CHAPTER 5

WHAT THE SURVEY SAYS

Showing Likert, Ranking, Check-All-That-Apply, and More

LEARNING OBJECTIVES

After reading this chapter, you will be able to:

- Modify traditional graphs and create potential alternatives
- Differentiate between graph types appropriate for sequential versus Likert-type questions
- Handle the Not Applicable data category
- Hack error bars to actually increase readability of a graph
- Visualize branching questions in several ways

This chapter shares a menu of options for displaying survey data, including ranking, rating, and check-all-that-apply questions. I'll show you common visualization options, point out what the critics say about them, and offer possible solutions and alternatives that might better meet your communication needs. Our goal here is to present survey data so that readers can better understand survey respondents and effectively use the data to take more informed action. We'll deal

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

with those pesky neutral categories and walk through options for communicating sample size.

Rating scales are usually shown in a stacked bar graph, which poses significant obstacles to being seen as reader friendly. However, there are some options: design it thoughtfully with an intentional color scheme and order, or convert the data to another graph type.

A diverging stacked bar graph repositions Likert scale data so that they diverge around a central point, with the positive affect falling out to one side and the negative affect falling out to the other. It handles some of the criticisms of the regular stacked bar graph by making it easier to compare the middle values in your rating scale.

Breaking the data out into small multiples is another possibility that is particularly useful when you really want to compare each of the response options in a rating scale. Each response option will essentially appear on its own baseline, making comparisons a snap.

Even though a rating scale might have an obnoxious nine response options, that doesn't mean we have to report on all of them. Often, our audience doesn't care about the difference between Agree and Somewhat Agree. If we aggregate the similar responses and graph those larger groups, then the traditional stacked bar doesn't appear so cumbersome.

One of these alternatives will fit your rating data, but ranking data is a different story.

The simplest way to show ranking data is through a column or bar chart, ordered by frequency from greatest to least. These charts work just fine, most of the time. When do they fall short? Well, when the values in your data set are all high, such as in the 80% to 90% range (out of 100%). Or when you have to show data on all U.S. states and territories. Then, a chart with a set of tall columns can be visually aggressive. In that case, try a lollipop graph, which looks like its name—a dot for the value that sits on top of a stick, connecting the dot to an axis.

Another way to handle ranking data is to isolate the top three or five items with the highest ranking and assign an icon or picture to each item. Pair the icon with the corresponding percentage in a large, interesting font, and you could be all set.

What if you need to show ranking data over time? A bump chart is the answer, and you'll make one by modifying a line graph with markers.

Branching questions are a whole different animal, because their structures are questions, within questions, within questions. For shallow branching, you can get away with an annotation inside another graph. For deeper branching, try a nested map that visualizes a whole series of branched questions, one inside the other.

Finally, this chapter addresses the three ways to visually handle Not Applicable data:

1. If the missing data are small and consistent, just note this in the subtitle.
2. If the missing data are large and consistent, add the sample size to your data labels.
3. If the missing data are large and inconsistent, graph them off to the side.

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

This is one hefty chapter, and by the end, your ninja skills are going to be off the charts (or, literally, *on* the charts).

WHAT STORIES CAN BE TOLD ABOUT WHAT THE SURVEY SAYS?

Goodness knows, it seems as if every other phone call on a landline is someone with a survey. This type of call is so widespread because a survey remains one of the best ways to systematically discover the opinions of representatives of a group. Surveys have been conducted on every topic imaginable, so stories can include these variations:

- Respondents said Option A was most favorable.
- Most of the respondents had a negative view of our service.
- Of the options we gave, respondents liked X, H, and O best.
- Customers said our top three services are résumé coaching, mentor matching, and job skills training.
- Polls showed that 47% of the students knew a Hispanic/Latino who had dropped out of school.
- Citizens believe A and J will be good candidates.

Survey data tell the story of what people think, feel, do, or believe.

HOW CAN I VISUALIZE WHAT THE SURVEY SAYS?

So much of the decision making in the chapter leans on the survey question type. We walk through options for ranking, rating, and branching question types. Be sure to check out the resources at the end of this chapter for color, which is very important in our examples, and some visualization support for specialized question types, such as network analysis.

Rating

Rating questions come in two basic forms: those that have sequential response options (think Poor to Excellent) and those that are diverging (such as Strongly Disagree to Strongly Disagree and other Likert types). We can use similar graph types to visualize these two kinds of questions, with minor tweaks for better emphasis.

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

Stacked Bar

A stacked bar graph can be a reliable friend. Most of the time, when reporting on percentages that make up a whole, stacked bars can be trusted to come through for you. But sometimes, just like your best friends, stacked bars show up to the party loud and confused, wearing a little too much makeup. Let's look at how to tame them into being supportive of your rating data set.

Default stacked bar graphs often look like Figure 5.1.

Figure 5.1 /// Default stacked bars can be hard to read.

Sit down and drink some water, default stacked bar. What's going right is that it is easy to see that all of the Yes, Sorta, and No responses add up to 100%. Good! It's easy to see that we asked seven questions in this section of the survey. I don't really know if that's important. But I'm working hard to pull out anything good from this visual. The stacked bar as a graph type isn't all that bad; it just needs some formatting help.

The common criticism about stacked bars is that they can be hard to interpret. Let's say you were mostly curious about the students in the middle of this stacked bar, those who said Sorta. While we are good at judging the length of bars, it can be hard to compare the Sorta responses for each survey question because they don't share a common baseline. In the example, do you think there are more students who reported Sorta on "I see leadership opportunities for Latinx youth" or for "My close friends have volunteered in the past 12 months"? It's even harder to tell because the organizational color scheme doesn't distinguish the segments from one another all that well. On top of that, we have a disconnected legend

Do not copy, post, or distribute
Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

that makes our brains have to bounce around a lot to understand what each segment represents.

While we can't completely eliminate all of those problems with the stacked bar, the suggestions in this section can address the ones that are most offensive.

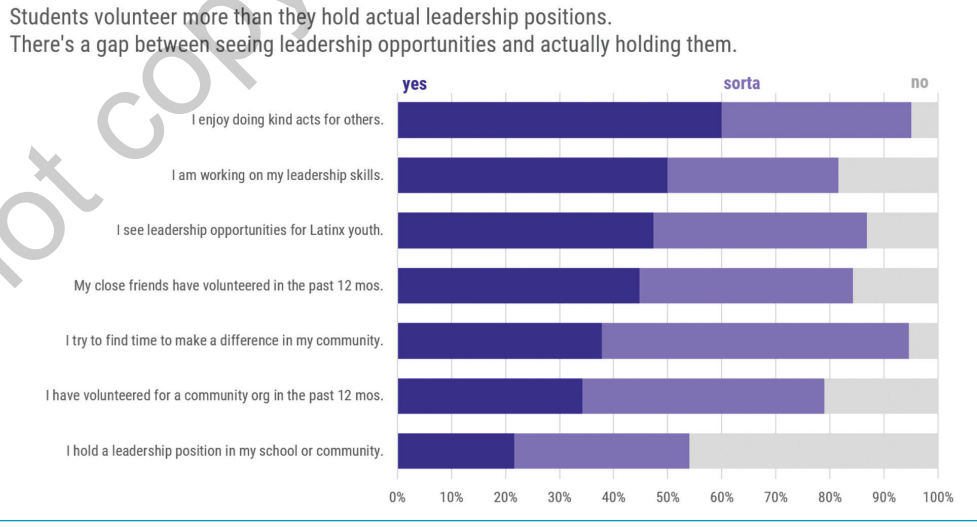
One way to handle these issues is to rock the formatting of the traditional stacked bar. Same data, better told.

Tweak 1: Sort the questions from greatest to least on the category that matters the most to your audience. In this case, we are assuming this is the Yes data. Others would be more likely to focus on the No data, or even the Sorta data. It depends on what you're trying to point out to your particular audience. Sorting the data makes it easier for readers to quickly assess what was most often reported and least often reported in this data set. The typical stacked bar has no intentional sort order—it's commonly the order that the questions were asked on the survey. The thing is, readers don't care about the order of the questions on the survey! If you sort your data table from smallest to greatest on your category of interest, Excel will graph your data greatest to least.

Excel Ninja
Level: 4

Tweak 2: Apply a sequential color scheme. Recolor each segment of the stack from darkest to lightest shades of one color. I usually apply the lightest color to the category of least importance (the No data, in this case). Matching a sequential color scheme to sequential data makes sense to readers and can help them mentally group the Yes data and the Sorta data if they want to. It also holds up better when reprinted in black and white.

Figure 5.2 /// Slight tweaks make the standard stacked bar chart easier to interpret.



Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Tweak 3: Reposition the legend. When the bar segments are horizontal but the default legend is vertical, it makes readers volley through a bunch of mental gymnastics to connect the information and interpret the graph. In this case, I just deleted the legend and inserted text boxes above each segment in the stacked bar. I linked the text boxes to the corresponding cells in my table. Remember how to do that?

Insert a text box. It will have a dashed line border and a blinking cursor (Figure 5.3), waiting for your words. Click on the border of the text box so the dashed lines become solid (Figure 5.4).

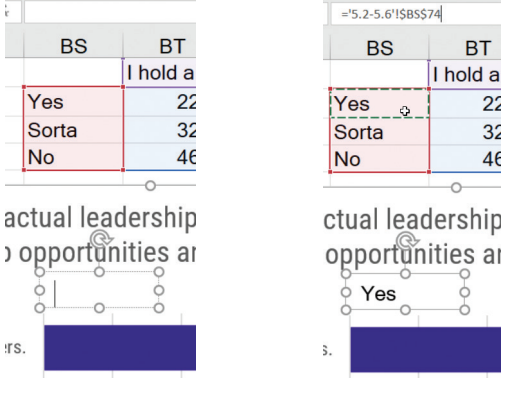
Then go straight to the formula bar in Excel and type in an equal-sign (Figure 5.4).

Then head to the cell in your table with the label you want to insert. In this case, it's the label that says Yes. Click inside that cell. Excel will add that cell's location to the formula bar, right after your equal-sign.

Hit Enter, and your text box will populate with the label. From there, you should change the color of the word "Yes" so that it matches the color of the Yes segments in the stacked bar below it. You can reposition the text box as needed, too. That kind of flexibility doesn't come with a default legend. So, if the legend isn't working for you to maximize your graph, try linked text boxes instead.

Those tweaks go a long way in making the default stacked bar work its best for your sequential data. The same strategies can help diverging response options, too. The only extra consideration involves the color-coding tweak. For diverging data sets, use a diverging color scheme. You'll need two shades of two colors and a neutral color for neutral.

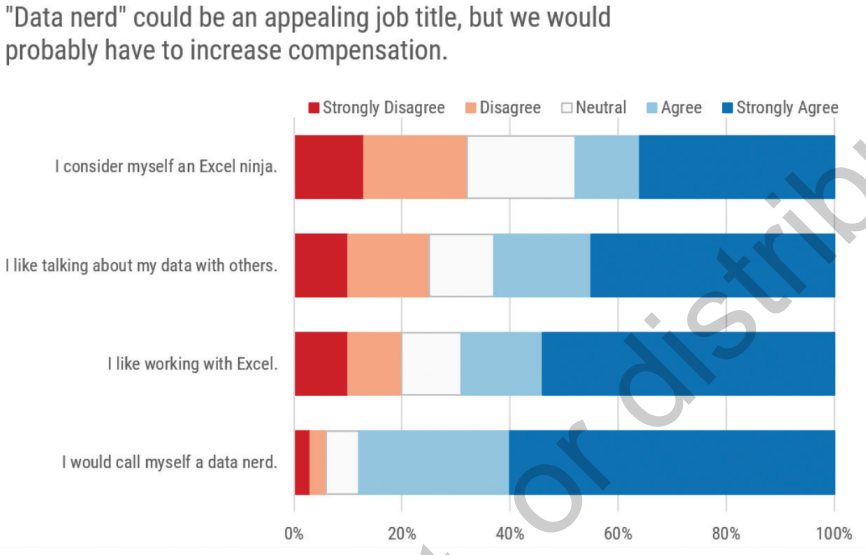
Figures 5.3 and 5.4 /// Click on the border of the text box until it changes from a dashed to a solid line. Type an equal-sign in the formula bar. Click in the cell you want as your label.



Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.5 /// Apply the darkest colors of each shade to the more strongly felt sentiments.



One end of the response set is cast in blues, with the darker blue on the outside and the lighter blue on the inside. The other end of the stacked bar gets two shades of another color (Figure 5.5). Neutral is in a neutral color. This works even better with culturally associated color schemes. In the United States, reds and oranges are viewed as more negative, so use them on the Disagree side of the data. Blues tend to be seen as more positive.

I pulled the color schemes for several of the graphs in this section from **ColorBrewer2.com**. It is a website developed by Cynthia Brewer's team, designed for cartographers (mapmakers!) who deal with situations where they need several colors to be distinguishable when placed right up next to one another. The examples in this chapter don't have anything to do with maps, but we can use the same tools because our data segments also touch each other. ColorBrewer has color schemes for both sequential and diverging data sets. The website gives you the full color scheme, along with the RGB (red, green, blue) color codes you need to customize your graph inside Excel.

Small Multiples

The previous tweaks get us further away from the loud and obnoxious default stacked bar, but I'm still unsettled because they don't entirely solve the problem of the difficulty comparing those middle values. This example and the next can give you a lift here. Let's start with small multiples and apply them to the

Do not copy, post, or distribute
Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Yes/Sorta/No data from earlier. We are going to use ninja skills to break up the stacked bar into three sets of regular bars, all in the same graph.

We could simply make three separate bar graphs, one for the Yes data, one for the Sorta data, and one for the No data. That could be totally fine if that's all the ninja skill you are up for today. The downside to that situation is the fidgeting you have to do to align all the graphs and labels and the care with which you must copy and paste them so they hang together as one visual. This next method takes a little more behind-the-scenes work, but it keeps the visuals in one tidy place. No matter which method you choose, Kong, Heer, and Agrawala (2010) found that small multiples are effective at breaking down complicated displays if you have fewer than 4,096 data points (precise, huh?).

Excel Ninja
Level: 6

We are going to use secret buffer columns to push our data so that each segment shares a common baseline. My table is in Figure 5.6.

Figure 5.6 /// Insert a column next to each actual data column in your table.

	Yes	Yes Buffer	Sorta	Sorta Buffer	No	No Buffer
I hold a leadership position in my school or community.	22%		32%		46%	
I have volunteered for a community org in the past 12 mos.	34%		45%		21%	
I try to find time to make a difference in my community.	38%		57%		5%	
My close friends have volunteered in the past 12 mos.	45%		39%		16%	
I see leadership opportunities for Latinx youth.	47%		39%		13%	
I am working on my leadership skills.	50%		32%		18%	
I enjoy doing kind acts for others.	60%		35%		5%	

My actual data are separated by a column I inserted next to each. Those inserted columns are for my secret ninja buffer data. Each actual data column gets its own buffer buddy next door. Together, the data column and its buffer buddy need to add up to 100%.

See how the Yes data and the Yes Buffer data add up to 100% for each survey question (Figure 5.7)? Compute this for each data and buffer duo.

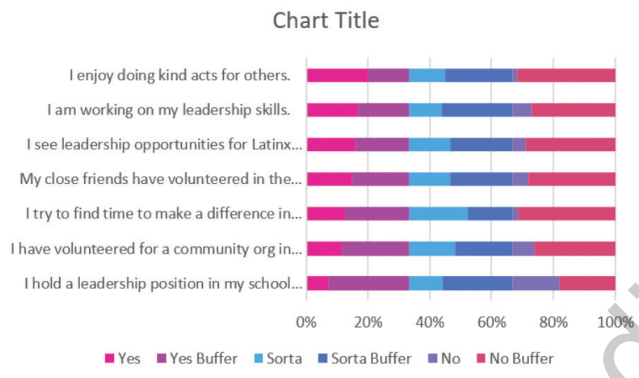
Figure 5.7 /// Subtract 100% from each value in your actual data column and put the remainder in the buffer column.

	Yes	Yes Buffer	Sorta	Sorta Buffer	No	No Buffer
I hold a leadership position in my school or community.	22%	78%	32%	68%	46%	54%
I have volunteered for a community org in the past 12 mos.	34%	66%	45%	55%	21%	79%
I try to find time to make a difference in my community.	38%	62%	57%	43%	5%	95%
My close friends have volunteered in the past 12 mos.	45%	55%	39%	61%	16%	84%
I see leadership opportunities for Latinx youth.	47%	53%	39%	61%	13%	87%
I am working on my leadership skills.	50%	50%	32%	68%	18%	82%
I enjoy doing kind acts for others.	60%	40%	35%	65%	5%	95%

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

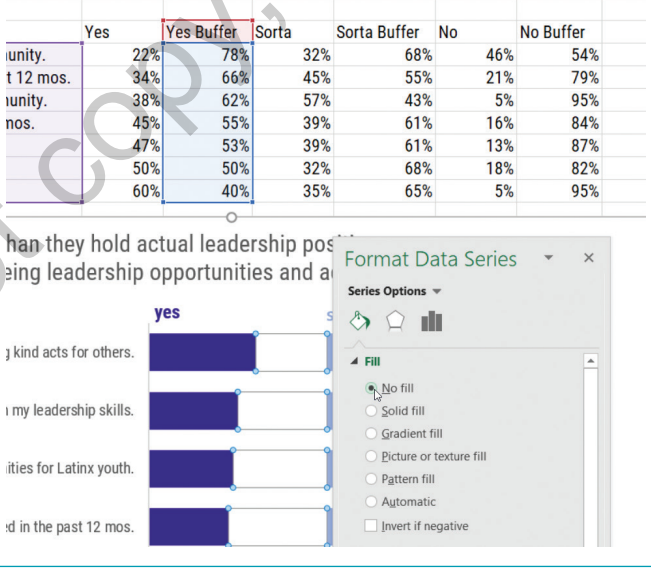
Figure 5.8 /// Select the whole table and insert a 100% stacked bar. It will look loud and obnoxious.



Now select all the data, even the headings, and insert a 100% stacked bar graph (the third option shown in the 2-D bar chart choices).

It's looking silly now, but there's not much left to do to make this chart look amazing. Like any other chart, change the font and colors, decrease the gap width, and add your awesome title and legends. Then, inside the graph, right-click on each of the buffer segments and change their color to No Fill (Figure 5.9). Choose Format Data Series and look for the paint bucket icon or the Fill menu.

Figure 5.9 /// Each buffer segment should not be filled with any color.



Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

You should already see the three separate bars in one graph! Don't stop; keep whipping this graph into shape by deleting the x-axis, gridlines, and y-axis line, since they no longer make much sense. You'll want to replace this with data labels inside each actual data segment of the stacked bar (Figure 5.10).

Figure 5.10 /// Small multiples of regular bars in the same graph make it easier to compare values that were in the middle of the stacked bar.

Now we can tell whether the Sorta responses are different for those two survey questions I mentioned earlier (they aren't). Yes, the data labels are there to tell us, but even if the labels were gone, we could compare the lengths of those Sorta bars because they now share a common baseline. It's also true that we don't necessarily need to have different colors on each set of bars, since they no longer touch one another. Make them all the same color if you aren't trying to emphasize one set of response options over another.

Supercool ninja trick, isn't it? These secret buffer zones are going to keep coming in handy. It's all about mastering the software and what it can do in order to make the best possible visualizations for your data.

Diverging Stacked Bar

Diverging stacked bar charts are great for showing the spread of negative and positive values, such as Strongly Disagree to Strongly Agree, and because they align to each other around the midpoint, they can handle some of the criticism

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

directed at regular stacked bar charts, which is that it is difficult to compare the values of the categories in the middle of the stack (Heiberger & Robbins, 2014; Talbot, Setlur, & Anand, 2014). Making a diverging stacked bar was approximately 8 billion times easier than I expected; it just takes a little ninja skill.

The secret again is hidden buffer values at either end of the bars. I told you they'd come in handy! We are just going to switch up the order of the data and the buffers in our spreadsheet.

Excel Ninja
Level: 7

Figure 5.11 is what my data table looks like after adding the buffer values.

Figure 5.11 /// I added some notes to the table to help you conceptualize how this is going to work.

	A	B	C	D	E	F	G
1		<---------- Add to 100%			Add to 100%	---------->	
2		Buffer	Strongly Agree	Agree	Disagree	Strongly Disagree	Buffer
3	I would call myself a data nerd.	15%	50%	35%	8%	7%	85%
4	I can manipulate Excel to do what I want it to do.	20%	70%	10%	10%	10%	80%
5	I should get paid well for my dataviz ninja skills.	25%	45%	30%	15%	10%	75%
6		^turn bars no fill					turn bars no fill^

Let's walk through it a bit. In the middle, in the dark text, are my actual data values, what I ultimately want to show the audience. On either side I have Buffer columns. These are the secret columns! In gray you can see my notes. So here's the plan. Within each row, all values from Disagree over to the right need to add up to 100%. Just mentally sum your Disagree and Strongly Disagree values and type the remainder in the Buffer column (or make a formula if you need to flex more Excel muscle). Same thing for Agree over to the left.

Starting in cell A2, select all the headings and values (but not my gray notes) and insert a 100% stacked bar graph. Mine looked like Figure 5.12 at first.

Not cool. I should see six segments in the stacked bar, not six bars. So look up in your toolbar for the button that says Switch Row/Column (or Switch Plot if you are on a Mac). Now it looks like Figure 5.13.

Bingo! So now it's time to make the two Buffer categories disappear. Just right-click directly on the Buffer bars and select No Fill color. When you're done, it should look like Figure 5.14.

Look at that! Now the only bars with color are the ones that encode our values. Still, it doesn't look that great, right? So, delete the legend (we'll add it back in later), delete the gridlines, and delete the y-axis line. Delete the x-axis, too . . . *but first!* We want more of the chart area taken up by our actual values. Right now, the stacked bars are squished together because Excel set the maximum of the x-axis to accommodate our Buffer zones. Right-click on the x-axis, select Format Axis, and in that dialogue box, pick a new max that is nearer to the end of the bar segments

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.12 /// Excel made six stacked bars instead of three.

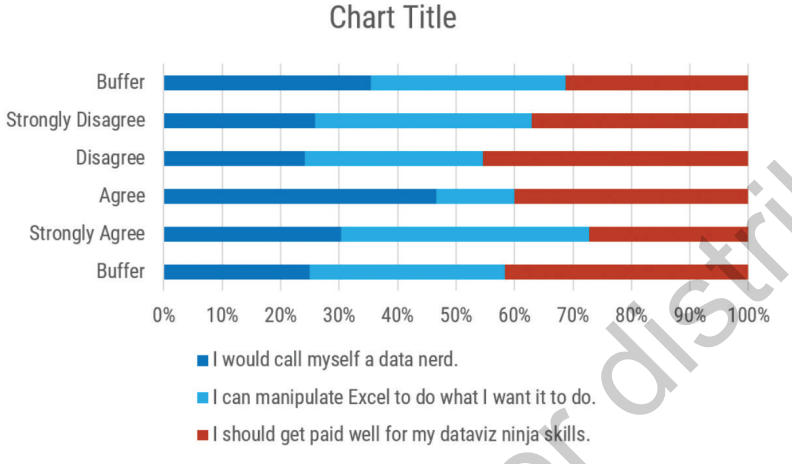
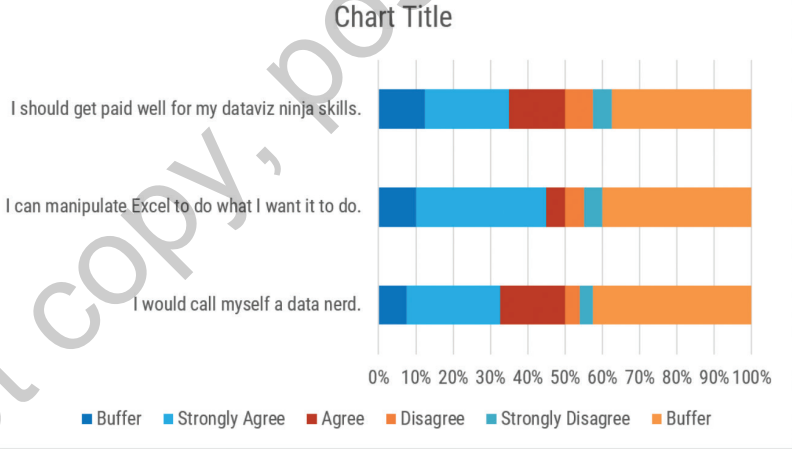


Figure 5.13 /// Each of the three stacked bars has six segments.



you want showing. I chose 65%. I then changed the axis minimum to 5% (or 0.05). Once the right scale is in place, click back on the axis and delete it.

I also added data labels and changed the colors so that the positive values were bluer and the negative values were redder, to reflect what I discussed about colors in the last section (Figure 5.15). Whatever data are on the right end of this graph are going to get more attention because they are surrounded by more white space, so

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

Figure 5.14 /// The Buffer bars essentially disappear—they are really there only to support the structure of the data visualization.

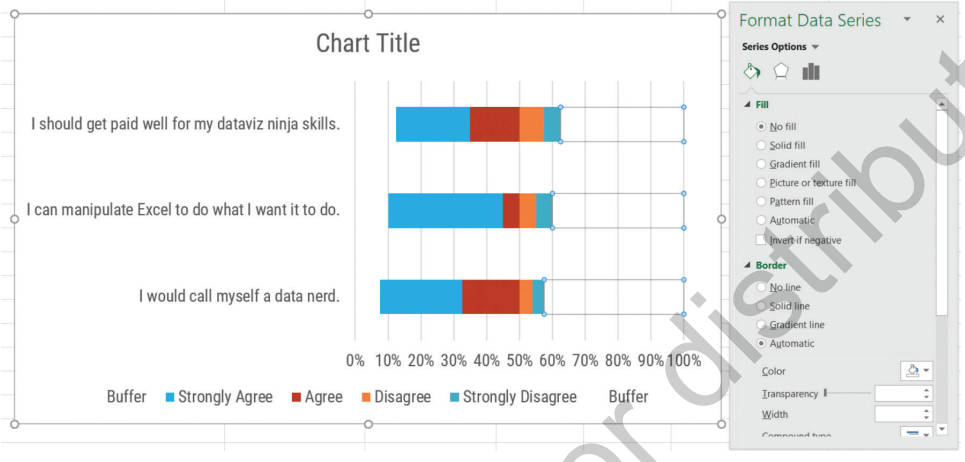
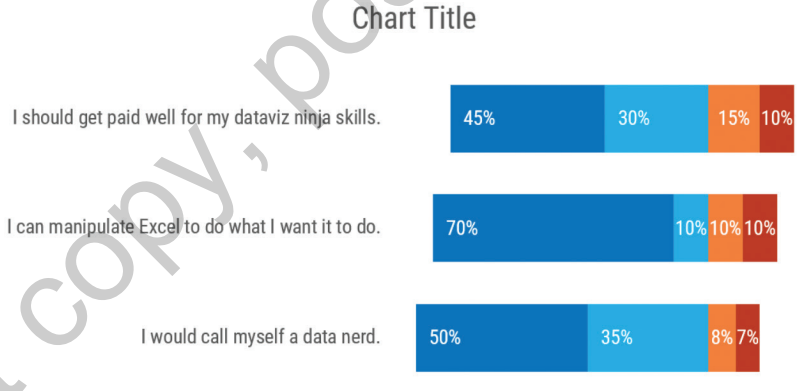


Figure 5.15 /// Use shades of two colors to help the audience mentally group each half of the diverging stacked bar.



that's why I have the Disagree categories on that side of the chart space even though this might go against the convention of how many of us were taught to write survey questions. That's also why my data have been sorted by the Disagree category. Want to hear something interesting? Even though academic convention has been to put negative on the left and positive on the right, could it just be that our professors were right-handed? Some research shows that though we tend to consider left = bad and right = good, left-handed people tend to think left = good and right = bad (Casasanto, 2009). Question everything!

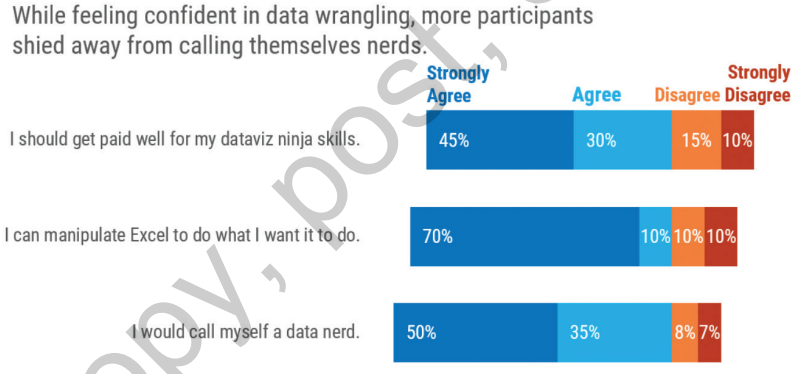
Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

Some people really want a line down the middle, to mark where the divergence occurs. To do that, click on the Insert tab. Look in the Illustrations group, and you should see an icon that looks like a group of shapes. Open it and click on the line. You've probably done this plenty of times in PowerPoint or Word. Then hold down the control key (this keeps the line perfectly straight) and drag to draw the line on top of the graph.

To finish it off, I just need to add back my legend (across the top, using linked text boxes) and give it a descriptive chart title as in Figure 5.16. It should be obvious why condensed fonts are handy in data visualization with this example. Cramming some of those labels into tiny segments or getting the Strongly Disagree text box to fit in that squished space of the graph is only possible because the letters are tall and skinny.

Pow! That's right! It takes just a little forethought and some basic math to better represent diverging survey data and circumvent some of the issues of a basic stacked bar.

Figure 5.16 /// The diverging stacked bar makes it easier to compare all positive responses to all negative responses.



What About Neutral?

Can I still use the diverging stacked bar chart if I have a neutral category (e.g., a five-point scale instead of a four-point scale)?

—Meredith

Yes, Meredith, you can! Go back to the table and insert two columns in the middle of the table. Now, divide your neutral value by two and type that number into each column such that half of neutral belongs with your positive values and half

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

with your negative. This keeps the neutral category aligned to the middle, sort of straddling the midline. It does get a little tricky to add data labels to each segment, though, because Excel wants to label each half. You'll want to insert a text box and type in your total neutral value. You may also decide that the line isn't as impactful or necessary. Personally, I don't think it has quite the same effect, but try it out and let me know your take on the neutral situation!

But here's the thing: Diverging stacked bars diverge around your preferential cut point. For example, if you cared only about the Strongly Agree category in this last example, you might choose to diverge between Agree and Strongly Agree (and flip the scale). So in terms of neutral, I encourage you to think about whether neutral is what you want. In most survey scenarios, I am guessing that neutral is not desired. So I would group it with the Disagree and Strongly Disagree side of this divergence. Neutral should sit in the middle on our surveys so that respondents see that as a midpoint option where its position actually conveys neutrality. But when we are making use of that data for decision making, neutral is no longer neutral.

Aggregated Stacked Bar

Probably the most common way of visualizing rating survey data is via a stacked bar chart, just like the one in Figure 5.17 created by the team at the Hole in the Wall Gang Camp. (Long sidenote: Have you ever heard of this place? It's a camp, started by Paul Newman [swoon], for children with severe illnesses, such as cancer, hemophilia, sickle cell, and other terrible things that should never happen to children. Just reading their latest data report made me cry. Thinking about their data report later made me cry again. Just a little something in my eye right now; must be an eyelash.)

Their basic stacked bar chart reports on camper appreciation, and the team already wisely ordered the bars from greatest to least on Strongly Agree and made the title into a takeaway sentence. However, when looking at Figure 5.17, you can probably see a few things that aren't really working. The labels in the skinny parts are scrunched. The x-axis is labeled in 5% increments, which is overkill. And just overall, it feels like a lot to mentally process. What do we do?

My strategy was to first find out what they really wanted to say about their data.

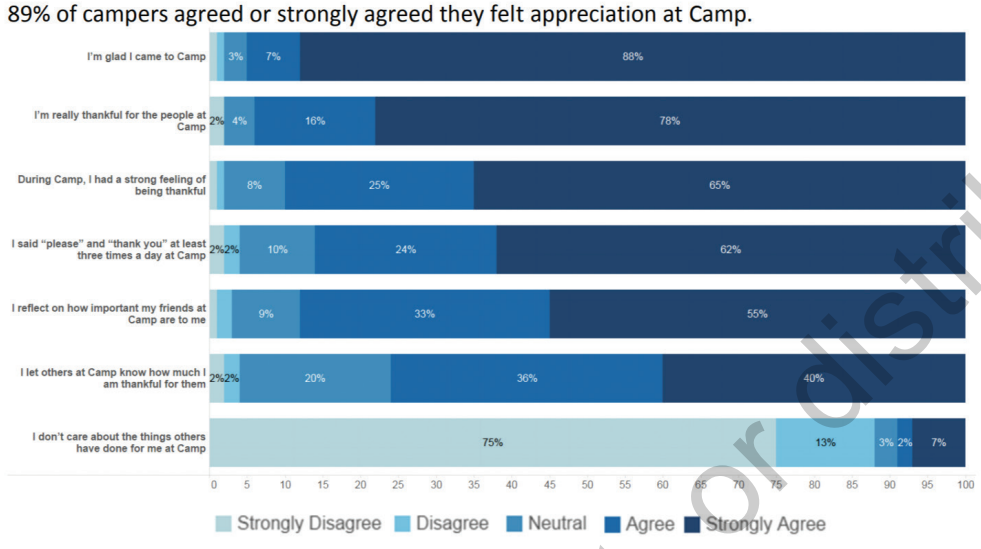
Ann Gillard, director of research and evaluation, said, "We were fortunate to see such strong positive outcomes and want to celebrate this as we share the story of Hole in the Wall with potential campers, families, funders, and friends."

Lucky for us, they neatly encapsulated it in their title. Notice how they are grouping Agreed and Strongly Agreed with their words? They are doing a bit of digestion for their readers by collapsing those categories into a number that's more meaningful.

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

Figure 5.17 /// The typical stacked bar gets a bit hard to read.



Source: Gillard, A. (2013). Camp survey [internal report]. Hole in the Wall Gang.

And Ann knew it. She said, "We have found that our readers like to see percentages of people rather than numbers of responses, and that readers want to know what people agree with rather than reading about averages. So, we moved away from reporting means in these charts and put the means into an appendix for the other data nerds to view." Ann is on the right track, plus she called them data nerds, so now she is my best friend.

I followed their lead and aggregated those categories in the graph by summing their values in the Excel spreadsheet. This produced the aggregated stacked bar seen in Figure 5.18.

You'll also notice I reversed their categories. Since they are emphasizing the positive responses in their messaging, I put those bar segments on the left, so they share a common baseline that's easier for the audience to read. I put number labels only on those positive responses. I also used their Hole in the Wall Gang brand color there and grayed out the rest to add further emphasis where they want it. Notice that I followed this pattern even for the last question, which was reverse-coded, where answering negatively is a good thing. I reduced the number of increments on the x-axis and color-coded the title a little more.

Excel Ninja
Level: 2

With those tweaks, the amazing things happening at Hole in the Wall Gang are even clearer to their readers (future campers, donors, etc.). Aggregated stacked bars are a fitting visualization choice when the level of detail in a Likert-type scale is unnecessary.

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

Figure 5.18 /// If the difference between Agree and Strongly Agree doesn't matter much to your readers, aggregate those categories in the graph.

Changing the Wording of the Survey

When you have one item that needs to be reverse-coded, would you consider rewording the question when you present the data? (I would probably have tried to word the question in a different way in the survey to avoid this, but there are certainly times where the question needs to be worded as it is).

—Sheila

Hi Sheila, I definitely wrestled with how to handle this reverse-scored question. One thing I might consider for the future is rewording this question in the chart/report with an asterisk explaining how this was originally reverse-worded and scored.

—Ann Gillard

I'll just chime in to say I do think researchers tend to be a little too loyal to the survey when it may not be ideal for the audience. I support changing the wording on the graph, carefully, and with a note, as Ann mentions. While we don't want to mislead anyone about the questions asked, no matter what, reverse-coded questions are a little confusing and there's no perfect way to handle their presentation, but I think slightly altering the wording in the graph probably does the least damage.

—Stephanie

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

I'm sure you have the hang of this by now, but it is also totally possible to collapse sequential categories into aggregate bars. Perhaps Poor and Fair combine on one side and Good and Excellent combine on the other. As long as it makes sense to your readers and you feel you are fairly representing the survey respondents, collapse away!

RANKING

Ranking questions are the kind of survey questions where you allow the respondents to select more than one answer. It could be that they can check all that apply. It could be that you are asking them to rate each answer option on its own scale. It could be that you're asking respondents to rank a given set of answers from Best to Worst. Whichever way, there are a couple of right ways and tons of wrong ways to visualize ranking data.

Column Graph

Let's say you want to poll your clients about their favorite taco toppings. You craft a survey with the question "Which taco toppings are your favorite? (Check all that apply)," and your survey lists 20 options plus an "Other" fill-in-the-blank option. Because respondents can pick more than one option, there's no way these responses will add up to 100%. No. Way. Take a look at Figure 5.19 for the table of data.

But often when people see a set of percentages that belong to one survey item, they are too quick to pull the pie chart trigger. Such errors of assumption produce visualizations that look like Figure 5.20.

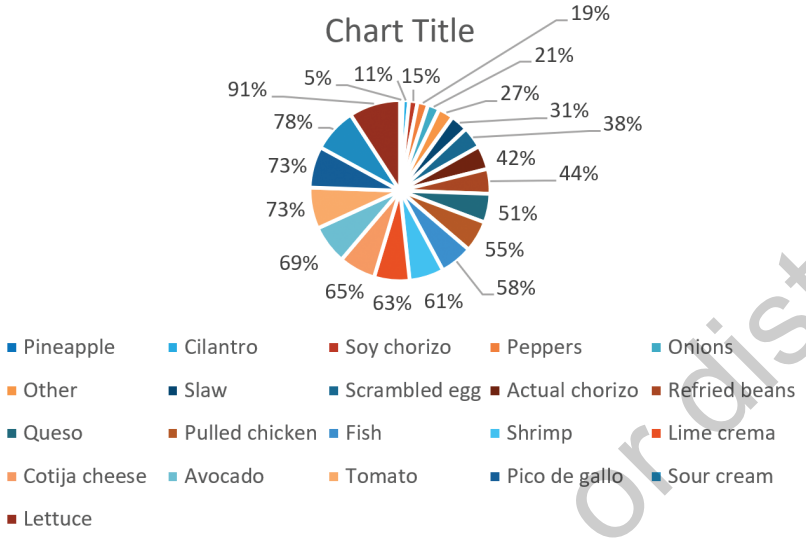
Figure 5.19 /// A quick sum of these values shows they add to 990%.

	A	B
1	Pineapple	5%
2	Cilantro	11%
3	Soy chorizo	15%
4	Peppers	19%
5	Onions	21%
6	Other	27%
7	Slaw	31%
8	Scrambled egg	38%
9	Actual chorizo	42%
10	Refried beans	44%
11	Queso	51%
12	Pulled chicken	55%
13	Fish	58%
14	Shrimp	61%
15	Lime crema	63%
16	Cotija cheese	65%
17	Avocado	69%
18	Tomato	73%
19	Pico de gallo	73%
20	Sour cream	78%
21	Lettuce	91%

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.20 /// A pie chart is not the right visualization for this data.



Excel shouldn't even allow pie charts to be made from check-all-that-apply data. It's all too easy to insert one and think the job is done. But this is how visualizations end up making the rounds on Twitter as jokes among the slightly savvy crowd or end up on **http://viz.wtf/**. The values do not add up to 100%, so a pie chart is the wrong visualization.

Excel Ninja
Level: 2

It makes much more sense to graph this data as a simple bar chart, and it makes much, *much* more sense to order those bars from greatest to least, as in Figure 5.21.

Remember back in Chapter 1 when we discussed how humans are pretty decent at comparing the lengths of things? Bar charts encode data by length, so it is less onerous for viewers to read a bar chart. But we can make it even easier by ordering the bars by greatest to least values (or least to greatest; I could hear arguments for either direction). This way, readers can quickly scan the graph and see what was most popular and what was least popular.

And on that note, it's entirely possible that people really care only about the most popular (or the least). This long list of possible areas of concern could very well be overkill. Sure, we need to provide a fairly exhaustive list of response options on the survey, but that doesn't mean we have to visualize all of those options for the reader. It may be more effective to limit the reporting to the top three or five concerns. Focusing the visual in this way can help readers hone in on what is most important, make the visualization less overwhelming, and more quickly compel action (see Figure 5.22).

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.21 /// A bar chart is a better fit for check-all-that-apply data.

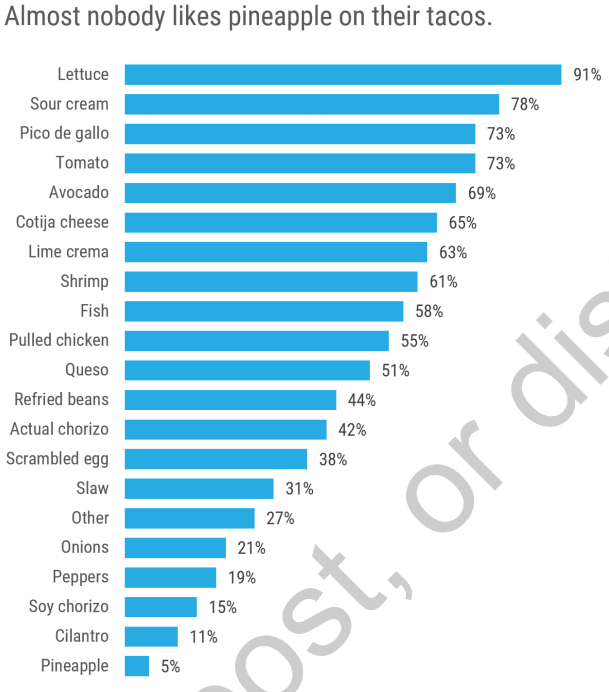
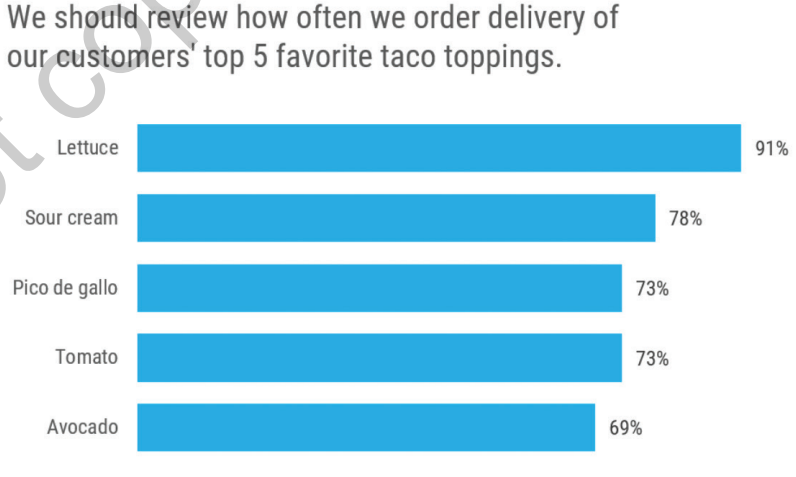


Figure 5.22 /// Zoom in on a handful of items that are the most frequently checked in a check-all-that-apply survey question.



Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

The Lollipop Variation

Another way to make the lengthy bar chart less visually overwhelming is to remove even more ink. We've already come a long way—we deleted tick marks, lightened gridlines, and removed an axis line. What else is left to remove? Well, what's left is all that ink making up each bar. I know it sounds ridiculous to think we could remove any of that ink because it is what encodes our data. But really all we care about on those bars is the very end, where it stops, since that's what tells us the value. Instead of an ink-heavy bar chart, we are going to make an ink-friendly lollipop graph. Some research (Blasio & Bisantz, 2002) shows that this move to ink-friendly design can impact readability of the data and lead to faster response times and greater accuracy (Gillan & Richman, 1994). Inbar, Tractinsky,

Figure 5.23 /// Add a column of data that will serve as your y-values—no one will ever see it.

	A	B	C
1	Topping	% of respondents	lollipop spacing
2	Pineapple	5%	1
3	Cilantro	11%	2
4	Soy chorizo	15%	3
5	Peppers	19%	4
6	Onions	21%	5
7	Other	27%	6
8	Slaw	31%	7
9	Scrambled egg	38%	8
10	Actual chorizo	42%	9
11	Refried beans	44%	10
12	Queso	51%	11
13	Pulled chicken	55%	12
14	Fish	58%	13
15	Shrimp	61%	14
16	Lime crema	63%	15
17	Cotija cheese	65%	16
18	Avocado	69%	17
19	Tomato	73%	18
20	Pico de gallo	73%	19
21	Sour cream	78%	20
22	Lettuce	91%	21

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

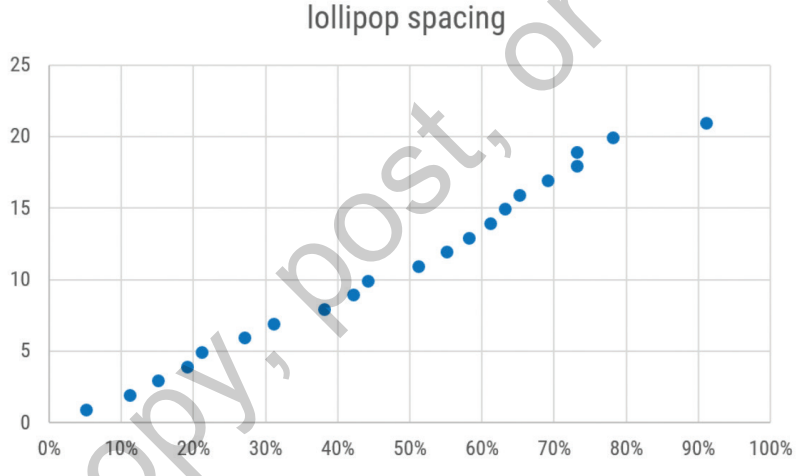
and Meyer (2007) showed that people did not rate a minimalist graph design, like the lollipop, to be as beautiful as a traditional bar, but they supposed this could be because lollipops are a new chart type for many to view and change can be hard.

The backbone of the lollipop graph is a scatterplot, which means we will need x-values and y-values. The x-values are easy—those are my survey responses. The y-values are going to be faked, just inserted to make each lollipop equidistant from the others. So Figure 5.23 shows the new data table, with my actual survey responses and a new column next door where I typed in placeholder data, from 1 to 21.

Excel Ninja
Level: 8

Now highlight both columns with numbers and insert a simple scatterplot (Figure 5.24). Don't highlight the data labels or the header row, or else the graph won't work out properly.

Figure 5.24 /// Your graph should be a series of dots.



The scatterplot process created dots—these are the lollipop heads! It is basically visualizing just the ends of the bars in the bar graph. Right now, I suggest you click on the horizontal gridlines and delete them. They can confuse things if we keep them in place any longer.

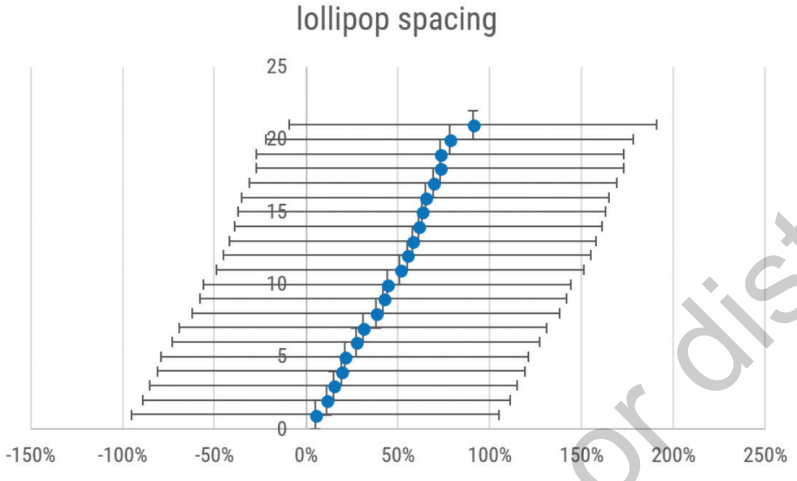
It's time to make the stick of the lollipop. We are going to create these by hacking error bars. Navigate up to the Chart Tools > Design tab and look for the button that says Add Chart Element. Click the drop-down arrow there, hover down to Error Bars, hover on its arrow to open another menu, and finally click on More Error Bars Options. In older versions of Excel, you should just see a button in the Chart Tools > Layout ribbon for Error Bars.

As soon as I clicked on More Error Bars Options, my graph got complicated (see Figure 5.25).

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

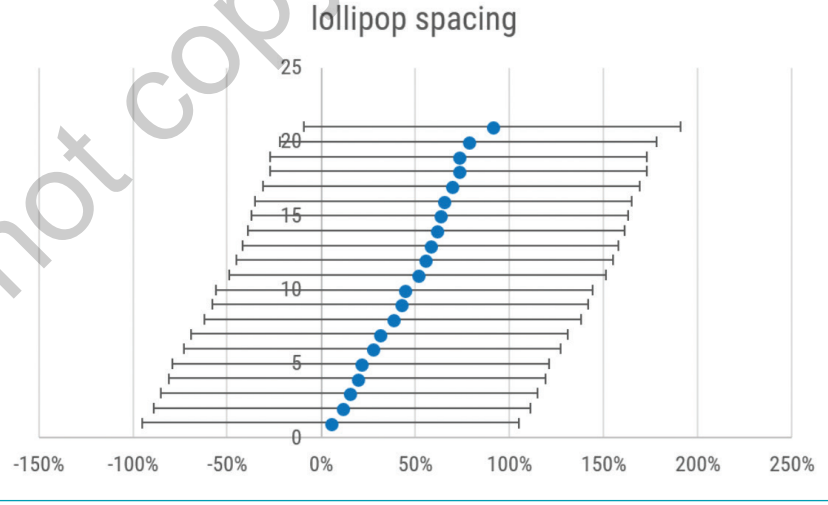
Figure 5.25 /// Excel added error bars in all four directions around each dot.



We do not need the vertical error bars at all. Just click on them and hit the Delete key on your keyboard. Your graph should look like Figure 5.26.

Now right-click on the horizontal error bars and select Format Error Bars.

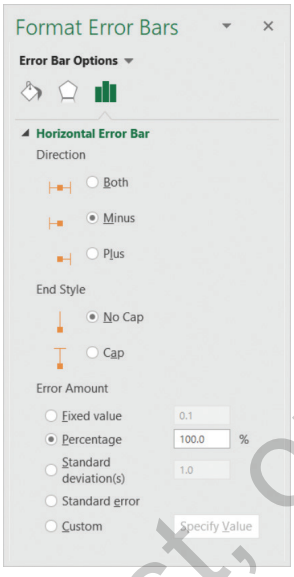
Figure 5.26 /// Delete the vertical error bars so only the horizontal ones show.



Do not copy, post, or distribute

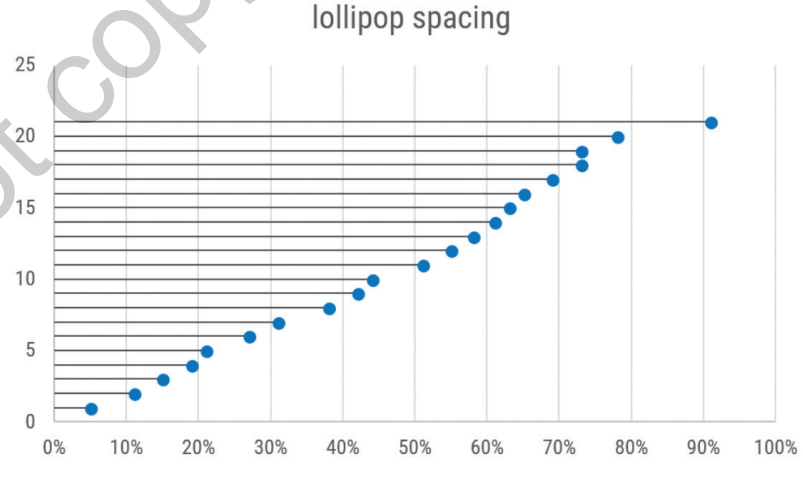
Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.27 /// Horizontal error bars can be adjusted to meet the axis line and look like sticks for your lollipops.



Here is where the magic happens. You'll want to adjust every bit of this dialogue box. To make our lollipops, we really only want the error bars that shoot out to the left, so in the Direction area, switch from Both to Minus (Figure 5.27).

Figure 5.28 /// The essential pieces of the lollipop graph are in place.



Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

In End Style, get rid of the cap. In Error Amount, click the radio button by Percentage and type in 100% (Figure 5.27). This will fix your x-axis so it starts at 0%, just as it should, and will extend the lollipop stick from the lollipop head to the y-axis (Figure 5.28).

Give yourself a fist pump, because that was awesome.

Next up, let's fix that y-axis, shall we? Right-click on it and adjust the maximum to 22—that is one more than the highest number we listed in our fake data. If, in doing so, Excel then adjusts your minimum, just change it back to zero. Now you can delete the y-axis labels altogether.

Let's get the data labels in place now, so we know which lollipop represents what. We are going to insert another series of fake data so that we can use its labels. Since we are still working with a scatterplot, the y-values for the data are already here—we will use the same y-values in the "lollipop spacing" column. But the x-values that we are adding will be new. So add another column of data, and this time just fill it up with zeros. This way, the new series of data will be a set of dots all the way over on

Figure 5.29 /// Add another column of data, full of zeros, to act as label placeholders.

	A	B	C	D
1	Topping	% of respondents	lollipop spacing	labels
2	Pineapple	5%	1	0
3	Cilantro	11%	2	0
4	Soy chorizo	15%	3	0
5	Peppers	19%	4	0
6	Onions	21%	5	0
7	Other	27%	6	0
8	Slaw	31%	7	0
9	Scrambled egg	38%	8	0
10	Actual chorizo	42%	9	0
11	Refried beans	44%	10	0
12	Queso	51%	11	0
13	Pulled chicken	55%	12	0
14	Fish	58%	13	0
15	Shrimp	61%	14	0
16	Lime crema	63%	15	0
17	Cotija cheese	65%	16	0
18	Avocado	69%	17	0
19	Tomato	73%	18	0
20	Pico de gallo	73%	19	0
21	Sour cream	78%	20	0
22	Lettuce	91%	21	0

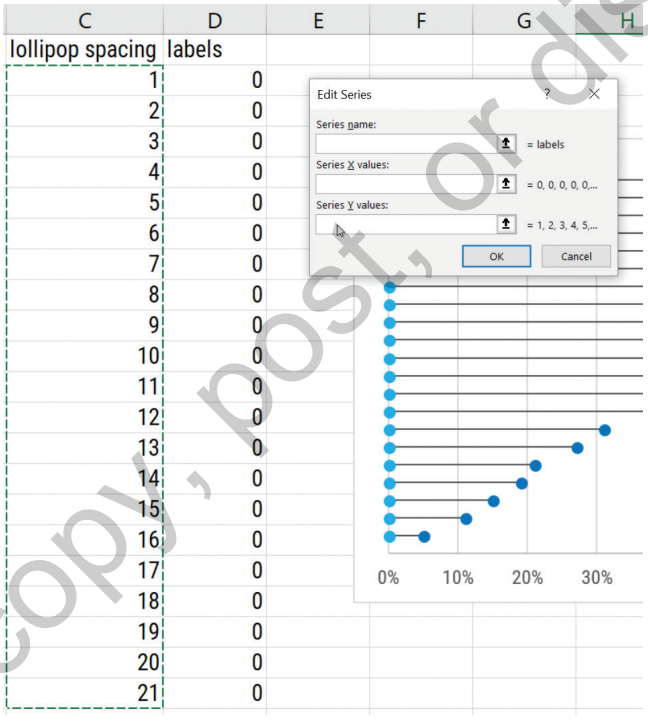
Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

the left side of the graph and the labels we eventually put there will look like regular, standard, proper labels for the lollipops. The table will now look like Figure 5.29.

Right-click on the graph and click Select Data. In the dialogue box that opens, click on Add in the Legend Entries section. Use the cell picker icons in the new box to select labels as your Series Name, all those zeros as the Series X Values, and all the numbers underneath "lollipop spacing" as your Series Y Values, as shown in Figure 5.30.

Figure 5.30 /// Add another series of data, composed of the data for labels and lollipop spacing.

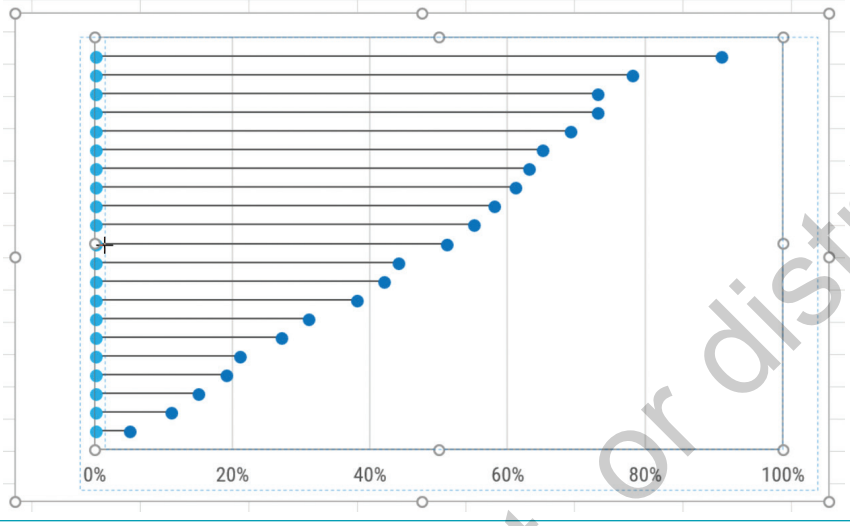


Click OK, and your graph now has a second set of dots going up the left side of the graph. Sweet! We are going to add data labels to the dots. But there's no room for the data labels yet. So, click inside the graph, on the white background (Excel calls this the plot area). Its border will become active, and you should see a little marker, midway down on the left, in the midst of your new series of dots. Click on that baby and drag it over to the right so that you are shrinking the plot area and making room for your labels (Figure 5.31).

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

Figure 5.31 /// Drag the side handle over to the right so you have space for the labels.



Now that there's room to breathe, right-click on the new set of dots and select Add Data Labels. Excel is going to give you the y-axis values, which are a secret. But click on each one and link it to the cell with the label you desire using the magic linked text box trick we discussed earlier in this book. Be sure you select the Left label position. You may need to readjust your plot area here or enlarge your entire graph to make room for all those labels (I did both).

Labels are in place; lollipop and stick are ready to go. It's time to get rid of the second set of dots that our labels are attached to. Right-click on those dots and select Format Data Series. In the Marker Options window, select None to eliminate the marker dots altogether.

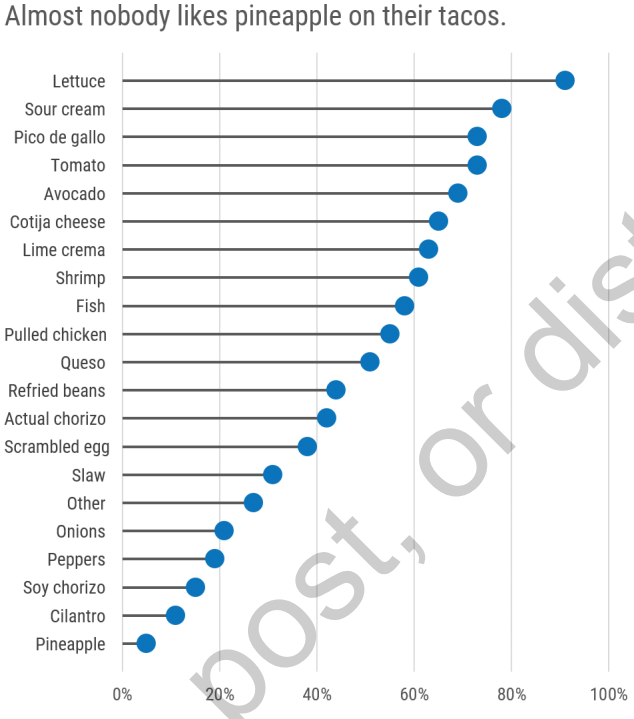
Add an awesome title and make any color adjustments. I ultimately decided to make my lollipop heads a little larger, my lollipop sticks a little thicker, and I removed my x-axis line, all of which can be accessed by right-clicking on that chart element and selecting Format.

Go ahead and compare this lollipop graph (Figure 5.32) to the original bar graph of the same data a few pages back (Figure 5.21). Less visually demanding, isn't it? Even though it took us several more steps to make the lollipop chart, we put in that elbow grease to make life easier for our readers. Plus you learned about the secret power of error bars in this process!

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.32 /// The lollipop graph is less visually overwhelming than a traditional bar graph.



Lollipops can also be vertical, and thankfully that doesn't take as much work. We are talking about Excel ninja level 3 here. You'll simply highlight your actual data values and insert a line graph with markers (instead of a scatterplot—so also no need for extra columns of fake data [I hear you cheering]). Then delete the line. For the lollipop sticks, you'll just pop in Drop Lines. Look in the same tab where you added your Error Bars, open the menu for Lines, and select Drop Lines. Dang, that's easy! I usually like to increase the size of the markers to something really big, like 20, and position my data labels smack in the center (see Figure 5.33).

Excel Ninja Level: 3

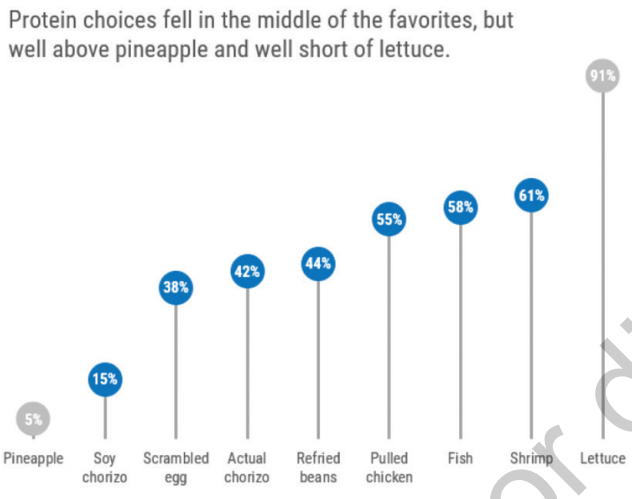
Yummy enough to eat, I know! Just be thoughtful here, because the size of the marker technically spans a range of values, not just the one in its label. This kind of lollipop is a little less visually accurate but has all kinds of appeal for its simplicity.

Lollipop charts are decent alternatives to standard bar and column charts anytime, for any kind of data that would work in a bar or column, not just for check-all-that-apply data. They are especially helpful when the visualization is overpowering due to its massive ink, because the lollipop focuses attention at

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.33 /// Lollipop graphs can replace column graphs, too.



the value. My U.S.-based clients who have data on all 50 states plus the U.S.-colonized territories tend to favor lollipops because any bar graph of all that data is going to be intense.

Large Number With Icon

I saw this next example in a magazine (Figure 5.34) and thought of you.

Figure 5.34 /// Try combining a single large number with an icon that represents that category.

Source: © George, P. (2014, March). How to save money on triple play cable services. *Consumer Reports.* Used with permission.

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

We don't have the survey question *Consumer Reports* asked, but it was probably something like, "When you tried to negotiate a better bundle package, which of these deals did you receive? (Check all that apply)." The fact that the respondents could check multiple answers is a clue that the five answers we can see here add to more than 100%. Rather than try any sort of graph, *Consumer Reports* chose to display the five answers using large numbers paired with a simple icon.

The most tedious part of developing a graphic like this is simply finding the icons that match your underlying category. Most of us have probably spent hours searching through websites trying to find suitable images. Add to that task the challenge of finding a set of images that all look like they belong together! Let me rescue your afternoon with two very helpful websites:

> The Noun Project—accessed at **https://thenounproject.com/**—was started by folks who dreamed of having an icon for everything. It's a super-exhaustive site with a powerhouse search system that makes finding icons an easy task. The icons generally have a similar look and feel to them, too—meaning they mostly look like they were drawn by the same hand (or hand on a mouse). The icons come in black and white for easy recoloring. Many are free, or at most they cost US$1.
>
> Flat Icon—accessed at **http://www.flaticon.com/**. The awesomeness of Flat Icon can be summed up in two words: Free. Vector. Vector means that you can resize the image as large as you'd like and it won't blur or fuzz on you. The icon you create for your graphic in your report can also appear on your slides in a much larger and more visible form and it will still look crisp and clean. The icons on this site are no charge, but many do require attribution in that you need to give the icon designer credit somewhere in your report or on your site, wherever the icon appears.

Alternatively, you could simply make your own icons. It wouldn't be all that difficult or time-consuming to construct the icon for Nothing (though maybe think a little harder to come up with something more effective than a circle with a line through Thing). To construct an entire graphic like this one, use any kind of blank canvas software—even PowerPoint would do the trick. You just insert text boxes and icons and be happy.

I wouldn't want to see a report full of the large-number-with-icon visuals, but it can work well in limited doses for check-all-that-apply data.

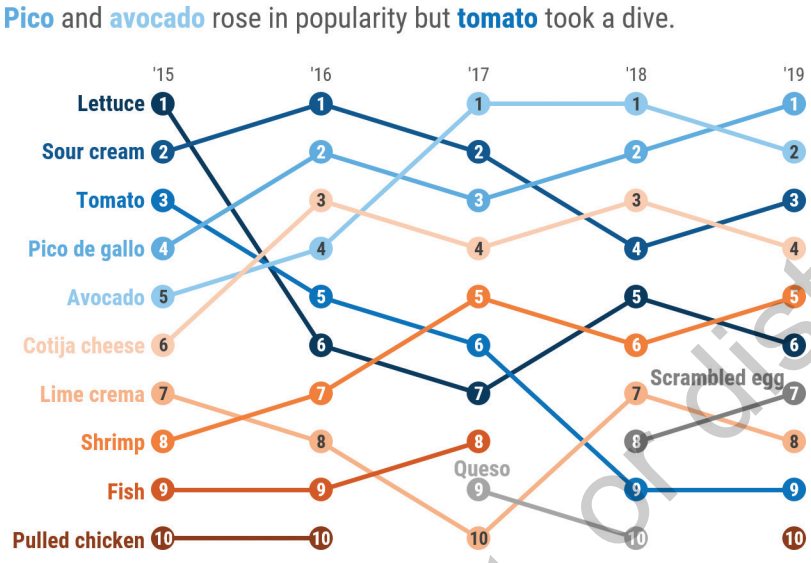
Bump Chart

So far, these ranking visualization choices have shown data for a single point in time, as if we only gave the survey once. More likely, we are surveying our respondents on the same questions over time to establish longitudinal trends. Rank over time is best shown as a bump chart. In this visual, we prioritize the rank position over showing the exact percentage of respondents who favored this particular taco topping.

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.35 /// A bump chart shows change in rank over time.



Excel Ninja
Level: 4

The table setup lists the rank position for each year, leaving blank any that fell out of the top 10 favorite toppings (Figure 5.36).

Figure 5.36 /// Ditch the exact percentages and just list the rank order each year.

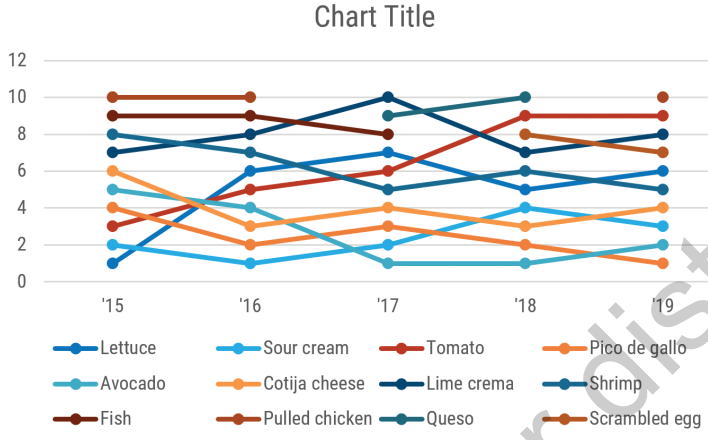
	A	B	C	D	E	F
1	Topping	'15	'16	'17	'18	'19
2	Lettuce	1	6	7	5	6
3	Sour cream	2	1	2	4	3
4	Tomato	3	5	6	9	9
5	Pico de gallo	4	2	3	2	1
6	Avocado	5	4	1	1	2
7	Cotija cheese	6	3	4	3	4
8	Lime crema	7	8	10	7	8
9	Shrimp	8	7	5	6	5
10	Fish	9	9	8		
11	Pulled chicken	10	10			10
12	Queso			9	10	
13	Scrambled egg				8	7

Highlight all the data and insert a line with markers graph (Figure 5.37). It won't look right at first, so click the Switch Row/Column button.

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

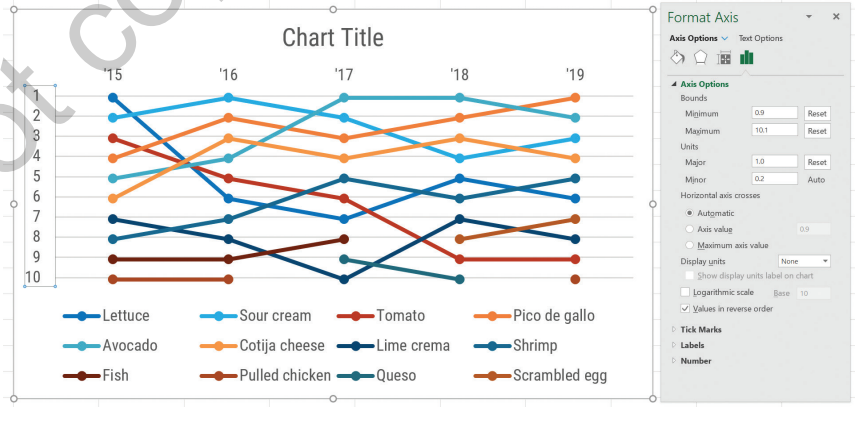
Do not copy, post, or distribute

Figure 5.37 /// A bump chart is just a modified line graph with markers.



Excel typically graphs the data such that the y-axis starts with the lowest value at the bottom. But in a bump chart, number one is the best and should be at the top. So let's flip the y-axis. Just right-click on it and select Format Axis. Then click the radio button that says Values in Reverse Order. While here, let's adjust the y-axis scale. We really only need it to run from 1 to 10. But if you put 1 in the Minimum box and 10 in the Maximum box, you'll see that the lines in the graph got a little cut off. So I add a little extra space in the axis scale and run it from 0.9 to 10.1 (Figure 5.38).

Figure 5.38 /// Adjust the y-axis so it is inclusive of your data and the values are in reverse order.



Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Once the y-axis is set, delete it and its gridlines. The x-axis is now at the top of the plot area, so delete its line, too.

The visual that's left can still be pretty hard to read, since the lines are so tangled. The best way to handle that is through careful formatting. Shades of one color—a sequential coloring scheme—would be best to communicate sequential data, like ranks. But finding 10 discernible shades of one color is impossible. In fact, ColorBrewer turns up no options for us. So I chose to work with shades of three colors—blues for the top five, oranges for the bottom five, and grays for the newcomers in the top 10 list. Then I inserted text boxes for each of the ingredient names, using the magic linked strategy we have applied over and over again. Each label's proximity to its corresponding data and its color match is important to helping the reader make connections in this dense graph. Some people might also choose to label the right side of the graph as well. Finally, I added data labels to the center of each marker so that the reader can trace the popularity of a single topping over time. Formatting is key in making this graph as readable as possible.

Try a bump chart any time you think your story is about the top 5 or 10 or *x* over time. It's just a line chart with markers and a reverse axis and some patient coloring and labeling. When you are done, you can walk down the hall, giving out bump chart fist bumps.

BRANCHING

Branching questions are the type of question where if a respondent answers one way, then they are given an additional set of questions that branch out from the first answer. Anytime you hear someone describing the survey responses by saying something like, " . . . and of those, 45% said . . . ," then he or she is talking about subsets of respondents who were given particular questions based on their answers to previous questions. For example, when the survey says, "If you said *Yes*, proceed to question 32," the path through the survey branched. Sure, you could just graph those answers with one of the other options presented in this chapter. But it isn't likely that your readers will quickly grasp that you have isolated a subset of respondents for examination. Let's look at a few other visualizations that make it clearer.

Annotated Graph

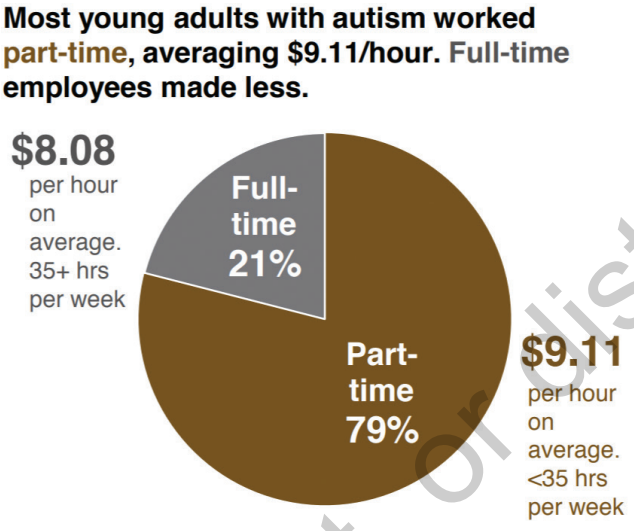
If your branches don't extend very far or the data you need to report are limited, consider the simplest charting method: add an annotation inside another graph. The note provides just a little extra explanation—the data from your branch.

The data in Figure 5.39 came from a survey where young adults with autism were asked whether they worked part-time or full-time. The survey branched from there, probing about hourly pay, benefits, and promotions, depending on which side of the branch the respondent followed. Both groups were asked about hourly pay at some point, and it was this extra data that we wanted to display for both groups.

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

Figure 5.39 /// Embed extra data from down the branch into the original graph to give it more impact.



Source: Roux, A. M., Shattuck, P. T., Rast, J. E., Rava, J. A., & Anderson, K. A. (2015). *National autism indicators report: Transition into young adulthood.* Philadelphia, PA: Drexel University, A. J. Drexel Autism Institute, Life Course Outcomes Research Program. Used with permission.

The annotated graph strategy can be applied to any kind of graph type, but a pie chart works here because we have only two slices. Each slice has an annotation, detailing the average hourly pay for full-time and part-time workers, including a note that defines full- and part-time hours. The annotations were inserted with text boxes—and several of them. I inserted one text box for the hourly dollar amount and linked it to the cell in the table, so that when those numbers change with next year's data, the text box will update. I inserted a second text box below each hourly rate and typed in the text "per hour on average" and the definition.

Excel Ninja Level: 2

Clearly, there's only so much extra text we can pack into one data visualization, so the annotated graph strategy for branching data can handle limited depth down the branch.

Nested Area Graph

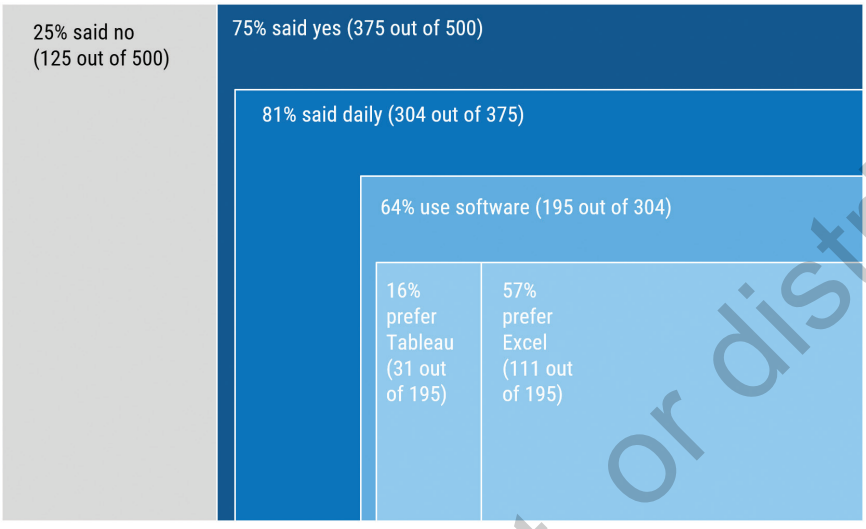
Figure 5.40, my friends, is a nested area graph.

See how it works? In this case, 25% of the survey respondents said no (and they skipped to the next section). Seventy-five percent of the respondents said yes, they

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.40 /// The inner rectangles represent the subsets of respondents.



visualize data. The people who do visualize data were of most interest in this study, so they were given an extra set of questions. Of that 75% who visualize data, 81% reported "daily." This software company is only curious about primary users, those who visualize daily, so those folks saw even more questions pop up in their survey window. And of those who visualize data daily, 64% said they use software. You get it? It's a pretty intuitive visualization. Nested area graphs can handle more levels of branching than the pie chart plus large number example.

This is graphing by area—which is tricky. Remember how Chapter 1 pointed out that humans are bad at interpreting area? Turns out the difficulty with judging area isn't any different if the area is in a circle (like a bubble graph) or a square (um . . . that's a nested area graph; Heer & Bostock, 2010). Therefore, the rule is that the area must be proportionate to the data it's representing—a reader should be able to put down a ruler and calculate the area of the Daily square, and it had better be 81% of the area of the Yes square. And it is.

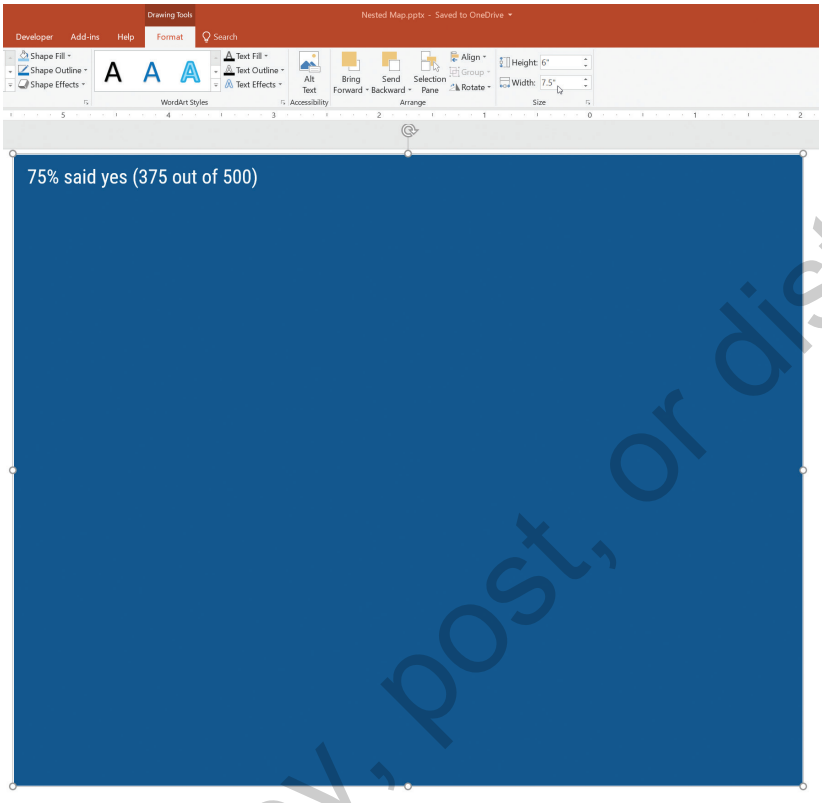
Excel Ninja Level: 3 (the math is the hardest part)

I made this diagram in PowerPoint. I started by calculating the total rectangle area by just throwing out some dimensions that would be easy to work with. Area is height times width (oh, middle school math class). I said 6 inches by 10 inches, or 60 square inches, would be the area to represent all 500 respondents. The dark-blue Yes rectangle had to be 75% of that size, or .75 multiplied by 60 square inches, which is 45 square inches. I wanted the Yes rectangle to be the same height as the overall rectangle, which was 6 inches. So, I divided 45 square inches by the 6-inch height and got 7.5 inches for my width. So far I've just been using a calculator. Now it's time to actually visualize! Insert a rectangle in PowerPoint (see Figure 5.41).

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Figure 5.41 /// Over in the right of the Drawing Tools section, you'll see where to enter the height and width dimensions you calculated.



Look in the orange Drawing Tools section, inside the Format tab, for the Size area where you can punch in those dimensions you just calculated.

To get to the blue daily rectangle, I calculated 81% of the size of the Yes rectangle, which you'll recall was 45 square inches. The product is 36.45 square inches. I wanted it to nest inside the Yes rectangle, so the height needed to be smaller than 6 inches. I chose to lay this out with 5 inches in height, which leaves a width of 7.29 inches. So, I inserted another rectangle and sized it up properly. And so on.

Sounds complicated, but it really isn't too hard. You just pick an area to start with for your total set of respondents and divide it up in proportion to the data it needs to represent.

What makes the nested area graph seen in Figure 5.40 work best: listing the *n*s alongside the main heading in each rectangle. This will help people who are

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

having a hard time wrapping their brains around the visualization because they can see how the *n*s carry down into each section, such that the numerator for the parent rectangle becomes the denominator for the child. Main headings can be inserted with text boxes.

What doesn't work: using PowerPoint's Scale Height and Scale Width features. You might have started exploring PowerPoint and noticed that if you right-click on the rectangle, you see a menu option called Size and Position. If you open it, you'll see a place to enter height and width, just as we've been doing. You'll also see spots that say Scale Height and Scale Width, where PowerPoint offers the ability to scale the rectangle by a specific percentage. You may be tempted to shortcut the above process and simply copy one rectangle and type 81% into the Scale Height and Scale Width boxes. At first, it may even appear to work, but once we get down to the inner rectangles of the nested area graph, it becomes clear that the scale option is a little wonky. Try it if you don't believe me. The surefire way to make a proportionally representative nested area graph is to break out the calculator and crunch those numbers by hand.

VISUALIZING NOT APPLICABLE OR MISSING DATA

Yes, I know the jig is up. All my examples in this chapter so far have been pretty tidy, as if every response option were addressed by every single respondent. The truth is that life and data collection are messy. How can we show that different questions have different sample sizes? The most appropriate visualization method will depend on the severity and inconsistency of your problem.

Note Small Consistent Missing Data

Excel Ninja Level: 0

The easiest solution when data are acting like tiny but equal-opportunity absentees for every response option is to just make a note of it somewhere in the graph. My preferred location is in a subheading, underneath the main takeaway point of the visualization. Make it smaller (in a report, like 9-point font) and gray (see Figure 5.42).

If the data are missing consistently, it is good to note, but this isn't a super-critical issue. This treatment marks the issue but relegates its importance to a background matter.

Add Sample Size for Large Consistent Missing Data

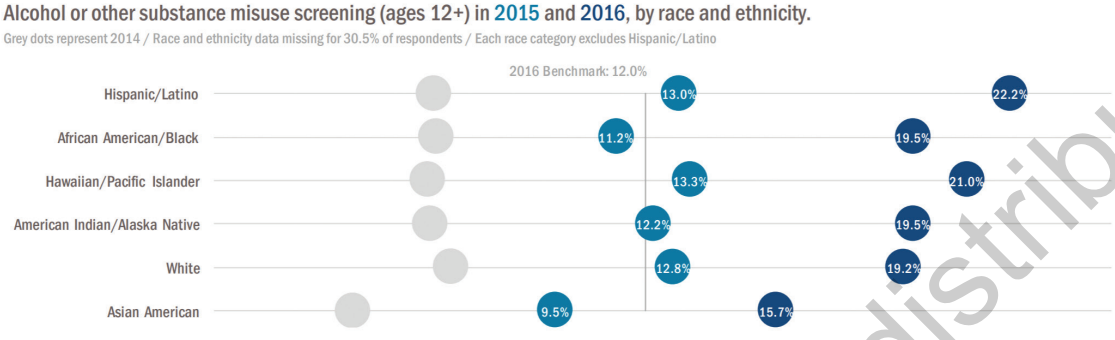
Excel Ninja Level: 1

Here's everyone's favorite nightmare: You sent your reliable research assistants out to collect data with some paper and pencil surveys . . . only one page of the survey was missing. It'll be okay because you have more data, but what to do with those questions that have many fewer responses? The answer: Note the sample size in the data label.

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

Figure 5.42 /// Add a subtitle to the graph to explain the missing or removed data, if it is consistent across all categories.



Source: Malone, M. (2017). SBIRT graph. Oregon Health Authority Office of Health Analytics. Used with permission.

To do this in Excel, you can add the sample size to each data label cell in your table. Make a line break within a cell (on a PC) by holding down the Alt key while you hit Enter.

In the formula bar, the data appear broken onto two lines (Figure 5.43).

In the graph, the "missing" data show up as part of your data label, under its corresponding question. In this case, the prenatal question was asked on a different page, but it was grouped with other healthcare data from the missing page for reporting purposes. The data label makes it clear that the sizes are inconsistent among the questions in this graph (Figure 5.44).

The same strategy can be used for large amounts of Not Applicable data. You can delete it from the graph, but the remaining data labels need to note the sample size. Again, the strategy here is to note the absence in a clear but diminished way.

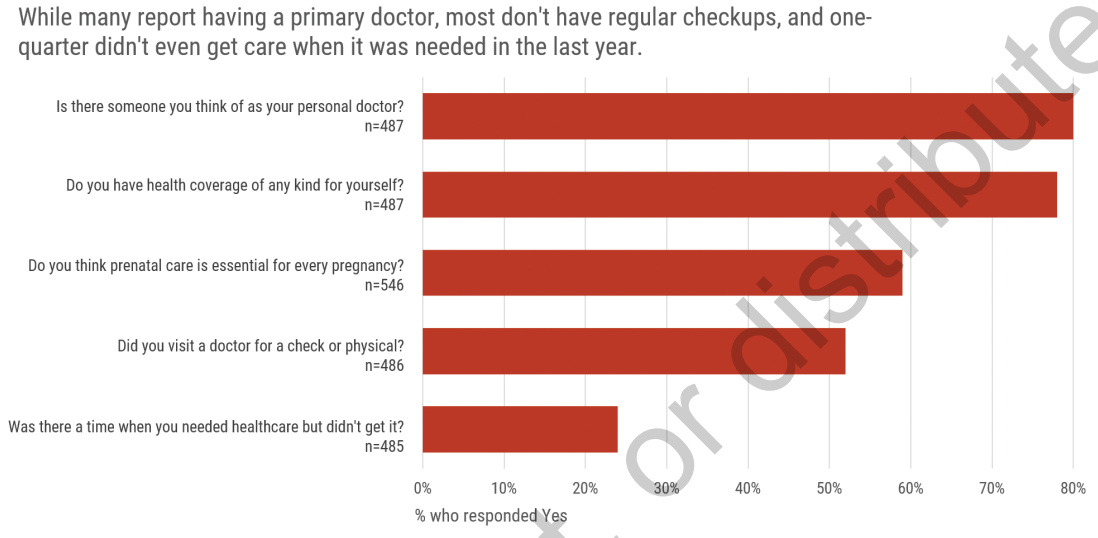
Figure 5.43 /// Type Alt + Enter at the end of a line to create a line break within a cell.

Do you have health coverage of any kind for yourself?
n=487

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

Do not copy, post, or distribute

Figure 5.44 /// Add the sample size for each question when there is significant missing data.



Add a Graph on the Side for Large Inconsistent Missing Data

The messiest data to deal with are those cases when lots of respondents skipped questions that weren't applicable to them. The most accurate way to handle such data is to be super honest that it is gone and show it in the graph—just off to the side.

Excel Ninja Level: 9

To do this in Excel, we are going to combine two graph methods we already discussed in this chapter: diverging stacked bars and small multiples.

The left portion of this table should remind you of the way we set up the diverging stacked bar table. We add secret hush-hush buffer zones to either end of the Likert scale so that both halves add to 100%. Not Applicable answers are way over to the right, followed by a secret buffer column that pairs up with the Not Applicable data to also add to 100%. Adding across each row here should total 300% (see Figure 5.45). I know it goes against your survey methodologist instincts, but see this through.

Graph everything—the whole table, including the headings—as a 100% stacked bar graph. As with the other secret buffer graphs, you'll mark the secret buffer bars as No Fill, and you should have three of those buffer values represented in each

Do not copy, post, or distribute

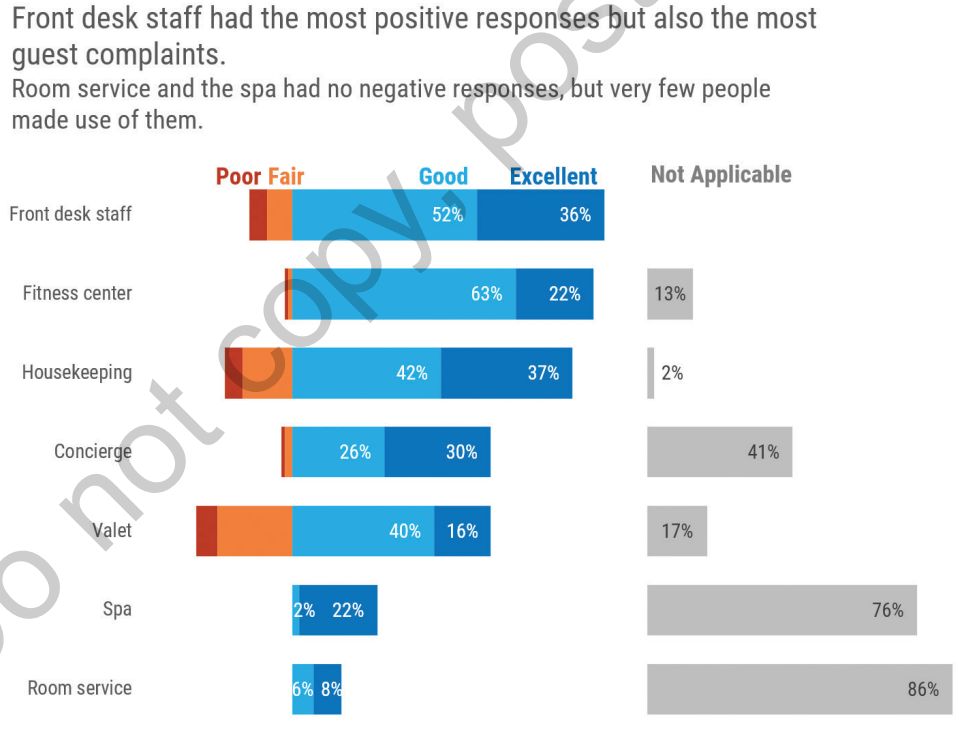
Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.

stack. You have now created a diverging stacked bar graph illustrating your main responses and a separate graph to the right with your Not Applicable responses (see Figure 5.46).

Figure 5.45 /// Add secret buffer columns for the negative responses, the positive responses, and the Not Applicable responses.

Please rate your overall experience with:								
	Bad Buffer	Poor	Fair	Good	Excellent	Good Buffer	Not Applicable	N/A Buffer
Room service	100%	0%	0%	6%	8%	86%	86%	14%
Spa	100%	0%	0%	2%	22%	76%	76%	24%
Valet	73%	6%	21%	40%	16%	44%	17%	83%
Concierge	97%	1%	2%	26%	30%	44%	41%	59%
Housekeeping	81%	5%	14%	42%	37%	21%	2%	98%
Fitness center	98%	1%	1%	63%	22%	15%	13%	87%
Front desk staff	88%	5%	7%	52%	36%	12%	0%	100%

Figure 5.46 /// Turn the buffer segments No Fill and you are almost there.



Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

If you have forgotten any of the steps, review the sections that discuss diverging stacked bar and small multiples graphs. But really it's just formatting from here, folks. I used strong color-coding and inserted linked text boxes for labels at the top of each response option.

The title here is important because it needs to explain why both graphs are included and needed for proper interpretation. It's a complicated concept. If it becomes difficult to encapsulate both graphs in a single title, it's probably a good sign that you are trying to convey too much information and that perhaps the Not Applicable data should be dropped and mentioned in another way. If you do drop the Not Applicable data, then go back to the previous section and be sure to add the sample size to each label.

The overarching point of this section is that you should treat the missing or Not Applicable data differently from the main data you collected. Your primary data need to be seen in their own light.

/// EXERCISES

Head over to **ColorBrewer2.com** and snag a color scheme for your favorite rating data. Look for one that is close to your organizational color scheme, or one that makes semantic sense to you (as in reds for Strongly Disagree). If you don't see the RGB color codes right away, look for a drop-down menu that may be set to HEX (those are web codes). Select RGB to get the codes you need for Excel. The numbers are listed in order so that the first one is for R, the second for G, and the third for B.

Look up wtfviz at **http://viz.wtf/** and have a hearty chuckle at the illogical data visualizations that actually made it to print and screen. Then pick one that you laugh at the hardest and remake it so that it's better. You will likely need to rethink the graph type and review the formatting in terms of color and legend. Sketch it first on paper and then try to make your redesign on a computer.

Network diagrams stem from specific types of surveys, where you essentially research the strength of the networks within a group of people. Try this: Make a list of all the students in your class and distribute that list to everyone in class, asking them to mark the students with whom they've had at least two previous classes together. Collect all the data by tallying responses. This data would be visualized in a network diagram.

Analyzing these connections takes a specialized software program: NodeXL is a handy one because it is free, acts as a plug-in for Excel, and allows plenty of visual customization. It visualizes the data by assigning each person in the survey list as a dot (or node) and draws connections between the people in the network via lines, where the thickness of the line represents more connections. The network diagram also groups together those who report hanging together.

Use pencil and paper or try out NodeXL (can be accessed at **https://nodexl.codeplex.com/**) or another software program and visualize your class survey data. Even with their potential to get a bit complex, network diagrams are becoming more popular as a concise way to show many interconnections with specific check-all-that-apply data.

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Do not copy, post, or distribute

/// RESOURCES

Visualizing social network diagrams can be clunky using some traditional analysis programs. The most elegant network diagrams are often made through specialized software and perhaps some programming skills. For a great overview of the software available for network diagramming, check out Andy Kirk's collection of resources, which can be retrieved from **http://www.visualisingdata.com/resources/**.

The nested area graph example first came in front of my eyes via Innovation Network's "State of Evaluation" report. It's an amazing, visual forward report, and even if you have no interest in its subject matter, you'll want to click through it to stare in awe at the thoughtful (and proportionate) data visualizations. Go to **https://www.innonet.org/media/2016-State_of_Evaluation.pdf**.

/// REFERENCES

Blasio, A. J., & Bisantz, A. M. (2002). A comparison of the effects of data–ink ratio on performance with dynamic displays in a monitoring task. *International Journal of Industrial Ergonomics*, *30*(2), 89–101.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351–367.

Gillan, D. J., & Richman, E. H. (1994). Minimalism and the syntax of graphs. *Human Factors*, *36*(4), 619–644.

Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 203–212). New York, NY: ACM.

Heiberger, R. M., & Robbins, N. B. (2014). Design of diverging stacked bar charts for Likert scales and other applications. *Journal of Statistical Software*, *57*, 1–32.

Inbar, O., Tractinsky, N., & Meyer, J. (2007, August). Minimalism in information visualization: Attitudes towards maximizing the data–ink ratio. In *Proceedings of the 14th European Conference on Cognitive Ergonomics: Invent! Explore!* (pp. 185–188). New York, NY: ACM.

Kong, N., Heer, J., & Agrawala, M. (2010). Perceptual guidelines for creating rectangular treemaps. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 990–998.

Talbot, J., Setlur, V., & Anand, A. (2014). Four experiments on the perception of bar charts. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2152–2160.

Do not copy, post, or distribute

Copyright ©2020 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.