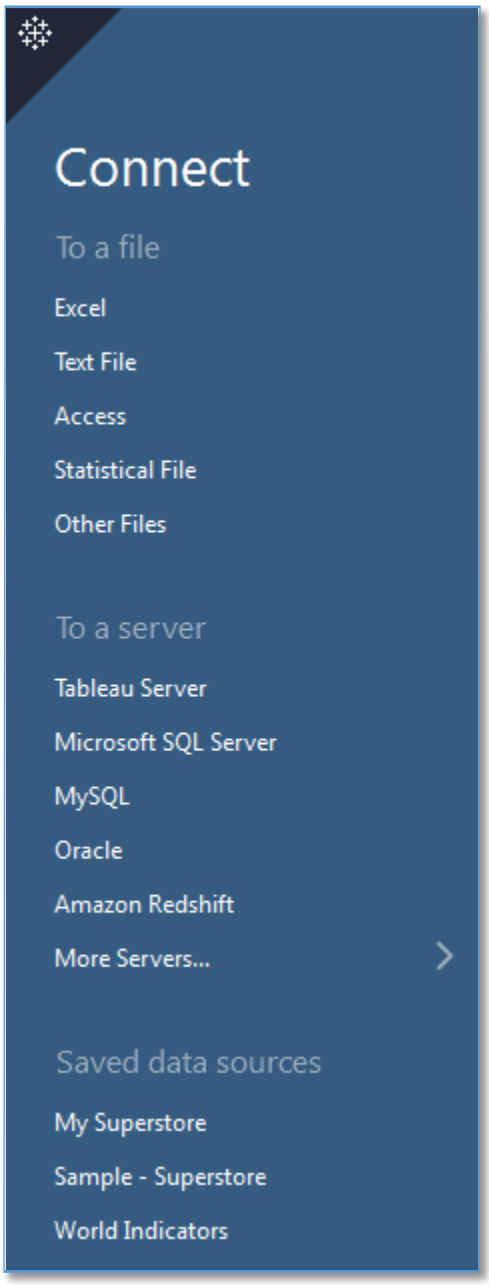
Understanding Tableau Data Sources

A data source is a reusable connection to data including connections to relational databases, cloud-hosted databases, spreadsheets, and more. When using Tableau Desktop, you may select data from a variety of locations including your local computer, server hosted or even from the cloud. In order to share a data source with others, you would first connect to it in Tableau Desktop and then publish it to the Tableau Server. The published data sources can include data or connections to live databases. The published data sources can also include layers of customizations including calculations, groups, and sets.



You should publish a data source when you want users to connect

to

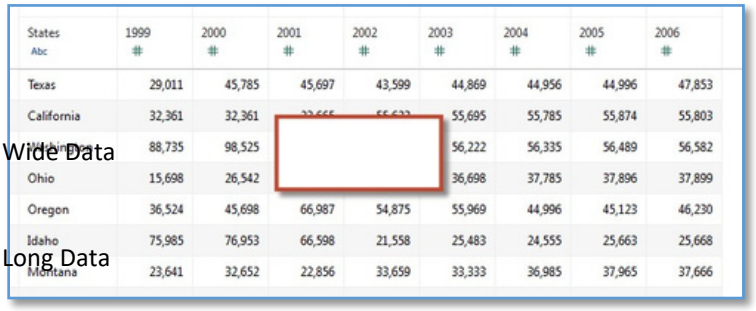
the same data source from multiple workbooks. When a published data source is refreshed, workbooks using the source will reflect the changes. This facilitates consistency and accuracy in reporting.

It is often useful to create a data extract in Tableau. This pulls the data from a data source into a stand-alone data set for use in Tableau Desktop. This is especially useful if performance is an issue or if you need to distribute to users without access to the data. It is important to note, however, that you cannot create an extract of a cube data source.

If you only want users to connect to a data source from a single workbook, you should embed the data source in a workbook but do not publish the source. Every published workbook has at least one embedded data source.

Tableau Data Format

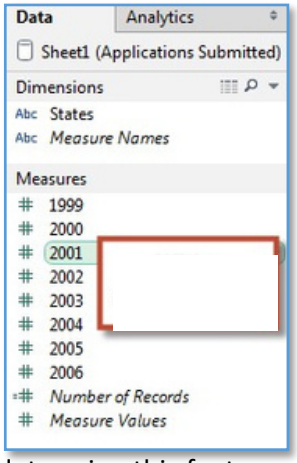
Tableau can work with data in varying formats. However to get the most out of Tableau, it is best if your data is at a detail level or "long" with a separate record for every element you are counting as opposed to compressed or "wide" with many details for each element.



States	1999	2000	2001	2002	2003	2004	2005	2006
Texas	29,011	45,785	45,697	43,599	44,869	44,956	44,996	47,853
California	32,361	32,361			55,695	55,785	55,874	55,803
Washington	88,735	98,525			56,222	56,335	56,489	56,582
Ohio	15,698	26,542			36,698	37,785	37,896	37,899
Oregon	36,524	45,698	66,987	54,875	55,969	44,996	45,123	46,230
Idaho	75,985	76,953	66,598	21,558	25,483	24,555	25,663	25,668
Montana	23,641	32,652	22,856	33,659	33,333	36,985	37,965	37,666

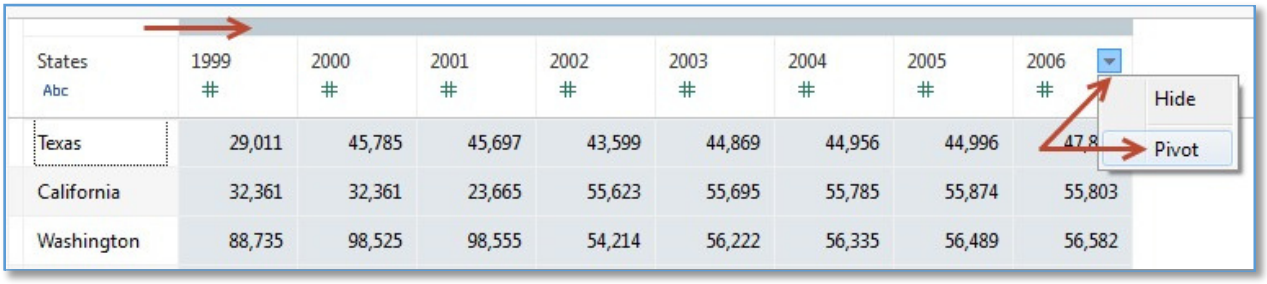
Wide Data

Long Data



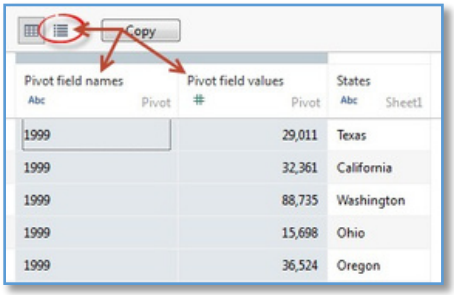
While you can work with "wide" data in Tableau, you will have reduced reporting capabilities.

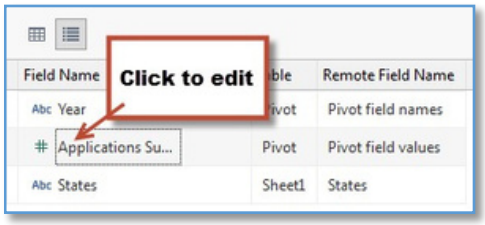
If you're starting with "wide" data, you do have the option to **Pivot** your data using this feature on the Data Connection page. To do this, select the columns you wish to transpose, hover over one of the selected columns, and select Pivot from the drop down arrow.



Once pivoted, your data will be in the detail or "long" format making it easier to work with.

If you have taken these steps, you will want to ensure your field names have been updated to reflect the new format appropriately, before moving to your initial worksheet.

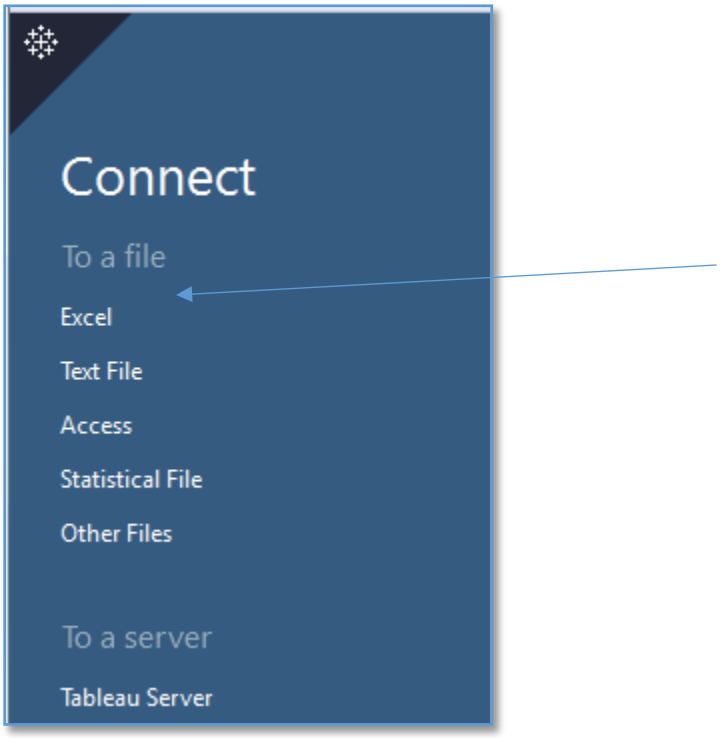




Connect to an Excel Data Source

Tableau works well with spreadsheets. In order to get familiar with the tool, you can download data from the Data Warehouse and save it locally in excel format. Once you have done this, you are ready to get started with the tool. In order to help you become familiar with the mechanics of the process you'll need, our first example will create a connection to a local excel file.

1. Start by opening Tableau. From the start page, select the database type of **Excel**.

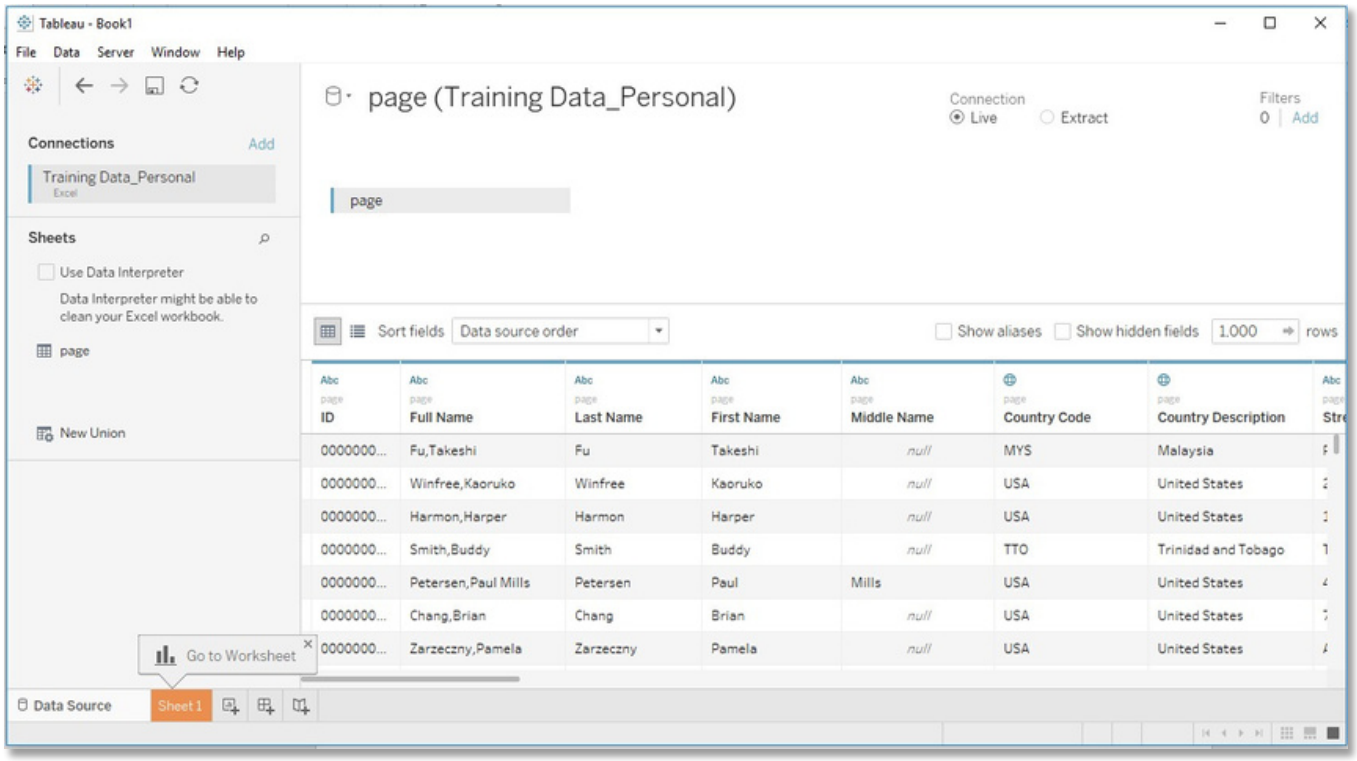


2. Choose the file you'd like to use and click **Open**. For this example, use **Training Data Personal** in the **Introduction to Tableau** folder on your desktop.
3. This will bring you to the Data Connection page. This page will appear any time you select a file, server, or saved data source on the start page.
   The left side of the page shows information about the data connection, including the data source type and tables within the data source.
   The white box in the upper area of the page shows the tables that are being used and includes options for adjusting the join type and fields used, if any.

Below this, in the area toward the bottom of the page is where you can preview the data.

In the top area, under Connections, you can select whether to use a live connection or an extract.

In the top right area, under **Filters**, you have the option to add filters to the data source to restrict the data included.

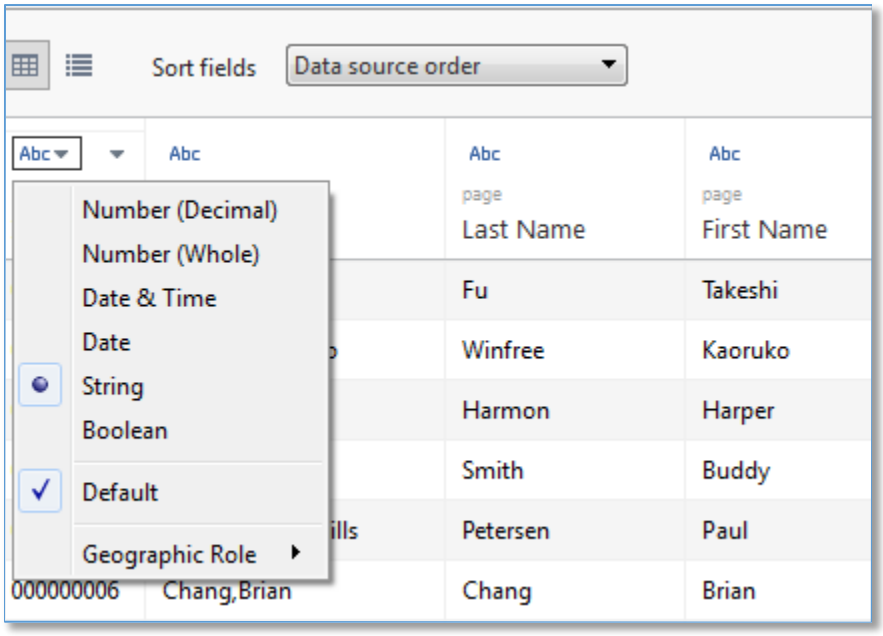


4. If you are opening a spreadsheet that has only one tab, you will see one sheet listed and the data connection will automatically be made. If there are multiple tabs, you will see a table listed for each tab. You'll need to double-click the sheet name for the connection you'd like to make.

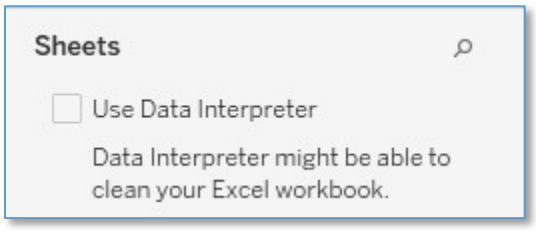
   Note that the data connection is automatically set as **Live**. If you want to create an extract, you should change the connection type.

5. Tableau has several options to ensure your data is properly imported. By default, it will import your data with the source format.

If you notice that data doesn't appear as expected on import, you can make adjustments to the definitions directly in the Metadata Grid. For example, an employee id might be created as a number or a date may have been imported as a string. You can change these to the appropriate definitions before going any further.



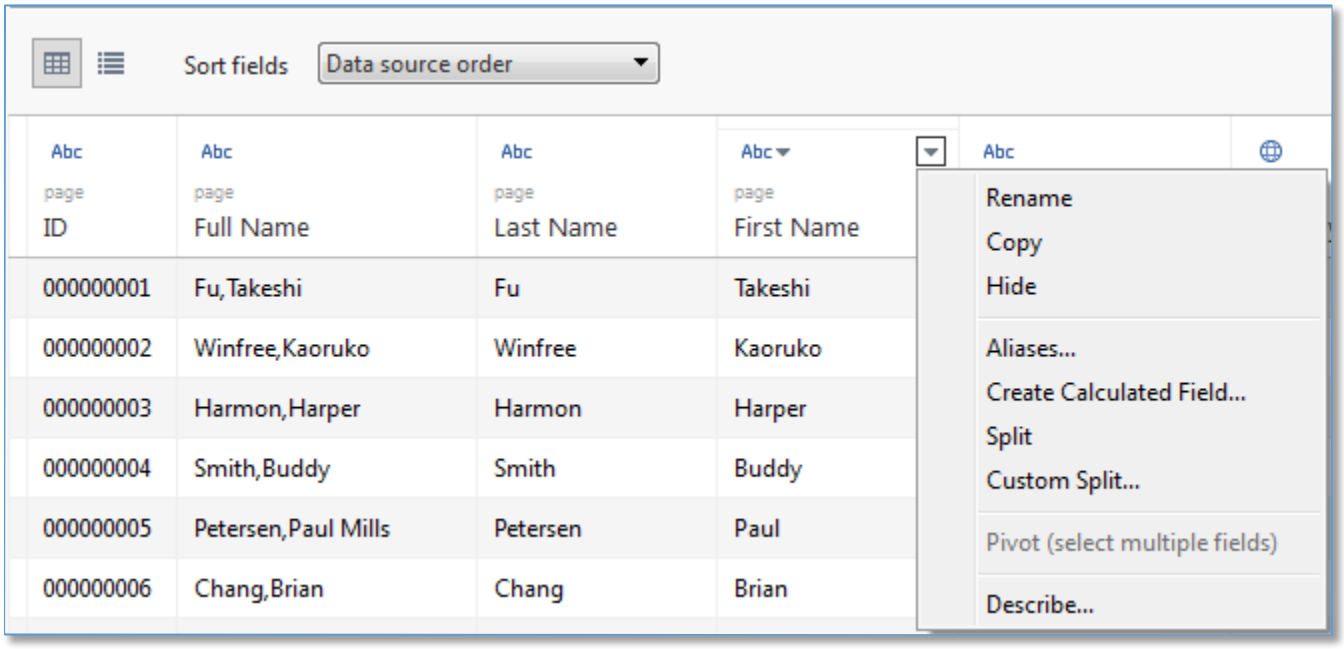
You can also turn on the Tableau Data Interpreter and it will attempt to interpret the data from your spreadsheet. When on, it will show warnings and a preview of its interpretations.



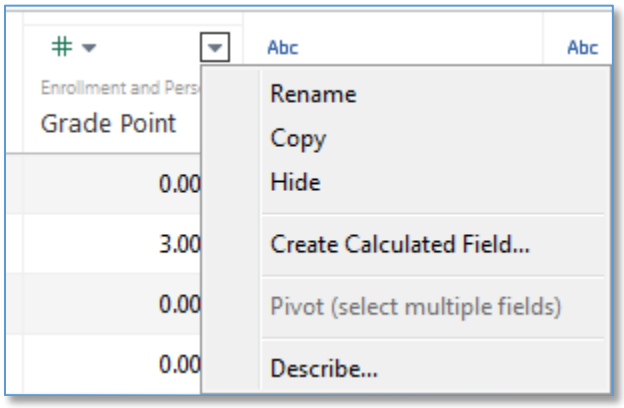
If you are connected to a Server data source, the data should already be properly defined. As a result, only the Geographical Role will be available for the field definition and the Tableau Data Interpreter will be turned off.

6. You can edit metadata information directly on the **Data Connection** page using the **Metadata Grid**. Alternatively, you can modify data attributes in the Data window or your worksheet.
   Using the **Metadata Grid**, you can rename, copy, hide, manage aliases, and create calculated fields from the view. This view makes it easy to see all of the fields in a simple format and to make quick changes to the data source before going to the worksheet.

Additional options from this view are Split, Custom Split, and Pivot. The menu options will vary depending upon the selected field type and your data source type.



The options for measures are limited to relevant actions including rename, copy, hide, and create calculated field.



Please note that if you are using a server data source, you will not be able to manage aliases and that option is removed.

7. Once you have set your connection as desired, you are ready to start working with your data. At the bottom of the page, click the sheet under **Go to Worksheet**.