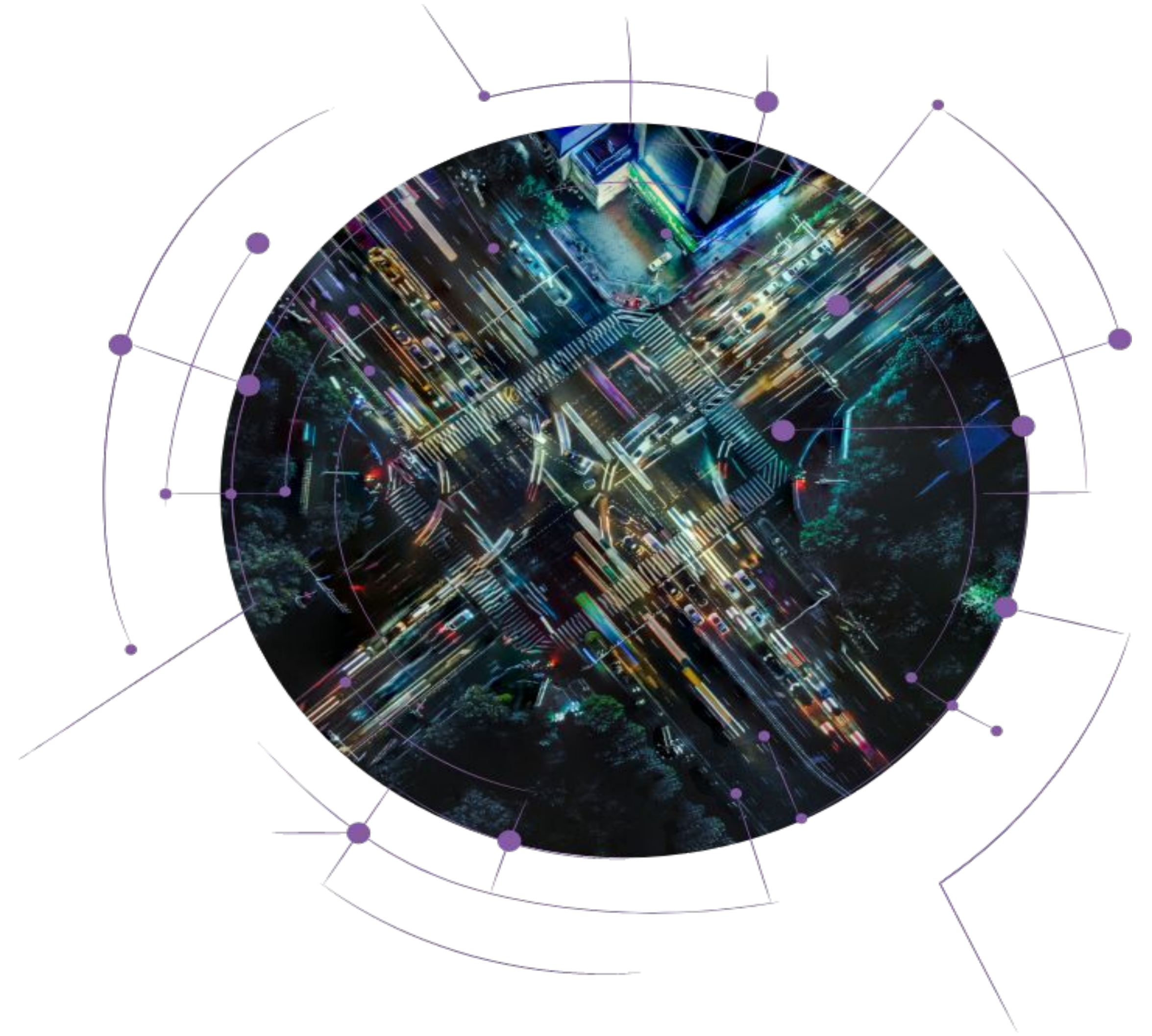


# DATA SOCIETY:

## Introduction to Tableau

Part 1



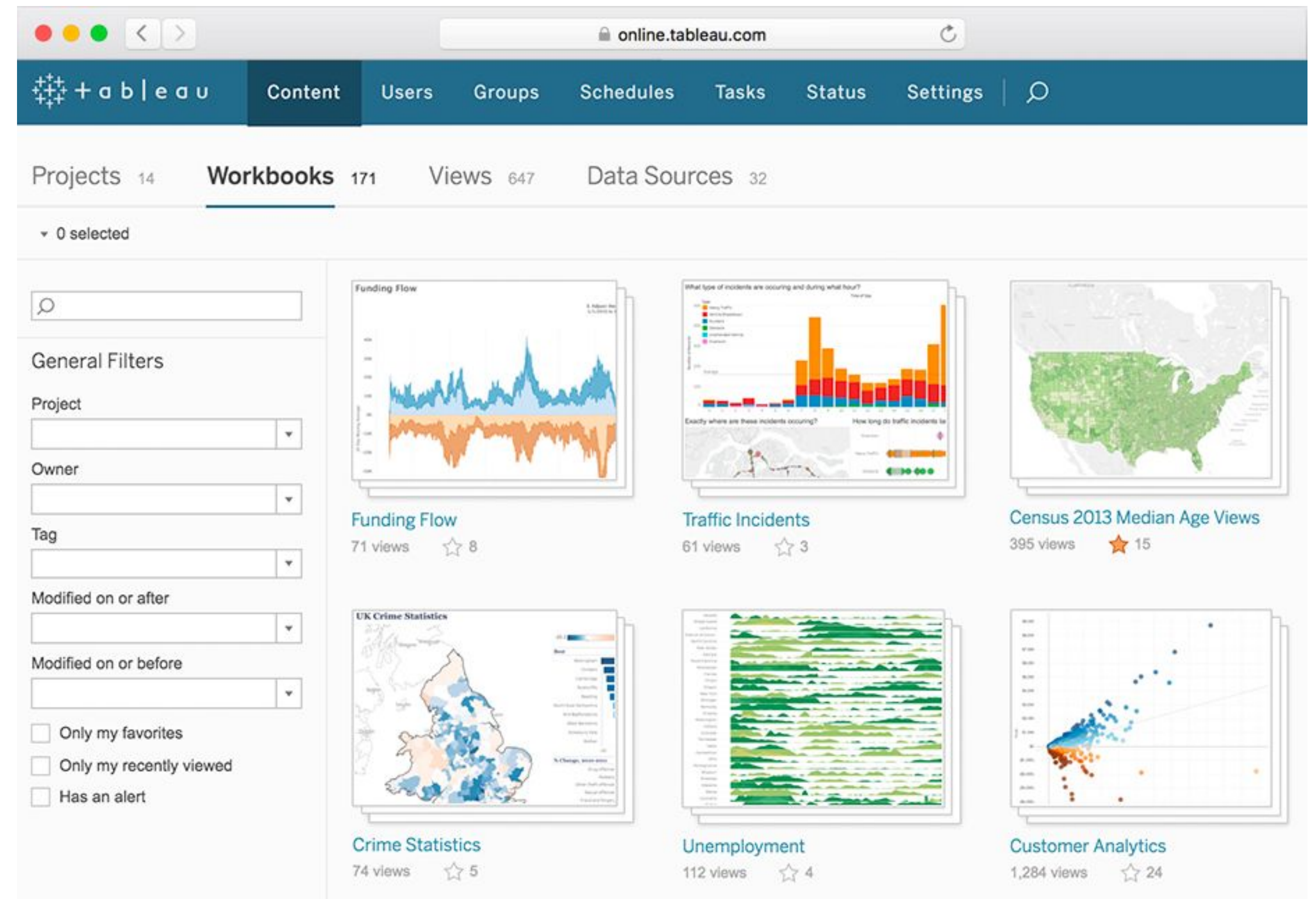
# Module completion checklist

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
<b>Purpose</b>	Spreadsheet application used for manipulating data	Visualization tool used for data analysis
<b>Usage</b>	Most suitable for statistical analysis of structured data	Most suitable for quick and easy representation of large datasets, which helps resolve big data issues
<b>Performance</b>	Moderate speed, with no option to speed up	Moderate speed, with options to optimize and enhance the progress of an operation
<b>Security</b>	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
<b>User Interface</b>	Macro and Visual Basic scripting knowledge required to maximize tool potential	Tool can be used without any coding knowledge
<b>Business Need</b>	Best for preparing one-off reports with small datasets	Best while working with big data
<b>Products</b>	Bundled with MS Office Tools	Comes with different versions, such as Tableau server, cloud, and desktop
<b>Integration</b>	Integrates with ~60 applications	Integrates with ~250 applications
<b>Real-time Data Exploration</b>	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](#).

The screenshot shows the Tableau website's product release page for Tableau Desktop 2020.2. The page features the Tableau logo and navigation links (Why Tableau, Products, Solutions, Resources, Partners, COVID-19) in the top left. On the top right, there are links for PRICING, SIGN IN, and a search icon, along with 'TRY NOW' and 'BUY NOW' buttons. The main heading is 'Tableau Desktop 2020.2' under the sub-header 'PRODUCT RELEASE AND DOWNLOAD'. A recommendation states: 'We recommend using the newest maintenance release of this version, 2020.2.10, which contains additional fixes.' Below this is a prominent 'DOWNLOAD TABLEAU DESKTOP 2020.2.10' button. A 'Download Link Notice' section at the bottom explains that the Windows version has been removed due to certificate issues and directs users to a Knowledge Base article for more details.

Tableau

Why Tableau Products Solutions **Resources** Partners COVID-19

PRICING SIGN IN

TRY NOW BUY NOW

PRODUCT RELEASE AND DOWNLOAD

## Tableau Desktop 2020.2

We recommend using the newest maintenance release of this version, [2020.2.10](#), which contains additional fixes.

[DOWNLOAD TABLEAU DESKTOP 2020.2.10](#)

### Download Link Notice

The Windows version has been removed due to an issue with certificates preventing it from being able to install correctly.  
For more details, please see [this Knowledge Base article](#).

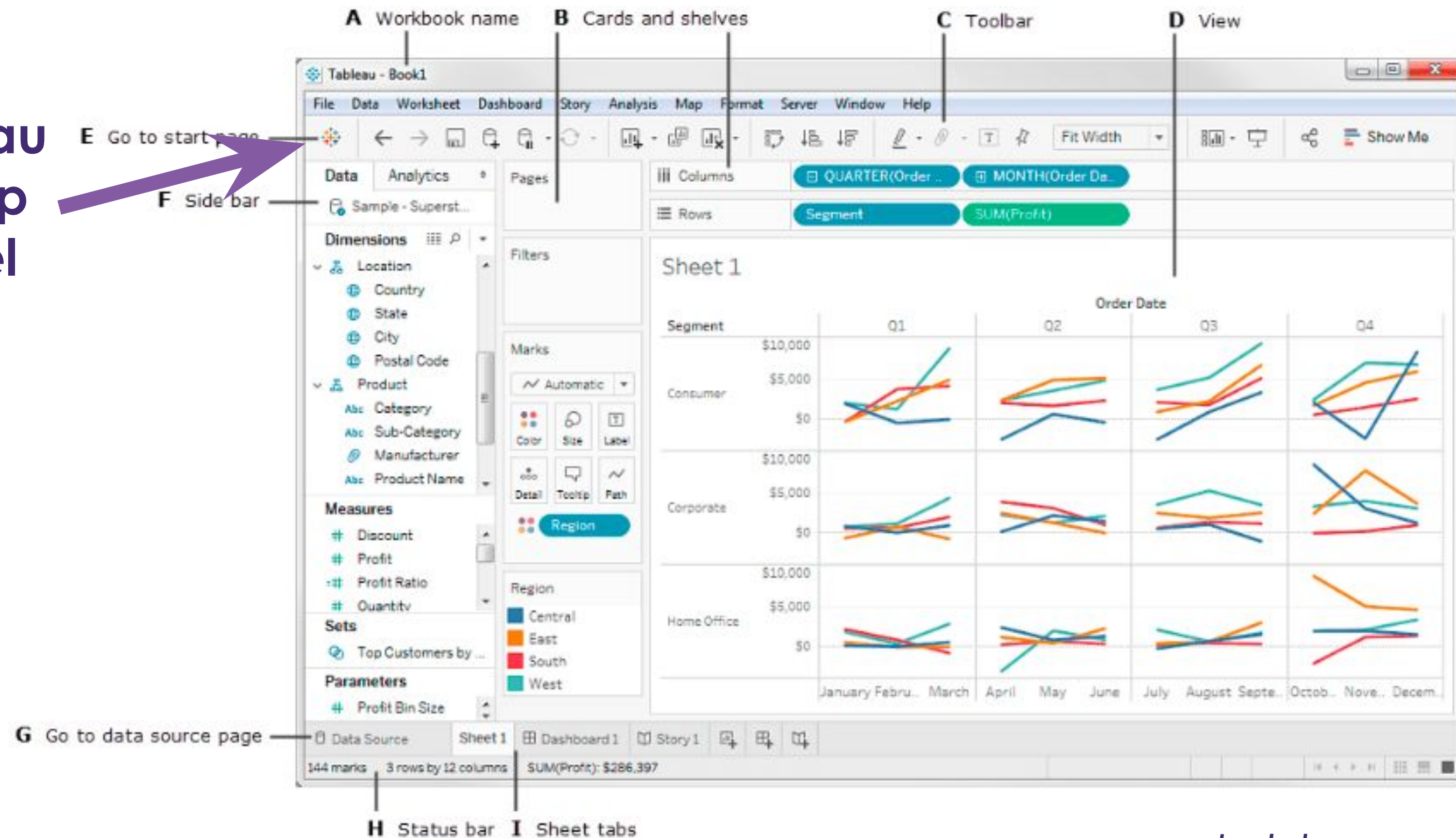
Build number

Release date

Product support

# Tableau overview

Click the Tableau logo to bring up Connect panel



source: [tableau.com](https://tableau.com)

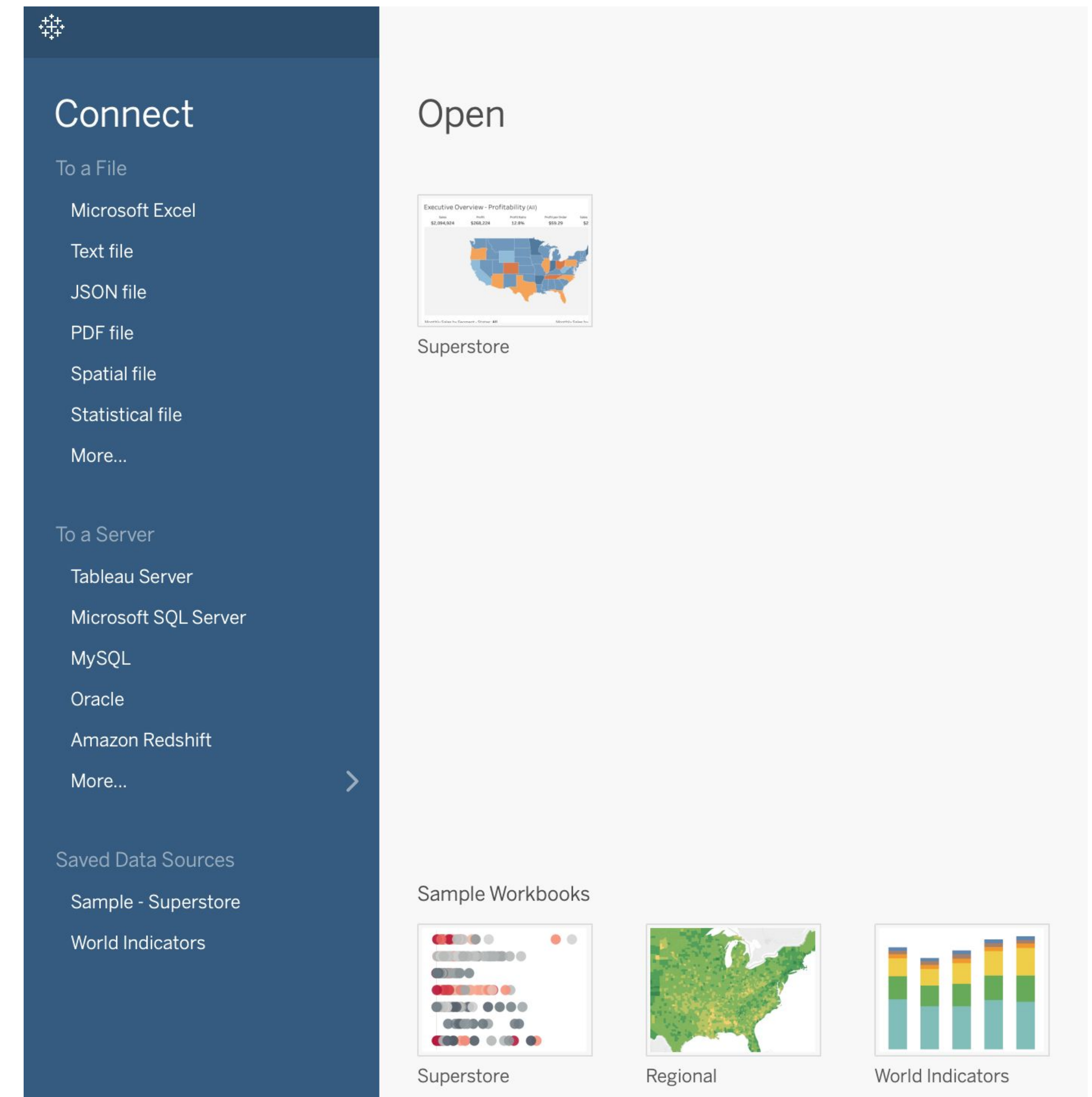


# Module completion checklist

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	

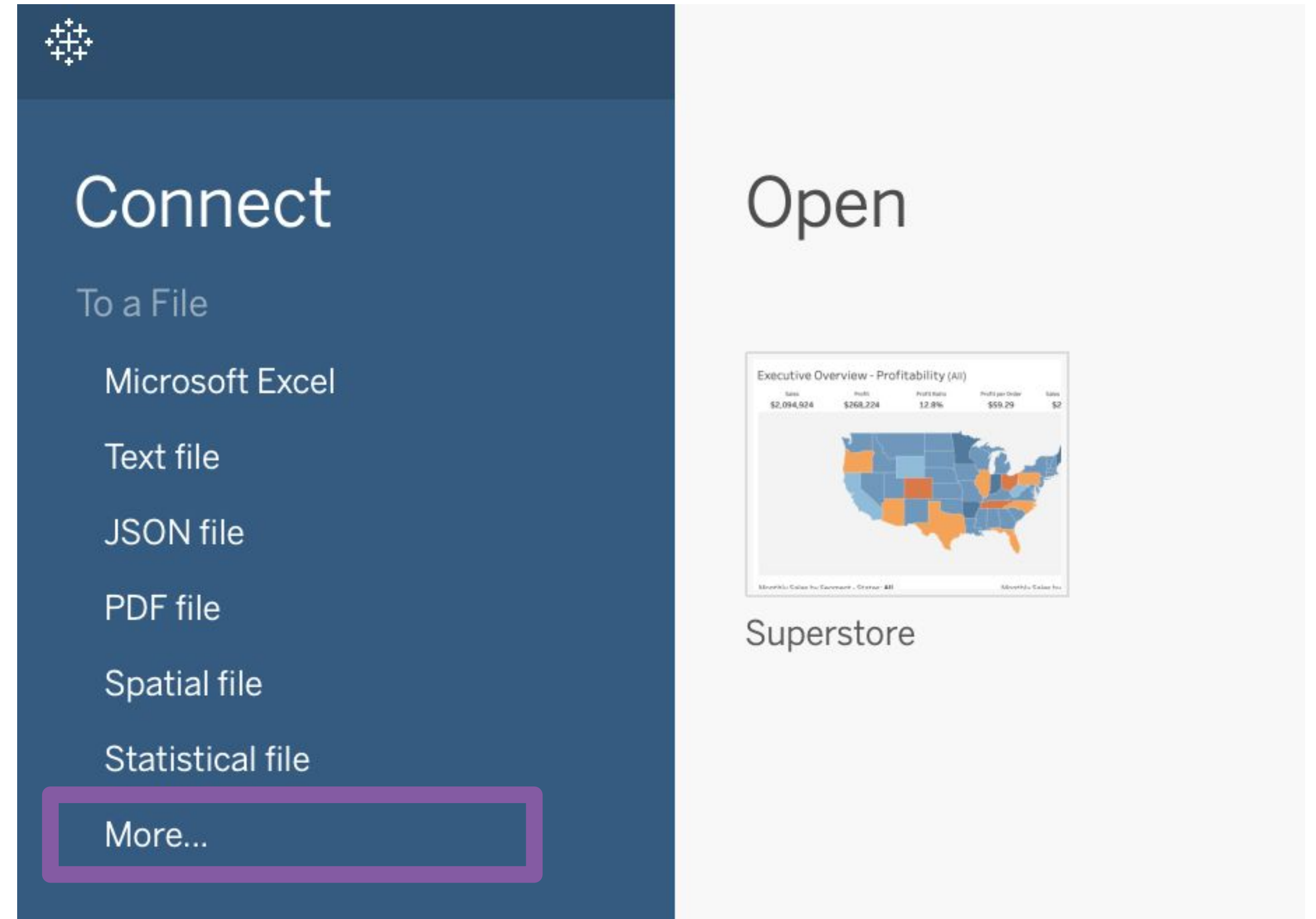
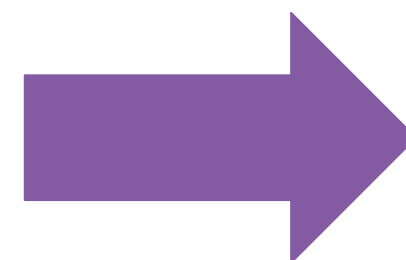
# Importing data

- 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.

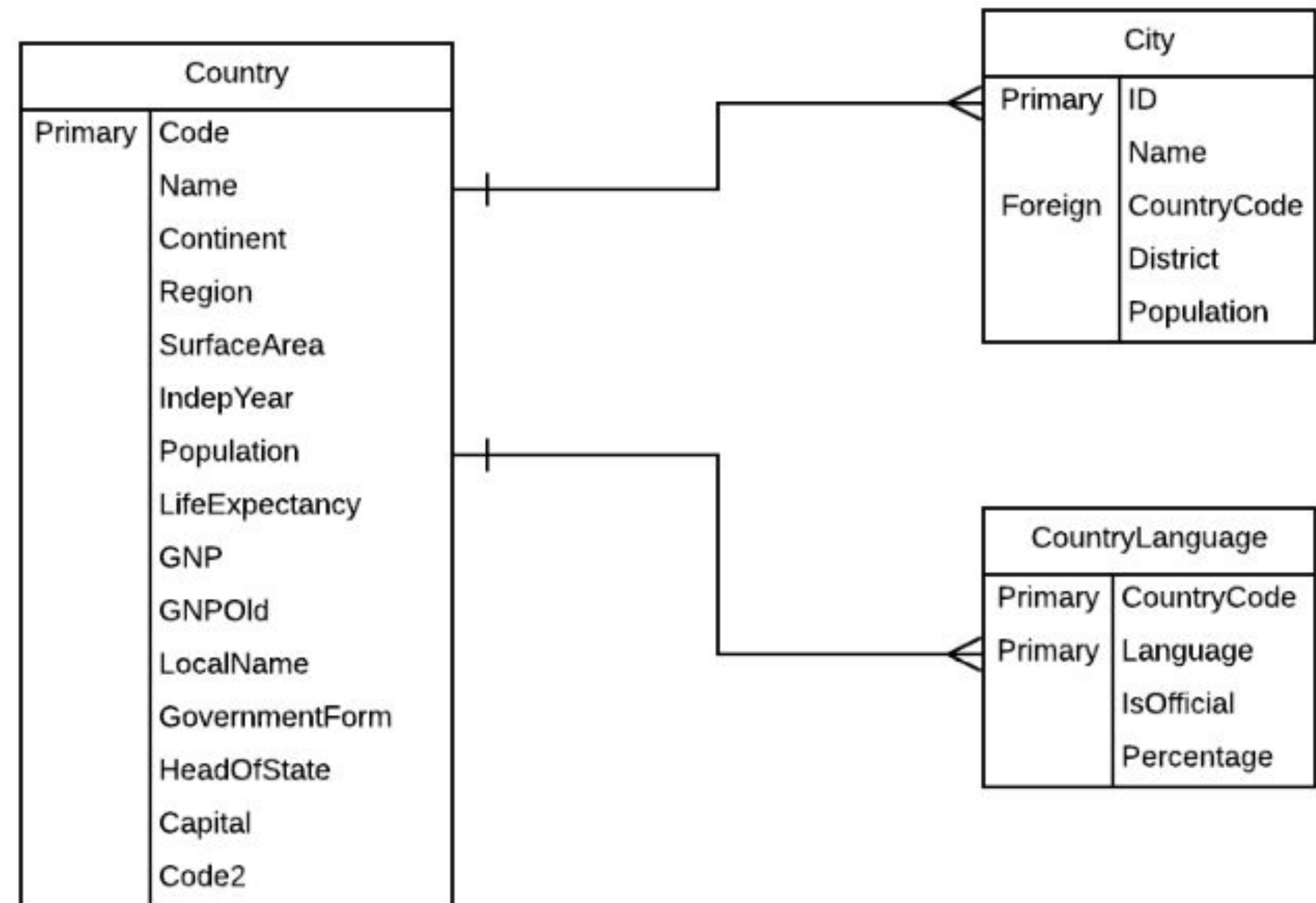




# World database

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

World Database ERD



# 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	

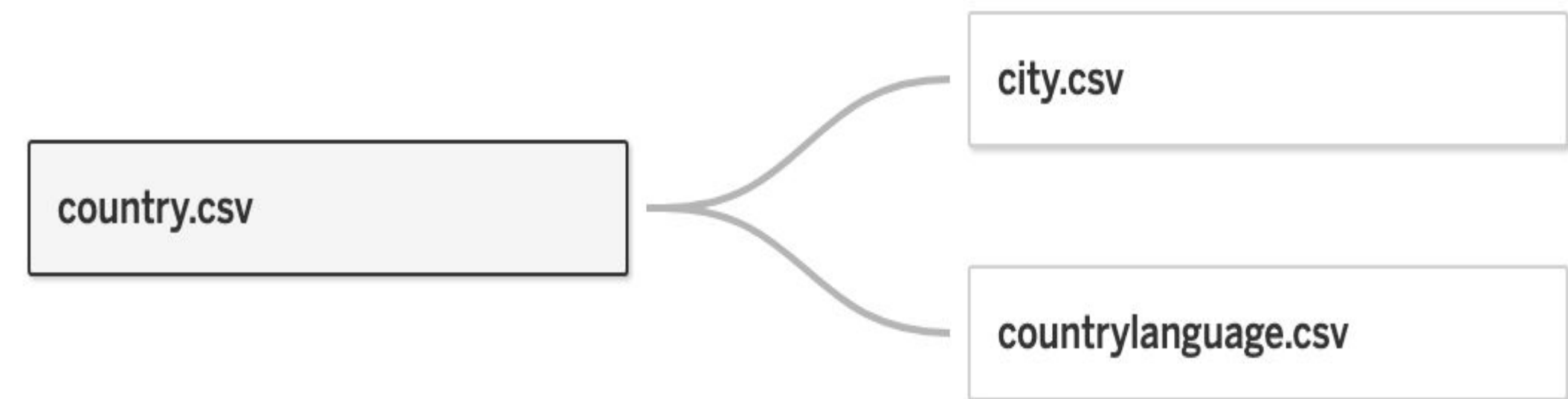
# Joining datasets using Relationships

- 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.



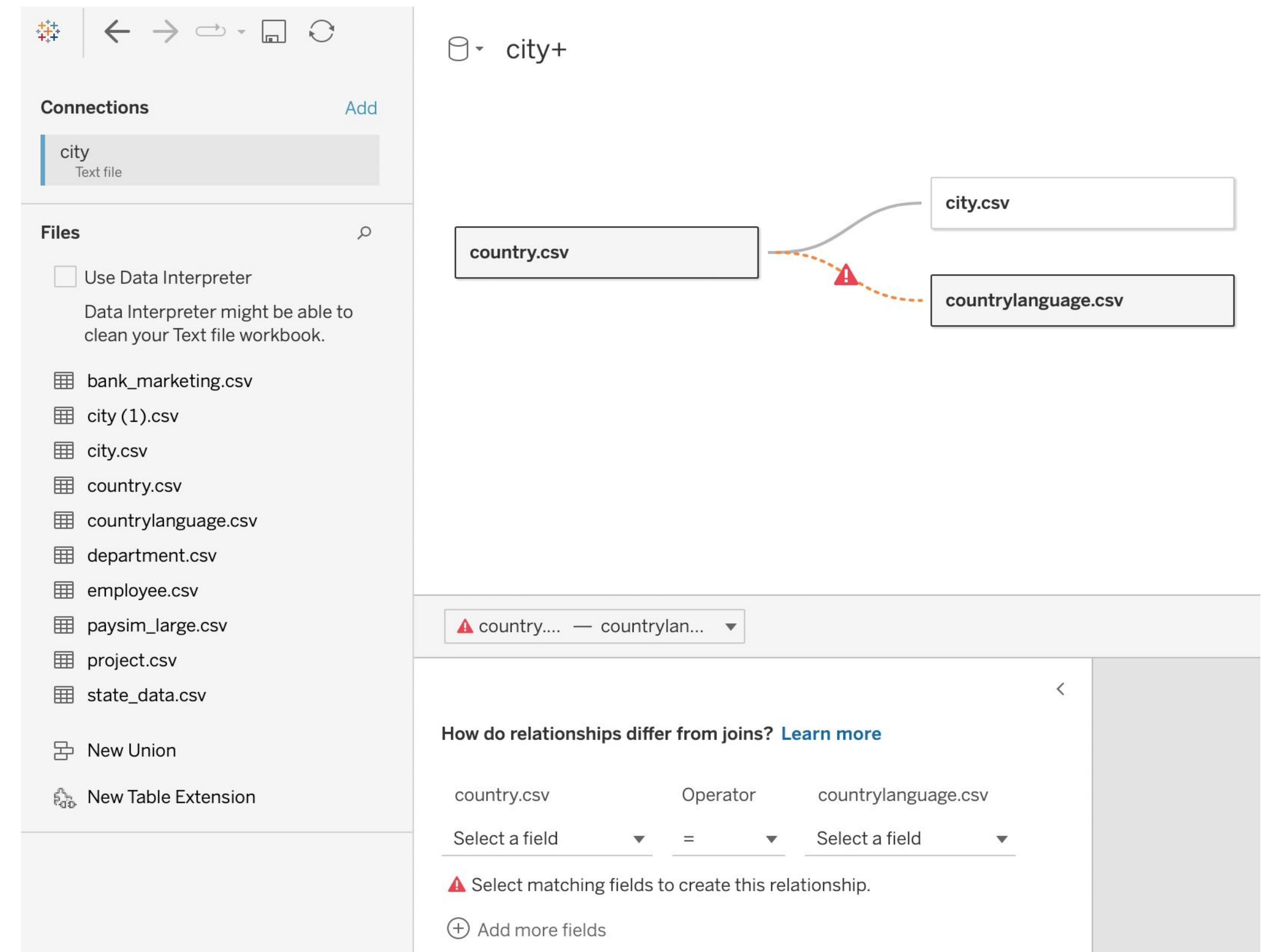
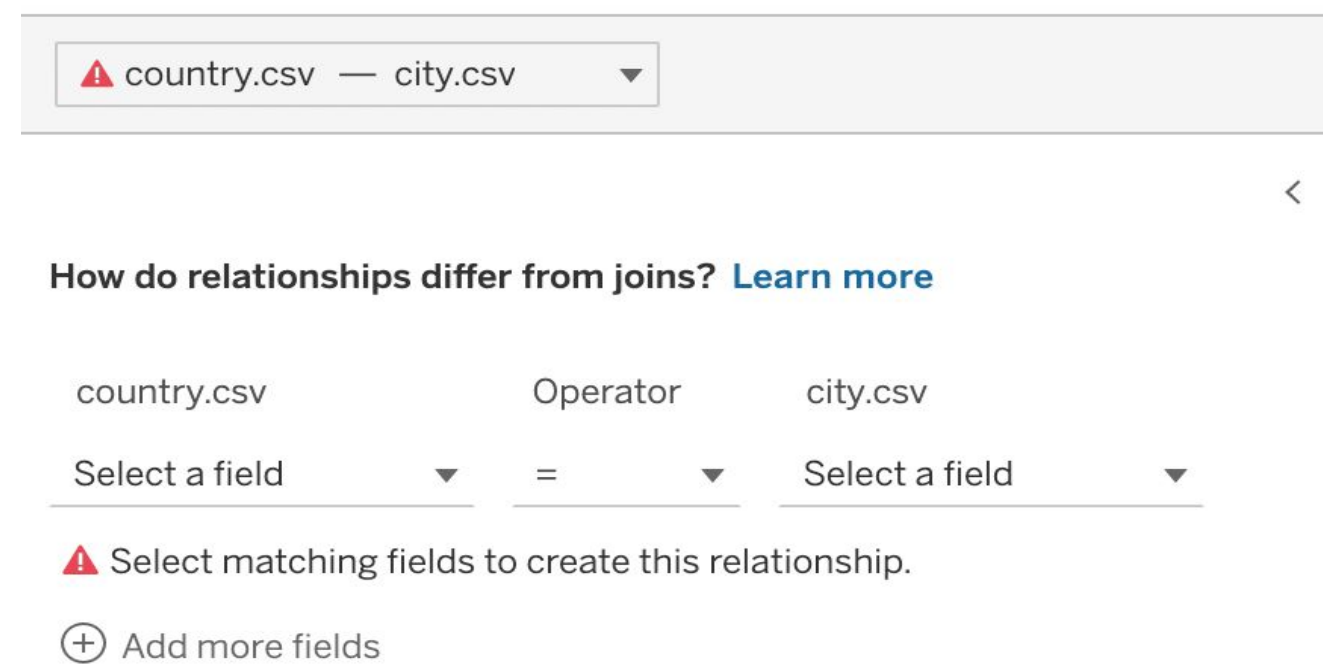
# Relationships are contextual joins

- 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).



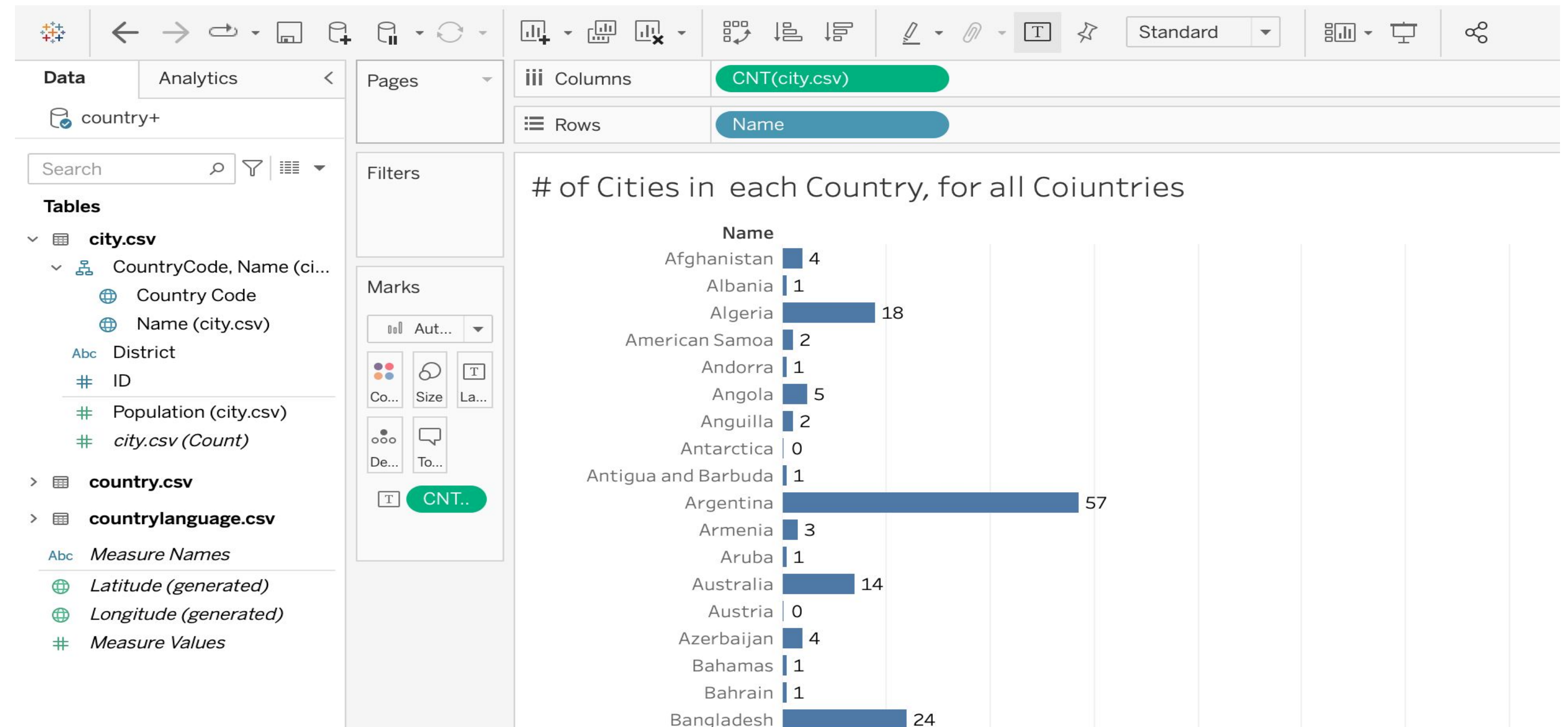
# Select fields to create Relationships

- 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

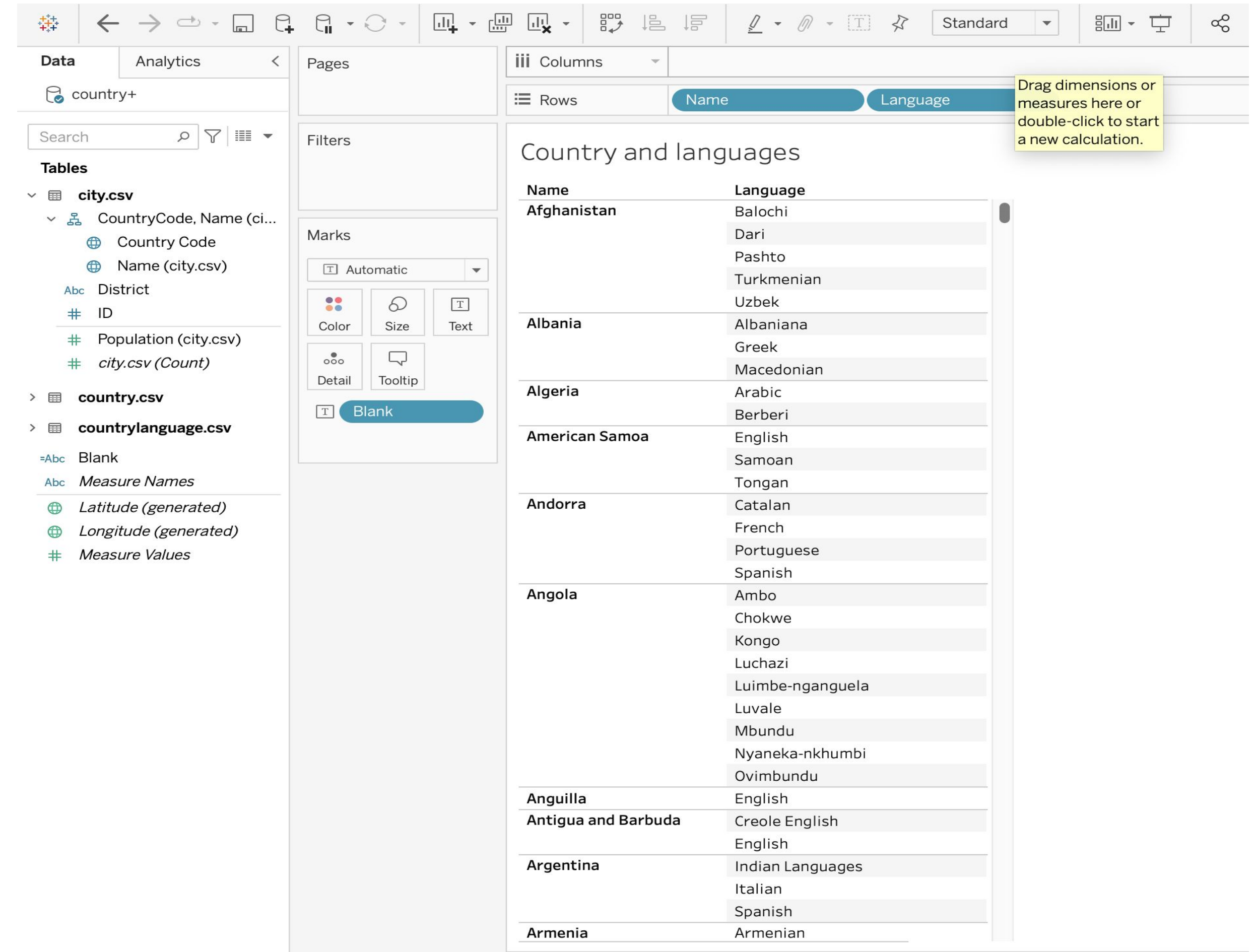
- Relationships allow for unmatched entries in one dataset to display as 0s.
- Visualizing the count of cities by country shows **all** countries, even those without any corresponding cities.
- Displaying all countries helps to avoid undercounting the number of countries.





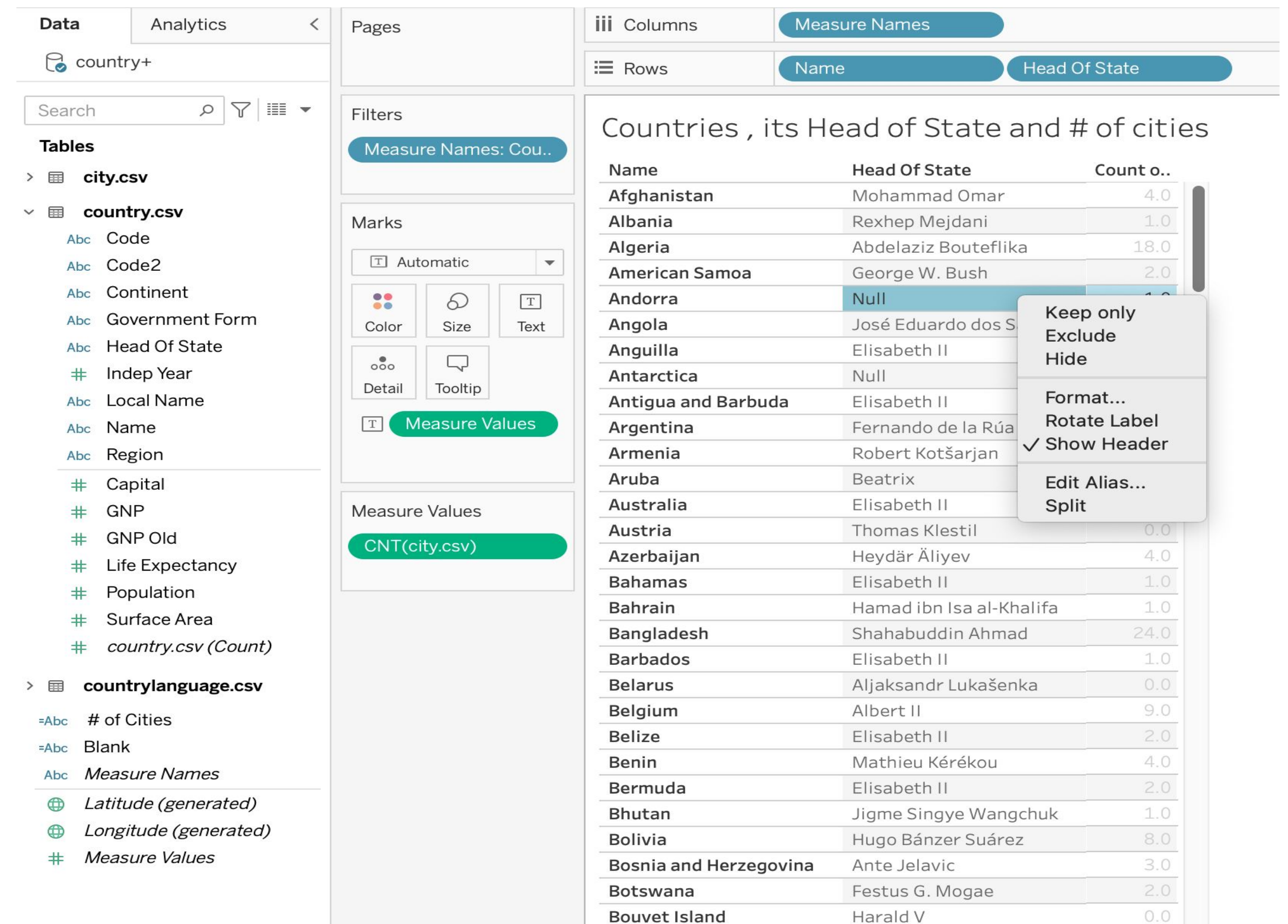
# Relevant dimensions across tables

- 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.



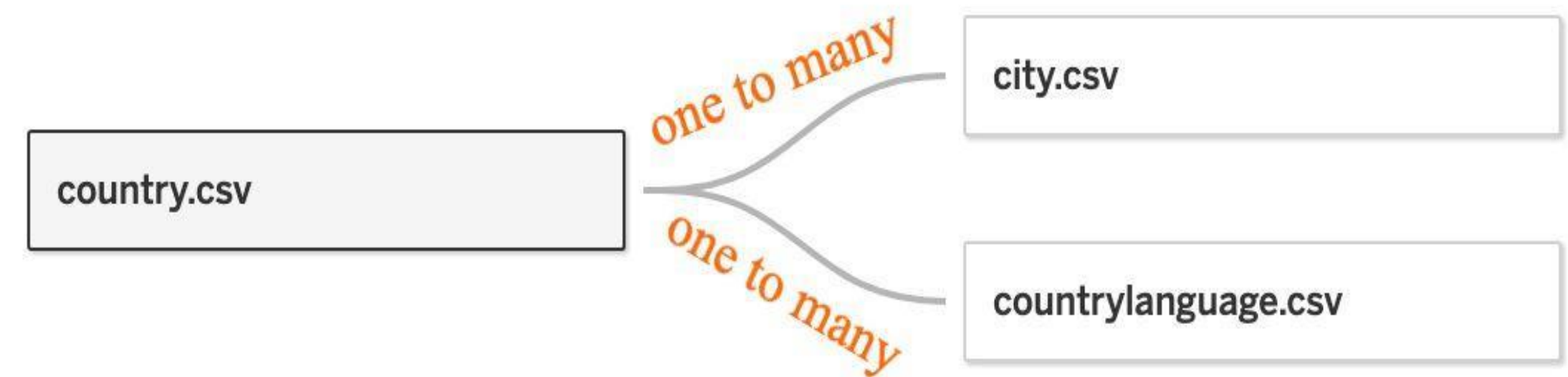
# Retaining unmatched values

- A **null** appears to represent a country 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.

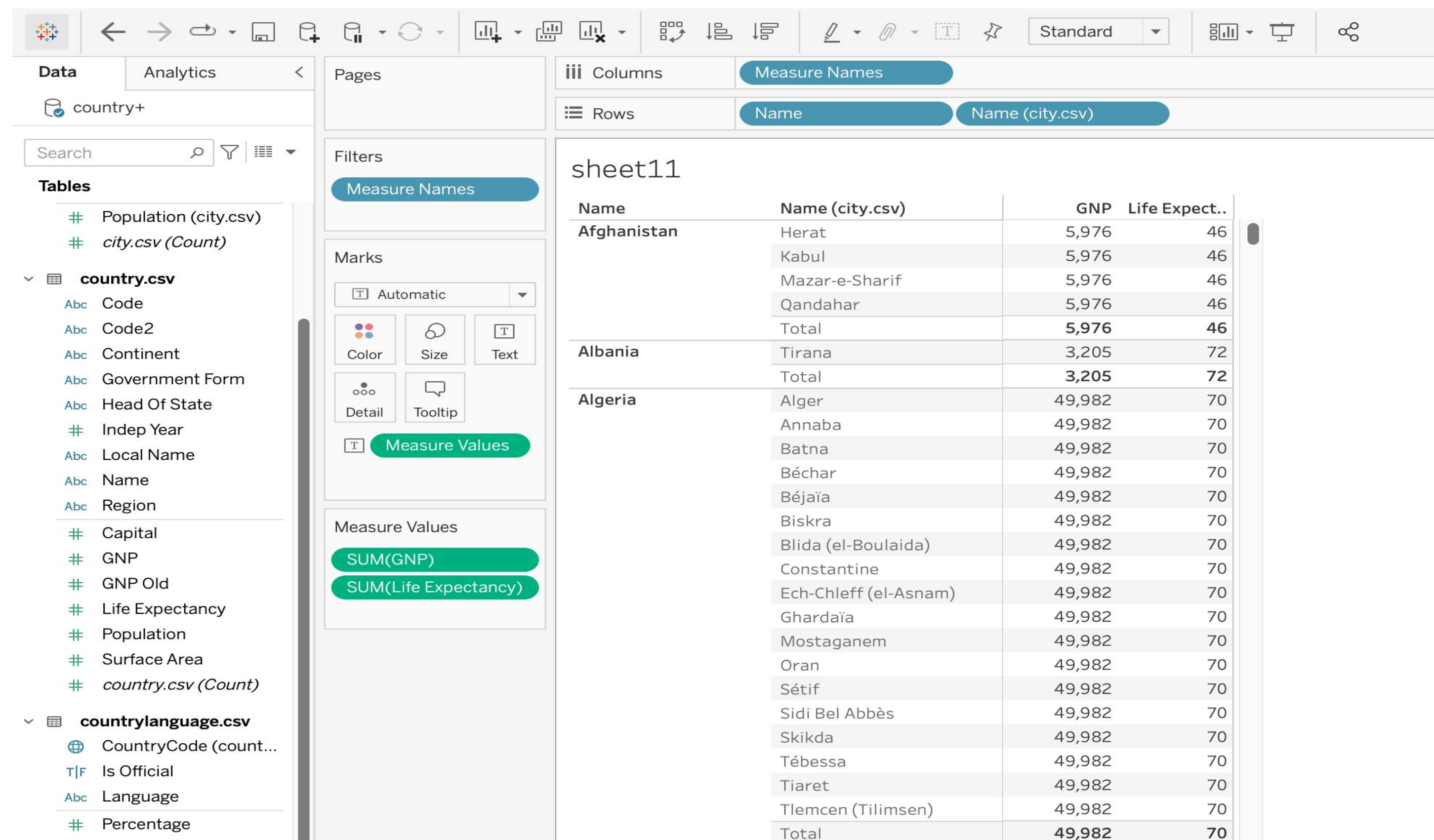






# Relationships and level of detail, cont'd.

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



# 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	

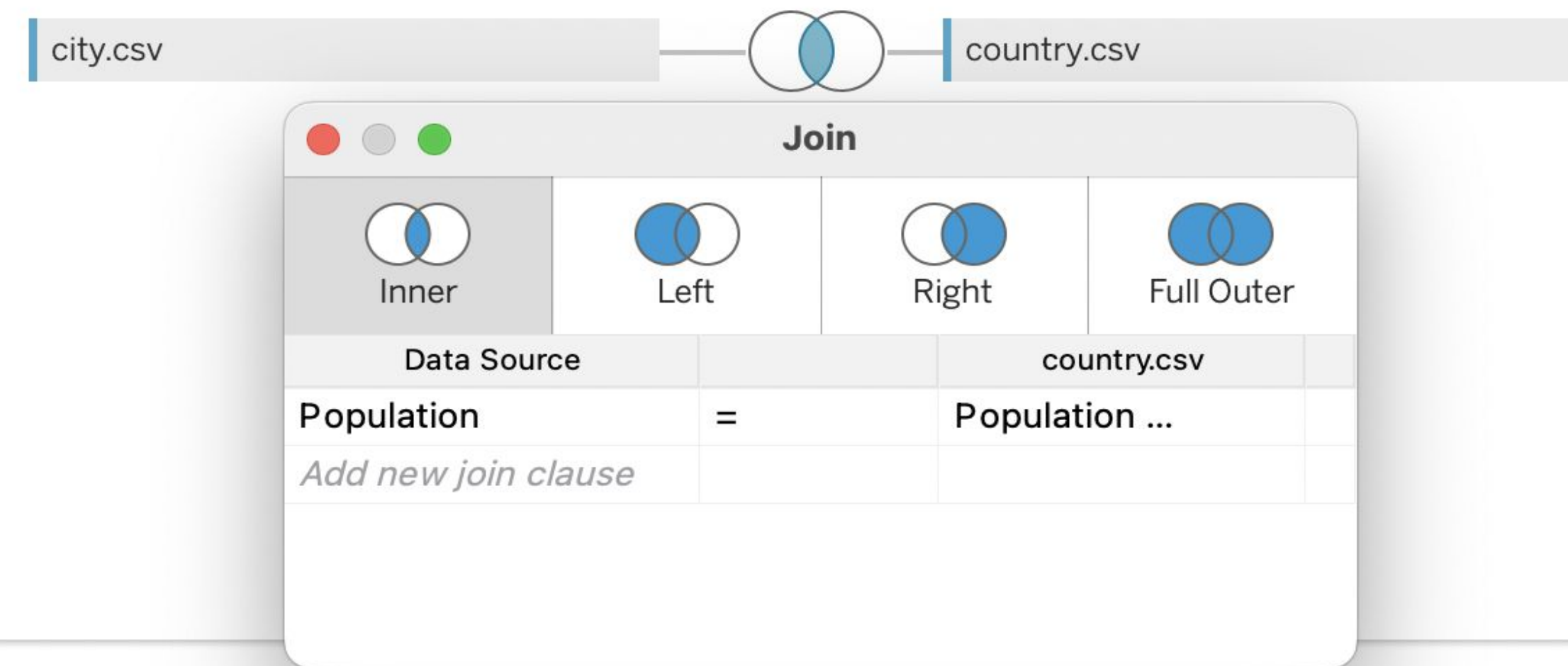


# Manually joining datasets

- 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.

city.csv+ (Multiple Connections)

city.csv is made of 2 tables. ⓘ



# 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**.

The screenshot displays the Tableau interface for a 'country.csv' data source. On the left, the 'Connections' pane shows 'country' as a 'Text file'. Below it, the 'Files' pane lists several CSV files: city.csv, country.csv, countrylanguage.csv, department.csv, employee.csv, project.csv, and state\_data.csv. A 'New Union' button is at the bottom of this list. The main area shows the 'country.csv' data source with a dropdown menu indicating '15 fields 239 rows'. Below this, the 'Name' field is set to 'country.csv'. The 'Fields' section contains a table with columns: Type, Field Name, Physical Table, and Remote Fiel... (truncated). The table lists fields: Code, Name, Continent, Region, and Surface Area, all mapped to 'country.csv'. To the right of the 'Fields' table, there is a preview of the data with columns: Code, Name, and Continent. The preview shows rows for ABW (Aruba, North America), AFG (Afghanistan, Asia), AGO (Angola, Africa), AIA (Anguilla, North America), ALB (Albania, Europe), AND (Andorra, Europe), ANT (Netherlands Antilles, North America), and ARE (United Arab Emirates, Asia). At the bottom, there is a 'Go to Worksheet' button and a 'Data Source' tab with 'Sheet 1' selected.

country

country.csv

country.csv

15 fields 239 rows

Name  
country.csv

Fields

Type	Field Name	Physical Table	Remote Fiel...
Abc	Code	country.csv	Code
Abc	Name	country.csv	Name
Abc	Continent	country.csv	Continent
Abc	Region	country.csv	Region
#	Surface Area	country.csv	SurfaceArea

Abc country.csv  
Code

Abc country.csv  
Name

Abc country.csv  
Continent

ABW Aruba North America

AFG Afghanistan Asia

AGO Angola Africa

AIA Anguilla North America

ALB Albania Europe

AND Andorra Europe

ANT Netherlands Antilles North America

ARE United Arab Emirates Asia

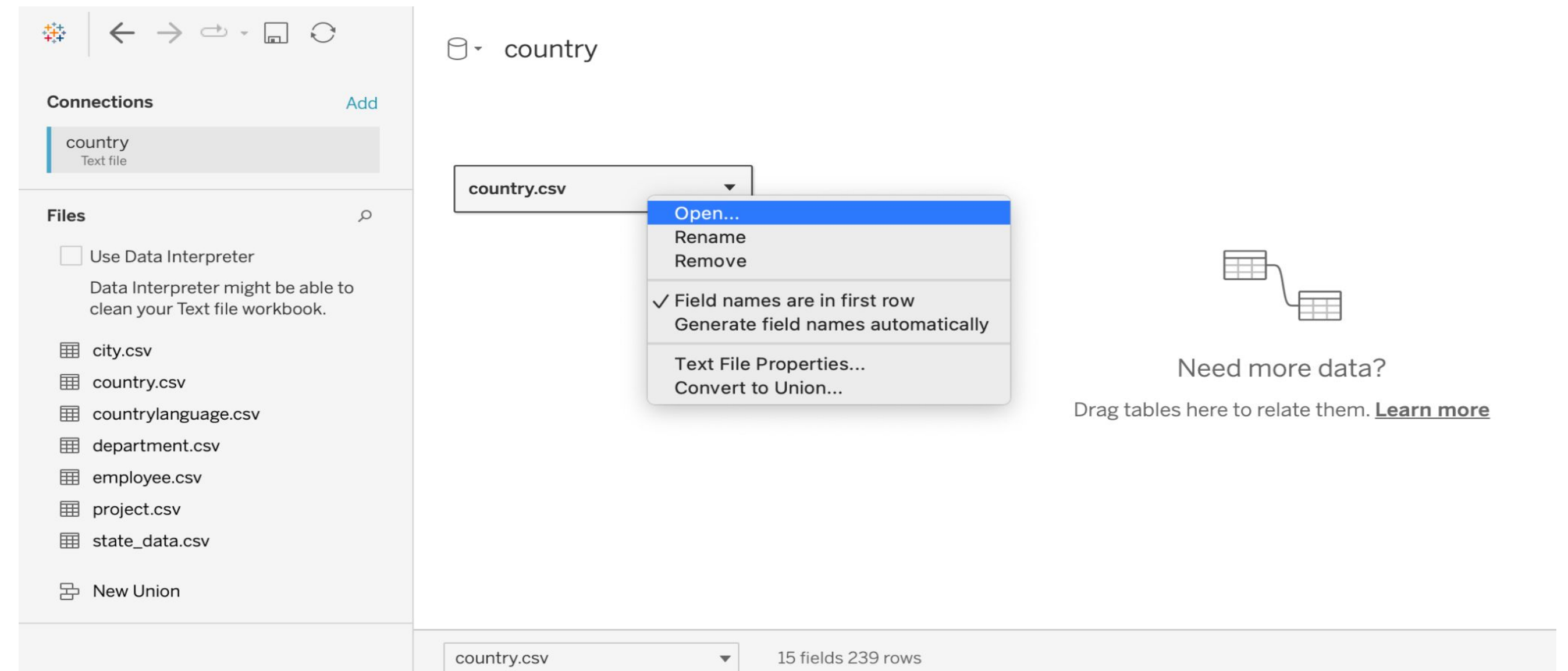
Go to Worksheet

Data Source Sheet 1

Need more data?  
Drag tables here to relate them. [Learn more](#)

# 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.



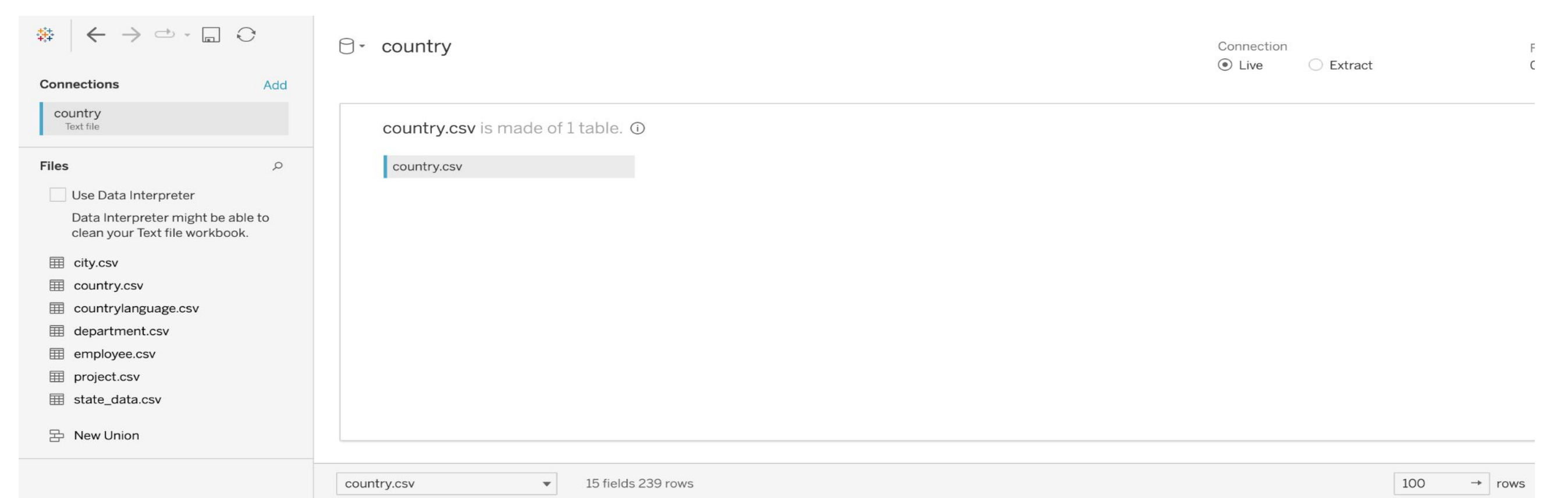
country

country.csv

Open...  
Rename  
Remove  
✓ Field names are in first row  
Generate field names automatically  
Text File Properties...  
Convert to Union...

Need more data?  
Drag tables here to relate them. [Learn more](#)

country.csv 15 fields 239 rows



country

country.csv is made of 1 table. ⓘ

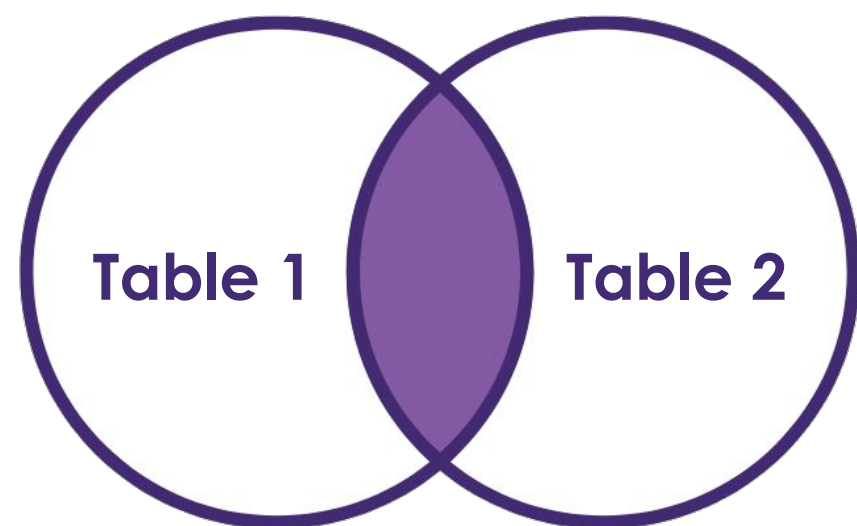
country.csv

country.csv 15 fields 239 rows

100 rows

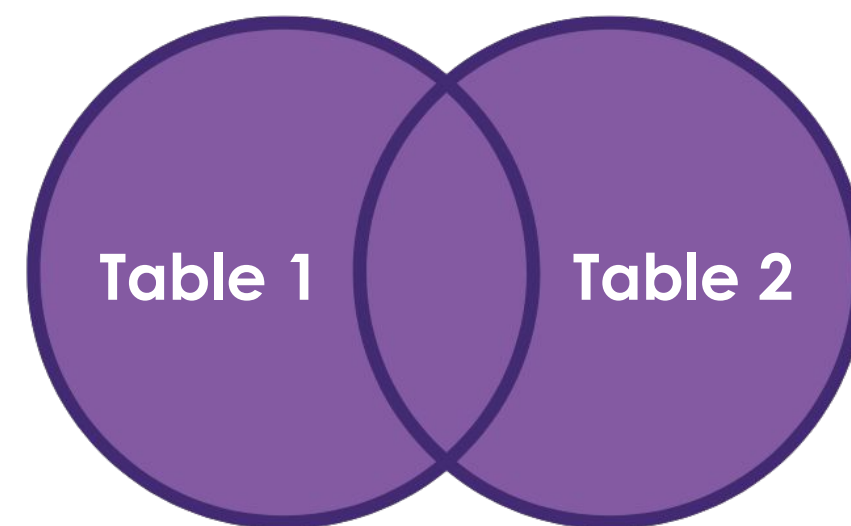


# Types of joins



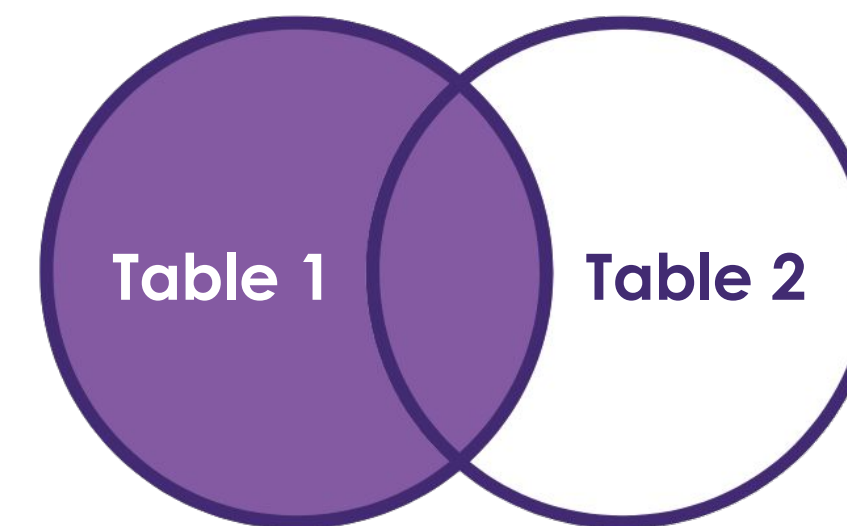
## Inner join

includes matching records from both datasets



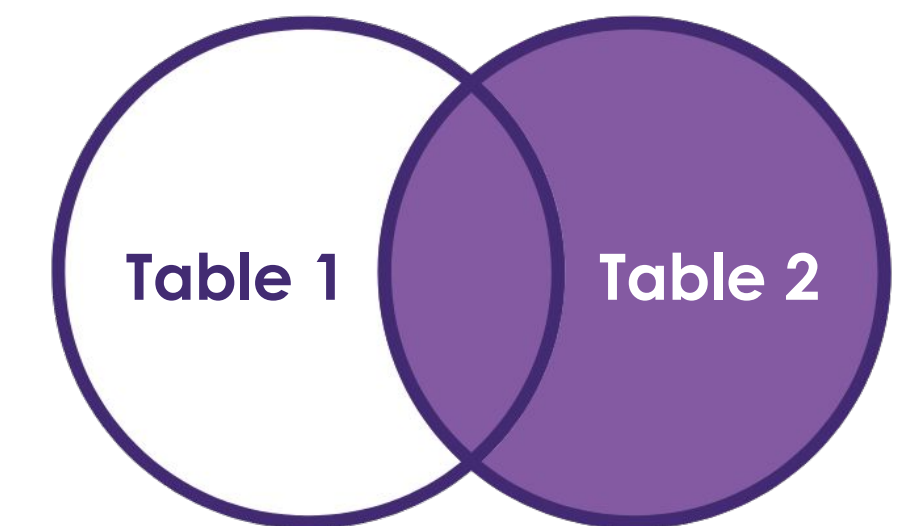
## Full outer join

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.

city+

Connection ☒ Live ☐ Extract Filter 0

city.csv is made of 2 tables. ⓘ

city.csv — country.csv1

**Join**

☒ Inner ☐ Left ☐ Right ☐ Full Outer

Data Source		country.csv1
Population	=	Population ...
<a href="#">Add new join clause</a>		

city.csv 20 fields 6 rows 6 rows

#	city.csv	city.csv	city.csv	city.csv	city.csv	country.csv1
ID	Name	Country Code	District	Population	Code (country.csv1)	
34	Tirana	ALB	Tirana	270,000	BRB	
481	Portsmouth	GBR	England	190,000	VUT	
485	Swindon	GBR	England	180,000	WSM	
509	Ipswich	GBR	England	114,000	VCT	
537	Road Town	VGB	Tortola	8,000	AIA	
927	Bissau	GNB	Bissau	241,000	BLZ	

Type	Field Name	Phys...	Rem...
#	ID	city.csv	ID
Abc	Name	city.csv	Name
🌐	Country Code	city.csv	Count...





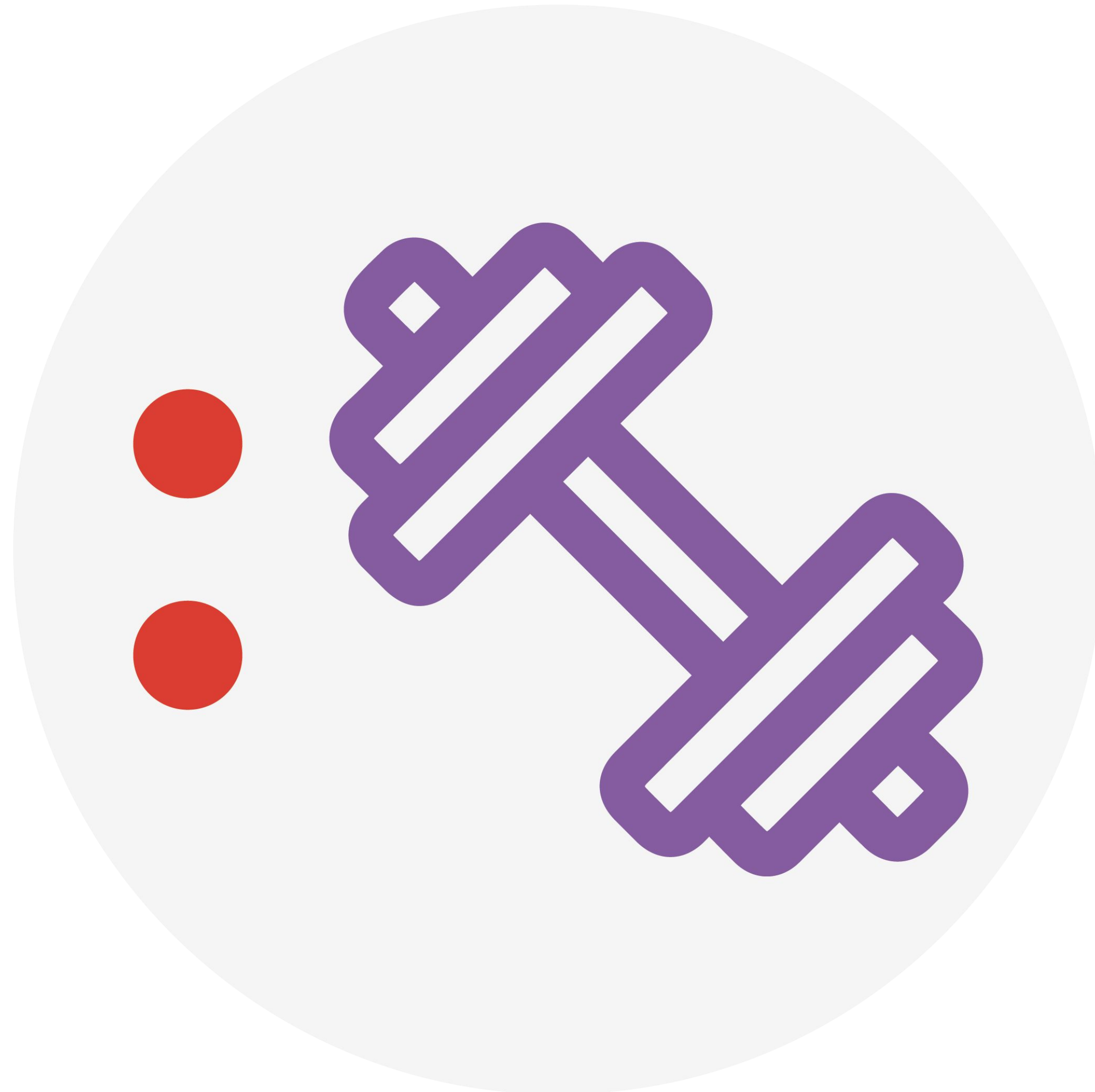
# Relationships vs. joins

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 want 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

