# Tableau

Public version: <https://public.tableau.com/s/download>

Student version: <https://www.tableau.com/academic/students>

Trial version: <https://www.tableau.com/products/trial>

Paid version: <https://www.tableau.com/pricing/individual>

# Lesson 1: Connect To Excel

1. Open Tableau
2. Connect
   1. To a File
      1. Excel
         1. Browse and select file
         2. Drag and drop the sheet you want to load
         3. Create a “Sheet”
         4. Select “Dimensions” & “Measures”
         5. “Save” [Public vs Desktop”]
      2. Text File
      3. Access
      4. Statistical File
   2. To a Server
      1. OData
      2. More…

# Lesson 2: Pie Charts

## Dimensions:

The parameters using which you would categorize/visualize. Categorical Values

## Measures:

The actual measurements. Numerical Values

## Show Me:

1. This shows the various types of visualization
2. And also the dimension & measures required for each visualization

## Filter Section:

1. Add Dimension to be filtered with, and filtered based on available values

## Row & Columns:

1. Display a dimension/Measure as row/column

## Marks Section:

1. For Colors of the slice/data
2. Labels on the slice
3. Size of data
4. Detail
5. Tool Tip

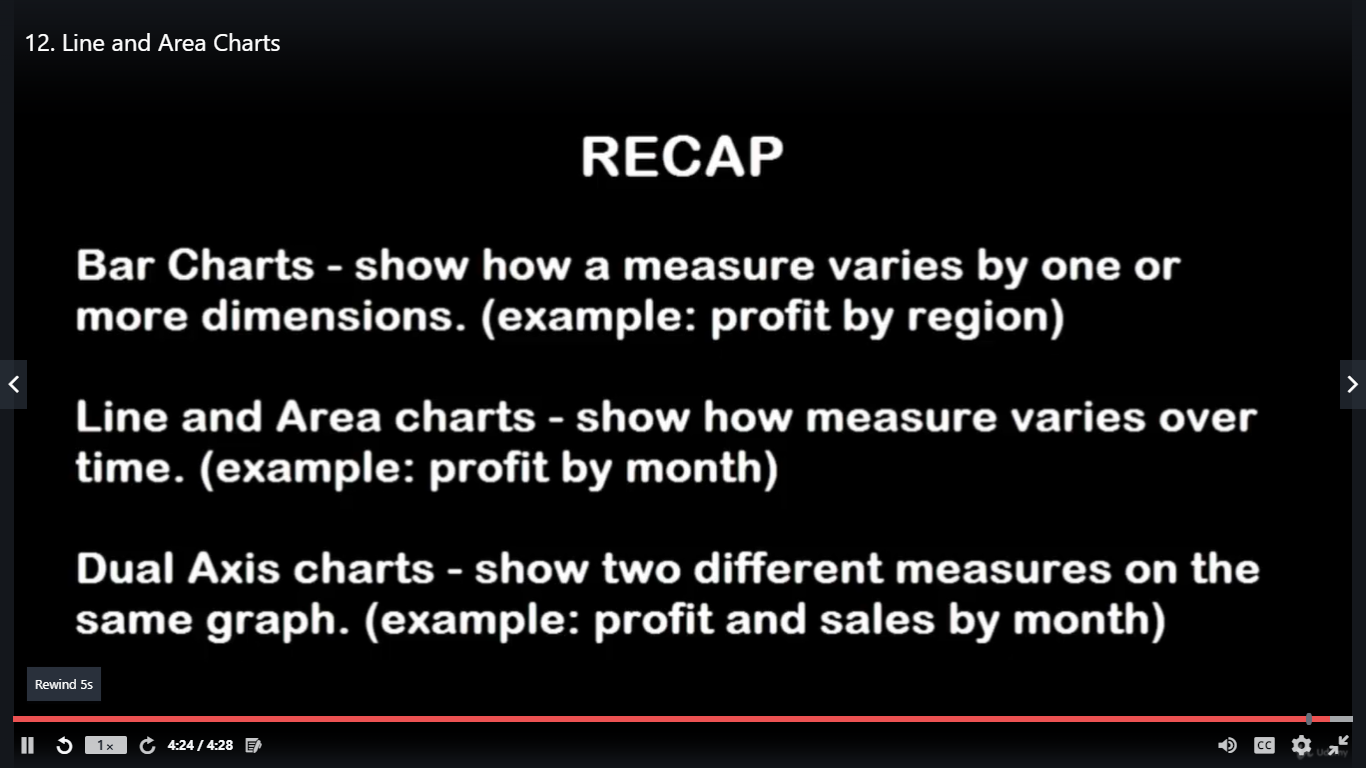
Ctrl+Z: Undo operation

# Lesson 3: Bar Charts [Measure varies with dimension]

1. Stacked Bar Chart
2. Side-By-Side Bar Chart

Sort: To sort the graph in asc/desc

# Lesson 4: Line & Areas Charts [Measure varies over time]



Dual Axis Charts : More than 2 measure and one dimension

Hands-on: Sales on 2 different years

# Lesson 5: Box Charts & Whisker plots

<https://www.dummies.com/programming/big-data/data-science/box-plots-graphical-technique-for-statistical-data/>

A box plot is designed to show several key statistics for a dataset in the form of a vertical rectangle or box. The statistics it can show include the following:

1. Minimum value
2. Maximum value
3. First quartile (Q1)
4. Second quartile (Q2)
5. Third quartile (Q3)
6. Interquartile range (IQR)

Box-Plots require Disaggregate Values

## Percentile

<https://www.dummies.com/education/math/statistics/how-to-calculate-percentiles-in-statistics/>

The kth percentile is a value in a data set that splits the data into two pieces: The lower piece contains k percent of the data, and the upper piece contains the rest of the data (which amounts to [100 – k] percent, because the total amount of data is 100%). **Note:** k is any number between 0 and 100.

The **median** is the 50th percentile: the point in the data where 50% of the data fall below that point, and 50% fall above it.

To calculate the kth percentile (where k is any number between zero and one hundred), do the following steps:

1. Order all the values in the data set from smallest to largest.
2. Multiply k percent by the total number of values, n.This number is called the index.
3. If the index obtained in Step 2 is not a whole number, round it up to the nearest whole number and go to Step 4a. If the index obtained in Step 2 is a whole number, go to Step 4b.
   1. Count the values in your data set from left to right (from the smallest to the largest value) until you reach the number indicated by Step 3. The corresponding value in your data set is the k**th** percentile.
   2. Count the values in your data set from left to right until you reach the number indicated by Step 2. The kth percentile is the average of that corresponding value in your data set and the value that directly follows it.

For example, suppose you have 25 test scores, and in order from lowest to highest they look like this: 43, 54, 56, 61, 62, 66, 68, 69, 69, 70, 71, 72, 77, 78, 79, 85, 87, 88, 89, 93, 95, 96, 98, 99, 99. To find the 90th percentile for these (ordered) scores, start by multiplying 90% times the total number of scores, which gives 90% ∗ 25 = 0.90 ∗ 25 = 22.5 (the index). Rounding up to the nearest whole number, you get 23.

Counting from left to right (from the smallest to the largest value in the data set), you go until you find the 23rd value in the data set. That value is 98, and it’s the 90th percentile for this data set.

Now say you want to find the 20th percentile. Start by taking 0.20 x 25 = 5 (the index); this is a whole number, so proceed from Step 3 to Step 4b, which tells you the 20th percentile is the average of the 5th and 6th values in the ordered data set (62 and 66). The 20th percentile then comes to (62 + 66) ÷ 2 = 64.

The median (the 50th percentile) for the test scores is the 13th score: 77.

The steps shown here demonstrate one way of [**calculating percentiles**](https://www.dummies.com/education/math/statistics/how-to-calculate-percentiles-in-statistics/), but there are several other acceptable methods. Do not be too alarmed if your calculator or a friend gives you a value close to but different from what these steps would give.

## Quartile/Hinge

*Quartiles* split up a data set into four equal parts, each consisting of 25 percent of the sorted values in the data set. Quartiles are related to percentiles like so:

First quartile (Q1) = 25th percentile

Second quartile (Q2) = 50th percentile

Third quartile (Q3) = 75th percentile

Because the second quartile is the 50th percentile, it’s also the **median** of a data set. The fourth quartile usually isn’t used because its value is greater than every element in a data set, so what’s the point?

One commonly used approach for calculating quartiles follows these two steps:

1. Split the data into a lower half and an upper half (leaving out the median).
2. Compute the median of the lower half and the upper half.
3. After you’ve split the data into lower and upper halves, you figure out the quartiles as follows:
   1. Q1= the median of the lower half
   2. Q2 = the median of the entire data set
   3. Q3 = the median of the upper half

**Eg:**

The following data represent a sample of eight stock returns for Gamma Industries:

5, 7, 6, 3, 0, –2, 4, 3

The sorted values are:

–2, 0, 3, 3, 4, 5, 6, 7

In this example, you have eight elements. Because 8 is an even number, the median is the average of the fourth and fifth elements: –2, 0, 3, 3, 4, 5, 6, 7

(3 + 4)/2 = 3.5. Therefore, the second quartile (Q2) is 3.5.

The values below the median constitute the lower half of the sorted sample

–2, 0, 3, 3

The values above the median constitute the upper half of the sorted sample

4, 5, 6, 7

Both the lower and upper halves have four sample elements. Because 4 is an even number, the median is the average of the second and third elements.

For the lower half, the median is: (0 + 3)/2 = 1.5. This is the average value of the two middle elements. Therefore, the first quartile (Q1) is 1.5.

For the upper half, the median is (5 + 6)/2 = 5.5. Therefore, the third quartile (Q3) is 5.5.

As with percentiles, Microsoft Excel uses a different approach to computing quartiles; if you use the QUARTILE function, you will get 3.5 for Q2, but you will also get

2.25 for Q1 (instead of 1.5)

5.25 for Q3 (instead of 5.5)

## Interquartile Range:

The interquartile range (IQR) is the difference between the third quartile and first quartile: **IQR = Q3 – Q1.**

The interquartile range is a measure of dispersion; it shows how much spread there is between the elements in the middle 50 percent of a dataset.

## Outlier

A box plot is drawn so that

1. The top of the box represents the third quartile (Q3) of the data.
2. The bottom of the box represents the first quartile (Q1) of the data.
3. The middle of the box (shown with a line) represents the second quartile (Q2).

In addition, there’s a line above the box to indicate the maximum value in the data that doesn’t exceed **Q3 + 1.5 x IQR** and a line below the box to indicate the minimum value in the data that doesn’t fall below **Q1 – 1.5 x IQR**. Values outside of this range are **outliers** and are shown on the box plot as individual points.\

# Lesson 6: Histograms

<https://www.dummies.com/programming/big-data/data-science/histograms-graphical-technique-for-statistical-data/>

A histogram is a graph that represents the probability distribution of a dataset. A histogram has a series of vertical bars where each bar represents a single value or a range of values for a variable. The heights of the bars indicate the frequencies or probabilities for the different values or ranges of values.

Edit Bin: Could be used to edit the range of the bin

Zoom In:

View data:

Second Measure:

# Lesson 7: Heat Map & Tree Map

## Heat Map:

Density of distribution

Scatter Plot

Density/Magnitude – using color

2 methods:

1. Density
2. “Heat Map” template

## Tree Map

Uses Color and Size for depicting the density.

Nested Trees

# Lesson 8: Scatter Plot

1. 2 Dimensions required.
2. Disaggregate data
3. “Density” – heat map

# Tips for the Certified Test

Common Questions: Cross Tab/Bar Chart

Questions on Dates: Line Chart

Questions on distribution – Histogram