

Data Visualisation

Chapter 1: Design and Integrity

Dr James Baglin

How to use these slides

Viewing slides...

- Press 'f' enable fullscreen mode
- Press 'o' or 'Esc' to enable overview mode
- Pressing 'Esc' exits all of these modes.
- Hold down 'alt' and click on any element to zoom in. 'Alt' + click anywhere to zoom back out.
- Use the Search box (top right) to search keywords in presentation

Printing slides...

- Click here to open a [printable version of these slides](#).
- Right click and print from browser or save as PDF (e.g. Chrome)

Learning Objectives

The learning objectives of this chapter are as follows:

- Define data visualisation
- List and explain the different types of data and why data types are important to data visualisation
- Identify the various components of a basic data visualisation plot
- Explain the data visualisation design process including the following three stages:
 - Identifying a targeted audience and a data visualisation design objective
 - Focusing, justifying and choosing methods
 - Construction and evaluation

Learning Objectives Cont.

- Apply the Trifecta Check-up to guide the critique of a data visualisation
- Apply the Data Visualisation Check-list to produce publication quality data visualisations
- Discuss ethical principles and data integrity as it relates to the practice of data visualisation
- Locate and identify reliable and reputable sources of data for visualisation

What is data visualisation?

Andy Kirk (2012) defined data visualisation as follows:

“the representation and presentation of data that exploits our visual perception abilities in order to amplify cognition” (p.17)

- Representation
- Presentation
- Data
- Visual perception
- Amplify cognition

Why do we visualise data?

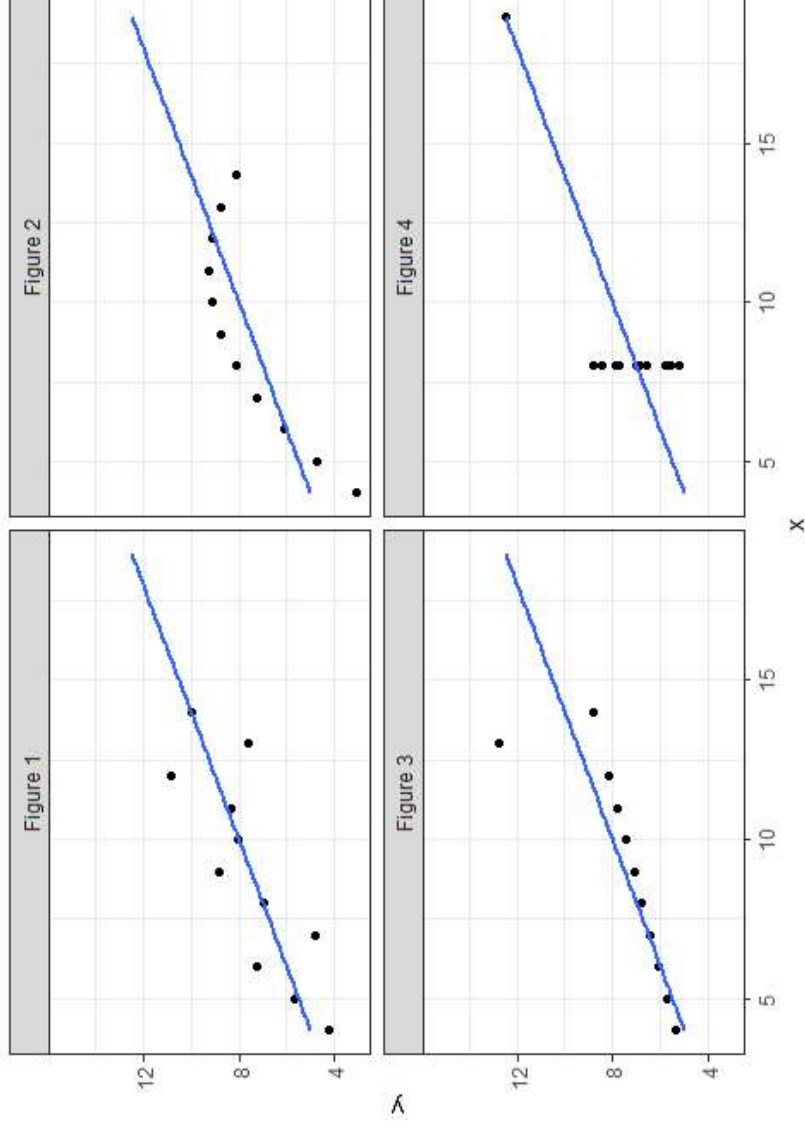
The greatest value of a picture is when it forces us to notice what we never expected to see - John W Tukey, Exploratory Data Analysis, (1977)



By Source, [Fair use](#), [Link](#)

Anscombe's quartet

- All scatter plots have approximately the same correlation...



Why do we visualise data? Cont.

- Why do we visualise?
 - Exploration - identifying interesting and important features (Buja et al. 2009)
 - Assist in communicating key data insights to our targetted audience

Turning Tables

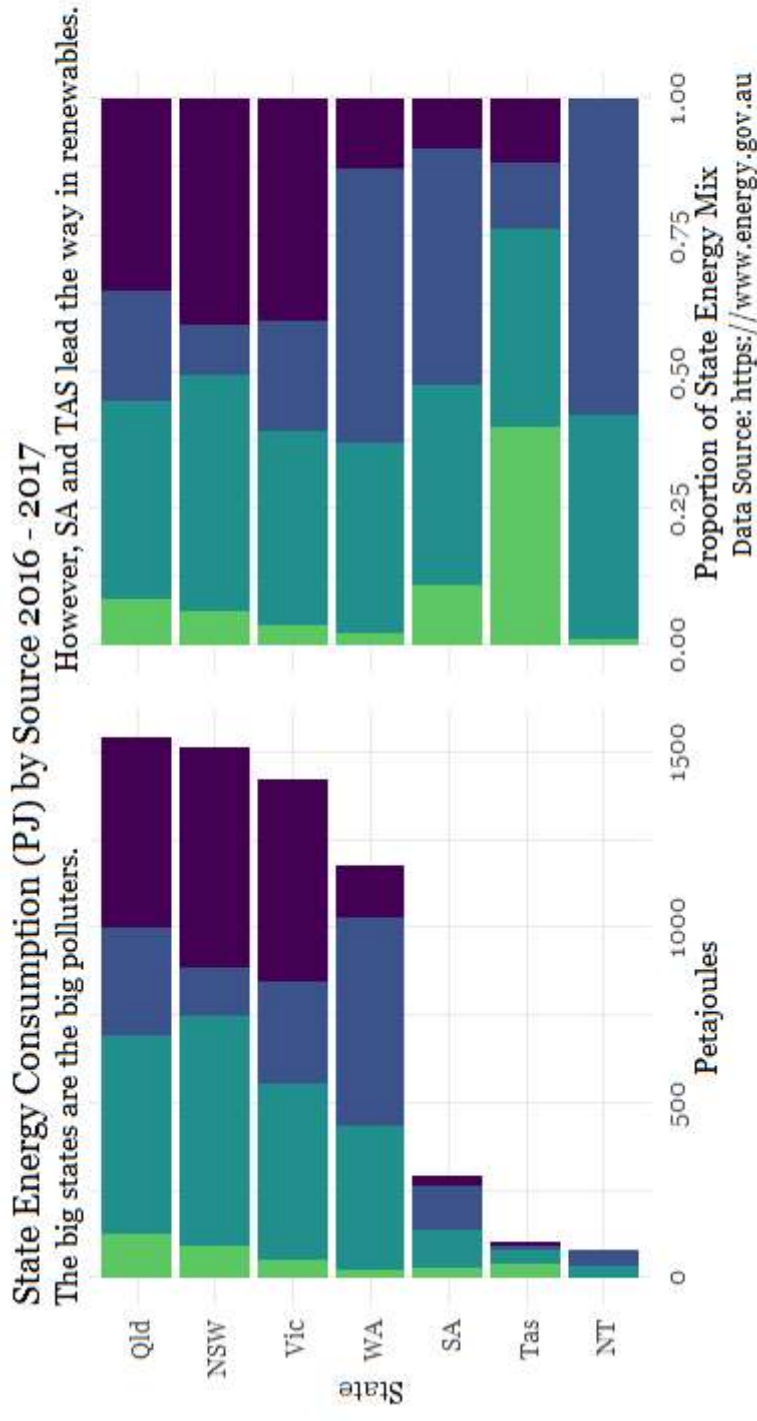
- The following table was taken from the Department of the Environment and Energy website - [Australian Energy Update Report 2018](#)

Energy use by State 2018 : Sheet1

Australian Energy Statistics, Table C						
Australian energy consumption (PJ), by state, by fuel, energy units						
Department of the Environment and Energy, Australian Energy Statistics, Table C, August 2018						
2016-2017						
State	Coal	Oil	Gas	Renewables	Statistical Discrepancy	Total
NSW	624.70	656.00	139.30	94.90	15.80	1,530.60
Vic	577.7	503.8	286.8	55.3	-19.3	1404.3
Sheet1						

- How do the states compare in terms of renewable energy use?

Turning Tables



A Visual Design Process - Andy Kirk (2012)

- **Guiding Principles**
 - Strive for form and function
 - Justify everything you do
 - Keep it accessible and intuitive
 - Avoid deceiving the viewer
- **Identifying a targeted audience and a data visualisation design objective**
 - Audience: Who are you communicating with?
 - Objective: the “so what?” question.
- **Focusing, justifying and choosing methods**
- **Construction and evaluation**

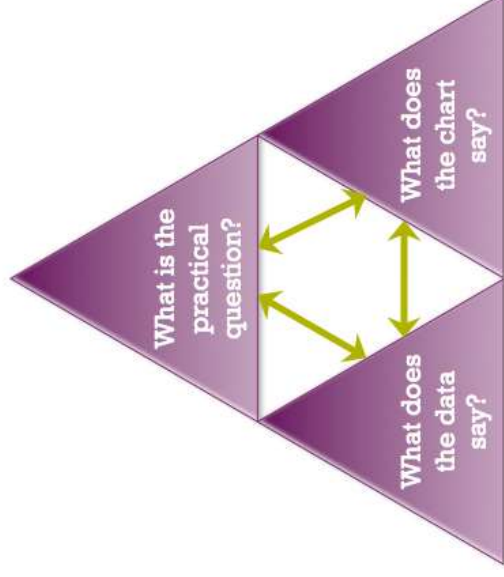
Audience and Objective

- Identify the audience and objective of the following visualisation:
 - 110 years of Australian temperatures -
<http://www.bom.gov.au/climate/history/temperature/>
(Bureau of Meteorology 2020)

Trifecta Check-up

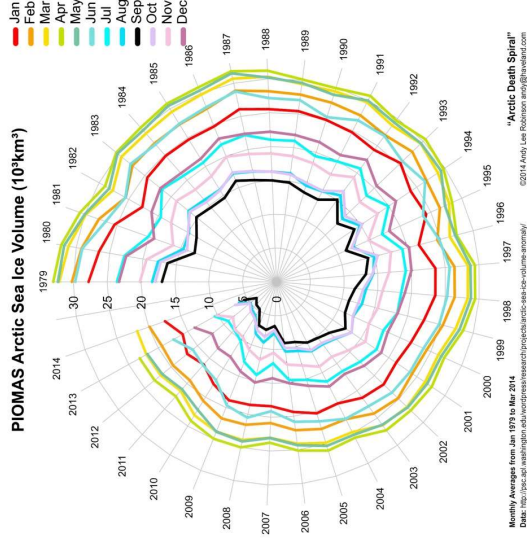
Kaiser Fung, author of [Junk Charts](#), provides a very simple and powerful framework, called the Trifecta Check-up, to use when evaluating a data visualisation.

Junk Charts Trifecta Checkup



Critique

- Critique the following data visualisation by Andy Lee Robinson according to the Trifecta check-up



Arctic Death Spiral

- You can read Kaiser's review [here](#).

Data Vis Check-list

- Evergreen and Emery (2016)'s [check-list](#) is a useful tool for avoiding common data visualisation pitfalls.
- A data visualisation is scored according to five criteria:
 - Text
 - Arrangement
 - Colour
 - Lines
 - Overall

Data Vis Check-list Cont.

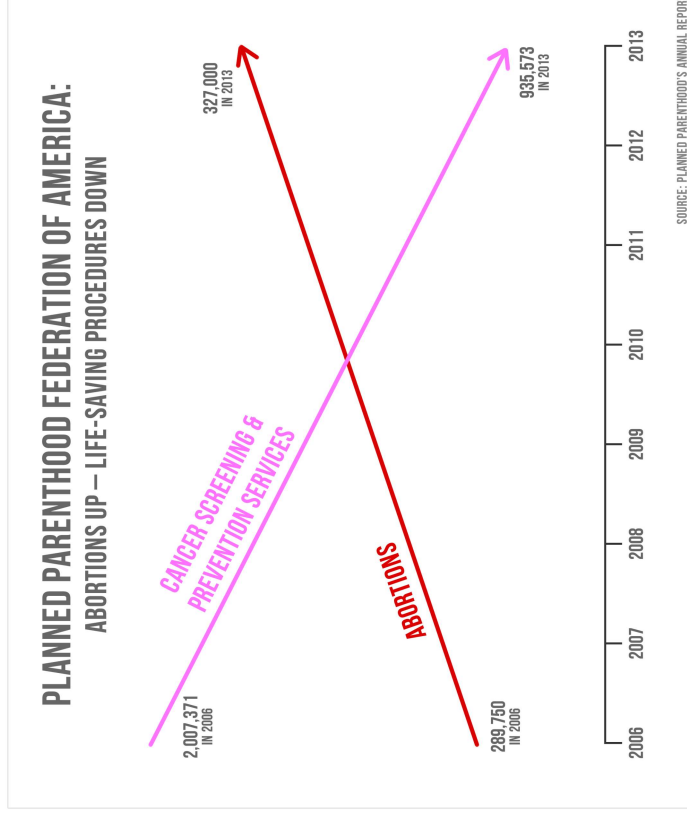
- Let's apply the check-list to the following data visualisation
- by Norbury (2018)
- We will use the [online check-list tool](#)



Ethical Principles of Data Visualisation

- **Ethics** can be defined as “*a set of beliefs about what is morally right and wrong*”
- Data visualisation ethical principles include the following:
 - Beneficence
 - Transparency
 - Accuracy
 - Objectivity
 - Respect
 - Accountability

Ethical Example - Paprocki (2015)



- Abortions increased from 289,750 in 2006 to 327,000 in 2013 = 13% increase
- Screening and preventative services decreased from 2,007,371 in 2006 to 935,573 in 2013 = 53% decrease

Data Integrity

- **Integrity** can be defined as “*the quality of being honest and having strong moral principles that you refuse to change*”
- Data integrity issues include the following:
 - Permission
 - Security
 - Consent
 - Privacy and sensitive information
 - Data quality
 - Citation

Strava Case Study

- Read the article [Strava's fitness heatmaps are a 'potential catastrophe'](#)
- Current Strava Heatmap - <https://www.strava.com/heatmap>
- Think about the ethical and integrity issues raised

References

- Buja, A., D. Cook, H. Hofmann, M. Lawrence, E.-K. Lee, D. F. Swayne, and H. Wickham. 2009. "Statistical inference for exploratory data analysis and model diagnostics." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1906): 4361–83. <https://doi.org/10.1098/rsta.2009.0120>.
- Bureau of Meteorology. 2020. "110 years of Australian temperatures." <http://www.bom.gov.au/climate/history/temperature/>.
- Evergreen, S., and A. K. Emery. 2016. "Updated data visualization checklist." <https://stephanieevergreen.com/updated-data-visualization-checklist/>