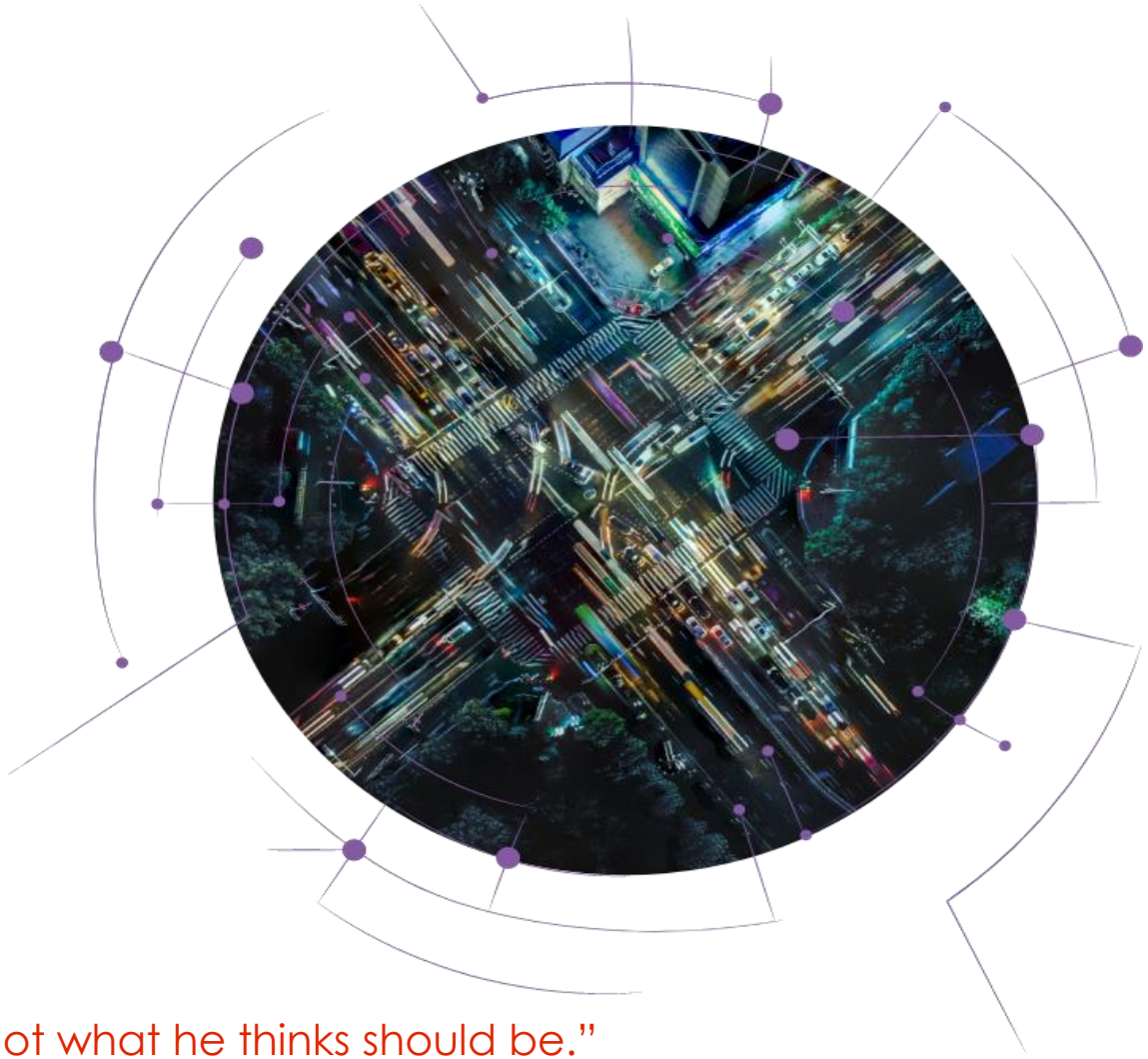
DATASOCIETY:

Intro to Tableau

Day 3



"One should look for what is and not what he thinks should be."

- Albert Einstein

Warm-up chat question

- There are many public datasets out there available for use and experimentation
- Have a look at the list at the following address: https://github.com/awesomedata/awesome-public-datasets

Which of these datasets might you like to explore?

Agenda

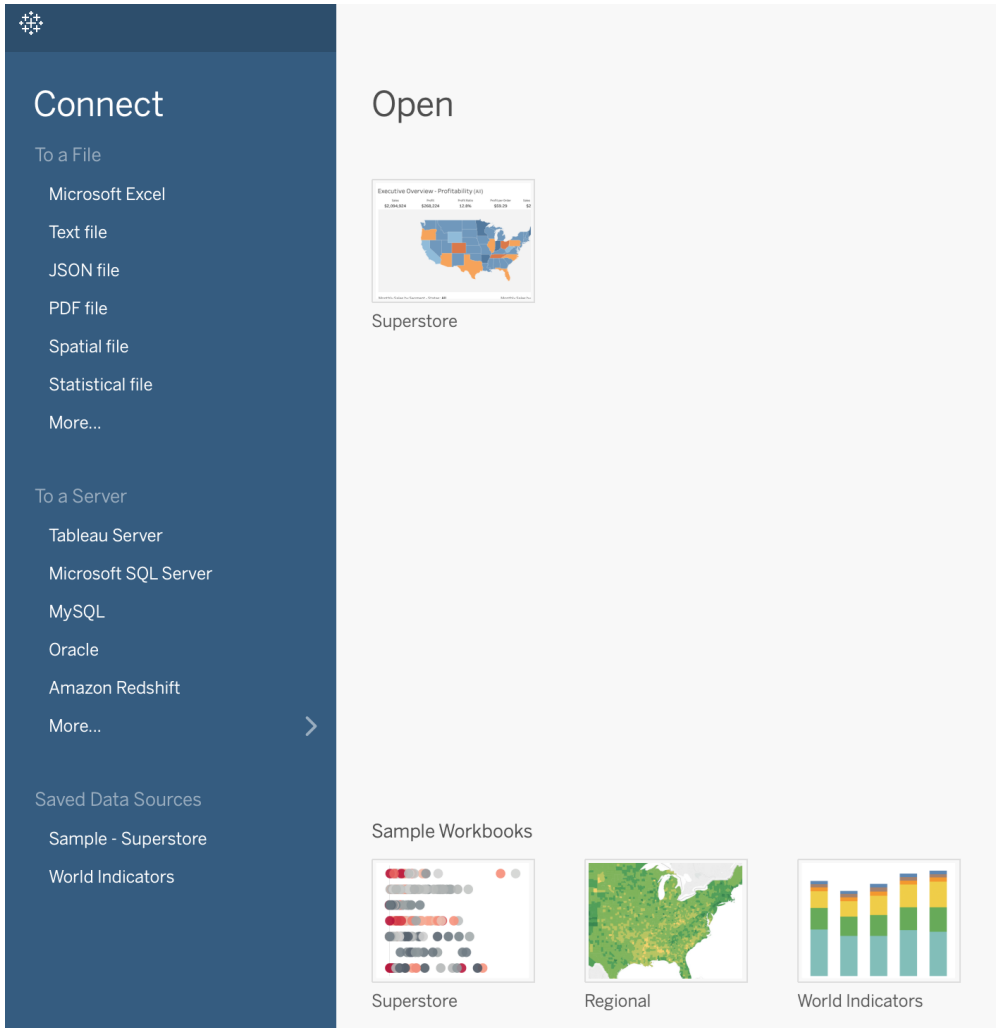
- Discuss filtering and formatting capabilities in Tableau
- Explain the concept of functions
- Implement basic functions on the dataset
- Introduce and implement table calculations

Recap: importing data

- Import data with the **Connect** panel

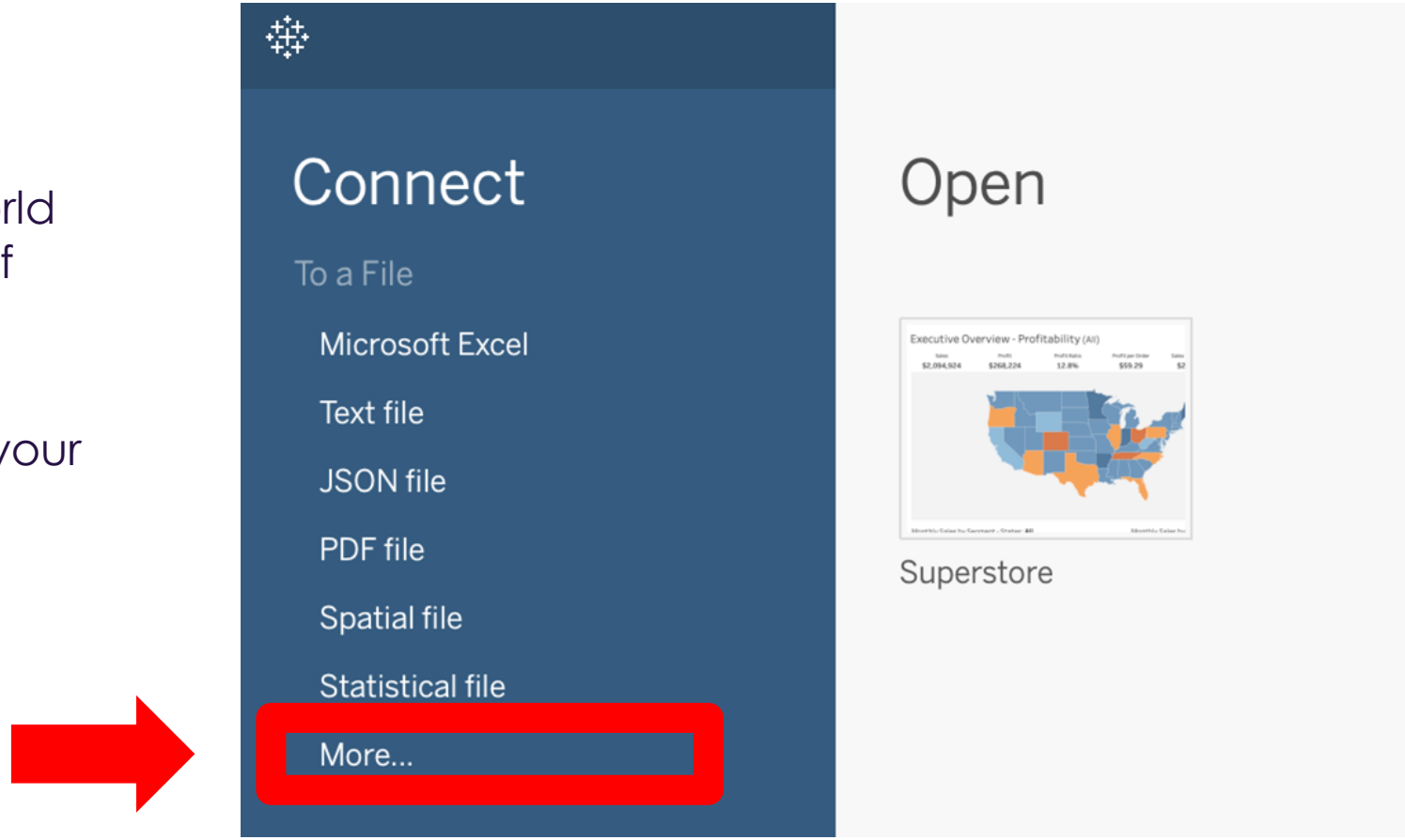
- Supports multiple formats such as:
  - Microsoft Excel (.xlsx)
  - Text (.txt, .csv)
  - JSON (.json)
  - PDF (.pdf)
  - R data format (.RData)

- Supports Database Connections such as:
  - MySQL
  - Oracle
  - Redshift



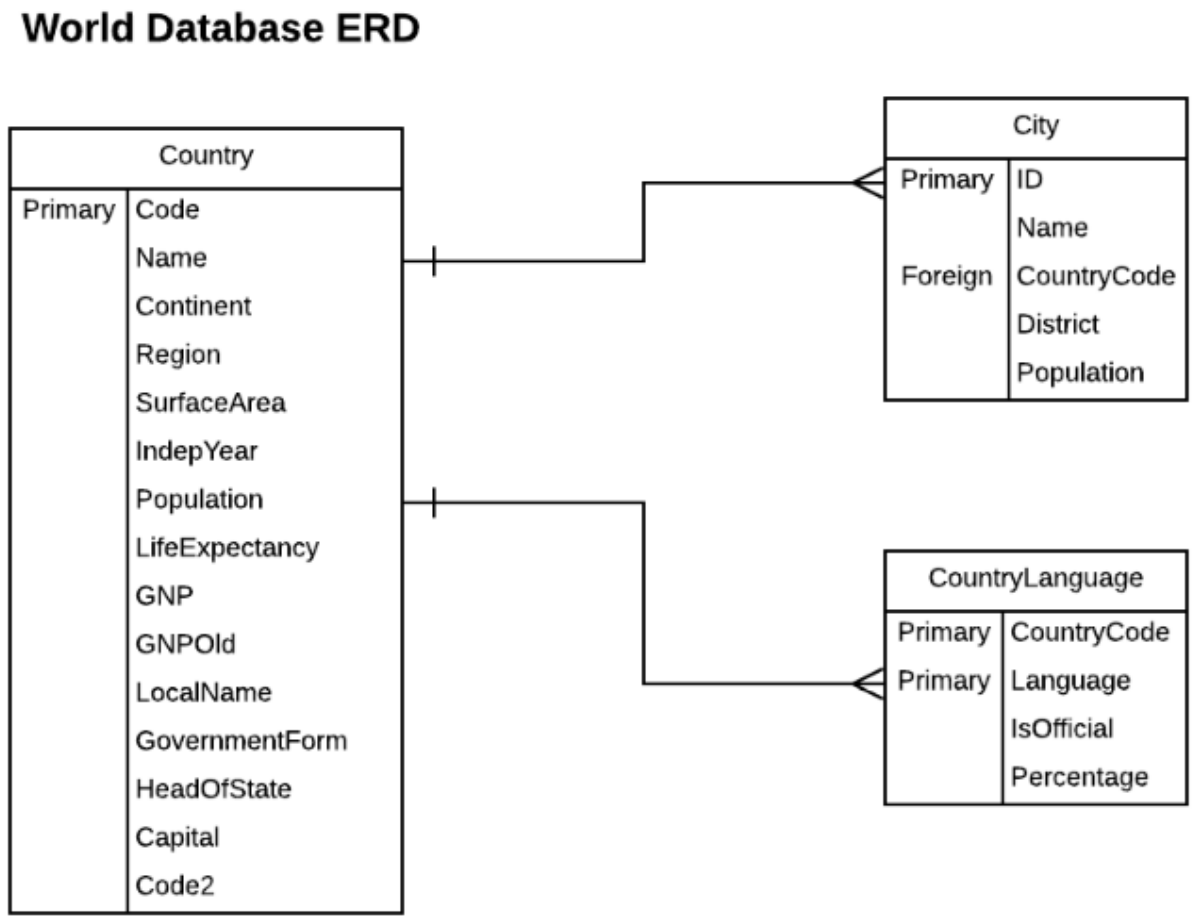
Import World Data : CSV

- Let's import some pieces of the world dataset today and see what sort of insights we can reveal
- Click the **"More..."** item to browse your local CSV files



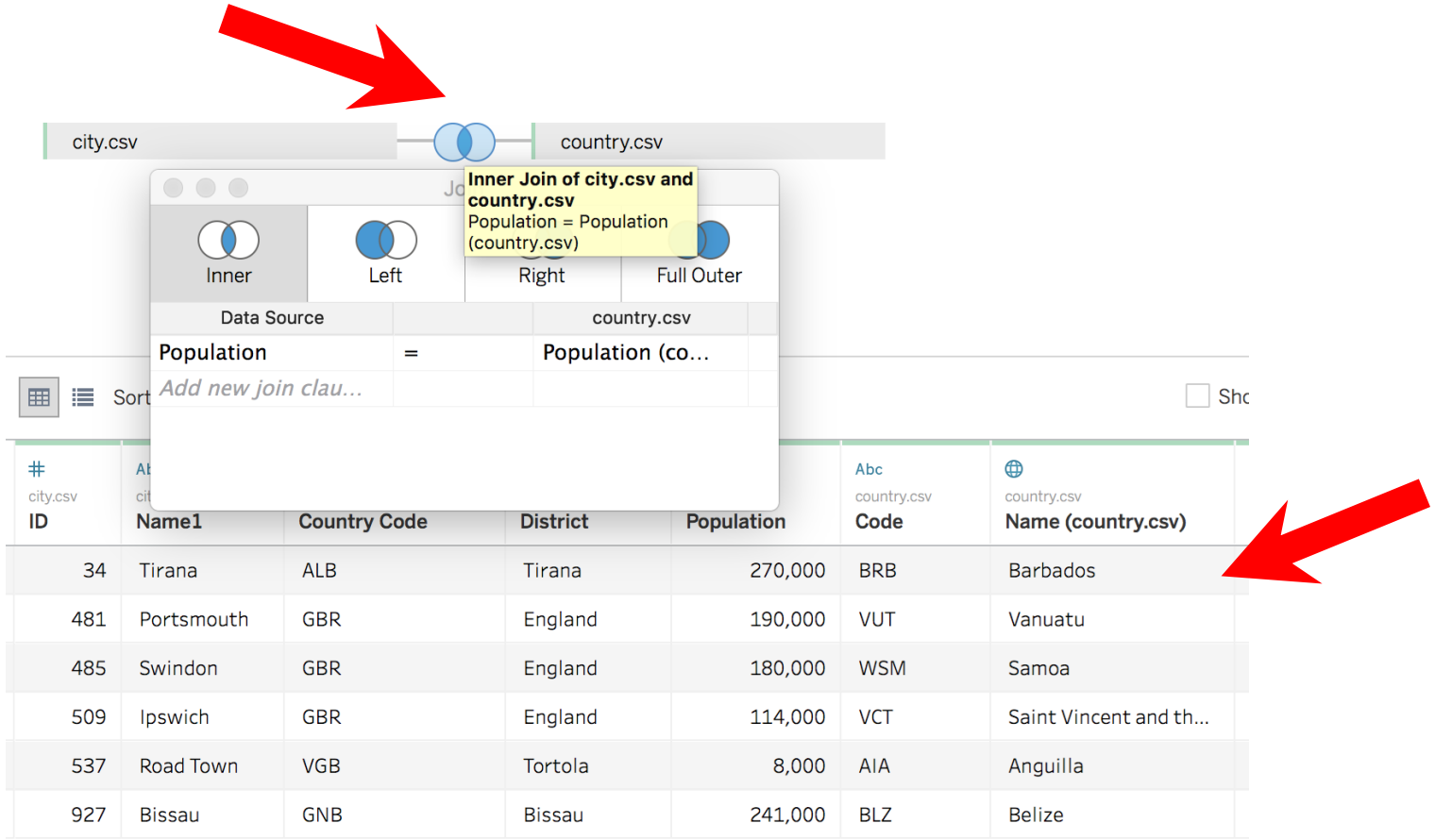
Recap: world database

- For now, import the following three CSV files:
  - **country.csv**
  - **city.csv**
  - **countrylanguage.csv**
- We'll use the other CSV files during our Exercises



Recap: auto-join error

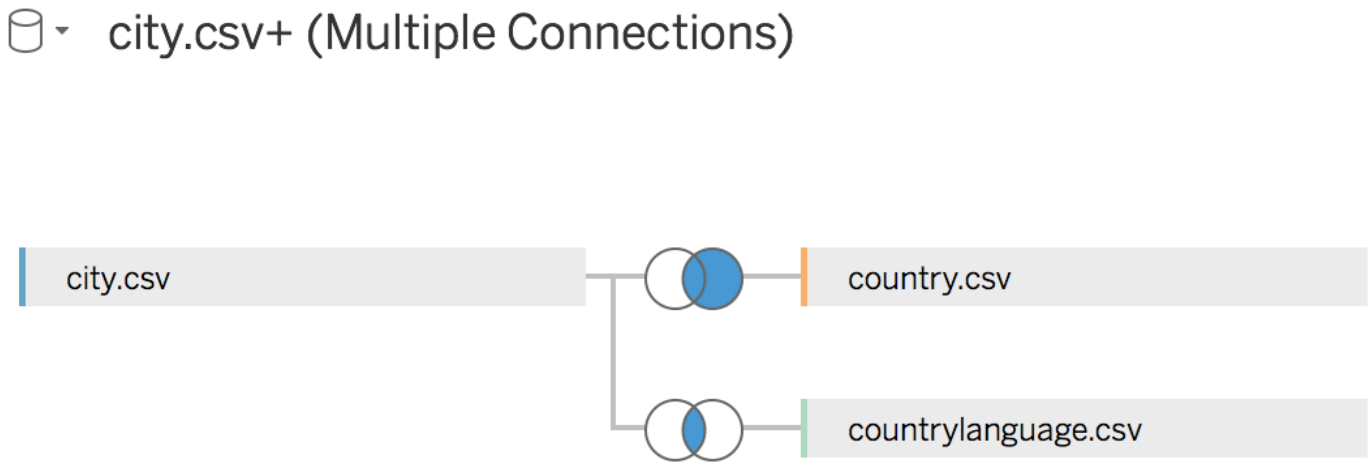
- Where did the countries go?
- Change to **inner join** if you want to see the intersection between the two tables
- Press the join drawing to see what is being joined and **fix it if necessary**



city.csv ID	Name1	Country Code	District	Population	country.csv Code	country.csv Name (country.csv)
34	Tirana	ALB	Tirana	270,000	BRB	Barbados
481	Portsmouth	GBR	England	190,000	VUT	Vanuatu
485	Swindon	GBR	England	180,000	WSM	Samoa
509	Ipswich	GBR	England	114,000	VCT	Saint Vincent and th...
537	Road Town	VGB	Tortola	8,000	AIA	Anguilla
927	Bissau	GNB	Bissau	241,000	BLZ	Belize

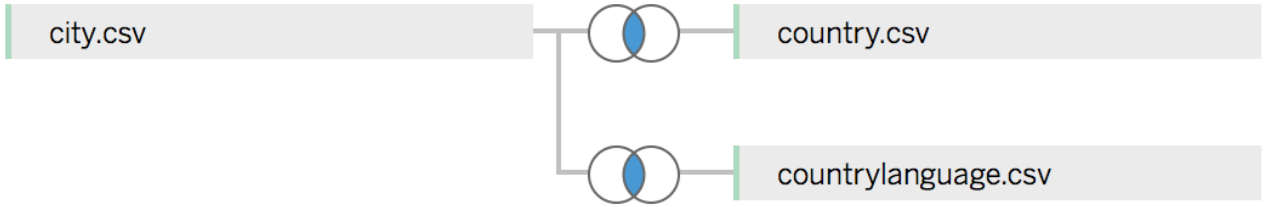
Recap: add the third table

- Now try joining the **country language** table
- Which join best combines the three datasets?
- Does the order in which you import tables matter?
- Why did you choose that order and those joins?



Recap: sequencing joins

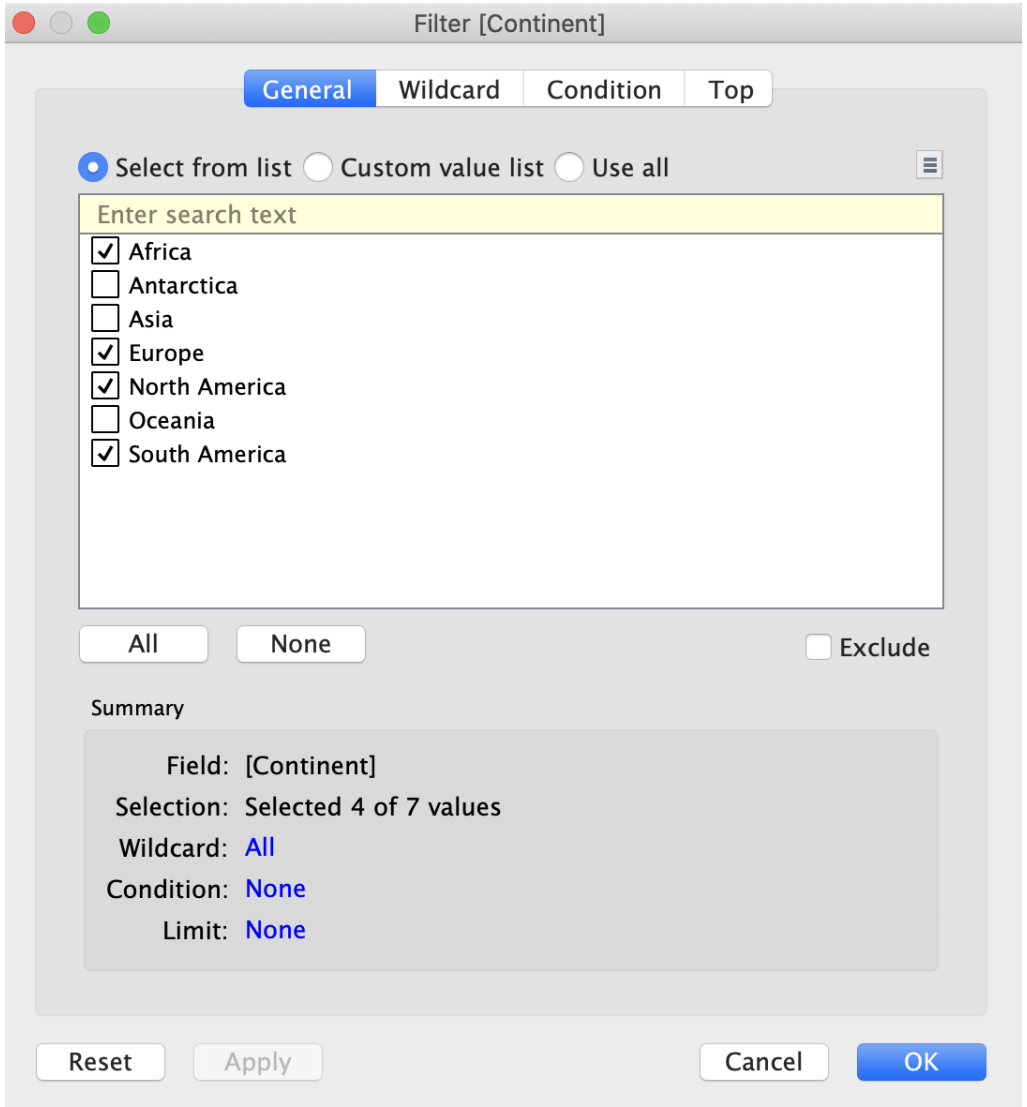
- Try out this sequence of joins:
  - First, an inner join of city and country using the country code
  - Next, an inner join of country and country language using country codes, as well



- Why is it okay to join country language to city?

Cleaning and focusing: filters

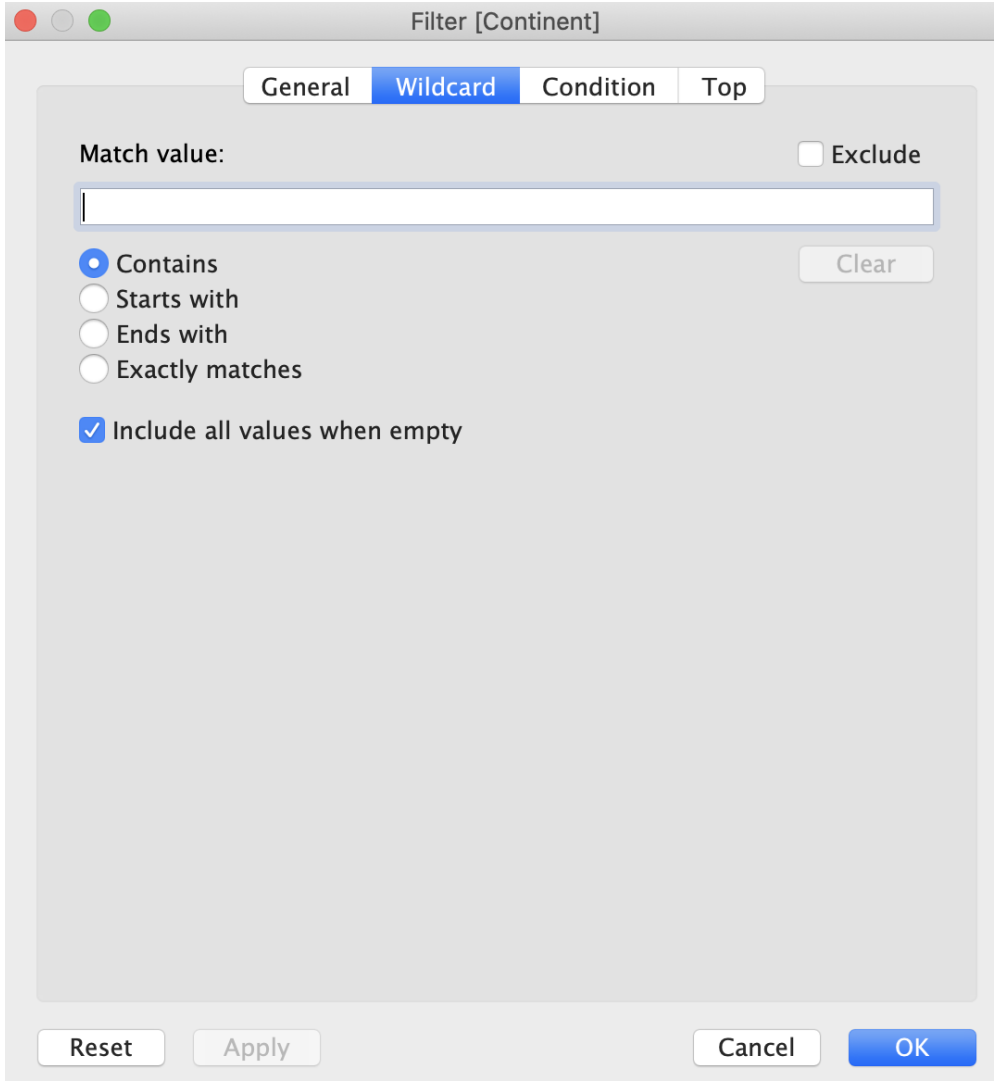
- It is often appropriate to graph only a portion of the data when:
  - there are credible **outliers**
  - there are **bad data** such as input errors, null values or coded values
    - i.e. missing values are often coded as integers and need to be filtered
  - the question only addresses a certain **interval**
    - i.e. "did the administration of a drug reduce mortality in the *following month*?"



Filtering the world data: match value

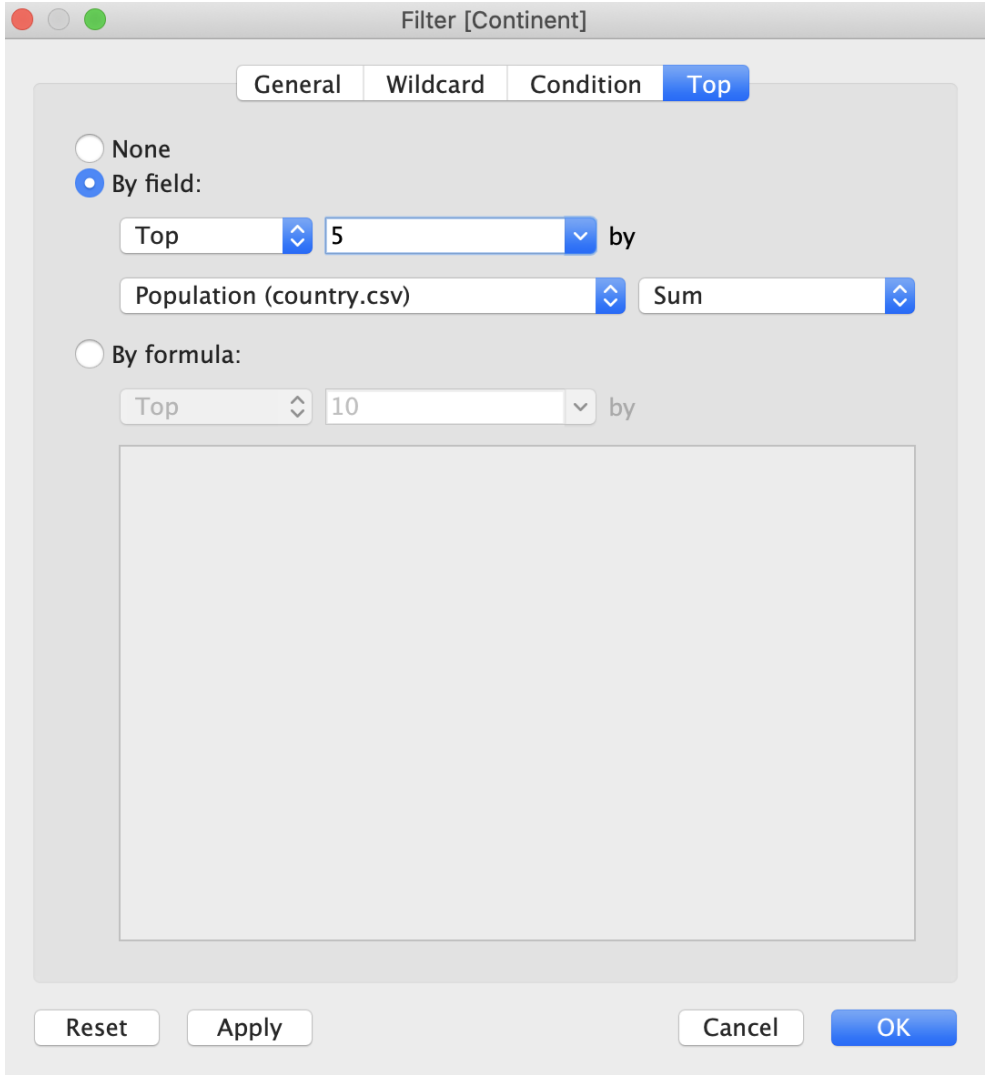
- We can use different type of **filters** to get more specific data

- Adding a match value will only include or exclude those columns that match that value



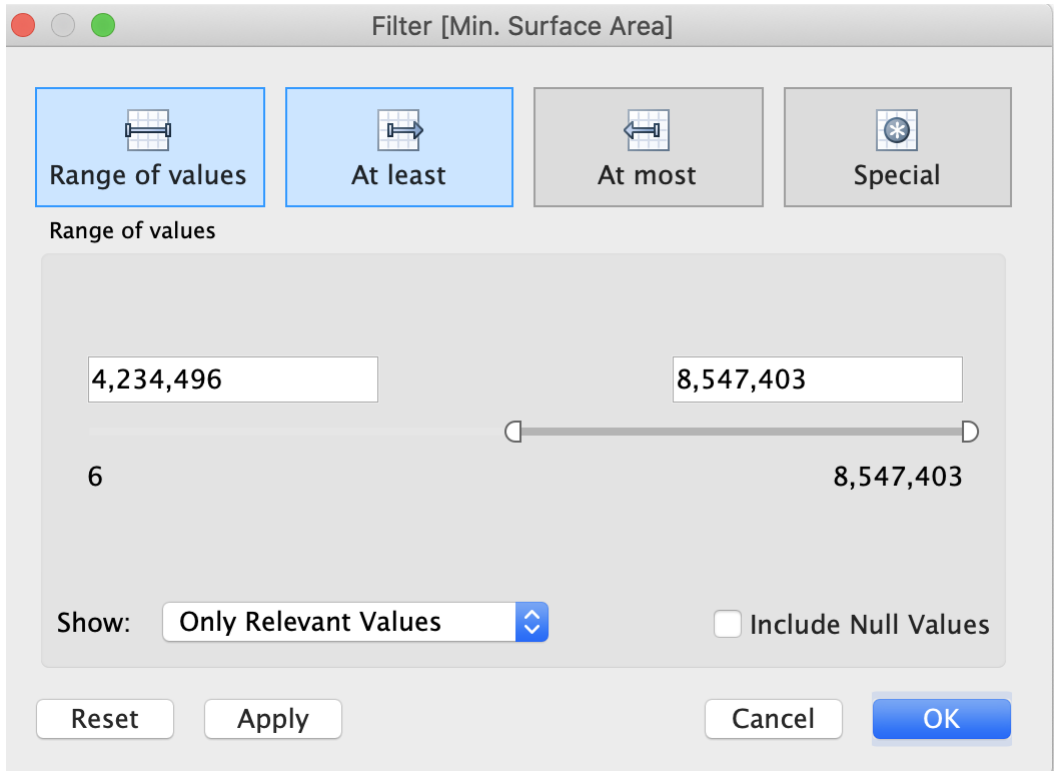
Filtering the world data: top values

- If we want to display only top results we can use the **Top** option and enter the top number of values that we want to see returned



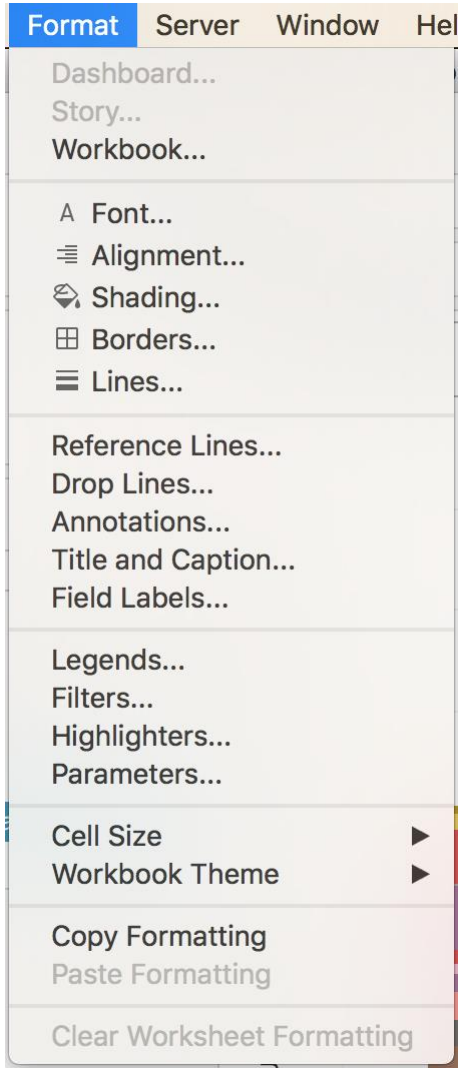
Filtering the world data: ranges

- We can also use specific ranges to filter our data. This only works when our data is quantitative.
- There are different ways of filtering by range
- We can also choose to filter out null values.



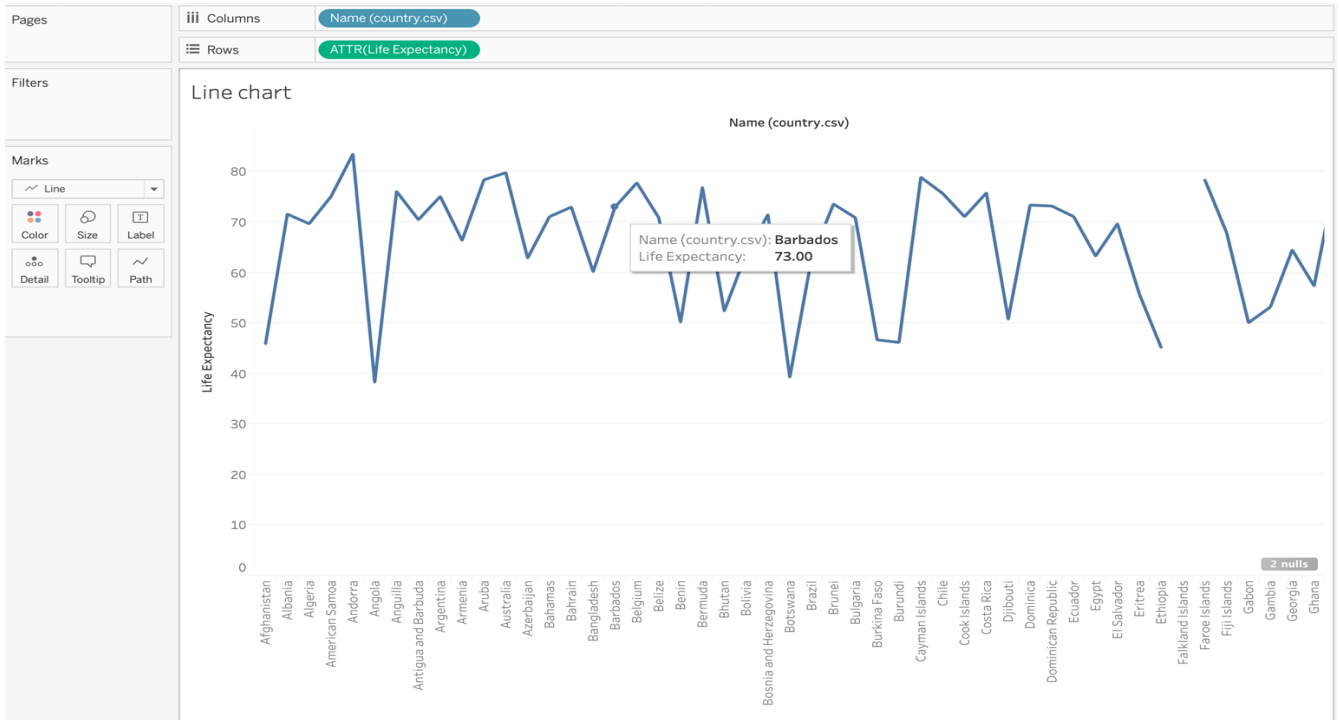
Formatting your figures

- Open formatting pane with the **Format** menu
- The formatting pane is contextual, meaning it changes based on what is highlighted
- Notice the **different elements** Tableau will let you format, from font and line type, to annotations, to labels and legends



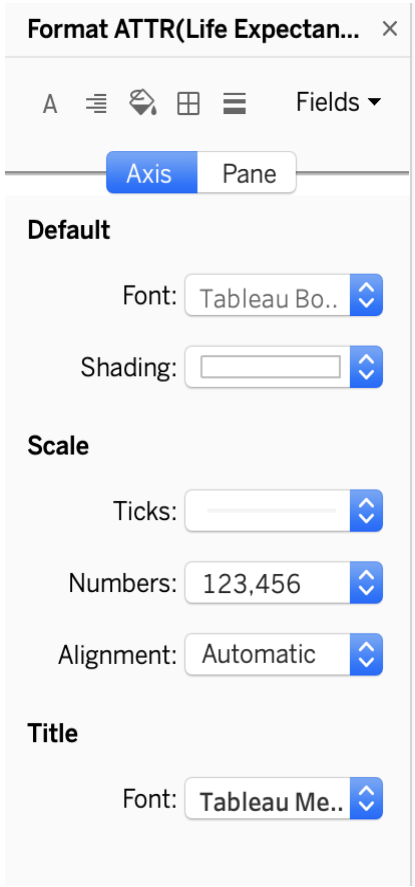
Formatting your figures

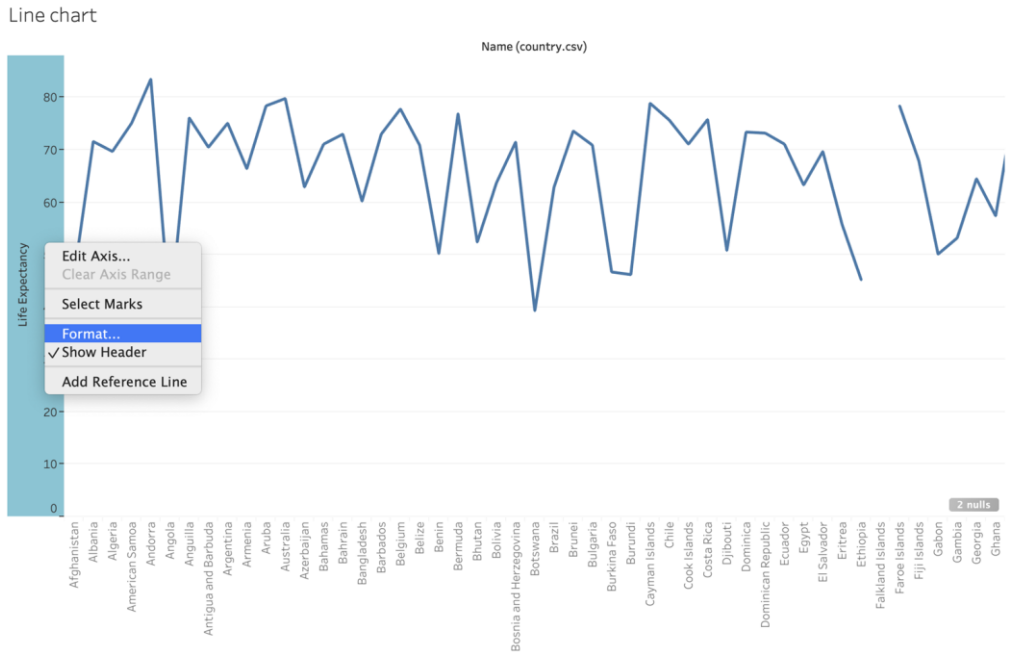
- Let's start with the line chart that we made for showing life expectancy
- What would you do to **improve** this?
  - What is being shown here? What aspects are most important?
  - Can you think of any existing variable that would add insight?
  - Now what about font and colors?



Formatting : right-clicking an element

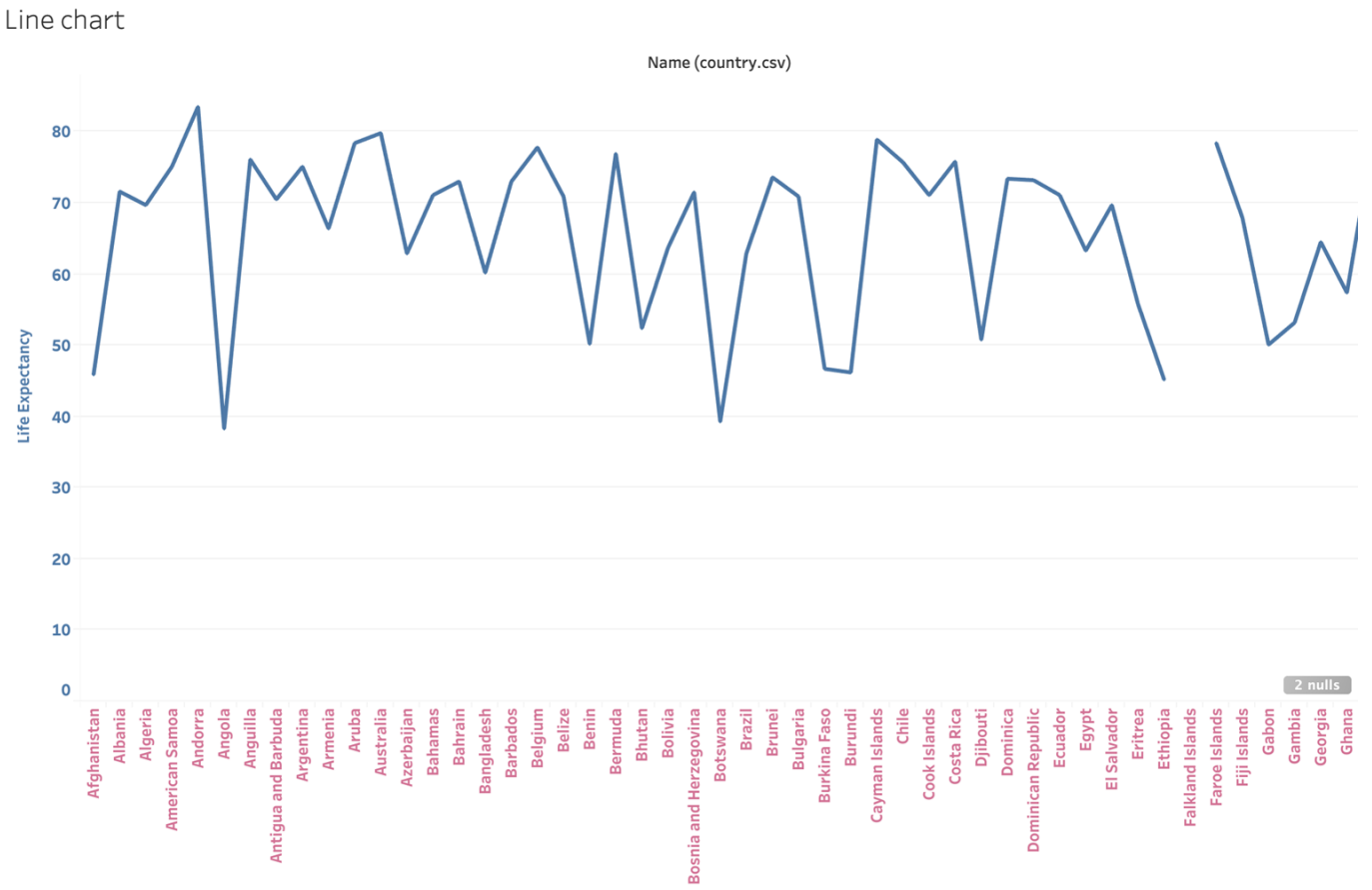
- In addition to using the Format menu, we can **open the formatting pane by right-clicking** on any view element
- Start by selecting the y axis of our "Life expectancy" view
- Note that the contextual format pane pops up on the left





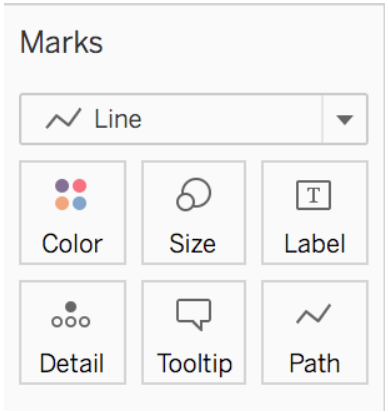
Formatting : customizing an axis

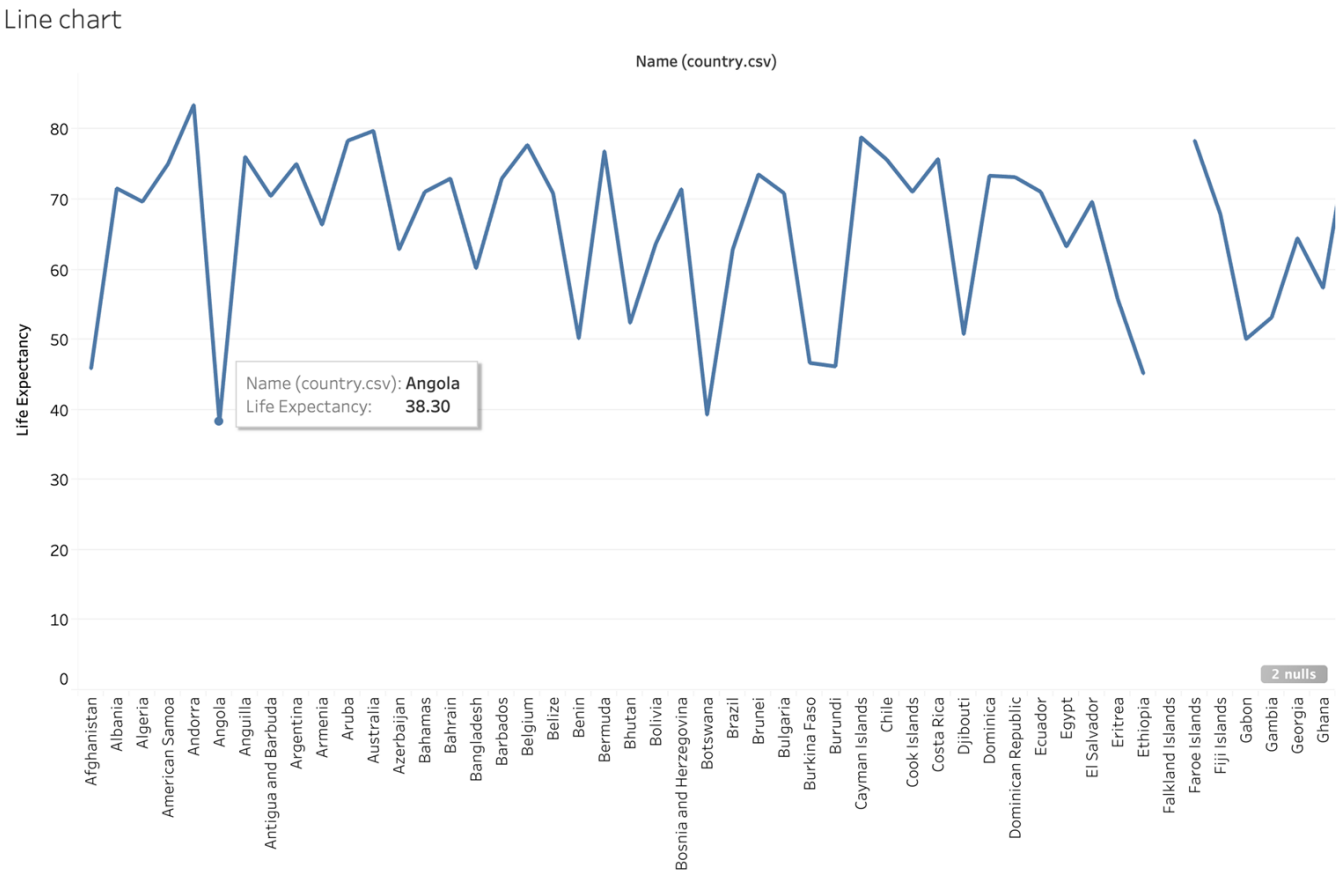
- Let's change the font, size and color of the axis
- Experiment with these elements in order to
  - Make the numbers easier to read
  - Make the year stand out
- What kinds of changes did you make?



Formatting : the Marks card

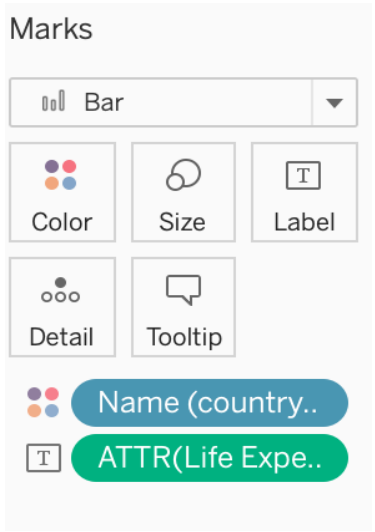
- The marks card is a powerful tool that allows you to add "marks"
  - Marks visually **highlight** certain features of the data
  - We've started with a very basic version of this figure with no marks

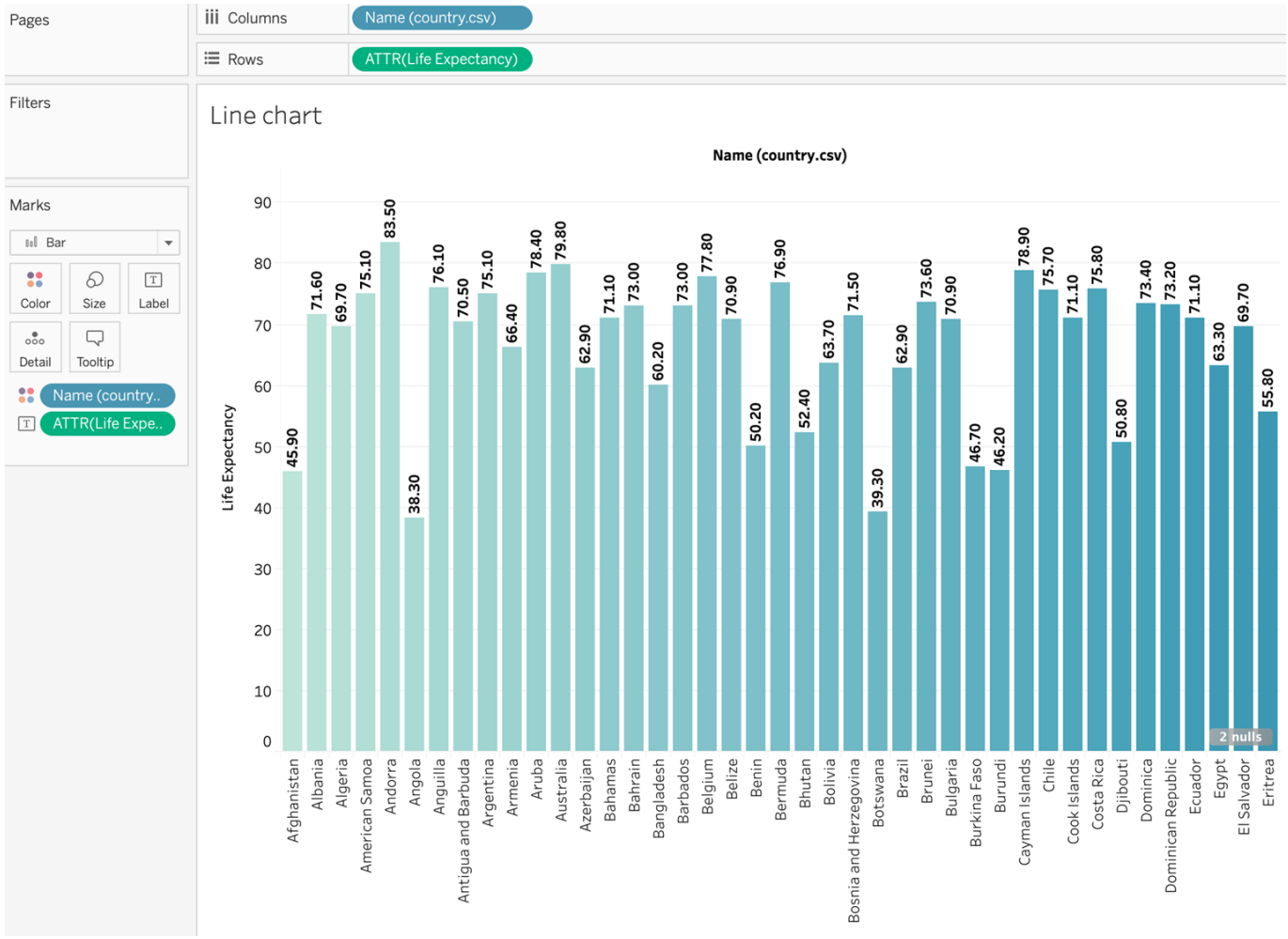




Formatting with the marks card

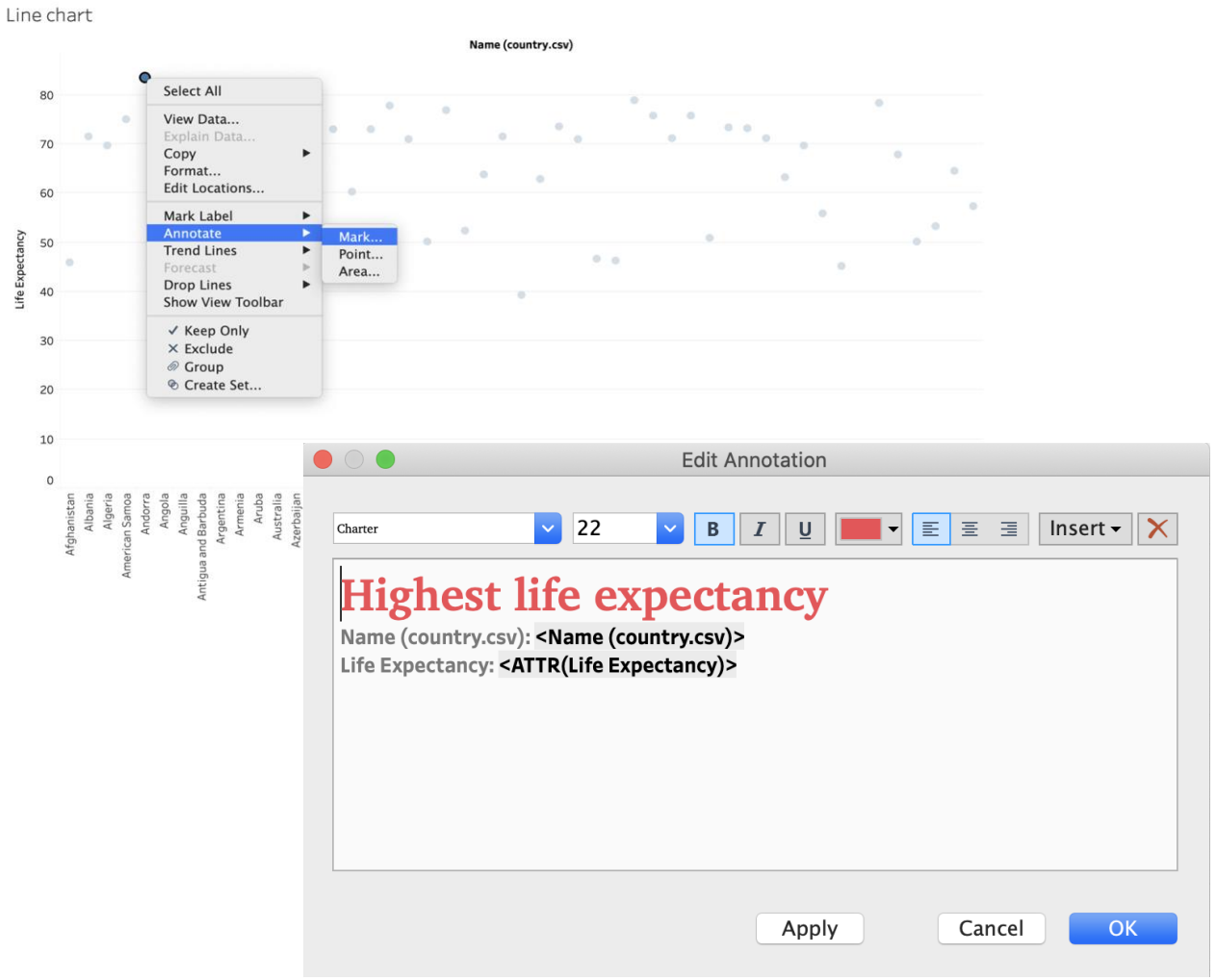
- With 3 changes to the Marks card, we can:
  - produce a visually interesting **color**
  - add a count **label**
  - convert the line chart into a **bar chart** appropriate for discrete events





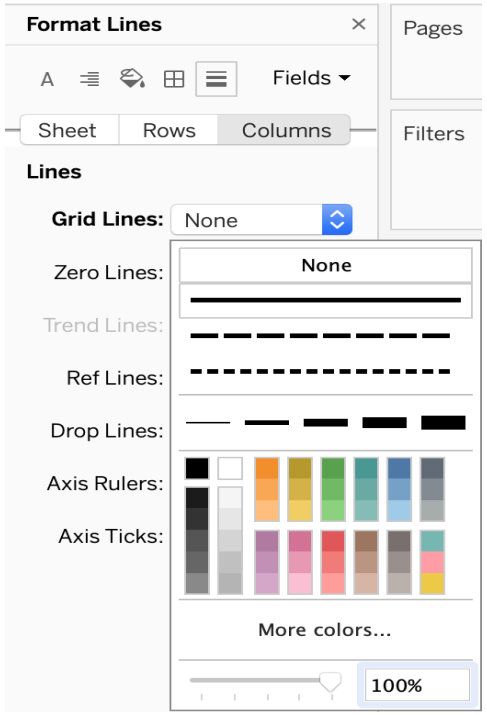
Annotating a point

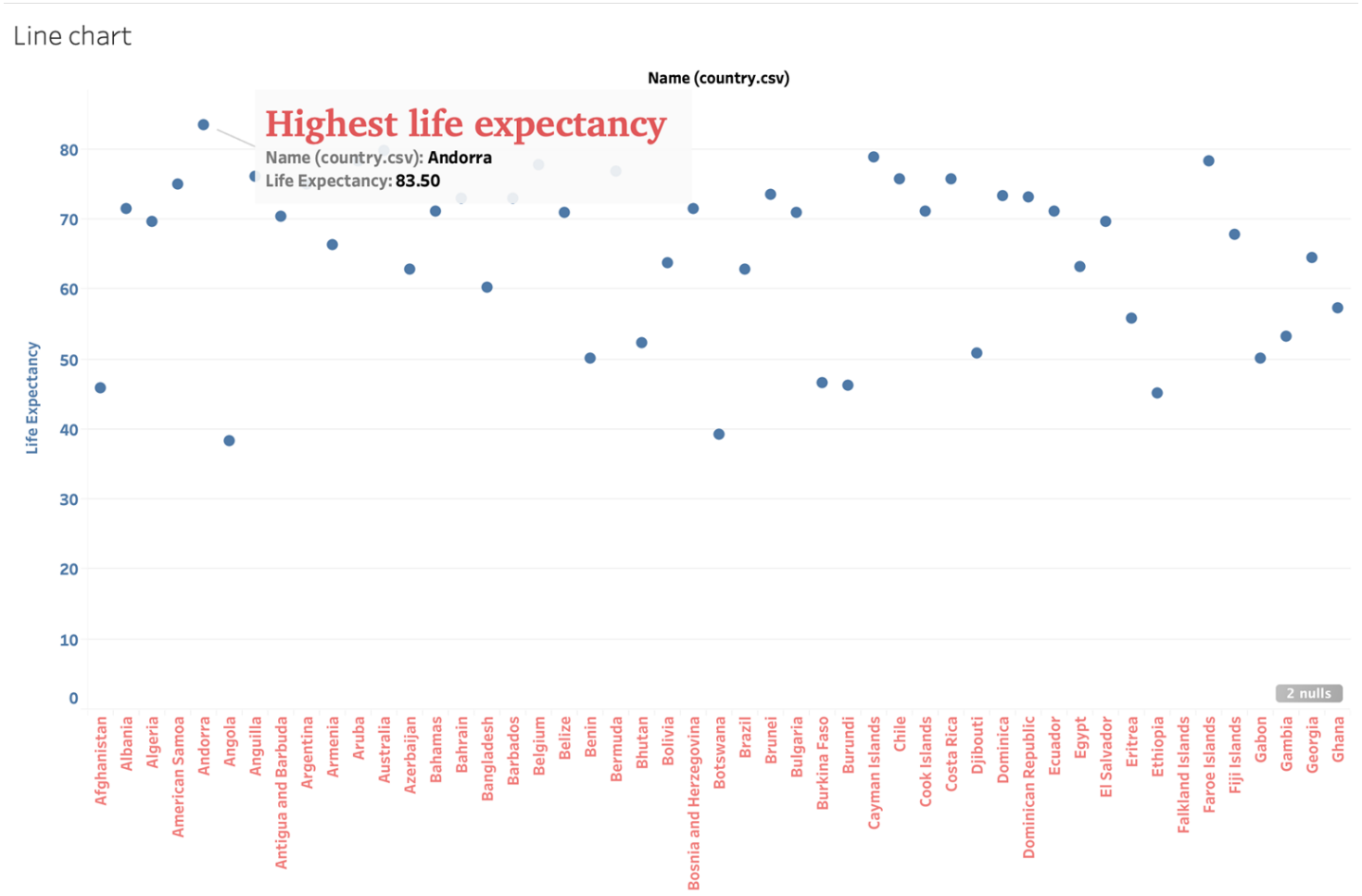
- Let's use the Marks card to convert the bar chart to a **point chart**
- Then, add an annotation to the country with the **highest life expectancy** by right-clicking on the point with highest life expectancy
  - Add a call-out by annotating with a mark
  - Add a custom header
  - Customize the data that is shown in the annotation



Formatting : Annotate Marks

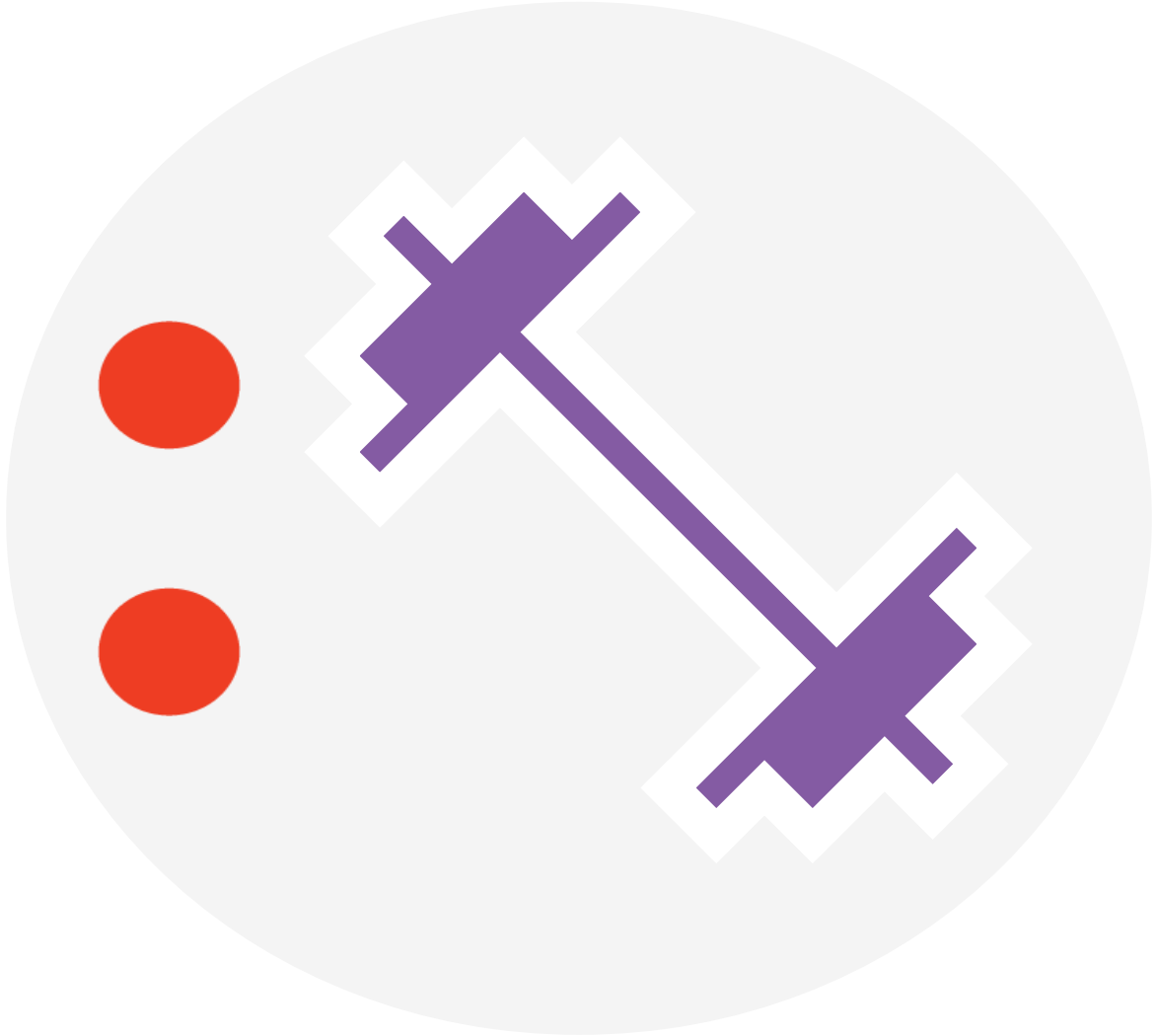
- Finally, add some **grid lines** to the figure space





Knowledge check 1

Exercise 1



Agenda

- Discuss filtering and formatting capabilities in Tableau
- Explain the concept of functions
- Implement basic functions on the dataset
- Introduce and implement table calculations

Tableau functions

- Tableau's **built-in functions** are essential for being able to customize your data representation
- Just as we spent time learning Excel commands and SQL functions, we will need to get familiar with Tableau's syntax
- For a complete listing of Tableau's function repertoire, you can have a look at the **documentation** on the tool's website

Function types

- Just like SQL functions, Tableau functions are classified into several types
- These include:

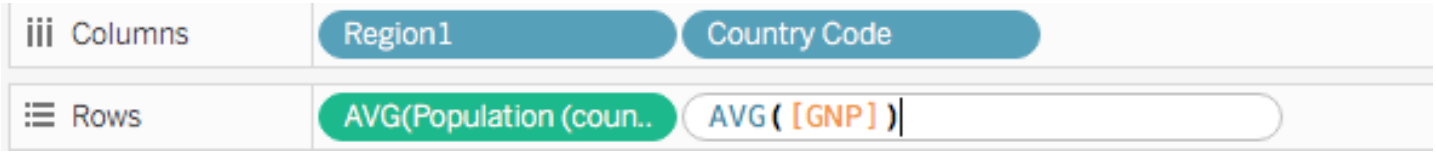
- Number Functions
- String Functions
- Date Functions
- Type Conversion
- Logical Functions
- Aggregate Functions

- Pass-Through Functions (RAWSQL)
- User Functions
- Table Calculation Functions
- Spatial Functions
- Additional Functions

- To see Tableau functions separated by type, visit **this page**

Creating a function: Option 1

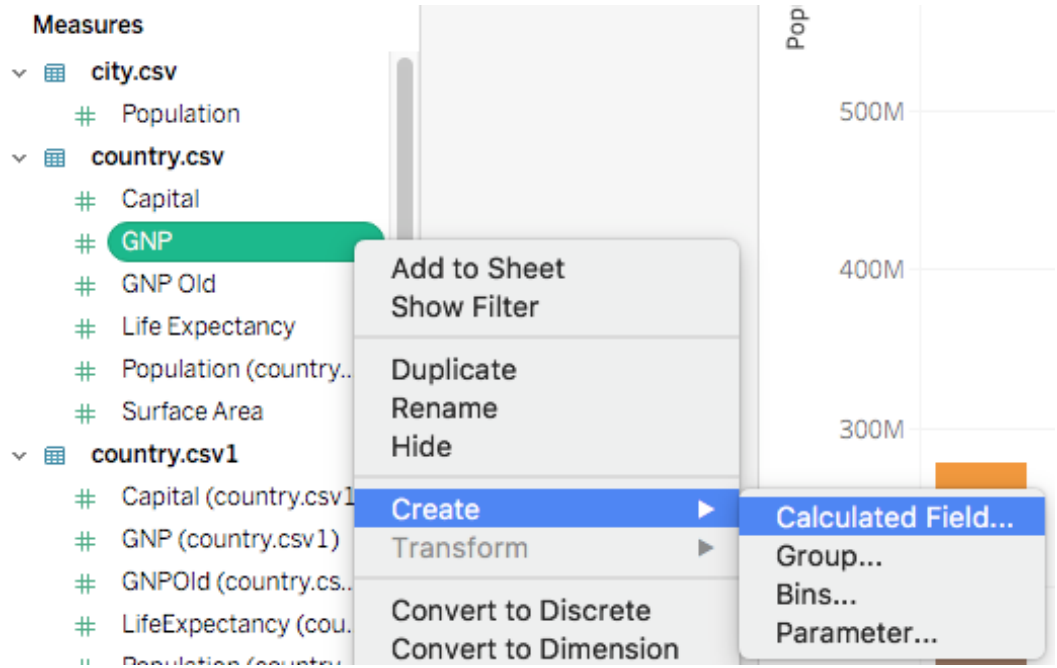
- There are two ways to create functions in Tableau
- **Option 1:** Specify directly in the **"shelf"**



- Useful for simple calculations
- This creates data in the view without adding a new column
- **Note:** You can create a new column by dragging it to the data pane

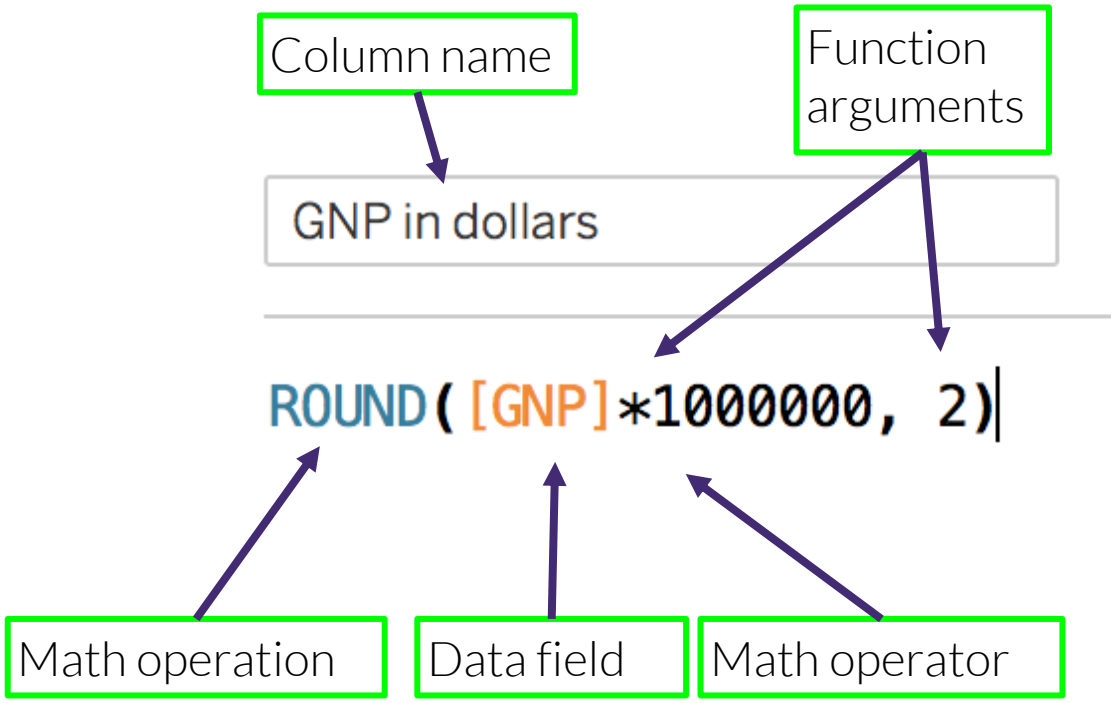
Creating a function: Option 2

- **Option 2:** Specify with the **menu**
  - Useful for more complex calculations
  - Creates a new column in data



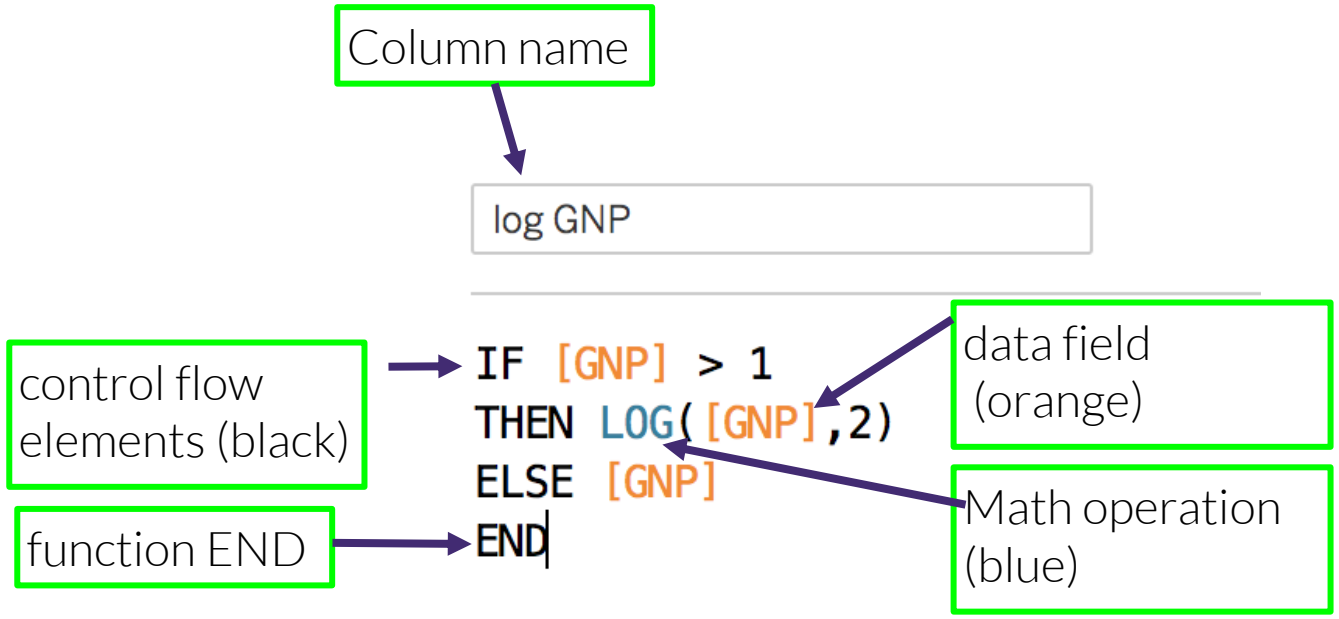
Basic function syntax

- **Function parts:**
  - Built in functions
    - control flow **(black)**
    - math operations (=,-,/,*...)
    - logical operators (<, >, =)
    - math functions **(blue)**
  - Fields from data pane **(orange)**
  - Strings (demarcated by “ ”)
  - Integers



Control flow using function syntax

- This is a simple function that gets the log of the GNP data column
  - Base 2 log for values > 1
  - Otherwise it takes the value

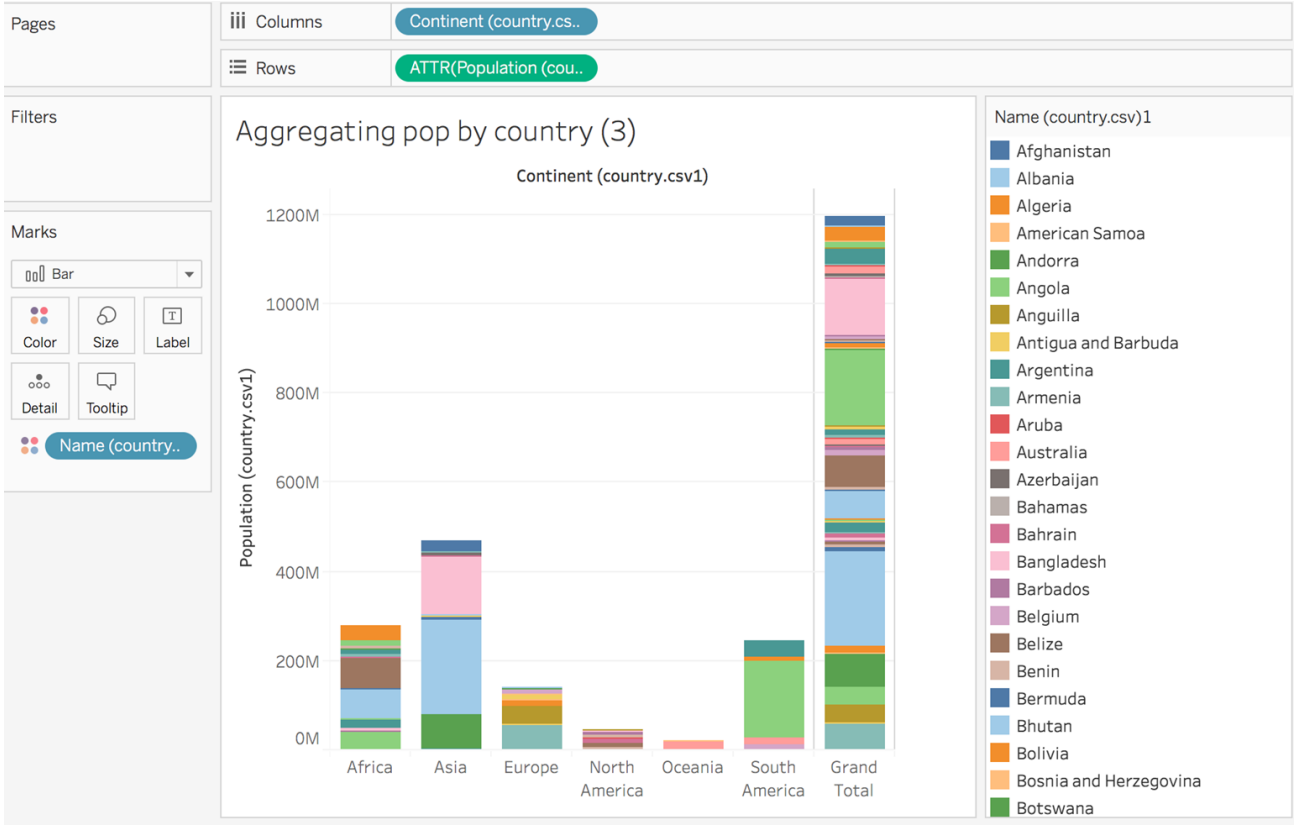


Agenda

- Discuss filtering and formatting capabilities in Tableau
- Explain the concept of functions
- Implement basic functions on the dataset
- Introduce and implement table calculations

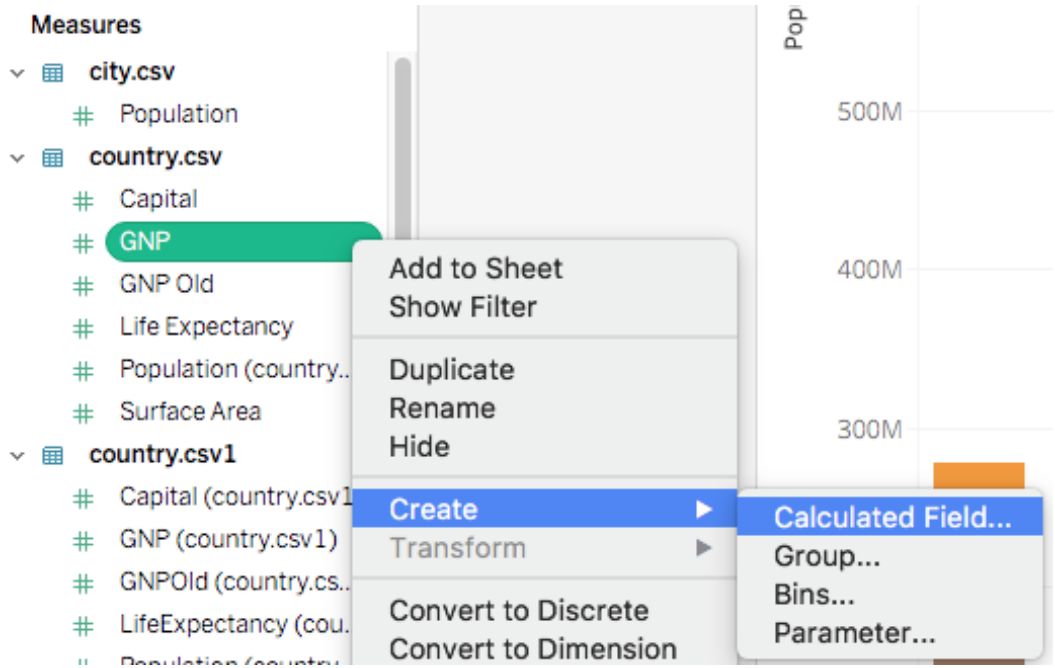
Create a basic population chart

- Before creating and implementing our first function, let's start by creating a **continent-level population analysis graph**



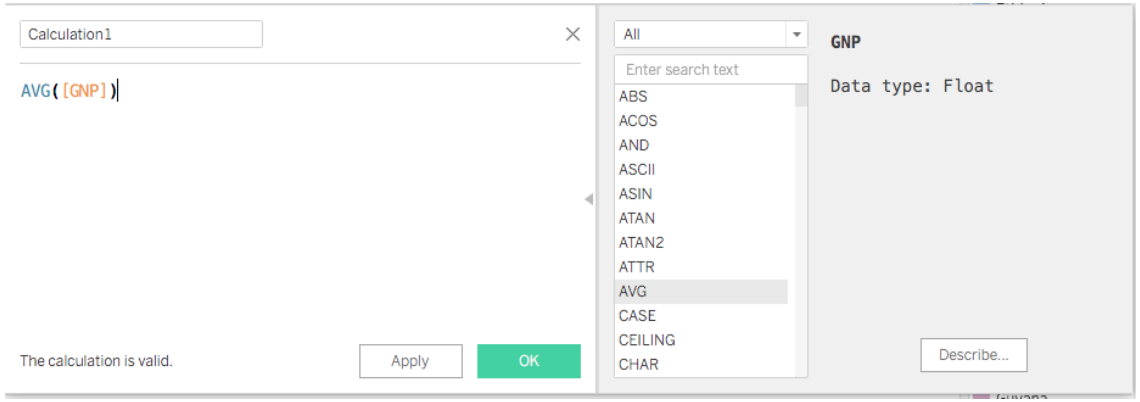
Creating a function

- We will now create our first function in Tableau using **GNP**
- Right click on the GNP field to get to the calculation pane
  - Create > Calculated Field...

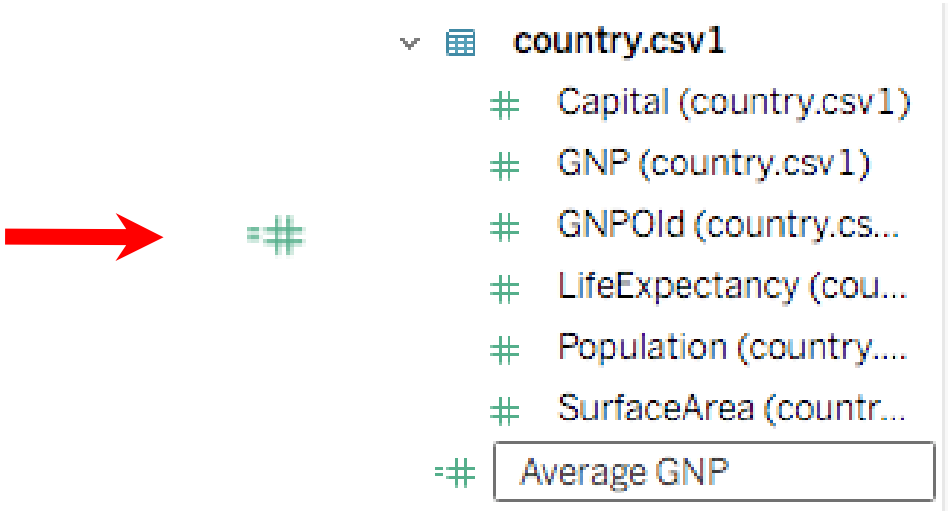


Validating a function

- We're going to calculate the **average GNP** grouped by country
- Input the formula and note if calculation is "valid"

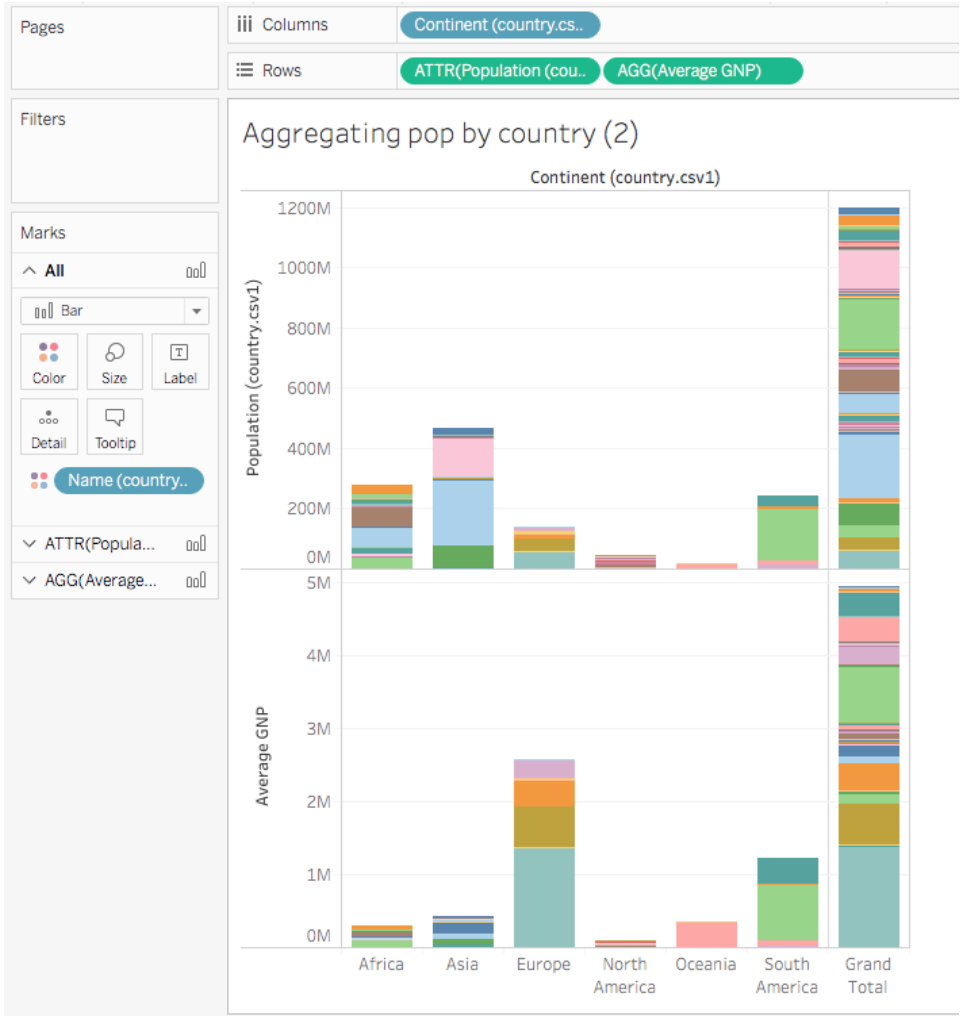


- A new column appears in the data pane with a "=" prefix, which tells us that the function is calculated



Modifying a visualization via function

- Then, drag **average GNP** to the new field into our analysis shelf (notice it auto aggregates at the continent level!)
- Average GNP now appears as a separate pane within our visualization



Agenda

- Discuss filtering and formatting capabilities in Tableau
- Explain the concept of functions
- Implement basic functions on the dataset
- Introduce and implement table calculations

Preview: table calculations

- **Table calculations** give you calculations derived from the **view**, such as:
    - Rank
    - Percentile
    - Moving average
    - Difference
    - Running total
- These calculations do not consider any measures or dimensions that are filtered out of the visualization
- Today, we'll learn how to implement table calculations on our dataset

Video: table calculations

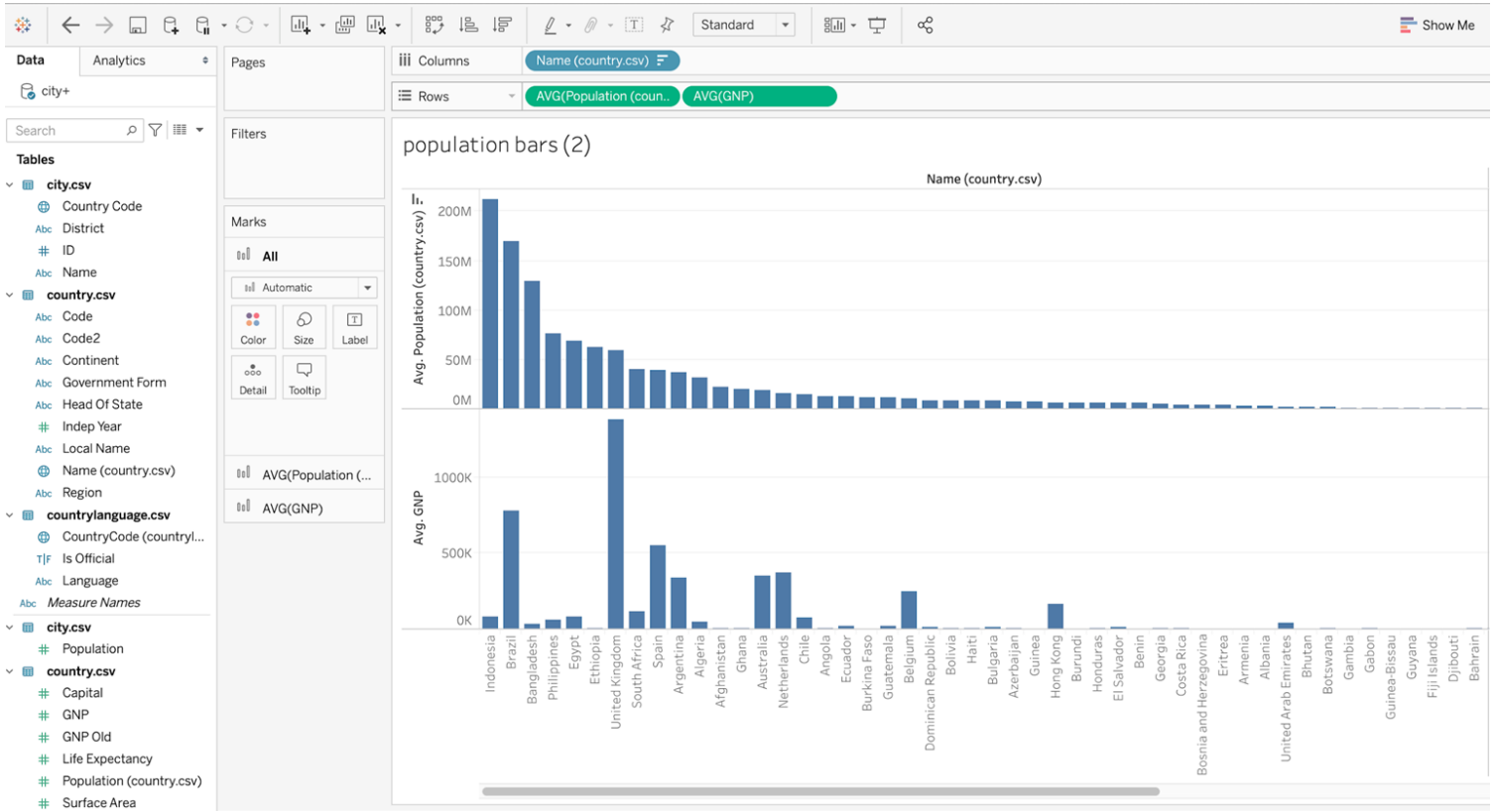
- **Table calculations** are an extremely useful way to perform basic operations on your dataset within Tableau

- Before we practice using them together, click the link below and watch a short video from Tableau walking through using both **pre-built (Quick)** and **custom** table calculations

https://help.tableau.com/current/pro/desktop/en-us/calculations_tablecalculations.htm

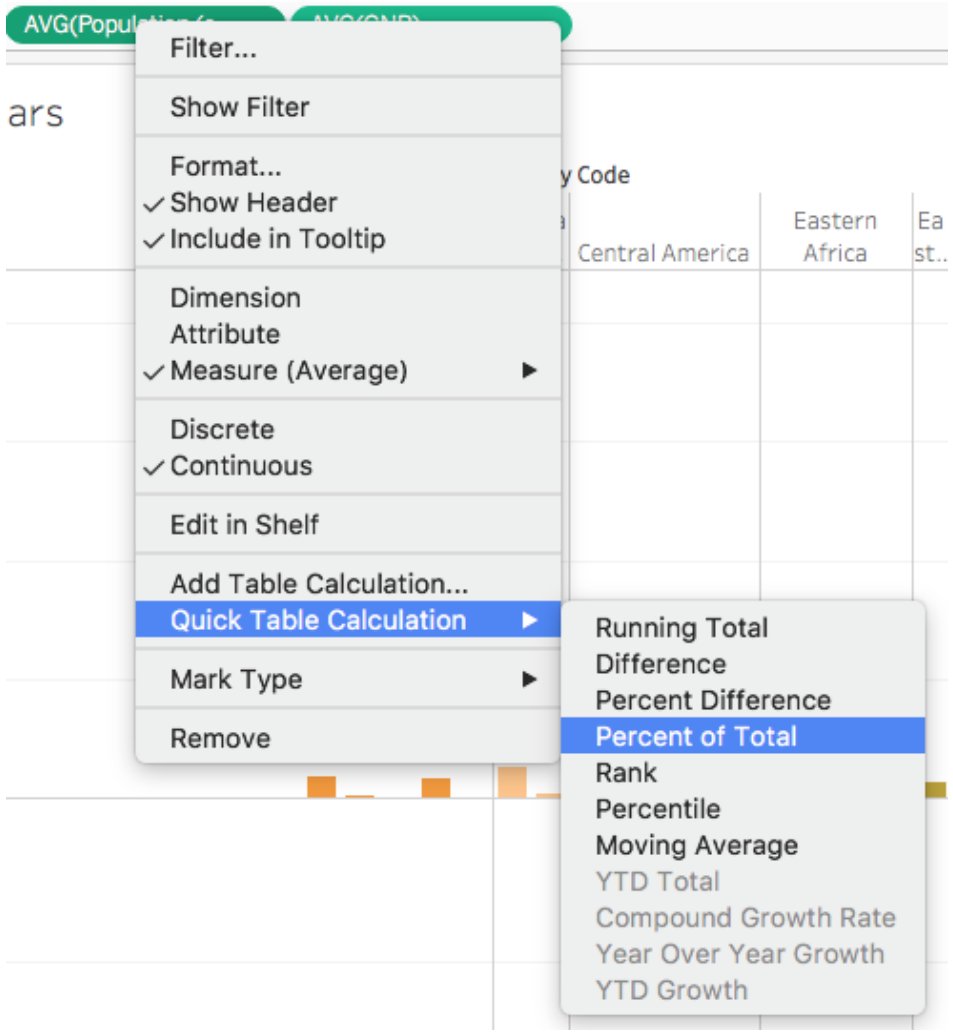
Create population chart by country

- Drag the **Name** field from the country table and drop it in the Columns shelf
- Drag and drop the **average population and average GNP** fields into the Rows shelf



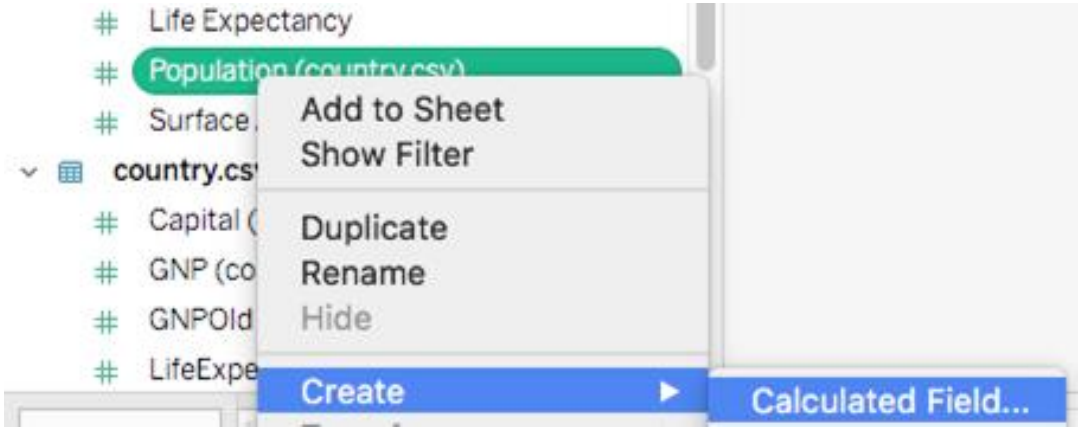
Create a table function (Option 1)

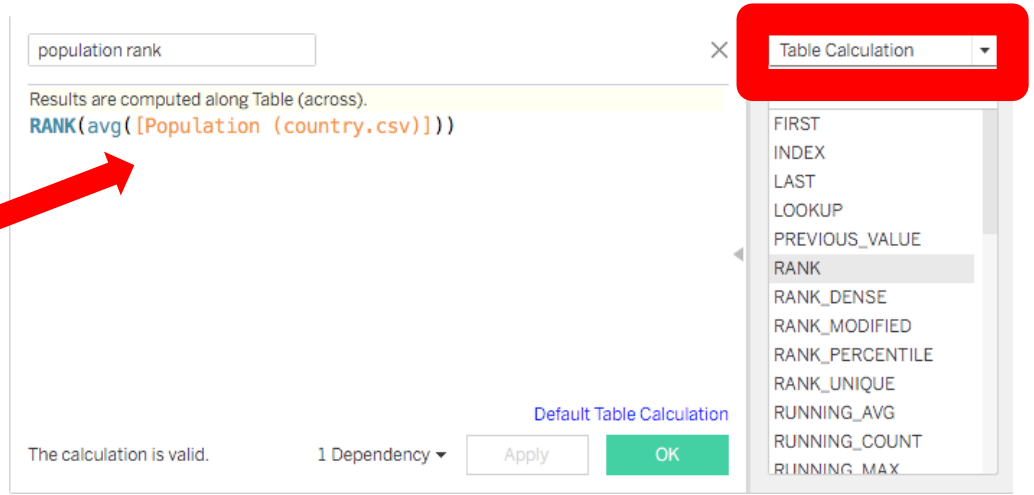
- There are two ways to implement table calculations in Tableau
- **Option 1:** Quick table calculations
  - Right click on the appropriate pill and get a menu of quick calculations
  - Note that the options change based on the selected element



Create a table function (Option 2)

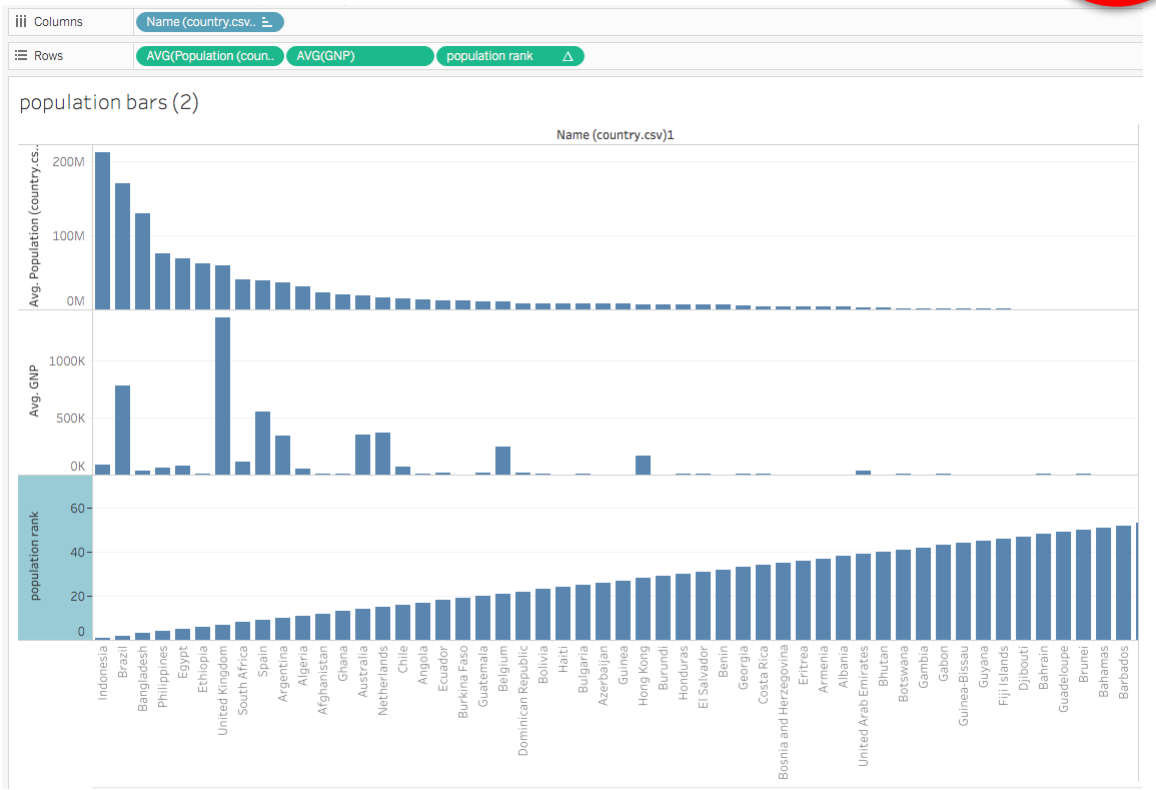
- **Option 2:** Manual calculations
  - Right click a Measure > Create > Calculated Field
  - From calculations dropdown menu at top right, select "Table Calculation"
  - Write formula for required calculation
  - Let's calculate a ranked order list of **average population by country**





Adding a table function to the plot

- Let's use the ranked population formula that we just made
- When we drag it to the data shelf, a triangle indicates that this is a table field
- We can add population rank to our population visualization and sort it
- What do you infer from this visualization?

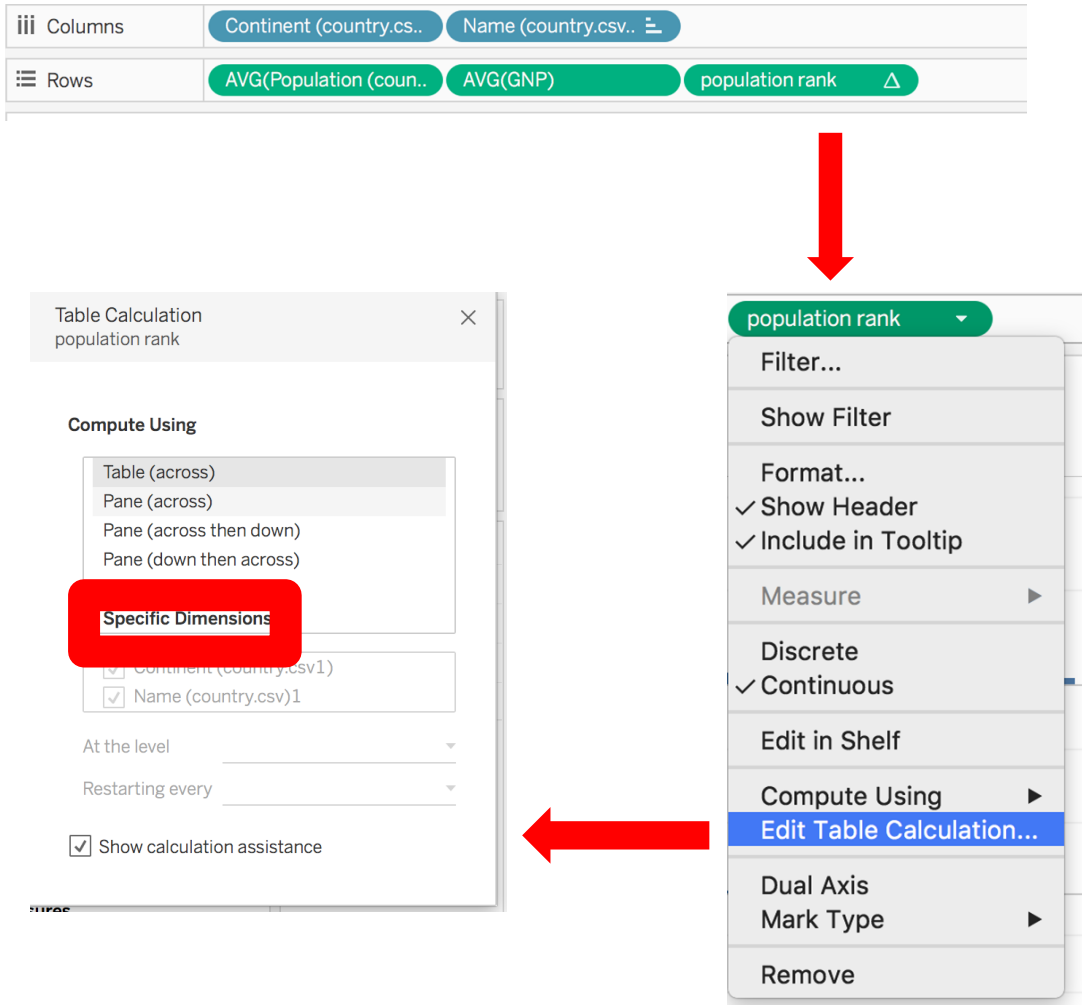


Step 4: Split by continent

- Drag the continent dimension to the columns shelf
- Add continent to the colors mark table
- This will allow us to see how table calculations work

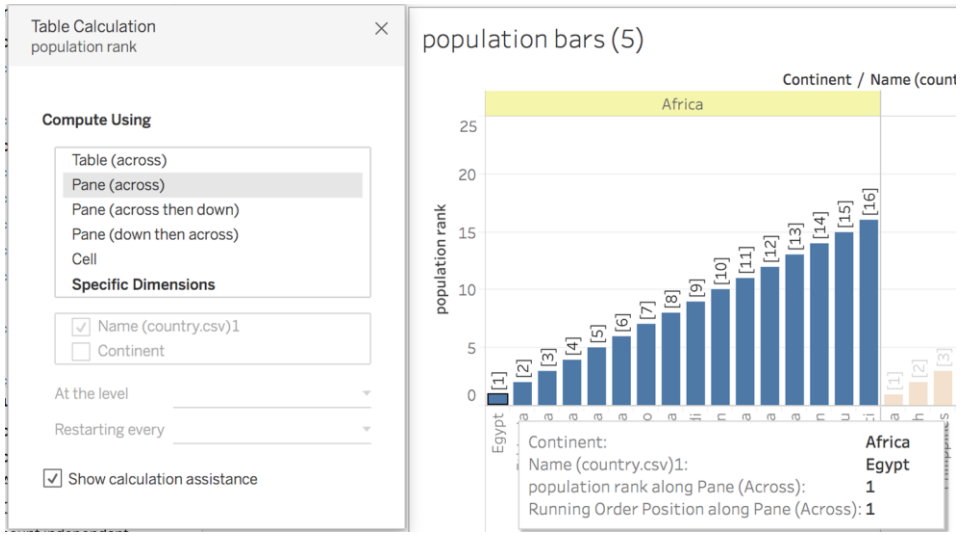
Step 5: Modify the table calculation

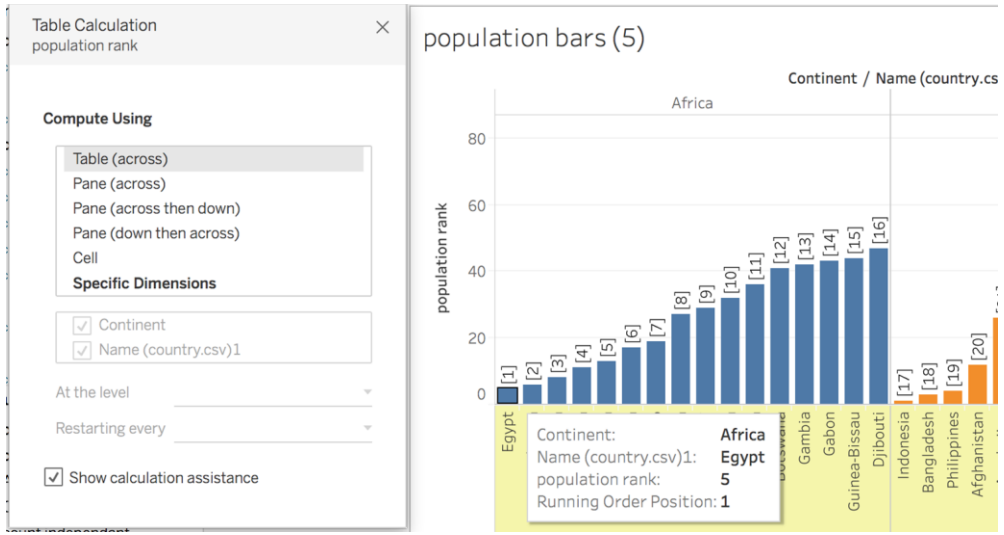
- Our view now uses color to separate the dataset by continent
- Click the **population rank pill** with the triangle
- Look at "Compute Using", then go to "Edit Table Calculation" to see a live demo
- Toggle between two **"Specific Dimensions"** options: **checking** and **unchecking Continent**



How table functions affect data

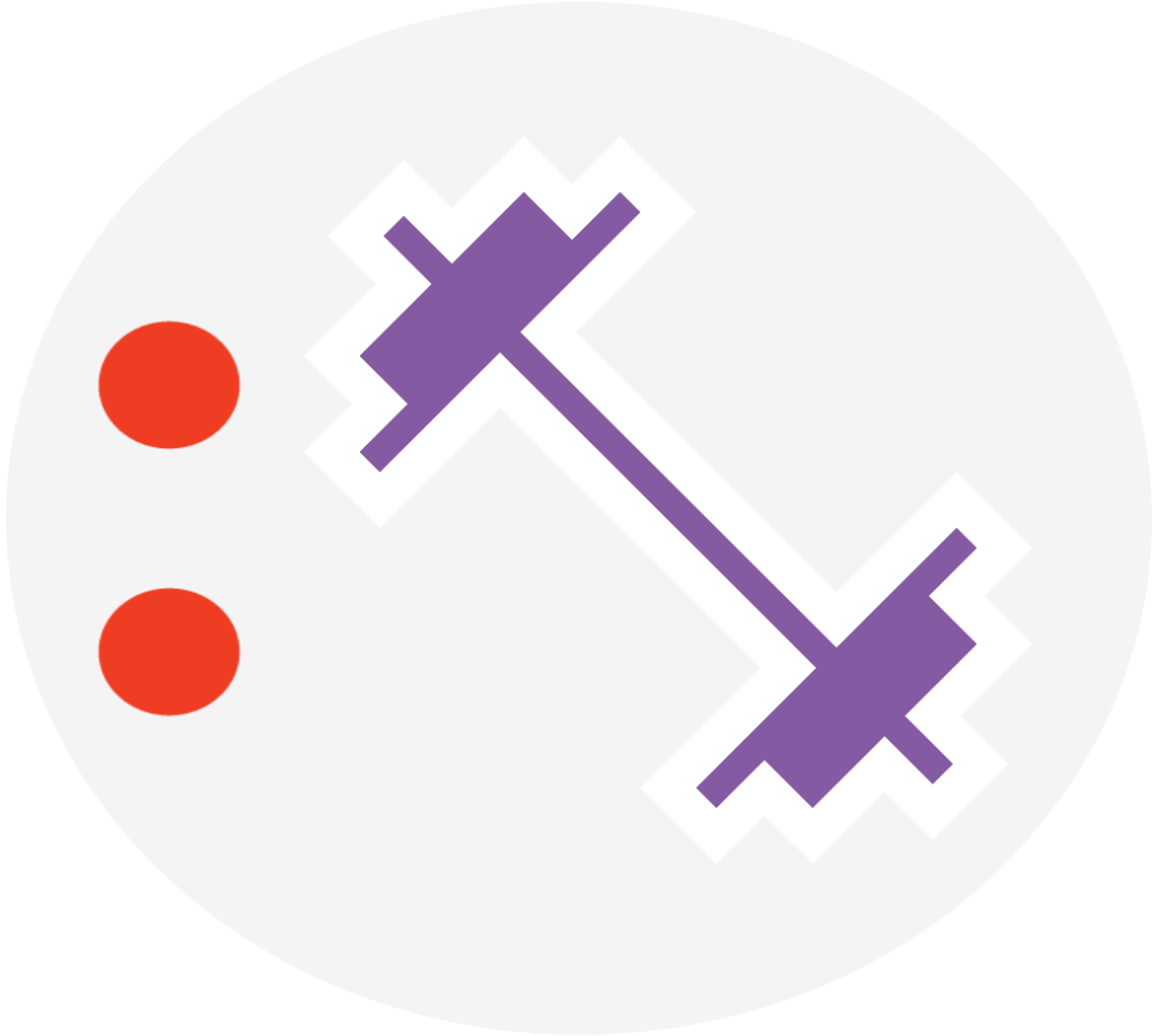
- The difference in this case is subtle, but you can see it if you focus on **Egypt**
  - Within Africa (pane), Egypt's population is ranked **#1**
  - Across the whole dataset (table), Egypt's population is ranked **#5**
  - The **table function** computes across the length of the entire table





Knowledge check 2

Exercise 2



What we covered today

- Cleaning and focusing with filters
  - Match value
  - Top values
  - Ranges
- Formatting figures
- Tableau functions
  - Basic function syntax
  - Creating a function via shelf
  - Creating a function via menu
  - Implementing a function
- Table calculations

Upcoming module

- In the next module, we will cover:
    - Level of detail (LOD) functions
    - Number functions
    - Aggregate calculations

DATASOCIETY:

Thank you!