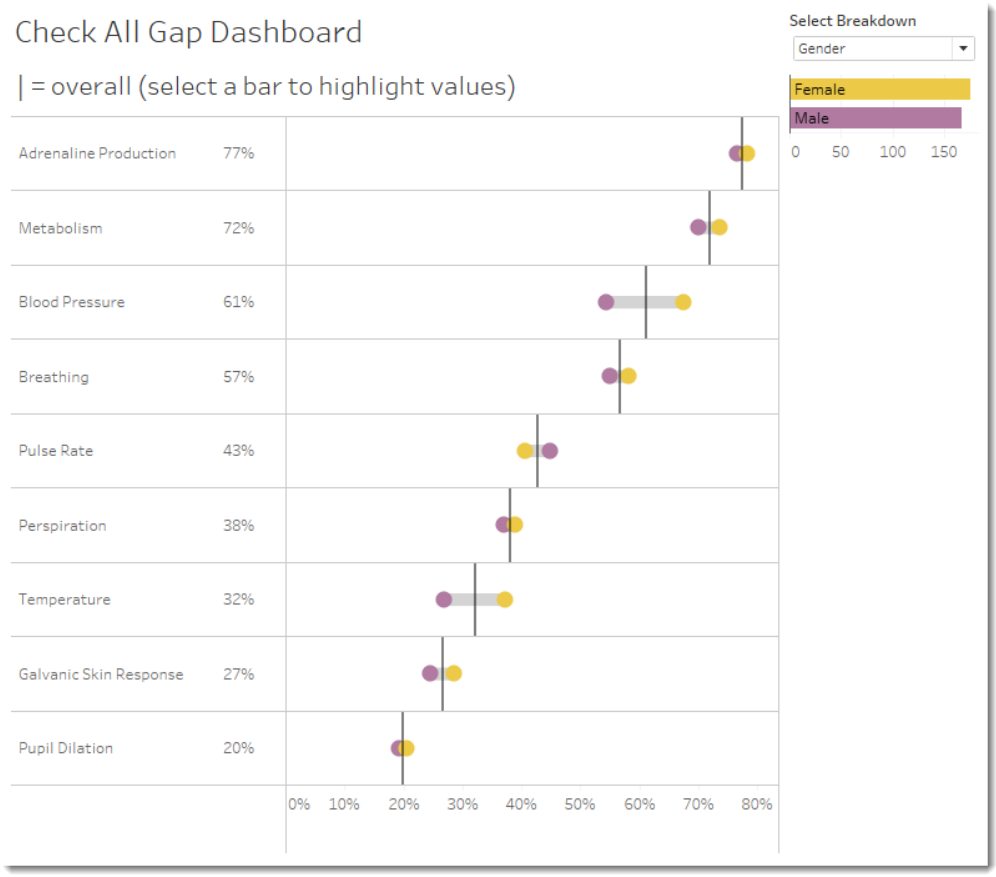
Visualizing Survey Data Using Tableau

Attendee Guide



www.DataRevelations.com

914.945.0567

For Tableau 2020.2 and later – October 2021

Contents

I. Getting Started

Understanding the Relationship between the Survey and the Data

The sample survey is fairly small. 845 respondents provided demographic information (age, location, and gender) and answered several types of questions (single punch, select all that apply, Likert, and benchmarking.)

Here's what some of the survey questions looked like to people taking the survey.

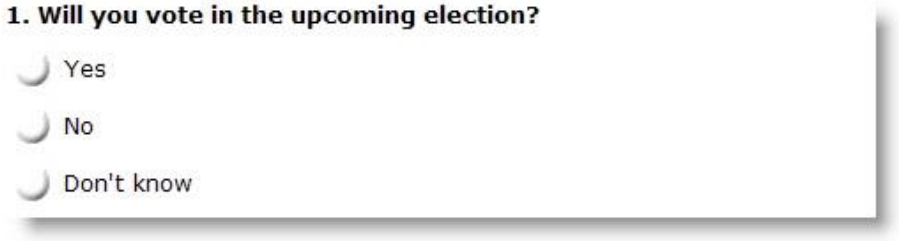


Figure 1 – A Yes / No / Maybe question

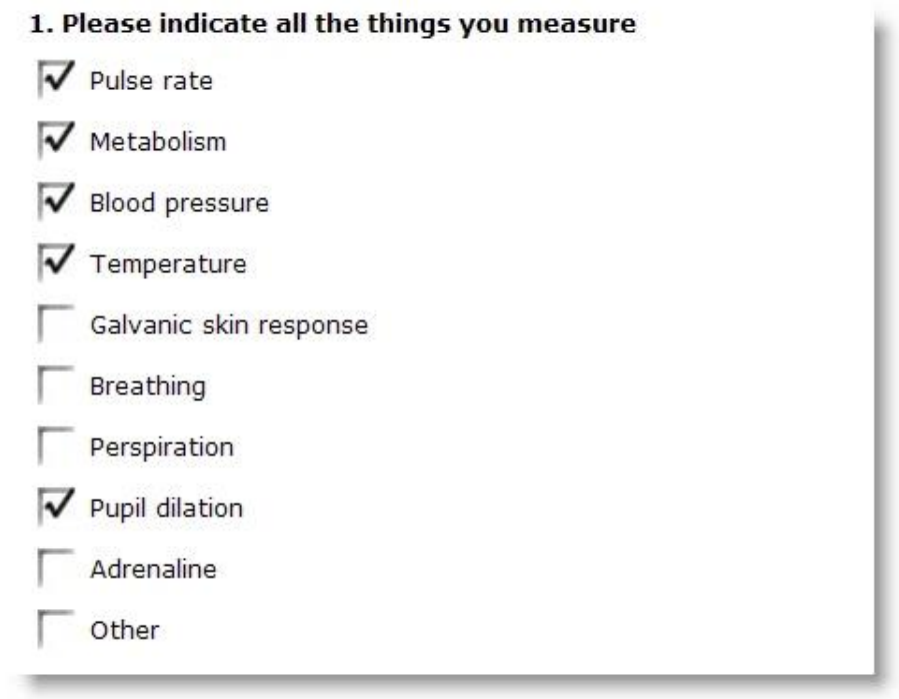
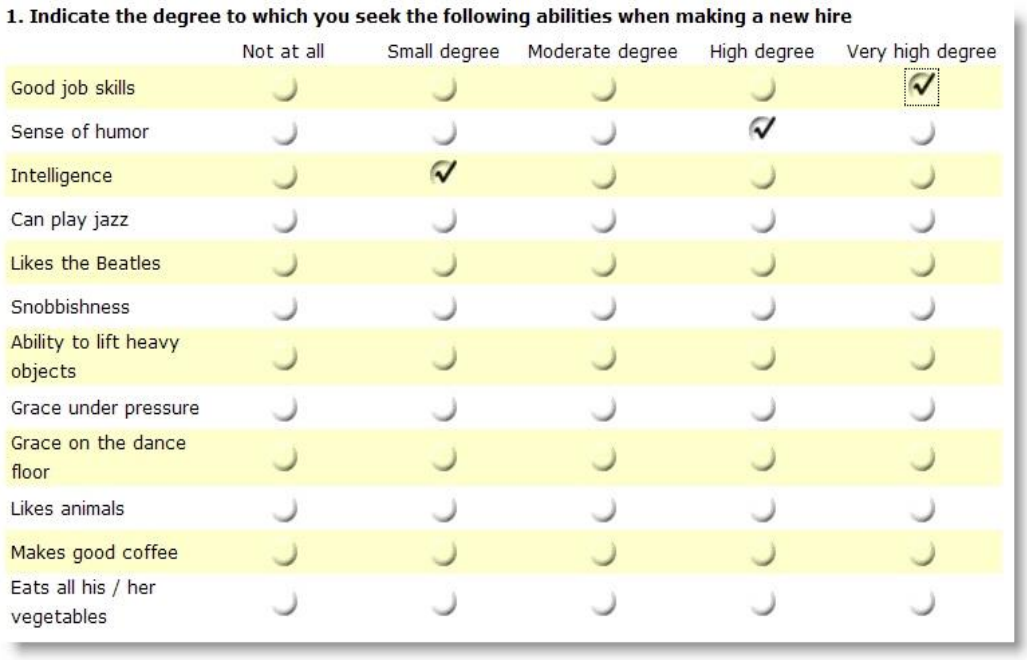


Figure 2 – A Check-all-that-apply question



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 3 – A collection of Likert scale questions

Let's see how these questions map to the data.

To Explore the Survey Data

1. Open the file 1_**DataRevelations_SurveyData_V4.XLSX** from the **Source** folder.
2. With the first tab selected (the one called "**Data Labels**"), see if you can find the column that corresponds to the **Yes / No / Don't know** question.
3. See if you can find which columns correspond to the **Check-all-that-apply** question and the **Likert scale** questions.
4. There's also a question that asks folks to enter their annual salary; see if you can find that one.
5. Click the second tab in the Excel workbook (the one called "**Data Numbers**") and consider the similarities and differences between the survey responses when rendered as labels and the survey responses as numbers.
6. Have a look at the third tab (the one called "Question Helper") and consider how the question IDs are mapped and grouped.

Answers: *The Yes / No / Don't know is Q1 in column F; the check-all-that-apply question is spread across Q2_1 – Q2_9 in columns H through P; the Likert scale questions are represented by Q3_1 through Q3_9 in columns Q through Y; the Salary question is Q100 in column G. By the way, there's some more Likert questions further to the right.*

Topics for Discussion

- Why have a tab for label responses and another for numeric responses?
- Why is there a Question Helper tab?

A Preview of Data Reshaping

Without a doubt the number one impediment for success with Tableau is arranging your data so that it "plays nicely" with Tableau.

Survey data is no exception, and we are doomed to failure if we leave the data as it is with one row for each respondent and a separate column for each question.

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

One row per respondent.

	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	Millenials	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	Millenials	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	Millenials	1	Yes	$ 110,462	No	Yes	Yes	Yes
22	37	Female	North America	Millenials	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 4 -- "Flat" data as downloaded from survey system where there is a separate column for each question.

… we want something that looks like this:

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

	A	B	C	D	E	I	K
1	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

Figure 5 -- "Reshaped" data where there are only two columns for each question, 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 "Q2_8" into something meaningful.

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

	A	B	C	D	E	F	G	H	I	J	K
1	RespID	Q0_Gender	Q0_Generation	Q0_Location	Q0_Weight	Qtype	Question Grouping	Wording	Question ID	Value	Labels
2	2	Male	Generation X	South America	1	Single-Punch	Vote	Vote in the upcoming election?	Q1	0	No
3	2	Male	Generation X	South America	1	Benchmark	Salary	What is your salary?	Q100	98037.68	98037.68
4	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Pulse Rate	Q2_1	0	No
5	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Metabolism	Q2_2	0	No
6	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Blood Pressure	Q2_3	1	Yes
7	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Temperature	Q2_4	0	No
8	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Galvanic Skin Response	Q2_5	1	Yes
9	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Breathing	Q2_6	0	No
10	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Perspiration	Q2_7	0	No
11	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Pupil Dilation	Q2_8	0	No
12	2	Male	Generation X	South America	1	Multi-Punch	What do you measure	Adrenaline Production	Q2_9	1	Yes
13	2	Male	Generation X	South America	1	Likert	Importance	Price	Q28_IMP	5	Very Important
14	2	Male	Generation X	South America	1	Likert	Satisfaction	Price	Q28_SAT	1	Not at all satisfied
15	2	Male	Generation X	South America	1	Likert	Importance	Response Time	Q29_IMP	5	Very Important
16	2	Male	Generation X	South America	1	Likert	Satisfaction	Response Time	Q29_SAT	1	Not at all satisfied
17	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	Good Job Skills	Q3_1	2	Small degree
18	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	Good Sense of Humor	Q3_2	2	Small degree
19	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	High Intelligence	Q3_3	1	Not at all
20	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	Can Play Jazz	Q3_4	2	Small degree
21	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	Likes the Beatles	Q3_5	3	Moderate degree
22	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	Good Ability to lift heavy objects	Q3_6	2	Small degree
23	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	Has grace under pressure	Q3_7	4	High degree
24	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	Is Kind to animals	Q3_8	2	Small degree
25	2	Male	Generation X	South America	1	Likert	Indicate degree to which you agree	Makes good coffee	Q3_9	5	Very high degree
26	2	Male	Generation X	South America	1	Likert	Importance	24-7 Support	Q30_IMP	5	Very Important
27	2	Male	Generation X	South America	1	Likert	Satisfaction	24-7 Support	Q30_SAT	1	Not at all satisfied
28	2	Male	Generation X	South America	1	Likert	Importance	Ease of Use	Q31_IMP	5	Very Important
29	2	Male	Generation X	South America	1	Likert	Satisfaction	Ease of Use	Q31_SAT	1	Not at all satisfied
30	2	Male	Generation X	South America	1	Likert	Importance	Ability to Customize UI	Q32_IMP	5	Very Important
31	2	Male	Generation X	South America	1	Likert	Satisfaction	Ability to Customize UI	Q32_SAT	1	Not at all satisfied
32	2	Male	Generation X	South America	1	Likert	Importance	Ability to filter based on role	Q33_IMP	2	Of Little Importance
33	2	Male	Generation X	South America	1	Likert	Satisfaction	Ability to filter based on role	Q33_SAT	1	Not at all satisfied

Figure 6 -- Survey data nirvana. Label responses, numeric responses, logical groupings, and mapping of Question IDs to something that is human readable.

While we don't have to have both labels and numbers, the grouping, and the question wording, having these elements will make our work much faster and easier.

We'll see in the section called "Getting Your Data Setup" on page 52 how to take the flat data and the question helper and convert them into the fully rendered data set we show above. For the next set of examples, we will work with data that has been converted for us.

Some Useful Shortcut Keys

Here are some shortcuts we'll be using in the class:

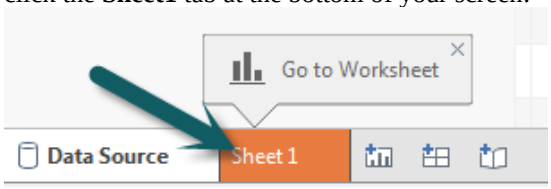
Description	Windows	Mac
Make chart less wide	CTRL + Left Arrow	Control-Command-Left Arrow
Make chart wider	CTRL + Right arrow	Control-Command-Right Arrow
Opens the Drop Field menu	Right-click + Drag to shelf	Option-Drag to shelf
Copies a field in the view to be placed on another shelf or card	Ctrl + Drag	Command-Drag

Connecting to Data

For the next series of examples, we will connect to the data contained in **2_DataRevelatins_Reshaped_V4.xlsx**. (This is the data represented in the screenshot in Figure 6 above.)

To Connect to The Survey Data Source and Create an Extract

1. From within Tableau click **Connect to data**.
2. Click **Microsoft Excel**.
3. Navigate to the folder called **Source**.
4. Select **2_DataRevelatins_Reshaped_V4.xlsx** (this is different from the one we just looked at) and click **Open**. As there is only one sheet that contains any data Tableau automatically places that into the active table area.
5. Indicate that you want to connect via an **Extract** (it's on the upper right corner) and click the **Sheet1** tab at the bottom of your screen.



6. Tableau will offer to create a .hyper file. Accept the default name and click **Save**. Your screen should look like the one shown below.

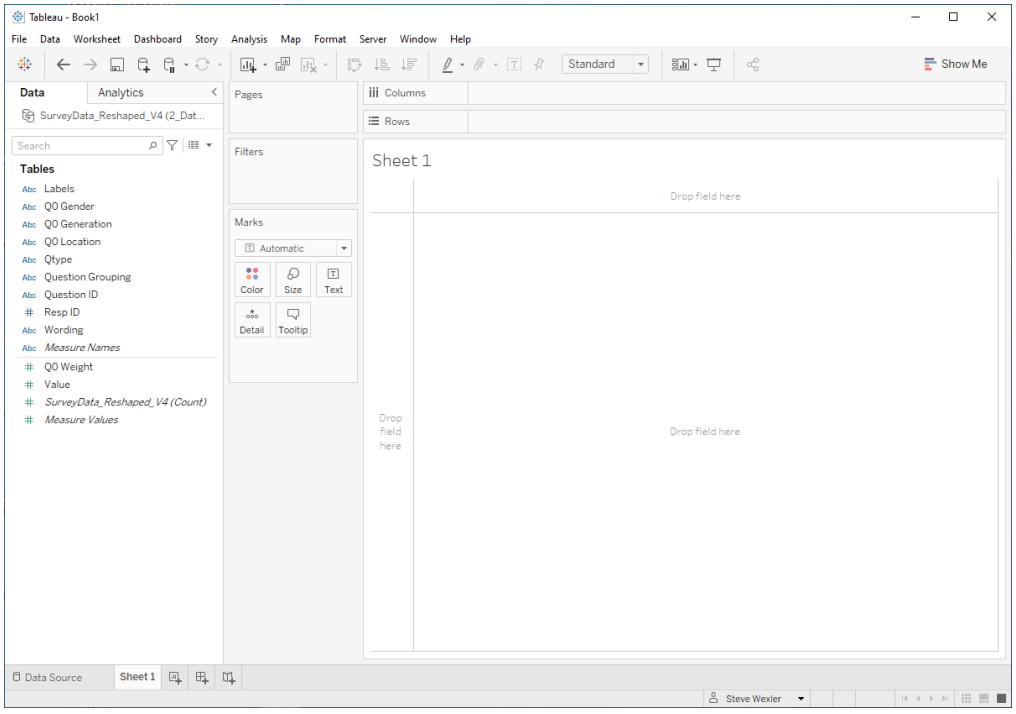


Figure 7 -- Tableau after connecting to the survey data source.

Exploring the Demographics

Before we plow into analyzing responses to questions let's first get a sense of just who is in our respondent pool. We will also create an "n=" display to use on a dashboard so we can see just how many respondents are in play when we apply various filters.

Using a Mac? It's Option-Drag to Columns shelf.

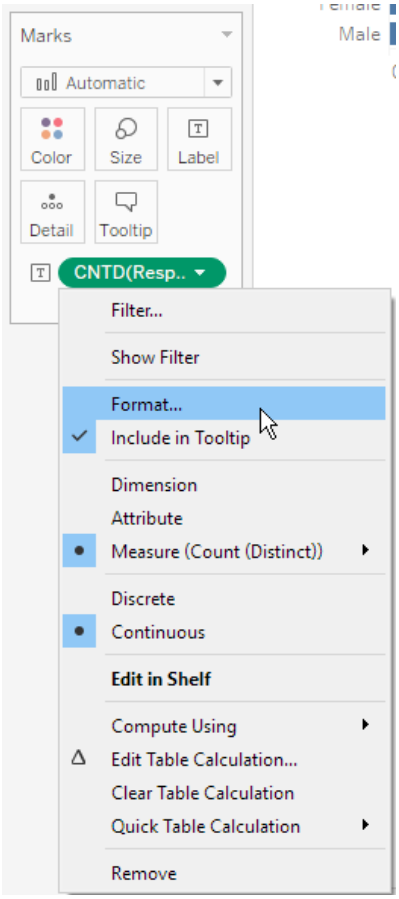
To Show Respondent Breakdown by Gender

1. Right-drag **Resp ID** to the Columns shelf and select **CNTD(Resp ID)** from the Drop Field dialog box.
2. Drag **Q0 Gender** onto Rows.
3. Right-click CNTD(Resp ID) on the Columns shelf, select **Quick Table Calculation**, and then select **Percent of Total**.

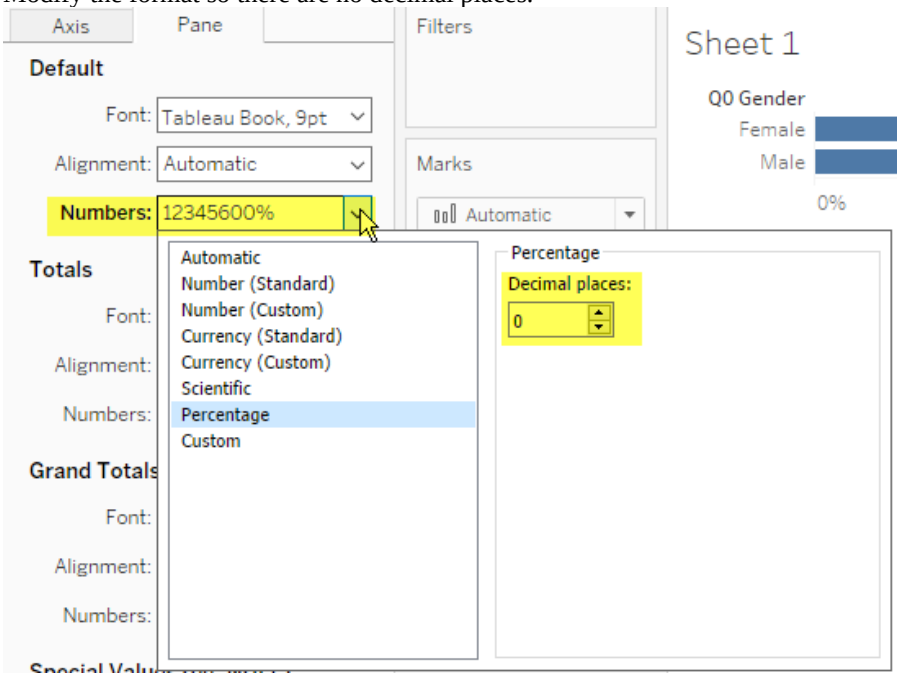
Using a Mac? It's Command-Drag.

4. Control-drag **CNTD(Resp ID)** to the **Label** button.

5. Right click the newly-dropped pill, and select **Format** from the popup menu.



6. Modify the format so there are no decimal places.



7. Close the Format pane (look for a little X to click)

8. Right drag Resp ID from the Data pane (not the columns shelf) onto the **Rows** shelf and select **CNTD(Resp ID)** from the Drop Field dialog. Yes, the visualization will look dumb.

9. Right-click the just-dropped field you placed on the Rows shelf and select **Discrete**. This will turn the green pill into a blue pill and make the viz look a lot better, as shown here.

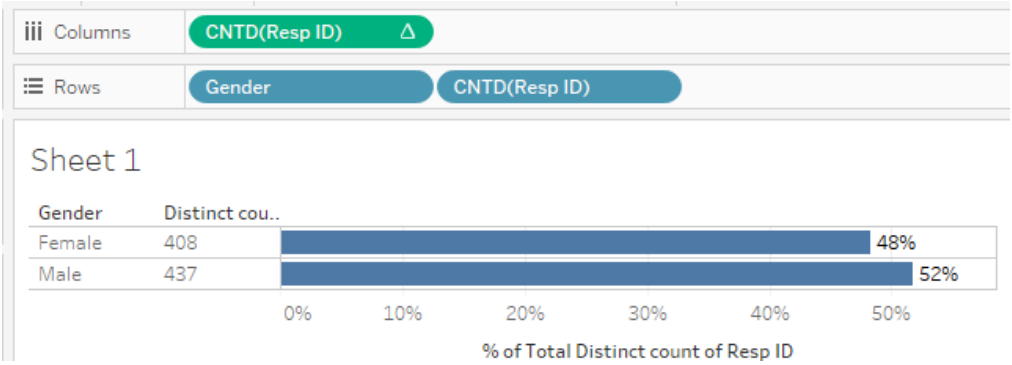


Figure 8 -- What happens when you make a measure discrete.

10. Right-click the **Axis** and de-select **Show Header**.

11. Right-click the Field labels (the small text that has the words "Q0 Gender" and "Distinct…" and select **Hide Field Labels for Rows**.

12. Modify the formatting to suit your tastes (e.g., remove the rule between Female and Male.) And definitely make the chart less wide (there's a really good shortcut key for this).

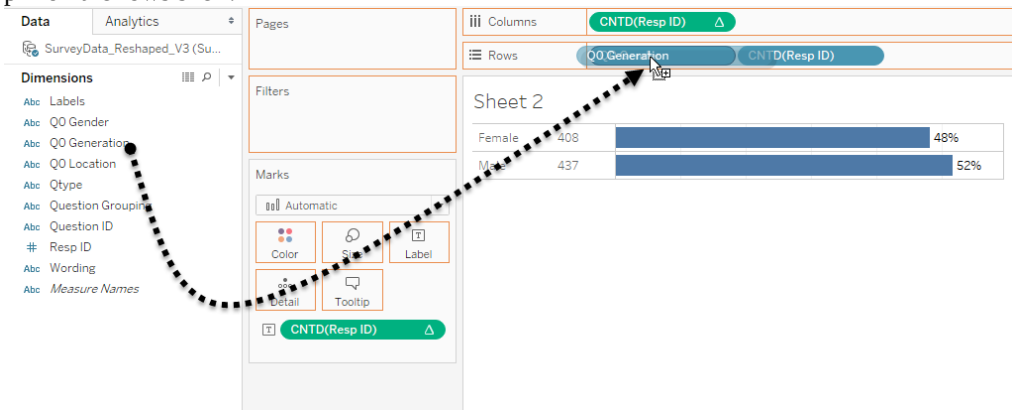
13. Rename the tab on the bottom left of your screen to **Gender**.

14. Save your work to *as a packaged workbook*. You can name the file whatever you want and save it wherever you want; just make sure you can find it later.

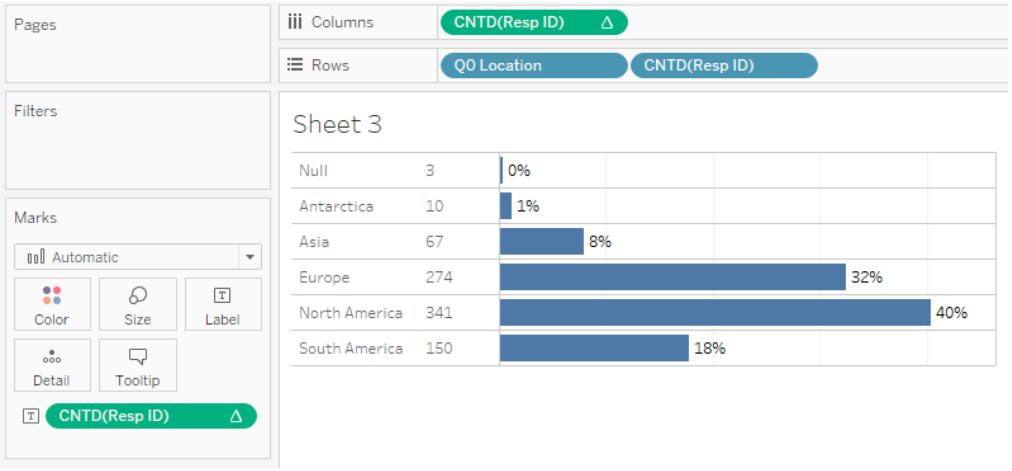
To Show Respondent Breakdown by Generation and Location

Note: If you had difficulty completing the previous exercise, open the file **1a_GenderDemographic.twbx** from the **Starter** folder and work with that.

1. Right-click the tab named Gender and select **Duplicate**.
2. Drag **Q0 Generation** from the list of Dimensions and place it on top of the Gender pill on the rows shelf.

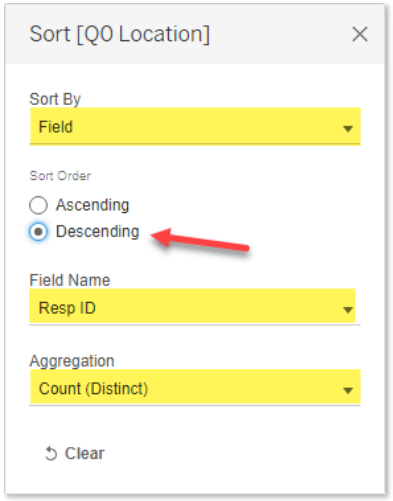


3. Rename the sheet **Generation**.
4. Right-click the sheet Generation and select **Duplicate**.
5. Drag **Q0 Location** from Dimensions and place it on top of the Generation pill on the rows shelf.



6. Right-click **Null**, select **Edit Alias**, and change the Alias to **Not Specified**.
7. Right-click the **Q0 Location** pill on the Rows shelf and select **Sort**.

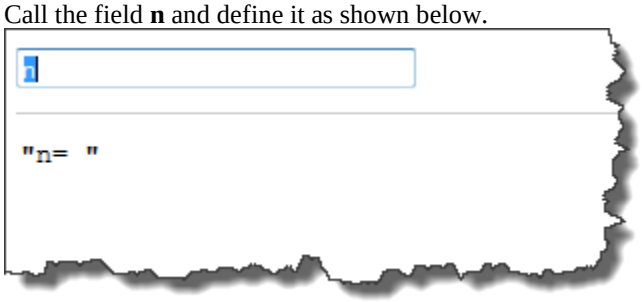
8. Indicate you want to sort by the Field Resp ID, in descending order, using Count (Distinct) as the aggregation, as shown below.



9. Rename the tab **Location** and save your work.

To Create an "n=" Visualization

1. Create a new sheet.
2. Right-click in a blank portion of the Data window (where the list of measures and dimensions are) and select **Create Calculated Field**.
3. Call the field **n** and define it as shown below.



4. Click **OK**.
5. Drag this newly-created field to the **Rows** shelf.
6. Right-drag **Resp ID** to the Text button (on the Marks Card area) and select **CNTD(Resp ID)** from the Drop Field dialog box.
7. Hide the field label for rows.
8. Format the worksheet so that everything is **14-point Tableau Book**.
9. Rename the sheet **Sample Size**.
10. Save your work.

Building a Demographics Dashboard

Now that we have the individual components and overall sample size visualization rendered, let's see how to fashion a demographics dashboard with action filters.

Our goal is to create an interactive dashboard that looks like this.

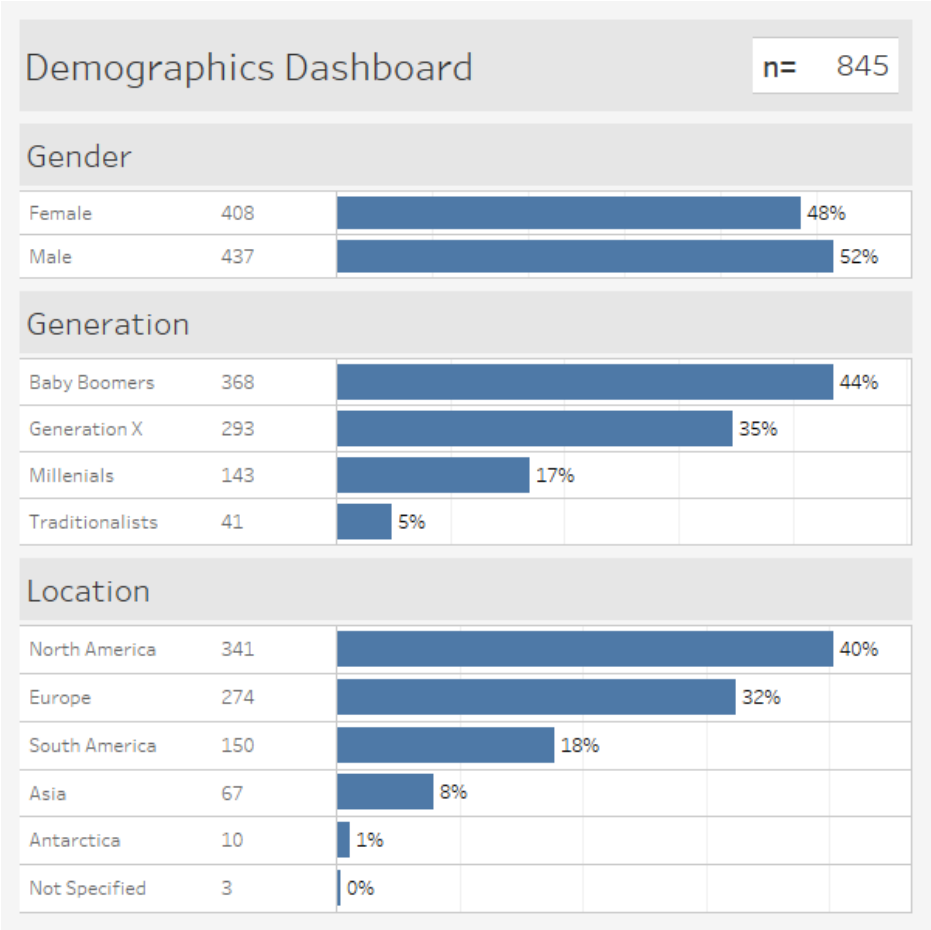
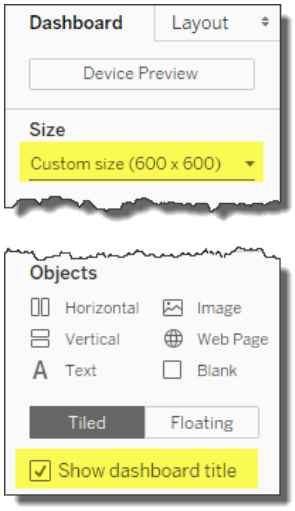


Figure 9 -- Simple demographics dashboard

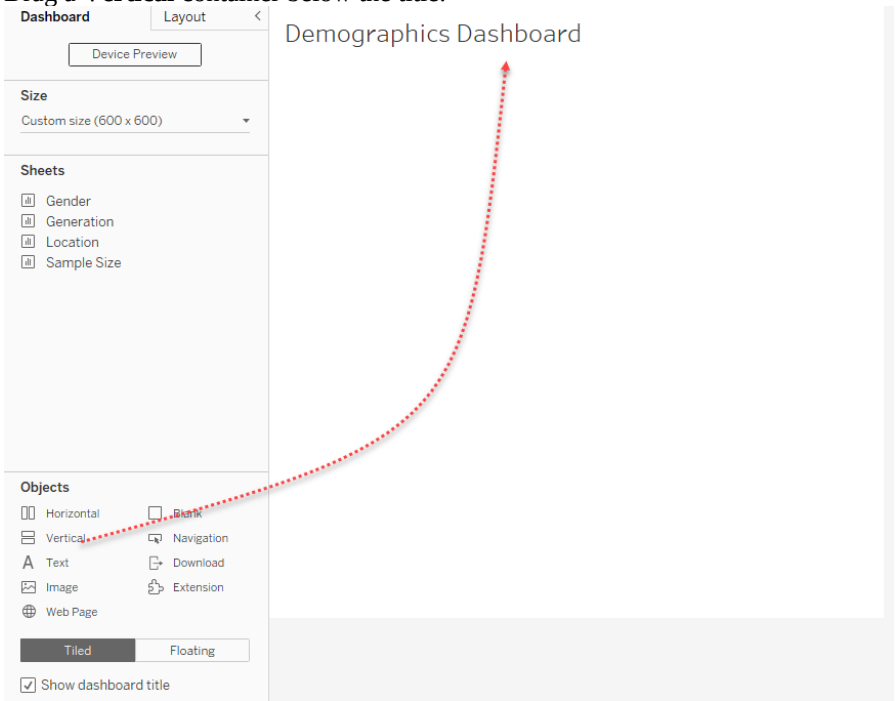
To Build the Demographics Dashboard

Note: If you had difficulty completing the previous exercise, open the file **1b_AllDemographics.twbx** from the **Starter** folder and work with that.

1. Create a New Dashboard.
2. Make the dashboard size Exactly 600 x 600 and indicate that you want to show the title.

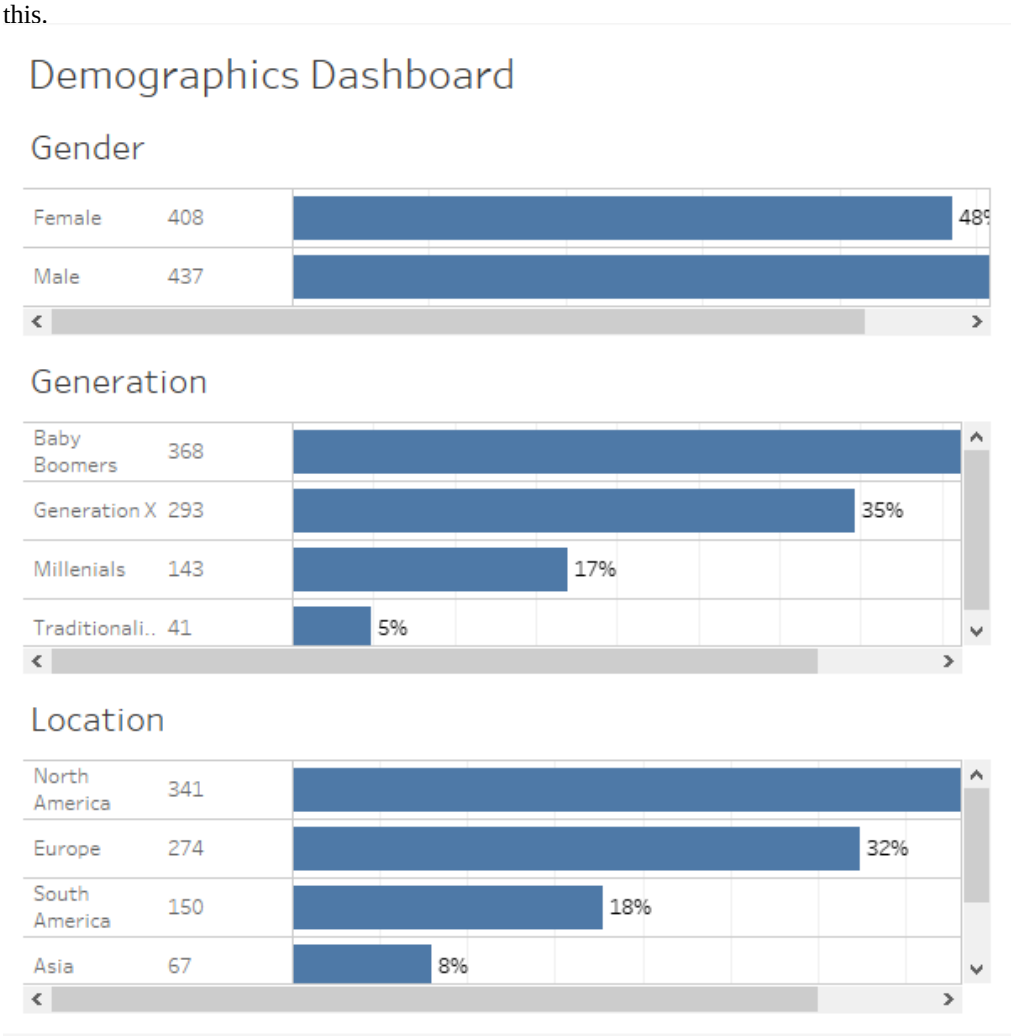


3. Rename the tab **Demographics Dashboard.**
4. Drag a **Vertical** container below the title.

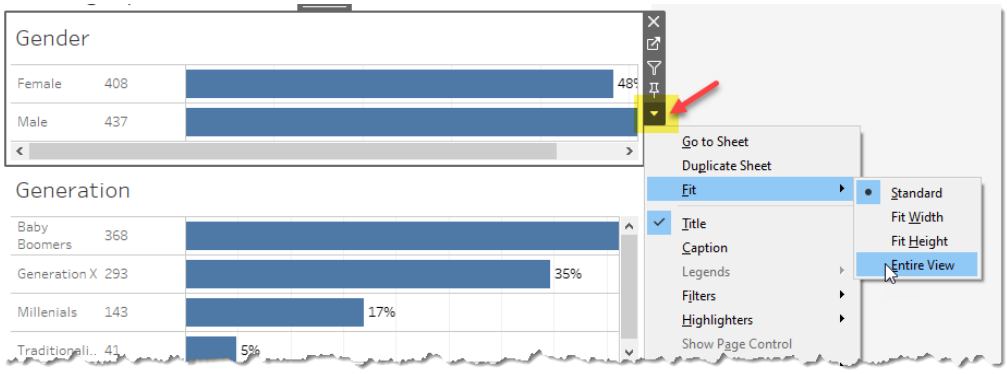


5. Drag **Gender** into the newly-dropped vertical container.
6. Drag **Generation** below Gender.

7. Drag **Location** below Generation. You may have something that looks messy, like this.



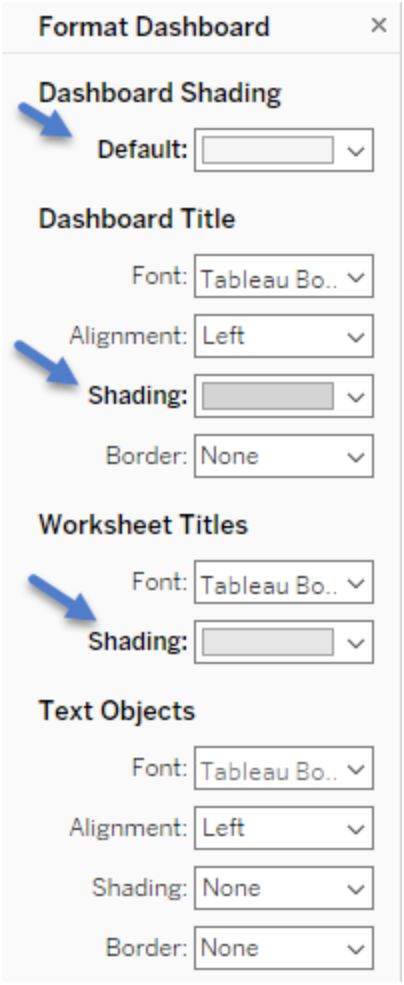
8. Click anywhere in the Gender sheet to activate it, then click the caret menu, and select **Fit** and **Entire view**.



9. Repeat this for the **Generation** and **Location** worksheets.

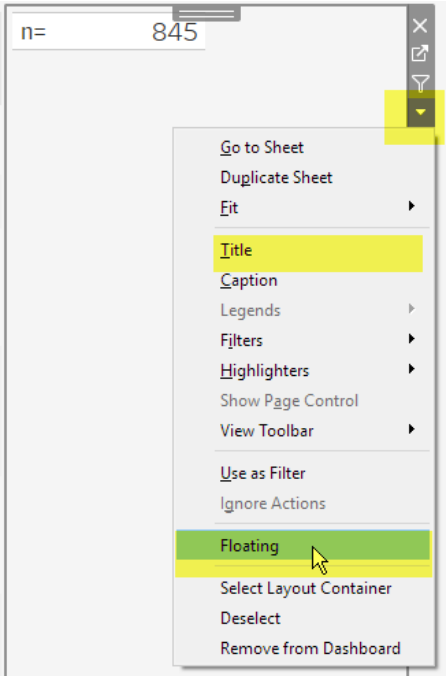
10. From the **Dashboard** menu select **Format**.

11. Make the Default shading a very light gray; left-align the dashboard title (if it is not already left aligned), and shade it medium gray, and make the Worksheet Titles a medium light gray.



12. Close the Format Dashboard pane.
13. Drag **Sample Size** to the right side of the dashboard. Don't worry if things don't look great at the moment.

14. Select the Sample Size window and click the arrow in the top right corner. Then indicate you want the element to be **Floating** and you do not want to display a title, as shown below.



15. Move the Sample Size window to the upper right and resize it so that it, well, looks good. You may need to select the Entire View option so that it fits nicely within the size you have selected.

!

Important: Your dashboard should look like the one page 12. Have a look at the length of the bars in each demographic segment. Anything look a little bit off?

To Add Dashboard Actions

1. Select the **Gender** window and click the carat (down arrow) in the upper right corner.
2. Select **Use as Filter**.
3. Repeat for the **Generation** and **Location** windows.
4. Try selecting elements from different windows (e.g., Female from the Gender window and Baby Boomers from the Generation window). Notice that the visualizations change, and the sample size changes.
5. Press the **Esc** key to clear your selections.
6. Save your work.

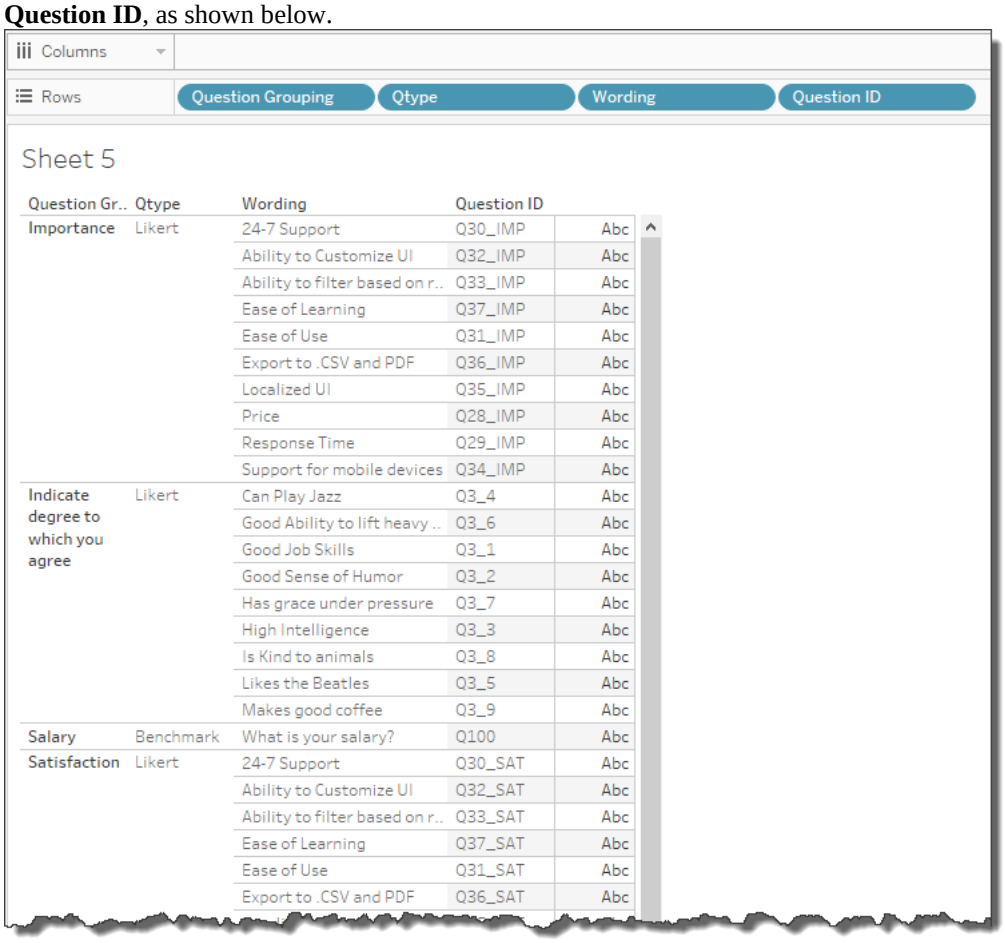
Taking Inventory – Mapping All Questions with All Possible Responses

Before visualizing any responses, let's get a good handle on questions, responses, and seeing if there was anything that was poorly coded on our data set.

To Map All Questions and Responses

Note: If you had difficulty completing the previous exercise, open the file **1c_DemographicsDashboard.twbx** from the **Starter** folder and work with that.

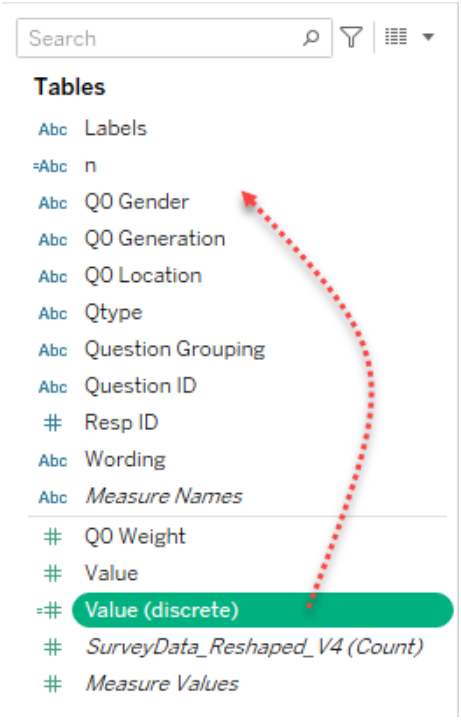
1. Create a new sheet.
2. Drag **Question Grouping** onto rows, followed by **Qtype**, then **Wording**, and then **Question ID**, as shown below.



Question Gr..	Qtype	Wording	Question ID	
Importance	Likert	24-7 Support	Q30_IMP	Abc
		Ability to Customize UI	Q32_IMP	Abc
		Ability to filter based on r..	Q33_IMP	Abc
		Ease of Learning	Q37_IMP	Abc
		Ease of Use	Q31_IMP	Abc
		Export to .CSV and PDF	Q36_IMP	Abc
		Localized UI	Q35_IMP	Abc
		Price	Q28_IMP	Abc
		Response Time	Q29_IMP	Abc
		Support for mobile devices	Q34_IMP	Abc
Indicate degree to which you agree	Likert	Can Play Jazz	Q3_4	Abc
		Good Ability to lift heavy ..	Q3_6	Abc
		Good Job Skills	Q3_1	Abc
		Good Sense of Humor	Q3_2	Abc
		Has grace under pressure	Q3_7	Abc
		High Intelligence	Q3_3	Abc
		Is Kind to animals	Q3_8	Abc
		Likes the Beatles	Q3_5	Abc
		Makes good coffee	Q3_9	Abc
Salary	Benchmark	What is your salary?	Q100	Abc
Satisfaction	Likert	24-7 Support	Q30_SAT	Abc
		Ability to Customize UI	Q32_SAT	Abc
		Ability to filter based on r..	Q33_SAT	Abc
		Ease of Learning	Q37_SAT	Abc
		Ease of Use	Q31_SAT	Abc
		Export to .CSV and PDF	Q36_SAT	Abc

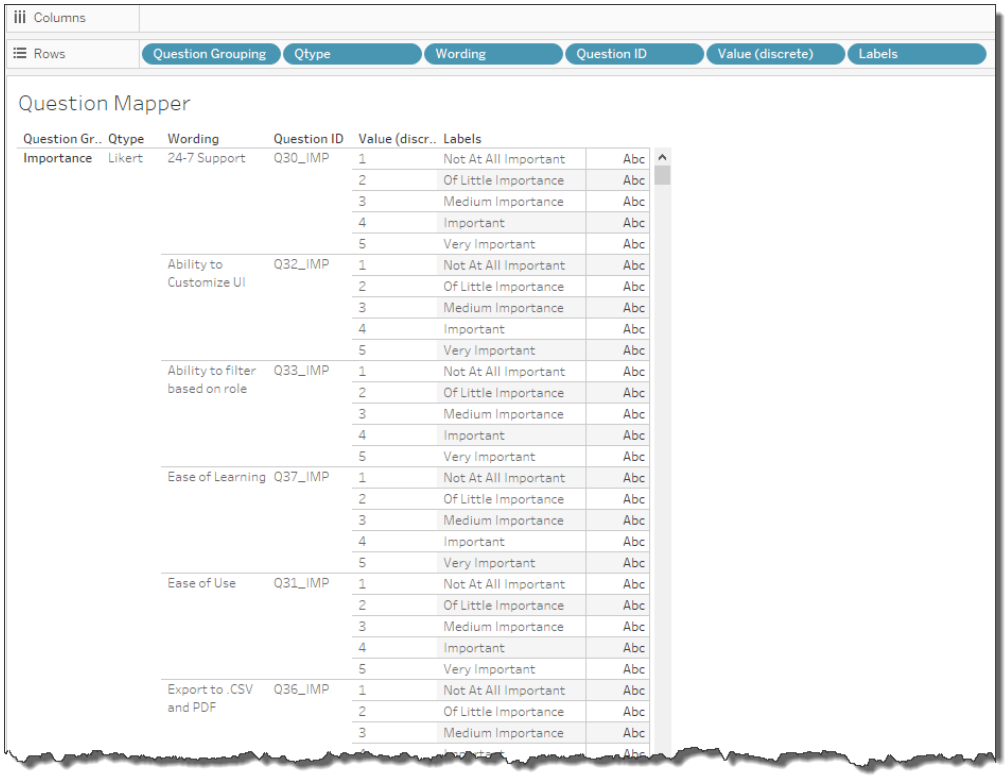
3. Right-click the measure called **Value** and select **Duplicate**.
4. Rename the newly-created field **Value (discrete)**.
5. Drag the measure into the top area. This will make Tableau treat the field as something that is by default discrete (notice the blue color instead of green once it's

dropped).



6. Drag this newly-created dimension to Rows, after the other four pills.

7. Drag **Labels** to Rows. Your screen should look like the one shown below.



8. Rename the sheet **Question Mapper** and save your work.

So, just what do we have here?

You can see from the portion of the screen that you have a bunch of questions about "Importance" and can also see that the possible values go from 1 to 5 where 1 maps to "Not At All important", 1 maps to "Of Little Importance", etc.

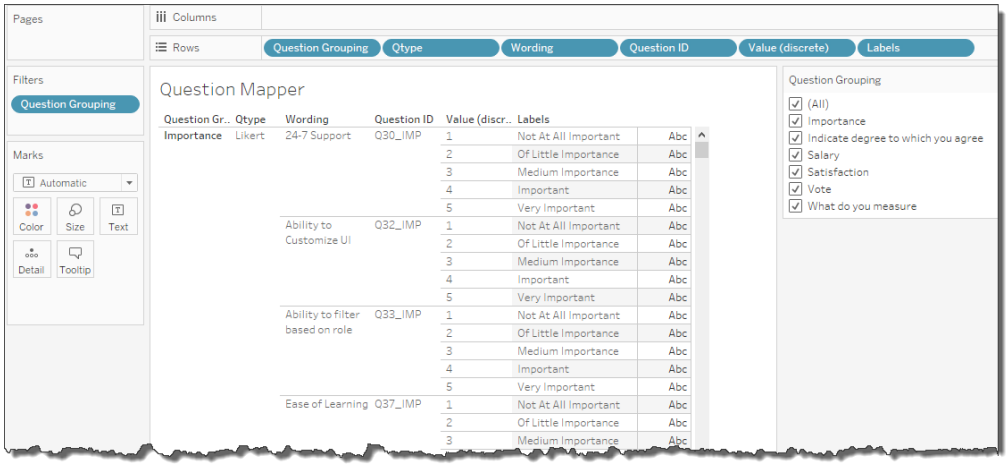
At this point you should be looking for any stray values, say a value of 6.

If you scroll down a little bit (next page) you'll see a question grouping called "Indicate the degree to which you agree" where you again have values of 1 through 5 but this time 1 maps to "Not at all", 2 maps to "Small degree", etc.

We should be pleased as it appears that our Likert questions consistently go from 1 through 5. This means we won't have to craft multiple sets of calculated fields to deal with different numeric scales.

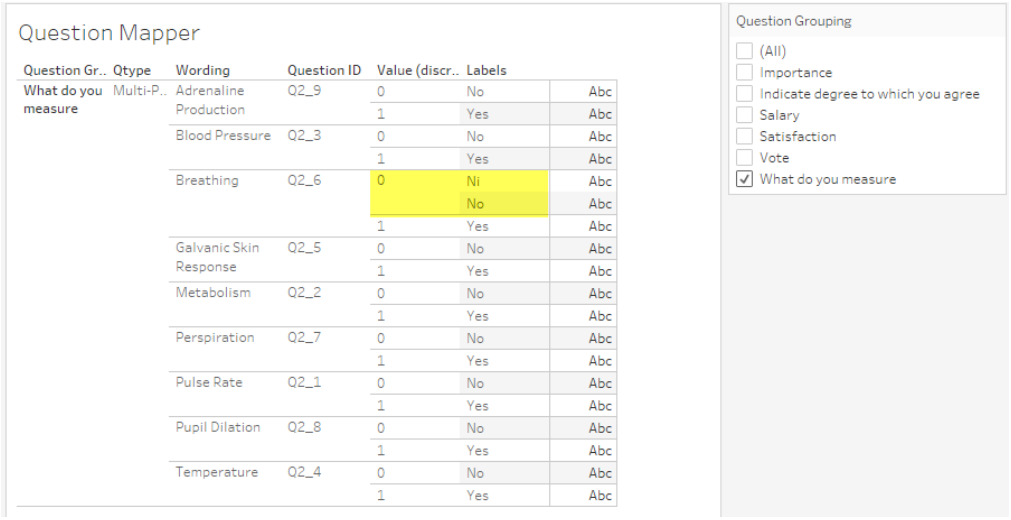
Indicate degree to which you agree	Likert	Good Job Skills	Q3_1	1	Not at all
				2	Small degree
				3	Moderate degree
				4	High degree
				5	Very high degree
		Good Sense of Humor	Q3_2	1	Not at all
				2	Small degree
				3	Moderate degree
				4	High degree
				5	Very high degree
		Has grace under pressure	Q3_7	1	Not at all
				2	Small degree
				3	Moderate degree
				4	High degree
				5	Very high degree
		High Intelligence	Q3_3	1	Not at all
				2	Small degree
				3	Moderate degree
				4	High degree
				5	Very high degree
		Is Kind to animals	Q3_8	1	Not at all
				2	Small degree
				3	Moderate degree
				4	High degree
				5	Very high degree
		Likes the Beatles	Q3_5	1	Not at all
				2	Small degree
				3	Moderate degree

At this point it might be useful to add a filter so you can focus on only certain question groups. You can do this by filtering by Question Grouping as shown below.



Spotting questions that have coding errors

In case you're wondering what a coding error looks like, see what happens if we just focus on the "What do you measure" questions, as shown below.



So, for all questions the universe of possible values is 0 and 1. Except for Question Q2_6 (Breathing), 0 maps to "No" and 1 maps to "Yes."

The miscoding of "Ni" instead of "No" will only present a problem if our calculated field for determining the percentage of people that checked an item were to use Labels instead of Values.

We'll discuss this in a little bit.

What happened to [Number of Records] and why you should care

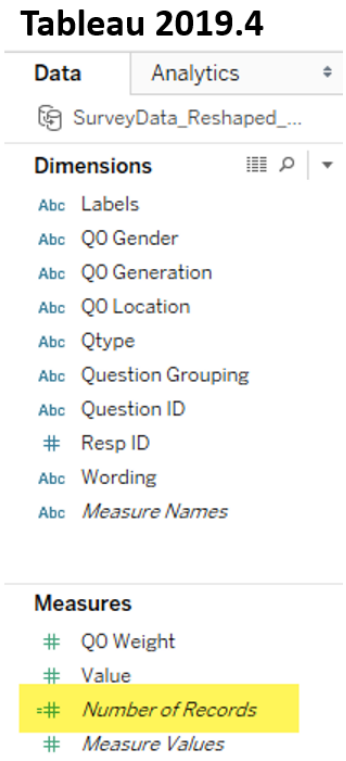
I've written almost ten years' worth of blog posts on visualizing survey data using Tableau (see https://www.datarevelations.com/visualizing-survey-data/). The good news is that dozens of how-to articles that are currently on my website are relevant and the techniques work.

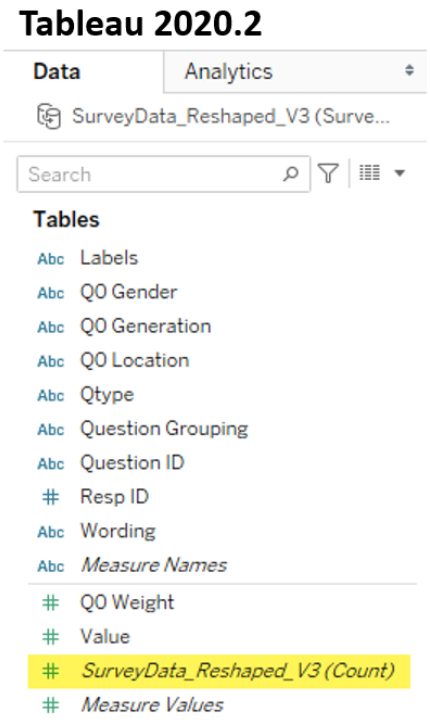
That said, Tableau made a HUGE change in how you can model data with the release of version 2020.2. With this release, a very useful field [Number of records] is no longer created automatically. This can be head-scratching if you try to apply the techniques and calculations in the how-to articles only to wonder "yes, but I don't have that field available. What am I supposed to do?"

The good news is that the workaround for this is ridiculously easy.

What it looked like then, and what it looks like now

I use a standard data set for many of the examples in the how-to articles (you can find it here, assuming you can access DropBox). Here's what how the data pane appears when you in Tableau 2019.4 vs. Tableau 2020.2 and later.

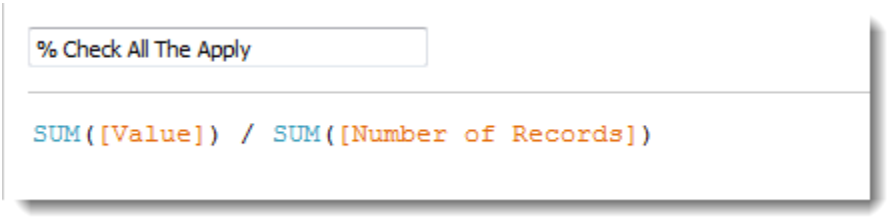




So, what happened to [Number of Records] and how do you make all the how-to articles that reference that field work?

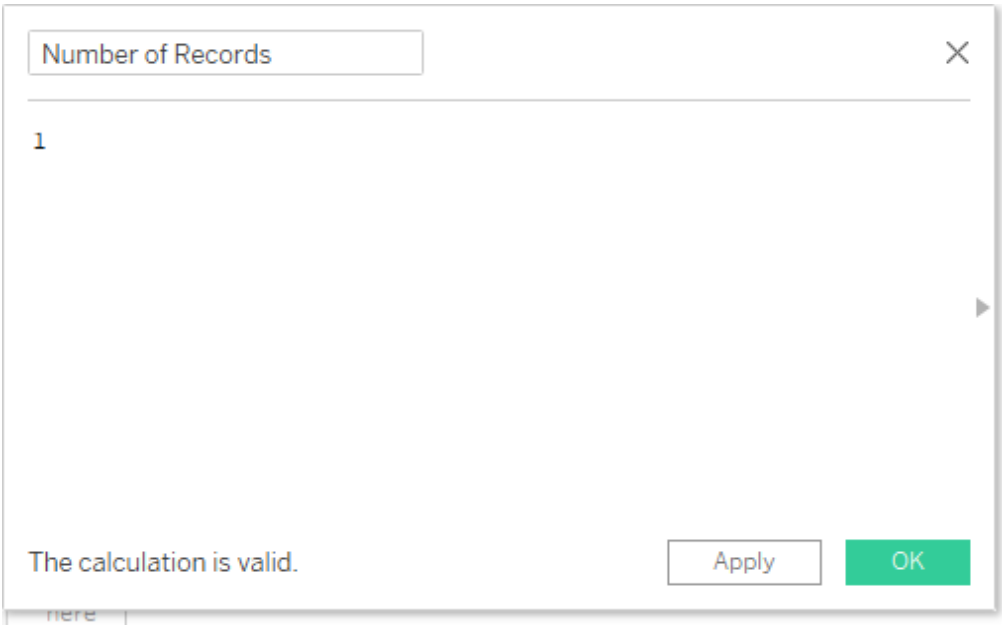
Making all the how-to articles work

Let's say you're reading the article "How to visualize check-all-that-apply questions using Tableau" and you see this formula.



What are you to do?

The easiest thing to do is to "roll your own" [Number of Records] field and define it as follows.



Yes, that's all there is to it. This field places a "1" in every row of your source data so Tableau will add up all the relevant rows to get the correct denominator when it performs the % Check All That Apply calculation.

Important: This works if you follow my advice on getting your data "just so" using Tableau Prep Conductor, Alteryx, EasyMorph, etc. This will not work if you start noodling with Tableau's new Relationship model (aka, "the noodle").

For the time being, I don't think you will need to use the new model as I think getting your data "just so" using one of the above-mentioned tools will work perfectly in most cases.

II. Visualizing Single-Response, Gap Charts, and Check-All-That-Apply Questions

Overview

Now that we have our demographics dashboard in place, we're ready to look at actual survey responses. For this next series of exercises, we'll look at the "Do you plan to vote in the up-coming election" responses and the "Indicate all the things you measure" responses.

Visualizing the Yes / No / Don't Know Question

From here on in our approach to all of the question sets will be similar in that we'll first filter by the question set that interests us and then craft a visualization around that set of questions.

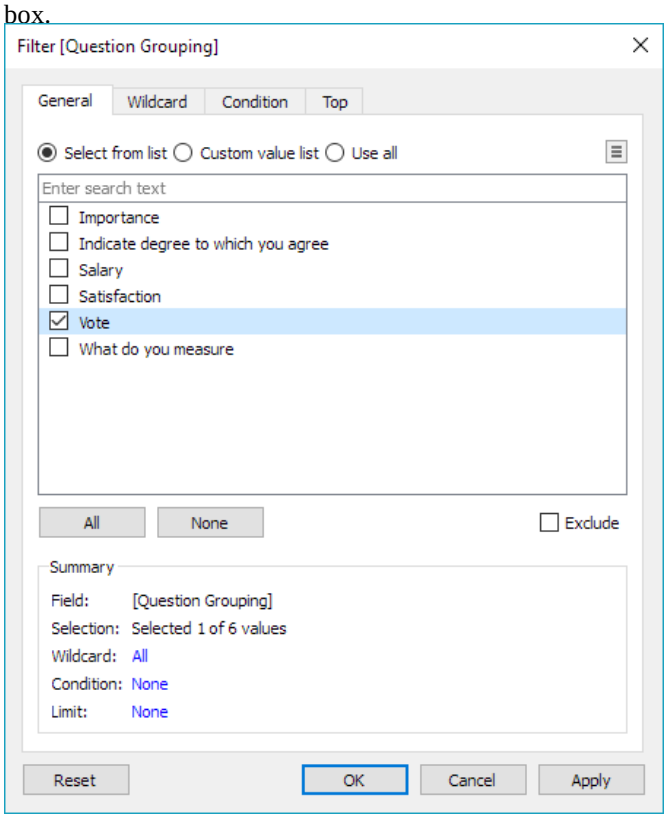
We'll also look at breaking down the result by the different demographic dimensions (Gender, Generation, and Location.)

To Visualize the Yes / No / Maybe Question (Raw Count)

Note: If you had difficulty completing the previous exercise, open the file **1d_DemographicsDashboard.twbx** from the **Starter** folder and work with that.

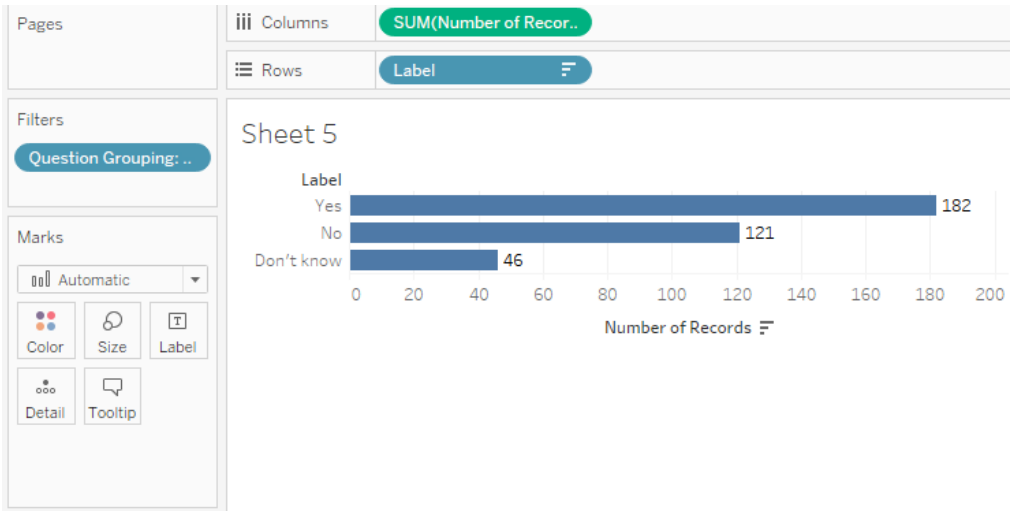
1. Create a new sheet.

2. Drag **Question Grouping** to the Filters shelf and select **Vote** from the Filter dialog box.



3. Click **OK**.
4. Drag **Labels** to the Rows shelf and **Number of Records** to the Columns shelf. **Important!** You can also drag *SurveyData_Reshaped+V4 (Count*) to the Columns shelf.
5. Sort the bars in descending order and turn **Mark Labels** on (click the  icon in the toolbar).

Your screen should look like this.



6. Rename the sheet **Do you plan to vote (raw)**.

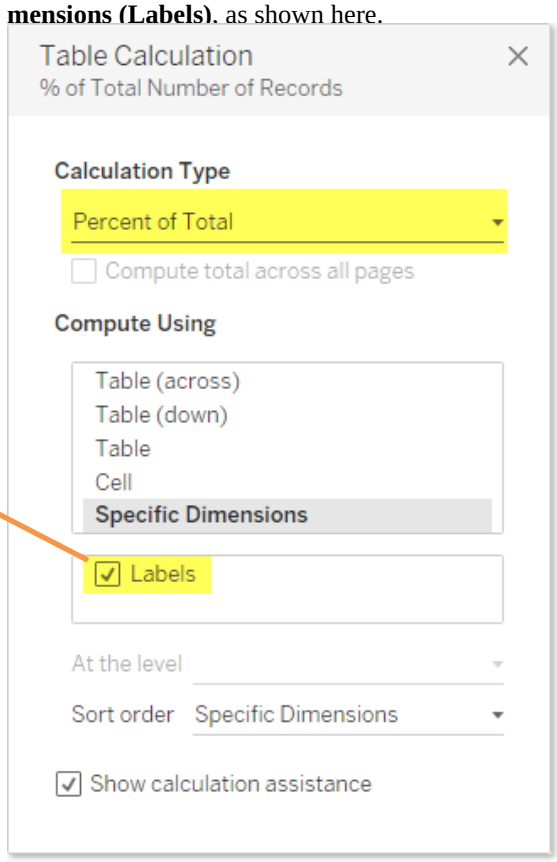
Let's see how to show a percentage rather as well as the raw vote count.

To Visualize the Yes / No / Maybe Question as a Percentage

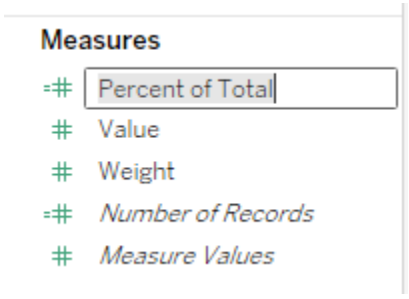
1. Right-click the **Do you plan to vote (raw)** tab and select **Duplicate**.
2. Right-click the **SUM(Number of Records)** pill on the Columns shelf and select **Add Table Calculation.**

Specifying the scope (vs. just selecting Table Down) will ensure things work if we add more dimension to the visualization.

3. Change the Calculation Type to **Percent of Total** and Compute using to **Specific Dimensions (Labels)**, as shown here.



4. Close the Table Calculation dialog box by clicking the X.
5. Drag the SUM(Number of Records) pill -- it now has a delta sign on it-- from the Columns shelf and drag it towards the bottom of the Data pane (where all the fields are along the left side).
6. When given the opportunity, name the just dragged field Percent of Total, as shown below.

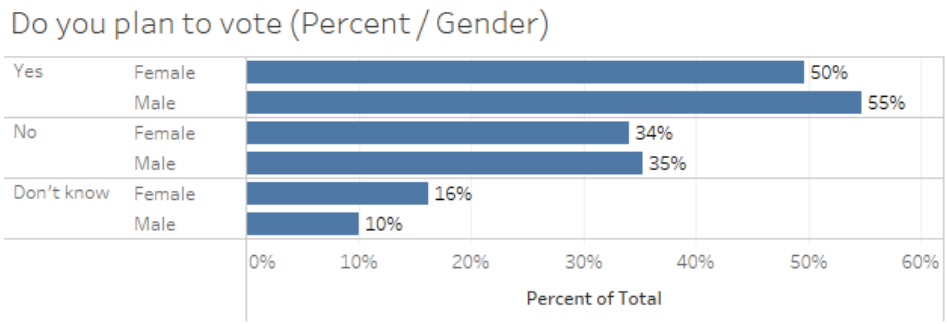


7. Right-click the newly created field **Percent of Total** field (along the left side, bellow the thin dotted line).
8. Select **Default Properties** and select **Number Format**.
9. Change the number format to **Percentage** with **0** decimal places.

10. Click **OK**.
11. Rename the tab **Do you plan to vote (percent).**
12. Save your work.

To Break Down by Gender

1. Right-click the tab labeled **Do you plan to vote (Percent)** and click **Duplicate**.
2. Drag **Q0 Gender** to the right of **Labels** on the Rows shelf.
3. Hide the field labels (right click **Q0 Gender** and select **Hide Field Labels for Rows**).
4. Rename the tab **Do you plan to vote (Percent / Gender).**
5. Save your work. Your screen should look like this.



Topics for Discussion

- How would you show a breakdown by Generation? By Location?
- Is there a cool way to simply swap these dimensions, versus having three separate visualizations?
- Is there a better way to show the differences between men and women?

Dealing with Weighted Responses

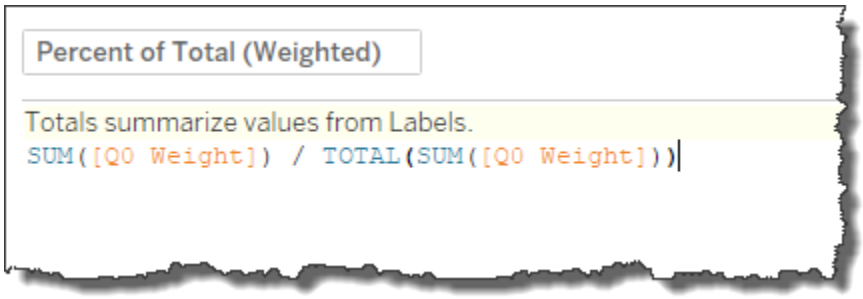
Each of the different question types we'll be exploring will need their own special formulas for dealing with weighted responses.

Let's see how to modify the Percent of Total formula so that it handles weighted responses.

To Create a Weighted Percent of Total Calculation

1. Right-click the **Do you plan to vote (Percent / Gender)** tab and select **Duplicate**.
2. Under Measures, Right-click **Percent of Total** and select **Duplicate**.
3. Right-click the newly created duplicate field and select **Edit**.

4. Change the name to **Percent of Total (Weighted)** and modify the formula so it read as follows.



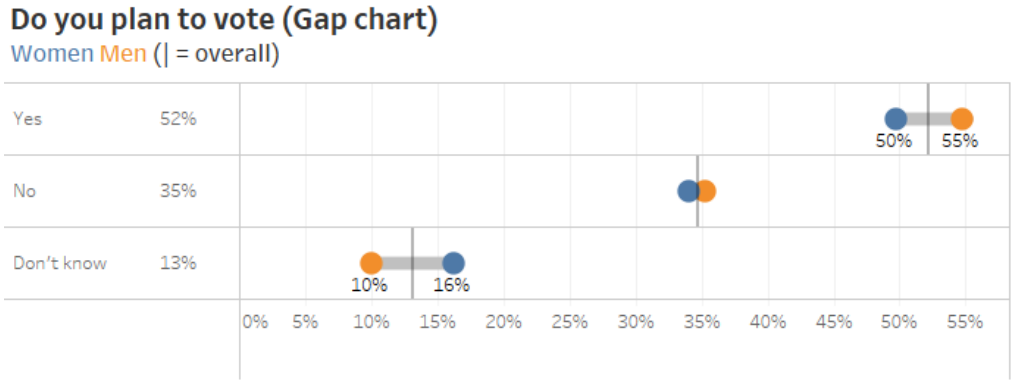
5. Click **OK**.
6. Drag **Percent of Total (Weighted)** on top of the **Percent of Total** pill on the Columns shelf. The bar lengths and the numbers may not be correct; we should check the "Compute Using" setting.
7. Right-click the **Percent of Total (Weighted)** pill on the Columns shelf, select **Compute Using** and select **Labels**.
8. Rename the sheet **Do you plan to vote (Weighted / Gender**) and save your work.

Creating a Gap Chart (connected dot plot)

Aka, a connected dot plot, a barbell chart, or a dumbbell chart.

That "Yes / No / Don't know" chart we showed earlier. Is there a better way to show it?

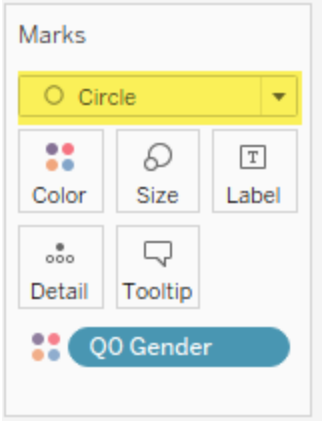
I think yes. Consider the chart below.



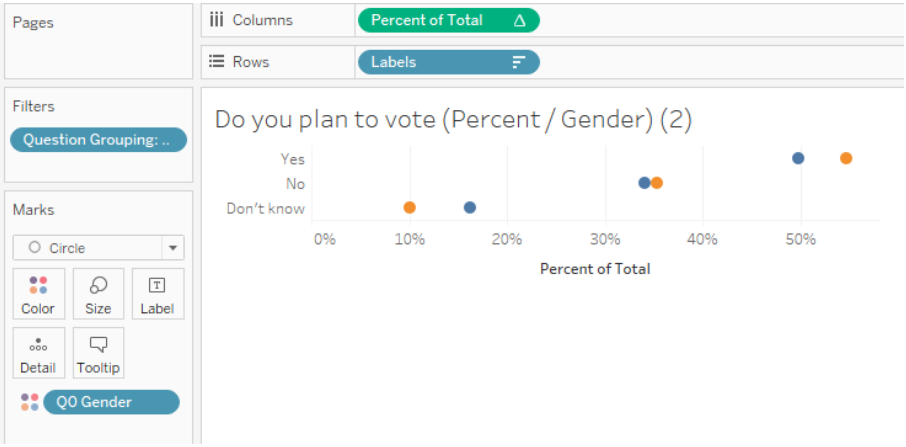
Note: If you had difficulty completing the previous exercise, open the file **2a_Vote.twbx** from the **Starter** folder and work with that.

To create a gap chart

1. Right-click **Do you plan to vote (Percent / Gender)** and select **Duplicate** (note this is NOT the weighted version you just explored).
2. Rename the **sheet Do you plan to vote (Generalized).**
3. Drag Q0 Gender from the Rows shelf onto Color.
4. Turn **Mark Labels** off (click the  icon in the toolbar).
5. Chance the Mark type to Circle.

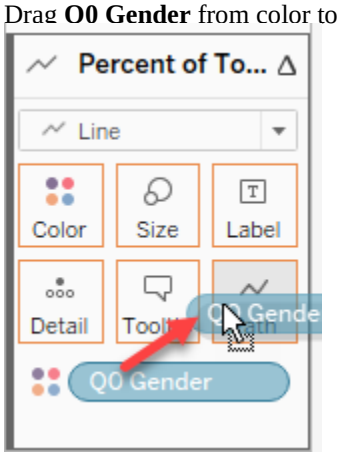


Your screen should look like this.

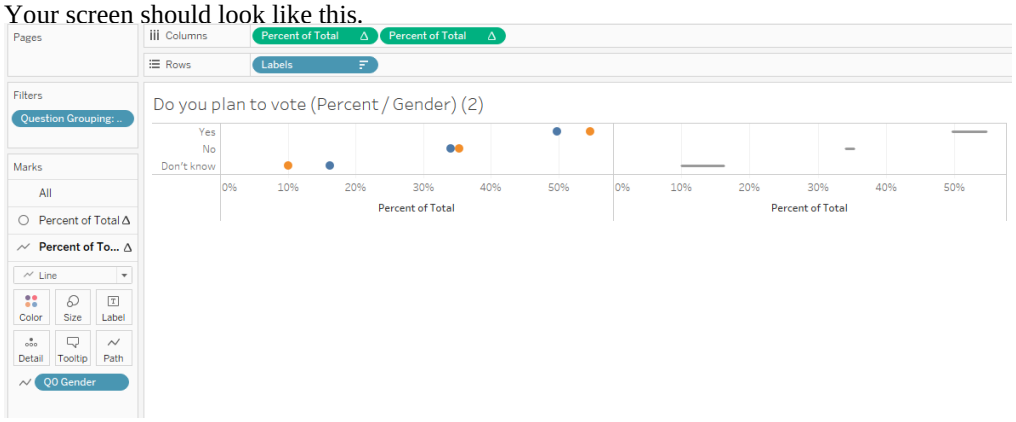


6. While pressing the CTRL key, select **Percent of Tot**al on the Columns shelf and drag to the right to duplicate the chart (yes, it looks redundant).
7. Select the second **Percent of Total** field.
8. Change the Mark type to Line. Notice that the chart looks even sillier.

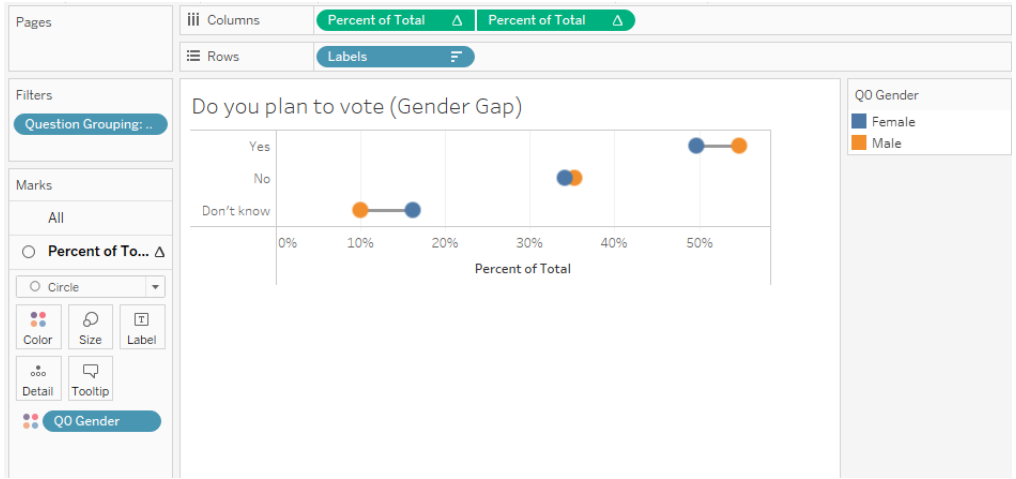
9. Drag **O0 Gender** from color to Path.



Your screen should look like this.



10. Right-click the second **Percent of Total** field on Columns and select **Dual Axis**.
11. Right-click the axis along the top and select **Synchronize axis**.
12. Right-click the axis along the top and select **Move marks to back**.
13. Right-click the axis along the top and de-select **Show Header**.
14. Rename the sheet **Do you plan to vote (Gender gap)** and save your work. Your screen should look like this.



Creating a Flexible Break Down using a Parameter

One of the more interesting things to explore with survey data is to see if there are differences when you cut the responses by various demographics (e.g., Gender, Generation, etc.).

In this next example we'll make it so that we can select a parameter that will allow us to see the survey results by any of the following dimensions:

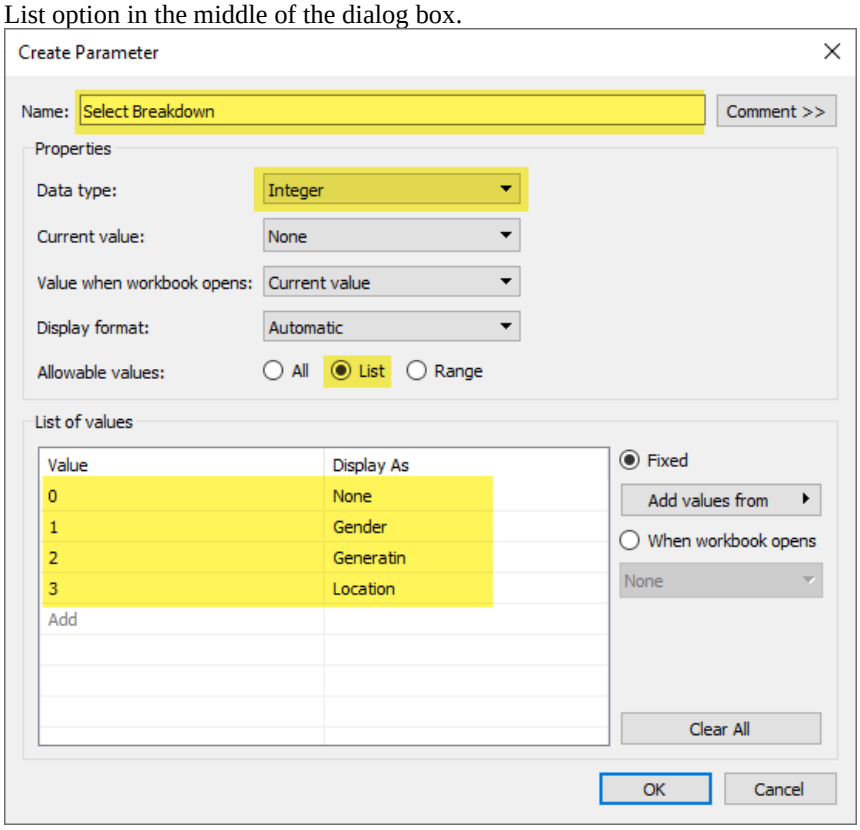
- Gender
- Generation
- Location

Note: The following examples will work for both the weighted and unweighted examples.

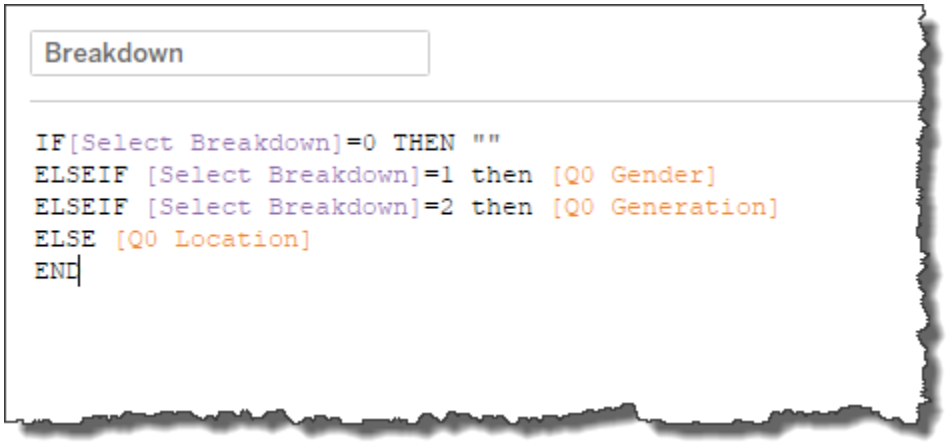
Note: If you had difficulty completing the previous exercise, open the file **2b_Vote_Gender-Gap.twbx** from the **Starter** folder and work with that.

To Create a Parameter-Driven Breakdown

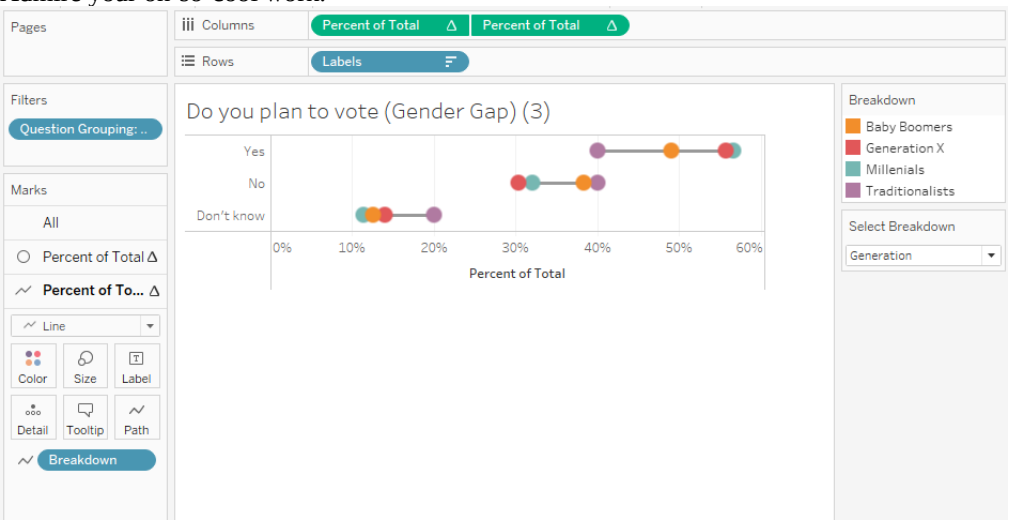
1. Right-click the current sheet and select **Duplicate**.
2. Create a parameter called **Select Breakdown** and define it as follows and check the List option in the middle of the dialog box.



3. Click **OK** when you are done editing the parameter.
4. Right-click the newly-created **Select Breakdown** parameter in the bottom left portion of your screen and select **Show Parameter Control**.
5. Create a new calculated field called **Breakdown** and define it as follows.



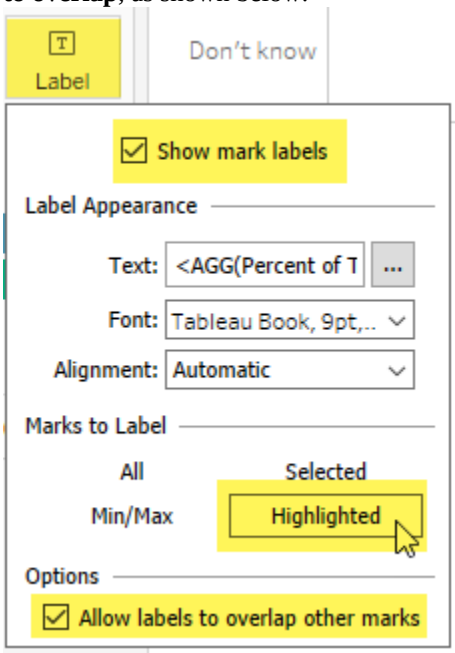
6. Click **OK** when you are finished editing.
7. Select the first Percent of Total field on the Columns shelf.
8. Drag **Breakdown** from the Data Pane and place it on top of Q0 Gender to replace it.
9. Select the second Percent of Total field on the Columns shelf.
10. Drag Breakdown onto Path.
11. Change the **Select Breakdown** parameter to **Generation**.
12. Admire your oh-so-cool work.



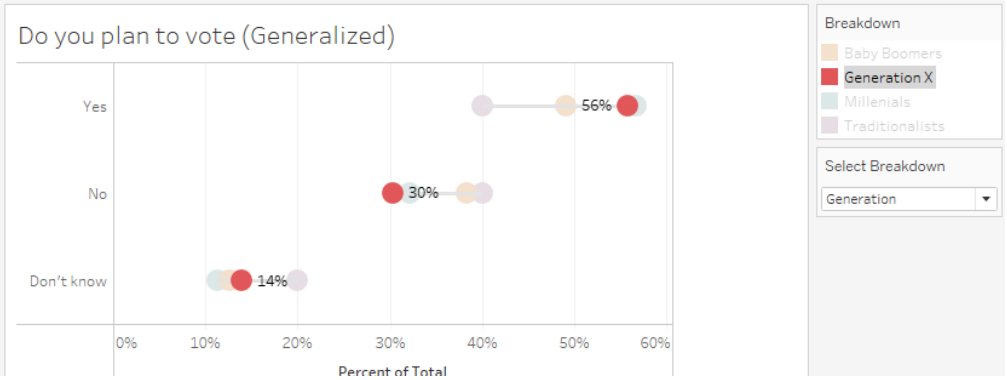
13. Save your work.

Bonus exercise: Using the color legend to highlight marks and show measures

1. Select the first Percent of Total pill on the Columns shelf.
2. Ctrl-drag the pill onto Label on the Marks card.
3. Click the Label button.
4. Indicate you want to **Show mark labels** when **highlighted**, and **allow mark labels to overlap**, as shown below.



5. Select an item from the color legend and see how both the mark and mark label get highlighted.



Topics for Discussion

- Would the labels look better if they were underneath, instead of next to the marks? How can you change that?

- How could we also show the overall percentages (like what happens if we change the breakdown to none)?
- Some of these segments have low response counts (e.g., Antarctica). How can we see this?
- Did you notice the null values?

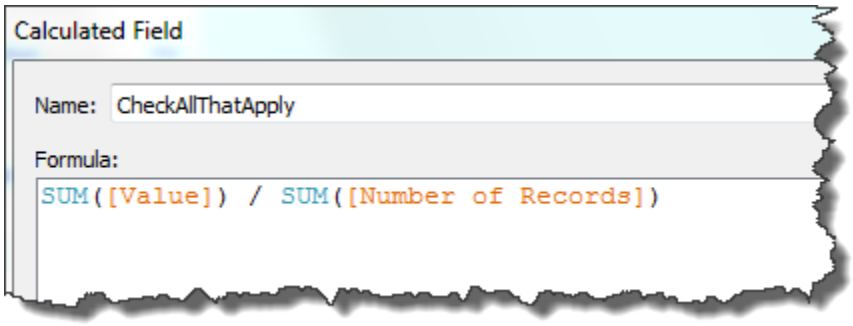
Visualizing Check-All-That-Apply Questions

If you recall looking at the underlying data, the possible values for the check-all-apply questions are 0 and 1. This will make fashioning a calculated field very easy. Note that if we only had label results we would have to use conditional statements (e.g., "IF Label = 'Yes' then 1 else 0 END).

To Visualize a Check-All-That-Apply Question

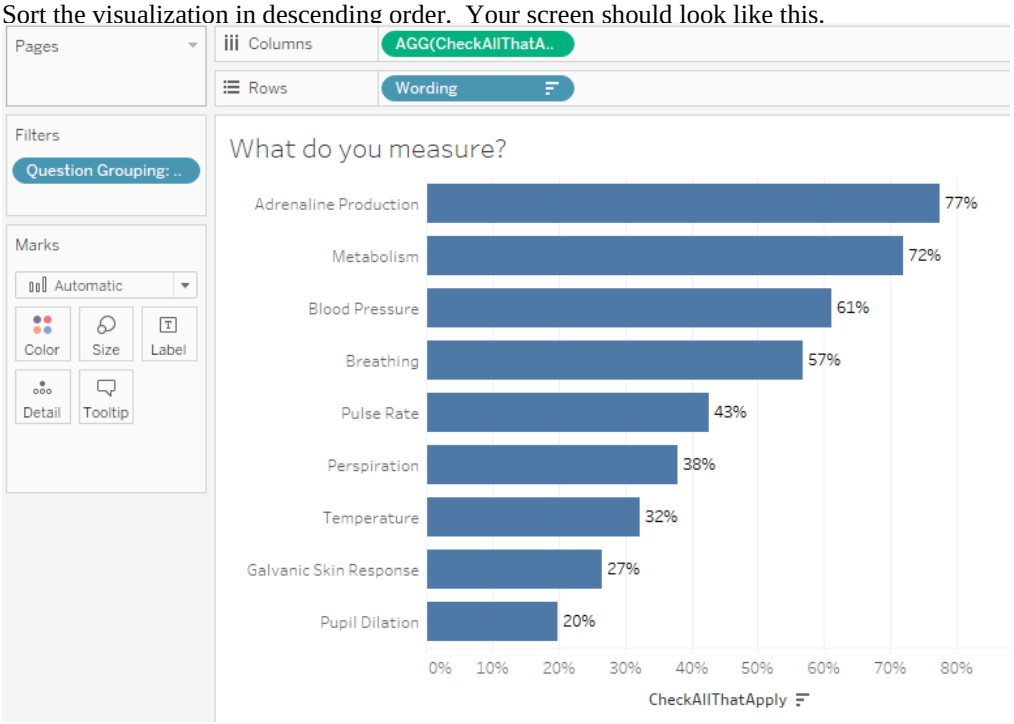
! **Note:** If you had difficulty completing the previous exercise, open the file **2c_VoteGeneralized.twbx** from the **Starter** folder and work with that.

1. Create a new sheet.
2. Drag **Question Grouping** to the Filters shelf and select **What do you measure** from the dialog box.
3. Drag **Wording** to the Rows shelf.
4. Create a new calculated field called **CheckAllThatApply** and define it as follows.



5. Drag **CheckAllThatApply** to the Columns shelf.
6. Turn **Mark Labels** on (click the little "T" icon at the top of your screen).
7. Right-click **CheckAllThatApply** under Measures and select **Default Properties | Number Format**.
8. Change the Default Number Format to **Percentage, 0 Decimal** places and click **OK**.

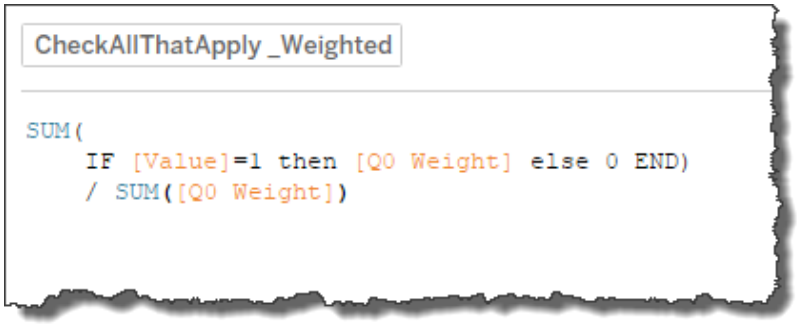
9. Sort the visualization in descending order. Your screen should look like this.



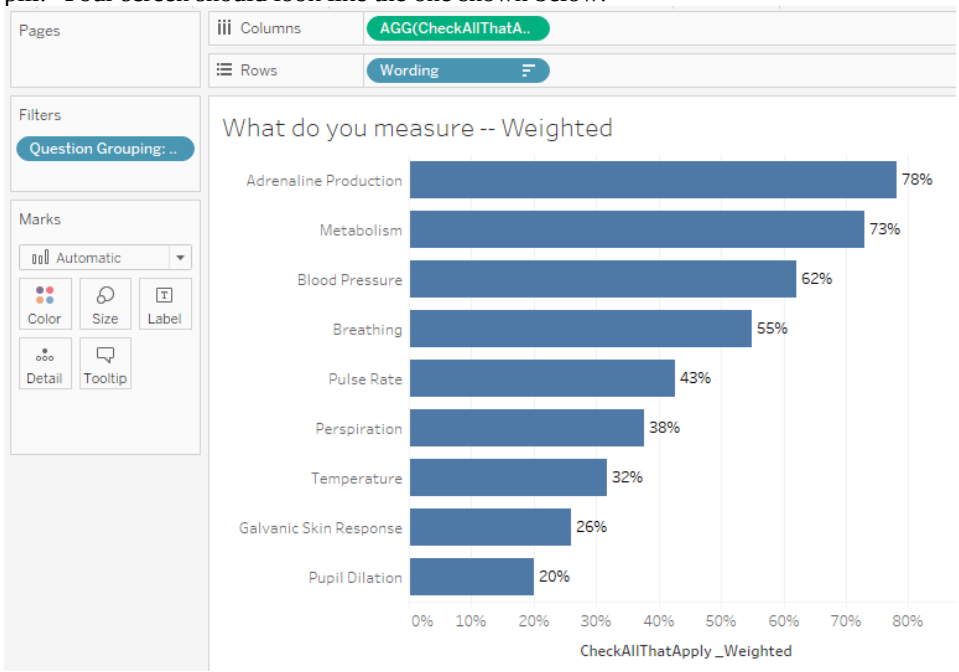
10. Rename the tab **What do you measure** and click save your work.

To Create a Weighted Check-All-That-Apply Field

1. Right-click the What do you measure tab and select **Duplicate**.
2. Right-click CheckAllThatApply under Measures and select **Duplicate**.
3. Right-click the newly created field and select **Edit**.
4. Change the name to **CheckAllThatApply_Weighted** and modify the calculation as shown below.



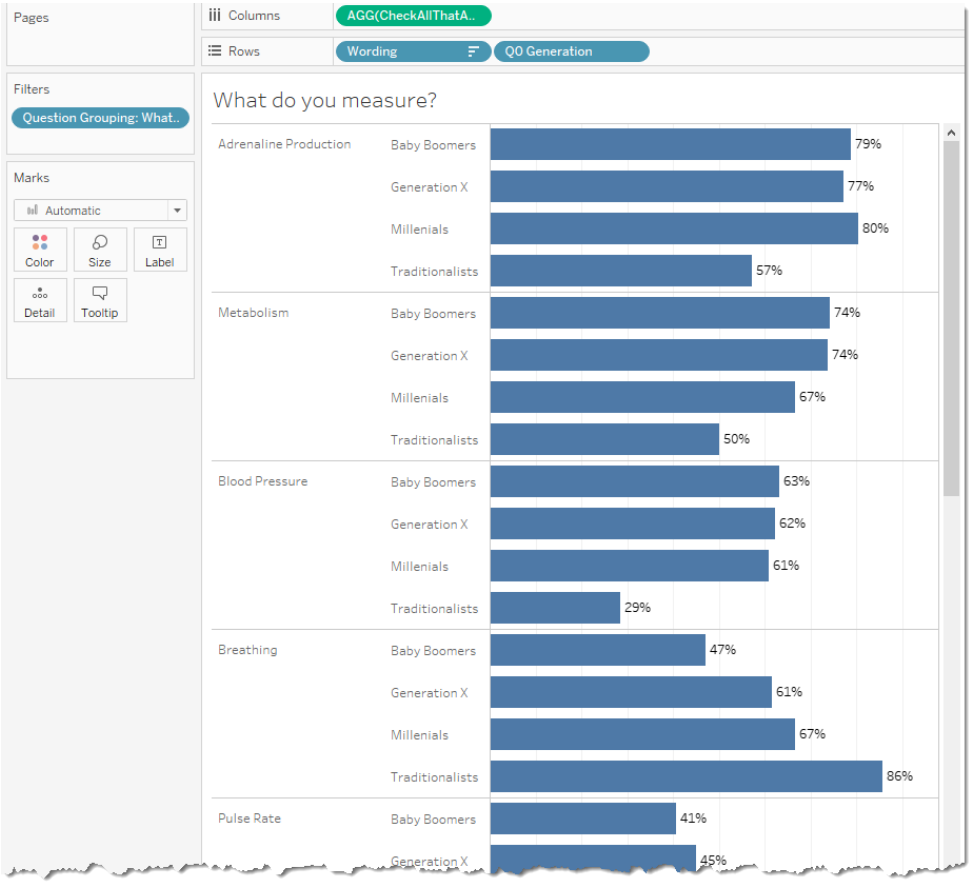
5. Click **OK**.
6. Drag the newly-created field to the Columns shelf so that you replace the previous pill. Your screen should look like the one shown below.



7. Rename the tab **What do you measure – Weighted** and save your work.

Creating a Gap Chart for a Check-All-That-Apply Question

If you were to drag Gender or Generation onto the Rows shelf, you'll see there are some pretty big differences in responses from the different groups.



All the bars are difficult to interpret (and they take up a lot of screen real estate). As the Gap chart worked so well for the Vote question, let's try it for the "what do you measure" question as well.

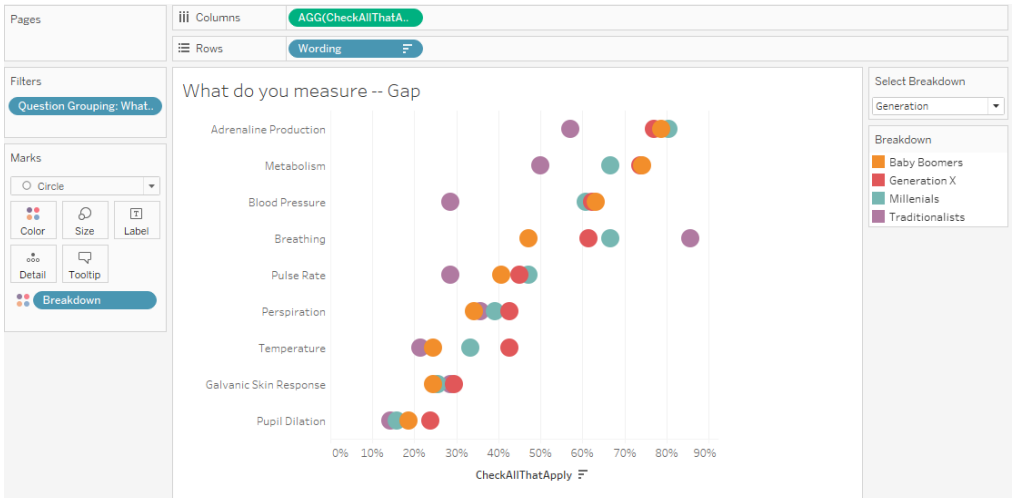
To Create a Gap Chart for a Check-All-That-Apply Question

!

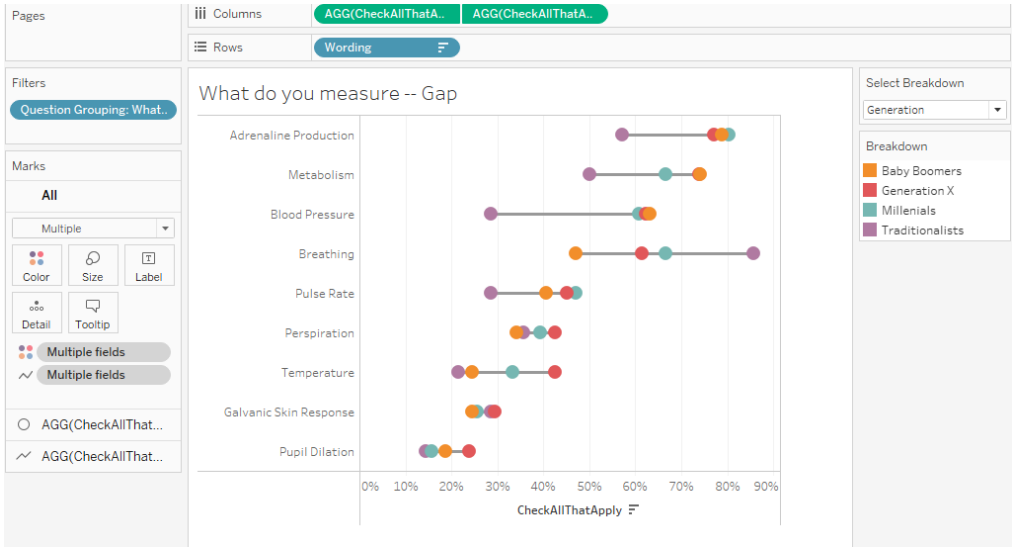
Note: If you had difficulty completing the previous exercise, open the file **2d_CheckAll.twbx** from the **Starter** folder and work with that.

1. Right click the **What do you measure** tab and selected **Duplicate**.
2. Rename the sheet **What do you measure – Gap**.
3. Right-click the **Select Breakdown** parameter and choose **Show Parameter.**
4. Change the Mark type to **Circle**.
5. Turn Mark Labels off (the little "T" icon on the toolbar.

6. Drag **Breakdown** onto **Color**. Your screen should look like this.

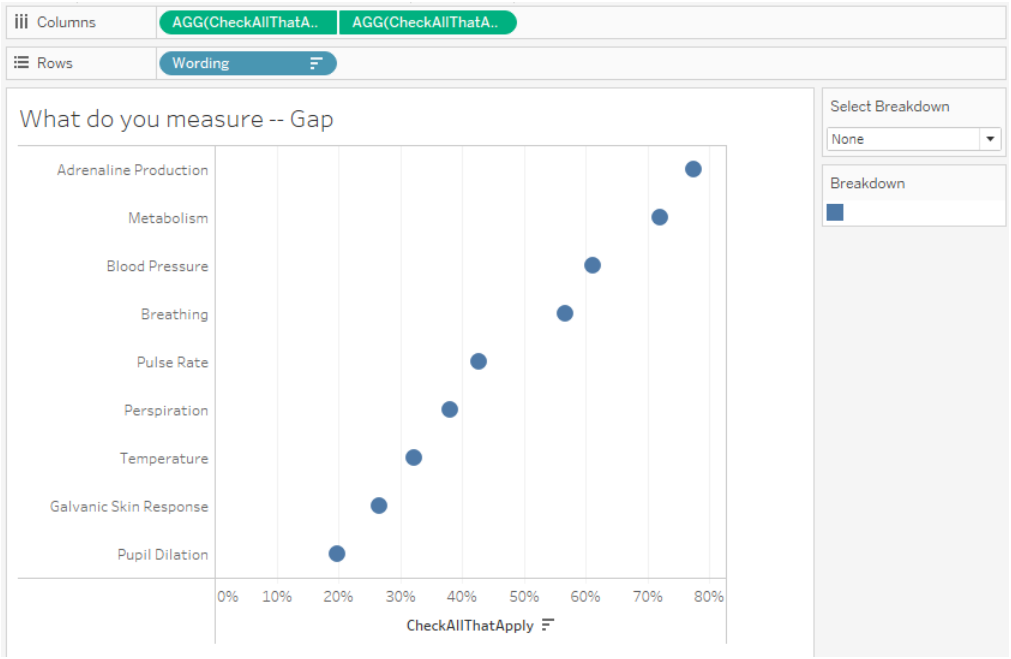


7. Ctrl drag the **CheckAllThatApply** pill that is on Columns to the right so that you have the field shown twice.
8. Select the second instance of **CheckAllThatApply** and change the mark type to **Line**. Yes, the second instance now looks strange.
9. For the Line chart, drag **Breakdown** from Color and place it on **Path**.
10. Right-click on the axis for the line chart and select **Dual Axis**.
11. Right-click the axis along the top and select **Synchronize Axis**.
12. Right-click the axis along the top and select **Move marks to back**.
13. Right-click the axis along the top and de-select **Show Header**. Your screen should look like this (note that you may have to adjust the size of the circles using the Size button on the Marks card).



Adding an Overall Reference Line

Supposed we wanted to see the overall responses. We could change the Select Breakdown parameter to none, as shown here.



But supposed you wanted to be able to see this when you were exploring the gaps by Gender, Generation, etc.?

A great way to do this is to add an overall reference line. Let's see how to do that using a Level of Detail (LoD) expression.

To Add an Overall Reference Line

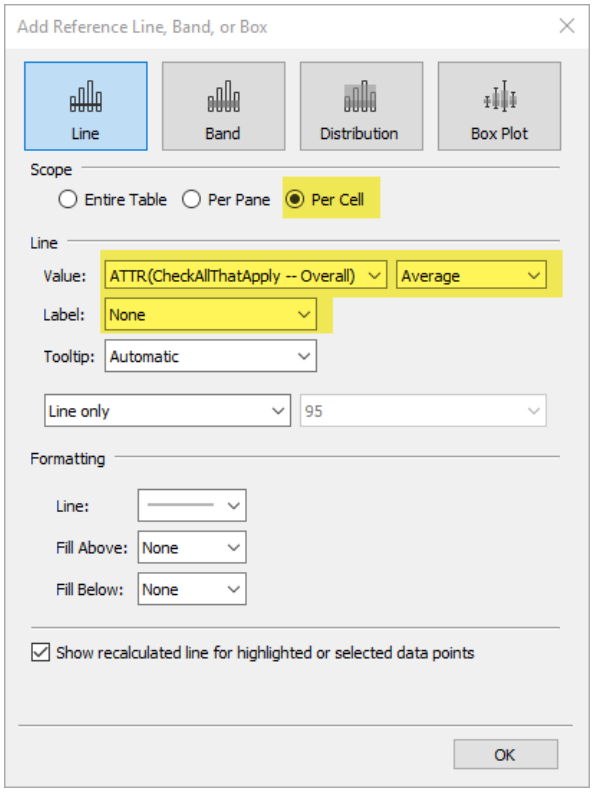
1. Right-click the field **CheckAllThatApply** and select **Create,** then **Calculated Fiel**d.
2. Create a new field called **CheckAllThatApply – Overall** and define it like this.

   CheckAllThatApply -- Overall

   ```
   {Exclude [Breakdown]: [CheckAllThatApply]}
   ```

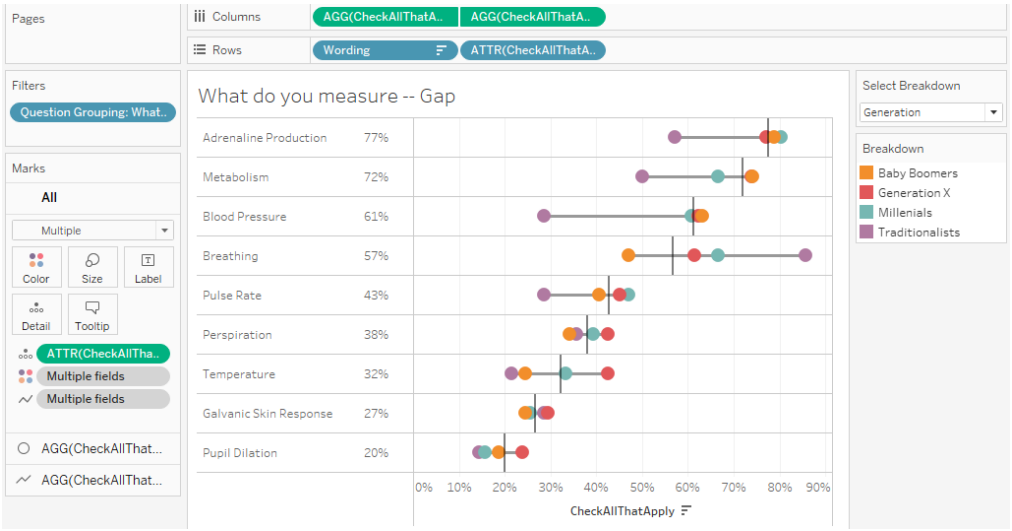
   This translates as "yes, there will be separate dots for the whatever demographic breakdown you selected, but for this calculation ignore them and imagine there is no breakdown."
3. Drag this new field onto the **Detail** button.
4. Right-click in the axis along the bottom and select **Add Reference Line**.

5. Change the Scope to **Per Cell**.
6. Change the Value to **ATTR(CheckAllThatApply – Overall)** and the Label to **None**. The dialog box should look like this.



7. Click **OK**.
8. Drag **CheckAllThatApply – Overall** onto Rows.
9. Right-click this field and select **Discrete**.

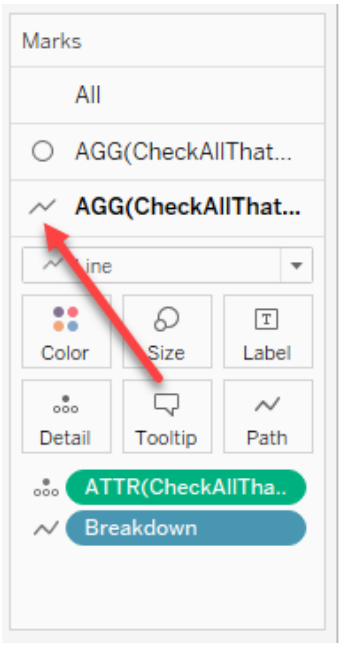
10. Right-click **CheckAllThatApply – Overall** in the Data pane and change the default number format to Percentage with 0 decimal places. Your screen should look like this.



11. Save your work.

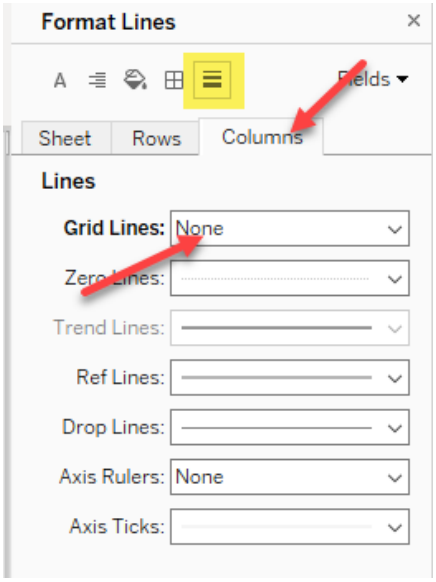
Bonus Exercise – Tweaking the Formatting

1. Select the Second instance of CheckAllThatApply so that you are dealing with the Line chart and not the circle chart.

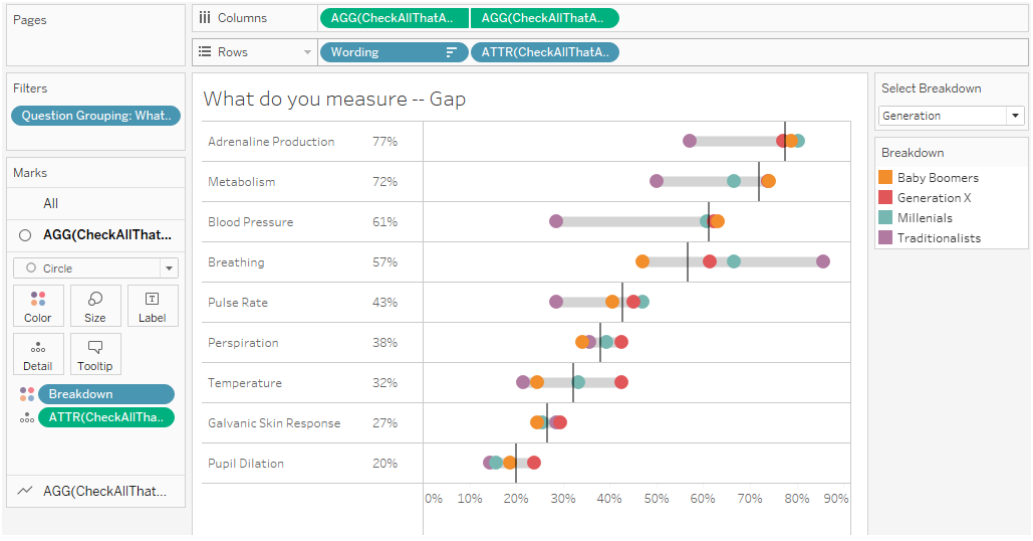


2. Click the Size button on the Marks card and drag it to the right so that the line is almost as wide as the circles.

3. Click the Color button and change the line color to a light gray.
4. Right click the axis along the bottom and select Edit Axis.
5. Remove the Axis title and close the dialog box.
6. Right-click in the middle of the chart and select Format.
7. Change the Line formatting so that there are no gridlines for Columns.



8. Close the Formatting pane.
9. Save your works. Your screen should look like this.



!

Note: Before continuing, see if you can make it so that when you click an item in the color legend the appropriate circle – and the amount associated with it – appears.

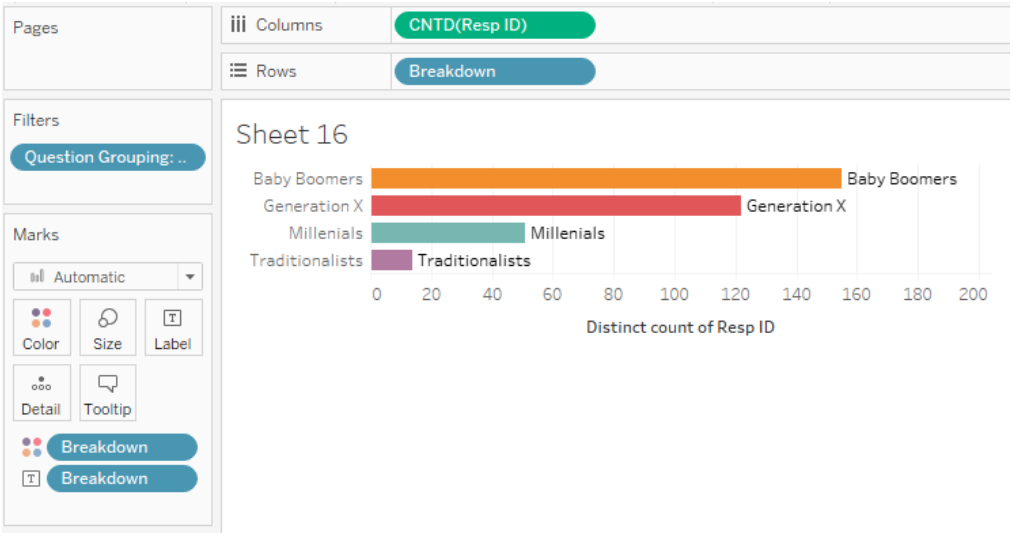
Creating a Check-All-That-Apply Dashboard

For this next series of examples, we'll create bar chart showing the number of responses for the check-all-that-apply question and combine it with the gap chart to fashion an interactive dashboard.

To Create the Demographics Bar Chart

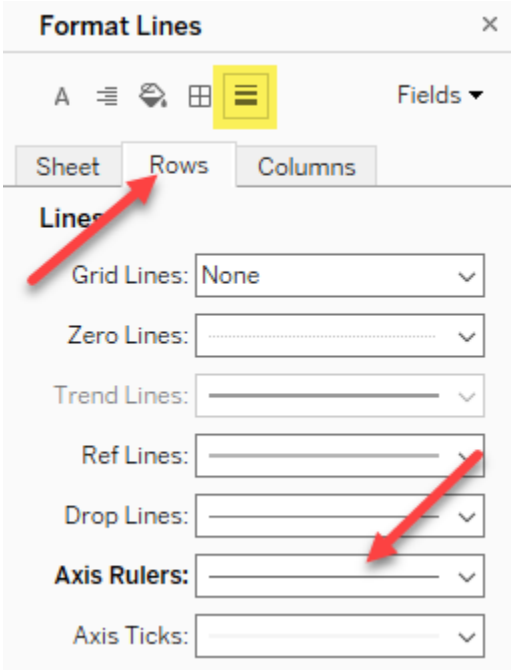
Note: If you had difficulty completing the previous exercise, open the file **2e_Check-All_Gap.twbx** from the **Starter** folder and work with that.

1. Create a new sheet.
2. Add **Question Grouping** to the Filters shelf and select **What do you measure**.
3. Place Breakdown on Rows.
4. Right-drag **Resp ID** onto Columns and select **CNTD(Resp ID)** from the Drop Field dialog box.
5. Drag **Breakdown** onto **Color**.
6. Drag **Breakdown** onto **Label**. Your screen should look like this.

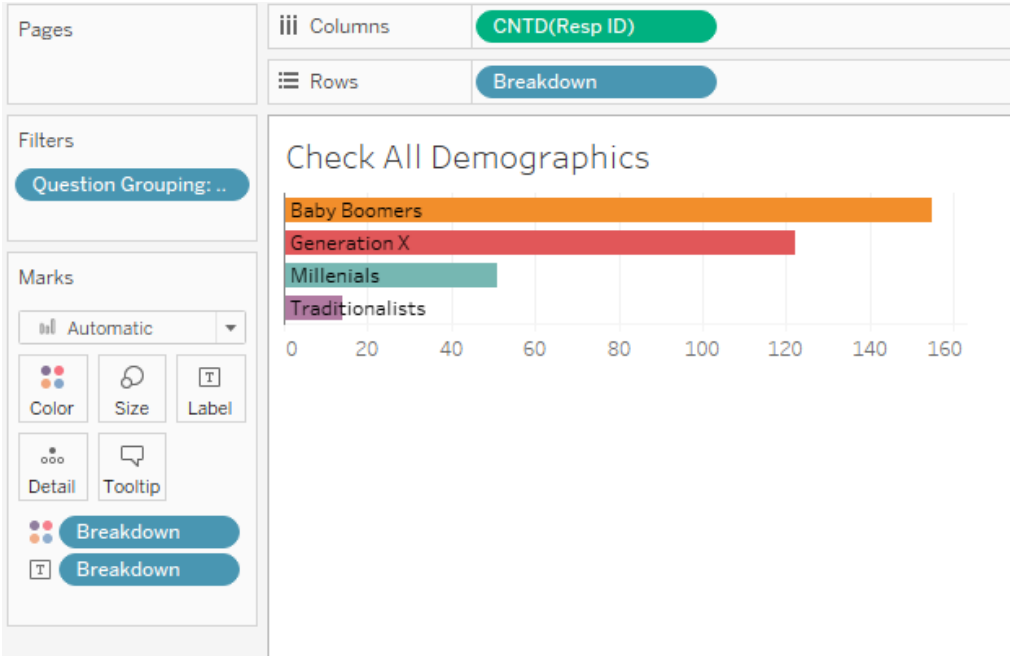


7. Right-click in the "header" along the left side of the chart (where is reads "Baby Boomers", "Generation X," etc.) and de-select **Show Header**.
8. Click the **Labels** button on the marks card and change the alignment to **Left**.
9. Rename the sheet **Check All Demographics**.
10. Right-click the Axis and select **Edit Axis**.
11. Delete the Axis title and close the dialog box.
12. Right-click in the middle of the chart and select **Format**.

13. In the **Format Lines** section, select Rows and indicate that you want a **solid dark gray Axis Ruler**.



14. Close the Format pane. Your screen should look like this.

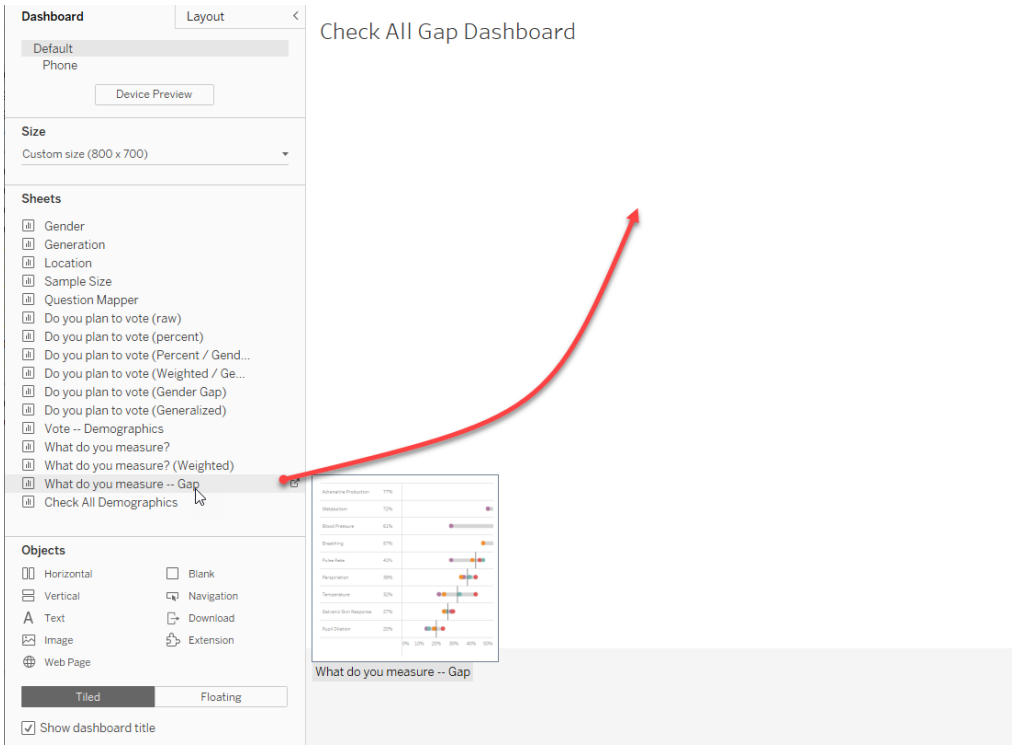


15. Save your work.

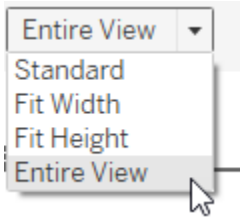
To Create the Check All That Apply Dashboard

Note: If you had difficulty completing the previous exercise, open the file **2f_Check-All_Gap_Demo.twbx** from the **Starter** folder and work with that.

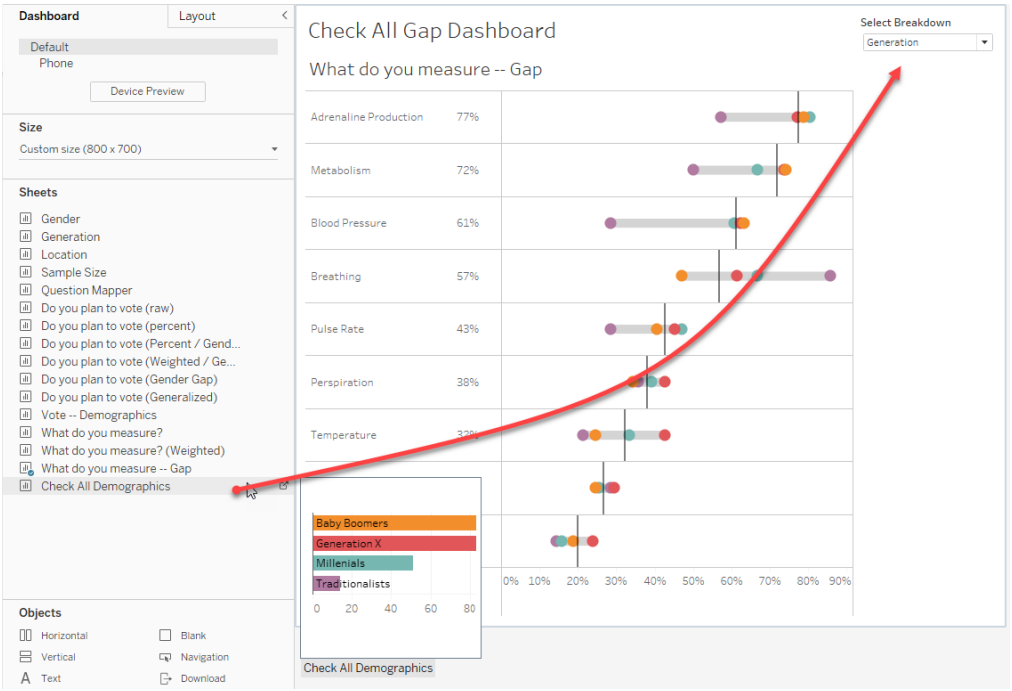
1. Create a new Dashboard and indicate you want the size to be 800 wide by 700 tall.
2. Indicate you want to **Show the dashboard title**.
3. Rename the dashboard tab **Check All Gap Dashboard.**
4. Drag **What do you Measure – Gap** from the Sheets section into the Dashboard.



5. Remove the color legend.
6. Click in the What do you measure – Gap sheet and select **Entire View** from the drop-down menu on the tool bar.

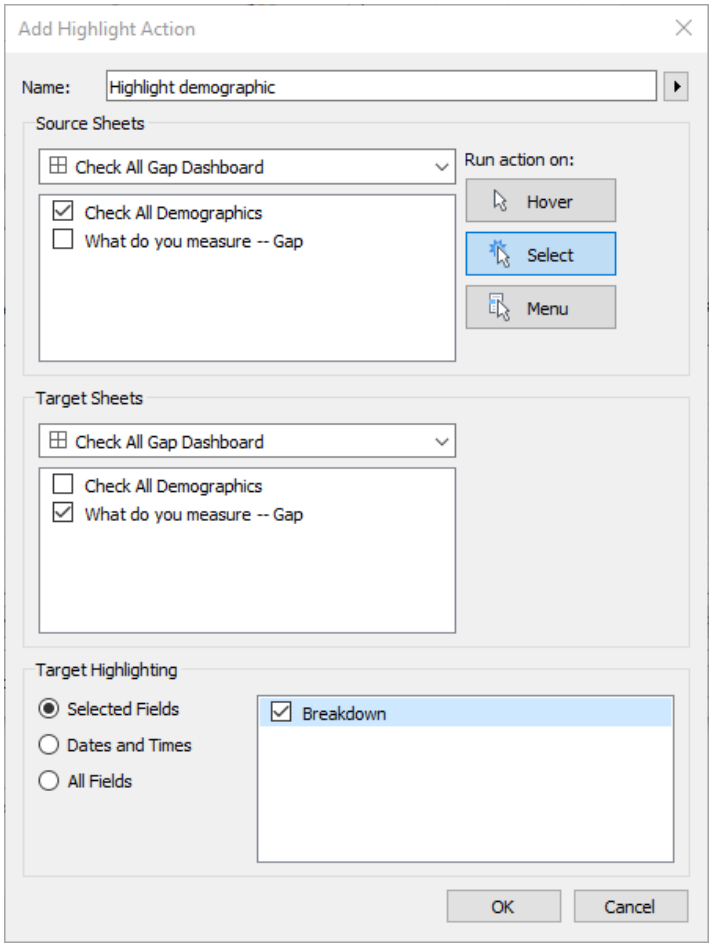


7. Drag **Check All Demographics** below the Select Breakdown parameter.



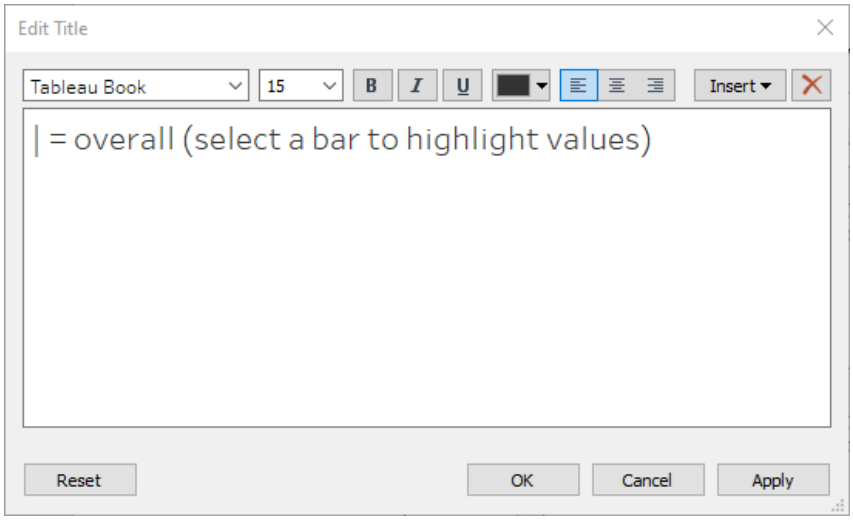
8. Right-click the title for the sheet you just dropped and select **Hide Title**.
9. Select **Fit Width** from the fit drop-down menu in the tool bar.
10. Remove the color legend.
11. Click the **Dashboard** menu and select **Actions**.
12. Click **Add Action** and select **Highlight**.

13. Fill in the Add Action dialog box as shown below.



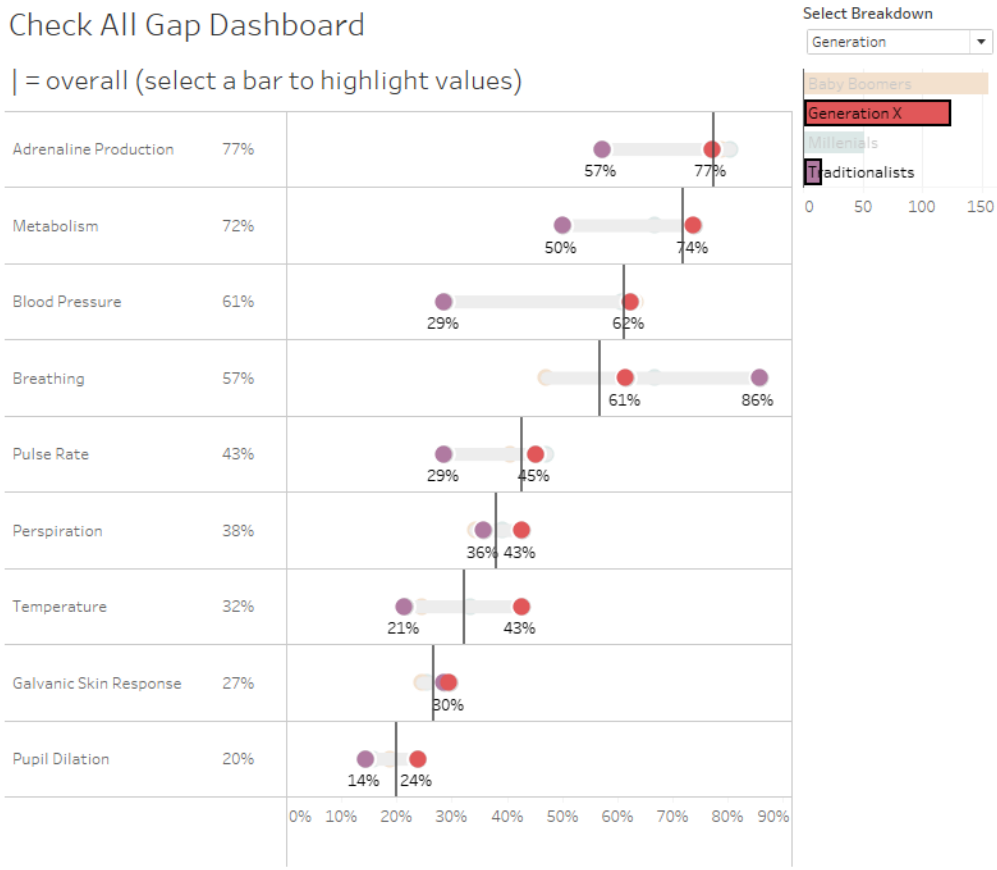
14. Click **OK**, then click **OK** again.

15. Double-click in the sheet title and edit it as shown below.



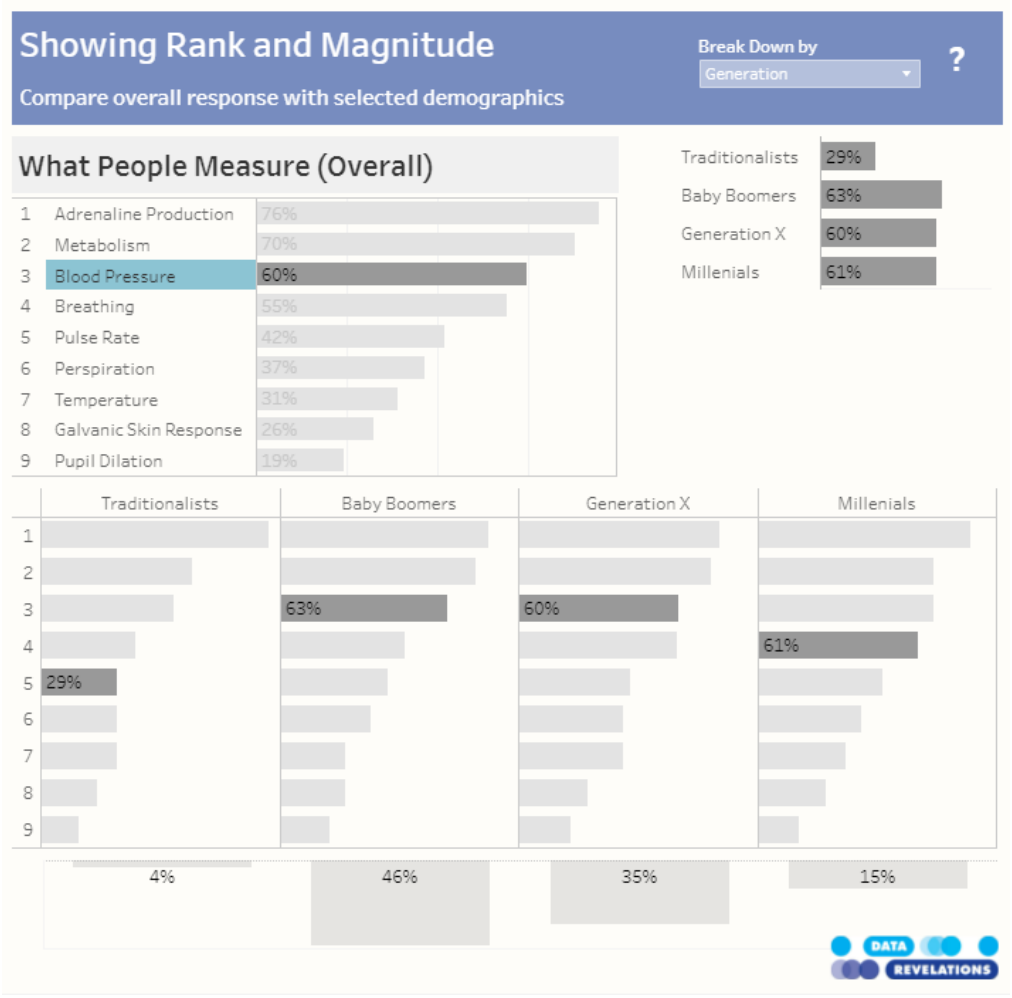
16. Click **OK** and save your work.

Now, take your own advice and try clicking some bars and changing the Select Breakdown parameter. Your screen should look like this.



Are There Other Ways to Show This?

Yes. You could build an interactive dashboard that allows you to show rank and magnitude, like the one shown here.



I prefer the gap chart / connected dot plot, but if you want to explore the rank and magnitude approach, see

https://www.datarevelations.com/resources/howmany/

and

https://www.datarevelations.com/resources/visual-ranking-within-a-category/

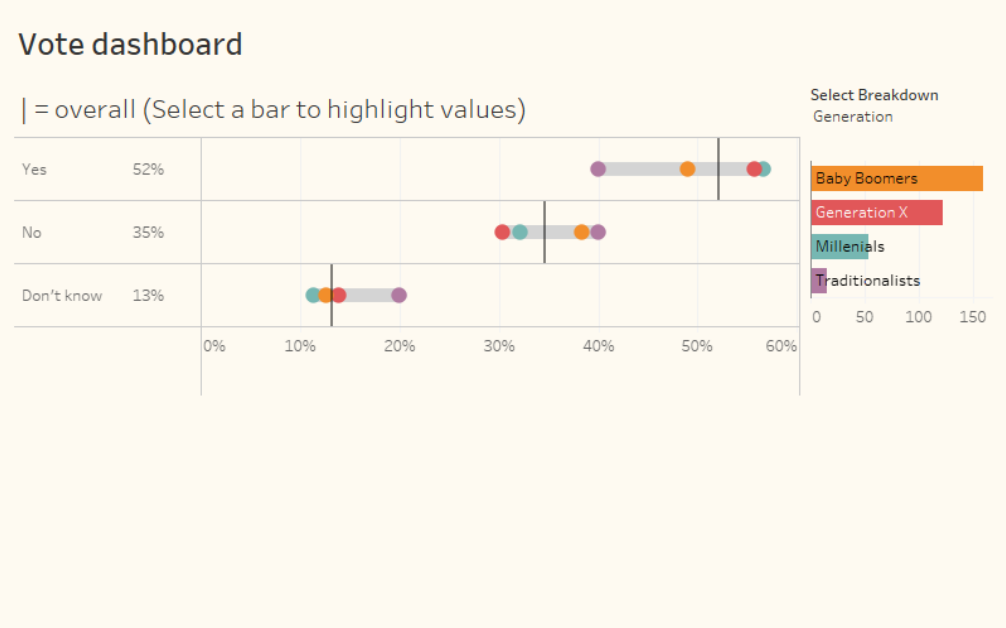
Topics for Discussion

The overall reference line is a valuable addition, as is the bar chart showing the response count by the selected demographic.

Would this same approach be valuable for the single punch question we looked at earlier?

Absolutely.

If you go into the Completed folder you will find a packaged workbook called **DataRevelations_SurveyData_Completed_V4.twbx**. Open this file and go to the tab called **Vote dashboard**.



If we look under the hood, we'll see that there's a somewhat more complicated field that figures out the overall percentage for Yes, No, and Don't know:

```
SUM({EXCLUDE  [Breakdown]: SUM([Number of Records])}) /
SUM({Exclude [Labels],  [Breakdown]: SUM([Number of Records])})
```

The numerator portion translates as "please ignore the dots [Breakdown] and determine how many responses there are for "Yes", how many for "No", etc.

The denominator portion translates as "ignore the dots and ignore the separate rows; just figure out the total number of people that answered the question."

Could we have used a table calculation for this? Yes, but it's tricker as you'll need one field for the numerator and another for the denominator so you can control the scope of the calculation.

You could also use a different LoD expression (one that uses {FIXED}).

Mind the Gap? Mind the Margin of Error!

The Gap chart and its ability to compare demographics is great but realize that if your response count for demographic segments is low, the margin of error is going to be **high**.

If you have several thousand people responding to a survey and have 400+ in each of the demographics you are comparing, then you are probably fine. If the sample response count is low – and 41 for traditionalists is low – then reporting on gaps may not be warranted. PLEASE visit www.datarevelations.com and do a search on "confidence" and "margin of error"!

III. Getting Your Data Setup "Just So"

Overview

Over the past several in writing about visualizing survey data using Tableau I've found that the number one impediment to success is getting the data in the right format.

We'll look at using Tableau Prep for setting up the data.

What do I mean by "just so"

When I deal with survey data there are usually four different elements that need to fit together:

1. The demographic information (e.g., age of respondents, gender, etc.)
2. Survey responses in text format
3. Survey responses in numeric format
4. Meta data that describes the survey data.

Let's see what the four elements look like using an Excel sample data set (click here to download).

Demographic data

Here's what the demographic data looks like.

	A	B	C	D	E
1	RespID	Q0_Gender	Q0_Location	Q0_Generation	Q0_Weight
2	2	Male	South America	Generation X	1
3	4	Female	South America	Baby Boomers	1.44
4	5	Female	South America	Generation X	1
5	6	Male	Antarctica	Baby Boomers	1.44
6	9	Female	Europe	Baby Boomers	1.32
7	12	Female	Europe	Baby Boomers	1.56
8	15	Male	North America	Baby Boomers	1.56
9	16	Male	Antarctica	Baby Boomers	1.44
10	17	Female	Europe	Baby Boomers	1.32
11	18	Male	North America	Traditionalists	0.595
12	22	Male	South America	Generation X	1.32
13	25	Female	South America	Generation X	1.32

Survey responses in text format

Here are several columns of survey responses in text format. Column F contains data for a Yes / No / Don't know question. Column G contains responses for a question about salary. Columns H through P are responses for check-all-that apply questions and columns Q and R contain Likert scale responses.

	A	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	RespID	Q1	Q100	Q2_1	Q2_2	Q2_3	Q2_4	Q2_5	Q2_6	Q2_7	Q2_8	Q2_9	Q3_1	Q3_2
2	2	No	$ 98,038	No	No	Yes	No	Yes	No	No	No	Yes	Small degree	Small degree
3	4	No	$ 138,936	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Very high degree	Very high degree
4	5	Yes	$ 84,471	No	Yes	Yes	Yes	No	No	No	No	Yes	Very high degree	High degree
5	6	Don't know	$ 138,534	No	Yes	Yes	No	No	No	No	No	Yes	Very high degree	High degree
6	9	Yes	$ 68,944	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Very high degree	Very high degree
7	12	No	$ 100,663	No	No	Yes	Yes	No	No	No	No	Yes	Moderate degree	Moderate degree
8	15		$ 122,481											
9	16	Yes	$ 106,036	Yes	Yes	No	No	No	Yes	No	No	Yes	Moderate degree	Very high degree
10	17	Don't know	$ 81,681	Yes	Yes	Yes	No	No	No	No	No	Yes	High degree	High degree
11	18	No	$ 104,200	No	Yes	No	No	No	Yes	No	No	Yes	Moderate degree	High degree
12	22	No	$ 172,723	No	No	Yes	Yes	No	No	No	No	No	High degree	High degree
13	25	Yes	$ 153,410	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Small degree	High degree
14	26	Yes	$ 93,194	No	Yes	No	No	No	Yes	No	No	Yes	High degree	High degree
15	27	Yes	$ 101,662	Yes	Yes	Yes	No	No	No	No	Yes	No	Very high degree	High degree
16	29		$ 114,216										Very high degree	Very high degree
17	30	No	$ 97,354	No	No	No	No	Yes	No	No	No	No	Moderate degree	High degree

Survey responses in text format

Survey responses in numeric format

Here are the same responses but in numeric format.

	A	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	RespID	Q1	Q100	Q2_1	Q2_2	Q2_3	Q2_4	Q2_5	Q2_6	Q2_7	Q2_8	Q2_9	Q3_1	Q3_2
2	2	0	98,037.68	0	0	1	0	1	0	0	0	1	2	2
3	4	0	138,935.50	1	1	1	0	0	0	1	0	1	5	5
4	5	1	84,471.00	0	1	1	1	0	0	0	0	1	5	4
5	6	2	138,533.55	0	1	1	0	0	0	0	0	1	5	4
6	9	1	68,943.60	1	1	1	1	0	1	1	1	0	5	5
7	12	0	100,663.20	0	0	1	1	0	0	0	0	1	3	3
8	15		122,480.76											
9	16	1	106,035.60	1	1	0	0	0	1	0	0	1	3	5
10	17	2	81,681.30	1	1	1	0	0	0	0	0	1	4	4
11	18	0	104,199.70	0	1	0	0	0	1	0	0	1	3	4
12	22	0	172,723.10	0	0	1	1	0	0	0	0	0	4	4
13	25	1	153,410.40	1	1	1	1	1	1	1	0	1	2	4
14	26	1	93,194.00	0	1	0	0	0	1	0	0	1	4	4
15	27	1	101,661.78	1	1	1	0	0	0	0	1	0	5	4
16	29		114,215.85										5	5
17	30	0	97,353.72	0	0	0	0	1	0	0	0	0	3	4
18	31	1	120,061.37	0	0	0	0	0	0	1	0	0	3	4

Survey responses in numeric format

I'll explain why it's so useful to have the survey responses in both text and numeric format in a bit.

Meta Data

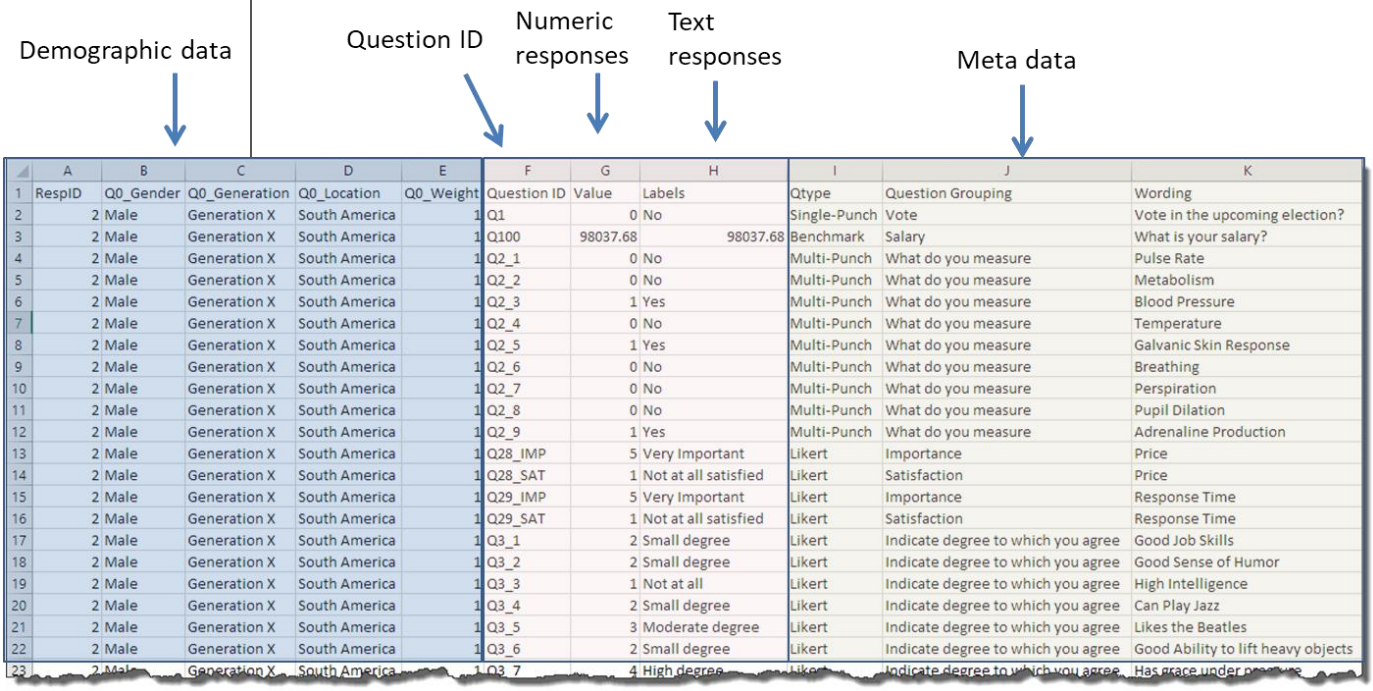
Here's some data that I usually prepare by hand as most survey tools won't product it for me automatically. Having this helps me understand the data and will greatly streamline my work in Tableau.

	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

Survey data meta data. This doesn't take long to create and will be a huge time saver once we get the data into Tableau.

What does "just so" look like?

Our goal is to combine and reshape the various elements so that they look like this.



	A	B	C	D	E	F	G	H	I	J	K
1	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	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	Price
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	Response Time
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 grace under pressure

The key thing is that I no longer have a separate column for each survey response. Indeed, I've reduced the number of columns from 45 to just 11, but I've also increased the number of rows from 845 to well over 20,000. That is a good thing.

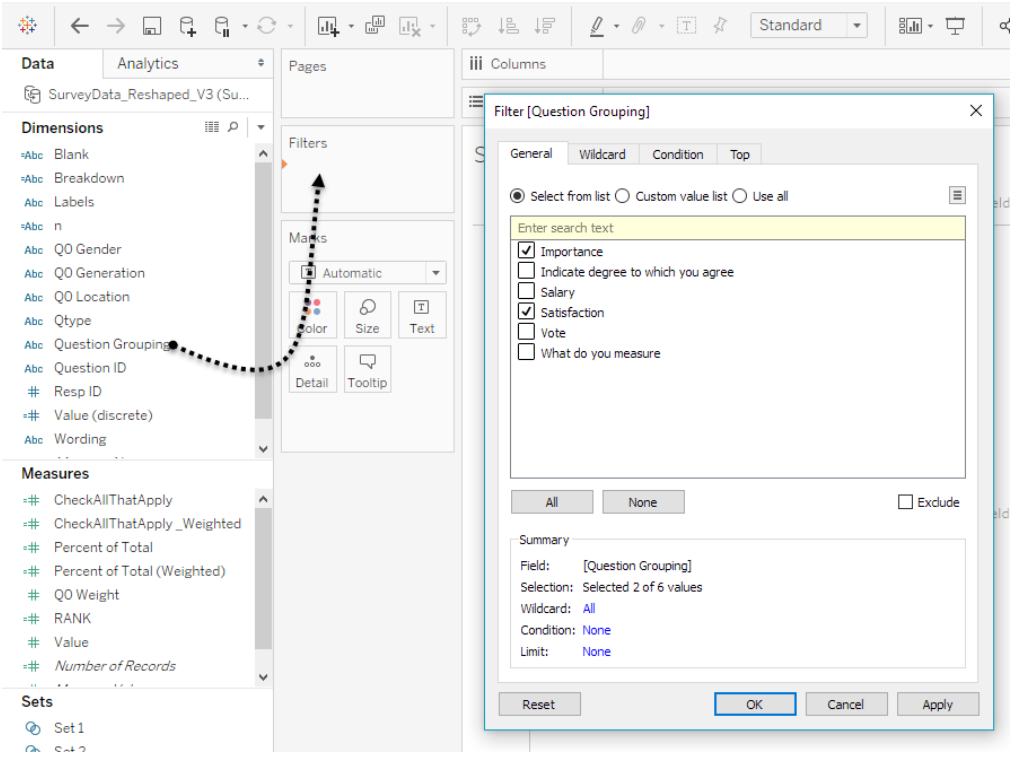
Why this works so well with Tableau

Our goal is to use Tableau Prep (or another tool) to get the data in this format, not to actually use the data, but if you need convincing on why the meta data is so helpful, consider the following example.

Let's say that in your survey you ask people to indicate the importance and satisfaction about certain services, as shown here.

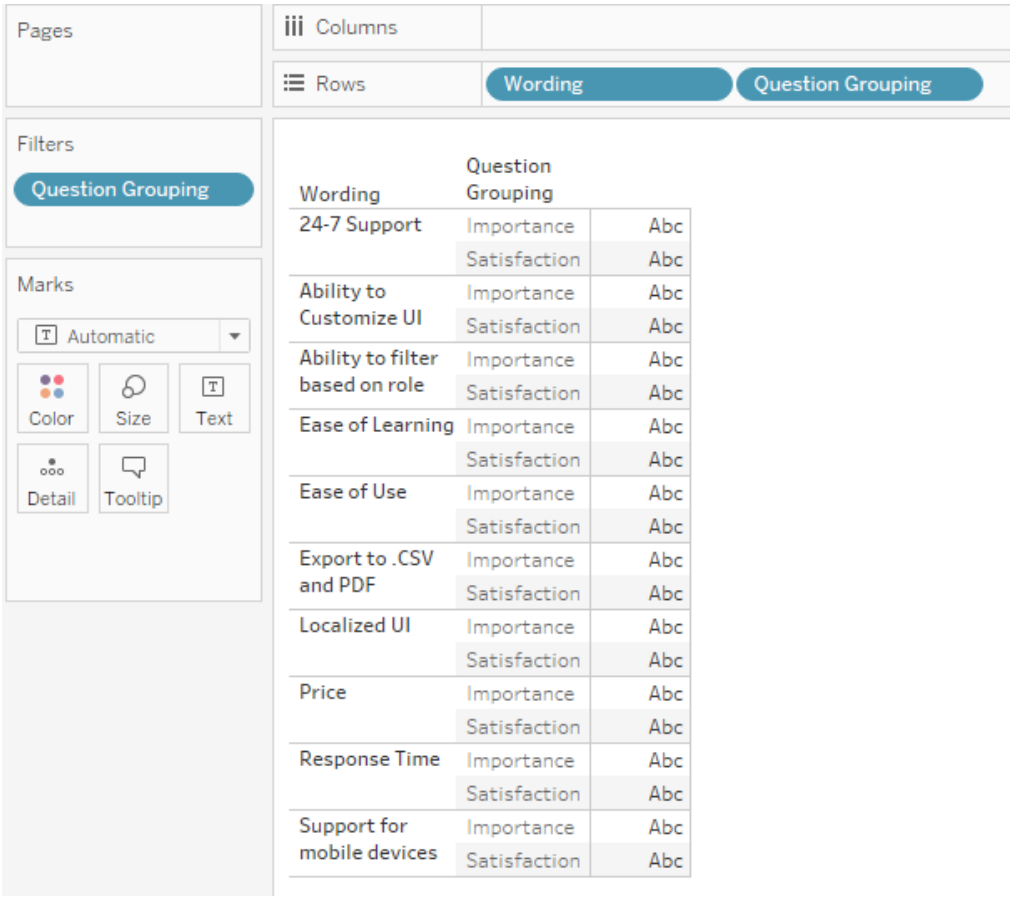
	Importance	Satisfaction
Price	Critical	Satisfied
Response Time	Important	Satisfied
24-7 Support	Critical	Disappointed
Ease of Use	Very Important	Satisfied
Ability to Customize UI	A little impotant	Neutral

With the data set up "just so" conducting this comparison in Tableau becomes easy. First we can drag Question Grouping into Filters and indicate that we just want to look at Importance and Satisfaction questions.



Using the Question Grouping field to just focus on Importance and Satisfaction questions

Then we can drag Wording and Question Grouping onto the Rows shelf which gives us the framework for comparing importance and satisfaction across ten different questions. No more having to "look up" which questions we want to explore and no more having to alias question IDs. I love this!



The helper file meta data provides the framework for comparing questions and building visualizations.

Why do we need both text and numeric results?

We don't really need them, but I know I certainly want them.

Consider all of the Likert scale question results. The universe of possible values are

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 as sometimes a 5 means "Strongly agree", for other questions it means "Critical" and for others it means "Extremely satisfied".

Without having both numeric and text results we will have to write A LOT of IF / CASE statements and I, for one, do not want to do that.

So, now that we understand how and why we want the data "just so" we'll see how to get it that way using Alteryx and using Tableau's Excel Add-In.

Reviewing the Data

Before we plot ahead, let's look at the source data. The file is called **DataRevelations_SurveyData_V4.xlsx** and you can find it in the Source folder on your **USB** drive.

Notice that the file contains three different tabs:

Data Labels

	A	B	C	D	E	F	G	H	I
1	RespID	Q0_Gender	Q0_Location	Q0_Generation	Q0_Weight	Q1	Q100	Q2_1	Q2_2
2	2	Male	South America	Generation X	1	No	$ 98,038	No	No
3	4	Female	South America	Baby Boomers	1.44	No	$ 138,936	Yes	Yes
4	5	Female	South America	Generation X	1	Yes	$ 84,471	No	Yes
5	6	Male	Antarctica	Baby Boomers	1.44	Don't know	$ 138,534	No	Yes
6	9	Female	Europe	Baby Boomers	1.32	Yes	$ 68,944	Yes	Yes
7	12	Female	Europe	Baby Boomers	1.56	No	$ 100,663	No	No
8	15	Male	North America	Baby Boomers	1.56		$ 122,481		
9	16	Male	Antarctica	Baby Boomers	1.44	Yes	$ 106,036	Yes	Yes
10	17	Female	Europe	Baby Boomers	1.32	Don't know	$ 81,681	Yes	Yes
11	18	Male	North America	Traditionalists	0.595	No	$ 104,200	No	Yes
12	22	Male	South America	Generation X	1.32	No	$ 172,723	No	No
13	25	Female	South America	Generation X	1.32	Yes	$ 153,410	Yes	Yes
14	26	Female	South America	Millenials	0.765	Yes	$ 93,194	No	Yes
15	27	Male	Europe	Baby Boomers	1.56	Yes	$ 101,662	Yes	Yes
16	29	Male	Europe	Generation X	1		$ 114,216		
17	30	Male	Europe	Baby Boomers	1.32	No	$ 97,354	No	No
18	31	Male	Europe	Millenials	0.68	Yes	$ 120,061	No	No
19	33	Male	North America	Generation X	1		$ 134,308		

Data Numbers

	A	B	C	D	E	F	G	H	I	J	K
1	RespID	Q0_Gender	Q0_Location	Q0_Generation	Q0_Weig	Q1	Q100	Q2_1	Q2_2	Q2_3	Q2_4
2	2	Male	South America	Generation X	1	0	98,037.68	0	0	1	0
3	4	Female	South America	Baby Boomers	1.44	0	138,935.50	1	1	1	0
4	5	Female	South America	Generation X	1	1	84,471.00	0	1	1	1
5	6	Male	Antarctica	Baby Boomers	1.44	2	138,533.55	0	1	1	0
6	9	Female	Europe	Baby Boomers	1.32	1	68,943.60	1	1	1	1
7	12	Female	Europe	Baby Boomers	1.56	0	100,663.20	0	0	1	1
8	15	Male	North America	Baby Boomers	1.56		122,480.76				
9	16	Male	Antarctica	Baby Boomers	1.44	1	106,035.60	1	1	0	0
10	17	Female	Europe	Baby Boomers	1.32	2	81,681.30	1	1	1	0
11	18	Male	North America	Traditionalists	0.595	0	104,199.70	0	1	0	0
12	22	Male	South America	Generation X	1.32	0	172,723.10	0	0	1	1
13	25	Female	South America	Generation X	1.32	1	153,410.40	1	1	1	1
14	26	Female	South America	Millenials	0.765	1	93,194.00	0	1	0	0
15	27	Male	Europe	Baby Boomers	1.56	1	101,661.78	1	1	1	0
16	29	Male	Europe	Generation X	1		114,215.85				
17	30	Male	Europe	Baby Boomers	1.32	0	97,353.72	0	0	0	0

Data Helper

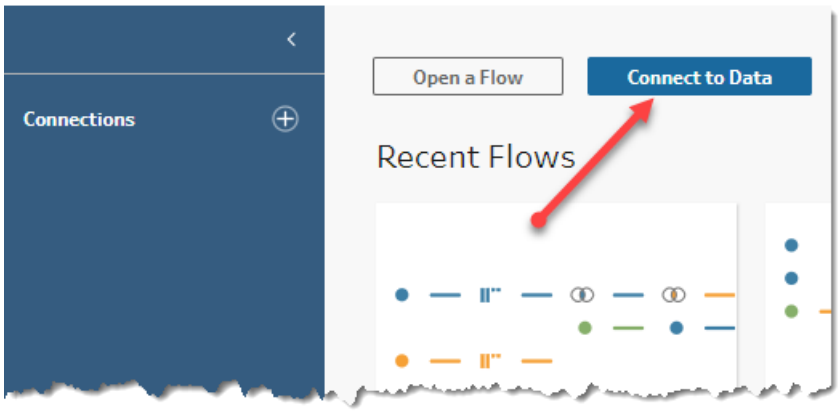
	A	B	C	D
1	QuestionID	Wording	Question Grouping	Qtype
2	Q1	Vote in the upcoming election?	Vote	Single-Punch
3	Q100	What is your salary?	Salary	Benchmark
4	Q2_1	Pulse Rate	What do you measure	Multi-Punch
5	Q2_2	Metabolism	What do you measure	Multi-Punch
6	Q2_3	Blood Pressure	What do you measure	Multi-Punch
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

We need to combine and reshape the data so that we get something that looks like what we saw on page 5.

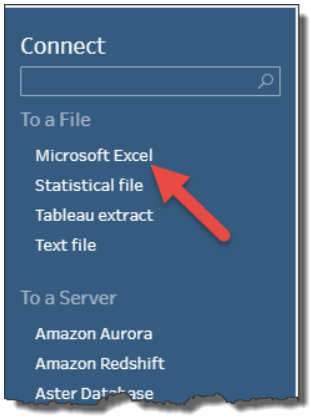
Getting Data "Just So" Using Tableau Prep

To Start Tableau Prep Builder and Connect to the Source File

1. Start Tableau Prep Builder and click **Connect to Data**.

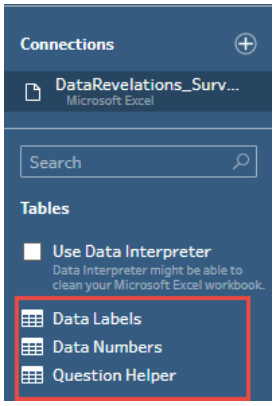


2. Select **Microsoft Excel** from the To a File list.



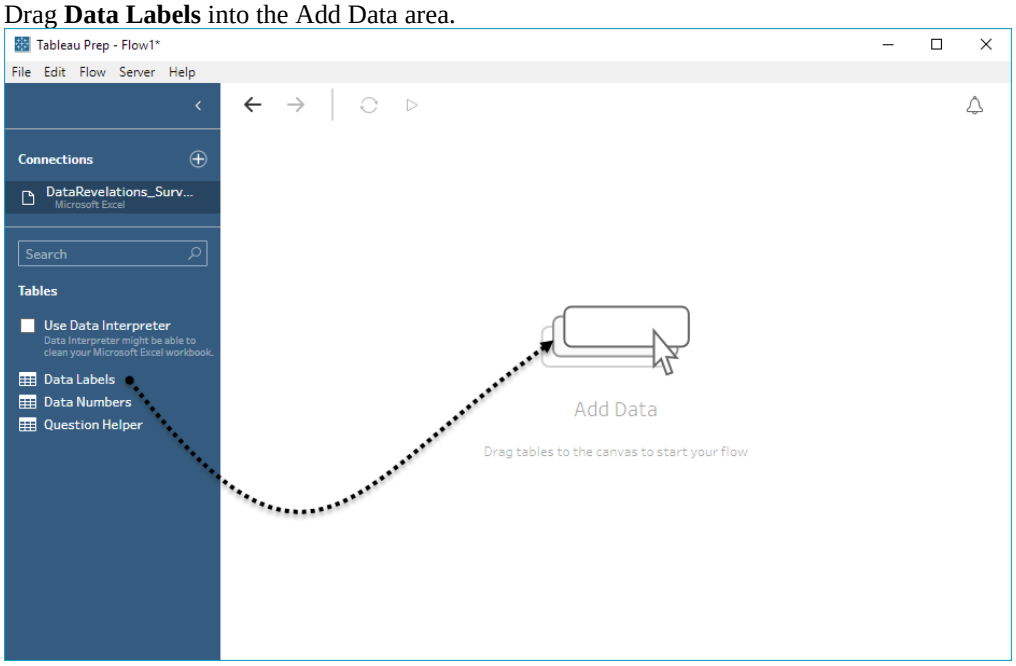
3. Go to the **Source** folder, select the file **DataRevelations_SurveyData_V4.xlsx**, and click **Open**.

   You'll see a list of the three "tables" that comprise the Excel file, as shown below.

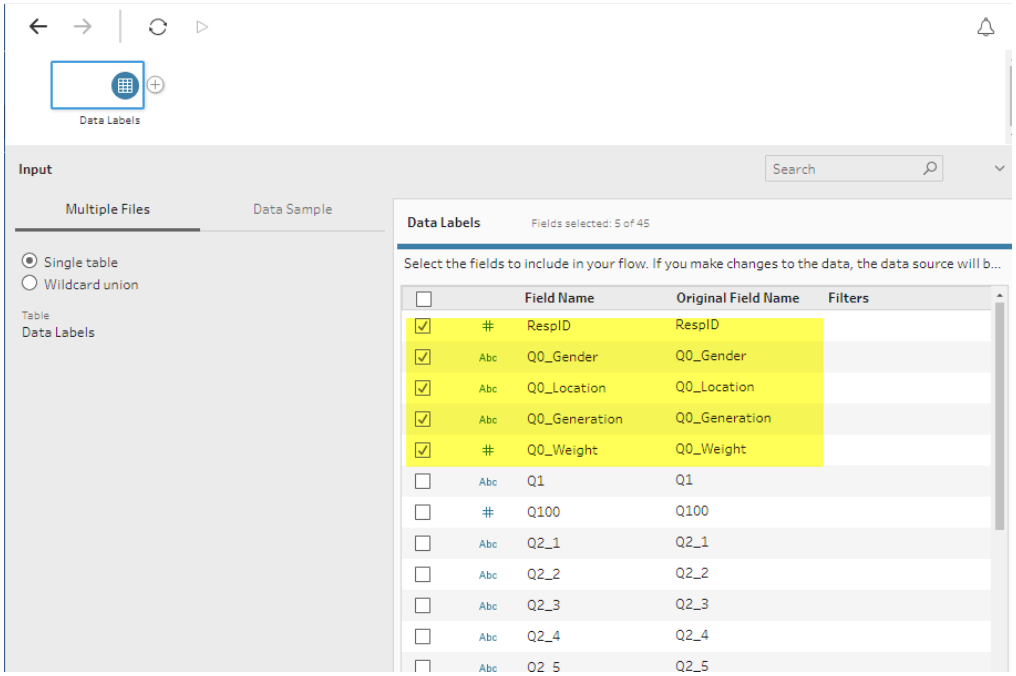


To Specify the Demographic Components

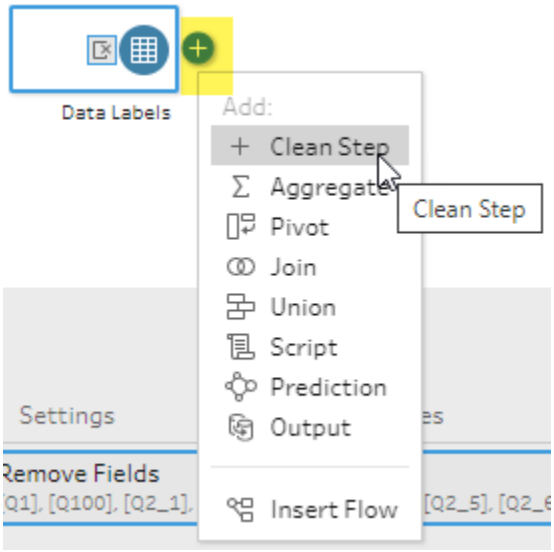
1. Drag **Data Labels** into the Add Data area.



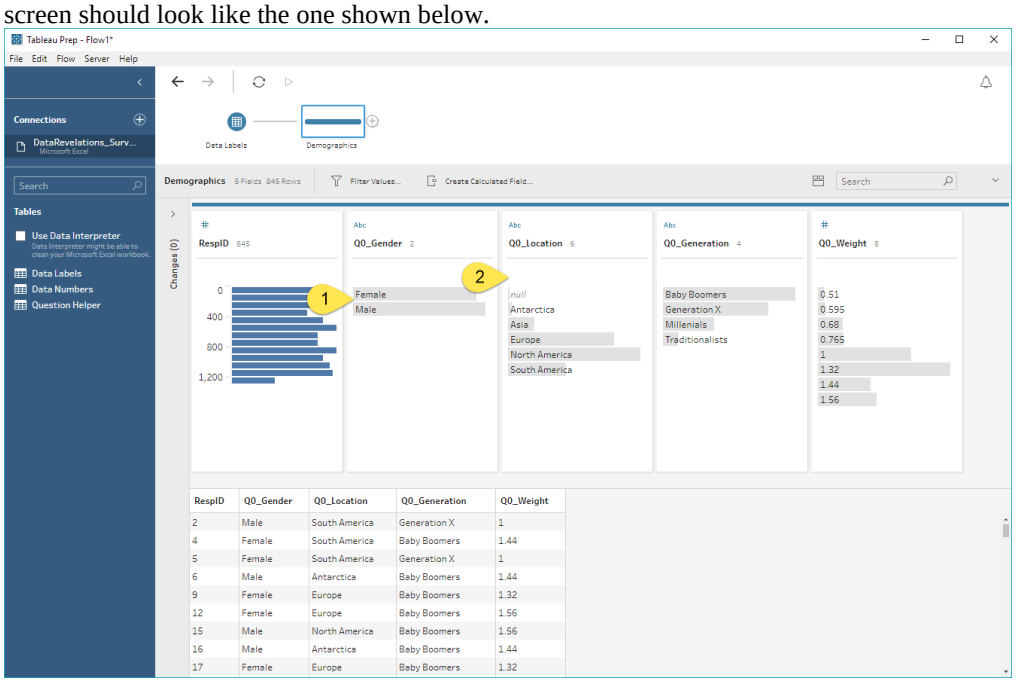
2. Indicate the fields you want to include (Resp ID and the demographic fields) as shown below.



3. Click the "+" sign next to the Data Labels step in the upper left corner and indicate you want to add a **Clean Step**.



4. Right-click the step currently named "Clean 1" and rename it **Demographics**. Your screen should look like the one shown below.



Inspecting the Merchandise

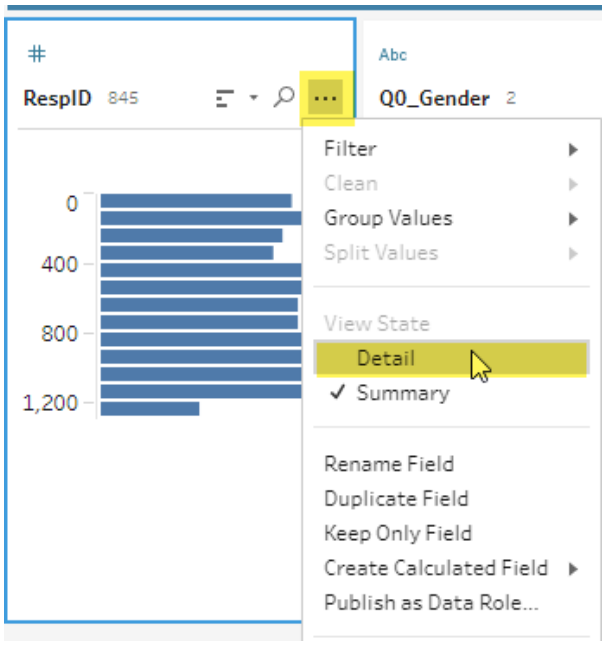
Note that you can now see both a profile view and a data grid view.

Look at the stuff in the middle! You can see a distribution of all the responses to each question, before you even analyze things in Tableau! For example, we can see that more men than

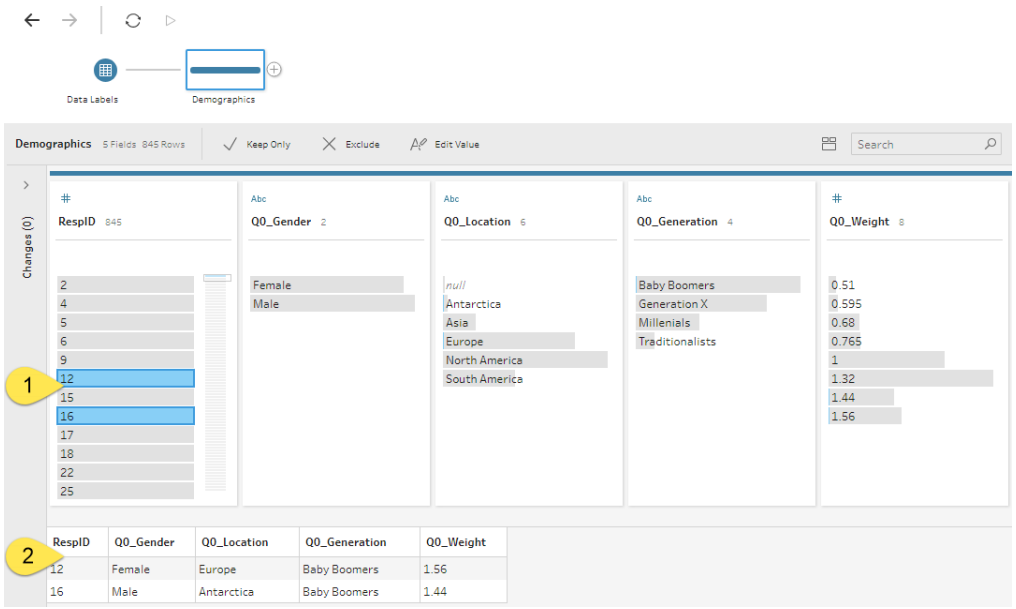
women took this survey (1) and that there were a handful of people that did not specify where they live (2).

Let's look at some individual responses.

1. Click the ellipsis (. . .) in the upper right corner of the RespID field and indicate that you want to change the view state to **Detail**.



2. Select RespIDs 12 and 16. As shown below.

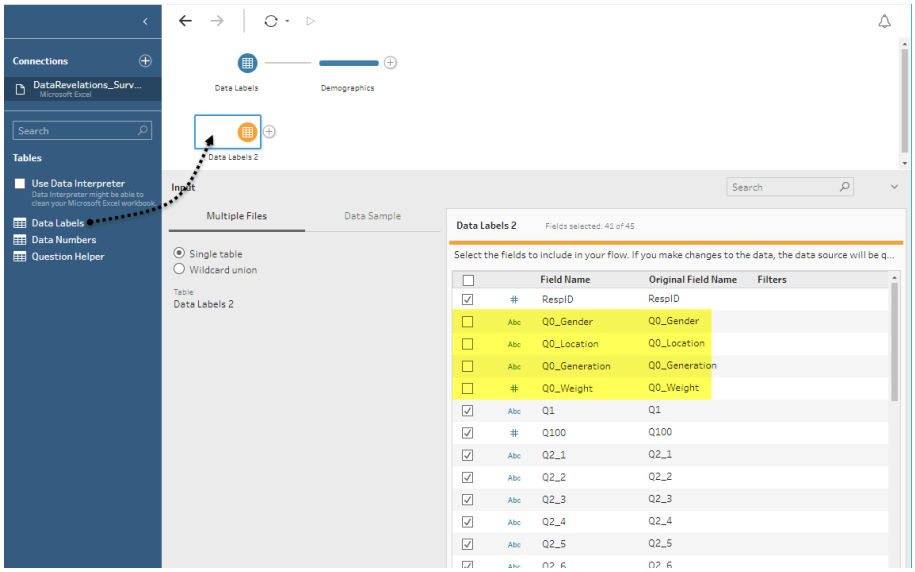


Selecting items in the profile view (1) shows associated records in the data grid (2).

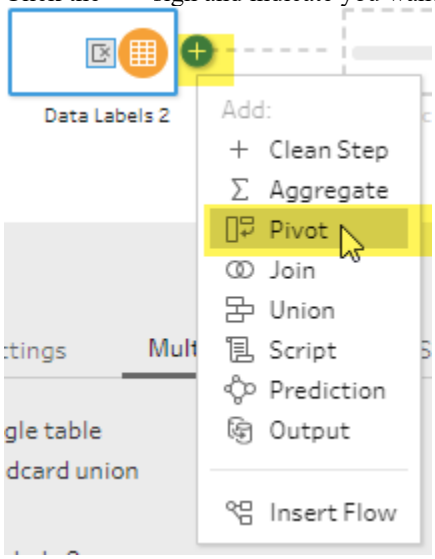
3. **Save your work**, either as a Tableau flow or packaged Tableau flow (you can save it anywhere you want; just remember where you put it.)

To Import and Reshape the Label Responses

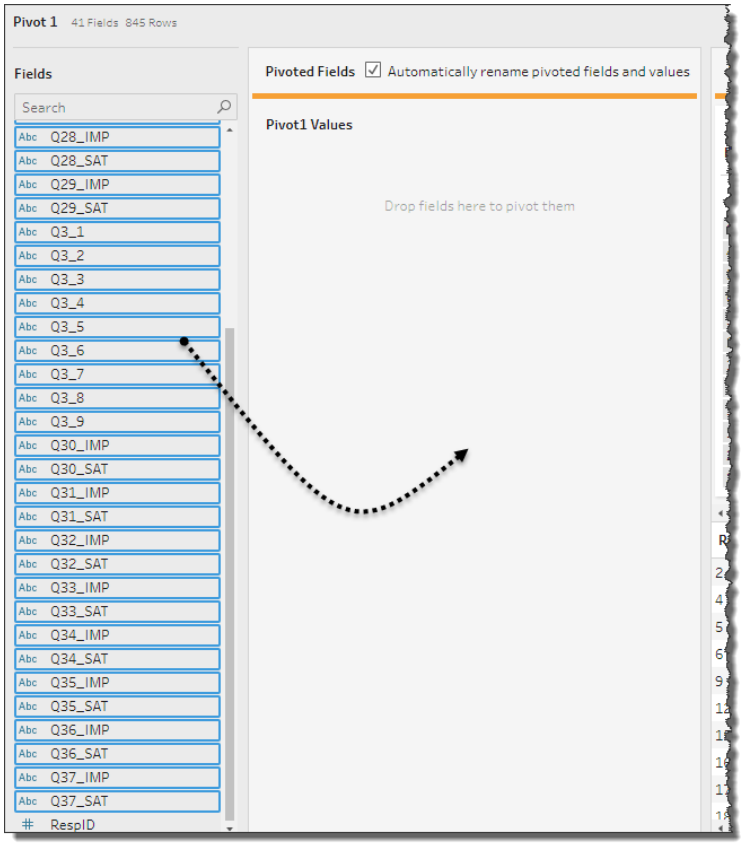
1. Drag **Data Labels** into the data area a second time and indicate that you do **not** want to include the demographics fields, as shown below. Make sure that you DO include RespID.



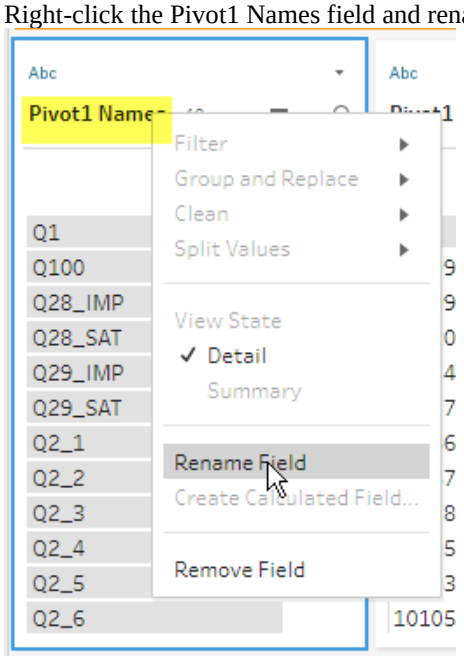
2. Click the "+" sign and indicate you want to add a Pivot.



3. Drag all of the fields except RespID into the Pivot1 Values area.



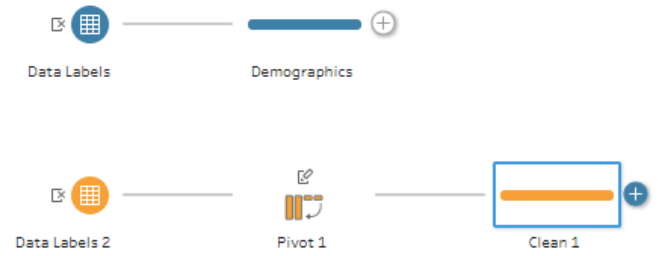
4. Right-click the Pivot1 Names field and rename it **Question ID**.

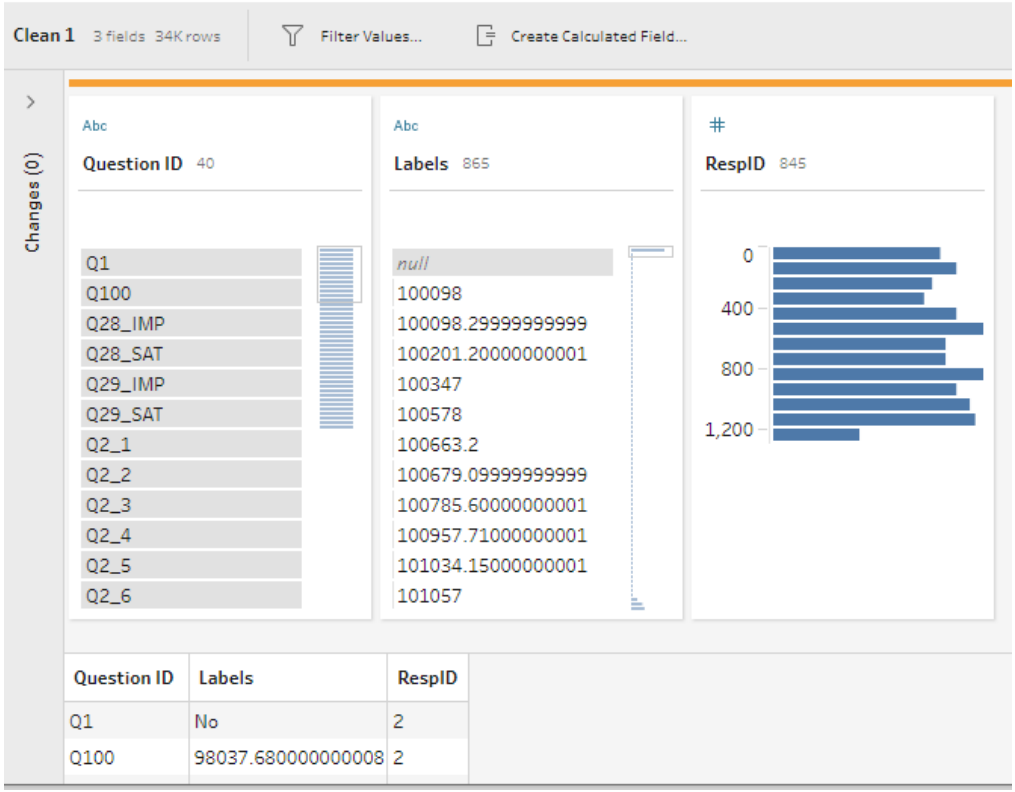


5. Right-click the Pivot1 Values field and rename it Labels.

Labels

6. Add a **Clean Step**. Your screen should look like the one shown below.





7. Save your work.

To Join the Demographic and Reshaped Label Data

1. Drag the Clean 1 step to the right of the Demographics step and indicate you want to do a Join.

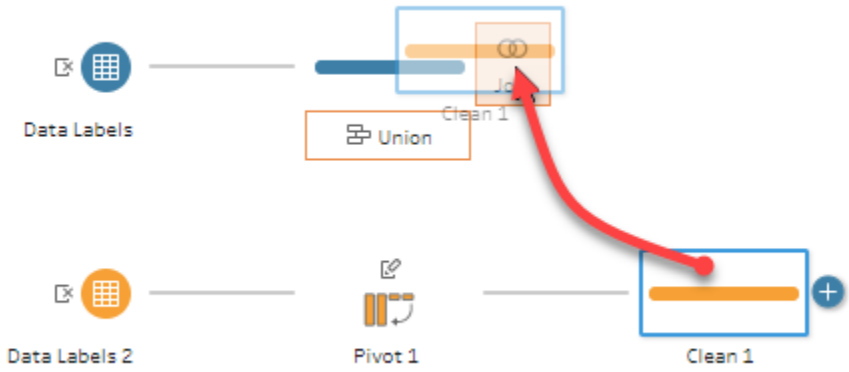
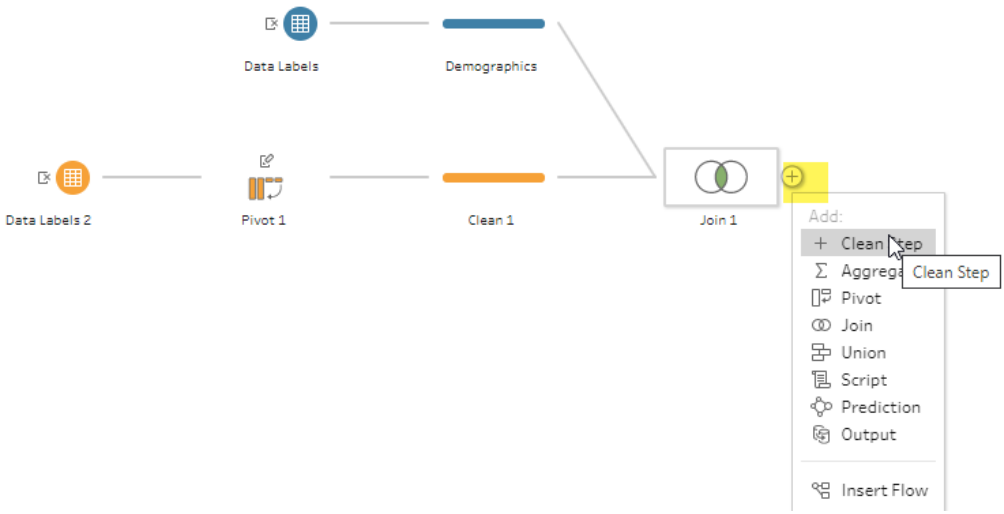


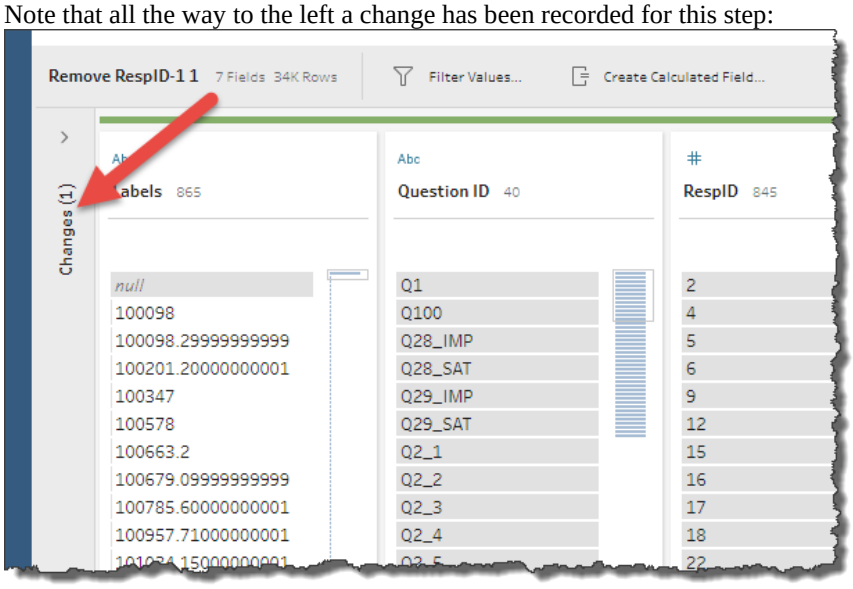
Tableau will guess that you want to perform an inner join on the RespID field (this is indeed what we want to do.)

2. Indicate you want to add a **Clean Step** after the Join 1 step.

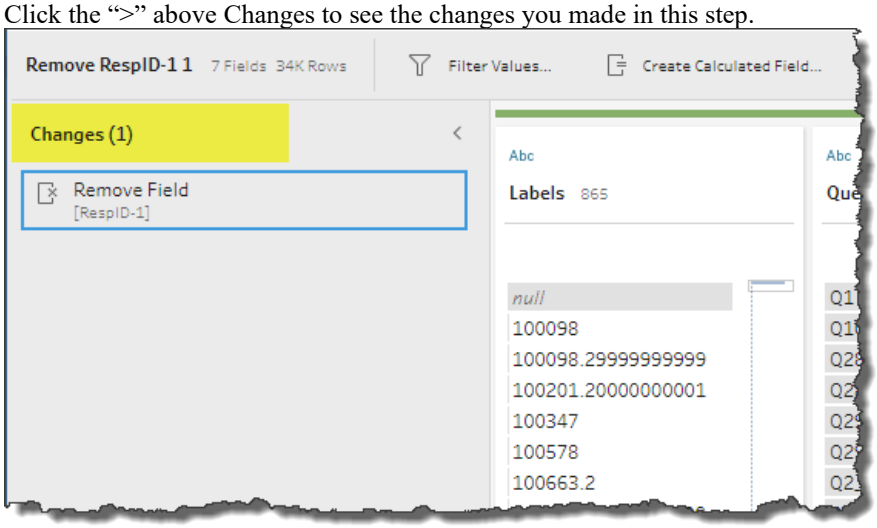


3. Find the **RespID-1** field (it's probably all the way to the right), right-click it, and indicate you want to Remove it.

Note that all the way to the left a change has been recorded for this step:



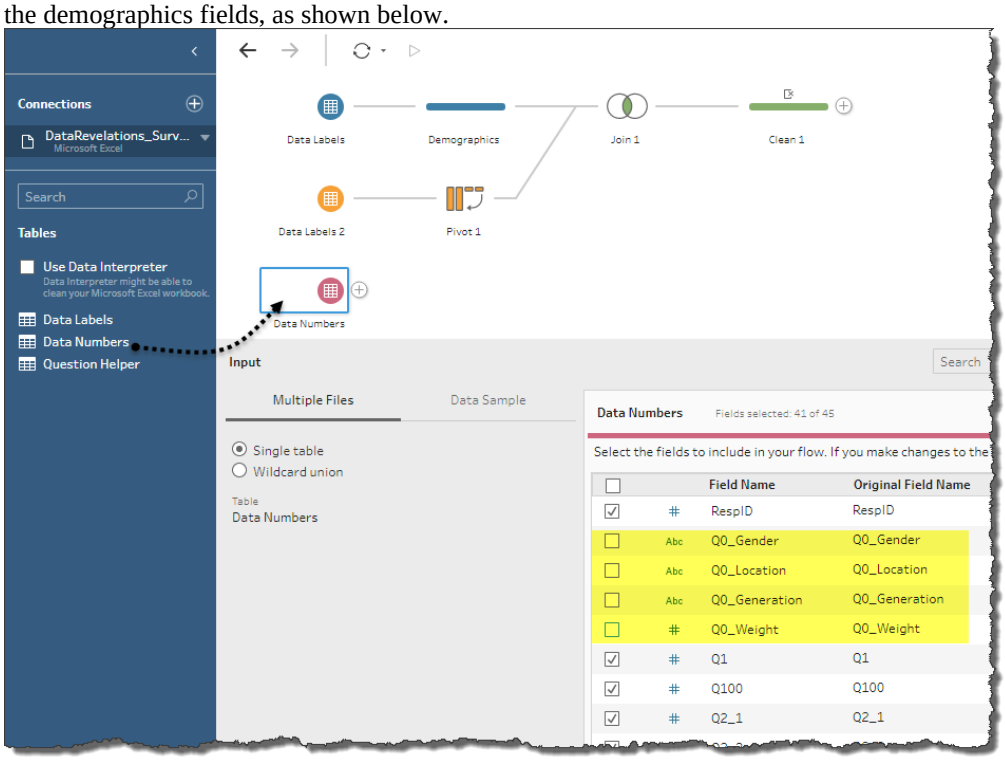
4. Click the ">" above Changes to see the changes you made in this step.



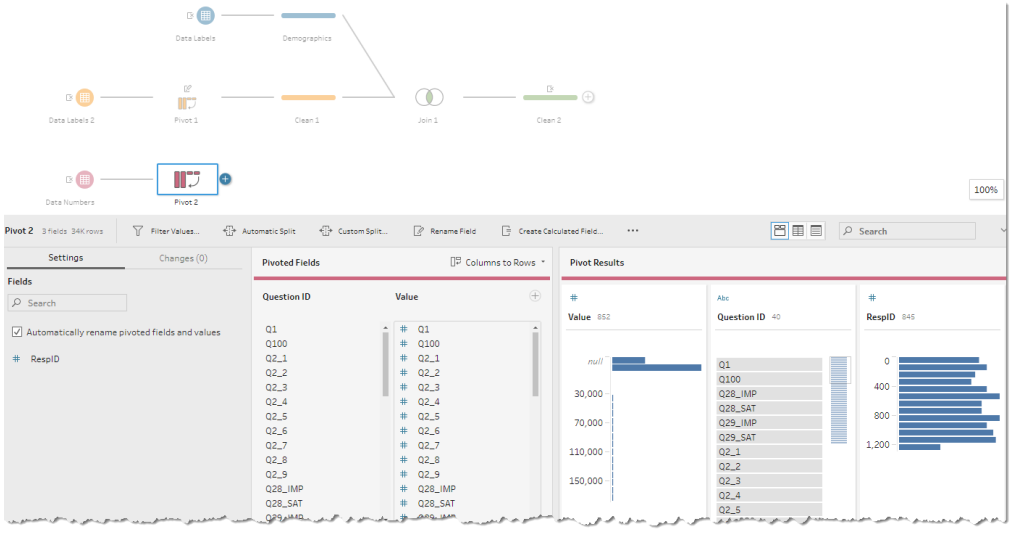
5. Save your work.

To Import and Reshape the Numeric Survey Responses

1. Drag **Data Numbers** into the data area and indicate that you do not want to include the demographics fields, as shown below.



2. Click the "+" and indicate you want to add a **Pivot**.
3. Drag all the fields except RespID into the Pivot1 Values area.
4. Rename Pivot2 Names to **Question ID**.
5. Rename Pivot2 Values to **Value**. Your screen should look like the one shown below.

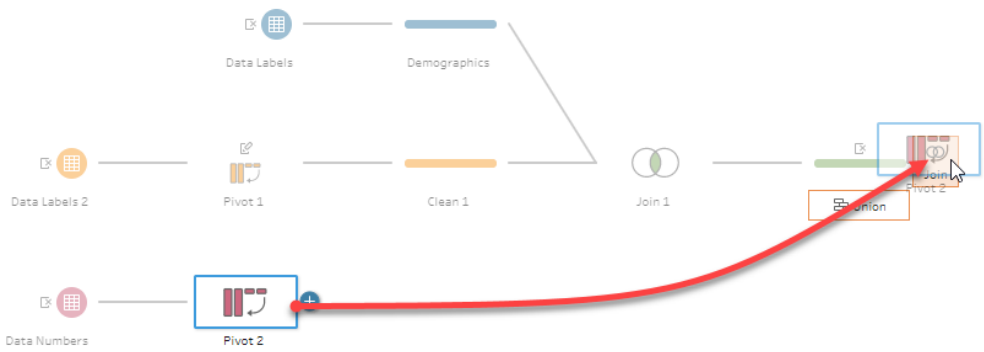


To Join the Demographic, Label, and Numeric Data Together

Now that we have both the text and numeric results pivoted / reshaped we need to merge the data so that all the numeric responses line up with all the text responses. That is, for every Resp ID we want to make sure the text and numeric responses for each Question ID line up properly.

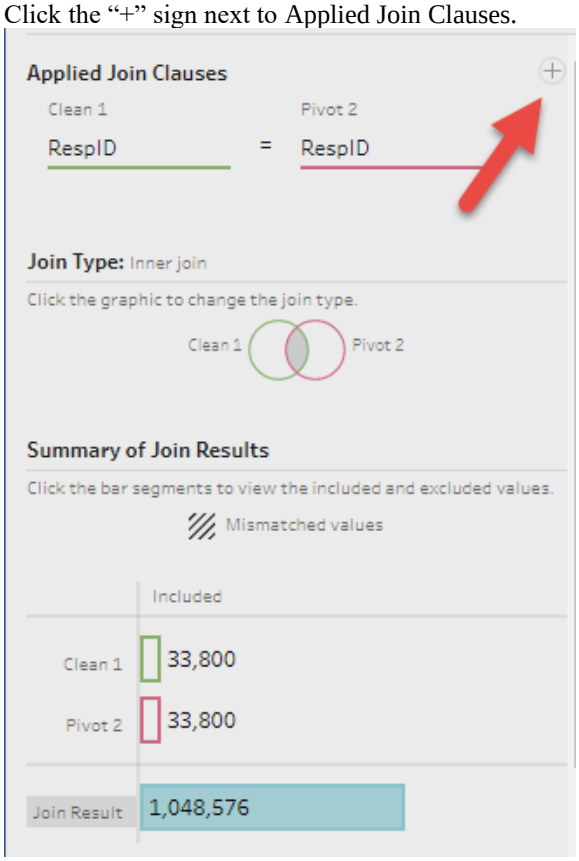
We'll do this by joining the two reshaped data sources.

1. Drag the Pivot step in the Data Numbers flow to the right of the last step in the first Data Labels flow.

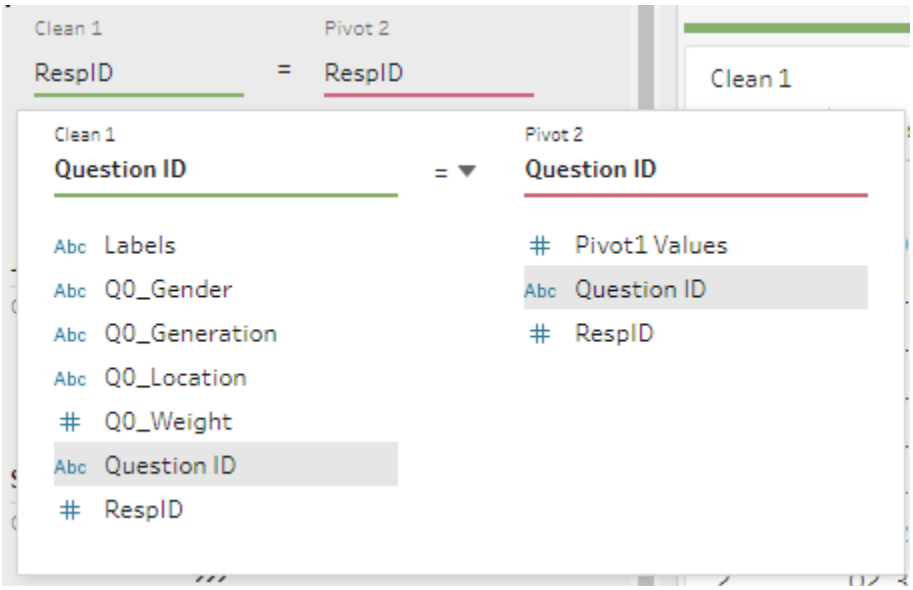


Note that we need t join on both RespID and on Question ID.

2. Click the "+" sign next to Applied Join Clauses.



3. Indicate that you also want to join **Question ID** form Clean 2 with Question ID from Pivot 2, as shown here.

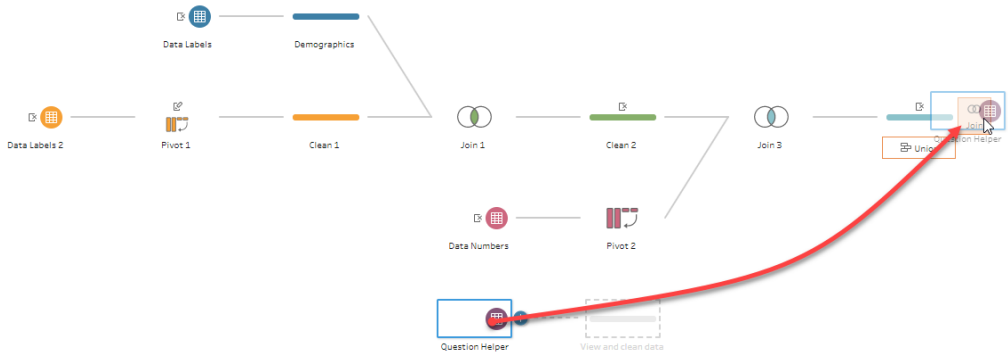


4. Add a new step and remove the unnecessary fields Question ID-1 and RespID-1.

Note: You can rearrange the order of the fields by dragging them left and right. This would just be to help you understand the data; Tableau and Tableau Prep don't care about the field order.

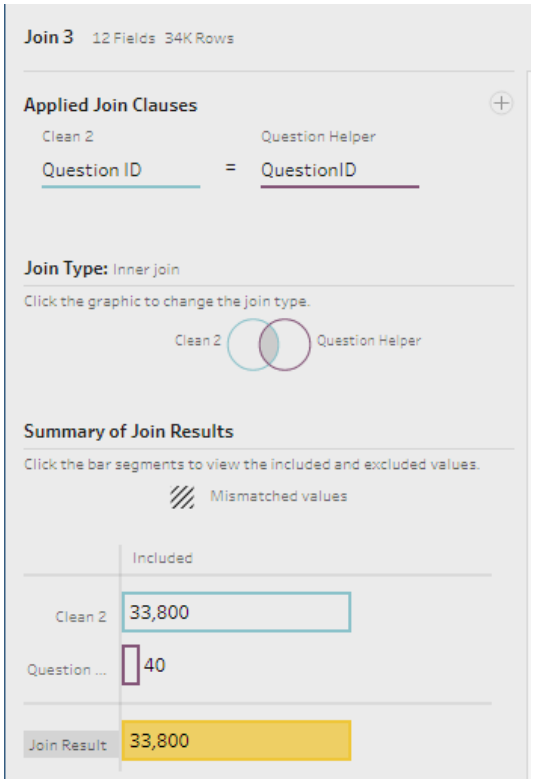
To Import and Connect to the Meta Info (the Helper File)

1. Drag Question Helper into the Add Data area.
2. Drag the Question Helper step to the right of the Clean 3 step (part of the first Data Labels flow).



Even though the field names are not identical, Tableau Prep will guess that you want

to join Question ID from Clean 3 with QuestionID (no space) from Question Helper.



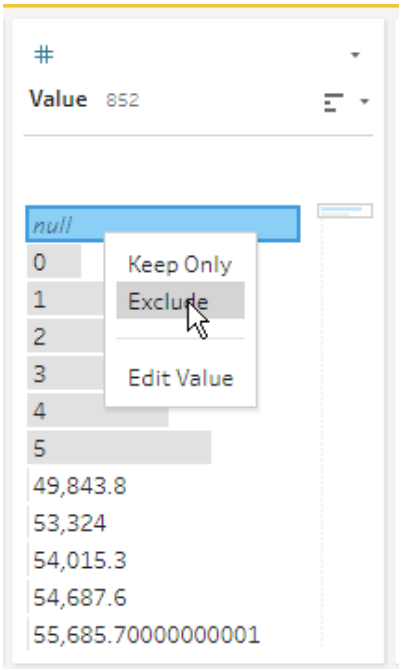
3. Add a new step so you can easily inspect your results.
4. Save your work.

Removing Null Values

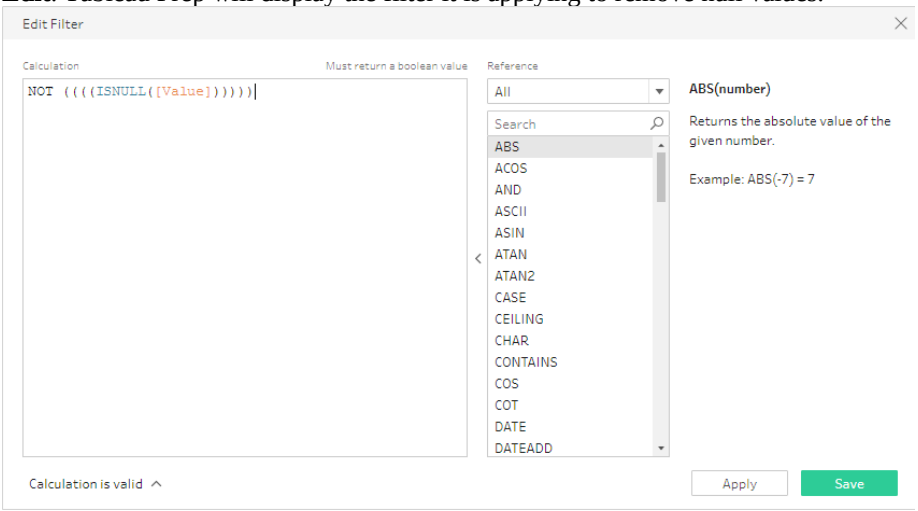
If you know your data well and know that it has been coded correctly, you can probably trim down your data set by removing Null values (i.e., rows where a respondent completely skipped answering a question.)

1. With the latest step in the flow selected, find the Value field.

2. Right-click null and select **Exclude**.



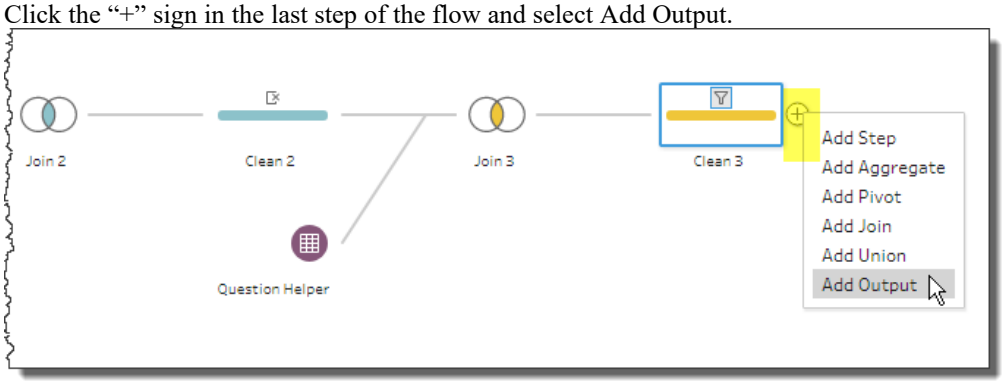
3. In the Changes portion of the screen, right-click the newly-created filter and select **Edit**. Tableau Prep will display the filter it is applying to remove null values.



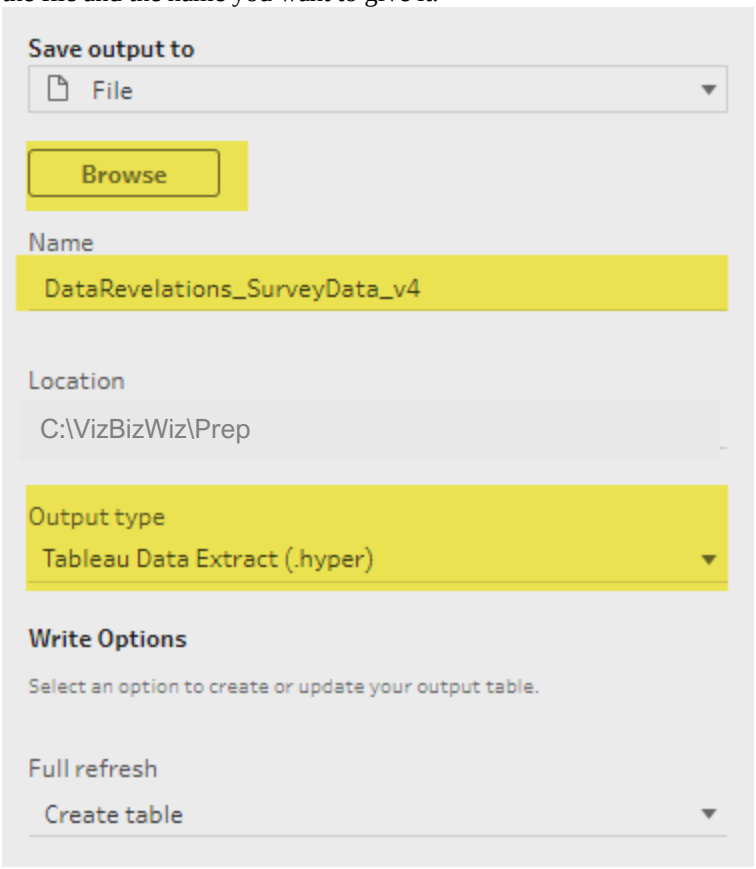
4. Click **Save**.

To Export the Results

1. Click the "+" sign in the last step of the flow and select Add Output.



2. Indicate the Output type (Hyper, Excel. Or CSV) as well as where you want to save the file and the name you want to give it.



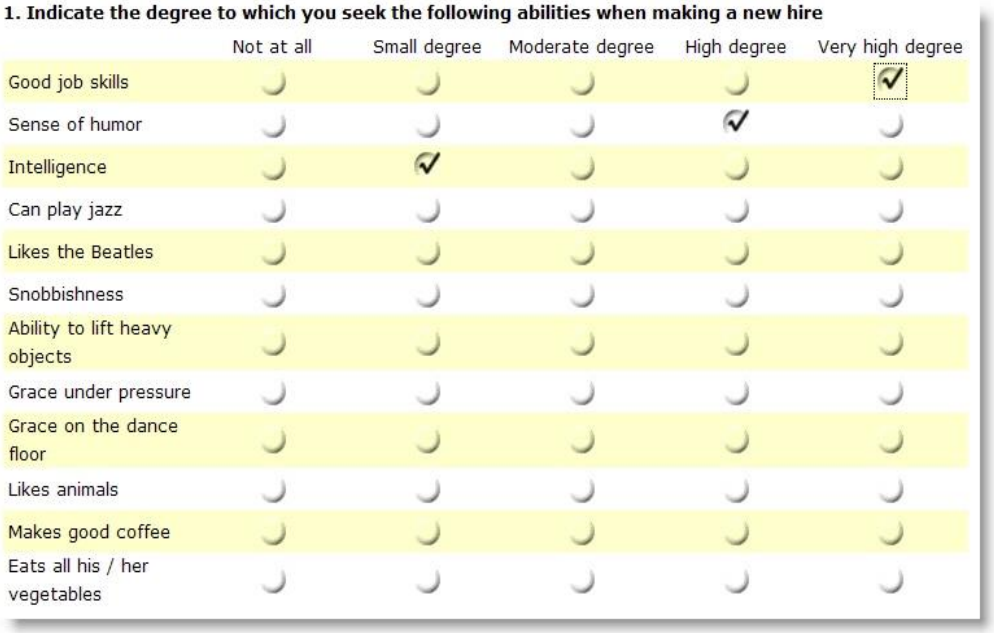
3. Click the Run Flow button.

4. When the flow completes, click **Done**.
5. Save your work.

IV. Visualizing Likert Scale Question

Overview

If your surveys are like the vast majority of those I've seen there are probably several sets of Likert scale questions that look like this:



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	○	○	○	○	○

Let's look at several different ways to visualize this type of question group, starting with a really ugly, hard-to-grok 100% stacked bar chart.

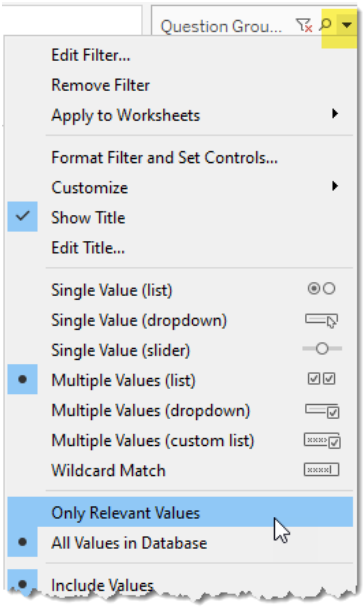
Why start with something that's ugly and hard to understand? Building this first will help us build visualizations that are beautiful and easy to understand.

To Create a Likert Scale 100% Stacked Bar Chart

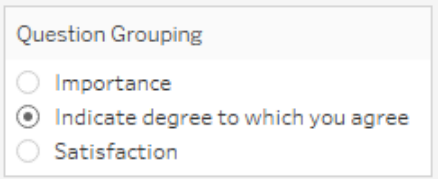
Note: If you had difficulty completing the previous exercise, open the file **2g_Check-All_Dashboard_Complete.twbx** from the **Starter** folder and work with that.

1. Create a new worksheet.
2. Drag **QType** to the Filters shelf and select **Likert**.
3. Drag **Question Grouping** to the Filters shelf and select **Indicate degree to which you agree**.
4. Right-click Question Grouping in the Filters shelf and select **Show Filter**.

5. Click the carat symbol on the upper right of the Question Grouping filter and select **Only Relevant Values**.



6. Modify the Filter setting so that it shows a **Single Value (list)** and that **(All)** is not an option.

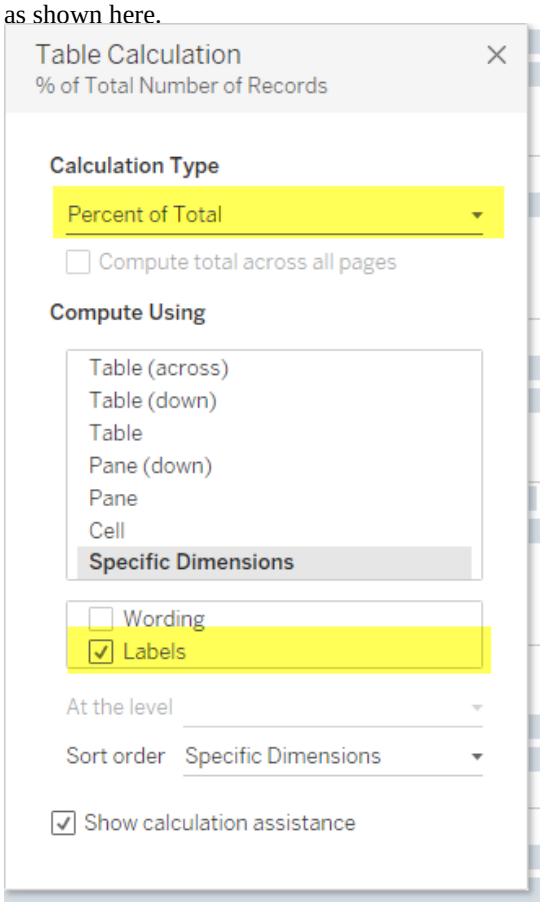


7. Drag **Wording** to the Rows shelf.
8. Drag **Labels** to the Rows shelf.
9. Manually reorder the items so that they appear in this order:

Wording	Label
Can Play Jazz	Very high degree
	High degree
	Moderate degree
	Small degree
	Not at all

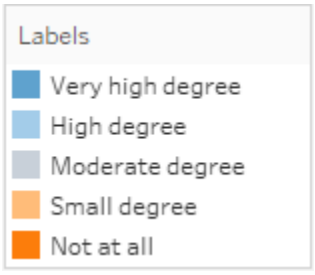
10. Drag **Number of Records** onto the Columns shelf.
11. Right-click **SUM(Number of Records)** on the Columns shelf and select **Add Table Calculation**.

12. Indicate you want to calculate the Percent of Total summarizing values from **Label** as shown here.

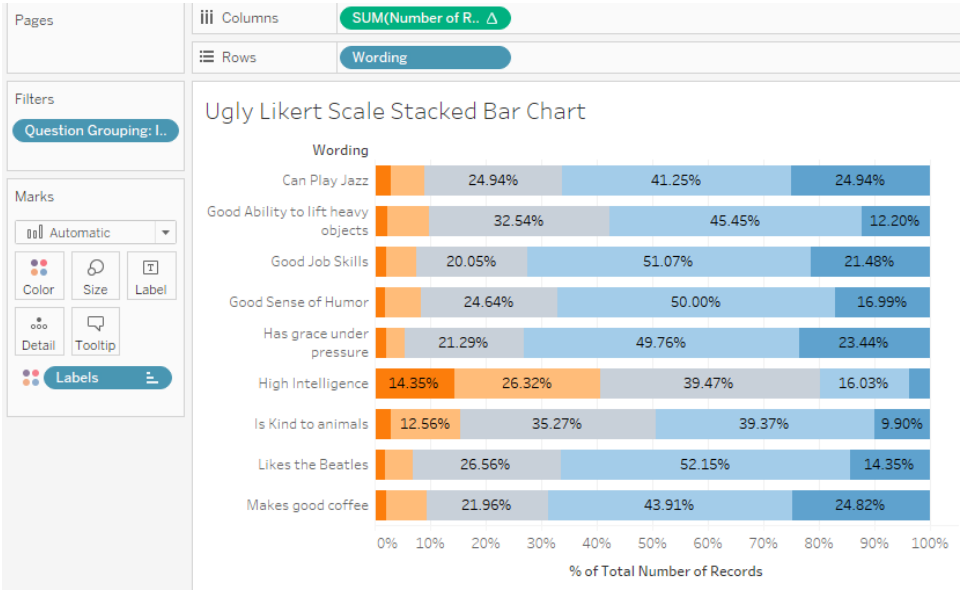


13. Close the Table Calculation dialog box.
14. Turn mark labels on by clicking the **"T" (Text)** icon.

15. Drag **Labels** from the Rows shelf to the **Color** button.
16. Edit the colors using the color blind palette, as shown below.



17. Rename the worksheet tab **Ugly Likert Scale Stacked Bar Chart**.
    Your screen should look like the one shown below.



18. Save your work.

Topics for Discussion

- Do you consider this chart ugly and/or hard to interpret?
- What would make it easier to understand?
- See what happens when you change the Question Grouping filter. Why do the colors change?

Percent Top Two Boxes

I think the popularity of this next view stems from how hard it is to interpret the 100% stacked bar chart. This next chart will just focus on the percent of people that selected either "Very High Degree" or "High Degree".

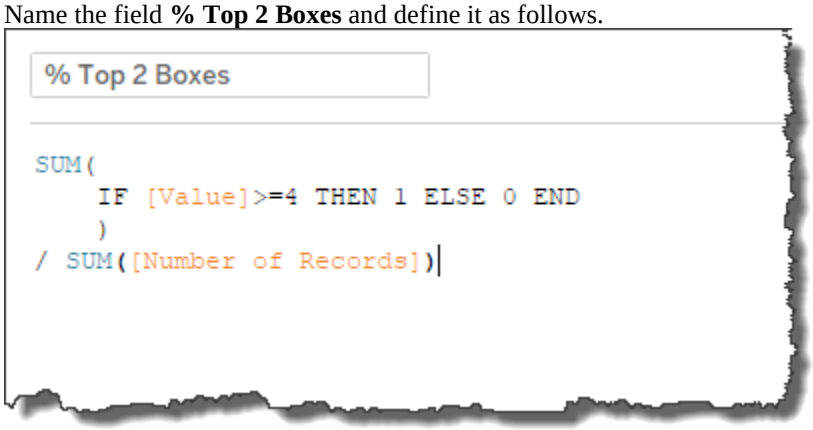
Note – if you examine the data the values for the Likert scale questions go from 1 to 5 where "Very High Degree" is 5 and "High Degree" is 4.

To Create a Percent Top Two Boxes Chart

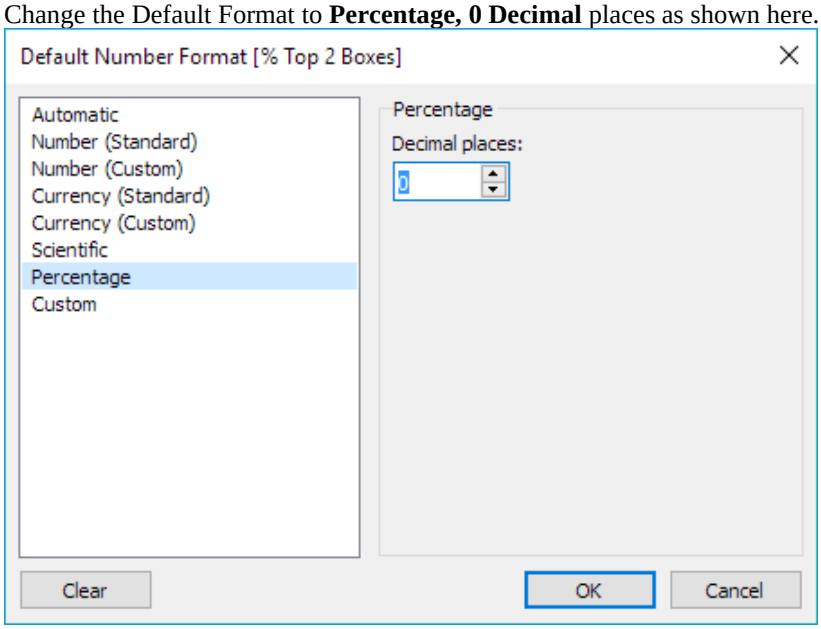
! **Note:** If you had difficulty completing the previous exercise, open the file **4a_Likert.twbx** from the **Starter** folder and work with that.

1. Right-click the **Ugly Likert Scale Stacked Bar** tab and select **Duplicate**.

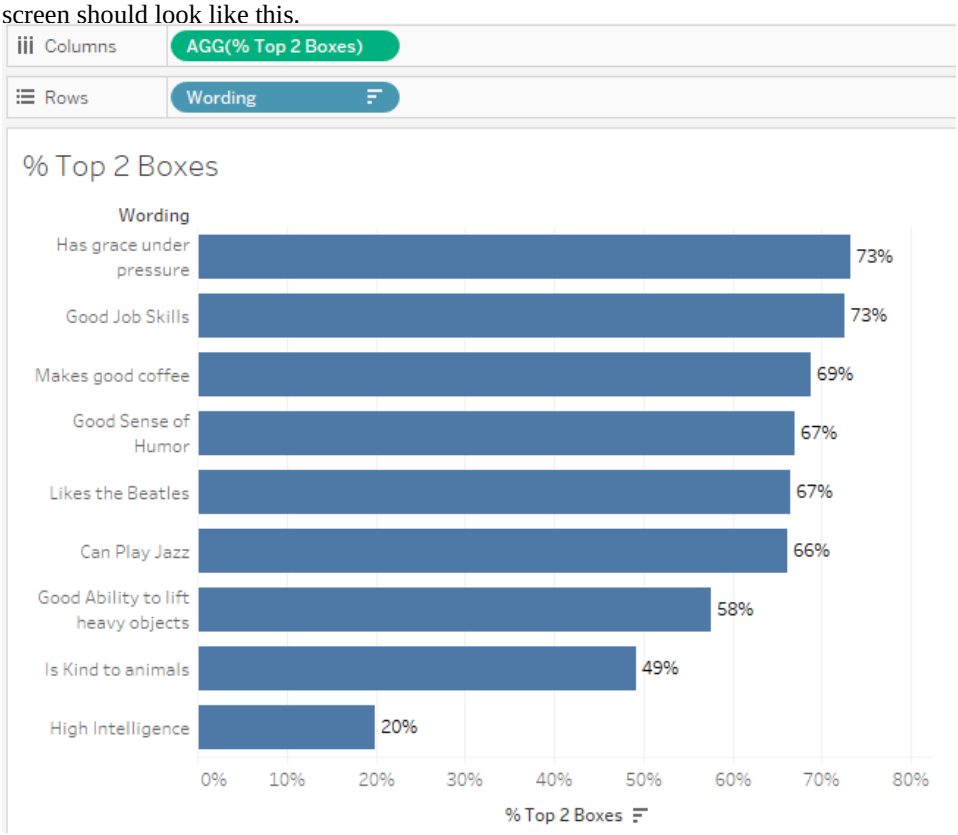
2. Remove **SUM(Number of Records)** from the Columns shelf and **Labels** from the Color shelf.
3. Right-click in the Measures area and select **Create Calculated Field**.
4. Name the field **% Top 2 Boxes** and define it as follows.



5. Click **OK**.
6. Right-click the newly created field in the Measures area and select **Default Properties | Number Format**.
7. Change the Default Format to **Percentage, 0 Decimal** places as shown here.



8. Drag **% Top 2 Boxes** to the Columns shelf and sort in descending order. Your screen should look like this.



9. Rename the sheet **% Top 2 Boxes** and save your work.

Note: If you want to take weights into account modify the % Top 2 Boxes formula so that it looks like this:

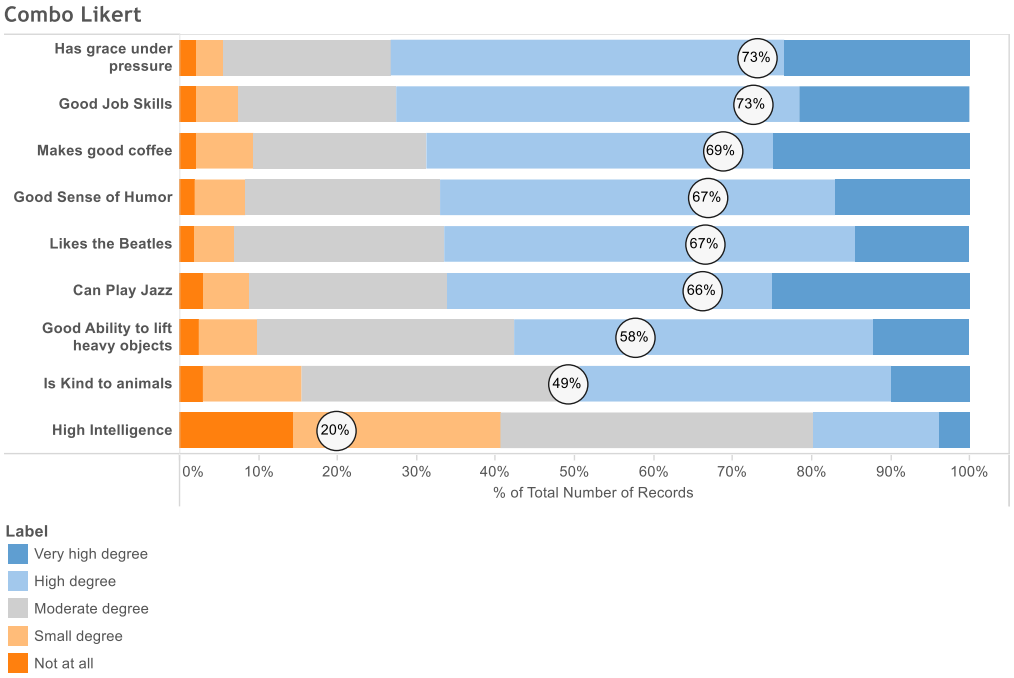
```
SUM(
    IF [Value]>=4 then [Weight] ELSE 0 END
    )
/ SUM([Weight])
```

Topics for Discussion

- Why is this chart easier to "grok" than the 100% stacked bar?
- Suppose you wanted to show Percent Top Box or Percent Top Three Boxes; is there an easy way to do this?

Combination Stacked Bar and Circle Chart

While not "Likert Nirvana" the combination stacked bar / circle chart does make it easy to see and rank sentiment.



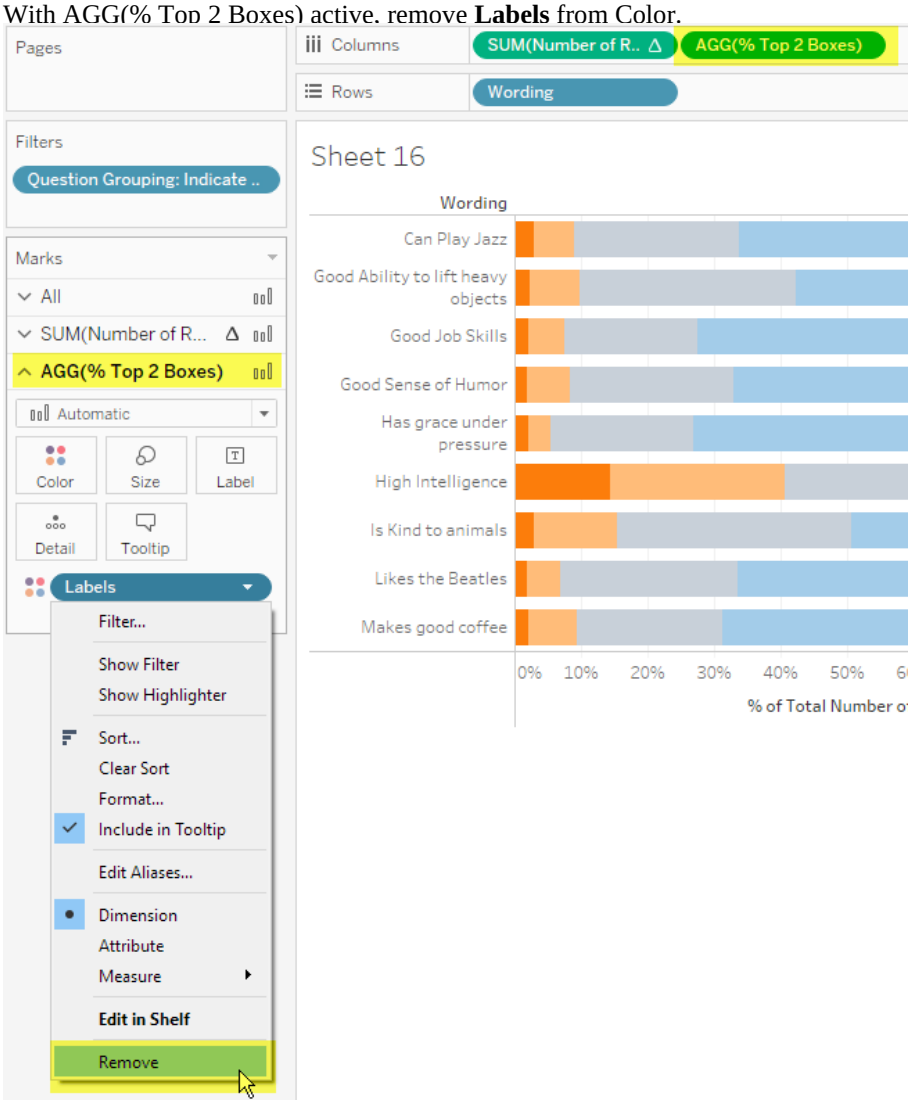
Let's see how we can go about building this.

To Create a Combination Stacked Bar Circle Chart for Likert Scale Data

Note: If you had difficulty completing the previous exercise, open the file **4b_Likert.twbx** from the **Starter** folder and work with that.

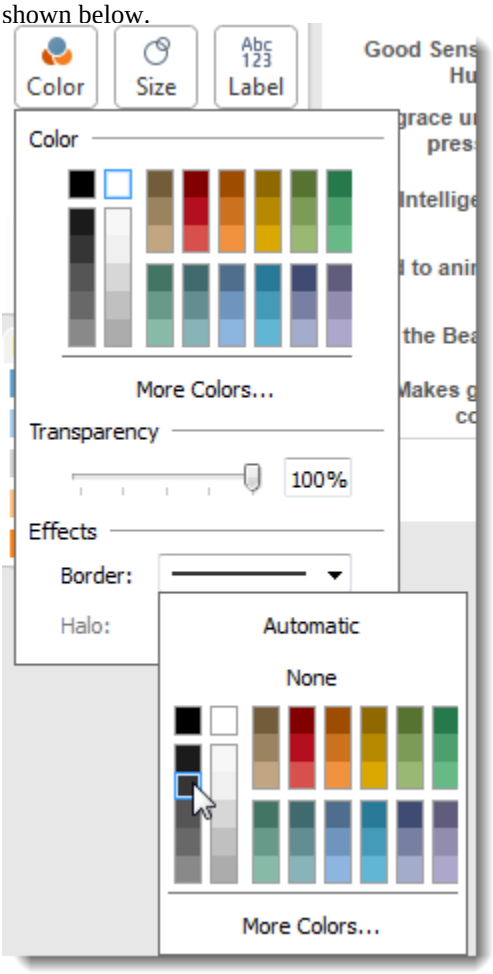
1. Right-click the **Ugly Likert Scale Stacked Bar** tab and select **Duplicate**.
2. Turn off Mark Labels (the "T" icon).
3. Drag **% Top 2 Boxes** to the right of SUM(Number of Records) on the Columns shelf.
4. Select the second green pill (the one you just placed on the Columns shelf) to make that chart active.

5. With AGG(% Top 2 Boxes) active, remove **Labels** from Color.

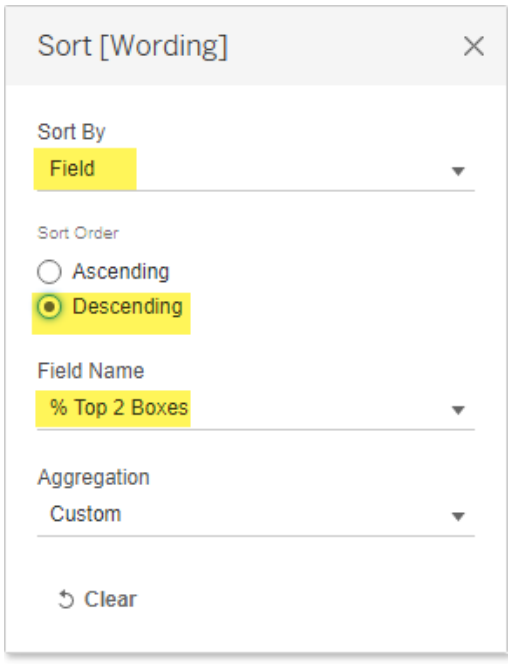


6. Right-click the second green pill on Columns and change the Mark Type to **Circle**.
7. Right-click the second green pill on Columns and select **Dual Axis**.
8. Right click the first green pill on Columns and change the Mark Type to **Bar**.
9. Click the second green pill to make the circle chart active.

10. Click the **Color** button and change the color to white with a dark gray border, as shown below.



11. Click the **Size** button and move the slider to the right a bit.
12. While holding down the **CTRL** key drag the second green pill to the **Label** button (Command-Drag on a Mac).
13. Click the **Label** button and change the **Alignment** to **Centered**.
14. Right-click the secondary axis (the axis along the top) and select **Synchronize Axis**.
15. Right-click **Wording** on Rows and select **Sort**.
16. Indicate that you want to sort by the Field % Top 2 Boxes in Descending order, as shown below.



17. Close the Sort dialog box.
18. Right-click the secondary axis and deselect **Show Header**.
19. Rename the tab **Combo Likert** and save your work.

Divergent Bar Charts

While there are many ways you can show Likert scale data, my "go to" is usually some form of divergent bar chart where the positive sentiment displays to the right and negative sentiment displays to the left.

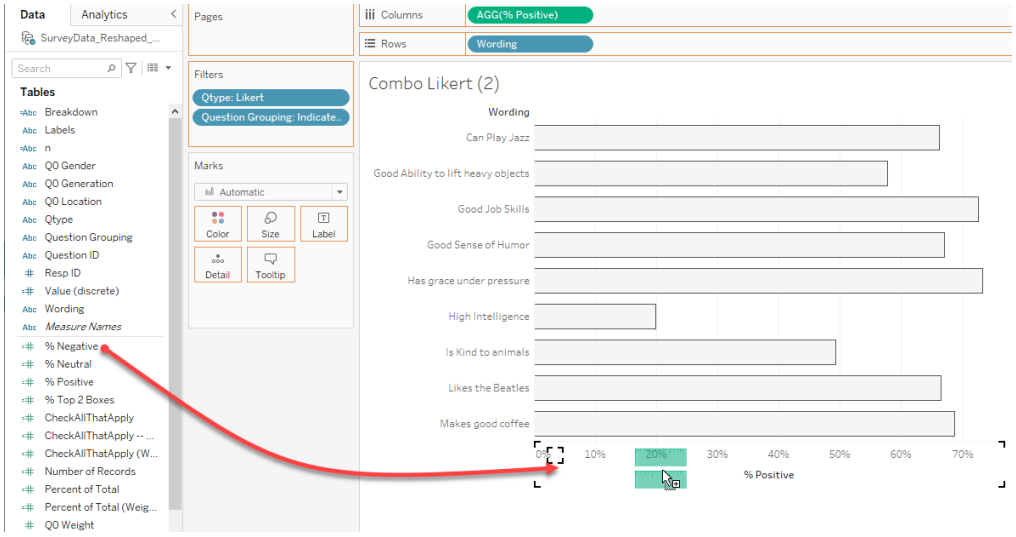
There are A LOT of different ways we can do this (and I'll make sure to show you alternatives) but for this next exercise we'll create a simple three-way chart (positives, neutrals, and negatives) combined with a five-element chart in a tool tip.

To Create the Positives, Negatives, and Neutrals

Note: If you had difficulty completing the previous exercise, open the file **4c_Likert.twbx** from the **Starter** folder and work with that.

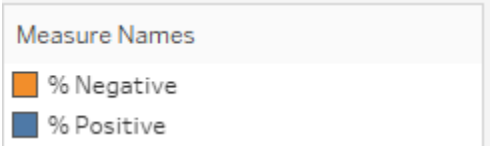
1. Duplicate the Combo Likert worksheet you created in the previous example.
2. Remove everything from Columns, Rows, and Labels (just leave what is on the Filters shelf).

3. Create a calculated field called **% Positive** and define it as follows:
   SUM(
   IF [Value]>=4 then 1
   ELSE 0
   END) /
   SUM([Number of Records])
4. Create a calculated field called **% Negative** and define it as follows:
   SUM(
   IF [Value]<=2 then **-1**
   ELSE 0
   END) /
   SUM([Number of Records])
5. Create a calculated field called **% Neutral** and define it as follows:
   SUM(
   IF [Value]=3 then 1
   ELSE 0
   END) /
   SUM([Number of Records])
6. Change the Mark type to **Automatic.**
7. Drag **Wording** onto Rows.
8. Drag **% Positive** into Columns
9. Drag **% Negative** to the value axis. This will invoke a "Measure Names / Measure Values" construct.



10. Drag **Measure Names** from Rows and place it on **Color**.

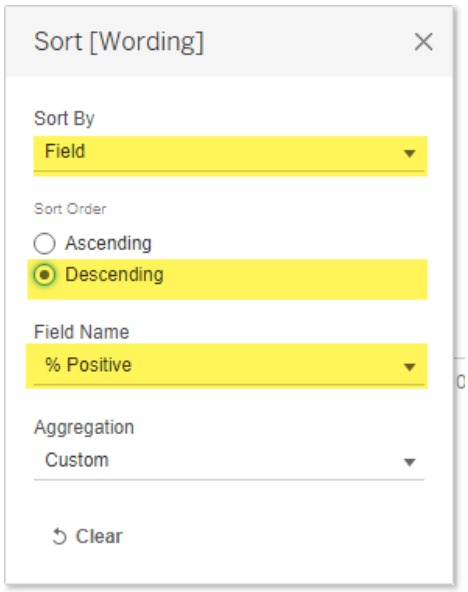
11. Edit the colors so that % Positive is blue and % Negative is Orange.



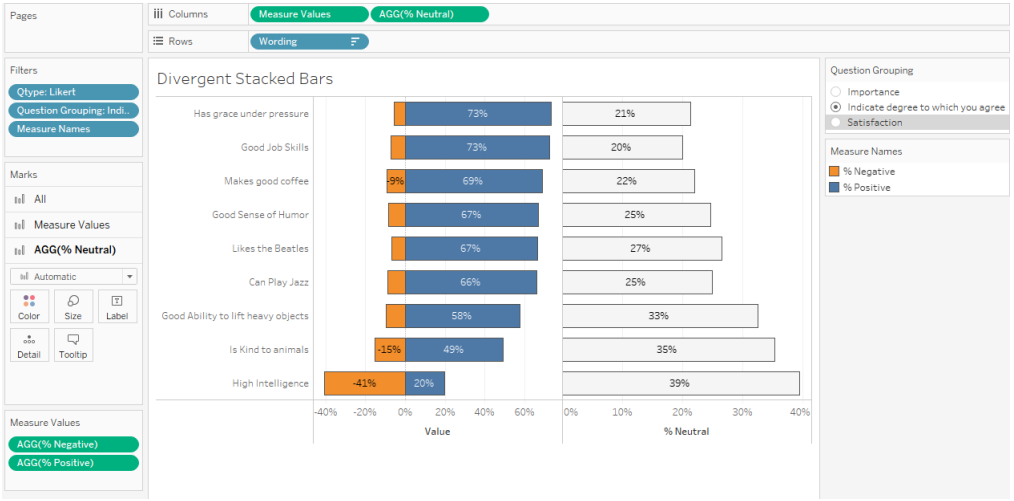
12. Drag **% Neutral** onto Columns.
13. Select the % Neutral pill and remove Measure Names from Color. Your screen should look like this.

Note: if the % Neutral bars aren't gray, click the Color button (with the appropriate measures selected) and change the color.

14. Right click the **Wording** pill on Rows and select **Sort**.
15. Change the sort so that you are sorting by the field **% Positive** in **Descending** order.



16. Close the Sort dialog box.
17. Change the default number format for **% Negative, % Neutral**, and **% Positive** so that they are all displays with percentage with zero decimal places.
18. Click the **Show Mark Labels** (T) in the toolbar to show the numbers in the bars.
19. Rename the sheet **Divergent Stacked Bars** and save your work. Your screen should look like this.

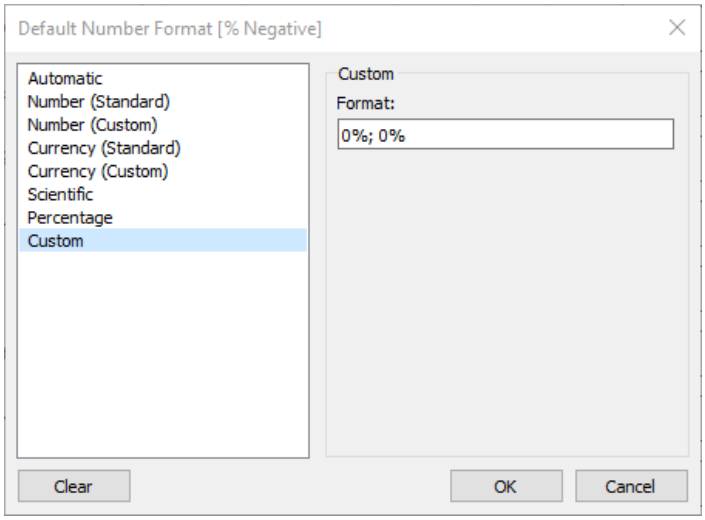


Topics for Discussion

- Look at the length of the bars for the Neutrals and compare with the Positives and Negatives.
- The negative numbers in the orange bars: will your audience / stakeholders have an issue with this?

To Make the Negative Numbers Look They are Positive

1. In the Data Pane, right click **% Negative** and select **Default Properties | Number Format.**
2. Change the format to **Custom** and define it like this.



This translates as "If the number is positive, present it like this; if the number is negative, present it the same way."

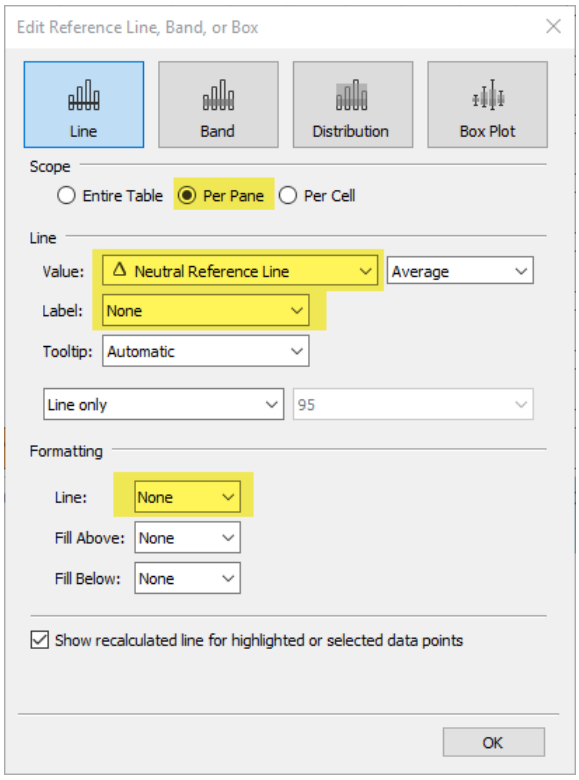
3. Save your work.

To Create an Auto-Adjusting Reference Line

1. Create a new Calculated Field called **Neutral Reference Line** and define is as follows:

```
WINDOW_MAX([% Positive]) + ABS(WINDOW_MIN([% Negative]))
```

2. Select The **% Neutral** pill on Columns to make sure it is active.
3. Drag the Neutral Reference Line onto **Detail**.
4. Right click the axis for % Neutral and select **Add Reference Line**.
5. With **Per Pane** selected, change the Value to **Neutral Reference Line** and change Label to **None** and Line to **None**, as shown here.



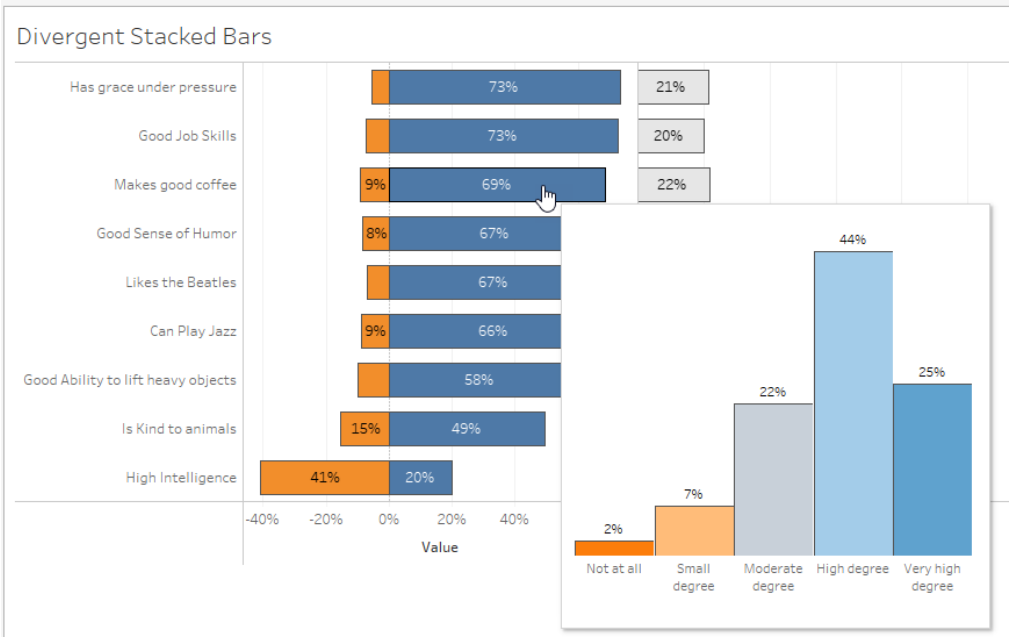
6. Click **OK**.
7. Try changing the Question Grouping selection (i.e., select "Importance" or "Satisfaction).
8. Save your work.

Hey! Remember the demographics dashboard we created earlier and how the bar lengths are not consistent? Placing a hidden reference line is one way we can address that problem.

Showing the Distribution in a Tool Tip

My friend, colleague, and fellow author of **The Big Book of Dashboards**, Andy Cotgreave, states that every chart is a compromise. The divergent stacked bar chart we just created is clear and compact, but it only shows three levels of sentiment, not all five that could be specified by survey participants.

In this next exercise we will create a histogram that will allow us to see the full distribution when we hover over a bar.

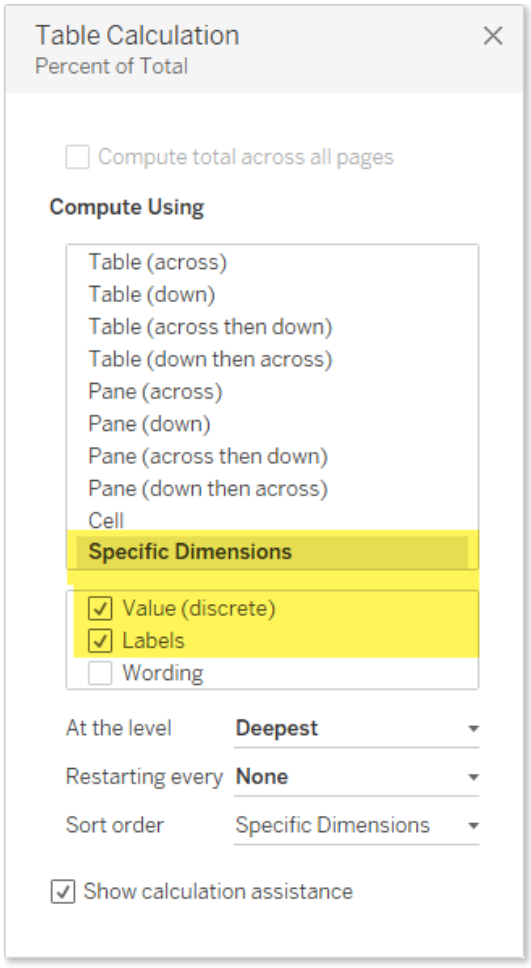


To Create the Histogram

Note: If you had difficulty completing the previous exercise, open the file **4d_Likert.twbx** from the **Starter** folder and work with that.

1. Duplicate the Divergent Stacked Bar sheet.
2. Rename the sheet **Histogram**.
3. Remove **Measure Values** and **% Neutral** from Columns.
4. Remove **Neutral Reference Line** from details.
5. Place **Value (discrete)** and **Labels** on Columns.
6. Place **Percent of Total** on Rows.
7. Place **Value (discrete)** on Color.
8. Right-click **Percent of Total** on Rows and select **Edit Table Calculation**.

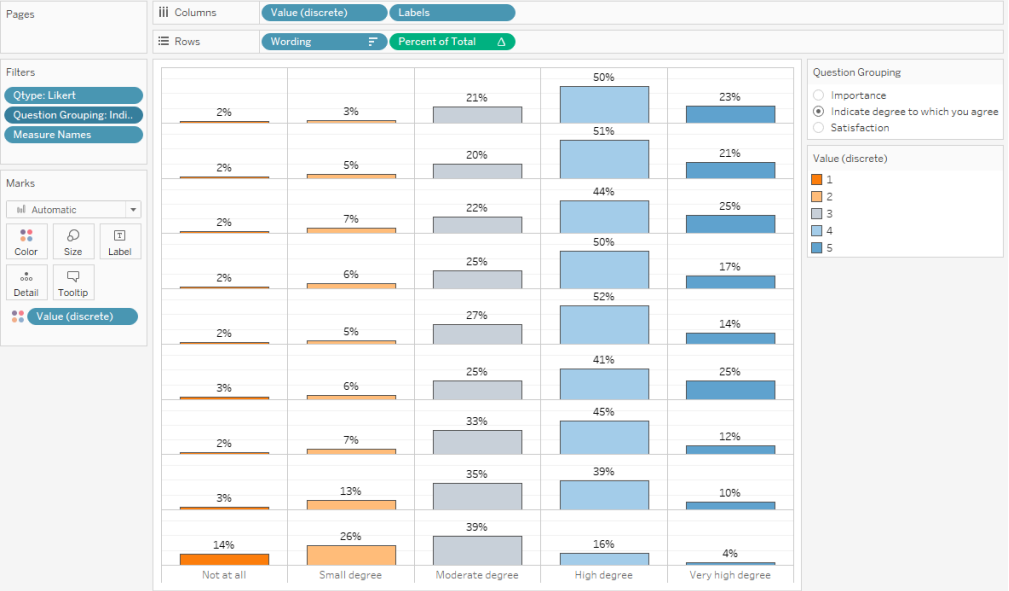
9. Change the scope to **Value (discrete)** and **Labels**.



10. Modify the colors using the Colorblind palette.

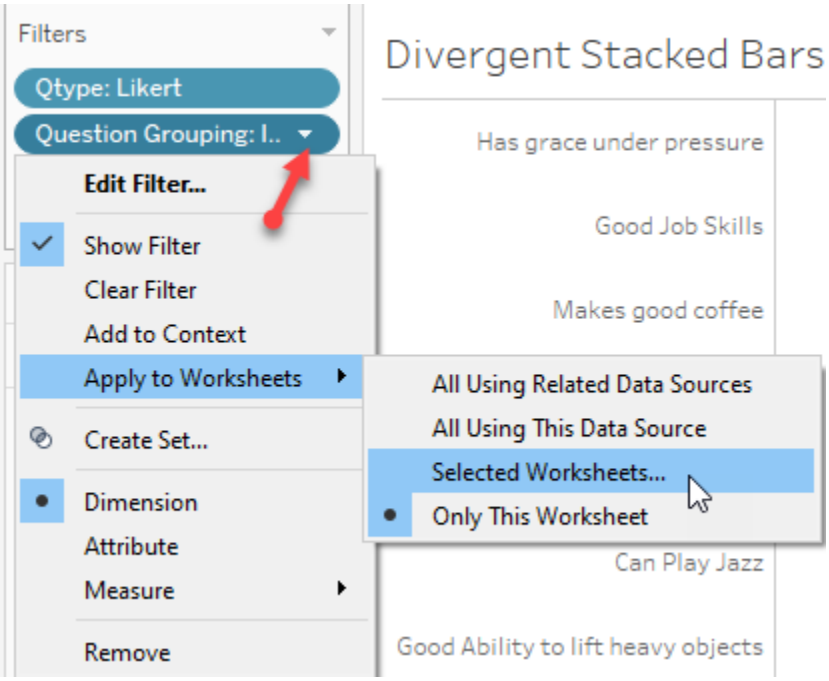
11. Right click in one of the vertical axes ("Percent of Total") and de-select **Show Header**.
12. Right click the text "Value (discrete) / Labels" near the top of the chart and select **Hide Field Labels for Columns**.
13. Right-click the numbers along the top and de-select **Show Header**.
14. Right-click the labels along the left (i.e., "Has grace under pressure")and de-select **Show Header**.

15. Hide the title bar.
16. Change the view from Standard to **Entire View**. Your screen should look like this.

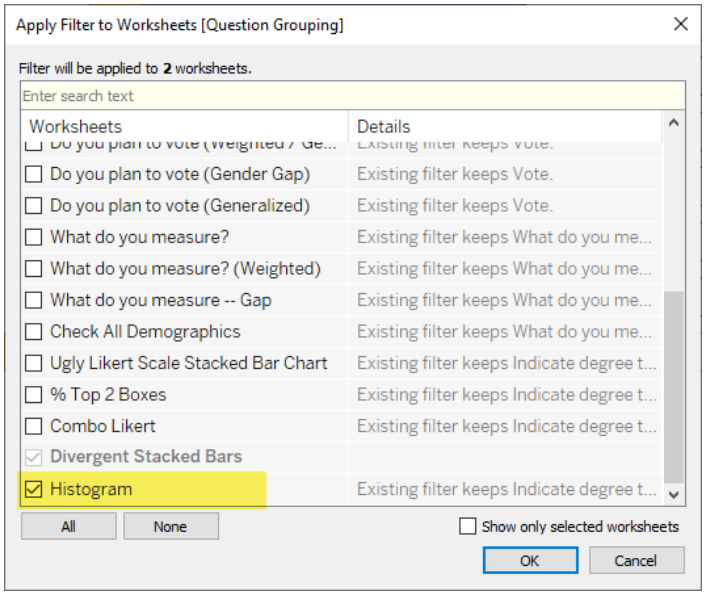


To Create the Tool Tip

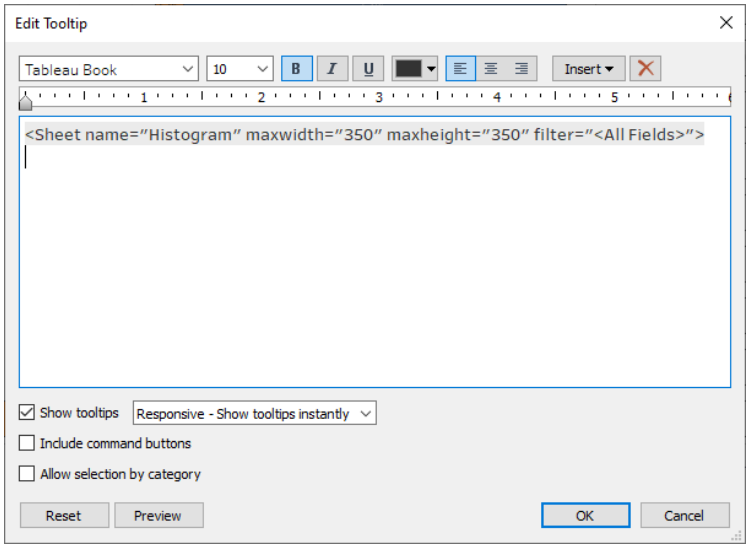
1. Go back to the Divergent Stacked Bar sheet.
2. Click **Question Grouping** in Filters and choose **Apply to Worksheets | Selected Worksheets**.



3. Make sure **Histogram** is selected (this is very important!)



4. Click **OK**.
5. Click the **Tooltip** button.
6. Delete everything that is currently in the Edit Tooltip dialog box.
7. De-select Include command buttons and Allow selection by category.
8. Click the **Insert** button at the top of the screen and select **Sheets | Histogram**.
9. Change the maxwidth and maxheight settings to 350. The dialog box should look like this.



10. Click **OK**.

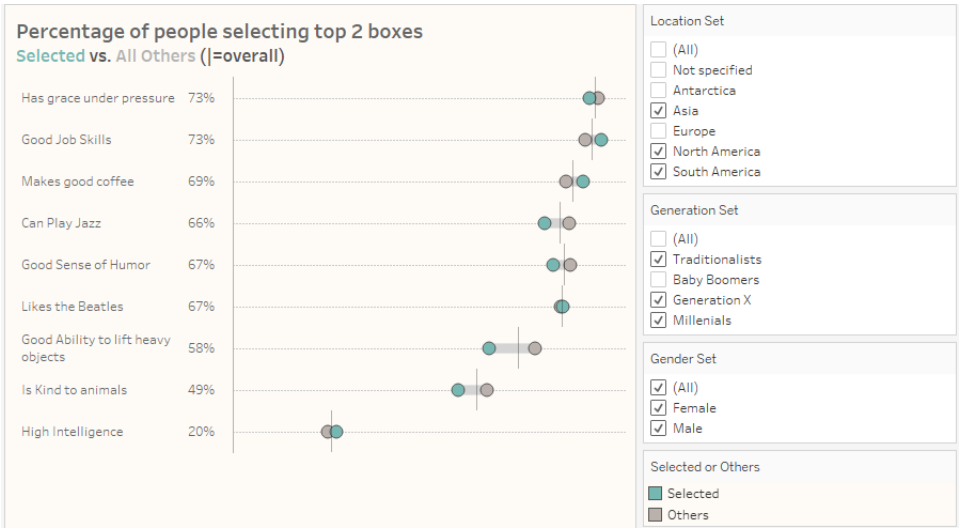
11. Hover over a bar and admire your work.
12. Save your work.

Topics for Discussion

We looked at gap charts for the single and multi-punch questions. Would something like this work for Likert-scale questions?

YES! First, please visit https://www.datarevelations.com/resources/setcontrols/.

Then, go into the **Likert Examples** folder that's in the **Completed** folder and open Data-Revelations_SetControls_Top2Boxes.twbx.



There are lots of other ways to visualize Likert-scale data. Please have a look at these posts:

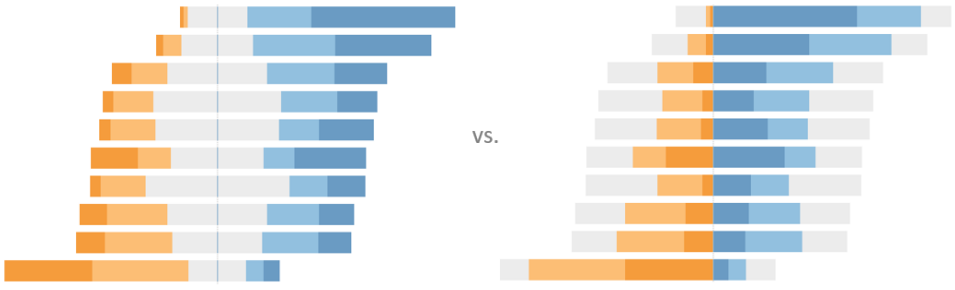
https://www.datarevelations.com/resources/got-likert-data-neutrals/

https://www.datarevelations.com/resources/rethinkingdivergent/

and

https://www.datarevelations.com/resources/howto-likert/ (this is my latest thinking on the subject)

then go into the **Likert Examples** folder that's in the **Completed** folder and open the other files… and explore!

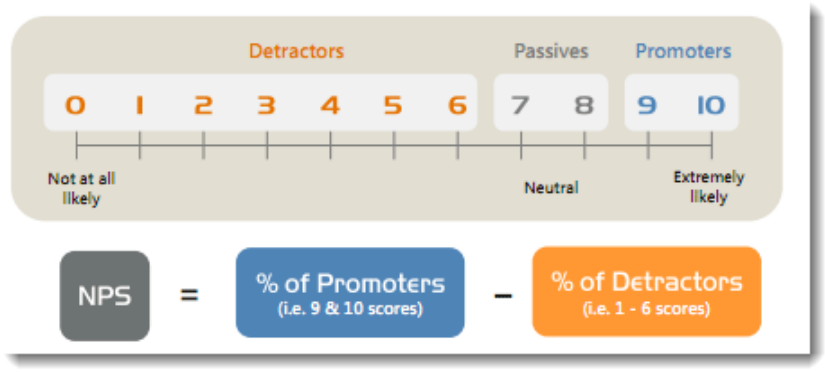


What about Net Promoter Score (NPS)?

In an NPS survey, respondents are presented with the question "Using a scale from 0 to 10, would you recommend this product / service to a friend or colleague?"

- Anyone that responds with a 0 through 6 is considered a Detractor.
- Anyone that responds with a 7 or 8 is considered a Passive (or Neutral).
- Anyone that responds with a 9 or 10 is considered a Promoter.

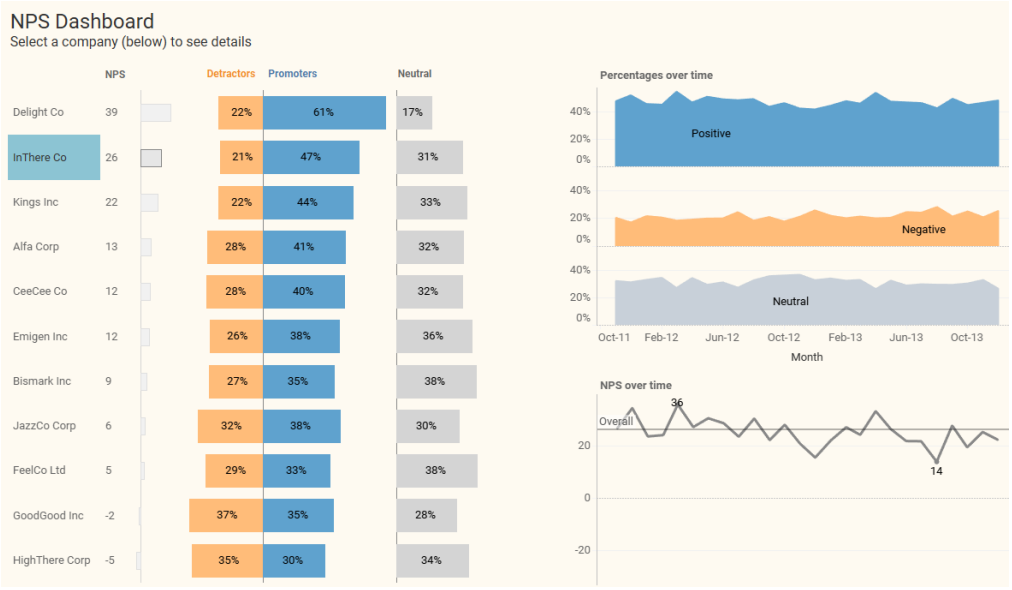
The Net Promoter Score (NPS) is computed by taking the percentage of people that are Promoters, subtracting the percentage of people that at Detractors, and multiplying that number by 100.



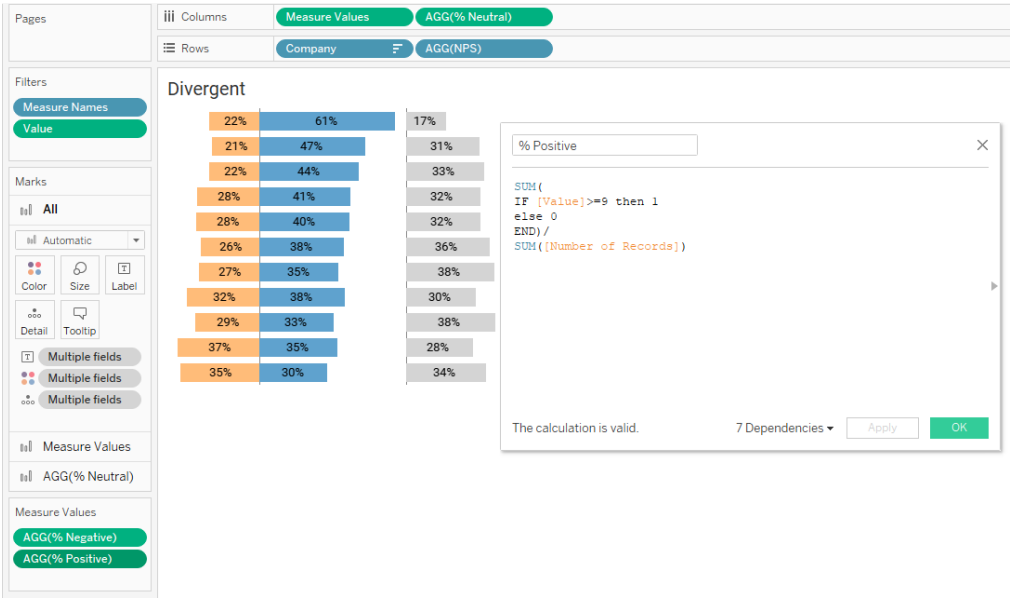
How would you present data like this in Tableau?

I think NPS is closely related to Likert data in that you are showing positive and negative sentiment (and neutrals, aka "passives.").

In the Completed folder there is a folder called Net Promoter Score. Open the file called **DataRevelations_NetPromoterScore.twbx**.



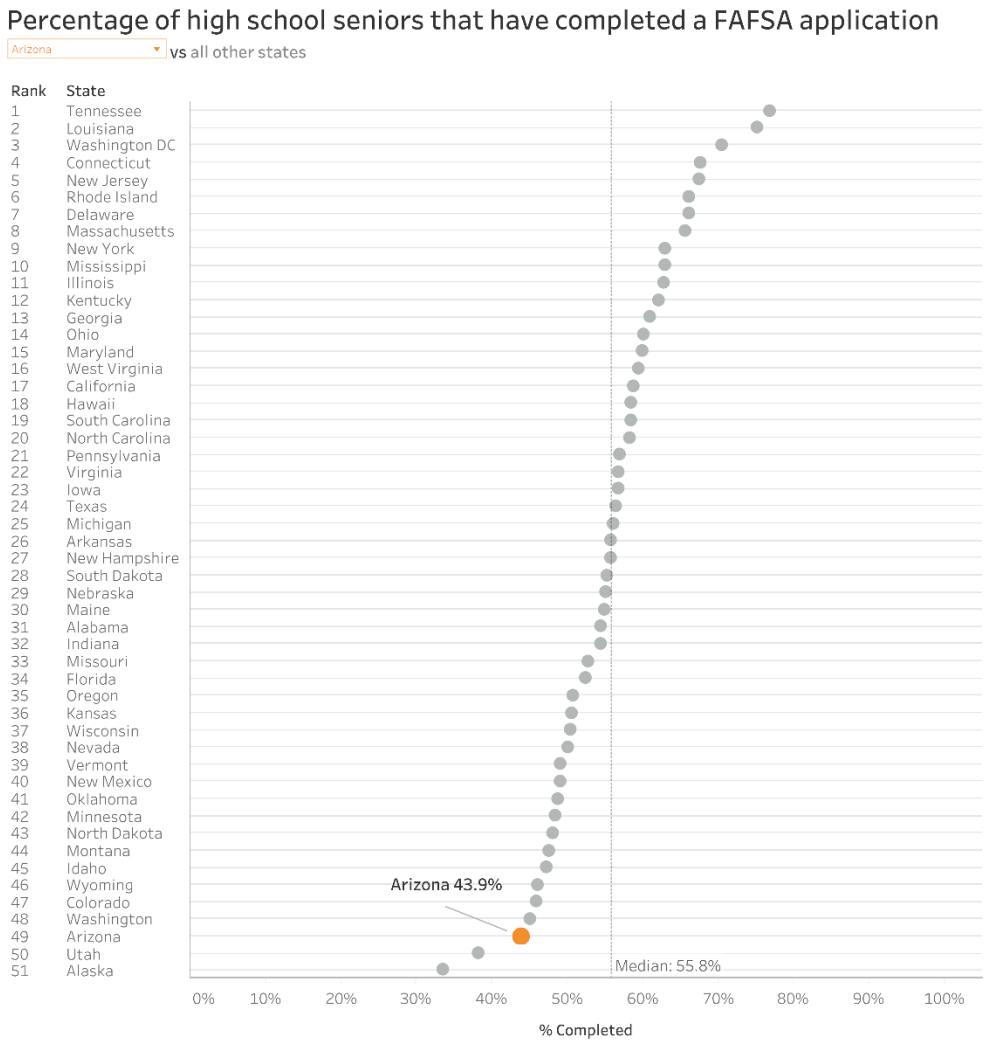
If you look under the hood, you'll see that there are very similar calculated fields to the ones we created for Likert scale data, but the values are a little different.



V. Benchmarking

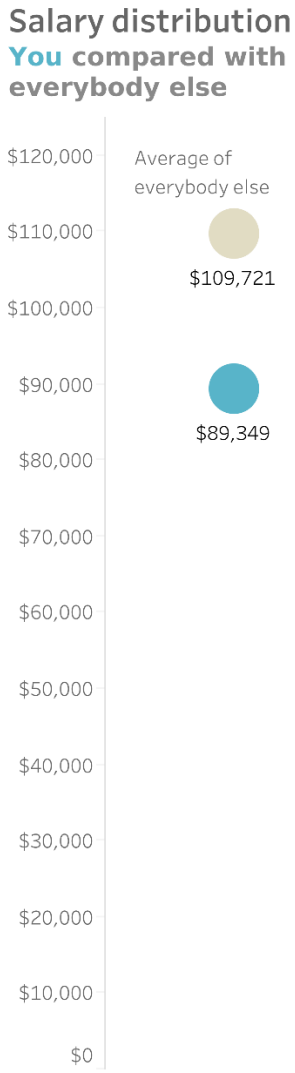
Overview

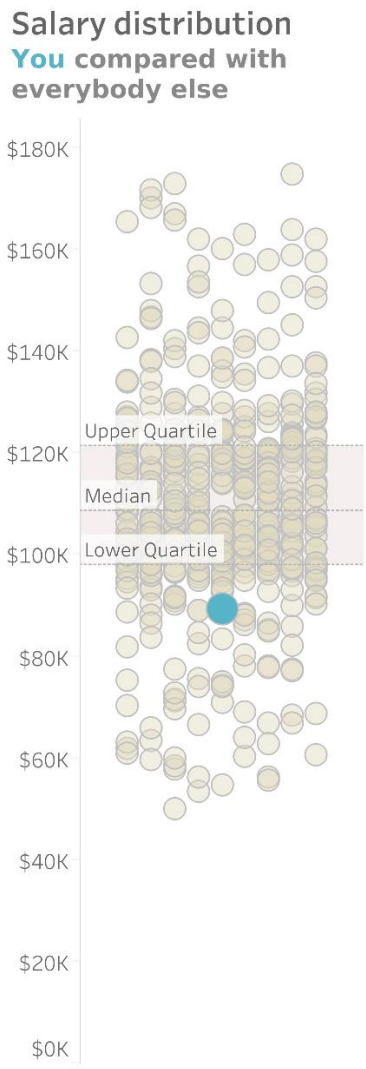
All of the results we've presented so far have been some type of summary statistics / aggregations. There may be times when you want to show how an individual respondents (or store, or school, or school district) compares with others, such as in this FAFSA completion dashboard shows how Arizona compares with other states.



As of July 2020.
Source: National College Attainment Network (NCAN.ORG)

Indeed, the intelligent display of individual respondents and shine light on data that a summary visualization obfuscates. Consider this data visualization that shows the salary of an individual compared with the average of all other survey respondents vs. a disaggregated view.





Same data, but what a different takeaway.

The chart on the right is called a jitterplot. Let's see how to make on.

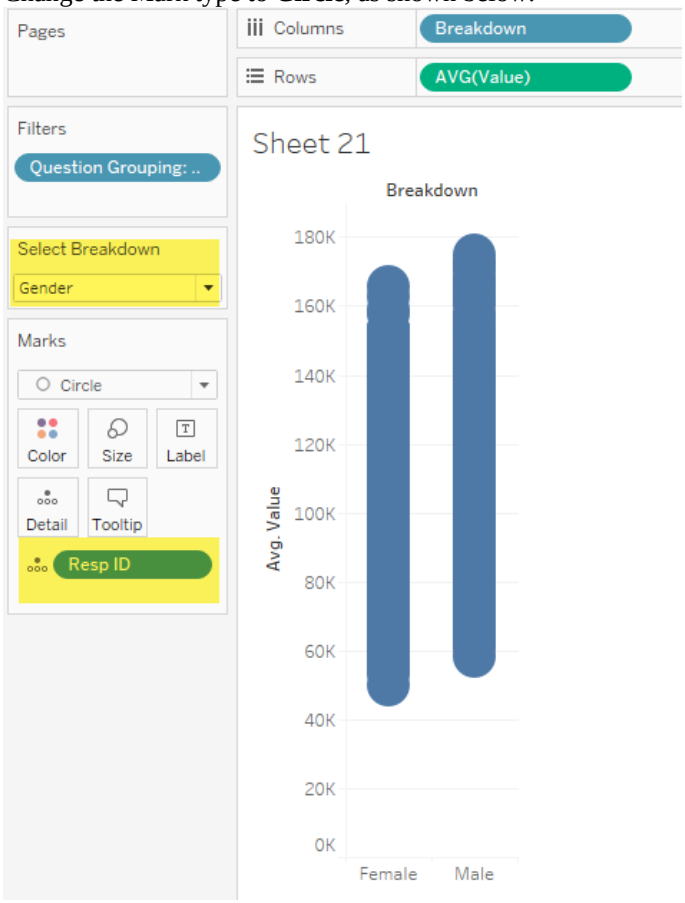
To Create a Simple Jitterplot

!

Note: If you had difficulty completing the previous exercise, open the file **4e_Likert.twbx** from the **Starter** folder and work with that.

1. Create a new worksheet.
2. Drag **Question Grouping** to the Filters shelf, select **Salary** and click **OK**.
3. Drag **Breakdown** on to the Columns shelf.
4. Right-drag **Value** onto the Rows shelf and select **AVG(Value).**
5. Right click the **Select Breakdown** parameter and select **Show Parameter**.

6. Select Gender from the Select Breakdown drop down. You now have a visualization that will compare the average salary for females vs. the average salary for males.
7. Drag **Resp ID** onto Detail.
8. Change the Mark type to **Circle**, as shown below.

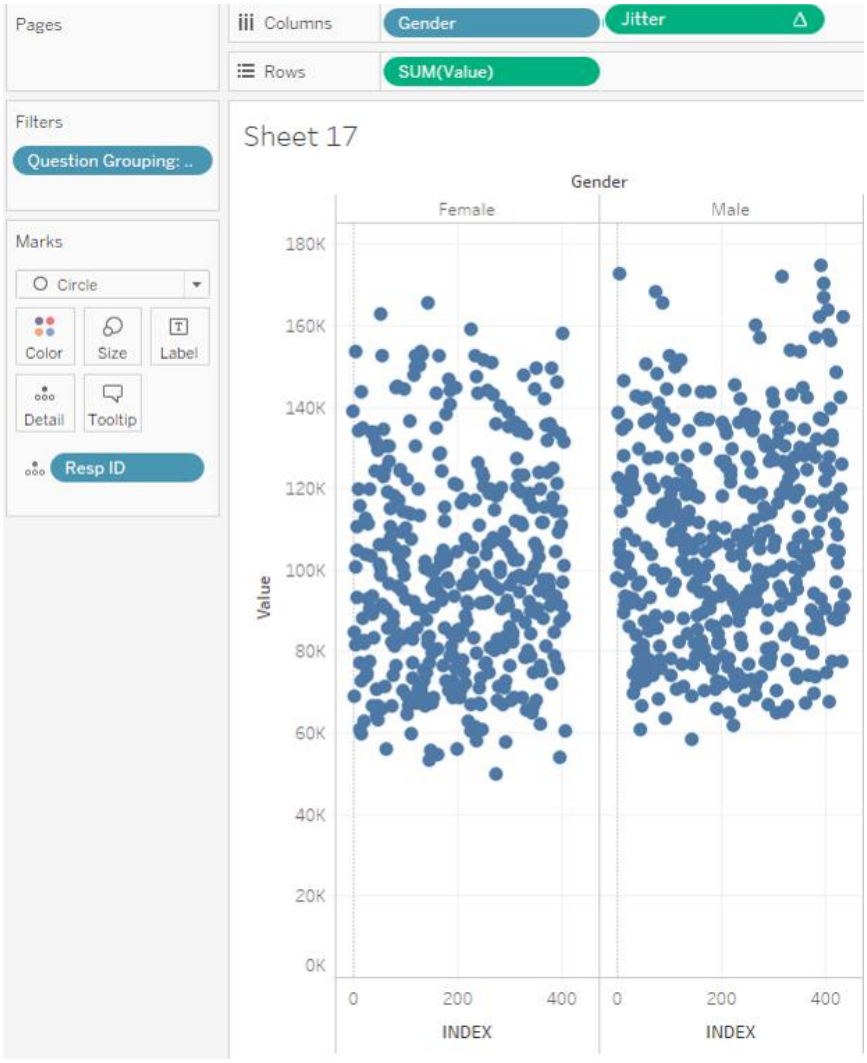


9. Create a new field called Jitter and define it as follows.

INDEX() *returns the current row in the partition. For example, for the first row* INDEX() *would be equal to 1; the second row it would be equal to 2.*

10. Drag **Jitter** to the Columns shelf.
11. Right-click **Jitter** on the Columns shelf, select **Compute Using**, and select **Resp ID**.

12. Make the size of the visualization narrower, as shown below.



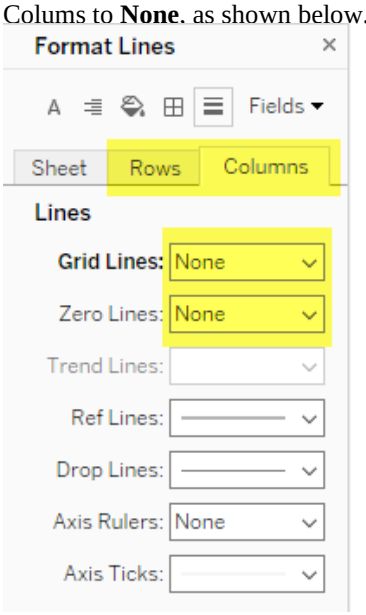
13. Rename the sheet **Simple Jitterplot** and save your work as a packaged workbook.

To Clean Up the Visualization and Add Reference Lines

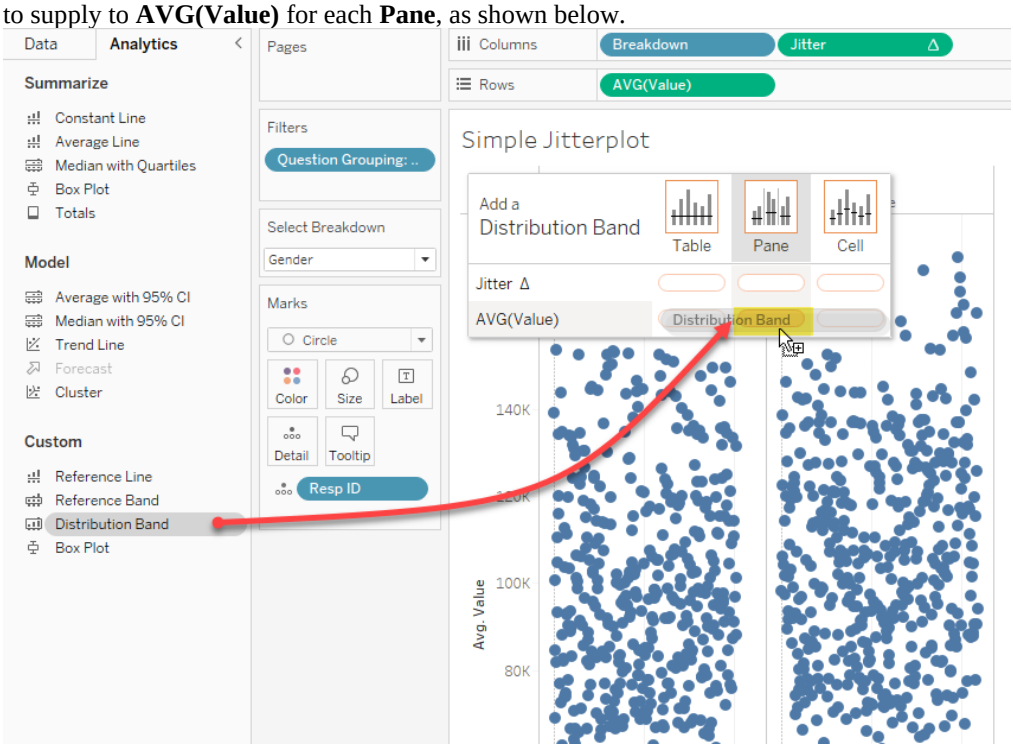
Note: If you had difficulty completing the previous exercise, open the file **5a_Benchmark.twbx** from the **Starter** folder and work with that.

1. Right click anywhere in the chart and select **Format** from the context menu.

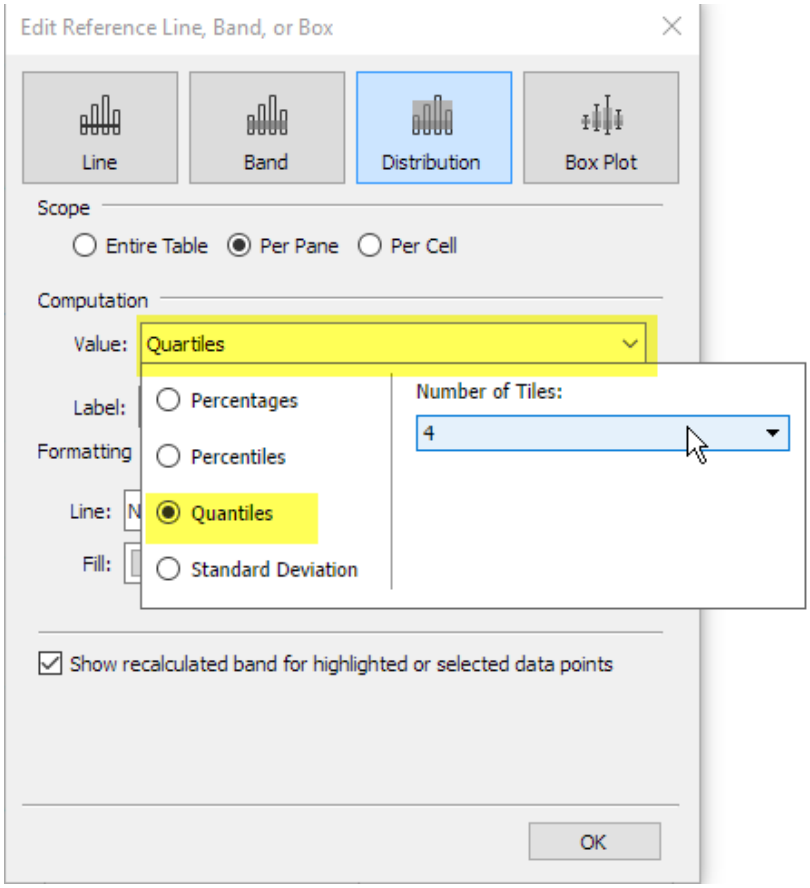
2. Click the **Lines** button and change **Grid Lines** and **Zero Lines** to for both Rows and Colums to **None**. as shown below.



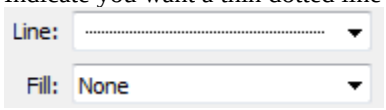
3. Close the Format pane by clicking the **X** in the upper right corner.
4. Right-click on the **Jitter** axis and turn off **Show Header**.
5. Click the **Analytics** tab.
6. Drag a **Distribution Band** from Custom onto the chart and indicate that you want it to supply to **AVG(Value)** for each **Pane**, as shown below.



7. Indicate that you want **Quantiles / 4** from the drop-down box, as shown below.

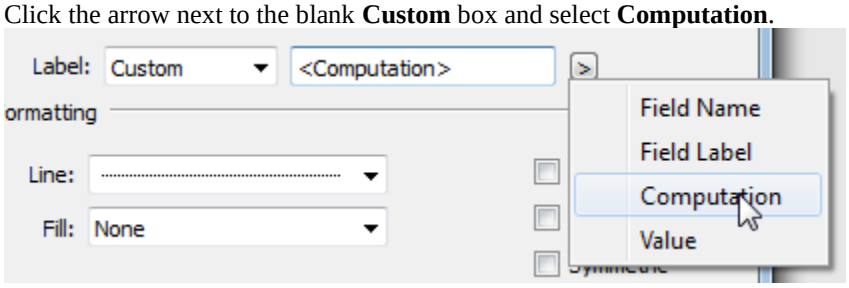


8. Indicate you want a thin dotted line and no Fill.



9. Click the **Label** drop down and select **Custom**.

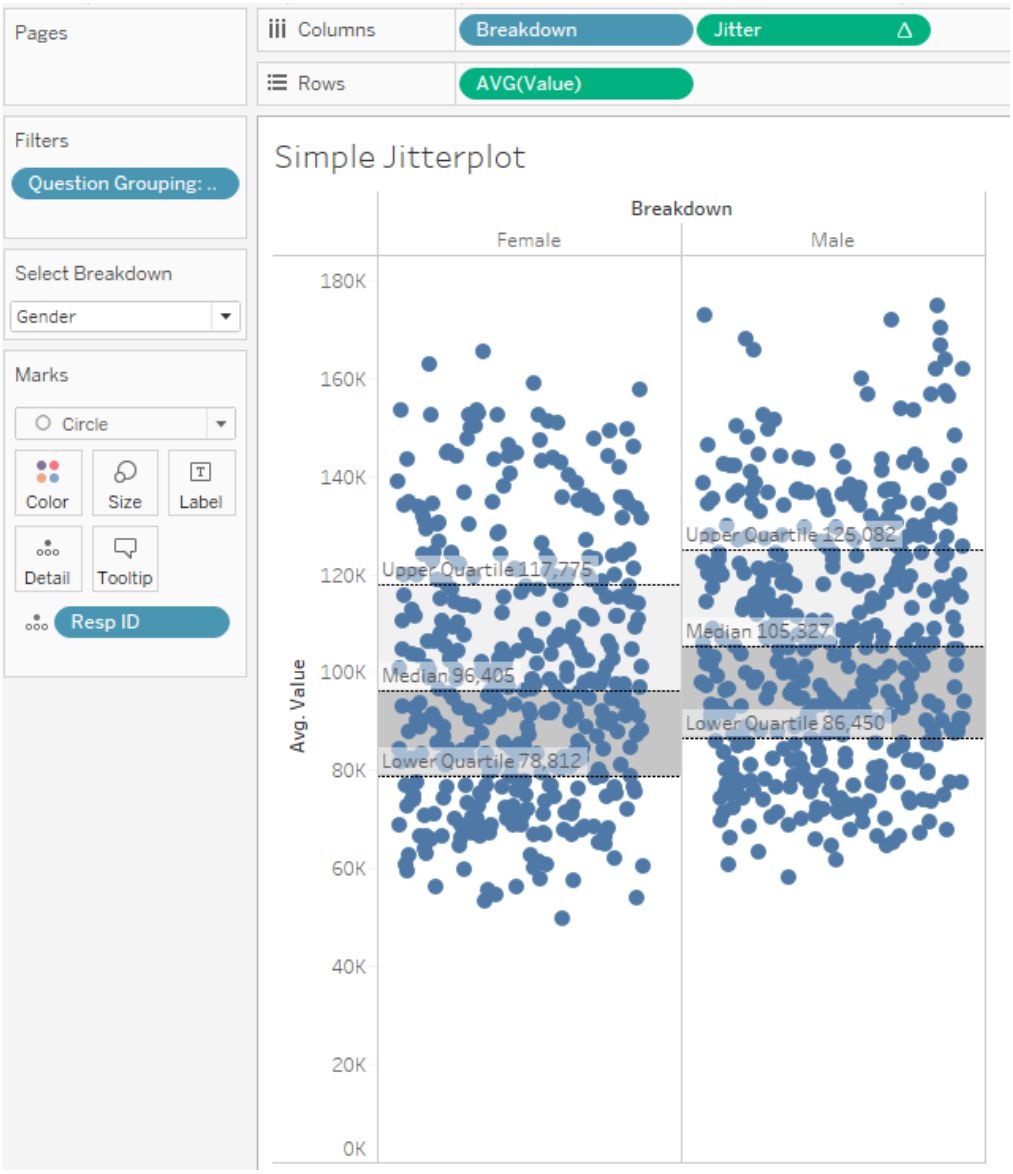
10. Click the arrow next to the blank **Custom** box and select **Computation**.



11. Type a space, click the arrow again, and select **Value**.

12. Click **OK** and save your work**.**

    Your screen should look like the one shown below.



Controlling Jittering

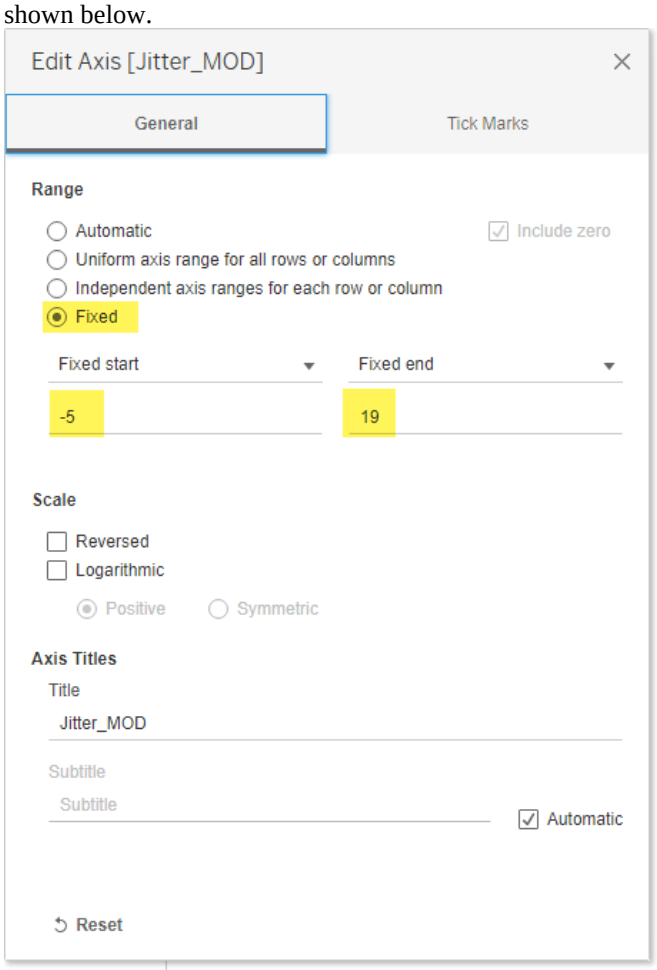
One of the reasons the jittering works here is that there is little or no relationship between the Resp IDs and their associated salaries. That is, Resp IDs are fairly random.

There may be time, however, when this is not the case and we can overcome this by forcing Tableau to put a mark into one of several pre-defined "slots" using a mod function.

There will be an added benefit in that this will allow us to specify a fixed axis and in doing so we can both narrow the strip and add some white space around the chart, making the reference lines easier to read.

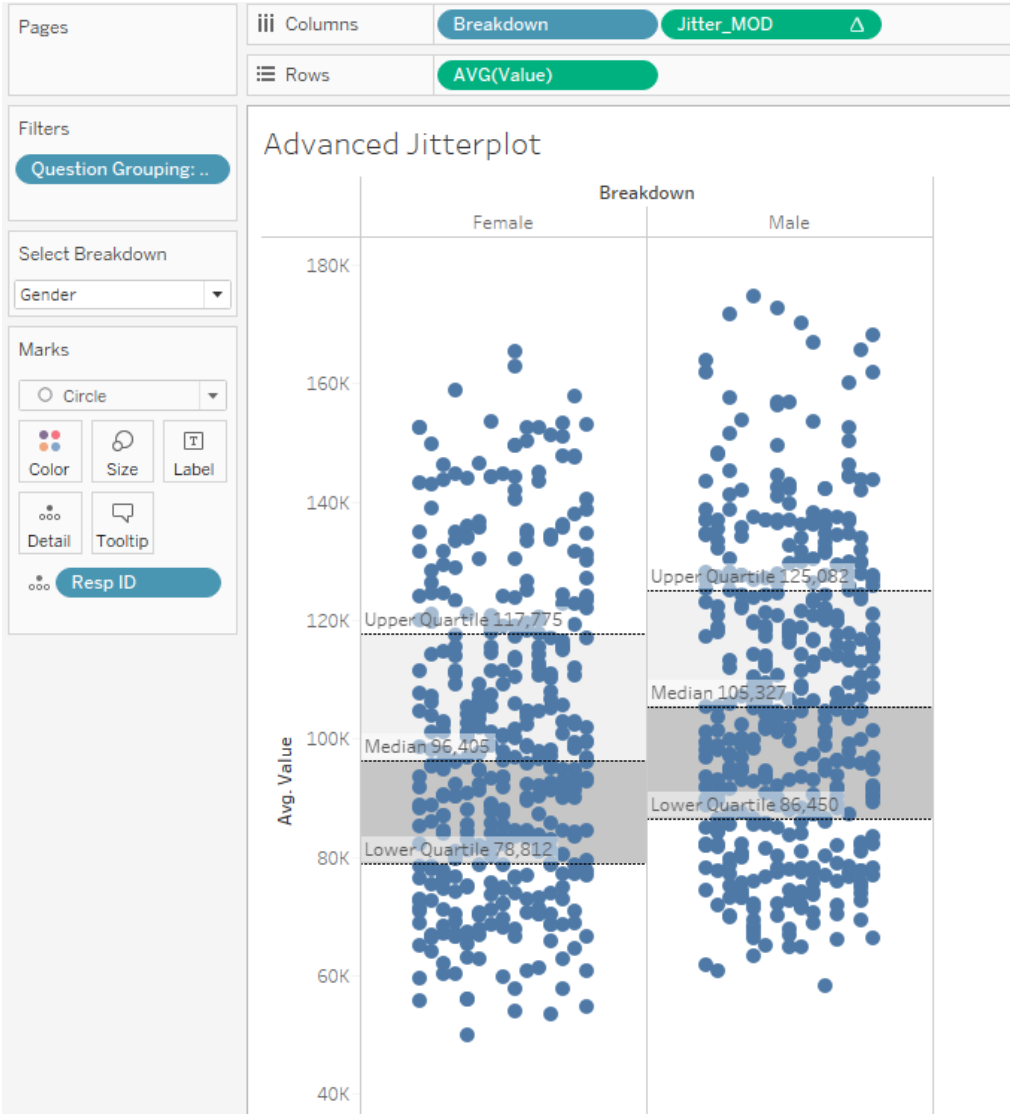
To Create a More Sophisticated Jitterplot

1. Right-click the current sheet and select **Duplicate**.
2. Rename the sheet **Advanced Jitterplot**.
3. Create a new field called **Jitter_MOD** and define it as follows.
   INDEX()%15
4. Replace Jitter on the Columns shelf by dragging **Jitter_MOD** on top of it.
5. Right-click **Jitter_MOD** on the rows shelf, select **Compute Using**, and select **Resp ID**.
6. Right-click in the **Jitter_MOD** axis and select **Edit Axis**.
7. Change the Range to **Fixed** and make the start value **-5** and the end value **19**, as shown below.



8. Close the dialog box.
9. Right-click the **Jitter_MOD** axis and turn off **Show Header**.
   Your screen should look like the one shown below.

INDEX()%15 *returns the remainder when you divide* INDEX() *by 15. For example, if the* INDEX *equals 1,* INDEX()%15 *will return 1. If the* INDEX *equals 2, then* INDEX()%15 *will return 2. If the* INDEX *equals 19 then* INDEX()%15 *will return 4. This means no matter what the ID number, the associated mark will be forced into one of 15 different slots.*



10. Save your work.

Comparing an Individual Response with All Others

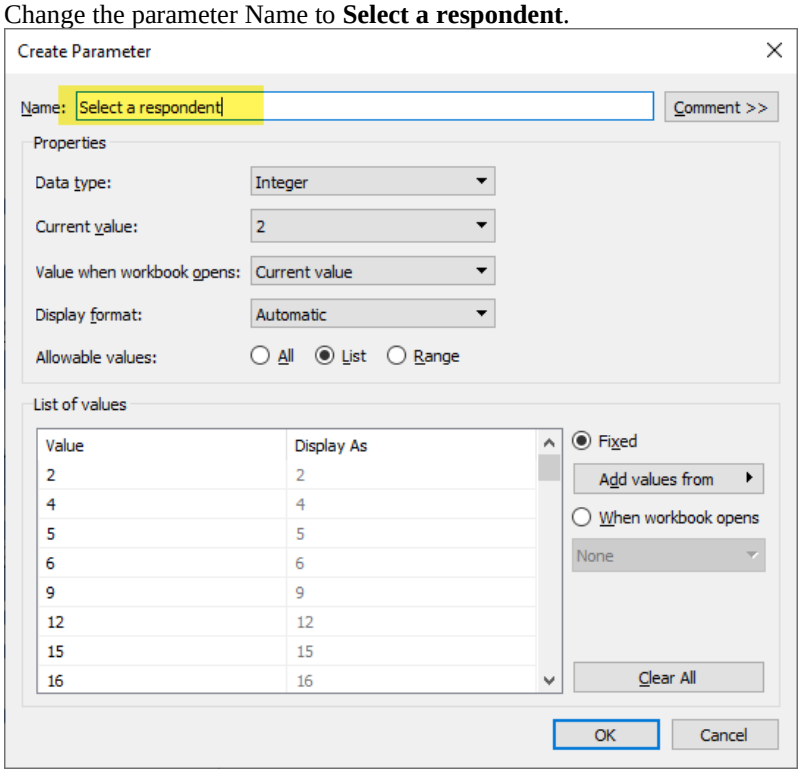
In the next series of exercises, we'll create a chart that will allow us to select a single respondent ID and see that person's salary compared with all others. We'll do that by populating a parameter list with all possible respondent IDs.

To Populate a Parameter List with Respondent IDs.

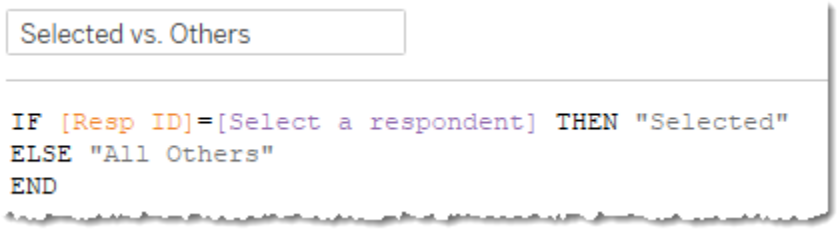
!

Note: If you had difficulty completing the previous exercise, open the file **5b_Benchmark.twbx** from the **Starter** folder and work with that.

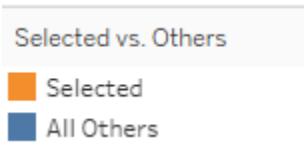
1. Right-click **Resp ID** in the Data pane and select **Create | Parameter**.
2. Change the parameter Name to **Select a respondent**.



3. Click **OK**.
4. Right-click the newly-created parameter and select **Show Parameter**.
5. Create a Calculated field called **Selected vs. Others** and define it as follow:



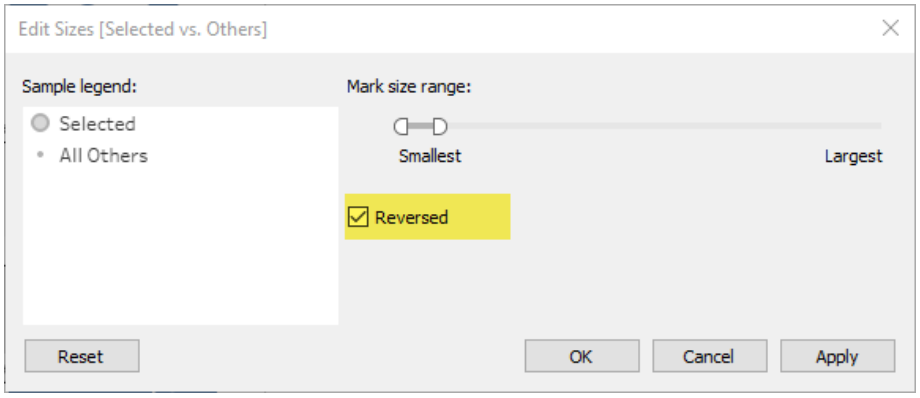
6. Drag **Selected vs. Others** onto the Color button. You probably won't see anything happen.
7. Within the color legend, change the order so that **Selected** comes before **All Others**.



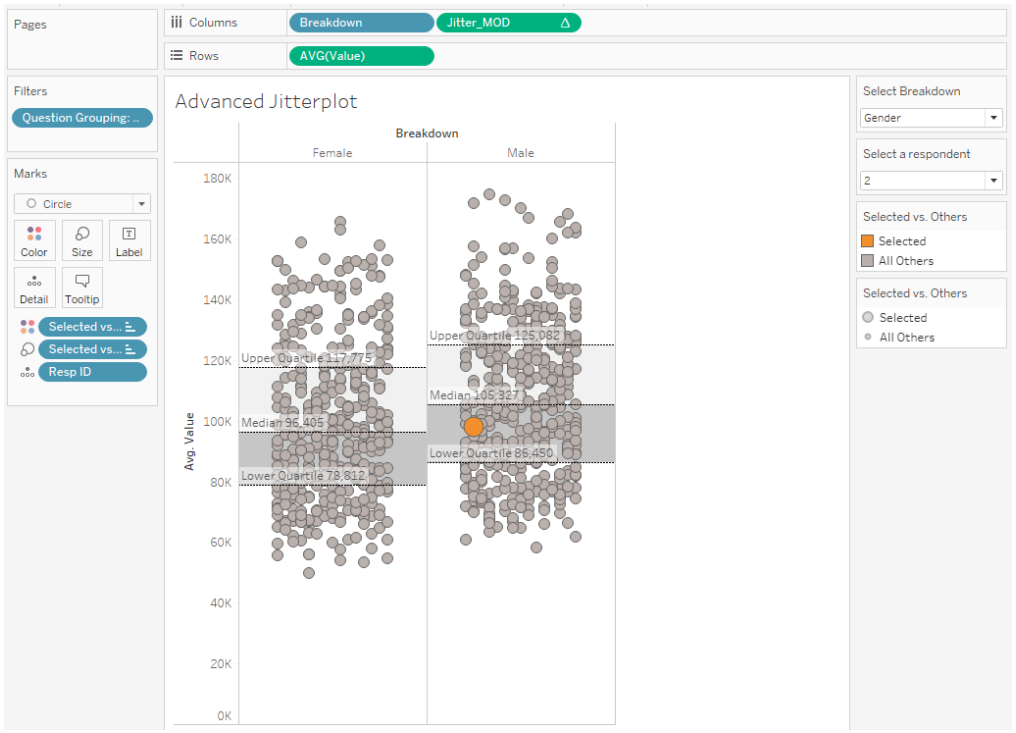
Now you should be able to see the "Selected" dot as you changed the Z-order of the elements so the selected dot will be on top.

8. Drag **Selected vs. Others** from the Data pane onto the Size button.

9. Edit the Size legend to that **Selected vs Others** is **reversed**.



10. Click **OK**.
    You may find it necessary to futz with the sizing a bit as the big dot may be too big and the other dots may be too small.
11. Edit the color legend so that **All Others** is a light gray and that all dots have a dark gray border.
12. Save your work.
    Your screen should look like the one shown below.



Bonus Exercise: Centering the Dot

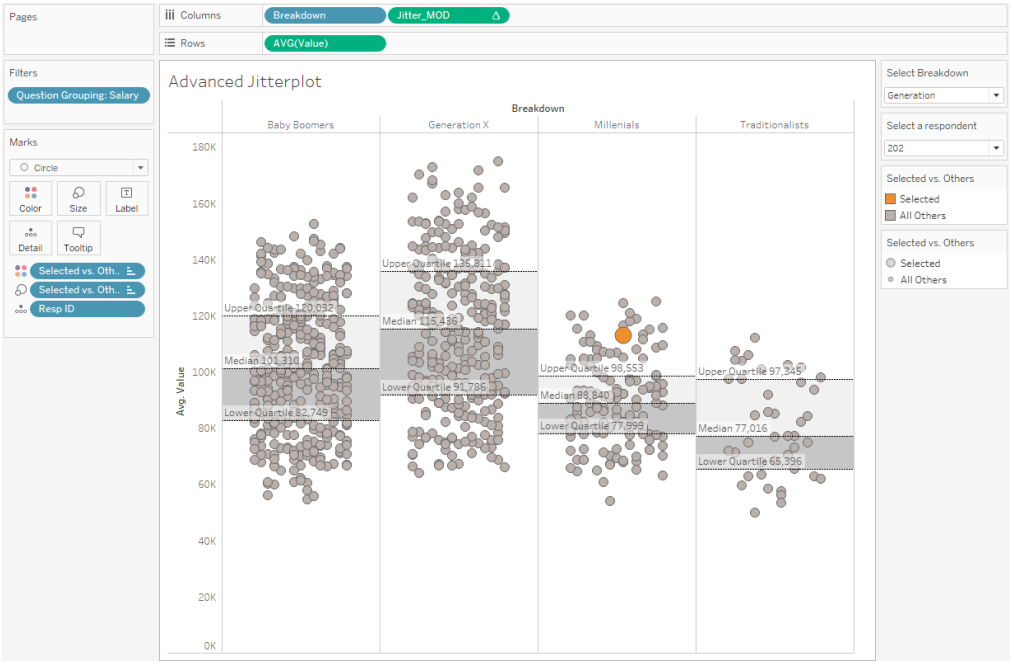
Modify the calculated field **Jitter_MOD** so it reads as follows:

Jitter_MOD

Results are computed along Table (across).

```
IF ATTR([Resp ID])= [Select a respondent] THEN 8
ELSE
INDEX()%15 END
```

This will place whatever is the dot in question in the center, as shown here.



Here's how to read the modified calculation:

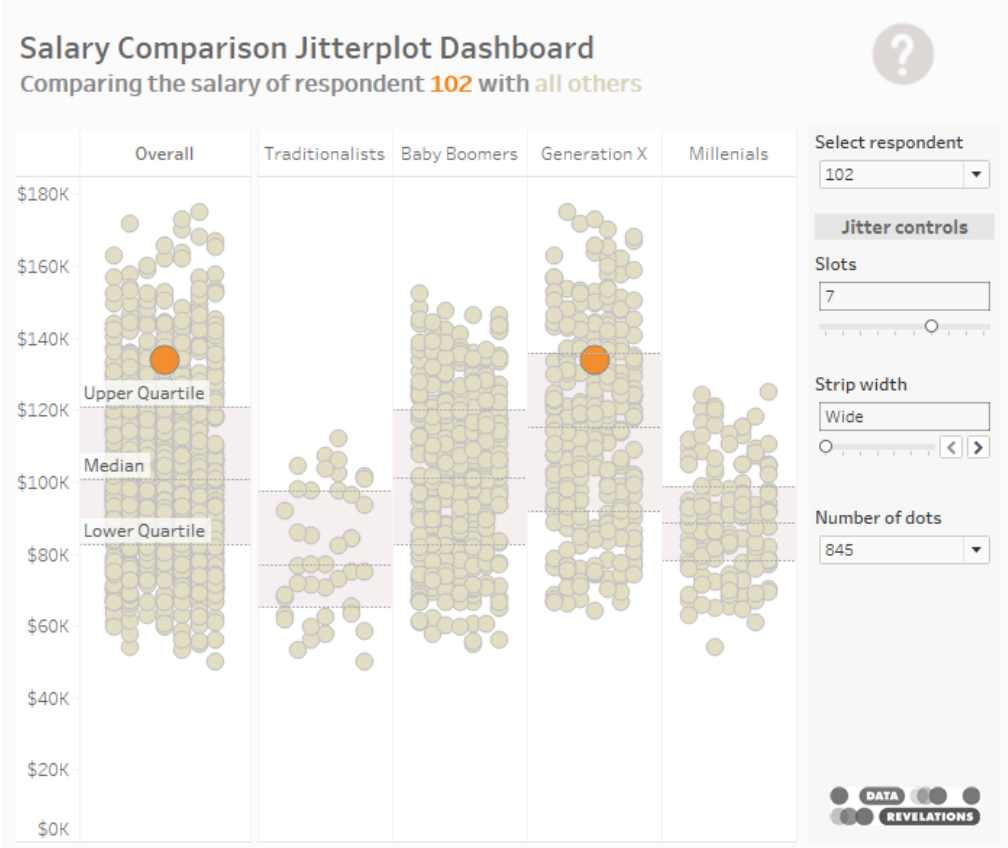
> If the Resp ID is the same as what is selected in the drop-down parameter, place that dot in the 8th position; if not, place it in some position between 0 and 15, depending on what row the item is in the data partition.

Why 8? It's the mid-point of 0 and 15.

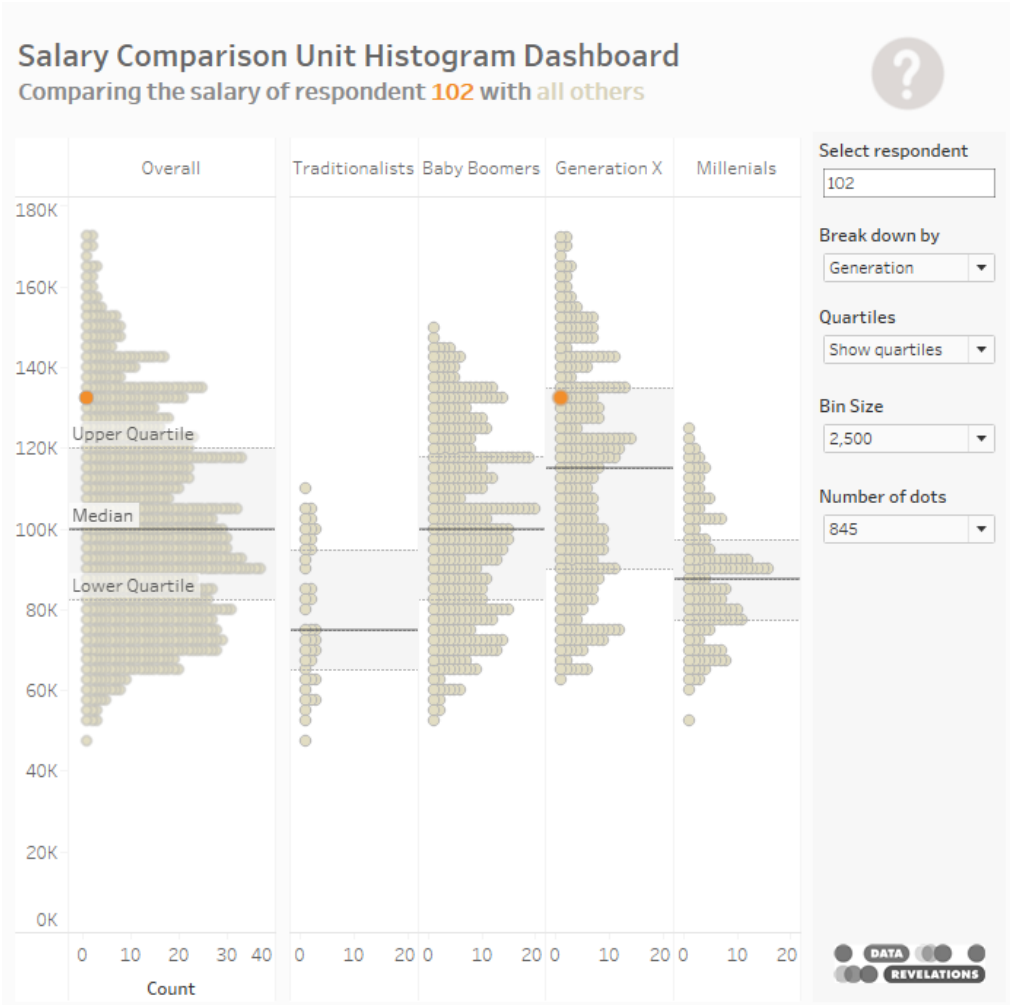
Why the ATTR() function? Because INDEX() is a table calculation which is a form of aggregation and Tableau can't compare an aggregate calculation and a row-level calculation.

Other Ways to Show This

Open the **file DataRevelations_Benchmarking.twbx** in the Benchmarking folder. You'll see variation on the visualization we just created.



If you click the next tab, you'll see a unit histogram (also called a Wilkinson dot plot). This is another useful way to present the same data.



You can read more about the jitterplot and unit histogram here:

https://www.datarevelations.com/resources/fresheyes/

https://www.datarevelations.com/resources/betterthanjitterplot/