

Data journalism and audience accessibility

Writing for my mum about statistics she's reading in the paper (and other things)

Anna Khoo

Data Journalism

Office for National Statistics



My mum (busy, interested member of the public)

Time poor

Reads on mobile

Consumes data content mostly via media (and me)

Has a lot of sensible questions

Well-educated and curious

Has no idea about credible intervals etc

Technical language my mum doesn't get

- 'Confidence intervals': morale-improving fitness training stints
- 'Index': at the back of a book
- 'Breakdown by age': what your car does
- 'Decomposition analysis': watching things rot
- 'Spatiotemporal model': a toy TARDIS

How to help your reader with technical terms

Avoid: Use simpler words

Buffer: Readable without understanding that term

Contain: Put technical detail somewhere else

(sub-section, drop-down, tool-tip, methodology)

Define: Explain briefly, 'this is called/is where...'

Explain: Talk through it in detail, maybe with examples

Format: Use your headers, labels, lists, charts

Simpler alternative language

Fictional example:

Spatiotemporal modelling at LAD level continues to show a high variance in death rates after adjustments for general health state and population density.

We looked at deaths over time and by local authority district. Accounting for general health in each area and population density, there were still large differences in death rates.

Helping people through denser detail (Buffer)

'**Black and Mixed children accounted for a higher percentage of children** (9% and 7%, respectively) **entering care** after the 2011 Census in the Growing up in England (GUiE) dataset **than those who did not go into care** (4% for both ethnic groups).'

'**Price rises for food and non-alcoholic drinks are having the greatest effect on record when it comes to driving up** the overall **inflation** rate, with a 1.51 percentage point contribution in November 2022, the largest contribution since the start of the National Statistics series in 2006.'

(emphasis added on main information)

NB: Frontloaded sentence highlights paragraph focus on a skim read
Use of short brackets to help
Longer data source detail is in the middle/end for easier reading.

Example – simple definitions

Deprived backgrounds

Around 8 in 10 children (81%) who went into care for the first time after the 2011 Census were living in a household that was [deprived in at least one dimension](#), compared with 50% among those who did not go into care.

A household is considered as deprived if it meets at least one of the following criteria over four dimensions:

- Employment deprivation is where any household member, who is not a full-time student, is either unemployed or long-term sick
- Education deprivation is where no household member has a Level 2 qualification (for example, GCSE at grades A* to C or equivalent) and no person aged 16 to 18 years is a full-time student

Section header gives context (Format)

‘Around 8 in 10’ is easier than 81% (Alternative)

First sentence makes sense without understanding ‘deprivation dimensions’ (Buffer)

Link from term to methodology (Contain)

Full definition immediately after main finding (Define/Explain)

Bullet list for legibility (and less scary) (Format)

Example – caveats in conversational tone

Retail, hospitality and arts and recreation industries among worst affected by coronavirus lockdowns

Industries that experienced the most disruption to the workforce during the coronavirus (COVID-19) pandemic, such as retail, hospitality, and arts and entertainment, have seen large swings in average pay and pay growth over recent years. This makes understanding trends in wage growth more complicated.

One factor that affected average wages during the coronavirus pandemic was workforce composition. A relatively large amount of low paid workers lost their jobs, which pushes up average wages even if no workers received pay rises. [This is known as a compositional effect.](#)

Section header gives main point on skim read (Format)

‘This is complicated’. Caveats and nuances coming next, it’s all useful and interesting and worth your time.

Explanation of what a technical term means before we introduce it so that’s not terrifying. (Define)

Link from term to methodology (Contain)

Findings before methodology

Some people don't need to know how you did it to understand why it's important. That's fine.

Areas with strong local businesses and adult education tended to have lower rates of loneliness, modelling found that local authorities in London particularly benefit from these.

Regression analysis

Unpacking models

Spelling it all out is worth it sometimes.

This means that, if a person who was eligible for free school meals as a child had had the same level of education, KS4 attainment, years of labour market experience, ethnicity, and went to secondary school in the same region as a person who attended an independent school, the model predicts that on average they would still earn around 20% less than that of an independent school student. That result was true for men and women.

Engaging a public audience

- **Be topical**, listen to what people are asking/assuming
 - Regularly updated summaries ([Coronavirus latest insights](#))
 - News style articles
- **Make it human**, we care about people
 - Personalise it ([How your area has changed in 10 years: Census 2021](#))
 - How this affects them ([Personal inflation calculator](#))
 - Quotes to ground numbers in personal experiences ([Living longer lives](#))
- **Invite curiosity**, people want to immerse themselves
 - Explanatory visual walk-throughs ('Scrolly' article on [local income deprivation](#))
 - Easy to use tools ([Census maps](#))
 - Games ([Census 2021 quiz: how well do you know your area?](#))

More project examples from ONS Digital Content: https://onsvisual.github.io/project_cairo