Data visualization: Part 1

Dr W.J. Macinnes

April 14th, 2017

Visualization

- More than just showing data
  - Data graphics vs information visualization
  - Points, lines, one or more coordinate system, numbers, symbols, words, shading and colour
  - Allows *reasoning* about data
- Important part of the story you are trying to tell
  - Or a small story in itself
  - More than just a substitute for small statistical table
- Mathematics goes back thousands of years
  - But visualization barely 200
  - William Playfair (1759-1823)
- Statistic, words and visualizations are your three key tools for telling your story
- Reference: Many images from : Edward Tufte. *The visual display of quantitative information.*

Golden rule

- Precision in writing?
- Most information in the fewest words
- Precision in visualization?
- Most information with the least ink

- Golden rule ‘Don’t let the visualization get in the way of understanding the data’

Principles

- Show the data
- Allow the viewer to think about substance or essence of data
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets easy to understand
- Encourage the eye to compare different pieces of data
- Show the data at multiple levels at the same time (detailed and overview)
- Have a clear purpose
  - Description
  - Decoration
    - Not always least ink, but beware of ducks
  - Exploration
- Be closely aligned to the statistic and verbal descriptions

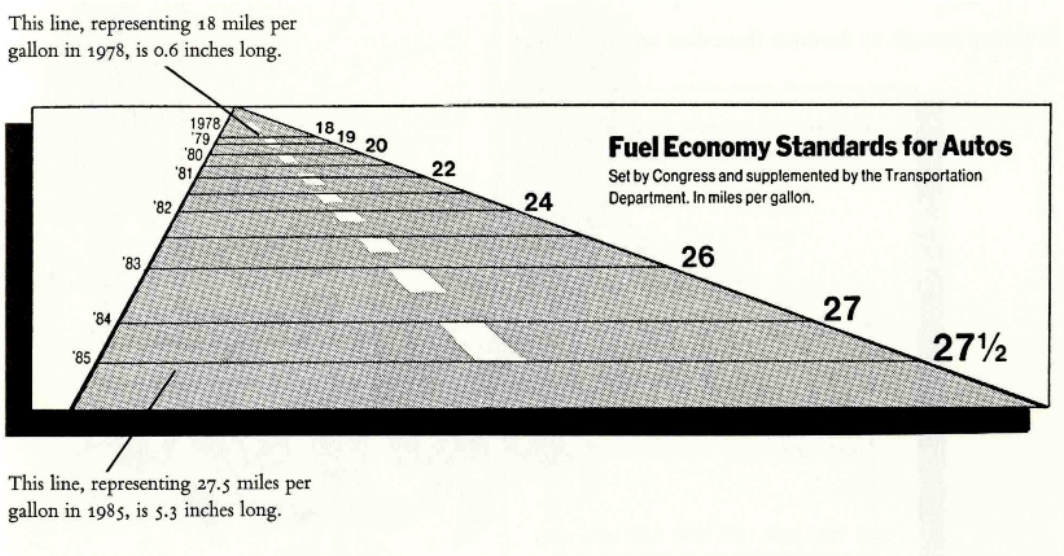
Ducks (AKA Chart-junk)

Visual diagnostics

- Cholera outbreak, London, 1854
- Dr John Snow plotted a dot for address of each death, and a cross for location of water pumps
- After creating this visual, he removed the handle from the pump at Broad street (circled) and ended the epidemic that killed 500 people in that neighbourhood

Lie factor

- Lie factor = size of effect in graphic/size of effect in data
- Graph lines increase 758%, data increases 53%
  - 783/53= 14.8
- Other problems
  - The future is usually depicted as 'ahead' of us, but this graph reverses that standard to help the exaggeration
  - Dates on the left maintain the same font size despite the perspective that changes everything else
  - Line size is shrinking from both perspective and smaller data numbers. The viewer cannot separate these two effects

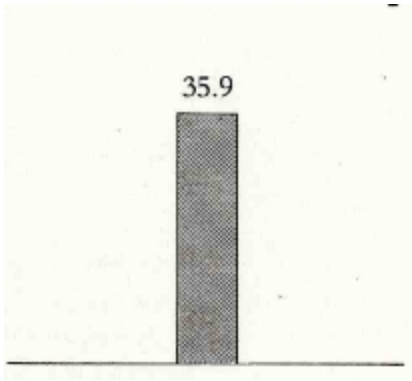


Principles

1. Representation of numbers, as measured by graph surface, should be proportional to the real data
2. Graphs require clear and detailed labelling, including key events
3. Show data variation, not design variation
4. Ensure your data only display a single influence. Use standardized units if the data change over time due to causes other than your primary effect
5. Don't use more dimensions in the graph than in the data
6. Do not quote data out of context

Repetitive ink

- Some redundancy in graphics is good, but too much is wasted ink
  - A basic bar chart repeats height information in six different ways
  - Eliminate any 5 and you still get the point across

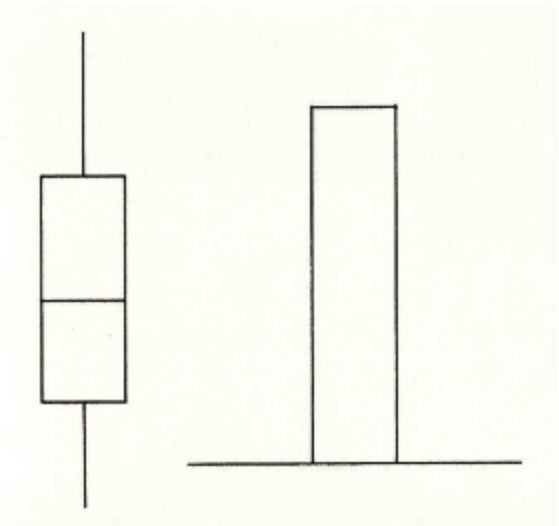


Value/magnitude is portrayed by

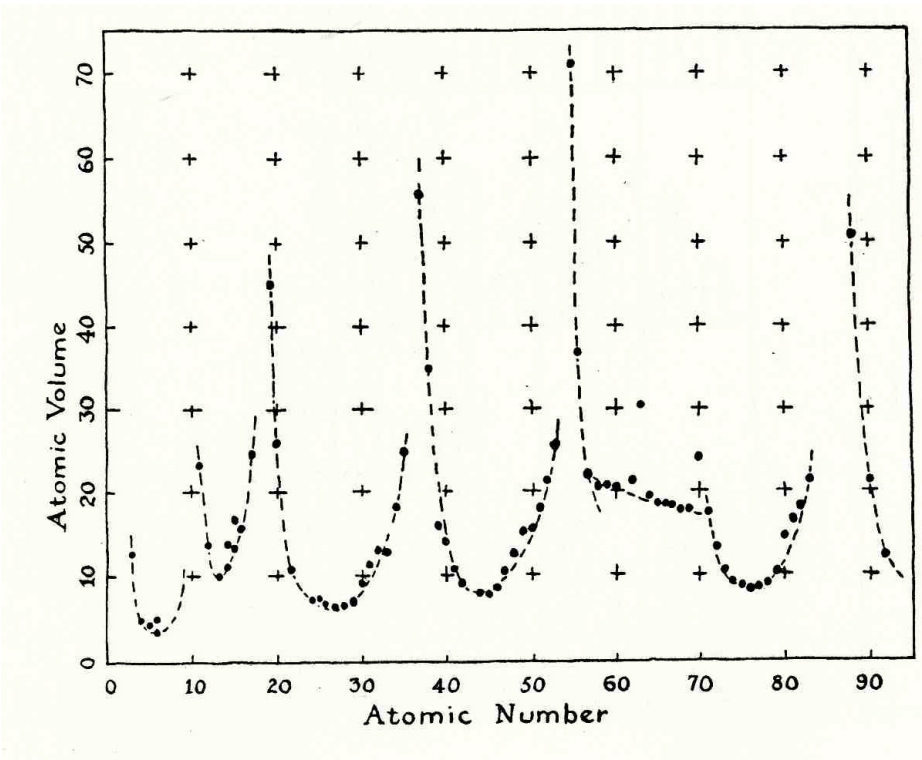
1. Height of left line
2. Height of right line
3. Height of shaded area
4. Position of top horizontal line
5. The number itself
6. The location of the number

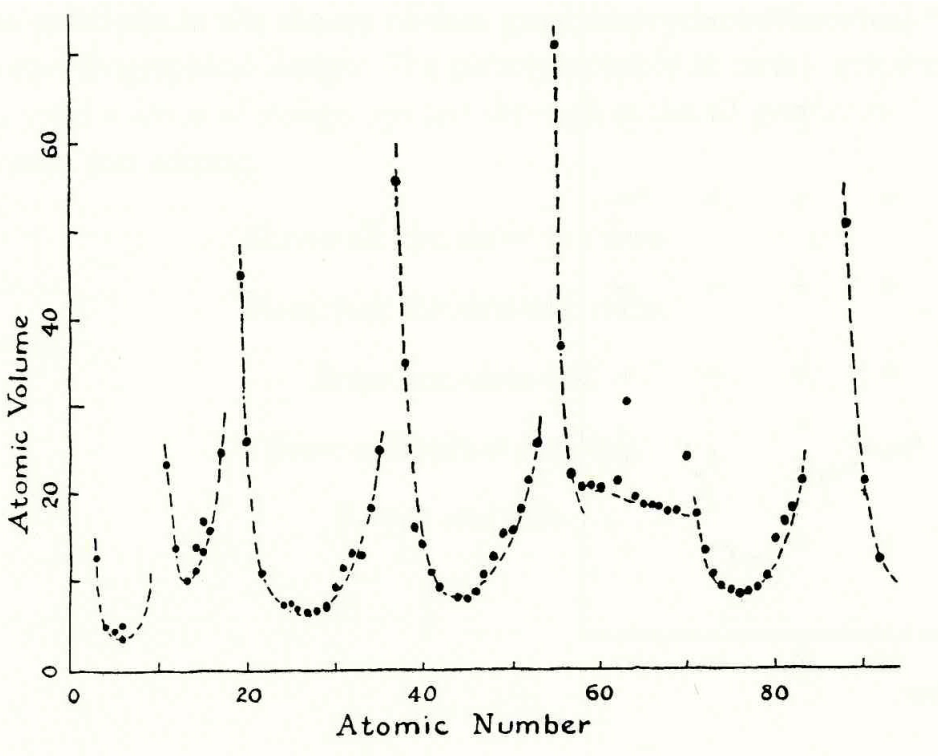
Remove ink or add data

- Box plot vs bar chart
- Same ink and both portray mean
  - but box plot adds variance and 90%/ 10% confidence intervals

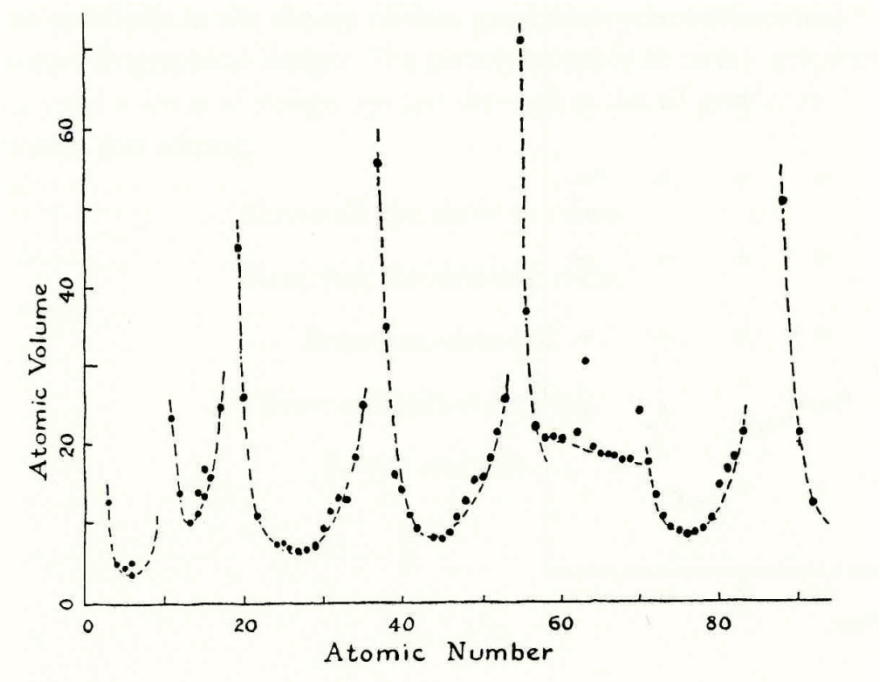


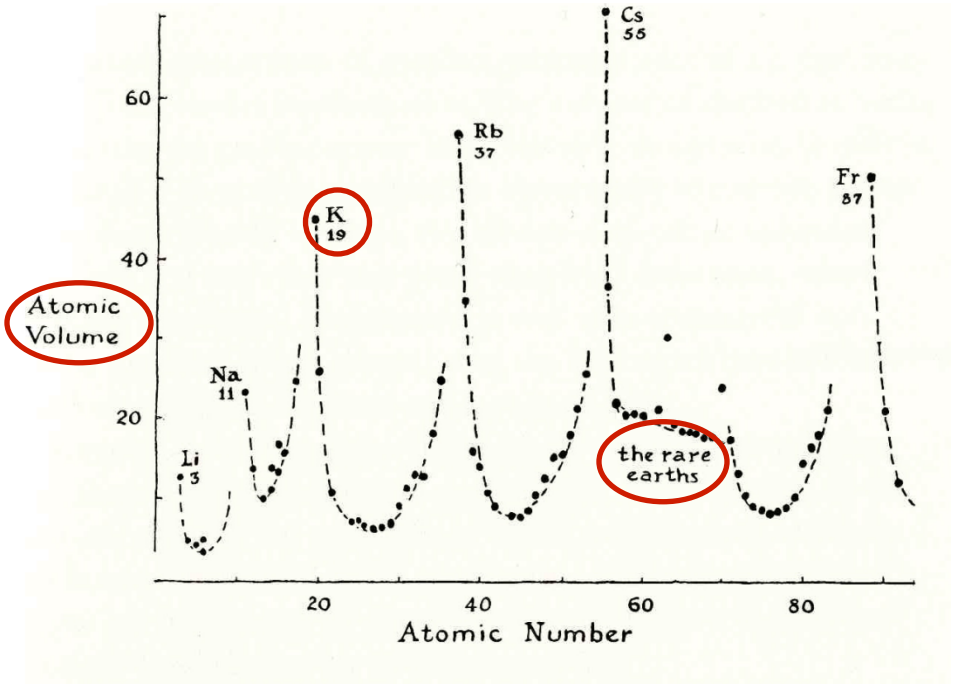
Reduction


Atomic Volume
0
10
20
30
40
50
60
70
0
10
20
30
40
50
60
70
80
90
Atomic Number


Atomic Volume
0
20
40
60
0
20
40
60
80
Atomic Number

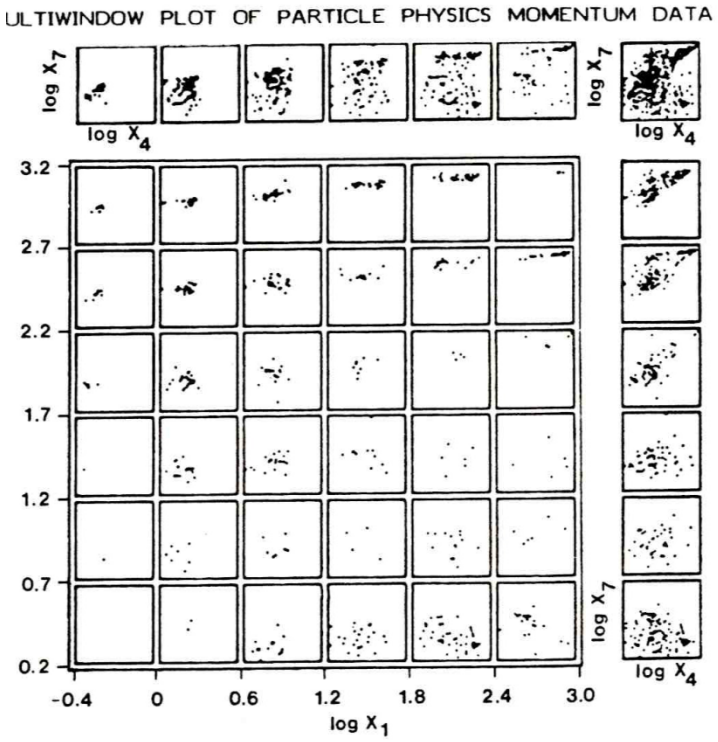
And then addition

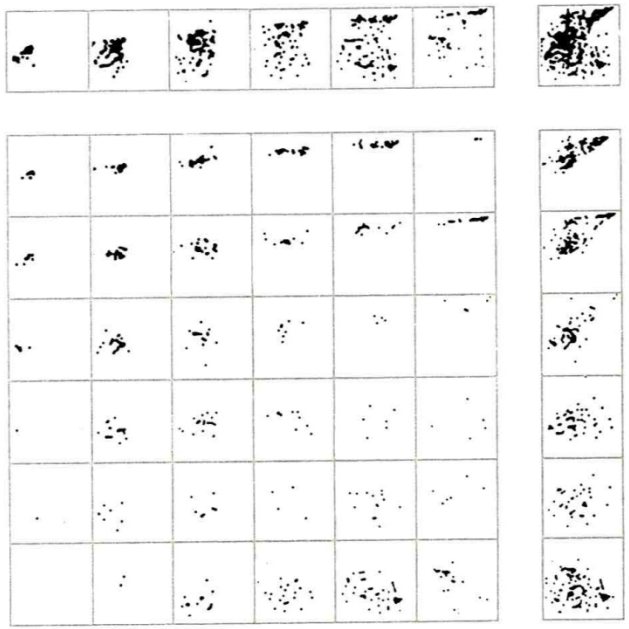




Darkness

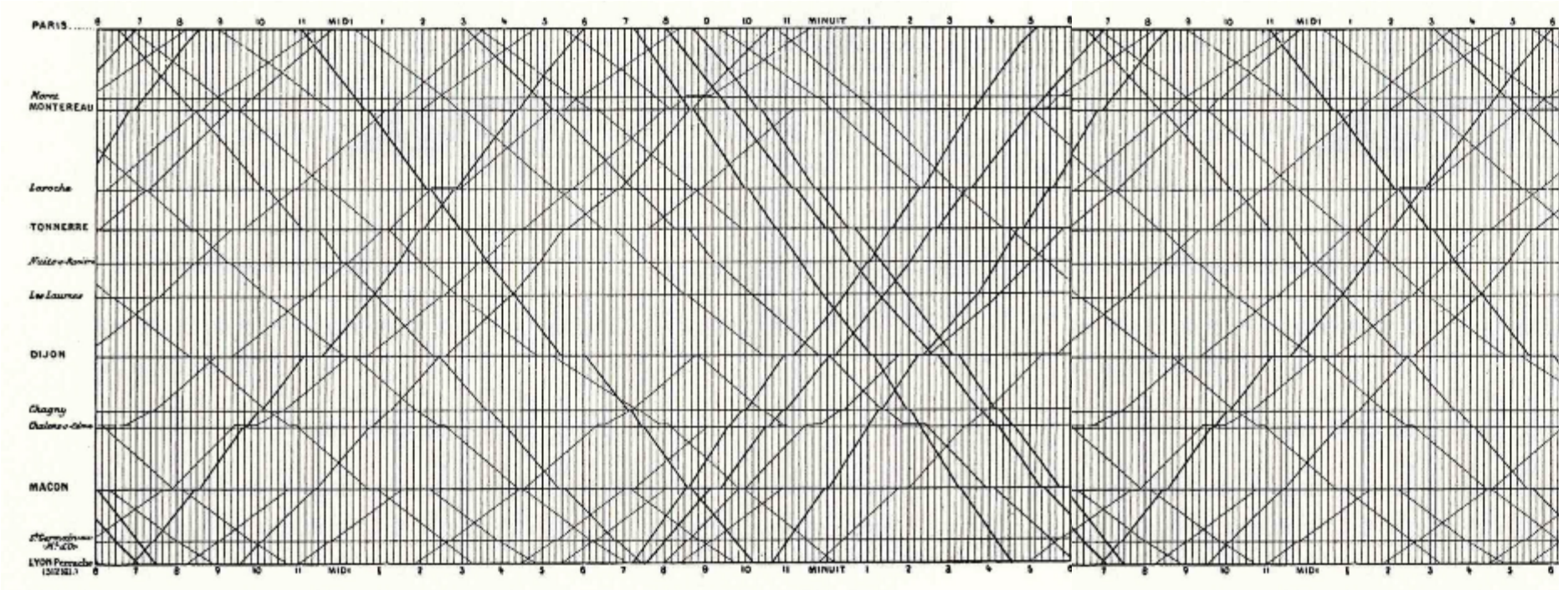
- Light grey and muted colour is your new best friend





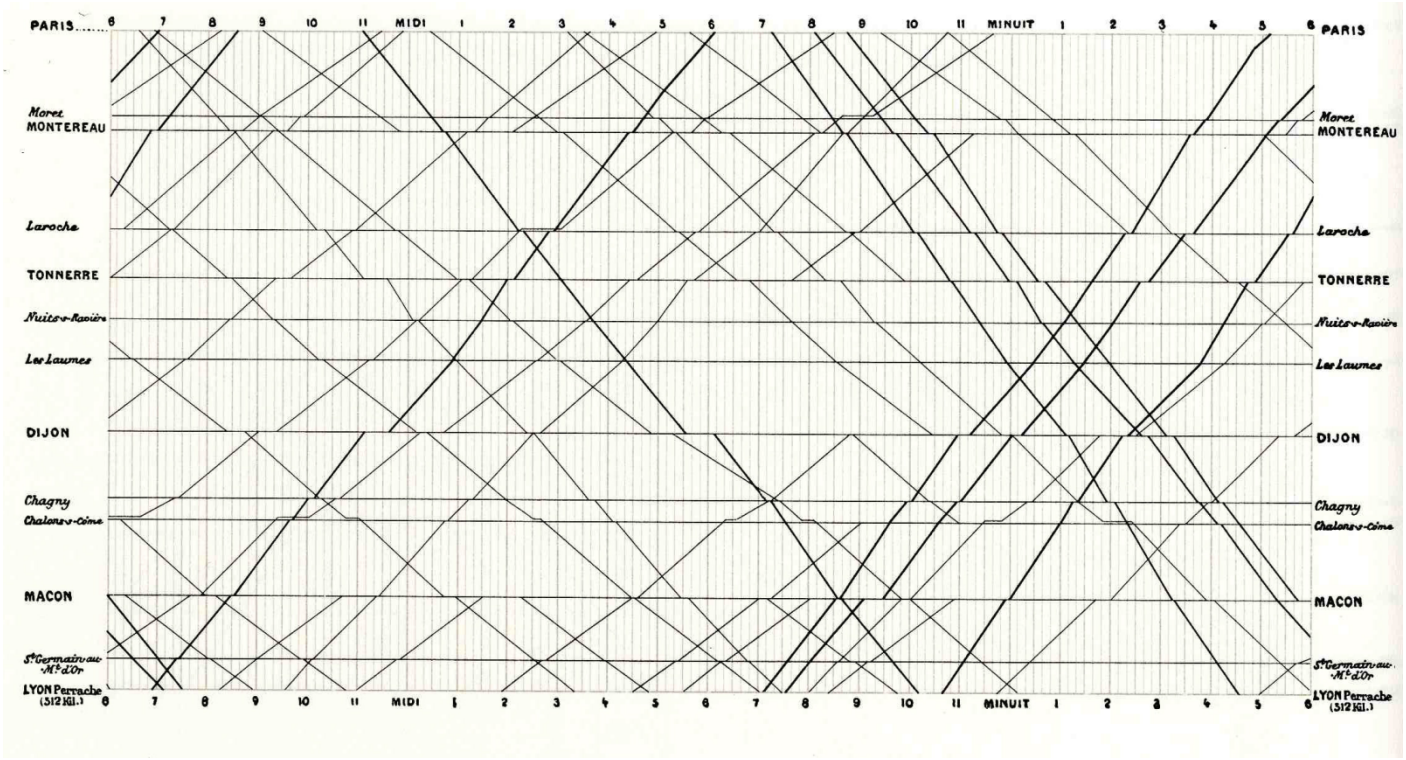
Good redundancy

- Repeating part of the day in the French train example
  - Travellers don't have to find their line on the left after it disappears on the right
  - Or a clever designer could mount this on a cylinder instead!



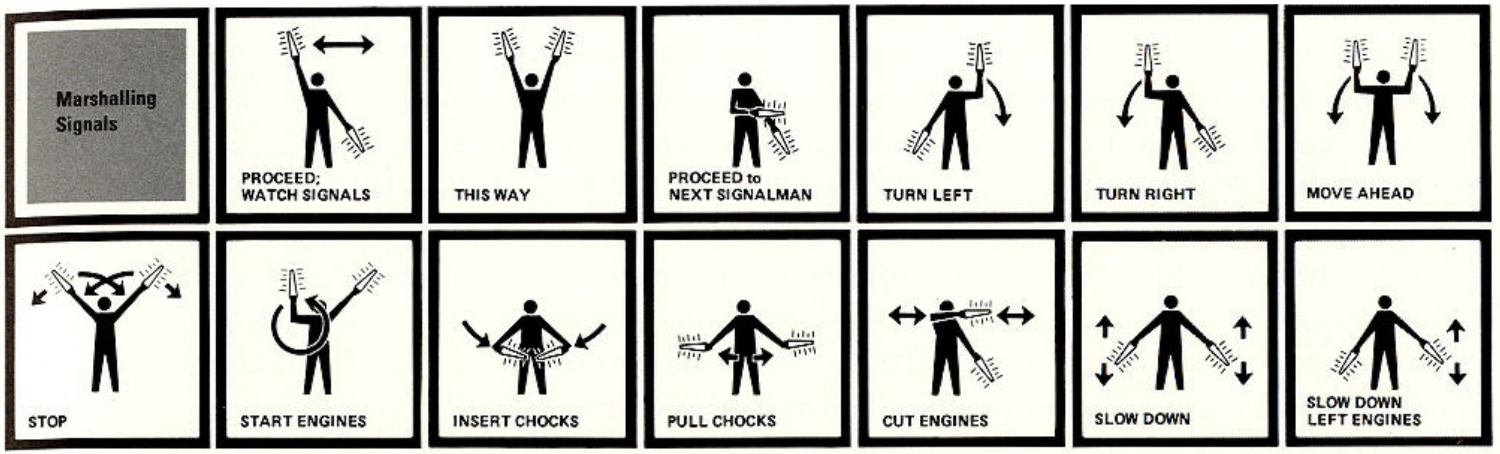
Thin grey

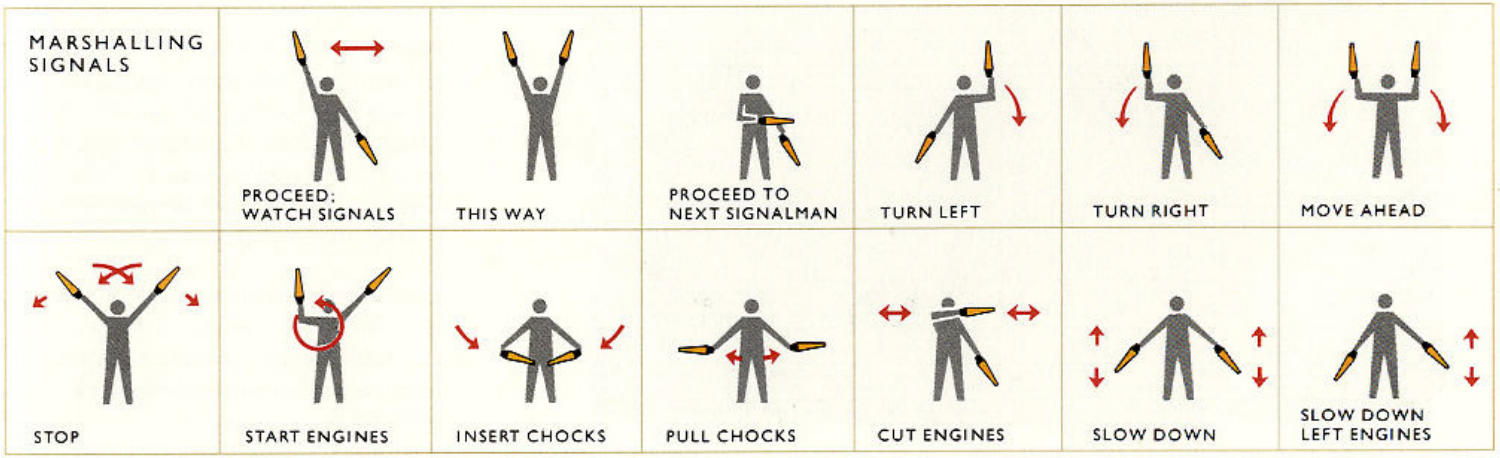
- Remember our train schedule?



Colour : other

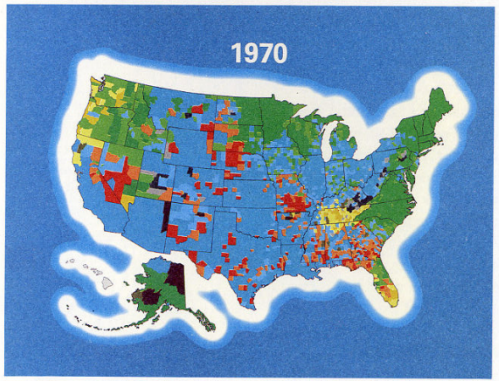
- Be VERY conservative using colour
- Small amounts of intense, saturated colour can draw attention to key aspects
  - Motion, significant results
  - Too much colour will attract attention everywhere

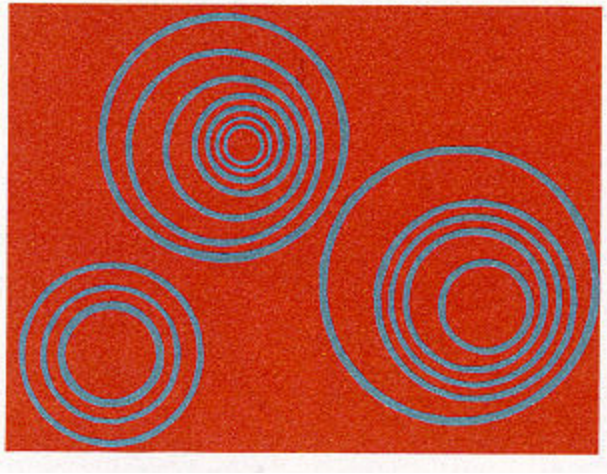


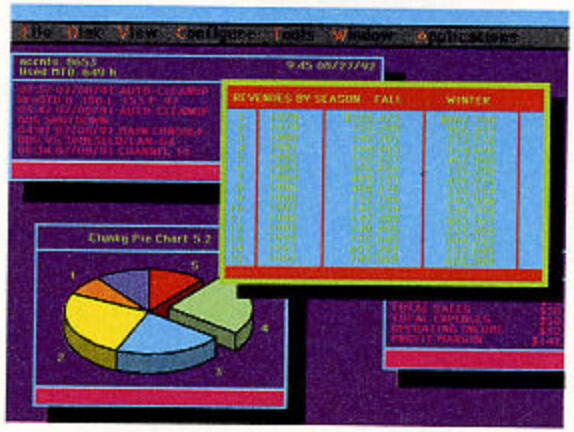


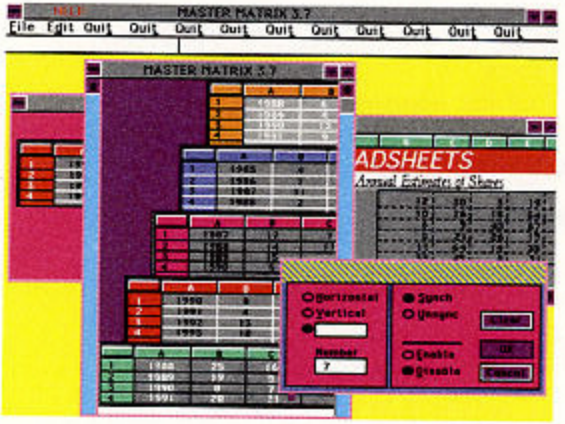
Colour excess

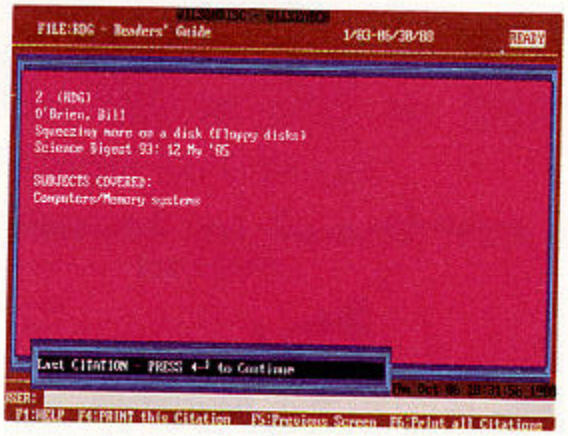
- Avoid colour for colour's sake
- Blue and red in particular do not mix





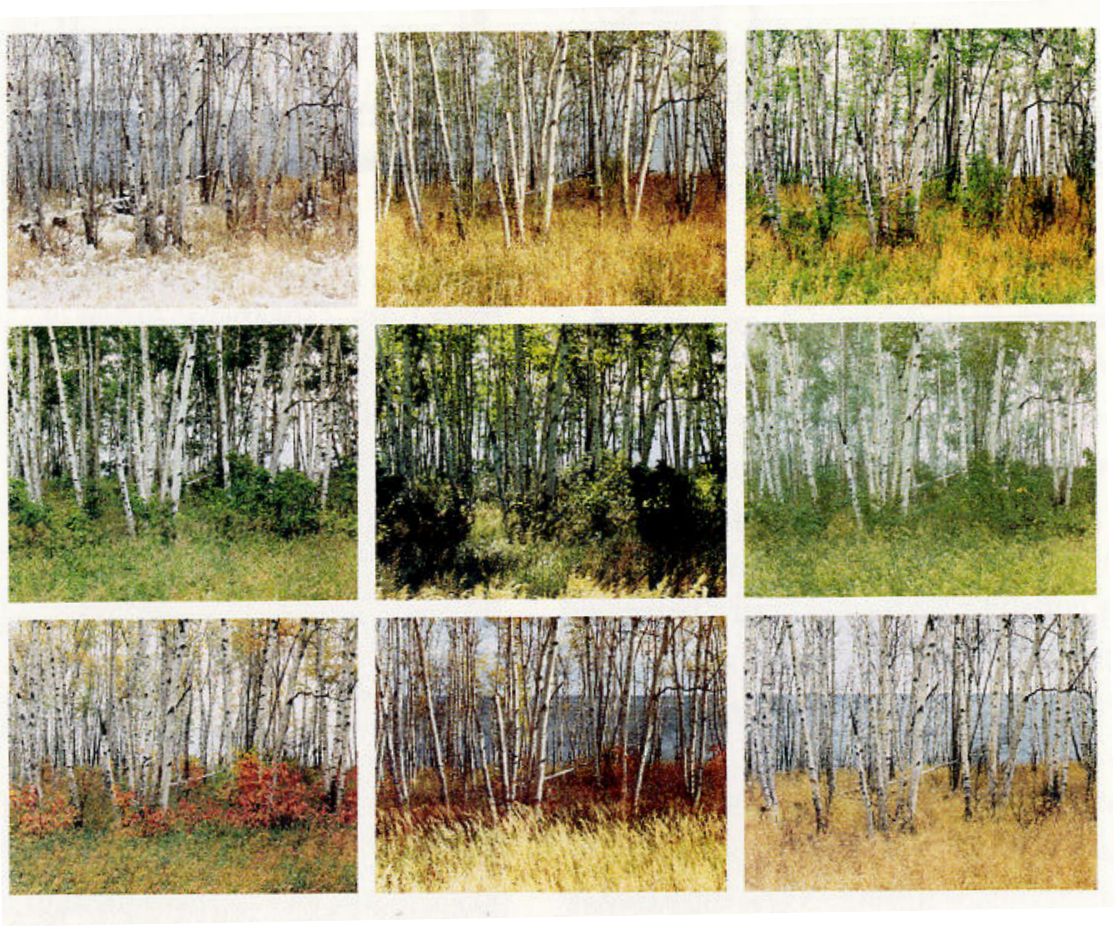






Colour palette

- If you want to group colours, look to nature



Colour palette

- And ‘accent’ colours in Powerpoint
- Colour is extremely potent in grabbing attention
  - Be subtle

Black outline, Red fill	Black outline, Accent 2 fill	Black outline, Dark Blue fill	Black outline, Blue Accent 1 fill
Grey outline, Red fill	Grey outline, Red accent 2 fill	Grey outline, Dark Blue fill	Grey outline, Blue Accent 1 fill
No outline, Red fill	No outline, Red accent 2 fill	No outline, Dark Blue fill	No outline, Blue accent 1 fill

Graph layouts

- Use the 2-page view option in word to get overview of document
- We will cover more of this in Posters