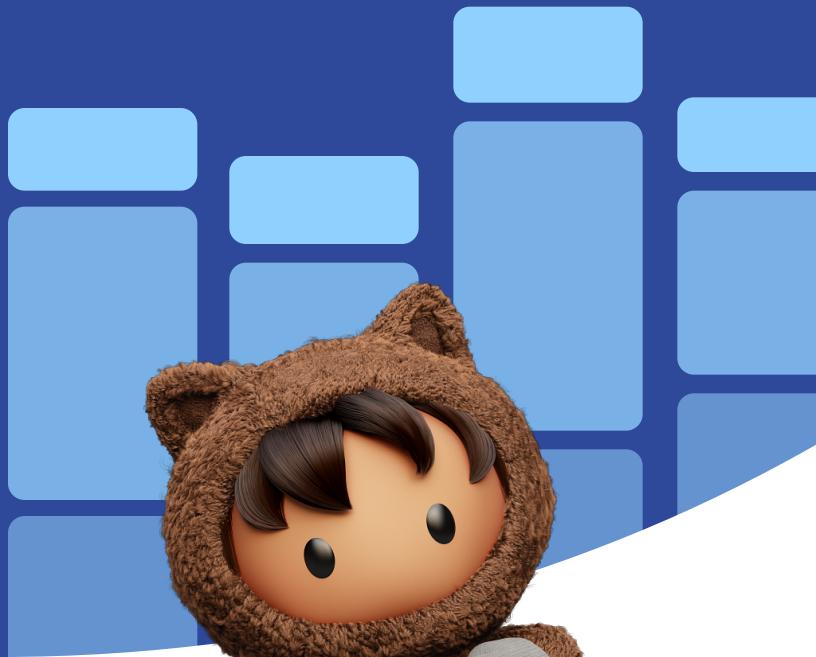
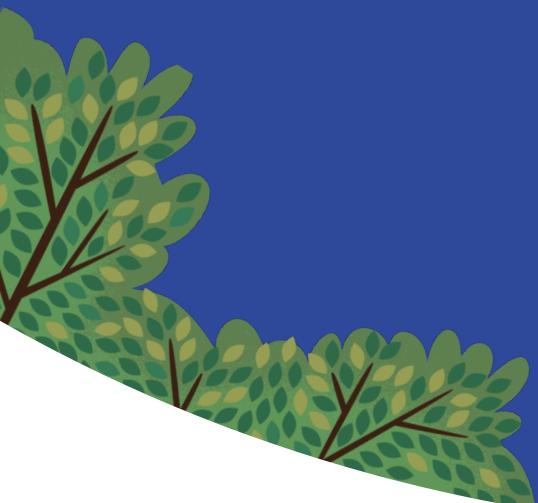




Visualizing Survey Data



In partnership with



About the author

Steve Wexler is the founder of *Data Revelations*, author of *The Big Picture: How to Use Data Visualization to Make Better Decisions—Faster* (winner of Data Literacy's Most Innovative Data Book) and co-author of *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios*.

Steve has worked with ADP, Gallup, Johnson & Johnson, Deloitte, ExxonMobil, Convergys, Consumer Reports, The Economist, SurveyMonkey, Con Edison, D&B, Marist, Cornell University, Stanford University, Tradeweb, Tiffany, McKinsey & Company, and many other organizations to help them understand and visualize their data. A Tableau Visionary (Hall of Fame) and Iron Viz winner, Steve also serves on the advisory board to the Data Visualization Society and is co-host of the popular webinar series Chart Chat (winner of Data Literacy's Most Interesting Podcast award).

His presentations and training classes combine an extraordinary level of product mastery with real-world experience gained through developing thousands of visualizations for dozens of clients. Steve has taught thousands of people in both large and small organizations and is known for conducting his workshops and presentations with clarity, patience, and humor.

Steve offers an on-demand class on visualizing survey data using Tableau on his website. [See bigpic.me/course](https://www.bigpic.me/course)

Steve also offers many free resources on visualizing survey data on his website. [See bigpic.me/surveydata](https://www.bigpic.me/surveydata)

Steve Wexler
DataRevelations.com
sbewxler@DataRevelations.com
 914.945.0567



Contents

Overview.....	4
The Business Case for Using Visualization Tools with Survey Data.....	5
Better Insights through Ad Hoc Exploration.....	5
Going Beyond Exploration and Creating Insightful Visualizations	9
Longitudinal Analysis.....	15
Preparing Individualized Dashboards for Stakeholders and Participants.....	16
Addressing Uncertainty and Margin of Error	18
Testing for Statistical Significance	21
Creating Visualizations that Blend Your Survey Data with Third-Party Data.....	22
Reshaping Survey Data	23
Why Having Both Text and Numeric Responses Is Useful.....	25
Tools to Get Your Data in the Right Format	27
Building Visualizations—Some More Examples.....	31
Conclusion.....	43
Where to Learn More	43
About Tableau and Additional Resources.....	44



Overview

Many commercial survey tools allow you to build complex surveys and collect data. But these tools are not as robust when it comes to analyzing and visualizing the collected data. All too often the best stories in the survey data remain hidden behind canned reports that are too difficult or expensive to customize.

In this whitepaper, we will look at the advantages of using visualization tools with survey data. We'll share how you can use these tools to explore, derive, visualize, and share key insights.

The examples in this report use Tableau, but the concepts and techniques can be applied to many visualization tools. With whatever tool you use, make sure it provides a flexible and interactive user experience that lets you easily shift perspectives and combine data sources.

This whitepaper is divided into three sections:

Section	Summary	Page
The Business Case for Using Visualization Tools with Survey Data	Learn how visualization tools can help you find the best stories in your data and see examples of visualizations that best tell these stories.	5
Reshaping Data	Learn how to get your data in the optimal format so that the data is easy to work with in your data visualization tool.	23
Building Visualizations	See additional examples of the types of visualizations you can create.	31



I'm borrowing this image from my book, *The Big Picture: How to Use Data Visualization to Make Better Decisions Faster*. If you see this icon it means you shouldn't use this type of data visualization.

The Business Case for Using Visualization Tools with Survey Data

In this section we'll look at some of the benefits that come from using a strong visualization tool with survey data. We'll explore:

- Better Insights through Ad Hoc Exploration
- Creating Insightful Visualizations
- Longitudinal Analysis
- Preparing Individualized Dashboards for Stakeholders and Participants
- Addressing Uncertainty and Margin of Error
- Testing for Statistical Significance
- Integrating with Additional Data that Live Outside Your Survey Tool

Better Insights through Ad Hoc Exploration

You will probably come to your survey analysis with some preconceived notions and hypotheses about your data. As you vet your hypotheses, you should be conscious of three things:

1. What is the relationship/correlation among the elements you're exploring?
2. Are there any large differences in survey responses among different demographics? For example, do older people respond differently than younger people?
3. Are the differences statistically significant? How can we show the margin of error to our stakeholders?

To answer these questions, you need a tool that lets you quickly see whether there is any validity to your theories and explorations. In other words, you don't want to have to spend hours generating a report or building a dashboard only to discover what you thought would yield insights did not.

The example on the next page shows some simplified salary survey data in Tableau. The view shows the average salary for all respondents.

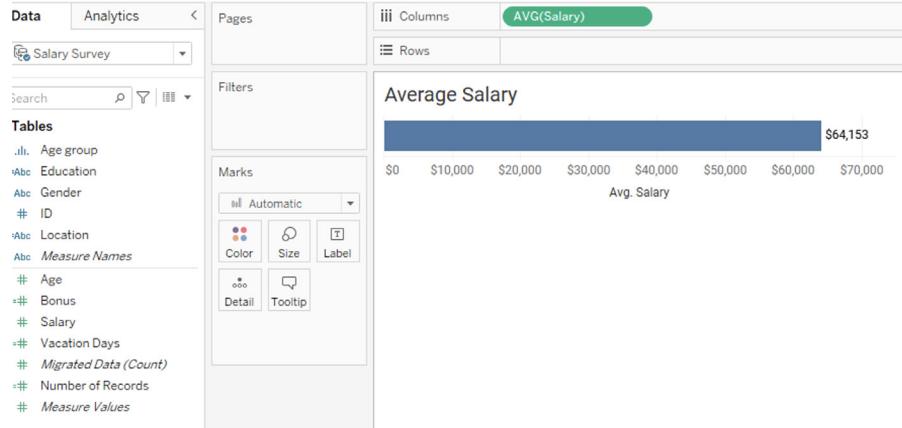


Figure 1: This simple bar chart shows the average salary for all respondents.

Upon seeing a chart like this, you may want to know things like:

- What is the salary breakdown by gender?
- What is the salary breakdown by age?
- What is the salary breakdown by location?

A good visualization tool will allow you to answer these questions quickly. In Tableau, you need only drag the applicable category (aka “dimension”) to the Rows shelf. Here’s what the resulting visualization looks like when we partition the average salary by gender.

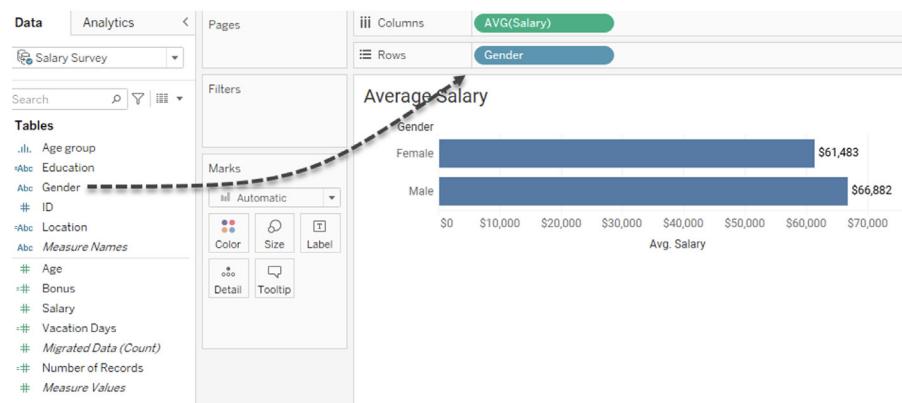


Figure 2: By dragging Gender to the Rows shelf, we can see that men, on average, make about \$2,500 more than women.

Here's what happens if instead of Gender you drag Age (bin) to the Rows shelf.

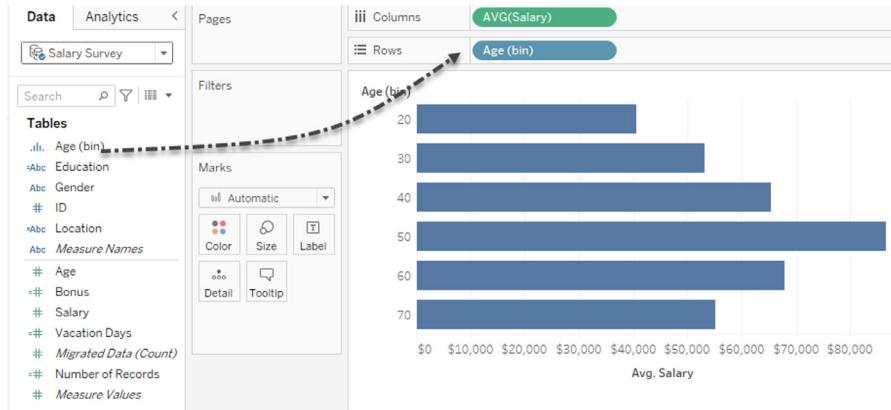


Figure 3: This view shows salary broken down by age bins. We can see that the average salary crests between the ages of 50 and 59.

Of course, you'll want a tool that allows you to combine both inquiries into a single visualization, as we see below.

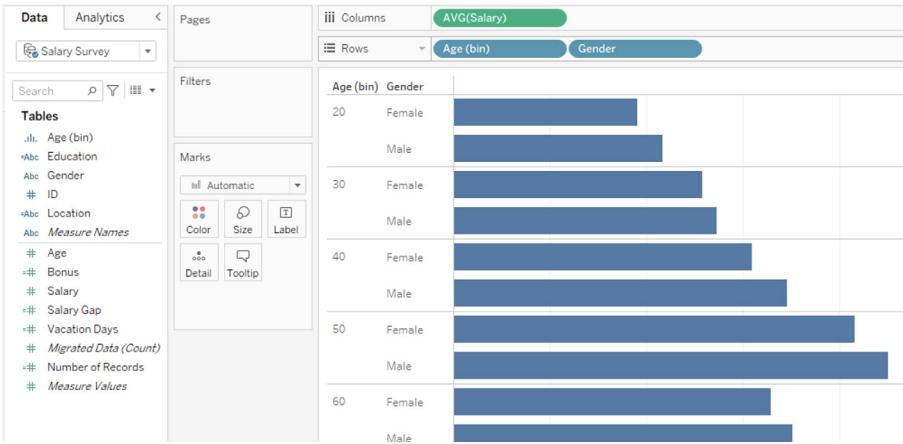


Figure 4: Here we see salary broken down by age bins and gender.

Showing the Differences with Gap Charts

It appears there's a salary gender gap across all age groups. Which group has the largest gap and which has the smallest?

Seeing this with groups of paired bars in Figure 4 can be a little difficult. Indeed, if we were comparing locations instead of gender then we'd have four or five bars for each age bucket, rather than two for gender, and that would be very difficult to parse.

In situations like this, organizations such as Pew Research or the *The Economist* will use a connected dot plot (also called a barbell chart, a dumbbell chart, or the name I prefer, a gap chart.) I've not seen any survey tool offer this capability, but you can make such a chart using many data visualization tools. Here's a gap chart created using Tableau that shows the differences in salaries for men and women across different age bins.

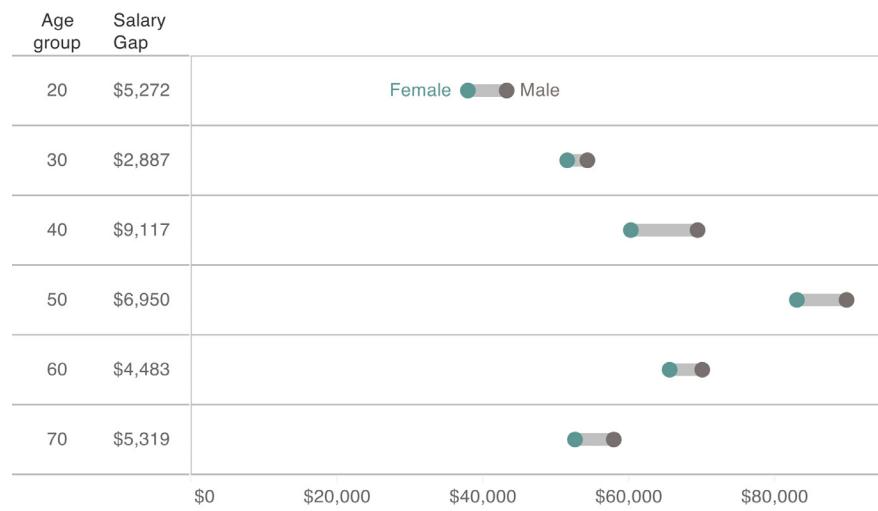


Figure 5: A connected dot plot, also known as a gap chart. It's easy to see that there are gaps across all age groups and that the gap for people between the ages of 30 and 39 is smallest and between 40 and 49 is largest.

As you experiment, you might find that many of your explorations will not yield anything particularly insightful, and that's fine. A good visualization tool rewards exploration, while limiting wasted time on dead ends or fruitless hypotheses.

Going Beyond Exploration and Creating Insightful Visualizations

Gleaning insight from the data is important, but how you communicate that insight to others is also critical and you will want to find the visualization that gets your audience to “aha” faster. Here are some examples.

Showing Sentiment: Likert-Scale Questions

There are certain types of survey questions that have been traditionally difficult to visualize, with Likert-scale questions being at or near the top of the list. Here’s a typical set of Likert- scale questions as they would appear in a survey.

	Not at all satisfied	Somewhat dissatisfied	Neutral	Satisfied	Very Satisfied
24-7 Support	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Ability to Customize UI	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Ability to filter based on role	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ease of Learning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Ease of Use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Export to .CSV and PDF	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Localized UI	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Price	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Response Time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Support for mobile devices	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 6: Here’s a set of Likert-scale questions as they would appear to a person taking a survey.

Below is an example of an attempt to visualize a collection of Likert-scale questions, in this case from a survey about people's satisfaction with particulars of a company's product.

Satisfaction Findings -- Bar Chart

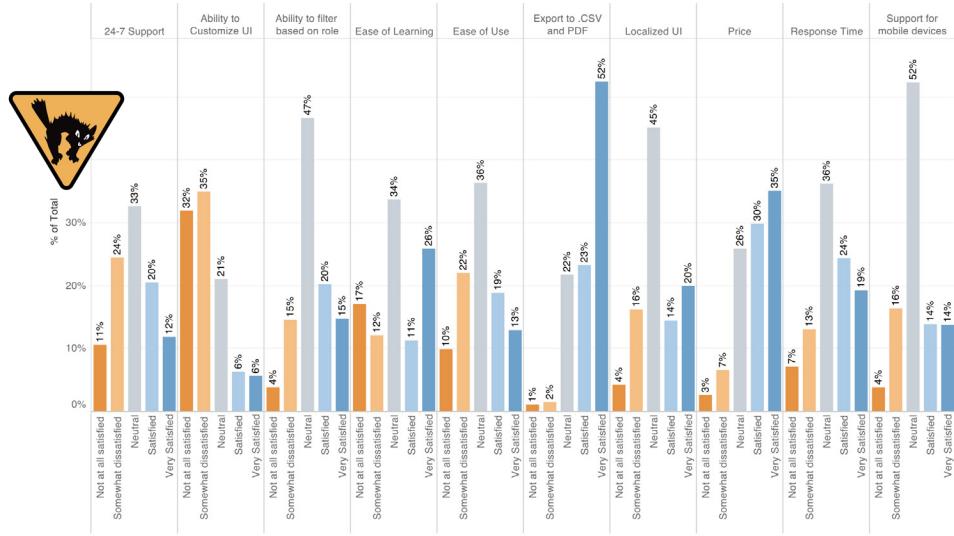


Figure 7: This view shows a typical but difficult-to-interpret bar chart (so many bars, segments, colors, and vertical labels).

Here is a somewhat better attempt, using a stacked bar chart.

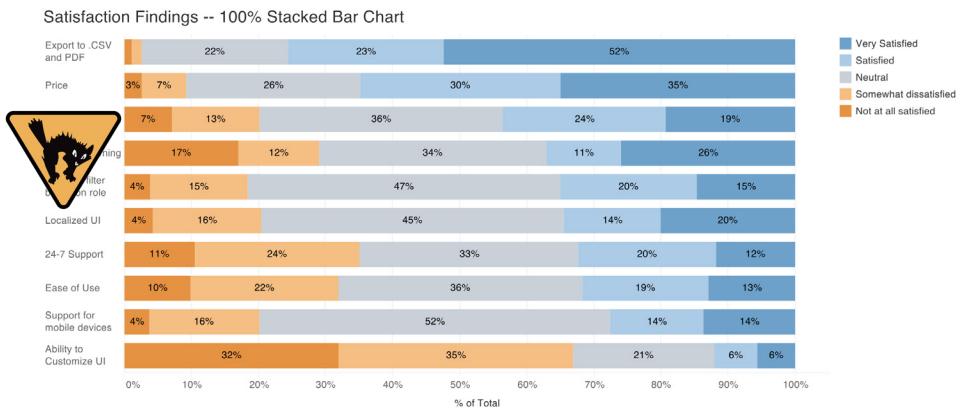


Figure 8: This 100% stacked bar chart is also difficult to read.

Using a divergent stacked bar chart, we can present a much more insightful story with respect to respondent tendencies. In the example below we see three levels of sentiment instead of five (we're combining “Satisfied” with “Very Satisfied”, and “Not at all Satisfied” with “Somewhat Dissatisfied”).

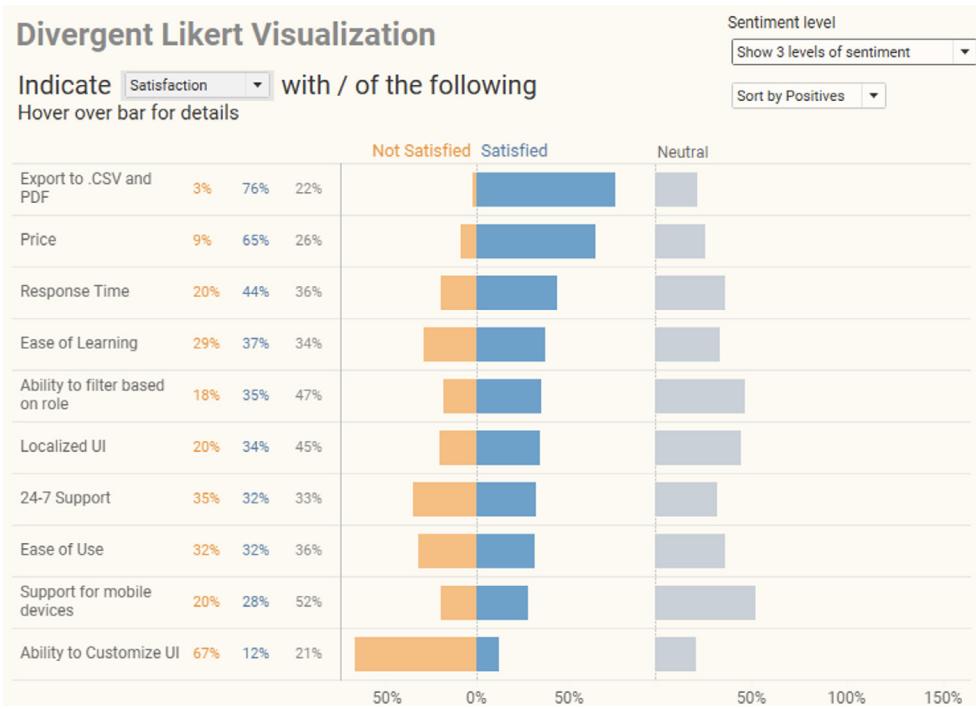


Figure 9: This divergent stacked bar chart shows three levels of sentiment. Notice that we can easily compare negative, positive, and neutral sentiment.

With an interactive dashboard you can give your audience the ability to see details on demand. For example, in the figure below, we can show different levels of sentiment (1), change the sort order (2), and hover over a bar to see details (3).



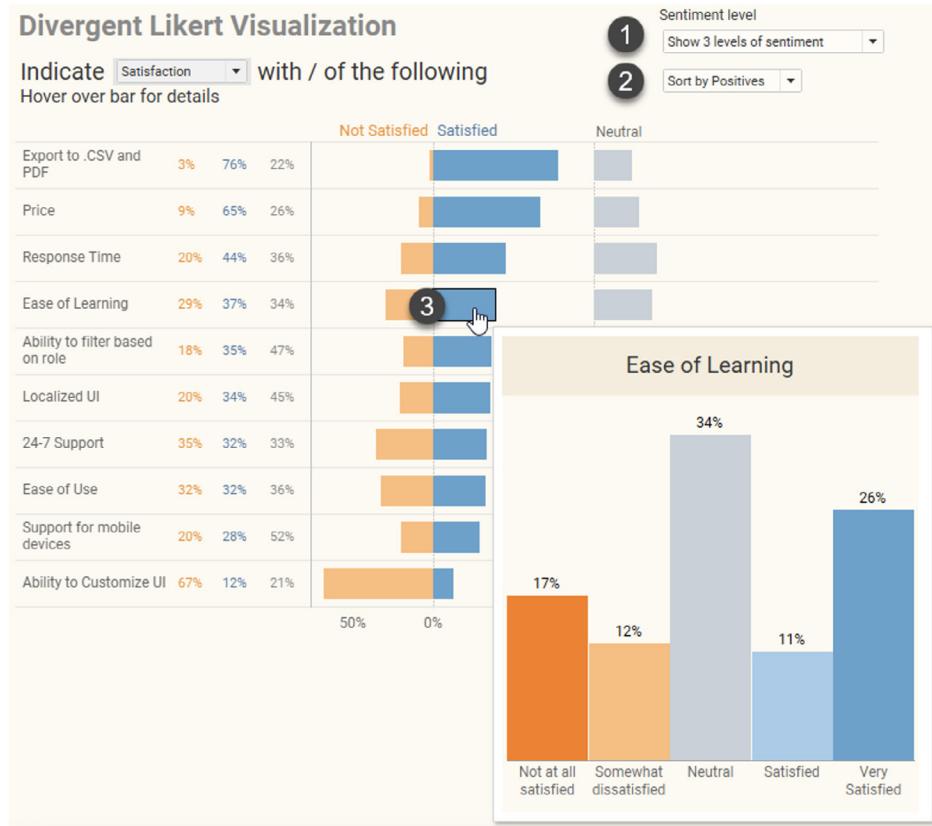


Figure 10: A Tableau dashboard giving the audience different views of the detail and the ability to see details on demand.

The Challenges of Visualizing Likert-Scale Data

Why show only three-levels of sentiment? Why are the neutrals over to the side? What about showing percent top two boxes only; i.e., just the folks who selected a 4 or 5 using a 5-point scale?

One could write a thick whitepaper just on Likert-scale data so if this has piqued your interest, please see these articles on why I think certain approaches work best:

www.datarevelations.com/howto-likert/

www.datarevelations.com/got-likert-data-neutrals/

www.datarevelations.com/rethinkingdivergent/

Net Promoter Score (NPS) Analysis

A common way to gauge consumer sentiment is to ask survey respondents to indicate, based on a 0-to-10 scale, whether they would recommend a product or service.

Respondents that provide a rating of 0 to 6 are considered “detractors” while those that give a rating of 9 or 10 are considered “promoters.” Those giving a rating of 7 or 8 are considered “neutral.” The Net Promoter Score is computed by taking the percentage of promoters minus the percentage of detractors and multiplying by 100.

Here's a bar chart showing NPS across various brands.

Net Promoter Score for Various Brands

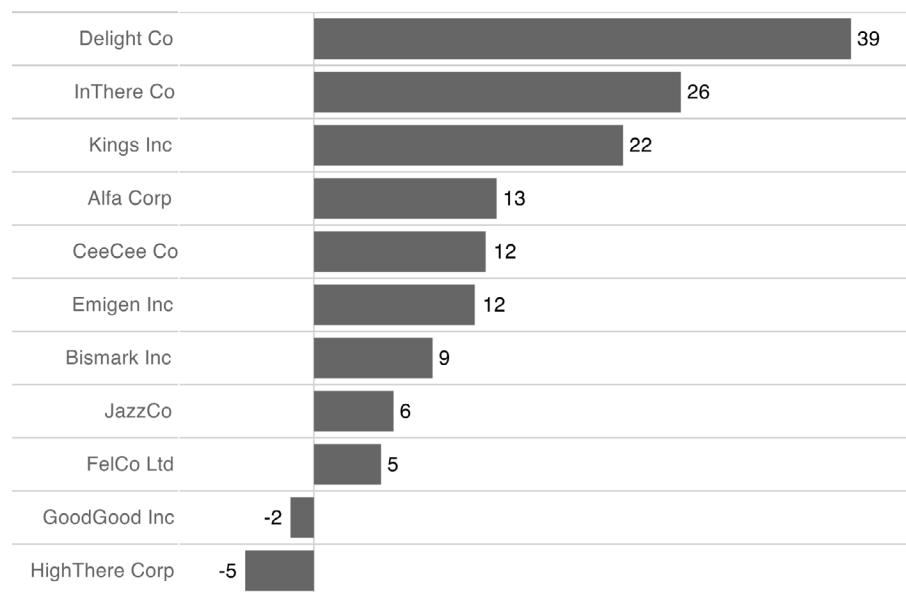


Figure 11: A bar chart showing Net Promoter Score (NPS). Of course, you'll want a tool that allows you to combine both inquiries into a single visualization, as we see below.

Consider the brand FelCo Ltd with its NPS of 5 (it's third from the bottom). There are a lot of ways to have a 5 NPS score. For example, maybe a lot of people love the brand (52.5 % promoters) and almost as many people hate the brand (47.5% detractors). Or maybe there are a lot of neutrals and if the company can win them over, they'll have a lot more people touting the brand.

Here's a considerably more potent visualization that helps people see the extent to which people are much more likely to recommend or not recommend various brands.

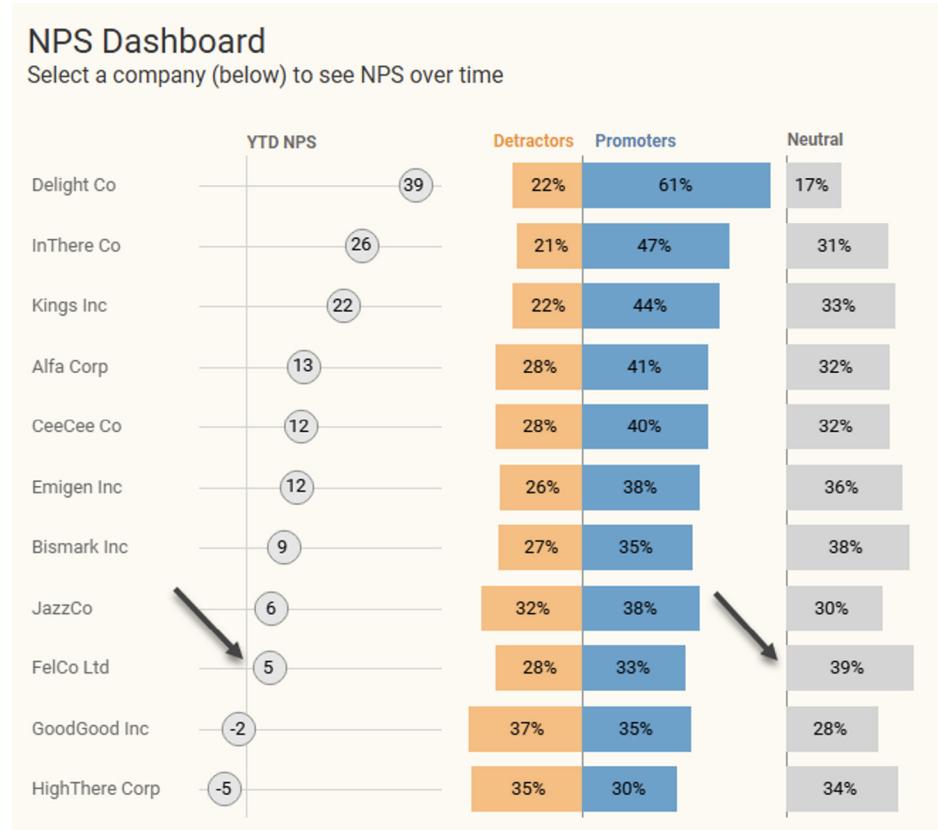


Figure 11: NPS dot plot paired with a stacked bar chart. This is easy to create and does a much better job of showing what comprises the score. FelCo has a lot of people who are “on the fence” (the neutrals), the most of any of the brands listed.



Longitudinal Analysis

The NPS dashboard from the previous example shows year-to-date net promoter score. Suppose you want to see how the NPS for a particular company has changed or not changed so far this year. Clicking that company will show how both the scores (1) and components that make up the scores (2) have changed over time.

Indeed, while FelCo's NPS may be relatively flat over time (1), we can see that the recent dip is because the percentage of neutrals and promoters has decreased and the number of detractors has increased (2).

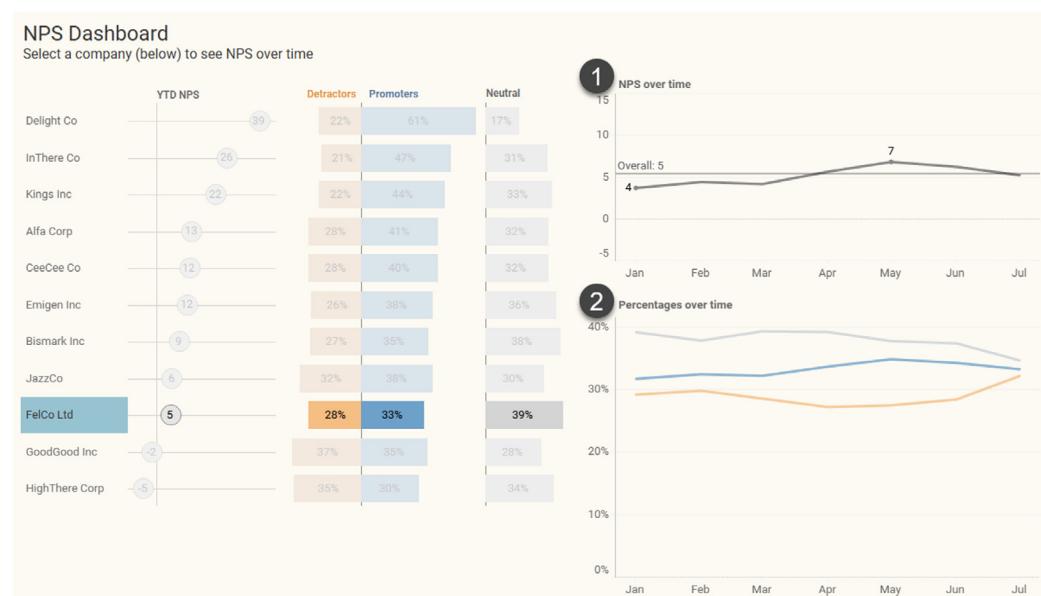


Figure 12: Selecting a company will show score changes over time.

Preparing Individualized Dashboards for Stakeholders and Participants

Consider the following types of questions a stakeholder or participant might ask with respect to a survey on performance:

- How do our customer satisfaction scores this year compare to last year's?
- How is our division doing compared to other divisions?
- How is the store I manage doing compared with other stores?

A powerful visualization tool makes it fast and easy to answer these types of questions and generate individualized reports and dashboards.

The example below comes from my book *The Big Picture*.

Here we see a personalized dashboard that shows how the manager responsible for store S34 is doing compared with other stores. Note that every manager gets this personalized view where their dot is the centerpiece of the story.

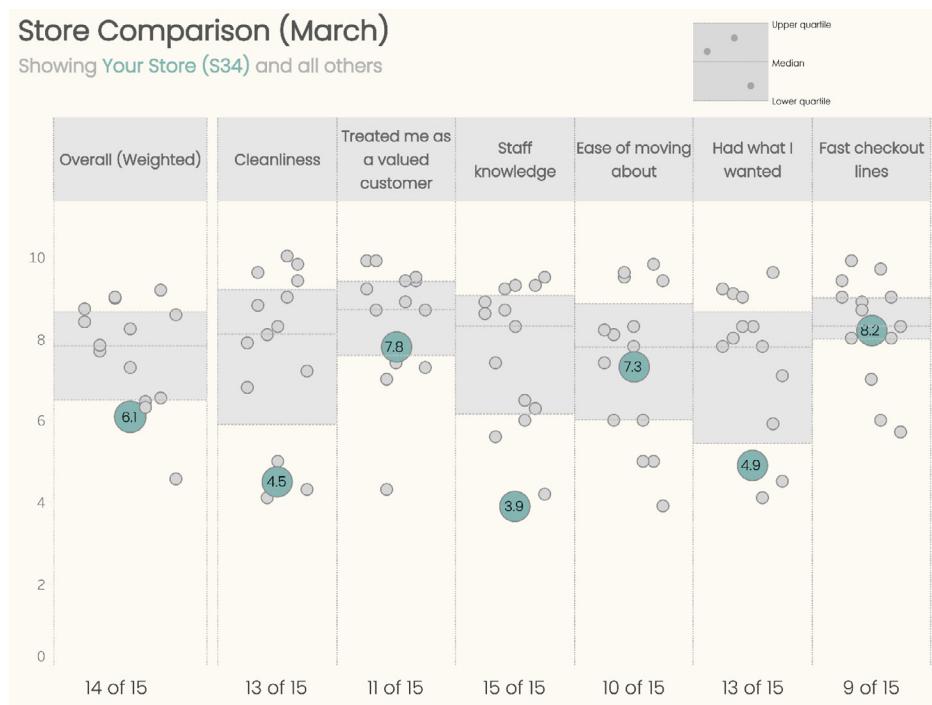


Figure 13: Store comparison dashboard for March.

The manager of this store will likely be concerned as the store received an overall weighted score of 6.1 out of 10 and is ranked 14 out of 15. It's easy to see that this store is performing poorly with respect to Cleanliness, Staff knowledge, and Had what I wanted (just look for the green dots that are significantly lower than the other dots.)

This is a chance for the manager's manager (e.g., the director) to help to improve performance.

You may be thinking, "goodness, I would not want the public humiliation of everyone being able to see how poorly my store is performing." Realize that in this example nobody except the director and senior management knows which dots belong to which stores. As a manager you can see that there are other stores, with some doing well and some doing poorly, but you cannot tell which stores (other than your own) each dot represents.

If I were the director, I would look at which stores were doing well and consider having managers of those stores confer with the manager of store S34. I'd also create clear goals and incentives to improve performance. Imagine how good the manager of S34 will feel in April when the dashboard shows significant progress from the previous month (Figure 14.)

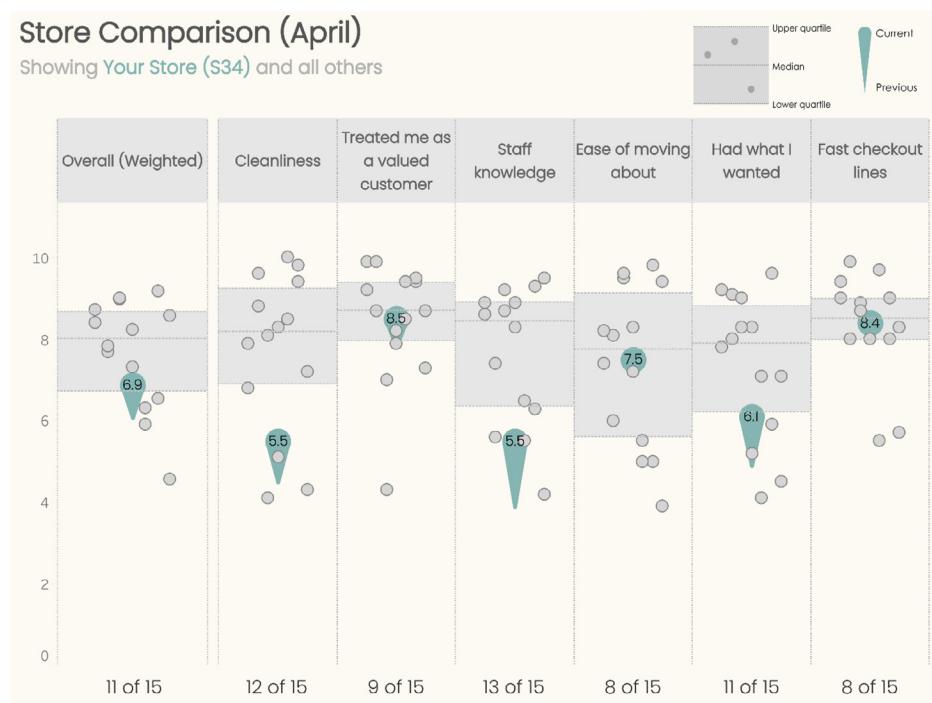


Figure 14: Store comparison dashboard for April showing changes for from March for the highlighted store.

At some point the organization may be in a place that all the stores are performing well, even the store ranked last. If this happens maybe it would make sense to compare store performance against some other quantifiable measure such as improvement from the previous year, comparison with a competitor, etc.

Note: Every time I've seen this type of comparative dashboard used, it has resulted in those who are not performing well to strive to do better. Yes, it can be upsetting to see "your dot" in a bad location, but it can be uplifting to see the placement of the dot improve over time, and that is exactly the type of impact you want to have with data visualization.

Addressing Uncertainty and Margin of Error

Imagine you've just fashioned a dashboard showing the percentages of people that responded to a check-all-that-apply question and you're quite pleased with how clear and functional it is. You even have an on-demand parameter that will allow you to break down the results by gender, generation, and so on.

Percent of respondents selecting symptom

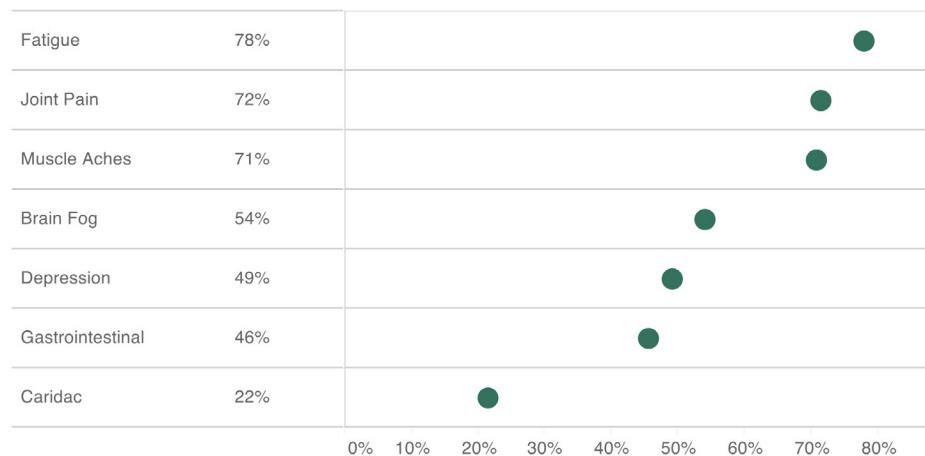


Figure 15 A dot plot showing the percentage of respondents reporting having a symptom.

In your preliminary write-up you note that people reporting Fatigue is ranked highest with 78% and Joint pain is second with 72%. Before presenting your findings, you ask one of your colleagues to review your work and she asks you "what is the margin of error for these results?"

Not sure what she means, you ask her to clarify. "If you were to conduct this survey again with a similarly-sized group of people, how confident are you that you would get the same result?"

Clueless on how to determine this, your colleague explains some very useful statistical methods built around the mean limit theorem and how these simple formulas can help you display a range of values for the survey results.

You thank her profusely, do a little research, then come back and tell her that, with a confidence interval of 95%, the range of values for Fatigue is 72% to 84% and the range for Joint Pain is 66% to 78%.

Besides reporting this, how can you make it easy for your audience to see that the so-called “true” value is somewhere within a range of values?

Here's an approach that uses error bars to indicate the range of values that we can be 95% confident contain the population wide rate.

Percent of respondents selecting symptom (95% confidence interval with 205 responses)

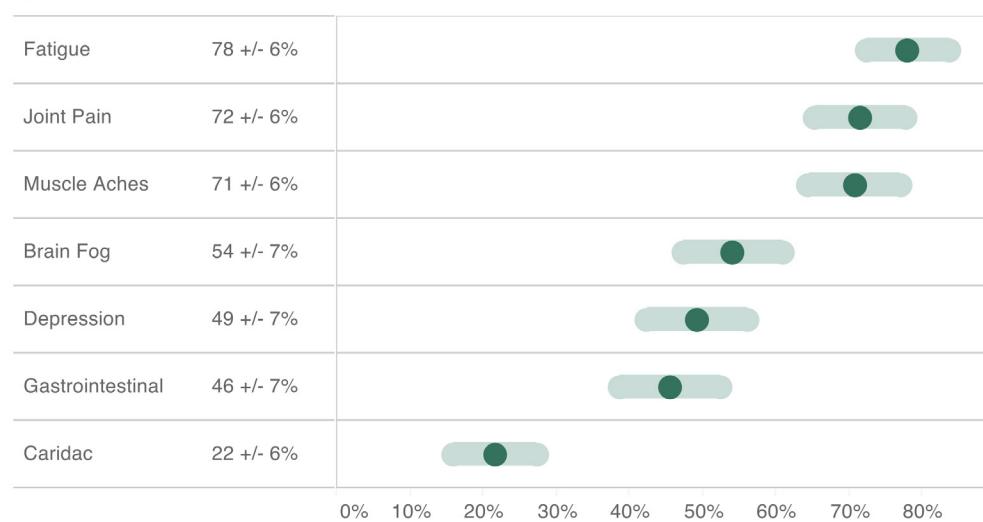


Figure 16: A dot plot with error bars showing the range of possible values.

Also note that we're showing the number of responses in the title (205, often reported as “N=205”). Indeed, it's the relatively small number of responses that lead to the relatively wide margin of error. Compare the error range in the image above with the one below where we have 1,005 responses.

**Percent of respondents selecting symptom
(95% confidence interval with 1,005 responses)**

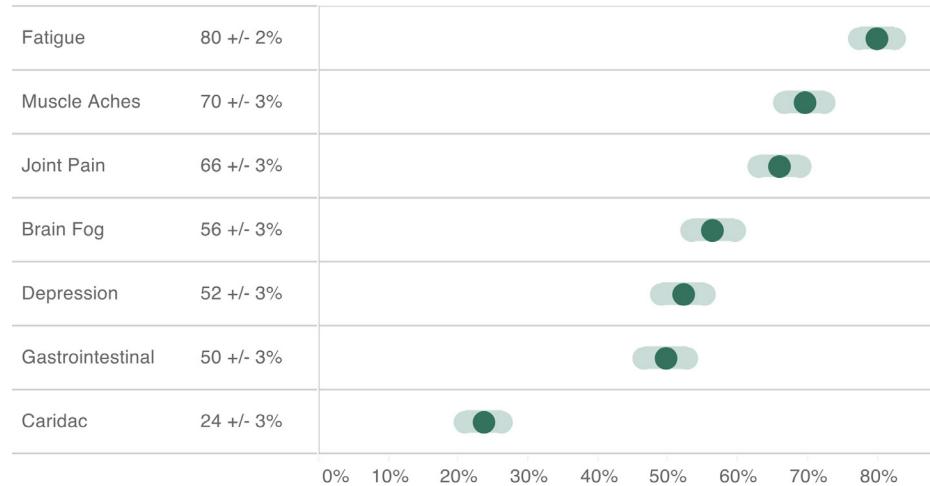


Figure 17: When the response count is larger the margin of error will be smaller.

For more thoughts on dealing with uncertainty and margin of error in reporting and visualizing survey results, please see these two links.

www.datarevelations.com/showing-uncertainty/

www.datarevelations.com/marginoferror/

Testing for Statistical Significance

In addition to margin of error, a question that everyone who analyzes survey data should ask is whether the results are statistically significant. For example, many people conduct longitudinal studies to see how people's opinion changes over time. It's very common to compare results from the current period with a previous period (e.g., this year versus last year, this quarter versus last quarter, etc.).

A common concern is if and when there is a change from a previous period, is that change statistically significant? Your visualization tool should be able to fashion different types of formulas that test for statistical significance (e.g., Z-test and T-test).

In the visualization below, the bar chart to the far left shows a percentage of people considered "promoters" (because they responded with a 9 or a 10 on a Net Promoter Score survey). The second bar chart shows the percentage change from the previous period, and the sparklines show the trends over the past 12 periods. The red dot indicates that the change from the previous period was significant based on whatever statistical test the organization deemed appropriate.

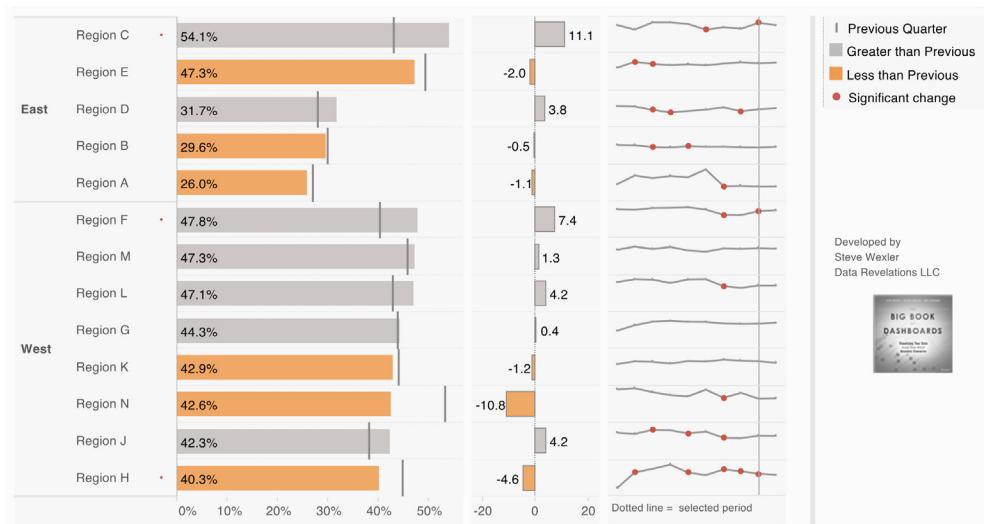


Figure 18: This combination chart shows percentage of promoters, difference from the previous period, and changes over time. The red dot indicates the change from the previous period was statistically significant.

If you need more sophisticated statistical analysis, look for a tool with this feature built in or a tool that integrates with a dedicated statistics tool (e.g., "R").



Creating Visualizations that Blend Your Survey Data with Third-Party Data

A very powerful capability that you should seek in a visualization tool is the ability to create insightful visualizations from several data sources. For example, suppose you want to combine your survey results with internal sales data, or with local census data.

Even if you use a data visualization tool that allows quick, ad hoc blending of data sources, you may want to combine your data using a data prep tool. Realize that the format your survey data is in when you download it is highly unlikely to play nicely with your data visualization tool.

The need to restructure the data brings us to our next section.



Reshaping Survey Data

Getting the Data in the Right Shape

The biggest impediment to visualizing survey data with any visualization tool is getting the source data in a format that plays nicely with the tool.

While most survey tools maintain the data in a normalized database for internal purposes, they render it in a one-row-per-respondent format for their clients.

This means you will probably receive an Excel file or a CSV file with one row per respondent and hundreds of columns, each column corresponding to a different question in the survey.

Instead, we need to pivot/reshape the data so all of the questions get merged into two columns. So instead of this...



	A	B	C	D	E	F	G	H	I	J	K
1	RespID	Q0_Gender	Q0_Location	Q0_Generation	Q0_Weight	Q1	Q100	Q2_1	Q2_2	Q2_3	Q2_4
2	2	Male	South America	Generation X	1	No	\$ 98,038	No	No	Yes	No
3	4	Female	South America	Baby Boomers	1.44	No	\$ 138,936	Yes	Yes	Yes	No
4	5	Female	South America	Generation X	1	Yes	\$ 84,471	No	Yes	Yes	Yes
5	6	Male	Antarctica	Baby Boomers	1.44	Don't know	\$ 138,534	No	Yes	Yes	No
6	9	Female	Europe	Baby Boomers	1.32	Yes	\$ 68,944	Yes	Yes	Yes	Yes
7	12	Female	Europe	Baby Boomers	1.56	No	\$ 100,663	No	No	Yes	Yes
8	15	Male	North America	Baby Boomers	1.56		\$ 122,481				
9	16	Male	Antarctica	Baby Boomers	1.44	Yes	\$ 106,036	Yes	Yes	No	No
10	17	Female	Europe	Baby Boomers	1.32	Don't know	\$ 81,681	Yes	Yes	Yes	No
11	18	Male	North America	Traditionalists	0.595	No	\$ 104,200	No	Yes	No	No
12	22	Male	South America	Generation X	1.32	No	\$ 172,723	No	No	Yes	Yes
13	25	Female	South America	Generation X	1.32	Yes	\$ 153,410	Yes	Yes	Yes	Yes
14	26	Female	South America	Millennials	0.765	Yes	\$ 93,194	No	Yes	No	No
15	27	Male	Europe	Baby Boomers	1.56	Yes	\$ 101,662	Yes	Yes	Yes	No
16	29	Male	Europe	Generation X	1		\$ 114,216				
17	30	Male	Europe	Baby Boomers	1.32	No	\$ 97,354	No	No	No	No
18	31	Male	Europe	Millennials	0.68	Yes	\$ 120,061	No	No	No	No
19	33	Male	North America	Generation X	1		\$ 134,308				
20	34	Male	North America	Generation X	1.32		\$ 146,227				
21	36	Female	North America	Millennials	1	Yes	\$ 110,462	No	Yes	Yes	Yes
22	37	Female	North America	Millennials	0.765		\$ 104,602				
23	38	Female	South America	Baby Boomers	1.32	Don't know	\$ 119,871	Yes	Yes	Yes	No
24	40	Female	North America	Baby Boomers	1.44	Don't know	\$ 134,158	Yes	No	No	No
25	42	Female	Europe	Baby Boomers	1.44		\$ 115,750				
26	43	Female	Europe	Baby Boomers	1.44		\$ 77,022				

Figure 19: This view shows data as downloaded from a survey system, which places each question in a separate column. In this case, the survey results were encoded as text ("yes," "no," "strongly agree," etc.) rather than numeric values (1, 0, 5, etc.)

... we want something that looks like this:

A	B	C	D	E	I	K
RespID	Q0_Gender	Q0_Generation	Q0_Location	Q0_Weight	Question ID	Labels
2	2 Male	Generation X	South America	1	Q1	No
3	2 Male	Generation X	South America	1	Q100	98037.68
4	2 Male	Generation X	South America	1	Q2_1	No
5	2 Male	Generation X	South America	1	Q2_2	No
6	2 Male	Generation X	South America	1	Q2_3	Yes
7	2 Male	Generation X	South America	1	Q2_4	No
8	2 Male	Generation X	South America	1	Q2_5	Yes
9	2 Male	Generation X	South America	1	Q2_6	No
10	2 Male	Generation X	South America	1	Q2_7	No
11	2 Male	Generation X	South America	1	Q2_8	No
12	2 Male	Generation X	South America	1	Q2_9	Yes
13	2 Male	Generation X	South America	1	Q28_IMP	Very Important
14	2 Male	Generation X	South America	1	Q28_SAT	Not at all satisfied
15	2 Male	Generation X	South America	1	Q29_IMP	Very Important
16	2 Male	Generation X	South America	1	Q29_SAT	Not at all satisfied
17	2 Male	Generation X	South America	1	Q3_1	Small degree
18	2 Male	Generation X	South America	1	Q3_2	Small degree
19	2 Male	Generation X	South America	1	Q3_3	Not at all
20	2 Male	Generation X	South America	1	Q3_4	Small degree
21	2 Male	Generation X	South America	1	Q3_5	Moderate degree
22	2 Male	Generation X	South America	1	Q3_6	Small degree
23	2 Male	Generation X	South America	1	Q3_7	High degree
24	2 Male	Generation X	South America	1	Q3_8	Small degree
25	2 Male	Generation X	South America	1	Q3_9	Very high degree
26	2 Male	Generation X	South America	1	Q30_IMP	Very Important
27	2 Male	Generation X	South America	1	Q30_SAT	Not at all satisfied
28	2 Male	Generation X	South America	1	Q31_IMP	Very Important
29	2 Male	Generation X	South America	1	Q31_SAT	Not at all satisfied
30	2 Male	Generation X	South America	1	Q32_IMP	Very Important
31	2 Male	Generation X	South America	1	Q32_SAT	Not at all satisfied
32	2 Male	Generation X	South America	1	Q33_IMP	Of Little Importance
33	2 Male	Generation X	South America	1	Q33_SAT	Not at all satisfied
34	2 Male	Generation X	South America	1	Q34_IMP	Of Little Importance

Each respondent is listed 40 times, one time for each question.

Figure 20: The data is reshaped with only two columns for each question (not including the demographic questions), resulting in many more rows. Note that we do not yet have anything in place that groups related questions together or that translates things like “Q7” into something meaningful. Of course, you’ll want a tool that allows you to combine both inquiries into a single visualization, as we see below.

Actually, what we really want is something that looks like this:

Demographic data Question ID numeric responses text responses Meta data

A	B	C	D	E	F	G	H	I	J	K
RespID	Q0_Gender	Q0_Generation	Q0_Location	Q0_Weight	Question ID	Value	Labels	Qtype	Question Grouping	Wording
2	2 Male	Generation X	South America	1	Q1	0 No	Single-Punch	Vote	Vote in the upcoming election?	
3	2 Male	Generation X	South America	1	Q100	98037.68		Benchmark	Salary	What is your salary?
4	2 Male	Generation X	South America	1	Q2_1	0 No	Multi-Punch	What do you measure	Pulse Rate	
5	2 Male	Generation X	South America	1	Q2_2	0 No	Multi-Punch	What do you measure	Metabolism	
6	2 Male	Generation X	South America	1	Q2_3	1 Yes	Multi-Punch	What do you measure	Blood Pressure	
7	2 Male	Generation X	South America	1	Q2_4	0 No	Multi-Punch	What do you measure	Temperature	
8	2 Male	Generation X	South America	1	Q2_5	1 Yes	Multi-Punch	What do you measure	Galvanic Skin Response	
9	2 Male	Generation X	South America	1	Q2_6	0 No	Multi-Punch	What do you measure	Breathing	
10	2 Male	Generation X	South America	1	Q2_7	0 No	Multi-Punch	What do you measure	Perspiration	
11	2 Male	Generation X	South America	1	Q2_8	0 No	Multi-Punch	What do you measure	Pupil Dilation	
12	2 Male	Generation X	South America	1	Q2_9	1 Yes	Multi-Punch	What do you measure	Adrenaline Production	
13	2 Male	Generation X	South America	1	Q28_IMP	5 Very Important	Likert	Importance	Price	
14	2 Male	Generation X	South America	1	Q28_SAT	1 Not at all satisfied	Likert	Satisfaction	Response Time	
15	2 Male	Generation X	South America	1	Q29_IMP	5 Very Important	Likert	Importance	Response Time	
16	2 Male	Generation X	South America	1	Q29_SAT	1 Not at all satisfied	Likert	Satisfaction	Good Sense of Humor	
17	2 Male	Generation X	South America	1	Q3_1	2 Small degree	Likert	Indicate degree to which you agree	Good Job Skills	
18	2 Male	Generation X	South America	1	Q3_2	2 Small degree	Likert	Indicate degree to which you agree	Good Sense of Humor	
19	2 Male	Generation X	South America	1	Q3_3	1 Not at all	Likert	Indicate degree to which you agree	High Intelligence	
20	2 Male	Generation X	South America	1	Q3_4	2 Small degree	Likert	Indicate degree to which you agree	Can Play Jazz	
21	2 Male	Generation X	South America	1	Q3_5	3 Moderate degree	Likert	Indicate degree to which you agree	Likes the Beatles	
22	2 Male	Generation X	South America	1	Q3_6	2 Small degree	Likert	Indicate degree to which you agree	Good Ability to lift heavy objects	
23	2 Male	Generation X	South America	1	Q3_7	4 High degree	Likert	Indicate degree to which you agree	Has stare under eye	

Figure 21: Behold, survey data nirvana! Label responses, numeric responses, logical groupings, and mapping of question IDs are all in a form that is human-readable.

While we don't have to have the labels, values, question grouping and wording fields, these elements will make our work much faster and easier (more on this later).

Why Having Both Text and Numeric Responses Is Useful

When it comes to questions based on the five-point Likert scale, the universe of possible numeric values is this:

1 2 3 4 5

Suppose we want to know just what each of the values (1, 2, 3, 4, and 5) stand for? The problem is that it depends on the question being asked. Sometimes a 5 means "strongly agree" while for other questions it means "critical" or perhaps "extremely satisfied."

Without having both the numeric and text results, we may have to write a lot of IF or CASE statements. This can become very cumbersome if you have a lot of survey questions.

By having both the numeric and text results we can use the numeric values to easily visualize all Likert questions, and let the text provide context for displayed questions.

Adding Useful Metadata to the Mix

While not essential, it can be enormously useful to create a table that "explains" each of the Question IDs and groups related questions together (this is metadata, or "data about the data"). For example, consider the column names in the survey data example in Figure 19 on page 23. You will see columns with names like these:

Q1
Q100
Q2_1
Q2_2
Q28_IMP
Q28_SAT
Q29_IMP
Q29_SAT

Just what does each stand for?

If you created the survey, you may know that Q28_IMP refers to the question “how important is response time?” But even if you created the survey yourself, imagine how difficult it would be to remember every question, especially if you had hundreds of them.

A great way to streamline your efforts would be to create a type of “helper” table that maps each Question ID to a human-readable form and groups related questions together. Check out the example below.

	A	B	C	D
1	QuestionID	Wording	Question Grouping	Qtype
2	Q1	Vote in the upcoming election?	Vote	Single-Punch
3	Q100	What is your salary?	Salary	Benchmark
4	Q2_1	Pulse Rate	What do you measure	Multi-Punch
5	Q2_2	Metabolism	What do you measure	Multi-Punch
6	Q2_3	Blood Pressure	What do you measure	Multi-Punch
7	Q2_4	Temperature	What do you measure	Multi-Punch
8	Q2_5	Galvanic Skin Response	What do you measure	Multi-Punch
9	Q2_6	Breathing	What do you measure	Multi-Punch
10	Q2_7	Perspiration	What do you measure	Multi-Punch
11	Q2_8	Pupil Dilation	What do you measure	Multi-Punch
12	Q2_9	Adrenaline Production	What do you measure	Multi-Punch
13	Q3_1	Good Job Skills	Indicate degree to which you agree	Likert
14	Q3_2	Good Sense of Humor	Indicate degree to which you agree	Likert
15	Q3_3	High Intelligence	Indicate degree to which you agree	Likert
16	Q3_4	Can Play Jazz	Indicate degree to which you agree	Likert
17	Q3_5	Likes the Beatles	Indicate degree to which you agree	Likert
18	Q3_6	Good Ability to lift heavy objects	Indicate degree to which you agree	Likert
19	Q3_7	Has grace under pressure	Indicate degree to which you agree	Likert
20	Q3_8	Is Kind to animals	Indicate degree to which you agree	Likert
21	Q3_9	Makes good coffee	Indicate degree to which you agree	Likert
22	Q28_IMP	Price	Importance	Likert
23	Q28_SAT	Price	Satisfaction	Likert

Figure 22: This “helper” file maps each Question ID to its human-readable form and groups together related questions.

Tools to Get Your Data in the Right Format

To get the most from your visualization tool, you will need a process, tool, or technology that will pivot both the numeric-encoded responses and the text-encoded responses and join these with the survey metadata.

Here's a summary of some of the more popular tools and technologies available to help you get your data set up for visualizing survey data.

Tableau's Built-In Pivot Feature

Description: All versions of Tableau starting with Tableau 9.0 have a built-in pivoting tool that converts data with a lot of columns into data that has fewer columns and many rows.

#	Abc	Abc	Abc	#	Abc	#
Data Numbers	Data Numbers	Data Numbers	Data Numbers	Data Numbers	Pivot	Pivot
Resp ID	Gender	Location	Generation	Weight	Question ID	Numeric Values
2	Male	South America	Generation X	1.00000	Q0	0.00
4	Female	South America	Baby Boomers	1.44000	Q0	0.00
5	Female	South America	Generation X	1.00000	Q0	1.00
6	Male	Antarctica	Baby Boomers	1.44000	Q0	2.00
9	Female	Europe	Baby Boomers	1.32000	Q0	1.00
12	Female	Europe	Baby Boomers	1.56000	Q0	0.00
15	Male	North America	Baby Boomers	1.56000	Q0	null
16	Male	Antarctica	Baby Boomers	1.44000	Q0	1.00
17	Female	Europe	Baby Boomers	1.32000	Q0	2.00

Figure 23: Use Tableau's pivot feature to wrangle your survey data.

Advantages: Easy to use, updates as the data updates, free with Tableau.

Disadvantages: Cannot join pivoted data so you cannot combine text-encoded survey results with numeric-encoded results. Metadata can only be added with a blend / relationship. Only works with Excel and CSV files.

Appropriate use: Surveys that do not have many questions.

More information: [Reshaping Survey Data with Tableau 9.0 and later](#)

Tableau Prep Builder

Description: This powerful data manipulation tool offers extract/transform/load capabilities as well as robust data blending and processing.

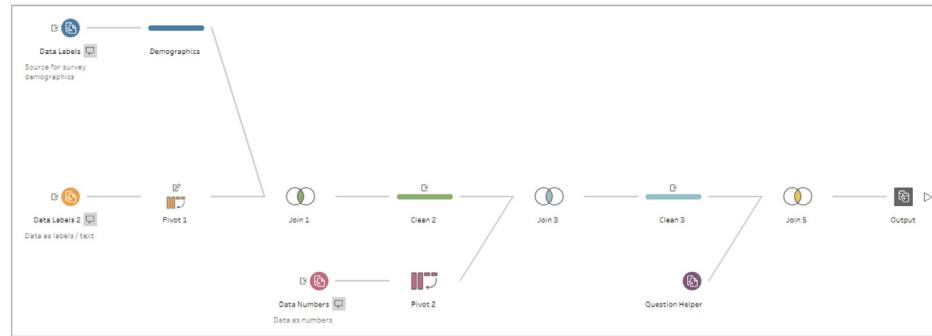


Figure 25: Tableau Prep Builder at work.

Advantages: Flexible and easy. Works with many data sources. Can combine multiple data sources and handle poorly-coded survey results. Can save as .HYPER file (local or published), CSV, or to a database. Workflow can be completely automated. Included with Tableau Creator license.

Disadvantages: Limited support for SPSS files (although you can save your SPSS files in CSV format).

More information: [Getting your survey data “just so” using Tableau Prep](#)



Alteryx Designer

Description: This powerful data manipulation tool offers extract/transform/load capabilities (ETL), geospatial analysis, and predictive modeling capabilities.

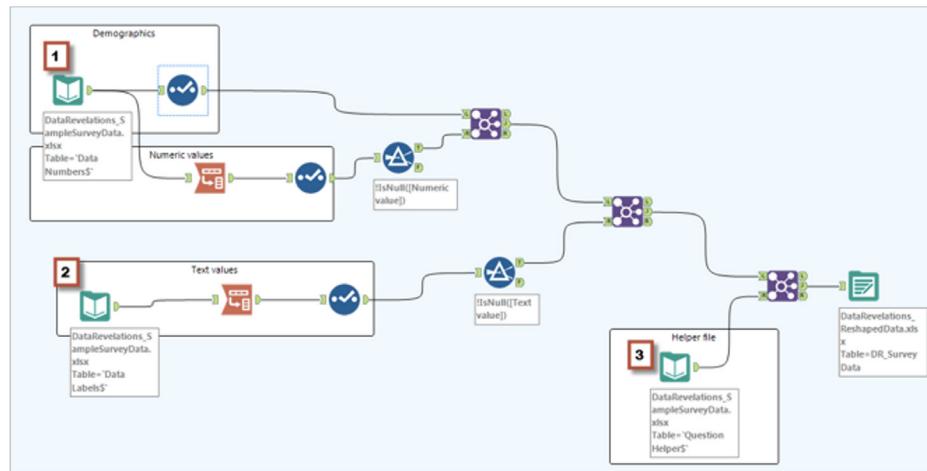


Figure 24: Alteryx Designer at work.

Advantages: Very flexible. Works with virtually any data source (including SPSS files). Can combine multiple data sources and handle poorly-coded survey results. Can save directly to a Tableau data extract file (HYPER file). Workflow can be completely automated.

Disadvantages: Learning curve. Price.

More information: [Getting Survey Data “Just So” Using Alteryx](#)



EasyMorph

Description: This very friendly tool offers extract/transform/load capabilities as well as robust data blending and processing. Great support for SPSS files.

The screenshot shows the EasyMorph application interface with three main windows:

- Imported table 1:** A large table view showing data from a survey. The columns include RespID, Gender, Location, Generation, Weight, Question ID, Labels, Values, Wording, and Question Grouping. The data shows responses for various questions like "Vote in the upcoming election?", "Response Time", etc., across different demographic groups.
- Imported table 3:** A smaller table view showing a subset of data related to physiological measurements. Columns include QuestionID, Wording, and Question Grouping. Rows include Q0 (Vote in the upcoming election?), Q1 (Pulse Rate), Q2 (Metabolism), Q3 (Blood Pressure), Q4 (Temperature), Q5 (Galvanic Skin Response), Q6 (Breathing), Q7 (Perspiration), Q8 (Pupil Dilatation), and Q9 (Adrenaline Production).
- Imported table 2:** A table view showing a mapping or summary of question IDs and values. Columns include RespID, Question ID, and Values. The data shows counts for each question ID, such as 2, 4, 5, 6, 9, 12, 15, 16, 17, 18, 22, and 2%.

The left sidebar of the main window displays a workflow history with steps like "Import all columns from 'SurveySampleData_Sourcefiles_EasyMorph.xlsx'", "Unpivot all columns except 5 into [Question ID] and [Labels]", "Merge [Values] from 'Imported table 2'", "Merge [Question Grouping] and 1 more column from 'Imported table 3'", "Sort by [RespID] in ascending order", "Filter rows where [Values] doesn't have 'empty'", and "Create the Hyper file 'SurveyData_ViaEasyMorph.hyper'".

Figure 25: EasyMorph at work.

Advantages: Flexible and friendly. Works with many data sources and has particularly strong SPSS support. Can combine multiple data sources and can handle poorly-coded survey results. Can save directly to a Tableau data extract file (HYPER file). Workflow can be completely automated. Very reasonably priced.

Disadvantages: Somewhat confusing interface.

More information: [Getting your survey data “just so” using EasyMorph](#)

Building Visualizations—Some More Examples

In the next series of examples, we will see how to visualize some typical survey data. While the examples will use Tableau, the concepts can be applied to other data visualization tools.

So, Just Who Are These People that Took the Survey?

Before visualizing any questions, you may want to get a sense of how many survey responses you received and how the responses are broken down by gender, location, and so on.

Here's how the reshaped data appears in Tableau.

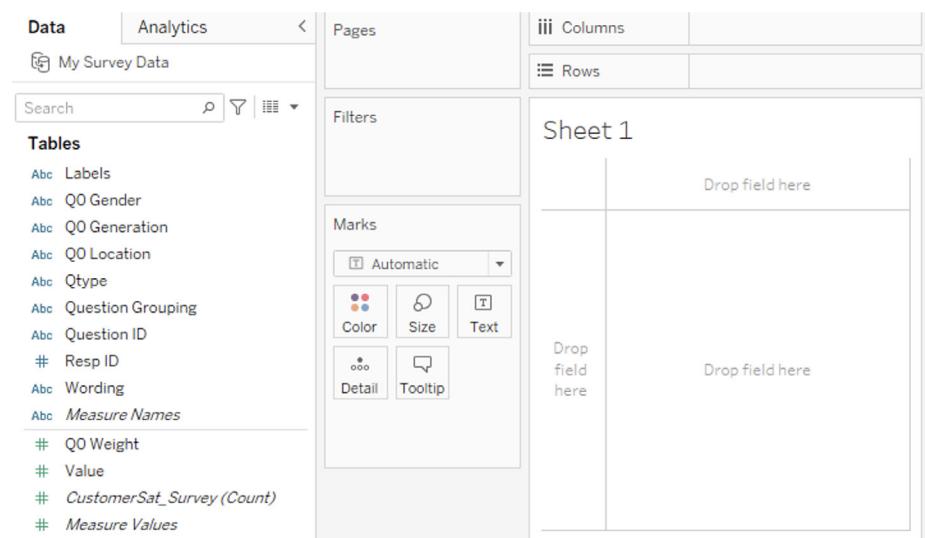


Figure 27: This view shows how the reshaped data fields appear in Tableau.

To determine just how many survey responses we received, we need to count the distinct number of Resp IDs there are in the data. You may recall that when we reshaped the data, we repeated the Resp ID for every question response, so we cannot perform a simple count.

In Tableau, we produce a distinct count using the CNTD() function as shown here.

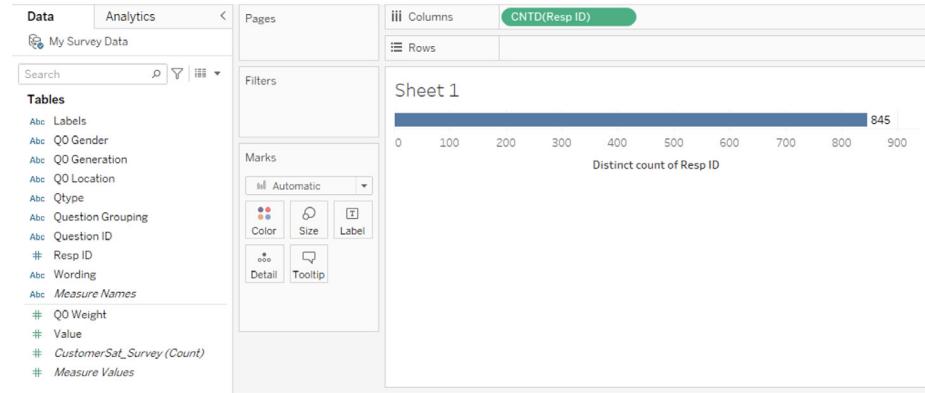


Figure 28: Using CNTD(), we see there were 845 survey responses.

If we want to see a breakdown by gender, we can simply drag Gender to the Rows shelf, as shown here.



Figure 29: This view shows the respondent count by gender.



You can repeat this exercise for all of the categories (dimensions) that interest you, then build an interactive dashboard showing the relationship among the different demographics elements, as shown here.

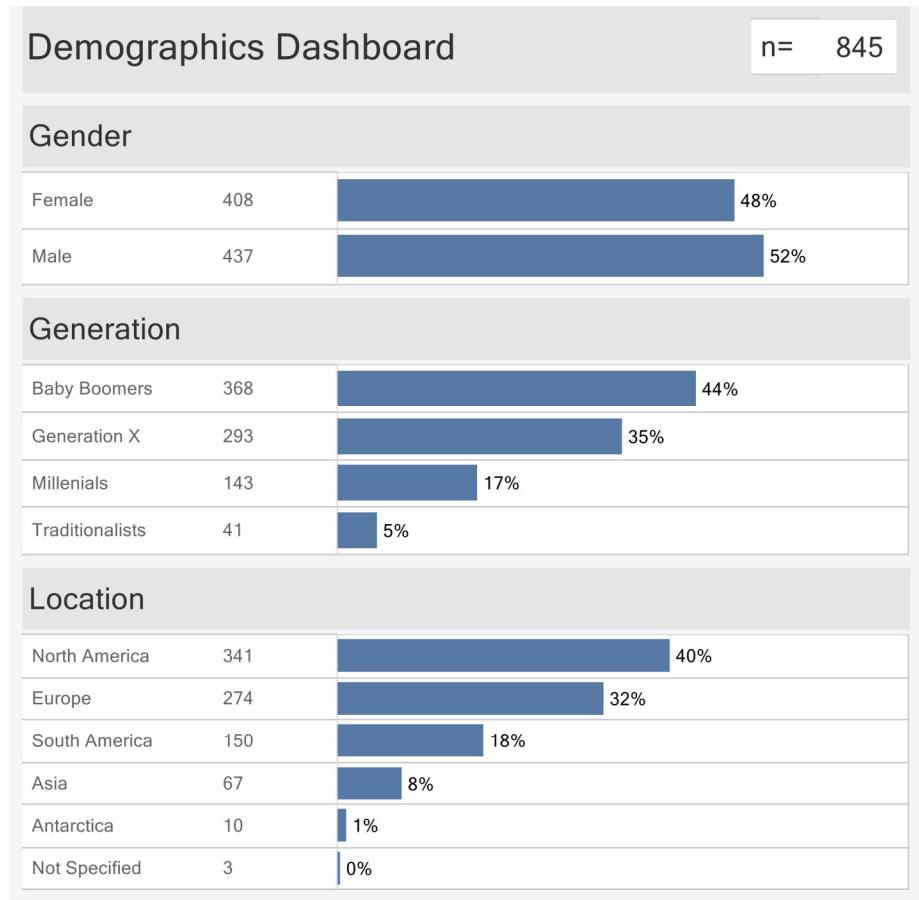


Figure 30: Here's a simple demographics dashboard. Selecting a bar or bars from one of the charts will filter the other charts.

Building a Check-All-That-Apply Visualization

Now that we have a good sense of who our survey takers are, we're ready to start visualizing their responses to different survey questions.

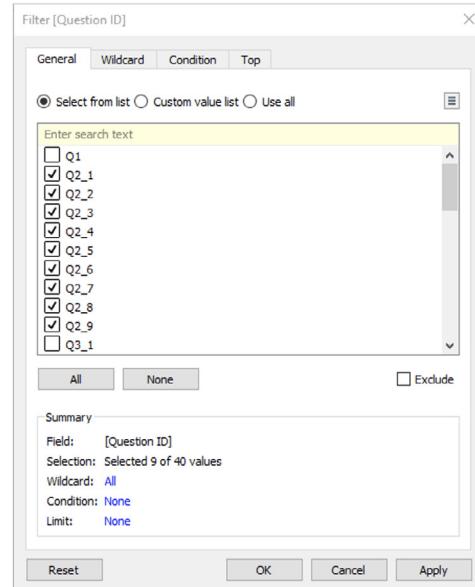
For this next example, we'll explore the responses to the question "indicate which of these things you measure; check all that apply."

So, just how are we going to focus on only those questions? This is where having that metadata will prove useful, but first let's see how to do this without the metadata.



Visualizing Survey Responses without Metadata

We will start first by dragging the Question ID field to the Filters shelf then selecting the questions that interest us.



Note:

While we'll be exploring the data with Tableau, the approach applies to other exploratory data visualization tools as well.



Figure 31: Indicate which questions we want to analyze.

Next we'll drag the Question ID onto the Rows shelf.

Question ID	Text
Q2_1	Abc
Q2_2	Abc
Q2_3	Abc
Q2_4	Abc
Q2_5	Abc
Q2_6	Abc
Q2_7	Abc
Q2_8	Abc
Q2_9	Abc

Figure 32: Here are the questions we want to analyze—but just what does “Q2_1,” “Q2_2,” and so on stand for?

If we keep using this approach, we would have to create an alias for every single question ID so that people looking at our survey results don't have to guess what question IDs such as "Q2_1" stand for.

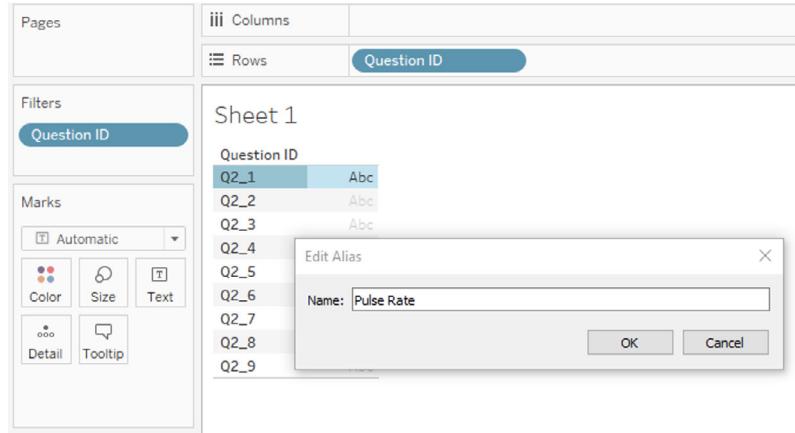


Figure 33: We need to provide a human-readable alias for each question.

While this approach will work, it will prove tedious with a lot of survey questions. Let's see how much easier it is using the metadata we looked at earlier.

Visualizing Survey Responses with Metadata

We'll start by dragging Question Grouping to the Filters shelf and selecting the group or related questions we want to visualize, in this case "what do you measure" as shown below.

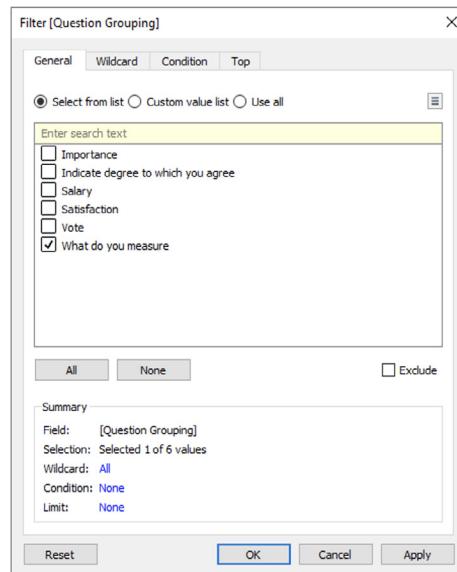


Figure 34: The metadata from the "helper" file makes it easy to select a group of related questions.

Now instead of dragging Question ID to the Rows shelf, we'll drag the Wording field.

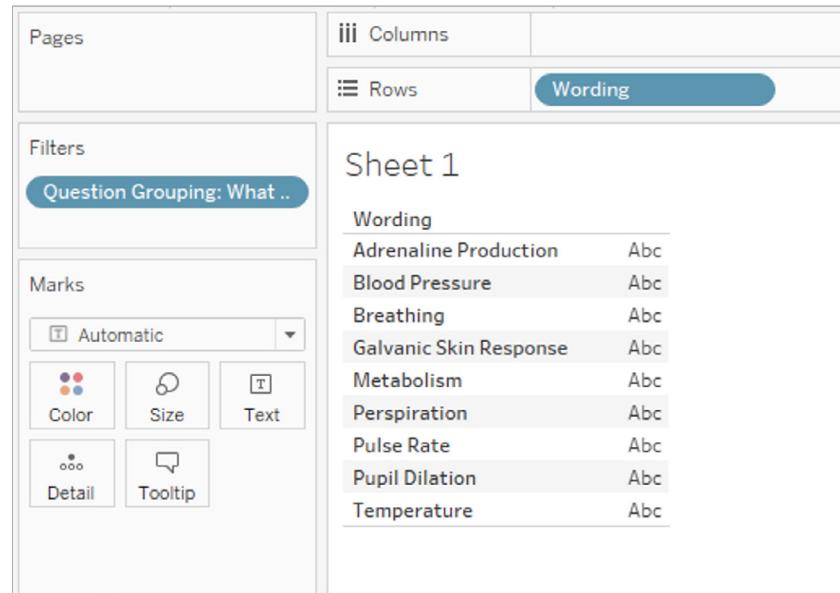


Figure 35: Now we see the questions we want to analyze in human-readable form.

Now we just need to figure out how to show the percentage of respondents that selected each item.

Computing the Percentage of People that Selected an Option

Just how the survey responses are encoded will depend on your survey tool. For this collection of questions, there are only three possibilities and their numeric values are encoded as follows:

1 (meaning yes)

0 (meaning no)

Blank (meaning they did not answer)

The algorithm for computing the percentage of people answering yes is:

Cycle through all the responses and add up everyone that selected yes (meaning they have a response of “1”). Take the total and divide by the number of people that answered the question.

Here's what the formula looks like in Tableau:

$\text{SUM}([\text{Value}]) / \text{COUNT}([\text{CustomerSat_Survey}])$ ²

²Newer versions of Tableau generate the field CustomerSat_Survey (Count) automatically. This field sums the number of records (or in this case, the number of responses) that are being counted. This field replaces [Number of Records]. Note that there are a lot of examples in the wild that refer to this now obsolete field. See <https://www.datarevelations.com/why-doesnt-this-survey-data-stuff-work/> on how to address this.

Note that if you only had text responses, the formula would look something like this:

```
SUM(  
    IF Label = "Yes" then 1  
    ELSEIF Label = "No" then 0  
    END) /  
    COUNT([CustomerSat_Survey])
```

Here's the resulting visualization, sorted from highest to lowest:

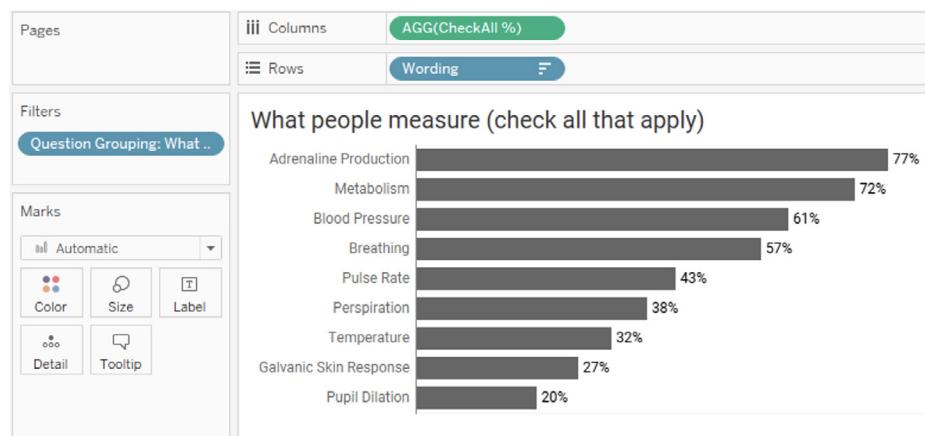


Figure 36: The resulting check-all-that apply visualization shows 77% of all respondents measure “adrenaline production” while only 20% measure “pupil dilation.”

■ Note: In preparing the data I removed all null responses. Had I not performed this task I would need to apply a filter that removes null (blank) responses.

With a little extra effort, you can also craft a dashboard that allows you to see how responses to a particular selection differ by demographics. In the dashboard below, we see a big difference between Traditionalists and Baby Boomers, particularly with respect to measuring Blood Pressure and Breathing.



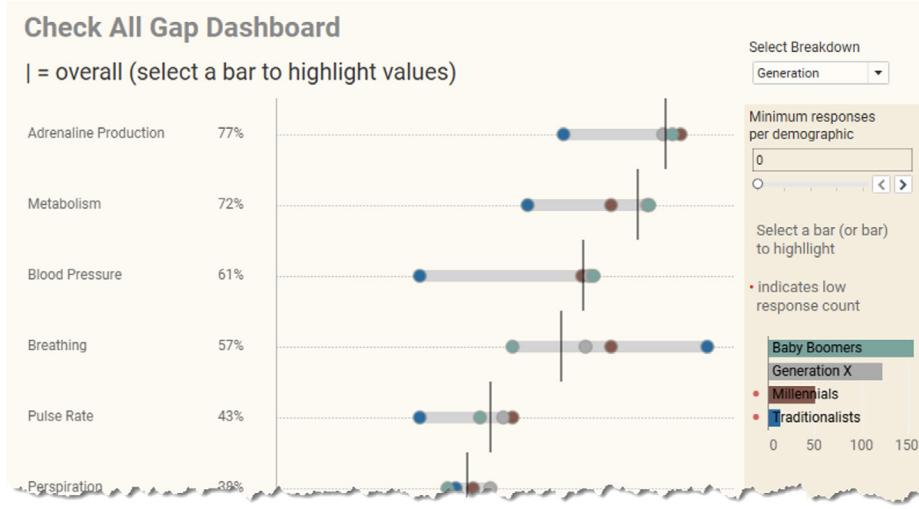


Figure 37: This dashboard shows overall responses (the vertical line) as well as responses based on a particular demographic, in this case, “Generation.”

The dashboard also shows the number of responses within the demographic (there are very few for Millennials and Traditionalists) and allows the user to easily highlight one or more demographic components, as we see below.

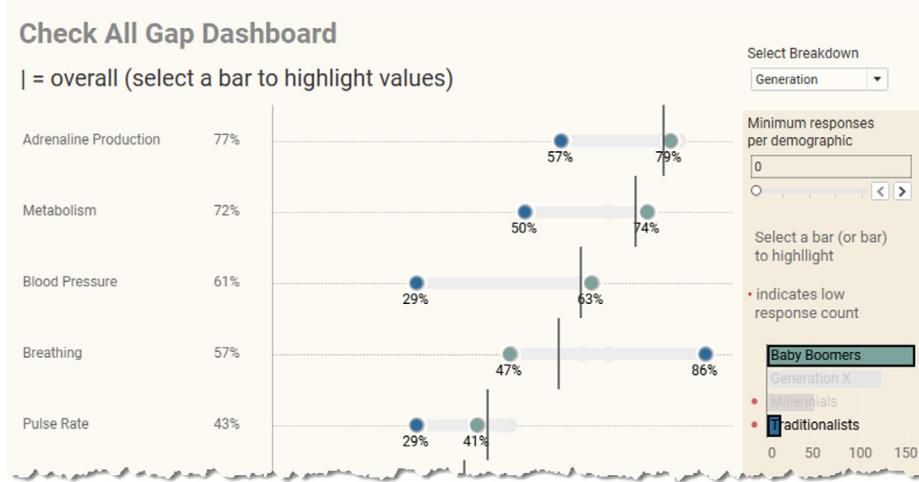
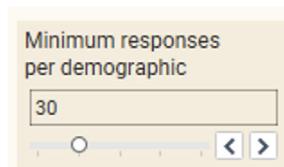


Figure 38: Interactivity allows the user to explore the gaps between Baby Boomers and Traditionalists.

Flagging Too Few Responses

Did you see the little red dots next to the bar charts in the previous two images? Did they concern you?

If the answer to both questions is “yes”, well good for you! It’s great that we have a compact and clear way to show the differences among respondents, but we don’t want to mislead our audience into thinking there are noteworthy differences when we have such a low response count for some of the segments as the margin or error will be very large.



In the previous examples we both flag the problem and also offer a way to remove items from the chart if the response count is too low.

But should this filter even be optional?

If you provide filtering capabilities in your dashboards it may be possible to apply so many filters that what are supposed to be anonymous findings are no longer anonymous. For example, let’s say you’re looking at a public HR survey dashboard and you just want to look at the salaries for instructional designers who live in Montana with over ten years’ experience. You might just get one result, and if you happen to be friends with an instructional designer who lives in Montana, you may have stumbled upon that person’s salary.

How can you address this? Consider adding a non-editable filter that hides results if there are fewer than a certain minimum number of responses.

For more information, see www.datarevelations.com/too-few-survey-responses/

Google Forms and Microsoft Forms

Both Google Forms and Microsoft Forms do a poor job of encoding check-all-that-apply data into a structure that tools like Tableau can use. If you are using either of these services for collecting survey data, please see www.datarevelations.com/google-forms-surveydata/

Building a Likert-Scale / Top-Two Boxes Visualization

In addition to the divergent stacked bar chart we saw in Figure 9 on page 11, a popular way to visualize Likert-scale data is to just focus on the percentage of respondents that selected the top 2 boxes; that is, a “4” or a “5” with a 5-point Likert scale question.

As with the previous examples, we first need to filter the question group to show the questions that are of interest. Then you need to display the filtered list, as we have here.

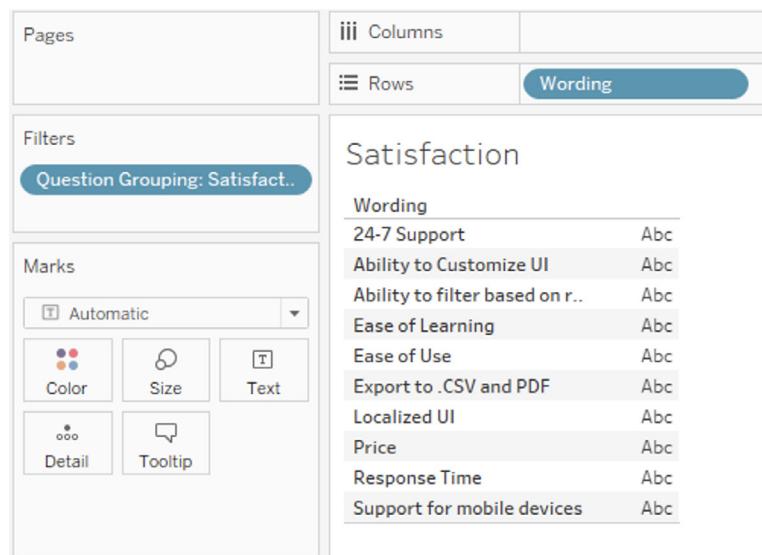


Figure 39: Each of the Likert-scale questions gets its own row.

Interestingly, if we just wanted to present a “top-two boxes” view, we could create a simple variation of the formula we employed for the check-all-that-apply question (assuming you are using a 5-point Likert scale):

```

SUM(
  If [Value]>=4 then 1 ELSE 0
)
/SUM([Number of Records])
  
```

The calculation would produce a visualization that looks like this:

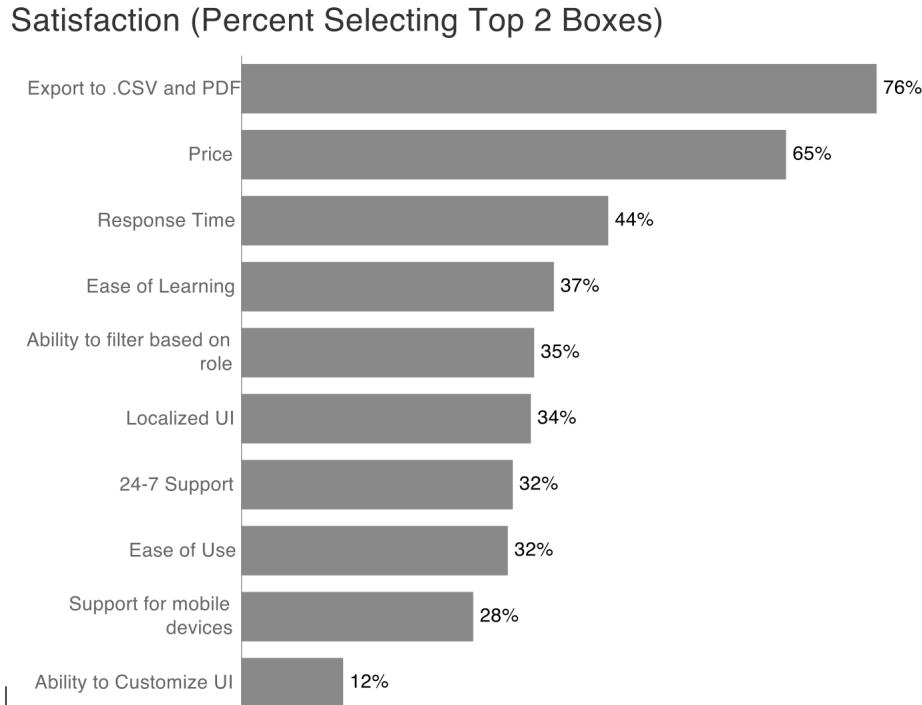


Figure 40: AHere's a top-two boxes visualization.

Whenever I share this Top 2 Boxes approach in my survey data classes I'm almost immediately hit with questions like these:

- Suppose I want to show the top ONE box only?
- Suppose I want to see the BOTTOM two boxes?
- Suppose I want to sort from lowest to highest instead of highest to lowest?
- Is there a way to apply this view to all the Likert-scale questions and not just the Satisfaction questions?

Yes, with a good visualization tool you should be able to accommodate all these requests.

Here's a dashboard that addresses most of these questions.

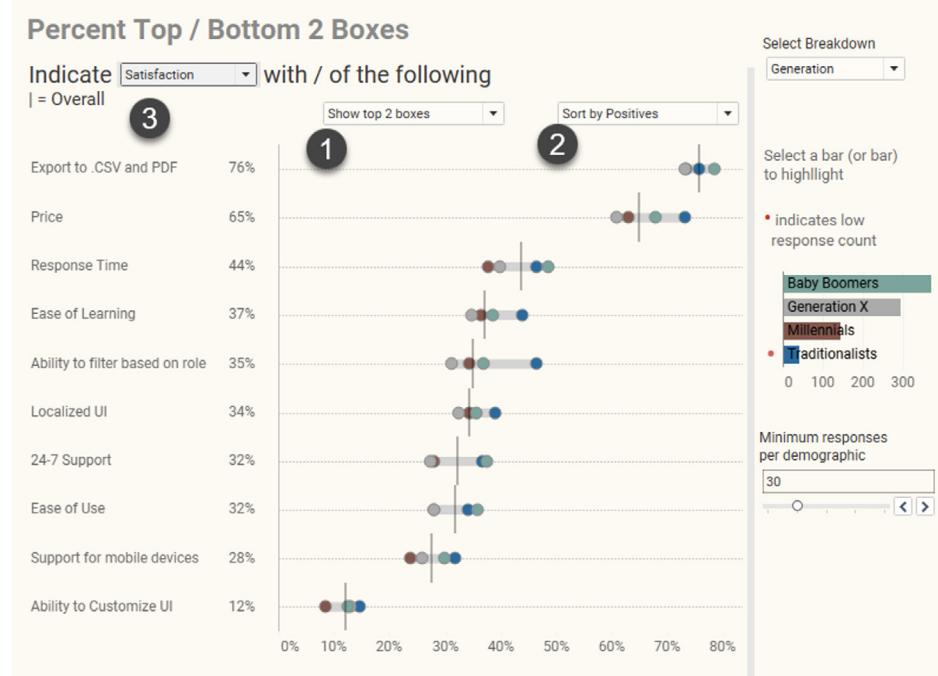


Figure 41: A connected dot plot (gap chart) embedded in a flexible dashboard that provides multiple ways to see and understand Likert-scale data.

Here you can change from Top 2 boxes to Bottom 2 boxes (1), change the sort order (2), and change to a different question group to analyze (3).

You can download this dashboard in Tableau format here: www.datarevelations.com/likert-demographics/

Conclusion

If you're going to go to the effort to build and deploy a great survey (or surveys), why rely on underpowered analytic tools and their attendant visualizations to glean and share insights?

There is so much more you can achieve, from discovering new relationships through ad hoc analysis to combining disparate data sources with blending, to crafting infinitely more interesting and insightful visualizations. With a little work to get the data set up properly and with the right tools to interrogate and visualize your data, you'll get a better, deeper understanding of your users that you'll be able to communicate to your common stakeholders.

Where to Learn More

I offer many free resources (and over 40 articles on visualizing survey data) that you can find at my web site. See bigpic.me/surveydata

I also offer a comprehensive on-demand course on visualizing survey data using Tableau where I walk through how to create demographic, single-punch, multi-punch, Likert-scale, sentiment, and benchmarking visualizations. See bigpic.me/course



About Tableau

Tableau Software helps people see and understand data. Offering a revolutionary new approach to business intelligence, Tableau allows you to quickly connect, visualize, and share data with a seamless experience from the PC to the iPad. Create and publish dashboards and share them with colleagues, partners, or customers—no programming skills required. See how Tableau can help your organization by starting your free trial at tableau.com/trial

Additional Resources

[Download Free Trial](#)

Explore Other Resources

[Product Demo](#)

[Training & Tutorials](#)

[Community & Support](#)

[Customer Stories](#)

[Solutions](#)





 + a b | e a u®

