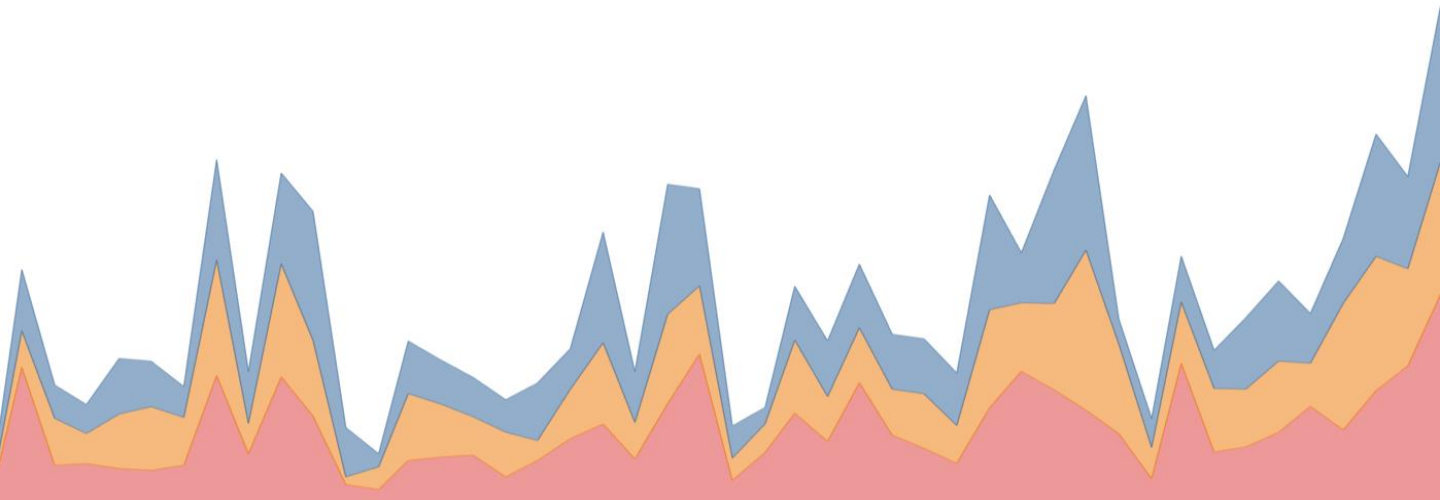
Tableau for Data Science and Business Analytics

Course Handbook
(v2021.01.06)

Farzan Sajahan



This handbook has been prepared as a supplementary reference material for the tableau for data science and business analytics course.

Two Possible Tableau Options

1. Tableau desktop developer
    - Costs about $70 / month
    - 14 day trial license
    - Student access for a year if approved
2. Install Tableau Public
    - Free to use as long as you like
    - Similar to desktop developer except for few limitations (e.g. files saved on cloud, limited data connections, no export etc.)
    - Not a limitation for learning especially for this course

How to install Tableau desktop?

- Go to https://www.tableau.com/products/trial
- Enter your email and download tableau desktop
- Install and use

How to get 1-year license for Tableau desktop?

- Go to https://www.tableau.com/academic/students#form
- Fill the form
- Wait for Tableau to verify your details – takes a couple of days
- Look for the 1 year product license key sent to your email
- Use the license key with your tableau desktop developer installation

How to install Tableau public app?

- Go to https://public.tableau.com/
- Enter your email and download tableau public app
- Install and use

Note:

In this course we do not cover tableau server. Some may use tableau server or tableau public from the cloud. You may find the interface different from the tableau versions used in the course and you may not follow the course. In order to avoid that, please use one of the versions suggested above.

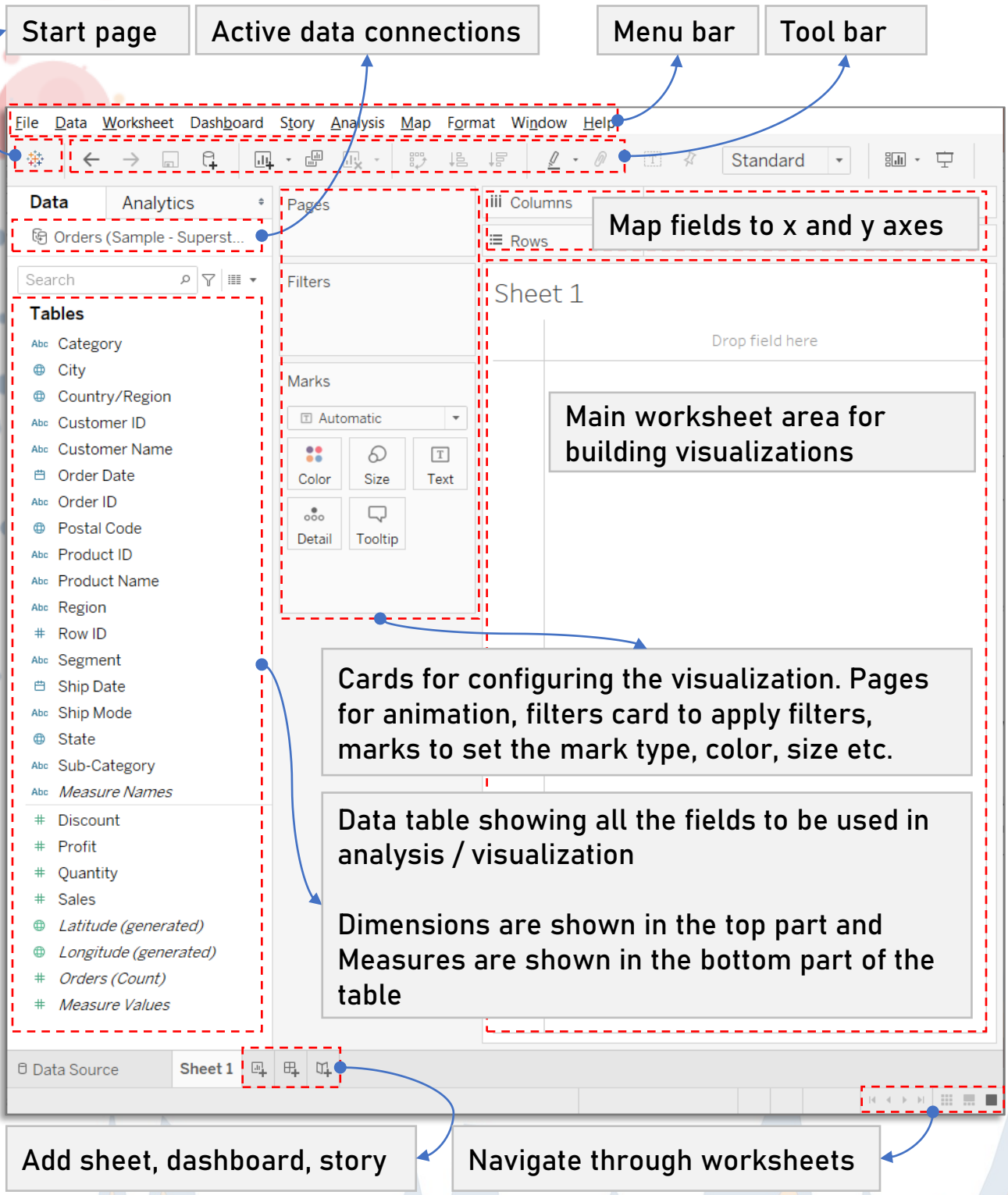
Superstores Dataset

- Dataset contains the sales info of Super stores.
- Use the dataset provided in the course to reproduce the same results and please do NOT use the dataset from the tableau standard installation.
- 3 sheets: orders – order info, returns – returns orders info, people – names of the sales people for each region

Connecting Data to Tableau

- Two major connection types: to a file and to a server
    - To a file: MS excel, text or comma separated values, stat files etc.
    - To a server: SQL, Google Sheets, SAP, dropbox etc.
- Add / edit connections from the connections pane
- Drag drop tables (sheets from the excel source) from the left to the data source pane on the right to get started and preview before getting started

Tableau Interface



Data Types and Terminologies

- 7 different data types in Tableau
    1. Number (Whole) – same as integer
    2. Number (Decimal) – same as float
    3. String
    4. Boolean
    5. Date
    6. Geographical
    7. Mixed
- Measure and Dimension fields
    - Measures: Anything numerical that can be aggregated using sum, average etc. e.g. Sales, Profit
    - Dimensions: Anything that allows grouping of numbers e.g. Category, Region etc.
- Continuous and Discrete fields
    - Continuous: Any field that has continuous values. All decimal numbers are continuous by definition. Continuous fields are shown in green color in the table pane.
    - Discrete: Any field that is distinct – all dimensions are discrete by definition. Depending on circumstances whole numbers and also dates can be treated as discrete fields. Discrete fields are shown in blue color in the table pane.

4 Ways to Generate and Share Output

1. Copy as image (applies to tableau desktop only) – image can be pasted anywhere
2. Save as TWBX (applies to tableau desktop only) – can be opened in tableau reader
3. Export as PDF / PPT (applies to tableau desktop only) – can be opened using PDF or PPT
4. Save on tableau public cloud and share the link (applies to both tableau desktop and tableau public)

Note: one could save file as TWB (tableau workbook). But when a TWB file is shared, one needs to connect to the dataset locally. This makes the workbook accessible only using tableau desktop / server developer program. However, TWBX packs the dataset and workbook together (tableau packed workbook) and can be opened using the free tableau reader program, or can be saved on tableau public as TWBX and opened through web browser or using tableau public app.

Aggregation

- Numerical Measures values need to be shown using aggregation because it is hard to visualize the raw values without summarizing.
- Several methods exist to summarize data (a.k.a. descriptive statistics): Sum, Average, Median, Count, Count(Distinct), Maximum, Minimum, Percentile, Standard Deviation, and Variance.
- Most commonly used aggregation methods are Sum and Average which show the sum and average respectively of all data points in the measure field used in the visualization.
- Default aggregation is 'Sum'. To change the aggregation, right click on the measure field in the visualization → Measure → Select aggregation method

Using Marks Card

- Colors can be used to show different members dimension e.g. category → color shows different categories in different color
- Legend card appears showing the mapping of colors with the field used in the color card
- Continuous fields on Colors are shown using continuous color scale
- Size can play a similar role as color. Only caveat is that size cannot be mapped to a negative number in principle.

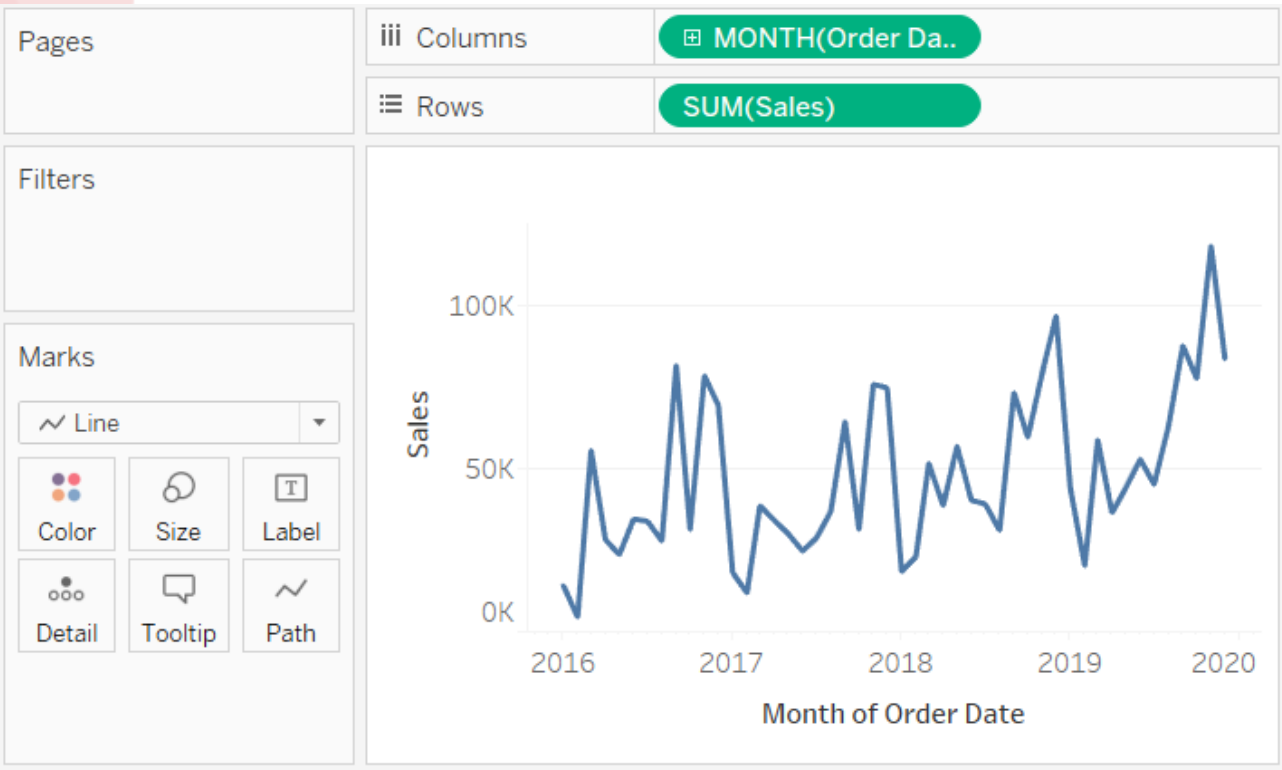
Charting Method

- Cartesian coordinate system (e.g. line chart, bar chart, scatter plot)
  - X and Y axes used to locate a point in the planes
  - Variable / fields mapped to the X and Y coordinates
  - 'Measures' on both axes → Scatter plot
  - 'Measure' and 'Dimension' on each of the axes → Bar chart
  - More fields can be mapped to other elements such as color and size of the marks
- Polar coordinate system (e.g. pie chart, donut chart)
  - 'Angle' and 'Radius' used to locate a point in the planes
  - Any 'measure' is converted into 'angle' by calculating the % of total and then calculating the angle out of 360°
  - Pie charts are the polar coordinate equivalent of simple bar charts
  - Charts constructed using polar coordinate system can be tricky to read

Line Chart

How to Construct?

Date field (as continuous field) → Columns
Measure → Rows
Mark → Line



Note

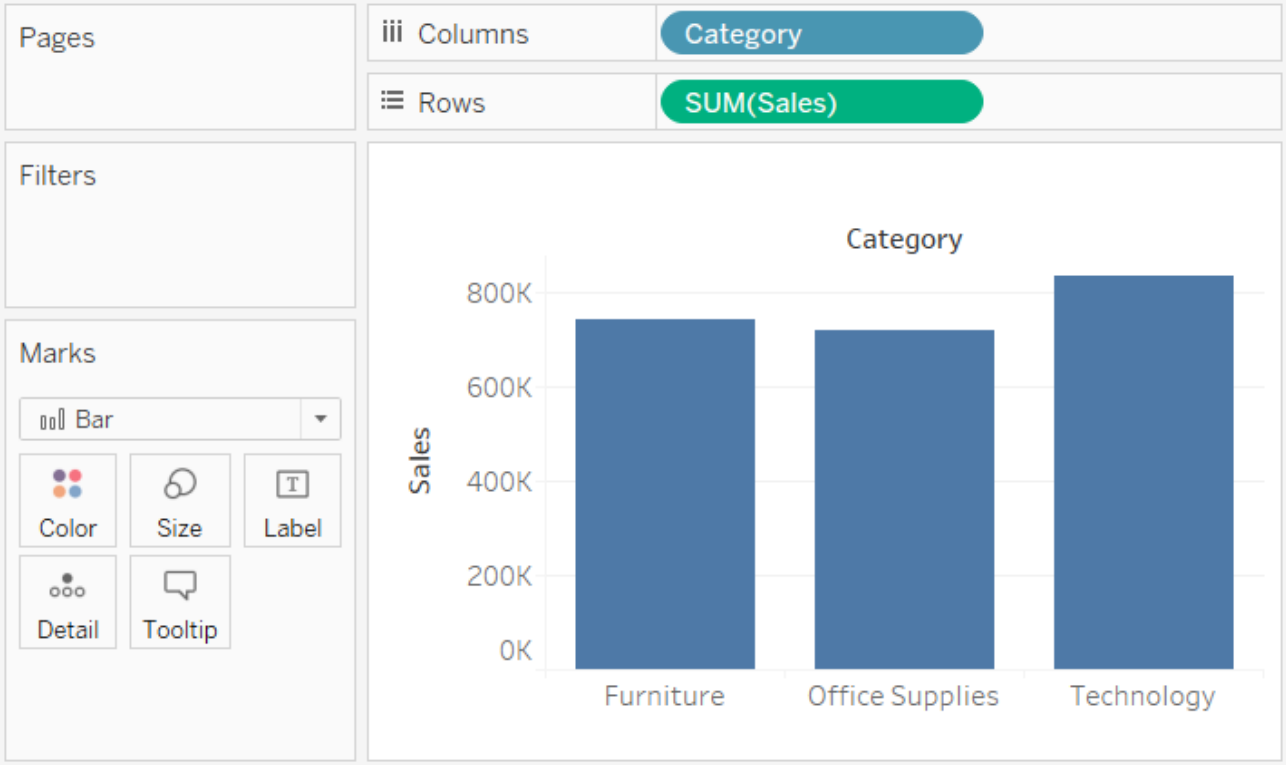
Best for showing trends

In a line chart, it is inherently assumed that the values in the x-axis is continuous

Vertical Bar Chart or Column Chart

How to Construct?

Dimension → Column
Measure → Rows
Mark → Bar



Note

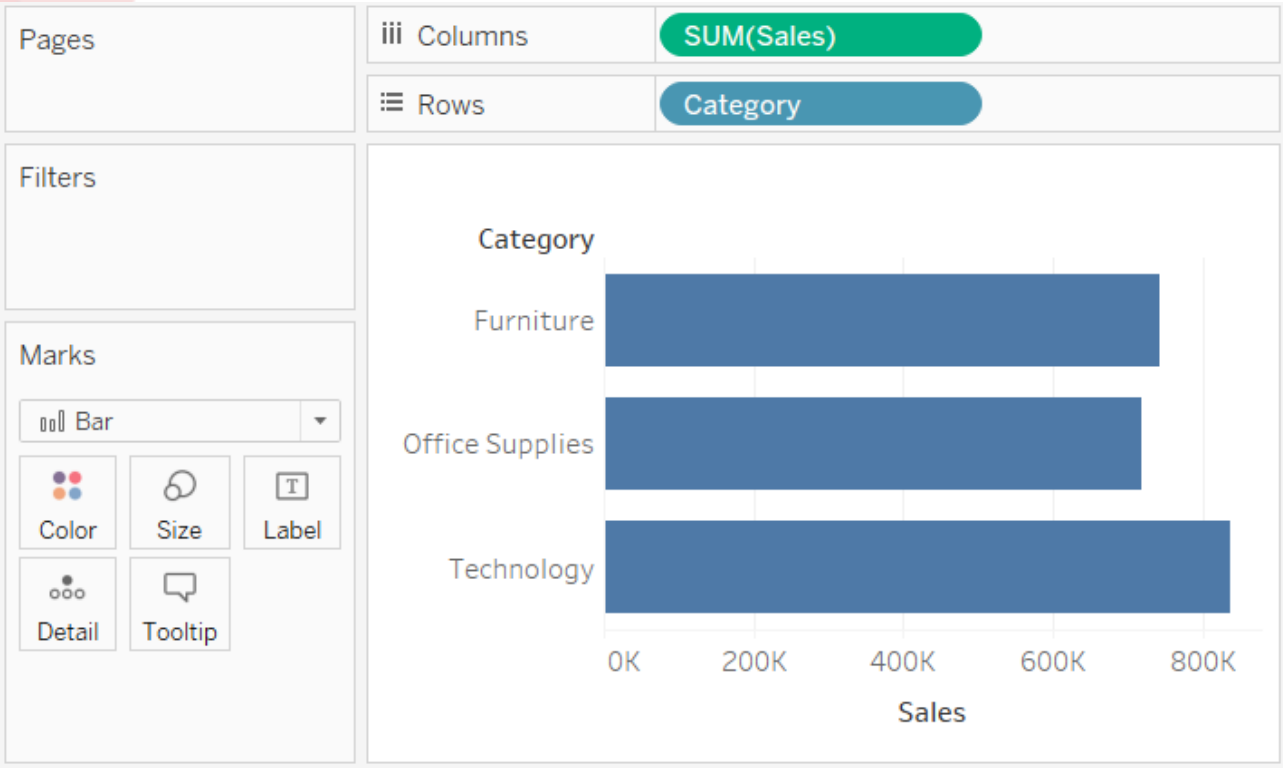
Best for categorical comparison of values. e.g. sales per category, sales per region, profit per region etc.

Depending on the real estate available, show chart as a column chart or bar chart. If the chart area is landscape and the names of the categories can be shown clearly, use a column chart.

Horizontal Bar Chart

How to Construct?

Dimension → Rows
Measure → Columns
Mark → Bar



Note

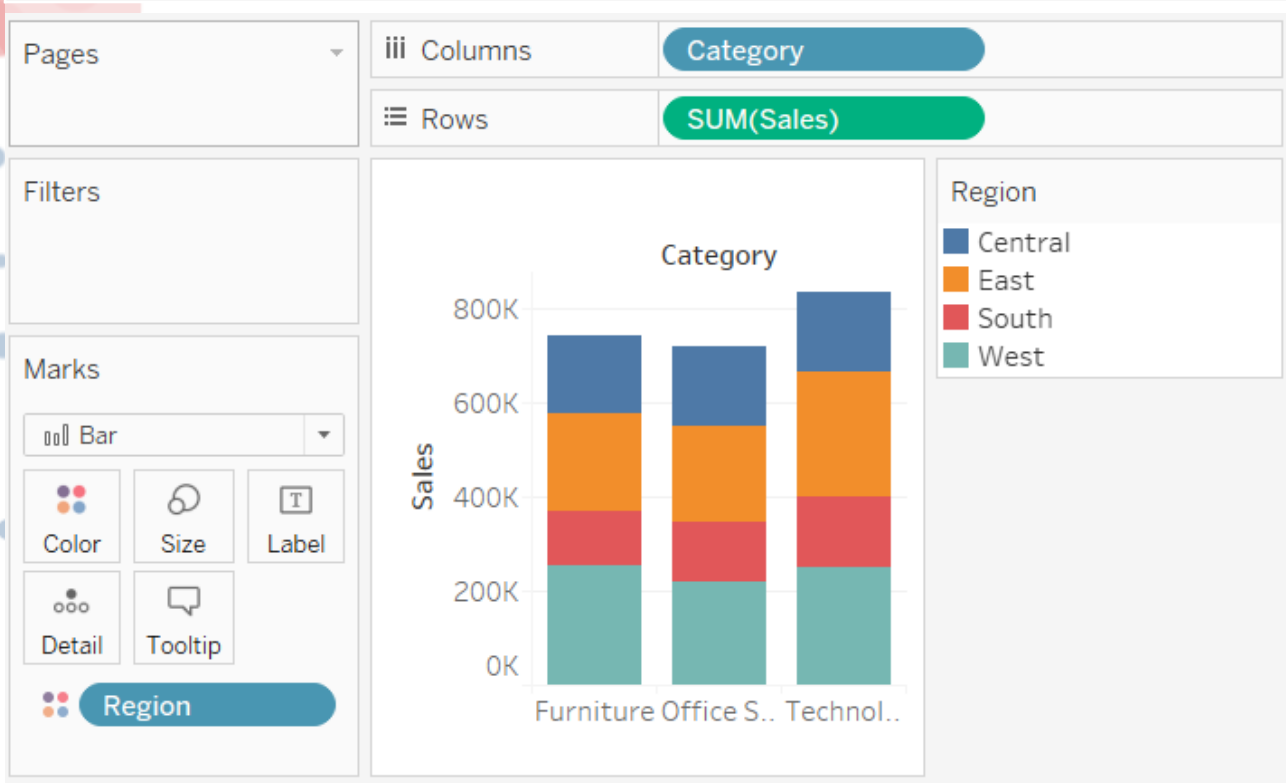
Best for categorical comparison of values. e.g. sales per category, sales per region, profit per region etc.

Depending on the real estate available, show chart as a column chart or bar chart. If the chart area is portrait and if there are many members within a dimension use a horizontal bar chart.

Stacked Bar Chart

How to Construct?

Same as bar chart +
Additional dimension → Color



Note

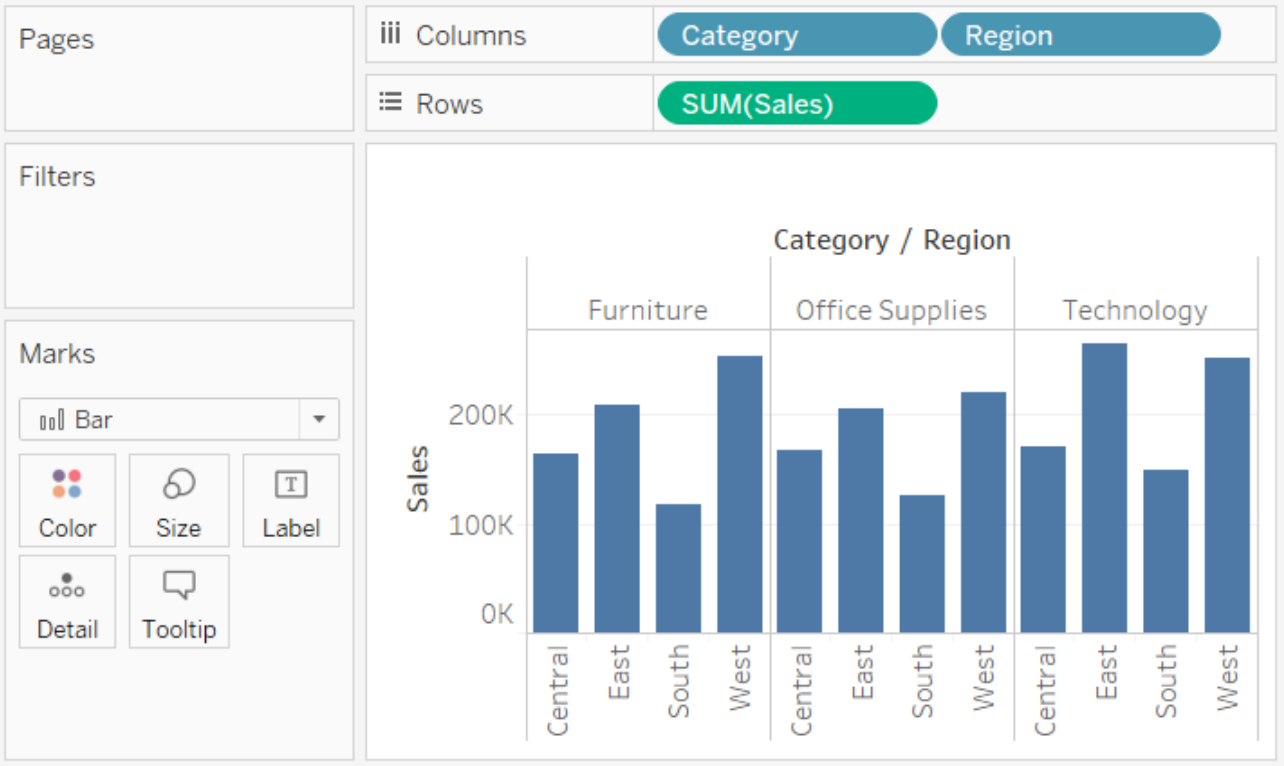
Best for categorical comparison of values across more than one dimension.

Primary variable for which the values are compared is used in the columns while the secondary variable used to compare within a column or to show the composition of the column can be used in the color.

Side-by-side Bar Chart

How to Construct?

Same as a bar chart +
Additional dimension → Columns / Rows depending on whether it is a vertical / horizontal side by side bar chart



Note

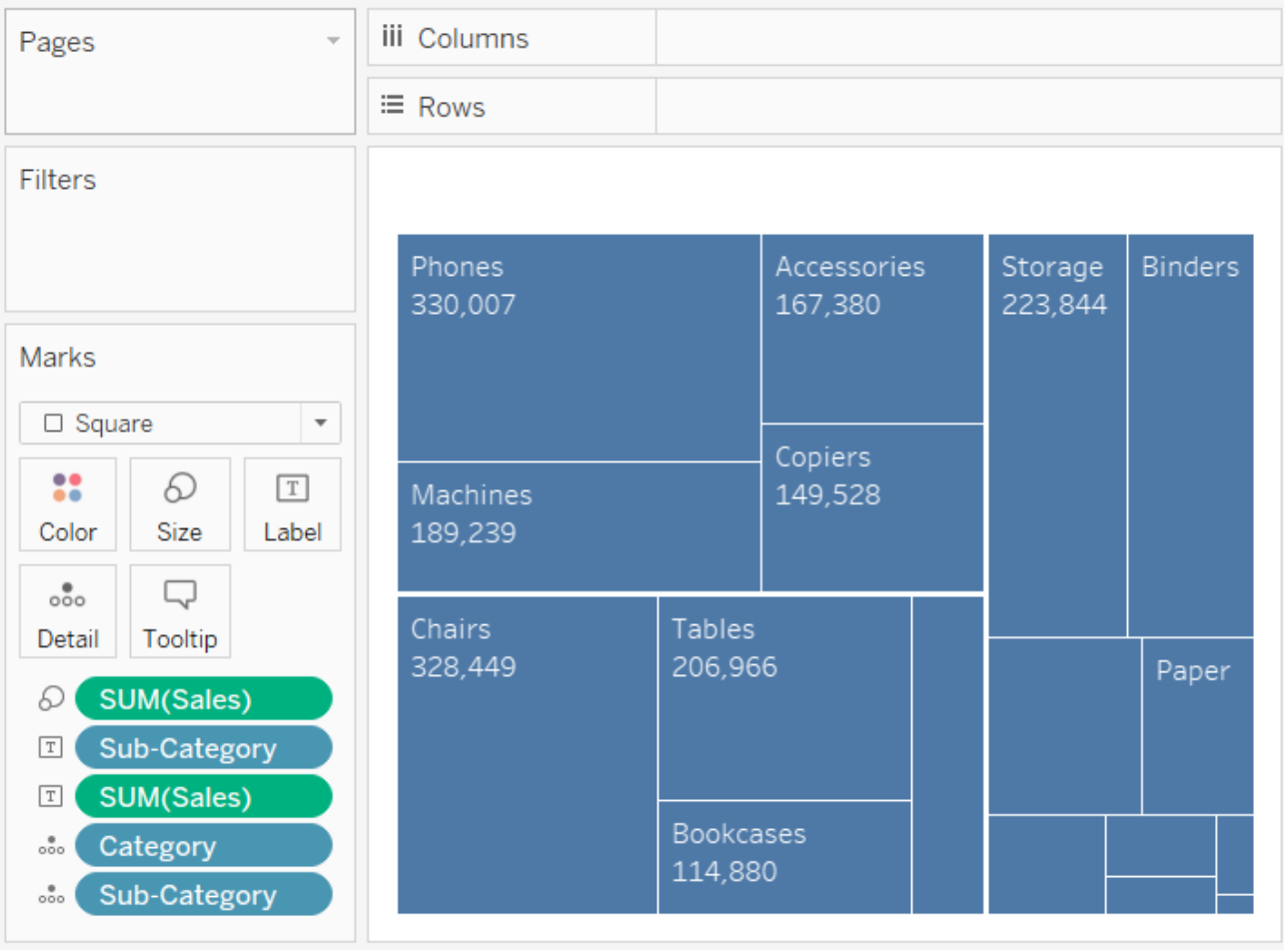
Best for categorical comparison of values across more than one dimension. Unlike stacked bar chart, side-by-side bar charts creates multiple bar charts for each of the member used as the main variable in the columns (in this case).

Compared to stacked bar chart, this can use more space but easier to read without any legend.

Tree Maps

How to Construct?

Dimension → Detail
Measure → Size
Mark → Square
Enable text labels with necessary fields
Additional Dimension → Detail
(in the order of hierarchy e.g. Category > Sub-category)



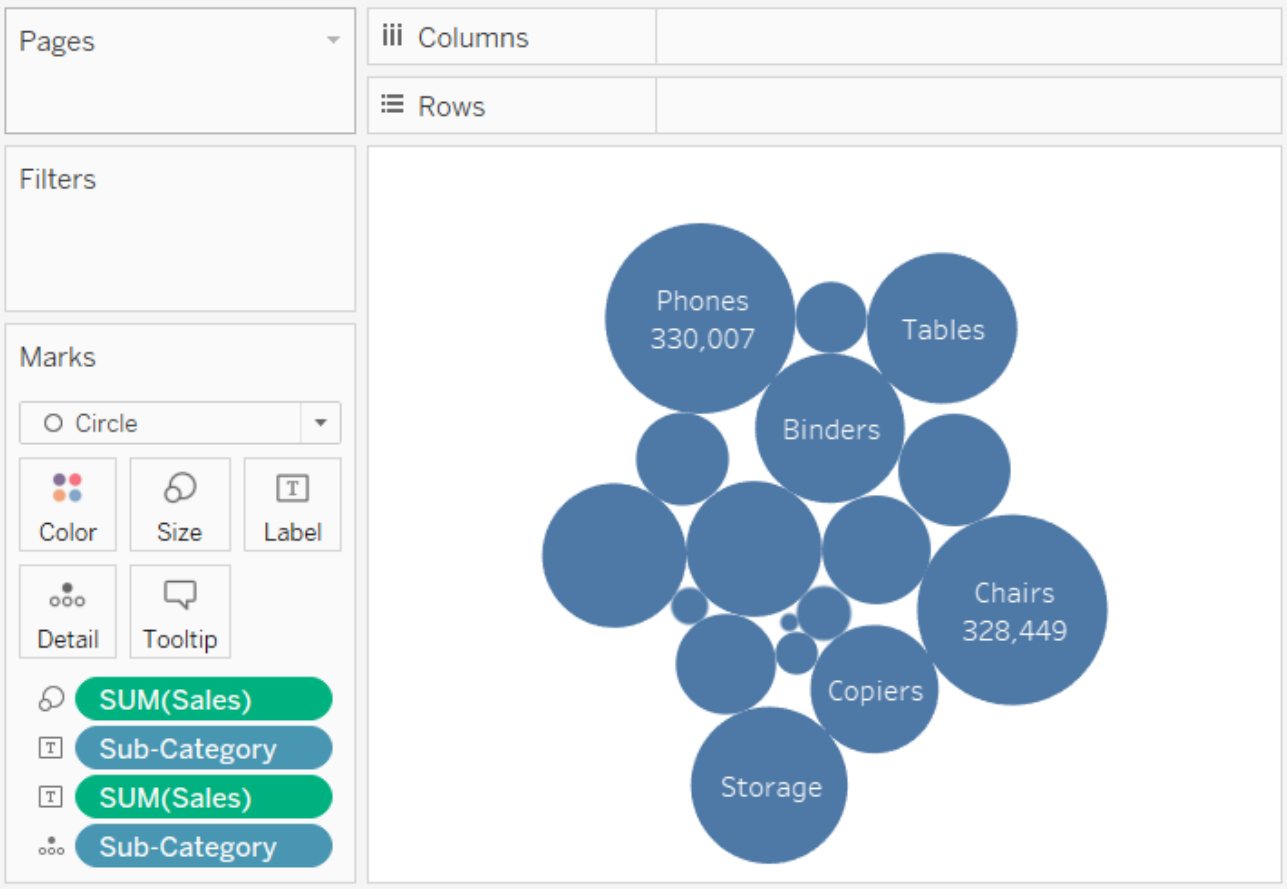
Note

Tree maps are great to compare values categorically. Though visually nicer, unlike bar charts they can be difficult to read with more categories. Works best when emphasizing the top contributing categories rather than reporting all categories.

Packed Bubble Chart

How to Construct?

Same as Tree maps with one slight difference:
Mark → Circle



Note

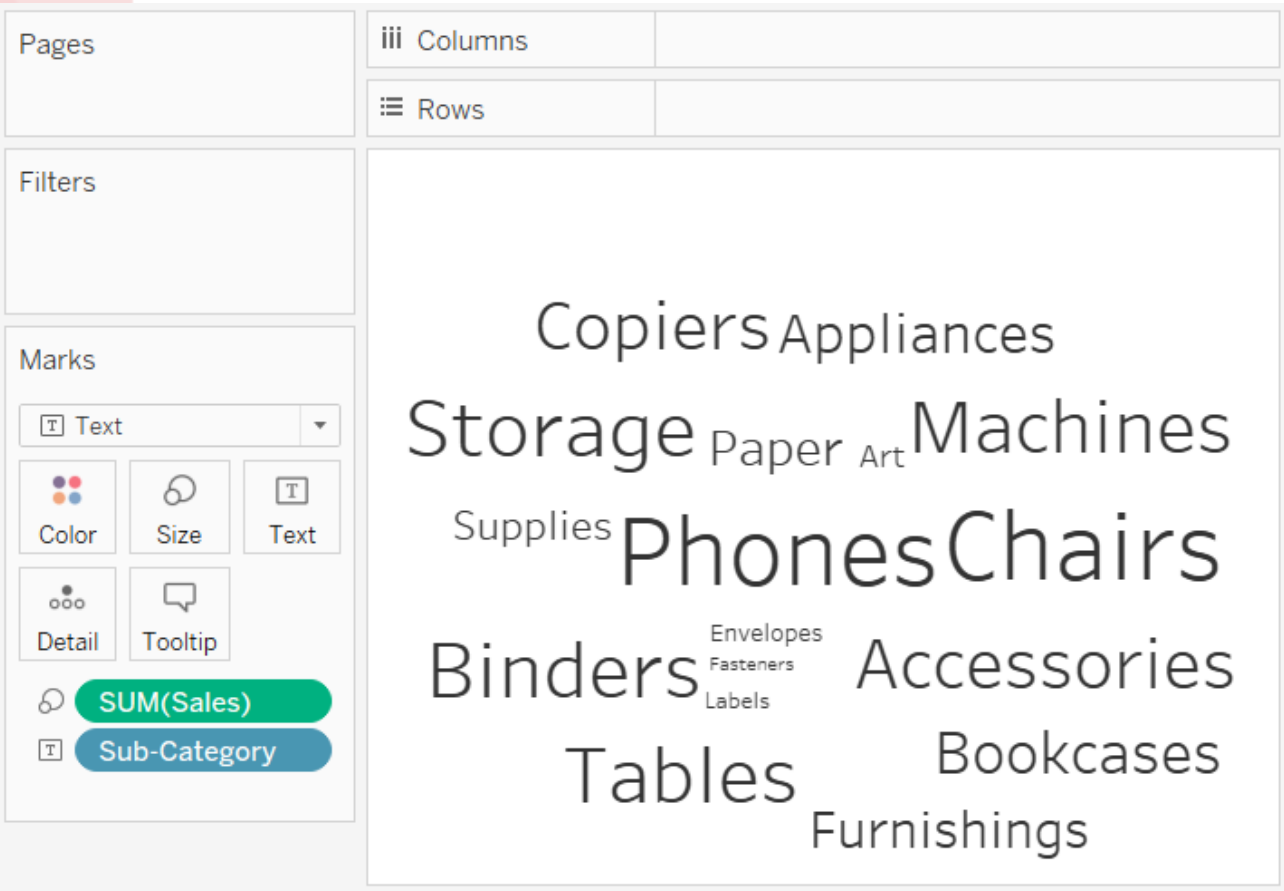
Works similar to the tree maps but these can be difficult to read especially because it is hard for human eyes to compare the size of the circles than the squares/rectangles. Best to highlight top items or used as filter.

Filters are explained later.

Word Cloud

How to Construct?

Mark → Text
Dimension → Text
Measure → Size



Note

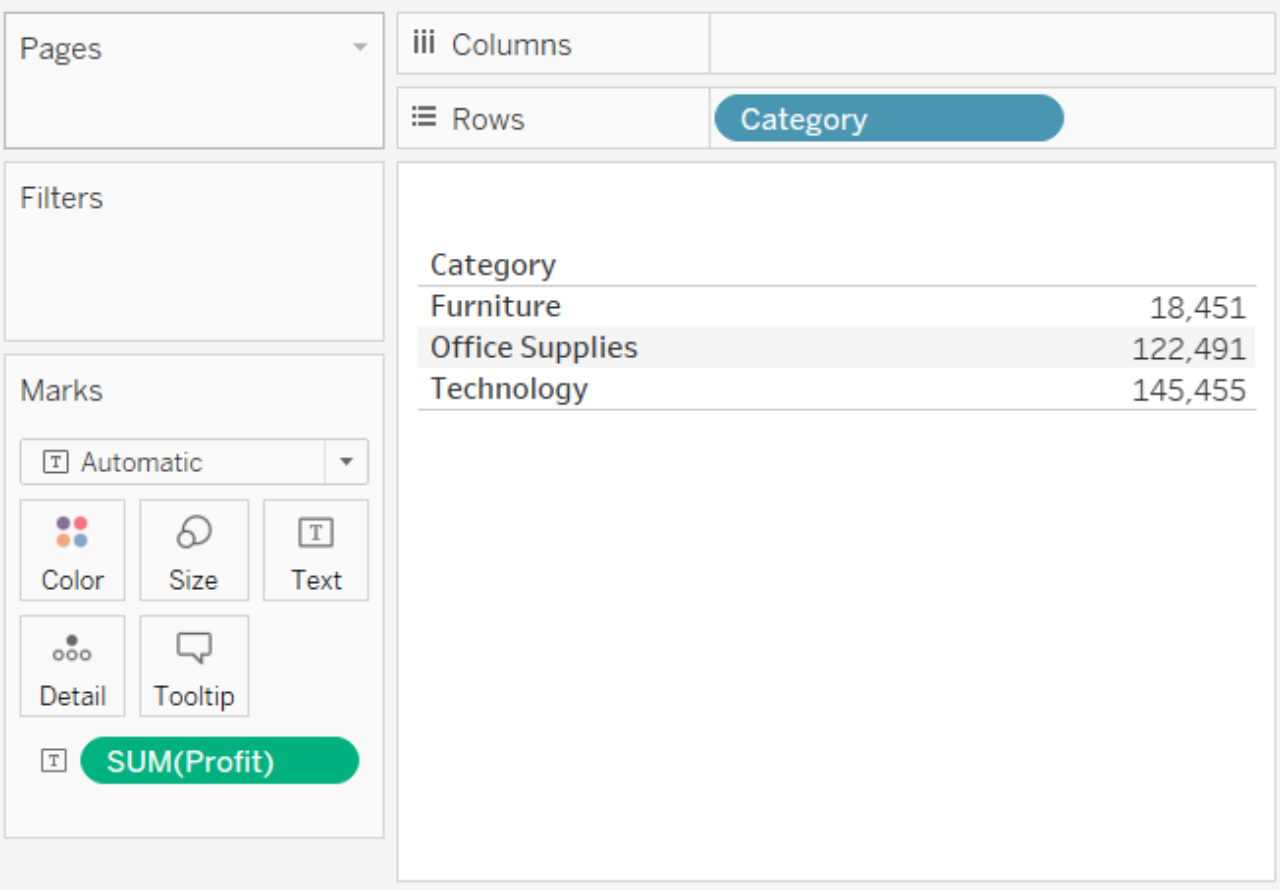
Works similar to the tree maps but these can be difficult to read especially because it is hard for human eyes to compare the size of the text than the squares/rectangles. Best to highlight top items or used as filter.

Filters are explained later.

Text Tables

How to Construct?

Dimension → Rows / Columns
Measure → Table area where it says 'Abc'



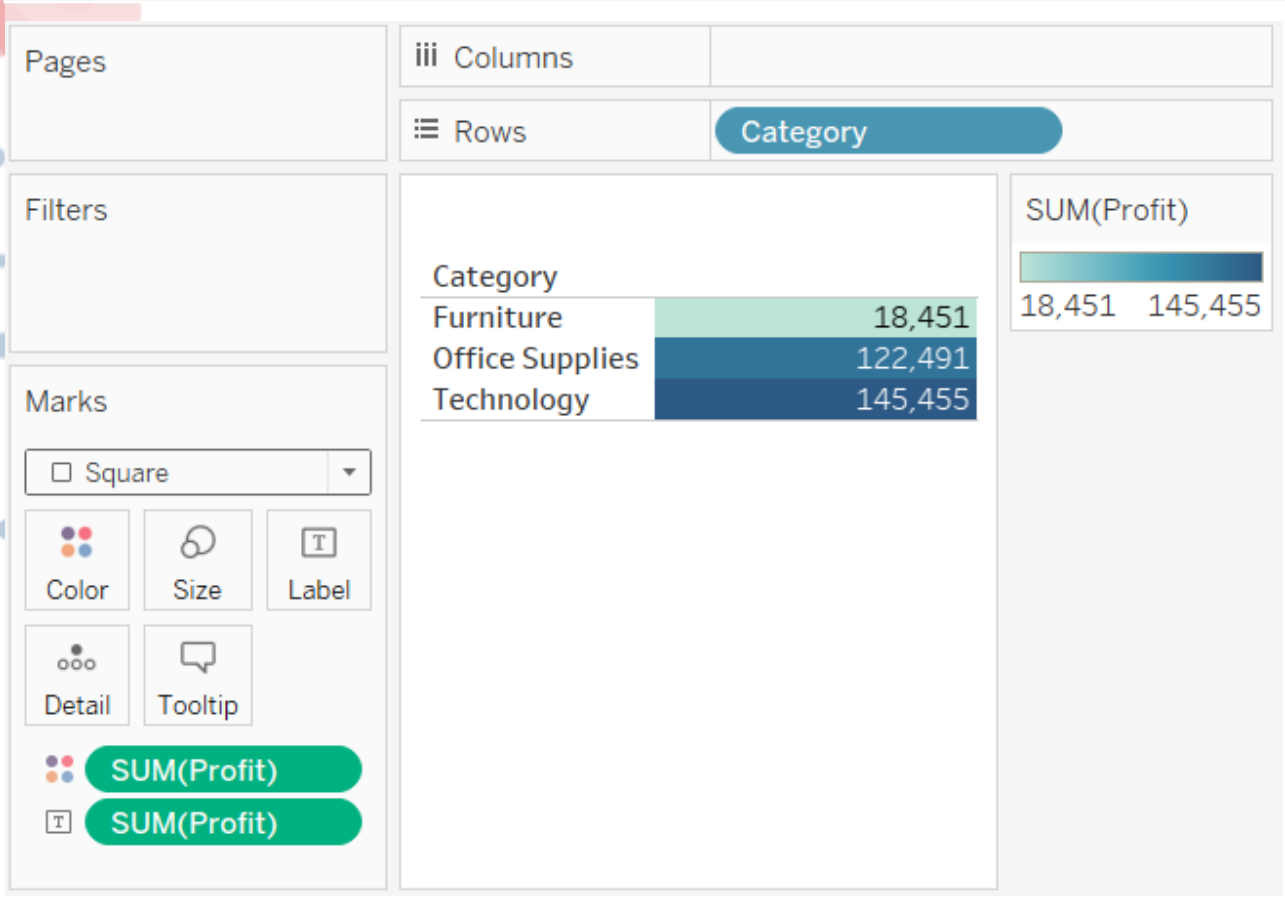
Note

Great for reporting precise values, however, with more values can clutter. Used mainly in financial reports where every line needs to be looked at and work well when sorted properly.

Highlight Tables

How to Construct?

Same as text tables +
Measure → Color
Mark → Square



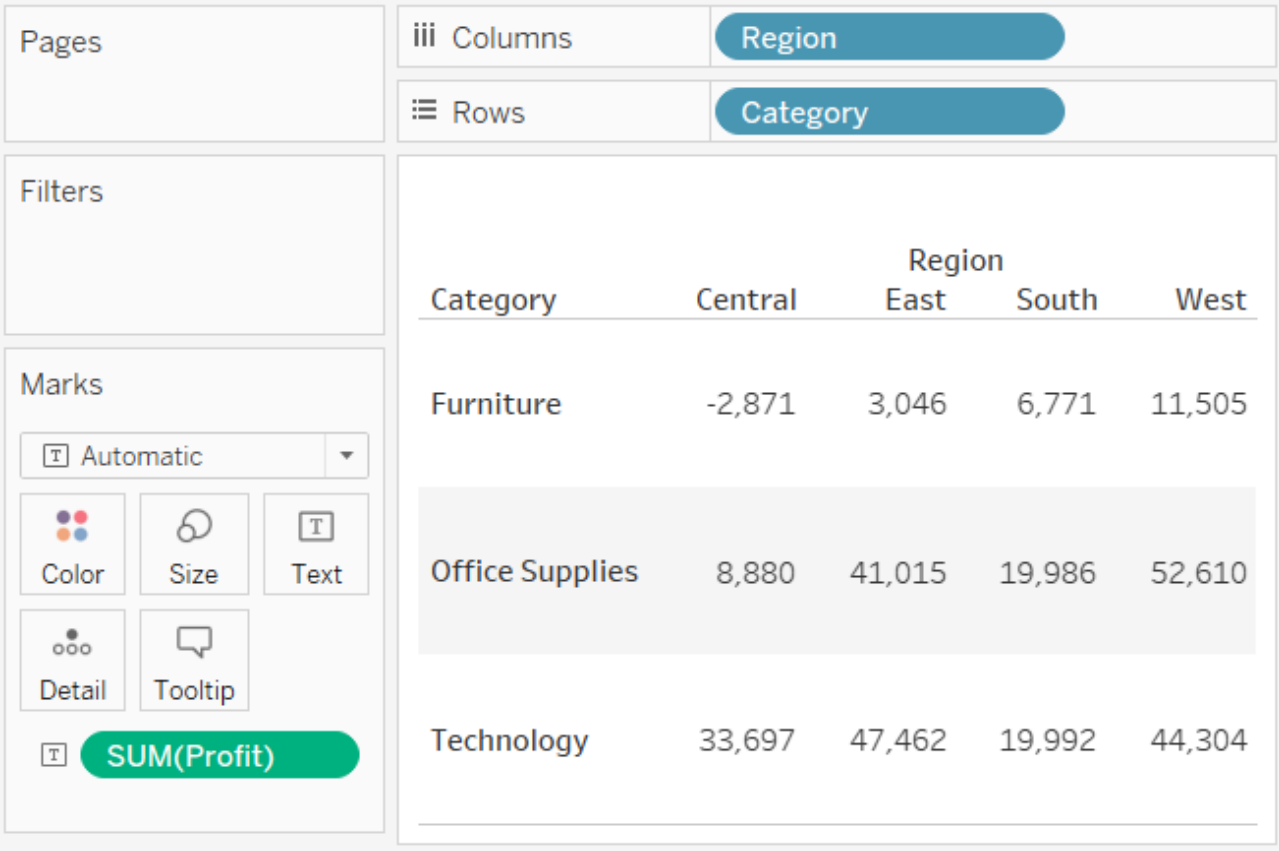
Note

Adding highlights draws attention to the most important values

Cross Tables

How to Construct?

Dimension1 → Rows
Dimension2 → Columns
Measure → Table area where it says 'Abc'
Highlight color if needed → same as highlight tables



Category	Region: Central	East	South	West
Furniture	-2,871	3,046	6,771	11,505
Office Supplies	8,880	41,015	19,986	52,610
Technology	33,697	47,462	19,992	44,304

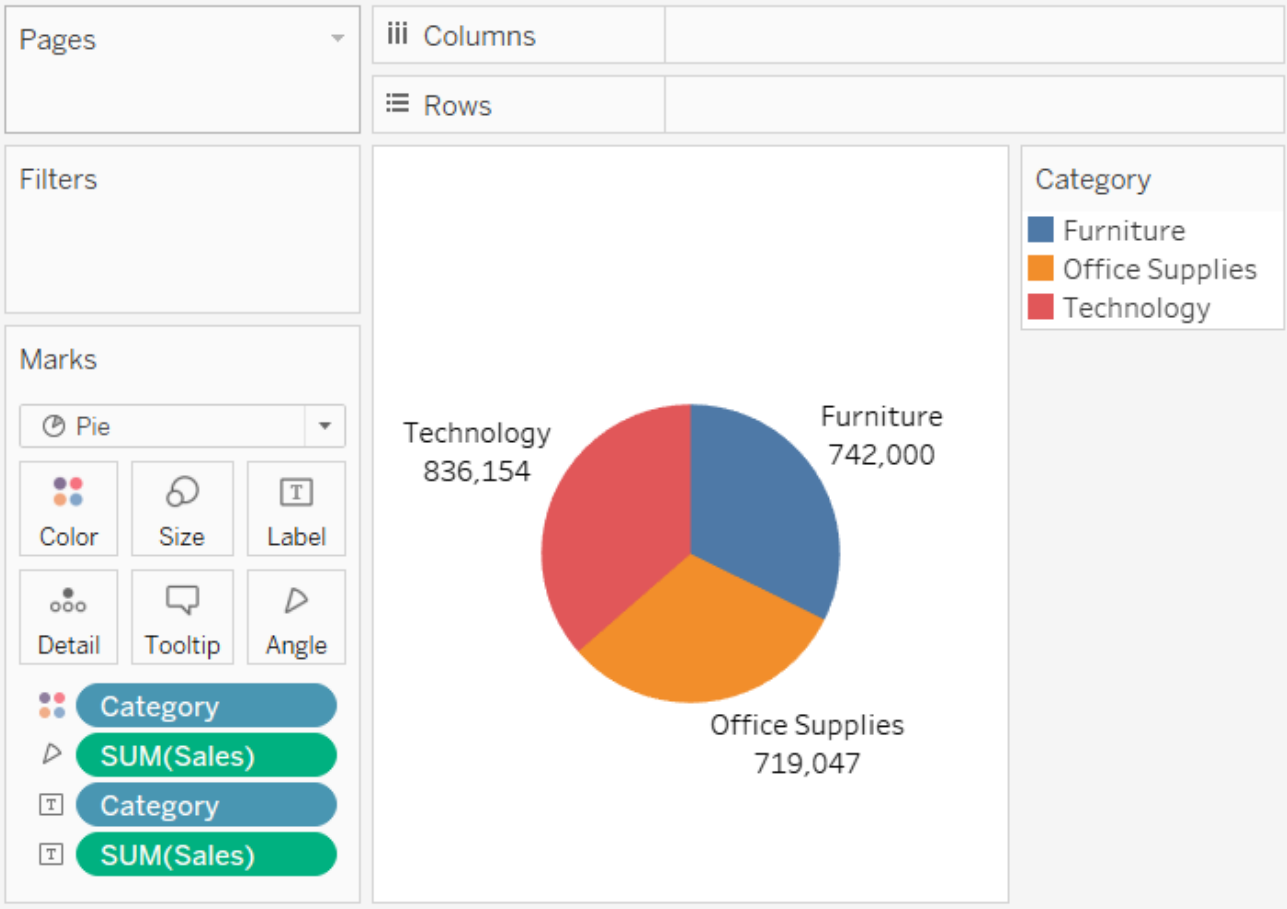
Note

Cross tables are great for reporting the values by comparing two dimensions simultaneously. Any other chart for this purpose would be hard to read compared to a cross table.

Pie Chart

How to Construct?

Mark → Pie
Dimension → Color
Measure → Angle
Enable text labels with necessary fields



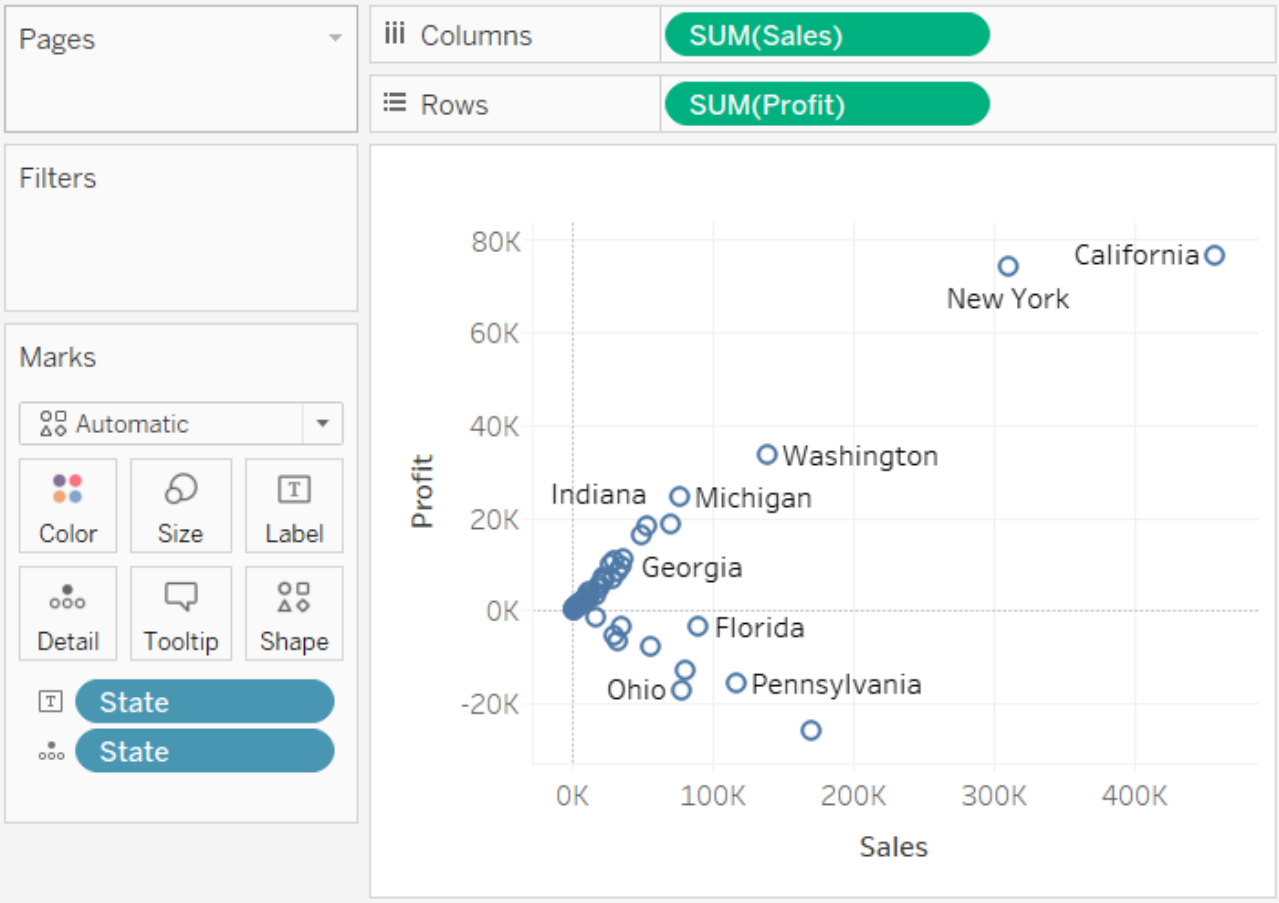
Note

Works similar to a simple bar chart. However, best to avoid when the number of categories becomes large limiting human perception of the sizes.

Scatter Plot

How to Construct?

Measure 1 (determining variable) → Columns
Measure 2 (resulting variable) → Rows
Analysis > uncheck Aggregate if every row is to be shown as a dot
if marks need to be grouped by a dimension: Dimension → Detail



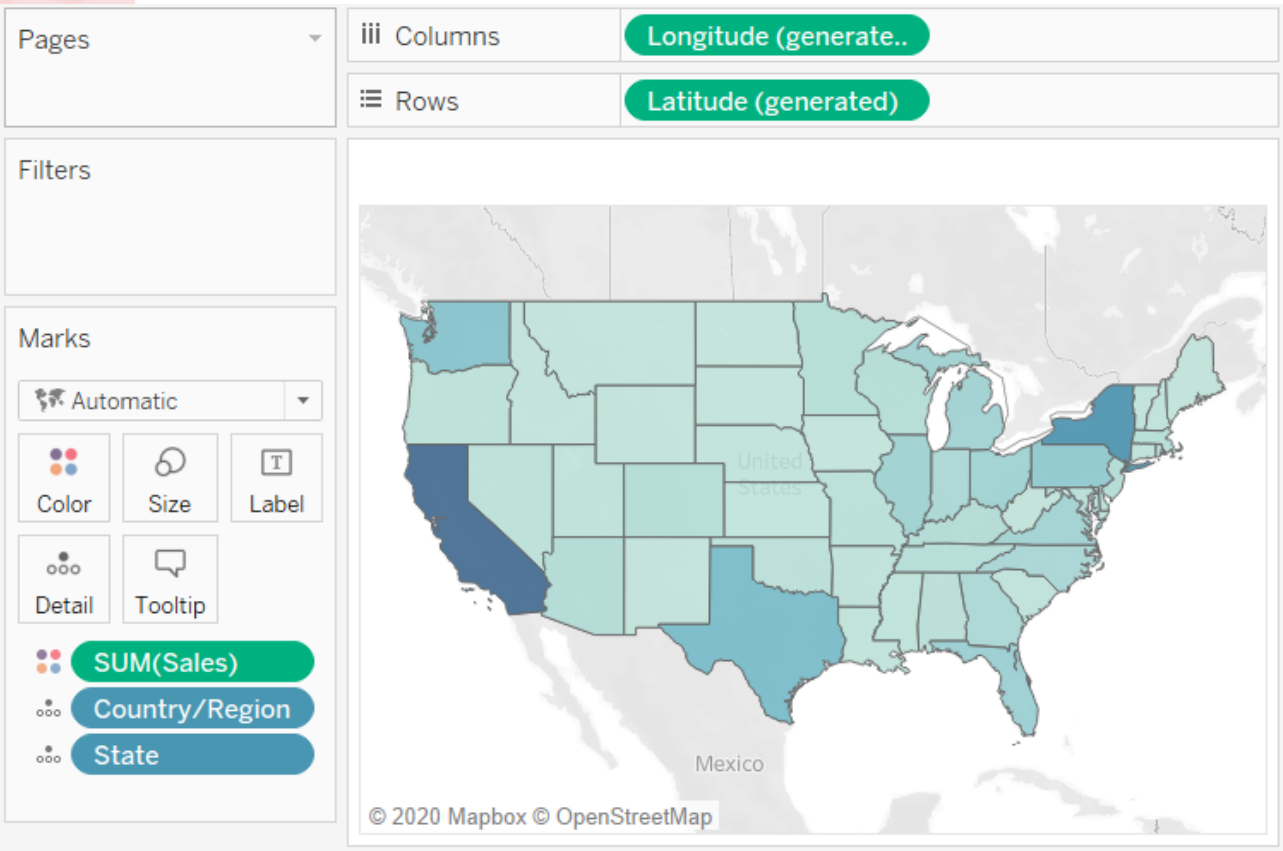
Note

Great for showing relationship between two numerical fields. 'Causing' field is shown in the x-axis and the 'Resulting' field is shown in the y-axis.

Fill Map

How to Construct?

Geographic Dimension Field(s) → Details (Use in hierarchy[1])
Measure → Color



Note

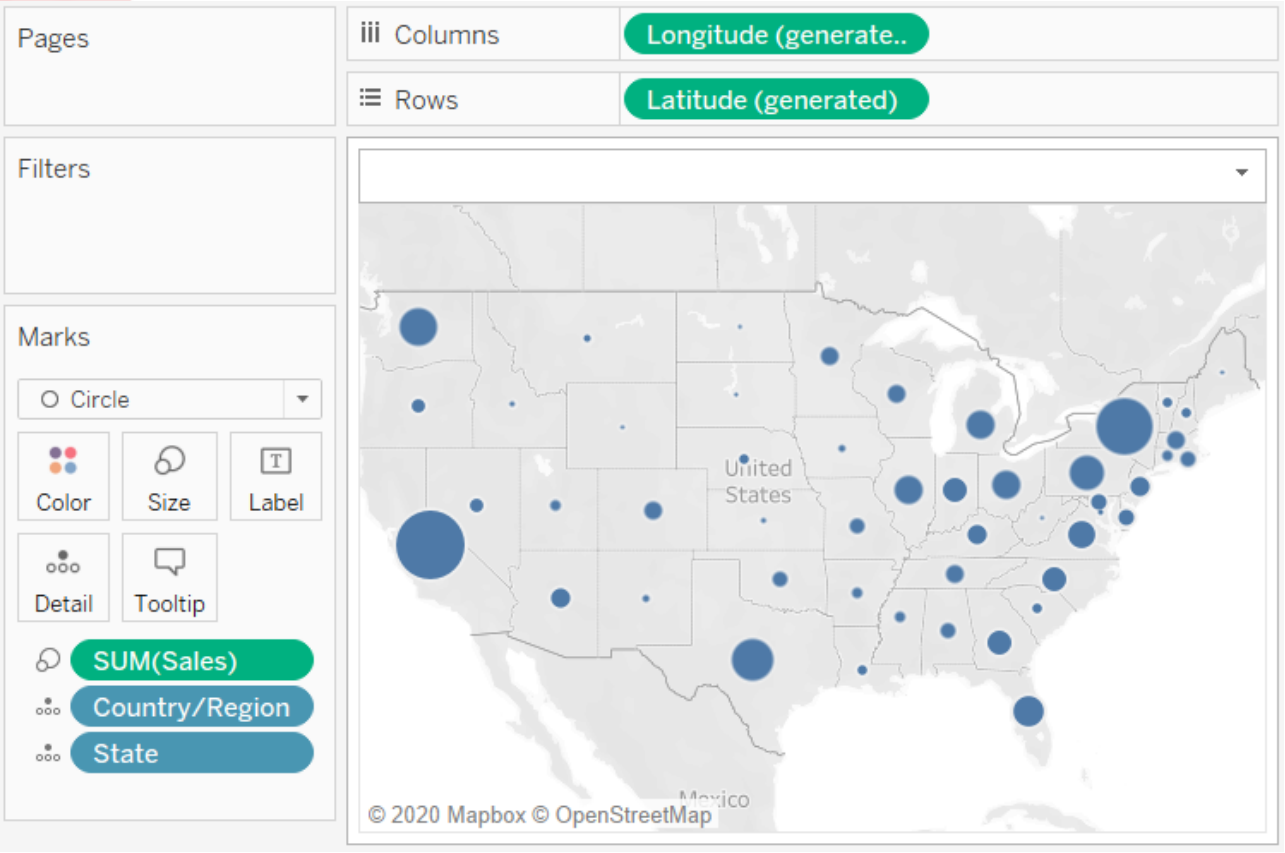
Maps are great for showing values across geography. For instance sales or profit across states or cities. Use fills where values could be negative as the size cannot be a negative value.

[1] Refer to 'fixing map errors' section if map doesn't appear

Symbol Map

How to Construct?

Geographic Dimension Field(s) → Details (Use in hierarchy[1])
Measure → Size and adjust the size

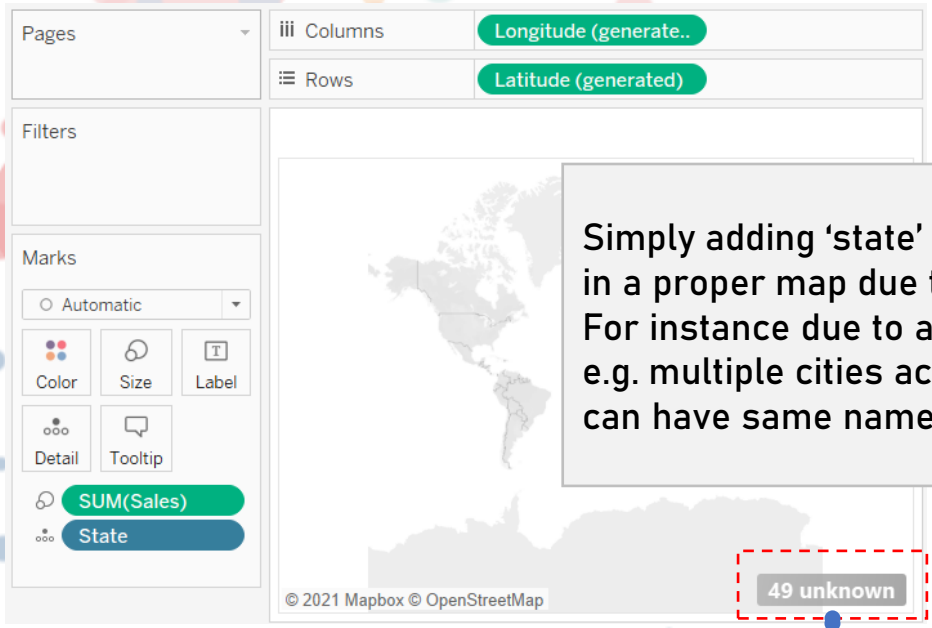


Note

Maps are great for showing values across geography. For instance sales across states or cities. Use symbol map where values are positive as the size cannot be a negative value.

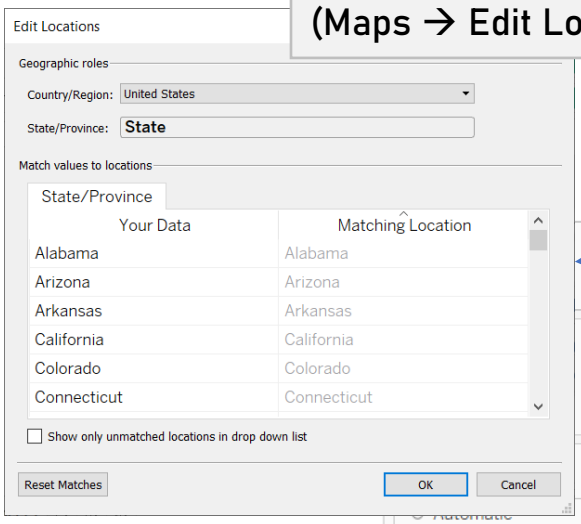
1 Refer to 'fixing map errors' section if map doesn't appear

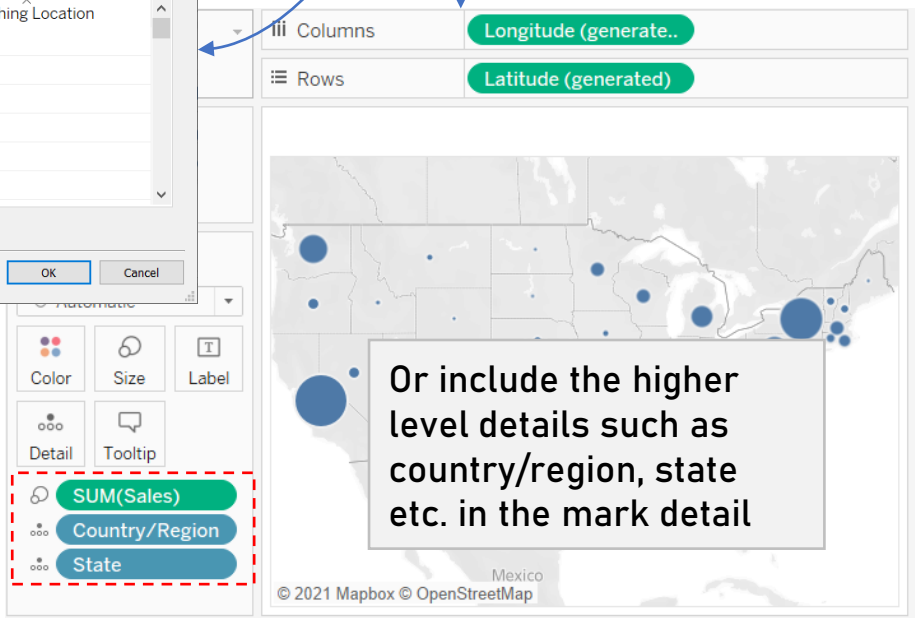
Fixing Map Errors



Simply adding 'state' field may not result in a proper map due to several reasons. For instance due to ambiguous names e.g. multiple cities across different states can have same names

Fix the Country/Region manually to set the higher detail in the map (Maps → Edit Locations)



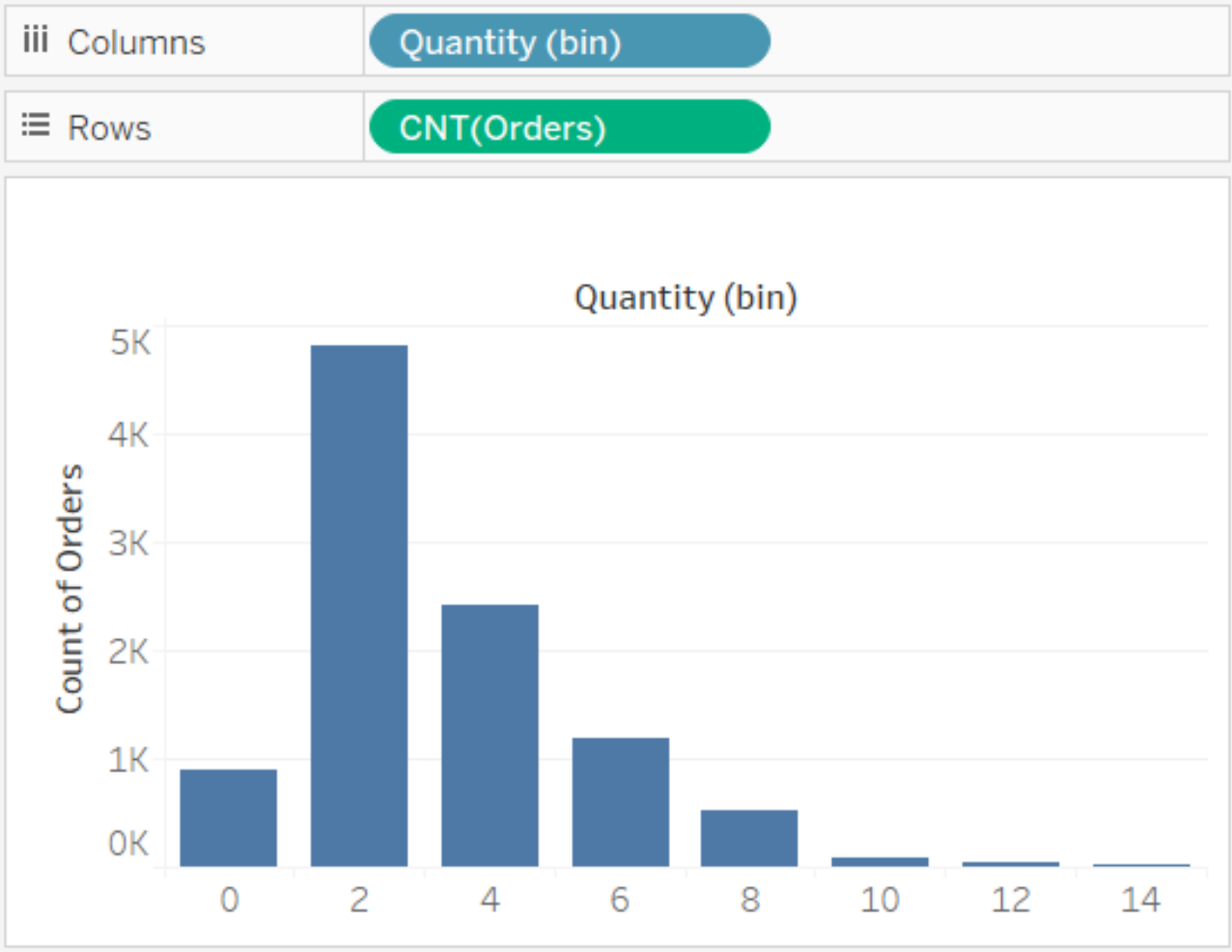


Or include the higher level details such as country/region, state etc. in the mark detail

Histogram

How to Construct?

Create bins: Right click on the measure > Create > Bins > Adjust the bin size if needed
Bin → Columns
Count of rows → Rows



Note

Histograms show how the data is distributed. One must carefully select the bin size so that the distribution is not too thinly spread not highly concentrated.

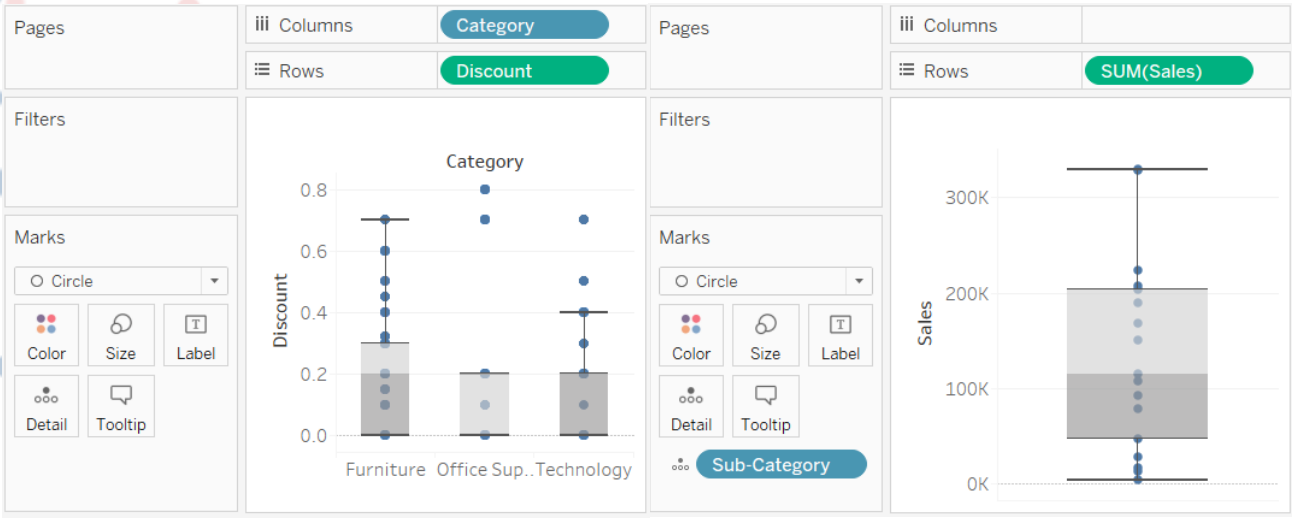
Boxplot

How to Construct unaggregated box plot?

Measure → Rows
Dimension → Columns
Menu > Analysis > Uncheck Aggregate Measures
Show me → Box and whisker plot

How to Construct aggregated box plot?

Measure → Rows
Dimension → Columns
Menu > Analysis > Uncheck Aggregate Measures
Show me → Box and whisker plot

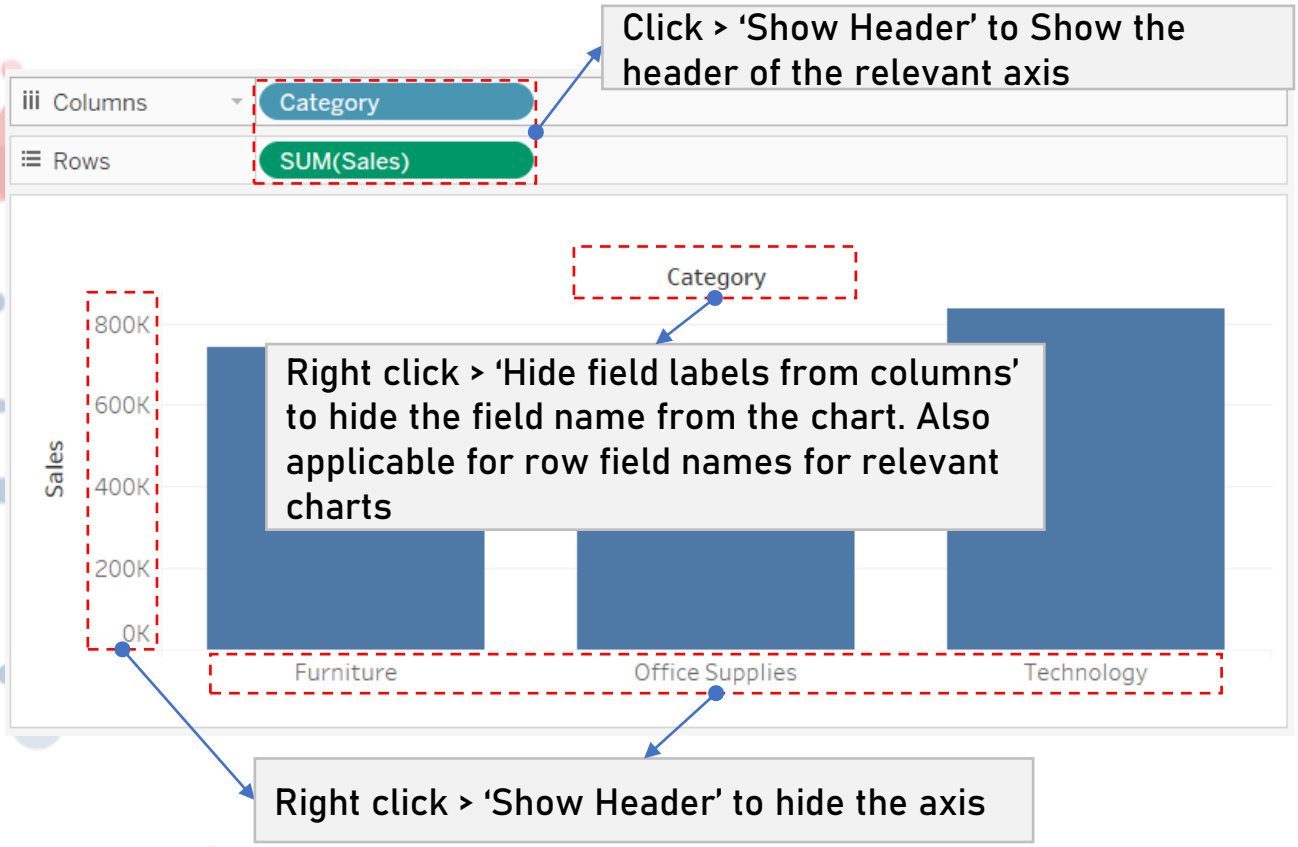


Note

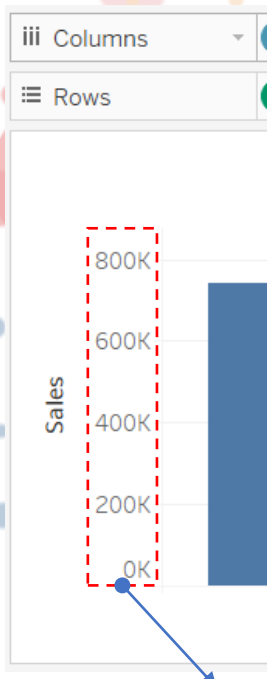
Boxplot show the distribution of the a measure field.

Boxplots show 5 point summary of the distribution: min, max, median, 25% and 75% percentile data points.

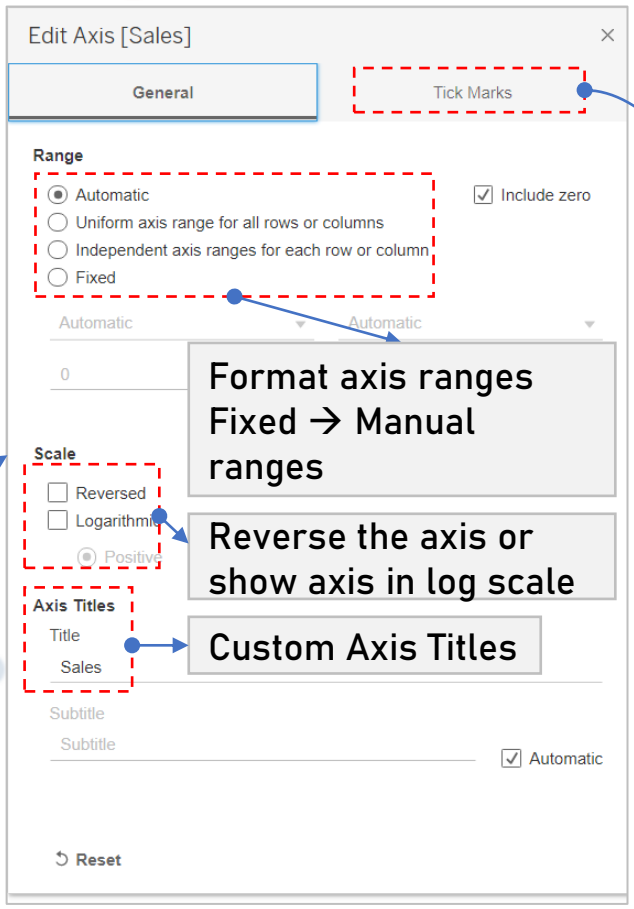
Formatting Chart Layout



Formatting Chart Axes



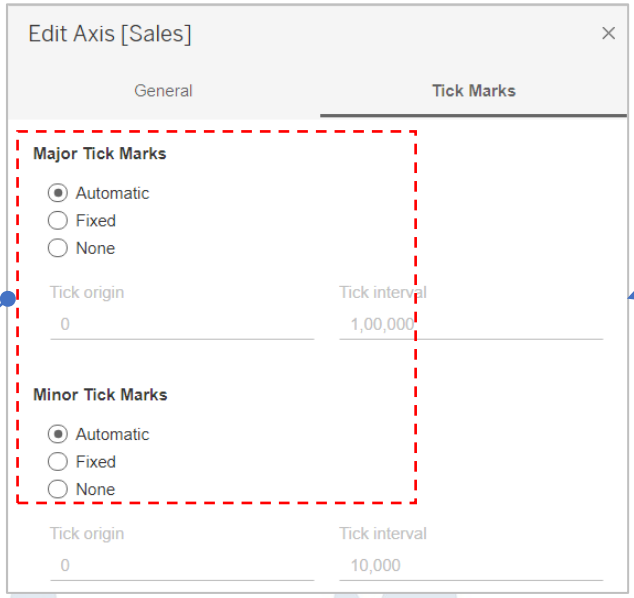
Right click >Edit Axis



Format axis ranges Fixed → Manual ranges

Reverse the axis or show axis in log scale

Custom Axis Titles

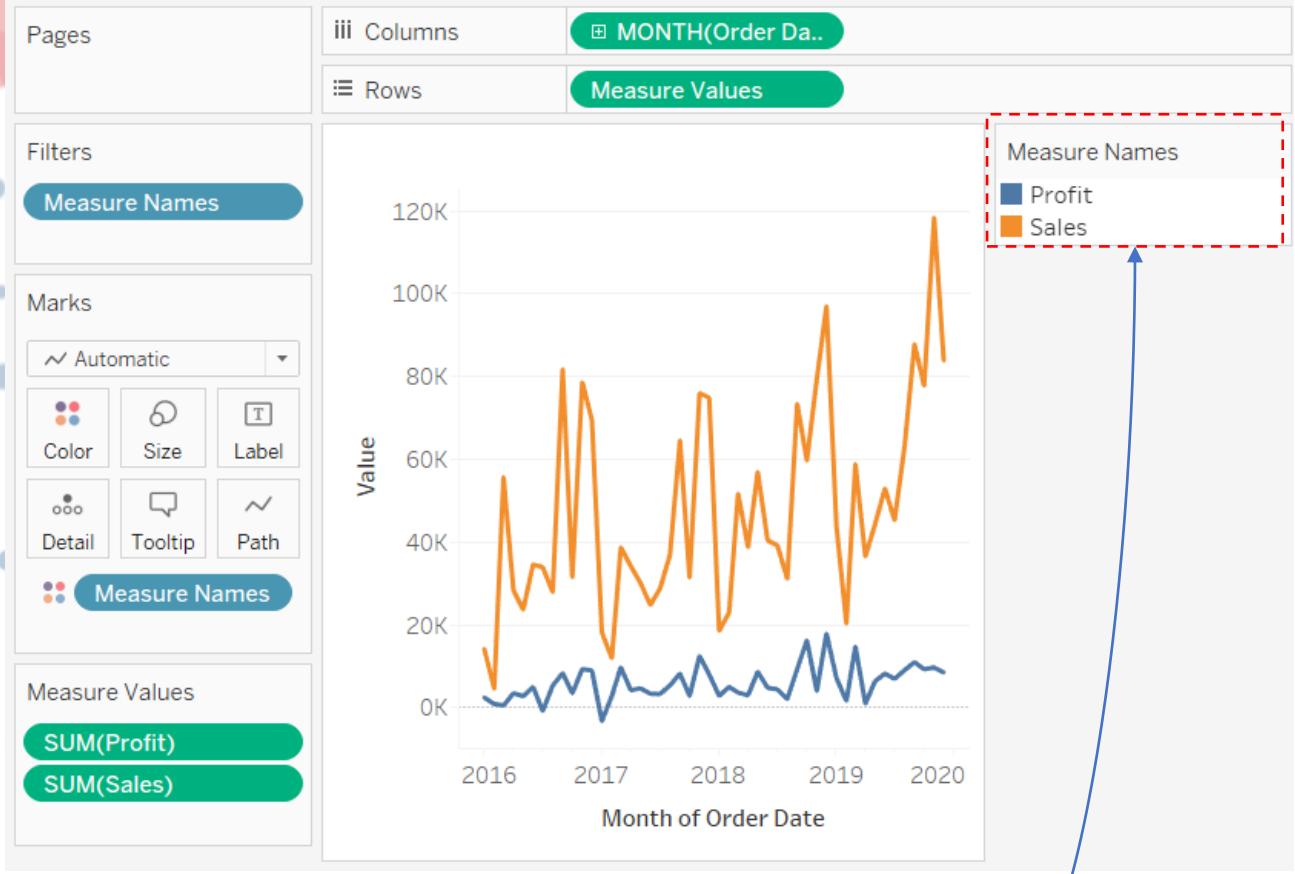


Adjust the axis tick marks

Combined Axis Chart

How to Construct?

Create a simple line chart with Measure 1
Measure 2 → Y-axis (not on the Rows shelf)



Note

Measures shown using different colors (see legend)

Measure Values are shown in the rows and Measure Names are shown in the filters. Only the measures used in the chart are selected in the filters

Any number of measures can be shown in this type of chart

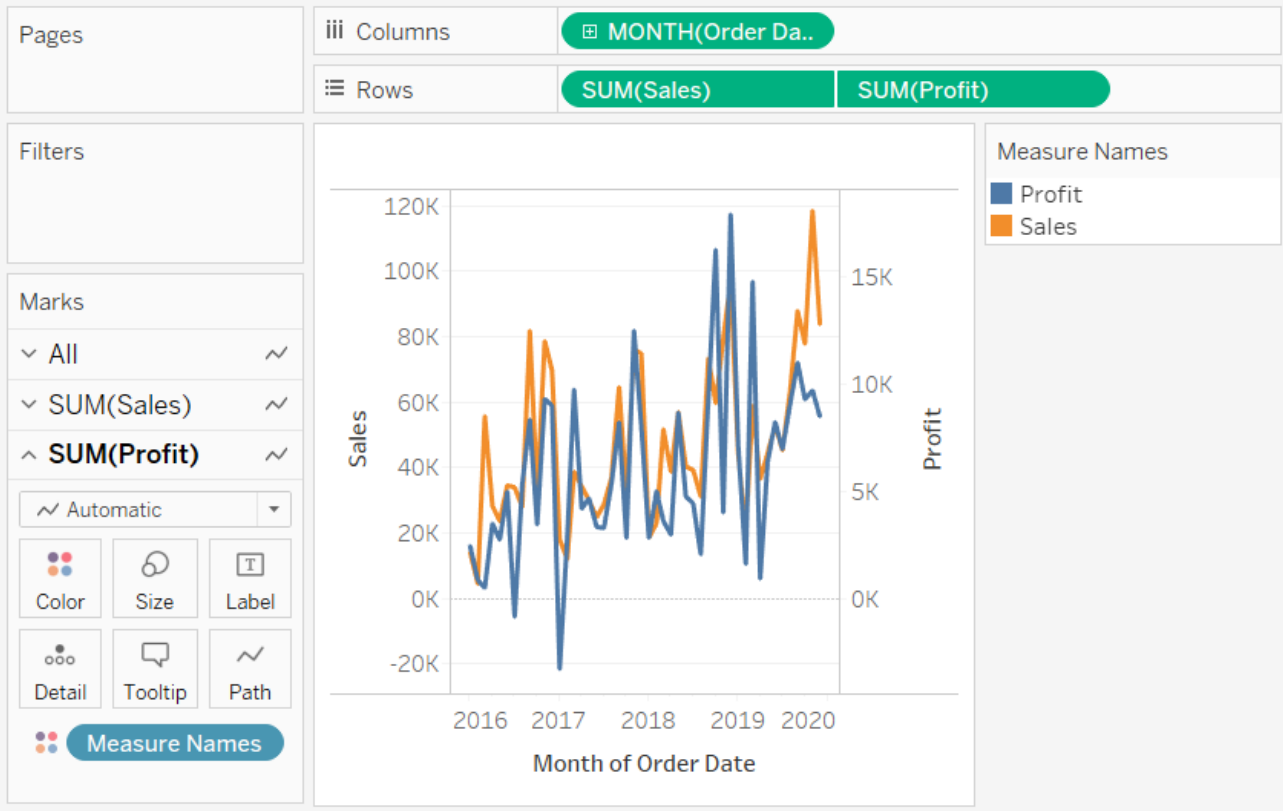
Note that since the axis is common for all measures, measures with large values may appear normal while the measures with smaller values might look compressed

Dual Axis Chart

How to Construct?

Create a simple line chart with Measure 1
Measure 2 → Rows (or columns is Measure 1 is in the columns)
This creates two independent charts using Measure 1 and Measure 2
Right click Measure 2 > Dual Axis

Optional: right click on any of the axis (right or left aka primary or secondary) and click synchronize axis. This makes both charts plotted using same axis ranges



Note

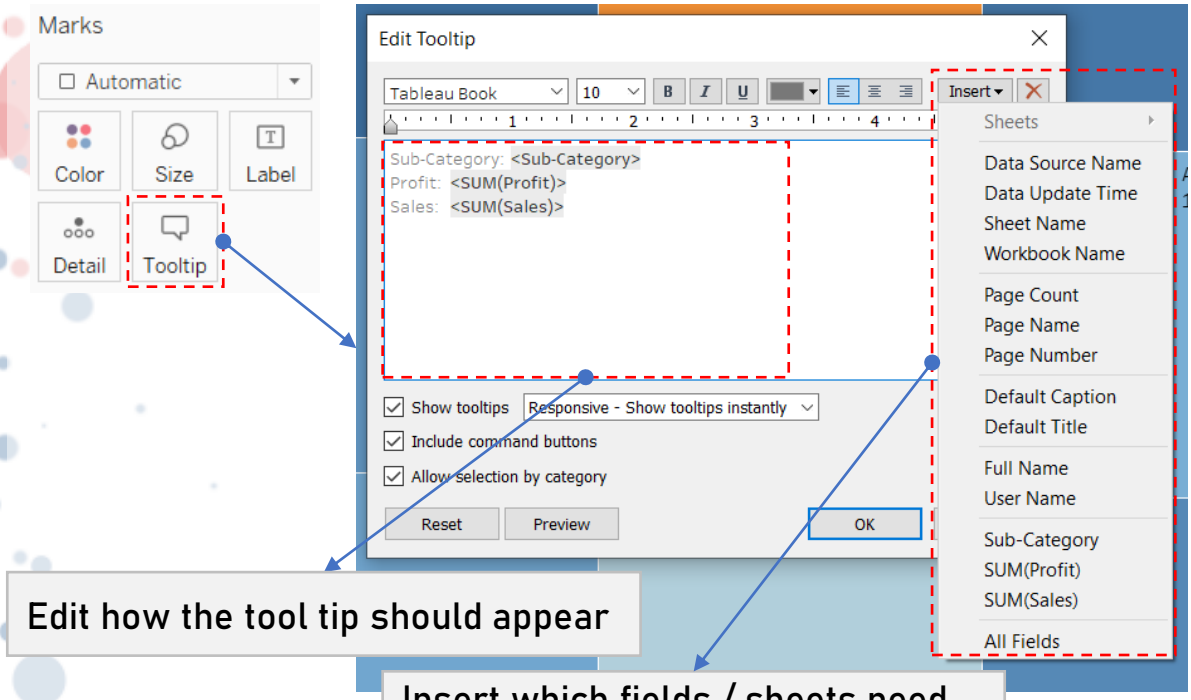
Dual axis charts, as the name goes, can only show two measures.

Synchronizing axes make the chart look like it was constructed using combined axes method.

If the value ranges of the measure fields used in the dual axis charts are too much apart, then it is best not to synchronize the axes.

Tool Tips

Tool tips help convey more information related to the chart interactively without taking up more space.

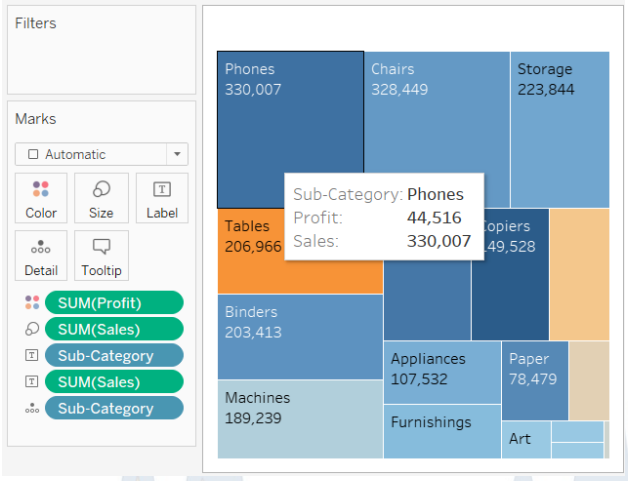


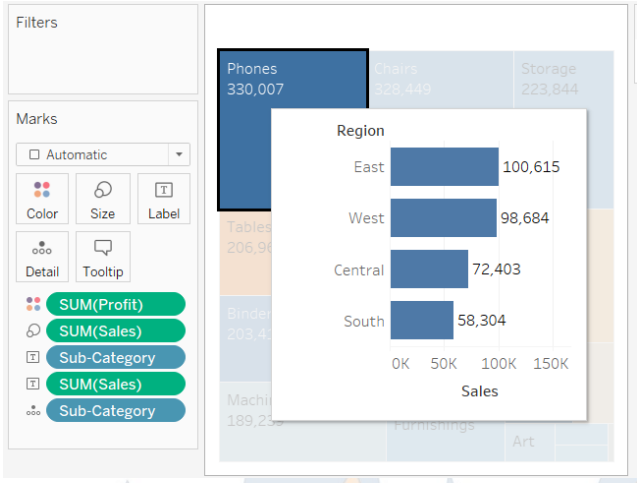
Edit how the tool tip should appear

Insert which fields / sheets need to be shown in the tool tip

Note: only the fields in the details are shown in the tool tip

Examples:



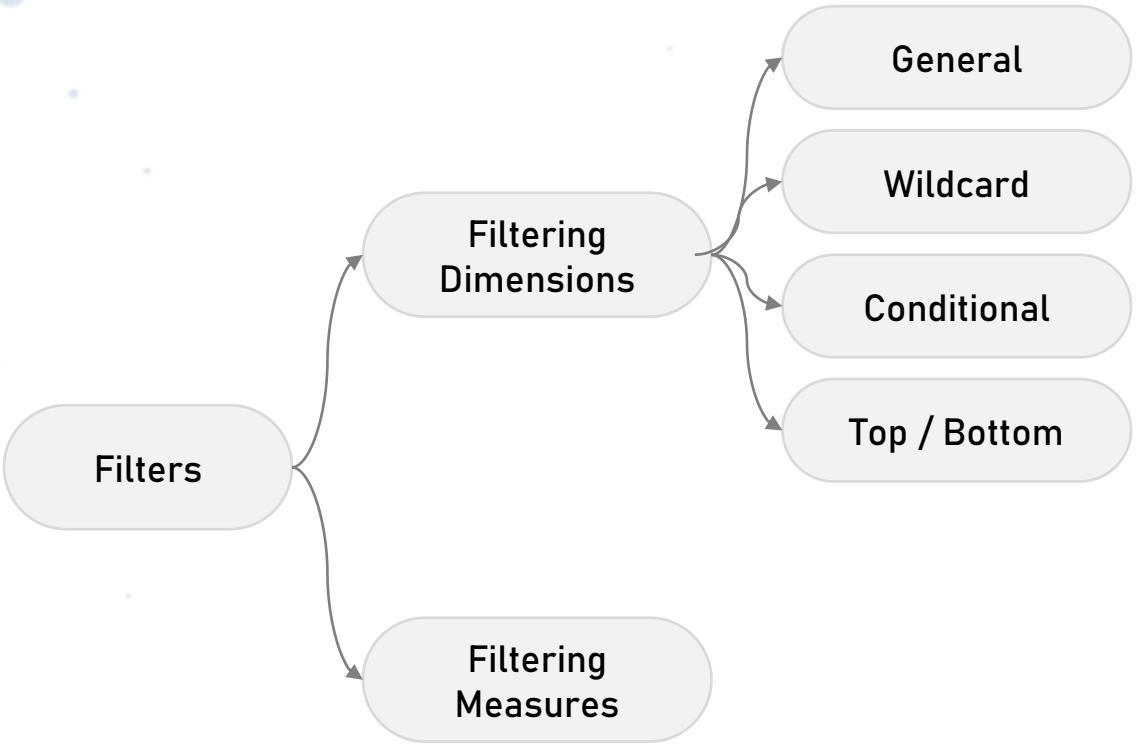


Filters

Filters restrict the data to be used in the chart depending on the criteria one applies. Both Measure and Dimension fields can be used in the filters.

For instance, sales trend can be shown only for the selected states using filters. Similarly all the transactions with worth $1 can be left out using filters.

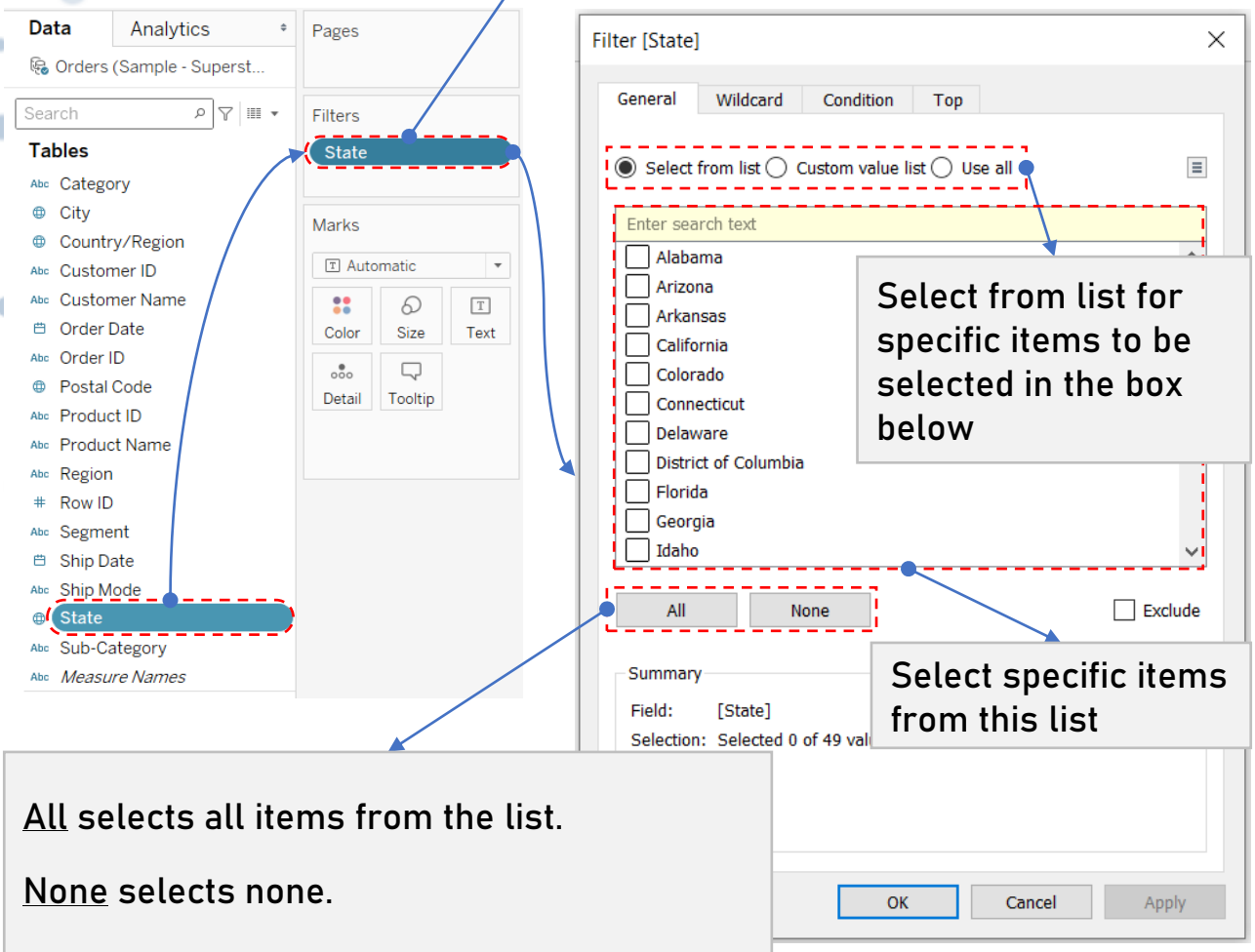
Types of Filters



General Filters

General filters as the name indicate can be used to filter any value from the dimension field. One could select the members of the dimension field from the list and apply filter. Also this list can be shown to the audience so that they can apply filter themselves.

Right click and show filter to show the members in the chart so that the user can control the filter



Select from list for specific items to be selected in the box below

Select specific items from this list

All selects all items from the list.

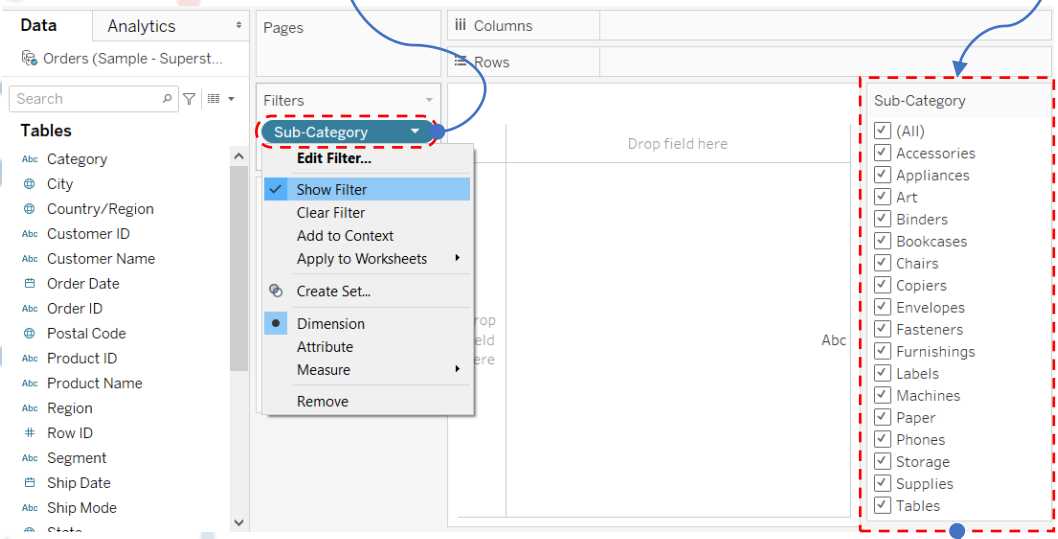
None selects none.

Note: if data gets updated later and if new members get added to the field, 'All' will not fetch the new members. But 'Use all' will.

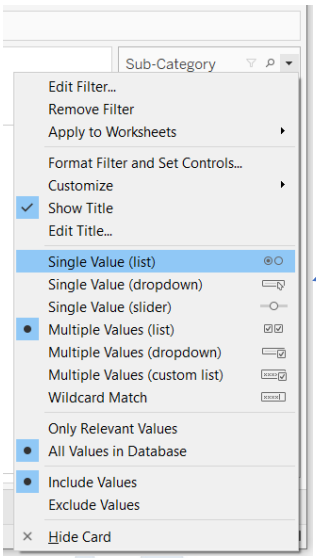
Allowing Audience to Interact with Filters

Filters are one of the ways to build interactivity in the charts. Once the filters are created, the items within the filter can be selected by the user depending on what they would like to analyze.

Right click on the field in the filter and click show filter to show filter control on the chart

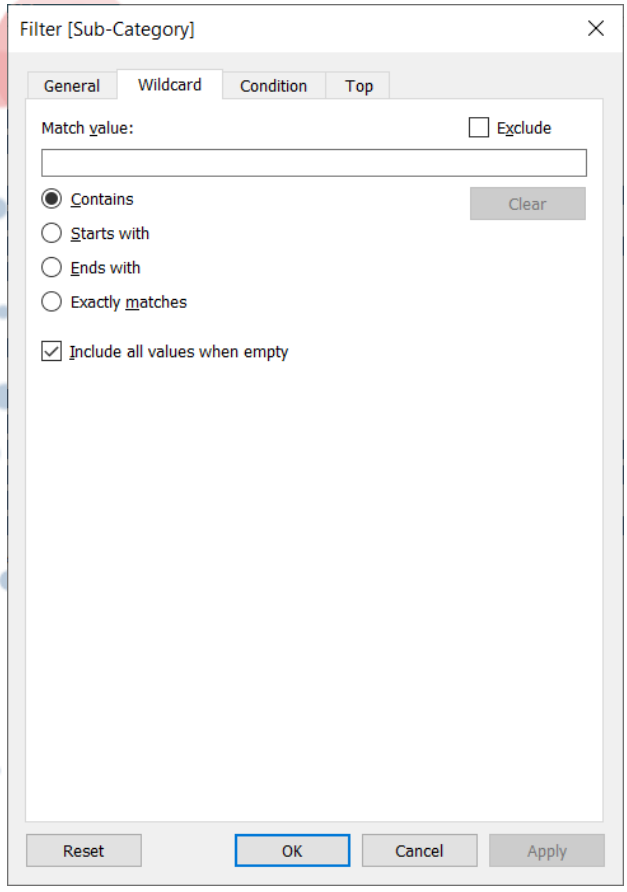


Click on the dropdown menu to control how filter needs to be shown e.g. list, dropdown etc.

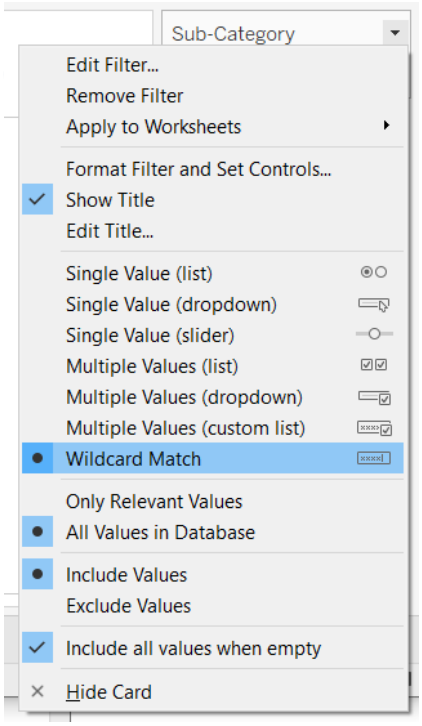


Wildcard Filter

Wildcard filters allow one to filter based on the text input and filters the items that contain, start with, end with or exactly matches the match value.

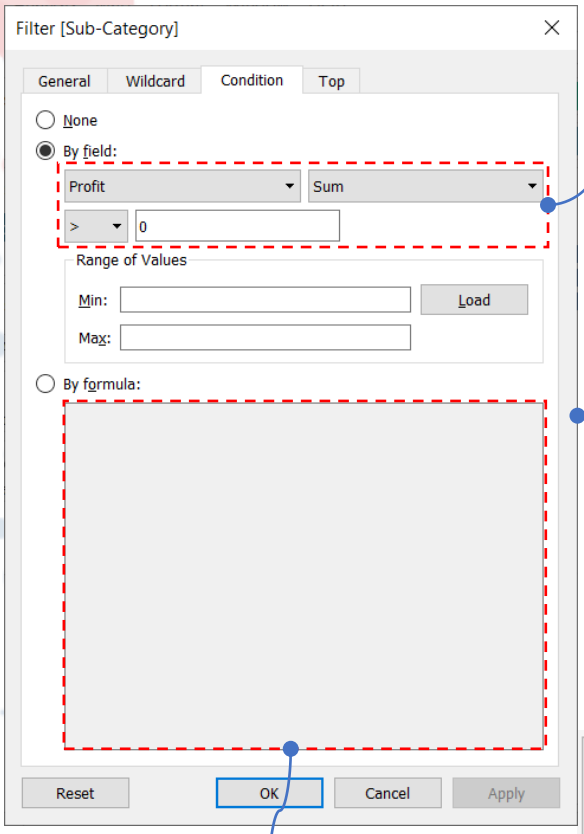


Use wildcard match from the show filter drop down menu for the wildcard filters



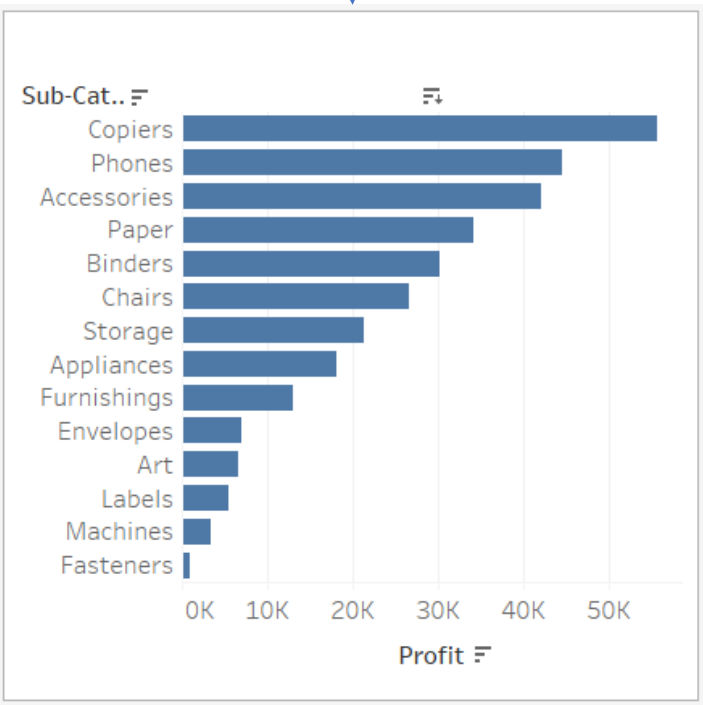
Conditional Filter

Instead of selecting the items from the filters directly, one could apply a condition to select the items. For instance, one could select all sub-categories that made a profit.

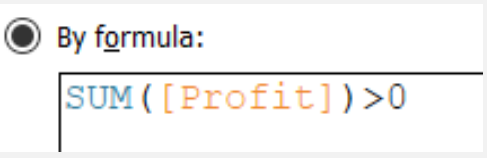


Build conditional filter using any field from the data source, selecting the appropriate aggregation method and specifying the condition

This filters and shows all sub-categories with profit > 0

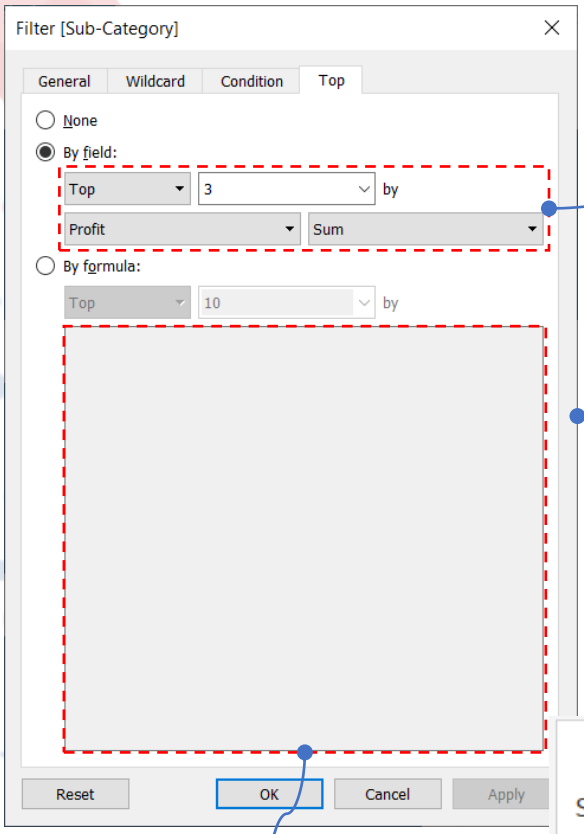


Not all conditions can be simply configured. In such cases formulas define conditions.



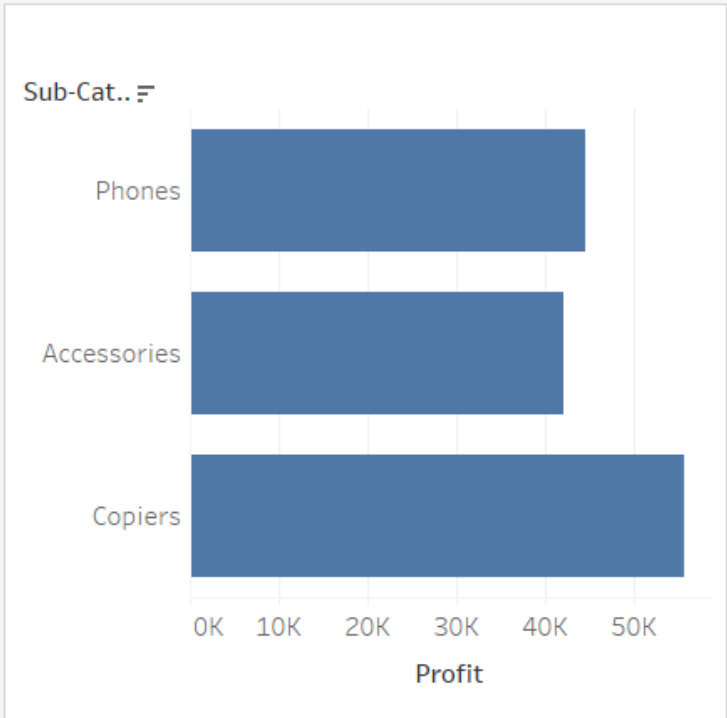
Top / Bottom Filter

Similar to conditional filters top / bottom filters allow one to select top or bottom 'n' number of items. For instance, in this case top 3 sub-categories in terms of total profits are shown.



Build top/bottom filter by specifying whether top or bottom items need to be shown, using any field from the data source, selecting the aggregation method and specifying the condition

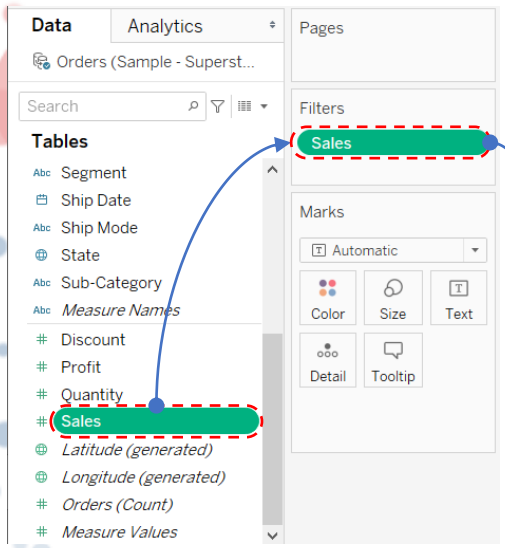
This filters and shows all sub-categories with profit > 0

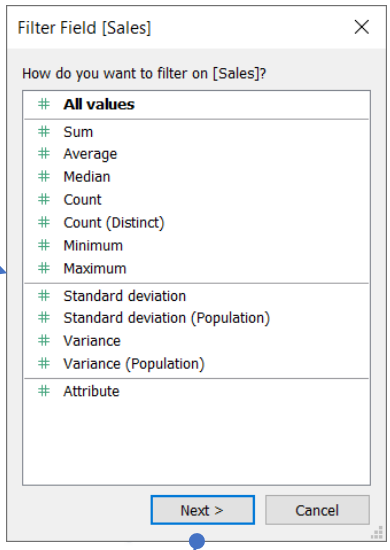


Similar to condition filters, formulas can be used in the top / bottom filters as well

Filtering Measures

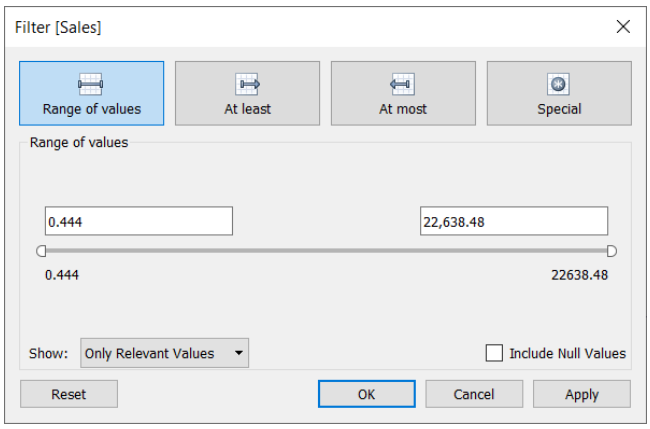
Similar to filtering the members of the dimension field, numerical values can be filtered using the values.

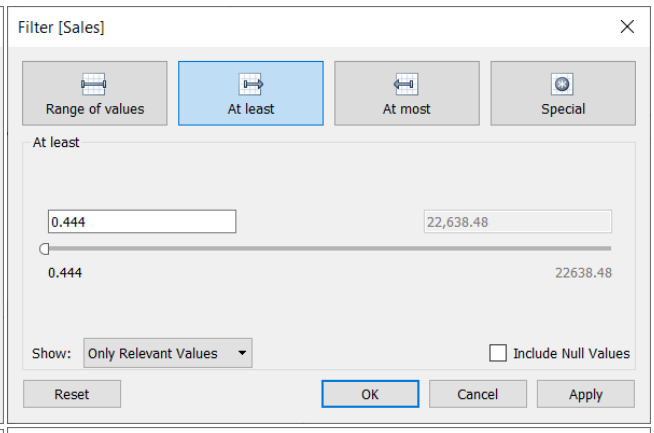


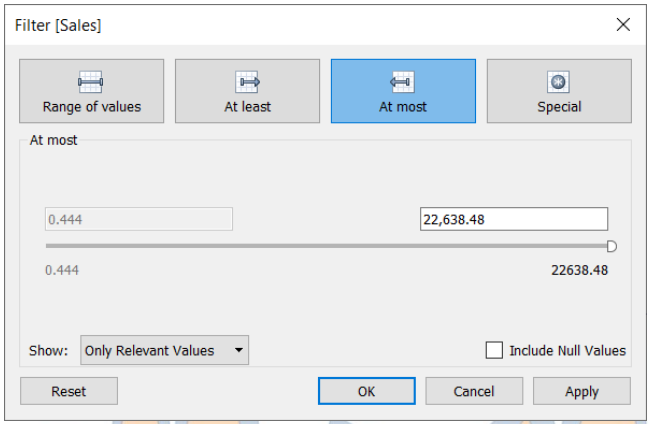


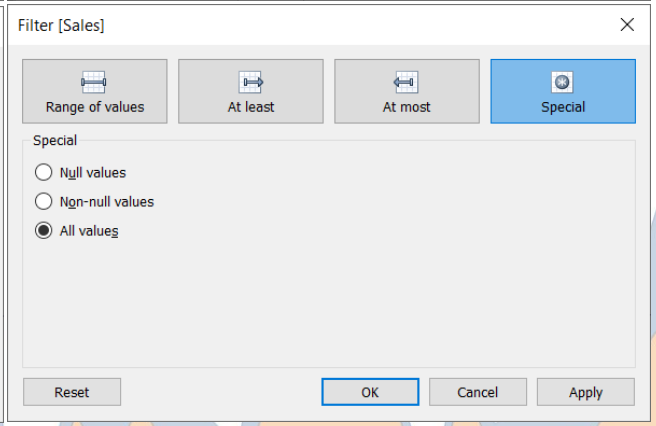
Select the aggregation type or select 'all values' is no aggregation is required

Select the type of the filter and specify the values to filter







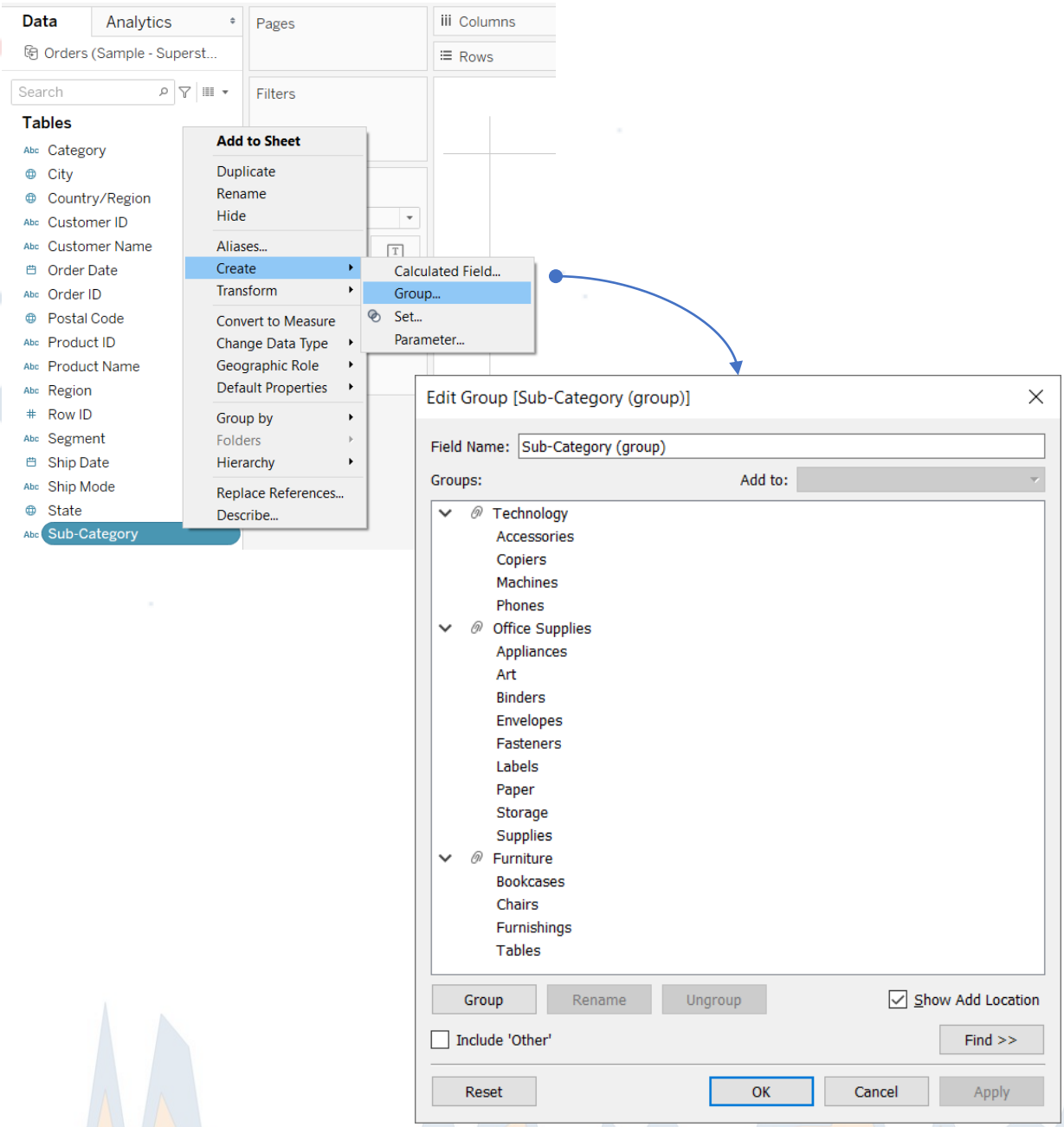


Groups

Groups allow one to group the members of the dimension field into a logical group and create a dimension at a higher level in the hierarchy.

In this case members of the sub-category field can be grouped together to create field same as the category field.

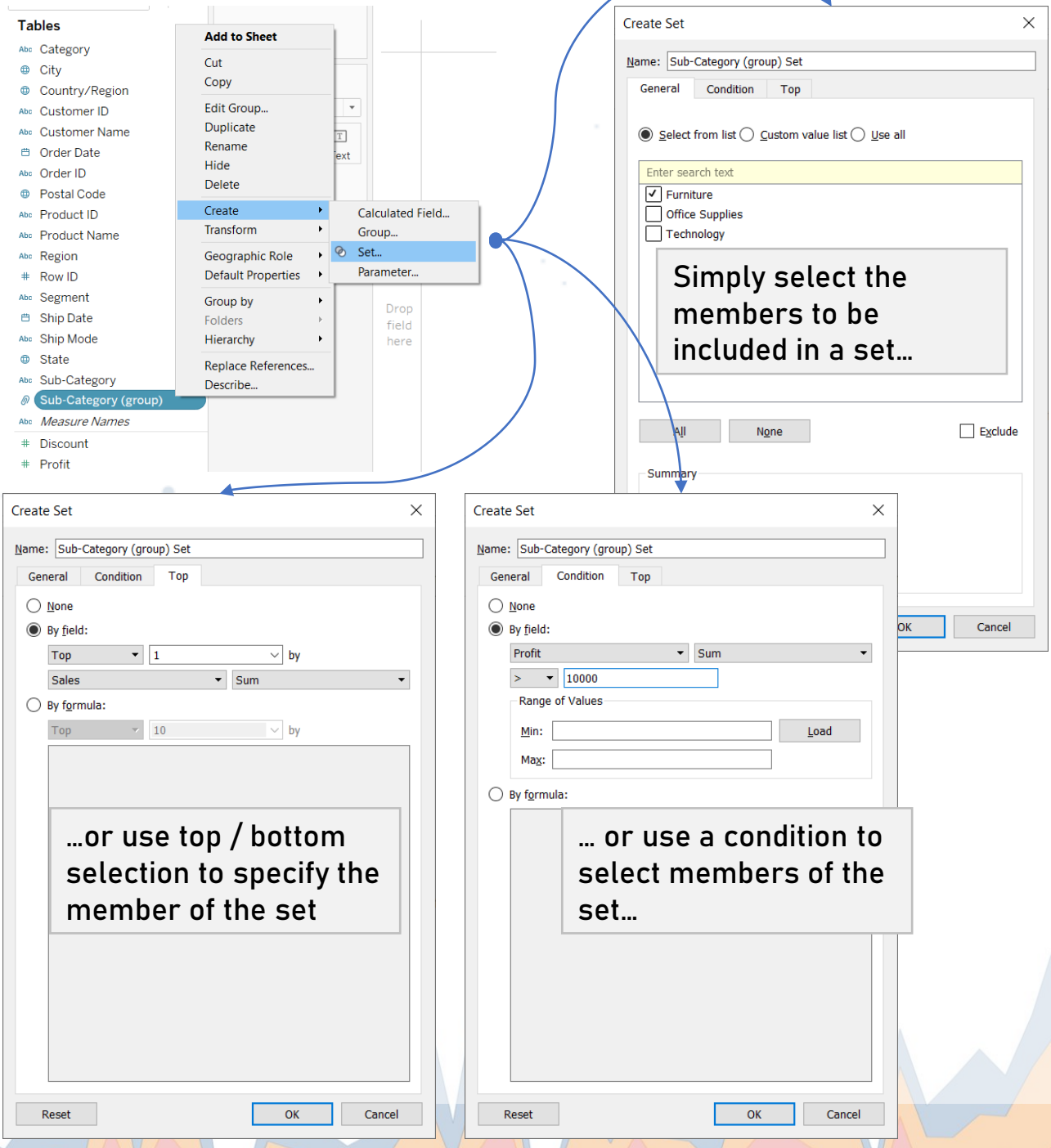
Grouped field is a new field and does not alter the original field



Sets

Sets are custom fields that create a subset of a dimension field depending on the given criteria similar to how a filter is created. Sets can be used in filters or colors to show only the set or color the sets differently.

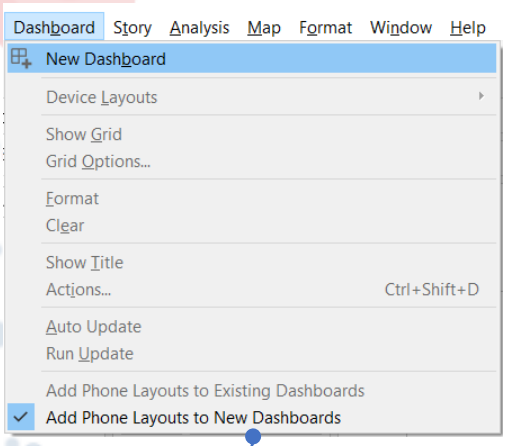
Even though Sets are similar to Groups by creating subset of the dimension field, they are different: Sets are binary and dynamic (gets updated with new data) while groups aren't.



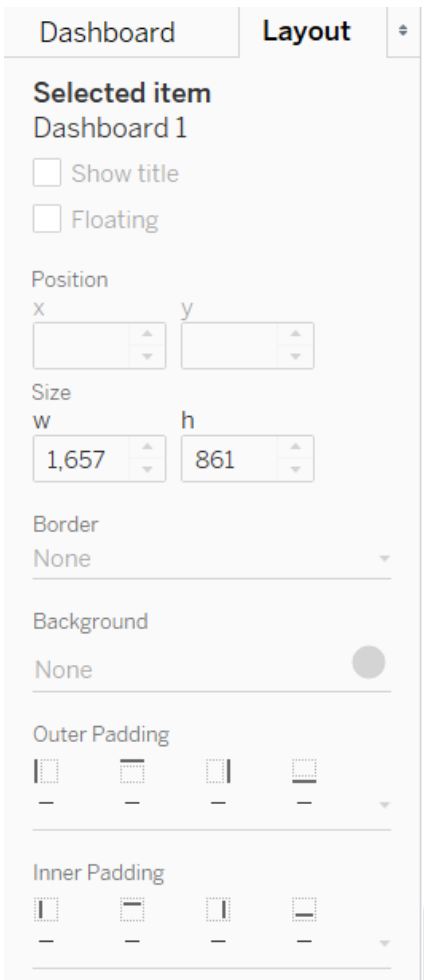
Dashboard Layout

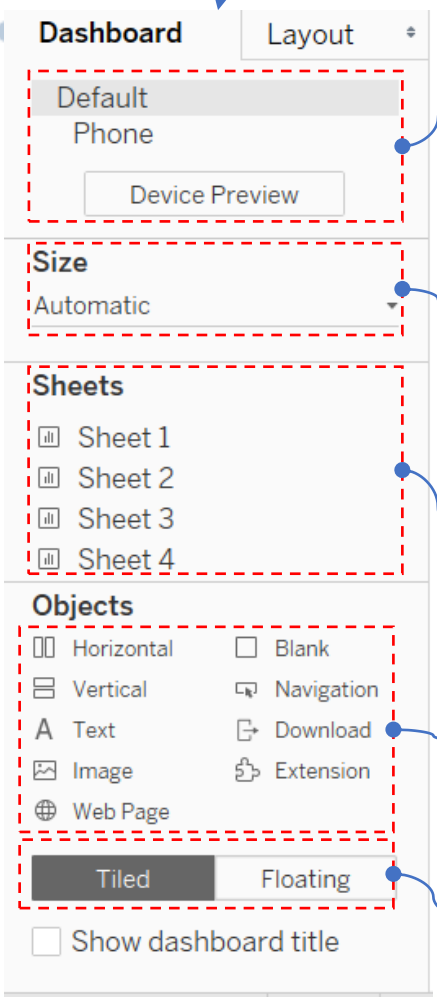
Dashboards are collection of charts to convey information from multiple charts simultaneously.

Create dashboard from the menu or from the 'create dashboard' icon from the bottom of the screen.



Format the layout of the dashboard from the layout pane





Preview dashboard in multiple device screens

Specify the size of the dashboard canvas. Automatic better with most screens

Worksheets can be dragged from here and dropped in the dashboard canvas

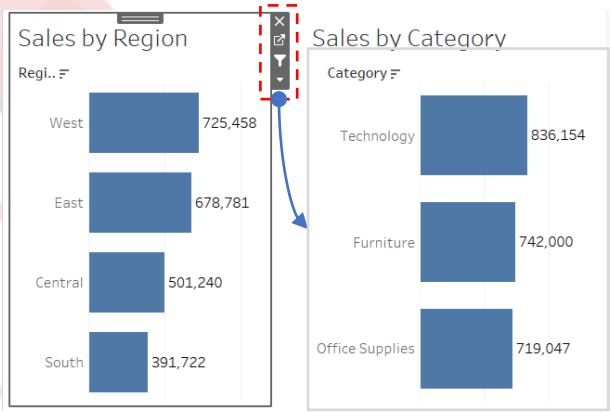
Dashboard objects / containers for the charts and other elements

Either align elements as tiles or allow them to float

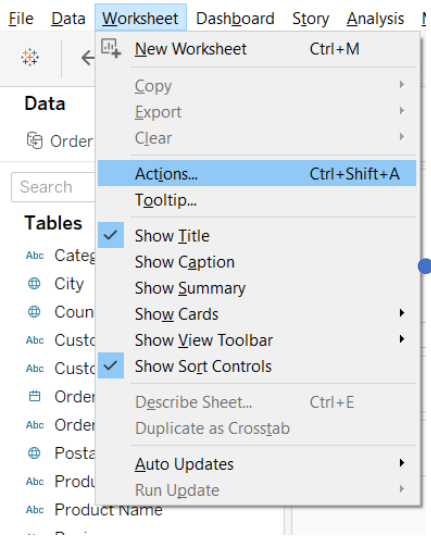
Action Filters

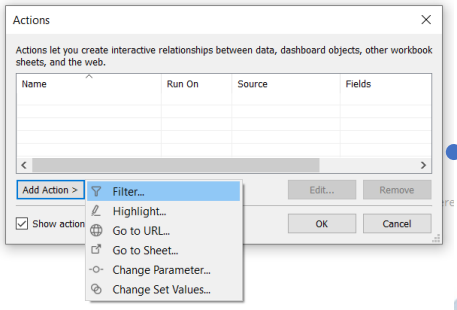
Action filters allow one to use a chart as a filter.

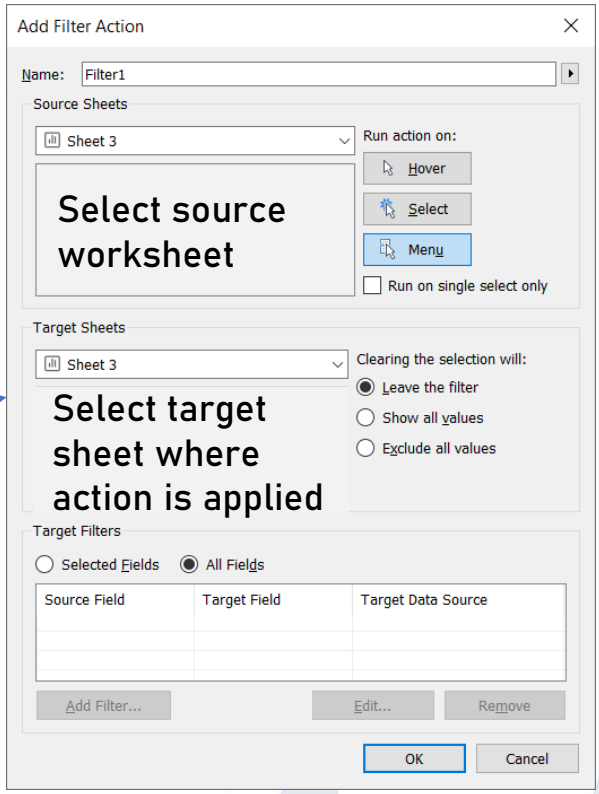
Quick action filter can be enabled from a dashboard by selecting the 'funnel' icon



Action filters can also be created from scratch using the 'Actions' from under the worksheet menu

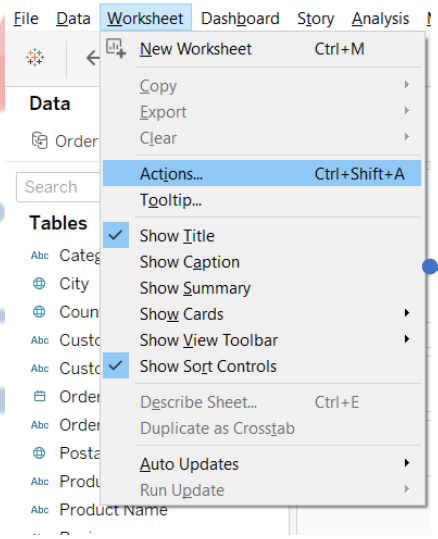


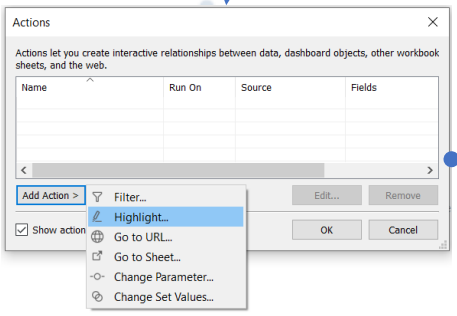


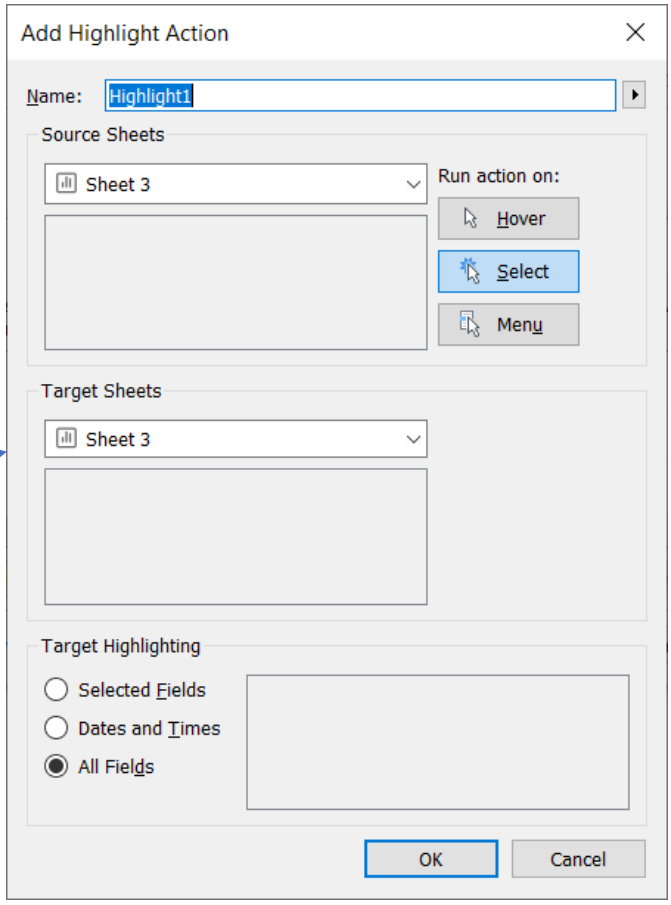


Highlight Actions

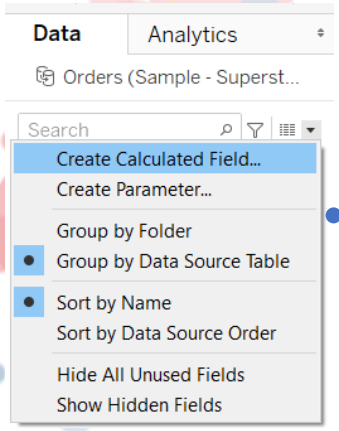
Similar to filter actions, highlight actions allow mark to be selected from a source sheet and the same is highlighted in the target sheet





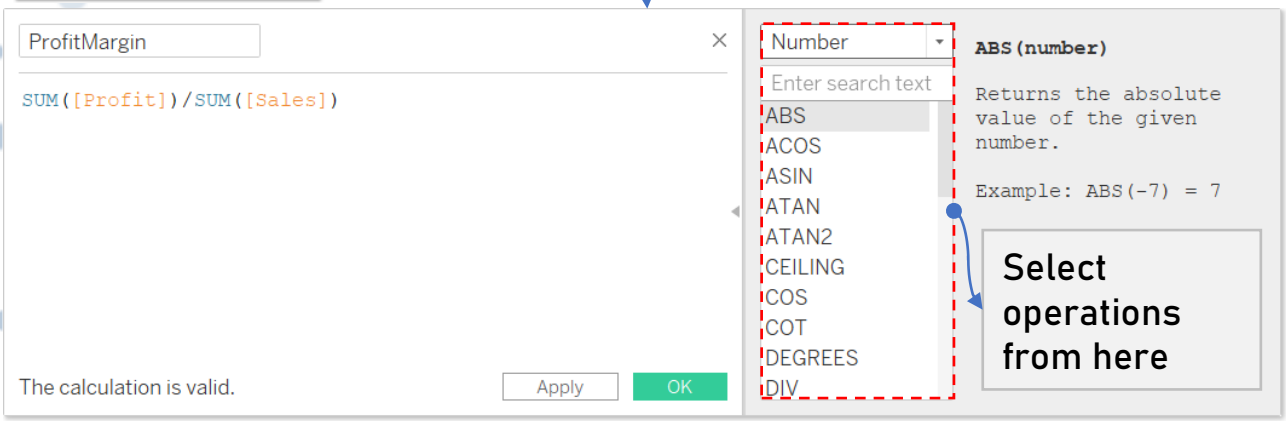


Number Calculations



In a lot of instances, one needs to create custom calculations. For instance, one may want to calculate the profit from the sales and costs or profit margin from the profits and sales etc. Calculated fields are useful for doing so.

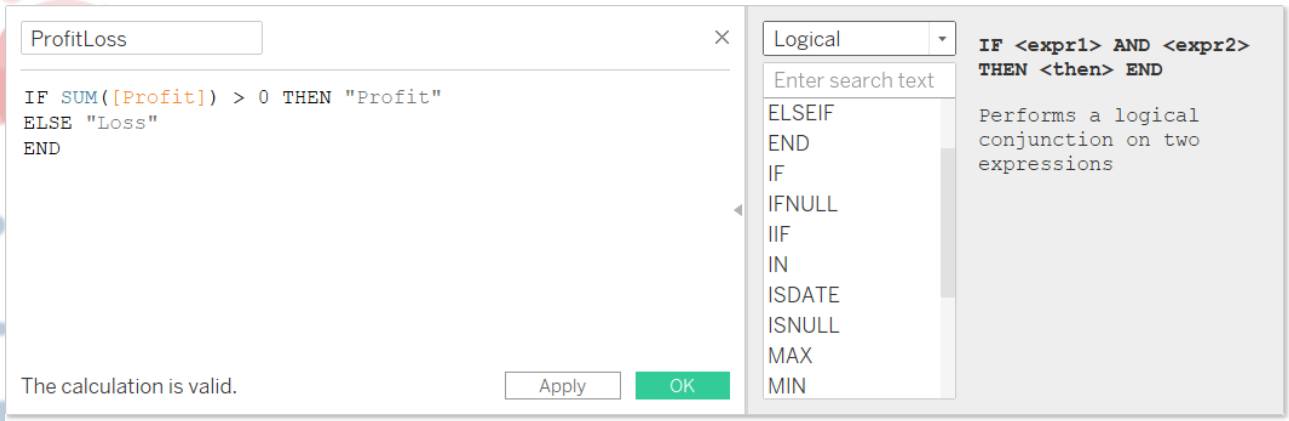
Number calculations are commonly used and they return numerical outputs. Several off-the-shelf functions are available in Tableau.



Common Number Calculations	What it does?
SUM, AVG, MEDIAN, COUNT	Aggregation calculations
Round, ceiling, floor	Rounds off to the digits specified, or to the higher integer, or lower integer respectively
SQRT	Returns square root of a given number
ZN	Returns expression if not NULL else zero
MIN, MAX	Returns minimum or maximum values
ABS	Returns absolute value
LOG, LN, EXP	Returns logarithm for a given base or natural logarithm or exponent of a given number

Logical Functions

Logical calculations are helpful in determining a given condition is true or not. For instance, one can determine if a given state made profit or loss by looking at whether sum of profit was greater or less than 0.



Common logical functions	What it does?
IF-THEN-ELSEIF-ELSE-END	IF-THEN-ELSE and ELSEIF allow one to construct a logical statement and test a series of expression. When the expression is true, returns what is after the THEN
CASE-WHEN-THEN-END	Works similar to IF-THEN statements but used for simple-listed conditions
IIF	A simpler version of IF-THEN-ELSE constructed in one line for simple conditions
AND, OR, NOT	Logical conjunction, disjunction and negation on two logical expressions

Common Calculations

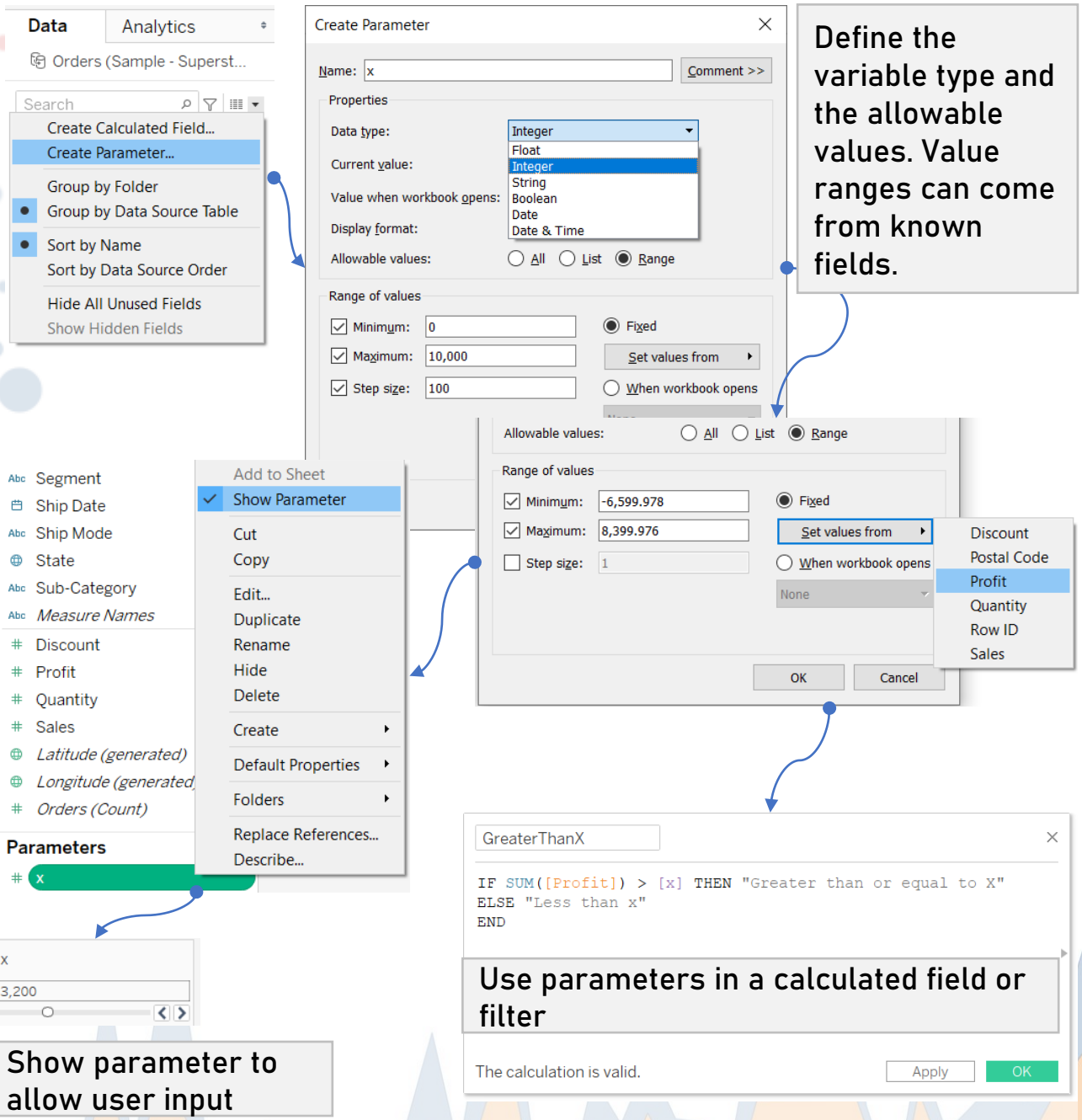
Apart from the numeric or logical calculations, there are many types of calculations in Tableau such as string, date, type conversion etc. Here are some of the commonly used calculations.

Operation	What it does?
LEFT, RIGHT, MID	Returns specified number of characters from the start, middle or from the end of the string
LOWER, UPPER	Converts the string to lower or upper case
SPLIT	Splits the string based on the specified delimiter
CONTAINS	Returns TRUE if string contains the specified substring
LEN	Returns the number of characters in the string
STARTSWITH, ENDSWITH	Returns TRUE if string starts or ends with the specified substring
LTRIM, TRIM, RTRIM	Returns the string with the spaces removed from the beginning, middle or the end
ISDATE	Returns TRUE if the given string is a valid date
DAY, WEEK, YEAR	Returns the day, week, or the year of the given date as an integer
DATEPART	Returns the part of the date (e.g. year, month, week etc.) as an integer
DATEADD	Adds the specified number of days to a given date and returns the incremented date

Parameters

Parameter is variable created within the workbook that can act as a user-input value in a calculation or filter.

For instance, let's say that we want to show the items that have made profits greater than lets say x and x can be adjusted depending on the user input. In such case, a parameter is created to get the value for x, and then x can be used in the calculated field.



Quick Table Calculations

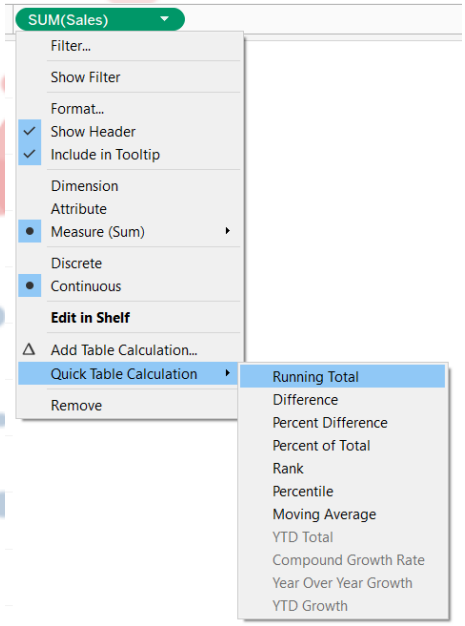


Table calculations are a special type of calculation that you can apply to the specifically to what is in the view.

These calculations will be not be seen in the data pane as a separate calculated field. You can consider them as a kind of wrapper on top of the existing field shown in the visualization.

Since table calculations are applied only on the view, any data point that has been filtered out is not considered.

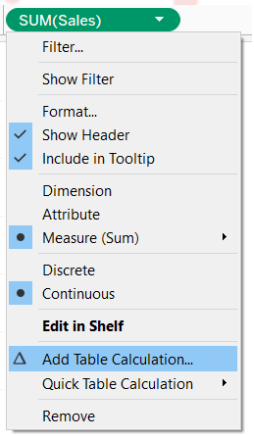
A common example is % of total.

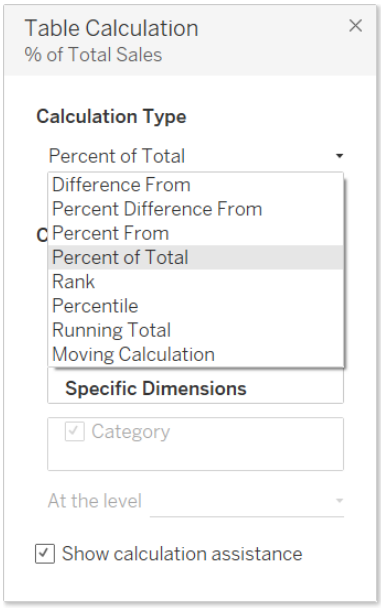
Quick Table Calculation	What it does?
Running Total	Calculates the running total or the sum of all values until each mark
Difference	Calculates the difference between two marks
Percent Difference	Calculates the difference between two marks in percentage
Percent of Total	Shows the percent of each value compared to the sum of all values in the view
Rank	Returns the rank compared to all values
Percentile	Shows the percentile of each value compared to the sum of all values in the view
Moving Average	Returns the moving average for specified number of marks

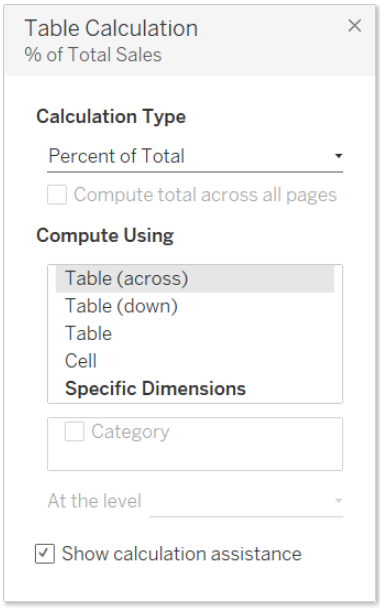
(Slightly) Advanced Table Calculations

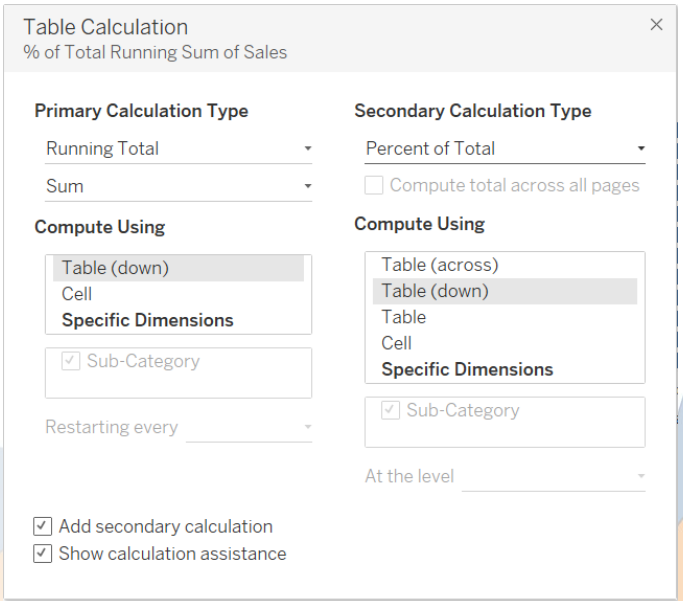
There are more ways to do table calculations. From the table calculation window, one could define

- What table calculation is to be applied (e.g. percent of total, difference, rank etc.)
- How to apply the table calculation (e.g. across the table or down) and
- What secondary calculation to be used in addition etc.









Level of Detail Expressions

Level of Detail Expressions also known as LOD expressions are some of the advanced calculations in tableau. Calculated fields generally make computations at the visualization level, while the LOD expressions are useful in making computation at both data source and visualization levels.

If one needs to compute a value that will not change depending on the visualization it is hard to do so with normal calculation.

For instance, if we need to calculate the overall profit margin using the following expression `SUM([Profit])/SUM([Sales])` and use it in any other computations using a normal calculation then what happens is that then this computation is done based on the dimensions used, filters applied to the worksheet and not based on all values from the data source.

However, if we use the following LOD expression instead, `{ FIXED [Row ID]: SUM([Profit])/SUM([Sales])}` then the profit margin is computed using all rows from the data source.

Note that instead of ROW ID we could use any dimension in this case for aggregating at a particular level of detail.

LOD Expression	What it does?
INCLUDE	Computes at the data source at the level of detail specified and also uses the dimensions used in the worksheet `{ INCLUDE [State]: SUM([Profit])/SUM([Sales])}`
EXCLUDE	Computes at the data source at the level of detail specified but excludes uses the dimensions used in the worksheet `{ EXCLUDE [City]: SUM([Profit])/SUM([Sales])}`
FIXED	Uses the specified dimension as the level of detail irrespective of what is in the worksheet `{ FIXED [Row ID]: SUM([Profit])/SUM([Sales])}`

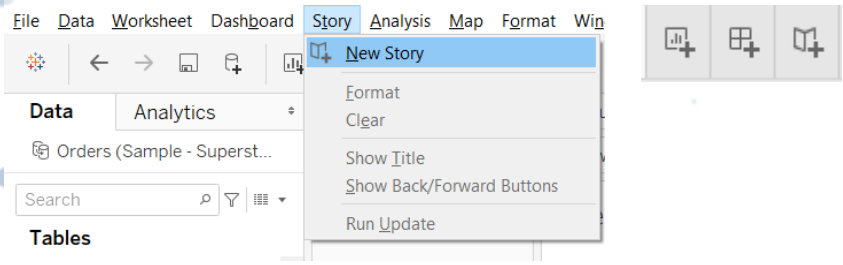
Stories

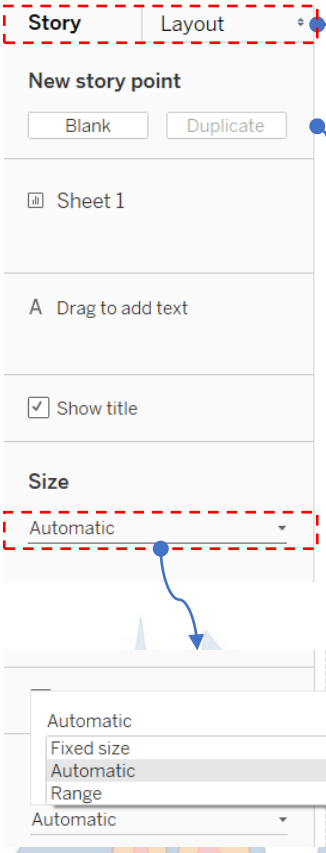
Story is a collection of sheets and dashboards, arranged in a sequence making a compelling case using data.

Constructing a story is similar to constructing a dashboard. So the methods used to create, name, and manage worksheets and dashboards also apply to stories.

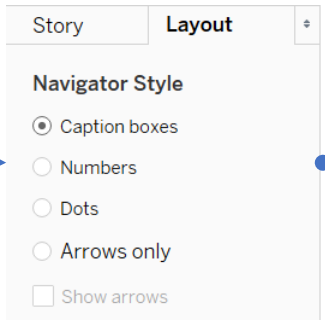
However, story conveys information with a scope larger than individual worksheets or dashboards in order to make a compelling case. Each individual sheet in a story is called a story point.

Story can be created from menu (Story → New Story) or from the set of icon on the bottom of the screen





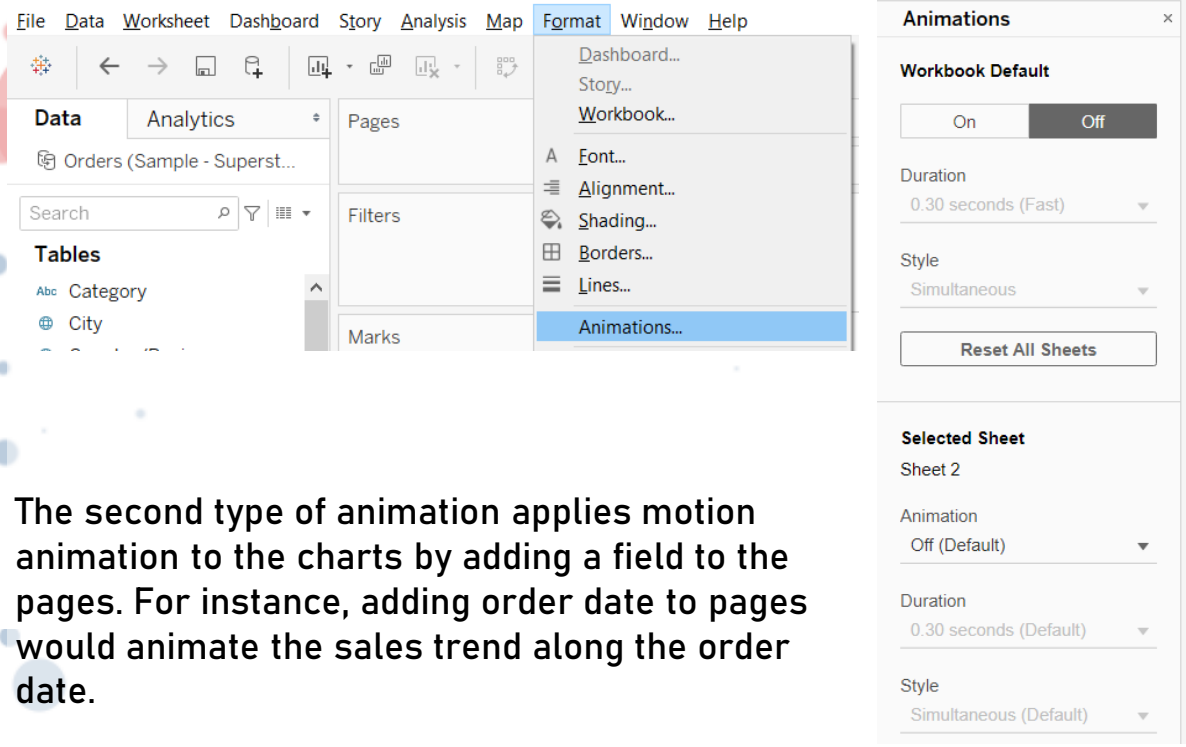
Story pane allows creating a new story point, include sheets/dashboard to the story point and changing the size of the story



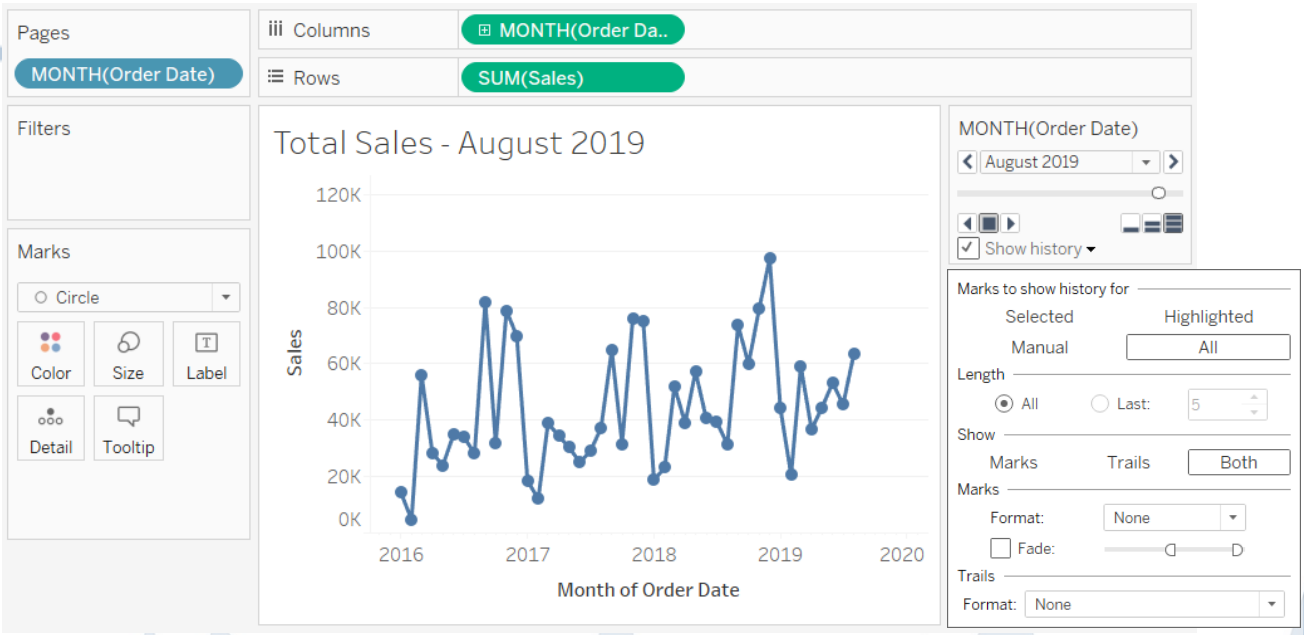
Adjust the navigation style from the from the layout tab

Two Types of Animations

By switching ON Animations in the sheets or the entire workbook the transition due to filters or parameters would become smooth.

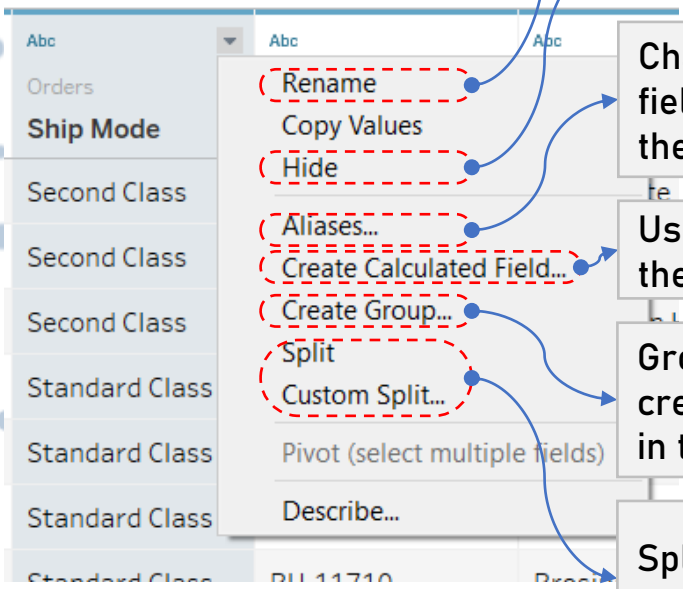


The second type of animation applies motion animation to the charts by adding a field to the pages. For instance, adding order date to pages would animate the sales trend along the order date.



Data Clean-Up Steps

Most often raw data is unusable for analysis due to multiple reasons. Several steps need to be applied to clean the data before any analysis.



Rename the field name

Hide the field from the data table

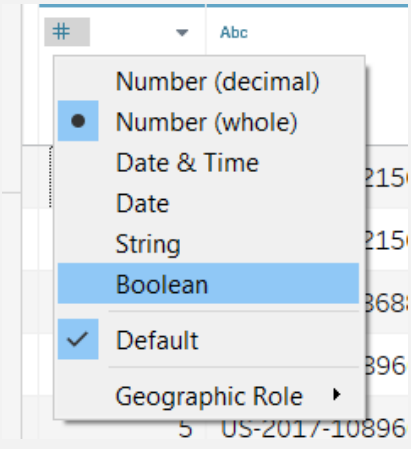
Change how members of a dimension field (a.k.a. aliases) should appear in the worksheets

Use the field in a calculation right at the stage of data preparation

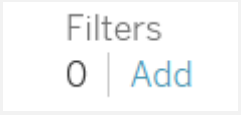
Group members of the dimension and create a dimension at a higher level in the hierarchy

Split a string field into components using a specified delimiter

Click on the field type icon to change the field type



Filter members of a dimension field or specific values in a measure fields using filters

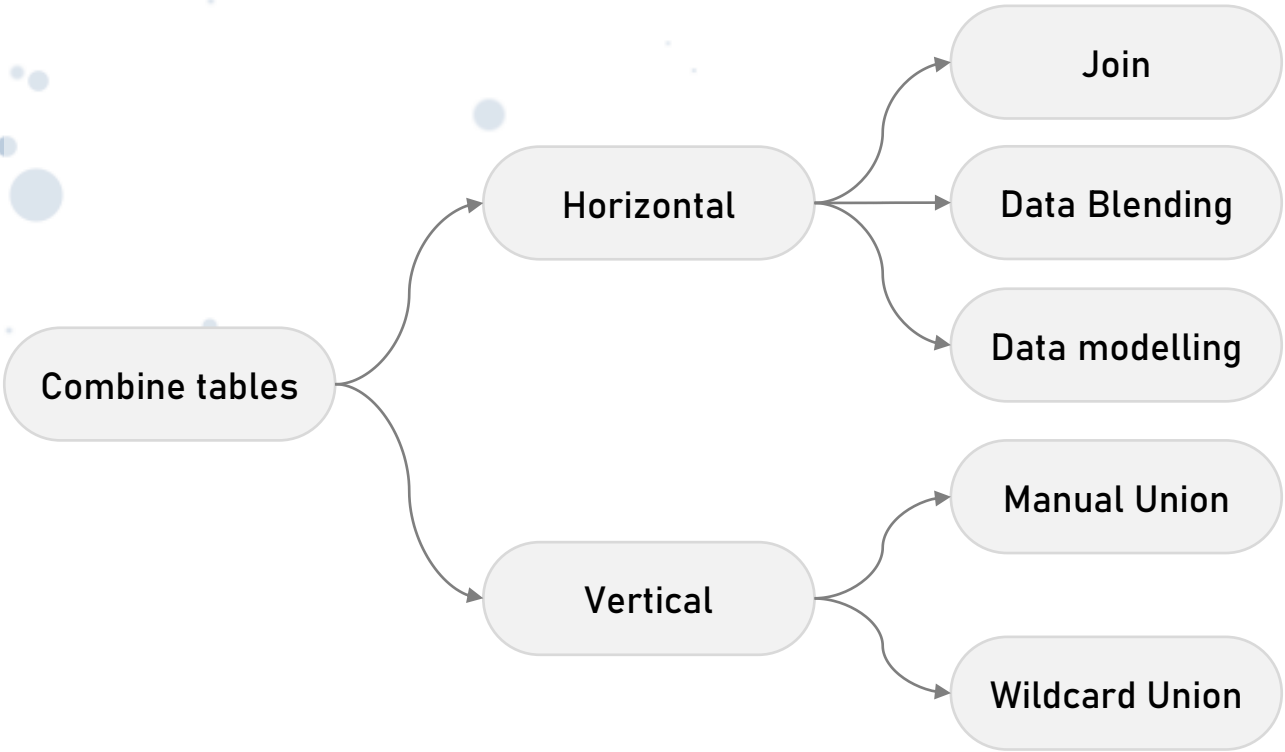


Combining Tables

Data often come from multiple sources in different shapes and forms and needs to be combined into a single data table for analysis. Several data combining techniques exist.

When the situation is to add new values to an existing dataset and when the fields are identical, we combine the data sources vertically, meaning we stack them up using unions.

When we have multiple data sources each having different set of columns then we combine then horizontally using join, blending or data modelling. For this we need at least one column to be present across all the data sources being merged.



Unions

Unions are used to combine new data to the exiting / old tables for the analysis or when data is reported by each segment.

Order ID	Sales ($)
001	15.00
002	11.00
003	10.00

+

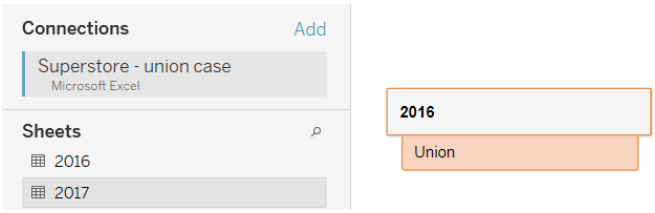
Order ID	Sales ($)
004	10.00
005	11.00
006	12.00

=

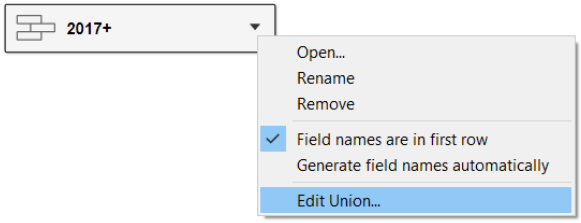
Order ID	Sales ($)
001	15.00
002	11.00
003	10.00
004	10.00
005	11.00
006	12.00

For instance, when each business unit report data in a separate table, union combines all of these data tables into one data table. The primary condition for union is that the structure of the data source must be same

In order to 'union' two data tables bring the second table right below the first data source in the tables.

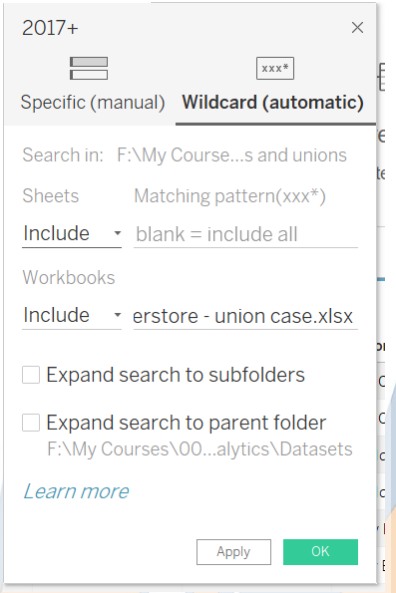


Alternatively, click all sheets from the connection and bring them to data source (right)



Right click on the union to edit union.

Wildcard or automatic union allows to combine all data sources from a specific location with specific names automatically.



Joins

Join is used to combine multiple data tables horizontally using a common field to match rows across multiple tables. Four types of joins are possible: left, right, inner, outer.

In the following example the first two tables are combined using the field 'Emp #'. Ideally when all rows are present across all tables the combined table using any of the four methods has no missing values.

Name	Emp #	Age
Adam	001	32
Ben	002	41
Chelsy	003	45

Emp #	Salary
001	60k
002	45k
003	62k

Name	Emp #	Age	Salary
Adam	001	32	60k
Ben	002	41	45k
Chelsy	003	45	62k

However, when few rows are missing in any of the input tables the output is table can have missing values depending on the type of join. In the following case, a left join retains all values from the table on the left while showing NA values for the matching rows for the fields that come from the right table.

Total number of rows in the resulting table would be same as the left table.

Name	Emp #	Age
Adam	001	32
Ben	002	41
Chelsy	003	45

Emp #	Salary
001	60k
003	62k

Name	Emp #	Age	Salary
Adam	001	32	60k
Ben	002	41	-
Chelsy	003	45	62k

Alternatively, when tables are combined using a right join, all rows from the right side of the table are retained, with NA values for the fields in the left table that have no values.

Total number of rows in the resulting table would be same as the right table.

Name	Emp #	Age
Adam	001	32
Ben	002	41

Emp #	Salary
001	60k
003	62k

Name	Emp #	Age	Salary
Adam	001	32	60k
-	003	-	62k

Joins

Inner join retains only the perfectly matching rows. Inner join would result in a table with the least number of rows.

Name	Emp #	Age
Adam	001	32
Ben	002	41

Emp #	Salary
001	60k
003	62k

Name	Emp #	Age	Salary
Adam	001	32	60k

While outer join combines all rows irrespective of missing values and the outer join would result in a table with the most number of rows.

Name	Emp #	Age
Adam	001	32
Ben	002	41

Emp #	Salary
001	60k
003	62k

Name	Emp #	Age	Salary
Adam	001	32	60k
Ben	002	41	-
-	003	-	62k

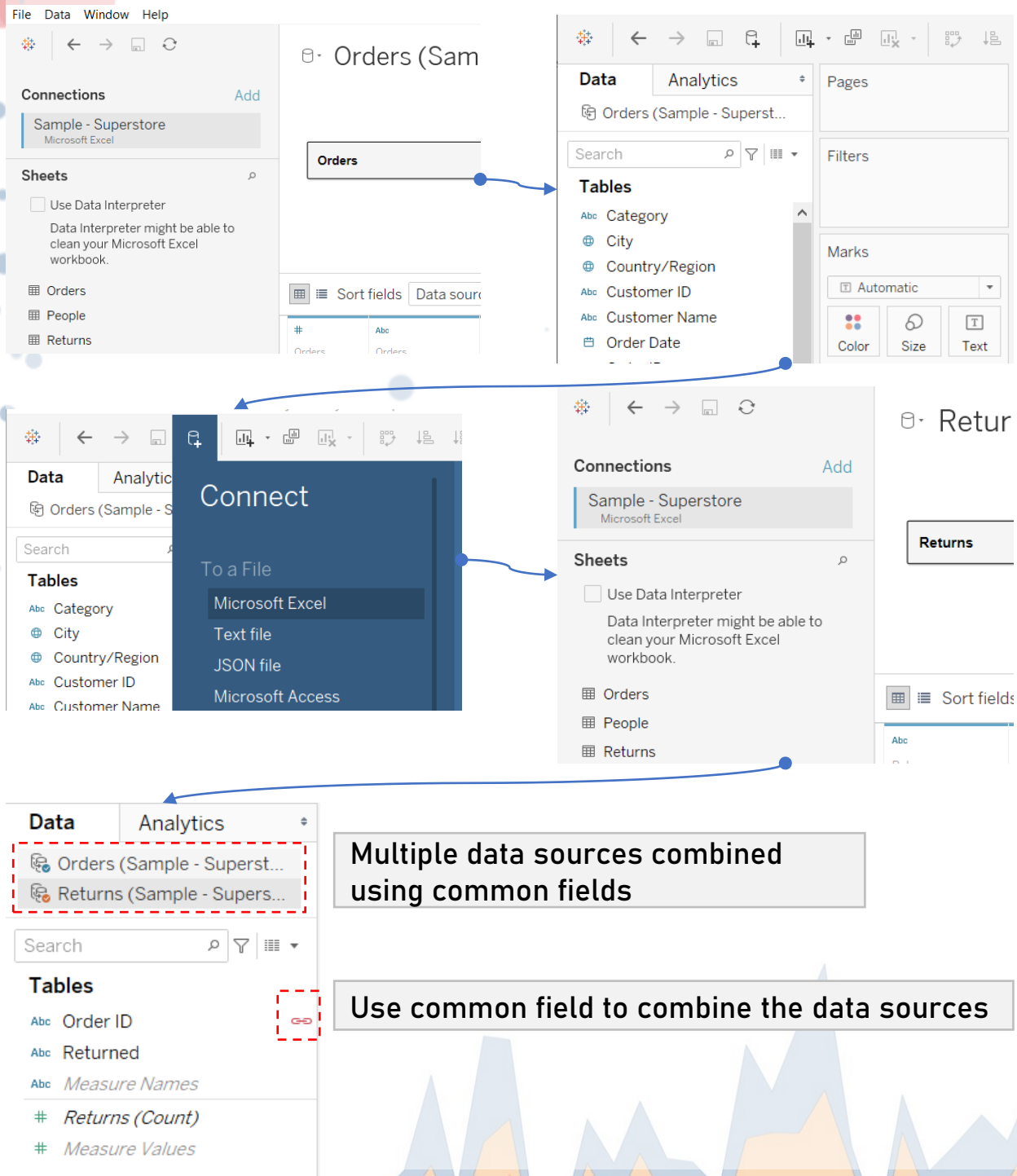
Note:

Joining two tables creates a combined physical table, which means that by combining two input tables, a third table is created.

This also means that the combined table needs to exist separately in the memory making join a memory intensive technique.

Data Blending

Unlike Join, data blending does not create a new table but queries the input table based on the fields used in the visualization. The combination is not physical but logical. The combination is done on the aggregated tables and not necessarily on the actual tables if only aggregated fields are required.



Data Relationships and Data Modeling

Data modeling is a new, simple and intuitive technique introduced in Tableau 2020 which is similar to data blending in terms of how tables are combined but similar to joins in terms of how constructed.