

Quick Study Guide

What is Tableau?

Tableau is a powerful, and fastest growing data visualization tool used in the business intelligence industry, it is helpful to drive visual meaning from your million rows of data.

Whenever we connect to a data source, tableau perform two things:

- Assign a data type
- Assign a role i.e. Categorized fields into Dimensions and Measures

Common Data Types in Tableau

- Number
- String
- Geographic
- Date
- Date Time
- Boolean

Important Concepts

Dimensions – The Qualitative fields that describe categories of data. These are the independent variables. In case we have multiple dimensions in a row or column, the first dimension creates the pane.

Measures – The Quantitative or numerical fields that measure categories of data. These are the dependent variables.

Measure Name – It is a dimension that contains a label for each measure in the data source.

Measure Value – It is a measure that contains numerical values of each measure in the data source.

Latitude and Longitude – These are the Tableau generated geo fields, and it is generated when we have geographic fields in our data source.

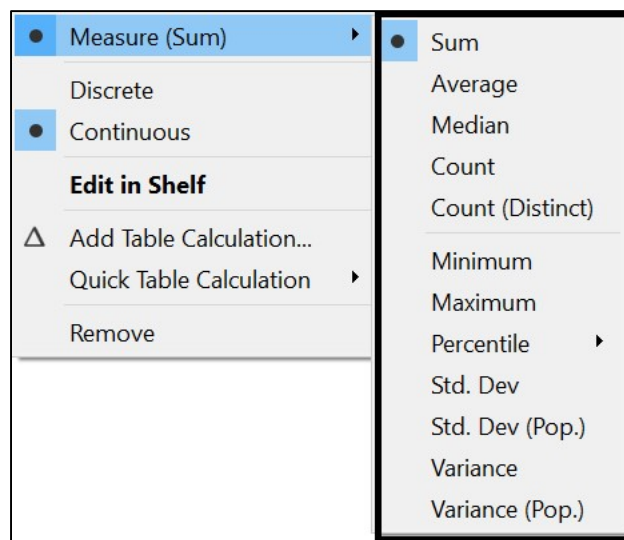
Discrete and Continuous Fields

Discrete fields	Continuous Fields
Create headers/labels	Create Axis
Can be sorted	Can't be sorted
Always represented in Blue Color	Always represented in Green Color

Aggregation and granularity

By default, measures are aggregated, and the default aggregation is SUM.

The following aggregations are available for measures:



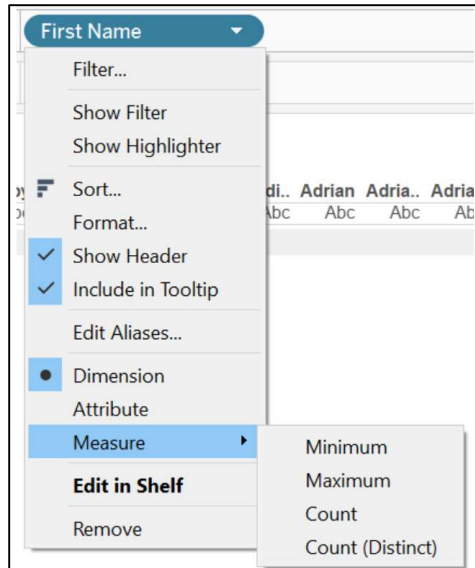
Dimensions break down the aggregated total into smaller totals by category or we can say that dimension provides the granularity in a chart.

Aggregating the Dimensions

You can aggregate a dimension in the view as **Minimum**, **Maximum**, **Count** or **Count (Distinct)**. When you aggregate a dimension, you create a new temporary measure column, so the dimension actually takes on the characteristics of a measure.

The only exception in this is if you have numeric dimension let's suppose EXAM ID, in this case you will find all the aggregate functions similar as measures.

For Reference, see the below image.

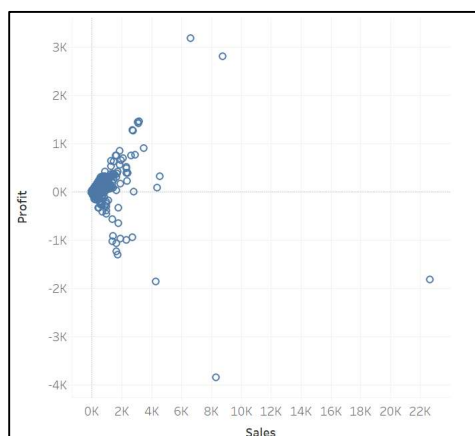


Ratio Calculations

$\text{SUM}(\text{Profit})/\text{SUM}(\text{Sales})$ sums the profits and sales to whatever the granularity of the view is, then computes the ratio at that aggregation. $\text{Profit}/\text{Sales}$ computes the profit ratio at the lowest level of granularity then sums the ratios to the requested aggregation of the view.

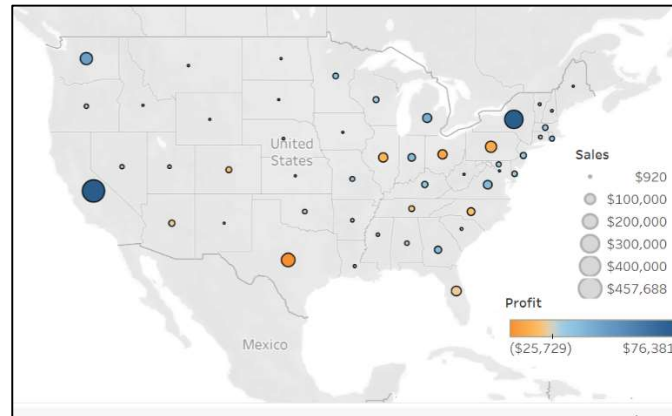
Basic Chart Types in Tableau

Scatterplot – Demonstrate the relationship between measures. Ex – Relationship between profit and sales.



Maps – Tableau has two maps types: symbol maps and filled maps. The symbol maps use the symbols to represent a central point of a geographic region whereas in filled map boundaries of a geographic region is filled with a color.

Below is an example of symbol map. The size of the symbol represents the sales, and color of symbol represent the profit.

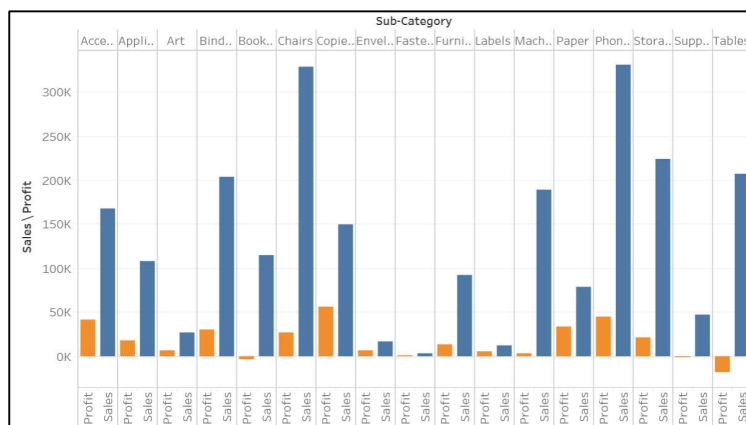


Add Background Image Map

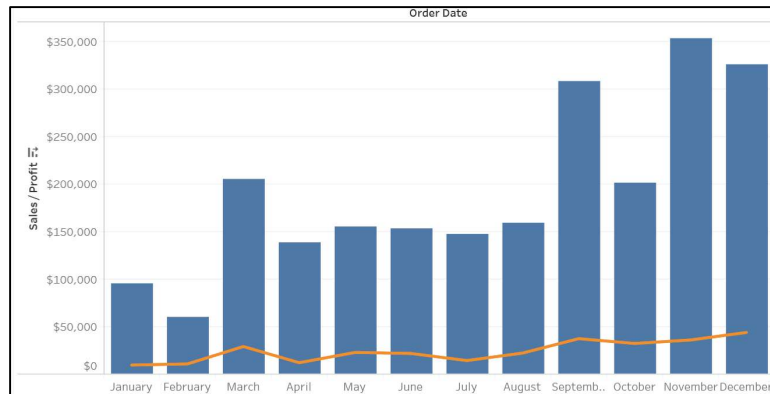
In your dataset you must have X and Y coordinates to plot the custom image and to show your data in the foreground.

To add an image, go to menu Map -> Background Images -> Add Image. Provide the path of the image and define the X and Y Fields.

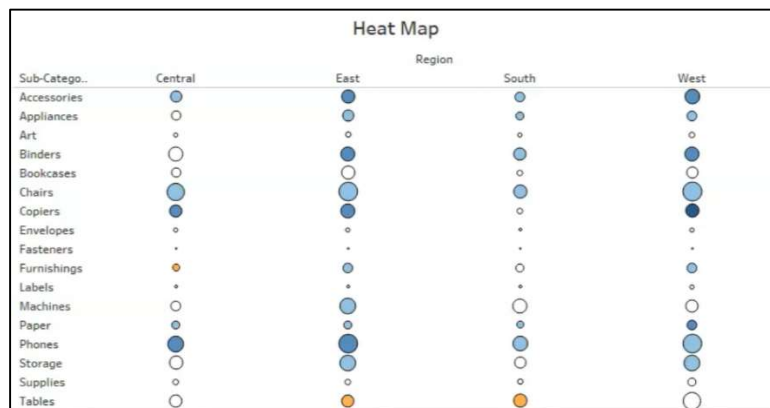
Combined Axis Chart – It is created with one shared axis, such as an X-axis for Sub-Category, and two separate axes, such as Y-axes for two different measures Profit and Sales. Placing measures on the same axis allows you to build a view that can easily compare those measures using one chart.



Dual Axis Chart – It allows you to compare multiple measures in a graph. Dual Axes are two independent axes. In this example, bars are showing the Sales and line is showing the profit. As both axes were showing the \$\$ value so we synchronize both the axes and change the marks type to a bar and a line respectively.



Heat Map – In a heat map, marks on the charts are represented as colors and these colors are displayed in a crosstab. In the below heat map, we are showing sales and profit for each sub-category.



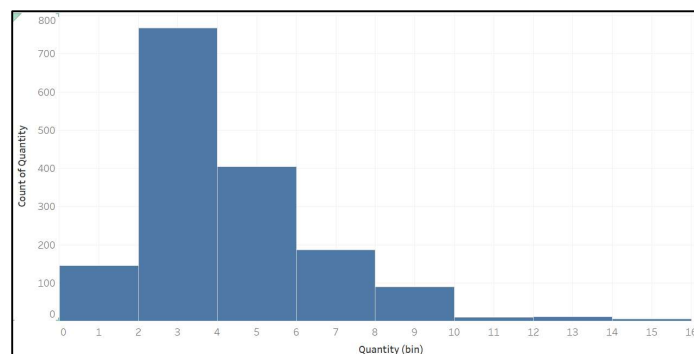
Tree Map – This chart type is used to display data in the nested rectangles. We use dimension to define the structure of the tree map and measure to define the size of the rectangles.



Crosstab, Highlight Table – Crosstabs are also known as Text tables. A highlight table encodes a measure on Text and color on Marks card. Ex- It shows the Sub-Categories profit in each region.

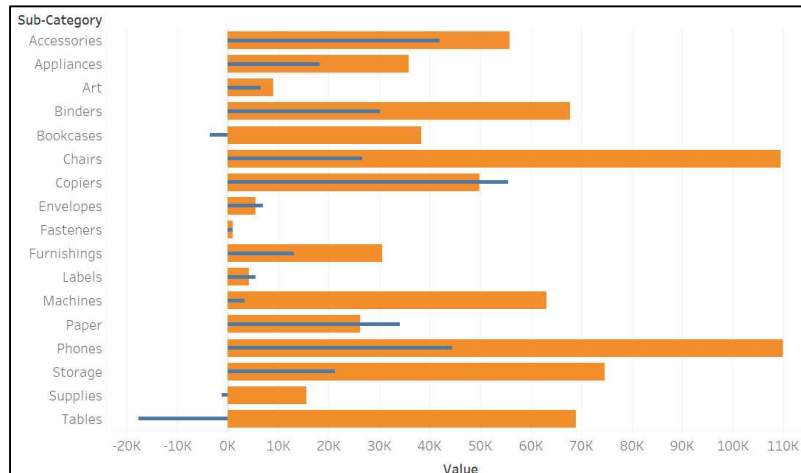
Sub-Catego..	Region			
	Central	East	South	West
Accessories	\$7,252	\$11,196	\$7,005	\$16,485
Appliances	(\$2,639)	\$8,391	\$4,124	\$8,261
Art	\$1,195	\$1,900	\$1,059	\$2,374
Binders	(\$1,044)	\$11,268	\$3,901	\$16,097
Bookcases	(\$1,998)	(\$1,168)	\$1,339	(\$1,647)
Chairs	\$6,593	\$9,358	\$6,612	\$4,028
Copiers	\$15,609	\$17,023	\$3,659	\$19,327
Envelopes	\$1,778	\$1,812	\$1,465	\$1,909
Fasteners	\$237	\$264	\$174	\$275
Furnishings	(\$3,906)	\$5,881	\$3,443	\$7,641

Histogram – Shows the frequency of distribution in your data. Or we can say that it displays the shape of the distribution. It looks like a bar chart but groups values for a continuous measure into bins (a range of values). Binning a measure creates a dimension, so we will find newly created bin in a dimension pane. It can always be created using a single measure. Ex – It shows the frequency distribution of a Quantity field.

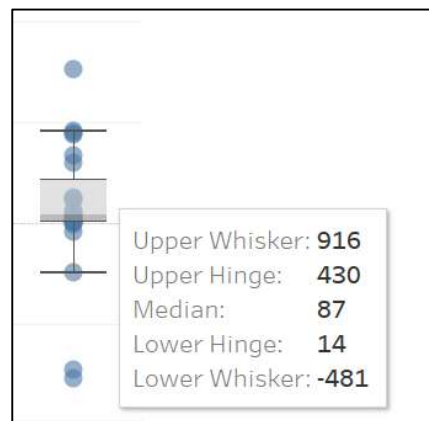


Bar in Bar Chart – It is used to compare the measures against the goals. It is similar like Bullet graph. To create this chart, first create the combined axis chart and then follow the below steps:

- Drop the measure names field from Rows/Columns to Color marks card.
- Go to Analysis menu -> Stack Marks -> Off
- Now, drop the measure names to the size so that both bars will be visible.



Box Plots – Are also known as Box and Whisker Plots, it helps us to compare distributions and identify outliers.



Median

If you have odd numbers, then $n/2$ and round up to get the position of the median number.

If you have even numbers, then $n/2$ and round up. Take the number in that position and average it with the number in the next higher position to get the median.

IQR (Interquartile range) is the height of the box. Formula is **Upper hinge – Lower hinge**.

The height of the whisker is **1.5 times of IQR**.

Upper whisker = Upper hinge + (1.5 * IQR)

The upper whisker of the box plot is the largest dataset number smaller than 1.5 times of IQR above the third quartile.

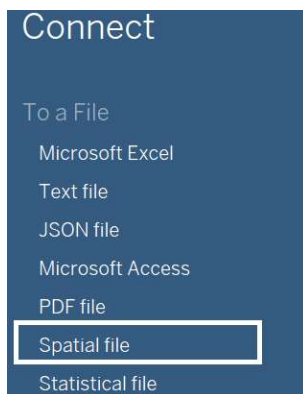
Lower whisker = Lower hinge – (1.5 * IQR)

The lower whisker of the box plot is the smallest dataset number larger than 1.5 times of IQR below the first quartile.

Connect to Spatial Files

Spatial files allow you to analyze geographic data that lies outside, and Tableau automatically recognize this data. The spatial files encode geographic features as geometric shapes, which we use to visualize and analyze geographic data.

Open Tableau, and then Select “Spatial file” to connect to a spatial file



Groups

Groups are denoted by paper clip. It lets you combine several members of a single dimension into categories that create a new dimension field that did not exist in the original data set. It can be created from dimensions in a data pane or directly from the visualization.

@ Sub-Category (group)

Sets

Sets are useful for viewing and highlighting data that meets specific criteria. Sets are always binary it means that you are either in the set or not. Once a set is created, it can be used as a filter or combine set to show either data that exists in one, but not in the other, or the combination of all data found in either set. You can **combine two sets** only if these are created using same dimension. It can be created from dimensions in a data pane or directly from the visualization.



Hierarchies

It allows us to organize the dimensions in our data. Usually, we create the hierarchy for related columns for example – Country, State, City, Zip.



- Country
- State
- City
- Zip

When we use hierarchy in a view, we can drill up and down along the fields of the hierarchy.

Hierarchies can be created in two ways:

- Select multiple dimensions and right click and select Create Hierarchy
- Drag and drop one field to another

Sorting

There are two types of Sorting:

Computed Sort – It is automatic update the sorted view if data changes.

Manual Sort – It maintains the order even if data changes.

Work with aggregation options

For any aggregate calculations, we cannot combine an aggregated and disaggregated values.

Some of the important points to note:

- $\text{SUM}(\text{Profit}) * [\text{Quantity}] \rightarrow$ is not valid
- $\text{SUM}(\text{Profit}) * 4 \rightarrow$ Is valid
- $\text{SUM}(\text{Profit}) * \text{SUM}(\text{Quantity}) \rightarrow$ Is valid
- $\text{MAX}(\text{SUM}(\text{Profit}), (\text{Sales})) \rightarrow$ is not valid

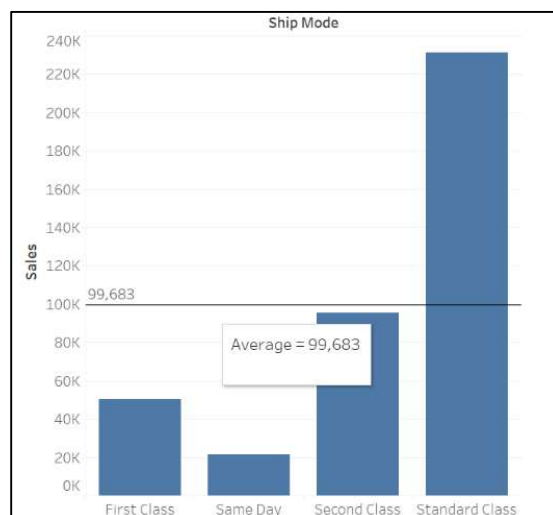
Analytics

Reference lines marks a specific value on an axis. We can only create a reference lines from the measures currently in a view.

There are two ways to create reference lines:

- Right click on axis and select Reference line
- Using Analytics Tab

In the below example, Reference lines show the Average sales across all ship modes.



Reference Bands – It shows the data that falls within a certain window value and display as shaded area between two values.

To add the reference band, go to Analytics pane and drop the Reference band to the chart and select the Scope, Band From and Band To values.

Edit Reference Line, Band, or Box

Line Band Distribution Box Plot

Scope

☐ Entire Table ☒ Per Pane ☐ Per Cell

Band From

Value: SUM(Profit) Minimum

Label: Computation

Tooltip: Automatic

Band To

Value: SUM(Profit) Maximum

Label: Computation

Tooltip: Automatic

Formatting

Line: [red line]

Fill: [orange box]

☒ Show recalculated band for highlighted or selected data points

OK

Reference Distribution – are a variation of reference bands, it shows the shading as per your selected scope to indicate the distribution of values along the axis. There are 4 different ranges of data distribution.

To add the reference distribution, go to Analytics pane and drop the Reference distribution to the chart and select the Scope, and Computation Value.

Edit Reference Line, Band, or Box

Line Band Distribution Box Plot

Scope

☐ Entire Table ☒ Per Pane ☐ Per Cell

Computation

Value: 60%,80% of Minimum

Label: ☒ Percentages Percentages: 60,80

Tooltip: ☐ Percentiles

Percent of: SUM(Sales)

Formatting

☐ Quantiles

☐ Standard Deviation

Minimum

☐ Symmetric

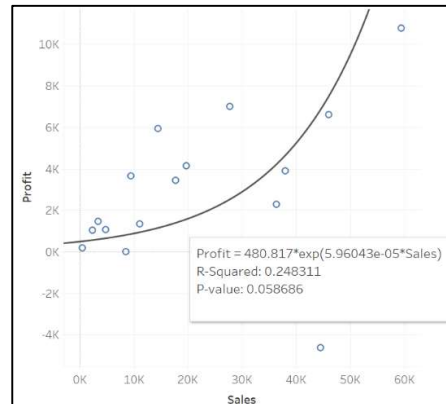
☒ Show recalculated band for highlighted or selected data points

OK

Trend Lines can be used to explore the relationships between two measures in your data. It can be added from Analytics tab.

Trend Lines could be:

- Linear
- Logarithmic
- Exponential
- Polynomial
- Power



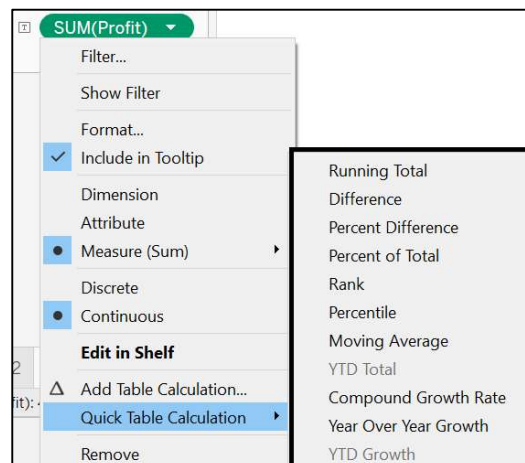
R-Squared is a statistical measure of how well the trend fits the data. Value of 1 or 100% is a perfect fit.

P-value is a probability value associated with significance. Smaller is better, ideally less than .05

Table Calculations

These are computations that are applied to a value in the data table in a view. These are performed after the results have been returned from database and are processed locally in Tableau. List of useful default calculations called “Quick Table Calculations”.

How to perform – Right click on measure in a marks card or rows/columns in a view, Click on Quick Table Calculations and select the calculation.



Level of detail (LOD) Expressions

Level of detail (LOD) defines the granularity of your data.

LOD is defined by the dimensions

If data is more aggregated, then it is less granular and vice-versa.

For ex – Drop profit into the view and then drop dimension and depending on the dimension you have chosen to add to the view, your data will be aggregated accordingly – to the Viz LOD. However, sometimes our requirement is to perform a calculation that is independent to the Viz LOD or we need data that is at a different level of detail than the view.

To solve these issues, LOD expressions comes in picture.

For a table scoped LOD expressions, we can wrap curly braces around any aggregated field. {AGG([Field])} like {SUM([Sales])}.

There are three keywords we can use in level of detail expressions:

- **FIXED**
- **INCLUDE**
- **EXCLUDE**

Syntax of LOD Expressions:

- An open curly bracket followed by keyword FIXED, INCLUDE or EXCLUDE.
- Followed by 0 or more dimensions you want to be fixed, included or excluded from the computation
- Followed by a colon “:”
- Followed by an aggregated measure
- In the end, closing curly bracket.

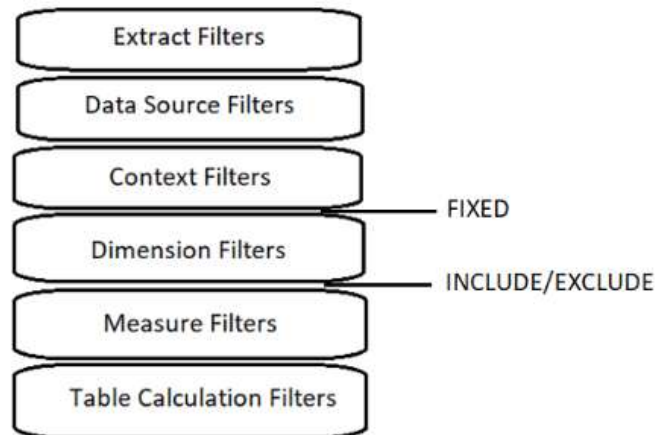
Ex – {EXCLUDE [Sub-Category] : AVG(Profit) }

Context filters

If we have multiple filters in filters shelf, then it works on AND clause. But by specifying the context filter we can make other filters dependent on the context. Context filters are queried first then other filters are queries against the result of that query. It specially helps when using Top or Bottom N filters or other combination of filter where the ordering of filter matters to the

outcome. To apply context filter, right click on the field in Filters shelf and select “Add to context”

There are several different kinds of filters in Tableau and they get executed in the following order from top to bottom:



Dashboard and its Action for Interactivity

In Tableau, dashboards are the collection of views that are built with different dimensions and measures, or even built from different data sources.

To add the actions, navigate to Dashboard menu -> Actions -> Add Actions

Actions can be run on either Select, Hover or Menu.

Create a story using dashboards or views

A Tableau story is a connected series of worksheets and dashboards, that further allows us to capture insights and share them in a sequential manner.

A Tableau story consists of one or more story points. One story point contains only one dashboard or worksheet that highlights specific details.

Visual Best Practices

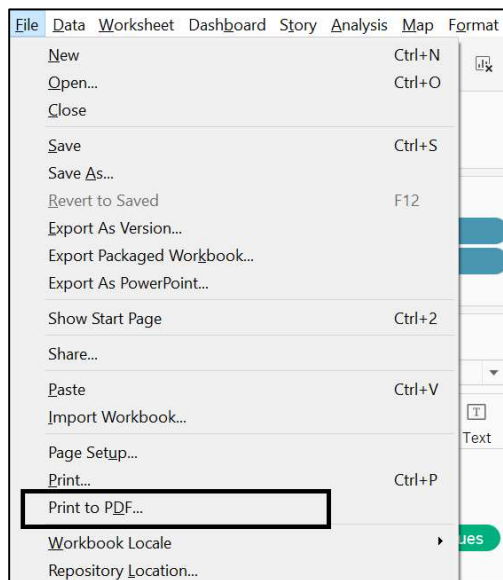
1. Arrangement of view and filters – The most important component should be placed at the upper left.
2. Instructions – Clear instructions for how to use the dashboard and each component.

3. Meaningful titles
4. Tooltip – Should describe data clearly
5. Consistent font formatting – Titles, instructions, filters, labels, headers, legend (all components in the dashboard must have consistent font)
6. Color – Use alarming color for critical data, keep the color range to a minimum.

Sharing Options

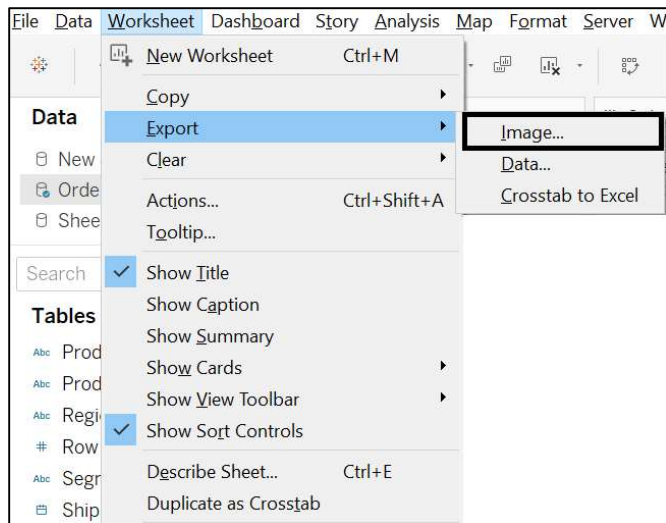
Share a twbx as a PDF

Go to file and click on Print to PDF



Share a twbx as an Image

Go Worksheet and click on Export and Select Image



Live Connection Vs Extract

By default, Tableau creates the Live connection

Live Connection	Extract
Queries the data from the database	Queries the data from the Tableau Data Engine
Data updates automatically	Manually refresh the extract.
Workbook performance is slow, if we have large amount of data.	Performance is fast but to refresh the extract is time consuming
Always connected to the data source for the real time updates	Accessed offline
If connected to a database, user need to enter the credentials	No need to enter the credentials

Joins

Joining is a method to combine data from multiple data sources or combine data from different tables in a single data source. When we refer data from multiple data sources then it is called as ***cross-database joins***.

There are 4 types of joins:

Inner – The resulting table contains values that have matches in both tables.

Left – The resulting table contains all values from the left table and corresponding matches from the right table.

Right - The resulting table contains all values from the right table and corresponding matches from the left table.

Full Outer – The resulting table contains all values from both tables.

Blending

Blend is a left join after aggregation at the level of granularity.

We need separate data connection for every participant in blend. One data source is considered as primary and the other is secondary. The primary data source is determined by the field you first add to a view. Tableau shows this by giving that data source a blue check mark in the data pane. If you then add a field from a secondary data source, then that data source will be considered as secondary and shown in orange color check mark in the data pane.

Data blending is worksheet specific. It means that we can change which data source to be considered as primary and which is secondary on different worksheet in a same workbook.

The major difference between join and blend - *Join combine data before aggregation at row level whereas blend combines data after aggregation.*

Unions

It is a way to combine multiple tables from single data source. For optimal results, the tables that we combine using a union must have the same structure i.e. same number of fields and that fields must have the same names and data types.

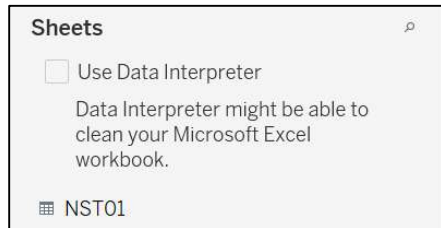
When we union multiple sheets, two new columns automatically generated, Sheet and TableName.

The difference between join and union – Join appends columns from one table to another whereas union appends rows.

Data Interpreter

Data Interpreter helps in cleaning your data. It can detect things like titles, notes, footers, empty cells, and so on and bypass them to identify the actual fields and values in your data set.

Drag a table to the canvas (if needed), then on the Data Source page, in the left pane, select the Use Data Interpreter check box to see if Data Interpreter can help clean up your data.



Happy Learning & All the best!!!