

Tracking How Federal Health Cuts Impact Your Communities

2025 National Fellowship
USC ANNENBERG
CENTER FOR HEALTH JOURNALISM

<https://github.com/statnews/uscfellowship25>

PRESENTED BY

J. Emory Parker
STAT News

EMAIL	emory.parker@statnews.com
BLUESKY	@jaspar.bsky.social
GITHUB	@jemory

**Backing up every
CDC dataset**

Deletion of data on CDC's site

Total number of results available for download on data.cdc.gov

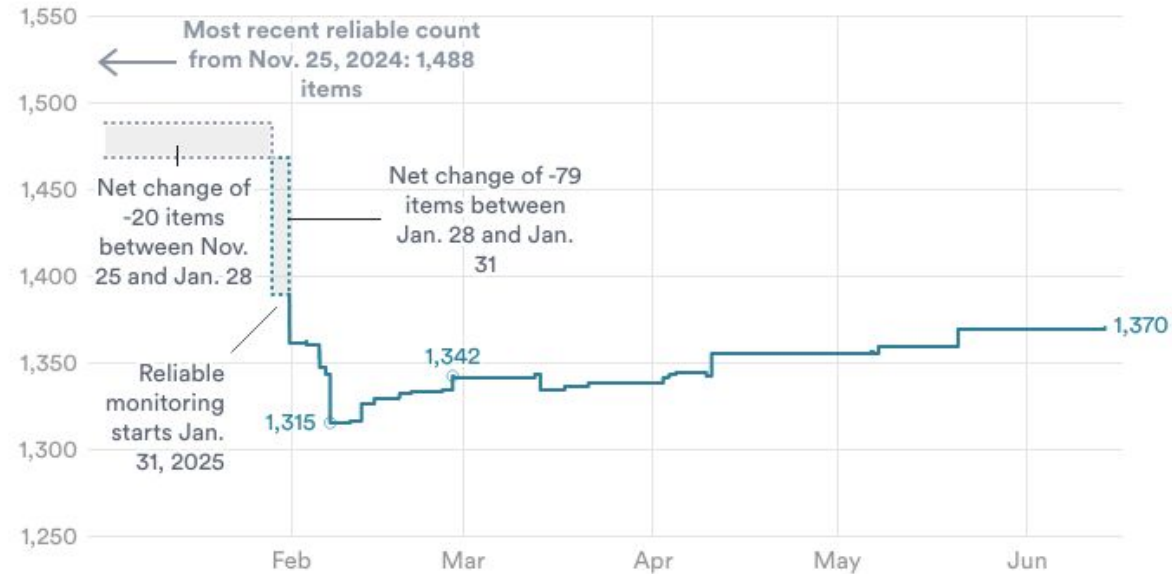


Chart: J. Emory Parker/STAT • Source: STAT Tracking of CDC APIs

Reconstructing the past

STAT Reporting from the frontiers
of health and medicine

Date ▼	Change	Name	ID
2025-06-13 12:30:33	Added	hcp_influenza	89x6-rgq5
2025-05-20 15:19:07	Added	DASH GSHS - No Fast Food (2015)	4bif-w52v
2025-05-20 15:19:07	Added	DASH YRBSS - Did not participate in 60 minutes of activity 1x/weekly (MS)	8u74-z25n
2025-05-20 15:19:07	Added	DASH YRBSS - Physical Fight (MS)	dr7j-7p7w
2025-05-20 13:21:14	Added	DASH YRBSS - Students who are Currently Sexually Active (HS)	5dm2-74uk
2025-05-20 13:21:14	Added	DASH - Youth Risk Behavior Surveillance System (YRBSS): Middle School	k5bc-k3g8
2025-05-20 13:21:14	Added	DASH - Global School-based Student Health Survey (GSHS)	pxpe-pgrg
2025-05-20 13:21:14	Added	DASH - Youth Risk Behavior Surveillance System (YRBSS): High School – Including Sexual Orientation	q6p7-56au

[Get the data](#)

Tracking changes as they happen

STAT Reporting from the frontiers
of health and medicine

Row	...	Topic	Data_Value	...	StratificationCategory1	Stratification1	StratificationCategory2	Stratification2
1	...	Heart Disease Mortality	182.4	...	GenderSex	Overall	Race/Ethnicity	Overall
2	...	Heart Disease Mortality	172.6	...	GenderSex	Overall	Race/Ethnicity	Overall
3	...	Heart Disease Mortality	255.6	...	GenderSex	Overall	Race/Ethnicity	Overall
4	...	Heart Disease Mortality	343.4	...	GenderSex	Overall	Race/Ethnicity	Overall
5	...	Heart Disease Mortality		...	GenderSex	Overall	Race/Ethnicity	Overall
6	...	Heart Disease Mortality	218.6	...	GenderSex	Overall	Race/Ethnicity	Overall
...
59091	...	Heart Disease Mortality		...	GenderSex	Male	Race/Ethnicity	Asian and Pacific Islander
59092	...	Heart Disease Mortality		...	GenderSex	Female	Race/Ethnicity	Asian and Pacific Islander
59093	...	Heart Disease Mortality	338.4	...	GenderSex	Overall	Race/Ethnicity	American Indian and Alaskan Native
59094	...	Heart Disease Mortality	432.2	...	GenderSex	Male	Race/Ethnicity	American Indian and Alaskan Native
59095	...	Heart Disease Mortality		...	GenderSex	Female	Race/Ethnicity	American Indian and Alaskan Native

Drilling down into changes

Backup copies of data that's been removed from data.cdc.gov

Page 1 of 6 >

id	Name ▲	Mentions "gender"	Download	File size	File type
vtwh-8kxg	Adult Tobacco Consumption In The U. S. Glossary And Methodology		Download	0.07 MB	xls
hfr9-rurv	Alzheimer's Disease and Healthy Aging Data	⚠	Download	120.26 MB	csv
xs7u-t3bn	Alzheimer's Disease and Healthy Aging Indicators: Caregiving	⚠	Download	10.21 MB	csv
jhd5-u276	Alzheimer's Disease and Healthy Aging Indicators: Cognitive Decline	⚠	Download	10.80 MB	csv
thir-stei	American Community Survey (ACS) – Vision and Eye Health Surveillance	⚠	Download	334.65 MB	csv
4ht3-nbmd	Archived Cumulative Data: Percent of pregnant people aged 18-49 years receiving at least one dose of a COVID-19 vaccine during pregnancy overall, by race/ethnicity, and date reported to CDC-Vaccine Safety Datalink, United States / December 20, 2020 – Jan		Download	0.04 MB	csv
fpp2-pp25	Behavioral Risk Factor Data: Tobacco Use (2010 And Prior)		Download	9.80 MB	csv
wsas-xwh5	Behavioral Risk Factor Data: Tobacco Use (2011 to present)		Download	11.80 MB	csv
ikwk-8git	Behavioral Risk Factor Surveillance System (BRFSS) - National Cardiovascular Disease Surveillance Data	⚠	Download	54.57 MB	csv
iuq5-y9ct	Behavioral Risk Factor Surveillance System (BRFSS) Historical Questions	⚠	Download	1.99 MB	csv

Making the data available

STAT Reporting from the frontiers
of health and medicine

Tracking deficits in NIH grant funding

NIH grant awards are lagging behind previous years

Cumulative amount of new grants awarded during the first 110 days of the year

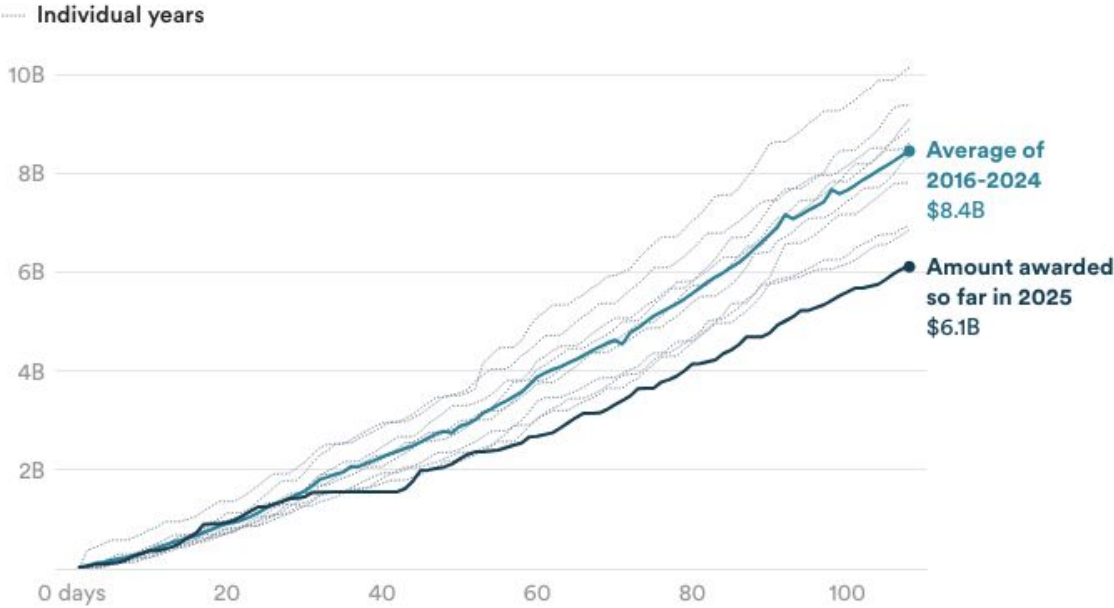


Chart: J. Emory Parker/STAT • Source: STAT Analysis of RePORTER data

NIH institutes and centers with largest funding declines

Cumulative amount of new NIH grants awarded in the first 100 days of 2025 compared with the average for the same period from 2016-2024.



Chart: J. Emory Parker/STAT • Source: STAT Analysis of RePORTER data

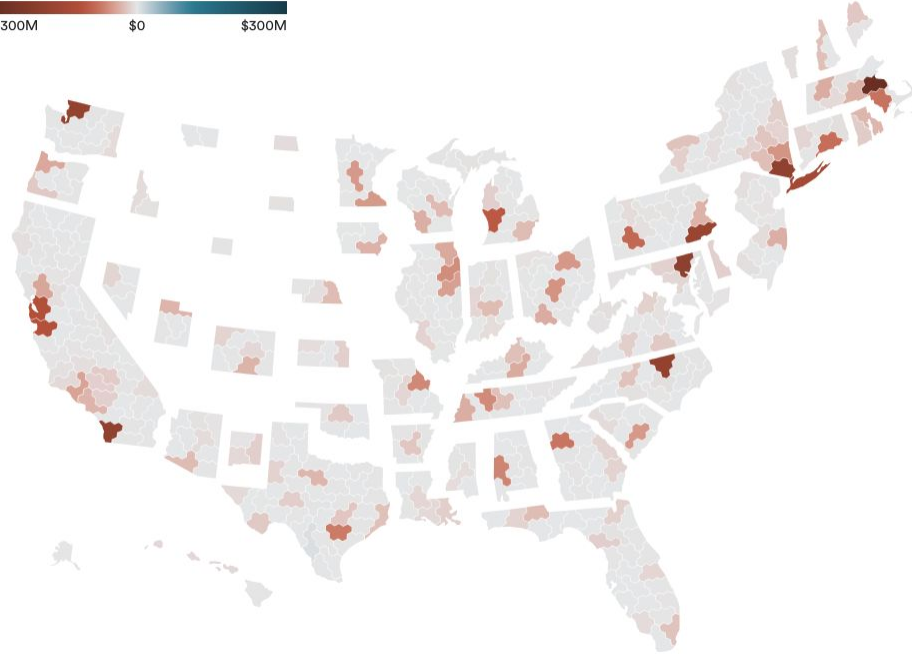
Drilling down

STAT Reporting from the frontiers
of health and medicine

Big cities and university towns lost the most from decline in grant awards

Change in the cumulative amount of new NIH grants awarded by congressional district in the first 110 days of 2025 compared with the average for the same period from 2016-2024.

-\$300M \$0 \$300M



Note: In this image the size of states has been scaled by population to make every district the same size. The shapes of districts have also been simplified for legibility. This exaggerates the geographic size of cities to enhance visibility. The position of districts may not match their real-world locations.

Map: J. Emory Parker/STAT • Source: STAT Analysis of RePORTER data

Localizing our findings

STAT Reporting from the frontiers
of health and medicine

10 institutions that had the largest drop in new NIH grants

Cumulative amount of new NIH grants awarded to research institutions in first 100 days of 2025 compared with the median for the same period in 2016-2024. It does not include grants that were terminated or frozen.



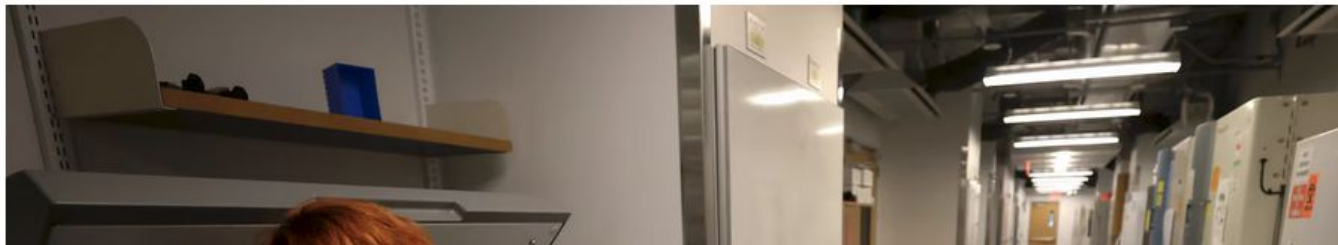
Chart: J. Emory Parker/STAT • Source: STAT Analysis of RePORTER data

Funding is drying up, and so are career options. Some budding scientists are debating abandoning their research.

Across New England and the country, thousands of budding scientists have awoken to a stark new reality

By [Chris Serres](#) and [J. Emory Parker](#) Globe Staff and STAT, Updated May 27, 2025, 9:58 a.m.

    169

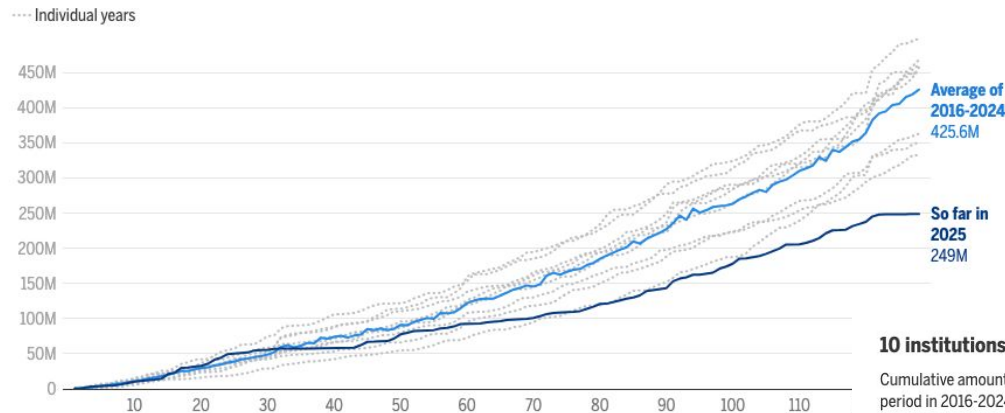


Working with The Globe

STAT Reporting from the frontiers
of health and medicine

\$177 million less awarded for early career grants so far this year

Cumulative dollar values of new grants awarded during the first 128 days of the year

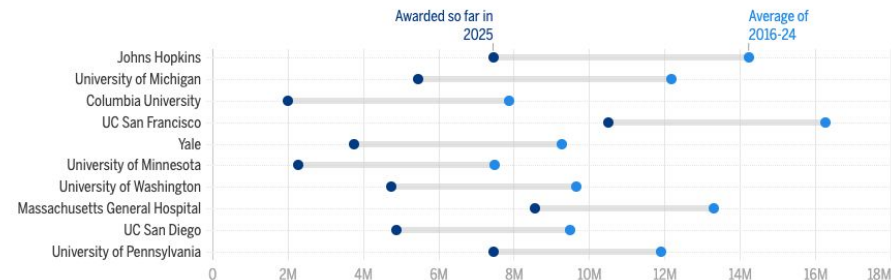


Early career grants have activity codes F30, F31, F32, T32, T34, T35, TL1, K01, K08, K12, K22, K23, K24, K43 and K99.

Chart: J. Emory Parker/STAT • Source: Analysis of RePORTER data

10 institutions that had the largest drop in early career grants

Cumulative amount of new NIH grants awarded to research institutions in first 128 days of 2025 compared with the median for the same period in 2016-2024. It does not include grants that were terminated or frozen.



Early career grants have activity codes F30, F31, F32, T32, T34, T35, TL1, K01, K08, K12, K22, K23, K24, K43 and K99.

Chart: J. Emory Parker/STAT • Source: Analysis of RePORTER data

Results of early career analysis

STAT Reporting from the frontiers
of health and medicine

Tabula can turn
PDFs into
spreadsheets

Tabula

Get it at tabula.technology

Runs entirely on your machine

Only works for text-based pdfs, will not work with scanned pdfs

HHS cancelled grants PDF

Shortened: <https://tinyurl.com/25efmb8z>

Full version:

https://tags.hhs.gov/Content/Data/HHS_Grants_Terminated.pdf

STAT Reporting from the frontiers
of health and medicine

Tabula



Tabula is a tool for liberating data tables locked inside PDF files.

[View the Project on GitHub](#)
[tabulapdf/tabula](#)



Current Version: **1.2.1**

Other Versions: [pre-releases & archives](#)

Need help? [Open an issue on Github.](#)

Donate: Help support this project by [backing us](#) on OpenCollective.

We'd love to hear from you! Say hi on Twitter at [@TabulaPDF](#)

Latest Version: Tabula 1.2.1

June 4, 2018

Tabula 1.2.1 fixes several bugs in the user interface and processing backend. (You can read about all the changes in [the release notes](#).)

Download Tabula below, or [on the release notes page](#).

Special thanks to our [OpenCollective](#) backers for supporting our work on Tabula; if you find Tabula useful in your work, please consider a [one-time](#) or [monthly](#) donation.

How Can Tabula Help Me?

If you've ever tried to do anything with data provided to you in PDFs, you know how painful it is — there's no easy way to copy-and-paste rows of data out of PDF files. Tabula allows you to extract that data into a CSV or Microsoft Excel spreadsheet using a simple, easy-to-use interface. Tabula works on Mac, Windows and Linux.

Who Uses Tabula?

Tabula is used to power investigative reporting at news organizations of all sizes, including [ProPublica](#), [The Times of London](#), [Foreign Policy](#), [La Nación \(Argentina\)](#), [The New York Times](#) and the [St. Paul \(MN\) Pioneer Press](#).

Grassroots organizations like [SchoolCuts.org](#) rely on Tabula to turn clunky documents into human-friendly public resources.

And researchers of all kinds use Tabula to turn PDF reports into Excel spreadsheets, CSVs, and JSON files for use in analysis and database applications.

**Let's give Tabula
a try...**

Tabula

Install and run Tabula. It will open in a browser window.

The main screen will allow you to import new PDFs and access PDFs you've imported in the past.

Select “Browse...” to locate a PDF on your system.

We'll be using the shortened

HHS_Grants_Terminated.pdf file for this example.

Then select “Import” to begin the process of extracting data.



The screenshot shows the Tabula web interface. At the top, it says "Import one or more PDFs". Below this is a "Browse..." button next to a text input field, and an "Import" button. Below the input field is a section titled "Imported PDFs" which contains a table of imported files.

File Name	Size	Pages	Date Added
HHS_Grants_Terminated.pdf	330 kB	4	24 Jun 2025 18:

Tabula

Draw a box around the table on the first page of the document.

If the document contains similar tables on every page, like this example, select “Repeat this Selection” to automatically copy the box to every page.

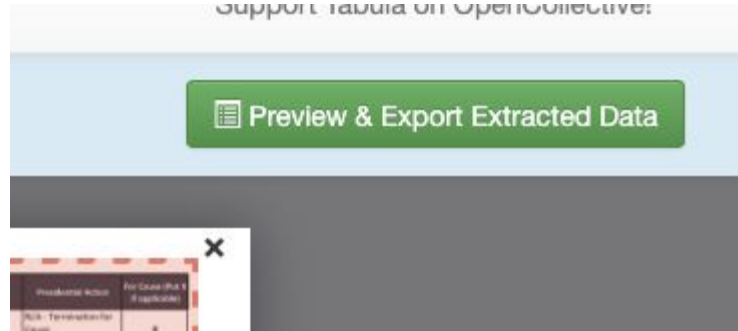
Manually adjust the boxes on each page, if necessary.

The screenshot displays the Tabula web application interface. At the top, there's a navigation bar with 'My Files', 'My Templates', 'About', 'Help', and 'Source Code'. The main area shows a document titled 'HHS_Grants_Terminated.pdf'. Below the document name, there are buttons for 'Templates', 'Clear All Selections', and 'Autodetect Tables'. On the left, a sidebar shows four thumbnails of the document pages, with the first page selected. The main view shows the first page of the document, which contains a large table. A red dashed box is drawn around the table. Below the table, there's a button labeled 'Repeat this Selection'. The bottom of the interface shows the date '06/26/2025' and the page number 'Page 1 of 74'.

06/26/2025 Page 1 of 74 Repeat this Selection

Tabula

Once you have all the tables selecting within red boxes, select “Preview & Export Extracted Data” to proceed to the next step.



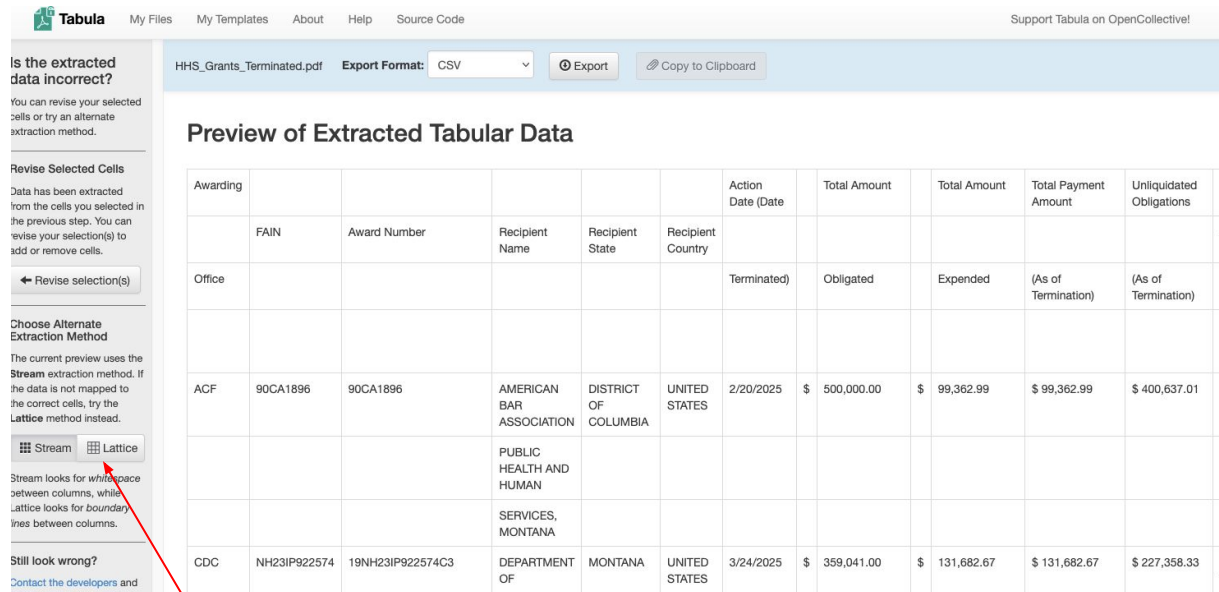
Tabula

You'll get a preview of what the extracted data will look like.

In this case, the preview doesn't look right.

In this document, the table is made up of lines so we need to change the extraction method to "lattice".

Use "stream" when there aren't lines in the document.



Tabula My Files My Templates About Help Source Code Support Tabula on OpenCollective!

is the extracted data incorrect?
You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells
Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.
← Revise selection(s)

Choose Alternate Extraction Method
The current preview uses the Stream extraction method. If the data is not mapped to the correct cells, try the Lattice method instead.

☒ Stream ☐ Lattice

Stream looks for white space between columns, while Lattice looks for boundary lines between columns.

Still look wrong?
[Contact the developers and](#)

HHS_Grants_Terminated.pdf Export Format: CSV Export Copy to Clipboard

Preview of Extracted Tabular Data

Awarding	FAIN	Award Number	Recipient Name	Recipient State	Recipient Country	Action Date (Date)	Total Amount	Total Amount	Total Payment Amount	Unliquidated Obligations
Office						Terminated)	Obligated	Expended	(As of Termination)	(As of Termination)
ACF	90CA1896	90CA1896	AMERICAN BAR ASSOCIATION	DISTRICT OF COLUMBIA	UNITED STATES	2/20/2025	\$ 500,000.00	\$ 99,362.99	\$ 99,362.99	\$ 400,637.01
			PUBLIC HEALTH AND HUMAN SERVICES, MONTANA							
CDC	NH23IP922574	19NH23IP922574C3	DEPARTMENT OF	MONTANA	UNITED STATES	3/24/2025	\$ 359,041.00	\$ 131,682.67	\$ 131,682.67	\$ 227,358.33

Tabula


That looks better!

All that's left is to select an export format (CSV should be fine for most cases) and select "Export".

Tabula will then save a .csv file wherever your browser usually saves downloads.

The CSV file can be opened in Excel or Google Sheets.

STAT Reporting from the frontiers of health and medicine

 **Tabula**

My FilesMy TemplatesAboutHelpSource Code

Support Tabula on OpenCollective

HHS_Grants_Terminated.pdfExport Format: CSVExportCopy to Clipboard

Is the extracted data incorrect?

You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells

Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

Revise selection(s)

Choose Alternate Extraction Method

The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice** method instead.

StreamLattice

Stream looks for whitespace between columns, while Lattice looks for boundary lines between columns.

Still look wrong?

Contact the developers and tell us what you tried to do that didn't work.

Preview of Extracted Tabular Data

Awarding Office	FAIN	Award Number	Recipient Name	Recipient State	Recipient Country	Action Date (Date Terminated)	Total Amount Obligated	Total Amount Expended	Total Payment Amount (As of Termination)	Unliquidated Obligations (As of Termination)	Award Title
ACF	90CA1896	90CA1896	AMERICAN BAR ASSOCIATION	DISTRICT OF COLUMBIA	UNITED STATES	2/20/2025	\$ 500,000.00	\$ 99,362.99	\$99,362.99	\$400,637.01	MICHIGAN OVERREP OUR PEO STOP
CDC	NH23IP922574	19NH23IP922574C3	PUBLIC HEALTH AND HUMAN SERVICES, MONTANA DEPARTMENT OF	MONTANA	UNITED STATES	3/24/2025	\$ 359,041.00	\$ 131,682.67	\$131,682.67	\$227,358.33	Immunizat Vaccines f Children
CDC	NH23IP922574	20NH23IP922574C3	PUBLIC HEALTH AND HUMAN SERVICES, MONTANA DEPARTMENT OF	MONTANA	UNITED STATES	3/24/2025	\$ 872,020.00	\$ 513,715.78	\$513,715.78	\$358,304.22	Immunizat Vaccines f Children
CDC	NH23IP922574	20NH23IP922574C5	PUBLIC HEALTH AND HUMAN SERVICES, MONTANA DEPARTMENT	MONTANA	UNITED STATES	3/24/2025	\$ 11,309,969.00	\$ 7,075,695.96	\$ 7,075,695.96	\$4,234,273.04	Immunizat Vaccines f Children

Export Format:

CSV

Export

Tabula

STAT Reporting from the frontiers
of health and medicine

NIH RePORTER is
useful for looking
tracking grants

RePORTER Basics

<https://reporter.nih.gov/>

Covers grants from NIH

Website is useful for searching for grants, but for more comprehensive analysis you'll need to use the API and have some code knowledge

<https://github.com/dhoconno/reporter>

This resource provides real-time analysis similar to STAT's

File and code here are open source and you're able to freely use and redistribute them

Quick links

All terminated grants:

https://raw.githubusercontent.com/dhoconno/reporter/refs/heads/main/data/processed/taggs/hhs_grants_terminated.csv

All grants:

https://github.com/dhoconno/reporter/raw/refs/heads/main/data/processed/reporter/nih_awards_all.csv.zst

All meetings in federal register:

https://github.com/dhoconno/reporter/raw/refs/heads/main/data/processed/federal_register/nih_fr_meetings_all.csv.zst

Note: the .zst files are compressed.

Uncompressing ZST files

On Mac

Open the Terminal application.

Navigate to the location where of the file.

Example: if it's in your “Downloads” folder

Type: `cd ~/Downloads`

Use the UNZSTD command

Example: to uncompress the file “nih_awards_all.csv.zst”

Type: `unzstd nih_awards_all.csv.zst`

This will create a new file “nih_awards_all.csv” in your Downloads folder

NIH and NSF cancelled grant trackers

Where to access Scott Delaney's data

NIH data

<https://grant-watch.us/nih-data.html>

Click “download all as CSV” to follow along

Access my completed Google Sheet here:

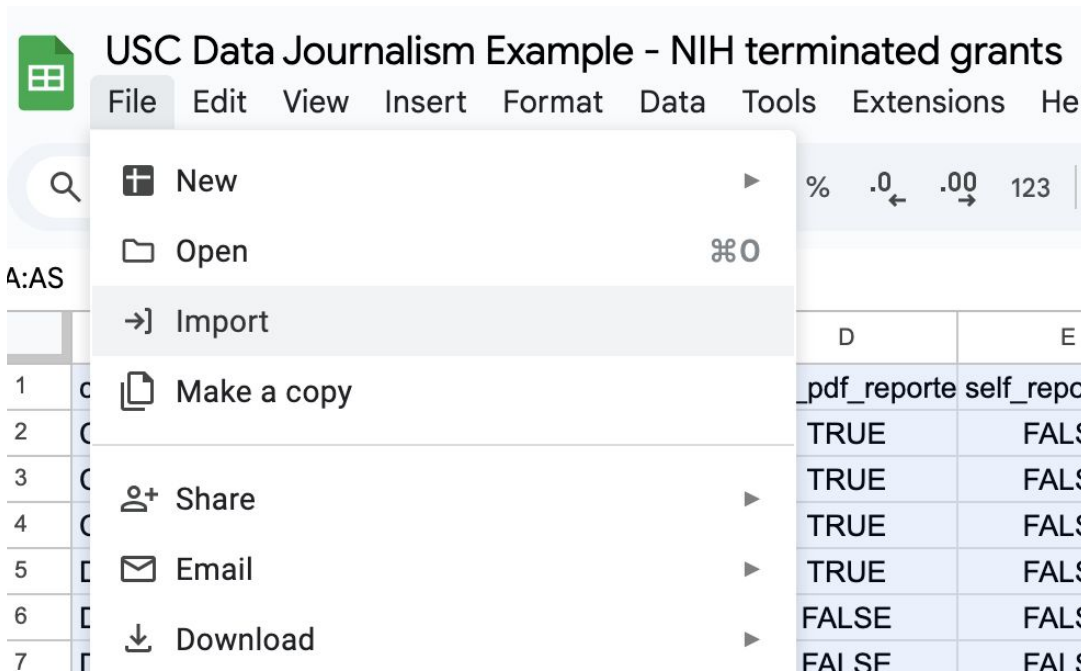
<https://docs.google.com/spreadsheets/d/1G6Y581pqxLy-hEgFUc2HPWvUCc12dbbUGO1theRdSkQ/edit?usp=sharing>

NSF data

<https://grant-watch.us/nsf-data.html>

**Let's try some
quick analysis of
NIH
terminations...**

Getting started

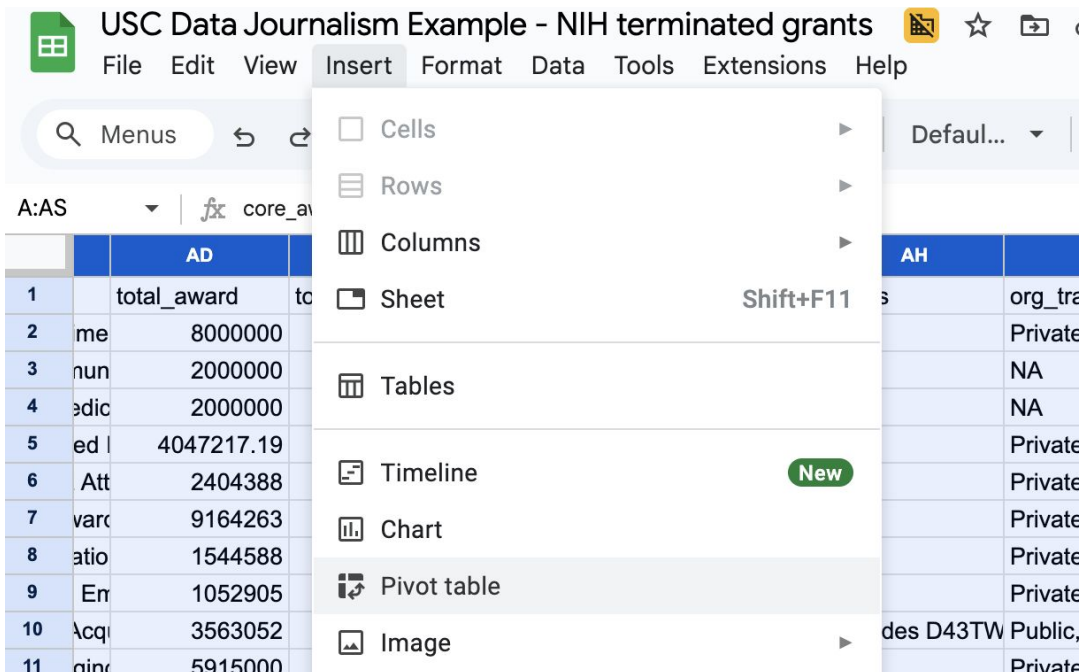


The screenshot shows the Google Sheets interface for a spreadsheet titled "USC Data Journalism Example - NIH terminated grants". The "File" menu is open, and the "Import" option is highlighted. The spreadsheet contains data with columns D and E. The visible data is as follows:

	D	E
1	_pdf_reporte	self_repc
2	TRUE	FALS
3	TRUE	FALS
4	TRUE	FALS
5	TRUE	FALS
6	FALSE	FALS
7	FAI SF	FAI S

Open Google Sheets and import the NIH terminated grants csv file from the previous slide.

Create a pivot table



The screenshot shows a Google Sheets interface for a spreadsheet titled "USC Data Journalism Example - NIH terminated grants". The "Insert" menu is open, and the "Pivot table" option is highlighted. The spreadsheet data is as follows:

	AD	
1	total_award	to
2	me	8000000
3	nun	2000000
4	edic	2000000
5	ed l	4047217.19
6	Att	2404388
7	varc	9164263
8	atio	1544588
9	En	1052905
10	Acq	3563052
11	qinc	5915000

The "Insert" menu options are: Cells, Rows, Columns, Sheet, Tables, Timeline, Chart, Pivot table, and Image. The "Pivot table" option is highlighted, and a "New" button is visible next to it.

Select all columns in the spreadsheet.

Select Insert > Pivot table

Select Insert to New sheet and click Create

Let's get the total \$ by city and state

ROWS Add

org_city ×

Order: Descen... Sort by: SUM of...

☐ Show totals

☐ Repeat row labels

org_state ×

Order: Ascen... Sort by: org_st...

☐ Show totals

Columns Add

Values Add

total_award ×

Summarize by: SUM Show as: Default

Cancellation_source
reinstated_indicator
reinstated_est_date
reinstatement_case
project_title
activity_code
org_name
org_type
dept_type
program_office
org_state
org_city
org_congdist
us_rep
us_rep_phone
flagged_words

Drag `org_city` and `org_state` into the Rows area.

Drag `total_award` into the Values area.

Set “Summarize by” to SUM.

Uncheck “Show totals” and “Repeat row labels” if checked.

Under `org_city`, set “Order” to Descending and “Sort by” to “SUM of total_award”.

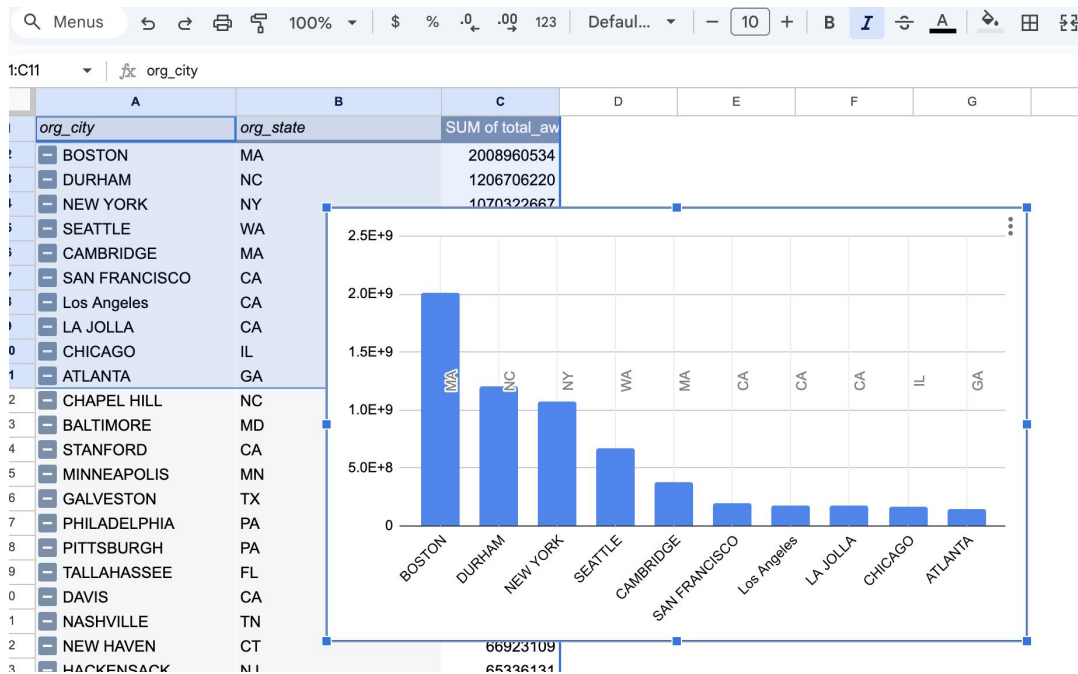
What are the most impacted cities?

1	<i>org_city</i>	<i>org_state</i>	SUM of total_aw
2	— BOSTON	MA	2008960534
3	— DURHAM	NC	1206706220
4	— NEW YORK	NY	1070322667
5	— SEATTLE	WA	673106722.3
6	— CAMBRIDGE	MA	377649851.4
7	— SAN FRANCISCO	CA	196858417.9
8	— Los Angeles	CA	176116814.8
9	— LA JOLLA	CA	173505245.2
10	— CHICAGO	IL	164050717.7
11	— ATLANTA	GA	144354675
12	— CHAPEL HILL	NC	135286849
13	— BALTIMORE	MD	127478894
14	— STANFORD	CA	118276300.7
15	— MINNEAPOLIS	MN	113511057

You should now have a table that looks like this.

What other variables should we look at?

Let's make a quick chart



Highlight the top 10 cities and select Insert > Chart.

You should now have a chart that looks like this.

Play with the settings in the “Chart editor” panel to the right.

How can we improve this chart?

What next?

Hint: we could replicate when we just did with the csv of all grants we uncompressed earlier.

This would tell us about areas that have historically benefitted from grants.

Could we then compare our analysis of historical grant data with our analysis of terminated grants?

Some additional resources

Using Google Cloud with Google Sheets
(when your dataset is very big!)

https://docs.google.com/presentation/d/1KSuvk23sUVTqH6fpe367_M3W2LGAYJrMroqLZyMdILc/edit?slide=id.p#slide=id.p

Using IF and VLOOKUP functions in Google Sheets

<https://resources.ire.org/tipsheets/20250309-30389.pdf>

Advanced pdf techniques (requires
command line)

https://github.com/chadday/nicar_ocr

<https://github.com/jsvine/nicar-2025-pdf-plumber-workshop>

Introduction to the command line

<https://github.com/ghing/nicar-2025-command-line>

Advanced resources for the adventurous

Python

Simple Python for Spreadsheets

<https://docs.google.com/document/d/1hzYXER2Cs7OY4j8NOA47RAoXQ0zATBCjEfCrDtHOPRI/edit?tab=t.0#heading=h.13xoyau58q91>

Intro to PANDAS

<https://docs.google.com/document/d/1uoFrbpwmQP4-Kddpa4VErdWDz2Pqz-NPnca9NvFm9OE/edit?tab=t.0#heading=h.93s6hba7j4u2>

R

Introduction to R and Rstudio

<https://github.com/ireapps/teaching-guide-R123>

Introduction to R walkthrough

<https://github.com/ireapps/nicar25-intro-to-R>