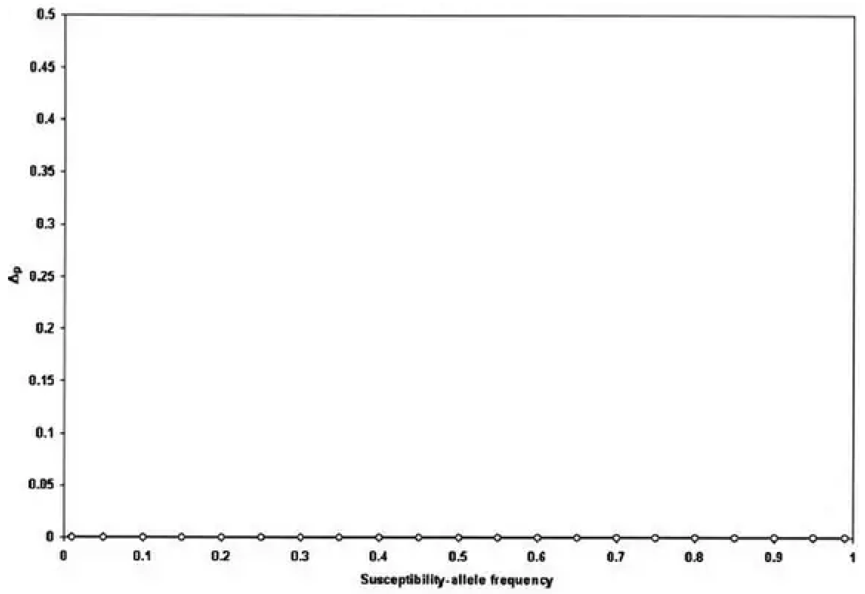
Data Visualization

ENTMLGY 6707 Entomological Techniques and Data Analysis



https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/

Learning objectives

1. Compare and contrast displays of quantitative information
2. Evaluate pitfalls and alternatives of commonly used displays

Figures

How to Display Data Badly

HOWARD WAINER*

The American Statistician, May 1984, Vol. 38, No. 2

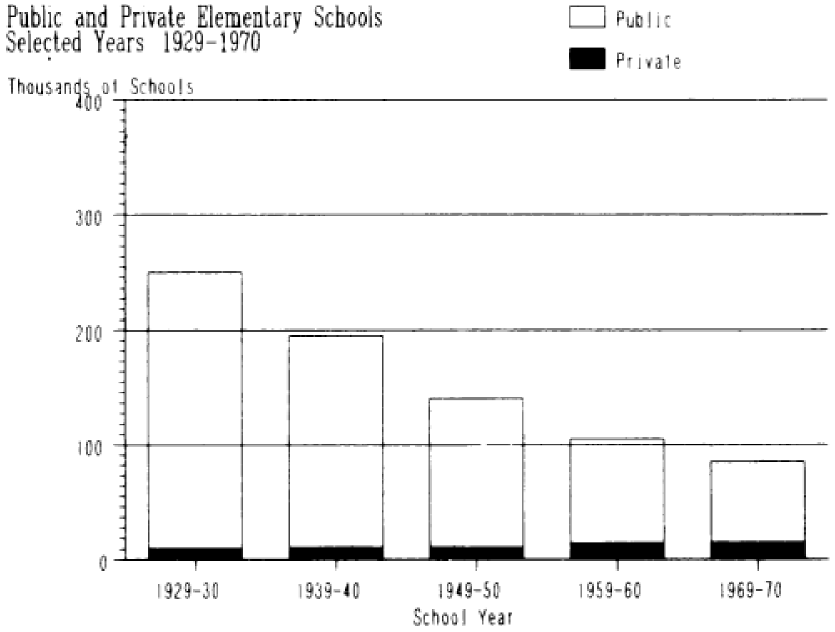


Figure 4. Hiding the data in the scale (from SI3).

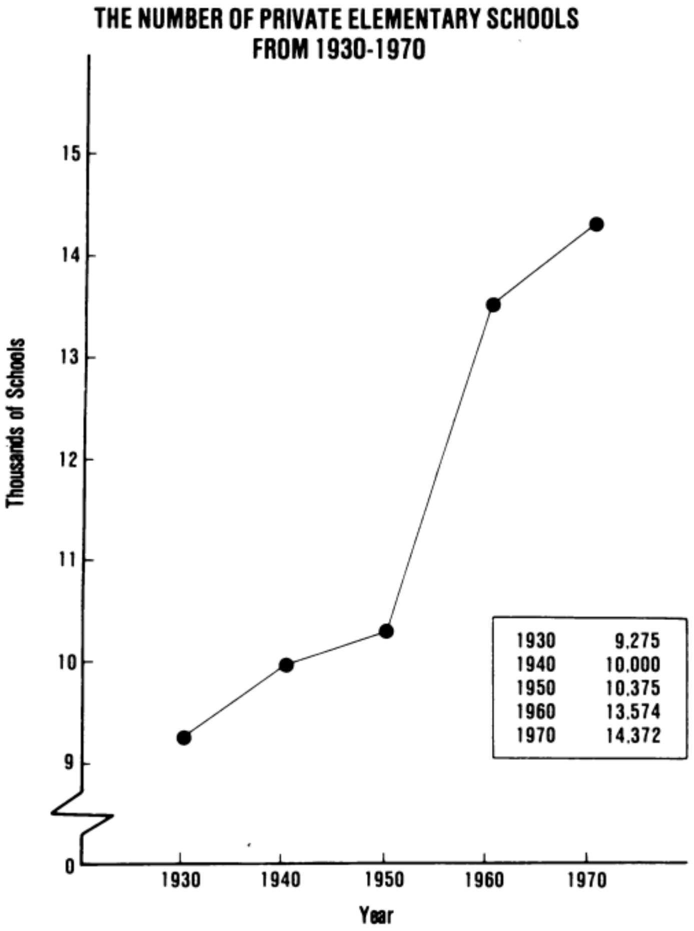


Figure 5. Expanding the scale and showing the data in Figure 4 (from SI3).

Wainer (1984) *The American Statistician* (38) 137-147

Figures

How to Display Data Badly

HOWARD WAINER*

The American Statistician, May 1984, Vol. 38, No. 2

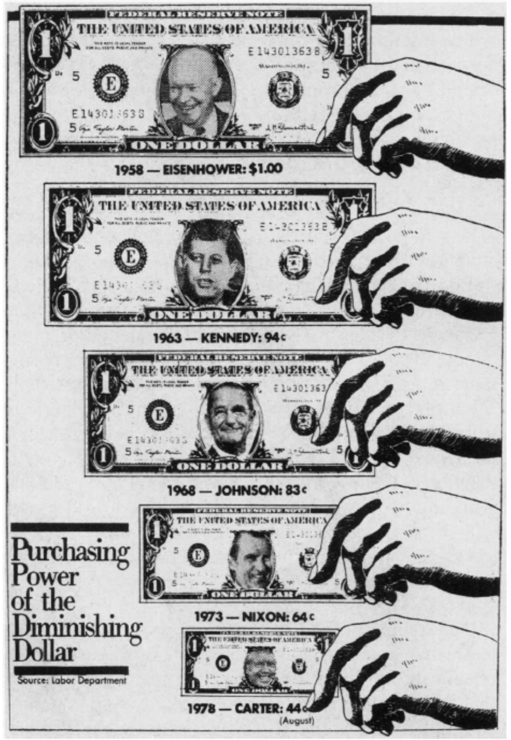


Figure 9. An example of how to goose up the effect by squaring the eyeball (© 1978, The Washington Post).

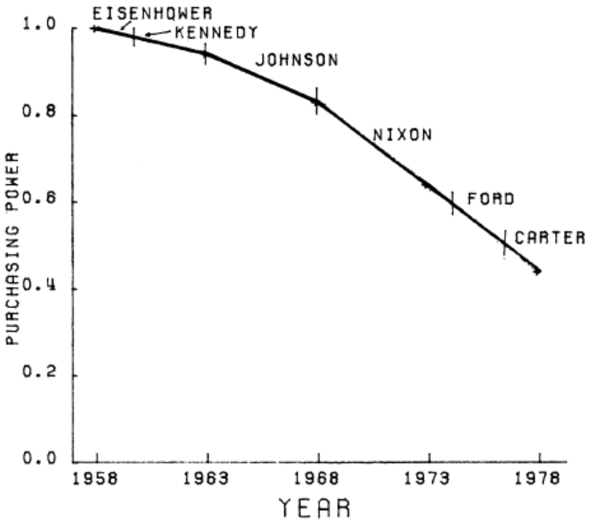


Figure 10. The data in Figure 9 as an unadorned line chart (from Wainer, 1980).

Tables and (…versus) Figures

Let's Practice What We Preach: Turning Tables into Graphs

Andrew GELMAN, Cristian PASARICA, and Rahul DODHIA

Raters' characterization of sentences (absolute score range)	*Percent*
Negative (1–1.9)	23.9
Neutral (2–2.9)	52.2
Positive ($\geq$ 3)	23.9
Total	100.0

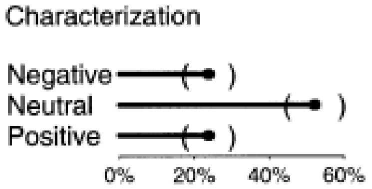


Figure 1. Top panel: Table from Ellenberg (2000) shows the relative frequencies of three categories in a set of 46 ratings of sentences. Bottom panel: Graphical display allows direct comparison without distractions of irrelevant decimal places. Parentheses show ± 1 standard error bounds based on the implicit binomial distribution with n = 46.

Profession	*Frequency of recent citations*	*1996 total employed (1,000)*	*Relative frequency*
Lawyers	8101	880	9.2
Economists	1201	148	8.1
Architects	1097	160	6.9
Physicians	3989	667	6.0
Statisticians	34	14	2.4
Psychologists	479	245	2.0
Dentists	165	137	1.2
Teachers (not university)	3938	4724	0.8
Engineers	934	1960	0.5
Accountants	628	1538	0.4
Computer programmers	91	561	0.2
Total	20,657	11,034	1.9

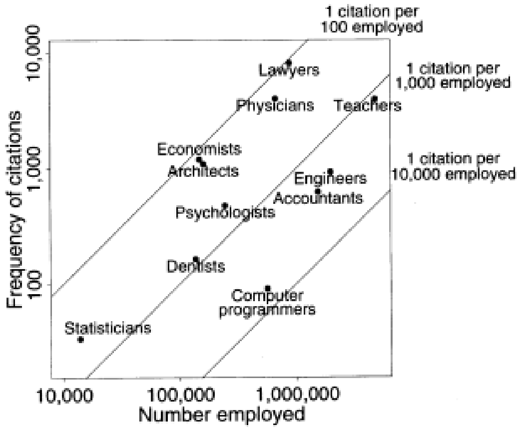


Figure 2. Top panel: Table from Ellenberg (2000) displays counts and rates of citations of various professions from the New York Times *database. Bottom panel: Display as a figure shows the relative positions of the different professions much more clearly. The log-log display allows comparison across several orders of magnitude. The x and y axes are on the same scale, so that any 45° line indicates a constant relative frequency.*

Tables

Table 1
Effect of sex, food, and silk protection on the mean (±SE) supercooling point (SCP) of *H. hebetor*

Comparison (stage)	n[a]	SCP (°C)±SE[b]	SCP range (°C) min, max	Shapiro–Wilk (W-value)[c]	P[d]
Sex category (adult)					
Female	51	−18.3±0.63 a	−26.4, −7.3	0.973	0.2905
Male	51	−17.7±0.63 a	−27.5, −7.8	0.972	0.2803
Food presence[e] *(adult)*					
Fed female	23	−16.5±0.51 a	−20.1, −10.9	0.965	0.5769
Unfed female	23	−24.3±0.50 b	−27.7, −20.6	0.899	0.0241
Silk protection (pupa)					
Naked pupa	17	−25.4±0.55 a	−27.5, −18.7	0.727	0.0002
Silk-encased pupa	17	−25.8±0.58 a	−29.0, −19.6	0.857	0.0136

[a] Pairs measured during the same cooling period using a cooling rate of ca. 1 °C min^{-1}.
[b] Mean SCPs, for each pair, followed by the same lowercase letter, are not significantly different ($P > 0.05$).
[c] Test for normality. $0 < W\text{-value} \leqslant 1$.
[d] Probability that the observed data came from a normal distribution.
[e] Females were fed honey 24 h before SCP measurement.

ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Insect Physiology 51 (2005) 759–768

Journal of Insect Physiology
www.elsevier.com/locate/jinsphys

Cold hardiness of *Habrobracon hebetor* (Say) (Hymenoptera: Braconidae), a parasitoid of pyralid moths

M.A. Carrillo*, G.E. Heimpel, R.D. Moon, C.A. Cannon[1], W.D. Hutchison

Department of Entomology, University of Minnesota, 219 Hodson Hall, 1980 Folwell Avenue, St. Paul, MN 55108, USA

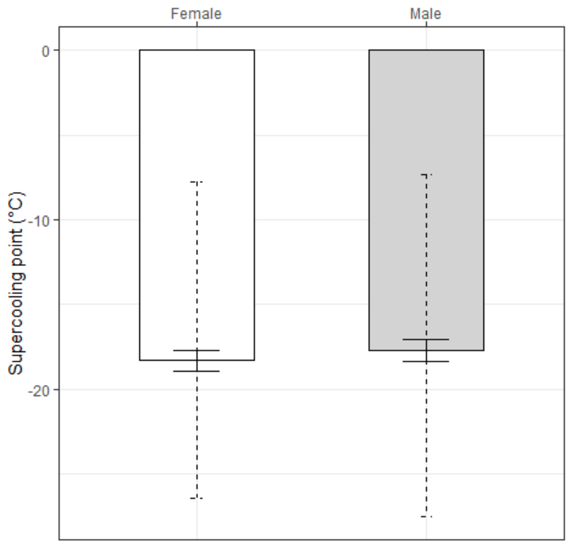
Received 6 December 2004; received in revised form 10 March 2005; accepted 10 March 2005

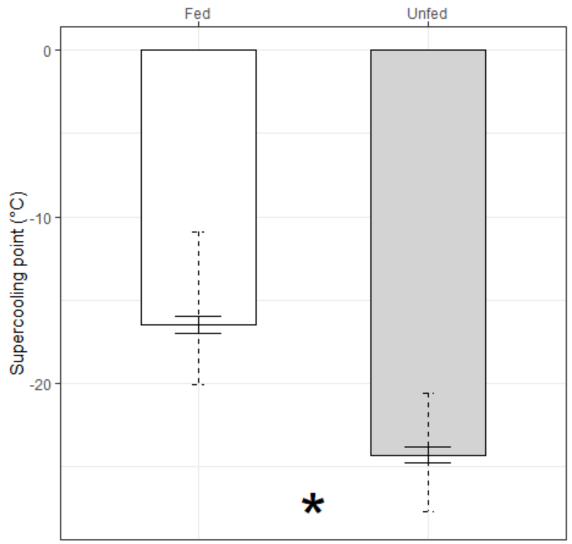
Tables

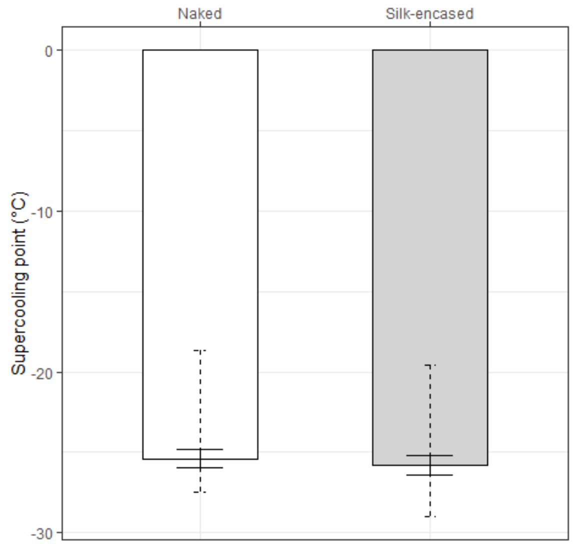
Table 1
Effect of sex, food, and silk protection on the mean (±SE) supercooling point (SCP) of *H. hebetor*

Comparison (stage)	n[a]	SCP (°C)±SE[b]	SCP range (°C) min, max	Shapiro–Wilk (W-value)[c]	P[d]
Sex category (adult)					
Female	51	−18.3±0.63 a	−26.4, −7.3	0.973	0.2905
Male	51	−17.7±0.63 a	−27.5, −7.8	0.972	0.2803
Food presence[e] *(adult)*					
Fed female	23	−16.5±0.51 a	−20.1, −10.9	0.965	0.5769
Unfed female	23	−24.3±0.50 b	−27.7, −20.6	0.899	0.0241
Silk protection (pupa)					
Naked pupa	17	−25.4±0.55 a	−27.5, −18.7	0.727	0.0002
Silk-encased pupa	17	−25.8±0.58 a	−29.0, −19.6	0.857	0.0136

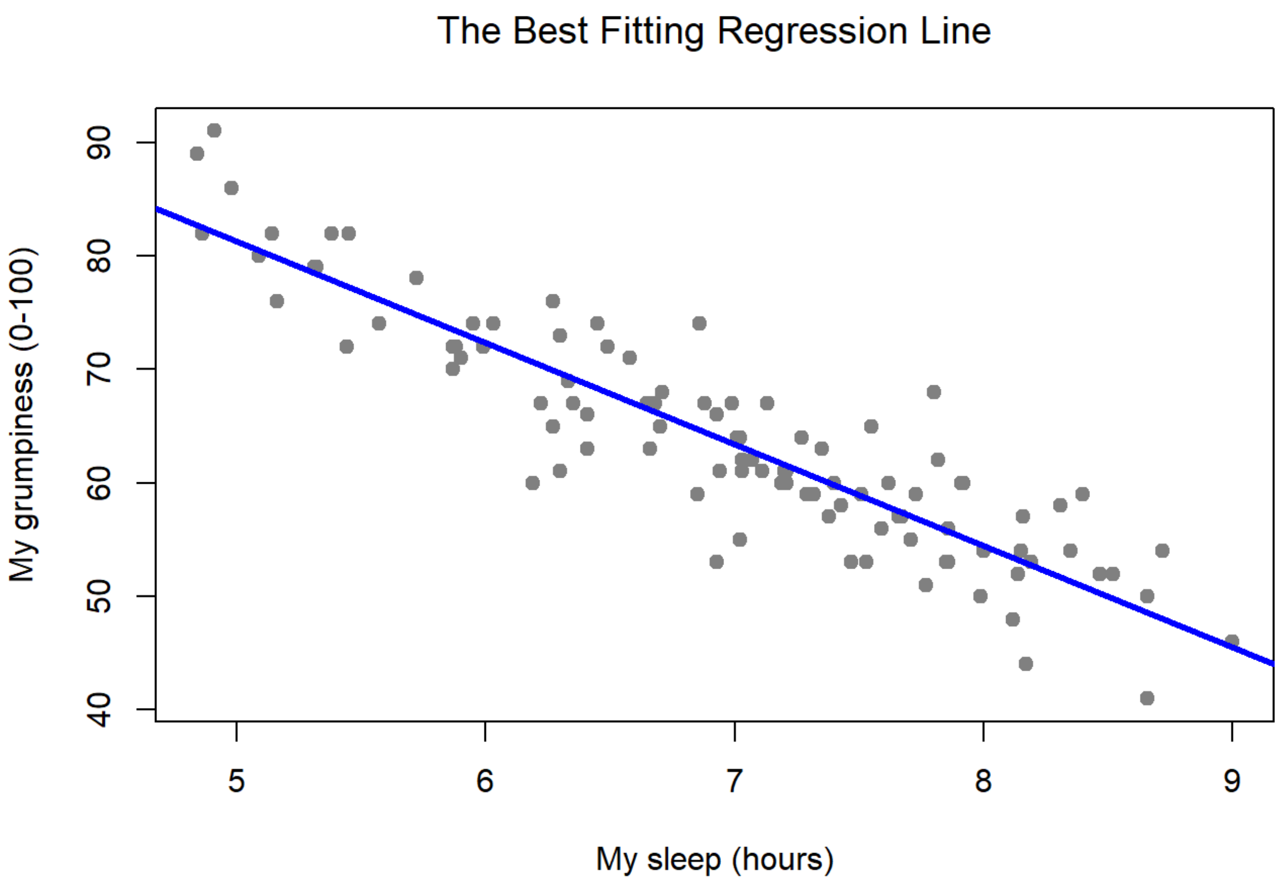
[a]Pairs measured during the same cooling period using a cooling rate of ca. 1 °C min^{-1}.
[b]Mean SCPs, for each pair, followed by the same lowercase letter, are not significantly different ($P>0.05$).
[c]Test for normality. $0<W\text{-value}\leq 1$.
[d]Probability that the observed data came from a normal distribution.
[e]Females were fed honey 24 h before SCP measurement.





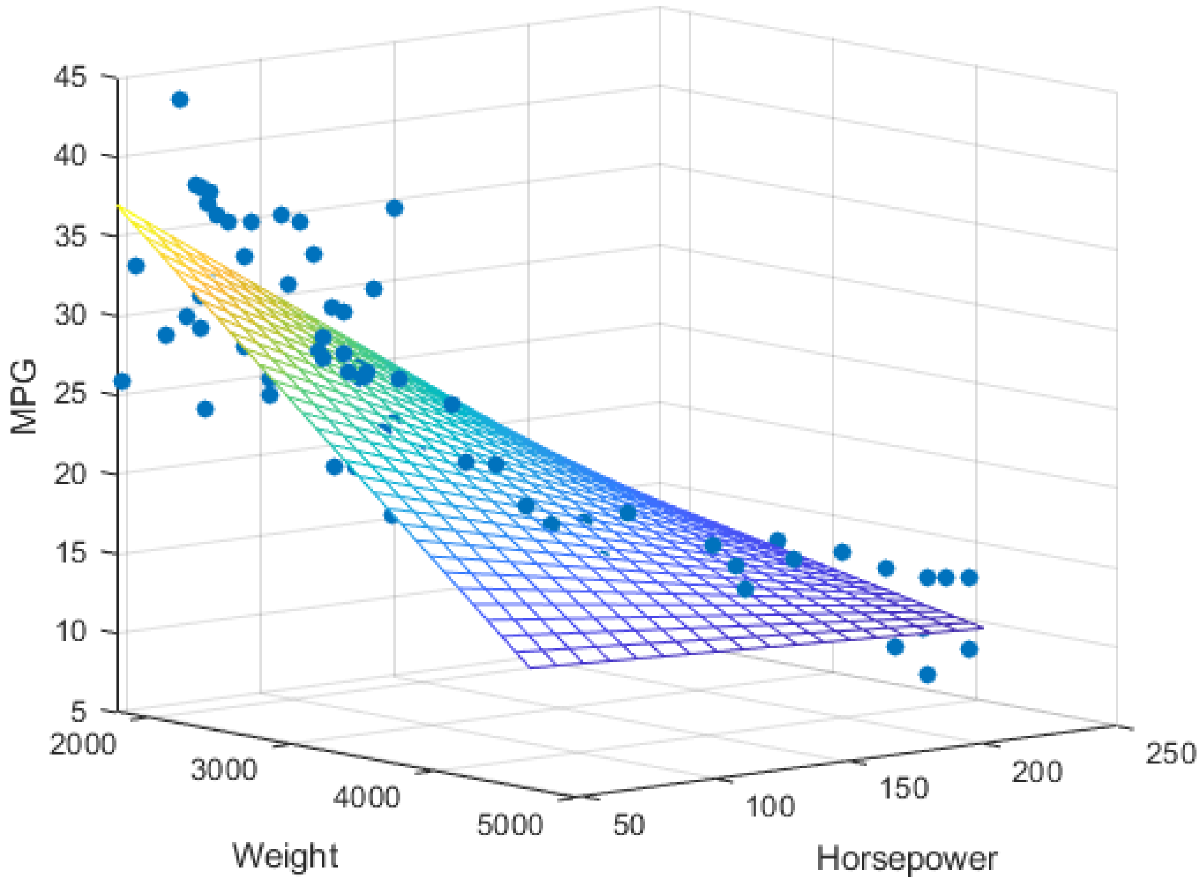


$$Y = \beta_0 + \beta_1 X$$



Source: https://learningstatisticswithr.com/book/regression.html

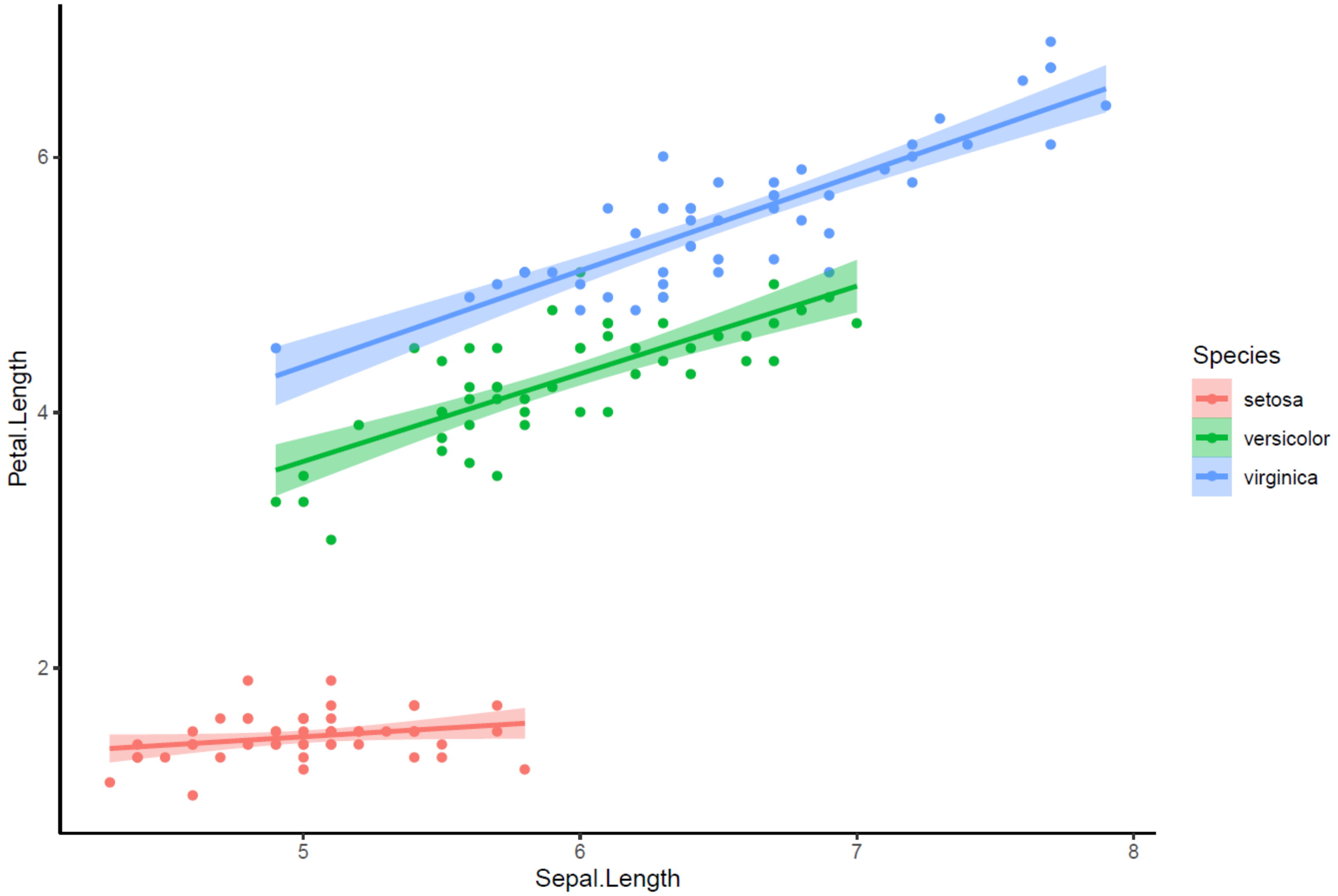
$$Y = \beta_0 + \beta_1 X + \beta_2 Z$$

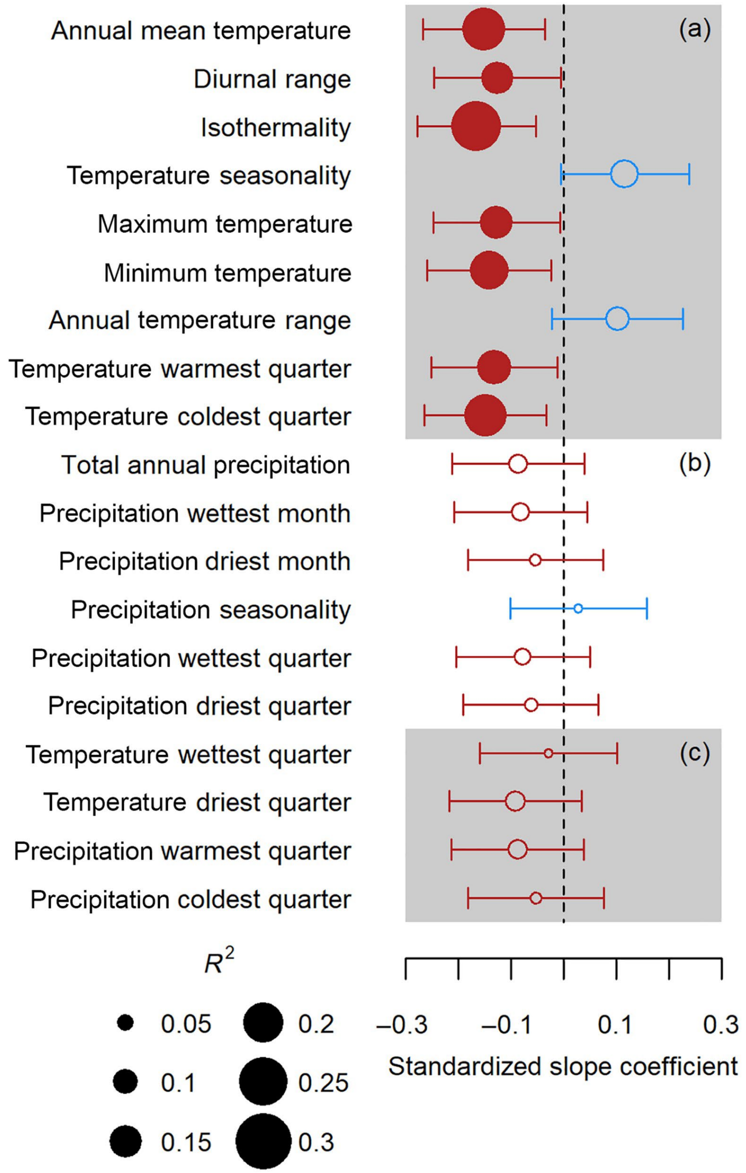


Source: Lekha Priya (medium.com)

$$Y = \beta_0 + \beta_1 X + \beta_2 Z$$

Analysis of covariance (ANCOVA)





Received: 14 January 2022 | Revised: 20 October 2022 | Accepted: 25 October 2022

DOI: 10.1002/ecy.3940

ARTICLE

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Phenological response to climate variation in a northern red oak plantation: Links to survival and productivity

Jonathan A. Knott[1,2] | Liang Liang[3] | Jeffrey S. Dukes[1,4] |
Robert K. Swihart[1] | Songlin Fei[1]

Tables are often necessary

Table 1. Candidate variables for autologistic regressions examining the probability of an outbreaking population of mountain pine beetle in a 12 × 12 km cell on the Chilcotin Plateau of British Columbia, 1972–1986.

Variable type	Variable	Explanation and rationale
Temporal	lag1	Presence/absence of mountain pine beetle in a cell the previous year.
	lag 2	Same, two years previous.
	lag 3	Same, three years previous.
Spatial	lst nbhd	First-order neighborhood (nearest four neighbours).
	2nd nbhd	Second-order neighborhood (nearest eight neighbours).
	infestations	Number of discrete infestations in each cell. This differs from the response variable, the presence/absence of red attack in each cell.
Environmental	tmin	Minimum temperature over calendar year.
	tmax	Maximum temperature over calendar year.
	tmean	Mean temperature over calendar year.
	cold[a]	Number of days cold enough to cause lethal mortality to overwintering larvae, after Wygant (1940). Temperatures less extreme than −37°C can be lethal early and late in the year, and complete mortality occurs when larvae are exposed to temperatures < −37°C for short periods (Wygant 1940, Somme 1964, Stahl et al. 2006b).
	warm[a]	Mean August temperature. Development and emergence of new mountain pine beetle adults is closely governed by temperature. Peak flight occurs in a narrow window in summer (McCambridge 1971, Safranyik 1978, Safranyik and Carroll 2006).
	ddegg[a]	Accumulated degree days above 5.5°C from August to end of growing season.
	dd[a]	Accumulated degree days above 5.5°C from August in previous year to current July.
	Pla[a]	0/1 indicator variable: sufficient heat accumulation to hatch 50% of eggs before winter (306°C degree days).
	Plb[a]	0/1 indicator variable: sufficient heat accumulation to develop and emerge on a univoltine life cycle (833°C degree days).
	P2[a]	0/1 indicator variable if minimum winter temperatures were higher than −40°C.
	elevation	Mean elevation of cell, based on 25 sampled points (regular design) within cell. This may be a useful proxy for host tree distribution, as lodgepole pine do not grow at high elevations over our study area.

[a]Derived from Safranyik et al. (1975).

Figures

PLOS | BIOLOGY

PERSPECTIVE

Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm

Tracey L. Weissgerber[1]*, Natasa M. Milic[1,2], Stacey J. Winham[3], Vesna D. Garovic[1]

1 Division of Nephrology & Hypertension, Mayo Clinic, Rochester, Minnesota, United States of America, **2** Department of Biostatistics, Medical Faculty, University of Belgrade, Belgrade, Serbia, **3** Division of Biomedical Statistic and Informatics, Mayo Clinic, Rochester, Minnesota, United States of America

* weissgerber.tracey@mayo.edu

CrossMark
click for updates

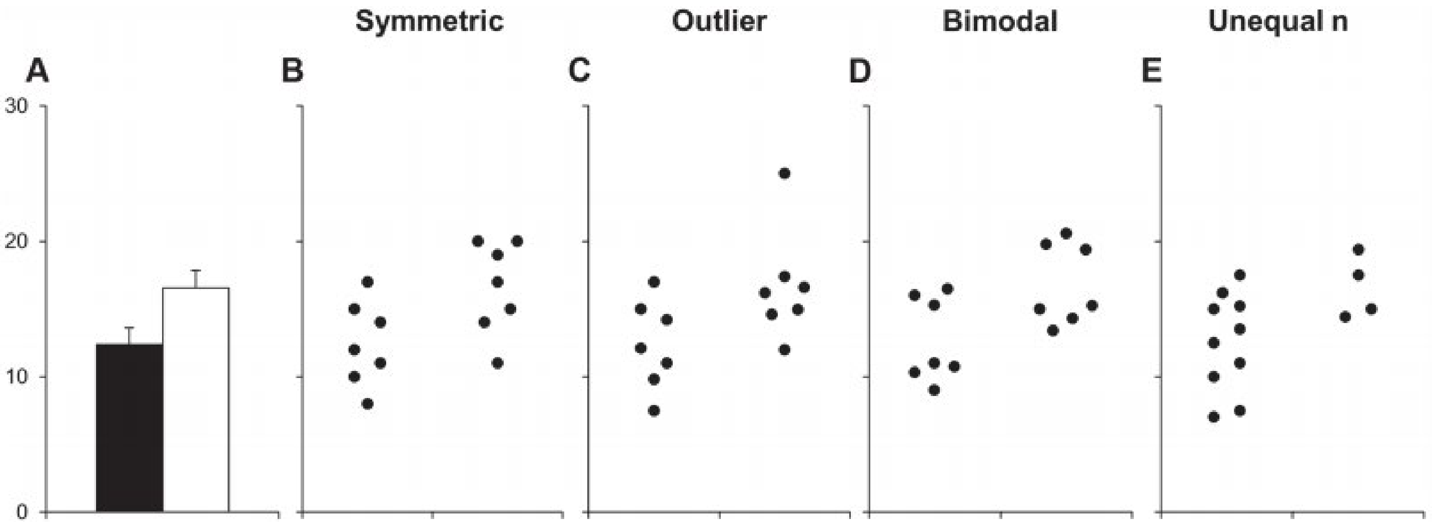
OPEN ACCESS

Citation: Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol 13(4): e1002128. doi:10.1371/journal.pbio.1002128

Published: April 22, 2015

Abstract

Figures in scientific publications are critically important because they often show the data supporting key findings. Our systematic review of research articles published in top physiology journals ($n = 703$) suggests that, as scientists, we urgently need to change our practices for presenting continuous data in small sample size studies. Papers rarely included scatterplots, box plots, and histograms that allow readers to critically evaluate continuous data. Most papers presented continuous data in bar and line graphs. This is problematic, as many different data distributions can lead to the same bar or line graph. The full data may suggest different conclusions from the summary statistics. We recommend training investigators in data presentation, encouraging a more complete presentation of data, and changing journal editorial policies. Investigators can quickly make univariate scatterplots for small sample size studies using our Excel templates.


Symmetric
Outlier
Bimodal
Unequal n
A
B
C
D
E
30
20
10
0

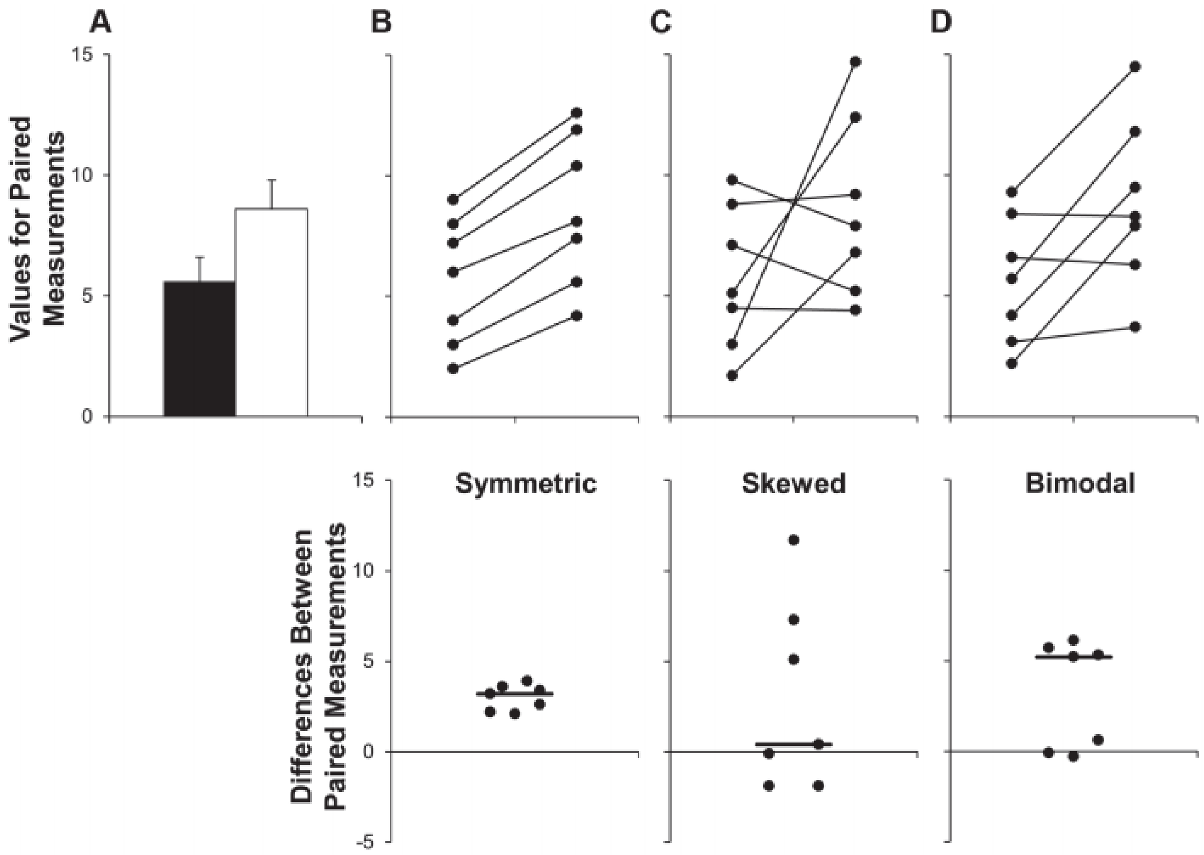
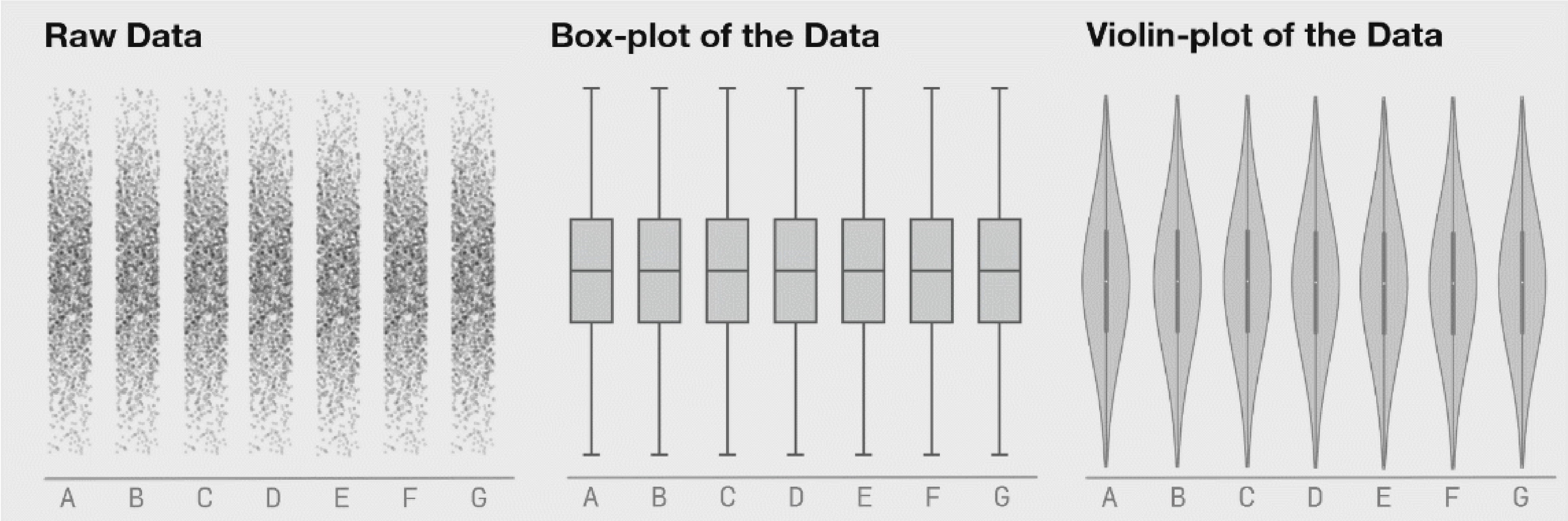
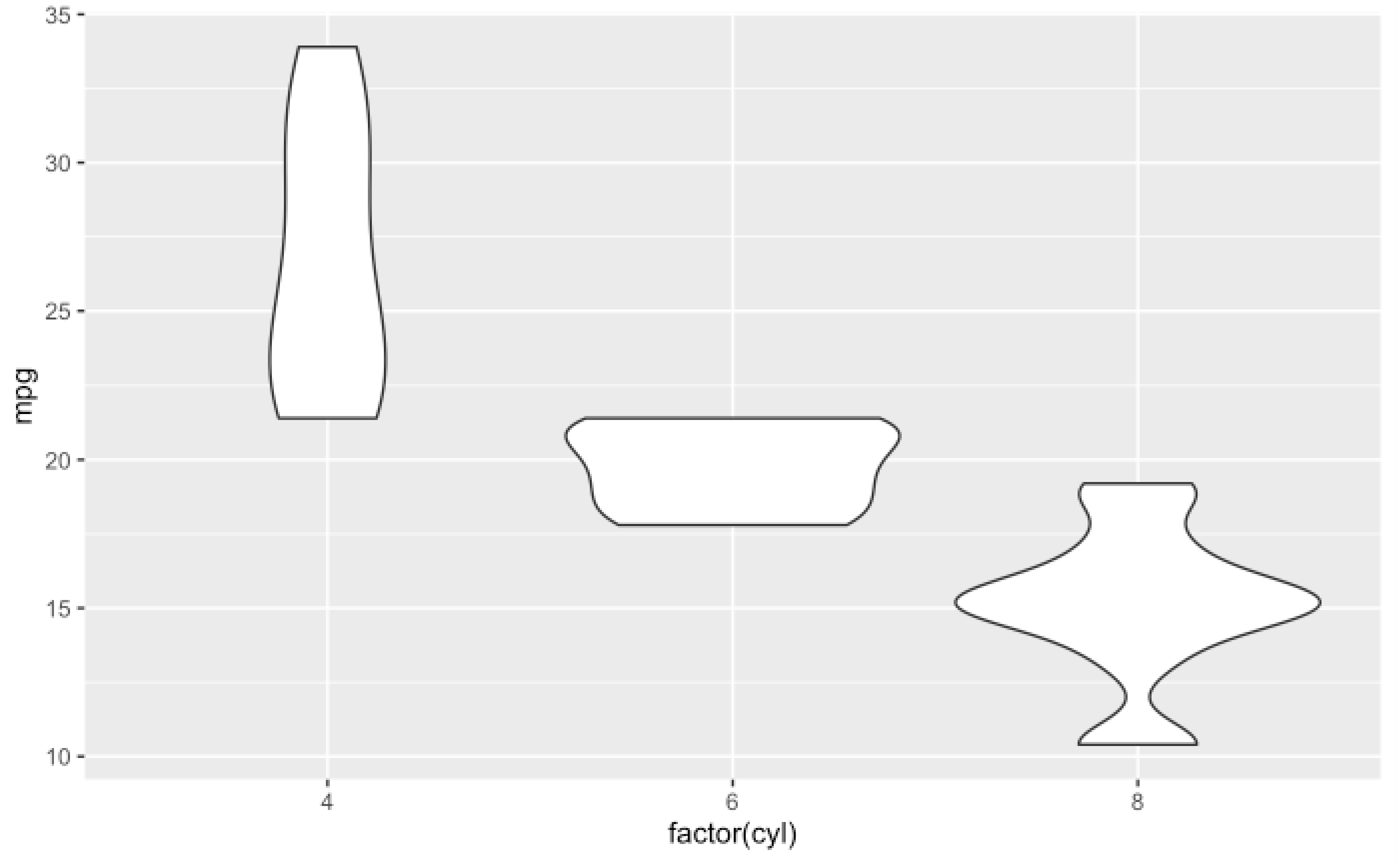


Fig 2. Additional problems with using bar graphs to show paired data. The bar graph (mean ± SE) suggests that the groups are independent and provides no information about whether changes are consistent across individuals (Panel A). The scatterplots shown in the Panels B–D clearly demonstrate that the data are paired. Each scatterplot reveals very different patterns of change, even though the means and SEs differ by less than 0.3 units. The lower scatterplots showing the differences between measurements allow readers to quickly assess the direction, magnitude, and distribution of the changes. The solid lines show the median difference. In Panel B, values for every subject are higher in the second condition. In Panel C, there are no consistent differences between the two conditions. Panel D suggests that there may be distinct subgroups of “responders” and “nonresponders.”

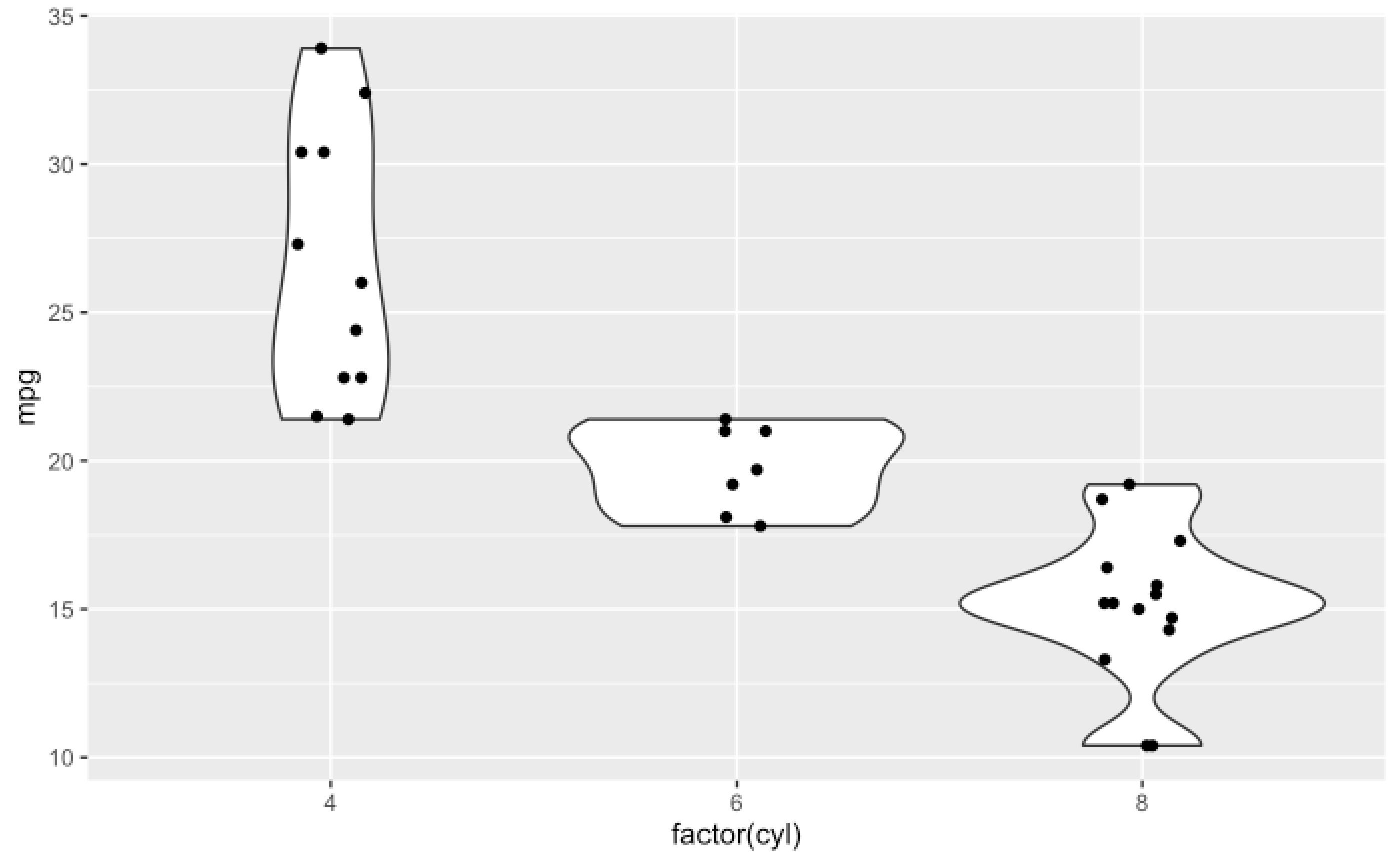


Source: autodeskresearch.com

```r
p <- ggplot(mtcars, aes(factor(cyl), mpg))
p + geom_violin()
```



```
p + geom_violin() + geom_jitter(height = 0, width = 0.1)
```



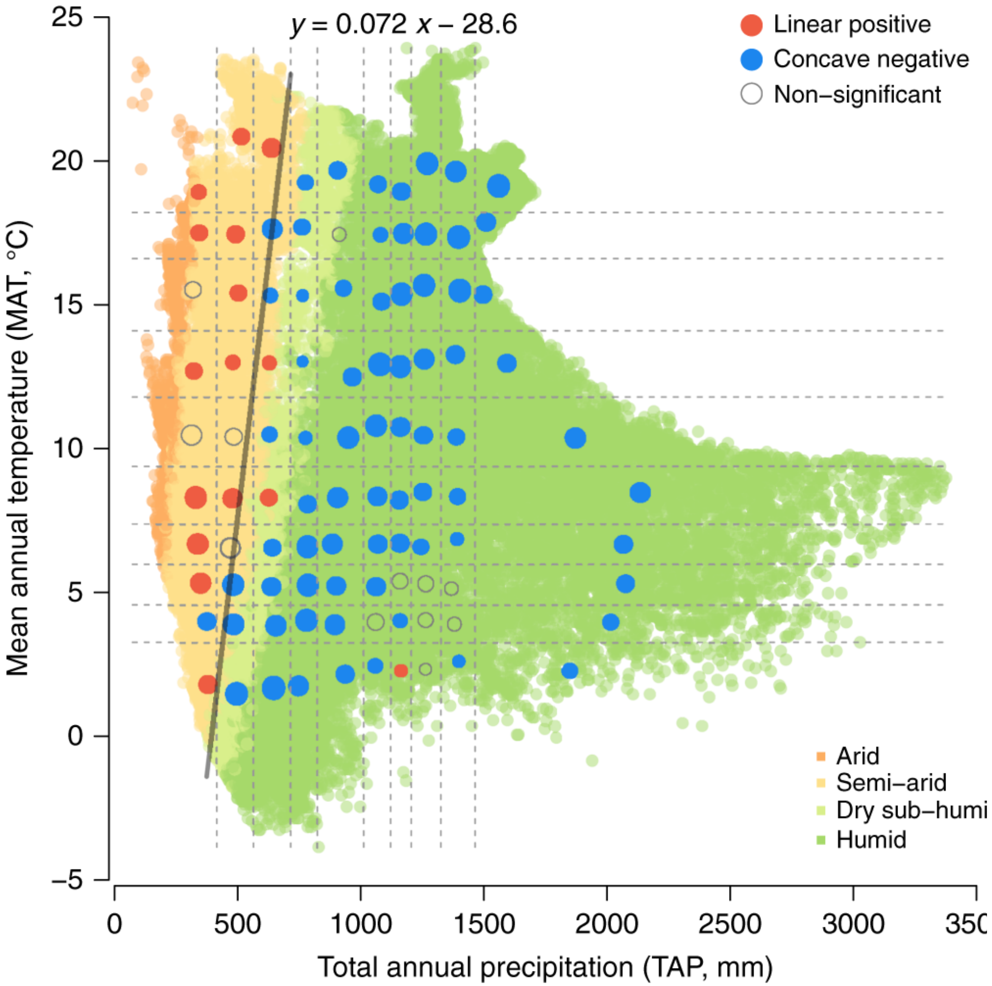
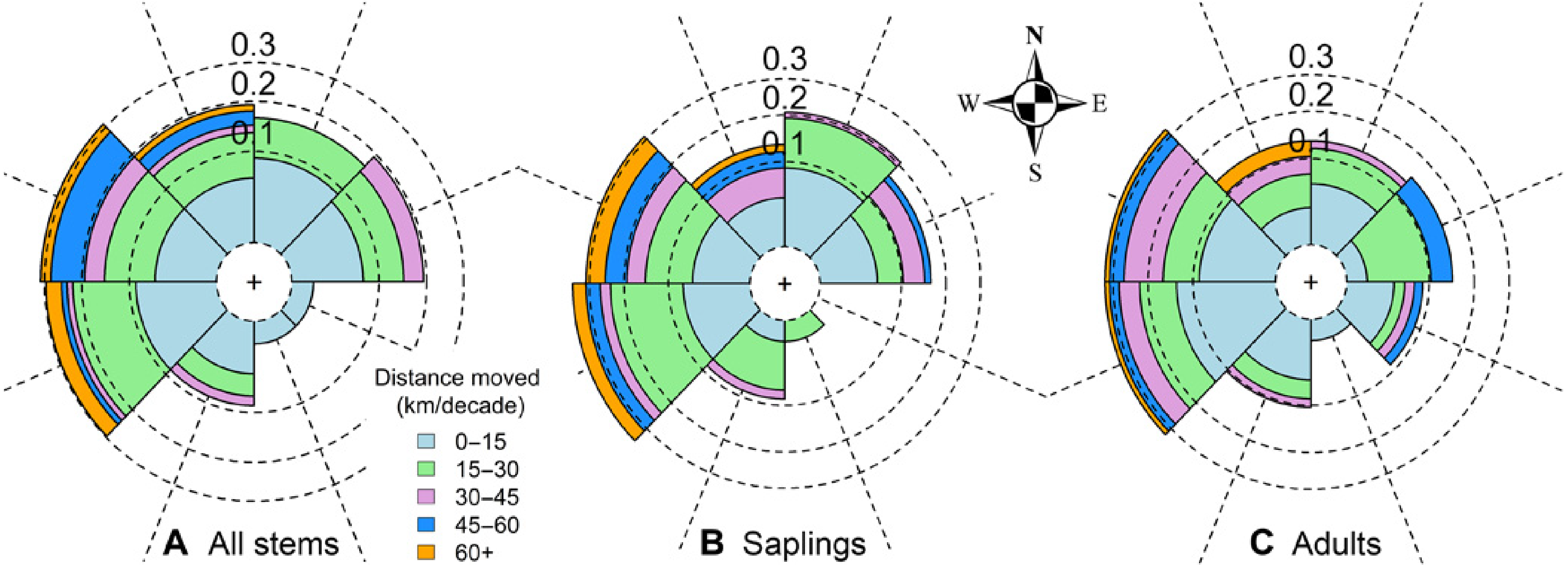


Fig. 1 Relationship between tree richness and productivity in different climatic units. The color of circles within each climatic unit indicates the form of the relationship determined by a generalized linear model between biodiversity and productivity, and circle radius is proportional to the number of forest plots (log-transformed) in each climatic unit. Forest plots were divided into 10 × 10 climatic units according to their MAT and TAP quantile classes based on WorldClim[25]; and are colored according to their aridity index (0.03-0.2, arid; 0.2-0.5, semi-arid; 0.5-0.65, dry sub-humid; > 0.65, humid) based on the Global Aridity Index[26]. The line between the red and blue points is the division between the linear-positive and concave-negative relationships based on a logistic regression for the two groups as a function of MAT and TAP

ARTICLE
https://doi.org/10.1038/s41467-018-07880-w OPEN

Impacts of climate on the biodiversity-productivity relationship in natural forests

Songlin Fei[1], Insu Jo[1], Qinfeng Guo[2], David A. Wardle[3,4], Jingyun Fang[5], Anping Chen[1], Christopher M. Oswalt[6] & Eckehard G. Brockerhoff[7,8]



RESEARCH ARTICLE | CLIMATE CHANGE

Divergence of species responses to climate change

Songlin Fei[1,2,*], Johanna M. Desprez[1], Kevin M. Potter[3], Insu Jo[1], Jonathan A. Knott[1] and Christopher M. Oswalt[4,5]
+ See all authors and affiliations

Science Advances 17 May 2017:
Vol. 3, no. 5, e1603055
DOI: 10.1126/sciadv.1603055