Some Ways to Increase the Use of Graphs throughout the Introductory Applied Statistics Course

John D. McKenzie, Jr., Babson College
John D. McKenzie, Jr., Babson College, Babson Park, MA 02457

KEY WORDS: GAISE, graphical communication, statistical displays, tables, textbooks

Abstract

The GAISE College Report states that students in an introductory applied statistics course should know "how to graph the data as a first step in analyzing data, and how to know when that's enough to answer the question of interest" and "how to interpret … graphical displays of data - both to answer questions and to check conditions". It maintains that students should be able to communicate the results of these analyses, which often is done through graphs. Sadly, many textbooks have limited graphical (and tabular) displays after their introduction in a second or third chapter and hence such displays are not reinforced in many introductory courses, even though with today's technology it is quite easy to generate quality displays and modify them. In this paper some ways to increase the use of graphs and tables for both analysis and presentation in these courses will be presented.

GAISE Recommendations

The GAISE College Report states that students in an introductory applied statistics course should know "how to graph the data as a first step in analyzing data, and how to know when that's enough to answer the question of interest" and "how to interpret … graphical displays of data - both to answer questions and to check conditions". It maintains that students should be able to communicate the results of these analyses, which often is done through graphs.

Strengths and Weaknesses of Typical Coverage

There are numerous strengths of the typical coverage of graphs in an introductory applied statistics course, as reflected in the textbooks used in these courses. These strengths are the same regardless of the type of course. For example, one for general education or one for business and economics. First, an overview of graphs is usually presented at the beginning of the text. Most of today's textbooks present this overview in the second chapter, an improvement from presenting the graphs in the third chapter after basic summary statistics. And, today this overview usually includes scatterplots instead of waiting until the chapter on simple linear regression to introduce these important displays to the students. Another strength is the early inclusion of contingency tables instead of waiting until the chapter on the chi-square tests of two qualitative variables. The inclusion of one section on deceptive graphs in most textbooks is another strength.

But, even with these strengths, there are still many weaknesses in how graphs are introduced in most textbooks. First, they do not cover some important graphical displays. Probably the worst example of such an omission is the lack of time series plots. Leaving these displays until the chapter on time series analysis, that is rarely covered means that most students will not be exposed to a display that they will encounter in the real world. Another serious omission is an explanation of how to handle missing data.

Another problem with the current coverage is that while students are exposed to a large number of graphical displays there are often at a loss to know when to use each display. To supplement the graphical coverage in the current text being used at my institution, I have constructed a two-way table with the number of the variables crossed with the type of data (quantitative and qualitative).

In addition to identifying the names of the appropriate display in each cell, I include the Minitab command (or commands) used to construct that display. And, while I do not recommend that a textbook be oriented towards one specific piece of software, I believe that mention of such software is better than advocating the use of pencil and paper to construct graphical displays. For example, one text that was recently used at my institution described in detail how to determine the width of a histogram interval: "To determine the width of a class interval, you divide the range (the highest value [minus] the lower value) of the data by the number of class groupings desired." It would have far more appropriate to critique the default histograms obtained from a variety of packages.

Another weakness in the standard textbook coverage is lack of explanations of how to make professional presentation displays. Most books only deal with displays for analysis. And, while Microsoft has improved many statistical functions with Excel 2003,

texts should alert the students to be cautious about using Microsoft Excel to construct default presentation displays, most of which violate good graphing principles. As one example, consider the interesting doughnut chart found in Exhibit 1 below. It was created by using raw data, instead of summarized data. The suggested use of three-dimensional charts is another weakness in Excel.

Finally, and most important, the biggest weakness in the treatment of graphical displays in most textbook is the lack of such displays after the second chapter. Hence it is not uncommon to believe that graphs are not important components of a modern statistical analysis which they are. This limited reinforcement is present not just in the text but also in the book's exercises. As the GAISE College Report mentions "quite often a simple graph is sufficient to tell a story".

Nine Suggestions for Reinforcement

Below are nine ways that an instructor can supplement the typical coverage present in most introductory applied statistics textbooks to better educate his or her students on the use of graphs.

1. Five-Minute Starters

I often start off a class with a graphical display from a newspaper or magazine. These displays are constructed from real data that are also interesting and current to my students. This is an attempt to reinforce one component of statistical literacy (understanding the basic language of statistics). Namely, being able to interpret or critique statistical graphs.

2. Exploratory Data Analysis

I attempt to present an analysis display, that is, not a display appropriate for presentation, whenever I can. Graphs should complement summary statistics. They should accompany every inferential statistical analysis. As David Hildebrand explained to his students, PTDD (Plot The Damn Data).

3. Checking Conditions for Inference

One reason for such displays before each inferential analysis is to check whether it is appropriate to use this confidence interval or hypothesis test. Here JMP does an excellent job in contrast to other packages such as Minitab. It forces the user to examine a display, while Minitab only provides a readily accessible option.

4. Understanding Concepts

Graphical displays are most helpful in getting students to understand fundamental concepts of statistics. Understanding the meaning of a linear relationship between two variables by only presenting an equation, such as SATV = 41.6 + 0.924 SATM, is not as good as requesting students to sketch the regression line onto the scatterplot, present in Exhibit 2. More often than not some students will present lines completely above or below the linear scatter of points.

Another excellent way to teach your students a concept is to let them use an applet. Many of these incorporate well-designed graphs.

5. Supplemental Exercise Queries

Consider the following two exercises taken from two popular textbooks.

"a. What do the boxplots suggest about any gender differences in pulse rates?
b. Is it appropriate to analyze these data using the methods of inference in this chapter? Explain."

"At the 0.05 level of significance, is there evidence of a difference in the mean computer anxiety experienced by the female and male business students?"

The first exercise does a good job at reinforcing graphs in a chapter that introduces inference. It does not need to be supplemented. The second exercise does not do a good job. I often ask my students to submit an appropriate graphical display with the requested analysis.

6. Quiz and Exam Questions

If one believes that graphs should be incorporated throughout an introductory applied statistics course, it is important to include questions about graphs on quizzes and exams. I continue to be disappointed by the lack of any graphs in the test questions constructed by my colleagues (and present in the test bank supplements that accompany most textbooks). If something is important to be covered in class and assigned as part of a homework assignment, it should be assessed.

7. Required Displays in Written Reports

If an instructor assigns his or her students written reports, he or she should require the students to include graphical displays in their final report. These presentation displays should incorporate good graphing principles such as keeping the choice of displays simple to make them more effective in communicating their message. For example, consider Cleveland's ordering of elementary tasks:

1. Position along a common scale
2. Position along identical, nonaligned scales
3. Length
4. Angle – Slope
5. Area
6. Volume
7. Color hue – Color saturation – Density

Keep It Simple Statistician!

Likewise, these presentation displays should incorporate some common components, often never mentioned in textbooks, and hence missing from student submissions. Among these are an informative title; well-labeled axes; well-designed scales; a legend, if necessary, and a source. For example, having a title tell a story is far better than the following four alternatives: no title; the name of the graphical display; the name of the graphical display with the column(s) containing the graphed data; and the name of the graphical display with the names of the variables containing the graphed data. An instructor may decide to identify these components in the five-minute starters mentioned above.

Another example of good graphics often not incorporated in presentation graphs is the choice of graphing symbols. With approximately 5% of males color-blind, it is not enough to use color alone to distinguish between two entities. For example, consider the choice of symbols by Minitab in the time series plot in Exhibit 4.

8. Required Displays in Oral Reports

Similar to requiring graphical displays in reports, an instructor should require well-designed displays in their students' oral reports. They should not accept default PowerPoint graphical displays.

9. Feedback Displays

Finally, an instructor may reinforce the importance of graphs throughout the course by presenting feedback displays. These are displays that may compare results of homework assignments, quizzes, and exams. Consider the Minitab boxplot present in Exhibit 4. Similar feedback displays can also be constructed to compare exam questions or multiple section exams.

Conclusion

In this paper I have attempted to present some ways to increase the use of graphs and tables for both analysis and presentation in the introductory applied statistics course. It is my belief that incorporating one or more of these suggestions will be a small step in satisfying portions of the GAISE College Report.

Exhibit 1: An interesting Excel doughnut chart

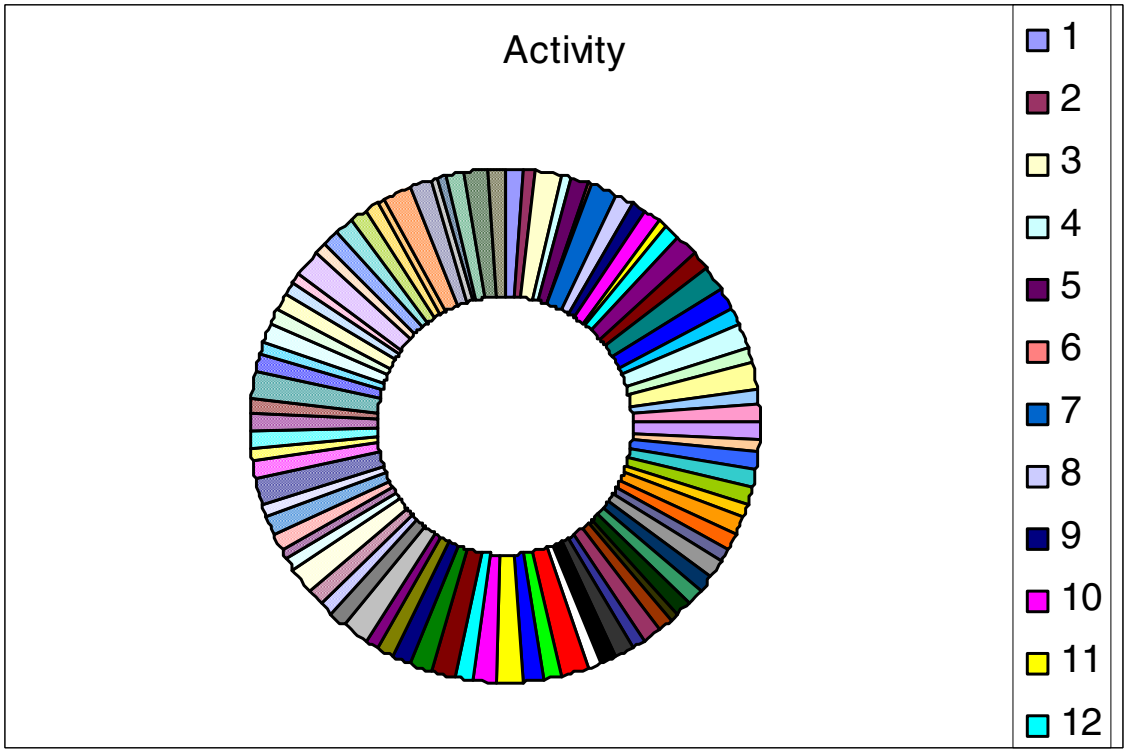

Exhibit 2: Minitab graphical display for students to plot SATV = 41.6 + 0.924 SATM equation

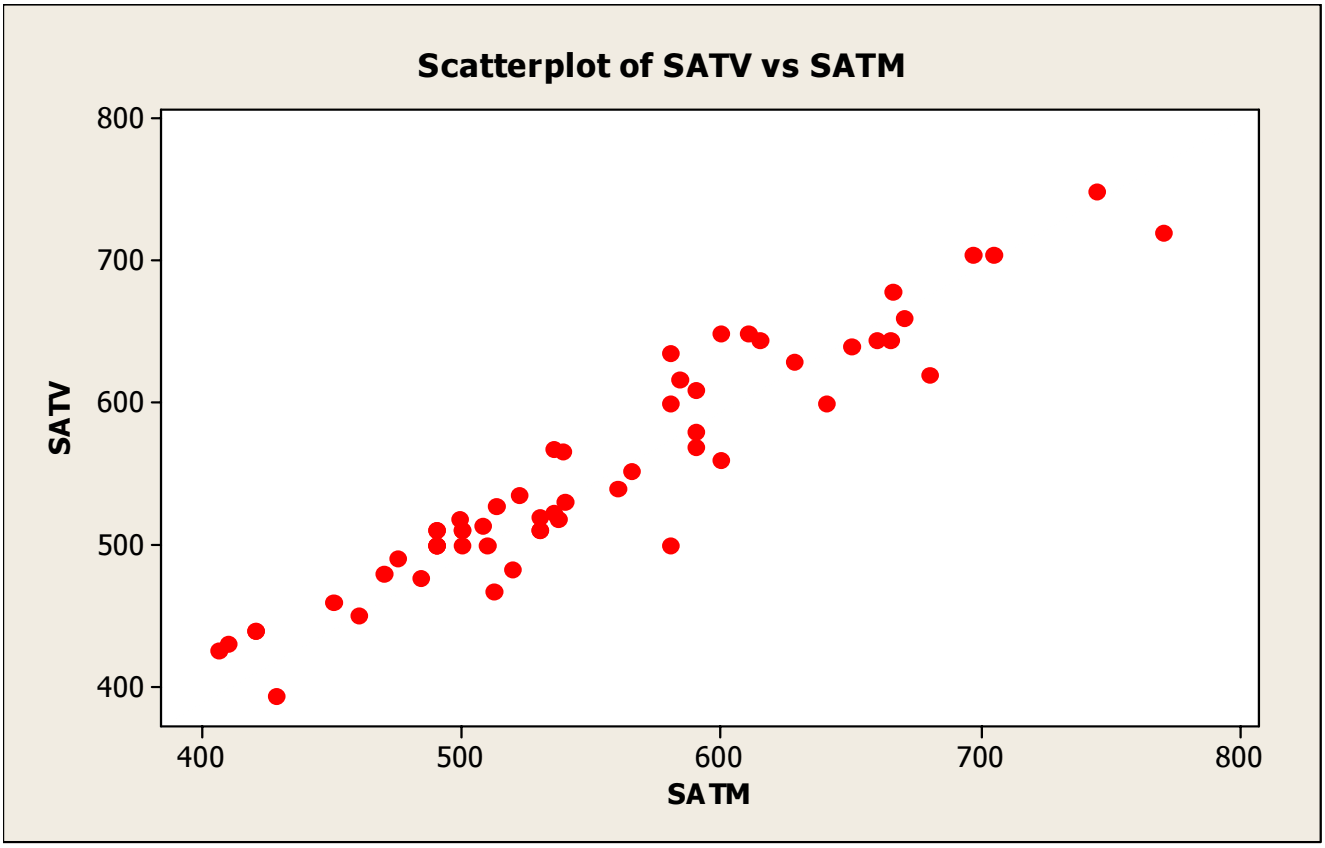

Exhibit 3: Minitab time series plot with correct choice of different symbols for time series plot

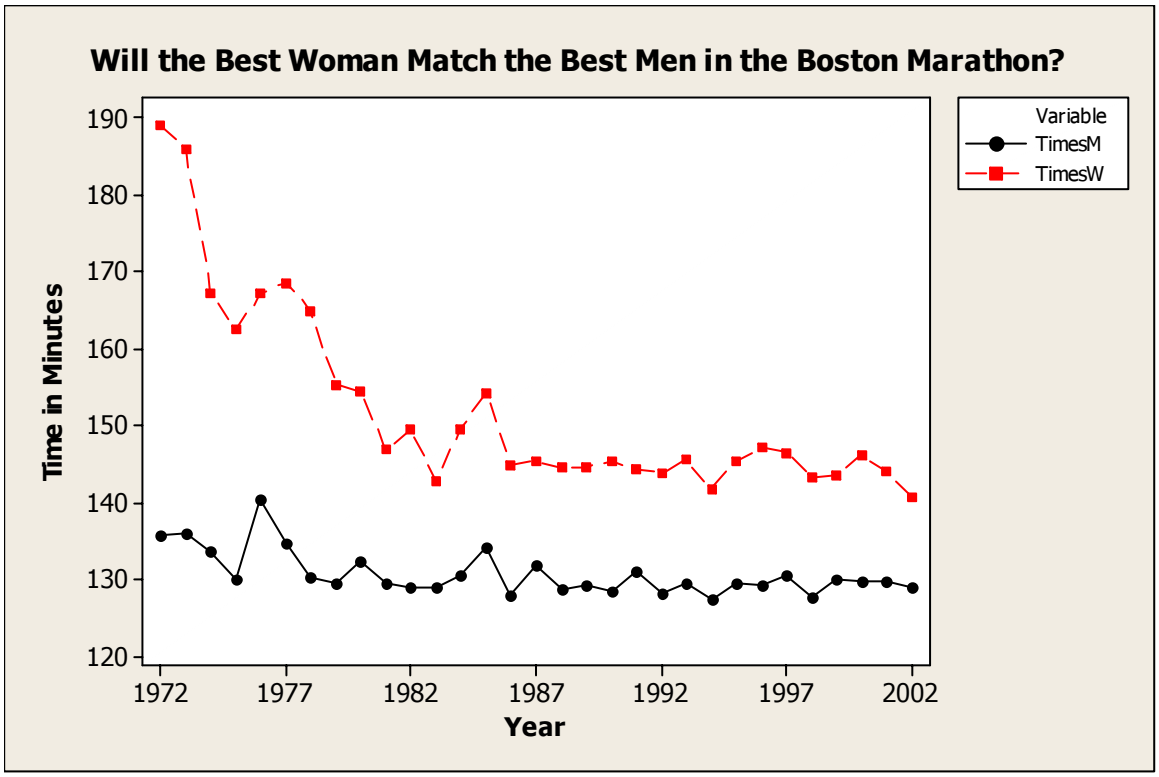

Exhibit 4: Minitab boxplot of adjusted homework assignments to illustrate feedback display