linksHTML7E6DE0

Notes on Tufte's principles

Sam Clifford

2019-10-31

Tufte's principles for visualising data

A quick tour

- Visual representations of data must tell the truth
- Good graphical representations maximize data-ink and erase as much non-data-ink as possible
- Avoid chartjunk: the excessive and unnecessary use of graphical effects in graphs
- Produce high data density graphs

Graphical integrity

Visual representations of data must tell the truth

1. The visual representation of numbers, as physically measured on the graph, should be directly proportional to the numerical quantities represented
    - Lie Factor = the size of the effect shown in the graphic divided by the size of the effect in the data.
    - Values over 1 overstate the effect & under 1 understate.
    - Don't make points bigger to emphasise them as important
2. Clear, detailed and thorough labelling helps avoid graphical distortion and ambiguity. Write explanations of data on the graph. Label important events from data.
3. Show variation through data, not through design.
4. Ensure appropriate standardization & comparisons are used, e.g. CPI-adjusted or seasonally-adjusted.
5. The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
    - In practice, this means not mapping the same variable to two graphical elements
6. Graphics must not quote data out of context.

Maximising `data:ink` ratio

Good graphical representations maximize data-ink and erase as much non-data-ink as possible

1. Above all else, show data
2. Maximise the data-ink ratio
    - ink on a graph that represents data
    - data-ink ratio = 1 minus the proportion of the graph that can be erased without loss of data-information
3. Erase non-data-ink

Figure 1: Improper emphasis of most and least fuel efficient cars. We only need look at vertical range to see these are the min and max

Figure 2: Fuel efficiency mapped to colour and y coordinate. Gradated pattern of colour implies there's another relationship with fuel efficiency being looked at.

Figure 3: Standard boxplot

4. Erase redundant data-ink
5. Revise and edit

Avoid Chartjunk

If an element of a graph is to be included, it should be because it aids understanding and reveals information, not because it looks pretty

- Examples of chartjunk include but are not restricted to:
  - hatching (patterns instead of colours)
  - heavy grids
  - equally spaced lines, too close together, as in bar charts, histograms, boxplots
  - self-promoting graphs that demonstrate the graphical ability of the designer rather than displaying the data
  - 3D graphics that distort perspective (e.g. 3D pie charts)

Maximising data density

1. Maximise data density and the amount of data shown, within reason
2. Apply the shrink principle
   - most graphs can be shrunk down very far without losing legibility or information
   - this is because most graphs are quite sparse
3. Exploit **small multiples** to provide for comparisons across groups
   - series of the same small graph repeated in one visual
   - can compare a main relationship across one or more grouping variables

Figure 4: Tufte-compliant boxplot maximising data ink

Figure 5: High dimensional data set with points coloured by number of forward gears and shape mapped to transmission type

Figure 6: High dimensional data set with small multiples faceting by number of forward gears and transmission type

- a great way to visualise large quantities of data, or when there are a high number of dimensions