



Data Visualisation: Charts & Other Choices

DISCOVER. LEARN. ANALYSE. SHAPE. REPEAT



#DatawiseLondon





Our mission is to help you unlock the value of your data.



Learning objectives

1. Identify factors that make data visualisations less/more successful
2. Understand principles that can be used to design a data visualisation
3. Understand types of visualisations commonly used, and type of data each is best suited to represent
4. Introduce some no-coding tools to get started



Agenda

1. Introductions
2. Exercise: What works/what doesn't
3. 5 steps for data visualisation
4. Thinking about colour
5. Accessibility considerations
6. No-code tools to get started



Go to www.menti.com and use the code 33 02 79 8



Exercise: what works, what doesn't?

https://jamboard.google.com/d/1dBwjxTJat9DY_upjCNjv1x_XINUR9wW7aox_uw2LAg0/viewer?f=0



Our principles



visualization

1. Decide your take home message
2. Define your audience
3. Prep the data
4. Choose the type of visualization
5. Refine!



1. Decide your take home message

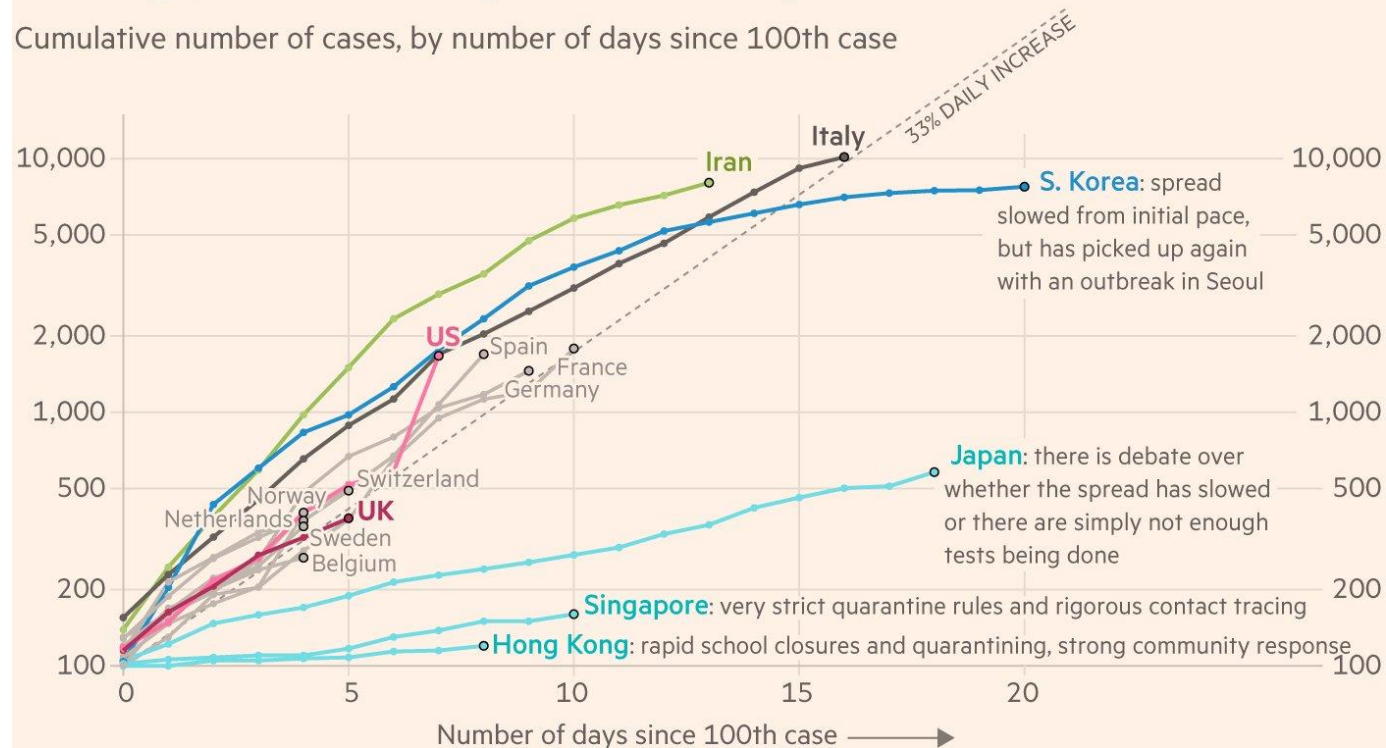
- ↘ What do you want your audience to do with this information?
 - ★ What do you want your audience to conclude?
 - ★ What action do you want your audience to take?
 - ★ How do you want your audience to feel?
- ↘ What would the newspaper headline for this visualisation say?



1. Decide your take home message

Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread

Cumulative number of cases, by number of days since 100th case



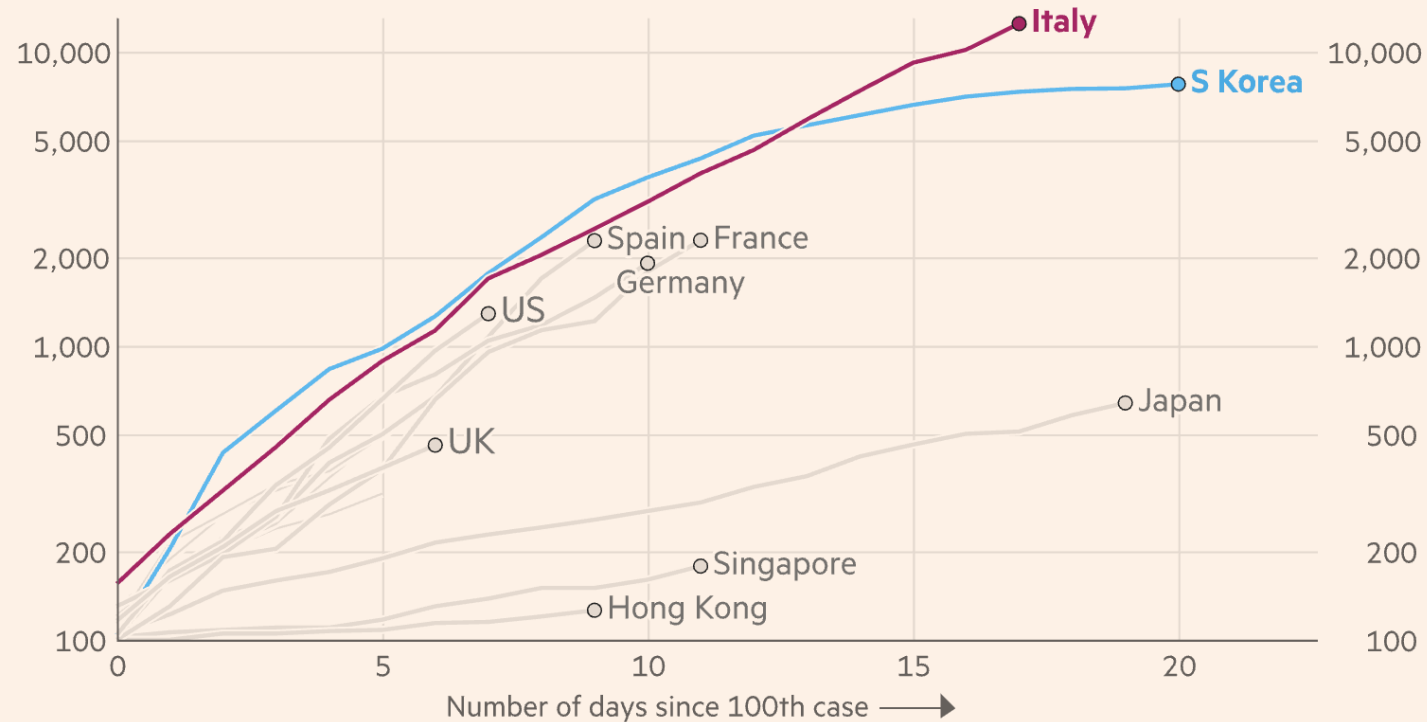
Source: FT analysis of Johns Hopkins University, CSSE
FT graphic: John Burn-Murdoch / @jburnmurdoch
© FT



1. Decide your take home message

In early March, Italy and South Korea appeared to be on similar paths

Cumulative number of coronavirus cases, by number of days since 100th case (data as of March 12)



2. Define your audience

- ↘ What background knowledge will this audience have?
- ↘ Will this be used internally or externally?
- ↘ What data literacy level will your audience have?
- ↘ What questions will the audience be looking to answer?
- ↘ What sorts of decisions will your audience be making with this visualization?



2. Define your audience

Coronavirus in the UK

Total deaths

81,431

Latest daily figure

563

new deaths

Two-month trend



Total cases

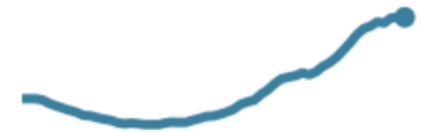
3,072,349

Latest daily figure

54,940

new cases

Two-month trend



Exercise: Who's your audience?

- Which audience(s) are you making visualisations for?
- What knowledge will they be bringing?
- What sorts of decisions/conclusions will they be looking to make?



3. Prep & explore the data

- ↘ Filter all but the data of interest
- ↘ Check and clean the data
 - ★ Remove duplicates
 - ★ Check for missing values
- ↘ Perform any summaries needed
- ↘ What level of detail do you need?
- ↘ What level of precision do you need?



Data Cleaning checklist

- ↘ How much data is missing?
- ↘ Are missing values encoded the same way? (e.g. NA, 'Unknown')
- ↘ Standardise and validate (e.g. My First Charity > My First Charity Ltd), check postcodes
- ↘ Are category options consistent (e.g. a computer will treat "Woman" and "woman" as distinct options)
- ↘ Are dates in the same format (e.g. dd/mm/yyyy vs mm/dd/yyyy)?
- ↘ Check for plausibility and errors/typos (does an organisation really have an income of £10bn?)
- ↘ Check for duplicates
- ↘ Check for PII



Data Exploration

- ↘ What's the average value for the thing you are measuring?
- ↘ What's the median? The range?
- ↘ How does this value vary by different groups?
- ↘ What other factor(s) could explain the variation?

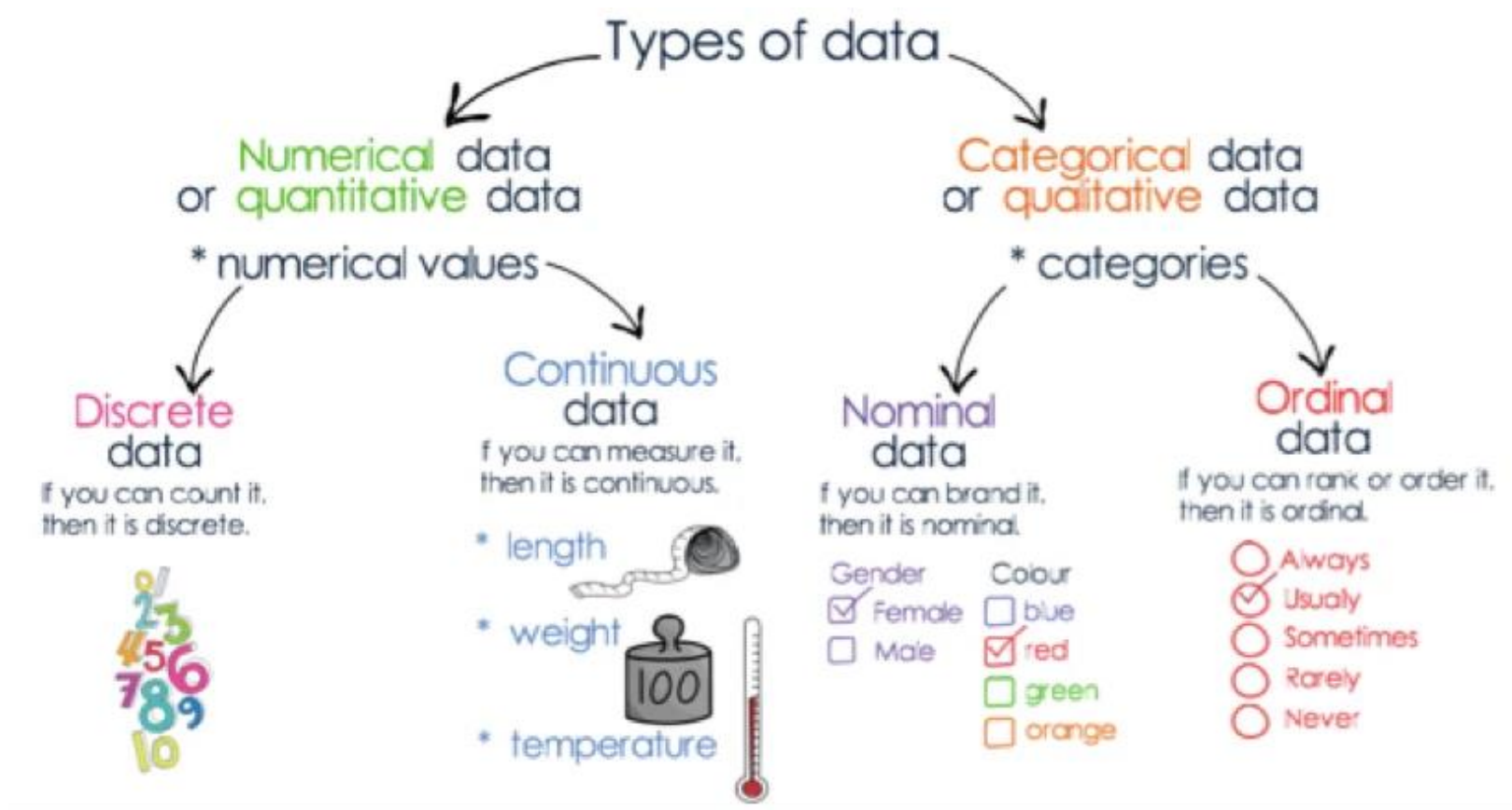


4. Choose the type of visualization

- ↘ What type data do you have?
- ↘ How much data do you have?
- ↘ What do you want to show?



What type of data are you using?



How much data do you have?

- How many pieces of data (variables) are you using?
 - **One:** e.g. Client age/gender/ethnicity, amount donated
 - **Two:** e.g. Total fundraising over time,
 - **Three or more:** e.g. Outcome by client sex and age band
- How many observations do you have?
 - **One:** e.g. total donated per individual or total donated per postcode district
 - **Two or more:** e.g. academic performance by school, each year for 10 years; monthly donation per individual



COMPARISON

RELATIONSHIP

What would you
like to show?

DISTRIBUTION

COMPOSITION



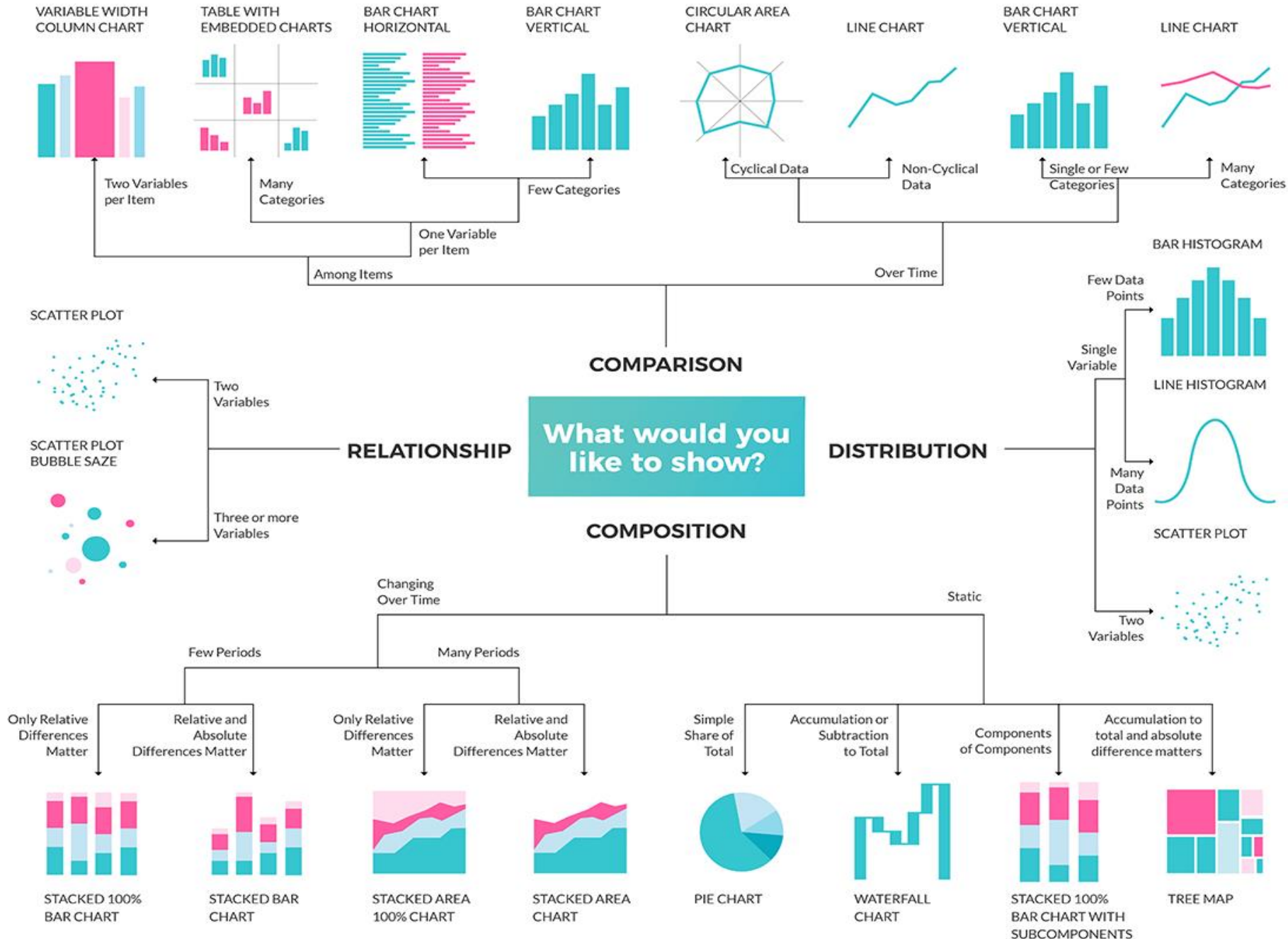
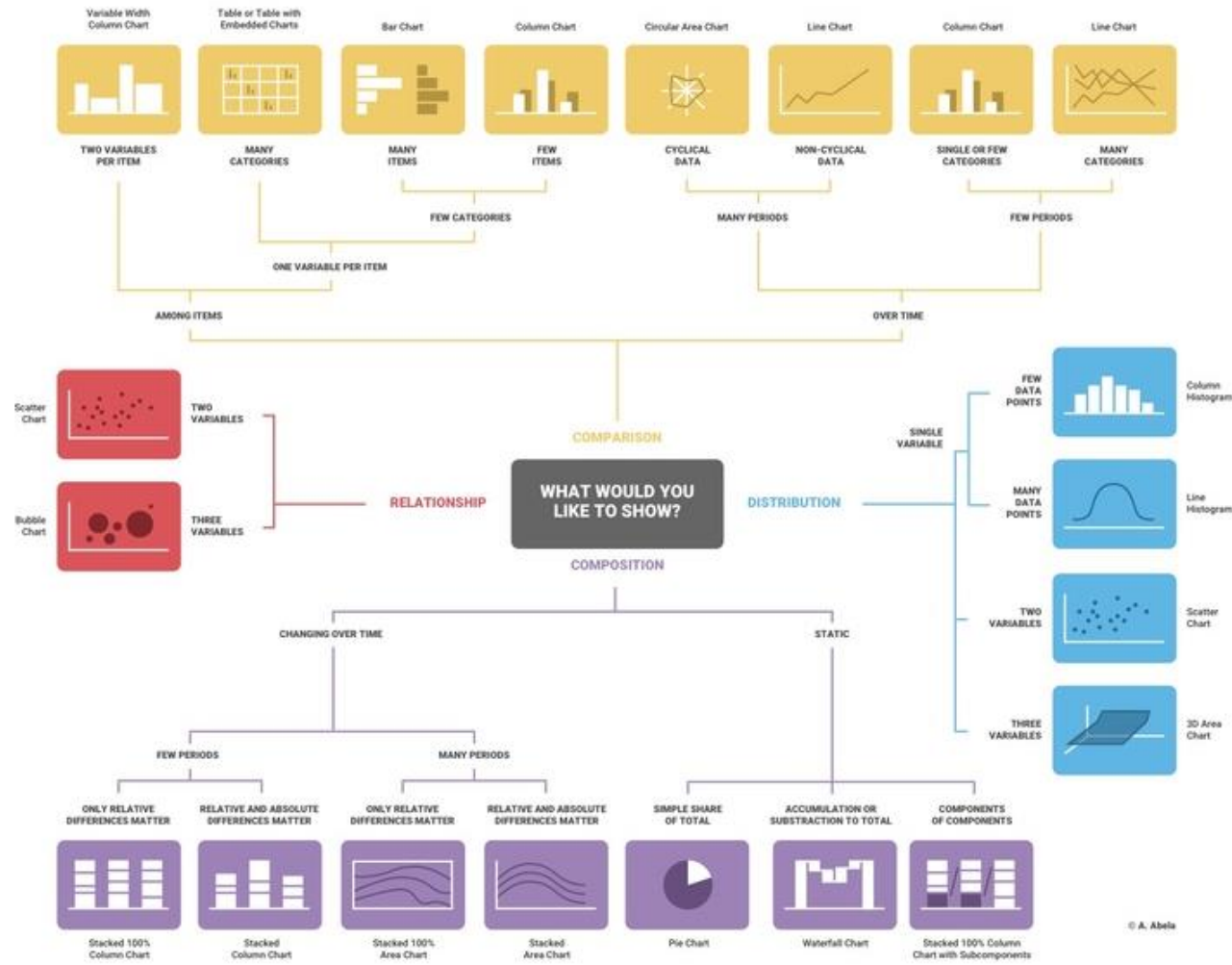


CHART SUGGESTIONS - A THOUGHT-STARTER



© A. Abela

<https://infogram.com/page/choose-the-right-chart-data-visualization>



What kind of data do you have? Pick the main type using the buttons below.

Then let the decision tree guide you toward your graphic possibilities.

Numeric

Categoric

Num & Cat

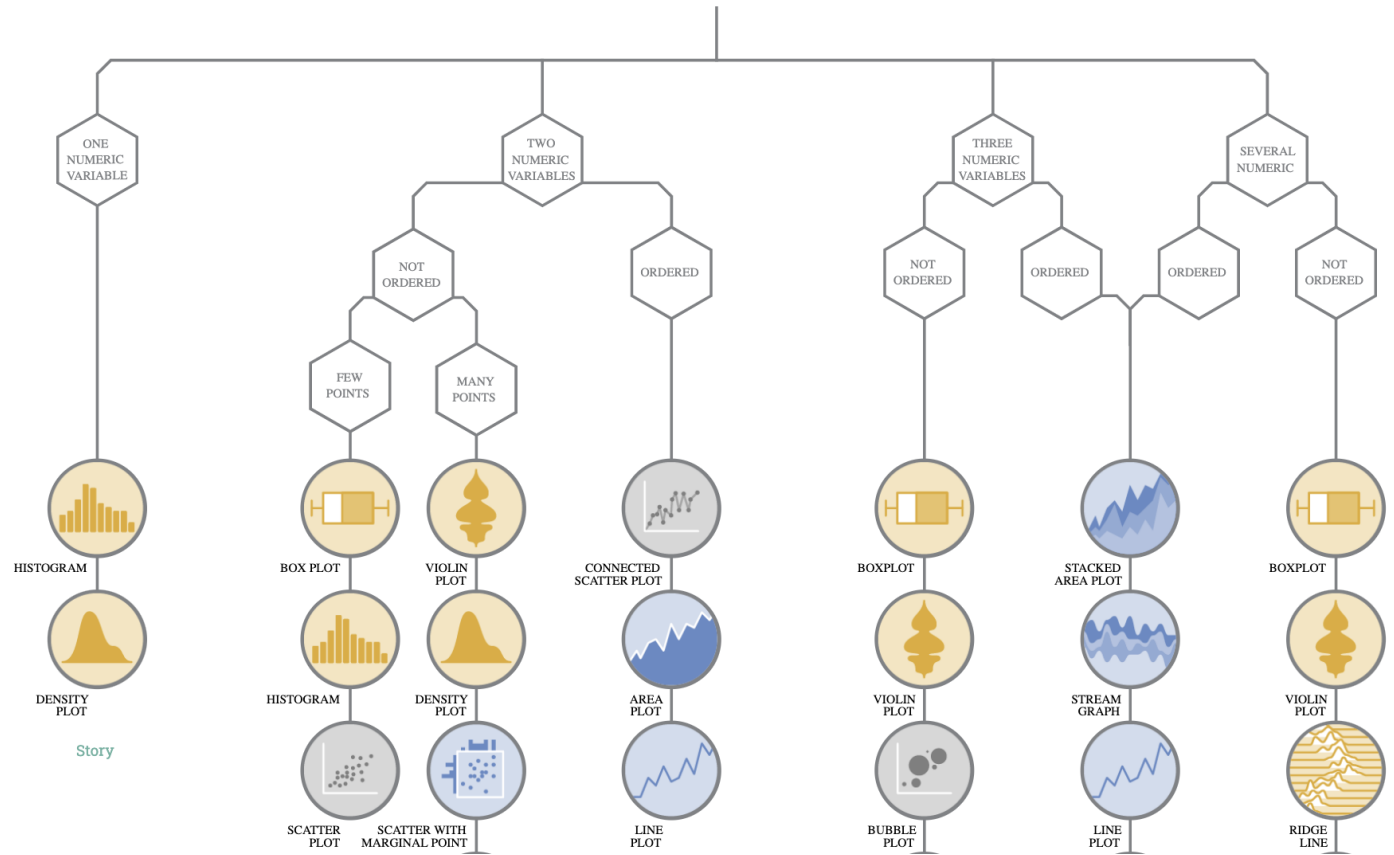
Maps

Network

Time series



from Data to Viz



<https://www.data-to-viz.com/>

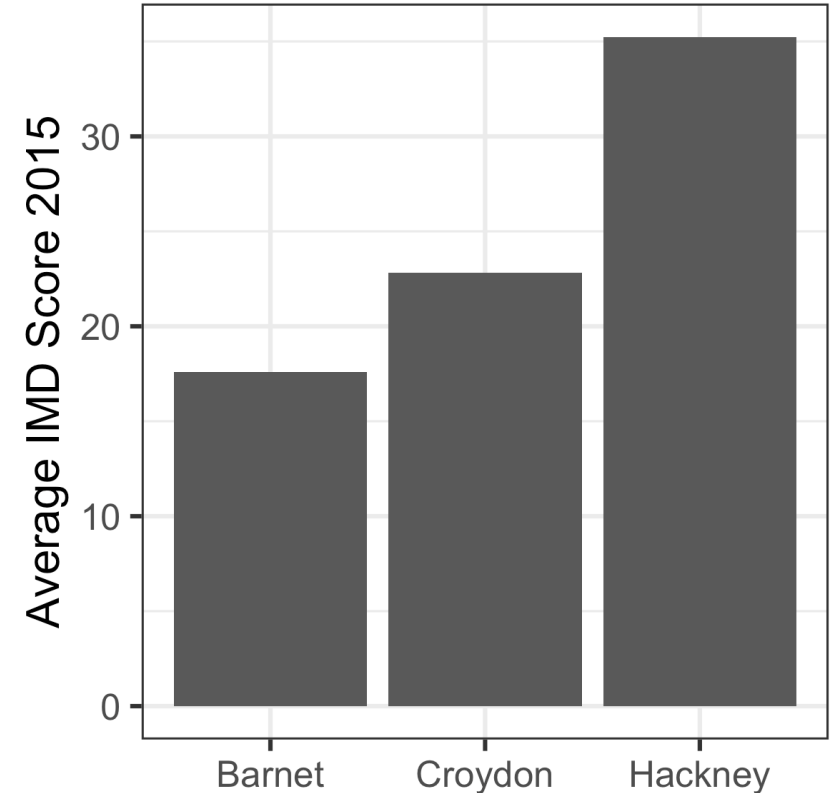


Comparing values



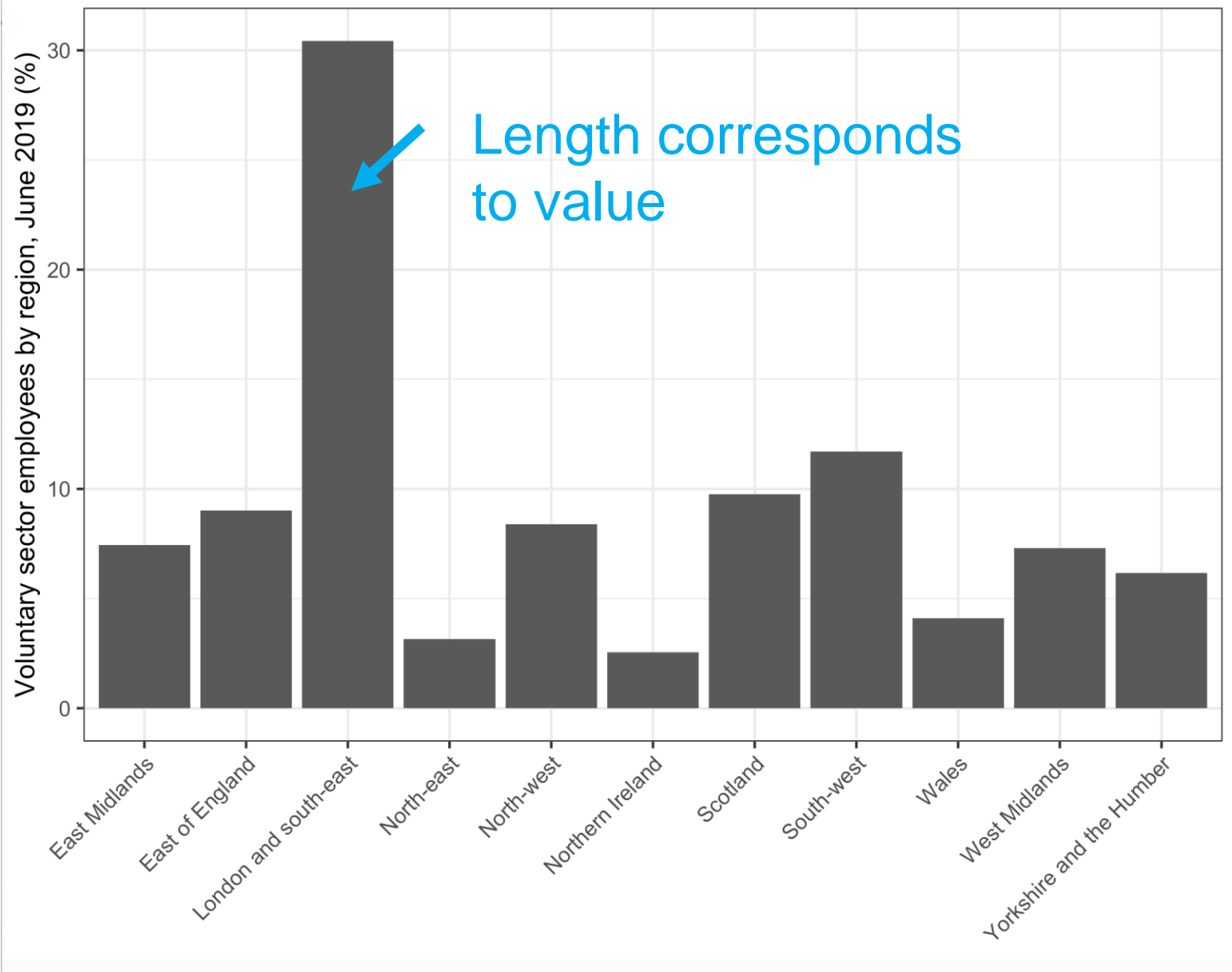
Bar charts

- **What is it?** The value for each entity in a category is shown as a bar, where the length corresponds to the value
- **Input data:** One numeric variable for 1 or more categories
- **Why use it?** To compare amounts

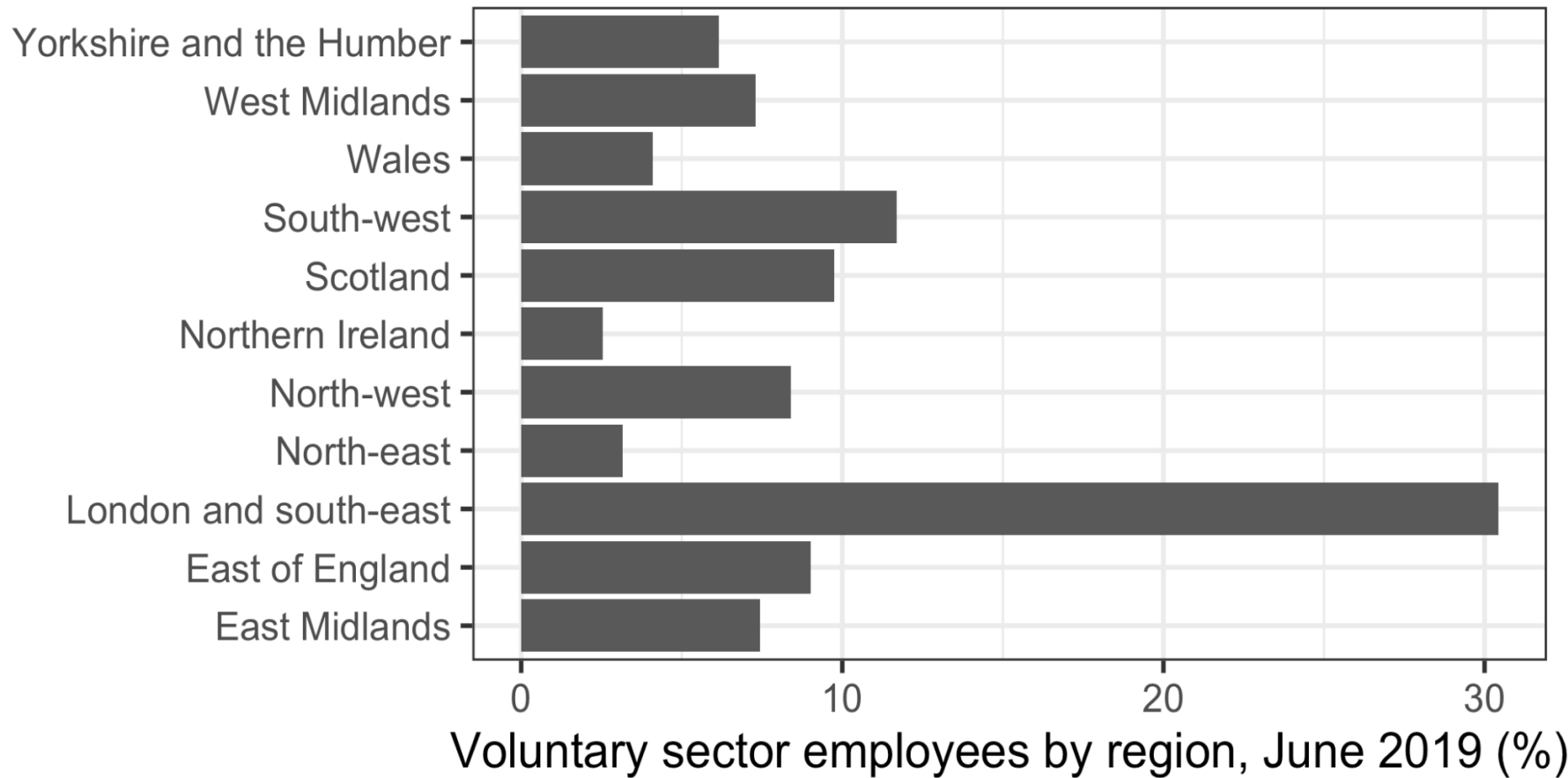


Bar charts

Region	Percent
London and south-east	30.422247
South-west	11.700077
Scotland	9.747461
East of England	9.023659
North-west	8.396877
East Midlands	7.434044
West Midlands	7.311723
Yorkshire and the Humber	6.161340
Wales	4.100153
North-east	3.159430
Northern Ireland	2.542988



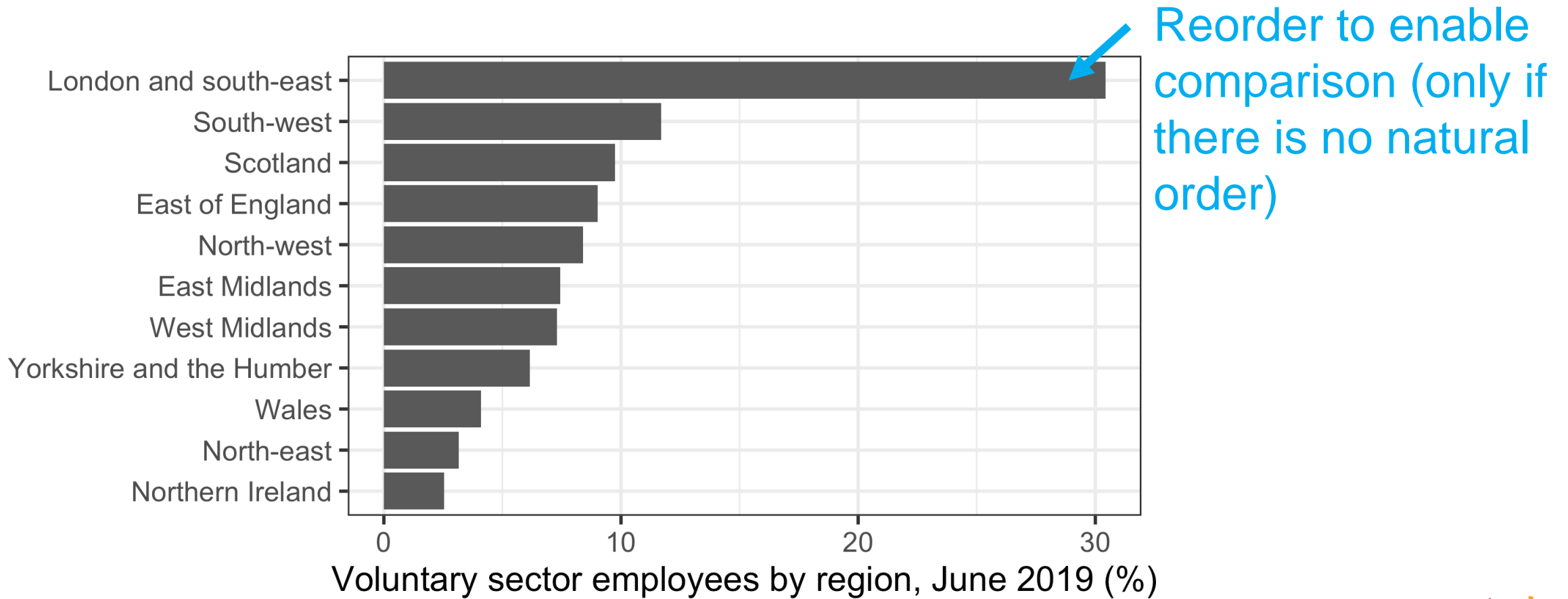
Bar chart tips



Flip charts with long labels sideways



Bar chart tips



Bar chart tips

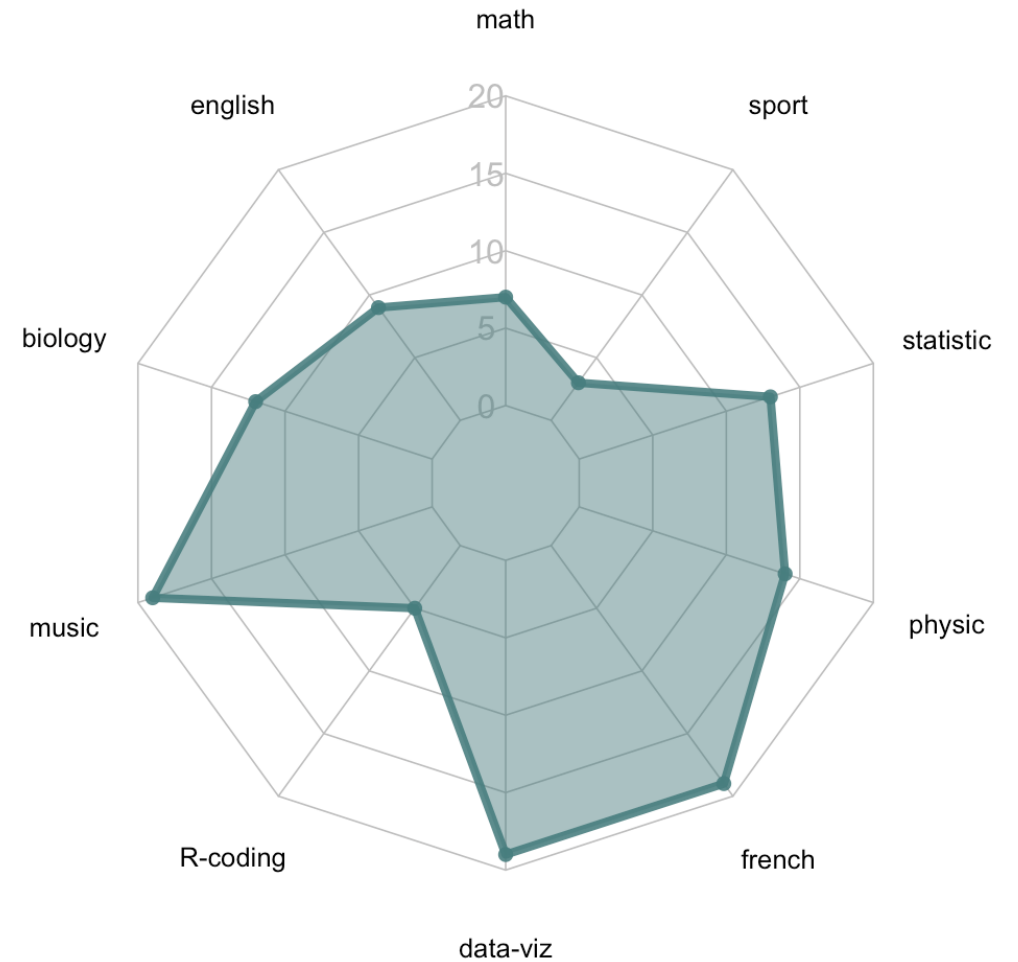


Add value to bar if precise value is important

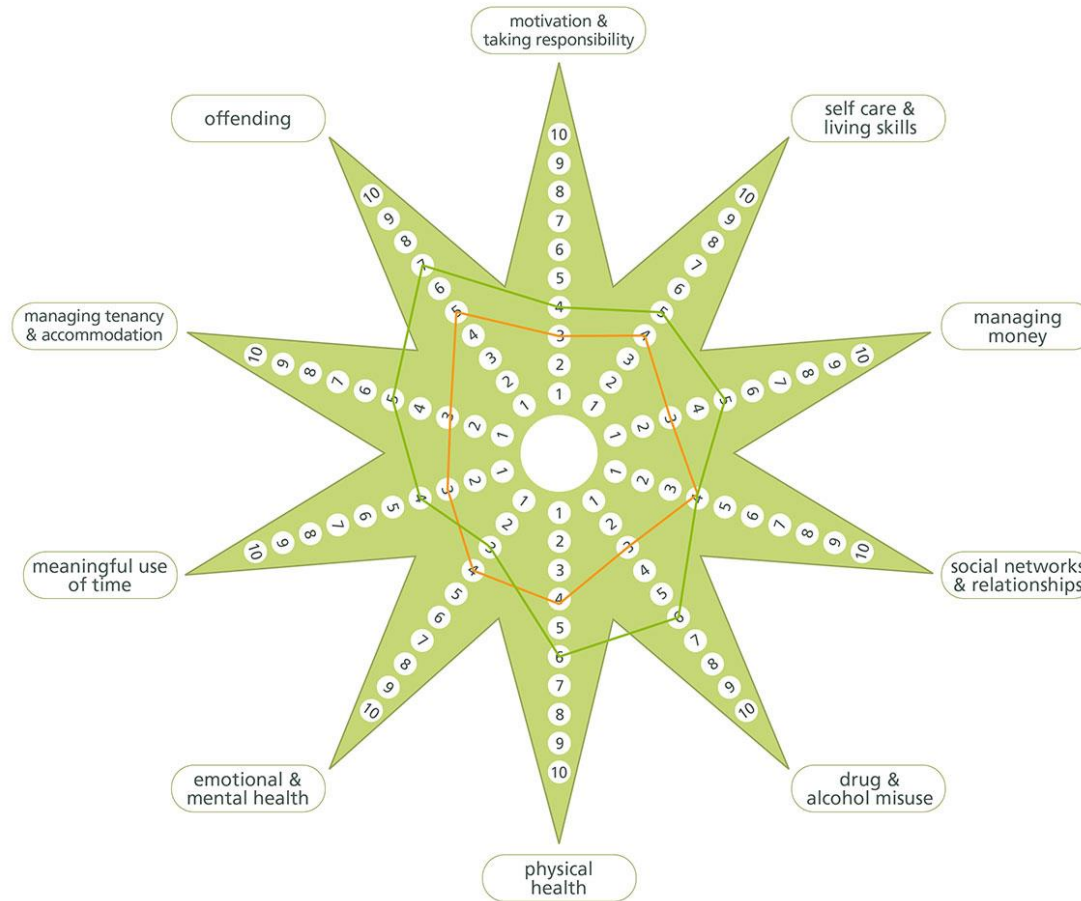


Spider Chart/Radar Chart

- **What is it?** A 2D chart that shows a series of values for *multiple variables*
- **Input data:** One number for >3 variables
- **Why use it?** To compare values across multiple domains



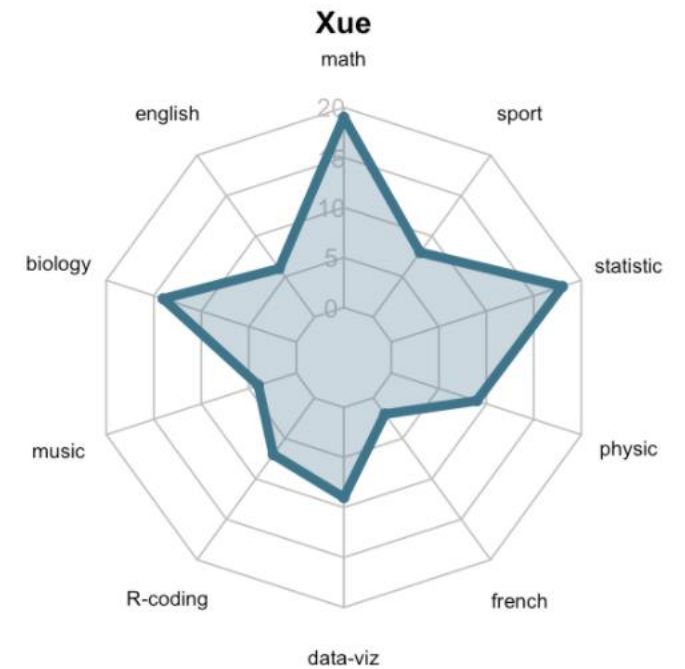
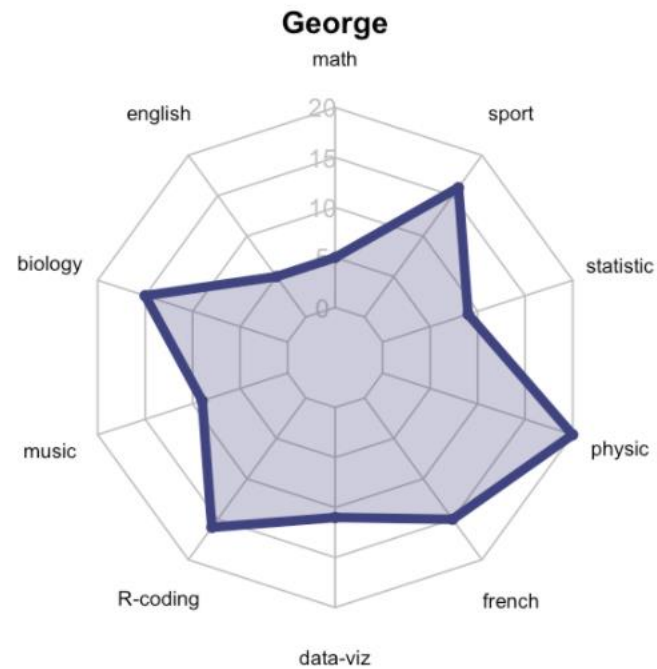
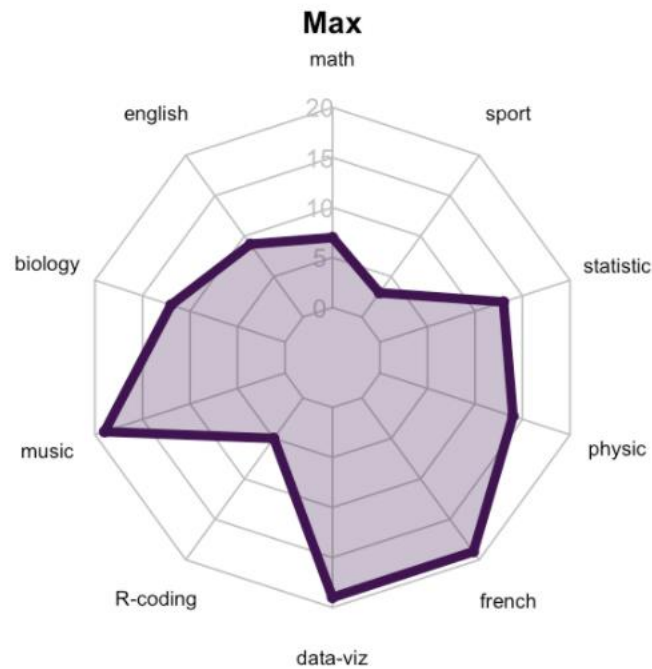
Outcome stars are a type of radar chart



Spider Chart/Radar Chart

Advantages:

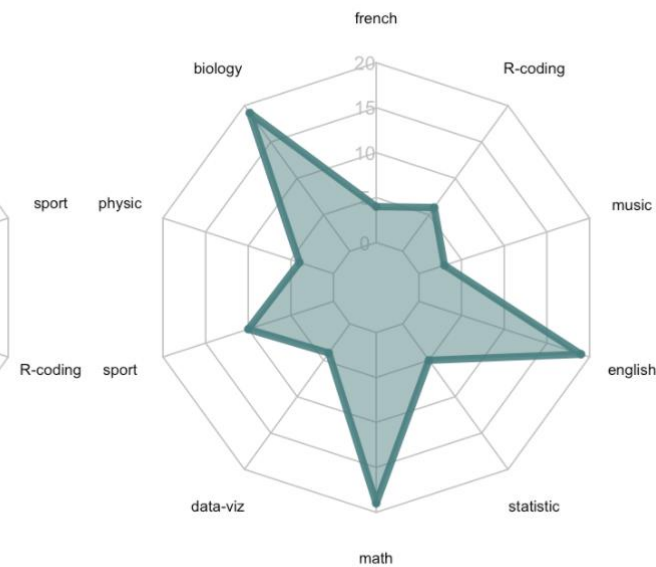
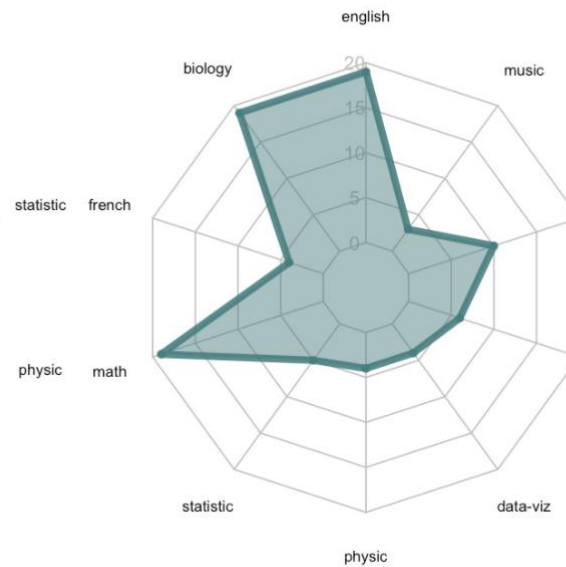
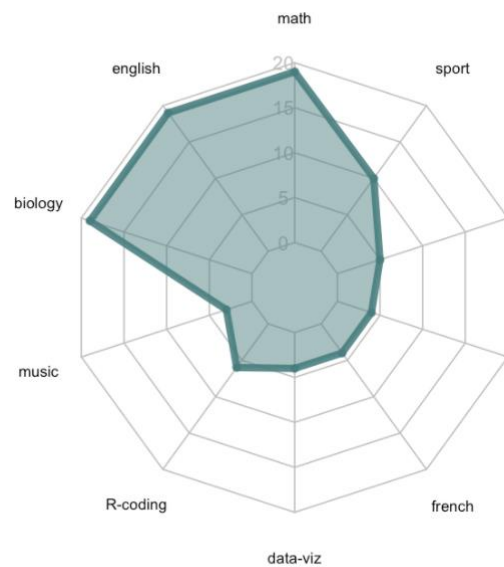
- Useful for individuals to compare scores across similarly measured



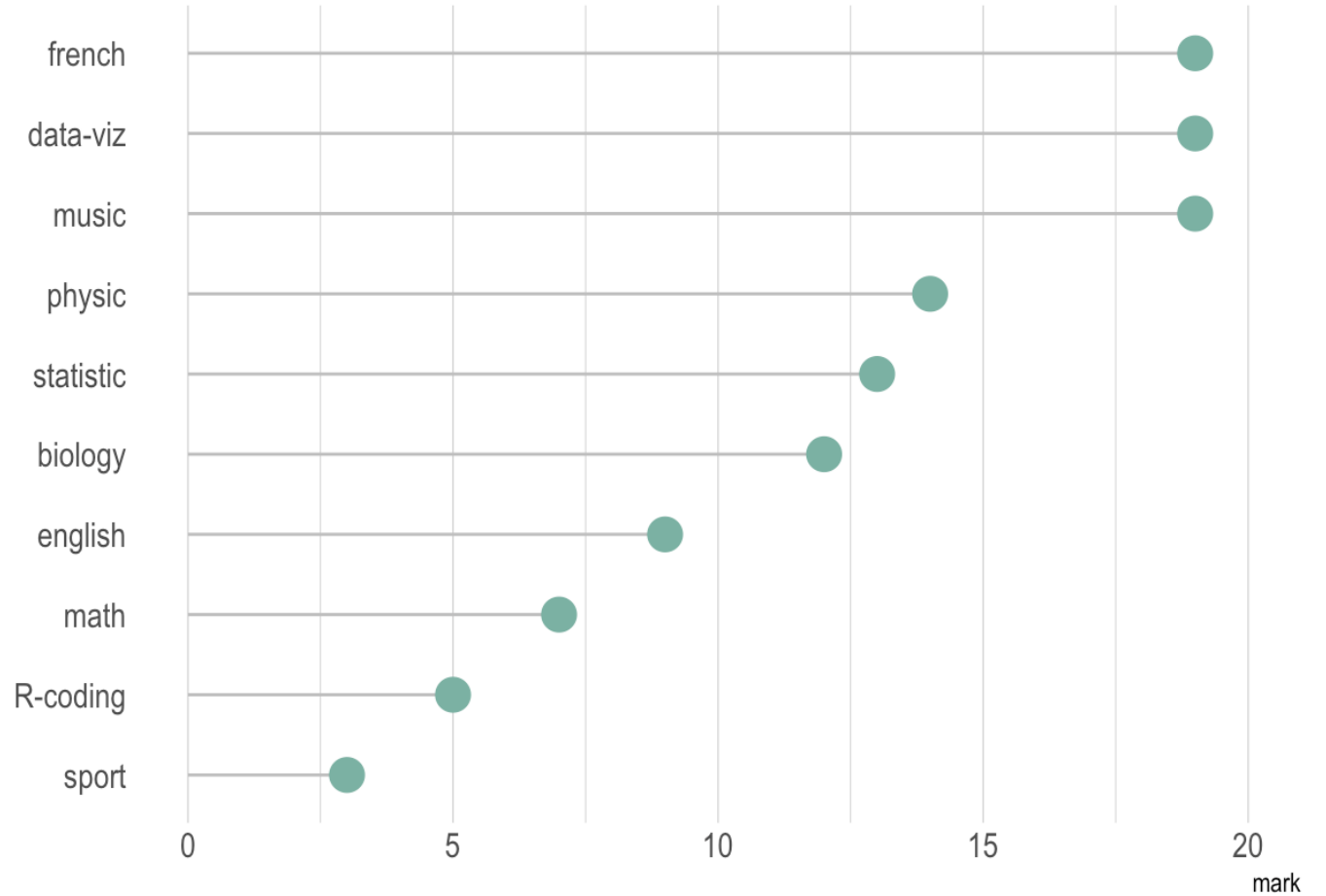
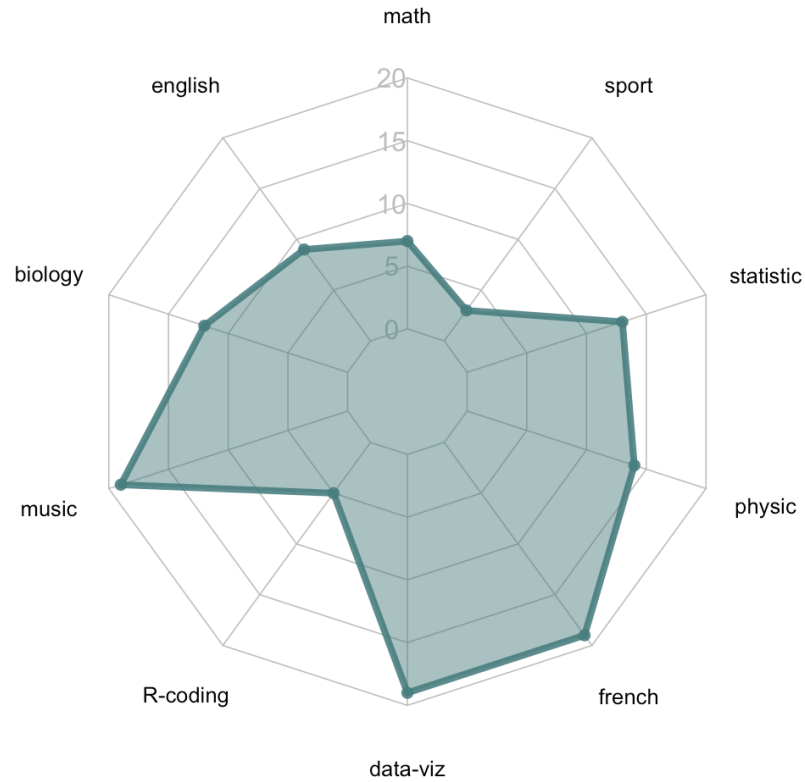
Spider Chart/Radar Chart

Disadvantages

- Circular layout is hard to read
- Category order affects the shape
- Hard to determine ranking
- Data needs to be on the same scale



Spider Chart/Radar Chart alternatives



Dumbbell Charts

- **What is it?** Compares 2 different data points for each group
- **Why use it?** To compare values over time or between viewpoints (e.g. Labour vs. Conservative)

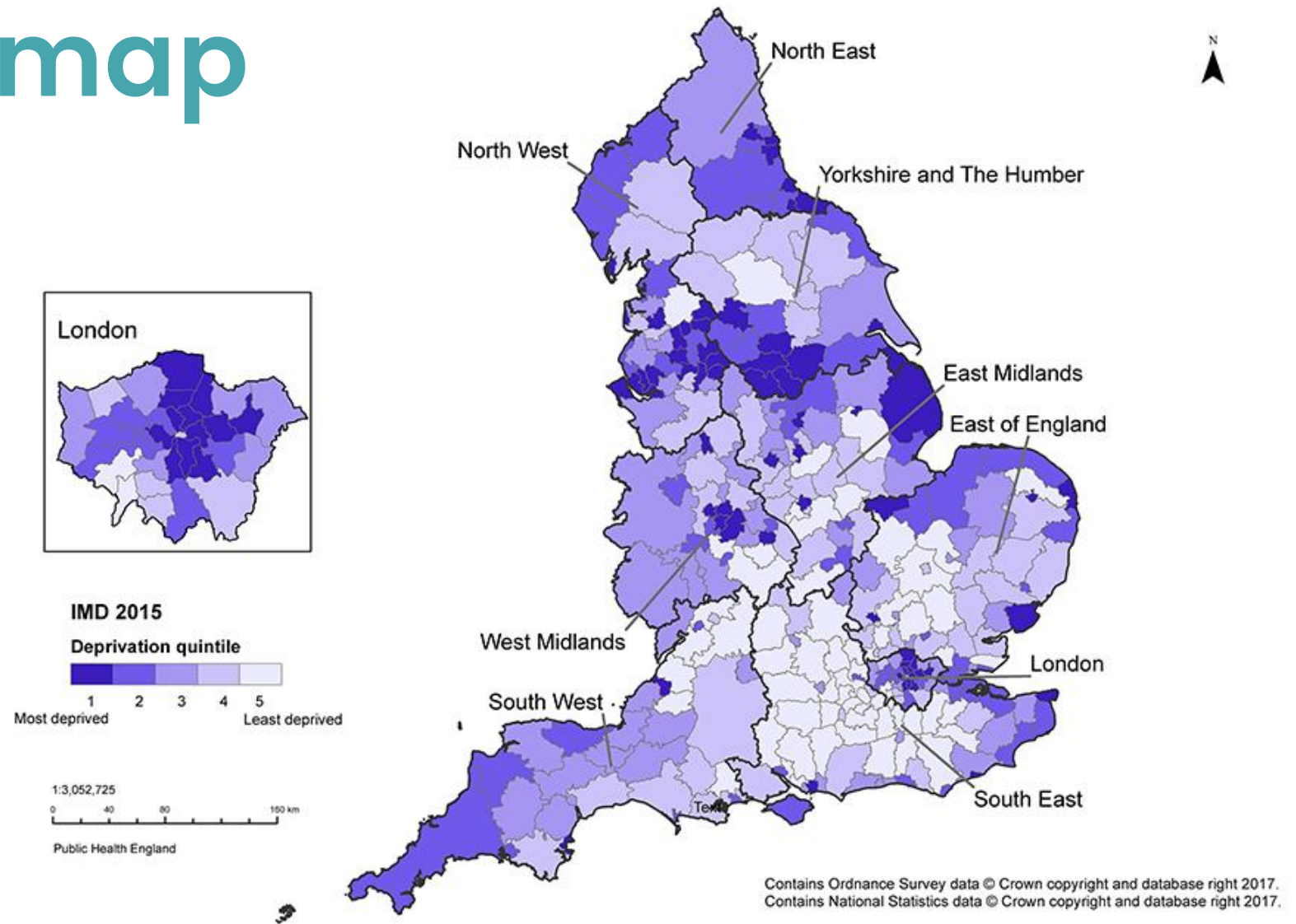
Even in London, LGBTQ+ people are more likely to struggle

Comparing survey responses from LGBTQ+ people and heterosexuals about life in London



Chloropleth map

- **What is it?** A map in which geographical areas are coloured to indicate different values
- **Why use it?** To understand how a metric varies across known geographical boundaries
- **Challenges:** Choosing the "right" geographical level as there may be more variation within an area



Visualising distributions: understanding variation



Histograms & density plots

What is it?

- shows the *distribution* of numeric data
- A *distribution* shows the probability of different possible values

Input data: one numeric variable

Why use it? To understand the frequency of different values in your data, e.g. if some values are more common than others



Histograms

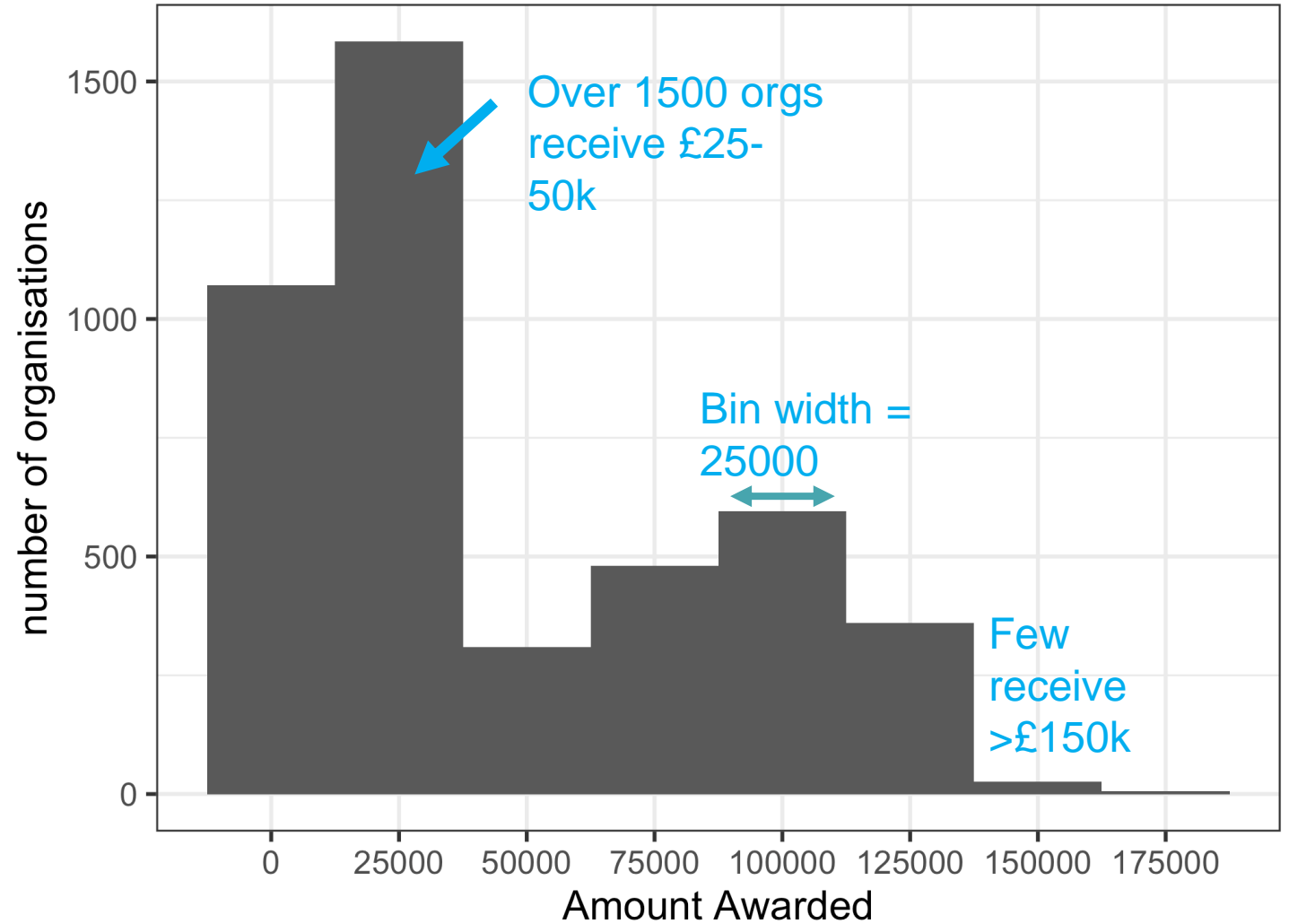
Input data: Grant funding awarded to different organisations

Amount Awarded
49495
55709
81483
88157
71514
83795
44654
59753
110000
51478

Each row is 1 grant

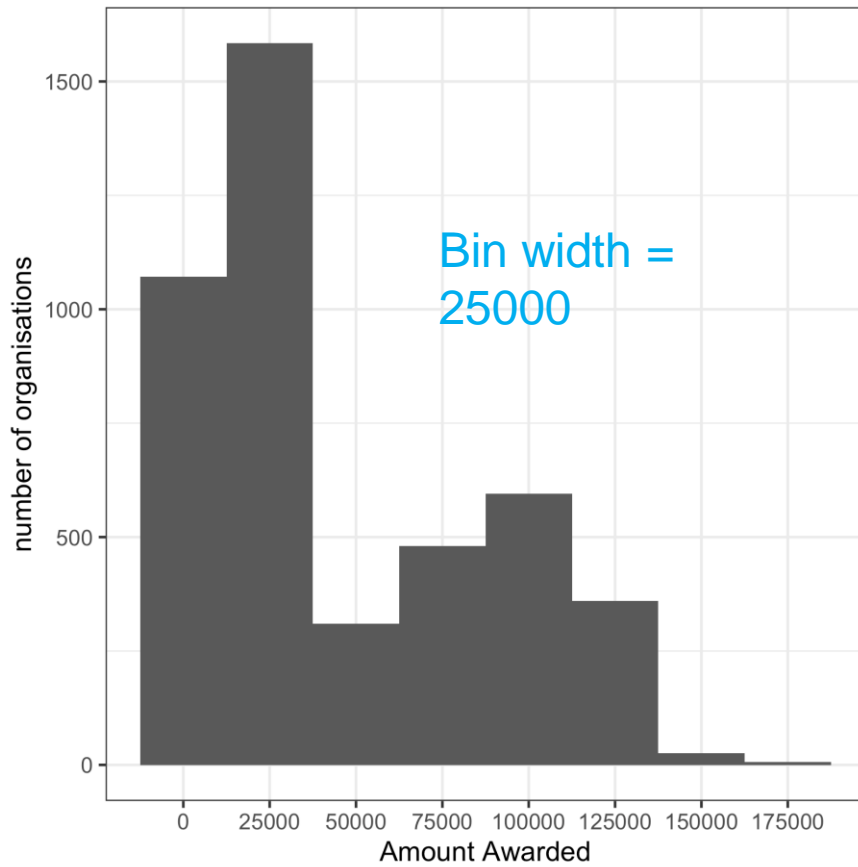
Within range 50000-75000

BBC Children in Needs Grants

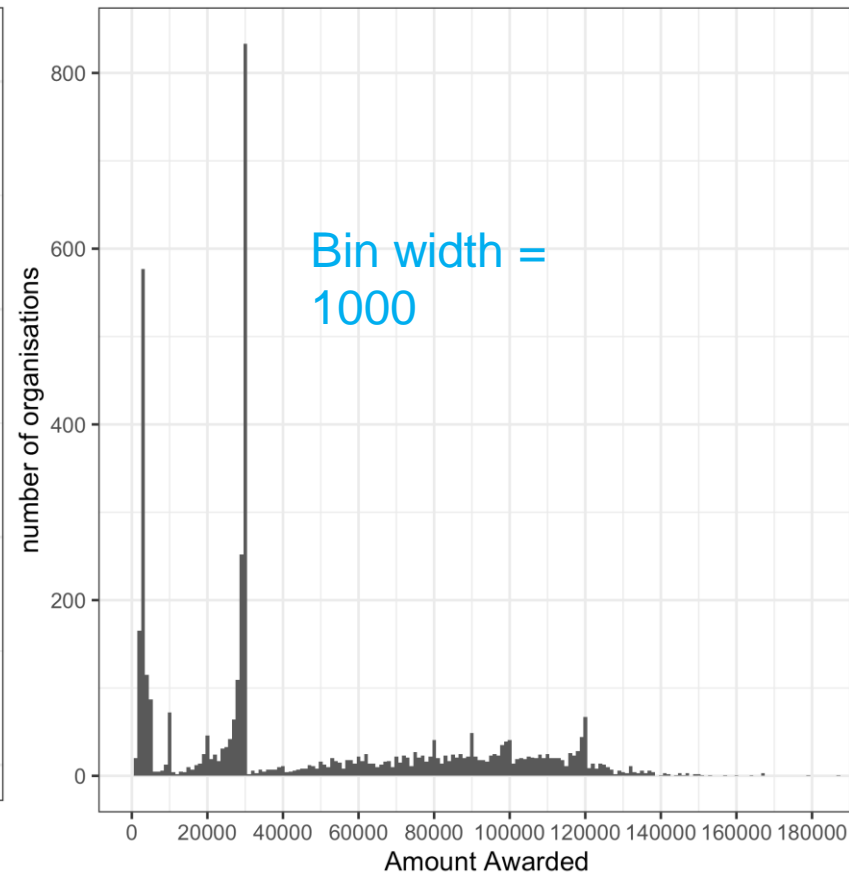


Common pitfalls: bin width

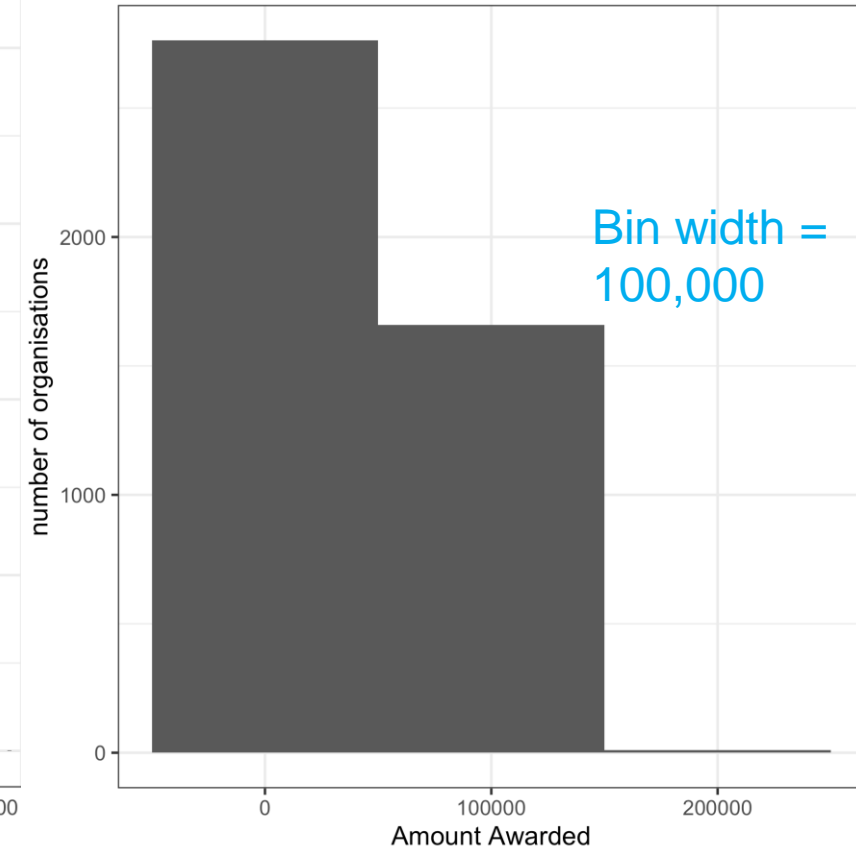
BBC Children in Needs Grants



BBC Children in Needs Grants



BBC Children in Needs Grants



BREAK



COMPARISON

RELATIONSHIP

What would you like to show?

DISTRIBUTION

COMPOSITION

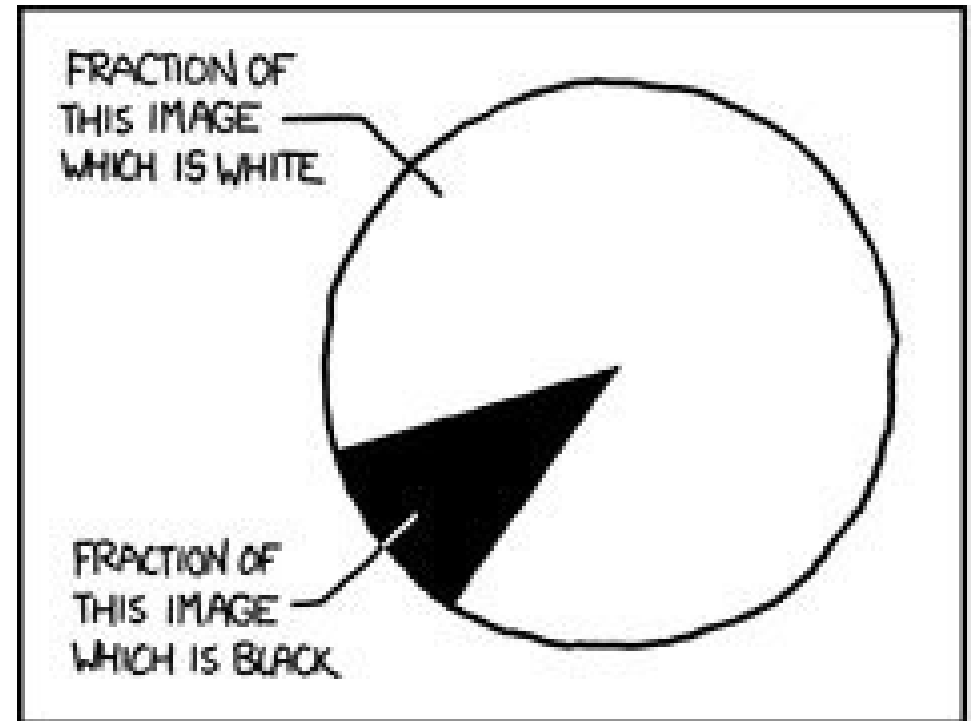


Understanding composition



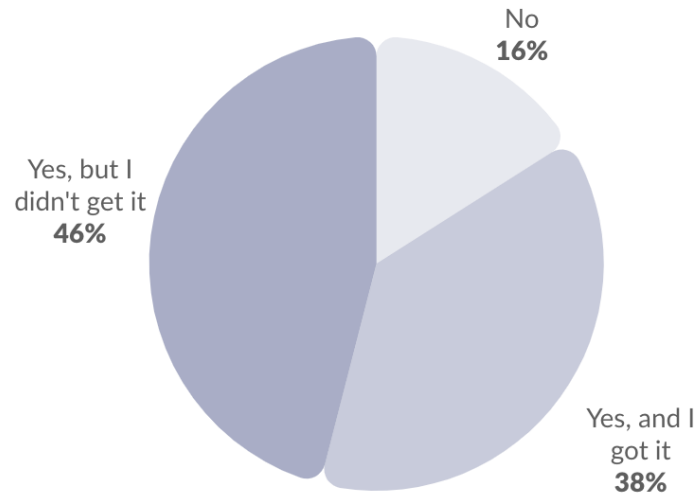
Pie charts and donut charts

- **What is it?**
 - A circle divided into sections where each piece of the pie represents the proportion
 - Donut chart uses ring vs circle
- **Input data:** Proportion/Percentage for each value in a category
- **Why use it?** To compare proportions



Pie charts and donut charts

In the past 12 months, have you wanted counseling from a mental health professional?



How much of our support went to organisations led by Black, Asian and minority ethnic communities?

We wanted to show what proportion of these grants went to organisations which self-defined as "BAME-led" when they applied for funding.



- Best used with small datasets
- Best used with simple fractions ($1/10$, $1/2$, $1/4$)



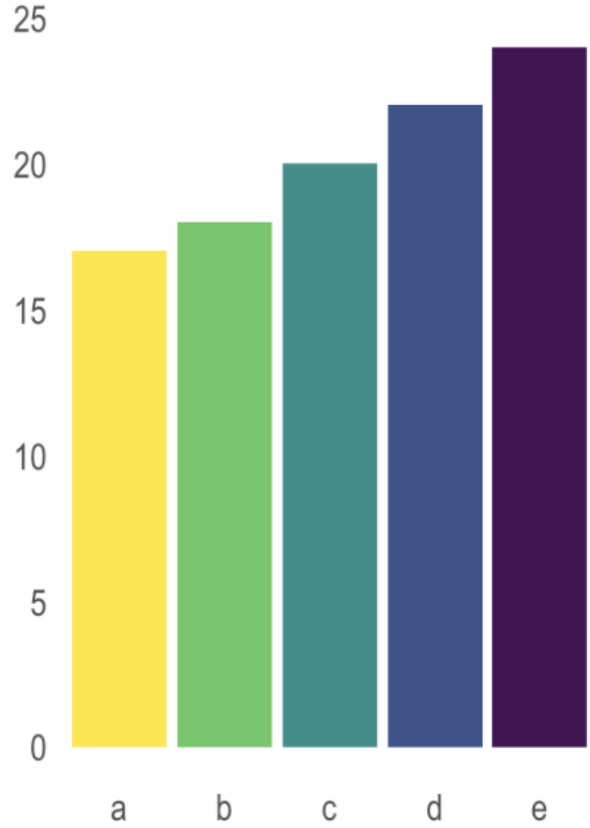
Pie charts – What does this tell you?



- Which slice is the biggest?
- Can you order these by value?



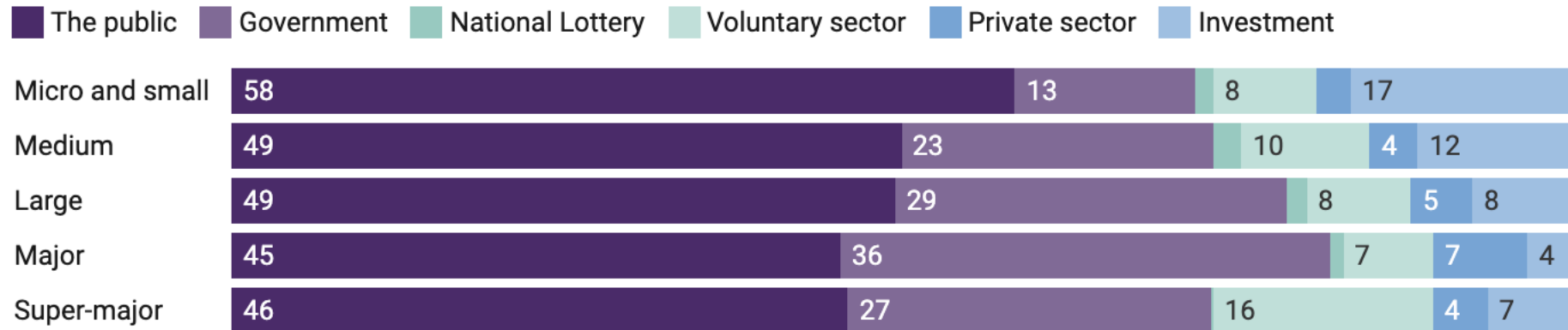
Pie chart alternatives: bar charts



Pie chart alternatives: stacked bar charts

Micro and small organisations receive a larger proportion of income from the public than other organisations

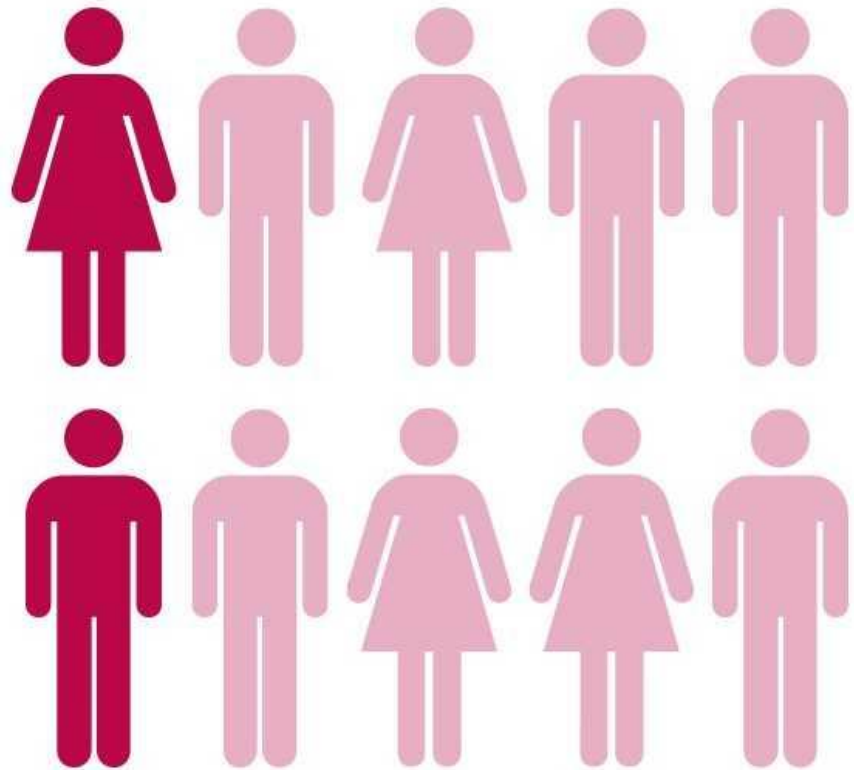
Breakdown of income by source and size of organisation, 2017/18 (%)



Source: NCVO, Charity Commission • [Get the data](#) • Created with [Datawrapper](#)



Pie chart alternatives: pictorial fraction chart



Only

19%



are aware that
people living with HIV
and on effective
HIV treatment
can't pass it on.

Reference: YouGov survey of 2,075 GN adults, June 2019



Understanding relationships



Scatterplots, line charts, & area charts

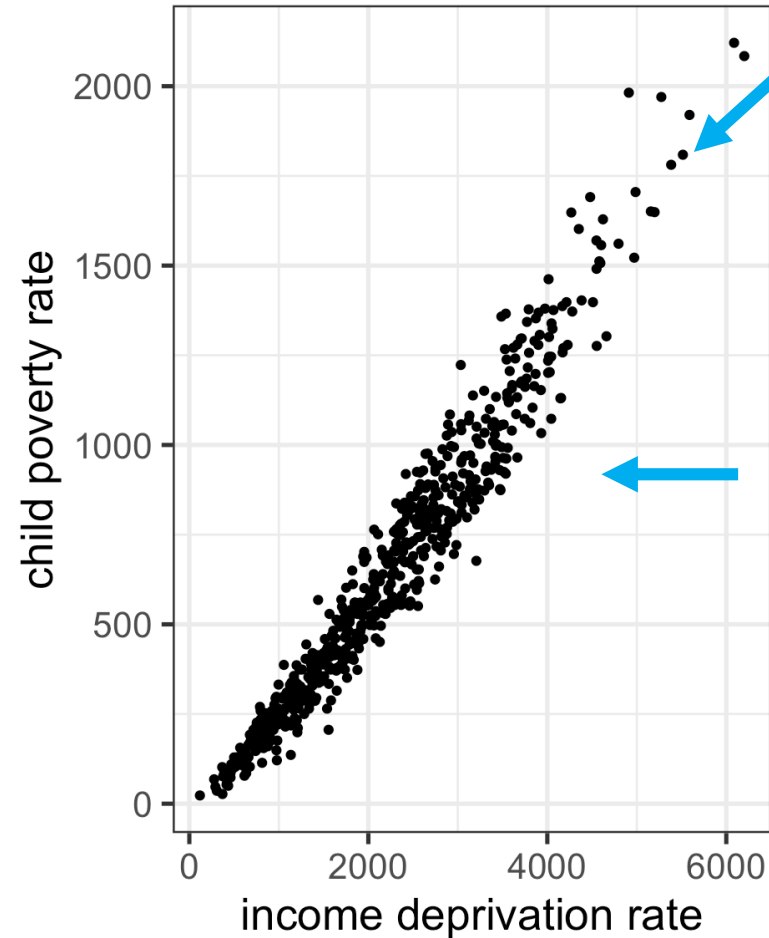
- **Scatterplots:** show how one factor (variable) changes as another changes, e.g. weight and height
- **Input data:** Two numeric variables
- **Why use it?**
 - To understand relationships between two factors



Scatterplots

income_deprivation	child_poverty
3036	1058
2891	1057
3130	1082
2377	822
2061	640
2931	1036
3536	1366
2924	997
3039	1041
1683	493
2718	955
2485	830
2613	929

Each row is data from one ward in London

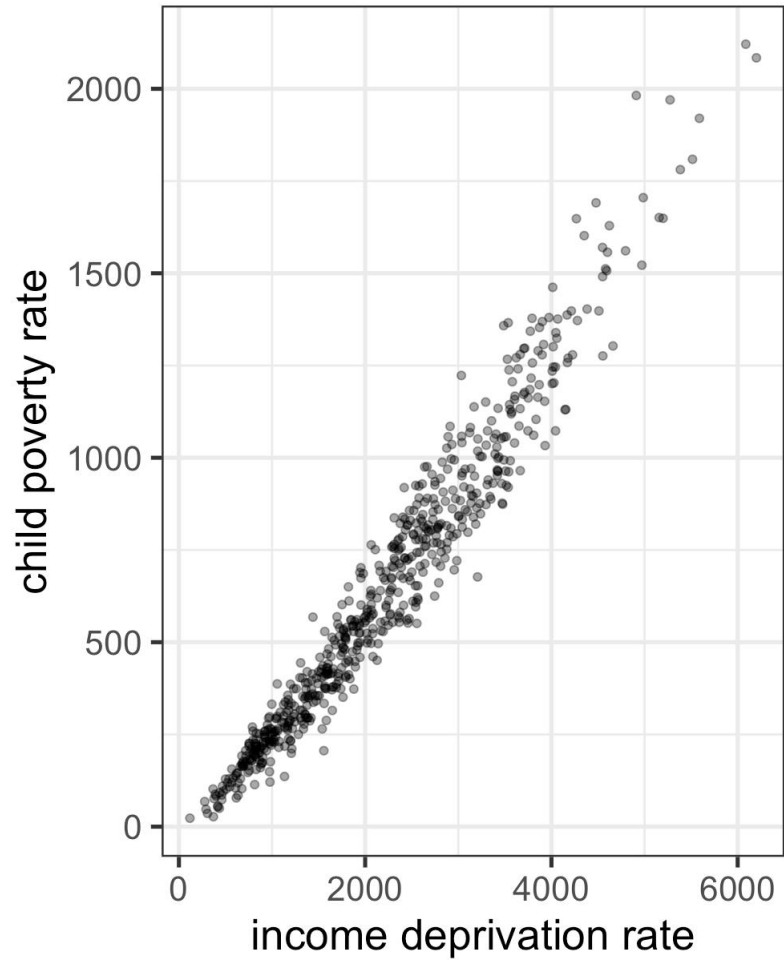
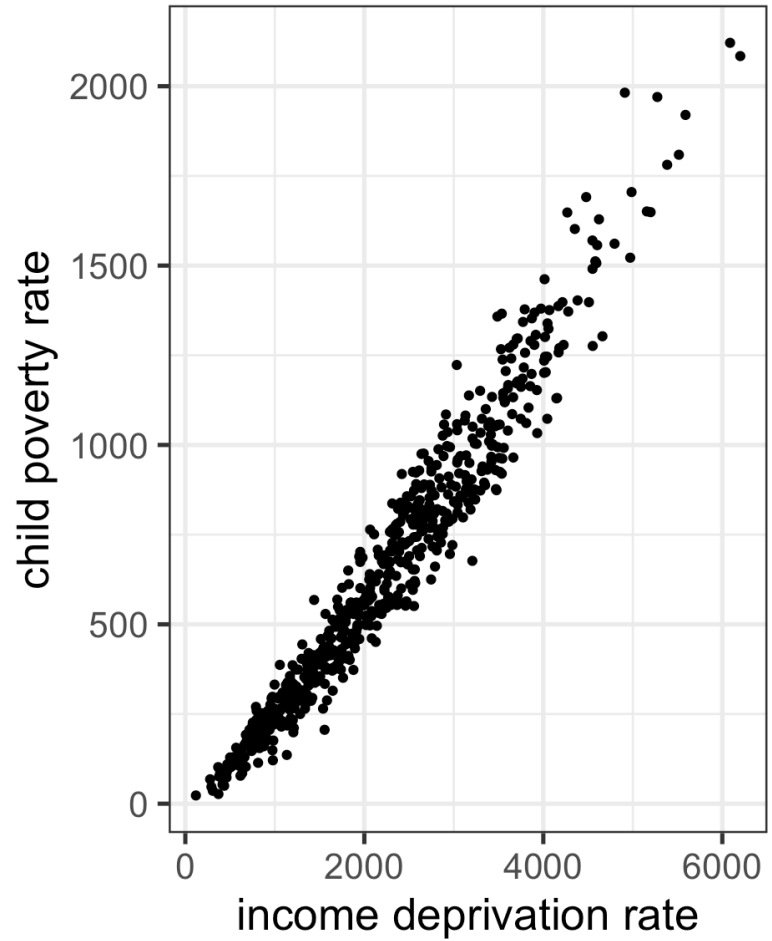


Each point is one row or sample

Overall pattern can tell you the type of relationship



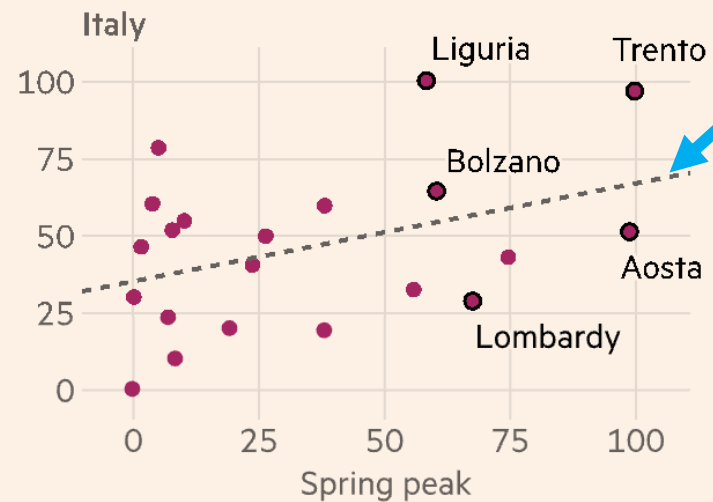
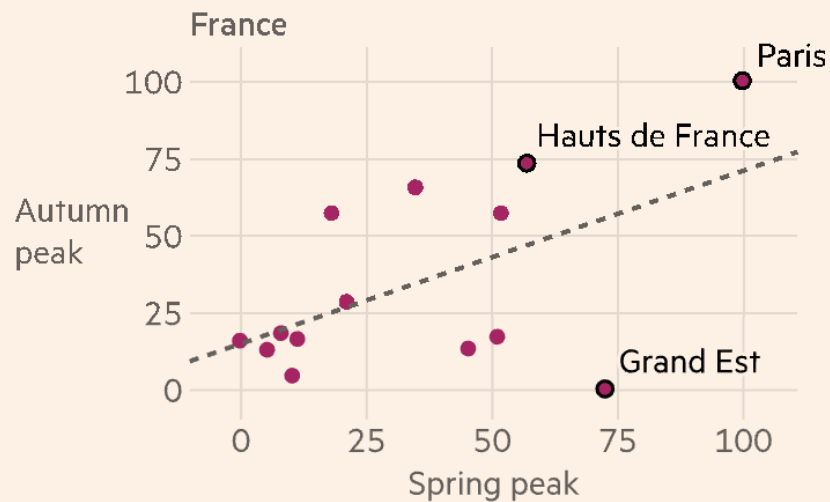
Common pitfalls: overlapping points



Scatterplots

On average, places that were hit hardest in the spring are suffering the most in the autumn

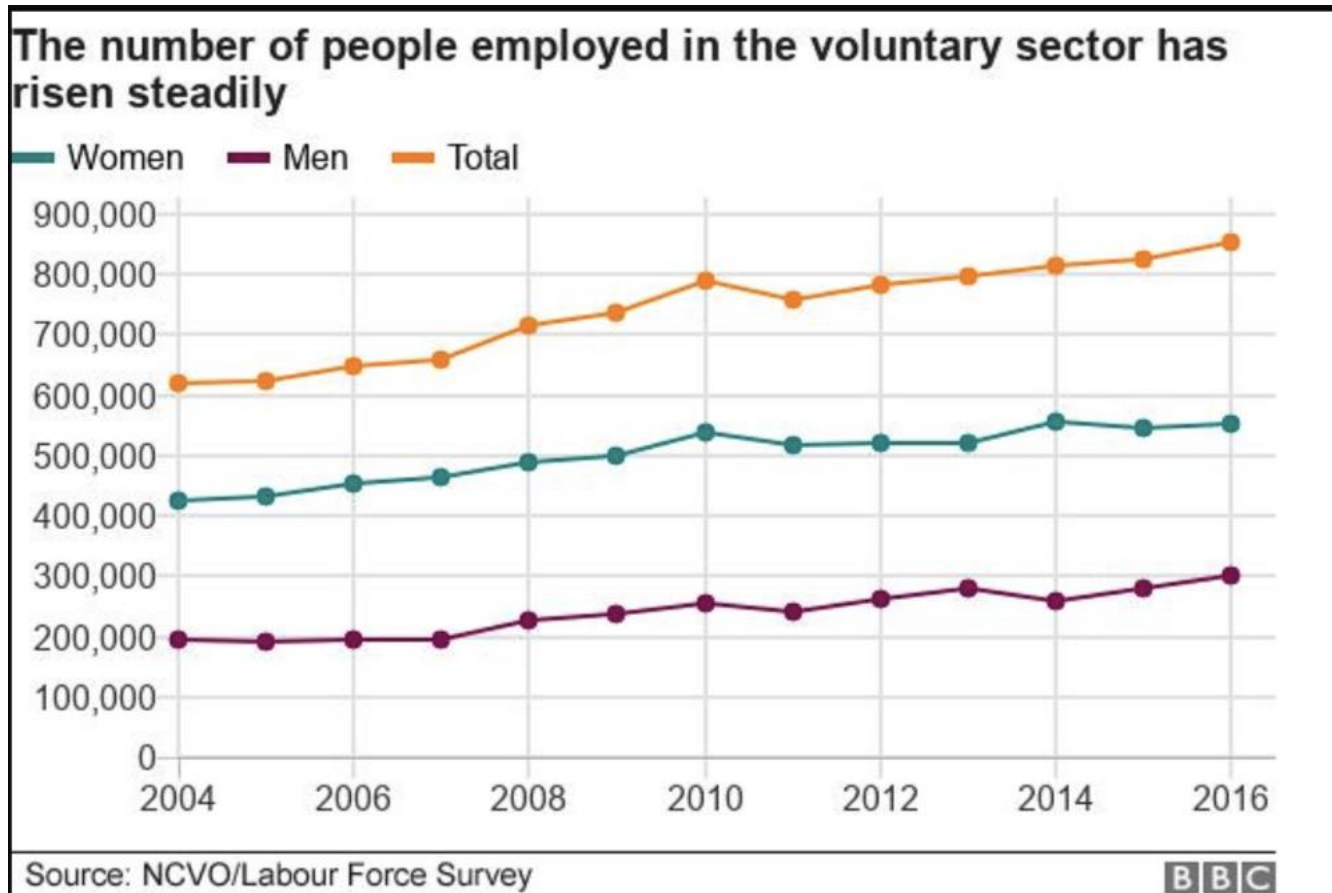
Weekly cases per 100,000 people*, spring peak vs autumn peak. Each circle represents one region



Line of best fit can help you compare trajectories



Line charts

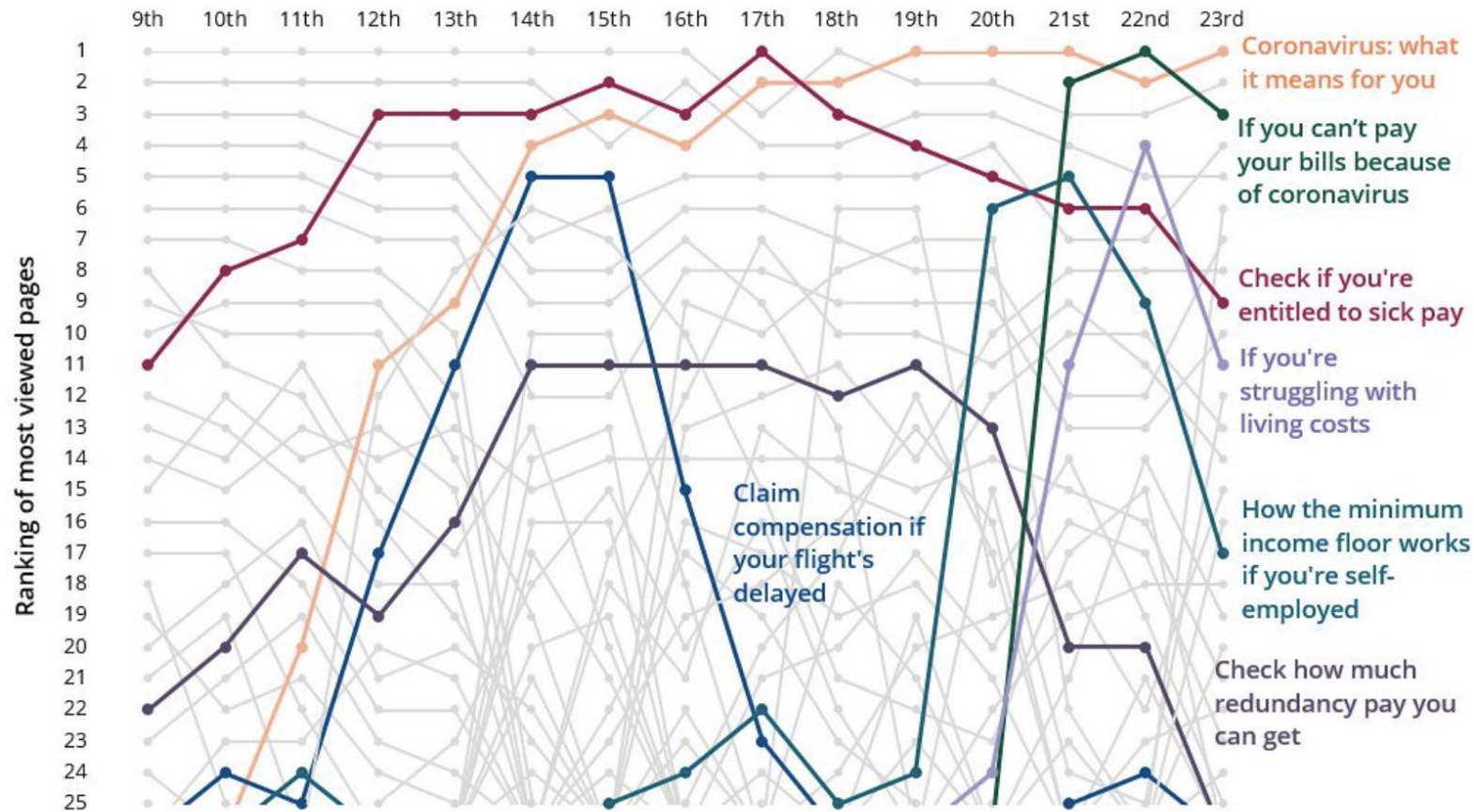


- Like a scatterplot, but the markers are connected by a line
- The x-axis (horizontal axis) values are ordered, often over time



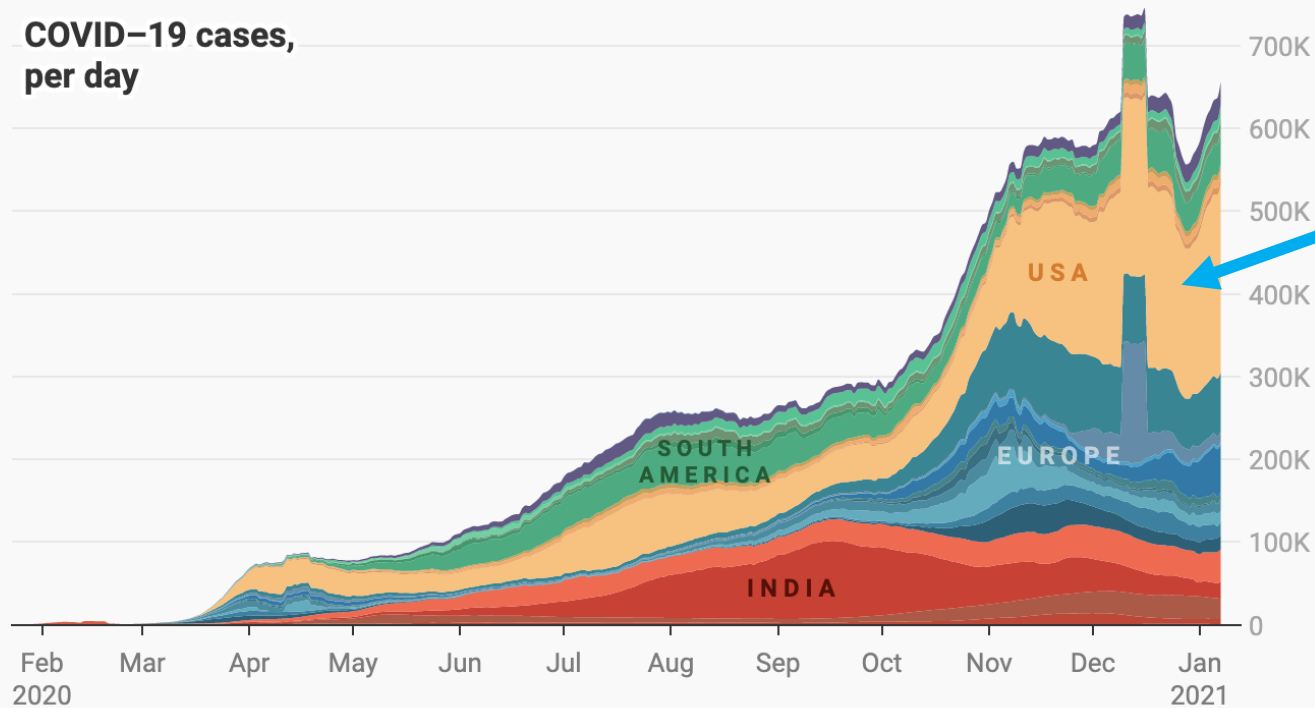
Common pitfalls: Spaghetti

Citizens Advice top 25 web pages March 2020



Area charts

COVID-19 cases,
per day



Like a line chart but
the area is filled in to
highlight the trend

Seven day rolling average of number of people confirmed to have COVID-19, per day (not including today). This chart gets updated once per day with data by Johns Hopkins. Johns Hopkins university doesn't provide reliable data for March 12 and March 13.

Source: [Johns Hopkins CSSE](#) • [Get the data](#) • Created with [Datawrapper](#)



**Pause for
Questions!**

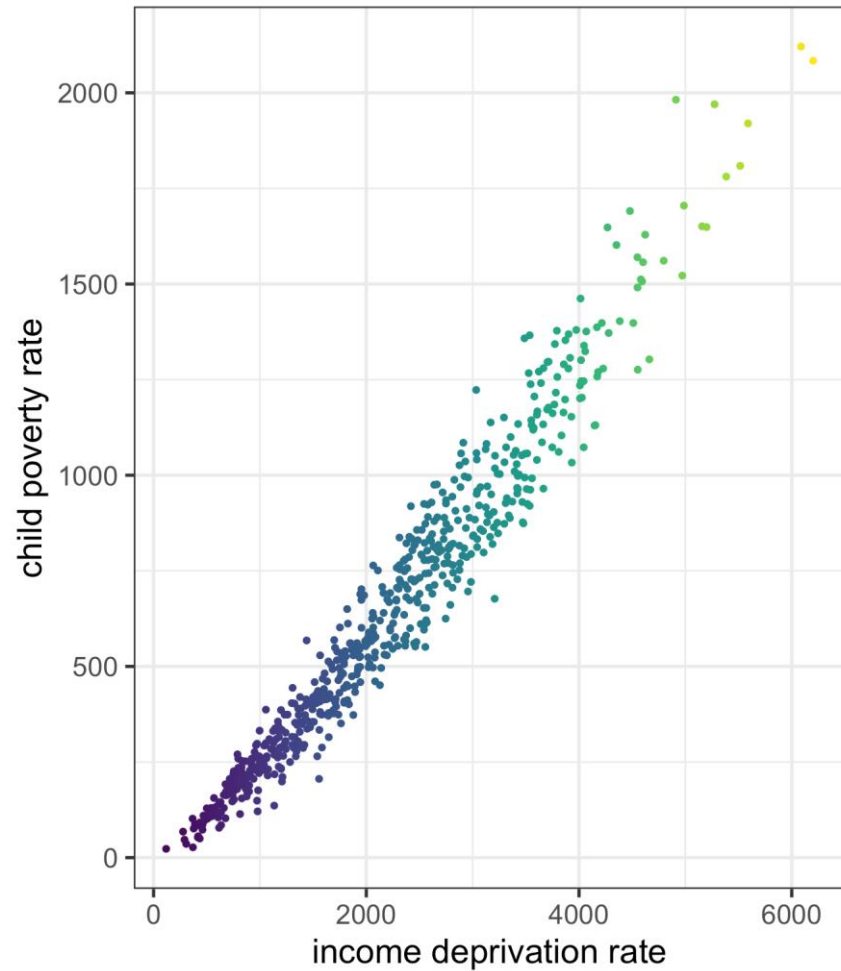


Get feedback & Refine

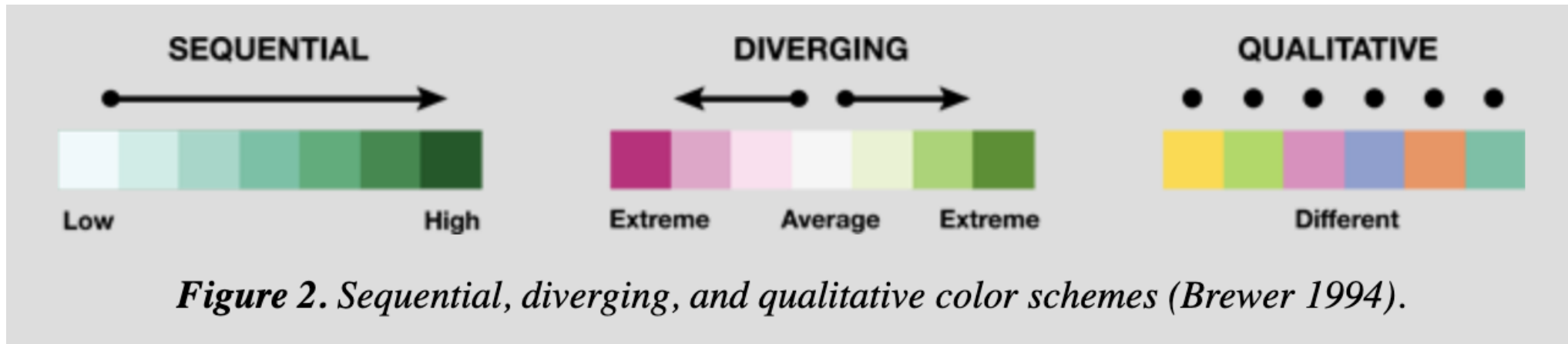
- ↘ Does every part of the figure have a purpose?
- ↘ What can you remove?
- ↘ Do you need additional text for storytelling?
- ↘ Consider colours
 - ★ Use consistently
 - ★ Use to highlight important information
 - ★ Consider accessibility



Avoid colouring for the sake of it



Use colour scales appropriately



<https://cartographicperspectives.org/index.php/journal/article/view/1538/1819>



Shots given

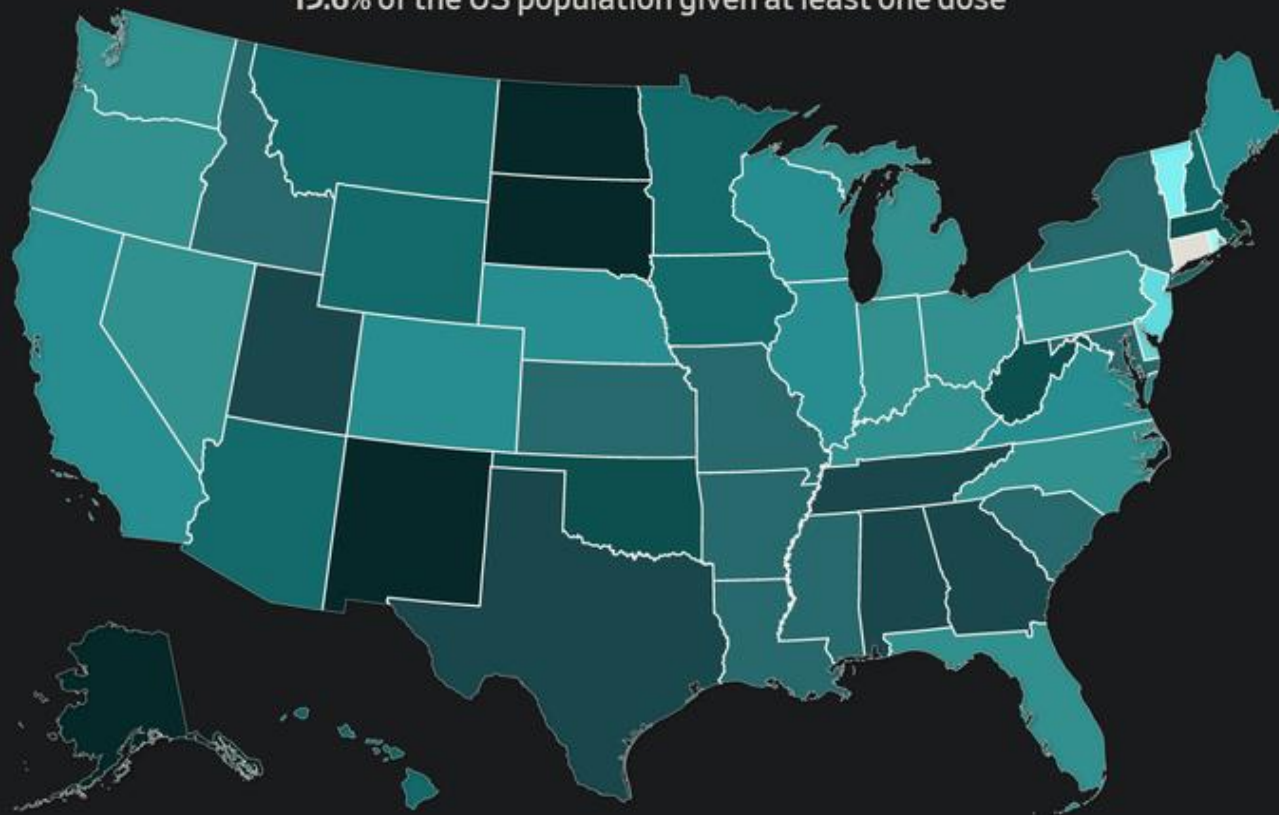
Doses distributed

Doses used

Share of state population given at least one dose

14 15 16 17 18 19%

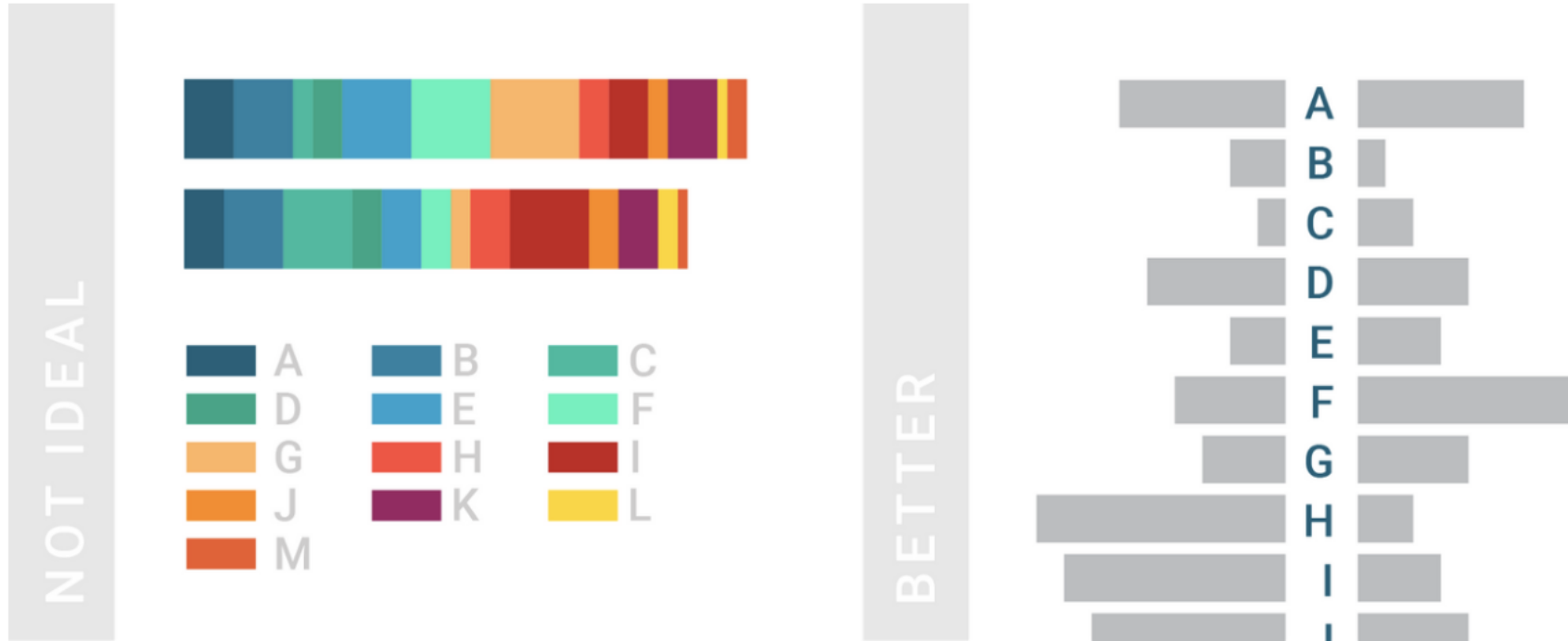
15.6% of the US population given at least one dose



Note: Total includes U.S. territories and federal agencies. Last updated March 2, at 5:00 p.m.
Source: Centers for Disease Control and Prevention



Avoid colouring too many categories



Example of an ideal and non-ideal chart. (Chartable.com, 2018)



Design for colour-vision deficiency

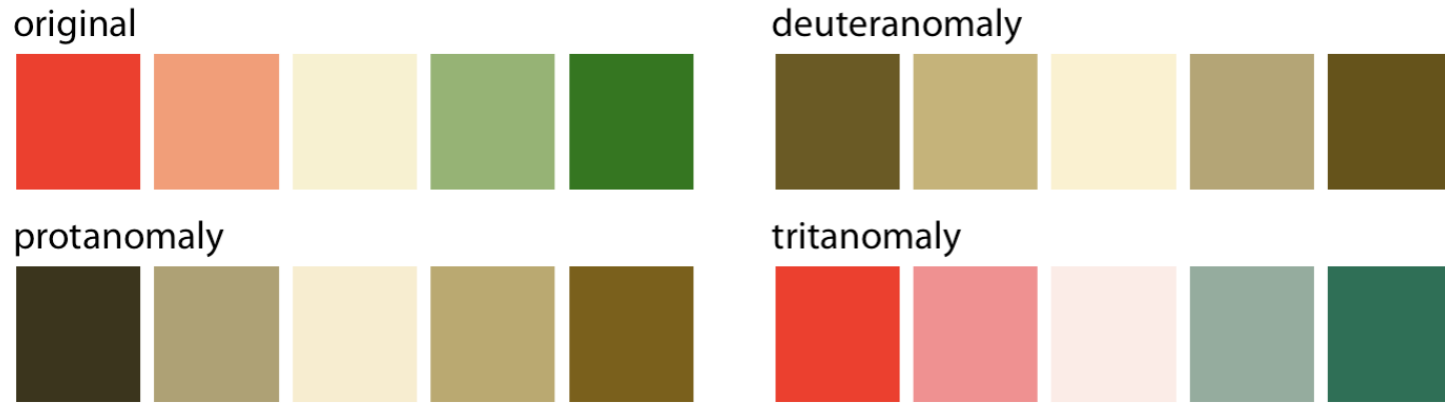


Figure 19.7: A red–green contrast becomes indistinguishable under red–green cvd (deuteranomaly or protanomaly).

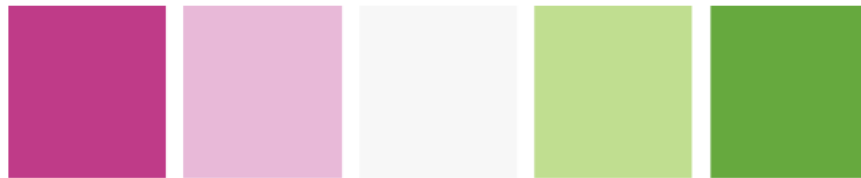


Figure 19.8: A blue–green contrast becomes indistinguishable under blue–yellow cvd (tritanomaly).

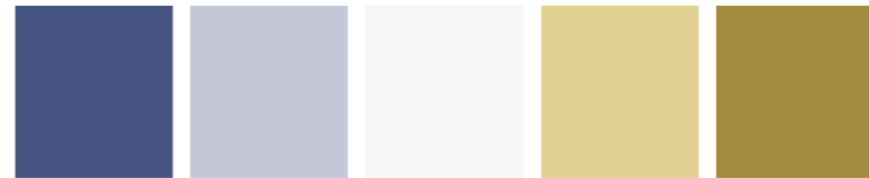


Design for colour-vision deficiency

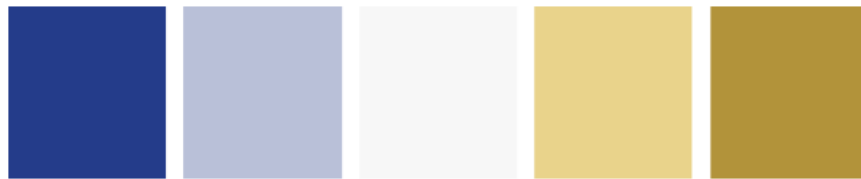
original



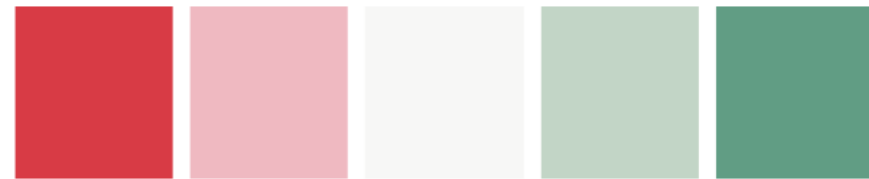
deuteranomaly



protanomaly



tritanomaly



<https://colorbrewer2.org>



<https://clauswilke.com/dataviz/color-pitfalls.html#not-designing-for-color-vision->

Treat your colours fairly*

*unless you deliberately don't want to



NOT IDEAL



BETTER



Treat your colours fairly*

*unless you deliberately don't want to

rainbow scale



rainbow converted to grayscale



<https://colorbrewer2.org>



Accessibility

- ↘ Consider using an extension like [colourblinding](#) to check colours
- ↘ Use shapes/line types in addition to colour
- ↘ Use alt-text and appropriate legends to explain your figures
- ↘ Use bigger font size for titles/labels



Accessibility

↘ <https://accessibility.blog.gov.uk/2016/09/02/dos-and-donts-on-designing-for-accessibility/>

Designing for users on the autistic spectrum

Do...	Don't...
use simple colours 	use bright contrasting colours 
write in plain English Do this.	use figures of speech and idioms 
use simple sentences and bullets 	create a wall of text 
make buttons descriptive 	make buttons vague and unpredictable 
build simple and consistent layouts 	build complex and cluttered layouts 

Designing for users of screen readers

Do...	Don't...
describe images and provide transcripts for video <code><alt></code>	only show information in an image or video 
follow a linear, logical layout 	spread content all over a page 
structure content using HTML5 <code><h1></code> <code><nav></code> <code><label></code>	rely on text size and placement for structure 36pt, bold Header 
build for keyboard use only 	force mouse or screen use 
write descriptive links and headings Contact us	write uninformative links and headings Click here

Designing for users with low vision

Do...	Don't...
use good colour contrasts and a readable font size Aa	use low colour contrasts and small font size Aa
publish all information on web pages 	bury information in downloads 
use a combination of colour, shapes and text Start >	only use colour to convey meaning 
follow a linear, logical layout 200% magnification 	spread content all over a page 200% magnification 
put buttons and notifications in context 	separate actions from their context 



Making your data viz

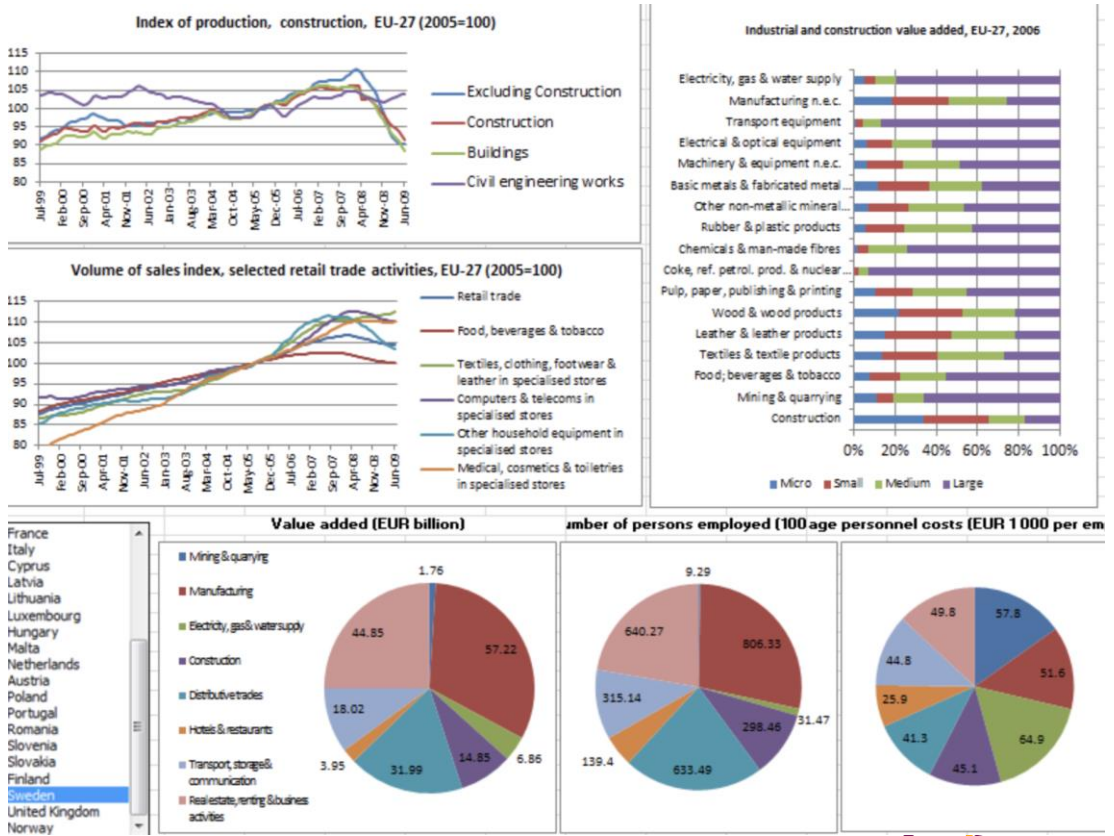


The best visualization software is the one that allows you to make the figures you need.





- **Good for:**
 - Quick basic charts (bar charts, scattercharts)
 - Can make dashboards
 - Using data you already have in excel
- **Limitations:**
 - Data has to be compiled/formatted, ready for visualisation
 - Some limits to types of charts
- **Cost:** Office 365 free for charities





- **Good for:**
 - Dashboards
 - Maps
 - Interactivity
 - Multiple charts
 - Can be embed in websites
- **Limitations:**
 - Might be more than you need
- **Cost:** Tableau Desktop free for charities

Client profiles by issues for the last 12 months compared to census 2011



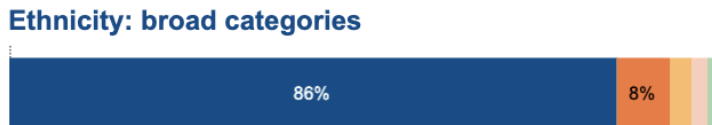
Region: (Multiple values) Category: (All) Subjects: (All)

Use this page to view the profile of people filtered by category & subjects. Select the category or subjects you are interested in and compare against census 2011 profiles. If you want to know more about the issues certain groups of people go to the issues by profile sheet.

Profile for selected issues



Census 2011 for comparison

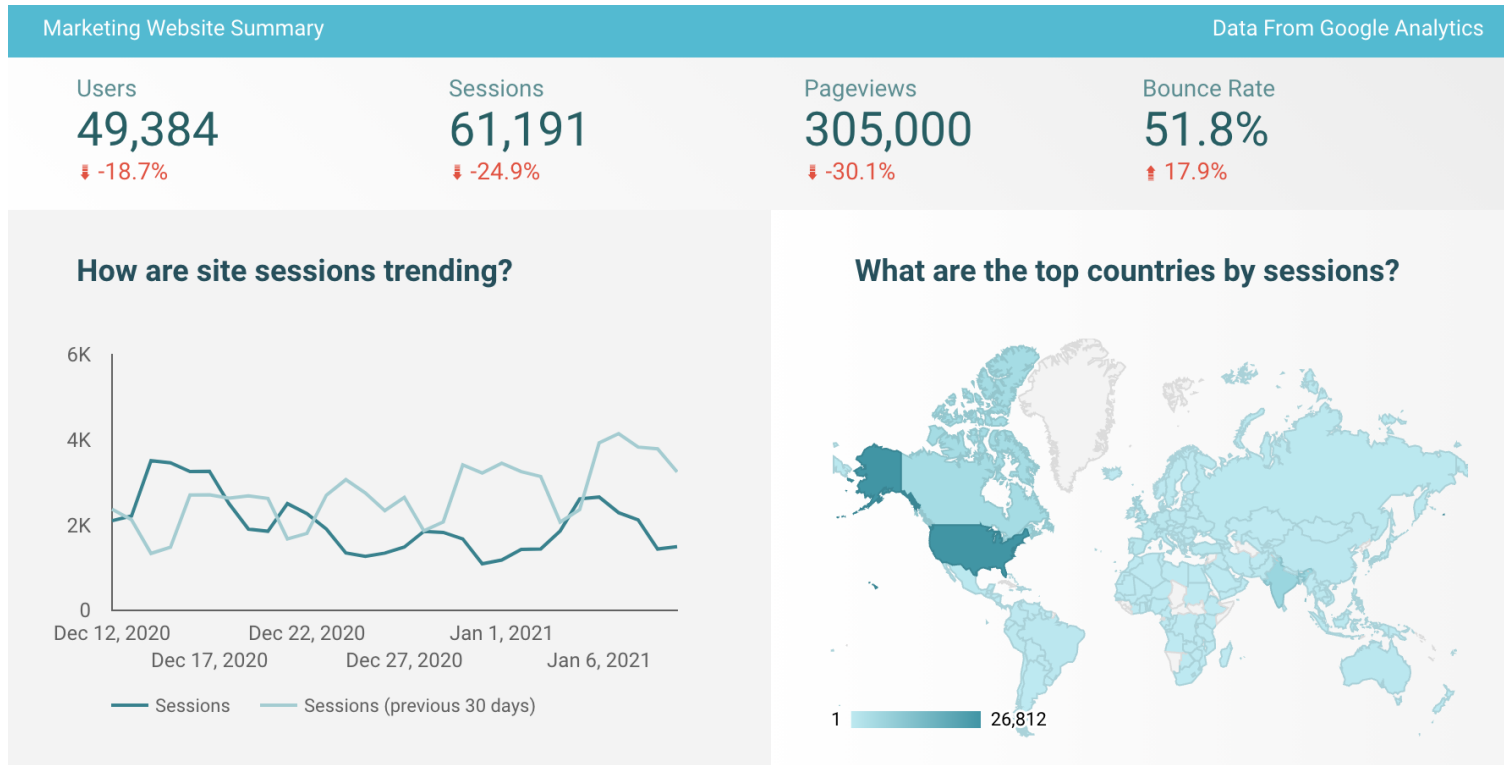


<https://public.tableau.com/en-gb/gallery/?tab=viz-of-the-day&type=viz-of-the-day>



Data Studio

- **Good for:**
 - Dashboards
 - Integrates with other google tools
- **Limitations:**
 - Fewer visualisation options than Tableau
- **Cost:** Gsuite is free for charities



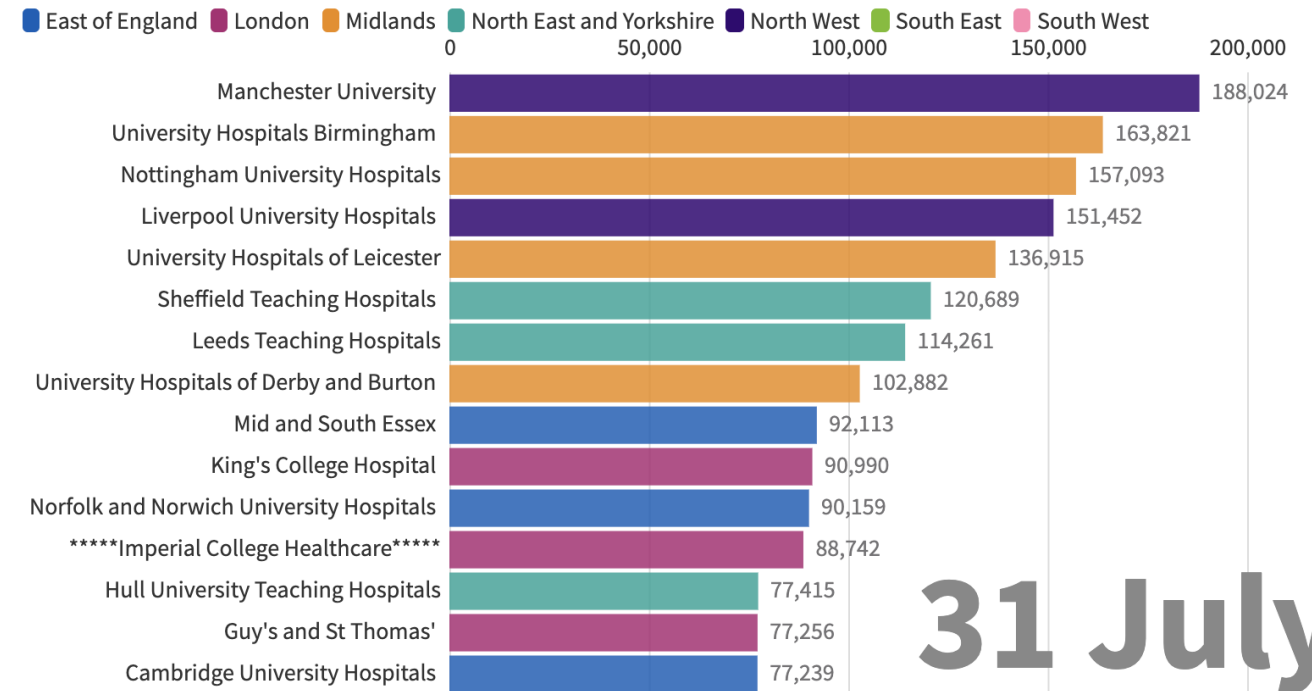
<https://datastudio.google.com/navigation/report>



- **Good for:**
 - Storytelling and presentations
 - Interactive charts
 - Embedding in websites
- **Limitations:**
 - Free plan makes data and projects public
- **Cost:** Free option, non-profit discount

Top 15 NHS Trusts with Covid-19 staff shortages

Day-by-day cumulative numbers of staff on Covid-19 related sick leave by NHS Trust



31 July

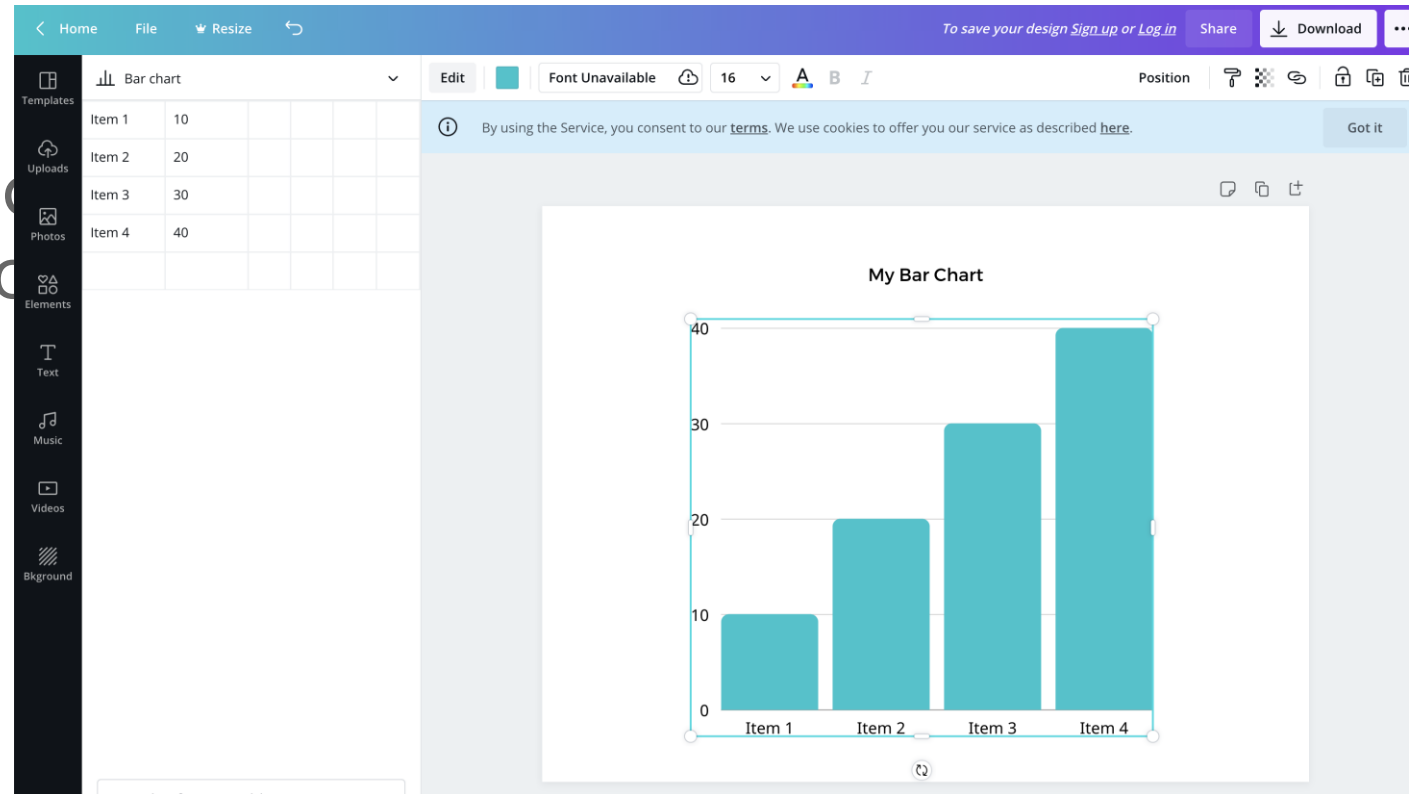
Replay

<https://www.threesixtygiving.org/2020/11/19/my-first-data-expedition-a-workshop-without-a-post-it-note-in-sight/>





- **Good for:**
 - Helps you choose your graphs
 - Simple charts/infographics
 - Drag-and-drop
- **Limitations:**
 - Data pasted from spreadsheet
- **Cost:** Free for non-profits



<https://www.canva.com/graphs/>



infogram

- **Good for:**
 - Interactive charts and maps
 - Presentations
 - Drag and drop
- **Limitations:**
 - Free version makes visualisations public
- **Cost:** Free option



<https://infogram.com/>



- **Good for:**
 - Infographics
 - Posters
 - Drag and drop
- **Limitations:**
 - Less suited to integrating with data sources
- **Cost:** Free option, non-profit discount



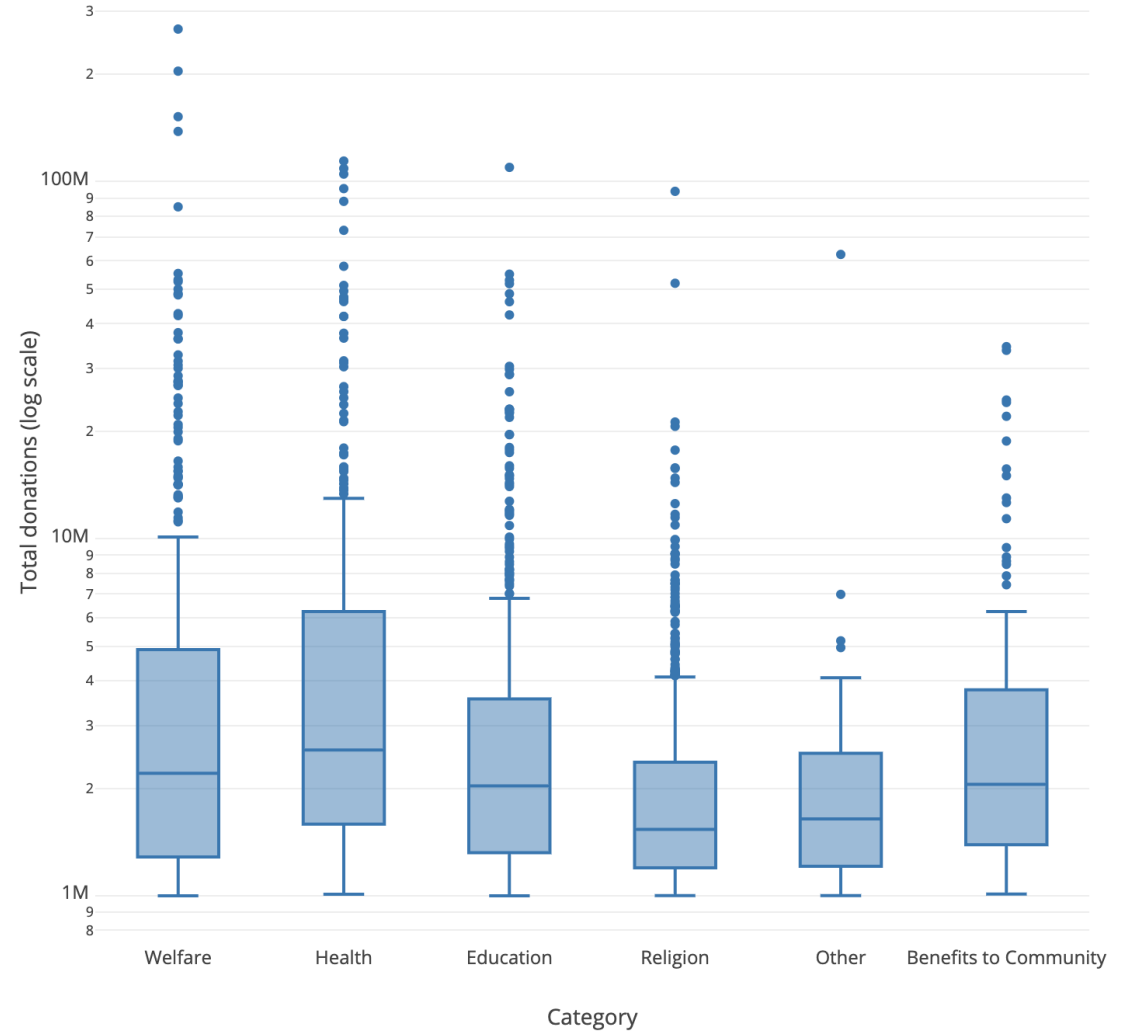
<https://piktochart.com/>





- **Good for:**
 - Interactive charts
 - Embedding in websites
 - Attractive default formats
- **Limitations:**
 - Paid options is very expensive
- **Cost:** Free option, but limited to public charts

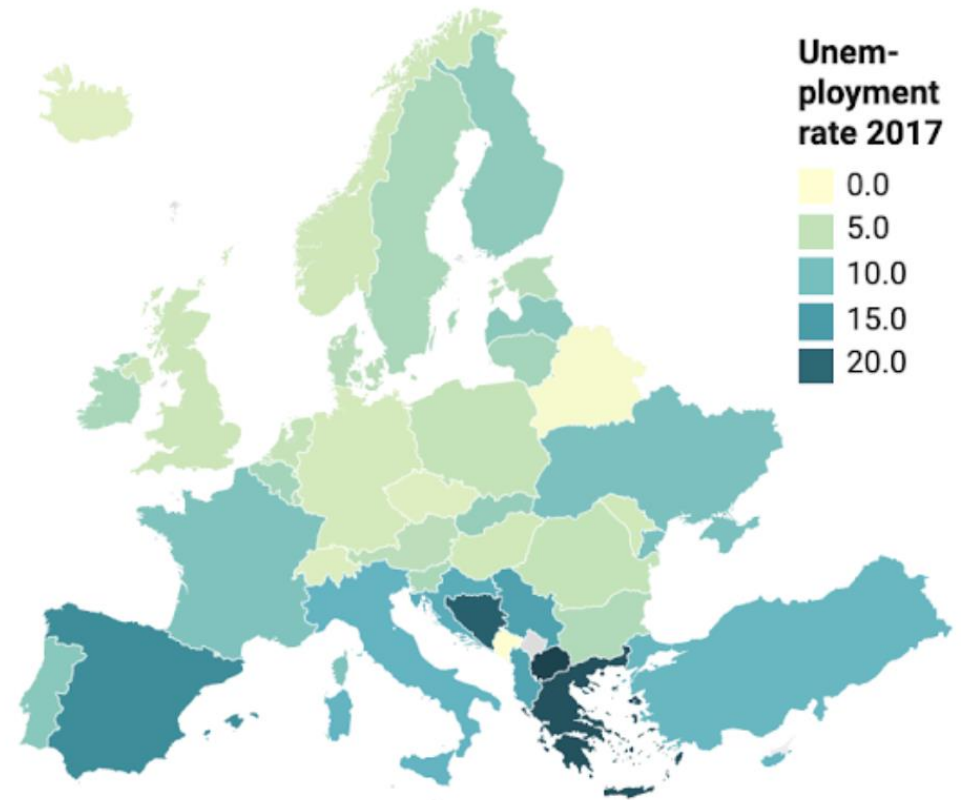
Canadian Charity Snapshot (donations of more than \$1M)



Datawrapper

- **Good for:**
 - Quick charts, maps, and tables
 - Interactive
 - Attractive default formats
- **Limitations:**
 - Limited formatting options with free version
- **Cost:** Free option

The unemployment rate is higher in South Europe than in North Europe



Map: Lisa • Source: IMF • Get the data • Created with Datawrapper

<https://www.datawrapper.de/>

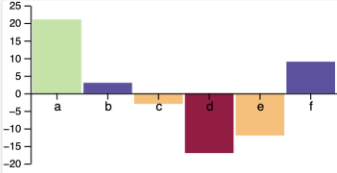


RAWGraphs

- **Good for:**
 - Lots of chart options
 - Can import data
 - Decent formatting options
- **Limitations:**
 - No maps or interactive figures
- **Cost:** Free
























2. Choose a chart

Show All charts ▾



Bar chart
It displays a categorical dimension and related amounts. Each bar represents a category, width is proportional to the quantitative dimension.

[Code](#) [Tutorial](#)

 Alluvial Diagram Correlations, proportions	 Arc Diagram Networks	 Bar chart Correlations
 Multi-set bar chart Correlations, proportions	 Stacked bar chart Correlations, proportions	 Beeswarm plot Distributions, time series, proportions
 Box plot Distributions	 Bubble chart Correlations, proportions	 Bumpchart Correlations, proportions
 Circle Packing Hierarchies, proportions	 Circular dendrogram Hierarchies, proportions	 Contour plot Correlation, density
 Convex hull Correlations, proportions	 Linear dendrogram Hierarchies, proportions	 Hexagonal binning Correlation, density
 Line chart Time series, correlations	 Matrix Plot Correlations, time series, proportions	 Radar Chart Correlations
 Sankey Diagram Flows, networks	 Streamgraph (area chart) Correlations, proportions	 Sunburst diagram Hierarchies, proportions
 Treemap Hierarchies, proportions	 Violin plot Distributions	



Datawrapper

- **Datawrapper:** <https://www.datawrapper.de/>
- **Data:** https://docs.google.com/spreadsheets/d/1kTuJ-rYtu0h8l8Cl40_w0YqSrljBO6SmfD8sgoW01rE/edit#gid=276530633



Data visualisation next steps

Resources to help with all things data...

Next steps

- ✓ Complete our session evaluation
- ✓ [Register for Datawise London support](#)
- ✓ Check training opportunities [on our Eventbrite page](#)
- ✓ [Sign up to our eNews](#)
- ✓ [Let us know if you have a digital support needs](#)



Resources

General guides

<https://clauswilke.com/dataviz/index.html>

<https://www.toptal.com/designers/data-visualization/data-visualization-best-practices>

<https://depictdatastudio.com/data-visualization-design-process-step-by-step-guide-for-beginners/>

<https://towardsdatascience.com/10-free-tools-to-instantly-get-started-with-data-visualisation-d7fadb5f6dce>

Choosing a chart

<https://www.data-to-viz.com/>

<https://datavizproject.com/>

Accessibility

<https://medium.com/nightingale/writing-alt-text-for-data-visualization-2a218ef43f81>

<https://fossheim.io/writing/posts/accessible-dataviz-design/>

<https://blog.datawrapper.de/colors/>

<https://support.infogram.com/hc/en-us/articles/360013046634-Creating-accessible-content>

<https://accessibility.blog.gov.uk/2016/09/02/dos-and-donts-on-designing-for-accessibility/>





Thank you for listening

DULCIE VOUSDEN
SORREL PARSONS

info@superhighways.org.uk

[@SuperhighwaysUK](https://twitter.com/SuperhighwaysUK)

[@DataKindUK](https://twitter.com/DataKindUK)

www.datawise.london

#DatawiseLondon

