

Data Visualisation

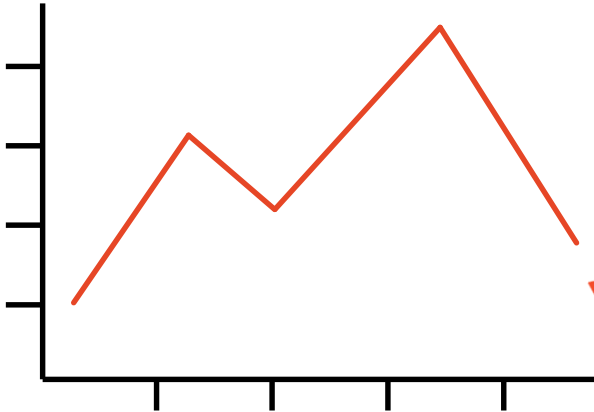
CIVL3704

28 March 2022

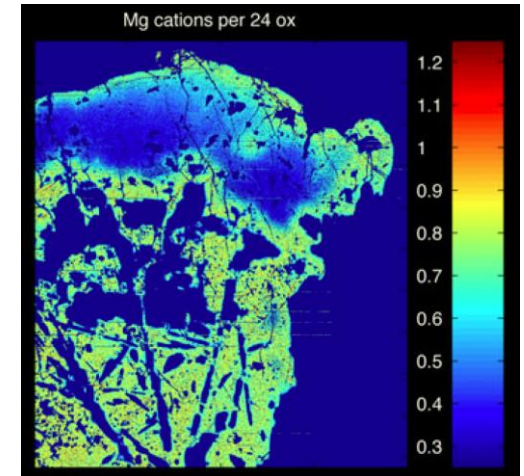
Presented by Nathaniel Butterworth
Slides courtesy of Kayla Maloney

<https://www.sydney.edu.au/research/facilities/sydney-informatics-hub.html>

What is data visualisation?

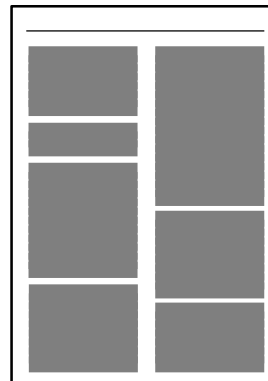


ID	X	Y
tl_001	3.7	0.42
tl_002	2.2	0.45
tl_003	6.5	0.48
tl_004	3.9	0.51

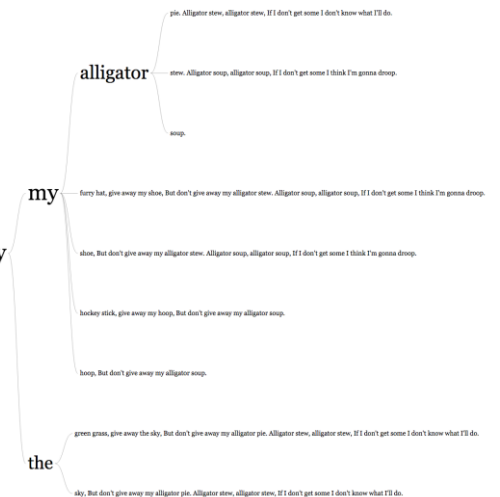


Maloney et al. (2011)

[doi:10.1111/j.1525-1314.2011.00934.x](https://doi.org/10.1111/j.1525-1314.2011.00934.x)

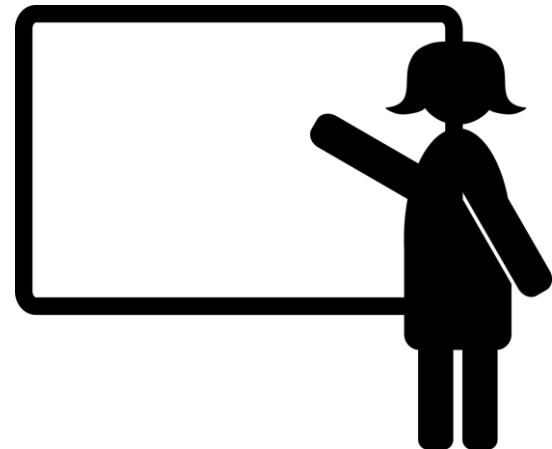
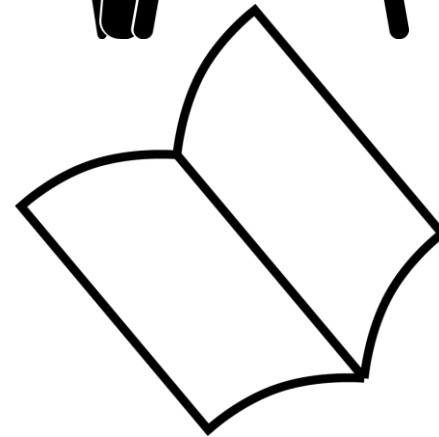


Give away



Who is it for?

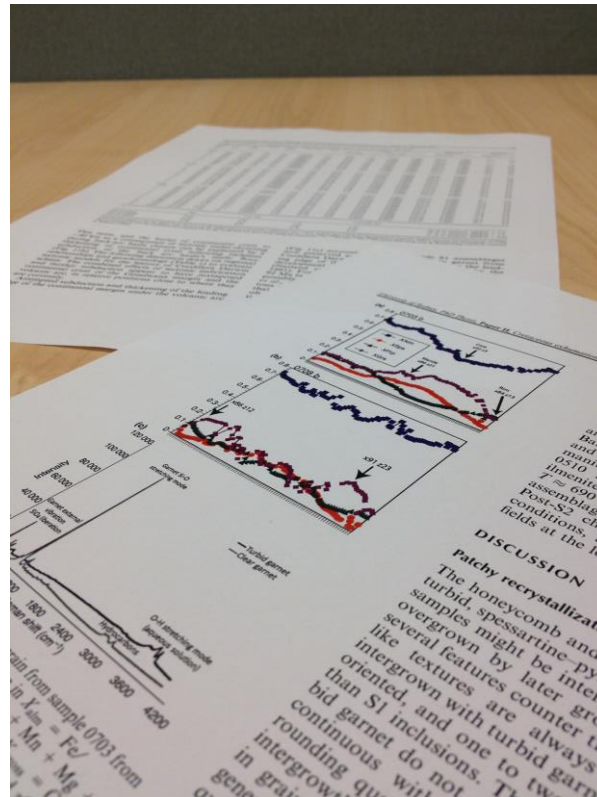
- Yourself
 - As part of the analysis of your data
- Others in your field
 - At conferences or as figures in papers
- Non-specialists
 - As part of public outreach
 - As part of reporting requirements



Presentation medium

How are you going to present your visualisation?

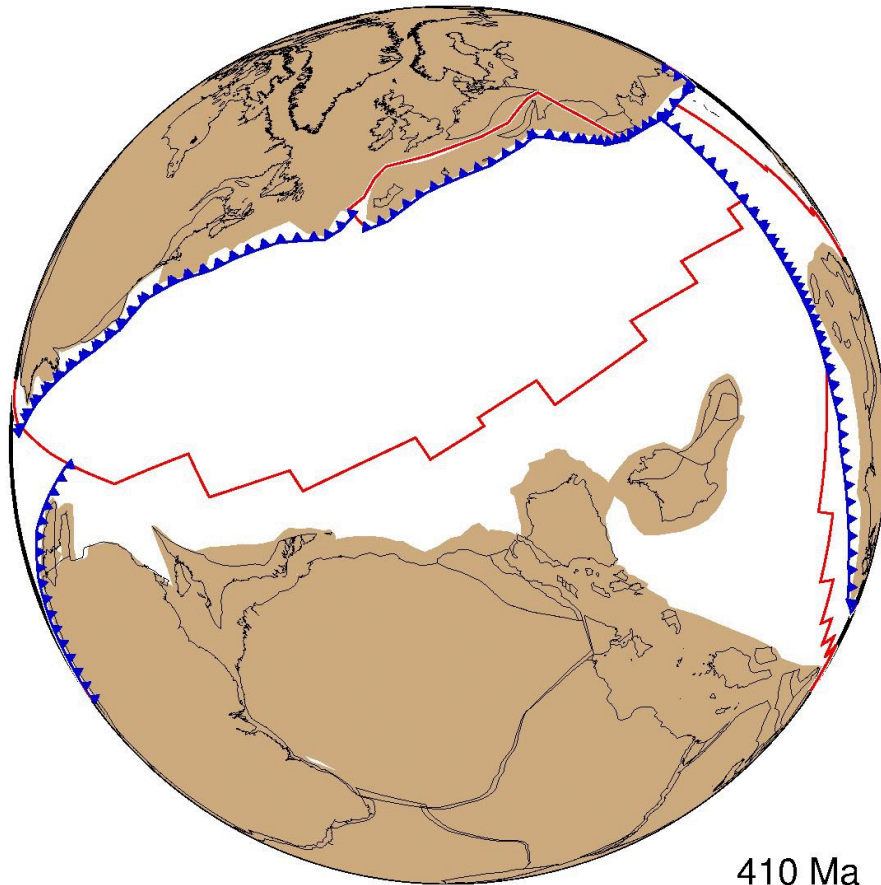
- Figure or table in a paper or poster



Presentation medium

How are you going to present your visualisation?

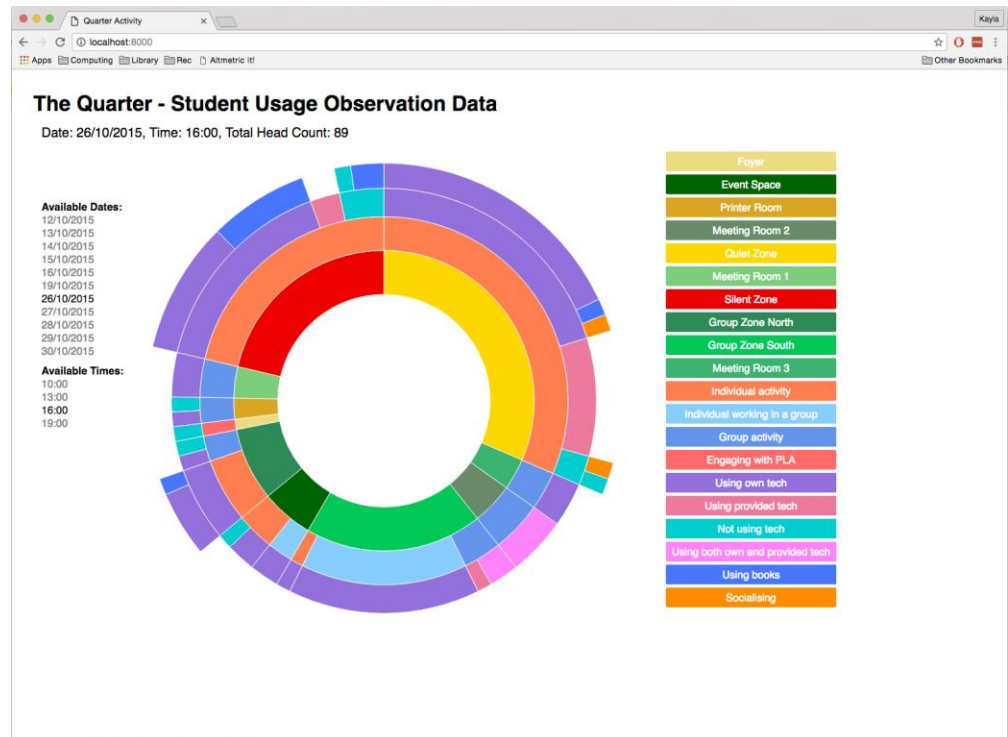
- Figure or table in a paper or poster
- On a slide in a talk








Presentation medium

How are you going to present your visualisation?

- Figure or table in a paper or poster
- On a slide in a talk
- On a website



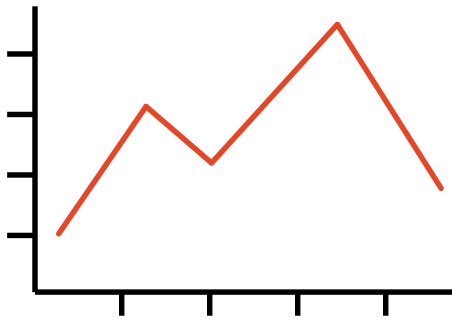
Principles of Data Visualisation

- Focus on the primary message of the visualisation 
- Keep it as simple and uncluttered as possible 
- Provide all information necessary to understand the visualisation 
- Make comparisons easy 
- Avoid any misleading presentation 

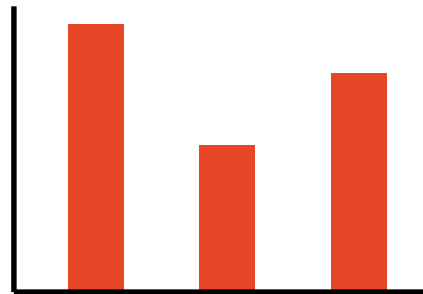
Focus on the primary message

Pick a figure type appropriate to your data and your message

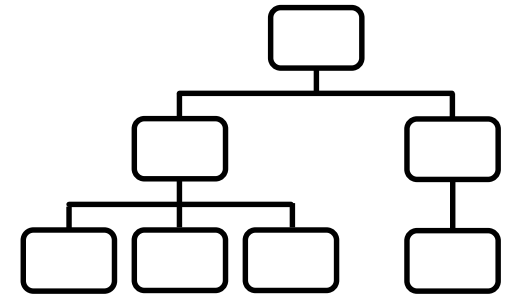
Numerical



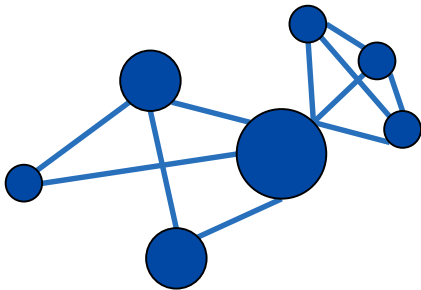
Categorical



Hierarchical



Network



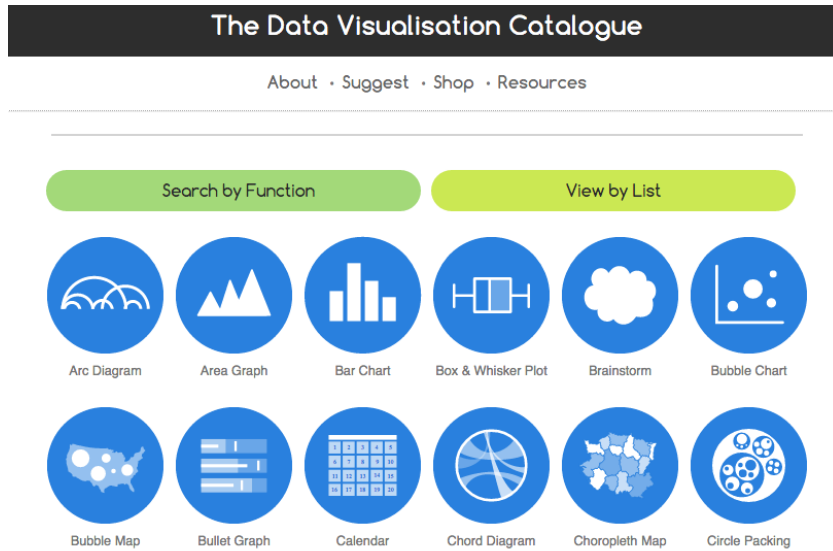
Geographic/geospatial



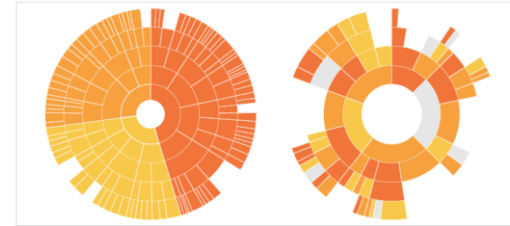
Textual



Focus on the primary message



Sunburst Diagram



Description

As known as Sunburst Chart, Ring Chart, Multi-level Pie Chart, Belt Chart, Radial Treemap.

This type of visualisation shows hierarchy through a series of rings, that are sliced for each category node. Each ring corresponds to a level in the hierarchy, with the central circle representing the root node and the hierarchy moving outwards from it.

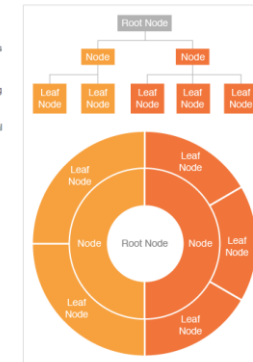
Rings are sliced up and divided based on their hierarchical relationship to the parent slice. The angle of each slice is either divided equally under its parent node or can be made proportional to a value.

Colour can be used to highlight hierarchical groupings or specific categories.

Functions

Hierarchy Part-to-a-whole

Anatomy



The Data Visualisation Catalogue
can be found at:

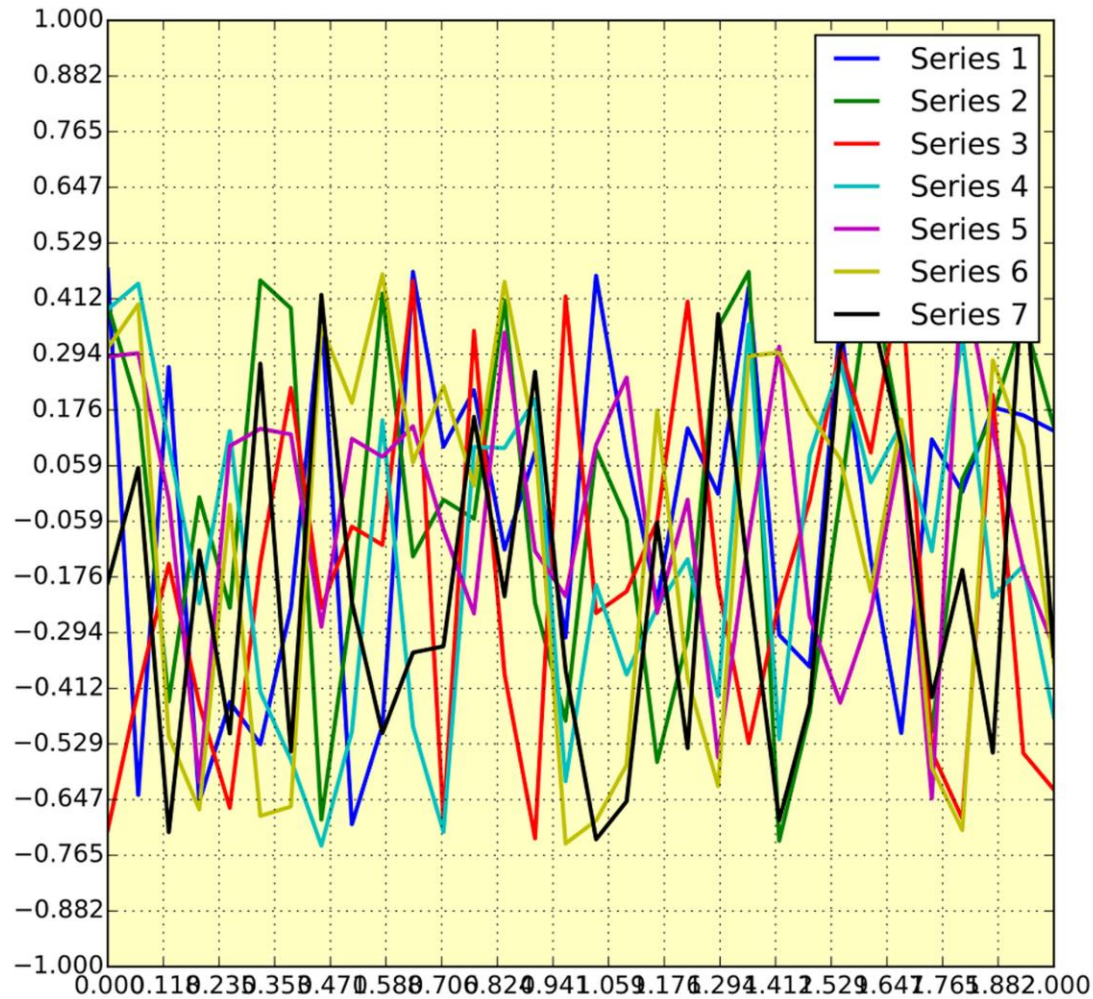
<http://www.datavizcatalogue.com/>

Or

<https://datavizproject.com/>

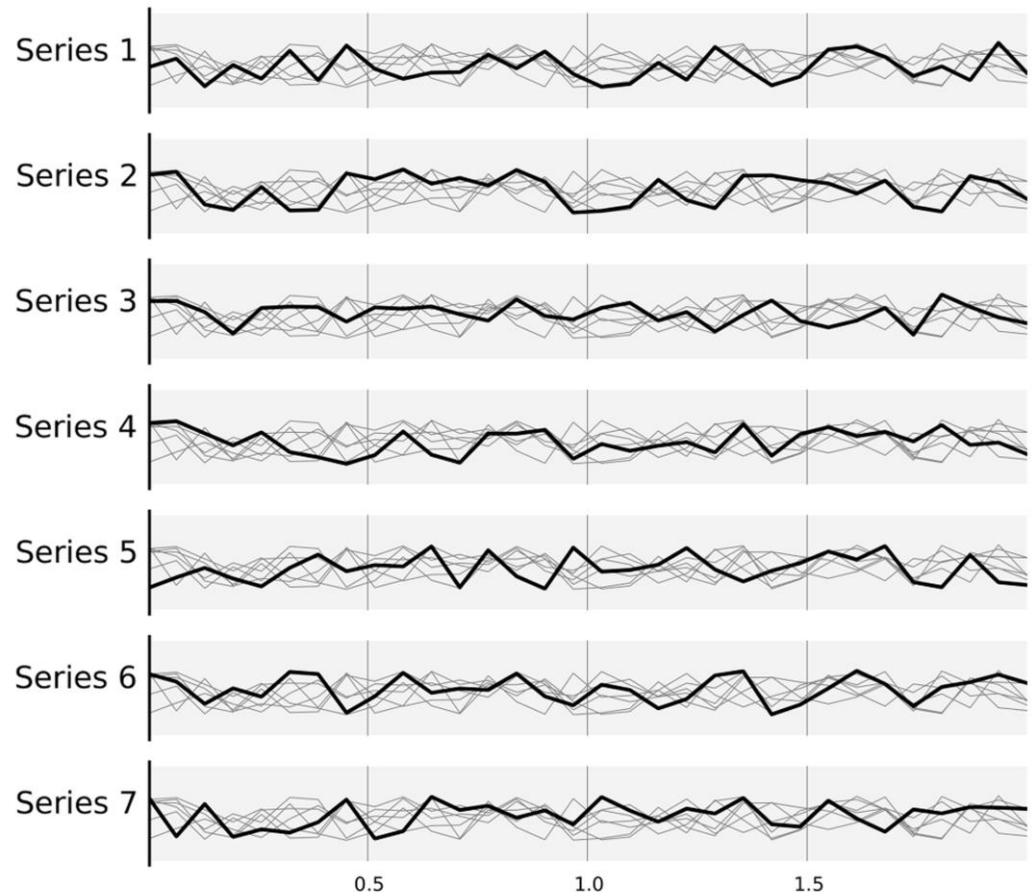
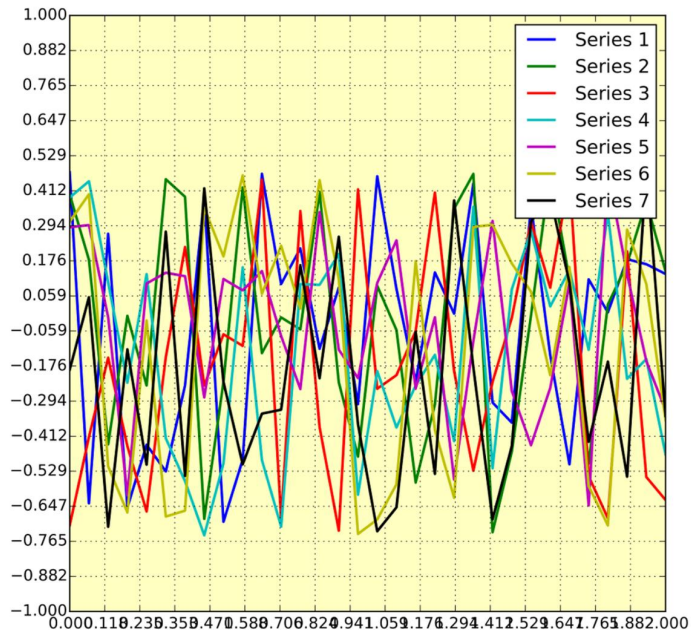
Keep it simple and uncluttered

What's wrong with this graph?



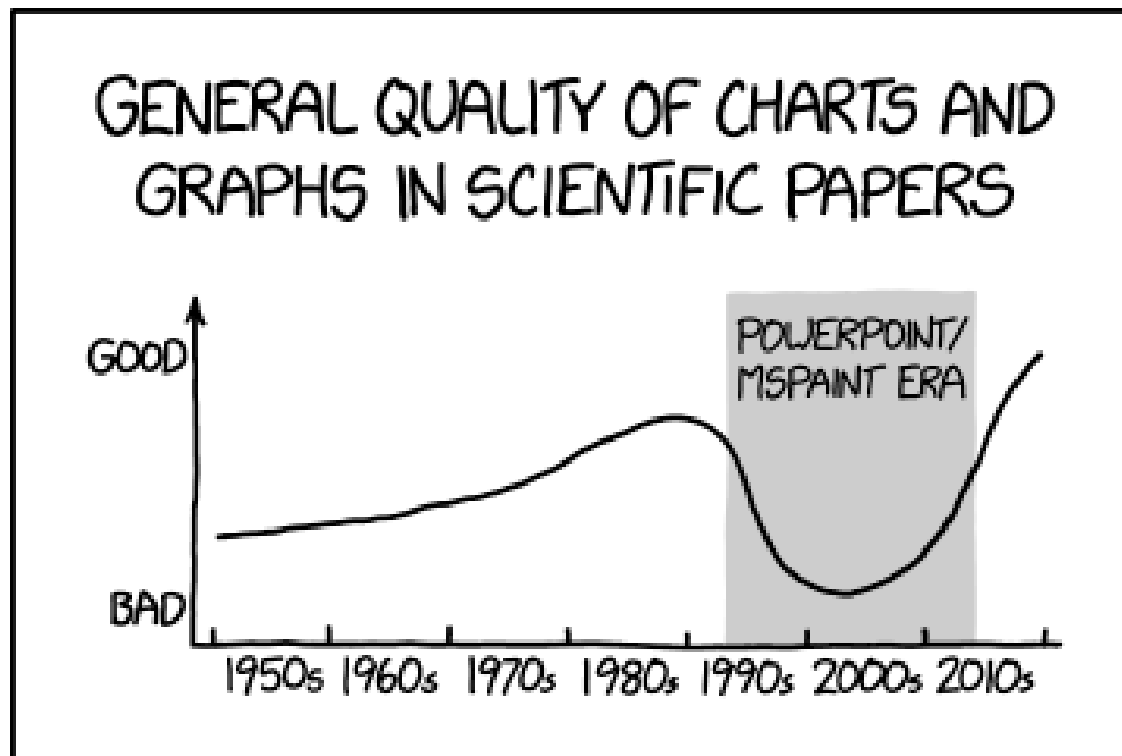
Keep it simple and uncluttered

Consider your message when simplifying and tidying.



Keep it simple and uncluttered

Plotting program defaults

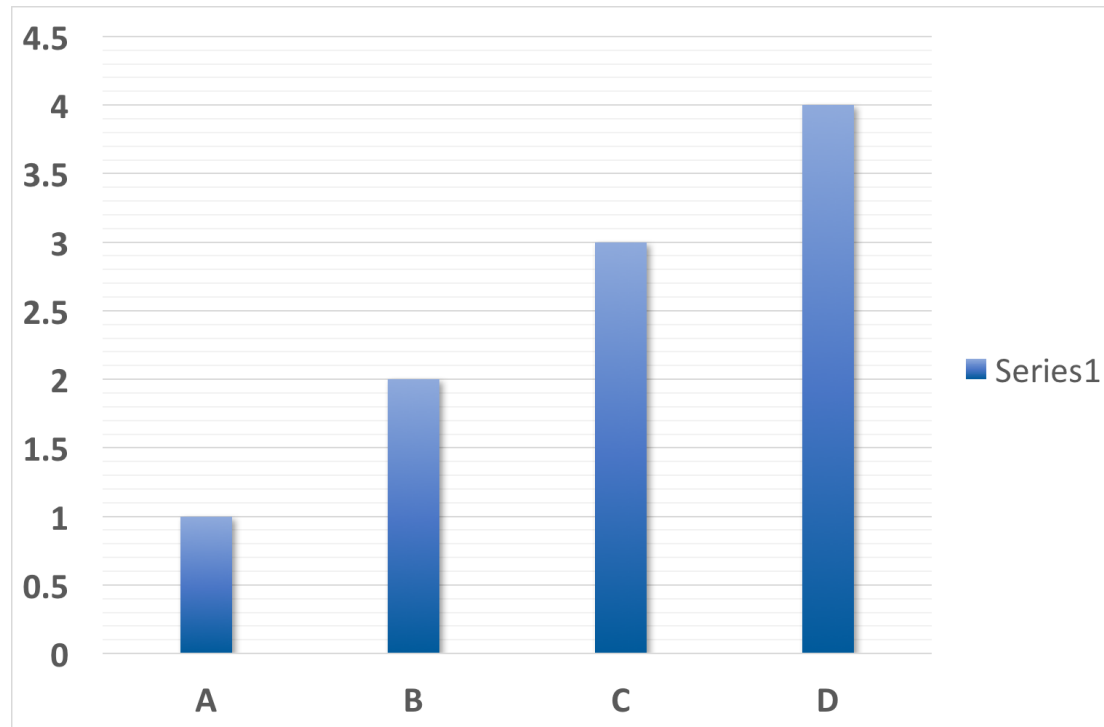


[Scientific Paper Graph Quality](#) by Randall Munroe [xkcd.com](#) [CC BY-NC 2.5]

Keep it simple and uncluttered

Plotting program defaults

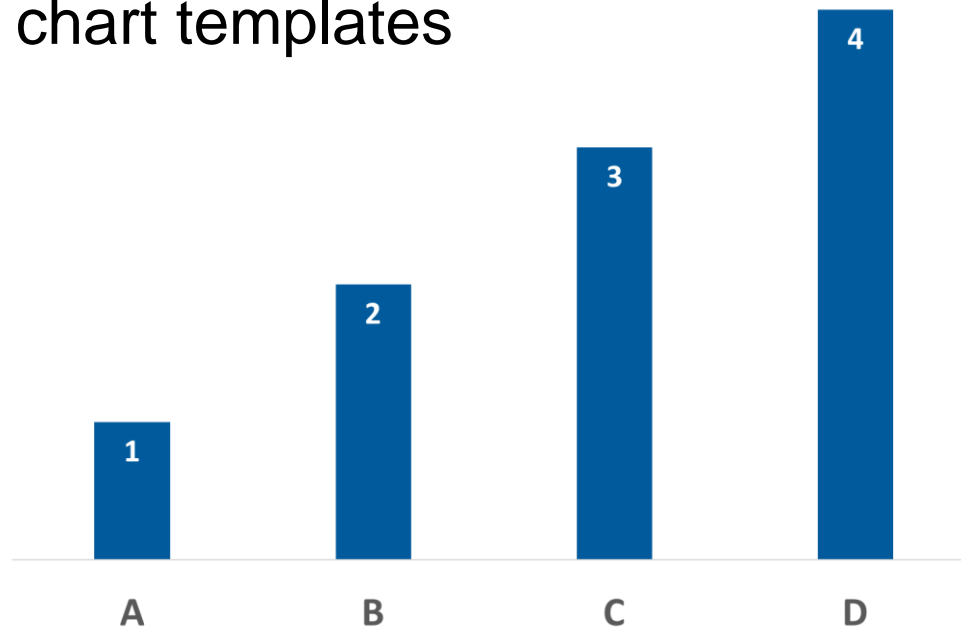
- Gridlines, gradients, shadows, excessive annotations add clutter



Keep it simple and uncluttered

Plotting program defaults

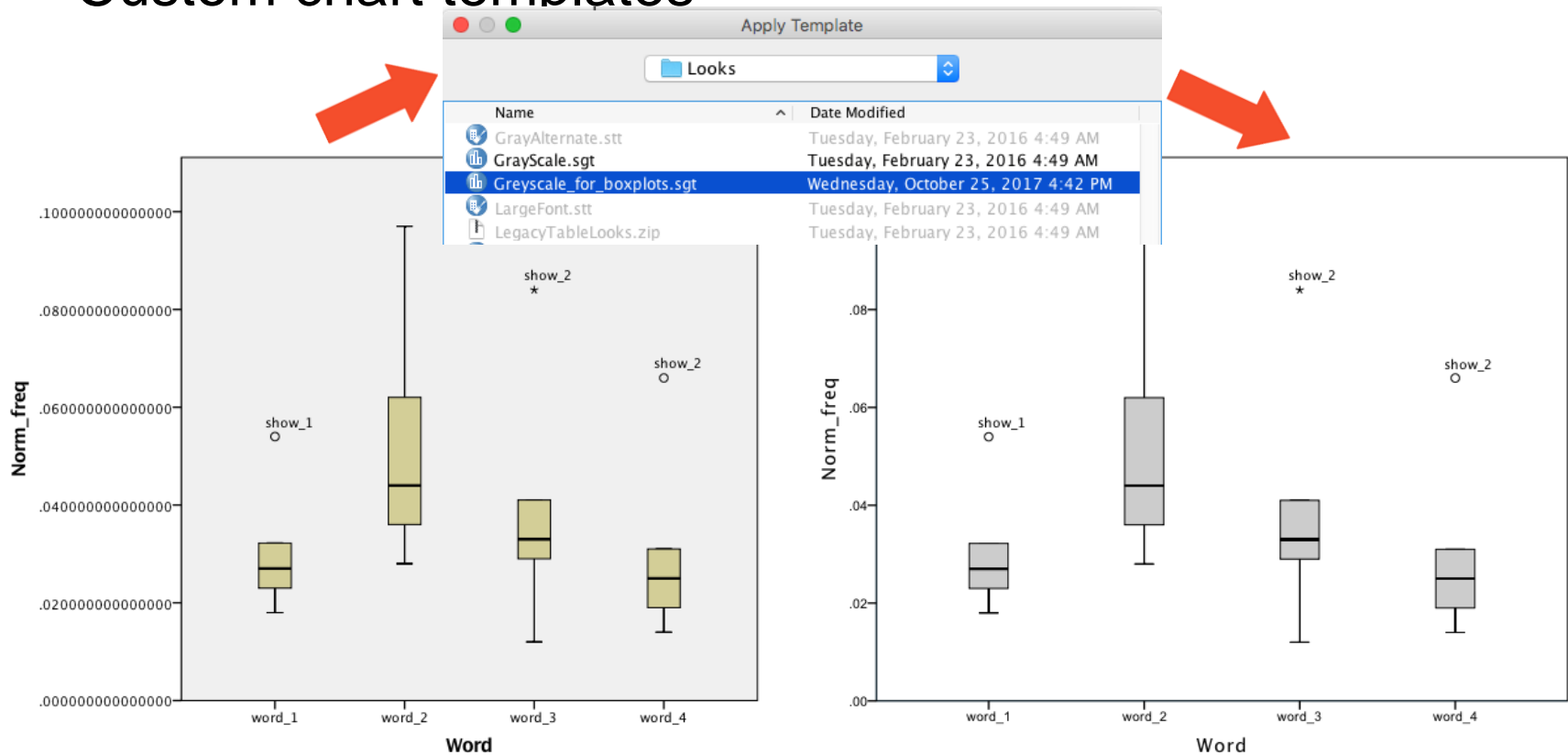
- Gridlines, gradients, shadows, excessive annotations add clutter
- Direct labelling useful for simple plots
- Custom chart templates



Keep it simple and uncluttered

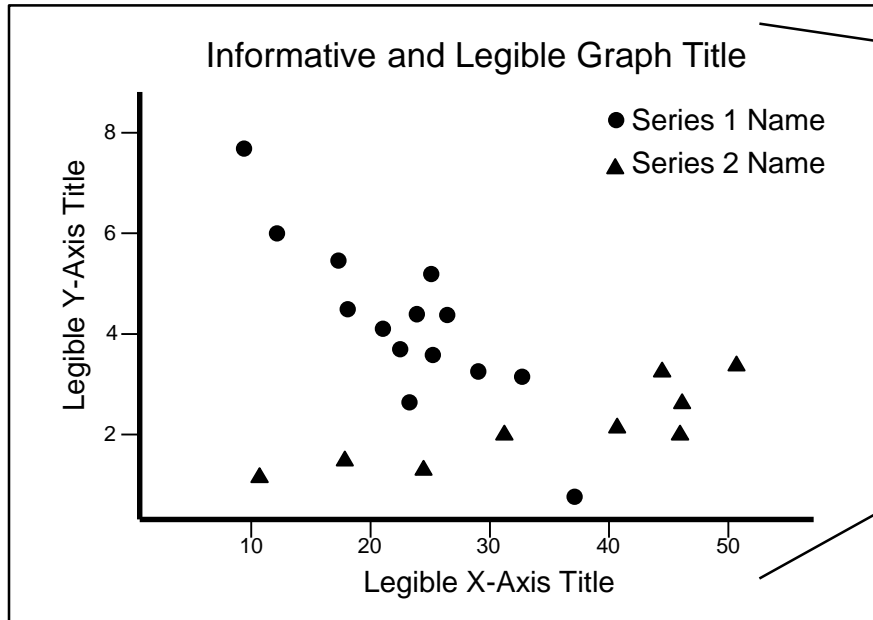
Plotting program defaults

– Custom chart templates

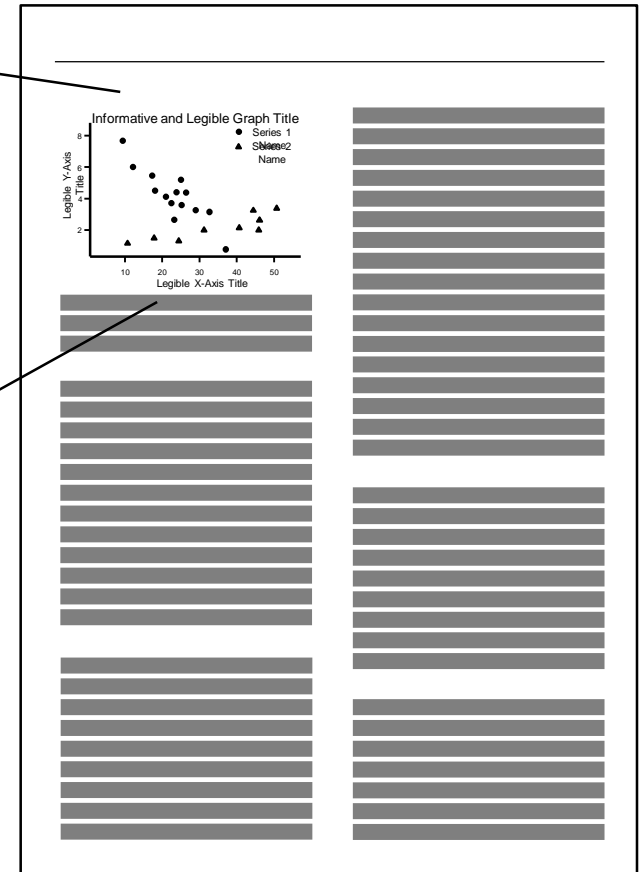


Include all necessary information

Ensure all parts are easily legible



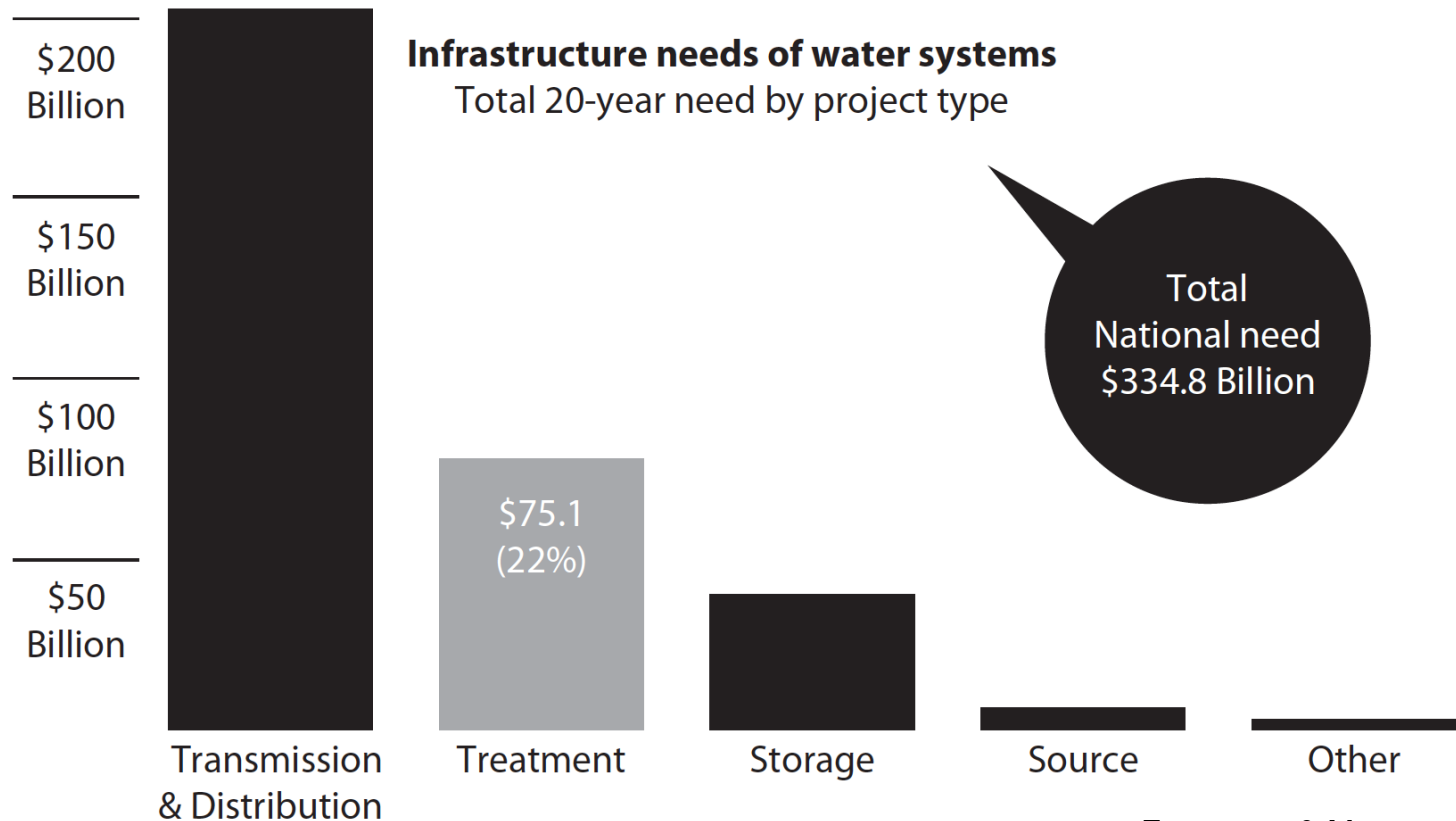
- Keep font sizes large-ish
- Choose symbols that are easy to discriminate between
- Draft your visual at the final size



Include all necessary information

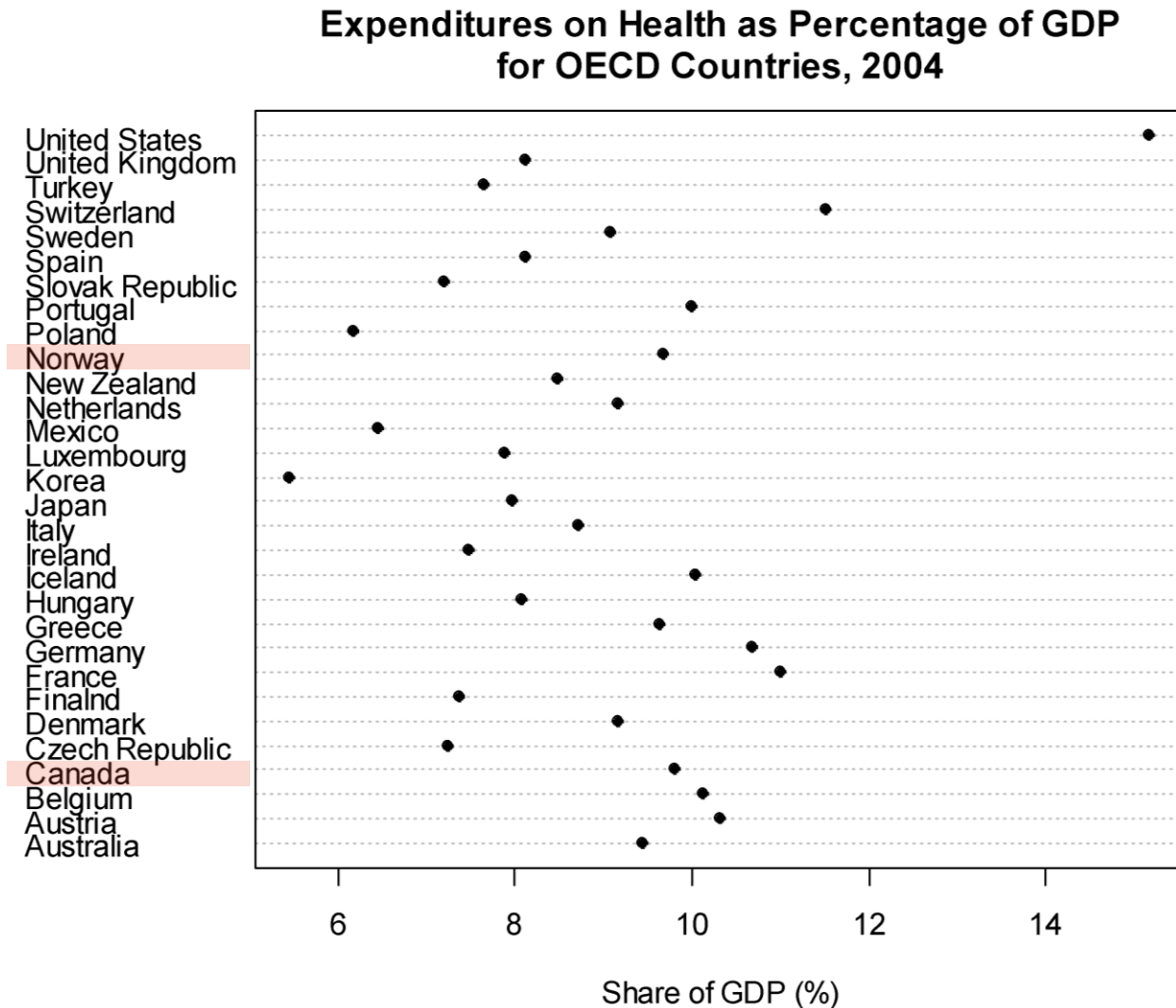
Emphasise important details

- Use sparingly! You don't want to re-clutter your visual



Evergreen & Metzner
(2013)

Make comparisons easy

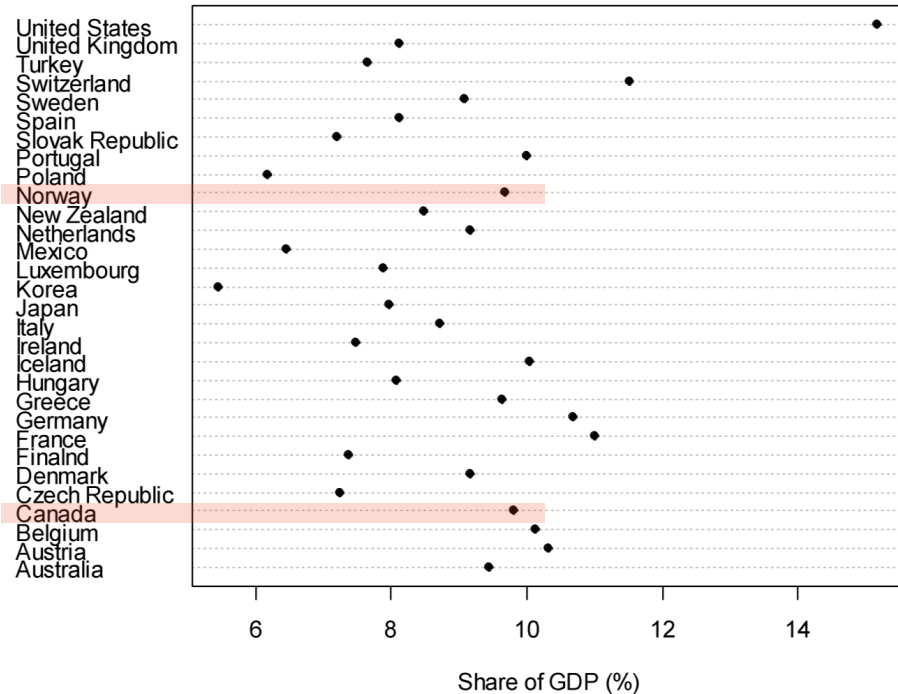


Michael Friendly
(2001)
[CC-BY-NC 3.0]

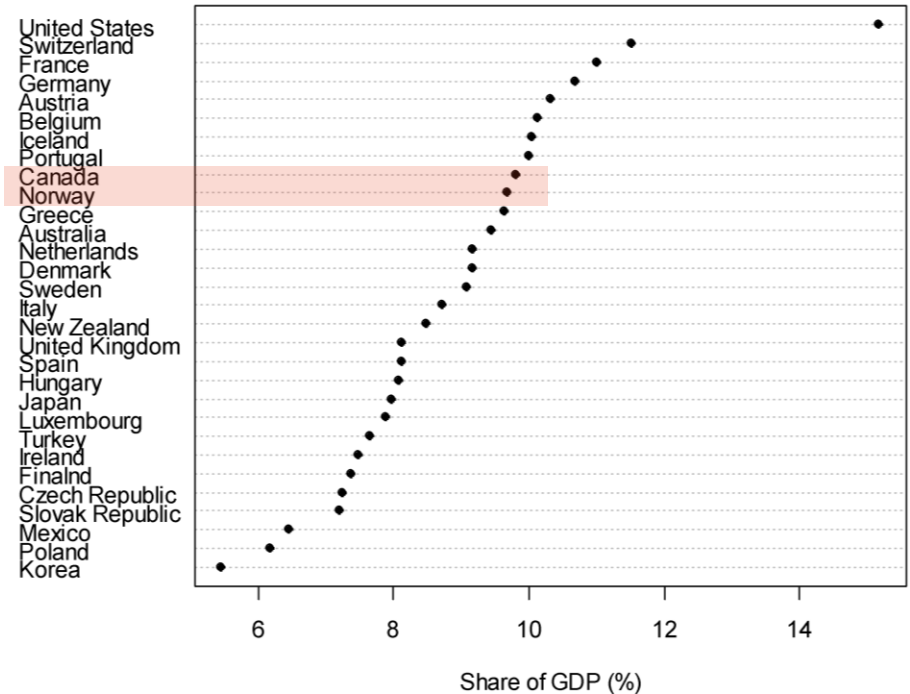
Make comparisons easy

Order data in a meaningful way

Expenditures on Health as Percentage of GDP
for OECD Countries, 2004



Expenditures on Health as Percentage of GDP
for OECD Countries, 2004



Make comparisons easy

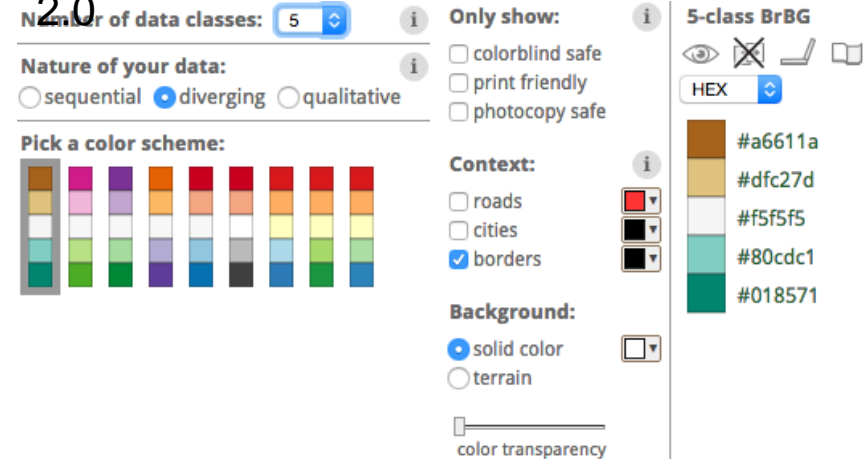
Colour resources

- ColorBrewer 2.0 – helps you create colour palettes and notes whether they are colour-blind, print and photocopy friendly
colorbrewer2.org

- Color Oracle – a colour-blindness simulator that applies a filter to your screen to demonstrate how it would be perceived by people with several types of colour impaired vision
colororacle.org

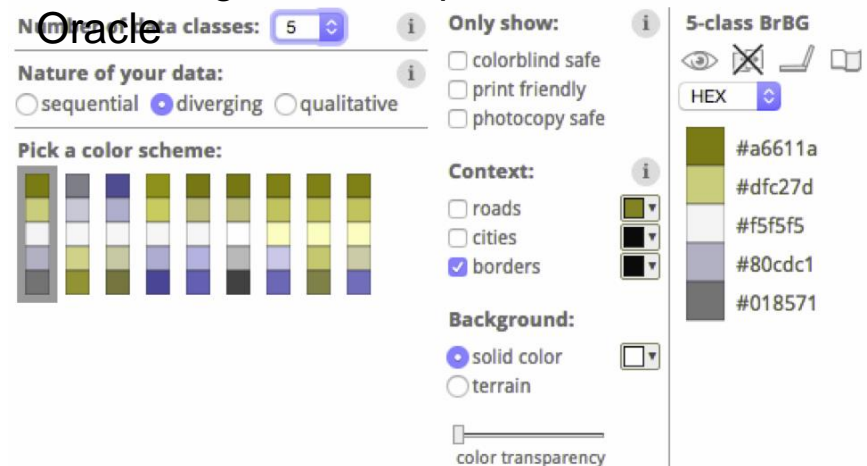
Picking a colour scale using ColorBrewer

2.0



Simulating deuteranopia with Color

Oracle

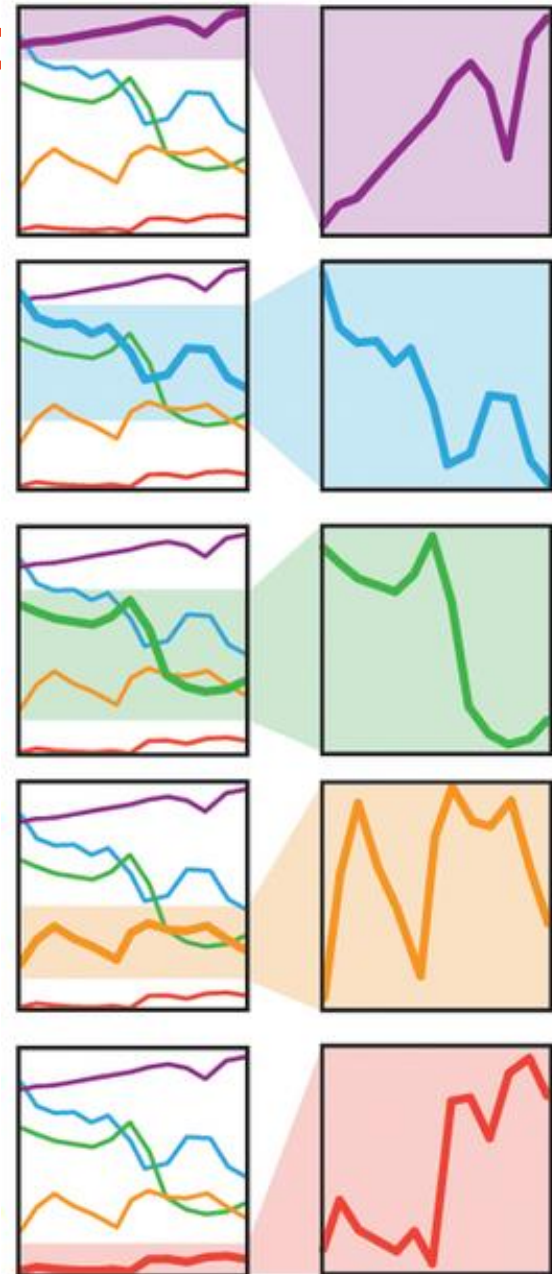


Avoid misleading presentation

Ensure your scale is appropriate

- Different scales can wildly change how your data is perceived
- Including zero shows the overall trend
- Clipping the y-scale around the data highlights the actual data values and the variation within your data

****Always include zero on bar or area plots****



McInerny
&
Krzywinski
(2015)

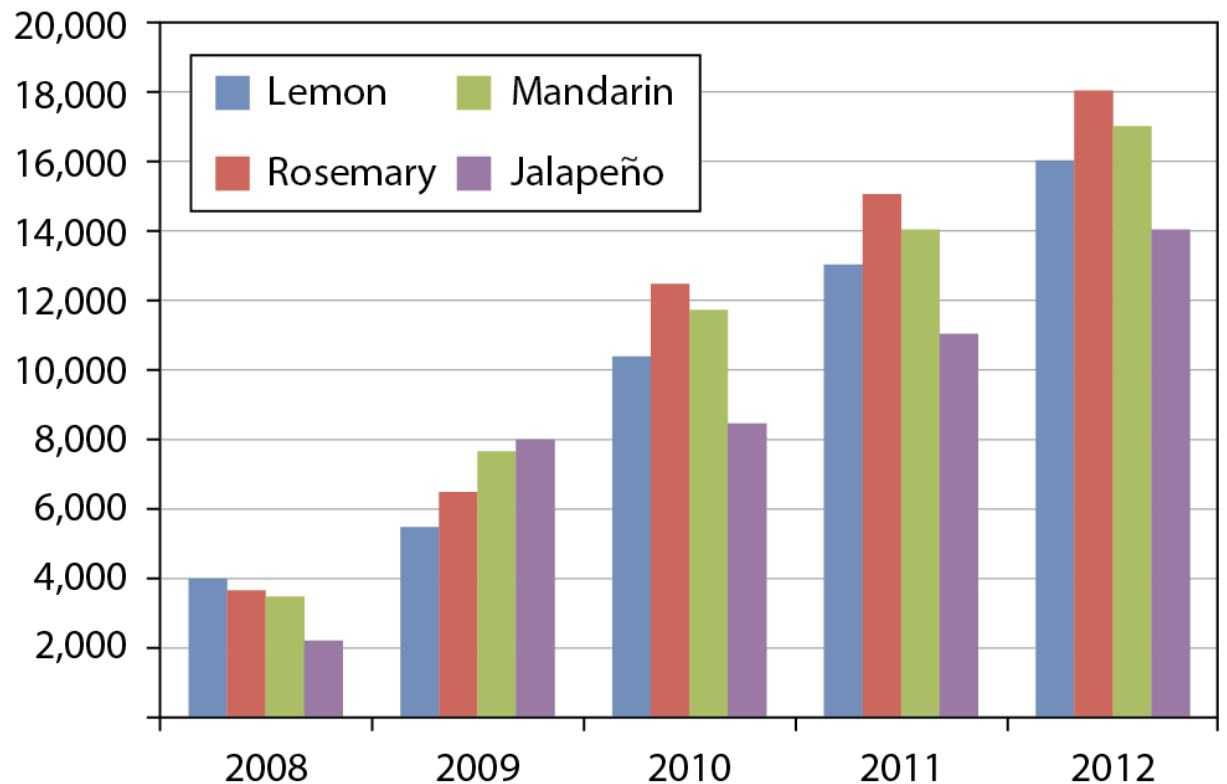
Avoid misleading presentation

Consider your use of colour:

- Pre-existing colour associations

Importing infusion scents

Quantities rise as programs help increase production



Evergreen &
Metzner (2013)

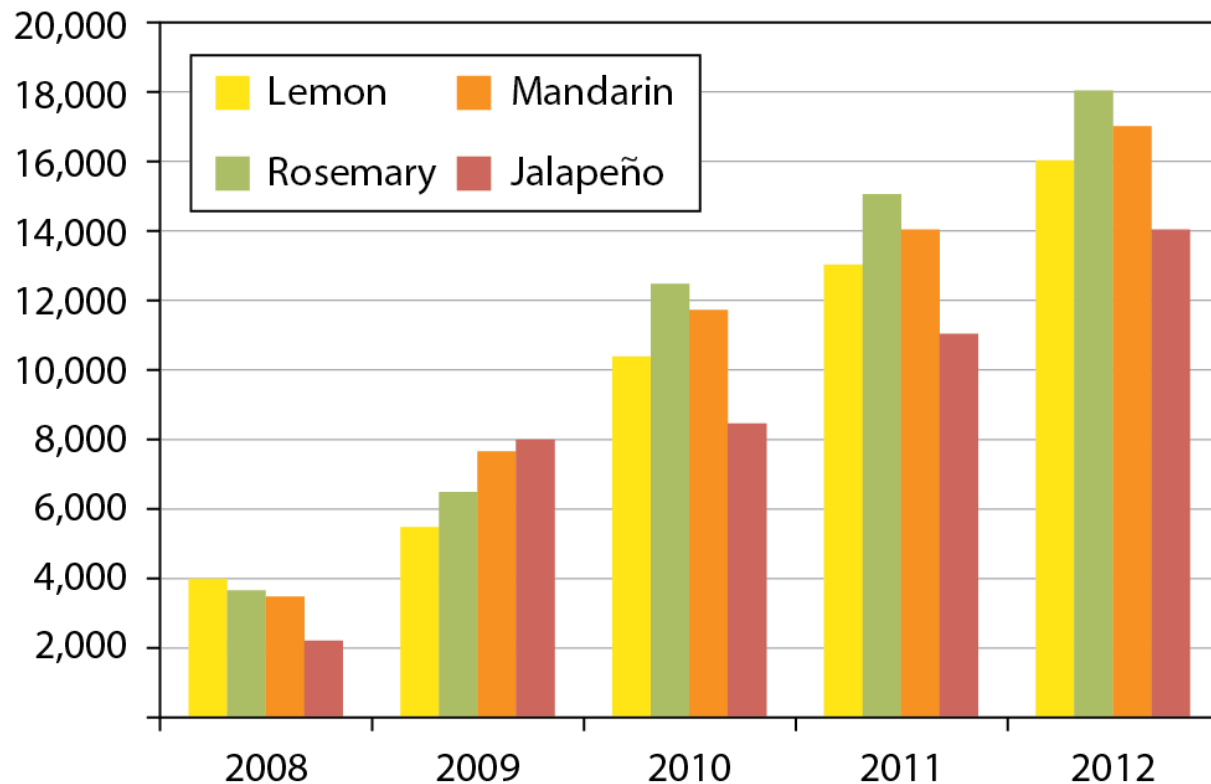
Avoid misleading presentation

Consider your use of colour:

- Pre-existing colour associations

Importing infusion scents

Quantities rise as programs help increase production

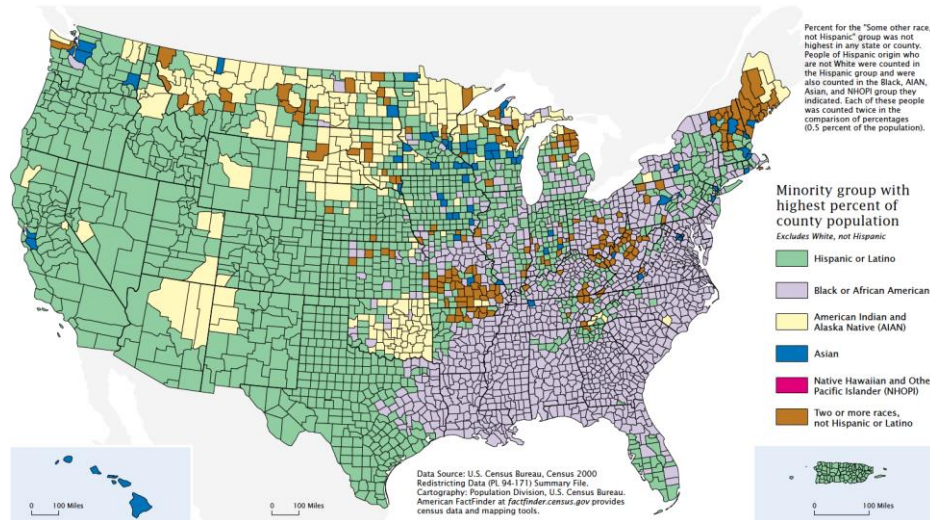


Recoloured from
Evergreen & Metzner
(2013)

Avoid misleading presentation

Consider your use of colour:

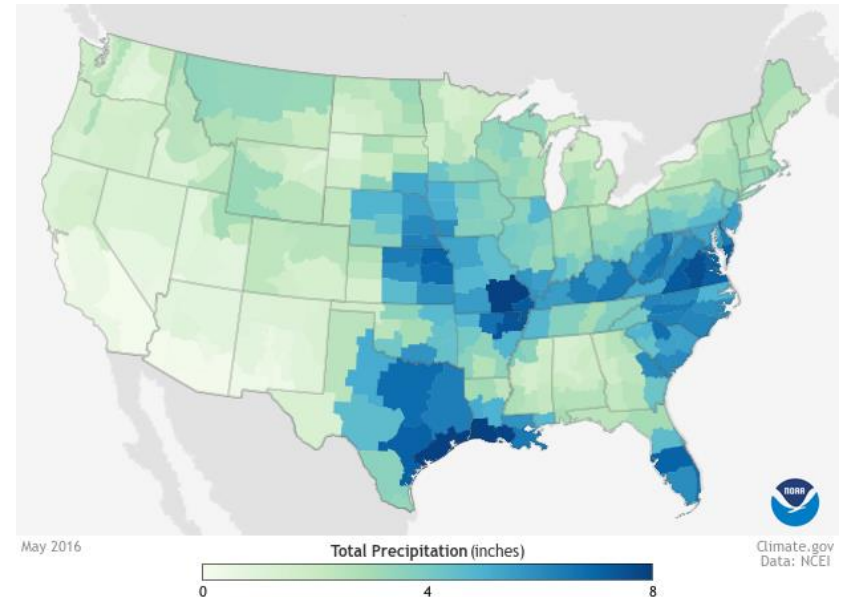
- Does it imply nominal or sequential data?



US Census
Bureau, [Census](https://www.census.gov)
[2000](https://www.census.gov)

Nominal

- Unordered
- Represented by different, distinct colours

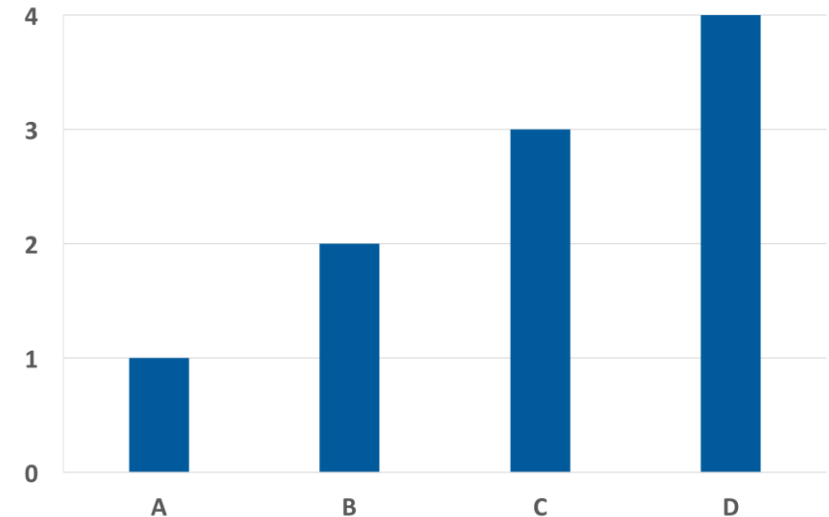
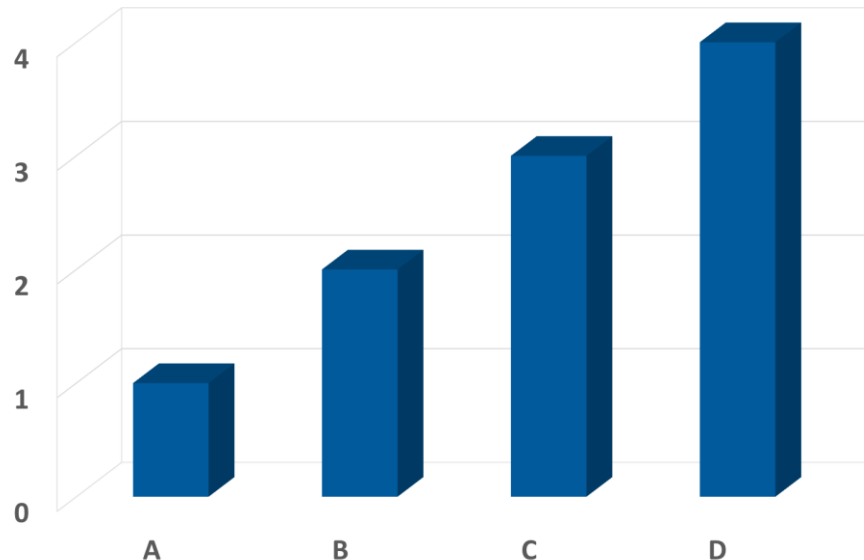


Sequential

- Ordered
- Represent by a gradient, generally in hue or intensity

Avoid misleading presentation

Do not use pseudo-3D effects



- 3D effects for non-3D data distort your data
- These bar plots were made using the exact same data!
- Could you tell that the data values are 1, 2, 3, 4 for the pseudo 3D version?

So, to create an effective data visualisation...

Before you start, consider:

- Who is it for?
- How will it be presented?
- What do you need to say?

Remember the Principles while making your visualization:

- Focus on the primary message of the visualisation
- Keep it as simple and uncluttered as possible
- Provide all information necessary to understand the visualisation
- Make comparisons easy
- Avoid any misleading presentation

Then get a pal to look it over.

Python example!

- <https://github.com/natbutter/SydneyTolls>

Where next?

Data Analysis and Visualisation Guide

<http://libguides.library.usyd.edu.au/data-analysis-visualisation>

Data Analysis and Visualisation: Visualisation Tools

[Home](#)[Creating a Visualisation](#)[Visualisation Tips](#)[Analysis and Visualisation Toolkit](#)

Visualisation tools available at the University of Sydney

The University of Sydney provides licences to some commercial software packages for staff and students. Please visit the [list of software](#) available through the university to see what software you are eligible to access, and for information on how to obtain access to the available packages.

Tools specific to data visualisation that are available at the University are listed and described below. Keep in mind that analysis and visualisation are often overlapping activities, so be sure to check both the analysis and visualisation sections to ensure that you don't miss the ideal tool for your data!

Adobe Illustrator – A vector graphics editor that can be extremely useful for creating figures or touching up those made in other programs, including adding annotations. Illustrator can also be useful in creating outreach visuals, such as infographics. Illustrator is part of the Adobe Creative Suite. [Get Illustrator from ICT.](#)

Microsoft Visio – A diagramming and vector graphics application that can help you simplify complexity with dynamic, data-driven visuals. Visio shapes can be linked to multiple data sources including Microsoft Excel. [Get Visio from ICT.](#)

Freely available visualisation tools

In addition to commercial software, a host of open-source and/or freely available tools exist for data visualisation. We have collected a short list of some of the more widely used, or easy to use tools that are available. Keep in mind that analysis and visualisation are often overlapping activities, so be sure to check both the analysis and visualisation sections to ensure that you don't miss the ideal tool for your data!

Interactive Visualisation

Tableau Public – A tool that allows you to create and share interactive charts and graphs, maps, live dashboards and applications. All work will be saved to your Tableau Public profile, visible online.

D3.js – A JavaScript library for manipulating documents based on data. D3 allows you to create interactive web-based visualisations of your data, and helps you bring data to life using HTML, SVG, and CSS.

Where next?

University training in specific tools, including Excel, MATLAB, R, Python

Training schedule available at:

<https://informatics.sydney.edu.au/services/training/>

Hacky Hour – A regular drop-in meetup for any questions about data, coding, analyses, visualisation, modelling, and research problems.

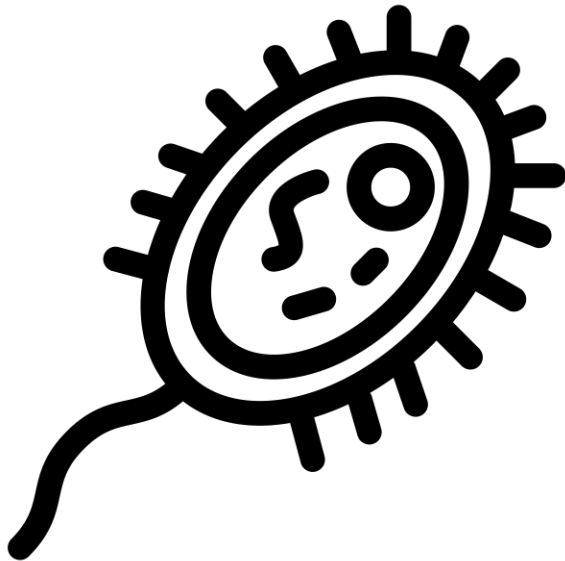
Consultations – come and talk to someone about your data, and what you want to visualise!

Email us: researchdatasupport@sydney.edu.au

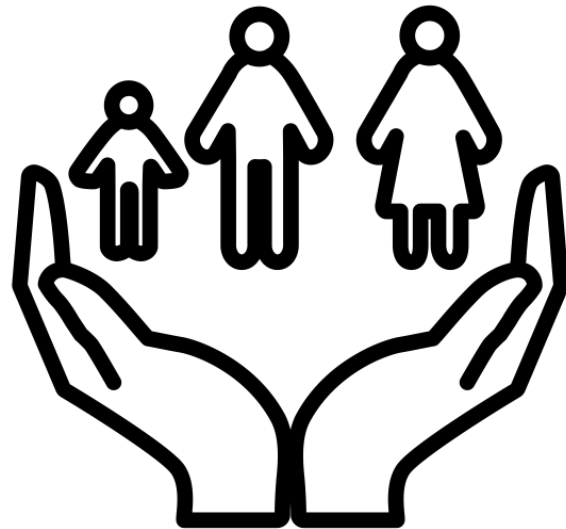
Other Resources

The Noun Project and Flaticon

- Libraries of icons free to use under Creative Commons licences



Bacteria by Adnen Kadri via the Noun Project [CC-BY 3.0 US]

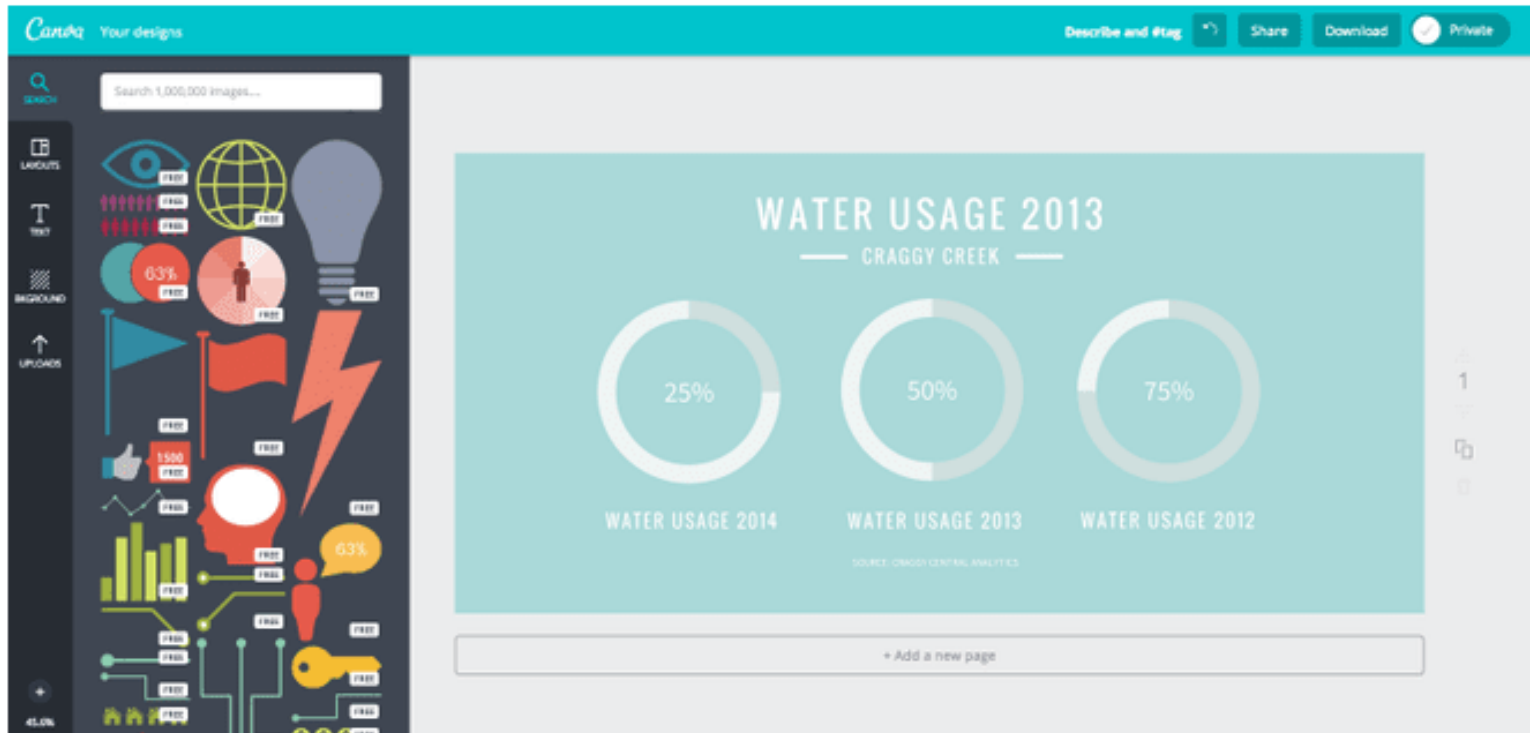


family insurance by H Alberto Gongora via the Noun Project [CC-BY 3.0]

Other Resources

Canva

- Browser-based graphic design tool
- Includes a drag-and-drop infographic maker

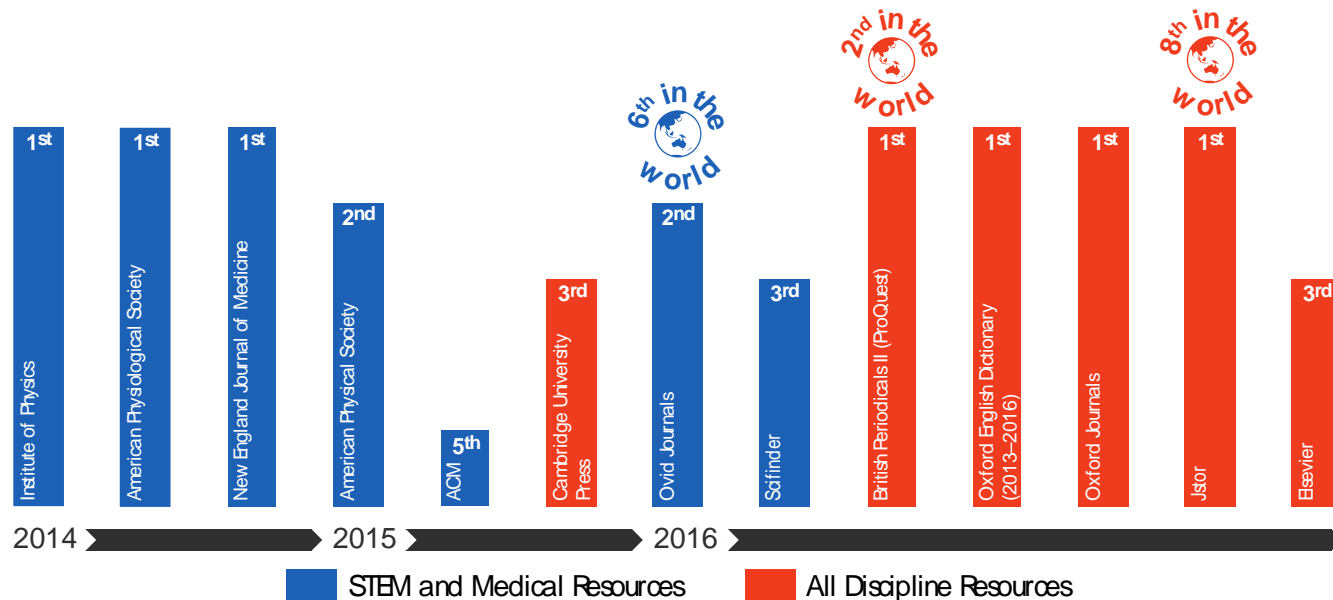


Other Resources

Adobe Illustrator

- Tidy up visualisations created with other tools
- Make icons, infographics from scratch

The University of Sydney's Usage Rank in Australia and New Zealand for Online Scholarly Resources



Over **11 million** uses of the more than **1 million** online scholarly resources in our collection in 2016

Other Resources

Online training videos across a wealth of topics and tools, available to anyone with a UniKey

Access to LinkedIn Learning (formerly Lynda.com):

1. Go to <https://www.linkedin.com/learning>
2. Sign In with SSO
3. Login using your UniKey and password

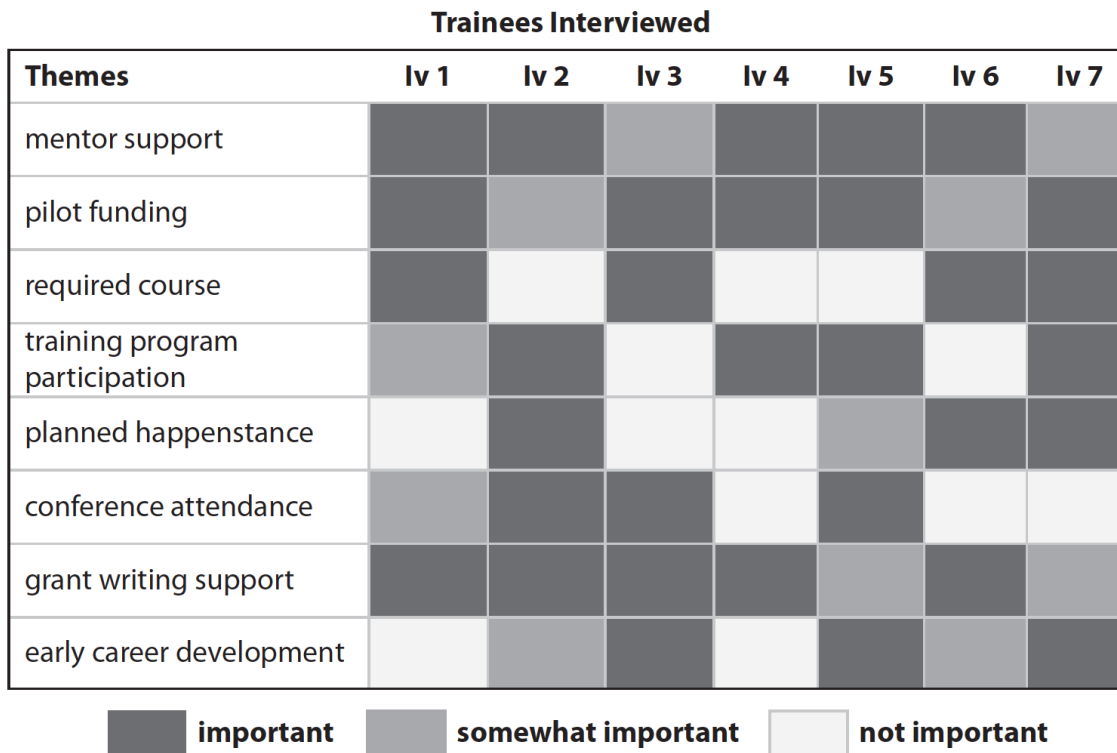
References

- Ahrens et al. (2012) Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature*, 485(7399), 471-477. [doi:10.1038/nature11057](https://doi.org/10.1038/nature11057)
- Evergreen & Metzner (2013) Design principles for data visualisation in evaluation. In Azzam & Evergreen (Eds.) *Data visualization, part 2. New Directions for Evaluation*, 140, 5-20. [doi:10.1002/ev.20071](https://doi.org/10.1002/ev.20071)
- Friendly (2001) Gallery of Data Visualization, Electronic document, accessed June 2016. <http://www.datavis.ca/gallery/>
- Henderson & Segal (2013). Visualizing qualitative data in evaluation research. In Azzam & Evergreen (Eds.) *Data visualization, part 1. New Directions for Evaluation*, 139, 53-71. [doi:10.1002/ev.20067](https://doi.org/10.1002/ev.20067)
- Maloney et al. (2011) Crustal growth during back-arc closure: Cretaceous exhumation history of Cordillera Darwin, southern Patagonia. *Journal of Metamorphic Geology*, 29(6), 649-672. [doi:10.1111/j.1525-1314.2011.00934.x](https://doi.org/10.1111/j.1525-1314.2011.00934.x)
- McInerny & Krzywinski (2015) Points of View: Unentangling complex plots. *Nature Methods*, 12(7), 591. [doi:10.1038/nmeth.3451](https://doi.org/10.1038/nmeth.3451)
- National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Information, Electronic document, accessed July 2016. <https://www.climate.gov/maps-data>
- Reese et al. (2012) Living kidney donation: outcomes, ethics, and uncertainty. *The Lancet*, 385(9981), 2003-2013. [doi:10.1016/S0140-6736\(14\)62484-3](https://doi.org/10.1016/S0140-6736(14)62484-3)
- Rougier et al. (2014) Ten Simple Rules for Better Figures. *PLoS Computational Biology*, 10(9), e1003833. [doi:10.1371/journal.pcbi.1003833](https://doi.org/10.1371/journal.pcbi.1003833)
- Tripto et al. (2013) Mapping What They Know: Concept Maps as an Effective Tool for Assessing Students' Systems Thinking. *American Journal of Operations Research*, 3(1A), 245-258. [doi:10.4236/ajor.2013.31A022](https://doi.org/10.4236/ajor.2013.31A022)
- Tobacco and the Economy, Gale, Foreman, and Capehart, Agriculture Economic Report No. 789, September 2000.
- U.S. Census Bureau, Census 2000 Redistricting Data (PL 94-171) Summary File. Available at <http://www.census.gov/population/www/cen2000/atlas/index.html>

Include all necessary information

Make visualisations self-contained

Figure 3.5. Matrix Displaying the Level of Importance of Themes Uncovered in Interviews With Training Program Participants



Note: Iv# is the number of each trainee. Darker shades indicate increased importance of theme.

- Figure captions
- Informative labels and/or title
- Colours or symbols defined
- Abbreviations or acronyms defined
- Information on how to interpret visualisation provided if necessary

Henderson & Segal
(2013)