Elemental Graphics for Analysis of Variance using the R Package granova

Robert M. Pruzek
Department of Educational and Counseling Psychology
1400 Washington Ave.
State University of New York at Albany
Albany, N.Y. 12222
E-mail: rmpruzek@yahoo.com
http://rmpruzek.com

James E. Helmreich
Department of Mathematics
3399 North Rd.
Marist College
Poughkeepsie, NY 12601
E-mail: James.Helmreich@Marist.edu
http://foxweb.marist.edu/users/james.helmreich

November 28, 2010

Abstract

We illustrate the general concept of what we describe as *elemental graphics*: displays that afford direct visualizations of data with respect to fundamental questions that drive particular statistical methods. Attention is directed to methods that have both descriptive and inferential features through a focus on what can be seen as *elemental graphics* for analysis of variance. Our graphics are partly pedagogical in nature, highlighting the fundamental comparisons and outcomes made in the conduct of an ANOVA; however, their use readily extends well beyond pedagogy. These graphics can be helpful in the analysis of nearly any data set for which the corresponding method is applicable. The graphics were created using our graphical ANOVA R package granova.[1]

1 Introduction

To be most informative in practice, statistical methods should help users see relationships, hypothesized or discovered, among sets of variates or between groups with respect to one or more variates. Particular methods are usually defined using specific models, or by relational questions. Some procedures pose explicit questions that can be approached both descriptively and through

[1]This paper grew out of notes prepared for a talk given by R. Pruzek at the 2008 meeting of the Society of Multivariate Experimental Psychology in Montreal, Ontario, Canada.

inference. We provide examples and discussion of what we describe as *elemental graphics* for a variety of methods supported by the `R` package `granova`.

An elemental graphic facilitates *direct visualization* of data in a way that illuminates the questions that a particular statistical method seeks to answer. For example, a simple scatterplot may be considered the canonical example of an elemental graphic for the association of two variables, and by incorporating superposition of a regression line, especially if augmented with vertical lines (from the points to the regression line) that show 'errors', such a plot can become an elemental graphic for a linear prediction. Here we consider elemental graphics for analysis of variance (ANOVA) methods.

It is interesting to note that despite being on the scene for more than three quarters of a century, analysis of variance continues to be central to many applied statistical sciences, and some modern statisticians continue to advocate for use of ANOVA methodology (Gelman [3]). In this paper we examine four classes of ANOVA methods. We provide examples of elemental graphics for each of the following methods or models: one and two-way ANOVA; contrast-based analyses of group differences and dependent (or paired) sample comparisons. We make use of four functions in our `R` [10] package `granova` [8] (for graphical ANOVA); just as for all `R` functions, `granova` is freely available and documented in `R`.

A key feature of elemental graphics is that beyond their obvious value in displaying data their use generally provides details showing how data points and related summaries play out in the context of specific questions that drive particular analytic methods. Frequently, either virtues and limitations of particular methods will be highlighted in some form by elemental graphics. When methods are closely related (e.g., those based on ranking, and their parametric counterparts) it can, as we shall demonstrate, be especially helpful to investigate the details about how such methods work when used to analyze the "same" data.

For any particular statistical analysis of data, an elemental graphic will not be unique. A set of elemental graphics for a particular combination of method and data are likely to differ at least in details. Elemental graphics can differ in how they represent data with respect to questions central to a method, and also particular features of given data. Tastes of the graphic-analyst, with respect to the intended audience can also influence displays, such as when different colors, symbols and/or line-types are used; more generally there are many ways to use ink on a page or on a screen. Differences in details can make differences in appearance and in effectiveness discernible, perhaps for specific audiences, or for data sets of various sizes and complexities. Furthermore, the tools, or software itself, available to implement particular graphics will generally play a role in graphic construction. A given static graphic depicting the criterion for the choice of regression line might be different from one implemented via an interactive applet. For a wide variety of examples, some of which may be described as elemental graphics with respect to particular statistical models, see Graph Gallery; note especially the examples showing association plots and various lattice plots, where all examples include the relevant `R` source code.

In general, elemental graphics can deepen one's understanding of both data and methods, since the same basic method can be passed over a variety of data sets, and different methods can be passed over the same data. Variations on this theme can be particularly helpful for students. Put another way, elemental graphics, when they are feasible, provide a means for users to *see an analysis*, such as a one- or two-way ANOVA, using data for which they may have notable experience or interest.

Compared with conventional non-elemental graphics, those that seem not to have any direct correspondence with particular (descriptive and inferential) questions that drive statistical methods, elemental graphics can facilitate more useful or informative evaluations of methods, or data, in relation to one another. This is chiefly because elemental graphics provide more incisive information about how particular data points play out in analyses. There appear to us to be relatively few extant statistical graphics that are reasonably described as elemental, so there may be numerous opportunities for further development of elemental graphics. We have found it interesting to query whether elemental graphics exist or have been developed for particular methods. But we recognize

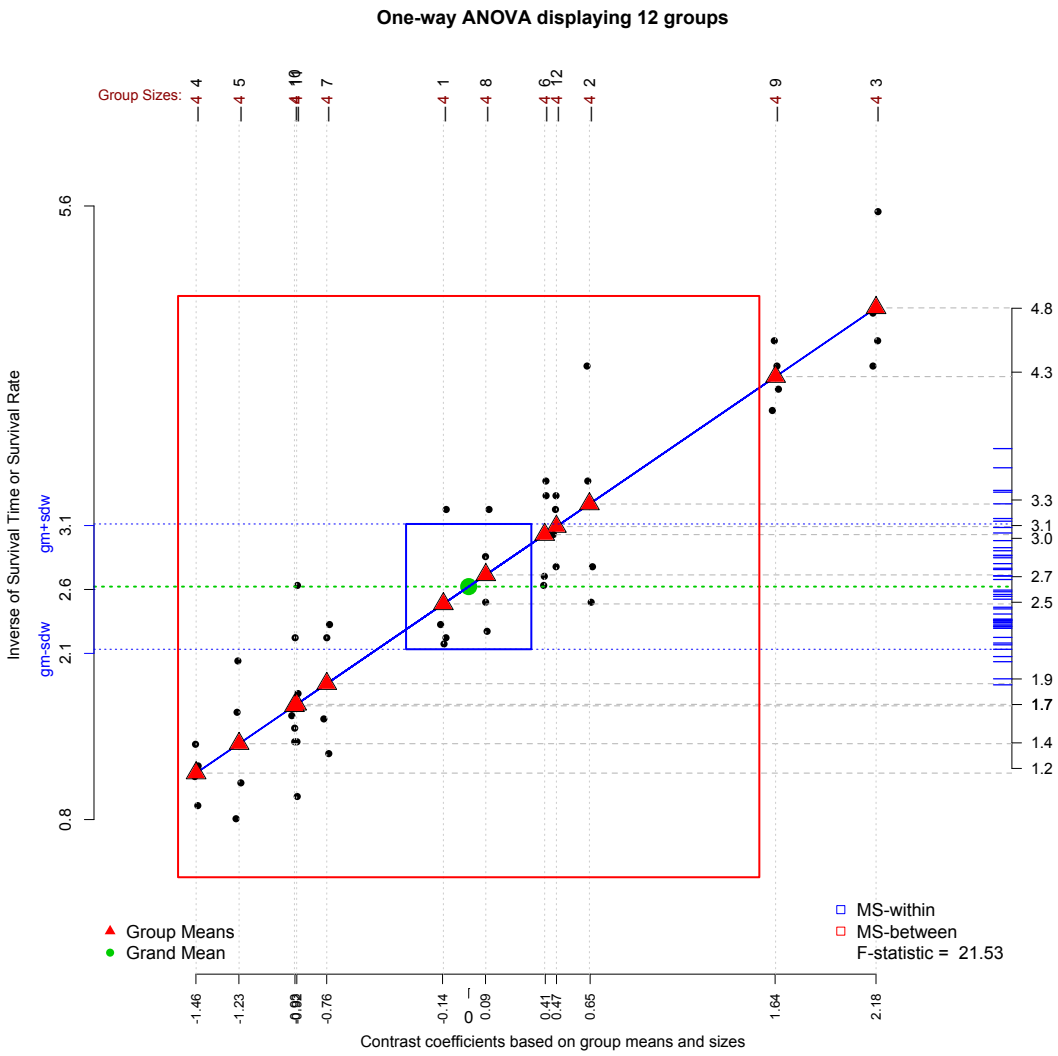


Figure 1: Inverse survival times for 12 treatments from Box, Hunter & Hunter [2]. Note that the variances across groups are similar for this metric.

that especially for relatively complex methods of analysis, or high dimensional data, that it may be difficult – at least with existing software – to construct elemental graphics.

2 A Graphic for One-way ANOVA

The (omnibus) F statistic at the heart of any inferential application of one-way ANOVA implies a particular way to compare means, one relying on data-based contrasts. The elemental graphic we have developed is based on this particular comparison of means. An example is shown in Figure 1, where it seen that the elemental graphic is essentially a specialized scatterplot; this figure was generated using the function `granova.1w` in `R`. The baseline (horizontal axis) is labeled in terms of contrast coefficients c_j for the J groups, ordered from smallest to largest, each of which is a group mean minus the grand mean: $c_j = M_j - M$, $j = 1 \ldots J$. Scores y_{ij} and group means M_j for the groups are printed on the vertical axis. In this case, $J = 12$ and all group sizes are the same, $n = 4$. Symbols representing the group means (red triangles) necessarily fall on a straight line since contrast coefficients and group means differ only by a constant. Score values y_{ij} for each group are plotted with slight horizontal jittering in vertical columns aligned with the contrast coefficients.

(For the functions `granova.1w`, `granova.2w` and `granova.contr` the amount of jittering can be controlled by the user.) Numerical values for the individual group means M_j are shown on the right side of the graphic, as are the full set of residuals (the aggregation of $y_{ij} - M_j$ differences), using a rug plot. The grand mean M and (nominal) range of all scores are shown on the left side; group sizes n_j are provided along the top of the plot, as are group identification symbols.

The mean square error within groups (MSW) and the mean square error between groups (MSB) are readily visualized in terms of areas of squares. Both squares are centered on the grand mean (large green dot). The blue square corresponds to the MSW. Its side is based on the standard deviation of residuals (pooled within group mean deviations), shown in a rug plot at the right margin. Given the statistic MSB (= SSB/(df-between)), where SSB $= n \sum (M_j - M)^2 = n \sum c_j M_j$, SSB can be viewed as a linear combination of the group means where the coefficients are the c_j's. By taking its square root, the MSB determines the length of the side of the red square.[2]

In other words, this graphic depicts individual observations plotted in columns whose positions are determined by the c_j so that both within and between group variation can be visualized and compared. Standard summary statistics are represented as the areas of the blue and red squares. Finally the F statistic can be seen as the ratio of the area of the red to the blue square. In this case MSB > MSW, so F > 1. The F-statistic is printed as a legend at the lower right. (If variation between groups is sufficiently large in relation to within group variance, the red square may be too large to appear in the plot.)

The graphic generated by `granova.1w` provides all basic information associated with a one-way ANOVA; it accommodates any number of groups, which may be of varying sizes. (The analyst can inspect the code to see how the computation is done.) In some cases, especially when there are many groups, group means, and consequently some c_j, may be either identical or nearly so. This is true for groups 10 and 11 shown in Figure 1. In such cases it may be difficult to distinguish the corresponding columns. When this happens one may wish to modify input data slightly by adding or subtracting a small value from all scores in designated groups; this can improve the graphic and generally results in only trivial changes in summary statistics.

2.1 Details for a Specific Example: Poison Data for 12 Unstructured Groups

The data shown in Figure 1 were presented initially in Box and Cox [1], then discussed in detail by Box, Hunter and Hunter [2], hereafter called BHH. The scores are survival times in units of 10 hours; the data are discussed by BHH using both a one-way and two-way ANOVAs. (These data are reexamined again in the next section.) The data were collected for three poisons, each subject to four treatments, in a balanced full-factorial design, with 4 replications for each cell. Central to an understanding of these data is that survival times themselves are far less satisfactory for ANOVA applications than the reciprocal of these times because as BHH make clear the assumption of constant population variances that underpins use of the F-statistic (21.53) is far more reasonable after this non-linear transformation.[3] Finally, numeric output from the function provides the standard one-way ANOVA table, a table with group statistics ordered according to the contrast coefficients in the graphic, weighted and 20% trimmed means, and finally, variances and standard deviations.

If survival times (or reciprocals) are converted to ranks, we can proceed with a rank-based version of one-way ANOVA, commonly described as non-parametric ANOVA. Figure 2 shows such an analysis using the transformed version of ranks given in our poison data set. In this case scores

[2]For technical reasons, both squares have sides based on twice these standard deviations, so the areas of each square are four times the corresponding MS's; of course the multiplier has no effect on the ratio of these areas.

[3]The reader is encouraged to reanalyze these data which may be found in package `granova`, as the dataset `poison`; advantages of using this version of the data include the fact that two alternative transformations of the initial survival times, as well as contrast coefficients (see below) are provided.

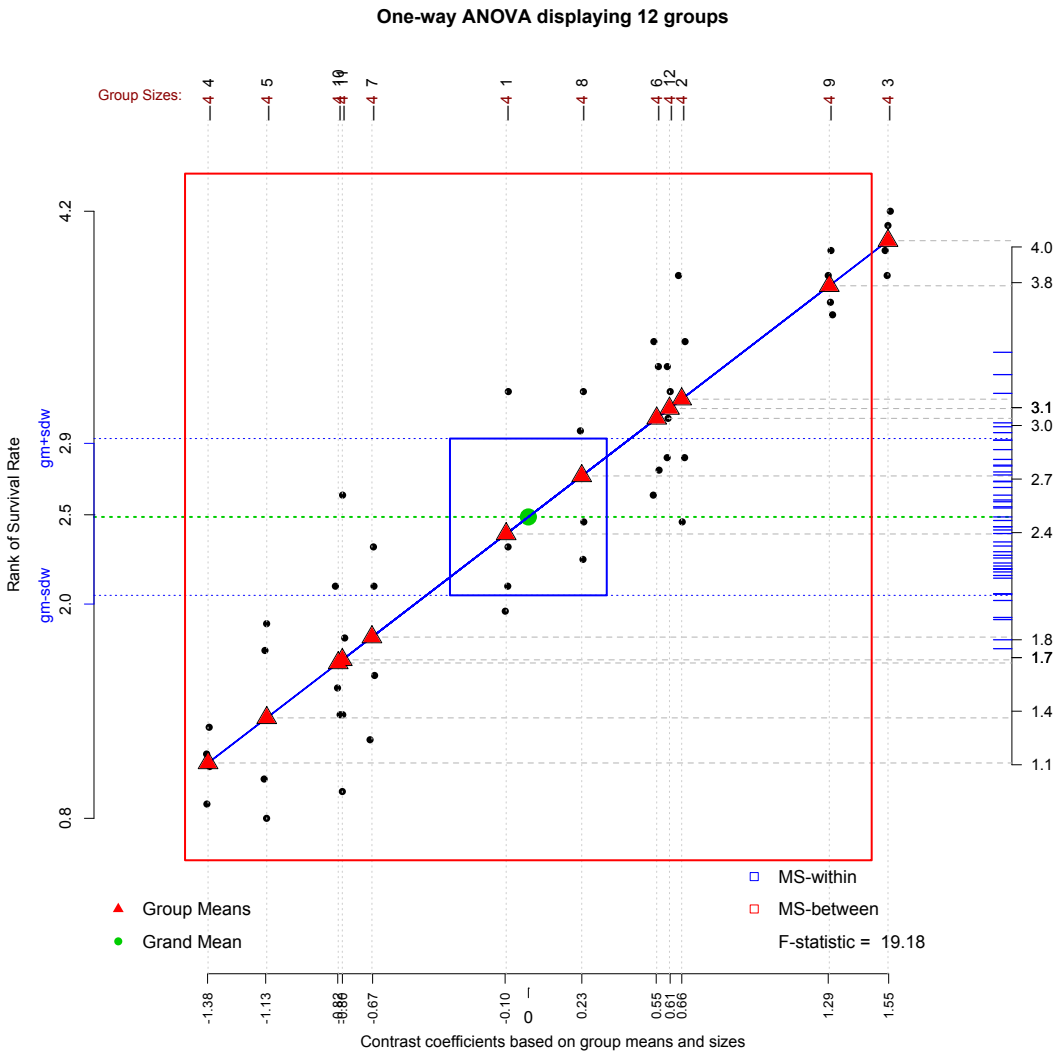


Figure 2: Inverse Survival Time Ranks for 12 treatments from Box, Hunter & Hunter [2]. Note again that variances across groups are similar for this metric.

have been converted to ranks, and also rescaled to have the same median as the original data (*i.e.*, 1/SurvTimes), and a spread comparable to that of the original data. (Because this constitutes a linear transformation of the ranks it has no effect on the rank-based test statistic, but by making the metrics compatible, helps in the graphic to show similarities between the preceding analysis and the one based on ranks.) The standard Kruskal-Wallis test is based on ranks; it asks whether the mean ranks differ statistically across groups and it entails computing a χ^2 statistic. For these data, $\chi^2 = 40.2$, with a p-value of $3.32 * 10^{-5}$. However, the graphic shows more. Note that the ordering of the group means based on ranks conforms exactly to the order based on reciprocals of survival times. This is because the highest scores have the lowest ranks, and vice versa – and since the ranks have been linearly rescaled, the ordering holds. In this example there are no identifiable outliers for the rank data, however that is in general possible even though most authors who discuss rank-based ANOVA do not make this point.

The function `granova.1w` provides various summary statistics. We reproduce these and the specific `R` command for each graphic in Appendix A 7.

In the following section we use the same data as given in BHH (see Table 8) for our illustration of two-way ANOVA. While basic graphic information will be presented, the reader is advised to consider use of the `R` software directly to generate the graphics because the key graphic is dynamic:

it consists of a three-dimensional data display the position of which is controlled by the mouse.

3 A Graphic for Two-Way ANOVA

Moving to two-way ANOVA requires a reconceptualization, one that is informed by consideration of contrasts as shown in Figure 1. In particular, account must be taken of structure imposed on the groups. The two-way display is best seen with reference to a row by column table where means M_{jk} appear in cells, and row and column means are made available, say, on the right and lower margins respectively. Ideally, as in the case of our first example, the design will be balanced; in any case the set of means for rows and columns can generally be used to generate contrast coefficients (usually called row and column 'effects'). The standard way to generate contrast coefficients cum effects is to compute them by subtracting the grand mean from the row and column marginal means respectively. For rows, $c_{j.} = M_{j.} - M$; for columns, $c_{.k} = M_{.k} - M$. The `granova.2w` function can (in principle) accommodate any number of rows or columns.

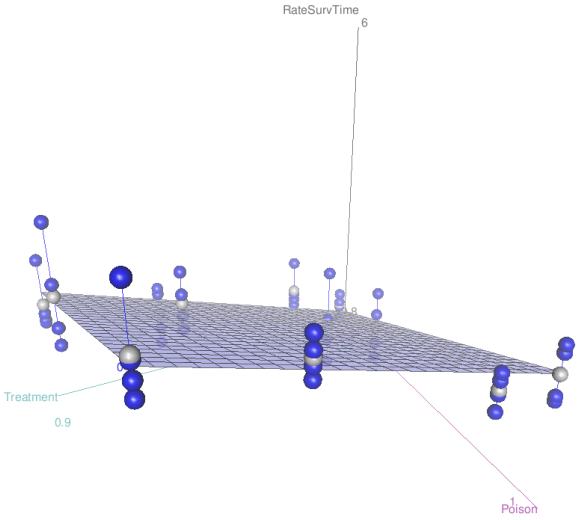
The two-way ANOVA analog of the straight line that shows the locus of means for one-way ANOVA is a flat surface, where grid-lines can be used to show how the $c_{j.}$ and $c_{.k}$ identify the ordered (and properly spaced) rows and columns. In general, as in the one-way graphic, row and column effects are ordered and a flat surface always follows from fitting additively. The degree of 'tilt' of the surface (most evident in the initial display) corresponds generally to the magnitude of the corresponding effect. Initially, before one begins to rotate the graphic, factor labels are shown on the lower axes; response values are shown on vertical axis. When used interactively, rotation generally affords clear visualization of the data in all cells, whatever their sizes. Such dynamic graphics connect directly to basic questions that underpin two-way ANOVA, as they facilitate visualization of effects.

When cell sizes vary, and especially when the design is not balanced, effects generally take the cell sizes into account. In the case of unbalanced data, however, many students of ANOVA methodology hold differing opinions about how analyses should proceed. That is, different analysts routinely advocate different models to analyze unbalanced data. For this reason, numerical results produced by `granova.2w` may not satisfy some data analysts. It follows that the likelihood that the `granova.2w` graphic will be judged as 'satisfactory' may depend on how close the row by column dataset is to being balanced. Furthermore, individual groups need not be of equal sizes, and indeed cell counts of zero are permitted.

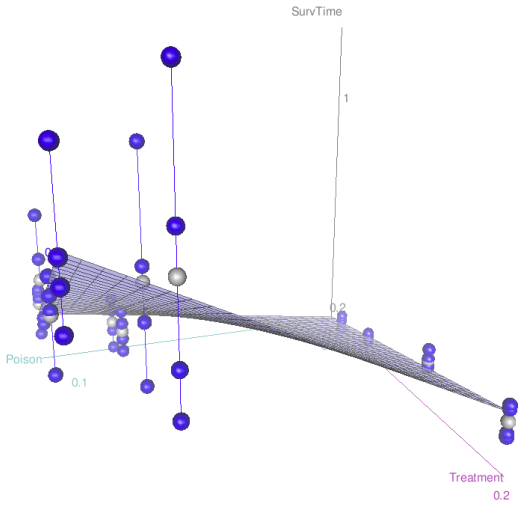
3.1 Details for the Two-Way ANOVA Example: A 3×4 Structure with 4 replications per Group

Figure 3(a) exhibits a snapshot for inverse survival time data for a two-way ANOVA corresponding to the preferred analysis, as discussed in BHH. Each of the twelve within-cell distributions is depicted as a set of (blue) spheres, (initially) vertically stacked for individual row-column combinations. The mean for each row-column combination corresponds to a white sphere. (Note that if the mean for any group is close to one of the data points, then the mean depicted by the white sphere may be masked or obscured.)

The flat surface seen in Figure 3(a) corresponds to the "fitted means" for an additive model analysis. Tacitly, use of an additive model implies that cell means lie sufficiently close to the flat surface, which means no evidence of interaction between the rows and columns. Indeed, the test statistic for interaction in this example is not statistically different from zero. Each interaction term (for any cell of such a two-way table) can be written as $M_{jk} - M_{j.} - M_{.k} + M$. When the deviations, shown by the difference between the cell means and the corresponding fitted values (the algebraic sum of the trailing three terms in this expression) tend to be large, this is generally



(a) Additive fitted model to survival rates (inverse survival times).



(b) Quadratic fitted model to untransformed survival times.

Figure 3: Screen capture of three dimensional plots with fitted models of the poison survival rates and times from Box, Hunter & Hunter [2]. Note how the quadratic model differs from the additive version; graphics are interactively zoomable and orientable.

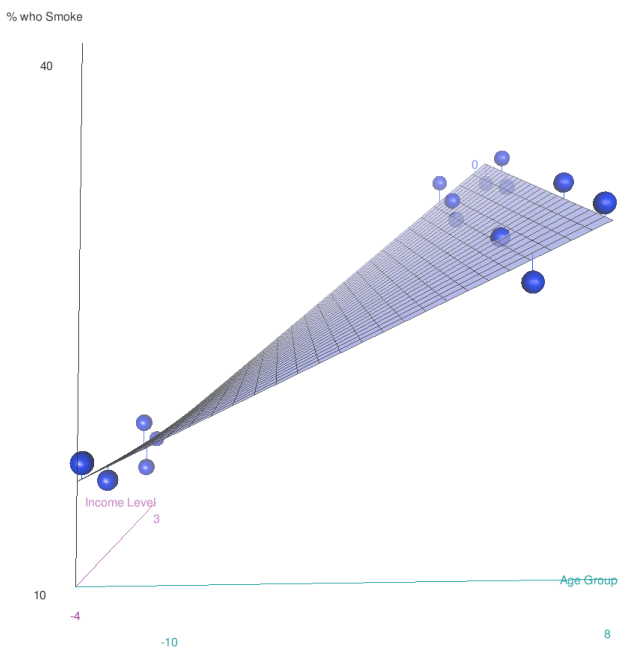


Figure 4: Screen capture of the smokers dataset with quadratic model fit.

taken as evidence of interaction between the row and column effects. This effect will show up as deviations, possibly systematic, of the white spheres from the flat surface when an additive model has been fit.

Figure 3(b) shows a graphic (snapshot) where the response is the untransformed survival times with a quadratic fit superimposed; the curvature is strongly evident (compare to Figure 3(a)). That is, Figure 3(b) shows the data in its original metric (SurvTime) and provides a surface based on a quadratic model fit. The comparison of these two elemental graphics for the transformed and the original response points up how much difference a non-linear transformation can make; particularly, how much simplification can be lent to the analysis by such a transformation that homogenizes group variances.

3.2 Details for a Second Two-Way ANOVA Example: A 3×5 Structure with No Replications

Another example is shown here, using data from Hoaglin, Mosteller and Tukey (1991) [4] (HMT). In this case the purpose was to describe the relationships among five income groups, for three age groups and the percent of the population who smoke. There is only one response for each Income/Age combination: the percent in the population who smoke. (The original source of the data is the Health Interview Survey conducted by the National Center for Health Statistics between 1978 and 1980, as excerpted by HMT in chapter 6.) Note that lacking replications within cells means that standard methods to test hypotheses about interaction do not work. However, the graphic provides information that suggests interaction in this case; see Figure 3. As in the case of the poison data, HMT, like BHH, examine and discuss these data more thoroughly than space permitted here.

Figure 4 shows a quadratic fit to percentages who smoke for the HMT data (with no replications within cells); note that this curved surface fits the data about as well as a simple model can. Indeed, the quadratic fit is equivalent to using Tukey's one-degree-of-freedom-for-non-additivity, as in the case of the survival time analysis (see Figure 3(b)). This shows that in cases like this the graphic

does more than permit the analyst to "see" the data, it carries a visual message of what it means to fit Tukey's one-degree-of-freedom-for-non-additivity model in a two-way ANOVA. For percent data, non-linear transformations can, as seen in the analysis of the poisons data, help to simplify analyses; for more details see chapter 13 of HMT [4].

Finally, we note that numerical results provided in this run of the `granova.2w` graphic are wholly straightforward, and conform with those given by BHH. However, as noted above, for unbalanced data, or datasets where there are no observations in some cells, numerical results given by `granova.2w` may be inappropriate (at least by some standards). Our experience suggests that when some cells are empty, it may be helpful to *impute* data for those cells and proceed, taking care to see how the pseudo-values play out in the graphic as well as in the numerical analysis. Of course strategies like this can be used for unbalanced data generally. Naturally, imputation can be done in a variety of ways, so the analyst may want to compare graphics across different imputations. There is a large literature on the analysis of unbalanced designs and imputation; for an especially informative discussion of this issue, see Venables [11].

4 A Graphic for Comparing Groups Using Contrasts

Contrasts, or contrast vectors, show comparisons between groups, or linear functions of group means; individual coefficients, negative versus positive, show exactly which groups are "contrasted with" others. Our elemental graphic for this case has been developed for depicting how groups of scores, not just group means, compare across the (non-zero) contrast coefficients with respect to responses. The set of comparisons, one for each contrast vector, can be seen as constituting an elemental graphic for the set of contrast vectors that the user has supplied. This function can facilitate analyses of higher order fixed effects designs especially those with relatively few levels for each factor; it can work for crossed or factorial designs as well as those that entail nesting, or mixtures thereof.

The function `granova.contr` accommodates up to $J-1$ contrasts for J groups (or "cells"). If fewer than $J-1$ contrasts are specified, then the number per group must be provided in the initial call. For the current version of this function group sizes must be equal. Contrast vectors need not be mutually orthogonal, however a number of virtues derive from mutual orthogonality. The elemental graphic(s) generates as many panels as there are specified contrasts, with data values jittered so as to reduce overlapping of points. (The amount of jittering can be controlled by the user.)

In addition to the jittered points, each panel exhibits a straight line that connects response means for that contrast as linearly combined for the negative (left) and positive (right) coefficients in the corresponding contrast vector. Evidence of non-zero effects correspond to slopes of these lines that depart notably from horizontal (slope = 0). Standardization of contrast vectors ensures that negative coefficient means (generally linear combinations) are always compared with positive coefficient means; this provides the generality to accommodate virtually any pre-specified contrasts, and gives a framework for comparing standardized effect sizes over contrasts. A set of prescribed contrasts defines any analysis. In general outliers and/or skewed distributions are readily displayed in some or all of the panels.

In addition to the displays for individual contrasts, a numerical summary is provided with the graphic showing the pattern of means across groups (in the order given). Numerical results can help inform the user about 'significance' in the case of each panel using a regression approach to the analysis, where the usual normal theory assumptions are (tacitly) invoked. Each contrast vector is a predictor, and the set of these independent variables are used to predict the response. The usual lines of the standard ANOVA or linear model summary table for contrast-based designs are isomorphic with the panel displays.

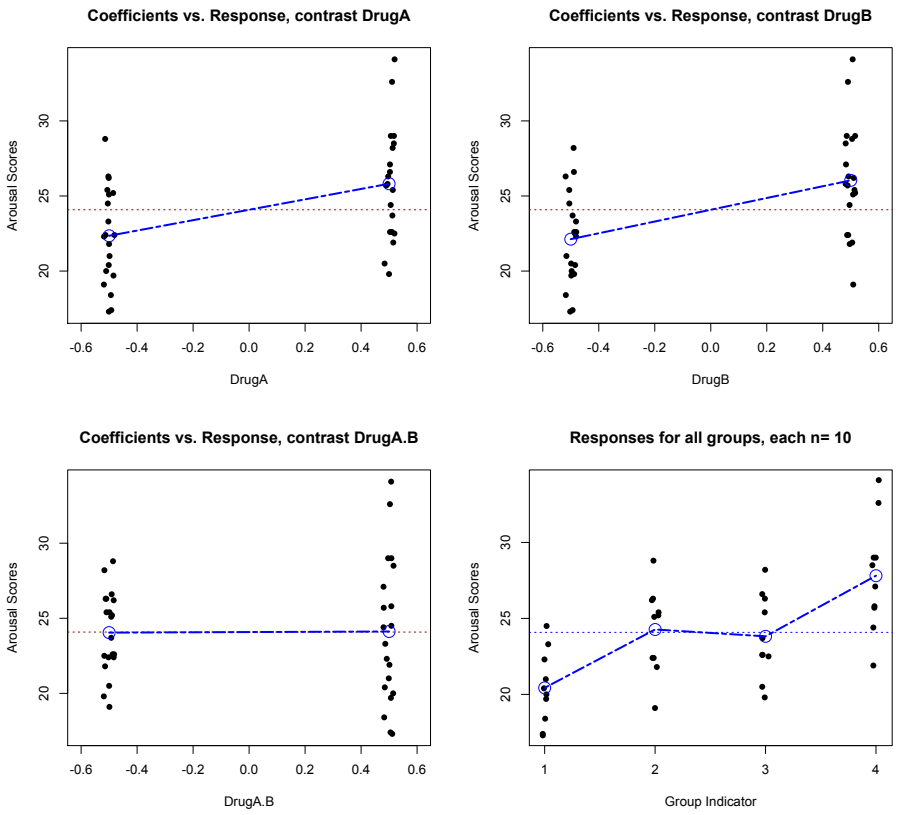


Figure 5: Contrasts for Lowry Dataset [6]

4.1 Details for a Contrasts-based Analysis of Physiological Arousal Data: A 2×2 Structure with Replications.

Figure 5 depicts an analysis based on three contrasts for a 2×2 design, where the contrast vectors show how groups are compared in what has become a standard analysis. These kinds of contrasts readily generalize to virtually any 2^k factorial or 2^{k-m} fractional factorial design, and are easily implemented in `R`. Lowry [6] provides these data, which pertain to arousal levels in rats, and he carries out all standard computations to help link our results with those that are standardly reported for this ANOVA.

The first panel contrasts the two levels of Drug A; the second shows the effect of two levels of Drug B on arousal; these correspond to so-called main effects. The blue dashed lines that compare means suggest that both main effects are non-zero for both drugs. The numerical analysis reinforces this interpretation as it yields t-statistics with magnitudes above 3 for both main effects; for each, the standardized effect size exceeds unity (in magnitude). The third panel corresponds to a contrast defined as the product of the first two (DrugA.B) and this suggests no interaction; indeed the t-statistic is near zero as is the standardized effect size. Finally, the fourth panel does not concern contrasts, but instead exhibits scores and means of all (four) treatment groups. The last panel might be compared an output from a one-way treatment using `granova.1w`, although we do not conduct that analysis.

4.2 Details for Contrasts for the Poison Data: Data Based Contrasts

The poisons data set, using inverse survival time responses, has been used for the illustrations in Figure 6. The goal of this illustration is to show how a contrasts-based graphic might effectively

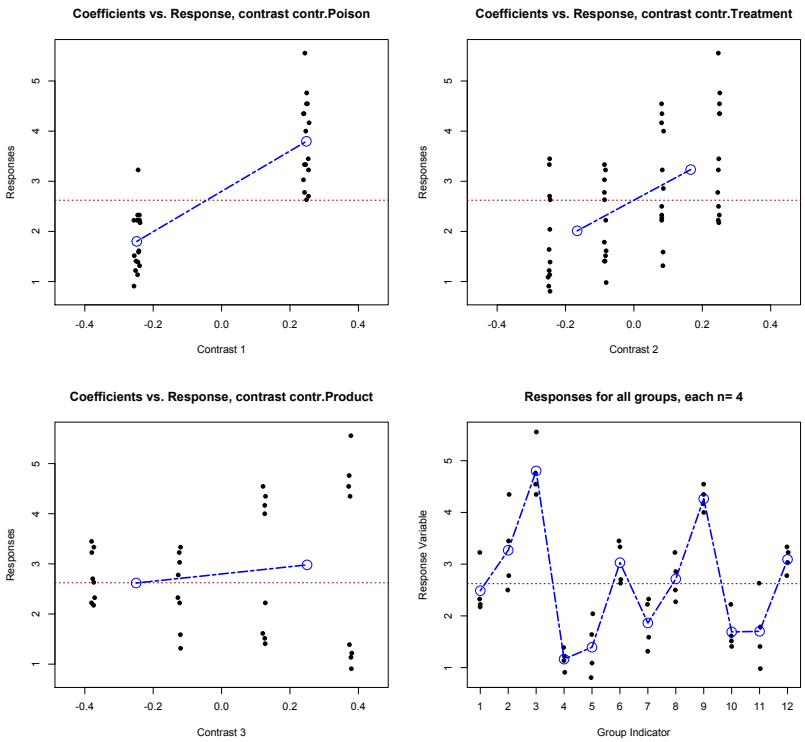


Figure 6: Contrasts for poison data from Box, Hunter & Hunter [2].

be generated following a two-way ANOVA, at least for a balanced design. Note the caveats below, however, as related to the numerical summary. Construction of the specific contrasts was based on information provided in the output of the initial two-way ANOVA shown in Section 3. In particular, given the *ordering* of the means for the row and column factors, and using just a simple "linear" contrast for the row and column groups, these first two data-based contrast vectors show that nearly all variance of the means for this 4×3 design can be summarized or accounted for using only two contrasts based on mean ordering, one for each factor.

A third contrast is also shown, computed using individual products of the respective contrast vectors, where the interaction effect which for the reciprocal survival time metric (same as survival rate) is notably smaller than that for the row and column contrasts. Had the experimenter *prespecified* these two linear contrasts, for rows and columns respectively, then the analysis would constitute a planned-comparison among groups, and the numerical (test statistic) results, and corresponding probabilities, would (ostensibly) have been meaningful. As it is, given that the initial data analysis results were used to construct the contrast coefficients, the display shown in Figure 6 is nevertheless useful for summarization of variation among group means. The qualification holds that probabilistic results (printed in Appendix A 7) for such a run are not meaningful since in fact these contrasts were not planned. We have found that graphics like this based on contrasts often show no discernible effects with respect to mean differences, but they may, as in the case of right side of the southwest panel in Figure 6, show that groups associated with contrasts may have point-sets that differ noticeably from one another in ways we might not have become aware of sans graphics.

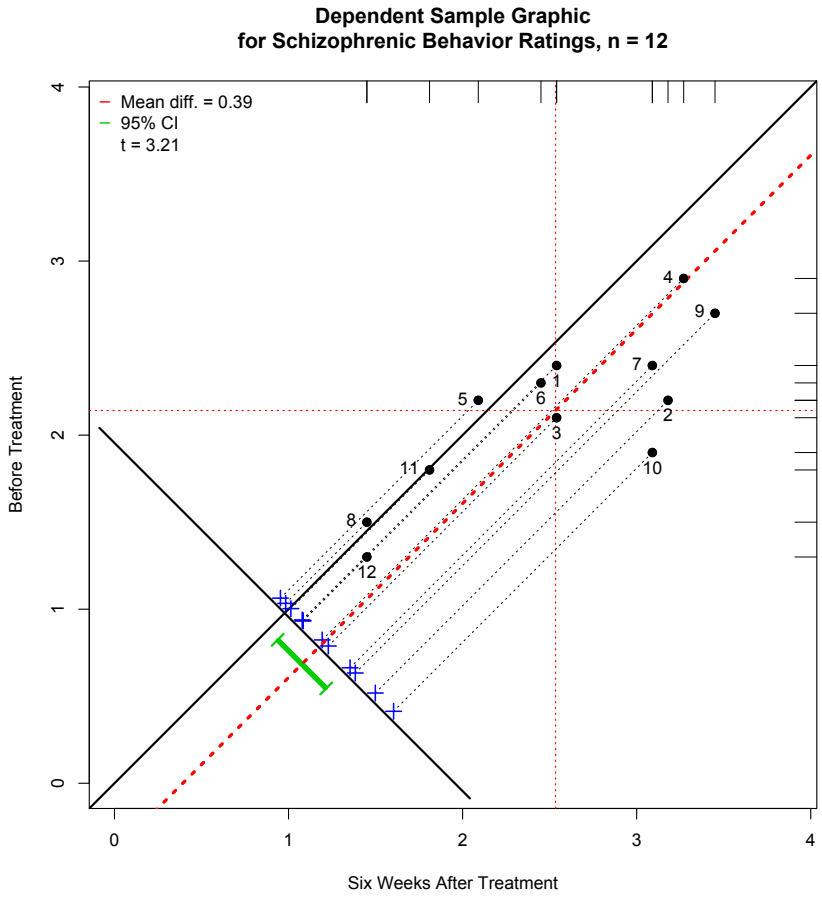


Figure 7: Dependent sample assessment plot using schizophrenic behavior ratings from Stanley and Walton [9], via Lehman [5]

5 Dependent Sample Assessment Plots

Next, we introduce an elemental graphic for analysis of two dependent samples. We call this a *Dependent sample assessment plot*, where an example is shown in Figure 7.[4]. As discussed in Pruzek and Helmreich (2009) [7], there are at least four distinctive ways to acquire dependent sample data, two of which entail repeated measures for units or individuals, and two of which entail either blocking or matching to define pairs. In all cases, X and Y scores are used to generate difference scores D that are the usual focus of an analysis. The graphic we present shows all scores, X, Y, and D, and how they relate to one another. Note that the central question that usually drives a dependent sample analysis concerns the average difference score, and especially how much that differs from zero. Typically, the use of graphics in the dependent sample case, if used at all, is to check parametric assumptions about difference scores; we shall demonstrate some advantages of adding graphical results for such data.

5.1 Details for a Repeated Measures Problem: Examining Effects of a Placebo Using Difference Scores

The graphic shown in Figure 7 focuses on a scatterplot of two dependent samples (paired (X, Y) values), here for a sample of size $n = 12$. Each pair of scores corresponds to before and after

[4]A more detailed examination of the dependent sample paradigm, especially via the elemental graphic presented here, is presented in Pruzek & Helmreich[7]

behavior ratings of the same person diagnosed as a chronic schizophrenic. In this case, Y values show ratings before taking a 'placebo' as if it were a drug, X values depict ratings six weeks after treatment. Data are from an Stanley and Walton [9]; note that this example is discussed by Lehmann [5]. For each (X, Y) point, one can see three projections: one to the 'north' forming a rug plot showing the X score distribution, another to the 'east' showing the distribution via rug tufts for Y scores; and finally to the 'southwest' where projections are shown by thin dotted lines to crosses on a line segment at the lower-left. The (narrow, red) dashed vertical and horizontal lines (that intersect with the heavier diagonal (dashed red) line) correspond to the means for the X and Y (marginal) distributions respectively. The 12 crosses depict a (version of a) stripchart of the D's, where each $D = X - Y$, i.e., the distribution of difference scores. The mean of the distribution of difference scores (crosses) corresponds to the heavy dashed (red) line that is parallel to the identity line (showing $X = Y$); the green line segment below the difference score crosses shows a standard 95% confidence interval, which in this case is rather short and does not span zero. We have included (optional) reference labels for each individual to assist the discussion.

While a standard analysis focuses on the difference scores and either a t-statistic or a confidence interval, considerably more information is readily seen in the graphic. We distinguish two groups of subjects in Figure 7: A = (8, 11, 12, 1, 5, 6) and B = (2, 3, 4, 7, 9, 10). The six subjects in group A showed almost no change after treatment (all points fall near or on the $Y = X$ line). The B group differs from A, since it is for these individuals that the strongest placebo effects appear. In fact, the summary significance shown by the t-statistic of 3.21 rests almost entirely on effects for group B, since the effects for individuals in group A are near zero. Although such a post-hoc interpretation risks over-interpretation of data, it seems likely that the investigator who is concerned about placebo effects would want to know more about how subjects in group B might differ from those in A since the B persons seemed more susceptible to placebo effects. Such questions are representative of the kind of post-hoc queries that might be of interest to the applied researcher, and they are far more readily brought to light when using this type of graphics-approach to analysis than would be the case for an exclusively numerical analysis.

5.2 Details for a Randomized Block Dataset: Hypothetical Data Comparing Pairs of Diabetics

An extensive search of more than three dozen textbooks and several articles turned up no useful examples of what we deem to be a particularly promising approach to experimental design and analysis. In particular, as will be illustrated using hypothetical data, we have strong reason to believe that relatively few researchers recognize the potential of randomized block designs for increasing design efficiency when comparing treatments. (We have also come to believe that these designs are generally not effectively taught in most statistics textbooks.) However simple, our example illustrates the key points.

Suppose we aim to compare two treatments (that could be two diets, drugs or exercise regimens, etc.). We begin using simulated data for a sample of 30 diabetics. We shall refer to the two groups as 'Treatment' and 'Control', respectively. Suppose further before we began the treatments, we obtained A1c measures of blood glucose, this being a standard measure that reflects a blood glucose levels over the most recent 3 months. Given these pre-experimental A1c values, we rank all persons, and then form blocks of size two: the two highest scores have been assigned to block 1, the next pair to block 2, until all 30 persons are assigned to 15 blocks. For randomized blocks the next step is to randomly allocate individuals within blocks to the 'Treatment' or 'Control' groups.

Over a period of several months the experiment will be assumed to have been carefully run (possibly using double blind methodology) after which A1c measures will again be obtained for all individuals. The hypothetical A1c values we use for analysis represent a realistic range of values,

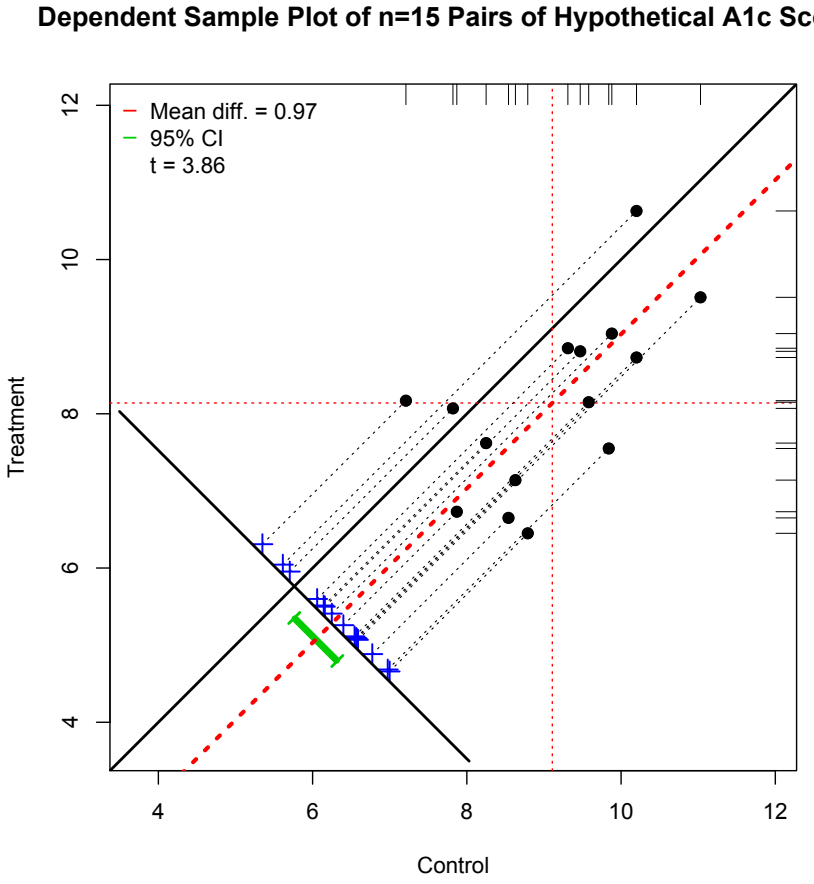


Figure 8: Dependent sample assessment plot of simulated A1c scores for 15 randomized blocks of diabetic patients.

as plotted in Figure 8, using function `granova.ds`. As in the case of the preceding example each data point corresponds to a row in the original data set, but in this case these rows correspond to pairs (blocks) of individuals for the 'Treatment' and 'Control' groups respectively. In this case the Treatment A1c scores are notably lower than their Control counterparts, which suggests a desirable effect of the treatment. Indeed, as shown by the 95 percent CI (green line segment, lower left), these data are inconsistent with the hypothesis of equal mean A1c scores for the two putative populations. Since A1c scores generally correlate quite highly with one another over time periods of several months, and blocking was based on initial A1c scores, this design has controlled one of the key individual differences for participants in this experiment. This general approach, based on standard blocking methodology, can increase – perhaps quite substantially – the likelihood of finding statistically reliable effects. The method can also be helpful in identifying interactions between treatments and individual difference variables. Generally speaking the stronger the correlation between the initial ranking variable and the ultimate response (putting aside treatment effects), the greater the improvement in design efficiency as compared with independent sample designs that do not employ blocking.

Treatment effects can often be demonstrated using such a design, even for small samples, whenever an effective means can be found for ranking individuals before blocking. In this case, the positive correlation (.62) between scores in the 'Treatment' and 'Control' columns reflects the dependency introduced by blocking. Pruzek and Helmreich (2009) discuss various other possibilities in some detail, including the case where dependent sample assessment plots can be employed to advantage in the analysis of observational data.

Note further that when an elemental graphic such as that in Figure 8 is used to display data for a dependent sample analysis, the analyst is provided with a variety of visual information to help understand such data. A little reflection makes clear that attention need not be restricted to inference based on (somewhat arbitrary) summary statistics; and indeed, description of data may turn out to be at least as interesting or useful as inference. In this case, the graphic shows that 3 of the 15 pairs exhibit effects contrary to the mean difference between Treatment and Control groups. At least two possible interpretations seem relevant in such situations: one is that outcome measures – such as the A1c scores – may not be wholly reliable. It is common for outcome measures to vary more or less randomly over time. Second, there may be reason to suspect that scores vary systematically with respect to grouping variables. (It is known that A1c scores tend to differ systematically by ethnicity.) That is, covariate differences may be associated with individual differences within or between blocks. In general, graphics may show clusters, outliers, or trends across blocks. In general, many such visually identified 'irregularities' may be seen as evidence of interactions between treatments and covariates, interactions that are unlikely to have been identified when analyses do not go beyond numerical summaries. The function `granova.ds` permits use of labels for points that can facilitate explorations of such possibilities. Finally, note that marginal distributions for the variate can be informative (shown as rug plots in Figure 8), as can the shape of the distribution of difference scores (which is often assumed to be normal in the parent population).

6 Concluding Remarks

Although it has become standard practice to include graphics for many statistical analyses, particularly those based on regression and ANOVA, with few exceptions such graphics tend not to be strongly connected or closely linked to corresponding numerical analyses. By introducing *elemental graphics*, we aim to strengthen the connection between the questions that drive particular methods of analysis and graphics for displaying data. Experience suggests that this step often helps to understand the applicability of chosen methods for particular data sets, and also to aid better understanding of methods generally. Of course it can also help the analyst understand extant data. While we have concentrated on graphics for various ANOVA methods because we have focused on documenting a certain package in `R`, it is clear that elemental graphics have been or could be developed for many other methods. Graphics produced using the `assoc` function in the `vcd` package in `R` provide good examples of the possibilities. While elemental graphics may have special value for pedagogic uses, they can be broadly useful in many situations where one aims to learn just how particular data points play out in the context of particular methods of analysis.

We have concentrated on the four functions: `granova.1w`, `granova.2w`, `granova.contr` and `granova.ds`, which are available in the `R` package `granova`. In the foregoing, we have tried to illustrate principal aspects of each elemental graphic, but the reader should recognize that far more insight derives from hands-on experiences of using such graphical methods with one's own data, where the particulars of context and applicability may become central.

A broader challenge is to encourage more focused and deeper thinking about strengths and weaknesses of data analytic methods, perhaps especially in the light of graphic display methods for wide varieties of data. In our experience, pedagogical value often leads to practical value in wide ranges of situations. As the estimable John Tukey so often emphasized, applied science is best served when methods are used in the service of understanding data, not as ends in themselves.

7 Appendix A: Command and Numeric Output from Functions

7.1 Command and Numeric Output Figure 1:

```
R> granova.1w(poison$RateSurvTime, poison$Group,
     ylab = "Inverse of Survival Time or Survival Rate",
     resid = TRUE, top.dot = .15, kx = 1.4, px = 1.4)

$grandsum
 Grandmean df.bet df.with MS.bet MS.with F.stat F.prob SS.bet/SS.tot
      2.62  11.00   36.00   5.17    0.24  21.53   0.00          0.87
```

7.2 Command and Numeric Output Figure 2:

```
R> granova.1w(poison$RankRateSurvTime, poison$Group,
    ylab = "Inverse of Survival Time or Survival Rate",
    resid = TRUE, top.dot = .15, kx = 1.4, px = 1.4)

$grandsum
 Grandmean df.bet df.with MS.bet MS.with F.stat F.prob SS.bet/SS.tot
      2.49  11.00   36.00   3.70    0.19  19.18   0.00          0.85
```

7.3 Command and Numeric Output Figure 3(a):

```
R> granova.2w(poison[, c(4, 2, 1)])
[1] SurvTime ~ Treatment * Poison
$Treatment.effects
      A       C       D       B
-0.1650 -0.0869  0.0548  0.1970

$Poison.effects
   III     II      I
-0.203  0.065  0.138

$CellCounts.Reordered
         Poison
Treatment III II I
        A   4  4 4
        C   4  4 4
        D   4  4 4
        B   4  4 4

$CellMeans.Reordered
         Poison
Treatment   III    II     I
        A 0.210 0.320 0.412
        C 0.235 0.375 0.568
        D 0.325 0.668 0.610
        B 0.335 0.815 0.880
```

```
$aov.summary
                 Df  Sum Sq Mean Sq F value    Pr(>F)
Treatment         3 0.92121 0.30707 13.8056 3.777e-06 ***
Poison            2 1.03301 0.51651 23.2217 3.331e-07 ***
Treatment:Poison  6 0.25014 0.04169  1.8743    0.1123
Residuals        36 0.80072 0.02224
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

7.4 Command and Numeric Output Figure 3(b):

```
R> granova.2w(poison[, c(4, 2, 1)], fit = "quadratic")
[1] SurvTime ~ Treatment * Poison
$Treatment.effects
      A       C       D       B
-0.1650 -0.0869  0.0548  0.1970

$Poison.effects
   III     II      I
-0.203  0.065  0.138

$CellCounts.Reordered
         Poison
Treatment III II I
        A   4  4 4
        C   4  4 4
        D   4  4 4
        B   4  4 4

$CellMeans.Reordered
         Poison
Treatment   III    II     I
        A 0.210 0.320 0.412
        C 0.235 0.375 0.568
        D 0.325 0.668 0.610
        B 0.335 0.815 0.880

$aov.summary
                 Df  Sum Sq Mean Sq F value    Pr(>F)
Treatment         3 0.92121 0.30707 13.8056 3.777e-06 ***
Poison            2 1.03301 0.51651 23.2217 3.331e-07 ***
Treatment:Poison  6 0.25014 0.04169  1.8743    0.1123
Residuals        36 0.80072 0.02224
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

7.5 Command and Numeric Output Figure 4:

```
R> granova.2w(smokers, fit = 'quadratic')
[1] percent ~ income * age
$income.effects
     E:>25000 D:15000-24999 C:10000-14999 A:<5000 B:5000-9999
```

```
        -3.73         -2.07          1.27    1.60        2.93

$age.effects
   65+  17-30  31-65
-13.70   5.27   8.47

$CellCounts.Reordered
               age
income          65+ 17-30 31-65
  E:>25000        1     1     1
  D:15000-24999   1     1     1
  C:10000-14999   1     1     1
  A:<5000         1     1     1
  B:5000-9999     1     1     1

$CellMeans.Reordered
               age
income          65+ 17-30 31-65
  E:>25000       17    28    33
  D:15000-24999  15    32    36
  C:10000-14999  18    36    39
  A:<5000        14    38    42
  B:5000-9999    16    41    41

$aov.summary
            Df  Sum Sq Mean Sq
income       4   92.93   23.23
age          2 1440.13  720.07
income:age   8   75.87    9.48
```

7.6 Command and Numeric Output Figure 5:

```
R> granova.contr(arousal, contrasts = contrasts22, ylab = "Arousal Scores",
   xlab = names(contrasts22))
$summary.lm

Call:
lm(formula = resp ~ contrst)

Residuals:
   Min     1Q Median     3Q    Max
-5.910 -2.015 -0.075  1.885  6.290

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.0825     0.4657  51.712  < 2e-16 ***
contrst1      3.4650     0.9314   3.720 0.000676 ***
contrst2      3.9150     0.9314   4.203 0.000166 ***
contrst3      0.0750     0.9314   0.081 0.936267
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.945 on 36 degrees of freedom
```

```
Multiple R-squared: 0.4668,Adjusted R-squared: 0.4223
F-statistic:  10.5 on 3 and 36 DF,  p-value: 4.173e-05


$means.pos.neg.coeff
           neg   pos diff stEftSze
Drug.A   22.35 25.82 3.46     1.18
Drug.B   22.12 26.04 3.91     1.33
Drug.A.B 24.05 24.12 0.07     0.03

$contrasts
     Drug.A Drug.B Drug.A.B
[1,]   -0.5   -0.5      0.5
[2,]   -0.5    0.5     -0.5
[3,]    0.5   -0.5     -0.5
[4,]    0.5    0.5      0.5

$group.means.sds
       [,1]  [,2]  [,3]  [,4]
Means 20.43 24.27 23.82 27.81
S.D.s  2.41  2.81  2.74  3.67
```

7.7 Command and Numeric Output Figure 6:

```
R> con.poison
      contr.Poison contr.Treatment contr.Product
 [1,]           -1               3            -3
 [2,]            0               3             0
 [3,]            1               3             3
 [4,]           -1              -3             3
 [5,]            0              -3             0
 [6,]            1              -3            -3
 [7,]           -1               1            -1
 [8,]            0               1             0
 [9,]            1               1             1
[10,]           -1              -1             1
[11,]            0              -1             0
[12,]            1              -1            -1
R> granova.contr(poison$RateSurvTime, con = con.poison)
$summary.lm

Call:
lm(formula = resp ~ contrst)

Residuals:
     Min       1Q   Median       3Q      Max
-1.34972 -0.28122  0.02386  0.30390  0.91469

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.62238    0.07891  33.231  < 2e-16 ***
contrst1     4.02505    0.38971  10.328 2.44e-13 ***
```

```
contrst2     3.47925    0.42621   8.163 2.35e-10 ***
contrst3     0.46785    0.34522   1.355    0.182
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.5467 on 44 degrees of freedom
Multiple R-squared: 0.7992,Adjusted R-squared: 0.7855
F-statistic: 58.38 on 3 and 44 DF,  p-value: 2.21e-15

$means.pos.neg.coeff
                 neg  pos diff stEftSze
contr.Poison    1.80 3.80 2.00     4.07
contr.Treatment 2.01 3.23 1.22     2.49
contr.Product   2.62 2.98 0.36     0.74

$contrasts
      contr.Poison contr.Treatment contr.Product
 [1,]        -0.25           0.250        -0.375
 [2,]         0.00           0.250         0.000
 [3,]         0.25           0.250         0.375
 [4,]        -0.25          -0.250         0.375
 [5,]         0.00          -0.250         0.000
 [6,]         0.25          -0.250        -0.375
 [7,]        -0.25           0.083        -0.125
 [8,]         0.00           0.083         0.000
 [9,]         0.25           0.083         0.125
[10,]        -0.25          -0.083         0.125
[11,]         0.00          -0.083         0.000
[12,]         0.25          -0.083        -0.125

$group.means.sds
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
Means 2.49 3.27 4.80 1.16 1.39 3.03 1.86 2.71 4.26  1.69   1.7  3.09
S.D.s 0.50 0.82 0.53 0.20 0.55 0.42 0.49 0.42 0.23  0.36   0.7  0.24
```

7.8 Command and Numeric Output Figure 7:

```
R> granova.ds(schiz, rev = TRUE, ident = TRUE, main = "Dependent Sample Graphic
   for Schizophrenic Behavior Ratings, n = 12")

            Summary Stats
n                  12.000
mean(x)             2.534
mean(y)             2.142
mean(D=x-y)         0.392
SD(D)               0.424
ES(D)               0.926
r(x,y)              0.817
r(x+y,d)            0.604
LL 95\%CI           0.123
UL 95\%CI           0.662
t(D-bar)            3.208
```

```
df.t               11.000
pval.t              0.008
```

7.9 Command and Numeric Output Figure 8:

```
R> granova.ds(a1c, revc = TRUE, main = "Dependent Sample Plot of n=15
   Pairs of Hypothetical A1c Scores")

            Summary Stats
n                  15.000
mean(x)             9.108
mean(y)             8.140
mean(D=x-y)         0.968
SD(D)               0.971
ES(D)               0.997
r(x,y)              0.622
r(x+y,d)           -0.114
LL 95\%CI           0.430
UL 95\%CI           1.506
t(D-bar)            3.862
df.t               14.000
pval.t              0.002
```

8 Appendix B: Tables

	percent	income	age
1	38.00	A:<5000	17-30
2	41.00	B:5000-9999	17-30
3	36.00	C:10000-14999	17-30
4	32.00	D:15000-24999	17-30
5	28.00	E:>25000	17-30
6	42.00	A:<5000	31-65
7	41.00	B:5000-9999	31-65
8	39.00	C:10000-14999	31-65
9	36.00	D:15000-24999	31-65
10	33.00	E:>25000	31-65
11	14.00	A:<5000	65+
12	16.00	B:5000-9999	65+
13	18.00	C:10000-14999	65+
14	15.00	D:15000-24999	65+
15	17.00	E:>25000	65+

Table 1: The smokers dataset from HMT [4]

	Before Treatment	Six Weeks After Treatment
1	2.40	2.54
2	2.20	3.18
3	2.10	2.54
4	2.90	3.27
5	2.20	2.09
6	2.30	2.45
7	2.40	3.09
8	1.50	1.45
9	2.70	3.45
10	1.90	3.09
11	1.80	1.81
12	1.30	1.45

Table 2: The schizophrenia dataset; measures of schizophrenic behavior before and six weeks after treatment with a placebo. From an article by Stanley and Walton [9], obtained via Lehmann [5].

Row	Poison	Treat	Grp	SrvT	RtSrvT	RkRtSrvT	c.Pois	c.Tr	c.Prd
1	I	A	1	0.31	3.23	3.19	-1	3	-3
2	I	A	1	0.45	2.22	2.10	-1	3	-3
3	I	A	1	0.46	2.17	1.96	-1	3	-3
4	I	A	1	0.43	2.33	2.32	-1	3	-3
5	II	A	2	0.36	2.78	2.82	0	3	0
6	II	A	2	0.29	3.45	3.47	0	3	0
7	II	A	2	0.40	2.50	2.46	0	3	0
8	II	A	2	0.23	4.35	3.84	0	3	0
9	III	A	3	0.22	4.55	3.98	1	3	3
10	III	A	3	0.21	4.76	4.12	1	3	3
11	III	A	3	0.18	5.56	4.20	1	3	3
12	III	A	3	0.23	4.35	3.84	1	3	3
13	I	B	4	0.82	1.22	1.16	-1	-3	3
14	I	B	4	1.10	0.91	0.88	-1	-3	3
15	I	B	4	0.88	1.14	1.09	-1	-3	3
16	I	B	4	0.72	1.39	1.31	-1	-3	3
17	II	B	5	0.92	1.09	1.02	0	-3	0
18	II	B	5	0.61	1.64	1.74	0	-3	0
19	II	B	5	0.49	2.04	1.89	0	-3	0
20	II	B	5	1.24	0.81	0.80	0	-3	0
21	III	B	6	0.30	3.33	3.33	1	-3	-3
22	III	B	6	0.37	2.70	2.75	1	-3	-3
23	III	B	6	0.38	2.63	2.61	1	-3	-3
24	III	B	6	0.29	3.45	3.47	1	-3	-3
25	I	C	7	0.43	2.33	2.32	-1	1	-1
26	I	C	7	0.45	2.22	2.10	-1	1	-1
27	I	C	7	0.63	1.59	1.60	-1	1	-1
28	I	C	7	0.76	1.32	1.24	-1	1	-1
29	II	C	8	0.44	2.27	2.25	0	1	0
30	II	C	8	0.35	2.86	2.97	0	1	0
31	II	C	8	0.31	3.23	3.19	0	1	0
32	II	C	8	0.40	2.50	2.46	0	1	0
33	III	C	9	0.23	4.35	3.84	1	1	1
34	III	C	9	0.25	4.00	3.62	1	1	1
35	III	C	9	0.24	4.17	3.69	1	1	1
36	III	C	9	0.22	4.55	3.98	1	1	1
37	I	D	10	0.45	2.22	2.10	-1	-1	1
38	I	D	10	0.71	1.41	1.38	-1	-1	1
39	I	D	10	0.66	1.52	1.53	-1	-1	1
40	I	D	10	0.62	1.61	1.67	-1	-1	1
41	II	D	11	0.56	1.79	1.81	0	-1	0
42	II	D	11	1.02	0.98	0.95	0	-1	0
43	II	D	11	0.71	1.41	1.38	0	-1	0
44	II	D	11	0.38	2.63	2.61	0	-1	0
45	III	D	12	0.30	3.33	3.33	1	-1	-1
46	III	D	12	0.36	2.78	2.82	1	-1	-1
47	III	D	12	0.31	3.23	3.19	1	-1	-1
48	III	D	12	0.33	3.03	3.04	1	-1	-1

Table 3: The poison dataset from BHH [2]. The columns are respectively: Row - case number; Poison - type of poison; Treat - treatment method; Grp - group numbers; SrvT: survival time; RtSrvT: survival time rate; that is the inverse of SrvT; RkRtSrvT: ranked rate of survial; c.Pois: poison contrast; c.Tr: treatment contrast; c.Prd: the product of c.Pois and c.Tr for interactions.

	Arousal	FactorA	FactorB
1	20.40	1	1
2	20.00	1	1
3	24.50	1	1
4	19.70	1	1
5	17.30	1	1
6	17.40	1	1
7	18.40	1	1
8	21.00	1	1
9	22.30	1	1
10	23.30	1	1
11	22.40	1	2
12	22.40	1	2
13	26.20	1	2
14	28.80	1	2
15	26.30	1	2
16	19.10	1	2
17	25.40	1	2
18	25.10	1	2
19	21.80	1	2
20	25.20	1	2
21	20.50	2	1
22	26.60	2	1
23	25.40	2	1
24	22.60	2	1
25	22.50	2	1
26	26.30	2	1
27	19.80	2	1
28	28.20	2	1
29	23.70	2	1
30	22.60	2	1
31	34.10	2	2
32	32.60	2	2
33	29.00	2	2
34	29.00	2	2
35	25.70	2	2
36	21.90	2	2
37	28.50	2	2
38	25.80	2	2
39	27.10	2	2
40	24.40	2	2

Table 4: Data from Lowry [6]; pertain to arousal levels in rats.

	Treatment	Control
1	7.14	8.63
2	8.15	9.58
3	8.07	7.82
4	7.62	8.25
5	7.55	9.84
6	6.73	7.87
7	8.17	7.21
8	8.85	9.31
9	9.04	9.88
10	8.81	9.47
11	9.51	11.03
12	10.63	10.20
13	6.45	8.79
14	8.73	10.20
15	6.65	8.54

Table 5: Simulated A1c scores for 15 randomized blocks of diabetic patients.

References

[1] G. E. P. Box and D. R. Cox. An analysis of transformations (with discussions). *Journal of the Royal Statistical Society, Series A*, 143:383–430, 1964.

[2] G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters*. Wiley, New York, 1978.

[3] Andrew Gelman. Analysis of variance: why it is more important than ever. *The Annals of Statistics*, 33(1):1–53, 2005.

[4] D. C. Hoaglin, F. Mosteller, and J. W. Tukey, editors. *Fundamentals of Exploratory Analysis of Variance*. Wiley & Sons, New York, 1991.

[5] Erich L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, 1975.

[6] Richard Lowry. *Concepts & Applications of Inferential Statistics*. Vassar College, Poughkeepsie, N.Y., 2010. http://faculty.vassar.edu/lowry/webtext.html.

[7] R. M. Pruzek and J. E. Helmreich. Enhancing dependent sample analyses with graphics. *Journal of Statistics Education*, 17(1), March 2009. http://www.amstat.org/publications/jse/v17n1/helmreich.pdf.

[8] R. M. Pruzek and J. E. Helmreich. `granova`*: Graphical Analysis of Variance*, 2010. `R` package version 2.0.

[9] W. J. Stanley and D. Walton. Trifluoperazine ("stelazine") a controlled clinical trial in chronic schizophrenia. *Journal of Mental Science*, 107:250 – 257, 1961. Table I.

[10] `R` Development Core Team. `R`: *A Language and Environment for Statistical Computing*. `R` Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[11] W. N. Venables. Exegeses on linear models. In *Proceedings of the 1998 International S-PLUS Users Conference*, 2000. http://www.stats.ox.ac.uk/pub/MASS3/Exegeses.pdf.