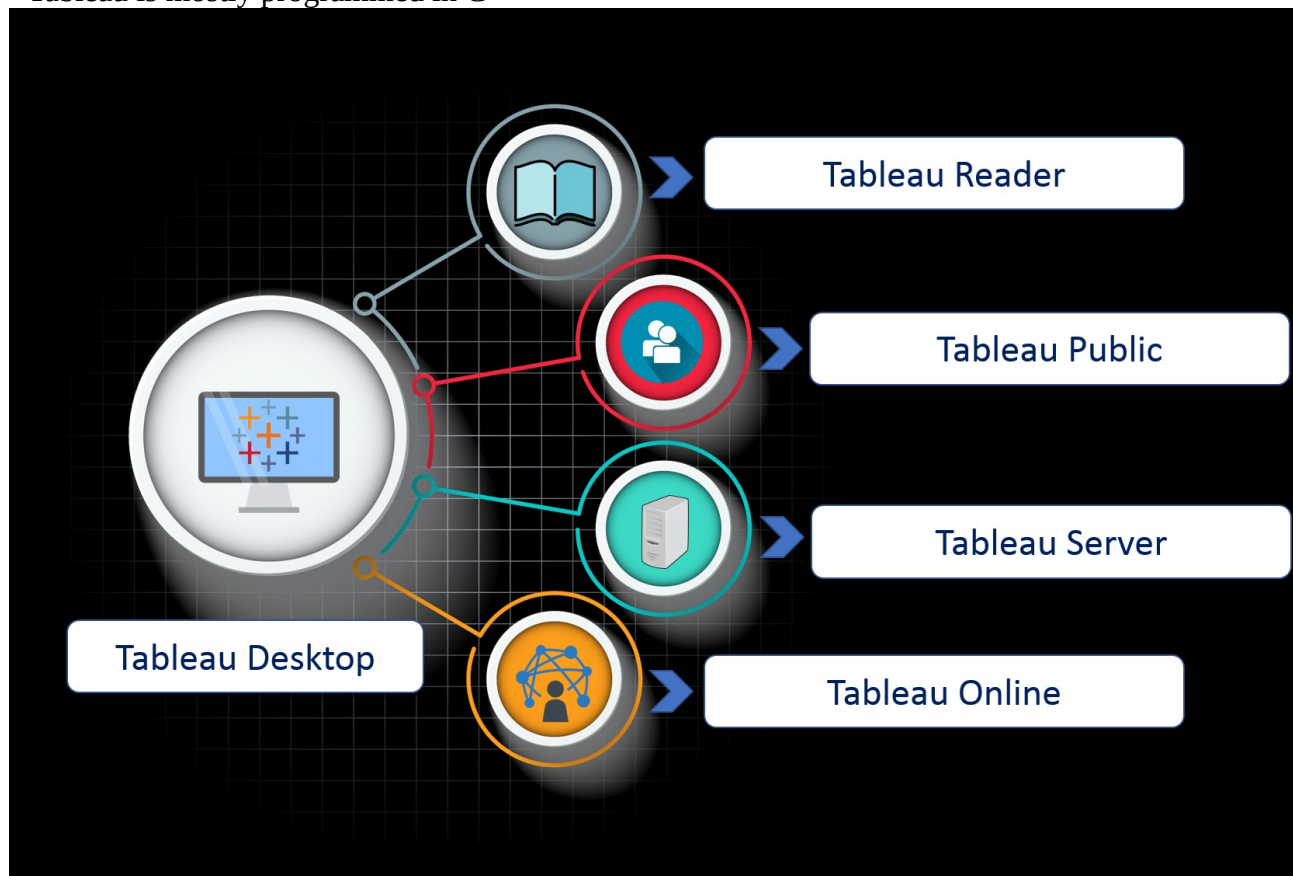# Tableau

Data Visualization is an art of presenting the data in a manner that even a non-analyst can understand it. A perfect blend of aesthetic elements like colors, dimensions, labels can create visual masterpieces, hence revealing surprising business insights which in turn helps businesses to make informed decisions.

What is Tableau?

- Tableau helps in getting insights of data with visualization, helping achieve it from connecting data from number of sources be it flat files to data warehouse.
- *Tableau* can help anyone see and understand their data. Connect to almost any database, drag and drop to create visualizations, and share with a click.

Tableau product family
- Tableau is mostly programmed in C++

	Key Features	Other Features	Operating System	License
Tableau Desktop	Creating Dashboards and Stories locally	Tableau Personal - limited data sources, non connectivity to Tableau Server Tableau Professional - Full enterprise capabilities	Windows, Mac	Personal - $999 Professional - $1999
Tableau Public	A Massive, public, non commercial Tableau Server	All data published in public	-	Free
Tableau Online	Creating Dashboards and Stories on the Cloud	Live Connections	-	$500 per year per
Tableau Reader	View Dashboards and Sheets locally	Cannot modify workbooks or connect to the server	Windows, Mac	Free
Tableau Server	Connect to Data sources and share Dashboards	Users can directly interact with Dashboards via browser	Windows	Core Licensing

1. Installation

Out of the five above mentioned products, Tableau Desktop, Public and Online offer Data Visual Creation.

Tableau Desktop

It is available in the following three formats :

1. Free trial for 14 days

2. If you are a student or a teacher, you get free access to the Desktop for a full year.

3. Purchase Tableau

Tableau Public

Tableau Public is purely free of all costs and does not require any licence. But it comes with a limitation that all of your data and workbooks are made public to all Tableau users.

Tableau Online

Tableau Online is the best option for you, if you wish to make your Workbooks on the Cloud and be able to access them from anywhere.

2. Connecting to Data

Tableau can connect with various data sources such as text, excel file, databases to big data queries also. In this section, we will look at the basics and advance feature of data connectivity with different sources. Here we will also look at Join types, Data Blending, connection with cubes, custom sql and Google Analytics.
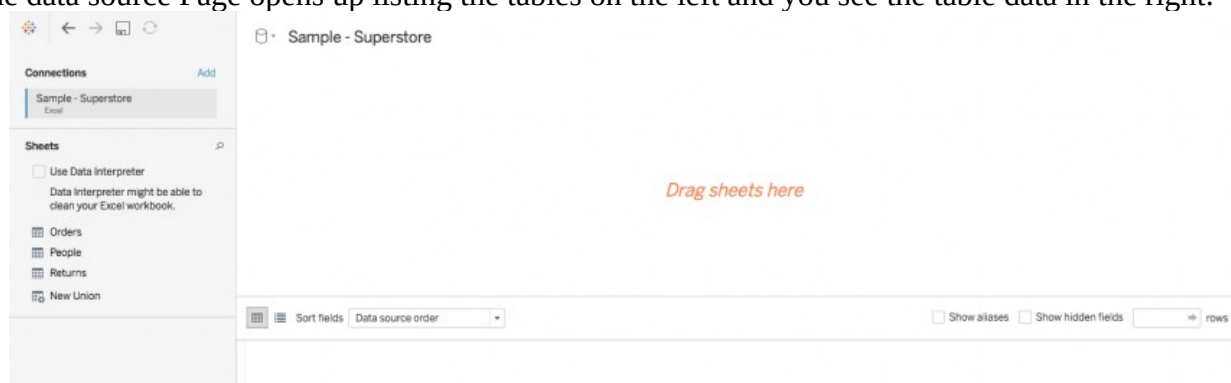
When you open Tableau, you will see a screen that looks like this, where you have the option to choose your data connection:
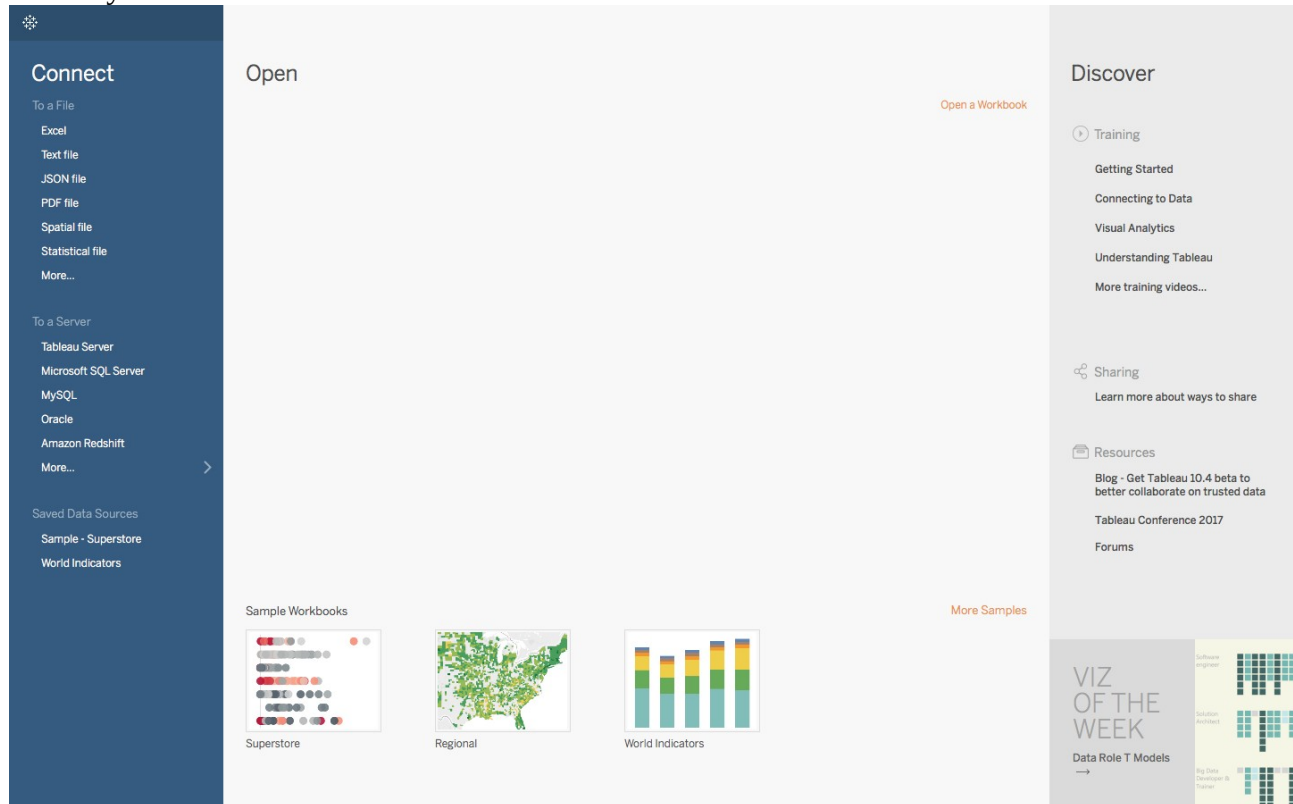
This is where you import your data. As is visible, there are multiple formats that your data can be in. It can be in a flat file such as Excel, CSV or you can directly load it from data servers too.

Let's start with using a sample data set of a superstore, which can be found here
The data is that of a United States' Superstore which is deliberating over its expansion. It wishes to know the prospective regions of the country. Let's see how can we visualize the same.

2.1. Import the data, as it is an excel file, click on Connect to a file and choose Excel, once imported the data source Page opens up listing the tables on the left and you see the table data in the right.

2.2. Now drag among the sheet to work on, so let's pull Orders here.

We see there are discrepancies and unwanted rows in the table, to clean the data, we can use the option of **Data Interpreter**, which allows tableau to clean the data automatically.

Here we go! Data looks all aligned now.

In Tableau, the Excel workbook is treated as a database and the individual tabs are treated as individual tables within that database. For this reason, you can join tabs to each other if they have at least one field in common. When you join tables, you are appending additional fields to your data source based on shared fields. To do so, simply drag the table (i.e., tab) that you want to join into the data editing interface and tell Tableau what the two tabs have in common.

So, here we have left join Orders with Returns sheet, to fetch the matching returns with all the orders, the one which are not available are marked as null.

Add: Can be used to add more source data connections and join the

New Union: Unions can be used when we have multiple sheets and we know that all column headers are same, instead of joining we can use Union. A new column is added that tells what sheet the data came from.

After retrieving the data, we have option for each column which can be used for Data preparation.

Rename: Can change the column name.

Copy Values: To copy the initial values in the consecutive rows.

Hide: Hide the entire column

Aliases: New names to each distinct row value.

Create Calculated field: Allows to create a new field with the respective column.

Split and Custom split: eg: Name into Firstname and lastname

** Also you can change the Data type of a column by clicking on data type icon, but with caution, because if the data in columns does not match the type, then results in null values.
** and the Changes made to the data at this stage only change the metadata and has no impact on the underlying data source.

Types of connecting the data
We have two types of connections available Live or Extract.

Live connections are advisable when we are working with large datasets and/or datasets held on powerful, in-memory databases.

Extracts create a snapshot of data at the point they are created. Only a part of the whole data is used and so provides a faster approach compared to live connections. Extracts allow us to work when offline and doesn't require authentication for working on the data each time. Whereas, the data should be refreshed periodically, so the update data is available. This can be refresh tasks can be automated with Tableau server/Online.

Filters
This Filters option next to the connection type, is a data source filter. It helps in filtering out the data you don't need for you analysis. Also, helps in ensuring the security and visibility of data to the required groups only.

3. Tableau Visualisation Interface

To start with visualization we need to get familiar to the Tableau Terminology, which will help us in knowing the interface better, before we start creating Viz.
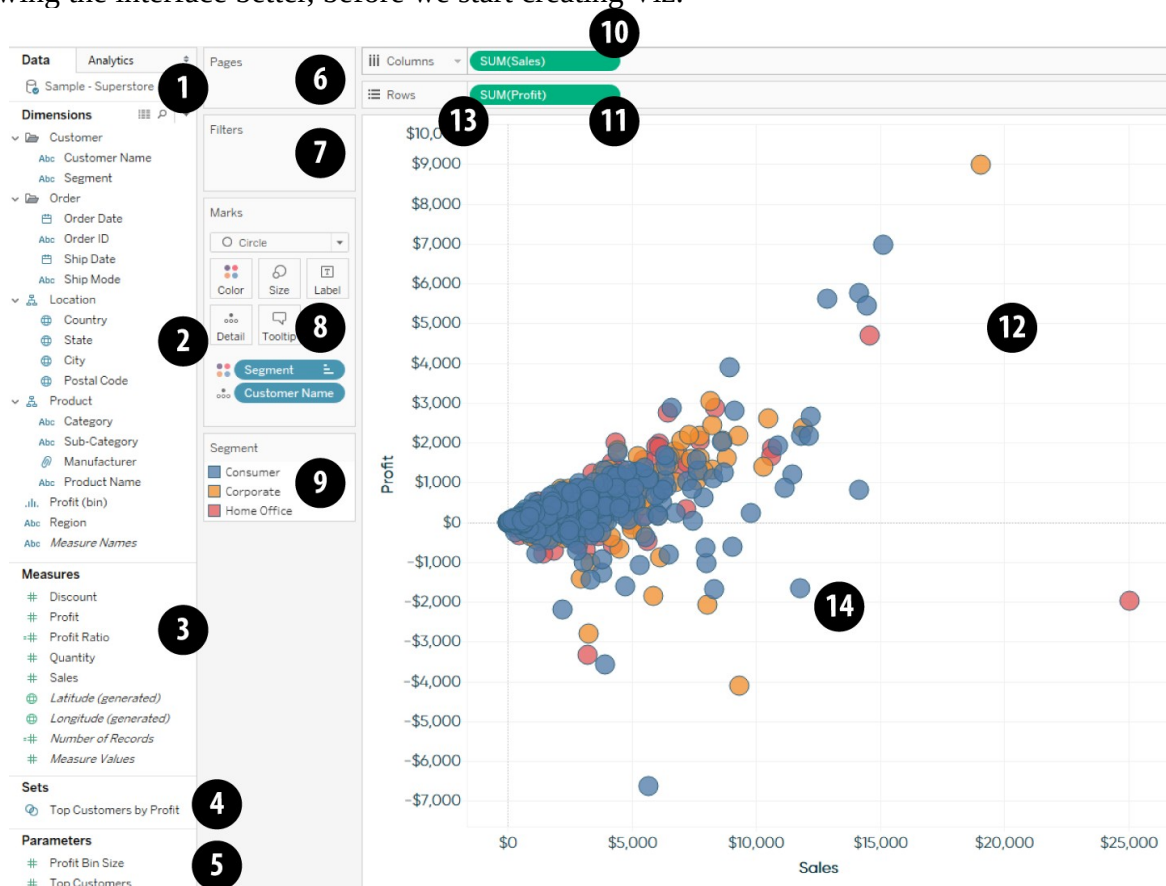
1. **Data Sources**: Displays all of the data connections in the workbook. Note that only one data connection (Sample – Superstore) is being displayed in this example, but when you connect to more than one data source at a time, they appear and by clicking on each of then, we can bring the following fields related to that data source..

2. **Dimensions area of the Data pane**: A list of all of the fields in the data source classified as dimensions.

Dimension
Dimensions are the fields that you cannot aggregate.
They are generally string values, or Discrete values, and are mostly used as column or row headings.
E.g. CountryName, ProductName etc

Measure
Measures, as the name suggests are the fields that you can measure or aggregate.
They are generally numerical values, or Continuous and are mostly used as plotting values.
E.g. CountryPopulation, ProductSales etc

3. **Measures area of the Data pane**: A list of all the fields in the data source classified as measures.

4. **Sets area of the Data pane**: If the data source you are using contains at least one set, or if you have created one or more sets, they will show up here.

5. **Parameters area of the Data pane**: If the workbook you are using contains at least one parameter, or if you have created one or more parameters, they will show up here.
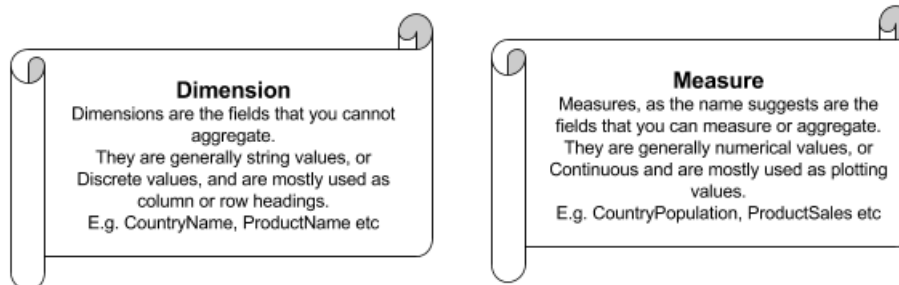
6. **Pages Shelf**: The Pages Shelf allows you to "flip" through a "page" for each dimension member and/or add animation to a view. For example, you can put a dimension for Month of Order Date onto the Pages Shelf and have the view rotate through one month of data at a time.

7. **Filters Shelf**: Any dimension or measure that you filter a view by will be displayed here.

8. **Marks Card (Marks Shelf)**: Each square in this area is called a Marks Card, which are called that because they influence the marks on the view. Each Marks Card resides on the Marks Shelf. Note that <u>depending on the chart type you are creating, additional cards will show up, such as for Shape or Path</u>.

9. **Legend**: There are several different legends that will appear here to show how the marks are encoded, including Color (pictured), Size, and Shape.

10. **Columns Shelf**: Fields placed here will create columns on the view.

11. **Rows Shelf**: Fields placed here will create rows on the view.

12. **Worksheet/View**: Each tab in a Tableau workbook is called a worksheet and the area that displays a data visualization is a view.

13. **"Pill"**: The slang term for fields being used on a worksheet. This term is used due to the oblong shape dimensions and measures inherit once they are placed on a shelf or Marks Card.

14. **Mark**: Each data point on the view.

15. On the Top left corner, we have a **Show Me** tab, which suggests type of charts based on dimension and measures chosen.

How to view the underlying data at worksheet level?

- Right click the data connection in the data window and choose "view data".

- Click the data source tab and view the data.

- To know what data is contained in the dimension/measure, right click and choose "Describe" which will provide the information of the field and on clicking "Load" you get to see what data the field contains.

4. Discrete vs Continuous

Another way to classify field in Tableau is as discrete or continuous.
Blue color of the field indicates that it is **discrete**.
Green color of the field indicates that it is **continuous**.

Two rules of thumb, while considering if the field is discrete or continuous:
1. *Discrete fields draw headers; Continuous fields draw axes*
2. *Discrete fields can be sorted: Continuous fields cannot.*

Following visualizations for sales by month:

Sales by Month: Discrete
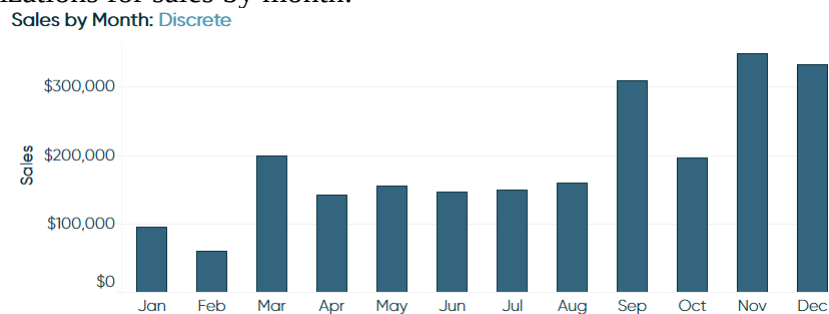
Here, the Month is discrete and could sort the bars highest sales to the lowest.

But in the below image, we see the time axis is continuous and so the order of the dates cannot be changed as they follow the chronological order from oldest on the left to the most decent on the right.

This depicts how we can use a field in discrete or continuous way to achieve the trend.

Sales by Month: Continuous

5. Aggregation in Tableau

By Default, every measure in Tableau is aggregated in some way.

SUM – All of the sales add up together

AVG – All of the sales add up, divided by the number of records

MEDIAN - When sorted, the sales amount for the record in the exact middle of your data

CNT – A count of all records with sales

CNTD – A count of distinct sales amounts

MIN – The smallest sales values of data

MAX – The Largest sales values of data.