AC 2009-315: REASONING ABOUT CATEGORICAL DATA: MULTIWAY PLOTS AS USEFUL RESEARCH TOOLS

Richard Layton, Rose-Hulman Institute of Technology
Richard A. Layton is the Associate Director of the Center for the Practice and Scholarship of Education and an Associate Professor of Mechanical Engineering at Rose-Hulman Institute of Technology. His areas of scholarship include student team management, assessment, education, and remediation, laboratory reform focused on student learning, visualization of quantitative data, and engineering system dynamics. He is a guitarist and songwriter in the alternative rock band “Whisper Down”.

Susan Lord, University of San Diego
Susan M. Lord received a B.S. from Cornell University and the M.S. and Ph.D. from Stanford University. She is currently Professor and Coordinator of Electrical Engineering at the University of San Diego. Her teaching and research interests include electronics, optoelectronics, materials science, first year engineering courses, feminist and liberative pedagogies, and student autonomy. Dr. Lord served as General Co-Chair of the 2006 Frontiers in Education Conference. She has been awarded NSF CAREER and ILI grants. She is currently working on a collaborative NSF-funded Gender in Science and Engineering project investigating persistence of women in engineering undergraduate programs. Dr. Lord’s industrial experience includes AT&T Bell Laboratories, General Motors Laboratories, NASA Goddard Space Flight Center, and SPAWAR Systems Center.

Matthew Ohland, Purdue University
Matthew W. Ohland is an Associate Professor in the School of Engineering Education at Purdue University and is the Past President of Tau Beta Pi, the engineering honor society. He received his Ph.D. in Civil Engineering from the University of Florida in 1996. Previously, he served as Assistant Director of the NSF-sponsored SUCCEED Engineering Education Coalition. He studies longitudinal student records in engineering education, team-member effectiveness, and the implementation of high-engagement teaching methods.
Reasoning About Categorical Data:
Multiway Plots as Useful Research Tools

Key words: categorical data display, multiway plot, research methods

Abstract

In this paper, we help our audience learn to create and interpret “multiway plots”—a powerful tool for exploring and presenting categorical data. We use eighth-semester undergraduate persistence data as a case study of how multiway plots are used, but do not explore a particular research question. Instead, our goal is to disseminate a powerful, yet underutilized, research tool that facilitates an iterative process of reasoning about one’s categorical data: design the display/reason about the data/redesign the display/reason about the data etc., until the logic of one’s display is consistent with the logic of one’s analysis. Our case study begins with familiar column graphs and bar graphs typically used in engineering education journals. We then show how the same information is transformed into a multiway plot. Each step in the transformation is illustrated and explained. The results are two very different visualizations of the same data: clustered-column or bar graph (“before”) and multiway plot (“after”). Specific elements of the before and after graphs are highlighted to let the reader experience the perceptual advantages and to assess the utility of multiway plots in drawing meaningful conclusions from categorical data. We also alert our audience to the technical issues involved in creating multiway plots including software resources. Through this work, we hope to raise the awareness of the engineering education community of the benefits of multiway plots for visualizing, exploring, and presenting categorical data. In doing so, we hope to contribute to the continued enhancement of research quality in our discipline.

Introduction

Multiway plots are powerful tools for exploring and presenting categorical data. Developed by statistician William Cleveland\textsuperscript{1} based on work on human perception of quantitative data, multiway plots are respected by experts in data presentation such as Naomi Robbins\textsuperscript{2} and are consistent with graphical design principles advocated by Edward Tufte.\textsuperscript{3} However, a quick review of journals in the field of engineering education reveals that column and bar charts are the dominant form of data display. Our goal in this paper is to bring the form and function of the multiway plot to the attention of the engineering education community. We illustrate the advantages of this type of display, compared to clustered-column charts, for interpreting and presenting categorical data.

Categorical data is regularly encountered in engineering education research. For example, in studying the effects of gender and race on undergraduate student pathways, we might categorize students by sex, race, and their academic major in the eighth semester. Each of these categories has two or more levels that are mutually exclusive. For example, the category “Sex” has the levels Female and Male; the category “Race” might include the levels Asian, Black, Hispanic, and others; the category “Eighth-Semester Major” might have the levels Arts and Humanities, Business, Computer Science, Engineering, and others.
What distinguishes multiway data is that there is a single quantitative value for every combination of levels: one level from each category. In the student pathways example, the single quantitative value is the count (or number) of students of a particular combination of sex, race, and eighth-semester major—for example, the number of Black Females in Engineering. The goal of the multiway plot is to facilitate the study of how these counts relate to the categories. The method is especially useful when the categories have more than two levels. For example, a two-dimensional scatter plot suffices for comparing the behavior of women to men (in the aggregate) since this categorization by sex has only two levels and hence only two dimensions. However, as soon as we disaggregate the data by gender and race, the number of dimensions multiplies. The multiway plot facilitates presenting and exploring such multilevel data.

Goals. The main goal of this work is to bring to the attention of the engineering-education community the method for displaying categorical data called the multiway plot, and to illustrate the advantages of this type of display, compared to clustered-column charts, for the purposes of thinking about one’s data, displaying one’s data, and facilitating the iterative process of using the display to reason further about one’s data, leading to new displays. Also we hope to help our audience learn to interpret multiway plots in general and to alert the community to issues faced when using software to create these graphs.

Background and motivation. To make meaningful comparisons and conclusions from a graphical representation of data, a viewer has to decode the representation. Research in human perception shows that we do some of these decoding tasks with more accuracy than others. Cleveland and McGill\(^4\) identified several elementary graphical perception tasks and experimentally determined the accuracy with which subjects perform the tasks. In order of decreasing accuracy, these perceptual tasks are (as summarized by Robbins\(^2\)):

1. Position along a common scale
2. Position along identical, nonaligned scales
3. Length
4. Angle or slope
5. Area
6. Volume
7. Color hue, color saturation, density

This list helps us identify the problems associated with bar (and column) charts. Horizontal bars (or vertical columns) define areas on a graph, and the list shows that our perceptions of area are less accurate than our perception of other graphical elements. Moreover, the area of a bar contains no data—often only the endpoint of the bar represents data. The area of the bar just gets in the way of our ability to make comparisons.

This last sentiment echoes one of Tufte’s principles in his theory of data graphics: to “maximize the data-ink ratio”, that is, devoting as large a portion as possible of the ink in a graphic to the display of data information.\(^3\) Thus, in a bar (or column chart), the ink used to create the sides of the bar and the ink used to color the area of the bar does not necessarily convey data information. Tufte tells us, “above all else, show the data” and “erase non-data ink”.


The dot plot and multiway plot designed by Cleveland\(^1\) are both consistent with Tufte’s design principles and take advantage of the two perception tasks we do most accurately: comparing positions along a common scale and comparing positions along identical, nonaligned scales. Multiway plots are particularly useful for categorical data, as illustrated in this paper. Robbins agrees that dot plots and multiway plots are superior alternatives to bar and column charts for categorical data.\(^2\)

A multiway plot is an ordered array of individual dot plots. As a preview, a sample of an individual dot plot is shown in Figure 1, showing the number of White Female undergraduates—all of whom matriculated in an engineering major—who, by their eighth-semester, were enrolled in the majors shown. For example, about 9000 students remained in Engineering, about 1000 students switched to Other STM (Science, Technology, and Mathematics) fields, and about 100 students switched to Computer Science. The graph feature to notice is that the dots are organized to facilitate comparisons of position on a common scale—the graphical perception task we perform with the greatest accuracy.

A quick review of journals of engineering education reveals that column and bar charts are the dominant form of data display—the community seems to be unaware of the benefits of multiway plots in both thinking about data and in presenting data.

In the next section, we show how the ubiquitous bar or column chart can be transformed to a multiway plot, achieving two ends: first, playing to the strengths of the viewer by giving the viewer a perceptual task that can be accomplished with accuracy—thereby helping us (researchers) consider new ways to present our data; and second, creating a graph that is designed to reveal the logic of one’s analysis—thereby helping us consider new ways to think about our data. As Tufte says, “if displays of data are to be truthful and revealing, then the logic of the display design must reflect the logic of analysis”.\(^5\) Multiway plots assist us in extracting the story the data tell.
**Method and results: transforming column charts to multiway plots**

_Eighth-semester persistence data._ To interpret multiway plots in contexts that speak to engineering education audiences, we use categorical data from MIDFIELD (the Multiple-Institution Database for Investigating Engineering Longitudinal Development) on eighth semester persistence disaggregated by race and gender. MIDFIELD data are described in detail in other work.⁶

Table 1 shows the data we use in all our subsequent displays. There is one quantitative variable, count, and two categorical variables, eighth-semester destination (8 levels) and race-gender type (10 levels). What distinguishes multiway data is the cross-classification of the categorical variables¹; there is a count for each combination of levels of destination and race-gender type. The goal is to study how the count relates to the categories. The context for these data is briefly outlined in the footnote to the table.

<table>
<thead>
<tr>
<th>Eighth-semester destinations</th>
<th>Asian Female</th>
<th>Asian Male</th>
<th>Black Female</th>
<th>Black Male</th>
<th>Hispanic Female</th>
<th>Hispanic Male</th>
<th>Nat-Am Female</th>
<th>Nat-Am Male</th>
<th>White Female</th>
<th>White Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts &amp; Humanities</td>
<td>25</td>
<td>38</td>
<td>66</td>
<td>85</td>
<td>15</td>
<td>26</td>
<td>1</td>
<td>3</td>
<td>329</td>
<td>791</td>
</tr>
<tr>
<td>Business</td>
<td>52</td>
<td>166</td>
<td>137</td>
<td>208</td>
<td>21</td>
<td>71</td>
<td>1</td>
<td>18</td>
<td>689</td>
<td>2688</td>
</tr>
<tr>
<td>Computer Science</td>
<td>19</td>
<td>110</td>
<td>39</td>
<td>82</td>
<td>2</td>
<td>29</td>
<td>1</td>
<td>2</td>
<td>87</td>
<td>836</td>
</tr>
<tr>
<td>Engineering</td>
<td>699</td>
<td>2716</td>
<td>2363</td>
<td>3788</td>
<td>250</td>
<td>1011</td>
<td>37</td>
<td>148</td>
<td>6182</td>
<td>27722</td>
</tr>
<tr>
<td>Other Non-STM</td>
<td>10</td>
<td>15</td>
<td>52</td>
<td>109</td>
<td>10</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>195</td>
<td>523</td>
</tr>
<tr>
<td>Other STM</td>
<td>56</td>
<td>124</td>
<td>143</td>
<td>125</td>
<td>22</td>
<td>39</td>
<td>8</td>
<td>10</td>
<td>774</td>
<td>1774</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>31</td>
<td>53</td>
<td>134</td>
<td>192</td>
<td>13</td>
<td>53</td>
<td>3</td>
<td>6</td>
<td>413</td>
<td>1227</td>
</tr>
<tr>
<td>TOLEDO</td>
<td>162</td>
<td>803</td>
<td>1113</td>
<td>2112</td>
<td>88</td>
<td>406</td>
<td>23</td>
<td>85</td>
<td>2213</td>
<td>10995</td>
</tr>
</tbody>
</table>

Note: The population (75686 students) comprises first-time-in-college (native) undergraduate students, matriculating in an engineering major, in the cohorts from 1988/89 through 1998/99, disaggregated by race and gender. The table shows their eighth–semester destination majors and the number of students (the count) going to each. The destination “TOLEDO” is an acronym for “Trajectory Of Leaving Education, Destination Obscure”, accounting for students who have dropped out, stopped out, or otherwise left the dataset.

_Graphing the data._ The data from Table 1 are graphed in a clustered-column graph in Figure 2. The first thing we notice is that the largest count (27722 White Males in Engineering) is four orders of magnitude greater than the smallest count (for example, the single-digit number of Native American Females in several destination majors).
With differences of this magnitude, we change the vertical axis to a logarithmic scale (base 10 in this case), yielding Figure 3. The distance between gridlines in Fig. 3 represents a factor of 10 difference in the counts of students. The smaller-magnitude counts are much easier to see than in Figure 2, though counts of zero or one are still undetectable because the lower limit of the vertical axis is one. It should also be noted that the need for using ten different colors for identification has caused colors to be used that would be impossible to distinguish if printed or photocopied in black and white. This is of particular concern given that most archival journals (including the *Journal of Engineering Education*) do not permit the use of color.

To make the destination labels easier to read, we exchange the horizontal and vertical axes, yielding Figure 4. The alphabetical ordering of categories from the bottom up in both the destination majors and the race-gender categories is a software default—and not a particularly useful default.
To help us discover patterns in the data, we order the categories in something other than alphabetical order. To Table 1 we add a column of row totals and a row of column totals and sort the rows and columns by decreasing membership count, as shown in Table 2. Engineering is now in the topmost row, indicating that it is the predominant eighth-semester destination of all students matriculating in Engineering. Similarly, White Males are now in the leftmost column, indicating that they are the majority race-gender category of all Engineering matriculates.

### Table 2: Categories ordered by decreasing magnitude of group membership.

<table>
<thead>
<tr>
<th>Eighth-semester destinations</th>
<th>White Male</th>
<th>White Female</th>
<th>Black Male</th>
<th>Black Female</th>
<th>Asian Male</th>
<th>Asian Female</th>
<th>Hispanic Male</th>
<th>Hispanic Female</th>
<th>Nat-Am Male</th>
<th>Nat-Am Female</th>
<th>Row totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineering</td>
<td>27722</td>
<td>6182</td>
<td>3788</td>
<td>2363</td>
<td>2716</td>
<td>1011</td>
<td>699</td>
<td>250</td>
<td>148</td>
<td>37</td>
<td>44916</td>
</tr>
<tr>
<td>TOLEDO</td>
<td>10995</td>
<td>2213</td>
<td>2112</td>
<td>1113</td>
<td>803</td>
<td>406</td>
<td>162</td>
<td>88</td>
<td>85</td>
<td>23</td>
<td>18000</td>
</tr>
<tr>
<td>Business</td>
<td>2688</td>
<td>689</td>
<td>208</td>
<td>137</td>
<td>166</td>
<td>71</td>
<td>52</td>
<td>21</td>
<td>18</td>
<td>1</td>
<td>4051</td>
</tr>
<tr>
<td>Other STM</td>
<td>1774</td>
<td>774</td>
<td>125</td>
<td>143</td>
<td>124</td>
<td>39</td>
<td>56</td>
<td>22</td>
<td>10</td>
<td>8</td>
<td>3075</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>1227</td>
<td>413</td>
<td>192</td>
<td>134</td>
<td>53</td>
<td>53</td>
<td>31</td>
<td>13</td>
<td>6</td>
<td>3</td>
<td>2125</td>
</tr>
<tr>
<td>Arts &amp; Hum</td>
<td>791</td>
<td>329</td>
<td>85</td>
<td>66</td>
<td>38</td>
<td>26</td>
<td>25</td>
<td>15</td>
<td>3</td>
<td>1</td>
<td>1379</td>
</tr>
<tr>
<td>Computer Science</td>
<td>836</td>
<td>87</td>
<td>82</td>
<td>39</td>
<td>110</td>
<td>29</td>
<td>19</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1207</td>
</tr>
<tr>
<td>Other Non-STM</td>
<td>523</td>
<td>195</td>
<td>109</td>
<td>52</td>
<td>15</td>
<td>18</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>933</td>
</tr>
<tr>
<td>Column totals</td>
<td>46556</td>
<td>10882</td>
<td>6701</td>
<td>4047</td>
<td>4025</td>
<td>1653</td>
<td>1054</td>
<td>421</td>
<td>272</td>
<td>75</td>
<td>75686</td>
</tr>
</tbody>
</table>

Using this new ordering, we redo Figure 4 to obtain Figure 5. We begin to see some patterns in the data. For example, White Males and White Females have the highest count in each destination, but not always in the same proportion. However, our ability to extract meaning is confounded by the cognitive tasks this chart asks of us:

- Comparing race-gender counts by destination.
- Comparing destination counts by race-gender group.
The difficulty of accurately making such comparisons is the primary problem plaguing all clustered-bar and clustered column charts. One solution is to separate the two cognitive tasks by creating two graphs—one for each type of comparison.

### Figure 5: Sorting in order of decreasing category subtotals.

Comparing race-gender counts by destination. In the first graph type, the cognitive task is to compare race-gender counts for each destination, one destination at a time. For example, if we extract from Figure 5 the bars representing the Engineering destination, we can stack them for comparison as shown in Figure 6.

### Figure 6: Bar chart of race-gender counts for one destination.

Next we consider Cleveland and McGill’s results on graphical perception tasks. Figure 6 has a common scale (first on Cleveland and McGill’s list) that we keep. The data we’re trying to show is the count of students—the position of the endpoints of the bars. But a bar has both length...
In Figure 7, the top row is the White Male group because it is the race-gender group with the largest total membership (not because it has the largest membership remaining in Engineering, although that is also true in this case). Similarly, Native American Females are in the bottom row because they are the group with the smallest total membership.

The dot plot has the advantage, compared to the bar chart, of satisfying Tufte’s advice to make the data prominent (the count stands out) and to increase, as much as practicable, the ratio of data to ink (we’ve eliminated the ink used to outline and fill the bar). Most importantly, we are better able to distinguish small differences in the count. Notice how easy it is to discern that Asian Males stand out slightly in this plot, persisting in engineering in higher numbers than the size of their group would suggest. Cleveland calls such points “visual asymmetries”. Knowing to look for this, one can return to Figure 5 and see the same thing. It’s there, it just cannot be seen easily because the chart is neither organized nor thoughtfully designed to help us make such comparisons.

We’re ready now to create a dot plot for each destination major. The array of destination dot plots is Figure 8: a multiway plot. This is a complete, though not unique, visualization of the ordered data in Table 2. To facilitate comparisons, all the horizontal scales are identical and the ordering of the race-gender groups is identical.
**Ordering the categories.** The ordering of the categories is “crucial to the perception of effects.” The ordering we’ve used is the increasing total number of students in a category. In Figure 8, the race-gender rows (or *levels*) are arranged bottom-to-top in order of increasing total number of students of that type group (Table 2 column order). The destination plots (or *panels*) are arranged in order of increasing total number of students going to each destination (Table 2 row order), starting at the lower left corner and increasing from left to right, up a row, left to right, and so forth, with the panel having the highest total count (Engineering) appearing in the top-rightmost position. The arrows in Figure 9 illustrate this ordering.
Other orderings are possible and should be considered in data exploration. The selection of an ordering scheme should be informed by one’s research question. However, not all orderings produce meaningful insights. For example, Figure 10 is a multiway plot with the categories in alphabetical order. No patterns emerge, no symmetries appear, and no visual asymmetries stand out either. With these data, as with many published examples of categorical data, alphabetical ordering enhances neither our reasoning about, nor our display of, the data.
Comparing destination counts by race-gender group. In the second graph type, the cognitive task is to compare destination counts for each race-gender group, one race-gender group at a time. For example, if we extract from Figure 5 all the bars representing White Females, we can stack them for comparison as shown in Figure 11. Here destinations are assigned to the levels and race-gender groups are assigned to the panels. This plot might be considered the “transpose” of Figure 6.
This figure too is improved by placing a dot at the endpoint and eliminating the bars, yielding Figure 12. Again we observe that the dot plot brings out the visual asymmetries. For example, Computer Science stands out as an eighth-semester destination that attracts White Females in smaller numbers than the size of their group would suggest. The plot also raises a new question: do White Females change majors to Business in smaller than expected numbers, or do they change majors to Other STM fields in greater than expected numbers? From this plot alone we cannot answer this question (nor are such research questions our focus here), but highlighting that a dot plot can bring such questions to our attention is our focus. The dot plot assists us in thinking about our data in ways that standard bar graphs, column graphs, and data tables do not.

In Figure 12, Engineering is in the top row because among destinations it is the group with the largest total membership (again, not because it has the largest number of White Females, although that is the case in this dataset). The category Other Non-STM is the bottom row because it is the destination with the smallest total membership.

Repeating for all the race-gender groups and assembling the ordered dot charts yields the multiway plot in Figure 13. The figure shows us the same data as shown in Figure 8—organized differently to facilitate different comparisons. As Cleveland says, “We can more effectively compare values within a panel than values between panels.” Thus, in Figure 8 the six race-gender counts for one destination are readily compared because they are graphed on a common scale. But in Figure 8 we cannot as effectively compare the eighth-semester destination counts...
for a given race-gender group. This is done far better in Figure 13, where race-gender is assigned to the panels and destination is assigned to the levels.

In Figure 13, White Male is the upper right panel because among race-gender groups it is the group with the largest membership. Similarly, Native American Females are in the lower left panel because they are the group with the smallest total membership.

![Multiway plot with race-gender groups assigned to panels.](image)

Figure 13: Multiway plot with race-gender groups assigned to panels.

In the next figure, we illustrate how the research question might inform the ordering of the panels. Suppose the research question involves an investigation of the experiences of female students in the different race categories. To make it easier to compare the destinations of female students of different races, we can modify Figure 13—arrange the panels so that females are all...
in the left-hand column, males in the right-hand column, and align males and females of the same racial group in the same row. This arrangement is shown in Figure 14. We can now more easily compare the eighth-semester destinations of males and females of the same racial group, or compare the destinations of all females or all males. For example, comparing the eighth-semester destinations of Hispanic males and females, we see similarities except for Other STM, Social Sciences, and Computer Science majors. Note that the display only tells us that the differences exist, not why they exist.

Figure 14: Reordering the panels to facilitate comparisons by sex.
Software. The column and bar graphs in this paper were made with Microsoft Excel 2007. However, dot plots and multiway plots are not in Excel’s native vocabulary. The options for creating dot plots and multiway plots include:

1. **Excel Add-Ins.** Robbins\(^2\) points readers to an Excel macro written by Kenneth Klein for making individual dot plots and Vidmar\(^7\) shows how to use Excel for both dot plots and multiway plots. We have not tested these add-ins and so offer no opinion on their utility.

2. **SAS:** This popular commercial package for statistical analysis has a “multivariate” toolkit that includes multiway plots.

3. **R:** an open source software version of the commercial package called S. Both the R and S packages are used and recommended by statisticians. S is the package used to produce most of the graphs in both Cleveland’s and Robbin’s texts. These are programming environments with extensive libraries of functions for dot plots, multiway plots, scatter plot matrices and other tools for exploratory data analysis and visualization.\(^8\)

4. **Programming languages** such as MATLAB. There are free downloadable MATLAB toolboxes for multiway plots.\(^9\) However, these may not offer sufficient fine control of all aspects of the final display. All the dot plots and multiway plots in this paper were made using MATLAB programs written by Layton. This code allows precise placement of every feature of the graph.

5. **Consultants:** Researchers lacking the time to invest in learning how to use these resources have the option of obtaining the help of consultants specializing in data displays.

Our experience with using multiway plots to investigate our categorical data is that, like any worthwhile effort, it takes time, study, and intellectual effort. The cost is in developing the necessary software skills to produce graph designs that are not part of conventional software suites. The benefit is in obtaining visualizations of the data that help us to think about the data in ways that tables of data like Table 1 and easy-to-obtain clustered-bar charts like Figures 1 through 5 do not facilitate. Having new ways to think about the data also help us communicate those findings effectively to our audience.

**Discussion and conclusions**

The graphical transformation in the previous section demonstrates the perceptual advantages of the multiway plot compared to clustered-bar or column charts or to the tables of categorical data on which the graphs are based. For example, the data we’ve been discussing is completely contained in Table 2. Even with the data organized by magnitude of the row and column totals however, the table does not help us visualize the comparisons between numbers. (This is not to say that the table could not be redesigned—Tufte has several recommendations for table designs that we have not explored here.\(^3\))

Even the best of the clustered-bar charts, Figure 5, does not facilitate easy comparison of counts of students going to specific destinations. Bars, as data encoding artifacts, add two unnecessary visual cues—length and area—that we do not perceive as accurately as we do the position of the endpoint. Dot plots and multiway plots take advantage of our ability to accurately compare positions along a common scale. And as Cleveland notes, the dots in the multiway plots make visual asymmetries stand out, inviting us to ask “why,” perhaps leading us to the next research
question and to the next re-design of the data display. Thus the display method facilitates both the way we think about the data and the way we present the data.

Multiway plots have the additional advantage that by reversing panels and levels—e.g., compare Figure 8 and Figure 13—different comparisons come to the forefront, telling the different stories that the data may have to tell. For example, consideration of Figure 13 led us to rearrange the panels to create Figure 14, giving us a clearer picture of comparisons by sex.

Multiway plots also give us a complete and evenhanded picture of the data. First, the data themselves are shown. Note that even small counts such as the one Native American female in Computer Science shows clearly in both Figure 8 and Figure 13—a useful feature of multiway plots for representations of data with small sample sizes. Second, no data point is given greater “weight” by virtue of its greater length or area (as in bar and column charts). All data are given equal graphical importance because they are represented by dots of equal size and hue.

In addition, the highlighted comparison of specific elements of the “before” graph (the bar chart) and “after” graph (the multiway plot) emphasize Tufte’s design principles to erase non-data ink and to make the data prominent. For example, Figure 5 hints at the possibility of some patterns in the data, but Figure 8 makes those patterns clear and Figure 13 gives a “transposed” view of the same data. Likewise, the bar chart of Figure 11 seems adequate until a comparison with the dot plot of Figure 12 shows us how perceptually intrusive are the length and area of the bars compared to relative sparseness of the dots.

Through this work, we hope to raise the awareness of the engineering education community of the benefits of multiway plots for visualizing, exploring, and presenting categorical data. In doing so, we also hope to contribute to the continued enhancement of research quality in our discipline. We have shown the advantages of the multiway plot compared to the ubiquitous clustered-bar and column charts we find in our literature for displaying categorical data. Thinking about how one will display one’s data is a critical aspect of the research process. Perhaps the most important message is to thoughtfully select one’s own data display rather than using default displays and possibly miss an opportunity to more fully understand a dataset and the stories it has to tell.

Bibliography