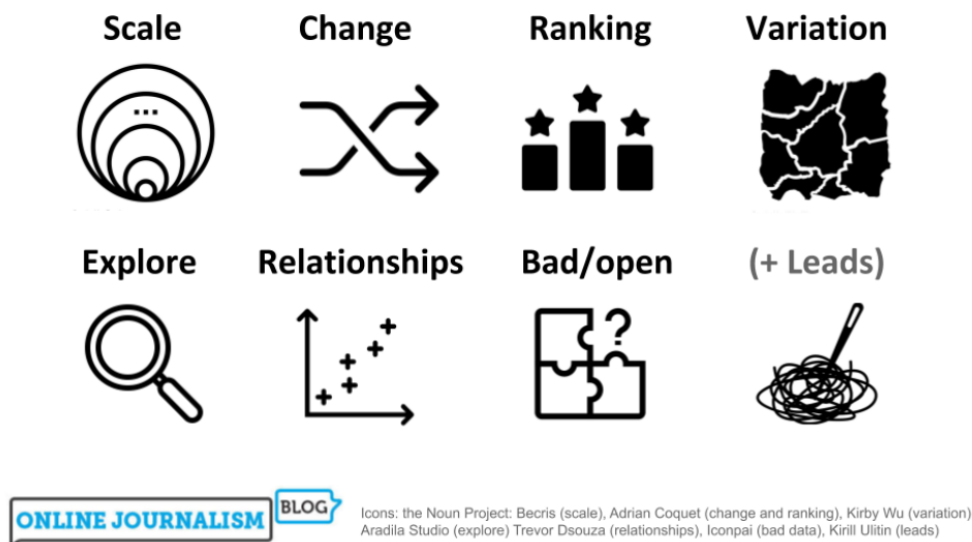


DATA JOURNALISM

From Relationships to Ranking: Angles for Your Next Data Story

By **Paul Bradshaw** | 6 hours ago

7 common angles for data stories



In my data journalism teaching and training, I often talk about common types of stories that can be found in datasets. So I thought I would take 100 pieces of data journalism and analyze them to see if it was possible to identify how often each of those story angles is used.

I found that there are actually broadly seven core data story angles. Many incorporate other angles as secondary dimensions in the storytelling (a “change” story might go on to talk about the scale of something, for example), but all the data journalism stories I looked at took one of these as its lead.

I walk through how the most common angles can help you identify story ideas, the variety of their execution, and the considerations to bear in mind.

Data Angle 1: Scale — This is How Big a Problem Is

Perhaps the most common type of story found in data is the **scale story**. These are stories that identify a big problem, or the size of an issue which has become topical.

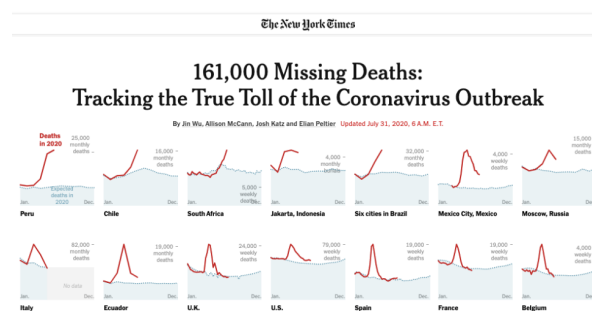
At their most simple, scale stories provide an **update** on new numbers being released: It could be the latest **unemployment figures** (<https://www.smh.com.au/politics/nsw/nsw-unemployment-hits-almost-8-per-cent-as-the-covid-recovery-starts-20200615-p552tk.html>), the **amount of crime** (<https://www.agriland.co.uk/farming-news/cost-of-rural-crime-hits-50-million-in-the-uk/>), air pollution, money spent on some area, births, deaths, or marriages.

During the first months of the pandemic, for example, we had daily scale stories on the numbers of cases, deaths, and tests, among other things.

Examples of scale stories include “[Death toll in UK care homes from coronavirus may be 6,000, study estimates](https://www.ft.com/content/9d6b46e2-55f4-4de1-ba21-f1ab9f14bcbf)” (<https://www.ft.com/content/9d6b46e2-55f4-4de1-ba21-f1ab9f14bcbf>),” but also stories like “[Unduly lenient sentences review scheme inadequate](https://www.bbc.co.uk/news/uk-47879288)” (<https://www.bbc.co.uk/news/uk-47879288>) where the lead is based on **reaction** to the scale of an issue you have identified.

Sometimes scale is provided as **background** to a single-event story, as in “[Drone causes Gatwick Airport disruption](https://www.bbc.co.uk/news/uk-40476264)” (<https://www.bbc.co.uk/news/uk-40476264>) (how many near misses are there?) or to a policy proposal, such as in “[New drivers could be banned from driving at night, ministers say](https://www.bbc.co.uk/news/uk-49026171)” (<https://www.bbc.co.uk/news/uk-49026171>)” (how many new drivers are under 19).

Scale stories are one of the easier genres to write: in many cases, no calculation is needed.



This New York Times piece outlines the scale of the death toll during the pandemic. Screenshot.

Indeed, the main work involved is likely to be in setting **context** to that scale — at its worst, a scale story simply becomes a “big number” story (“[A lot of money was spent on stuff](https://www.bbc.co.uk/news/uk-england-london-16616790)” (<https://www.bbc.co.uk/news/uk-england-london-16616790>), or “Something happens to a lot of people”), and the reader is left unclear whether this is actually newsworthy or just normal.

For that reason, it’s important to put scale into context by using **percentages or proportions** (e.g. “one in five”) or comparisons and analogies (“The money spent on the scheme is the equivalent of the wages of 500 teachers”).

You might also bring in change and/or variation as a **secondary angle**: establishing historical context to the scale you’ve just outlined, or how that scale varies.

In the [New York Times piece](https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html) (<https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html>), above, for example, the “true toll” (scale) of the coronavirus outbreak is immediately contextualized by charts which show how that has changed since the start of the year, in different parts of the country.

Data Angle 2: Change and Stasis — Things are Going Up, Things are Going Down, Things aren’t Happening

Change stories are almost as common as scale stories — and probably more straightforward to pitch.

After all, change is inherently newsworthy and gives you the verb (“rises,” “plummets,” “goes up”) that you need in a headline.

Once you’ve identified some sort of change in your data, it’s likely you will need further reporting to answer the **why?** question. Why are those numbers going up or down?

You might also add a secondary angle to your story which explores **variation** in that trend — the areas where those numbers have gone up or dropped; the most and least.

This can help you direct your reporting on “why?” because chances are that the areas affected most will be those most aware of the issue, and able to comment on it.



This story in the Belfast Telegraph illustrates the power of a “change” data angle. Screenshot.

When reporting on change, it’s important to be aware of two considerations: **seasonality** and **margins of error**.

Seasonality is the role that (typically predictable and normal, and therefore non-newsworthy) seasonal factors can play in numbers, such as the end of a financial year or school term, the release of new cars, or simply changing temperatures. Year-on-year comparisons (this August compared to last August, for example) or seasonal adjustment (<https://vizstas.com/2020/08/05/what-is-seasonal-adjustment-part-1/>) is often used to prevent this effect.

The **margin of error**, meanwhile, is the range within which the *real* numbers actually lie. Because many datasets are based on *samples* which are then generalized to the rest of the population being looked at, a margin of error (or confidence interval (<https://socratic.org/questions/what-is-the-difference-between-the-confidence-interval-and-margin-of-error#:~:text=The%20margin%20of%20error%20is,%C2%B1%20the%20margin%20of%20error.>)) is used to indicate how accurate that generalization actually is. If any change is within that margin of error, then we can’t really report that anything has changed (<https://www.theguardian.com/commentisfree/2011/aug/19/bad-science-unemployment-statistical-noise>).

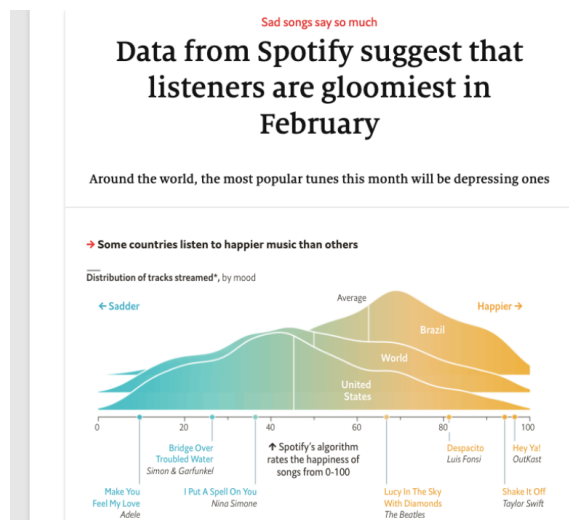
A variation of the change story is the **lack of change** angle. This story on company insolvencies (<https://www.bbc.co.uk/news/uk-53417948>), for example, looks for change where you would expect it, but identifies the absence of any increase in companies going bust during the pandemic and seeks expert comment for this counter-intuitive finding.

Data Angle 3: Ranking and Outliers — Who’s Best and Who’s Worst? Who’s Unusual and Why?

Ranking stories are all about who or what comes out **worst or best** in a dataset, or where a particular entity of interest (the local police force, schools or teams, or an industry if it’s the specialist press) sits in comparison to others.

Typical stories in this category might include “Local area one of worst areas for crime” or “Local school among best performers in the country.”

You might focus on the places “worst-hit,” as in “The parts of Birmingham in top 10 UK areas worst-hit by Universal Credit advances (<https://www.birminghammail.co.uk/news/midlands-news/part-birmingham-top-10-uk-17759533>),” or you might look at where your sector compares to others, as in “Construction is third most dangerous UK industry (<https://www.khl.com/construction-europe/construction-is-third-most-dangerous-uk-industry/141341.article>).”



This [Economist article](https://www.economist.com/graphic-detail/2020/02/08/data-from-spotify-suggest-that-listeners-are-gloomiest-in-february) (<https://www.economist.com/graphic-detail/2020/02/08/data-from-spotify-suggest-that-listeners-are-gloomiest-in-february>) is a “ranking” story because it identifies the “gloomiest” month. Screenshot

But ranking stories can also be about the best or worst **times, places, or categories** that a dataset reveals.

The [Economist article](https://www.economist.com/graphic-detail/2020/02/08/data-from-spotify-suggest-that-listeners-are-gloomiest-in-february) above (<https://www.economist.com/graphic-detail/2020/02/08/data-from-spotify-suggest-that-listeners-are-gloomiest-in-february>), for example, is about the top-ranked month for listening to gloomy songs. A Birmingham Live story, on the other hand, leads on “[The most common crimes in Sandwell – and where you’re most likely to be a victim](https://www.birminghammail.co.uk/black-country/most-common-crimes-black-country-17692682)” (<https://www.birminghammail.co.uk/black-country/most-common-crimes-black-country-17692682>).”

Data Angle 4: Variation — Postcode Lotteries, Maps, and Distributions

Variation stories work best when we expect equal treatment, or when we seek to hold a mirror up to a part of life.

The classic example uses a **choropleth map or heatmap** to show how some parts of a country have less access to something, or more demand for something, than other parts.

The phrase “postcode lottery,” for example, reflects the sense that a person’s access to something which is supposed to be equally distributed is actually a game of chance.

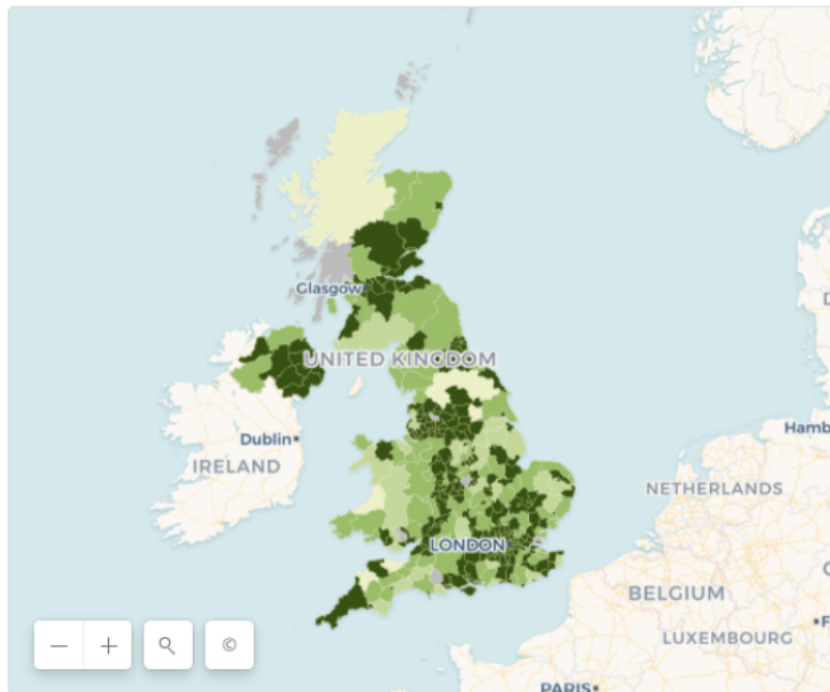
The BBC data unit story “[IVF: NHS couples face social rationing](https://www.bbc.co.uk/news/uk-england-44884503)” (<https://www.bbc.co.uk/news/uk-england-44884503>), for example, maps out how where you live in England can mean the difference between being able to access fertility treatment or not.

A variation story may be revealing that the unfairness exists — or, if people are aware of it, precisely how and where it plays out (particularly in their area).

How prepared is your area for electric cars?

Average distance between charging locations

Tap or click for more details



This BBC Shared Data Unit story by Aimee Stanton focused on the variation in access to electric car charging point. Screenshot.

Algorithmic accountability stories such as [ProPublica's Machine Bias series](https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing)

(<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>) are often about variation and the unfairness that is revealed when an algorithm is unpicked. It may be people being sentenced differently, or [given different insurance quotes](https://themarkup.org/allstates-algorithm/2020/02/25/car-insurance-suckers-list)

(<https://themarkup.org/allstates-algorithm/2020/02/25/car-insurance-suckers-list>), despite no meaningful difference between them on the dimensions that matter.

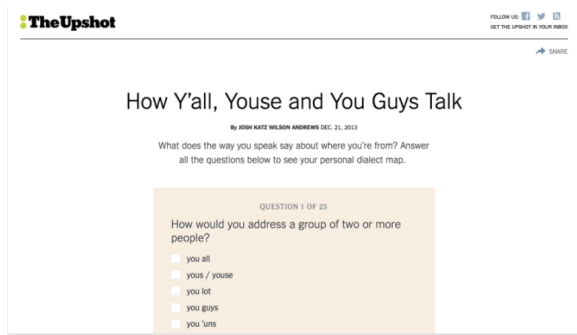
A variation story can equally be used to highlight areas of underserved demand, or lack of supply. [One story that I worked on for the BBC Shared Data Unit about electric car charging points](https://www.bbc.co.uk/news/uk-47696839) (<https://www.bbc.co.uk/news/uk-47696839>) involved identifying how much infrastructure existed in the country, and where. The picture that the data painted provided a foundation for case studies and reaction.

Data Angle 5. Explore: Tools, Interactivity — and Art

Exploratory angles are largely web-native. Its selling point is often characterized by a call to action, such as “Explore,” “Play,” or “Take the quiz.” Alternatively, it might sell the comprehensiveness of the analysis in the way that it is *mappe*”, or documents “every X that ever happened,” or simply answers the question “who/how/where”.

Readers are often invited to explore the data in an explanatory story to generate a **personal version** of it, from **quizzes** such as the BBC's “[7 billion people and you: What's your number?](https://www.bbc.co.uk/news/world-15391515)” (<https://www.bbc.co.uk/news/world-15391515>) and The New York Times's “[Take the quiz: Could you manage as a poor American?](https://www.nytimes.com/interactive/2020/01/28/upshot/administrative-burden-quiz.html)”

(<https://www.nytimes.com/interactive/2020/01/28/upshot/administrative-burden-quiz.html>) ‘ to interactive **maps** such as the LA Times's “[Every shot Kobe Bryant ever took. All 30,699 of them](https://graphics.latimes.com/kobe-every-shot-ever/)” (<https://graphics.latimes.com/kobe-every-shot-ever/>).”



This New York Times interactive (<https://www.nytimes.com/interactive/2014/upshot/dialect-quiz-map.html>) became one of their most-read stories of all time. Screenshot.

This category also takes in **simulators** like the Washington Post’s record-breaking “[Why outbreaks like coronavirus spread exponentially, and how to flatten the curve](https://www.washingtonpost.com/graphics/2020/world/corona-simulator/)” and Matt Korostoff’s “[Wealth shown to scale](https://mkorostoff.github.io/1-pixel-wealth/),” as well as **games**, **calculators**, and **chatbots**, among other formats.

Exploratory stories don’t have to be interactive: The Guardian’s “[Who is dying from coronavirus and in which NHS trusts?](https://www.theguardian.com/world/2020/mar/25/what-nhs-trust-data-tells-us-about-spread-coronavirus-uk)”’s exploratory in offering broad insights and a static map, but does not offer any particular control to the reader, and Bloomberg’s excellent narrated visualizations such as “[How Americans Die](https://www.bloomberg.com/graphics/dataview/how-americans-die/)” allow some interaction with the charts but are strongly author-driven.

They can also be quite quirky — even a form of art. “[Sweet Love: Popular Wedding Songs Reimagined As Cupcakes](https://medium.com/nightingale/popular-wedding-songs-reimagined-as-data-viz-cupcakes-2fd925920b2),” for example, is simply an exploration of what happens when playlists are treated as data and that data is visualized in a particular way.

Data Angle 6. Relationships and Debunking: When Things are Connected — Or Not

Journalists often seek to establish **relationships** by looking at data, but this can be problematic: correlation is, of course, not the same as causation, so even though two things may be going up or down at the same time it does not mean that the two are related — as The Guardian’s “[Is rise in violent crime due to cuts to neighborhood policing?](https://www.theguardian.com/uk-news/2018/apr/09/rise-in-violent-not-due-to-police-cuts-alone-figures-show)” explores.

For that reason it is perhaps just as often that you will see a story **debunking** a relationship between two data points as you will see one trying to establish that X is causing Y.

Channel 4 News’s FactCheck, for example, looked at data to answer the question “[Are migrants causing the A&E crisis?](https://www.channel4.com/news/factcheck/high-immigration-nhs-crisis)” and found no relationship between the size of non-British population in an area and A&E performance.

The complexity involved means that any relationship-angled story is likely to be quite **explanatory** — or at least *need* to be explanatory in order to communicate the caveats involved in any exercise.

How an obsession with home ownership can ruin ...



The Economist's "How an obsession with home ownership can ruin the economy," for example, spends over 12 minutes exploring the relationship between those two variables rather than pitching the story as a simple claim that home ownership is ruining the economy.

Relationship stories don't have to be about correlation; **network analysis** offers another way of telling stories which are grounded in factually established relationships such as donations, directorships, family connections, social media follows or other interactions.

The project "[Investigating Google's revolving door](https://exposingtheinvisible.org/resources/inspiration/revolving-door-google/)

(<https://exposingtheinvisible.org/resources/inspiration/revolving-door-google/>), for example, uses data to expose the number of people moving between the tech giant and government bodies, while

The Seattle Times's "[In Seattle art world, women run the show](https://www.seattletimes.com/entertainment/visual-arts/in-seattle-art-world-women-run-the-show/)"

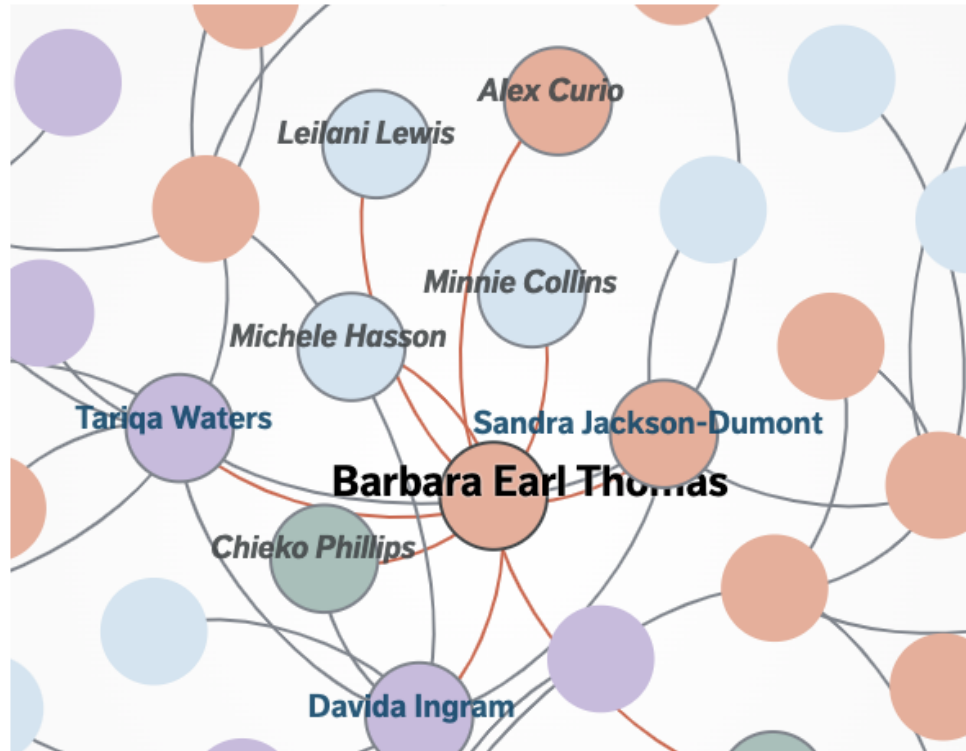
(<https://www.seattletimes.com/entertainment/visual-arts/in-seattle-art-world-women-run-the-show/>) uses a network diagram visualization to tell the story of connections within the local art scene (built by asking women "in the Seattle-area art community to name mentors, collaborators and peers who have influenced their careers.")

Making connections

Intro ● ● ● Mimi Gardner Gates ● ● Davida Ingram ● Diana Adams ● ●

Barbara Earl Thomas, an artist and the former director at the Northwest African American Museum, has been a force in Seattle's art world for 40 years. She joined the Seattle Arts Commission, now called the Office of Arts and Culture, in 1979.

[Next »](#)



This Seattle Times story walks you through the connections in a local network but also allows you to explore them. Screenshot.

But even network analysis can lack concreteness: a connection between two people or a movement between organizations is **rarely proof alone of corruption or causation** of suspect-looking decisions.

For that reason network analysis is also often presented as an exploratory story ("[Australian political donations 2016-17: who gave what to which parties](#)")

(<https://www.theguardian.com/australia-news/datablog/ng-interactive/2018/feb/02/australian-political-donations-2016-17-who-gave-what-to-which-parties#?entityID=6>), as part of an explanatory feature

("Radiohead's corporate empire: inside the band's dollars and cents")

(<https://www.theguardian.com/music/2016/apr/29/radiohead-corporate-structure-firms>), or the process is

used to identify a single data point based on its relationships, which then informs further

reporting ("[Top Tory has family link with offshore banker who gave party £800,000](#)")

(<https://www.theguardian.com/politics/2014/jul/08/andrea-leadsom-family-links-offshore-bank-donations-tories>).").

Many of the stories to come out of the **Panama Papers**, for example, come under this

latter category, and the ICIJ's [The Power Players](https://www.icij.org/investigations/panama-papers/the-power-players/) (<https://www.icij.org/investigations/panama-papers/the-power-players/>) is an exploratory angle on the various stories that the documents

helped uncover (<https://source.opennews.org/articles/people-and-tech-behind-panama-papers/>).

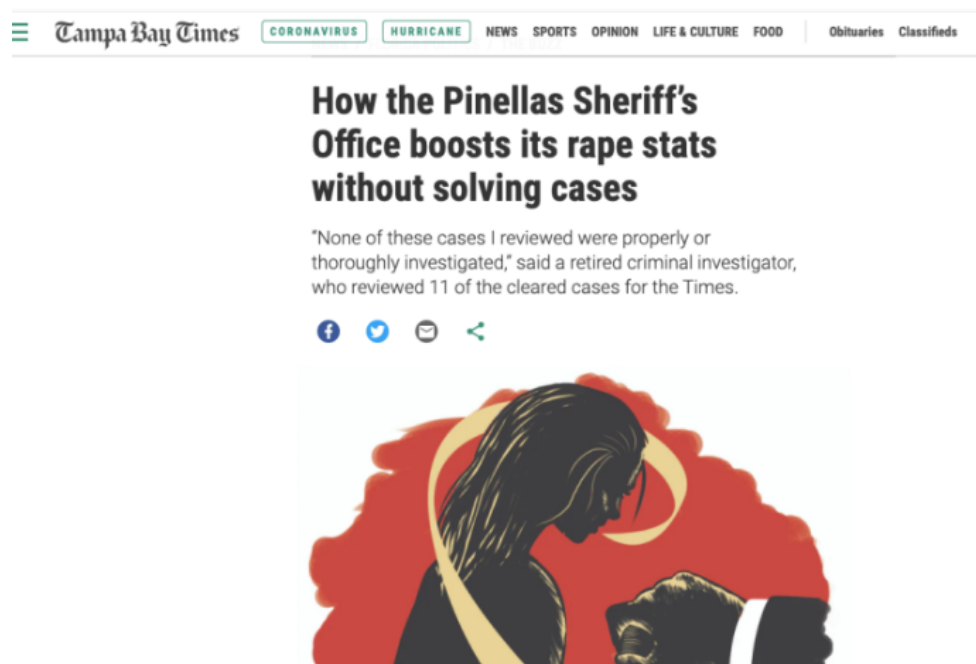
Reuters's [Connected China](http://china.fathom.info/) (<http://china.fathom.info/>) is one of the best examples of navigating the challenges of network-based storytelling, leading you carefully through the structures of power it has mapped out and allowing you to explore along the way.

Data Angle 7: Problems and Solutions: Bad Data, No Data, and Get the Data Stories

The final category of story may feel a little bit ‘meta’ — stories about data itself: **the lack of it, the problems with it, or simply its availability** — but that doesn’t mean it might not be a good story.

Bad data stories can be incredibly important: power is exercised, money spent, and lives affected on the basis of data, so if the data is flawed then the exercise of power is likely to be flawed too. [Algorithmic accountability](https://datajournalism.com/read/handbook/two/investigating-data-platforms-and-algorithms/the-algorithms-beat-angles-and-methods-for-investigation) (<https://datajournalism.com/read/handbook/two/investigating-data-platforms-and-algorithms/the-algorithms-beat-angles-and-methods-for-investigation>) stories might also, for example, lead on the flawed data underpinning those algorithms, as in Der Spiegel’s “[Increased Risk](https://web.br.de/interaktiv/erhoehtes-risiko/english/index.html) (<https://web.br.de/interaktiv/erhoehtes-risiko/english/index.html>),” the credit reference agency Schufa, it reports, “knows far less about many people than one might think – and yet dares to calculate an exact score.”

Data also often forms the basis of **political claims of success or accusations of failure**, so if a police force is [underreporting crime](https://www.latimes.com/local/cityhall/la-me-crime-stats-20151015-story.html) (<https://www.latimes.com/local/cityhall/la-me-crime-stats-20151015-story.html>), a hurricane-hit country is [inaccurately reporting deaths](https://datajournalismawards.org/projects/hurricane-marias-dead/) (<https://datajournalismawards.org/projects/hurricane-marias-dead/>), or [tests sent are being counted as tests conducted](https://onlinejournalismblog.com/2020/05/04/how-should-journalists-report-fiddling-the-figures-on-coronavirus-tests/) (<https://onlinejournalismblog.com/2020/05/04/how-should-journalists-report-fiddling-the-figures-on-coronavirus-tests/>), then there is the risk of those in power making **false claims unchallenged**.



This Tampa Bay Times story exposes flaws in official data. Screenshot.

Bad data story ideas can come from those claims, from hearing about someone who has been through the system and seen its gaps, or from simply scrutinizing existing data for problematic features: both [this Guardian story on homelessness data being “not fit for purpose.”](https://www.theguardian.com/society/2020/feb/27/questions-raised-as-england-rough-sleeping-figures-show-9-fall) (<https://www.theguardian.com/society/2020/feb/27/questions-raised-as-england-rough-sleeping-figures-show-9-fall>) and [this BBC article on concerns over gender pay data](https://www.bbc.co.uk/news/uk-england-42611365) (<https://www.bbc.co.uk/news/uk-england-42611365>) came from journalists noticing red flags in the data.

They can also lead to follow-up stories where better data is sought through alternative sources, as in [this story on homelessness figures by my BBC colleague Dan Wainwright](https://www.bbc.co.uk/news/uk-england-51398425) (<https://www.bbc.co.uk/news/uk-england-51398425>).

A related data problem angle is the **no data** story: often the lack of data on an issue represents a lack of political interest in that issue, or will to address it.

‘No data’ stories typically report on **concerns being raised** about the lack of information or transparency.

The BMJ’s investigation “[Are medical schools turning a blind eye to racism?](https://www.bmj.com/content/368/bmj.m420)” for example, leads on the line “Medical schools in the UK are ill prepared to deal with the racism and racial harassment experienced by ethnic minority students,” while a [BIJ investigation into farm emissions](https://www.thebureauinvestigates.com/stories/2019-06-13/deadly-gas-ammonia-cutting-farm-emissions-could-save-3000-lives-a-year) lists as one of its main findings that “The government only monitors ammonia emissions from the largest intensive poultry and pig farms, completely missing the biggest polluters — beef and dairy farms.”

If the missing data was *previously* published then the story may focus on **the decision to stop** publishing (The Tampa Bay Times’s “[Florida medical examiners were releasing coronavirus death data. The state made them stop](https://www.tampabay.com/news/health/2020/04/29/florida-medical-examiners-were-releasing-coronavirus-death-data-the-state-made-them-stop/)” is one example), and occasionally it may form the basis of an editorial, such as [this one by The Chicago Reporter](https://www.chicagoreporter.com/chicago-police-department-arrest-api-shutdown-is-its-own-kind-of-cover-up/) on the police force shutting down its API.

Occasionally the lack of data may compel a news organization, journalist or activist to compile their own data — at which point you have a **get the data story**.



The Washington Post’s Fatal Force project attempted to fill the gap in data around police shootings. Screenshot.

Perhaps the most famous examples of these – the Guardian’s [The Counted](https://www.theguardian.com/us-news/series/counted-us-police-killings) and Washington Post’s [Fatal Force](https://www.washingtonpost.com/graphics/investigations/police-shootings-database/) — revolve around people killed by law enforcement, and I have written previously about similar examples in “[Giving a voice to the \(literally\) voiceless: data journalism and the dead](https://onlinejournalismblog.com/2015/10/13/giving-a-voice-to-the-voiceless-data-journalism-and-the-dead/).”

Others focus on the **activists** trying to solve the problem, in what can become part-human interest story: “[Nobody Accurately Tracks Health Care Workers Lost to COVID-19. So She Stays Up At Night Cataloging the Dead](https://www.propublica.org/article/nobody-accurately-tracks-health-care-workers-lost-to-covid-19-so-she-stays-up-at-night-cataloging-the-dead)” by ProPublica and [this story about Brazilian coders building alternative Covid-19 tallies](https://restofworld.org/2020/brazil-data-transparency-covid19/) are two recent examples.

Get the data stories don't have to be so ambitious or personal, however. In the early part of the last decade many of the **Guardian Datablog's** earliest stories led with the call to "Get the data" as they opened up datasets that they had found, cleaned, or combined.

Of course, the novelty of merely making data available has faded over the last decade, and some journalists just put their data on GitHub rather than write an article about it.

Get the data: how UK migration patterns changed over the past year

There was a significant rise in immigration from the EU but little evidence of the predicted influx from Bulgarians and Romanians

DATA: get the full spreadsheet



▲ The data is based on the International Passenger Survey, which questions people entering through UK borders. Photograph: REX

Many Guardian Datablog stories invited the reader to "Get the data." Screenshot.

But if you've obtained some interesting data which isn't available elsewhere — for example through combining multiple datasets, using FOI [freedom of information], or scraping — then there is a lot of value in opening that up to your audience as **an act of relationship-building** and — in the best cases — community-building.

The point to remember here is that if you do want to build a community the data alone won't do the work for you: engaging with potential collaborators (for example through organizing hackdays) can make it more likely that they build on top of your valuable groundwork.

One More Angle: Finding Stories Through, Not In, Data

Having mapped these seven likely angles in data-driven stories, it is worth emphasizing that there is another type of story that none of them may cover: the stories where data is the **means** by which an **interview, event, document, or relationship** is uncovered or highlighted, enabling the journalist to find the story behind that.

You might call this the **single data point or needle in the haystack** story.

In a previous blog post I explored some of these techniques

(<https://onlinejournalismblog.com/2020/02/05/its-not-all-about-numbers-6-ways-that-data-can-give-you-a-story-lead/>), from interviewing individuals connected to a single data point (for example someone running a church in the UK's least religious city

(https://www.vice.com/en_uk/article/avaez5/norwich-least-religious-city-radical-christian-movements), to seeking an update from an organization responsible for tackling the numbers you are looking at.

Many investigative stories, for example, use data journalism techniques to help direct their interview and FOI requests, or choose which locations to visit, as they build a bigger picture of a systemic problem.

So while those seven angles are useful for sparking ideas and editorial brainstorming when faced with a dataset, they shouldn't be the limit of your options: any story can benefit from data journalism techniques.

This piece was first posted (<https://onlinejournalismblog.com/2020/08/11/here-are-the-7-types-of-stories-most-often-found-in-data/>) on the website of the [Online Journalism Blog](https://onlinejournalismblog.com/) (<https://onlinejournalismblog.com/>), and is reproduced here with permission.



Paul Bradshaw (<https://twitter.com/paulbradshaw>) leads both the [MA in Data Journalism](http://www.bcu.ac.uk/media/courses/data-journalism-ma-2018-19) (<http://www.bcu.ac.uk/media/courses/data-journalism-ma-2018-19>) and the [MA in Multiplatform and Mobile Journalism](http://www.bcu.ac.uk/courses/multiplatform-and-mobile-journalism-ma-2018-19) (<http://www.bcu.ac.uk/courses/multiplatform-and-mobile-journalism-ma-2018-19>) at Birmingham City University, and works as a consultant data journalist in the BBC England data unit.
