

Gathering and Analyzing Publicly-Available Data

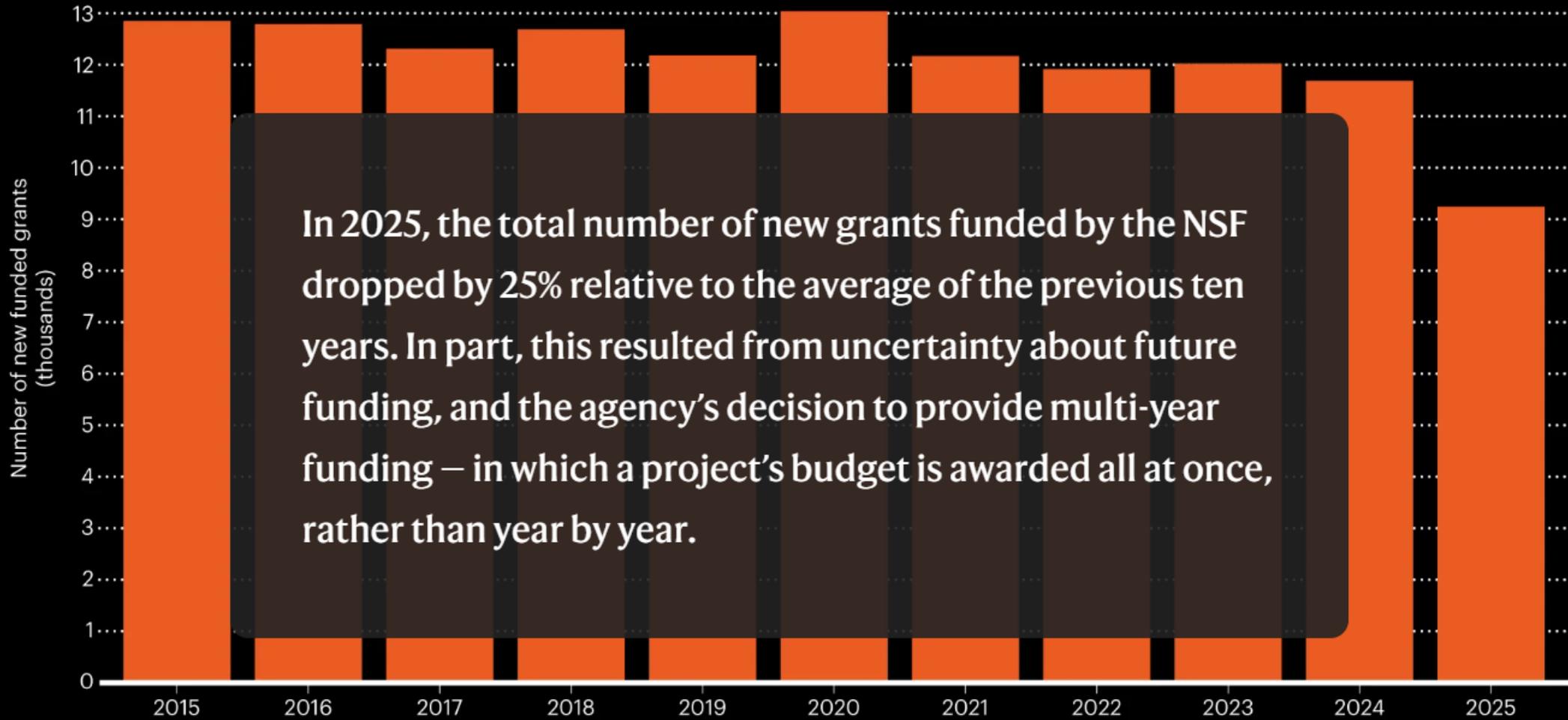
The State of Data Today

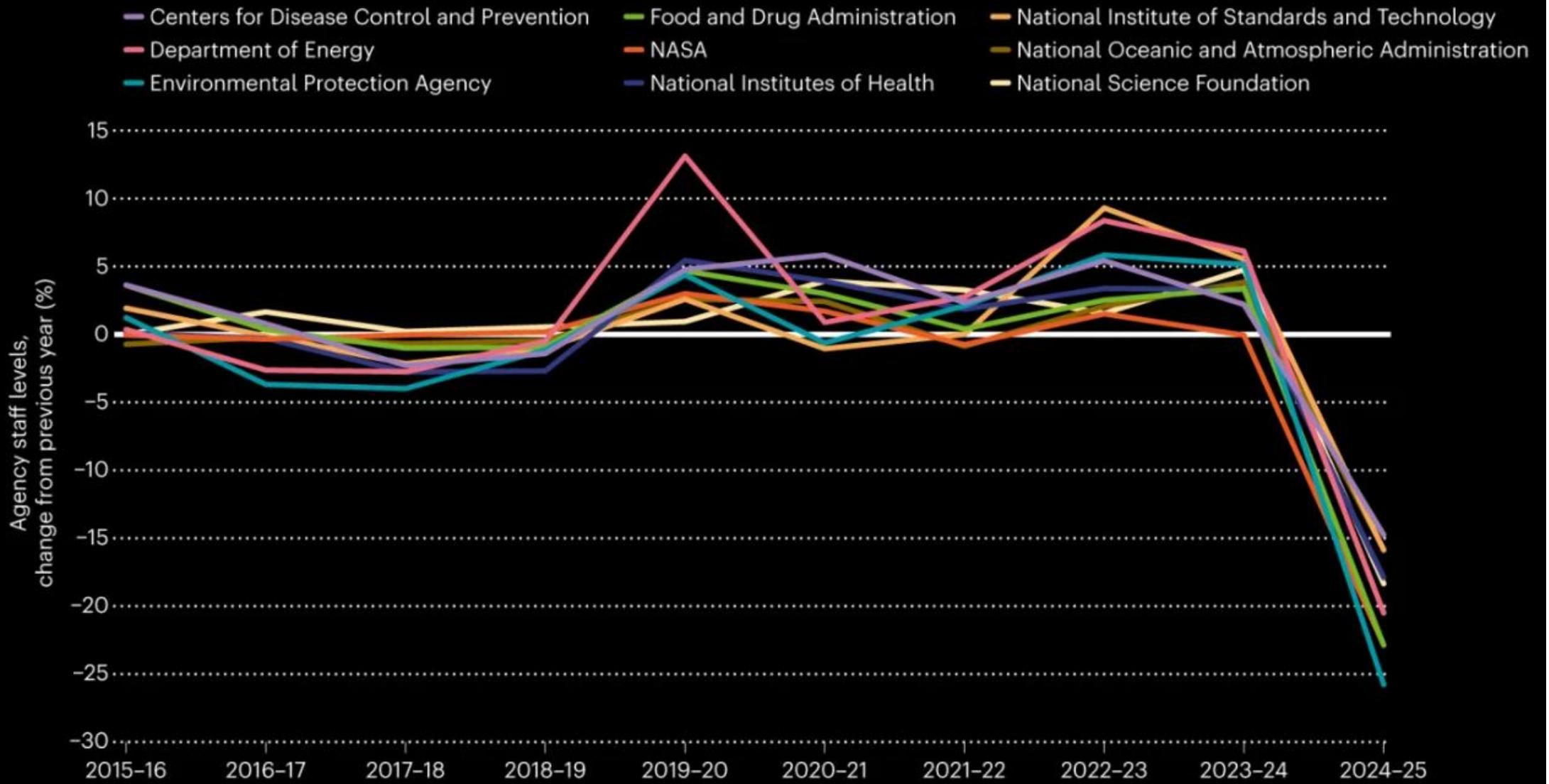
nature

US science after a year of Trump

A series of graphics reveals how the Trump administration has sought historic cuts to science and the research workforce.

National Science Foundation





Adjusted for inflation

Proposed

Non-defense research and development budget
(Constant 2017 US\$, billions)



What about the Data?

- The data (and personnel) landscape is always changing
- Different agencies feel differently about their data
- People have started to monitor or counter Trump data initiatives
- Everyone is in the same boat
- Trump's administration has a strong emphasis on AI and there are plenty of private companies that are ready to step in



[August 28, 2025 Tracking Changes in America's Data and Evaluation Infrastructure](#)

August Pulse Report - As part of ongoing monitoring of the changes to America's evidence infrastructure, the Data Foundation has identified several cr... [More...](#)

National Data Sources

- [Bureau of Labor Statistics](#)
- [FRED](#)
- [Bureau of Economic Analysis](#)
- [Data.Gov](#)
- [Census.Gov](#)
- Longitudinal Data: [NLSY](#), [SIPP](#), [PSID](#)
- [ICSPR](#)
- [CDC](#)
- [PolicyMap](#)

Federal gov survey data
ICMA (local gov)

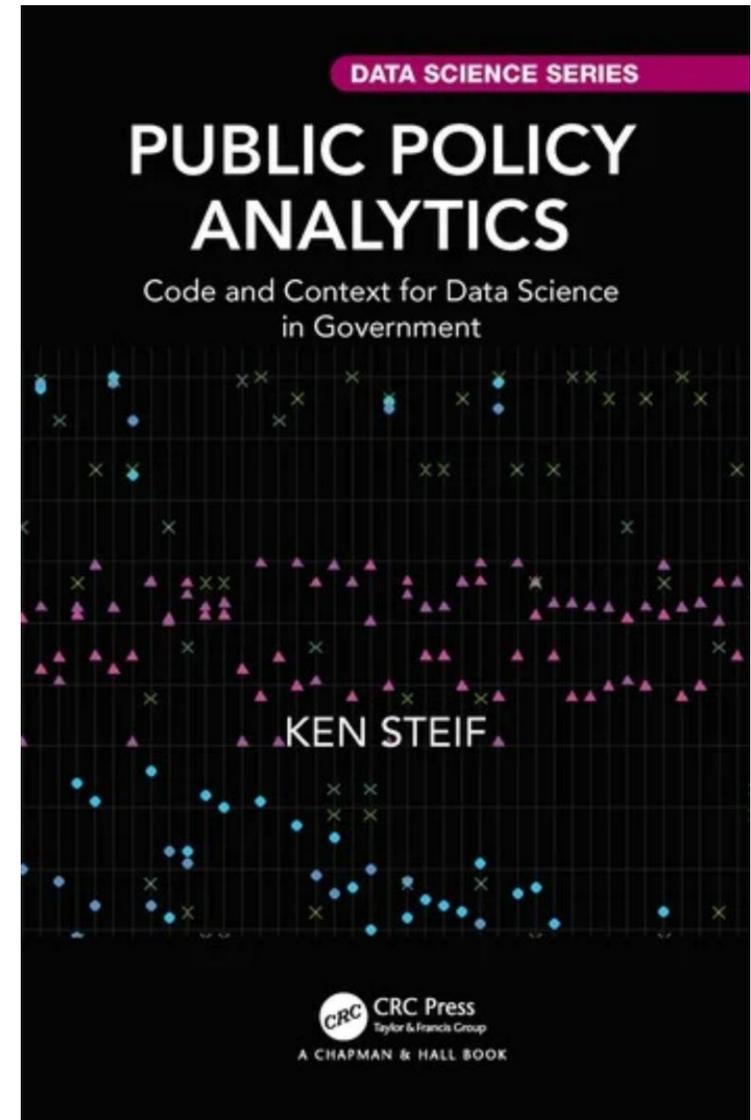
International Data Sources

- [World Value Survey](#)
 - [World Bank Data](#)
 - [UN Data](#)
 - [IMF Data](#)
 - [OECD Explorer](#)
 - [WHO](#)
- *Be careful combining multiple data sources

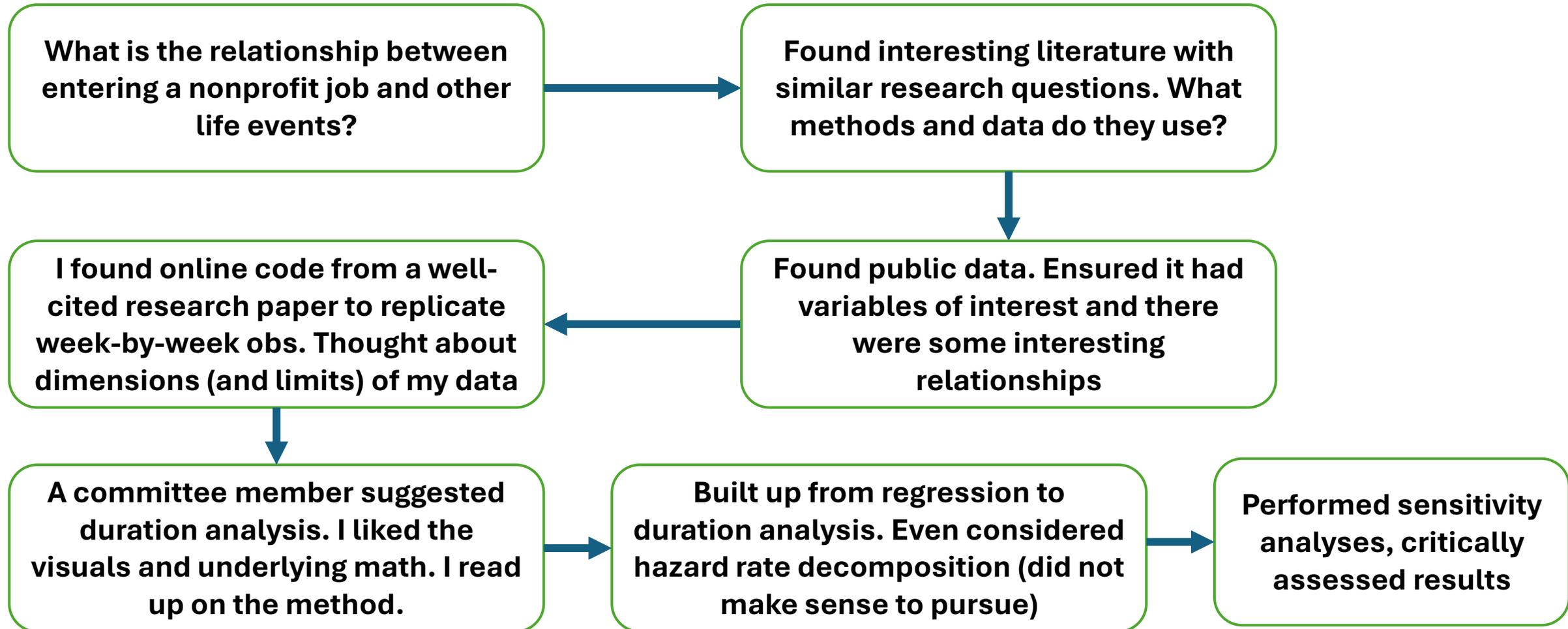
GSS
ISS

Preferred Method of Analysis

- Be good at the basics
- Let the research question drive the method
- Engage in an iterative process early on
- Think about your audience
- Build a case on why we should trust you



Preferred Method of Analysis

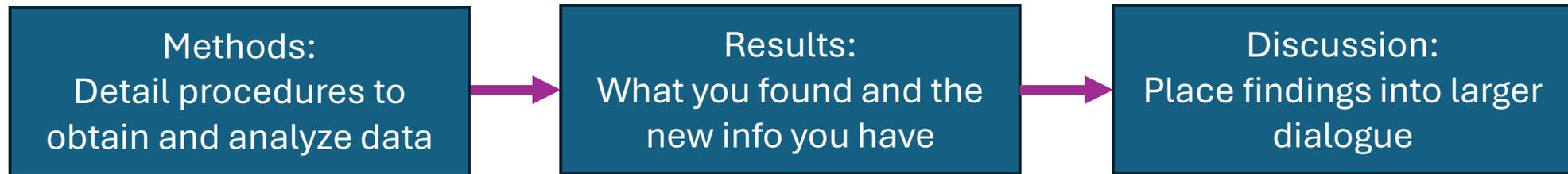


Replication Repositories

- [Harvard Dataverse](#)
- [ZBW Database](#)
- [OSF Registries](#)
- [OpenICSPR](#)

Writing a Methods Section

- Three main sections:



- Find a methods section for the journal you're aiming at, bonus if you can find an article utilizing the same approach as you
- Copy the general structure, input your content
- Don't repeat material, be as concise as possible
- Think about your audience

- Source: "Writing Your Journal Article in Twelve Weeks"

More Importantly: Presenting your Evidence

- You do not need to present all method details or all results
 - Write like a lawyer not a detective
 - E.g., Organize findings by themes (not by order of analyses you ran)
 - Consider letting a central image or theme drive your structure
 - Write in past tense
 - Do not repeat content across sections
 - Keep it short, prioritize good visuals
 - Discussion: Link (all) results to existing work and engage with the new findings
- Source: “Writing Your Journal Article in Twelve Weeks”

Good Practices for PhD Students

- Become comfortable with picking up new datasets
- Find published code for cleaning and replication
 - Don't re-invent the wheel
 - Look to other disciplines for inspiration
- For datasets for significantly high sunk costs, try to get 2-3 projects out of them
- Utilize AI ASAP! Limit your time coding but ensure you enact necessary guardrails. Original content should be yours, use it as a highly efficient TA

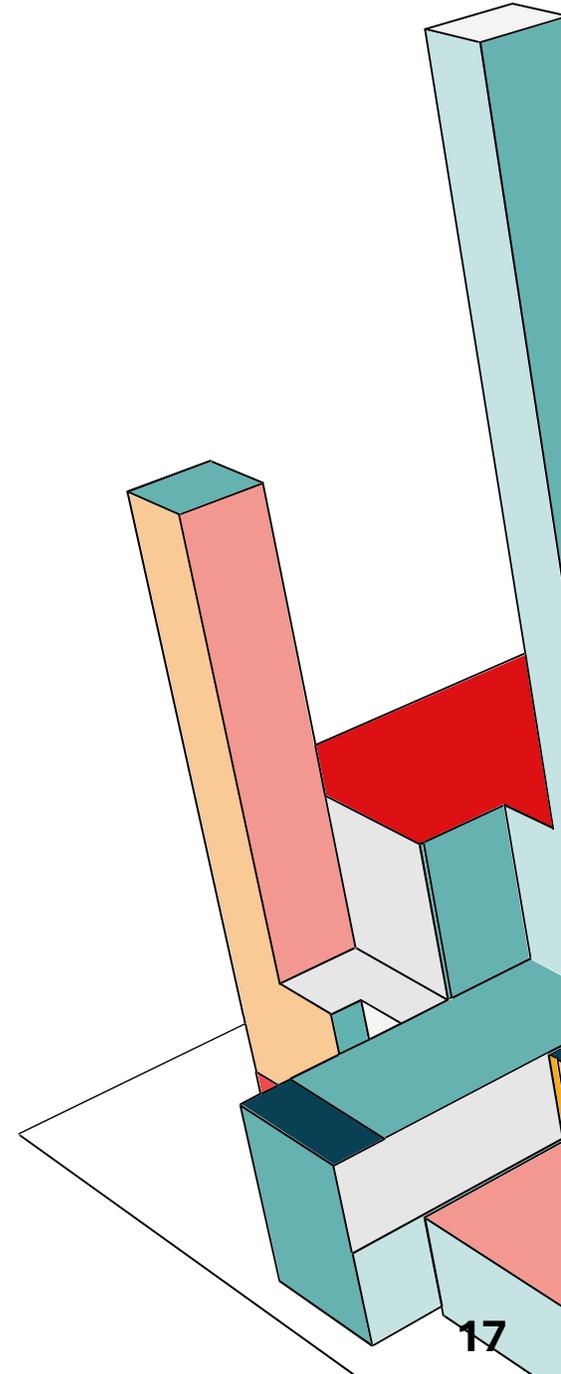
Utilizing AI Tools

- Start with original ideas and content, always validate
 - Consider it “co-Intelligence”
 - Spend at least 10 hours getting familiar with it
 - Use Opus and download the app. Consider using Claude Code
 - Check out [Scott Cunning's](#) and [Ethan Mollick's](#) work
 - Reddit post about [Claude desktop and research project](#)
-
- *Significant portions of this content is from conversations with Sebastian Jilke and Greg P.

Nonprofit Data

Going Beyond the BMF

- [IRS Main Landing Page](#)
 - More options of raw data
- [ProPublica Nonprofit Explorer](#)
 - 2012-2019 990 summary data, digital records of 990 documents, audits
- [Candid](#)
 - Data from 990, other publicly available sources, and self-reports
- [GivingTuesday Data Commons](#)
 - Dataset on 990 filings in addition to other social sector data
- [Urban Institute's National Center for Charitable Statistics](#)
 - Visuals and downloads of 990 data



990 Form Types

Form Type	Qualification	Description
990-N (postcard)	Receipts normally \leq \$50,000	Electronic notice (e-Postcard) for tax-exempt organizations not required to file other types
990-EZ	Gross receipts $<$ \$200,000, and Total assets $<$ \$500,000 (can also file 990)	Short form return of organization exempt from income tax
990	Gross receipts \geq \$200,000, or Total assets \geq \$500,000	Return of organization exempt from income tax
990-PF	Private Foundation	Return of private foundation
Schedules	Specific organization or activity type	Disclosure of specifics

Accessing the Data

- [Main Landing Page](#)
- [Tax Exempt Organization Search](#)
- [Bulk Data Downloads](#)
- [Business Master Files](#)
- [Statistics of Income Information](#)
- [Educational Resources](#)



NCCS Resources

- [NCCS Nonprofit Data Landing Page](#)
- [Urban Institute GitHub](#)
- [Nonprofit Open Data Collective](#) (with specialized R packages)
- Remember
 - Be mindful of who is **not** in your sample
 - Many researchers are using the same data
 - There is a high sunk cost with 990 data, and it has limited potential, but most nonprofit scholars expect you to be able to traverse it