

# Cartesian displays of many interval estimates\*

**Mario Peruggia**

*Department of Statistics  
The Ohio State University  
Columbus, Ohio*  
e-mail: [peruggia@stat.osu.edu](mailto:peruggia@stat.osu.edu)

**Jason Hsu**

*Department of Statistics  
The Ohio State University  
Columbus, Ohio*  
and  
*Eli Lilly and Company  
Indianapolis, Indiana*  
e-mail: [jch@stat.osu.edu](mailto:jch@stat.osu.edu)

and

**Yifan Huang**

*Pfizer, Inc.  
Groton, Connecticut*  
e-mail: [yifan.huang@pfizer.com](mailto:yifan.huang@pfizer.com)

**Abstract:** We consider the problem of constructing static graphical representations of a large number of interval estimates. Because of clutter, traditional graphical summaries are visually ineffective for representing more than a few intervals. The Cartesian displays introduced in this article overcome the limitations stemming from visual clutter and can represent effectively very many intervals. The construction of a Cartesian display for symmetric intervals is first presented in the context of a multiple comparisons application. Generalizations involving the representation of asymmetric intervals are then introduced and used to summarize aspects of the posterior distributions of numerous parameter contrasts in two hierarchical Bayes models.

**Keywords and phrases:** Hierarchical Bayes models, midpoint-radius representation, multiple comparisons.

Received March 2012.

## 1. Introduction

In many statistical applications inferential conclusions are summarized numerically by interval estimates. Sometimes intervals may be formed by combining

---

\*This material is based upon work supported by the National Science Foundation under Grants No. SES-0214574, SES-0437251, and SES-1024709.

point estimates of location with corresponding measures of accuracy. Perhaps most commonly this happens when sample means and standard errors are computed for several groups of observations. In other circumstances, the intervals may arise from direct calculation of their end-points, as is the case for confidence intervals constructed by inverting a test procedure. In this article we address the issue of constructing *static* graphical representations of very many interval estimates. The goal is to offer a synoptic representation of the intervals that is more readily amenable to making comparisons and drawing meaningful conclusions than a simple listing of their numerical values.

Typical graphical displays represent intervals estimates as line segments or some other two-dimensional entity. Examples from the multiple comparison literature are the line-by-line display implemented by numerous software packages, the *mean-mean scatterplot* of Hsu and Peruggia (1994) and the *comparison circles* of Sall (1992). These types of representations tend to become cluttered and break down visually when the number of entities to be plotted is larger than about ten. From a practical perspective, the principal reason why these displays do not scale up to represent hundreds or thousands of intervals lies in the size of the graphical entities that are being plotted (intervals, circles, etc.).

We illustrate how the smallest possible graphical entities, *points*, can be used to represent intervals visually. The basic premise underlying the representations that we introduce is that any interval can be summarized by two numbers (in infinitely many ways, in fact). Once a well defined rule for transforming an interval into a unique pair of numbers has been established, the elements of the pair can be regarded as the coordinates of a point to be plotted in the Cartesian plane. The shift of focus from elements of Euclidean geometry (the segments) to elements of analytic geometry (the corresponding points in the Cartesian plane) yields static graphical representations with distinctive features that are effective for handling large numbers of intervals.

The rest of the article is organized as follows. In Section 2 we identify three salient features that should be captured by any graphical representation of interval estimates. Within the context of multiple comparisons applied to a well known data set on automobile fuel efficiency, we review how the popular line-by-line display summarizes the three salient features and we identify some shortcomings. We describe Cartesian displays in Section 3 (for symmetric intervals) and in Section 4 (for asymmetric intervals). We motivate the basic display in Sections 3.1 and illustrate its construction in Section 3.2 using the same fuel efficiency data set. This example gives us the opportunity to examine the pros and cons of the display and to compare them to those of the line-by-line display in a situation with a limited number of intervals. Then, in Sections 4.1 and 4.2, we give two substantive examples centered on Bayesian hierarchical modeling of examination scores and of response times, illustrating the great potential benefits that accrue from the use of Cartesian displays in situations involving very many intervals. In Section 5 we summarize the benefits of Cartesian displays, discuss their limitations, and outline some additional applications and possible improvements.

## 2. Summarizing salient features of interval estimates with a traditional display: The fuel consumption example

We illustrate the basic ideas in the context of a multiple comparison application to the Fuel data from the data frame `fuel.frame` contained in the R library `SemiPar`. These are data on 117 makes of cars published in the April 1990 issue of Consumer Reports. The factor `Type` classifies makes of cars into six general categories: `Small`, `Sporty`, `Compact`, `Medium`, `Large`, and `Van`. To analyze how the response variable `Fuel` (which represents the gallons of fuel consumed by each make of car to travel 100 miles) is affected by the factor `Type`, we fit a one-way anova model and computed Tukey-Kramer 95% simultaneous confidence intervals for all pairwise comparisons using the R function `TukeyHSD()`. Because the factor `Type` has six levels, this yields  $(6 \times 5)/2 = 15$  non-redundant intervals, where, for a given comparison, we call an interval redundant if the interval for the negative of that comparison has already been constructed.

The top panel of Figure 1 exhibits the line-by-line display of the 15 simultaneous confidence intervals produced by R. In this particular application the intervals to be displayed are symmetric. For each pairwise comparison, there are three salient features to be conveyed:

- (a) The *position* of the interval. Because in this example the intervals are symmetric, the center is a natural summary, but any other well identified point, such as the lower bound of the interval, would do as well.
- (b) The *extent* of the interval. This could be summarized by the width of the interval, or, for the case of a symmetric interval, by its half-width.
- (c) The *identification* of the interval. In this example, appropriate labeling should enable one to identify the specific car categories involved in the comparison.

Quite generally, any effective graphical representation of interval estimates should convey these three features. How they are conveyed and the relative importance given to them will affect the eventual look of the display. In the line-by-line display of Figure 1, position and extent can both be read off the horizontal axis. The vertical axis is reserved for identification and the intervals are listed in lexicographical order. The latter choice emphasizes ease of look-up for specific comparisons, but the visual appearance of the display depends on the choice of labels. For example, the look of the display is not invariant with respect to translation of the labels into another language. In addition, comparisons involving “Compact” vehicles against all other types are easier to make, say, than comparisons involving “Van.” This is because the intervals for comparisons against “Compact” are all grouped together, while those for comparisons against “Van” are scattered throughout the display and, as noted in Cleveland and McGill (1984), objects close together can be compared more easily. This difficulty could be overcome by displaying the set of intervals for the redundant comparisons, at the expense of taking up twice as much vertical space, or of doubling the density of displayed intervals, or of some combination thereof.

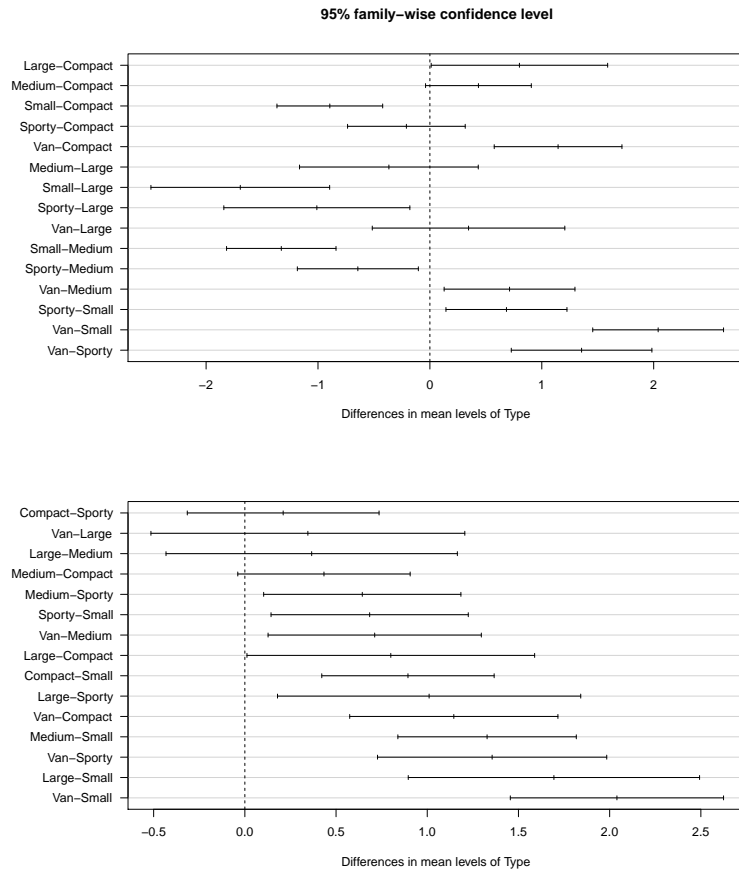


FIG 1. *Line-by-line display of the 15 simultaneous confidence intervals for the fuel consumption data. The intervals in the top panel are ordered lexicographically and those in the bottom panel are ordered according to the values of the point estimates of the contrasts.*

This example illustrates some of the consequences of using the vertical axis for identification. On the plus side, the lexicographical ordering offers ease of look-up of specific comparisons, especially when the total number of comparisons is small. On the minus side, the visual appearance of the display is not invariant with respect to the choice of labeling. However, if the goal is to show ordering with respect to some specific inferential feature rather than to facilitate look-up, it is possible to conjure up other meaningful arrangements of the intervals that are invariant with respect to relabeling. One way to do so, illustrated in the bottom panel of Figure 1, is to consider the 15 contrasts with positive estimated values and order the corresponding intervals according to the size of such values. This display originates from a hybrid use of the vertical axis, in which the intervals are arranged according to the order statistics of the summary

measure used to capture location. The identification labels must still be read off the vertical axis and the actual values of location and extent must still be read off the horizontal axis. The ordering has the advantage of making certain characteristics of the data more obvious. For example, the four non-significant comparisons are more easily identified in the bottom panel of Figure 1.

### 3. Cartesian displays for symmetric intervals

With regard to the features identified in Section 2, the Cartesian displays that we introduce in this article reserve the horizontal and vertical axes for the representation of location and extent. In general, these displays are based on the specification of a well-defined rule to map an interval into a point. For example, as shown in Sections 3.2, symmetric multiple comparison intervals can be summarized by their centers and half-widths in what we call *location-spread displays*. Alternative rules for mapping non-symmetric intervals into points are illustrated in Section 4.

#### 3.1. Kulpa's midpoint-radius representation

The location-spread display was suggested to us by the midpoint-radius representation introduced in Kulpa (2003) to describe interval arithmetic. A summary review is contained in Hayes (2003) and the basic definition is as follows.

**Definition 3.1.** The *midpoint-radius representation* is the function that maps a compact interval  $[a, b]$  into the point of coordinates  $((a + b)/2, (b - a)/2)$ .

The first coordinate (the midpoint) is the center of the interval and the second coordinate is its half-width (the radius). Figure 2 illustrates the mapping and can be used as a guide to understand its properties. For the statistical applications that we consider the following properties are most relevant.

- An interval  $[a, b]$  is mapped into the point of intersection between the ray of slope +1 emanating from the point  $(a, 0)$  and the ray of slope  $-1$  emanating from the point  $(b, 0)$ .

Once the point corresponding to an interval is given, the previous property yields a straightforward means of determining if a given reference value  $r$  is contained in the interval. In fact:

- All points included in the closure of the quadrant of vertex  $(r, 0)$  and delimited by the two diagonal rays of slopes +1 and  $-1$  correspond to intervals that contain  $r$ . All remaining points correspond to intervals that do not contain  $r$ .

Often, to determine statistical significance, a reference value of interest is  $r = 0$ . Then, to assess if the interval corresponding to a given point contains or does not contain  $r = 0$ , it is enough to ascertain the location of the point relative to the quadrant defined by the main diagonal rays emanating from  $(0, 0)$ . For

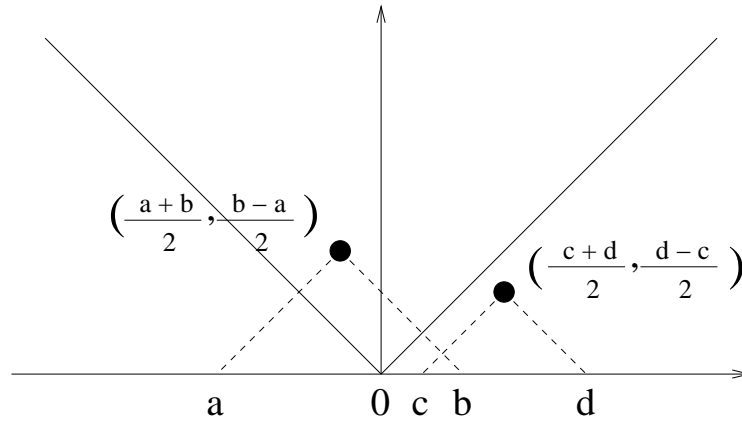


FIG 2. Illustration of the midpoint-radius mapping. A segment is mapped into the vertex opposite the base of an isosceles triangle with 45 degree base angles. The base of the triangle is given by the segment itself.

example, in Figure 2, the point  $((a+b)/2, (b-a)/2)$  lies inside the quadrant and its corresponding interval  $[a, b]$  contains  $r = 0$  while the point  $((c+d)/2, (d-c)/2)$  lies outside the quadrant and its corresponding interval  $[c, d]$  does not contain  $r = 0$ . Reference values  $r \neq 0$  can also be of interest, for example in practical equivalence applications.

### 3.2. Location-spread displays: The fuel consumption example revisited

Figure 3 exhibits the location-spread display of the simultaneous confidence intervals for the Fuel consumption data. (Sample R code for generating color versions of the display can be downloaded from the first author's web page ([www.stat.osu.edu/~peruggia/papers/lsp.txt](http://www.stat.osu.edu/~peruggia/papers/lsp.txt)). The display is constructed by plotting the center of each interval, summarizing location, along the horizontal axis and the corresponding half-width, summarizing extent, along the vertical axis and is thus the graph of the images of the 15 simultaneous confidence intervals under the midpoint-radius mapping. The points in Figure 3 are all plotted to the right of the vertical line through the origin because they correspond to the intervals for the same 15 pairwise contrasts with positive estimated values displayed in the bottom panel of Figure 1. In this case the left part of the figure could be omitted, but we decided to keep it to illustrate the general look of a location-spread display.

We used shading and other graphical symbols to emphasize the properties outlined above. In particular, the *reference* quadrant delimited by the two main diagonals through the origin is shaded as in Figure 4 of Hayes (2003). The four points falling inside the reference quadrant represent non-significant differences in mean fuel consumption according to car type (the corresponding

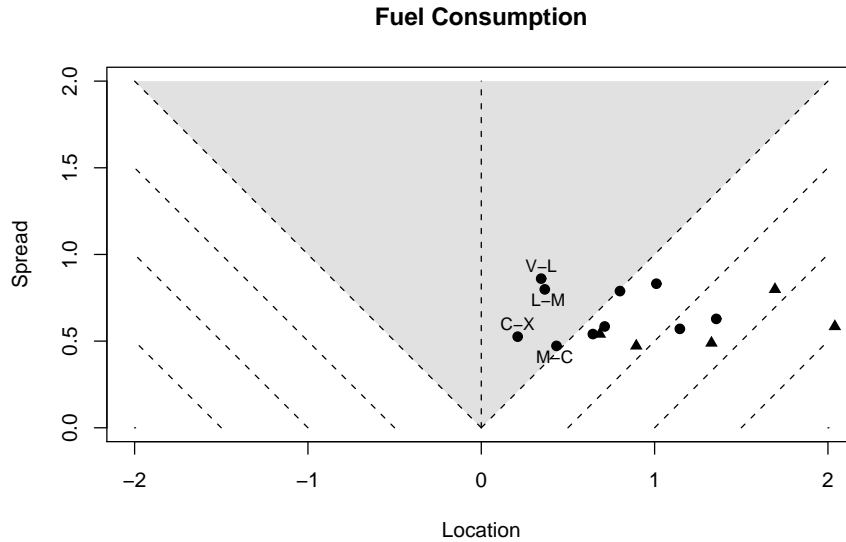


FIG 3. The location-spread display for the fuel consumption data. For each point in the display the horizontal dimension is the center of a Tukey-Kramer simultaneous confidence interval and the vertical dimension is the half-width of the interval. The comparisons involving **Small** cars are represented by the triangular symbols.

intervals contain zero). All remaining points lie outside the quadrant and represent significant differences. Comparisons with reference values other than zero is facilitated by the drawing of equally spaced rays of slope 1 to the right of the origin and of slope  $-1$  to the left of the origin. (In a color display, each ray can be color-coded based on how far from the origin the ray intersects the horizontal axis.) Following the rays in Figure 3 we can see, for example, that only one interval lies entirely to the right of the reference value  $r = 1$ .

Having reserved the horizontal and vertical axes for representing location and extent, we are left with the task of identification. One possible approach is direct labeling. For example, in Figure 3, we used direct labeling to identify the four non-significant comparisons. To reduce clutter we shortened the labels, denoting the six vehicle categories **Small**, **Sporty**, **Compact**, **Medium**, **Large**, and **Van** by **S**, **X**, **C**, **M**, **L**, and **V**, respectively. Still, direct labeling of all plotted points would render the graph too hard to decipher and only interesting subsets of points should be labeled. Many software packages provide tools for doing this interactively. Different plotting symbols (and/or color) can also be used effectively. For example, if one wishes to emphasize one or more subsets of comparisons, a subset-specific symbol can be used to plot the points corresponding to the comparisons in each subset. This is done in Figure 3, where all comparisons involving car type **Small** are represented by a triangle.

As a consequence of the fact that well-defined summaries of location and extent are mapped into the horizontal and vertical axes, the location-spread repre-

sentation is invariant with respect to relabeling. In addition, because points can be drawn as smaller graphical entities than segments, the display can represent a large number of interval estimates and still remain intelligible. Situations such as those illustrated in the remainder of the article, in which very large numbers of interval estimates are involved, make it clear that Cartesian displays can be much more effective than traditional displays like the line-by-line plots of Figures 1.

#### 4. Cartesian displays for asymmetric intervals

In this section we present two Cartesian displays that are variants of the location-spread display and are useful to represent intervals that are not symmetric about a point estimate.

##### 4.1. *M-M2Q displays: A school comparison survey*

We consider a large number of comparisons concerning inner London schools based on a Bayesian hierarchical model developed in Spiegelhalter et al. (1996) using a subset of the data originally analyzed in Goldstein et al. (1993). The response variable is an examination achievement score, averaged over study subjects, collected on 1978 pupils attending 38 different schools. Spiegelhalter et al. (1996) use both pupil level and school level covariates to specify a normal linear regression model for the examination achievement scores. For each school the regression model contains a random intercept term,  $\alpha_j$ ,  $j = 1, \dots, 38$ , ( $\alpha_{1,j}$  in the original notation) that the authors regard as the residual school effect after adjusting for the covariates.

Based on a set of  $M$  MCMC draws from the posterior distributions of the school specific intercepts, Spiegelhalter et al. (1996) construct estimates of the posterior marginal distribution of each school's ranking, which they summarize by a point estimate and a 95% credible interval. The school specific intercepts can also be used to estimate contrasts of interest involving specific schools. For the purpose of illustration we consider estimation of the posterior distributions of all pairwise contrasts  $\alpha_j - \alpha_k$ ,  $j \neq k$ ,  $j = 1, \dots, 38$ ,  $k = 1, \dots, 38$ . The empirical distribution of each contrast can be simply constructed by combining the differences  $\alpha_j^{(m)} - \alpha_k^{(m)}$ ,  $m = 1, \dots, M$ , based on the output at iteration  $m$  of the MCMC algorithm. This produces a total of  $38 \times 37 = 1,406$  non-degenerate empirical distributions that we would like to summarize graphically. There is, of course, redundancy in this set of distributions because  $\alpha_j^{(m)} - \alpha_k^{(m)} = -(\alpha_k^{(m)} - \alpha_j^{(m)})$ , but eliminating this redundancy would still leave 703 distributions, a very large number given that we wish to present a graphical summary of all of them at once.

We construct a global summary of interesting aspects of all distributions by means of a Cartesian display as follows. First, for each contrast  $\alpha_j - \alpha_k$ , we compute the 0.025 quantile,  $q_{j,k}(0.025)$ , the median,  $q_{j,k}(0.5)$ , and the 0.975 quantile,  $q_{j,k}(0.975)$ , of the corresponding empirical distribution. We want to



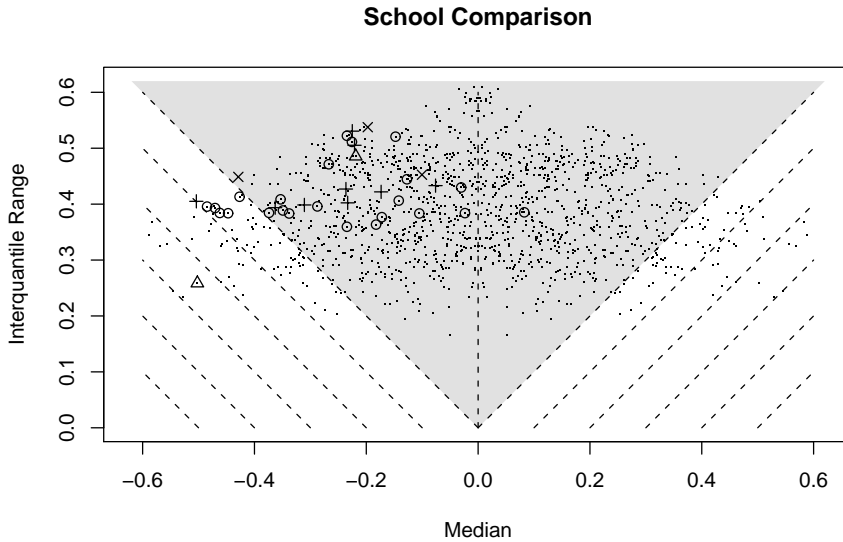


FIG 4. The *M-M2Q* display for the school comparison data. For each point in the display the horizontal dimension is the median of the posterior distribution of a pairwise contrast and the vertical dimension is either the median minus the 0.025 quantile or the 0.975 quantile minus the median, depending on whether the point is to the right or to the left of the vertical ray through the origin. The contrasts involving School 17 are plotted using different symbols depending to the denomination of the other school entering the comparison.

be able to determine from the Cartesian display which 95% credible intervals  $[q_{j,k}(0.025), q_{j,k}(0.975)]$  lie entirely to the left or to the right of zero (with a slight abuse of terminology, we will call such intervals *significant*). We use the horizontal axis to capture location and the vertical axis to capture extent. Precisely, if the median is positive, we summarize the credible interval by the position-extent pair given by  $(P_{j,k} = q_{j,k}(0.5), E_{j,k} = q_{j,k}(0.5) - q_{j,k}(0.025))$  (because the median is positive, the interval is entirely to the right of zero if and only if  $E_{j,k} < P_{j,k}$ ). If the median is negative, we summarize the credible interval by the position-extent pair given by  $(P_{j,k} = q_{j,k}(0.5), E_{j,k} = q_{j,k}(0.975) - q_{j,k}(0.5))$  (because the median is negative, the interval is entirely to the left of zero if and only if  $E_{j,k} < -P_{j,k}$ ). The scatterplot of the position-extent pairs thus determined is then constructed, yielding what we call the *M-M2Q display* shown in Figure 4. (“M-M2Q” stands for “Median-Median to Quantile”.)

There are a few features of Figure 4 worth mentioning. First, the chosen summary measures of position and extent do not identify uniquely the credible intervals  $[q_{j,k}(0.025), q_{j,k}(0.975)]$ , because the intervals need not be symmetric about  $q_{j,k}(0.5)$ . Yet, as explained in the previous paragraph, if a point in the M-M2Q display lies outside of the shaded reference quadrant, then the credible interval does not contain zero, and vice versa. Thus, this choice of  $(P_{j,k}, E_{j,k})$  pairs accomplishes the important goals of conveying direct visual information

about the point estimate of a given contrast (the median value  $P_{j,k}$ ) and about whether or not the contrast is significant.

Figure 4 represents summaries of all 1,406 non-degenerate contrasts (including the redundant ones). To reduce the blob-like effect of overplotting we used a very small plotting symbol. The foremost visual message conveyed by the figure is that the vast majority of points lie inside the reference quadrant. This is an indication that most differences between schools are of little importance, confirming the statement in Goldstein et al. (1993) that “few schools can be separated reliably.”

Simple graphical devices allow us to use the M-M2Q display to concentrate easily on contrasts involving specific schools. In Figure 4 we employed larger plotting characters to emphasize all pairwise comparisons of the type  $\alpha_{17} - \alpha_k$ ,  $k \neq 17$ ,  $k = 1, \dots, 38$ , involving School 17. School 17 is the school that attains the second to last posterior median ranking using the approach of Spiegelhalter et al. (1996), with only School 5 attaining a worse posterior median ranking. This is reflected in the fact that all but one of the median estimates for the contrasts under consideration are negative and, consequently, all but one of the larger symbols in the display are plotted to the left of the vertical line through the origin. However, only a small number of these contrasts have a posterior distribution that is concentrated away from zero, as can be seen from the fact that most of the points fall inside the reference quadrant.

In addition, specific plotting characters are used to code the school denomination of the other school entering the comparison (triangle for **Church of England**, plus sign for **Roman Catholic**, circle for **State school**, and times sign for **other**). From this plotting character coding, some inferences can be readily made at a visual level. For example, it is immediate to notice that School 17 (a **Church of England** school) is not significantly different than any of the schools in the other category. Also, interestingly, the most prominent difference is the one with another **Church of England** school, corresponding to the triangle toward the left hand side of the display.

#### 4.2. Lower and upper bound displays

In this section we illustrate the use of Cartesian displays to represent interval estimates computed in the analysis of a large set of response time data. The data were collected on four subjects over ten non-consecutive days in a series of recognition memory trials designed to measure how long it would take the subjects to react to certain stimuli. The details of the experiment, the Bayesian hierarchical model fit to the data, and various inferential issues are described in Peruggia (2007) and Craigmile, Peruggia and Van Zandt (2012). For the purpose of this discussion it suffices to know that, on any given day, each subject was presented with two lists of 40 stimuli. The Bayesian hierarchical model assumes a Weibull likelihood for the (shifted) response times, with parameters related to the subjects and to specific experimental conditions. In particular, the model contains 80 shape parameters  $r_{i,d,l}$ , where  $i$ ,  $1 \leq i \leq 4$ , indexes subject,  $d$ ,  $1 \leq d \leq 10$ , indexes day, and  $l$ ,  $1 \leq l \leq 2$ , indexes list.

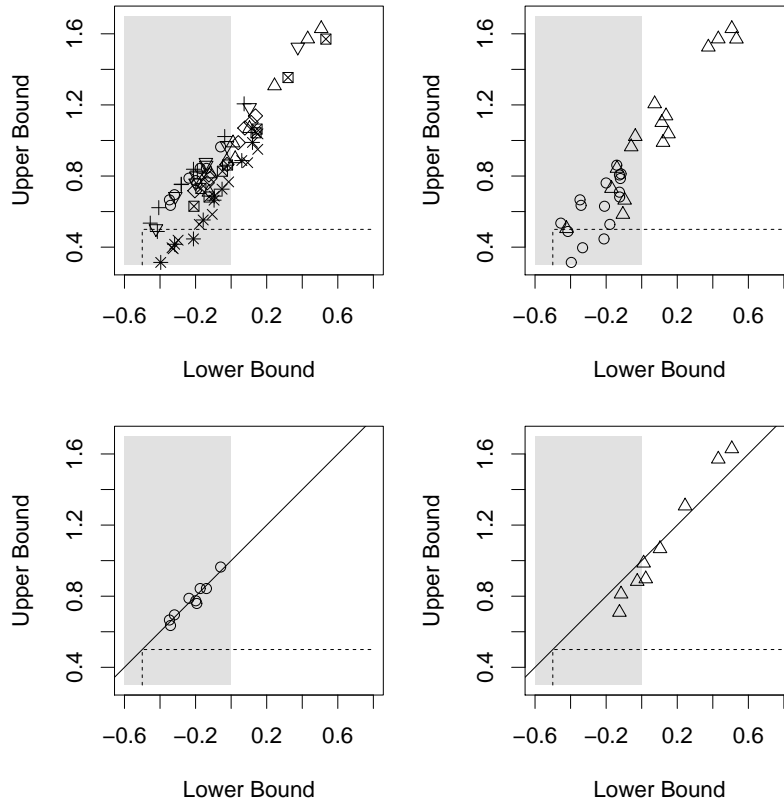


FIG 5. The lower and upper bound display for the shape parameter contrasts in the response time data example. The top left panel represents all 72 contrasts, labeled according to the subject  $\times$  list combination. The top right panel represents contrasts for earlier days (circles) and later days (triangles). The bottom panels represent all list 1 contrasts for subject 1 (left) and subject 2 (right).

In these types of experiments, as subjects become more accustomed to the tasks they are required to perform, their response times tend to become shorter and more regular. So, it is interesting to determine if the posterior distributions of the model parameters provide any evidence of that. Here we focus on the shape parameters of the Weibull distributions and consider all  $4 \times 9 \times 2 = 72$  contrasts of the type  $\Delta_{i,d,l} = r_{i,d,l} - r_{i,1,l}$ , measuring, for a given subject and list, the departure of the shape parameter on day  $d$  from the shape parameter on day 1.

In the top-left panel of Figure 5 we display the equal-tailed 95% posterior probability intervals for the 72 contrasts by plotting the upper limit  $U_{i,d,l}$  of each interval against its lower limit  $L_{i,d,l}$ . We call this Cartesian plot a *lower and upper bound display*. Compared to the M-M2Q display of Section 4.1, there is now a unique correspondence between the points in the plot and the corre-

sponding intervals, but the point estimates of the contrasts are not displayed. This representation uses the horizontal and vertical axes in concert to summarize location and extent. Identification is accomplished by using a different plotting character for each subject  $\times$  list combination.

The figure outlines several characteristics of the contrasts. First, quite a few points lie outside the shaded region representing the portion of the NW reference quadrant of vertex  $(0, 0)$  contained in the plotting region. This fact indicates that the corresponding intervals do not contain zero and underscores the presence of important differences in the values of the shape parameters entering the contrasts. Second, the dashed lines delineate a quadrant with vertex  $(-0.5, 0.5)$ . We are interested in points falling inside this quadrant because the interval  $(-0.5, 0.5)$  covers a small (arbitrary) region centered at zero and, if the 95% probability interval for a contrast is entirely contained in this small interval, the two shape parameters involved in the contrast can reasonably be regarded as essentially equivalent. (In this respect, this is an illustration of the application of the methodology to a “practical equivalence,” rather than a “significance” problem.) On this basis we can conclude that only a handful of shape parameters are essentially equivalent and that most of these equivalent parameters are for the same subject  $\times$  list combination denoted by an asterisk.

In the top-right panel of Figure 5 we display the equal-tailed 95% posterior probability intervals for the 16 contrasts between the shape parameters for day 1 and days 2 and 3 (circles) and the 16 contrasts between the shape parameters for day 1 and days 9 and 10 (triangles). Because of the relative locations of the triangles and the circles, it is apparent that the differences in shape parameters from baseline tend to become larger as days go by.

The lower left panel depicts all list 1 contrasts for subject 1 and the lower right panel depicts all list 1 contrasts for subject 2. Both panels contain the slanted reference line of equation  $UB = 1 + LB$  in upper bound vs. lower bound space. Points falling below, on, or above the line correspond to intervals of width smaller, equal, or larger than 1, respectively. The two panels have noticeably different characteristics. All intervals for subject 1 are located closely and have roughly the same width. The intervals for subject 2 exhibit more scatter in their locations and have a tendency to become wider as their lower bounds increase.

## 5. Discussion

In this article we introduced a class of graphical tools, the Cartesian displays, that can be used to represent statically a very large number of interval estimates and we illustrated their effectiveness with three applications. There are, of course, many other situations in which these displays might prove useful. Gene expression analysis and other applications in bioinformatics often involve a large number of inferences that one wishes to represent graphically. In drug discovery, for example, thousands or tens of thousands of molecular compounds may be screened for activities against cancer cell lines. We have also found these displays very useful for illustrating to medical researchers the impact of

the common practice of dichotomizing a continuous response and fitting a logistic regression model for the derived 0-1 variable. For example, in this case, a location-spread display can be constructed by dichotomizing the response at several different levels and plotting selected estimated coefficients against the half-widths of their corresponding confidence intervals. By doing so, it often becomes apparent that the coefficient for a given term might be significant at one cutoff level and non-significant at a different level. The display can also illustrate clearly that, in many cases, accuracy decreases as changes in the cutpoint render the grouping of the observations more unbalanced. In particular, if the rate of increase of the standard errors is roughly linear in the size of the estimates, it is easy to show that the points in the display will tend to fall along straight lines.

Quite naturally, the method has some limitations. As previously mentioned, direct labeling of the points is required for proper identification and this can often be accomplished effectively through the use of different colors and plotting characters. If many of the point estimates on which the intervals are based share a common estimated accuracy (which often happens for balanced designs), then the displays will present unappealing horizontal or diagonal streaks. Clearly, overplotting may become an issue when thousands of points are plotted. Reducing the size of the plotting character and increasing the size of the display is often helpful, as is jittering. However, more specialized plotting techniques, such as binning (Carr, 1991), will be required to deal with very large sets of location-spread pairs.

In conclusion, we observe that Cartesian displays are convenient graphical summary devices, with several features that make them powerful exploratory data analysis tools. The main factor motivating our development of these displays was the need for a static graphical display of many interval estimates. As noted before, the physical size of the plotting area may at times impede direct labeling of all the points. In such situations it may be possible to divide the points into meaningful subgroups, perhaps based on the values of some available covariates, create multiple Cartesian displays, and plot them in a trellis-like arrangement.

For those cases in which no static version results in a satisfactory display, we can envision useful interactive implementations. The simplest interactive version can be implemented in R by starting with an unlabeled display and using the point-and-click function `identify()` to generate the labels of individual points of interest. A direct generalization of the `identify()` function would allow the user to click on a point to trigger the display of the label and additional summary information concerning the associated interval that cannot be readily inferred from the display. For the case of the M-M2Q displays of Section 4.1, for example, the additional information could include the numerical values of the median, quantiles and endpoints of the asymmetric interval. More sophisticated implementations should include devices that can provide real-time visual feedback to the user. For displays with considerable overplotting, a zooming feature may be helpful. Also, a brushing tool and dynamic linking to a data table could help the user to visualize interesting groups of points. A menu driven interface could then let the user label the identified points interactively for creation of a

final display for a report. As another example, when constructing the M-M2Q display of Section 4.1 a reference value other than zero could be of interest, in which case one could simply subtract off such a value from the median and the two quantiles corresponding to each credible interval and redraw the M-M2Q display using the translated values. An interactive version of this procedure could be implemented, where a slider is used to adjust the desired reference value and the display is redrawn on the fly.

## References

- CARR, D. B. (1991). Looking at large data sets using binned data plots. In *Computing and Graphics in Statistics* (A. Buja and P. Tukey, eds.) 7–39. New-York: Springer-Verlag.
- CLEVELAND, W. S. and MCGILL, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association* **79** 531–554. [MR0770274](#)
- CRAIGMILE, P. F., PERUGGIA, M. and VAN ZANDT, T. (2012). A Bayesian Hierarchical Model for Response Time Data Providing Evidence for Criteria Changes Over Time. In *Current Issues in the Theory and Application of Latent Variable Models* (M. Edwards and R. MacCallum, eds.) 42–61. New York: Routledge.
- GOLDSTEIN, H., RASBASH, J., YANG, M., WOODHOUSE, G., PAN, H., NUTTALL, D. and THOMAS, S. (1993). A multilevel analysis of school examination results. *Oxford Review of Education* **19** 425–433.
- HAYES, B. (2003). A Lucid Interval. *American Scientist* **91** 484–488.
- HSU, J. C. and PERUGGIA, M. (1994). Graphical Representations of Tukey’s Multiple Comparison Method. *Journal of Computational and Graphical Statistics* **3** 143–161.
- KULPA, Z. (2003). Diagrammatic Analysis of Interval Linear Equations. *Reliable Computing* **9** 1–20. [MR1952126](#)
- PERUGGIA, M. (2007). Bayesian Model Diagnostics Based on Artificial Autoregressive Errors. *Bayesian Analysis* **2** 817–842. [MR2361976](#)
- SALL, J. (1992). Graphical comparison of means. *Statistical Computing and Statistical Graphics Newsletter* **3(1)** 27–32.
- SPIEGELHALTER, J., THOMAS, A., BEST, N. G. and GILKS, W. R. (1996). *BUGS Examples Volume 2, Version 0.5, (version ii)*. Cambridge, UK: MRC Biostatistics Unit.