



Tableau

Understanding Basic Data Concepts

Common types of data sets

What is a data set?

A data set (sometimes referred to as data source, or database) in the context of Tableau, contains the data used to build visualizations. Every bar chart, scatter plot, or line chart you see in Tableau has a connected database or spreadsheet that supplies the data.

Types

- Spreadsheets
- Relational databases
- Cloud data
- Other types of connections

What happens when Tableau Desktop connects to a data set?

When users connect to Tableau, the data fields in their data set are automatically assigned a *role* and a *type*.

Role	Type
A field can be assigned to a <i>Dimension</i> role or a <i>Measure</i> role. We'll learn more about dimensions and measures in the next lesson.	The field's data type defines if the field is, for example, a string, integer, or date. Data types can be changed by the user if Tableau does not assign the type appropriately. Any changes are saved in a Tableau data source (.tds) file as <i>metadata</i> —which is a set of data that describes other data.

A closer look at fields

What is a field?

A field, also known as a column, is a single piece of information from a record in a data set.

For example, if you were to collect data on how many times a week a commuter chose different modes of transportation, your data set could include:

Commuter name	Mode	Days per week
Neil	Walk	2
Neil	Bus	3

Lin	Bicycle	1
Lin	Carpool or Vanpool	3
Lin	Bus	1
Mae	Car (Single Occupancy Vehicle)	2
Mae	Bicycle	3

Commuter name, *Mode*, and *Days per week* are all fields. Commuter name and Mode are qualitative fields, while Days per week is a quantitative field.

The members of the Commuter name and Mode field are limited to relevant categories. You would not include *apple* as an option in Mode because it doesn't match the definition of the field.

The Days per week uses a range of 1-7 rather than categorical members. If you wanted to collect data on modes that people did not use, your range might include 0, too.

You would not include negative values in the Days per week range, since you can't measure a negative trip. If you were measuring temperatures or profit, however, you could have a range that included negative numbers.

In Tableau, quantitative fields are referred to as *Measures*, and qualitative fields are referred to as *Dimensions*.

Qualitative Fields (Dimensions)

- Describes or categorizes data
- Tells you what, when, or who
- Slices the quantitative data

Quantitative fields (Measures)

- Numerical data
- Provides the measurement for qualitative category
- Can be used in calculations

Fields in Tableau

When Tableau connected to this data set, it assigned the fields to either Dimensions or Measures.

- The qualitative fields that describe categories of data are in the top part of the pane, under **Dimensions**.
- The quantitative fields that measure categories of data are in the bottom part of the pane, under **Measures**.

What is a row-level record?

"What does a row of data contain?" is a simple but important question, and it can have a complex answer. Remember the commuter data we reviewed in the previous lesson?

Commuter name	Mode	Days per week
Neil	Walk	2
Neil	Bus	3
Lin	Bicycle	1
Lin	Carpool or Vanpool	3
Lin	Bus	1
Mae	Car (Single Occupancy Vehicle)	2
Mae	Bicycle	3

Each row captured the number of days per week a commuter used a particular mode of transportation to get to work, but it did not include a lot of details about the trips. For example, most people take two commuting trips a day if they go to work, and they might choose a different mode of transportation for each. Commuters might

use multiple modes on a single trip to get to work or to get home. Our first data set doesn't capture those details.

What if you wanted to know:

- How many commuters used multiple modes of transportation in a single day?
- How did commutes to work compare against commutes from work?
- How long did different commutes take?
- What was the length of each trip?

If you wanted to answer those questions, you would need to collect more detailed data, like this:

Commuter name	Trip	Time start	Time end	Date	Mode	Est. miles
Mae	To work	7:45 am	8:50 am	7/17/2019	Bicycle	9
Mae	From work	5:15 pm	5:45 pm	7/17/2019	Bus	6
Mae	From work	5:45 pm	6:15 pm	7/17/2019	Bicycle	3

In the second data set, a row of data is more detailed, or, *granular*, than in the first data set.

Why does knowing data granularity matter?

Data granularity refers to the level of detail for a piece of data, wherever you are looking. As data becomes less granular, we might describe it as an *aggregation*, or as *aggregated data*. Aggregation refers to how data is combined. The level of granularity or aggregation in a row or chart affects the questions we can ask of the data, and the discoveries we can make.

Granularity and aggregation in Tableau

When you move dimensions and measures in and out of a view, the view's level of detail changes.

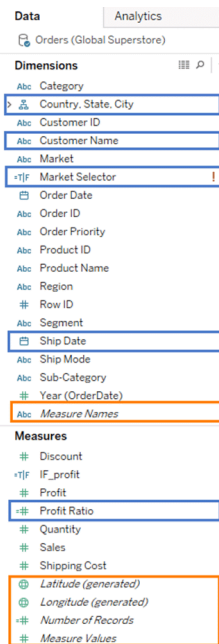
- By default, measures placed in a view are aggregated by SUM, which means that the data for that field in all of the rows is combined.
- Measures can also be aggregated as average, median, count, or count distinct.
- Dimensions break down the aggregated total into smaller totals by category.

The "Aggregation, Granularity, and Ratio Calculations" video from Tableau goes into more detail about aggregation and data granularity. It uses the example of looking at profit ratios and can help you understand the behavior of Tableau charts.

[aggregation_granularity_and_ratio_calculations_transcript.pdf](#)

How data is represented in Tableau

Each field in the Tableau **Data** pane has an icon to the left showing the field's data type. Fields are assigned both a data type and role automatically when you connect to data in Tableau. You can manually change the type in the data grid or **Data** pane.



1. Hierarchy

The Country, State, and City fields have been organized into a hierarchy. This is useful for creating drill-downs in charts.

2. Blue text field

- Customer name is a categorical, or qualitative, field.
- A blue icon indicates that the field is discrete, which means it is data that contains separate parts.
- Most dimensions are discrete fields.

3. Invalid field

The red exclamation point on this calculated field indicates a problem. The red exclamation point can mean a field is missing, or, in this case, that the calculated field is broken.

4. Date

The Ship Date icon is a date field. Dates can be discrete or continuous. The date in this image is blue, which tells us it is discrete.

If you are building charts that use date fields, it's important to understand how Tableau processes dates.

5. Calculated field

Profit Ratio is a calculated field. The green icon indicates it is a continuous field.

Continuous means data value in the set can be any value in the range.

Most measures are continuous fields.

6. Tableau-generated fields

The fields highlighted in orange are all Tableau-generated fields.

Sometimes they are added to the view automatically, depending on the type of chart you are building. As you become more proficient at building charts, you will learn how to use these fields without Tableau's help.

Abc

Text or string
values.



Discrete date
and time field.



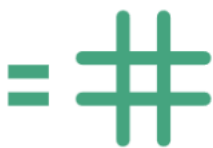
Discrete date
field.



Geographic field,
such as State or
Zip Code.



Continuous
numeric value.



The equal sign indicates
this is a calculated field.