

Summary

Visualisation using Tableau

Tableau is a business intelligence software that allows anyone to connect easily to data, visualise and create interactive dashboards. Tableau is quite simple to use, characterised by a drag and drop interface. It automatically displays column names, cleans data of any discrepancy and provides for an easy learning experience.

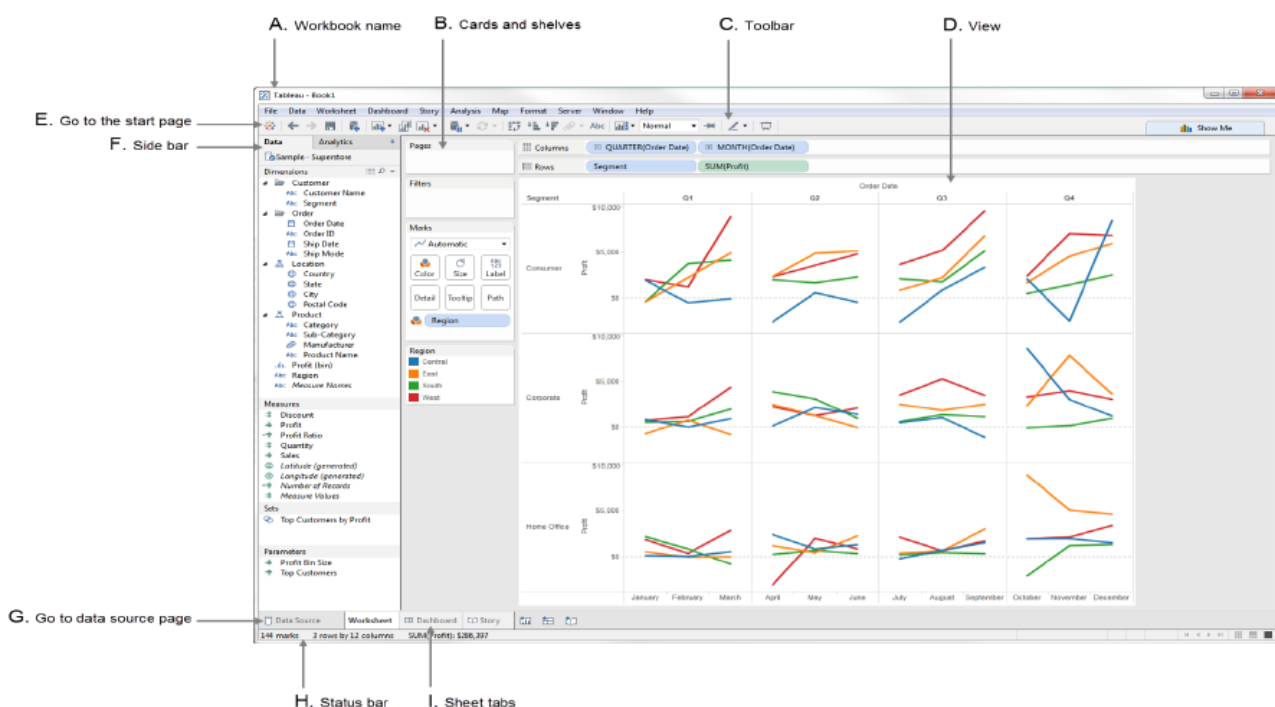
Data Exploration in Tableau

The Tableau interface has menus at the top. Below it is the toolbar, with buttons like *undo* and *save*; Tableau worksheets are not automatically saved, so worksheets need to be saved periodically.

The data window is to the left of the screen. On the data tab, the top lists all open data connections, and depending on which one is selected, the fields from that data source are listed below, broken out into dimensions and measures. The difference between dimensions and measures is that dimensions contain all those attributes or columns whose values are categorical; for example, gender, i.e. Male/Female. When the values are quantitative, they get assigned to the measures tab. For example, sales figures, population, percentages, and so on. Visualisations are created on measures based on the dimensions.

The interface also consists of shelves where various variables and parameters can be dragged and dropped.

These shelves are of rows, columns, filters, marks and pages.



It is very important to convert data into the correct format. Usually Tableau recognises the format of data itself but it can be changed manually using a simple drop down menu.

Tableau gives you the freedom to import data from different data sources such as excel files, text files, SQL, web etc. If the data belongs to multiple files then it can be collated together. If the data belongs to the same data source but different data files then those data files can be merged. To merge two files there must be a common join field between the two files. If the names of the join fields are same then Tableau will recognise them itself. Otherwise, you must specify the join field. There are 4 types of joins/merges:

1. Inner merge
2. Outer merge
3. Left merge
4. Right merge

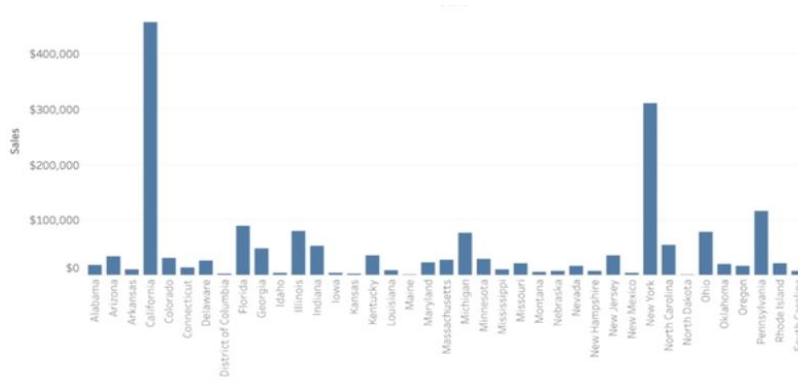
If the data belongs to different data sources itself then data blending is carried out. Once you have connected to the data, it can then be prepared. Tableau removes discrepancies from the data itself upon using the **data interpreter** option. To create new columns out of the existing ones, you can use the **split** and the **custom split** option. A normal split is carried out in a column if the delimiter on which the split is carried out is present in all the fields of the column. If not, a custom split will be carried out where you will specify the delimiter to carry out the split. Similarly, you can filter on the data too.

Hierarchies are a very powerful option in Tableau. Tableau creates known hierarchies itself such as Country – State – City etc. if the column names are intuitive but you can create hierarchies yourself by dragging a column onto other in the data shelf. Hierarchies can then be plotted directly with measures and gives you the freedom to drill down in the visualisation itself for deeper analyses.

Visualising and Analysing Data in Tableau

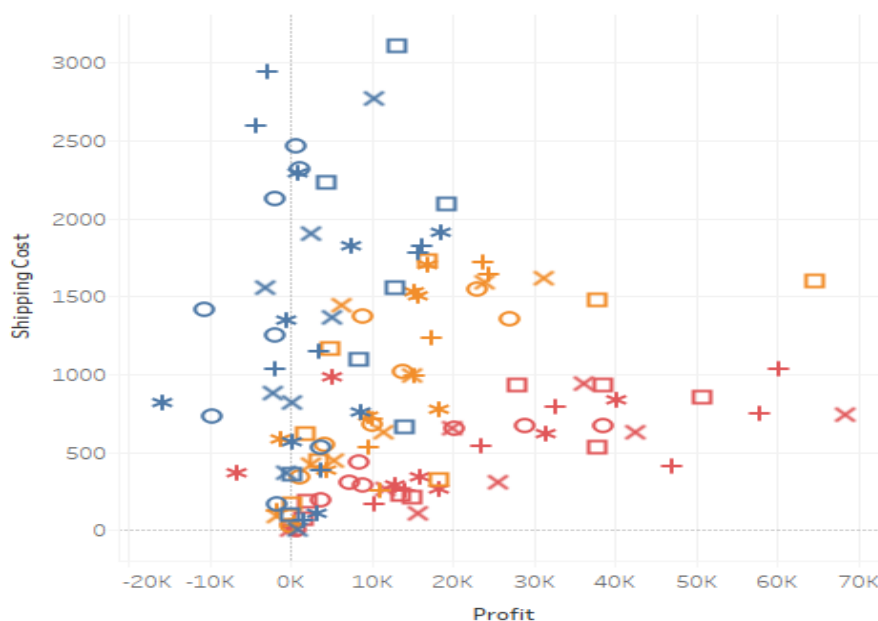
Visualisations are created based on the type of data you have. But you must know which combination of features (dimensions and measures) you can plot together to create the type of visualisations you want. Tableau automatically selects the best visualisation technique based on the features you have selected, but it can always be changed manually. You can leverage elements of Tableau, such as colour, size and shape to create interactive visualisations. There are several charts that can be created in Tableau. Some are discussed below:

1. Bar Chart: A bar chart is a very simple chart and is most useful when you want to compare data across dimensions. For example, visualising sales across regions.



It is mostly preferred when you want to create a visualisation for one categorical and one numeric feature. You can add other dimension variables to the colour option in the Marks card to make a stacked bar chart and to make the visualisation more interactive. A grouped or a stacked bar chart are plotted when you want to visualise a multi categorical dimension with a common quantitative variable.

2. Scatter Plot: A scatter plot is used when you want to visualise the relationship between two quantitative variables. It gives you a very good idea of how one variable changes or reacts to different values of the other variable. Scatter plots can be made very interactive by adding other dimension variables to the colour and shape options in the Marks card.

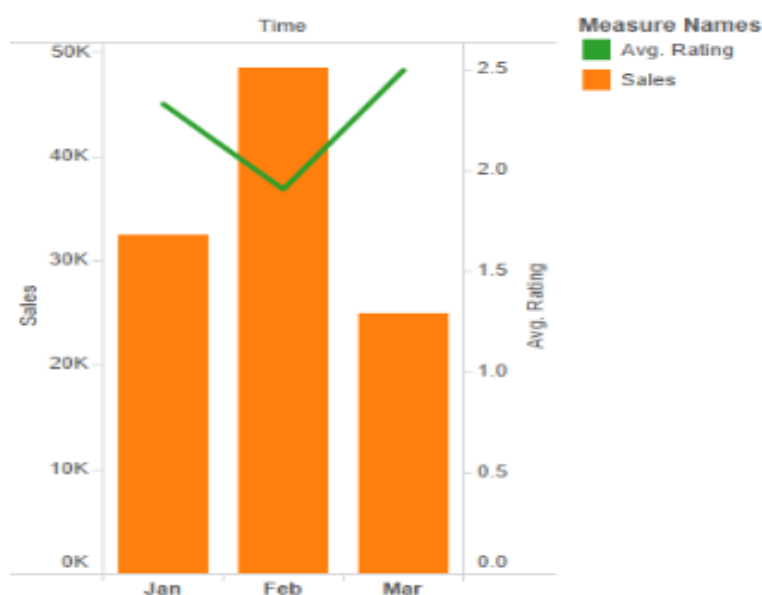


3. Tree Maps: In scatter plots, when you use a dimension to display colour, Tableau uses different colours for each category in the dimension. However, you can use a measure for colour too and Tableau will present a scale of colour ranging from light to dark depending on the measure's values. The same concept is applied in tree maps which are used to visualise two quantitative and one categorical variable. The quantitative variables exploit the size and the colour parameters of the tree map.

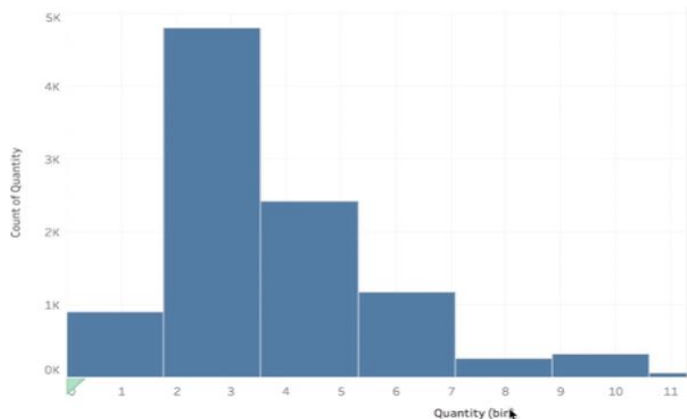


In the tree map above the colour represents the profit, the size represents the sales and the dimension variable is the month. A tree map can only accommodate 2 quantitative variables, one for size and the other for colour but can accommodate any number of dimension variables as that would only increase the number of boxes in the tree map.

4. Dual Axes: Dual axes charts is a very powerful way to visualise quantitative variables of different scale together. In these charts there are 2 y-axes representing one quantitative variable and 1 x-axis representing the dimension variable. These are created when the ranges of the quantitative variables differ a lot.



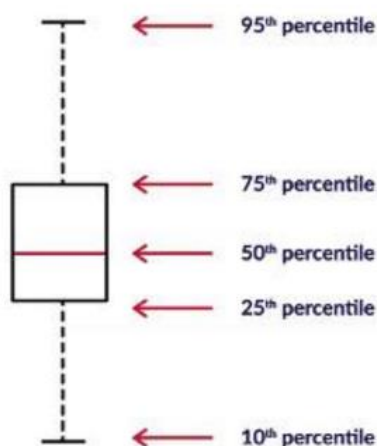
5. Histograms: Histograms plot quantitative data with ranges of data grouped into bins which are decided automatically by Tableau. Histograms show the distribution of a single variable across the bins. The bin width can also be changed manually.



The histogram above shows the Count of Quantity on the y-axis and the Quantity bins on the x-axis with a bin width of 1. Such a histogram can be used to manage inventory of a store.

- Box Plots: Box plots enable us to study the distributional characteristics of a variable and show the overall patterns of a variable. They tell you about the existence of **outliers** and the **overall spread** of the data. They are also very effective when you want to compare the distribution of two quantitative variables. The line

inside the box denotes the median value: 50% of the data lies both above and below that value. The boundaries of the box denote the 25th and the 75th percentile, with the whiskers denoting the lowest and the highest values in the data set. Any observations lying below or above the whiskers are outliers. However, box plots can only be created for quantitative variables.



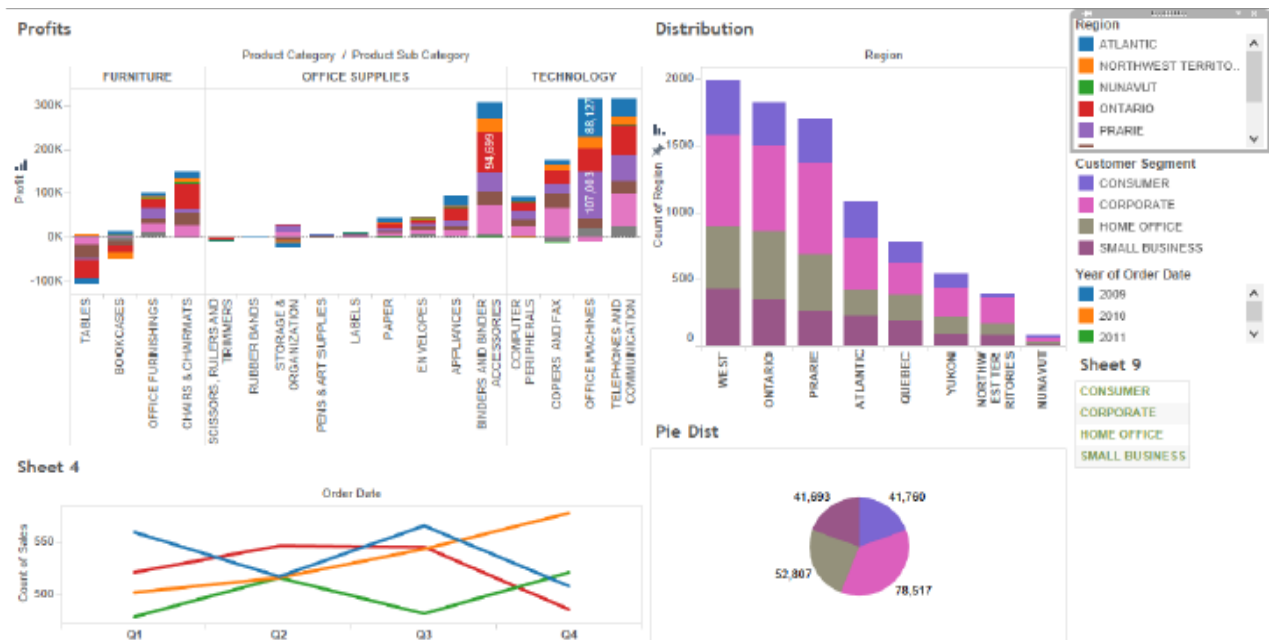
- Dashboard and Stories: Dashboards and stories are an effective way to present your finding about the data you have visualised.

To create a dashboard, click on the new dashboard icon at the bottom. The size of the dashboard can be selected from the bottom left. Sheets can be added by simply dragging and dropping.

To create a filter, simply select any chart on dashboard and select the filter icon on its upper right that says “Use as filter”. This enables some options in the menu. Go to Analysis, select filters, and the variables of that chart are available for filtering.

To create a custom filter, create a new sheet and add the variable name to rows and on the labels card on the marks shelf. On the tab in the rows shelf, right click and uncheck “Show header”. To modify the font, right click on the tab, select format and choose any color. Go to the dashboard and this sheet can be dragged. Select it and use as filter.

The show title at the bottom left toggles the visibility of titles for each chart and the dashboard, depending upon the selection.



A new story can be created by clicking on the new story icon at the bottom. New slides or points can be created by clicking on “New Blank Point”. Duplicate will create a copy of the existing sheet. Captions can be used as a part of the story. To mention other interesting facts about the data, use annotate to highlight areas and enter the text. A story can also contain dashboards. If the dashboard looks too cluttered, go to the dashboard and in the size option, select “Fit to story”. The titles can also be disabled to make more space in the story.

Sales Segmentation

Distribution Of Orders
Against Segments

Region Segment Wise
Distribution Of
Customers!

Region Wise
Profitability, Prarie has
the highest profit in

Ship Mode and Order
Priority. Most losses
via Delivery Truck!

The Dashboard

New Blank Point

Duplicate

