Investigating inequalities:

3 things

to know

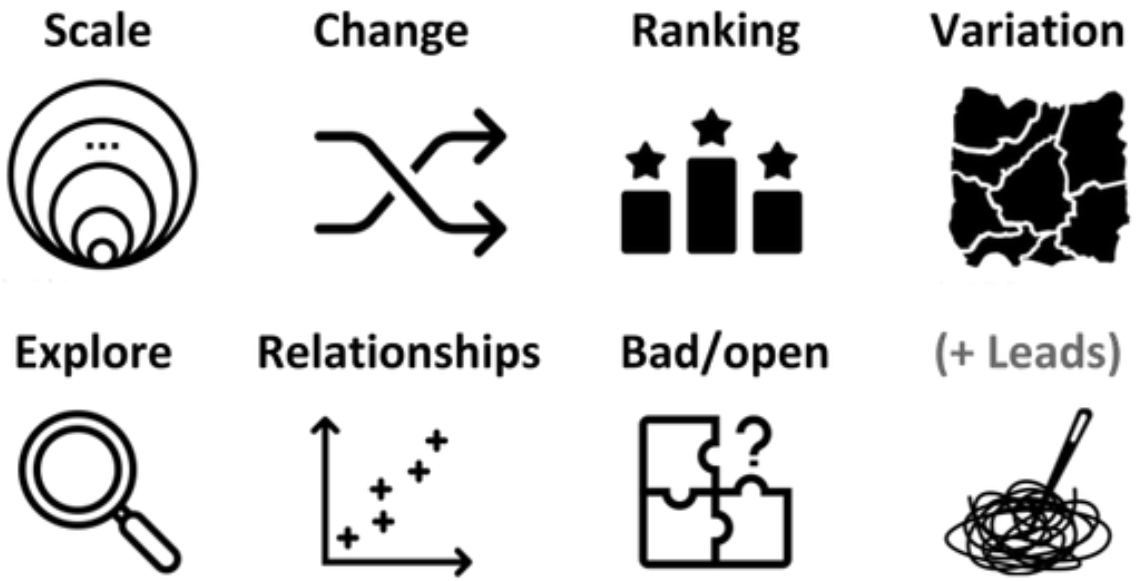
@PaulBradshaw, Birmingham City University, BBC
Author: Scraping for Journalists, Finding Stories in Spreadsheets, Data Journalism Heist, Online Journalism Handbook

What three things?

1. Consider **different angles** (don’t just look for relationships)
2. **Combine data** to identify potential inequalities
3. **Manage the risks** (watch out!)

1. Those angles.

7 common angles for data stories



ONLINE JOURNALISM BLOG

Icons: the Noun Project: Becris (scale), Adrian Coquet (change and ranking), Kirby Wu (variation), Aradila Studio (explore) Trevor Dsouza (relationships), Iconpai (bad data), Kirill Ulitin (leads)

https://onlinejournalismblog.com/2020/08/11/here-are-the-7-types-of-stories-most-often-found-in-data/

Inequality: news angles

- You might ‘reveal’ **variation** (inequality) or…
- Reveal the **scale** of inequality or…
- Which group/area **ranks** worst for a metric
- Reveal things are **getting worse** (or better) or…
- A ***failure* to improve**
- One group is **more likely** to be affected than another
- Concerns/calls around **lack of/flawed data**?

Inequality news stories

- Black men in England more likely to be diagnosed with late-stage prostate cancer, analysis shows
- Positive assessment rates for autism vary widely between parts of England
- Menopause: ‘postcode lottery’ for HRT treatment
- Black mothers in England twice as likely to have NHS birth investigated (FOI)
- Older people less likely to be given antibiotics, despite higher risk – study
- Poverty link to early death “scandalous” (correlation coefficient of 0.83)
- Ethnic minorities dying of Covid-19 in disproportionately high numbers (multiple regression analysis)
- Coronavirus: Higher death rate in poorer areas
- Menopause drugs spend varies widely across England
- Excess deaths for females 38% higher than males during July heat wave
- Government refuses to release gender abortion figures

Feature/investigation angles

- **Charticles**: *four graphs showing…*
- Raise a question and **explore** that through multiple angles (scale, ranking, change, no data) and case studies
- Look for a single data point(s) that **leads** to a case study
- Or a multi-interview feature (outliers, the ‘most typical’)

Inequality features

- The housing pandemic: four graphs showing the link between COVID-19 deaths and the housing crisis
- Are migrants causing the A&E crisis? (factcheck)
- ‘It’s like an oven’: Life in Britain’s hottest neighbourhoods
- Exclusive: Shame of all-white NHS leaders
- COVID-19’s Toll on People of Color Is Worse Than We Knew
- Financial inequality: the ethnicity gap in pay, wealth and property
- How coronavirus is worsening Britain’s racial wealth gap (Twitter thread)
- The Black American Amputation Epidemic (ProPublica scrollytell)
- Is it easier to get a job if you're Adam or Mohamed?
- Racial inequality is still a massive problem in Greater Manchester - and these are the facts that prove it

2. Practicalities.

FOI, EIR, SAR, XLOOKUP

Where’s the inequality?

- Does the data include **demographics** like ethnicity, deprivation, gender, etc?
- What demographics does it **not include?**
- Can you **FOI** that?
- Does the data include **proxies** such as the area?
- Can you **find the data** for areas to explore potential variation?

Useful datasets

- Gov.uk research & stats filtered to topic ‘health and social care’
- Prescription data
- National Child Measurement Programme
- Cancer dashboards / Rapid Cancer Registration Data
- Fingertips / Inequality Tools / Health Inequalities Dashboard
- NHS Workforce Race Equality Standard
- Indices of multiple deprivation (IMD): England, Wales, Scotland, NI
- Census (ethnicity, age, gender, employment)
- Patients registered at a GP practice by LSOA
- NHS workforce data; GP practice workforce
- GP patient survey
- Housing, pollution, lifestyle, employment, education, earnings — what else?

“**The proportion of health records containing the patient’s ethnicity code was high**, with 87% of the over 17 million inpatient spells having a valid ethnic group recorded in 2019/20, a slightly higher proportion than for outpatient attendances (83% of over 96 million) and A&E attendances (86% of over 19 million). In addition, 8.5% of inpatient records had a code of ‘not stated’, which, although a permitted code, is not useful for analysis purposes. However, 8.8% of inpatient spells had an ‘other’ ethnic group coded. These proportions have increased since 2010/11, from 6.1% (not stated) and 7.2% (‘other’ ethnic groups)”

Ethnicity Coding in English Health Service Datasets

FOI tips

- Not all data is published — ask what isn't
- Ask for **data dictionaries** and **asset registers**
- Anticipate **exemptions**:
  - s40 personal information (numbers <5 policy)
  - s21 information already published
- **Documents**: investigations, complaints, correspondence
- Ask for **reports** where analysis was conducted internally
- Ask for **impact assessments**

Beyond FOI

- **Advanced search**, e.g. for impact assessments

```
"impact assessment" -guidance site:rdash.nhs.uk filetype:pdf
```

- **Environmental Impact Regulations**: fewer exemptions than FOI if it relates to environment (e.g. building)
- Sources can make a **Subject Access Request** for their data
- Teach **genAI** principles and use to help with drafts and feedback
- Use **NotebookLM** or Google Pinpoint to search multiple documents

Combining data to investigate inequality

- Both datasets need to have a **field in common**, e.g. LSOA code, authority name, surgery address
+ *You may need to **convert** data, e.g. postcode to LSOA or authority*
- Use **XLOOKUP or VLOOKUP** in one dataset to look for a match in a second dataset - and fetch associated data…

Example: a simple lookup

- From the NHS Workforce Race Equality Standard page, I've downloaded *NHS Workforce Race Equality Standard (WRES) raw data*
- The *Staff survey data* sheet contains measures of harassment and discrimination
- But the measures are indicated by a code
- The *Contents* sheet contains each code and its description (e.g. "Percentage of staff experiencing harassment, bullying or abuse from staff")
- We need to fetch the descriptions that match each code
- A formula to do this would look like this:
  `=XLOOKUP(G2,Contents!B:B,Contents!D:D)`

The XLOOKUP function

```
=XLOOKUP(G2,Contents!B:B,Contents!D:D)
```

- The formula starts with an equals sign
- **XLOOKUP** is a function (recipe) to look up data
- All functions are followed by brackets containing any ingredients. This function has 3 ingredients:
  - What is being ‘looked up’ (**G2**)
  - Where is it being looked for? (**Contents!B:B**)
  - Where is the data you want to fetch? (**Contents!D:D**)
- **B:B** or **D:D** means look down the whole of column B or D
- **Contents!** means look in the sheet called ‘Contents’, not this one

Try it: using XLOOKUP to fetch data

We want to see if vaccination rates vary based on an area's deprivation…

- **Make a copy** of the pre-prepared dataset of vaccination and deprivation data
- Skip to **step 3** of the walkthrough
- In **Table5** create a new column for the lookup data, and write a lookup formula in the column's first row of data that looks for the code in the IMD sheet and brings back the value in column G, e.g:
  `=XLOOKUP(A7,IMD!A:A,IMD!G:G)`
- **Copy the formula down** the column for all area codes
- Bonus: start from step 1, or try a lookup with ONS ethnicity data as well

Other things you can do

- Calculate a correlation coefficient
- Create a scatterplot (deprivation, % black)
- Calculate a ratio (e.g. average in least deprived vs most deprived)
- Multiple regression analysis

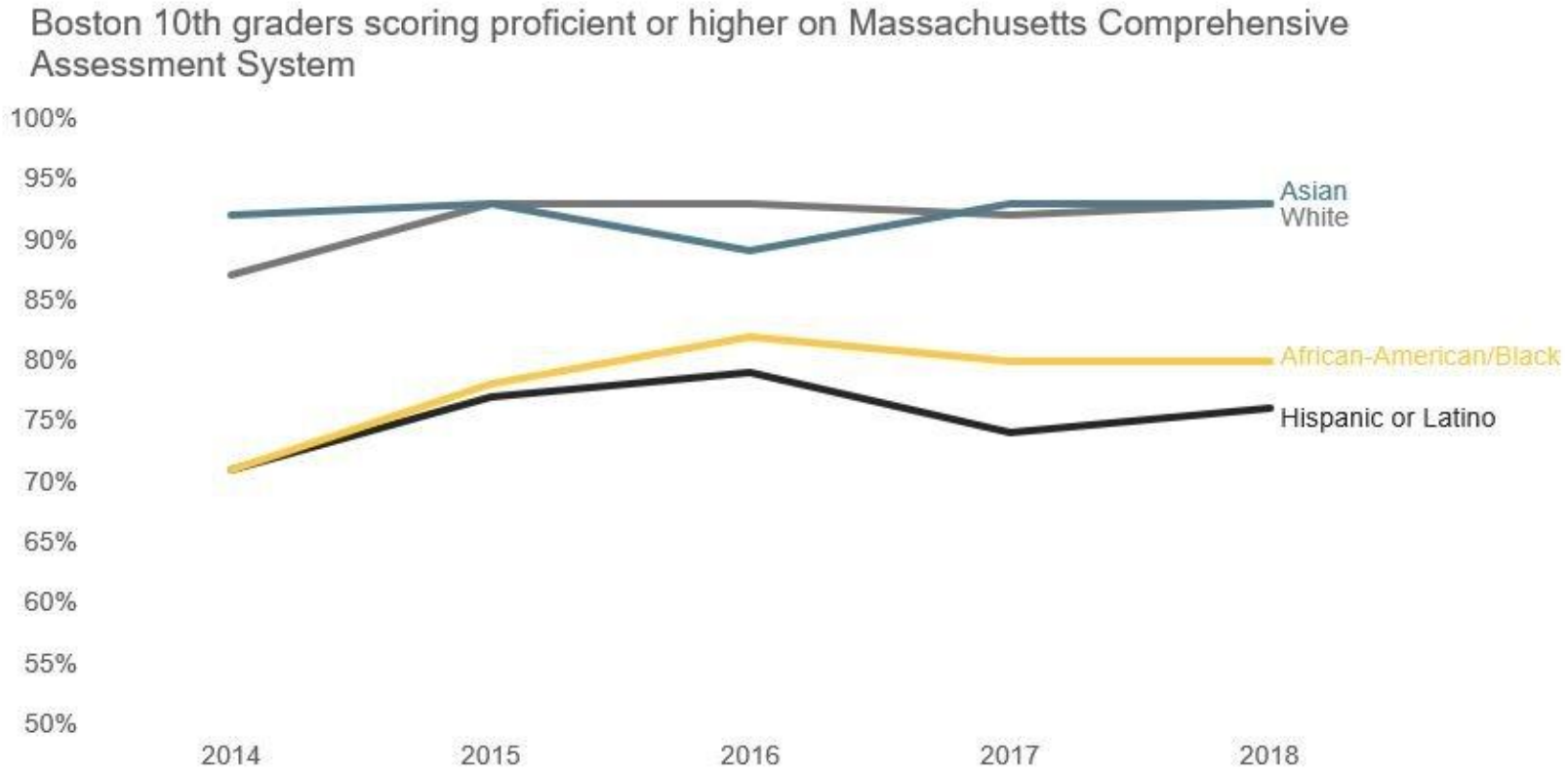
3. Manage risk.

Considerations and risks

- Data tells you what is happening, but not **why** - ask an expert
- Correlation vs **causation**: other factors, e.g. housing, social class, age profile
- Watch out for p-hacking (aka data dredging)
- Smaller groups are underrepresented in the data = more **variation**
- Communicating the quality of ethnicity data
- Tendency to **group non-white groups** into single ‘BME’
- Communicating **uncertainty** (e.g. margins of error)
- Deficit-based narratives
- Data doesn’t show the **lived experience** - case studies needed

“[Deficit narratives] reduce a group or culture to its “problems,” rather than portraying it with the strengths, creativity, and agency that people from those cultures possess. For example Maggie Walter and Chris Anderson describe how statistics used by settler colonial groups to describe Indigenous populations have mainly functioned as “documentation of difference, deficit, and dysfunction.””

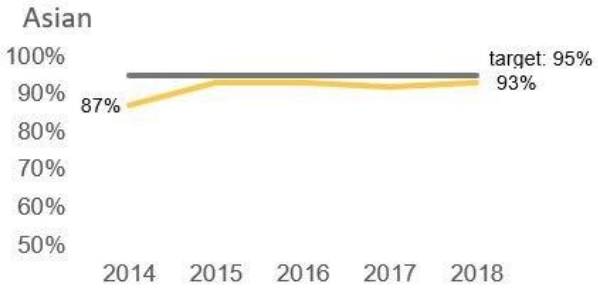
D'Ignazio and Klein (2010) Data Feminism

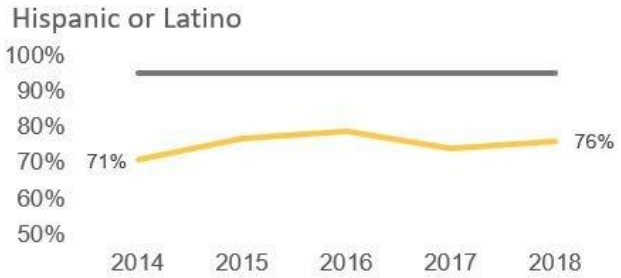


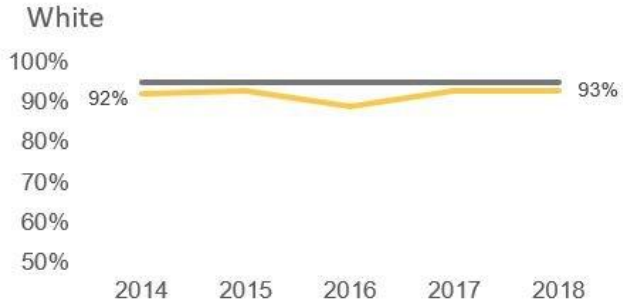
http://blog.pietablakely.com/presenting-data-for-a-targeted-universalist-approach

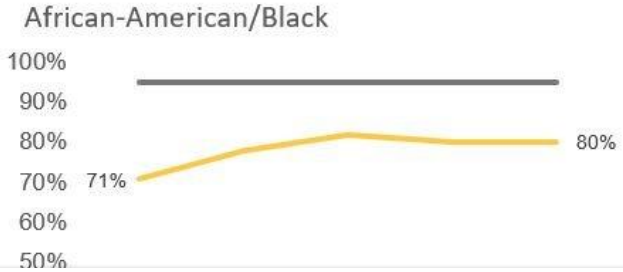
“As one step to counteract [women being reductively portrayed as victims of violent crimes like murder, rape, or intimate partner violence], Blecker chose to publish an example from Uruguay that didn’t focus on violence, but rather on quantifying women’s unseen contributions to the economy.”

Boston 10th graders scoring proficient or higher on Massachusetts Comprehensive Assessment System









Small multiples

http://blog.pietablakely.com/presenting-data-for-a-targeted-universalist-approach

Where next?

Key points

1. If data doesn’t have a breakdown by demographic, **consider matching** to demographic data
2. If you learn one skill, **learn XLOOKUP** to join datasets
3. Use data and interviews to complement the other’s blind spots and **manage risks**