

CSE 91

What Happened?

Joining Data

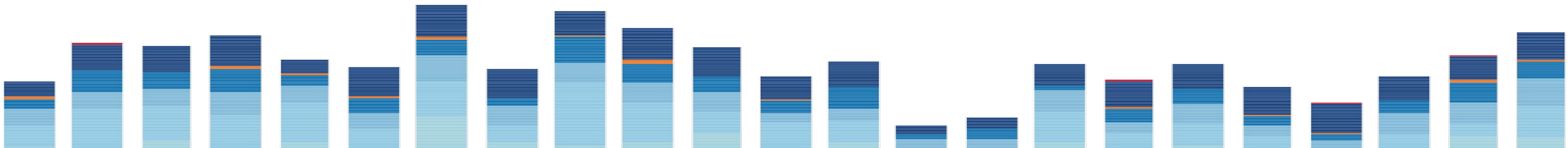


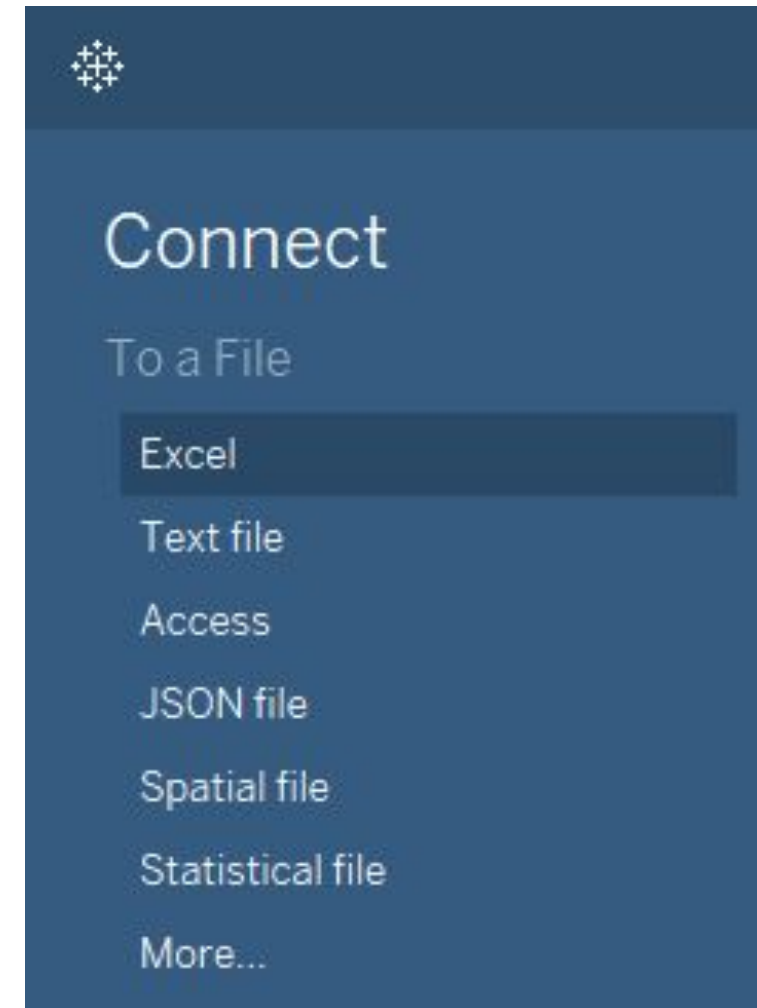
Tableau basics

- Importing
- Joining data
- More with calculations and filters
- Finding the story – more with chart types

Getting started

Let's start by creating a new workbook and importing our data again. But this time we're going to add a twist to it.

Remember how we wanted to account for the size of the city or county? We can do that if we import population data too and join it to our lobbying data. Lucky for us, our data has a column for a FIPS code. Every jurisdiction has its own code or Federal Information Processing Standard. We can use that to join in the population data to the fields where there is a FIPS code – the cities and counties.



The power of joins

The best journalists know how to make connections between seemingly disparate things. One of the first data journalism projects back when using computers was a new thing involved a list of felons and a list of school bus drivers. The journalist – Elliot Jaspin – used a computer to join the two databases and found felons who were not supposed to be near children driving the buses. But first, he needed to learn about joins.

- Open a new workbook and connect to lobbyingdata.xlsx
- Next, we add our censuswa.xlsx file. Do that by clicking **Add** in the connections area of your screen.

Note: In this case, we want an extract of the data-sets, not just to make a live connection.



Connection
☐ Live ☒ Extract [Edit](#) [Refresh](#)
Extract will include all data.

The power of joins

Tableau automatically previews the first data set you selected – LOBBYINGDATA1.

But how do we bring in the second data set? Just drag it into the area next to sheet1 from the lobbying data.

Tableau - tableaudj-module3

File Data Server Window Help

Connections Add

- lobbying Excel
- Census-WA Excel

Sheets

☐ Use Data Interpreter
Data Interpreter might be able to clean your Excel workbook.

- lobbying
- census1**
- New Union

lobbying+ (Multiple Connections)

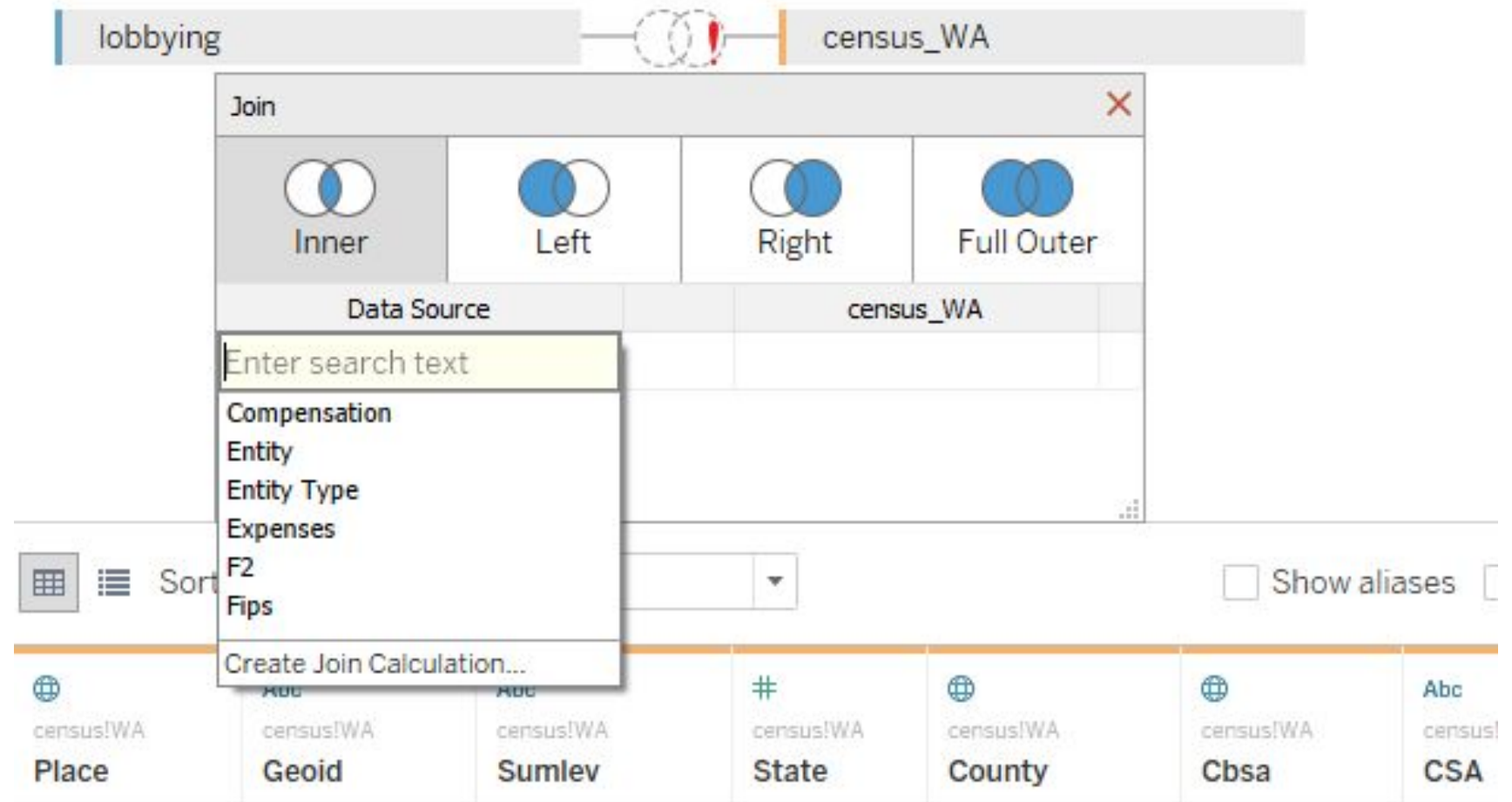
Connection: ☐ Live ☒ Extract | Edit Refresh Filters: 0 | Add
Extract will include all data.

Sort fields: Data source order ☐ Show aliases ☐ Show hidden fields 131 rows

lobbying Fips	lobbying Entity	lobbying Compensation	lobbying Expenses	lobbying Entity Type
5301290	Algona	18,000.00	0.00	CITIES
5302585	Arlington	13,687.50	0.00	CITIES
5303180	Auburn	1,900.00	0.00	CITIES
5304475	Battle Ground	12,000.00	0.00	CITIES
5305210	Bellevue	19,000.00	0.00	CITIES
5305280	Bellingham	5,733.32	4,332.68	CITIES
5308850	Burien	18,000.00	0.00	CITIES

The power of joins

When you do that, Tableau will open a dialogue box for you to select how to join this data. There are several options for different types of joins.

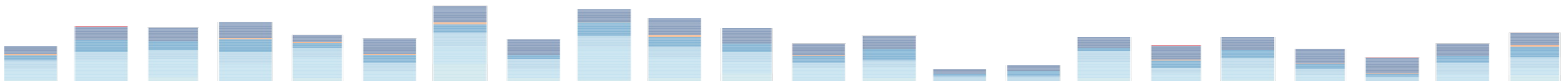


The power of joins

When you do that, Tableau will open a dialogue box for you to select how to join this data. There are several options. And these hold true for any software you may use to join data.

Let's go over the options:

- **Inner Join.** The records where the IDs match in both data sets. *Example: Only the records where a felon matches with a school bus driver.*
- **Left Join.** You get all the records from the data on the left side of your equation and any time the IDs match, you also get the records from the right side of the equation. *Example: All the school bus drivers and only records from felons when there is a match.*
- **Right Join.** You get all the records from the data on the right side of your equation and any time the IDs match, you also get the records from the right side of the equation. *Example: All the felons and only records from the school bus drivers when they match.*
- **Outer Join.** You add all the records from each data set together, even when there is no join. *(used rarely)*

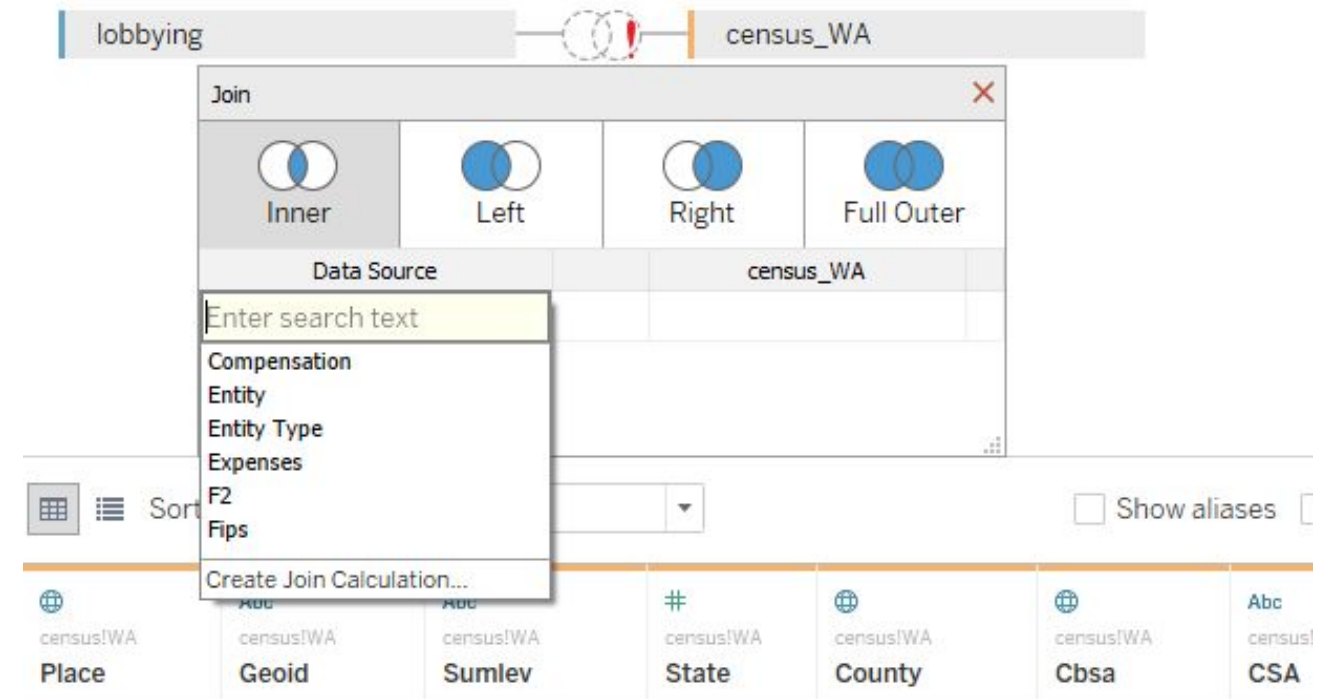


The power of joins

But before we can do this join, we have to fix an issue that sometimes occurs.

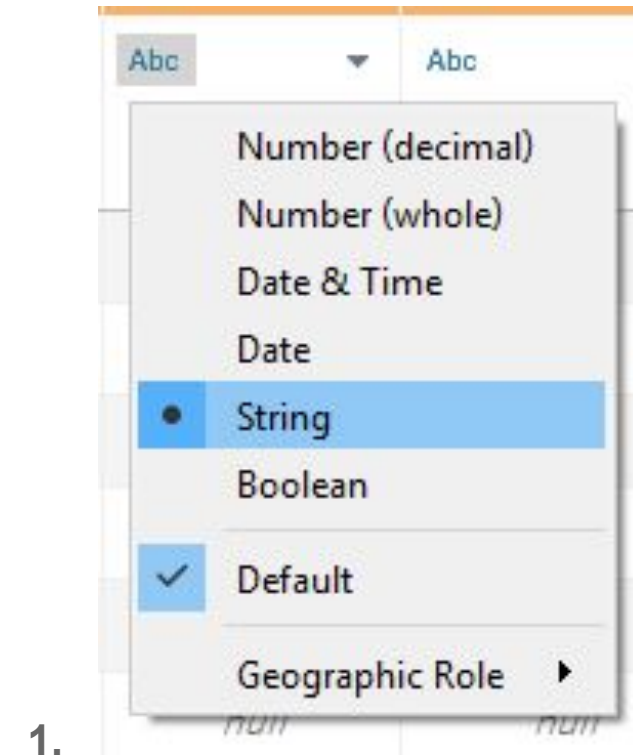
You'll note a red exclamation point just to the left of the census_WA sheet. That's telling us there is a problem. The problem in this case is a type mismatch. That means that the ID field on which we will join these two different data sets doesn't match because in one set of data, it is formatted as text and in the other, it's formatted as a number.

Even though the ID is made up of numbers, it shouldn't be formatted that way because we don't do math with it.



The power of joins

1. Okay, let's fix it. Click on the field for the ID and make sure **string** is selected. The field we are connecting the two datasets on are FIPS in sheet1 and GEOID in census_WA. Both should be formatted as string fields.
2. Once you format the fields, you'll notice that the exclamation point disappears.
3. You should also select extract instead of live – that will improve your performance within Tableau.



3.

Connection
○ Live ● Extract | Edit Refresh
Extract will include all data.

The power of joins

Tableau does a great job of explaining joins in a visual way. Let's take a look:

- Go ahead and select sheet 1 in the bottom left of your window to go to the Tableau Workbook. Now, that you have two different data sets, you will see both of them represented on the left hand side under Dimensions and Measures.

Now, how can you take advantage of the Census data we now have attached to our lobbying information?

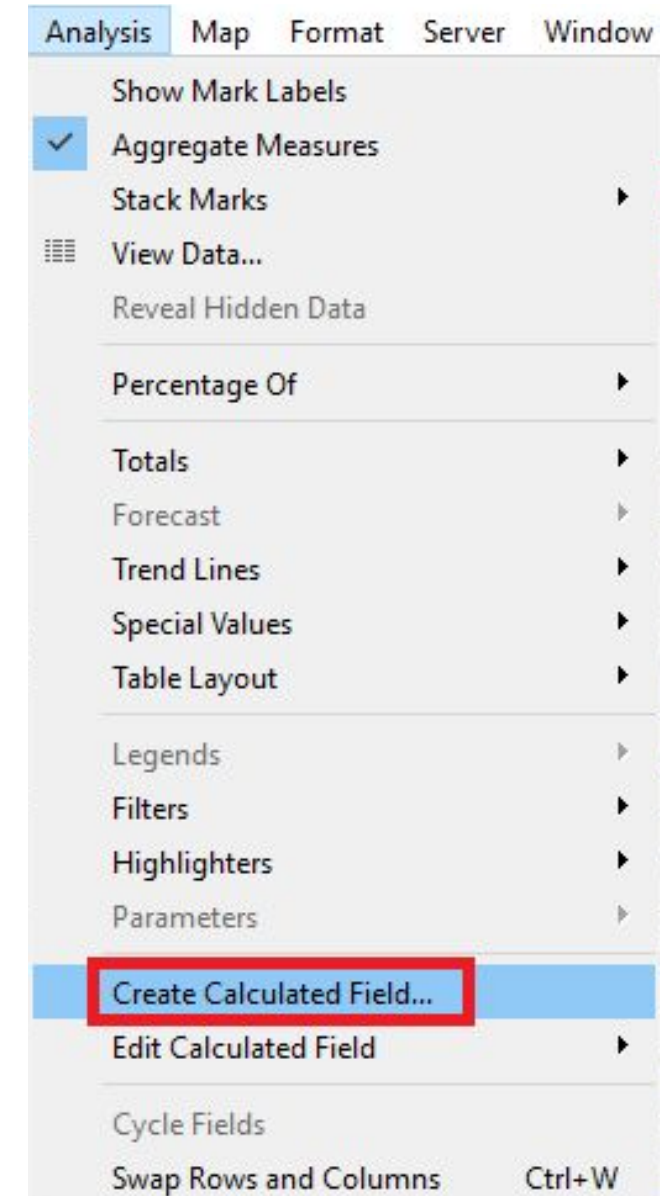
The image shows the Tableau interface with two data sources joined: 'census_WA' and 'lobbying'. The 'Dimensions' shelf on the left contains fields from both sources, including geographical and demographic data from 'census_WA' and organizational data from 'lobbying'. The 'Measures' shelf on the right contains fields from both sources, including population metrics from 'census_WA' and financial data from 'lobbying'.

Dimensions	Measures
▼ census_WA	▼ census_WA
Cbsa	# POP2010
Cnecta	# State
County	▼ lobbying
CSA	# Compensation
Geoid	# Expenses
Necta	=# per_capita
Place	=# total
# POP2000	# Latitude (generated)
Sumlev	# Longitude (generated)
Type	=# Number of Records
▼ lobbying	# Measure Values
Entity	
Entity Type	
Fips	
total (bin)	
Measure Names	

The power of joins

Let's create some calculated fields:

- We will re-create that calculated field for total. See if you can remember how.
- Now, let's create a per capita – a rate – so we can tell how much was spent on lobbying per person in the cities and counties.



The power of joins

Let's redo our calculated field for total:

×

[Compensation]+[Expenses]

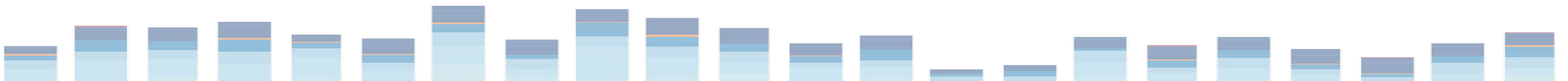
▶

The calculation is valid.

Sheets Affected ▼

Apply

OK



The power of joins

Let's calculate the money spent on lobbying per person:

×

`[total] / [POP2010]`

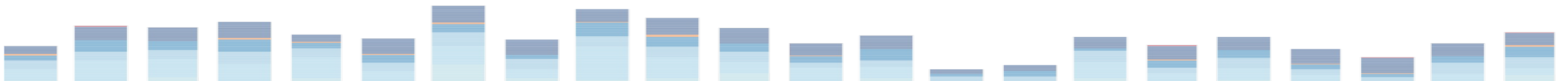
▶

The calculation is valid.

Sheets Affected ▼

Apply

OK

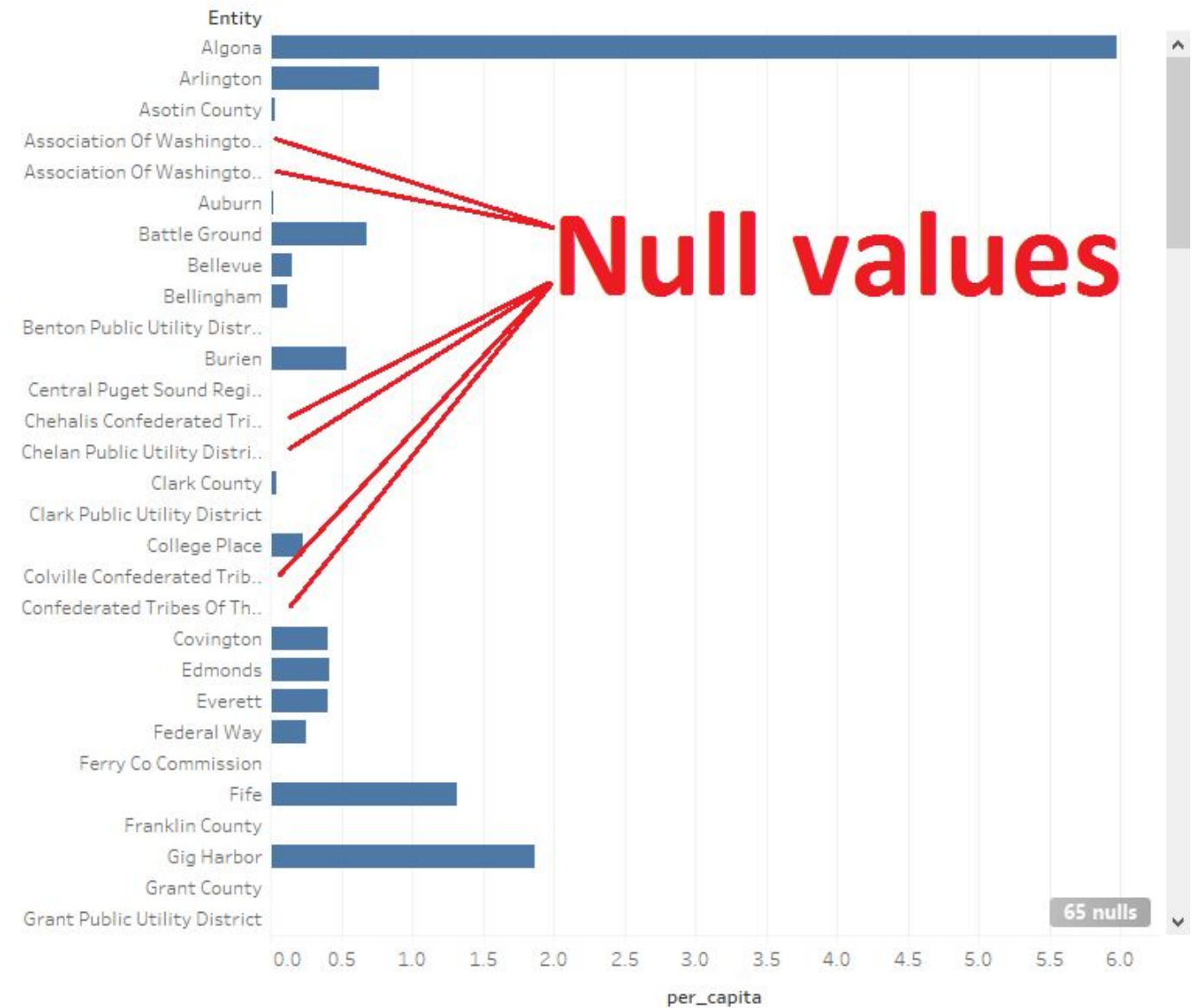


The power of joins

Let's drag **per_capita** onto the columns shelf and entity onto the rows shelf.

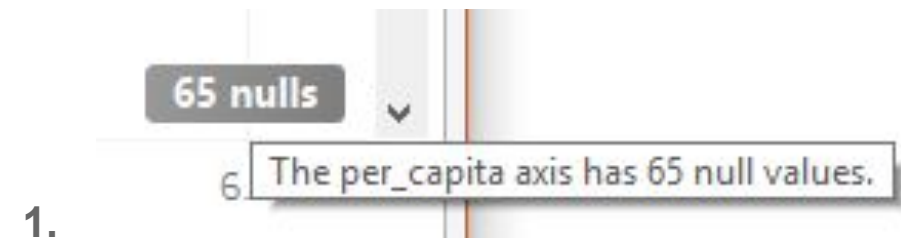
Note that there is only a rate calculated where there is a population amount. Some entities don't have population and so there is nothing in the per capita field.

Sheet 1



The power of joins

We can look at the data without those records. Look at the bottom right of your canvas or workbook. You'll see that Tableau has already done the hard work for you and identified 65 records with null values.



1. Click on the grey pill that says **65 nulls**.
2. A dialogue box will pop up and you can filter out the null values.



Navigating Tableau – the power of joins

Now, you'll see that the null values are gone. Let's filter that per_capita in descending order.

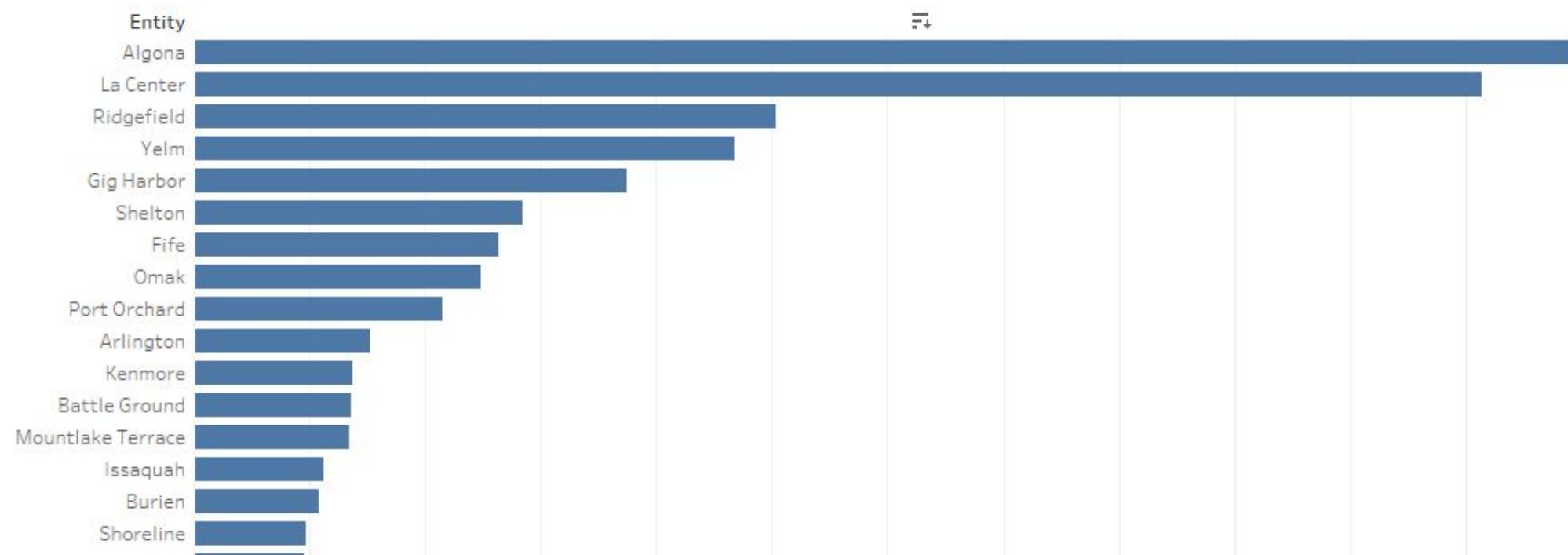
Sheet 1



The power of joins

Now, you'll see that the null values are gone. Let's filter that per_capita in descending order.

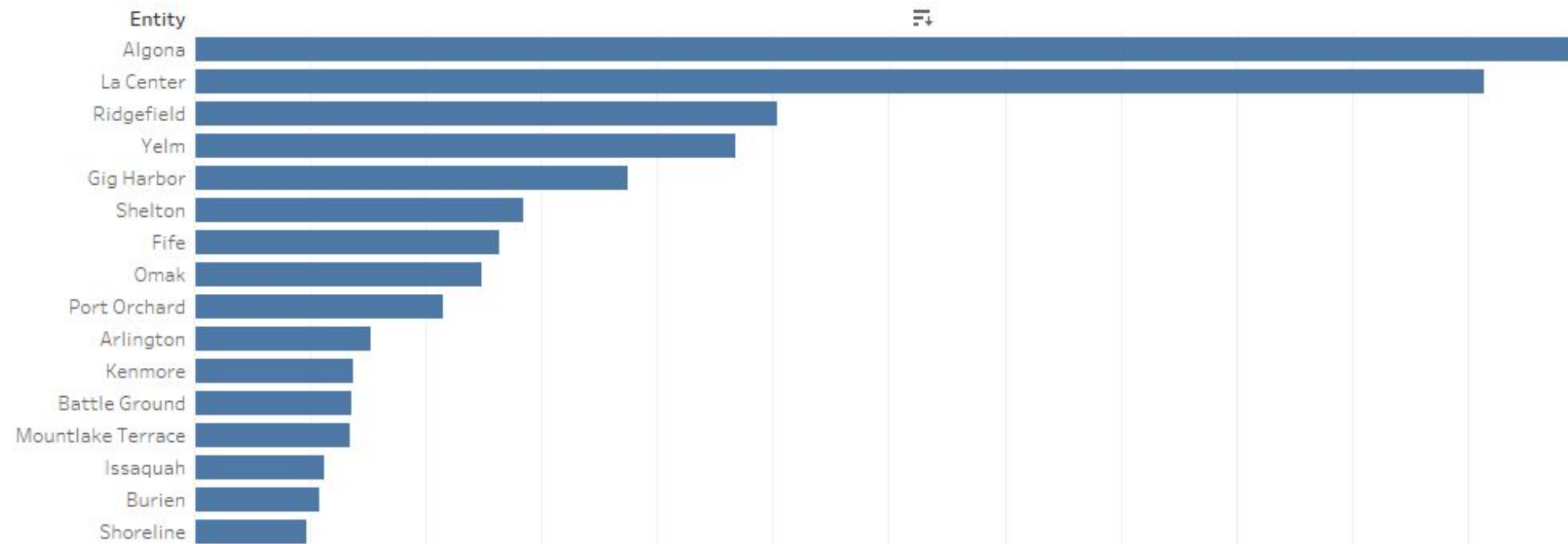
Sheet 1



The power of joins

So, do we think we know what the story might be?

Sheet 1



The power of joins

Let's add in population to the tool tip so we can take a closer look. This is easy to do.

- Drag Pop2010 onto the tooltip icon on the **marks** shelf.

That field will show up below the icons with the visual cue to the left that it's associated with the tooltip. Now, when you hover over the bar chart, you can get a sense of the population of the community as well as the rate.

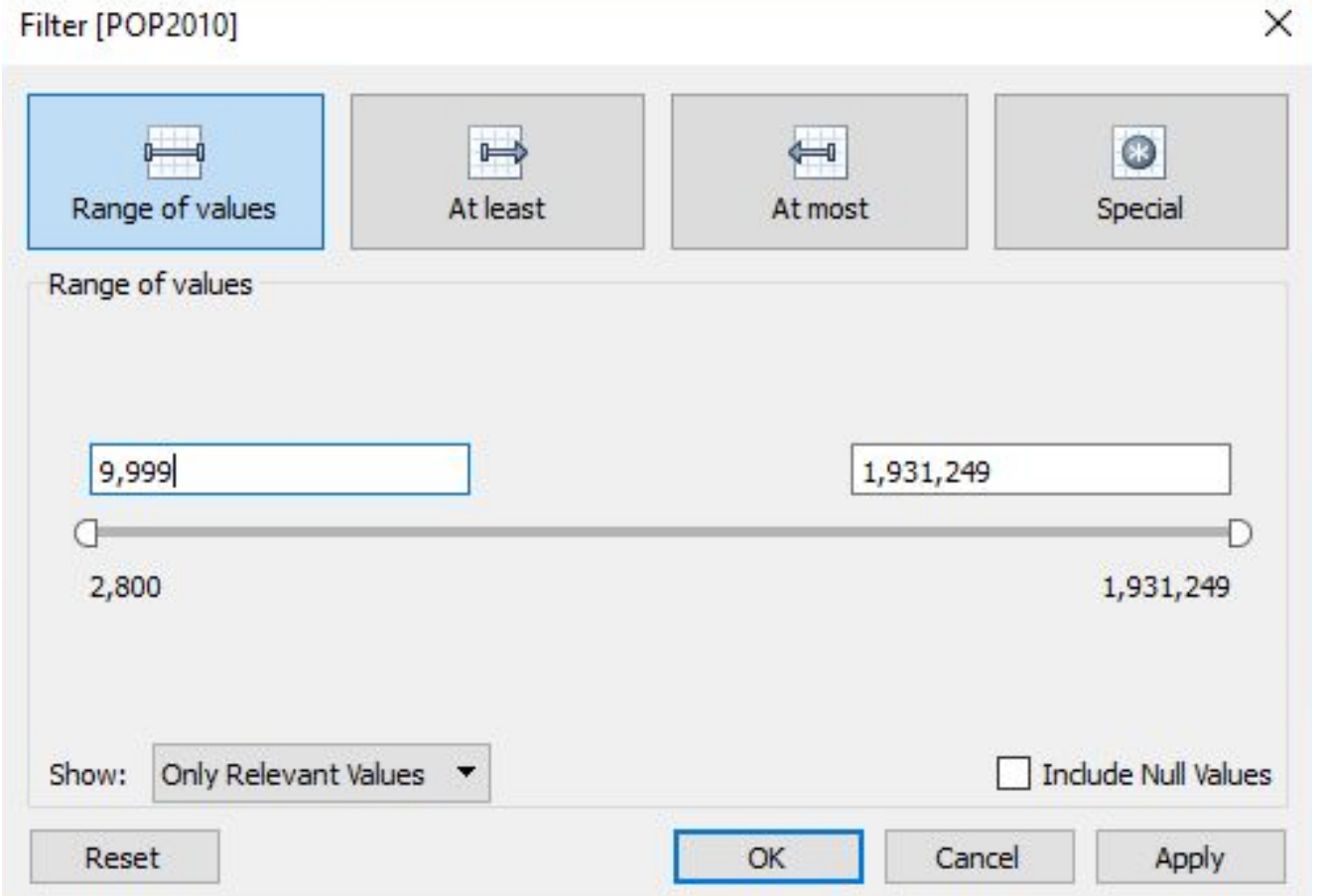


The power of joins

We used filters a little bit in an earlier module, but now we're going to do even more with filters to get a little more familiar with how to make use of them. As we just discussed, let's try filtering out those communities with smaller populations.

- Drag **Pop2010** onto the filters shelf.

When you do, a dialogue box will open. We want to set the minimum population at 9999. That way, we will filter out any towns and counties with populations under 10,000 where the rate would be skewed because the population is so small.

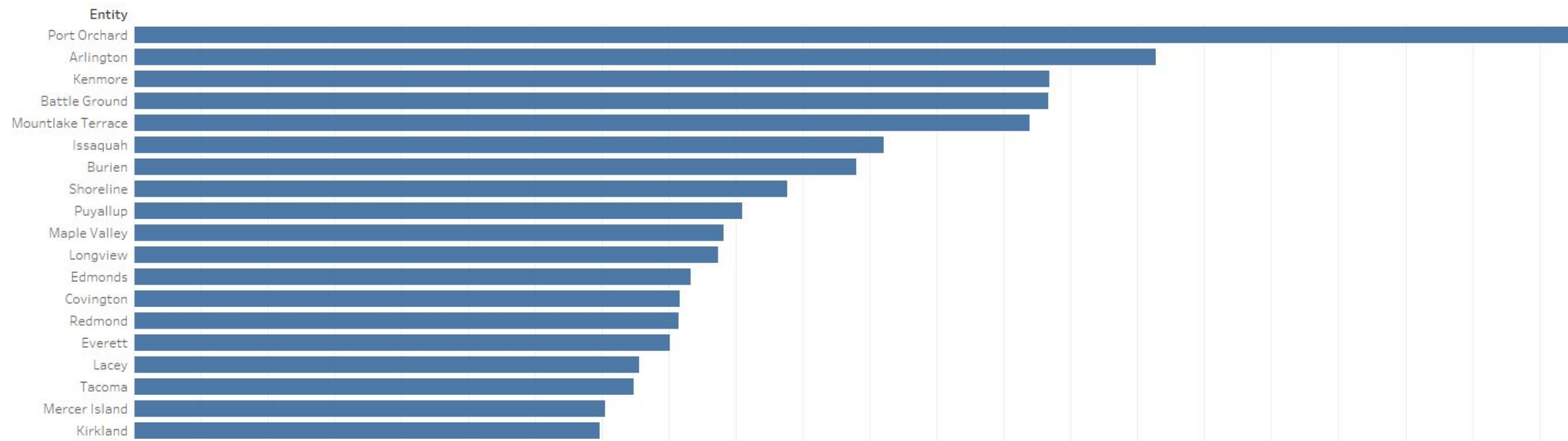


The image shows a dialog box titled "Filter [POP2010]" with a close button (X) in the top right corner. The dialog has four tabs: "Range of values" (selected), "At least", "At most", and "Special". Under the "Range of values" tab, there is a "Range of values" section with two input fields: the first contains "9,999" and the second contains "1,931,249". Below these fields is a horizontal slider with a handle in the middle, and the values "2,800" and "1,931,249" are displayed at the ends of the slider. At the bottom of the dialog, there is a "Show:" dropdown menu set to "Only Relevant Values", a checkbox for "Include Null Values" which is unchecked, and three buttons: "Reset", "OK" (highlighted with a blue border), and "Apply".

The power of joins

Finding the story: Now, we see that the city of Port Orchard has the highest spending per person on lobbyists. How else could we look at these data to find possible stories?

Sheet 1



The power of joins

Let's look at apples to apples now and filter so we can look just at cities and then just at counties.

- Drag **entity type** onto the filter shelf.

When you do that, a dialogue box will open.

- Select **OK**.

Filter [Entity Type]

General Wildcard Condition Top

☒ Select from list ☐ Custom value list ☐ Use all

Enter search text

- ☐ CITIES
- ☐ COUNTIES
- ☐ OTHER
- ☐ PORTS
- ☐ PUBLIC FACILITIES DISTRICTS
- ☐ SCHOOL DISTRICTS
- ☐ TRIBES
- ☐ UTILITY DISTRICTS

All None ☐ Exclude

Summary

Field: [Entity Type]
Selection: Selected 0 of 8 values
Wildcard: All
Condition: None
Limit: None

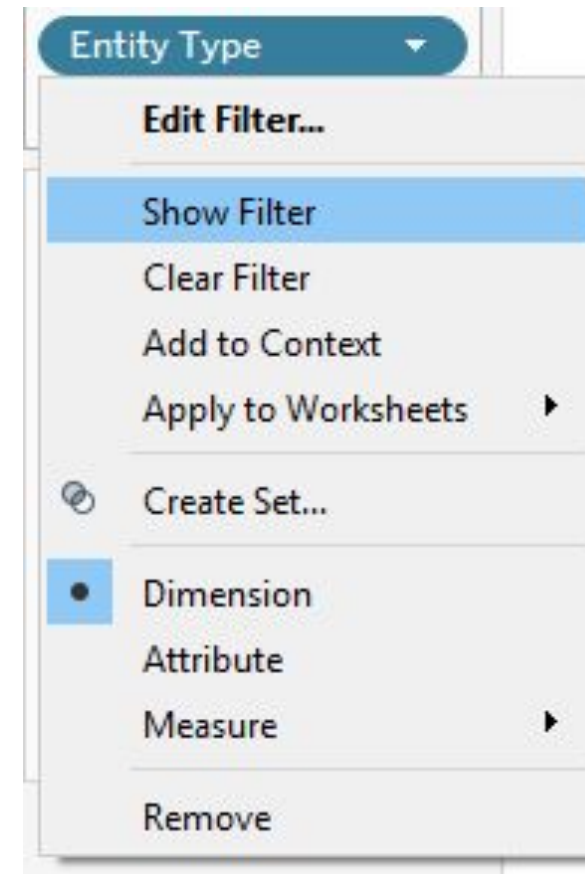
Reset OK Cancel Apply

The power of joins

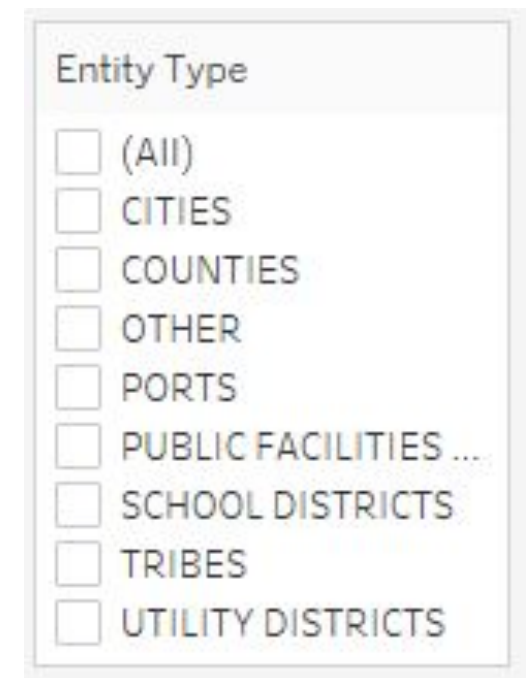
Now, we'll go back to the filters shelf and set it up so you can easily select between entity type.

- Click on the caret and then select **Show Filter**.

Once you've done that, you can easily then filter by entity type right from the worksheet.



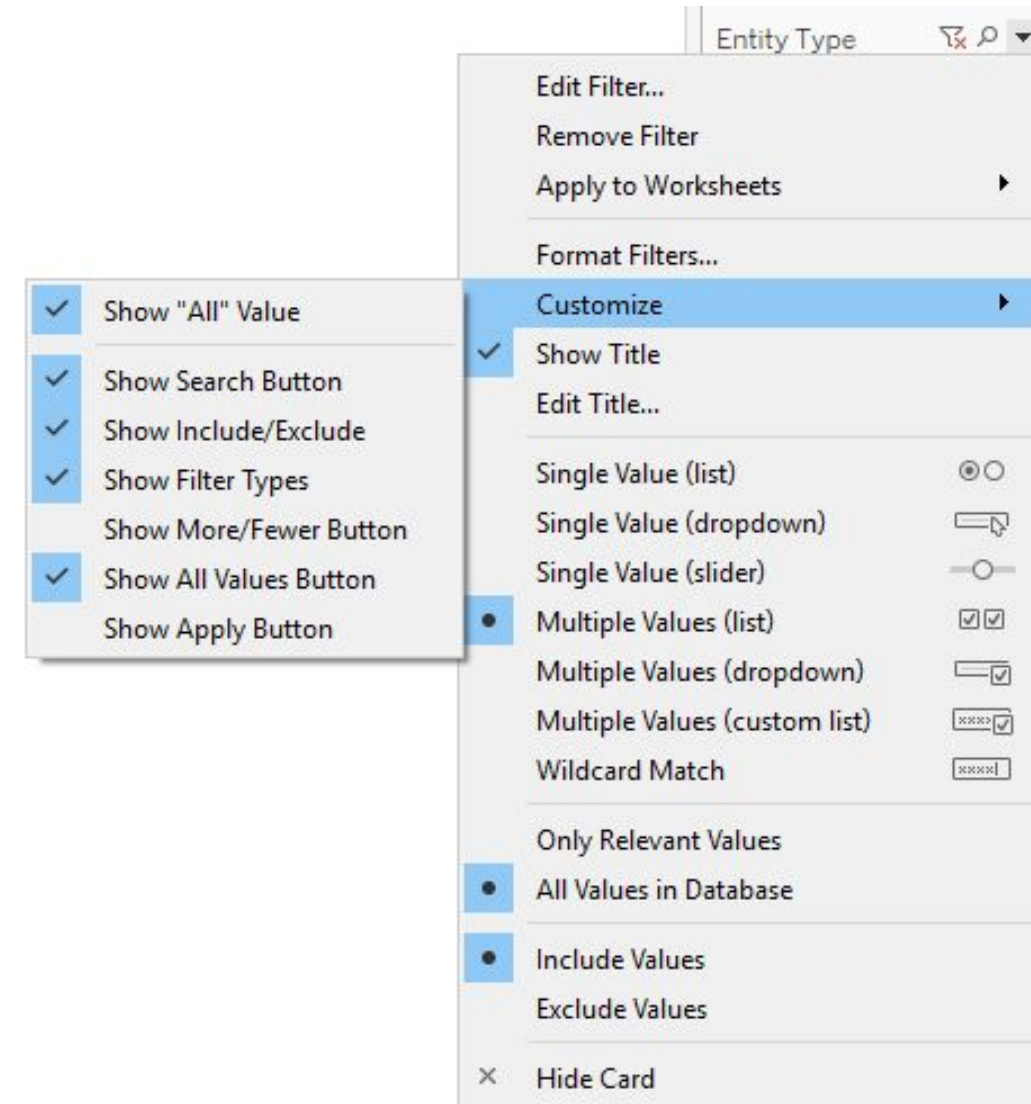
The filter will appear on the right side of the worksheet:



The power of joins

Now you can easily adjust the way you make our selections in the filter by editing the filter.

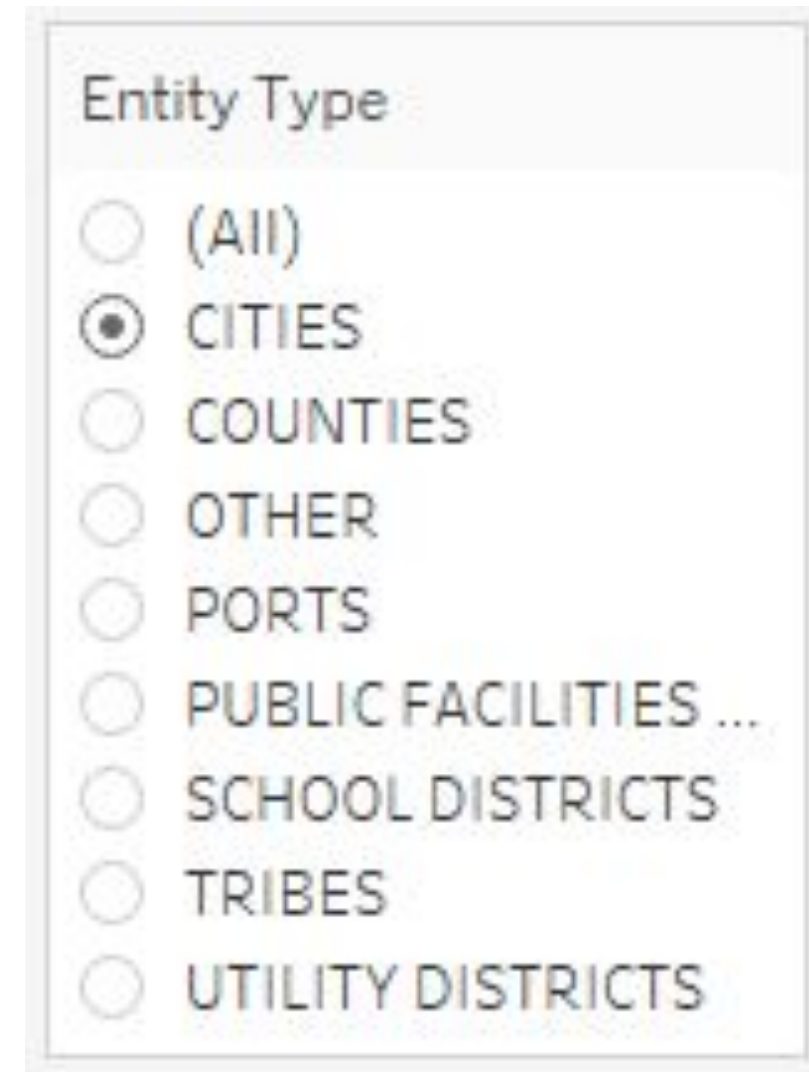
- Select the caret at the top right corner of the filter and a dialogue box will open.
- From there, you can select the way the filter appears and also customize it.



The power of filters

- Select **Single Value List**.

Now we can easily look at like categories.



Entity Type

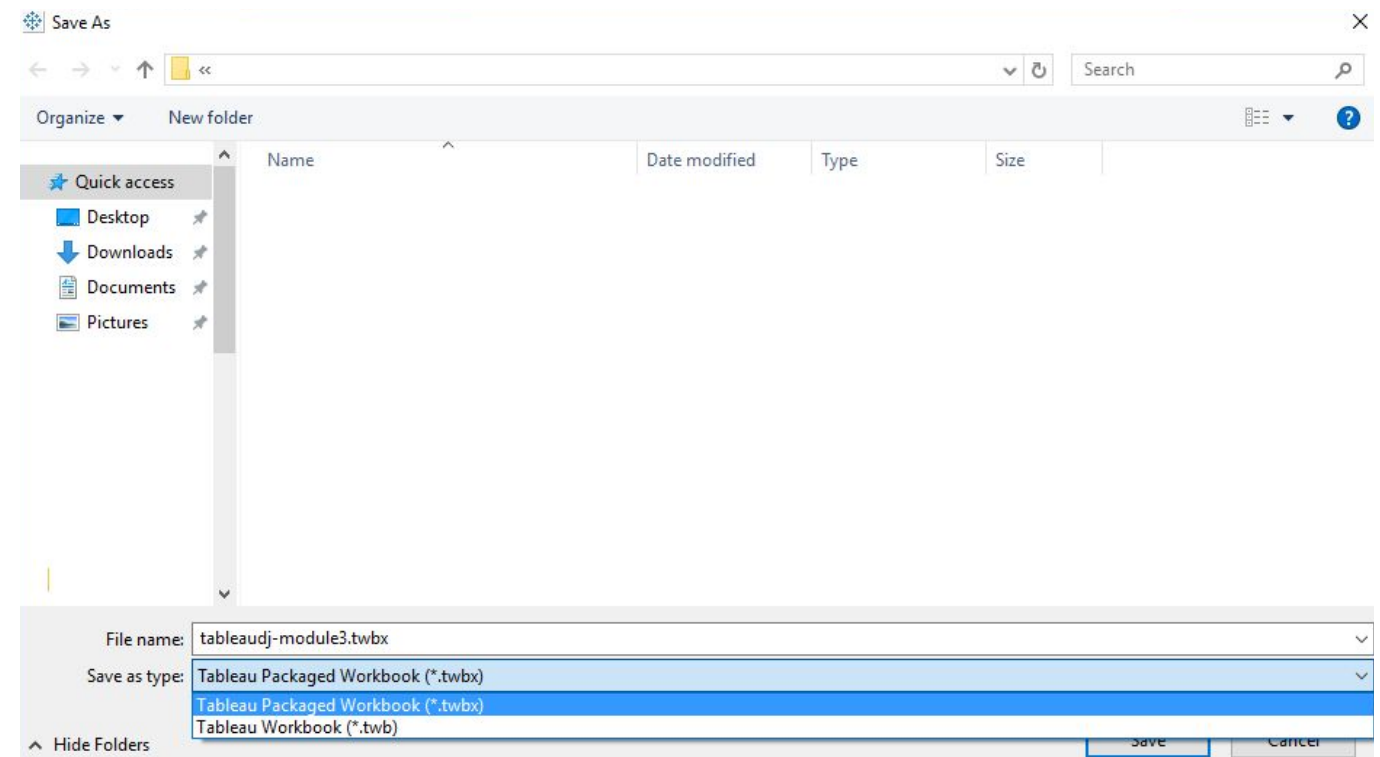
- ☐ (All)
- ☒ CITIES
- ☐ COUNTIES
- ☐ OTHER
- ☐ PORTS
- ☐ PUBLIC FACILITIES ...
- ☐ SCHOOL DISTRICTS
- ☐ TRIBES
- ☐ UTILITY DISTRICTS

The power of joins

In the next module, we'll begin to explore data geographically and pull all of our work together into one Tableau Dashboard. For now, though, let's save our work.

- Select **File/Save As**
- Then name your workbook with your lastname_lobbying

We will also save this workbook as a packaged workbook. I like to do this as a matter of course because Tableau makes sure you have all the data you need packaged up within the workbook itself.



SDPD Traffic Stop Data

Join the 2016 vehicle stop data from last week with

- vehicle_stops_race_codes.csv
- vehicle_stops_search_details_2016.csv

Explore the new questions you can answer!

