

Festival of Economics.

1-3 October 2024

Book talks, panels & live podcast recordings – your chance to ask experts from around the world the key economic questions of our time!

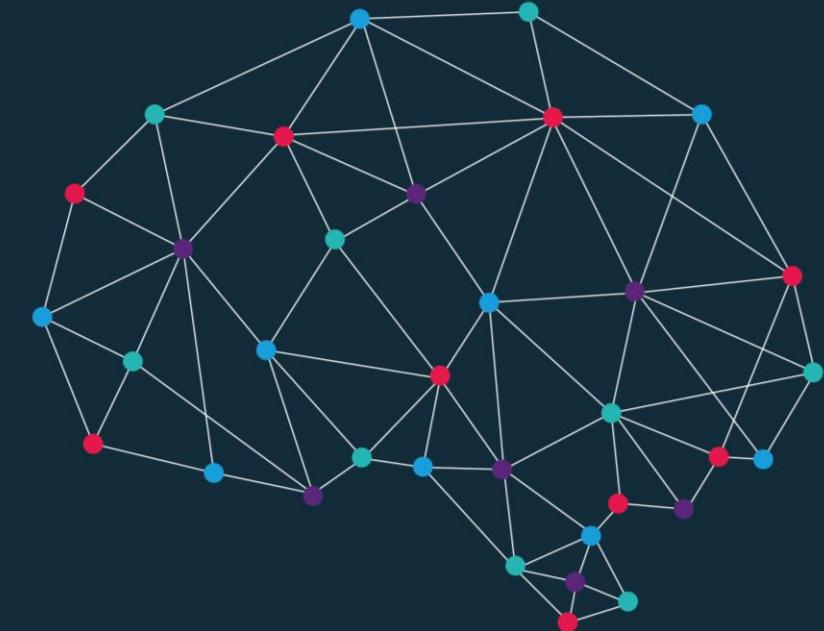
Scan for tickets and the programme.



FREE TICKETS for:

- **Legacies of Empire** (Tues 1 October, 7pm)
- **Inequalities** (Thursday 3 Oct, 7pm)

in the Wills Memorial Building and are free to UoB students



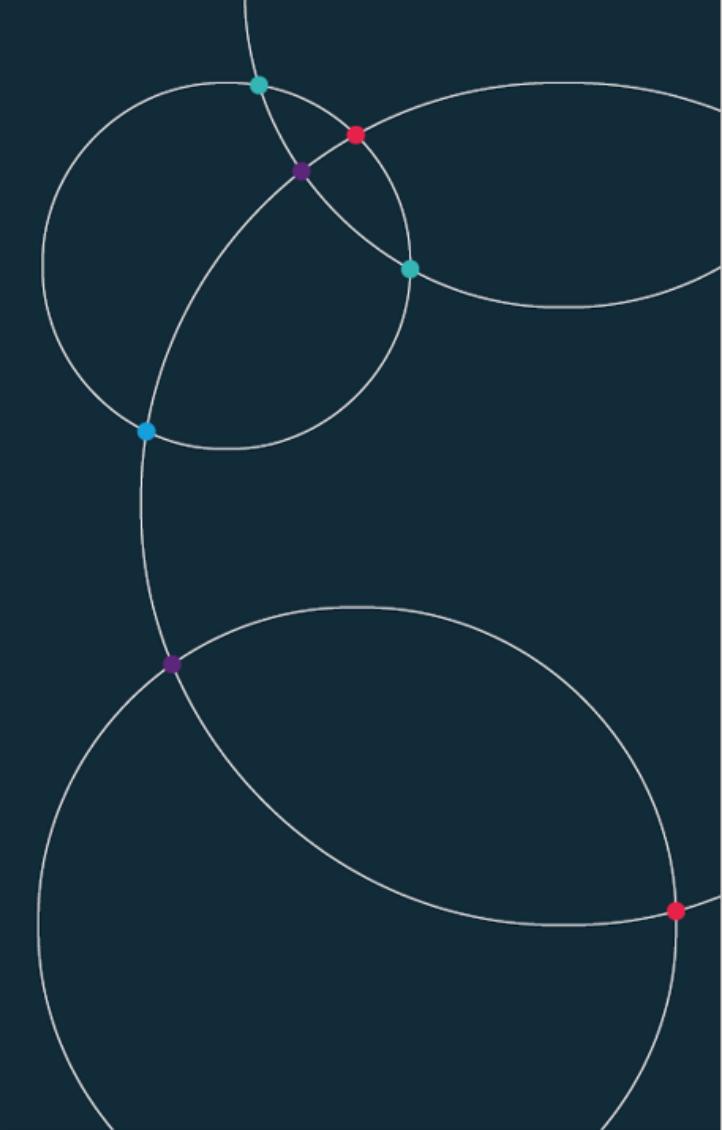
With the election of a new government after 14 years, the theme of the Festival is **looking back, to look forward:**

Legacies of Empire, Inequalities, The Tyranny of Nostalgia, Sustainable Futures, Labour's first 100 days, Building a fair society, Britain's place in the world , Social mobility, The kids aren't alright, UK productivity, Tim Harford podcast ... and more!

Data Science.

Week 3.

University of Bristol
Autumn 2024



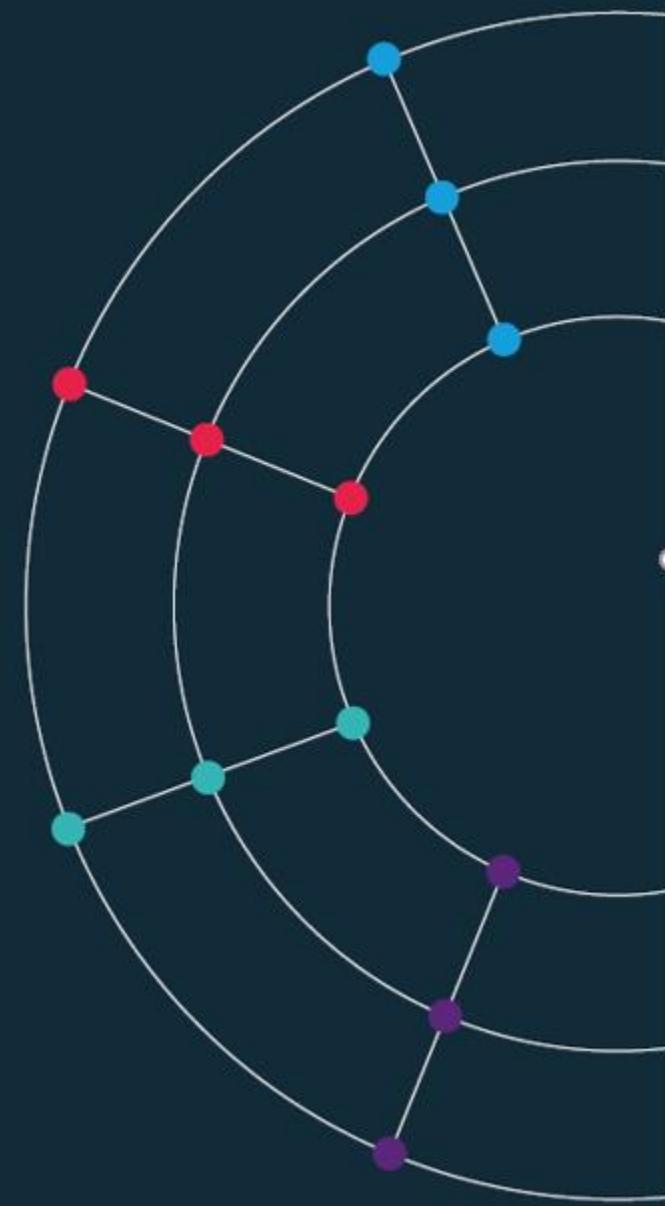
Office hours.

- Richard. Tuesday, 14.00-15.00 (TBC)
- Denes. Thursday, TBC (online)
- Finn. Friday, 13.30-14.30 (online)
- Josh. Thursday, 10.00-11.00 (TBC)

Week 3.

- **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 3.** Making a chart from scratch.

Bad, Ugly ... and Good.



When charts go wrong.



When charts go wrong.

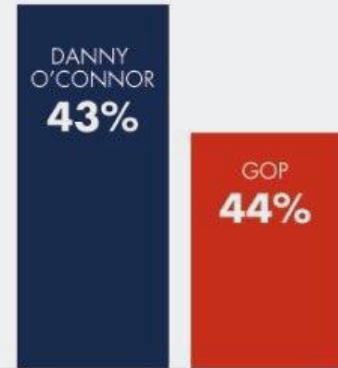


↓ *Nate Silver's TOP RATED pollster just BLEW. OUR. MINDS* ↓

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



SOURCE:
MONMOUTH POLL



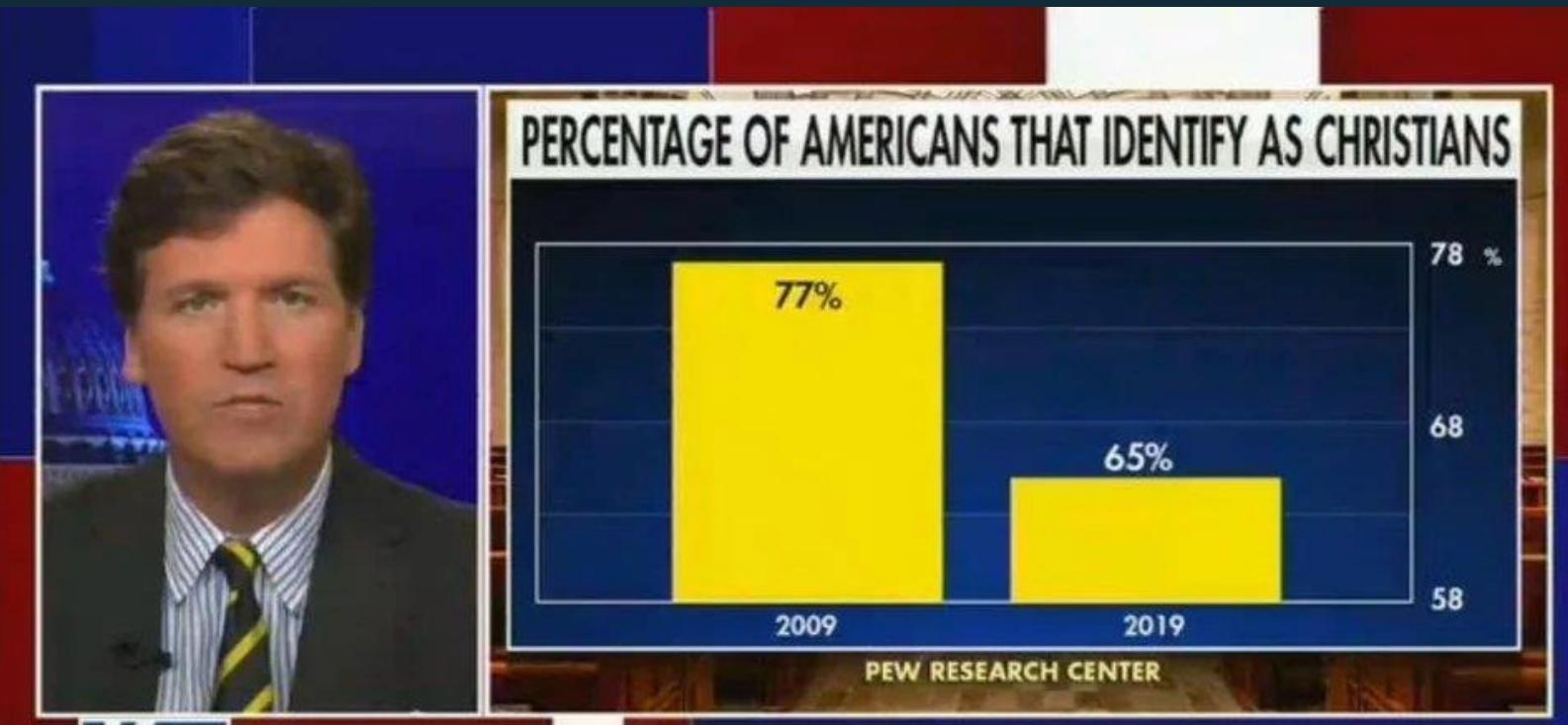
IT'S HAPPENING:
DEMOCRATS ARE ON THE CUSP OF WINNING THE OHIO SPECIAL ELECTION!

*Votes cast in English
Local Elections June 2009*



**Lib Dems
overtake
Labour**

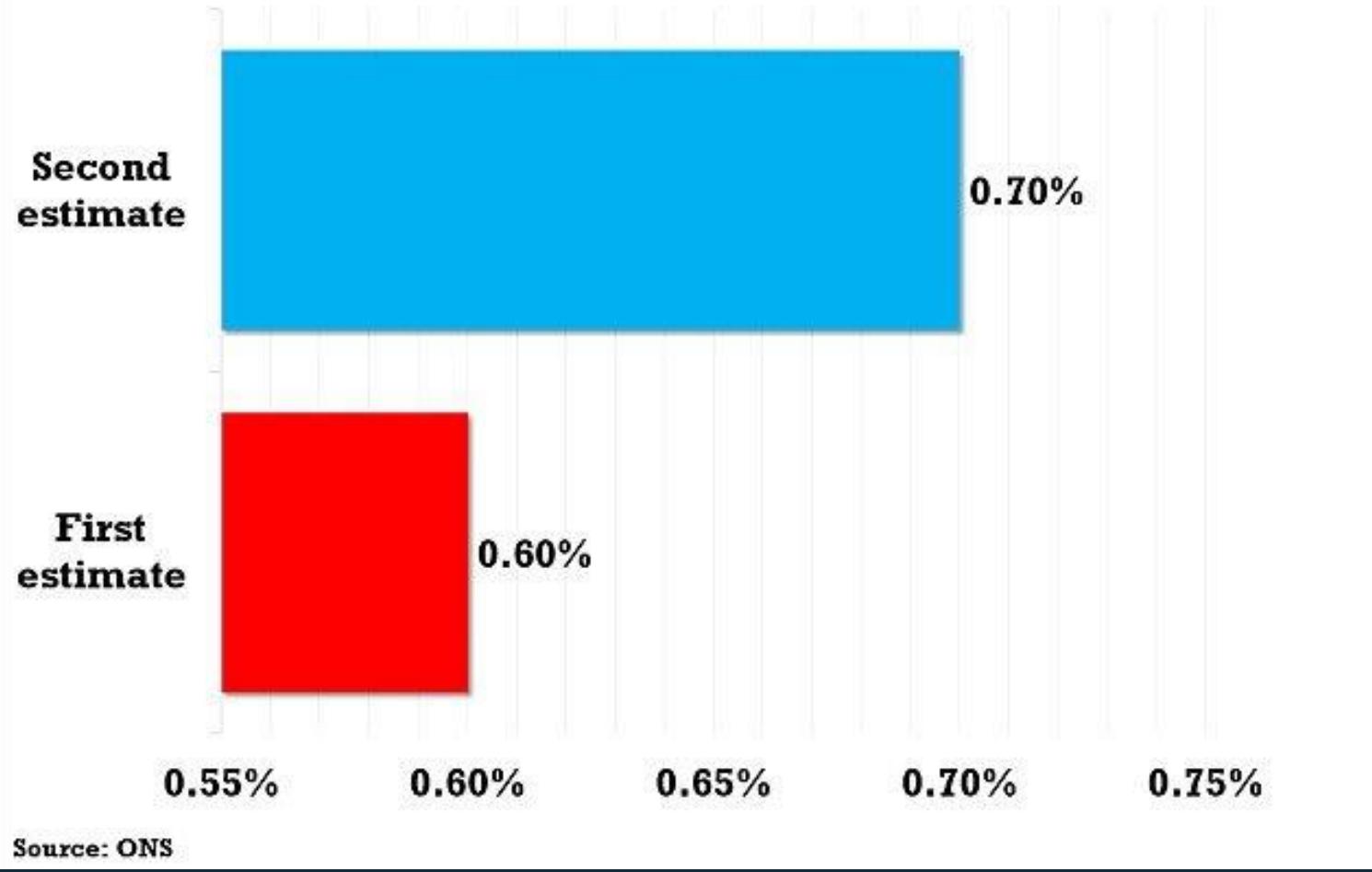
**SEE
PAGE
TWO**



WE USED TO BE AN ENTHUSIASTICALLY CHRISTIAN NATION

TUCKER CARLSON • TONIGHT •

2016 Q4 GROWTH UPGRADED

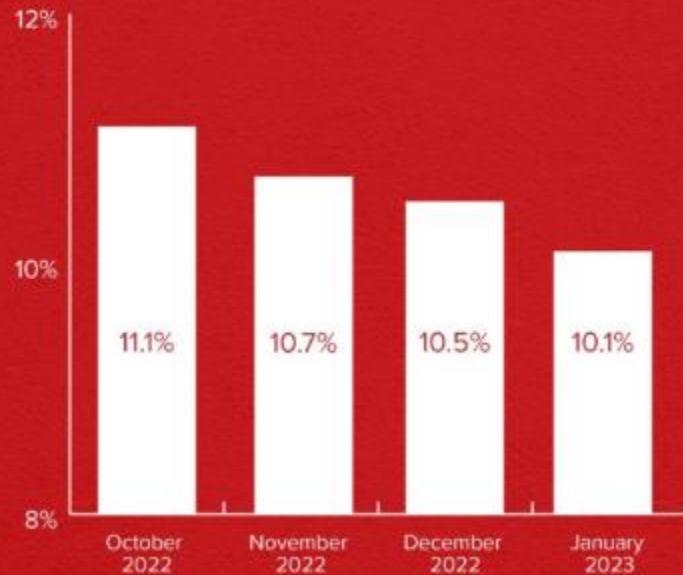




HM Treasury

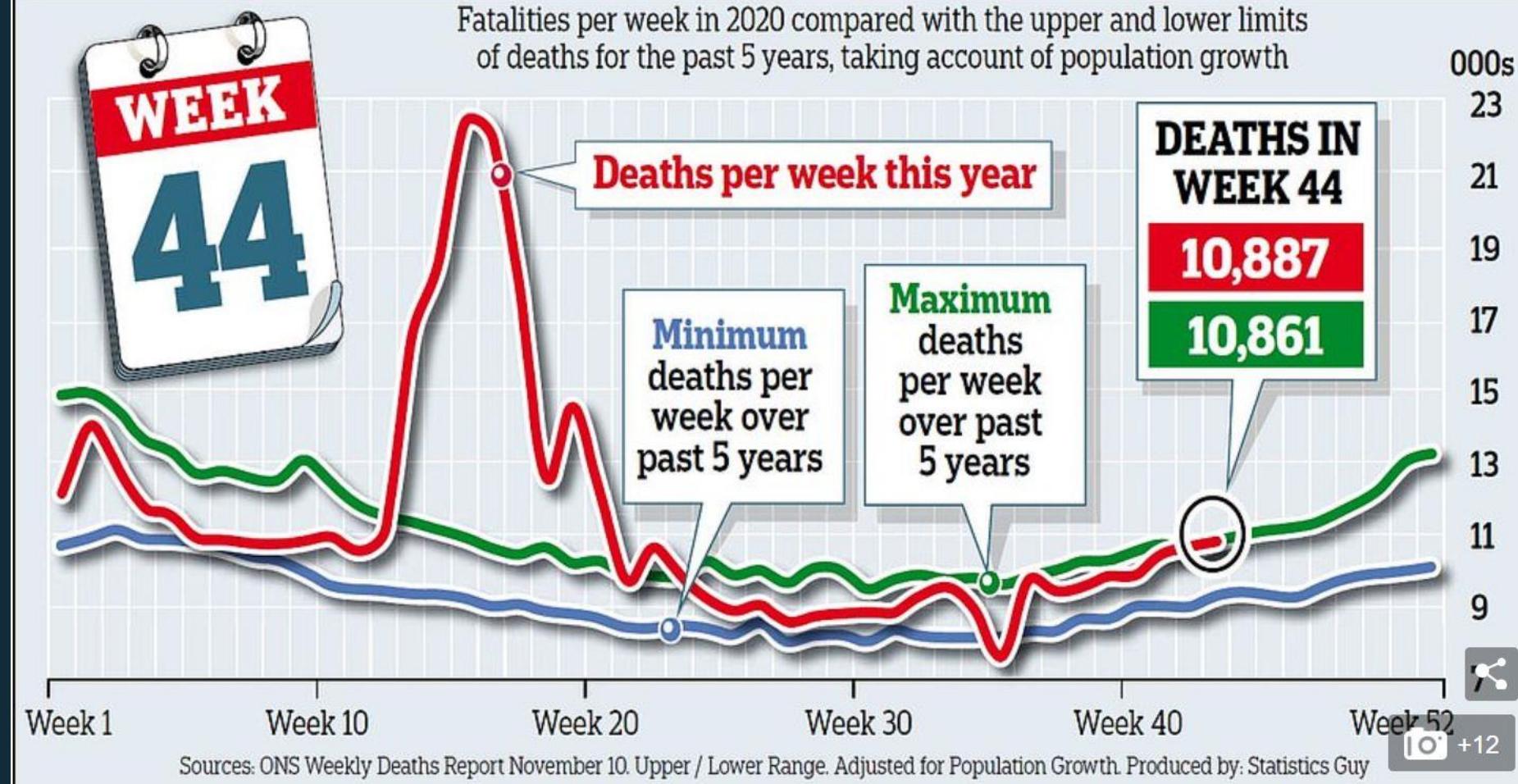


AN UPDATE ON JANUARY INFLATION STATISTICS



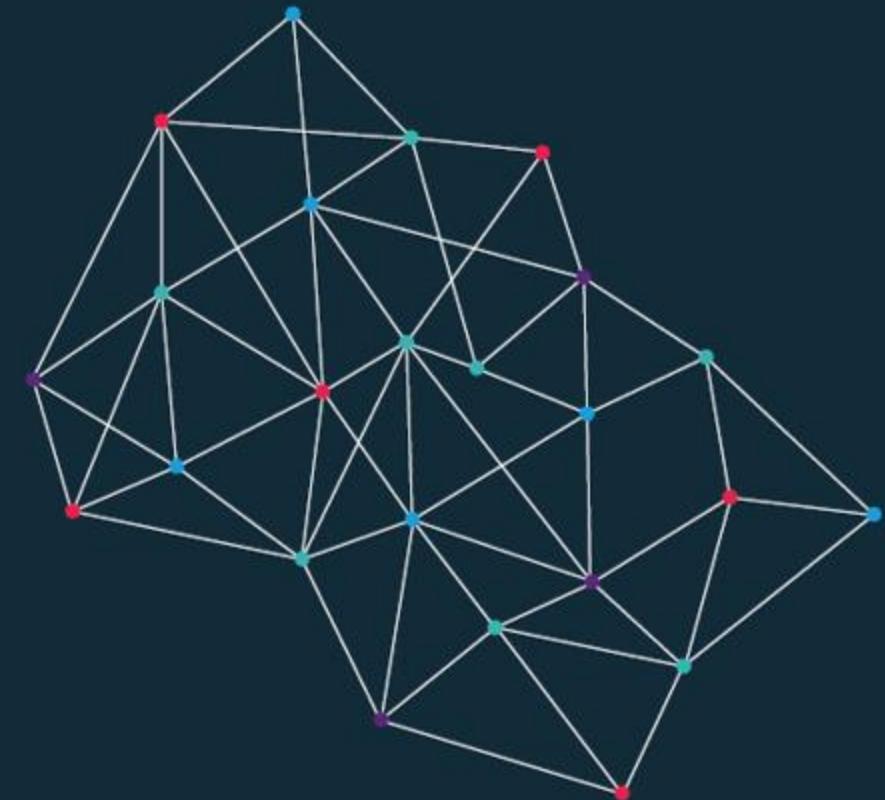
Source: Office for National Statistics

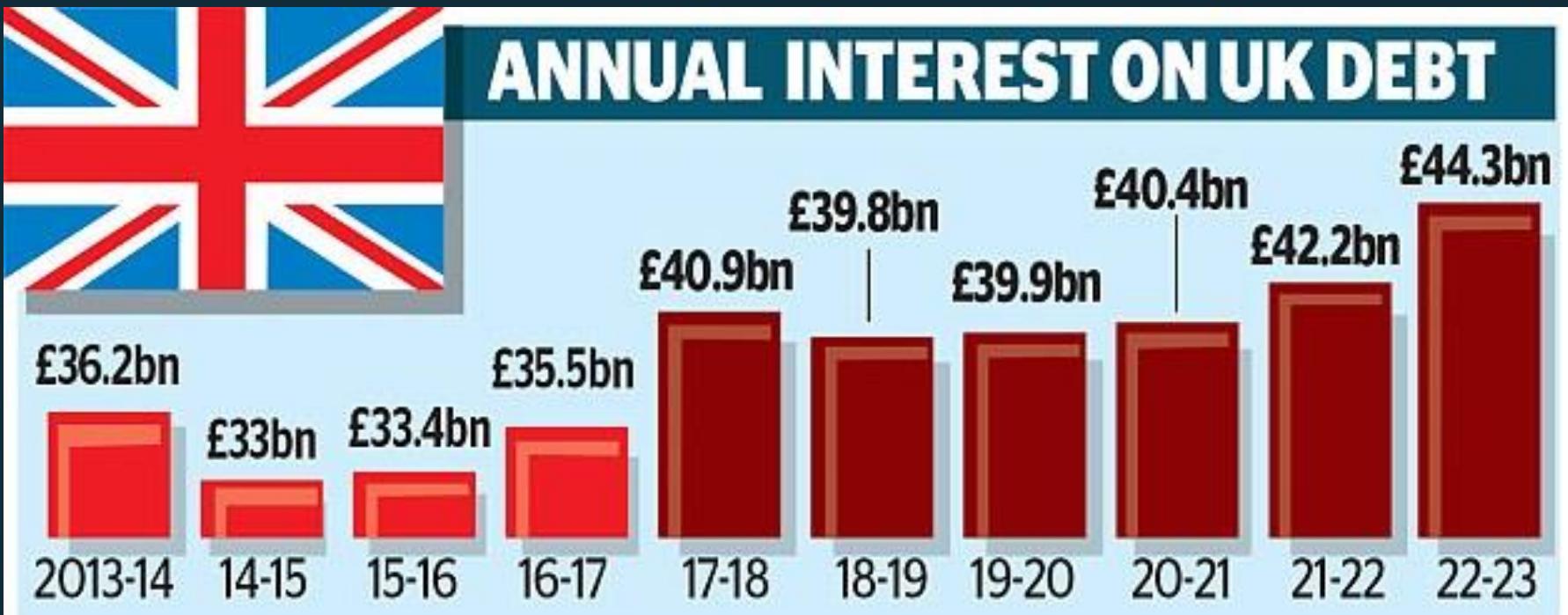
FATALITIES ARE BARELY ANY HIGHER



Source: Daily Mail, November 20th 2020

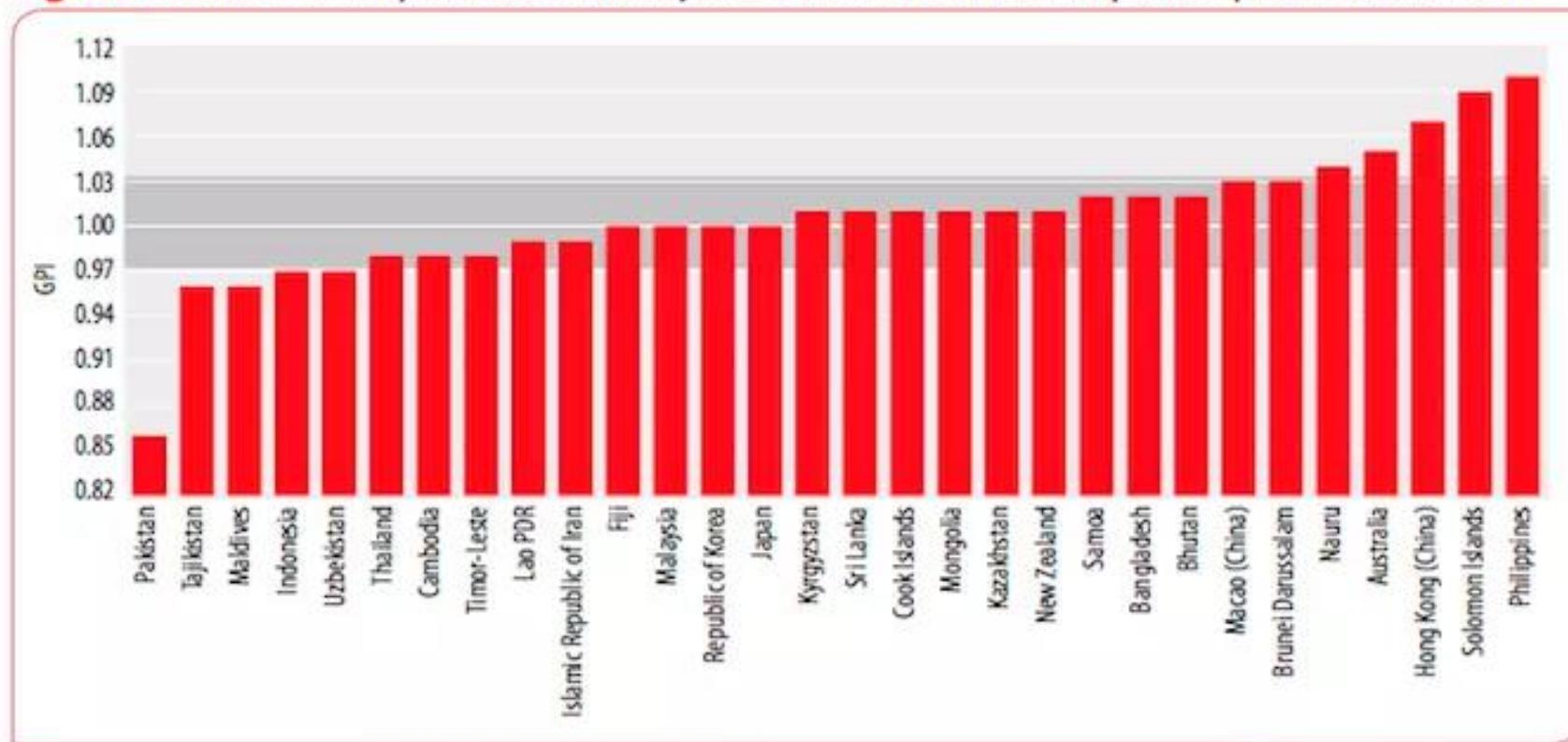
Two examples.





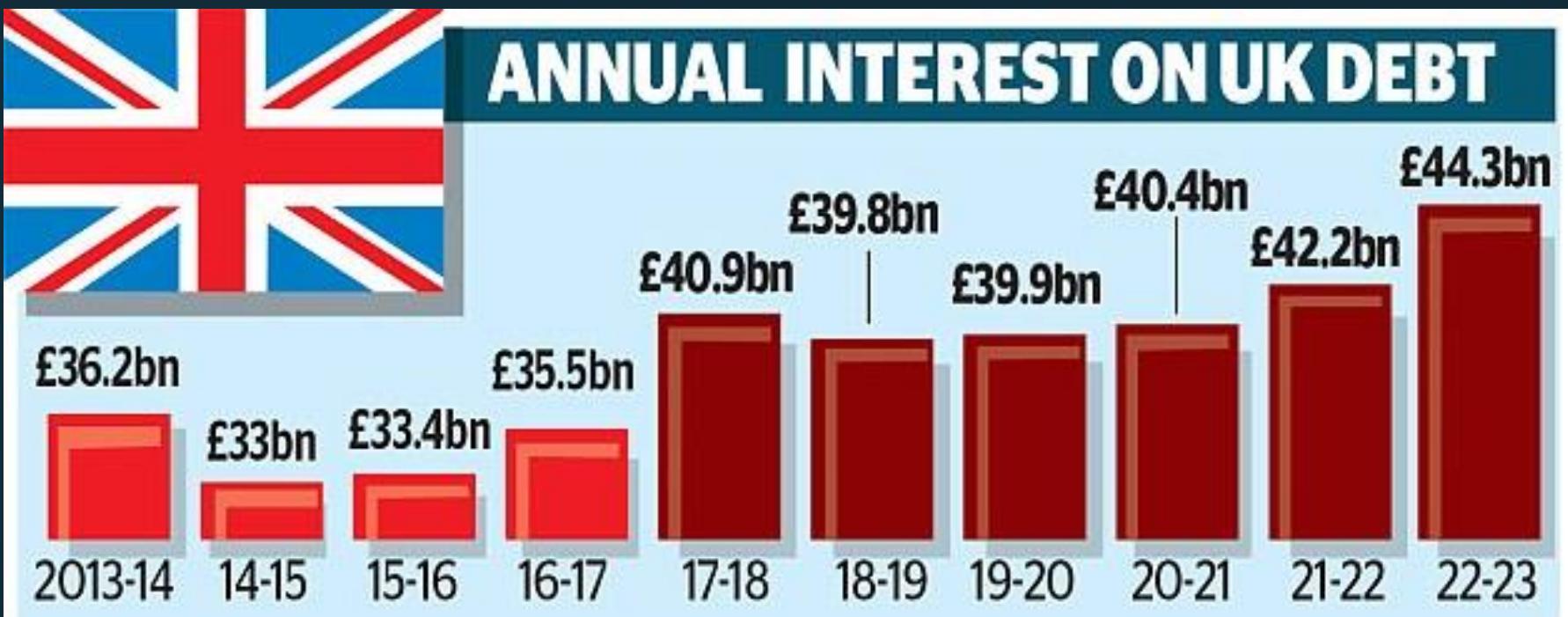
Source: Daily Mail, 2 Jan 2018

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

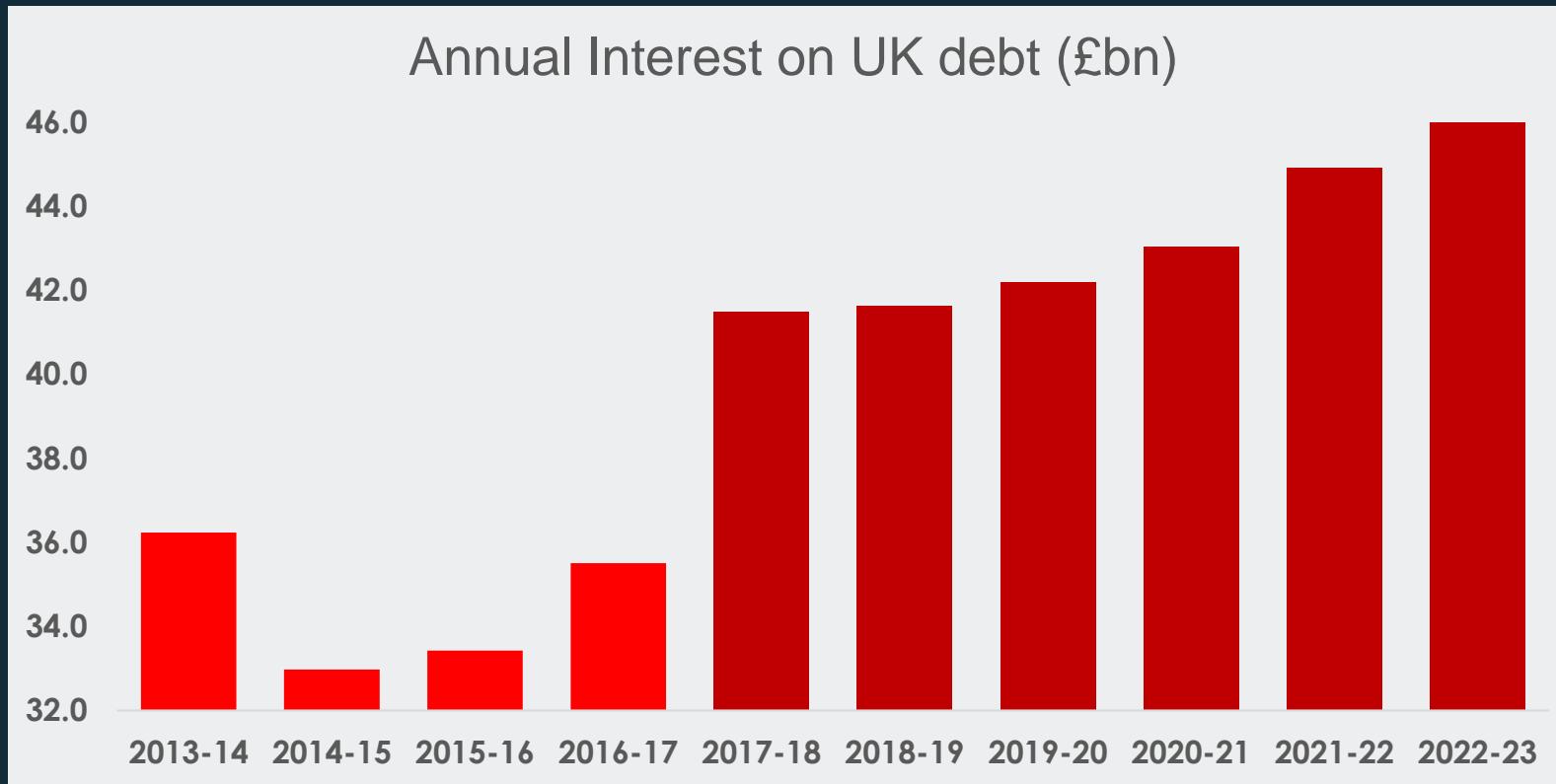


Source: UIS, 2011, Statistical Table 2.

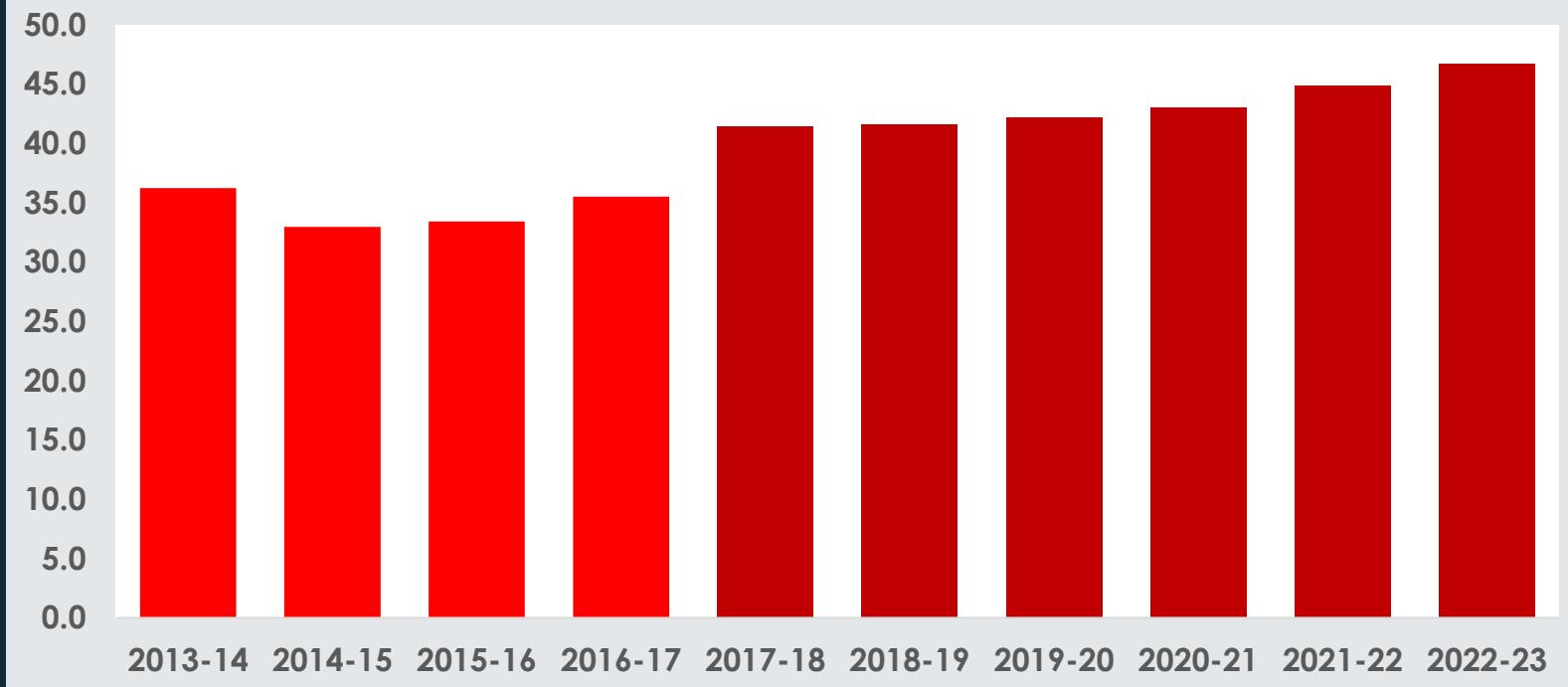
The same data: Different
stories



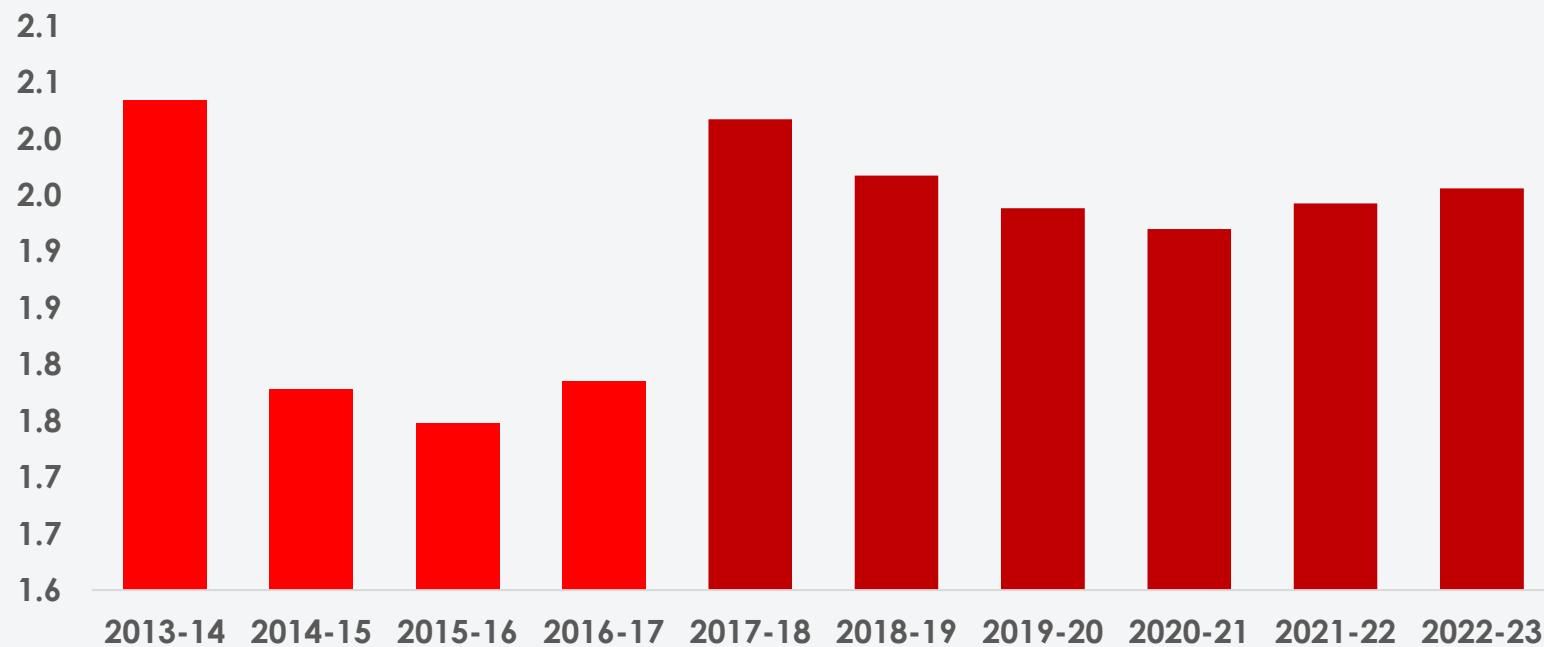
Source: Daily Mail, 2 Jan 2018

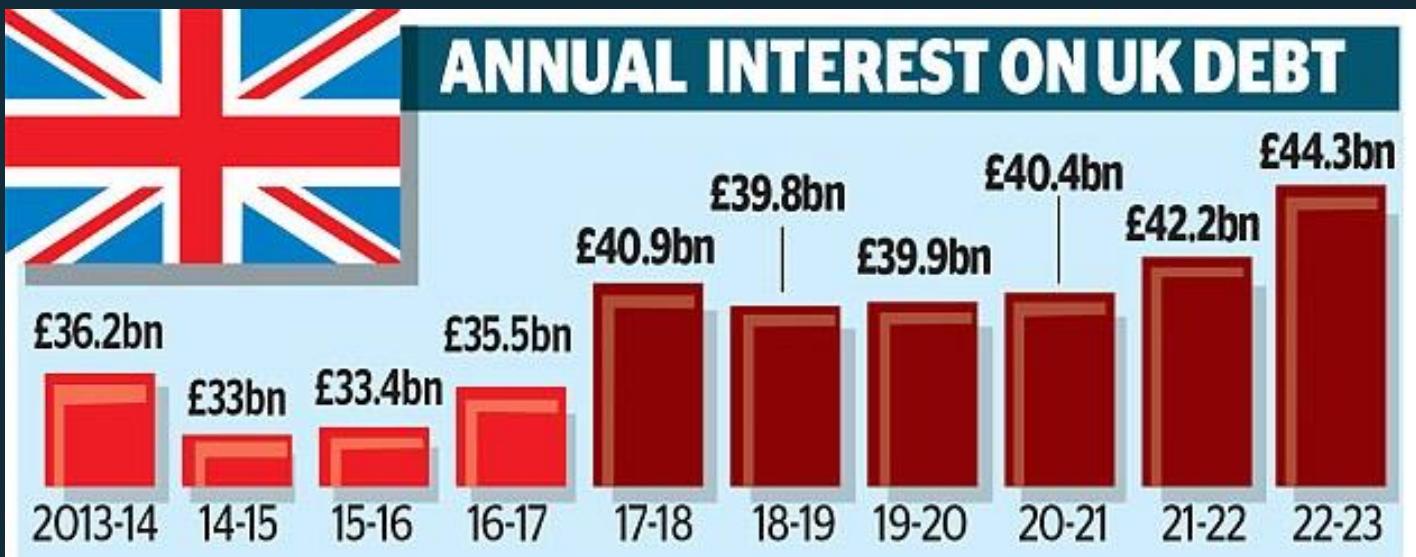


Annual Interest on UK debt (£bn)

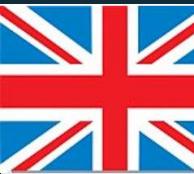


Debt interest
as % of GDP

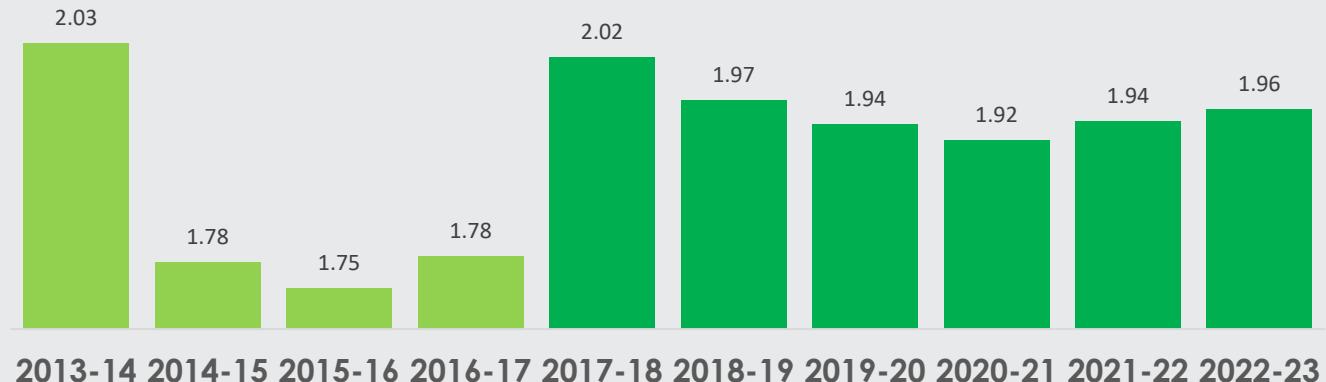




Source: Daily Mail, 2 Jan 2018



SHARE OF NATIONAL INCOME SPENT ON UK DEBT



Always 2% - the remarkable stability of the UK debt position

Interest payments on UK national debt as a % of GDP

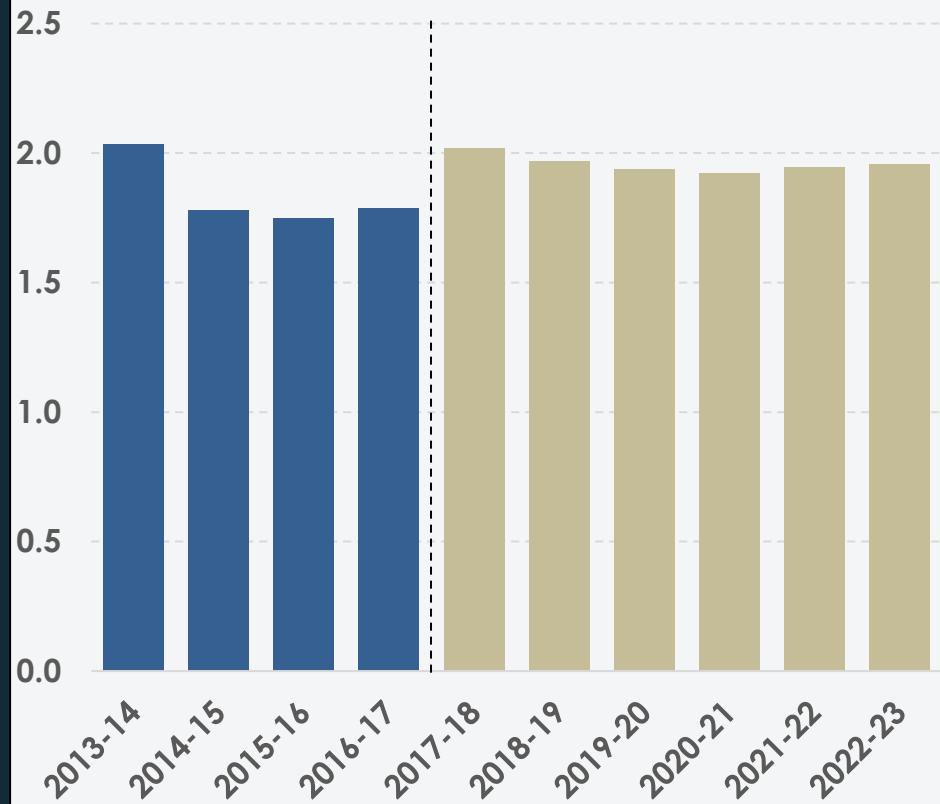
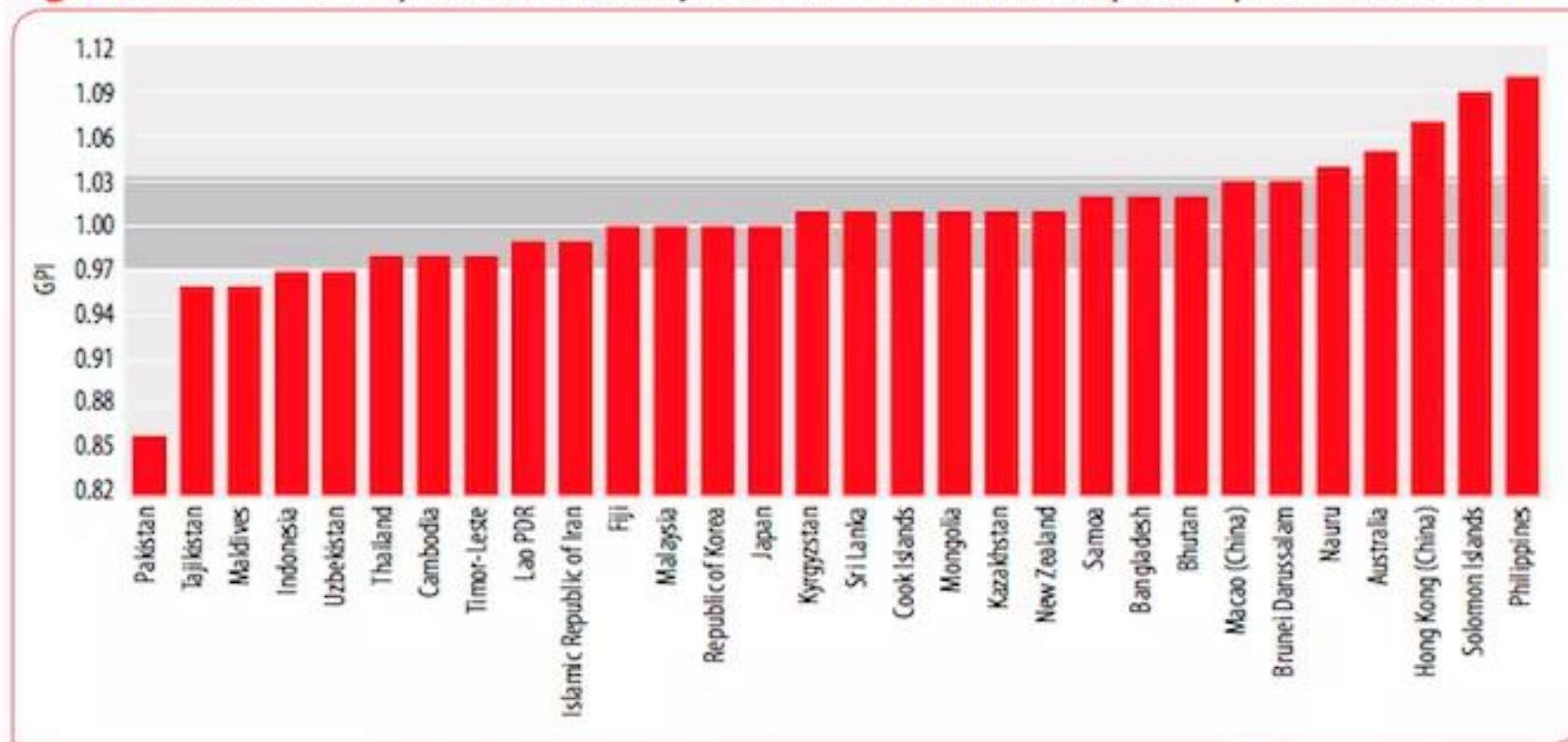
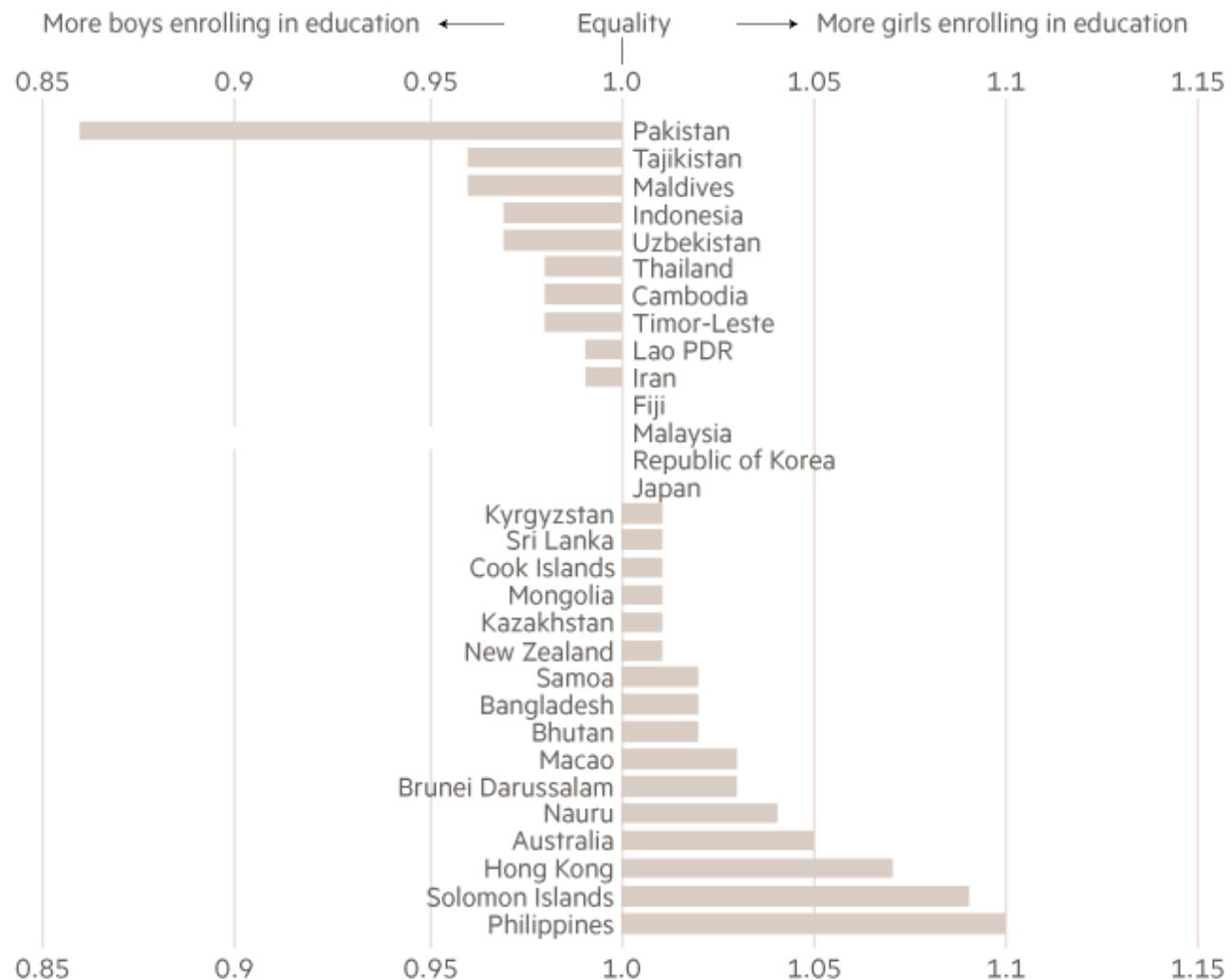


Figure 7: Gender Parity Index of the adjusted net intake rate in primary education, 2009



Source: UIS, 2011, Statistical Table 2.

Gender parity index*, 2009, Asia-Pacific region

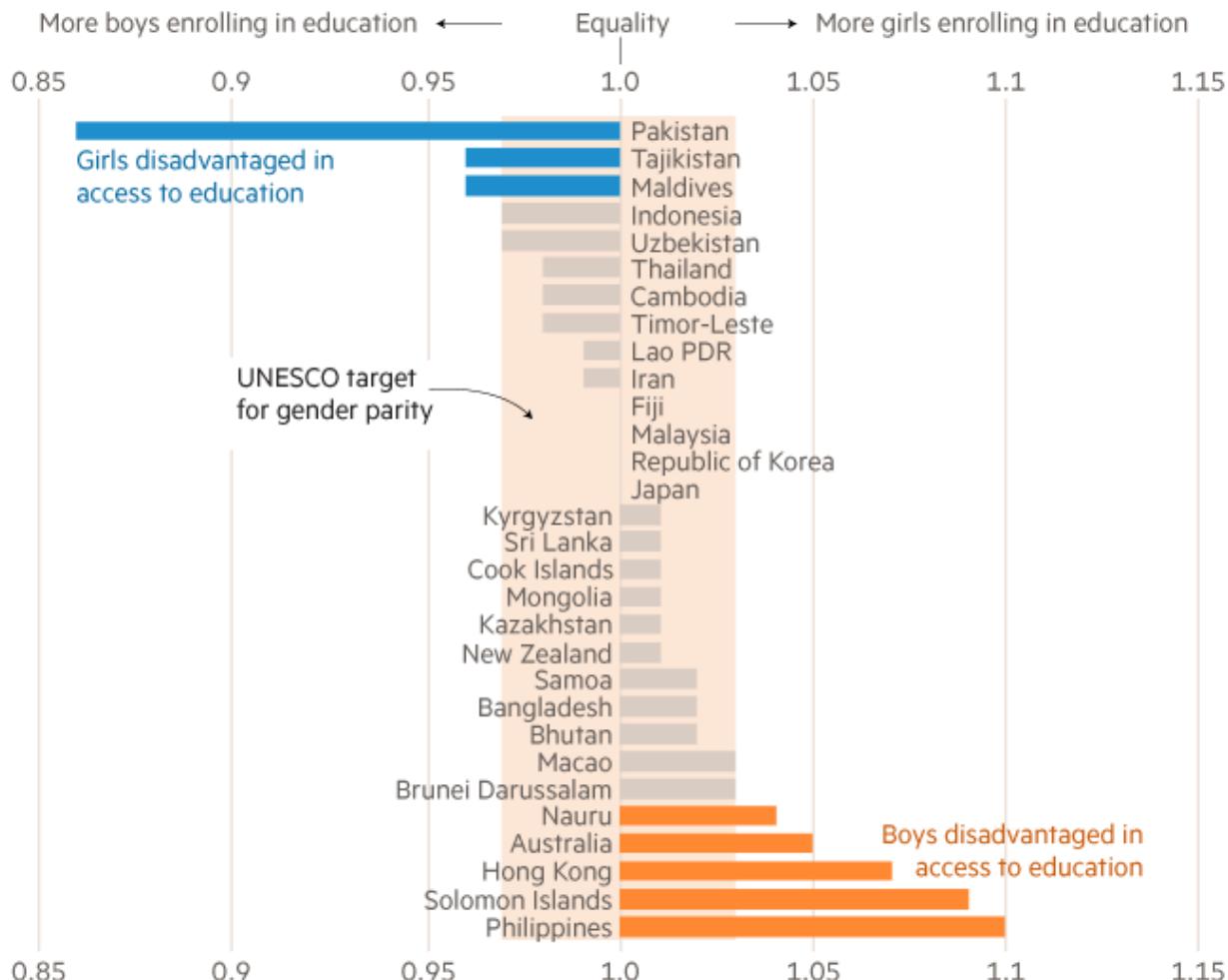


* Of adjusted net intake rate Source: UIS, 2011, Statistical table 2 Graphic by Alan Smith

© FT

Gender equality in primary education is not yet a reality for all

Gender parity index*, 2009, Asia-Pacific region



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

© FT

Chartjunk: a gallery.

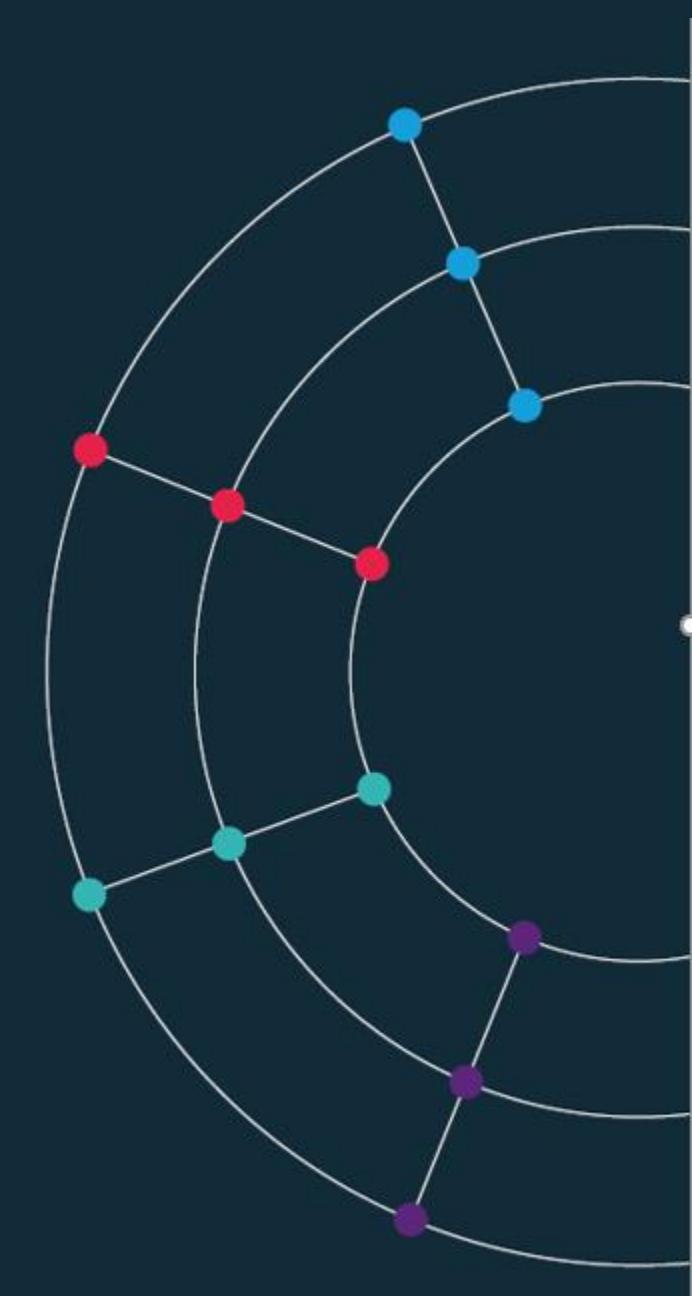
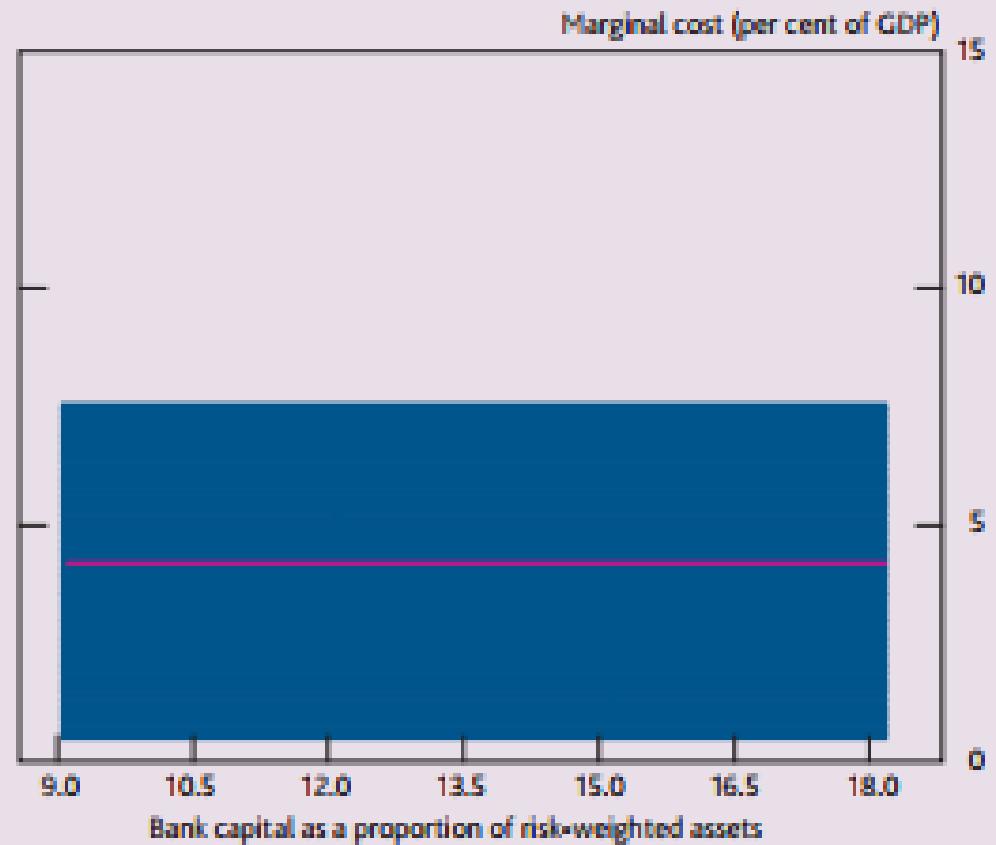
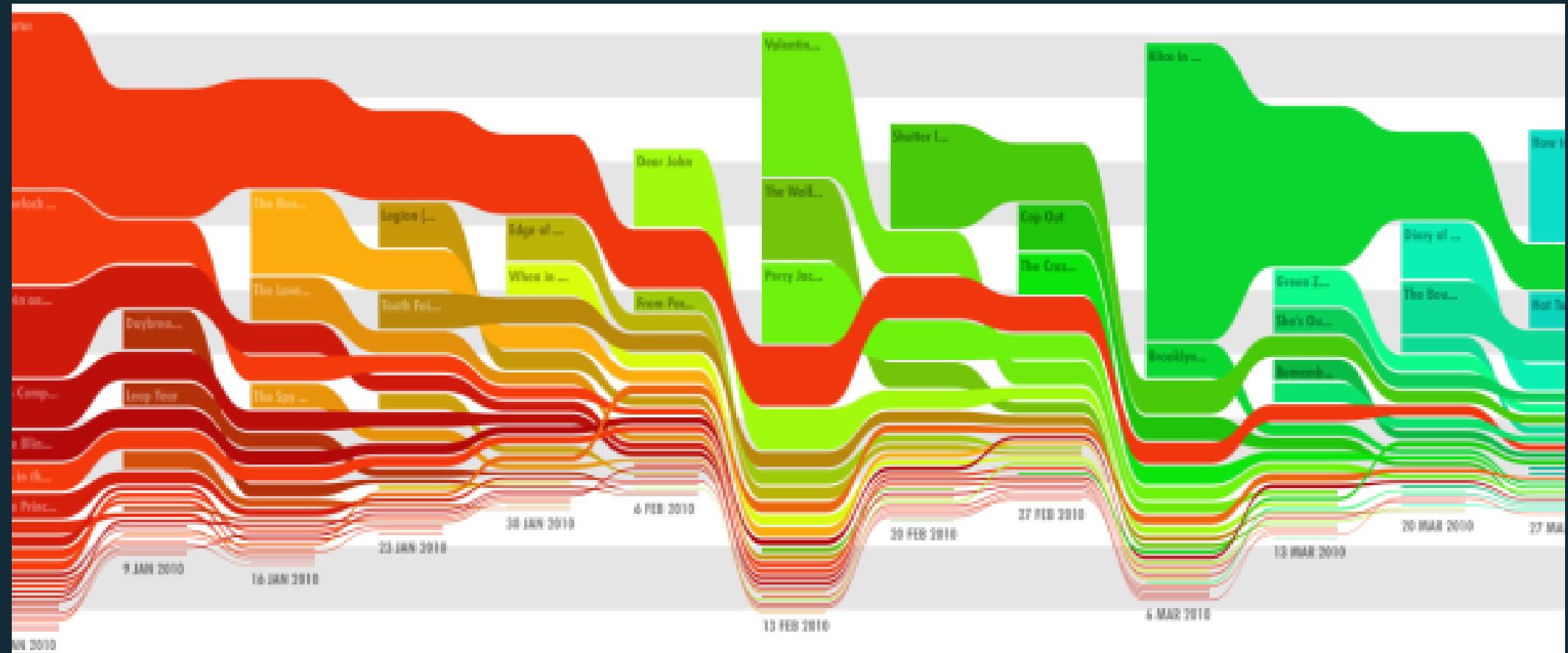


Chart A Marginal cost from higher capital levels



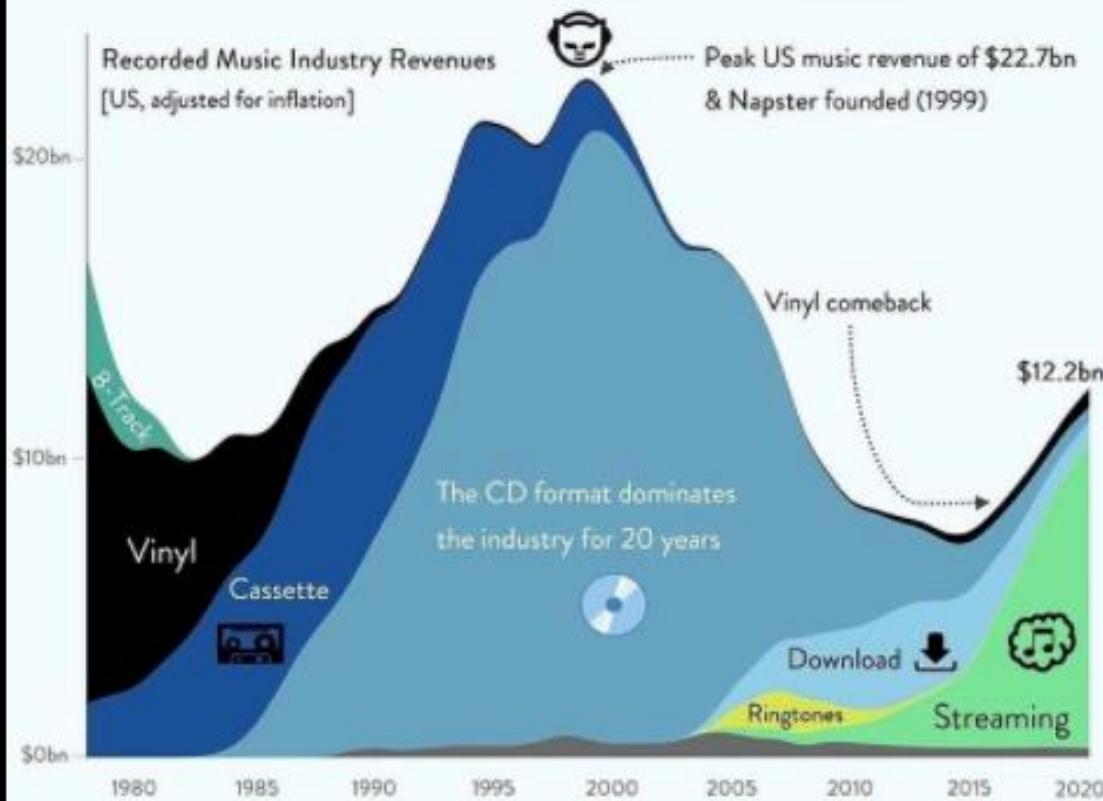
Source: Bank calculations.





...

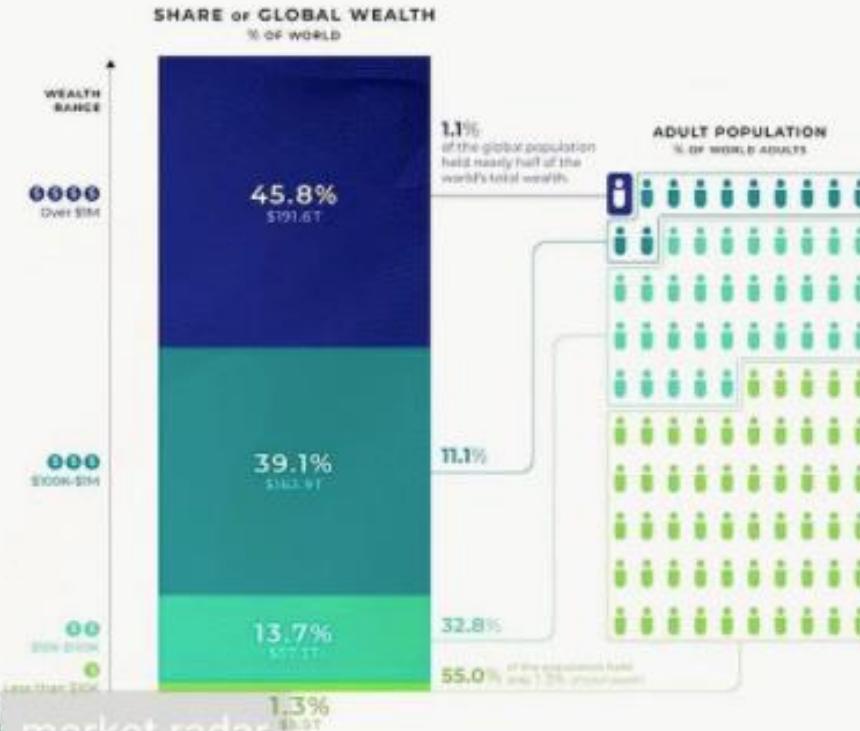
Here Comes The Sun: Streaming Revives The Music Industry



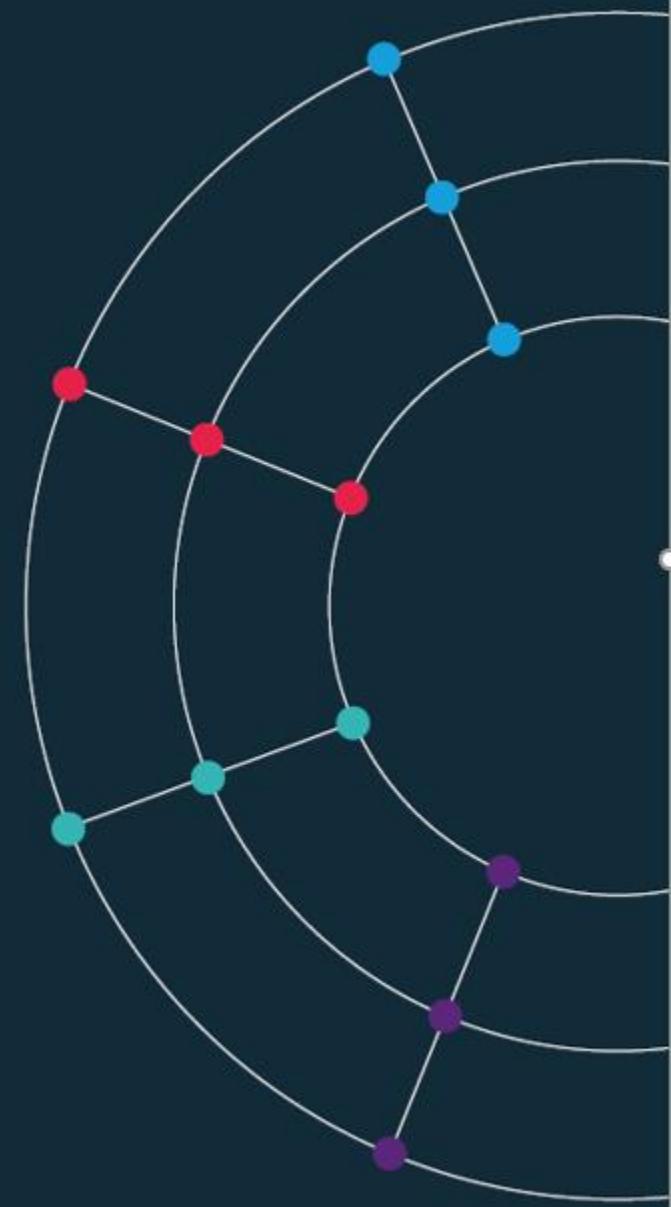
...

1% of adult population holds 46% of global wealth

THE GLOBAL WEALTH DISTRIBUTION



Dataviz heroes.



William Playfair.

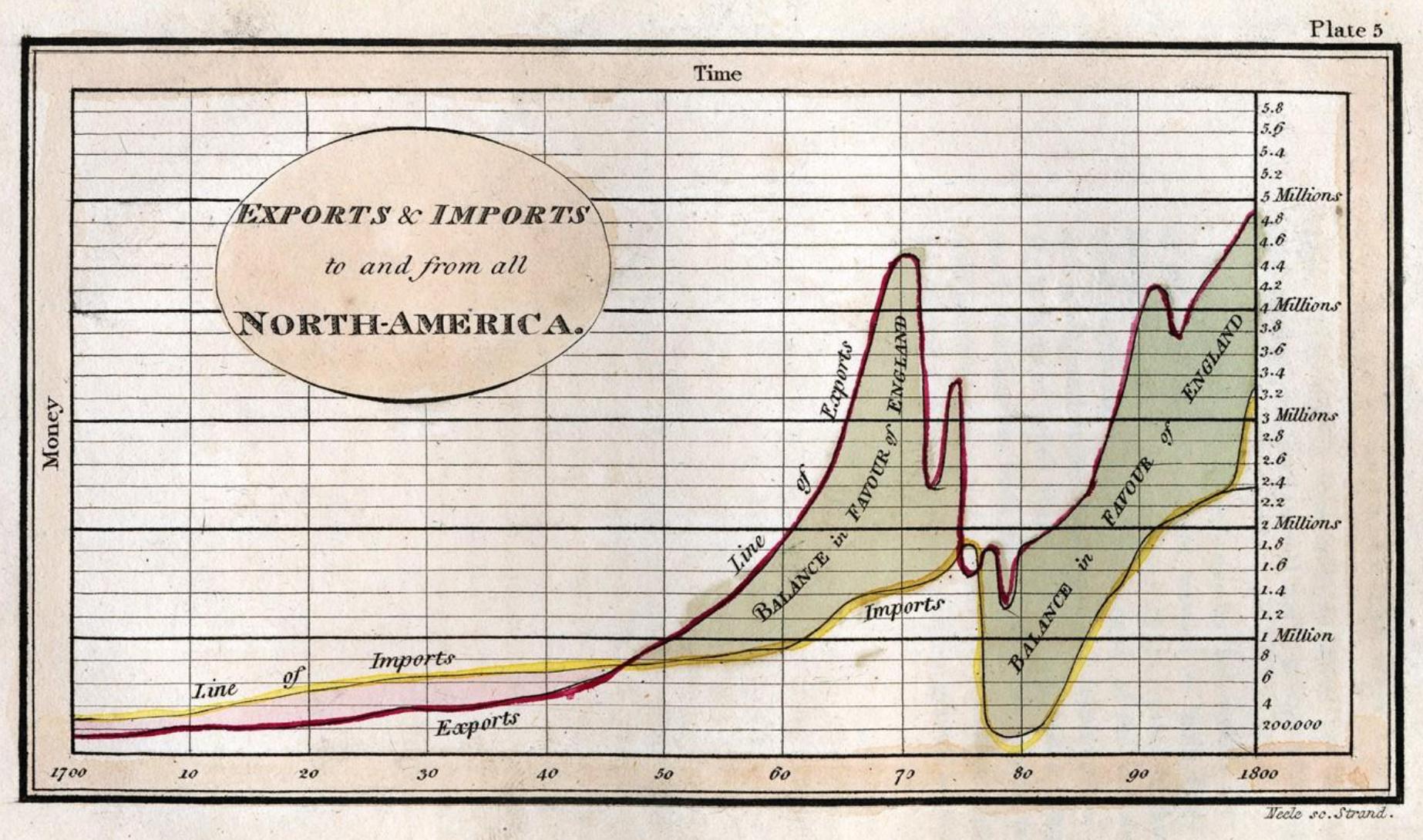
- Born in 1759, William Playfair invented the chart.
- *The Commercial and Political Atlas* (1785) which featured 40 charts covering trade, debt, expenditure and other economic matters
- Playfair was the first to add the vertical axis (previous charts did not have this feature, meaning the way data were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



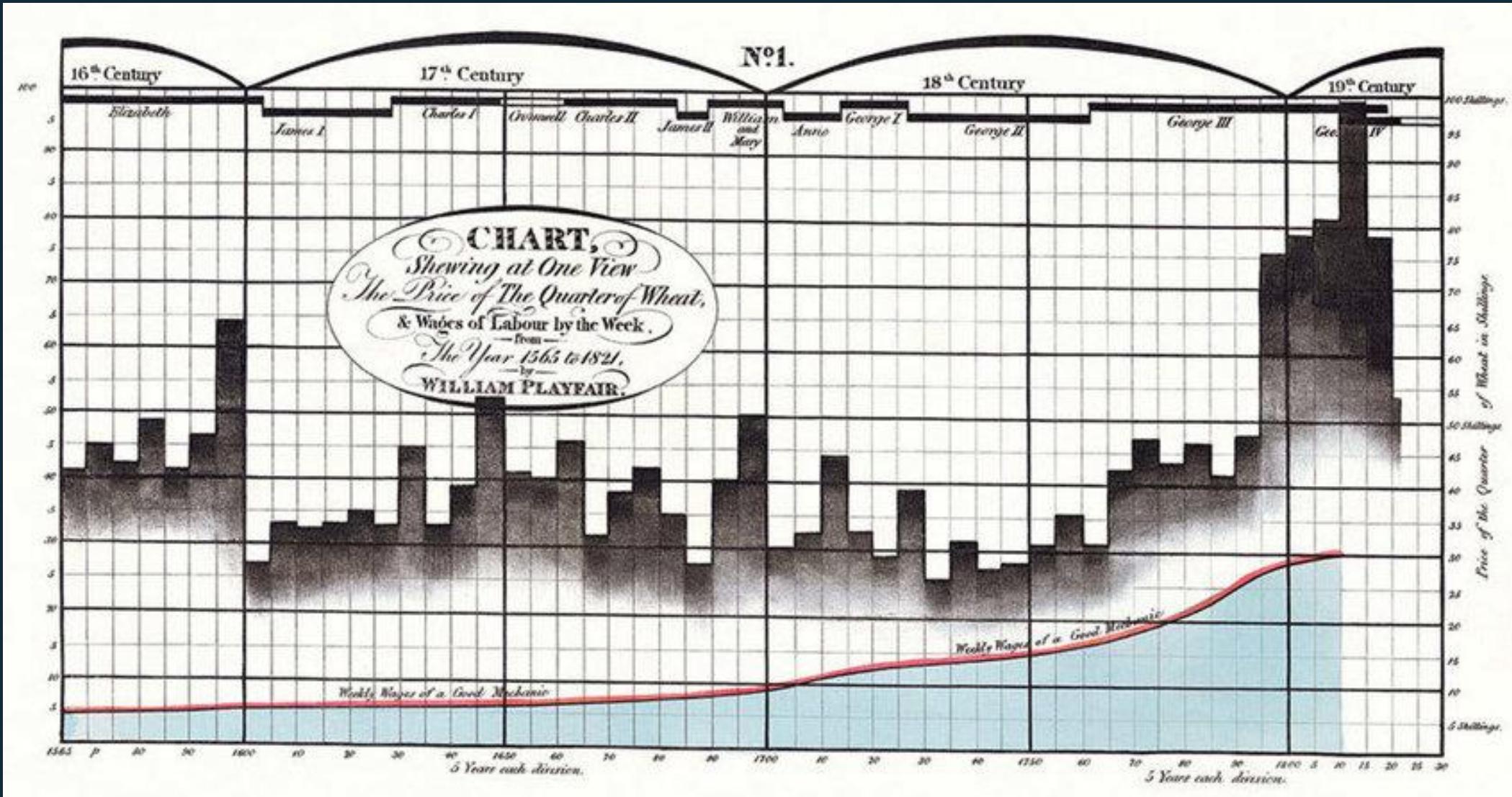
"Information, that is imperfectly acquired, is generally as imperfectly retained; and a man who has carefully investigated a printed table, finds, when done, that he has only a very faint and partial idea of what he has read; and that like a figure imprinted on sand, is soon totally erased and defaced."

Politics and trade.

Plate 5

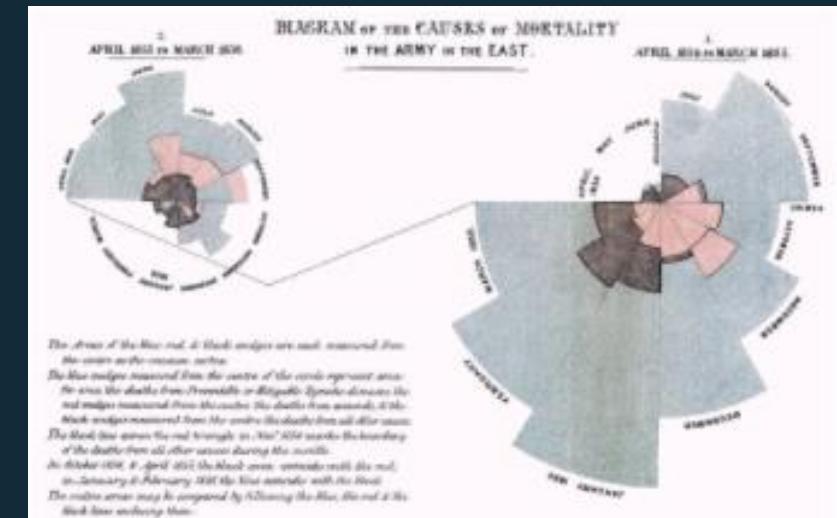


Real wages.



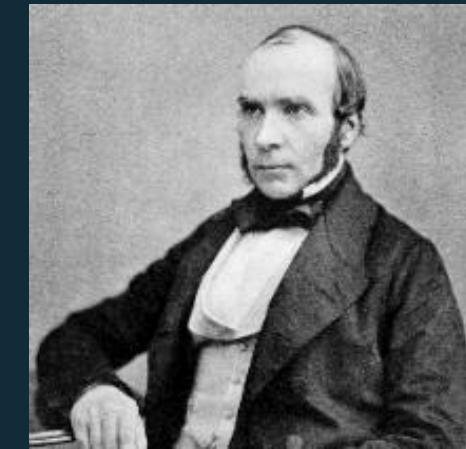
Florence Nightingale.

- 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 he disliked most coined a new word: “chartjunk”.



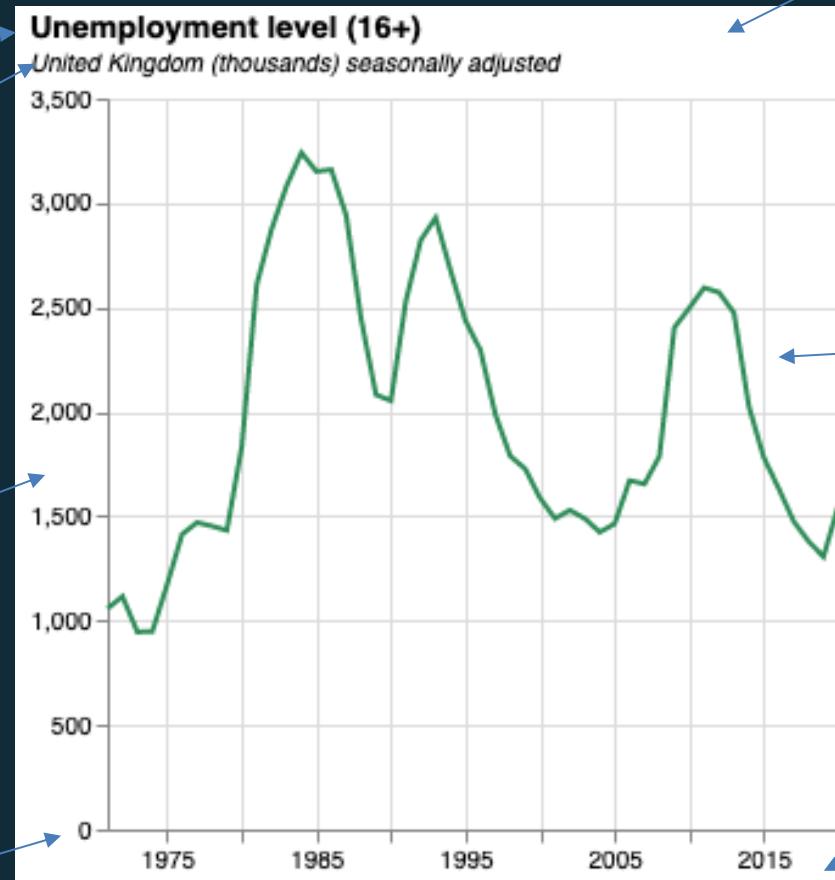
Anatomy of a good chart.

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

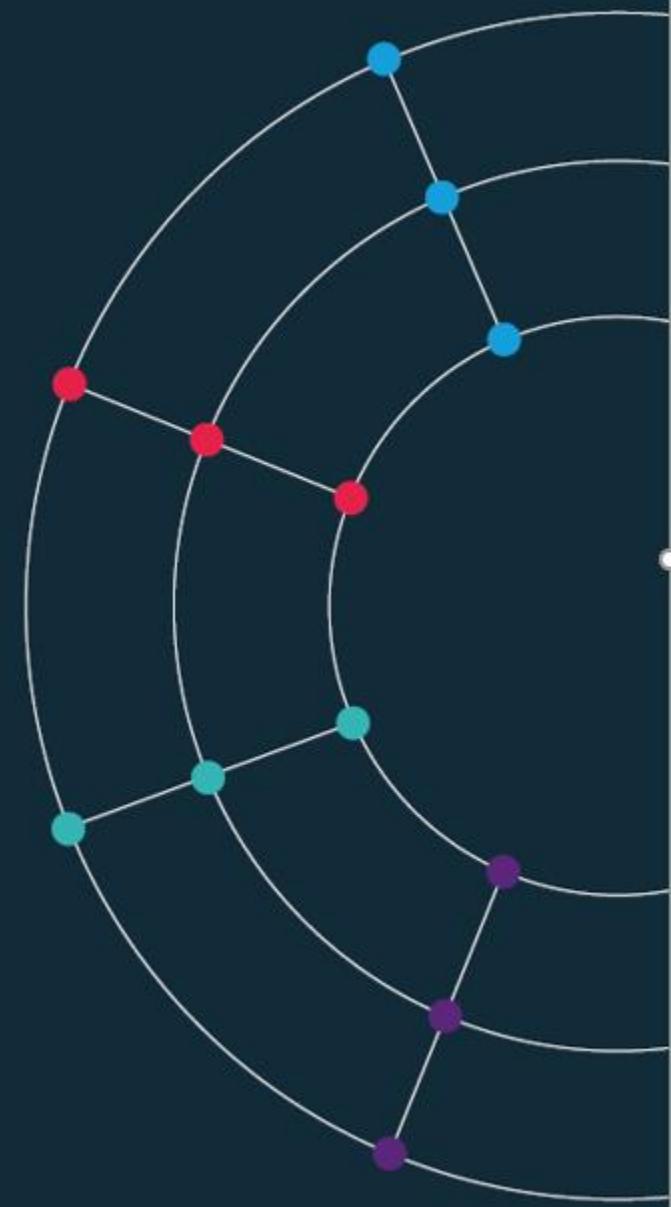


Size. The original version of this chart is 300px by 300px. This is perfect for mobile.

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

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

A design classic.



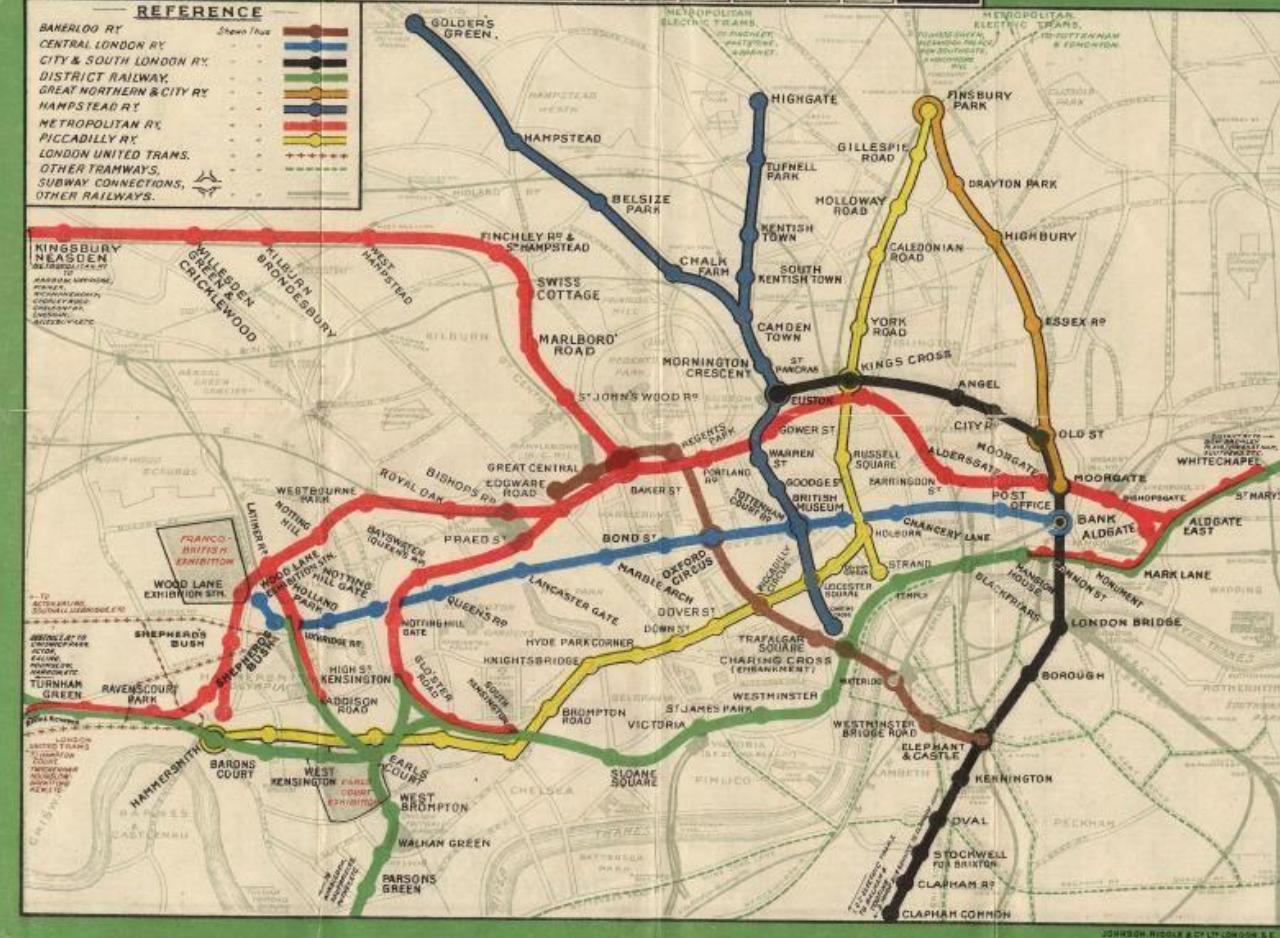
LONDON

UNDERGROUND

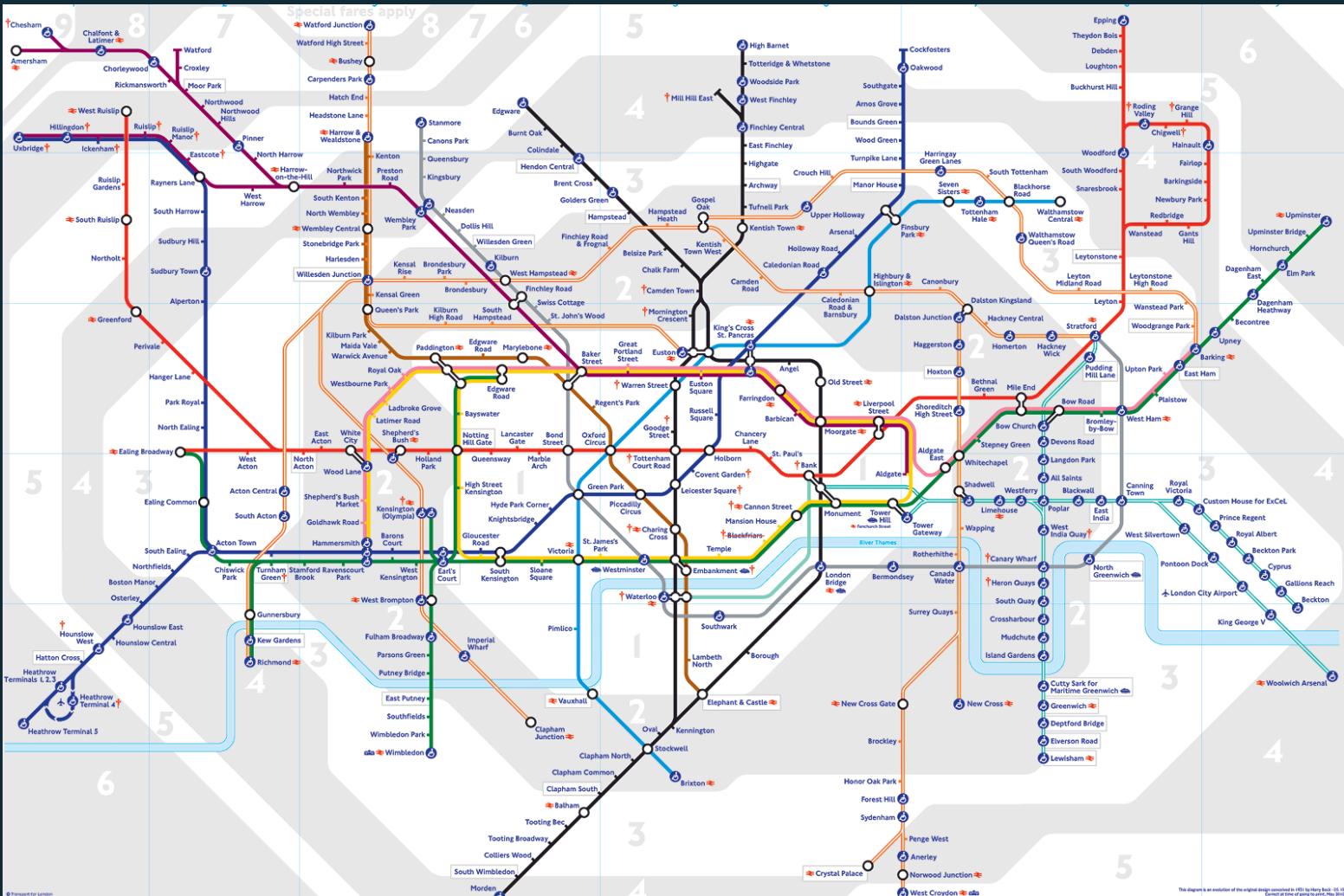
RAILWAYS

— REFERENCE

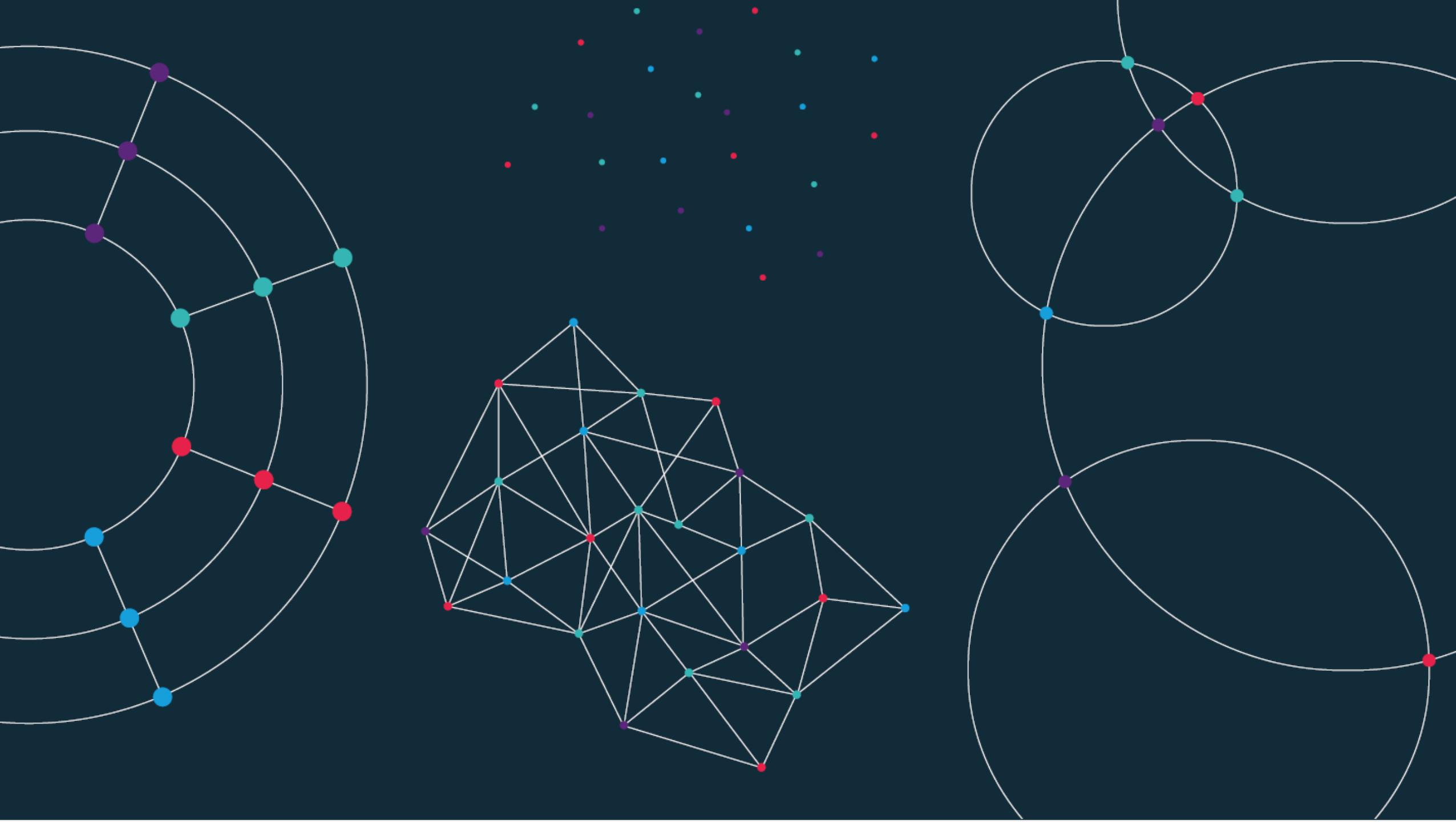
- BAMERLOO RY.
CENTRAL LONDON RY.
CITY & SOUTH LONDON RY.
DISTRICT RAILWAY.
GREAT NORTHERN & CITY RY.
HAMPSTEAD RY.
METROPOLITAN RY.
PICCADILLY RY.
LONDON UNITED TRAMS.
OTHER TRAMWAYS.
SUBWAY CONNECTIONS,
OTHER RAILWAYS.







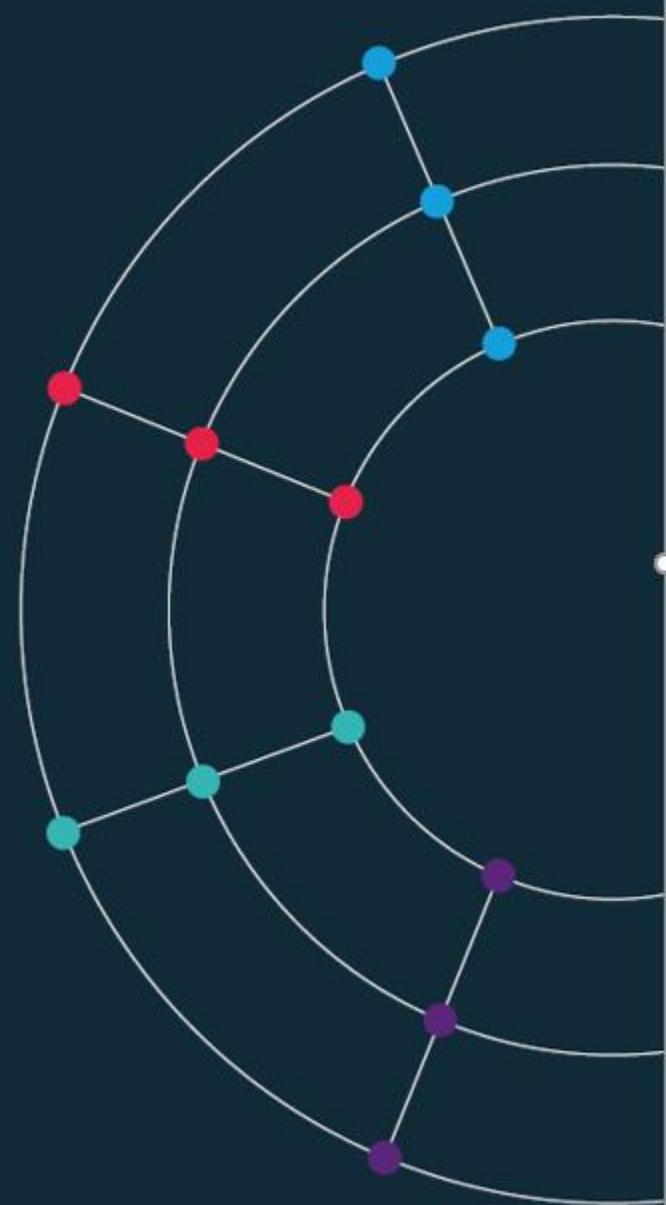




Practical 3.

From data to a chart

1. Finding the dataset
2. Transforming the series
3. Uploading the data
4. Adapting a Vega-lite example
5. Presenting your chart



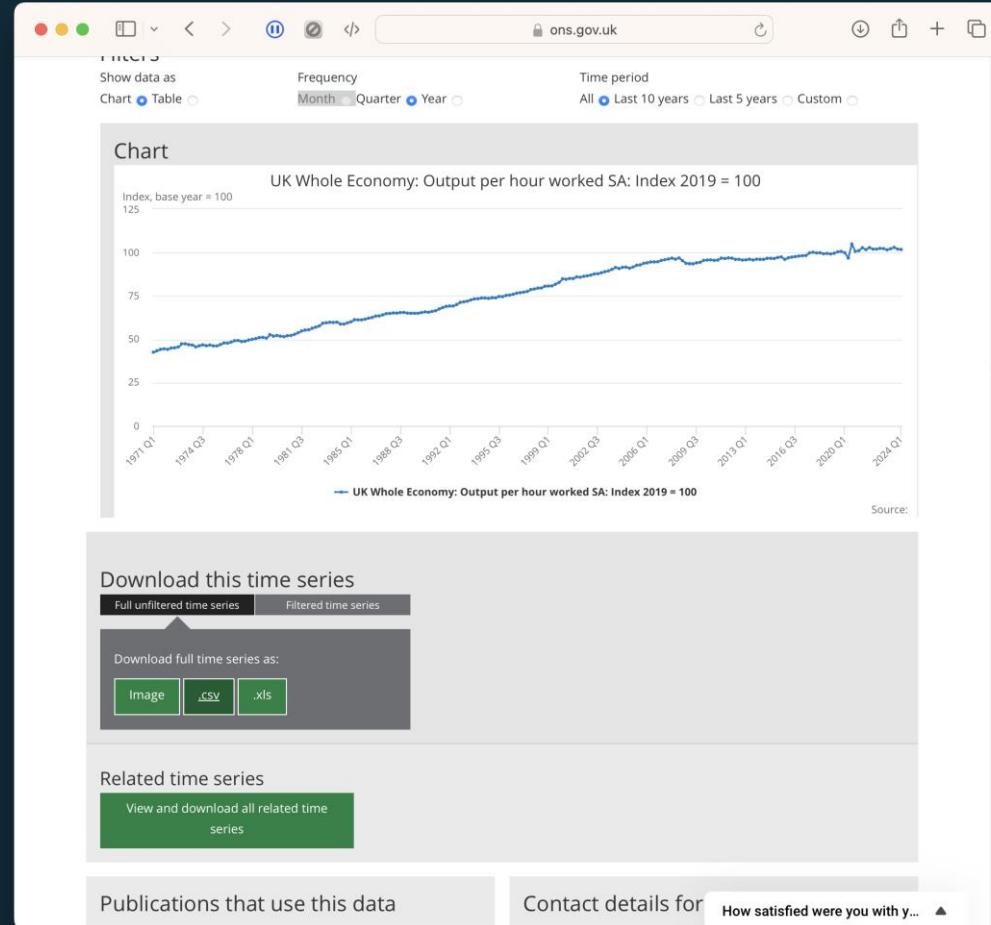
Finding data.

The screenshot shows the ONS website with a dark blue header. The top navigation bar includes links for Home, Business, industry and trade, Economy, Employment and labour market, People, population and community, and Taking part in a survey? A search bar is located below the navigation. A purple banner at the top says "census 2021 Data and analysis from Census 2021". Below the banner, the breadcrumb navigation shows: Home > Employment and labour market > People in work > Labour productivity. The main content area has a title "Labour productivity" and a subtitle "Efficiency of the UK workforce, including output per worker, per job and per hour. Data are available by industry and by region." It features two "Spotlight" cards: "Productivity flash estimate and overview" and "UK productivity flash".

Let's chart UK productivity.
This data is available from the ONS.

Finding data.

We need to download the CSV.

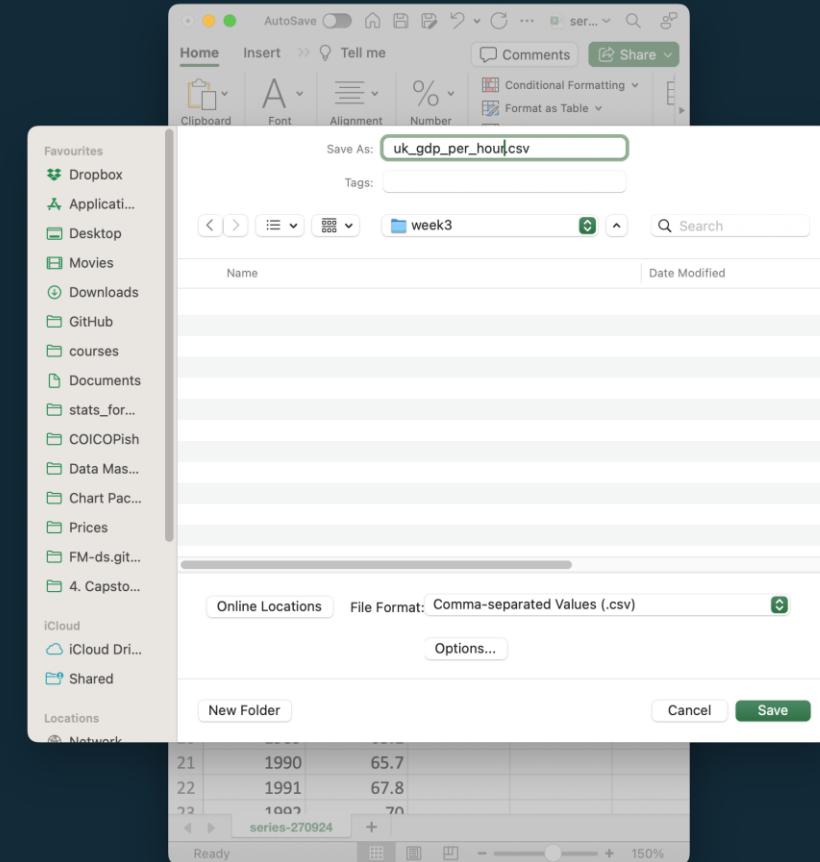


Transforming data.

Standardising column names.

| | | |
|----|-----------------|--|
| 1 | Title | UK Whole Economy: Output per hour worked SA: Index |
| 2 | CDID | LZVB |
| 3 | Source data | PRDY |
| 4 | PreUnit | |
| 5 | Unit | Index, base year = 100 |
| 6 | Release date | 15/08/2024 |
| 7 | Next release | 15-Nov-24 |
| 8 | Important notes | |
| 9 | 1971 | 43.7 |
| 10 | 1972 | 45 |
| 11 | 1973 | 47.1 |
| 12 | 1974 | 46.2 |
| 13 | 1975 | 46.5 |
| 14 | 1976 | 48.3 |
| 15 | 1977 | 49.1 |
| 16 | 1978 | 50.6 |
| 17 | 1979 | 51.9 |
| 18 | 1980 | 51.9 |
| 19 | 1981 | 54.2 |
| 20 | 1982 | 56.6 |
| 21 | 1983 | 59.6 |
| 22 | 1984 | 59.2 |
| 23 | 100% | 60.0 |

| | | |
|----|------|-------|
| 1 | date | value |
| 2 | 1971 | 43.7 |
| 3 | 1972 | 45 |
| 4 | 1973 | 47.1 |
| 5 | 1974 | 46.2 |
| 6 | 1975 | 46.5 |
| 7 | 1976 | 48.3 |
| 8 | 1977 | 49.1 |
| 9 | 1978 | 50.6 |
| 10 | 1979 | 51.9 |
| 11 | 1980 | 51.9 |
| 12 | 1981 | 54.2 |
| 13 | 1982 | 56.6 |
| 14 | 1983 | 59.6 |
| 15 | 1984 | 59.2 |
| 16 | 1985 | 60.9 |
| 17 | 1986 | 62.5 |
| 18 | 1987 | 64.3 |
| 19 | 1988 | 65.3 |
| 20 | 1989 | 65.1 |
| 21 | 1990 | 65.7 |
| 22 | 1991 | 67.8 |
| 23 | 100% | 70 |



And saving our output.

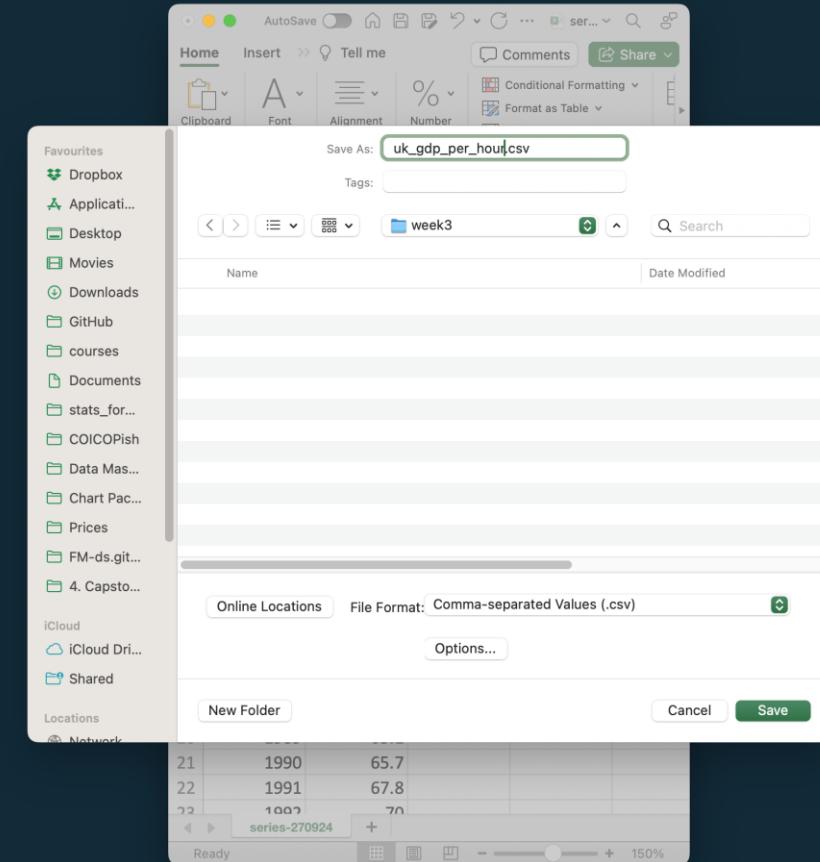
Is there a better way of doing this?

Transforming data.

Standardising column names.

| | | |
|----|-----------------|--|
| 1 | Title | UK Whole Economy: Output per hour worked SA: Index |
| 2 | CDID | LZVB |
| 3 | Source data | PRDY |
| 4 | PreUnit | |
| 5 | Unit | Index, base year = 100 |
| 6 | Release date | 15/08/2024 |
| 7 | Next release | 15-Nov-24 |
| 8 | Important notes | |
| 9 | 1971 | 43.7 |
| 10 | 1972 | 45 |
| 11 | 1973 | 47.1 |
| 12 | 1974 | 46.2 |
| 13 | 1975 | 46.5 |
| 14 | 1976 | 48.3 |
| 15 | 1977 | 49.1 |
| 16 | 1978 | 50.6 |
| 17 | 1979 | 51.9 |
| 18 | 1980 | 51.9 |
| 19 | 1981 | 54.2 |
| 20 | 1982 | 56.6 |
| 21 | 1983 | 59.6 |
| 22 | 1984 | 59.2 |
| 23 | 1985 | 60.0 |

| | | |
|----|------|-------|
| 1 | date | value |
| 2 | 1971 | 43.7 |
| 3 | 1972 | 45 |
| 4 | 1973 | 47.1 |
| 5 | 1974 | 46.2 |
| 6 | 1975 | 46.5 |
| 7 | 1976 | 48.3 |
| 8 | 1977 | 49.1 |
| 9 | 1978 | 50.6 |
| 10 | 1979 | 51.9 |
| 11 | 1980 | 51.9 |
| 12 | 1981 | 54.2 |
| 13 | 1982 | 56.6 |
| 14 | 1983 | 59.6 |
| 15 | 1984 | 59.2 |
| 16 | 1985 | 60.9 |
| 17 | 1986 | 62.5 |
| 18 | 1987 | 64.3 |
| 19 | 1988 | 65.3 |
| 20 | 1989 | 65.1 |
| 21 | 1990 | 65.7 |
| 22 | 1991 | 67.8 |
| 23 | 1992 | 70 |



And saving our output.

Is there a better way of doing this?

Uploading data.

Uploading the cleaned data.

A screenshot of a GitHub repository page for "mclass-user/mclass-user.github.io". The main navigation bar at the top includes "Code", "Issues", "Pull requests", "Actions", "Projects", "Wiki", and "Security". Below this, a dropdown menu shows "main" selected. A message "File successfully deleted." is displayed above a list of files. The list includes:

| Name | Last commit message | Last commit date |
|-----------------|----------------------|------------------|
| README.md | Initial commit | last week |
| index.css | Starting the project | last week |
| index.html | Add files via upload | last week |
| wk1_chart1.json | Starting the project | last week |
| wk1_chart2.json | Starting the project | last week |

Below the table is a section for "README.md". At the bottom of the page, there is a "mclass-user.github.io" section with a text input field containing "https://github.com/mclass-user/mclass-user.github.io/upload/main". A context menu is open over the "Upload files" button, showing options: "+ Create new file" and "Upload files".

A screenshot of the GitHub "Upload files" dialog for the same repository. The dialog has a header "Upload files - mclass-user/mc..." and a sub-header "mclass-user.github.io /". It features a large text area with a placeholder "Drag additional files here to add them to your repository" and a "choose your files" link. Below this is a list of files: "uk_gdp_per_hour.csv". A "Commit changes" section follows, containing fields for "Add files via upload" and "Add an optional extended description...". At the bottom, two radio buttons are shown: one selected for "Commit directly to the main branch" and another for "Create a new branch for this commit and start a pull request". Buttons for "Commit changes" and "Cancel" are at the very bottom.

Uploading data.

We need the ‘Raw’ URL for our data.

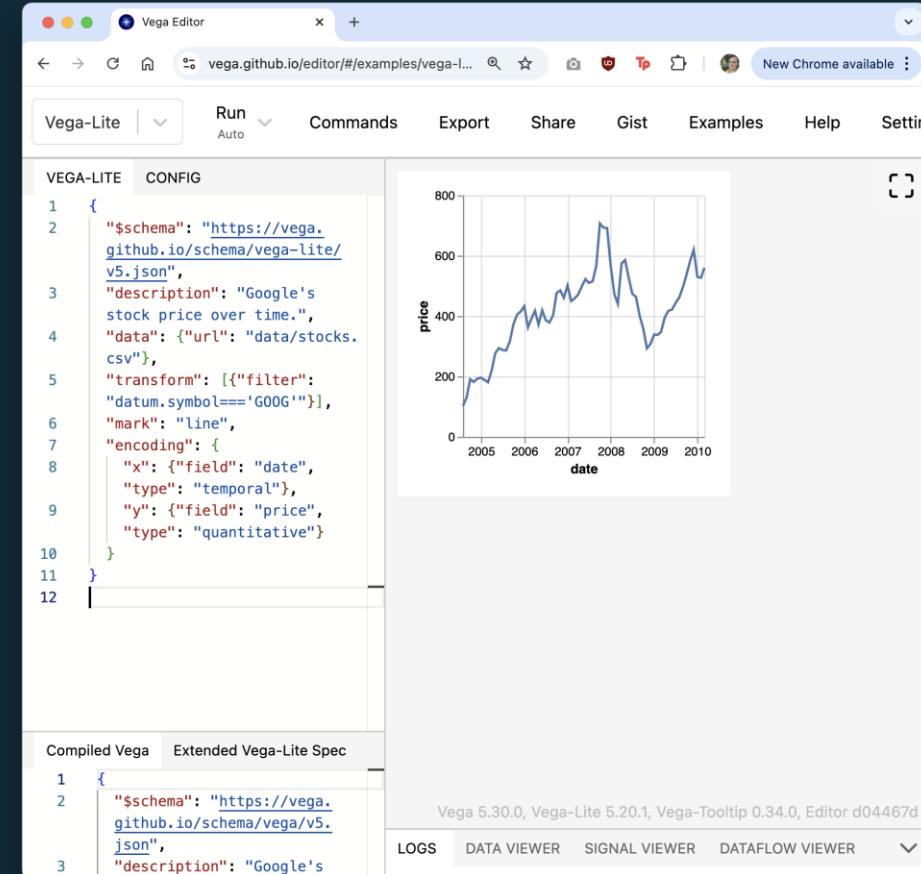
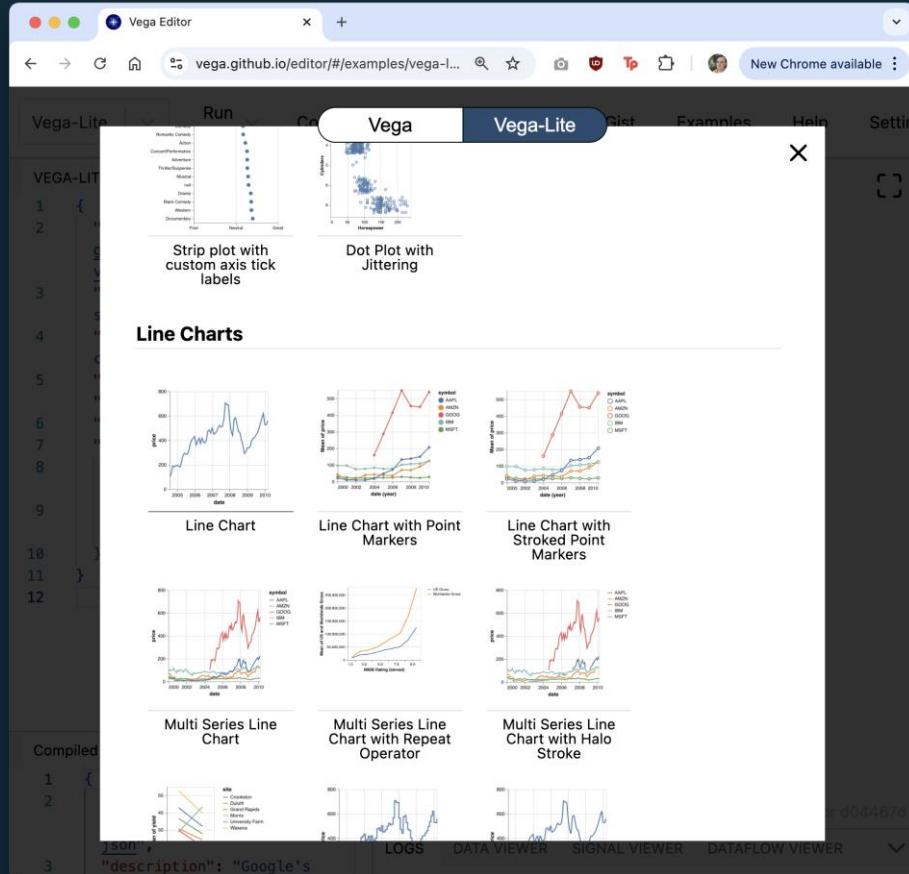
The image shows two side-by-side browser windows. The left window is a GitHub repository page for 'mclass-user/mclass-user.github.io'. It displays a file named 'uk_gdp_per_hour.csv' with 54 lines of data. The right window is a raw.githubusercontent.com page showing the contents of the same CSV file. The 'Raw' button in the GitHub preview panel is highlighted with a yellow box.

| | date | value |
|----|------|-------|
| 1 | 1971 | 43.7 |
| 2 | 1972 | 45 |
| 3 | 1973 | 47.1 |
| 4 | 1974 | 46.2 |
| 5 | 1975 | 46.5 |
| 6 | 1976 | 48.3 |
| 7 | 1977 | 49.1 |
| 8 | 1978 | 50.6 |
| 9 | 1979 | 51.9 |
| 10 | 1980 | 51.9 |
| 11 | 1981 | 54.2 |
| 12 | 1982 | 56.6 |
| 13 | | |

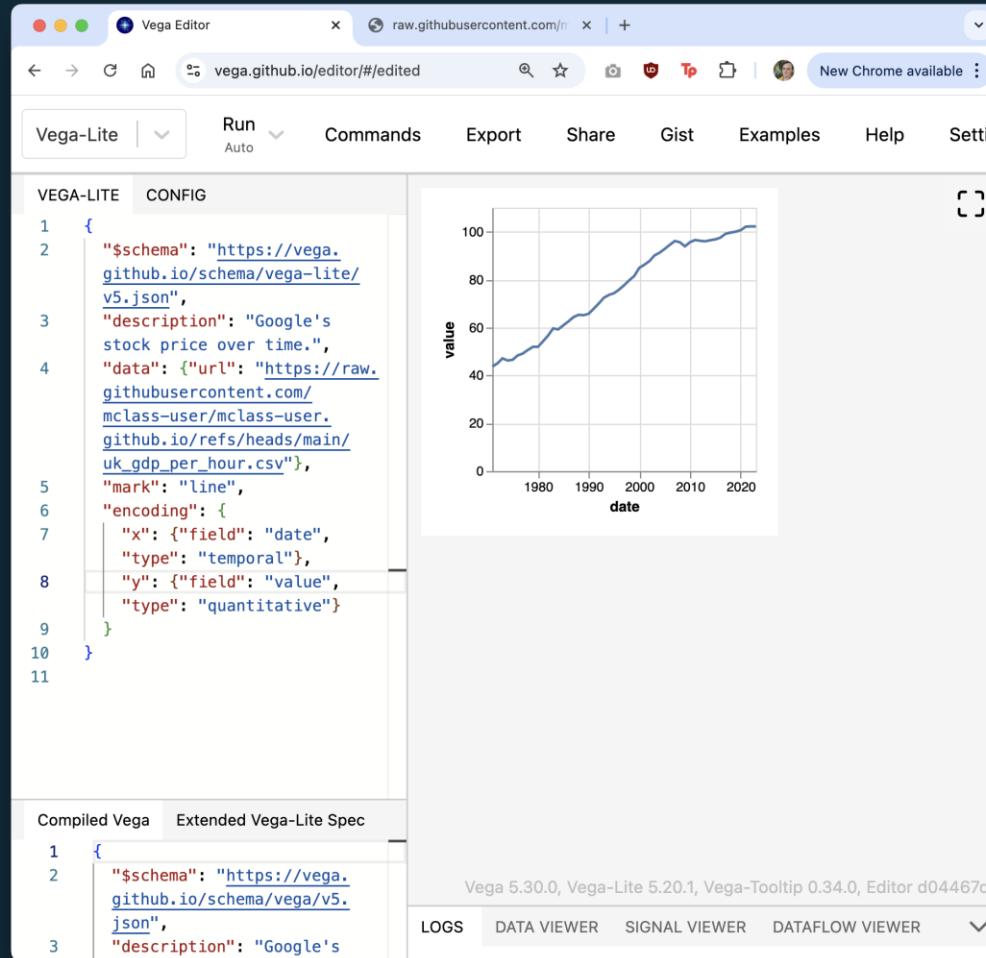
```
date,value
1971,43.7
1972,45
1973,47.1
1974,46.2
1975,46.5
1976,48.3
1977,49.1
1978,50.6
1979,51.9
1980,51.9
1981,54.2
1982,56.6
1983,59.6
1984,59.2
1985,60.9
1986,62.5
1987,64.3
1988,65.3
1989,65.1
1990,65.7
1991,67.8
1992,70
1993,72.4
1994,73.6
1995,74.3
1996,75.8
1997,77.6
1998,79.7
1999,81.5
2000,84.9
2001,86.2
2002,87.7
2003,90.1
2004,91.2
2005,92.8
2006,94.5
2007,96.1
2008,95.6
2009,93.9
2010,95.6
2011,96.5
2012,96.2
2013,96
2014,96.4
2015,96.8
2016,97.5
2017,99.1
2018,99.6
2019,100
2020,100.6
2021,102.1
2022,102.2
2023,102.2
```

Adapting a Vega-lite Example.

We can start with an example chart.



Adapting a Vega-lite Example.



The screenshot shows the Vega Editor interface. On the left, the VEGA-LITE tab displays a JSON configuration for a line chart. The chart plots 'value' against 'date', showing a steady increase from approximately 45 in 1980 to about 100 in 2020. On the right, the CONFIG tab shows the corresponding Vega JSON code. Below these tabs, the Compiled Vega and Extended Vega-Lite Spec sections show the generated JSON. The bottom status bar indicates the version: Vega 5.30.0, Vega-Lite 5.20.1, Vega-Tooltip 0.34.0, Editor d04467d.

```
VEGA-LITE
1 { "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
2   "description": "Google's stock price over time.",
3   "data": {"url": "https://raw.githubusercontent.com/mclass-user/mclass-user.github.io/refs/heads/main/uk_gdp_per_hour.csv"},
4   "mark": "line",
5   "encoding": {
6     "x": {"field": "date", "type": "temporal"},
7     "y": {"field": "value", "type": "quantitative"}
8   }
9 }
10
11
```

```
CONFIG
1 {
2   "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
3   "description": "Google's stock price over time.",
4   "data": {"url": "https://raw.githubusercontent.com/mclass-user/mclass-user.github.io/refs/heads/main/uk_gdp_per_hour.csv"},
5   "mark": "line",
6   "encoding": {
7     "x": {"field": "date", "type": "temporal"},
8     "y": {"field": "value", "type": "quantitative"}
9   }
10 }
```

```
Compiled Vega
1 {
2   "$schema": "https://vega.github.io/schema/vega/v5.json",
3   "description": "Google's stock price over time."}
```

```
Extended Vega-Lite Spec
1 {
2   "$schema": "https://vega.github.io/schema/vega/v5.json",
3   "description": "Google's stock price over time."}
```

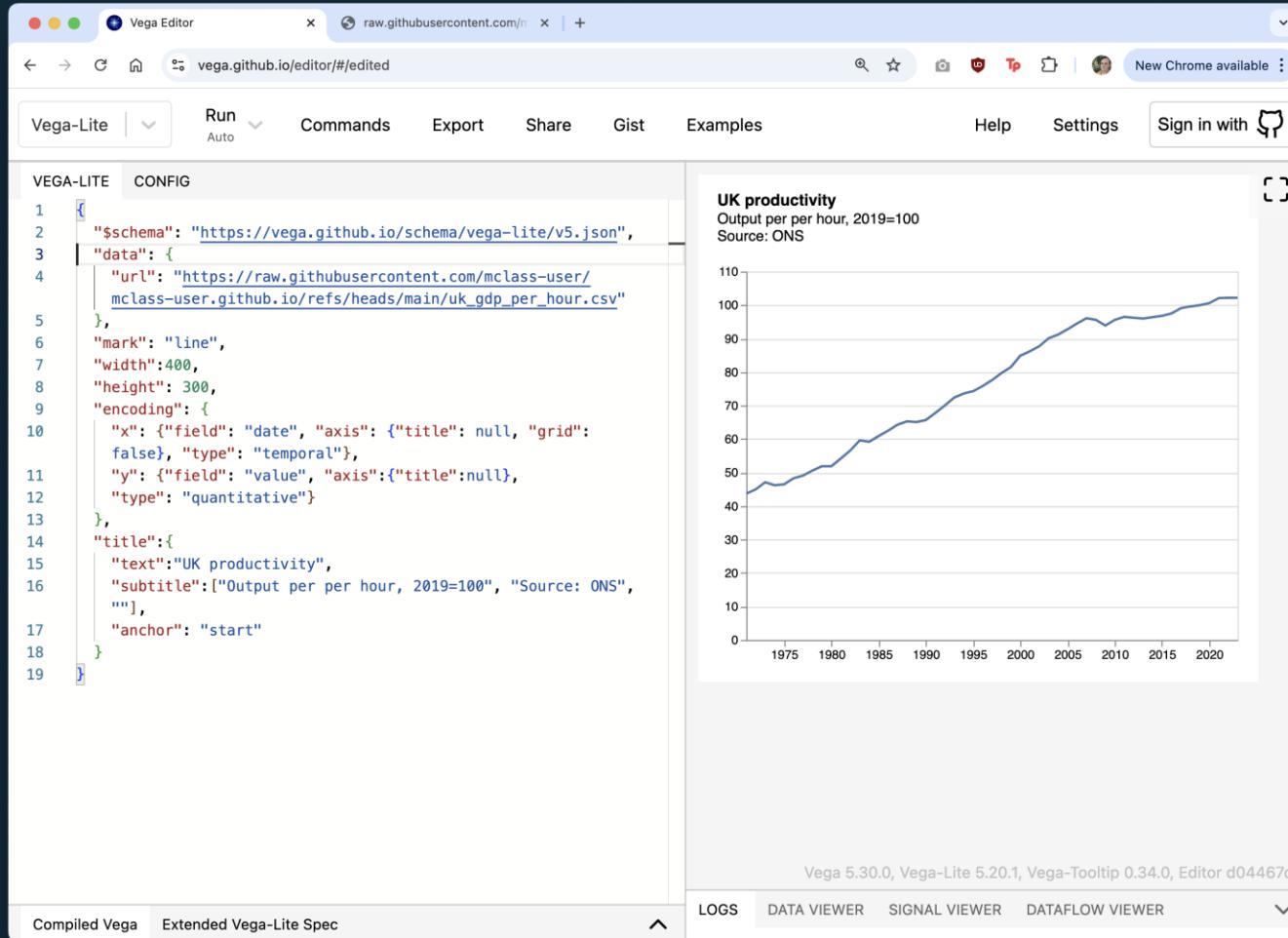
Vega 5.30.0, Vega-Lite 5.20.1, Vega-Tooltip 0.34.0, Editor d04467d

LOGS DATA VIEWER SIGNAL VIEWER DATAFLOW VIEWER

We adapt it to our needs by changing fields.

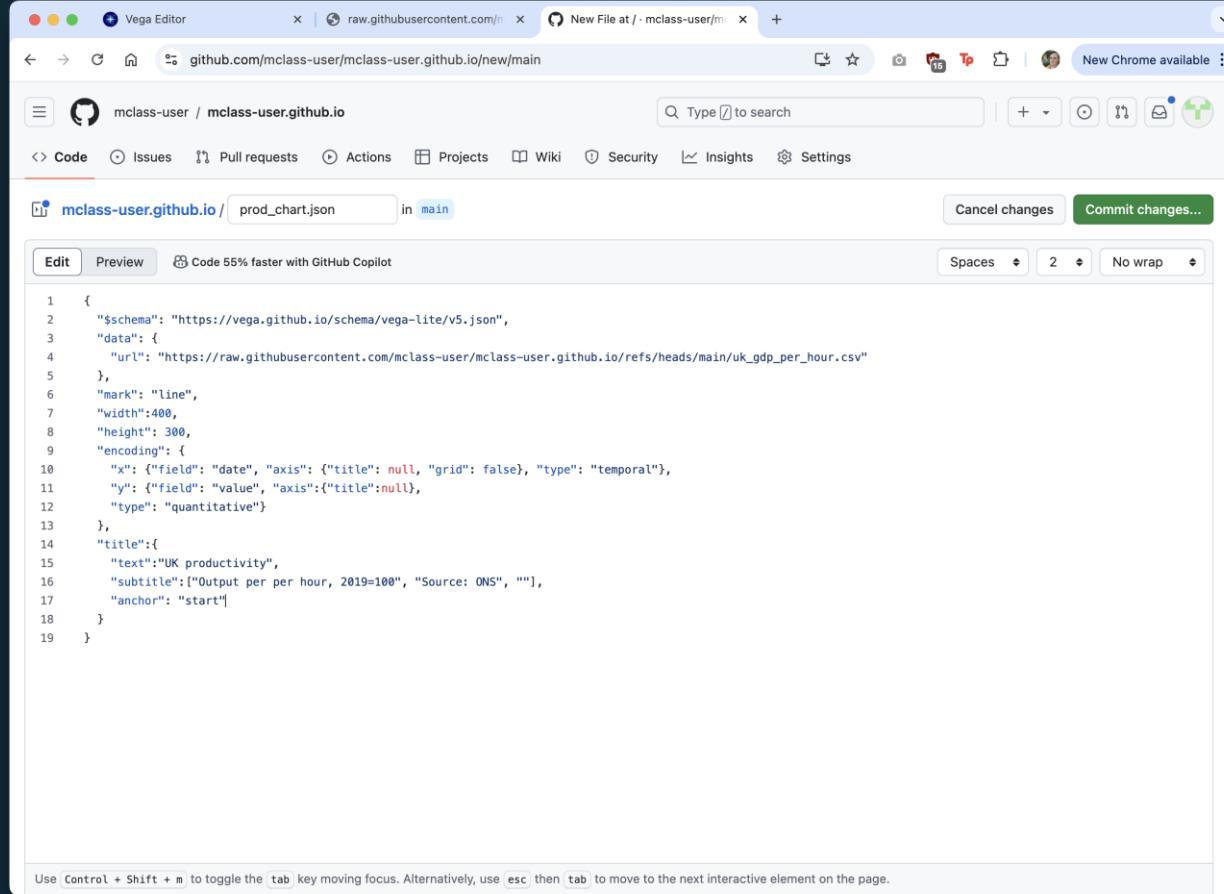
- Changing column names
- Adding our **Raw** data link
- Removing extraneous code

Adapting a Vega-lite Example.



The chart could do with work.
Let's customise the axes and add a title.

Uploading the chart.



A screenshot of a GitHub code editor window. The URL in the address bar is `github.com/mclass-user/mclass-user.github.io/main`. The file being edited is `prod_chart.json` in the `main` branch. The code editor shows the following JSON configuration:

```
1  {
2    "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
3    "data": {
4      "url": "https://raw.githubusercontent.com/mclass-user/mclass-user.github.io/main/uk_gdp_per_hour.csv"
5    },
6    "mark": "line",
7    "width": 400,
8    "height": 300,
9    "encoding": {
10      "x": {"field": "date", "axis": {"title": null, "grid": false}, "type": "temporal"},
11      "y": {"field": "value", "axis": {"title": null}, "type": "quantitative"}
12    },
13    "title": {
14      "text": "UK productivity",
15      "subtitle": ["Output per hour, 2019=100", "Source: ONS", ""],
16      "anchor": "start"
17    }
18  }
19 }
```

The GitHub Copilot logo is visible in the top right corner of the code editor. Below the code editor, a note says: "Use Control + Shift + m to toggle the tab key moving focus. Alternatively, use esc then tab to move to the next interactive element on the page."

Finally, we can add the finished chart to the repo.

Festival of Economics.

1-3 October 2024

Book talks, panels & live podcast recordings – your chance to ask experts from around the world the key economic questions of our time!

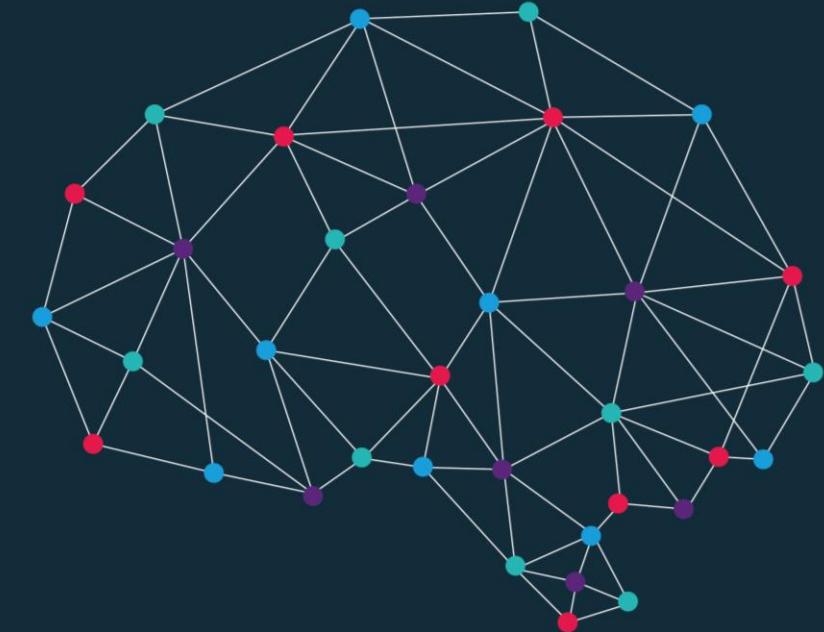
Scan for tickets and the programme.



FREE TICKETS for:

- **Legacies of Empire** (Tues 1 October, 7pm)
- **Inequalities** (Thursday 3 Oct, 7pm)

in the Wills Memorial Building and are free to UoB students



With the election of a new government after 14 years, the theme of the Festival is **looking back, to look forward:**

Legacies of Empire, Inequalities, The Tyranny of Nostalgia, Sustainable Futures, Labour's first 100 days, Building a fair society, Britain's place in the world , Social mobility, The kids aren't alright, UK productivity, Tim Harford podcast ... and more!

