Expressing Proportions

Stephen Few, Perceptual Edge
Visual Business Intelligence Newsletter
April/May/June 2016

Expressing the proportional relationship of one value to another is a common task when working with quantitative data. We often need to communicate either the degree to which one value compares to a greater value, such as expenses that equal 79% of the budget, or the degree to which one value compares to a lesser value, such as expenses that are 123% of the budget (i.e., 23% over budget). Proportions can be tricky to understand when they aren't expressed in the clearest, most informative way. In fact, they can be misleading and often are.

The old line that I recall from Crest toothpaste ads when I was a kid—"Four out of five dentists recommend Crest"—seemed straightforward and simple, but was far from it. The reference class wasn't declared (Which dentists and how many?). The ad might have been referring to five dentists at a dental convention sponsored by Crest. Even when our intentions are honest and unbiased, we can easily create confusion without intending to when we express proportions. For example, to say that this year's cases of a rare disease are 200% of last year's cases would create unnecessary alarm if the increase was from only one to two. We should express proportions with care. We can do this by following a few simple guidelines.

Forms of Expression

Proportions may be expressed in each of the following forms:

Frequency	7 out of 10
Ratio	7:10
Fraction	7/10
Rate	0.7
Percentage	70%

Each has its place, but frequencies and percentages are used most often, except in particular fields that favor other expressions. We're initially introduced to proportions when we're young as frequencies. This is because we begin to understand the basic concepts of *quantity* by counting. A children's arithmetic book might include the following figure to illustrate the proportion 7 out of 10:

For this reason, frequencies remain the form of expressing proportions that is most understandable by those who haven't become familiar with other forms through study or experience. Consequently, frequencies usually work best when communicating proportions to the general public.

If you think that everyone is comfortable with percentages, you're mistaken. Even though percentages are relatively easy to learn, they remain somewhat unfamiliar and intimidating to many people. Gerd Gigerenzer,

who has written several wonderful books about decision making, including *Risk Savvy: How to Make Good Decisions*, makes a good case for expressing proportions as frequencies when communicating information such as the risks of particular health problems and the effects of interventions, not only to the general public, but also to medical doctors. Even though most doctors are comfortable with percentages and rates, the ultra-simplicity of frequencies can still, at times, support greater understanding.

Frequencies are not the most efficient way to communicate proportions, however, so we usually choose to use other forms when we're communicating to audiences who are familiar with them. If you're giving a speech at a bookie convention, you would express the proportional chance of winning a bet as an odds ratio. If you're speaking at an epidemiological convention, however, you would probably use rates or percentages.

Because frequencies and percentages are the two most common forms of proportional expression, it's worthwhile to consider a few simple guidelines for their use, which we'll do in the next two sections.

Guidelines for Expressing Frequencies

To express a proportion as a frequency, we must choose a number to serve as the total to which the count relates. In a previous example, I chose 10 as the total to express a 70% proportion as 7 out of 10. But why 10 rather than 5, 20, 50, 100, or 1000, all shown below?

<div align="center">

7 out of 10

3.5 out of 5

14 out of 20

35 out of 50

70 out of 100

700 out of 1000

</div>

As you can see, "7 out of 10" is easier to understand than any of the other expressions. The expression "3.5 out of 5" involves a fraction, which unnecessarily complicates matters. Sticking with whole numbers usually works best. The expression "14 out of 20" and all expressions that follow are overly complicated because the quantities are larger and therefore more difficult to imagine, but there is another problem as well. The final expression illustrates this additional problem best. Notice that "700 out of 1000" suggests a level of precision that might not be appropriate. If the proportion was derived from a sample of 10 items, it might be misleading to express it using a total that exceeds 10.

Another complication that we should usually avoid is illustrated by the expression "17 out of 23." This would work fine if we're reporting a proportion that's specifically related to a total of 23, such as 17 out of the 23 subjects in a study, but in other cases it usually works best to choose a total that is a multiple of 10. This is because our number system is based on 10, which has led our brains to think most easily in multiples of 10. This is somewhat arbitrary. Not all number systems are based on 10. The fact that many are is probably due to the fact that many early counting systems evolved from using the built-in counters that we humans possess: our thumbs and fingers, totaling ten digits. Rather than saying "17 out of 23," unless 23 is specifically relevant, it would typically work better to express this proportion as a frequency of "15 out of 20."
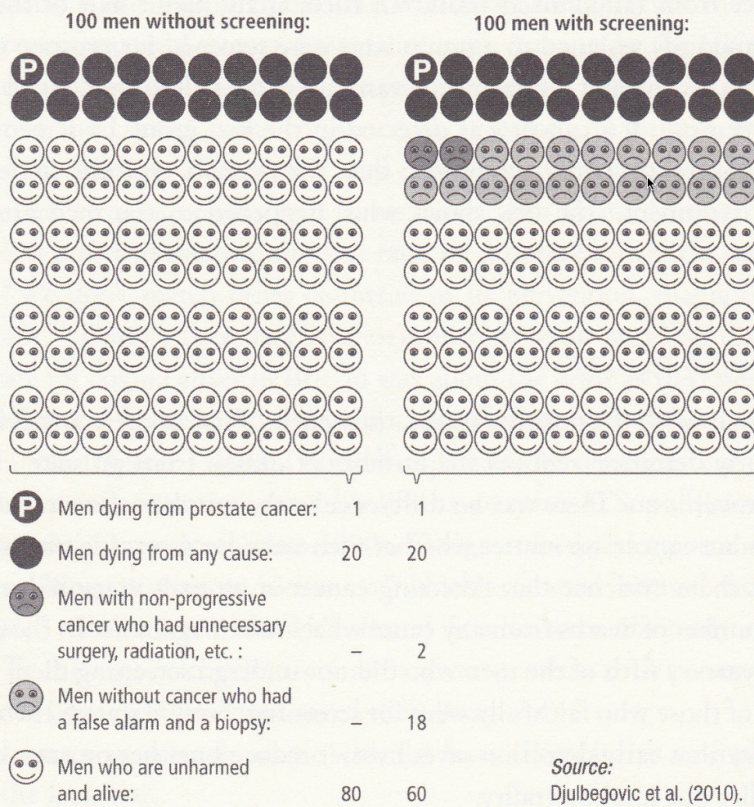
Along with his advocacy for the use of frequencies when expressing proportions, especially to people in the general public, Gigerenzer often favors the use of unit charts—charts based on frequencies that use simple icons to represent individual items—as the means to display them, such as the following example from the Harding Center for Risk Literacy:

Prostate Cancer Early Detection

by PSA screening and digital rectal examination.

Numbers are for men aged 50 years and older, not participating vs. participating in screening for 10 years.

HARDING CENTER FOR
RISK LITERACY

100 men without screening:

100 men with screening:

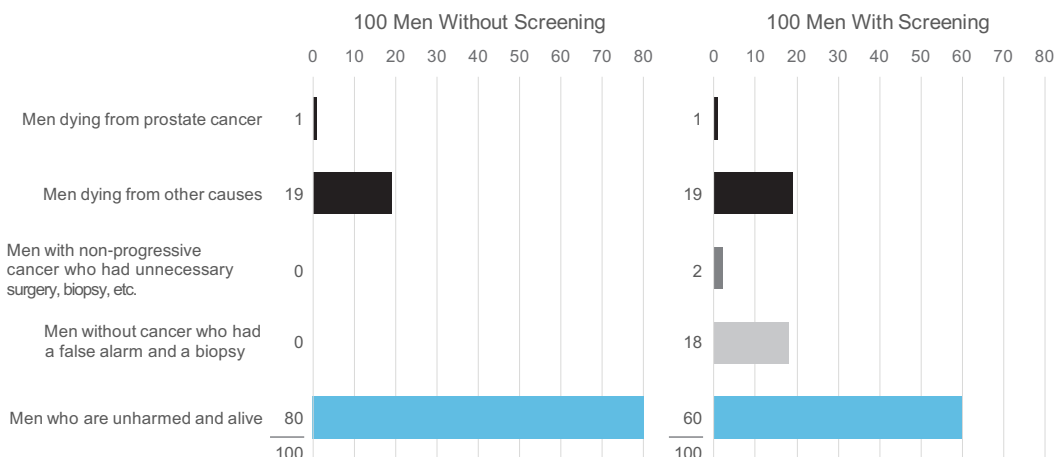|	Without	With
(P) Men dying from prostate cancer:	1	1
● Men dying from any cause:	20	20
☹ Men with non-progressive cancer who had unnecessary surgery, radiation, etc. :	–	2
☹ Men without cancer who had a false alarm and a biopsy:	–	18
☺ Men who are unharmed and alive:	80	60

Source:
Djulbegovic et al. (2010).

This unit chart was designed to help men understand the risks of prostate cancer relative to PSA (prostate-specific antigen) screening. Two sets of 100 icons each are used to show the risks faced by men who are not screened versus those who are. I'm not a fan of unit charts. I explained my reasons in the article "Unit Charts Are For Kids." This information about PSA screening could have been displayed as clearly and more efficiently using a bar graph, such as the one below.

Prostate Cancer Early Detection

by PSA screening and digital rectal examination.

These numbers are for men aged 50 years and older, not participating vs. participating in screening for 10 years.

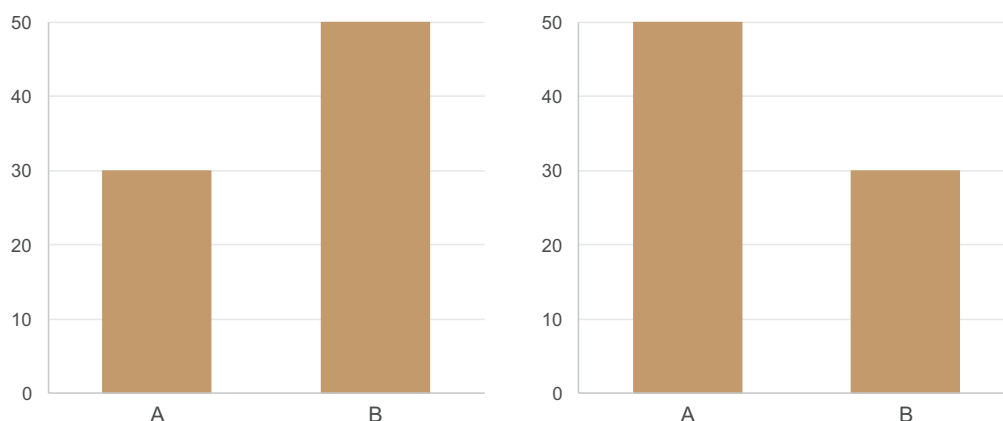We'll look further at graphs for displaying proportions later on.

Guidelines for Expressing Percentages

Unlike a frequency, we don't need to choose a total when expressing a proportion as a percentage or rate, because the total is always 100% for a percentage and 1 for a rate. This consistent standard is useful, but it creates a concern that is often overlooked. Usually, we should not express proportions as percentages that are based on data sets that consist of fewer than 100 items. For example, if we determine the proportion of people who are voting in favor of Proposition A based on a random sample of five people, of whom four say that they are voting for the proposition, we would not be justified in saying that 80% of voters are in favor of it. At best we could say that four out of a random sample of five people were planning to vote for it. Obviously, in this case, a sample of five lacks statistical power and therefore should not be considered significant. Although I haven't always been sensitive to this problem of expressing proportions as percentages when referring to small data sets, I now try to keep this concern in mind.

Another problem can occur when using percentages to express changes. What is the percentage change if the proportion of people with a particular disease increases from 10% to 15%? Is this a change of 5% or 50%? This is often expressed ambiguously, leaving us to wonder, or worse, leaving us misinformed. Unambiguously stated, this is an increase of 5 percentage points (i.e., 15% - 10% = 5%) and a 50% increase (i.e., 15% / 10% = 150% – 100% = 50%). Even when expressed in this manner, however, people often misunderstand. If 1% of the population suffered from a particular disease last year and this year it increased to 2%, that's a 100% increase. That seems like a lot, but it is still just 2% of the population. Even worse, if 0.001% of the population suffered from a disease last year and this year it increased to 0.002%, that's also a 100% increase, but it is only two thousandths of a percent.

When expressing percentage change, we should also be careful to clearly identify the intervals of time that are being compared. If you see on a dashboard that revenues went up by 12%, are you comparing yesterday to today, this week to last week, this month to last month, this quarter to last quarter, this year so far to last year on this date, this month compared to this same month last year at the same number of days into the month, and on the list of possibilities goes. Don't leave people guessing, for they will often guess wrong.

When expressing the proportion of one value to another that doesn't involve change through time, it is also helpful if you can to lead with the smaller value of the two. For example, if item A has a value of 50 and item B has a value of 100, say that A is 50% of B, rather than B is 100% greater than A. People find it more difficult to understanding proportions of one value to another when the first value is larger than the second. This is easy to illustrate. In the example below, for each graph, determine the percentage of bar A to bar B.
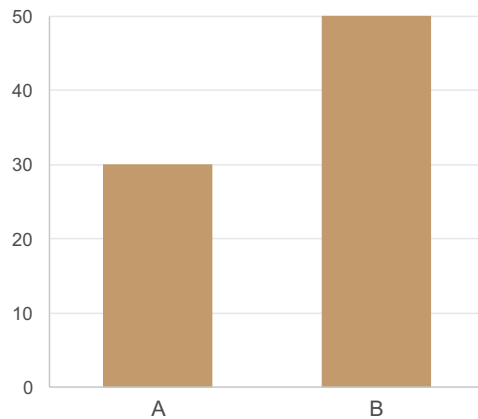

The comparison on the left was easy, wasn't it? A, with a value of 30, is 60% of B, with a value of 50. The comparison on the right, however, was more cognitively complex. A is 166% of B (i.e., 66% greater), but figuring this out involves a multi-step process. We must first determine how much greater A is than B, which is a difference of 20 units (i.e., 50 – 30 = 20). To get the full proportional difference, we must move on to compare the amount of A's value that exceeds B's to the value of B (i.e., 20 compared to 30) to come up with a proportion of 66%, and then add 100% to 66% to get the total proportional comparison of 166%. As you can

see, if the order of the two values doesn't matter, it is much easier for people to understand the proportion of one value to another when the first value is smaller than the second, so it works best to express it in this way. This option isn't always available, however, because some values should always be compared in a particular order. For example, when we compare actual expenses to the expense budget, if actual expenses exceeded the budget we would still compare actual expenses to the budget, never the other way around.

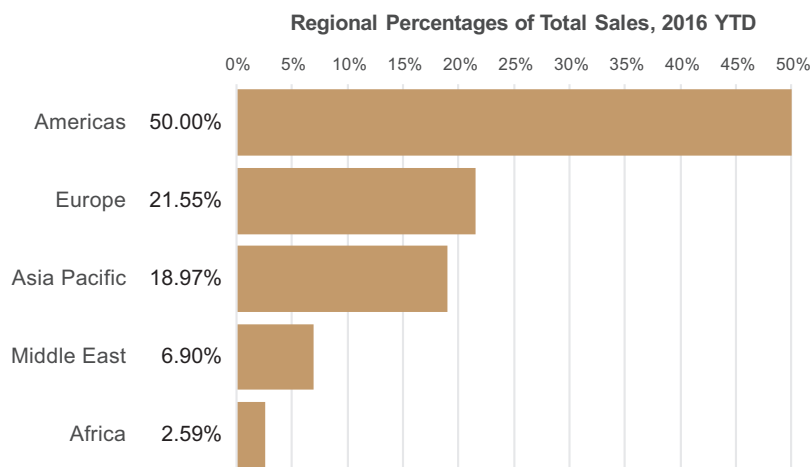## Best Ways to Present Proportions Graphically

If I were presenting a comparison of two values only, I would not necessarily display the information graphically. For example, rather than using the graph below, it would usually work just as well or better to use words and a number to simply say "A is 60% of B."
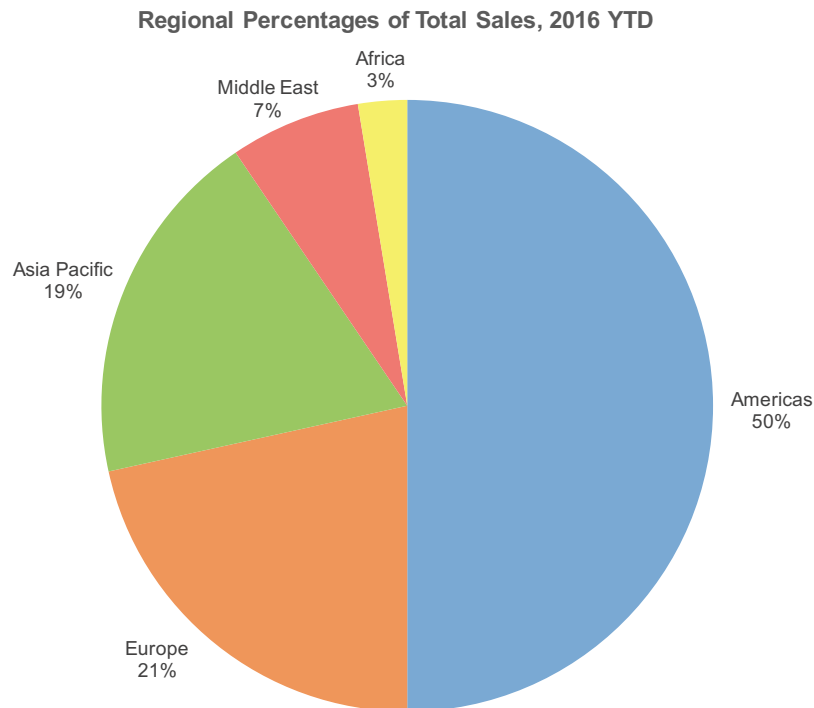

I might use a graph if I wanted to give the difference between the two values additional oomph, but generally not. Graphs are typically more useful when more than two values must be compared. If I want to show the portion of sales that each of the following regions contribute to total sales (i.e., a part-to-whole comparison), I could present the information in a simple table or I could do it in a bar graph.

Regional Percentages of Total Sales, 2016 YTD

Region	% of Sales
Americas	50.00%
Europe	21.55%
Asia Pacific	18.97%
Middle East	6.90%
Africa	2.59%

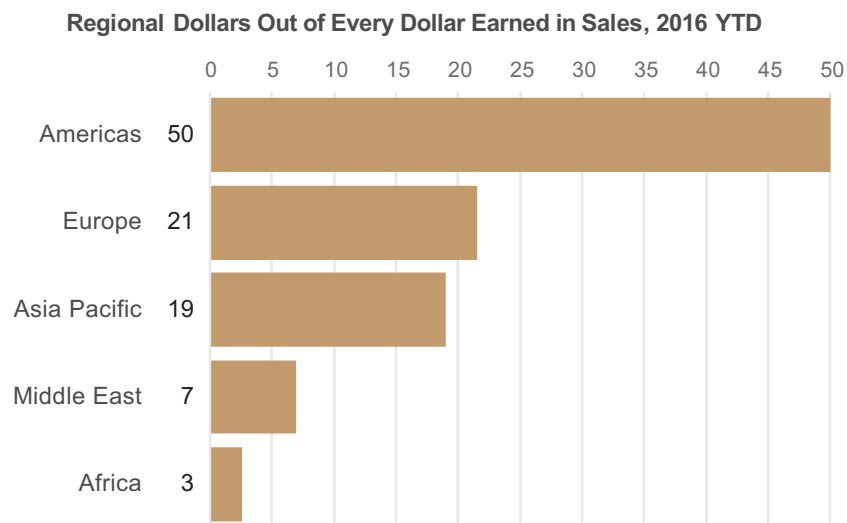

Regional Percentages of Total Sales, 2016 YTD

Although it certainly isn't necessary to present this proportional information graphically, it would be useful to make particular differences among the values more apparent (e.g., the fact that sales in the Americas are over twice sales in Europe) and to enable more rapid comparisons. When presented graphically, proportions are usually best displayed as bar graphs. Using a pie chart (see below), as most people do, is not a choice that makes proportional comparisons easy for our brains.

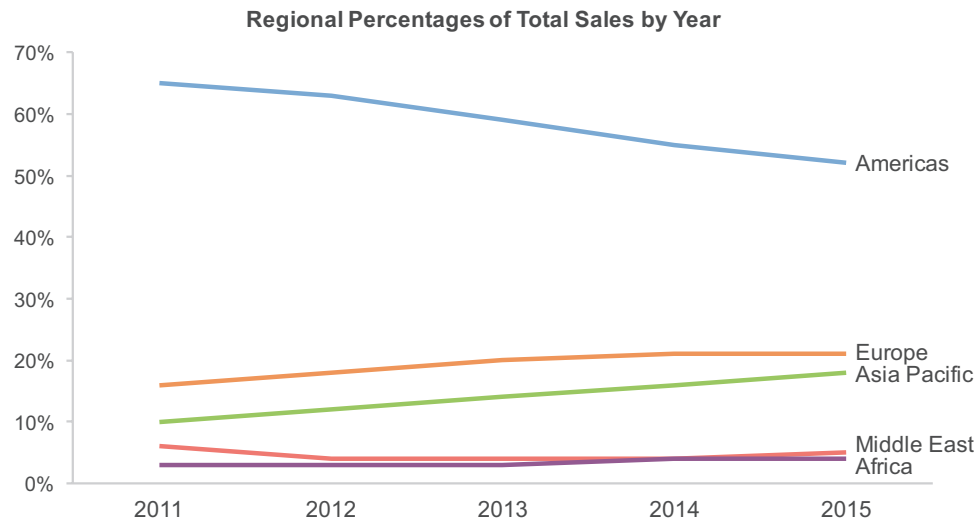**Regional Percentages of Total Sales, 2016 YTD**

If you don't understand why this is the case, I won't say again here what I've said so many times in the past. Instead, I'll refer you to an article that I wrote long ago titled "Save the Pies for Dessert."

Bar graphs can also be used when we choose to express proportions as frequencies rather than percentages, illustrated below.

Regional Dollars Out of Every Dollar Earned in Sales, 2016 YTD

The one exception to using bars for proportional comparisons is when you need to show how proportions changed through time. For example, did the regional part-to-whole sales relationship that we saw above remain

consistent through the last few years, or did the proportions change? The line graph below tells the story.

Regional Percentages of Total Sales by Year

Sales in the Americas have consistently declined for the last five years. This loss in proportion corresponds to sales increases in Europe and Asia Pacific. Neither a regular bar graph, nor a stacked bar graph, nor a stacked area graph would reveal this change in proportions as clearly as this ordinary line graph. For more about line graphs for displaying changes in proportions through time, I recommend that you read my article titled "Quantitative Displays for Combining Time-Series and Part-to-Whole Relationships."

Proportion is a fundamental quantitative relationship. We should understand proportions intimately. Skill in presenting data graphically requires an understanding of quantitative concepts and relationships, not merely skill in visual design. Many of the mistakes that people make result from understanding one but not the other. To communicate quantitative information effectively, we must integrate numeracy (skill with numbers), graphicacy (skill with graphics), and literacy (skill with words)—all three—in roughly equal proportions.

Discuss this Article

Share your thoughts about this article by visiting the Expressing Proportions thread in our discussion forum.

About the Author

Stephen Few has worked for over 30 years as an IT innovator, consultant, and teacher. Today, as Principal of the consultancy Perceptual Edge, Stephen focuses on data visualization for analyzing and communicating quantitative business information. He provides training and consulting services, writes the quarterly *Visual Business Intelligence Newsletter*, and speaks frequently at conferences. He is the author of four books: *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Second Edition, *Information Dashboard Design: Displaying Data for at-a-Glance Monitoring*, Second Edition, *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, and *Signal: Understanding What Matters in a World of Noise*. You can learn more about Stephen's work and access an entire library of articles at www.perceptualedge.com. Between articles, you can read Stephen's thoughts on the industry in his blog.