Announcements

- Readings:
  - In https://github.com/gquer/dsc-96_winter19/blob/master/03_joining_mapping/readings.md
  - By Wednesday Jan. 23 at 6.00pm (**tomorrow**!)
  - To gquer@ucsd.edu
  - **Subject: [DSC 96 W03 SecA|C Journal]: Name LastName**

DSC 96

What Happened?

Joining Data

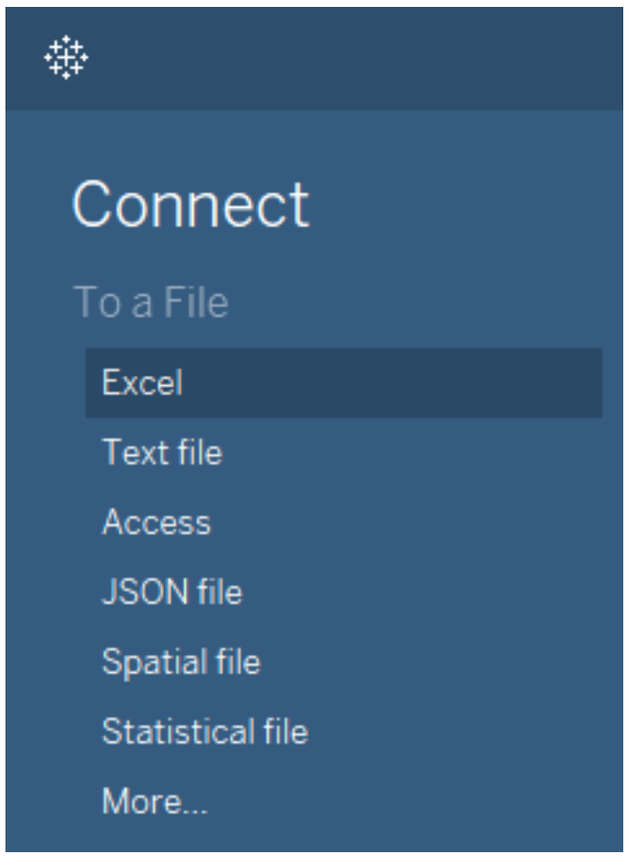
Tableau basics

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

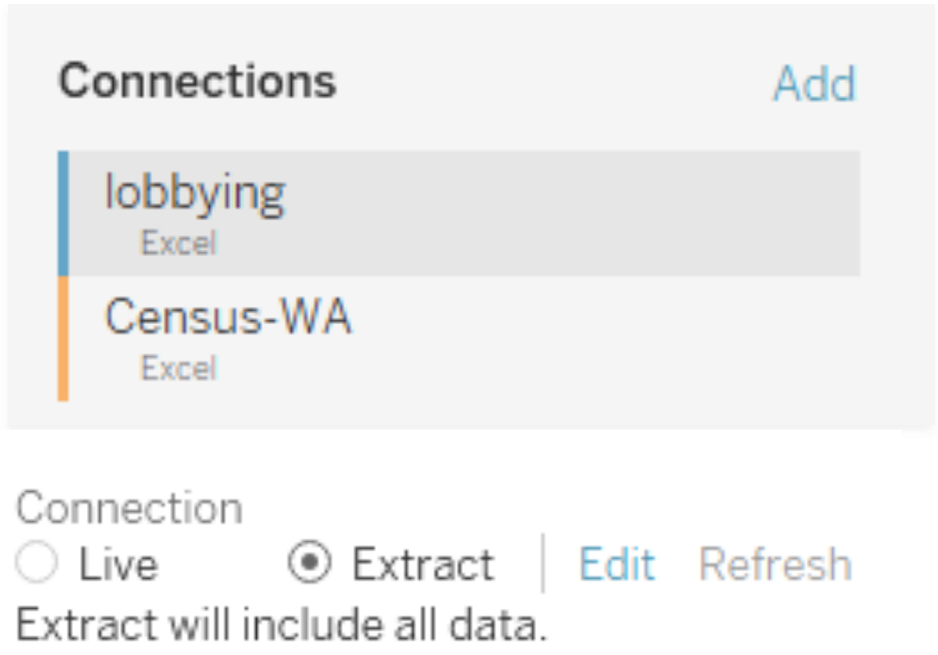
Getting started

- Lobbyists dataset
  - city/county population?

- new dataset!
- there is a column for a FIPS (Federal Information Processing Standard) code
  - every jurisdiction has a FIPS code
  - use that to join in the population data to the cities and counties



The power of joins



- Open a new workbook
- Import to lobbyingdata.xlsx
- add censuswa.xlsx file
  - click **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.

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
Extract will include all data.

Filters
0 Add

lobbying

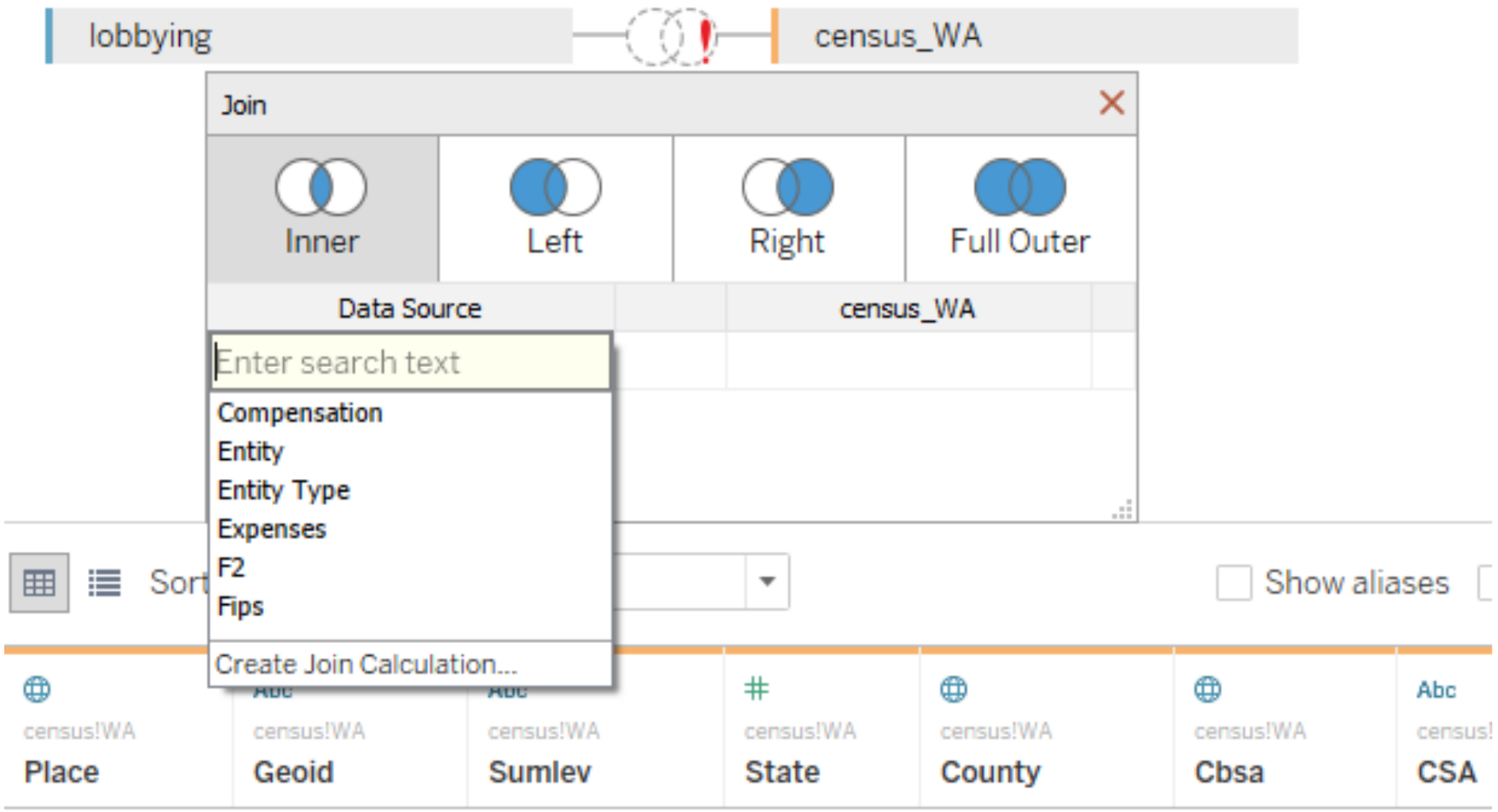
Sort fields Data source order

Show aliases Show hidden fields 131 rows

Abc lobbying Fips	Abc lobbying Entity	# lobbying Compensation	# lobbying Expenses	Abc 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

- how to join this data?
- several options for different types of joins



The power of joins

- **Example**
  - **DS1:** students, driving license ID for each student, student marks
  - **DS2:** soccer players, driving license ID for each player, soccer statistics

- **Inner Join.** The records where the IDs match in both data sets. *Example: Only the records where the ID of the student matches with a soccer player*
- **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 students and only records from soccer players 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 soccer players and only records from the students when they match.*
- **Outer Join.** You add all the records from each data set together, even when there is no join. *(all information)*

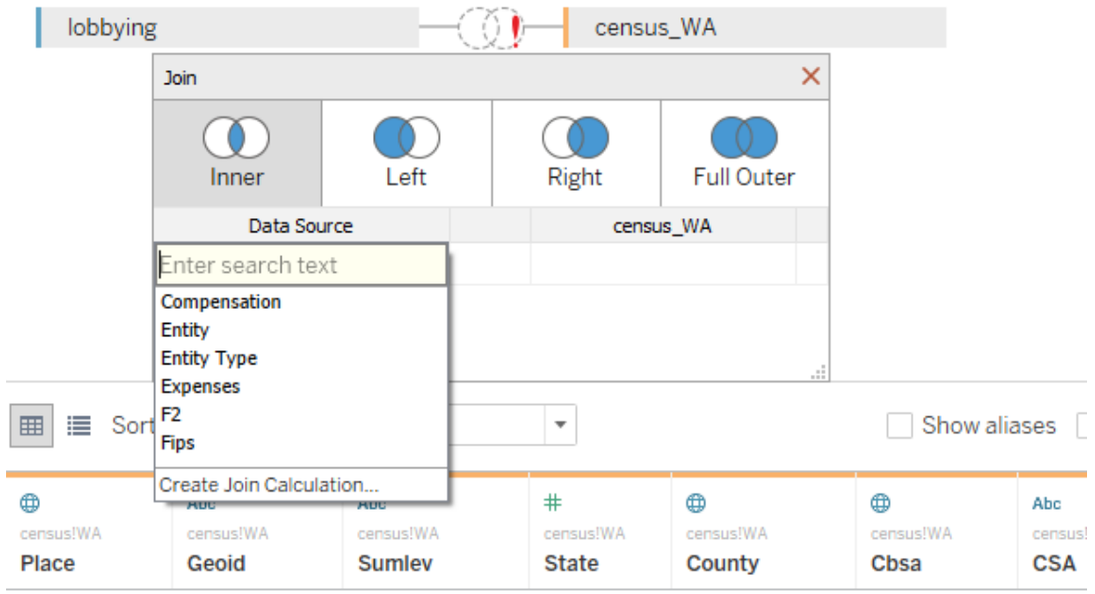
The power of joins

Attention! There may be issues!

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.



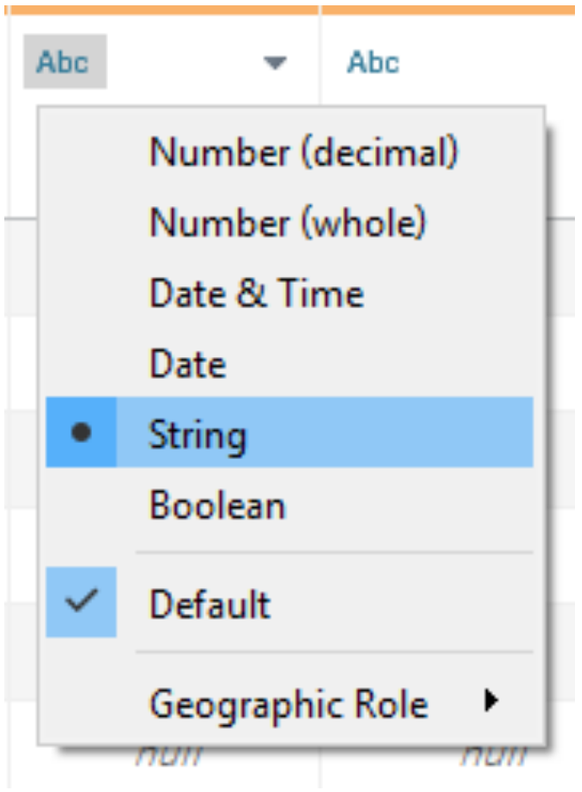
The power of joins

1. Click on the field for the ID
    1. make sure **string** is selected. The field we are connecting the two datasets on are FIPS in sheet1 and GEOID in census_WA
    2. Both should be formatted as string fields

2. Once you format the fields, the exclamation point disappears

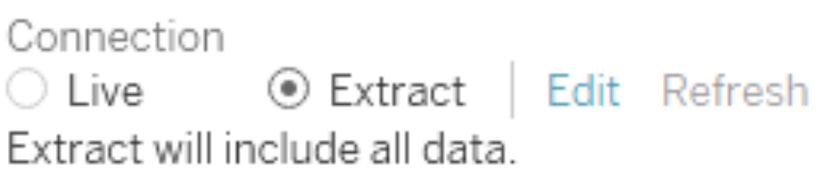
3. You should also select extract instead of live – that will improve your performance within Tableau.

1.

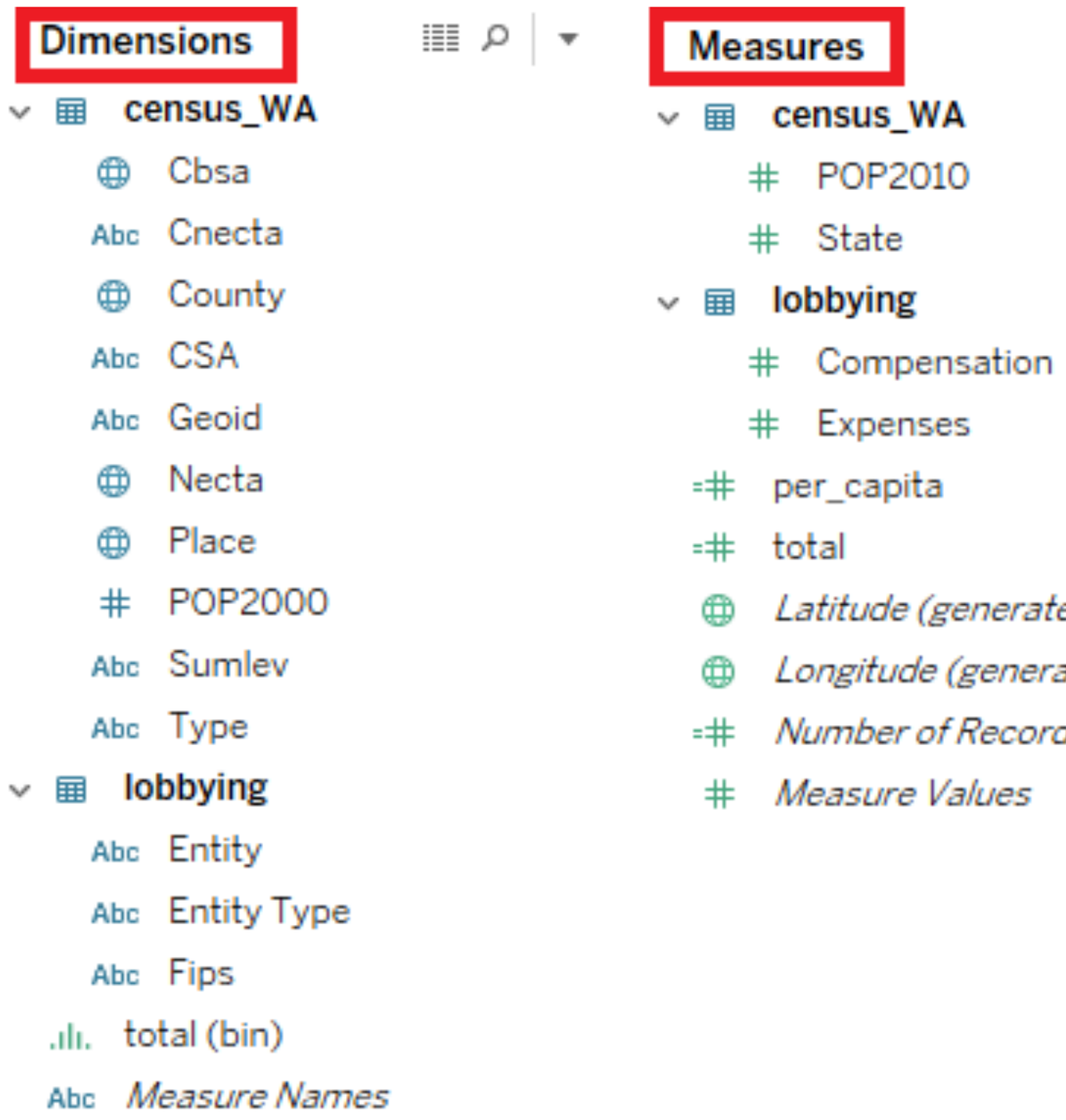


2.

3.

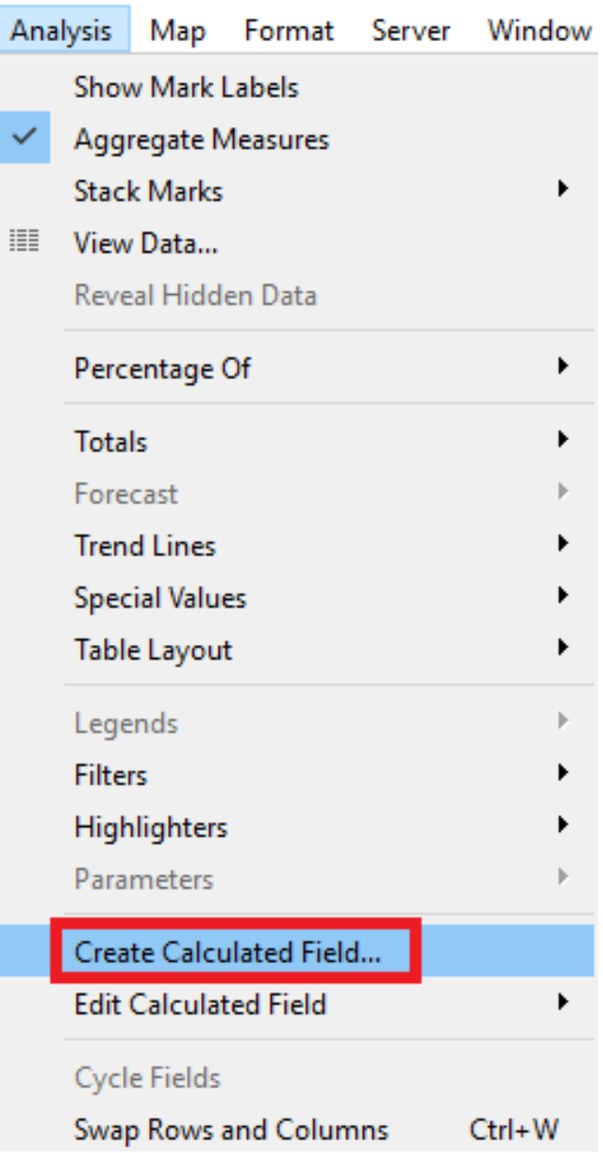


The power of joins: visual explanation



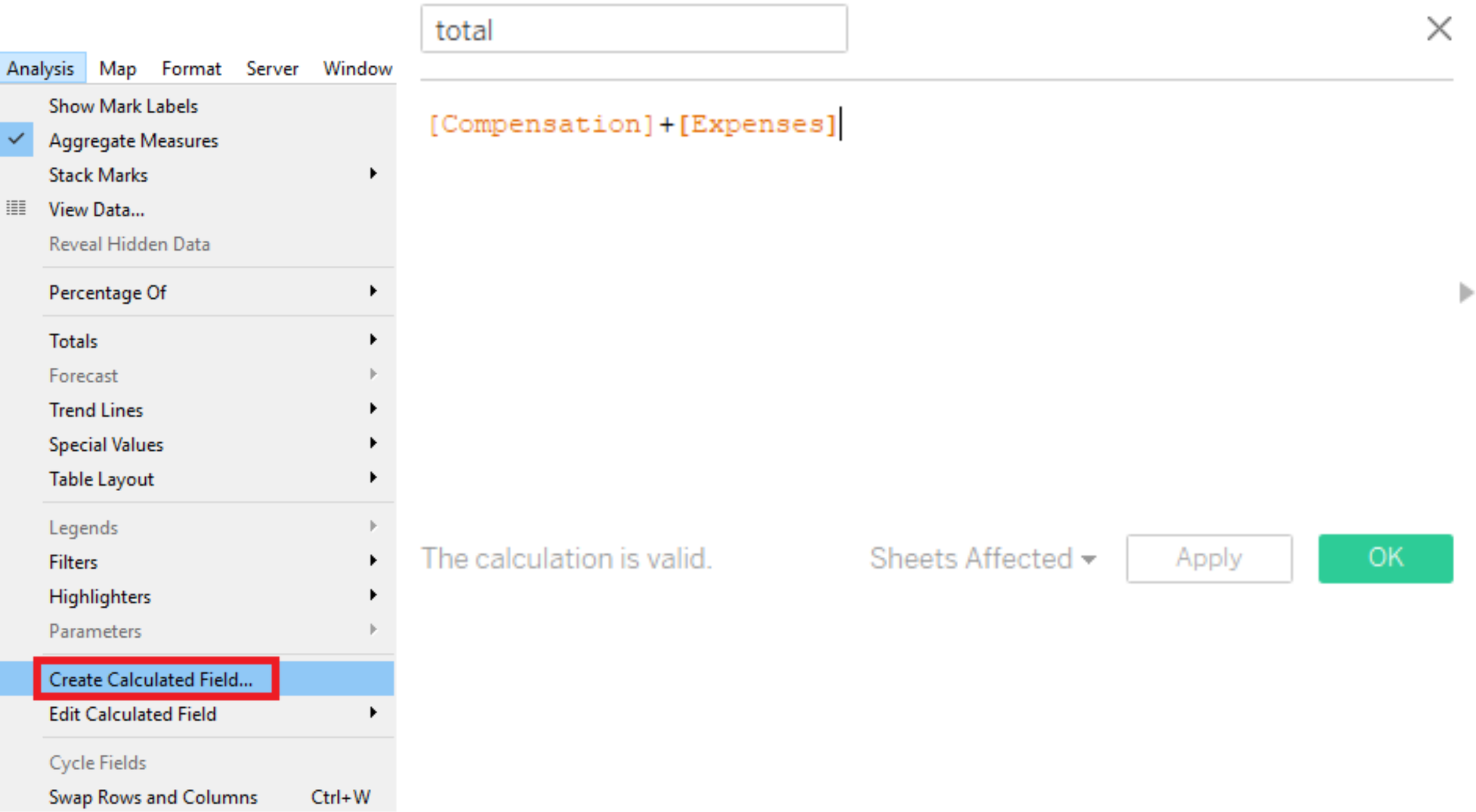
The power of joins

- Create a calculated field
- Total = Compensation + Expenses
- Per capita = Total/Pop2010



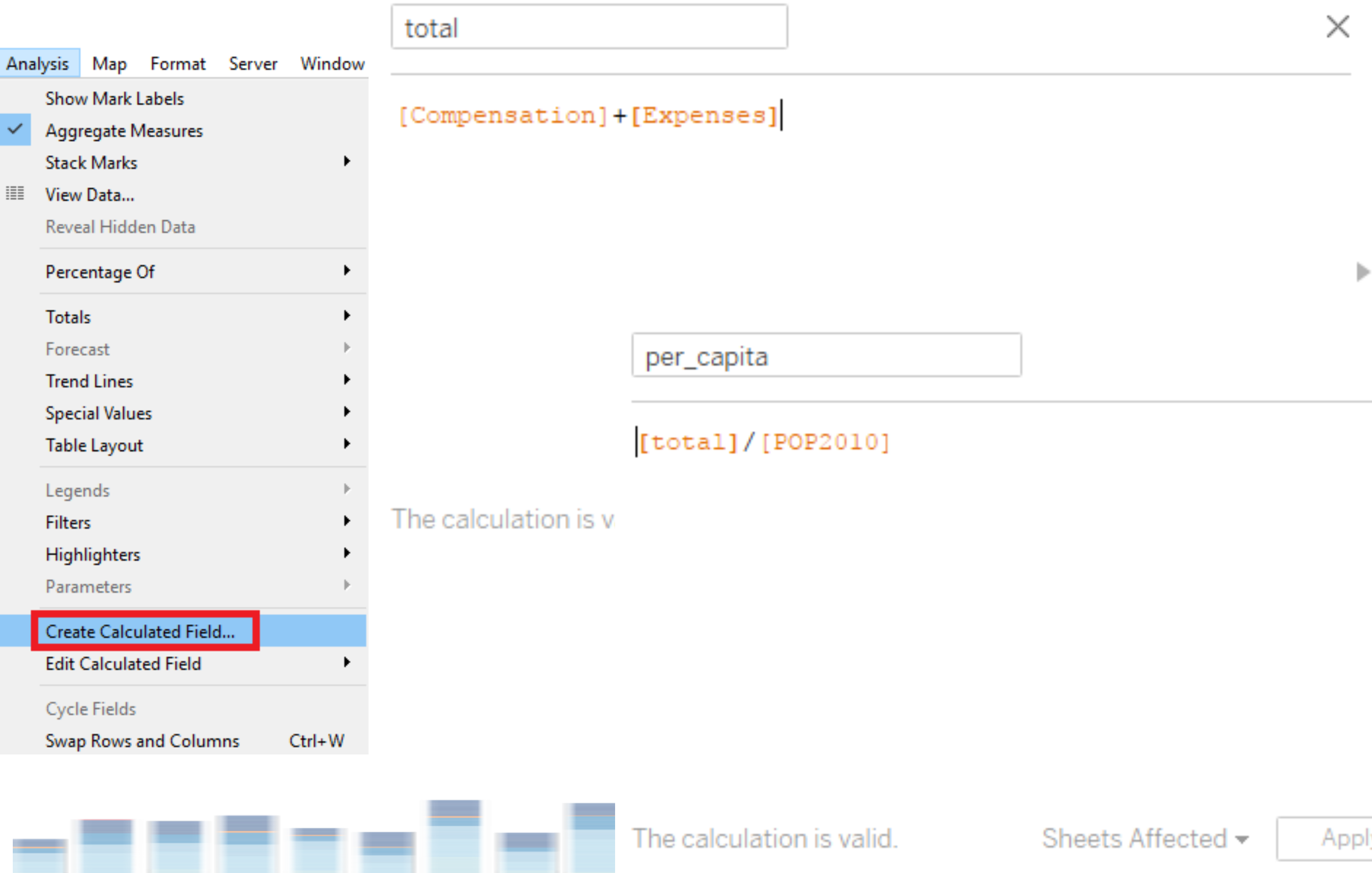
The power of joins

- Create a calculated field
- Total = Compensation + Expenses
- Per capita = Total/Pop2010



The power of joins

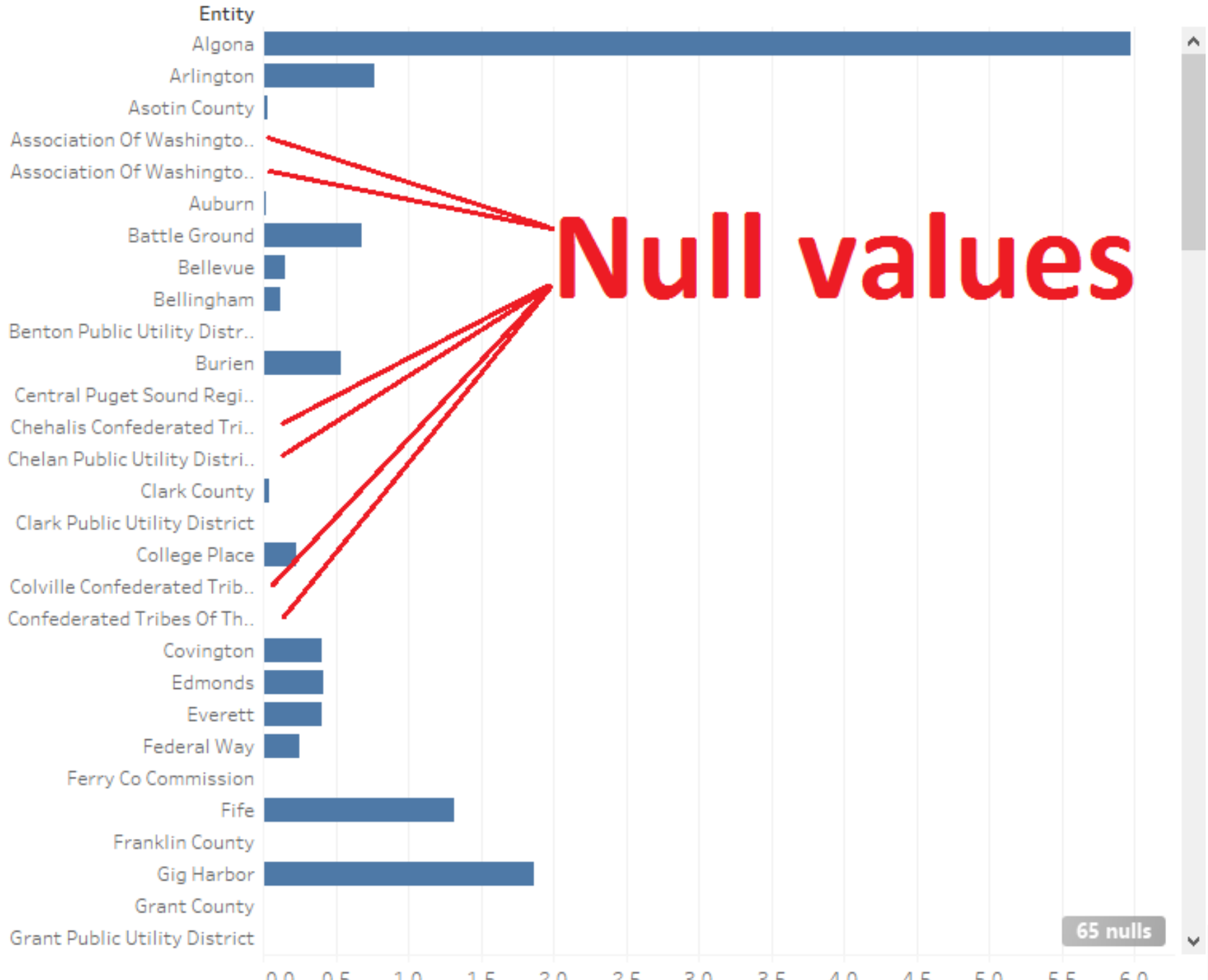
- Create a calculated field
- Total = Compensation + Expenses
- Per capita = Total/Pop2010



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.



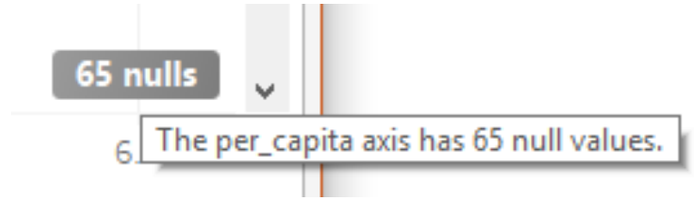
Eliminate the Null values

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.

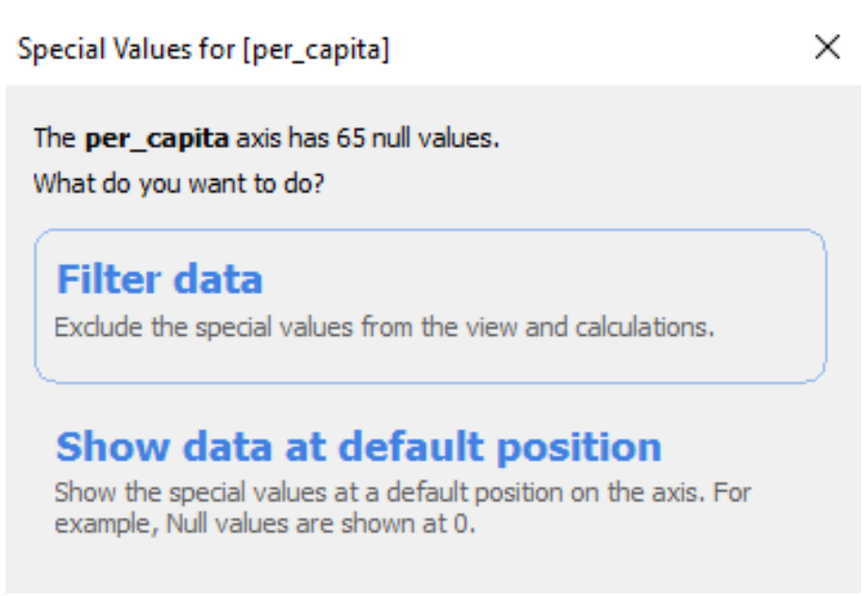
1. Click on the grey pill that says **65 nulls**.

2. A dialogue box will pop up and you can filter out the null values.

1.

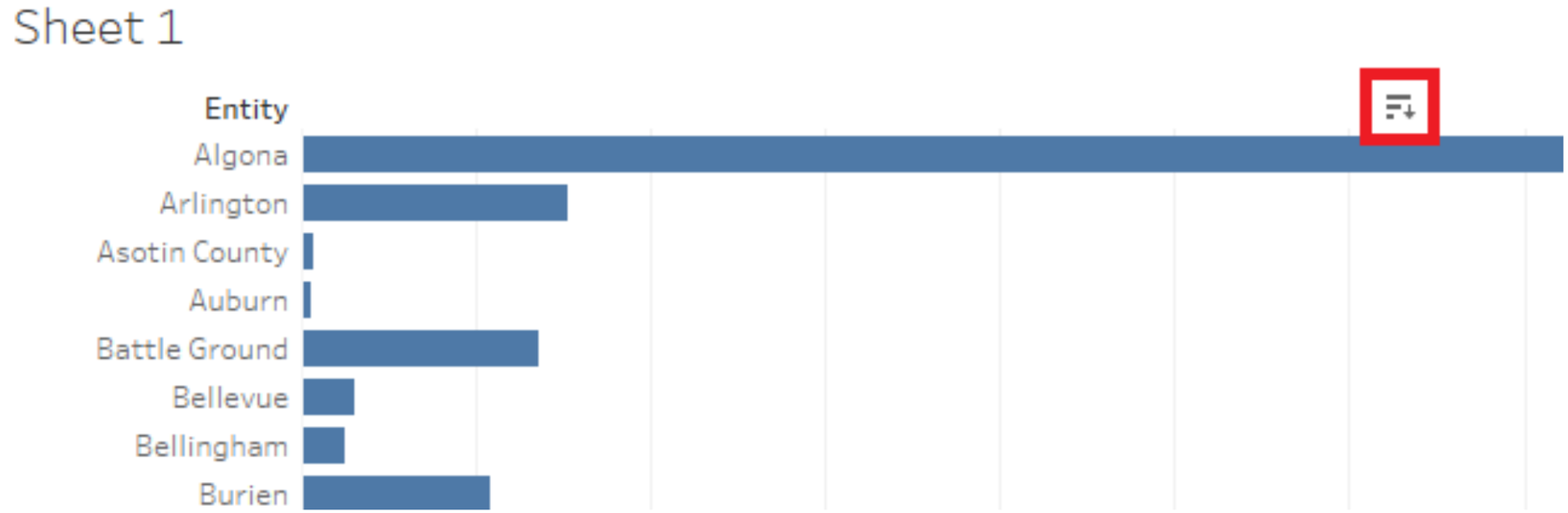


2.



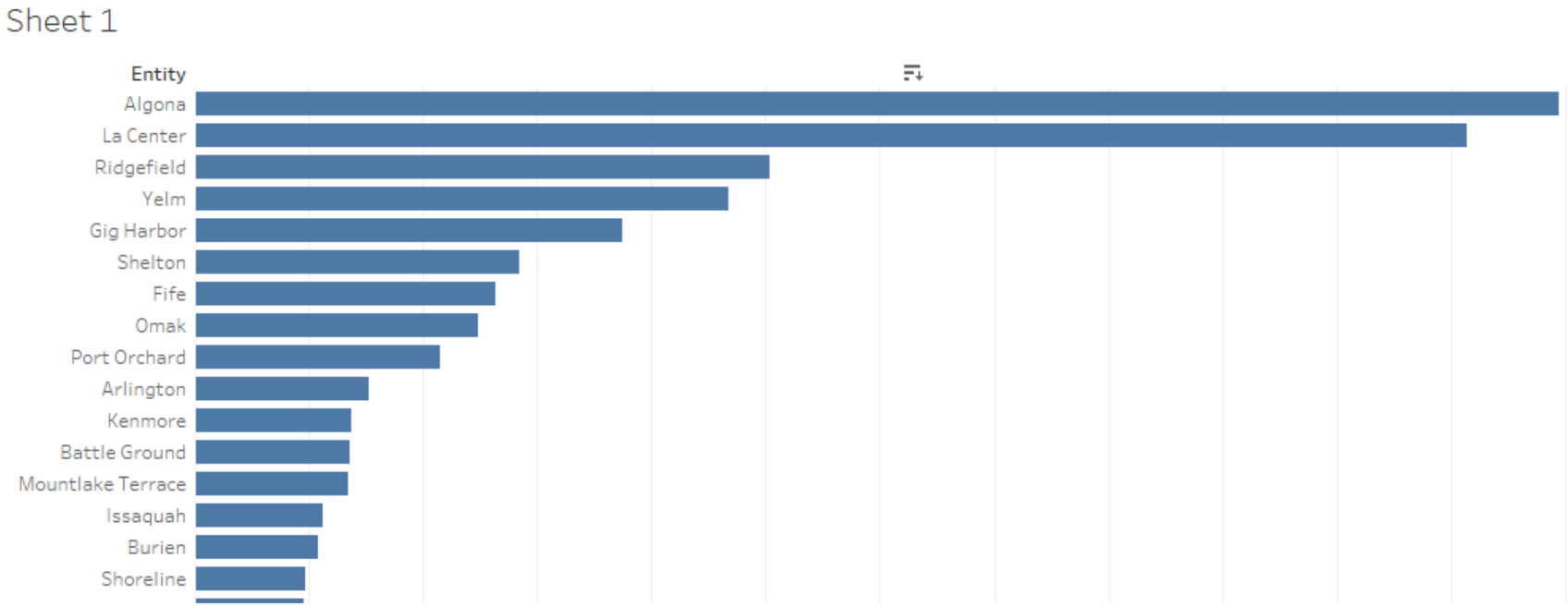
The power of joins

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



The power of joins

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

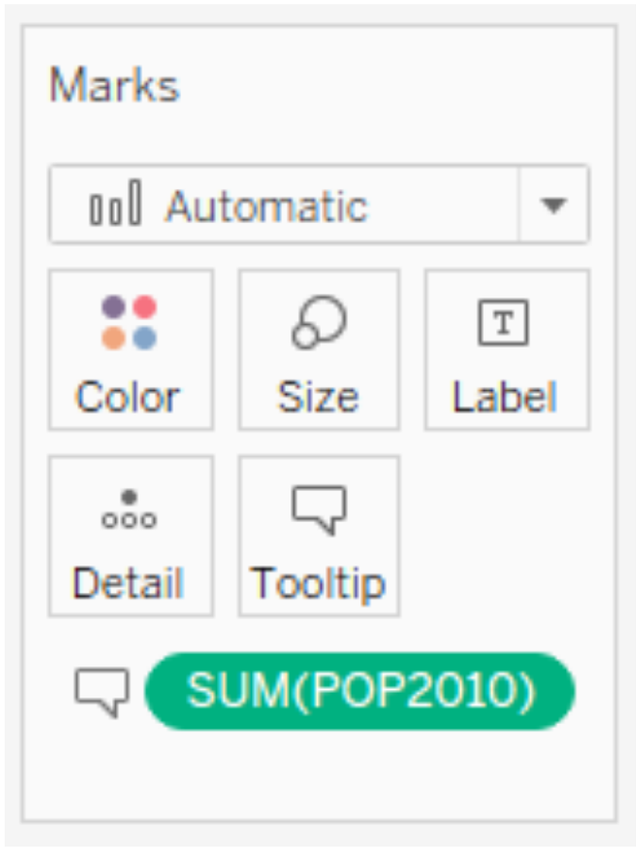


The power of joins

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.

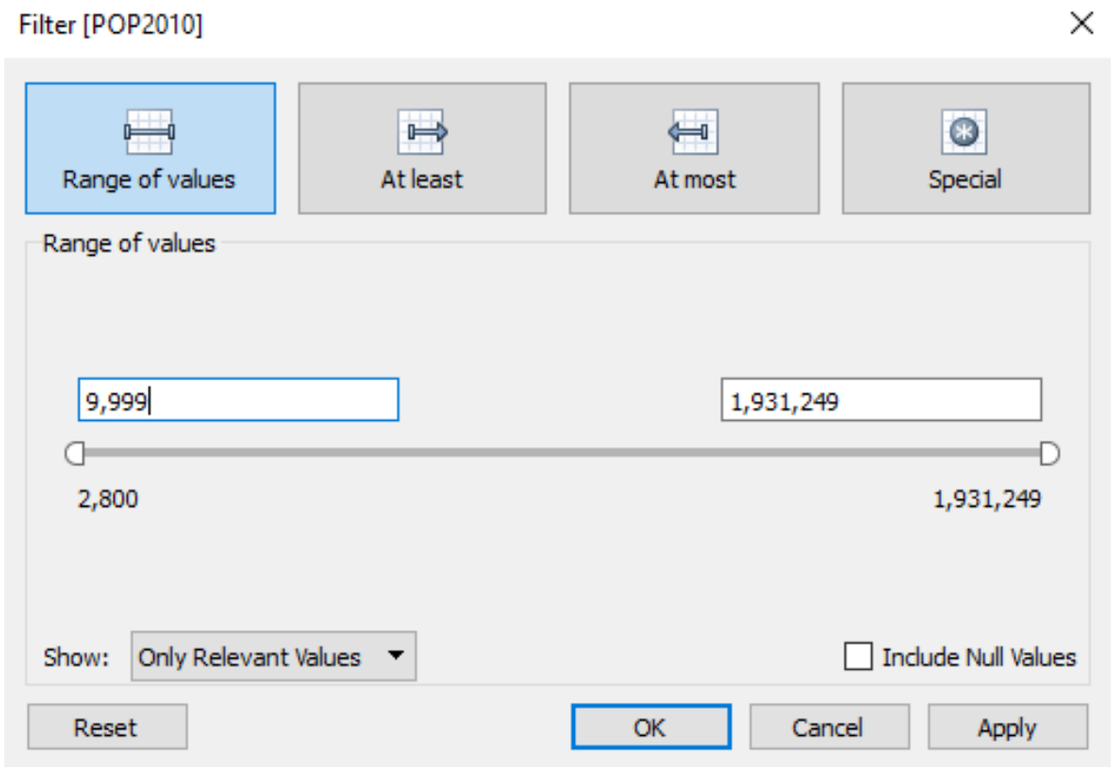
That field will show up below the icons with the visual cue to the left that it’s associated with the tooltip. Now, when you hover over the bar chart, you can get a sense of the population of the community as well as the rate.



The power of joins

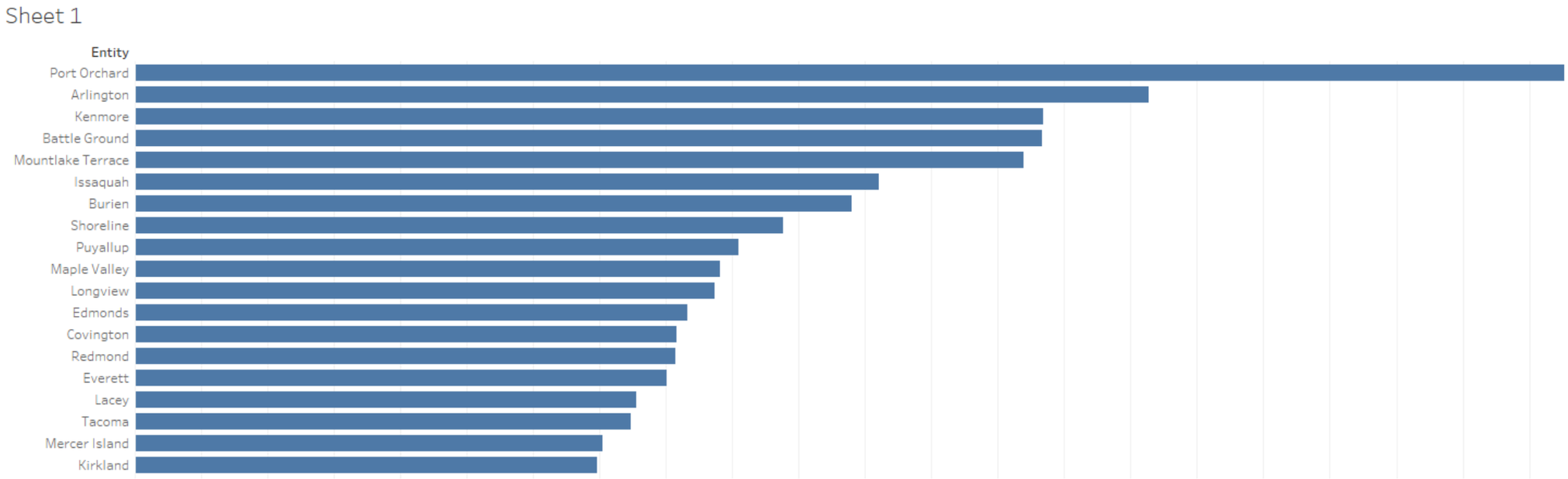
More on Filters:
let's try filtering out those communities with smaller populations.

- Drag Pop2010 onto the filters shelf
- 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 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?



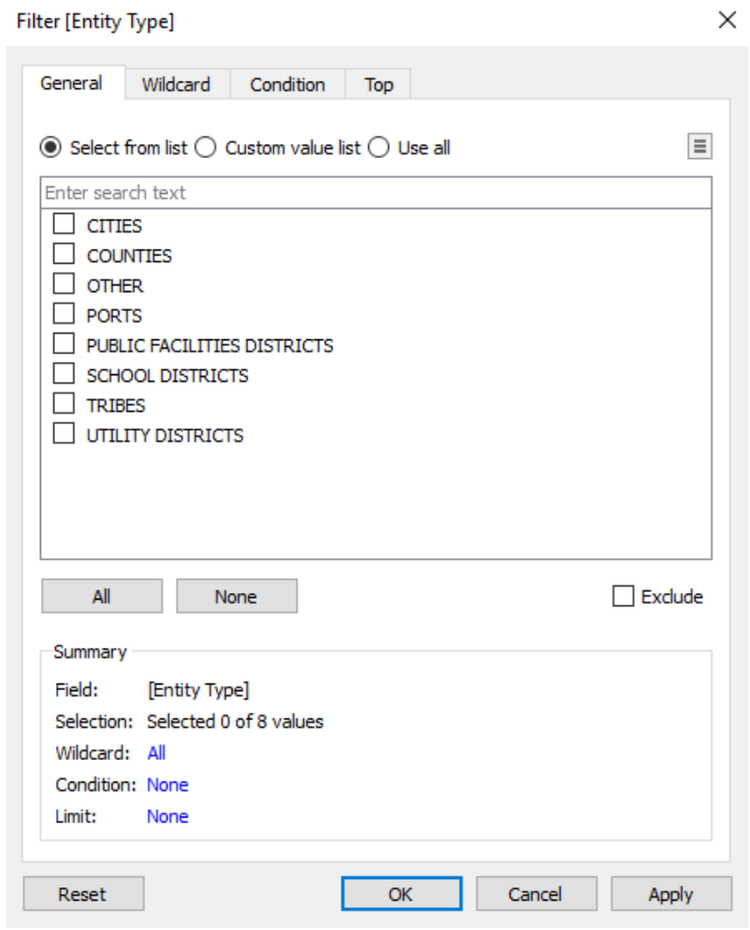
The power of joins

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

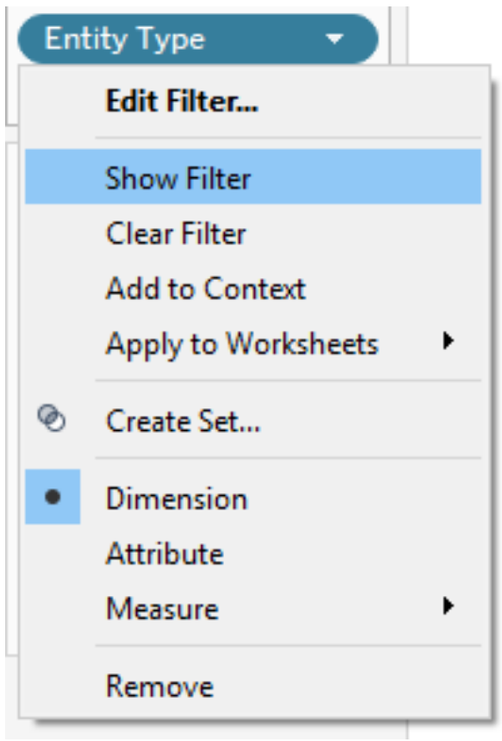


The power of joins

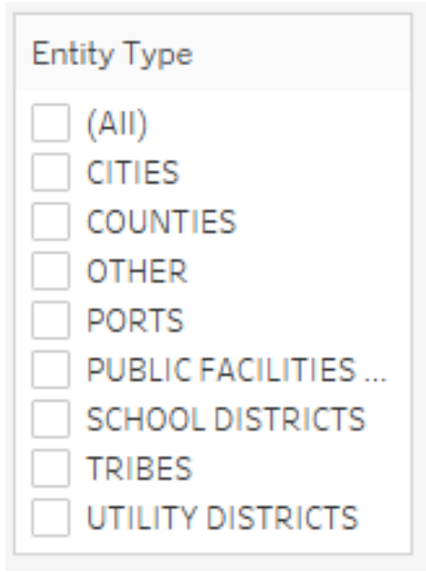
Now, we'll go back to the filters shelf and set it up so you can easily select between entity type.

- 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**.



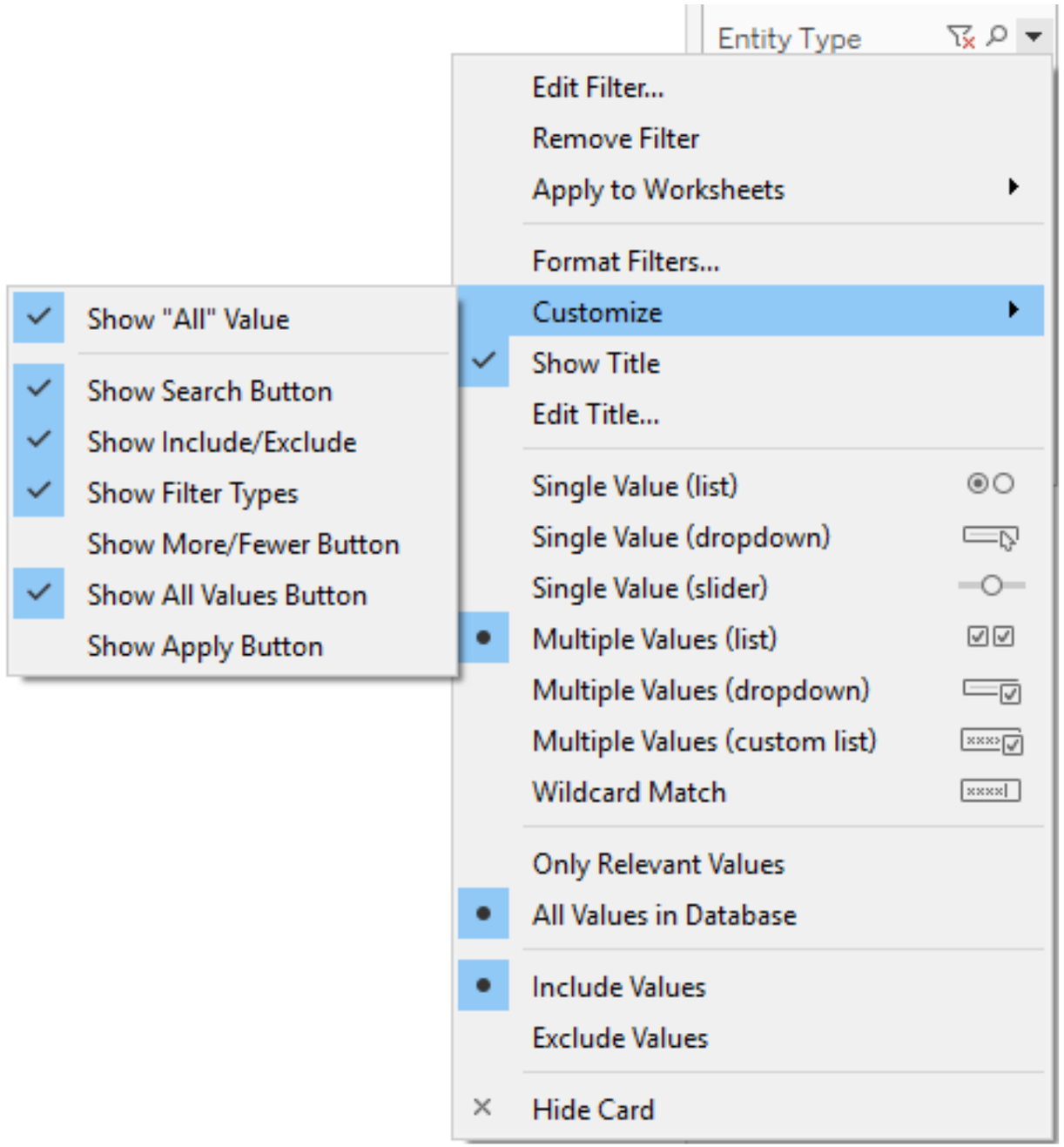
The filter will appear on the right side of the worksheet:



Customize the Filter

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.



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!

ASSIGNMENT:

- Due: Wednesday 1/30 at 6 PM
  - In SectionA/02_SD_vehicles
  - 1 workbook with your Name_LastName
  - As many worksheets as you want
    - Give a title to each of them
- 1 dashboard with the story you want to tell
  - race/age
  - timeseries
  - all three dataset
  - Mapping