

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/320016072>

# When One Shape Does Not Fit All: A Commentary Essay on the Use of Graphs in Psychological Research

Article in *Frontiers in Psychology* · September 2017

DOI: 10.3389/fpsyg.2017.01666

CITATIONS

8

READS

139

3 authors:



**Massimiliano Pastore**

University of Padova

107 PUBLICATIONS 1,467 CITATIONS

[SEE PROFILE](#)



**Francesca Lionetti**

University G. d'Annunzio Chieti-Pescara, Italy

71 PUBLICATIONS 600 CITATIONS

[SEE PROFILE](#)



**Gianmarco Altoè**

University of Padova

132 PUBLICATIONS 2,610 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Longitudinal trajectories of attachment relationships during middle childhood: A bayesian approach [View project](#)



Network Psychometrics [View project](#)



# When One Shape Does Not Fit All: A Commentary Essay on the Use of Graphs in Psychological Research

Massimiliano Pastore<sup>1\*</sup>, Francesca Lionetti<sup>2</sup> and Gianmarco Altoè<sup>1</sup>

<sup>1</sup> Department of Developmental and Social Psychology, University of Padova, Padova, Italy, <sup>2</sup> Department of Biological and Experimental Psychology, Queen Mary University of London, London, United Kingdom

**Keywords:** statistical reasoning, bar chart and box plot, graphical representation, exploratory data analysis, credibility crisis

## 1. INTRODUCTION

The interest of psychology in graphs is anything but new. It was 1972 when John Wilder Tukey, one of the fathers of the statistic of the XX century, distinguished among three categories of graphs: (1) *propaganda graphs*, that are intended to show what already can be learned using data-analysis and inferential testing, (2) *analytical graphs*, that allow to understand data over and above what inferential statistic has already shown, and (3) the *substitute for tables*, that are graphs from which numbers are to be read off (Tukey, 1977). From this classification it appears evident Tukey's recommendation for analytical graphs. Twenty years later, Leland Wilkinson and the Task Force on Statistical Inference of the American Psychological Association, similarly posited: "Before you compute any statistics, look at your data. (...) If you assess hypotheses without examining your data, you risk publishing nonsense" (Wilkinson and Task Force on Statistical Inference, 1999). Nowadays, visual inspection continuous to be largely recommended for understanding data set's meaning in exploratory data analysis, and is considered more useful than a solely strictly adherence to statistical testing to answer questions prompted by the experiment (Wixted and Pashler, 2002; Marmolejo-Ramos and Matsunaga, 2009). Also student's books and papers addressing mechanisms underpinning statistical reasoning have introduced a shift of perspective from *drawing graphs* to *using graphs* for making sense of data and evaluating hypotheses (Moore, 1998; Wild and Pfannkuch, 1999; Konold and Pollatsek, 2002; Bakker, 2004; Bakker and Gravemeijer, 2004; Pfannkuch, 2005; Watson, 2005; Garfield and Ben-Zvi, 2008; Matejka and Fitzmaurice, 2017). However, as we review below, a vast majority of research papers continue to adopt non optimal graphical representations. Also, though graphs could make data transparent, increasing the reliability of research findings (Tay et al., 2016), among guidelines proposed for promoting transparency in research (Nosek et al., 2015), no specific reference is made upon the relevance of adequate graphical representations.

## 2. STATIC GRAPHS AND BAR CHARTS

A recent systematic review (Weissgerber et al., 2015) of research articles published in top physiology journals in 2014 showed that the most often used graphical representations are *static graphs* and, among these, the widely known bar chart. Bar charts, useful for depicting frequencies and the occurrence of categorical variables, summarize means and standard deviations without depicting the underlying distribution of data. It results that nothing else is provided beyond what already the statistics show, increasing the risk to misinterpret research findings and to not detect important information (Cooper et al., 2002; Schriger et al., 2006; Saxon, 2015; Gelman, 2017). For example, the presence of anomalous outliers or of marked asymmetry cannot be inferred. Though a systematic review on this topic has not been published yet in psychology, research articles published between

### OPEN ACCESS

**Edited by:**

Pietro Cipresso,  
Istituto Auxologico Italiano (IRCCS),  
Italy

**Reviewed by:**

Kristian Kersting,  
Technische Universität Darmstadt,  
Germany

**\*Correspondence:**

Massimiliano Pastore  
massimiliano.pastore@unipd.it

**Specialty section:**

This article was submitted to  
Quantitative Psychology and  
Measurement,  
a section of the journal  
Frontiers in Psychology

**Received:** 18 August 2017

**Accepted:** 11 September 2017

**Published:** 25 September 2017

**Citation:**

Pastore M, Lionetti F and Altoè G  
(2017) When One Shape Does Not Fit  
All: A Commentary Essay on the Use  
of Graphs in Psychological Research.  
Front. Psychol. 8:1666.  
doi: 10.3389/fpsyg.2017.01666

January and June 2016 in four high impact psychology journals (*Behaviour Research Methods*, *Cognitive Psychology*, *Psychological Science*, and *Trends in Cognitive Science*) suggest a state of the art that does not differ much from that of other disciplines using statistical methods, with a significant presence of bar chart graphs in a field of research where continuous variables (e.g., reaction times, psychological test scores) are almost the norm (see also Bar Bar Plots Project, 2017). Specifically, on 131 research papers examined, bar charts were about 55% of 104 presented graphs.

To explain the impact of inadequate graphs more clearly and to understand the practical implications of this, we provide three examples below. To make our examples clear, we designed the three vignettes so that there is always a comparison between two experimental conditions and the analysis of variance (ANOVA) is used to statistically compare groups. The exemplifications that we provide show that an adequate use of graphs as a reasoning tool leads to results that differ from those that would have been reached if only summary statistics and static bar chart would had been adopted. We conclude that appropriate graphical representations can increase reliability in research findings and promote transparency in the way scientific information is shared and disseminated.

### 3. THE THREE VIGNETTES

#### 3.1. Example 1: The Hidden Difference

Let's suppose that 200 students are recruited from two different classes ( $a, b$ ) and randomly assigned by a researcher to two independent experimental conditions ( $x, y$ ). In each condition, subjects' performance on a specific experimental task is assessed. The research hypothesis is that there is a significant difference between the two experimental conditions.

In **Figure 1A**, number of subjects belonging to each class is depicted: 104 subjects belong to class  $a$ , and 96 to class  $b$ . The height of each bar correctly represents frequency of subjects in each condition and, therefore, bar chart is informative and pertinent. In **Figure 1B**, using the same graph, subjects' mean scores (and associated standard errors) in condition  $x$  and  $y$  are depicted: 1.8 and 1.73 respectively. Contrary to the research hypothesis, bar chart suggests no difference between the two experimental conditions, as confirmed also by ANOVA:  $F_{(1, 198)} = 0.24, p = 0.63, \text{Cohen's } d = 0.07$ .

Representing the same data using a box plot (**Figure 1C**) and a histogram (**Figure 1D**), we end up with a different conclusion. Condition  $x$  shows a skewed distribution while, on the contrary, data from condition  $y$  are more symmetrically distributed, suggesting that the two experimental conditions are not equivalent. The research hypothesis is now supported. Without the use of graphs as reasoning tools for exploring data, it would had never been possible to detect this difference in the two experimental conditions and the researcher would had not supported, inappropriately, his research hypothesis.

#### 3.2. Example 2: The Feigned Difference

Let's consider now a second example, in which we compare again two experimental conditions, but this time with less subjects per

each (small samples are common in experimental psychology) and, specifically, 50. Mean scores and standard errors are depicted in the bar chart in **Figure 1E**. Bar chart suggests that the two experimental conditions are different. Similarly, also ANOVA results suggest a significant difference between the two conditions:  $F_{(1, 98)} = 4.38, p = 0.04, \text{Cohen's } d = 0.42$ . When the same data are depicted using a box plot in the place of a bar chart, the presence of outliers appears evident (**Figure 1F**). Excluding these outlier values, the difference between the two experimental conditions is not significant anymore:  $F_{(1, 95)} = 1.87, p = 0.17, \text{Cohen's } d = 0.28$ . In relative small sample size studies, outliers may strongly influence statistical results and can be easily identified using adequate graph representations.

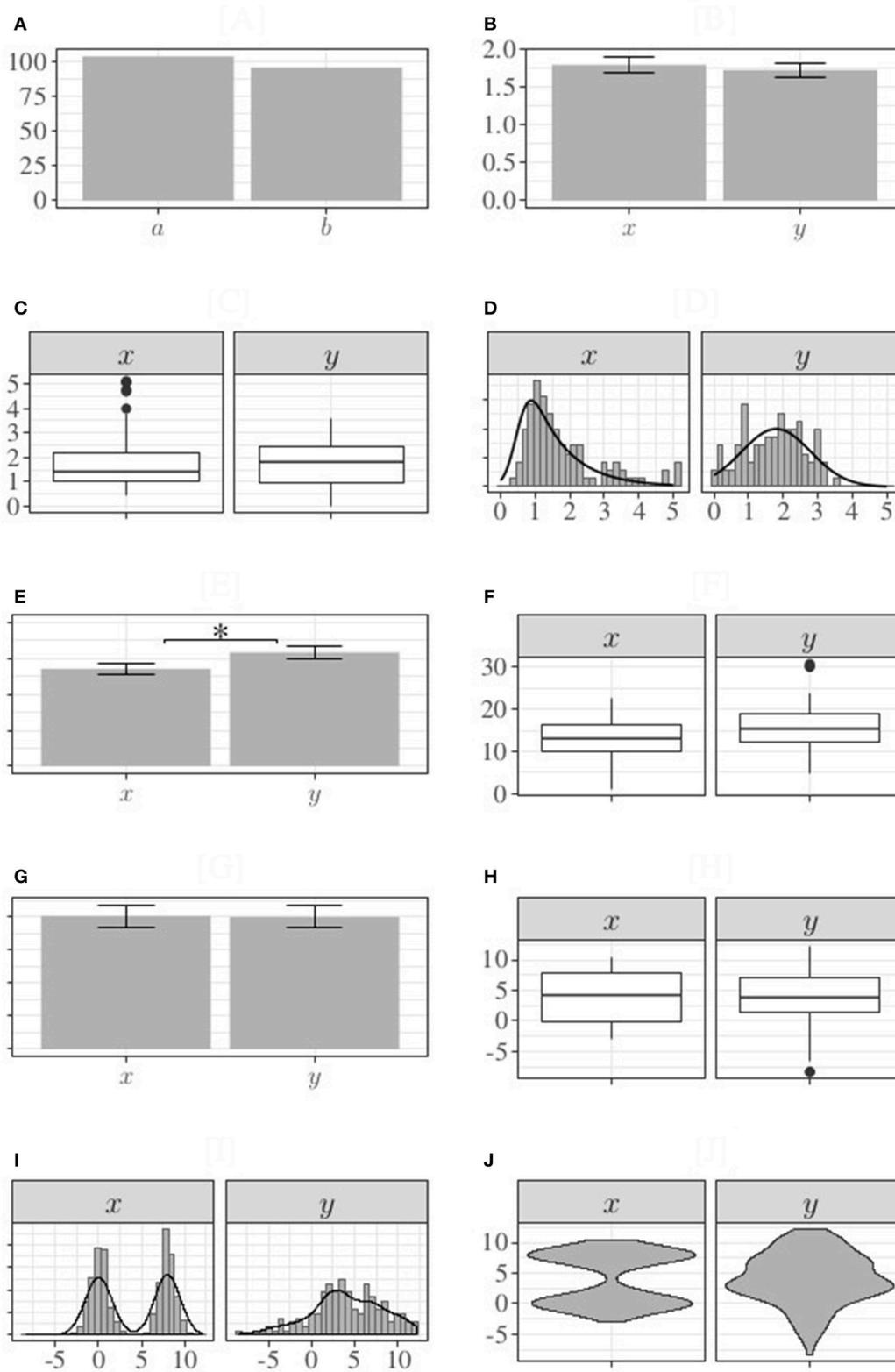
#### 3.3. Example 3: When One Graph Does Not Fit All

In the two examples reported above, the use of bar charts did not allow to properly detect false negative (example 1) and false positive (example 2) research findings. In both cases, adopting two alternative graph options (box plot for summarizing data, and histogram for plotting individual values) it was possible to accurately explore pattern of data that were otherwise concealed by bar charts. Is box plot the best graphical option able to adequately fit any type of data? Unfortunately, this is not the case. Box plots, more informative than bar charts for representing summary statistics of continuous variables, may fail as well in specific conditions, as we discuss in the current example.

Let's consider again two experimental conditions: means and standard errors are depicted in **Figure 1G**. It's easy to see that mean values (and standard errors) are comparable [see also  $F_{(1, 598)} = 0.01, p = 0.94, \text{Cohen's } d = 0.01$ ]. When a box plot is adopted in the place of a bar chart, a difference in variability emerged (see **Figure 1H**). Using a histogram for further exploring the pattern of data (see **Figure 1I**) a bimodal distribution in the condition  $x$  becomes easily identifiable. Box plots, more informative than bar charts (as we demonstrated with Example 1 and Example 2), may be not enough for gaining full information of data because do not allow to detect multimodal distribution. In this case the more adequate representation is the violin plot (see **Figure 1J**). Such type of graph, including information about densities of the distributions, allow to detect even multimodal distribution of data.

### 4. CONCLUDING REMARKS

Graphical representations are useful to become familiar with and understand the concept of variation, as well as to investigate the sources and the impact of variance on observed data, which are among the main aims of psychological research. Graphs remind us that the process of statistical inference is not mechanical (Gigerenzer and Marewski, 2015; McElreath, 2016). This process often involves subjective decisions (e.g., the evaluation, exploration, and/or deletion of outliers) which are an integral part of the analysis. Thus, graphs are among the most appropriate tools to enhance transparency and confer plausibility to such decisions.



**FIGURE 1** | Graphical representation of examples provided in the text. (A–D) Refer to example 1, section The Hidden Difference; (E,F) refer to example 2, section The Feigned Difference; (G–J) refer to example 3, section When One Graph Does Not Fit All. \*Indicates a significant difference between means at 0.05 level.

As we demonstrated with three simple examples, an accurate visual representation of data plays a pivotal role in the interpretation of research findings, representing a truly inferential statistic tool. Bar chart graph does not allow to fully explore data distribution, and may conceal important information increasing the risk to publish unreliable findings that fail at replication. Also, not adding anything more than what already summary statistics show, bar chart is of limited utility for promoting statistical reasoning on data, increasing the risk of a mechanical approach to data analysis. Widely adopted in psychology, bar chart is useful for depicting frequencies and categorical variables but may be misleading if adopted to represent summary statistics of continuous variables, as a companion of *t*-test and ANOVA. Histograms, allowing to depict the distribution of all data, are recommended, but do not offer the opportunity of summarizing data in a clear and effective way. Other graphical representations, more informative but less disseminated in our field, as box plot and violin plot, could represent a more informative option. More generally, combining various graphical techniques could allow researchers

to know more about full data, and promote the access to relevant information otherwise concealed by static data graphs. Improved data representation techniques could be one of the way for enhancing students' understanding of statistical reasoning, the scientific community understanding of published data, and a critical evaluation of research findings. Also, making data transparent, graphs may represent one of the answers to the crisis of credibility in psychology (see: Ioannidis, 2005; Pashler and Wagenmakers, 2012; Open Science Collaboration, 2015). The increased availability of powerful statistical software has opened the possibility of using new and more sophisticated analytic approaches, which are testified by recent scientific publications. However, this progress has yet to be fully integrated in graphical representations of data.

## AUTHOR CONTRIBUTIONS

MP proposed the topic and designed the vignettes; FL and GA contributed to further develop the idea; MP, FL, and GA wrote the paper.

## REFERENCES

- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Stat. Educ. Res.* 3, 64–83.
- Bakker, A., and Gravemeijer, K. P. (2004). "Learning to reason about distribution," in *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, ed B. V. Dordrecht (Springer Science+Business Media), 147–168.
- Bar Bar Plots Project (2017). #barbarplots. Available online at: <https://www.kickstarter.com/projects/1474588473/barbarplots>. (Accessed 2017-08-15).
- Cooper, R. J., Schriger, D. L., and Close, R. J. (2002). Graphical literacy: the quality of graphs in a large-circulation journal. *Ann. Emerg. Med.* 40, 317–322. doi: 10.1067/mem.2002.127327
- Garfield, J., and Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*, ed B. V. Dordrecht (Springer Science+Business Media)
- Gelman, A. (2017). Graphs as comparisons: a case study. *Blog Post*. Available online at: <http://andrewgelman.com/2017/07/16/graphs-comparisons-case-study>
- Gigerenzer, G., and Marewski, J. N. (2015). Surrogate science the idol of a universal method for scientific inference. *J. Manage.* 41, 421–440. doi: 10.1177/0149206314547522
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124
- Konold, C., and Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *J. Res. Mathe. Educ.* 33, 259–289.
- Marmolejo-Ramos, F., and Matsunaga, M. (2009). Getting the most from your curves: exploring and reporting data using informative graphical techniques. *Tutor. Quantit. Methods Psychol.* 5, 40–50. doi: 10.20982/tqmp.05.2.p040
- Matejka, J., and Fitzmaurice, G. (2017). "Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, CO: ACM), 1290–1294.
- McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: CRC Press.
- Moore, D. S. (1998). Statistics among the liberal arts. *J. Am. Stat. Assoc.* 93, 1253–1259. doi: 10.1080/01621459.1998.10473786
- Nosek, B. A., Alter, G., Banks, G., Borsboom, D., Bowman, S., Breckler, S., et al. (2015). Promoting an open research culture. *Science* 348, 1422–1425. doi: 10.1126/science.aab2374
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716
- Pashler, H., and Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science a crisis of confidence? *Perspect. Psychol. Sci.* 7, 528–530. doi: 10.1177/1745691612465253
- Pfannkuch, M. (2005). "Probability and statistical inference: how can teachers enable learners to make the connection?" in *Exploring Probability in School* (New York, NY: Springer), 267–294.
- Saxon, E. (2015). Beyond bar charts. *BMC Biol.* 13:60. doi: 10.1186/s12915-015-0169-6
- Schriger, D. L., Sinha, R., Schroter, S., Liu, P. Y., and Altman, D. G. (2006). From submission to publication: a retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. *Ann. Emerg. Med.* 48 750–756. doi: 10.1016/j.annemergmed.2006.06.017
- Tay, L., Parrigon, S., Huang, Q., and LeBreton, J. M. (2016). Graphical descriptives: a way to improve data transparency and methodological rigor in psychology. *Perspect. Psychol. Sci.* 11, 692–701. doi: 10.1177/1745691616663875
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Watson, J. (2005). The probabilistic reasoning of middle school students. *Math. Educ. Lib.* 40:145. doi: 10.1007/b105829
- Weissgerber, T. L., Milic, N. M., Winham, S. J., and Garovic, V. D. (2015). Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol.* 13:e1002128. doi: 10.1371/journal.pbio.1002128
- Wild, C. J., and Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *Int. Stat. Rev.* 67, 223–248. doi: 10.1111/j.1751-5823.1999.tb00442.x
- Wilkinson, L., and Task Force on Statistical Inference (1999). Statistical methods in psychology journals - guidelines and explanations. *Am. Psychol.* 54, 594–604. doi: 10.1037/0003-066X.54.8.594
- Wixted, J., and Pashler, H. (eds.). (2002). *Stevens' Handbook of Experimental Psychology*. New York, NY: Wiley.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Pastore, Lionetti and Altoè. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.