Tufte, Chapter 4: Data-Ink

By Z. Trozenski

Introduction and Key Terms

Graphics should represent the substance and sense of the data and not draw attention away from the data and towards the design. You don't want your aesthetics to compete with the information you're trying to present. **Data-ink** refers to the ink dedicated towards representing the actual data or, according to Tufte, "the **non-erasable** core of a graphic, the **non-redundant** ink arranged in response to variation in the numbers represented".
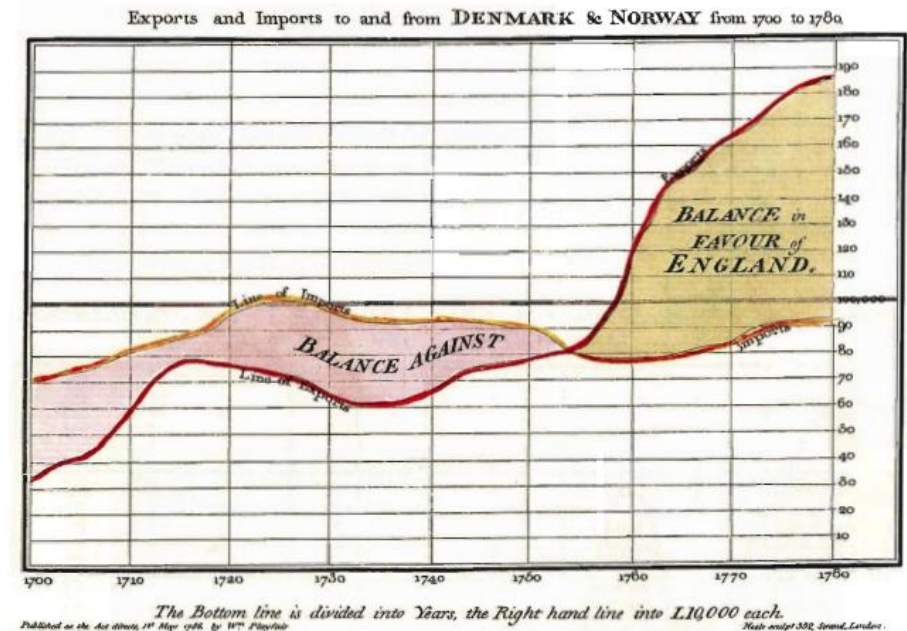
Concepts

#1 – "Above all else, show the data"

- A large share of ink on a graphic should present data-information, changing as the data changes.
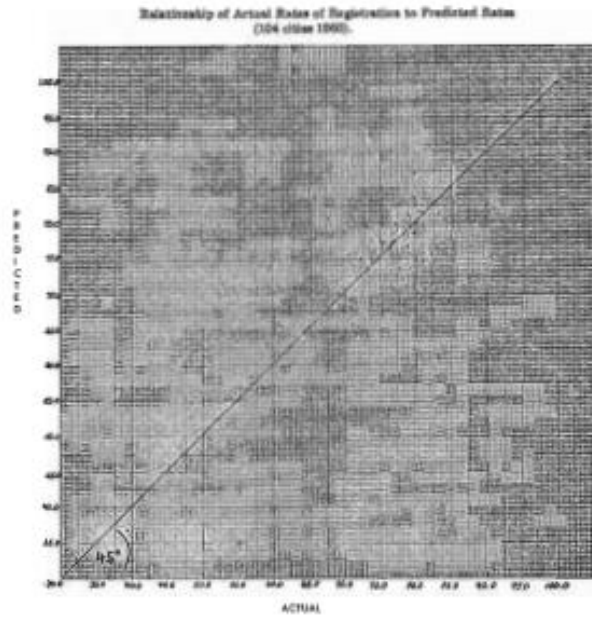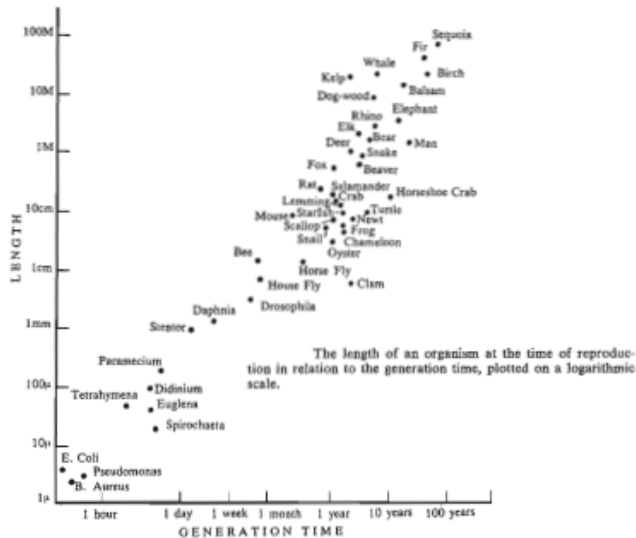
Bad:
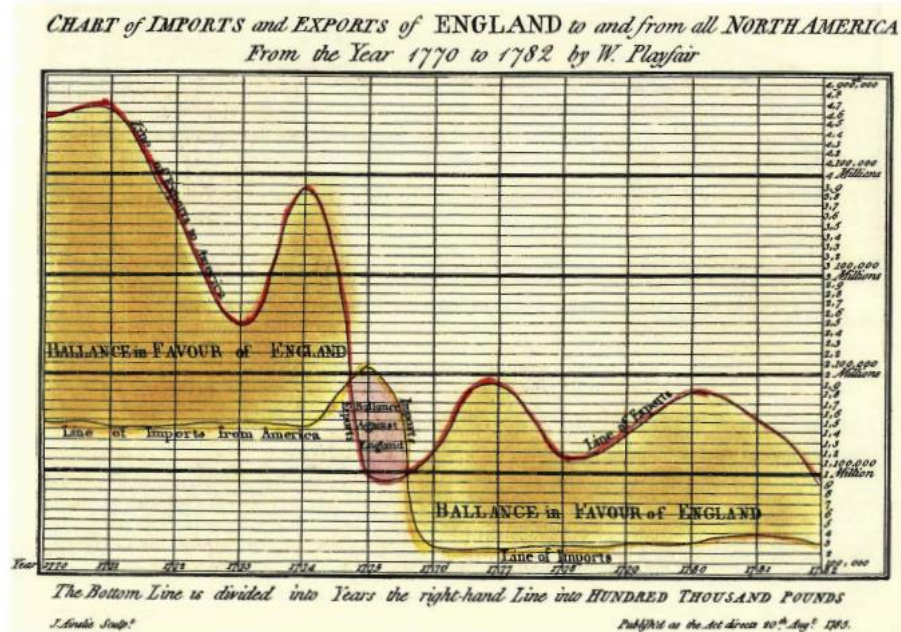
Good:

#2 – "Maximize the data-ink ratio"

- You should devote the most ink to displaying the information found within your data, with as little as possible devoted to non-data (i.e. legends, ticks, axes, etc.)
- Every blot of ink on a graphic requires a reason to be there, the most compelling reason is that the ink shows new information. Better yet, is to think about representing the most amount of data-information with the least amount of ink.
- Tufte's ratio:
  - Draw-ink rato = data ink / total ink used to print the graphic
  - = proportion of a graphics ink devoted to the non-redundant display of data information
  - = 1.0 - proportion of a graphic that can be erased without loss of data information

Bad:

Relationship of Actual Rates of Registration to Predicted Rates
(104 cities 1960).

Good:

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

- When you have ink devoted to items in your graphic which is not explicitly devoted to depicting statistical information it clutters the graphic and muddies its clarity.
- Redundant data-ink:
  - Depicting the same number over and over (i.e. labels shaded bars, legends, etc.)
  - Gratuitous decoration and reinforcement of the data
- Tufte claims that bilateral symmetry is also redundant (though I disagree with this point) in box plots, open bars, faces.

Bad:
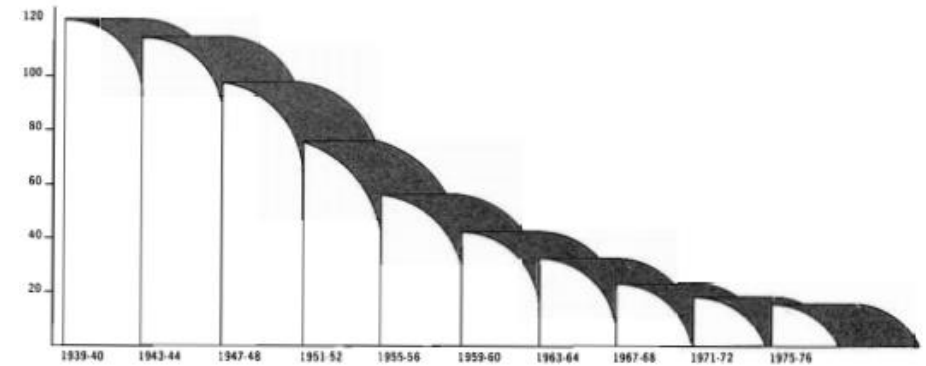
#4 – "Erase redundant data-ink."

- Redundancy does have a usefulness in that in can order complexity and establish context, allowing for comparisons over different parts of the data, and creating an aesthetic balance.
- Sometimes cycles in repeated data allow for a reader to navigate the graphic better without having to start over at the beginning of the graphic.
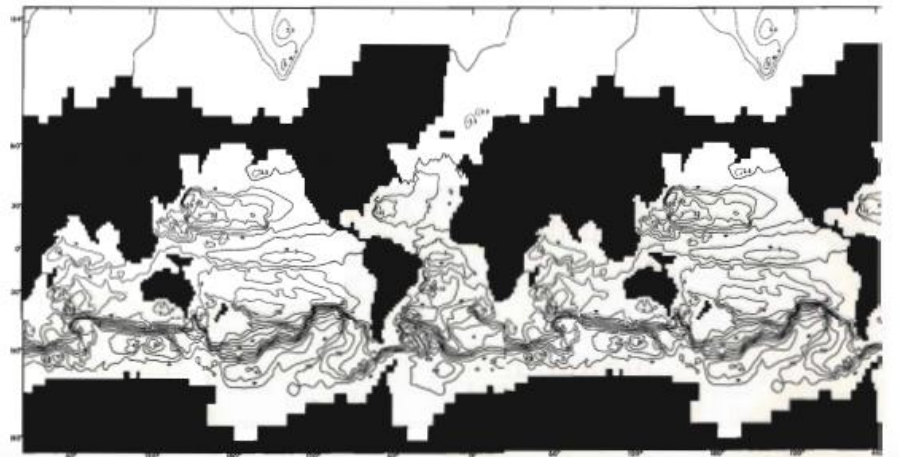
Good:

Bad:
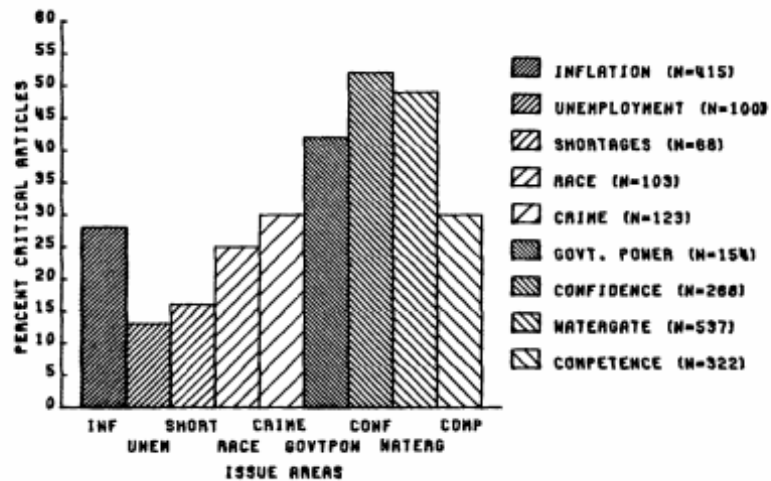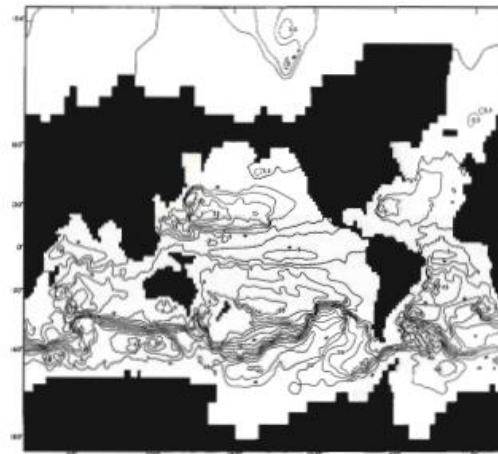
#5 – "Revise and edit."

- You should feel empowered to revisit your graphics and prune out unnecessary data.
 Some best practices to consider:
  - o Over multiple iterations of the graphic, remove different elements of reference ink (gridlines, precisions marks, axes, etc.) one at a time and examine the graph.
    If the removed element creates confusion for the reader either by revoking context or establishing order, then that element should remain.
  - o Any element, which when removed, is not hindering the comprehension of the data can be left out. This may also lead to more efficient uses of reference ink.
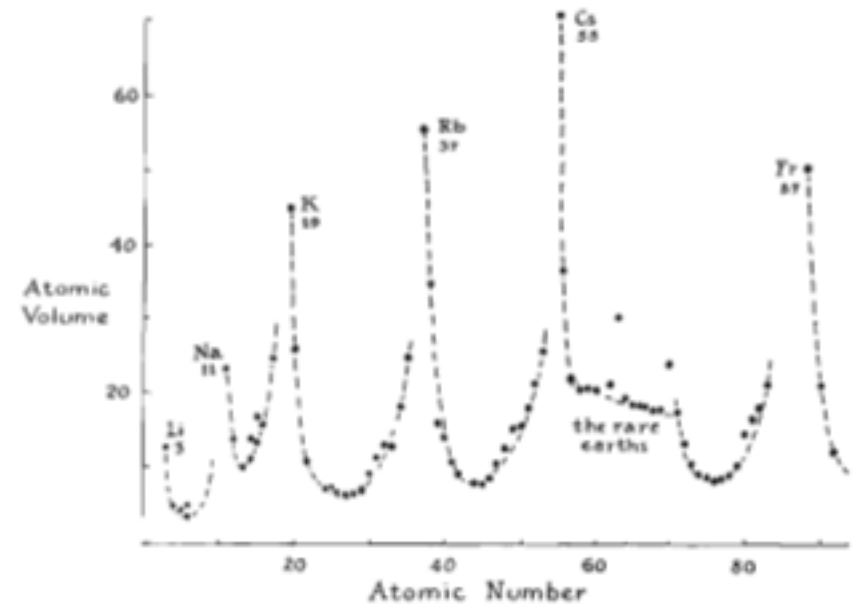
Example:
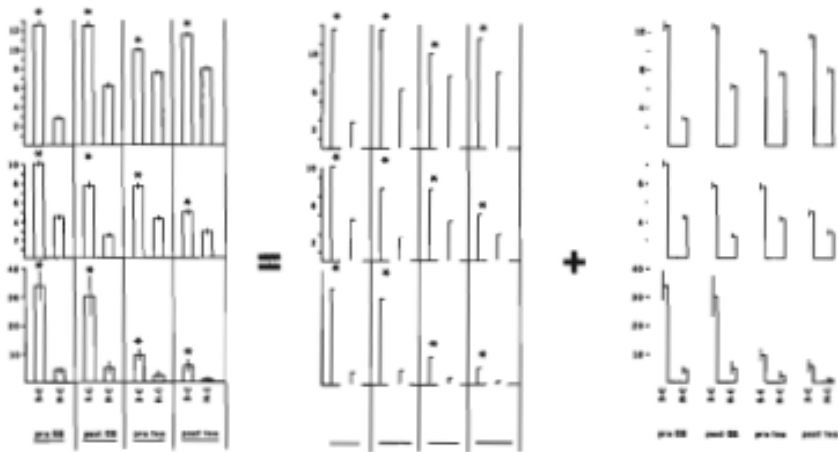
In this example we see on the far left, the original graph. Immediately to the right of the "=" sign is the ink Tufte considers unnecessary. To the right of the "+" is all the ink Tufte deems necessary to illustrate the data.

Before:
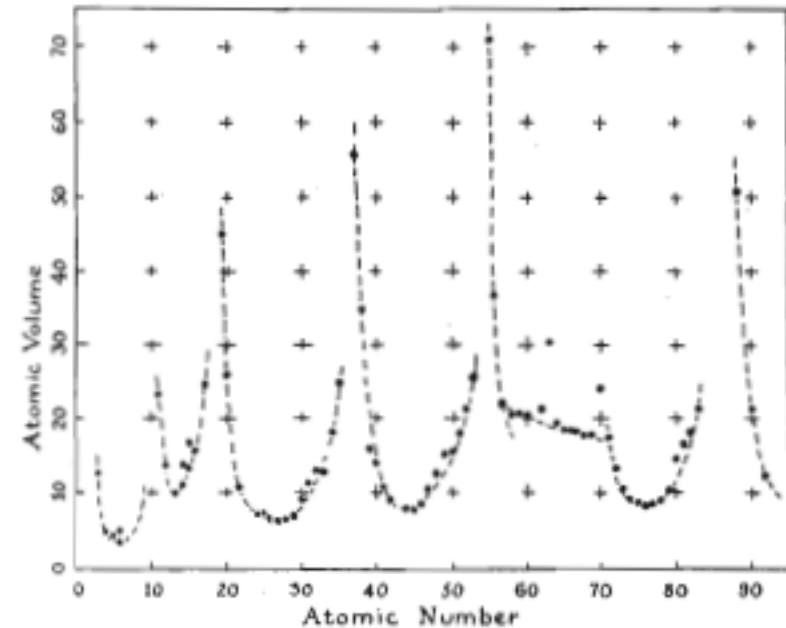
After:

Activity:
Please click through to this Google Form to complete an evaluation and ranking task based on the principles we've just discussed.

Link: https://forms.gle/VRLQVduPEAXfKbKR6

References:
Tufte, E. R. (2013). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

http://www-personal.umich.edu/~jpboyd/eng403_chap2_tuftegospel.pdf

Thanks for listening! Here's a picture of my cat: