

The shifting boxplot. A boxplot based on essential summary statistics around the mean.

El diagrama de caja cambiante. Un diagrama de caja basado en sumarios estadísticos alrededor de la media

Fernando Marmolejo-Ramos
University of Adelaide
Universidad del Valle

Tian Siva Tian
University of Houston

ABSTRACT

Boxplots are a useful and widely used graphical technique to explore data in order to better understand the information we are working with. Boxplots display the first, second and third quartile as well as the interquartile range and outliers of a data set. The information displayed by the boxplot, and most of its variations, is based on the data's median. However, much of scientific applications analyse and report data using the mean. In this paper, we propose a variation of the classical boxplot that displays information around the mean. Some information about the median is displayed as well.

Key words: Boxplots, confidence intervals, parametric tests, bootstrap, R.

RESUMEN

Los diagramas de caja son una técnica gráfica útil y ampliamente usada para explorar datos y así entender mejor la información con la que se está trabajando. Los diagramas de caja muestran el primer, segundo, y tercer cuartil como también la amplitud intercuartil y los valores extremos de un grupo de datos. La información mostrada en los diagramas de caja, y en varias de sus variaciones, está basada en la mediana de los datos. Sin embargo, la gran mayoría de las aplicaciones científicas analizan y reportan datos usando la media. En este artículo se propone una variación del diagrama de caja que presenta información alrededor de la media de los datos. La presente variación también presenta alguna información acerca de la mediana de los datos.

Palabras clave: diagramas de caja, intervalos de confianza, pruebas paramétricas, bootstrap, R.

Article received/Artículo recibido: December 15, 2009/Diciembre 15, 2009, Article accepted/Artículo aceptado: March 15, 2010/Marzo 15/2010

Dirección correspondencia/Mail Address:

Fernando Marmolejo-Ramos, Fernando Marmolejo-Ramos, School of Psychology, Faculty of Health Sciences, University of Adelaide, Australia and Institute of Psychology, Universidad del Valle, Colombia. Email: fernando.marmolejoramos@adelaide.edu.au, web page: www.firthurands.com

Tian Siva Tian, Departments of Psychology & Electrical and Computer Engineering, University of Houston. Email: ttian@uh.edu, web page: <http://www.uh.edu/~ttian/>

INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH esta incluida en PSERINFO, CENTRO DE INFORMACION PSICOLOGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET y GOOGLE SCHOLARS. Algunos de sus artículos aparecen en SOCIAL SCIENCE RESEARCH NETWORK y está en proceso de inclusion en diversas fuentes y bases de datos internacionales.

INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH is included in PSERINFO, CENTRO DE INFORMACIÓN PSICOLÓGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET and GOOGLE SCHOLARS. Some of its articles are in SOCIAL SCIENCE RESEARCH NETWORK, and it is in the process of inclusion in a variety of sources and international databases.

1. INTRODUCTION

Most researchers in several sciences, including Psychology, rely on results of parametric tests, like ANOVA and *t*-test. Parametric tests depend on two major assumptions in order to give unbiased results: homogeneity of variance and normality of data. It has been demonstrated that even small violations of those assumptions can cause the tests to give biased results (Wilcox, 1998). Furthermore, not only the results of parametric tests but also those of non-parametric tests can be affected when assumptions are violated (Zimmerman, 1998).

The homogeneity assumption requires that variances among batches of data are similar as indicated by homogeneity tests (e.g., Levene's test for the mean and Brown-Forsythe test for the median). The normality assumption requires that batches of data have normal distributions as indicated by normality tests (e.g., D'Agostino normality test, Kolmogorov-Smirnov test, and Shapiro-Wilk test). The mean (\bar{x}) and the standard deviation (*s*) are intrinsic components in the computations sustaining parametric tests. That is so, because the mean and the standard deviation are the most efficient unbiased estimators of location and scale, respectively, for normally distributed data (see Rosenberg & Gasko, 1983). So, it is no wonder that those estimators are frequently used in analysing and reporting data submitted to parametric tests.

A recommended approach to reporting data is via graphical methods. Graphs do not aim to convey numbers with decimal places, but rather to help the researcher find and report *patterns* in the data (Cleveland & McGill, 1984). Furthermore, research on graphical techniques suggests that graphs are more useful than tables at communicating comparisons among batches of data (Gelman, Pasarica, & Dohia, 2002). One of the most popular graphical techniques mainly used to explore data is called the boxplot and it presents summary statistics around the median, not the mean.

In this paper we present a variation of the boxplot which displays summary statistics based on the mean. The first section describes the traditional boxplot, some of the proposed variations, and its advantages and disadvantages. The second section presents a new variation, and the final section covers some conclusions and recommendations. The paper focuses more on the graphical performance of the present variation than on the technicalities of the computations supporting its construction. It is important to note that *the featured boxplot is designed to report data which have homogeneous variances and follow a normal distribution*. Nonetheless, it can be used to explore and report non-normal data so long as a robust estimator of

central tendency is added to the display and the interest is on the data's mean.

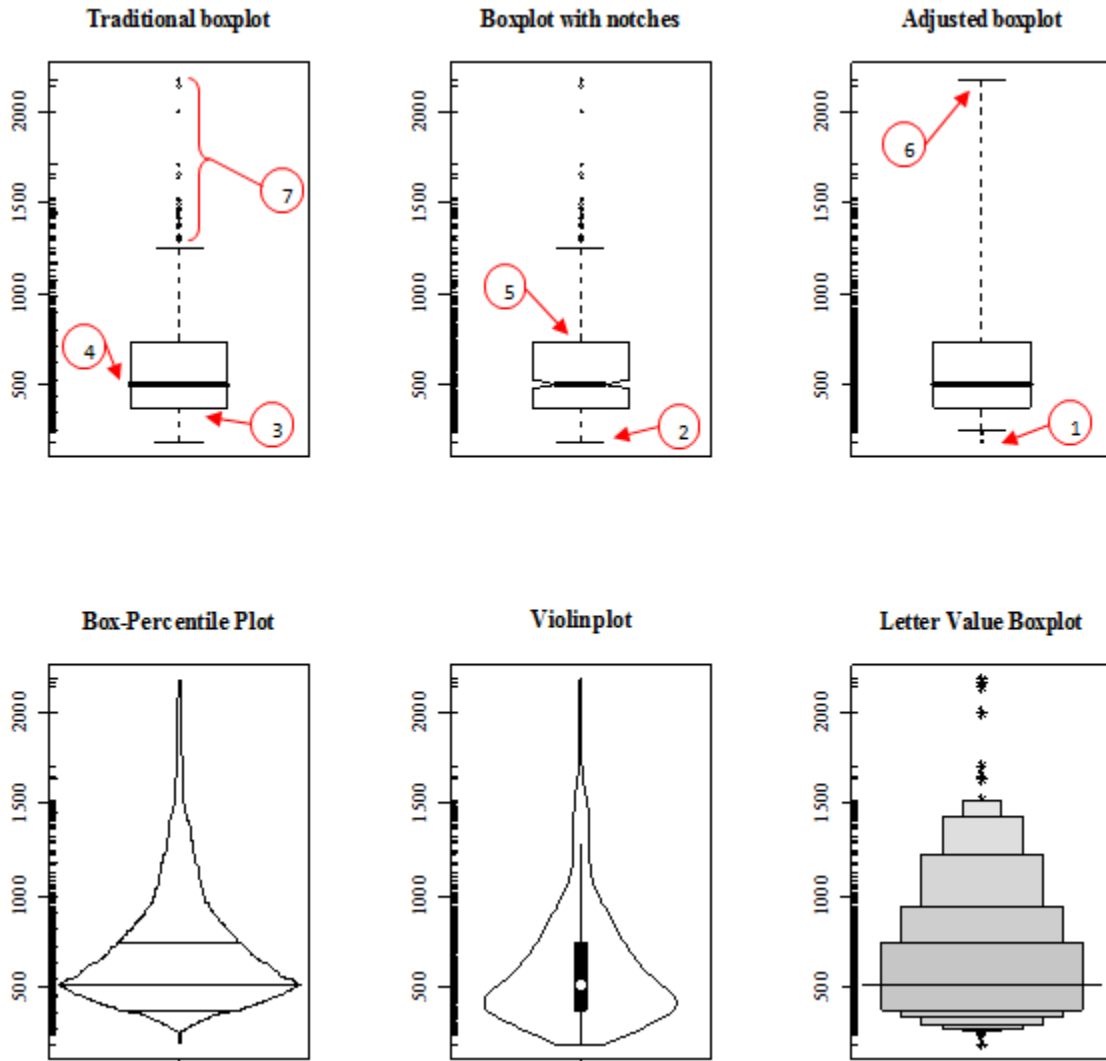
2. THE BOXPLOT AS A GRAPHICAL METHOD TO EXPLORE DATA

Graphical methods comprise any form of visualisation of quantitative data. Such methods are the basis for the exploration of data and have been used for the communication of results as well. Graphical methods assist in making statistical decisions, selecting methods to analyse data, and evaluating the limitations of typical null hypothesis tests (see Loftus, 1993; Marmolejo-Ramos & Matsunaga, 2009). One of the methods most commonly used in the visual inspection of data is the boxplot.

The boxplot is a graphical technique that depicts, in its traditional form, five numeric summaries about a data set in order to visualise its dispersion and skewness (McGill, Tukey, & Larsen, 1978). Those summaries are based on the median and correspond to the smallest observation, the median of the first half of the data (first quartile, Q_1), the median (second quartile, Q_2), the median of the second half of the data (third quartile, Q_3), and the largest observation. The area between the first and third quartile is known as the interquartile range and it gives an indication of spread in the data ($IQR = Q_3 - Q_1$). The *IQR* corresponds visually with the only box in the display and it covers approximately 50% of the observations closer to the median. The smallest and largest observations are those that fall outside the lines (or *whiskers*) that connect the *IQR* to the smallest or largest value that is not an outlier (e.g., within $1.5 \times IQR$). Additionally, sometimes the traditional boxplot can include an approximate 95% confidence interval around the median; also called the "notch" (see Figure 1).

Variations of the boxplot exist and some of them can be found in Potter (2006) plus a short explanation of their roots. Nevertheless, more variations have recently been proposed. For example, Hubert and Vandervieren (2008) proposed an adjusted boxplot based on the medcouple measure of skewness to determine the length of the whiskers. This variation enables the whiskers to signal outliers without making any assumption about the data's distribution. Similarly, Schwertman, Owens, and Adnah (2004) present a boxplot which modifies the computations of whiskers' limits under the assumption of normality and large samples (see also Sim, Gan, & Chang, 2005). Another variation termed "letter-value boxplot" has been proposed to show more information when presenting large data sets (Hofmann, Kafadar, & Wickham, 2009) (see Figure 1). Also, there are variations specifically designed for other sciences (e.g., Graedel, 1977).

Figure 1. Six different boxplot variations. All the boxplots are representing an Ex-Gaussian distribution of size 500 and parameters $\bar{x} = 300$, $\sigma = 50$, and $\tau = 300$. Actual observations are represented by a rug on the left axis in each plot. Traditional numeric summaries are indicated by circled numbers: 1. smallest observations, 2. lower limit of whiskers, 3. first quartile, 4. median, 5. third quartile, 6. upper limit of whiskers, and 7. largest observation. The box-percentile plot is proposed by Esty and Banfield (2003; as cited in Potter, 2006). The violinplot is discussed later in this paper and elsewhere in more detail (see Marmolejo-Ramos & Matsunaga, 2009).



Graphical demonstrations suggest that the advantage of the boxplot is that it can differentiate between the shapes of the normal and other distributions, but it fails to do so between the bimodal and uniform distributions. Another disadvantage of the boxplot is that it cannot inform about clusters of data (Hintze & Nelson, 1998). Thus, the variation presented in this paper carries by default both the advantages and the disadvantages of the traditional boxplot. However, in actual research, data sets tend not to present unimodal distributions, but rather positively skewed

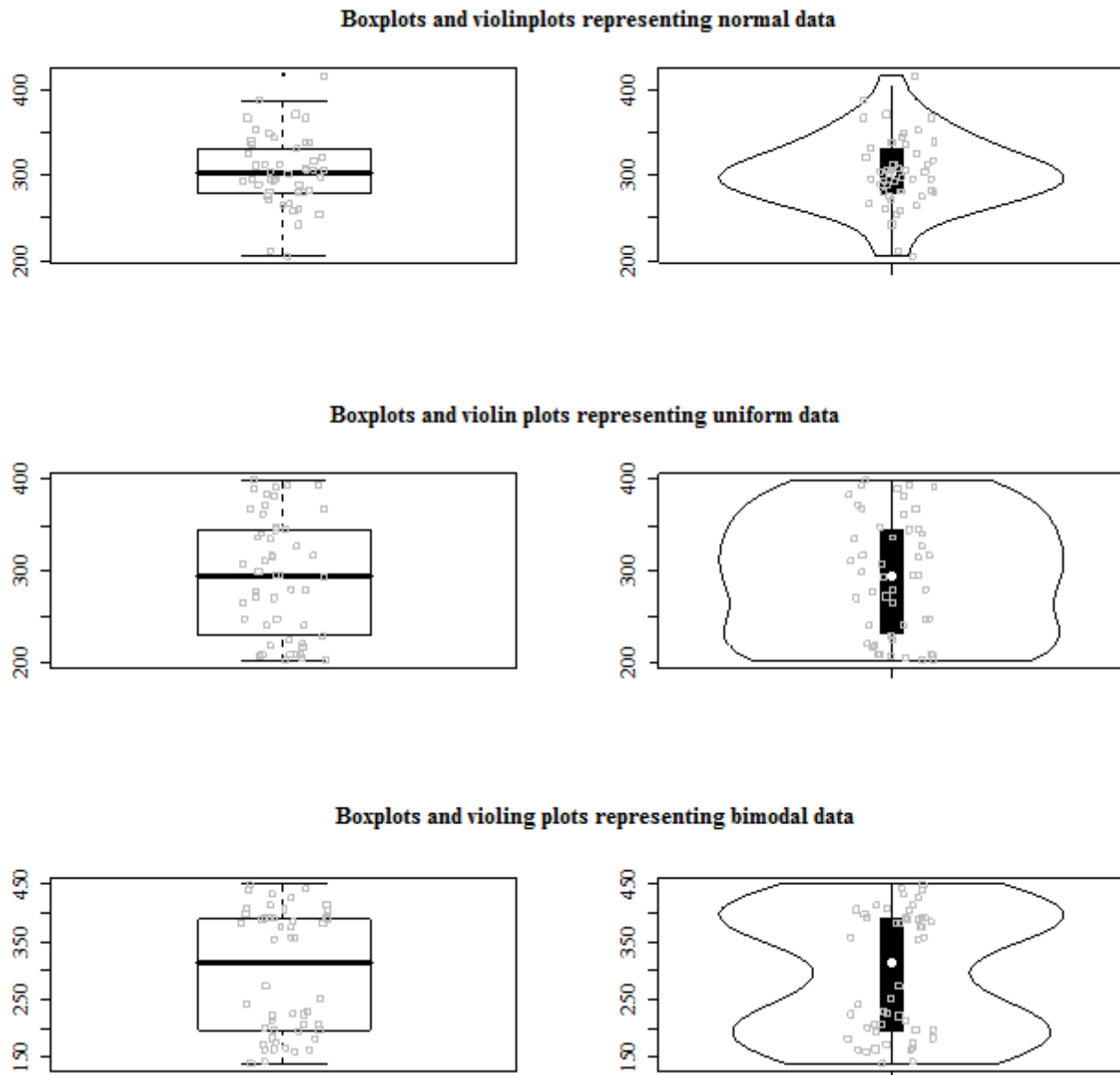
distributions (e.g., reaction times data). Also, normally distributed data can occur in research (e.g., Likert-type ratings).

A graphical method to differentiate between bimodal and uniform distributions, if encountered in actual research, is to use instead of the boxplot, the violinplot. The violinplot presents a boxplot surrounded by the underlying distribution of the data (Hintze & Nelson, 1998). This graphical method stems from the traditional boxplot,

therefore it presents summary statistics around the median, but it can be customised to display the mean and a 95% confidence interval around it (see Marmolejo-Ramos & Matsunaga, 2009). A simple and efficient way to deal with

the visualisation of clusters of data is to add a scatterplot to the boxplot display (see Figure 2).

Figure 2. Boxplots and violinplots representing three different distributions. The first row represents a normal distribution of size 50 and parameters $\bar{x} = 300$, $\sigma = 50$. The second row represents a uniform distribution of size 50 and parameters $x_{min} = 200$ and $x_{max} = 400$. The third row represents a bimodal distribution of size 50 and parameters $\bar{x} = 200$ and $\sigma = 30$ ($n = 25$) for the first normal distribution, and parameters $\bar{x} = 400$ and $\sigma = 30$ ($n = 25$) for the second normal distribution. Actual observations are displayed as scatterplots.



All of the reviewed boxplot variations use the median as the basis for their summaries and just a few variations present summary statistics around the mean or a value close to it (see Frigge, Hoaglin, & Iglewicz, 1989). The reason why the median is used as the basis for the computations is because it gives a reliable approximation of central tendency in non-normal distributions and is resistant to outliers (Rosenberg & Gasko, 1983). Such characteristics render the boxplot a powerful technique for gaining insight into data prior to formal analyses, especially when no assumption of normality is made. At the same time, though, such characteristics might preclude using the boxplot in the communication of results since most data are analysed and reported using summary statistics around the mean.

It was mentioned earlier that current variations of the boxplot use different computations for the length of the whiskers. That lack of agreement results in different statistical software displaying different boxplots for the same batch of data (see for example the difference in displays between the adjusted boxplot and the traditional boxplot shown in Figure 1). Another factor that adds to such divergence in displays is that the computation of the quartiles can be performed using different formulas (see Frigge et al., 1989). Moreover, other researchers recommend the use of quantiles instead of quartiles, but even so there are also different ways of computing them (see Hyndman & Fan, 1996). Therefore, it is advantageous to propose a boxplot which is based on commonly used summary statistics with computations which have great agreement among researchers.

3. THE SHIFTING BOXPLOT. A BOXPLOT DESIGNED TO REPORT DATA SUITABLE FOR PARAMETRIC TESTS

It is important to propose innovative graphical techniques, or create variations of classical ones, in order to promote newer ways of visual reasoning about data. Thus, a variation of a well-known graphical technique is practical so long it has broad applicability and is based on well-founded computations. Furthermore, any graphical variation should aim to be an efficient means to present and analyse information in a scientific manner (see Best, Smith, & Stubbs, 2001). The present variation meets those requirements through a modification of the traditional boxplot in order to display information around the mean.

The mean is selected as the core component of the present boxplot variation since it is part of most statistical tests and it is the most commonly reported estimator of location in research. In addition, it is also known that the mean and its associated estimator of scale (the standard deviation) are unbiased estimators for normal distributions (Rosenberg & Gasko, 1983). The normality assumption is one of the characteristics which parametric tests rely on,

along with the homogeneity of variance assumption. When such assumptions are violated both parametric tests and non-parametric tests (in which no assumptions are made) give biased results (see Wilcox, 1998; Zimmerman, 1998). Therefore, if the mean is to be reported as the location estimator of a data set, it is desirable that groups fulfil parametric assumptions in order to obtain correct estimations. As these assumptions apply principally to parametric tests, the shifting boxplot (here *SB*) is aimed to report results obtained via such tests.

The *SB* displays nine numeric summaries about a normally distributed batch of data:

1. Smallest outlying observation.
2. Minimum value within a $\pm 2 s$ range
3. Mean of first half of the data ($Q1_{\bar{x}}$)
4. Lower 95% CI limit for the mean
5. Mean ($Q2_{\bar{x}}$)
6. Upper 95% CI limit for the mean
7. Mean of second half of the data ($Q3_{\bar{x}}$)
8. Maximum value within a $\pm 2 s$ range
9. Largest outlying observation.

It is common among researchers to consider observations with less than 5% frequency as outlying observations (see Cowles & Davies, 1982), i.e., it is common to use a 2 standard deviations fence. In the *SB*, outlying observations are those that fall below $-2 s$ (1) and above $+2 s$ (9) and are represented by dashes.

Observations which are equal to or above $-2 s$ (2) and equal or below $+2 s$ (8) are enclosed by the innermost box (the long thin box) in the display. This box covers approximately 95% of the data and might have a similar role to that performed by the whiskers in the traditional boxplot. Note, however, that the whiskers in the traditional boxplot cover approximately 99% of the data.

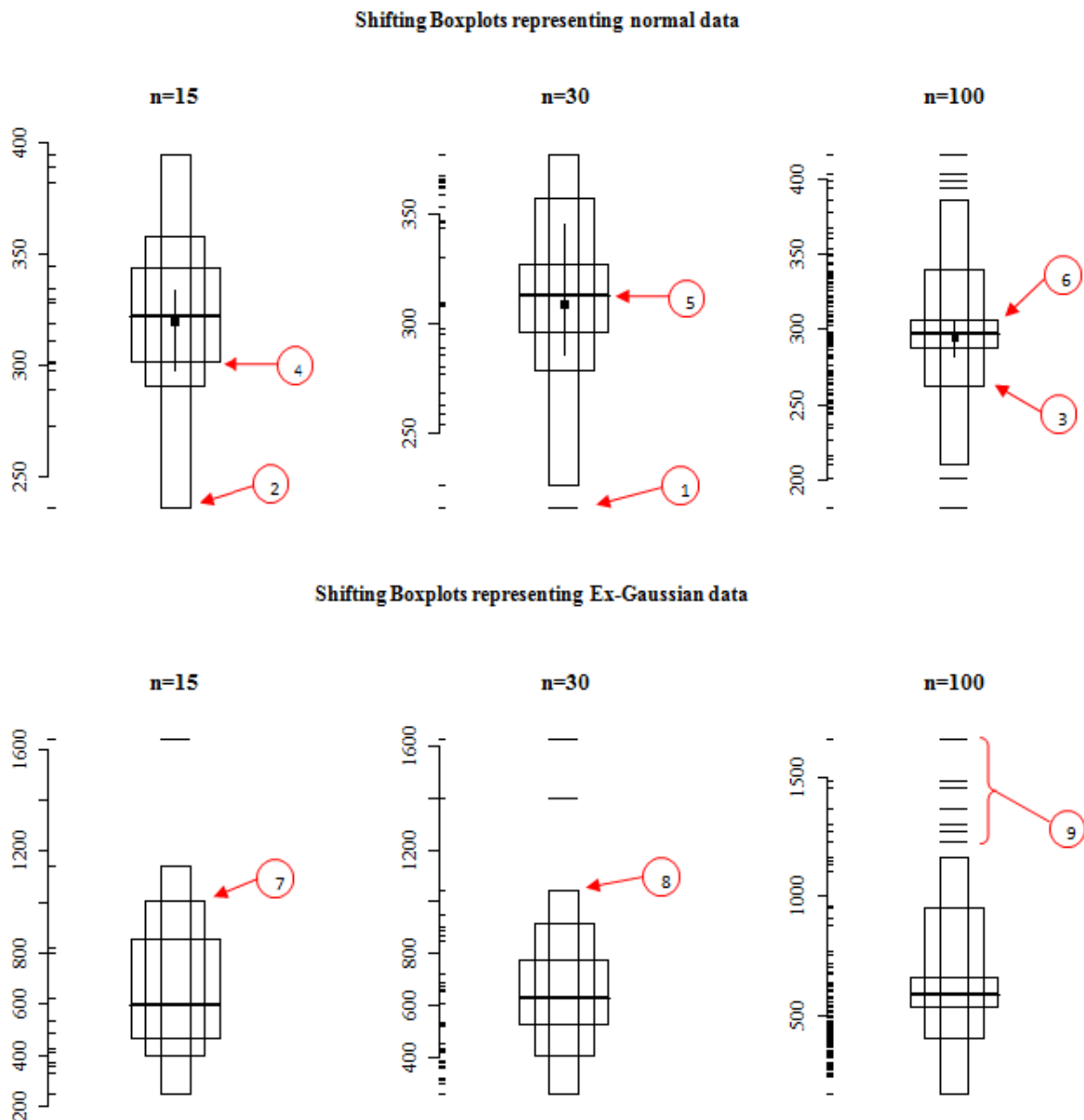
Observations that fall between the mean of the first half of the data (3) and the mean of the second half of the data (7) are enclosed by the middle box. This box might cover approximately 50% of those observations closer to the mean. We use the notation $Q1_{\bar{x}}$ and $Q3_{\bar{x}}$ to indicate that the limits of this box are quartiles based on the mean. As with the traditional quartiles, the mean of the data is not included to compute the means of the halves.

The thick horizontal line joining the borders of the outermost box represents the mean of the data set (5) and the horizontal limits of this box represent the lower (4) and upper (6) limits of the confidence interval (or CI). Thus, the *SB* is visually represented by 3 boxes and as many dashes as outlying observations exist (see Figure 3). It is worth noting that the information displayed in the present boxplot is in line with other graphical techniques used for

pedagogical and exploratory purposes (see Doane & Tracy, 2000). Thus, the *SB* could also be used as a tool to teach statistics, foster understanding of the distribution of

observations, and facilitate the comparison of distributions.

Figure 3. The shifting boxplot representing two different types of distributions with three different sample sizes. The first row represents normal distributions of sizes 15, 30 and 100 with parameters $\bar{x} = 300$, $\sigma = 50$. The second row represents Ex-Gaussian distributions of sizes 15, 30, and 100 with parameters $\bar{x} = 300$, $\sigma = 50$, and $\tau = 300$. Circled numbers indicate the nine numeric summaries that the boxplot displays (see text). The median and its 95% CI_{BCa} are also displayed as a small square with whiskers. Actual observations are represented by a rug on the left axis in each plot.



3.1. A note on confidence intervals and their interpretation in the shifting boxplot

Researchers recommend reporting CIs, since they not only give an idea of an experiment's power (see Loftus,

1996), but also enable one to make inferences about the data (Levine, Weber, Park, & Hullett, 2008). One of the core advantages of using confidence intervals is that they are inferential statistics that allow researchers to draw theoretical conclusions based on *data patterns* (Loftus,

1993) without the need of typical null hypothesis statistical testing (or NHST) (see Cumming & Finch, 2005; Masson & Loftus, 2003; Tryon, 2001). Indeed, one of the major criticisms to NHST is that it can only reject the null or fail to reject it but never gives support to it (e.g., Gallistel, 2009). Hence, CIs are perhaps the only available statistical estimator that would enable researchers not only to reject (or fail to reject) the null but also to propose inferences about why the null should be supported.

Traditional 95% CIs around the mean are computed by adding and subtracting to the mean the result of the multiplication of the standard error times 1.96 (or z score) ($\bar{x} \pm z_{95} \cdot \frac{\sigma}{\sqrt{n}}$). A more precise 95% CI can be obtained by multiplying the standard error times the t critical value for a two-tailed test with the respective degrees of freedom ($\bar{x} \pm t_{95,df} \cdot \frac{s}{\sqrt{n}}$). The computation of CIs for independent measures can be carried out for each data set in the manner presented above. However, it has been proposed that a correction for the computation of CIs should be performed when dealing with dependent measures. Such computation basically removes the between-groups variance since it is of no avail in the statistical analysis of dependent-measures designs (Cousineau, 2005; Jarmasz & Hollands, 2009; Loftus & Masson, 1994; Masson & Loftus, 2003; Morey, 2008). In other words, not taking into account the between-groups variance does not alter the results of a parametric test for dependent measures.

In the case of independent measures a rule of thumb would indicate that when two groups have significantly different means there will be less than 50% of overlap between the CIs of the groups (see Marmolejo-Ramos & Matsunaga, 2009). In graphical terms, there could be a significant difference between 2 independent groups when the edge of one of the data sets' CIs would be less than half-way through from reaching the other data set's mean. We assume that the correction applied to dependent-measures designs will give CIs an approximately similar behaviour. Thus, the graphical advantage of the correction is that the degree of overlap between CIs of different groups will roughly correspond to the results of the p value given by the statistical test in the case of dependent measures. Thus, users of the *SB* should take into account these considerations not only when plotting and interpreting CIs in the boxplot, but also when employing CIs in general.

3.1.2. Implementing more accurate confidence intervals

Although traditional CIs are broadly used and well-known, it has been suggested that they are not very accurate. Simulation studies indicate that exact CIs are asymmetrical rather than symmetrical, particularly when some skewness is evident, so CIs should adapt to the actual

distribution of the observations. To obtain more precise CIs it is recommended to use bootstrap confidence intervals (here CI_{boot}). The advantage of CI_{boot} is that they aim to produce CIs that are closer to those expected in the population and have better coverage than usual CIs (see Efron, 1987). The way to achieve that is by producing many theoretical sub-samples with replacement from the available sample, estimating the mean for each of them, and then computing CIs for the set of obtained re-sampled means. Note that the size and range parameters of the subsamples are those of the original sample. Production of CI_{boot} is a computer-intensive method and accurate results depend ultimately on the quality of the available data set (see Efron, 1988; Wood, 2005)

One of the existent CI_{boot} is called bias-corrected and accelerated CI (here CI_{BCa}) which incorporates a bias constant and an acceleration constant to produce accurate CIs. The constants enable CIs to adapt to the skewness of the sample while using some mathematical transformations to introduce normality and stabilise variance (technicalities can be found in Efron, 1987). The appeal of the CI_{BCa} is that although it uses non-parametric re-sampling, it does perform well under parametric assumptions. Thus, the CI_{BCa} can be estimated for practically any data set regardless of the type of distribution or the sample size. Given these reasons, the *SB* adopts the CI_{BCa} in the representation of data using by default 2000 re-samples.

3.1.3. Addition of a robust estimator of central tendency to explore non-normal data

The addition of the median and its confidence interval renders the *SB* suitable to explore data that is suspected to distribute non-normally. As the median is a robust estimator of central tendency in non-normal distributions, it will tend to fall close to where the true central location falls (see Sartori, 2006). On the other hand, as the mean is inefficient in non-normal distributions, it will tend to be pulled towards where most of the extreme observations fall (see Rosenberg & Gasko, 1983). As a result, the gap created between the mean and the median will give a visual indication of the degree of skewness and spread of the data. In addition, the degree of asymmetry of the boxes in relation to the mean will also assist in visualising the data's level of normality.

These are important visual properties that a boxplot should have since they warn the user about potential lack of normality in the data that could be due to factors other than a small sample size (see Choonpradub & McNeil, 2005).

The location estimator selected to explore non-normal data is the median and it is accompanied by its CI_{BCa} in the shifting boxplot. There are various estimators

of location robust to non-normal cases (e.g. Huber, Hampel, and Andrews), but the median is chosen because i) it is a well-known estimator of central tendency, ii) it is easy to compute, and iii) it performs well in several non-normal distributions (see Goodall, 1983). An interesting fact of adding a 95% CI_{BCa} around the median is that a low degree of overlap between CIs would suggest that a test on medians could yield significant results when CIs around the mean might suggest otherwise.

For this reason, it is recommended that the median and its CIs are used in the exploration of data prior to formal analysis. Additionally, it does not come amiss to display the median and the CIs when reporting data subjected to parametric tests since it would endorse the results or flag potential discrepancies at the median level. The *SB* displays the mean, the median, and their CIs by default and the median and its CIs can be suppressed if requested (see second row in Figure 3).

CONCLUSIONS

Although the graphical variation of the boxplot presented here has properties which make it useful both for exploring and reporting data (see Cox, 1978), we believe it is more appropriate to report data suitable for parametric tests. When non-normally distributed data need to be reported, it is appropriate to use the traditional boxplot. Nonetheless, the *SB* displaying the median and its CIs could be another reasonable approach to reporting non-normal data.

The *SB* is a useful graphical tool in that it keeps the graphical simplicity of the traditional boxplot and displays summary statistics around a broadly used estimator of location. Thus, the *SB* displays more information about data sets than graphical methods that are limited to conveying only means and error bars (e.g., bar plots or dynamite plots) (see Lane & Sándor, 2009). Additionally, the *SB* features accurate CIs that assist in making solid statistical inferences about results in the population. Future modifications of the *SB* could aim to implement other types of bootstrap CIs and even traditional symmetrical CIs.

Regarding the visual design of the present boxplot, like the traditional boxplot, the *SB* could be perceived as a single perceptual unit. In particular, the *SB* displays nine summary numbers around the mean that are visually bound by means of three boxes. More importantly, the perceptual junction of the boxes and the dashes representing outliers forms a compact visual enclosure that enhances the interpretability of data (see Carr, 1994). In this regard, behavioural experiments should be conducted to determine whether the present version facilitates data analysis and interpretation vis-à-vis other versions of the boxplot or even other graphical techniques.

To conclude, it is important that the data reported using the *SB*, or any other graphical method, are accompanied by results of homogeneity and normality tests in order to corroborate visual displays and substantiate the use of parametric tests. Also, given that the computer code that generates the boxplot can be customised, it is advisable that any modification be reported along with the plot when presenting results.

ACKNOWLEDGMENTS

The *SB* was implemented using R programming language. The authors thank Mia Hubert, David Lane, Kimihiro Noguchi, and Kristi Potter for their comments on the content of this paper. The authors also thank Rosalyn Shute and Anastasia Ejova for their comments on the structure of the paper. F.M.-R. thanks Yuka Toyama for her help with the visual arrangement of some of the figures. The code that produces the *SB* can be obtained from the journal's website or either of the authors.

REFERENCES

- Best, L. A., Smith, L. D., & Stubbs, A. (2001). Graph use in psychology and other sciences. *Behavioural Processes*, 54, 155-165.
- Carr, D. B. (1994). A colorful variation on boxplots. *Statistical Computing & Statistical Graphics Newsletter*, 5 (3), 19-23.
- Choonpradub, C., & McNeil, D. (2005). Can the boxplot be improved? *Songklanakarinn Journal of Science and Technology*, 27 (3), 649-657.
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79 (387), 531-554.
- Coles, M., & Davies, C. (1982). On the origins of the .05 level of statistical significance. *American Psychologist*, 37 (5), 553-558.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1 (1), 42-45
- Cox, D. R. (1978). Some remarks on the role in statistics of graphical methods. *Applied Statistics*, 27 (1), 4-9.
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60 (2), 170-180.
- Doane, D. P., & Tracy, R. L. (2000). Using beam and fulcrum displays to explore data. *American Statistician*, 54 (4), 289-290.
- Efron, B. (1987). Better bootstrap confidence intervals [with discussion]. *Journal of the American Statistical Association*, 82 (397), 171-200.

- Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, 104 (2), 293-296.
- Frigge, M., Hoaglin, D. C., & Iglewicz, B. (1989). Some implementations of the boxplot. *The American Statistician*, 43 (1), 50-54.
- Gallistel, C. R. (2009). Importance of proving the null. *Psychological Review*, 116 (2), 439-453.
- Gelman, A., Pasarica, C., & Dodhia, R. (2002). Let's practice what we preach: turning tables into graphs. *The American Statistician*, 56 (2), 121-130.
- Goodall, C. (1983). *M*-estimators of location: An outline of the theory. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 339-403). N.Y.: Wiley.
- Graedel, T. E. (1977). The wind boxplot: An improved wind rose. *Journal of Meteorology*, 16 (4), 448-450.
- Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *American Statistician*, 52 (2), 181-184.
- Hofmann, H., Kafadar, K., & Wickham, H. (2009). *Letter value boxplots*. Manuscript in preparation.
- Hubert, M., & Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis*, 52, 5186-5201.
- Hyndman, R. J., & Fan, Y. (1996) Sample quantiles in statistical packages. *The American Statistician*, 50 (4), 361-365.
- Jarmasz, J., & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology*, 63 (2), 124-138.
- Lane, D. M., & Sándor, A. (2009). Designing better graphs by including distributional information and integrating words, numbers, and images. *Psychological Methods*, 14 (3), 239-257.
- Levine, T. R., Weber, R., Park, H. S., & Hullett, C. R. (2008). A communication researchers' guide to null hypothesis significance testing and alternatives. *Human Communication Research*, 34, 188-209.
- Loftus, G. R. (1993). A picture is worth a thousand *p* values. On the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instruments, & Computers*, 25 (2), 250-256.
- Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyse data. *Current Directions in Psychological Science*, 5 (6), 161-171.
- Loftus, G. R., & Masson, M. J. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, 1 (4), 476-490.
- Marmolejo-Ramos, F., & Matsunaga, M. (2009). Getting the most from your curves: Exploring and reporting data using informative graphical techniques. *Tutorials in Quantitative Methods for Psychology*, 5 (2), 40-50.
- Masson, E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57 (3), 203-220.
- McGill, R., Tukey, J., & Larsen, W. A. (1978). Variations of boxplots. *American Statistician*, 32 (1), 12-16.
- Morey, R. D. (2008). Confidence intervals from normalised data: A correction to Cousineau. *Tutorials in Quantitative Methods for Psychology*, 4 (2), 61-64.
- Potter, K. (2006). Methods for presenting statistical information: The boxplot. In H. Hagen, A. Kerren, and P. Dannemann (Eds.), *Visualisation of large and unstructured data sets, Lecture Notes in Informatics (LNI), vol. S-4* (pp. 97-106). GI-Edition.
- Rosenberg, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 297-338). N.Y.: Wiley.
- Sartori, R. (2006). The bell curve in psychological research and practice: Myth or reality? *Quality & Quantity*, 40, 407-418.
- Schwertman, N. C., Owens, M. A., & Adnah, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics & Data Analysis*, 47, 165-174
- Sim, C. H., Gan, F. F., & Chang, T. C. (2005). Outlier labelling with boxplot procedures. *Journal of the American Statistical Association*, 100 (470), 642-652.
- Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical test. *Psychological Methods*, 6 (4), 371-386.
- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53 (3), 300-314.
- Wood, M. (2005). Bootstrapped confidence intervals as an approach to statistical inference. *Organizational Research Methods*, 8 (4), 454-470.
- Zimmerman, D. W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. *The Journal of Experimental Education*, 67 (1), 55-68