Logistics

- Ask for help! Hoss and I both have office hours that you can come to for help.
- Final exam time: 12:30pm May 7th.
- You should have completed your initial exploration and decided on the questions you want to investigate.
- This week you should be defining the hypotheses you will use to answer your questions.
- You should also be deciding on an experimental design.
- Think about potential extraneous and nuisance variables.
- Look into the Graduate Student Resource Center

https://unmgrc.unm.edu/support-services/individual-consultations.php

Charts and Plots

MATTHEW FRICKE

1.0 – SEND CORRECTIONS TO MFRICKE@UNM.EDU

Plots should…

- Show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data

Plots should…

- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set

Graphics should reveal data

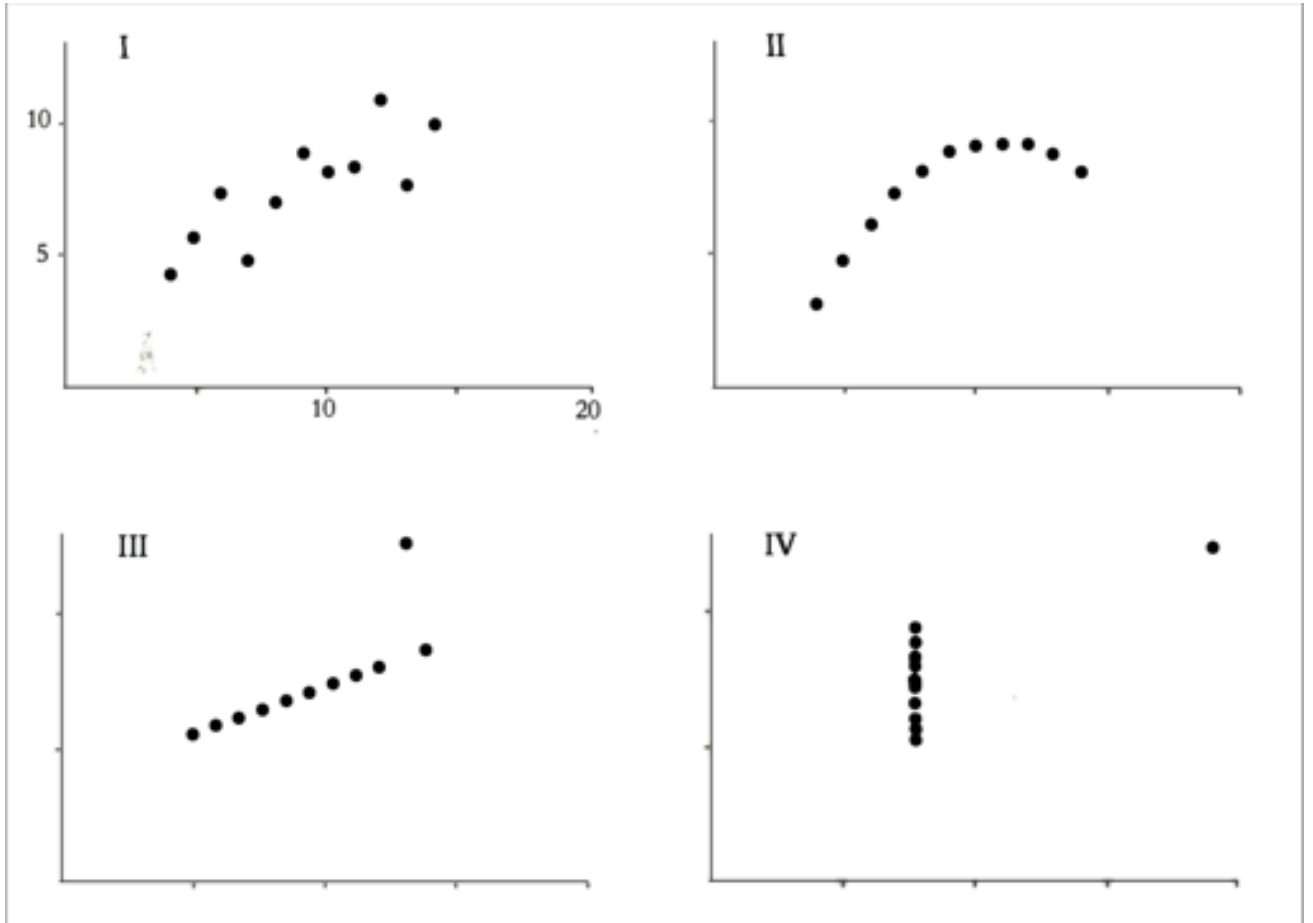
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

→

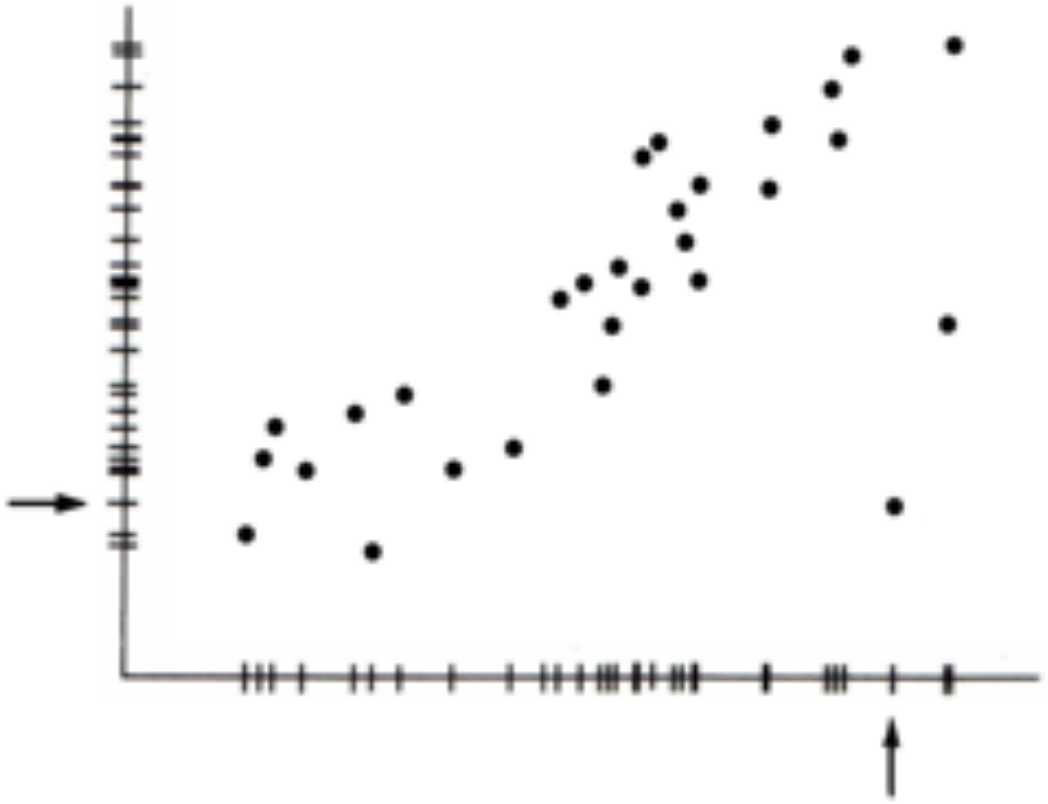
N = 11
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope = 0.118
t = 4.24
sum of squares $X - \overline{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
$r^2 = .67$

Graphics should reveal data

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

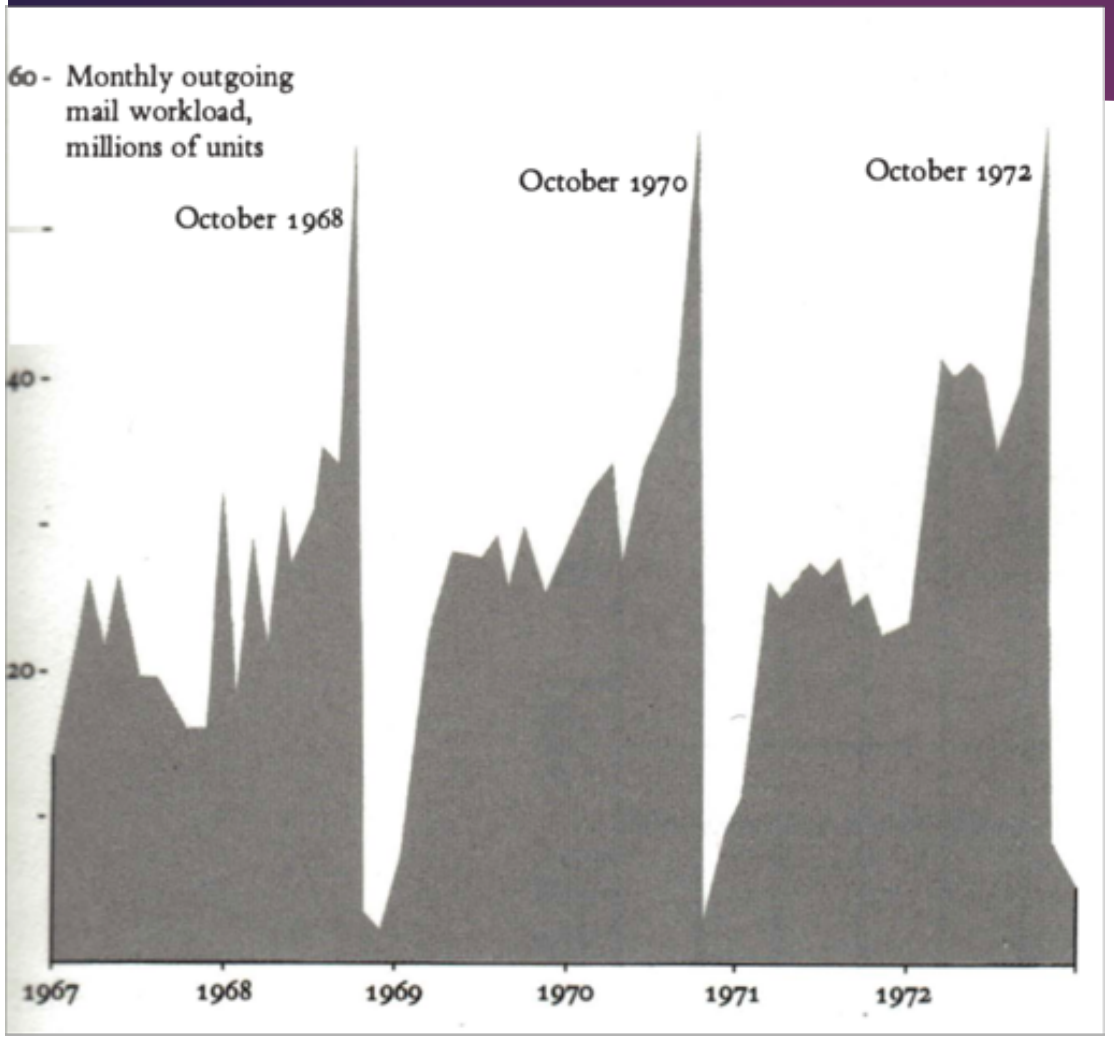


Graphics should reveal data



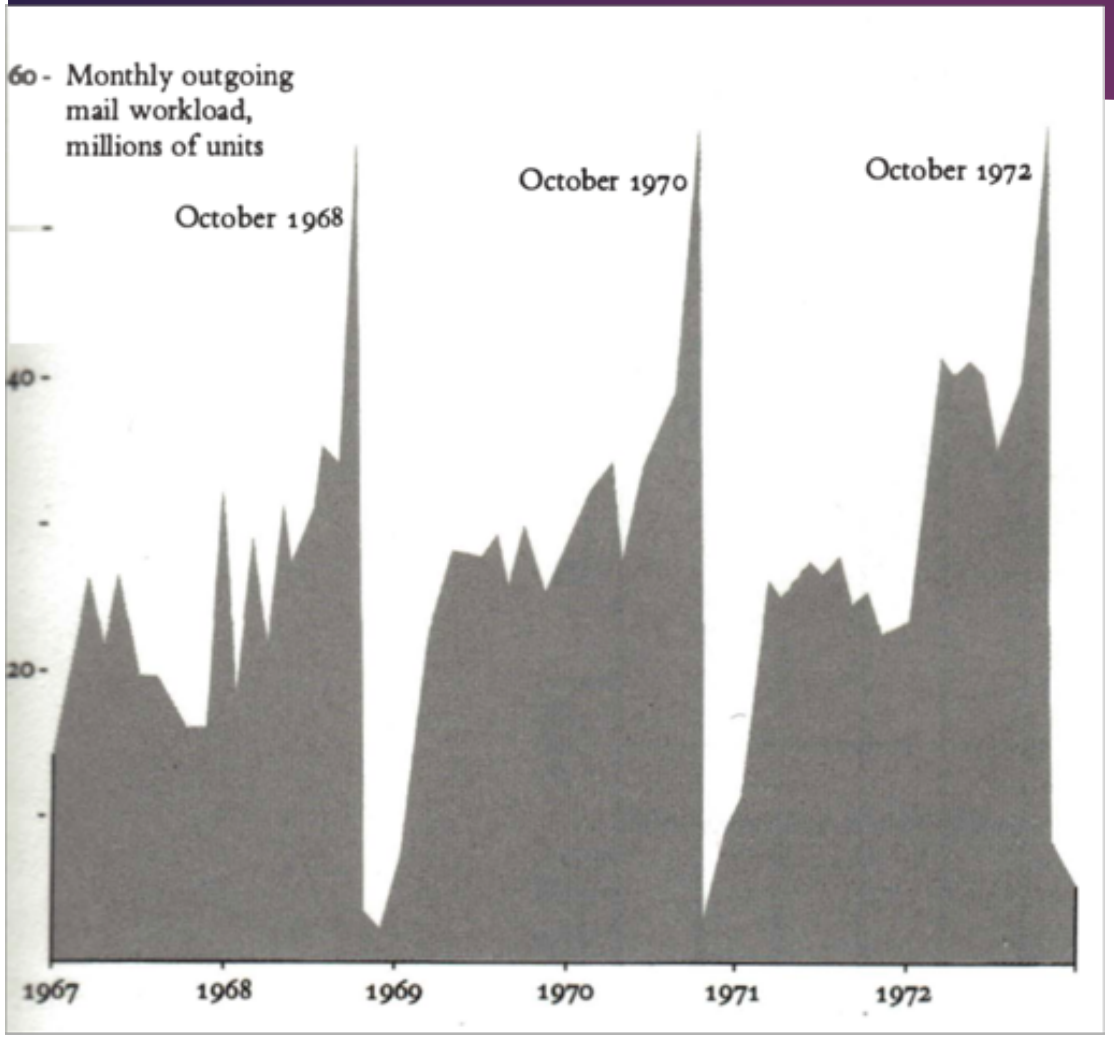
Stephen S. Brier and Stephen E. Fienberg, "Recent Econometric Modelling of Crime and Punishment: Support for the Deterrence Hypothesis?" in Stephen E. Fienberg and Albert J. Reiss, Jr., eds., *Indicators of Crime and Criminal Justice: Quantitative Studies* (Washington, D.C., 1980), p. 89.

Time series



Outgoing congressional mail (tax payer expense)

Time series

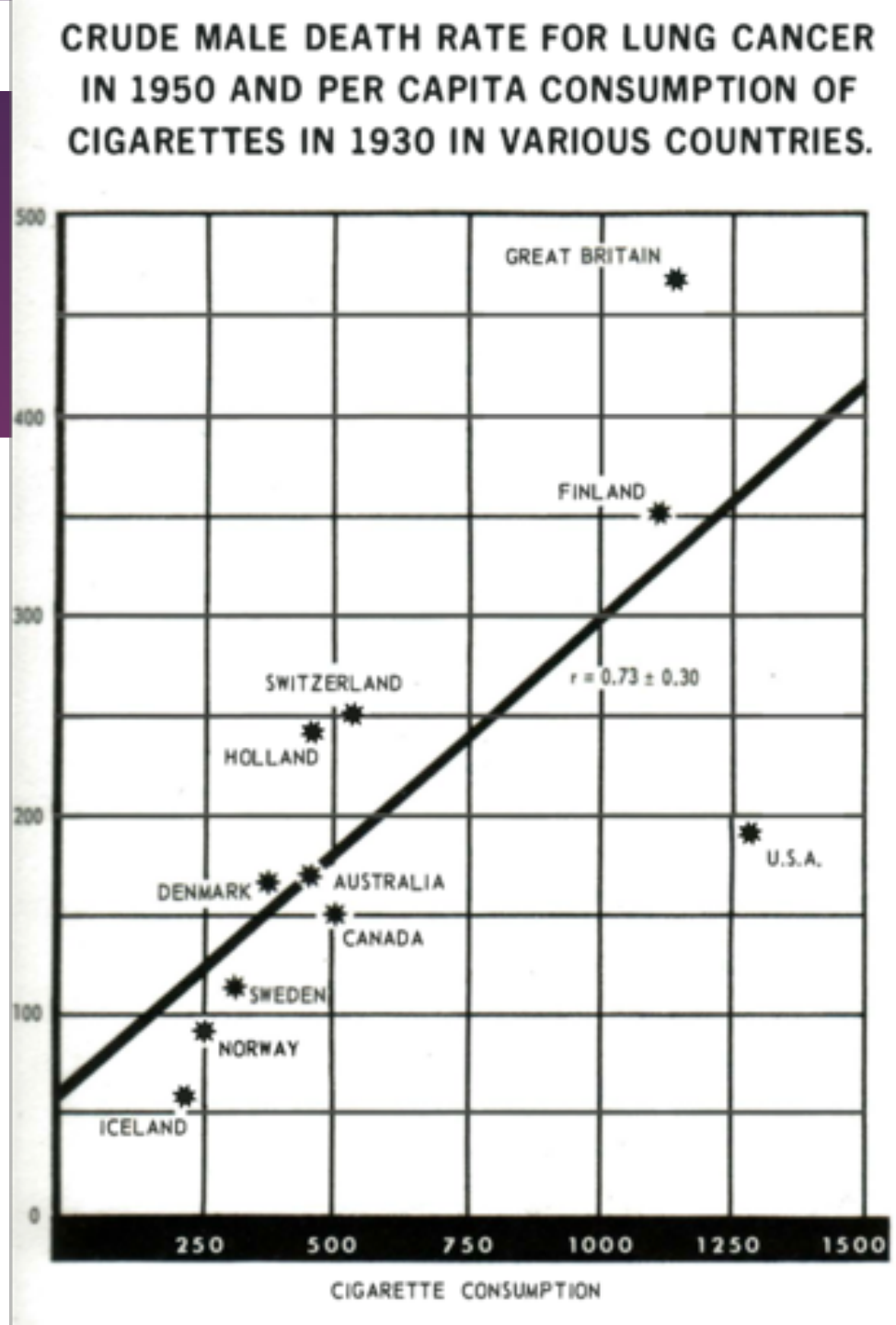


Outgoing congressional mail (tax payer expense)

Elections are in October

Relational Graphs

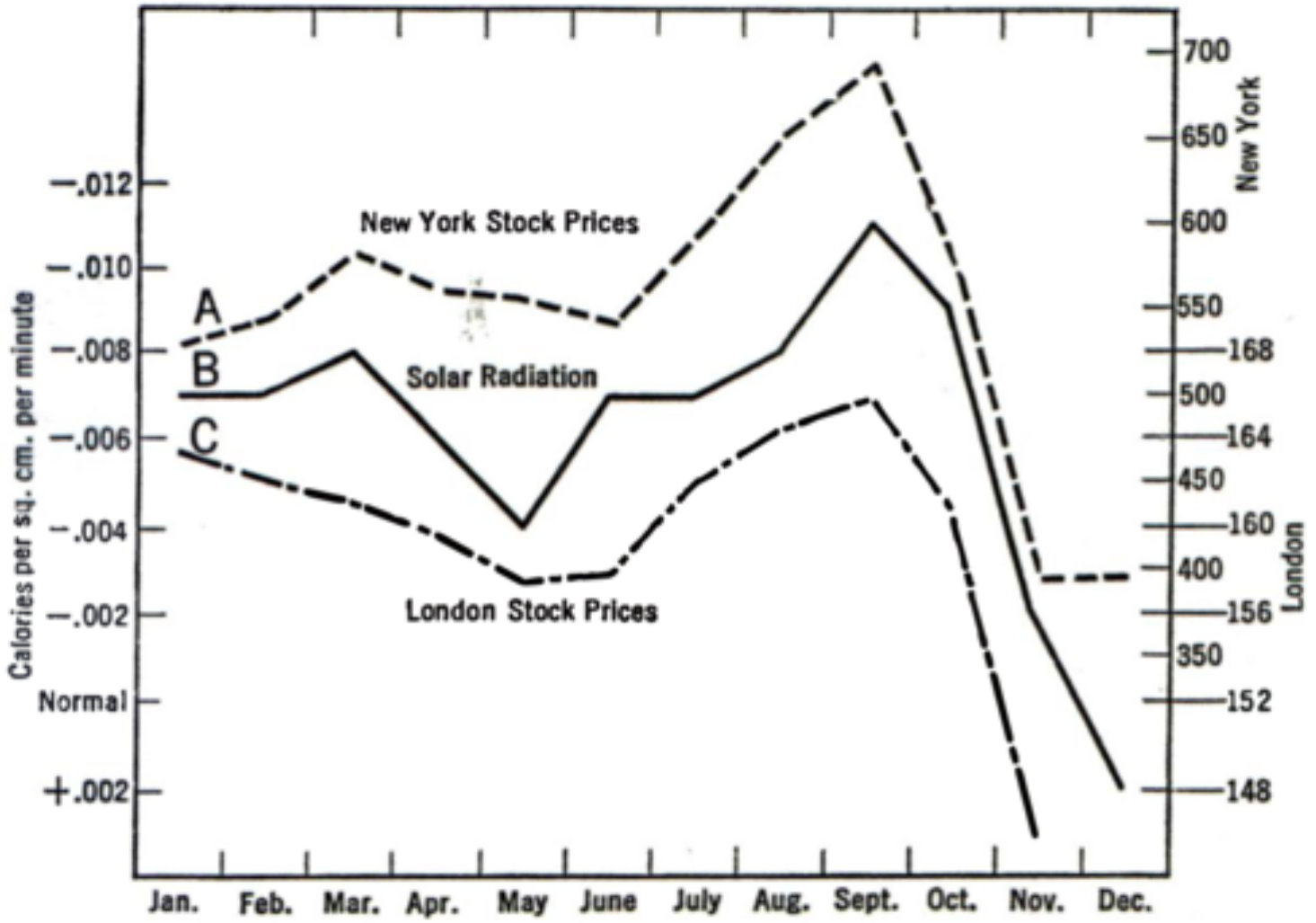
Scatter plots almost force you to see causation



Report of the Advisory Committee to the Surgeon General, Smoking and Health (Washington, D.C., 1964), p. 176; based on R. Doll, "Etiology of Lung Cancer," Advances in Cancer Research, 3 (1955), 1-50.

But can also …

Scatter plots almost force you to see causation



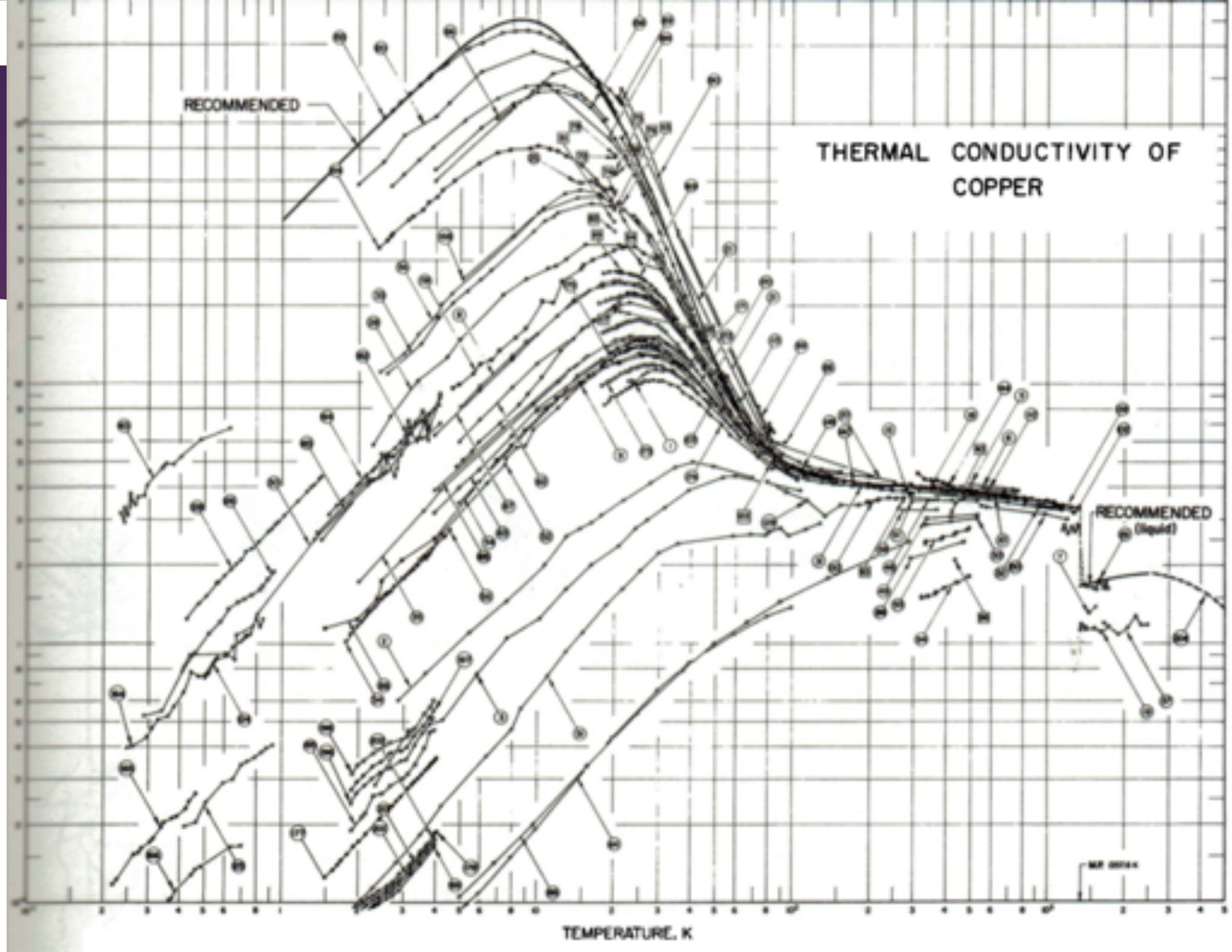
SOLAR RADIATION AND STOCK PRICES

A. New York stock prices (Barron's average). B. Solar Radiation, inverted, and C. London stock prices, all by months, 1929 (after Garcia-Mata and Shaffner).

Edward R. Dewey and Edwin F. Dakin, Cycles: The Science of Prediction (New York, 1947), p. 144

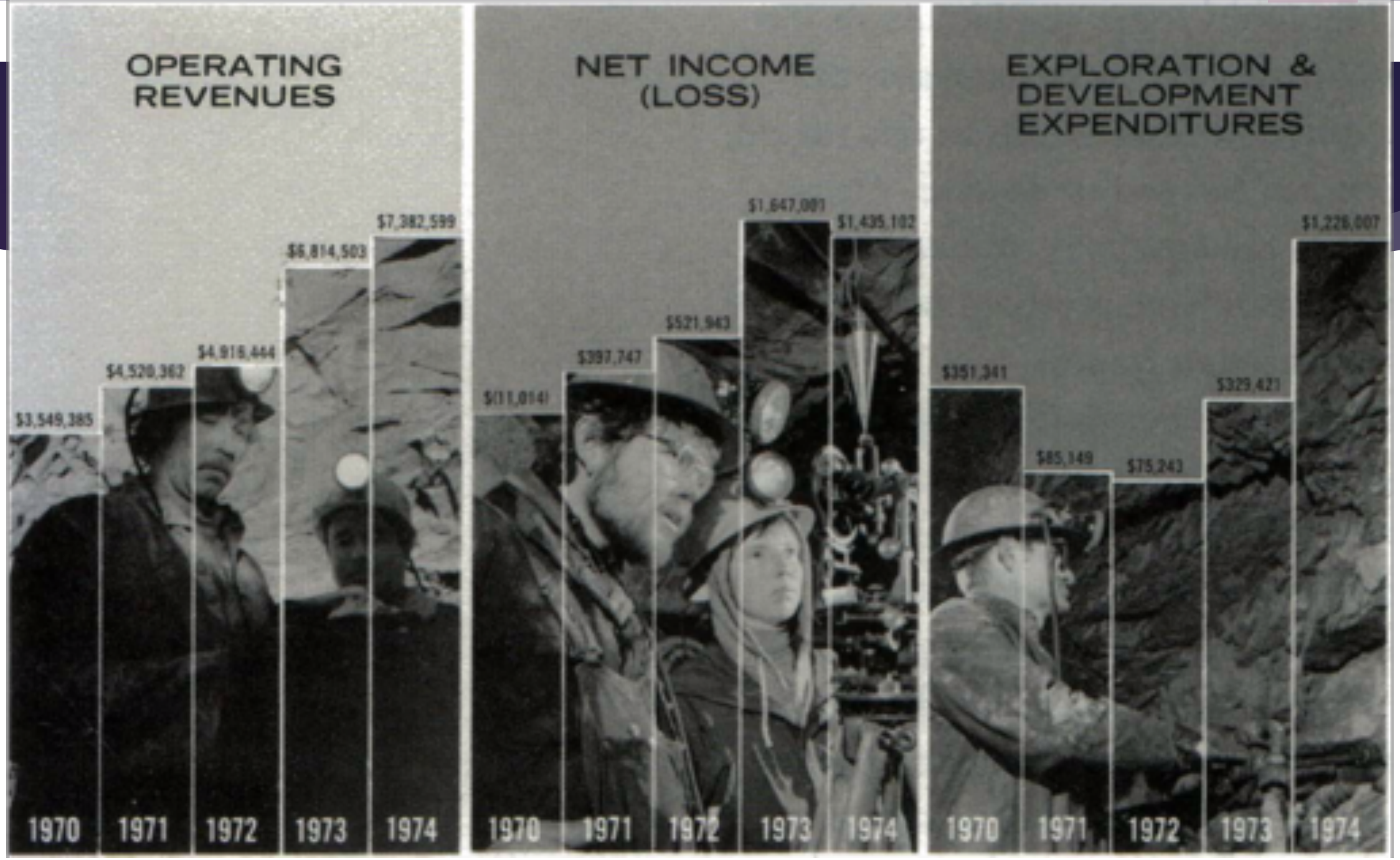
Each Curve is a different publication

The circles contain the publication ID

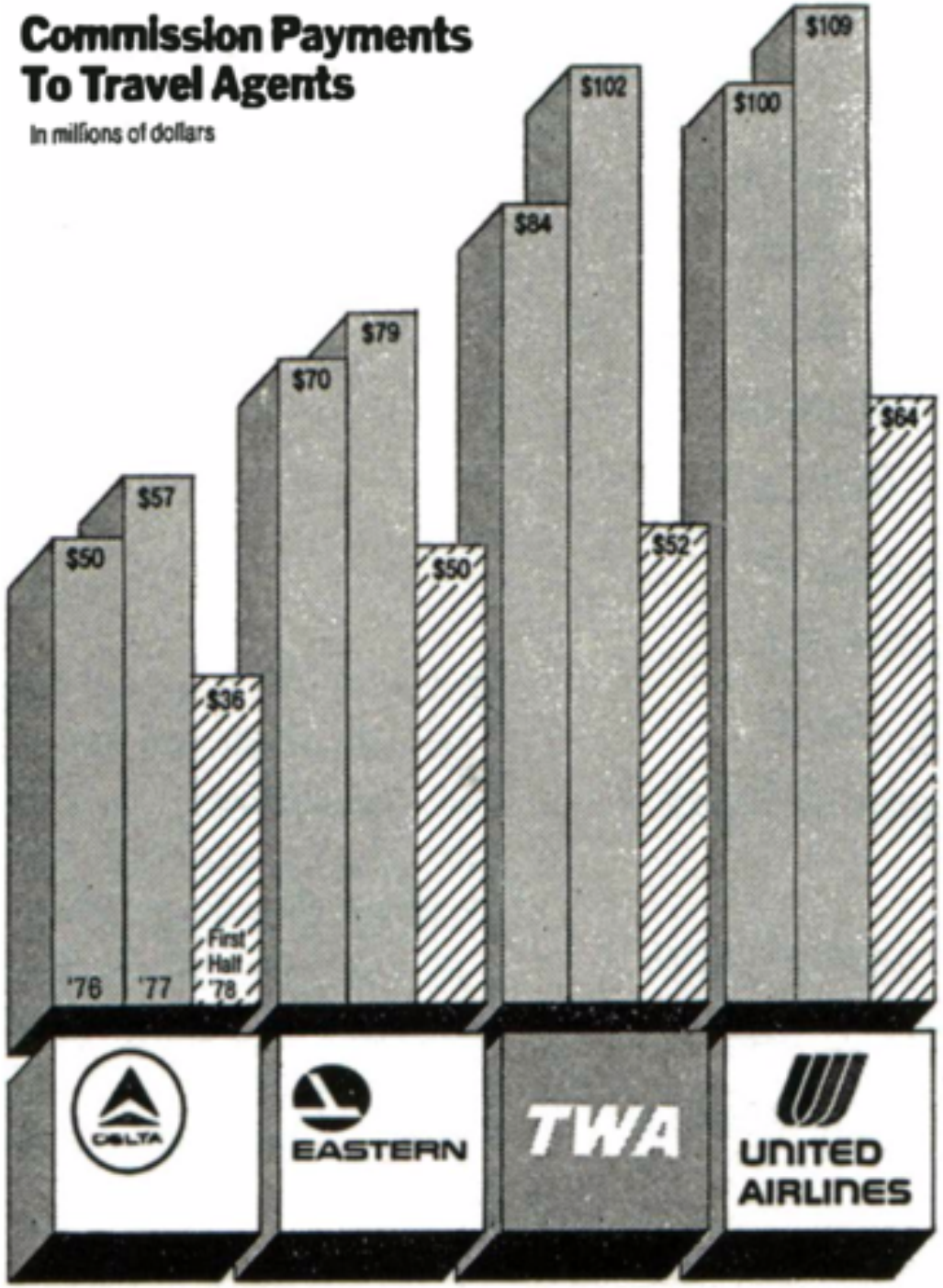


Principles of Graphical Excellence

- The well-designed presentation of interesting data: a matter of substance, of statistics, and of design
- Consists of complex ideas communicated with clarity, precision, and efficiency
- Gives to the viewer the greatest umber of ideas in the shortest time with the least [pixels] in the smallest space
- Nearly always multivariate
- Tells the truth about the data
- **graphics are instruments for reasoning about quantitative information.**

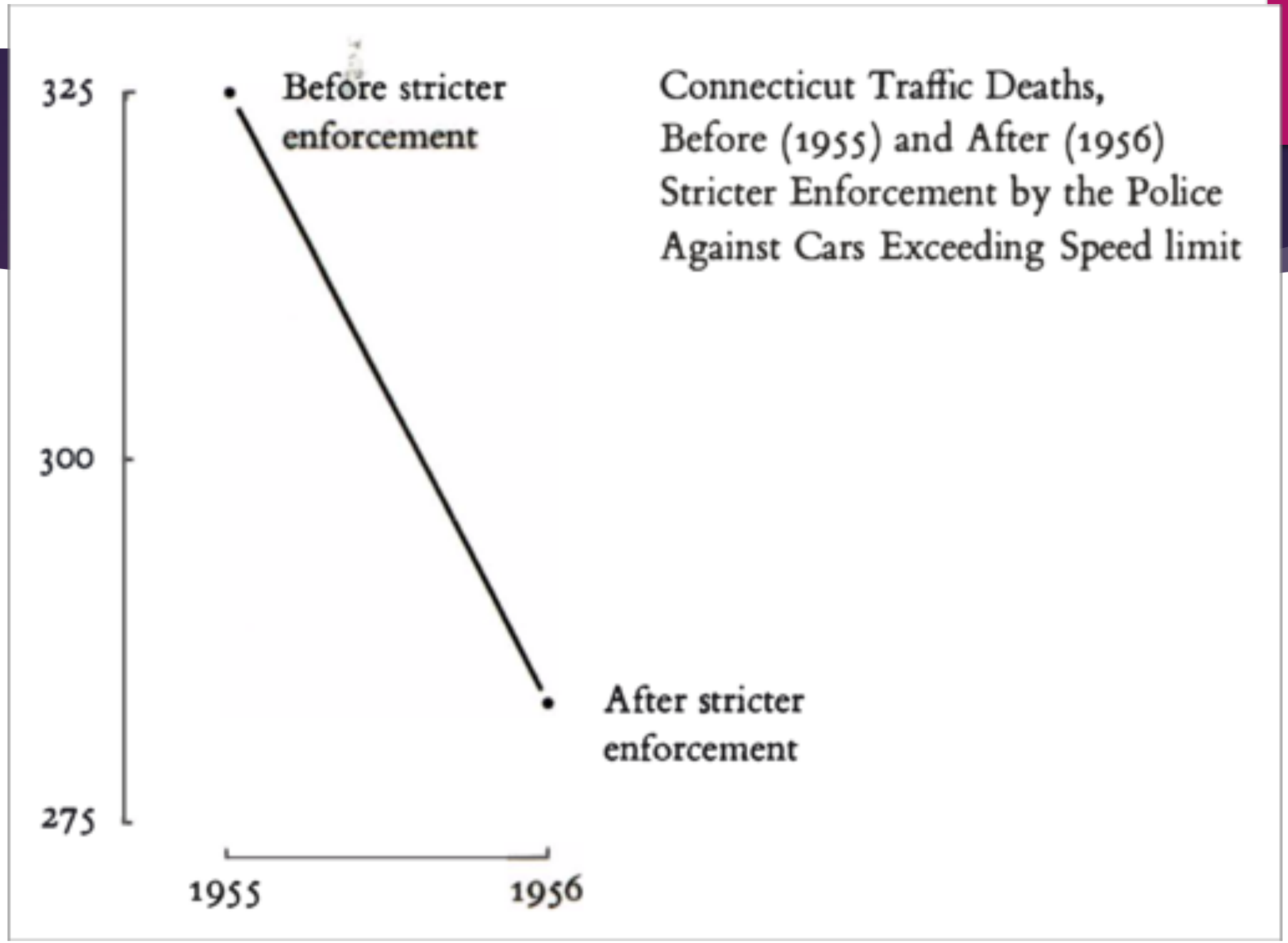


Day Mines, Inc., 1974 Annual Report, p. 1.

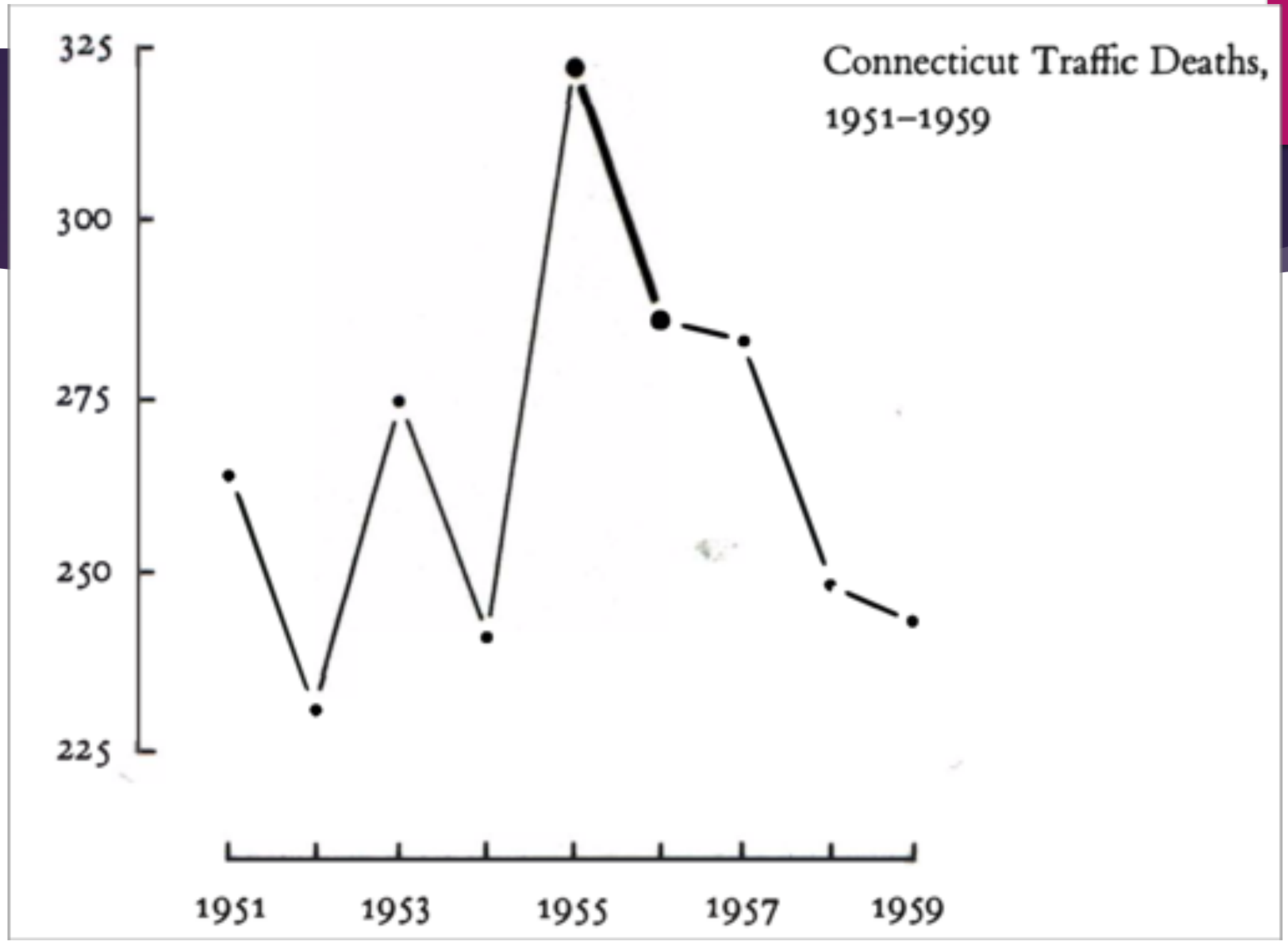


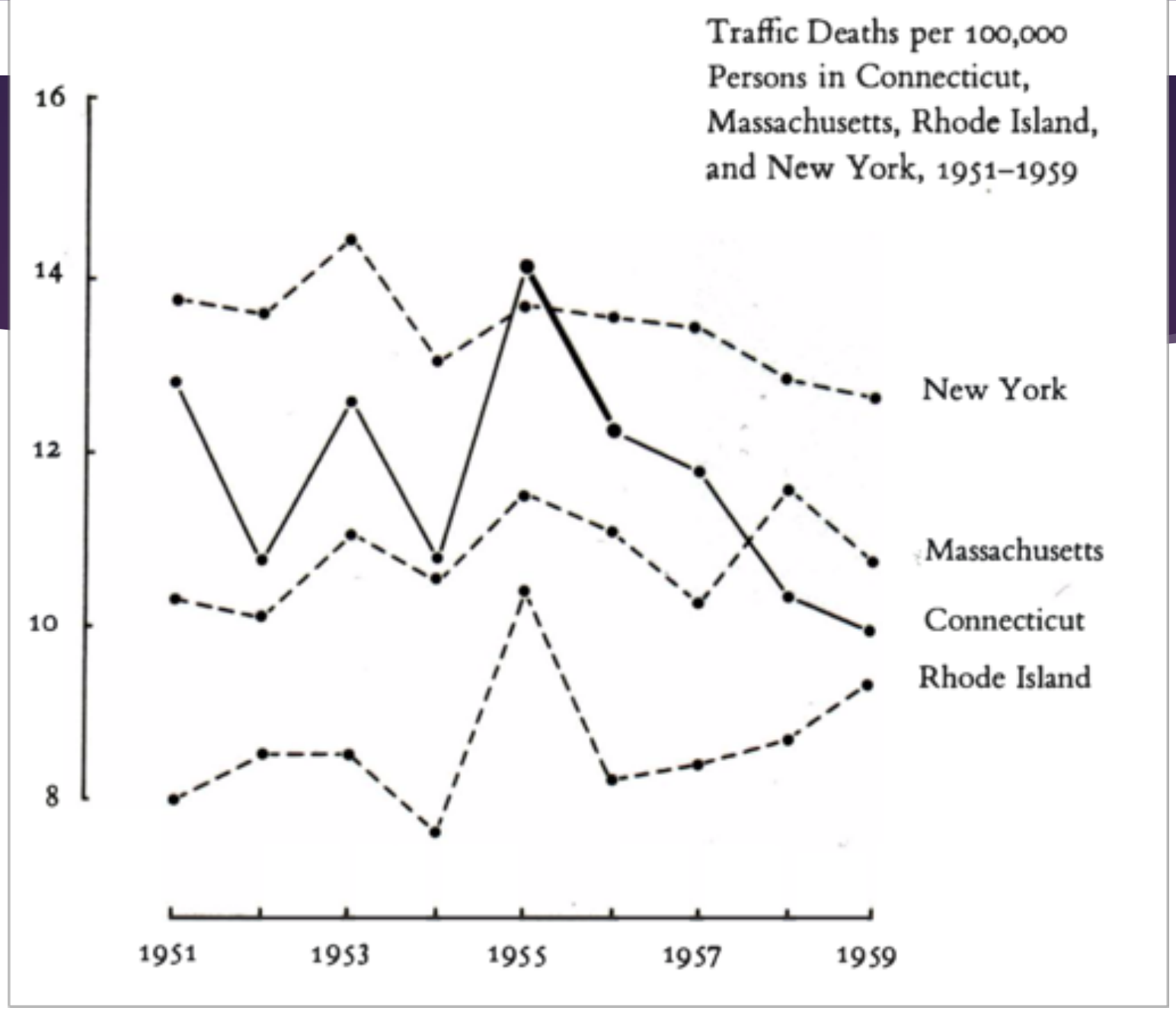
In a New York Time's article about declining travel agent commissions.

New York Times, August 8, 1978, p. D-1.



Connecticut Traffic Deaths, Before (1955) and After (1956) Stricter Enforcement by the Police Against Cars Exceeding Speed limit


Connecticut Traffic Deaths,
1951–1959
325
300
275
250
225
1951
1953
1955
1957
1959



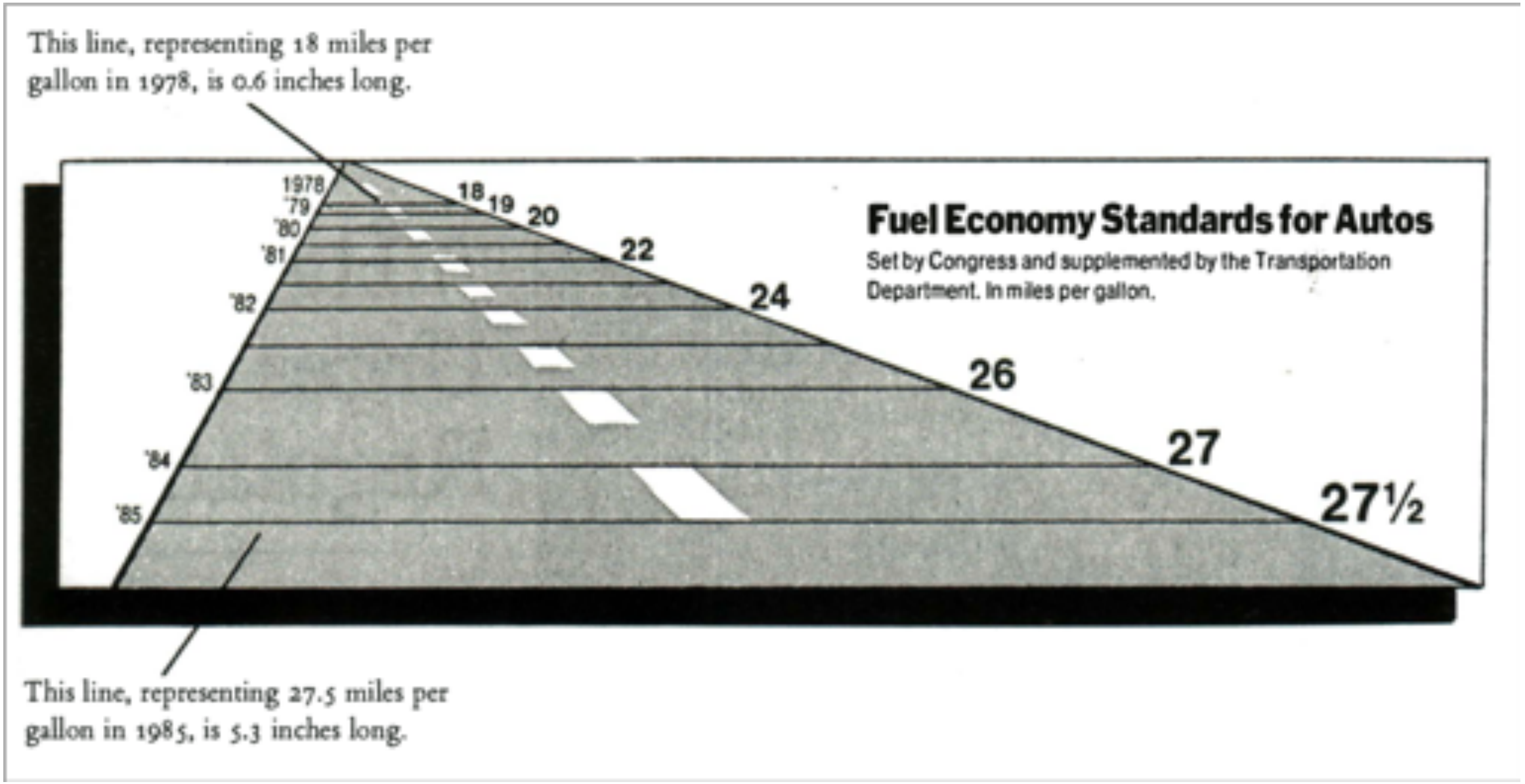
Donald T.Campbell and H. Laurence Ross, "The Connecticut Crackdown on Speeding: Time Series Data in Quasi Experimental Analysis," in Edward R.. Tufte, ed., The Q□umtitative A11alysis of Social Problems (Reading, Mass., 1970), 110-125.

Рост продукции промышленности (1922 г. = 1).

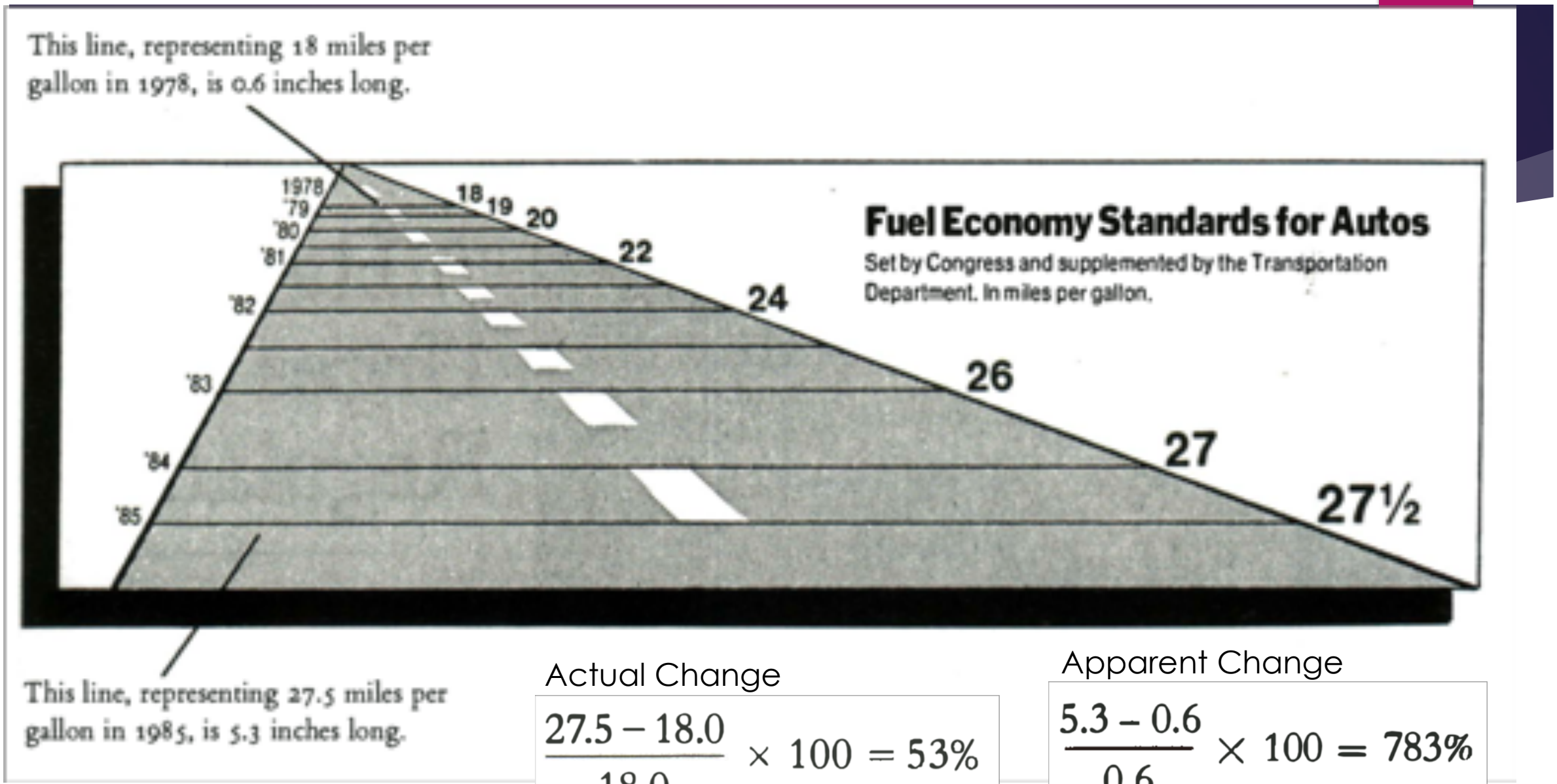
Pravda, May 24, 1982, p. 2.

Lie Factor

$$\text{Lie factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$



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



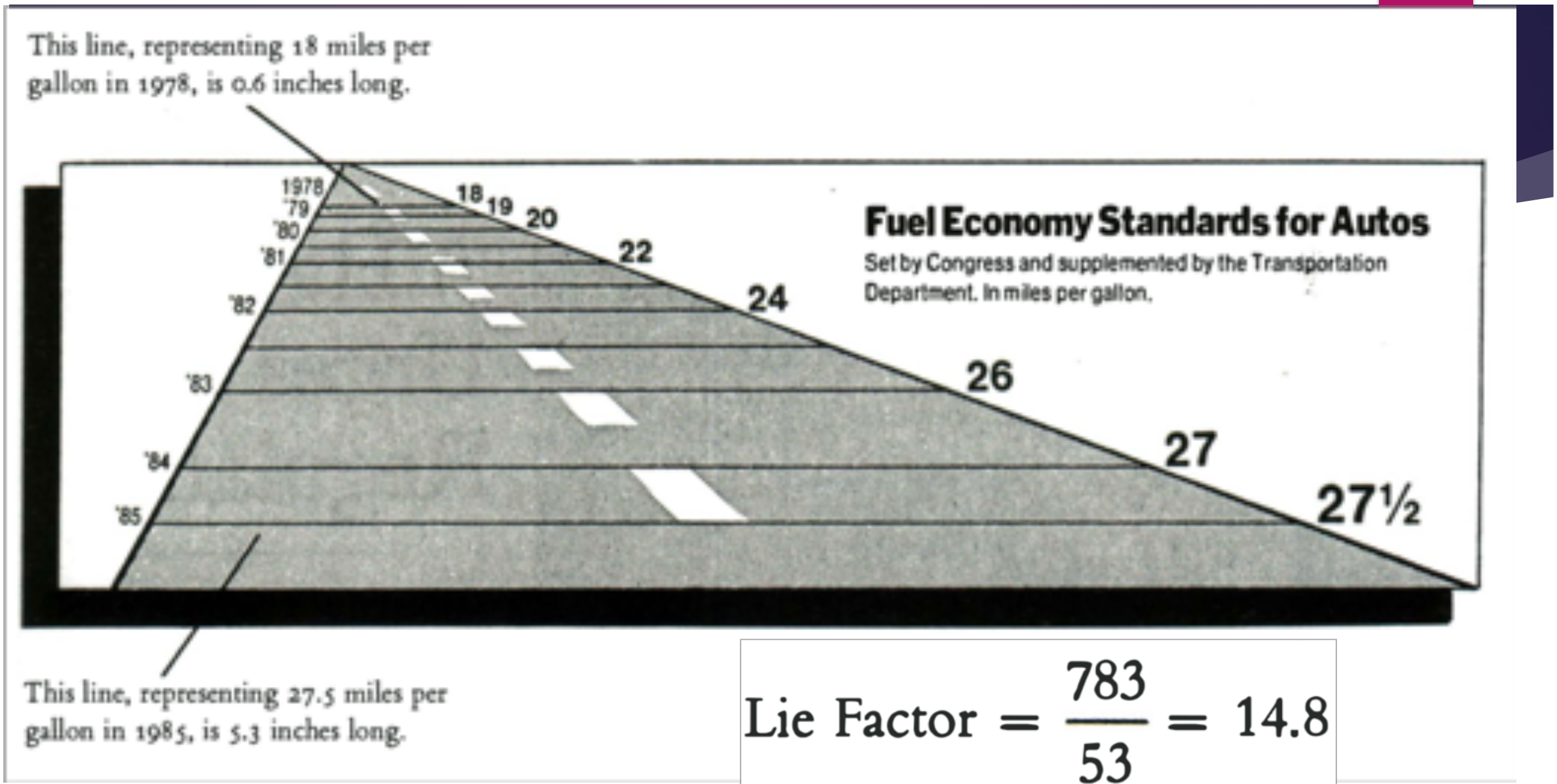
Actual Change

$$\frac{27.5 - 18.0}{18.0} \times 100 = 53\%$$

Apparent Change

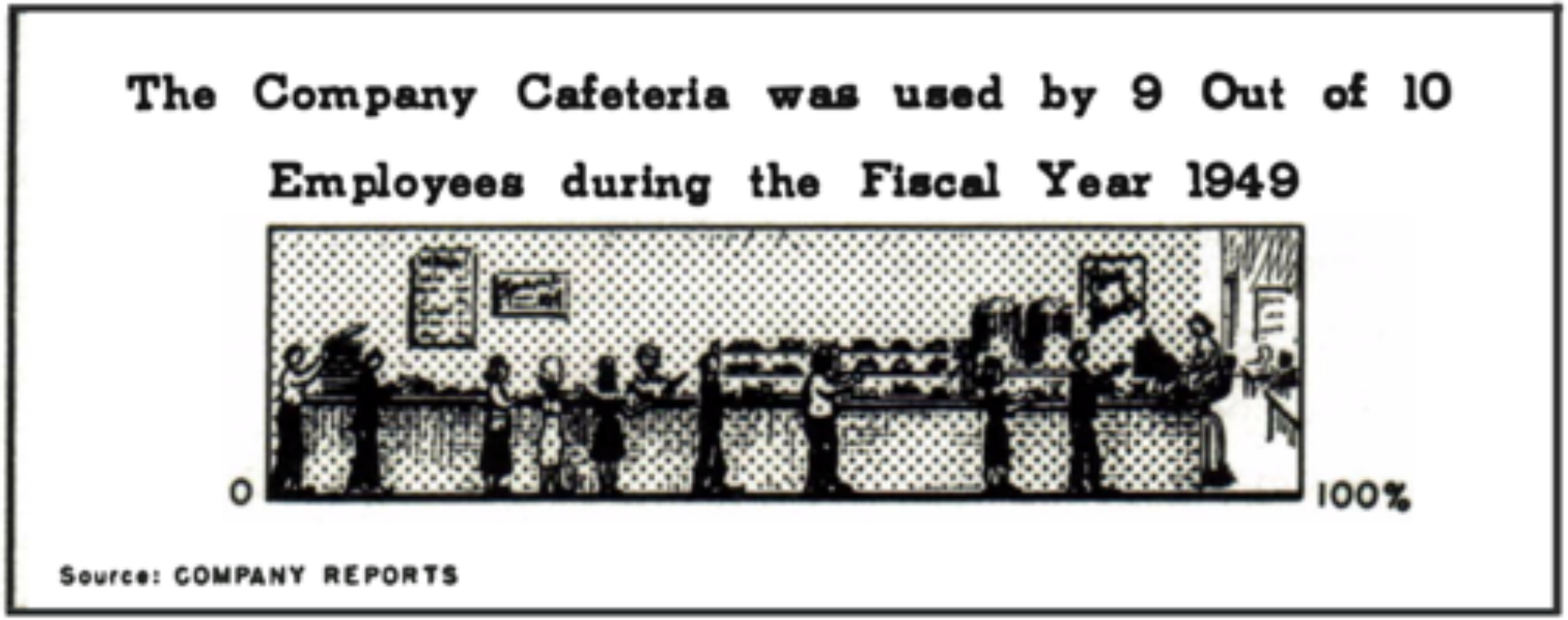
$$\frac{5.3 - 0.6}{0.6} \times 100 = 783\%$$

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



$$\text{Lie Factor} = \frac{783}{53} = 14.8$$

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



Mary Eleanor Spear, Charting Statistics (New York, 1952), p. 5

“Everyone spoke of an information overload, but what there was in fact was a non-information overload.”

- Richard Saul Wurman, *What-If, Could-Be* (Philadelphia, 1976)

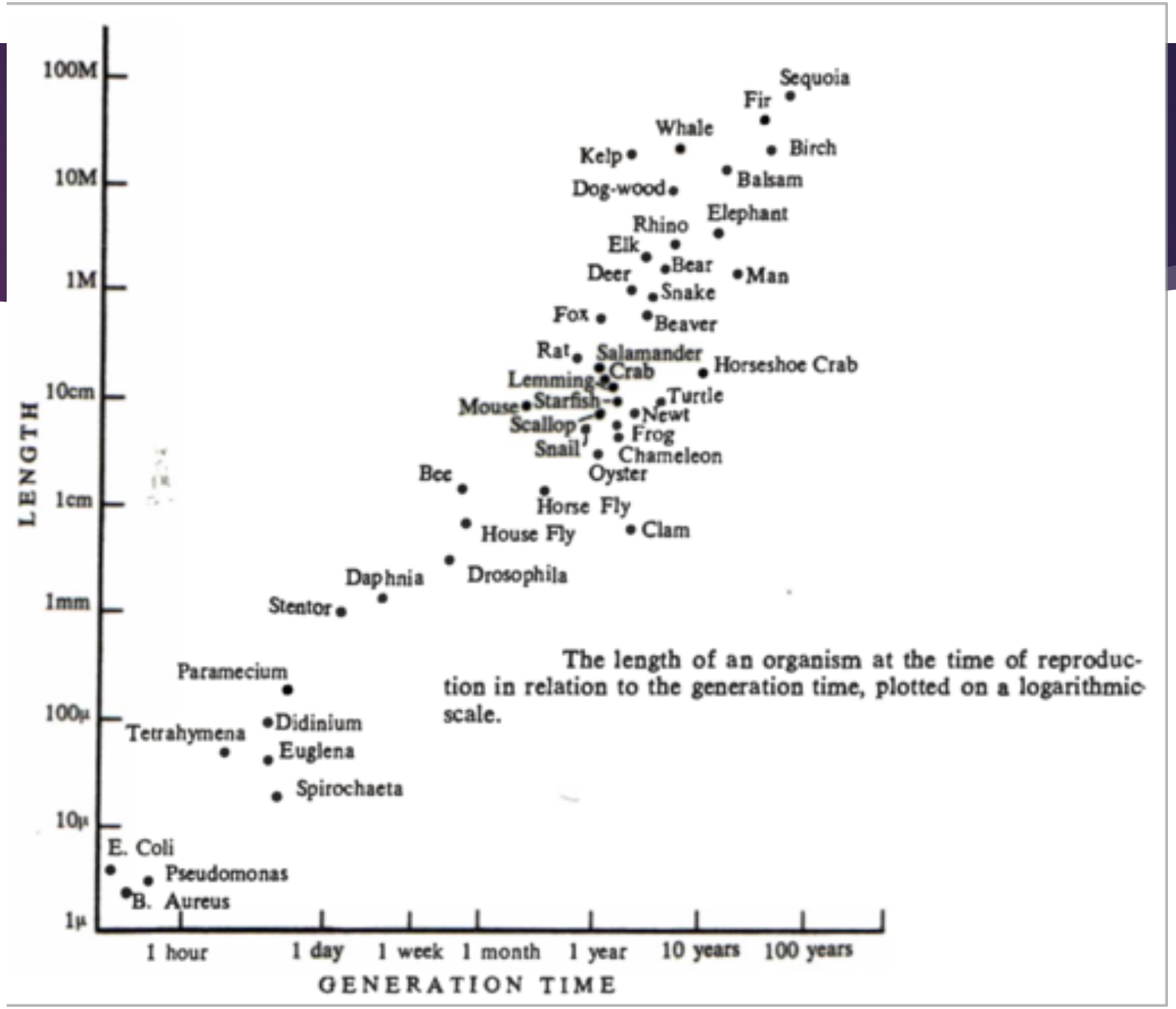
Data-ink Ratio

$$\text{Data-ink ratio} = \frac{\text{data-ink}}{\text{total ink}}$$

= proportion of a graphic's 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.

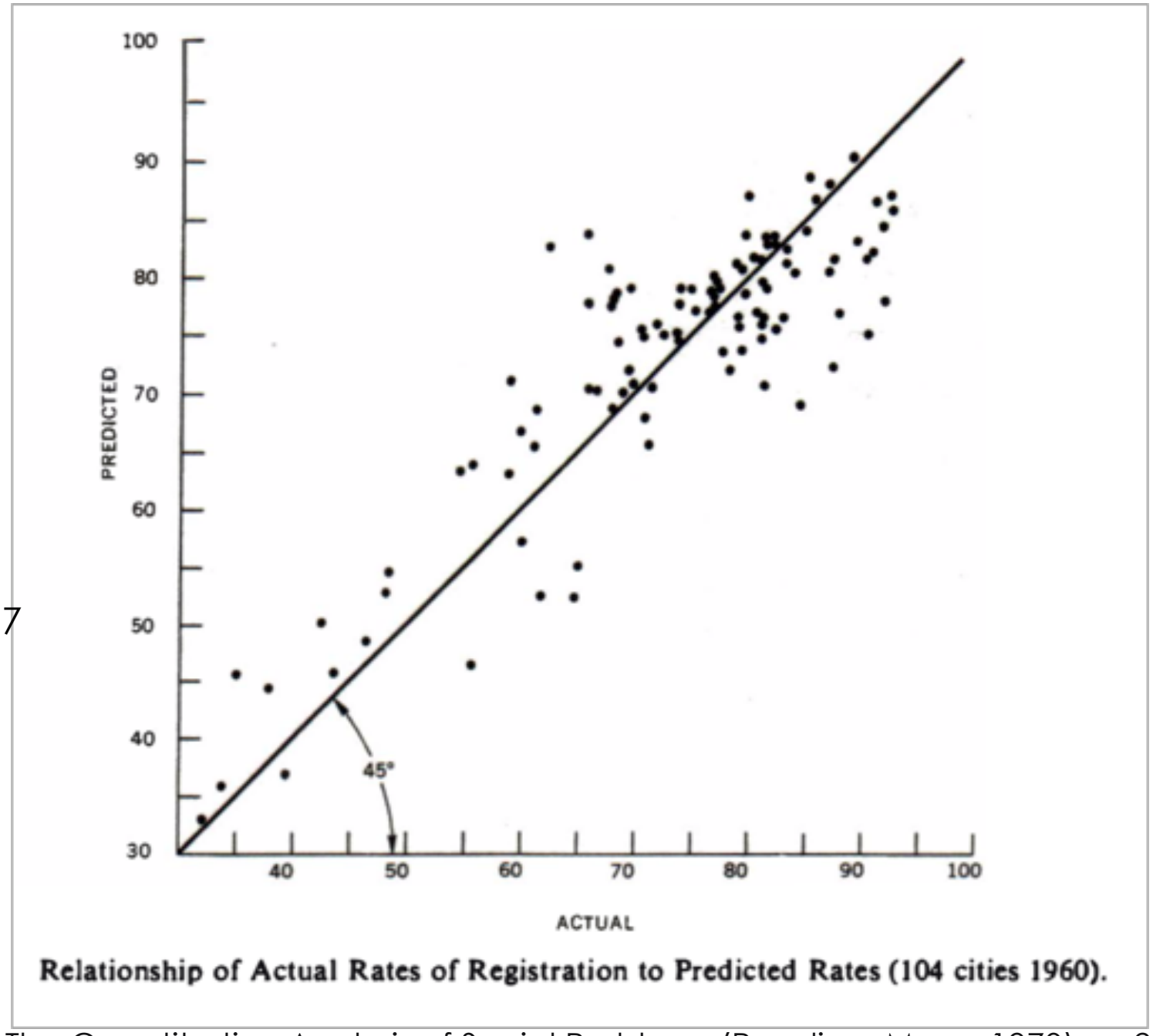
Data-ink ratio=0.8-0.9



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

John Tyler Bonner, Size and Cycle: An Essay on the Structure of Biology (Prince ton, 1965), p. 17.

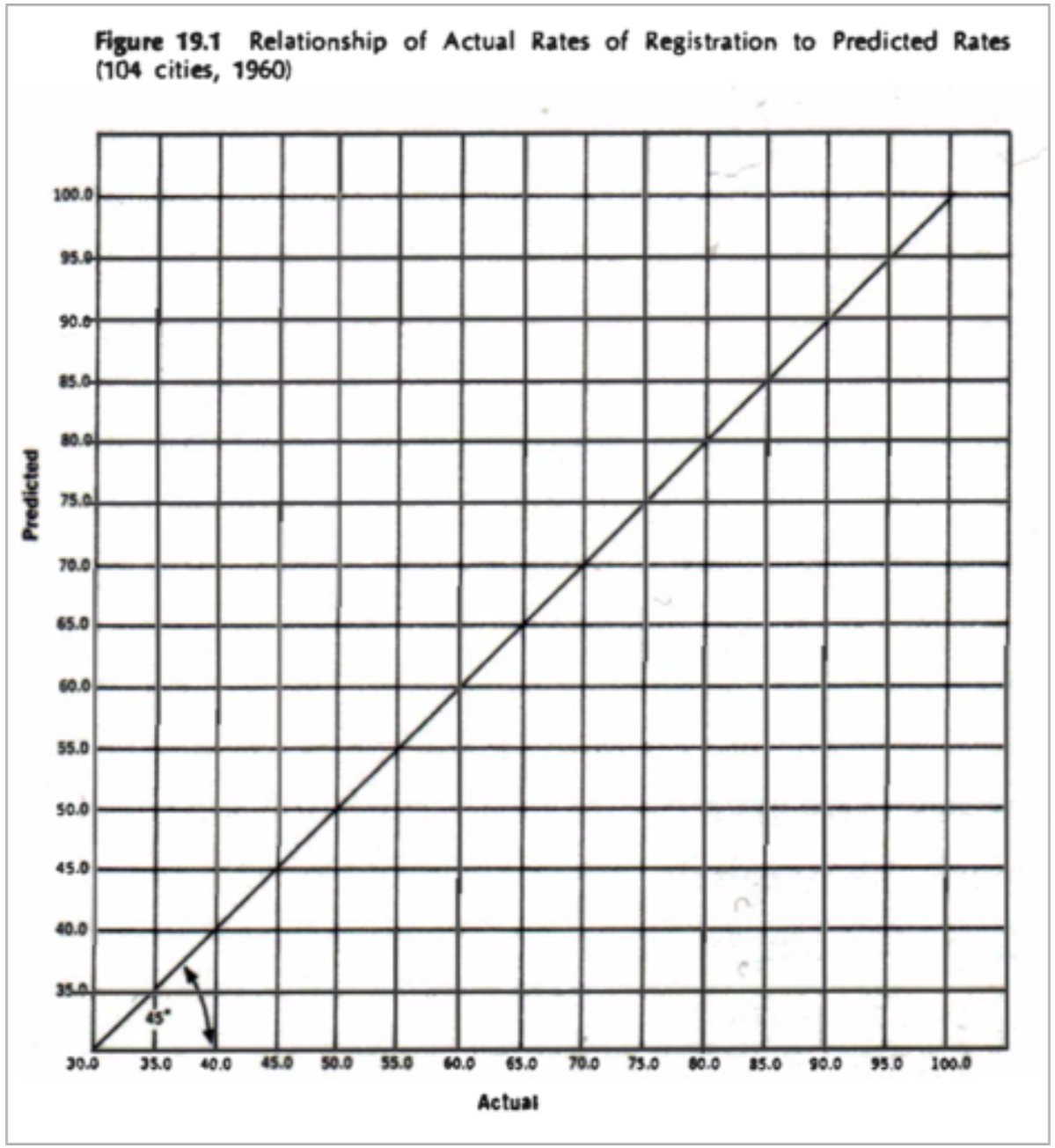
Data-ink ratio=0.7



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

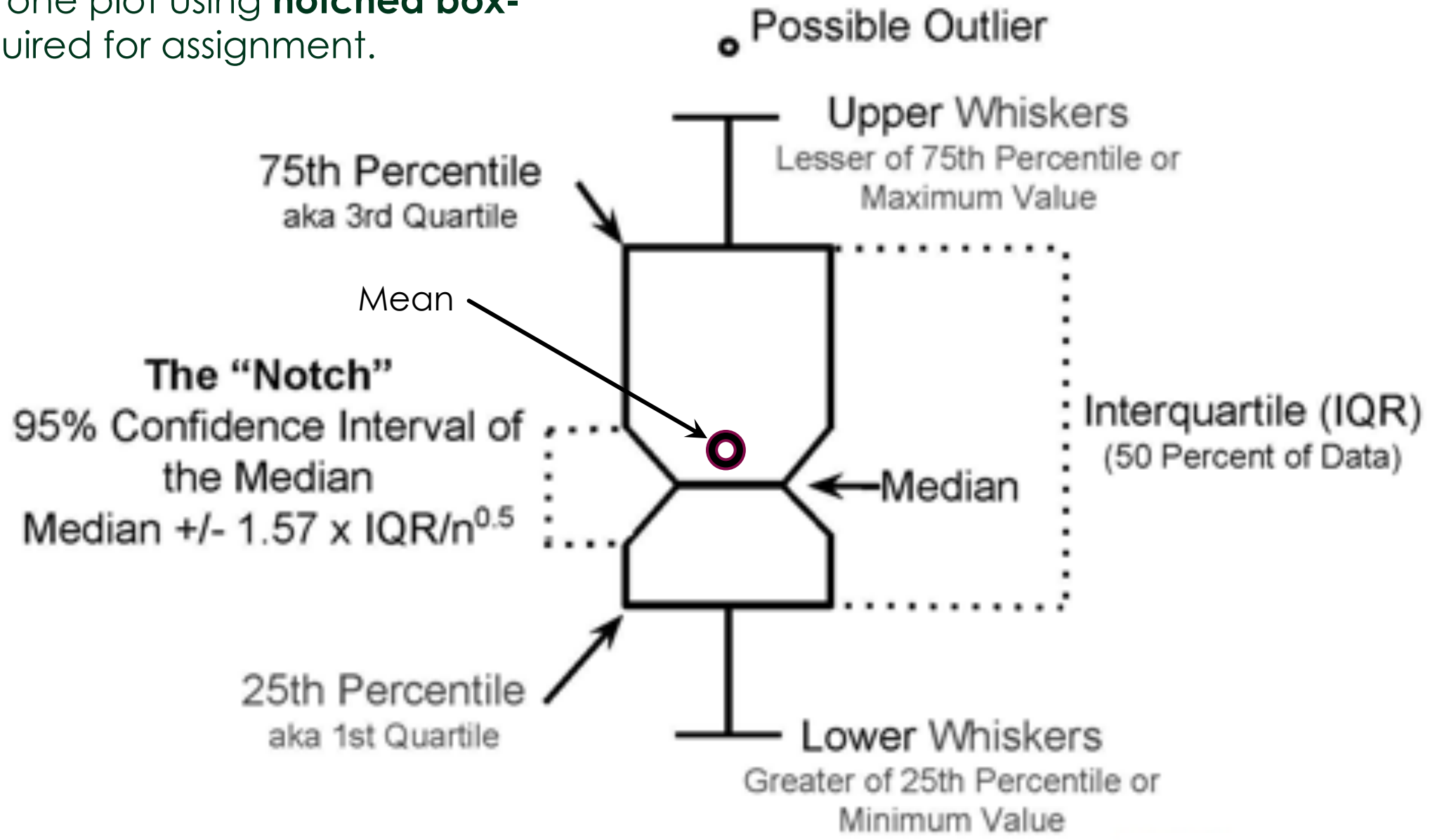
Edward R. Tufte, ed., The Quantita tive Analysis of Social Problems (Reading, Mass., 1970), p. 267

Data-ink ratio=0

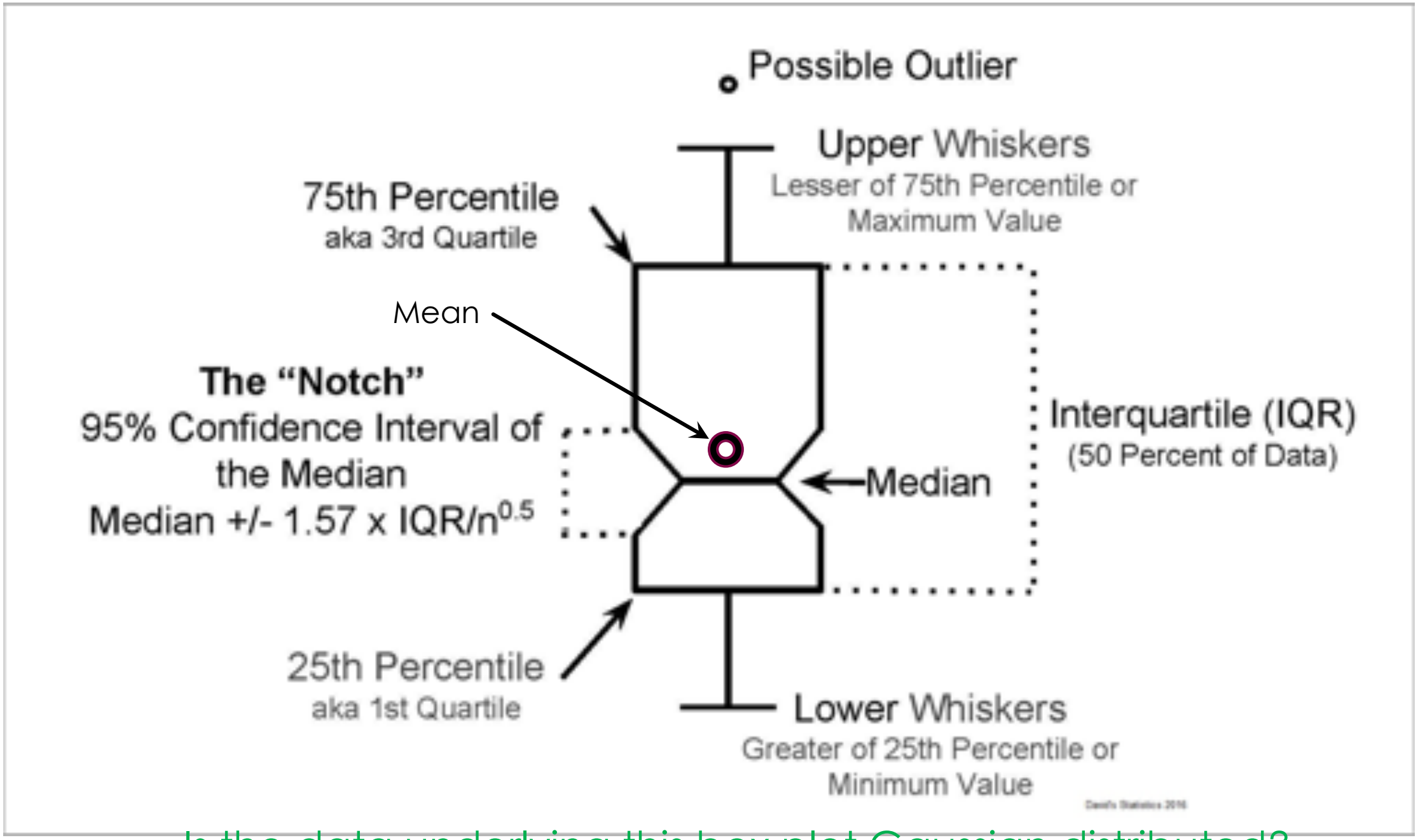


William]. Crotty, ed., Public Opinion and Politics: A Reader (New York, 1970), p. 364.

At least one plot using **notched box-plot** required for assignment.



Communicate a lot about the data concisely.



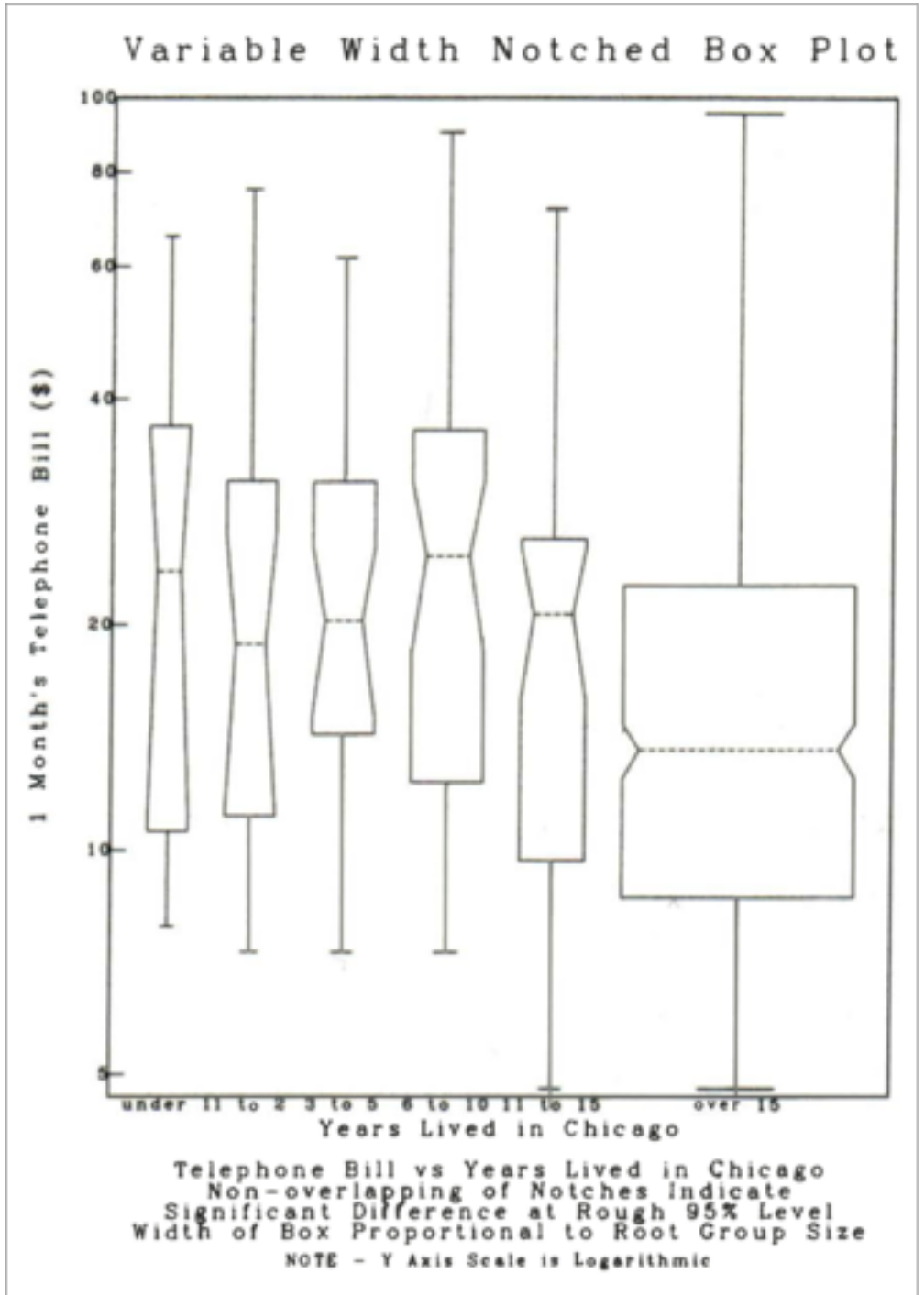
Is the data underlying this box plot Gaussian distributed?

Assigned Reading:

Robert McGill, John W. Tukey, and Wayne A. Larsen, "Variations of Box Plots," American Statistician, 32 (1978), 12-16.

https://www.tandfonline.com/doi/pdf/10.1080/00031305.1978.10479236

What would you conclude from this plot?



Logistics

- Assignment 3 Due in:

- 4 days, 6 hours, 3 minutes, 33 seconds
- 102.06 h
- 6123.55 min
- 367413 s

Violin plots show the distribution even more clearly

Showing p-values

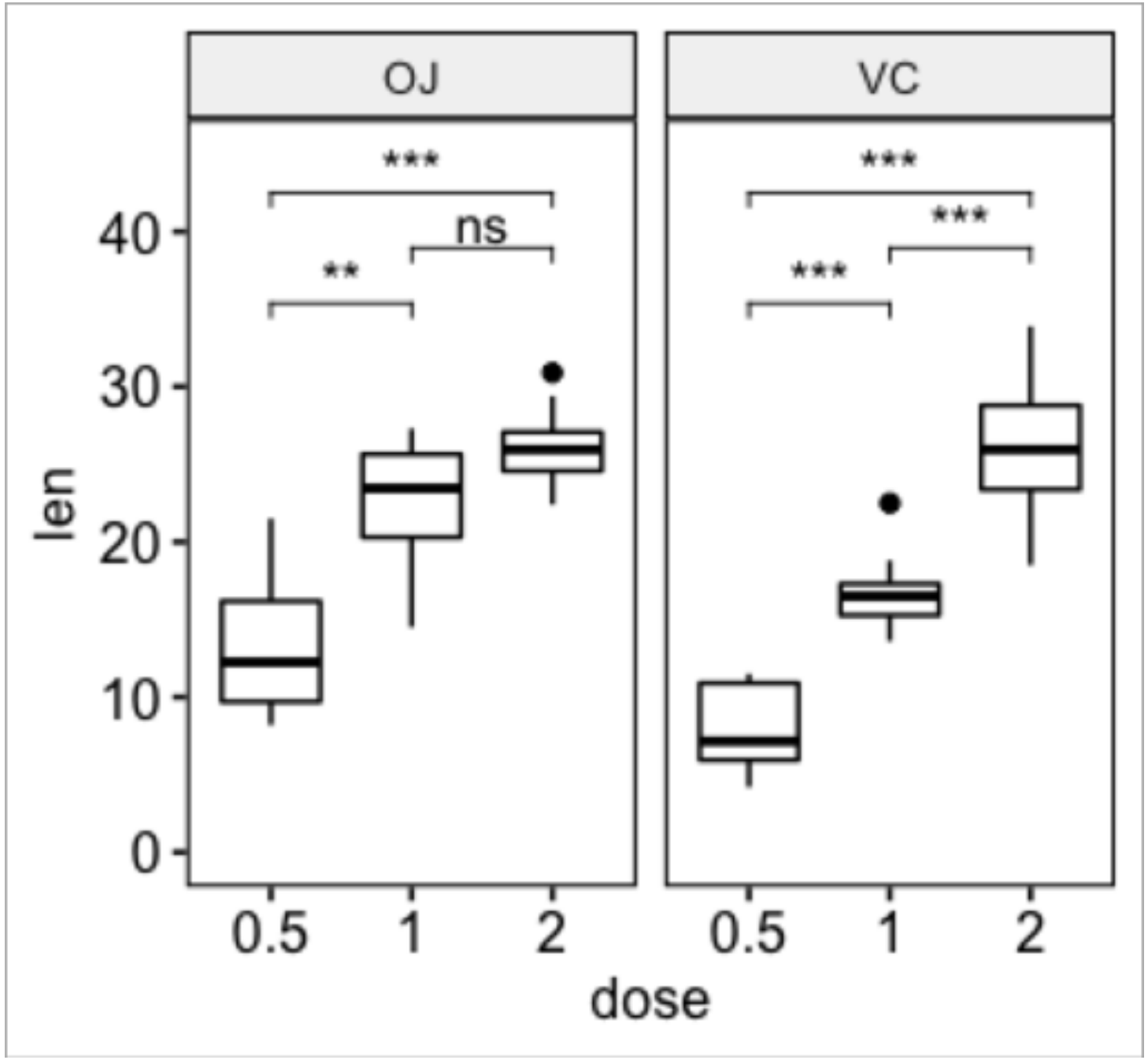
NS: $p \geq 0.05$

*: $p < 0.05$

**: $p < 0.01$

***: $p < 0.001$

****: $p < 0.0001$

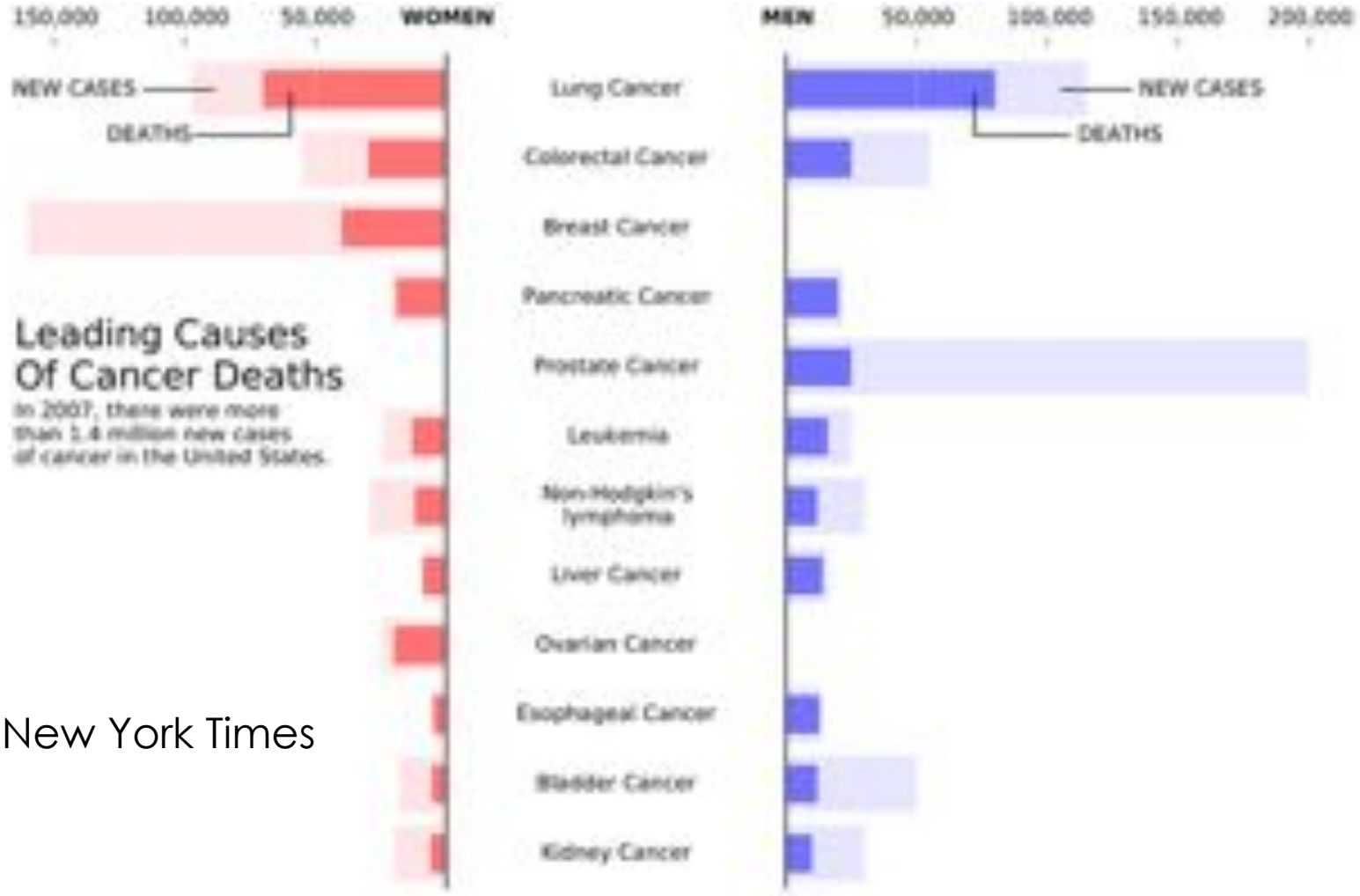


12 Simple Rules for Better Figures

Rule 1: Know your audience
Rule 2: Identify your message
Rule 3: Adapt the figure to the support medium
Rule 4: Captions are not optional
Rule 5: Do not trust the defaults
Rule 6: Use color effectively
Rule 7: Do not mislead the reader
Rule 8: Avoid "Chartjunk"
Rule 9: Message trumps beauty
Rule 10: Get the right tool
Rule 11: Make sure every element is legible when printed
Rule 12: DO NOT CUT AND PASTE BITMAP GRAPHICS

Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLoS Comput Biol 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833

Figure 1. Know your audience.



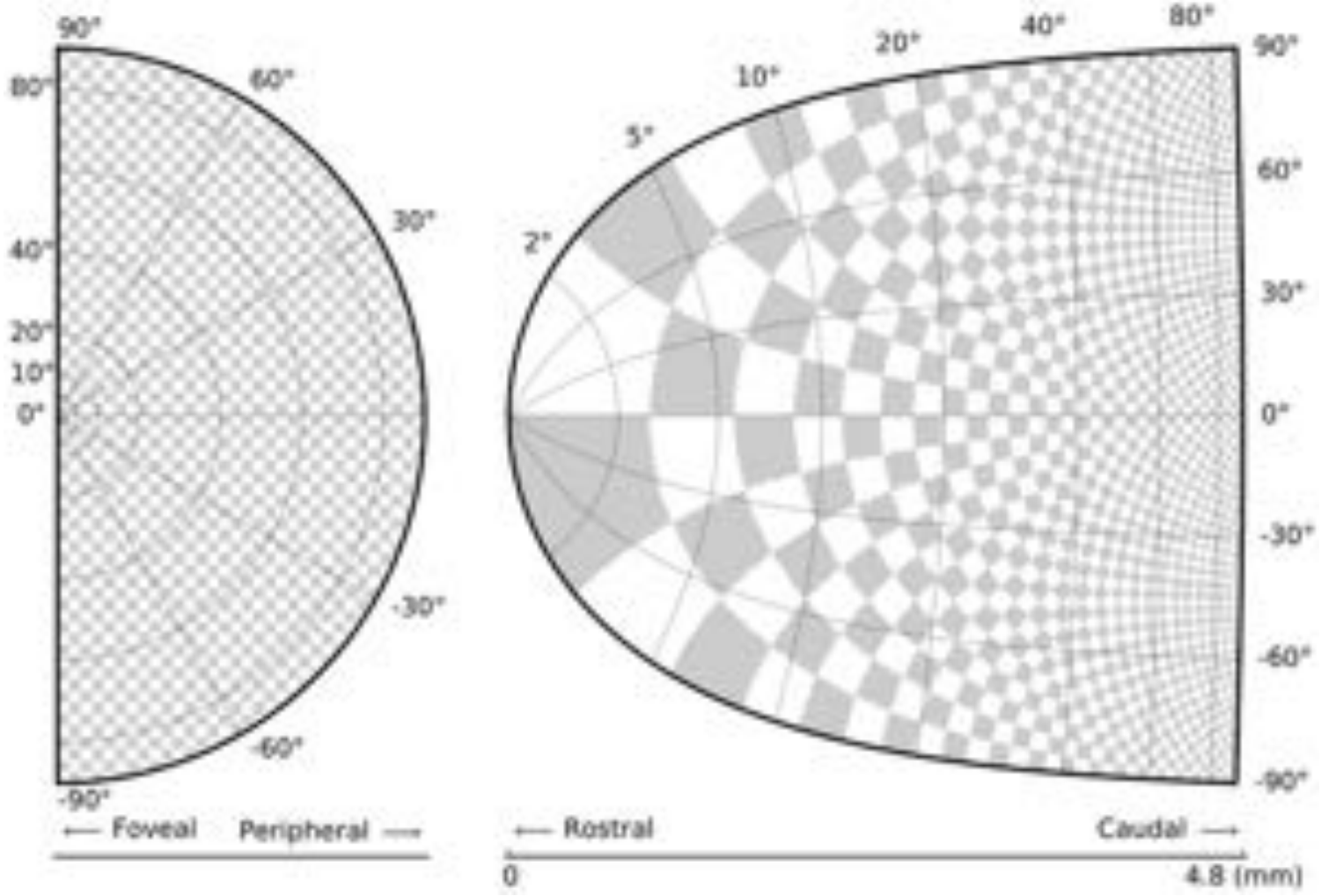
Published in The New York Times

Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

PLOS | COMPUTATIONAL BIOLOGY

Figure 2. Identify your message.

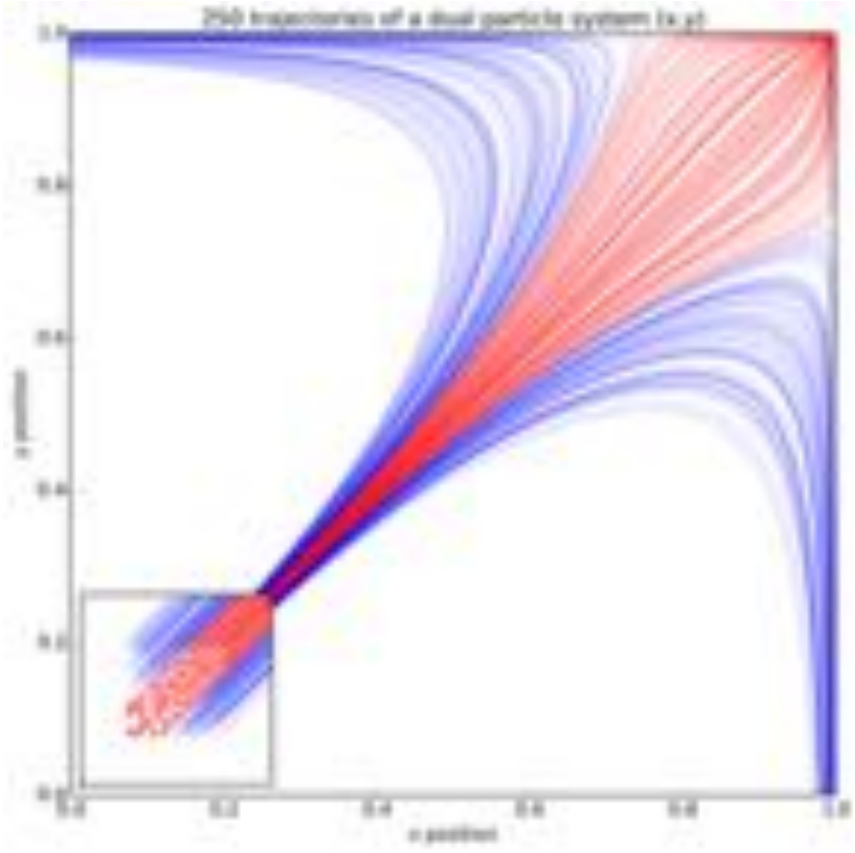
Figure 2. The superior colliculus (SC) is a brainstem structure at the crossroads of multiple functional pathways. The relationship between retina and the neurons is logarithmic.

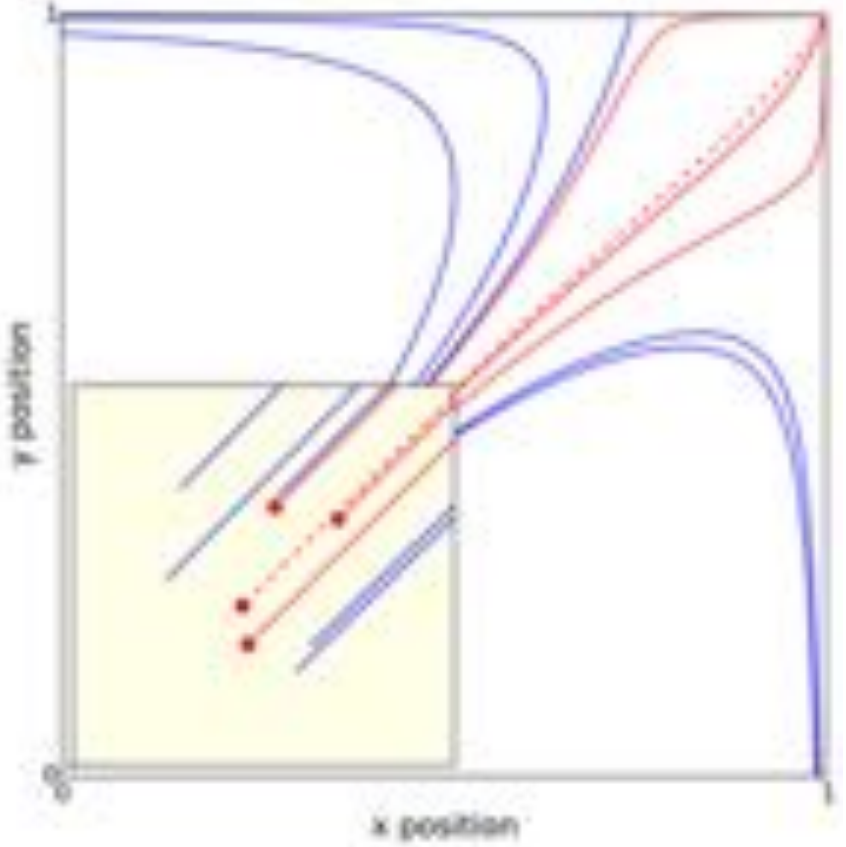


Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

PLOS | COMPUTATIONAL BIOLOGY

Figure 3. Adapt the figure to the support medium.





Journal Paper

Oral Presentation

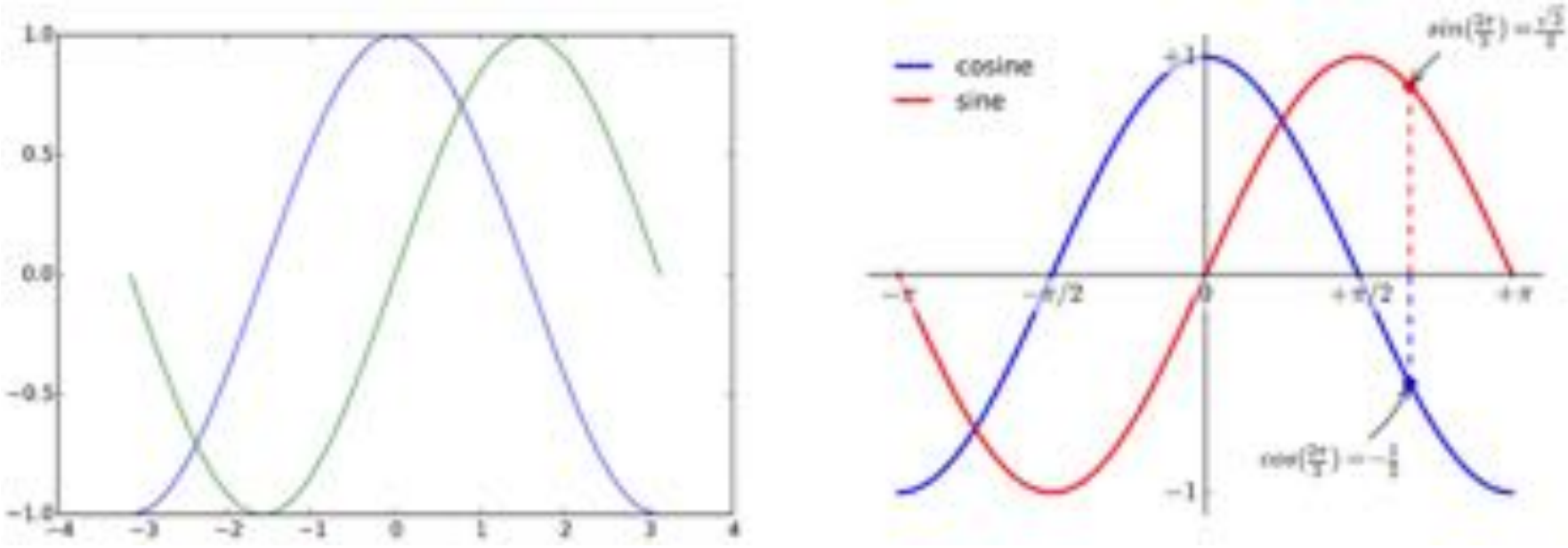
Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

PLOS | COMPUTATIONAL BIOLOGY

Rule 4: Captions are NOT optional

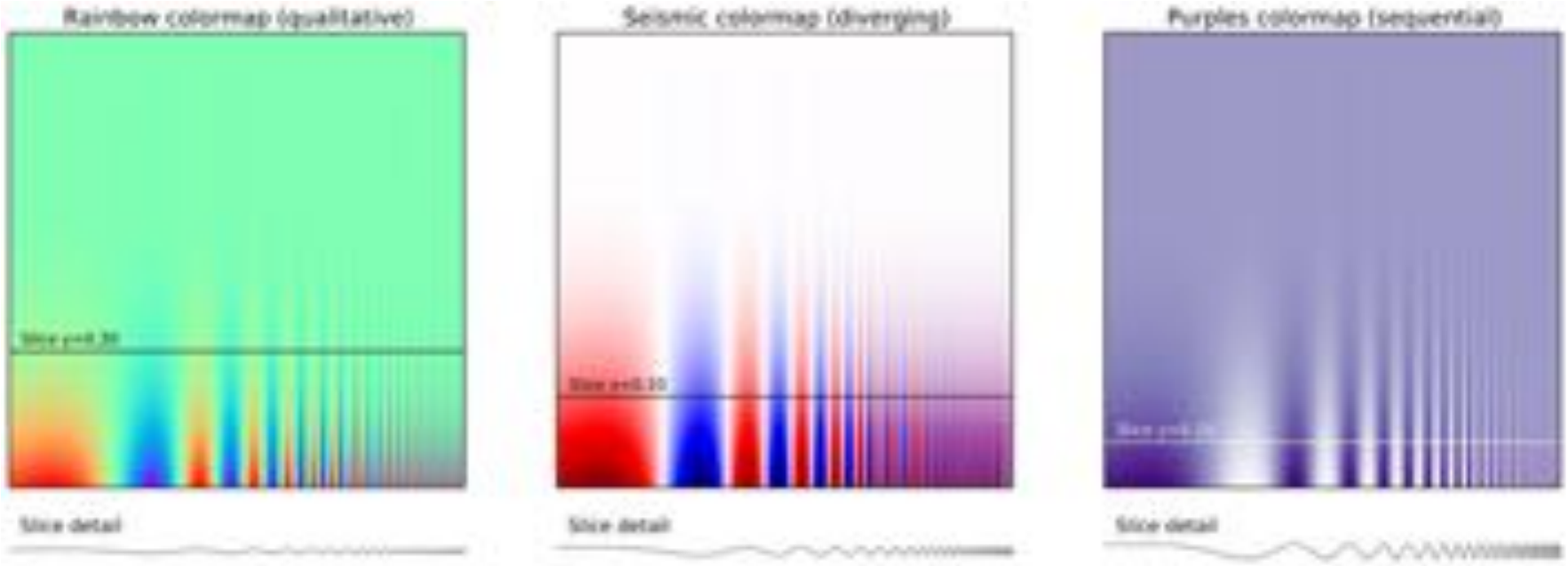
This is where you explain how your figure is supposed to work. Also include the message your figure conveys.

Figure 5. Do not trust the defaults.



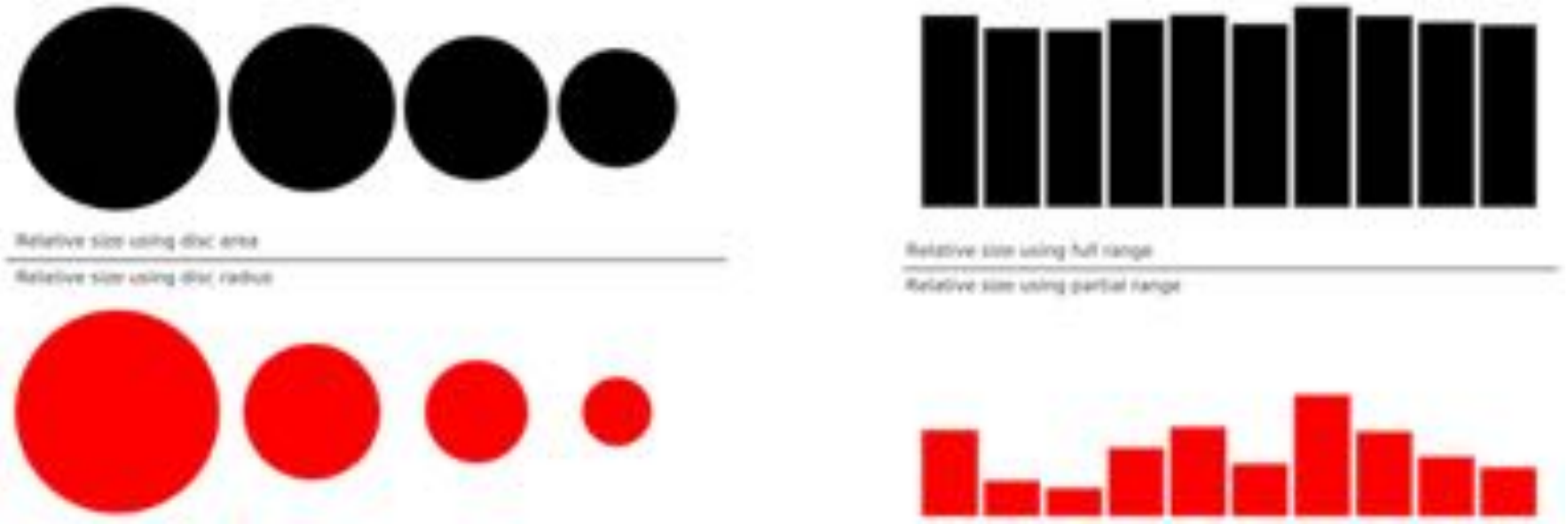
Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

Figure 6. Use color effectively.



Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

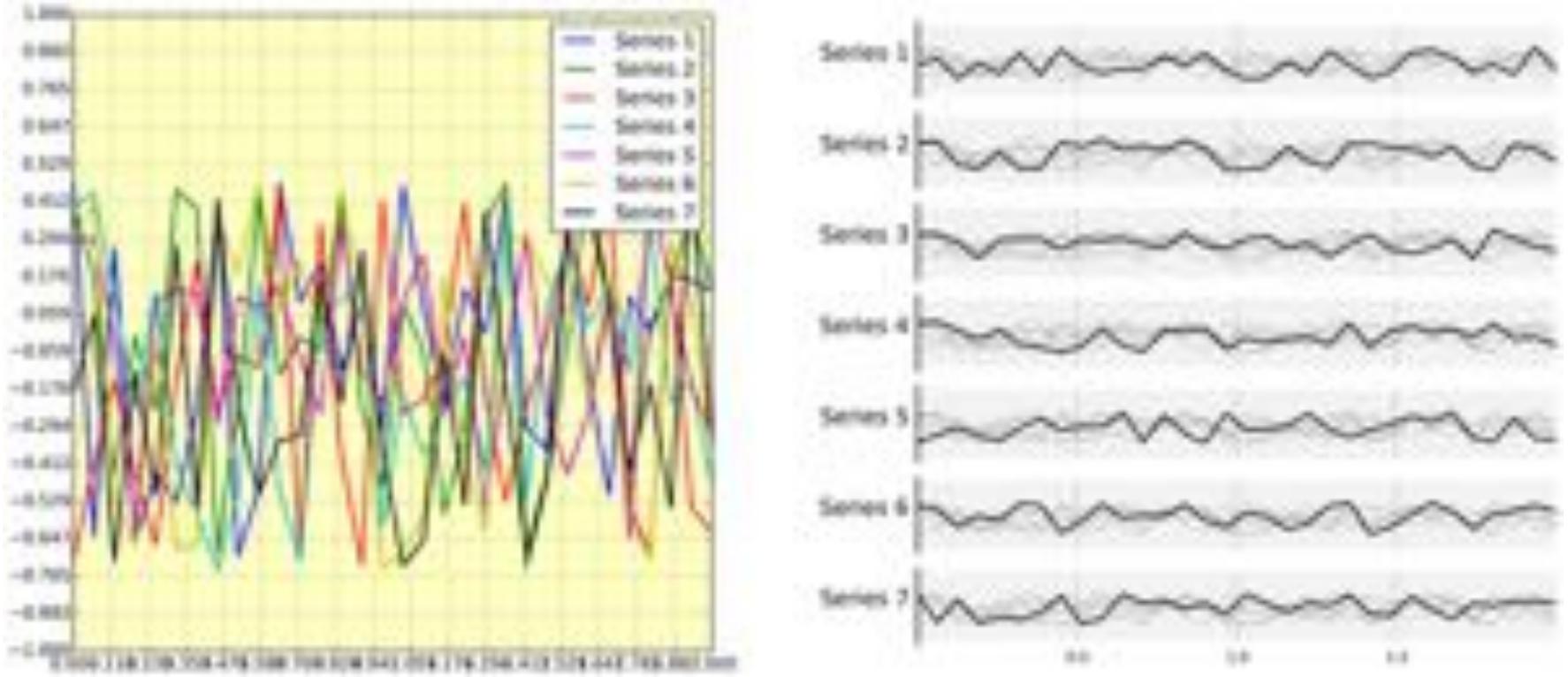
Figure 7. Do not mislead the reader.



Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

PLOS | COMPUTATIONAL BIOLOGY

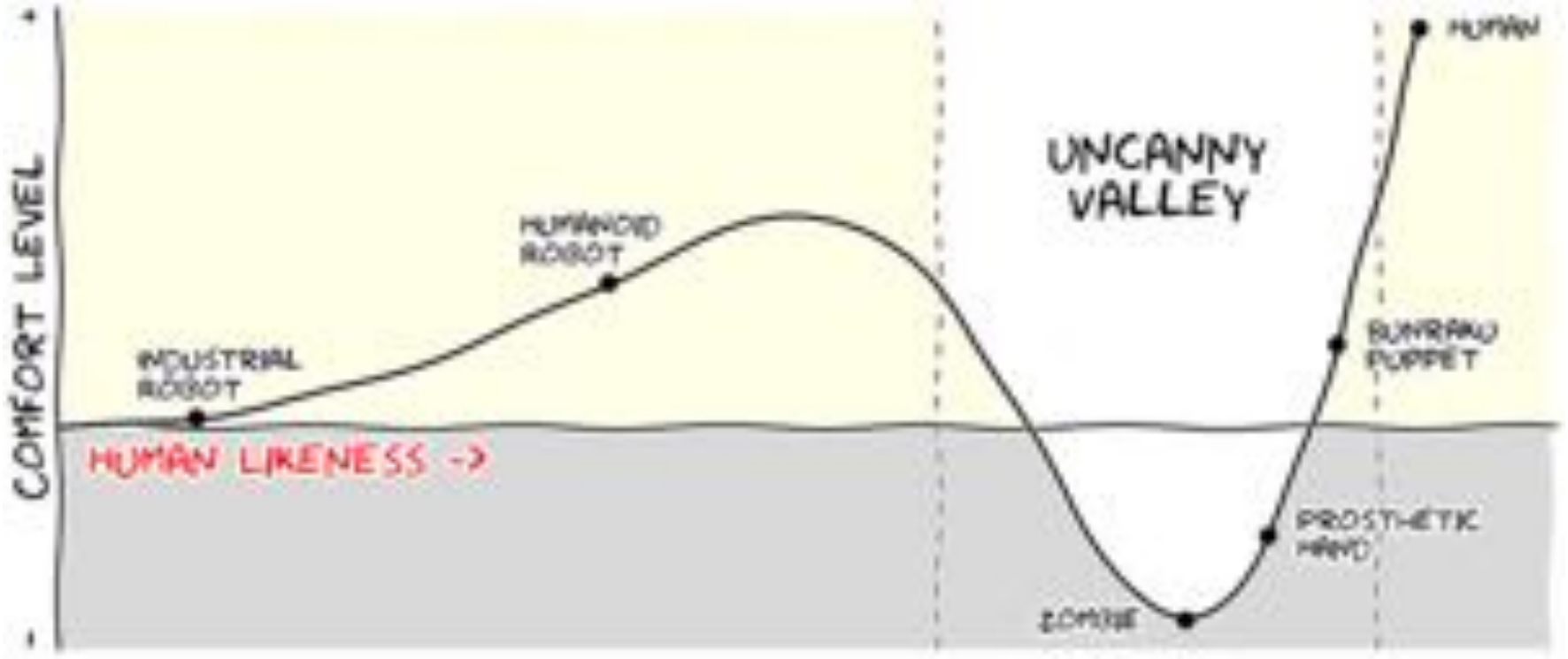
Figure 8. Avoid chartjunk.



Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

PLOS | COMPUTATIONAL BIOLOGY

Figure 9. Message trumps beauty.



Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

PLOS | COMPUTATIONAL BIOLOGY

10: Get the right tool

- Matlab
- Matplotlib
- R
- Inkscape
- Gimp
- GNUPlot
- Latex Pgfplots
- Adobe Illustrator (great but expensive and ldarning curve)

Next…

Analysis of Variation
Tukey's Honestly Significantly Difference Test
Cohen's Effect Size Calculation