An Improved Boxplot for Univariate Data
Author(s): M. L. Walker, Y. H. Dovoedo, S. Chakraborti and C. W. Hilton
Source: *The American Statistician*, Vol. 72, No. 4 (NOVEMBER 2018), pp. 348-353
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/45116983
Accessed: 03-01-2025 23:59 UTC

THE AMERICAN STATISTICIAN
2018, VOL. 72, NO. 4, 348–353: Statistical Practice
https://doi.org/10.1080/00031305.2018.1448891

Taylor & Francis
Taylor & Francis Group



# An Improved Boxplot for Univariate Data

M. L. Walker[a], Y. H. Dovoedo[b], S. Chakraborti[c], and C. W. Hilton[c]

[a]Department of Mathematics, Pensacola State College, Pensacola, FL; [b]Ally Financial Inc., Charlotte, NC; [c]Department of Information Systems, Statistics, and Management Science, The University of Alabama, Tuscaloosa, AL

**ABSTRACT**
The boxplot is an effective data-visualization tool useful in diverse applications and disciplines. Although more sophisticated graphical methods exist, the boxplot remains relevant due to its simplicity, interpretability, and usefulness, even in the age of big data. This article highlights the origins and developments of the boxplot that is now widely viewed as an industry standard as well as its inherent limitations when dealing with data from skewed distributions, particularly when detecting outliers. The proposed ratio-skewed boxplot is shown to be practical and suitable for outlier labeling across several parametric distributions.

**ARTICLE HISTORY**
Received October 2016
Revised January 2018

**KEYWORDS**
Asymmetric distribution; Bowley's coefficient; Fence constant; Outliers; Skewness

## 1. Introduction

The boxplot, first introduced by Spear (1952) and later formalized by Tukey (1977), is a convenient and useful tool for the data analyst. This article highlights the origins and progression of the boxplot as well as its inherent limitations in outlier labeling when dealing with skewed data. This brief background is necessary for understanding the ultimate aim of the article, which is to present a new modification to the traditional boxplot, the Ratio-Skewed boxplot, for use with any univariate dataset, symmetric or skewed, regardless of sample size. The proposed adjustment to the boxplot fences account for the underlying distribution's skewness, hence improves its ability to label outliers.

The boxplot is widely used today as a quick graphical method to display the data distribution. It is constructed with a lower and an upper fence, respectively, beyond which data points are flagged as potential outliers. These fences are constructed using the sample quartiles and a fence constant, $k$. The lower and upper fences of the traditional boxplot are given by

$$f_L^T = q_1 - k(q_3 - q_1) \text{ and } f_U^T = q_3 + k(q_3 - q_1) \quad (1)$$

where $q_1$ and $q_3$ are the first and the third sample quartile, respectively. This particular fence construction method has evolved over time, through trial and error, from earlier attempts at formalizing the boxplot methodology.

The essential features of what we now know as the boxplot were first introduced by Spear (1952) some 65 years ago, who discussed potential variations of what was then introduced as the range bar chart. The final modification shows an early version of the boxplot later popularized by Tukey, complete with box and whiskers, highlighting the five number summary from a set of data with the whiskers extending to the minimum and maximum values. Twenty years later, Tukey (1972) introduced a modified version of the range bar chart, known then as the schematic plot that could highlight potential outliers in a sample of data. In this early incarnation of the now ubiquitous boxplot, a data point was defined as within the "hinges" (the interquartile range), within the "sides" (the inner fences), and within the "corners" (the outer fences). In this version, Tukey (1972) essentially used $k = 1.0$ as an inner fence constant to highlight potential outliers and $k = 2.0$ as an outer fence constant to highlight extreme outliers, thus using two sets of fences in the boxplot construction. Later, perhaps through repetition and experience, Tukey (1977) altered and finalized these constants to be $k = 1.5$ and $k = 3.0$ for the inner and outer fence constants, respectively, by the time he released his book, *Exploratory Data Analysis* (EDA).

Within a decade, the boxplot had become a widely used tool for EDA and included in several statistical software packages. However, while Tukey (1977) had settled on inner and outer fence constants of $k = 1.5$ and $k = 3.0$, respectively, these were not universally accepted and many software packages used different fence constants as well as slightly differing definitions in finding the quartiles. The question of which fence constant $k$ to use in the boxplot seemed to be more of a subjective choice rather than one based on a mathematical logic, relying on the type of data, the analyst's experience and the needs of the analysis. The aim of Frigge, Hoaglin, and Iglewicz (1989) was to standardize different fence constant selection and methods of boxplot construction for use across statistical software platforms. They settled on the now standard method of using Tukey's (1977) fence constants of $k = 1.5$ and 3.0 as the default values, though the outer fence has largely been abandoned in favor of a single set of fences constructed using $k = 1.5$ to highlight any potential outliers. While this constant is now the norm in the construction of the traditional boxplot, the resulting fences given in Equation (1) are not without limitations as explored in the following sections.

**CONTACT** S. Chakraborti *schakrab@cba.ua.edu* Department of Information Systems, Statistics and Management Science, University of Alabama, P.O. Box 870226, Tuscaloosa, AL 35487.

## 2. Boxplot Fences

The boxplot's usefulness can be hindered by the violation of at least an implicit assumption of symmetry of the underlying distribution. For distributions that are symmetric or nearly symmetric, a traditional boxplot works well. When applying the traditional boxplot construction methodology to the standard normal distribution, the inner fences may be located at approximately ±2.7, leaving an area of 0.0035 in each tail of the distribution. Thus, we would expect that only 0.7% of sample data would be flagged as potential outliers. While there is not one unique formal definition of an outlier, this seems to be a reasonable representation of an extreme data value. Another rule of thumb for outlier identification is the three-sigma rule, which states that outliers are values that are greater than three standard deviations away from the mean of a distribution. When applying this rule to the standard normal distribution, we have an area of 0.00135 in each tail, with the expectation that 0.27% of data in a sample from a standard normal distribution may be expected as extreme values. Had the three-sigma rule been used in the formulation of boxplot outlier rules, the constant to align the boxplot's inner fences at these points would have required a value of $k = 1.7239$. Although not formally stated, Tukey may have had this idea in mind when finalizing the fence constants, possibly settling on $k = 1.5$ (one-half of 3.0) for intuitive reasons, leading to the traditional boxplot lower and upper fences. A recent formal justification of the use of $k = 1.5$ as the fence constant was presented by Dumbgen and Riedwyl (2007), though this value is only justified for symmetric distributions.

However, when applied to data from skewed distributions, the proportion of observations flagged as extreme values by the traditional boxplot tends to increase beyond commonly acceptable levels. For instance, while only 0.7% of the area under the standard normal distribution falls beyond the established fences (±2.7), the same methodology applied to a Chi-square distribution, with one degree of freedom, results in a lower fence at −1.731, that falls below the lower bound of the distribution, and an upper fence at 3.156, that leaves 7.57% of the distribution beyond the upper fence. This suggests that for data from skewed or highly skewed distributions, a traditional boxplot can flag an alarmingly high number of observations as potential outliers, most of which are "false alarms," expected to occur naturally in that distribution. This is the aspect of the traditional boxplot we are concerned with and is the primary focus of this article.

In recognition of this shortcoming, much research has been dedicated in recent years to modifying the boxplot to account for data from skewed distributions, with the earliest notable contribution from Kimber (1990). Data from a right-skewed distribution should naturally have $(q_3 - q_2)$ larger than $(q_2 - q_1)$ and the opposite holds true for left-skewed distributions. This reasoning led to Kimber's (1990) modified fences,

$$f_L^K = q_1 - c[2(q_2 - q_1)] \quad \text{and} \quad f_U^K = q_3 + c[2(q_3 - q_2)] \quad (2)$$

where $q_2 - q_1$ is the lower semi-interquartile range ($SIQR_L$), $q_3 - q_2$ is the upper semi-interquartile range ($SIQR_U$) and c is a constant generally taken to be equal to 1.5. By splitting the interquartile range at the median, and using the two resulting semi-interquartile ranges, the fences are constructed to utilize the skewness observed within the interquartile range to account for overall skewness in the data. Note that if the distribution is symmetric, these fences become identical to those in the traditional boxplot.

While the modification in Equation (2) improves the performance of the traditional boxplot for outlier labeling with skewed distributions, the improvement is only slight. For example, applying this fence construction procedure to the Chi-square distribution with one degree of freedom, the lower fence is found to be located at $x = -0.959$, still slightly below the range of the distribution, and the upper fence is located at $x = 3.928$, leaving 4.75% of the distribution beyond the upper fence. This is a marginal improvement but it is still partly unacceptable. The idea of splitting the interquartile range to account for skewness has been the basis for much of the ensuing research on the subject, including the modification proposed by Dumbgen and Riedwyl (2007) mentioned earlier. Further research into refining the boxplot fences for use with skewed distributions, mainly using Kimber's (1990) ideas as a basis, includes modifications that make assumptions on the underlying distribution, rely on additional variables that need to be estimated, or tend to be specific to certain data types. Of note are modifications proposed by Barnett and Cohen (2000) using distributional assumptions of lifetime data, Carling (2000), using variable fence constants, Schwertman and de Silva (2006) using sequential fence constants, Hubert and Vandervieren (2008) using the medcouple (*MC*, a robust measure of skewness introduced by Brys Hubert and Struyf (2004)), and Dovoedo and Chakraborti (2012, 2014), using fence constants obtained by controlling the probability that at least one sample observation is wrongly classified as an outlier. While these and other recent modifications may perform well in certain cases, many are not generalizable or may be too complex to be implemented widely in practice, as the greatest value of the boxplot is in its use as a simple general-purpose tool for EDA.

## 3. Improved Boxplot Fences for Univariate Data

The idea of splitting the interquartile range at the median by Kimber (1990) plays an important role in our proposal for the boxplot fences for skewed distributions. Skewness in a distribution is expected to produce differing values for the lower and upper parts of the interquartile range. Recall that the interquartile splits, referred to as the semi-interquartile ranges (SIQR), are denoted as

$$\text{SIQR}_L = (q_2 - q_1) \quad \text{and} \quad \text{SIQR}_U = (q_3 - q_2) \quad (3)$$

For data from a symmetric distribution, the values of $\text{SIQR}_\text{L}$ and $\text{SIQR}_\text{U}$ will be roughly equal. However for data from a right-skewed distribution, the $\text{SIQR}_\text{U}$ will typically be greater than the $\text{SIQR}_\text{L}$, and conversely for data from a left-skewed distribution. While Kimber (1990) proposed replacing the IQR with $2(\text{SIQR}_\text{L})$ in creating the lower fence and $2(\text{SIQR}_\text{U})$ in creating the upper fence, respectively, this adjustment only slightly accounts for the underlying skewness. We build on this idea and consider new boxplot fences that are more receptive to skewness and yet work as the traditional fences when the underlying distribution is symmetric.

One way to do this is to introduce a sample quartile-based measure of skewness in the fence construction, which is a function of the upper and the lower semi-interquartile ranges. It turns out there is in fact such a measure introduced by Bowley (1920), known as Bowley's Coefficient. This measure makes use of the sample quartiles to give an indication of skewness, given by

$$B_c = \frac{q_3 + q_1 - 2q_2}{q_3 - q_1} = \frac{\text{SIQR}_U - \text{SIQR}_L}{\text{IQR}} = \frac{\text{SIQR}_U - \text{SIQR}_L}{\text{SIQR}_U + \text{SIQR}_L} \tag{4}$$

Note that $B_C$ is centered at zero, takes on values between $-1$ and 1, where negative and positive values indicate negative and positive skewness, respectively. For sample data from symmetric distributions, the two semi-interquartile ranges should be nearly equal, in which case $B_C$ will be nearly equal to zero.

If Bowley's coefficient is recentered at one, it could be used as a multiplier in the formation of the boxplot fences to adjust skewness. This is consistent with our goal is to introduce an adjustment factor on the fences that equals one for symmetric distributions. To this end, note that the values $1 + B_C$ and $1 - B_C$ will be close to 1 for data from a symmetric distribution. For data from a right-skewed distribution, $B_C$ will be positive, so that $1 + B_C$ will be greater than 1 and $1 - B_C$ will be less than 1. Alternatively, for data from a left-skewed distribution, $B_C$ will be negative, so that $1 + B_C$ will be less than 1 and $1 - B_C$ will be greater than 1. Thus $1 + B_C$ and $1 - B_C$ can be applied as skewness-adjustment factors in the fence construction, $1 + B_C$ to the upper fence and $1 - B_C$ to the lower fence, respectively. Hence, adjusted boxplot fences can be given by

$$f_L^{\text{adj}} = q_1 - 1.5\text{IQR}(1 - B_c) \text{ and } f_U^{\text{adj}} = q_3 + 1.5\text{IQR}(1 + B_c) \tag{5}$$

Since $1 - B_c = 2\text{SIQR}_L/\text{IQR}$ and $1 + B_c = 2\text{SIQR}_U/\text{IQR}$, it follows that this fence adjustment is identical to that proposed by Kimber (1990) shown in Equation (2).

The skewness observed within the interquartile range should not be considered as absolute, but relative. For data from a right-skewed distribution, the $\text{SIQR}_U$ will be larger than $\text{SIQR}_L$, with the opposite holding true for data from a left-skewed distribution. In this sense, for data from a right-skewed distribution, the ratio of $\text{SIQR}_U$ to $\text{SIQR}_L$ would be a better measure of skewness. Thus, using Tukey's (1977) traditional boxplot fences with Kimber's (1990) idea as a foundation, we chose to work with these ratios to adjust the upper and lower fences, in order to account for any underlying skewness. The proposed modified fences are given by

$$f_L^{\text{RS}} = q_1 - 1.5\text{IQR}\left(\frac{\text{SIQR}_L}{\text{SIQR}_U}\right) \quad \text{and} \quad f_U^{\text{RS}} = q_3 + 1.5\text{IQR}\left(\frac{\text{SIQR}_U}{\text{SIQR}_L}\right) \tag{6}$$

respectively. The ratios $\frac{\text{SIQR}_L}{\text{SIQR}_U}$ and $\frac{\text{SIQR}_U}{\text{SIQR}_L}$ can be seen to be measures of skewness themselves, centered around 1. For data from a symmetric distribution, these two ratios should be nearly equal, each equal to 1, leading to the adjustment factors both nearly equal to 1, which then yields the equations for the traditional fences. Likewise, differences in the splits will increase or decrease the ratios slightly above or below one, extending or retracting the fences in order to control unnecessary flagging of regular observations as outliers, or false positives (alarms).

The proposed lower and upper fence skewness adjustment factors can be written more compactly in terms of Bowley's Coefficient as follows:

$$R_L = \frac{1 - B_c}{1 + B_c} \quad \text{and} \quad R_U = \frac{1 + B_c}{1 - B_c}. \tag{7}$$

Redefining the ratios in this way, the proposed boxplot fences given in Equation (6) can be equivalently expressed as

$$f_L^{\text{RS}} = q_1 - 1.5\text{IQR}\,R_L \quad \text{and} \quad f_U^{\text{RS}} = q_3 + 1.5\text{IQR}\,R_U \tag{8}$$

This equivalent form of the proposed skewness-adjusted fences further shows that they incorporate a measure of skewness into the upper and lower fences of the boxplot. To emphasize, for data from a symmetric distribution, the value of Bowley's coefficient, $B_C$ should be close to 0, leading to these adjustment factors to be nearly equal to 1, in which case the proposed fences will revert back to the traditional fences.

More interesting properties of these adjustment factors become apparent while examining their sum and differences. It can be shown that the sum of these ratios has a minimum value of 2, for symmetric distributions, a value that increases as the distribution becomes more asymmetric, essentially providing another measure of the underlying skewness. Likewise, the difference of these ratios can also be seen as a measure of skewness, which can be shown to be related to Bowley's coefficient with an additional multiplier, given by

$$R_U - R_L = (R_U + R_L + 2)B_c. \tag{9}$$

Thus, the difference of the ratios yields Bowley's coefficient for Skewness multiplied by the sum of the ratios shifted by 2. As the sum of the ratios has a minimum value of 2, the multiplying factor on the right-hand side of Equation (9) has a minimum value of 4. Since Bowley's coefficient is centered at 0, this multiplying factor affects the spread; so the difference of the ratios itself is a measure of skewness centered at 0 with a spread of at least four times that of Bowley's coefficient, meaning that it is more sensitive to deviations from symmetry.

As an example, a random sample of $n = 99$ observations was generated from the chi-square distribution, with one degree of freedom. The boxplots in Figure 1 illustrate the differences in fence creation and labeling of potential outliers among the Tukey, Kimber, and the proposed ratio-skewed boxplots. The traditional Tukey boxplot flags a large number of outliers just

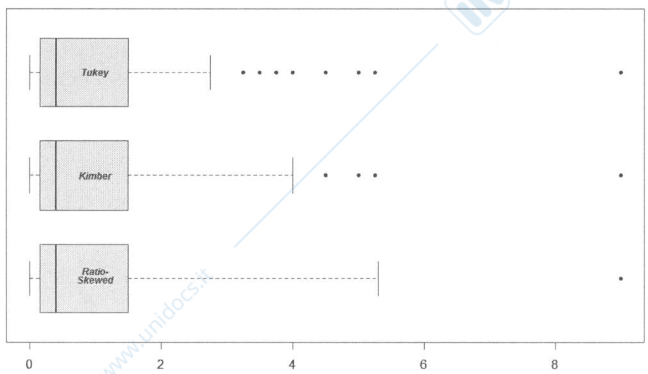


**Figure 1.** A boxplot comparison of the Tukey, Kimber, and ratio-skewed methods for data from a chi-square (1) distribution.

beyond the upper fence, which, for this right-skewed distribution, should by no means be considered outliers. Kimber's (1990) adjustment moves the upper fence farther out to account for right-skewness, yet still leaves a number of data points collectively just outside the upper fence. The proposed ratio-skewed modification, on the other hand, moves the upper fence even farther due to the relative skewness witnessed within the interquartile range, leaving just one (extreme) data point as a potential outlier. With a sample of $n = 99$, it seems that leaving one data point as a potential outlier (incorrectly) could reasonably be expected regardless of the underlying and often unknown distribution.

## 4. Simulation Study

In order to study the efficacy of the proposed ratio-skewed modification, we compare it to the Tukey (1977) and the Kimber (1990) methods, for both uncontaminated and contaminated data, in a simulation study. The results are discussed below.

### 4.1. Case of Uncontaminated Data

Four families of distributions are included in the study of uncontaminated data (data without any outliers). These include the Normal, the chi-square, the Gamma, and the *F* distribution. For each family of skewed distributions, we consider five varying levels of skewness. Datasets of sizes $n = 10$, 20, 30, 50, 100, and 150 are generated from each distribution and 10,000 simulations are run. The mean proportions of observations falsely (incorrectly) flagged as outliers are calculated. A subset of the results is displayed in Figure 2.

For the most skewed distribution, namely the chi-square(1) distribution (with a skewness coefficient of 2.828), the effectiveness of the ratio-skewed method at identifying potential outliers is clearly demonstrated, regardless of the number of observations generated. This better performance of the ratio-skewed procedure is seen across all moderate to highly skewed distributions and smaller size samples (in a larger study not shown here). However, when the distribution exhibits low to no skewness, the Tukey and Kimber methods tend to outperform the ratio-skewed method. Interestingly, the mean proportion of observations falsely flagged is seen to be negatively related to the number of observations used in the analysis, regardless of the underlying skewness. The order of preference among the three methods, in terms of outlier labeling, are Tukey, Kimber, and ratio-skewed when skewness and sample size are low. As the sample size grows, all three procedures perform comparably well under lower skewness.

**Table 1.** Distributions, skewness coefficients, and contaminating outliers.

| Distribution | Skewness | Outliers 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Normal (0,1) | 0.000 | 3.72 | 3.77 | 3.82 | 3.87 | 3.92 |
| Chi-square (4) | 1.414 | 23.51 | 23.56 | 23.61 | 23.66 | 23.71 |
| Chi-square (15) | 0.730 | 44.26 | 44.31 | 44.36 | 44.41 | 44.46 |
| Gamma (5, 1) | 0.894 | 17.78 | 17.83 | 17.88 | 17.93 | 17.98 |
| Gamma (2, 1) | 1.414 | 11.76 | 11.81 | 11.86 | 11.91 | 11.96 |
| F(30, 60) | 0.900 | 3.09 | 3.14 | 3.19 | 3.24 | 3.29 |
| F(20, 25) | 1.466 | 5.07 | 5.12 | 5.17 | 5.22 | 5.27 |

### 4.2. Case of Contaminated Data

Next, we study the case of contaminated data, where outliers are introduced in the data generated in the simulation study. The distributions, the associated skewness coefficients and the outliers used as contaminators in the study are shown in Table 1. In each case, we generated $n = 50$, $n = 100$, and $n = 150$ observations from each distribution. The smallest value $b$ an outlier is allowed to take is chosen in such a way that the probability that an observation from the underlying population falls beyond $b$ is very small (approximately 0.001 in our simulations). For

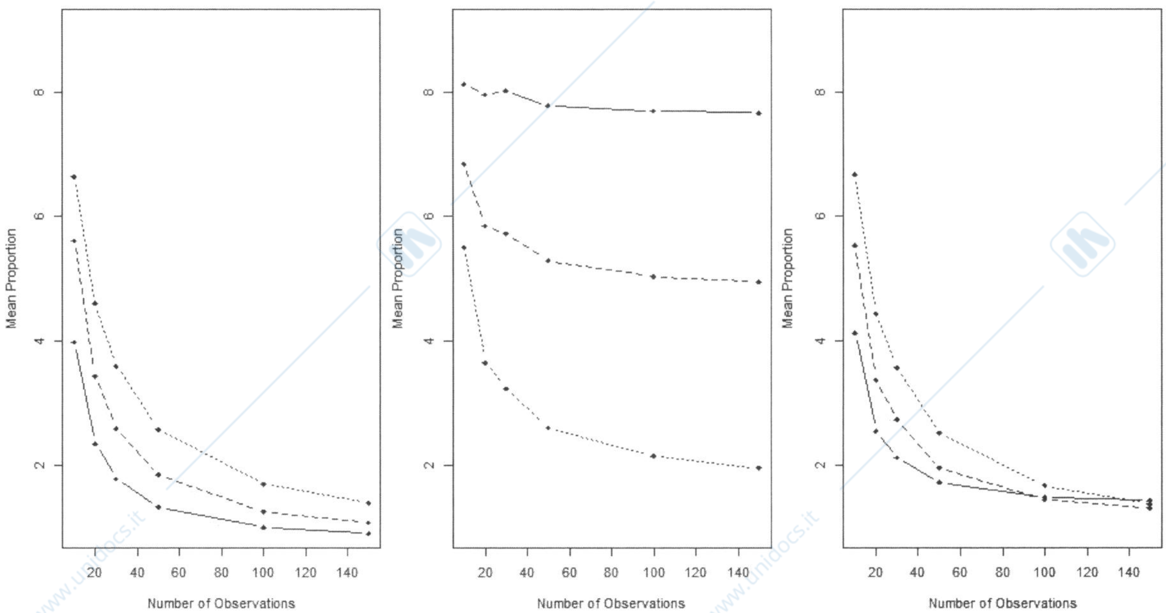


**Figure 2.** Mean proportion of observations falsely flagged for the Tukey method (solid line), the Kimber method (dashed line), and the ratio-skewed method (the dotted line) against the number of observations for three cases, Normal (0,1), chi-square (1), and chi-Square (25) distribution, respectively, in panels one to three from left to right.

$n = 50$, we introduced up to three of the listed potential outliers (the first one, the first two, or all three). For $n = 100$ and $n = 150$, we introduced up to five of the listed potential outliers (the first one, the first two, the first three, the first four, or all five). The outliers are simulated from the normal distribution with mean equal to the 0.999 quantile of the relevant distribution (such as Chi-square) and a standard deviation of 0.1. This provides the desired number of outliers to test the robustness of each boxplot method in the simulation study. Again, 10,000 simulations are run, and the mean proportion of outliers detected (POD) as well as the mean proportion of regular observations detected as outliers (PRDO) are calculated.

Figure 3 presents a visualization of a subset of the results, for two distributions, Normal (0, 1) and chi-square (15), with $n$ equal to 100, for an illustration. Note that our findings are also supported from the results for the larger study, for each distribution exhibiting right skewness, which we summarize here. All three methods have comparable POD values for the chi-square (15) case. In cases where the distributions have low to no skewness, the POD values for the ratio-skewed method tend to be lower than those of the Kimber and Tukey methods. The PRDO values for the ratio-skewed method under the Normal (0,1) distribution in Figure 3 are higher than those for the Tukey and Kimber methods. In general, it is seen (not reported here) that for less skewed distributions, such as the chi-square (15), the Gamma (5,1) and the $F$ (30,60), the ratio-skewed method has higher PRDO values than the Tukey and Kimber methods, but for more right-skewed distributions, such as the chi-square (4), the Gamma (2, 1), and the $F$ (20, 25), the ratio-skewed method has lower PRDO values regardless of the sample size. In general, for moderate to highly right-skewed distributions, the proposed ratio-skewed method is seen to be preferable to the Tukey and Kimber methods.

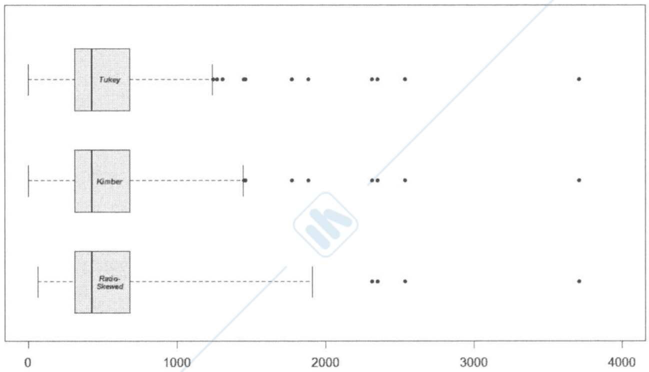


**Figure 4.** A comparison of the three boxplot methods using the North American Rivers dataset.

## 5. Illustrations

In addition to studying the performance of the ratio-skewed modification of the boxplot with simulated data, we also applied it to a well-known dataset available in the literature. The methods of Kimber (1990) and Tukey (1977) are also applied for comparison purposes.

### 5.1. Rivers Dataset

We use the dataset, "rivers" (McNeil 1977), readily available in the R package, "datasets," which gives the lengths in miles of 141 major North American rivers. It exhibits positive skewness (3.15) with several recognizable outliers. The Ratio-Skewed method, the Kimber method, and the Tukey method were applied to these data. Figure 4 shows the boxplots and the potential outliers identified as dots above the upper fences.

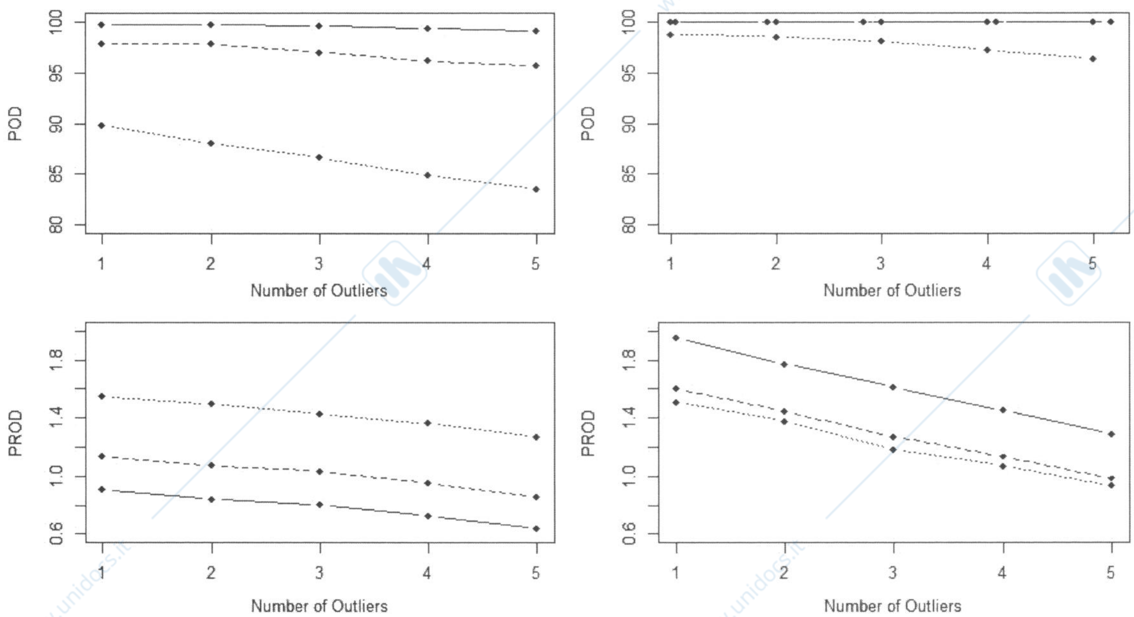


**Figure 3.** The POD and PRDO values against the number of outliers for the Normal (0, 1) and chi-square (15) distribution, respectively, in the left and the right panels. The solid line represents the Tukey method, the dashed line the Kimber method and the dotted line ratio-skewed method.

The number of observations labeled as potential outliers are 4, 11, and 8 for the ratio-skewed, the Tukey (1977), and the Kimber (1990) method, respectively. However, note that 3 of the outliers labeled by the Kimber method and 6 by the Tukey method lie within the right-fence of the proposed Ratio-Skewed method. This demonstrates the differences between the performance of the three methods and the potential relative usefulness of the Ratio-Skewed method when applied to positively skewed data.

## 6. Conclusion

In the 41 years since Tukey's (1977) *Exploratory Data Analysis*, the boxplot has emerged as a simple yet effective data visualization tool, ubiquitous in various fields of study. While there are more sophisticated graphical methods today, the boxplot remains relevant due to its simplicity, ease of interpretation, and relative effectiveness. However, the traditional boxplot is not without its limitations, particularly its ability to effectively label potential outliers in data from skewed distributions may be questionable. This issue has been studied for many years, resulting in various modifications to the traditional boxplot. While many of these complex proposals have definite applications when dealing with specific types and sizes of data due to the underlying assumptions, attempts at finding a modification of the boxplot for general use seem to have increasingly lost touch with Tukey's original intentions of simplicity and ease of interpretation in boxplots and EDA in general.

The proposed ratio-skewed modification presented in this paper is a method to improve the boxplot with fences that work effectively regardless of the underlying skewness or sample size. Through simulations and a practical application, this modification has been shown to work well over a wide variety of distributions, particularly right-skewed distributions. By retaining the basic structure of the traditional fence construction method, the addition of the SIQR ratios to the fences in order to measure and adjust for underlying skewness, is the key to the proposed ratio-skewed method that can be easily implemented in practice.


## Acknowledgments

The authors are grateful for the interest and the detailed comments of the editor, the associate editor, and two anonymous referees that contributed to an improved presentation. Assistance from the editorial office is also acknowledged.


## References


Barnett, O., and Cohen, A. (2000), "The Histogram and Boxplot for the Display of Lifetime Data," *Journal of Computational and Graphical Statistics*, 9, 759–78. [349]

Bowley, A. L. (1920), *Elements of Statistics* (4th ed), New York: Charles Scribner's Sons. [350]

Brys, G., Hubert, M., and Struyf, A. (2004), "A Robust Measure of Skewness," *Journal of Computational and Graphical Statistics*, 13, 996–1017. [349]

Carling, K. (2000), "Resistant Outlier Rules and the Non-Gaussian Case," *Computational Statistics and Data Analysis*, 33, 249–258. [349]

Dovoedo, Y. H., and Chakraborti, S. (2012), "Boxplot-Based Phase I Control Charts for Time Between Events," *Quality and Reliability Engineering International*, 28, 123–130. [349]

——— (2014), "Boxplot-Based Outlier Detection for the Location-Scale Family," *Communications in Statistics – Simulation and Computation*. [349]

Dumbgen, L., and Riedwyl, H. (2007), "On Fences and Asymmetry in Box-and-Whiskers Plots," *The American Statistician*, 61, 356–359. [349]

Frigge, M., Hoaglin, D. C., and Iglewicz, B. (1989), "Some Implementation of the Boxplot," *The American Statistician*, 43, 50–54. [348]

Hubert, M., and Vandervieren, E. (2008), "An Adjusted Boxplot for Skewed Distributions," *Computational Statistics and Data Analysis*, 52, 5186–5201. [349]

Kimber, A. C. (1990), "Exploratory Data Analysis for Possibly Censored Data from Skewed Distributions," *Applied Statistics*, 39, 21–30. [349,350,351,352,353]

McNeil, D. R. (1977), *Interactive Data Analysis*, New York: Wiley. [352]

Schwertman, N. C., and de Silva, R. (2006), "Identifying Outliers With Sequential Fences," *Computational Statistics and Data Analysis*, 51, 3800–3810. [349]

Spear, M. E. (1952), *Charting Statistics*, New York: McGraw-Hill. [348]

Tukey, J. W. (1972), "Some Graphic and Semigraphic Displays," *Statistical Papers in Honor of George W. Snedecor*, ed. T. A. Bancroft, Ames, IA: The Iowa State University Press, pp. 293–316. [348]

——— (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley. [348,350,351,352,353]