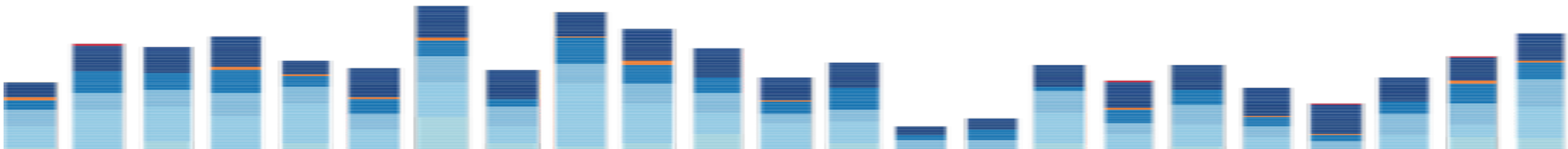# DSC 96

What Happened?

Answers on Day 1

Giorgio Quer (from Colin Jemmott)

Yesterday
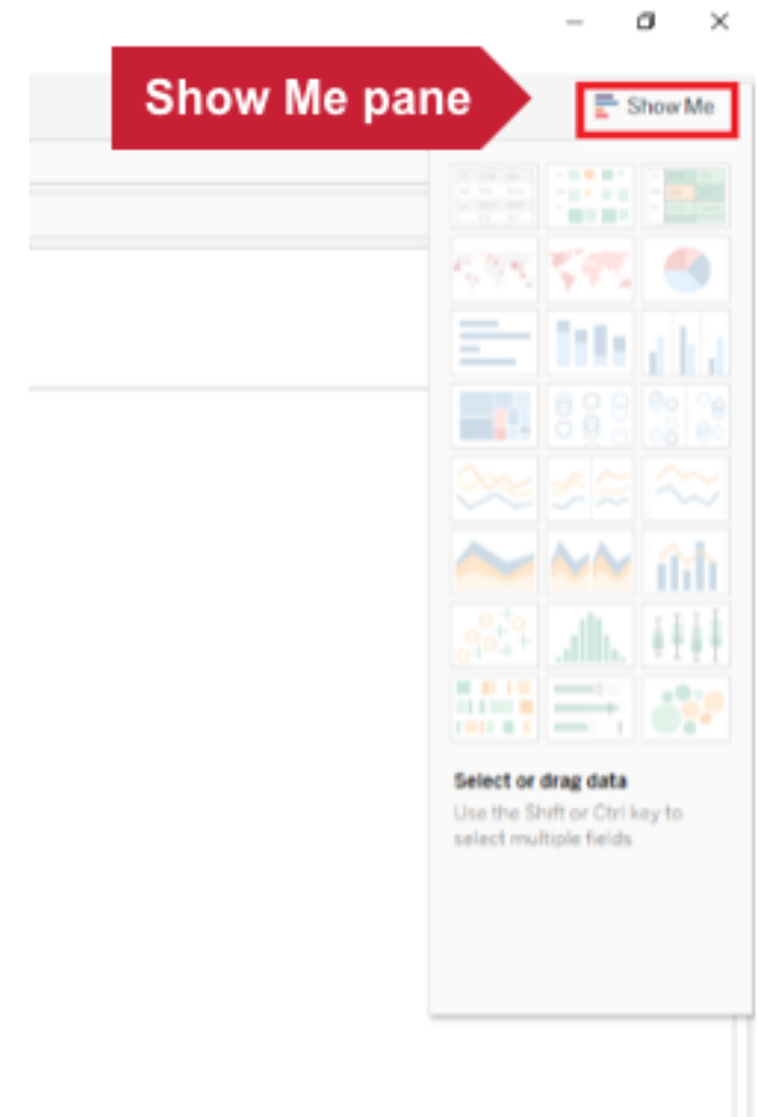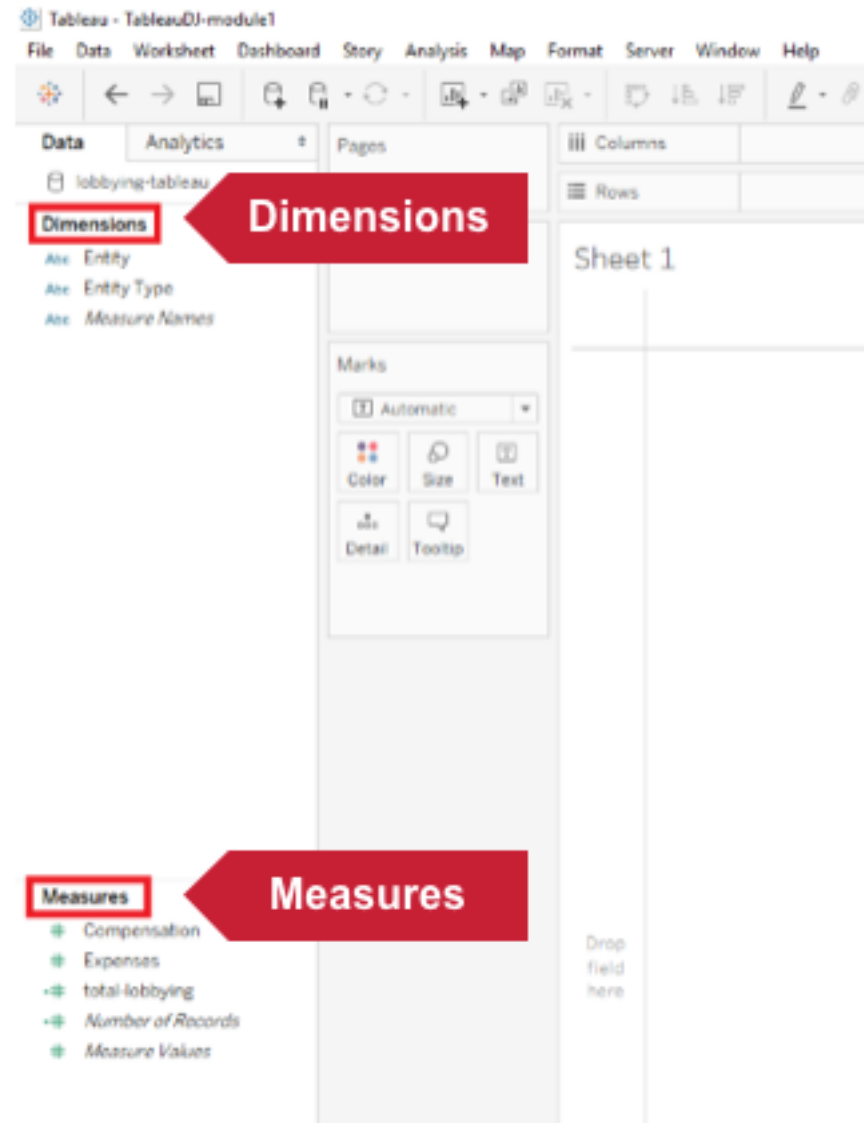
- Dimensions
- Measures
- Show me
- Analysis

Tableau journalism basics

- Review of importing and data analysis

- Aggregation

- Filters

- Do we have enough to understand the story?

Total of a variable

○ Drag the total-lobbying calculation from last time onto **row** and select the **tabular feature** on the **Show Me** pane to get the number: $2,523,090.

That's the total spent by local governments for lobbyists.

Sheet 5

2,523,090

Marks

T Automatic ▼

Color   Size   Text

Detail   Tooltip

T SUM(total)
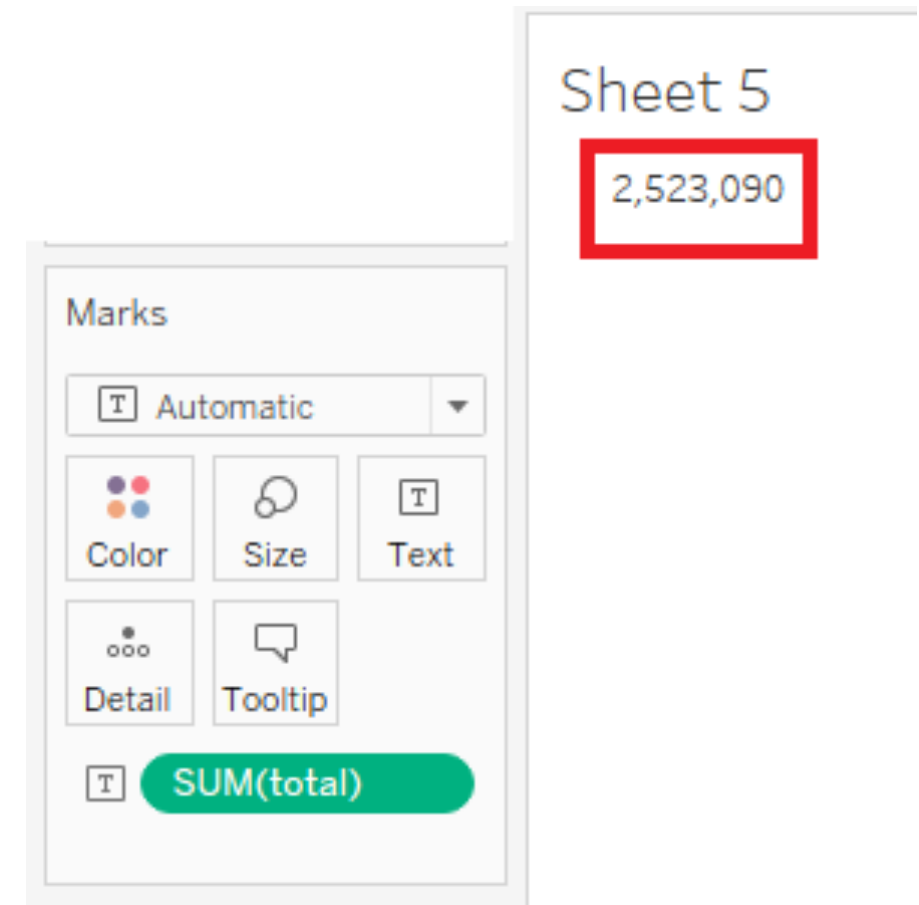
Total of a variable

With Tableau, as in many software programs, there are multiple ways to do the same thing.

○ Select the bar chart option on the **Show Me** feature
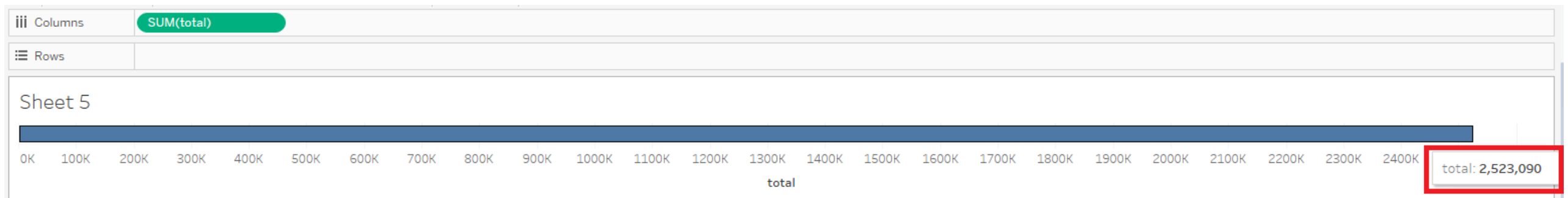
Sum of two variables

Here's a reminder:

○ On the menu, select **Analysis** and then **Calculated Field**.

○ Give your new calculated field a name: total. Then type the calculation you want to do.

○ Add **Compensation** and **Expenses:**

   [Compensations] + [Expenses]
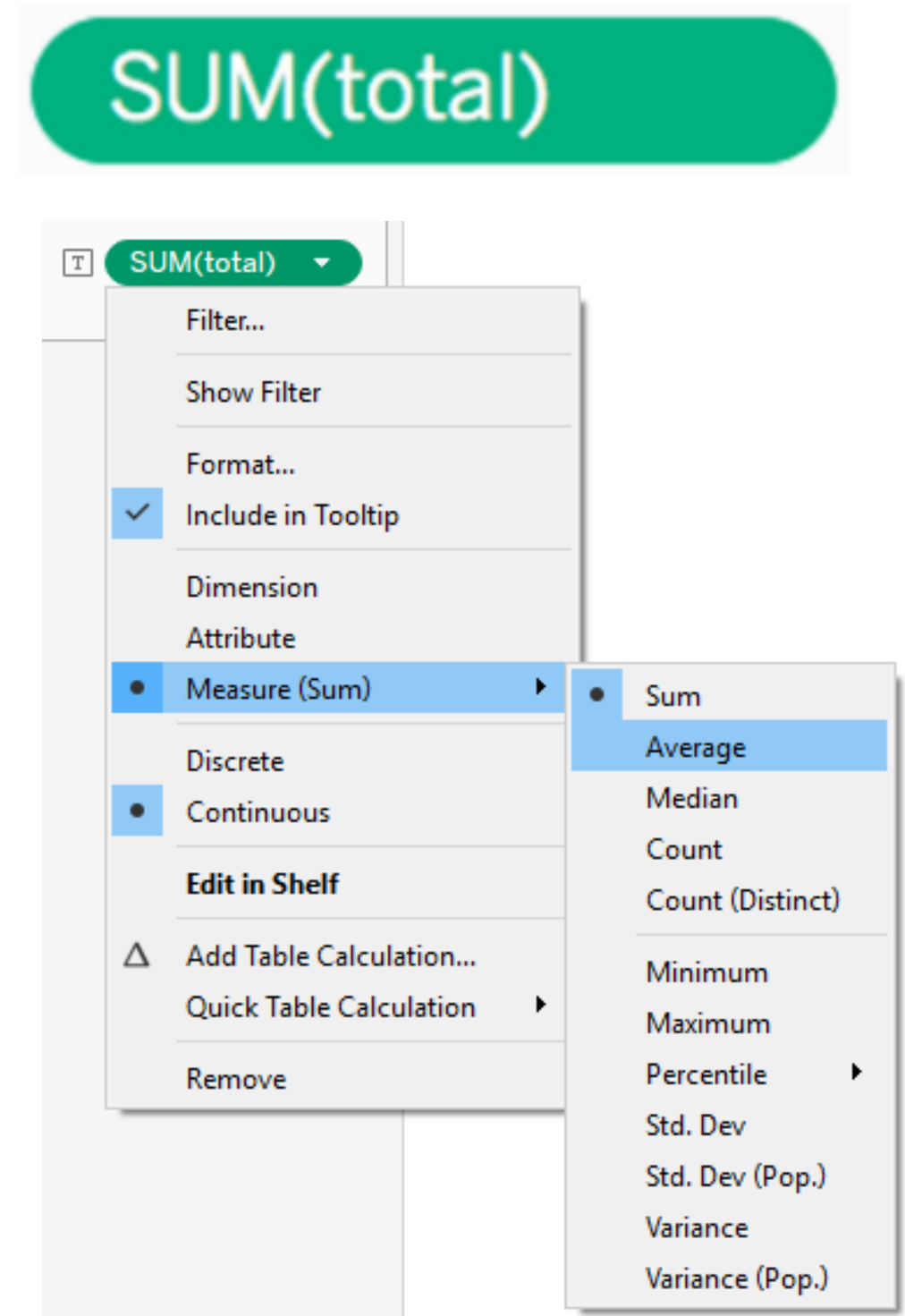
Aggregation: **sum, median, average**

Tableau automatically aggregates your data by the **sum**.
If you look at the green pill on the marks shelf, you'll see the word SUM in front of the total field: SUM(total)

Other aggregation methods?

How could we see average? How about the median? It's as easy as selecting the caret, or down arrow or triangle on the right side of the pill.

Can we compare average and median?

What if we want to get a sense of the median compared to the average? And why would that matter anyway?

To give you a sense of how you might use this difference between the average and the median, let's examine the type of governmental body – that's the entity type field – that is paying lobbyists and look at the gap between the median and the average.
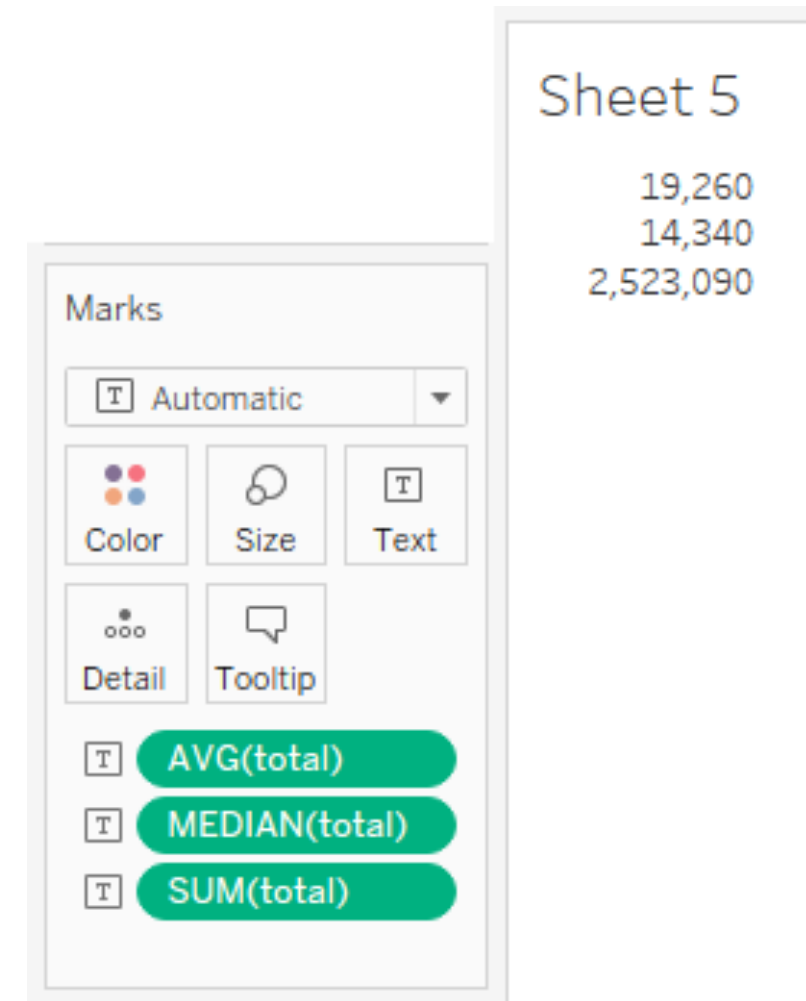
Average, median, sum

In addition to looking at the median and average, we can also look at the total.

- ○ Drag total onto the marks shelf three different times. Select three different aggregation levels: **SUM(total)**, **AVG(total)** and **MEDIAN(total)**.

Make sure that the mark just to the left of each pill is selected to be **Text (T)**. Then the actual numbers will show up on the workbook.

Sheet 5

19,260
14,340
2,523,090

Marks

⊤ Automatic ▼

Color   Size   Text

Detail   Tooltip

⊤ AVG(total)
⊤ MEDIAN(total)
⊤ SUM(total)

Average, median, sum: visualization

Of course, in Tableau, you can also look at your information visually. How would we do that?

Now, let's just look at the median compared to the average.

○ Drag off the SUM(TOTAL) pill.

You can imagine if we left total on there, it would show us one long bar for total and two very tiny bars for median and average. That wouldn't be all that useful.

○ Select the Bar Chart option on the Show Me Pane. You'll get something like this:
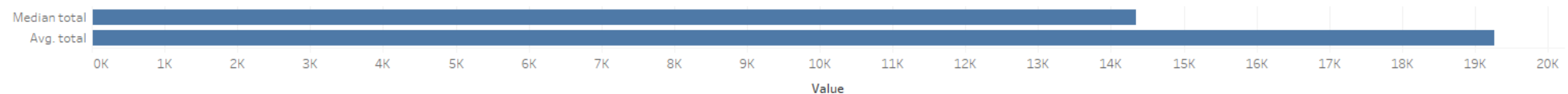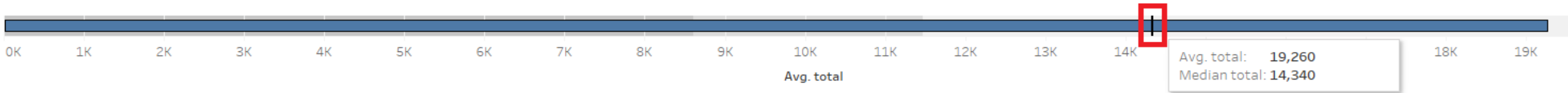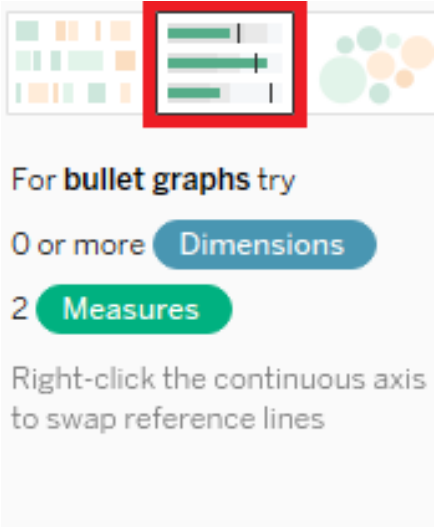
Average, median, sum: visualization

The **Show Me** pane shows you the options available to visualize your data.

If an option is **greyed out**, that means it's **not possible** to use it for the data you have. One other chart type that isn't greyed out here is a bullet chart (or graph), which often is used to look at a target VS the reality. In this case, we will tweak that traditional use of a bullet chart to look at the median compared to the average.

○   Select the bullet graph option.

You'll see a bar chart with a line marked on it. The thin black line is the median, $14,340. And the blue bar represents the average of $19,260:

For **bullet graphs** try

0 or more Dimensions

2 Measures

Right-click the continuous axis to swap reference lines

0K	1K	2K	3K	4K	5K	6K	7K	8K	9K	10K	11K	12K	13K	14K			18K	19K

Avg. total

Avg. total: 19,260
Median total: 14,340

Visualization per entity

The first bullet chart gave us a useful summary of the median lobbying payment and the average, but let's take a closer look by the type of entity or government. The bullet chart in this case is not really something we would want to build out as a chart for our readers. But it does tell us which entity types we might want to examine more closely.
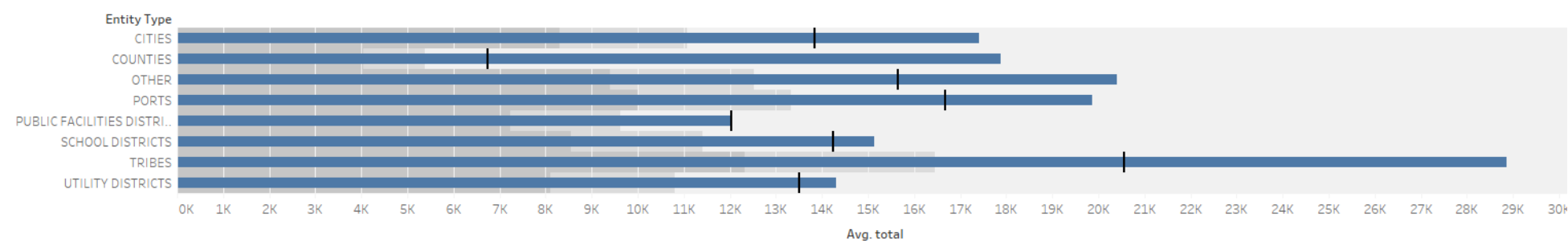

Where we see a big difference, that indicates that just a few local governments in that entity type accounted for much of the money going to lobbyists. For example, look at the difference between the mean and median in counties. Only two counties spent more than $60,000 on lobbying efforts. They also are the two most populous counties in the state: King County and Pierce County.
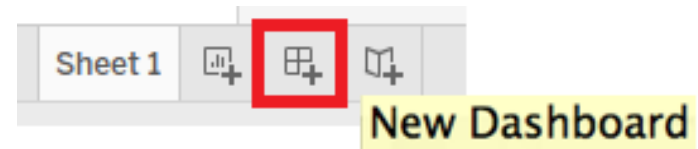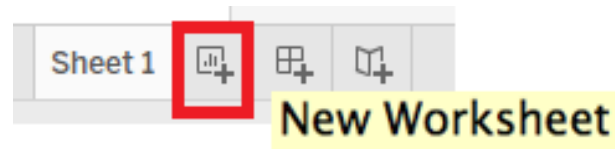
New analysis / visualizations?

- **undo button** in the top left corner of your screen to get back to the previous visualization

- add a new worksheet
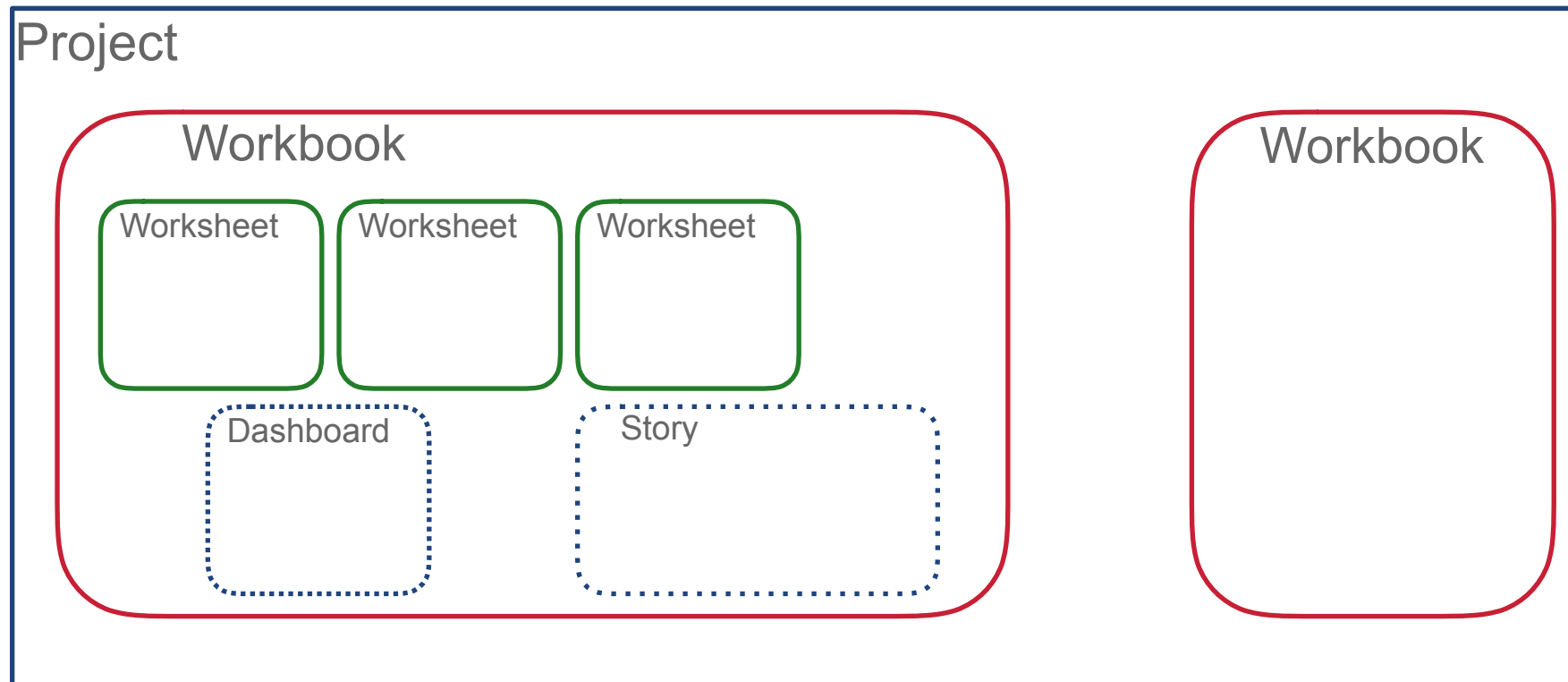
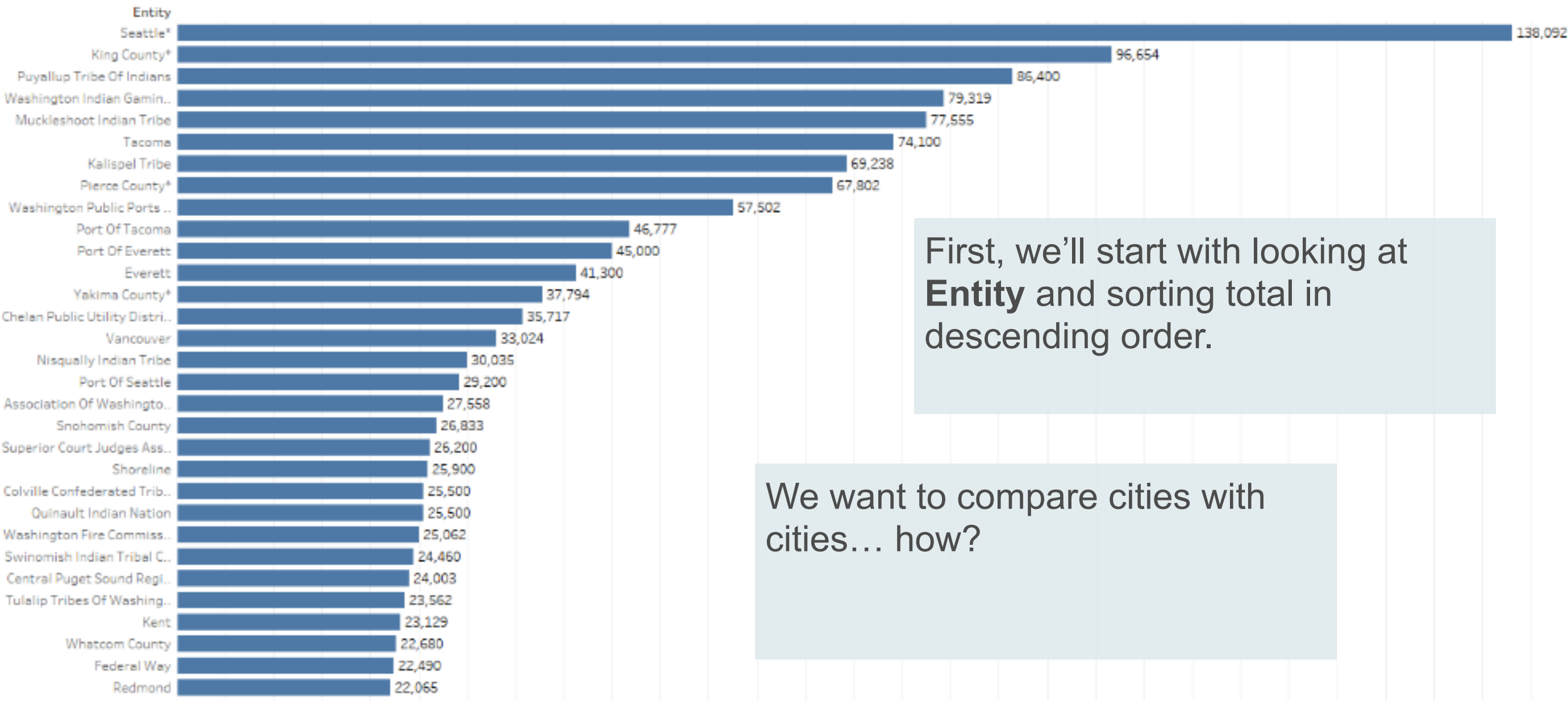- Copy the worksheet, and change the analysis
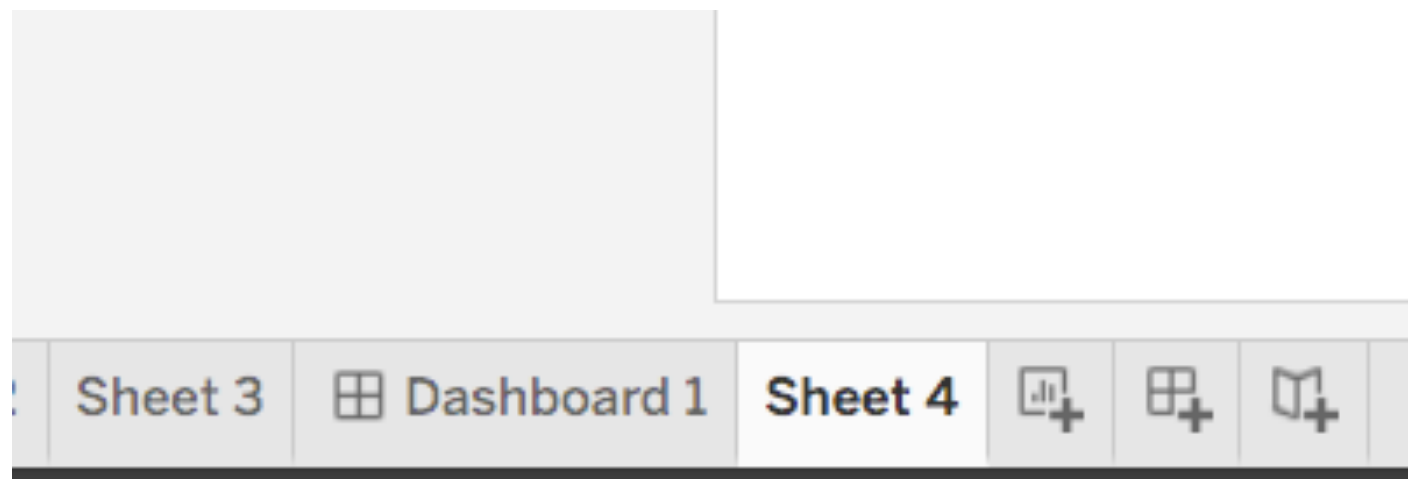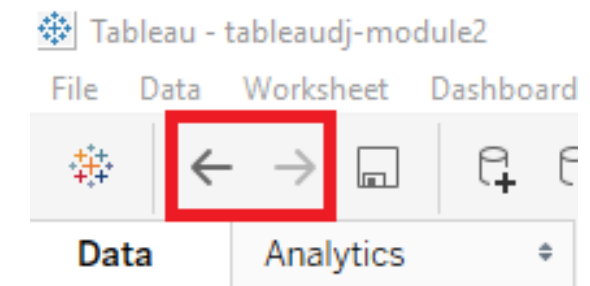
Worksheet - Dashboard - Story

○ For now, click on the plus sign for a new sheet.

Double-click on **Sheet 1** and then type the name you want for this sheet.
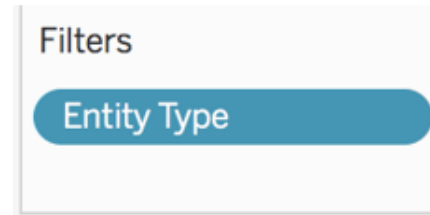
Filters

Entity	
Seattle*	138,092
King County*	96,654
Puyallup Tribe Of Indians	86,400
Washington Indian Gamin..	79,319
Muckleshoot Indian Tribe	77,555
Tacoma	74,100
Kalispel Tribe	69,238
Pierce County*	67,802
Washington Public Ports ..	57,502
Port Of Tacoma	46,777
Port Of Everett	45,000
Everett	41,300
Yakima County*	37,794
Chelan Public Utility Distri..	35,717
Vancouver	33,024
Nisqually Indian Tribe	30,035
Port Of Seattle	29,200
Association Of Washingto..	27,558
Snohomish County	26,833
Superior Court Judges Ass..	26,200
Shoreline	25,900
Colville Confederated Trib..	25,500
Quinault Indian Nation	25,500
Washington Fire Commiss..	25,062
Swinomish Indian Tribal C..	24,460
Central Puget Sound Regi..	24,003
Tulalip Tribes Of Washing..	23,562
Kent	23,129
Whatcom County	22,680
Federal Way	22,490
Redmond	22,065

First, we'll start with looking at **Entity** and sorting total in descending order.

We want to compare cities with cities… how?

Filters

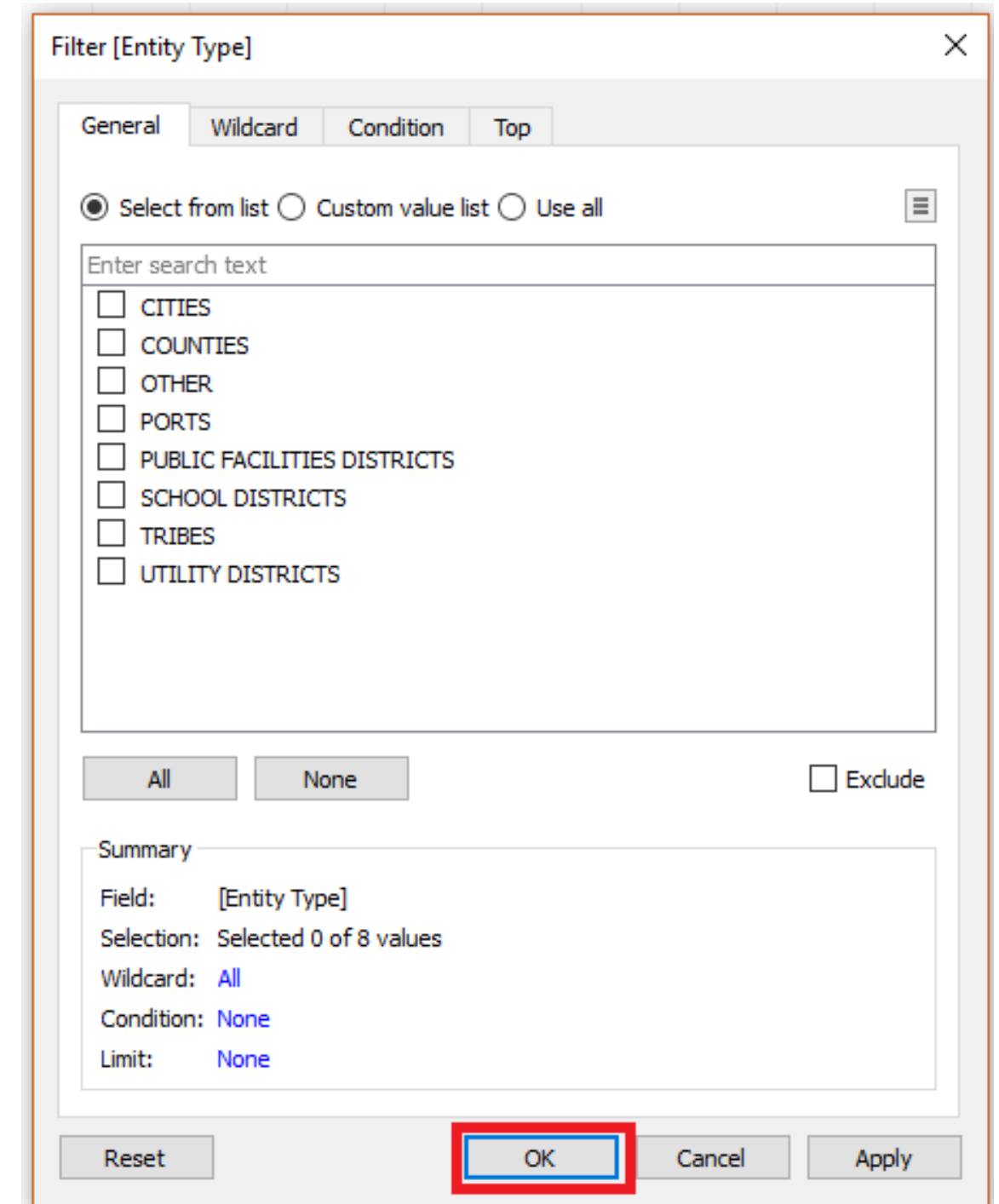○ Drag **Entity Type** onto the filter shelf:

Filters

Entity Type

When you do that, a dialogue box will pop up.

○ Click **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

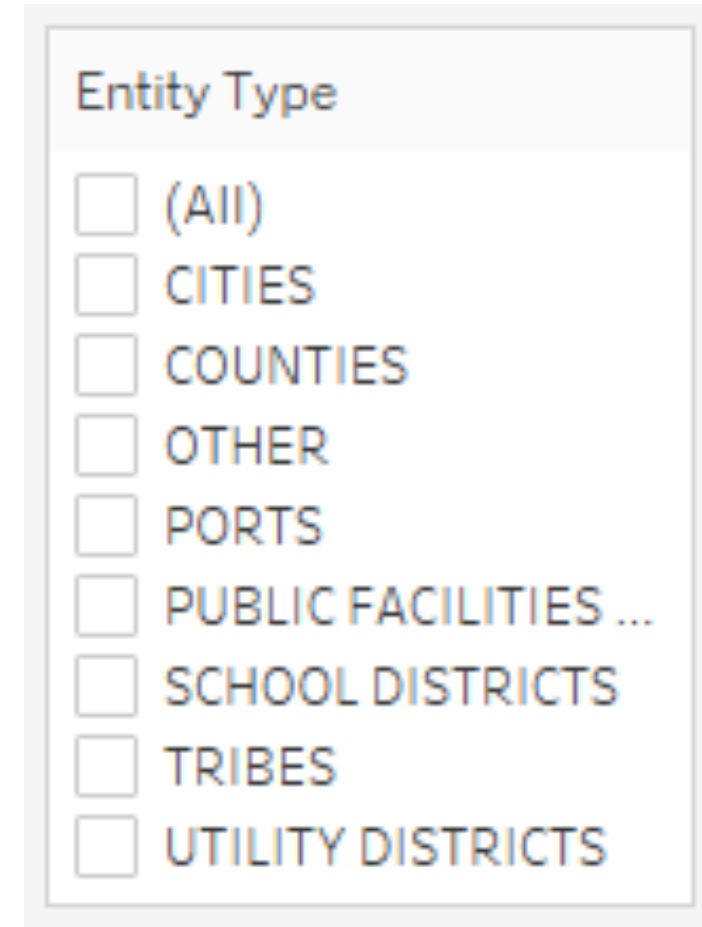Reset   OK   Cancel   Apply

Filters

○ Click on the caret on the right side of the **Entity Type** pill. The drop-down will offer you some choices.

○ Select **Show Filter**:
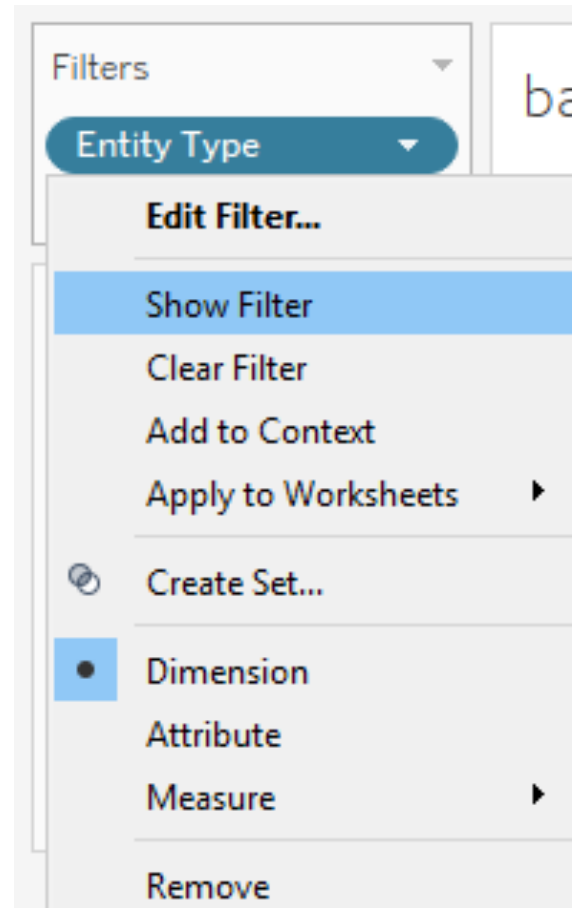
When you do that, a box will appear on the right-hand side of your workspace.

By default, Tableau will provide empty check boxes. You can select all or any of the options.

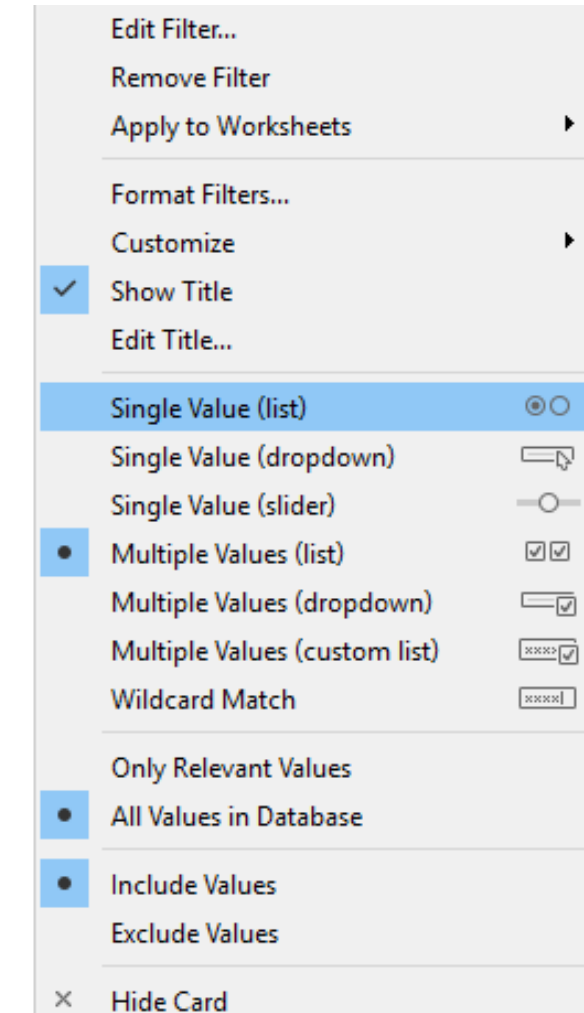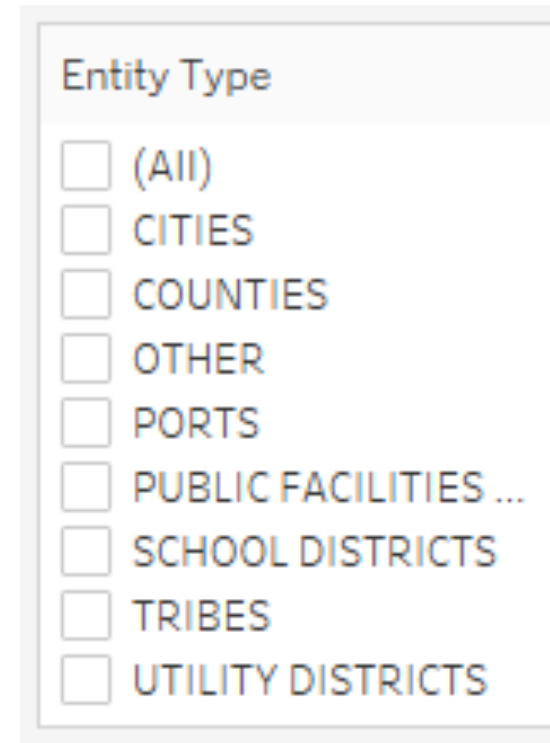Go ahead and practice to see what happens.

Filters

Different look?

○ Click on the caret on the right side of your filter, to the right of entity type.

When you do that, a box will appear. Tableau gives us a lot of options.

○ For now, we're going to select **Single Value (list)**.

Entity Type
☐ (All)
☐ CITIES
☐ COUNTIES
☐ OTHER
☐ PORTS
☐ PUBLIC FACILITIES ...
☐ SCHOOL DISTRICTS
☐ TRIBES
☐ UTILITY DISTRICTS

Edit Filter...
Remove Filter
Apply to Worksheets   ▸

Format Filters...
Customize   ▸
✓ Show Title
Edit Title...

Single Value (list)   ◉○
Single Value (dropdown)
Single Value (slider)
● Multiple Values (list)
Multiple Values (dropdown)
Multiple Values (custom list)
Wildcard Match

Only Relevant Values
● All Values in Database

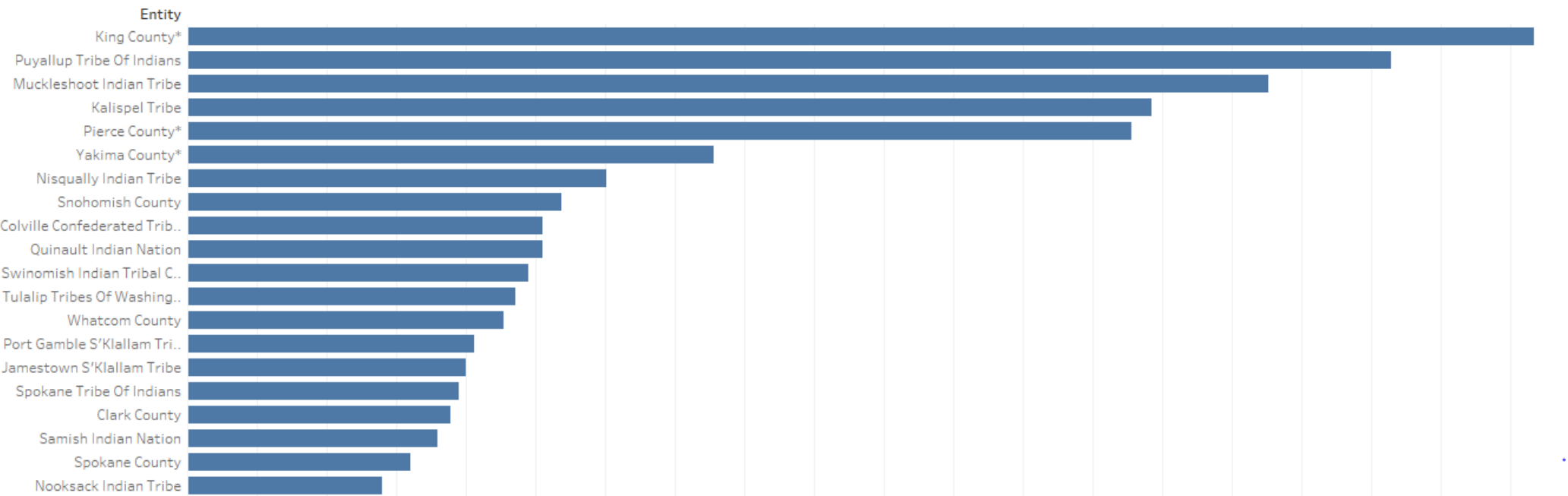● Include Values
Exclude Values

✕ Hide Card

Filters

Now, practice selecting and re-selecting different entity types.

You can easily see the amount of lobbying by the various local governments but filtered by the type of government.

Let's practice that for a minute.

Sheet 1

Entity

King County*	
Puyallup Tribe Of Indians	
Muckleshoot Indian Tribe	
Kalispel Tribe	
Pierce County*	
Yakima County*	
Nisqually Indian Tribe	
Snohomish County	
Colville Confederated Trib..	
Quinault Indian Nation	
Swinomish Indian Tribal C..	
Tulalip Tribes Of Washing..	
Whatcom County	
Port Gamble S'Klallam Tri..	
Jamestown S'Klallam Tribe	
Spokane Tribe Of Indians	
Clark County	
Samish Indian Nation	
Spokane County	
Nooksack Indian Tribe	

Basic filters and analyzing the data for story

Now, as additional practice, how many entity types do I have?

Which type of entity accounted for the most in lobbying spending, and how much was that?

How much each of these gave as a percentage out of the total?

Should I do another calculated field? What other ways, based on what we learned in the last module, could I contemplate exploring that?

Tree map

Tableau also has some nice shortcuts. Instead of dragging the pill out to a certain shelf. Try double-clicking total. Then double-click **Entity Type**. You should end up with a tabular chart as a default.
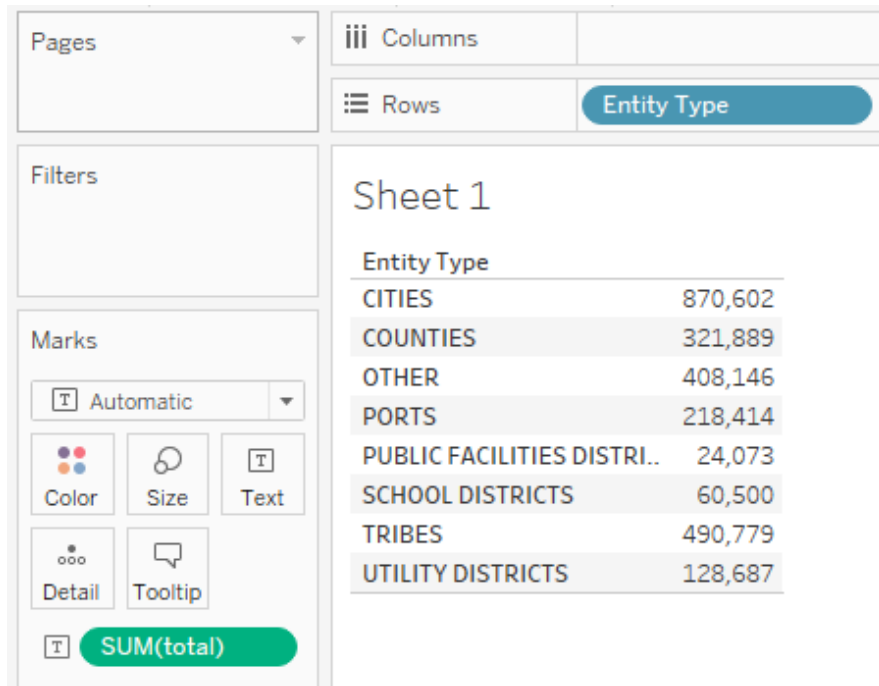
The **Show Me** pane will show you the other visualization options.
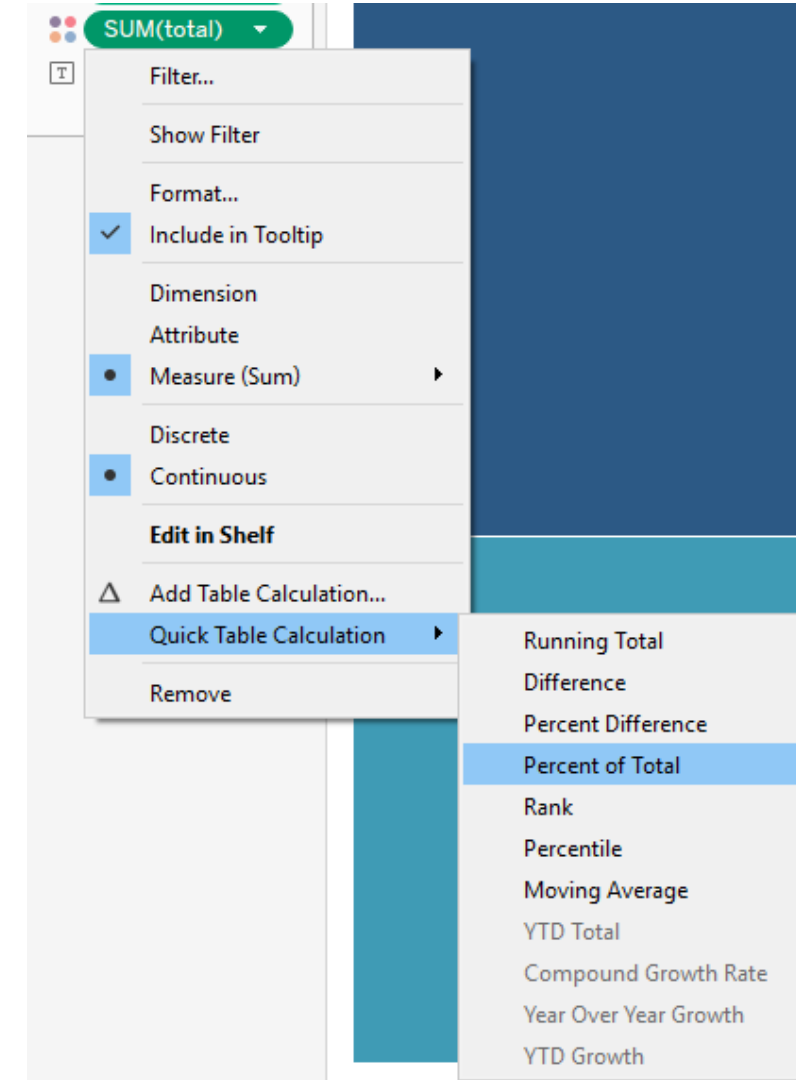
○ Select **Tree Map**.

Tree map

If we hover over each section, we can see the total in lobbying to that entity type.

But we also can add a calculation – the percentage out of the total would be useful to know.
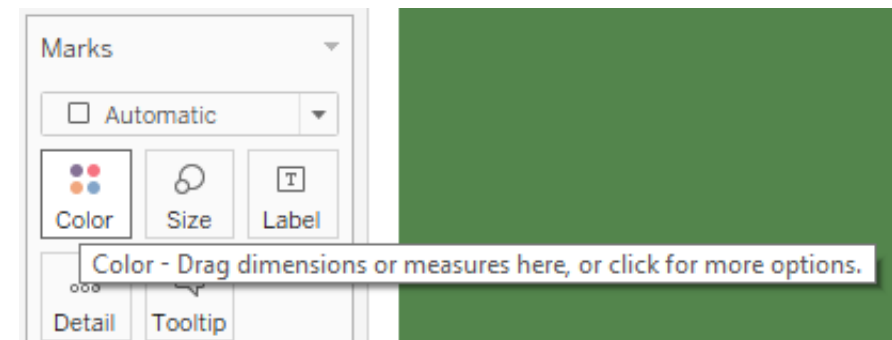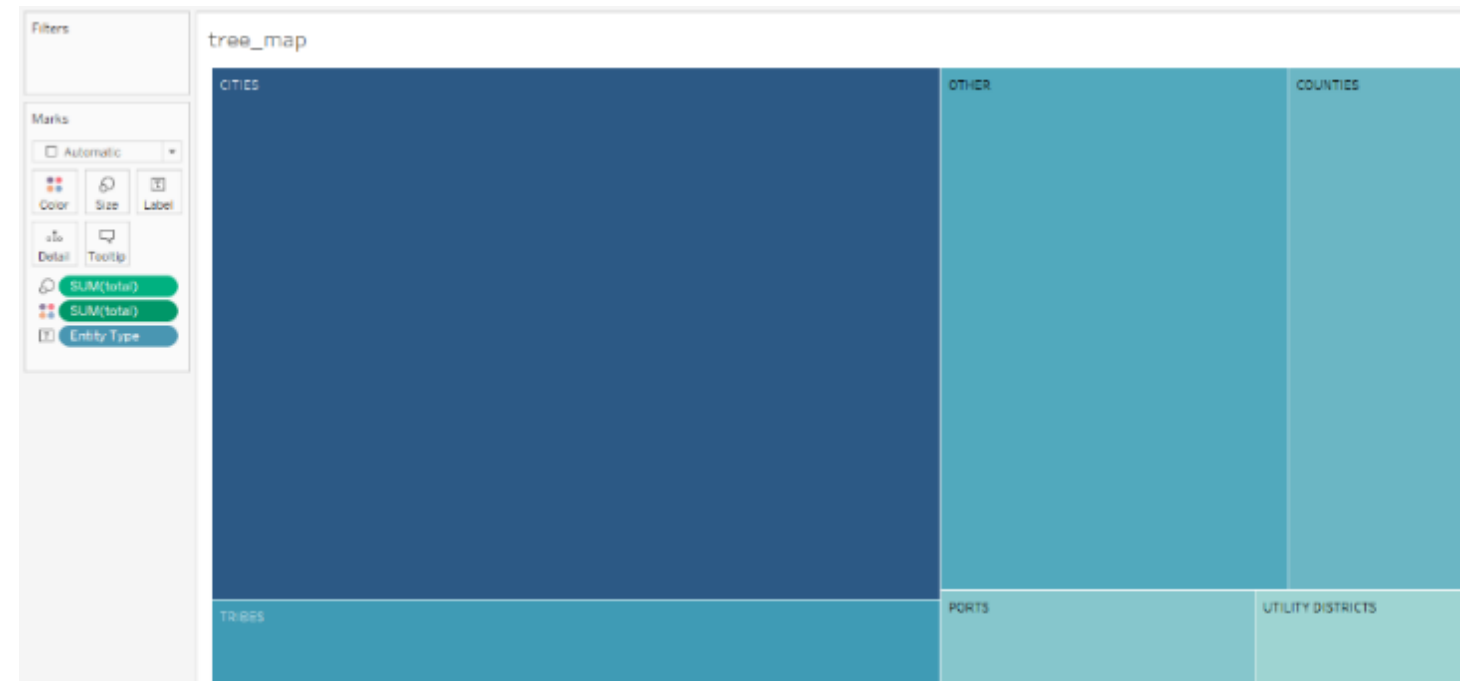
Then, when you hover over the tree map, you will see the type of government, the total and the percentage of the total.

Tree map

Adjustments in Color. You can change the color scheme of your tree map by double-clicking on color and selecting edit. Go ahead and try out some other options for the tree map.

After you've played around a bit, change the color scheme to green.

Tree map

Adjustments in tooltip styles.

Since we are using the tooltip with this tree map, let's take the opportunity to learn how to edit a tooltip.

○ Double-click on the **tooltip icon** in the marks shelf. A box will open. You can edit the font and text right from the tooltip box.

○ Go ahead and change **Entity Type** to **Government**, and **% of Total along Table (Across)** to **% of Total SUM (total).**

○ Click **OK** and look at your tooltip. We'll do more with tooltips in future modules.
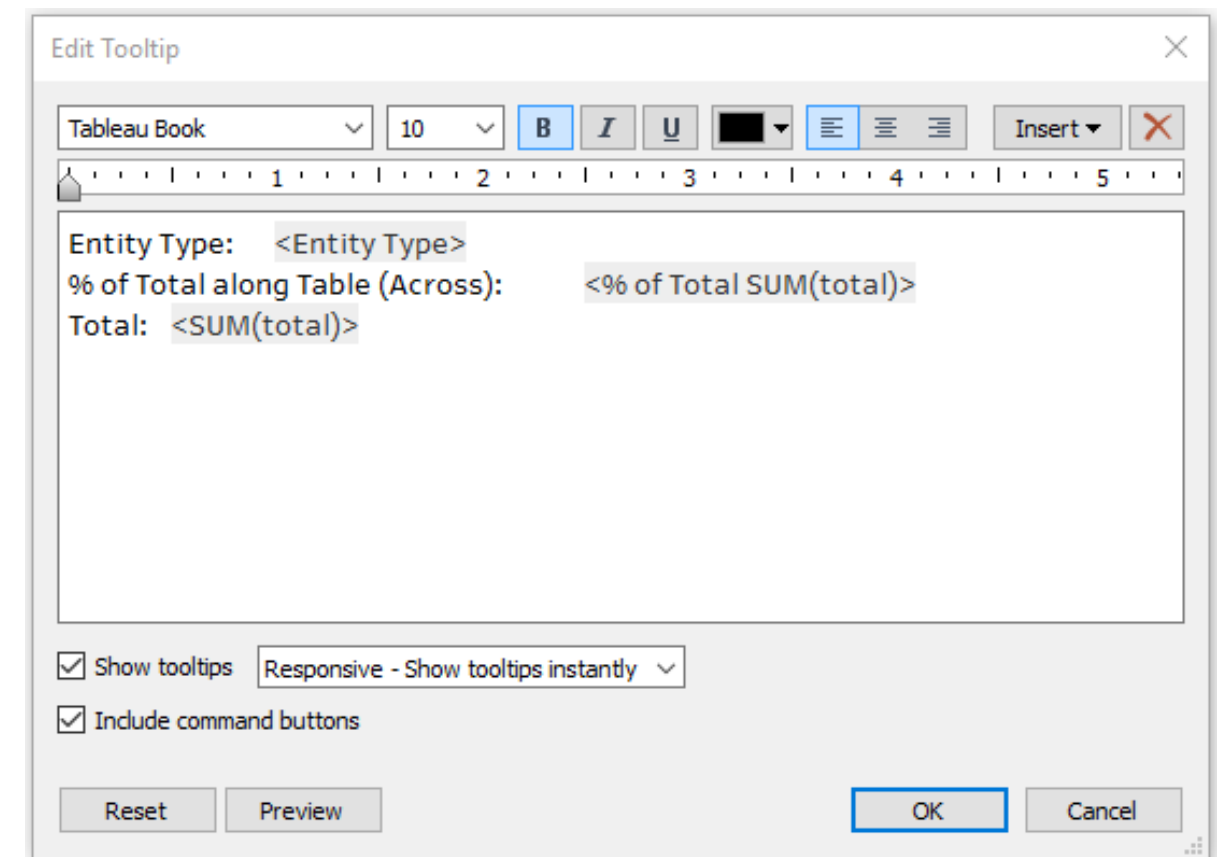
Basic filters and analyzing the data for story

- Now, as additional practice, how many entity types do I have?

- Which type of entity accounted for the most in lobbying spending?

- And how much was that?

Basic filters and analyzing the data for story

- So, is it newsworthy if the most populous areas, such as King County, or the city of Seattle, are responsible for most of the lobbying to the legislature?

- What about smaller governmental bodies and how much are they paying lobbyists?

- How could we look at that in a way that would be enlightening?
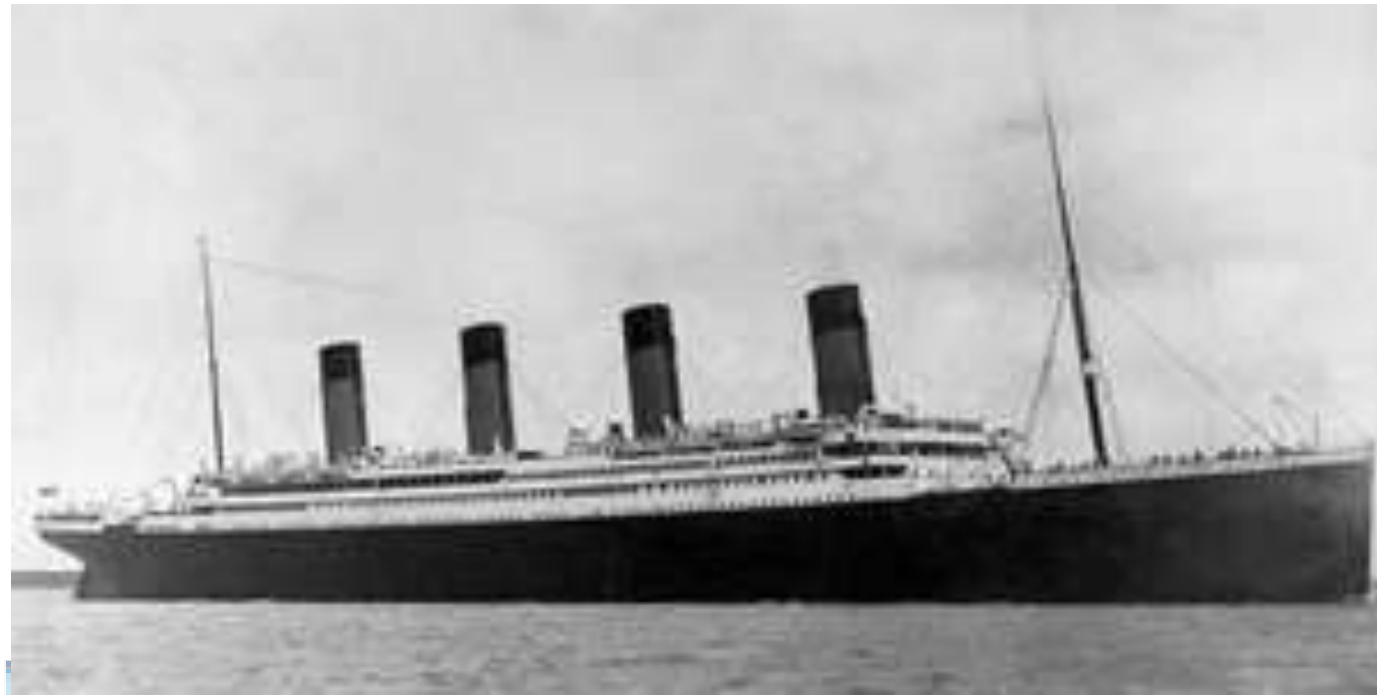
- **What additional data might we need?**

Titanic!

Titanic dataset and cool things you can do with it: https://www.kaggle.com/c/titanic

Tableau official training: https://www.tableau.com/learn

Tableau examples with Titanic data: https://public.tableau.com/search/all/titanic

Titanic Questions

- Can you attribute survival to a single primary trait?
- In other words, can you attempt to tease apart the confounding effects of pclass/age/sex, when assessing one?

- Can you explain why the distribution of fares by pclass seems off? why are some 3rd class tickets more expensive than first class?

- Does group size have an effect on survival rate?
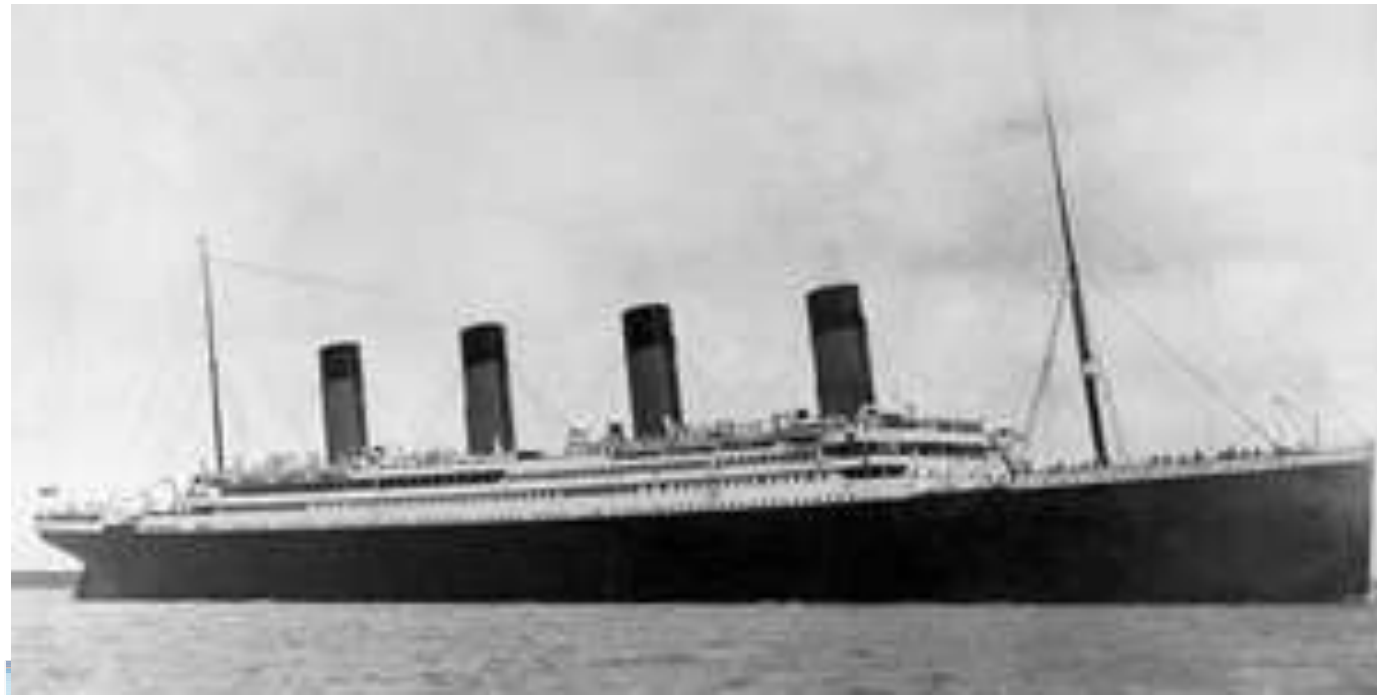
Reading for next week

- In https://github.com/gquer/dsc-96_winter19/blob/master/02_data_messy/readings.md
- By Wednesday Jan. 16 at 6.00pm
- To gquer@ucsd.edu
- Subject: [DSC 96 W02 SecA Journal]: *Name LastName*