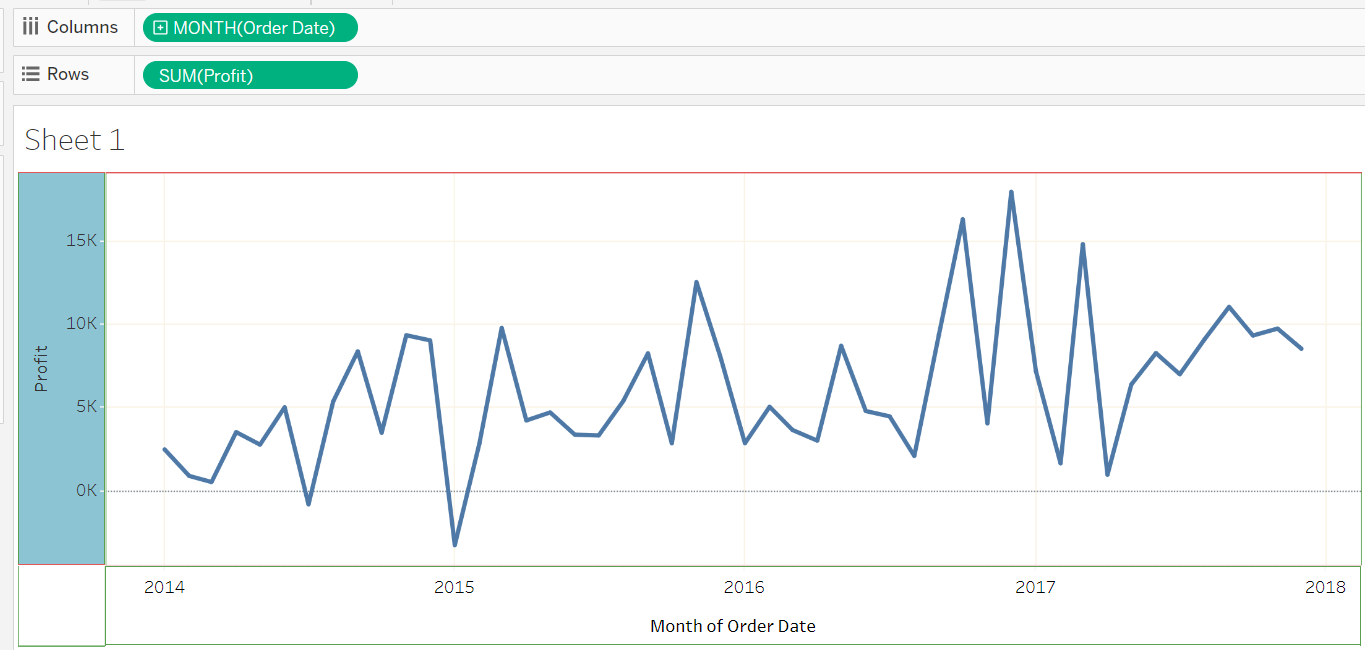
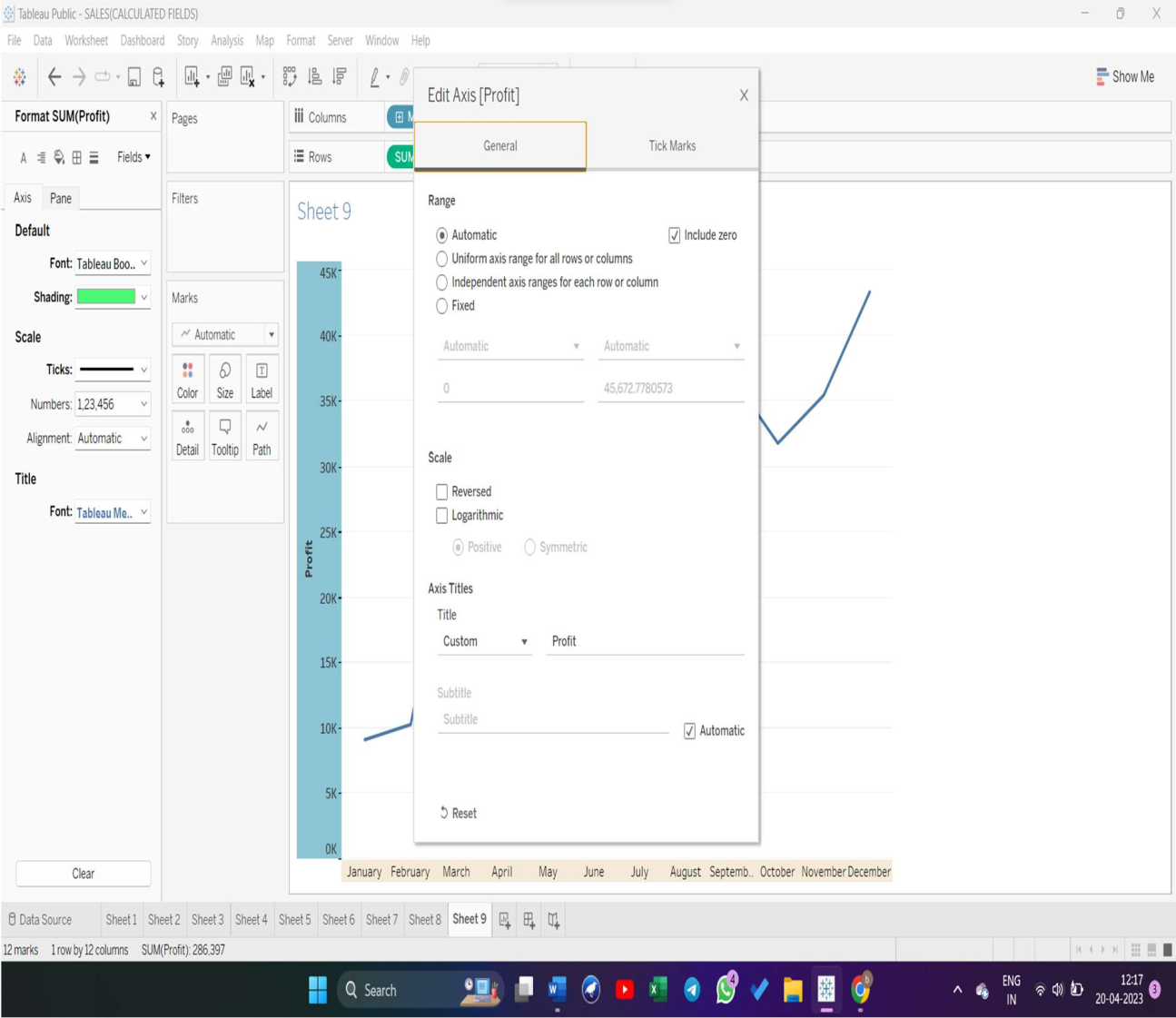
WEEK-5

EDITING AND FORMATTING AXES

When you add a continuous ﬁeld (a ﬁeld with a green background) to the Columns or Rows shelf, it adds an axis to the view. An axis shows data points that lie within a range of values. For each axis, you can specify the range, scale, and tick mark properties.



To edit an axis range, double-click the axis that you want to edit.



DATA MANIPULATION

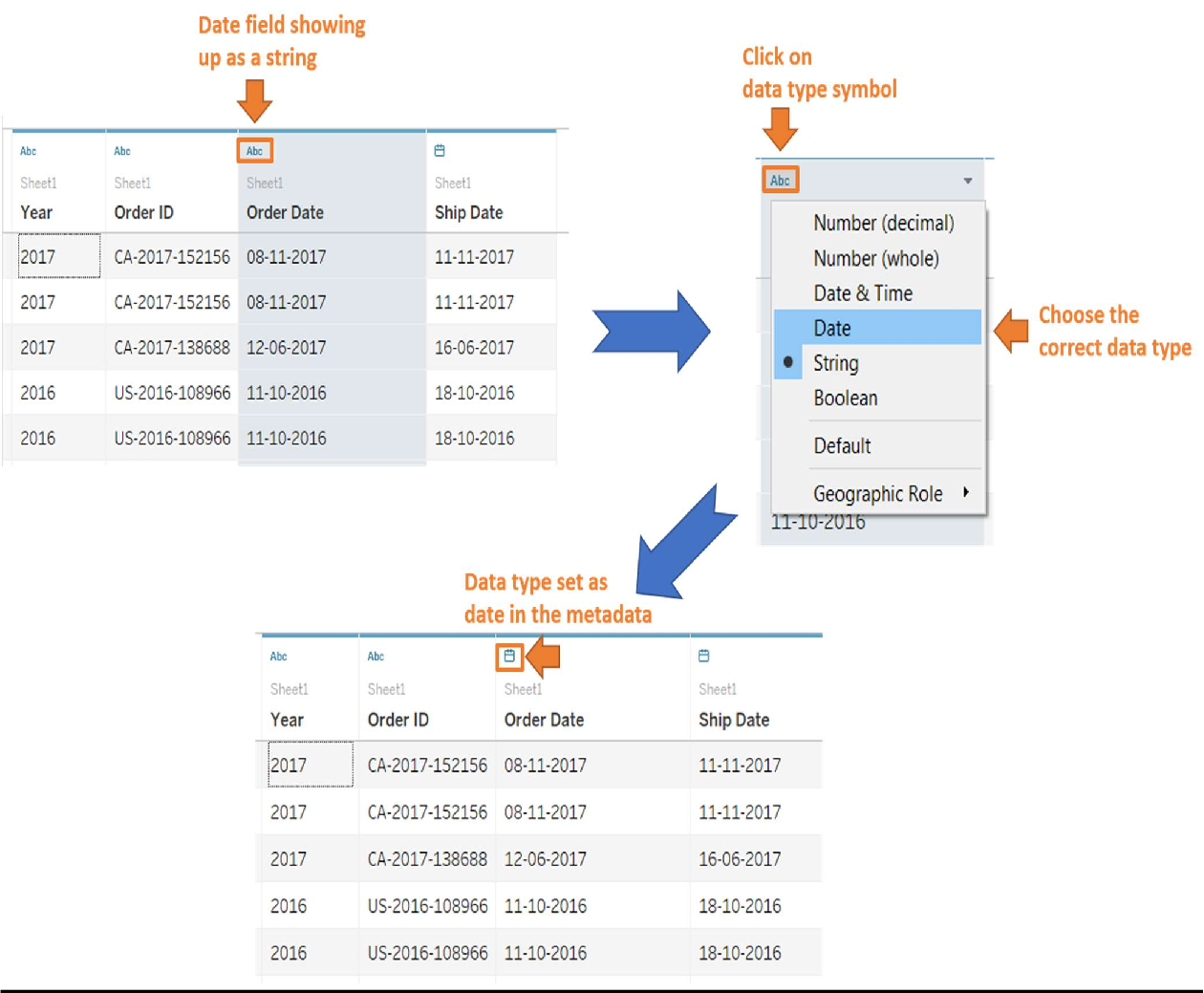
### Multiple data manipulation operations can be carried out to make data fit for visualization in the data preview pane or the data pane in the worksheet view. The data manipulation options vary by the data type of the columns.

For example, the split option only comes up for string datatype columns, the create bins option only comes up for numeric datatype columns, etc.

*Note: All data manipulation is done at the metadata level, which is limited to Tableau only. The changes are not reflected in the original data file or data source.*

CHANGE THE DATATYPE

If Tableau has inferred a wrong data type for a column, the data type can be changed by clicking on the data type symbol in the column header, as shown below.



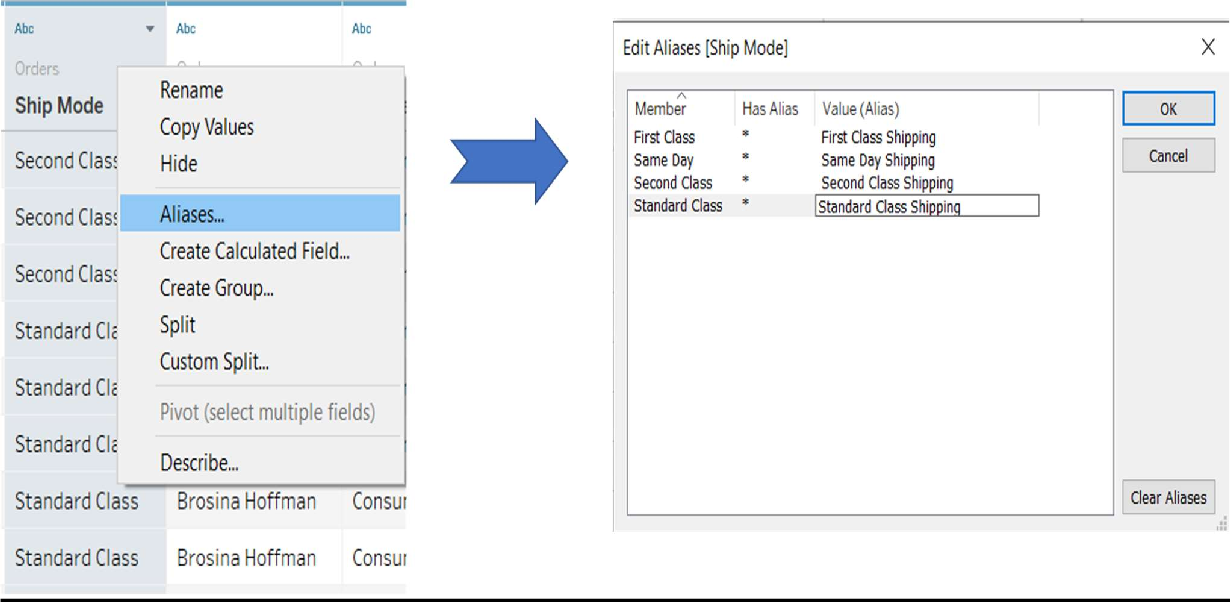
HIDE

Right-click on the column header and select hide to hide an unwanted column from the data view. It is best practice to hide unwanted columns before extract creation in order to reduce the data extract size and reduce latency in load time. Unwanted columns can also be hidden just to prevent any unwanted clutter in the data pane.

*Note: Multiple columns can be selected by holding the control key and left- clicking on the column header. If multiple columns are to be selected in a continuous sequence, hold the shift key and left-click on the first and last*

*column of the sequence. Next, right-click and choose hide to hide them collectively.*

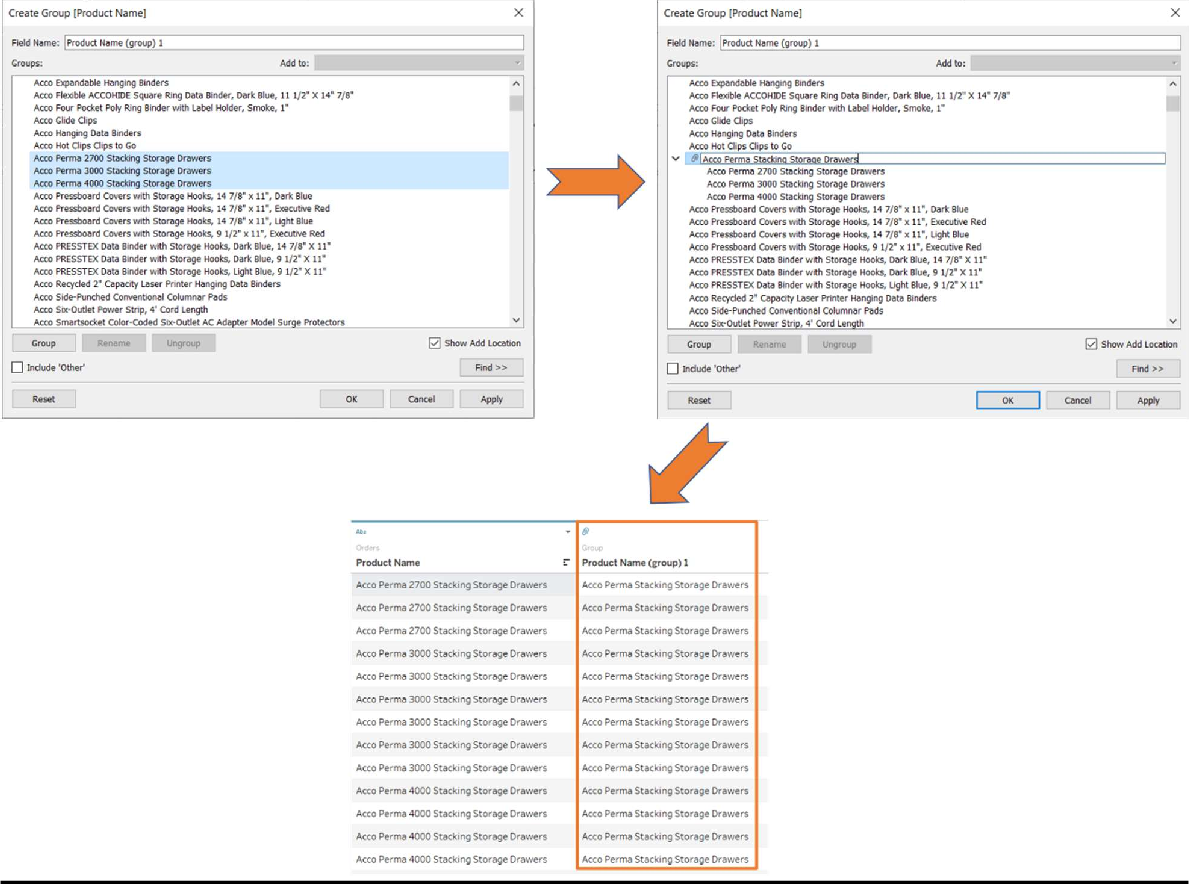
ALIASES

If certain string data values are to be aliased for visuals to make them more intuitive, right-click on the header or the small downward arrow of the column in which data values are to be aliases and choose alias.

An edit alias window pops up showing the distinct values in the Member column, as shown above. You can add an alias for each member to the corresponding values column.

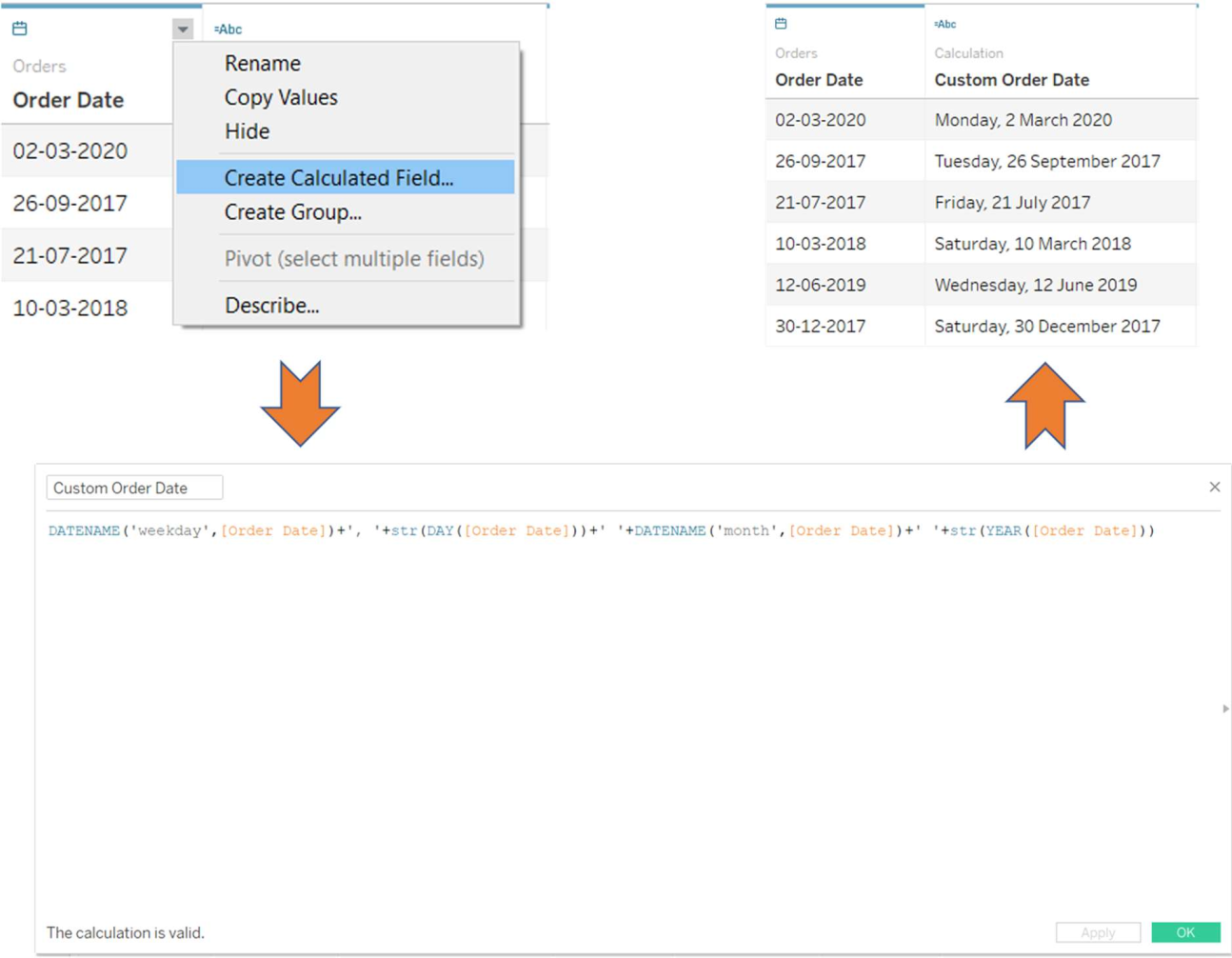
GROUP

Groups can be created in Tableau for string datatype column values in order to assign the same alias for multiple values or reduce the granularity of values for a column. To create a group for a column, right click on its header and choose create group. Next, a create group window pops up in which you can select multiple values. Then click on Group. This results in grouping those values with a common group name, which can be edited. Once you've finished creating the groups, click ok.



As shown above, a new column gets created with the datatype group. This new column with grouped values can be used in the visualization instead of the original column.

CALCULATED FIELD

Calculated fields can be used if you need to create customized logic for manipulating certain data types or data values. There are a large-range of functions available in Tableau that can used individually or collectively for data manipulation. For example, if you want the date format to include the weekday and month in separate columns, a calculated field will need to be created using the formula shown below.

# WEEK – 6

STRUCTURE DATA FOR ANALYSIS

There are certain concepts that are fundamental to understanding data prep and how to structure data for analysis. Data can be generated, captured, and stored in a dizzying variety of formats, but when it comes to analysis, not all data formats are created equal.

Data preparation is the process of getting well formatted data into a single table or multiple related tables so it can be analyzed in Tableau. This includes both the structure, i.e. rows and columns, as well as aspects of data cleanliness, such correct data types and correct data values.

# HOW STRUCTURE IMPACT ANALYSIS

The structure of your data may not be something you can control. The rest of this topic assumes you have access to the raw data and the tools needed to shape it, such as Tableau Prep Builder. However, there may be situations when you can't pivot or aggregate your data as desired. It is often still possible to perform the analysis but you may need to change your calculations or how you approach the data. For an example of how to perform the same analysis with different data structures, see Tableau Prep Day in the Life Scenarios: Analysis with the Second Date in Tableau

Desktop(Link opens in a new window). But if you can optimize the data structure it will likely make your analysis much easier.

##### DATA STRUCTURE

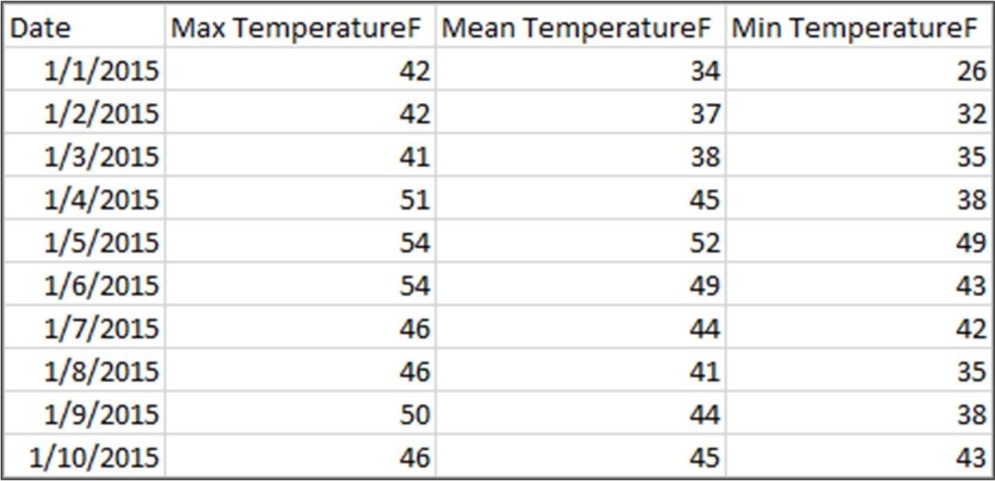
Tableau Desktop works best with data that is in tables formatted like a spreadsheet. That is, data stored in rows and columns, with column headers in the first row. So what should be a row or column?

What is a row?

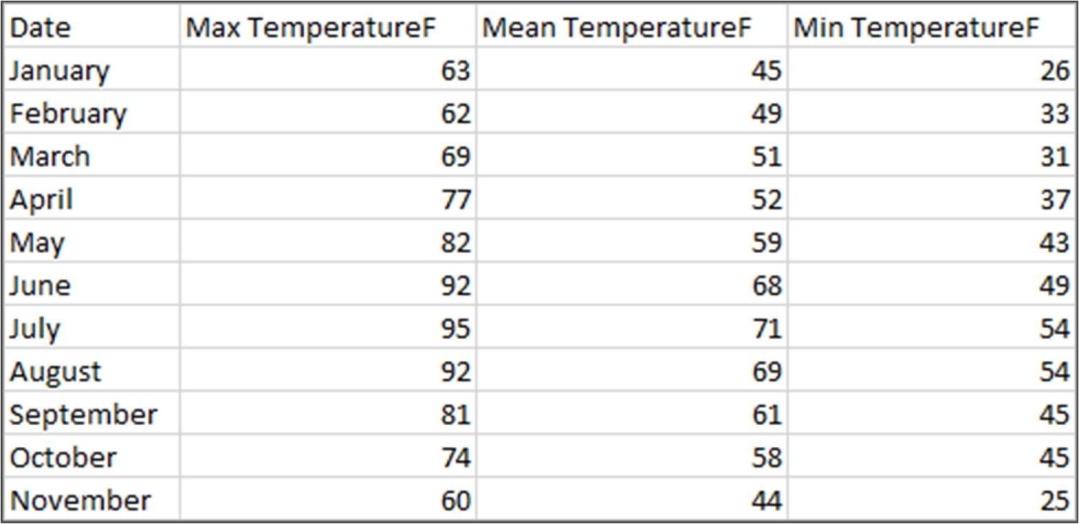
A row, or record, can be anything from information around a transaction at a retail store, to weather measurements at a specific location, or stats about a social media post.

It's important to know what a record (row) in the data represents. This is the *granularity* of the data.

Here each record is a day



Here each record is a month



Try to make sure you can answer the question "What does a row in the data set represent?". This is the same as answering "What does the TableName(Count) field represent?". If you can't articulate that, the data might be structured poorly for analysis.

A concept related to what makes up a row is the idea of aggregation and granularity, which are opposite ends of a spectrum.

##### A ggregation

* refers to how multiple data values are brought together into a single value, such as counting all the Google searches for Pumpkin Spice or taking the average of all the temperature readings around Seattle on a given day.
* By default, measures in Tableau are always aggregated. The default aggregation is SUM. You can change the aggregation to options like Average, Median, Count Distinct, Minimum, etc.

##### Granularity

* refers to how detailed the data is. What does a row or record in the data set represent? A person with malaria? A provinces' total cases of malaria for the month? That's the granularity.
* Knowing the granularity of the data is crucial to working with level of detail (LOD) expressions.

Understanding aggregation and granularity is a critical concept for many reasons; it impacts things like finding useful data sets, building the visualization you want, relating or joining data correctly, and using LOD expressions.

What is column?

A *column* of data in a table comes into Tableau Desktop as a *field* in the data pane, but they are essentially interchangeable terms. (We save the

term *column* in Tableau Desktop for use in the columns and rows shelf and to describe certain visualizations.) A field of data should contain items that can be grouped into a larger relationship. The items themselves are

called *values* or *members* (only discrete dimensions contain members).

##### Data types

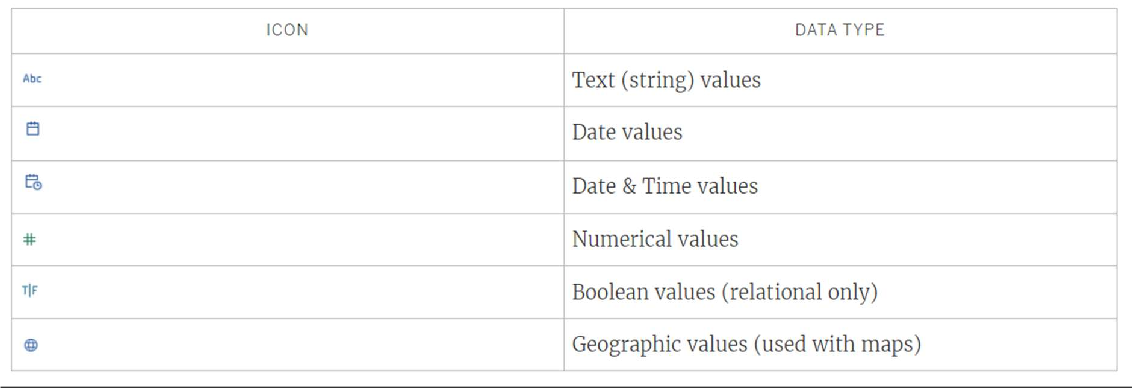
Databases, unlike spreadsheets, usually enforce strict rules on data types. Data types classify the data in a given field and provide information about how the data should be formatted, interpreted, and what operations can be done to that data. For example, numerical fields can have mathematical operations applied to them and geographic fields can be mapped.

Tableau Desktop assigns whether a field is a dimension or measure, but fields have other characteristics that depend on their data type. These are indicated by the icon each field has (though some types share an icon).

Tableau Prep uses the same data types. If data type is enforced on a column and an existing value doesn't match its assigned data type, it may be displayed as null (because "purple" doesn't mean anything as a number).

Some functions require specific data types. For example, you cannot use CONTAINS with a numerical field. Type functions are used to change the data type of a field. For example, DATEPARSE can take a text date in a

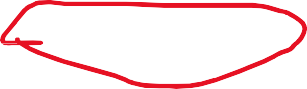
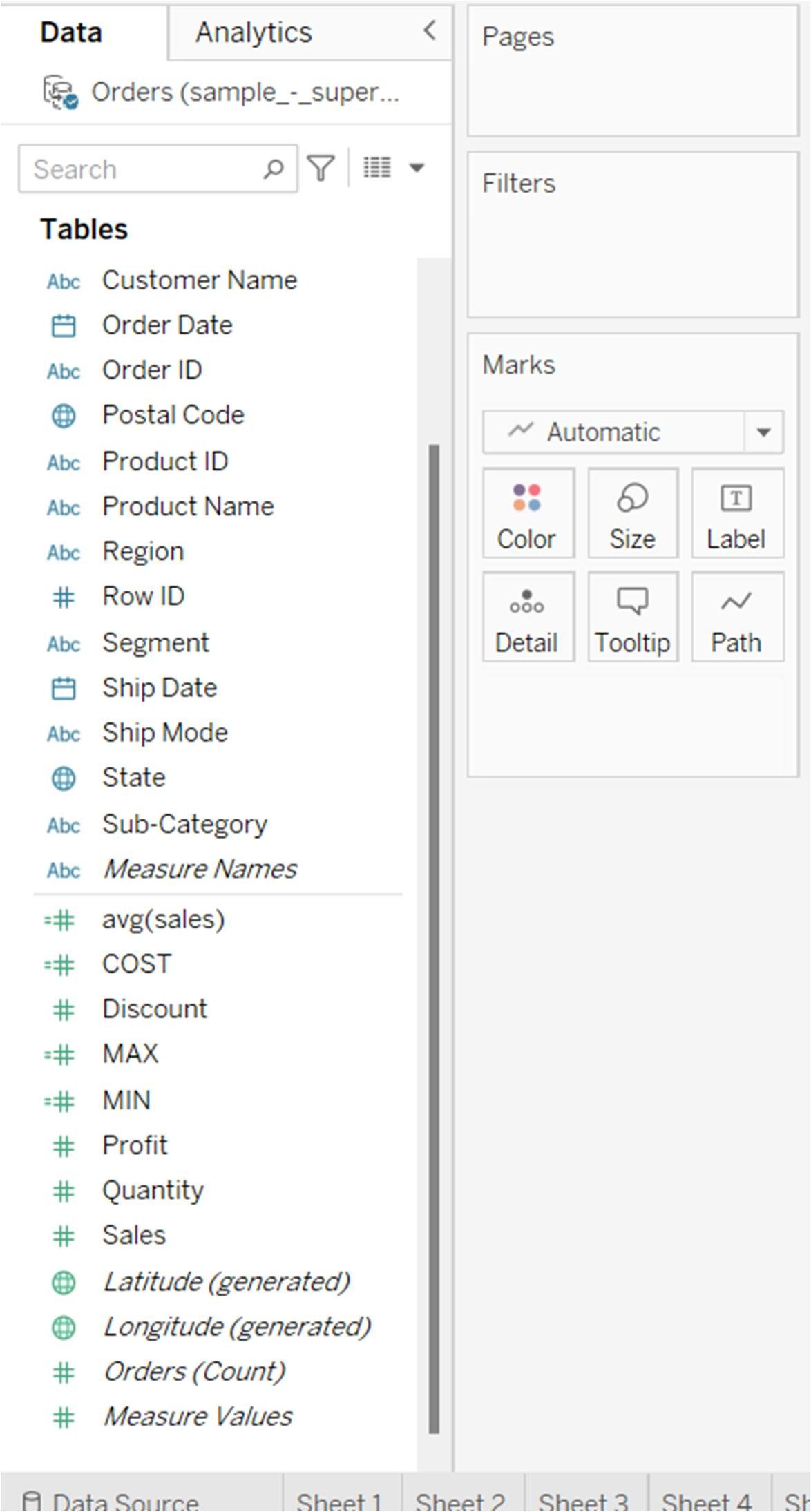
specific format and make it a date, thus enabling things like automatic drill down in the view.



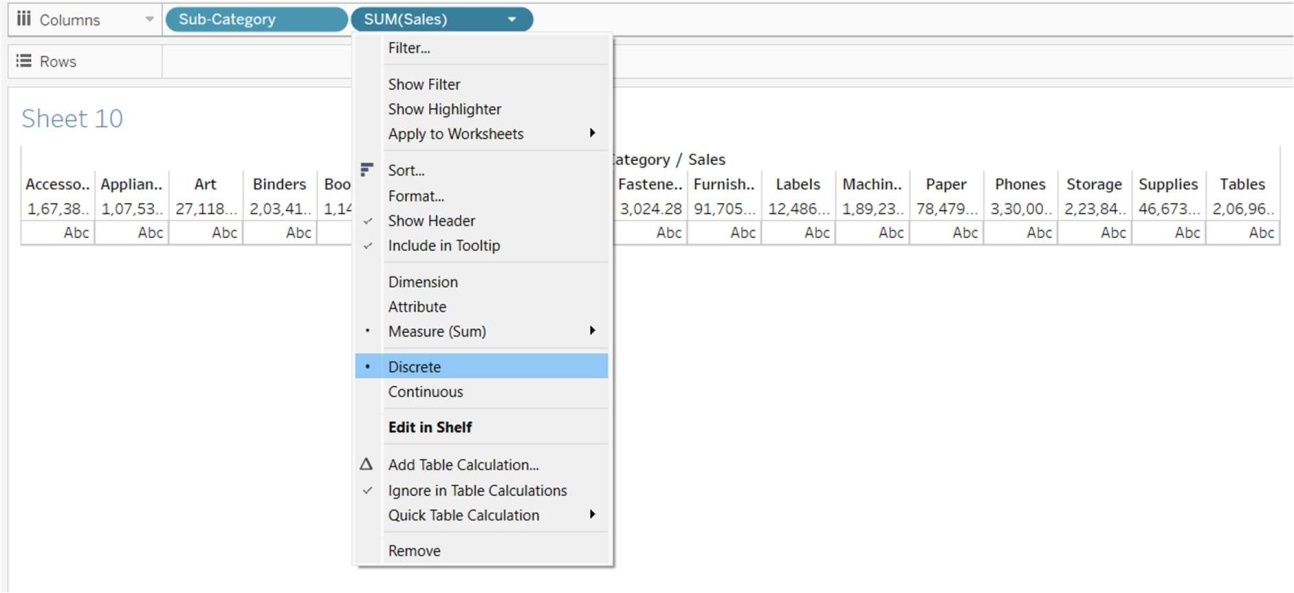
# Categorizing fields

Each column in the data table comes into Tableau Desktop as a field, which appears in the Data pane. Fields in Tableau Desktop must be either a dimension or measure (separated by a line within tables in the Data pane) and either discrete or continuous (color coded: blue fields are discrete and green fields are continuous).

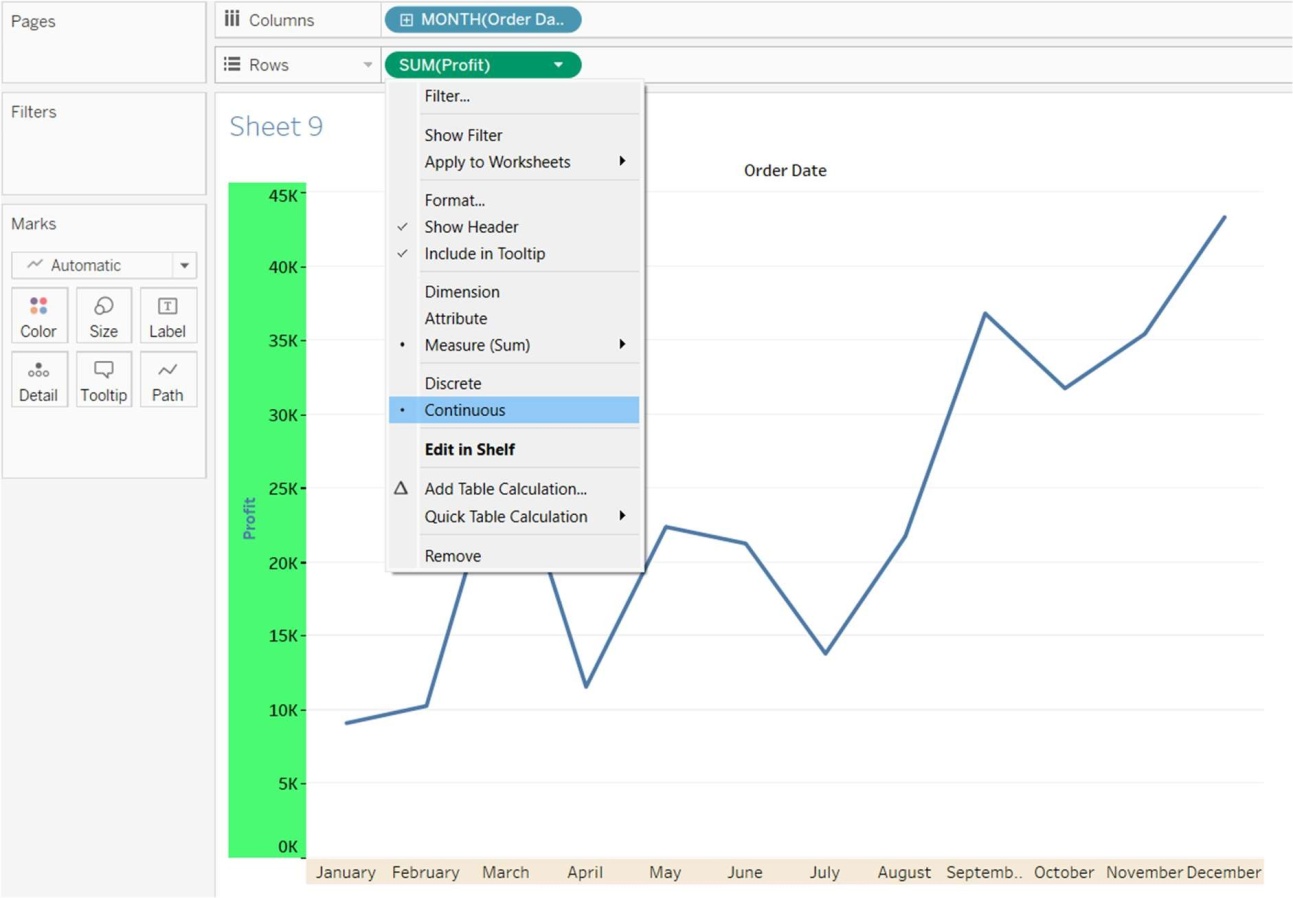
* *Dimensions*are qualitative, meaning they can't be measured but are instead described. Dimensions are often things like city or country, eye color, category, team name, etc. Dimensions are usually discrete.
* *Measures*are quantitative, meaning they can be measured and recorded with numbers. Measures can be things like sales, height, clicks, etc. In Tableau Desktop, measures are automatically aggregated; the default aggregation is SUM. Measures are usually continuous.



* *Discrete*means individually separate or distinct. Toyota is distinct from Mazda. In Tableau Desktop, discrete values come into the view as a label and they create panes.



* *Continuous*means forming an unbroken, continuous whole. 7 is followed by 8 and then it's the same distance to 9, and 7.5 would fall midway between 7 and 8. In Tableau Desktop, continuous values come into the view as an axis.



* Dimensions are usually discrete, and measures are usually continuous. However, this is not always the case. Dates can be either discrete or continuous.
  + Dates are dimensions and automatically come into the view as discrete (aka date parts, such as "August", which considers the month of August without considering other information like the year). A trend line applied to a timeline with discrete dates will be broken into multiple trend lines, one per pane.
  + We can chose to use continuous dates if preferred (aka date truncations, such as "August 2024", which is different than "August 2025"). A trend line applied to a timeline with continuous dates will have a single trend line for the entire date axis.

In Tableau Prep, no distinction is made for dimensions or measures. Understanding the concepts behind discrete or continuous are important, however, for things like understanding the detail versus summary presentation of data in the profile pane.

* Detail: the detail view shows every domain element as a discrete label and has a visual scrollbar to provide a visual overview of all the data.
* Summary: the summary view shows the values as binned on a continuous axis as a histogram.

# SORTING THE TABLEAU DATA

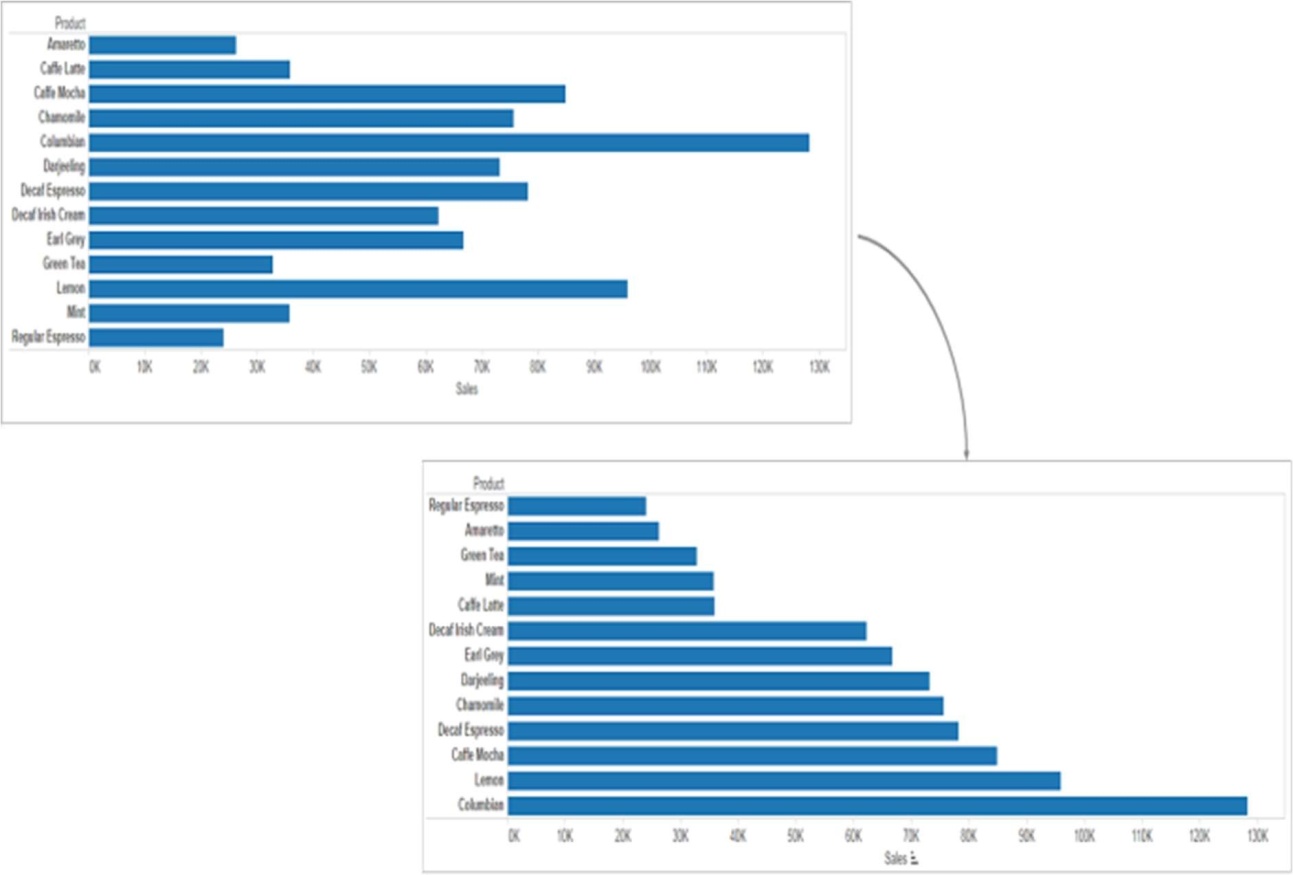
you can sort items in a table based on a measure used in the view. Sorting gives you control over the order that dimension members are listed in a table and can often reveal relationships between ﬁelds.

*To sort items in a table:*

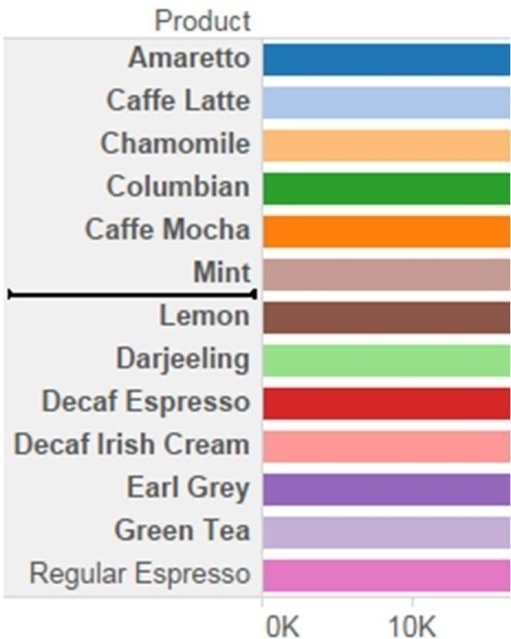
1. Select the items you want to sort.
2. Click one of the sort buttons  on the toolbar.

Note: The tool tips for the sort buttons tell you how the table is going to be sorted.

For example, click the Sort Ascending button to reorder the bars in the chart from lowest sales to highest sales, as shown below.

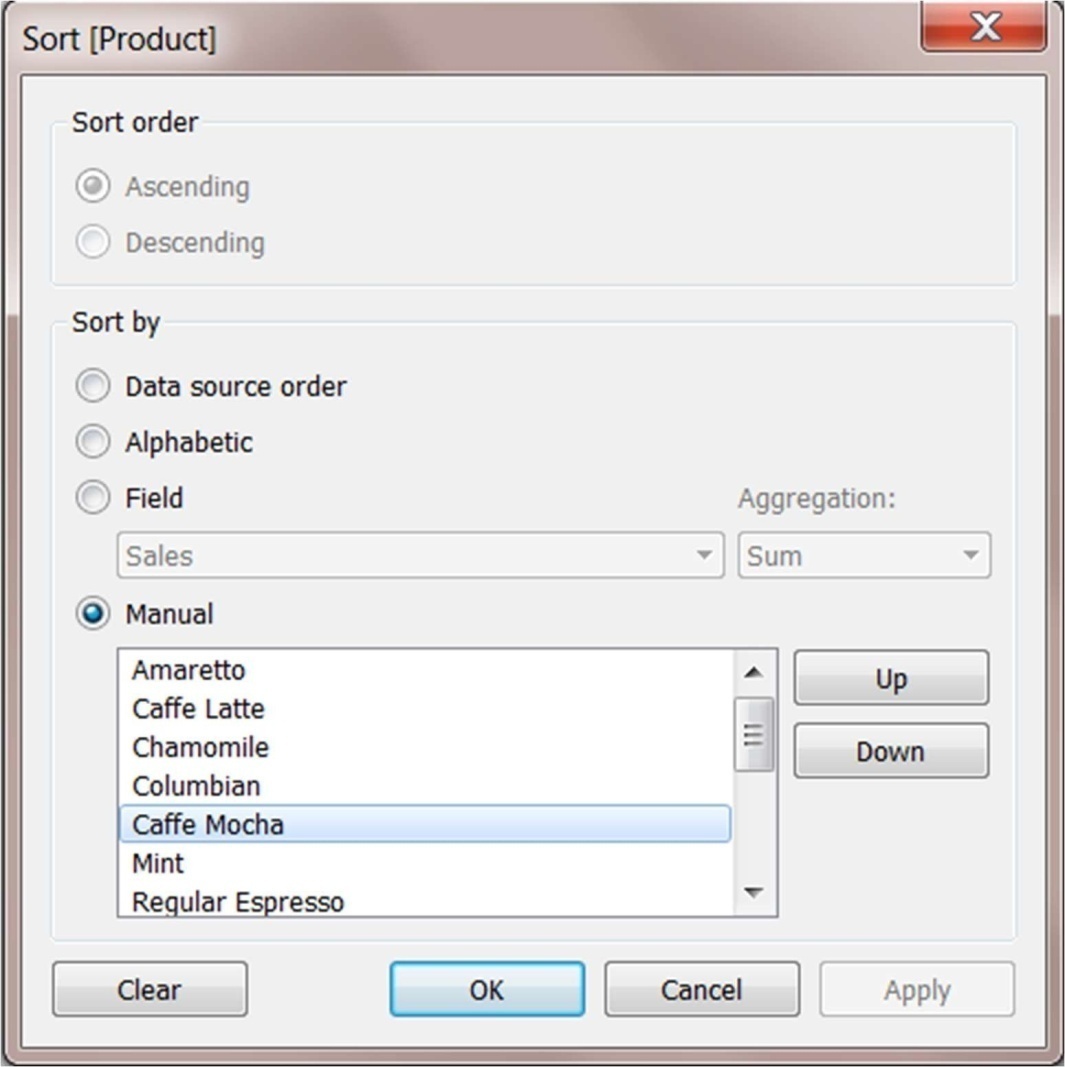


You can also sort items by dragging the ﬁeld labels in a column. In the example below, Regular Espresso is dragged to move below Mint. The new location is indicated by the black line.



You can also manually sort items in the view using the Legend. To manually sort items do the following steps:

1. In the Legend, right-click anywhere in the white space and select Sort from the context menu.
2. In the Sort dialog, in the Manual section, select items that you want to reorder and then use the Up and Down buttons to move items in the list.



# RESETTING THE SORT ORDER

You can reset how a table is sorted by either clicking the Undo button or by reverting the workbook to its saved state.

##### FILTERING TABLEAU DATA (CATEGORICAL DATA)

Dimensions contain discrete categorical data, so filtering this type of field generally involves selecting the values to include or exclude.

When you drag a dimension from the Data pane to the Filters shelf in Tableau Desktop, the following Filter dialog box appears:

* General: Use the General tab to select the values you want to include or exclude.
* Wildcard (Tableau Desktop only): Use the Wildcard tab to define a pattern to filter on. For example, when filtering on email addresses you might want to only include emails from a specific domain. You can define a wildcard filter that ends with "@gmail.com" to only include Google email addresses.
* Condition: Use the Condition tab in the Filter dialog box to define rules to filter by. For example, in a view showing the average Unit Price for a collection of products, you may want to only show the Products that have an average unit price that is greater than or equal to $25. You can use the built-in controls to write a condition or you can write a custom formula.
* Top: Use the Top tab in the Filter dialog box to define a formula that computes the data that will be included in the view. For example, in a view that shows the average Time to Ship for a collection of products, you can decide to only show the top 15 (or bottom) products by Sales. Rather than having to define a specific range for Sales (e.g., greater than $100,000), you can define a limit (top 15) that is relative to the other members in the field (products).



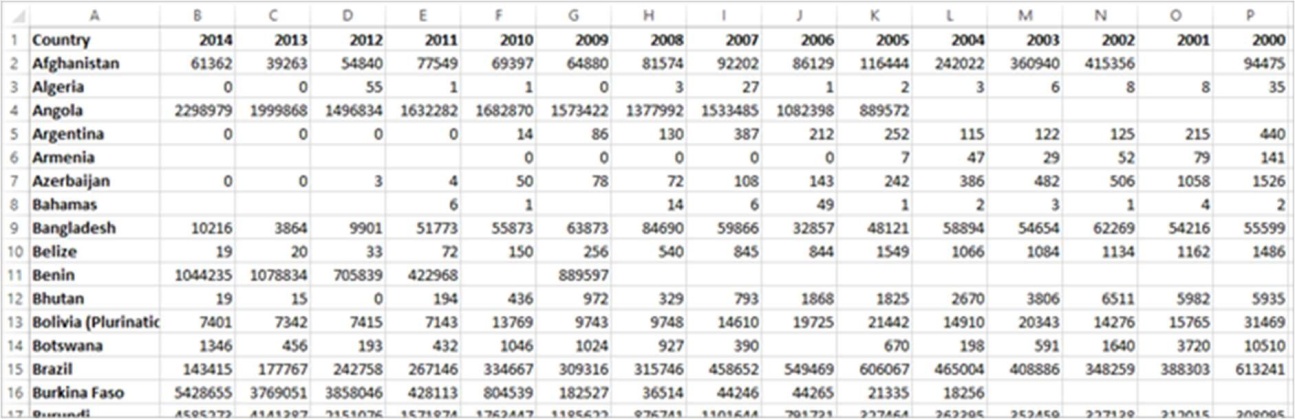
Sheet I1



##### P ivoting and Unpivoting Data

People-friendly data is often captured and recorded in a wide format, with many columns. Machine-readable data, like Tableau prefers, is better in a tall format, with fewer columns and more rows.

W IDE DATA

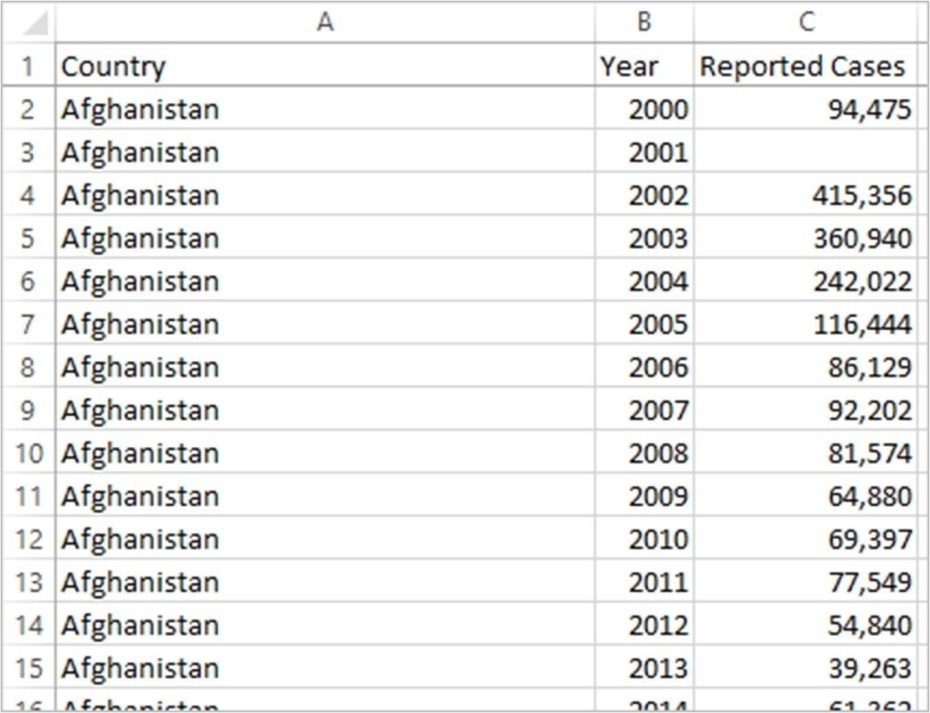


It's easy for a person to read and understand this format. However, if we were to bring this data into Tableau Desktop, we get a ﬁeld per column. We have a ﬁeld for 2000, a ﬁeld for 2001, a ﬁeld for 2002, etc.



T ALL DATA

If we pivot the data, we reshape the data from wide to tall. Now, rather than having a column for each year, we have a single column, Year, and a new column, Reported Cases. In this format we have 1,606 rows and 3 columns. This data format is taller rather than wider.



Now in Tableau Desktop, we have a field for Year and a field for Reported Cases as well as the original Country field. It's much easier to do analysis because each field represents a unique quality about the data set—location, time, and value.

