# Getting Started with Data Visualisation

Dulcie Vousden, DataKind
UK

DISCOVER. LEARN. ANALYSE. SHAPE. REPEAT

datawise
LONDON

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

superhighways
harnessing technology for community benefit

COALITION
FOR EFFICIENCY

Mokerble

DataKind UK

hear
humanity. equality. rights.

london plus

# 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

# Exercise: What works, what doesn't?

https://jamboard.google.com/u/0/d/1DzX2m369TvBLWVtiQArADh5pjomo6G7evzwvBiaBftQ/viewer

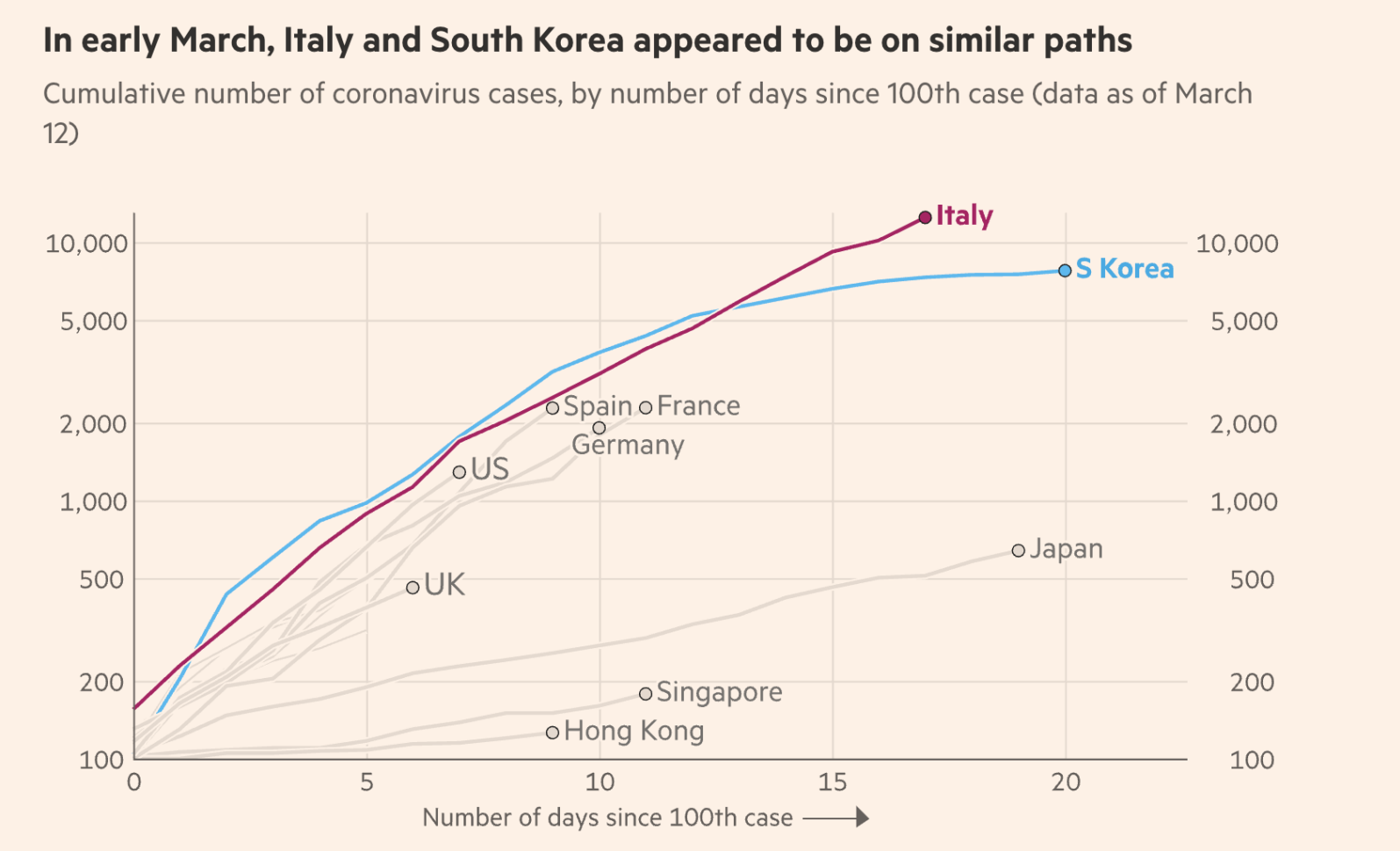
# visualization

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

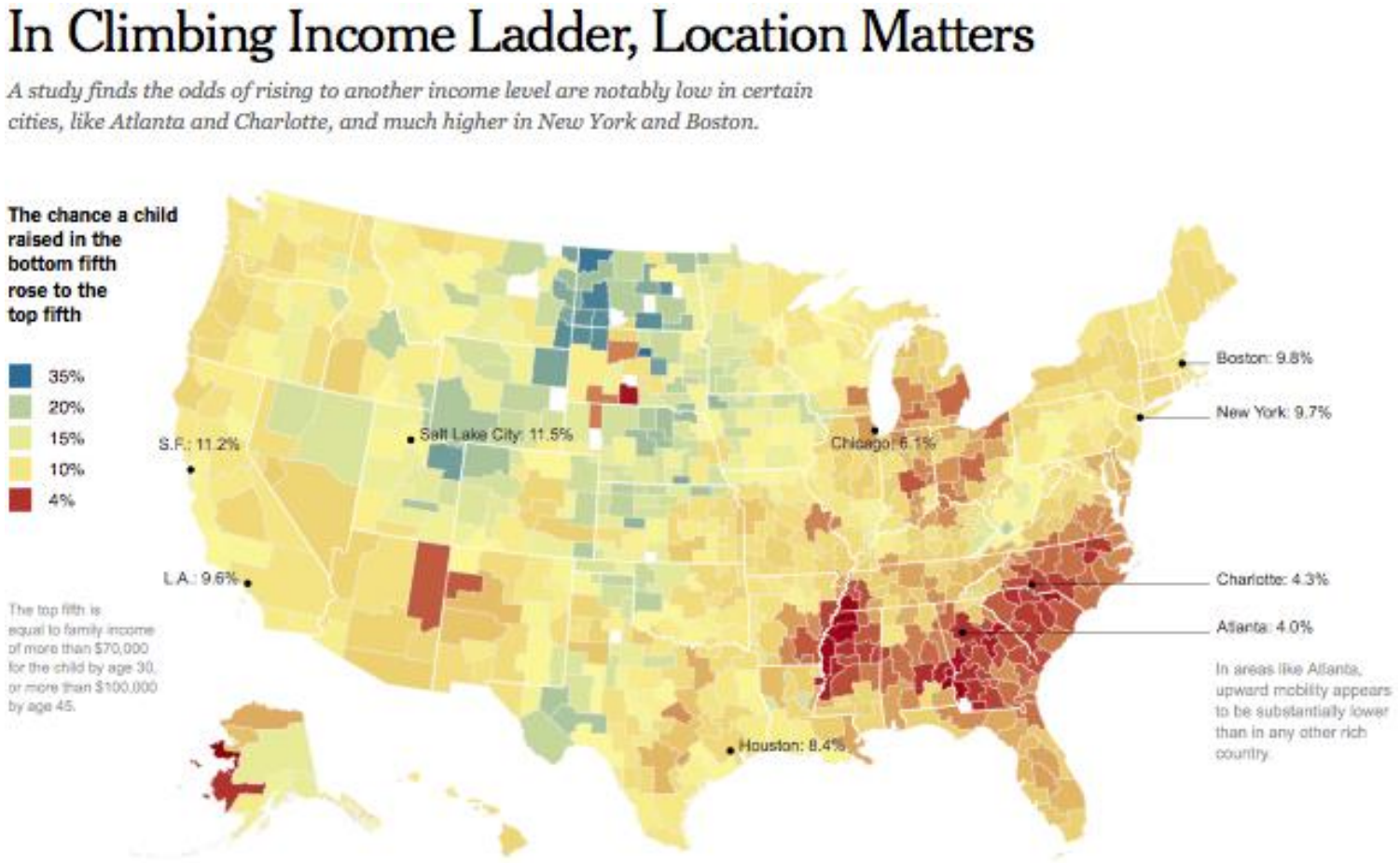
- What would the newspaper headline for this visualisation say?
- What do you want your audience to do with this information?
- What do you want your audience to conclude?
- What action do you want your audience to take?

# 1. Decide your take home message




https://ig.ft.com/coronavirus-global-d

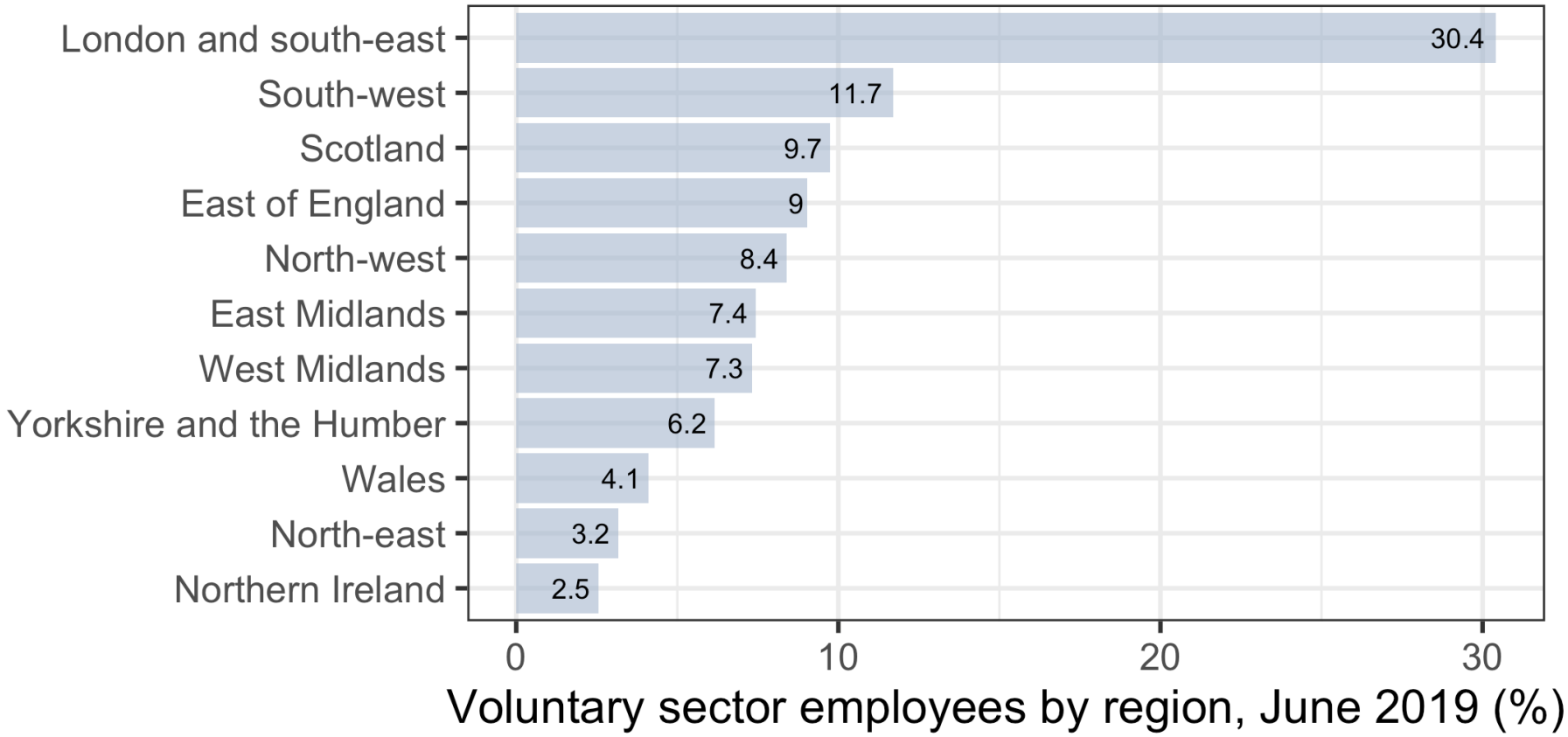
# 1. Decide your take home message




Source: MotherJones and https://depictdatastudio.com/data-visualization-design-

# Voluntary sector employees over-represented in London

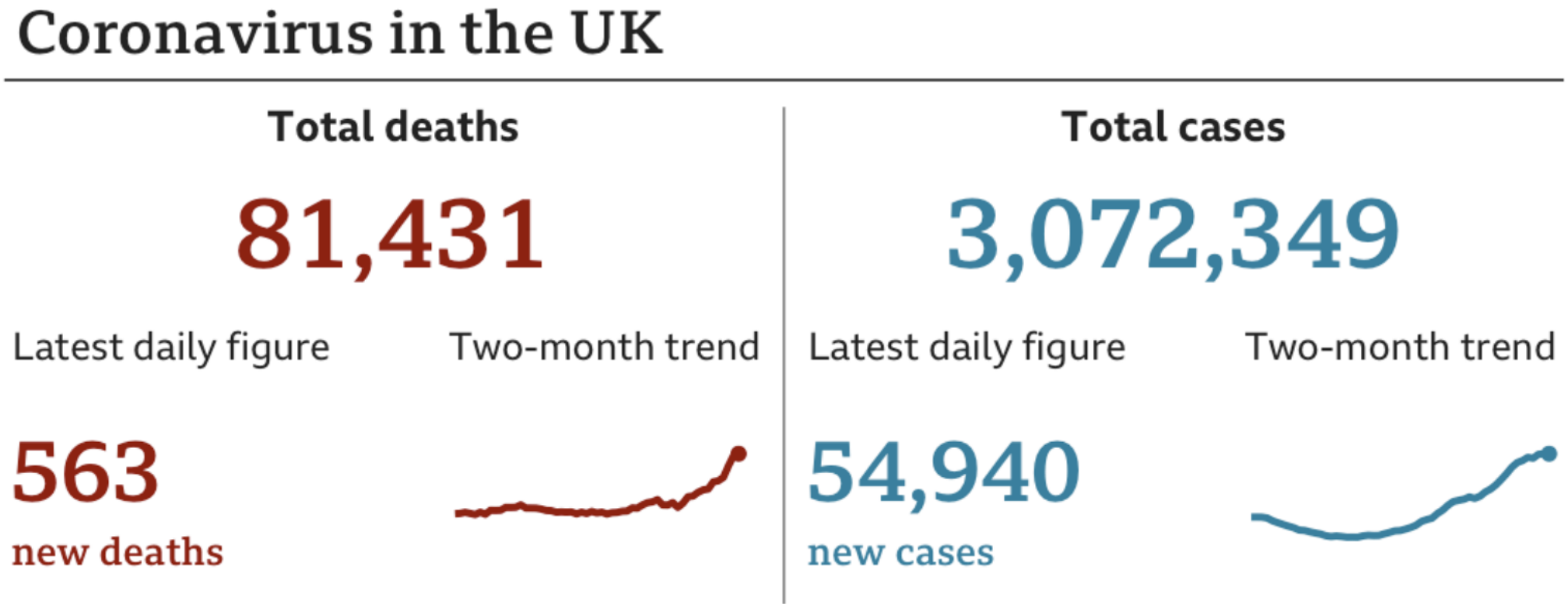
# Voluntary sector employees by region



# 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



# Exercise: Who's your audience?

**Think-pair-share**

- 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?

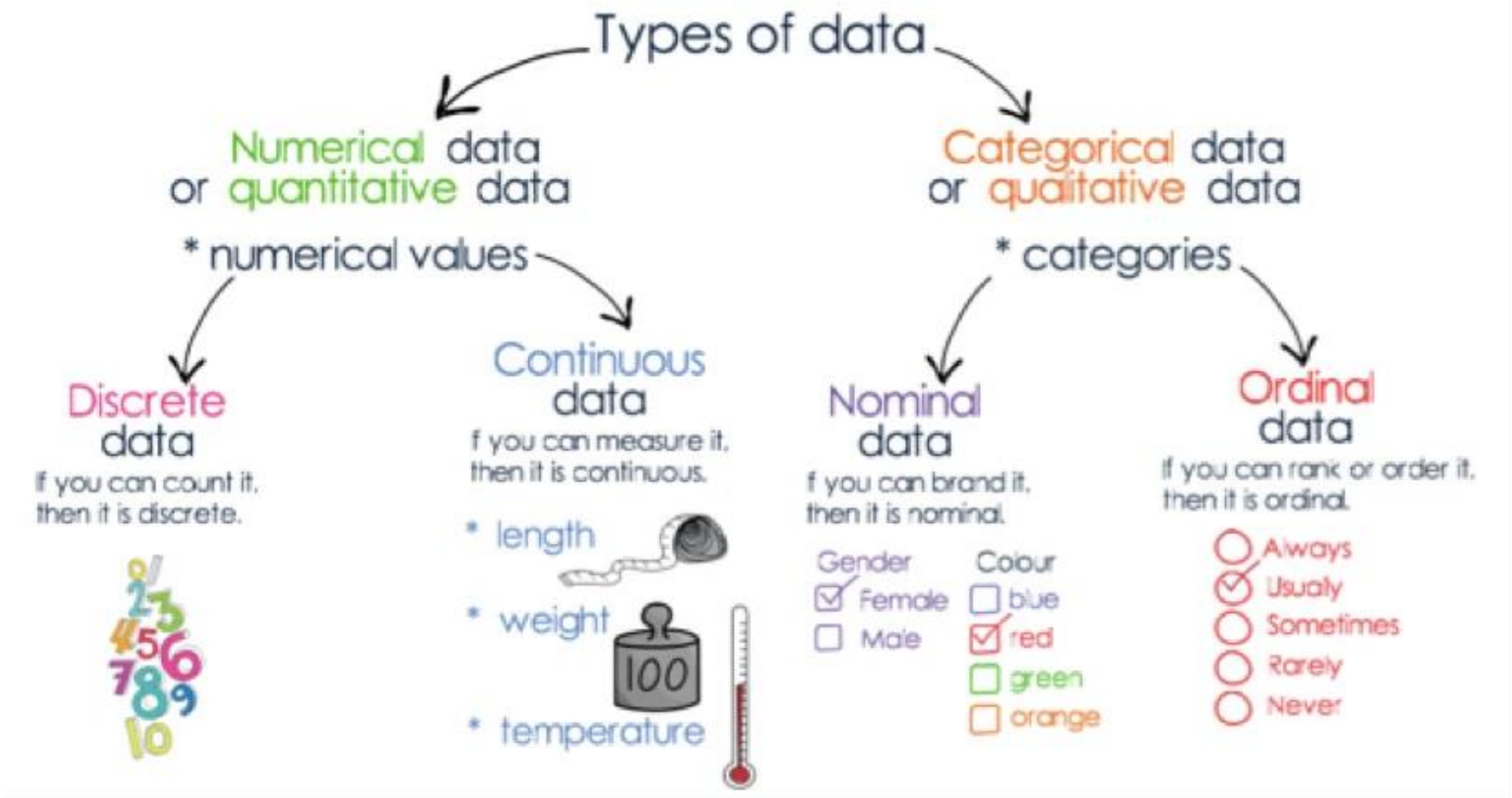
# 3. Prep the data

- Filter all but the data of interest
- Perform any summaries needed
- What level of detail do you need?
- What level of precision do you need?

# 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?

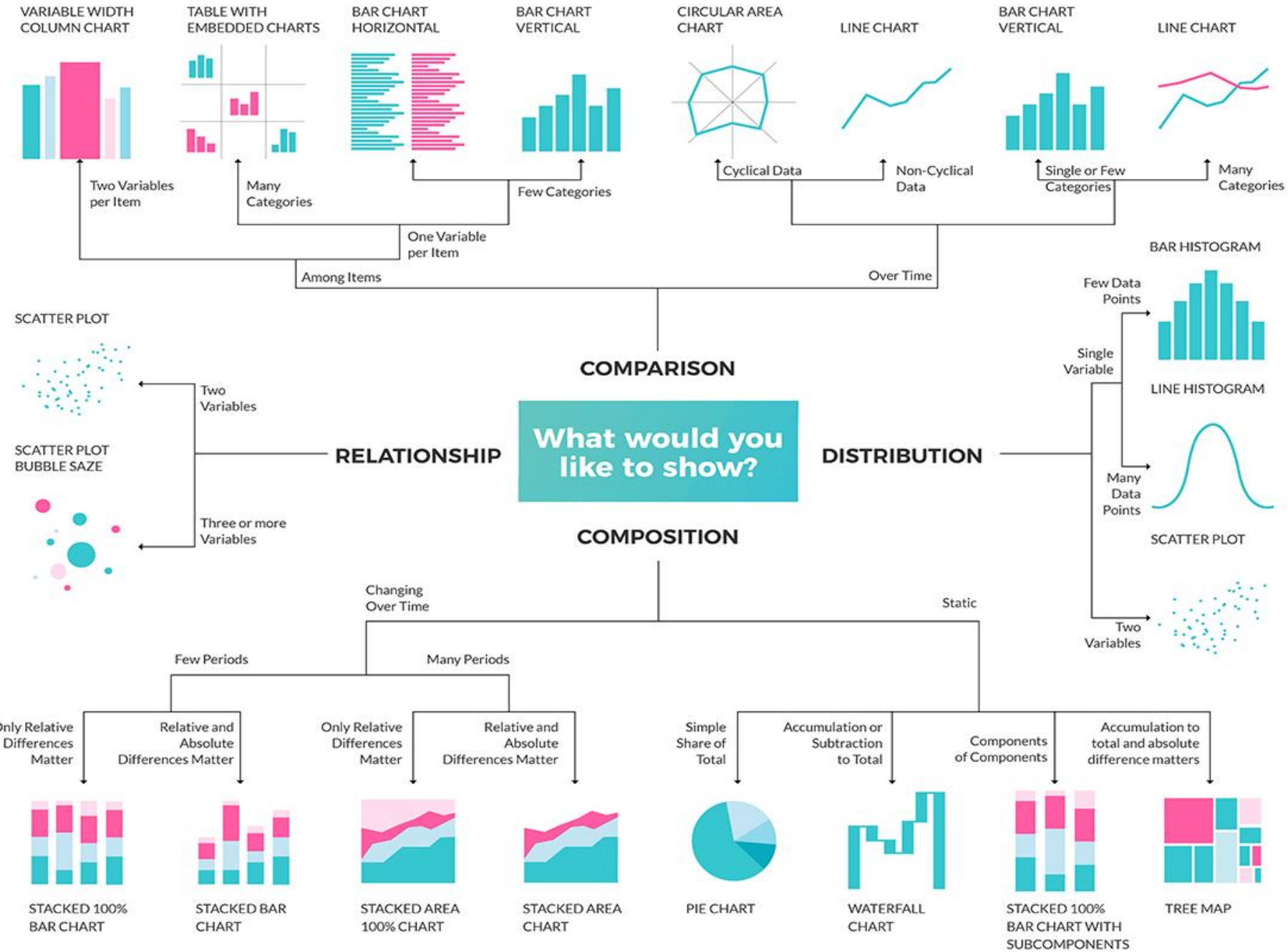



https://legac.com.au/blogs/further-mathematics-exam-revision/further-mathematics-unit-3-

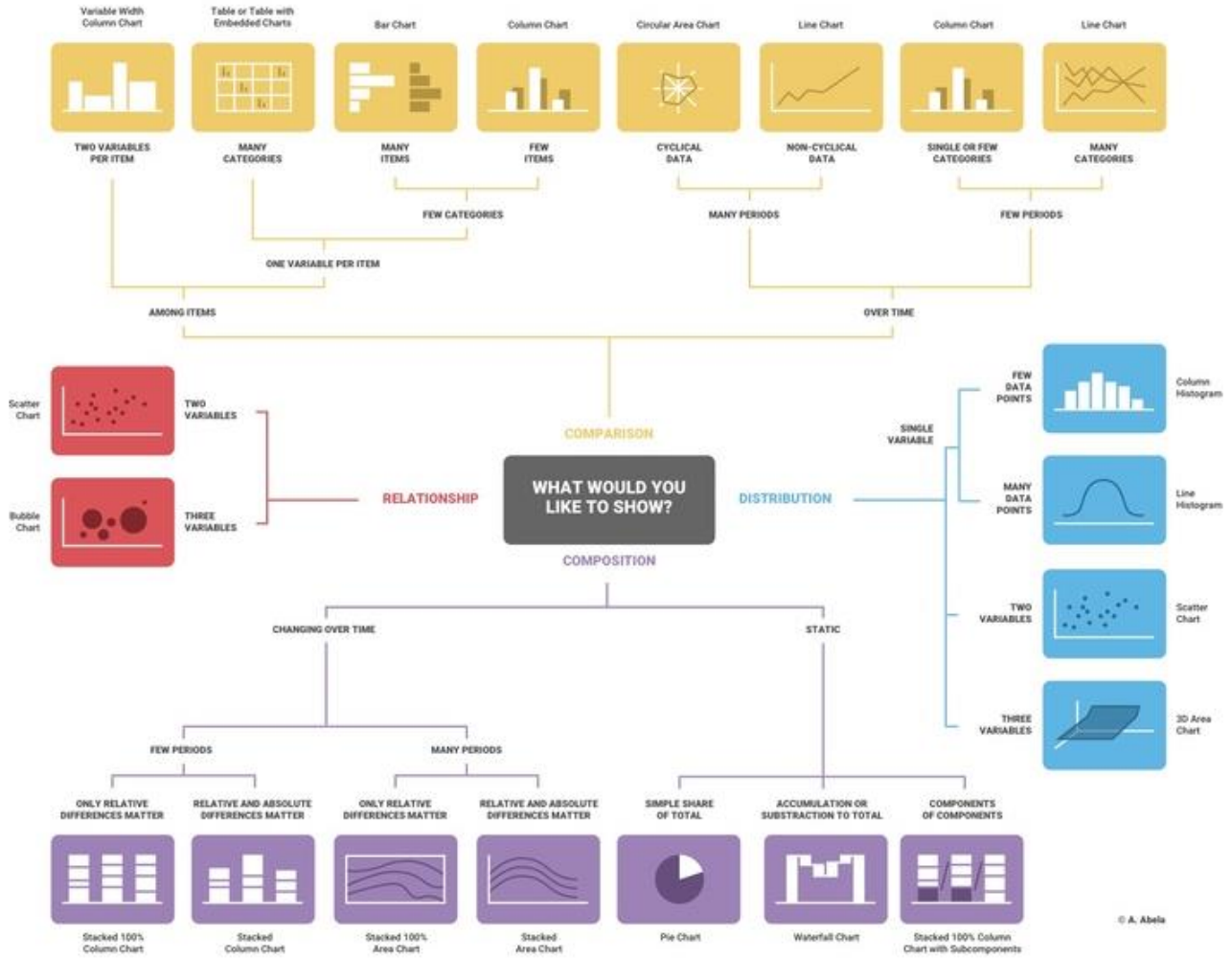
# How much data do you have?

- 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


What would you like to show?
COMPARISON
RELATIONSHIP
DISTRIBUTION
COMPOSITION
VARIABLE WIDTH COLUMN CHART
TABLE WITH EMBEDDED CHARTS
BAR CHART HORIZONTAL
BAR CHART VERTICAL
CIRCULAR AREA CHART
LINE CHART
BAR CHART VERTICAL
LINE CHART
Two Variables per Item
Many Categories
Few Categories
One Variable per Item
Among Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
Over Time
BAR HISTOGRAM
Few Data Points
Single Variable
LINE HISTOGRAM
Many Data Points
SCATTER PLOT
Two Variables
SCATTER PLOT
SCATTER PLOT BUBBLE SAZE
Two Variables
Three or more Variables
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Accumulation to total and absolute difference matters
STACKED 100% BAR CHART
STACKED BAR CHART
STACKED AREA 100% CHART
STACKED AREA CHART
PIE CHART
WATERFALL CHART
STACKED 100% BAR CHART WITH SUBCOMPONENTS
TREE MAP

# CHART SUGGESTIONS - A THOUGHT-STARTER



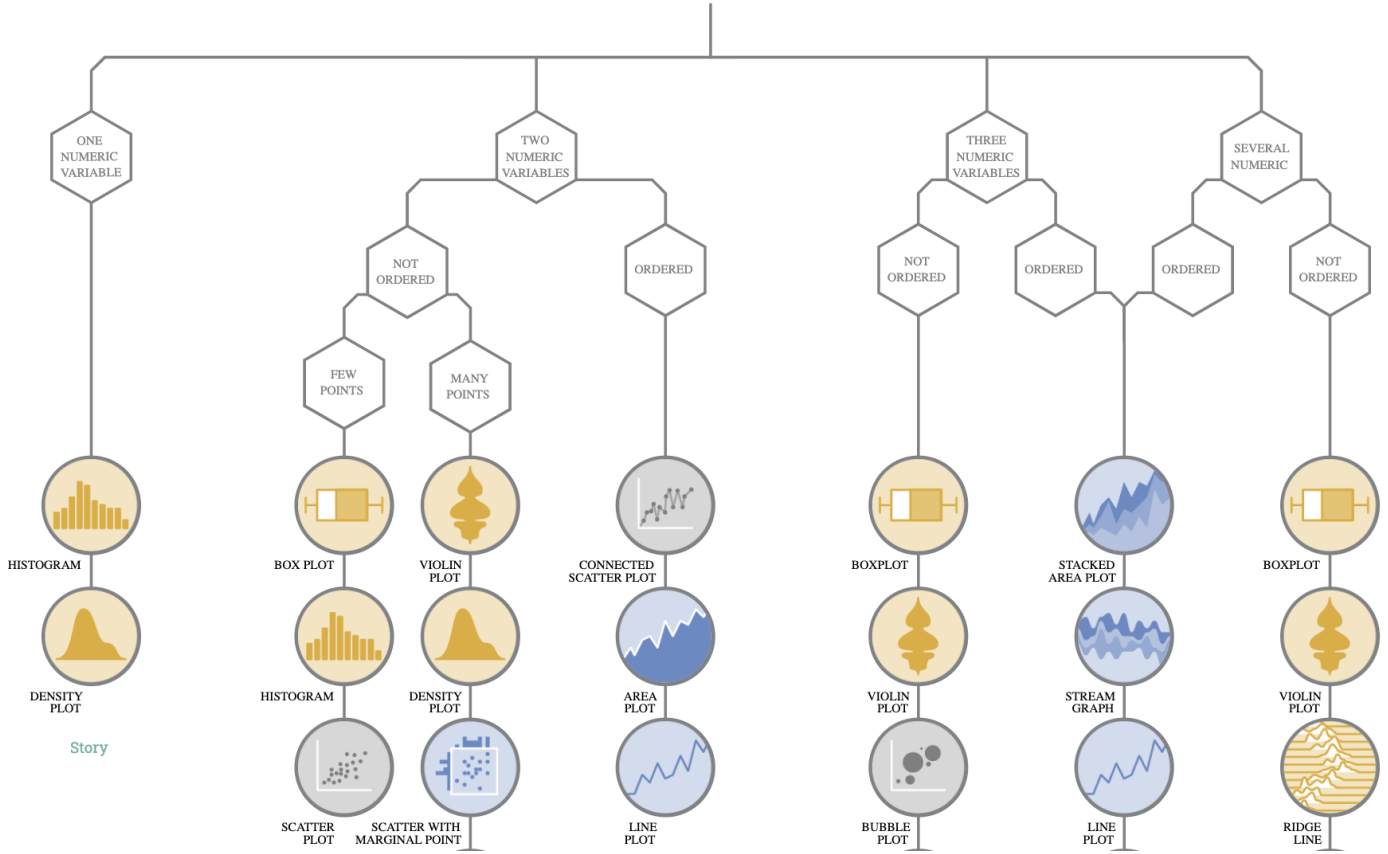

https://infogram.com/page/choose-the-right-chart-data-visualization

from Data to Viz

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

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.

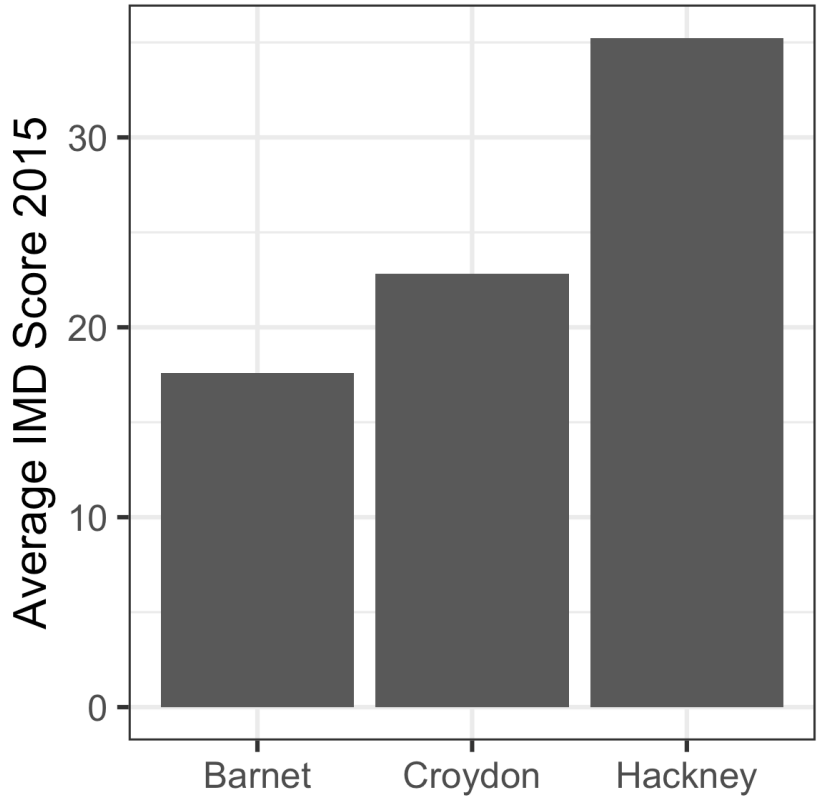
Numeric | Categoric | Num & Cat | Maps | Network | Time series



# Comparing values

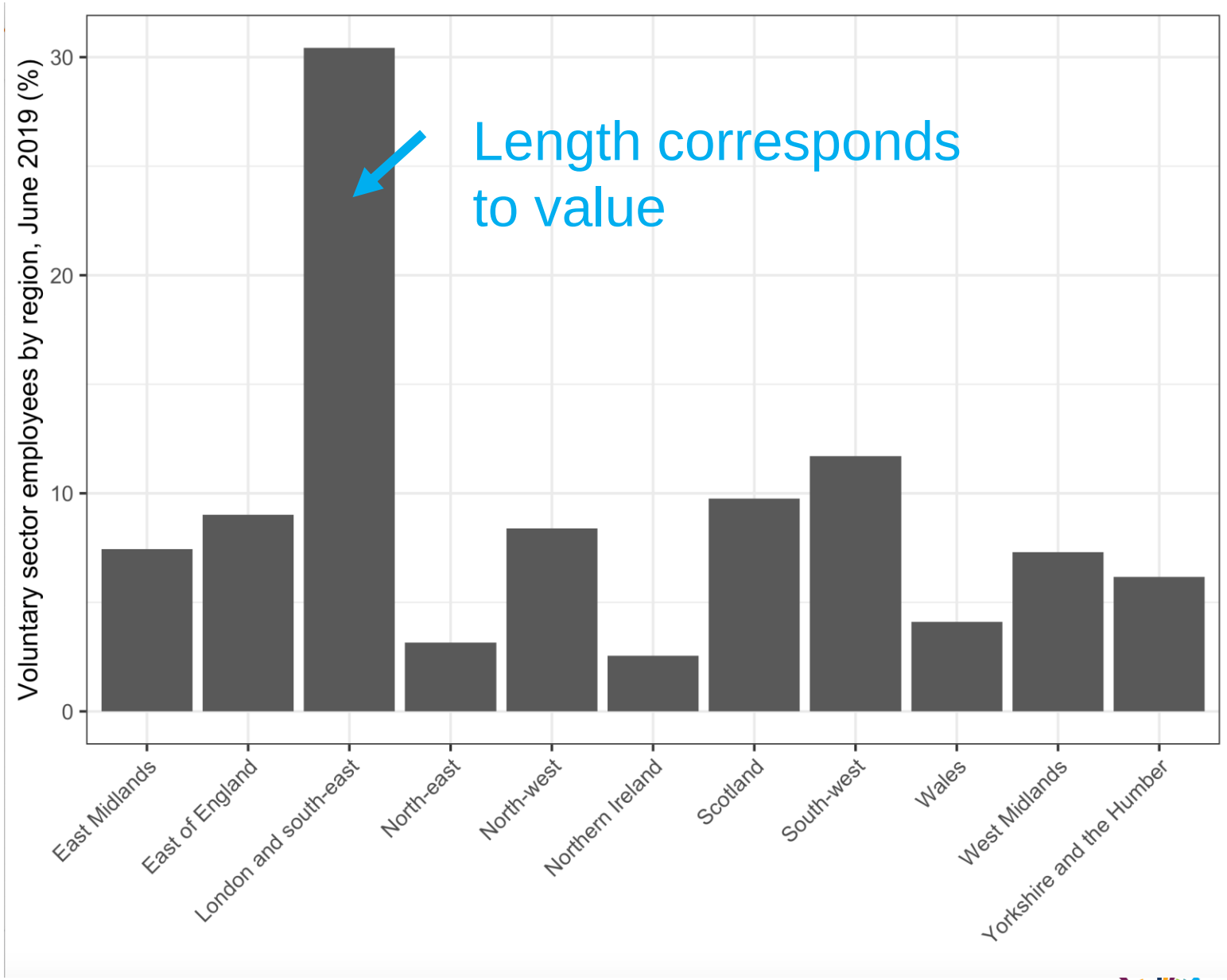
# Bar charts

- **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

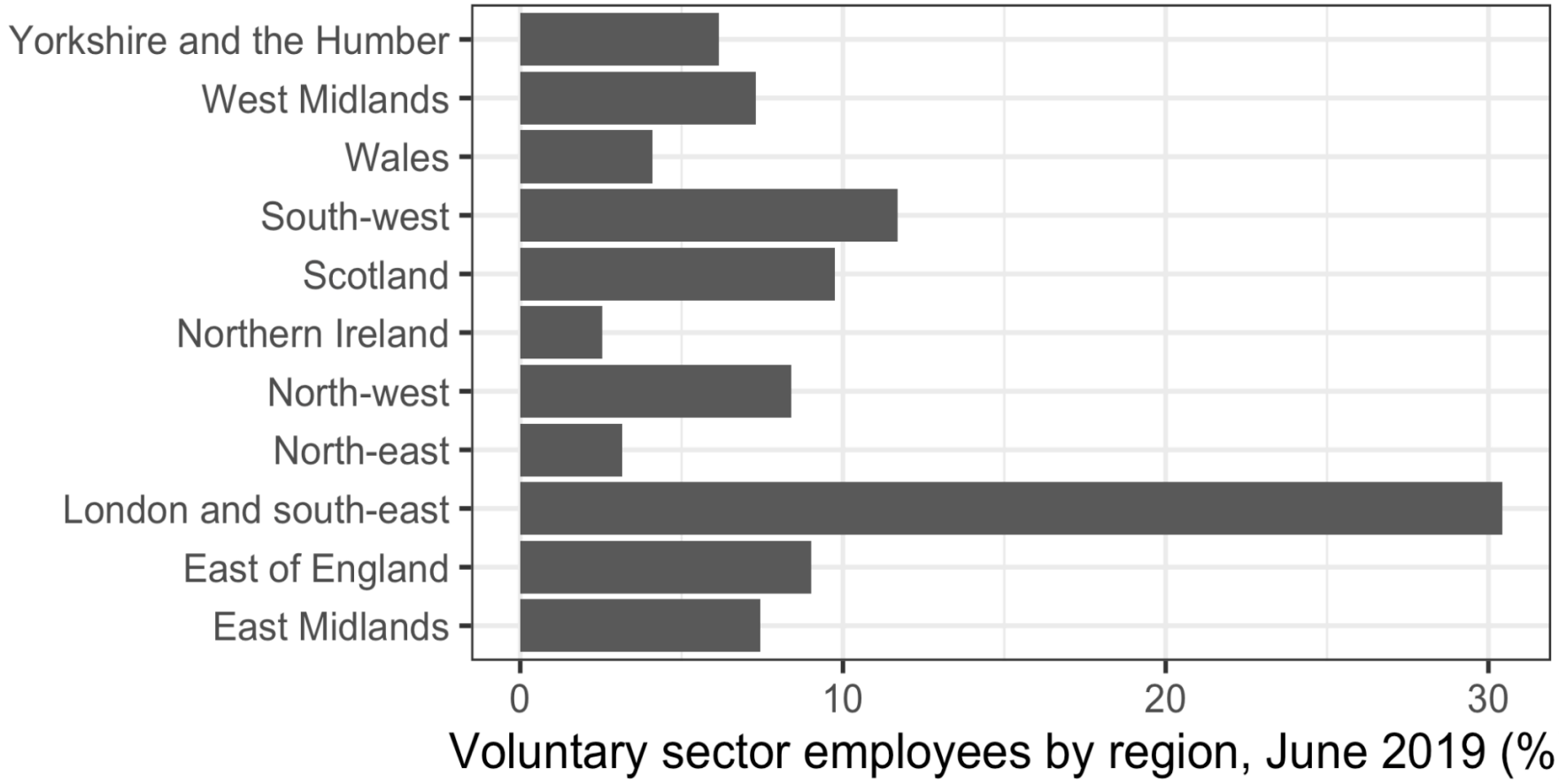


# Bar charts

| 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 |

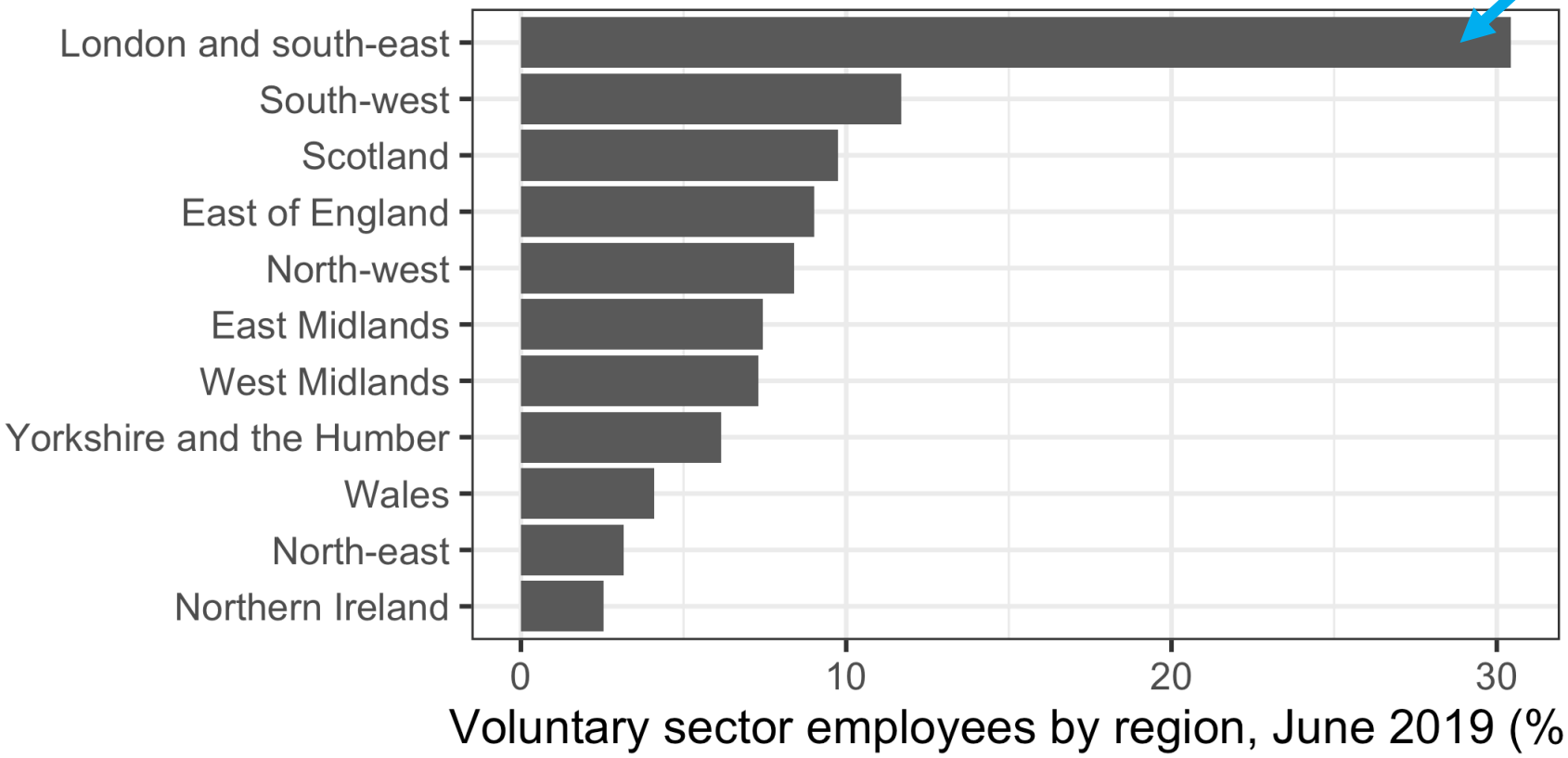


# Bar chart tips



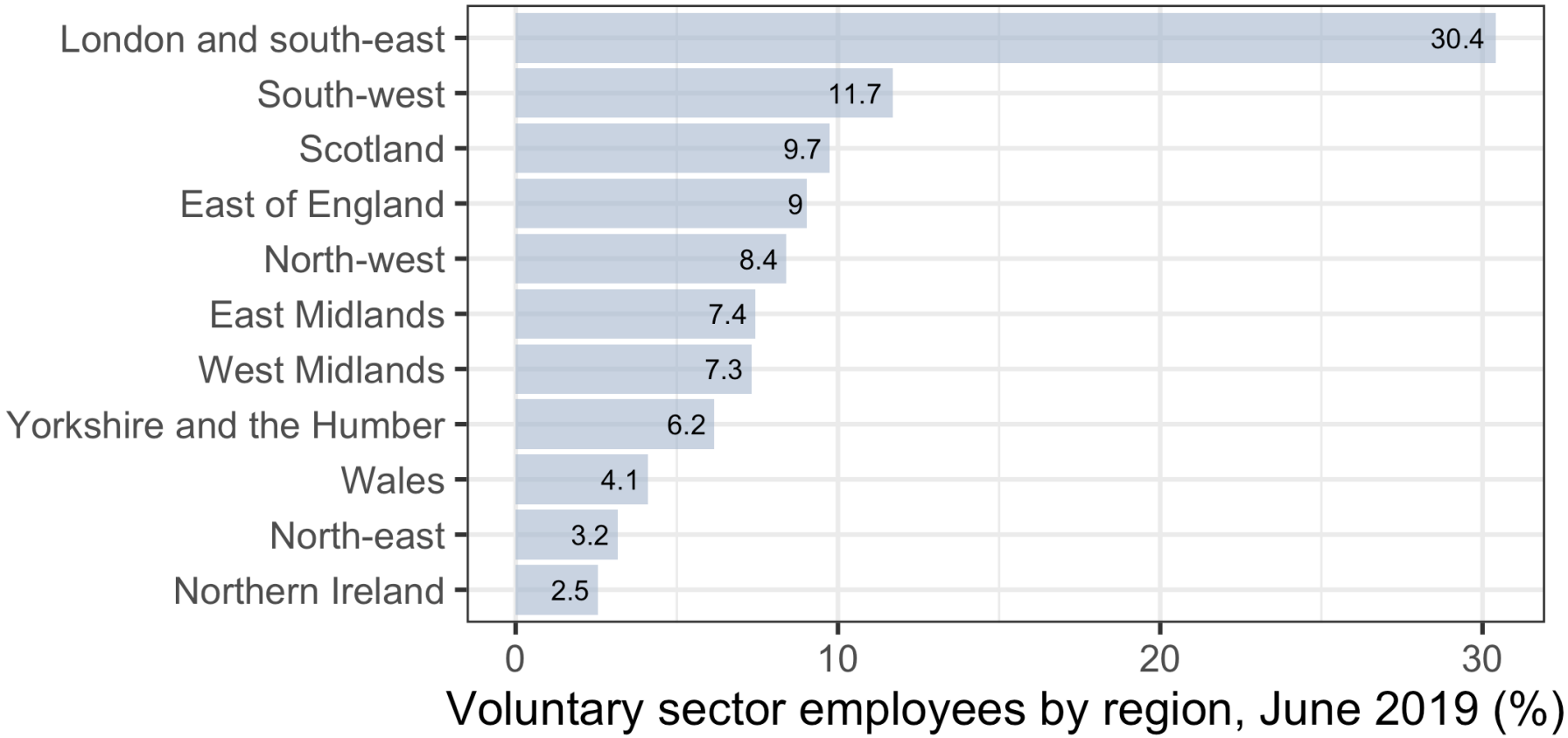

Flip charts with long labels sideways

# Bar chart tips




Reorder to enable comparison (only if there is no natural order)

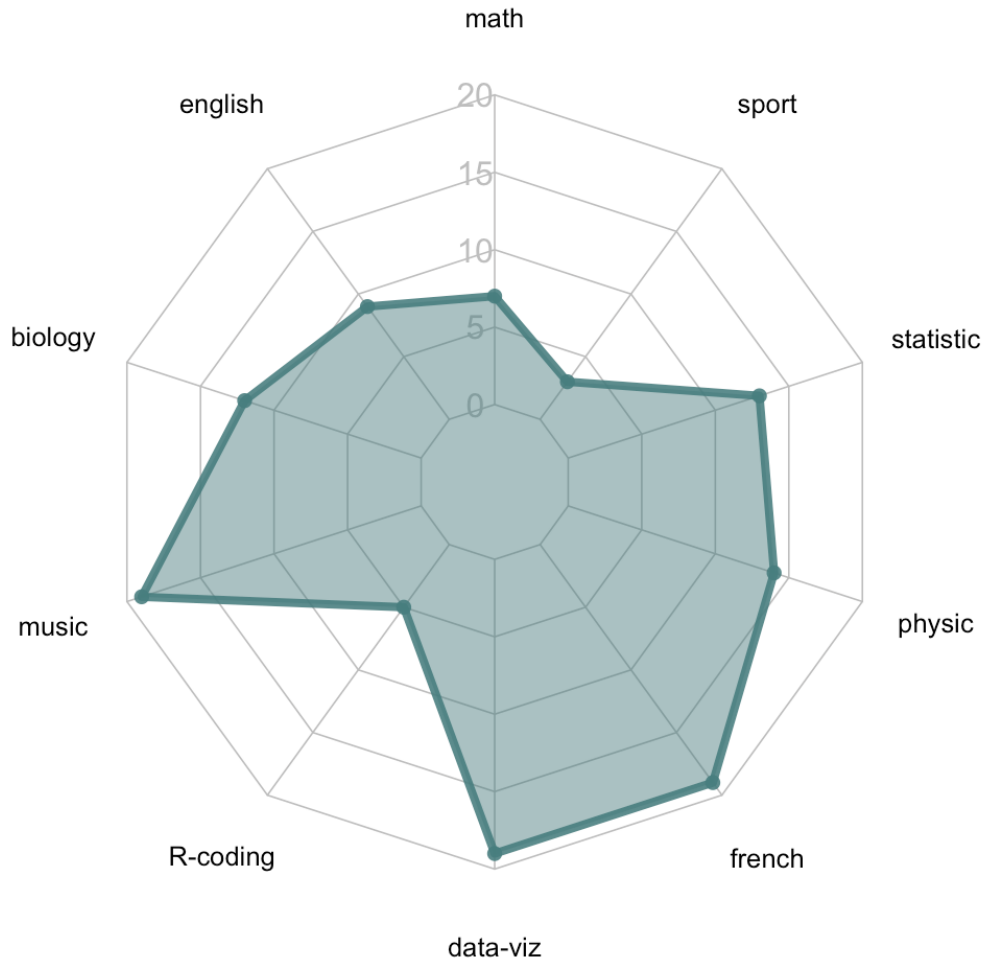
# Bar chart tips




Add value to bar if precise value is important

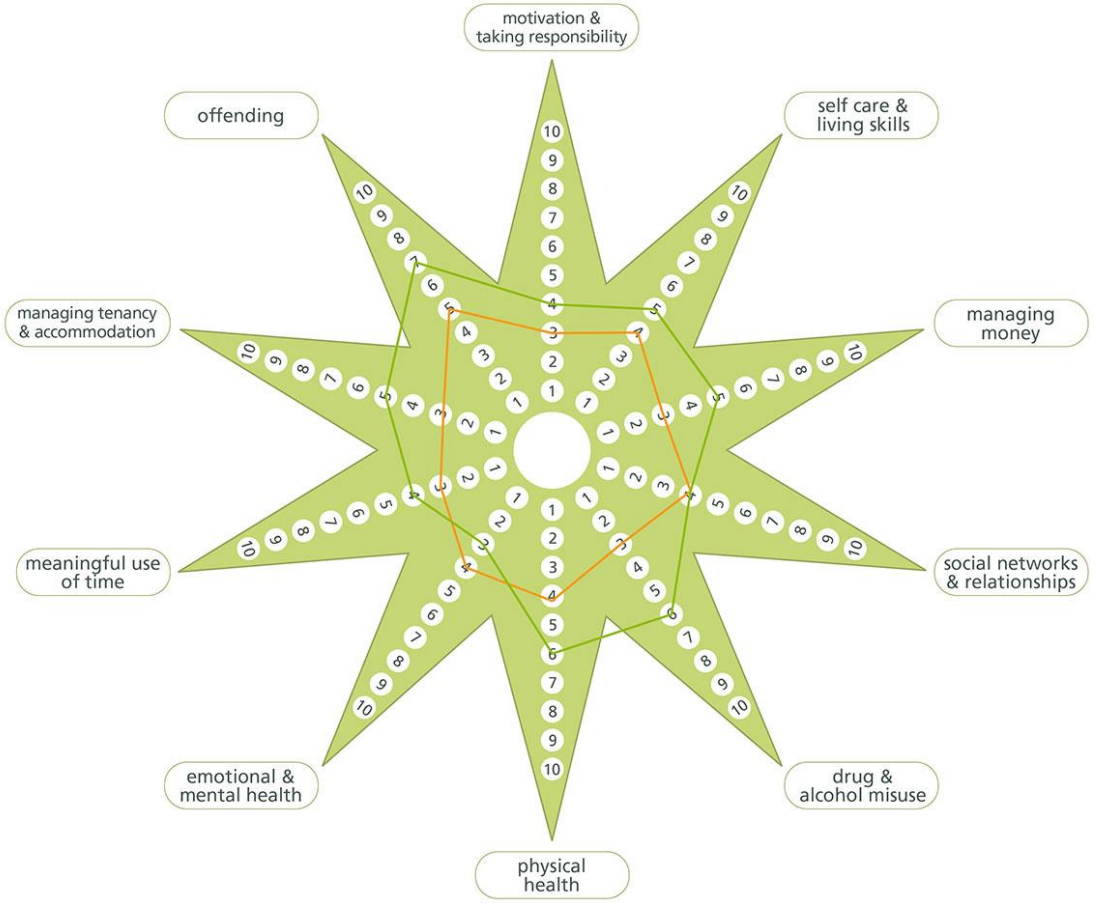
# Spider Chart/Radar Chart

- **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




https://www.data-to-viz.com/caveat/spider

# Outcome stars are a type of radar chart

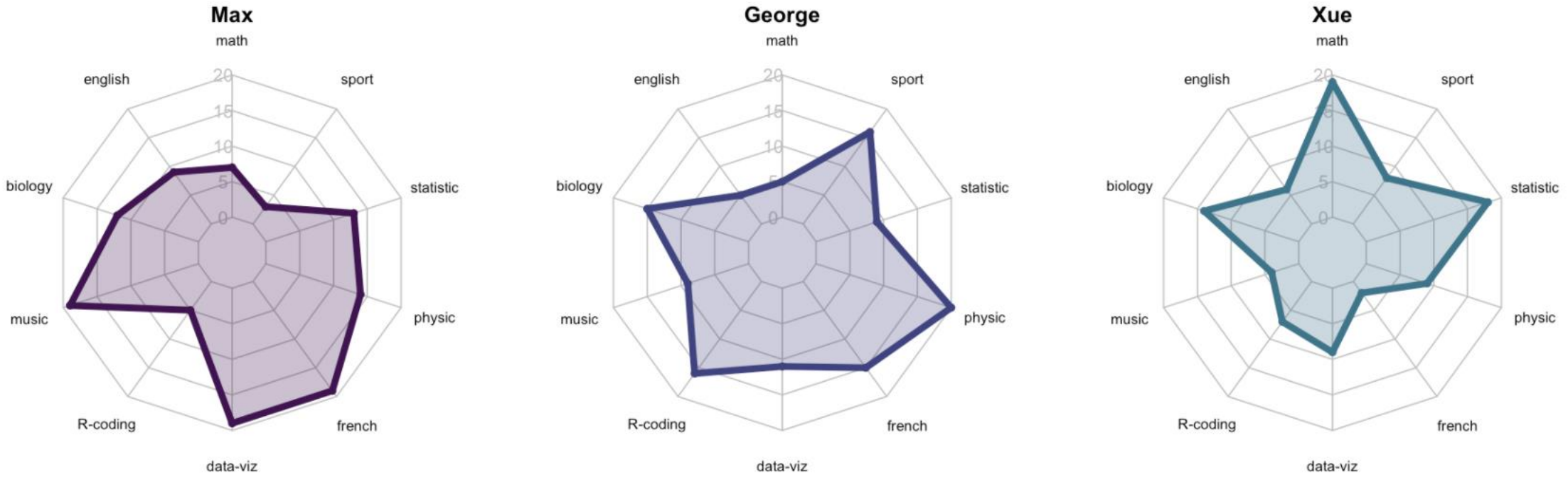



Homelessness Star™ © Triangle Consulting Social Enterprise Ltd
Authors: Sara Burns and Joy MacKeith
**www.outcomesstar.org.uk**

# Spider Chart/Radar Chart

**Advantages:**

- Useful for individuals to compare scores across similarly measured

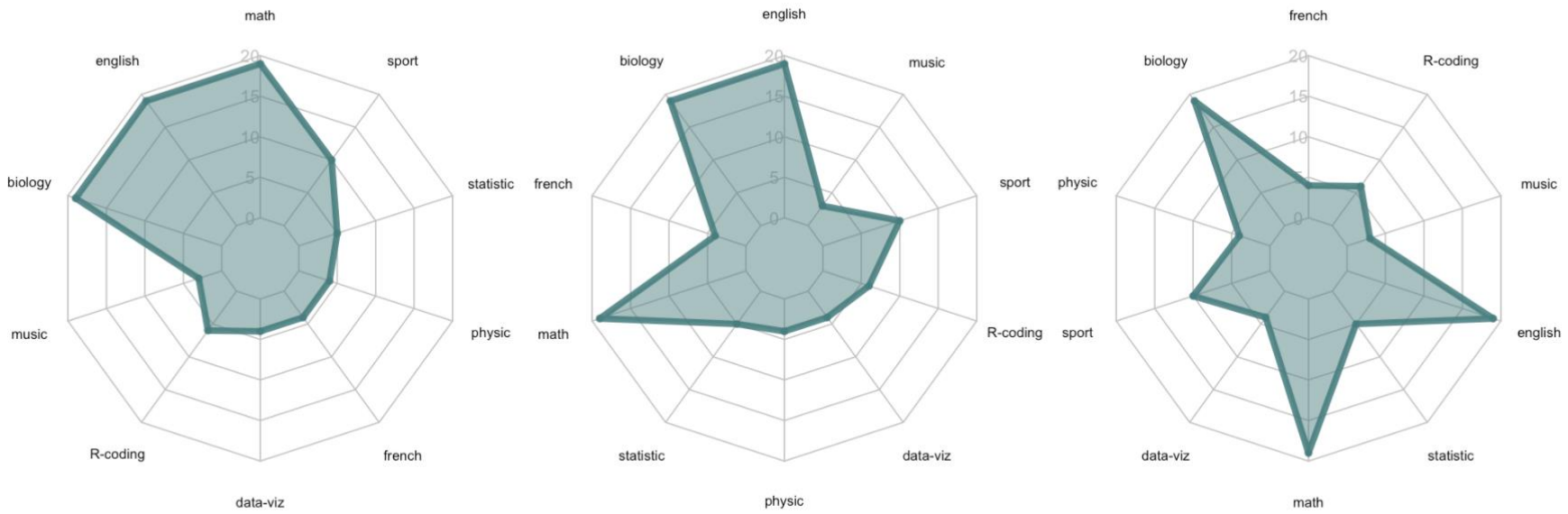



https://www.data-to-viz.com/caveat/spider

# Spider Chart/Radar Chart

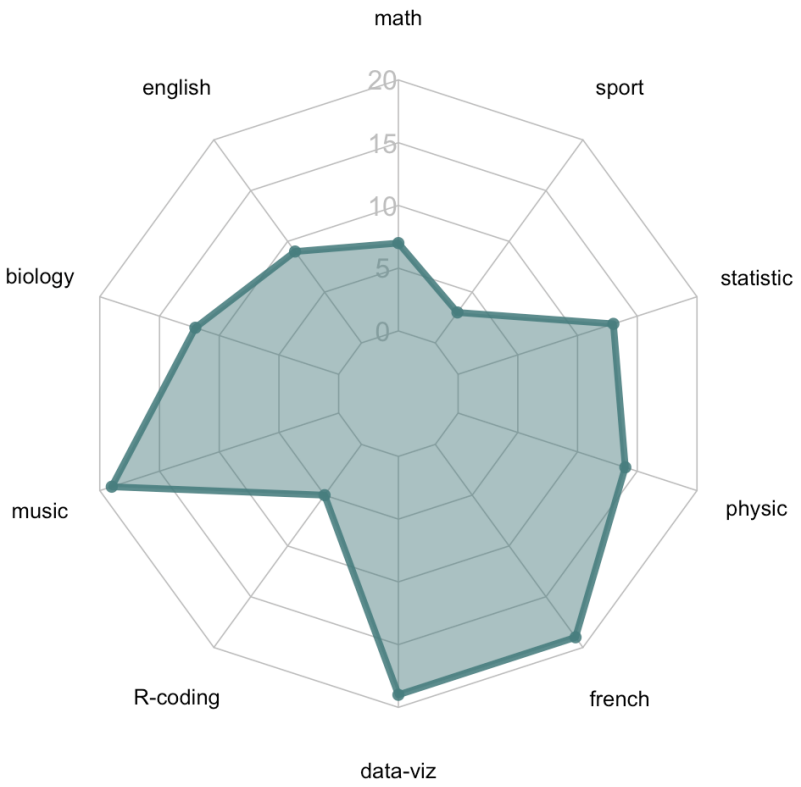
## Disadvantages

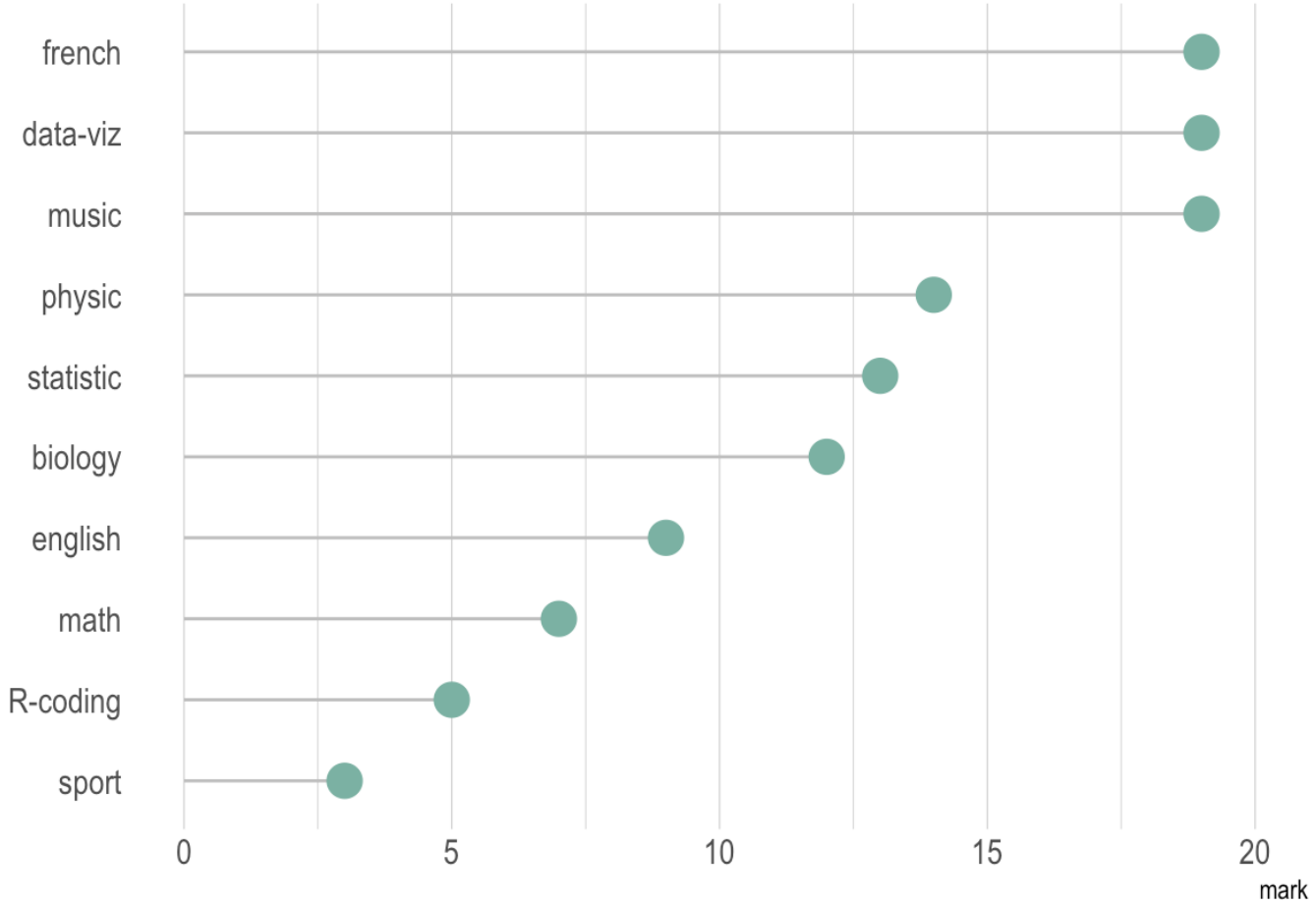
- Circular layout is hard to read
- Hard to determine ranking
- Category order affects the shape
- Data needs to be on the same scale




https://www.data-to-viz.com/caveat/spide

# Spider Chart/Radar Chart alternatives




https://www.data-to-viz.com/caveat/spider

# 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

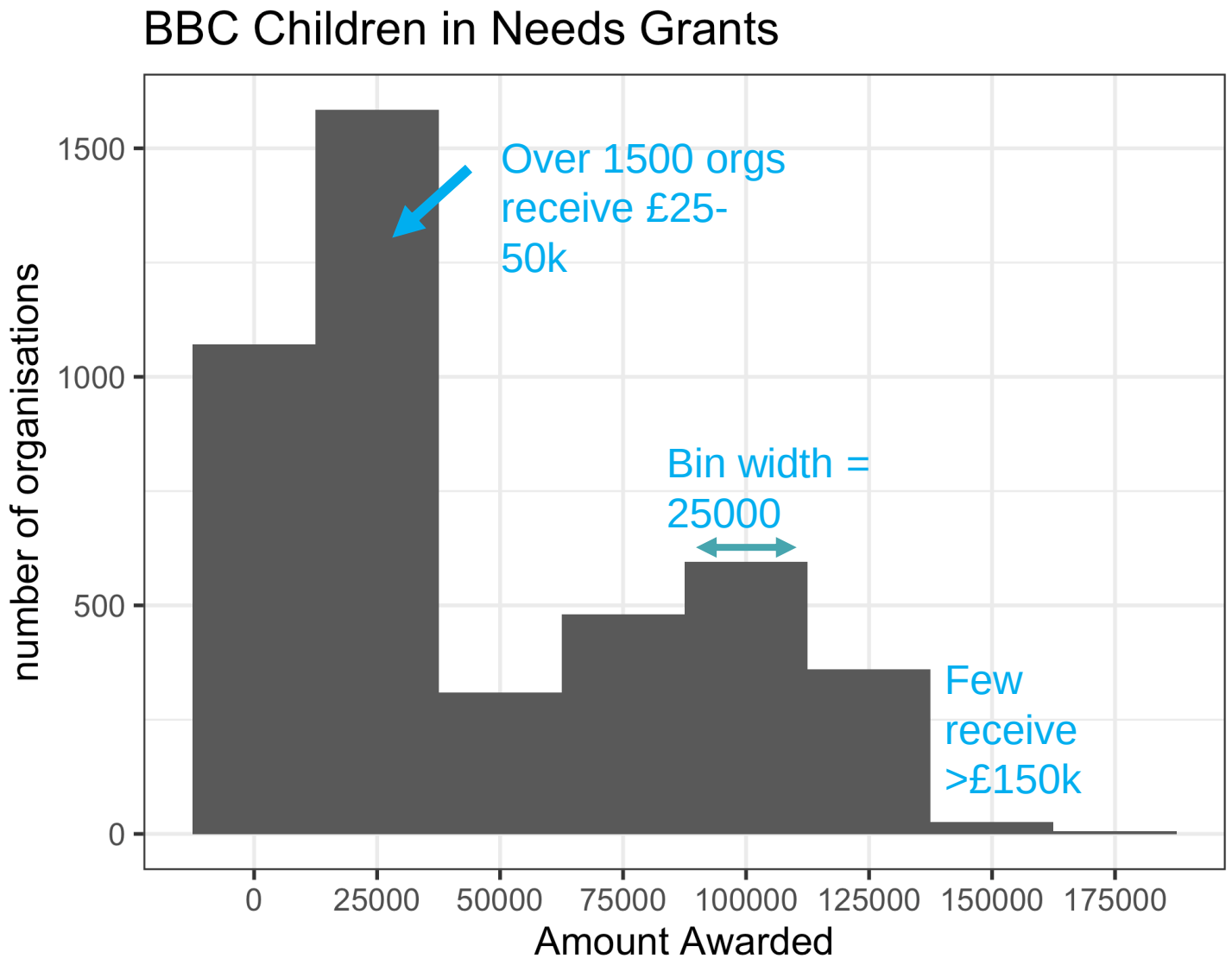
# Histograms

**Input data:** Grant funding awarded to different organisations

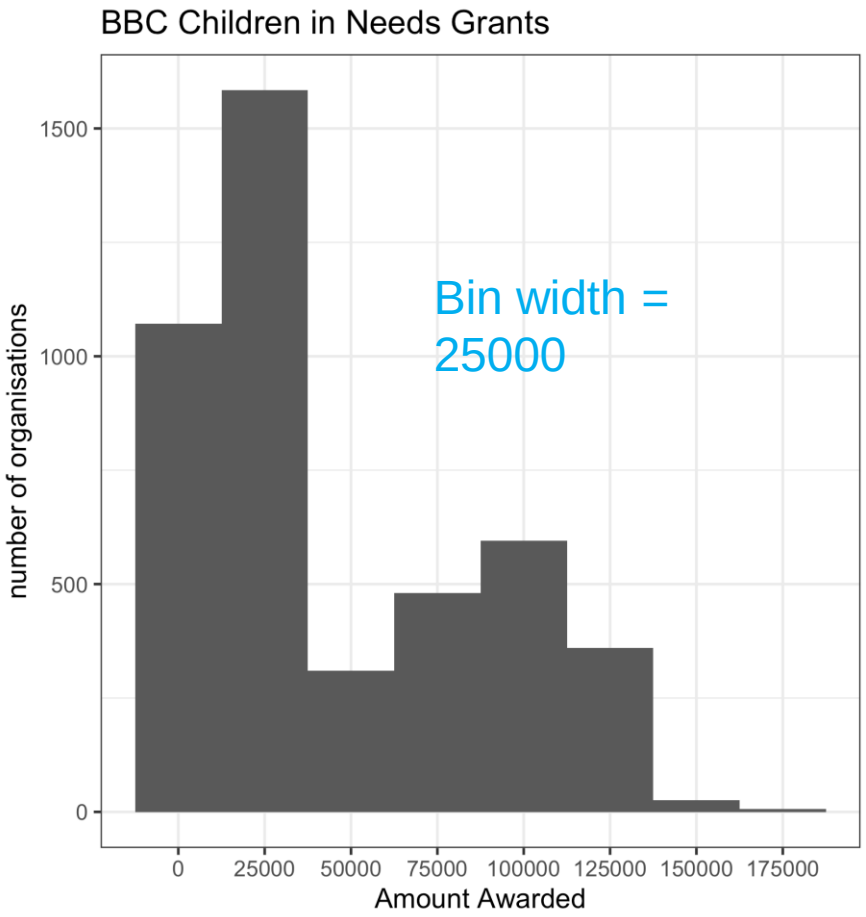
| Amount Awarded |
|---|
| 49495 |
| 55709 |
| 81483 |
| 88157 |
| 71514 |
| 83795 |
| 44654 |
| 59753 |
| 110000 |
| 51478 |

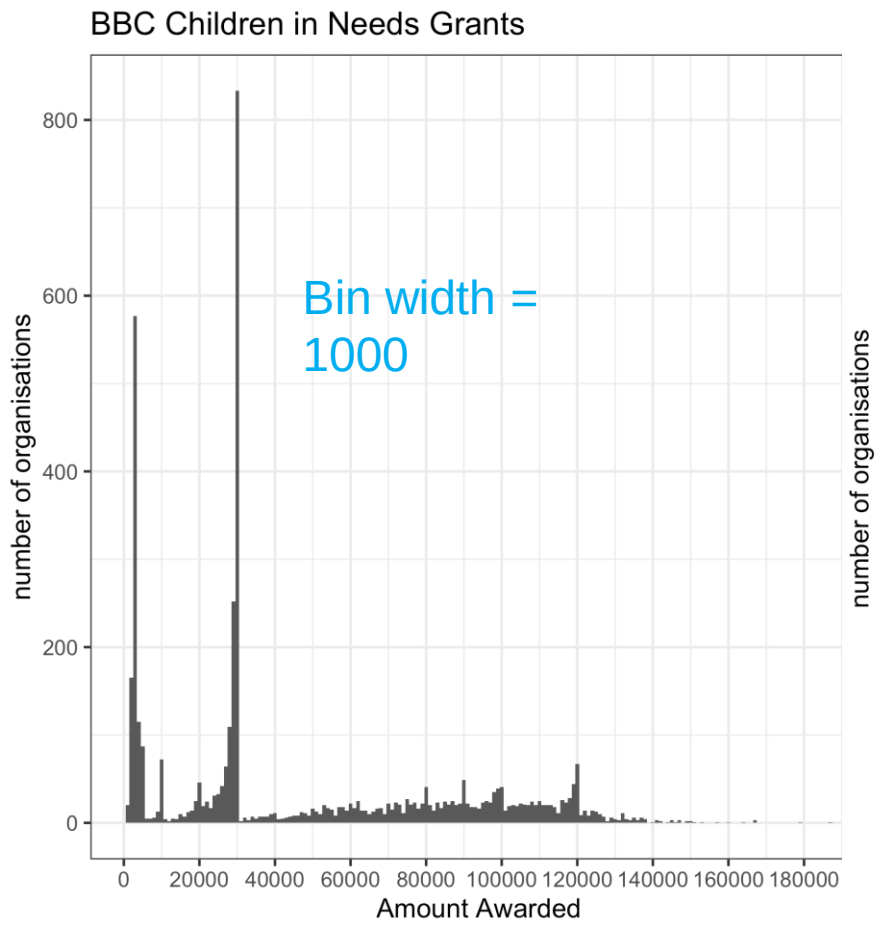
Each row is 1 grant

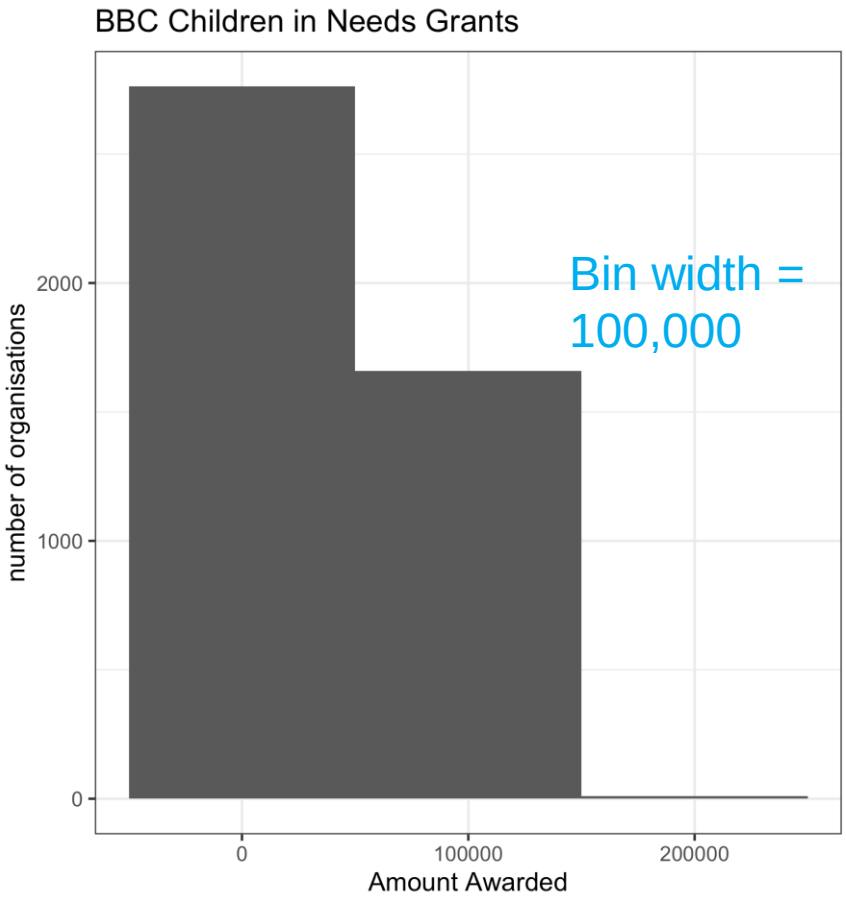
Within range 50000-75000



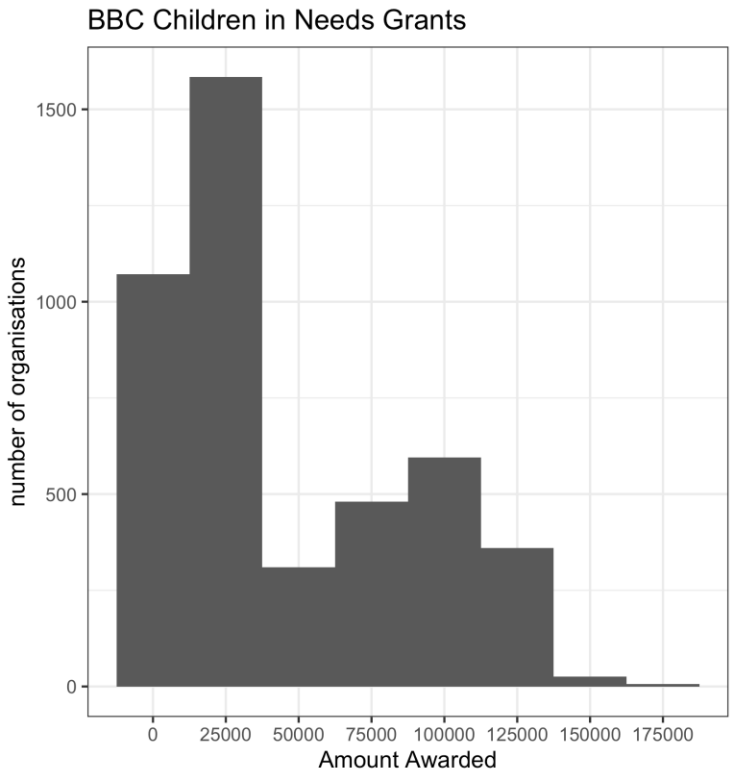
# Common pitfalls: bin width

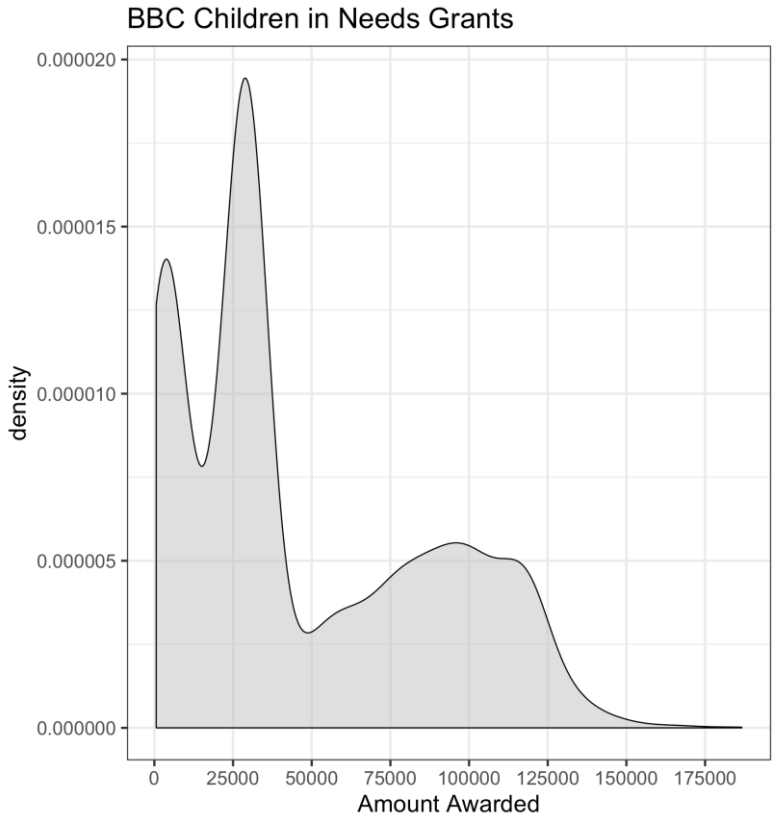


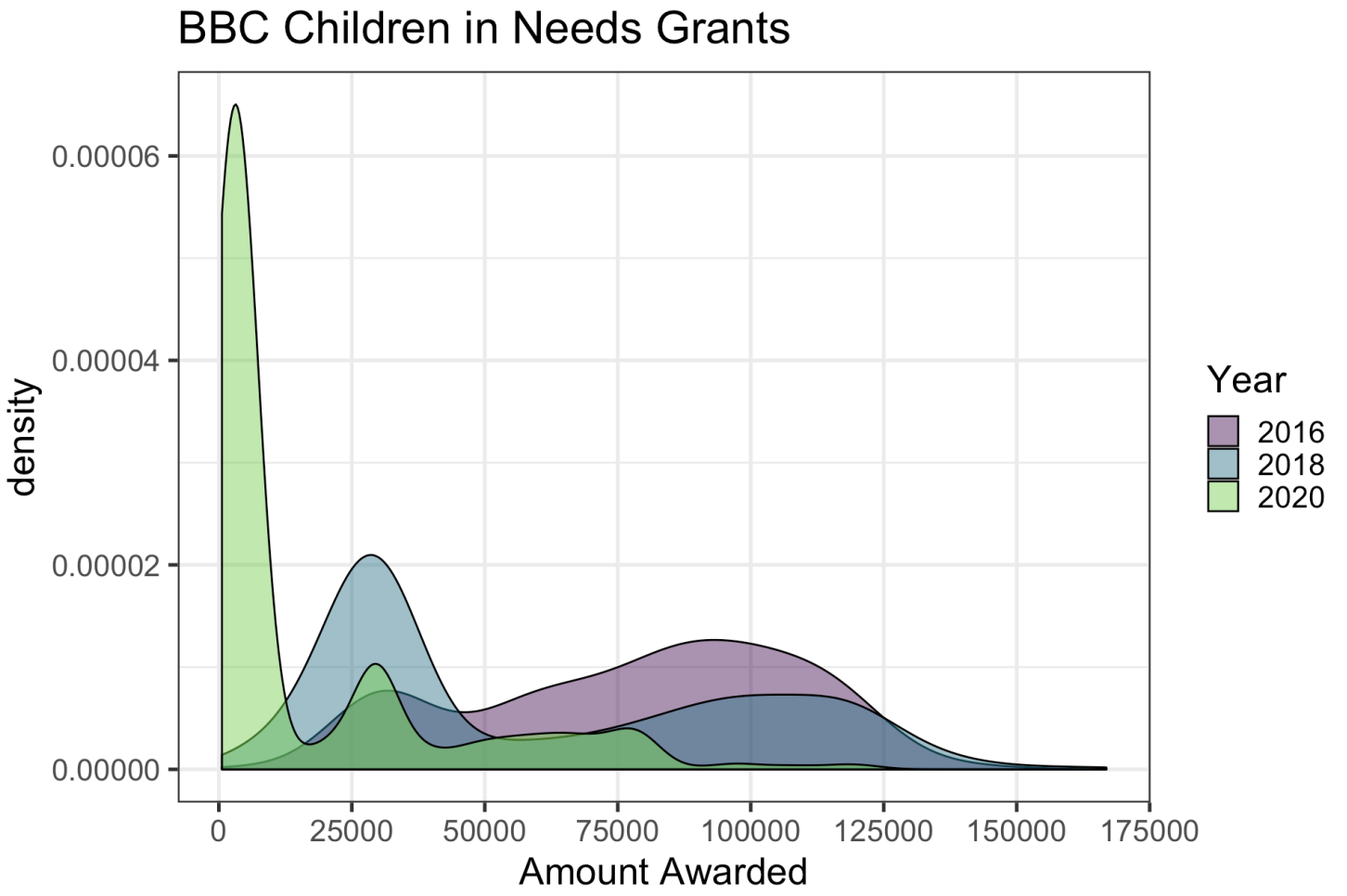
# Histograms vs density plots




- **Density plots** estimate the underlying distribution of the data over a continuous range
- **Peaks** show where values are concentrated (total area under the curve adds up to 1)
- Better used with **large** datasets
- Useful for **comparing distributions**

# Histograms vs density plots




- **Density plots** estimate the underlying distribution of the data over a continuous range
- **Peaks** show where values are concentrated (total area under the curve adds up to 1)
- Better used with **large** datasets
- Useful for **comparing distributions**

# Boxplots

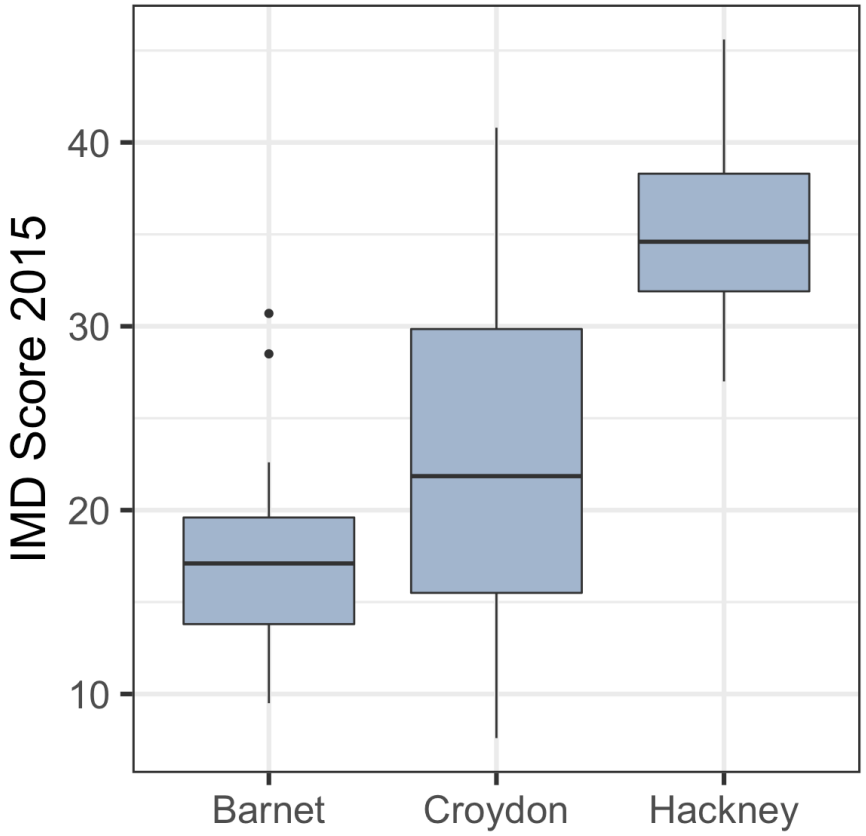
**What is it?**

- Summarises one or more numeric variables

**Input data**: 2+ numeric variables (e.g. Outcome over Time), or 1 numeric variable for many categories (e.g. Score per Region)

**Why use it?**

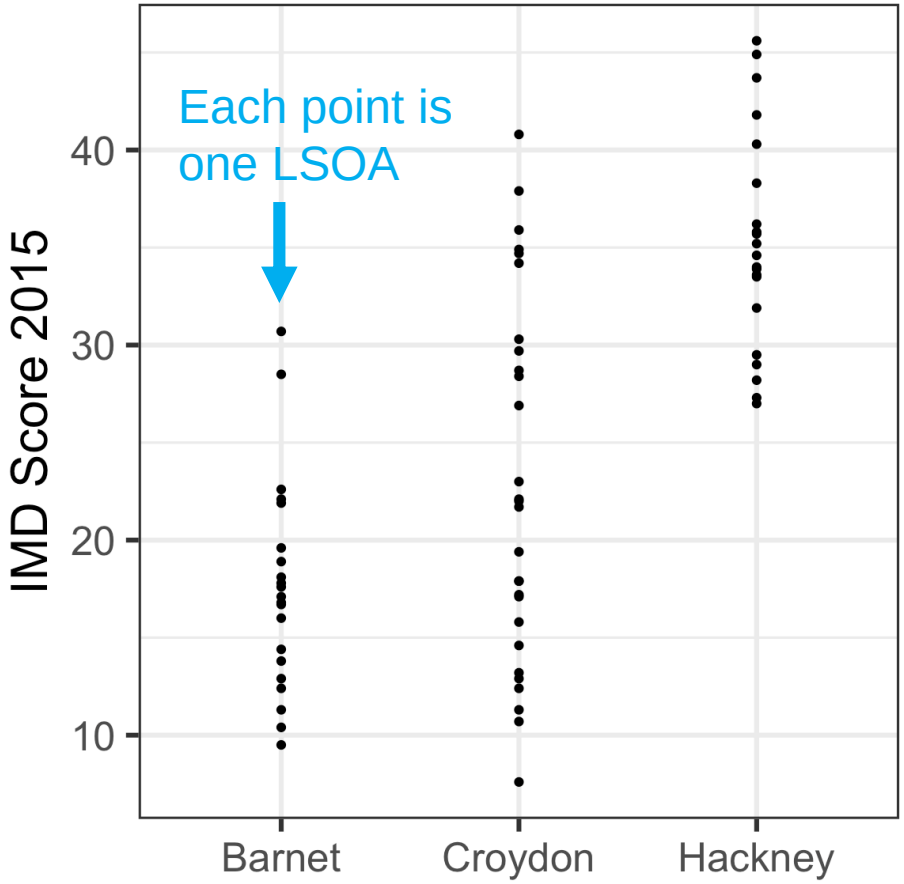
- To visualise many distributions at once
- To understand the range and middle values in your data



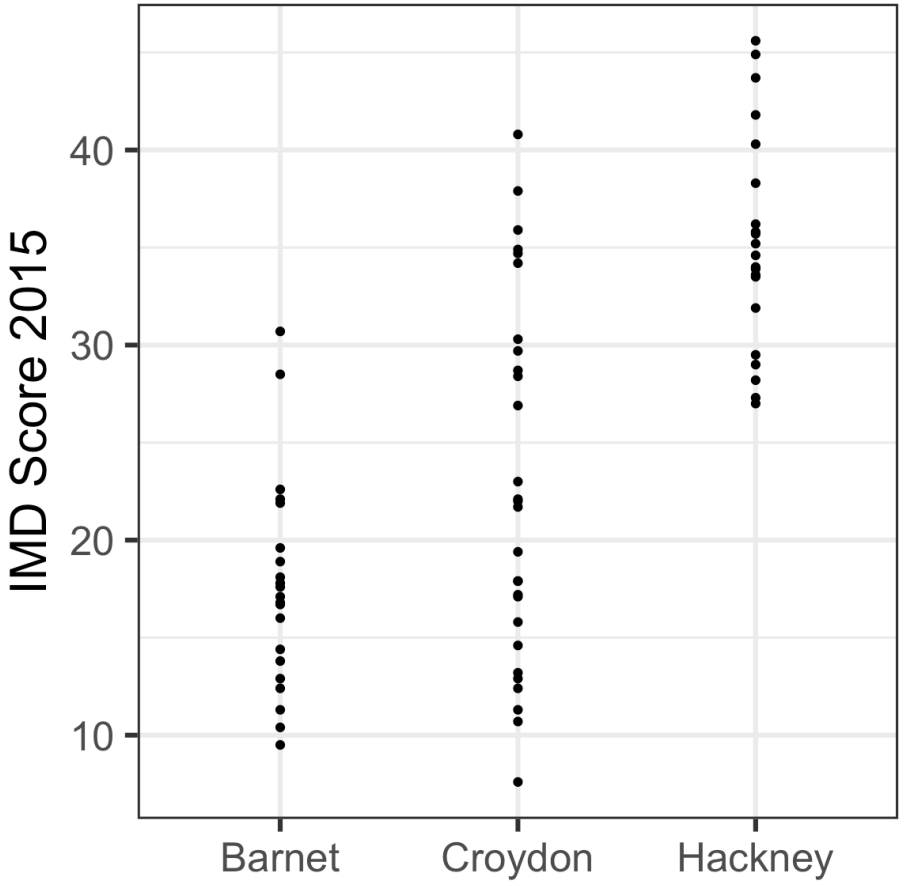
# Why make a boxplot?

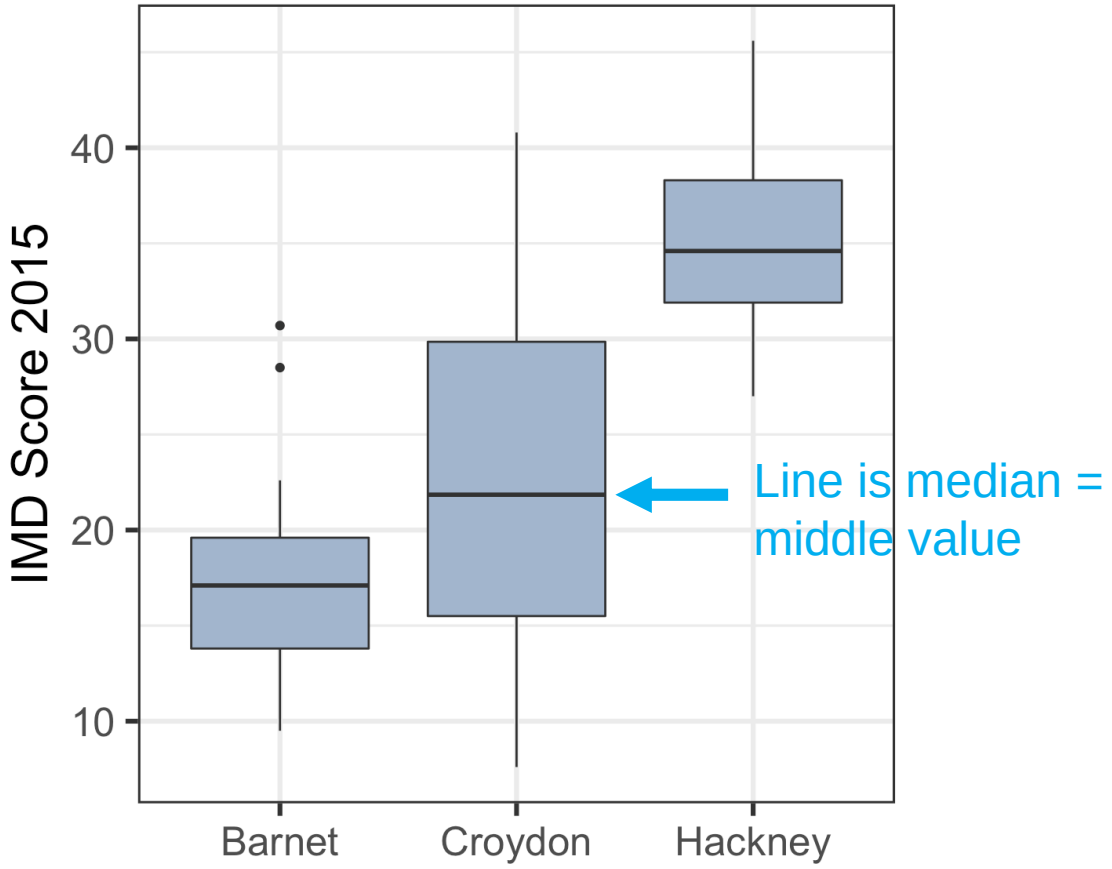
| Code | local_authority | index_of_multiple_deprivation_imd_score_2015 |
|---|---|---|
| E05000043 | Barnet | 14.4 |
| E05000044 | Barnet | 30.7 |
| E05000045 | Barnet | 21.9 |
| E05000046 | Barnet | 28.5 |
| E05000047 | Barnet | 19.6 |
| E05000048 | Barnet | 16.0 |

Each row is one LSOA

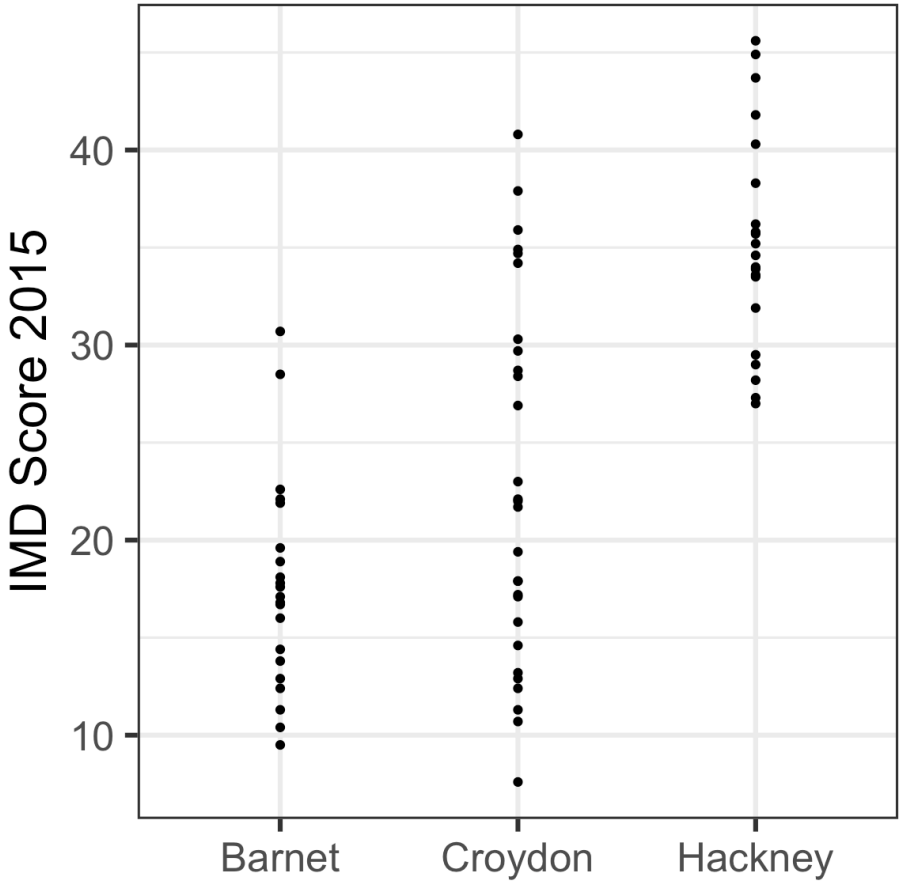


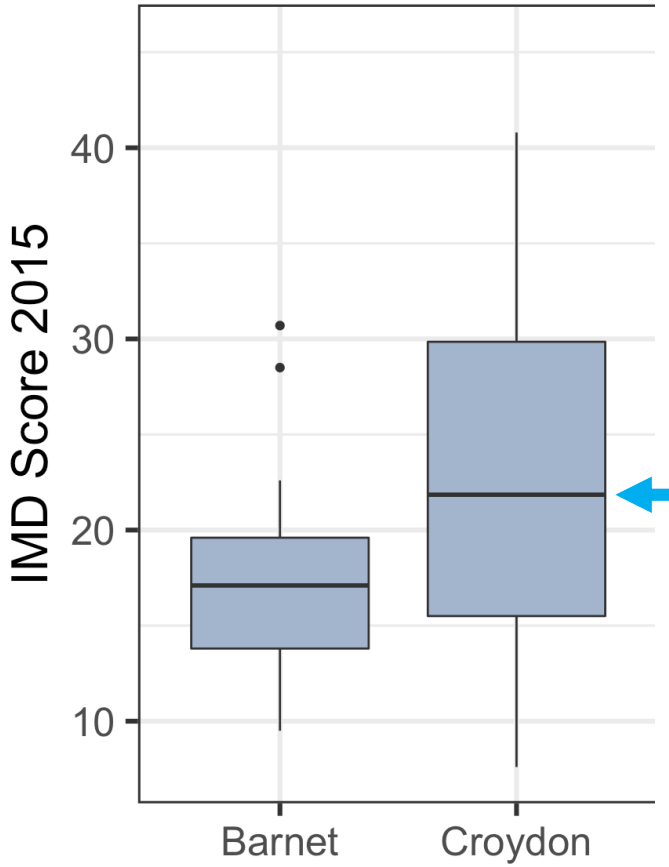
# Anatomy of a boxplot



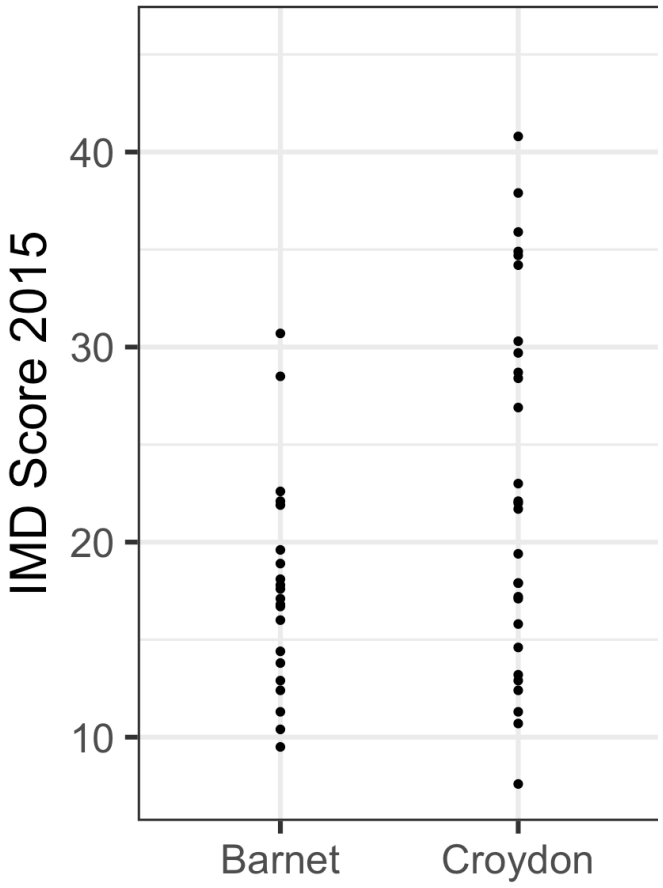
# Anatomy of a boxplot

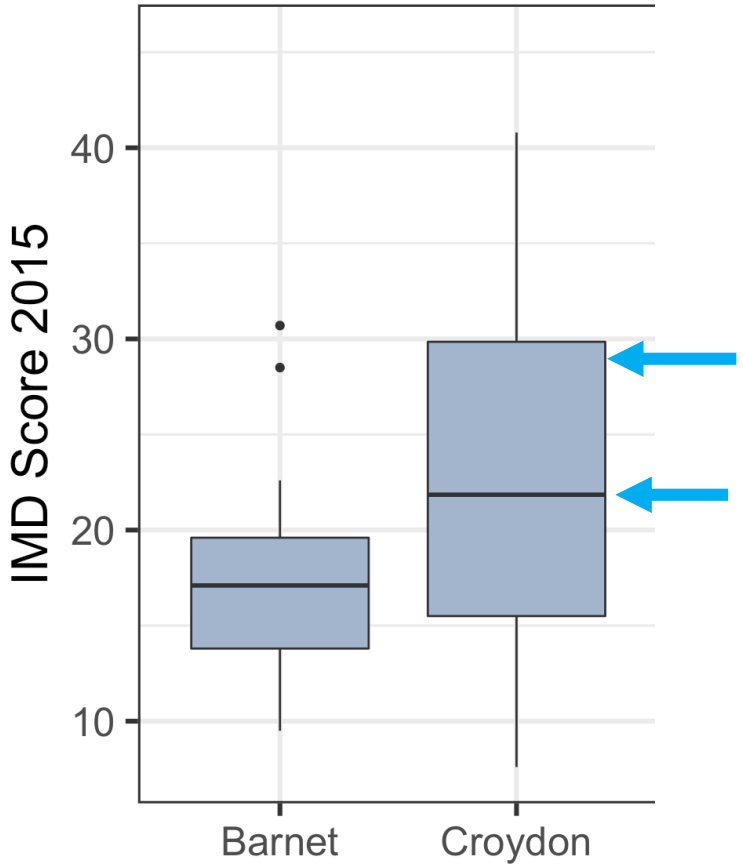



Line is median = middle value

# Anatomy of a boxplot

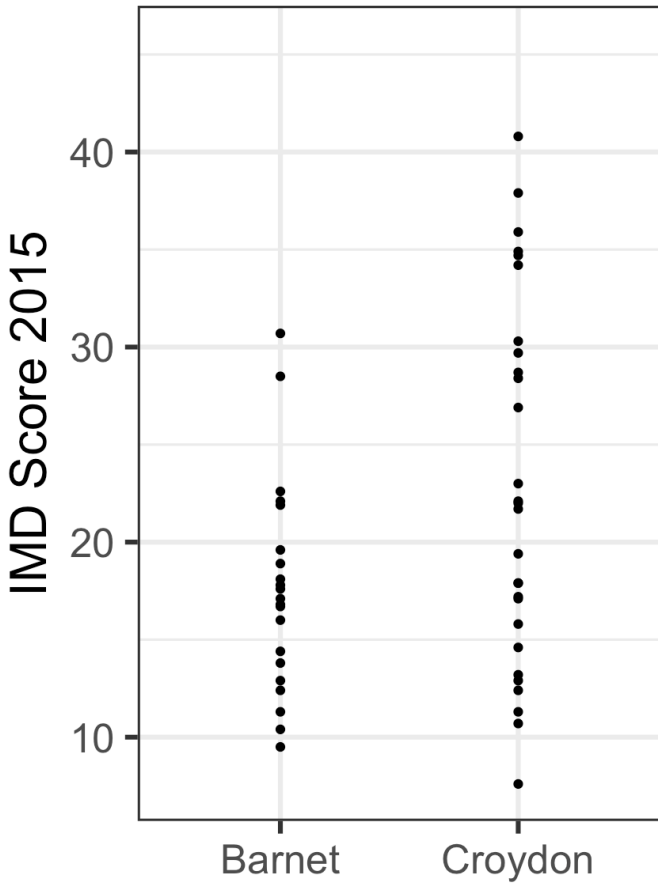


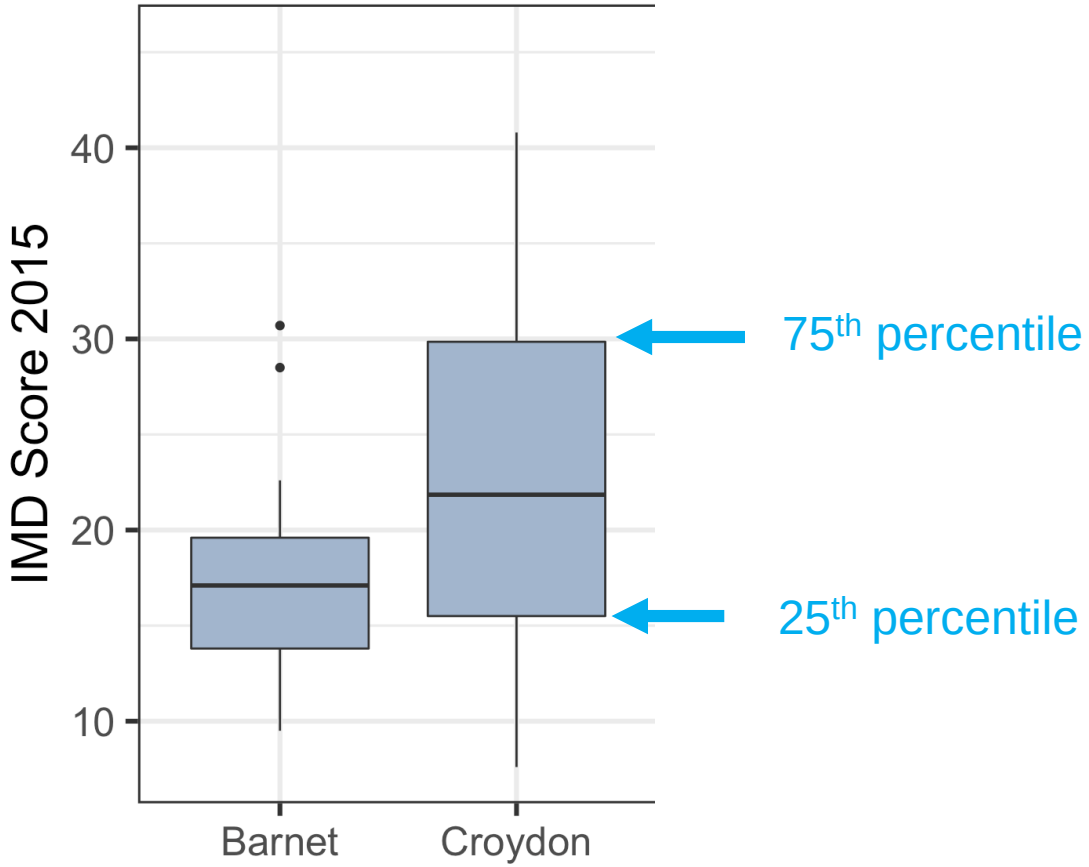




Box contains the middle 50% of values

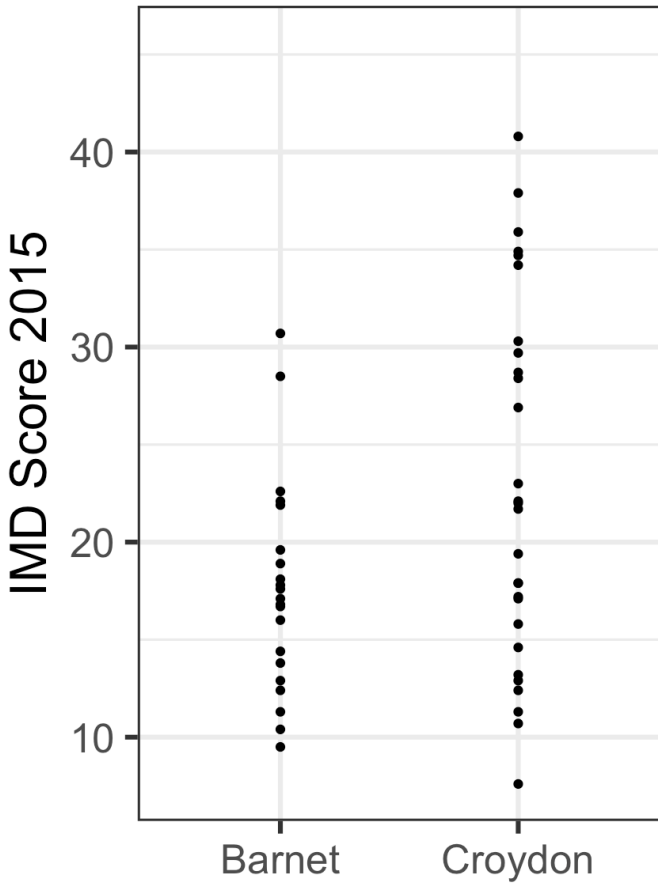
Line is median = middle value

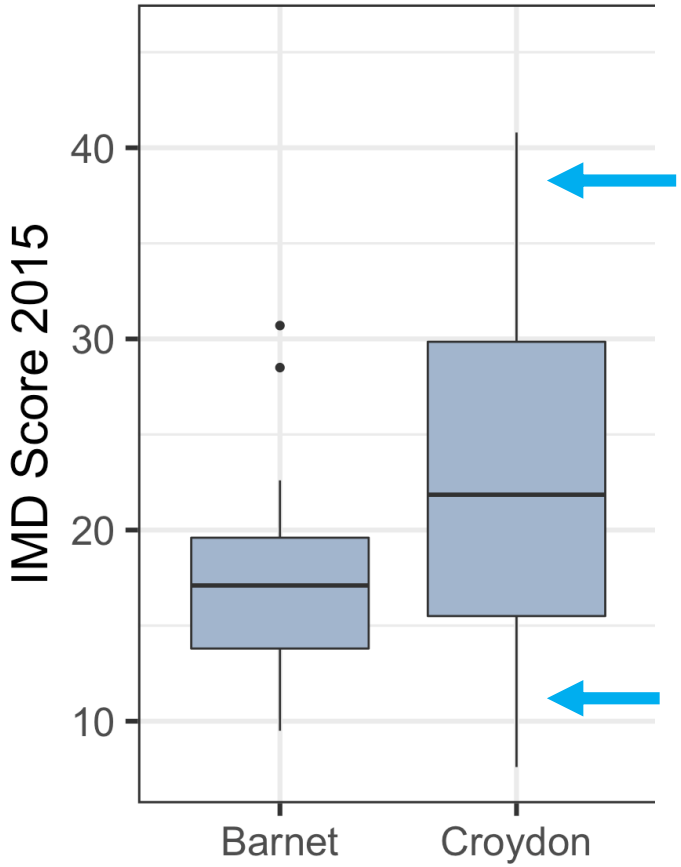
# Anatomy of a boxplot



# Anatomy of a boxplot

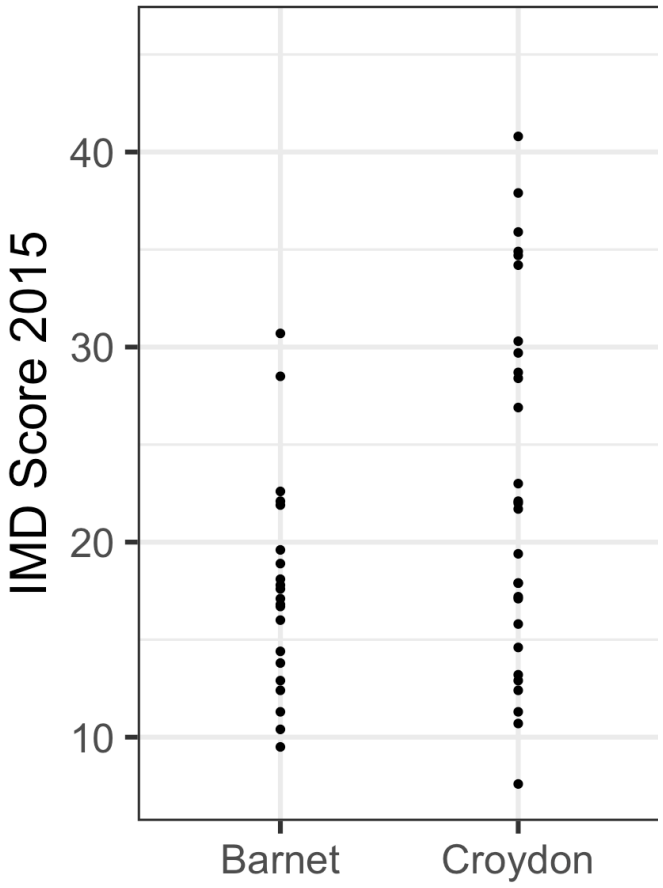


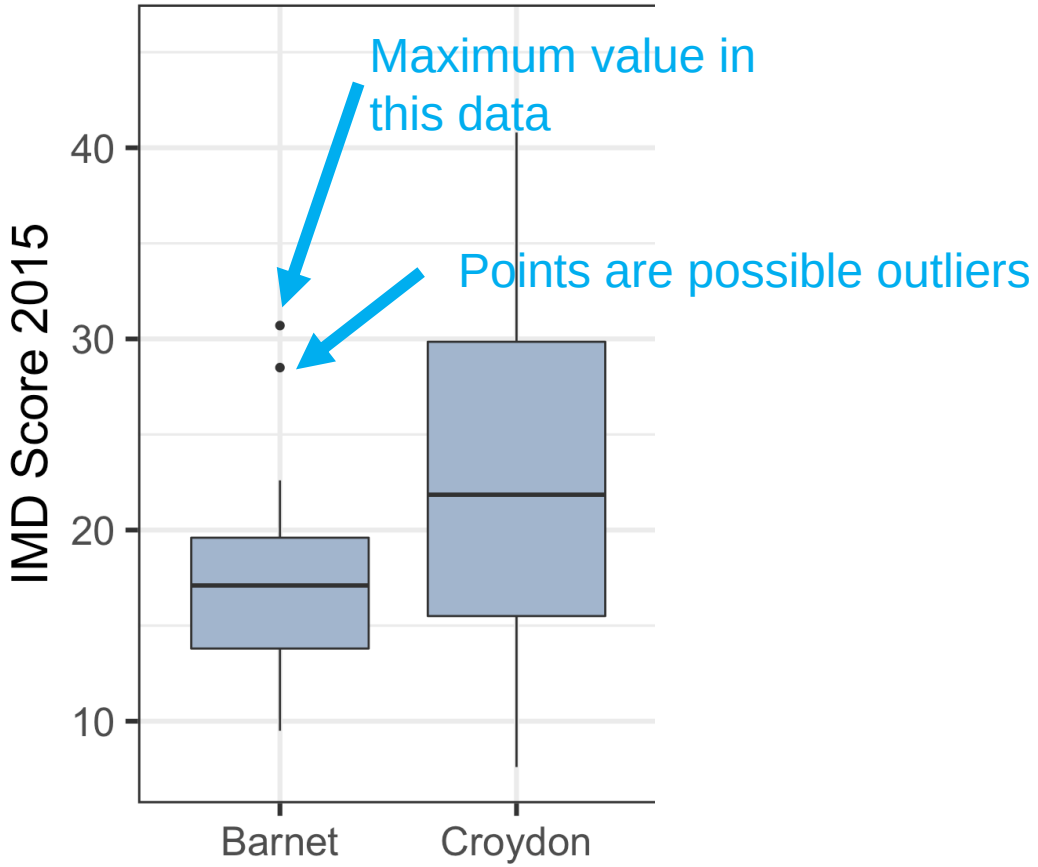




“whiskers” contain expected range

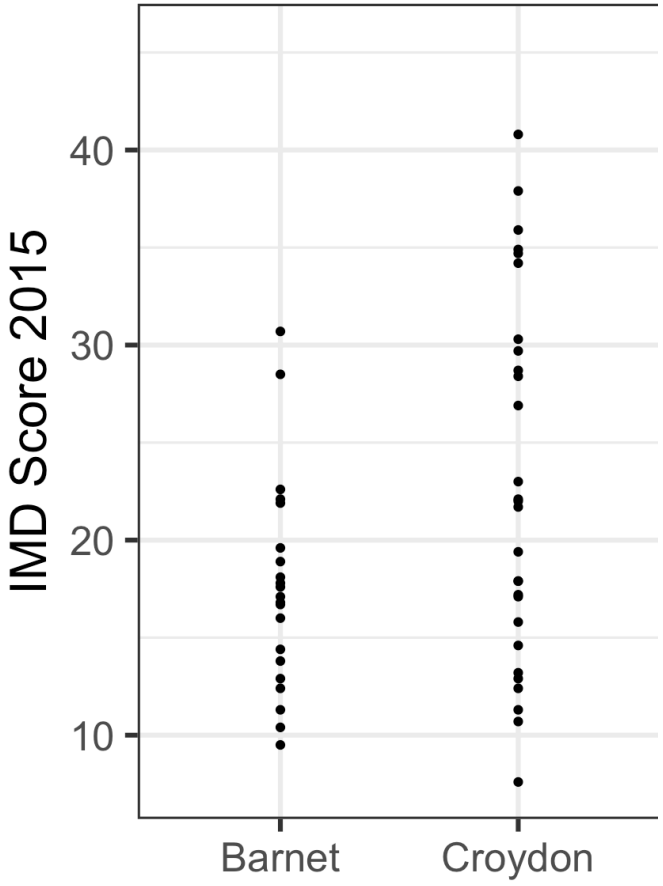
“whiskers” contain expected range

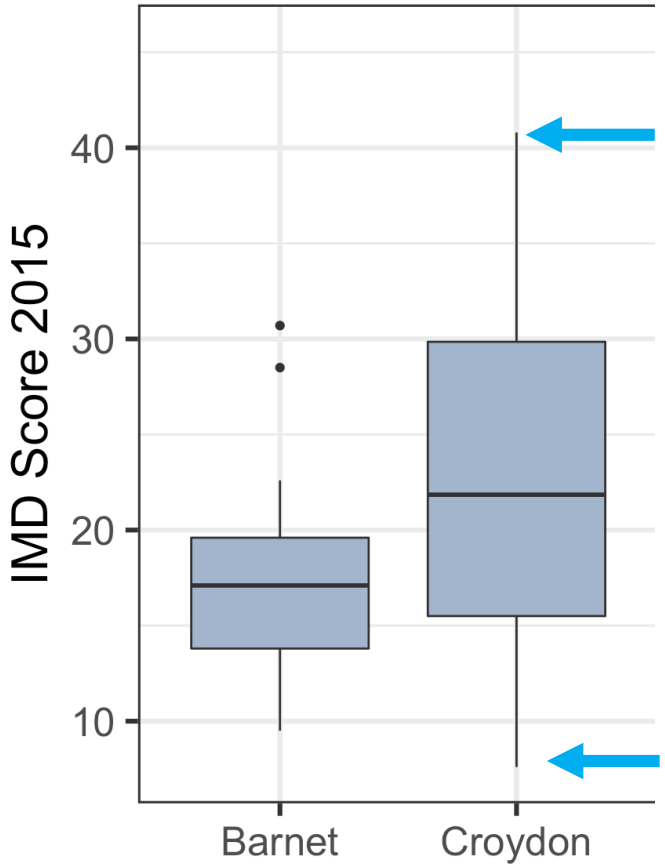
# Anatomy of a boxplot



# Anatomy of a boxplot

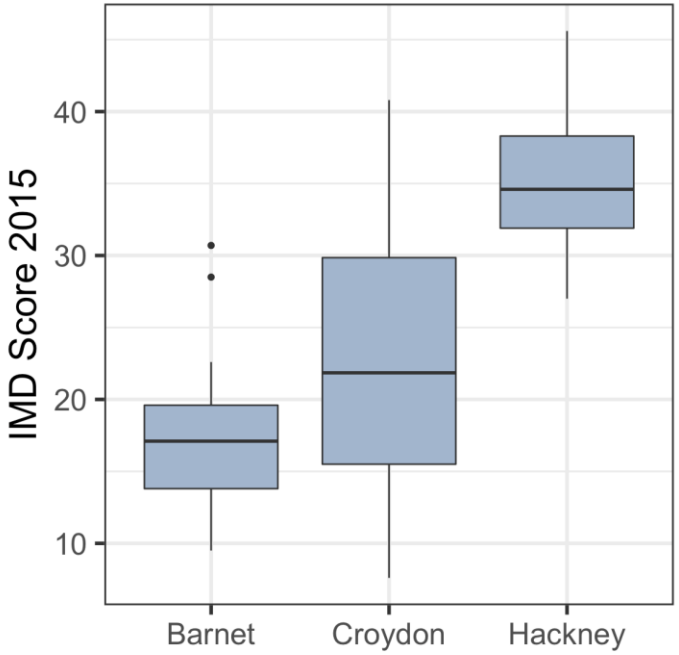


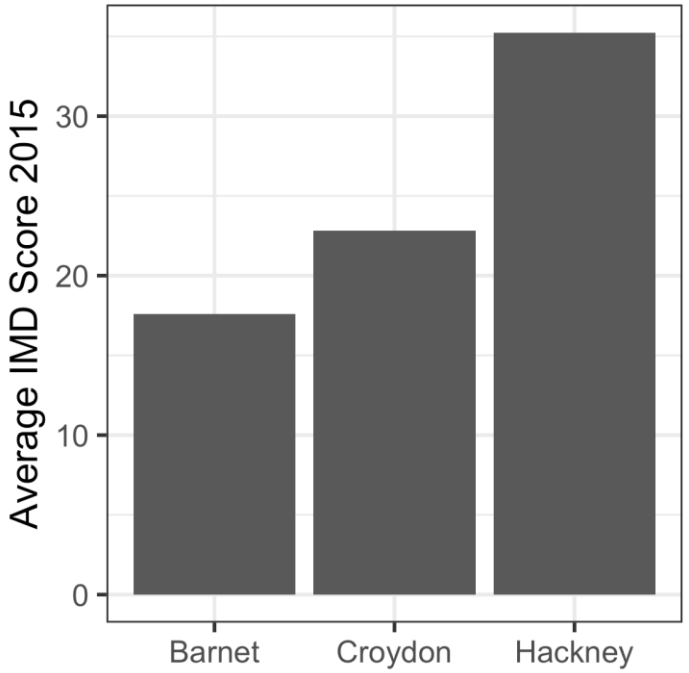




Maximum value in data

Minimum value in data

# Exercise: Think-pair-share



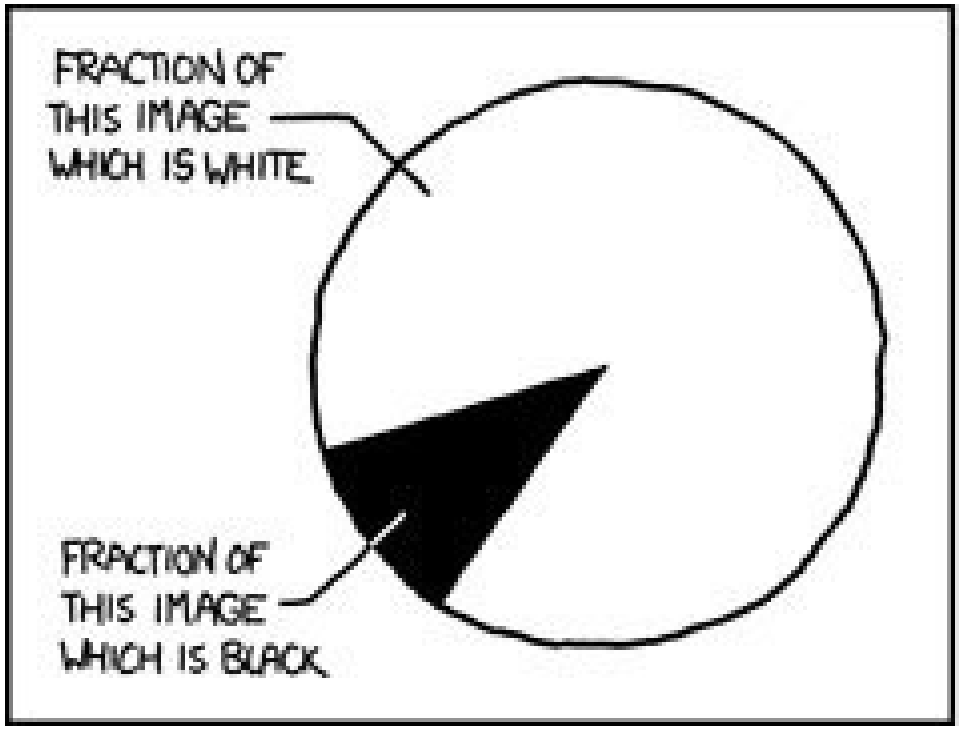

- What's something you learn from the boxplot that you don't see in a bar plot?
- When might you choose a bar plot?
- When might you choose a box plot?

# BREAK

# Understanding composition

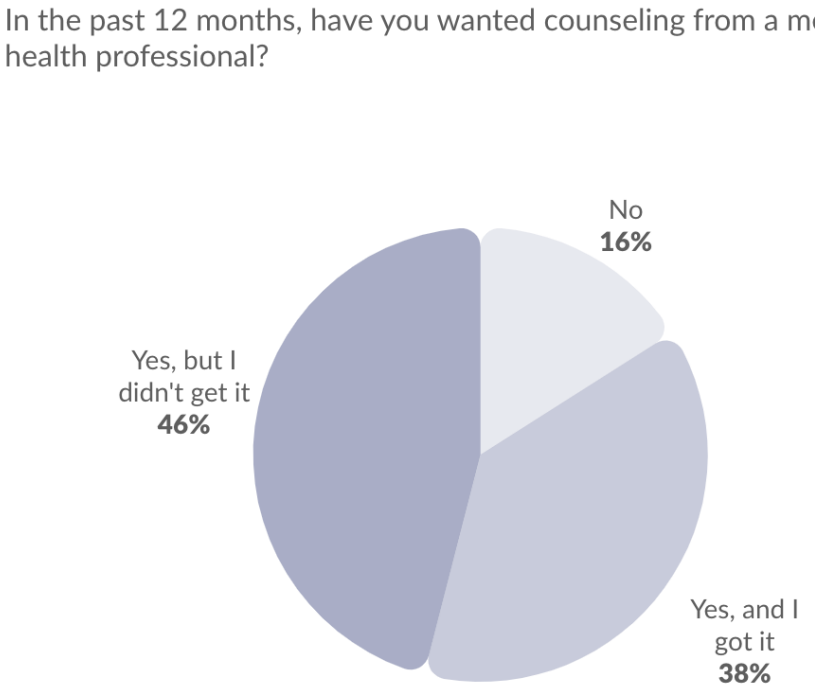
# Pie charts and donut charts

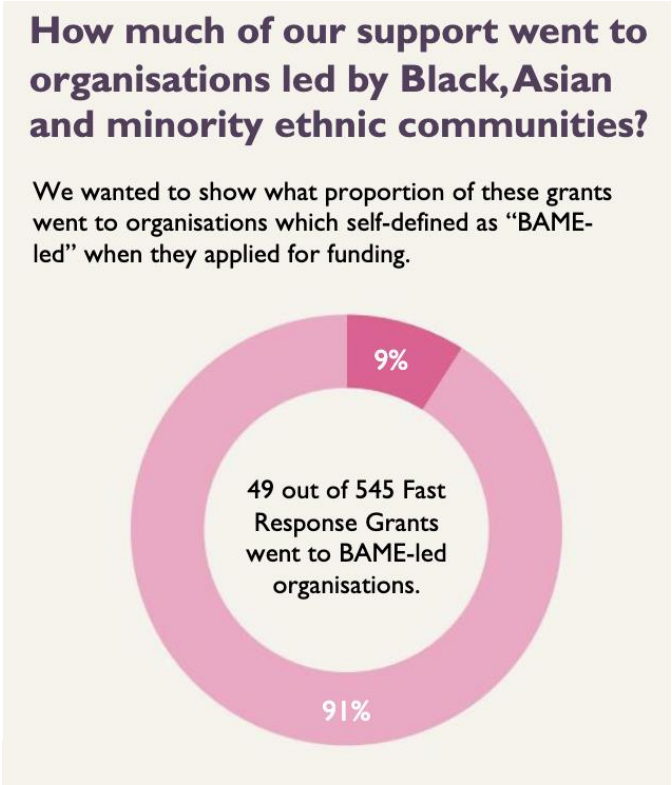
- **What is it?**
  - A circle divided into sections where each piece of the pie represents the proportion
  - Donut chart uses ring vs circle
- **Input data:** Proportion/Percentage for each value in a category
- **Why use it?** To compare proportions




Xkcd.com

# Pie charts and donut charts

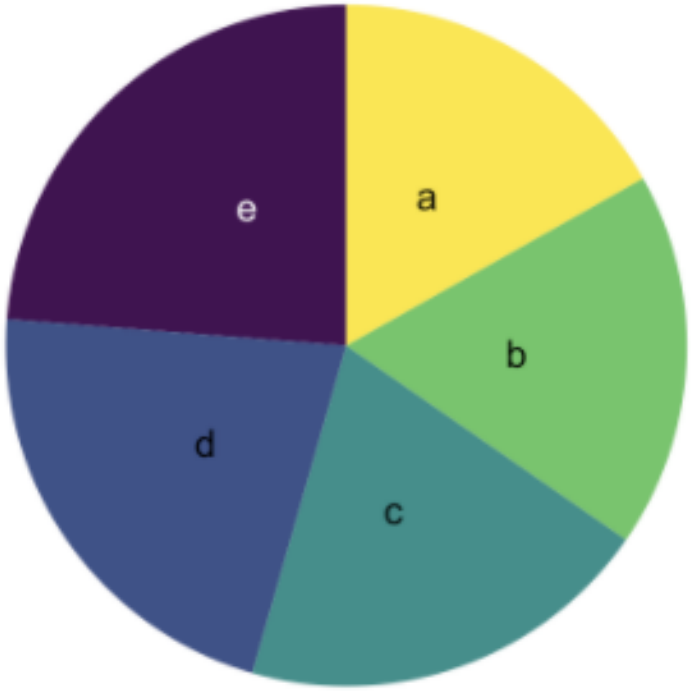



- Best used with small datasets
- Best used with simple fractions (1/10, ½, ¼)

https://esmeefairbairn.org.uk/latest-news/insights-response-covid-19/
https://www.thetrevorproject.org/survey-2020/?section=Suicide-Mental-Health

# Pie charts – exercises



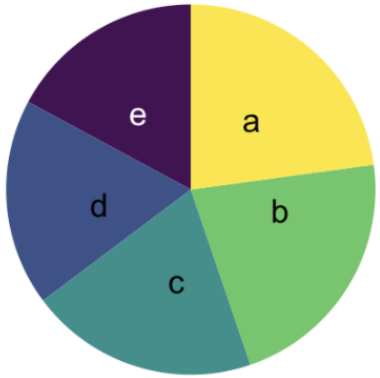

- Which slice is the biggest?
- Can you order these by value?

# Pie charts – exercises

- What is the take-home message?
- How do the values change by year?

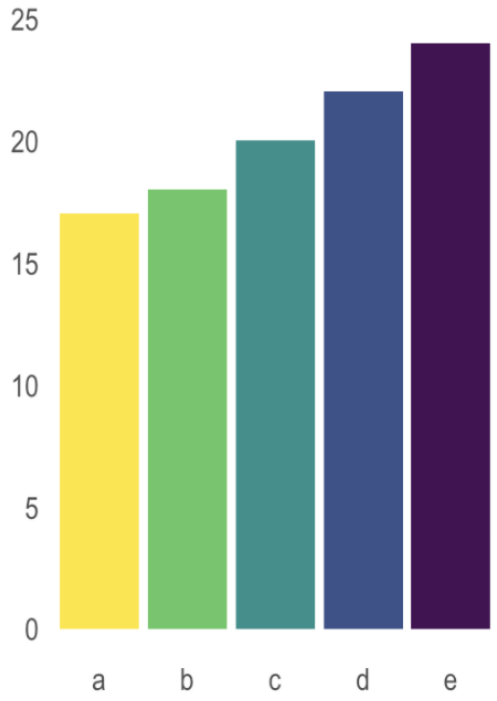
2017

2018

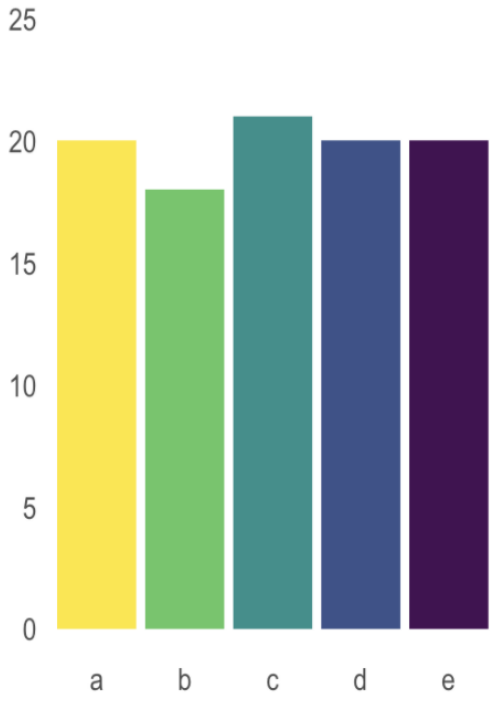



2019

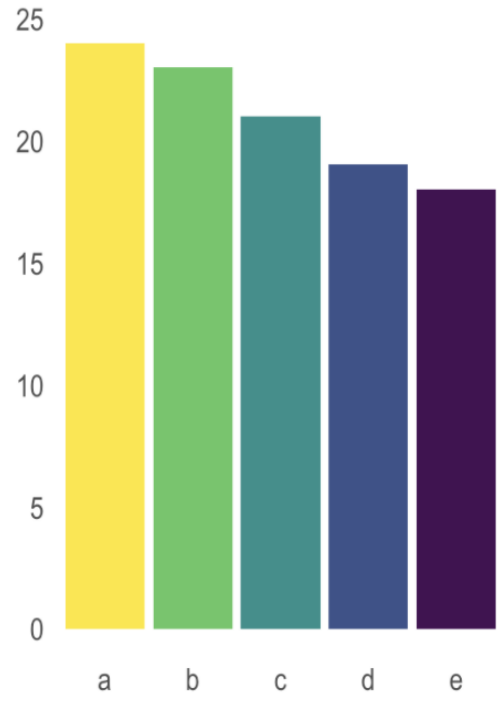
# Pie chart alternatives: grouped bar charts




2017



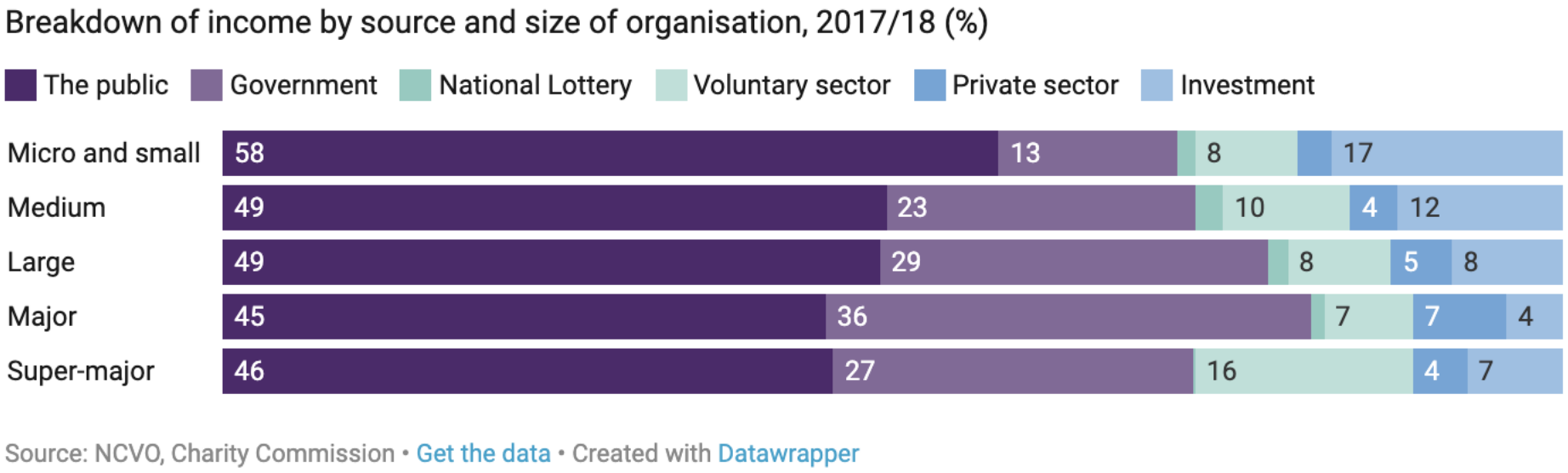

2018




2019

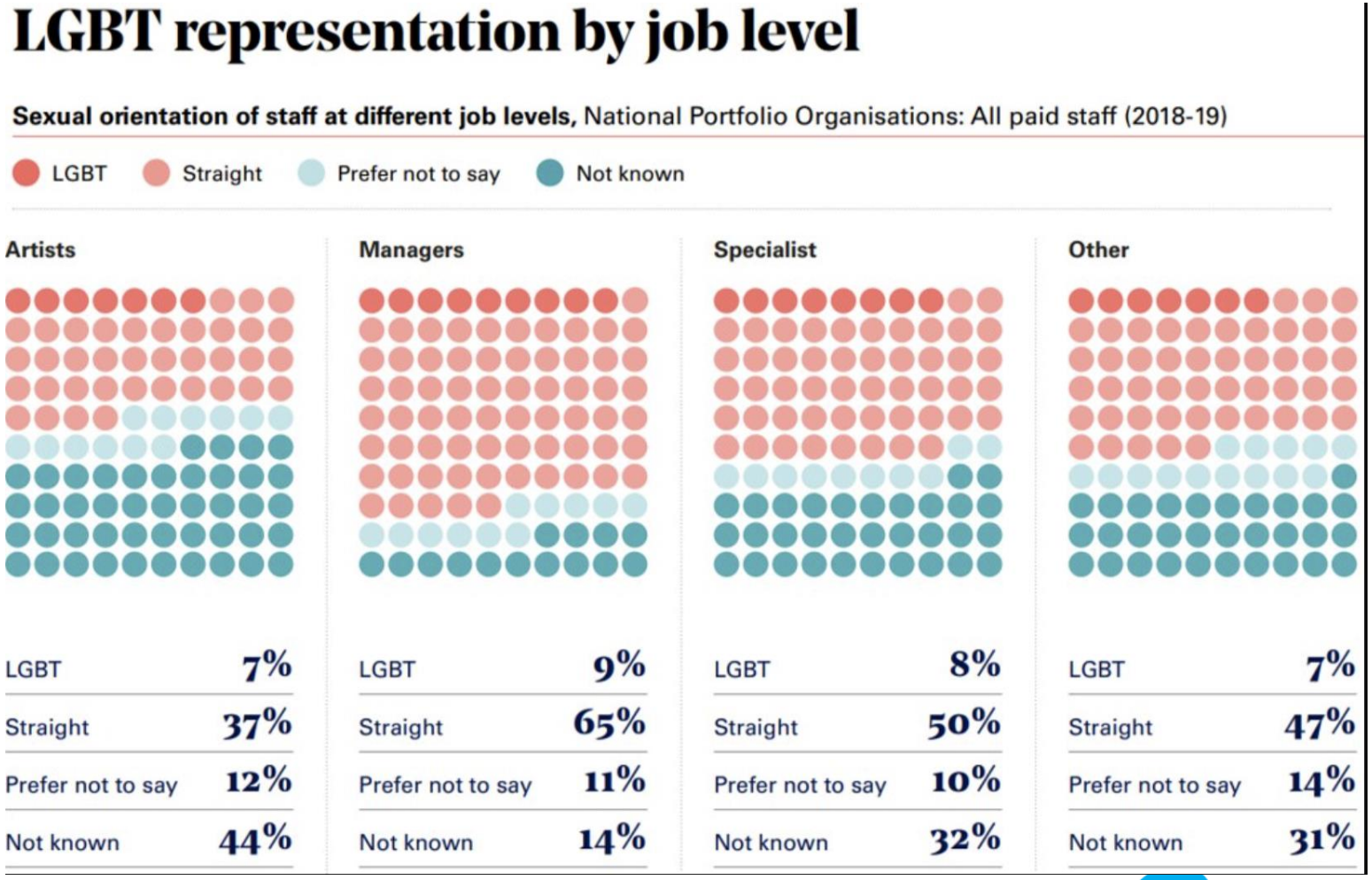
# Pie chart alternatives: stacked bar charts

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




Source: NCVO, Charity Commission • Get the data • Created with Datawrapper

# Pie chart alternatives: pictorial fraction chart



# Understanding relationships

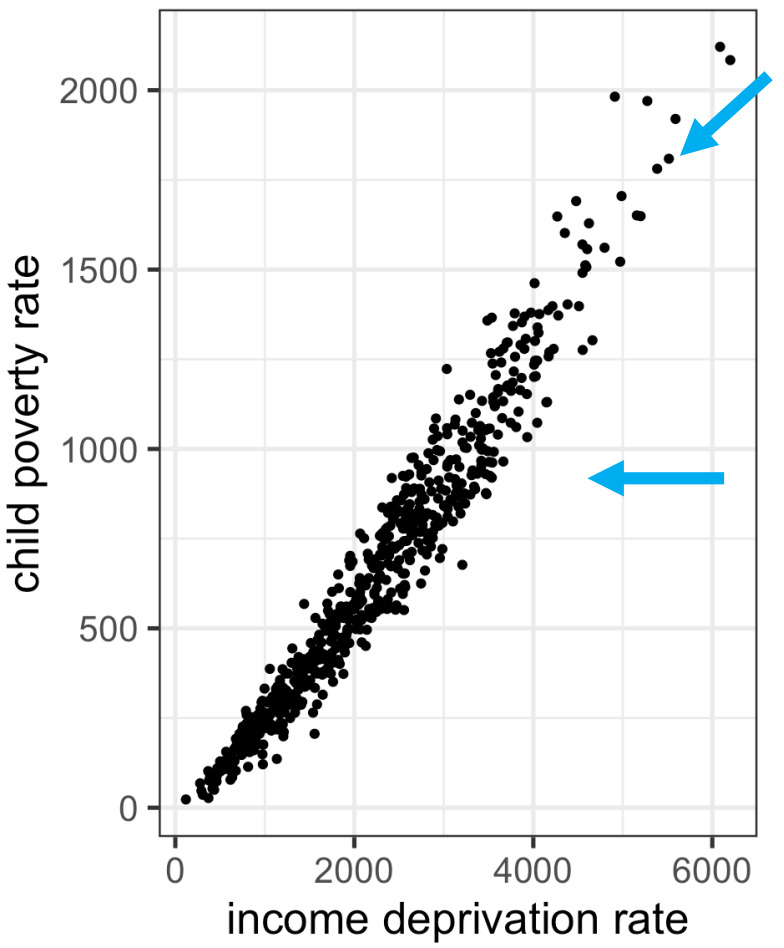
# Scatterplots, line charts, & area charts

- **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 |

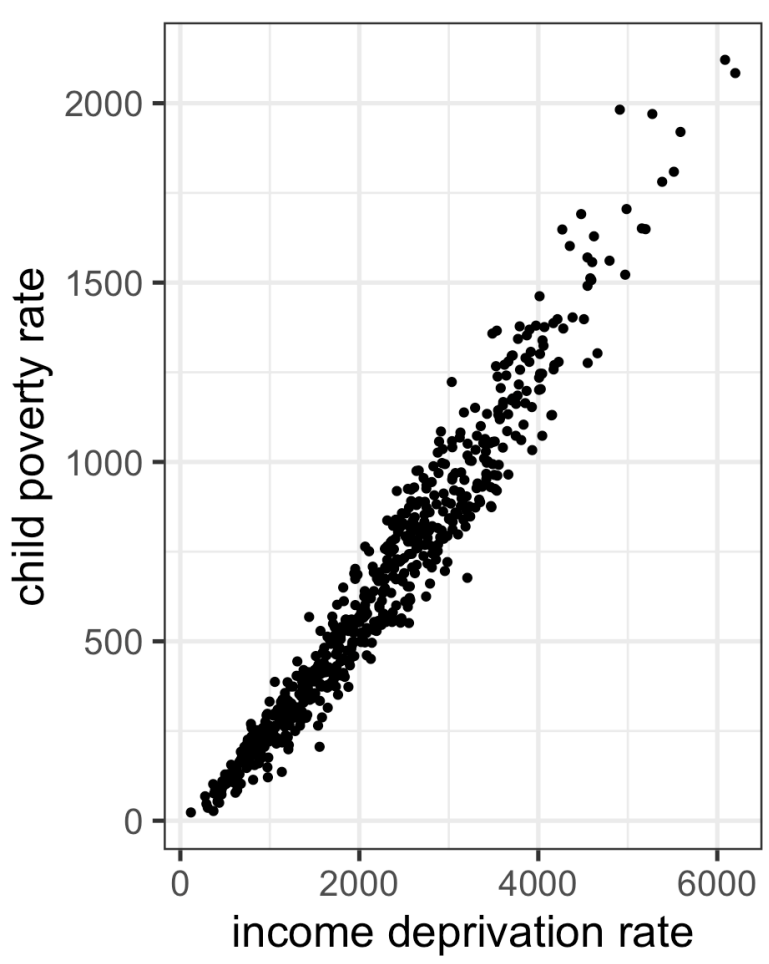
Each row is data from one ward in London

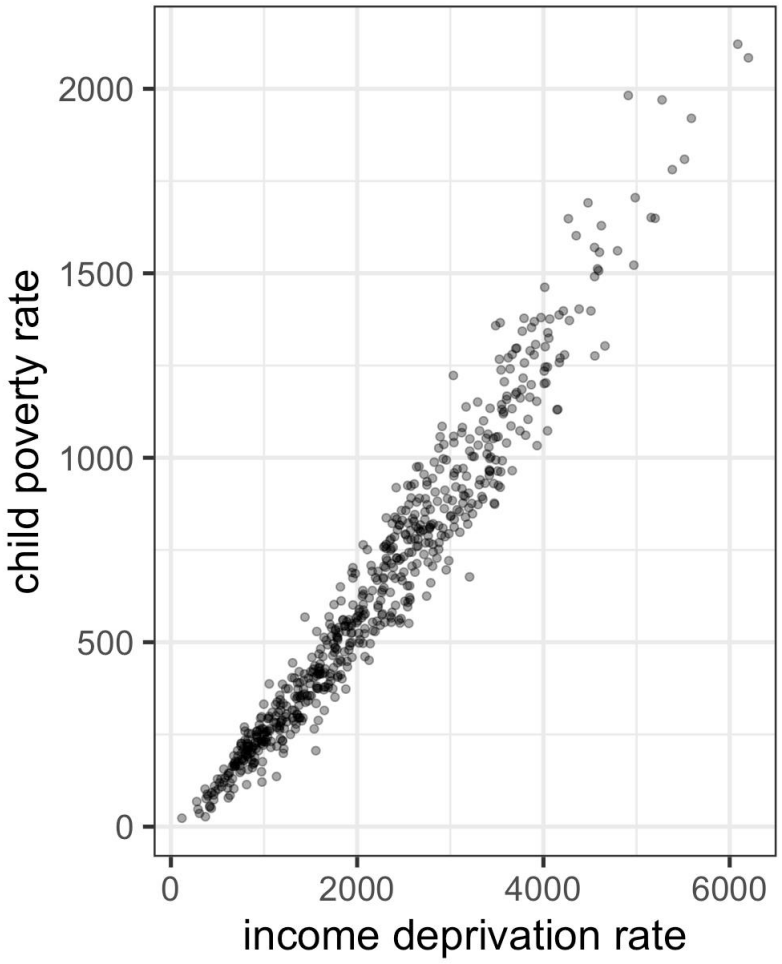



Each point is one row or sample

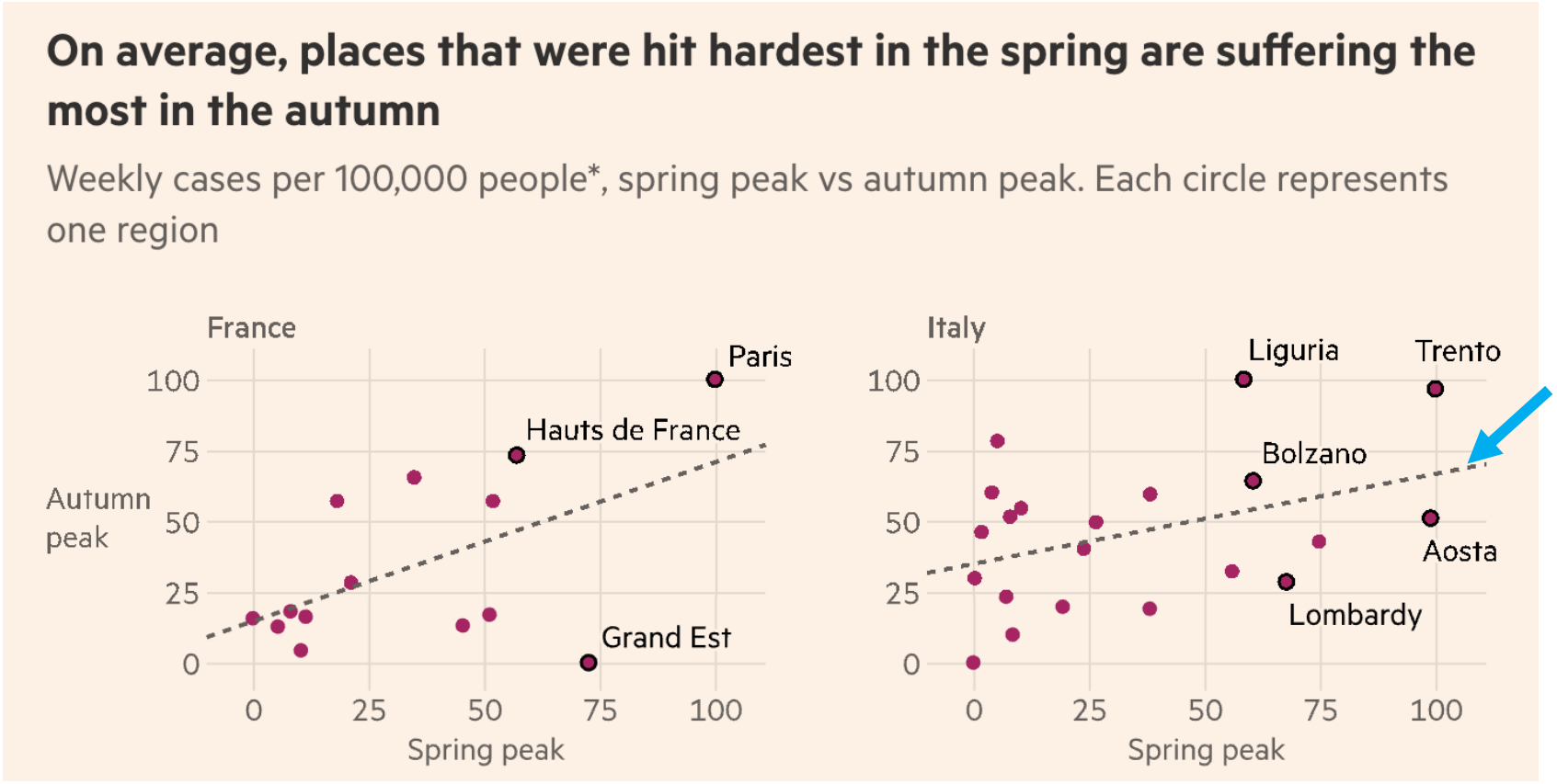
Overall pattern can tell you the type of relationship

# Common pitfalls: overlapping points



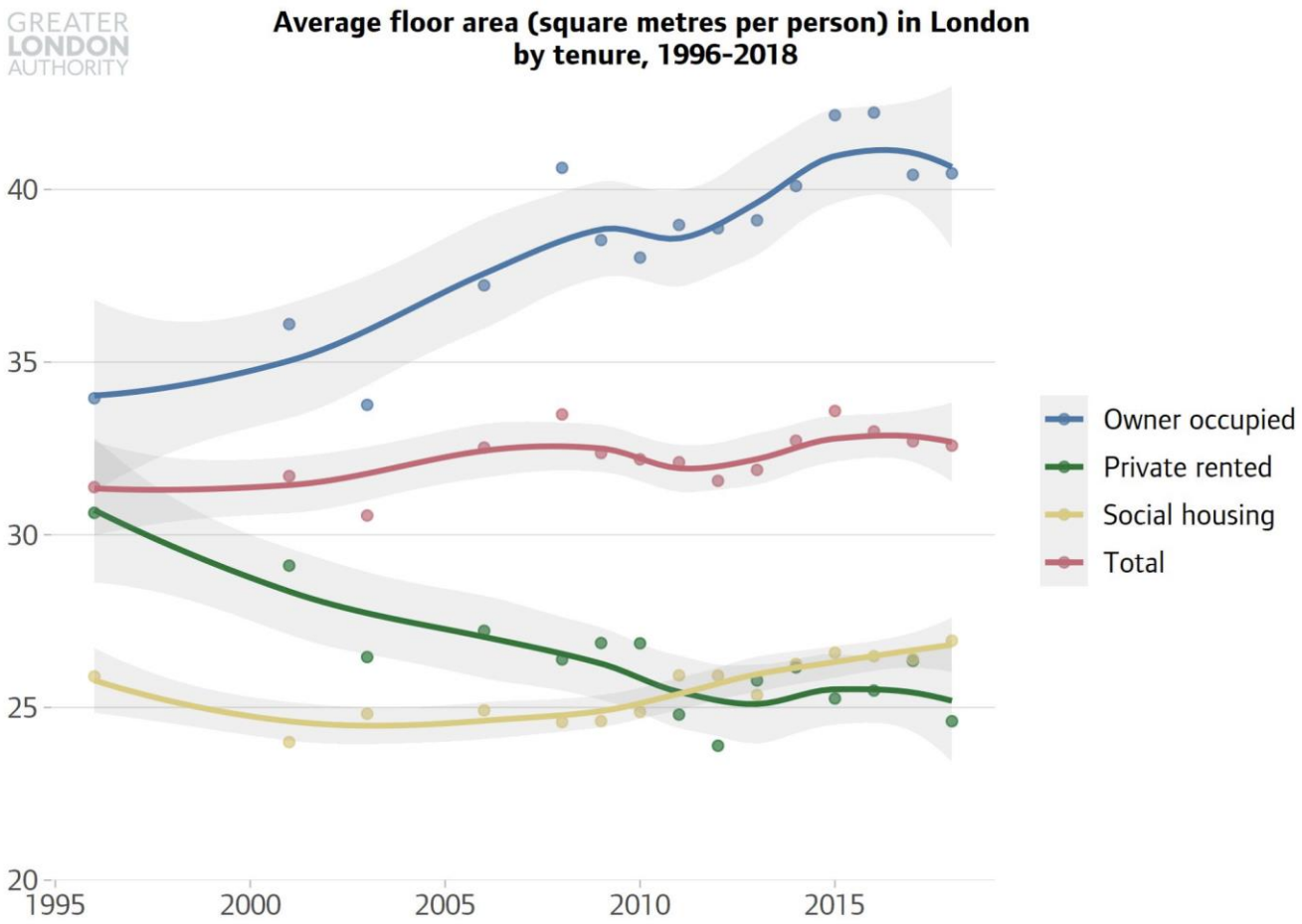



# Scatterplots

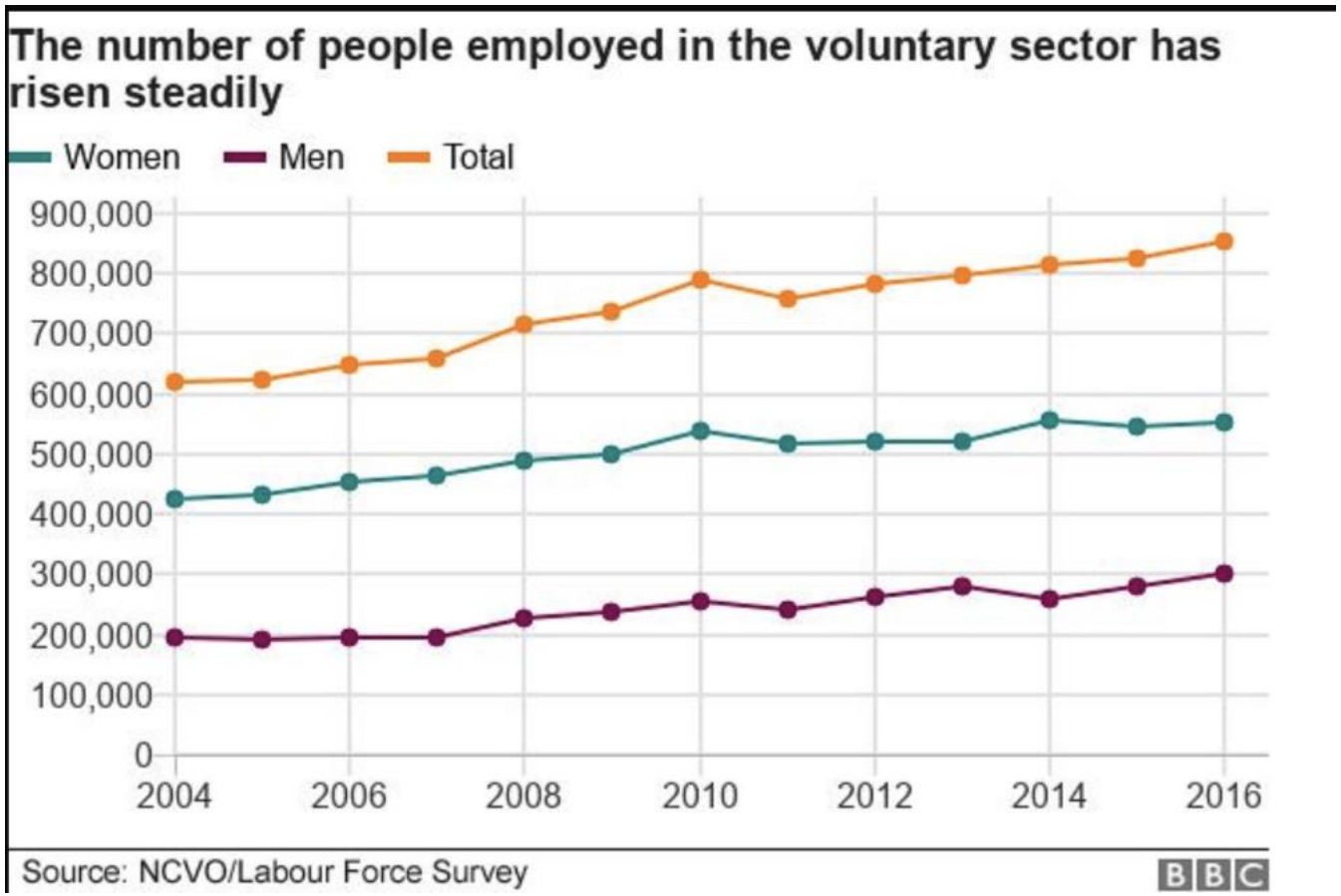



Line of best fit can help you compare trajectories

# Scatterplots

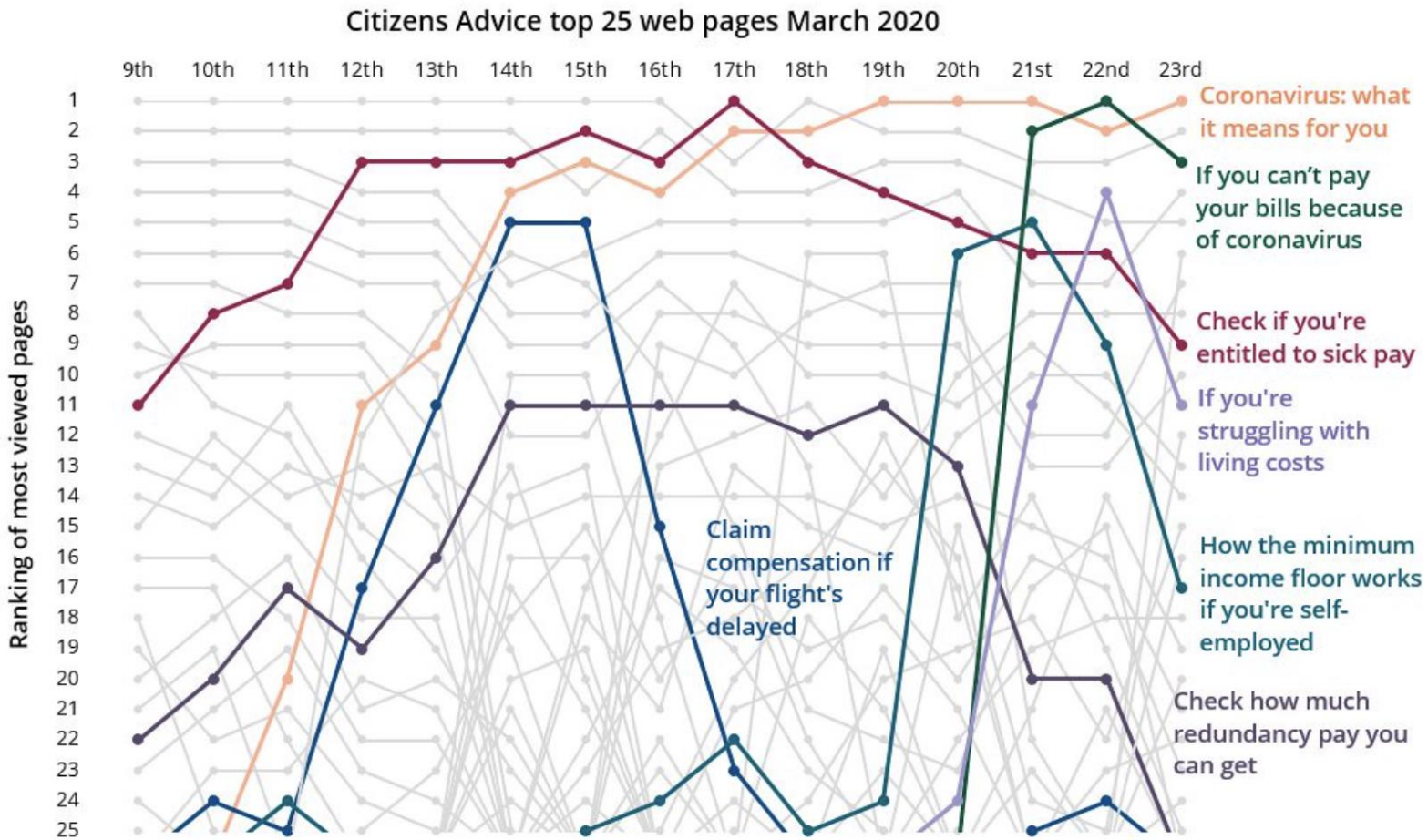


# Line charts

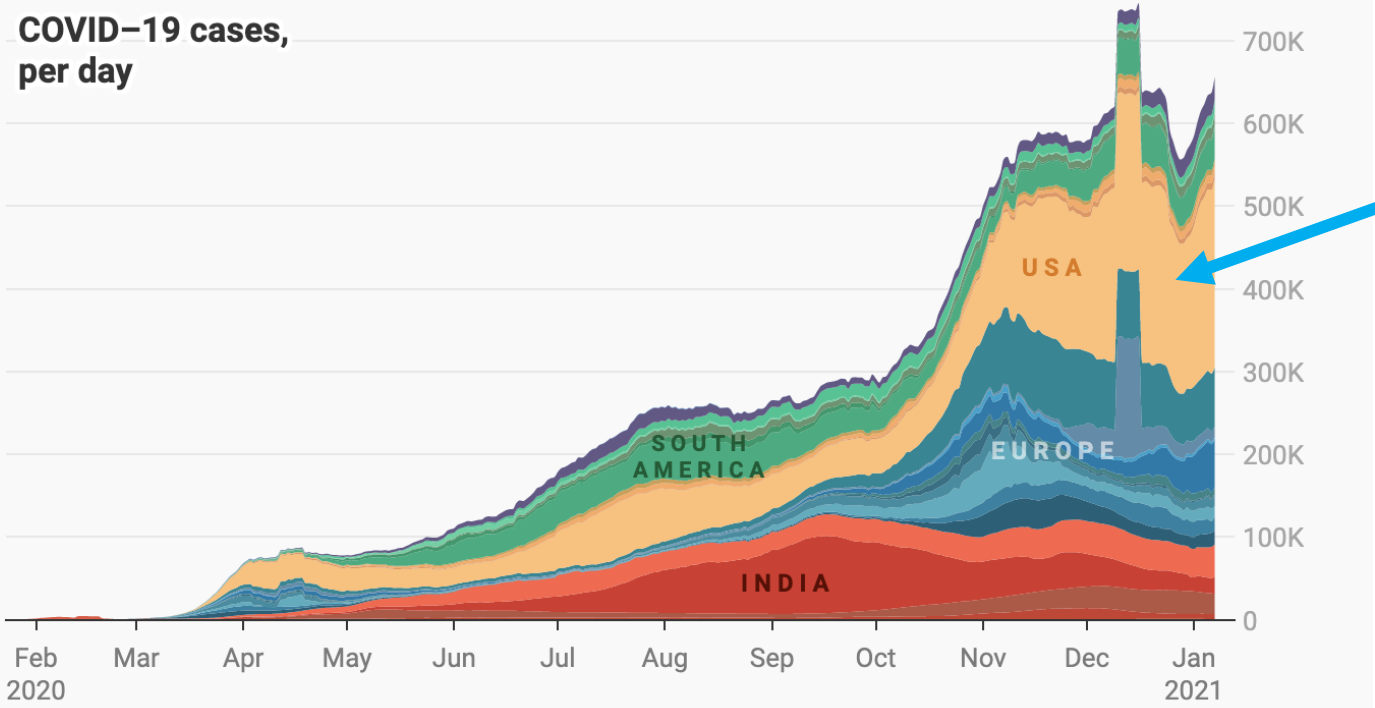



- Like a scatterplot, but the markers are connected by a line
- The x-axis (horizontal axis) values are ordered, often over time

# Common pitfalls: Spaghetti



# Area charts




*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

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

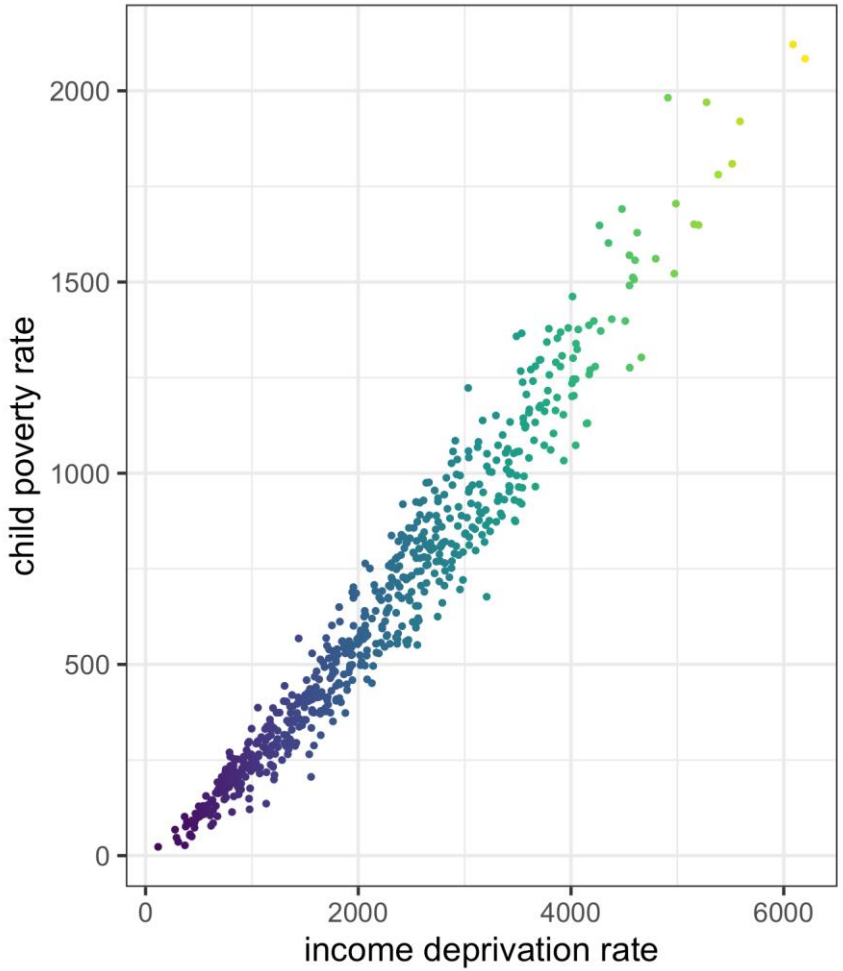
# Pause for Questions!

# Refine

- Does every part of the figure have a purpose?
- Consider colours
  - Use consistently
  - Use to highlight important information
  - Consider accessibility

# Thinking about colour

# Avoid colouring for the sake of it



# Use colour scales appropriately

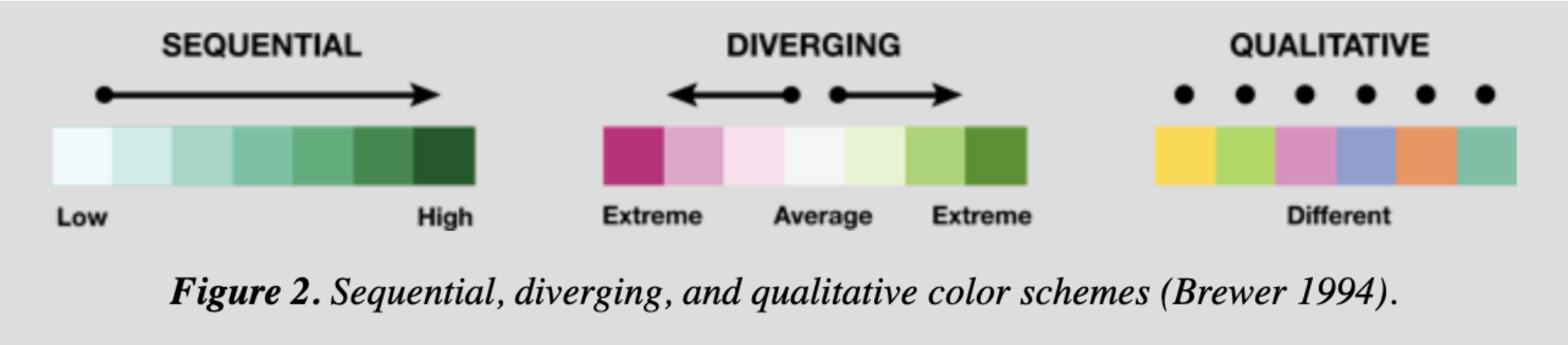
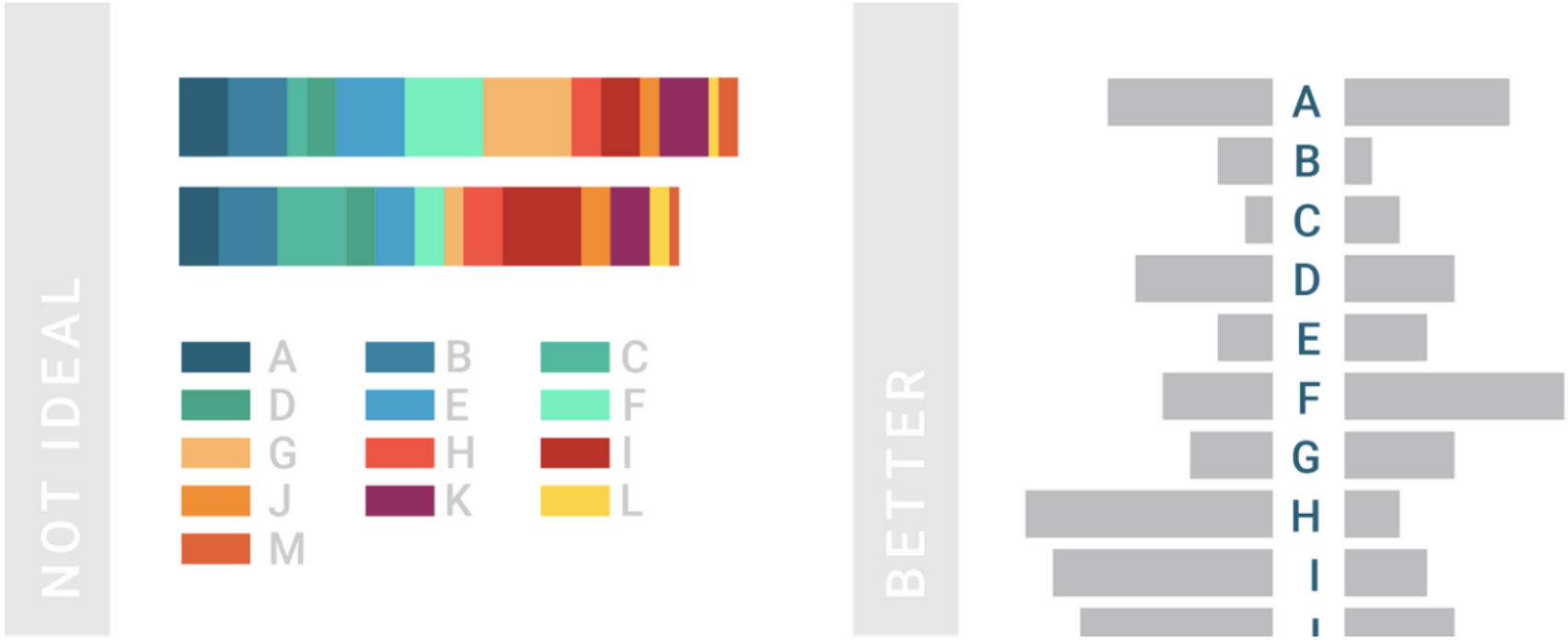



***Figure 2.*** *Sequential, diverging, and qualitative color schemes (Brewer 1994).*

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

# Avoid colouring too many categories




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

# 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

# Design for colour-vision deficiency

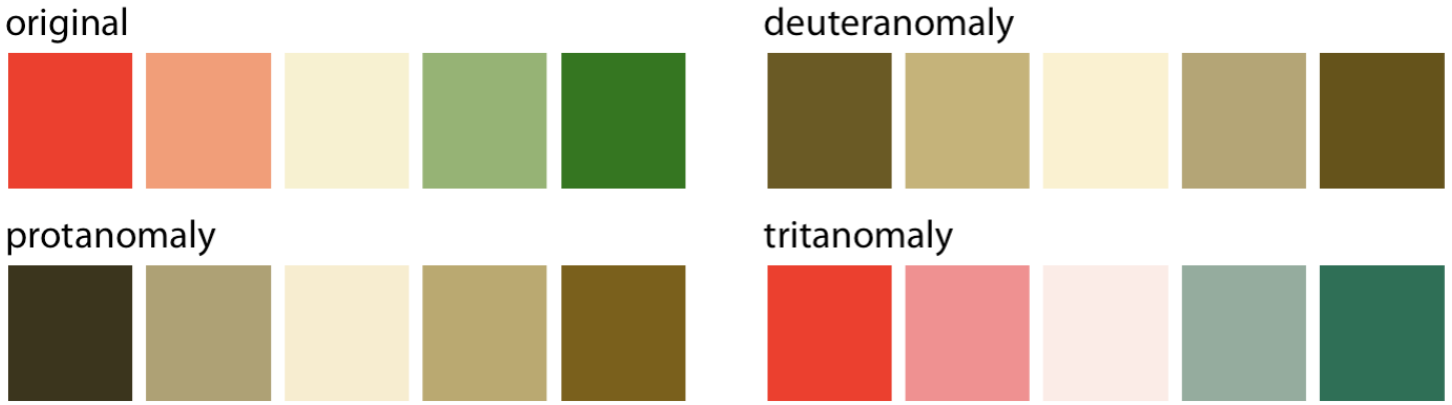



Figure 19.7: A red–green contrast becomes indistinguishable under red–green cvd (deuteranomaly or protanomaly).

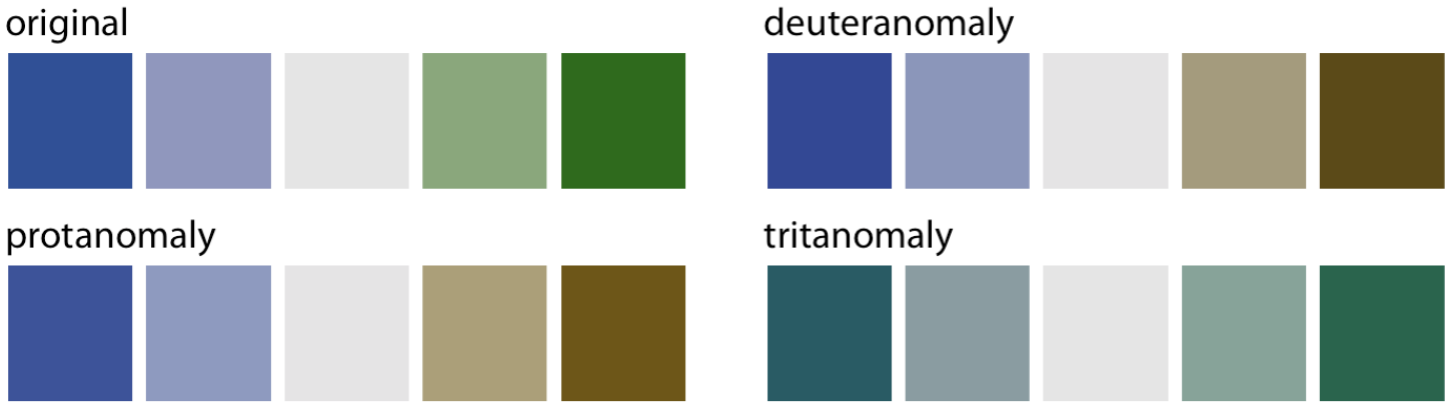
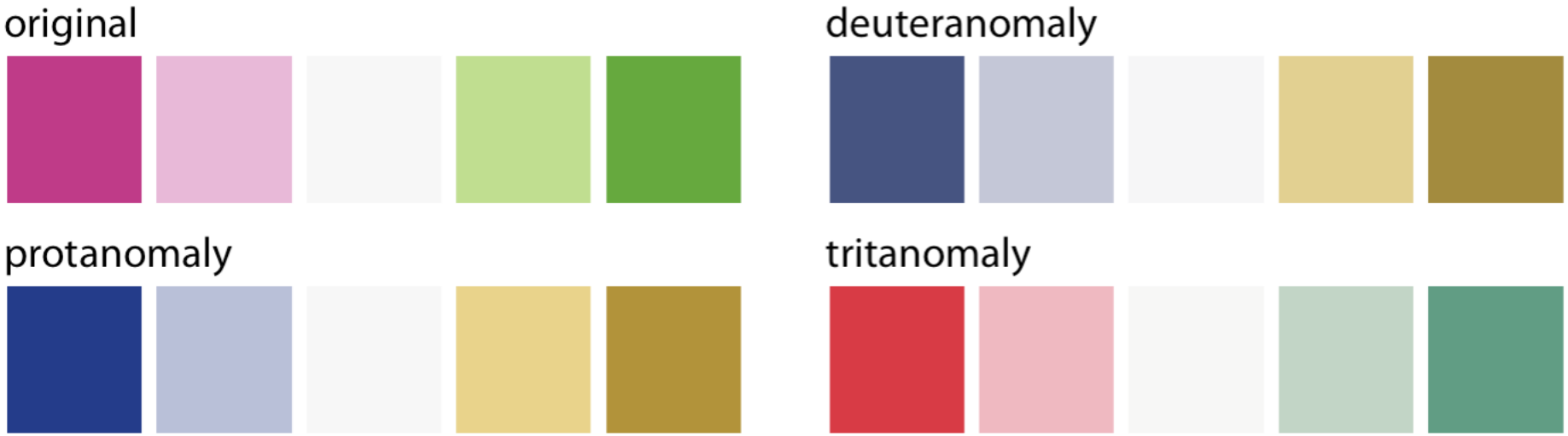



Figure 19.8: A blue–green contrast becomes indistinguishable under blue–yellow cvd (tritanomaly).

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

# Design for colour-vision deficiency




https://colorbrewer2.org

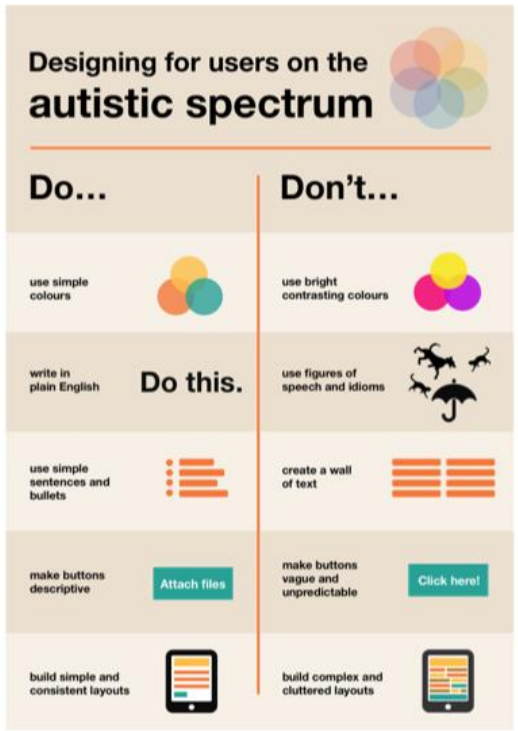
https://clauswilke.com/dataviz/color-pitfalls.html#not-designing-for-color-vision-

# 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

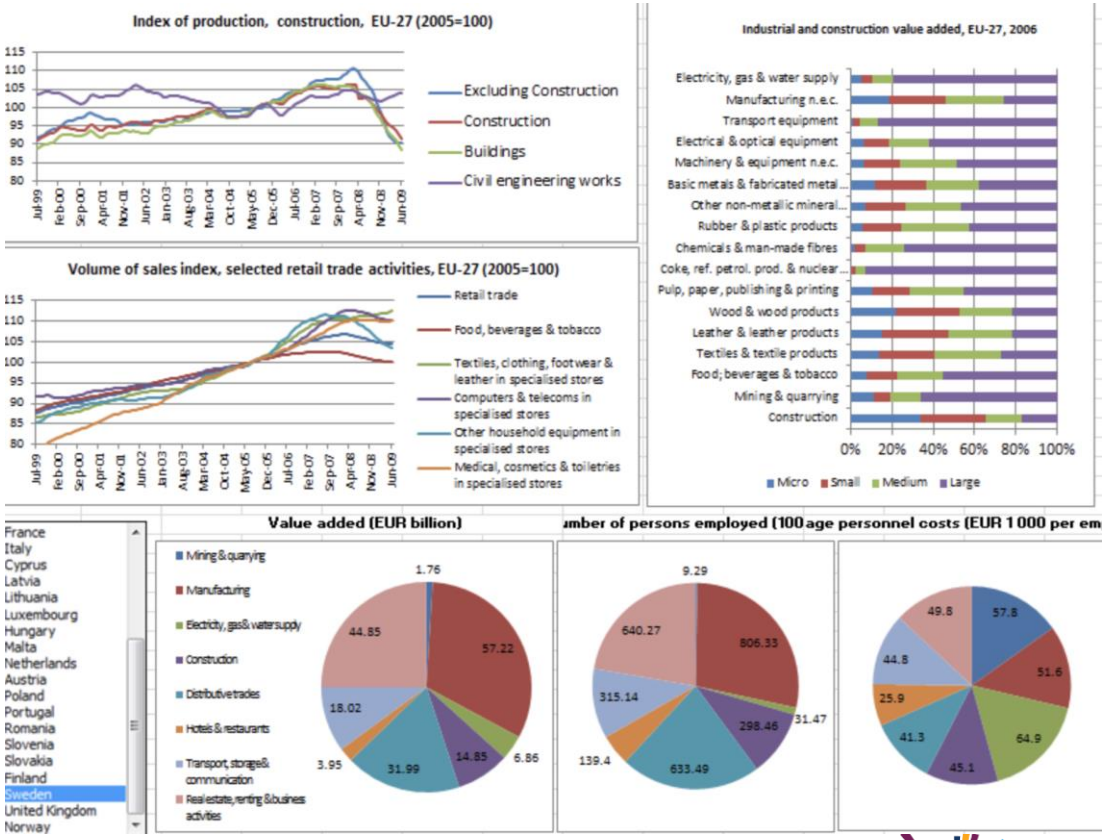
↘ https://accessibility.blog.gov.uk/2016/09/02/dos-and-donts-on-designing-for-accessibility/



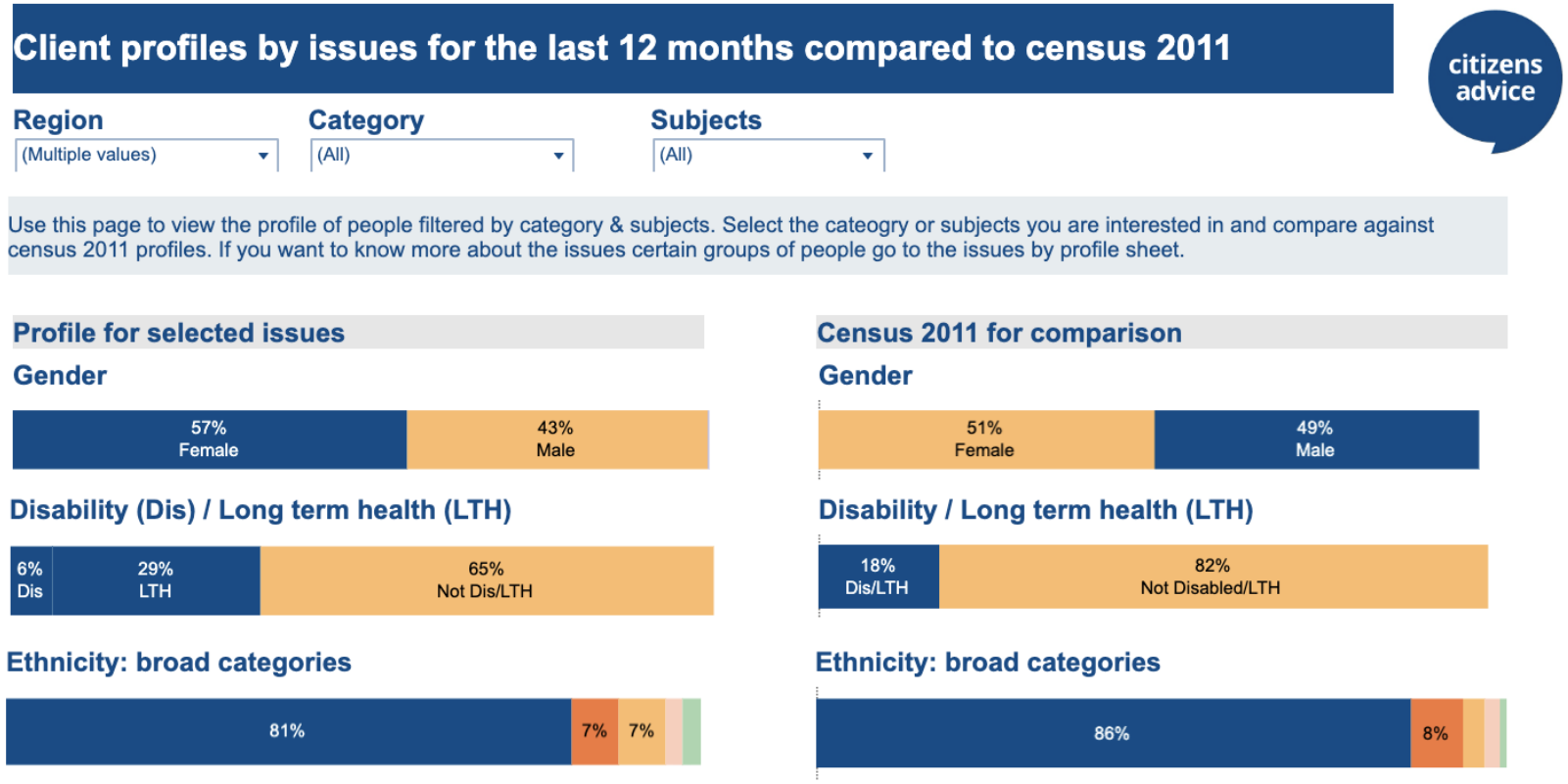
# Making your data viz

**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

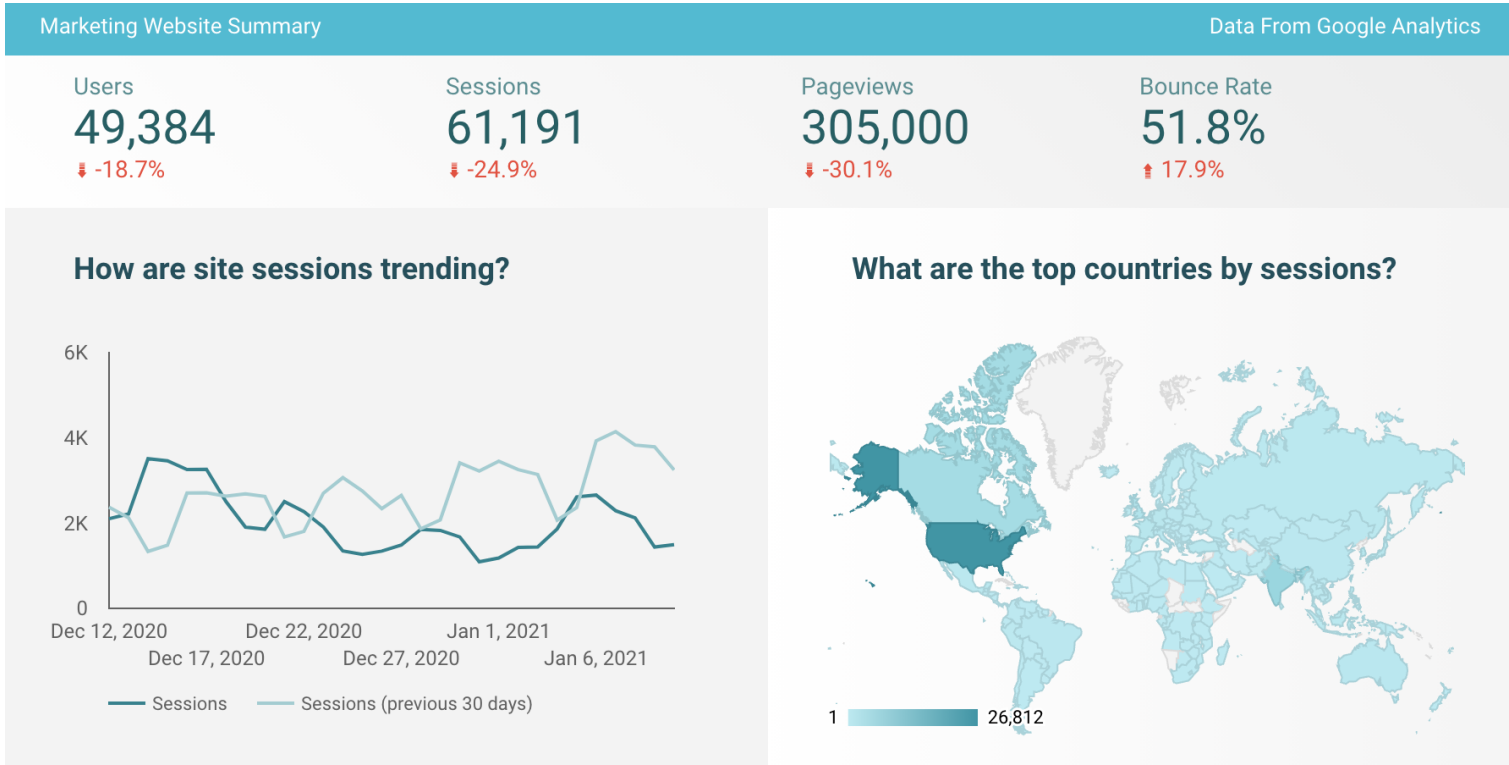



https://public.tableau.com/en-gb/gallery/?tab=viz-of-the-day&type=viz-of-the-day

# Data Studio

- **Good for:**
  - Dashboards
  - Integrates with other google tools
- **Limitations:**
  - Fewer visualisation options than Tableau

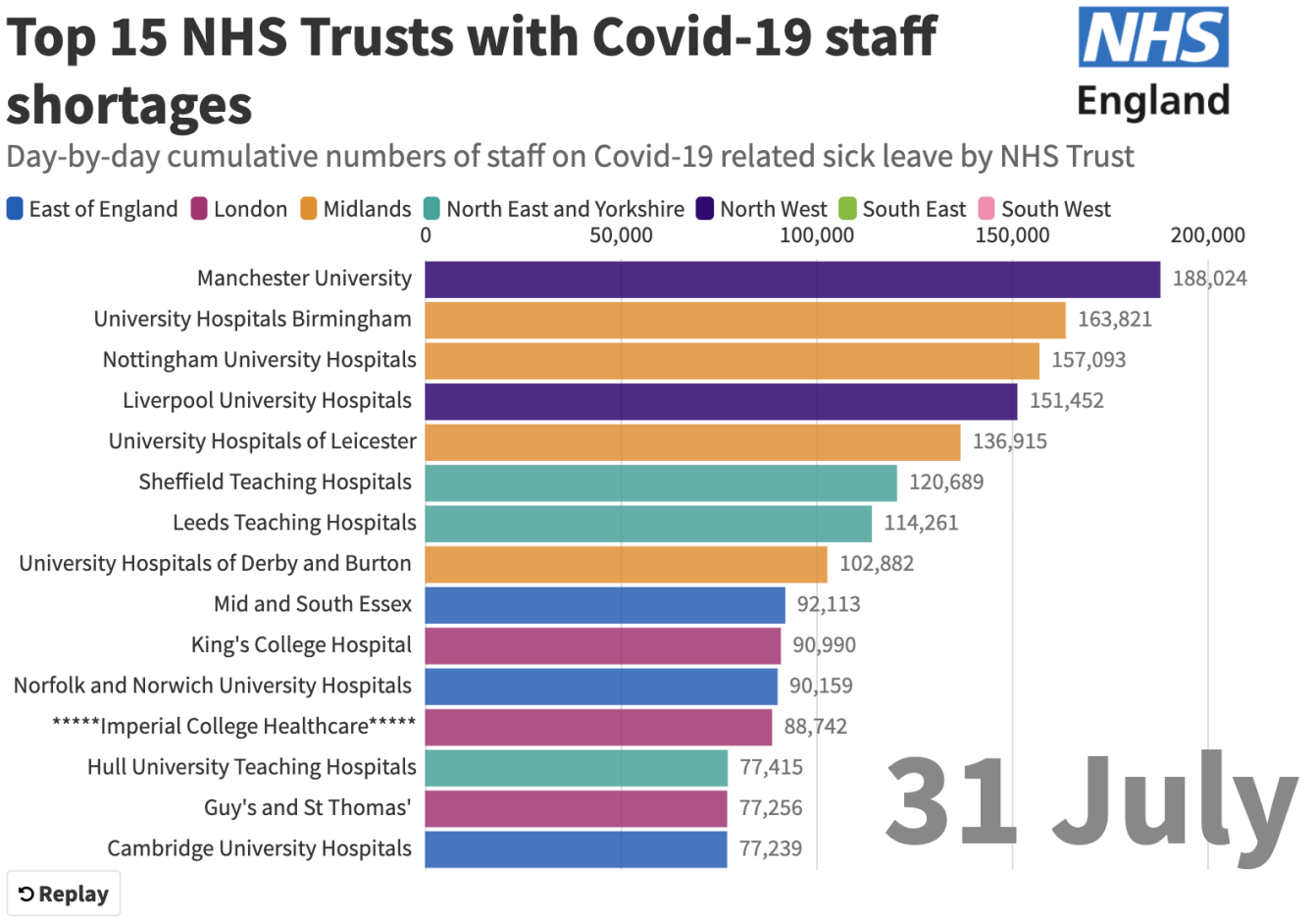
- **Cost:** Gsuite is free for charities




https://datastudio.google.com/navigation/rep

# Flourish

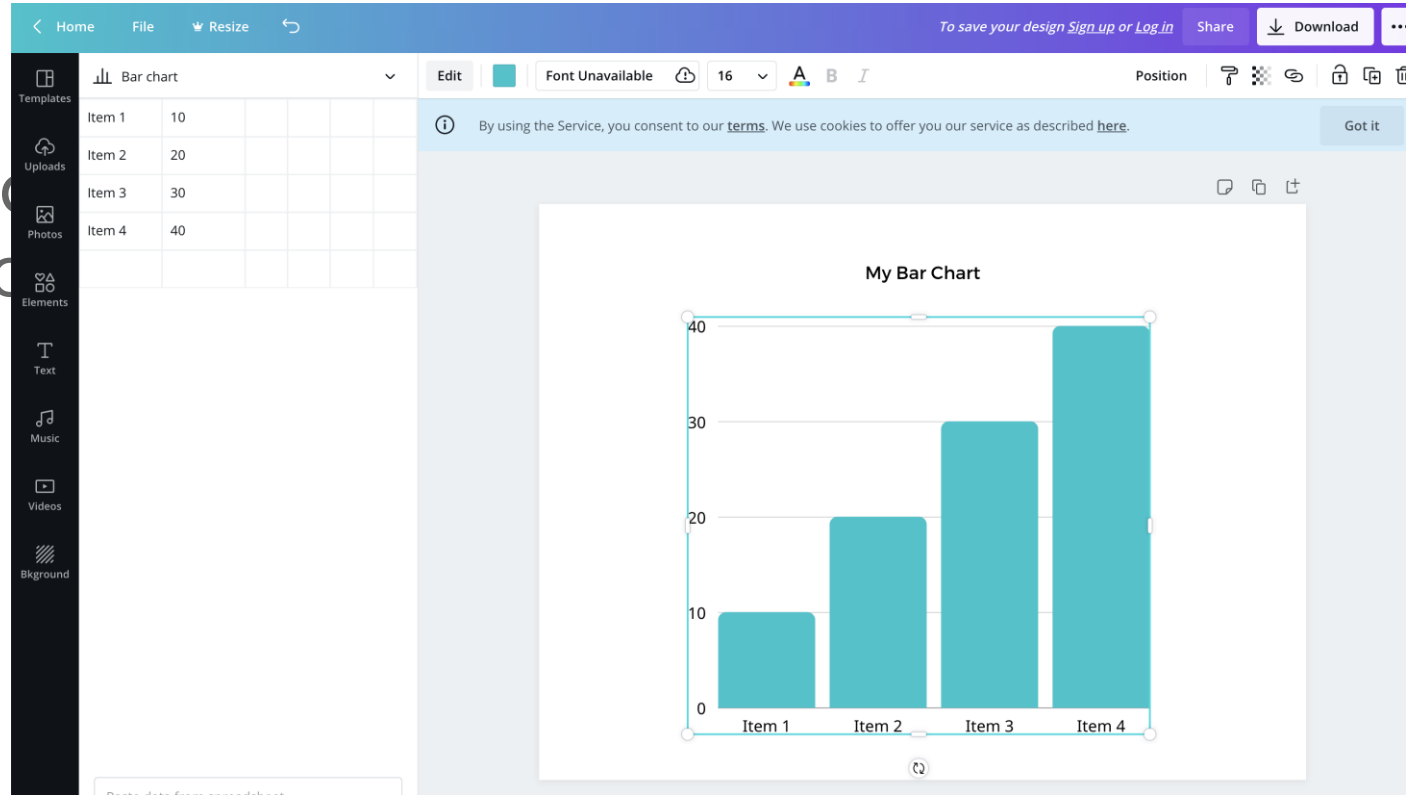
- **Good for:**
  - Storytelling and presentations
  - Interactive charts
  - Embedding in websites
- **Limitations:**
  - Free plan makes data and projects public
- **Cost:** Free option, non-profit discount




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 gr
  - Simple charts/infographic
  - Drag-and-drop
- **Limitations:**
  - Data pasted from spreadsheet

- **Cost:** Free for non-profits




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

infogram

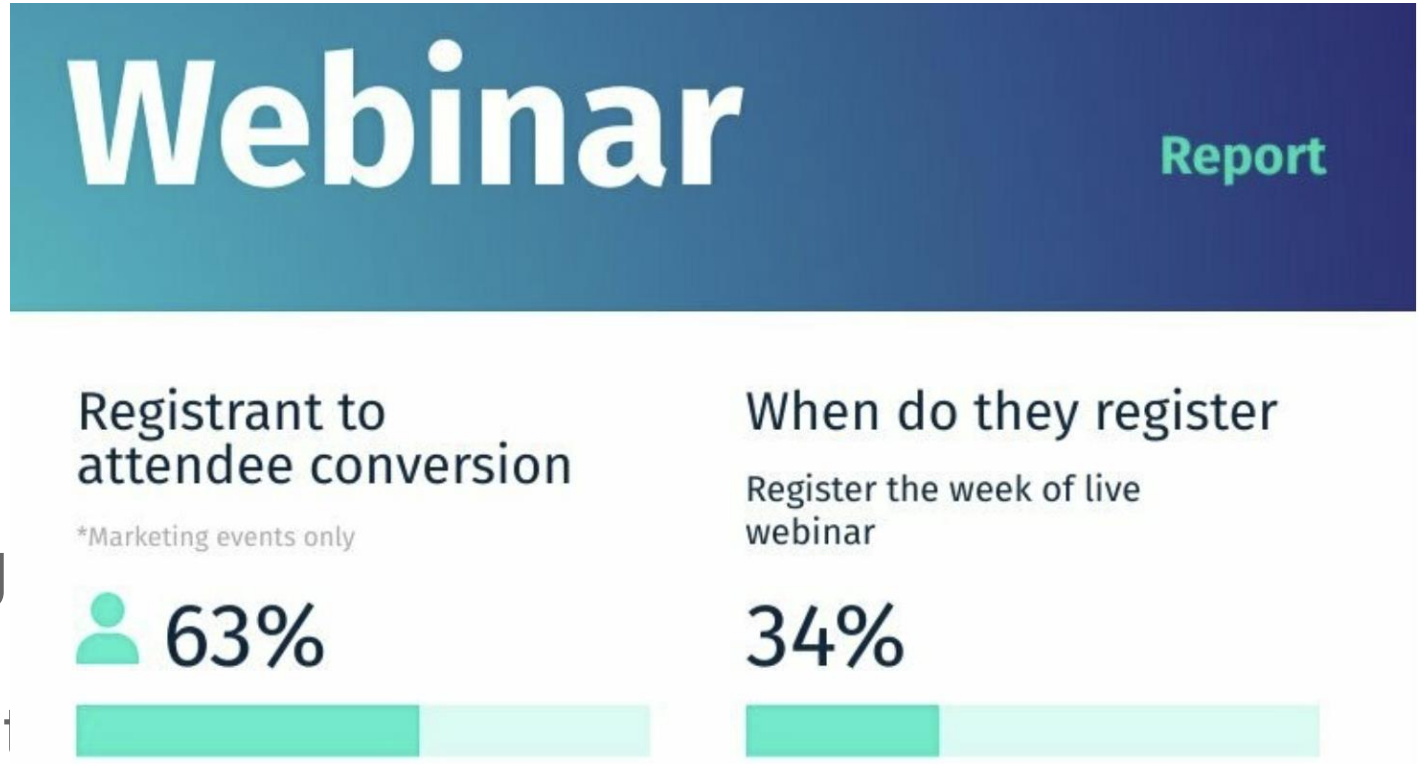
- **Good for:**
  - Interactive charts and maps
  - Presentations
  - Drag and drop
- **Limitations:**
  - Free version makes visualisations public

- **Cost:** Free option

https://infogram.com/

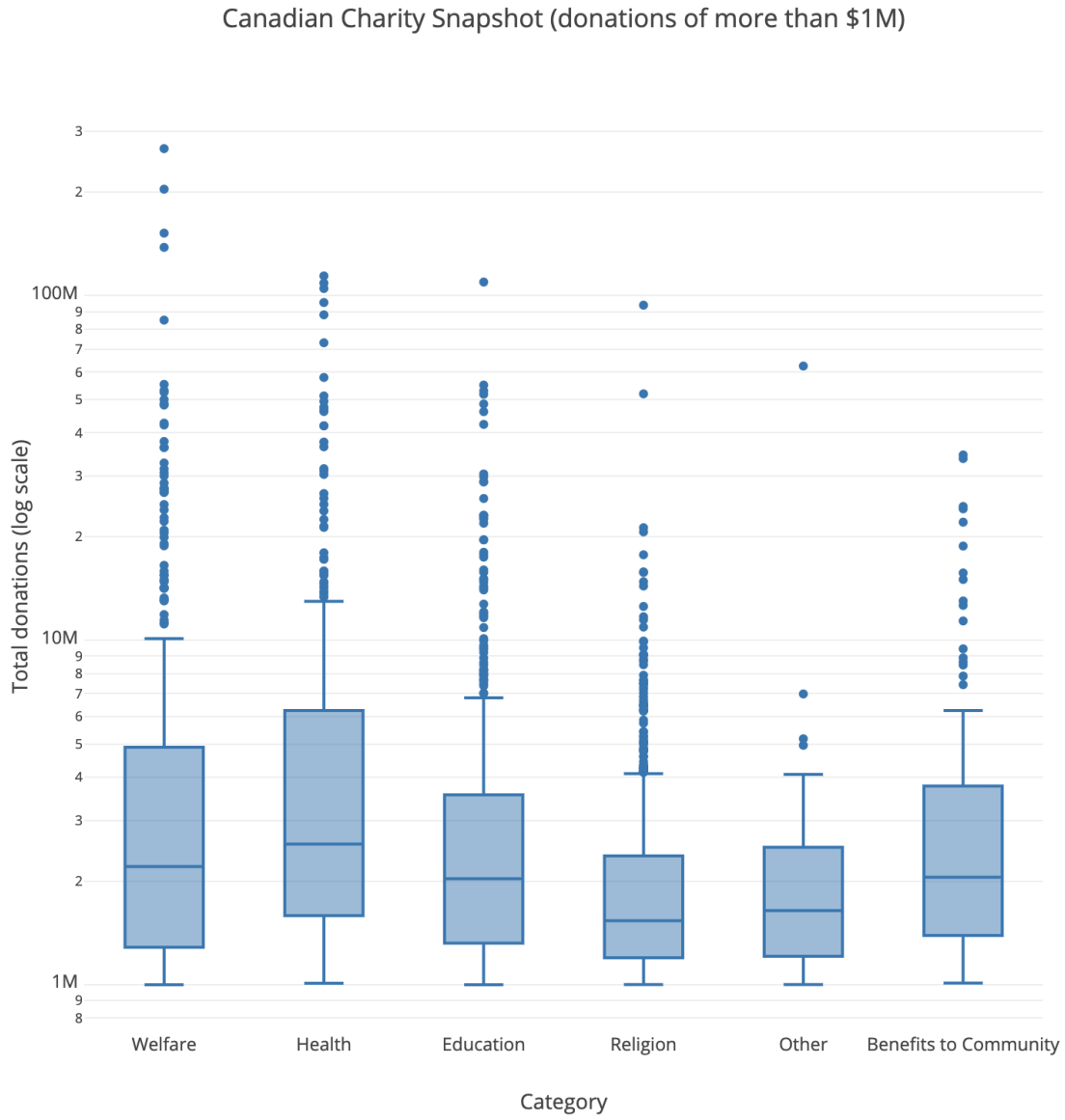
PIKTOCHART

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




https://piktochart.com/

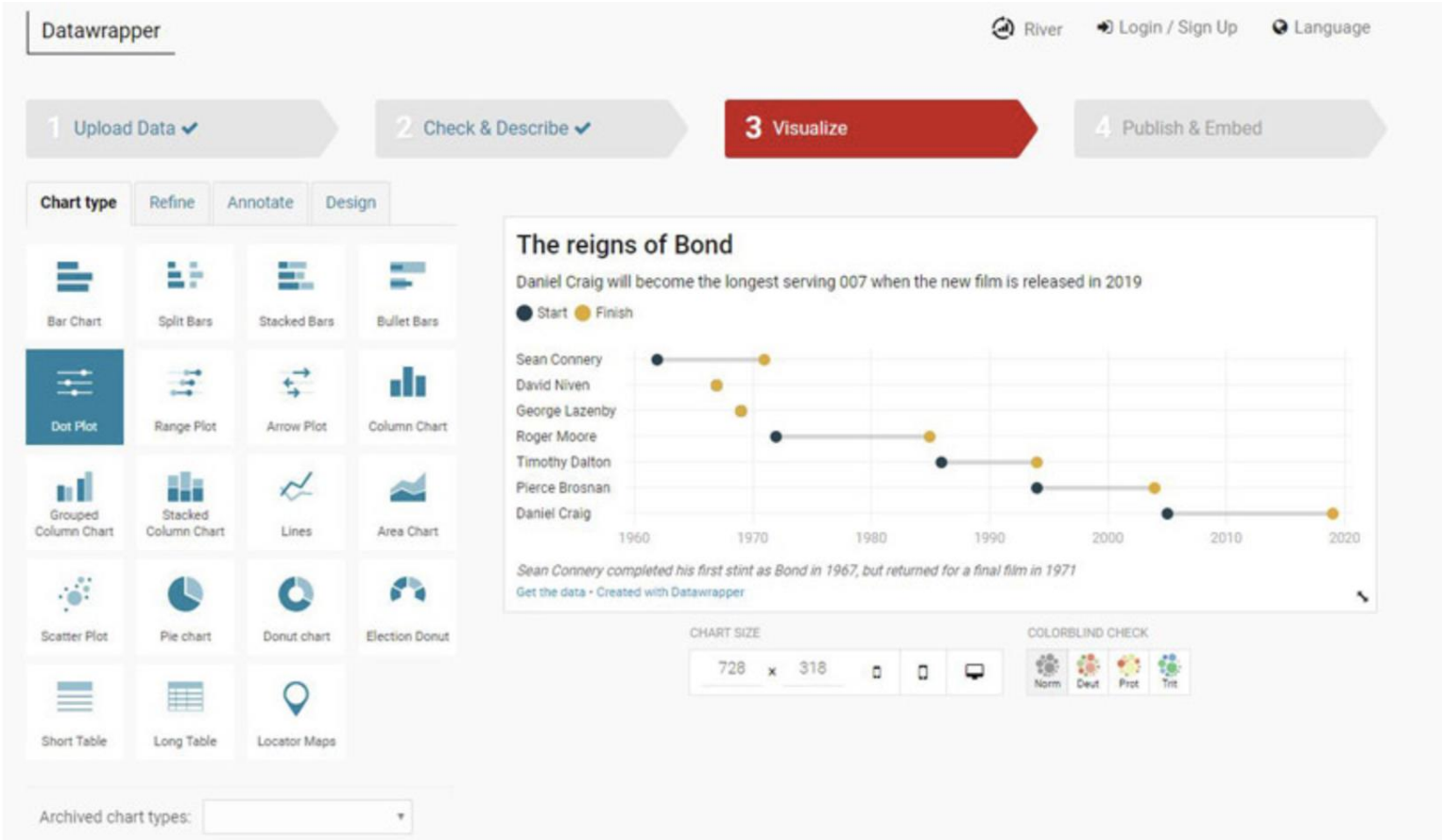
- **Good for:**
  - Interactive charts
  - Embedding in websites
  - Attractive default formats
- **Limitations:**
  - Paid options is very expensive
- **Cost:** Free option, but limited to public charts




https://plotly.com/chart-studio/

# Datawrapper

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




https://www.datawrapper.de/

# Datawrapper

- **Datawrapper:** https://www.datawrapper.de/
- **Data:** https://docs.google.com/spreadsheets/d/1ZUR1tlAzfU1vxwOhYrVy0Z_BebokX63NSNnPIR2MmQo/edit?usp=sharing

# Resources

## Office Hours

Our Office Hours are free support and advice sessions for any social change organisation with a data question. Drop by and talk to a data expert for an hour, whether your issue is operational or strategic, specific or speculative.

https://datakind.org.uk/what-we-do/office-hours/

# 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, DATAKIND UK

dulcie@datakind.org.uk
@DataKindUK