Biostatistics: why and how we should teach better

Natasa Milic, MD, PhD
Professor of Medical Statistics
School of Medicine University of Belgrade
Department for Internal Medicine
Mayo Clinic, Rochester, USA

Rejection of rejection: a novel approach to overcoming barriers to publication

Tired of rejections? **Cath Chapman** and **Tim Slade** offer a simple way to fight back

Cath Chapman *senior research fellow*, Tim Slade *associate professor*

NHMRC Centre of Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia

✓All academics aim to publish in high impact journals.

✓However, many leading medical journals reject more than 80% of the manuscripts they receive,

✓making rejection the biggest barrier to publication in high quality journals.

✓Recently, two authors proposed a novel solution to this problem: the rejection of rejection letter – which aims to overcome the leading barrier to publication—manuscript rejection.

Rejection of rejection letter

[*insert university emblem here*]

Dear Professor [*insert name of editor*]

[*Re: MS 2015_XXXX Insert title of ground-breaking study here*]

Thank you for your rejection of the above manuscript.

Unfortunately we are not able to accept it at this time. As you are probably aware we receive many rejections each year and are simply not able to accept them all. In fact, with increasing pressure on citation rates and fiercely competitive funding structures we typically accept fewer than 30% of the rejections we receive. Please don't take this as a reflection of your work. The standard of some of the rejections we receive is very high.

In terms of the specific factors influencing our decision the failure by Assessor 1 to realise the brilliance of the study was certainly one of them. Simply stating "this study is neither novel nor interesting and does not extend knowledge in this area" is not reason enough. This, coupled with the use of Latin quotes by Assessor 2, rendered an acceptance of your rejection extremely unlikely.

We do wish you and your editorial team every success with your rejections in the future and hope they find safe harbour elsewhere. To this end, may we suggest you send one to [*insert name of rival research group*] for consideration. They accept rejections from some very influential journals.

Please understand that our decision regarding your rejection is final. We have uploaded the final manuscript in its original form, along with the signed copyright transfer form.

We look forward to receiving the proofs and to working with you in the future.

Yours sincerely

Dr [*insert name here*]

[*Insert research group acronym here*]

[*Insert university here*]

[*Insert country here—that is, Australia/New Zealand/small European Country/Canada*]

MAYO CLINIC

Misuse of statistical methods

- is common in biomedical science research, even among papers published in high impact journals. This includes:

- using incorrect or suboptimal tests,

- summarizing data that were analyzed by nonparametric techniques as mean and standard deviation or standard error,

- reporting *p*-values that are inconsistent with the test statistic,

- p-hacking,

- and analyzing nonindependent data as though they are independent.

Additional problems

- arise from inadequate reporting of statistical methods. This may include

- failing to provide a power calculation,

- not reporting which statistical test was used, or not providing adequate detail about the test,

- not addressing whether the assumptions of the statistical tests were examined,

- or not specifying how replicates were treated in the analysis.

The reliance on null hypothesis testing and p-values has been heavily questioned, and a variety of alternate approaches have been proposed.

Increased quantification of scientific research

- have expanded the scope of applications of statistical methods

- This has created new avenues for scientific progress, but it also brings concerns about conclusions drawn from research data.

- Underpinning many published scientific conclusions is the concept of "statistical significance," typically assessed with an index called the p-value.

- While the p-value can be a useful statistical measure, it is commonly misused and misinterpreted.

In February, 2014, George Cobb, Professor Emeritus of Mathematics and Statistics at Mount Holyoke College, posed these questions to an ASA discussion forum:

- Q: Why do so many colleges and grad schools teach p = 0.05?

- A: Because that's still what the scientific community and journal editors use.

- Q: Why do so many people still use p = 0.05?

- A: Because that's what they were taught in college or grad school.

Cobb's concern

- Cobb's concern was a long-worrisome circularity in the sociology of science based on the use of bright lines such as $P < 0.05$:

- "We teach it because it's what we do; we do it because it's what we teach."

Highly visible discussions happened over the last few years

- ScienceNews (Siegfried, 2010) wrote: "It's science's dirtiest secret: The 'scientific method' of testing hypotheses by statistical analysis stands on a flimsy foundation."

- Phys.org Science News Wire (2013) cited: "numerous deep flaws" in null hypothesis significance testing

- ScienceNews (Siegfried, 2014) said: "statistical techniques for testing hypotheses…have more flaws than Facebook's privacy policies."

- "Simply Statistics" (Leek, 2014) responded: "The problem is not that people use P-values poorly, it is that the vast majority of data analysis is not performed by people properly trained to perform data analysis"

- Regina Nuzzo published an article in *Nature* entitled "Scientific method: statistical errors" (Nuzzo, 2014). That article is now one of the most highly viewed *Nature articles, as reported by altmetric.com*

nature
International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Au

Archive > Volume 520 > Issue 7549 > Research Highlights: Social Selection > Article

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

Bar graphs criticized for misrepresenting data

Researchers on Twitter denounce a common way to display data.

Dodgy data analysis has come under fire in 2015, and an article in *PLoS Biology* takes aim at another staple of statistics. It argues that bar graphs used to describe a continuum of data are often uninformative and misleading, and should be purged from much of the scientific literature.

MAYO CLINIC

Data presentation is the foundation of our collective scientific knowledge…
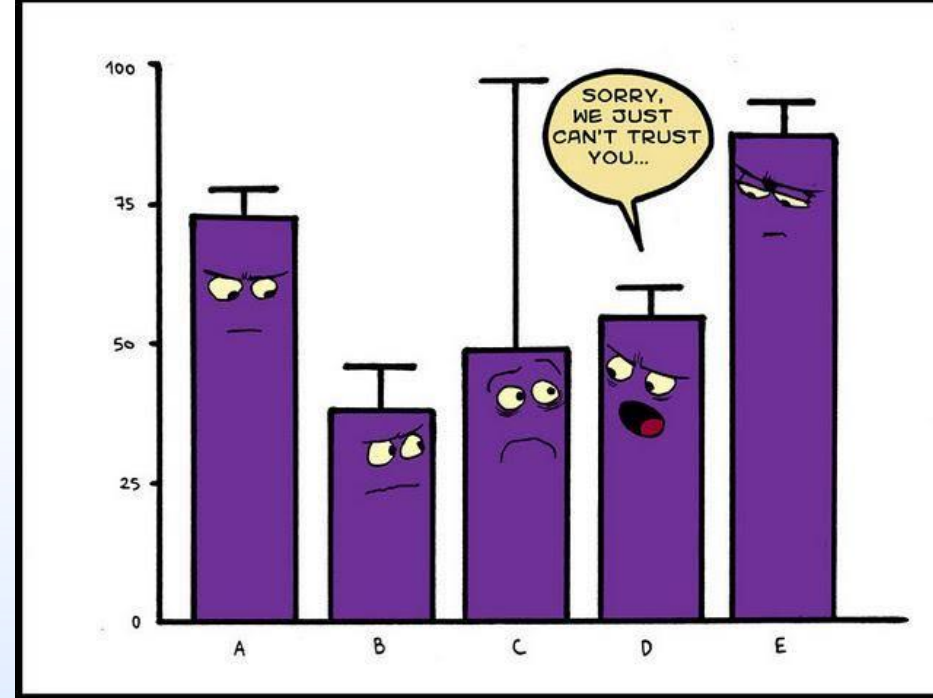
- We've all been there: the papers we read, presentations we attend, posters we see, they all use bar plots, and some of our close collegues do it too!
- But just because a practice is standard doesn't mean it should continue.
- While there exist a wide array of ways to display data, many people choose to use bar plots, a simple graph showing a group mean and standard error (or deviation).
- Unfortunately, most data aren't as clean as bar plots make them seem, and since bar plots reveal very little about the distribution of the data, this kind of visualization can be misleading.
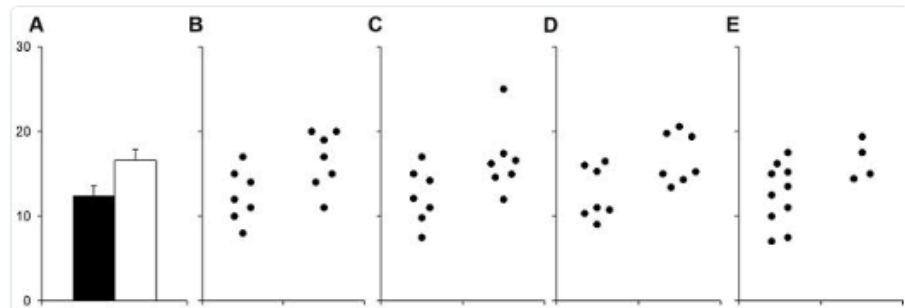
Researchers at the Mayo Clinic in Rochester, Minnesota

❖ looked at more than 600 research articles

❖ published in top physiology journals in early 2014

❖ and found that bar graphs were used to describe continuous   data

❖ in more than 85% of the articles

❖ They showed that very different data sets can be described by the same bar graph

❖ "A visually appealing figure is of little value if it is not suitable for the type of data being presented," they write

❖ As an alternative, they recommend dotplots showing every data point, especially for studies with small sample sizes
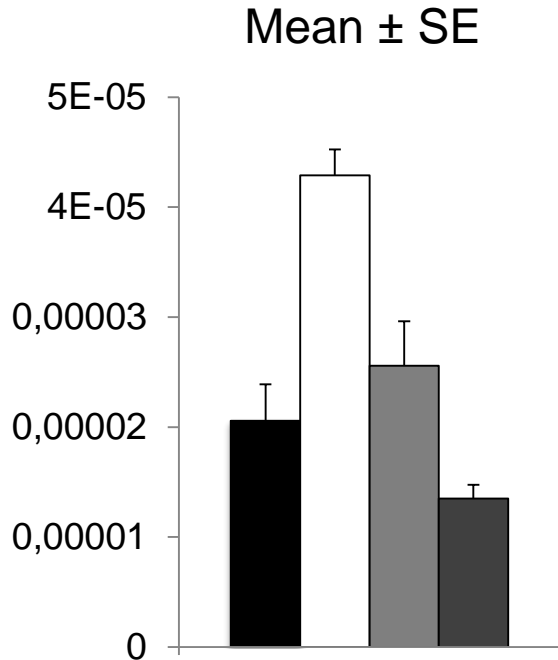
MAYO CLINIC

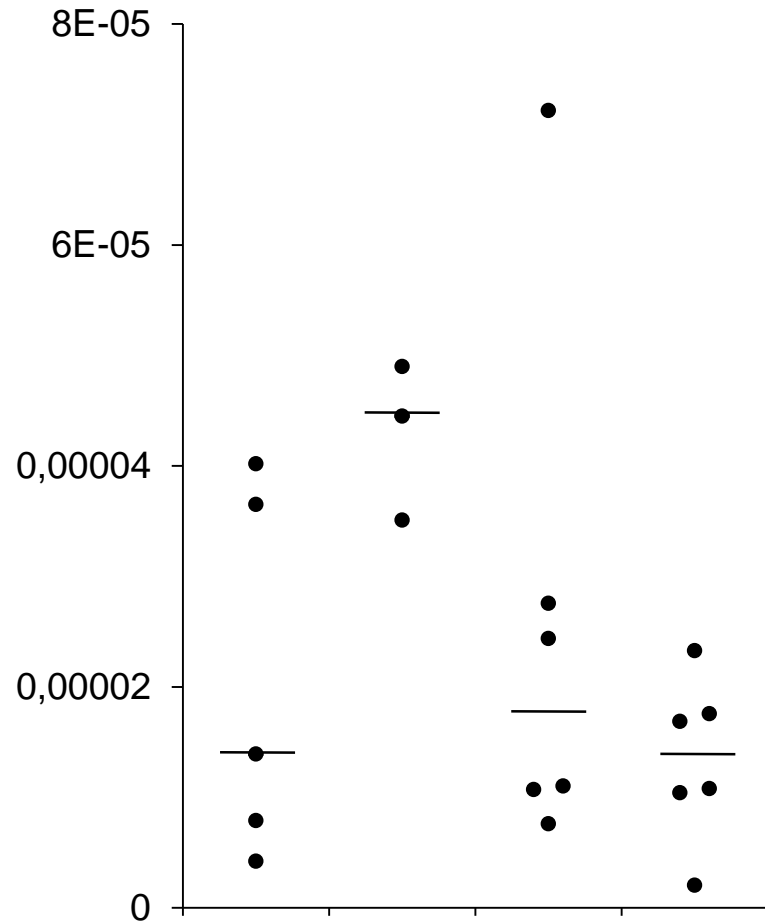# Many different data distributions can lead to the same bar graph…
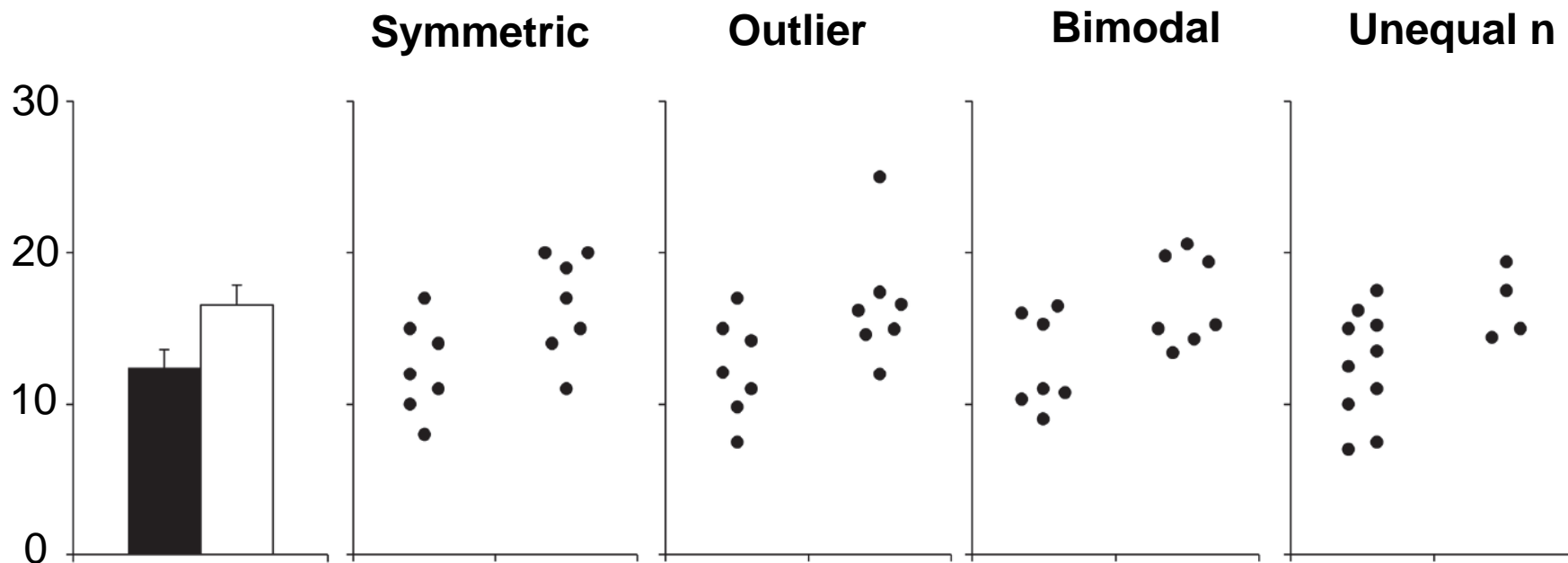
Test	p value			
	Symmetric	Outlier	Bimodal	Unequal n
T-test: equal var.	0.035	0.074	0.033	0.051
T-test: Unequal var.	0.035	0.076	0.033	0.035
Wilcoxon	0.056	0.10	0.173	0.067

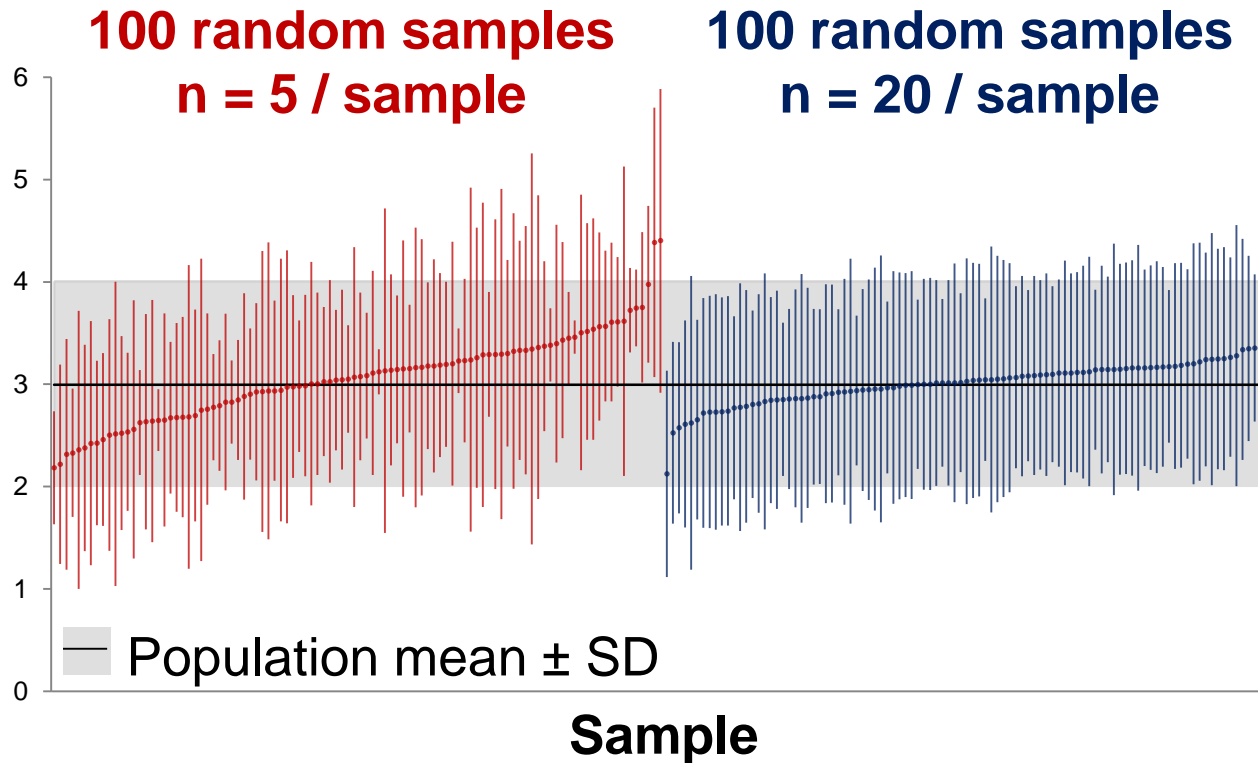# Our interpretation depends on what we see

Reader is a passive observer

Reader is an active participant

Summary statistics are only meaningful when there are enough data to summarize

100 random samples n = 5 / sample

100 random samples n = 20 / sample

Population mean ± SD

Sample

With small sample sizes, means and SDs can be very inaccurate

Tom Oates @toates_19 · Apr 22

Everyone in the world with access to numbers and plot making software should have to read this paper

Hofmann Lab @HofmannLabU_TX · May 2

Finally! Paper in @PLOSBiology makes it plainly obvious why bar graphs hide the data and (often) mislead the reader bit.ly/1KiLOrV

Tim Doherty @TimDoherty_ · May 12

Just made a bunch of bar plots for my ICCB poster, but then realised I'll probably get shunned after the @PLOSBiology paper

P. Sean McDonald
@pseanmc

Weissgerber et al., which could be retitled, "Beyond bar graphs: why Sean's graphs suck". ow.ly/McvM6

Taha Yasseri @TahaYasseri · Apr 26

Wondering why I'm going to reject your next paper? Reason explained here bit.ly/1K9vIGo awesome @PLOSBiology

Must try harder

- Low number of cancer-research studies have been converted into clinical success

- Major factor is the overall poor quality of published preclinical data

- A warning sign should be the "shocking" number of research papers in the field for which the main findings could not be reproduced

- The finding resonates with a growing sense of unease among specialist editors and not just in the field of oncology

- Across the life sciences, handling corrections that have arisen from avoidable errors in manuscripts has become an uncomfortable part of the publishing process

Effects

- ## Journal policy changes
  - PLOS Biology
  - Journal of Biological Chemistry
  - Kidney International
  - Journal of Neuroscience Research
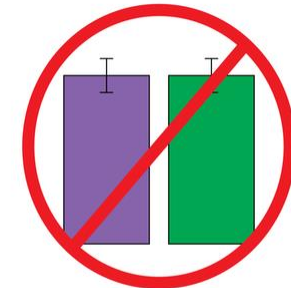
#barbarplots

- ## Editors & reviewers using the paper when requesting improved data visualization

- ## "Bar bar plots" Kickstarter campaign

A group of young scientists interested in improving scientific communication started #barbarplots initiative. Specifically, they have made t-shirts reproducing a widely-shared meme, with the goal of sending them to the editors of the journals.
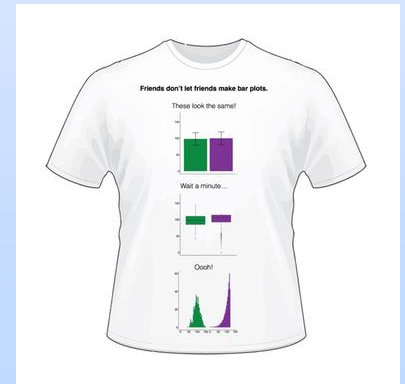
Why this happen?

Statistics are essential,

but training

is not always required

Latest articles from *PLOS Biology*

_____

April 8, 2016
Gauging the Purported Costs of P...
by Simon Robin Evans

_____

April 8, 2016
Reinventing Biostatistics Educatio...
by Natasa M. Milic

PLOS | BIOLOGY

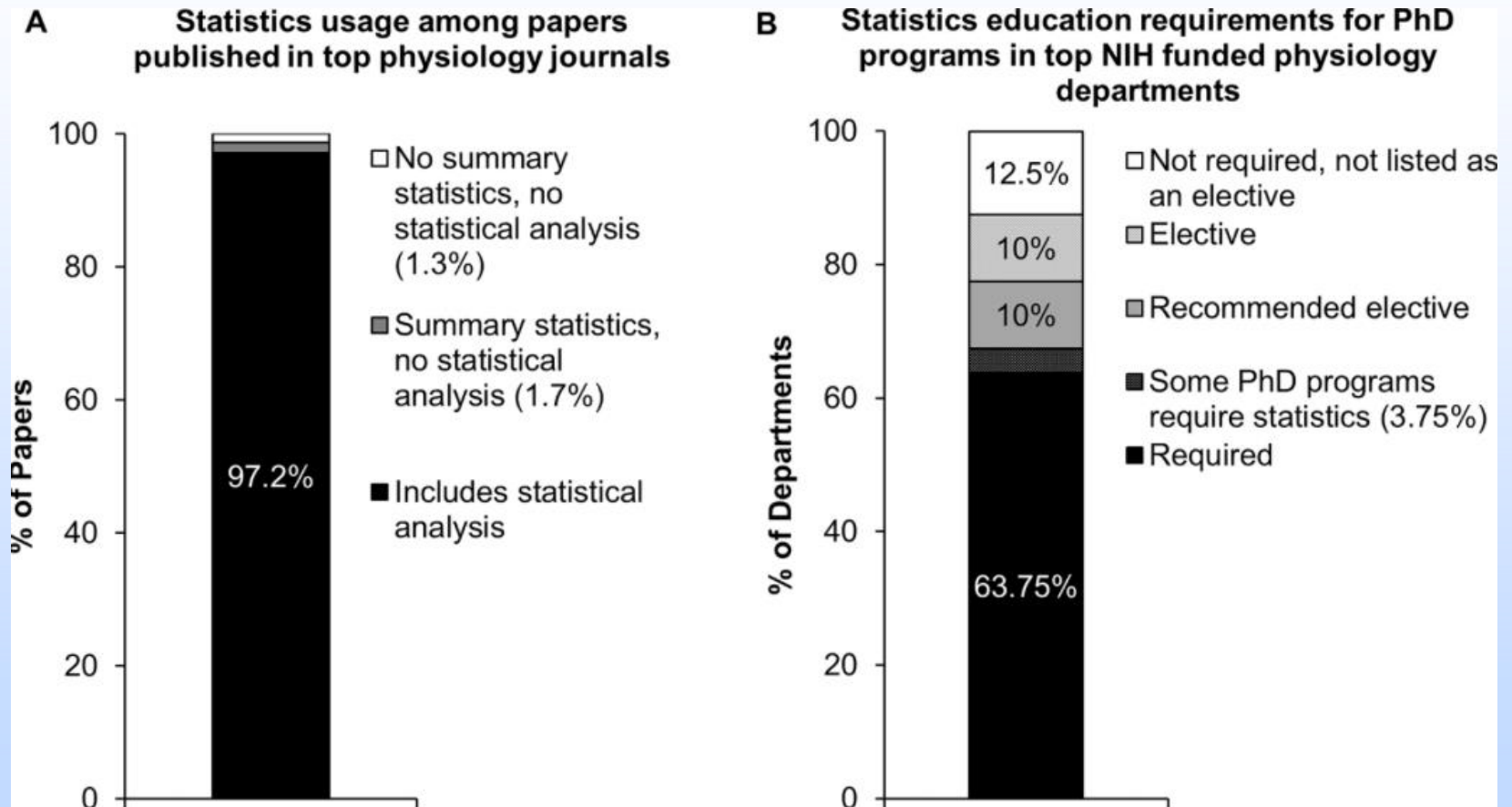| 164 Save | 8 Citation |
| 17,428 View | 550 Share |

PERSPECTIVE

Reinventing Biostatistics Education for Basic Scientists

Tracey L. Weissgerber[1]*, Vesna D. Garovic[1], Jelena S. Milin-Lazovic[2], Stacey J. Winham[3], Zoran Obradovic[4], Jerome P. Trzeciakowski[5], Natasa M. Milic[1,2]

Congratulations,

Out of all 2016 PLOS Biology articles, yours was in the top 50 most downloaded.

# Courses often not targeted towards students field of interest

A Statistics usage among papers published in top physiology journals

□ No summary statistics, no statistical analysis (1.3%)

■ Summary statistics, no statistical analysis (1.7%)

■ Includes statistical analysis

97.2%

B Statistics education requirements for PhD programs in top NIH funded physiology departments

12.5%
10%
10%
63.75%

□ Not required, not listed as an elective

□ Elective

□ Recommended elective

■ Some PhD programs require statistics (3.75%)

■ Required

Recommendations for improving statistics training for biomedical students include:

1) Developing new curriculum:

- Data visualization first, statistics second

- Target misconceptions & missed skills

- Visual approach to learning

2) Encouraging departments to require statistics training,

3) Tailoring coursework to the student's field of research,

4) Developing tools and strategies to promote education and dissemination of statistical knowledge.

Reproducibility in focus

- No one denies that irreproducibility of scientific research is a serious problem

- The problem affects everyone in the scientific community: academic labs that are attempting to build on and extend others' work, biopharmaceutical companies that are searching the scientific literature for new drug targets to pursue and journal editors who are deciding which papers to publish.

- Work that is not reproducible saps time, money and energy

- Statistical community has been deeply concerned

- Radical choices: Basic and Applied Social Psychology - baned p-values (Trafimow and Marks, 2015)

- Misunderstanding or misuse of statistical inference is only one cause of the "Reproducibility crisis" (Peng, 2015)

- This concern was brought to the attention of the ASA Board
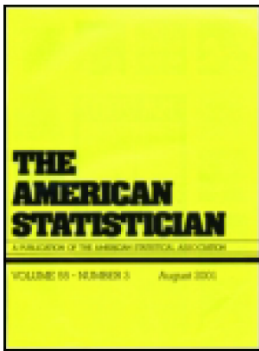
The ASA's statement on p-values: context, process, and purpose

The ASA has not previously taken positions on specific matters of statistical practice.

When the ASA Board decided to take up the challenge of developing a policy statement on p values and statistical significance, it did so recognizing this was not a lightly taken step.

By contrast, the Board envisioned that the ASA statement on p-values and statistical significance would shed light on an aspect of our field that is too often misunderstood and misused in the broader research community.

The intended audience would be researchers, practitioners and science writers who are not primarily statisticians. Thus, this statement would be quite different from anything previously attempted.

Let's be clear

- Nothing in the ASA statement is new.

- Statisticians have been sounding the alarm about these matters for decades, to little avail.

- We hoped that a statement from the world's largest professional association of statisticians would open a fresh discussion and draw renewed and vigorous attention to changing the practice of science with regards to the use of statistical inference.

What is a p-value?

- Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (for example, the sample mean difference between two compared groups) would be equal to or more extreme than its observed value.

Principles

- *P-values can indicate how incompatible the data are with a specified statistical model*

- *P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.*

- *Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.*

Principles continued

- *Proper inference requires full reporting and transparency*

- *A p-value, or statistical significance, does not measure the size of an effect or the importance of a result*

- *By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.*

Other approaches

- Methods that emphasize estimation over testing, such as

- confidence, credibility, or prediction intervals;

- Bayesian methods;

- alternative measures of evidence, such as likelihood ratios or Bayes Factors;

- and other approaches such as decision-theoretic modeling and

- false discovery rates

Editor's Perspective

The End of Journals

Harlan M. Krumholz, MD, SM

Too Static

The journal publication is currently a static product, presented as a singular contribution rather than as a living document. It can be corrected or retracted, but it is not interactive and has no capacity for iterative change spurred by input from the larger audience. Many scientific projects might be better presented as an interactive website with the opportunity for the community to probe the findings and provide feedback. Creative visualizations of data are often best presented in ways that allow images to be rotated and manipulated for better understanding. In the future, novel strategies for conveying knowledge and engaging readers will probably emerge, leaving behind the static presentation of results that offers limited options for interactive understanding.

Krumholz, Circulation CQO 2015

COMMUNITY PAGE

From Static to Interactive: Transforming Data Visualization to Improve Transparency

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

Objective:

Create tools needed to transform scientific publications from

static reports into interactive datasets

Effective figures should:

1. Immediately convey information about the study design

2. Illustrate important findings

3. Allow the reader to critically evaluate the data
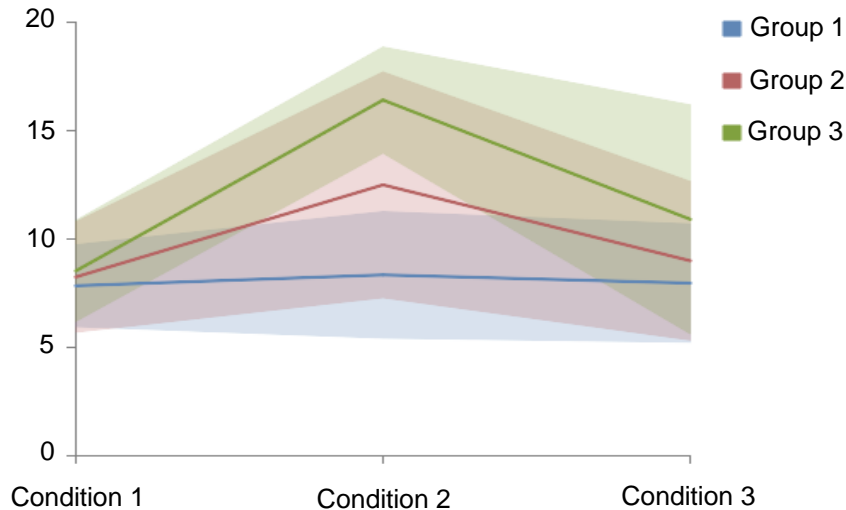
Link to video tutorial on site
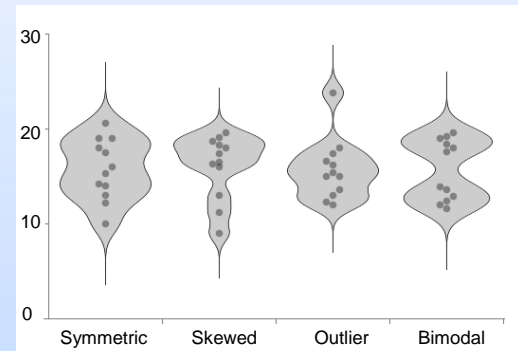
Solution: Interactive Line Graph
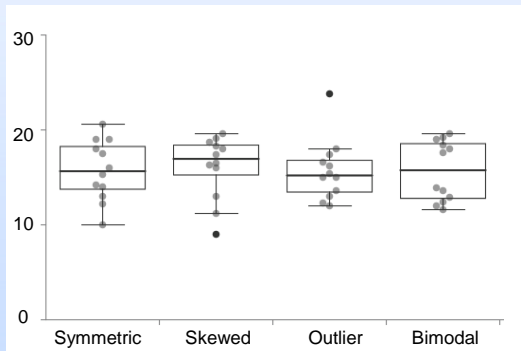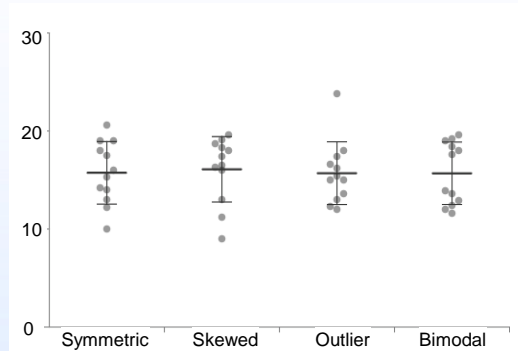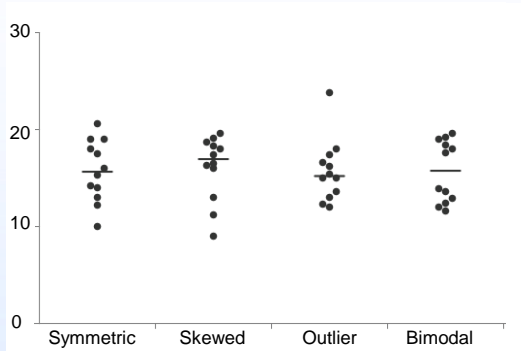http://statistika.mfub.bg.ac.rs/interactive-graph/

The interactive line graph allows you to do four things:

1. Examine different summary statistics

2. Display lines for some or all individuals in each group

3. View a subset of groups, conditions or time points

4. View change scores for any two conditions or time points

Interactive Dotplot (JBC 2018)
http://statistika.mfub.bg.ac.rs/interactive-dotplot/

Changing data presentation is critical to promote transparency

- Dear Dr. Milic,

- I'm writing to let you know that your article "Data visualization, bar naked: A free tool for creating interactive graphics" has been selected as the representative 'Methods and resources' article for our 2017 retrospective collection called "The year in JBC: 2017". Congratulations! The associate editors who took on this project looked through hundreds of papers to come up with what they felt best represented the exciting advances reported in JBC last year, and were very pleased to include your paper in the group!

All data presentation methods are a reflection of reality…
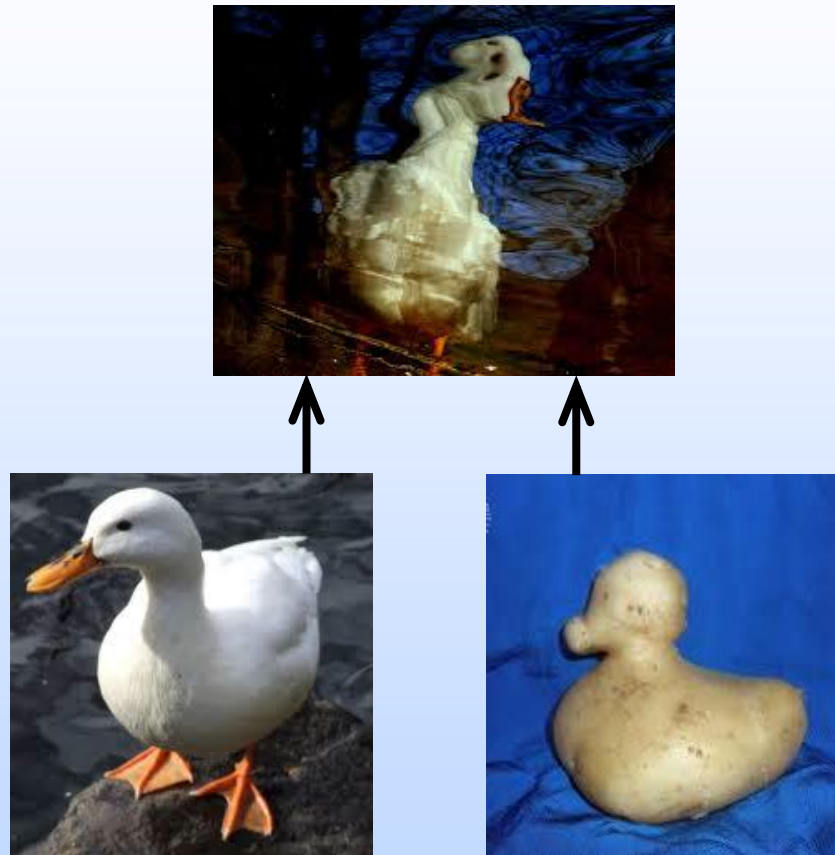
Select methods that minimize distortion

What can you do?

- Banish bar graphs from your papers and talks

- Reviewers & editors: Request figures that show data distributions

- Talk to editors about improving data presentation in their journals

- Work with statistics instructors to organize data presentation training for trainees, junior investigators and senior researchers
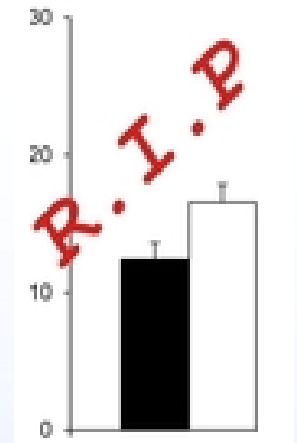
Announcement: Towards greater reproducibility for life-sciences research in *Nature*

This week we go further. Alongside every life-sciences manuscript, we will publish a new reporting-summary document.

This is another step in encouraging transparency, in ensuring that papers contain sufficient methodological detail, and in improving statistics reviewing and reporting.

New policy includes a variety of steps to improve reporting & transparency, including discouraging the use of bar graphs and encouraging authors to show the data distribution.

Efforts are under way: **Shrink the _P_ Value for Significance, Raise the Bar for Research: A Renewed Call**
March 22, 2018 _JAMA,_

- The _P_ value of .05 has once again been questioned as a threshold for clinical significance in medical research, this time in a commentary that offers a way to ease toward more relevant alternatives.

- That _P_ values are currently "misinterpreted, overtrusted, and misused" means that a research finding within the .05 standard "is wrongly equated with a finding or an outcome (eg, an association or a treatment effect) being true, valid, and worth acting on," according to John PA Ioannidis, MD, DSc, Stanford University, California

"Drowning in a Flood of Statistical Significance"

- "These misconceptions affect researchers, journals, readers, and users of research articles, and even media and the public who consume scientific information.

- A better metric, one that would serve the needs of clinicians, would reflect whether there is a treatment effect, one large enough to be clinically meaningful.

- More useful are hazard ratios (or relative risks or odds ratios) with confidence intervals that convey effect sizes that can show whether a treatment outcome may be clinically appealing. Those metrics don't simply dichotomize results in terms of significance vs nonsignificance.

Conclusion:
What Are We Trying to Answer?

- "We need to think for each study and each question that we are asking: why are we doing it, and what are we trying to answer? And then we can select the metric and the tool that will specifically look at what we want to answer„

- **No single index should substitute for scientific reasoning**

Acknowledgements

Mayo Clinic

Vesna Garovic

Tracey Weissgerber

Stacey Winham

University of Belgrade

Marko Savic

Jelena Milin-Lazovic

Temple University

Zoran Obradovic