Cleveland’s hierarchy

Cleveland’s Graphical Features Hierarchy

Best information transfer

* * *
Position along a common scale: bar charts, time series, scatter plot

* * *
Position along nonaligned axis: ?????

Length: stacked bar charts, waterfall charts

Angle / Slope: pie charts

Area: pie charts, matrix charts, radar charts

Volume: 3d charts

Worst information transfer

Colour

Scales

- The top of the hierarchy involves putting things on scales
- But what scale do we use?
  - Are our data anchored to zero?
    - If so, are we interested in differences or ratios?
  - Are they anchored somewhere else?

1 Anchors

Utter Problem Algae
Global temperature
Global temperature
Climate lessons

- Choosing an anchor is a scientific decision
- Remember: graphic design is communication

Magazine circulation (advertisement)

Magazine circulation (absolute amount)
Magazine circulation (trend)

Area and volume
How to Lie with Statistics

Advertisement lessons

- Use area to indicate fair comparisons
  - On a physical scale

- Areas that can be compared linearly should be preferred
  - Depends on importance of feature

- Avoid using (or hinting at) volume

2 Transformations

Physical quantities

- 1 is to 10 as 10 is to what?
  -
  -

- The log scale is often good for physical quantities:
  - When zero means zero
Log vs. linear

Making room

Data shape

- There are a lot of different ways to show data shape
- Choices will depend on your data set:
  - Overall size
  - Number of replicates
- Number of levels, predictor variables, etc.

**Showing points**

![Boxplot]

**Boxplot**

![Violin plot]

**Violin plot**
Orchard lessons

• Choices about log vs. linear scale are scientific choices
  – Neither is more valid, or closer to the data

• You can also make choices about
  – sending a simple message
  – providing more information about shape

• Log scales are almost never physical
  – Don’t mislead with area information on a log scale

Probabilities

• 1% is to 2% as 50% is to what?

  –

  –

• The natural distance to use on a probability scale is log odds

  –

  –

  –
Odds

- Odds are a ratio between the probability of something and the probability of its opposite:
  \[ o = \frac{p}{1 - p} \]

- Log odds give a natural distance on probability space

Extreme values

- Our transformations take extreme values to infinity.
- Use link functions: this is like using estimated values instead of observed; they are rarely infinite
- Extend the scale (e.g., use \( \log(1 + x) \) instead of \( \log(x) \))
  - This usually involves arbitrary choices
  - Should often be avoided for analysis
  - Usually OK for visualization