DATA SOCIETY

Intro to Tableau

Day 2

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

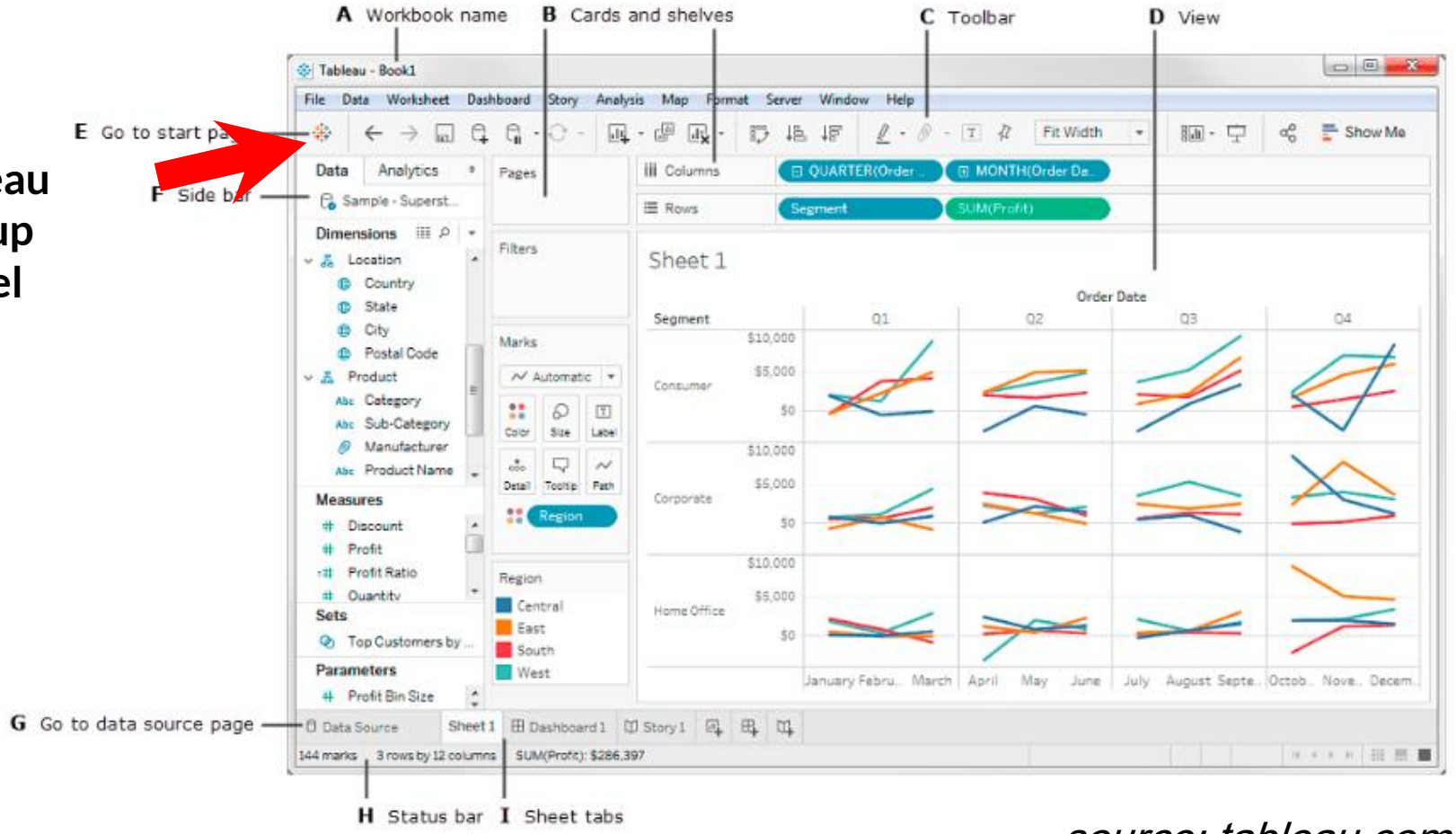
- Albert Einstein

Agenda

- Import the given dataset into Tableau and explain the concept of joins
- Explore the Tableau platform layout
- Create basic visuals using the World Data
- Introduce the concept of aggregating, binning, and grouping

Tableau overview

Click the Tableau logo to bring up Connect panel



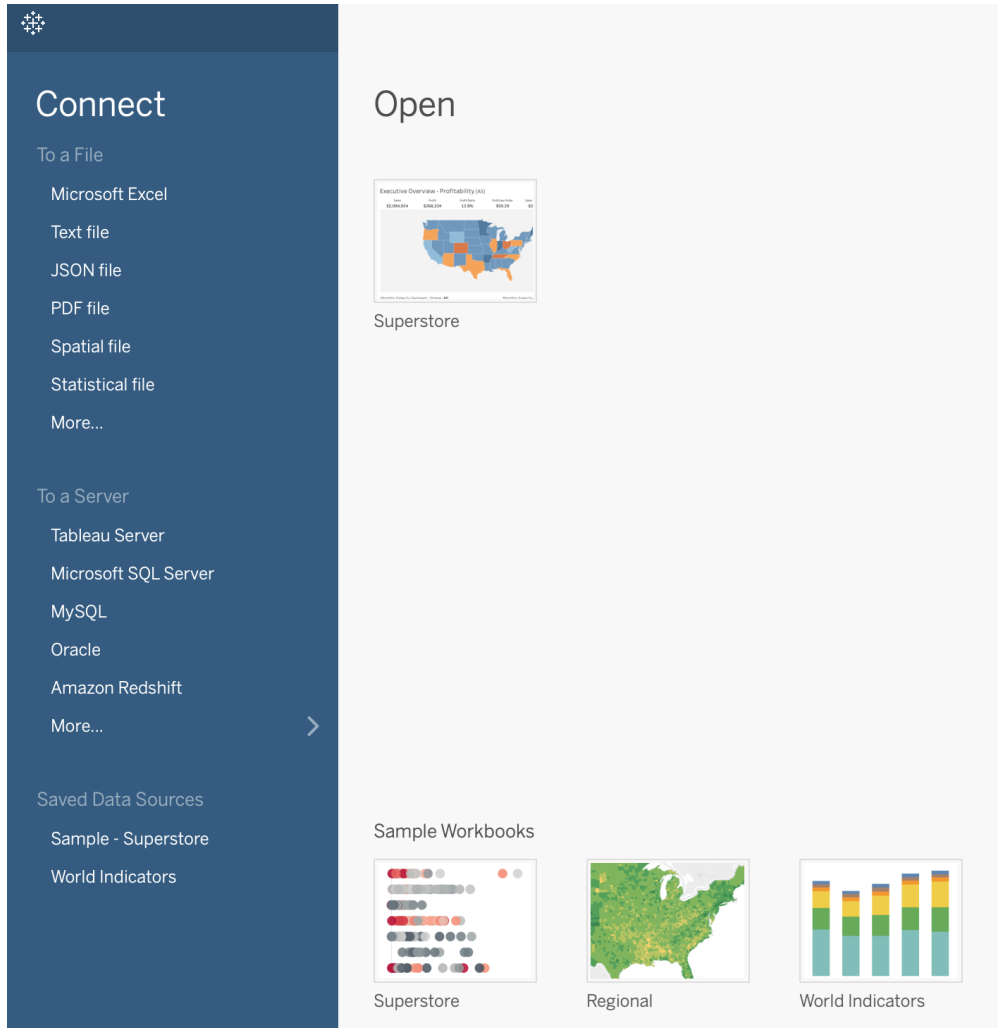
source: tableau.com

Importing data

- Import data with the **Connect** panel

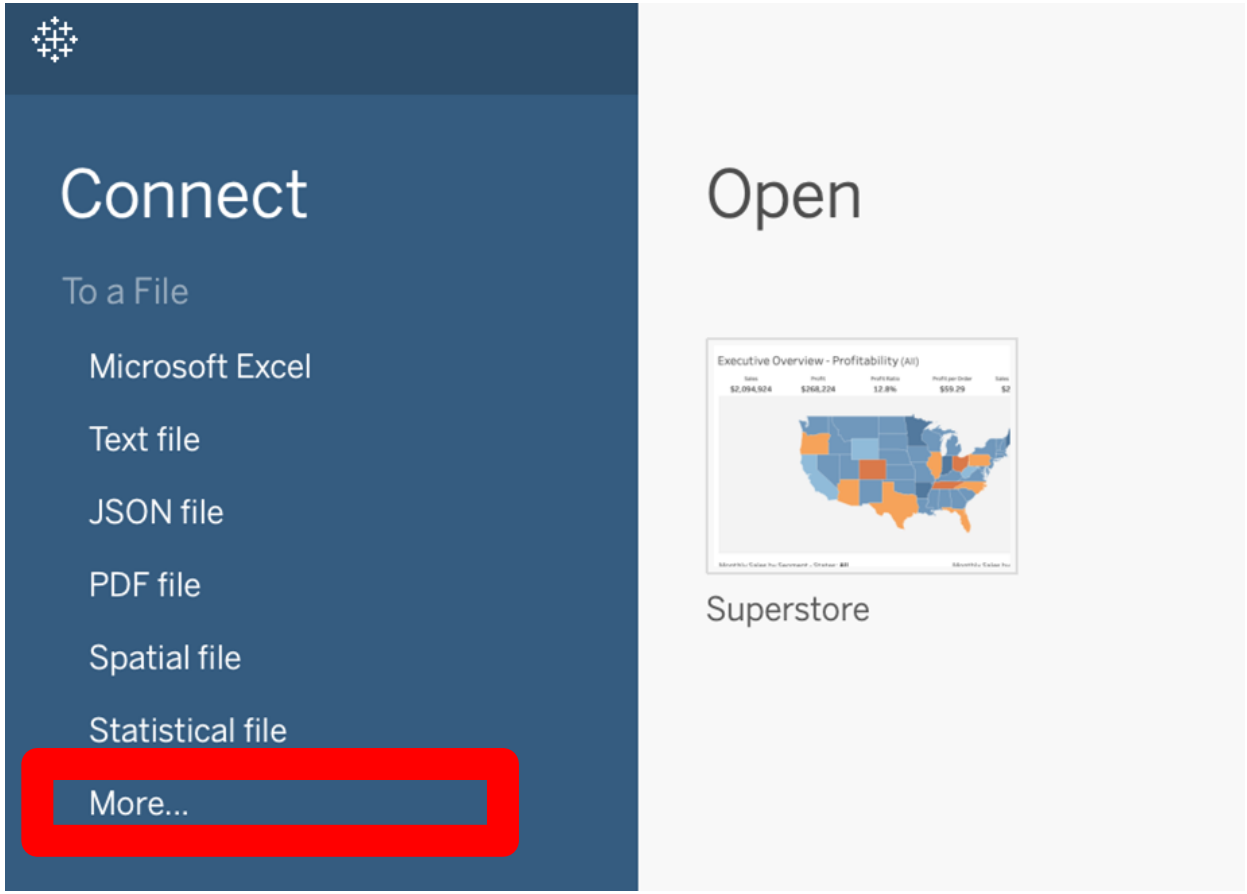
- Supports multiple formats such as:
  - Microsoft Excel (.xlsx)
  - Text (.txt, .csv)
  - JSON (.json)
  - PDF (.pdf)
  - R data format (.RData)

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



Import World Data : CSV

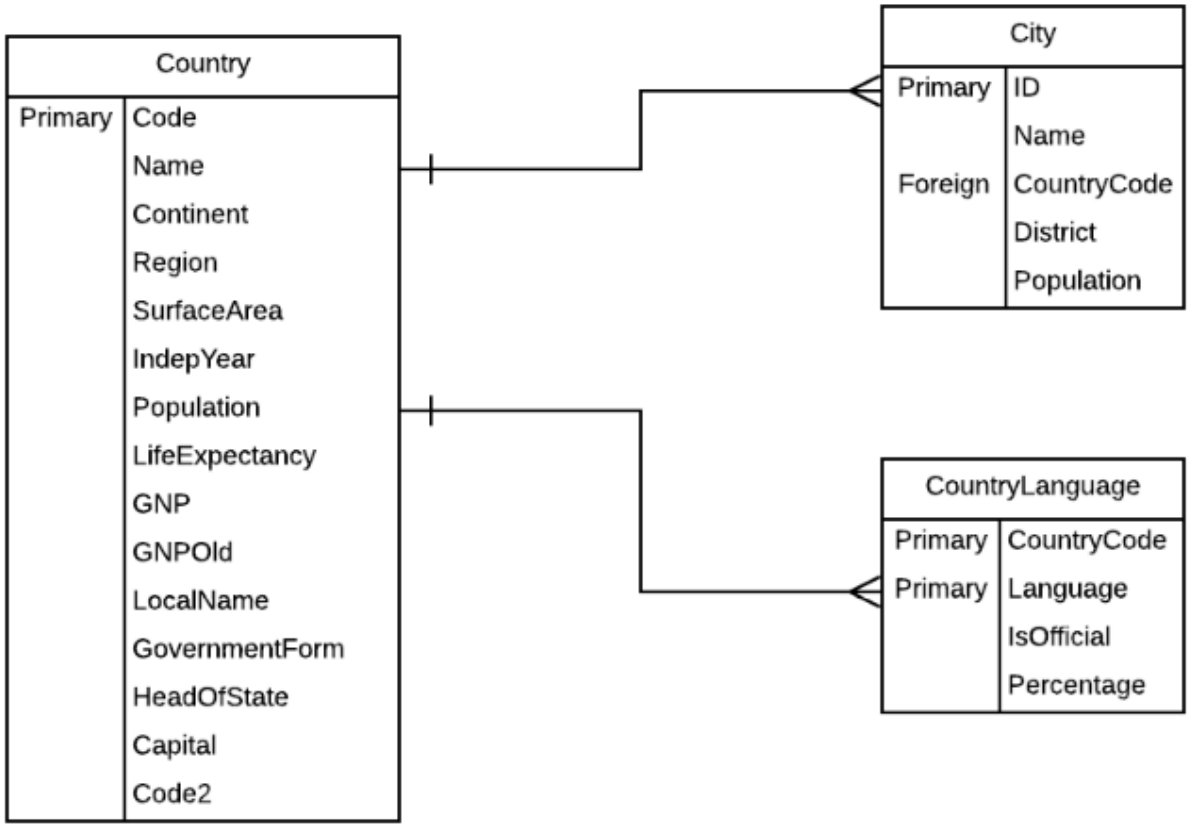
- Let's import some pieces of the world dataset today and see what sort of insights we can reveal
- Click the **"More..."** item to browse your local CSV files



World database

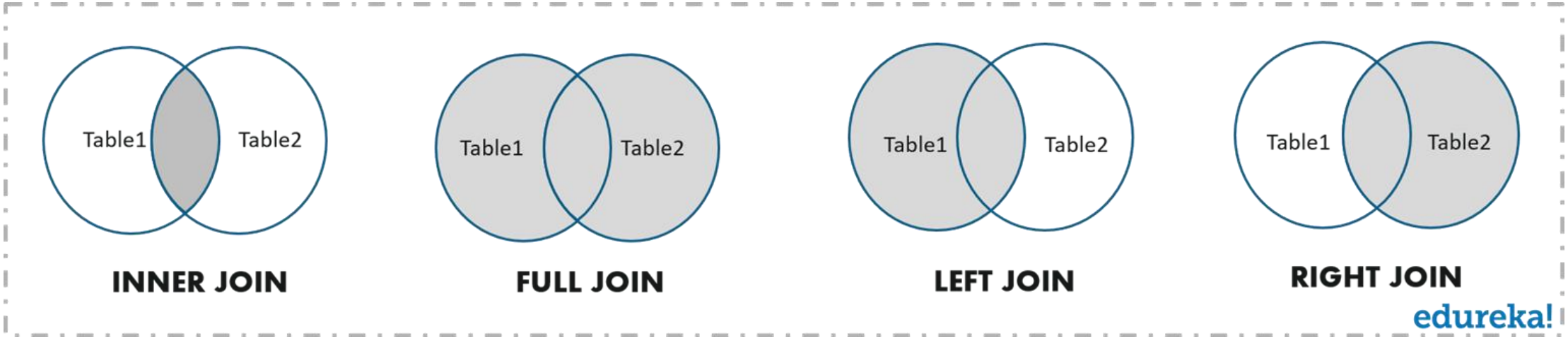
- For now, import the following three CSV files:
  - **country.csv**
  - **city.csv**
  - **countrylanguage.csv**
- We'll use the other CSV files during our Exercises

World Database ERD



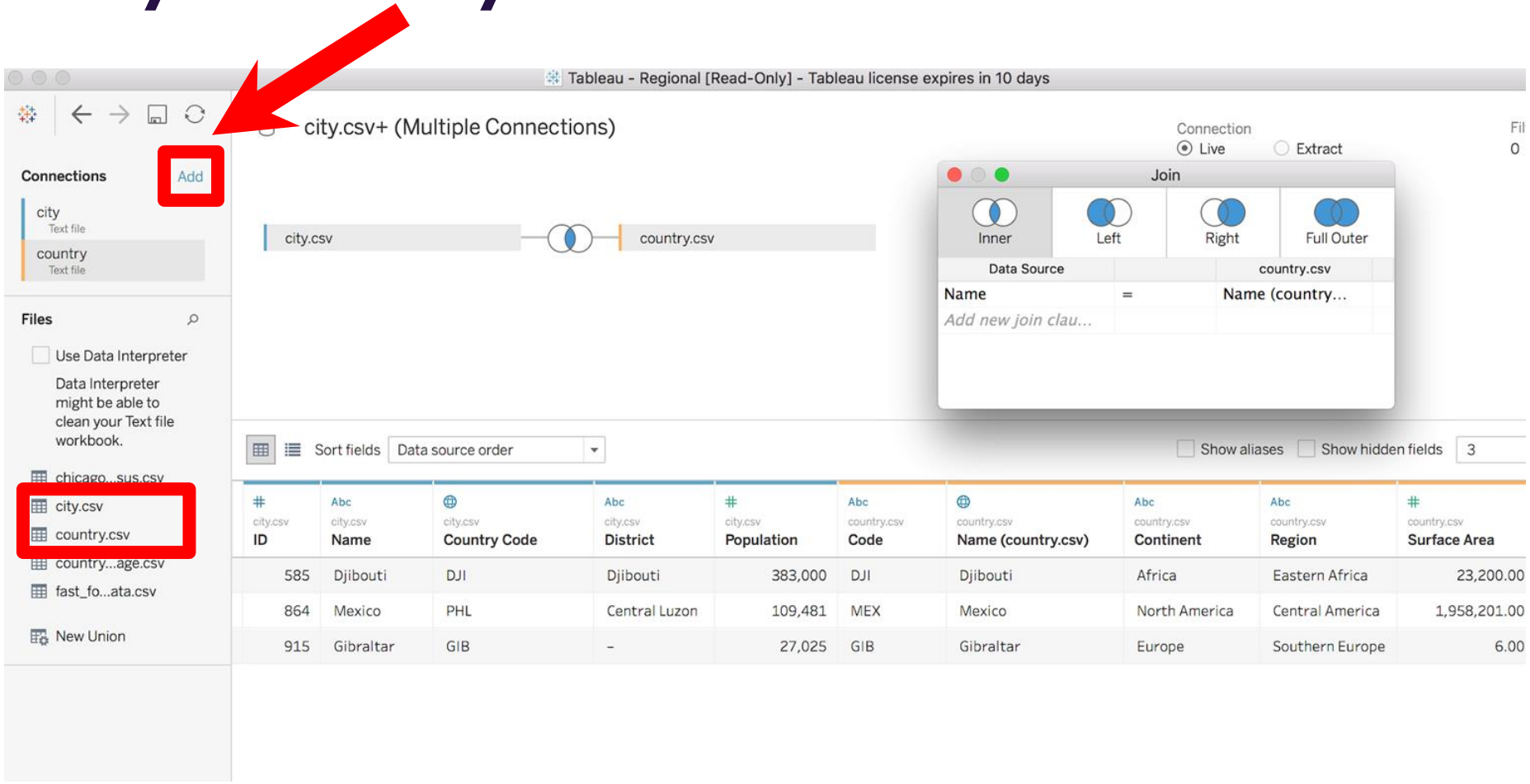
Join the tables

- Recap from SQL - Let's go over **joins** in Tableau
- We need to join the imported tables
- Combining different data sources is a key to effective **data science**



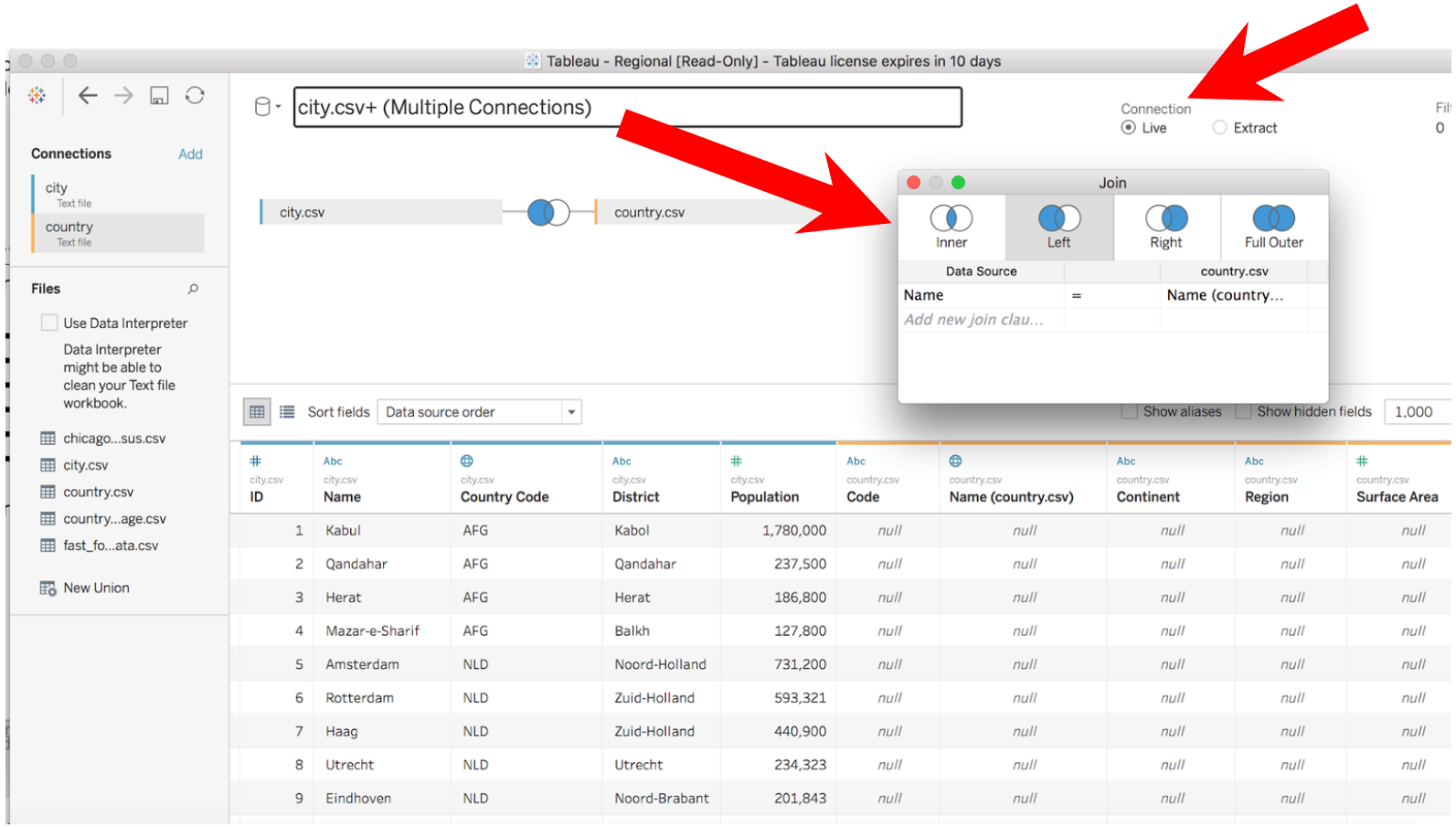
Joining country and city tables

- Let's start by joining the country and city tables
- Use **"Add"** in the Connections pane to see the default auto-join
- You can also drag the table from the **Files** drawer below



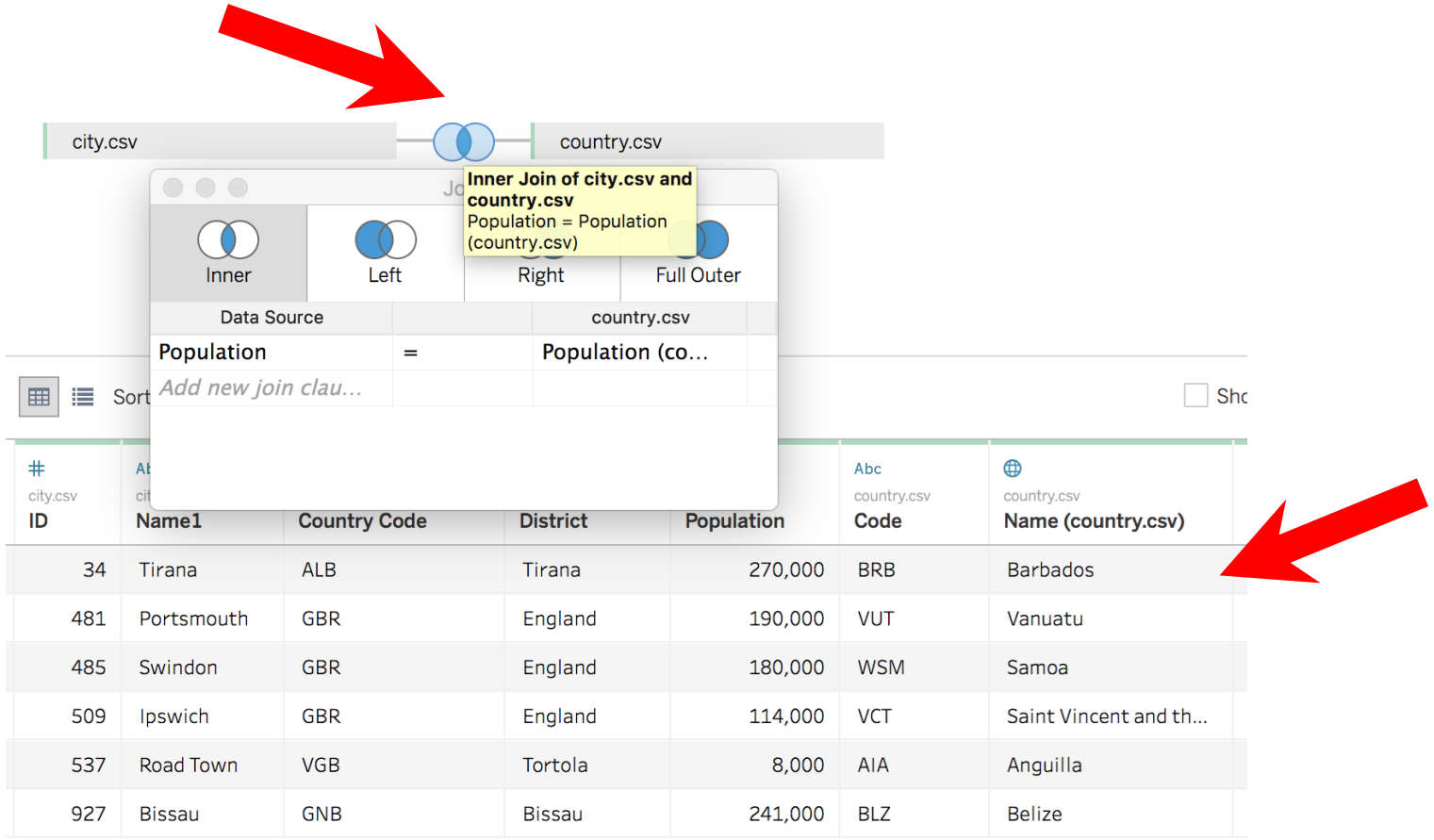
Experiment with joins

- Try out each of the 4 options in the **Join** window
- What do you see when you try each of these options?



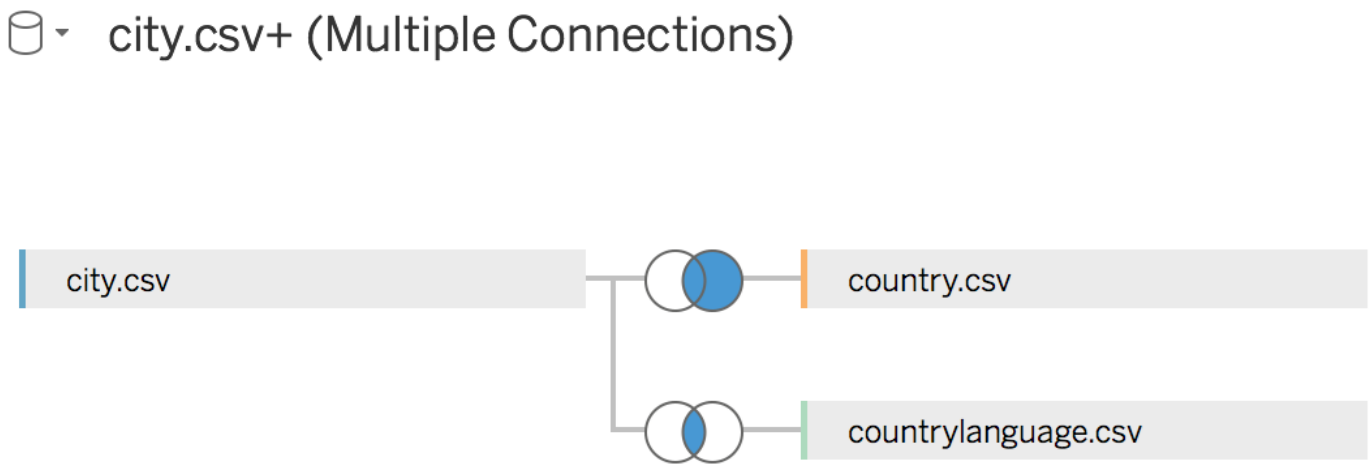
Did the auto-join work?

- 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
  - **Auto-joining sometimes fails**, depending on how Tableau interprets the data!



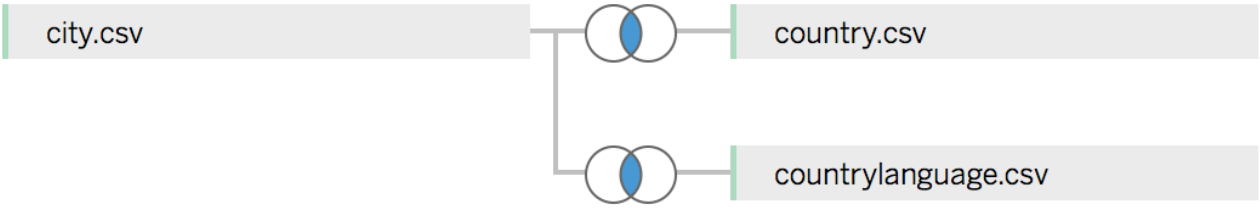
Add the third table

- 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?



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



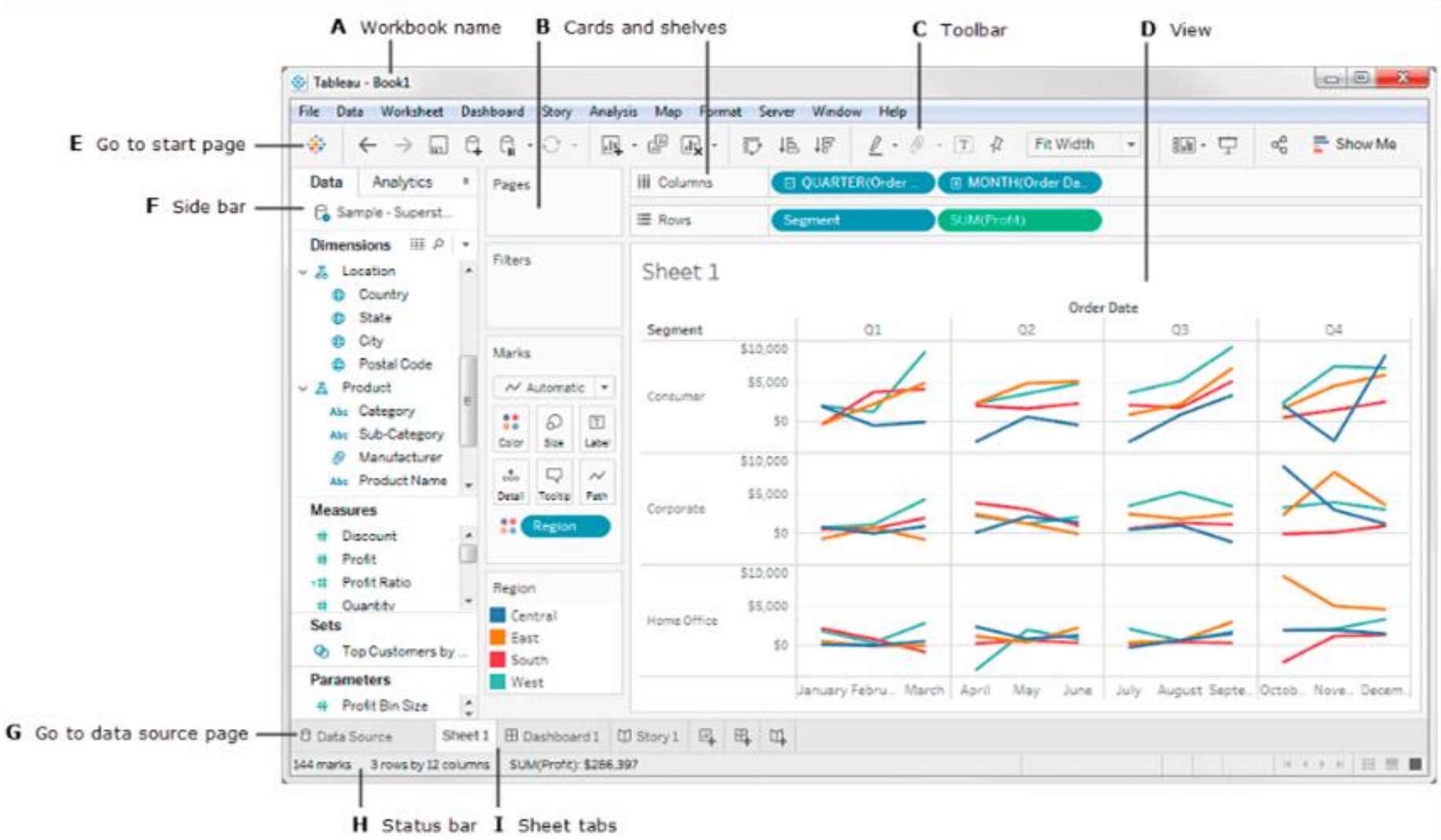
- Why is it okay to join country language to city?

Agenda

- Import the given dataset into Tableau and explain the concept of joins
- Explore the Tableau platform layout
- Create basic visuals using the World Data
- Introduce the concept of aggregating, binning, and grouping

Overview: key parts of Tableau UI

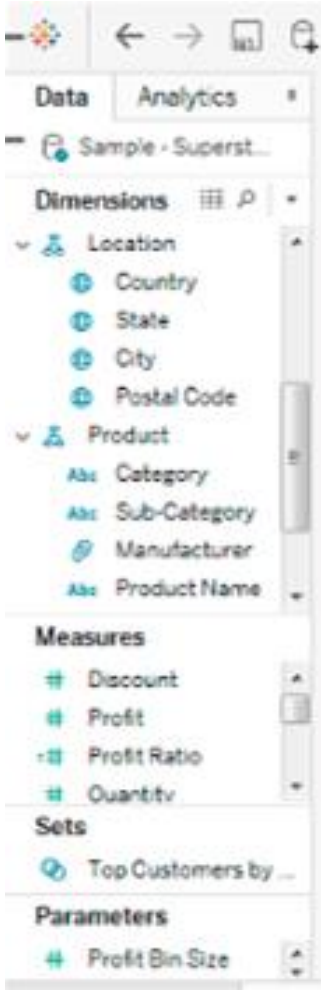
- Start button
- Data and analytics views
- Sheets view
- Marks panel
- Story tab
- Dashboard tab
- Columns and row shelves
- Variable "pills"
- "Show Me" panel

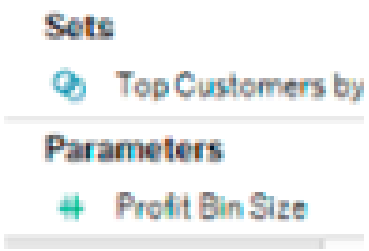


The data tab

- The **data tab** shows several key pieces of information:
  - Dimensions and Measures variables
  - Loaded databases
  - Sets
  - Parameters

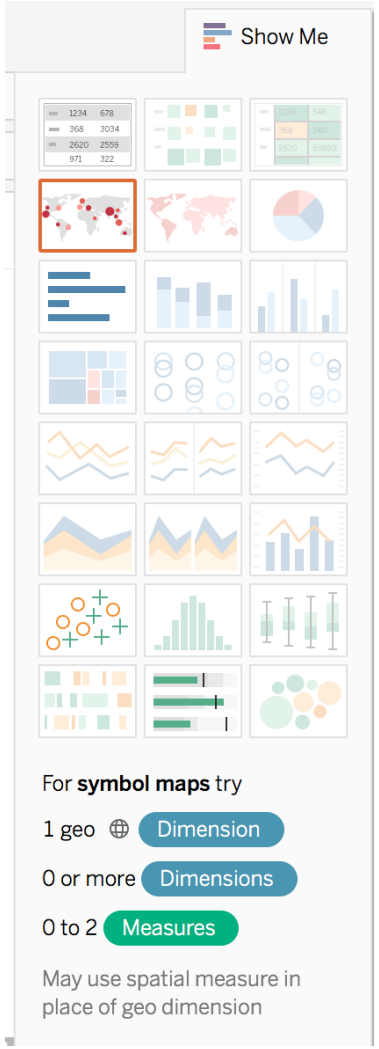
- Note that type of variable is noted to the left of the variable name in the form of an **icon**





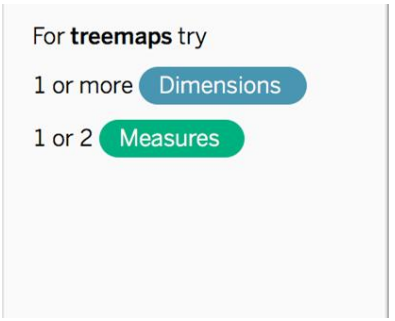
“Show Me” palette

- The “Show Me” palette makes it easy to choose the visualization that you want
- Tableau automatically adjusts dimensions and measures to better fit your data to the map
- It also suggests which visualizations might best suit the data you are working with



Grayed-out “Show Me” options

- Grayed-out visualizations cannot be generated from the given data
- When selecting a grayed-out visualization type, pay attention to the suggestions on the bottom



Knowledge check 1

Exercise 1

Save your work!

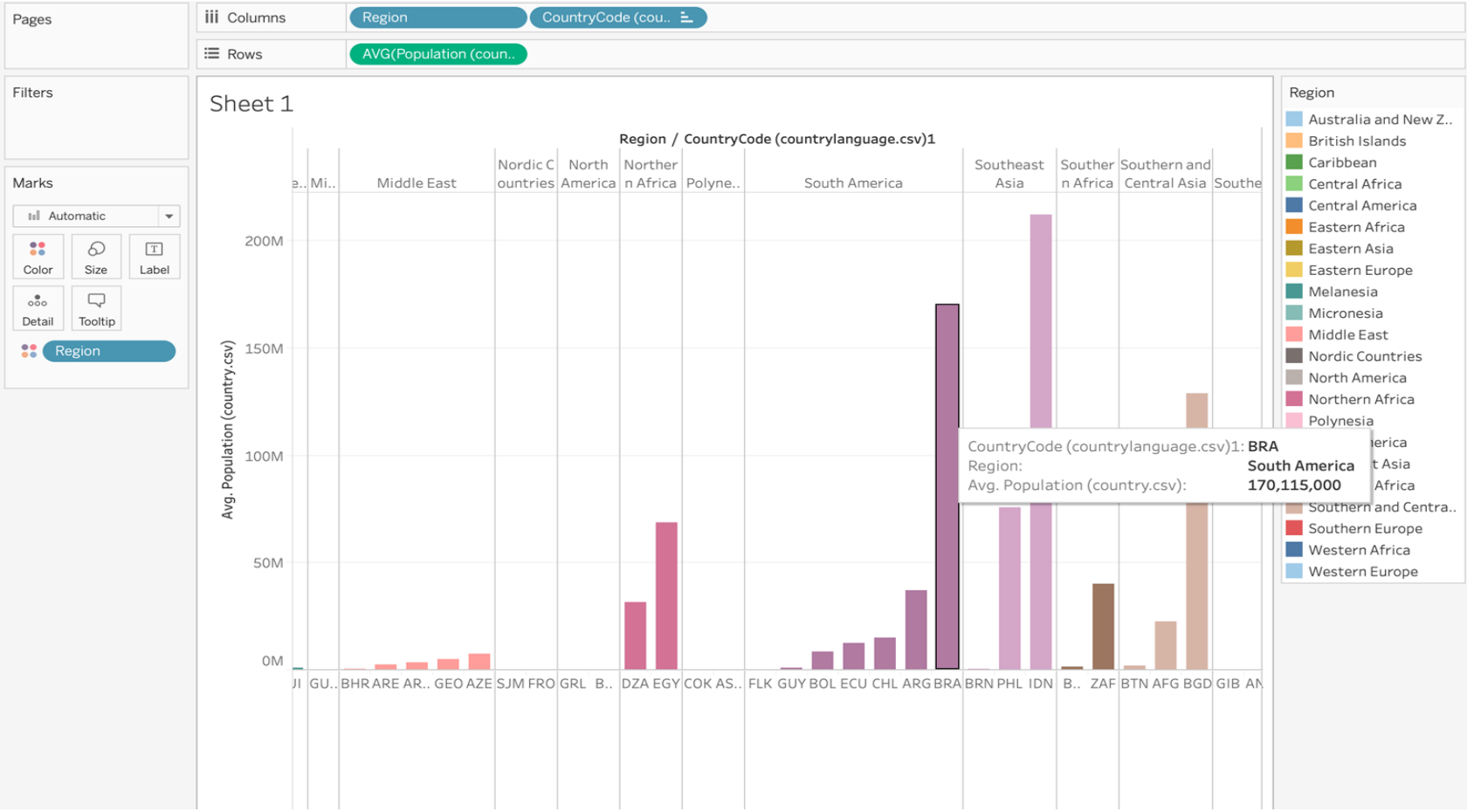
- We will now start creating visualizations in Tableau
- We will see a lot of different insights from the data as we learn more about Tableau
- Make sure to save all your classwork (including exercises) on your local drive!

Agenda

- Import the given dataset into Tableau and explain the concept of joins
- Explore the Tableau platform layout
- Create basic visuals using the World Data
- Introduce the concept of aggregating, binning, and grouping

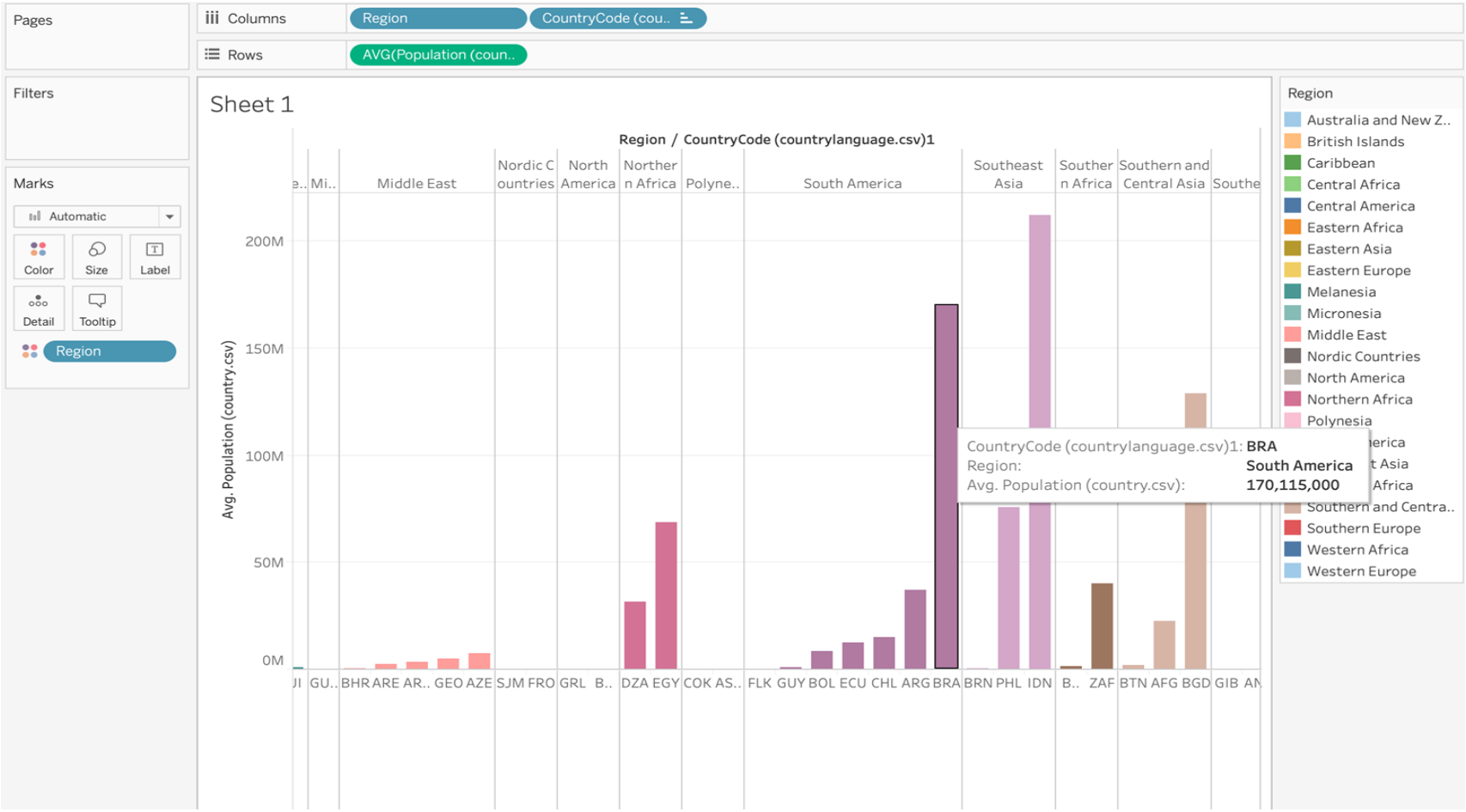
Data visualization: bar charts

- Let's plot average population by country and categorize the bar chart by region



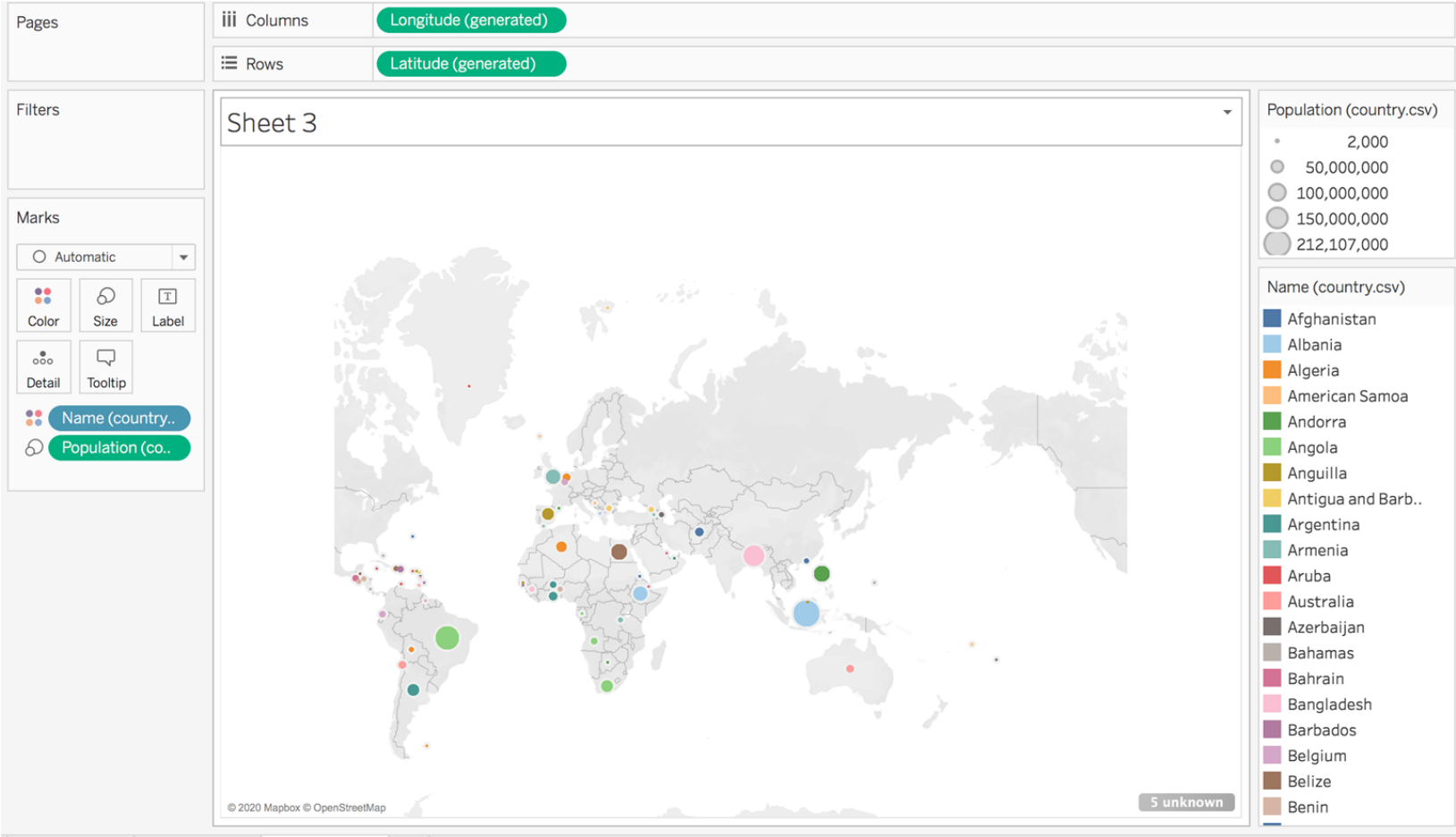
Evaluating our bar chart

- What do you see from the graph?
- Is there anything that you would change?
- What follow on visualizations would you do?



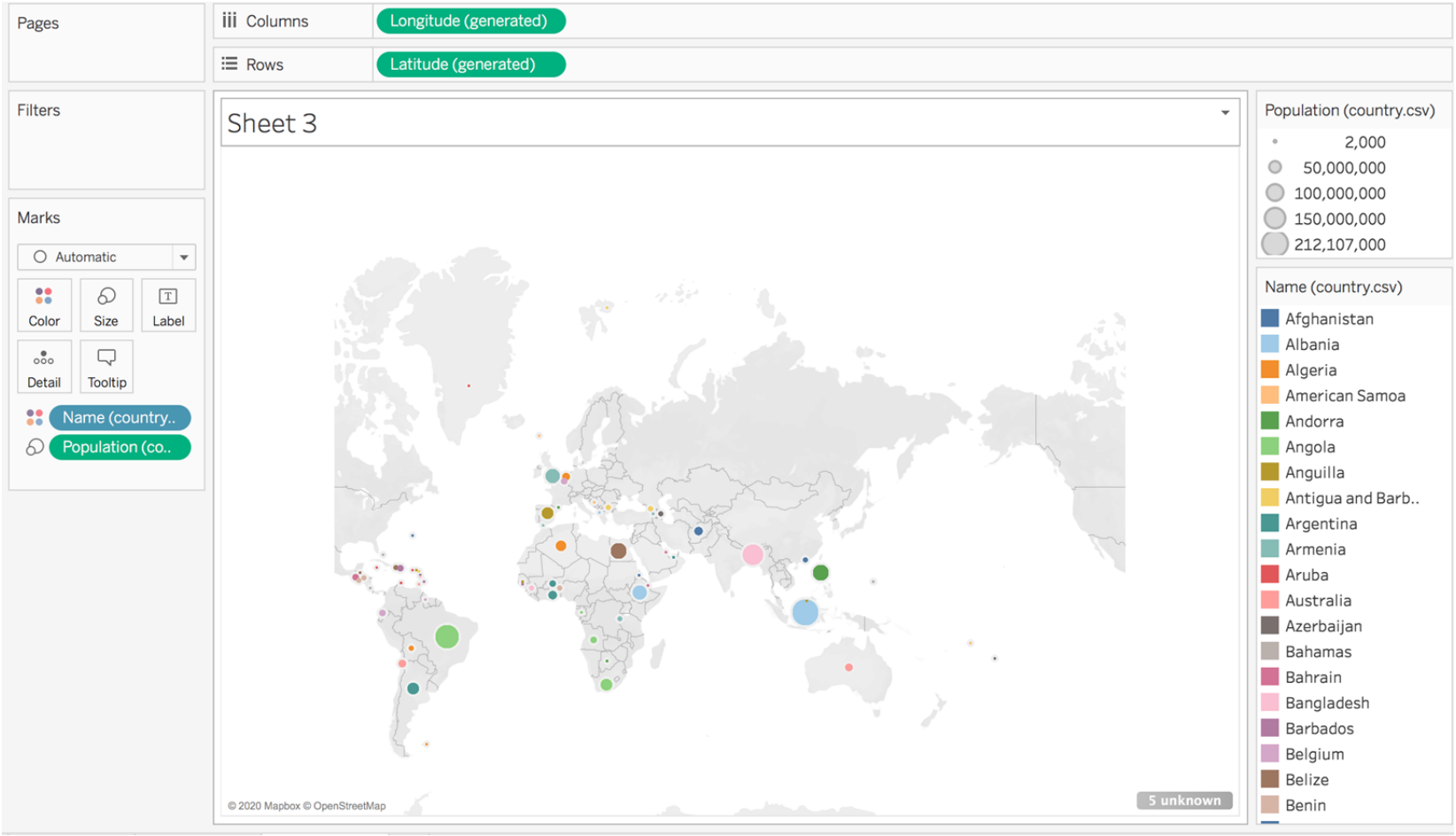
Data visualization: symbol map

- We will now plot the same information on a **map**



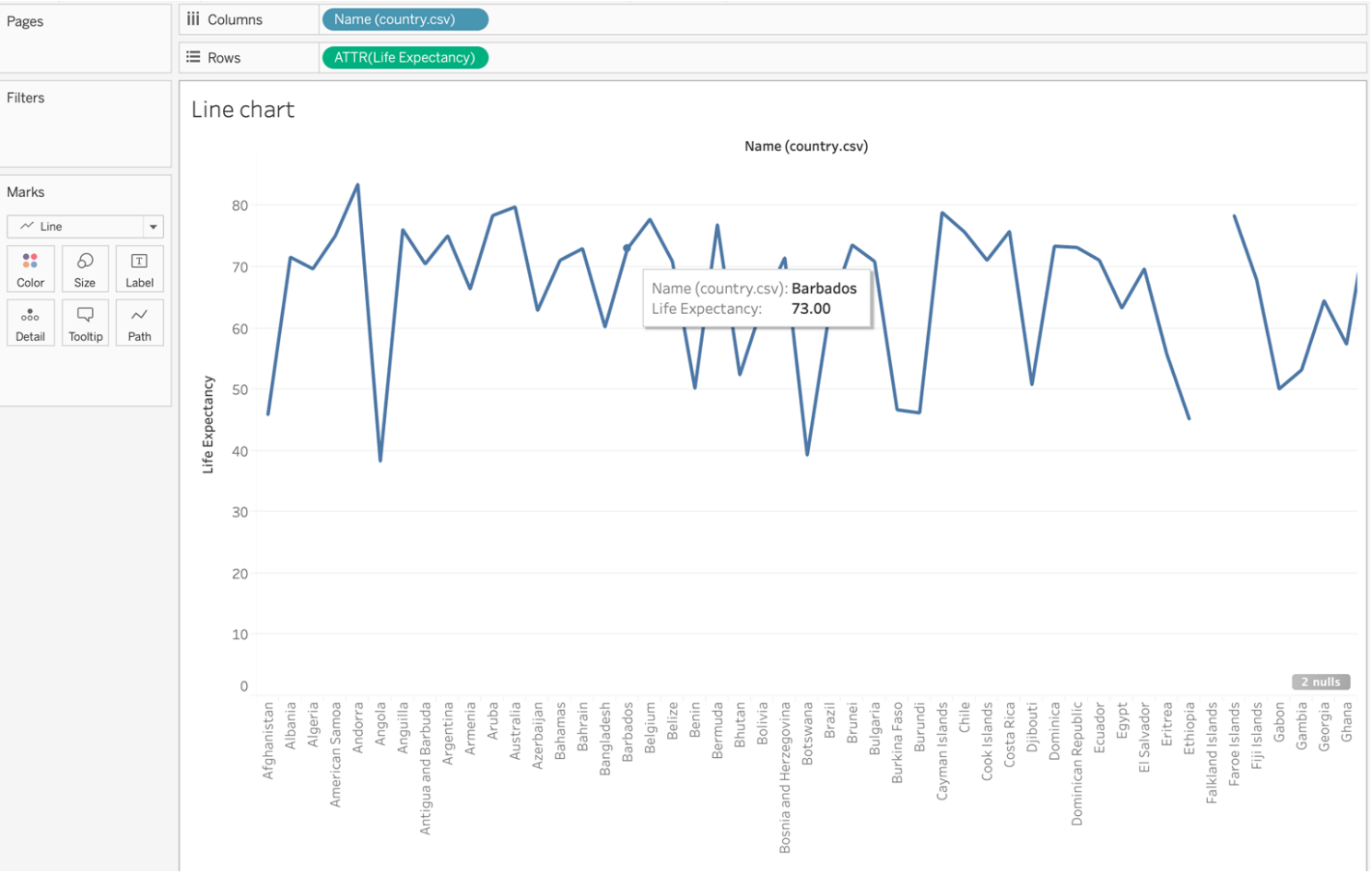
Evaluating our symbol map

- Is this view of the data better? Worse?
- Is there anything that is missing from the data?
- How would you fix it?



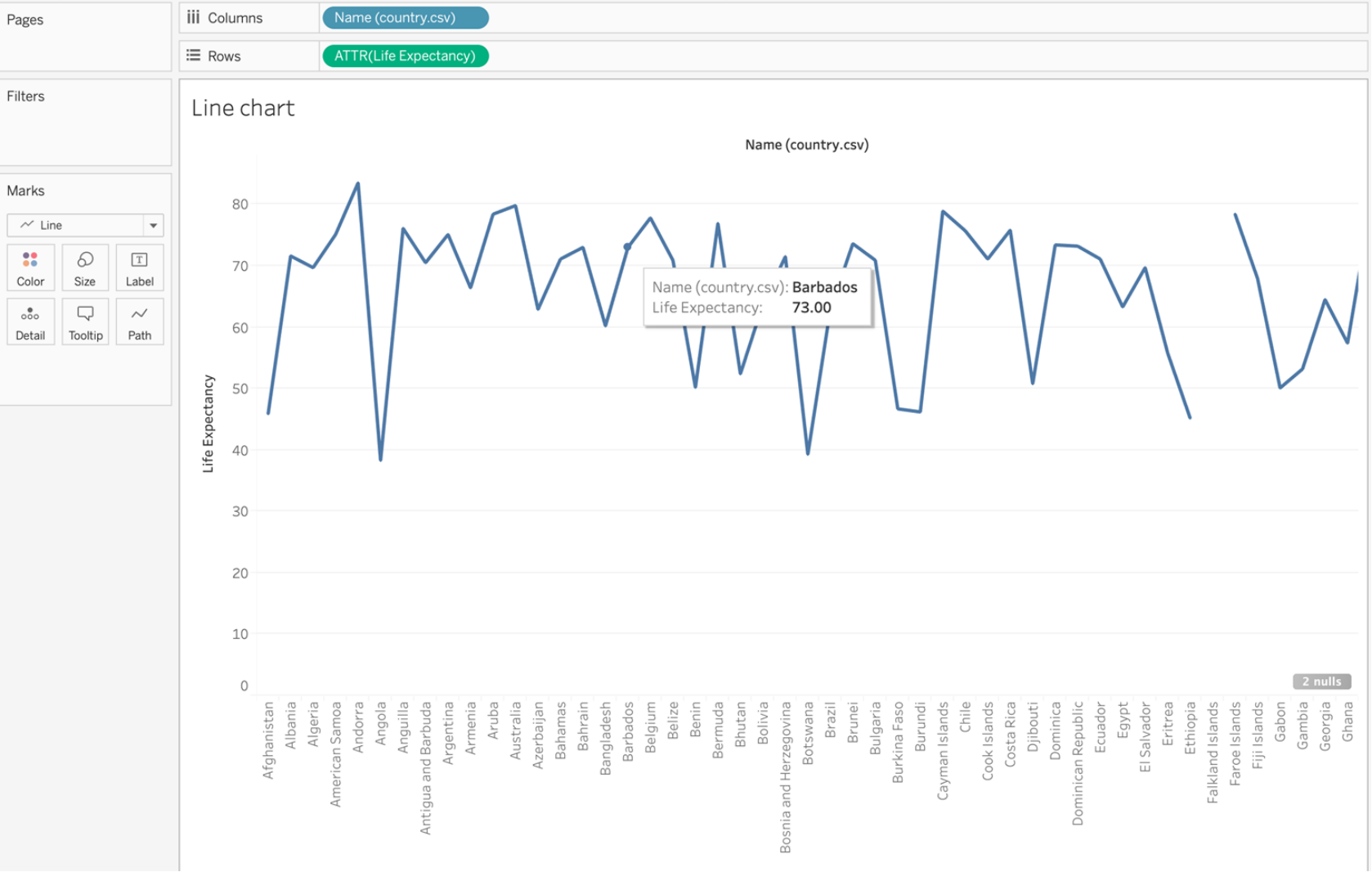
Data visualization: line chart

- We will now make a third graph, **life expectancy by country**



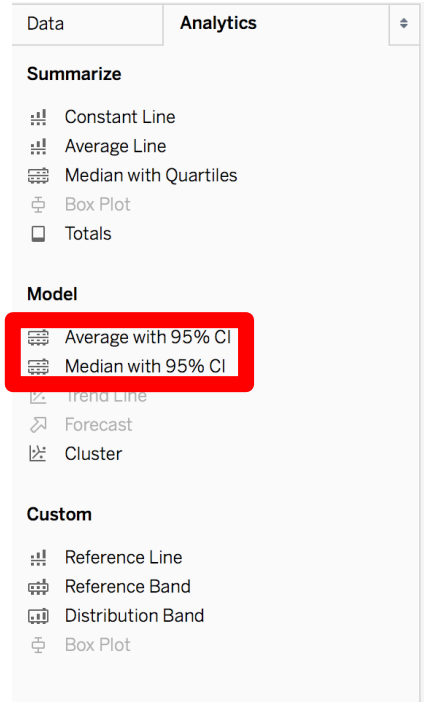
Data Visualization: Line Chart

- What do you see from the graph?
- Is there anything that you would change?
- What follow on visualizations would you do?

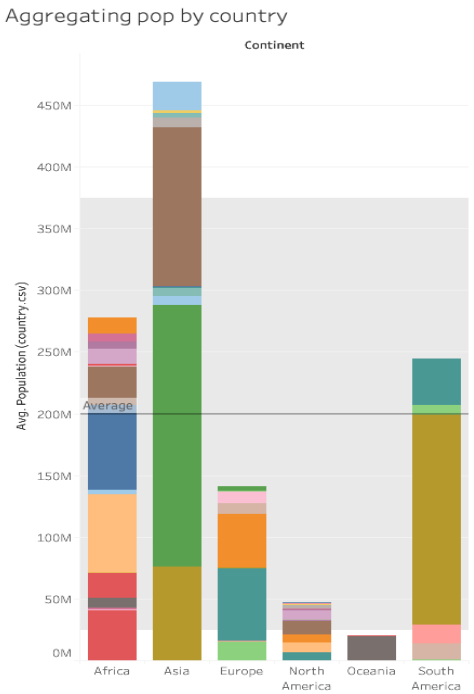


Annotating with the Analytics tab

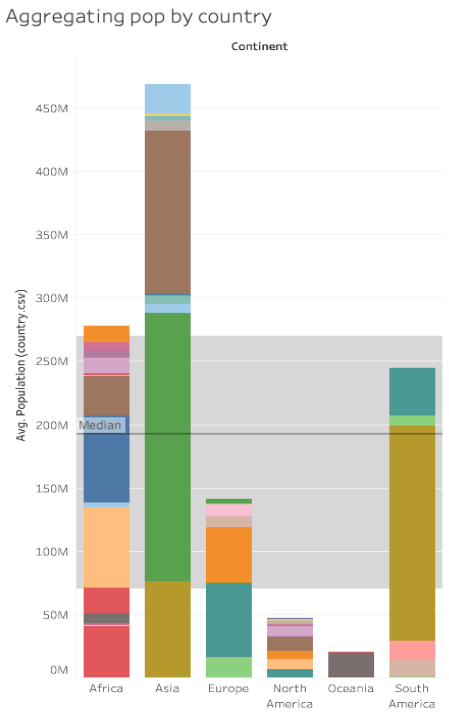
- From the **Analytics** tab, you can annotate features like central tendency and distribution



- Median with quartiles with 95% CI

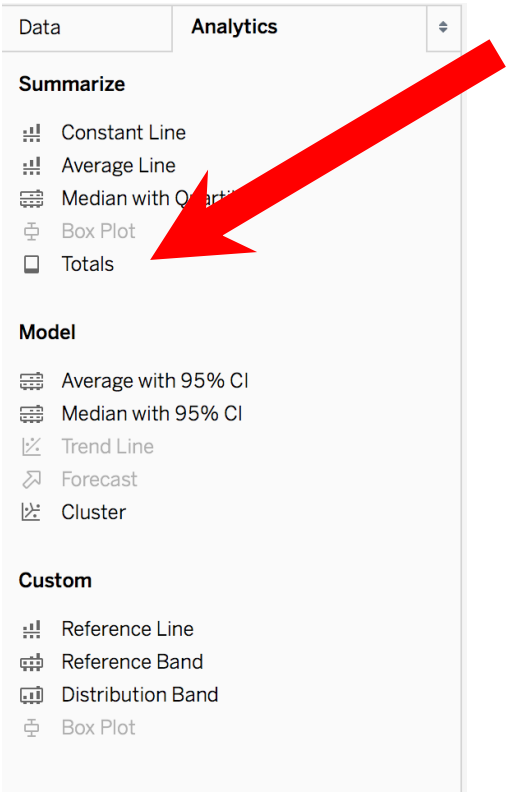


- Mean line with 95% confidence interval



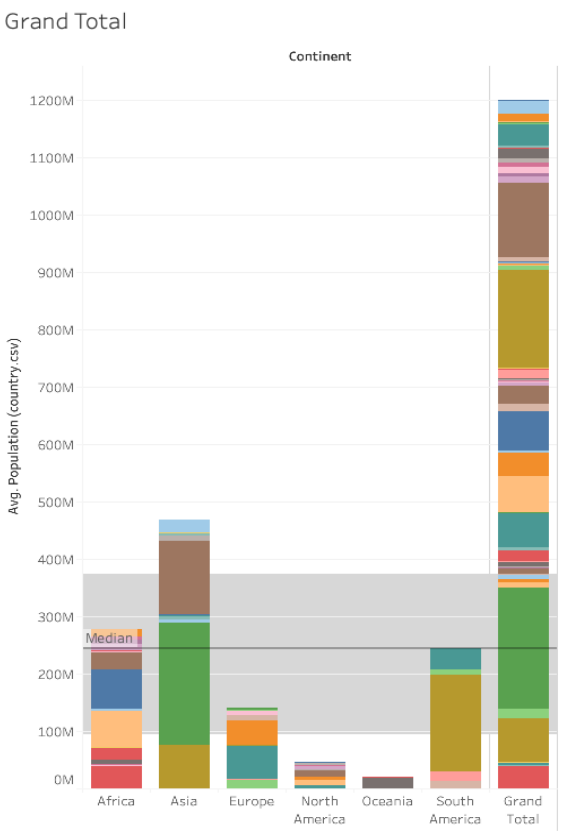
- Or even manually added lines

Totals in the Analytics tab

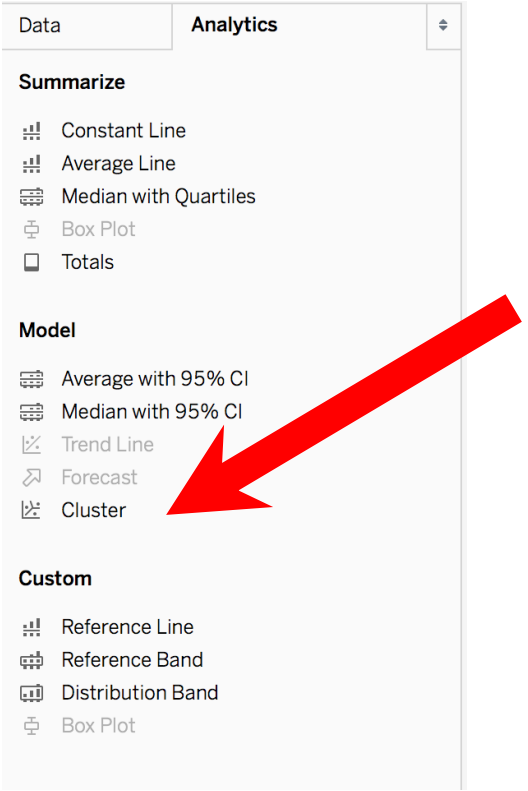


- We can also get a grand **total** column appended to the end of our visualization
- Can anyone spot a red flag to look out for after the totals are graphed?

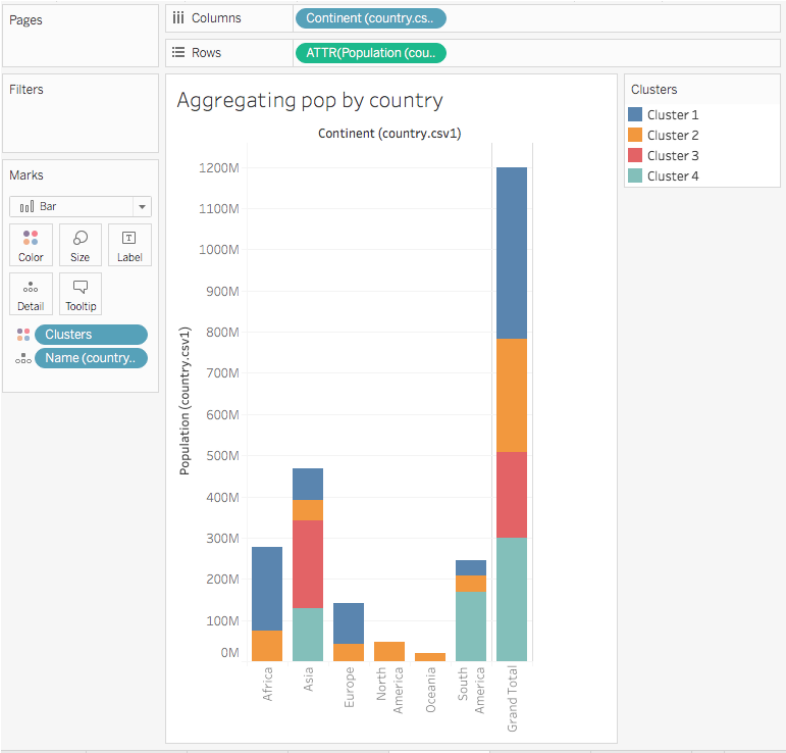
Grand total



Clustering in the Analytics tab



- We can use **clustering** to automatically cluster by any attribute
- Here we automatically clustered by attribute: population size
- Cluster the data on your analysis and see what this means by mousing over the totals column



Summing up the world data

- Let's go through our **critical insights** from this analysis
  - Bar chart of populations
  - Map of populations
  - Line chart - life expectancy

- What are some **next steps** in this analysis?
  - What analysis would you do next?
  - What data would you like to have that you do not have?

Data integrity

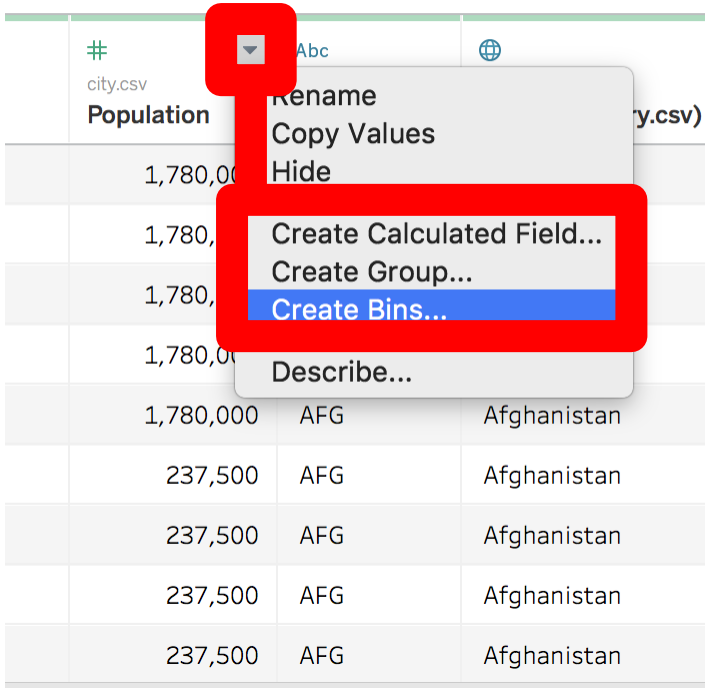
- Are there any **data integrity** issues that you can see?
- How would you deal with these?
- What did you check?
- Is there anything that you should check but did not have time to?

Agenda

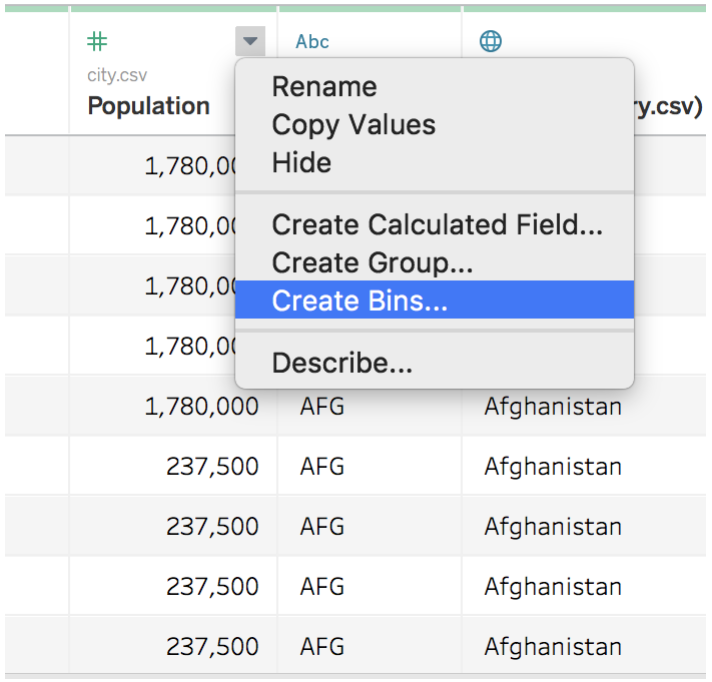
- Import the given dataset into Tableau and explain the concept of joins
- Explore the Tableau platform layout
- Create basic visuals using the World Data
- Introduce the concept of aggregating, binning, and grouping

Aggregating, binning, and grouping

- Sometimes it makes sense to format a column into chunks
  - **Aggregating:** Using a formula to calculate on some grouping of the data
  - **Binning:** Sorting continuous data into bins by value
  - **Grouping:** Using manual assignment to categorize data
- Apply these with the dropdown menu to the right of each column



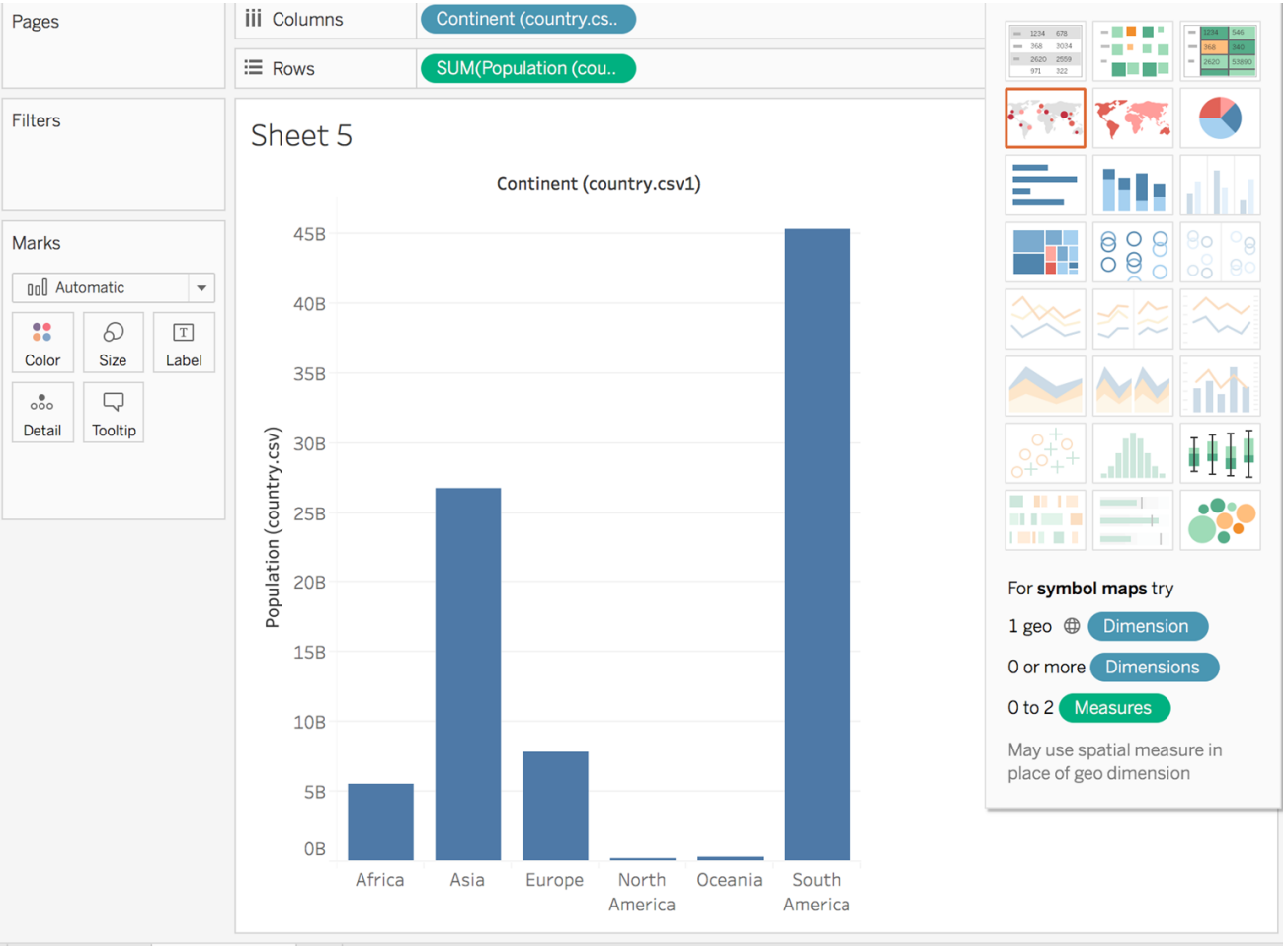
Where do these groups appear?



- Selecting any of these options will make new columns from the original column with some sort of summary of the column
- Let's try this out on some of the world data

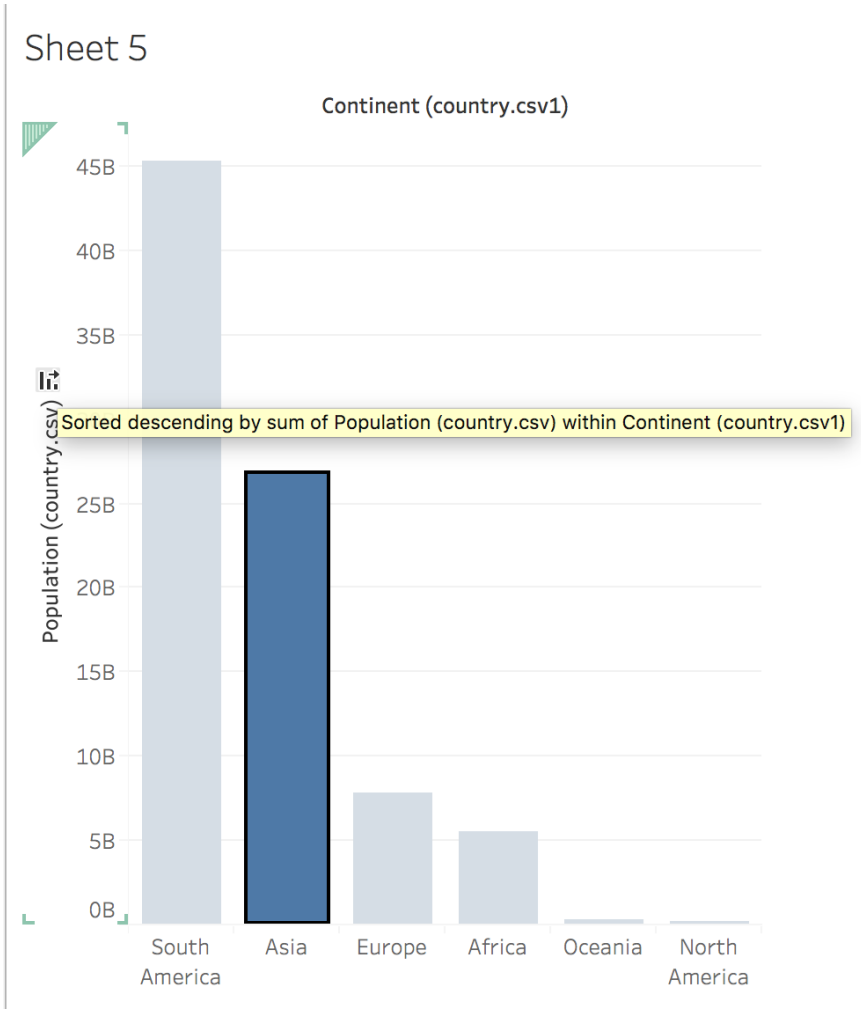
Data visualization: bar chart

- We will look at the **total population by continent**



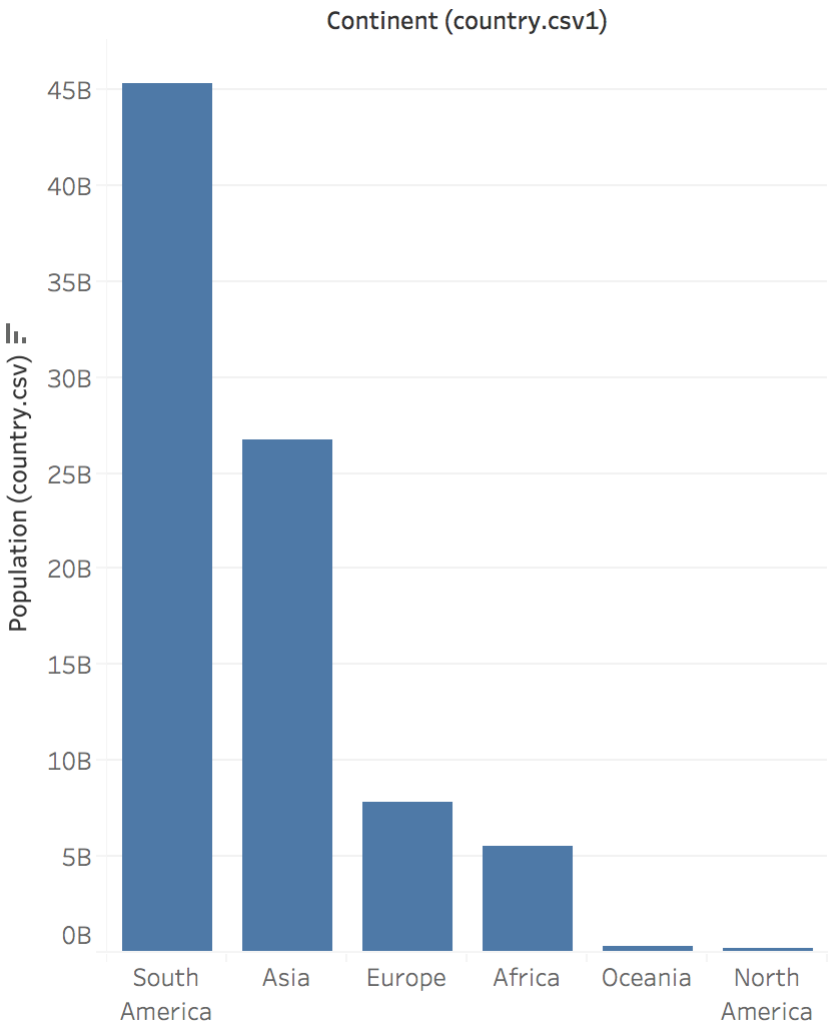
Sorting

- We can **sort by total population** by using this icon, located on the appropriate axis



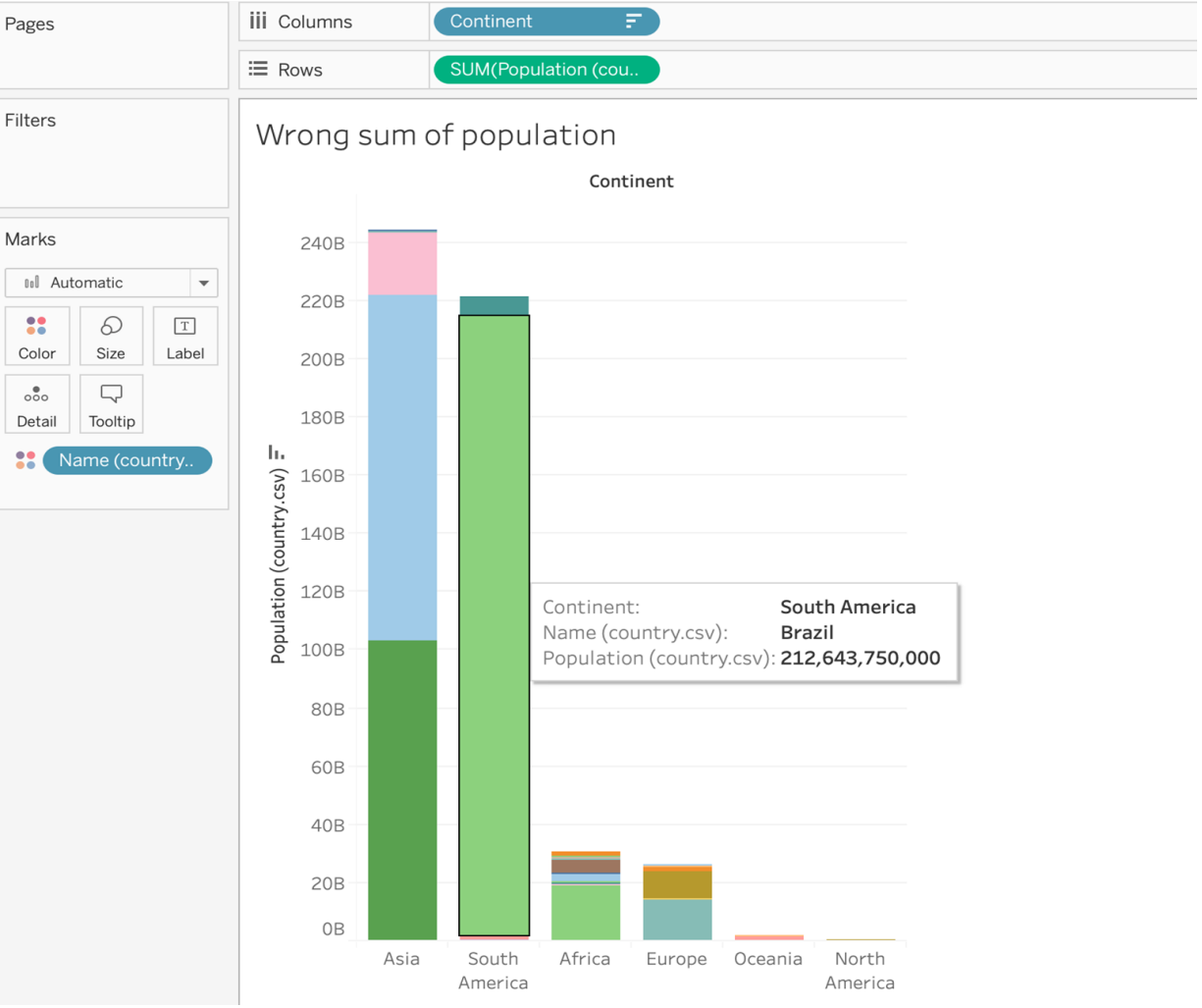
Data integrity

- Does this data make sense?
  - Are there more people in South America than Asia?
- What could be wrong?
- In today's Exercises, we will ask you to:
  - Rebuild these figures in your book
  - Look for data integrity issues
- We will come back to fix them



Data integrity, ctd.

- Let's look at the summary **population data by country**
- Does this data make sense?
  - Is the population supposed to be in billions?
  - Are there 212 billion people in Brazil?
- What could be wrong?



Aggregation in the data

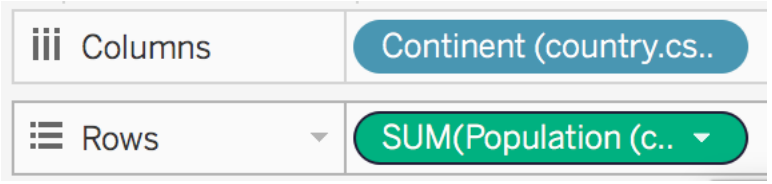
- Check the data – the only way such high numbers are arising is that the **aggregation** of population is wrong
- We want to calculate the population per continent with each country only represented *once*
- We can apply the **ATTR (Attribute)** argument to tell Tableau that the populations are an attribute of each country

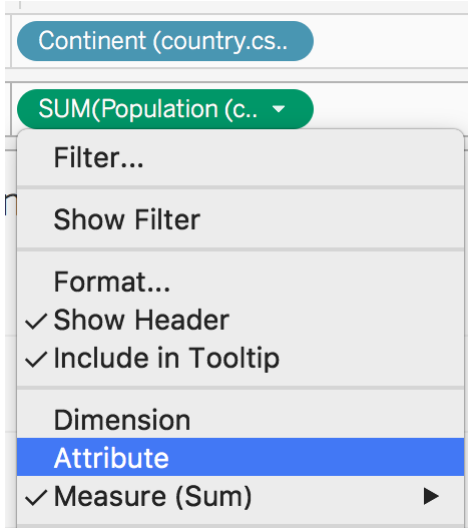
Continent	Name (country.csv)	Population (country.csv)
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000
South America	Brazil	170,115,000

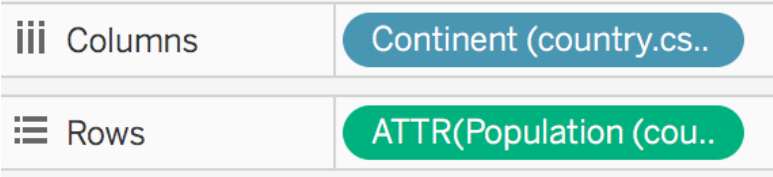
Note that Brazil has 170 M People

Aggregation and attribute

- Look closely at the row and column shelves in the pills above the graph
- Population is being aggregated as a sum across all rows of each country
- Use the drop down in the pill and switch the aggregation to **"Attribute"**
- This means that the country level values are taken as an attribute of the data rather than being further aggregated

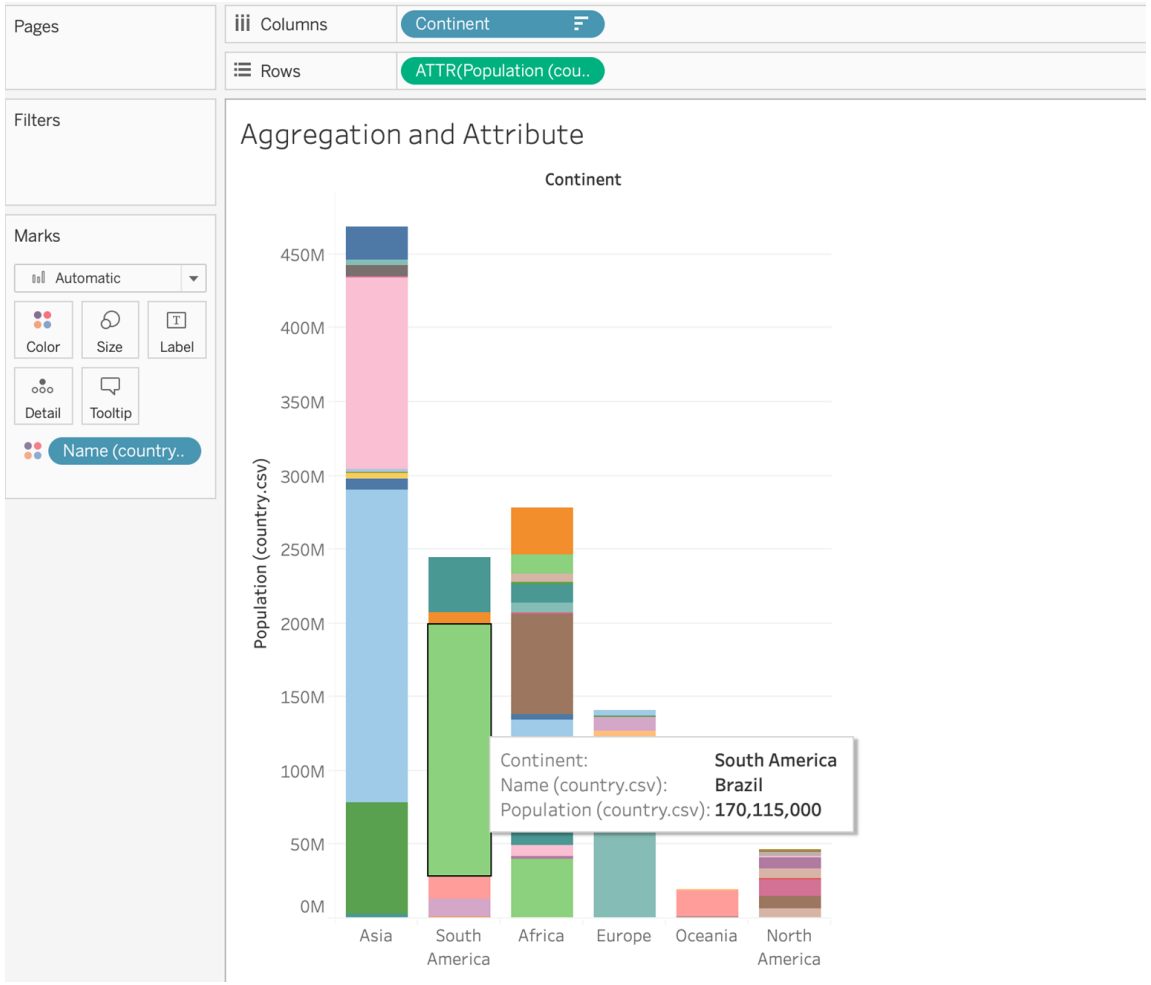






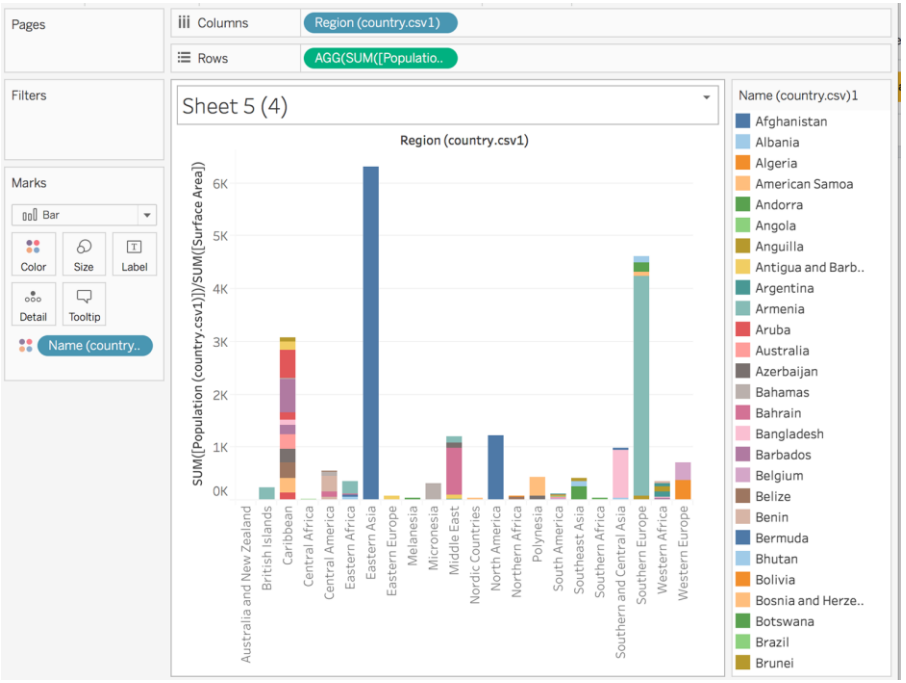
How has our visualization changed?

- Now we are treating the country populations on each row as the whole population of each country, in other words as an **attribute** of the country
- Before, Tableau was treating each row as a part of each country's population and summing it together
- We verify this by noting that the country populations are now as expected



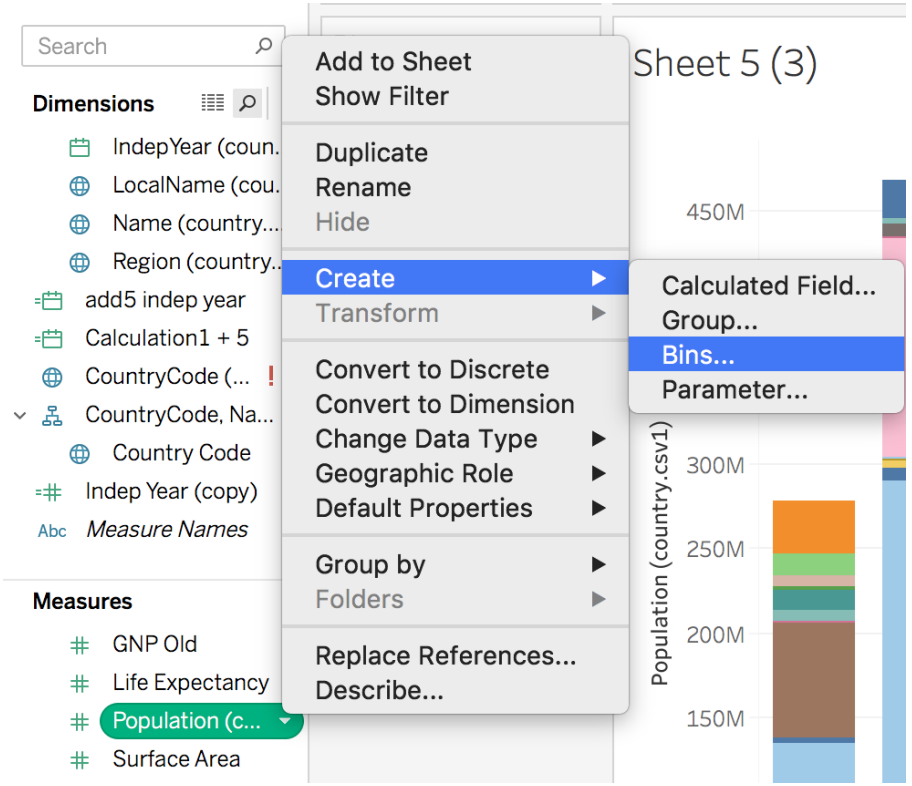
Aggregation: get population density

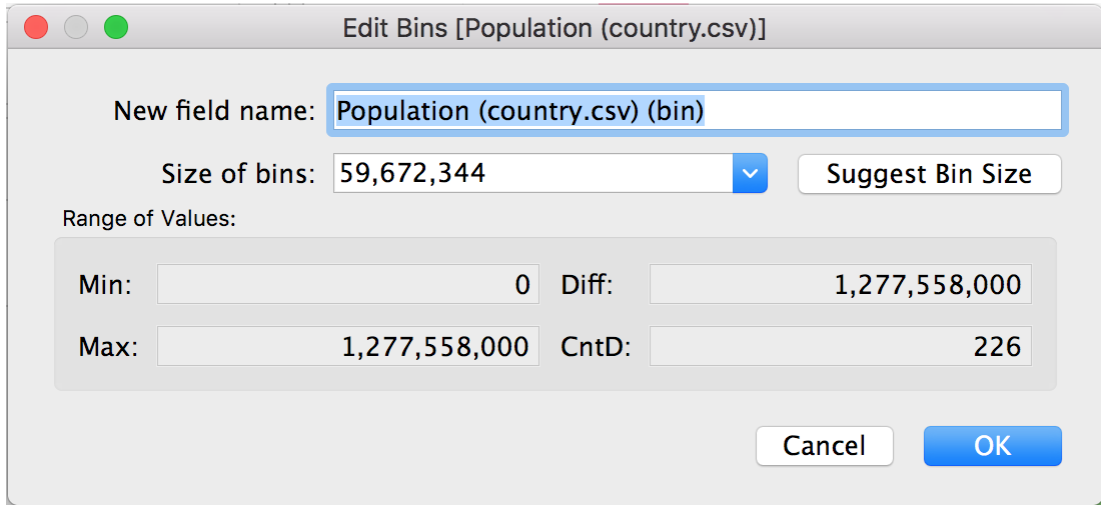
- Let's try plotting **population density:**
  - First, we add an aggregating dimension in the **Columns** field
  - Then we write the aggregating formula in the **Rows** field
  - When we press enter, Tableau makes an aggregation formula that calculates the new value per region
  - **Note** that this is not a new column – rather, it is cast as **"AGG()"**



Binning

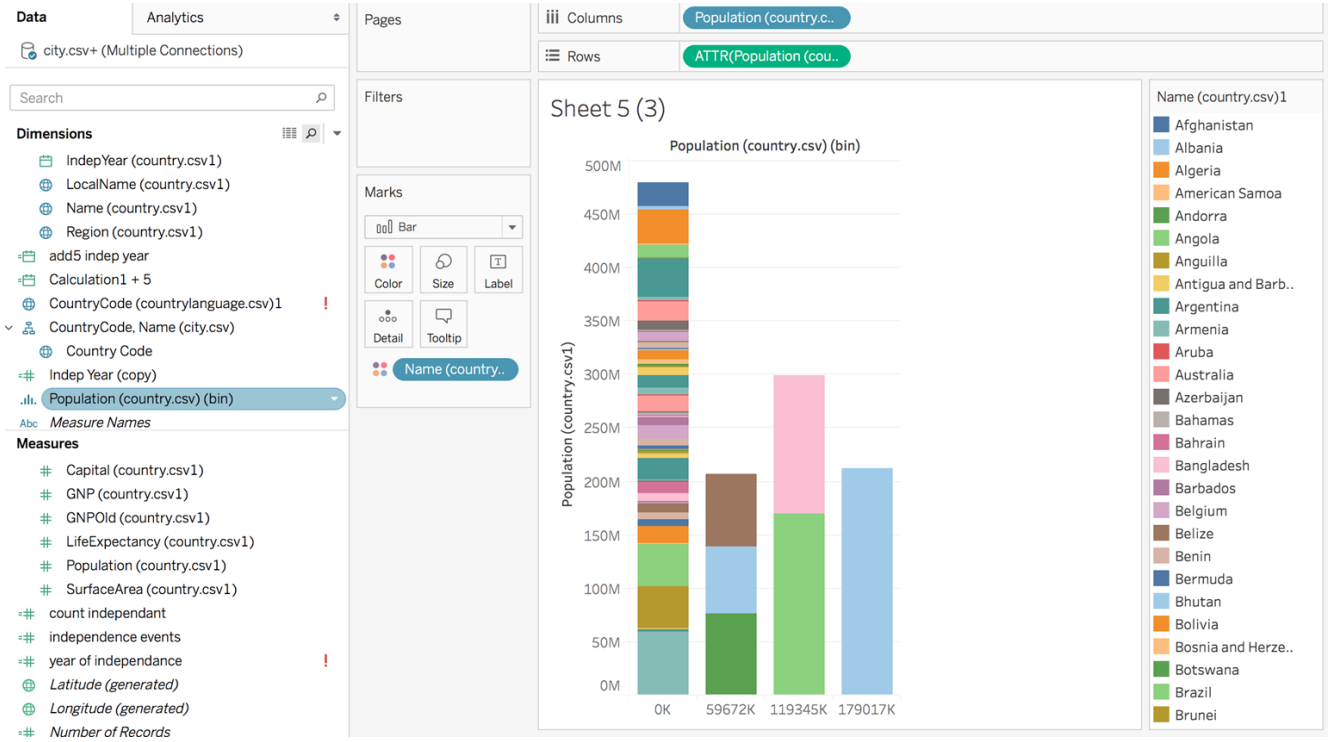
- Now we will create **bins** based on the country's population
- Binning can help us group many **continuous values** into smaller groups of bins for easier analysis





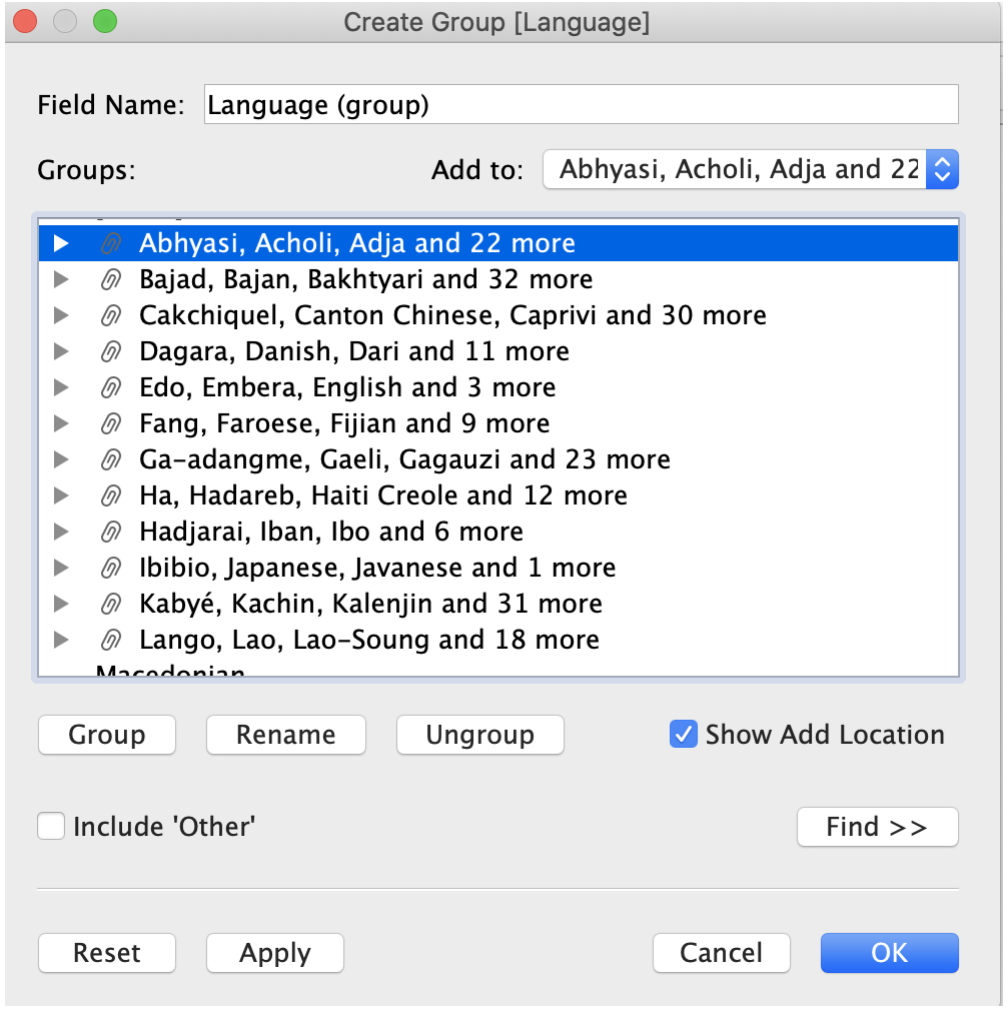
Binning

- We now can see that most countries fall into the smallest of four bins
- Note that we are using the population attribute to keep Tableau from counting countries multiple times



Grouping

- If your data can be classed into obvious or natural categories, you might want to organize them using **grouping**
- You can specify groups manually
- For instance, languages might be grouped into different classes alphabetically



Knowledge check 2

Exercise 2

What we covered today

- Importing data
  - CSV
  - SQL server
- Tableau parts
- Data integrity
- Dimensions and Measures
- “Show Me” palette
- Charts and Figures
  - bar chart
  - symbol map
  - shape plot
  - line plot
- Analytics tab
  - annotations
  - clusters
  - Total
- Aggregating, binning and grouping

Upcoming module

In the next module, we will cover:

- Filtering options
- Formatting options
- Functions in Tableau

DATASOCIETY:

Thank you!