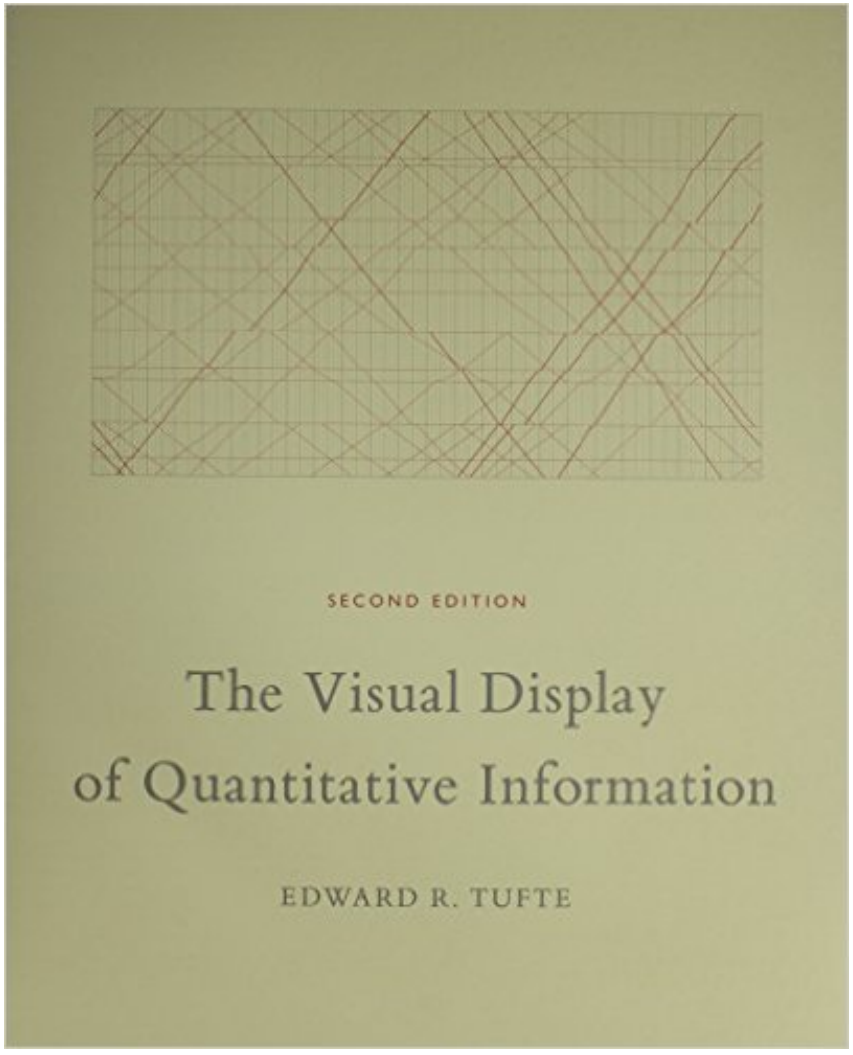
Bio 204: Biological Data Analysis

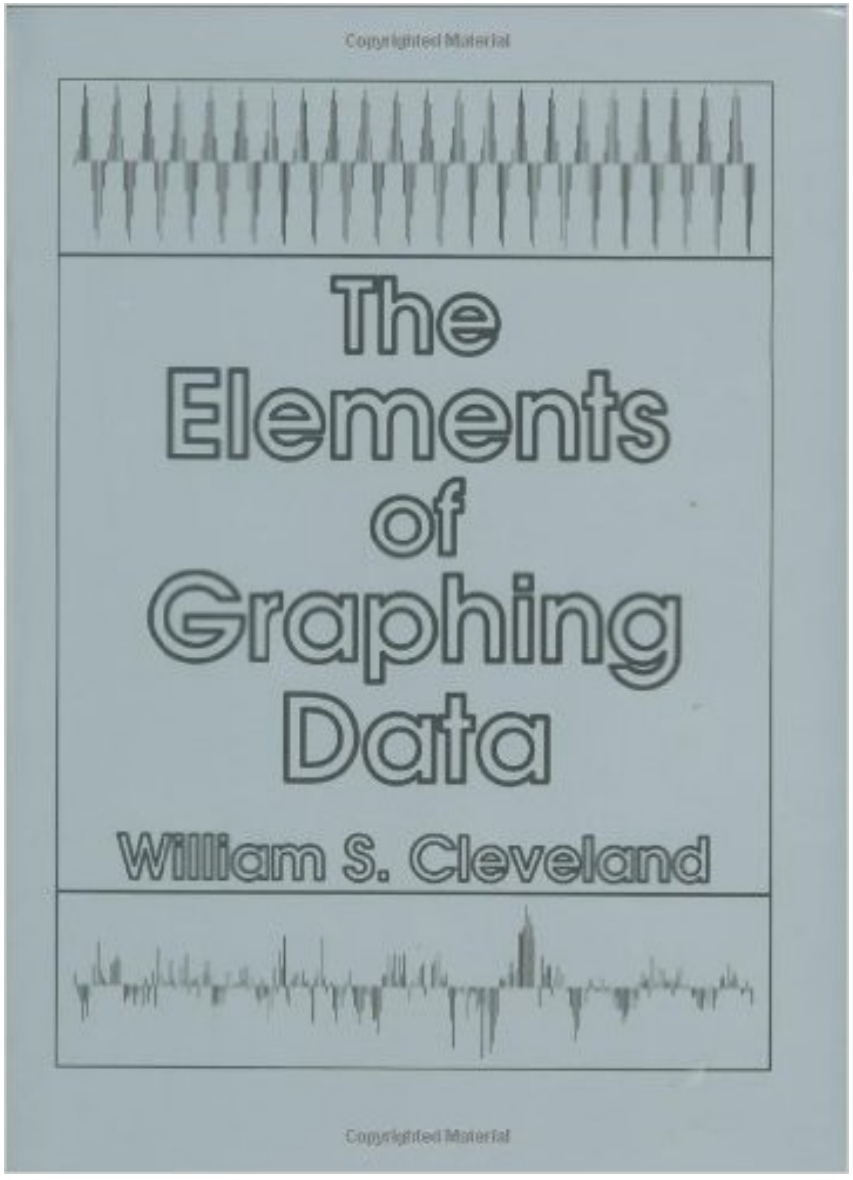
Introduction to Visualization

Instructor: Paul M. Magwene

Fall 2016

Book recommendations





Statistical graphics should...

Paraphrasing Tufte, The Visual Display of Quantitative Information, Chapter 1:

- Show the data
- Induce the viewer to think about the substance of data, analysis or model
- Avoid distorting what the data have to say
- Present complex information compactly
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail
- Serve a purpose
- Be closely related integrated with the statistical and verbal descriptions of a data set

What visual representations do we have at our disposal when creating statistical graphics?

Perceptual hierarchy for quantitative information ala Cleveland and Williams

- Position along a common scale
- Positions along nonaligned, identical scales
- Length
- Direction, angle
- Area
- Volume, curvature
- Shading, color saturation

Judgements of angle and area are perceptually difficult

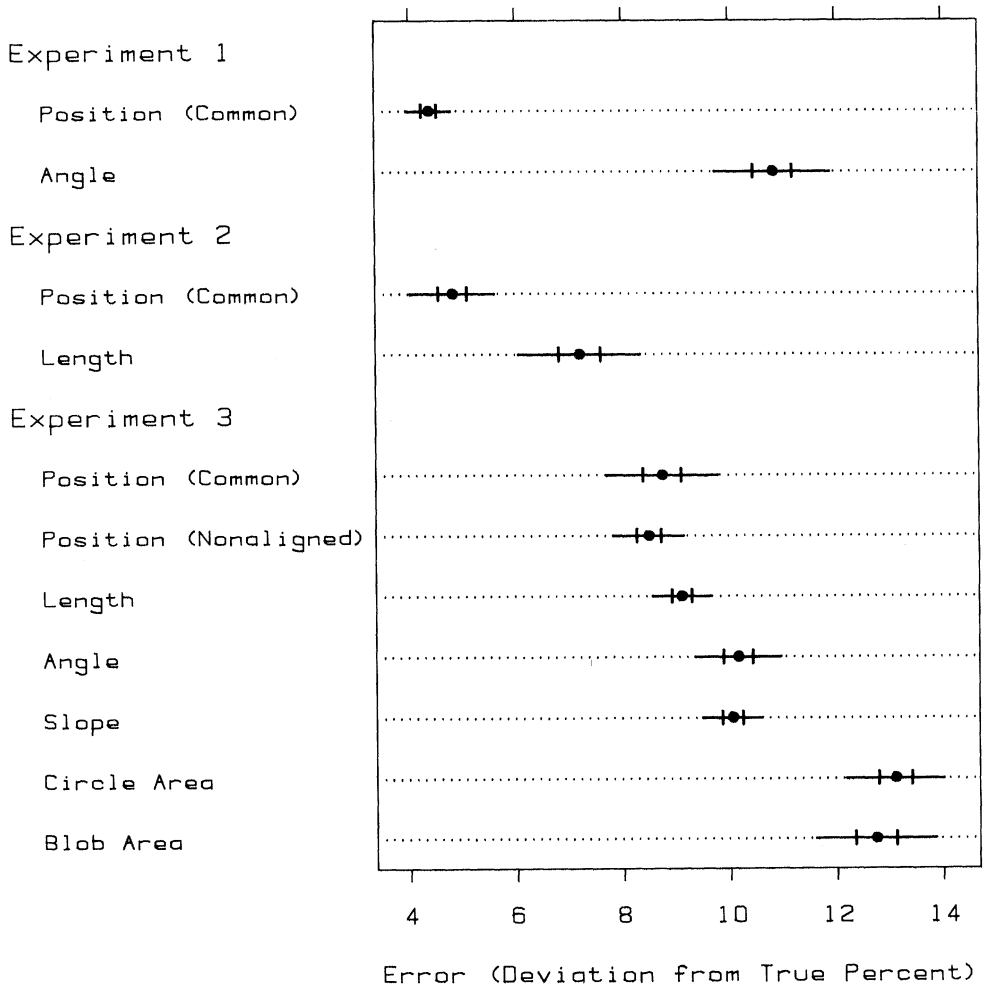


Figure: Figure from Cleveland and Williams (1985), demonstrating the judgement error associated with different types of visual percepetion.

Q: Why do pie charts suck?

A: They rely on both angle and area for intepretation.

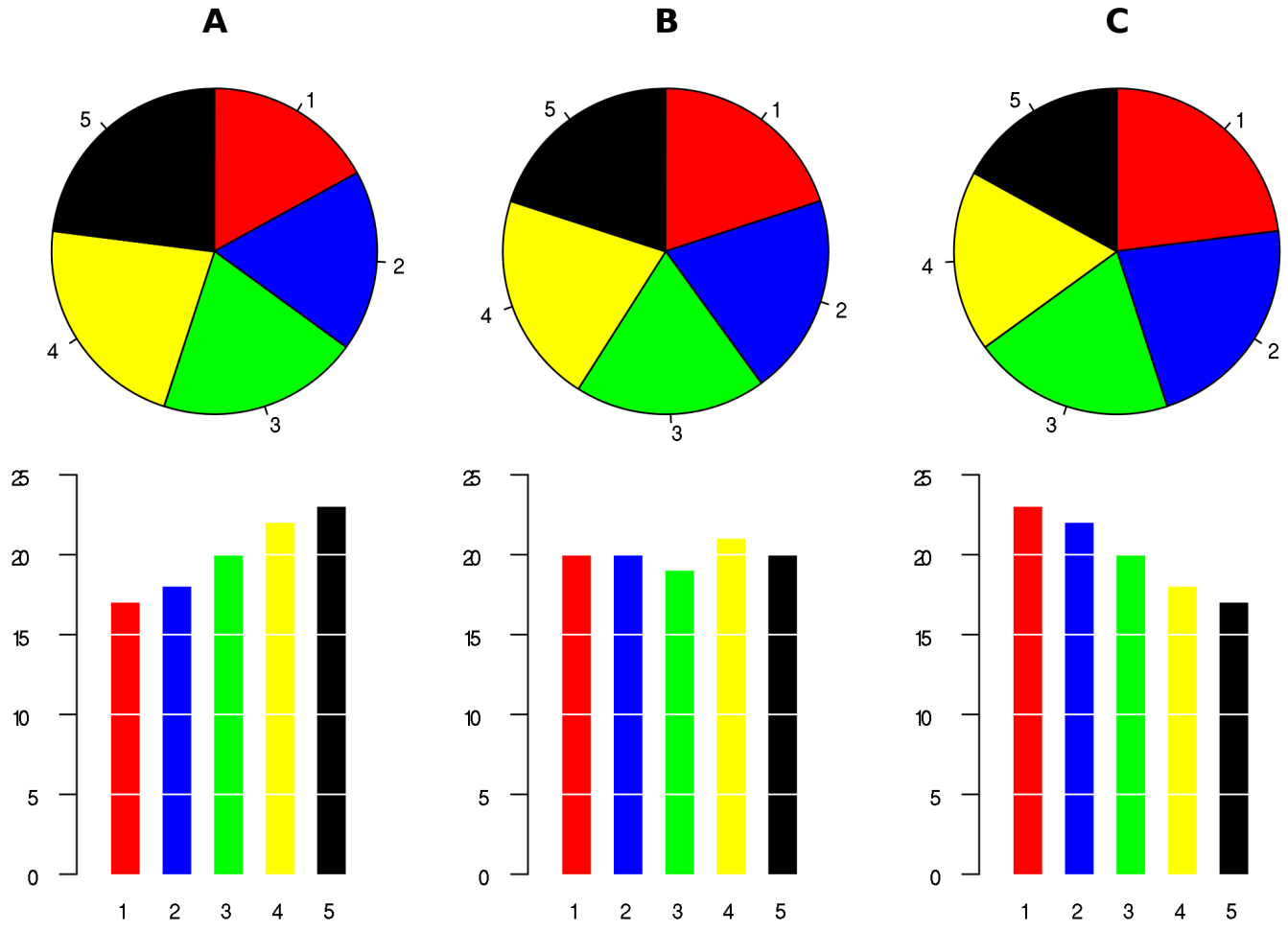
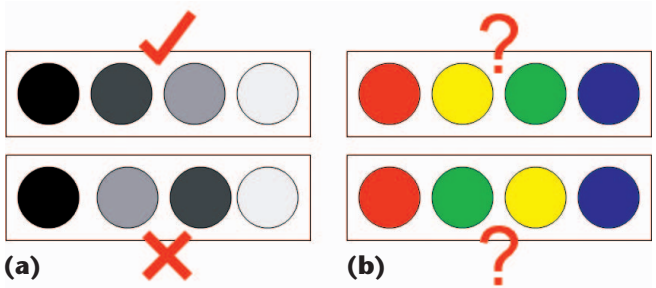
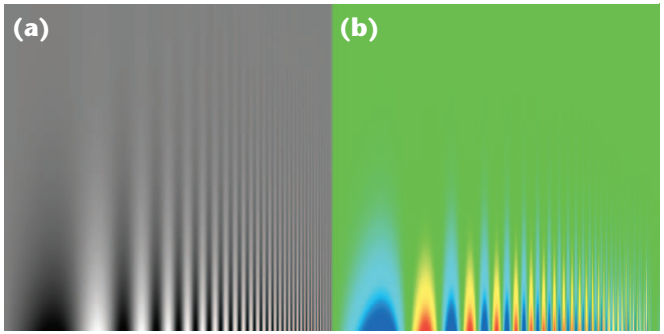


Figure: Pie charts are perceptually hard to interpret. Image from Wikipedia. This color scheme hurts my eyes.

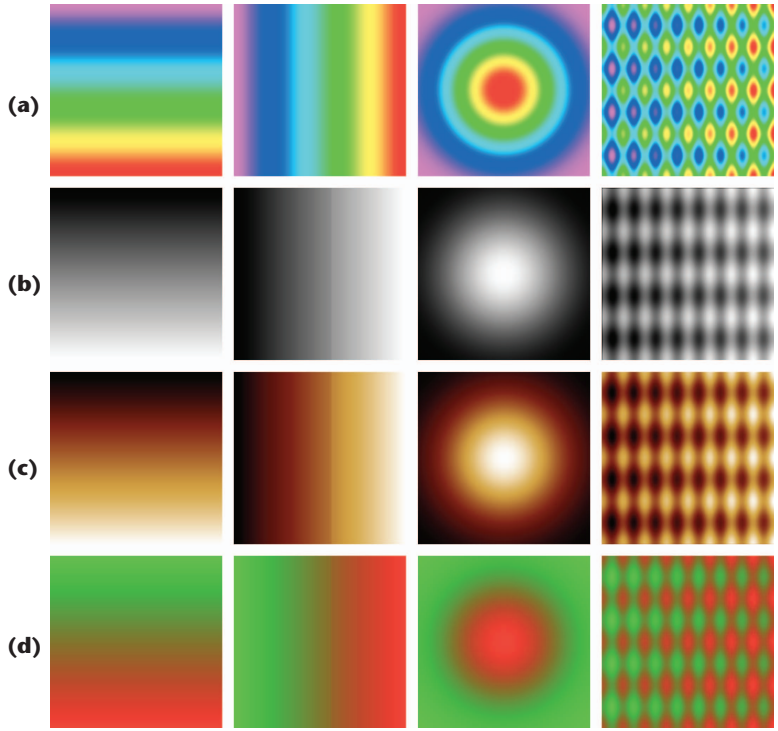
Ban the 'Rainbow' color scheme from your visualizations!



1 Perceptual ordering. (a) We can easily place the gray paint chips in order based on perception, (b) but cannot do this with the colored chips.



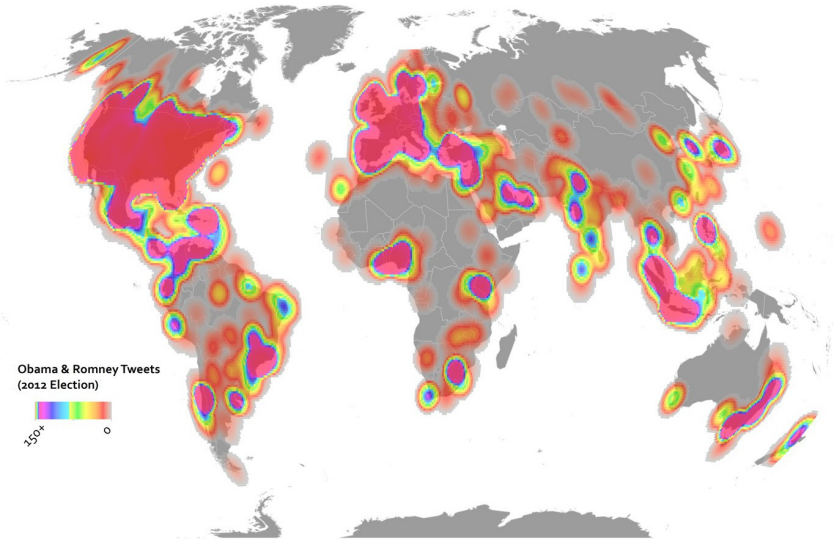
2 Spatial contrast sensitivity function. Frequency increases to the right and contrast increases toward the bottom of both images in the figure. We can see detail at much lower contrast in the (a) luminance-varying gray-scale image than with the (b) rainbow color map.

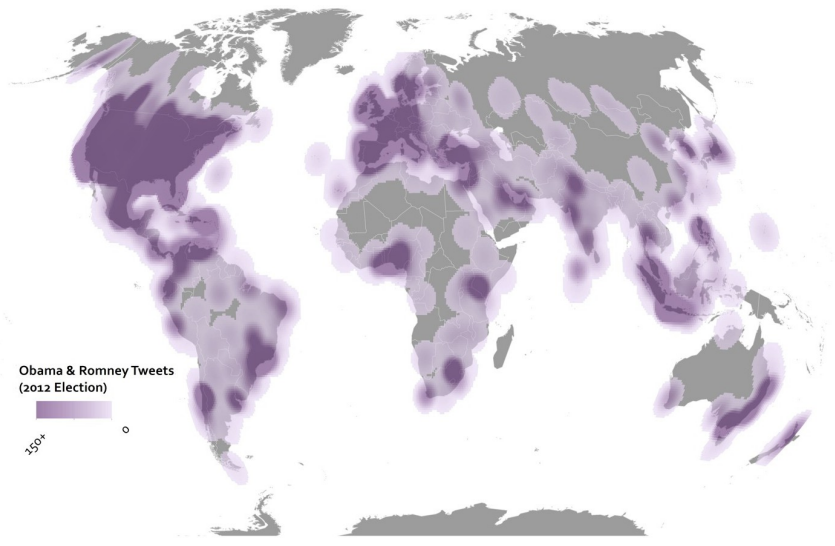


3 Four data sets visualized with (a) rainbow, (b) gray-scale, (c) black-body radiation, and (d) isoluminant green–red color maps. Apparent sharp gradients in the data in (a) are revealed as rainbow color map artifacts, not data features, by comparing this row with the same data viewed using the other color maps. Conversely, the sharp gradient found at the center of the second data set (see the second column) shown in the gray-scale and black-body radiation (and to a lesser extent, the isoluminate green–red) images is not found in the corresponding image with the rainbow color map.

Figures from Borland and Taylor (2007), "Rainbow Color Map (Still) Considered Harmful", IEEE Computer Graphics and Applications.

Same data, two different color schemes





Figures from Anthony C. Robinson, Penn State via this link.

Well designed color schemes from ColorBrewer

http://colorbrewer2.org

The ColorBrewer color schemes are available in R via the RColorBrewer package.

Get rid of chartjunk

- Tufte on chartjunk
- A presentation on improving data-ink ratios

https://speakerdeck.com/cherdarchuk/remove-to-improve-the-data-ink-ratio

Other good sources of data visualization advice

- http://guides.library.duke.edu/datavis
- http://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts