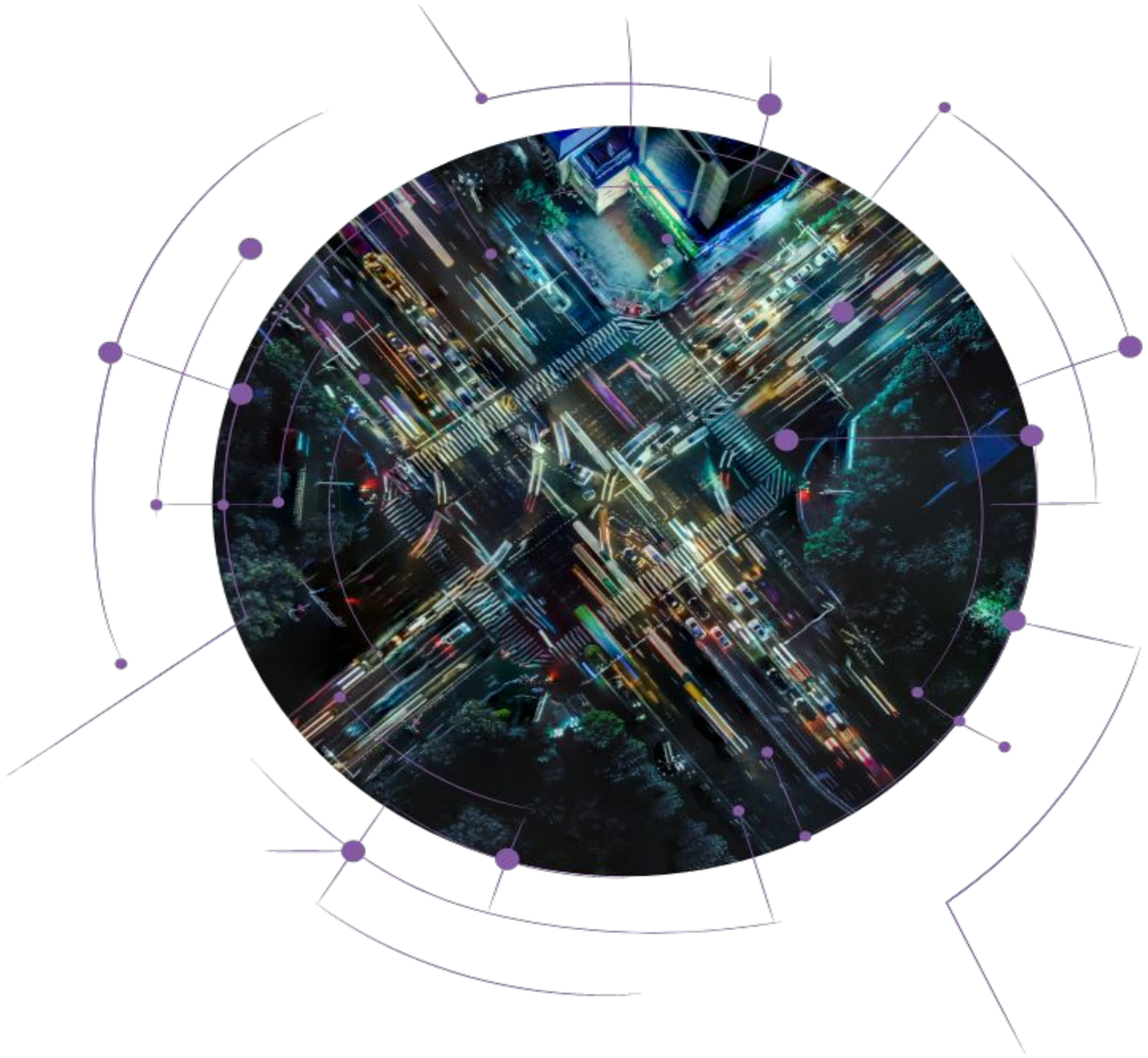
DATASOCIETY:

Introduction to Tableau

Part 1

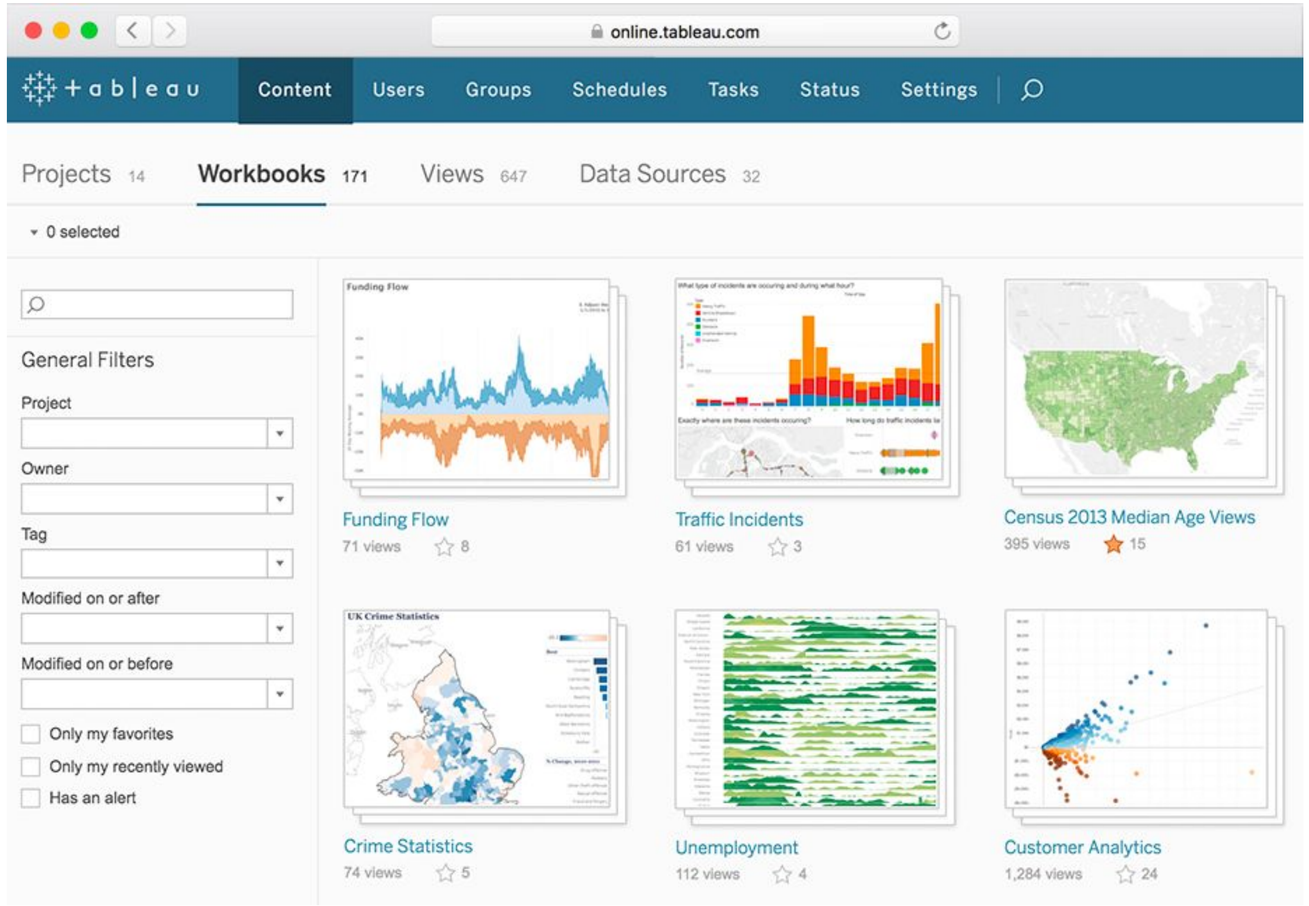


Module completion checklist

Objective	Complete
Explain the need for Tableau and describe its features	
Describe how data sources connect to Tableau	
Import the given dataset into Tableau	
Explain the concept of Relationships	
Discuss the use of joins in Tableau	

Why use Tableau?

- It offers a quick and easy way to create interactive visualizations and explore data.
- It is easy to integrate with multiple data sources.
- It is compatible with OS X, Windows, and Linux.
- It integrates with R and Python for advanced analysis.



Excel vs. Tableau

Parameters	Excel	Tableau
Purpose	Spreadsheet application used for manipulating data	Visualization tool used for data analysis
Usage	Most suitable for statistical analysis of structured data	Most suitable for quick and easy representation of large datasets, which helps resolve big data issues
Performance	Moderate speed, with no option to speed up	Moderate speed, with options to optimize and enhance the progress of an operation
Security	Relatively weak (compared with Tableau); needs regular updates	Extensive options to secure data without scripting; row-level security and permissions are built-in.

Excel vs. Tableau, cont'd.

Parameters	Excel	Tableau
User Interface	Macro and Visual Basic scripting knowledge required to maximize tool potential	Tool can be used without any coding knowledge
Business Need	Best for preparing one-off reports with small datasets	Best while working with big data
Products	Bundled with MS Office Tools	Comes with different versions, such as Tableau server, cloud, and desktop
Integration	Integrates with ~60 applications	Integrates with ~250 applications
Real-time Data Exploration	You need to have an idea of where your data leads to generate insights.	You are free to explore data without seeking concrete outcomes, especially with features like drill-down and data blending.

Module completion checklist

Objective	Complete
Explain the need for Tableau and describe its features	✓
Describe how data sources connect to Tableau	
Import the given dataset into Tableau	
Explain the concept of Relationships	
Discuss the use of joins in Tableau	

Installation

- The Tableau Desktop version we have used during development is 2020.2.19, available here: **Tableau Desktop 2020.2.19**.

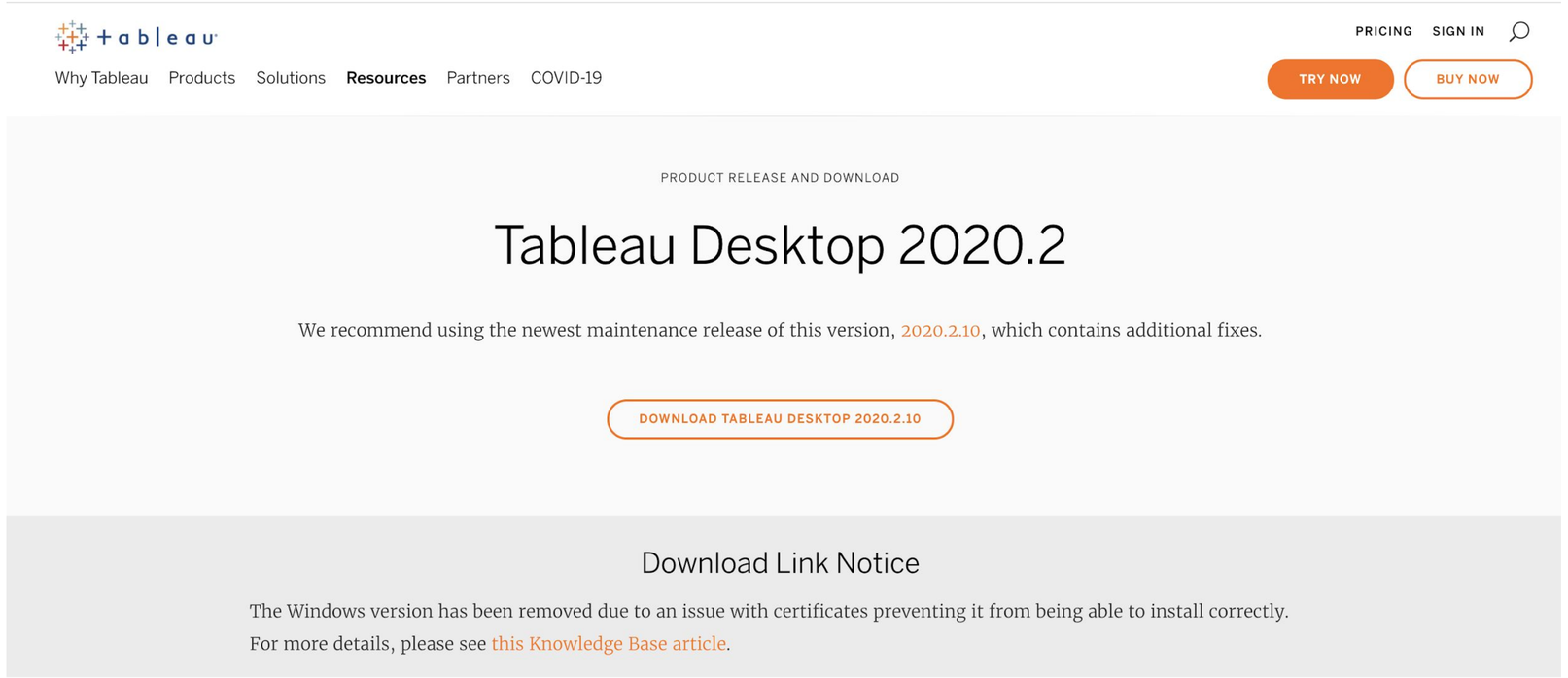
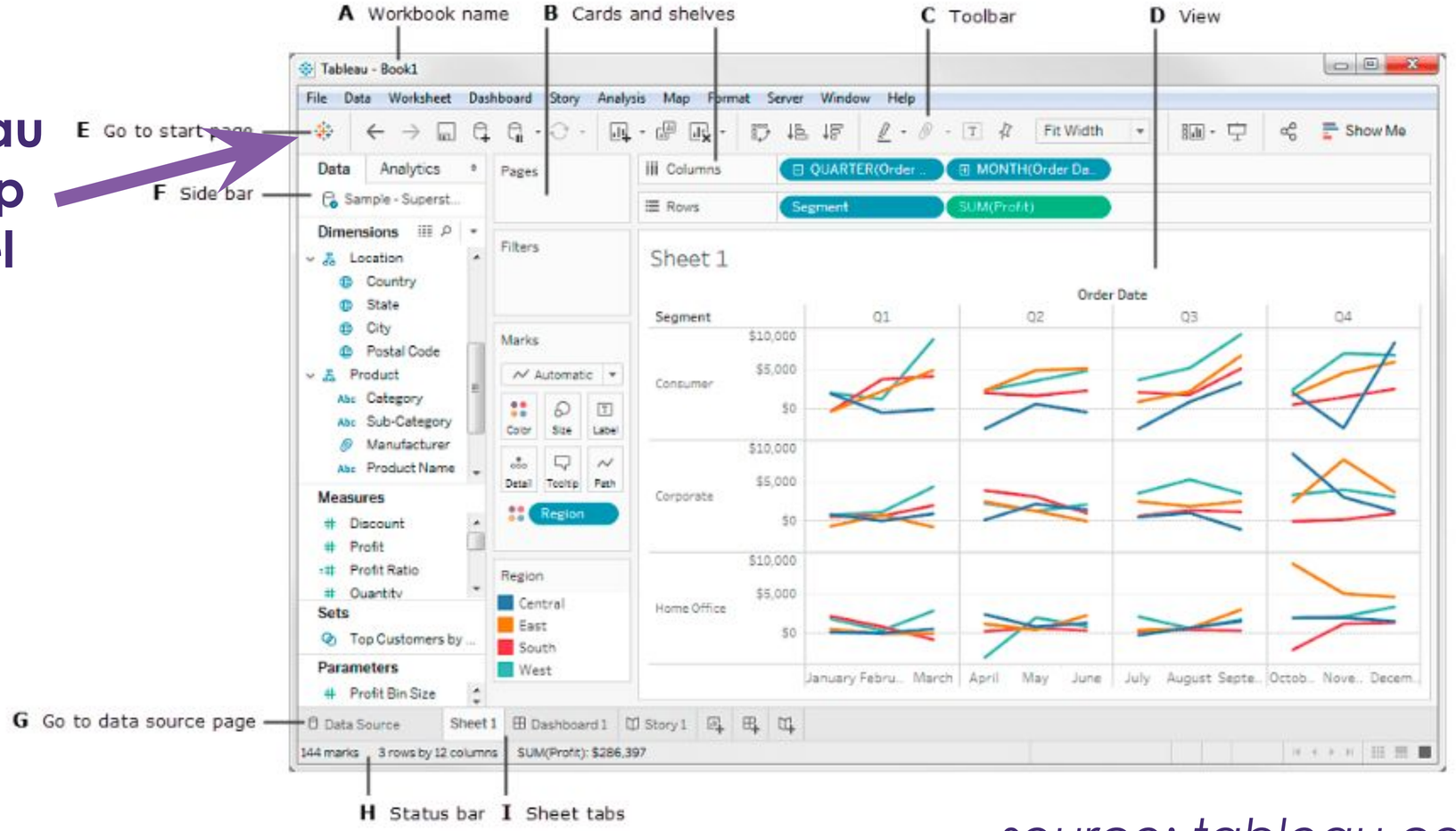


Tableau overview

Click the Tableau logo to bring up Connect panel



source: tableau.com

Module completion checklist

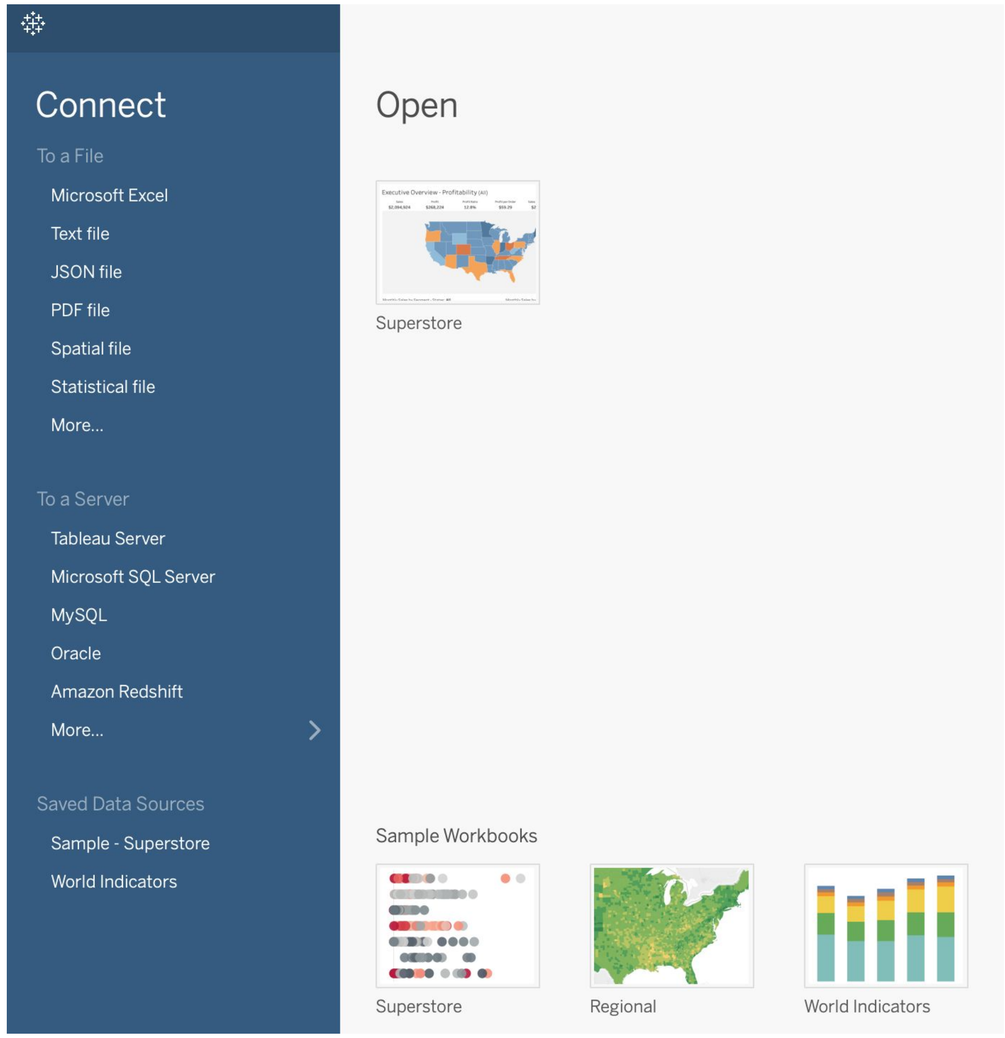
Objective	Complete
Explain the need for Tableau and describe its features	✓
Describe how data sources connect to Tableau	✓
Import the given dataset into Tableau	
Explain the concept of Relationships	
Discuss the use of joins in Tableau	

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).

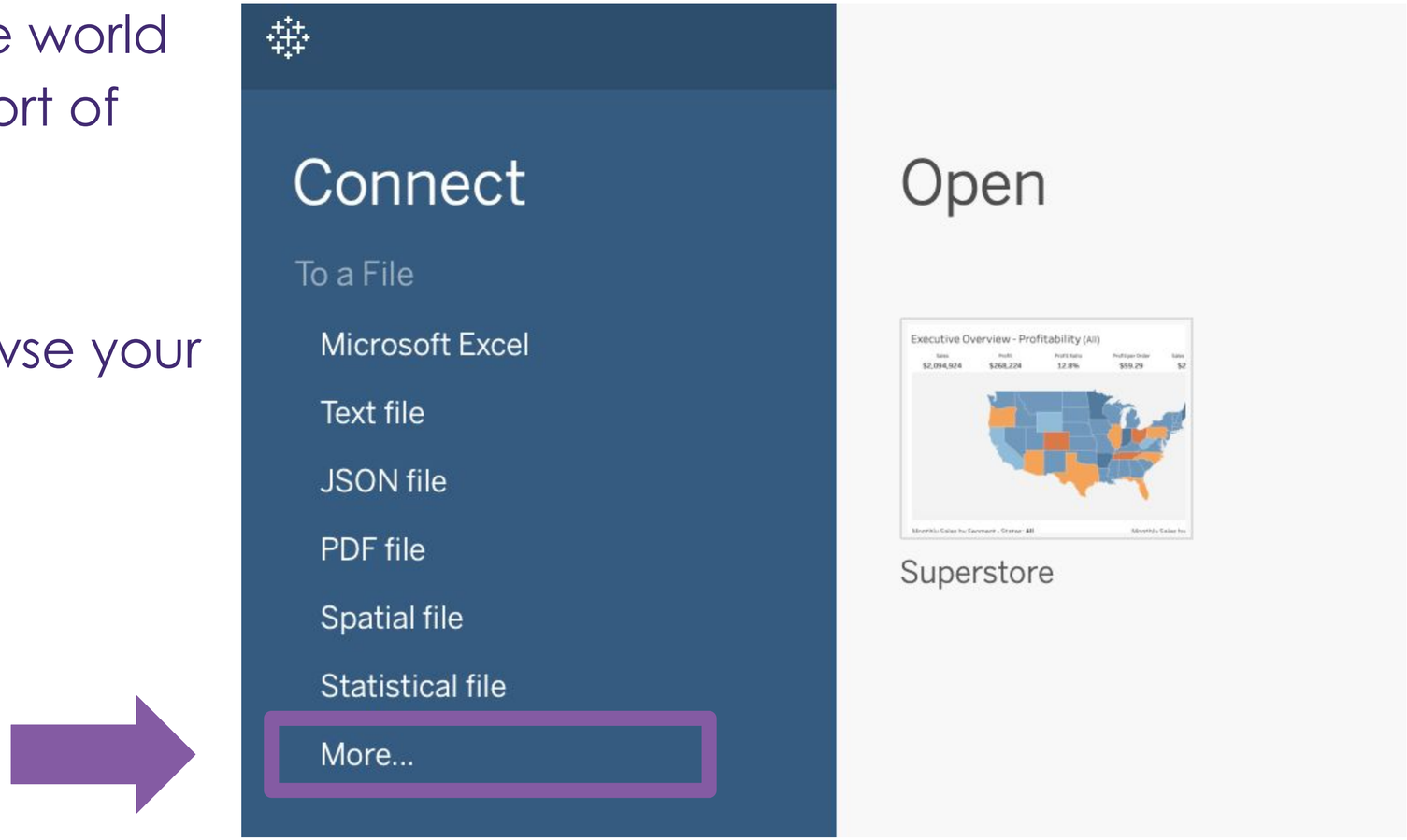
- Supports Database Connections such as:
  - MySQL.
  - Oracle.
  - Redshift.



Import world data : CSV

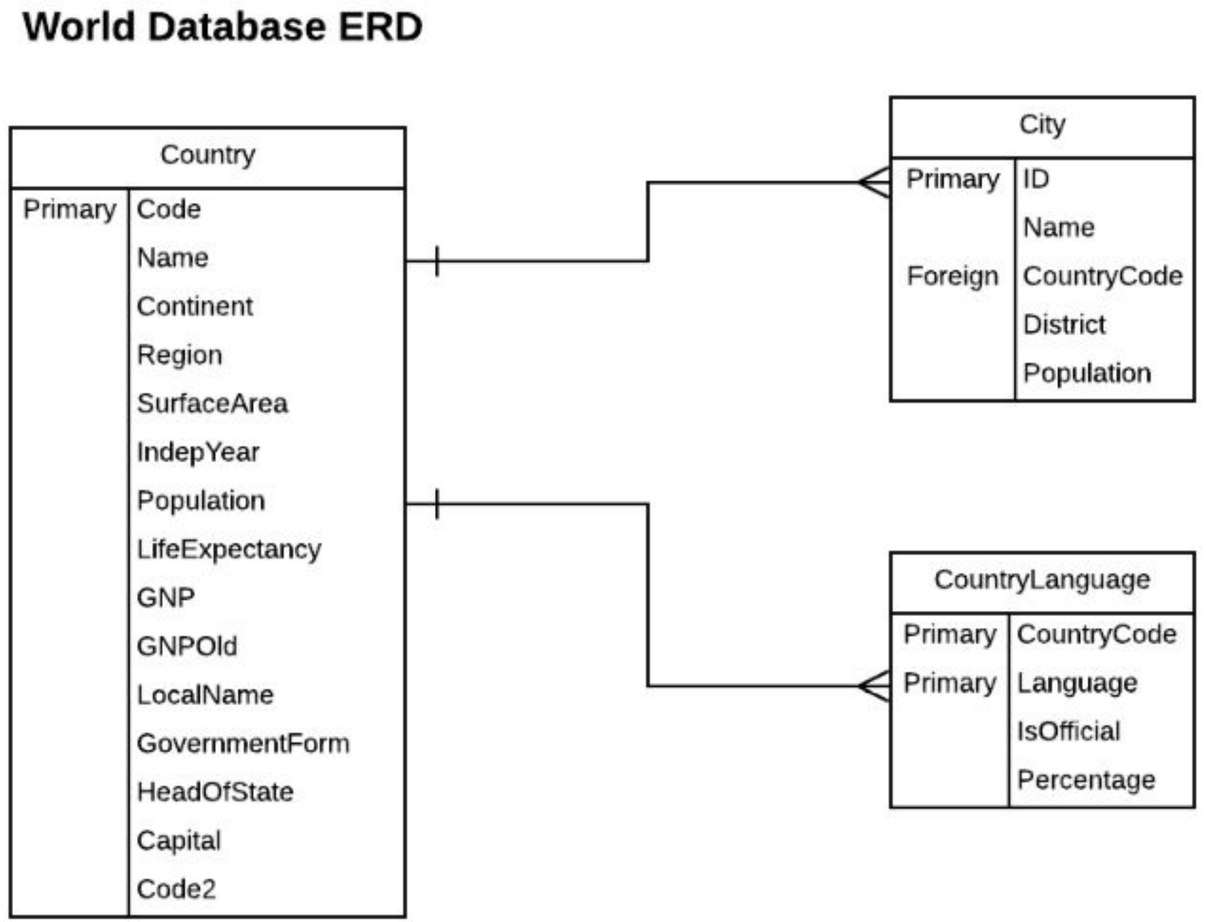
- Let's import some pieces of the world dataset today and see what sort of insights we can reveal.

- 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.



Module completion checklist

Objective	Complete
Explain the need for Tableau and describe its features	✓
Describe how data sources connect to Tableau	✓
Import the given dataset into Tableau	✓
Explain the concept of Relationships	
Discuss the use of joins in Tableau	

Joining datasets using Relationships

- It is often necessary to combine data from multiple places - different tables or even data sources - to perform a desired analysis.

- Tableau has introduced a feature called **Relationships** to minimize the amount of data pre-processing required to visualize related datasets in multiple ways.

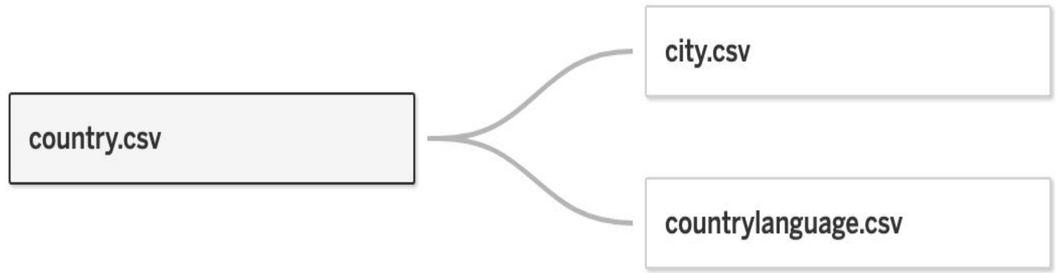
- To explore some of the concepts and features of Relationships, we're going to examine geospatial data from the World Database.

Relationships are contextual joins

- A single relationship will support all join types at the same time.

- For **Measures**, this means **all values are always retained**, even if they're unmatched nulls (not missing values).

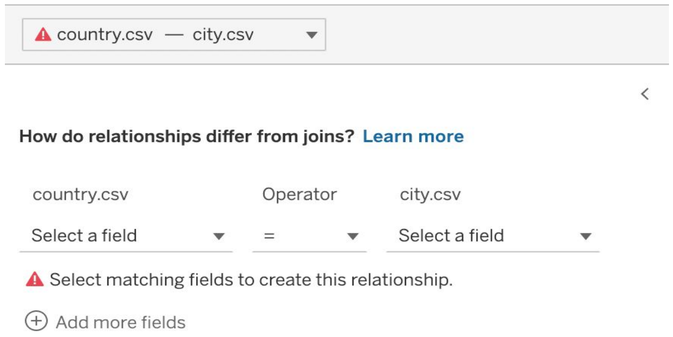
- For **Dimensions**, relevant domains will be displayed **across tables**, and Tableau will display all values in the domain by default (even if there are no matching Measure values).

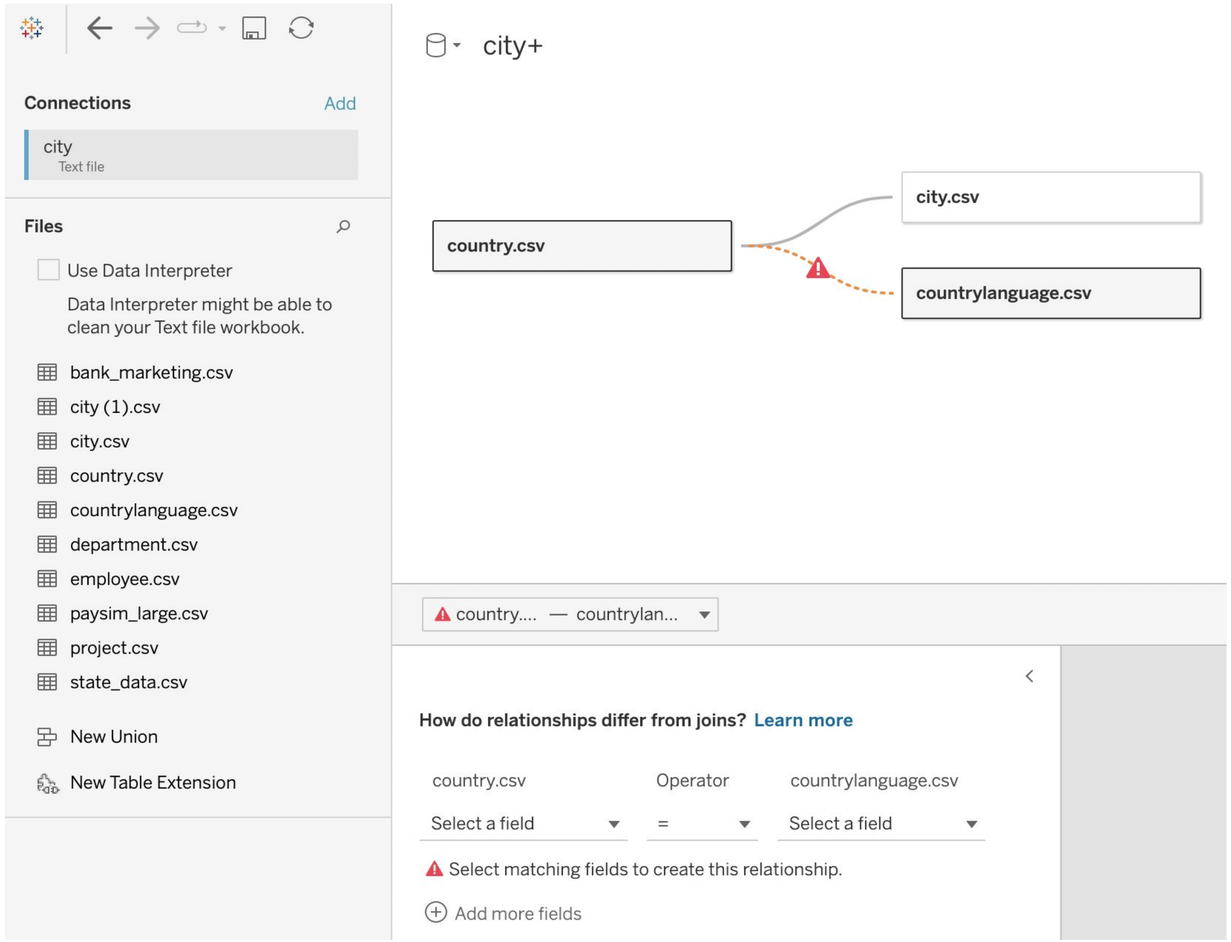


Select fields to create Relationships

- Tableau creates Relationships after **identifying the fields** you want to use to establish relationships.

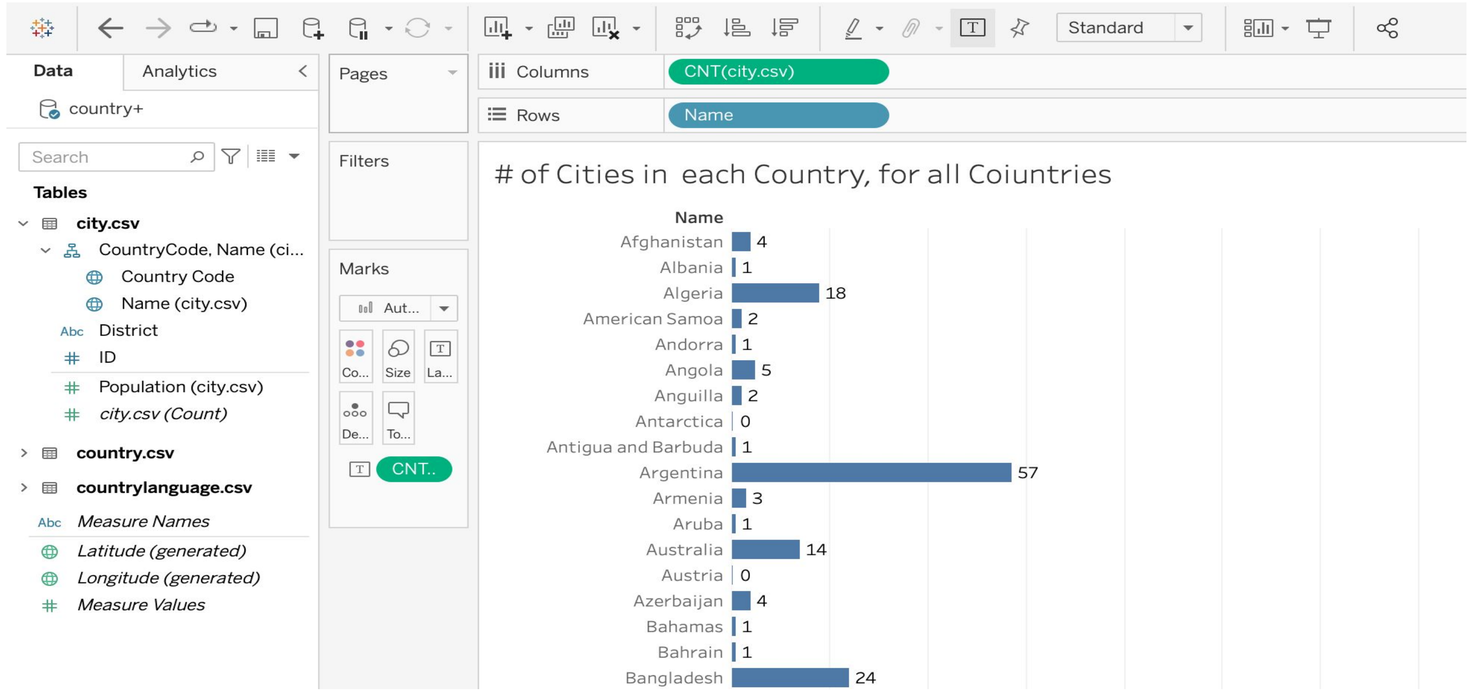
- These fields should have common values that can be used to connect records from one table to another.





Unmatched measures as zero

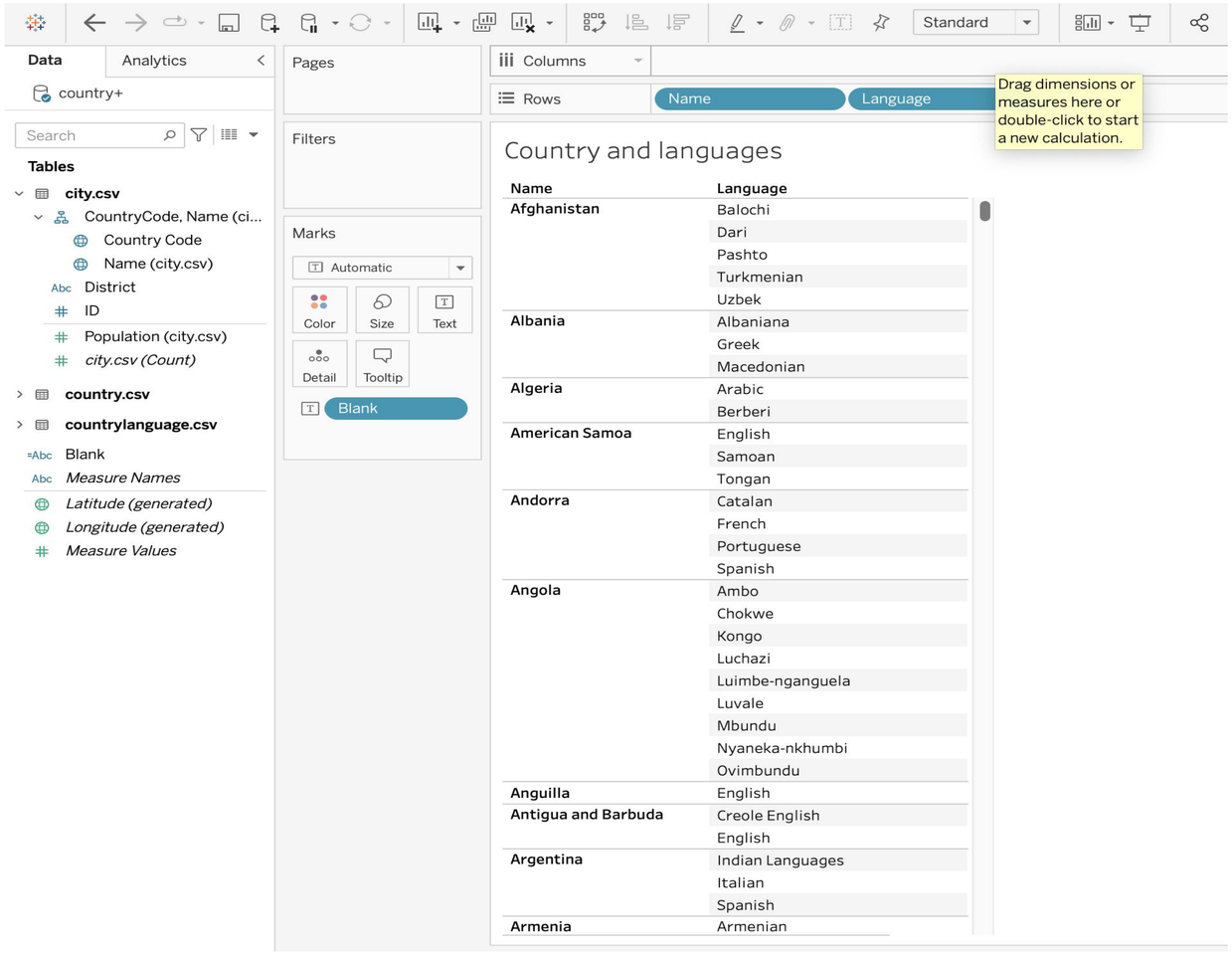
- Relationships allow for unmatched entries in one dataset to display as 0s.
- Visualizing the count of cities by country shows **all** countries, even those without any corresponding cities.
- Displaying all countries helps to avoid undercounting the number of countries.



Relevant dimensions across tables

- Additionally, Relationships affect how Dimensions will behave in visualizations involving different datasets.

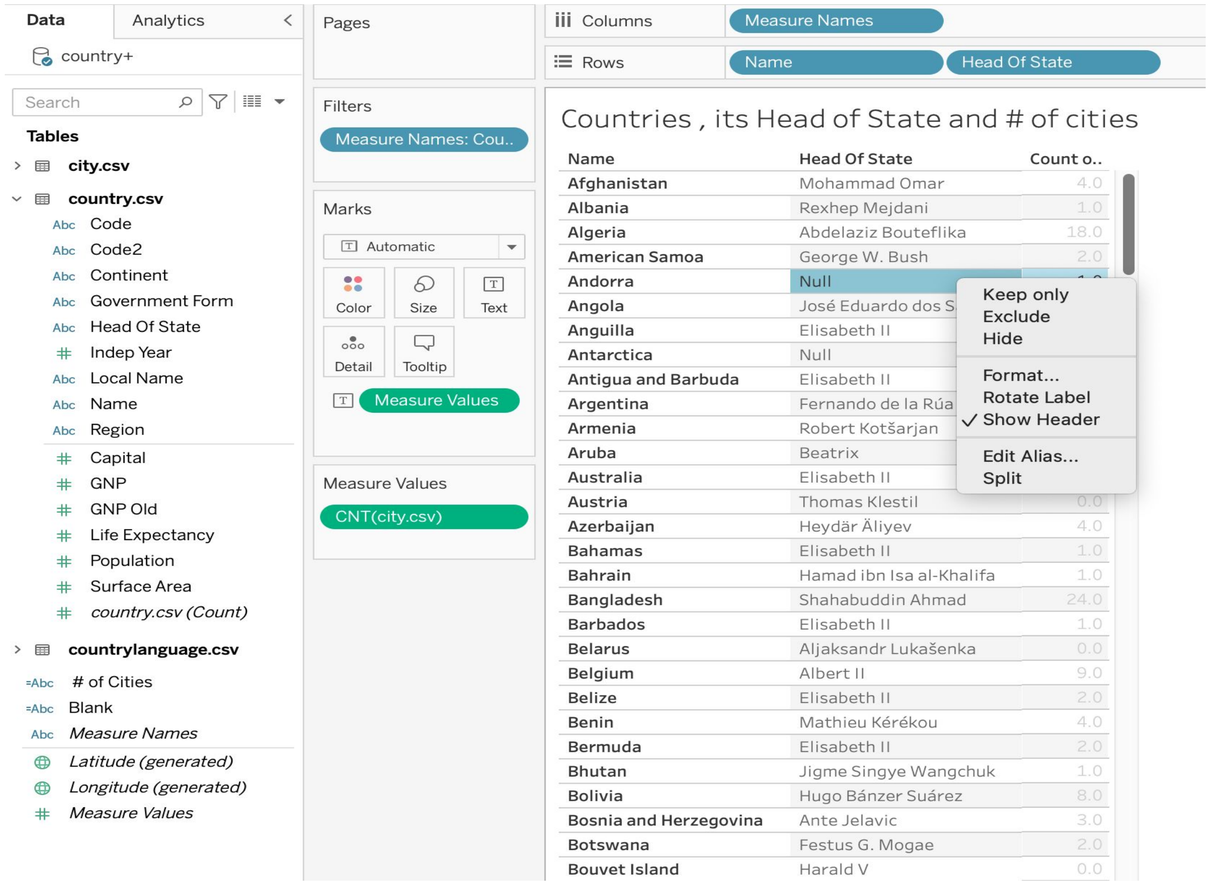
- When visualizing countries and languages, Tableau excludes territories like Antarctica, which may not have any languages.



Retaining unmatched values

- A **null** appears to represent a country a without Head of State (unmatched value)

- Retaining these unmatched values through the Relationship again helps avoid undercounting the number of countries in total.

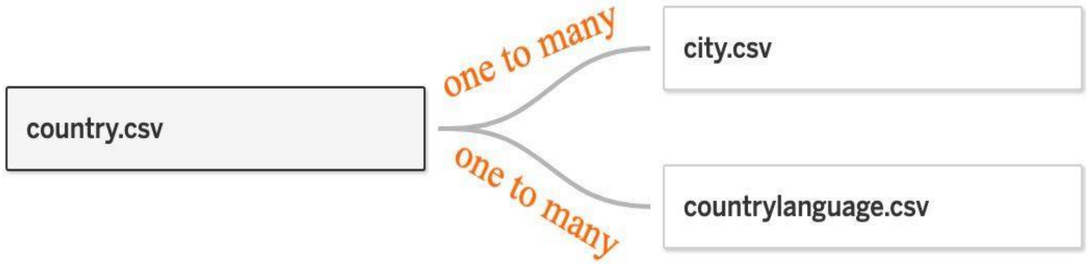


Relationships are smart aggregations

- Aggregations will resolve to the Measure's inherent level of detail from the source table.

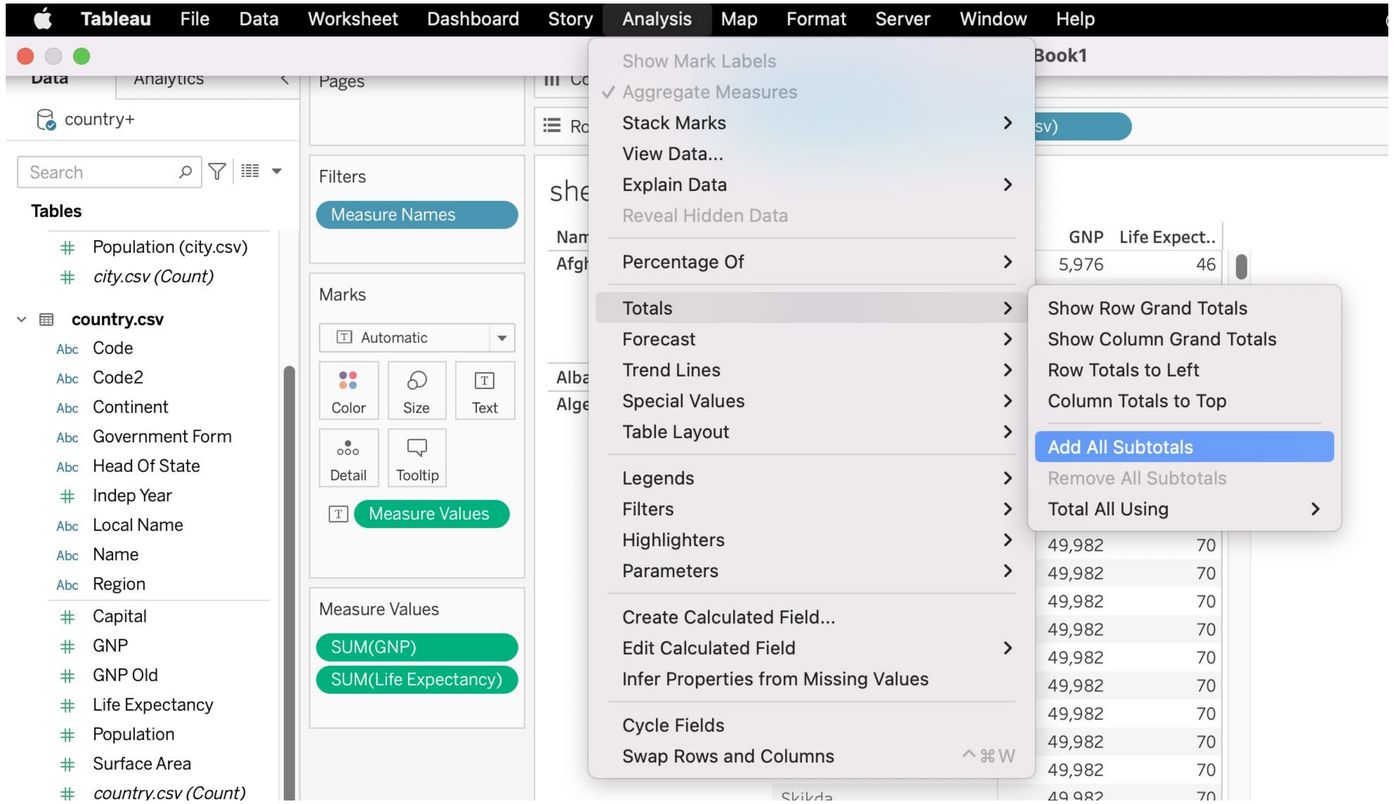
- Measures are only replicated across **lower** levels of detail in the visualization.

- This means it is much less likely to end up with unnecessary, meaningless duplicated data in visualizations.



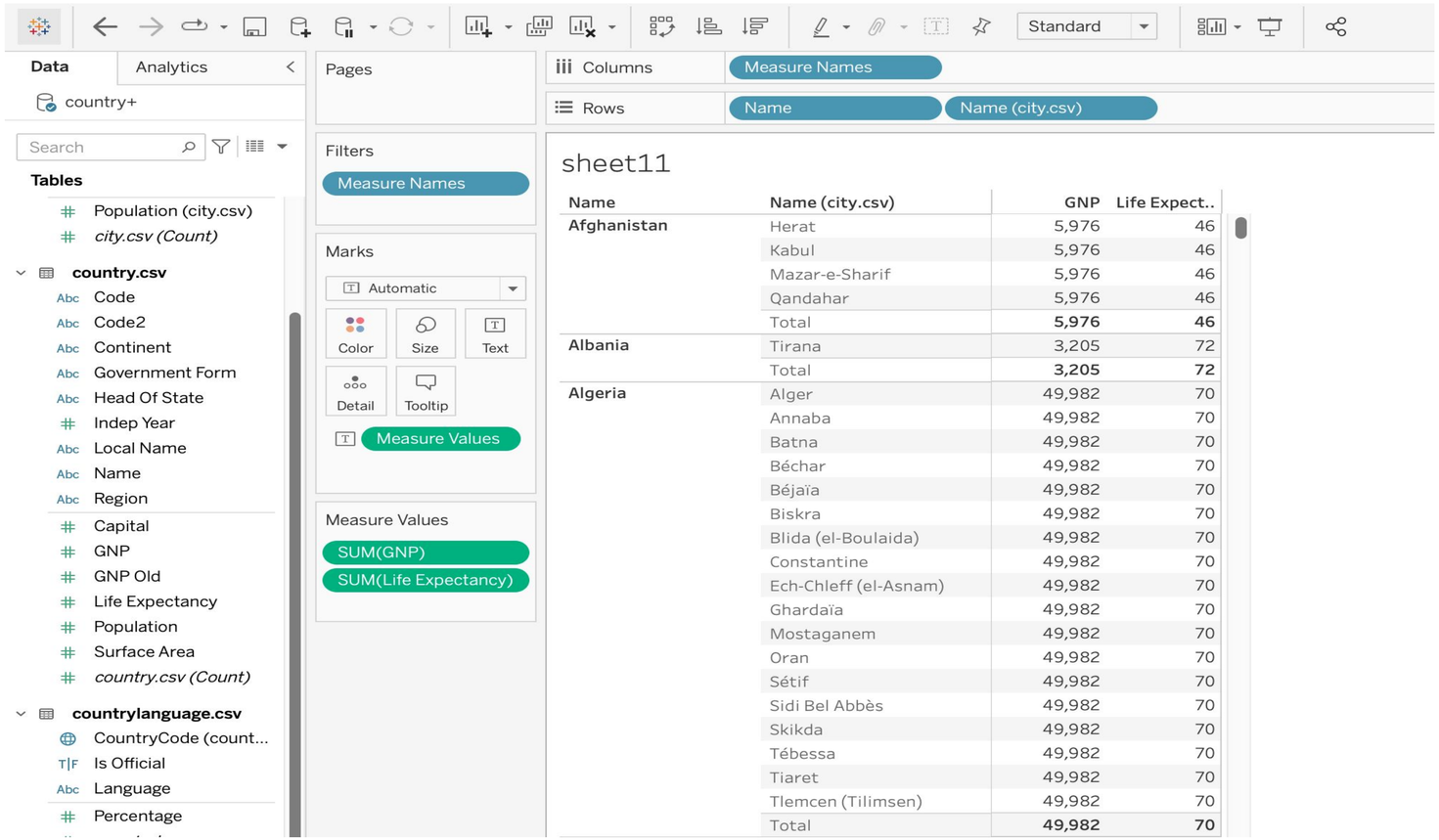
Relationships and level of detail

- Relationships resolve to a Measure's **level of detail**.
- For example, each country is linked to some cities, a GNP, and life expectancy data.
- When we add Country Name, City Name, GNP, and Life Expectancy, to a visualization, we obtain subtotals of GNP and life expectancy.



Relationships and level of detail, cont'd.

- Tableau aggregates life expectancy and GNP in accordance to Country Name.



Name	Name (city.csv)	GNP	Life Expect..
Afghanistan	Herat	5,976	46
	Kabul	5,976	46
	Mazar-e-Sharif	5,976	46
	Qandahar	5,976	46
	Total	**5,976**	**46**
Albania	Tirana	3,205	72
	Total	**3,205**	**72**
Algeria	Alger	49,982	70
	Annaba	49,982	70
	Batna	49,982	70
	Béchar	49,982	70
	Béjaïa	49,982	70
	Biskra	49,982	70
	Blida (el-Boulaida)	49,982	70
	Constantine	49,982	70
	Ech-Chleff (el-Asnam)	49,982	70
	Ghardaïa	49,982	70
	Mostaganem	49,982	70
	Oran	49,982	70
	Sétif	49,982	70
	Sidi Bel Abbès	49,982	70
	Skikda	49,982	70
	Tébessa	49,982	70
	Tiaret	49,982	70
	Tlemcen (Tilimsen)	49,982	70
	Total	**49,982**	**70**

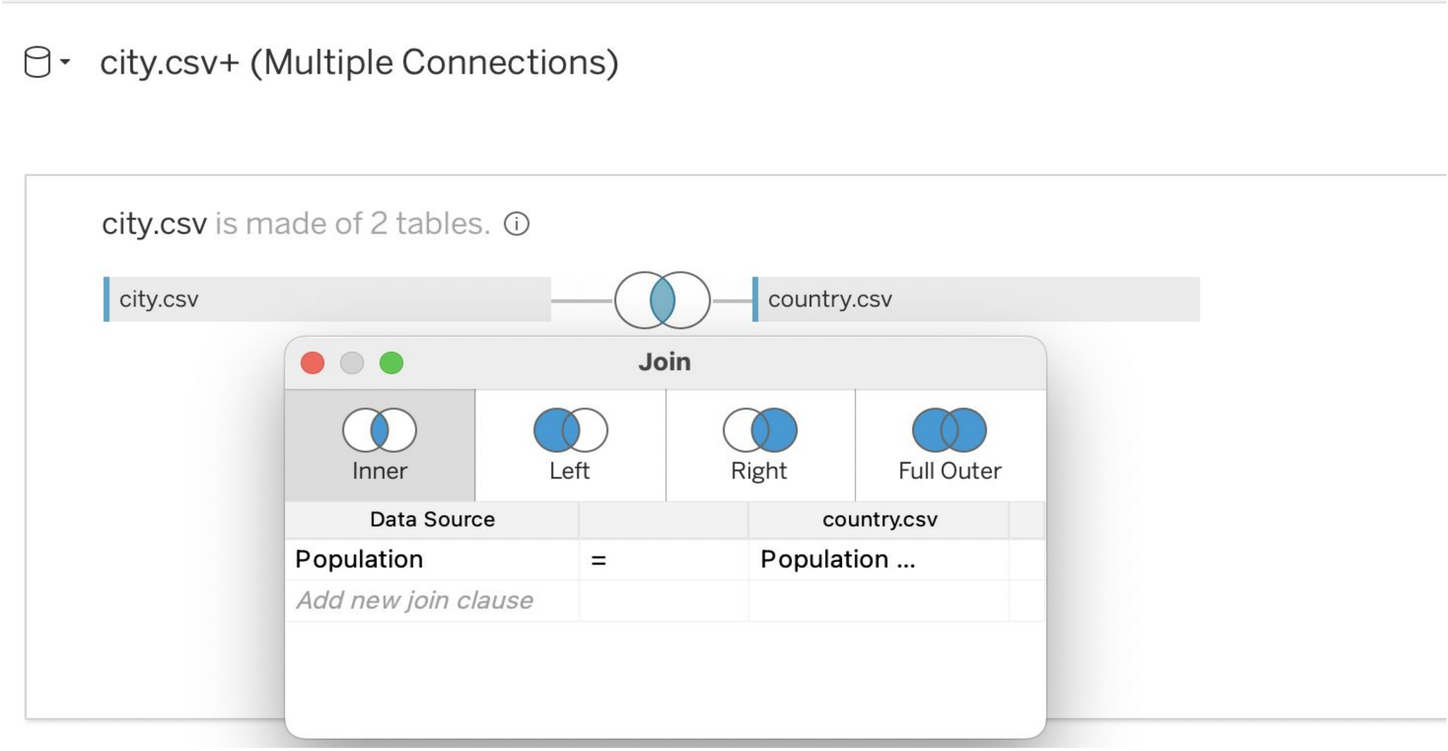
Module completion checklist

Objective	Complete
Explain the need for Tableau and describe its features	✓
Describe how data sources connect to Tableau	✓
Import the given dataset into Tableau	✓
Explain the concept of Relationships	✓
Discuss the use of joins in Tableau	

Manually joining datasets

- Previously, Tableau utilized **joins** in order to merge and combine data.

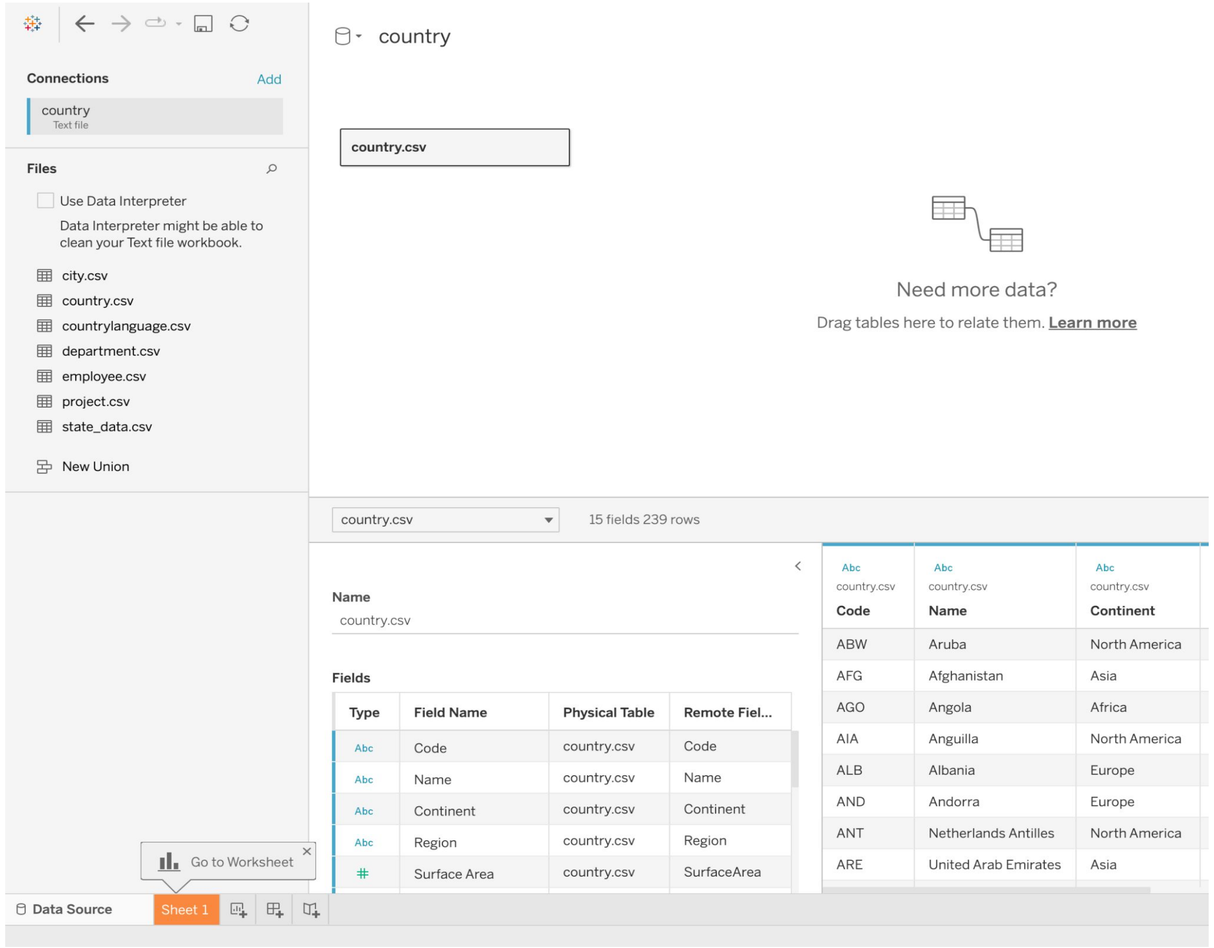
- It is still possible to join tables **manually** by specifying the precise variables the datasets will use as common keys, but this functionality is no longer the default.



Joining tables using joins

- To join tables using Relationships, we can use the default **relationship canvas** visible when viewing data sources.

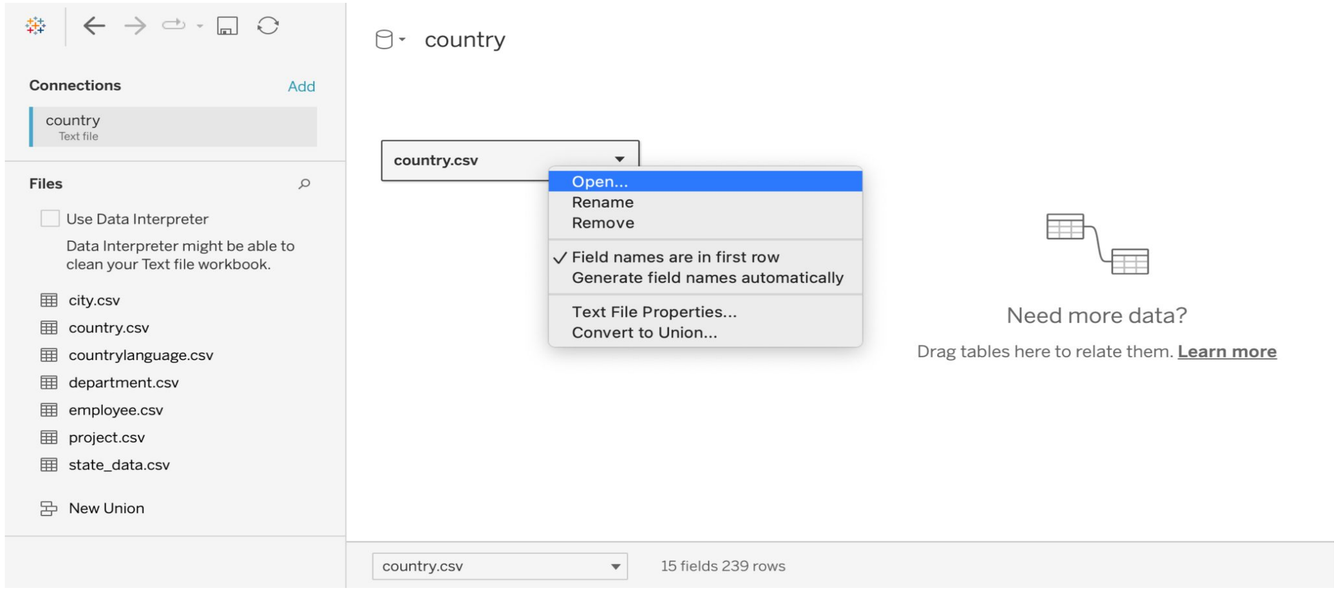
- But to join tables manually, we can access the legacy **join canvas**.

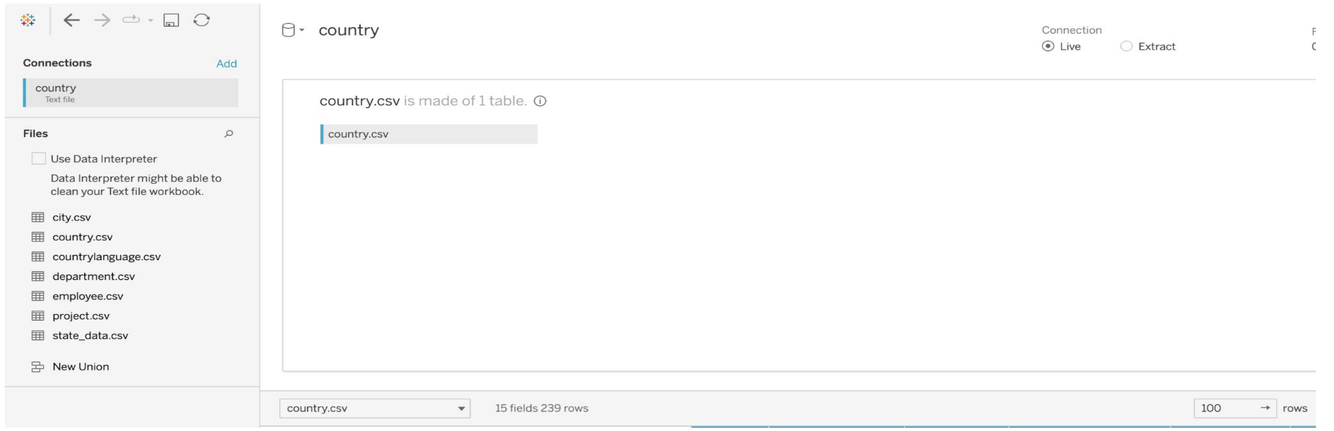


Joining tables using joins, cont'd.

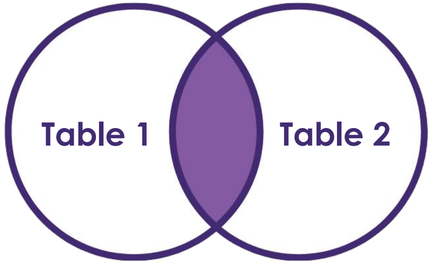
- To open the join canvas, select **Open** from the dropdown menu or **double-click** the first data source.

- In the join canvas, we can specify exactly how we want the sources to relate to one another based on four basic types of joins.

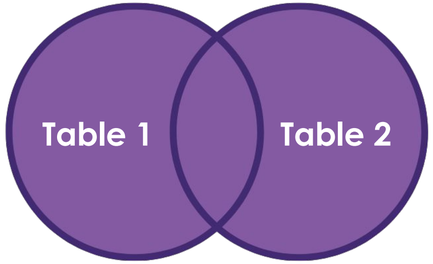




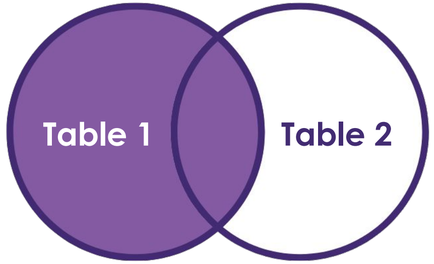
Types of joins



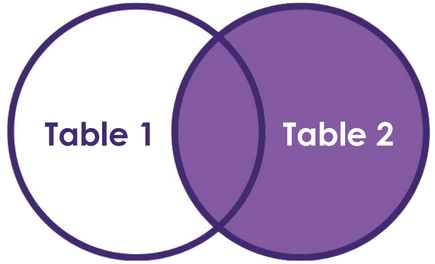
Inner join
includes matching records from both datasets



Full outer join
includes all records from both datasets



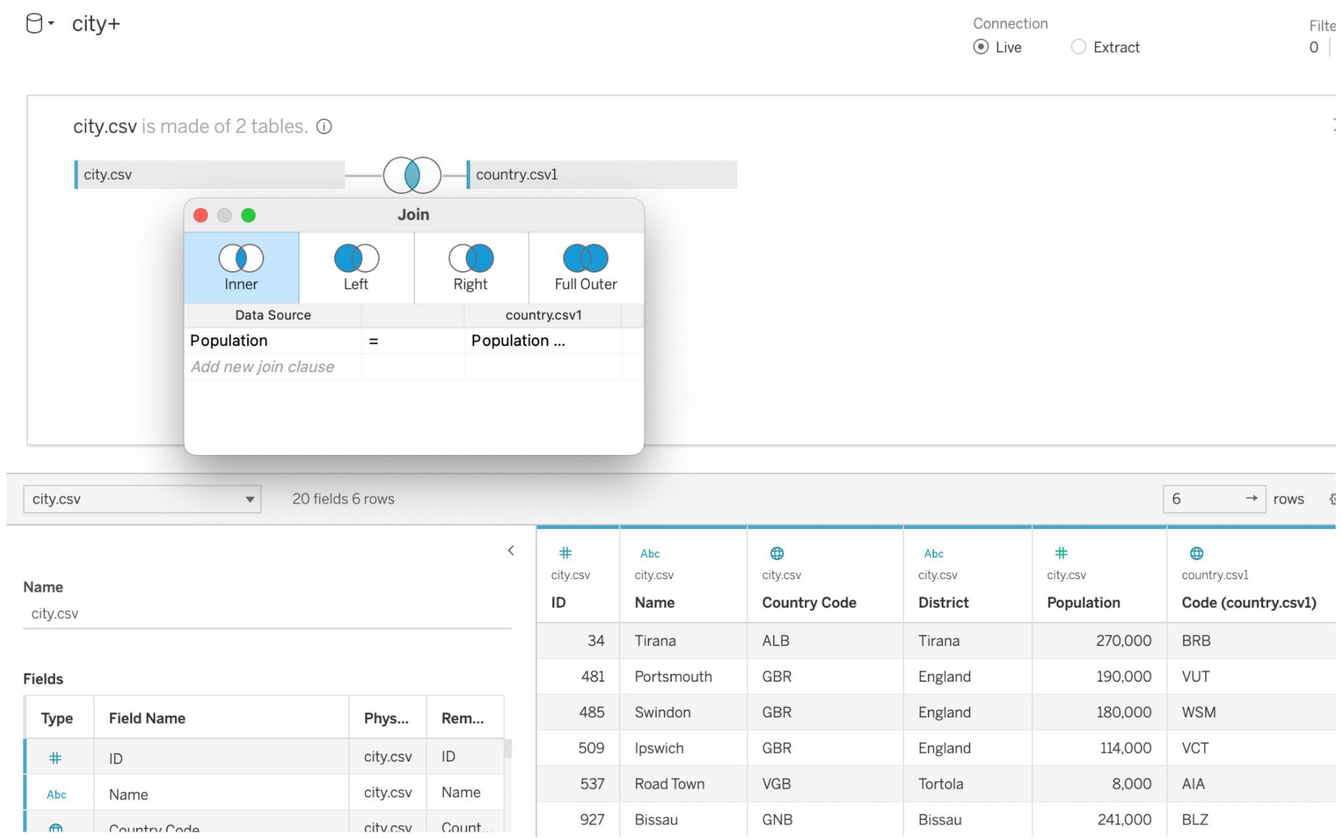
Left outer join
includes all records from left dataset and matching records from right dataset



Right outer join
includes all records from right dataset and matching records from left dataset

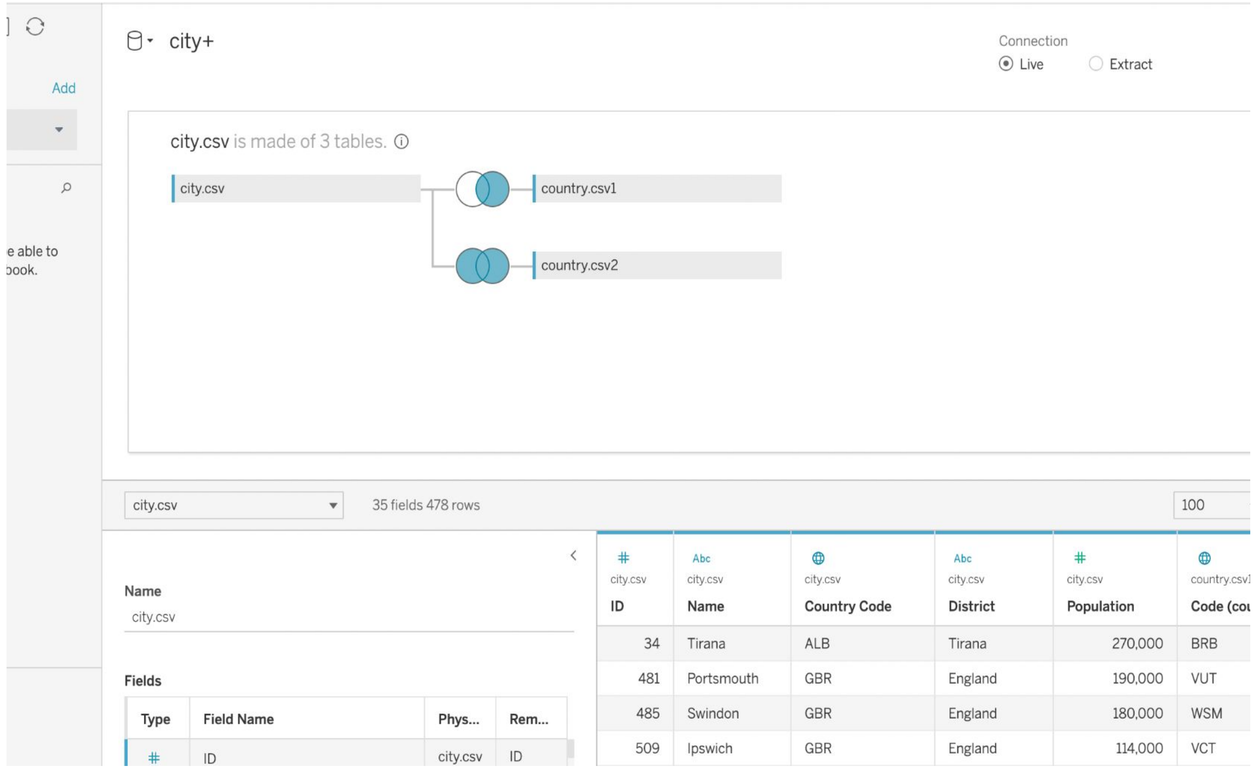
Best practices for using joins

- Start by joining the country and city tables.
- Here, inner join is the default join.
- Consider including left or right outer join to get all of the records in one database, even if there are no matching data points.
- Let's try all types of joins on the country and city dataset.



Join multiple datasets

- Now join the **countrylanguage** dataset.
- Does the order in which you import tables matter?
- Why did you choose that order?
- Why did you choose those types of joins?

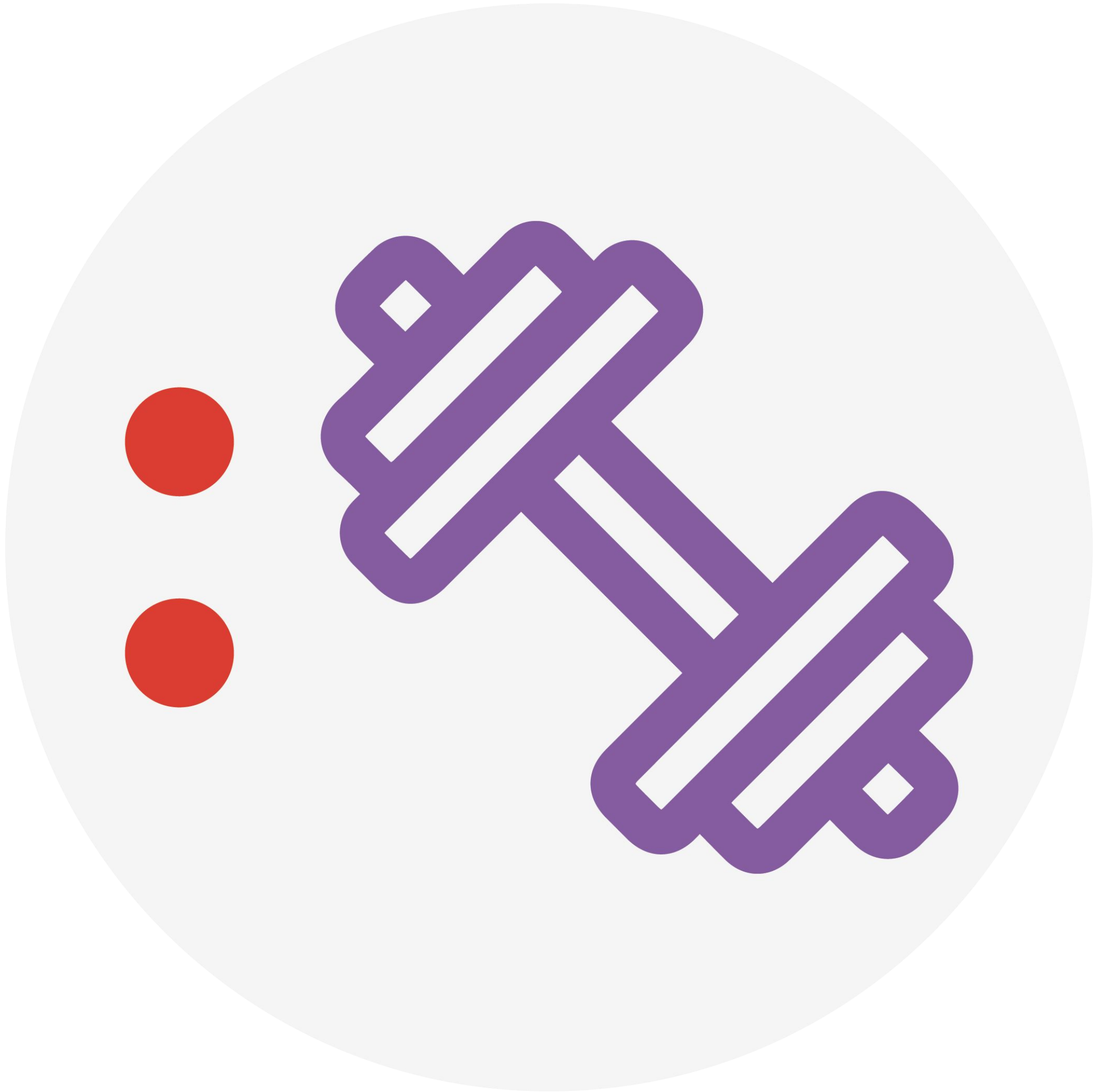


Relationships vs. joins

Relationships	Joins
Describes how two, independent, logical tables(in logical layer) are related to each other (the tables are not merged)	Combine two tables into one (in physical layer)
Maintains the same level of detail in the data sources	Will sometimes duplicate data stored at differing levels of detail
All measures are kept, even the ones that do not match	Some measures get filtered
Do not allow us to decide on the join type	Can select the way we went to join the data

Knowledge check 1

Exercise 1



Module completion checklist

Objective	Complete
Explain the need for Tableau and describe its features	✓
Describe how data sources connect to Tableau	✓
Import the given dataset into Tableau	✓
Explain the concept of Relationships	✓
Discuss the use of joins in Tableau	✓

End of Part 1