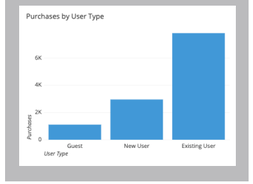
Types of Graphs

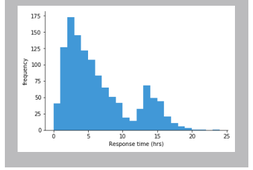
Bar Chart

A bar chart (aka bar graph, column chart) plots numeric values for levels of a categorical feature as bars. Levels are plotted on one chart axis, and values are plotted on the other axis. Each categorical value claims one bar, and the length of each bar corresponds to the bar's value. Bars are plotted on a common baseline to allow for easy comparison of values.



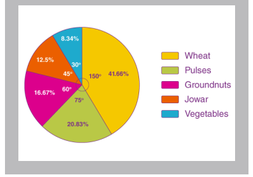
Histogram

A histogram is a chart that plots the distribution of a numeric variable's values as a series of bars. Each bar typically covers a range of numeric values called a bin or class; a bar's height indicates the frequency of data points with a value within the corresponding bin.



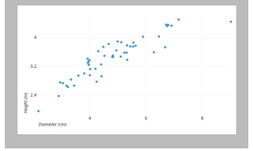
Pie Chart

A pie chart is a type of graph that represents the data in the circular graph. The slices of pie show the relative size of the data, and it is a type of pictorial representation of data. A pie chart requires a list of categorical variables and numerical variables. Here, the term "pie" represents the whole, and the "slices" represent the parts of the whole.



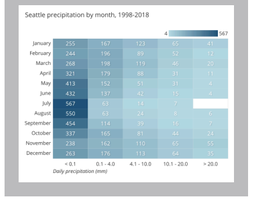
Scatter Chart

A scatter plot uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.



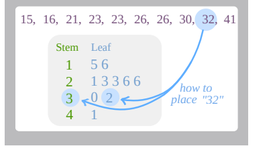
Heatmap

A heatmap depicts values for a main variable of interest across two axis variables as a grid of colored squares. The axis variables are divided into ranges like a bar chart or histogram, and each cell's color indicates the value of the main variable in the corresponding cell range.



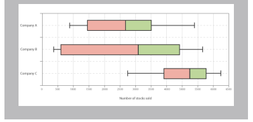
Stem and Leaf plot

A stem and leaf plot is used to organize data as they are collected. A stem and leaf plot looks something like a bar graph. Each number in the data is broken down into a stem and a leaf, thus the name. The stem of the number includes all but the last digit. The leaf of the number will always be a single digit.Mar 31, 2021



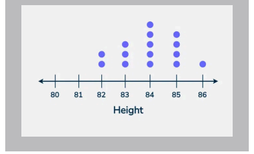
Box & Whisker plot

A Box and Whisker Plot is a convenient way of visually displaying the data distribution through their quartiles.
The lines extending parallel from the boxes are known as the "whiskers", which are used to indicate variability outside the upper and lower quartiles. Outliers are sometimes plotted as individual dots that are in-line with whiskers. Box Plots can be drawn either vertically or horizontally.



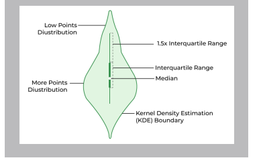
Dot plot

A dot plot is a simple form of data visualization that consists of data points plotted as dots on a graph with an x- and y-axis. These types of charts are used to graphically depict certain data trends or groupings. A dot plot is similar to a histogram in that it displays the number of data points that fall into each category or value on the axis, thus showing the distribution of a set of data. There are two types of dot plots: the Cleveland and Wilkinson dot plots. This type of charting method is commonly used by the Federal Open Market Committee (FOMC).



Violin plot

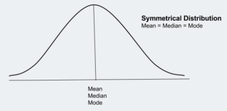
Violin Plot is a method to visualize the distribution of numerical data of different variables. It is quite similar to Box Plot but with a rotated plot on each side, giving more information about the density estimate on the y-axis. The density is mirrored and flipped over, and the resulting shape is filled in, creating an image resembling a violin. The advantage of a violin plot is that it can show nuances in the distribution that aren't perceptible in a boxplot. On the other hand, the boxplot more clearly shows the outliers in the data. Violin Plots hold more information than box plots, they are less popular. Because of their unpopularity, their meaning can be harder to grasp for many readers not familiar with the violin plot representation.



Symmetric & Skewness distribution

Symmetrical distribution

A symmetric distribution is one where the left and right sides of the histogram or probability density function (PDF) are mirror images of each other. In such distributions, the mean, median, and mode are usually equal or very close. A common example is the normal distribution (bell curve), where data is evenly distributed around the center. Another example is the uniform distribution, where all values occur with equal probability. Symmetric distributions are widely used in statistical analysis, as they simplify calculations and make predictions more reliable.

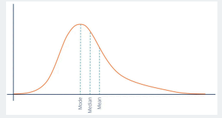


Skewness distribution

skewed distributions are asymmetrical, meaning data is more spread out on one side of the mean than the other. Right (positive) skewed distributions have a long tail on the right, with the mean greater than the median, such as income distribution, where a few high earners pull the mean upward. Left (negative) skewed distributions have a long tail on the left, with the mean less than the median, like the age at retirement, where most people retire at a certain age but some retire much earlier. Skewness is important in fields like finance and machine learning, as it affects risk assessment, forecasting, and model performance.

Positively Skewed Distribution

A positively skewed (or right-skewed) distribution is a type of distribution in which most values are clustered around the left tail of the distribution while the right tail of the distribution is longer.



Negatively Skewed Distribution

A negatively skewed (also known as left-skewed) distribution is a type of distribution in which more values are concentrated on the right side (tail) of the distribution graph while the left tail of the distribution graph is longer.