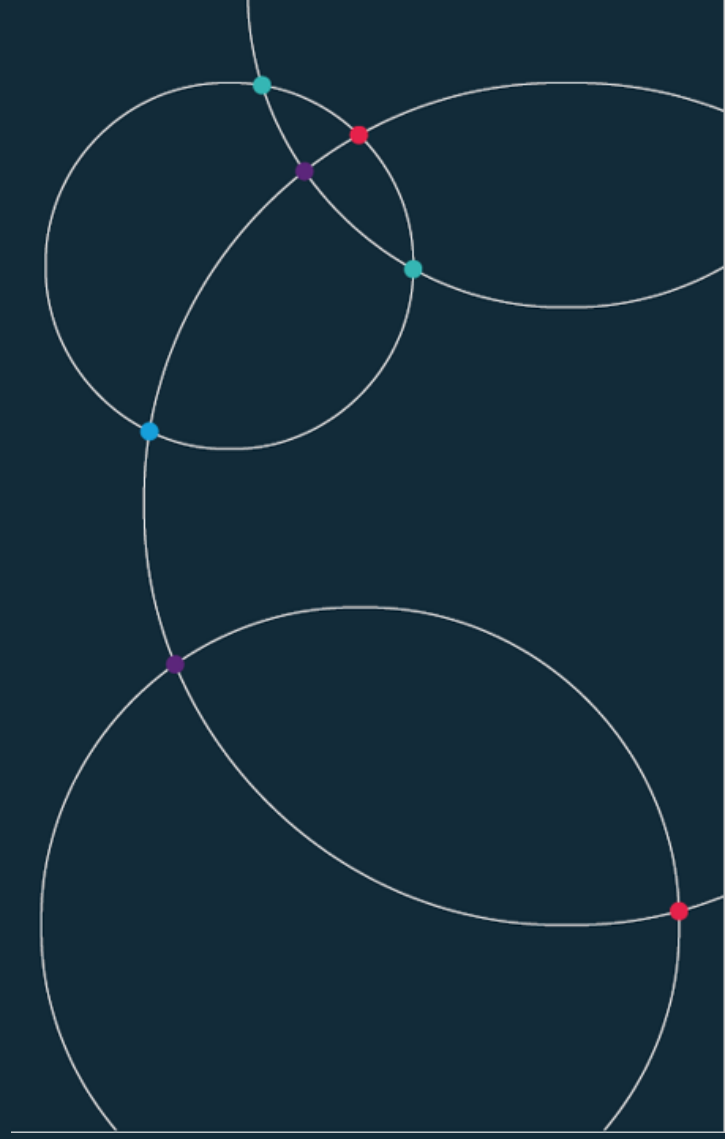
Data Science.

For economists

Richard Davies

Autumn 2022



Week 2.

- **Charts.** The good the bad and the ugly.
- **What makes a great chart?** And what to avoid.
- **Two ways to make a live chart.** APIs vs GitHub data.
- **Practical 2.** Taking a bad chart and making it good.

Reminder I: A vital diary date.

The festival of economics

- Part of your portfolio will be to comment on / analyse and argument made at the Festival of Economics.
- The dates for this are 15th -16th -17th November.
- The events are here: https://www.economicsobservatory.com/events
- You will need to attend at least one session. The cost is £5.*
- The event is an opportunity to meet professional economists and potential employers, and we look forward to seeing you there.

** If this is problem for anyone, please let us know, as the Observatory will have some sponsored tickets we may be able to provide free of charge.*

Reminder II: chart ideas and code.

Sign up to the Observatory Newsletter and social feeds

Data stories. Last week we wrote about the sterling crash, and made 10 charts.

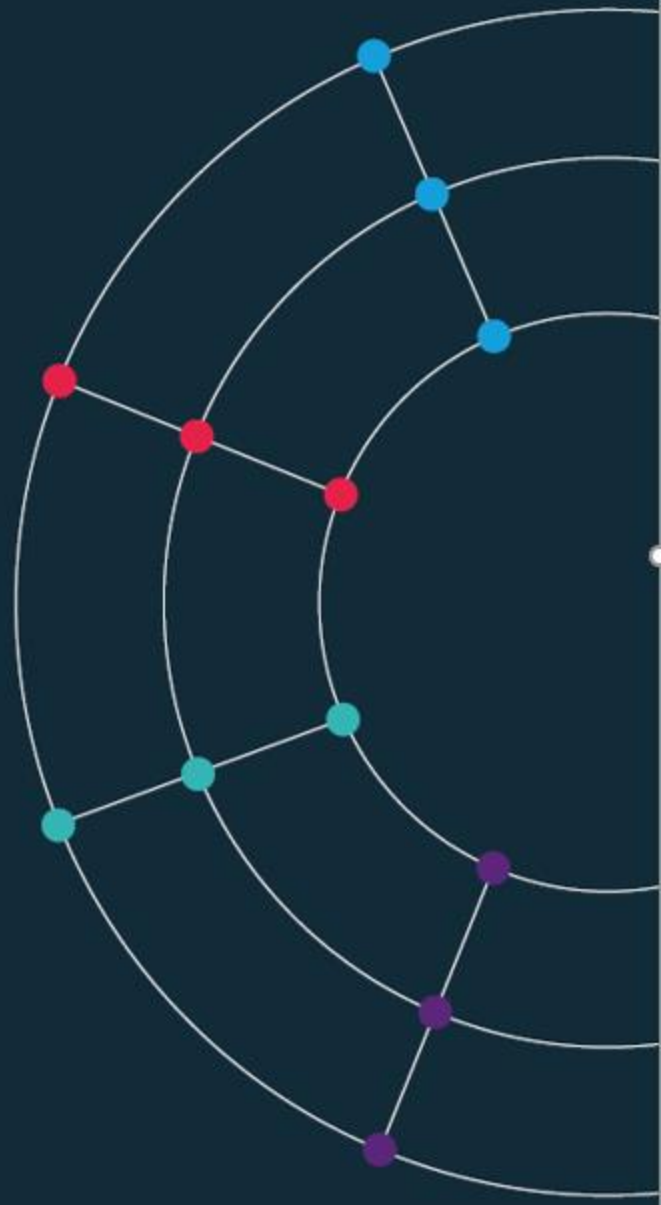
https://www.economicsobservatory.com/whats-happening-in-uk-markets-the-story-in-ten-charts

Economics Observatory. Sign up for our newsletter. We will be emphasising charts and data over coming months.

https://www.economicsobservatory.com/join-us

Social Accounts. Follow our social accounts. On IG: @RapidCharts, @EconomicsObservatory. On Twitter: @EconObservatory, @RD_Economist. We will post chart ideas, along with code, during the duration of the course. These will help with both your portfolio and project.

Bad, Ugly … and Good.

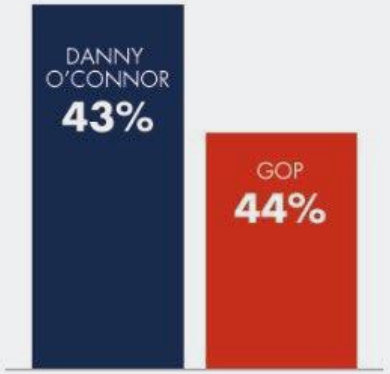


When charts go wrong.

↓ Nate Silver's TOP RATED pollster just BLEW. OUR. MINDS ↓

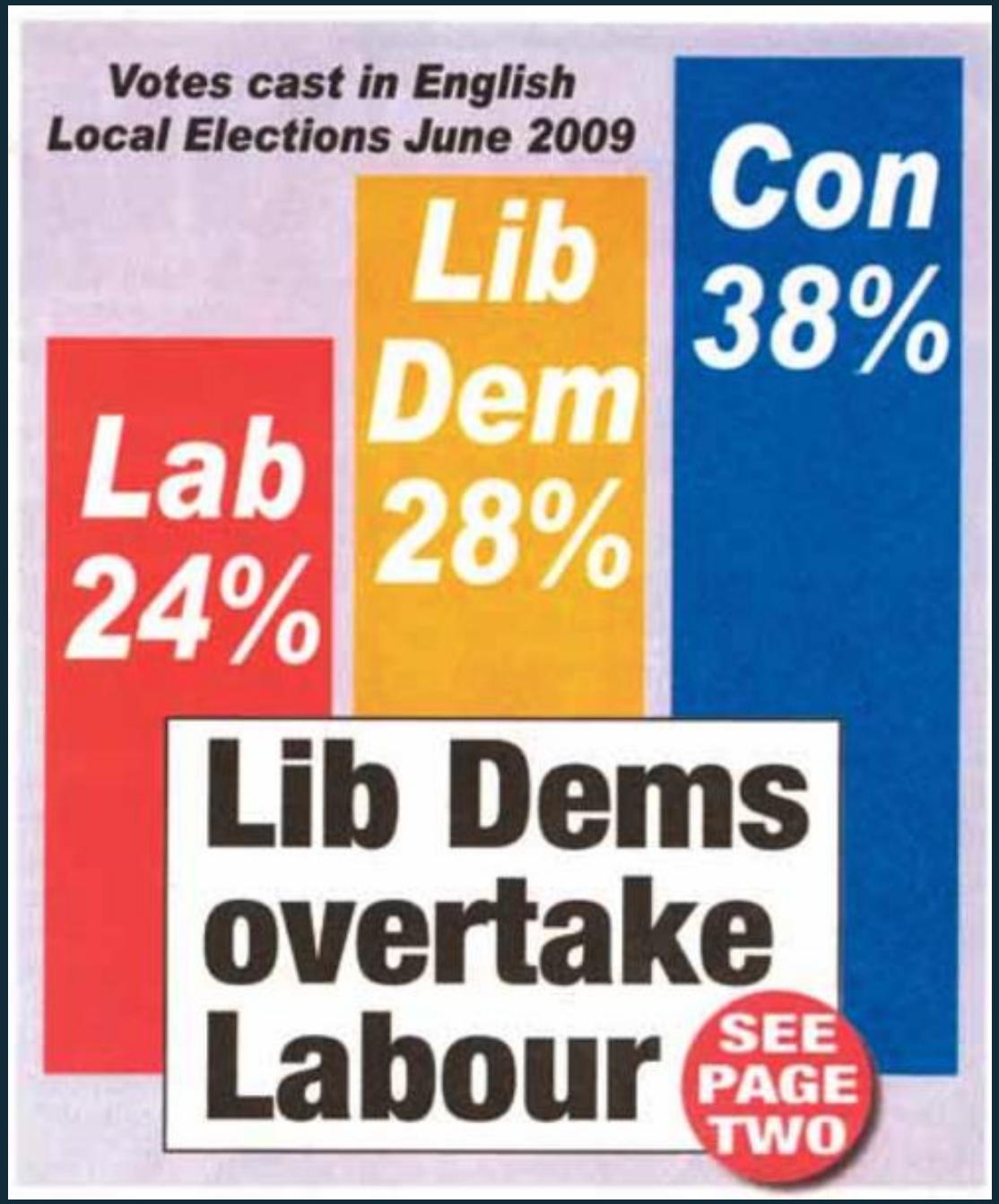
DANNY O'CONNOR ONLY DOWN BY 1%
in Ohio Special Election

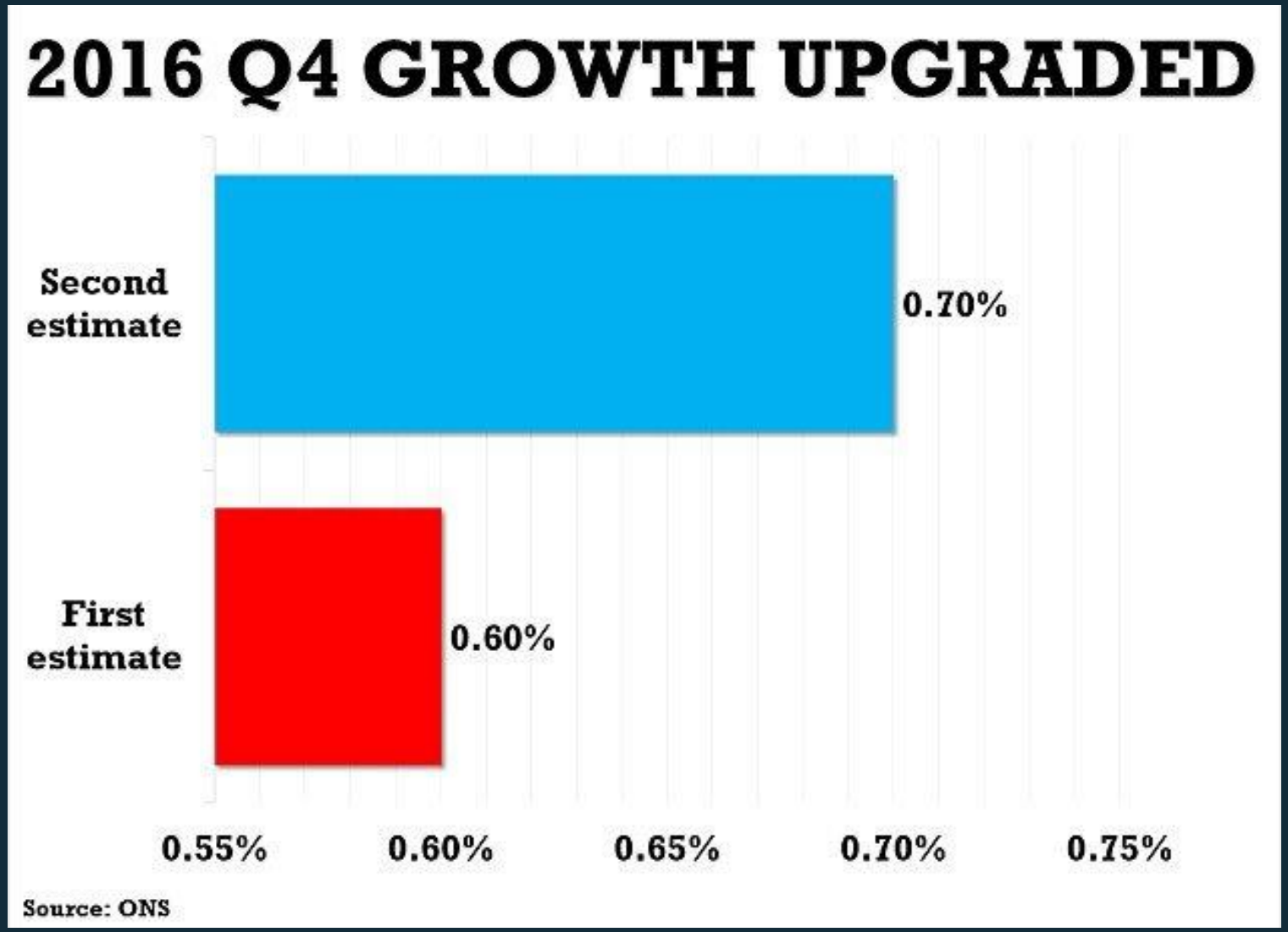
SOURCE:
MONMOUTH POLL



IT'S HAPPENING:

DEMOCRATS ARE ON THE CUSP OF WINNING THE OHIO SPECIAL ELECTION!

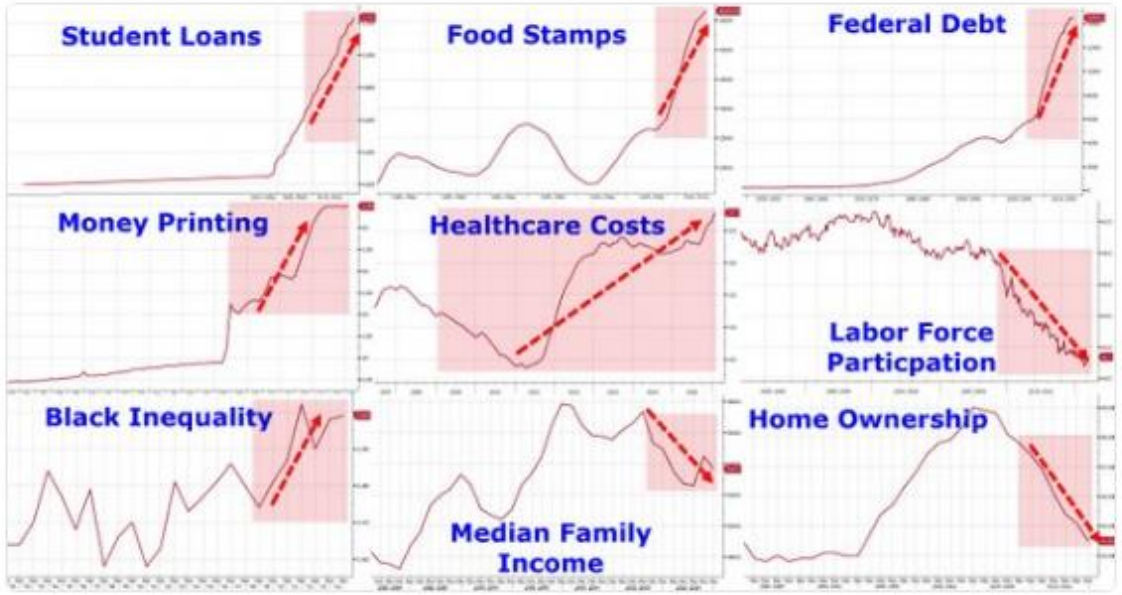

Votes cast in English
Local Elections June 2009
Con
38%
Lib
Dem
28%
Lab
24%
Lib Dems
overtake
Labour
SEE
PAGE
TWO


2016 Q4 GROWTH UPGRADED
Second estimate
0.70%
First estimate
0.60%
0.55%
0.60%
0.65%
0.70%
0.75%
Source: ONS

Donald J. Trump
@realDonaldTrump

Following

"@TaylorEdwards99: THIS IS @POTUS'S LEGACY! AN ABSOLUTE DISASTER!!! WE NEED @realDonaldTrump NOW!! #MAGA #TRUMP2016



RETWEETS 6,911 LIKES 13,785

11:20 PM - 2 Jun 2016

6.9K 14K

International Monetary Fund + Add to myFT

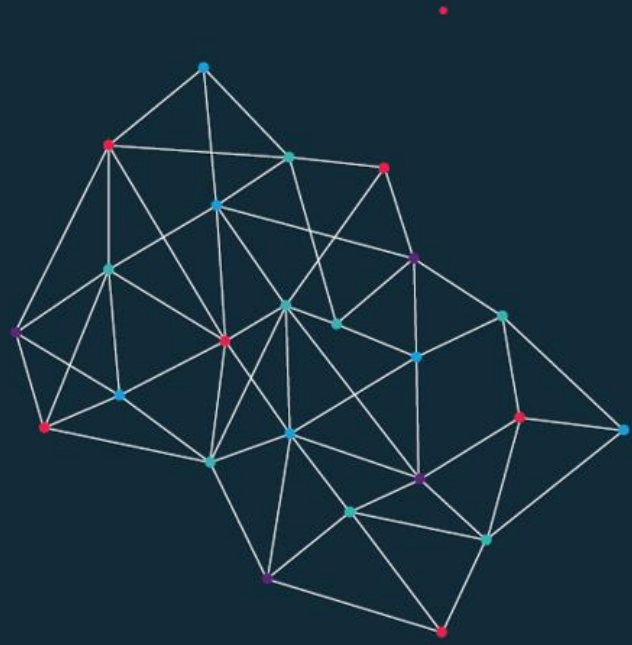
IMF bailouts hit record high as global economic outlook worsens

Fund hands out $140bn in loans as interest rate rises push up countries' borrowing costs

IMF managing director Kristalina Georgieva meets Argentina's economy minister Sergio Massa. Argentina is set to receive $3.9bn from the IMF in the next few weeks © Argentina's Economy Ministry/AFP via Getty Images

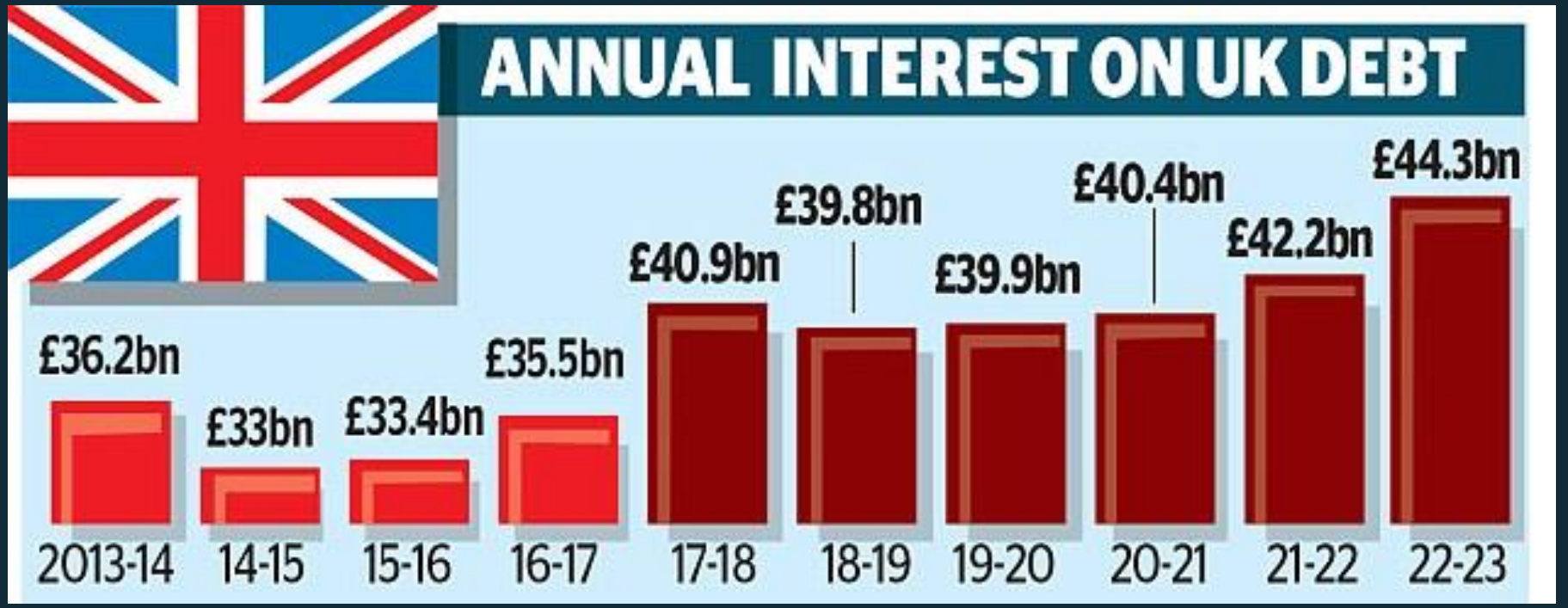
Jonathan Wheatley in London SEPTEMBER 25 2022 77

https://www.ft.com/content/eddedee3-669d-42cc-9597-33609a8bff99



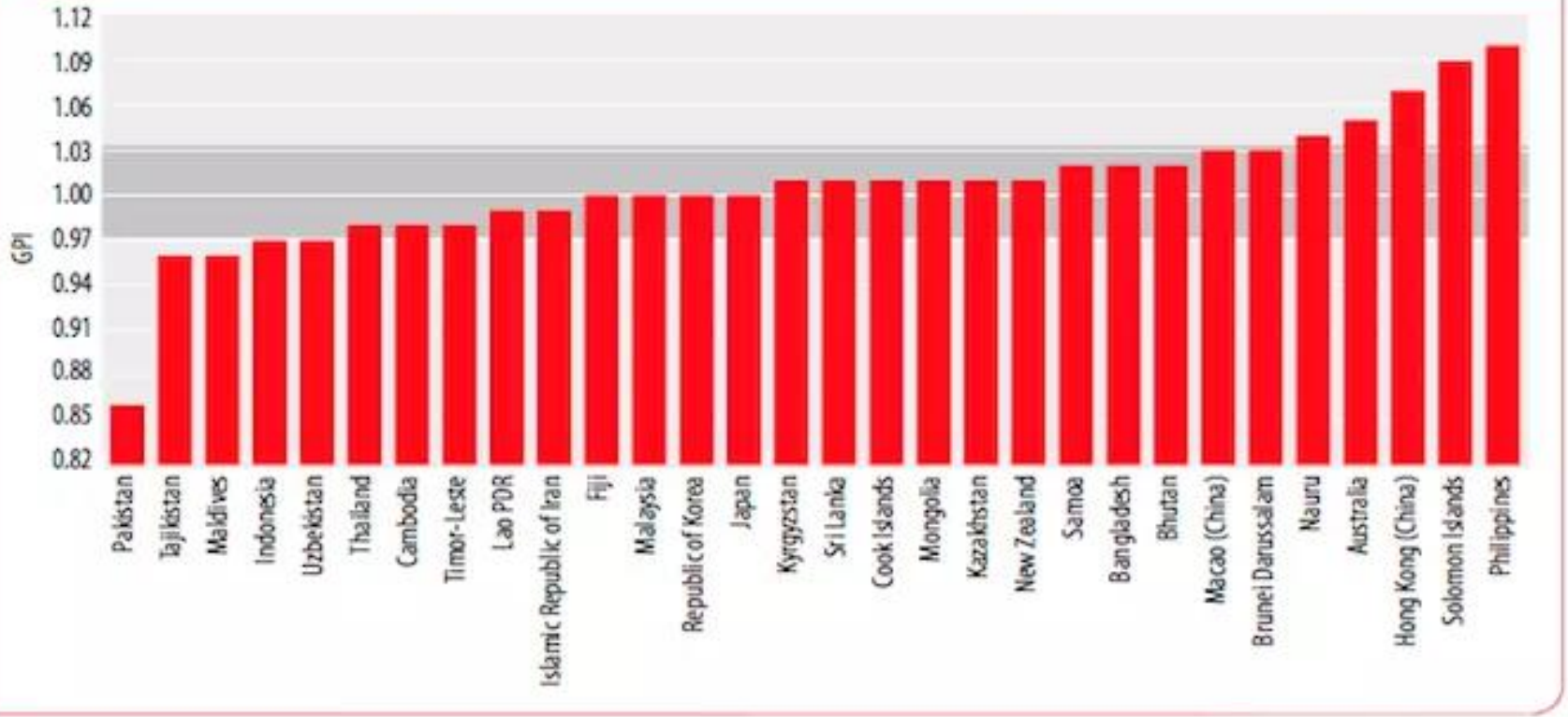
Two worked examples.

What is wrong here? And how can we improve things?



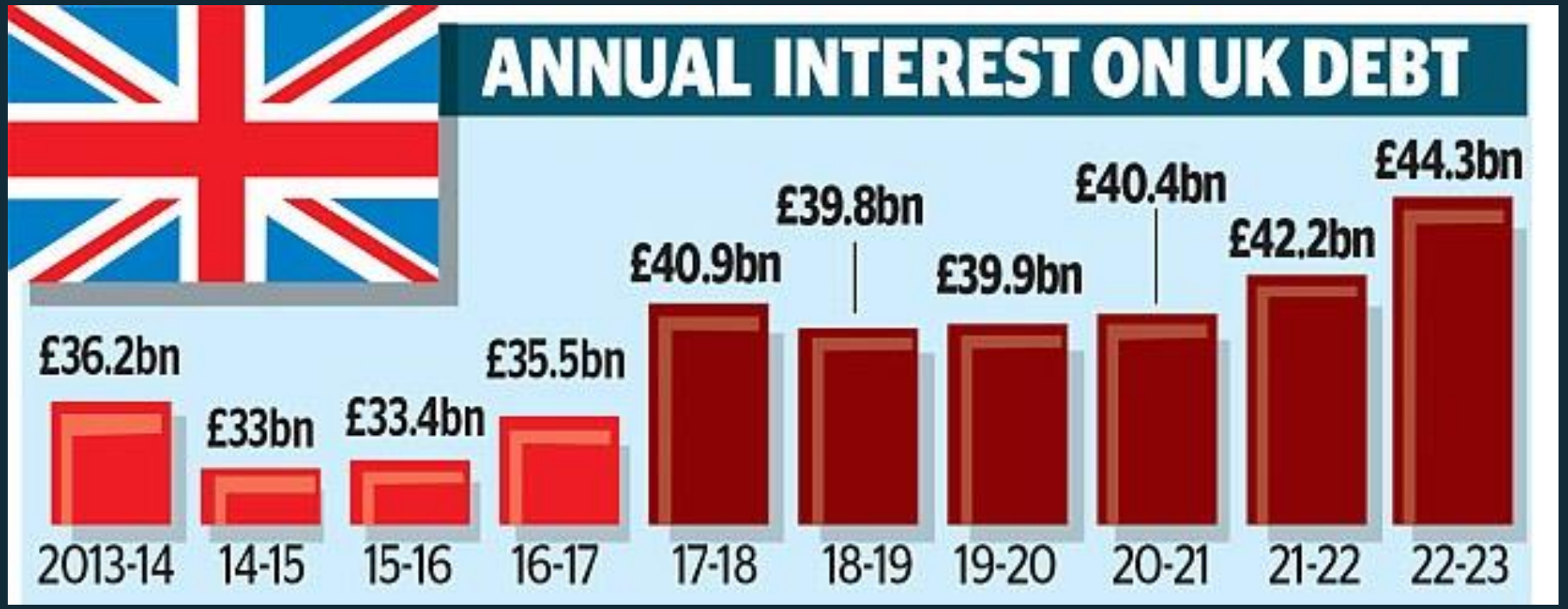
Source: Daily Mail, 2 Jan 2018

Figure 7: Gender Parity Index of the adjusted net intake rate in primary education, 2009

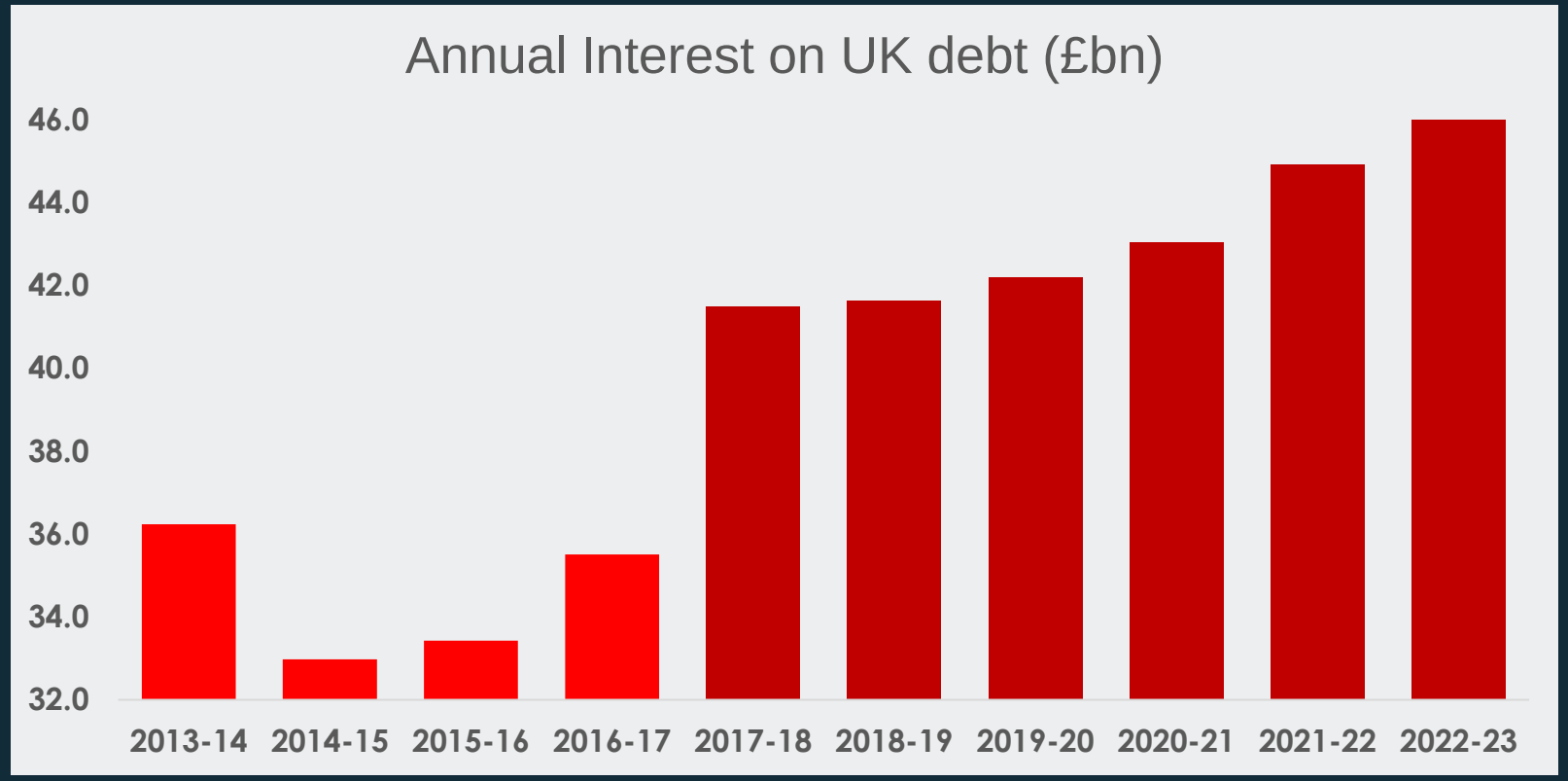


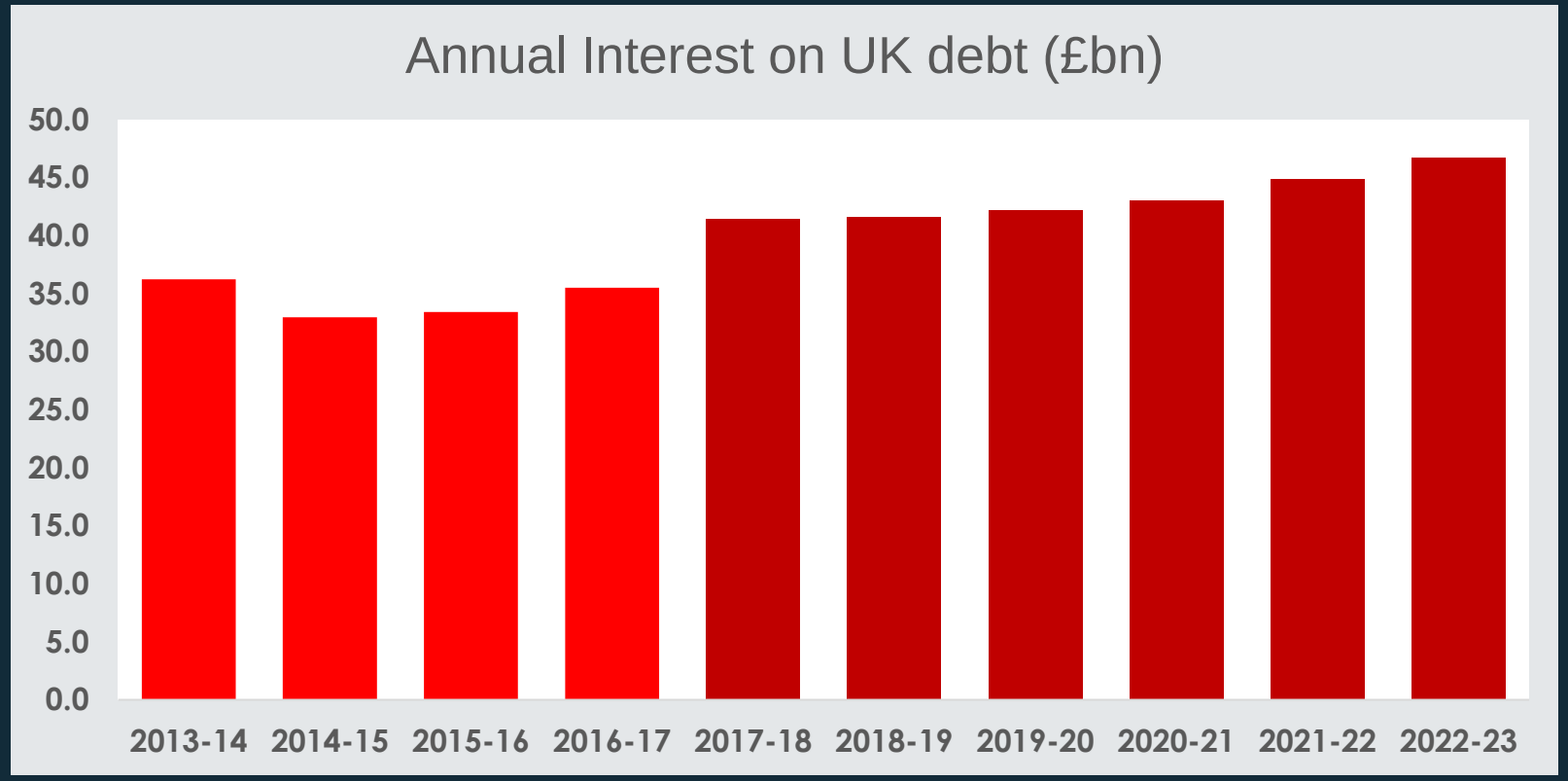
Source: UIS, 2011, Statistical Table 2.

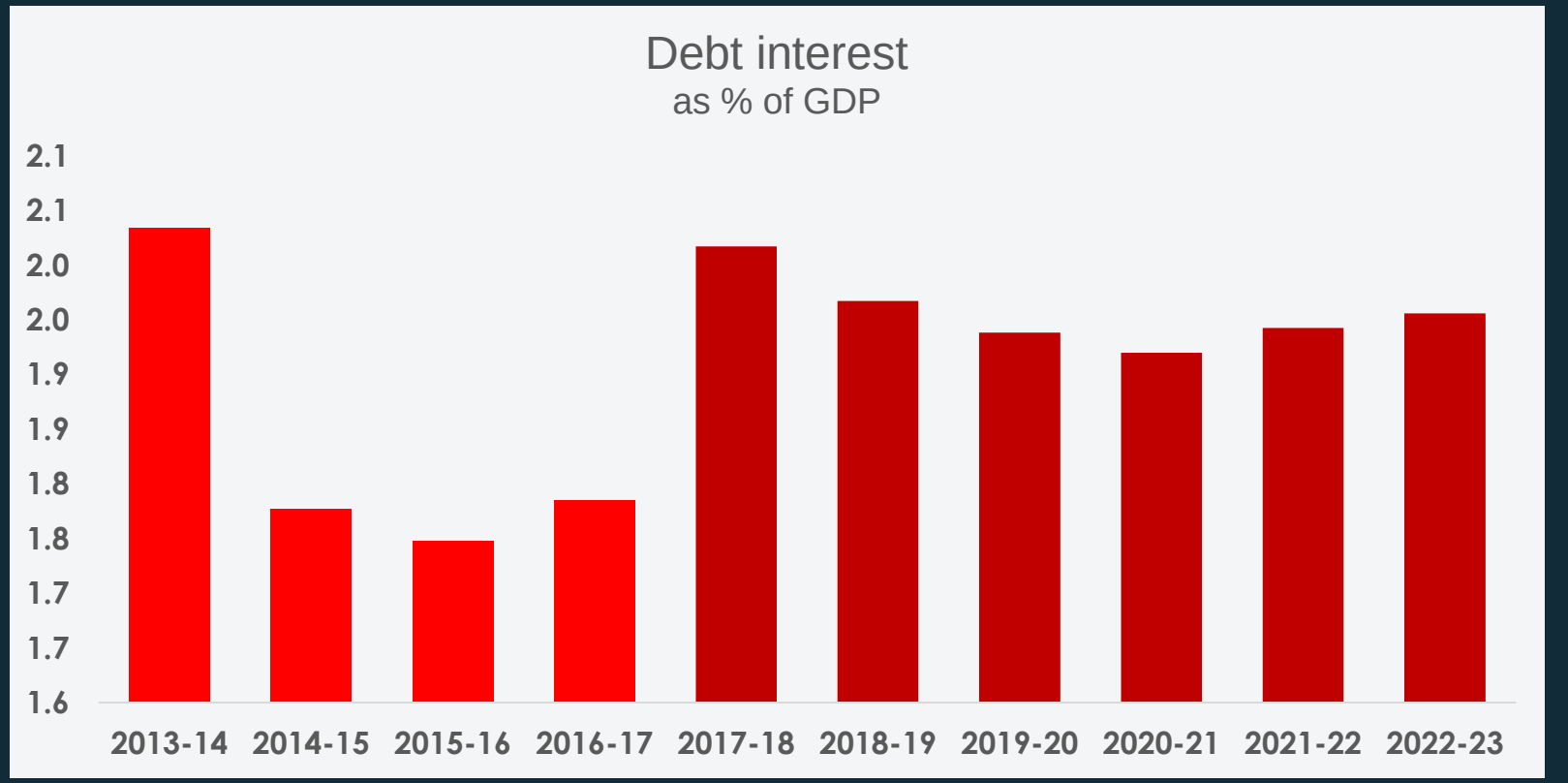
The same data: Different stories

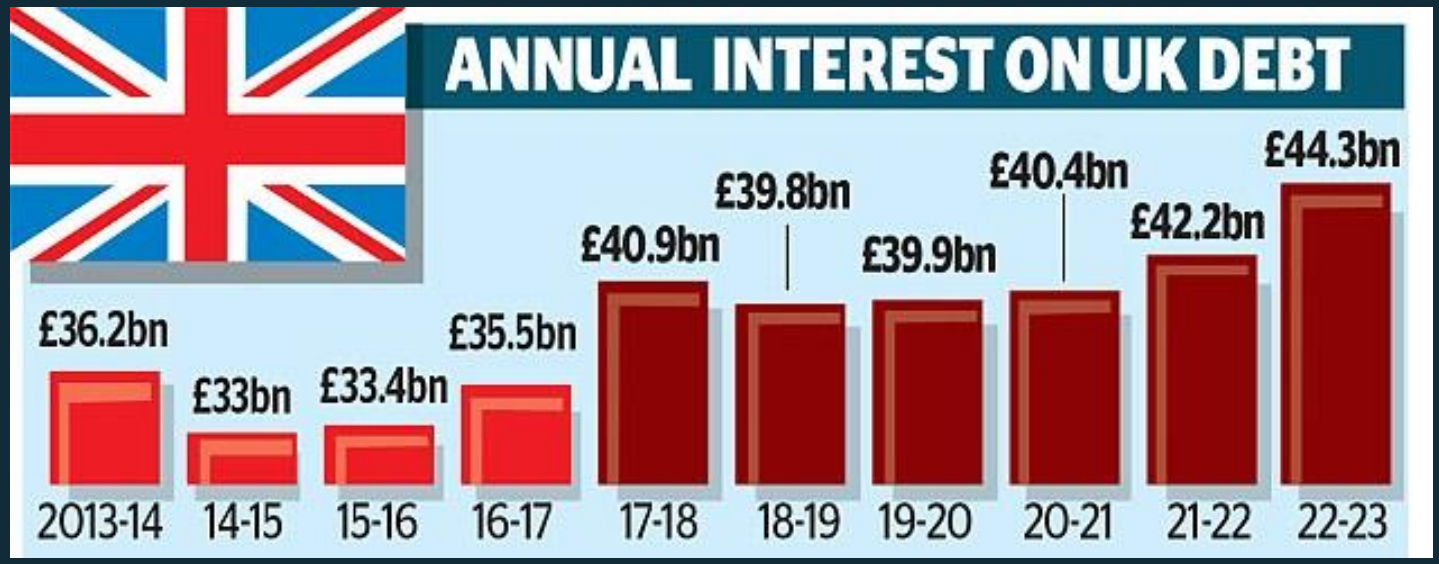


Source: Daily Mail, 2 Jan 2018

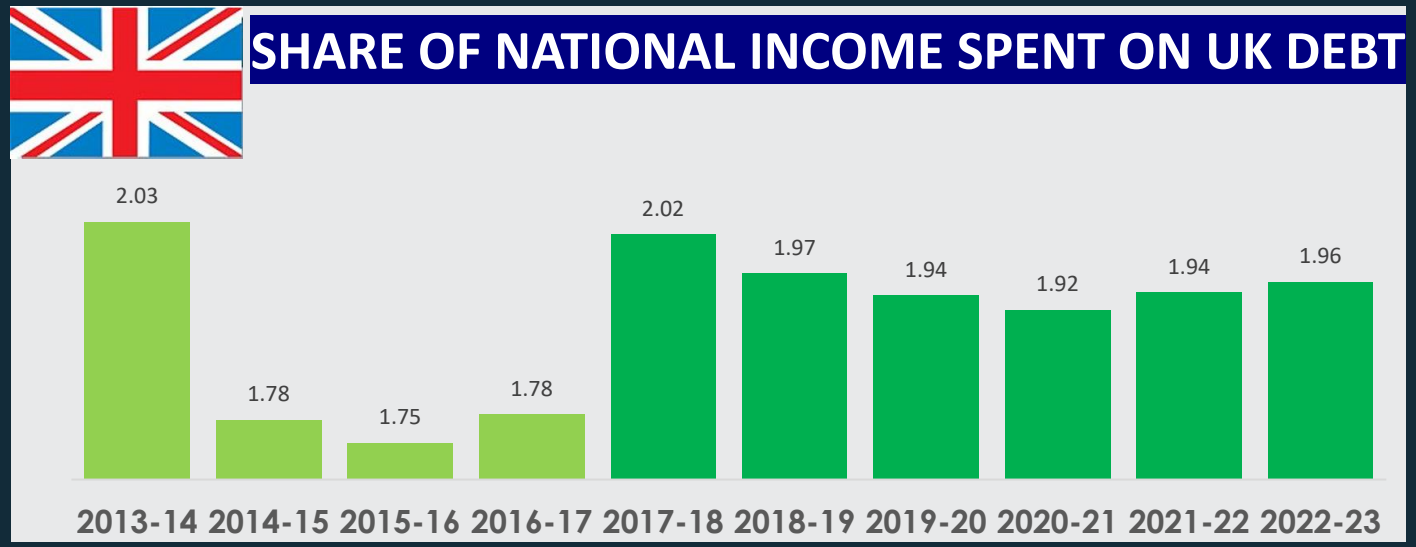

Annual Interest on UK debt (£bn)
46.0
44.0
42.0
40.0
38.0
36.0
34.0
32.0
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
2020-21
2021-22
2022-23


Annual Interest on UK debt (£bn)
50.0
45.0
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
2020-21
2021-22
2022-23


Debt interest
as % of GDP
2.1
2.1
2.0
2.0
1.9
1.9
1.8
1.8
1.7
1.7
1.6
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
2020-21
2021-22
2022-23



Source: Daily Mail, 2 Jan 2018


SHARE OF NATIONAL INCOME SPENT ON UK DEBT
2.03
1.78
1.75
1.78
2.02
1.97
1.94
1.92
1.94
1.96
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
2020-21
2021-22
2022-23

Always 2% - the remarkable stability of the UK

Interest payments on UK national debt as a % of GDP

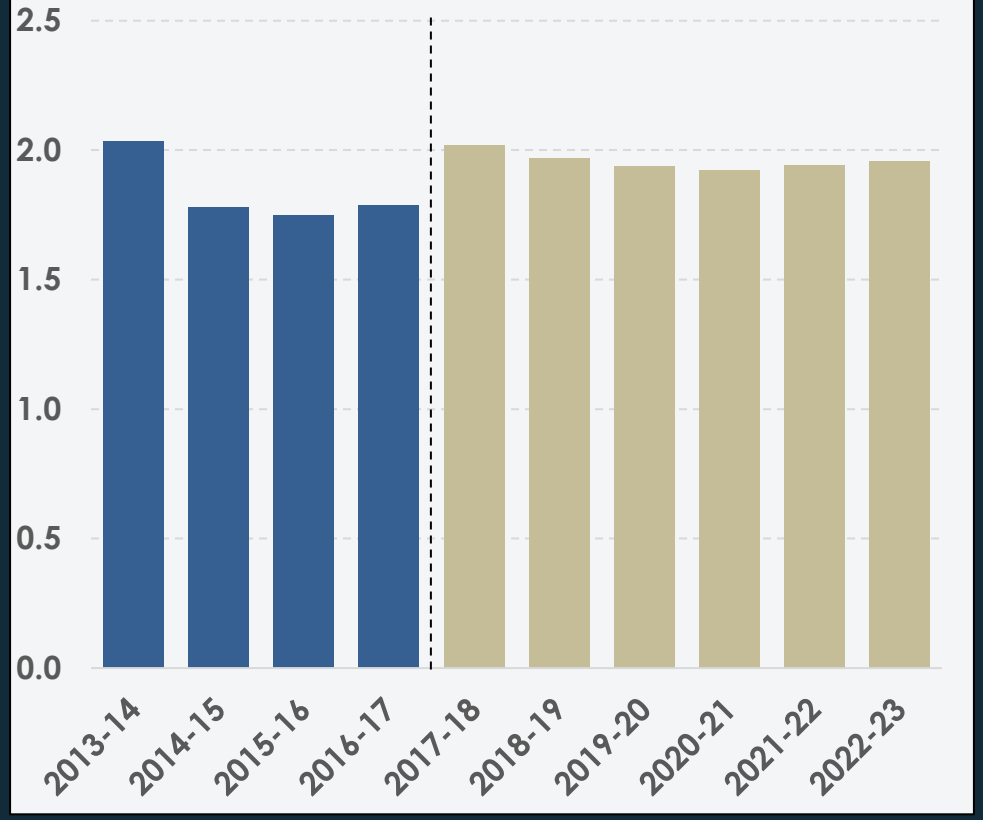
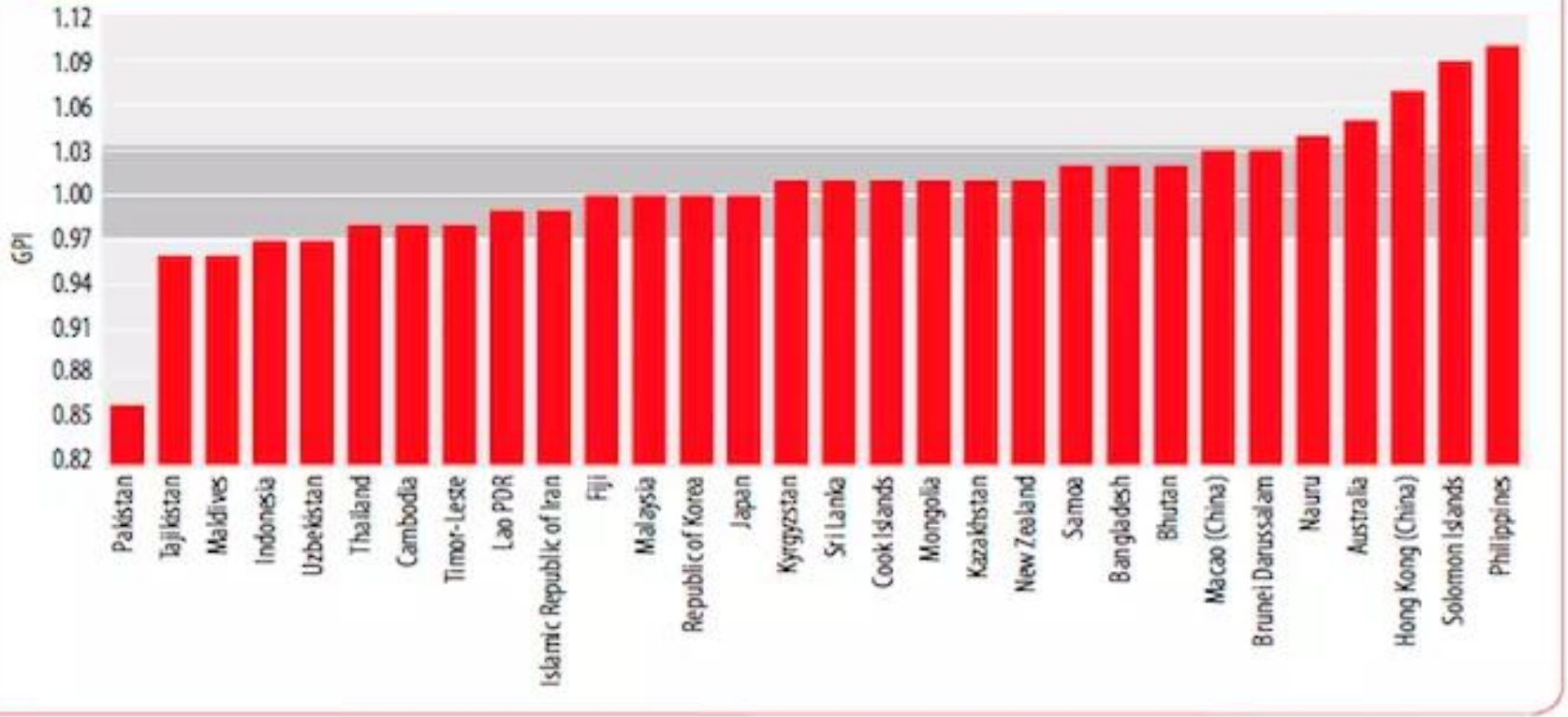
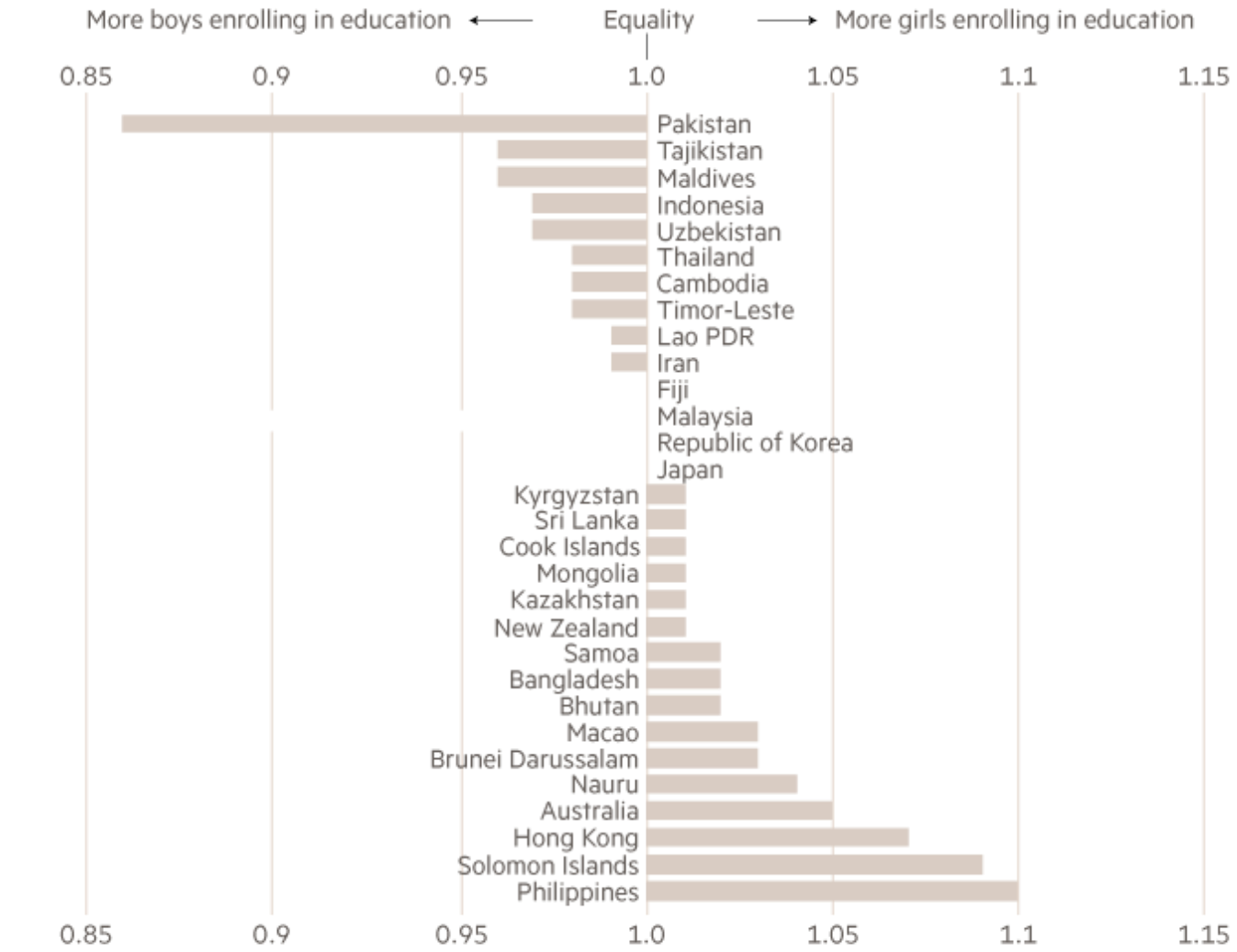


Figure 7: Gender Parity Index of the adjusted net intake rate in primary education, 2009



Source: UIS, 2011, Statistical Table 2.

Gender parity index*, 2009, Asia-Pacific region

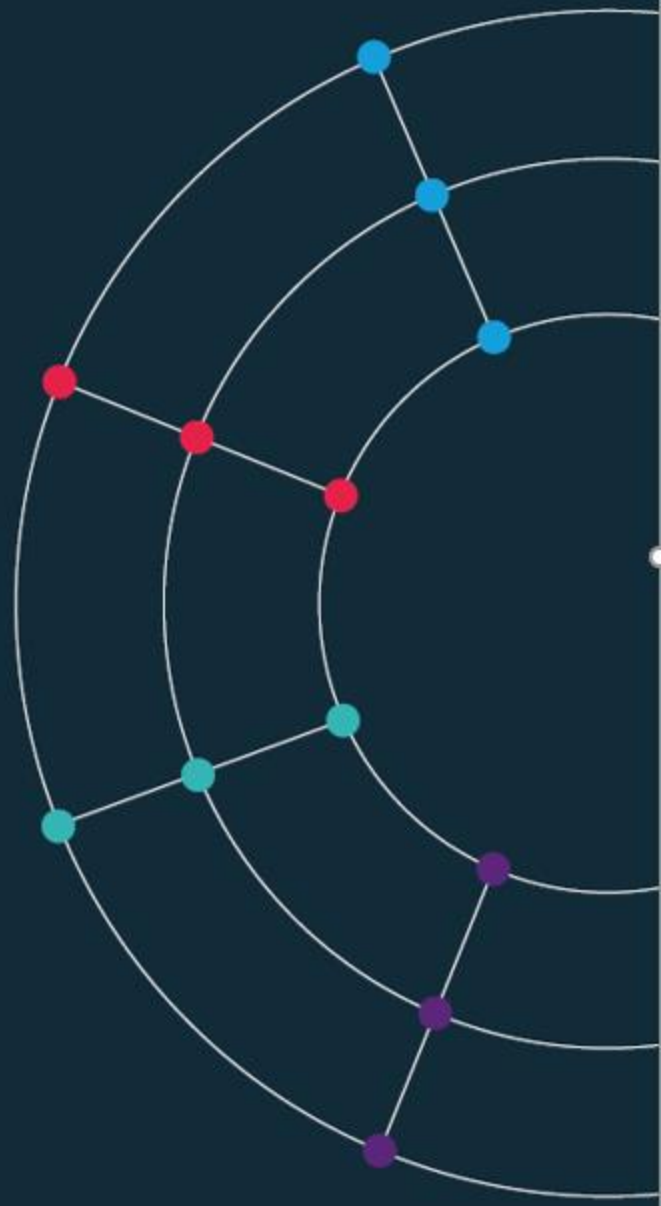


* Of adjusted net intake rate Source: UIS, 2011, Statistical table 2 Graphic by Alan Smith

© FT

Dataviz heroes.

Playfair, Nightingale and other pioneers

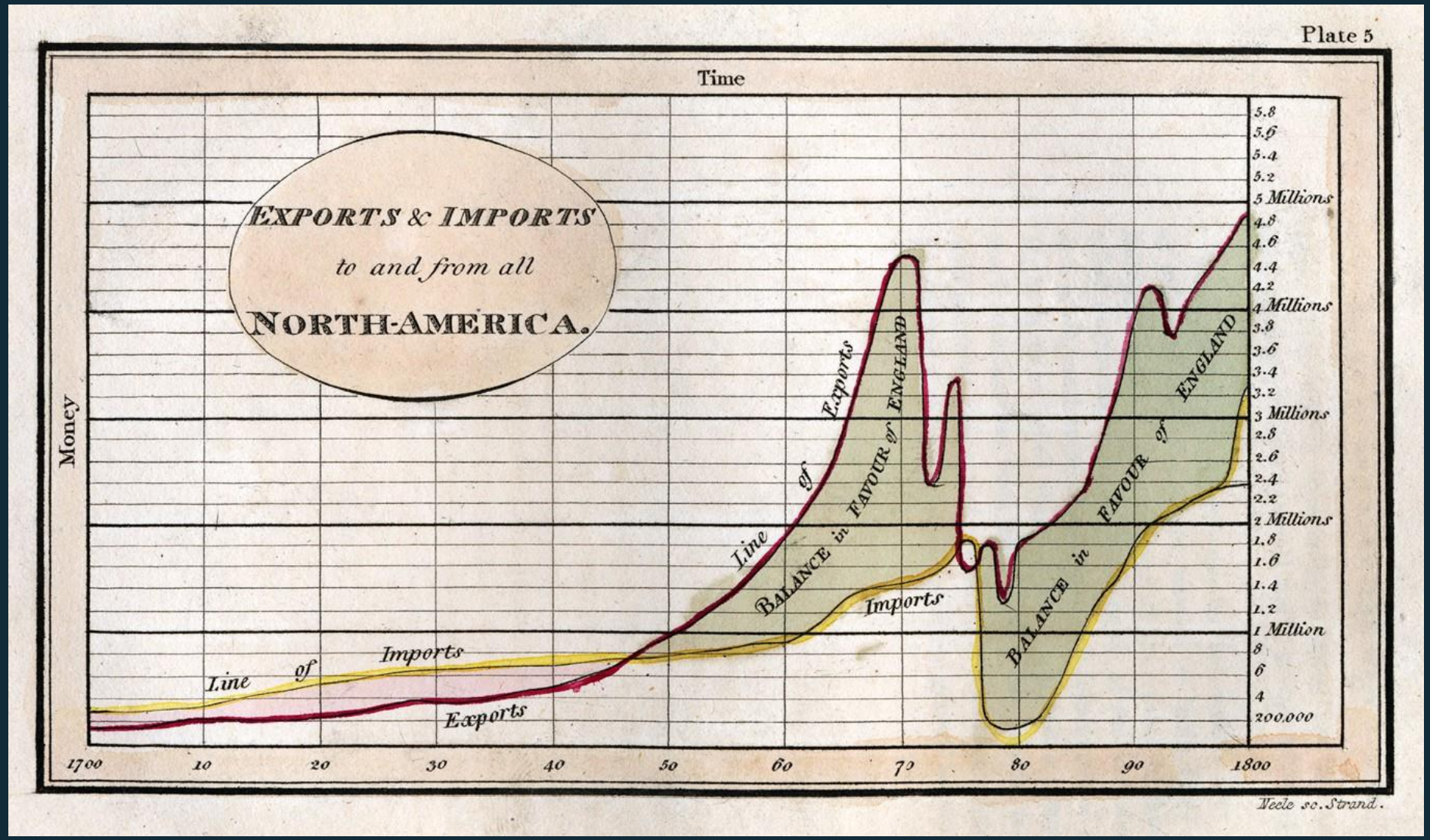


William Playfair.

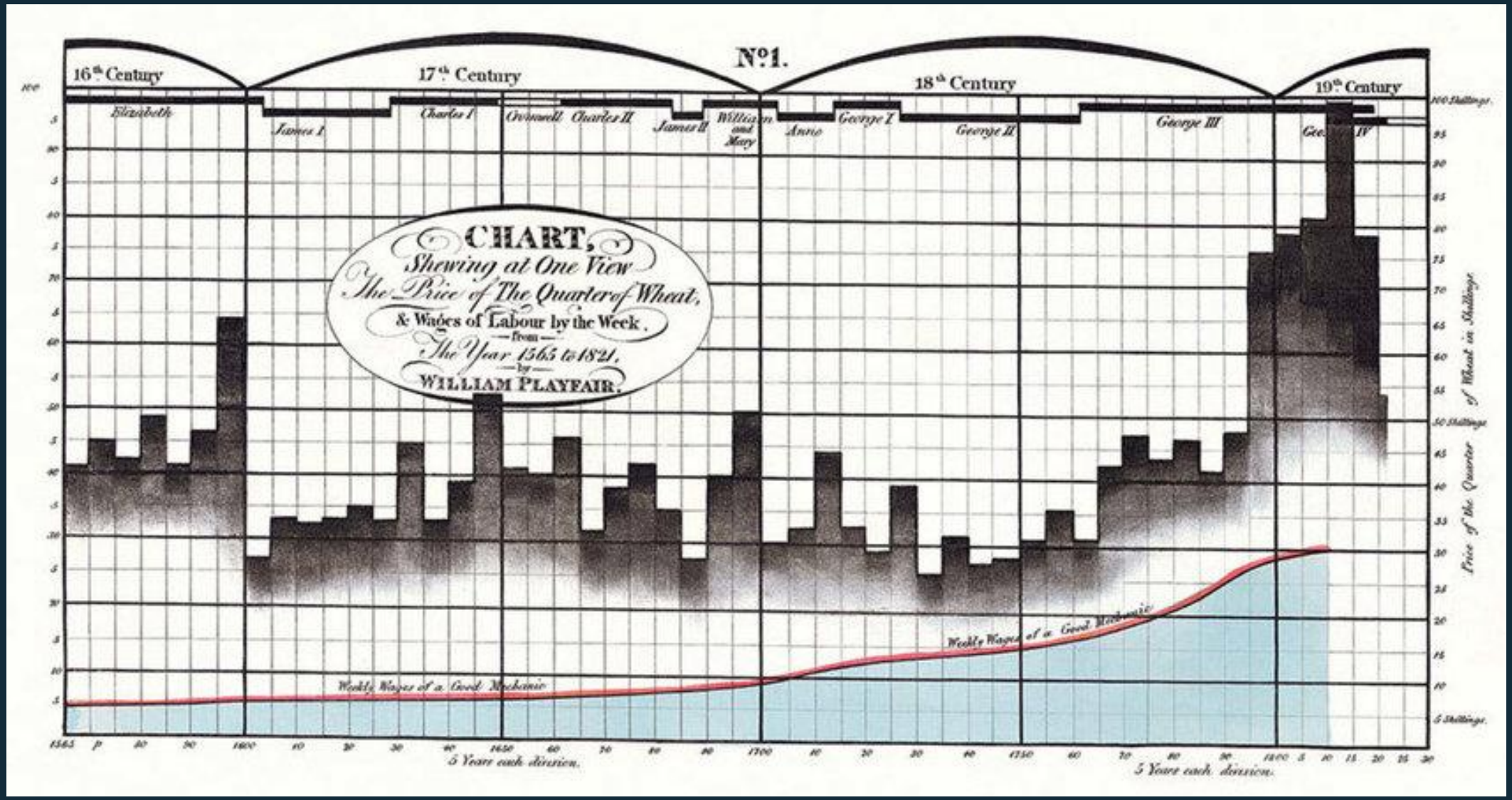
- Born in 1759, William Playfair invented the chart.
- *The Commercial and Political Atlas (1785)* which featured 40 charts covering trade, debt, expenditure and other economic matters
- Playfair was the first to add the vertical axis (previous charts did not have this feature, meaning the way data were stacked vertically meant nothing).
- His charts were relevant to policy. Playfair made charts that demand those in charge of policy to pay attention

"Information, that is imperfectly acquired, is generally as imperfectly retained; and a man who has carefully investigated a printed table, finds, when done, that he has only a very faint and partial idea of what he has read; and that like a figure imprinted on sand, is soon totally erased and defaced."

Politics and trade.

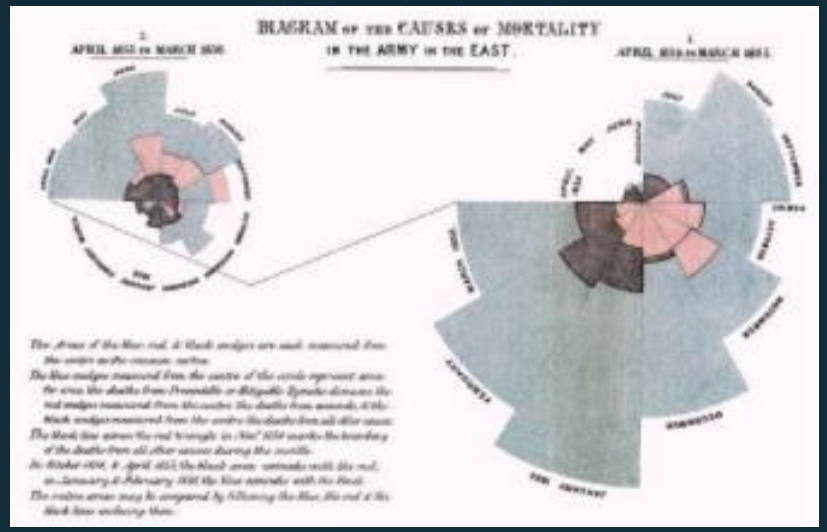


Real wages.

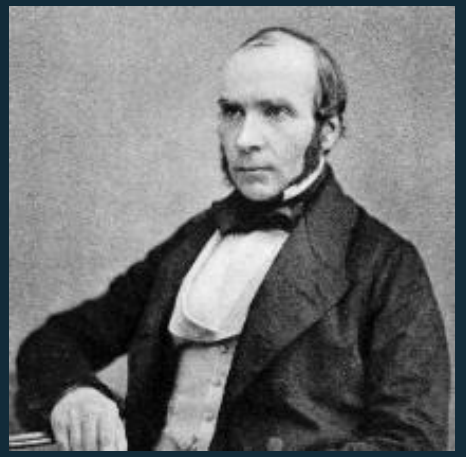

No1.
16th Century
17th Century
18th Century
19th Century
Elizabeth
James I.
Charles I
Oliver Cromwell
Charles II
James II
William and Mary
Anne
George I
George II
George III
CHART,
Shewing at One View
The Price of The Quarter of Wheat,
& Wages of Labour by the Week,
From
The Year 1565 to 1821,
By
WILLIAM PLAYFAIR.
Weekly Wages of a Good Mechanic
Weekly Wages of a Good Mechanic
Price of the Quarter of Wheat in Shillings
100 Shillings
50 Shillings
5 Shillings
5 Years each division.
5 Years each division.

Florence Nightingale.

- Born in 1820, she was a disruptor of both medicine and data
- Statistician, campaigner, nurse, and inventor, she is best known as the founder of modern nursing: the lady with the lamp
- After working on the frontline during the Crimean war, she decided to communicate the miserable conditions she witnessed, creating pioneering using data visualisations to make her case



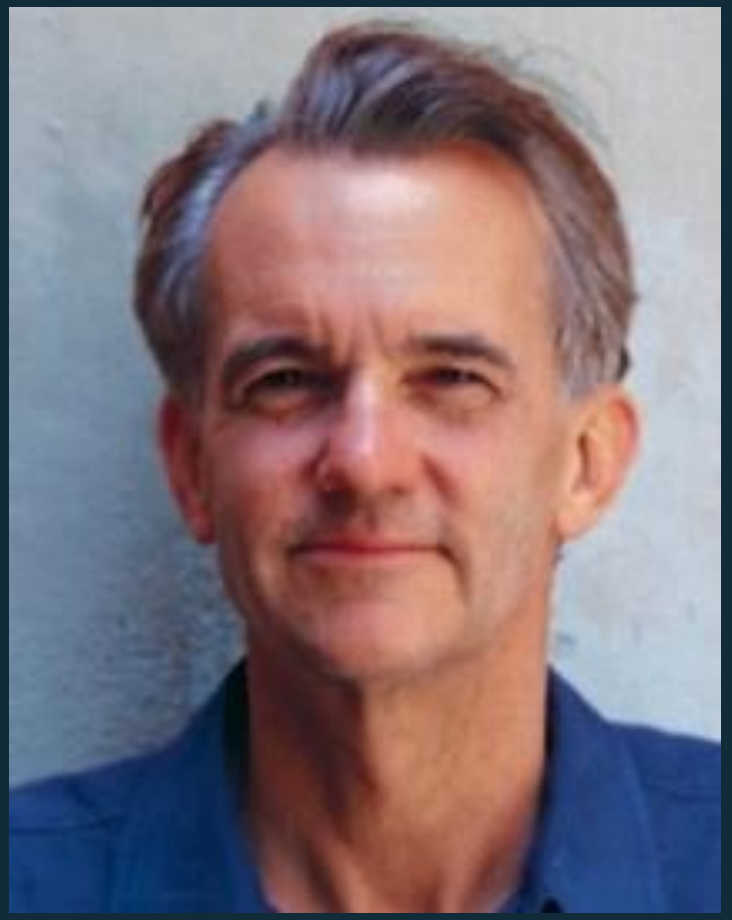
John Snow.



- John Snow was an English physician
- Studied 1854 cholera epidemic in London
- Founding father of epidemiology, a branch of medicine that studies the contraction, spreading and control of disease.
- Data pioneer, using carefully plotted maps to solve a problem that had defeated scientists of his day
- Story of the Broad Street pump

Edward Tufte.

- The American statistician and graphic design theorist.
- *The Visual Display of Quantitative Information*, April 1983
- Primary message: designers didn't have to sacrifice mathematical accuracy in the quest for elegance
- **"data-ink ratio"** to measure the proportion of a given graphic which was simply illustration rather than a representation of data.
- He questioned the "data density" of graphics. And for those graphics which his disliked most coined a new word: "chartjunk".



Anatomy of a good chart.

Y axis metric included in the title

Subtitle including extra detail (such as units)

Y axis label not needed because of title

Quantitative data should start from zero by default

Size. The original version of this chart is 300px by 300px. This is perfect for mobile.

Eye-catching colour. Make sure colours are consistent across your work

X axis label not needed because, in most cases, time is obvious

A useful checklist

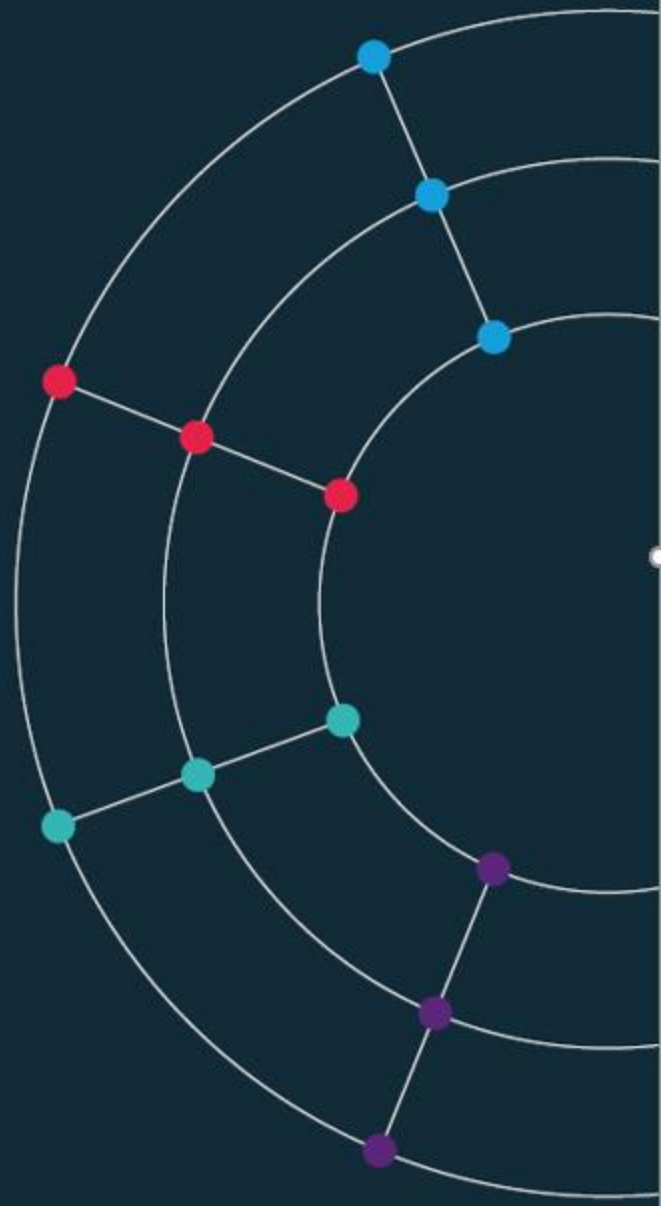
Play your Charts Right

A useful simple set of rules to follow

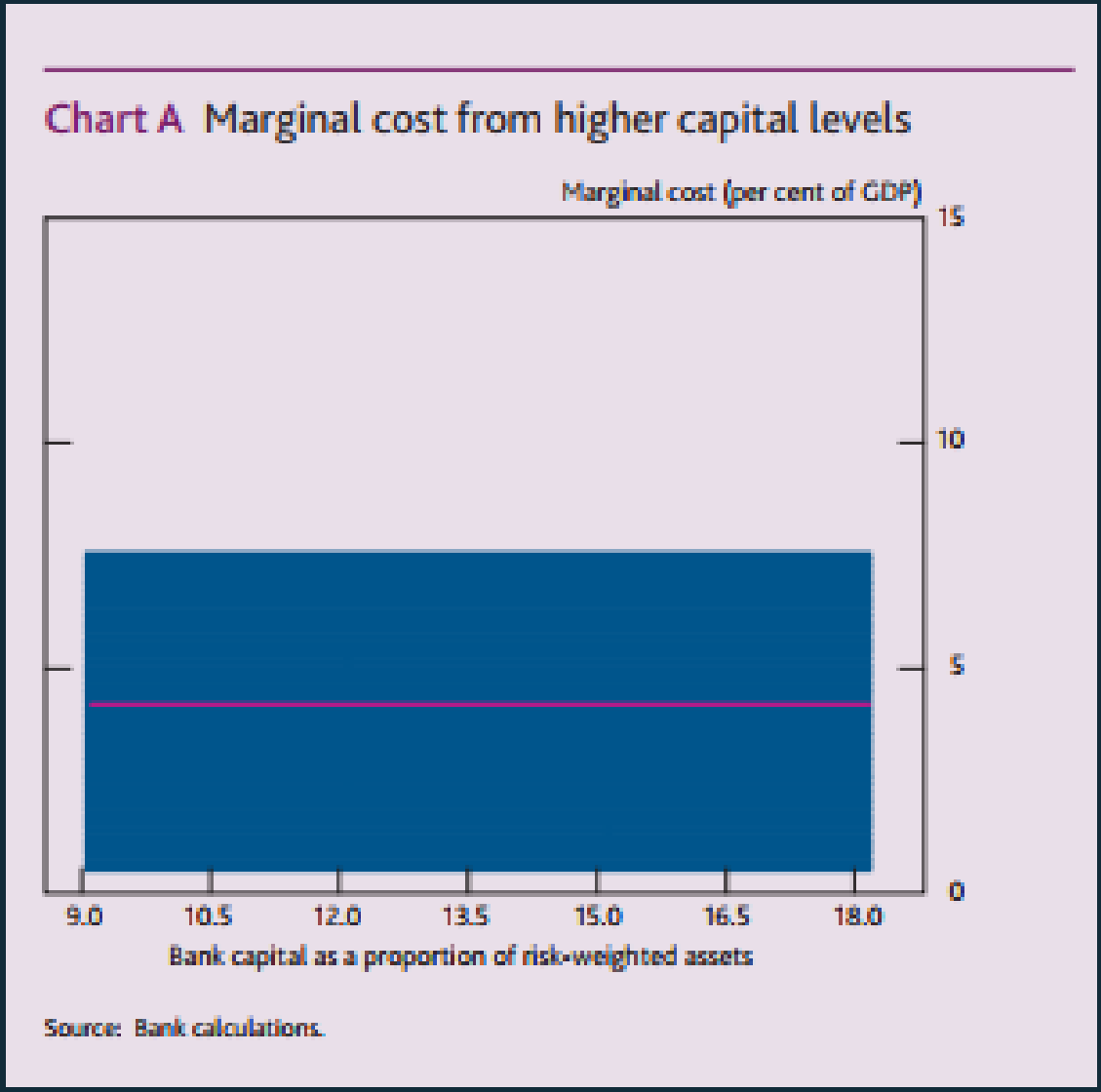
Available in the course folder: here

Killer charts vs Chartjunk.

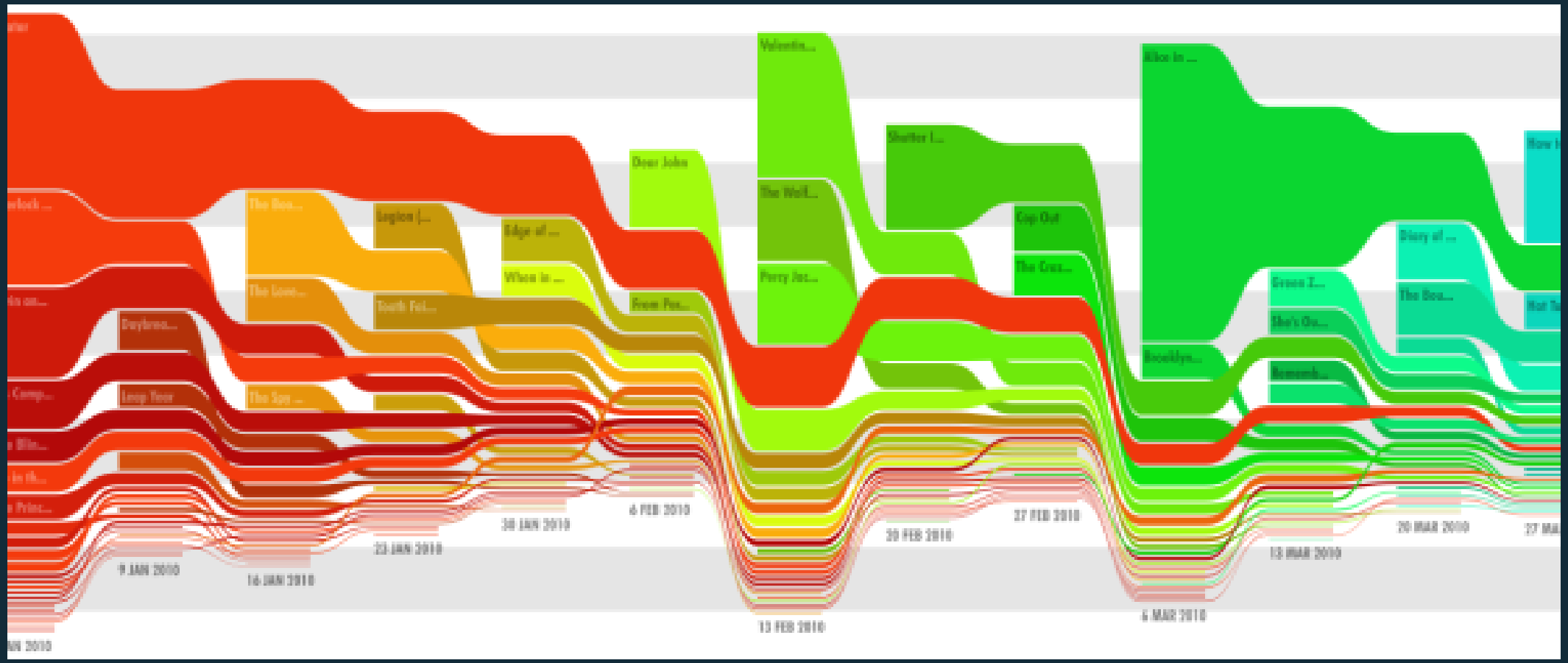
Some examples to discuss



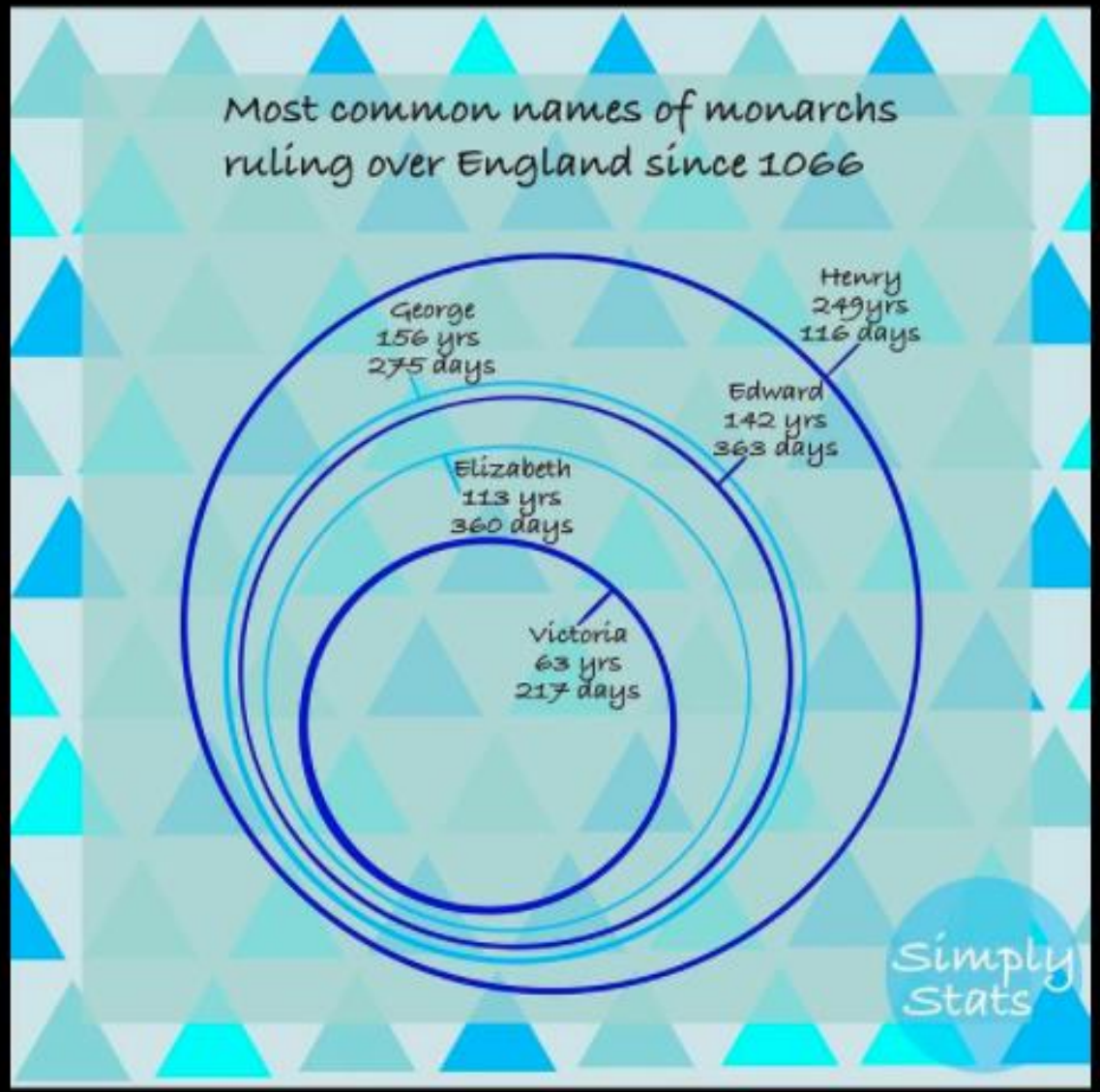
Not everything should be a chart.



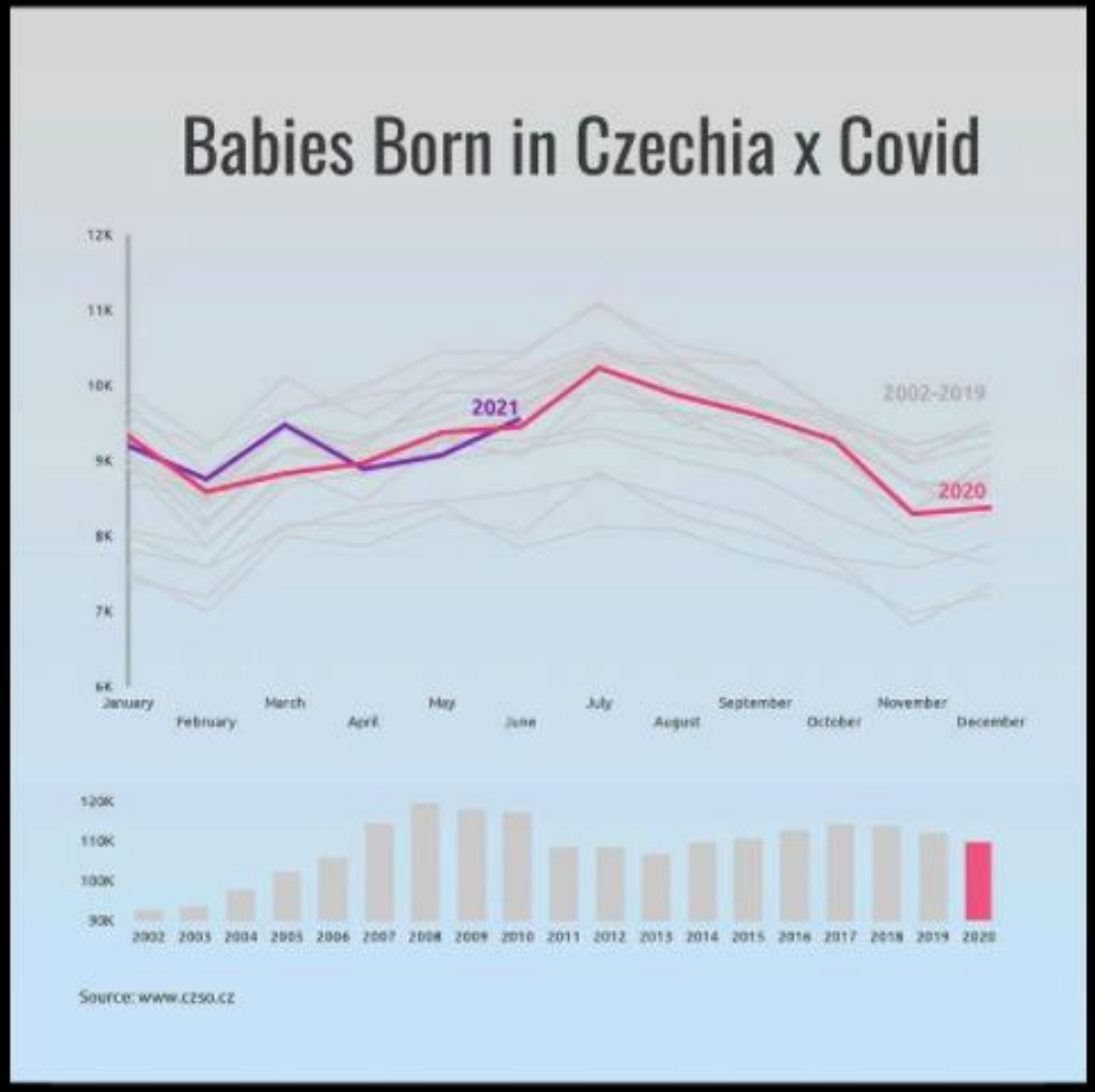
Pretty pictures do not always convey information.



stats_simply

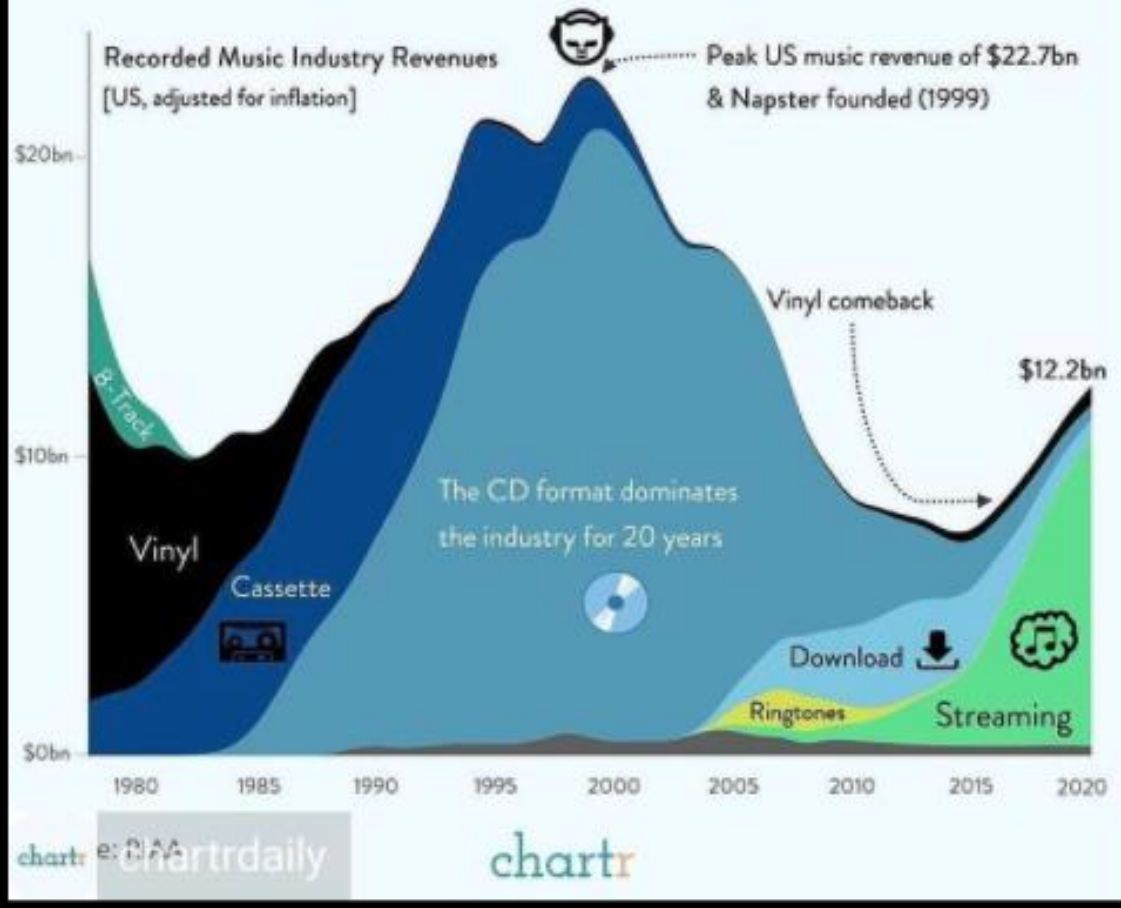


slicer_dicer



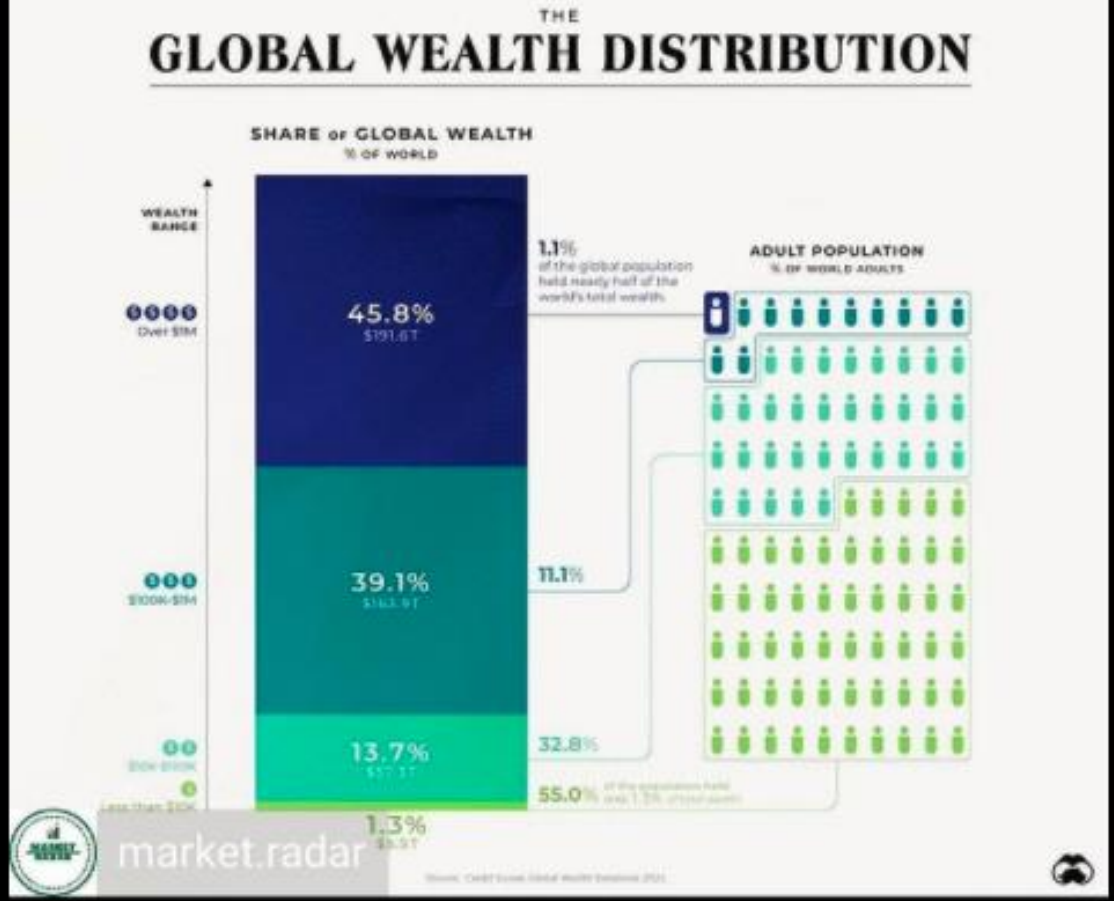
datalaradata
Toronto, Ontario

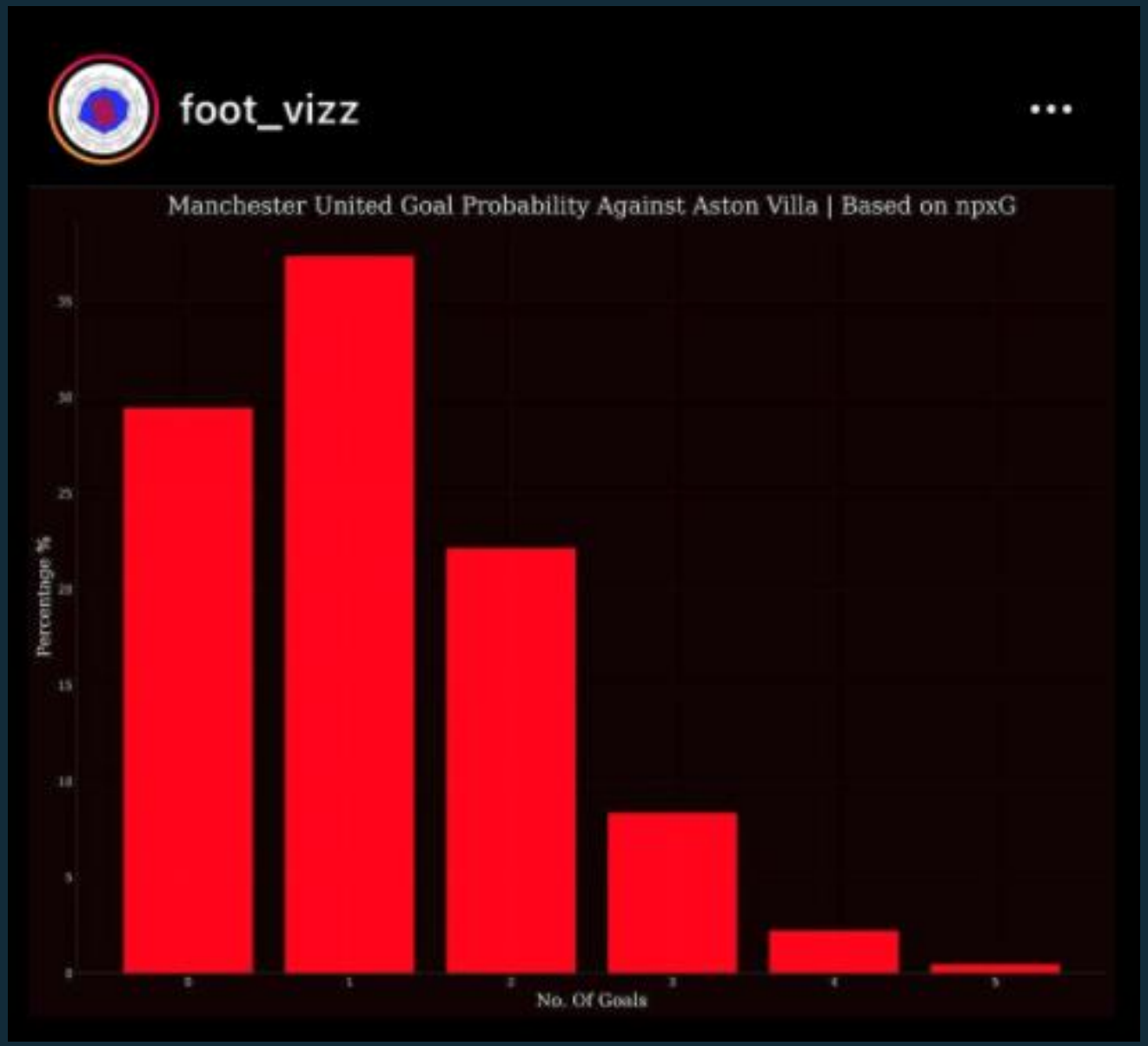
Here Comes The Sun: Streaming Revives The Music Industry



datalaradata
Toronto, Ontario

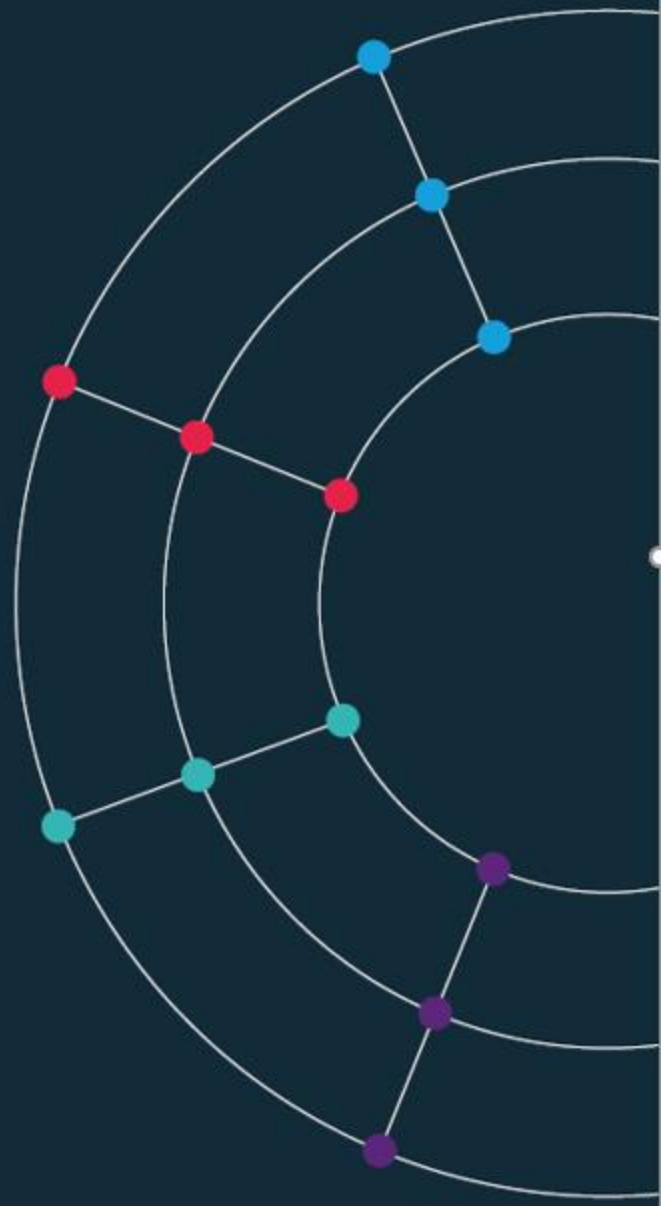
1% of adult population holds 46% of global wealth



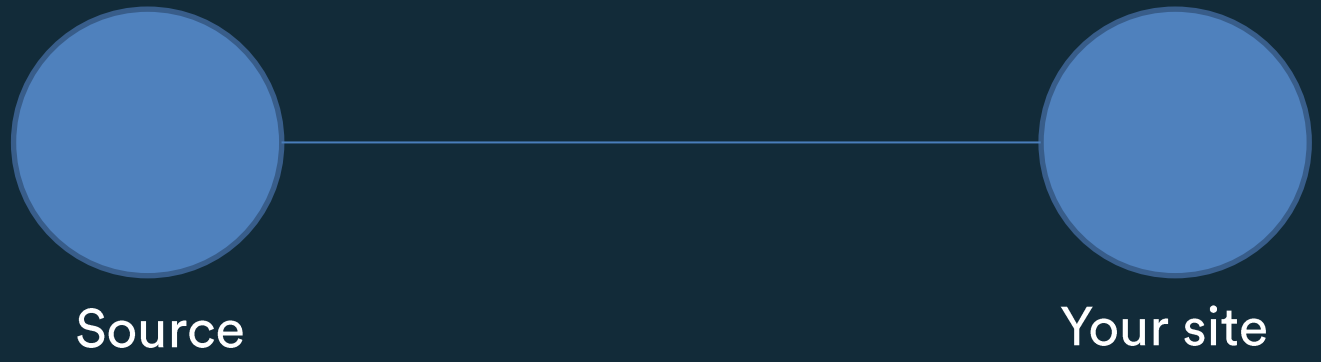

foot_vizz
Manchester United Goal Probability Against Aston Villa | Based on npxG
Percentage %
No. Of Goals

Data workflow.

Two approaches to sourcing and embedding data



Option 1. Direct link to API



Pros

- Auto-updating
- No data errors

Cons

- Slow/lags
- Multiple downloads: may be costly if a paid API.
- If API is unstable, your analysis becomes unstable.

Option 2. Link to live CSV file

Pros

- Fast
- Stable
- APIs don't always exist: ma
  be your only option.

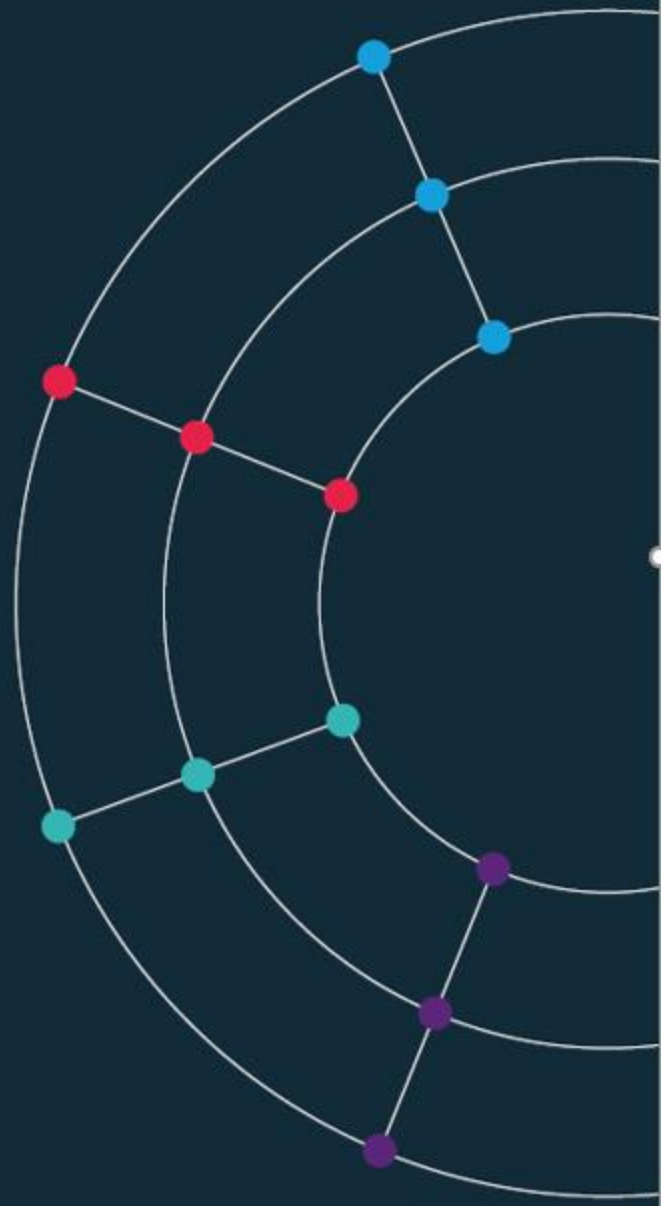
Cons

- Won't auto-update
- Requires maintenance

Practical 2.

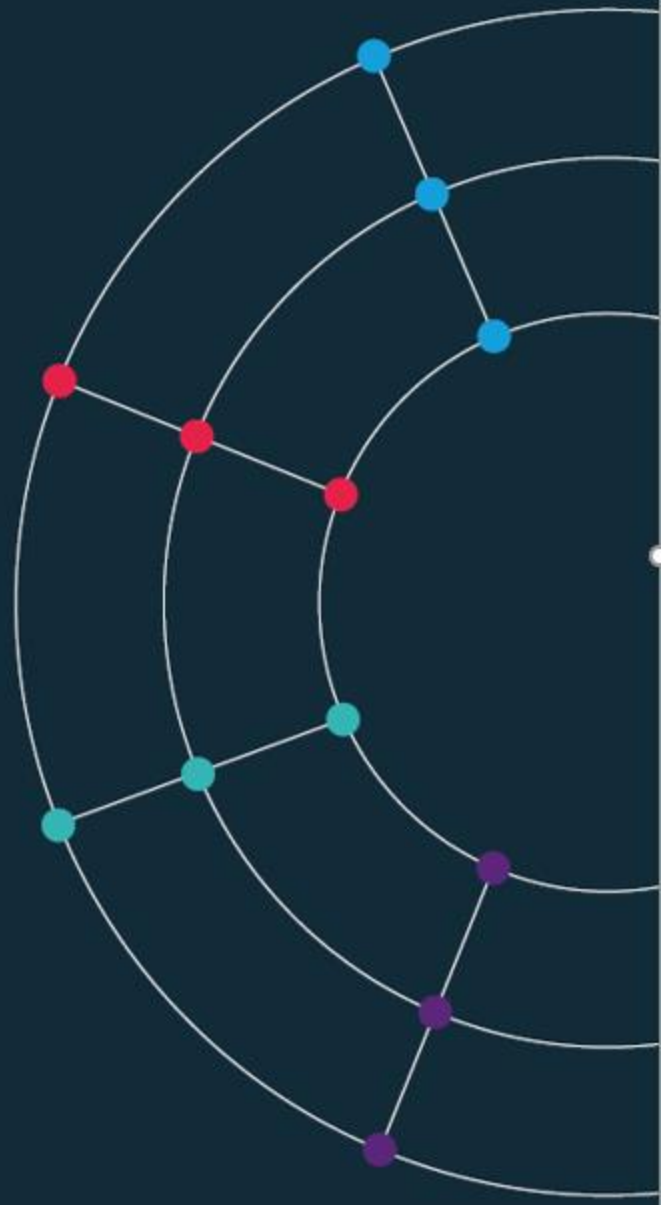
Embedding charts:

1. FT example: a problematic chart.
    1. Data cleaning.
        1. String functions
        2. Date-time functions
2. Embedding your chart.
3. Discussion: how to improve the chart.
4. API driven charts using ONS API.



Homework 2.

Week 2 – Three charts: an API, a GitHub, and your first original.

Clearly label your portfolio page so that we can see Week 2. Now add three charts:

1. A chart of IMF lending outstanding, hosted on your GitHub account. The data can be found here and here
2. A chart correcting the one used by the FT. The data for this will also be hosted on your GitHub account.
3. A chart which has data delivered by an API. We will discuss this in class.