Information Publication for Systems Engineers

(making engineering outputs more accessible)

Follow the presentation on your own screen by visiting:

extuitive.co.uk/ip

This will allow you to follow links and participate in interactive polls

Setting the scene

What are we not talking about?

- Giving presentations
- The fine details of the 'academic' publication process
- How to do techincal writing
- Being an artist
- Data analysis

So what *are* we talking about?

10:00 — 10:25	Scene setting and general concepts
10:25 — 10:40	**Data/Information management**
10:40 — 10:45	Mini-exercise
10:45 — 11:05	**Document presentation**
11:05 — 11:10	Break
11:10 — 11:25	Exercise
11:25 — 12:00	**Graphic presentation**
12:00 — 13:00	Lunch
13:00 — 13:20	Group exercise
13:20 — 13:40	**Publication**
13:40 — 14:00	**Advanced techniques**

Why do we care…

…as engineers?

…as Systems Engineers?

Pedantry, or professionalism?

Who am I?

Dr John Welford MEng PhD CEng IET MINCOSE

WSP New Zealand, Technical Principal Systems Engineer

(*Not* a data visualisation expert or a typographer)
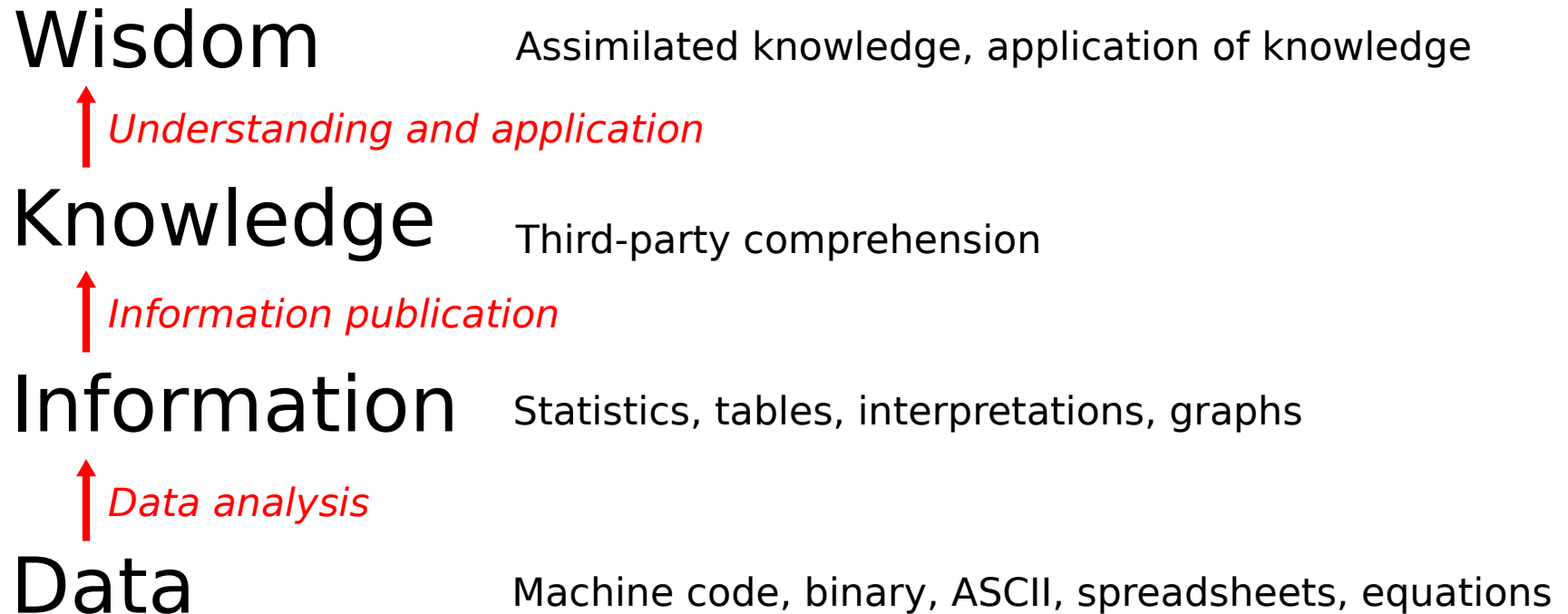
Who are you?

- Name
- Job/role
- Interest/experience in the area

General concepts

Data and information

The DIKW model:

Wisdom   Assimilated knowledge, application of knowledge

↑ *Understanding and application*

Knowledge   Third-party comprehension

↑ *Information publication*

Information   Statistics, tables, interpretations, graphs

↑ *Data analysis*

Data   Machine code, binary, ASCII, spreadsheets, equations

Think about your audience

- What do they already know? What will be familiar to them?
- What do they need to know?
- What is the inferential distance between them and you?
- What special needs might they have?

Presentation of abstraction

As engineers we are *often* working with an abstraction of the real system.

When we publish information we are *always* presenting an abstraction of the real system.

> *It seems that perfection is attained not when there is nothing more to add, but when there is nothing more to remove.*
>
> — *Antoine de Saint Exupéry*

It is therefore up to us to choose what to emphasize, for example…

(credit xkcd.com)

Separating content from presentation

Familiar to web-authors: HTML = content (+ structure),
CSS = presentation.

In DIKW terms: *information* = content (+ structure),
publication = presentation.

Ideally — first develop the content, then later develop the presentation.
Practically — development is often in parallel; however, content should always be prioritised.

Some tools provide a clear separation between content and presentation, others are more WYSIWYG. In either case, it pays to at least conceptually make the separation.

Dangers of conflating content with presentation

1. Distraction from the process of working with information
2. Reduced portability and compactness
3. Lack of proper information structure

(This presentation content is written in Markdown, the formatting is in CSS)

Additional considerations

Be conscious & intentional   Every aspect of presentation represents a decision. Every decision should be justifiable.

Be consistent   The same decision should have the same outcome each time it's made. There should be uniformity in the resulting publication.

Beauty vs. practicality   Ideally both! But (for engineering): $practicality >> beauty$.

Science, art, opinion

Be aware that most advice on the topic of information publication falls into one of three categories (including this tutorial):

1. Science (researched and peer-reviewed)
2. Art (general expert concensus)
3. Opinion (my own)

References

References cited or linked throughout, but top reference texts are:

The Elements of Typographic Style — Robert Bringhurst

Visualization Analysis & Design — Tamara Munzer

Show Me the Numbers: Designing Tables and Graphs to Enlighten — Stephen Few

Other authors to read or follow:

Edward Tufte, Naomi Robbins, Mike Bostock, Bret Victor

Data/information management

Examples of data from your own job

Fill in the questionnaire on your screen with an example of one source of data that you might utilise in your job.

Keep it brief, but submit multiple answers if you like.

The results will appear to everyone on the next slide in…

Your example data

Data structure

Likely to be out of your control:

- Results of a questionnaire
- Export from a simulation, model, or design tool
- Logs from monitoring equipment
- Scraping websites
- Verbal or written input from experts
- Previous work and general literature
- Experience and engineering judgement

The process of producing *information* from the *data* will involve some degree of analysis, but it is also where sensible choices can be made about the *information structure*.

Variations in data structure

Data can have many different structures. Part of data analysis is cleaning and tidying the data.

Consider...

	Treatment A	Treatment B
John Smith	—	2
Jane Doe	16	11
Mary Johnson	3	1

Transposed as…

	John Smith	Jane Doe	Mary Johnson
Treatment A	—	16	3
Treatment B	2	11	1

Or tidied as…

Name	Treatment	Result
John Smith	A	—
Jane Doe	A	16
Mary Johnson	A	13
John Smith	B	2
Jane Doe	B	11
Mary Johnson	B	1

Dont be afraid to change your data structure to support analysis and information presentation.

Information structure – focus on content

The output of analysis should yield *information*. At this stage it can be very tempting to start *presenting* the information, indeed you may wish to start considering the final published form prior to the information being complete.

However, it requires discipline to keep the concept of *structuring* the *content* separate from *presenting* it.

For example, setting up the structure of your report, is separate from deciding which levels of heading are going to be in bold.

Configuration control

Both data and information should be under some form of configuration management, ideally supporting:

- Versioning
- Change control
- Baselines
- Access control
- Backup/recovery

Also consider maintaining an auditable trail of how the *information* was generated from the *data*.

Data/Information tools

Spreadsheets: Excel, Calc, Numbers, Google Sheets

Databases: Access, DOORS, SQL, MBSE tools

Formats: CSV, JSON, XML

Languages: Matlab, Python, R, LaTeX, Markdown

Mini-exercise

You have 3 minutes to skim and rate a paper based on:

'How well you understood the paper'

'How useful you think it is'

Results

Document presentation

Assume the *information* we wish to publish is best presented in the form of a *document*.

The definition of 'best' here will depend on your audience and what the information is – as discussed previously.

We'll also assume that the information is already developed in terms of both *content* and *structure*. The format that this is captured in might vary depending on the tool that we choose to use.

For reference let's have a look at some examples of content and structure…

Content and structure – LibreOffice Writer

NB: Structure is less explicit here, as it is a WYSIWYG tool.

Content and structure – Markdown

```
Example document
================


Demonstrating the information *content* for a document, and a bunch of
different aspects of document *structure* (the funny-looking parts).

## Lists

Could be:
* Enumerated
* Unordered
  * Sub-lists

## Links
```

Content and structure – HTML

```html
<h1 id="example-document">Example document</h1>
<p>Demonstrating the information <em>content</em> for a document, and a
bunch of different aspects of document <em>structure</em> (the funny-looking
parts).</p>

<h2 id="lists">Lists</h2>
<p>Could be:</p>
<ul>
  <li>Enumerated</li>
  <li>Unordered
    <ul>
      <li>Sub-lists</li>
    </ul>
  </li>
</ul>
```

Content and structure – LaTeX

```
\section{Example document}\label{example-document}

Demonstrating the information \emph{content} for a document, and a bunch of
different aspects of document \emph{structure} (the funny-looking parts).

\subsection{Lists}\label{lists}
Could be:
\begin{itemize}
  \item Enumerated
  \item Unordered
  \begin{itemize}
    \item Sub-lists
  \end{itemize}
\end{itemize}
```

Presenting the document

In all the preceeding examples the content was the same, although the syntax used to provide the structure was different. However, presentation of that content depends on the *styling* that is added based on the structure.

Styling will also have it's own syntax that is tool/language specific. The capability of different tools to provide styling also varies significantly.

Next we will run through a bunch of different document presentation areas where we should make concious decisions about styling, this is better known as *typography*.

Typeface

A general 'font family', typefaces include many different fonts with variations in size and emphasis (bold, italic, etc).

Typefaces (and fonts) may be classified as either sans-serif, or serif.

Serifed fonts are considered easier to read in print than sans-serif. However, the science on this appears to be inconclusive.

Conversely sans-serif are sometimes preferred for on screen reading, as they scale better at low resolutions.

Interesting typefaces

Some typefaces have been designed to assist dyslexic readers, however their efficacy is disputed.

Sans Forgetica is designed to be intentionally difficult to read, as this prompts your brain to engage in deeper processing.

Headings

Typically a document may have many levels of heading.

Too many levels and you will lose the reader!

ALL CAPS, 'Title Case', and 'Sentence case' can be used at different levels of heading, along with changes in size, colour and font.

Normal text – size and spacing

Size: choose for legibility and to match the page (see later).

Letter spacing: best to stick with the default!

Setting: Either flush-left ragged-right, or justified. Justified text is achieved by the tool modifying the inter-word spacing, so choose justifed text only when the line is long enough — hyphenation may still be necessary to avoid sloppy spacing (if your tool supports this).

Advanced tools may support microtypography, which subtly adjusts other aspects of the text to improve readability and appearance.

Justified text is considered to be a problem for dyslexics, due to the uneven spacing and distracting 'rivers' of white space.

Kerning:
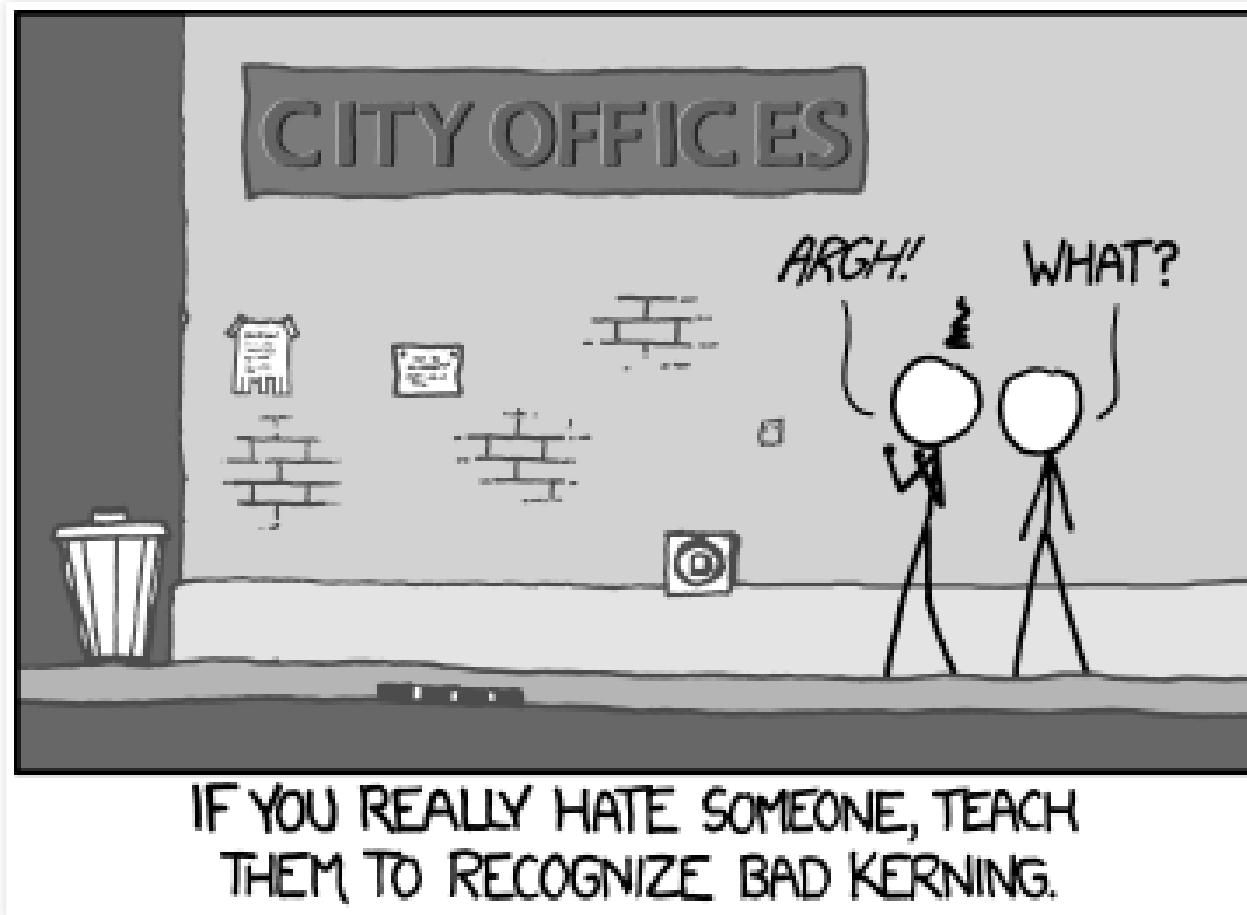
(credit xkcd.com)

Spaces between sentences

If you answered 'one space' then you're doing it right!

Double spacing is an artefact of victorian typewriter usage and is no longer relevant.

Paragraphs

Provide a pause in reading, and may be shown by either an initial indent or a slight space between blocks of text.

Indents are more common in printed literature, whilst spacing is more prevalent on the web (where there are less space constraints).

If you're working in a WYSIWYG tool then you shouldn't be inserting an extra carriage return between paragraphs. This is mixing up content with presentation.

Weighting and emphasis

Change one parameter at a time.
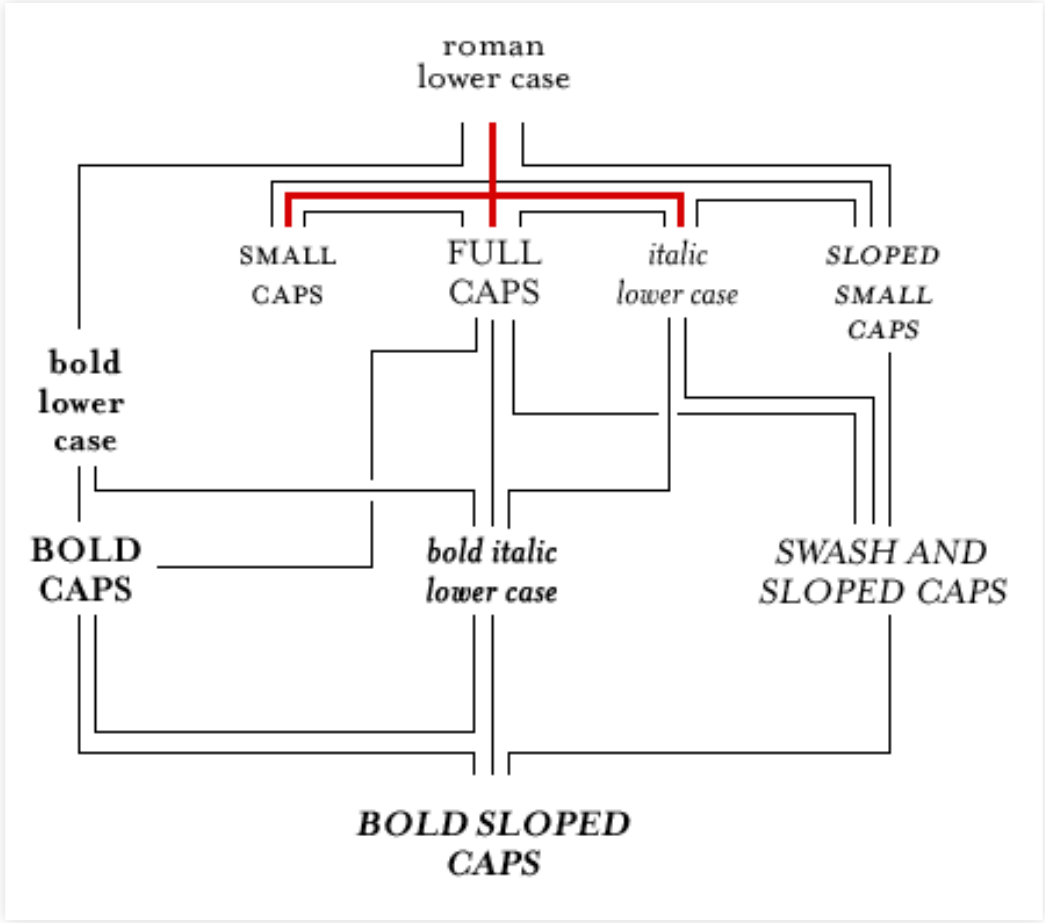
A roadmap of fonts in a family of type, originally from Bringhurst, reproduced by Boulton.

Page layout – lines

Line length: 66 characters is considered ideal, but anything 45 to 75 characters is ok (including punctuation and spaces). Longer might be ok for discontinuous texts (e.g. bibliographies).

Line spacing: 'leading' is usually slightly more than character height, giving a small gap between lines. Sometimes much larger spacing (1.5 or double-space) may be requested to allow for handwritten review comments.

Page layout – textblock

Margins: textblock width should be defined to achieve the right line length based on the typeface size and page size. Textblock height depends only on what size margins you leave — don't be stingy on the margins or your page will look ugly! Also worth considering are binding and on-screen reading.

Headers and **footers**: may carry information about the section and the page, or about the publication itself. The latter only seems necessary if there is a danger that pages may be reproduced out-of-context.

Lists

Avoid over-punctuating lists.

Be consistent in list structure and punctuation.

Enumerate lists when items have an order, or where they need to be referenced later (although be aware that this may imply a priority).

Special characters

Use non-breaking spaces where words should not be separated.

Dashes – come in various lengths, choose the correct one for your purpose:

- The hyphen - is used to join words.
- The en dash – is used in ranges.
- En dashes or em dashes — are used in a similar way to brackets.
- The minus sign — is a separate character altogether.

NB: These are content not presentation!

Notation, quantities and units

Try to stick with standard choices for symbols representing variables (but don't forget to also define them!).

Consider using a different typeface or italic font for variables.

Use SI units as far as possible.

Take care when typesetting numbers and units, some tips:

- Insert a small non-breaking space between numbers and units
- Ensure you have the SI prefix on your units correct (e.g. $10\,Kg$ is *not* ten kilograms, it is ten Kelvin grams!)
- Consider raising units to a negative power in place of the divide symbol

Units in tables and figures

In tables and figures consider using a slash to denote the units. This is recommended by the BIPM as the correct method of expressing values for multiple quantities.
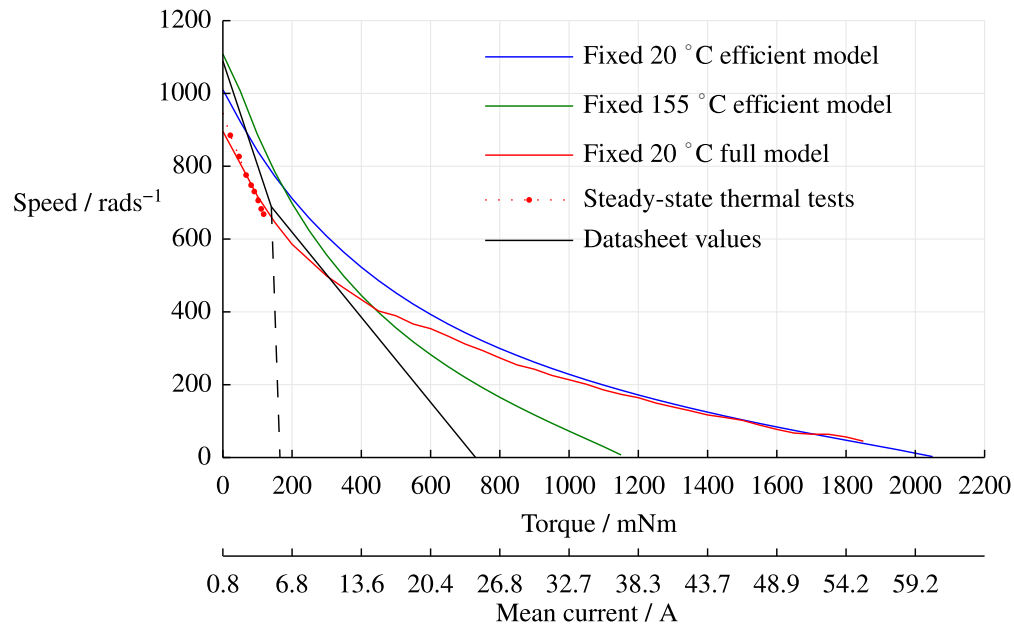
Figure 5.1: Speed-torque curve for fixed temperature simulations, datasheet and steady-state test values for the MMT motor at 24 V

Equations

Equations are usually set centred, with a reference number on the right-hand side of the page.

$$\theta_b = \max\left(\hat{\theta}_b, \min\left(-\hat{\theta}_b, \int \omega_b \, \mathrm{d}t\right)\right) \qquad (3.32)$$

Use a proper equation editor!

Abbreviations (acronyms & initialisms)

Very rarely do readers wish that an author had used *more* abbreviations!

Define abbreviations both the first time they are used, and within an abbreviations list.

Notes and references

Notes: use end-notes or side-notes for digressions that do not belong in the main text. References are a subset of notes.

References: Use proper reference management software; autogenerate bibliographies.

Consider hyperlinks if your document will be presented in digital format.

Don't cite fake references!

Cross-references

Where appropriate, cross-references within the document should be made, including:

- Chapters and sections
- Figures
- Tables
- Equations
- Pages
- Above/below

Always use your tool's cross-referencing functionality for these.

Figures

Figures should appear as soon as possible after, but not before, they are referenced in the text.

Text in figures should be horizontal (or at least oblique).

Text should be legible.

All figures should have captions below them.

See the Graphic Presentation section for more.

Tables

Tables should appear as soon as possible after, but not before, they are referenced in the text.

Numbers in columns should be aligned on the decimal.

Heavy gridlines are not necessary; a few horizontal rules are ok, but white space is usually better.

Quoted numeric precision should reflect accuracy of measurement.

A tables should have captions above them.

Making decisions on style

Start with prescribed document templates.

Check whether your organisation has a house style / style guide.

Choose another organisation's manual of style:

- The Chicago Manual of Style
- Wikipedia Manual of Style
- The Guardian style guide
- BBC News style guide

Document tools

Word processing: Word, Writer, LyX, Pages, Google Docs

Typesetting: InDesign, Scribus, Publisher, LaTeX

Content editing (any text editor): Notepad, Notepad++, Emacs, vim

For converting content and structure between tools try Pandoc.

Break

Five minutes only please!

Exercise

Review of the INCOSE system definition paper
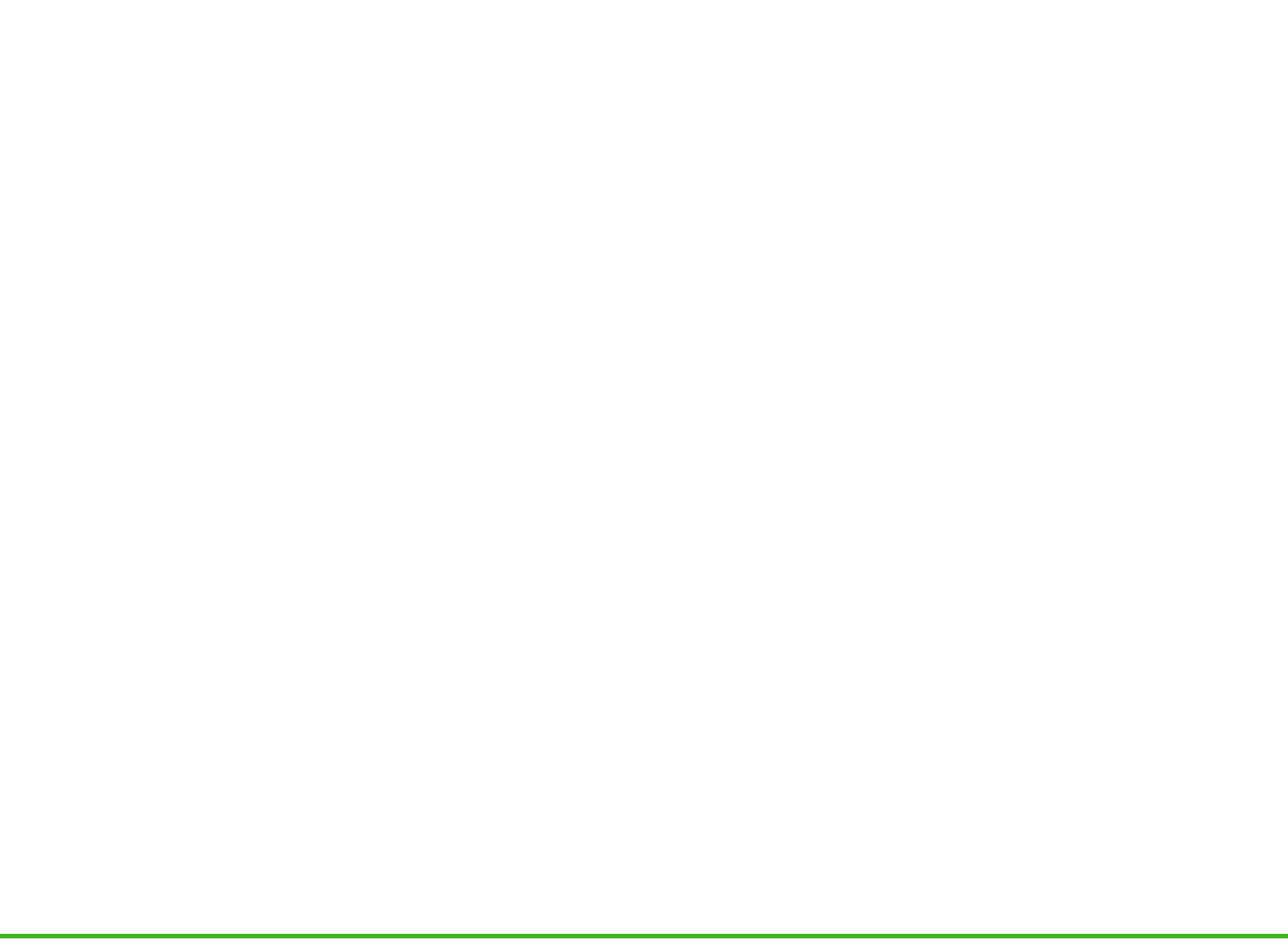
Graphic presentation

Visualizing MBTA Data

An interactive exploration of Boston's subway system

Mike Barry and Brian Card - June 10, 2014

Monday 2/3 7:06 am
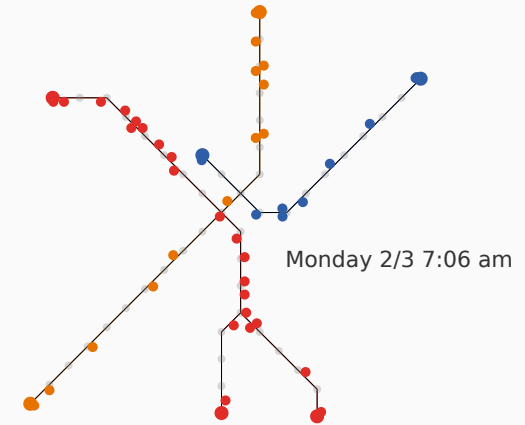
Boston's Massachusetts Bay Transit Authority (MBTA) operates the 4th busiest subway system in the U.S. after New York, Washington, and Chicago. If you live in or around the city you have probably ridden on it. The MBTA recently began publishing substantial amount of subway data through its public APIs. They provide the full schedule in General Transit Feed Specification (GTFS) format which powers Google's transit directions. They also publish realtime train locations for the Red, Orange, Blue, and Green lines. The following visualizations use data captured from these feeds for the entire month of February, 2014. Green Line data became available in October, 2014 so is not shown here. Also, working with the MBTA, we were able to acquire per-minute entry and exit counts at each station measured at the turnstiles used for payment.

We attempt to present this information to help people in Boston better understand the trains, how people use the trains, and how the people and trains interact with each other.

The Trains

In a typical weekday, trains make approximately 1150 trips on the red, orange, and blue lines starting at 5AM and continuing through 1AM the

Average Number of Trips per Day

Weekdays   Saturdays   Sundays

Is a diagram worth a thousand words?

Sometimes no!

Beacon Use Among Fatalities 2000–2012—United States

No Beacon 50%   Beacon 50%

Figure I-2. In the past 10 years in the United States about half of avalanche victims were not carrying basic rescue gear
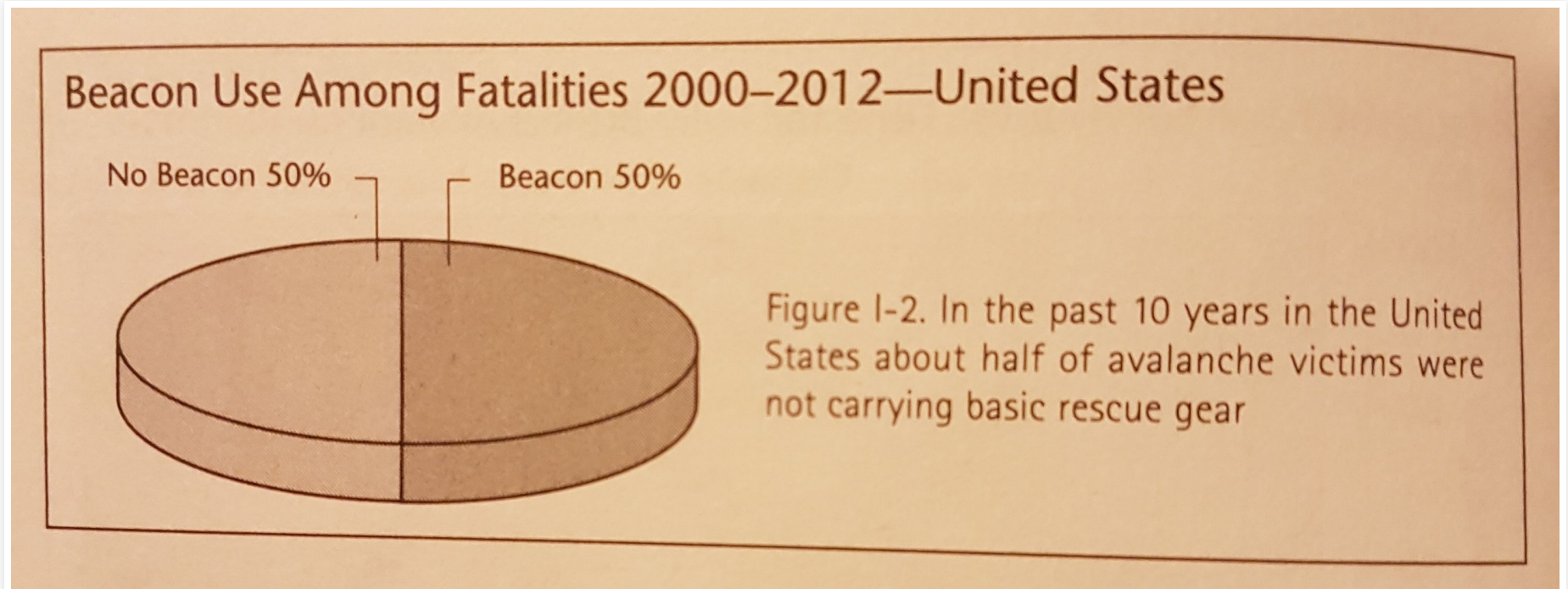
But sometimes yes!

Always consider text before tables, and tables before graphics.

Anscombes Quartet – data

| Set 1		Set 2		Set 3		Set 4	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

Anscombes Quartet – statistics

All sets have the same:

- Mean ($\bar{x} = 9.00, \bar{y} = 7.50$)
- Variance ($\sigma_x^2 = 10.00, \sigma_y^2 = 3.75$)
- Correlation ($\rho_{x,y} = 0.816$)

Anscombes Quartet – graphics
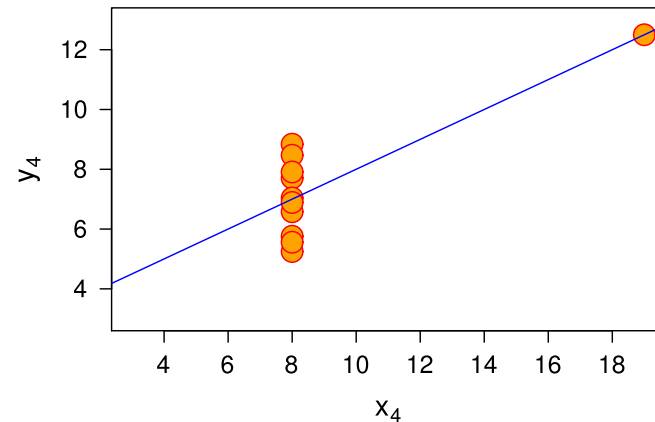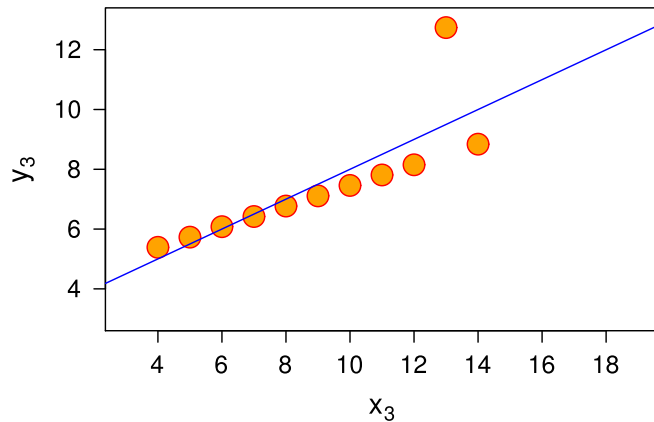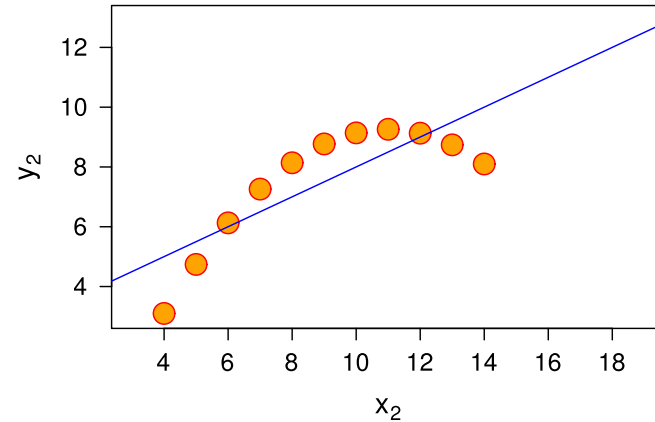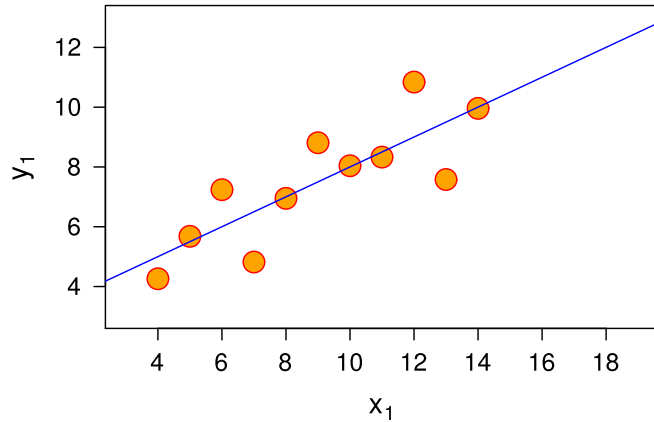
For another fun example check out the Anscombosaurus!

Two motivations for visualisation

Discovery:

- To explore and understand a dataset or problem space
- As a step towards automation (debugging)
- To aid human-in-the-loop processes
- *Working with* **data**

Presentation:

- For explanation
- *Presenting* **information**

The graphic may need to display the data, but the message should be the information

What is the most popular visualisation in Systems Engineering?
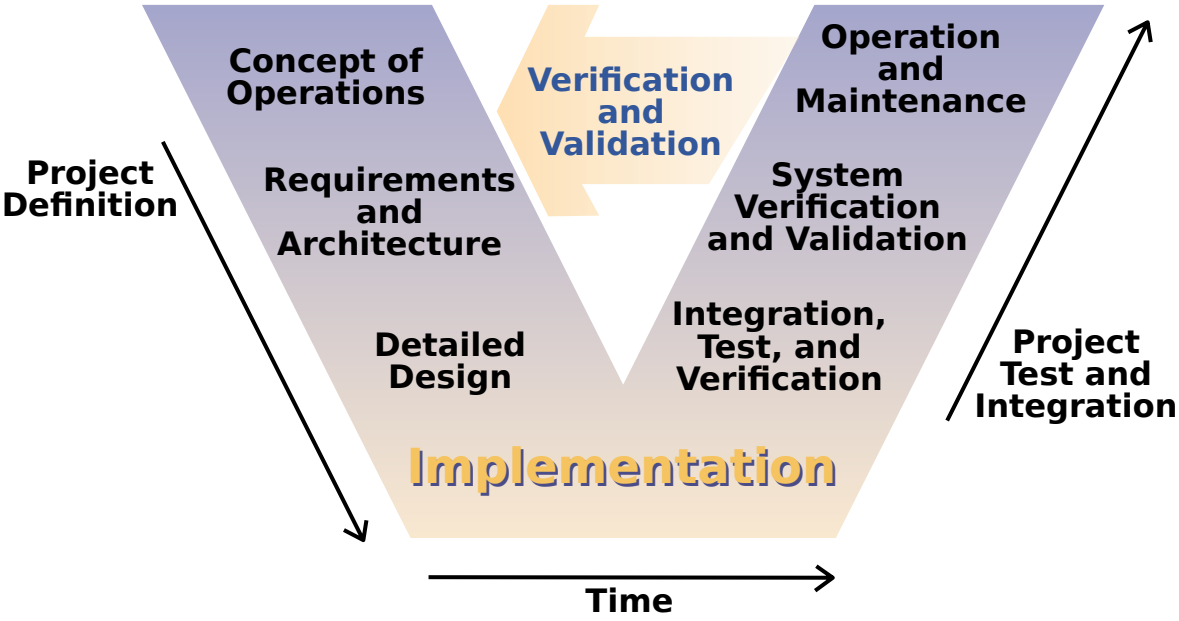
But why is it a 'V'?

Dataset types

Tables

Attributes (columns)

Items (rows)

Cell containing value

Networks

Link

Node (item)

Spatial (Geometry)

Position

Spatial (Fields)

Grid of positions

Cell

Attributes (columns)

Value in cell
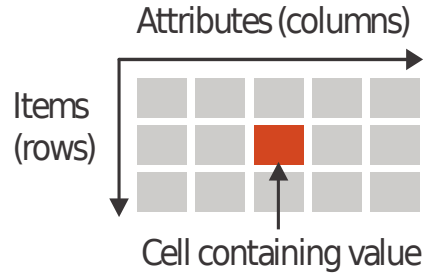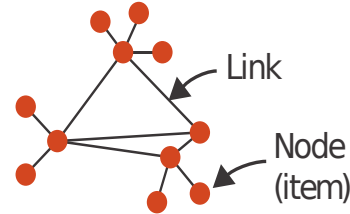
Multidimensional Table

Key 1

Key 2

Value in cell
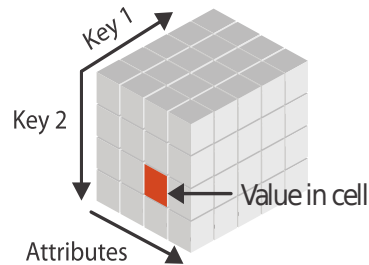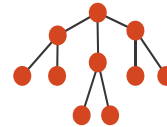
Attributes

Trees
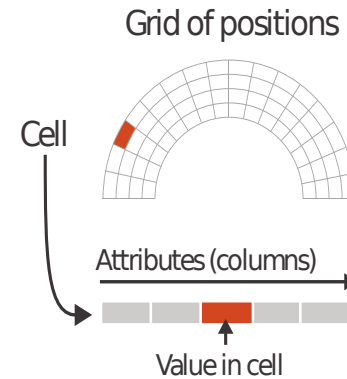
Items

Attributes

Items (nodes)

Links

Attributes

Grids/Items

Positions

Attributes

Attributes

Ordinal

Categorical

Quantitative
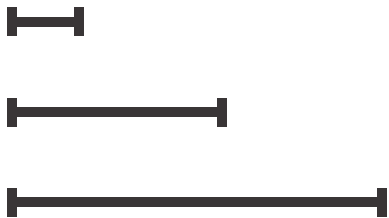
Encoding channels

Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Most

Effectiveness

Least

Same

Same

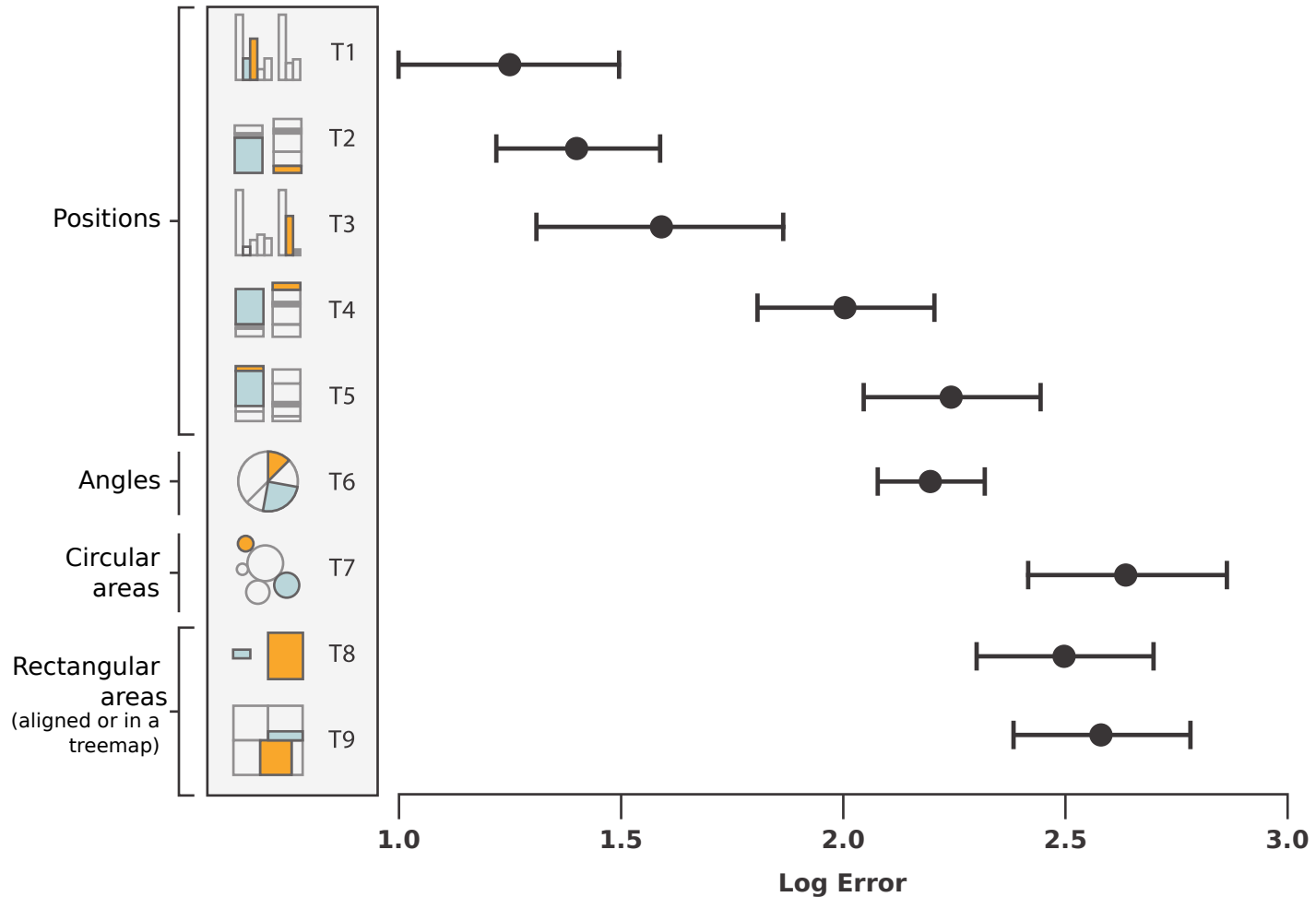## Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

Why are some channels more effective?
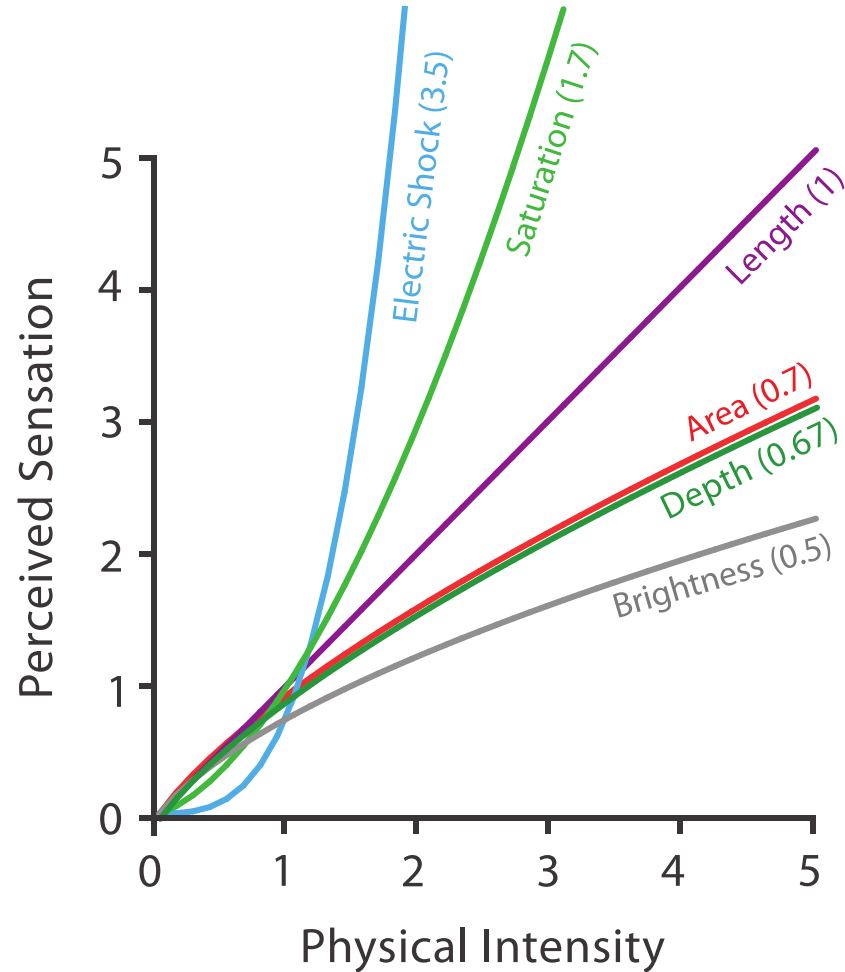
Stevens Psychophysical Power Law: $S = I^n$

Crowdsourced channel testing

Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design

Keys and values

Keys: Independent attributes (categorical or ordinal)

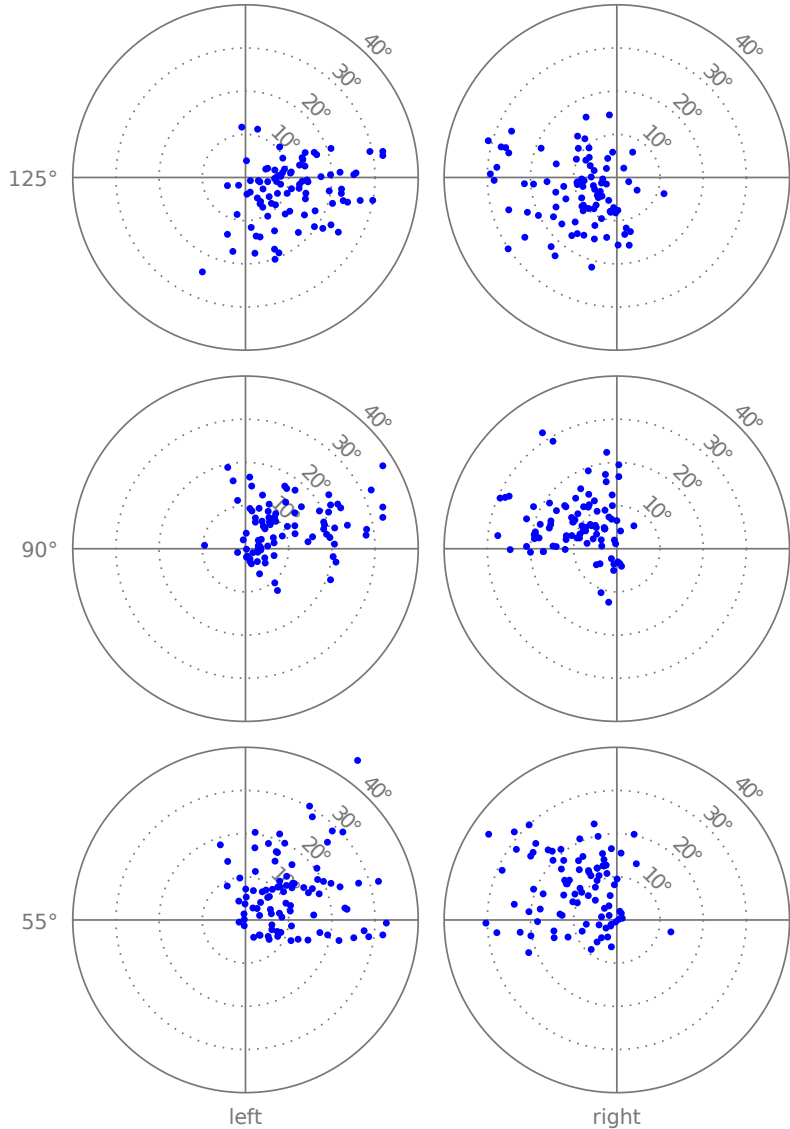**Values**: Dependent attributes (categorical, ordinal or quantitative)

Zero keys: Scatterplot

One key: Bar chart, Line chart, Dot charts, Coloured scatterplot

Two keys: Heatmap, Stacked bar chart, Coloured bar/line/dot charts
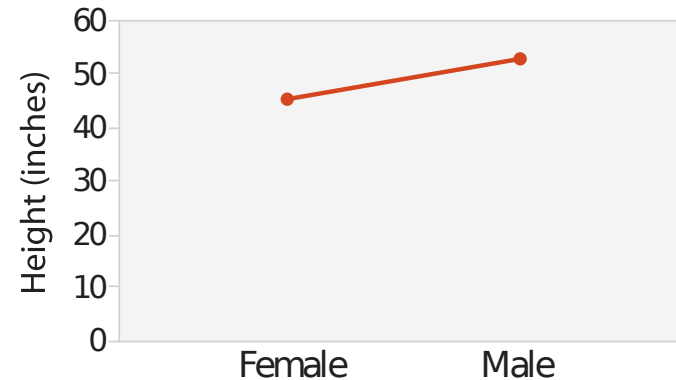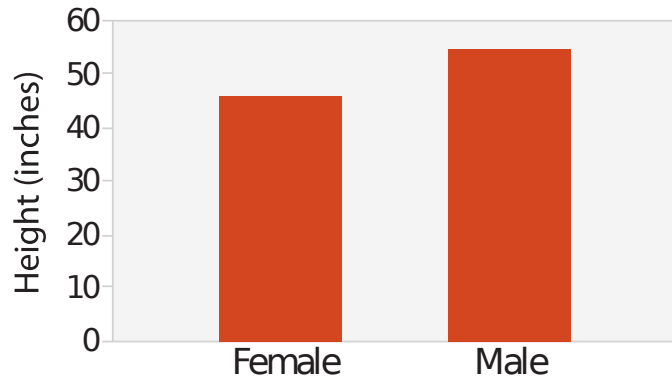
Three or more keys: 'Small multiples' of the above
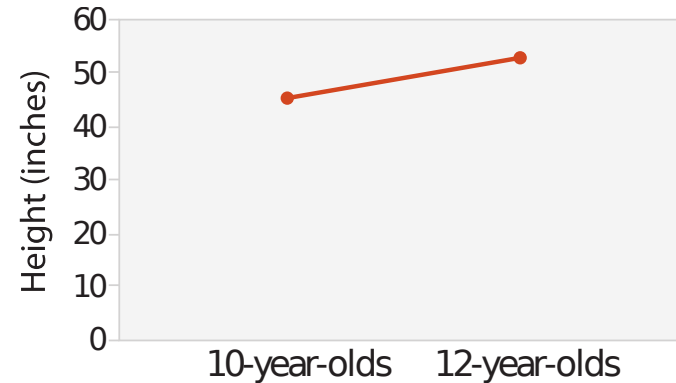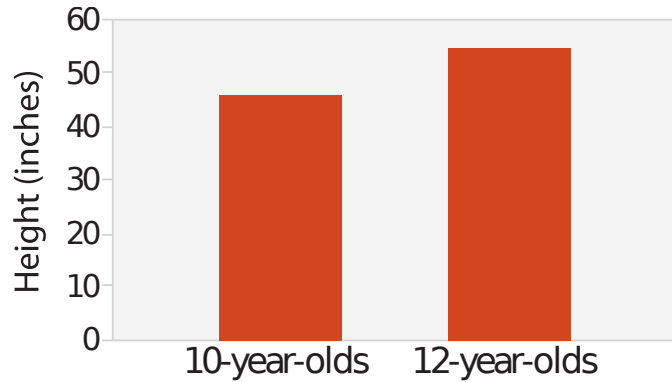
Small multiples

Categorical keys

Choose a sensible order for the key.

Don't connect points between categorical data.

Networks and trees

Both are highly relevant to Systems Engineering.

Networks:

- Visual modelling languages (SysML, LML, OPM, Simulink)
- Interfaces
- Stakeholder relationships

Trees:

- Work Breakdown Structures
- System Breakdown Structures
- Filesystems

Node–Link diagrams (networks and trees)

Intuitive, but limited in network size (consider interactivity or separate views).

Link density = number of links per node. Trees have link density of one. Maximum link density for effectiveness is around 3 or 4.

Consider layout:

- Automatic (e.g. force-directed) or manual
- Reading direction
- Information density (see later)
- Minimise edge crossings
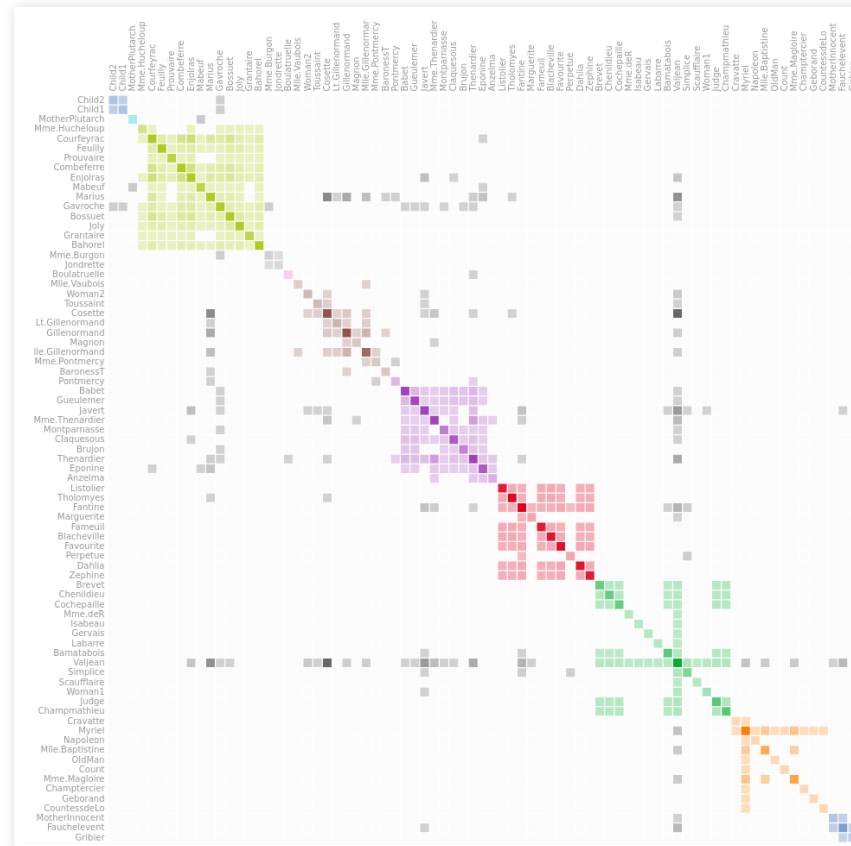- Minimise ratio between longest and shortest edges

Credit: Mike Bostock
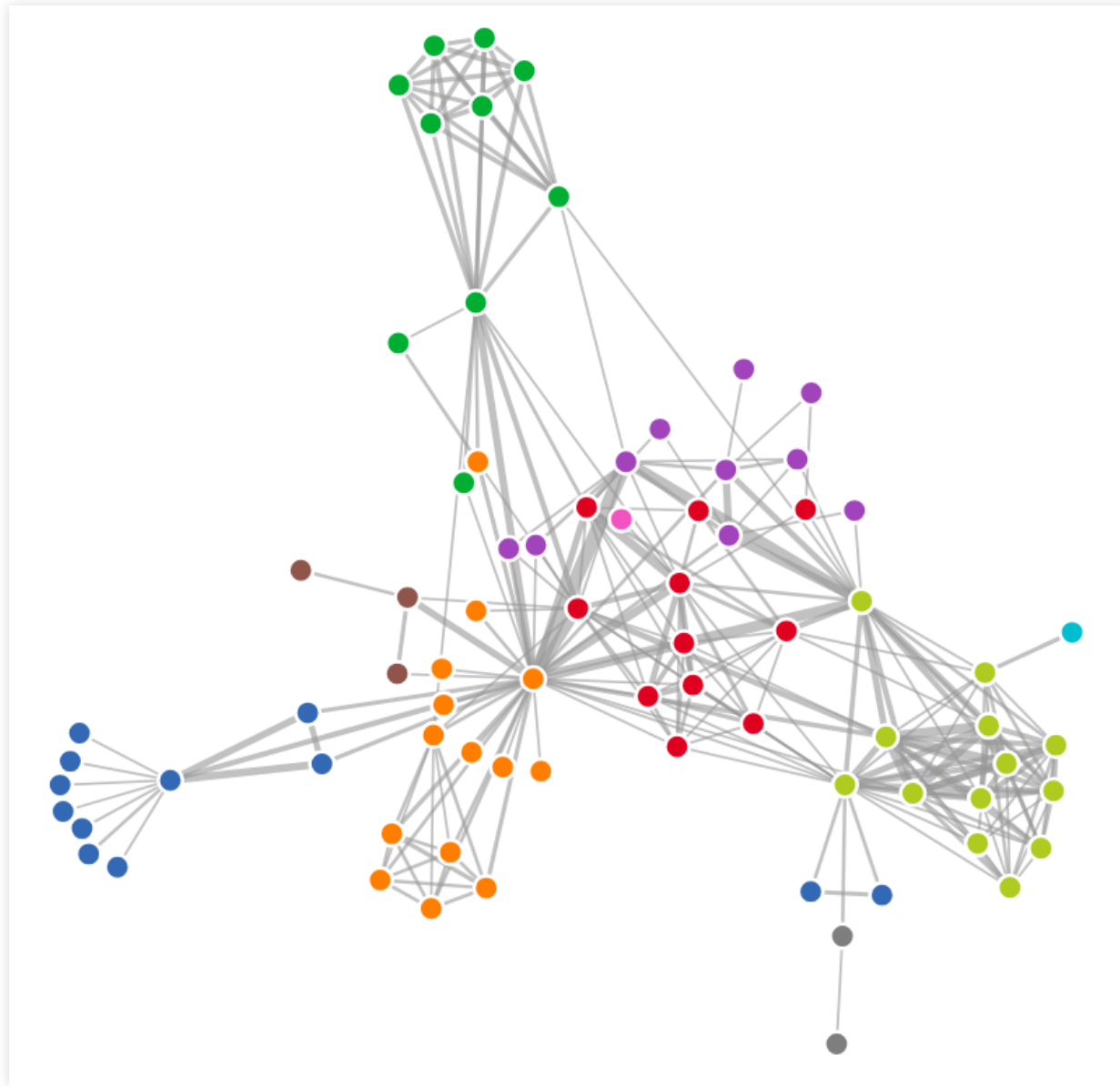
Credit: Mike Bostock

Matrix views (networks and trees)

In graphs of more than 20 vertices, matrix views typically perform better; the exception being where path finding is important.

Credit: Brian Staats

Enclosure (trees)

Show hierarchical structure through containment rather than connection.

Credit: Mike Bostock

Credit: Mike Bostock

Spatial data

Where data is spatial, it is usually beneficial to present it spatially.

Geometry: Chloropeth maps

Fields: Isocontours, Vector fields

Great Western Route Modernisation Programme
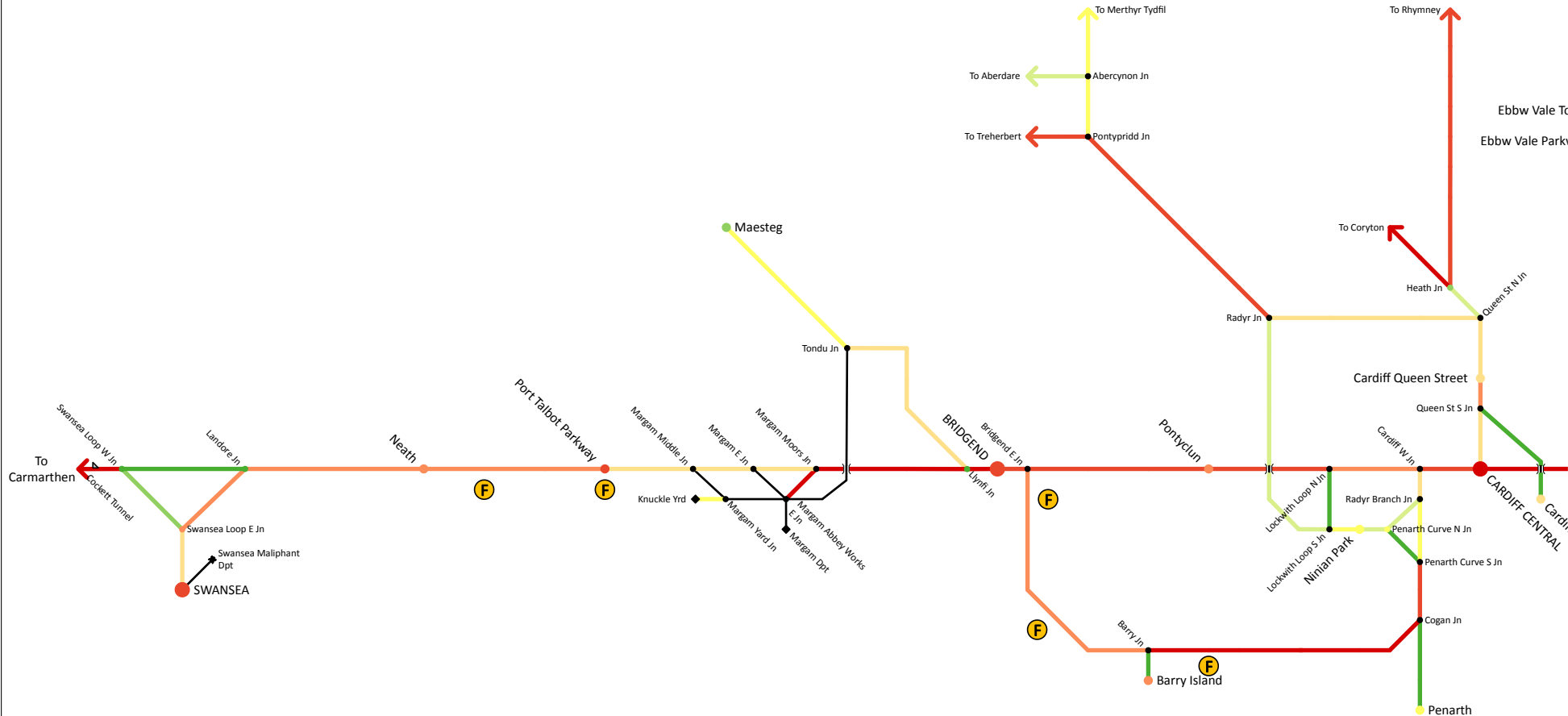
GW ISI Architecture: Interface Map - Wales

Performance View   Incidents   Combined Delays   Source: Standard   Hotspots: Hide

Prepared by WSP | Parsons Brinckerhoff for Network Rail

Instance Date: 01/01/2014

2014 2015 2016 2017 2018 2019 2020 2021

Key

No. of incidents

| 0 | 1+ | 44+ | 66+ | 110+ | 221+ | 353+ | 662+ | 1214+ |

To Merthyr Tydfil
To Aberdare
Abercynon Jn
To Rhymney
Ebbw Vale To
Ebbw Vale Parkw
To Treherbert
Pontypridd Jn
To Coryton
Heath Jn
Queen St N Jn
Radyr Jn
Maesteg
Cardiff Queen Street
Queen St S Jn
Tondu Jn
BRIDGEND
Bridgend E Jn
Pontyclun
Cardiff W Jn
Port Talbot Parkway
Margam Middle Jn
Margam E Jn
Margam Moors Jn
Radyr Branch Jn
Swansea Loop W Jn
Landore Jn
Neath
Llynfi Jn
CARDIFF CENTRAL
Cardi
To Carmarthen
Cockett Tunnel
Knuckle Yrd
Margam Yard Jn
Margam Abbey Works
Margam E Jn
Margam Dpt
Lockwith Loop N Jn
Lockwith Loop S Jn
Ninian Park
Penarth Curve N Jn
Penarth Curve S Jn
Swansea Loop E Jn
Swansea Maliphant Dpt
SWANSEA
Cogan Jn
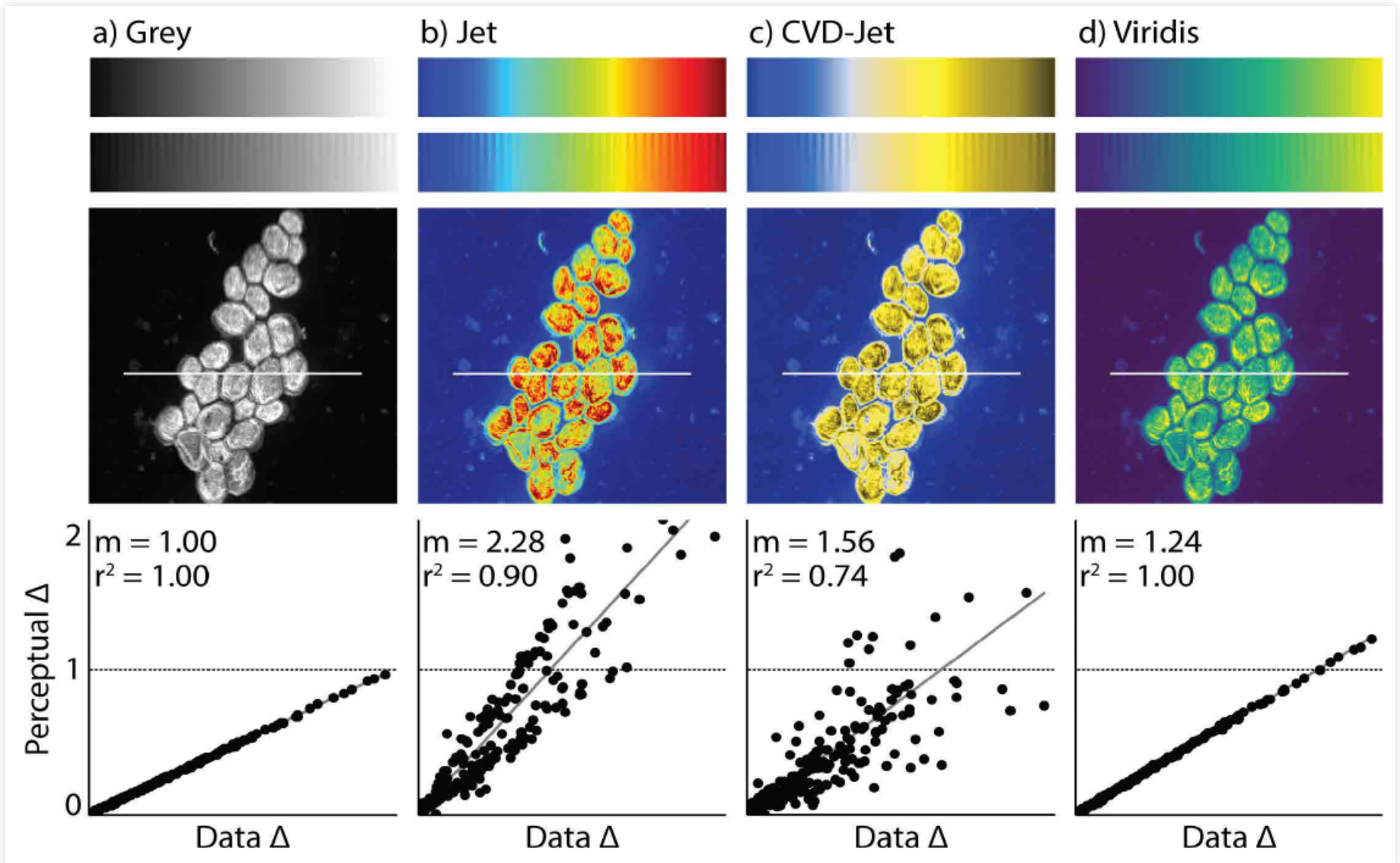Barry Jn
Barry Island
Penarth

Colour

Categorical: There are a limited number of discriminable bins (max 6–12), ensure that there is clear separation between them. Use variations in colour hue.

Ordered: Colour scales should be *perceptually* linear. Use variations in colour luminance, saturation and/or hue.

Rainbow colourmaps (e.g. 'Jet') are a poor default, as they are perceptually unordered and nonlinear. Try Viridis or Cividis.

Consider colour-blind users (Cividis is better).

Use online tools such as colorbrewer to choose a palette.

a) Grey

b) Jet

c) CVD-Jet

d) Viridis

Perceptual Δ

a) m = 1.00, r² = 1.00

b) m = 2.28, r² = 0.90

c) m = 1.56, r² = 0.74

d) m = 1.24, r² = 1.00

Data Δ

Credit: Optimizing colormaps with consideration for color vision deficiency to enable accurate interpretation of scientific data, Jamie R. Nuñez, Christopher R. Anderton, Ryan S. Renslow

Or, taken from a tweet by Matthias Bussonnier:
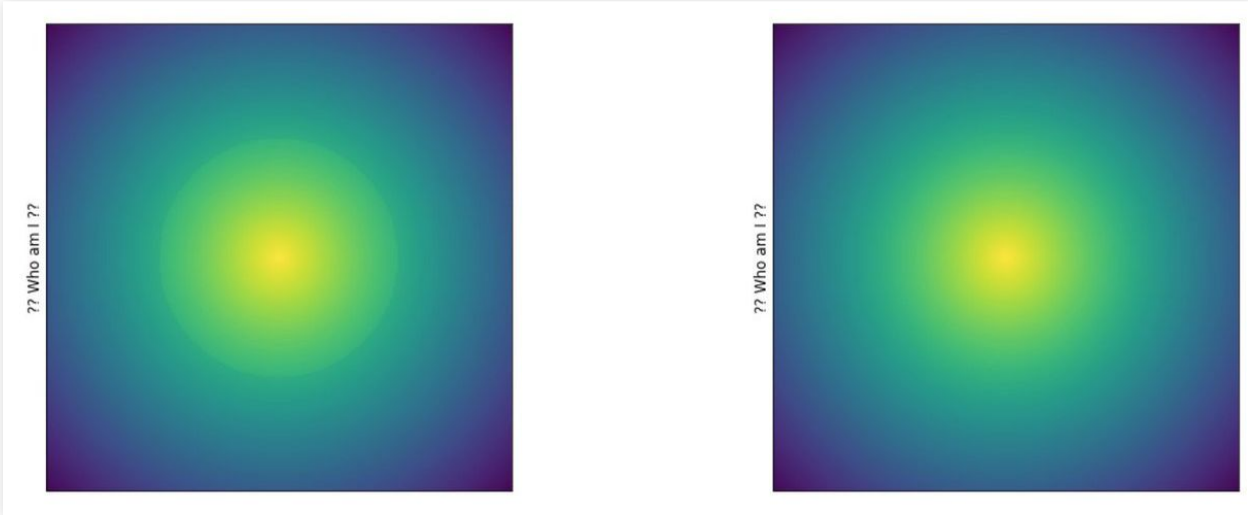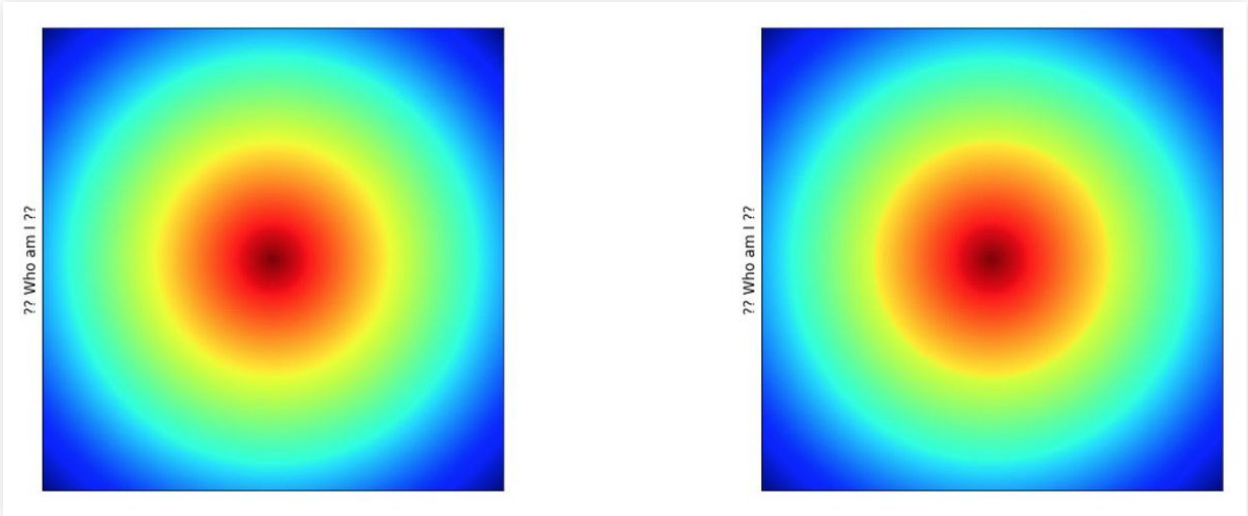
Information density

(Also called data-ink ratio)

> *Above all else show the data. — Edward Tufte*

Amount of information encoded, vs empty space in the graphic.

Generally higher information density is preferred.

- Subtract unnecessary ink (both data and non-data).
- De-emphasize remaining non-data ink. Emphasize remaining data ink.

Labelling graphics

Always label graphics! (items and attributes)

Ensure labels are legible (axis, font, size). Avoid abbreviations.

Use horizontal labels wherever possible. Rotating the graphic can make this easier. Oblique labels are a compromise.

Make sure all encoding is labelled. Legends are ok, directly labelling the data is better.

Producing graphics

1. For each item, choose the attribute(s) you wish to display
2. For each attribute to display, choose an appropriate channel to encode it
3. Choose a layout and labelling structure
4. If appropriate, choose how your item keys are sorted
5. Review and iterate

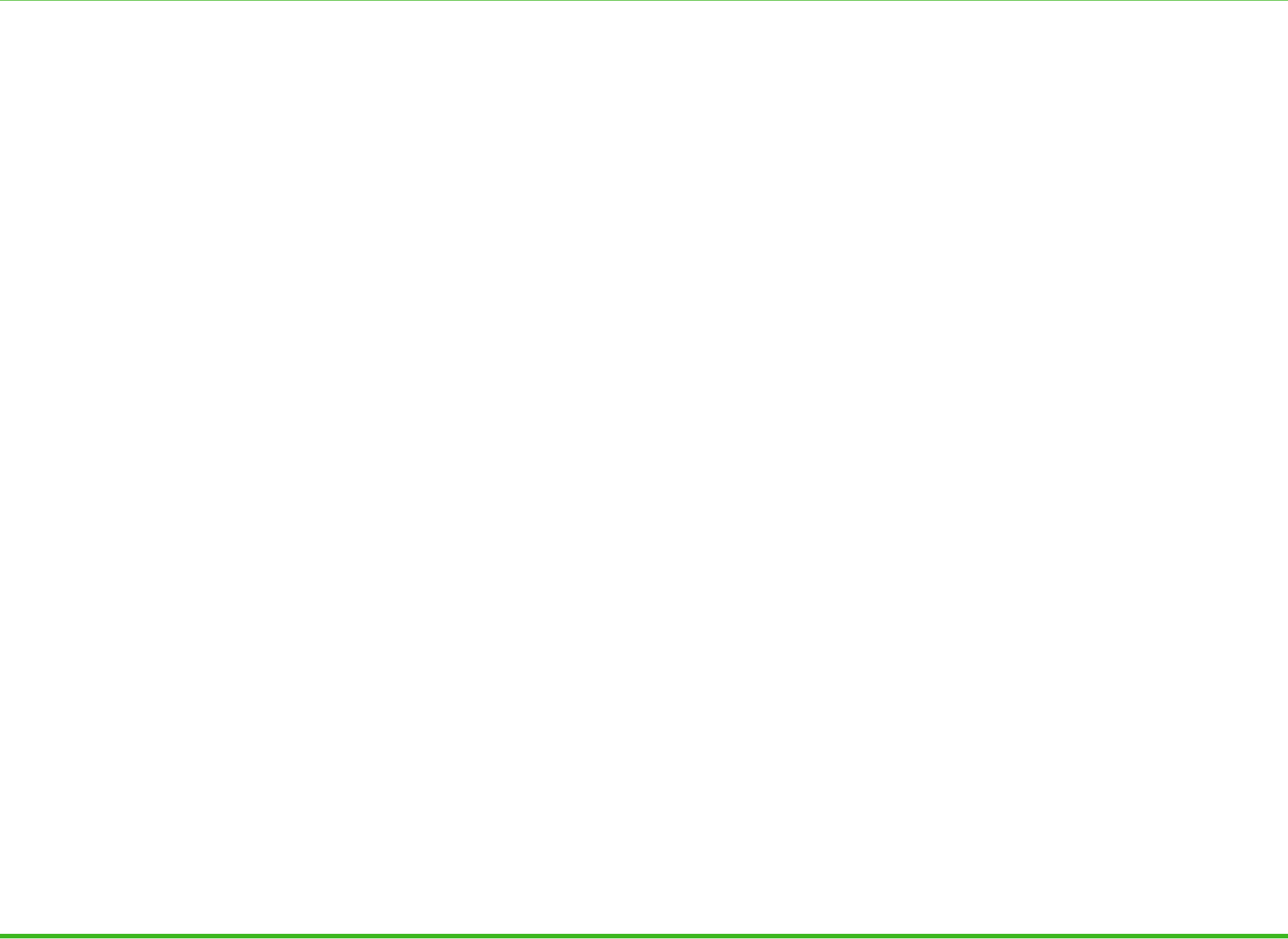# Graphics tools

Drawing: Visio, Inkscape

Data drawing: Tableau, Excel, Google charts

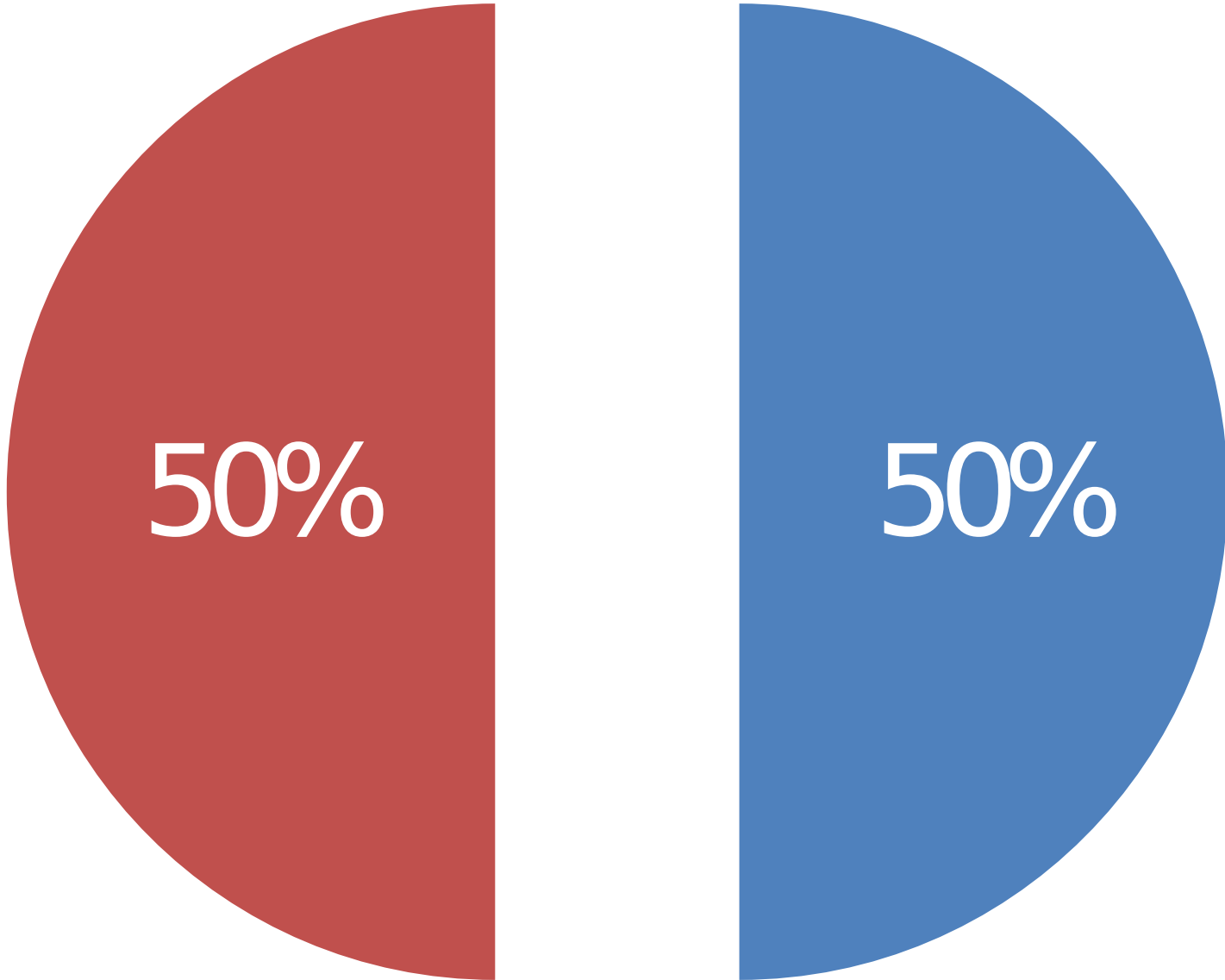**Data linking**: D3 (javascript), Shiny (R), Matlibplot (Python), Visio
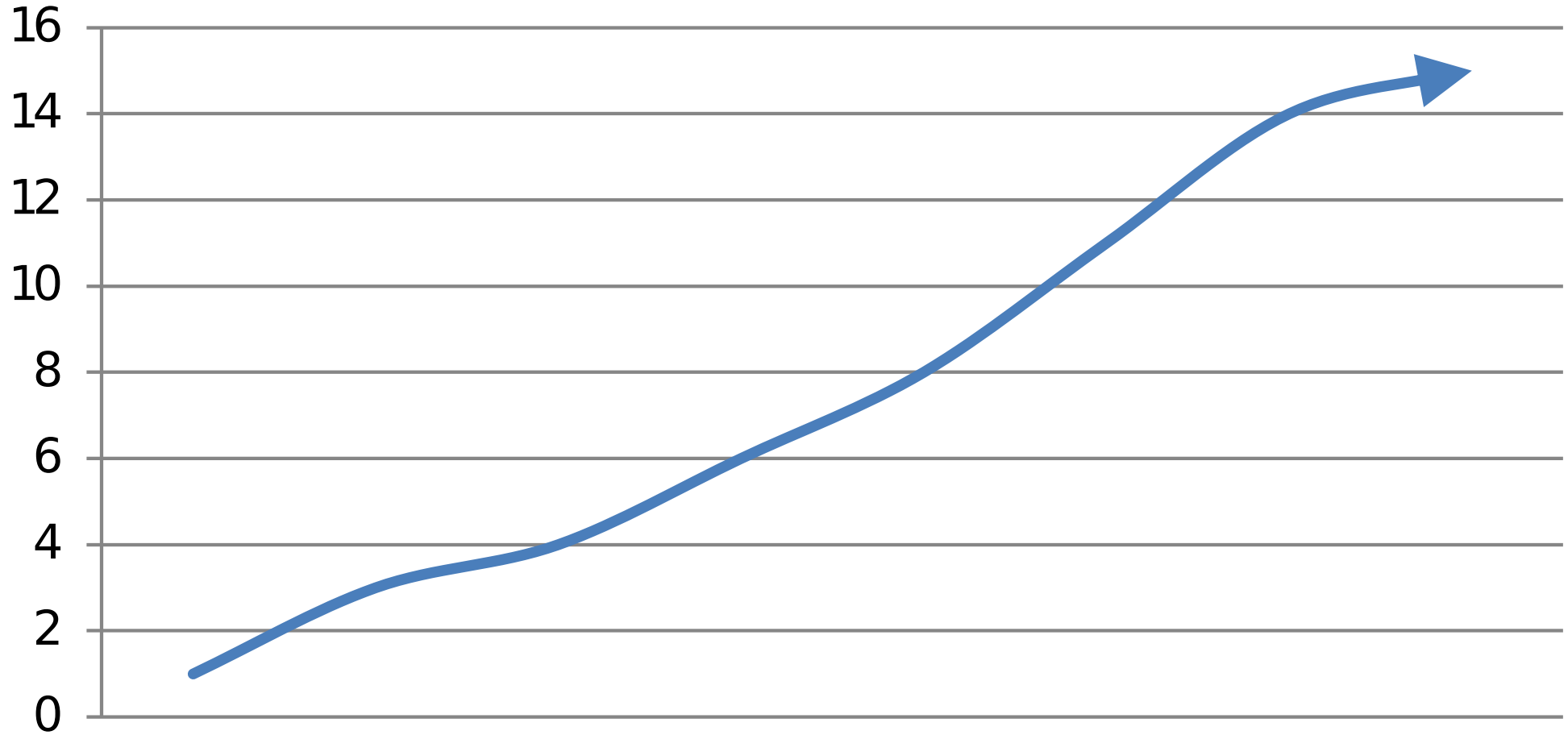
Lunch

extuitive.co.uk/ip

Personal graphical grievances
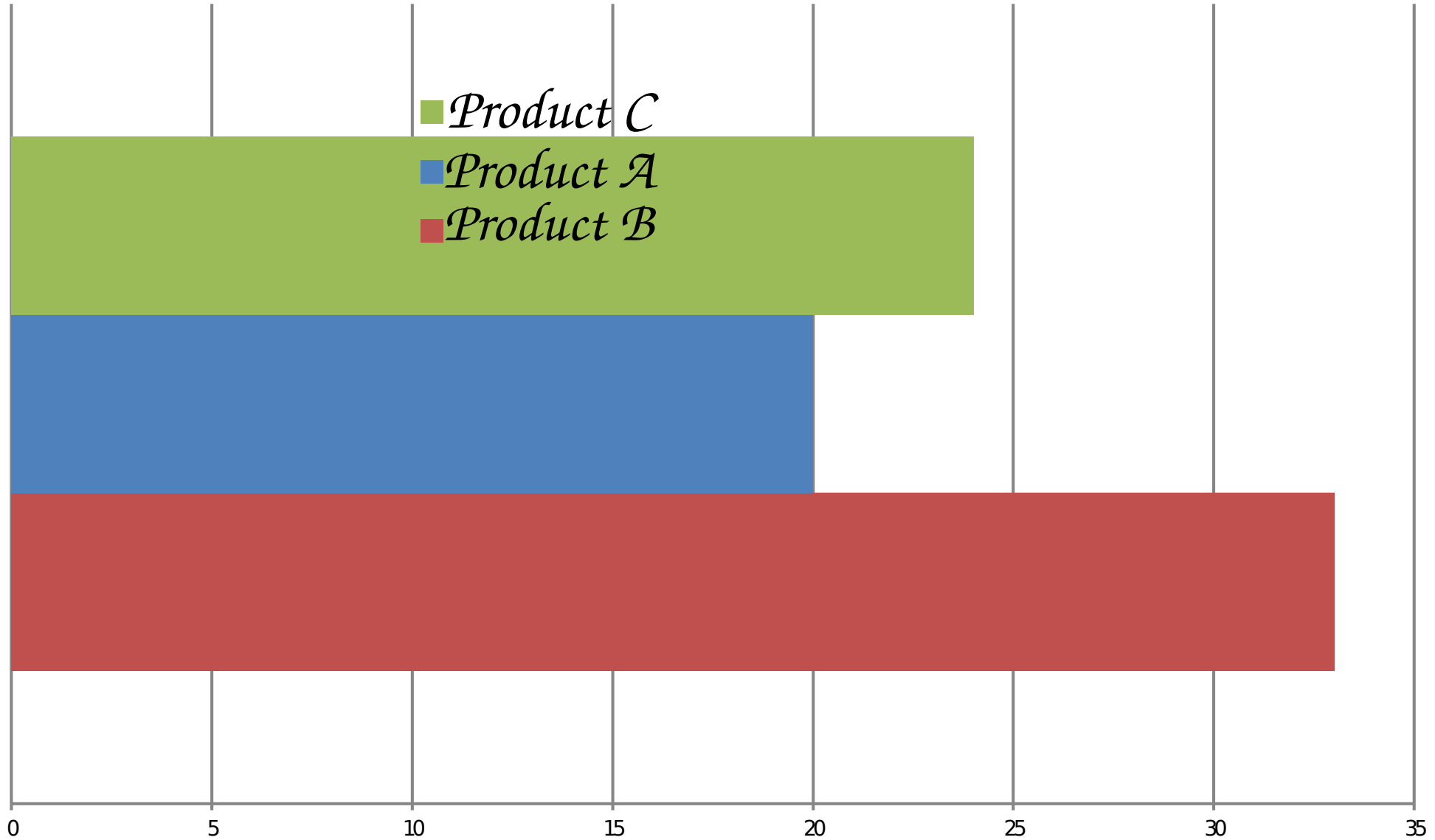
Unnecessary graphics
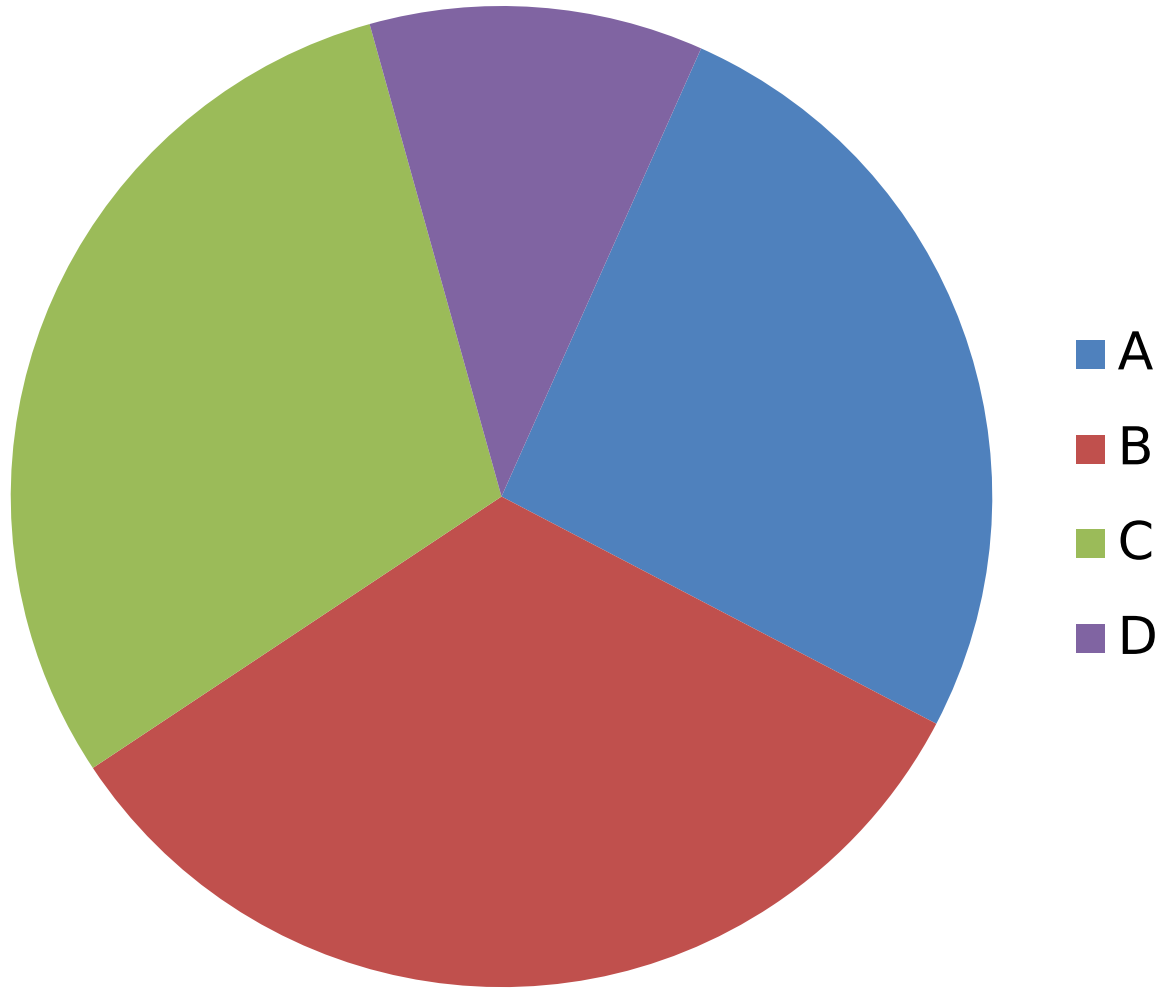
Lack of labelling

Poor fonts and/or legends

Product C

Product A
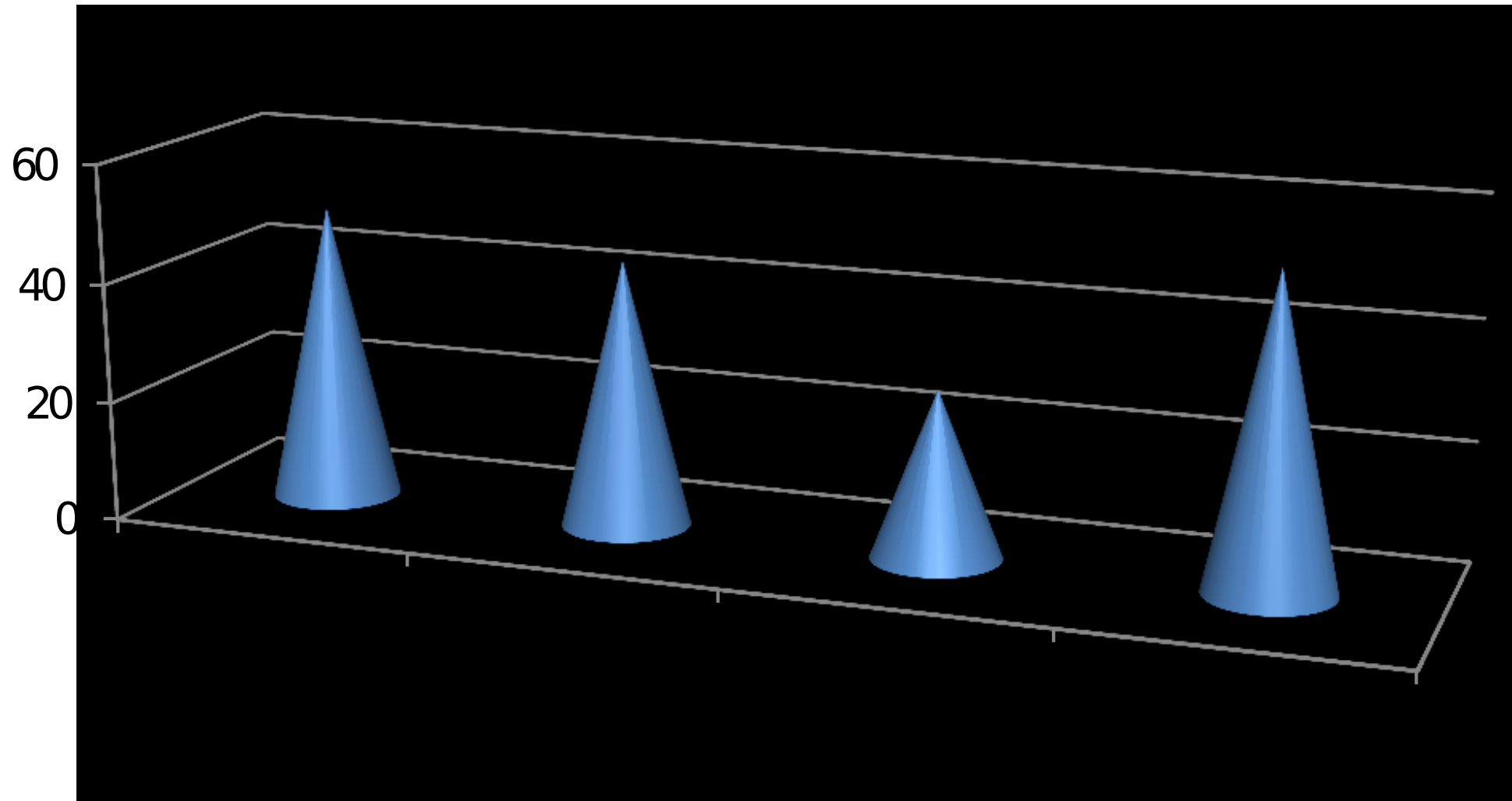
Product B

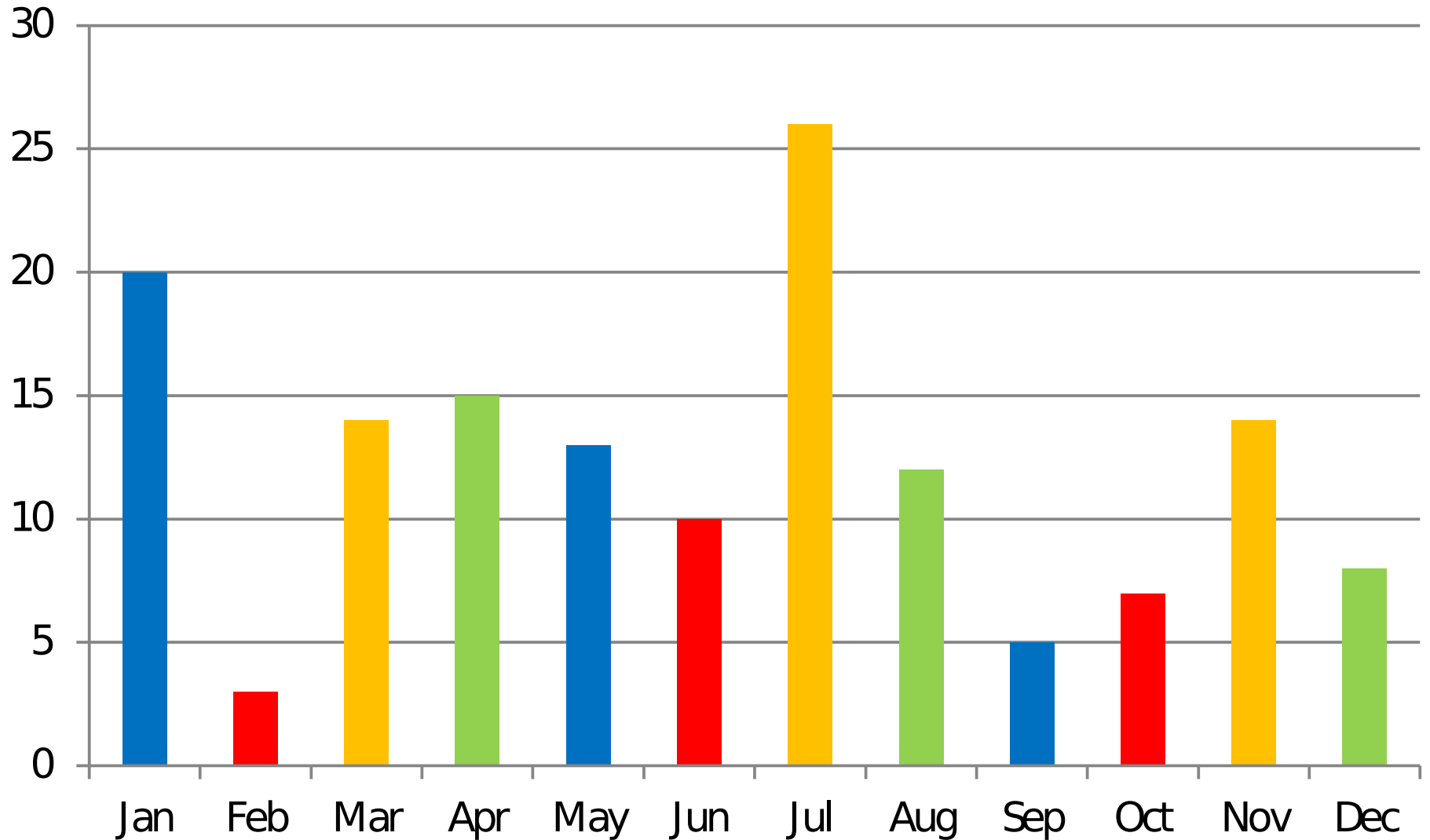0   5   10   15   20   25   30   35

Abbreviations
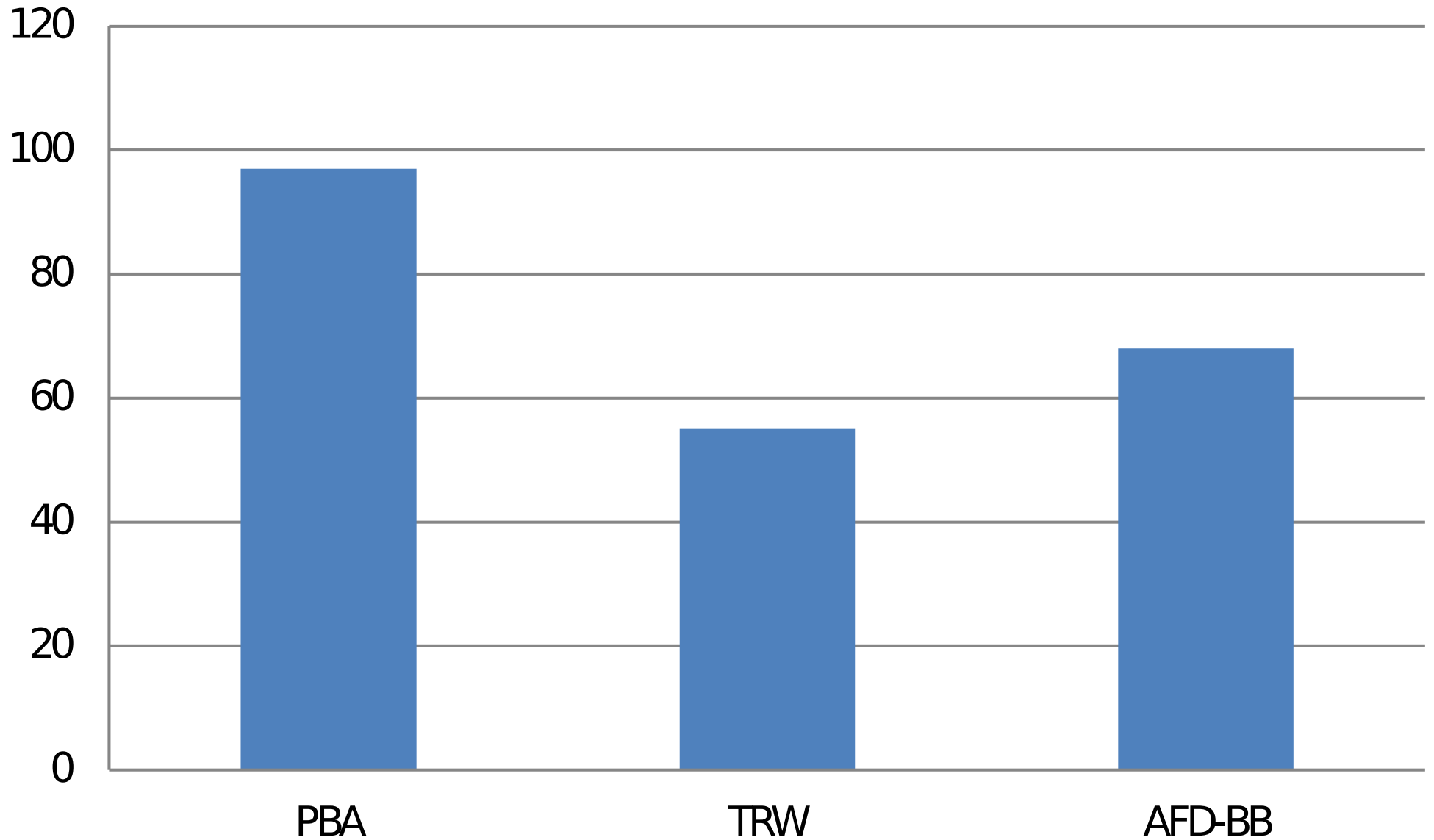
Use of angles or areas

Unnecessary dimensions
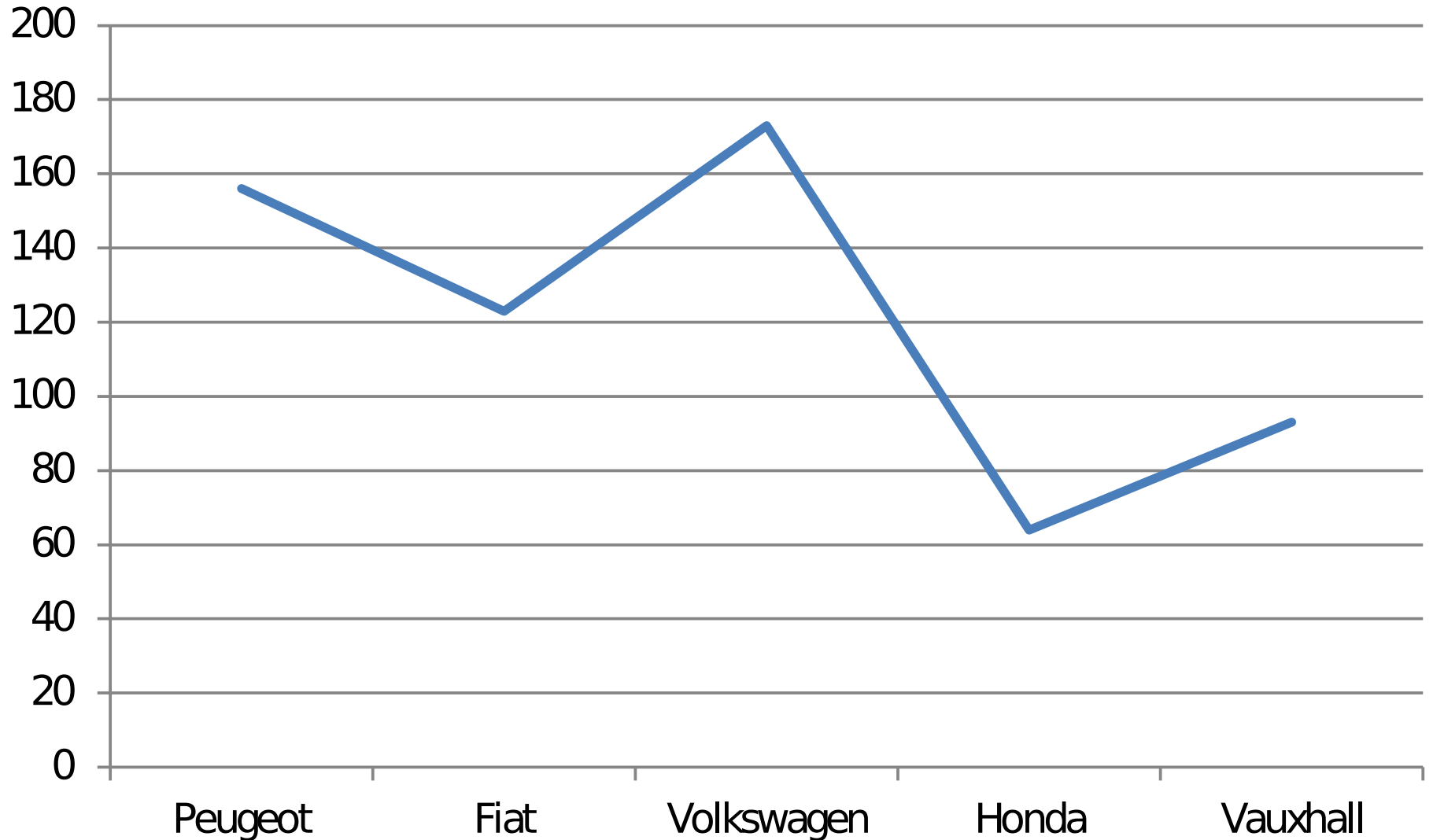
Irrelevant colours

Unsorted data
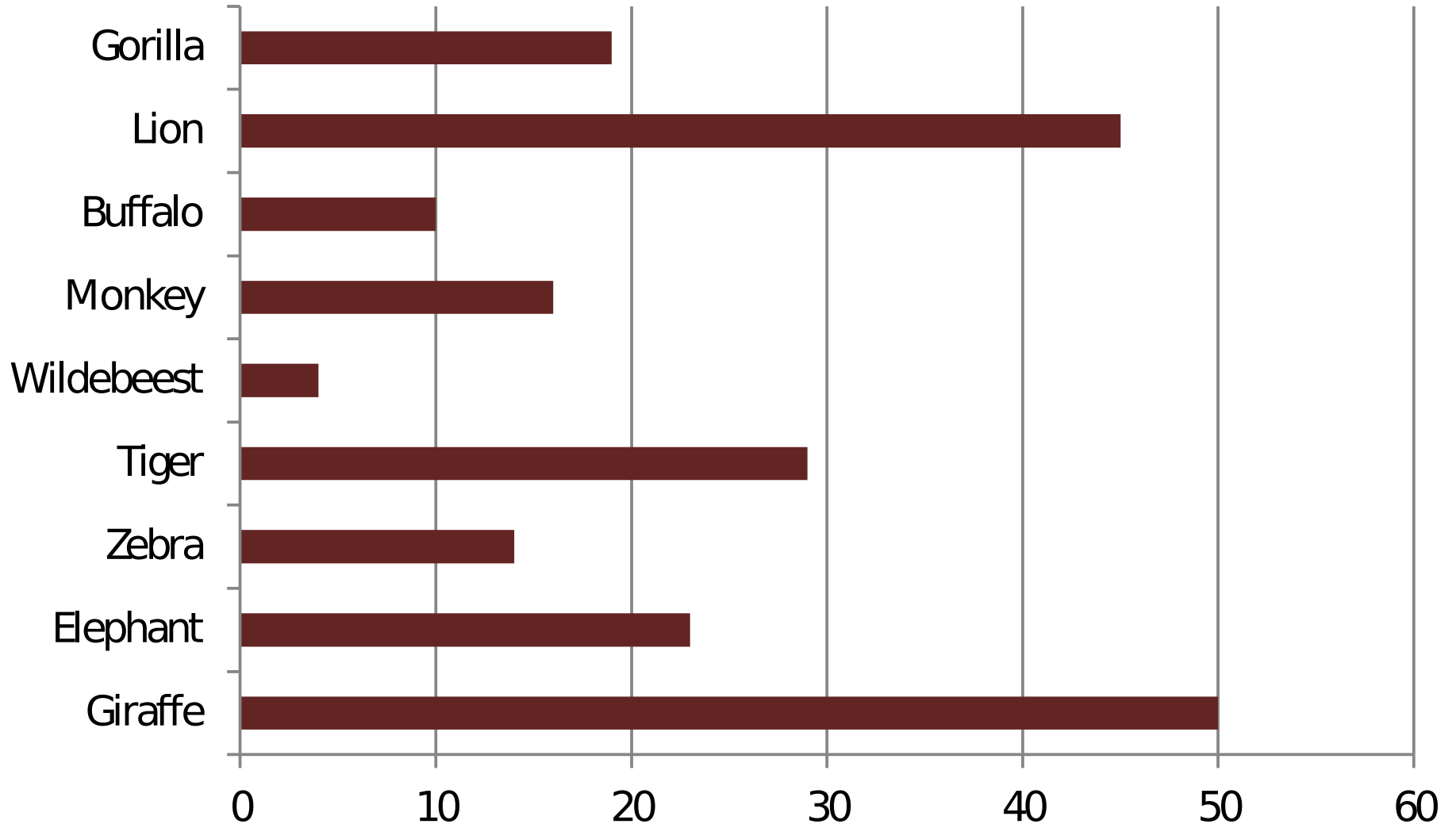
Scale or baseline issues
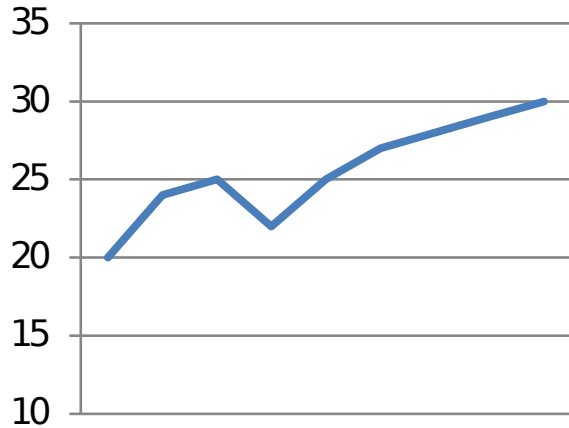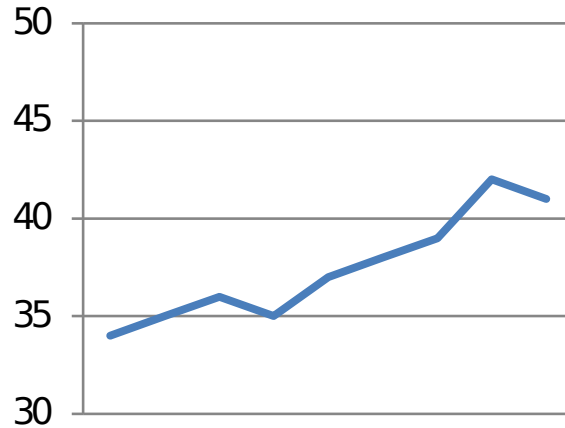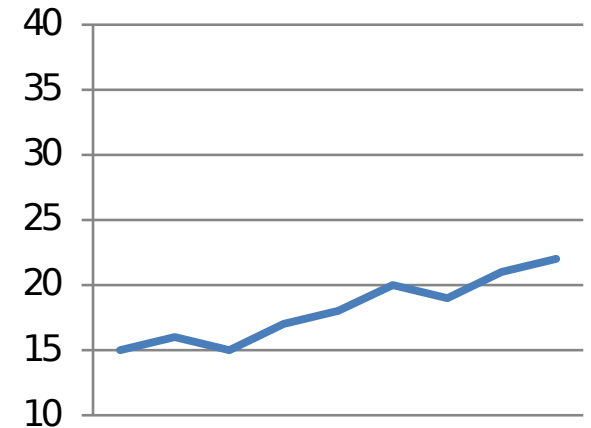
Group exercise

Improve upon specific graphics

A comparison of population and how people travel
Greater Sydney and its global city peers

Tokyo
13.6m
6,200 people/km²

Berlin
3.5m
3,900 people/km²

Los Angeles
13.0m
1,050 people/km²

Melbourne
4.3m
460 people/km²

Chicago
9.5m
3,400 people/km²

Greater Boston
4.7m
1,300 people/km²

Greater London
8.7m
5,800 people/km²

Greater Sydney today
5m
410 people/km²

New York (five boroughs)
8.4m
11,000 people/km²

Singapore
5.4m
7,000 people/km²

Paris
8.4m
6,000 people/km²

San Francisco (Bay area)
7.7m
425 people/km²

Greater Sydney in 2056
8m
610 people/km²

Global City Non-Car Mode Share

Major City Non-Car Mode Share

Car Mode Share

Figure 13: Private vehicle mode share – international comparison

Taken from the TfNSW Future Transport Strategy

SE Cost over "Successful" Programs
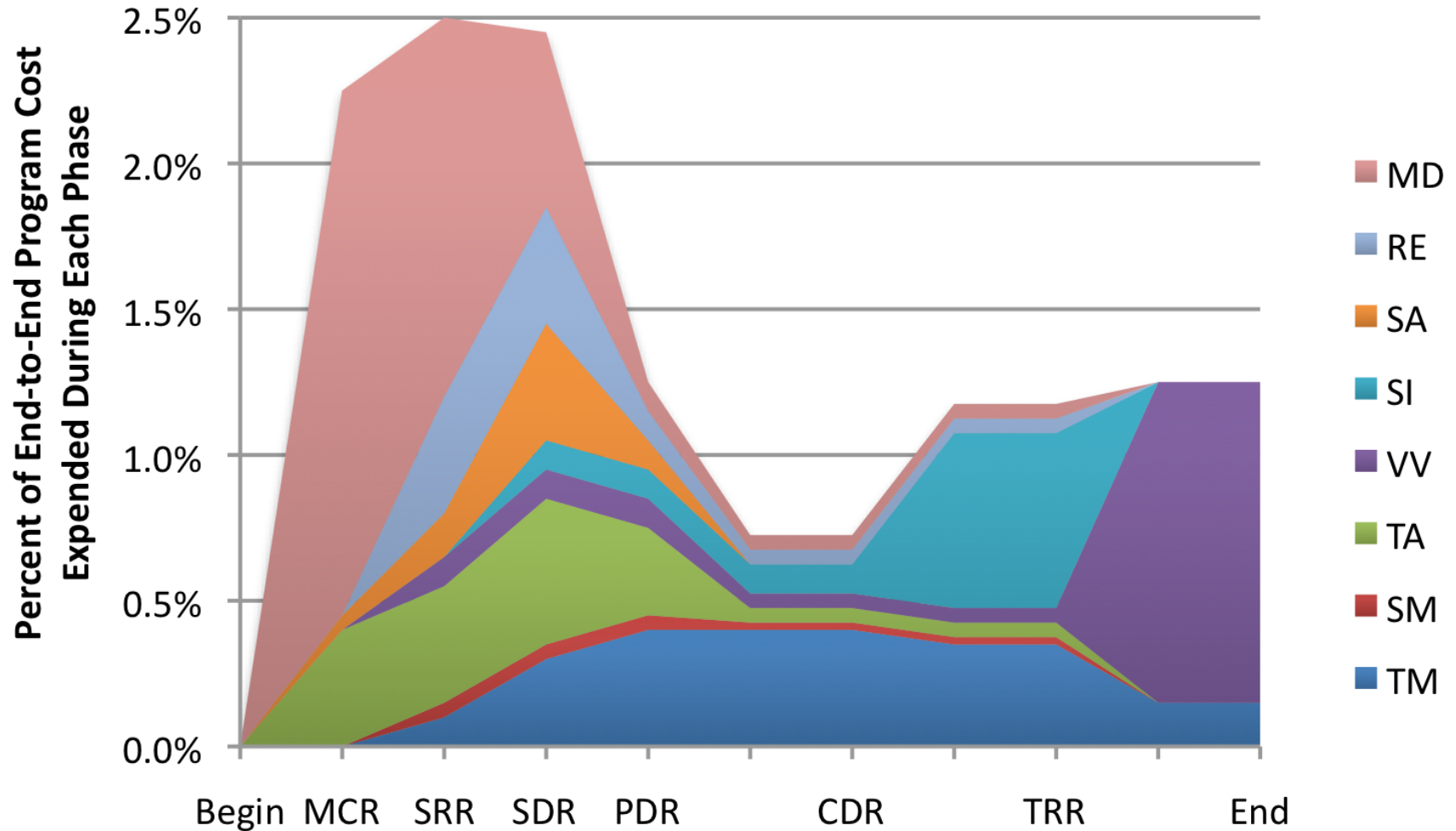
Taken from the seminal work of Eric Honour

ASEW

AUSTRALIAN SYSTEMS ENGINEERING WORKSHOP

ADELAIDE | 1-2 NOVEMBER 2018

(https://asew.com.au)

1. What do you think the overall message is? (*information*)
2. What is the background data? (*items* and *attributes*)
3. What type are the attributes? (*ordinal/quantitative/categorical*) (*keys/values*)
4. What are the channel encodings that have been used?
5. What different (/better) encodings could be used?
6. Sketch out a new version and label it.

Publication

Remember the basics!

Spell check

Check links and references

Set document properties

Remove comments and version history

Check formatting (in final publication format)

Work in a suitable format

Raster drawing tools manipulate pixels in an image. Except in rare cases, they do not embed the data.

Vector drawing tools manipulate shapes in a coordinate system. They allow data to be directly represented and embedded.

Benefits of vector formats:

- Vector → raster is easy, raster → vector is difficult and messy
- Vector files are typically smaller
- Vector files are easier to edit
- Vector images scale infinitely

Deliver in a suitable format

Word and Excel are editing tools, with reading modes bolted on.

Due to proprietary formats they also require the recipient to own a copy of the tool.

PDF is a better alternative. Delivering in PDF will:

- Reduce 'accidental' editing
- Reduce plagiarism
- Be more likely to reproduce correctly
- Be supported by many free viewers

HTML is emerging as an even better standard.

Image formats

What image format should I use?

Is part or all of it a vector image?

- No → Do you need animation?
  - No → Would you sacrifice some detail for a small file size?
    - No → portable network graphics .png
    - Yes → Do you want transparency?
      - Yes → portable network graphics .png
      - No → joint photographic experts group .jpeg / .jpg
  - Yes → graphics interchange format .gif
- Yes → Do you need it to work in Microsoft Office?
  - No → scalable vector graphics .svg
  - Yes → windows/enhanced metafile .wmf / .emf

bitmap .bmp

Advanced techniques

Interactive graphics

Interactivity is useful to handle complexity, it:

- Allows extra data to be shown without cluttering the graphic
- Allows the user to query the data
- Provides a lower level of detail
- Allows connection between views
- Engages and draws-in the reader

Avoid unnecessary interactivity, and do not rely on it.

Multiple views

A presentation of the data is only one possible *view* of it. Different views may be useful for:

- Different stakeholders
- Presenting different information (based on the same data)
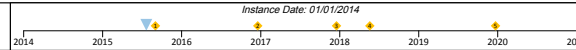- Showing different levels of abstraction to develop understanding

London North Western (LNW)

Platform length assessment for:
Class 158/0 2 car, 2 units (90.28 m)

Prepared by WSP | Parsons Brinckerhoff for Network Rail

Instance Date: 01/01/2014

DRAFT

Document Reference Number

Project	Originator	Doc.ID	Discipline	Seq. Number	Version

Legend:
- All platforms in station fit selected rolling stock
- Some platforms in station fit selected rolling stock
- No platforms in station fit selected rolling stock

(Click to change rolling stock type)
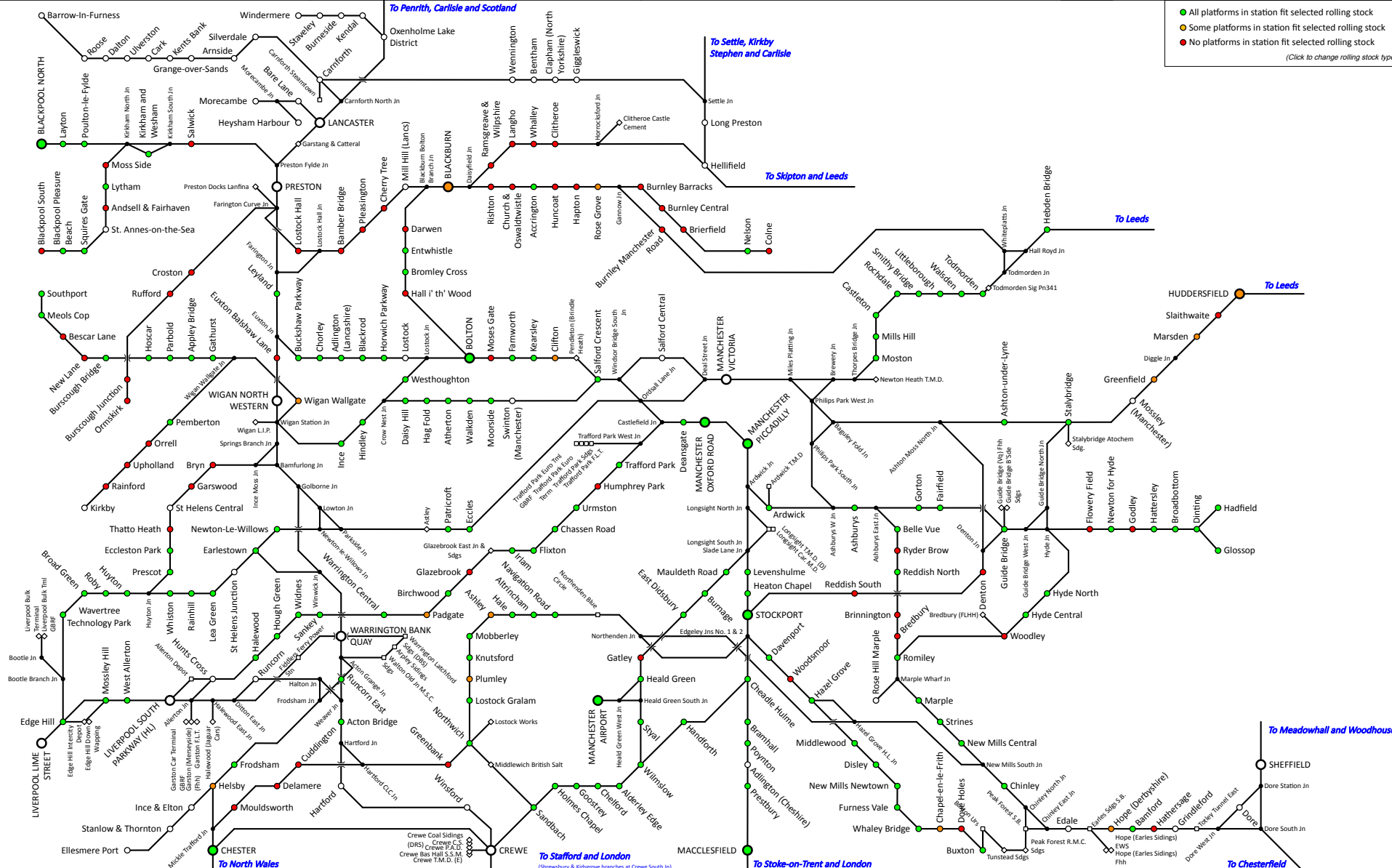
Drawn by:	Name	John Welford	Date	Checked by:	Name	Kevin Gedge	Date	Approved by:	Name		Date	Page
	Signature				Signature				Signature			1 of 1

Levels of abstraction

Present to reflect Systems Thinking – take the reader through multiple levels of abstraction…

Credit Bret Victor

Up and Down the Ladder of Abstraction

A Systematic Approach to Interactive Visualization

Bret Victor / October, 2011

"In science, if you know what you are doing, you should not be doing it. In engineering, if you do not know what you are doing, you should not be doing it. Of course, you seldom, if ever, see either pure state."

—Richard Hamming, *The Art of Doing Science and Engineering*

How can we design systems when we don't know what we're doing?

The most exciting engineering challenges lie on the **boundary of theory and the unknown**. Not so unknown that they're hopeless, but not enough theory to predict the results of our decisions. Systems at this boundary often rely on *emergent behavior* — high-level effects that arise indirectly from low-level interactions.

When designing at this boundary, the challenge lies not in constructing the system, but in understanding it. In the absence of theory, we must develop an *intuition* to guide our decisions. The design process is thus one of exploration and discovery.

How do we explore? If you move to a new city, you might learn the territory by walking around. Or you might peruse a map. But far more effective than either is *both together* — a street-level experience with higher-level guidance.

Likewise, the most powerful way to gain insight into a system is by *moving between levels of abstraction*. Many designers do this instinctively. But it's easy to get stuck on the ground, experiencing concrete systems with no higher-level view. It's also easy to get stuck in the clouds, working entirely with abstract equations or aggregate statistics.

This interactive essay presents the **ladder of abstraction**, a technique for thinking *explicitly* about these levels, so a designer can move among them consciously and confidently.

I believe that an essential skill of the modern system designer will be using the *interactive medium* to move fluidly around the ladder of abstraction.

Transitioning graphics

Animated transitions between views of the same data can be used to help aid understanding.

Research has shown that animated transitions between related data graphics significantly improve visual perception.

These benefits are greater if the animations are staged.

Analysis of climbers anatomy survey results

We show each respondent by a segment in the chart with two colours side-by-side, one colour represents whether they have [Palmaris Longus](#) in neither (■), one (■), or both hands (■), the other represents whether they suffer from [Dupuytren's contracture](#) in neither (■), one (■), or both hands (■).

Sorting the respondents by incidence of Palmaris Longus we can see that, of the 193 respondents, the majority (77%) had Palmaris Longus in both wrists, with 13% having it in only one wrist and 10% missing it altogether.

Palmaris Longus
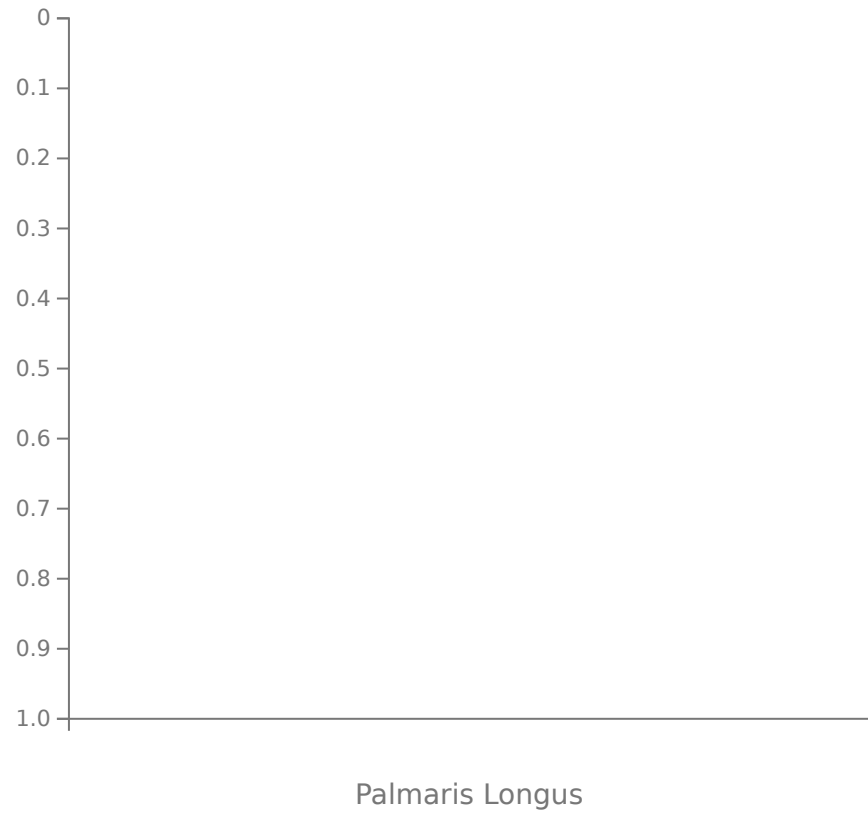
Explorable explanations

Explorable explanations are reactive and explorable documents that allow a reader to:

- Understand the authors underlying models
- Play with model assumptions
- Provide contextual information to learn related material
- Cross check the authors claims

There are a large number of examples available.

Most are built in HTML, however the Computable Document Format does exist as well.

A recent development is the Observeable coding system, which provides a reactive programming environment to support explorable explanations, visualisations and active reading.

johnwelford · Jul 6, 2018

♡
4

The Rock's "Skyscraper" jump

There appears to have been some speculation on the internet about whether Dwayne 'the Rock' Johnson will make the epic jump in his latest movie Skyscraper. This seems to have gotten serious, with 'Math Professors' being called in.

Who could resist the lure of a maths problem involving the Rock, so I took a look! I'm not going to claim that their analysis is wrong, but I definitely found it hard to follow, especially as they dive straight into parabolas and simultaneous equations. So I thought to myself, let's have a go at this and see if we can make it any clearer...

Let's start by getting a feel for the scale of the image. First we need the height of the Rock, which is quoted on Wikipedia as 1.96m. I'm not sure if this is correct, but if you want to play with things then go ahead and change the number below.

theRocksRealHeight > 1.96

Next, we can load the poster image into any old paint program and take some measurements from the image in pixels. Useful information will be: the height of the Rock (65 pixels), how far it is between the crane and the building (422 pixels), and maybe how high the crane is above the bottom of the window (213 pixels).

theRocksImageHeight > 65

craneToBuildingImageDistance > 422

craneToWindowImageHeight > 213

Using the Rock's real height, and how big he appears in the poster, we can get a measure of scale, and calculate that he will need to travel a massive 12.72m to get to the building – lucky this is the Rock we're

Speed reading

There are a few technologies around such as Spritz to facilitate speed reading.

Spritz

E-books

E-books are arguably the future of reading:

- 'Paperless office'
- Searchable
- Distributable
- Compact
- Resizable (reflowable text)

There are a variety of formats, readers and creation tools available, formats can be converted easily.

Summary

Discussion

The end

Feedback

✉ john.welford@wsp.com

🐦 @extuitive