

Data visualization

Carolyn J. Lawrence-Dill

Genetics, Development and Cell Biology

Interdepartmental Genetics and Genomics

Interdepartmental Bioinformatics and Computational Biology

Predictive Plant Phenomics Graduate Traineeship

- 1) Typography
- 2) Graphs & plots
- 3) Color
- 4) Figure legends/descriptions

Typography

sans serif font (Arial)

A b c j y

font size

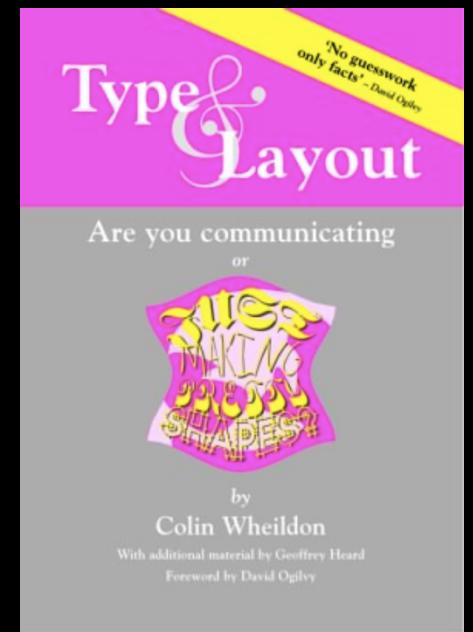
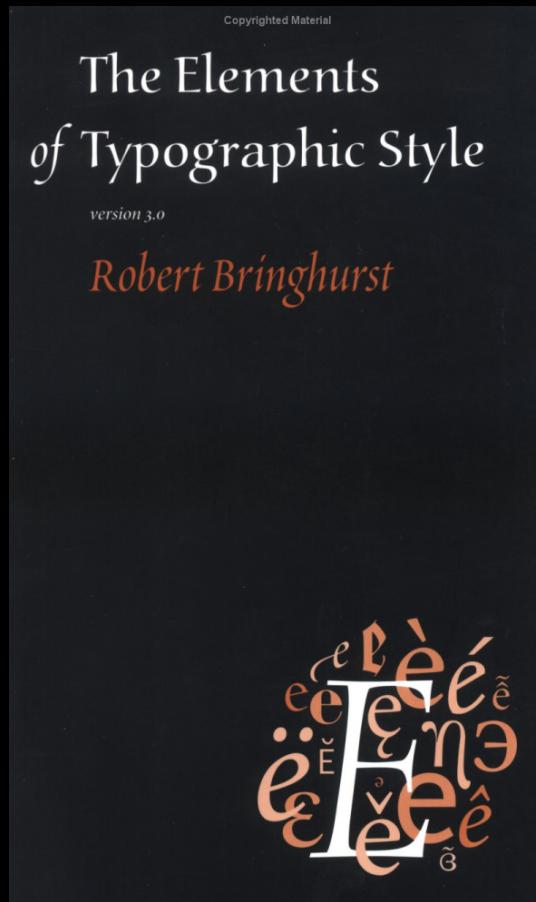
A b c j y

leading

serif font (Times New Roman)

How should you judge whether your typographic choice (e.g., type, type size, leading, number of columns) is a good one?

- a) You judge whether it looks appealing
- b) You base your decision on studies of reading comprehension



Fonts: serif or sans serif?

Minion is a serif font.

Arial is a sans serif font.

For continuous text, reading comprehension is greatly reduced by a sans serif font.

Percent of readers	Good	Fair	Poor
Serif font	67	19	14
Sans serif font	12	23	65

Justified

Is the stratosphere important for predicting changes in weather and climate? Although the role of the stratosphere has not been emphasized until recently, observations and models both indicate that the stratosphere acts to integrate high-frequency forcing from below, with long-lasting feedback effects.

Ragged right

Is the stratosphere important for predicting changes in weather and climate? Although the role of the stratosphere has not been emphasized until recently, observations and models both indicate that the stratosphere acts to integrate high-frequency forcing from below, with long-lasting feedback effects.

Should text be set justified or ragged right?

For continuous text, reading comprehension is greatly reduced by a ragged right margin.

Percent of readers	Good	Fair	Poor
Justified text	67	19	14
Ragged right text	38	22	40

Leading: the space between lines

Negative Leading

15/13: This climatology presents a picture consistent with the concepts of wavebreaking and vortex/surf-zone regions proposed by McIntyre and Palmer [1983, 1984]. In their conceptual model,

15/14: This climatology presents a picture consistent with the concepts of wavebreaking and vortex/surf-zone regions proposed by McIntyre and Palmer [1983, 1984]. In their conceptual model,

Solid

15/15: This climatology presents a picture consistent with the concepts of wavebreaking and vortex/surf-zone regions proposed by McIntyre and Palmer [1983, 1984]. In their conceptual model,

Positive Leading

15/16: This climatology presents a picture consistent with the concepts of wavebreaking and vortex/surf-zone regions proposed by McIntyre and Palmer [1983, 1984]. In their conceptual model,

15/17: This climatology presents a picture consistent with the concepts of wavebreaking and vortex/surf-zone regions proposed by McIntyre and Palmer [1983, 1984]. In their conceptual model,

15/18: This climatology presents a picture consistent with the concepts of wavebreaking and vortex/surf-zone regions proposed by McIntyre and Palmer [1983, 1984]. In their conceptual model,

What is the optimum size of type and leading?

Respondents were asked: which type size or sizes do you find easy to read as continuous text?

		Leading						
		9	10	11	12	13	14	15
Font Size	9	63	66	71				
	10		69	86	92			
	11			77	93	98		
	12				72	90	82	
	13					66	70	68

Percentages of respondents who found each combination of type size (down) and leading (across) easy to read.

11/13 was found to be best, with 10–12 point type and an extra 1–2 points of leading very good.

Line length: how long is too long?

38% of readers found that a line length over 60 characters is hard to read.

On an $8\frac{1}{2} \times 11$ page with 1" margins, how many characters per line result using 10–12 point fonts?

Characters/line	10	11	12
Single column	111	103	93
Double column	53	50	45

The scientific evidence is now overwhelming: climate change presents very serious global risks, and it demands an urgent global response.

This independent Review was commissioned by the Chancellor of the Exchequer, reporting to both the Chancellor and to the Prime Minister, as a contribution to assessing the evidence and building understanding of the economics of climate change.

The Review first examines the evidence on the economic impacts of climate change itself, and explores the economics of stabilising greenhouse gases in the atmosphere. The second half of the Review considers the complex policy challenges involved in managing the transition to a low-carbon economy and in ensuring that societies can adapt to the consequences of climate change that can no longer be avoided.

Sans Serif Font

The Review takes an international perspective. Climate change is global in its causes and consequences, and international collective action will be critical in driving an effective, efficient and equitable response on the scale required. This response will require deeper international co-operation in many areas - most notably in creating price signals and markets for carbon, spurring technology research, development and deployment, and promoting adaptation, particularly for developing countries.

Climate change presents a unique challenge for economics: it is the greatest and widest-ranging market failure ever seen. The economic analysis must therefore be global, deal with long time horizons, have the economics of risk and uncertainty at centre stage, and examine the possibility of major, non-marginal change. To meet these requirements, the Review draws on ideas and techniques from most of the important areas of economics, including many recent advances.

The benefits of strong, early action on climate change outweigh the costs

The effects of our actions now on future changes in the climate have long lead times. What we do now can have only a limited effect on the climate over the next 40 or 50 years. On the other hand what we do in the next 10 or 20 years can have a profound effect on the climate in the second half of this century and in the next.

No-one can predict the consequences of climate change with complete certainty; but we now know enough to understand the risks. Mitigation - taking strong action to reduce emissions - must be viewed as an investment, a cost incurred now and in the coming few decades to avoid the risks of very severe consequences in the future. If these investments are made wisely, the costs will be manageable, and there will be a wide range of opportunities for growth and development along the way. For this to work well, policy must promote sound market signals, overcome market failures and have equity and risk mitigation at its core. That essentially is the conceptual framework of this Review.

80

- 1) Typography
- 2) Graphs & plots
- 3) Color
- 4) Figure legends/descriptions

Displaying data well

- Be accurate and clear.
- Let the data speak.
 - Show as much information as possible, taking care not to obscure the message.
- Science not sales.
 - Avoid unnecessary frills — esp. gratuitous 3-D.

Graphs & Plots

Minimize non-data ink

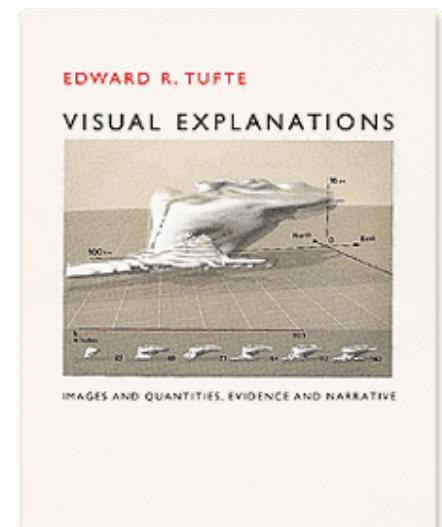
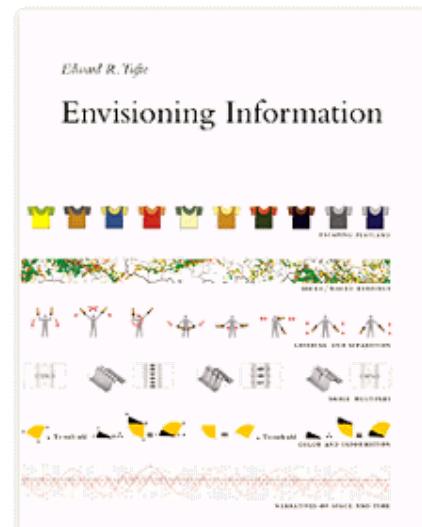
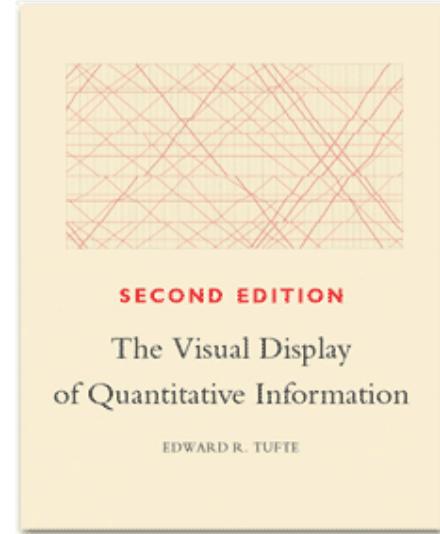
No “chartjunk” (e.g., gratuitous 3-D)

Avoid lying with data

Bank to 45°



Edward R. Tufte



Slide credit: Mark Baldwin, University of Reading

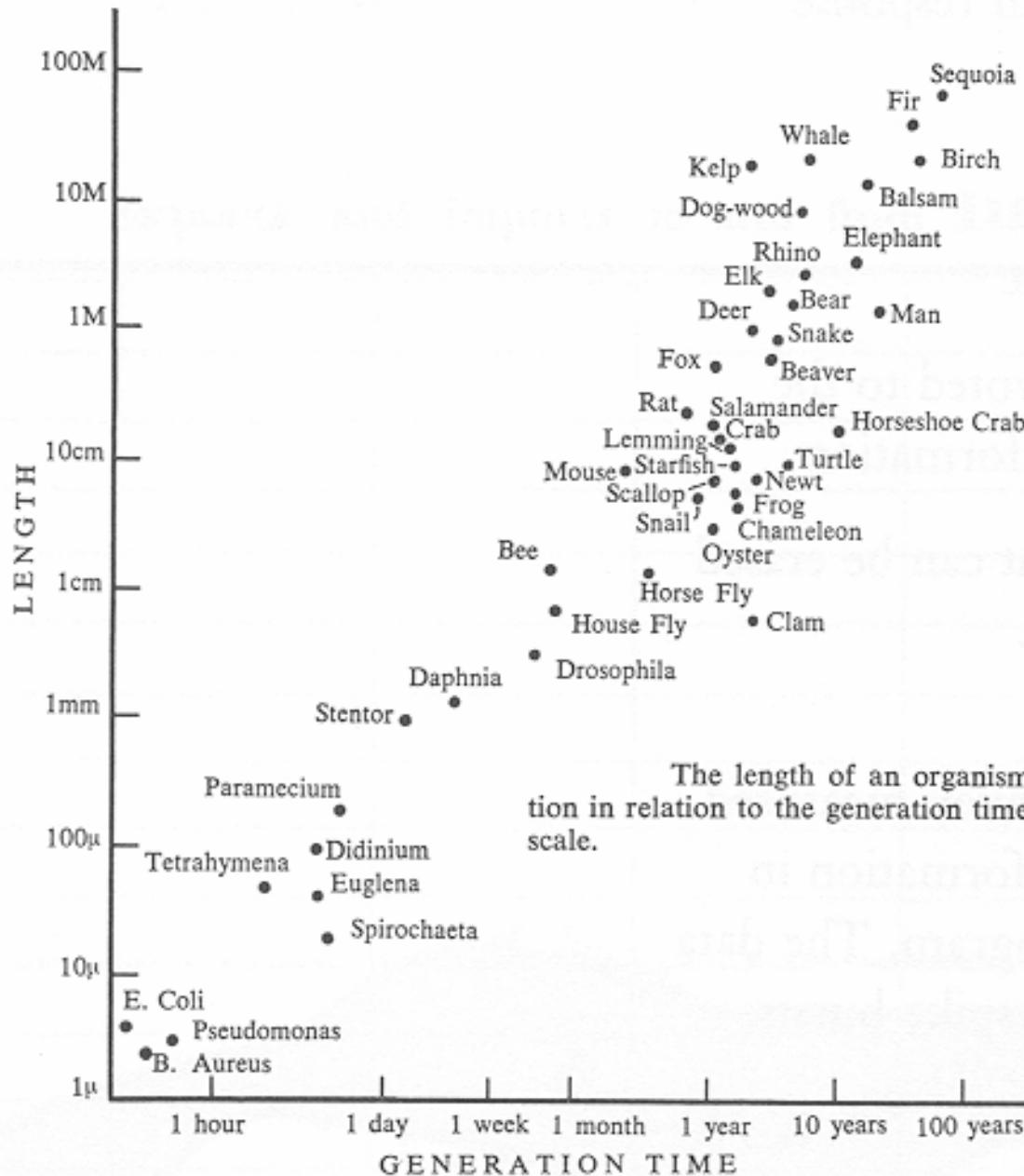
Maximize Data-ink; Minimize non-Data Ink

Edward Tufte defines the data ink ratio as:

Data Ink Ratio = (data ink)/(total ink in the plot)

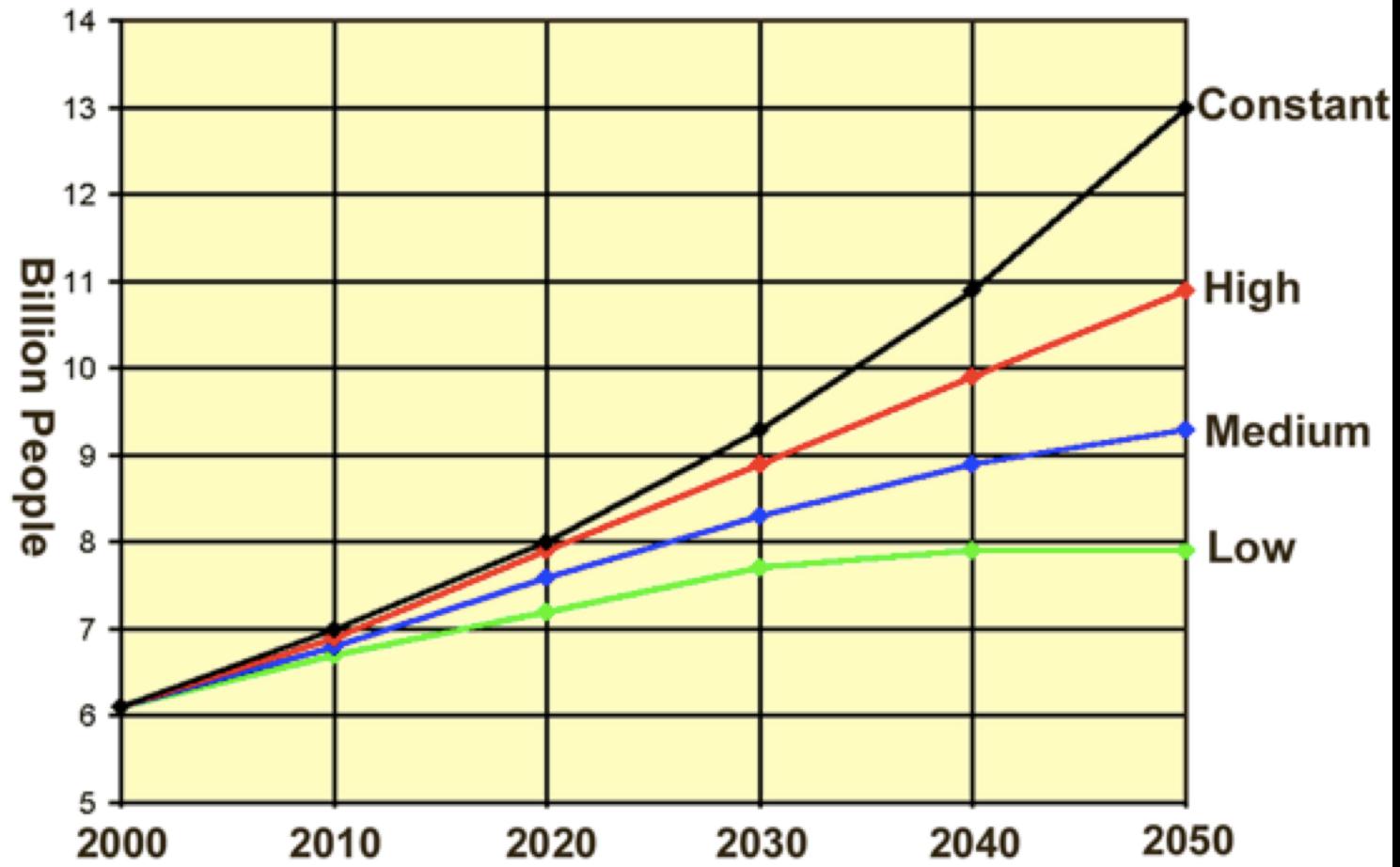
To do this you:

- Avoid heavy grids
- Replace enclosing box with an x/y grid
- Prune graphics by: replacing bars with single lines, erasing non-data ink; eliminating lines from axes; starting x/y axes at the data values
- Avoid over busy grids, excess ticks, redundant representation of simple data, boxes, shadows, pointers, legends.



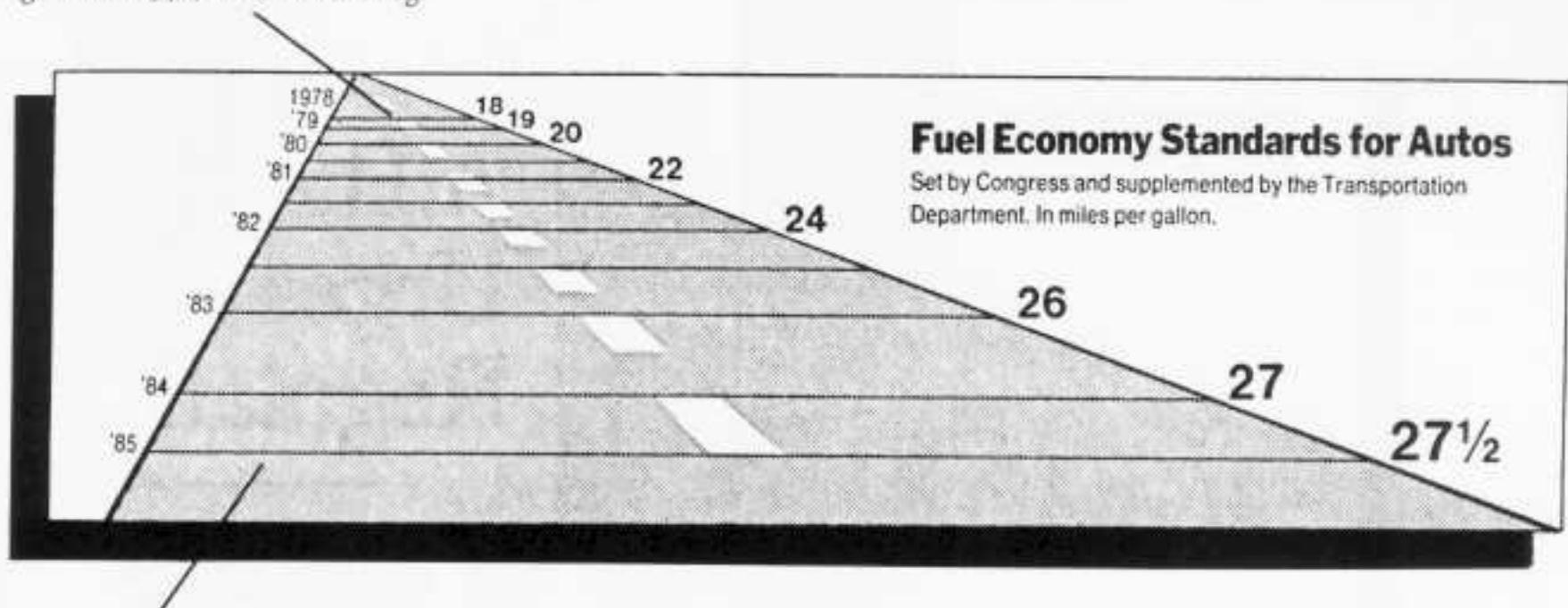
The length of an organism at the time of reproduction in relation to the generation time, plotted on a logarithmic scale.

World Population Projections



Data from United Nations, World Population Prospects 2000 Revision
Graphic copyright Facing the Future, 2001

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



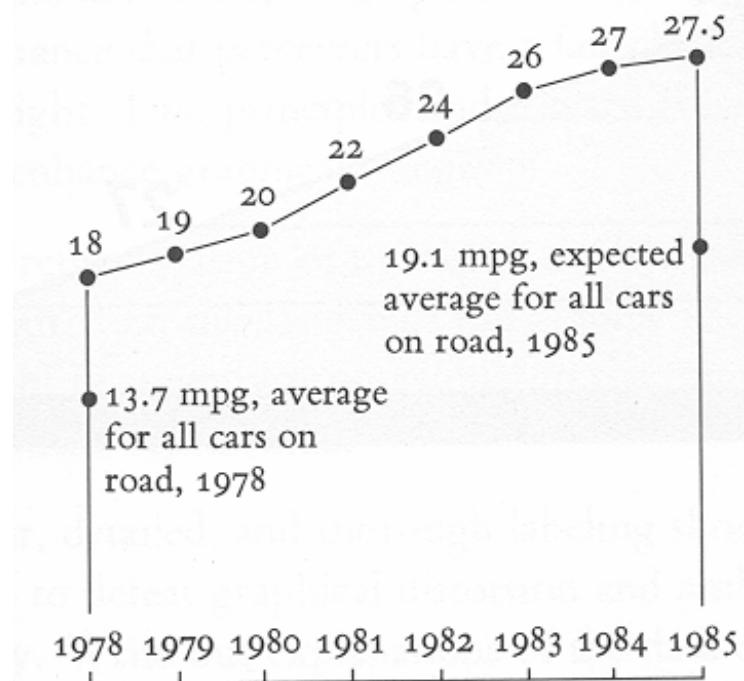
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Fuel Economy Standards for Autos

Set by Congress and supplemented by the Transportation Department. In miles per gallon.

New York Times, August 9, 1978, p. D-2.

REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985



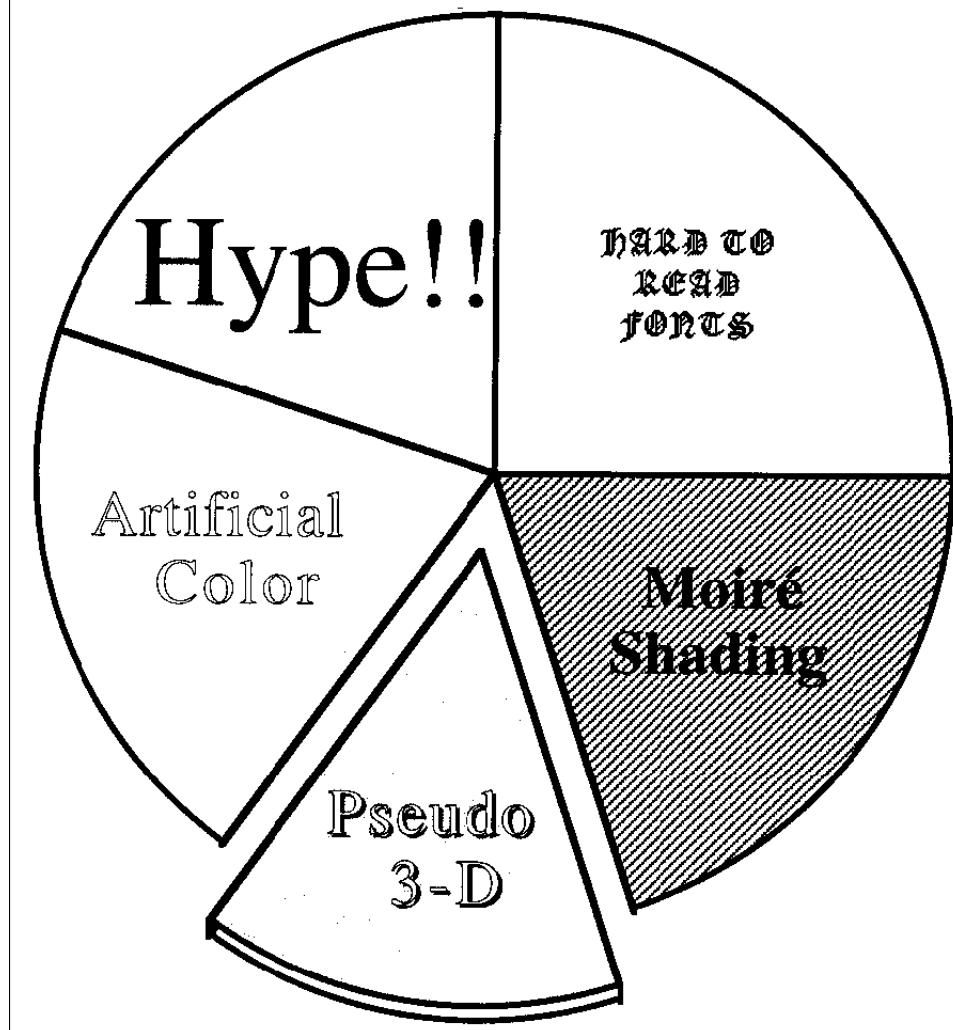
Chartjunk

Chartjunk consists of decorative elements that provide no data and cause confusion (e.g., fake 3-D).

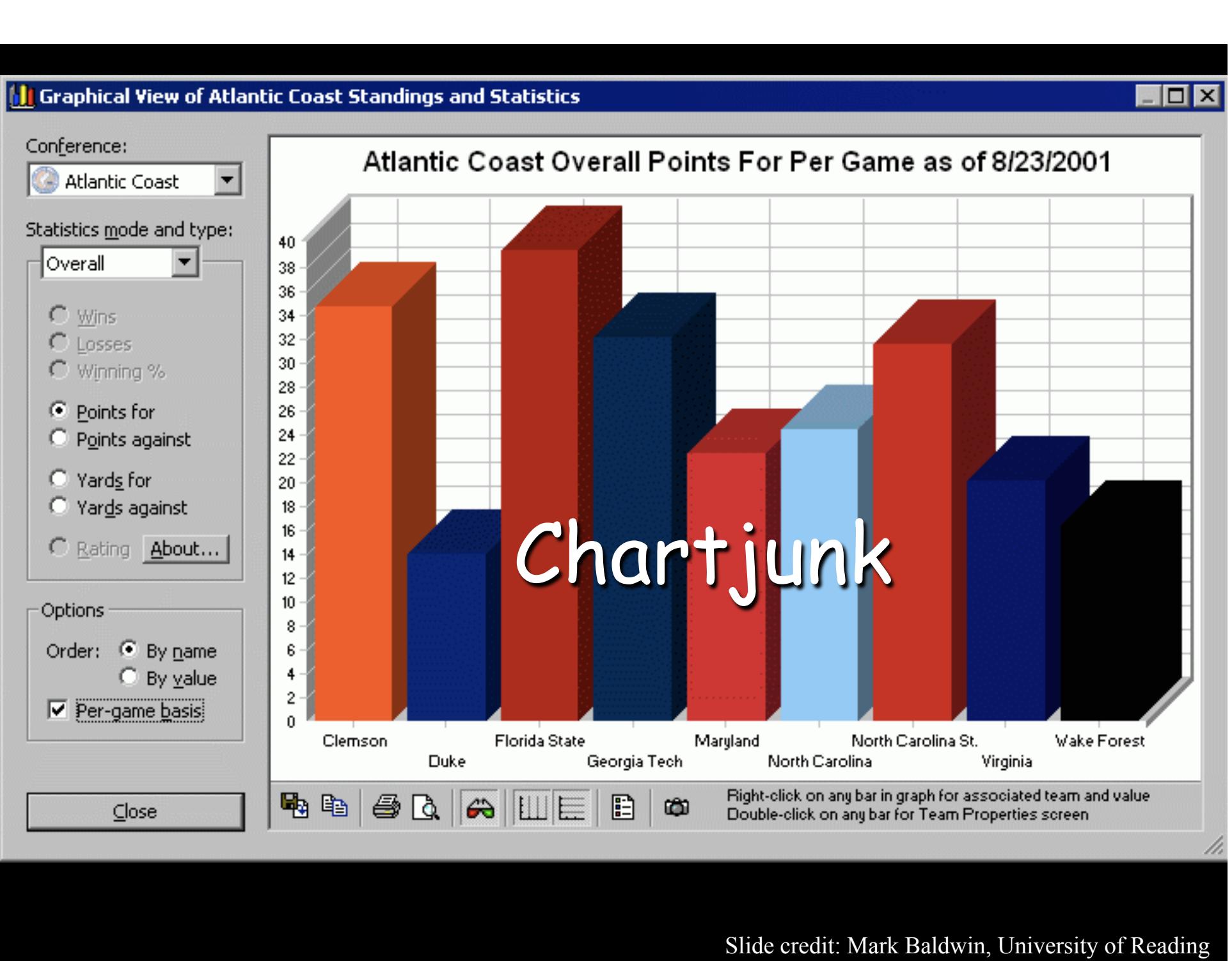
Elements in close proximity cause a visible interaction. Such interactions can be very fatiguing (e.g., moiré patterns, optical vibration) and can show information that is not really there.

In major science publications we see 2% to 20% moiré vibration. For example, in recent statistical and computer publications chartjunk ranges from 12% to 68%.

Some Components of Chartjunk

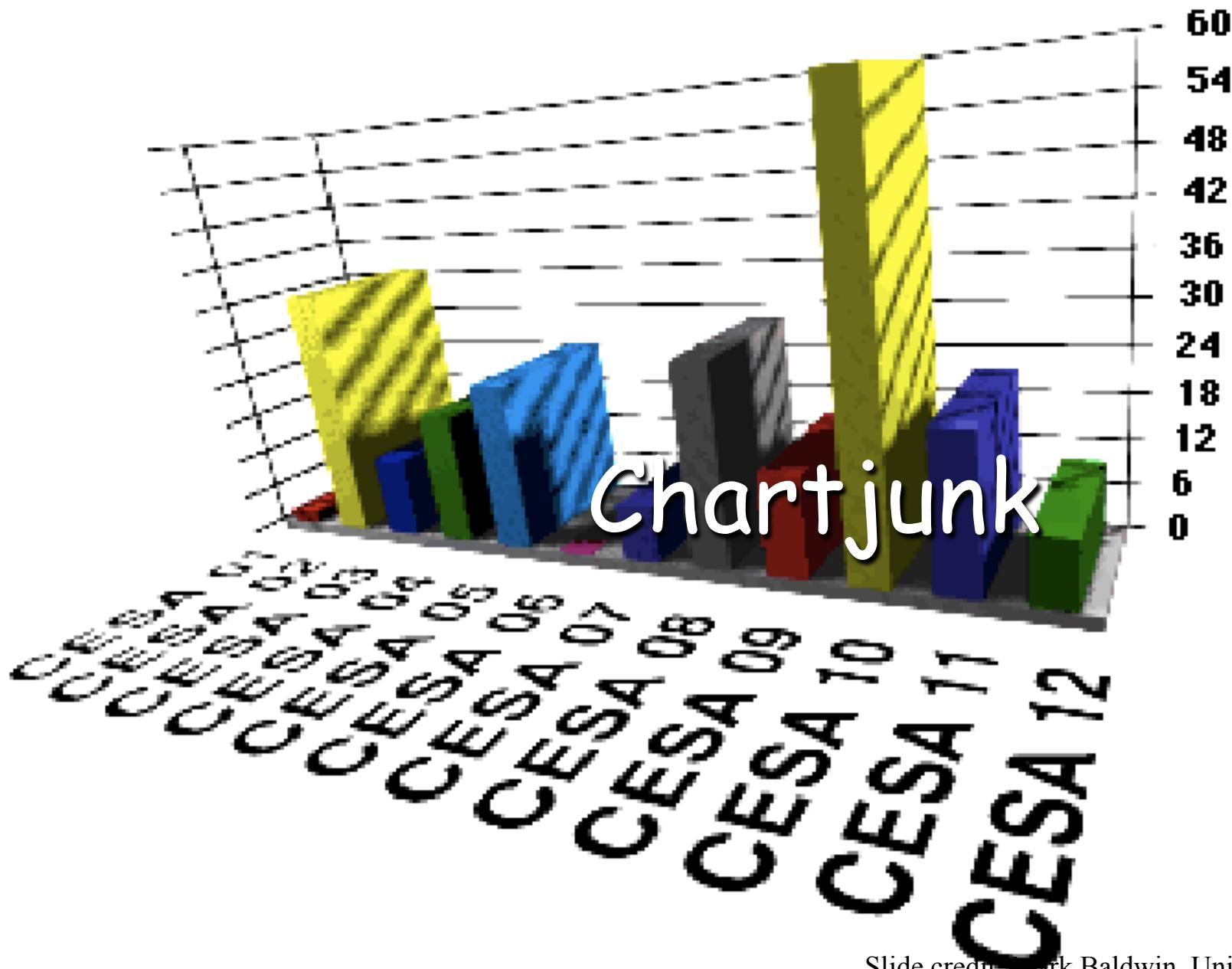


Slide credit: Mark Baldwin, University of Reading

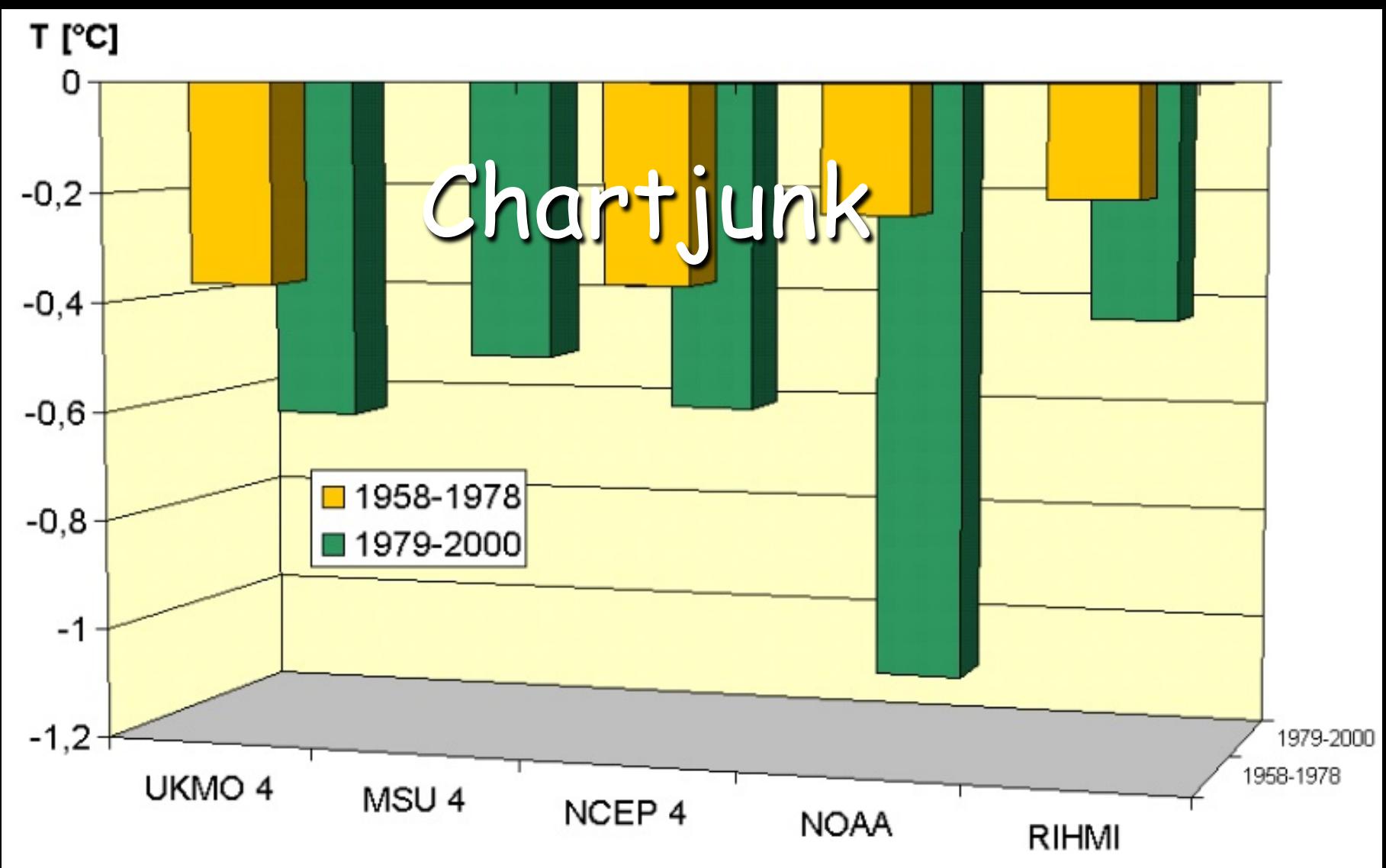


Slide credit: Mark Baldwin, University of Reading

No Computer Access

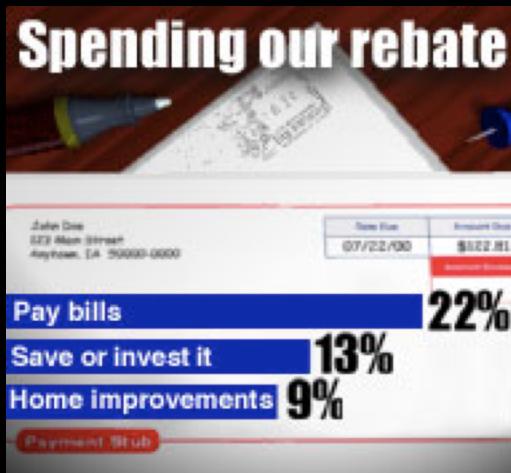
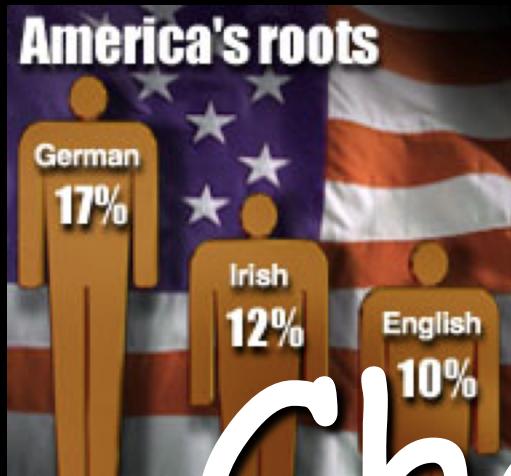


Slide credit: Mark Baldwin, University of Reading



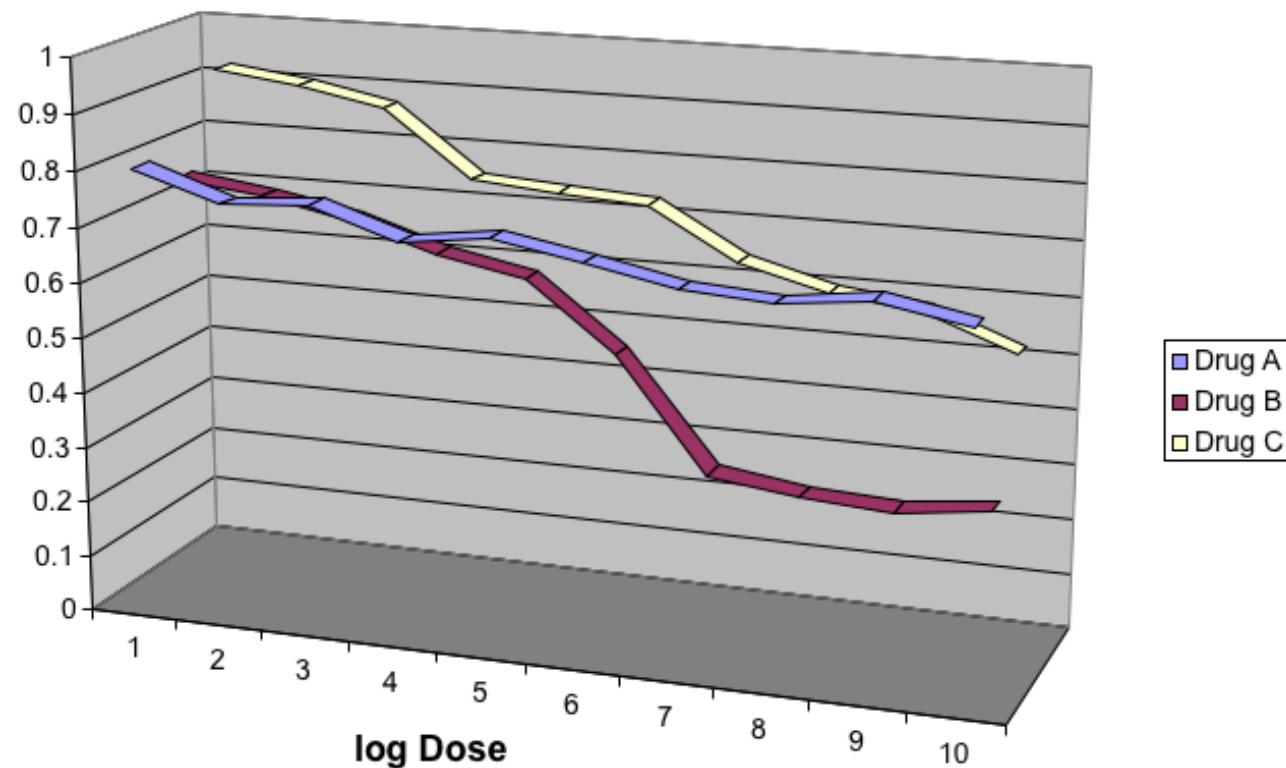
Slide credit: Mark Baldwin, University of Reading

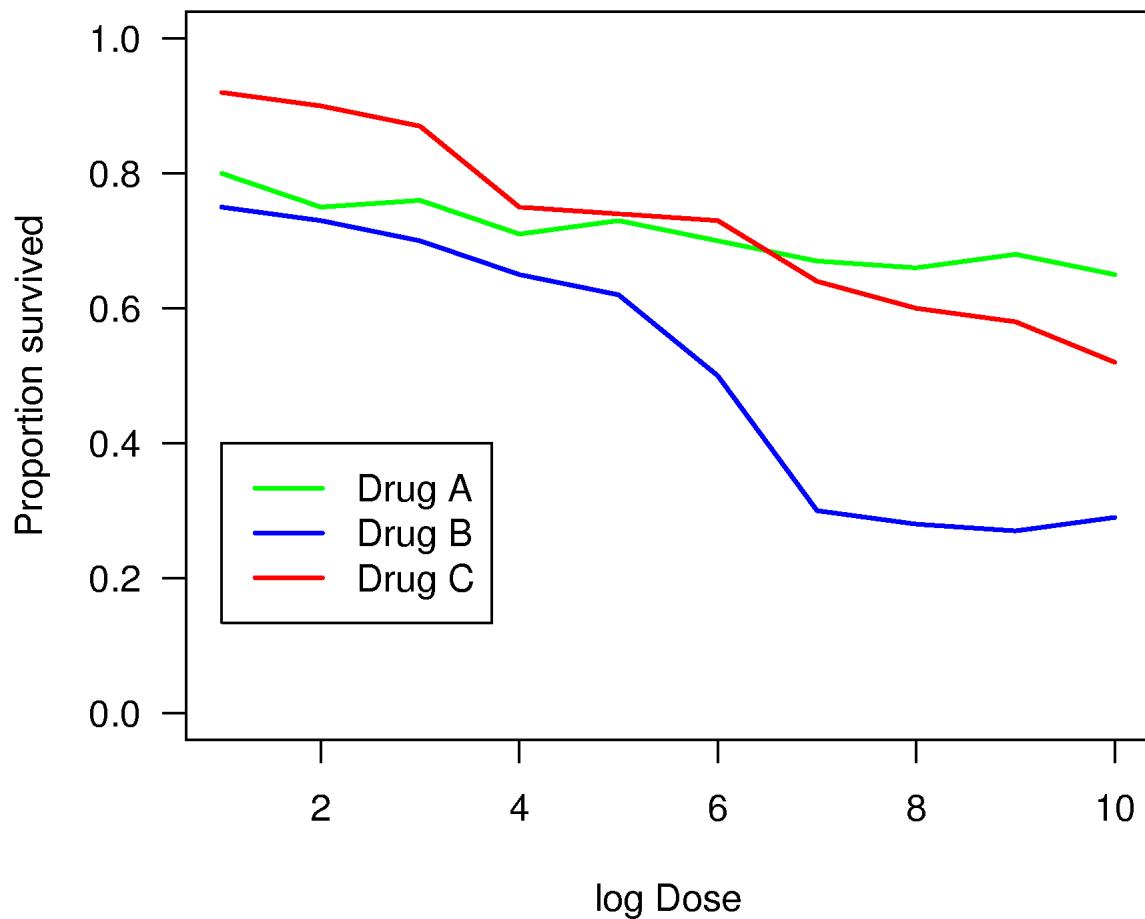
Chartjunk

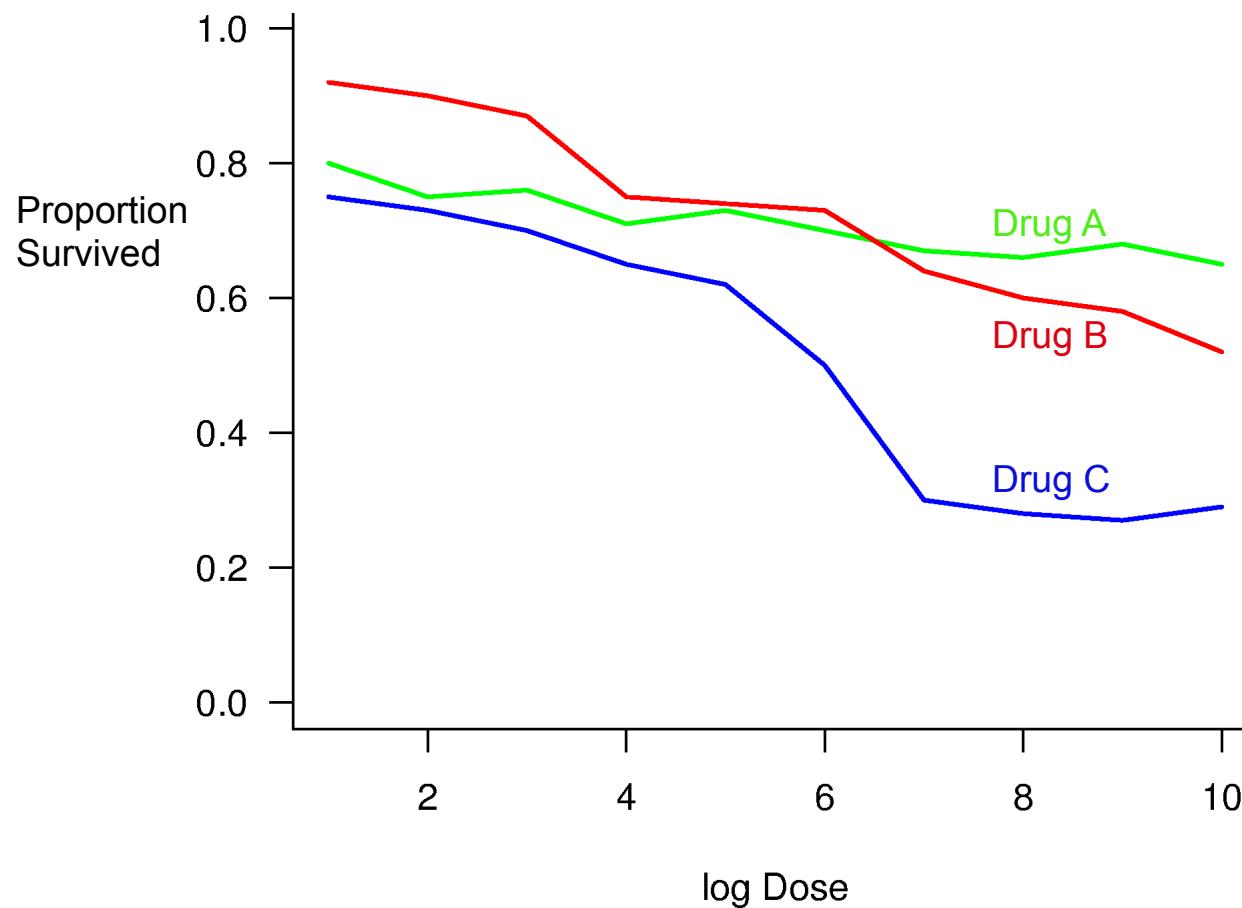


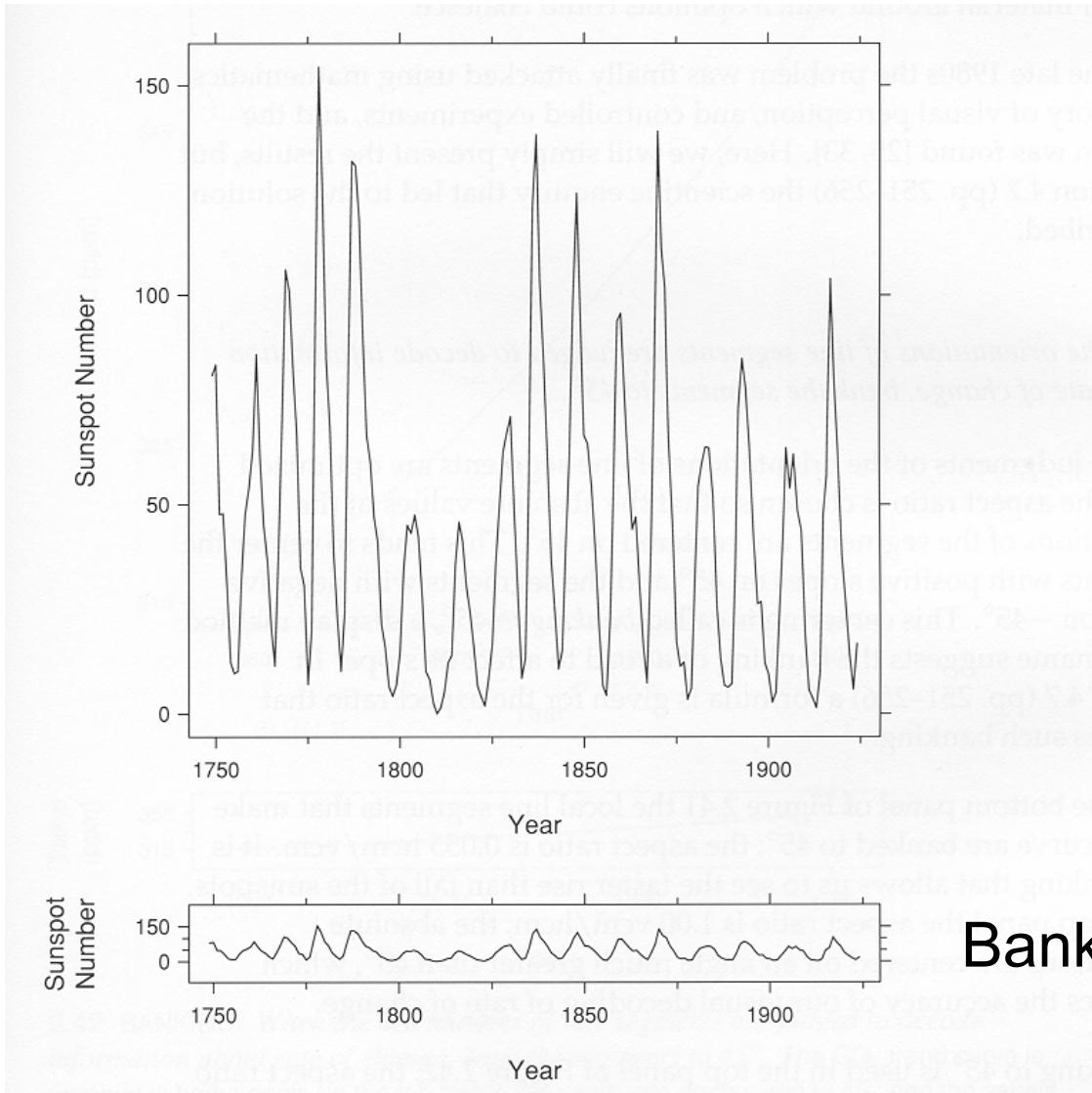
Slide credit: Mark Baldwin, University of Reading

Proportion survived









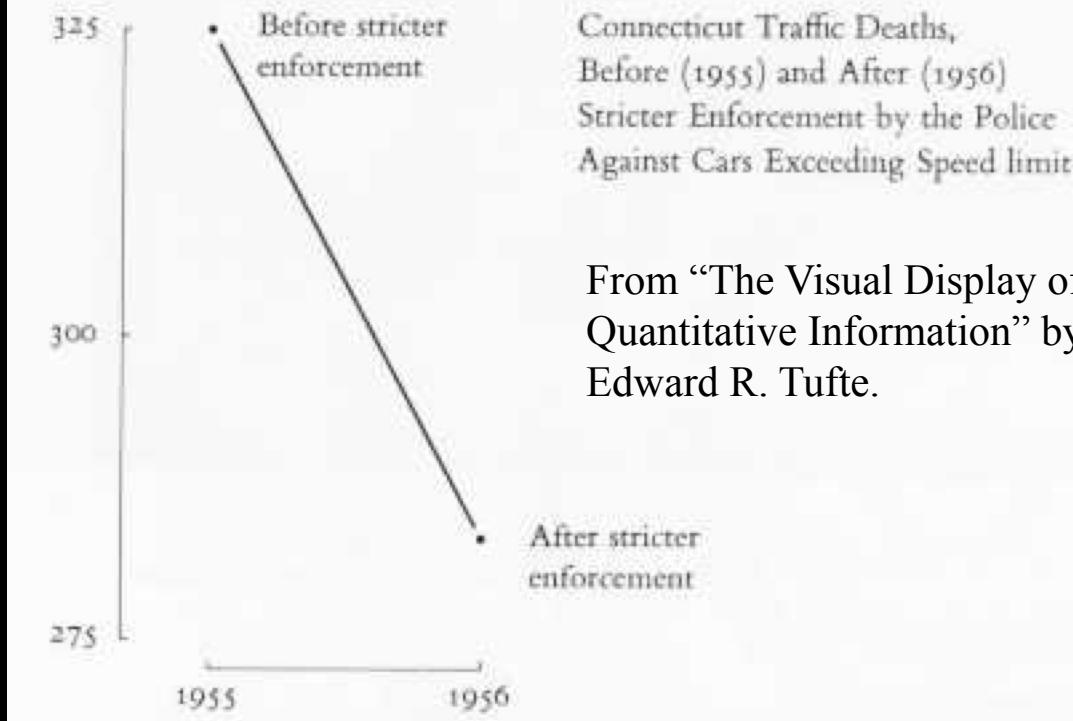
Bank to 45°



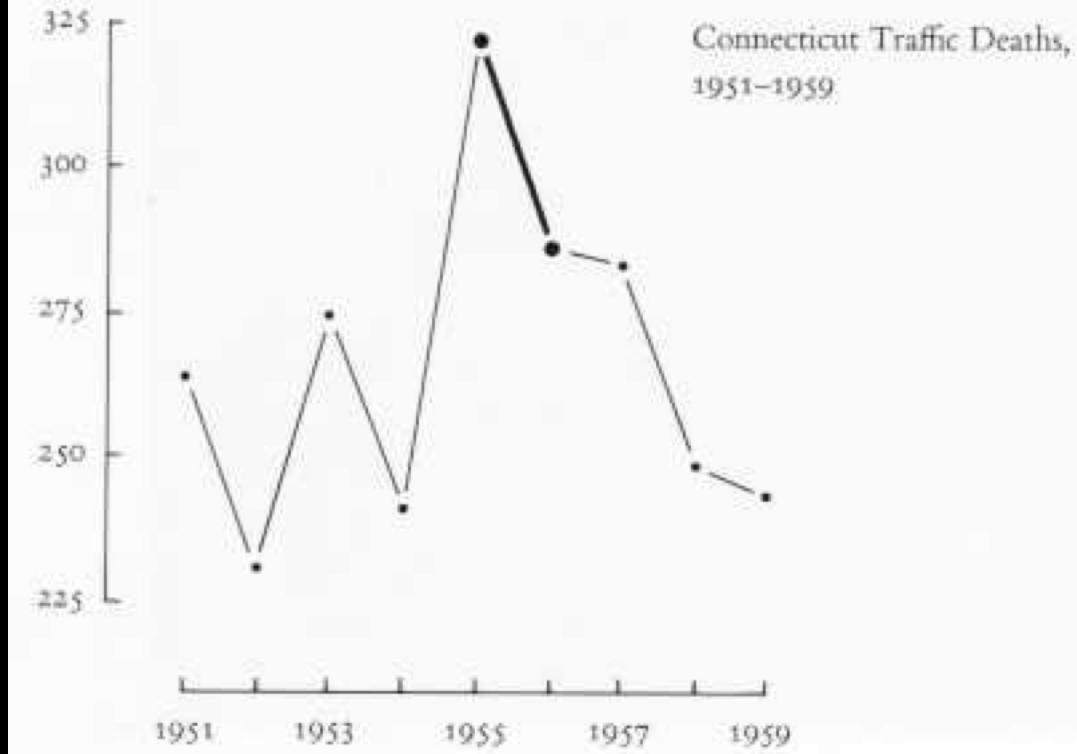
<https://eagereyes.org/basics/banking-45-degrees>

Graphics must not quote data out of context.

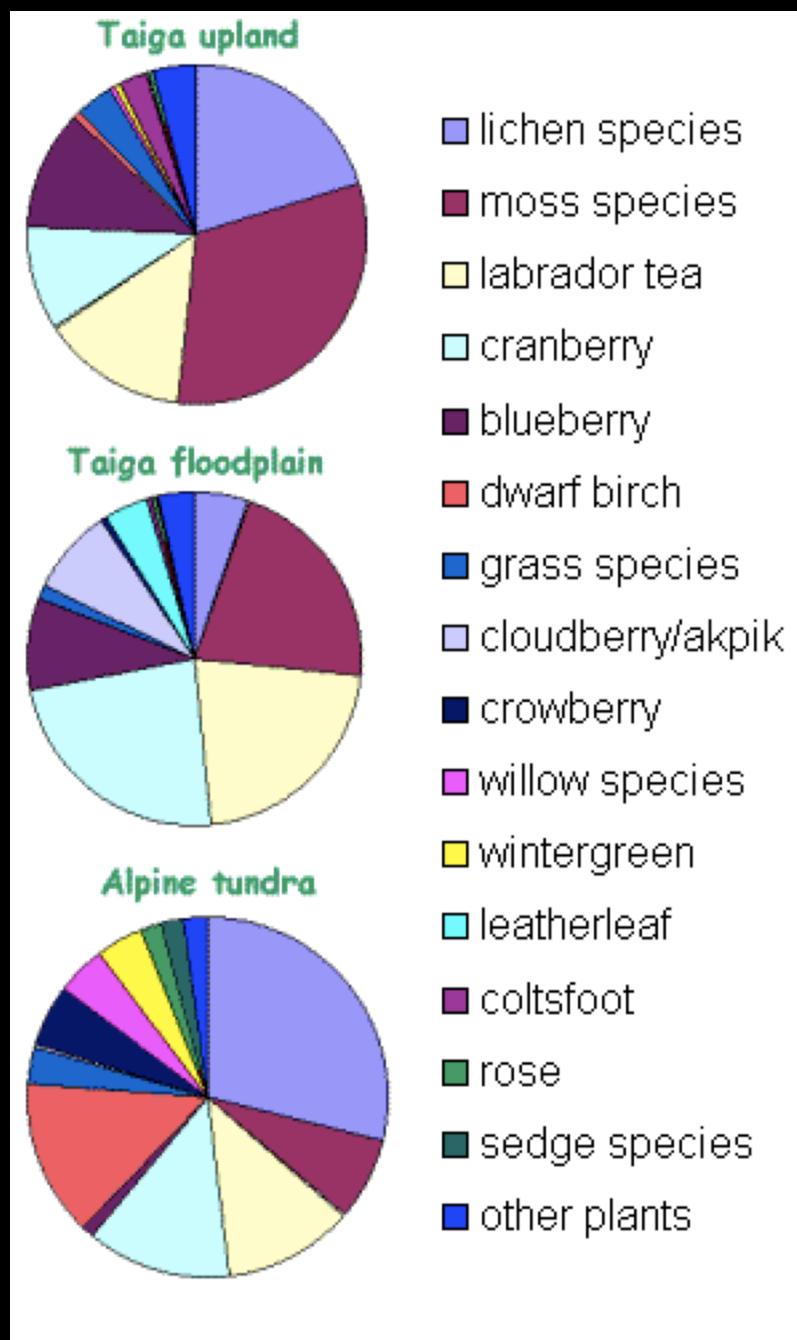
Nearly all the important questions are left unanswered by this display:



A few more data points add immensely to the account:

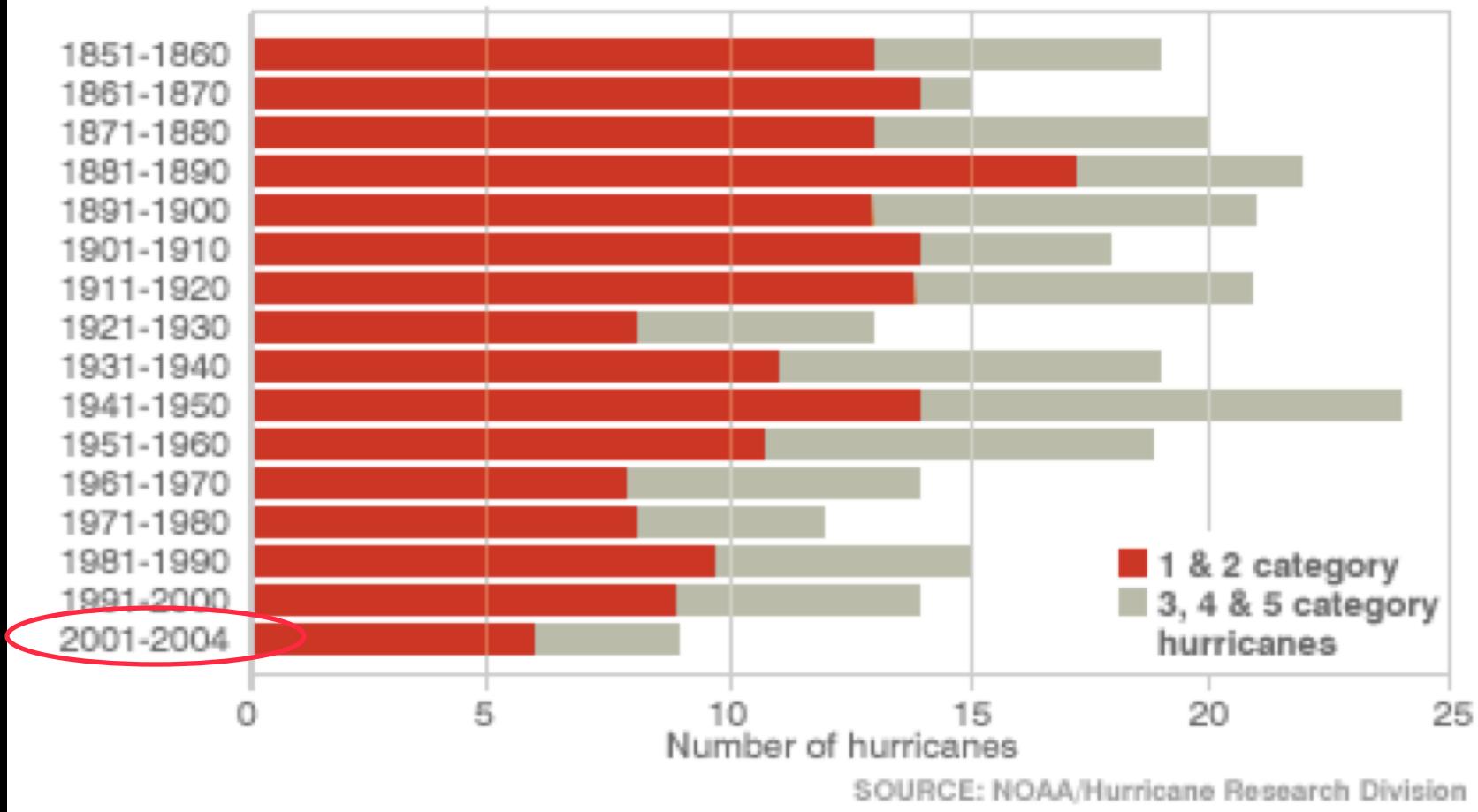


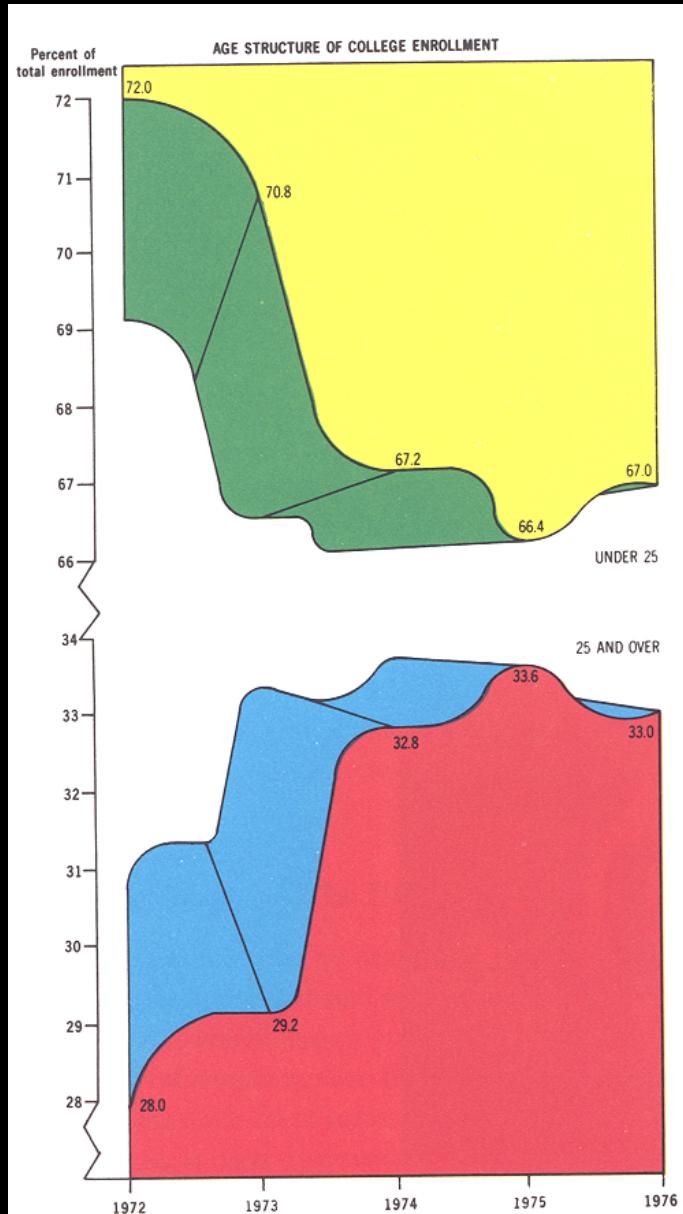
Avoid visual puzzles.



Slide credit: Mark Baldwin, University of Reading

HURRICANES STRIKING US MAINLAND EACH DECADE





AGE STRUCTURE OF COLLEGE ENROLLMENT

Percent of Total Enrollment 25 and Over

1972	● 28.0
1973	● 29.2
1974	● 32.6
1975	● 33.6
1976	● 33.0

- 1) Typography
- 2) Graphs & plots
- 3) Color
- 4) Figure legends/descriptions

What colors do you see?



full color



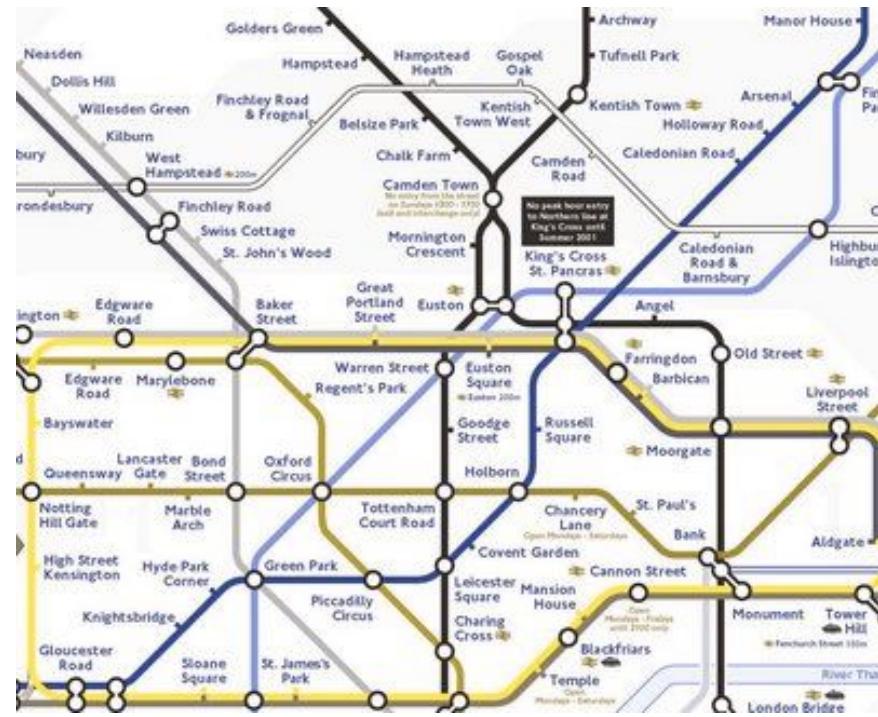
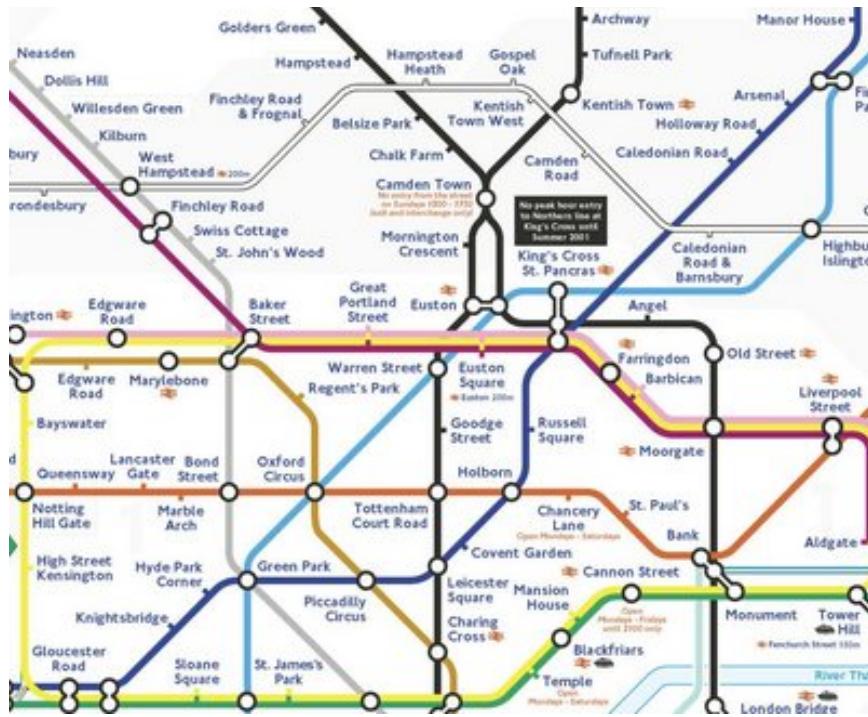
8% males, 0.05% females
deutanopic (red-green)



1% males protanopic

<http://www.vischeck.com>

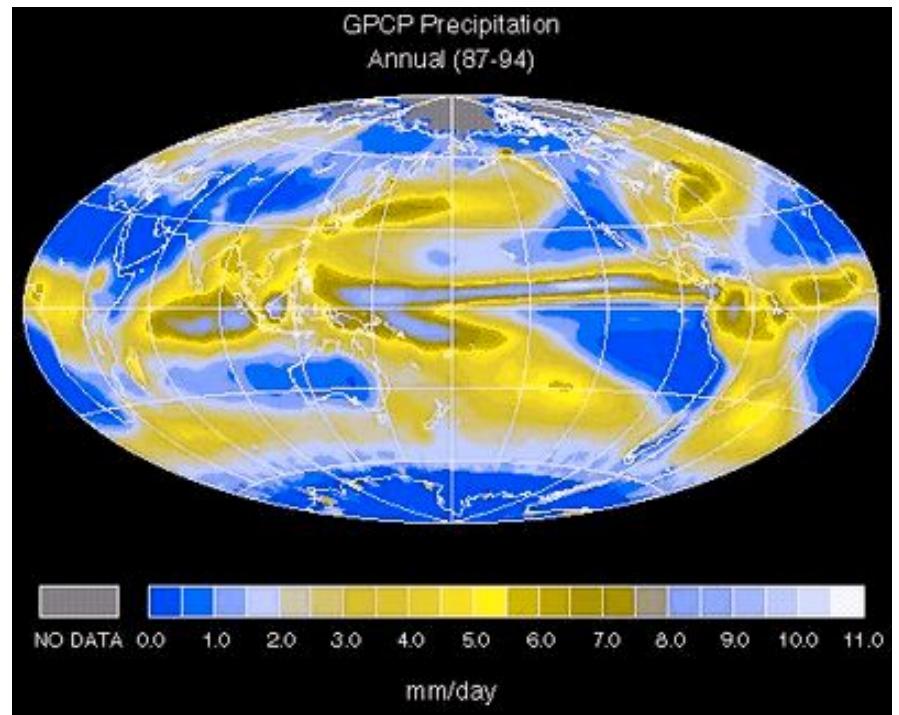
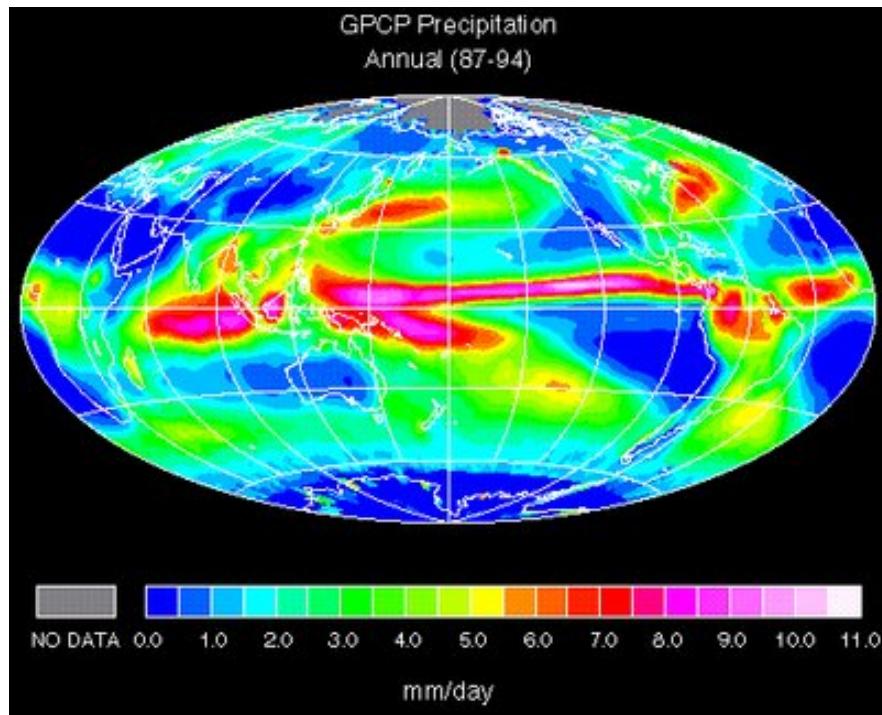
Can you see a difference?



6% of the male population can't

Slide credit: Mark Baldwin, University of Reading

Can you see a difference?



6% of the male population can't

- 1) Typography
- 2) Graphs & plots
- 3) Color
- 4) Figure legends/descriptions

Figure Legends: Write what it is, not what it means.
What it means belongs in the text of the article.

Figure X. Title statement.

Describe axes, color scheme, and symbols, if any.

Figure
number

Figure title: no more than 15 words, set in bold type,
using sentence case and ending with a period.

Figure 3. Vital rates observed during the experiment.

Vital rates from the first (a–c) and second (d–f) growing seasons. Letters denote statistically significant Tukey's HSD contrasts. In particular, snowmelt vs. control (x), warming vs. control (y), and warming vs. snowmelt (z). The y-axis in each graph reports the vital rate's unit of measure in brackets.

Figure legend (caption): no more than 300 words, describing the key message of the figure in such a way that readers can interpret the figure without referring to the text of the article.

Figure Legends: A good example!

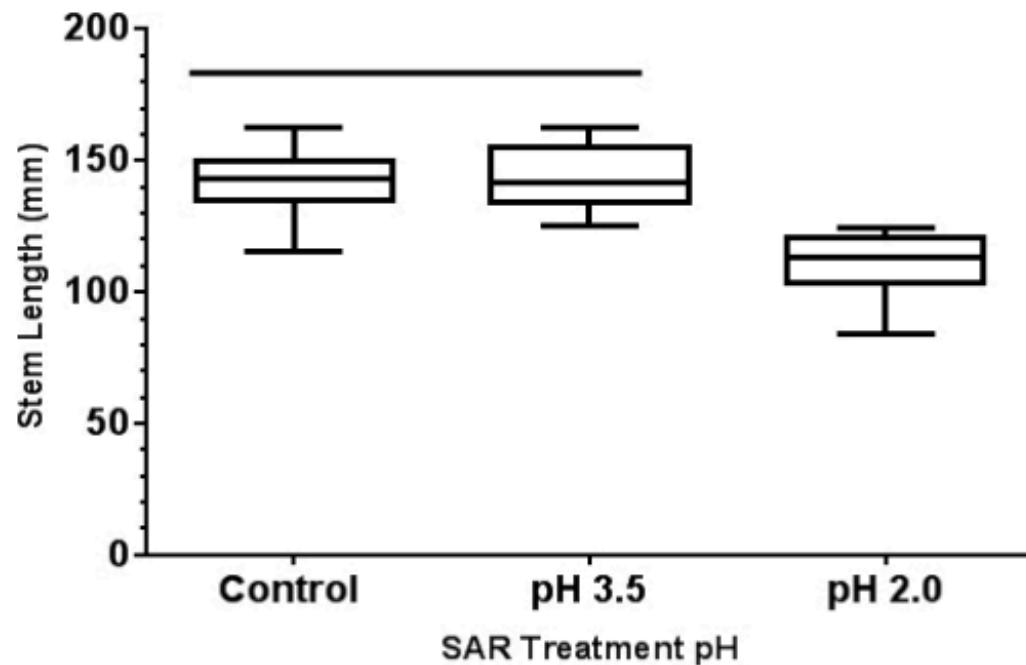


Figure 1. Median stem length of 30 d.o. sunflower seedlings following 20 daily treatments with simulated acid rain (SAR) solution in a greenhouse setting. Box = 25th and 75th percentiles; bars = min and max values. SAR solution prepared from tapwater with 2M sulfuric acid: 1M nitric acid. Number over bar indicates sample size. Line over bars indicates no significant differences. Kruskal-Wallis Test ($p < 0.0001$) and Dunn's post hoc test.

Conclusions

- 1) Pay attention to typography
- 2) Work to improve graphs and plots
- 3) Avoid “chartjunk”
- 4) Maximize data ink
- 5) Avoid visual puzzles
- 6) Use easy-to-decode color tables
- 7) Try to make the message in your graphic immediately apparent to others
- 8) In figure legends describe all aspects of the figure itself and leave interpretation to the body text