

# Data on inequality

Tips on where to find data and what to ask it from three Gannett reporters

**Jayne Fraser** (she/her)

USA TODAY Data Reporter  
@jaymekfraser  
[jfraser@gannett.com](mailto:jfraser@gannett.com)

**Nikie Mayo** (she/her)

Investigative Reporter for  
The Greenville News and the USA  
TODAY Network in South Carolina  
@nikiemayo  
[nmayo@gannett.com](mailto:nmayo@gannett.com)

**Hadley Barndollar** (she/her)

USA TODAY Network  
New England Regional Reporter  
@hbarndollar  
[hbarndollar@gannett.com](mailto:hbarndollar@gannett.com)

# Before we start, three statements of principle.

---

- We **ALL** have a responsibility to consider diverse perspectives in our reporting. Without them, our stories are incomplete, inaccurate and sometimes harmful.
- You have to acknowledge your own identity to understand what you don't know so you can strategize how to account for those gaps when reporting. Put in the work.
- There is a difference between **reporting about communities** and **reporting for communities**. Journalism at its best is a public service that empowers readers, not entertainment that tokenizes peoples' experiences.

# Refocus: Inequality is not just a special project topic.

---

- **DAILY:** What kinds of information **MUST** you share promptly so readers can make decisions or take action? What can you skip? How can you do that efficiently?
- **EXPLANATORY:** What questions do readers have? How can you help them make better decisions and get info they need?
- **ACCOUNTABILITY/ENTERPRISE:** What is your plan for tracking and measuring the work of leaders? How will you understand in a nuanced, detailed way how your community is affected by their action or inaction?
- **SOLUTIONS:** Who else in the world has done a good job? What strategies could work for your community? What are the challenges or limitations?
- **INVESTIGATIVE:** Compare the way things are against the way things should be according to professional standards, community norms, etc. Who is failing whom? Who is profiting at others' expense? Who opposed available solutions and why?

# Five tips to leave this presentation with:

— — —

1. **Admit what you don't know.** Throughout the process, check your assumptions and actively work to correct blind spots. Be willing to change your focus or approach based on source feedback. The problem (or solution) might not be what you thought it was initially.
2. **Be transparent in your reporting.** This applies to your process and final product.
3. **Experts: Who is one?** What is that person's proximity to the inequality? Who do we consider "official" sources and should we trust them more? Are some people considered experts, but their data is flawed?
4. **Highlight community resiliency and solutions.** No one's totality is "victim." Also, reporting on evidence-based solutions empowers readers because they know what they can demand from leaders.
5. **National stories about inequality are also LOCAL stories about inequality.** That said, national trends might have unique local nuances that are particularly critical to understand if communities are to apply effective solutions. Ex. State legislature conversations about trans athletes and critical race theory, Miami condo collapse.

**We suggest downloading the copy of this presentation shared in chat.**

At the end, we have extra slides with tips, links, and resources that we won't have time to go over in the presentation.

There is no such thing as a data story.

When we ask people questions, we don't call it an interview story.

**All good stories are about people.**

Data is just another source and a tool for organizing information about people in a way that makes it easier to interview.

# What is data anyway?

---

Information collected and structured in a way that it can be analyzed.

Translation: Information organized so that you can interview it.



**If you have ever made a grocery list, you have built a data set!**

# Types of data

---

**Quantitative** = numerical

- How many are there?
- What is this thing's value?
- How much \$ was spent?

**Qualitative** = categorical or descriptive

- What model of vehicle is it?
- Yes or no?
- What did the person see happen?



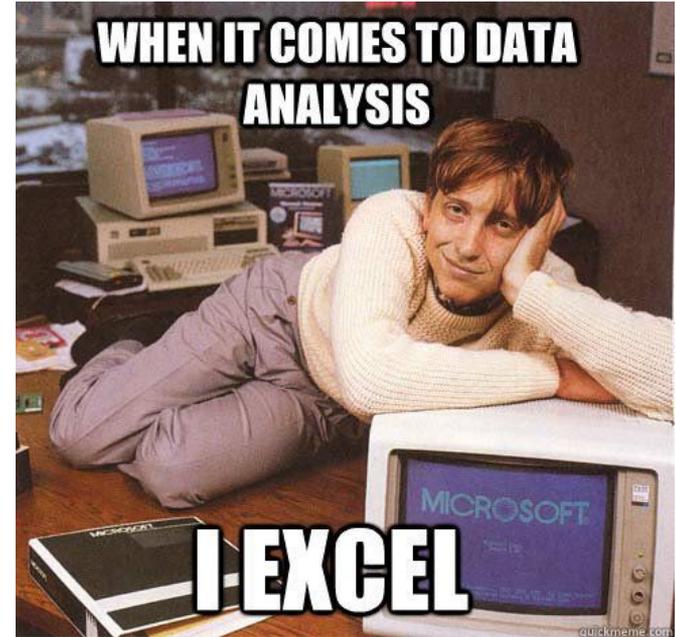
# How do you interview data?

— — —

- Count it
- Minimum, maximum, range, median, mode
- Organize or sort it
  - over time
  - by place
  - by category
- Compare it
- Test if A is cause of B, or just coincidence
- Are leaders using their own data to make smart decisions? Is this the best information they could gather to make those decisions?

**These are questions we ask human sources all the time:**

How often does this happen? Is this unusual? How do we compare? What caused it? How did you make that decision? Did you consider X, Y or Z?



# What is the role of data in reporting and writing a story?

— — —

- It's a source. We interview it like we do people. And like the people who make it, data can be flawed.
- Data informs decisions made by leaders and communities. Understanding how it is collected and how it is used is critical to knowing how systems work (or don't work).
- Data can add context or authority to anecdotal experiences that otherwise might be easy for readers to dismiss.
- Sound use of data can measure progress on problems that readers might otherwise ignore as intractable. Providing evidence of what is or isn't working empowers people to engage in issues that might seem insurmountable.
- **DATA CAN BE A NARRATIVE TOOL.** It's not all dry stats and numbers.

## Where to find data on your beat. It won't be labeled "inequality."

---

Inequality is everywhere. Consider all sectors, social issues, climate, money, government, etc. The inequality beat is ALL BEATS.

- **Demographics.** Any data set with this information will help you ask questions about differential outcomes or access.
- **Geography.** Are resources or programs equitably distributed? Who lives there or owns the land?
- **Outcomes.** How do you measure the "success" of a program?
- **Money.** Who is getting it? Who isn't?
- **Blank Forms.** Governments, hospitals, etc. all collect information. Some is converted to a data structure (or you can do that conversion yourself.)

# When requesting data, consider these dynamics.

— — —

- Official data may be shaped by the "power brokers."
- Marginalized communities may be further marginalized by the way data is collected, if they are fairly represented at all.
- It's important to see what official sources have, but ask:
  - a. Who is doing the counting and what are they counting?
  - b. If the data is about a perceived problem, who is defining the problem and what is the specific definition being used?
  - c. And while we're on definitions, ask for all of them. No matter what the data is supposed to represent, ask how that X is defined by officials. Does the official definition make sense or does it leave room for big gaps in the quality of the information collected?

# Be transparent. Address gaps and unknowns in your data.

— — —

**When the data you have has clear gaps in it, don't be afraid to say so.**

- Treat data the way you treat a human source. Interview it, scrutinize it, look for inconsistencies in the story it is telling.
- Explain the gaps in "official" data and how that affects the quality of decisions that can be made with it.
- Explain your process for filling the gaps. Use a consistent method and explain that method to your audience.
- If you're using what's referred to as "community data," make sure you understand and explain the methodology behind it. Vet the data, be transparent about where it came from and give proper attribution.

# If the data doesn't exist or the official data is flawed, build your own data set.

— — —

- **Build a “wish list” spreadsheet.**
  - What data points would you need to tell the story?
  - Are those data points quantitative or qualitative? Y/N? Sample or Census?

# If the data doesn't exist or the official data is flawed, build your own data set.

- **Use a tidy data format, at least to start.** Every variable has its own column, every observation its own row, and every value its own cell. Instead of tidy data, you have a “messy” format or a table if:
  - Column headers are values, not variable names. (“2017” instead of “year”, or “Montana” instead of “State”)
  - Multiple variables in one column. You see this a lot with Census data downloads: Non-Hispanic - white - females - 18-24yo (That’s four variables!)

Location	Age group	Data Type	2010	2011	2012	2013	2014	2015
Montana	0 to 5	Number	17,000	17,000	16,000	18,000	14,000	17,000
		Percent	24%	23%	23%	24%	19%	24%
	6 to 17	Number	26,000	26,000	27,000	29,000	27,000	26,000
		Percent	18%	18%	19%	20%	18%	17%
	less than 18	Number	44,000	43,000	44,000	47,000	41,000	43,000
		Percent	20%	20%	20%	21%	19%	19%

LocationType	Location	Age group	TimeFrame	DataFormat	Data
State	Montana	0 to 5	2019	Percent	0.15
State	Montana	6 to 17	2019	Percent	0.15
State	Montana	less than 18	2019	Percent	0.15
State	Montana	0 to 5	2019	Number	10000
State	Montana	6 to 17	2019	Number	23000
State	Montana	less than 18	2019	Number	33000
State	Montana	0 to 5	2018	Percent	0.18
State	Montana	6 to 17	2018	Percent	0.15
State	Montana	less than 18	2018	Percent	0.16

# If the data doesn't exist or the official data is flawed, build your own data set.

- **Now that you know what you want, brainstorm how you can get the data.**
  - **Government agencies** collect a ton of data. Learn which systems with funky acronyms people use in their daily jobs. They're almost always databases of some sort. And, regardless of what they try to tell you, all databases can be exported.
  - Or those agencies might not have a database of race, age, gender, etc. But do they collect that information in another format? **Ask for blank forms then categorize each line or checkbox.** If it's there, you can build your own dataset from the documents. Tedious, but tell your boss it's worth it because you will be THE ONLY ONE with the data!
  - **Build it yourself.**
    - Survey political candidates. Each question is a column!
    - Convert meeting agendas and minutes into a data set to track votes over time, attendance, etc.
    - Organize your research in a structured way. FOIA the officer-involved shooting policies for 10 area police departments. Build a spreadsheet with one row for each department and a column for each question you ask.
  - **Partner with your community to build the data.** Check out the notes form this [SRCCON presentation](#) for a walkthrough. ([Slideshow](#))
    - The school district couldn't tell The Philadelphia Inquirer which schools had lead and asbestos. So the newspaper partnered with teachers to collect samples from their school buildings then had those samples tested. They built their own data set.

# If the data doesn't exist or the official data is flawed, build your own data set.

— — —

- **Check your assumptions.** Alright, now that you have the data, take a minute to think critically about:
  - what definitions you used. Explain those to readers.
  - how you collected the information and whether that limits the scope or introduces biases.
  - what can be said and what can't. Explain limitations to readers.

**In other words, inspect and interrogate your data the same way you would a data set from a different source.**

# example

The Boston Globe's investigation into child sexual abuse was anchored by a simple spreadsheet. Reporters typed the names and locations of priests found in church directories.

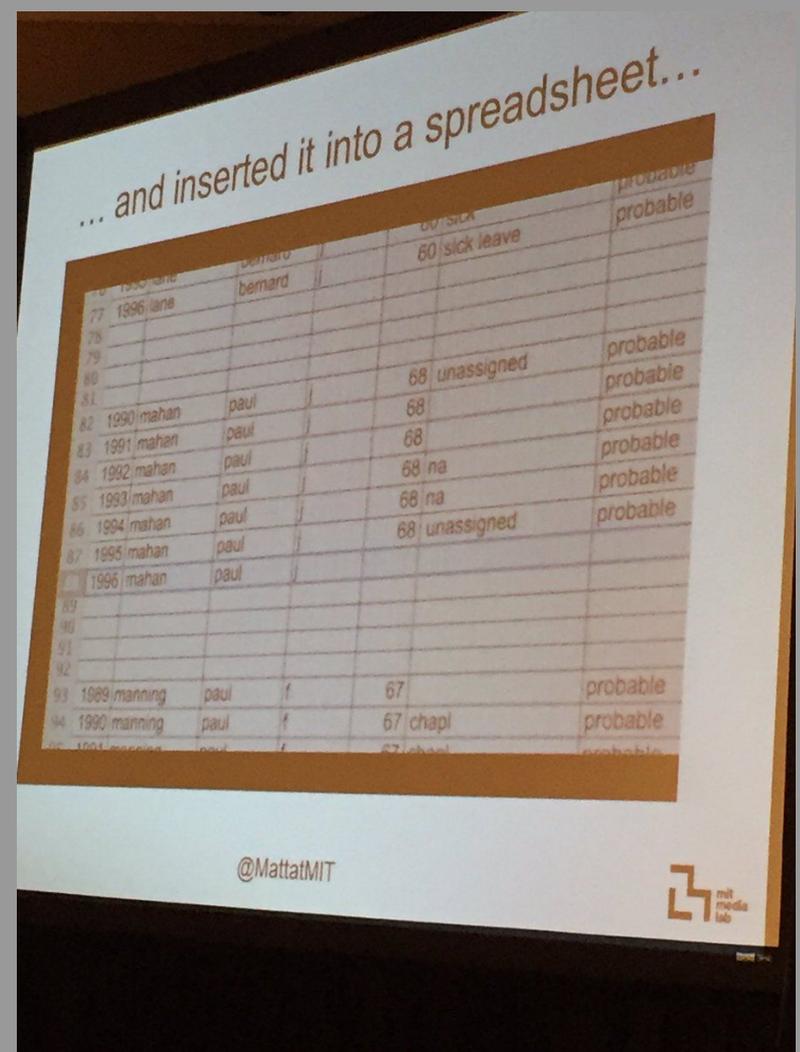
They didn't use pivot tables or sorting. They didn't do statistics.  
**They just organized information.**

That's it.

No fancy math, coding or "Big Data."

They just counted something  
no one else had counted before.

"The data was simple. It's basically one spreadsheet, 1,000 lines of text, but the impact was so profound. ... What it shows is the power of simple spreadsheets, simple databases. It doesn't have to be big or complicated." -- Matt Carroll



**Some examples from our work**

# Using overlapping data sets and geography to find striking disparities

MASSACHUSETTS

## 'By design or neglect': Flood, climate hazards threaten MA's redlined neighborhoods

Hadley Barndollar USA TODAY NETWORK

Published 5:03 a.m. ET Jun. 17, 2021 | Updated 9:10 p.m. ET Jun. 17, 2021

-Story used 1930s-era redlining maps, today's designated flood zones, and affordable housing data. Focused on MA and the city of Quincy.

Overlapping datasets and maps revealed that the country's historically-redlined neighborhoods – and a large portion of its affordable housing stock – are facing growing and disproportionate flooding threats. By 2050, the climate threats for these neighborhoods will triple.

Great technique to discover disparities in particular neighborhoods, cities, regions etc. Analyzing and overlaying health, climate, criminal justice data, etc. with geographical maps/zip codes. Ex. What parts of my coverage area have the worst health outcomes? Most juvenile arrests? Least number of Black/Latino homeowners?

# National stories on inequality are also LOCAL stories. Find the story in your backyard.

---

## 'The Maestro' of Anderson is among Black people dying disproportionately of the coronavirus

**Nikie Mayo** The Greenville News

Early in the pandemic, we knew national data showed that people of color were dying of the coronavirus at a disproportionately high rate. It was important for us to bring that home to our readers, to show them that South Carolina's Upstate was not unscathed.

We sought -- actually, fought for-- information about COVID-19 deaths from South Carolina's public health agency. Once we got the state's data, we compared it to records we got from our local coroner's office.

What we found:

The racial disparity in COVID-19 deaths was even more pronounced in Anderson County than it was statewide. Anderson County's population is 16% Black and is 80% white, according to the U.S. Census Bureau. But at the time of publication, 60% of the people who had died of COVID-19 there were Black, according to separate data from the Anderson County Coroner's Office.

# Accountability: Who we trust matters. How we count matters.

NEWS

## Leaders were slow to bring COVID-19 testing to Latino communities. Now people are sick.

Jayme Fraser, Erin Mansfield, Matt Wynn and Scott Linesburgh USA TODAY Network  
Published 6:10 a.m. ET Jul. 29, 2020 | Updated 1:15 p.m. ET Mar. 10, 2021



It was mid-June in California's Central Valley, and Dr. Patricia Iris was alarmed.

Every COVID-19 patient at Lodi Memorial that day was Latino, even though Latinos make up only 39% of the city's population.

**More:** [Leer en español](#)

Testing surveillance in San Joaquin County should have warned Iris that this was coming. But testing in Latino communities was so limited, it missed the oncoming wave.

Hospital beds swelled with Latino patients. On June 1, the county had just 23 patients hospitalized with COVID-19. By the end of the month, there were 140. Most, according to Iris, were Latino.

The numbers continue to rise. As of July 20, hospitals in the county were operating their ICUs at 132% capacity.

NEWS

## Tardan las autoridades en ofrecer pruebas de coronavirus a las comunidades latinas. Ahora hay personas enfermas.

Jayme Fraser, Erin Mansfield, Matt Wynn y Scott Linesburgh  
Published 4:46 p.m. ET Aug. 7, 2020 | Updated 1:38 p.m. ET Nov. 20, 2020



En el valle central de California, a mediados de junio, estaba alarmada la doctora Patricia Iris.

Ese día, cada uno de los pacientes de COVID-19 en el hospital Lodi Memorial era latino, a pesar de representar sólo el 39% de la población de la ciudad.

El seguimiento de las pruebas de COVID-19 en el condado de San Joaquín tendría que haberle indicado a la doctora Iris que este indicador se venía. Pero estaban tan limitadas las pruebas en la comunidad latina que el seguimiento no logró registrar la llegada de la ola.

Se llenaron las camas del hospital con pacientes latinos. El primero de junio, el condado tenía tan solo 23 pacientes internados con COVID-19. Para fines del mes, había 140. La mayoría, según la doctora Iris, era latina.

**Q+A**

**Extra resources, links and tips**

# Questions to ask your data.

## And different ways to measure inequality.

— — —

- Start with the basics: Who, what, when, where, why?
- Who is represented here, and how? Who is missing?
- What are you telling/showing me? What can this data NOT say and why?
- When was this collected? Is it still relevant?
- Where was this collected, and how? Understand the pros/cons of that method.
- What are the definitions use for this data?
- If there are major gaps in the information, why are they there?

# Questions to ask your data.

## And different ways to measure inequality.

---

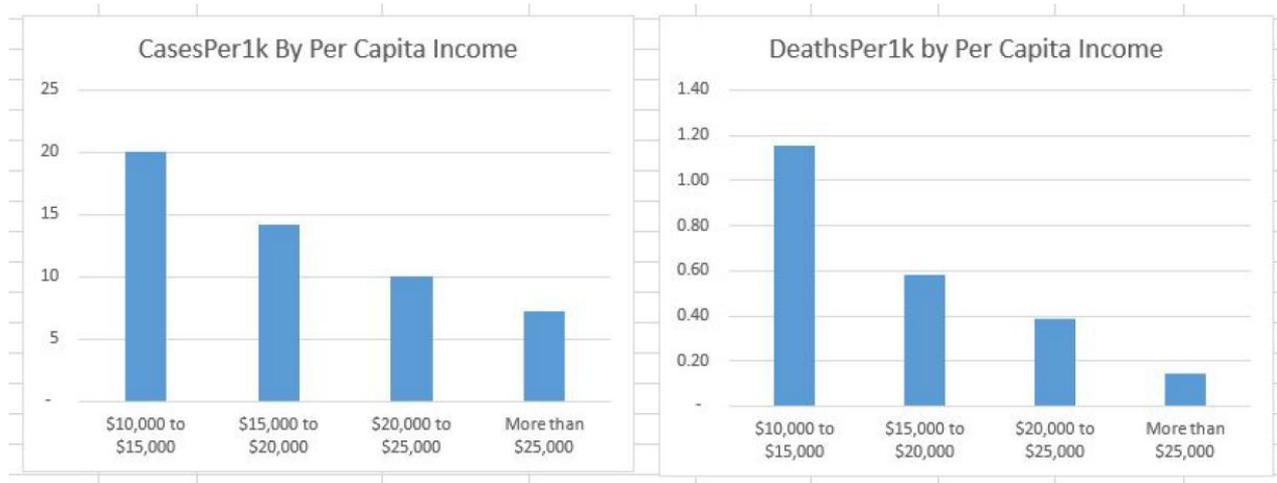
- What is your baseline or ideal? How do different groups compare to that ideal? Quantify the difference.
- Learn the power of ratios. 25% of CEOs are \_\_\_\_\_ means different things depending on how you fill the blank because the base population sizes are different.
  - Instead, try something like: 1 in every 88 white workers is a CEO. Only 1 in every 400 Black or Hispanic workers is a CEO.

# Questions to ask your data.

## And different ways to measure inequality.

— — —

- **How do you measure differences in outcomes?** What explains those differences? Can you control for “confounding variables”?
- **Is it proportional?** If Black people account for 11.2% of the U.S. workforce, do they hold 11.2% of CEO jobs?
- **Break your data into quartiles.** How does the top 25% compare to the bottom 25%? Is there a linear pattern from top to bottom?



# Questions to ask your data.

## And different ways to measure inequality.

— — —

Indexes weigh multiple data points to reach one summary number.

- **Social Vulnerability Index:** The [CDC uses 15 Census data points](#) to calculate which places are most/least resilient in the face of disasters. “A number of factors, including poverty, lack of access to transportation, and crowded housing may weaken a community’s ability to prevent human suffering and financial loss in a disaster.” ***Use this to test whether resources are being spent in the places that need the most support.***
- **Gini Index:** This is the most common statistical measure of [economic inequality](#). It compares the way resources are distributed in real life to the way resources would be distributed if it was perfectly equitable. Understand it well enough to talk with experts about it.
- **Dissimilarity Index:** This is a way to measure segregation. Specifically, it measures what percentage of Neighborhood A would have to move to Neighborhood B for two groups to be equally represented in each. [You could calculate this yourself](#), but vet it with a researcher to be sure you did it right and explain it correctly.
- **Exposure Index:** This is [another way](#) to measure segregation. It measures how the composition of the average resident's neighborhood varies according to that person's race (or other characteristic) and compares it to the city average. Again, you could calculate this yourself, but it’s best to work with experts.

# Pitfalls and Pro tips: Get the story right and earn trust.

— — —

- **Avoid the Monolith and Tropes:** The experience of all \_\_\_ people is like \_\_\_\_\_. Rural  $\neq$  white, uneducated, Republican. Poverty tropes: the victim, the criminal, the exception.
- **Do your research, but be humble.** Learn what you can in advance, but admit what you don't know. Be curious, courteous and ask lots of questions.
- **Learn the history.** Many American policies have roots in racism, ageism, sexism and/or ableism even if they aren't obvious today.
- **Be careful with the details you choose.** Avoid descriptions of songs, dance, spirituality, cultural symbols, fast food, smoking, etc. if they don't have a direct connection to the story.
- **Don't just write about the negative.** You'll find people in undercovered communities with solutions. And you should include diverse voices in your "regular" stories, not just token stories about a particular group.

# Pitfalls and Pro tips: Get the story right and earn trust.

— — —

- **Explain what you do:** Many people do not understand how journalists work. Distrust is particularly strong in undercovered communities. Be upfront about who you are, why you're there, the questions you're trying to answer, and how you'll use the information they share.
- **Give sources power.** Let them choose the place, time and method for the interview. Meet them somewhere they're comfortable. Show them your photos. Read your notes back to them and ask if you got it right.
- **Shut up and listen.** Have a conversation. Let it expand beyond your initial questions. Ask, "What else should I know?" "What do you people sometimes misunderstand about \_\_\_\_?" "What questions would you like to see my reporting answer?"
- **Ask where people get their information. Follow it, too.**

# Pitfalls and Pro tips: Get the story right and earn trust.

---

- **Ask for a tour.** My favorite interviews are walks around a neighborhood or a drive in their car, inspiring them to tell you things about their world you might never think to ask.
- **Find the helpers.** Which groups/people are filling the gaps in the official plan to make sure their community is informed, healthy, etc?
- **Partners, not sources.** Instead of extracting information from a community, make it a two-way conversation. This is easiest when you are clear about shared goals. This can be one-on-one or newsroom-with-organization.
- **End every interview with, “What questions do you have for me?”** (Note: It’s not, “Do you have any questions for me?”)
- **Use language that is accurate, respectful and precise.**

# Learn more about data equity

— — —

**We highly recommend Heather Krause, a data equity expert.**

- She taught a Knight online course last year that was fabulous.  
<https://knightcenter.utexas.edu/knight-centers-new-mooc-on-data-journalism-focuses-on-getting-your-data-right-with-equity-and-ethics-sign-up-now/>
- Insightful YouTube videos: <https://www.youtube.com/channel/UC-2kN-PC3dN7egPQkhyQ-wA>
- We All Count is a project to increase equity in data. The website is packed with resources on how to do better data analysis, build better data sets and how to explain that work to your bosses or readers.
  - Start with understanding the framework and all the places bias can be introduced into data: <https://weallcount.com/the-data-process/>
  - Then look at some of the compiled tools and resource lists: <https://weallcount.com/tools/>

# Resources for upping your data skills

— — —

- Mary Jo Webster's Data Journalism Academy is free (unless you want to donate). She is a veteran investigative data reporter and data journalism professor. (This is where Jayme got her start!) <https://sites.google.com/view/mj-basic-data-academy/home>
- Gannett's free training portal for anyone. (This presentation will be added eventually!) <https://training.usatodaynetwork.com/>
- Free peer data review from Open News to get feedback about your data analysis during development or before publication: <https://opennews.org/what/community/datareview/>
- Open News Scholarships+ are stipends to help you attend a conference, leadership course, university course or any thing else that will help you advance your career: <https://opennews.org/what/community/scholarships/>
- IRE (and it's nerd-focused NICAR) offer conferences, workshops, bootcamps, story guides and self-guided online learning portals. <https://www.ire.org/>

# Places to find data and data sets to consider

— — —

- U.S. Census Bureau -- It is especially good to form relationships with people in the regional offices. And the national PIOs are actually pretty responsive.
- Data.gov -- Most federal agencies host their public data sets on this website.
- [Data Is Plural](#) -- This weekly newsletter (and website) features all kinds of data sets, from TV show transcripts and Big Foot Sightings to the locations of bombs dropped during the Vietnam War and Medicaid drug utilization. It also includes links to stories that used that data. It's great for inspiration!
- U.S. Department of Housing and Urban Development
- State/federal agencies focused on education. The National Center for Education Statistics is the federal repository for ALL education-related data collected or used by the feds.
- The Annie E. Casey Foundation manages the [KIDS COUNT Data Center](#), which pulls data on kids and their welfare from a variety of state and federal sources. Often, it's available at the county level and for each of the last 10 years. Their spokespeople can help connect you with experts, too.
- [CountyHealthRankings.org](#) -- This program of the University of Wisconsin Population Health Institute brings together a variety of data sets on health conditions, outcomes, physical environment factors, care and access. Their online tools make it easy to compare counties and states. It also includes important contextual information for understanding each data point.

# Stories for inspiration

— — —

- A diet fueled by food stamps is making South Texans obese but leaving them hungry  
<https://www.washingtonpost.com/sf/national/2013/11/09/too-much-of-too-little/>
- Bridging the Gaps: A series reviewing of how ACA and a tribal component of the bill was supposed to improve access to health care for American Indians  
[https://missoulia.com/news/state-and-regional/special-report-bridging-the-gaps-left-by-indian-health-service/collection\\_791976bd-0d26-5f94-b222-91f7c94fac76.html](https://missoulia.com/news/state-and-regional/special-report-bridging-the-gaps-left-by-indian-health-service/collection_791976bd-0d26-5f94-b222-91f7c94fac76.html)
- Denied: How Texas keeps tens of thousands of children out of special education  
<https://www.pulitzer.org/finalists/houston-chronicle>
- How a school board turned five average schools into failure factories  
<https://projects.tampabay.com/projects/2015/investigations/pinellas-failure-factories/>
- Private and charter schools received more federal pandemic money than public schools by “double dipping”  
<https://www.goodjobsfirst.org/blog/50-state-breakdown-reveals-what-private-and-charter-schools-got-federal-pandemic-funds>
- Relief funding per student varies widely in Michigan school districts  
<https://www.bridgemi.com/talent-education/federal-covid-funding-varies-100-fold-across-michigan-schools>
- Which neighborhoods were neglected by the Paycheck Protection Program?  
<https://revealnews.org/article/which-neighborhoods-were-neglected-by-the-paycheck-protection-program/>

# Stories for inspiration

---

- Airlines are exempt from the ADA. And a separate federal law to protect the rights of people with disabilities has not been fully enforced. A series:  
<https://stories.usatodaynetwork.com/flying-while-disabled>
- Til Death Do Us Part: Why South Carolina is one of the deadliest states for women  
<https://www.pulitzer.org/winners/post-and-courier>
- The Philadelphia Inquirer partnered with teachers to collect data on lead and asbestos in their school buildings for htis investigative series:  
<https://www.inquirer.com/news/inq/toxic-city-philadelphia-inquirer-investigation-lead-asbestos-schools-20170618.html>

# Places to find data and data sets to consider

---

- For information on vulnerable populations in nursing homes: Centers for Medicare and Medicaid Services
- Court documents
- Death certificates -- ask for an "informational copy" in any state where a complete death certificate is not considered a public record.