ROUTE-TYPE AND LANDSCAPE-TYPE SOFTWARE FOR LEARNING STATISTICAL DATA ANALYSIS ®

Arthur Bakker
Utrecht University
The Netherlands

This paper contrasts two types of educational tools: a route-type series of so-called statistical minitools (Cobb et al., 1997) and a landscape-type construction tool, named Tinkerplots (Konold & Miller, 2001). The design of the minitools is based on a hypothetical learning trajectory (Simon, 1995). Tinkerplots is being designed in collaboration with five mathematics curricula and is open to different approaches. Citing experiences from classroom-based research with students aged ten to thirteen, I show how characteristics of the two types of tools influence the instructional decisions that software designers, curriculum authors, and teachers have to make.

1. INTRODUCTION

Professional statistical software packages are not suitable tools for young students to use to *learn* data analysis. How could 11-year-olds ever choose between a histogram, a box plot, or a pie chart, if they do not yet understand when these representations are useful? For this reason, special software has been designed for younger students. These tools, unlike professional packages, are designed to enhance learning and not just to get a data analysis job done. However, to design educational software we need to know the critical issues of learning data analysis in the classroom. In this paper, I contrast two recently developed tools, the statistical minitools (Cobb et al., 1997) and Tinkerplots (Konold & Miller, 2001), and examine what my research implies for the designing of, the teaching with, and curriculum writing for such data analysis tools.

Section 2 describes the essential features of the two tools. The intent of both tools is to allow students to start with what they can invent themselves and work towards the use of more conventional graphs and statistical notions. However, the two tools take different approaches to how this might be achieved. To describe these different approaches, I characterize the statistical minitools as a route-type tool series and Tinkerplots as a landscape-type tool.

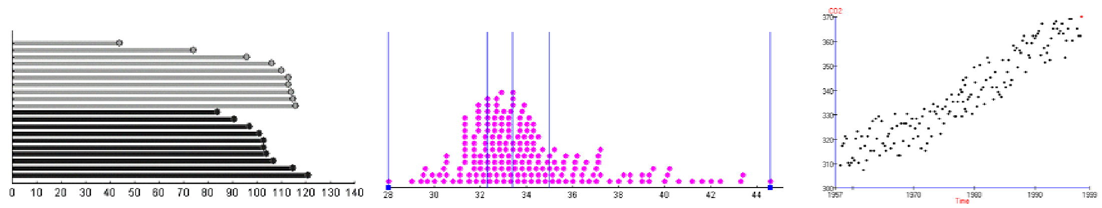
In section 3, I argue in which respects the route-type minitools need more flexibility, and in which situations the flexibility of Tinkerplots might cause difficulties for students, teachers, and curriculum authors. The argumentation is mainly based on teaching experiments with the minitool 1 and 2 in fourteen Dutch classes with four different teachers. The Dutch experiments consisted of a series of ten to fifteen lessons per class (age eleven to thirteen) in a normal school situation. The experience with Tinkerplots is limited to visits to three classrooms in the USA with students aged ten to twelve. In the last section, I reflect on the contrast between the tools and arrive at some recommendations.

2.1 ROUTE-TYPE SOFTWARE: STATISTICAL MINITOOLS

Gravemeijer, Cobb, and co-workers (1997) designed the three statistical minitools and accompanying activities with a hypothetical learning trajectory in mind (see Simon, 1995). In brief, a hypothetical learning trajectory entails possible starting points, end goals, and an anticipation of how every activity can contribute to progress towards the end goals from every previous step. The overarching idea that functioned as a guideline for the hypothetical learning trajectory was the notion of distribution (Cobb, 1999).

Figure 1 shows three different types of plots. The plot made by minitool 1 is called a value bar graph, because each case is signified by a bar whose relative length corresponds to the value of the case. Minitool 2 provides a stacked dot plot in which each case is represented as a dot positioned over a labeled number line. Minitool 3 shows a scatter plot in which the values of a case on two numeric variables are represented in a Cartesian coordinate system. Notice that minitool 1 does not provide the type of bar graph in which each bar represents a number of cases; here each bar is one case. The students can organize the data in minitool 1 with several options such as sorting by size or color, or making their own groups. In minitool 2, some of the grouping options form precursors to conventional plots such as histograms and box plots. Unconventional grouping options include making your own groups and using fixed group size. Minitool 3 allows, for instance, four equal groups within a chosen number of vertical slices, so students can compare a sequence of distributions.

The design of the minitools may be described by Lacan's semiotic theory (Lacan, 1965). According to this theory, once a student has developed one type of sign (a plot for example), its meaning can "slide over" the next signifier in the sequence, thus forming a new sign. In this way, complex meanings are developed in chains of signification. Applying this theory to instruction, students are guided in "reinventing" conventional plots such as histograms, box plots, and scatter plots. In the teaching experiments of Cobb, Gravemeijer, and co-workers, problems were typically solved with one minitool, with just one type of plot that was meaningful for the students at that stage of the hypothetical learning trajectory. With this background, I would say that the minitools support a *route type* of learning.

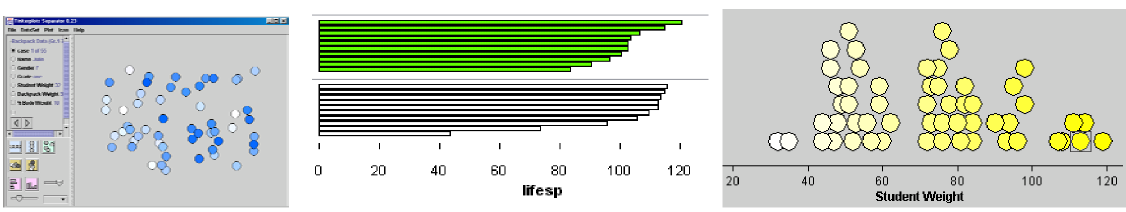


1) 2) 3)

Figure 1. Minitool 1, 2, and 3 with 1) a value bar graph, 2) a stacked dot plot with four equal groups, and 3) a scatter plot.

2.2. LANDSCAPE-TYPE SOFTWARE: TINKERPLOTS

The opening screen of Tinkerplots (Figure 2a) displays one icon for each case in a data set, haphazardly arranged in a space free of axes (cf. Tabletop, Hancock, 1995). Data sets are typically multivariate, whereas in the first two minitools usually two univariate data sets are compared. Students can separate, stack, and order the icons in two directions (up-down and left-right), and they can use different icon types, including value bars and dots. They can easily switch between value bar graphs, dot plots, and scatter plots without opening a new tool. Using the "fuse" option, students can also switch between plots composed of individual case icons (case value plots) and plots composed of aggregate cases, such as histograms and pie charts (aggregate plots).



a) b) c)

Figure 2. a) Opening Screen of Tinkerplots, b) a Value Bar Graph, and c) a Stacked Dot Plot.

The authors of Tinkerplots describe it as a construction tool, because it offers students many possibilities for making their own, often unconventional, graphs (Konold, 1998). Almost all options provided in the three minitools are available at once in Tinkerplots (see Figures 2b and 2c). In this way, teachers and authors of five collaborating mathematics curricula have the flexibility to make their own instructional sequences. Thus, in contrast to the minitools, the design of Tinkerplots does not assume a particular learning trajectory. The difference between the minitools and Tinkerplots can be characterized as a route versus a landscape approach. In this metaphor, Tinkerplots resembles a landscape that allows many routes (for the landscape metaphor see Fosnot & Dolk, 2001).

3.1 MORE FLEXIBILITY IN DESIGN AND USE OF THE MINITOOLS

Support for the Original Design

Some educators, after an introduction to the minitools, wondered why the first minitool was necessary, since they thought the second minitool was easier and closer to most conventional graphs. The experience, however, is that using minitool 1 in addition to minitool 2 makes a difference.

First, having students compare different graph types proved fruitful, for instance in developing an informal notion of distribution (Bakker, 2001). This finding is consistent with the recent trend of using multiple representations (e.g. Van Someren et al., 1998). Second, we reversed the order of minitool 1 and 2 in one class, as a way to test the importance of minitool 1. In this specific class, we introduced minitool 1 after two lessons with minitool 2. One student then publicly exclaimed, “but the bars are much easier, aren’t they”. This was in the context of working with means. In many mini-interviews I conducted in different classes, students said that finding the mean visually was easier with value bars than with dots. None of the interviewed students expressed a contradicting opinion. The visual estimation they had developed was a compensation strategy of cutting off the parts sticking out and “giving these to the shorter bars” (Bakker, 2001). This is one good reason to use bars (as in minitool 1) for estimating and understanding the mean instead of the balance model (in a dot plot like minitool 2). Additionally, many students preferred the bars to the dots in other contexts as well.

Third, extra support for having bars available comes from experiments with two sixth-grade classes using Tinkerplots (age about 11). As described in section 2.2, the Tinkerplots opening screen presents students with unorganized dots on the screen. The discussed data set included student weight, grade, backpack weight, and sex as variables—the question being whether older students tend to carry heavier backpacks. The students could choose different icons with bars as one of the options. Almost all students used value bars rather than dots for their final representations. The conclusion from these three points is that the bar representation is necessary in addition to dot plots, and that it makes sense to start with the value bars, at least for younger students.

There are several possible explanations why students find value bars easier to use and understand than dots. A first, simple explanation is that students have seen and made many bar graphs before, and probably few dot plots. A second explanation could be that ordered bars provide strong visual support whereas this ordering is lost with dots. One student said, “I find the bars clearer; with the dots it looks mixed up.” A third possible explanation is that students without a good understanding of the coordinate system find dots harder to interpret than value bars. Historically, bar graphs have been in use longer than dot plots. This may reveal that bar graphs are intuitively more accessible. I have found no dot plots before 1884 (Wilkinson, 1999), whereas Playfair had already made bar charts in 1786 (Tufte, 1983). See also Galton’s graph of Figure 3 for displaying the normal distribution. He represented twenty-one ideal data points with bars, so the normal distribution has an ogive shape instead of the now common bell shape. These three explanations and one historical consideration might elucidate why bars are easier to understand for young students than dots. However, more research is needed to validate such explanations.

This subsection demonstrated that classroom experiences, to a certain extent, support the initial design of minitool 1 and 2. The following subsections give reasoning why certain route-type characteristics need revision.

Minitool 1 should Have Vertical Bars as Well

At the moment, minitool 1 only provides horizontal bars. I first explain why Cobb and colleagues designed minitool 1 like this, and then argue why it should include vertical bars as well. The main reason to program only horizontal bars was to smooth the transition from minitool 1 to minitool 2. If the bars in minitool 1 disappear (this is possible in the revised version of minitool 1), then the dots only need to drop down the horizontal axis to get the representation of minitool 2. In the Dutch experiments, most students easily understood this transition. The choice of horizontal bars restricts the activity contexts to ones that beg to be organized by horizontal bars, such as breaking distance or life span.

There are several reasons to implement vertical bars as well. First, when asked to draw their own graphs, most students used vertical bars, even if they had only worked with horizontal bars in minitool 1. This probably means that vertical bars are closer to students' experiential reality than horizontal bars. They have probably seen more vertical bar graphs before.

Second, many phenomena also beg to be organized by vertical bars, such as height and growth. This limits the number of contexts that can be used in minitool 1 at the beginning of a learning sequence considerably and it limits the power of students to express their ideas visually.

Third, my conjecture is that vertical bars better evoke the use and understanding of the *median*. I will motivate this conjecture with different kinds of arguments. One is the type of mistakes students make when using horizontal bars: they often take the midrange (middle of the range) instead of the median. My hypothesis is that this originates in reading from left to right instead of top to bottom. When reading from left to right, students often see the middlemost value as the middle between the two extremes, the midrange. With vertical bars, reading from left to right would more readily evoke the median than the midrange. See also Galton's graph with the median and vertical bars in Figure 3.

As an experiment, readers are invited to determine the median in Figure 4, both with horizontal and vertical bars. Next recall whether you counted the bars or the end points of the bars, and what your reading direction was. Did you look halfway between the first and last bar? My sense is that counting the bases of vertical bars elicits the ordinal aspect of the data, whereas counting the end points stresses the rational aspect of the data, and thus distracts from the median as an ordinal characteristic. This conjecture certainly asks for future research.

A possible advantage of horizontal bars is that students would be less inclined to confuse horizontal value bars with vertical frequency bars in histograms. One could ask in opposition if this confusion should be avoided this way. Discussing and clarifying such confusion can also be a learning opportunity, provided there is not too much other confusion at the same time.

The conclusion from this line of reasoning is that in minitool 1 vertical bars should be implemented as well, both increasing (Figure 1.1) and decreasing (Figure 2b) in size. In my opinion, the initial minitool 1 was too limited and the designed learning trajectory too narrow. This is not just a technical detail, but indicates possible disadvantages of route-type tools in general.

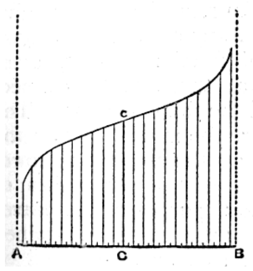


Figure 3. Galton's Graph of the Normal Distribution (1883) with Vertical Value Bars.

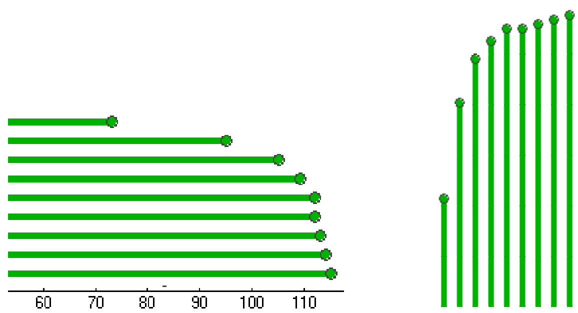


Figure 4. Find the Median with Horizontal or Vertical Bars.

Easier Switch Needed between Different Representations

Right now, the three minitools are different applets, which makes switching between representations cumbersome. There are good reasons to design smoother switching options between bars and dots in the minitools.

Different students take different routes, and teachers can benefit from this if they capitalize on different ideas for reflection during class discussions. The teachers and I therefore allowed students to select for themselves which minitool they would use. However, in practice students mostly stuck to one minitool, namely the one the materials suggested to try first. In a tool like Tinkerplots, students switch more easily between representations, including value bar graphs, dot plots, histograms, pie charts, and scatter plots. As already mentioned, contrasting various graphs turned out to be a fruitful activity for developing meaning of graphs, especially with respect to the notions of majority, outliers, and distribution (Bakker 2001). Easy transitions between different representations will increase the flexibility and expressive potential of the minitools. However, the students should make sense of the different representations before they can sensibly switch. This principle might cause difficulties when starting with a construction tool that allows many different representations.

3.2 POSSIBLE DIFFICULTIES IN USING TINKERPLOTS

Section 3.1 gave examples of possible disadvantages of a route-type tool. This section deals with possible difficulties of a landscape-type tool. In essence, the reasoning is as follows. If we want to start with students' own ideas and work towards conventional notions and graphs, then we should build in not only possibilities that students would use without prior statistical education, but also representations and grouping options that support the development towards conventional plots. For example, in both the minitools and Tinkerplots, students can group data in dot plots into equal intervals as preparation to the histogram. However, designing support for many possible learning routes might lead to an unworkable amount of possibilities. For teachers it would become difficult to know the tool so well that they can support students with very different explorations. With one example I illustrate that there can be good reasons for a limited environment at certain times.

Students were asked write to *Consumer Reports* regarding which of two battery brands they would recommend. Figures 1.1 and 2b show the data set with life span in hours of the two brands. This problem, with just one representation, caused a useful conflict. In most classes, about half of the students preferred one brand, because it had more batteries with a longer life span; the other half of the students preferred the other brand, because it had a higher mean, or was more reliable. Moreover, this single activity with just minitool 1 had already yielded an abundance of arguments and strategies that were useful as future leads towards more conventional notions such as mean, median, majority, spread, distribution, and even chance. This variety of student contributions already made it hard, especially for less experienced teachers, to orchestrate a good class discussion in which student arguments functioned as leads to new statistical ideas. Here, the constraint of just one type of representation turned out to be fruitful for student learning. On the other hand, comparing different representations can sometimes evoke a useful conflict. Then, what is being represented should be rather unproblematic for students. An example of this on weight data is given in Bakker (2001).

The key issue is whether students make sense of representations, and whether a statistical problem evokes a useful cognitive conflict (Fosnot, 1996). My conjecture is that, when using an open construction tool, too many cognitive conflicts can occur at the same time. The orchestration of class discussions can then become too diverse for students to understand the variety of ideas and representations, and for the teacher to move the class into a specific direction.

4. BEYOND THE CONTRAST: ROUTES THROUGH THE LANDSCAPE

The issues raised in the previous sections already indicate that this paper is not meant as an overview of the advantages of both tools. In this section, I reflect on some of the issues raised and arrive at some recommendations.

In practice, teachers follow the suggestions of the curriculum (school books). This saddles the curriculum authors with the challenge to keep students on task and treat possible conflicts systematically. A suggestion for the Tinkerplots design is therefore to have a gray-out option for either dots or bars. Then authors or teachers can decide whether to use the full landscape-type potential of Tinkerplots or the route-type constraints of the minitools. Perhaps students aged ten to twelve should begin working with constraint route-type software such as the minitools, and continue with landscape-type software like Tinkerplots. The choice of one or the other program also has to do with the aim. If we want to guide a class as a whole towards understanding specific notions and graphs, then a route-type tool might prove more suitable. When aiming at genuine data analysis with multivariate data sets from the start, a tool like Tinkerplots is more appropriate.

The contrast of route and landscape is not just characterizing software tools. It is also apparent in styles of teaching and designing. Cobb and colleagues investigated how specific mathematical practices evolved (Cobb, 1999) and used the notion of a hypothetical learning trajectory. Fosnot and Dolk (2001), on the other hand, promote the "landscape of learning" as opposed to linear approaches. For their type of education, a landscape-type tool seems a natural choice. However, it would be too simple to assume that a landscape-type tool like Tinkerplots would do the job, since Tinkerplots, as a tool, is rather neutral towards the type of learning it supports. The important issue is therefore how it is used.

In this paper, I have demonstrated that the described contrast influences the choices teachers, software designers, and curriculum authors have to make. I described how using a route-type tool led to seeking more flexibility in both the design and use of the minitools. How classroom-based research will further influence the design and use of a landscape-type tool such as Tinkerplots is a topic of future research.

ACKNOWLEDGMENTS

This research was funded by NWO, the Dutch National Science Foundation, under grant no. 575-36-03B. The opinions expressed do not necessarily reflect the views of the Foundation.

REFERENCES

Bakker, A. (2001). Symbolizing data into a ‘bump’. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th Conference of the International Group of the Psychology of Mathematics Education* (Vol. 2, p. 81-88). Utrecht, the Netherlands: Freudenthal Institute.

Cobb, P., Gravemeijer, K.P.E., Bowers, J., & McClain, K. (1997). *Statistical Minitools*. Designed for Vanderbilt University, TN, USA. Programmed and revised (2001) at the Freudenthal Institute, Utrecht University, the Netherlands.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1).

Fosnot, C.T. (1996). *Constructivism: Theory, perspectives, and practice*. New York: Teachers College Press.

Fosnot, C.T., & Dolk, M. (2001). *Young mathematicians at work: Constructing number sense, addition, and subtraction.* Portsmouth, NH: Heinemann.

Galton, F. (1883). *Inquiries into human faculty and its development* (2nd edn). London: J.M. Dent.

Hancock, C. (1995). *TableTop Software*. Novato, CA: Brøderbund Software Direct.

Konold, C. (1998). *Tinkerplots*: *Tools and curricula for enhancing data analysis in the middle school*. Grant proposal submitted to, and funded by, the National Science Foundation (# ESI-9818946). University of Massachusetts, Amherst: Author.

Konold, C., & Miller, C. (2001). *Tinkerplots* (version 0.23). Data Analysis Software. University of Massachusetts, Amherst (USA).

Lacan, J. (1965/1981). *The language of the self; the function of language in psychoanalysis*. Translated, with notes and comments, by Anthony Wilden (translation from 1968). Baltimore and London: John Hopkins University Press.

Simon, M.A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*, 114-145.

Tufte, E. (1983). *The visual display of quantitative information*. Cheshire: Graphics Press.

Van Someren, M.W., Boshuizen, H.P.A., Jong, T. de, & Reimann, P. (Eds.) (1998). *Learning with multiple representations*. Oxford: Elsevier Science Ltd.

Wilkinson, L. (1999). Dot plots. *The American Statistician*, *53,* 276-281.