Visualizing Survey Data

Steve Wexler

www.DataRevelations.com

swexler@DataRevelations.com

914.945.0567

Version 1.3 – June 14, 2014

About the author

Steve Wexler is the founder of [Data Revelations](), a data visualization consultancy that helps organizations better understand and share data. A Tableau Zen Master, Steve was the winner of the Strata / O'Reilly data visualization competition and the inaugural Tableau Iron Viz competition.

Steve also offers training classes in Tableau and visualizing survey data and enjoys an active following within the data visualization community with his blog posts reaching tens of thousands of unique users every year.

Steve attended Princeton University and was awarded a fellowship from the University of Miami.

Contents

Overview

While commercial survey tools allow you to build complex surveys and collect data, these tools are not nearly as robust with respect to analyzing and visualizing the collected data. All too often the best stories in the survey data remain hidden behind canned reports or reports that are too difficult or too expensive to customize.

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

The examples in this report use Tableau from Tableau Software, but the concepts and techniques can be applied to any visualization tool. With whatever tool you use, you should make sure it provides a flexible and interactive user experience, the ability to shift perspectives easily, and the ability to blend different data sources.

This whitepaper is divided into two sections:

Section	Comments	Page
The Business Case for Using Visualization Tools with Survey Data	Examples of how visualization tools can help you find the best stories in your data and examples of visualizations and dashboards that best tell these stories	2
Reshaping Data and Building Visualizations	A discussion of how survey data need to be prepared and how to create visualizations around the data	20

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. In particular, we will look at

- Better Insights Through Ad-Hoc Exploration

- Creating Meaningful Visualizations

- On-The-Fly Customization

- Intra-question Analysis

- Blending Survey Data with Third-Party Data

- Prepare Individualized Dashboard for Stakeholders and Participants

- Testing for Statistical Significance

- Interactivity

Better Insights Through Ad-Hoc Exploration

You will probably come to your survey analysis with some pre-conceived notions and theories about your data. As you vet your theories you should be conscious of three things:

1. What is the relationship / correlation among the elements you're exploring?

2. If there is a relationship, does it tell you anything useful?

3. Is the relationship statistically significant?

To be able to do this you need a tool that allows you to see quickly if there is any validity to your theories and explorations. That is, you don't want to have to spend hours generating a report only to discover what you thought would yield insights did not.

Consider the example below where we see some simplified salary survey data in Tableau and a basic visualization that shows the average salary for all respondents.
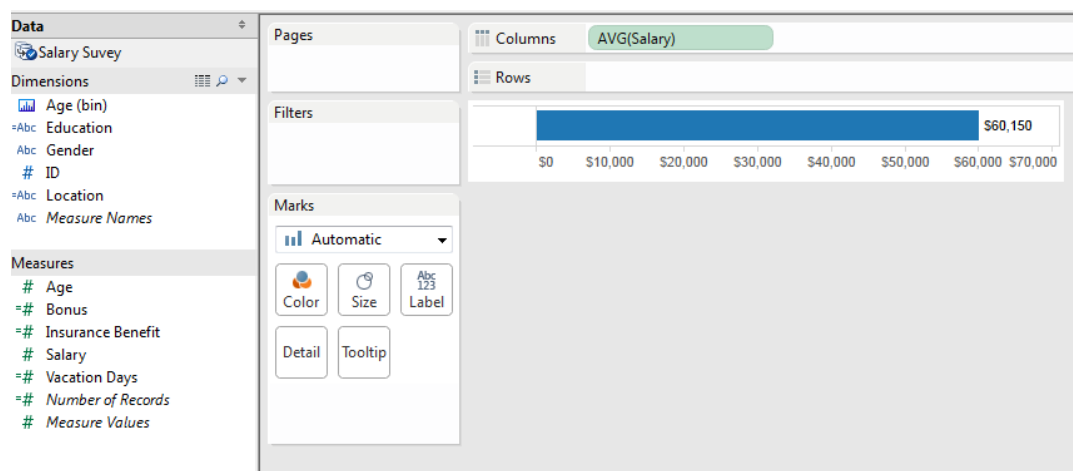
Figure 1 -- A simple bar chart that shows average salary for all respondents.

Upon seeing a chart like 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.
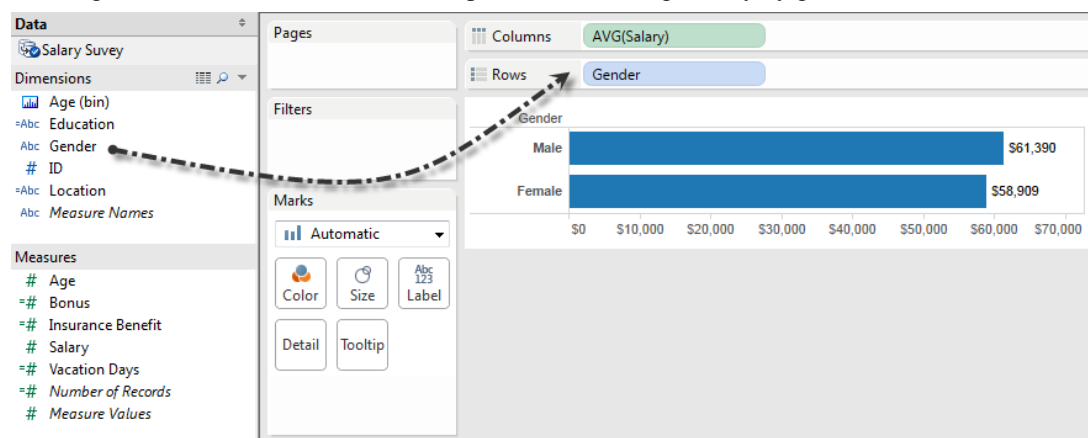

Figure 2 – By dragging Gender to the Rows Shelf we can see that men, on average, make around $2,500 more than women.

Here's what happens if instead of Gender you drag Age (bin) to the Rows Shelf.
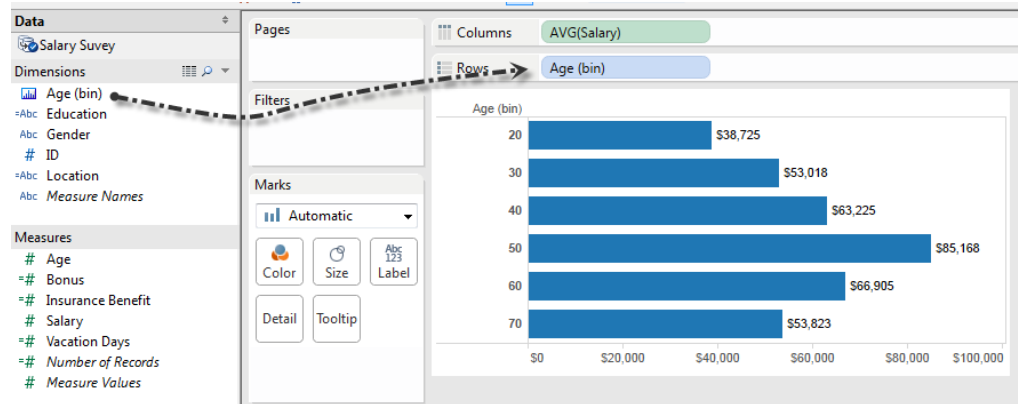
Figure 3 – Salary broken down by age bins. We can see that the average salary crests between the ages of 50 and 59.

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

Figure 4 – Salary broken down by age bins and gender.

As you experiment, you might realize that many of your explorations will not yield anything particularly insightful -- and that's fine. As long as you have a tool that allows you to see quickly if there is any value behind your theories, you should not be stymied by failure as it will not take long for you to find where the good stories reside.

Creating Meaningful Visualizations

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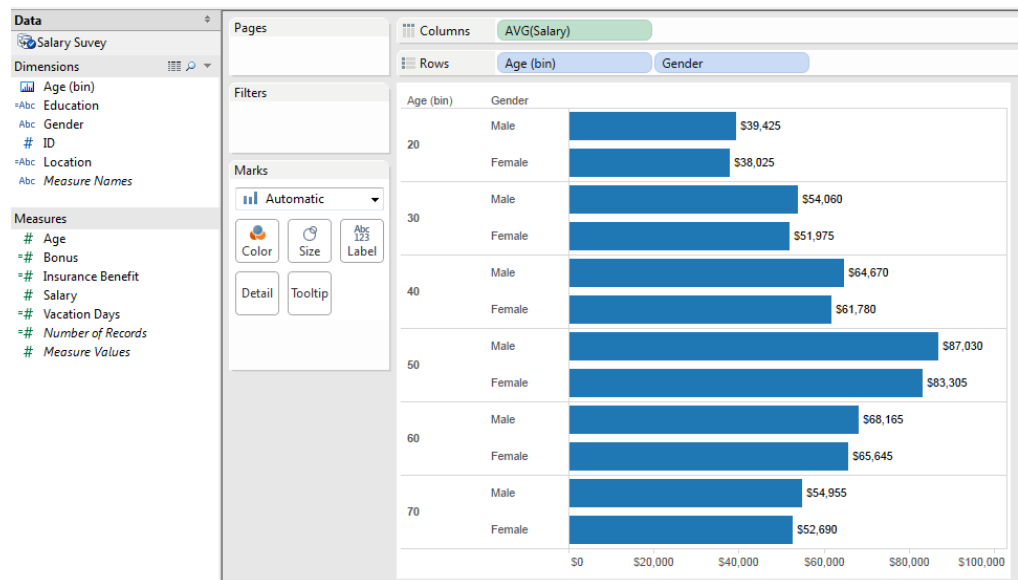
There are certain types of survey questions that traditionally have been difficult to visualize, with Likert-scale questions being at or near the top of the list. Here's a typical set of Likert-scale questions as it would appear in a survey.

Figure 5 -- A set of Likert-scale questions as they would appear to a person taking a survey

Here's an example of one attempt to visualize a collection of Likert-scale questions, in this case from a survey about learning using mobile devices.

Current level of interest in providing the following learning elements on mobile devices:
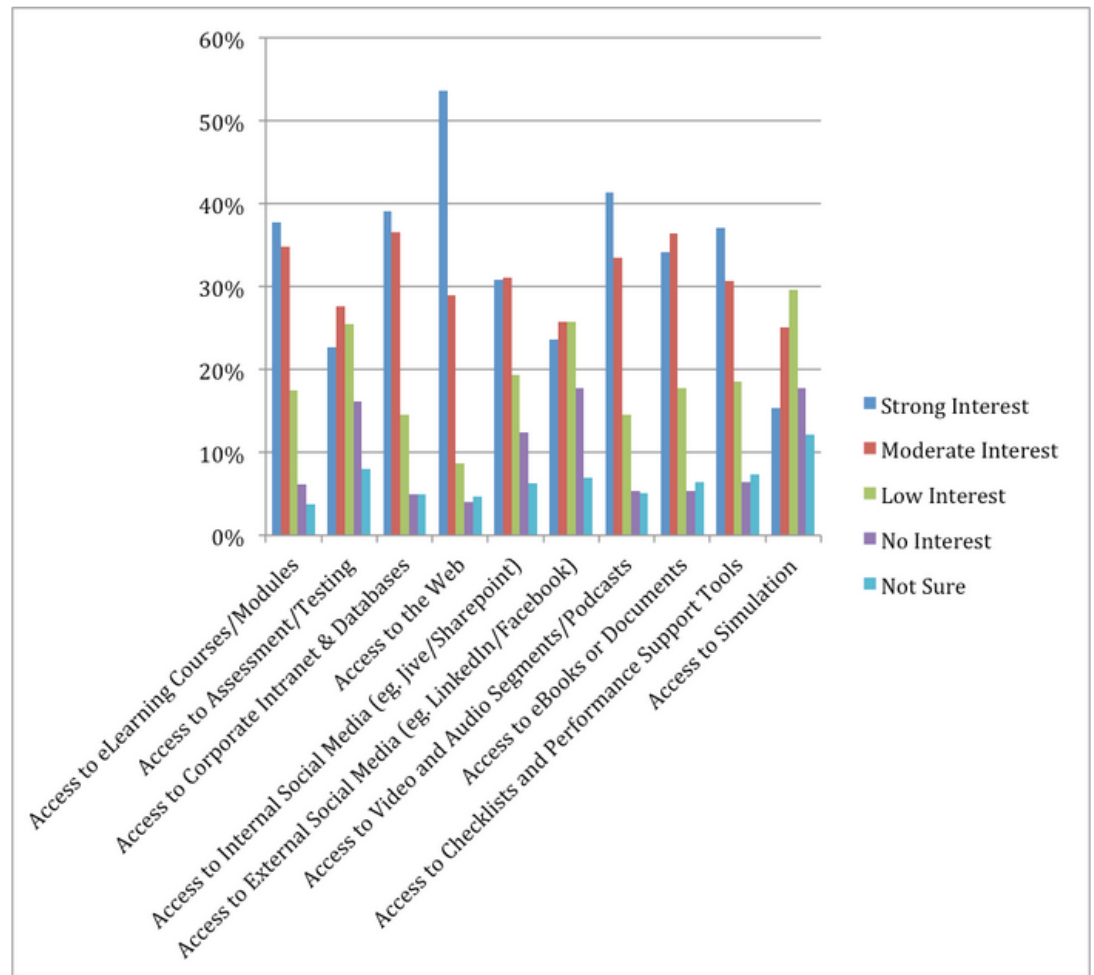
Figure 6 -- A difficult-to-interpret Excel chart from The MASIE Center's "Mobile Learning Pulse Survey", Fall 2012

Indeed, most people have so much difficulty visualizing this type of question that the survey industry has come to accept a "just show the top 2 boxes" chart as the norm for this type of question, as shown below.

Percentage Selecting Top 2 Boxes

Category	% Selecting Top 2 Boxes
Access to the Web	84%
Access to Corporate Intranet & Databases	78%
Access to Video and Audio Segments / Podcasts	76%
Access to eLearning Courses / Modules	72%
Access to Checklists and Performance Support Tools	69%
Access to eBooks or Documents	65%
Access to Internal Social Media (e.g., Jive / Sharepoint)	63%
Access to Assessment / Testing	50%
Access to External Social Media (e.g., LinkedIn / Facebook)	50%
Access to Simulations	41%

Figure 7 -- A visualization showing percent of respondents that selected the Top 2 Boxes.

While this is a good start in understanding respondent sentiments it only shows one part of the story -- the percentage of people that lean positively towards something.

Using a modified GANTT chart with Tableau we can tell the full story with respect to respondent proclivities.

Figure 8 -- A Divergent Stacked Bar Chart shows positive, negative, and neutral sentiment.

People accustomed to the "Top 2 Boxes" approach may at first resist the divergent stacked bar chart so you may want to craft a view that combines the best of both elements as shown in the dual-axis chart below.
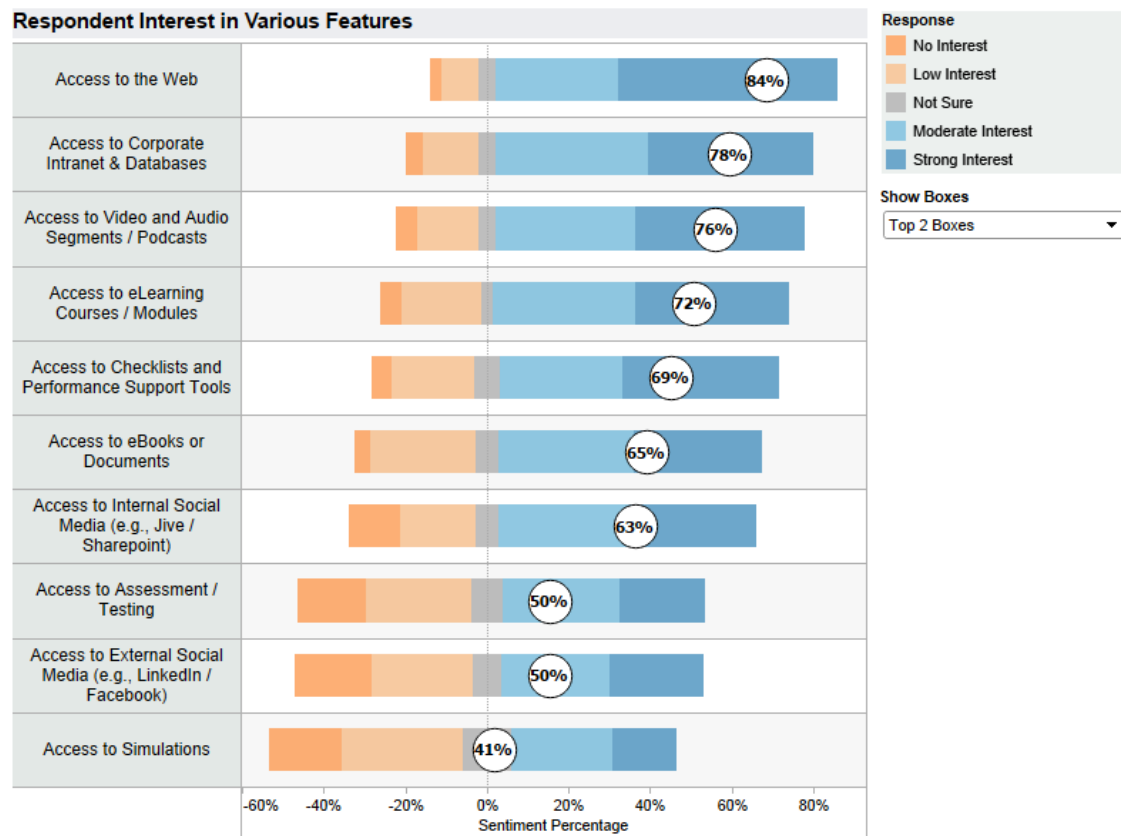
Figure 9 -- A dual axis chart that shows sentiment as a divergent stacked bar chart and percent Top 2 Boxes as a circle chart

Net Promoter Score (NPS) Analysis

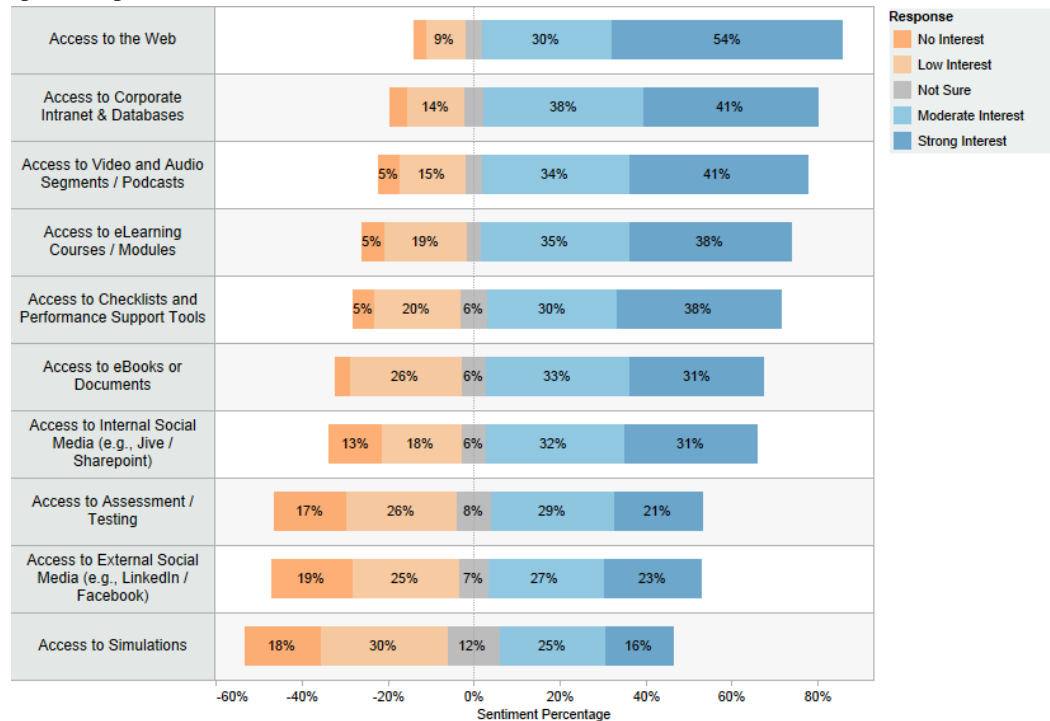A common way to gauge consumer sentiment is to ask survey respondents, using a 0 to 10 scale, to indicate whether or not they would recommend a product or service.

Respondents that provide a rating of 0 to 6 are considered "detractors" and those that give a rating of 9 or 10 are considered "promoters". The Net Promoter Score is computed by taking the percentage of promoters minus the percentage of detractors and multiplying by 100.

Here's a text-table rendering of NPS across various brands. It doesn't tell much of a story.

Brand X	40
Brand B	36
Brand Y	22
Brand Z	16
Brand A	-8
Brand D	-17
Brand Q	-20

Figure 10 -- A text table that shows Net Promoter Score (NPS)

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

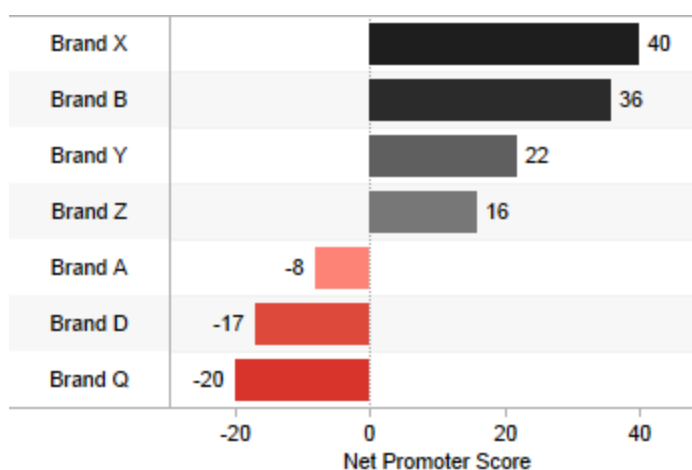*Figure 11 -- A bar chart showing NPS. This is very easy to create and does a much better job helping people see the wide gulf between the top and bottom products.*

While the bar chart is a big improvement over the text table, with a little more effort we can tell a richer story by creating a combination chart that shows both NPS and the actual promoter and detractor percentages.
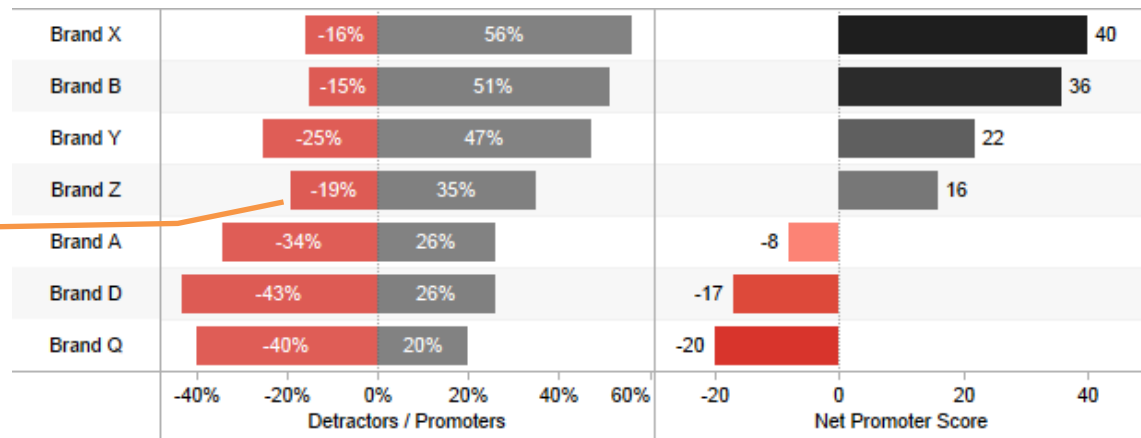
Figure 12 -- A combination chart showing Detractors, Promoters, and NPS. We can glean that Brand Z, despite a positive NPS, has many neutral respondents as the bars no not extend very wide. A major Internet / content provider and a top Human Resources vendor both use this type of visualization within their organizations.

On-The-Fly Customization

The moment you provide a visualization that shows the Top 2 Boxes someone will ask to see just the Top Box…

… or the Top 3 Boxes

… or the Bottom 2 Boxes

This is where interactivity and the ability to alter visualizations through user-controlled parameters becomes so valuable. In the Dual Axis chart on page 8 we showed the Top 2 Boxes as circles but with very little effort we can add a parameter menu that allows the user to control how many boxes (and what types of boxes) get displayed.

Figure 13 -- A parameter drop down menu that gives the user the ability to control what is displayed in the visualization

Intra-Question Analysis

As we will see in the second part of this whitepaper, the key to achieving insights into your survey data will be getting the data set up properly.

For the most part you will divide your survey responses into two broad categories

1. Dimensions you use to "cut" the data; i.e., buckets such as

    o   Gender

    o   Year born

    o   Geographic location

2. Reponses to questions that indicate opinions or intentions on certain subjects. Examples include:

    o   Will you vote in the next election?

    o   To what degree do you agree or disagree with the following statements?

    o   Indicate which of these products you own (check all that apply)

While partitioning by logical dimensions will comprise the majority of most visualizations, there will certainly be occasions where you will need to partition a set of questions by responses to another set of questions. That is, you will want to combine responses from this question…

Do you plan to vote in the upcoming election?

Yes	52%
No	34%
Don't know	13%
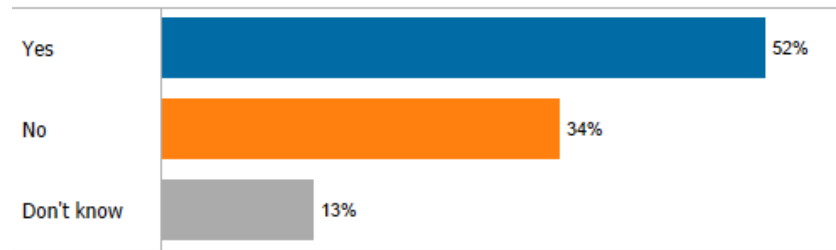
Figure 14 -- A visualization of a "yes / no / don't know" question

… with responses from this question…

Indicate which issues are important to you (check all that apply)

Economy	75%
Health Care	70%
Foreign Policy	60%
Immigration	57%
Education	41%
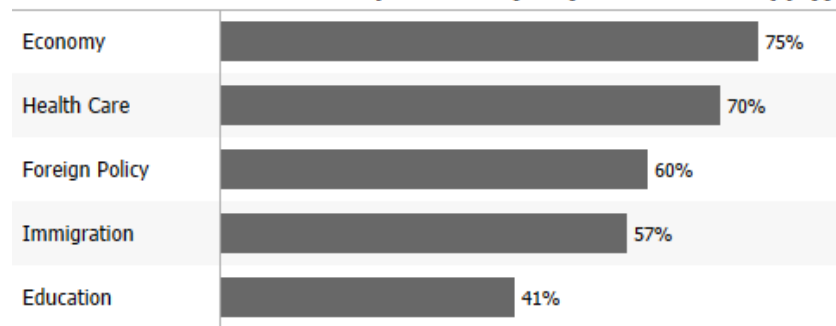
Figure 15 -- A visualization of a "check-al-that-apply" question

… to yield a product of the two questions that looks like this:

Indicate which issues are important to you (check all that apply) / plans to vote in upcoming election

Economy	Plan to vote	74%
	Don't Plan to Vote	75%
	Don't Know	83%
Health Care	Plan to vote	71%
	Don't Plan to Vote	67%
	Don't Know	72%
Foreign Policy	Plan to vote	60%
	Don't Plan to Vote	61%
	Don't Know	57%
Immigration	Plan to vote	58%
	Don't Plan to Vote	52%
	Don't Know	65%
Education	Plan to vote	42%
	Don't Plan to Vote	35%
	Don't Know	57%

Figure 16 -- A visualization of the "cuts" responses for one question by results for another question (See page 30 to understand how the percentage is computed)

There are several ways to fashion these intra-questions visualizations including joining a data table to itself, using named sets, or using actions (e.g., having the user select a bar or bars) to pass information about one visualization to another visualization.
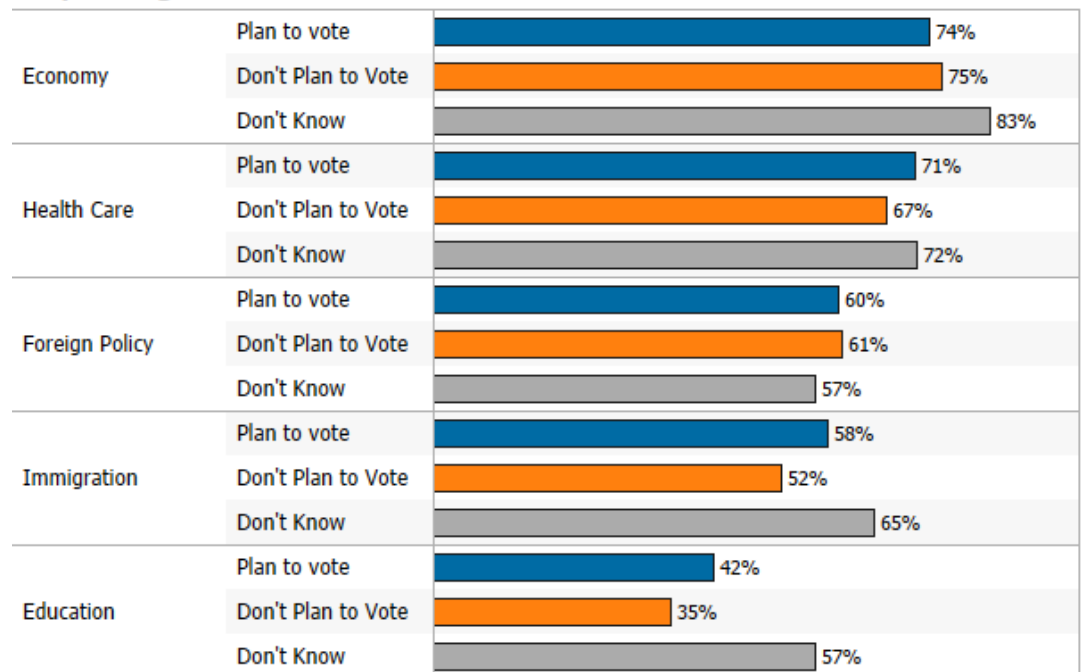
Creating Visualizations that Blend Your Survey Data with Third-Party Data

A very powerful capability that you should seek in a visualization tool is the ability to create insightful visualizations from several data sources without having to amass everything into a centralized data warehouse.

Consider the example below where we have data from three sources. The first data source comes from a public web site that shows 2012 election data for New York State, broken down by county.

	A	B	C	D	E
1	State	County	Winner	Obama	Romney
2	New York	Albany	Obama	64.49%	33.19%
3	New York	Allegany	Romney	36.21%	61.29%
4	New York	Bronx	Obama	91.45%	8.08%
5	New York	Broome	Obama	51.46%	46.15%
6	New York	Cattaraugus	Romney	42.49%	55.66%
7	New York	Cayuga	Obama	54.56%	43.16%
8	New York	Chautauqua	Romney	45.05%	52.92%
9	New York	Chemung	Romney	47.98%	50.31%
10	New York	Chenango	Romney	47.20%	50.29%
11	New York	Clinton	Obama	61.85%	36.26%
12	New York	Columbia	Obama	55.69%	41.97%
13	New York	Cortland	Obama	53.41%	44.31%
14	New York	Delaware	Romney	44.55%	53.32%
15	New York	Dutchess	Obama	52.80%	45.29%

Figure 17 -- 2012 New York State Election Results by County

Next we have survey data for people that live in New York State. Note that we do not have information about participants' political alignment but we do know the zip code in which they live.

ID	Gender	Generation	Zip Code	Q0. Vote in the upcoming election?	
4	Female	Baby Boomers	14572	0	
5	Female	Generation X	14801	1	
9	Female	Baby Boomers	10509	1	
12	Female	Baby Boomers	10512	0	
17	Female	Baby Boomers	10510	2	
25	Female	Generation X	10522	1	
26	Female	Millenials	12851	1	
36	Female	Millenials	12852	1	
37	Female	Millenials	14572		
38	Female	Baby Boomers	14801	2	
40	Female	Baby Boomers	10509	2	
42	Female	Baby Boomers	10512		
43	Female	Baby Boomers	10510		
46	Female	Millenials	10522	1	

Figure 18 -- Survey data broken down by Gender, Generation, and Zip Code

Finally, we have some "scaffolding" data from the US Postal Service that shows the relationship between Zip Code and County.

	A	B
1	Zip Code	County
767	12069	Montgomery
768	12070	Montgomery
769	12071	Schoharie
770	12072	Montgomery
771	12073	Schoharie
772	12074	Saratoga
773	12075	Columbia
774	12076	Schoharie
775	12077	Albany
776	12078	Fulton
777	12082	Rensselaer
778	12083	Greene

Figure 19 -- Zip Code and Country data from the US Postal Service

Blending Data to Answer Interesting Questions

In our survey we ask respondents if they plan to vote in the upcoming election. Suppose we would like to see results for people that live in particular counties? Better yet, suppose we

want to see results broken down by counties that voted for Romney in 2012 vs. counties that voted for Obama in 2012.

Using Tableau's ability to join and blend data from different sources we can build a dashboard like the one shown below.
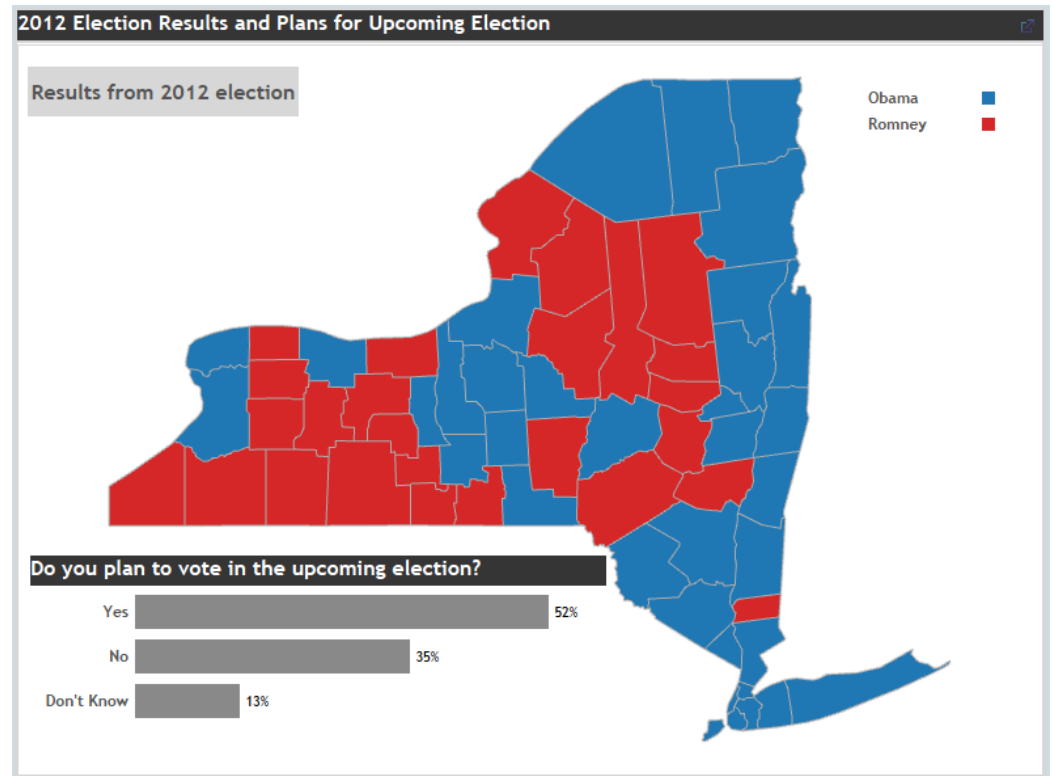

Figure 20 -- Dashboard showing a filled map with 2012 election results and a bar chart with answers to a survey question

Clicking a county will filter the bar chart to show responses from people that live in that county. Clicking either "Obama" or "Romney" in the map legend will filter the bar chart to show only results from counties that voted for the selected candidate in 2012.

Below we see results for respondents that live in counties that voted for Romney.

Selecting "Romney" here…

… filters the results in the map.

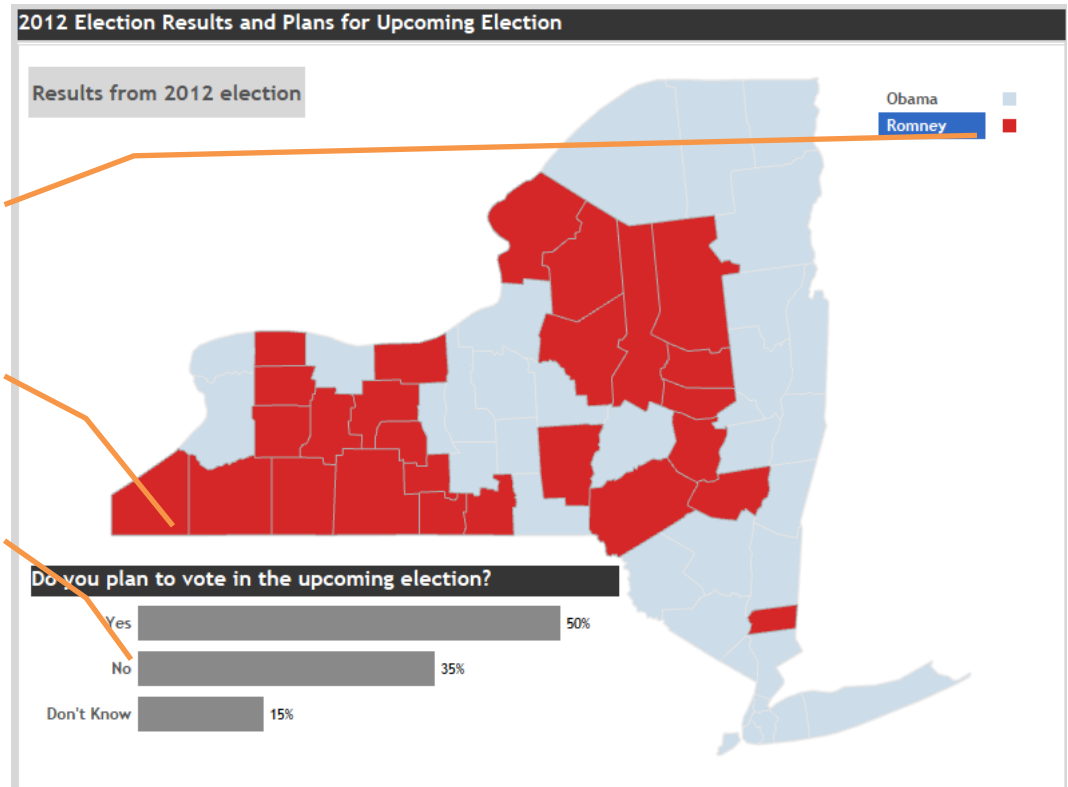*… and filters the results in the bar chart.*

Figure 21. Dashboard filtering survey results by counties that voted for Romney in 2012.

Prepare 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 is our performance this year compared with last year?

- How is our division doing compared with other divisions?

- How is my performance compared with the performance of my peers?

With a strong visualization tool it should be very easy to answer these types of questions and generate individualized reports and dashboards. Consider the example shown below that shows employee performance across six different areas. Here the manager has created a dashboard that allows her to generate a custom view for each of the 20 people that report to her.
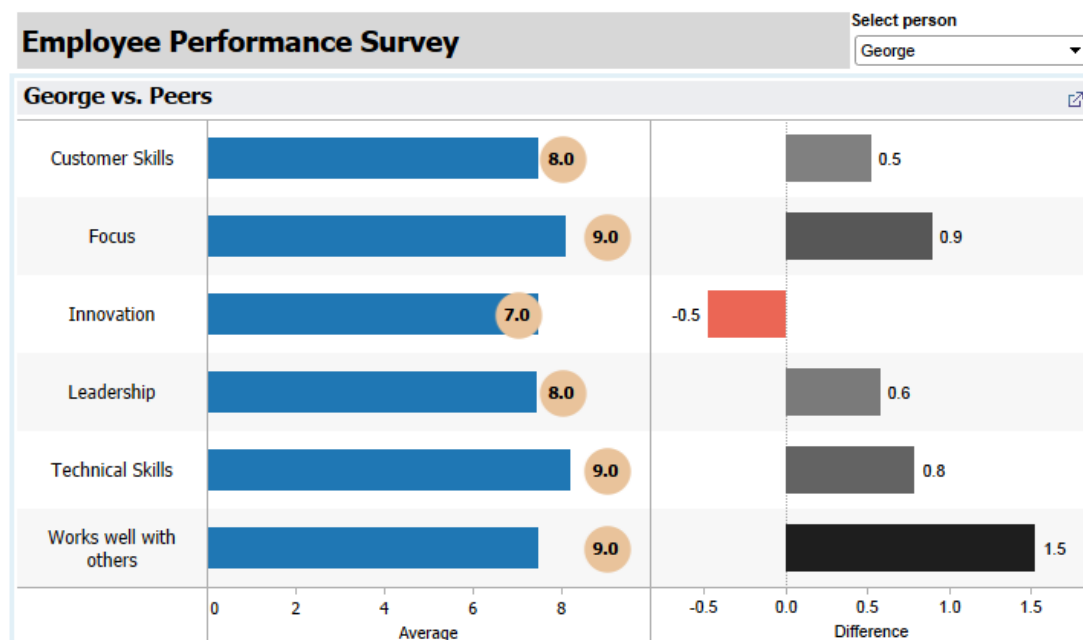
Figure 22 – A dashboard showing an individual's performance vs. that of his / her peers. The manager can select the individual from the pick list in the upper right corner.

Testing for Statistical Significance

A question that everyone who analyzes survey data should ask is whether or not the data, or results are statistically significant. For example, many people conduct longitudinal studies to see how people's opinion changes over time and it's very common to compare results from the current period with a previous period (e.g., this year vs. last year, this quarter vs. last quarter, etc.).

A common concern is that if / 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 bars show a percentage of people that respond "yes" when asked about particular attributes for a specific brand. The "positive 2" for Attribute 6 indicates that the brand is up 2 points from the previous period, but the increase is not statistically significant using a z-test formula. On the other hand, the decrease of three points for Attribute 4 *is* statistically significant.
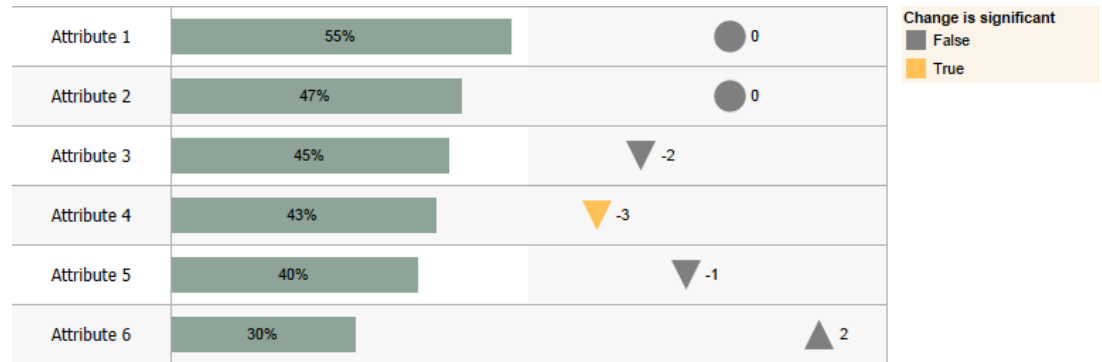
Figure 23 – A combination chart showing percent agreement for current period, increase / decrease from previous period, and whether or not the change is statistically significant.

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

Interactivity

It's great to have attractive, informative visualizations, but very often you may not be able to tell the story you want – or have the story resonate – through a static graphic that's copied into a PowerPoint presentation.

To really engage people you want to create visualizations and dashboard that are interactive.

Consider the example below where we see salary data broken down by gender and percentile.
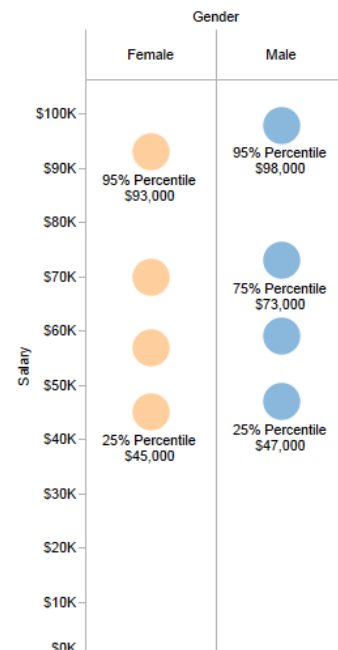

Figure 24 -- Static chart showing salary percentiles by gender.

Now let's take the same idea but create a dashboard that encourages the user to enter his / her salary and then filter the results by age and gender.
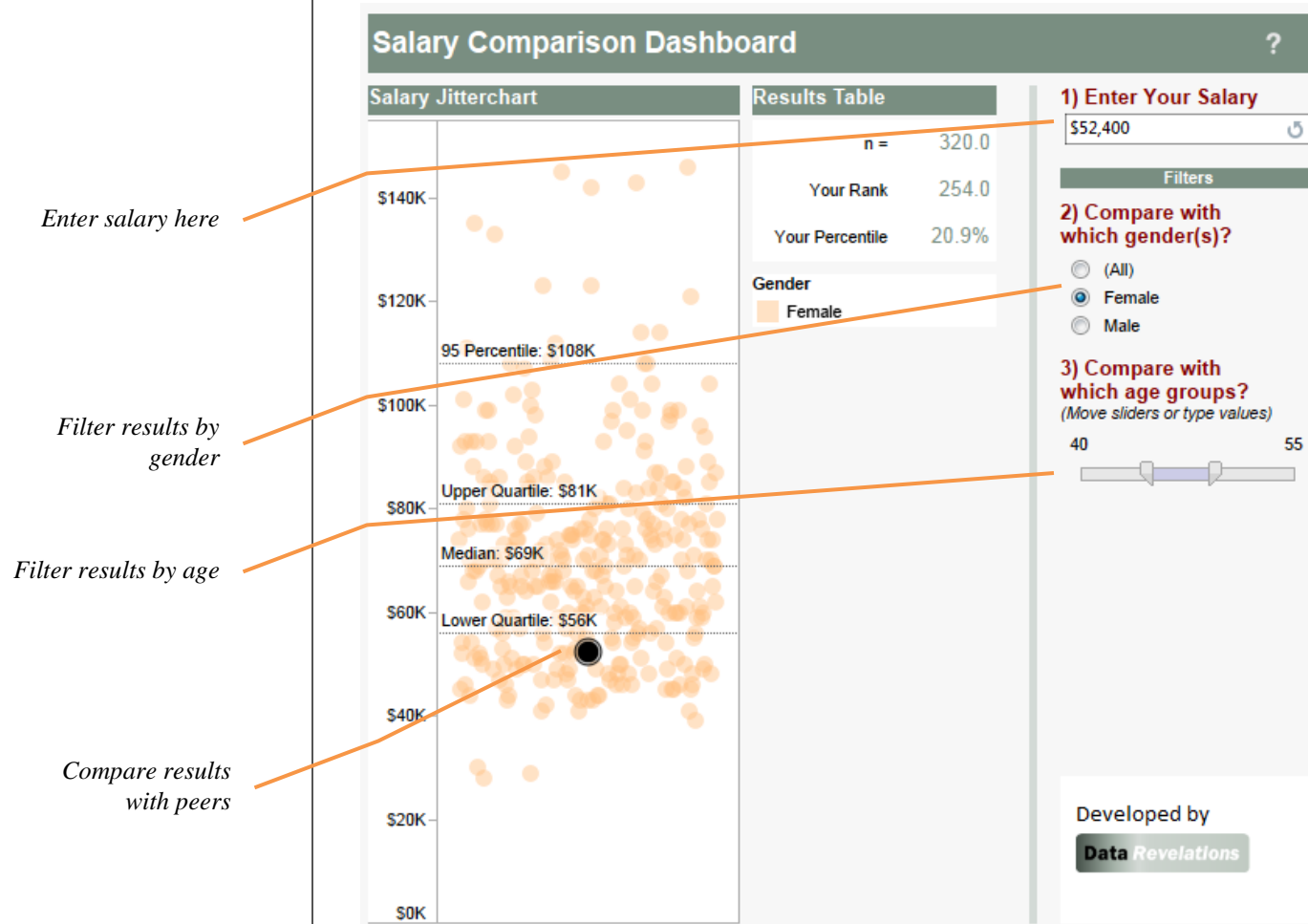

Enter salary here

Filter results by gender

Filter results by age

Compare results with peers

Figure 25 – Salary data presented in a personalized, interactive dashboard.

Reshaping Data and Building Visualizations

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 CSV file with one row per respondent and hundreds of columns, each column corresponding to a different question in the survey.

Before we discuss how the data should be arranged, let's look at some typical survey data.

Typical Survey Data

The screen shot below shows several rows and columns from a CSV file that might have come from one of several different survey tools (e.g., "Survey Monkey", "Zoomerang", etc.)

	A	B	C	D	E Q0. Vote in the upcoming election?	F Q1. Pulse Rate	G Q2. Metabolism	H Q3.Blood Pressure	I Q4. Temperature	J Q5. Galvanic Skin Response
1	ID	Gender	Country	Generation						
2	2	Male	South America	Generation X	0: No	0: No	0: No	1: Yes	0: No	1: Yes
3	4	Female	South America	Baby Boomers	0: No	1: Yes	1: Yes	1: Yes	0: No	0: No
4	5	Female	South America	Generation X	1: Yes	0: No	1: Yes	1: Yes	1: Yes	0: No
5	6	Male	Transportation	Baby Boomers	2: Don't know	0: No	1: Yes	1: Yes	0: No	0: No
6	9	Female	Europe	Baby Boomers	1: Yes	1: Yes	1: Yes	1: Yes	1: Yes	0: No
7	12	Female	Europe	Baby Boomers	0: No	0: No	0: No	1: Yes	1: Yes	0: No
8	15	Male	North America	Baby Boomers						
9	16	Male	Transportation	Baby Boomers	1: Yes	1: Yes	1: Yes	0: No	0: No	0: No
10	17	Female	Europe	Baby Boomers	2: Don't know	1: Yes	1: Yes	1: Yes	0: No	0: No
11	18	Male	North America	Traditionalists	0: No	0: No	1: Yes	0: No	0: No	0: No
12	22	Male	South America	Generation X	0: No	0: No	0: No	1: Yes	1: Yes	0: No
13	25	Female	South America	Generation X	1: Yes	1: Yes	1: Yes	1: Yes	1: Yes	1: Yes

Figure 26 -- Raw survey data in Excel

"Demographic" Questions

As we discussed earlier your survey questions will probably fall into two broad categories with the "demographic" questions providing logical buckets into which you can group responses to questions that pertain with the subject of your surveys.

In the screenshot above Columns B through D contain what one might call "demographic" questions. In this example we've just gathered gender, country of origin, and generation, but in your surveys you may see columns for ethnicity, education, marriage status, income, etc.

Yes / No / Maybe question

The data in column E, below…

E
Q0. Vote in the upcoming election?
0: No
0: No
1: Yes
2: Don't know
1: Yes
0: No
1: Yes
2: Don't know
0: No

Figure 27 -- Simple Yes/No/Maybe question

… represents the data that was gathered from this survey question.

1. Will you vote in the upcoming election?

- Yes
- No
- Don't know

Figure 28 -- Yes / No / Maybe question as it appears to the survey taker

"Check All That Apply" Questions

The data in columns F through N shown below…

F	G	H	I	J	K	L	M	N
Q1. Pulse Rate	Q2. Metabolism	Q3.Blood Pressure	Q4. Temperature	Q5. Galvanic Skin Response	Q6. Breathing	Q7. Perspiration	Q8.Pupil Dilation	Q9. Adrenaline Production
0: No	0: No	1: Yes	0: No	1: Yes	0: No	0: No	0: No	1: Yes
1: Yes	1: Yes	1: Yes	0: No	0: No	1: Yes	1: Yes	0: No	1: Yes
0: No	1: Yes	1: Yes	1: Yes	0: No	0: No	0: No	0: No	1: Yes
0: No	1: Yes	1: Yes	0: No	0: No	0: No	0: No	0: No	1: Yes
1: Yes	1: Yes	1: Yes	1: Yes	0: No	1: Yes	1: Yes	1: Yes	0: No
0: No	0: No	1: Yes	1: Yes	0: No	0: No	0: No	0: No	1: Yes
1: Yes	1: Yes	0: No	0: No	0: No	1: Yes	0: No	0: No	1: Yes
1: Yes	1: Yes	1: Yes	0: No	0: No	0: No	0: No	0: No	1: Yes
0: No	1: Yes	0: No	0: No	0: No	0: No	0: No	0: No	1: Yes
0: No	0: No	1: Yes	1: Yes	0: No	1: Yes	0: No	0: No	0: No
1: Yes	1: Yes	1: Yes	1: Yes	1: Yes	1: Yes	1: Yes	0: No	1: Yes
0: No	1: Yes	0: No	0: No	0: No	1: Yes	0: No	0: No	1: Yes

Figure 29 -- Check-all-that-apply question data

… represents the data that was gathered from this "check all that apply" survey question.

1. Please indicate all the things you measure

- ☑ Pulse rate
- ☑ Metabolism
- ☑ Blood pressure
- ☑ Temperature
- ☐ Galvanic skin response
- ☐ Breathing
- ☐ Perspiration
- ☑ Pupil dilation
- ☐ Adrenaline
- ☐ Other

Figure 30 – Check-all-that-apply questions as they appear to the survey taker

Notice that the only options for these columns are "1: Yes" and "0: No". A blank indicates that the survey respondent did not answer the question.

Likert Scale Questions

Columns O through Z …

Q134. Good Job Skills	Q135. Sense of Humor	Q136. Intelligence	Q137.Can Play Jazz	Q138.Likes the Beatles	Q139. Snobbiness	Q140.Ability to lift heavy objects	Q141.Grace under pressure	Q142.Grace dance
1: Small degree	1: Small degree	0: Not at all	1: Small degree	2: Moderate degree	1: Small degree	1: Small degree	3: High degree	2: Moderate
4: Very high degree	4: Very high degree	0: Not at all	4: Very high degree	4: Very high degree	1: Small degree	4: Very high degree	4: Very high degree	4: Very high
4: Very high degree	3: High degree	2: Moderate degree	4: Very high degree	3: High degree	3: High degree	3: High degree	4: Very high degree	4: Very high
4: Very high degree	3: High degree	1: Small degree	2: Moderate degree	3: High degree	2: Moderate degree	2: Moderate degree	3: High degree	3: High degree
4: Very high degree	4: Very high degree	2: Moderate degree	4: Very high degree	4: Very high degree	3: High degree	4: Very high degree	4: Very high degree	4: Very high
2: Moderate degree	2: Moderate degree	2: Moderate degree	3: High degree	3: High degree	4: Very high degree	3: High degree	3: High degree	4: Very high
2: Moderate degree	4: Very high degree	3: High degree	4: Very high degree	3: High degree	3: High degree	3: High degree	3: High degree	3: High degree
3: High degree	3: High degree	3: High degree	3: High degree	3: High degree	3: High degree	3: High degree	3: High degree	3: High degree
2: Moderate degree	3: High degree	3: High degree	4: Very high degree	2: Moderate degree	3: High degree	3: High degree	3: High degree	3: High degree
3: High degree	3: High degree	1: Small degree	3: High degree	3: High degree	3: High degree	3: High degree	3: High degree	3: High degree
1: Small degree	3: High degree	1: Small degree	3: High degree	2: Moderate degree	4: Very high degree	2: Moderate degree	4: Very high degree	3: High degree
3: High degree	3: High degree	2: Moderate	3: High degree	3: High degree	2: Moderate	3: High degree	3: High degree	2: Moderate

Figure 31 -- Likert scale question data

… contain Likert Scale data that was gathered from the following survey question.

1. Indicate the degree to which you seek the following abilities when making a new hire

	Not at all	Small degree	Moderate degree	High degree	Very high degree
Good job skills	◡	◡	◡	◡	✓
Sense of humor	◡	◡	◡	✓	◡
Intelligence	◡	✓	◡	◡	◡
Can play jazz	◡	◡	◡	◡	◡
Likes the Beatles	◡	◡	◡	◡	◡
Snobbishness	◡	◡	◡	◡	◡
Ability to lift heavy objects	◡	◡	◡	◡	◡
Grace under pressure	◡	◡	◡	◡	◡
Grace on the dance floor	◡	◡	◡	◡	◡
Likes animals	◡	◡	◡	◡	◡
Makes good coffee	◡	◡	◡	◡	◡
Eats all his / her vegetables	◡	◡	◡	◡	◡

Figure 32 – Likert scale questions as they appear to the survey taker

Here each response receives both a label "Very high degree", "High degree", etc., and a number rating from 1 to 5. It was Rensis Likert (pronounced "Lick-ert") that developed the idea

of assigning quantitative values to qualitative measures. If you're creating surveys, chances are you will be doing a lot of work with Likert scale questions.

Reshaped Data

While the raw data may be easy to manipulate in Excel you will have tremendous difficulty using the data in its raw form with a tool such as Tableau. To really get the most from you visualization tool you need to take the raw data that we saw in Figure 26 and convert it into something that looks like this:

	A	B	C	D	E	F
1	ID	Gender	Location	Generation	Question	Response
2	2	Male	South America	Generation X	Q0. Vote in the upcoming election?	0: No
3	2	Male	South America	Generation X	Q1. Pulse Rate	0: No
4	2	Male	South America	Generation X	Q2. Metabolism	0: No
5	2	Male	South America	Generation X	Q3.Blood Pressure	1: Yes
6	2	Male	South America	Generation X	Q4. Temperature	0: No
7	2	Male	South America	Generation X	Q5. Galvanic Skin Response	1: Yes
8	2	Male	South America	Generation X	Q6. Breathing	0: No
9	2	Male	South America	Generation X	Q7. Perspiration	0: No
10	2	Male	South America	Generation X	Q8.Pupil Dilation	0: No
11	2	Male	South America	Generation X	Q9. Adrenaline Production	1: Yes
12	2	Male	South America	Generation X	Q134. Good Job Skills	1: Small degree
13	2	Male	South America	Generation X	Q135. Sense of Humor	1: Small degree
14	2	Male	South America	Generation X	Q136. Intelligence	0: Not at all
15	2	Male	South America	Generation X	Q137.Can Play Jazz	1: Small degree
16	2	Male	South America	Generation X	Q138.Likes the Beatles	2: Moderate degree
17	2	Male	South America	Generation X	Q139. Snobbiness	1: Small degree
18	2	Male	South America	Generation X	Q140.Ability to lift heavy objects	1: Small degree
19	2	Male	South America	Generation X	Q141.Grace under pressure	3: High degree
20	2	Male	South America	Generation X	Q142.Grace on the dance floor	2: Moderate degree
21	2	Male	South America	Generation X	Q143.Likes animals	1: Small degree
22	2	Male	South America	Generation X	Q144.Makes good coffee	4: Very high degree
23	2	Male	South America	Generation X	Q145.Eats all his/her vegetables	1: Small degree
24	4	Female	South America	Baby Boomers	Q0. Vote in the upcoming election?	0: No
25	4	Female	South America	Baby Boomers	Q1. Pulse Rate	1: Yes
26	4	Female	South America	Baby Boomers	Q2. Metabolism	1: Yes
27	4	Female	South America	Baby Boomers	Q3.Blood Pressure	1: Yes
28	4	Female	South America	Baby Boomers	Q4. Temperature	0: No

Figure 33 -- Reshaped or "normalized" data

Notice that there is a separate row for each question in the survey and that the ID, Gender, Country, and Generation data is repeated for each survey respondent. This is what is called "normalized" data and it is a good thing.

There are a variety of tools you can use to turn the raw survey data into ready-to-visualize data. In this example we used the free Tableau add-in for Microsoft Excel but there are many ETL (extract, transform, and load) tools available. Indeed, if your data resides in a database server or if your data requires more complex transformation you may want to consider a more powerful tool such as Alteryx Designer.

Building Visualizations

In the next series of examples we will see how to visualize the three types of questions we reviewed in the previous section. While the example will use Tableau the concepts can be applied to other survey tools.

So, Just Who Are These People?

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.
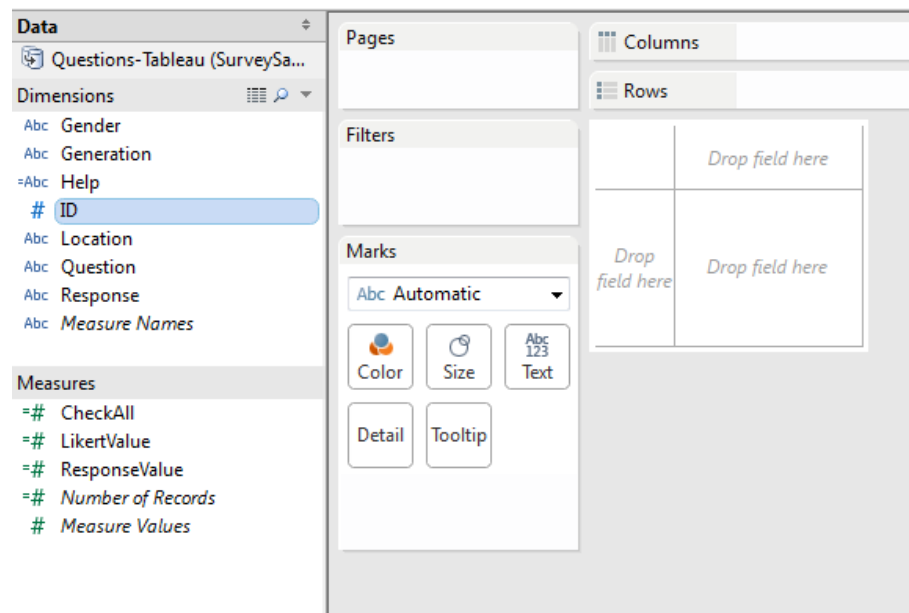
Here's how the reshaped data appears in Tableau.

Figure 34 -- 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 IDs there are in the data. You may recall that when we reshaped the data we re-peated the ID for every question response, so we cannot perform a simple count.

In Tableau we produce a distinct count using the COUNTD() function as shown here.
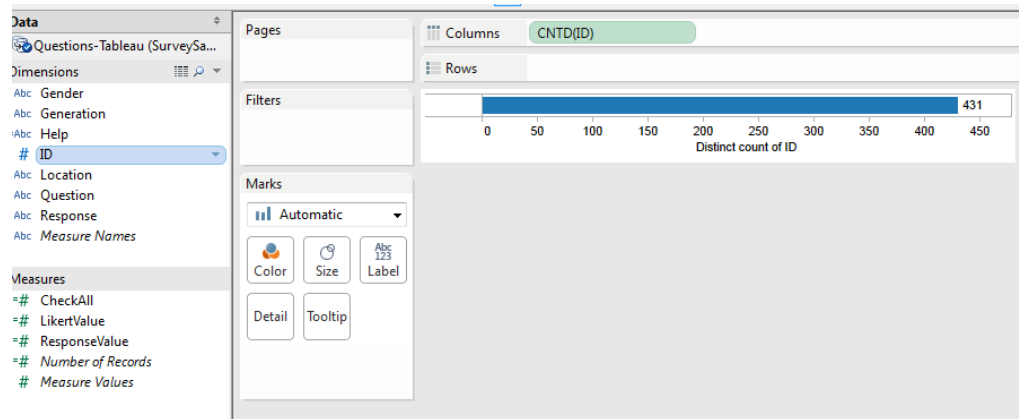
Figure 35 -- Using COUNTD() we see there were 431 survey responses

If we want to see a breakdown by gender we can simply drag Gender to the Rows shelf, as shown here.
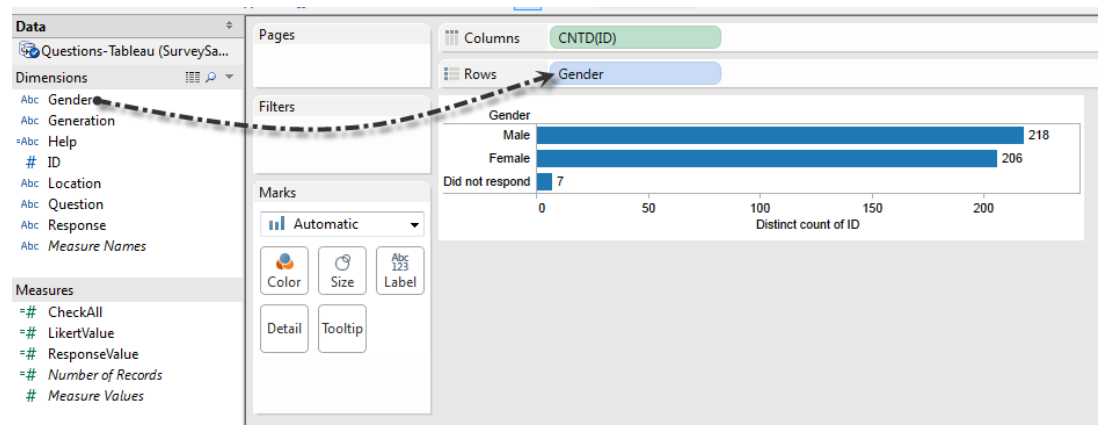
Figure 36 -- Respondent count by gender

You can repeat this exercise for all of the categories (dimensions) that interest you.

Building a Yes / No / Maybe Visualization

All of the survey questions are contained in the one "uber" category called "Question". As we don't want to see responses to every question at the same time we need a way to filter the Question field so that we are only working with the questions that interest us.

In Tableau you do this by dragging the Question field to the Filters shelf and then selecting the question (or questions) that interest you.
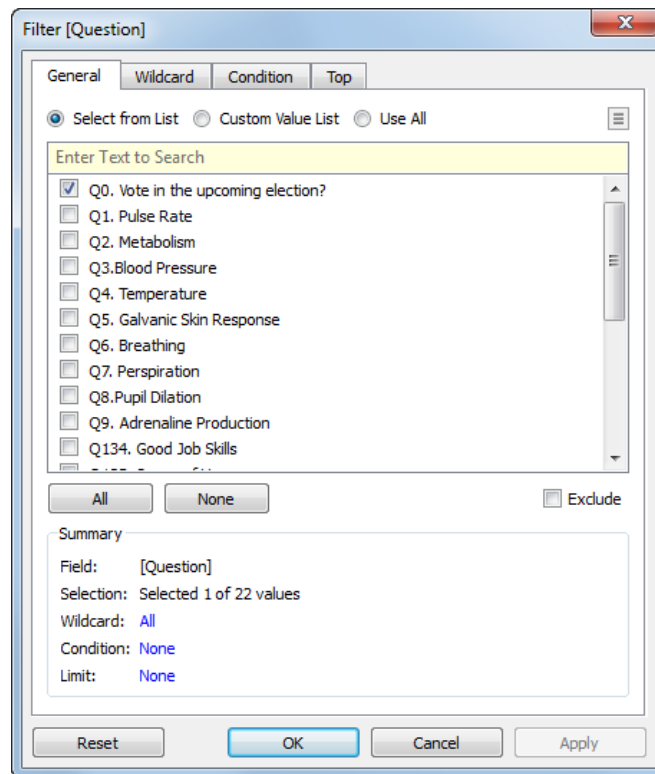
Figure 37 – We only want to see results for the first question in the list.

Now that we have told our visualization tool that we only want to see responses to one question, we just need to show the three possible responses and their attendant response count.
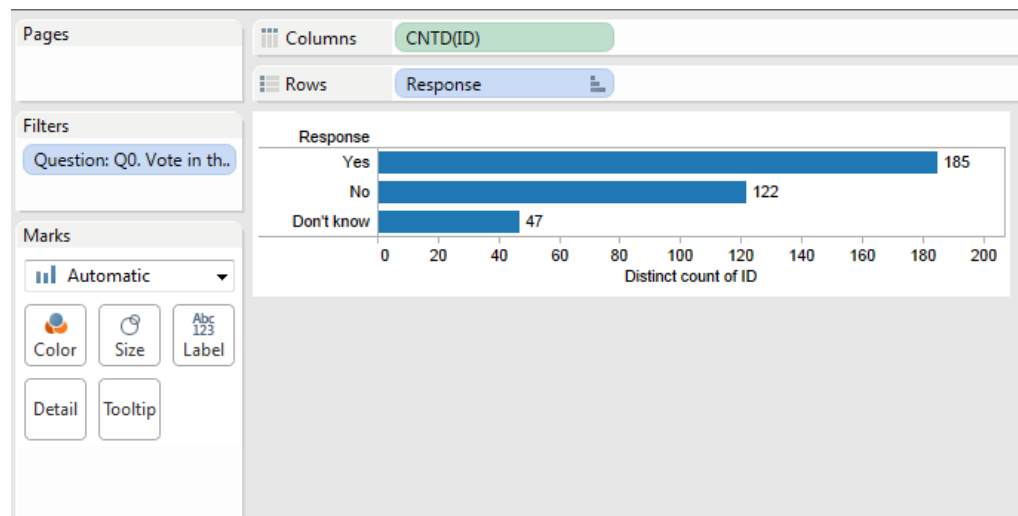
Figure 38 -- The Yes / No / Maybe data as a bar chart.

Depending on what visualization tool you use you may want to compute (and display) the results as a percentage of the total.
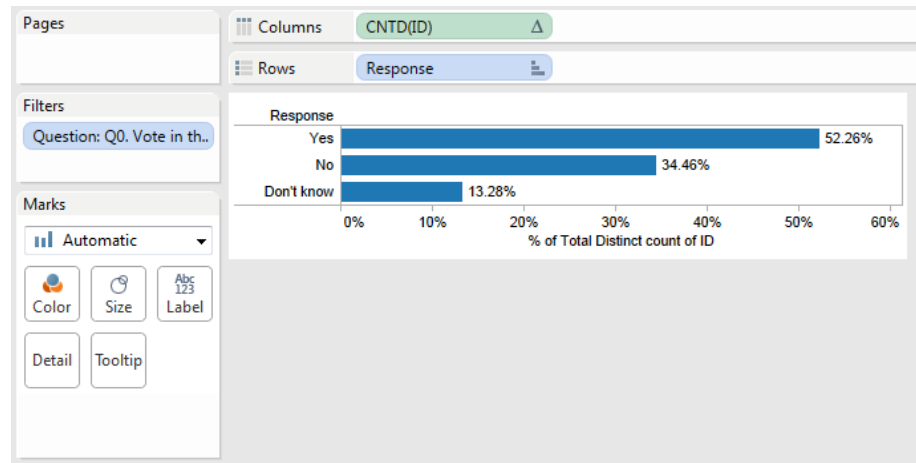
Figure 39 -- Responses as a percentage of total

Building a Check-All-That-Apply Visualization

As with the single "Yes / No / Maybe" visualization we first need to filter our data so that we only show results for the questions that interest us.
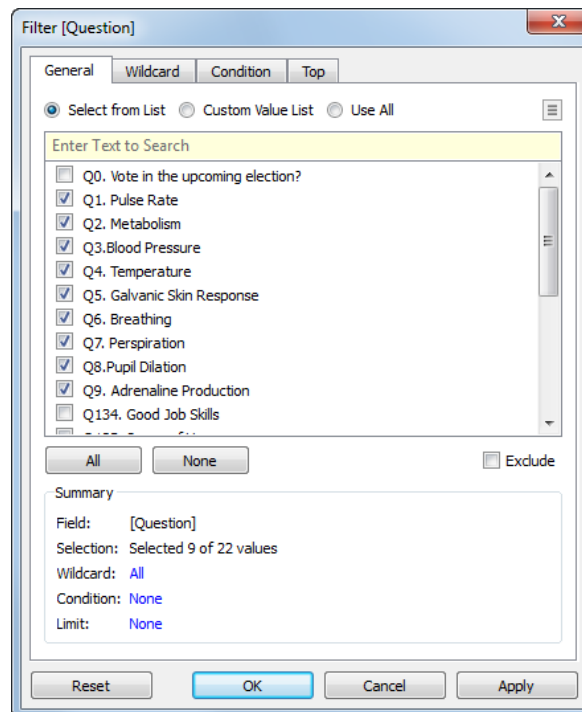
Figure 40 -- Using Tableau's filter dialog box so that we only see results for the Check-All-That Apply questions

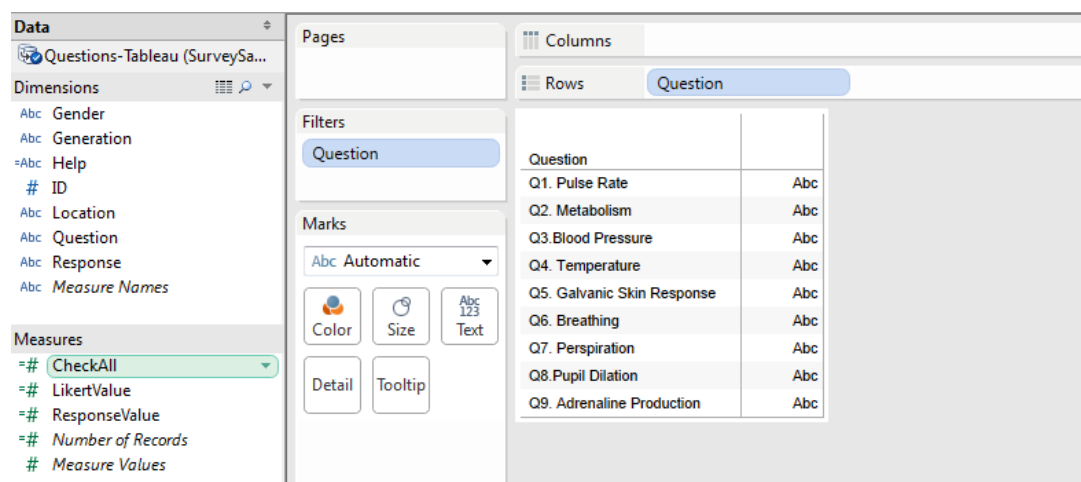We also need the question wording to be visible for each question so in Tableau we can just drag the Question field onto the Rows shelf.

Figure 41 -- A separate row for each of the filtered questions. We just need to figure out how the get the results.

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 they are encoded as follows:

- 1: Yes

- 0: No

- Blank (did not answer)

The algorithm for computing the percentage of people answering yes is

> *Cycle through all the responses and see if a respondent answered "yes" or "no". For each person that answered yes, add "1". Take the total and divide by the number of people that answered the question. Do not include people that left the question blank.*

Here's what the formula looks like in Tableau.

```
SUM(
    IIF (Left([Response],1)="1" ,1,0)
    )
/COUNTD([ID])
```

Here's the resulting visualization, sorted from highest to lowest.
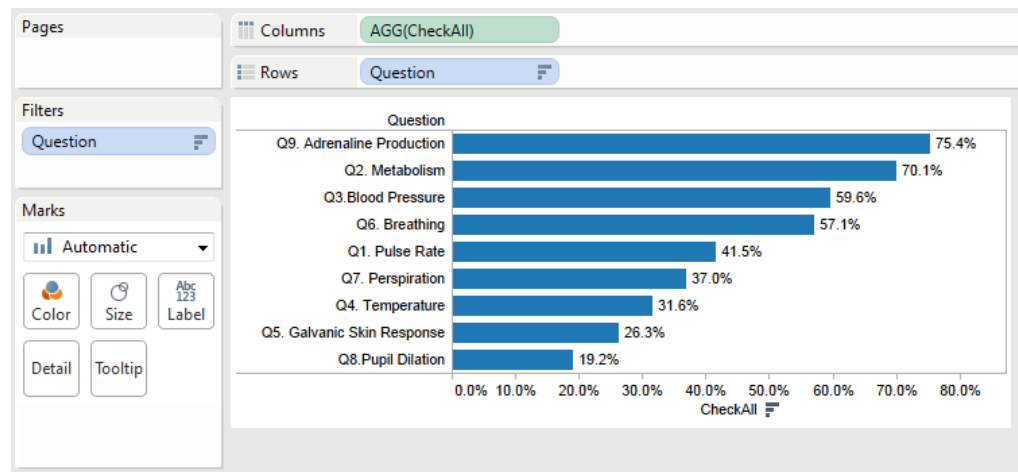
Figure 42. The resulting check-all-that apply visualization

Building a Likert Scale Question Visualization

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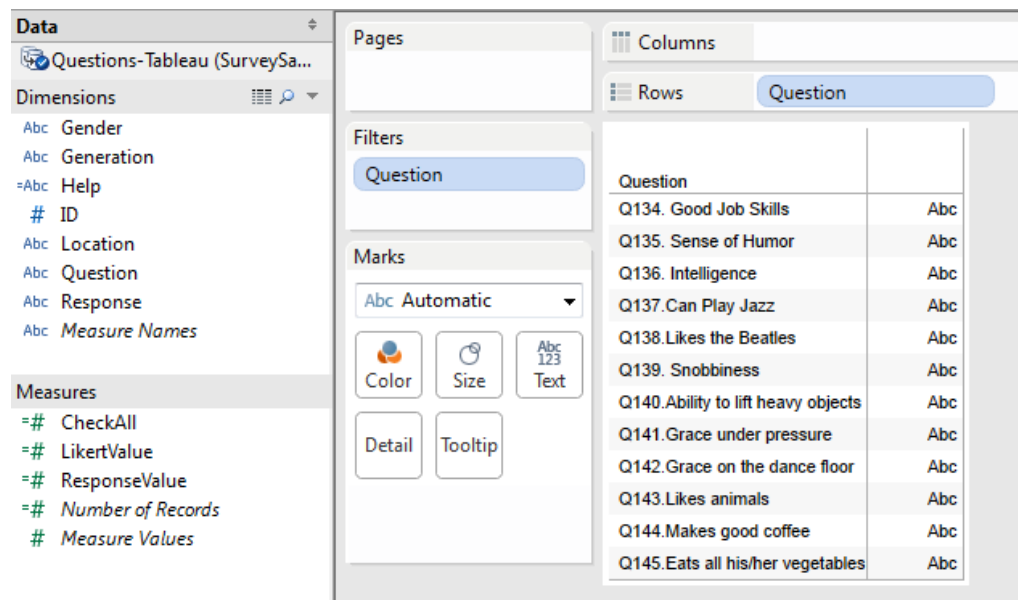

Figure 43 -- A separate row for each of the Likert scale questions

Interestingly, if we just wanted to present a "Top 2 Boxes" view we could create a simple variation of the formula we employed for the check-all-that-apply question:

```
SUM(

    If Left([Response],1)="1"  or Left([Response],1)="2"

then 1

    else 0
```

```
    end

    )

  /COUNTD([ID])
```

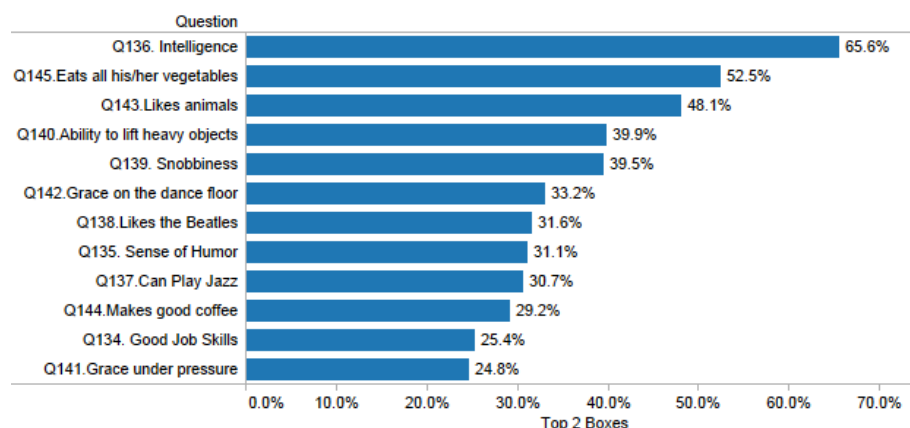This would produce a visualization that looks like this.

Figure 44 -- A Top 2 Boxes visualization

As we discussed earlier you may want to create visualizations that tell more of the story. One chart type that often proves useful is a stacked bar chart.

You create a stacked bar chart in Tableau by counting the total number of responses for each question and then creating a colored bar for each of the possible responses, as shown here.
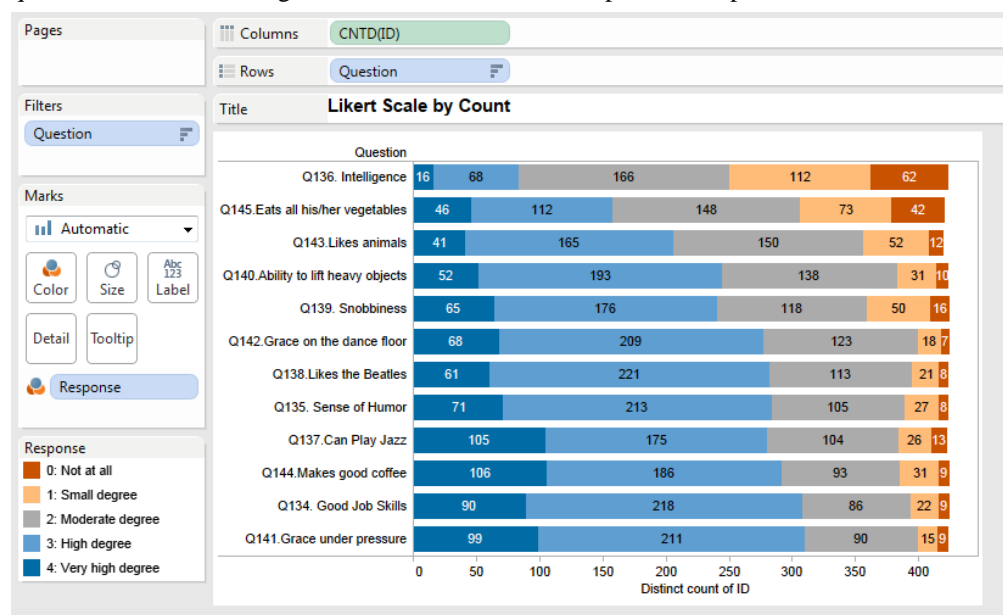

Figure 45 – Likert scale question results by count

Or you can calculate by percent of total, going across the bar, as shown here.
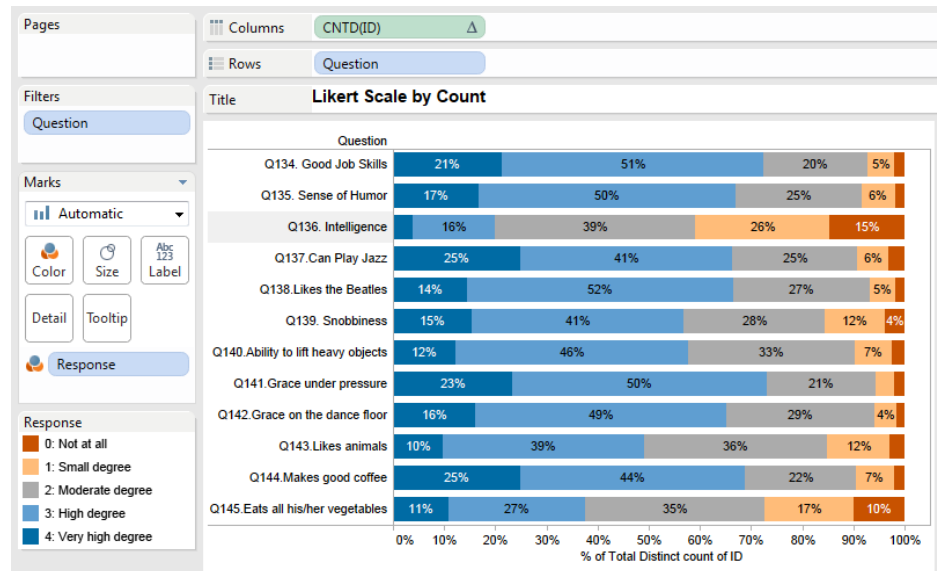
Figure 46 -- Likert scale questions by percentage

While colorful this is not a very good visualization as it's hard to sort by overall sentiment.

Here's an alternative approach that uses a divergent stacked bar with overlaying Likert scale.
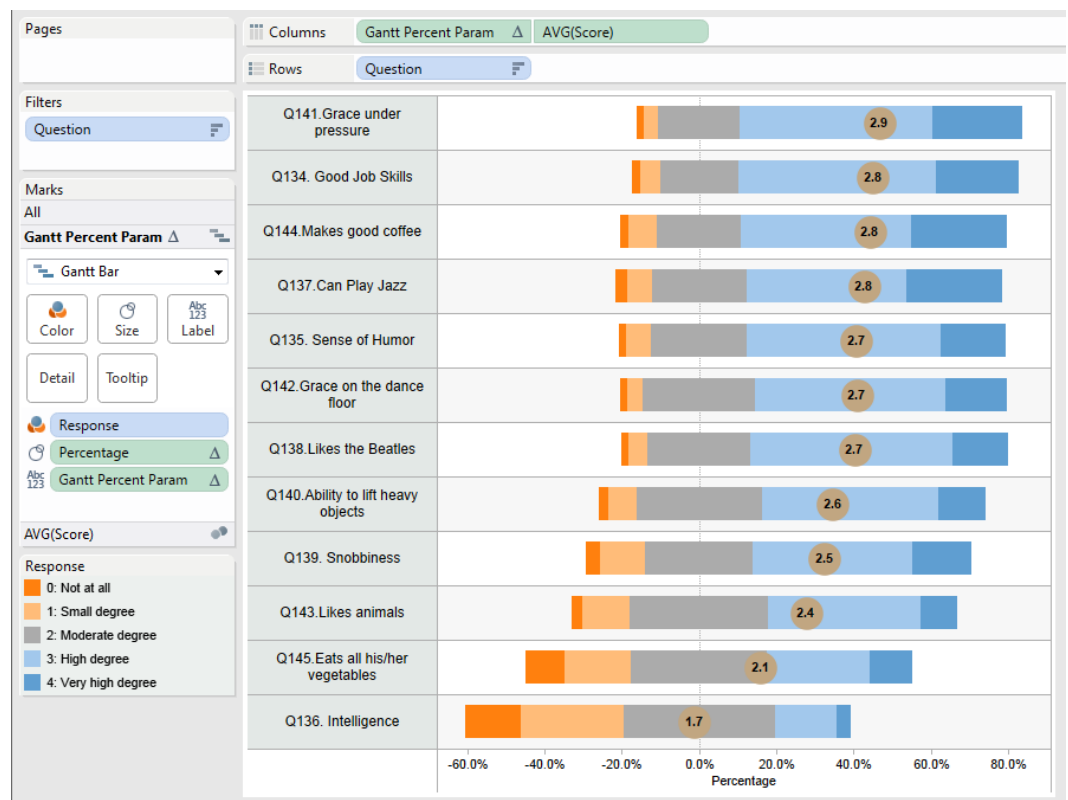

Figure 47 -- Divergent stacked bar chart with overlaying Likert scale (average of 0 to 4 for each question)

You can find the step-by-step instructions on how to create this visualization in Tableau at http://www.datarevelations.com/likert-scales-the-final-word.html.

Conclusion

High-end data analysis and visualization tools allow you to create more meaningful graphs and dashboards than what is readily available from survey tool vendors. In this whitepaper we have seen examples of some of the types of analytics you can do with tools such as Tableau, as well as how data can be structured effectively to get you started.