Statistical Graphics

Good and Bad

M. Huebner, November 2013

Ways to visualize two numbers: 4 and 8

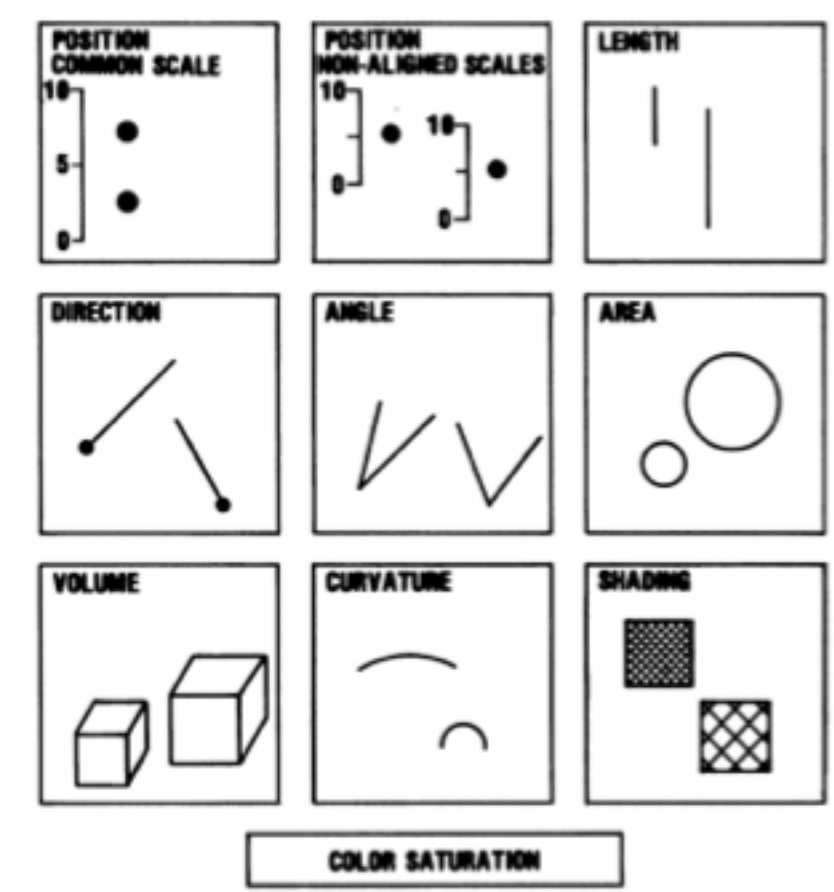


Figure 1. Elementary perceptual tasks.

Cleveland W and McGill R. Graphical Perception: Theory, experimentation and applications to the development of graphical methods. JASA 79 (387): 531-554; 1984.

Position and length

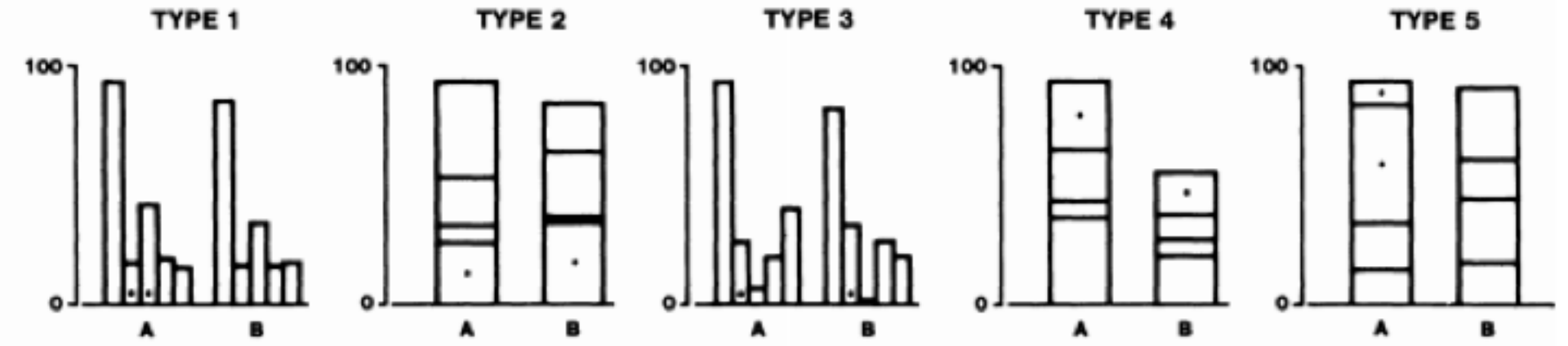


Figure 4. Graphs from position–length experiment.

Cleveland W and McGill R. Graphical Perception: Theory, experimentation and applications to the development of graphical methods. JASA 79 (387): 531-554; 1984.

Position versus angle

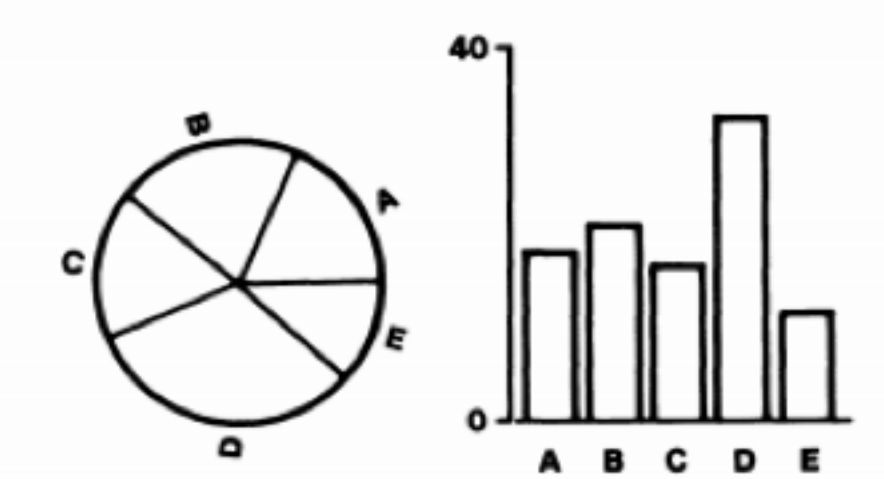
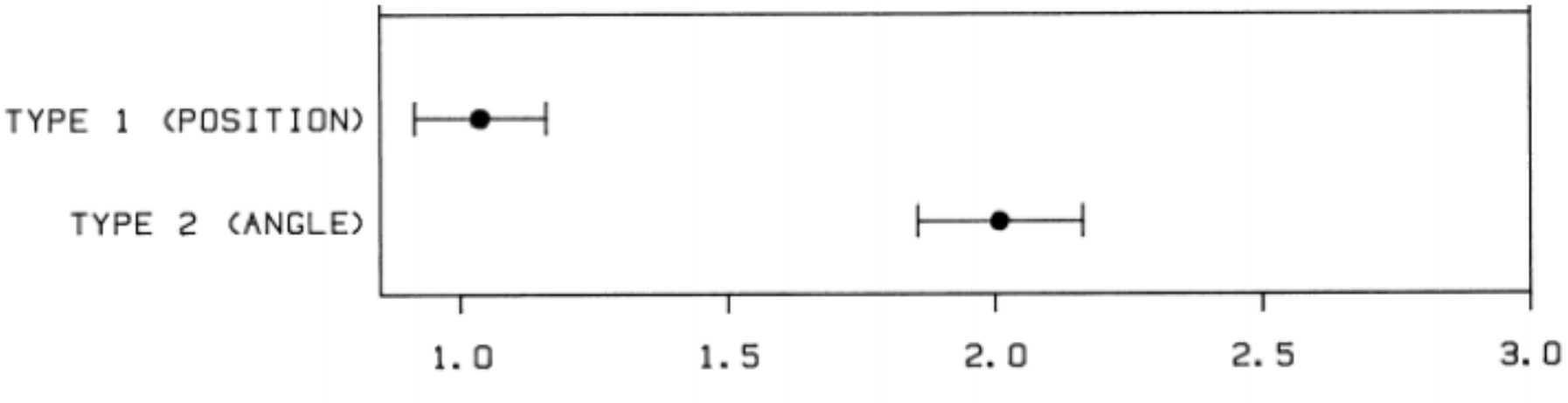


Figure 3. Graphs from position–angle experiment.

Cleveland W and McGill R. Graphical Perception: Theory, experimentation and applications to the development of graphical methods. JASA 79 (387): 531-554; 1984.

Accuracy of judgments: log absolute error and 95% confidence intervals

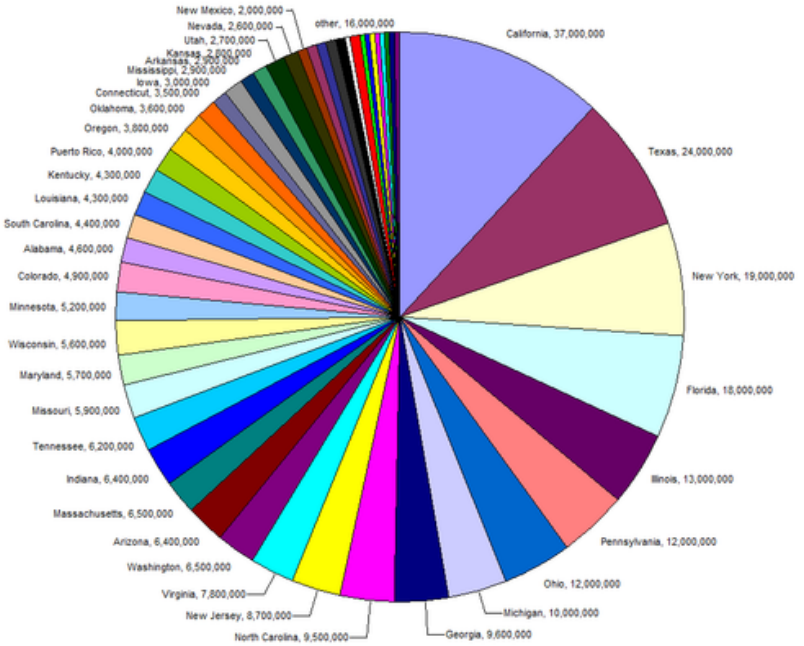


For each graph subjects were asked to indicate which (percentage, segment, ..) was smaller as a "quick visual judgment".

Accuracy= $\log_2$ (|judged % - true %| *0.125)

Pie charts

- Humans have a hard time judging relative angles.
- Pie charts can only use few data points.
- They can “dumb down” the message.



excelcharts.com

Pie charts

Takes lots of space for little information

Don't:

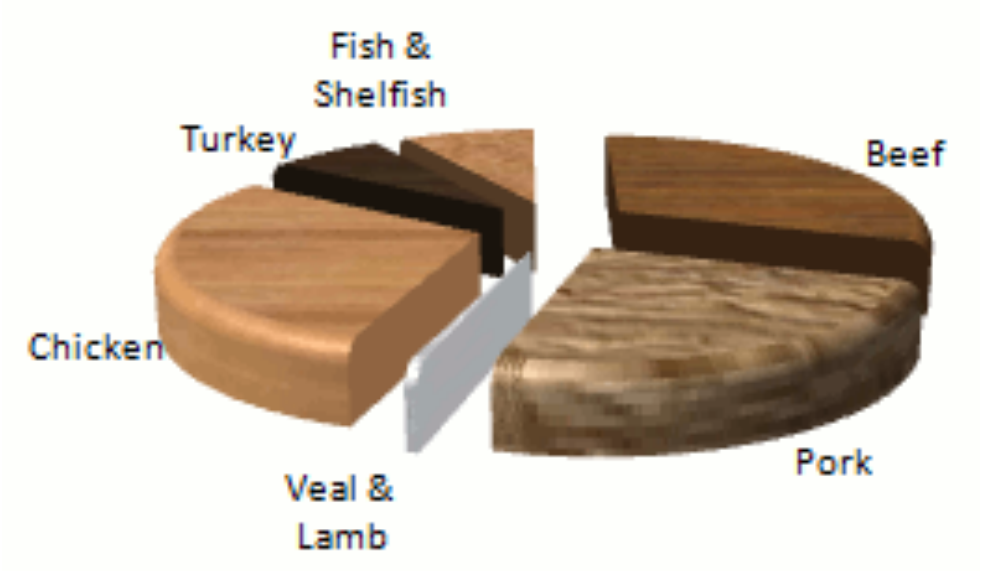
- It's difficult to judge one angle, but even harder to compare angels: no comparison of a series of data.
- 3D effects distort the true proportions

Do:

- Group slices into meaningful chunks
- Label the slices
- Gray out ("mute") small slices

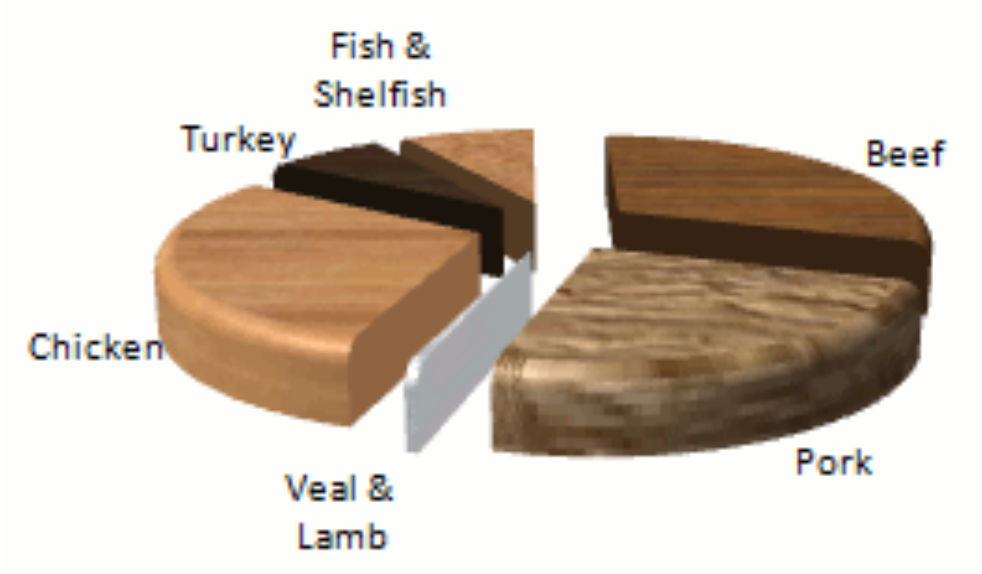
"Save pies for desserts." Stephen Few

“Exploded” pie charts



- The slice “Pork” seems to be the largest.
- Since the chart is “exploded”, it’s even harder to compare angles because they no longer share the same center.
- 3D and horizon!

“Exploded” pie charts



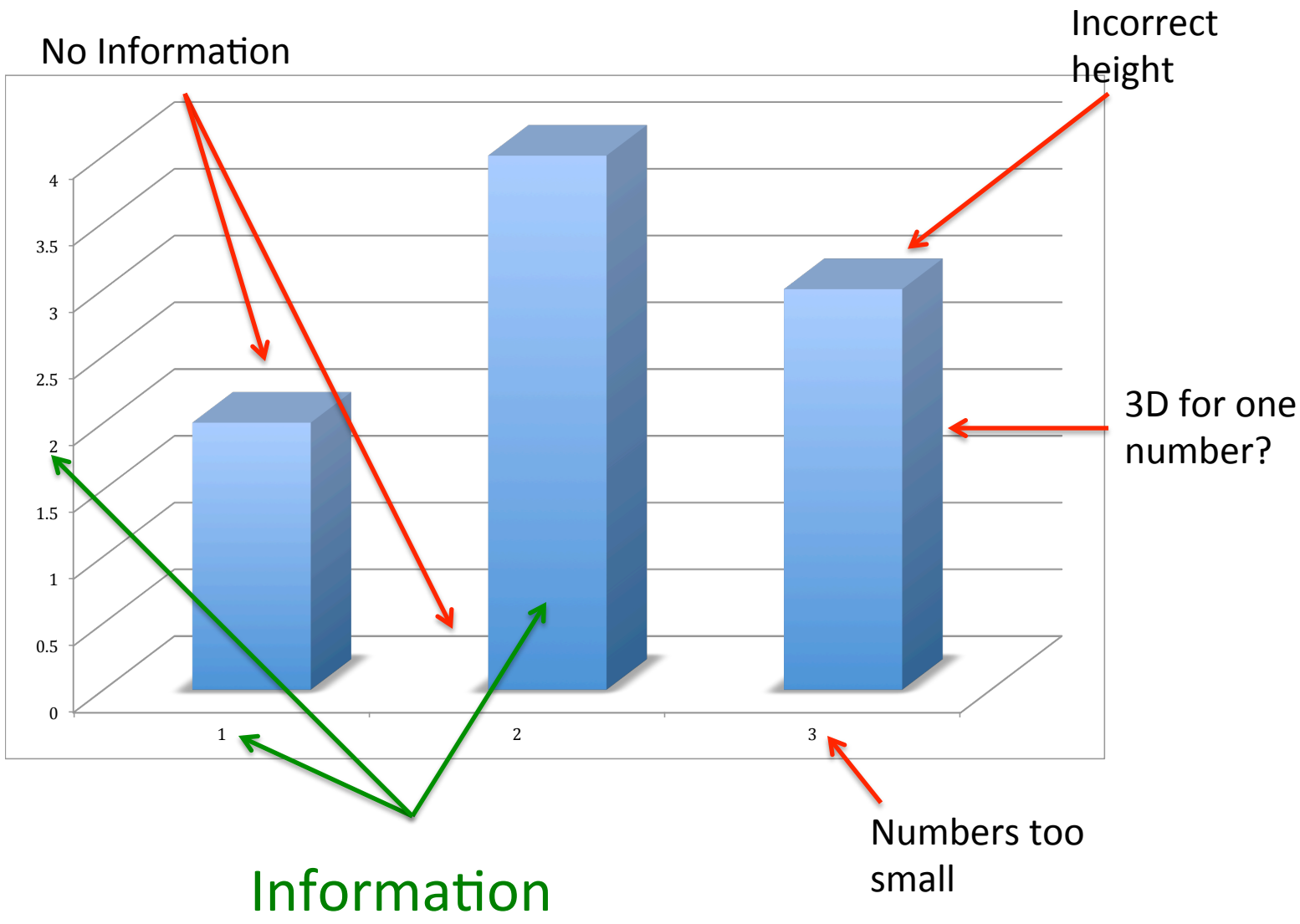
- The slice “Pork” seems to be the largest.
- Since the chart is “exploded”, it’s even harder to compare angles because they no longer share the same center.
- 3D and horizon!
- Pork is not the largest slice.

excelcharts.com

Avoid unnecessary complexity

- Irrelevant decoration
- 3D effects
- “chartjunk”
- color

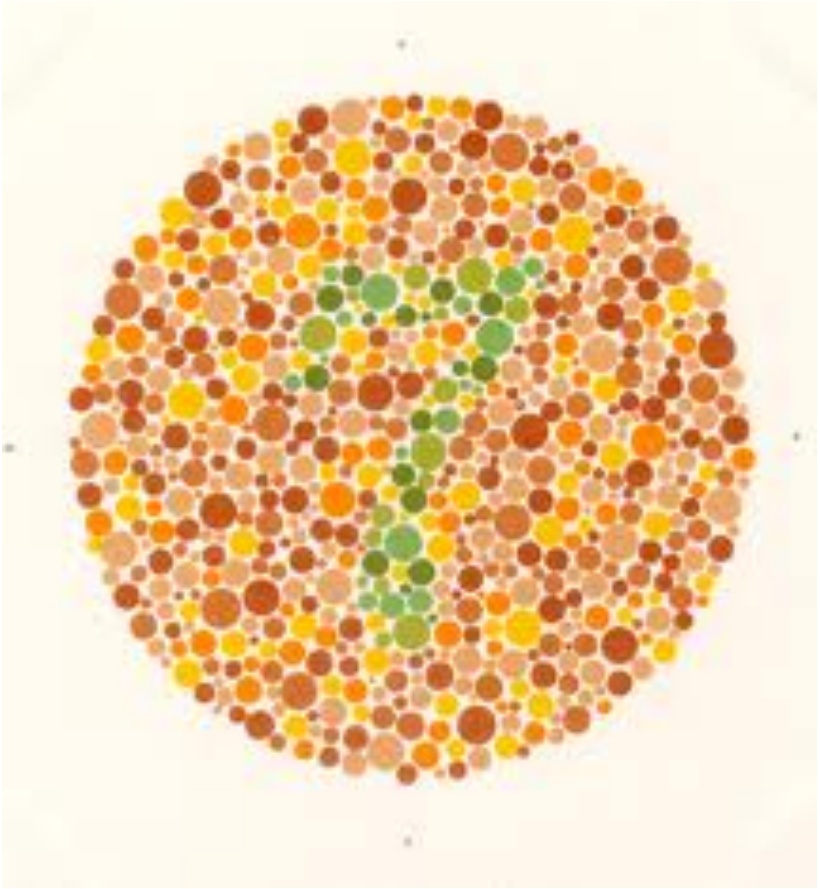
Chart junk


No Information
Incorrect height
4
3.5
3
2.5
2
1.5
1
0.5
0
1
2
3
3D for one number?
Numbers too small
Information

Color

- Humans do not naturally order color hues
- Only a limited number of hues can be discriminated in one graphic
- Optical illusions can be caused by:
  - hues, e.g., red is emotional. A red area may be perceived as larger.
  - shading; larger regions appear to be darker
  - orientation of pie chart with respect to the horizon

Color blindness



Shinobu Ishihara, a Japanese opthalmologist, devised a testing system for color vision abnormalities. The original plates were hand-painted and had hiragana symbols as the patterns, a later ‘international’ edition used arabic numerals.

Color blindness

- In R: >usepackage(dichromat)
- http://www.colorbrewer.org has ratings of the palettes for visibility under a variety of conditions, including red-green color blindness (e.g. use orange, blue, gray)

Common sources of distortion

- Use of 3D effects
- Linear scaling when displaying volume

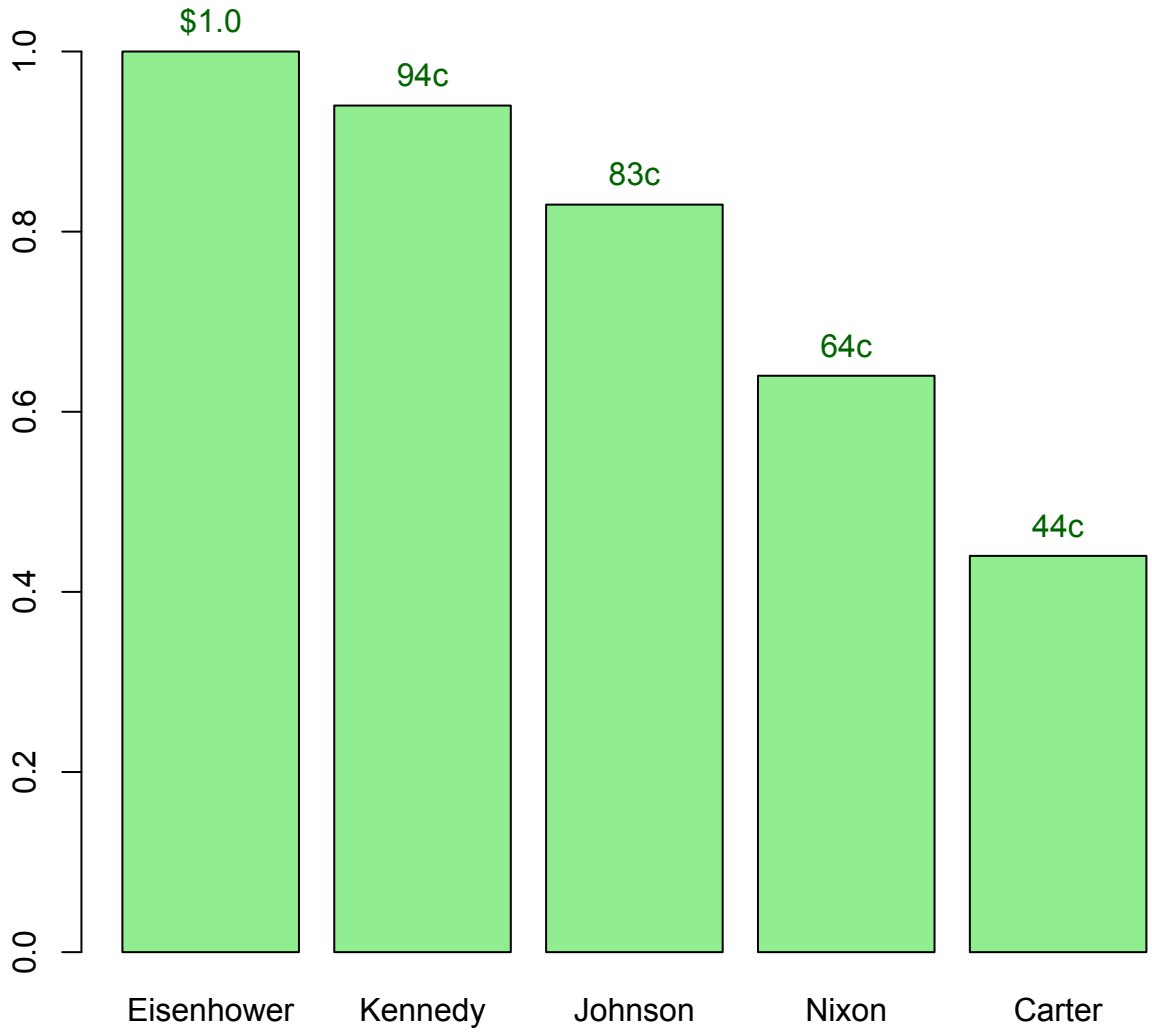
Inappropriate use of scaling when using areas or volume

The smallest bill, one 1978 dollar, is worth 44 cents of a 1958 dollar.

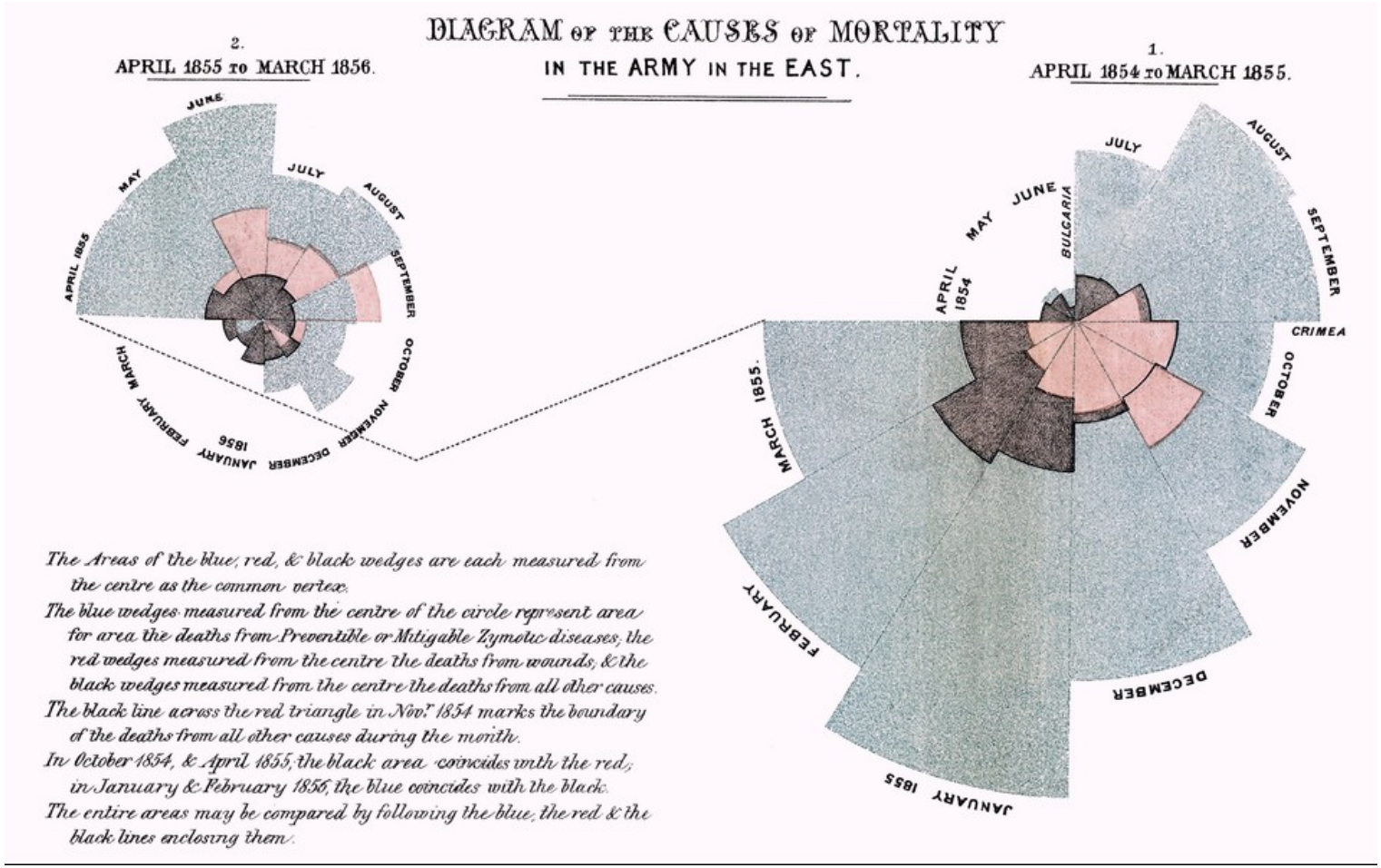
Its length is, correctly, 44% of the biggest one that represents a 1958 dollar.

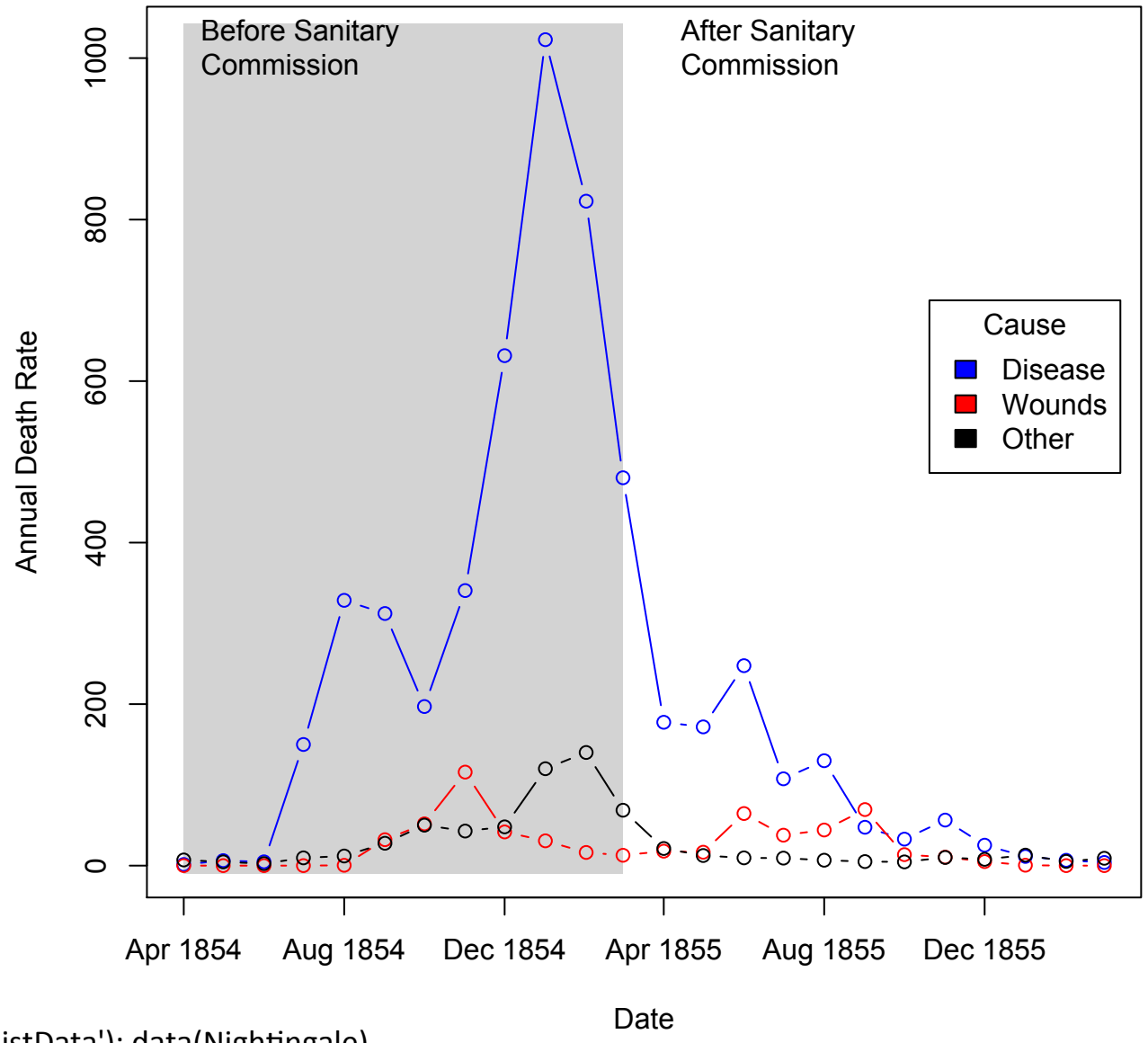
Its area is, incorrectly, 25% of the biggest dollar.

Books:
How to lie with charts by Gerald Jones
How to lie with Statistics by Darrell Huff
How to lie with maps by Mark Monmonier

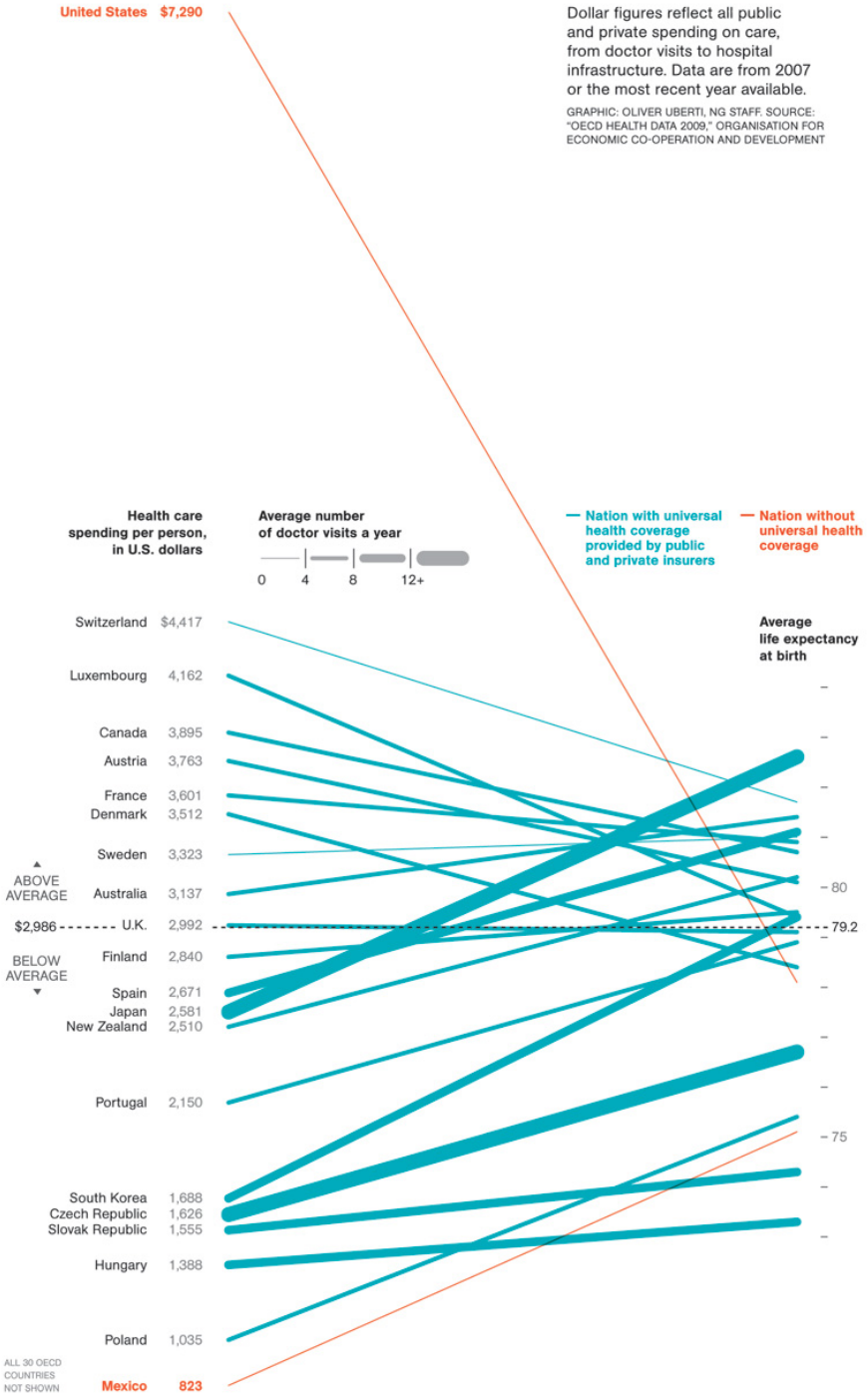

$1.0
94c
83c
64c
44c
0.0
0.2
0.4
0.6
0.8
1.0
Eisenhower
Kennedy
Johnson
Nixon
Carter

Florence Nightingale

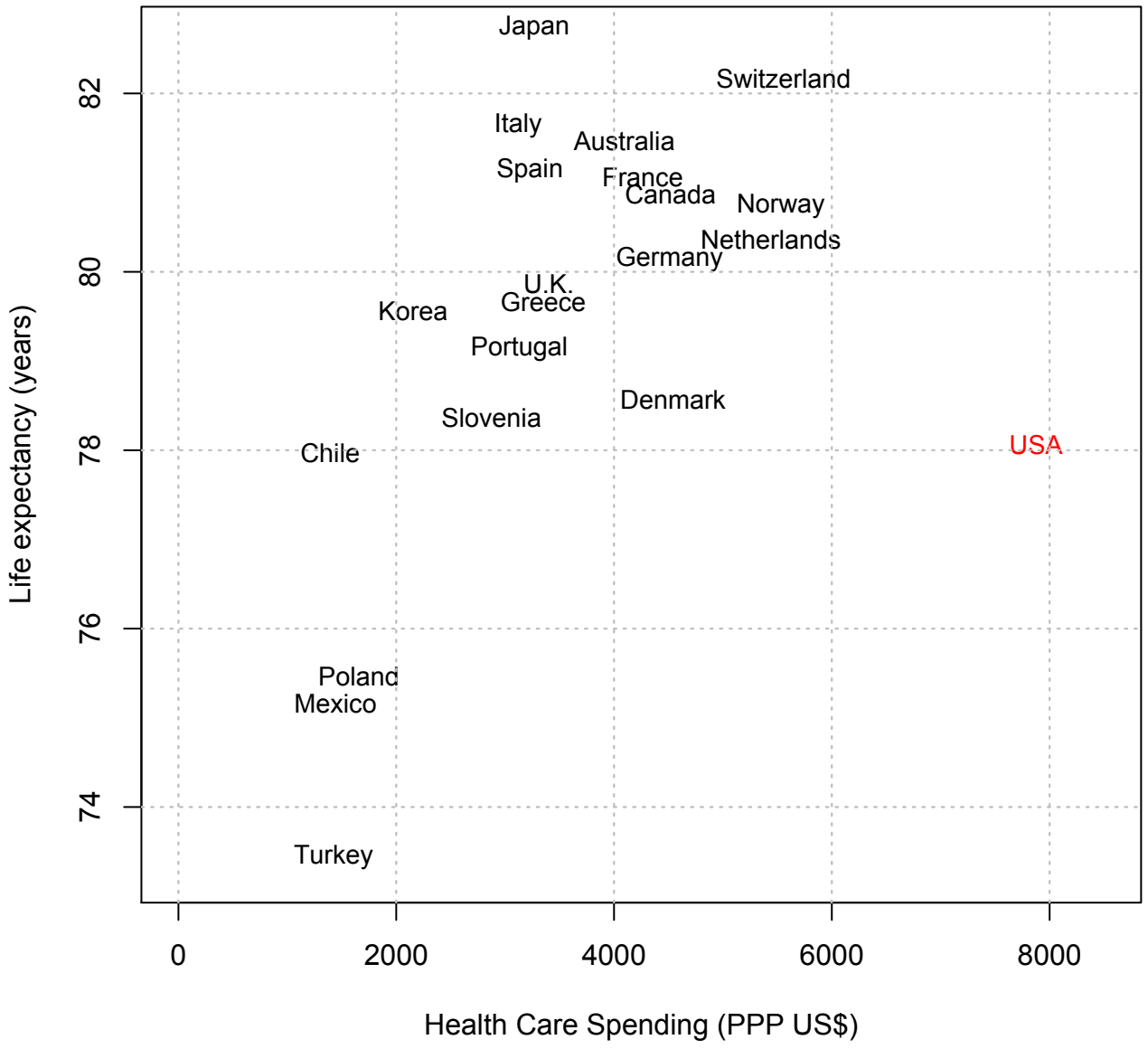




In R: library('HistData'); data(Nightingale)



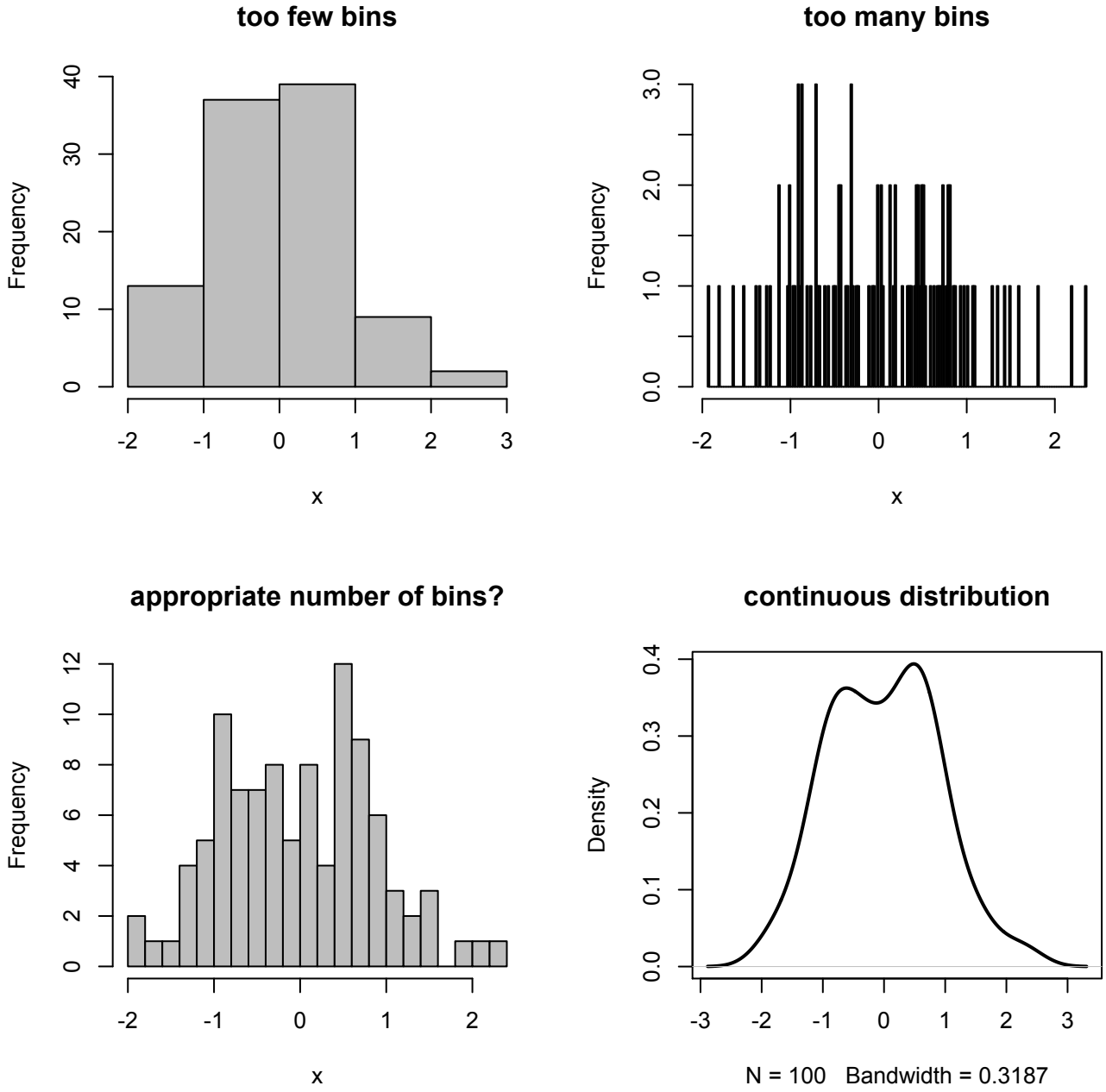
Uberti, Oliver. (2011) "The Cost of Care" in National Geographic



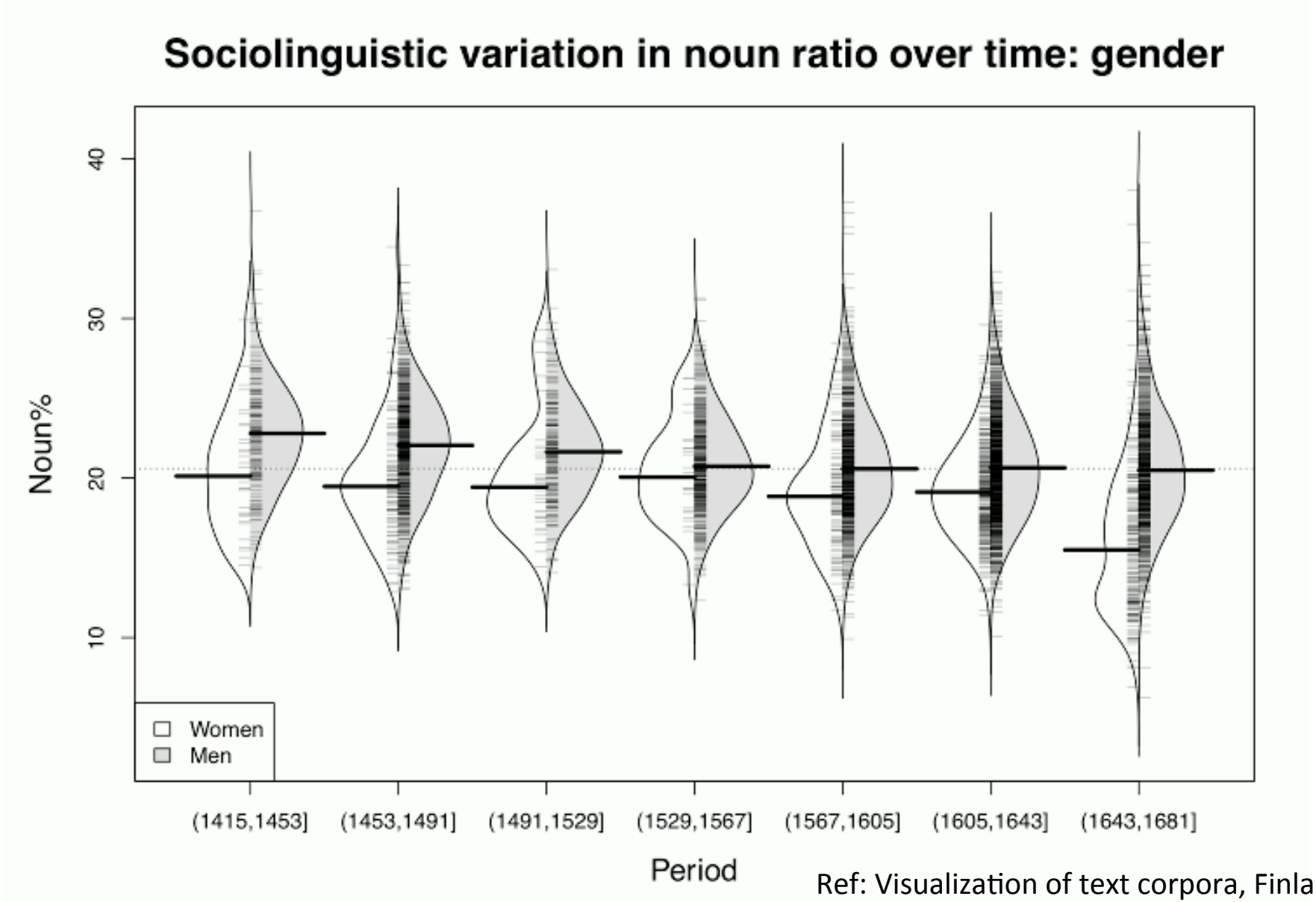
Version of A.Gelman's suggestion

Histograms

- Common choice for univariate, continuous data
- Categorizes continuous data
- 2D figure for 1D data
- Very dependent on how bins are formed, and number of bins
- Do density plots instead?


too few bins
Frequency
0
10
20
30
40
-2
-1
0
1
2
3
x
too many bins
Frequency
0.0
1.0
2.0
3.0
-2
-1
0
1
2
x
appropriate number of bins?
Frequency
0
2
4
6
8
10
12
-2
-1
0
1
2
x
continuous distribution
Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
N = 100 Bandwidth = 0.3187

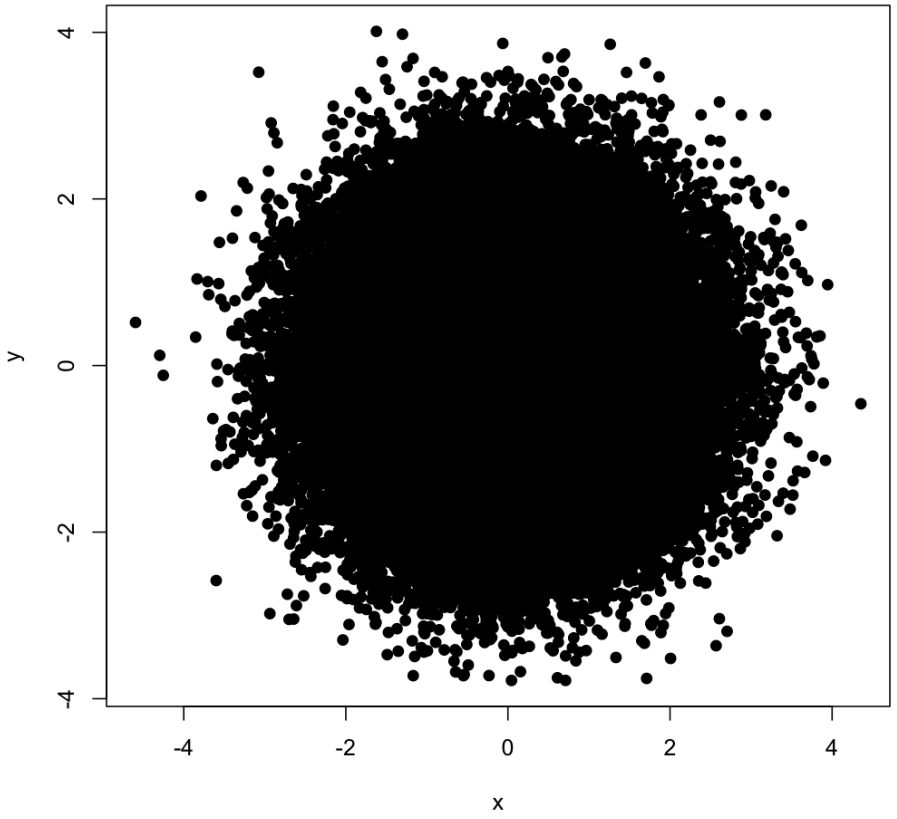
Alternatives to boxplots



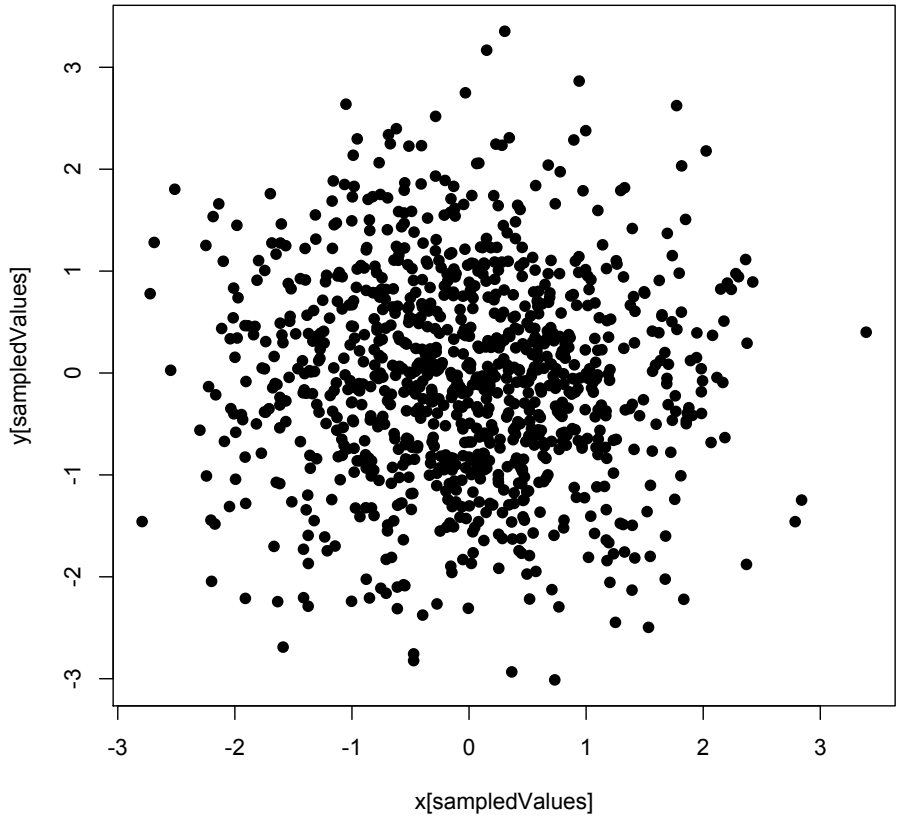
Ref: Visualization of text corpora, Finland

Large amounts of data

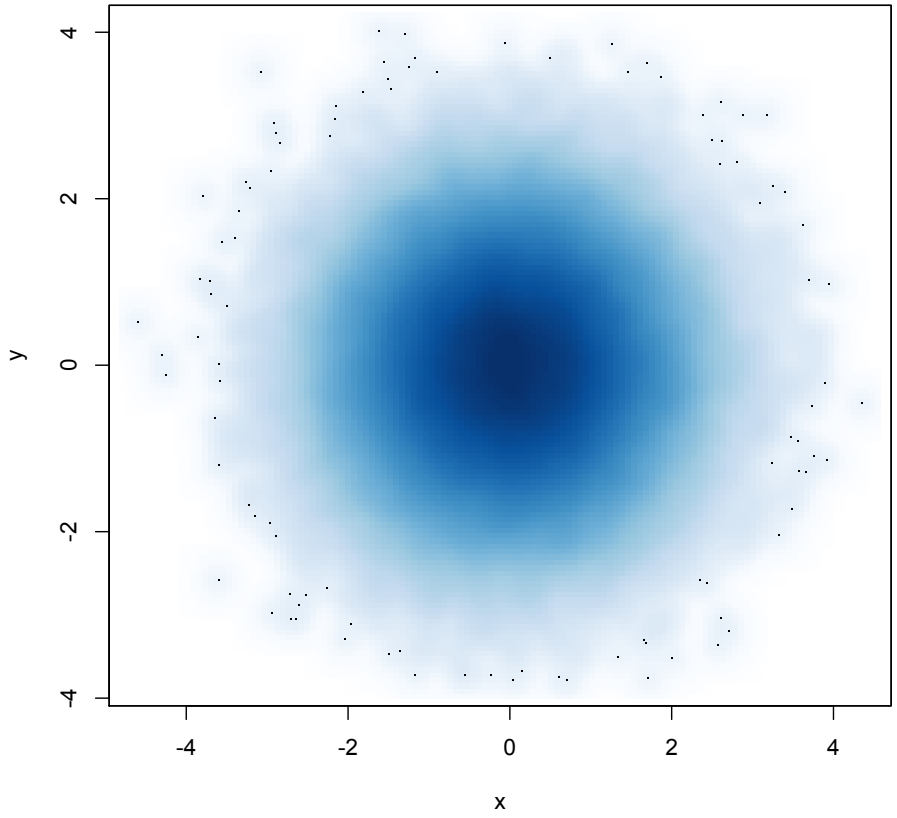
100000 x and 100000 y



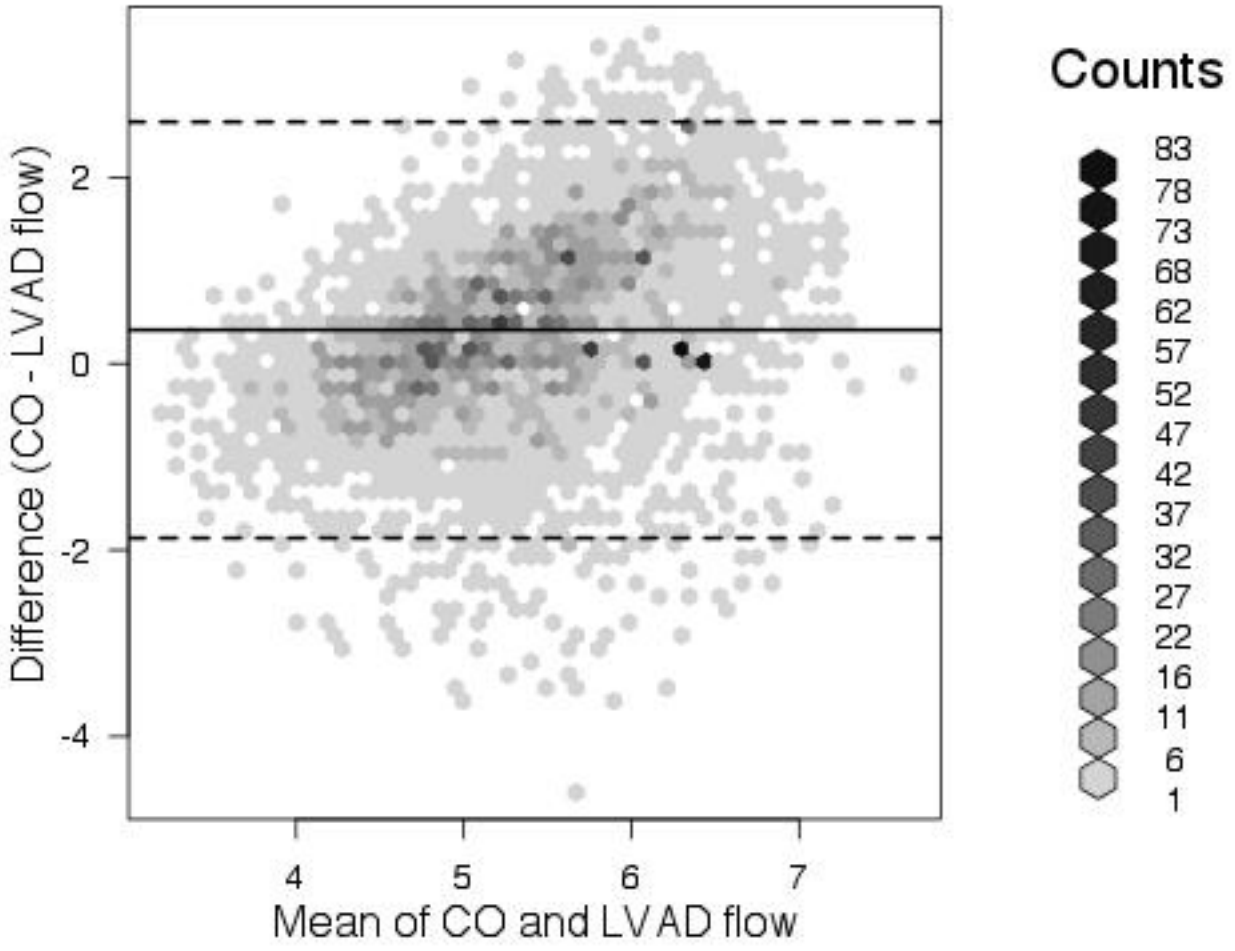
Large amounts of data - sampled



Large amounts of data – smooth scatter



Large amounts of data - hexbin



Hierarchy of human graphical perception abilities

1. Position along a common scale (most accurate task)
2. Position along identical nonaligned scales
3. Length
4. Angle and slope
5. Area
6. Volume
7. Color: hue (red, green, blue, etc.), saturation (pale/deep), and lightness– Hue can give good discrimination but poor ordering

Frank Harrell Statistical Graphics Course Handout

General suggestions

- Exclude unneeded dimensions
- Instead of showing too many actual data points, choose a random sample
- Omit chartjunk
- Keep continuous variables continuous (grouping ok for tables, but unnecessary for graphs)
- Don’t subset data finer than sample size supports

Frank Harrell Statistical Graphics Course Handout

Statistical Graphics References

- Cleveland, W. S. (1985). *The Elements of Graphing Data. Pacific Grove, Calif.: Wadsworth.*
- Cleveland W and McGill R. (1984). Graphical Perception: Theory, experimentation and applications to the development of graphical methods. JASA 79 (387): 531-554.
- Friendly, M. (2006). A brief history of data visualization. In *Handbook of Computational Statistics: Data Visualization, volume 3, ed. C. Chen, W. Hardle, and A. Unwin.*
- *Gelman A, Unwin A. (2012).* Infovis and Statistical Graphics: Different Goals, Different Looks. http://www.stat.columbia.edu/~gelman/research/published/vis14.pdf
- Tufte, E. R. (1983). *The Visual Display of Quantitative Information. Cheshire: Graphics Press.*
- Wainer, H. (1997). *Visual Revelations. New York: Springer.*
- Wickham, H., Cook, D., Hofmann, H., and Buja, A. (2010). Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics 16, 973-979.*
- Wilkinson, L. (2005). *The Grammar of Graphics. New York: Springer.*