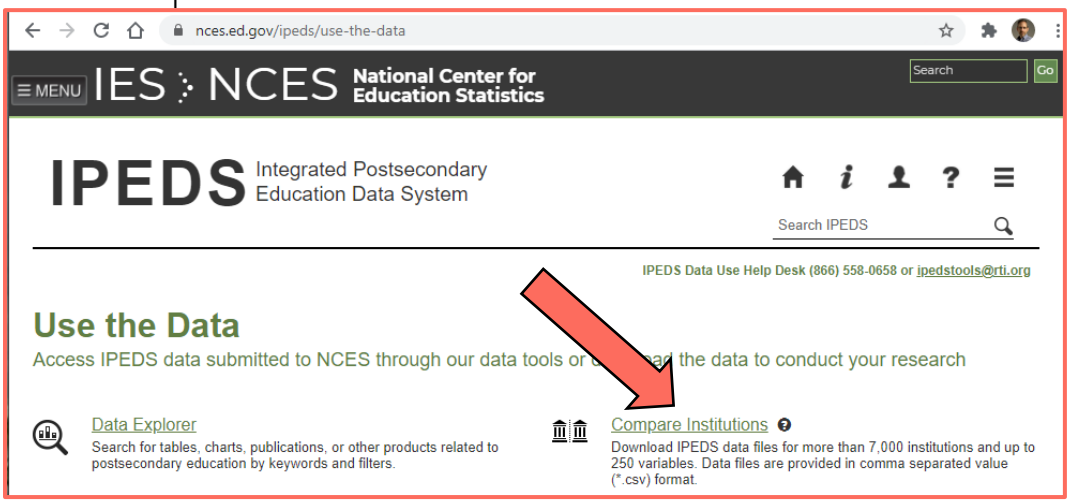
So let's dive into the IPEDS data and see if it points us to schools that have shown marked improvement in retention rates, and that might be worth a deeper look.

DOWNLOADING THE DATA

The home page for IPEDS, which is part of the National Center for Education Statistics, is available at https://nces.ed.gov/ipeds/. To access the data, click "Use The Data," as shown below:

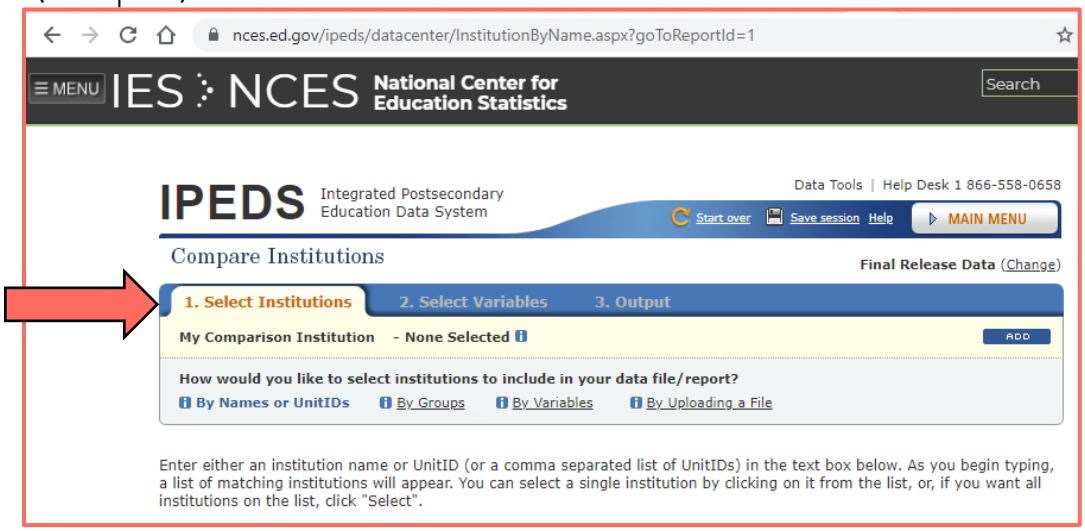


Next, Click "Compare Institutions":



That will bring up the portal we'll use to tell IPEDS the data we want to analyze. As shown below, there are three steps to building the dataset:

- Selecting the institutions to include in the dataset. ("Select Institutions")
- Selecting the details about each institution to show in the dataset. (Select Variables")
- Selecting whether to display the results on the screen or download them. ("Output")



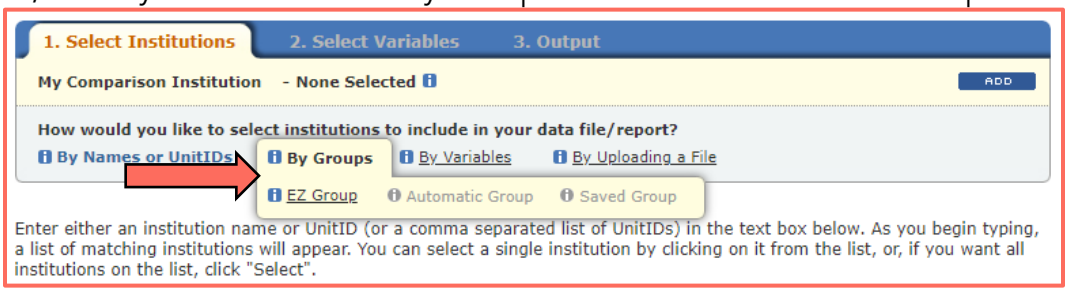
(For detailed instructions on using the "Compare Institutions" search, go to https://nces.ed.gov/ipeds/Help/View/1)

For this exercise, we'll look for schools that have improved their retention rate during the last decade. So we'll:

- compare retention rates for 2010 and 2018 (the most-recent year available); and
- restrict the search to four-year public and private non-profit schools that award a bachelors degree or above.

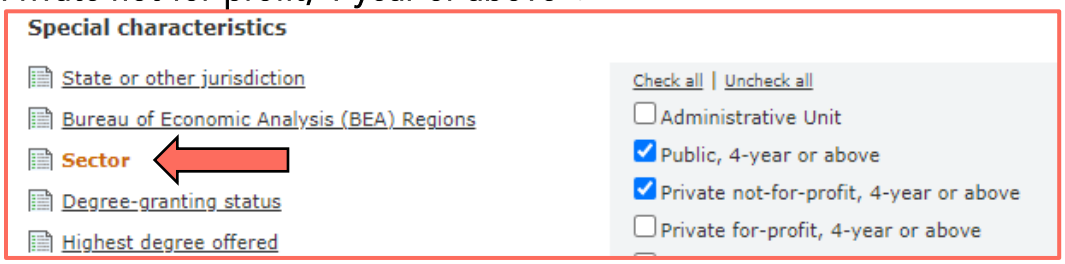
STEP ONE: SELECTING SCHOOLS

Begin by clicking the "Select Institutions" tab. Schools can be selected in a variety of ways, including selecting individual schools by name, and uploading a pre-determined list of schools. For this exercise, we'll select schools by the criteria described above. To start, hover your mouse over "By Groups" and then click on "EZ Group":

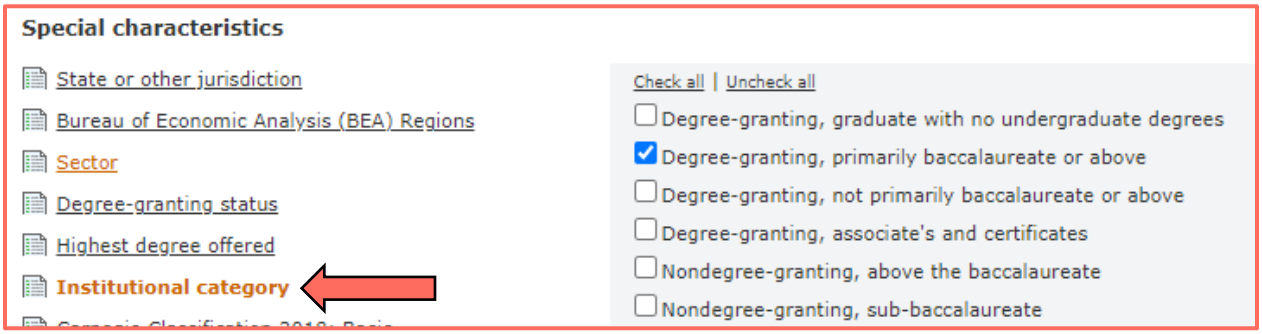


That will bring up a variety of "Special characteristics" to choose from. We'll use two of these.

First, **click on "Sector"** and put check marks in the boxes for **"Public, 4-year or above"** and **"Private not-for-profit, 4-year or above"**:



Next, **click "Institutional category"** and select **"Degree-granting, primarily baccalaureate and above"**:



Now click the "Search" button, which will display a list of the schools that match the criteria:

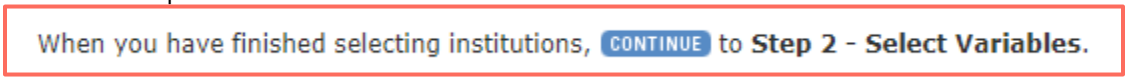
That's Step One. In the next step, we'll identify the datapoints, or "variables" that we want to include for each school. Each variable we add will appear in its own column in the spreadsheet we create. That will allow us to sort or filter by different variables.

STEP TWO: SELECTING VARIABLES

For this exercise, in addition to the name of the school and the "sector" (public or private), we'll include:

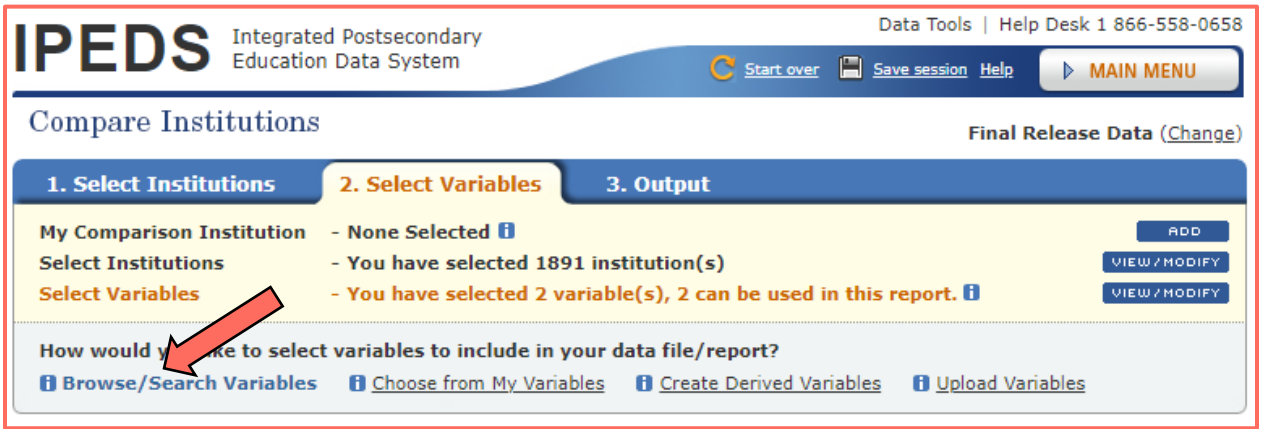
- The state abbreviation
- Whether the school is an HBCU
- The school's full-time enrollment for fall of 2010 and fall of 2018
- The school's retention rate for full-time students for the fall of 2010 and the fall of 2010

To start Step Two, click either the "Continue" button:

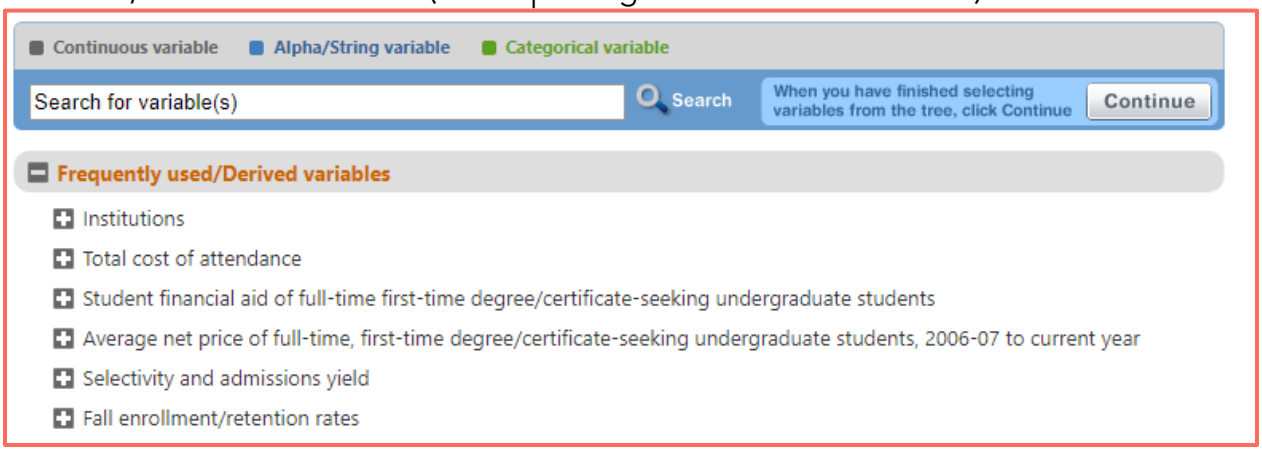


or click "Select Variables" at the top of the page.

Next click "Browse/Search Variables":



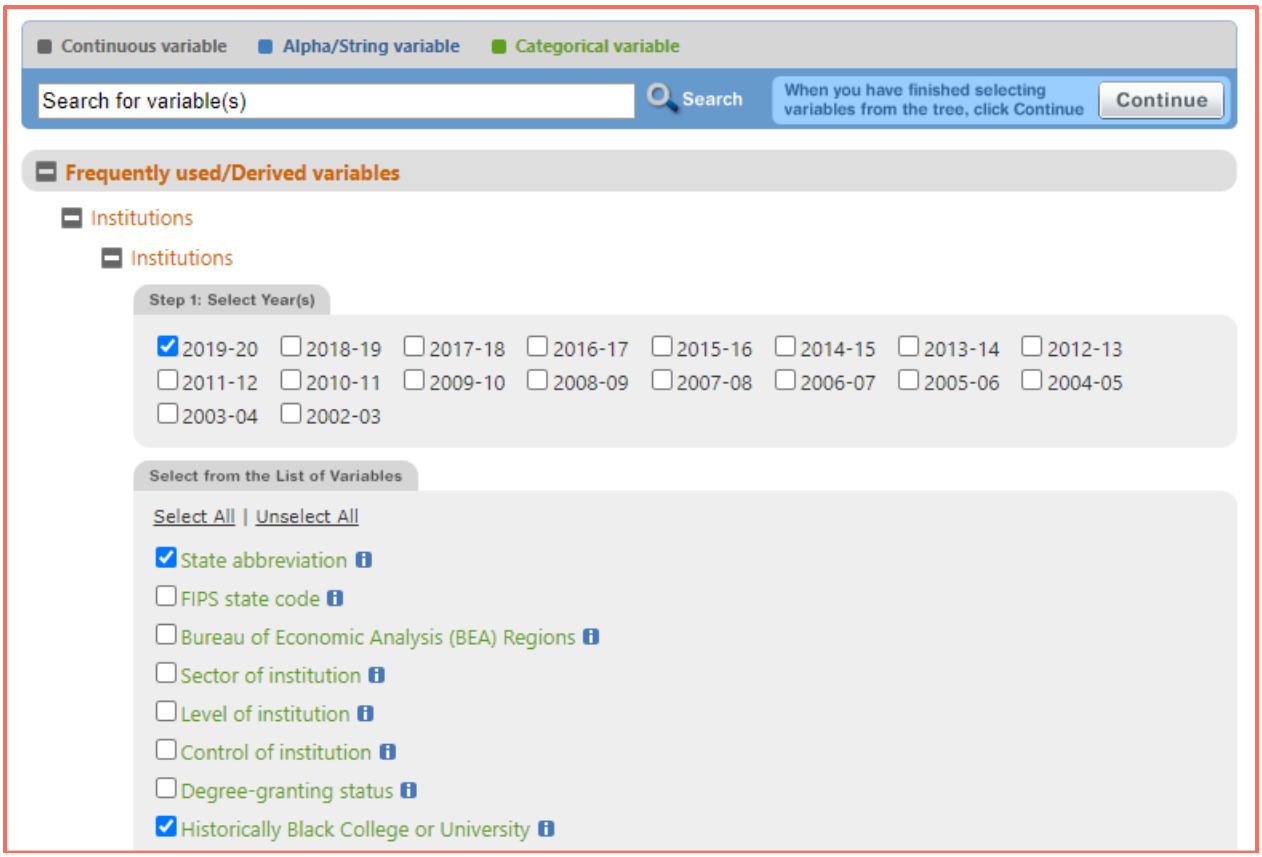
All of the data points we want can be found under "Frequently Used/Derived variables," so click that name (or the plus sign to the left of the label):



That will display a number of categories (some of which are shown above). Feel free to click through these to see the types of variable that can be added. For our exercise, we'll:

- Click "Institutions"
- Click the subcategory "Institutions"
- Select "2019-20"
- Select "State abbreviation"
- Select "Historically Black College or University"

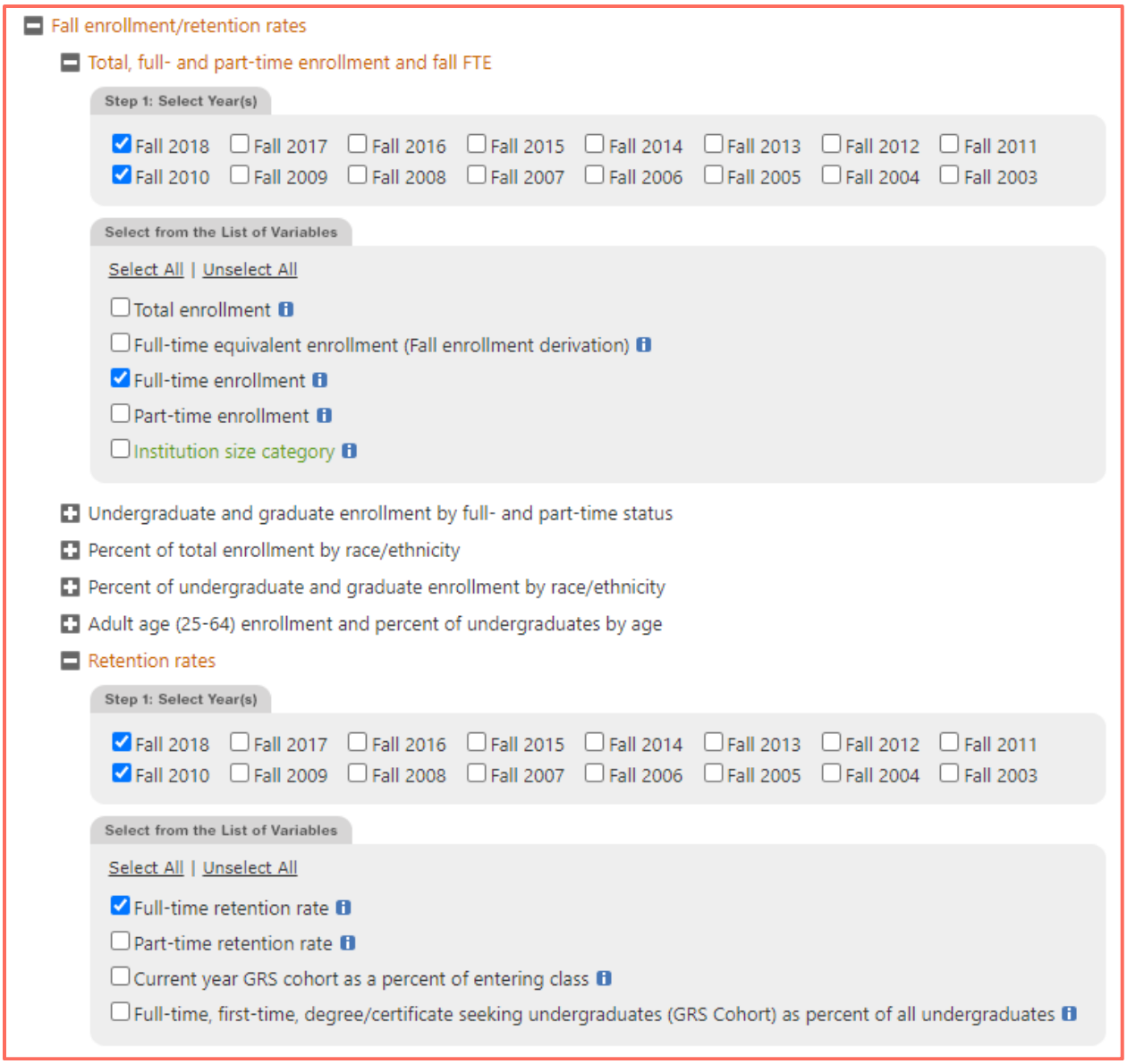
That will create columns in the spreadsheet showing the college's location and an indication of whether the school is an HBCU:



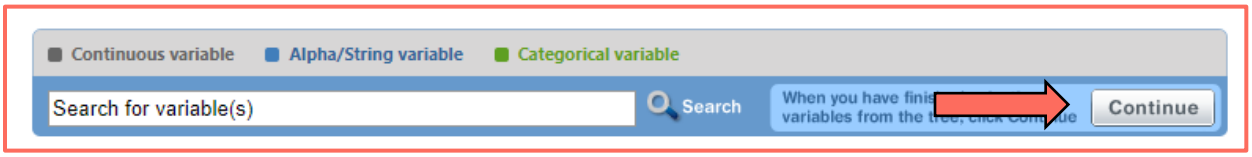
To include columns with the enrollment figures and retention rates, we will:

- Click "Fall enrollment/retention rates"
- Click "Total, full- and part-time enrollment and fall FTE"
- Select "Fall 2018" and "Fall 2010"
- Select "Full-time enrollment"

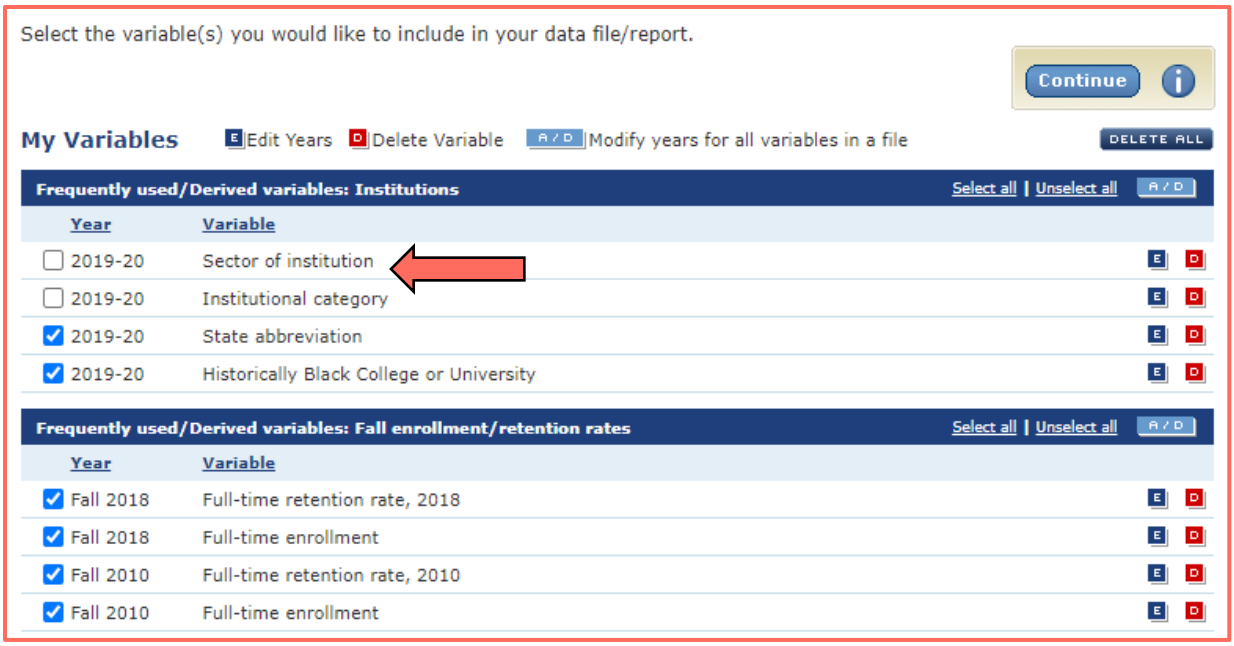
- Click "Retention rates"
- Select "Fall 2018" and "Fall 2010"
- Select "Full-time retention rate"



When the selections have been made in both the "Institutions" and "Fall enrollment/retention rates" categories, go back to the top of the page and click the "Continue" button:



That will display the variables you have selected, as well as the criteria used to narrow down the schools to include the spreadsheet (sector and Institutional category):



The selected variables will have check marks next to them, indicating that they will appear in the spreadsheet. But if you want to include the sector and institutional category, you'll have to click the boxes next to those labels. We only selected a single institutional category ("Degree-granting, primarily baccalaureate and above"), so there's no point including that value, but we'll click the box next to "Sector of Institution" to include an indication showing whether the school is public or private.

When that's done, click the "Continue" button to move to Step Three.

STEP THREE: SELECTING THE OUTPUT

In this step, we'll tell the IPEDS portal whether we want to view the data on the computer screen or download it as a file that can be viewed with a spreadsheet program. We want to analyze the data, so we'll click the button to "Download in comma separated format." That is a file format that can be read directly by Excel, Google Sheets and other programs.

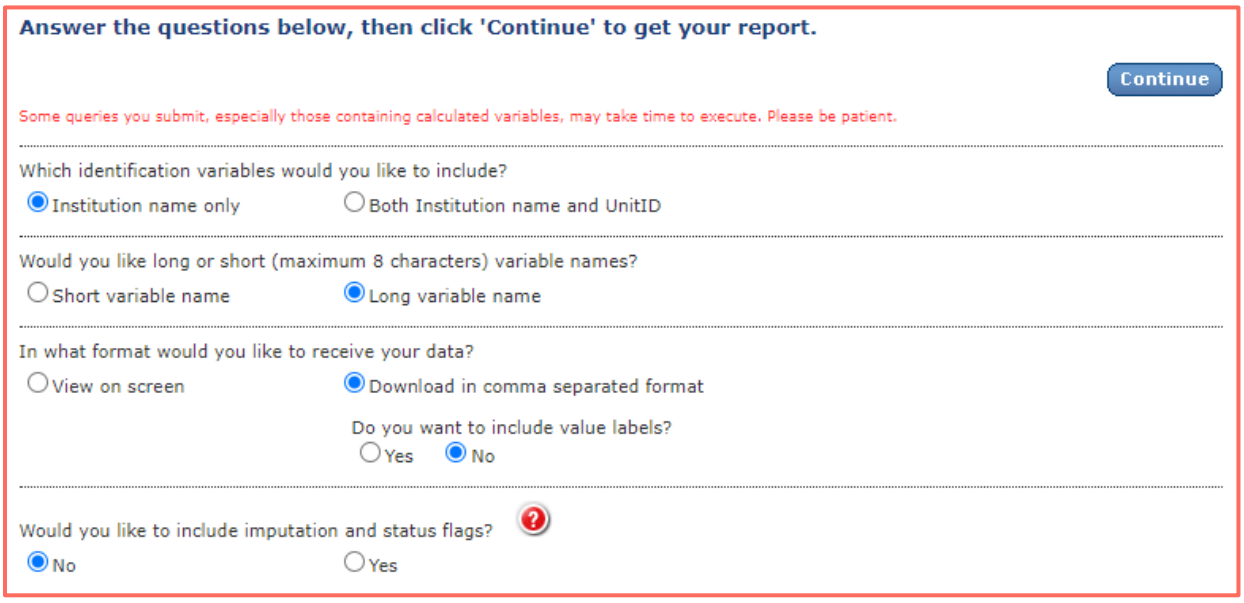
When making that selection, you'll also be asked: "Do you want to include value labels?" Selecting "yes" will download an additional file translating any codes used in the main data file. For our exercise, there are only two variables with codes, so to simplify things, select "No" and I'll list the codes here:

- For Institution Sector:
  - 1 = "Public, 4-year or above"
  - 2 = "Private not-for-profit, 4-year or above"
- For Historically Black College or University:
  - 1 = "Yes"
  - 2 = "No"

There are also options to:

- Include an ID number for the school in addition to the name (that won't be necessary)
- Choose short or long names for the variables (although we will later change the variable names, at this stage, long names will be easier to understand)
- Include "imputation and status flags" (these are not relevant to our research)

So the selections will look like this:

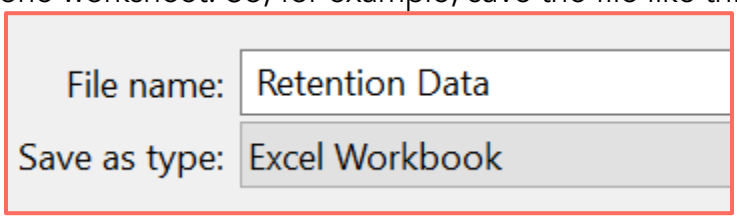


Now click the "Continue" button to download the .csv file – which will open in Excel if the program is on your computer.

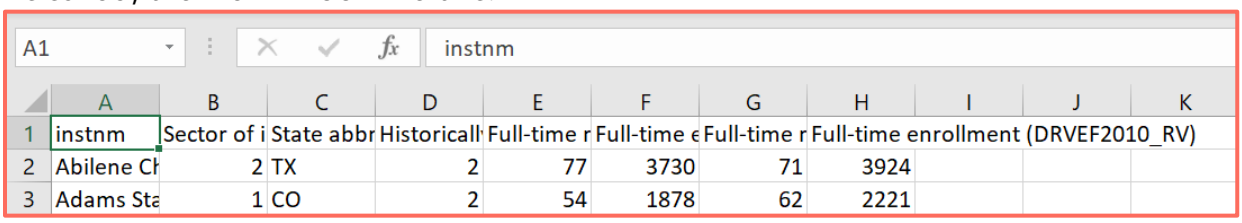
PREPARING THE WORKSHEET

Data from IPEDS tends to be very clean in the sense that you are not likely to find erroneous data, or data stored in the wrong column. But we'll make some adjustments to make the file easier to work with.

But first: If working in Excel, save the file – using the "Save As..." option and choosing "Excel Workbook" as the "Save as type" (while also giving the file an appropriate File name). Saving the file as an Excel Workbook is important because Excel cannot save multiple worksheets (the tabs at the bottom of the page) in a .csv file, and we'll be using more than one worksheet. So, for example, save the file like this:

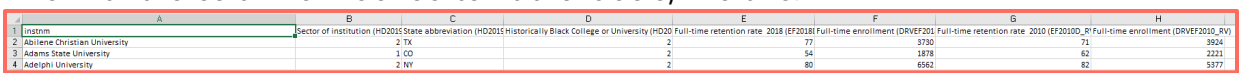


As saved, the file will look like this:



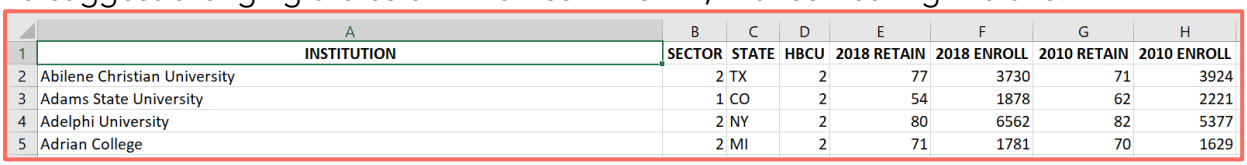
	A	B	C	D	E	F	G	H	I	J	K
1	instnm	Sector of i	State abbr	Historicall	Full-time r	Full-time e	Full-time r	Full-time enrollment (DRVEF2010_RV)			
2	Abilene Ch	2	TX	2	77	3730	71	3924			
3	Adams Sta	1	CO	2	54	1878	62	2221			

And with the columns widened to fit the labels, like this:



	A	B	C	D	E	F	G	H
1	instnm	Sector of institution (HD2015	State abbreviation (HD2015	Historically Black College or University (HD20	Full-time retention rate 2018 (EF2018	Full-time enrollment (DRVEF201	Full-time retention rate 2010 (EF2010D_R	Full-time enrollment (DRVEF2010_RV)
2	Abilene Christian University	2	TX	2	77	3730	71	3924
3	Adams State University	1	CO	2	54	1878	62	2221
4	Adelphi University	2	NY	2	80	6562	82	5377

I'd suggest changing the column names in Row 1, with something like this:

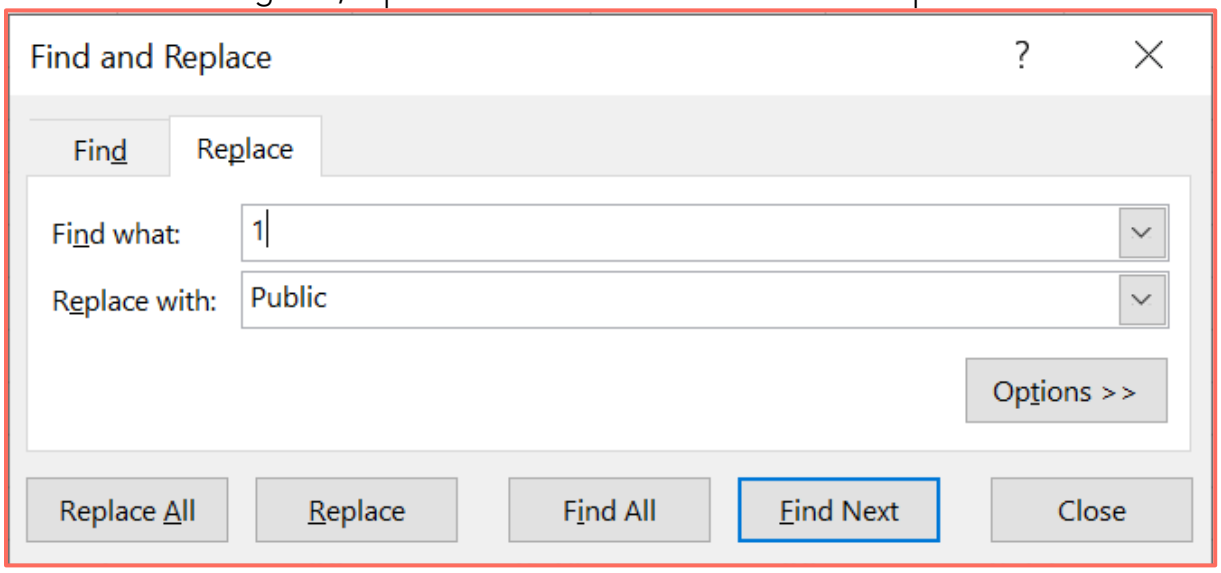


	A	B	C	D	E	F	G	H
1	INSTITUTION	SECTOR	STATE	HBCU	2018 RETAIN	2018 ENROLL	2010 RETAIN	2010 ENROLL
2	Abilene Christian University	2	TX	2	77	3730	71	3924
3	Adams State University	1	CO	2	54	1878	62	2221
4	Adelphi University	2	NY	2	80	6562	82	5377
5	Adrian College	2	MI	2	71	1781	70	1629

Optionally, you may also use the "Replace" function (found under "Find & Select" in Excel or "Find and Replace" in Google Sheets – or by typing CTRL-H with either program) to replace the numerical codes for the Sector and HBCU values.

For example:

- Select all of Column B (by clicking on the "B" at the top of the column)
- Select CTRL-H
- In the dialog box, replace "1" with "Public" and click "Replace All":



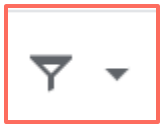
In the same box, replace "2" with "Private" and click "Replace All"

Then repeat the process with the HBCU column, replacing "1" with "Y" and "2" with "N."

We now have an easy-to-read file. But to assure a meaningful analysis, we will do some minor data manipulation by filtering out schools that are very small or have missing (or questionable) retention data. We'll do that by applying a filter to the columns to limit the visible schools to only those that meet the following criteria:

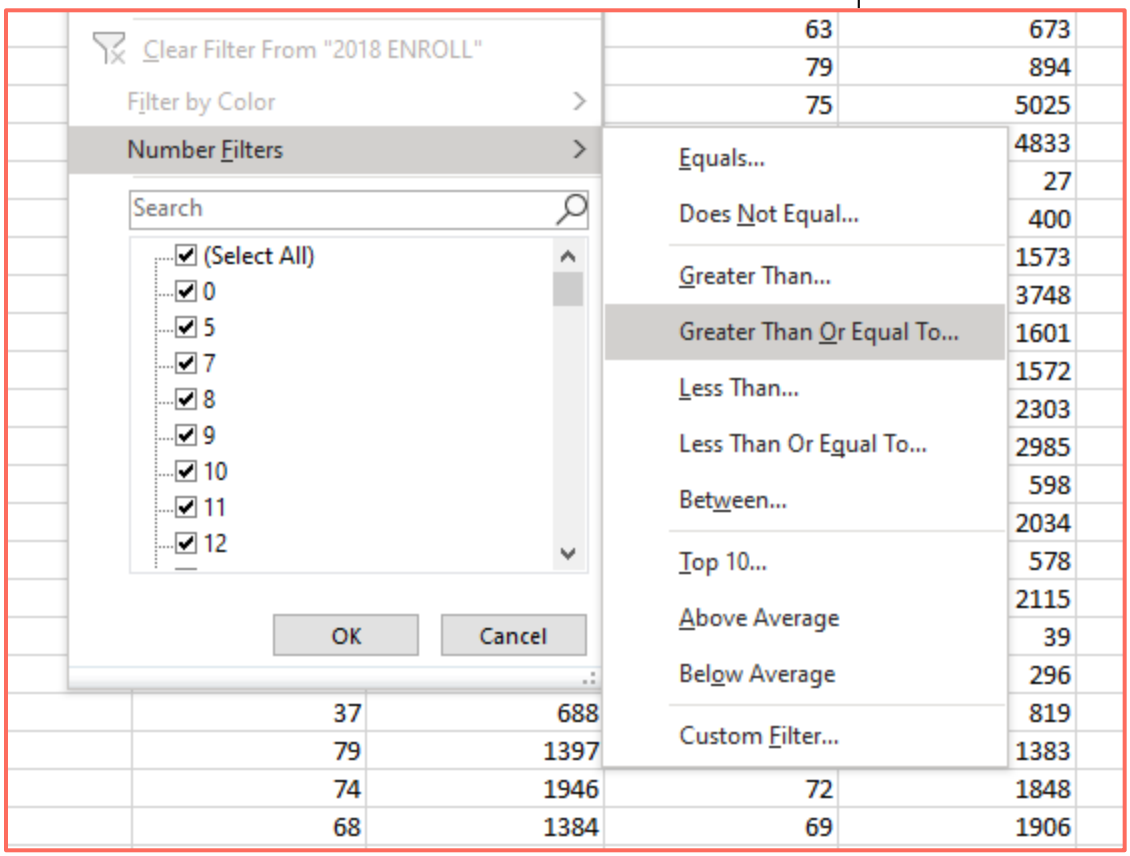
- Enrollment of at least 1,000 full-time students in both 2010 and 2018
- Retention figures are not blank or zero in either year.

In Excel, activate the filter by clicking any cell that contains data and clicking "Filter" from the Data tab (or clicking "Sort & Filter" from the Home tab and then selecting "Filter"). In Google Sheets, click the Filter icon:

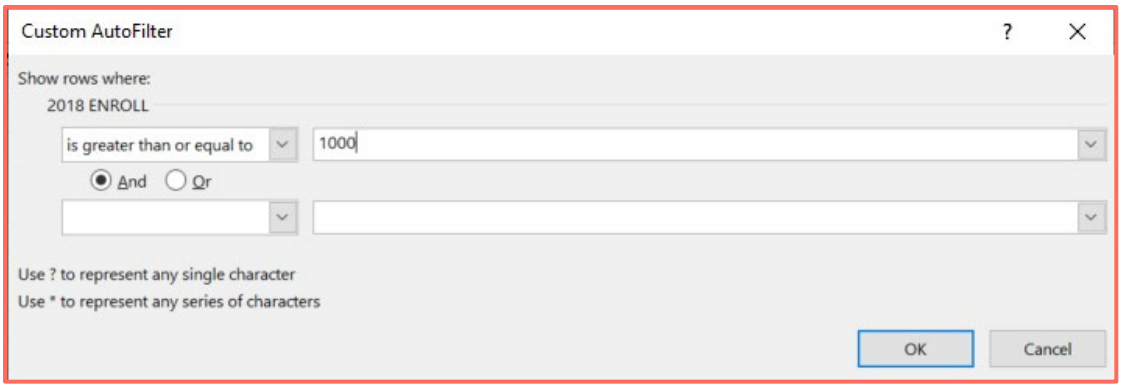


Small triangles will appear next to the column names in Row 1.

Click the arrow next to "2018 Enroll," and from the pull-down menu, hover the mouse over "Number Filters" and click on "Greater Than Or Equal To...":

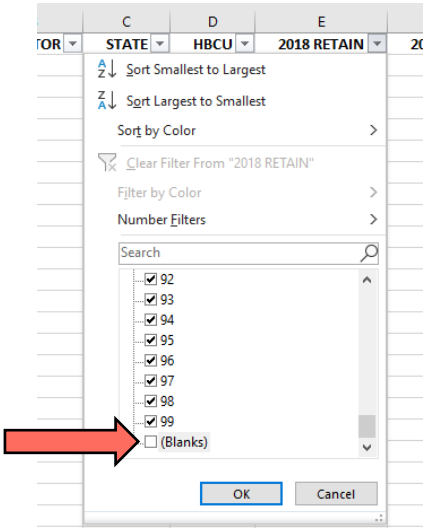


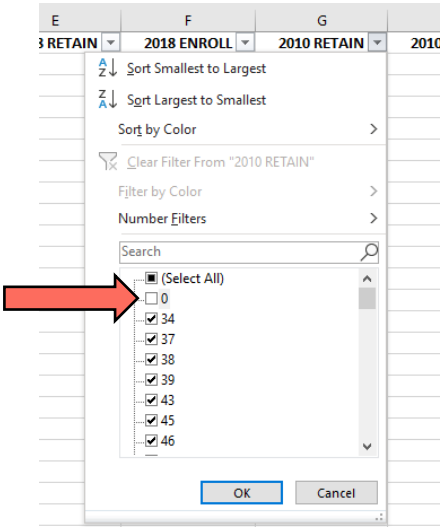
In the dialog box that appears, type "1000" and click "OK." This will tell Excel to temporarily hide all rows where the 2018 Enrollment is not at least 1,000:



Repeat that for the 2010 Enrollment.

Now click the filter arrow next to the "2018 RETAIN" label in Column E, and from the pull-down menu, scroll to the bottom and click the box next to "(Blanks)" to remove the checkmark, thereby telling Excel to not include blanks in the filtered spreadsheet. Click "OK" and repeat this task for the "2010 RETAIN" column – while also removing the checkmark in the box next to the "0" value:





Note that while the images above are from Excel, the steps are generally the same if working with Google Sheets.

At this point it would make sense to copy the filtered data to a new worksheet. Do that by:

- Clicking on the last cell with data in the filtered spreadsheet. (In the file I'm working with, it is Cell H1891.)
- Then use the scroll bars to move to the beginning of the sheet.
- Hold down the Shift key and click Cell A1. That will select the entire filtered data set.
- Copy the sheet by typing CTRL-C (or Command-C on a Mac).

- At the bottom of the screen (in either Excel or Google Sheets), click the + sign to add a new sheet.
- Click on Cell A1 of the new sheet
- Paste the copied data by typing CTRL-V (or Command-V on a Mac).

We now have a clean worksheet with all the schools that have at least 1,000 full-time students and have retention data for 2010 and 2018.

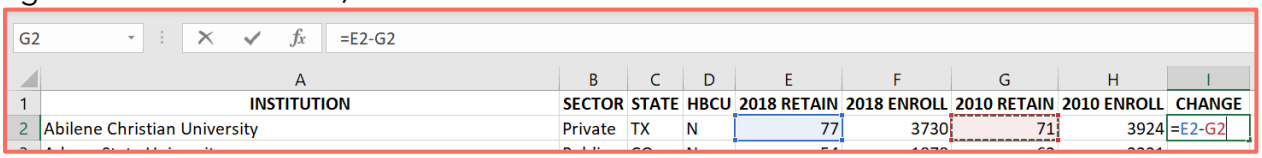
ANALYZING THE DATA

And at long last, we're ready to search for positive deviants, with a fairly simple analysis: We'll calculate the increase or decrease in the percentage of students retained in 2018 compared to 2010, and then sort the sheet so that the schools with the greatest increase are listed at the top.

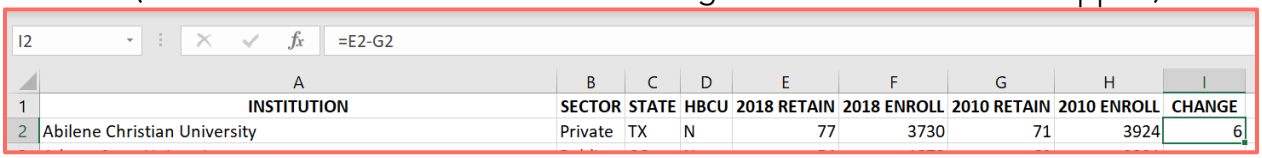
First, we'll add the calculation. In Cell I1, type "CHANGE" (or a similar label of your choosing. Then in Cell I2, type the formula that will take the 2018 retention figure and subtract the 2010 retention figure. That will produce a result showing the increase or decrease between the two years. As shown in the image below, the formula is:

=E2-B2

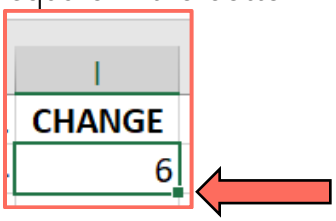
(assuming, as in my example, the 2018 retention figures are in Column E and the 2010 figures are in Column G).



Type that formula in Cell I2 and hit Enter, and it will calculate the change for the school in Row 2 (and the result would include a minus-sign if the retention rate dropped):



Click again in Cell I2, and notice the square in the bottom-right corner:



That is the Fill Handle and is used to copy a formula down the sheet in Excel or Google Sheets. Hover the cursor over the Fill Handle so that the cursor changes to a thin plus-

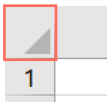
sign. Then double-click the Fill Handle, and the formula in Cell I2 will be copied down to all of the cells in Column I.

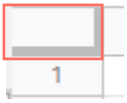
We've now calculated the increase or decrease in retention rates for all schools:

	A	B	C	D	E	F	G	H	I
1	INSTITUTION	SECTOR	STATE	HBCU	2018 RETAIN	2018 ENROLL	2010 RETAIN	2010 ENROLL	CHANGE
2	Abilene Christian University	Private	TX	N	77	3730	71	3924	6
3	Adams State University	Public	CO	N	54	1878	62	2221	-8
4	Adelphi University	Private	NY	N	80	6562	82	5377	-2
5	Adrian College	Private	MI	N	71	1781	70	1629	1
6	Alabama A & M University	Public	AL	Y	61	5492	75	5025	-14
7	Alabama State University	Public	AL	Y	59	3934	64	4833	-5
8	Albany College of Pharmacy and Health Sciences	Private	NY	N	81	1319	72	1573	9

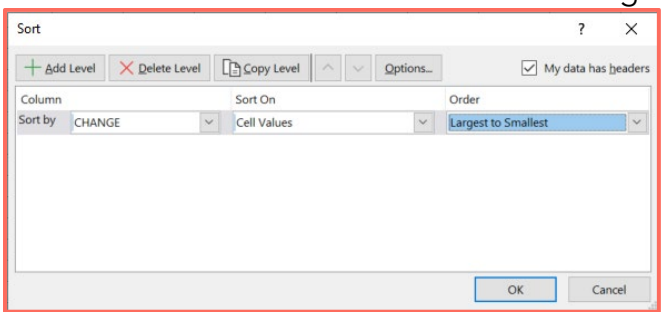
The final task will be to sort the data. To sort:

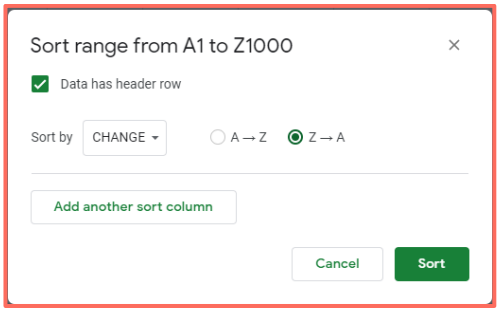
- Define the entire sheet by clicking the box that is above Row 1 and to the left of Column A:

Excel: 

Google Sheets: 

- From the Data tab, select "Sort" in Excel, or "Sort Range" in Google Sheets.
- From the dialog box, make sure there's a check mark in the box indicating the data has headers (these are the labels in Row 1).
- In the "Sort by" box, select the label for the column with the calculated change in retention rates (In my example, the column is labeled "CHANGE.")
- Select to sort the data from high to low, selecting "Largest to Smallest" under "Order" in Excel, or clicking the "Z → A" button in Google Sheets.
- Click "OK" in Excel or "Sort" in Google Sheets:





The sheet will now be sorted to show schools with the greatest increase in retention rates between the two years:

	A	B	C	D	E	F	G	H	I
1	INSTITUTION	SECTOR	STATE	HBCU	2018 RETAIN	2018 ENROLL	2010 RETAIN	2010 ENROLL	CHANGE
2	Bellevue University	Private	NE	N	77	5586	38	6639	39
3	Harris-Stowe State University	Public	MO	Y	65	1398	39	1233	26
4	Roosevelt University	Private	IL	N	74	3082	56	3844	18
5	Virginia Union University	Private	VA	Y	68	1400	50	1606	18
6	Jacksonville University	Private	FL	N	77	2702	60	2431	17
7	Regis College	Private	MA	N	85	1320	68	1031	17
8	Catawba College	Private	NC	N	83	1229	67	1234	16
9	Shawnee State University	Public	OH	N	70	2680	54	3892	16
10	The University of Texas at San Antonio	Public	TX	N	73	24415	57	22497	16

NEXT (AND IMPORTANT!) STEPS

The schools at the top of the list are your positive deviants – institutions that stand out in the data for their far-above-average performance. But as with all data journalism, finding positive deviants in data should be viewed as a "tip" – a lead to a possible story. There's still reporting to be done to investigate these schools more deeply, perhaps by viewing additional years to identify a steady trend rather than a fluke. And for Solutions Journalism in particular, you'll need to determine if the improvement is linked to actual policies, rather than simply good fortune or some external factor.

And you might want to examine certain types of schools, investigating the best-performing public institutions on the list, or focusing on HBCUs. Researching retention policies for the top-performing HBCU – Harris-Stowe State University in Missouri – could prove particularly fruitful.

Naturally, this exercise not only helps identify the schools with the greatest increase in retention, it also – at the bottom of the spreadsheet – identifies schools that have lost the most ground in keeping students enrolled. Reporting on those poor performers can also make for important journalism – and many reporters will be drawn to that more-traditional approach. But particularly if reporting on institutions with dismal track records, keep in mind that applying a solutions focus to the story – by including clear-eyed coverage of the better performers and how they found success – will always lead to better accountability and public-service journalism, with better audience engagement.