informs
http://pubsonline.informs.org/journal/inte/

INTERFACES
Vol. 47, No. 6, November–December 2017, pp. 473–488
ISSN 0092-2102 (print), ISSN 1526-551X (online)

A Practitioner's Guide to Best Practices in Data Visualization

Jeffrey D. Camm,[a] **Michael J. Fry,**[b] **Jeffrey Shaffer**[b]

[a] School of Business, Wake Forest University, Winston-Salem, North Carolina 27109; [b] Carl H. Lindner College of Business, University of Cincinnati, Cincinnati, Ohio 45221

Contact: cammjd@wfu.edu (JDC); mike.fry@uc.edu (MJF); shaffejy@ucmail.uc.edu (JS)

Received: June 27, 2016
Revised: November 8, 2016; April 7, 2017
Accepted: April 14, 2017

https://doi.org/10.1287/inte.2017.0916

Abstract. Data visualization is the process of visualizing data through tables, charts, graphs, maps, and other visual aids. Data visualization is often thought of as a descriptive tool, but it is also important for use in predictive and prescriptive analytics. It serves as a tool for descriptive data exploration, but also for communicating insights from data and from analytical models. In this tutorial, we provide a concise discussion of best practices for data visualization for effective communication.

History: This paper was refereed.

Keywords: professional: OR-MS • professional: implementation

Much has been written about the importance of communication in the application of operations research (OR) and in ensuring that the results from OR studies are accepted and implemented. Woolsey (1979, 1981), Levasseur (1991, 2013), and Keller and Kros (2000) provide examples. Data visualization is an important tool for data exploration; however, it is also critical for communicating results from analysis and recommendations from analytical studies.

From the point of view of good, available software, data visualization has never been easier. A variety of excellent data-visualization software packages are now available. A few of the leading specialized software packages include Tableau, SAS Visual Analytics, and Spotfire. The open source language R also provides excellent data-visualization capabilities, and Microsoft has recently updated the data-visualization capabilities in its ubiquitous Excel package to allow the creation of a wider range of charts. Excel also has a recommended-chart option, which provides the user with a list of suggested visualizations based on the selected data. Comparisons of data-visualization tools can be found in recent ratings by *PC Magazine* (Baker 2017) and *Predictive Analytics Today* (2017), and in Gartner's "Magic Quadrant for Business Intelligence and Analytics Platforms" (Sallam et al. 2017). However, software availability does not relieve the user of the responsibility of making intelligent choices—choices that impact the effectiveness of the visualization in communicating to the audience.

In this tutorial, we discuss best practices for effective data visualization, give examples of good and bad visualizations, and describe a variety of types of charts and graphs and the comparisons for which they are most effective.

The Goal and Three General Principles

The goal of a table or chart should be to display data in a manner that conveys the desired message to the intended audience. The input data to the visualization can be raw data, descriptive statistics from the raw data, or the output of analytical models. Hence, the principles we discuss in this paper are applicable across the full spectrum of analytics: descriptive, predictive, and prescriptive. For example, in descriptive analytics, scatter plots are useful for exploring and conveying possible relationships between pairs of variables. Overlaid line charts are often an excellent choice for the output of predictive models when the goal is to show the model results and prediction intervals to convey the uncertainty associated with those results. A network geographic map, with arcs showing distribution-center-to-customer-zone assignments, is an excellent way to convey the efficient structure of the solution from a supply-network optimization model.

Regardless of the type of input data (data or model output), the creation of the visualization must be guided by the intended audience, the interests of that audience, and the message that is to be conveyed. This requires an understanding of the makeup of the audience and an anticipation of questions audience members will pose. In short, empathy for your audience should be the main driver in designing your visualization.

Before embarking on a discussion of the three guiding principles for data visualization, it is instructive to know a bit about how we process information. Every second our brains are busy processing the information that our five senses are feeding us. The brain filters stimuli from our environment to process what is important in two main ways: unconscious and conscious, as Kahneman (2011) refers to as System 1 and System 2, respectively. System 1 represents the uncontrolled functions, such as reading people's facial expressions or the immediate reaction we have to an event. System 2 represents the controlled functions that require conscious effort, such as solving a math problem.

In data visualization, we leverage the pre-attentive attributes, which are System 1. We encode data in a certain way; for example, if we highlight something with color, the reader sees this and interprets it in 250 milliseconds without even thinking about it. Hence, by utilizing methods that are attuned to System 1, data-visualization tools, when used properly, can help a reader gain a quick and correct understanding of the data being presented. The three guiding principles, which we discuss next, are all about making sure we capitalize on how we process information using System 1.

The first general principle in data visualization is that *design and layout matter*. The design and layout of your visualization should facilitate ease of understanding your message. The design and layout, including the type of chart or table used, should draw attention to the parts of the visualization that are important in conveying your message to your intended audience. For example, axes should be labeled clearly and legends should be close enough to the display to facilitate easy comparisons.

The second general principle is to *avoid clutter*. In short, simpler is better. To make this point, Tufte (2001) introduced what he calls the data-ink ratio—the ratio of ink required to convey the intended meaning (data-ink) to the total amount of ink used in the table or chart. According to Tufte, the data-ink ratio should be as close to one as possible; that is, we should avoid ink that does not add information.

The third general principle, consistent with the notion that simpler is better, is *to use color purposely and effectively*. Although the use of color might be attractive to make the data visualization seem "prettier," it can be a distraction to the audience and hinder your message. As an example, color can be used effectively to draw attention to part of a visualization or to distinguish categories in the data; however, using too much color can distract the audience. Color should be used intentionally, that is, only if it assists in conveying your message.

In the following sections, we provide examples and more detail on these three general principles of good data visualization. For a more detailed discussion on how the brain processes images for data visualization, we refer the reader to Few (2009), particularly to Chapter 3.

An Example: Barring Pie Charts

Let us consider a simple descriptive analytics example to illustrate the use of the three general principles, as well as the iterative nature of the process for creating a good visualization. Like prototyping in the modeling process, visualization often benefits from incremental improvements over multiple iterations. Consider the following example.

We have data that have yielded the percentage of students declaring each of the 12 categories of majors (plus undesignated) in a college of business. Suppose these data are to be used as part of the recruiting process for the college; that is, the audience includes potential students, and in many cases, one or more parents of these potential students. Undoubtedly, the students and parents will have many questions about the majors; however, we know that their questions are likely to include the following:

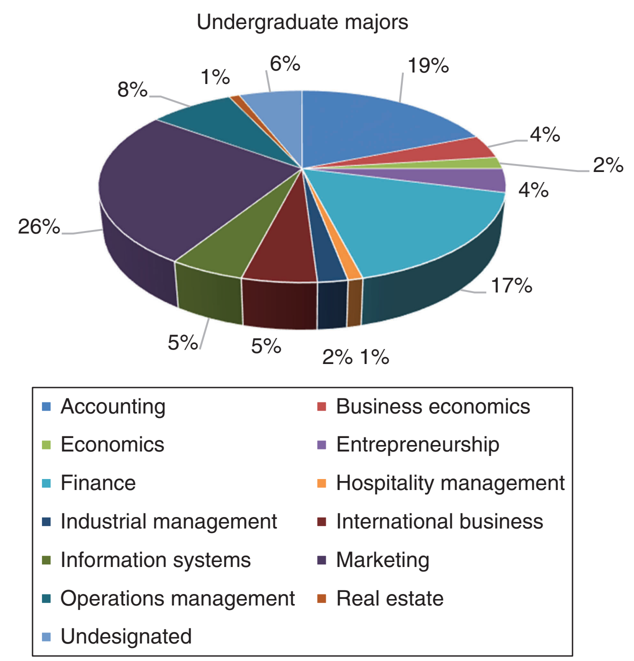
- What is the most popular major?
- What is the least popular major?
- What are the top three majors?
- How does the major I am considering compare to the other majors in terms of its popularity?
- What percentage of students are undesignated?

The data that we have available are the percentage of the total student population in each major, which illustrates a part-to-whole relationship—the percentage breakdown of the total for the variable under consideration. Pie charts are often touted as being useful for showing this part-to-whole relationship (Figure 1).

What are the potential problems with this choice of chart? One drawback is that it is difficult to see the legend and the colors associated with each major. The layout and design do not facilitate easy comparison because the viewer must move back and forth between the legend and the pie. With 13 categories, many of the colors are very similar, which also makes comparison difficult. It might be better to label each slice of the pie with its major; however, doing so would make the chart even more cluttered. Having the percentage next to each slice is helpful.

An interesting question is, "What does the third dimension add to this chart to help us better understand the data?" The answer is nothing. In this example, adding the third dimension makes comparing categories more difficult.

Figure 1. This Three-Dimensional Pie Chart Shows the Percentage of Undergraduates by Major for a College of Business



Note. The third dimension reduces the data-ink ratio and makes the chart more difficult to understand.

Consider Figure 2 in which we have eliminated the unnecessary third dimension.

Can we improve upon this chart? Studies have shown that humans are very good at measuring the length or height of an object on a common baseline. For example, when two people stand next to each other, we can easily tell which person is taller and by how much. Humans are also good at measuring the position of objects; for example, how far one person is standing from another. However, they are much less adept at accurately measuring angles, arcs, and area, and comparing them against each other. We can leverage this strength at measuring length in the design of a chart by carefully choosing the type of chart to use. Bar charts are a good choice for comparing categories. Given the questions we believe our audience will pose, that is, comparisons of the percentages of students in each major, a bar chart would be a better design for our purposes.

Figure 2. Compared to the Three-Dimensional Pie Chart in Figure 1, the Data-Ink Ratio in This Pie Chart Is Higher and the Chart Is Easier to Understand

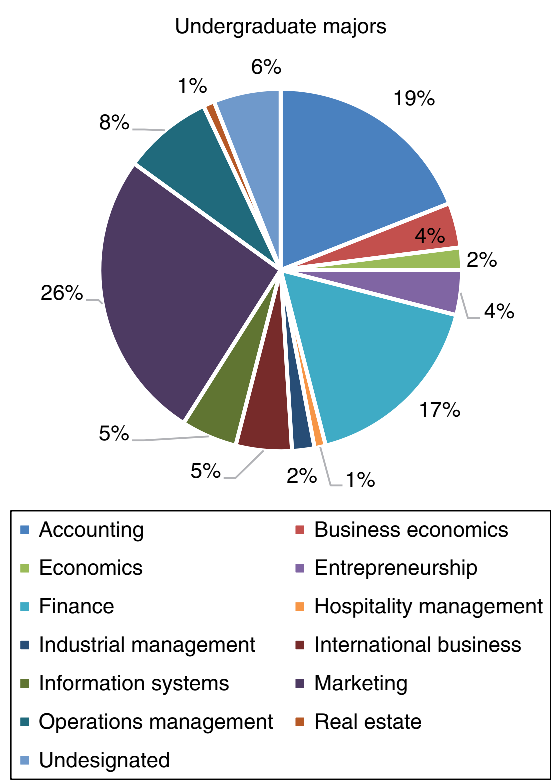


Figure 3 shows a vertical bar chart of the data on business college majors. The differences in college majors are much easier to visualize in this figure than in either Figures 1 or 2. For example, compare Accounting and Finance in Figures 3 and 2. Without the data labels, looking at Figure 2 and determining which major has a higher percentage of undergraduates is much harder. In Figure 3, even without data labels, we can relatively easily see that the bar for Finance is shorter than the bar for Accounting; thus, the percentage of Finance majors is smaller than the percentage of Accounting majors.

Can we improve upon Figure 3? If we think about how this chart will be used, perhaps sorting the data from high to low would make the chart easier to read and more useful for the reader. Figure 4 shows the same vertical bar chart sorted. Sorting allows the reader to more easily find the highest and lowest percentage majors and to compare majors that are close in percentage.

Finally, our adeptness at comparing lengths is not impaired by the use of a horizontal display versus a vertical display. Figure 5 displays a horizontal bar chart, which is a better design for the same data; the categorical text labels are easier to read because they are oriented horizontally. In addition, because we have labels next to each bar, we removed the horizontal axis for simplicity.

Figure 3. In This Bar Chart, Different Colors Are No Longer Necessary to Distinguish Between Majors; Thus, We Can Eliminate the Legend

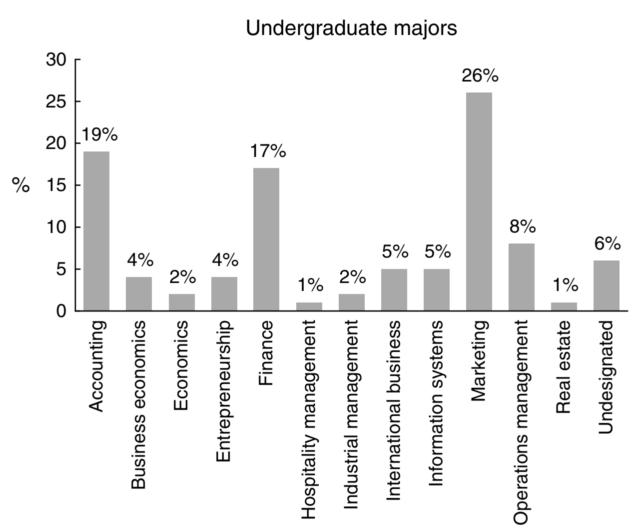
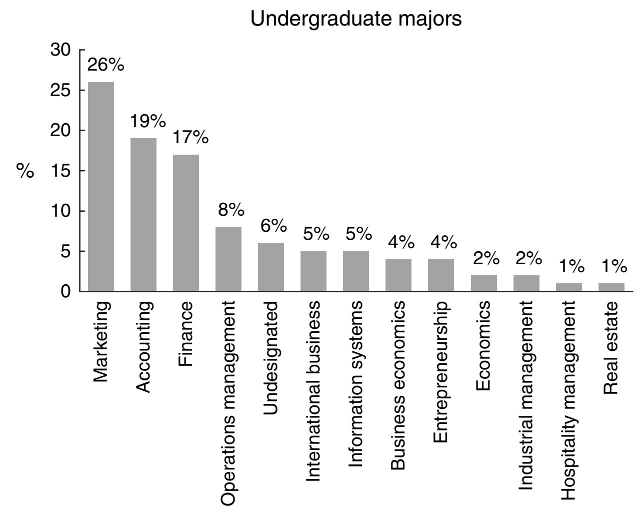


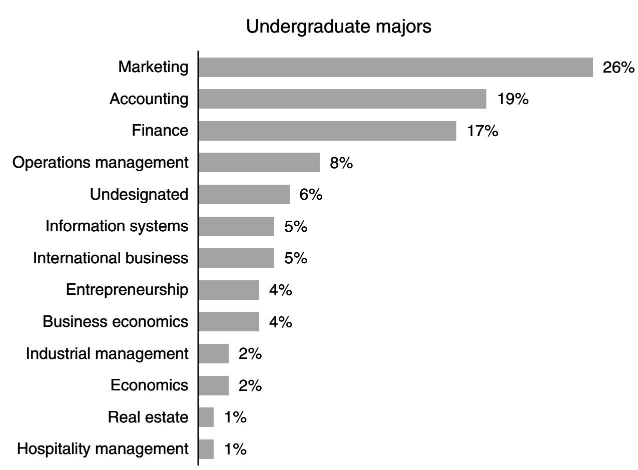
Figure 4. Sorting a Bar Chart Allows a Reader to Easily and Quickly Compare Differences Among Variables



What are the three largest majors? Compare the ease of answering this question by looking at Figure 5 versus Figure 2. We can also quickly determine the bottom three majors and the second highest; or we can compare the third highest with the fourth highest. With this simple example in Figures 1–5, we have illustrated the following three general principles:

- Design and layout matter: Use them effectively. We evolved from a pie chart to a bar chart to a sorted

Figure 5. The Categorical Text Labels Are Horizontal; Thus, the Chart Is Easier to Read Than the Vertical Bar Chart in Figure 4



bar chart, and we then rotated the bar chart for ease of reading.

- Avoid clutter: Too many labels, values, lines, and dimensions hinder the reader in interpreting the data. We eliminated the distracting third dimension.
- Use color purposely and effectively: Use it only when it is needed to distinguish categories or draw the attention of the reader. By choosing a bar chart, we were able to eliminate the use of different colors and the associated legend.

From this example, we can also learn the following:

- Avoid using a pie chart—use a bar chart instead.
- Sorting the data often makes relevant comparisons easier; in our example, it can help our audience to more easily answer the questions we anticipate.
- Avoid rotated text. If the data include long category names, consider rotating the chart.

In a more general way, this example illustrates what often happens in practice. We can rarely achieve the best visual display on the first attempt. Rather, iterative prototyping, based on the intended message, leads to a narrowing down of design options followed by successive improvements in labeling, the use of color, titles, and other details that improve the communication of our message. Berinato (2016) provides an expanded discussion of prototyping in the context of chart construction.

In the next three sections, we provide more detailed discussions of the three general principles.

Design and Layout Matter: Types of Comparisons and Chart Types

A number of primary comparisons can made when visualizing data. The type of comparison being made is critical when determining the type of data visualization or chart type.

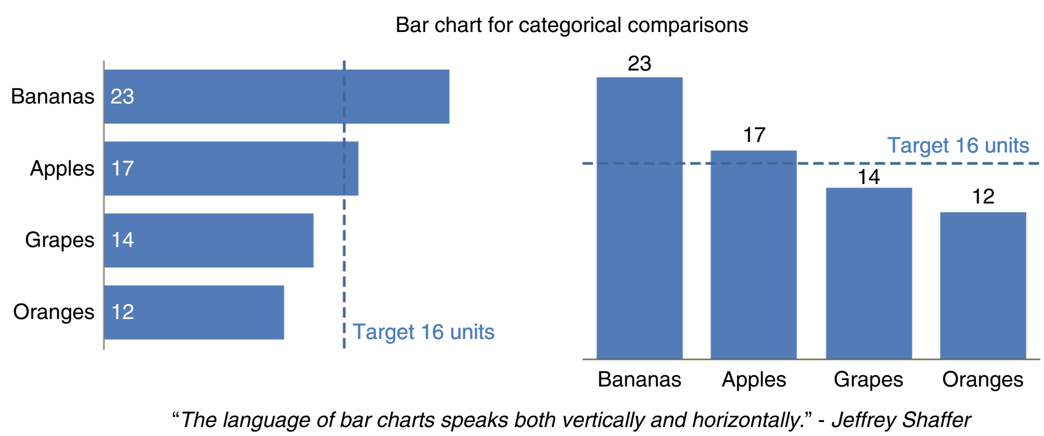
A categorical comparison is one of the more common comparisons, and several good choices are available. As we previously discuss, because humans are adept at measuring height and width, we can leverage this strength in our choice of chart type. As we saw in the first example, a bar chart, either vertically or horizontally, is a great choice for making quick and precise categorical comparisons. Figure 6 shows a simple example with two good approaches for a simple comparison (in this case, units of fruit sold). If our audience has a sales target of 16 units for each of these fruits, adding a target line helps the audience to easily compare how each fruit is doing relative to the target. In addition to a target line, other examples of this include using "average," "estimate," "threshold," and "previous year." Which of these to use should be driven by the problem context and questions you anticipate the audience will ask.

In designing a bar chart, consider the following:

- Order the data to provide additional context.
- For a small number of bars, consider data labels.

Position the labels such that the reader can easily see them. Notice the "data table" that is created when the

Figure 6. (Color online) Bar Charts for Categorical Comparisons Can Be Augmented with a Target Line



Notes. Target lines are often used to give the reader a reference point and context. Target lines can represent values, such as sales goals, previous year data, averages, and thresholds.

labels are aligned near the vertical axis of the chart. This makes it very easy for the eye to follow.

• If the chart must include many categories, then balance the use of axis labels with data labels. Both are unnecessary. In the example above, why use axis labels to give the reader approximate values, when the data labels provide the actual values?

• Avoid rotated text. If some category names are long, consider rotating the chart.

• Mute or remove gridlines and remove chart borders.

• Use a light font color on dark colors and dark font colors on light colors (e.g., white font color on the dark-colored bars in Figure 6).

• Always start the axis at zero on a bar chart or any other chart that encodes data using length or height (e.g., bar chart, histogram, lollipop chart, area chart). Because these types of charts use height or length to encode the data for comparison, the axis should begin at zero; otherwise, the visual comparison is incorrect and can be misleading.

In addition, the part-to-whole comparison is often necessary. Unfortunately, this is often shown as a pie chart, which for the reasons previously discussed, most data-visualization experts consider to be bad practice. A bar chart is recommended for these types of comparisons. To emphasize comparisons among categories and part-to-whole comparisons, the bar chart can be supplemented by a stacked bar chart (Figure 7).

Another very common comparison in data visualization is showing a trend over time. Time should generally be on the horizontal axis. Line charts are great for visualizing a trend, but a bar chart can also be used. Figure 8 shows two good approaches for times-series data. Note that the line chart does not show all data values because doing so would make the chart appear too cluttered. Here we simply show the highest value and the most recent value. For example, if our audience for Figure 8 is concerned about planning candy purchases, then having actual numbers, as we show on the bar chart in the figure, with the actual number of Halloween visitors, might be more useful. We could also enhance the figure by adding an average line, similar to our target line in Figure 6.

In designing a time-series chart, consider the following:

• Time should be on the horizontal axis; time should progress from earlier to later and read from left to right.

• If using bars to represent time, do not break the axis. Bars must begin at zero. It is acceptable to break the axis on a line chart; however, it has a magnifying

Figure 7. (Color online) A Stacked Bar Chart Can Enhance a Bar Chart and Is Generally Preferable to a Pie Chart for Making Part-to-Whole Comparisons

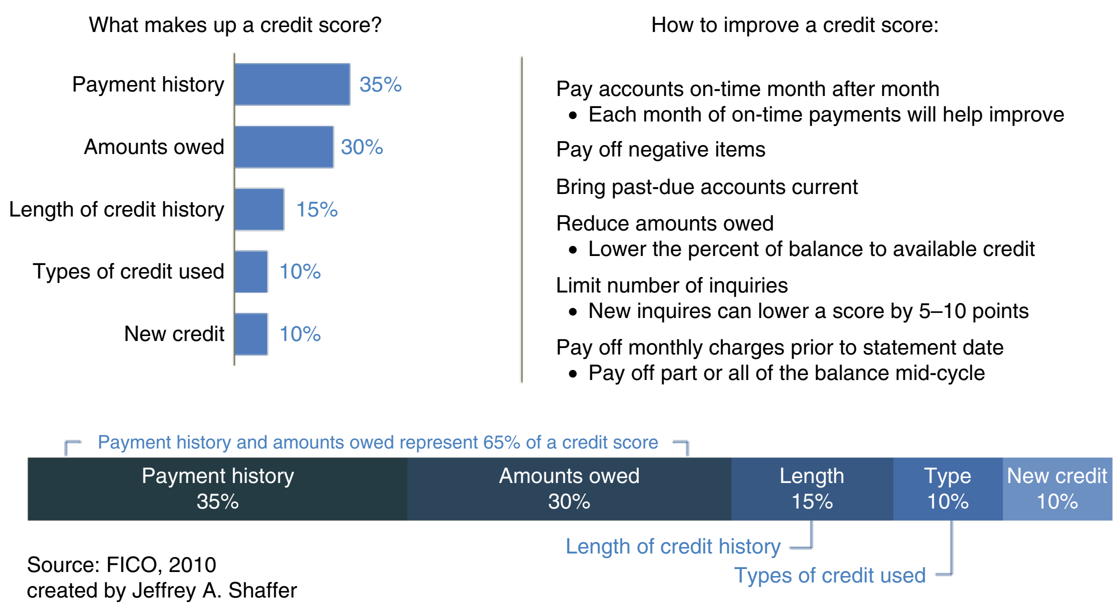
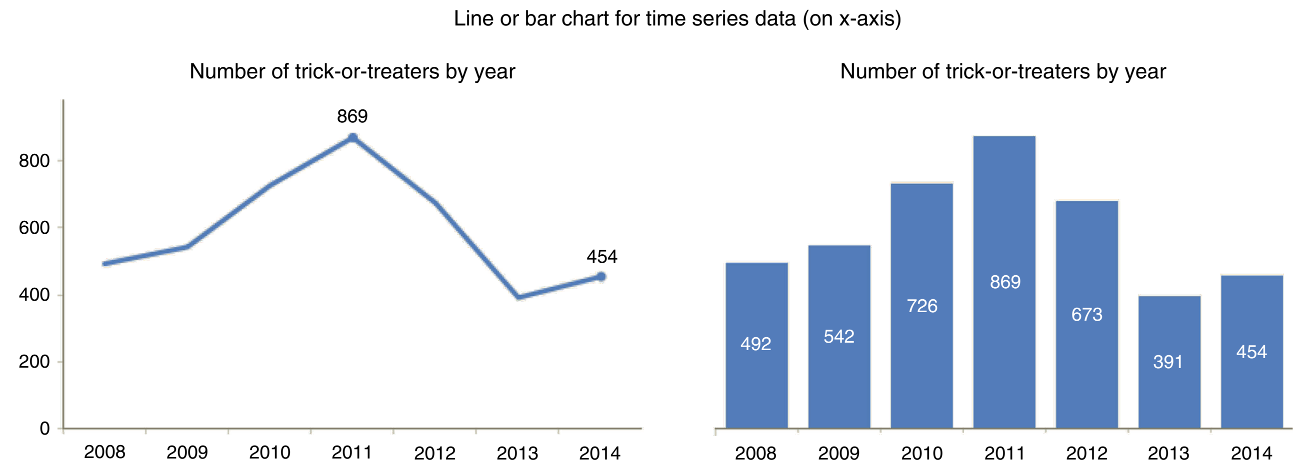


Figure 8. (Color online) Both Line Charts and Bar Charts Can Be Used Effectively for Time-Series Data



Note. The line chart emphasizes minimum and most recent values to the reader; the bar chart can more effectively convey information on values for each year without appearing too cluttered.

effect on the visualization, making the visual slope of the line appear much more dramatic as the magnification increases.

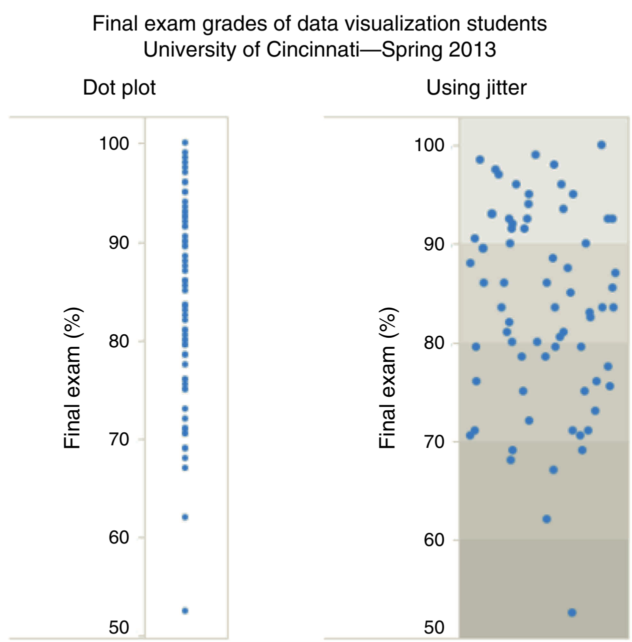
• Plot consistent time periods using a line chart. If time periods are missing or unequal, then note this and deal with it in the visualization; for example, use a dotted line where years are missing.

The use of dots can be effective in various visualization methods. This includes making comparisons, showing time-series data when an interval is missing, plotting a distribution, or showing a correlation between variables. If possible, avoid using shapes other than dots or circles. Circles and dots have a single center point that is clear to readers. Triangles, squares, asterisks, stars, or other shapes hinder the reader in making a quick and accurate determination of the center of the point.

Figure 9 shows an example of a dot plot, also known as a strip plot, to visualize the grades in a data-visualization class. A standard dot plot is on the left; the same plot using jitter and background shading for the performance bands (50 to 100 percent) is on the right. Jitter is a random value, or for our purposes, pseudo-random value, which is assigned to the dots to separate them so they are not plotted directly on top of each other. It is a great technique to use in dot plots, box plots with dots, and scatter plots, because it allows the reader to more easily observe differences in values that would otherwise be plotted on top of each other; thus, it helps the reader to generate insights from the chart.

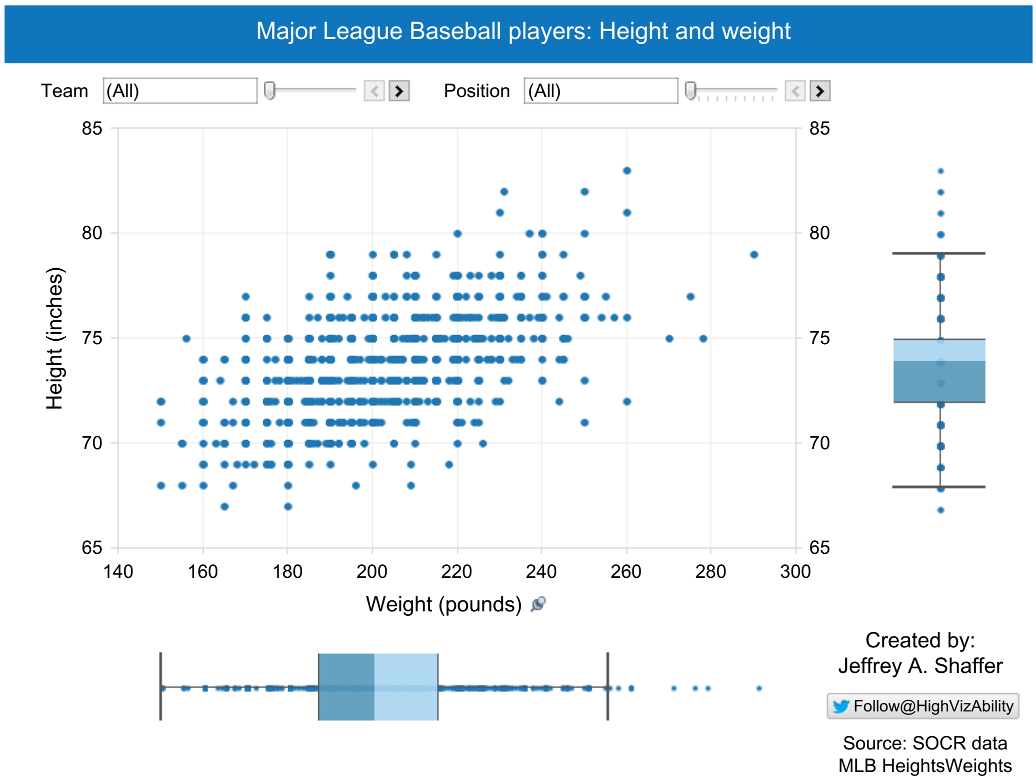
Dots can also be used in a scatter plot (with or without jitter). Figure 10 combines a traditional scatter plot

Figure 9. (Color online) Jitter Adds Some Randomness in the Horizontal Position of the Dots and Performance Bands Allow the Easy Designation of Different Grade Ranges



Notes. The use of jitter allows a viewer to easily differentiate the dots. We can see its value by comparing the chart on the right with the chart on the left in which all dots are in the same horizontal position.

Figure 10. (Color online) A Scatter Diagram and Box Plots Can Display the Relationship Between Two Variables and Are Useful for Displaying the Distributions of Correlated Variables



with a box plot on the horizontal and vertical axes to show the distribution of the height and weight of Major League Baseball players.

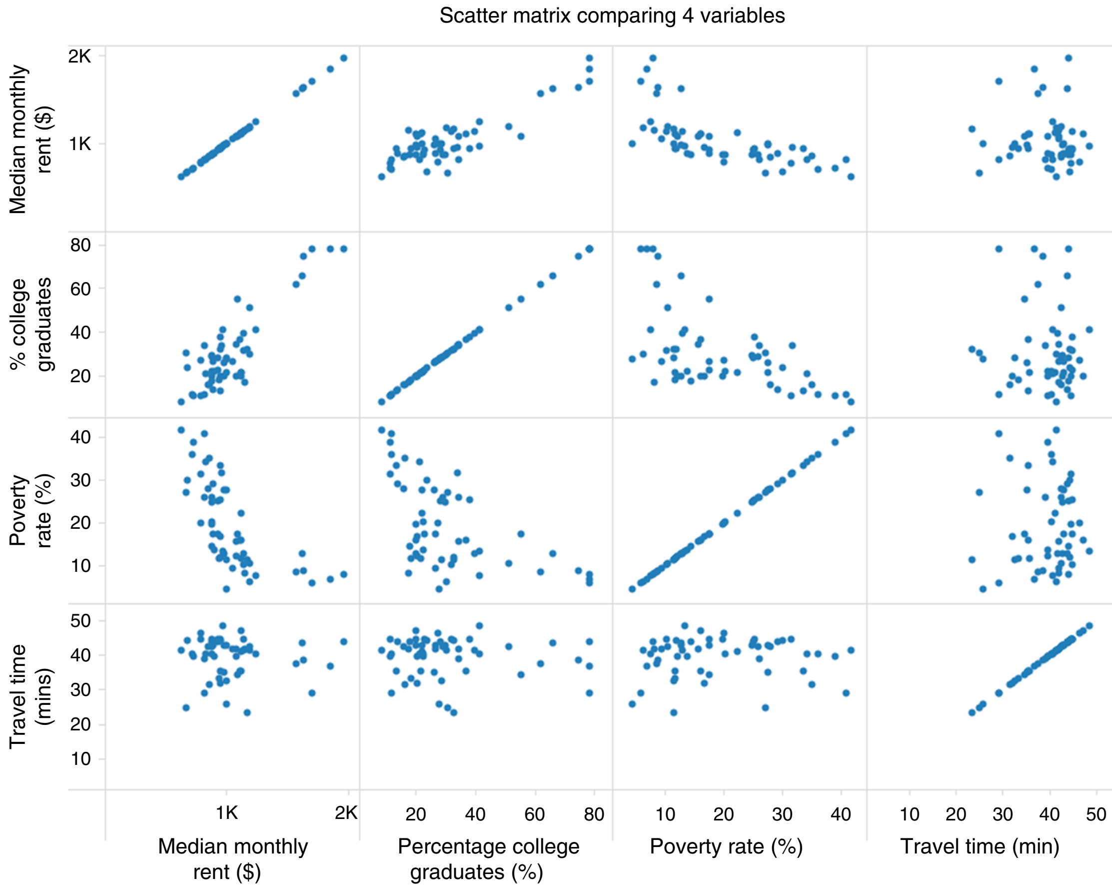
Scatter plots are also suitable in comparing many variables using a scatter matrix—a matrix that displays correlations among multiple variables. Figure 11 shows the correlation among several variables related to aspects of rental populations in different New York City boroughs. We see that the percentage of college graduates correlates positively with median monthly rent, but correlates inversely with poverty rate. Scatter matrices can be beneficial for displaying correlation among multiple variables and can often be helpful in designing more sophisticated descriptive and predictive models.

A histogram is a basic chart type that is used to display the distribution of data and compare frequencies across bins of ranges. Figure 12 shows a histogram of delivery times (in days) for a supplier. Operations managers can use this chart to quickly determine that most deliveries are in the ranges spanning three to six days, but can also see that one delivery took over 10 days.

Another common comparison measure is rank order. A slope graph is useful for this type of comparison. Figure 13 shows an example of the change in car production from Year 1 to Year 5 for car manufacturers. The axis on the left displays the production volume in Year 1; the axis on the right displays the production volume in Year 5. The line connecting the dots indicates the change (slope) between Years 1 and 5 for each car manufacturer.

Although many other chart types are appropriate in data visualization, listing them all and giving an example of each would be impossible. We suggest avoiding the following chart types, because better visualization tools, such as bar charts, stacked bar charts, or side-by-side bar charts, are usually available: pie chart, donut chart, funnel chart, gauges, pyramid chart, and radar chart.

Figure 11. (Color online) A Scatter Matrix Is Useful for Showing the Relationship Between Several Pairs of Variables and Visualizing Correlations Between Variables



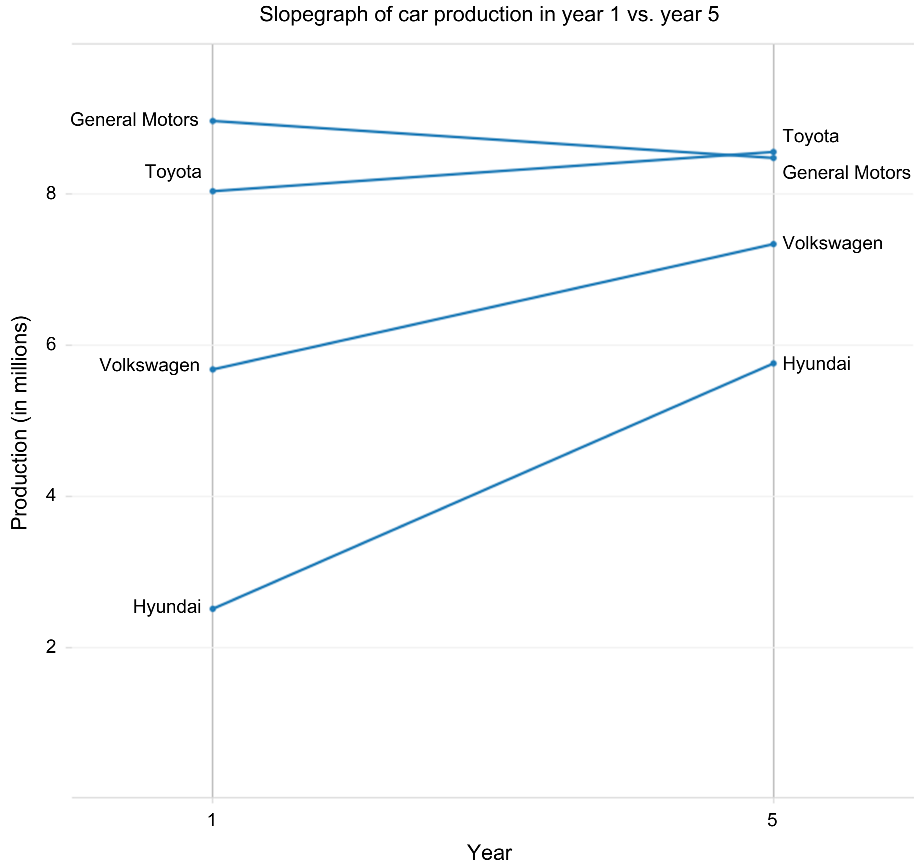
In addition, word clouds and bubble charts may be useful in certain instances, but they are not ideal for precise quantitative comparisons. Word clouds may be useful as a high-level comparison of the frequency of words occurring in a text source, but they are not useful when trying to quantify the exact number of words in text analysis. Bubble charts can be useful when the size of the bubble is a secondary factor in the analysis. Color and shapes used in conjunction with scatter charts can also convey additional dimensions in a visualization. However, as we discuss in the following sections, we must be careful to avoid too much clutter and overuse of color.

Figure 12. This Frequency Histogram of Supplier Performance Shows That Most Supplies Are Delivered Within Three to Six Days

Avoid Clutter

Any element that does not contribute to the intended message of a data visualization is clutter. Clutter distracts from the message of the visualization; therefore, it requires more mental effort on the part of the viewer. As we previously mention, Tufte's data-ink ratio is a way to measure clutter, and the elimination of clutter

Figure 13. (Color online) A Slopegraph Compares Changes in Ranking; Change in Values Is Indicated by the Line Connecting the Dots Between Year 1 and Year 5



will result in a data-ink ratio close to one (i.e., ink is used only to convey the message).

Knaflic (2015) discusses the Gestalt Principles of Visual Perception and how they can aid in eliminating clutter. These six principles, which allow us to create order from visual stimuli are:

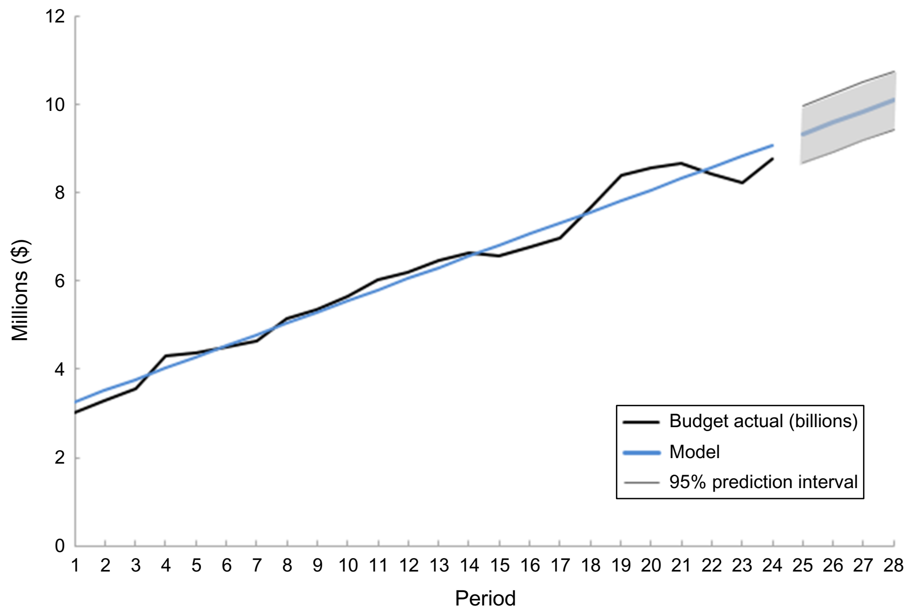
- Proximity: We place objects that are close together in a group;
- Similarity: We place objects with the same shape and (or) color in a group;
- Enclosure: We place objects that are enclosed in a group;
- Closure: If enough of a shape is shown, we map the shape to something familiar to us;
- Continuity: We expect objects that follow a path or curve to be more related than those that do not;
- Connection: We place objects that are connected in a group.

Understanding these principles often allows us to eliminate clutter and increase our data-ink ratio. The following example illustrates the principles of continuity, connection, and enclosure.

Assume we have 24 years (Periods 1–24) of budget data (measured in billions of dollars) and wish to predict the budget for the next four years (Periods 25–28). Using linear regression, we obtain a predictive model $budget_t = 0.2529 + 3.0135t$, where t = the period. Our objectives in constructing a graphic are to convey to our audience of budget analysts how well the predictive model fits historical data, to show the predictions from the model, and to emphasize the uncertainty associated with those predictions.

Figure 14 shows the results of this regression with an overlaid line chart. We use continuity with the darker (black) line giving the observed budget over time and the lighter (blue) line the estimated or fitted budget

Figure 14. (Color online) This Overlay Line Chart Illustrates the Results of a Predictive Model of Budget



from the model. We use connectedness and the break in the lines between Periods 24 and 25 to distinguish between the budgets of past periods and those in the future. Finally, we use enclosure to emphasize, based on the 95 percent prediction interval, the uncertainty of the predictions in Periods 25–28.

Effective and Sparing Use of Color

Color should not be used simply to "spice up" a visualization or to avoid a "boring" visualization. It should be used purposefully. For example, it can effectively draw the attention of the reader, highlight a portion of data, or to distinguish between different categories. For a discussion of the use of color in data visualizations, see Wexler et al. (2017, pp. 15–18).

Color can be used purposefully in data visualization in three primary ways: categorical, sequential, and diverging.

Categorical color uses different color hues to distinguish between different categories. Examples include categories that (1) involve apparel, such as shoes, socks, shirts, hats, and coats, and (2) vehicle types, such as cars, minivans, trucks, and SUVs. This color scheme can be used in many different chart types.

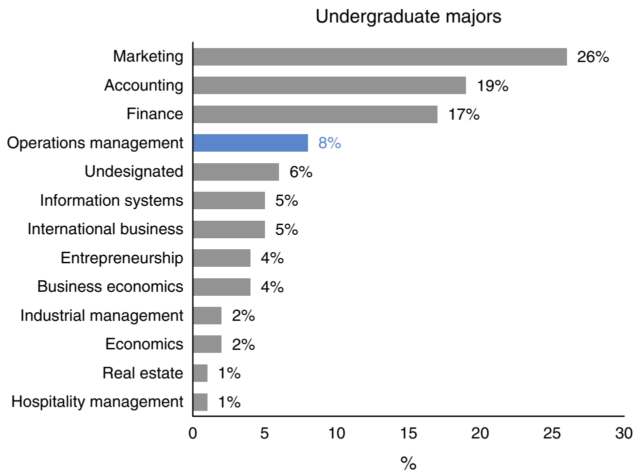
Sequential color can be used to encode a single measure from low to high, for example, using light to dark shades of one color to show intensity. For example, the quantities of parts in a warehouse could be encoded by shades of grey, where a darker grey indicates higher quantities and a lighter grey indicates lower quantities.

A diverging color palette is similar to sequential, but has a natural midpoint from which two sequential colors diverge. For example, a map that uses diverging colors to show profit and loss by state could represent profit by blue colors and loss by orange colors. Darker blue means higher profit and lighter blue means lower profit; a lighter shade of orange indicates smaller losses and a darker shade larger losses. Thus, divergence is really a combination of the first two types of color usage, two categories of sequential color diverging from a midpoint or zero.

Let us consider a simple example of the use of categorical color by extending our previous example of undergraduate majors (Figure 5). If we are preparing to use Figure 5 for an operations management faculty meeting and our goal is to convey the percentage of undergraduate operations management majors relative to other majors, we can easily use color to draw attention to operations management (Figure 15).

Figure 16 shows an example of sequential use of color—a map showing the population density

Figure 15. A Sorted Horizontal Bar Chart of the Percentage of Undergraduates by Major for a College of Business Emphasizes Operations Management



Note. This is an example of an effective use of categorical color to draw the reader's attention to a particular variable.

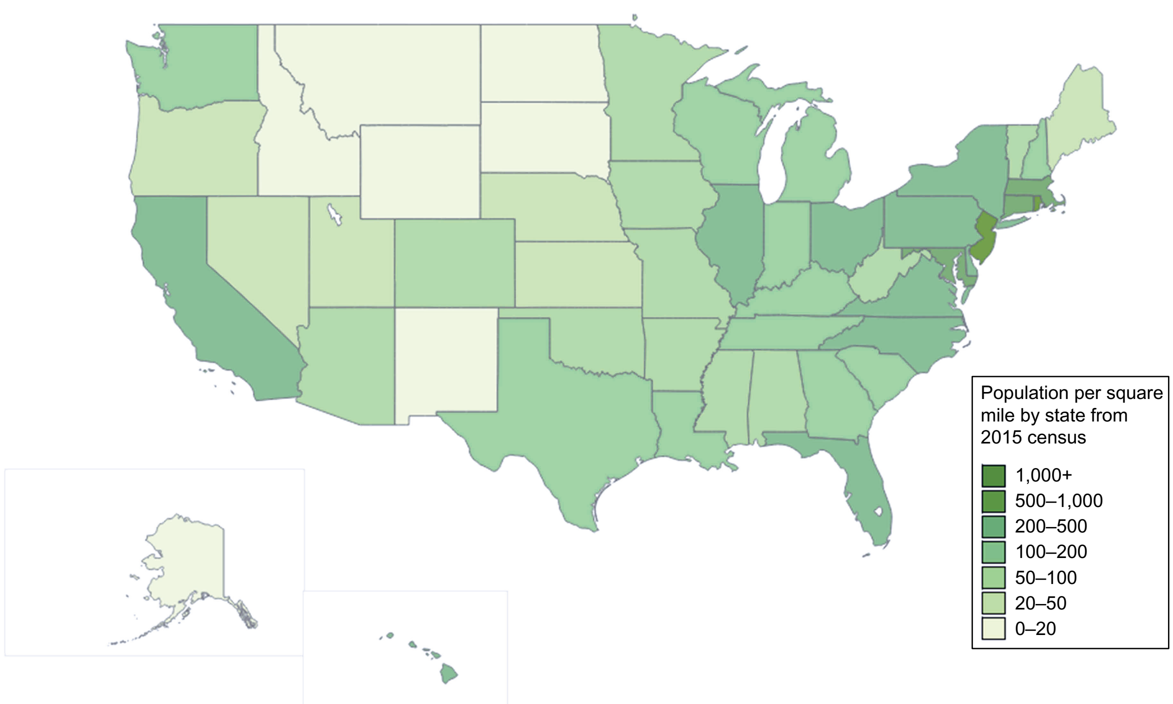
(population per square mile) for each state in the United States. From this map, a reader can quickly surmise that the population density is greatest in the Northeastern states and that the states in the Eastern part of the United States tend to have higher population density than those in the West.

Figure 17 shows a diverging use of color. The figure uses a blue-orange scale to show monthly same-store sales for 10 retail outlets.

As we have seen, color can be used effectively in delivering the message about data. However, when communicating with color, it is important to consider what we know about the presence of color blindness or color vision deficiency (CVD). Birch (1993) found that one in 12 men (eight percent) and one in 200 women (0.5 percent) have CVD, and their primary issue is distinguishing between red and green. Consequently, we advise against the use of red and green together in a data visualization.

One common solution among data-visualization practitioners is to rotate around the color wheel to blue-orange. Using blue instead of green to indicate "good" and orange instead of red to indicate "bad" works well; almost everyone (with very rare exceptions) can distinguish blue and orange from each other. Thus, the

Figure 16. A Map Showing Population Density (Population per Square Mile) for Each State of the United States Illustrates the Sequential Use of Color



Note. Darker shades of green indicate higher population density.

Figure 17. A Heat Map Shows Percentage Changes in Same-Store Sales for 10 Retail Outlets by Month

Percentage change in same-store sales

Store	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Anderson	5	3	2	7	8	7	8	5	8	10	9	8
Beechmont	3	4	5	9	4	2	5	3	2	0	1	2
Bellevue	7	7	7	14	12	8	5	9	10	9	7	6
Covington	12	14	13	17	12	11	9	8	7	8	5	3
Dayton	−2	−1	−1	0	2	4	3	5	6	7	8	8
Delhi	−6	−6	−8	−4	−5	−5	−4	−7	−4	−2	0	−2
Florence	−8	−5	−6	−2	4	7	9	12	11	12	14	12
Ludlow	5	3	0	6	4	4	5	6	10	11	8	7
Newport	−1	−2	−1	0	2	3	3	4	7	7	8	8
Walnut Hills	−4	−5	−3	−4	−1	−4	−5	−3	0	−2	0	−1

Notes. This heat map illustrates divergent use of color using blue and orange. Shades of blue indicate nonnegative percentage change in same-store sales (more intense shades of blue are higher positive same-store sales), whereas shades of orange indicate negative (more intense shades of orange are higher negative same-store sales).

blue-orange palette is often referred to as a "color-blind friendly" palette.

A Final Example: Using the Three Principles

We provide one final example that illustrates the use of the three data visualization principles and the use of a visualization for communicating visually the output and (or) recommendation from a prescriptive model, that is, a linear program.

The Calhoun Textile Mill Case (Camm et al. 1987), based on a real project, discusses a make-versus-buy decision that can be modeled as a linear program. For the upcoming quarter, Calhoun has known demand for each of its 15 fabrics, which can be made on a dobby loom or a regular loom. The number of each type of loom is given; therefore, the number of available loom hours of each type, ignoring changeovers and downtime, can be calculated easily. For a given fabric, although the production rate on a dobby or regular loom is the same, Fabrics 1–4 cannot be made on a regular loom, and total loom capacity is insufficient to satisfy demand. Therefore, some demand must be outsourced. The cost per yard to produce each fabric and the outsourced cost per yard for each fabric are given. Management wants to know the amount of each fabric to produce on each type of loom and the amount to outsource. Hence, once we have solved the problem, our objective is to develop a graphical display to communicate to management the model's recommendation and give insights on why the solution makes sense. For the curious, the full linear programming model is given in the appendix.

The first author, seduced by Excel's three-dimensional graphing capability, originally used Figure 18 to convey the recommendation and explain the solution.

This figure clearly violates all three of the guiding principles we presented. The design and layout do not facilitate a clear understanding of the recommendation. The third dimension adds no value in terms of conveying the message and, because of the use of three dimensions, some values are not even visible. Is any of Fabric 9 made on a dobby loom? This is impossible to

Figure 18. The Solution to the Calhoun Textile Mill Case Is Displayed As a Three-Dimensional Bar Chart

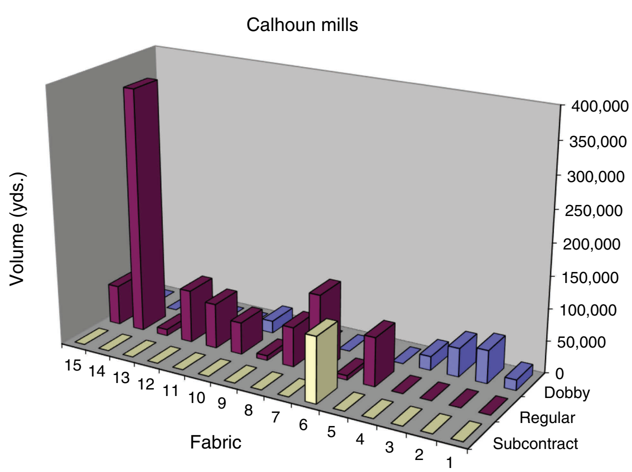
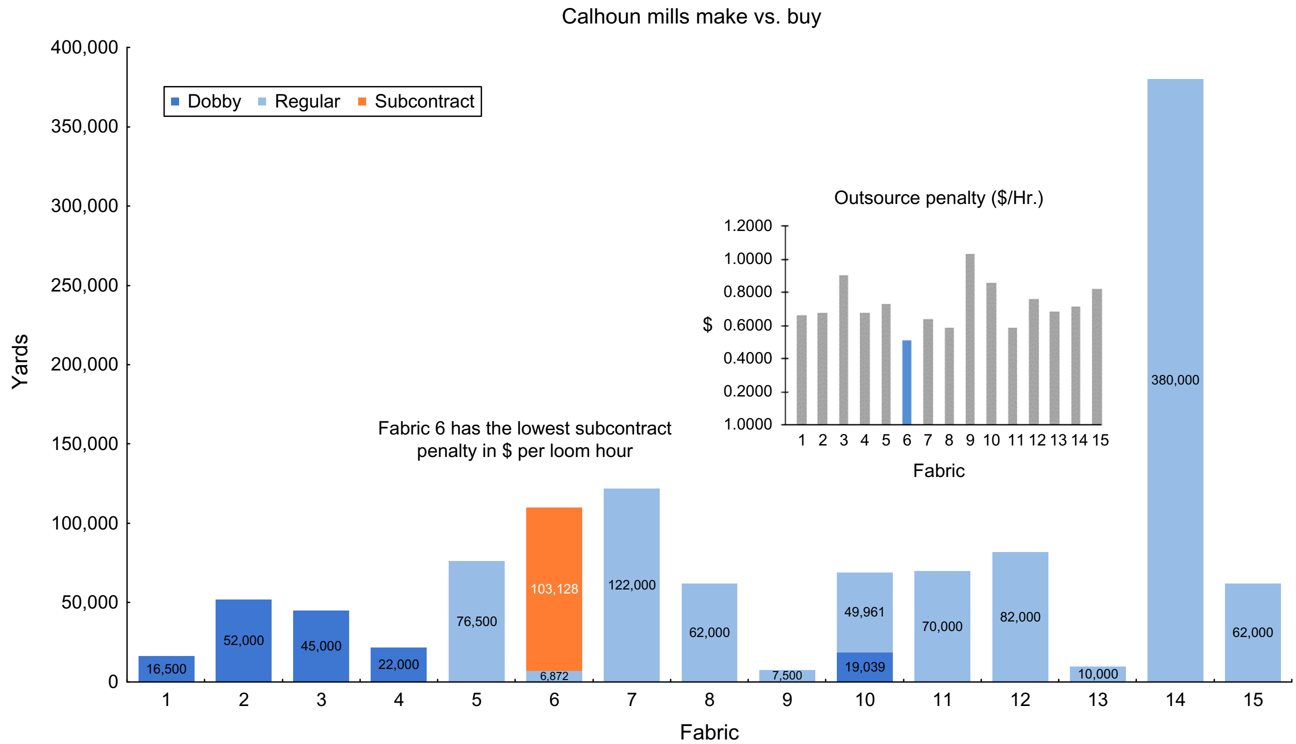


Figure 19. In a More Thoughtful Visualization of the Calhoun Textile Case Solution, We Use a Stacked Bar Chart to Show Production Amounts on the Dobby and Regular Looms, and on Subcontracted Amounts



Notes. We use color to draw attention to Fabric 6, which is the only fabric outsourced. The embedded bar chart explains why Fabric 6 is outsourced by displaying the outsource penalty per yard for each fabric.

determine from Figure 18. In addition to the unnecessary dimension, the unnecessary "wall" shading clutters the space and dramatically decreases the data-ink ratio. The label for the vertical axis is too far from the axis scale. While color is used somewhat sparingly, using color rectangles to indicate zero yards is unnecessary and distracting.

Compare Figures 18 and 19. In Figure 19, we use color and a stacked bar for each fabric. We can easily visually compare volumes by type of loom for each fabric. Nothing is hidden. We anticipate management's first question, "Why Fabric 6?" by drawing more attention to the fact that Fabric 6 is the only fabric outsourced, and we explain that it is the best choice because it has the lowest outsource penalty per loom hour of any of the fabrics. This final point is illustrated using an embedded bar chart showing the outsource penalty per loom hour for each fabric type. If this appears too cluttered, this last point could also be made in a separate graph. Figure 19 provides an example of trying to convey complicated information in a visualization in which we attempt to avoid overwhelming or confusing the reader. Knaflic (2015) includes an extended discussion of data visualization to reduce cognitive load for the reader.

Summary

Knowing your audience and the message you intend to deliver are of critical importance in creating an effective visualization. In this tutorial, we discussed best practices for data visualization, including three general principles: (1) *design and layout matter*: choose a design and layout that facilitate your message to your audience; (2) *avoid clutter*: the use of ink that does not contribute to your story will likely distract your audience and confuse the message; and (3) *use color purposely and effectively*: do not overuse color, but use it purposely to convey your message. On the surface, these principles might seem like common sense; however, in our experience, they are much like the rule that says "add comments to your code." Everyone agrees that following them is a good idea; however, doing so requires focus and a willingness to take the time to follow simple practices. In that good data visualization is a way to quickly and effectively communicate quantitative information, the principles apply to all types of analytics: descriptive, predictive, and prescriptive.

We recommend several excellent texts for further reading on best practices in data visualization. These include two classic texts, Tufte (1997, 2001), and the more recent texts, Cairo (2012) and Few (2009, 2012). Few (2006) also discusses data visualization in the context of data dashboards. Finally, for a treatment of data visualization and storytelling, we recommend Knaflic (2015).

Appendix

Let d_i = number of yds. of fabric i to produce on dobby looms $i = 1 \ldots 15$

r_i = number of yds. of fabric i to produce on regular looms $i = 5 \ldots 15$

o_i = number of yds. of fabric i to purchase outside $i = 1 \ldots 15$

Minimize

$$\begin{aligned}&0.6573d_1+0.555d_2+0.655d_3+0.5542d_4+0.6097d_5+0.6153d_6\\&+0.6477d_7+0.488d_8+0.5029d_9+0.4351d_{10}+0.6417d_{11}\\&+0.5675d_{12}+0.4952d_{13}+0.3128d_{14}+0.5029d_{15}+0.6097r_5\\&+0.6153r_6+0.6477r_7+0.488r_8+0.5029r_9+0.4351r_{10}\\&+0.6417r_{11}+0.5675r_{12}+0.4952r_{13}+0.3128r_{14}+0.5029r_{15}+0.8o_1\\&+0.7o_2+0.85o_3+0.7o_4+0.75o_5+0.75o_6+0.8o_7+0.6o_8+0.7o_9\\&+0.6o_{10}+0.8o_{11}+0.75o_{12}+0.65o_{13}+0.45o_{14}+0.7o_{15}\end{aligned}$$

subject to

$$\begin{aligned}d_1+o_1&=16{,}500\\d_2+o_2&=52{,}000\\d_3+o_3&=45{,}000\\d_4+o_4&=22{,}000\\r_5+d_5+o_5&=76{,}500\\r_6+d_6+o_6&=110{,}000\\r_7+d_7+o_7&=122{,}000\\r_8+d_8+o_8&=62{,}000\\r_9+d_9+o_9&=7{,}500\\r_{10}+d_{10}+o_{10}&=69{,}000\\r_{11}+d_{11}+o_{11}&=70{,}000\\r_{12}+d_{12}+o_{12}&=82{,}000\\r_{13}+d_{13}+o_{13}&=10{,}000\\r_{14}+d_{14}+o_{14}&=380{,}000\\r_{15}+d_{15}+o_{15}&=62{,}000\end{aligned}$$

$$\begin{aligned}&0.1925r_5+0.2625r_6+0.2389r_7+0.1911r_8+0.1911r_9\\&+0.1911r_{10}+0.2679r_{11}+0.2389r_{12}+0.2253r_{13}\\&+0.1911r_{14}+0.2389r_{15}\le 196{,}560\end{aligned}$$

$$\begin{aligned}&0.2149d_1+0.2149d_2+0.2149d_3+0.2149d_4+0.1925d_5\\&+0.2625d_6+0.2389d_7+0.1911d_8+0.1911d_9+0.1911d_{10}\\&+0.2679d_{11}+0.2389d_{12}+0.2253d_{13}+0.1911d_{14}\\&+0.2389d_{15}\le 32{,}760\end{aligned}$$

d_i, r_i, $o_i \ge 0$ for all i.

References

Baker P (2017) The best data visualization tools of 2017. Accessed October 17, 2017, http://www.pcmag.com/roundup/346417/the-best-data-visualization-tools.

Berinato S (2016) *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations* (Harvard Business School Publishing Corporation, Boston).

Birch J (1993) *Diagnosis of Defective Colour Vision* (Oxford University Press, New York).

Cairo A (2012) *The Functional Art: An Introduction to Information Graphics and Visualization* (New Riders, Berkeley, CA).

Camm JD, Dearing PM, Tadisina SK (1987) The Calhoun Textile Mill Case: An exercise on the significance of linear programming model formulation. *IIE Trans.* 19(1):23–28.

Few S (2006) *Information Dashboard Design* (O'Reilly Media, Sebastopol, CA).

Few S (2009) *Now You See It* (Analytics Press, Oakland, CA).

Few S (2012) *Show Me the Numbers*, 2nd ed. (Analytics Press, Oakland, CA).

Kahneman D (2011) *Thinking, Fast and Slow* (Farrar, Straus and Giroux, New York).

Keller C, Kros J (2000) Teaching communication in an MBA operations research/management science course. *J. Oper. Res. Soc.* 51(12):1433–1439.

Knaflic CN (2015) *Storytelling with Data* (John Wiley & Sons, New York).

Levasseur RE (1991) Effective communication—A critical skill for MS/OR professionals. *Interfaces* 21(2):22–24.

Levasseur RE (2013) People skills: Developing soft skills—A change management perspective. *Interfaces* 43(6):566–571.

Predictive Analytics Today (2017) Top 74 data visualization software. Accessed April 21, 2017, http://www.predictiveanalyticstoday.com/top-data-visualization-software/.

Sallam RL, Howson C, Idoine CJ, Oestreich TW, Richardson JL, Tapadinhas J (2017) Magic quadrant for business intelligence and analytics platforms. Report, Gartner, Stamford, CT.

Tufte ER (1997) *Visual Explanations* (Graphics Press, Cheshire, CT).

Tufte ER (2001) *The Visual Display of Quantitative Information*, 2nd ed. (Graphics Press, Cheshire, CT).

Wexler S, Shaffer J, Cotgreave A (2017) *The Big Book of Dashboards* (John Wiley & Sons, New York).

Woolsey G (1979) The fifth column—An exemplary essay on communication or corporate style, corporate substance, & the sting. *Interfaces* 9(3):10–12.

Woolsey G (1981) The fifth column—Yet another essay on effective communication or on learning where to begin, how to listen, and when to stop. *Interfaces* 11(1):10–12.

Jeffrey D. Camm is Associate Dean of Business Analytics and the Inmar Presidential Chair in Business Analytics at the Wake Forest University School of Business. He received his PhD in management science from Clemson University and a BS in mathematics from Xavier University. Prior to joining Wake Forest, he was a faculty member at the University of Cincinnati. Professor Camm's scholarship is on the application of optimization modeling to difficult decision problems in a diverse set of application areas including, operations planning and scheduling, supply chain optimization, product design, and conservation. He has received several teaching awards including the 2006 INFORMS Prize for the Teaching of OR/MS practice.

Michael J. Fry is Professor and Head of the Department of Operations, Business Analytics and Information Systems and has been named a Lindner Research Scholar in the Carl H. Lindner College of Business at the University of Cincinnati. Professor Fry has served as Interim Director of the University of Cincinnati Center for Business Analytics, and he has been a visiting professor at Cornell University and the University of British Columbia. He earned his PhD and MS degrees in industrial and operations engineering from the University of Michigan, and a BS degree from Texas A&M University. He has consulted for a variety of organizations including Great American Insurance Group, Procter & Gamble, Cardinal Health, Boeing, Copeland Corporation, Coleman Company, Vancouver City Savings Credit Union, and others.

Jeffrey Shaffer is Vice President of Information Technology and Analytics at Unifund. He is also adjunct professor at the University of Cincinnati teaching data visualization where he was the 2016 Outstanding Adjunct Professor of the Year in Operations, Business Analytics and Information Systems. He is the co-author of *The Big Book of Dashboards* and a regular speaker on the topic of data visualization, data mining, and Tableau training at conferences, symposiums, workshops, universities, and corporate training programs. He is one of 27 Tableau Zen Masters in the world and the winner of the 2014 Tableau Quantified Self Visualization Contest which led him to compete in the 2014 Tableau Iron Viz Contest.