THE UNIVERSITY OF MELBOURNE

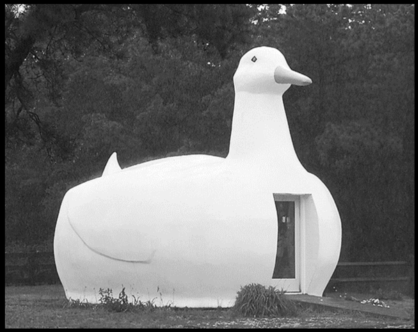


Image: Tufte (2001)

GEOM90007

SPATIAL VISUALISATION

LECTURE 4:

DATA GRAPHICS 2 - PRINCIPLES OF GOOD DATA GRAPHICS

REVISION

Word	Frequency
data	4009
one	1125
variables	1105
each	1103
graphics	1086
two	1067
visualisation	1039
plot	976
statistical	945
using	933
Document Total	**576,541**

WORD CLOUD

Which variables from the dataset are typically visualised?

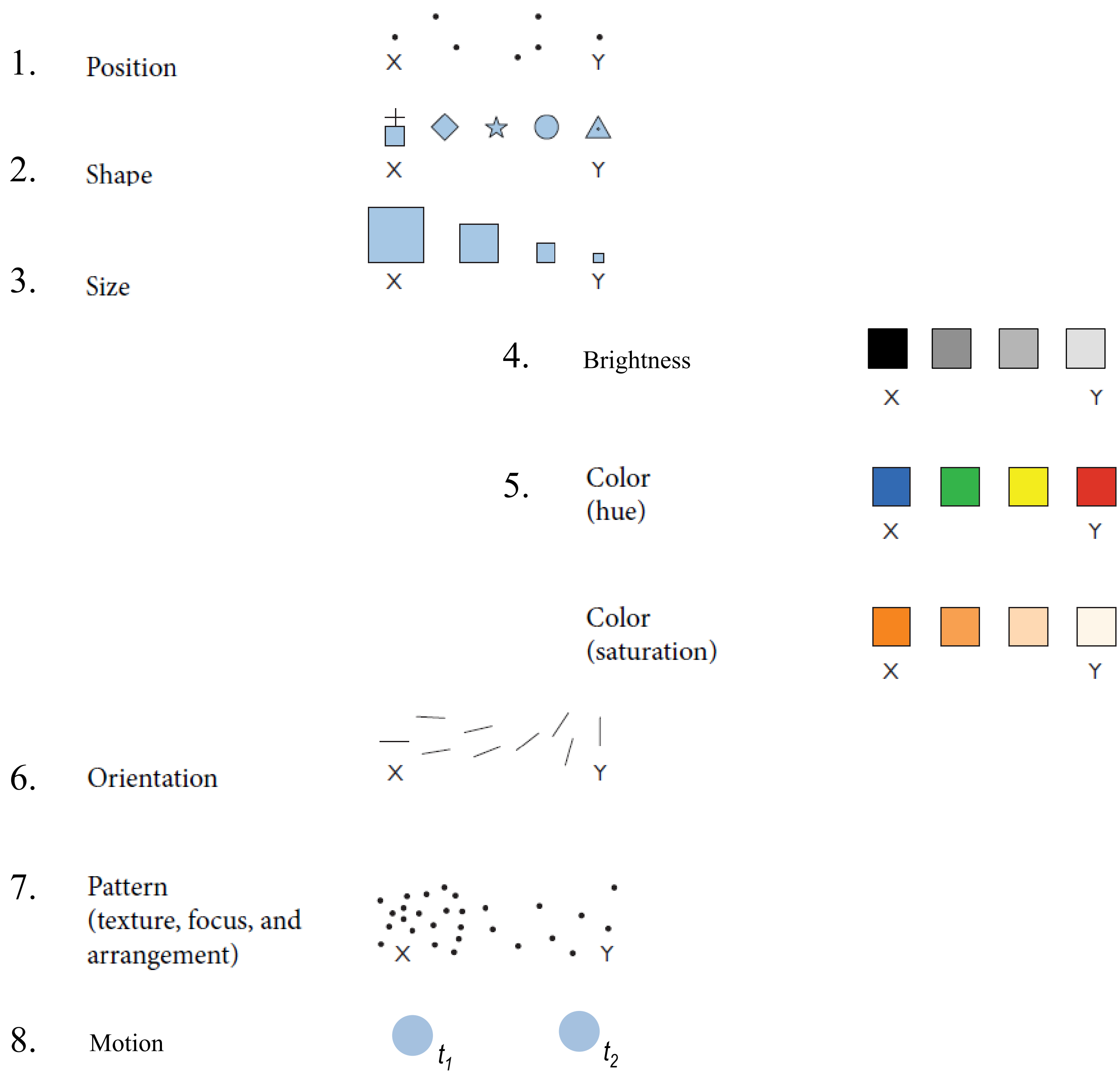
1. Size

What else could be visualised?

2. Position
3. Brightness
4. Colour (hue, saturation)
5. Orientation
6. Pattern (arrangement)

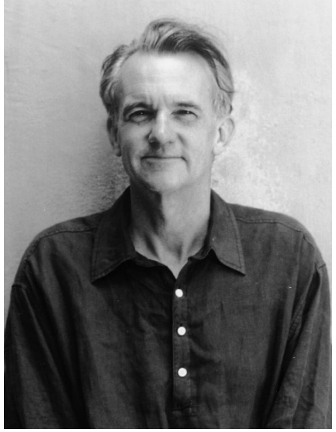
…

VISUAL VARIABLES



Images: Reformatted from Ward et al. (2010) to include 8 visual variables from Mackinlay (1986)

REVIEW (LECTURE 1)



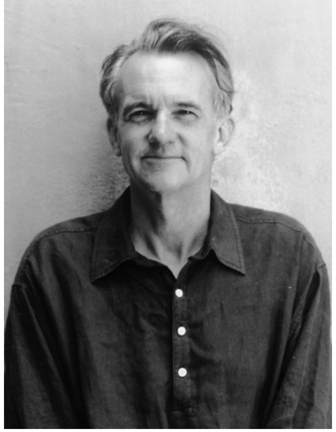
Excellence in statistical graphics consists of complex ideas communicated with:

- Clarity
- Precision
- Efficiency

Graphics *reveal* data.

(Tufte, 2001)

REVIEW (LECTURE 1)



Graphical excellence is that which gives to the viewer the ***greatest number of ideas*** in the ***shortest time*** with the ***least ink*** in the ***smallest space***

Graphical excellence is nearly always multivariate.

And graphical excellence requires telling the truth about the data.

(Tufte, 2001)

OVERVIEW

A. Data-ink ratio

B. Four principles of data graphics

C. Other topics: Iterative design

A. Data-ink ratio

“Data graphics should draw the viewer’s attention to the sense and substance of the data, not to something else”
(Tufte, 2001)

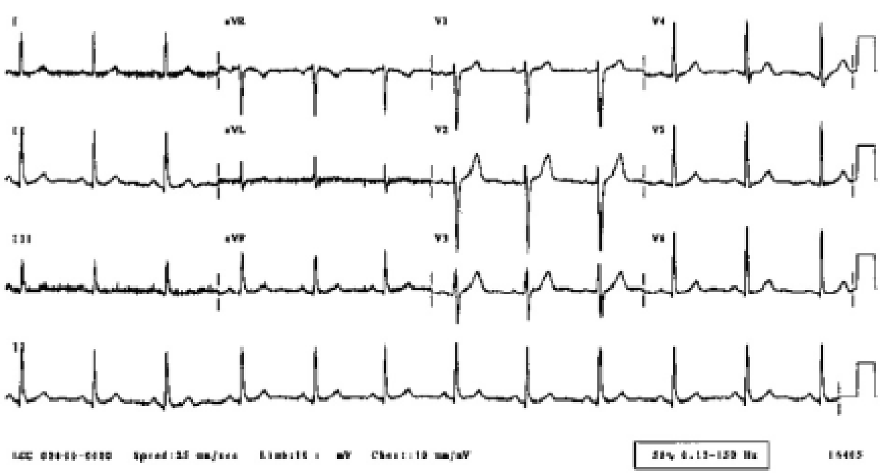
Therefore the majority of ink (or pixels) used in a graphic should present data-information.

$$Data - ink\ ratio^{*} \quad = \frac{ink\ (or\ pixels)\ used\ for\ data}{total\ ink\ (or\ pixels)\ used\ in\ graphic}$$

*Excludes background ink/pixels

Revisit the ECG example

A normal adult:



An 83-year-old adult with heart problems:

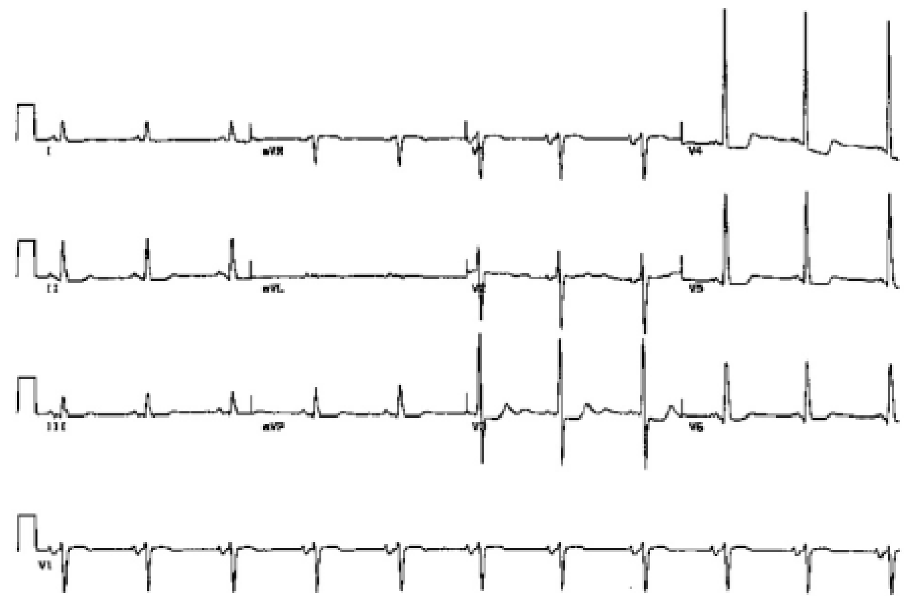
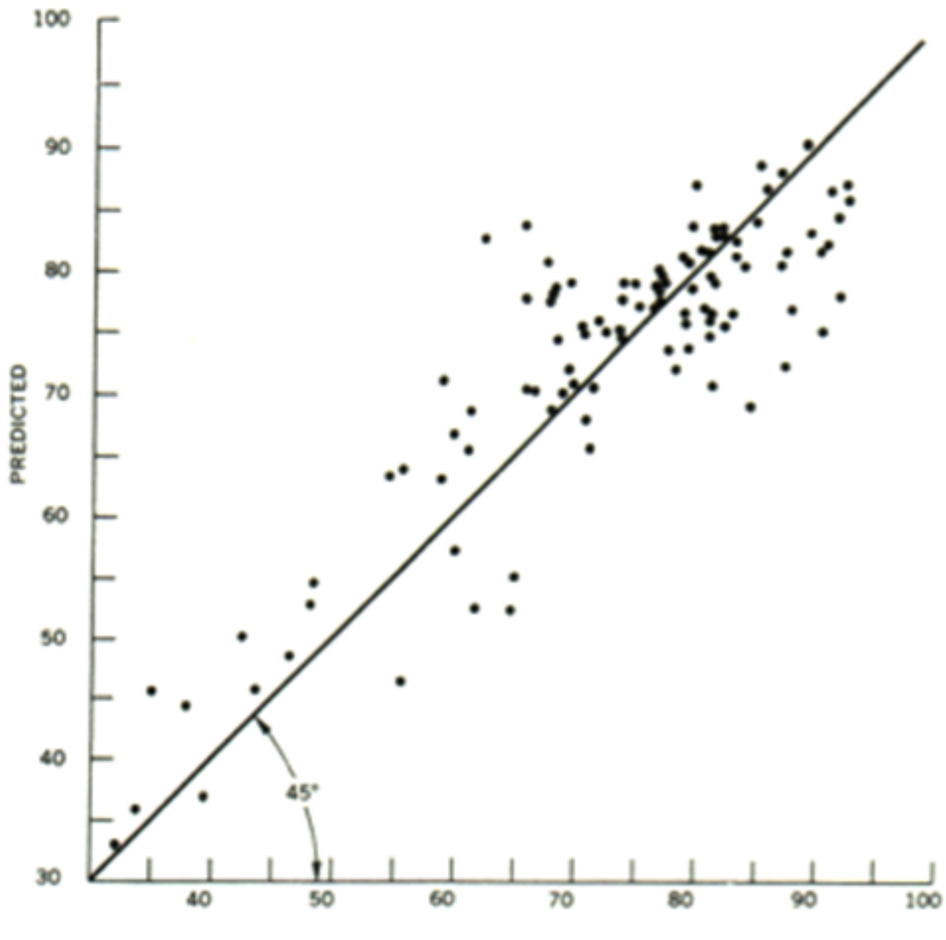


Image: http://www.ecglibrary.com/ecghome.html

All ink (or pixels) represent data direct from the ECG.
Nothing can be erased without losing information!

A. Data-ink example

Image: Kelley et al. (1970)



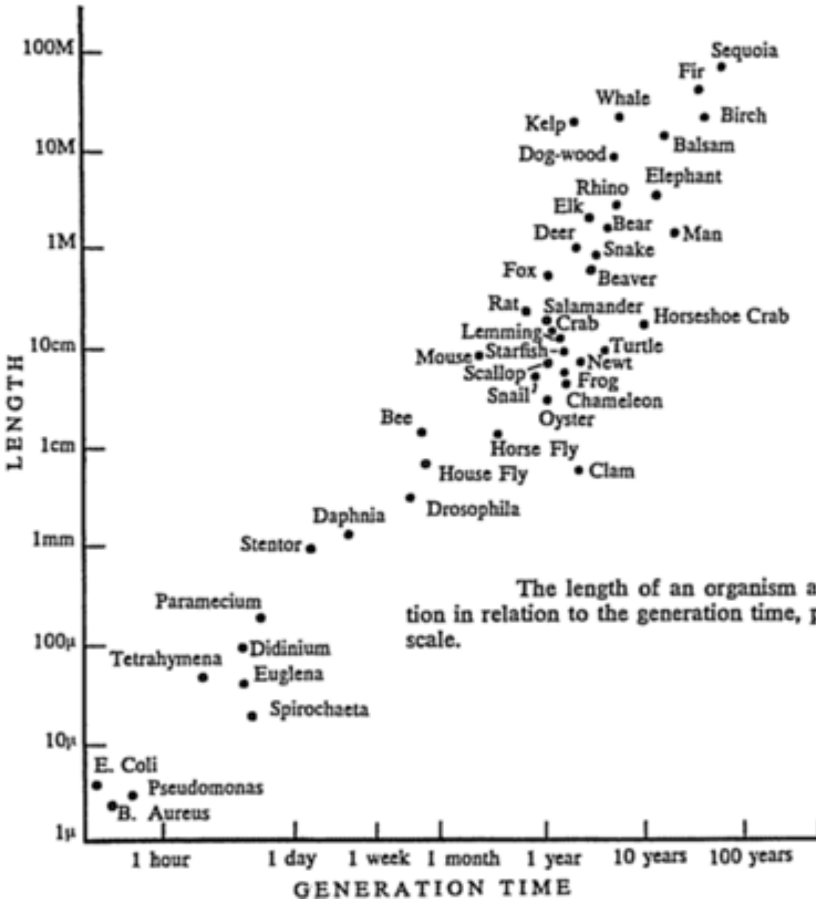
Very small data points.

Majority of ink is unrelated to data (background can be classified as junk)

Ratio < 0.1

A. Data-ink example

Image: John Tyler Bonner (1965)



The length of an organism at the time of reproduction in relation to the generation time, plotted on a logarithmic scale.

Improvement

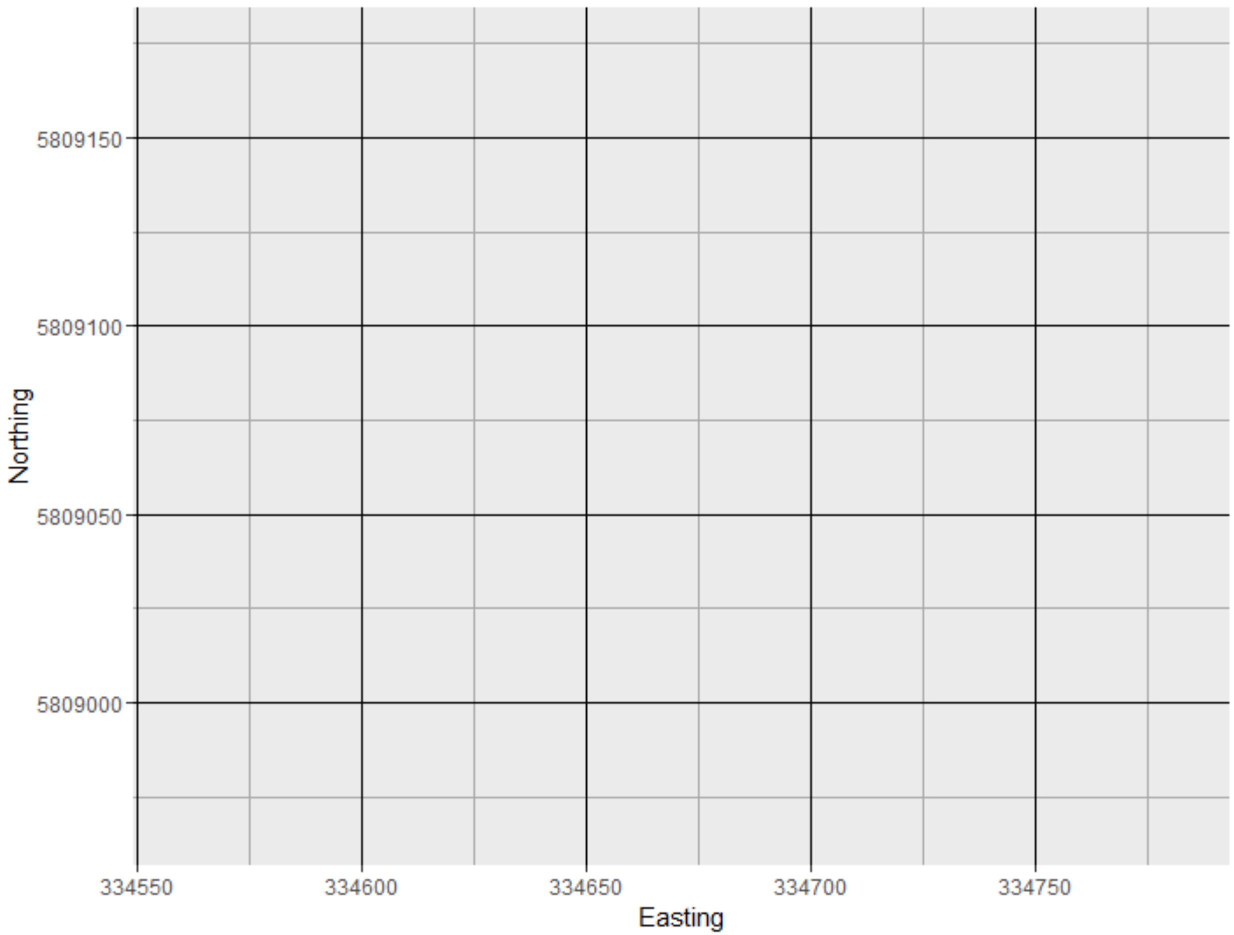
Majority of ink is related to data

Use of descriptive text is debateable

Ratio > 0.75

Image:

A. Data-ink example



Nothing

No ink related to data

Ratio = 0

Maximise the data-ink ratio

Every bit of ink (or pixel) in a data graphic must be there for a reason. Therefore **maximise** the data-ink ratio (within reason)

Erase non-data-ink

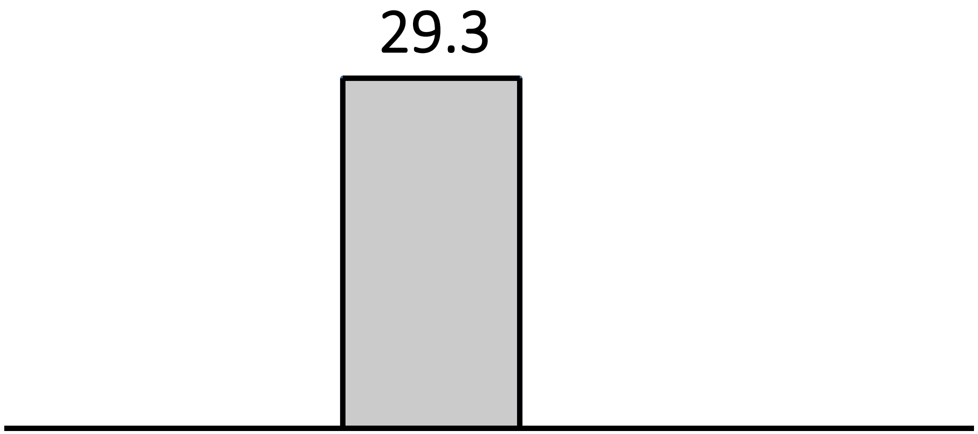
The erasing principle: **Erase** non-data-ink (within reason)

Such data can clutter the graphic (e.g., dense grid lines)

"For non-data-ink, less is more
For data-ink, less is a bore" (Tufte, 2001)

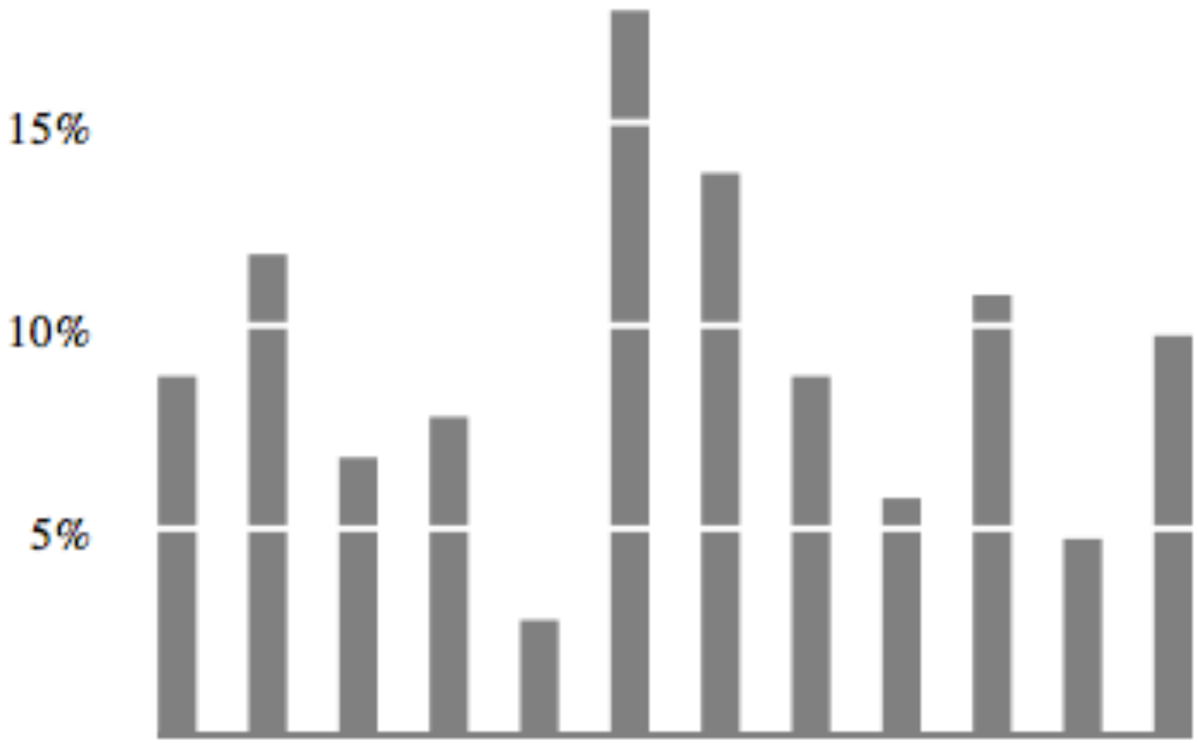
Example of redundancy with numerical data

Look for redundant data-ink, which are representations that are repeated and unnecessary

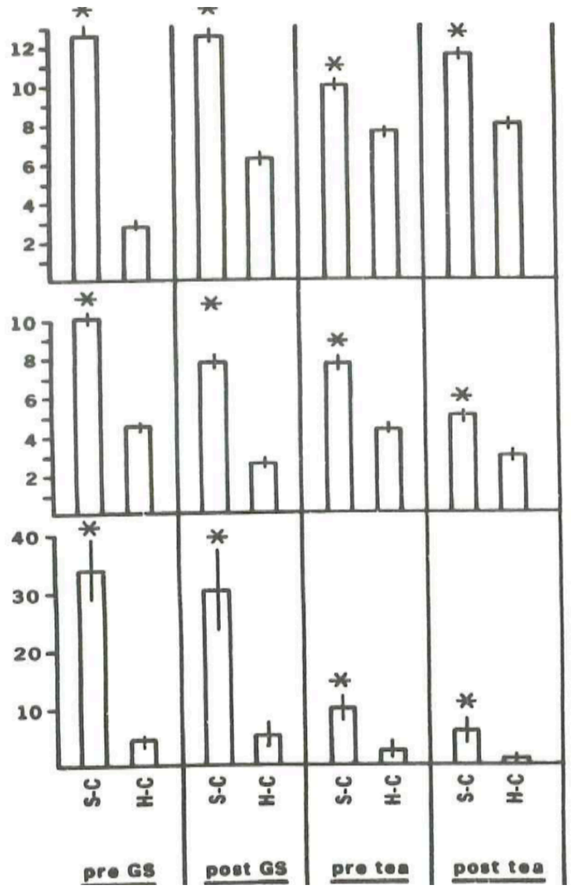


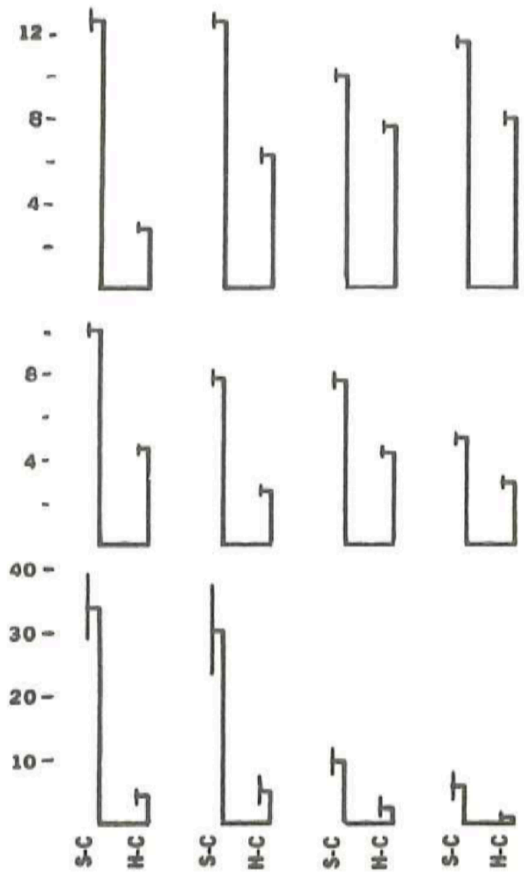
Any 5 of the following can be erased and the one remaining will still indicate height:
(1) Height of the left line, (2) Height of the right line, (3) Height of the shading,
(4) Position of the top line, (5) Position (not content) of the number, (6) the number itself

Redesign – Bar chart

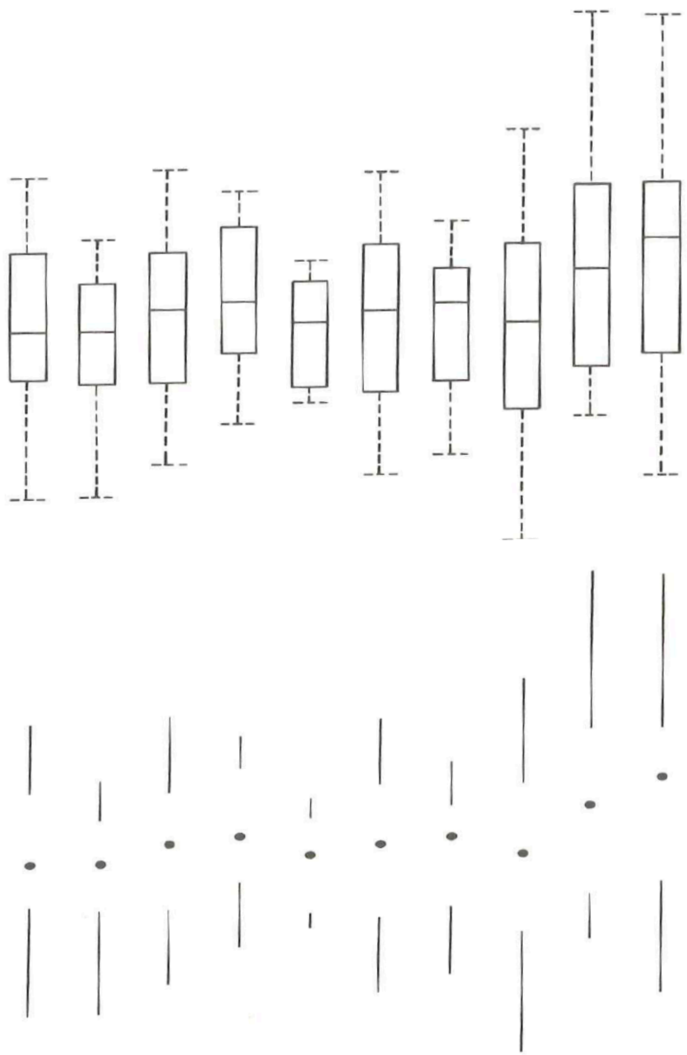


Redesign – Bar chart





Redesign – Box Plot



A. Data-ink summary

1. Above all else, show the data
2. Maximise the data-ink ratio
3. Erase non-data-ink
4. Erase redundant data-ink
5. Revise and edit

B. Four principles of data graphics

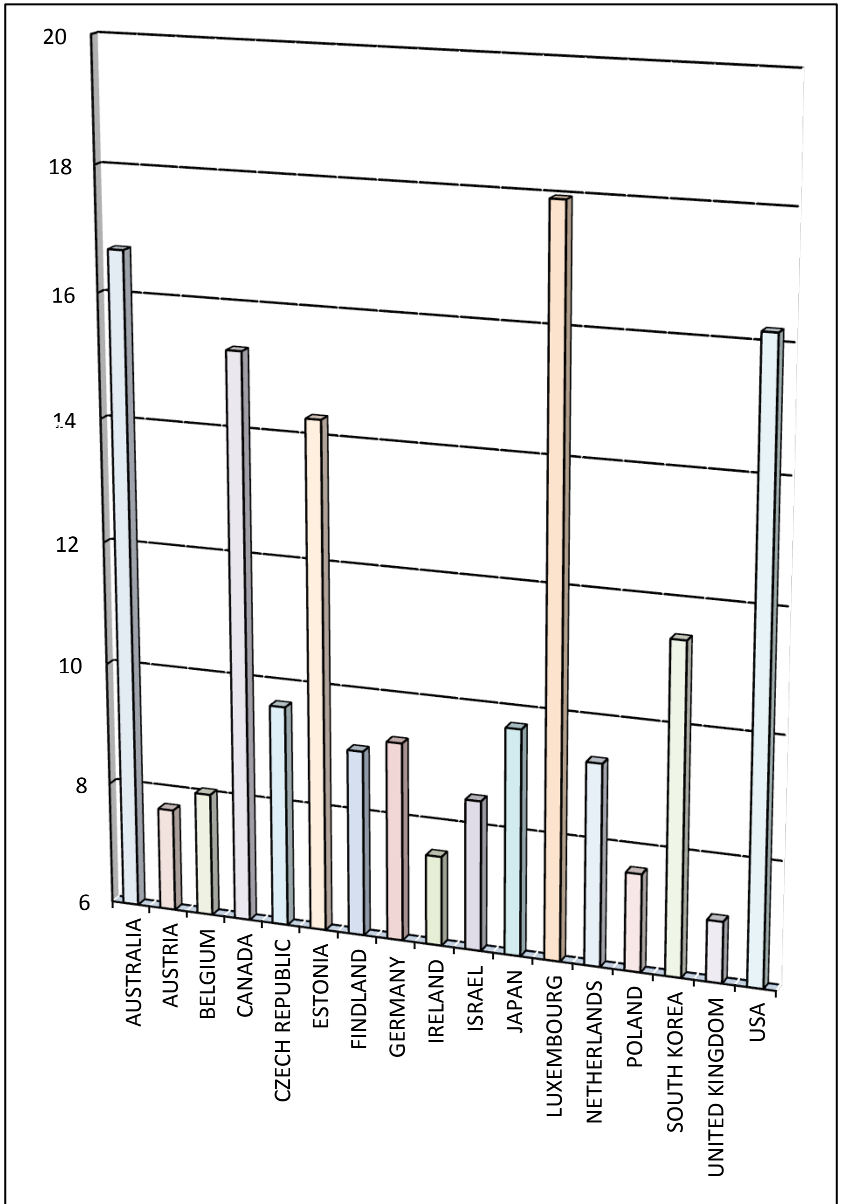
1. Data density
2. Data integrity
3. Data correspondence
4. Data aesthetics

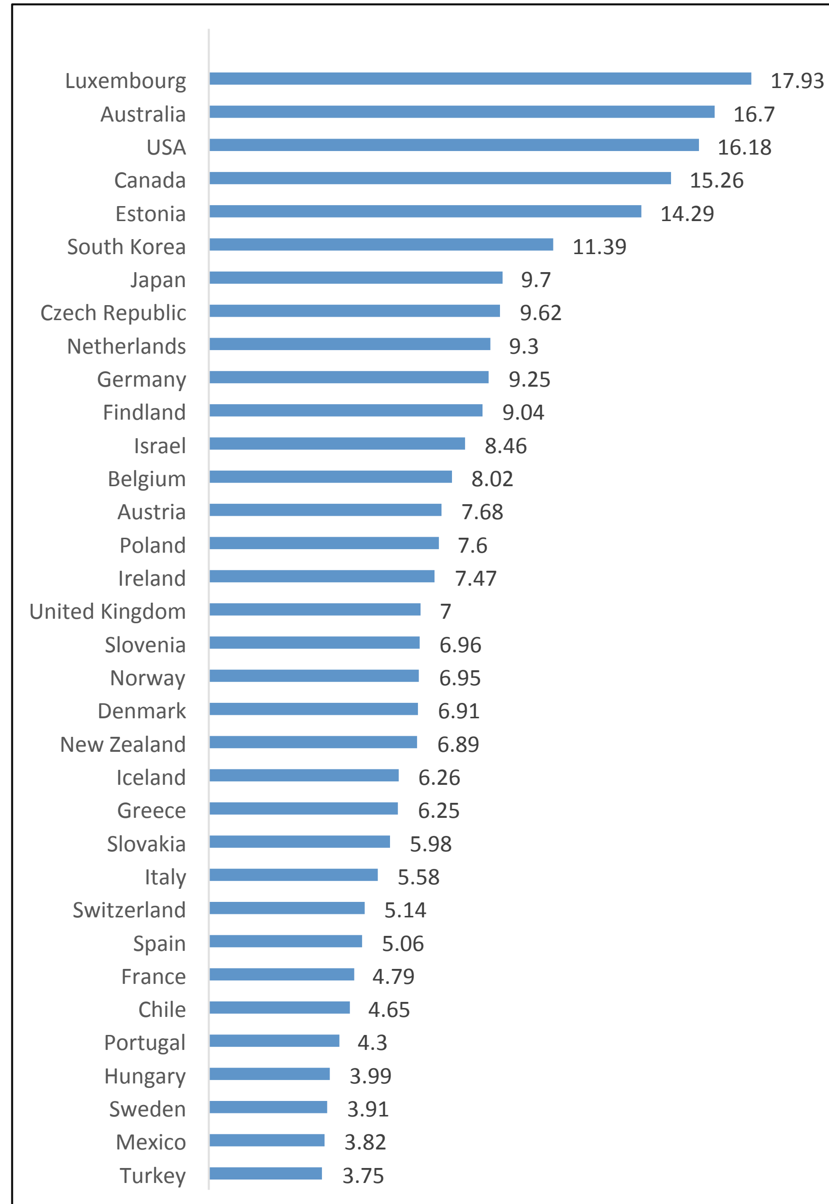
B. Example dataset: Table

Location	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Luxembourg	21.780	24.470	24.630	23.860	22.230	21.720	20.210	20.960	20.250	19.390	17.930
Australia	17.850	18.350	18.310	18.280	18.410	18.150	18.070	17.390	17.150	16.910	16.700
United States	19.300	19.390	19.260	18.750	18.850	18.100	16.660	17.290	16.730	16.010	16.180
Canada	16.890	16.480	16.610	16.100	16.850	16.210	14.990	15.150	15.270	15.080	15.260
Estonia	12.050	12.170	12.370	11.500	14.400	13.270	11.050	13.990	13.320	12.410	14.290
Korea	9.140	9.570	9.500	9.600	9.820	9.980	10.210	11.150	11.520	11.500	11.390
Japan	9.300	9.310	9.360	9.250	9.540	8.880	8.410	8.790	9.210	9.540	9.700
Czech Republic	11.900	11.940	11.580	11.580	11.740	11.120	10.400	10.590	10.520	10.050	9.620
Netherlands	10.310	10.380	10.000	9.840	9.880	9.990	9.550	10.130	9.400	9.360	9.300
Germany	9.950	9.750	9.540	9.700	9.320	9.440	8.800	9.280	8.940	9.090	9.250
Finland	13.620	12.790	10.410	12.570	12.130	10.500	9.980	11.480	10.060	8.990	9.040
Israel	9.080	8.950	8.450	8.750	8.900	8.770	8.530	8.980	8.700	9.460	8.460
Belgium	10.830	10.660	10.250	9.890	9.390	9.710	8.660	9.360	8.600	8.030	8.020

OECD C02 emissions (tonnes per capita) 2013
https://data.oecd.org/air/air-and-ghg-emissions.htm

B. Two graphics: Which do you prefer? Which do you think is better?





Approximately the same chart area

B.1 Data density

- The numbers represented in a data graphic can be represented in an abstract ‘data matrix’

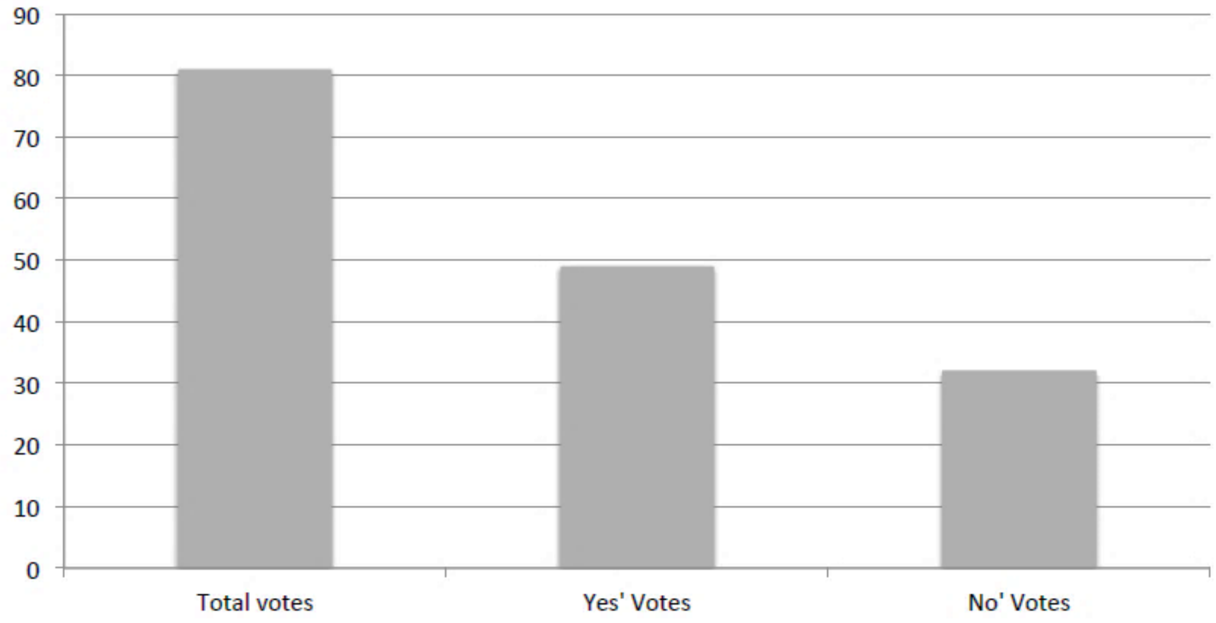
$$Data\ density = \frac{Number\ of\ entries\ in\ data\ matrix}{Available\ display\ area}$$

Example units: information per mm^2

- A data graphic that requires a description of 2000 words is typically more dense than one that requires 100 words

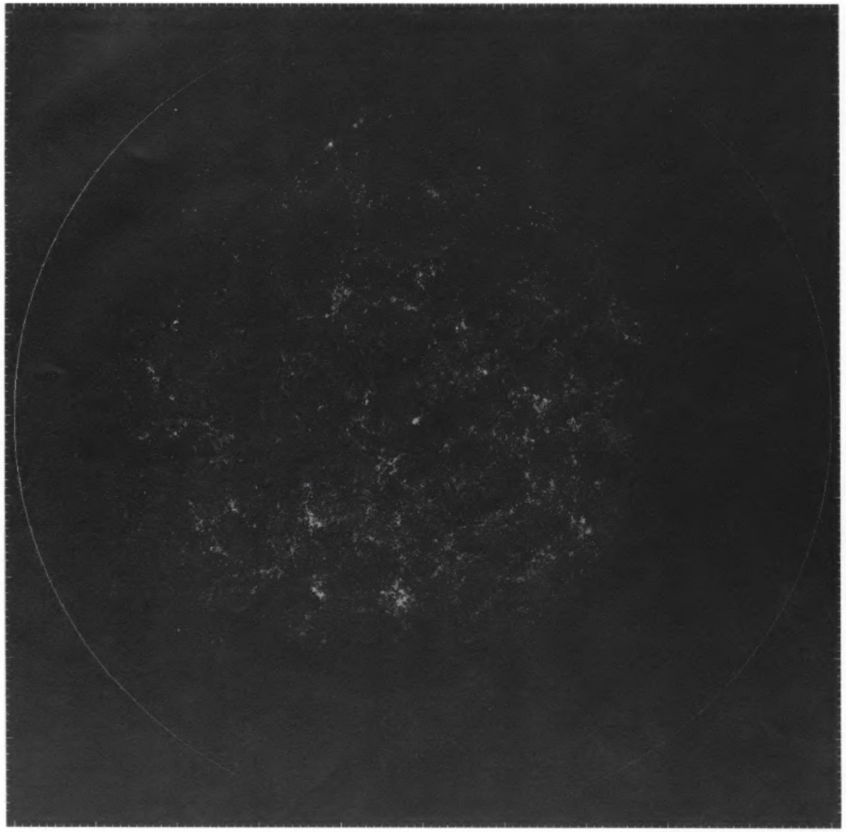
(Duckham, 2014)

B.1 Data density – Low example



- Data matrix contains 4 entries only: 2 numbers and 2 names
    - Total votes is a derived quantity
- Very low density

B.1 Data density – High example



Seldner et al. (1977)

Map area:

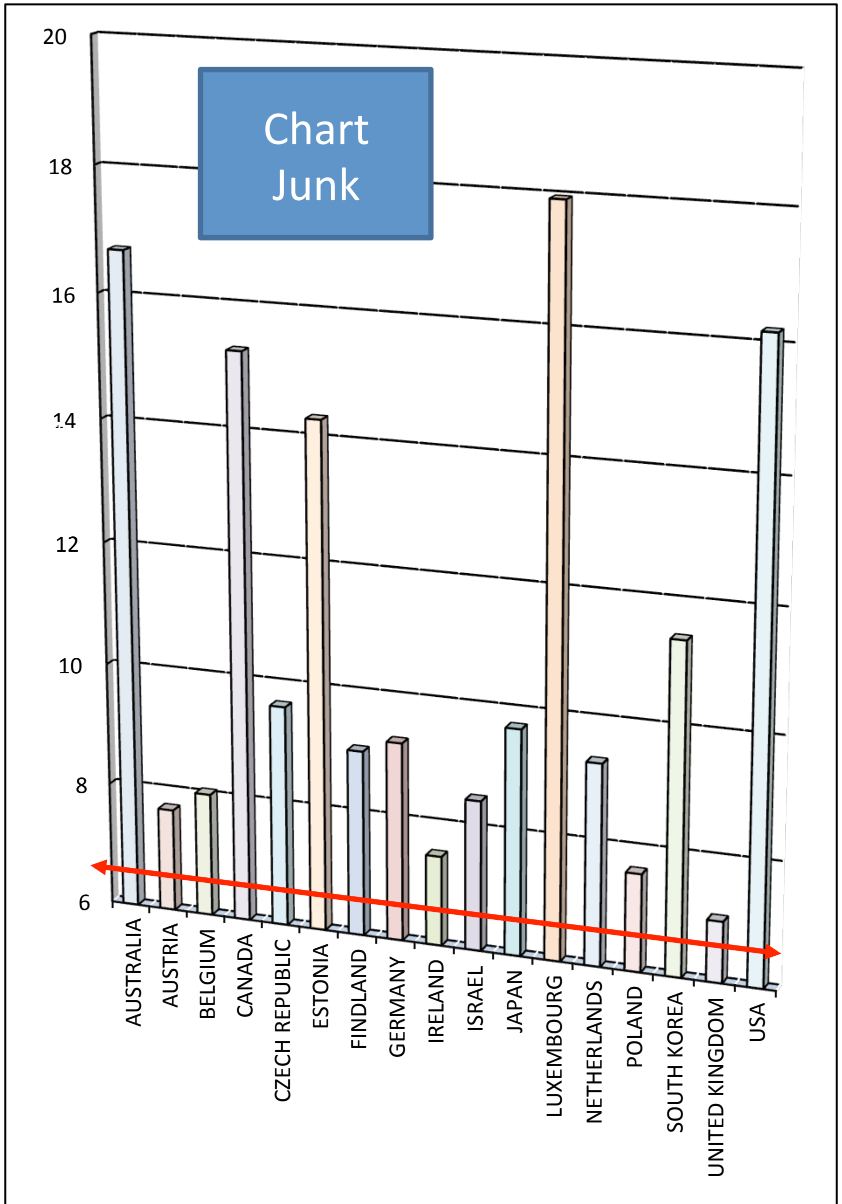
- 2D surface, 61 inches (393 cm^2)

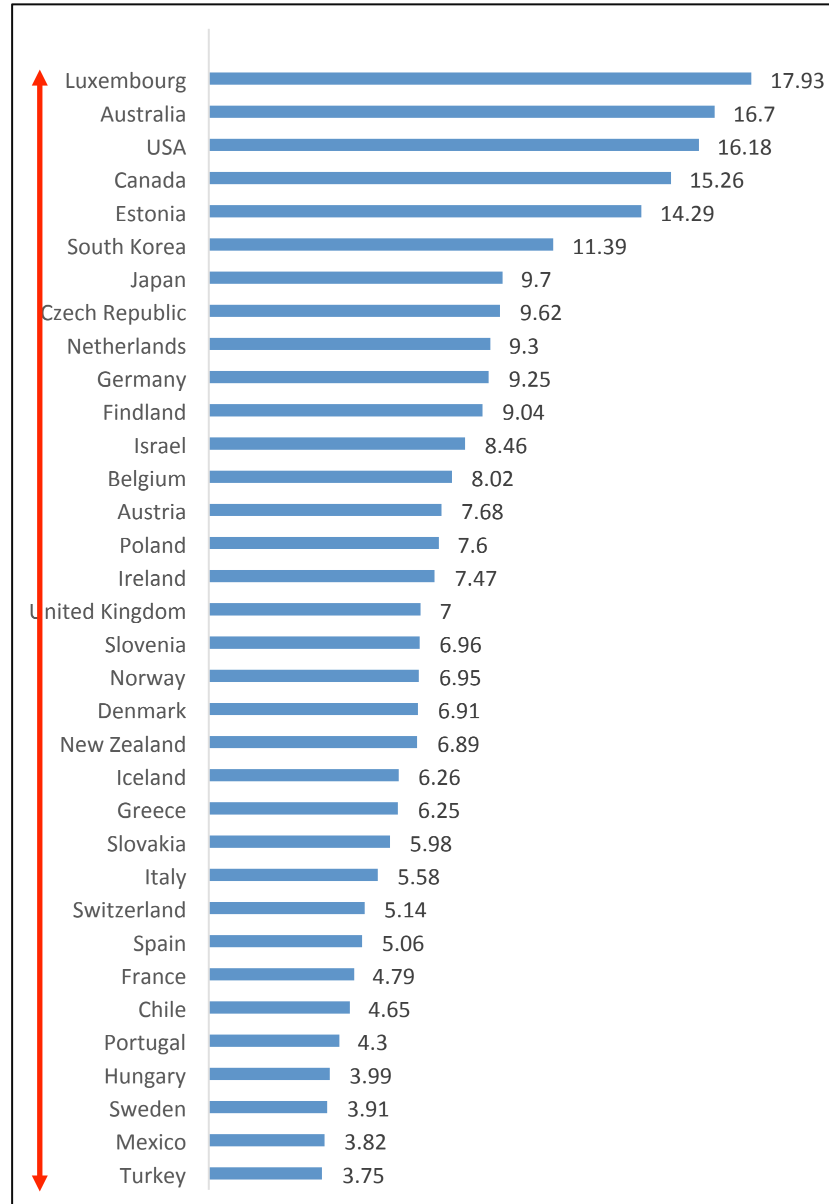
Data

- 1024 x 2222 rectangles (totalling 2,275,328), representing three numbers:
  - Two coordinates ~ position (x,y)
  - One galaxy count ~ brightness (grey)

- Data density
  - ~ 17,369 numbers per cm^2

B.1 Data density





B.2 Data integrity

Data graphics can be distorted which may mislead the viewer

1. Data scrubbing

Deliberately removing data (e.g., outliers) or deriving data (smoothing/resampling) to create a false message

2. Unbalanced scaling

Ratios (transformations) between raw data and visualisation

3. Range distortion

Move axis (at least one dimension) and not making this explicit

4. Abusing dimensionality

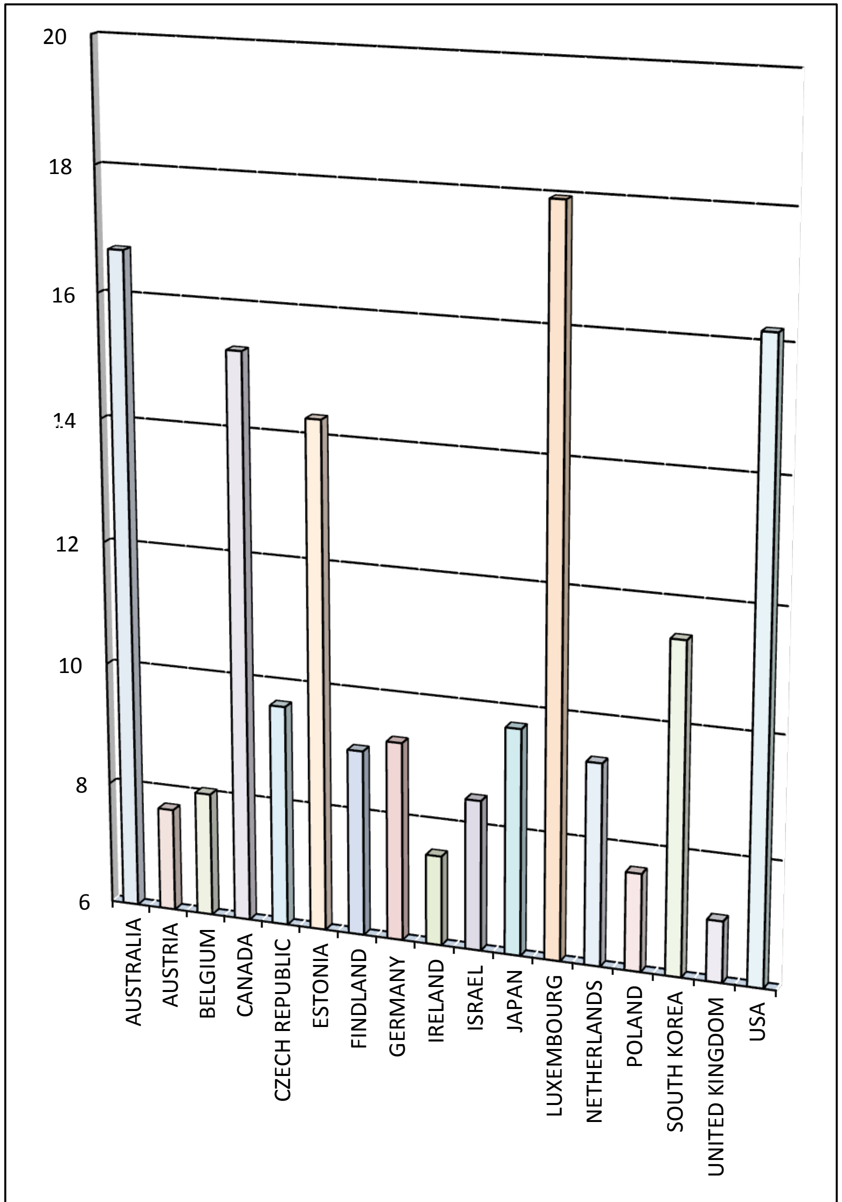
Errors resulting from dimensionality (rises with the power of the dimensions displayed: lines, areas, volumes)

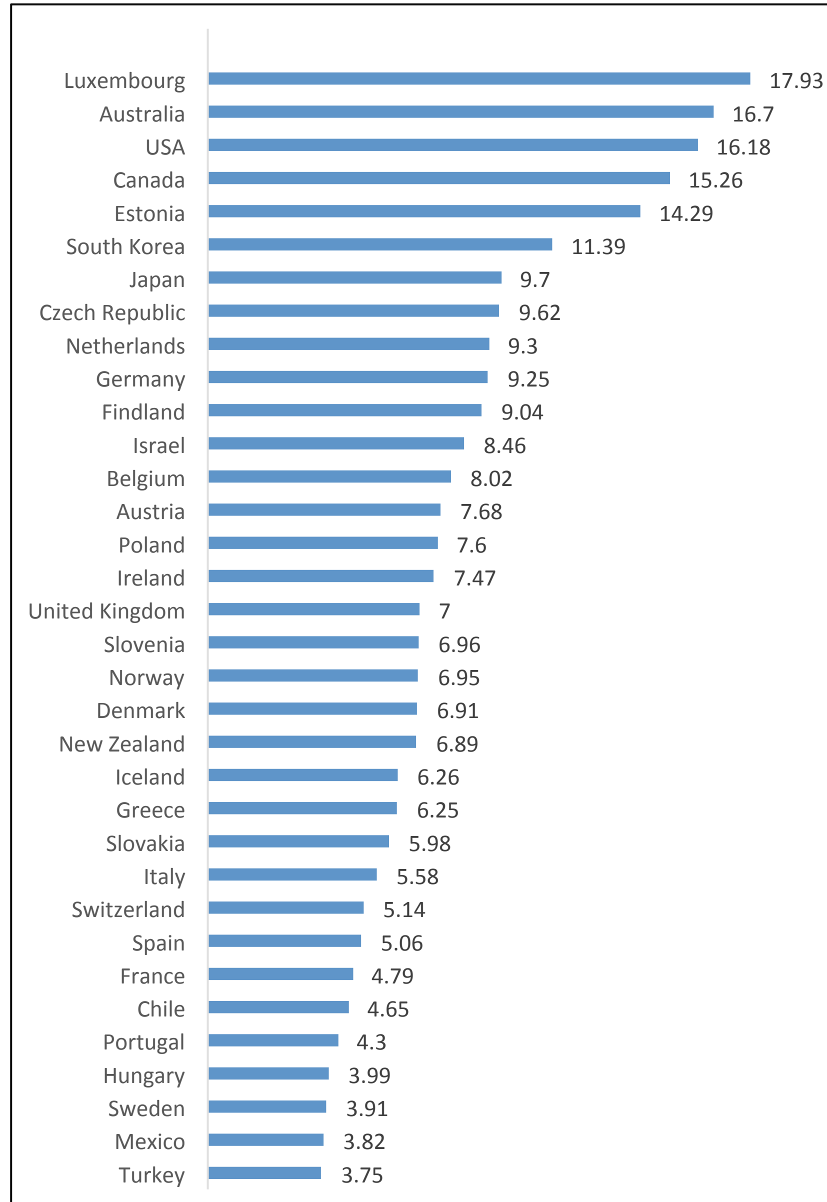
B.2 Data integrity

Data graphics can be distorted which may mislead the viewer

- Show entire scale & Maintain proper scale
- Avoid distortion
  - Lie factor = *size of effect shown in graphic/size of effect in data*
- Size encoding
- Additional principles
  - Clear, detailed labels to defeat distortion and ambiguity
  - Show data variation, not design variation
  - Account for inflation (time & money)

B.2 Data integrity

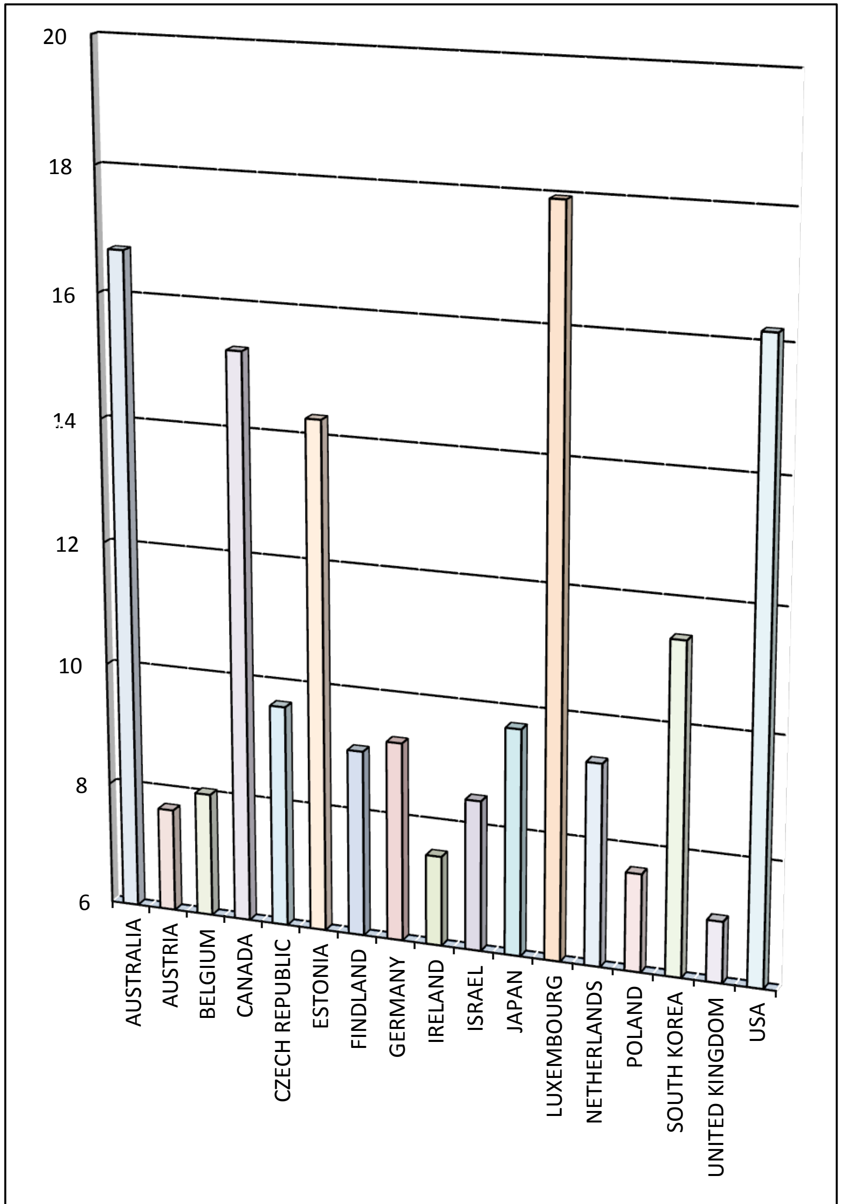


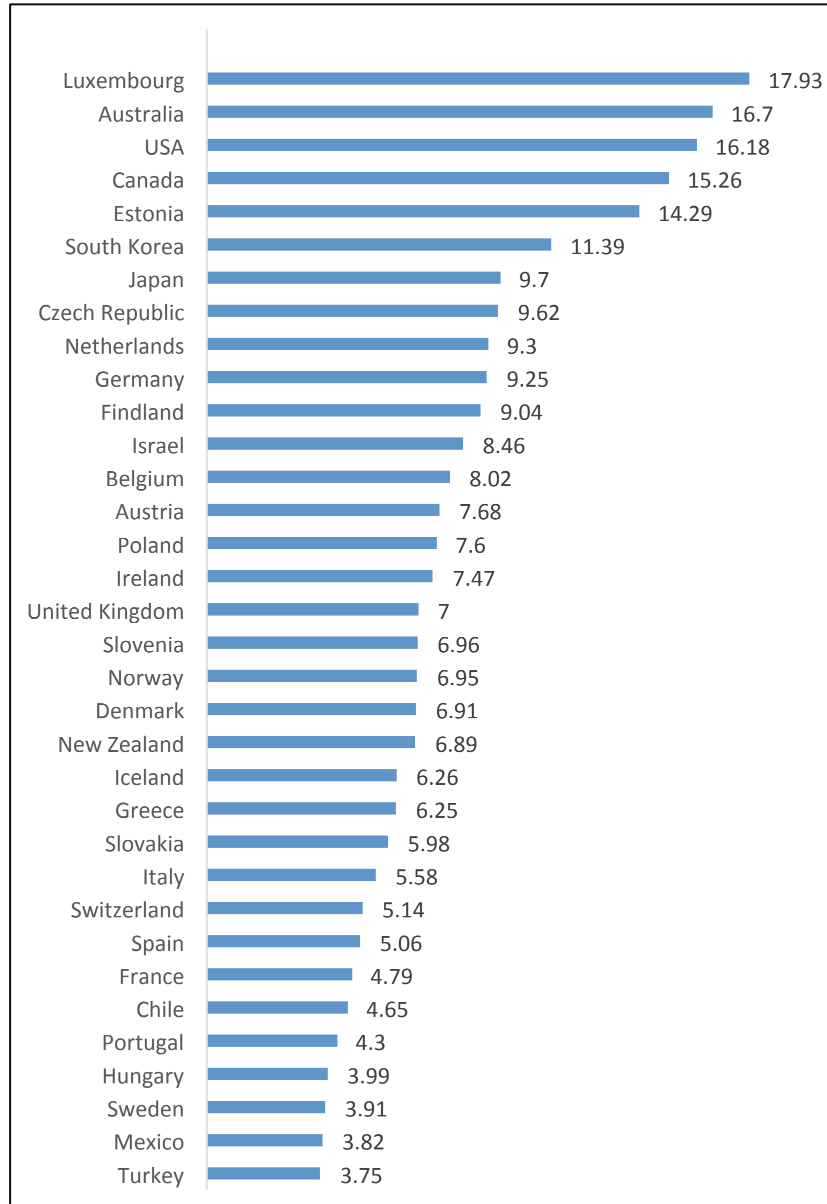


B.3 Data correspondence

- Good correspondence is helping the viewer understand the patterns in the data and form hypotheses

B.3 Data correspondence





B.4 Aesthetics

- Data graphics today should be aesthetically pleasing or… 'beautiful'
- Historically, aesthetics can be linked to the way art evokes an emotional response
- Aesthetic Science is examined across three broad fields:
  - Philosophy
  - Psychology
  - Neuroscience

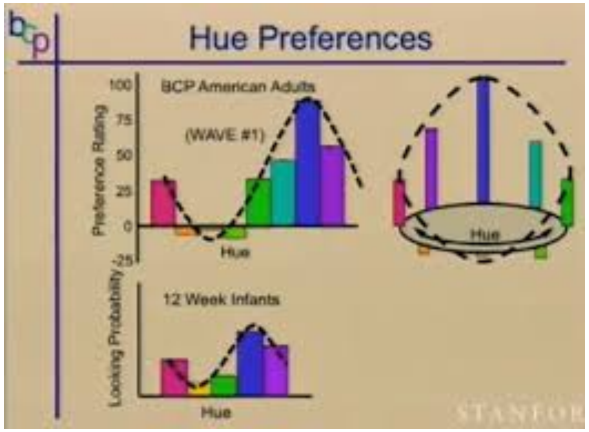
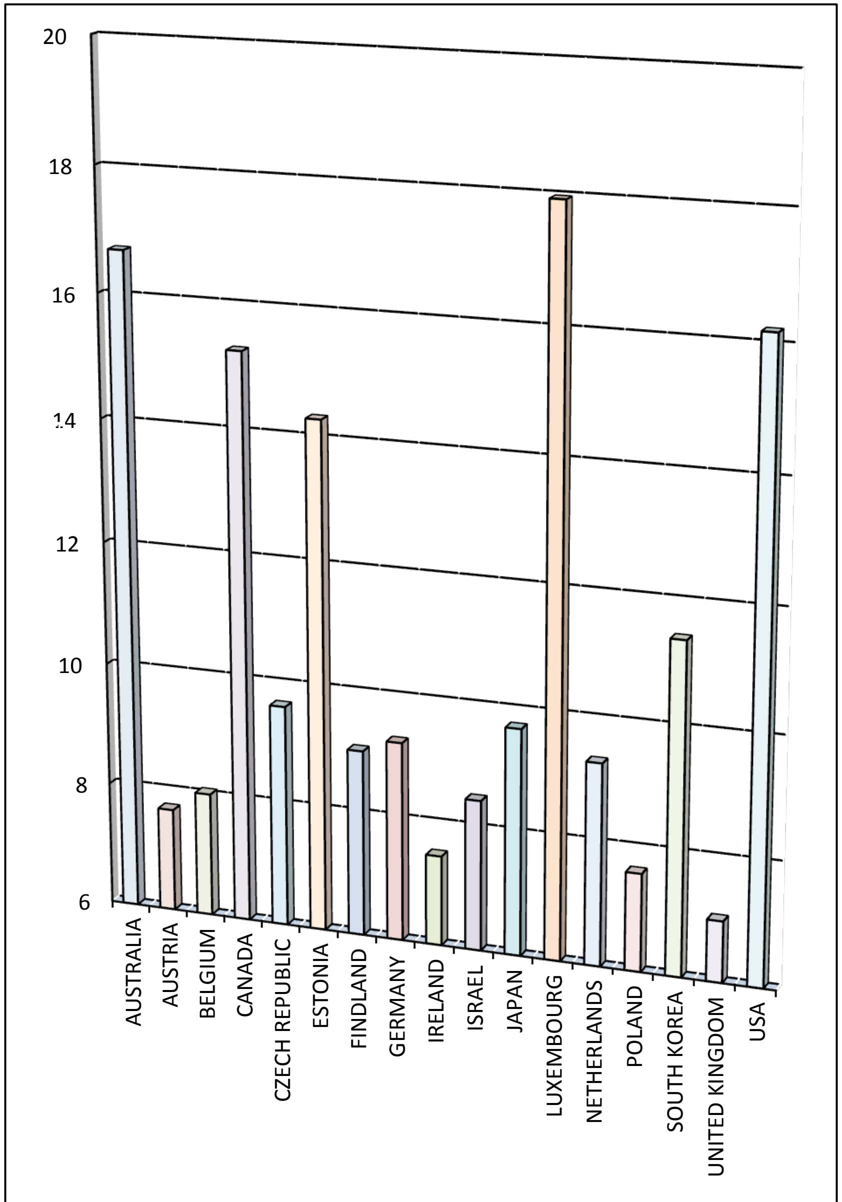
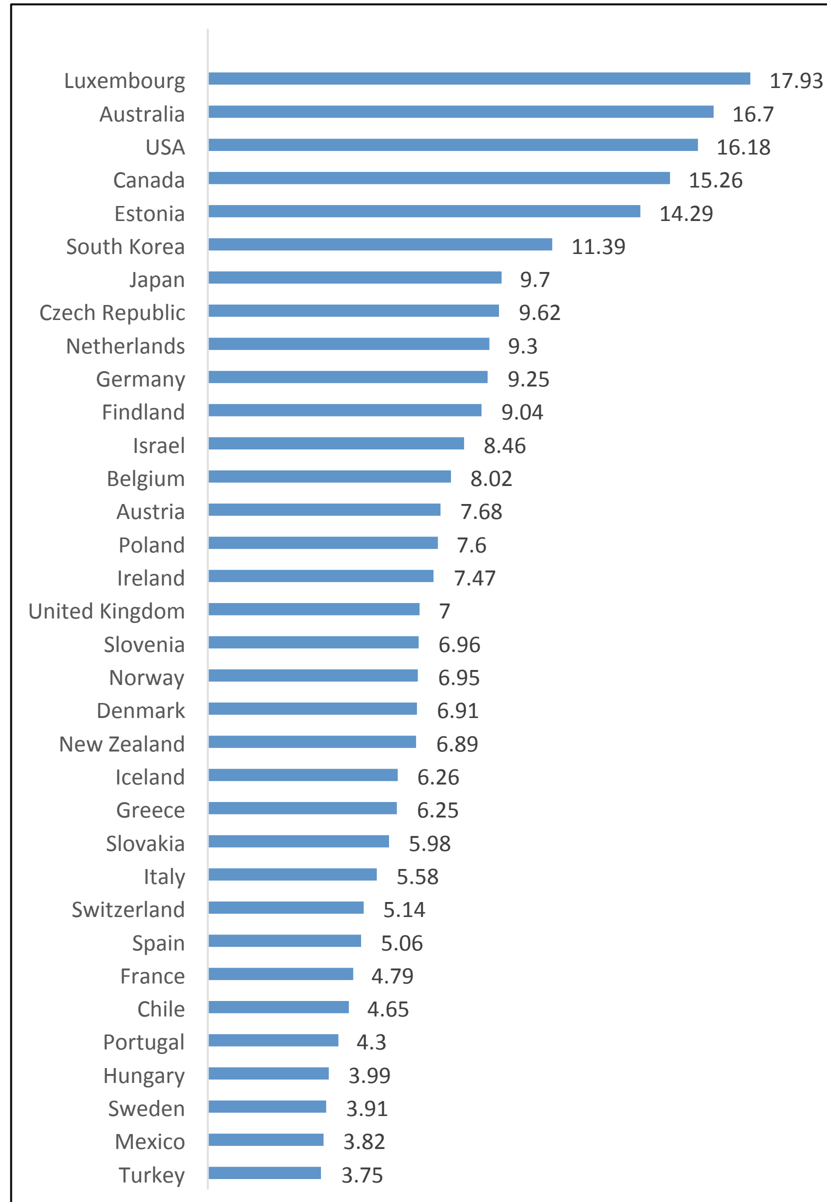


Image: Palmer (2009), continued Shimamura and Palmer (2013)

B.4 Aesthetics





B.4 Understanding aesthetics

- “Graphical excellence is often found in simplicity of design and complexity of data“
  (Tufte, 2001)

- Proven ‘graphic design principles’ may be adopted:
  - Balance (symmetry and use of display space)
  - Proportion and scale
  - Simplicity (don’t try to do too much)

https://www.google.com.au/?client=firefox-b#q=design+principles+edu

B.4 Making complexity accessible

Friendly

Words are spelled out, jargon is avoided

Words run left to right (usual direction for reading English)

Elaborate shading and symbols are avoided

Labels are placed on the graphic itself, no legend required

Colour deficiencies are considered

Type is clear and precise

Unfriendly

Multiple abbreviations, requiring viewer to re-read text to decode

Words run vertically, e.g., Y Axis, or in many directions

Obscure codings requiring the viewer to constantly revert to the legend

Chart-junk

Type is unreadable

B.4 Making complexity accessible

If the data suggests a certain shape, follow it, however…

Wider plots make it easier for the eye to follow from left to right
(Tukey, 1977)

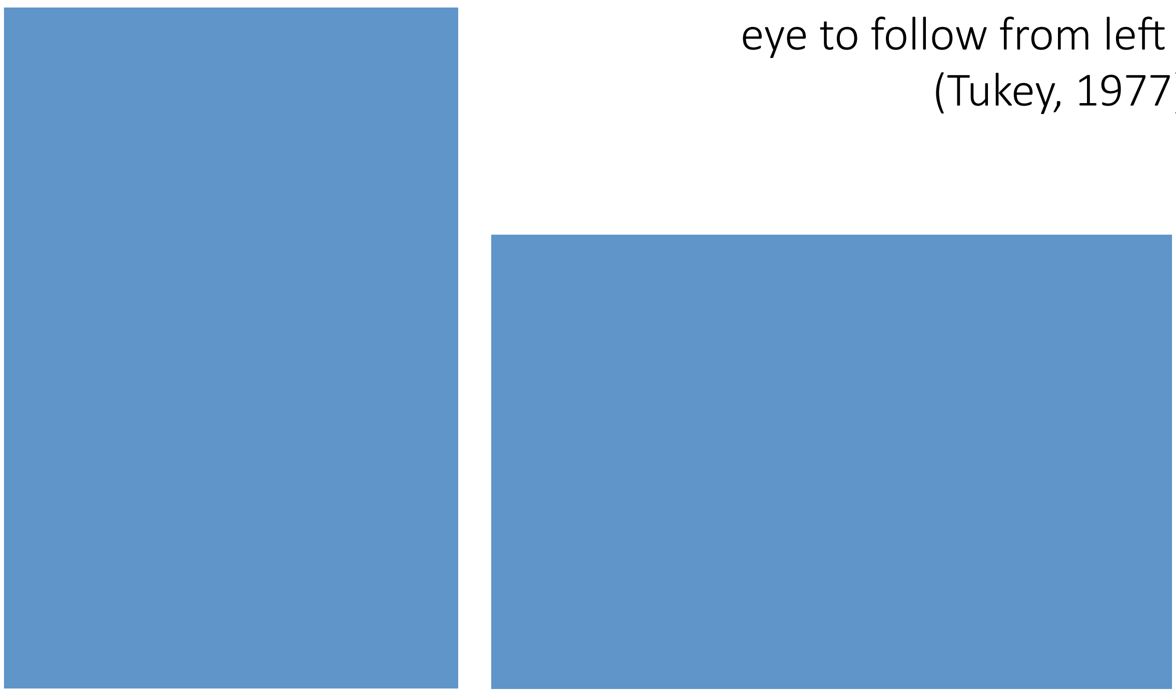


Chart junk

- "Decoration… to make the graphic appear more scientific and precise… to give the designer an opportunity to exercise artistic skills" (Tufte, 2001)

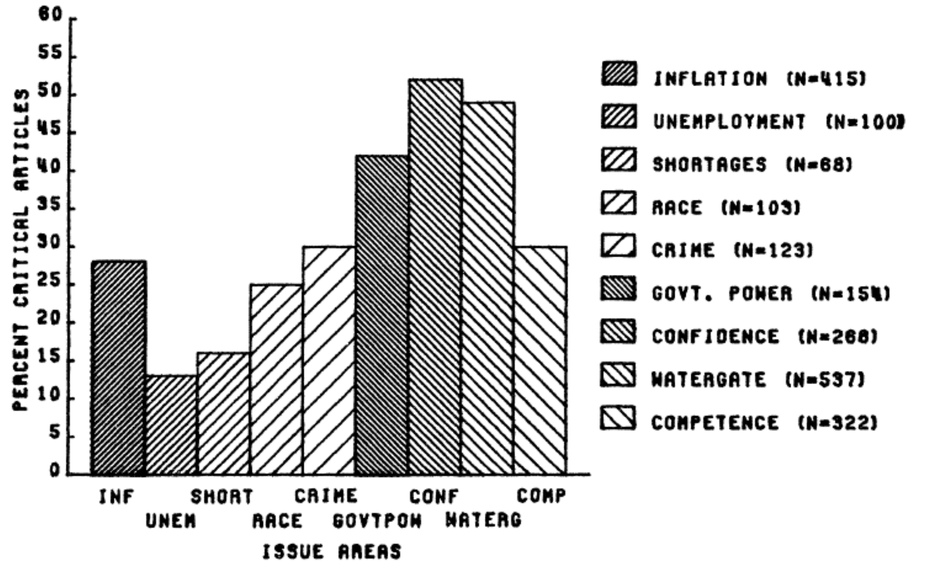
- Chart junk are **non-data ink** and **redundant marks**
  - May confound visual communication of your message!

- Innovative design can promote intrigue and curiosity – chart junk does not

Chart junk

1. Vibrations

Some spatial frequencies can cause vibration and movement that may distract the viewer's experience...

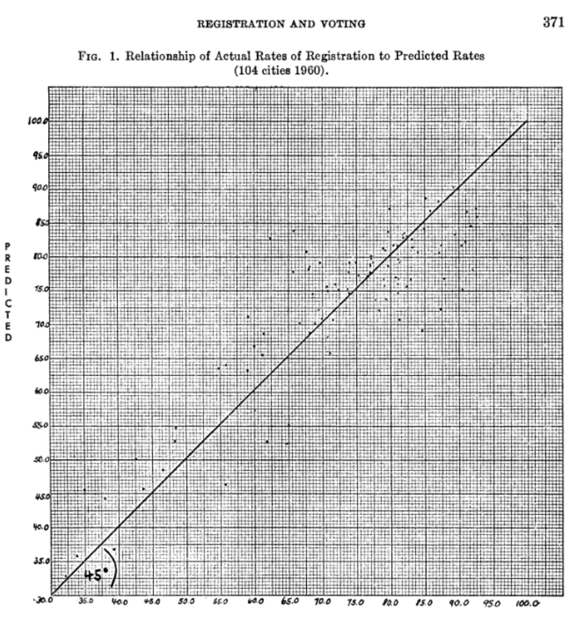


Source: Center for Political Studies Media Content Analysis Study, 1974; available through the University of Michigan, ICPSR. Not to be cited without full bibliographical reference to the present article.

Chart junk

2. Grid overload

Line properties (e.g., width and frequency) that compete with data



3. Ducks

Decorations that take over from quantitative data (i.e. form disastrously overtakes from function)

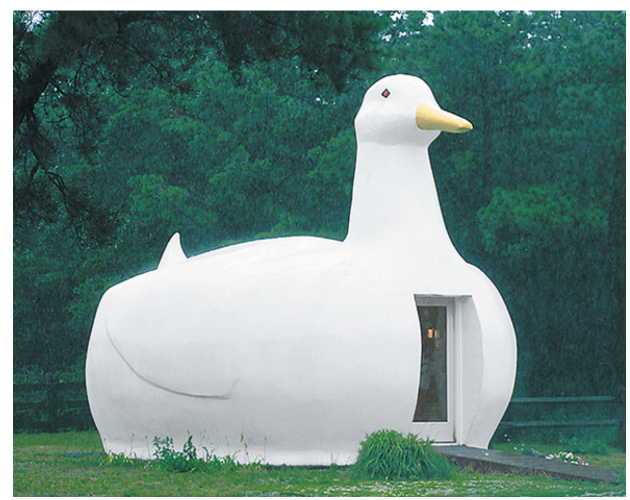


Image: Tufte (2000) Big Duck

Other chart-junk example

Image: Wikimedia commons

Other chart-junk example

DISPLAYING DATA GRAPHICS

SUPPORTING INFORMATION

Often insufficient information is included in data graphics. Supporting information should always include:

1. Caption

Mappings used for visual variables, data source

2. Title

Clear and concise

3. Axis

Always labelled with units, balanced distribution of marks

4. Legend (or key)

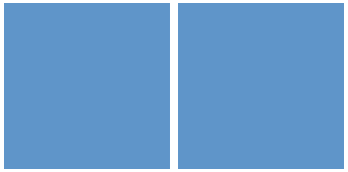
Description of the visual variables used, e.g., symbols

Consistency - multiple graphics

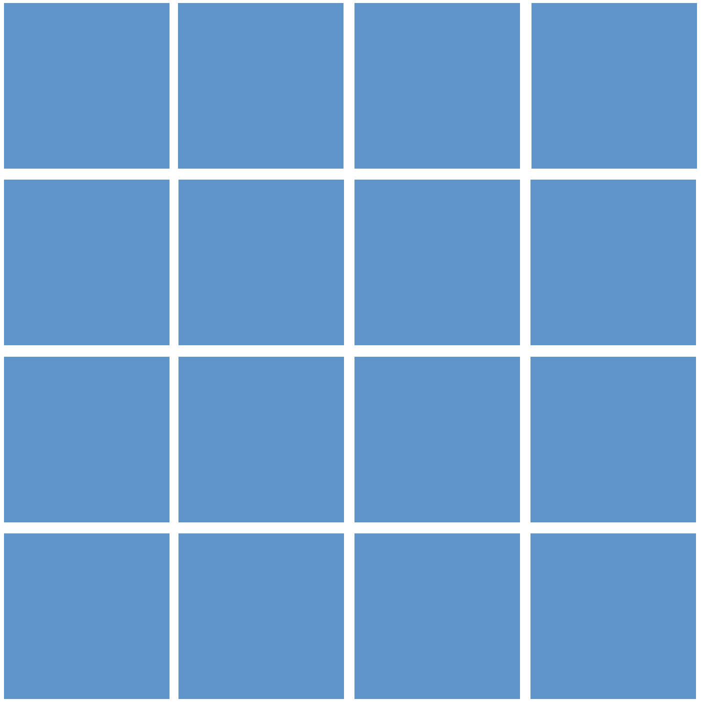
Consistent labelling, gridding and range values between multiple graphics

Use of visual variables must be consistent or otherwise sufficiently described!

One graphic



Two graphics
“two-up”



Matrix
(large data, information space)

IMPORTANT TERMINOLOGY

TAXONOMY

1. Classification of information visualisation techniques

Classification of standard (basic) data:

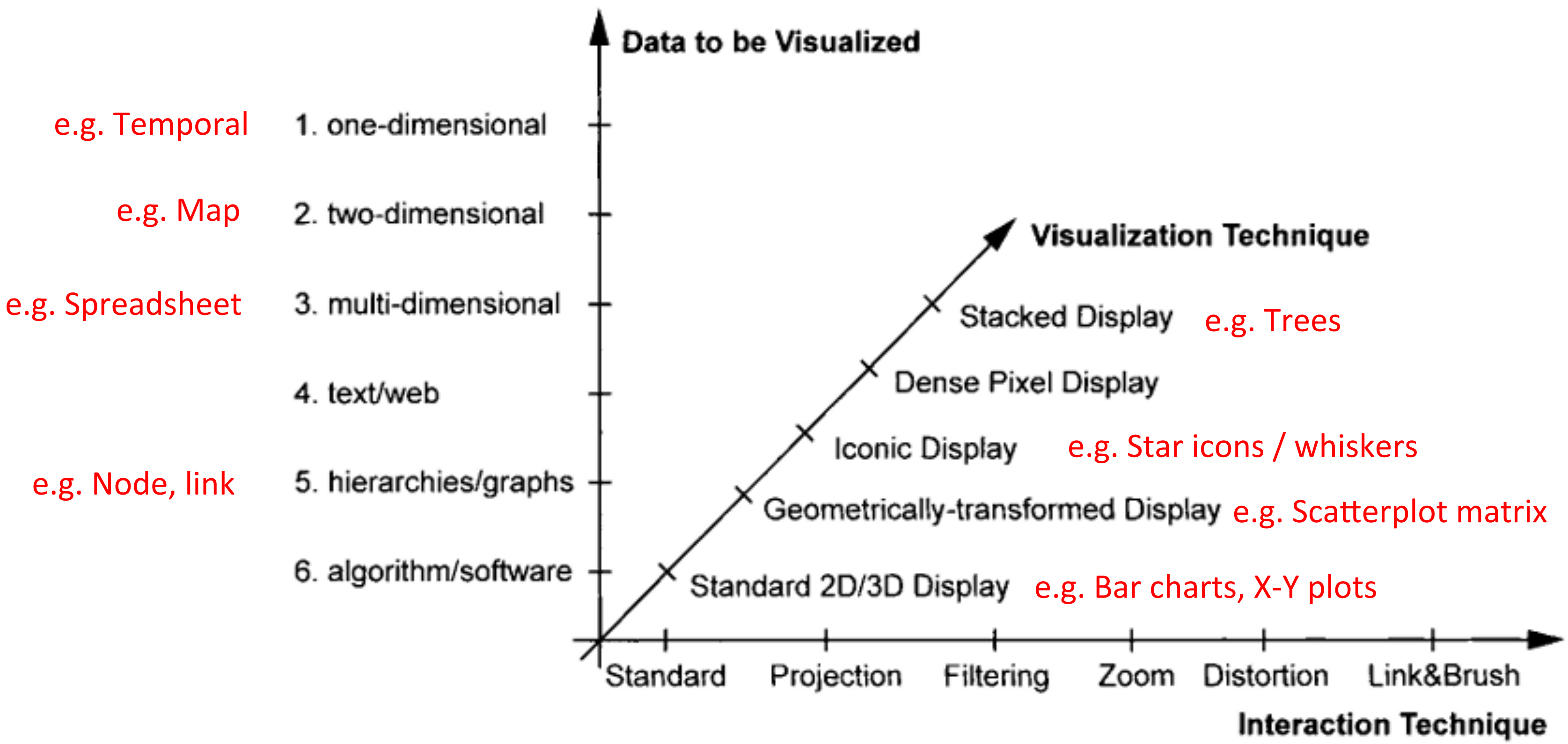
- Numerical
  - Discrete
  - Continuous
- Categorical
  - Nominal
  - Ordinal

Keim's data types, 2002

- One-dimensional (linear)
- Two-dimensional (map)
- Multi-dimensional
- Text
- Hierarchies/graphs
- Algorithms

TAXONOMY

2. Classification of information visualisation techniques



Keim et al., (2005) Scope, Techniques and Opportunities for Geovisualization. In *Exploring Geovisualization, Dykes et al. (ed.) Elsevier*

SALIENCE

- Prominence of a visual object within its surroundings
- Helps the viewer "to quickly rank large amounts of information by importance and thus give attention to that which is most important."
  (Kecskes I, 2012)
- Design task:
  Choice of visual variables must consider trade-offs (e.g., accuracy)

BRUSHING

Scatterplot matrix

Brushing is the process of **highlighting** data

Typically an interactive process to select features and visually link them across plots (e.g., two up, matrix)

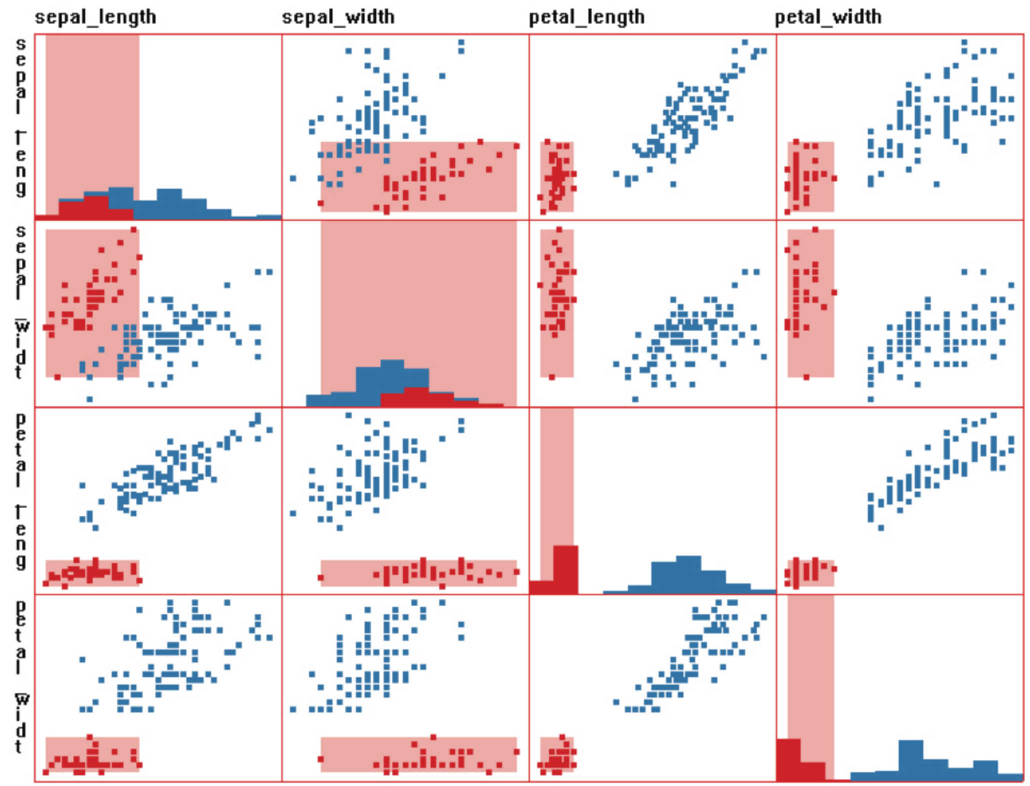


Image: Ward et al. (2010

SOFTWARE SUPPORTING VISUALISATION

- Commercial OTS
  - Tableau V9
  - Qlik
  - Alteryx
  - Echarts3
- Open Source
  - D3
  - Processing V3
  - Cinder

Question:

How does certain software fit within the visualisation pipeline?

ITERATIVE DESIGN PROCESS

- Creating a spatial visualisation is **design** – design is choice
- Data graphics will (almost always) improve with iteration
  - Data-ink
  - Visual variables
- Classic iterative design process

1. Analysis
2. Design
3. Implement
4. Test
5. Evaluate

NEXT LECTURE

- Cartography 1

READING

Agrawala, M., Li, W. and Berthouzoz, F. (2011) Design principles for visual communication. *Communications of the ACM,* 54(4), pages 60-69

Access: http://dl.acm.org.ezp.lib.unimelb.edu.au/citation.cfm?id=1924439