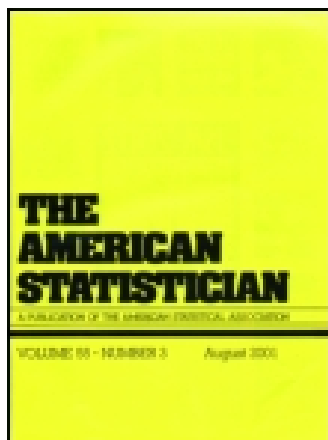# Dot Plots

Leland Wilkinson [a] [b]

[a] SYSTAT Products , SPSS Inc., 444 North Michigan Avenue, Chicago , IL , 60611 , USA

[b] Department of Statistics , Northwestern University , Evanston , IL , 60201 , USA
Published online: 17 Feb 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Statistical Computing and Graphics

Dot Plots

Leland WILKINSON

Dot plots represent individual observations in a batch of data with symbols, usually circular dots. They have been used for more than 100 years to depict distributions in detail. Hand-drawn examples show their authors' efforts to arrange symbols so that they are as near as possible to their proper locations on a scale without overlapping enough to obscure each other. Recent computer programs that attempt to reproduce these historical plots have unfortunately resorted to simple histogram binning instead of using methods that follow the rules for the hand-drawn examples. This article introduces an algorithm that more accurately represents the dot plots cited in the literature.

KEY WORDS: Dotplot; Graphics; Histogram; Kernel density estimation.

1. INTRODUCTION

Dot plots have a long history. Jevons (1884) used dot plots to graph the weight of British Sovereign coins by year. Dot plots have appeared in statistical texts such as Tippett (1944), Tukey (1977), Box, Hunter, and Hunter (1978), and Mosteller and Hoaglin (1991), as well as in various scientific sources (e.g., Uman 1969). They have been widely used in the medical literature (e.g., Krieg, Beck, and Bongiovanni 1988; Chastre et al. 1988).

The dot plot discussed in this article displays individual observations on a continuous scale using a dot or other symbol. Its distinguishing feature is the use of local displacement in a direction orthogonal to the scale in order to prevent dots from overlapping. This displacement is either symmetric, which produces a string of dots resembling a belt of beads, or asymmetric, which produces a stack of dots resembling a density. Figure 1 shows examples of these two types of dot plots: the *symmetric dot plot* and the *asymmetric dot plot* (or dot density). The

data are from Allison and Cicchetti (1976), available at http://lib.stat.cmu.edu/datasets/sleep. The variable plotted represents hours of slow-wave (nondreaming) sleep per day among 62 selected mammals. There is one dot in each plot for each of the 48 mammals with nonmissing data. The only difference between the two plots in the figure is that the dots in the upper panel are stacked on the axis and the dots in the lower are symmetrically aligned.

Dot plots now appear in several commercial statistical packages. These programs do not correctly reproduce the dot plots existing in the literature. Instead, they use regular binning to produce plots that resemble the line printer asterisk histograms in older mainframe statistics packages. Sasieni and Royston (1996), for example, described a dot plot algorithm that is based on regular binning. Their results are histograms, not dot plots. They can be reproduced by specifying a particular (large) number of bars to a histogram program. For reference, I will call these histogram plots (in their asymmetric or symmetric form) *histodot plots*, because they are histograms whose bars are drawn as stacks of dots. Histodot plots are recognizable by their regular horizontal spacing of the dots, which is determined by the histogram binning.

Figure 1. Asymmetric dot plot (upper panel), symmetric dot plot (lower panel).

Leland Wilkinson is Senior Vice President, SYSTAT Products, SPSS Inc., 444 North Michigan Avenue, Chicago, IL 60611 (Email: leland@spss. com); and Adjunct Professor, Department of Statistics, Northwestern University, Evanston, IL 60201. Laszlo Engelman, John Hartigan, and David Scott gave helpful comments. An anonymous reviewer provided the Tukey and Tukey reference. Stephen Stigler provided the Jevons reference. The graphics in this article were produced with SYSTAT.

Instead of histograms, it is more useful to think of dot plots as horizontal, one-dimensional scatterplots where tied values are perturbed or displaced vertically (Cleveland 1985). When points overlap, we may displace them by adding a small amount of uniform random error (Chambers, Cleveland, Kleiner, and Tukey 1983), we may displace them systematically in a textured pattern (Tukey and Tukey 1990), or we may displace them in increments of one dot width (the method used in the cited hand-drawn examples). These three methods are usually called, respectively, *jittered plots*, *textured dot strips*, and *dot plots*. Unlike histodot plots, all three of these methods position an outlier (or any case separated from the rest of the data) exactly where it should be on the scale rather than at a lattice point defined by binning.

Because dot plots are different from histograms, producing them on a computer involves different algorithms. I will present closed-form expressions for dot densities and discuss smoothing and determining dot size based on sample size. The result of these procedures (using symmetric or asymmetric displacement) is a plot that reproduces those appearing in the literature and which has a theoretical basis in the density estimation literature.

2. DENSITY PLOTS

To understand dot plots, it is most helpful to compare and contrast them with histograms and kernel density estimates. I will use a common notation that reveals similarities and differences in all three types of displays. For each, I will take the approach of representing a density by counts rather than framing it as a probability density estimation problem. Dot plots are designed and used for displaying data values directly; they were not intended as density estimators and would be ill-suited for this purpose.

2.1 Histograms

Histograms can be viewed as a binning procedure where the bins are connected intervals. In the following expression for the frequency histogram, the w function serves as an indicator for the bin in which a given value of x falls. In the ordinary histogram, the bin intervals are of equal width.

$$f(x) = \sum_{i=1}^{n} w\left(\frac{X_i - g(x)}{h}\right),$$

where $X_i \in \mathbf{X}$, a finite set of n values, with

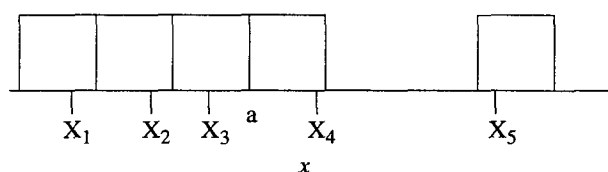$$w(u) = \begin{cases} 1 & \text{if } 0 \leq u < 1 \\ 0 & \text{otherwise} \end{cases},$$

and

$$g(x) = x_0 + h\left\lfloor \frac{x - x_0}{h} \right\rfloor,$$

and $h =$ bin width, and x_0 is a location parameter.

The histogram has an arbitrary location parameter, x_0. Ordinarily, this parameter is 0, so that one cutpoint is located at 0 on the scale. This location parameter affects the shape of the histogram. The *averaged shifted histogram* (ASH) was devised by Scott (1985) to ameliorate this scale-shape dependency. The ASH is produced by varying x_0 within an h-wide interval and averaging the heights of the bars resulting from each value of x_0.

Figure 2 shows a histogram for five observations using equal width bins and $x_0 = a$, an arbitrary location. All the bars are the same height for these data. If x_0 is set to approximately $a + h/2$, however, the second bar from the left will double in height because X_2 and X_3 are closer together than one bin width (h). Consequently, the third bar from the left will be empty and the histogram will have a different shape from that in the figure.

2.2 Kernel Densities

Kernel densities are produced by using a kernel function K for the w function. The simplest form is a uniform kernel, which acts as a local accumulator. Like the histogram, the uniform kernel produces a step function for the density. Unlike the ordinary histogram, the width of the steps produced is not uniform. The formula presented here has been called a "naive method" (Silverman 1986) because it uses a simple counting kernel instead of a probability density function.

$$f(x) = \sum_{i=1}^{n} w\left(\frac{X_i - g(x)}{h}\right),$$

where $X_i \in \mathbf{X}$, a finite set of n values, with

$$w(u) = \begin{cases} 1 & \text{if } |u| < .5 \\ 0 & \text{otherwise} \end{cases},$$

and $h =$ kernel width, and $g(x) = x$.

Figure 3 shows a uniform kernel data density for the same observations used in Figure 2. Notice that the kernel produces steps of different widths, depending on the location of the data.

2.3 Dot Densities

Dot densities use an indicator function for accumulation, like the uniform kernel and histogram, but the reference
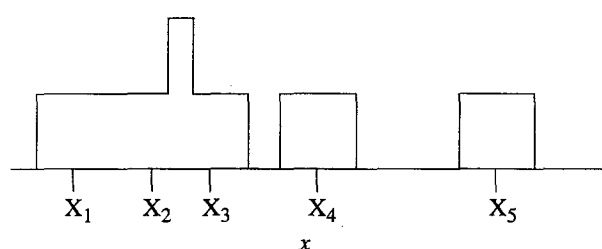

Figure 2. Histogram.
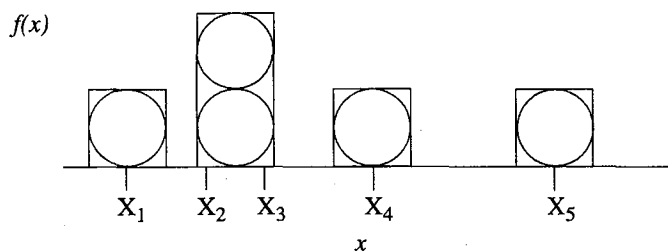

Figure 3. Uniform kernel density.

f(x)

Figure 4. Dot density.

point varies depending on the data. We use a locator function $g(x)$ to compute a reference point relative to any x. Specifically, for any x we look below it for a data point which is separated from all points below itself by more than h (which may end up being the smallest data point in the set). If this point is separated more than h from x, then we must move recursively toward x from the initially selected data point X_j in steps h wide until we find a data point X_k above it which is within h distance below x. The twiddle value v used in the locator function $g(x)$ centers a dot stack at the middle of the interval between X_k (the smallest value in the stack) and X_l (the largest value in the stack) as long as there is room to do so.

$$f(x) = \sum_{i=1}^{n} w\left(\frac{X_i - g(x)}{h}\right),$$

where $X_i \in \mathbf{X}$, a finite ordered set of $n + 1$ sorted values, including $X_0 \equiv -\infty$, and

$$w(u) = \begin{cases} 1 & \text{if } |u| < .5 \\ 0 & \text{otherwise} \end{cases},$$

h = dot diameter, and $g(x) = X_k + v$, where

$$X_j = \max_{1 \le i \le n} [X_i : X_i < x \quad \text{and} \quad X_i - X_{i-1} > h],$$

Hours of Sleep

Hours of Sleep
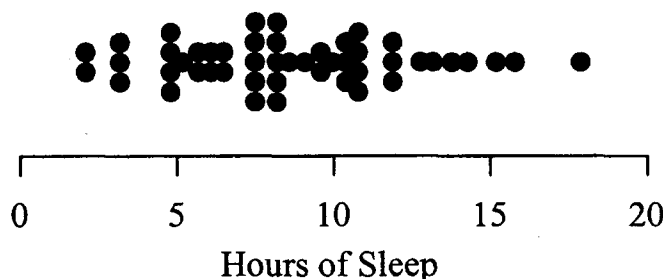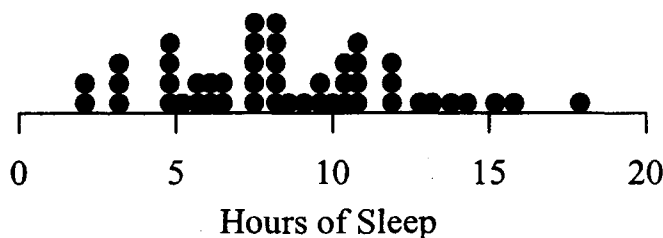
Figure 5. Overlapping asymmetric dot plot (upper panel), symmetric dot plot (lower panel).

and

$$X_k = \min_{1 \le i \le n} \left[X_i : x - X_j + h\left\lfloor\frac{X_i - X_j}{h}\right\rfloor < h\right],$$

and

$$X_l = \max_{1 \le i \le n} [X_i : X_i - X_k < h],$$

and

$$v = \begin{cases} (X_l - X_k)/2 & \text{if } X_k - X_{k-1} > h \\ 0 & \text{otherwise} \end{cases}.$$

Here is a summary of the algorithm:

1. Start with the smallest data value, $X_j = X_1$. The first stack of dots always begins here.
2. Count the number of values, n_j, within one dot's width (h) to the right of X_j.
3. Place n_j dots above X_j, or offset to the right of X_j by v if the n_j data values differ.
4. Move right to the next largest data value not included in the current stack of dots.
5. Repeat steps 2–4 until there are no data values left to plot.

There is one more step we can add to the algorithm. To reduce variance among the stacks (see Section 2.6), we can do a moving average smooth on the stacks of dots. This smoothing procedure serves a second purpose: it tends to remove differences between a left-to-right and right-to-left implementation of the segmentation algorithm. The smooth proceeds from left to right on all adjacent stacks of dots:

$$f_m(x) = f(x) + \left\lfloor\frac{f(x + h) - f(x)}{2}\right\rfloor,$$

for all $m = 1, \ldots, q$ adjacent stacks.

Adjacency is defined as two dot stacks within $h/4$ distance of each other. This amounts to exchanging dots between adjacent stacks to minimize differences. The floor function ensures that the exchange preserves the direction of the difference for odd counts of dot differences.

Figure 4 shows a dot density for the data used in the previous figures. Dots have been superimposed to illustrate their placement. Because there are no adjacent stacks, no averaging would occur if we were to apply the final averaging step.

2.4 Overlapping Dots

Some authors draw dot plots with partially overlapping dots. This helps locate dots closer to their actual data value on a scale and produces slightly different stacking patterns. Overlapping up to half a dot width is customary. Producing half-overlapping plots requires changing h to $h/2$ in the dot plot algorithm. The plotting symbols remain h wide and h tall. Figure 5 shows the result of this modification to the same data used in Figure 1. For small samples, overlapping distinguishes true dot plots even more from histodot plots. A similar modification could be made to allow dots to overlap vertically, perhaps to change the aspect ratio of the plot. I have not seen this done in published applications.
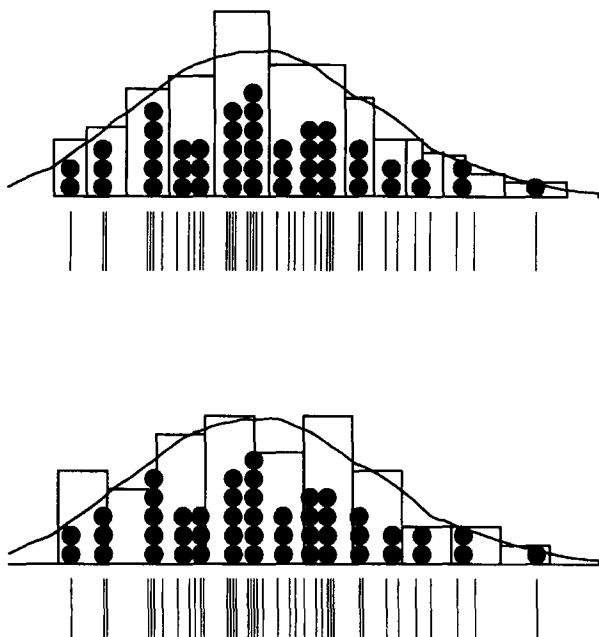
Figure 6. Dot densities, kernels, empirical densities, and histograms superimposed (histogram in upper panel is based on gap binning, lower is based on regular binning).

2.5 Comparing Densities

Figure 6 shows dot densities of the sleep data superimposed on other density graphics. Below each density is a stripe density plot (empirical density function) showing the location of the data. An Epanechnikov kernel smooth (Silverman 1986) is superimposed on both plots.

The lower plot, an ordinary equal-width bins histogram, fails to reveal the gaps in the data. The upper plot, a gap histogram, is produced by cutting the distribution at the k largest gaps between data points and using these cutpoints to construct bars whose areas are proportional to the counts in them. The gap histogram helps illustrate the spirit of the dot plot. Like the uniform kernel smooth, it follows the shape of the dot plot more closely than the ordinary histogram. The Epanechnikov probability kernel smooth, on the other hand, is intended as a density estimator. On small samples, dot plots reveal the data. On large samples, kernel smooths reveal the theoretical distribution. Combining smooths with dot plots on medium sized samples can exploit the advantages of both methods and replace histograms, except where regular bin cutpoints are needed to highlight fractiles of a distribution.

2.6 Choice of Dot Size (h)

In the context of density estimation, h is called a *smoothing parameter*, because it controls the smoothness of the fit-

ted density for a given sample size. The density estimation literature has investigated this parameter extensively for histograms (e.g., Sturges 1926; Doane 1976; Freedman and Diaconis 1981; Scott 1992) as well as for kernel estimates (Silverman 1986; Scott 1992; Wand and Jones 1995). The literature suggests that for roughly symmetric, unimodal densities, the optimal value of the histogram smoothing parameter is of order $n^{-1/3}$. With a normal distribution, for example, Scott (1979) suggested $h = 3.5sn^{-1/3}$, where s is the sample standard deviation. A smaller value is needed for asymmetric distributions.

We need to approach the problem from a different direction, however. Because dot plots are like tallies, the dot size parameter h affects not only the smoothness of the dot plot but also its shape. The larger the dots, the taller the dot density. We cannot change the aspect ratio of a dot plot without changing the shape of its dots. Assuming we wish to use circular dots, then we need a dot size on the order of $n^{-1/2}$. The argument for this is simple: the rectangular packing of n dots inside a unit square requires a dot diameter of $n^{-1/2}$. If we represent a standard normal density $\phi(z)$ inside the same square so that a z scale from -4 to 4 spans the bottom of the square and the maximum height of $\phi(z)$ is approximately $1/5$ its width (a fairly typical aspect ratio for normal distributions in textbooks), then we need a dot size of approximately $.25n^{-1/2}$. This calculation is based on rounding the ordinate of the normal density at 0 to $.4$ and its area between $(-4, 4)$ to 1. Although it is based on the normal, this dot size works well for a variety of distributions because it is small enough to prevent overflowing a plotting window in the face of moderate skewness. Nevertheless, a well-written dot plot program should automatically downsize dots when extreme overflow occurs.

Figure 7 shows the result of this dot size for n's up to 10,000 using a normal distribution. This strategy tends to produce more dot stacks than the optimal estimates suggested for histograms. This was the motivation for adding the smoothing procedure to remove some of the random spikes from the display.

2.7 Badness of Fit

The literature (e.g., Freedman and Diaconis 1981; Scott 1992) indicates that histograms have *mean integrated square error* (MISE) on the order of $n^{-2/3}$ and that kernel smooths have MISE on the order of $n^{-4/5}$. We should expect dot densities to do somewhat worse than both. Figure 8 shows the results of a Monte Carlo simulation using the dot plot algorithm presented previously on normal distributions with a dot size of $.25n^{-1/2}$. Five replications were done at
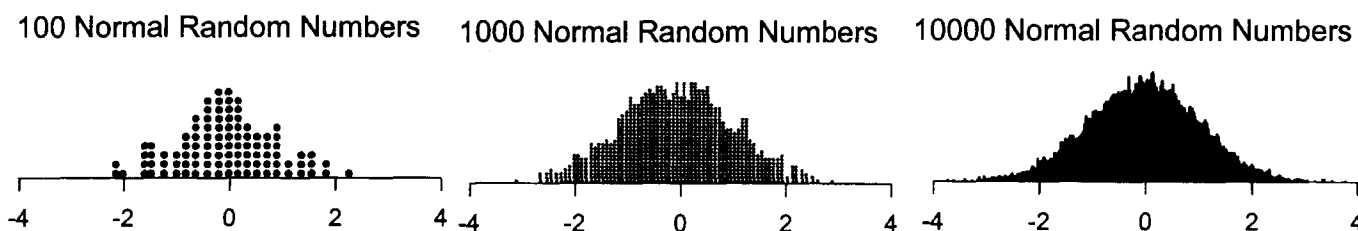
100 Normal Random Numbers　**1000 Normal Random Numbers**　**10000 Normal Random Numbers**
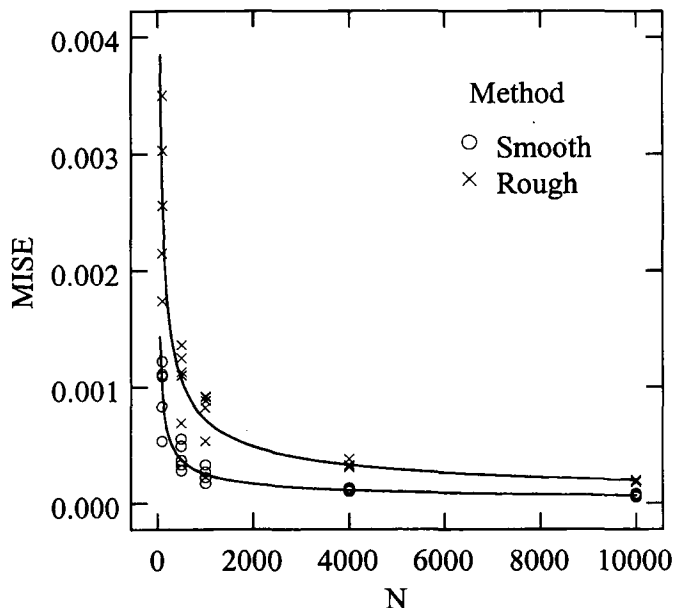

Figure 7. Large sample dot densities.

Figure 8. Dot density MISE as a function of N.

each sample size (100, 500, 1,000, 4,000, 10,000). The lower curve shows the results when the final moving average dot smoothing is applied and the upper curve shows the results when it is not. The curves were fit with a power function predicting MISE from N. The exponent for the upper curve is $-.56$ and for the lower, $-.53$. The results suggest that, for normal distributions, dot densities have MISE on the order of $n^{-1/2}$.

Following Freedman and Diaconis (1981) and Scott (1992, p. 54), we can separate MISE into two parts: MISE = IV + ISB, where IV is *integrated variance* and ISB is *integrated squared bias*. Scott shows that for histograms having bin width h of order $n^{-1/3}$,

$$\text{IV} \sim O\left(n^{-2/3}\right),$$

and

$$\text{ISB} \sim O\left(n^{-2/3}\right).$$

If we use h of order $n^{-1/2}$, as we do for dot plots, then Scott's formulas yield

$$\text{IV} \sim O\left(n^{-1/2}\right),$$

and

$$\text{ISB} \sim O\left(n^{-1}\right).$$

If dot plots follow roughly the histogram behavior, we should expect that the dot plot has high variance and low bias, exactly what we want for a data display as opposed to a density estimator.

3. EXAMPLE USES OF DOT PLOTS

Dot plots are especially suited for supplementing other graphics. Figure 9 shows dot-box plots (Wilkinson 1992) bordering scatterplots of the Allison and Cicchetti (1976) data on raw and decimal log scales. The scatterplots consist of body weights and brain weights of the mammals. The dot-box graphic superimposes a symmetric dot plot on a Tukey box plot. The symmetric dot plots contribute shape information to the box plots. Bordering the scatterplot with dot-box plots can help guide the search for a normalizing power transformation in an interactive computing environment. The dots fall at their proper locations regardless of the power transformation applied to the scale. Figure 9 also illustrates why histogram binning is unsatisfactory for constructing dot plots: a proper algorithm must ensure that the single dots in the border plots align with those in the scatterplot.

Figure 10 shows a classification tree (Breiman, Friedman, Olshen, and Stone 1984) analysis of the sleep data. The tree predicts a danger index (1 = unlikely to be killed by a predator, 5 = likely to be killed) from type of sleep (slow wave sleep and dreaming sleep) and the body and brain weight variables in Allison and Cicchetti (1976). In each frame node of the tree is a dot density. The advantage of dot densities in this context is that they work well for both continuous and categorical variables. Unlike ordi-
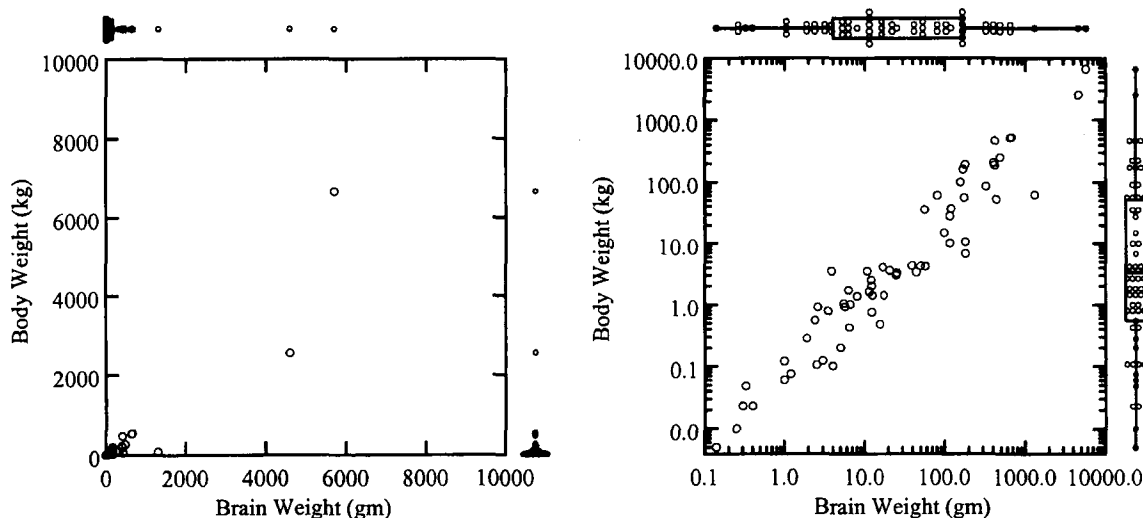

Figure 9. Dot-box bordered scatterplot.

nary histograms, dot densities have one stack of dots per category because they bin only where the data are.

This tree is called a *mobile* (Wilkinson 1999). This display format gets its name from the hanging sculptures created by Calder and other artists. If the squares were boxes, the dots marbles, the horizontal branches metal rods, and the vertical lines wires, the physical model would hang in a plane as shown in the figure. This graphical balancing format helps identify outlying splits in which only a few cases are separated from numerous others. Each box contains a dot density based on a proper subset of its parent's collection of dots. The scale at the bottom of each box is the danger index running from 1 (left) to 5 (right). This graphic is intended for a color display. Each dot is colored according to its terminal node at the bottom of the tree so that the distribution of predicted values can be recognized in the mixtures higher up in the tree.
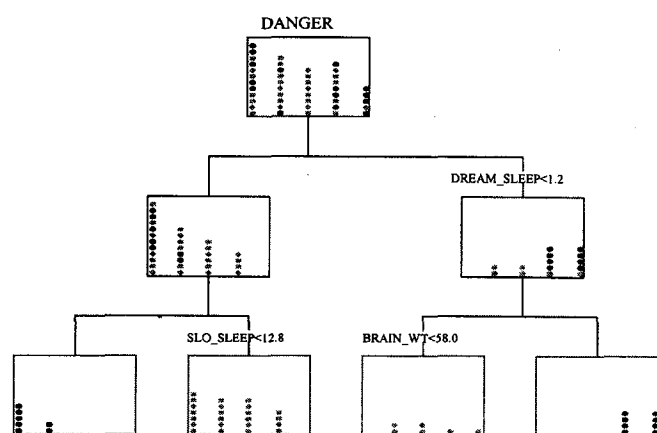

Figure 10. Classification tree with dot histograms.

4. CONCLUSION

Like stem-and-leaf diagrams, dot plots are easier to draw by hand than by computer. By defining dot plots using density-estimation notation, we can explicitly formulate an algorithm that produces plots that are closer to published hand drawings than those produced by binning methods. Moreover, this definition yields some insight into the behavior of dot plots. As high variance, low bias data representations, they are ideally suited for displaying moderate sized datasets when outliers and other irregularities are of interest.

[Received December 1996. Revised December 1998.]

REFERENCES

Allison, T., and Cicchetti, D. (1976), "Sleep in Mammals: Ecological and Constitutional Correlates," *Science*, 194, 732–734.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, New York: Wiley.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984), *Classification and Regression Trees*, Belmont, CA: Wadsworth.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983), *Graphical Methods for Data Analysis*, Belmont, CA: Wadsworth.

Chastre, J., Fagon, J. Y., Soler, P., Bornet, M., Domart, Y., Trouillet, J.-L., Gibert, C., and Hance, A. (1988), "Diagnosis of Nosocomial Bacterial Pneumonia in Intubated Patients Undergoing Ventilation: Comparison of the Usefulness of Bronchoalveolar Lavage and the Protected Specimen Brush," *The American Journal of Medicine*, 85, 499–506.

Cleveland, W. S. (1985), *The Elements of Graphing Data*, Monterey, CA: Wadsworth Advanced Books.

Doane, D. P. (1976), "Aesthetic Frequency Classifications," *The American Statistician*, 30, 181–183.

Freedman, D., and Diaconis, P. (1981), "On the Histogram as a Density Estimator: L_2 Theory," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandt Gebiete*, 57, 453–476.

Jevons, W. S. (1884), "On the Condition of the Gold Coinage of the United Kingdom, With Reference to the Question of International Currency," in *Investigations in Currency and Finance*, ed. W. S. Jevons, London: Macmillan (reprinted 1964 by Augustus M. Kelley, New York).

Krieg, A. F., Beck, J. R., and Bongiovanni, M. B. (1988), "The Dot Plot: A Starting Point for Evaluating Test Performance," *Journal of the American Medical Association*, 260, 3309–3312.

Mosteller, F., and Hoaglin, D. C. (1991), "Preliminary Examination of Data," in *Fundamentals of Exploratory Analysis of Variance*, eds. D. C. Hoaglin, F. Mosteller, and J. W. Tukey, New York: Wiley, p. 43.

Sasieni, P. D., and Royston, P. (1996), "Dotplots," *Applied Statistics*, 45, 219–234.

Scott, D. W. (1979), "On Optimal and Data-Based Histograms," *Biometrika*, 66, 605–610.

—— (1985), "Averaged Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions," *The Annals of Statistics*, 13, 1024–1040.

—— (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, New York: Wiley.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Sturges, H. A. (1926), "The Choice of a Class Interval," *Journal of the American Statistical Association*, 21, 65–66.

Tippett, L. H. C. (1944), *Statistics*, Cambridge, MA: Oxford University Press.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Tukey, J., and Tukey, P. (1990), "Strips Displaying Empirical Distributions: I. Textured Dot Strips," Technical Memorandum, Bellcore.

Uman, M. A. (1969), *Lightning*, New York: McGraw-Hill. Reprinted by Dover Books, 1984.

Wand, M. P., and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman & Hall.

Wilkinson, L. (1992), "Graphical Displays," *Statistical Methods in Medical Research*, 1, 3–25.

—— (1999), *The Grammar of Graphics*, New York: Springer-Verlag.