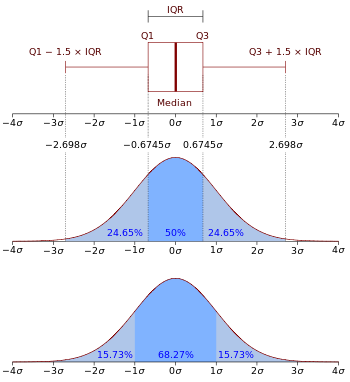
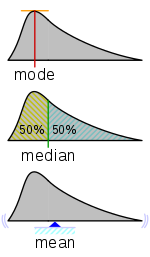
**Probability density function**

In [probability theory](https://en.wikipedia.org/wiki/Probability_theory), a probability density function (PDF), or density of a [continuous random variable](https://en.wikipedia.org/wiki/Continuous_random_variable), is a [function](https://en.wikipedia.org/wiki/Function_(mathematics)) whose value at any given sample (or point) in the [sample space](https://en.wikipedia.org/wiki/Sample_space) (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample. In other words, while the absolute likelihood for a continuous random variable to take on any particular value is 0 (since there are an infinite set of possible values to begin with), the value of the PDF at two different samples can be used to infer, in any particular draw of the random variable, how much more likely it is that the random variable would equal one sample compared to the other sample.

In a more precise sense, the PDF is used to specify the probability of the [random variable](https://en.wikipedia.org/wiki/Random_variable) falling within a particular range of values, as opposed to taking on any one value. This probability is given by the [integral](https://en.wikipedia.org/wiki/Integral) of this variable's PDF over that range—that is, it is given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range. The probability density function is nonnegative everywhere, and its integral over the entire space is equal to one.

The terms "probability distribution function" and "probability function"[[4]](https://en.wikipedia.org/wiki/Probability_density_function#cite_note-4) have also sometimes been used to denote the probability density function. However, this use is not standard among probabilists and statisticians. In other sources, "probability distribution function" may be used when the [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution) is defined as a function over general sets of values, or it may refer to the [cumulative distribution function](https://en.wikipedia.org/wiki/Cumulative_distribution_function), or it may be a [probability mass function](https://en.wikipedia.org/wiki/Probability_mass_function) (PMF) rather than the density. "Density function" itself is also used for the probability mass function, leading to further confusion.[[5]](https://en.wikipedia.org/wiki/Probability_density_function#cite_note-5) In general though, the PMF is used in the context of discrete random variables (random variables that take values on a discrete set), while PDF is used in the context of continuous random variables.



**Reference :** <https://en.wikipedia.org/wiki/Probability_density_function>

**Sample space**

In [probability theory](https://en.wikipedia.org/wiki/Probability_theory), the sample space (also called sample description space or possibility space) of an [experiment](https://en.wikipedia.org/wiki/Experiment_(probability_theory)) or random [trial](https://en.wikipedia.org/wiki/Trial_and_error) is the [set](https://en.wikipedia.org/wiki/Set_(mathematics)) of all possible [outcomes](https://en.wikipedia.org/wiki/Outcome_(probability)) or results of that experiment. A sample space is usually denoted using [set notation](https://en.wikipedia.org/wiki/Set_notation), and the possible ordered outcomes are listed as [elements](https://en.wikipedia.org/wiki/Element_(mathematics)) in the set. It is common to refer to a sample space by the labels S, Ω, or U (for "[universal set](https://en.wikipedia.org/wiki/Universe_(mathematics))").

For example, if the experiment is tossing a coin, the sample space is typically the set {head, tail}. For tossing two coins, the corresponding sample space would be {(head,head), (head,tail), (tail,head), (tail,tail)}, commonly written {HH, HT, TH, TT}. If the sample space is unordered, it becomes {{head,head}, {head,tail}, {tail,tail}}.

For tossing a single six-sided [die](https://en.wikipedia.org/wiki/Dice), the typical sample space is {1, 2, 3, 4, 5, 6} (in which the result of interest is the number of pips facing up)

**Experiment (probability theory)**

In probability theory, an experiment or trial (see below) is any procedure that can be infinitely repeated and has a well-defined [set](https://en.wikipedia.org/wiki/Set_(mathematics)) of possible [outcomes](https://en.wikipedia.org/wiki/Outcome_(probability)), known as the [sample space](https://en.wikipedia.org/wiki/Sample_space). An experiment is said to be random if it has more than one possible outcome, and deterministic if it has only one. A random experiment that has exactly two ([mutually exclusive](https://en.wikipedia.org/wiki/Mutually_exclusive_events)) possible outcomes is known as a [Bernoulli trial](https://en.wikipedia.org/wiki/Bernoulli_trial).

When an experiment is conducted, one (and only one) outcome results— although this outcome may be included in any number of [events](https://en.wikipedia.org/wiki/Event_(probability_theory)), all of which would be said to have occurred on that trial. After conducting many trials of the same experiment and pooling the results, an experimenter can begin to assess the [empirical probabilities](https://en.wikipedia.org/wiki/Empirical_probability) of the various outcomes and events that can occur in the experiment and apply the methods of [statistical analysis](https://en.wikipedia.org/wiki/Statistics).

**Empirical probability**

The empirical probability, [relative frequency](https://en.wikipedia.org/wiki/Frequency_(statistics)), or experimental probability of an event is the ratio of the number of outcomes in which a specified event occurs to the total number of trials, not in a theoretical sample space but in an actual experiment. In a more general sense, empirical probability estimates probabilities from [experience](https://en.wikipedia.org/wiki/Experience) and [observation](https://en.wikipedia.org/wiki/Observation).

Given an event A in a sample space, the relative frequency of A is the ratio m/n, m being the number of outcomes in which the event A occurs, and n being the total number of outcomes of the experiment.

In statistical terms, the empirical probability is an estimate or [estimator](https://en.wikipedia.org/wiki/Estimator) of a probability. In simple cases, where the result of a trial only determines whether or not the specified event has occurred, modelling using a [binomial distribution](https://en.wikipedia.org/wiki/Binomial_distribution) might be appropriate and then the empirical estimate is the [maximum likelihood estimate](https://en.wikipedia.org/wiki/Maximum_likelihood_estimate). It is the [Bayesian estimate](https://en.wikipedia.org/wiki/Bayesian_estimate) for the same case if certain assumptions are made for the [prior distribution](https://en.wikipedia.org/wiki/Prior_distribution) of the probability. If a trial yields more information, the empirical probability can be improved on by adopting further assumptions in the form of a [statistical model](https://en.wikipedia.org/wiki/Statistical_model): if such a model is fitted, it can be used to derive an estimate of the probability of the specified event

**Statistical model**

A statistical model is a [mathematical model](https://en.wikipedia.org/wiki/Mathematical_model) that embodies a set of [statistical assumptions](https://en.wikipedia.org/wiki/Statistical_assumptions) concerning the generation of [sample data](https://en.wikipedia.org/wiki/Sample_(statistics)) (and similar data from a larger [population](https://en.wikipedia.org/wiki/Statistical_population)). A statistical model represents, often in considerably idealized form, the data-generating process.

A statistical model is usually specified as a mathematical relationship between one or more [random variables](https://en.wikipedia.org/wiki/Random_variables) and other non-random variables. As such, a statistical model is "a formal representation of a theory" ([Herman Adèr](https://en.wikipedia.org/wiki/Herman_J._Ad%C3%A8r) quoting [Kenneth Bollen](https://en.wikipedia.org/wiki/Kenneth_A._Bollen)).

All [statistical hypothesis tests](https://en.wikipedia.org/wiki/Statistical_hypothesis_testing) and all [statistical estimators](https://en.wikipedia.org/wiki/Estimator) are derived via statistical models. More generally, statistical models are part of the foundation of [statistical inference](https://en.wikipedia.org/wiki/Statistical_inference).

**Statistical inference**

Statistical inference is the process of using [data analysis](https://en.wikipedia.org/wiki/Data_analysis) to deduce properties of an underlying [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution). Inferential statistical analysis infers properties of a [population](https://en.wikipedia.org/wiki/Statistical_population), for example by testing hypotheses and deriving estimates. It is assumed that the observed data set is [sampled](https://en.wikipedia.org/wiki/Sampling_(statistics)) from a larger population.

Inferential statistics can be contrasted with [descriptive statistics](https://en.wikipedia.org/wiki/Descriptive_statistics). Descriptive statistics is solely concerned with properties of the observed data, and it does not rest on the assumption that the data come from a larger population.

**Probability distribution**

In [probability theory](https://en.wikipedia.org/wiki/Probability_theory) and [statistics](https://en.wikipedia.org/wiki/Statistics), a probability distribution is a mathematical [function](https://en.wikipedia.org/wiki/Function_(mathematics)) that provides the probabilities of occurrence of different possible outcomes in an [experiment](https://en.wikipedia.org/wiki/Experiment_(probability_theory)). In more technical terms, the probability distribution is a description of a [random](https://en.wikipedia.org/wiki/Randomness) phenomenon in terms of the [probabilities](https://en.wikipedia.org/wiki/Probability) of [events](https://en.wikipedia.org/wiki/Event_(probability_theory)). For instance, if the [random variable](https://en.wikipedia.org/wiki/Random_variable) X is used to denote the outcome of a coin toss ("the experiment"), then the probability distribution of X would take the value 0.5 for X = heads, and 0.5 for X = tails (assuming the coin is fair). Examples of random phenomena can include the results of an [experiment](https://en.wikipedia.org/wiki/Experiment_(probability_theory)) or [survey](https://en.wikipedia.org/wiki/Survey_methodology).

A probability distribution is specified in terms of an underlying [sample space](https://en.wikipedia.org/wiki/Sample_space), which is the [set](https://en.wikipedia.org/wiki/Set_(mathematics)) of all possible [outcomes](https://en.wikipedia.org/wiki/Outcome_(probability)) of the random phenomenon being observed. The sample space may be the set of [real numbers](https://en.wikipedia.org/wiki/Real_numbers) or a set of [vectors](https://en.wikipedia.org/wiki/Vector_(mathematics)), or it may be a list of non-numerical values; for example, the sample space of a coin flip would be {heads, tails} .

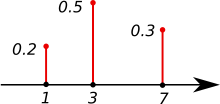
Probability distributions are generally divided into two classes. A discrete probability distribution (applicable to the scenarios where the set of possible outcomes is [discrete](https://en.wikipedia.org/wiki/Discrete_probability_distribution), such as a coin toss or a roll of dice) can be encoded by a discrete list of the probabilities of the outcomes, known as a [probability mass function](https://en.wikipedia.org/wiki/Probability_mass_function). On the other hand, a continuous probability distribution (applicable to the scenarios where the set of possible outcomes can take on values in a continuous range (e.g. real numbers), such as the temperature on a given day) is typically described by [probability density functions](https://en.wikipedia.org/wiki/Probability_density_function) (with the probability of any individual outcome actually being 0). The [normal distribution](https://en.wikipedia.org/wiki/Normal_distribution) is a commonly encountered continuous probability distribution. More complex experiments, such as those involving [stochastic processes](https://en.wikipedia.org/wiki/Stochastic_processes)defined in [continuous time](https://en.wikipedia.org/wiki/Continuous_time), may demand the use of more general [probability measures](https://en.wikipedia.org/wiki/Probability_measure).

**Probability mass function**

In [probability](https://en.wikipedia.org/wiki/Probability_theory) and [statistics](https://en.wikipedia.org/wiki/Statistics), a probability mass function (PMF) is a function that gives the probability that a [discrete random variable](https://en.wikipedia.org/wiki/Discrete_random_variable) is exactly equal to some value. The probability mass function is often the primary means of defining a [discrete probability distribution](https://en.wikipedia.org/wiki/Discrete_probability_distribution), and such functions exist for either [scalar](https://en.wikipedia.org/wiki/Scalar_variable) or [multivariate random variables](https://en.wikipedia.org/wiki/Multivariate_random_variable) whose [domain](https://en.wikipedia.org/wiki/Domain_of_a_function) is discrete.

A probability mass function differs from a [probability density function](https://en.wikipedia.org/wiki/Probability_density_function) (PDF) in that the latter is associated with continuous rather than discrete random variables; the values of the probability density function are not probabilities as such: a PDF must be [integrated](https://en.wikipedia.org/wiki/Integration_(mathematics)) over an interval to yield a probability.

The value of the random variable having the largest probability mass is called the [mode](https://en.wikipedia.org/wiki/Mode_(statistics))



**Descriptive statistics**

A descriptive statistic (in the [count noun](https://en.wikipedia.org/wiki/Count_noun) sense) is a [summary statistic](https://en.wikipedia.org/wiki/Summary_statistic) that quantitatively describes or summarizes features of a collection of [information](https://en.wikipedia.org/wiki/Information),[[1]](https://en.wikipedia.org/wiki/Descriptive_statistics#cite_note-1) while descriptive statistics in the [mass noun](https://en.wikipedia.org/wiki/Mass_noun)sense is the process of using and analyzing those statistics. Descriptive statistics is distinguished from [inferential statistics](https://en.wikipedia.org/wiki/Statistical_inference) (or inductive statistics), in that descriptive statistics aims to summarize a [sample](https://en.wikipedia.org/wiki/Sample_(statistics)), rather than use the data to learn about the [population](https://en.wikipedia.org/wiki/Statistical_population) that the sample of data is thought to represent. This generally means that descriptive statistics, unlike inferential statistics, is not developed on the basis of [probability theory](https://en.wikipedia.org/wiki/Probability_theory), and are frequently [nonparametric statistics](https://en.wikipedia.org/wiki/Nonparametric_statistics).[[2]](https://en.wikipedia.org/wiki/Descriptive_statistics#cite_note-2) Even when a data analysis draws its main conclusions using inferential statistics, descriptive statistics are generally also presented. For example, in papers reporting on human subjects, typically a table is included giving the overall [sample size](https://en.wikipedia.org/wiki/Sample_size), sample sizes in important subgroups (e.g., for each treatment or exposure group), and [demographic](https://en.wikipedia.org/wiki/Demographic" \o "Demographic)or clinical characteristics such as the [average](https://en.wikipedia.org/wiki/Average) age, the proportion of subjects of each sex, the proportion of subjects with related [comorbidities](https://en.wikipedia.org/wiki/Comorbidity), etc.

Some measures that are commonly used to describe a data set are measures of [central tendency](https://en.wikipedia.org/wiki/Central_tendency) and measures of variability or [dispersion](https://en.wikipedia.org/wiki/Statistical_dispersion). Measures of central tendency include the [mean](https://en.wikipedia.org/wiki/Mean), [median](https://en.wikipedia.org/wiki/Median) and [mode](https://en.wikipedia.org/wiki/Mode_(statistics)), while measures of variability include the [standard deviation](https://en.wikipedia.org/wiki/Standard_deviation) (or [variance](https://en.wikipedia.org/wiki/Variance)), the minimum and maximum values of the variables, [kurtosis](https://en.wikipedia.org/wiki/Kurtosis) and [skewness](https://en.wikipedia.org/wiki/Skewness).

**Central tendency**

In [statistics](https://en.wikipedia.org/wiki/Statistics), a central tendency (or measure of central tendency) is a central or typical value for a [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution). It may also be called a center or location of the distribution. Colloquially, measures of central tendency are often called [averages](https://en.wikipedia.org/wiki/Averages).

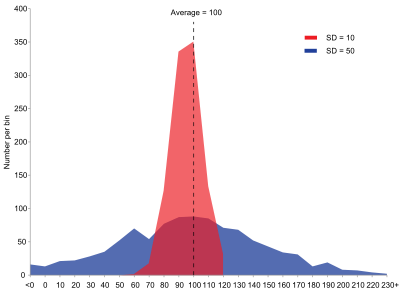
The most common measures of central tendency are the [arithmetic mean](https://en.wikipedia.org/wiki/Arithmetic_mean), the [median](https://en.wikipedia.org/wiki/Median) and the [mode](https://en.wikipedia.org/wiki/Mode_(statistics)). A central tendency can be calculated for either a finite set of values or for a theoretical distribution, such as the [normal distribution](https://en.wikipedia.org/wiki/Normal_distribution). Occasionally authors use central tendency to denote "the tendency of quantitative [data](https://en.wikipedia.org/wiki/Data) to cluster around some central value."

The central tendency of a distribution is typically contrasted with its [dispersion](https://en.wikipedia.org/wiki/Statistical_dispersion) or variability; dispersion and central tendency are the often characterized properties of distributions. Analysis may judge whether data has a strong or a weak central tendency based on its dispersion.

**Statistical dispersion**

In [statistics](https://en.wikipedia.org/wiki/Statistics), dispersion (also called variability, scatter, or spread) is the extent to which a [distribution](https://en.wikipedia.org/wiki/Probability_distribution) is stretched or squeezed. Common examples of measures of statistical dispersion are the [variance](https://en.wikipedia.org/wiki/Variance), [standard deviation](https://en.wikipedia.org/wiki/Standard_deviation), and [interquartile range](https://en.wikipedia.org/wiki/Interquartile_range).

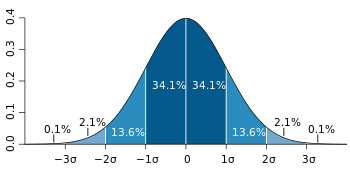
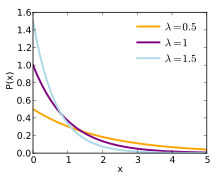
Dispersion is contrasted with location or [central tendency](https://en.wikipedia.org/wiki/Central_tendency), and together they are the most used properties of distributions.



**Shape of a probability distribution**

In [statistics](https://en.wikipedia.org/wiki/Statistics), the concept of the shape of a probability distribution arises in questions of finding an appropriate distribution to use to model the statistical properties of a population, given a sample from that population. The shape of a distribution may be considered either descriptively, using terms such as "J-shaped", or numerically, using quantitative measures such as [skewness](https://en.wikipedia.org/wiki/Skewness) and [kurtosis](https://en.wikipedia.org/wiki/Kurtosis).

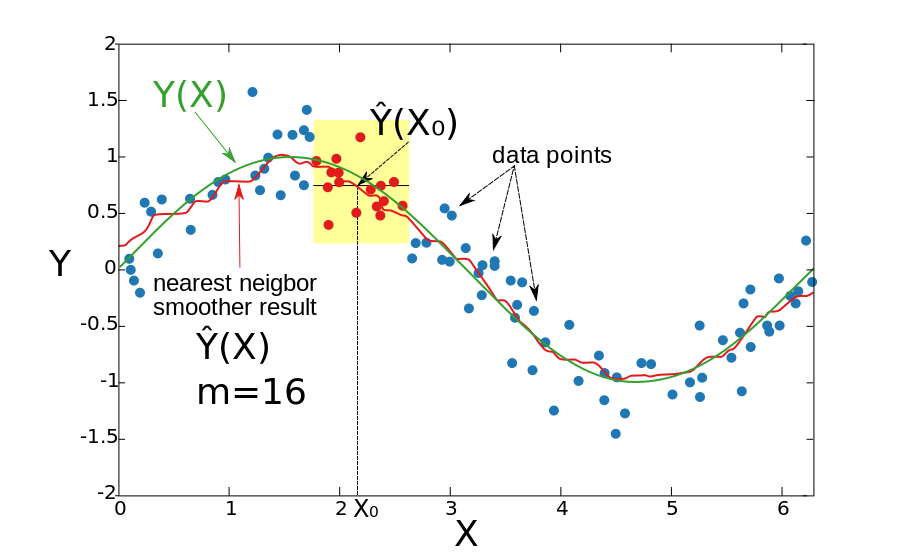
Considerations of the shape of a distribution arise in statistical [data analysis](https://en.wikipedia.org/wiki/Data_analysis), where simple quantitative descriptive statistics and plotting techniques such as [histograms](https://en.wikipedia.org/wiki/Histograms) can lead on to the selection of a particular family of distributions for modelling purposes.

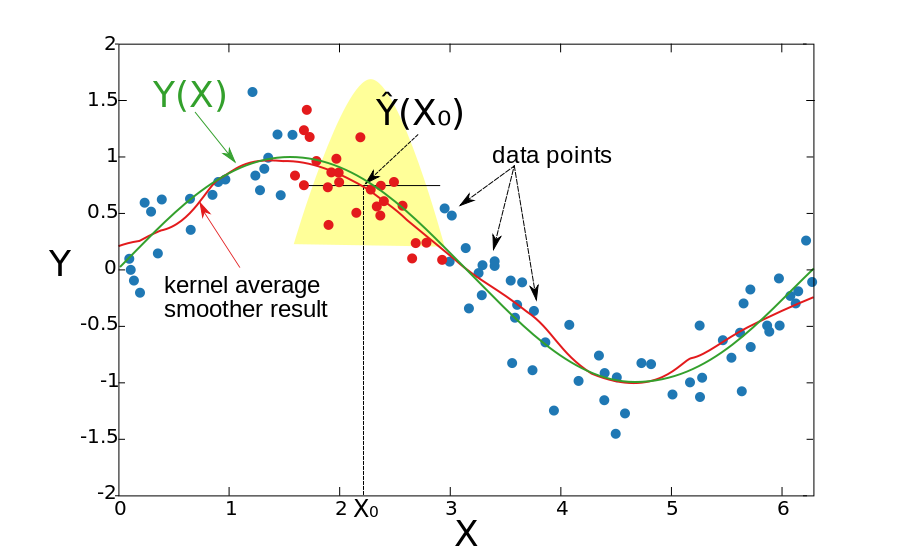
 

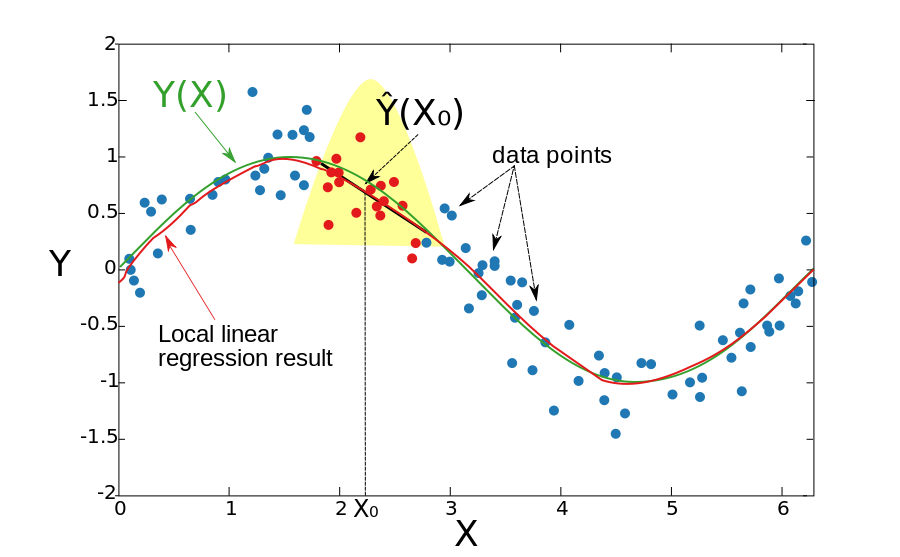
**Kernel smoother**

A kernel smoother is a [statistical](https://en.wikipedia.org/wiki/Statistics) technique to estimate a real valued [function](https://en.wikipedia.org/wiki/Function_(mathematics))  f : R ^p----🡪R as the weighted average of neighbouring observed data. The weight is defined by the kernel, such that closer points are given higher weights. The estimated function is smooth, and the level of smoothness is set by a single parameter.

This technique is most appropriate when the dimension of the predictor is low (p < 3), for example for data visualization.



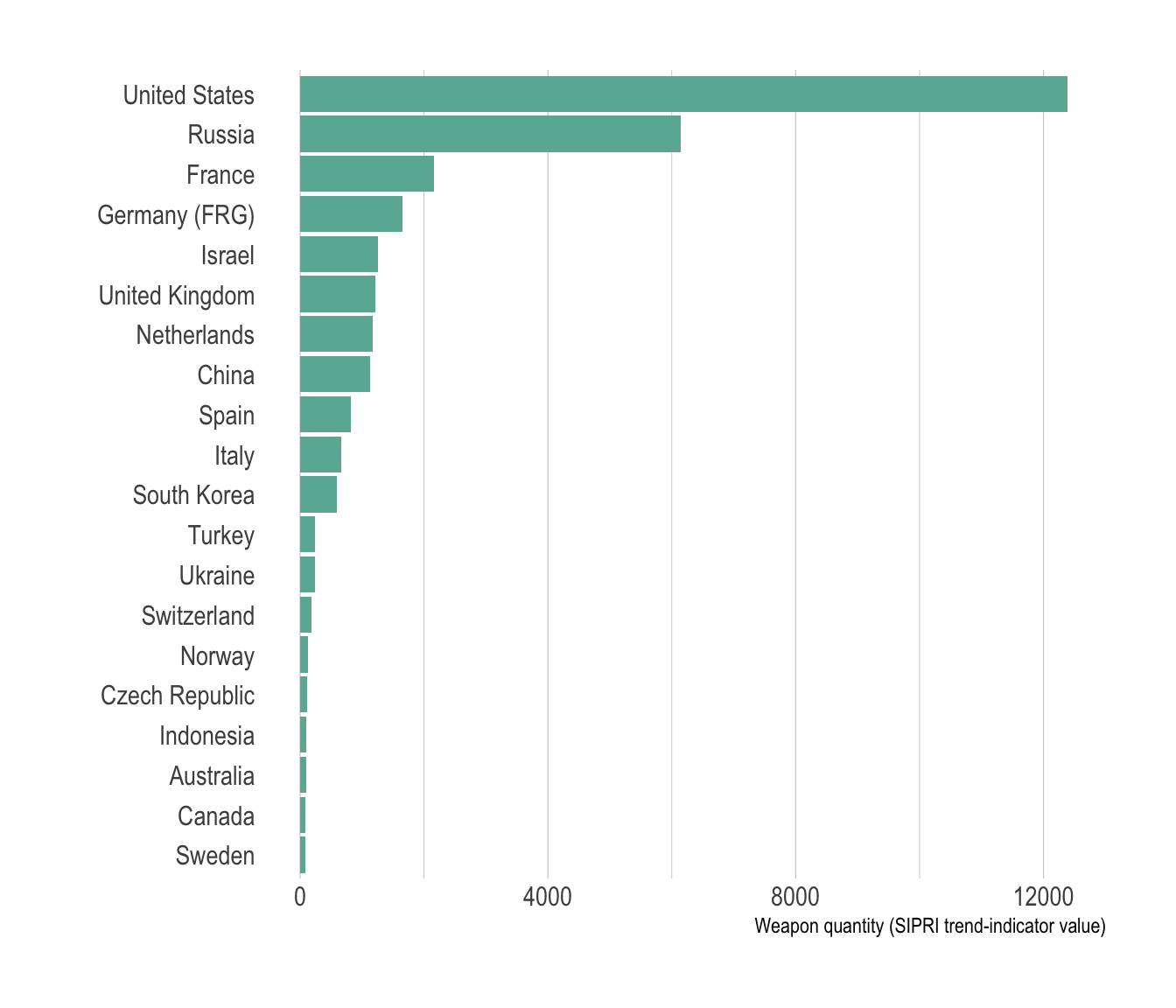




**Bar Plot**

A barplot (or barchart) is one of the most common types of graphic. It shows the relationship between a numeric and a categoric variable. Each entity of the categoric variable is represented as a bar. The size of the bar represents its numeric value.

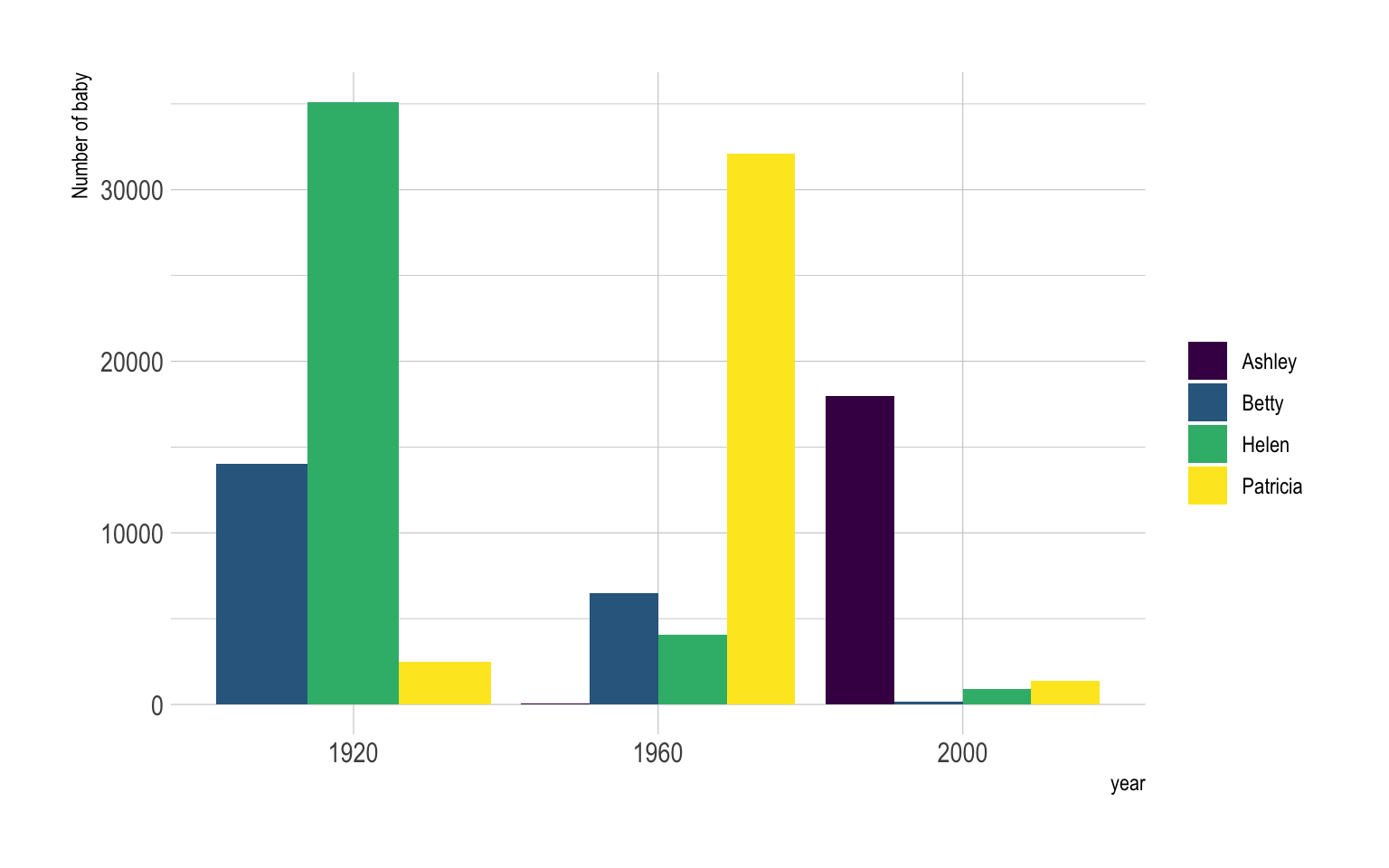
Here is an example showing the quantity of weapons exported by the top 20 largest exporters in 2017



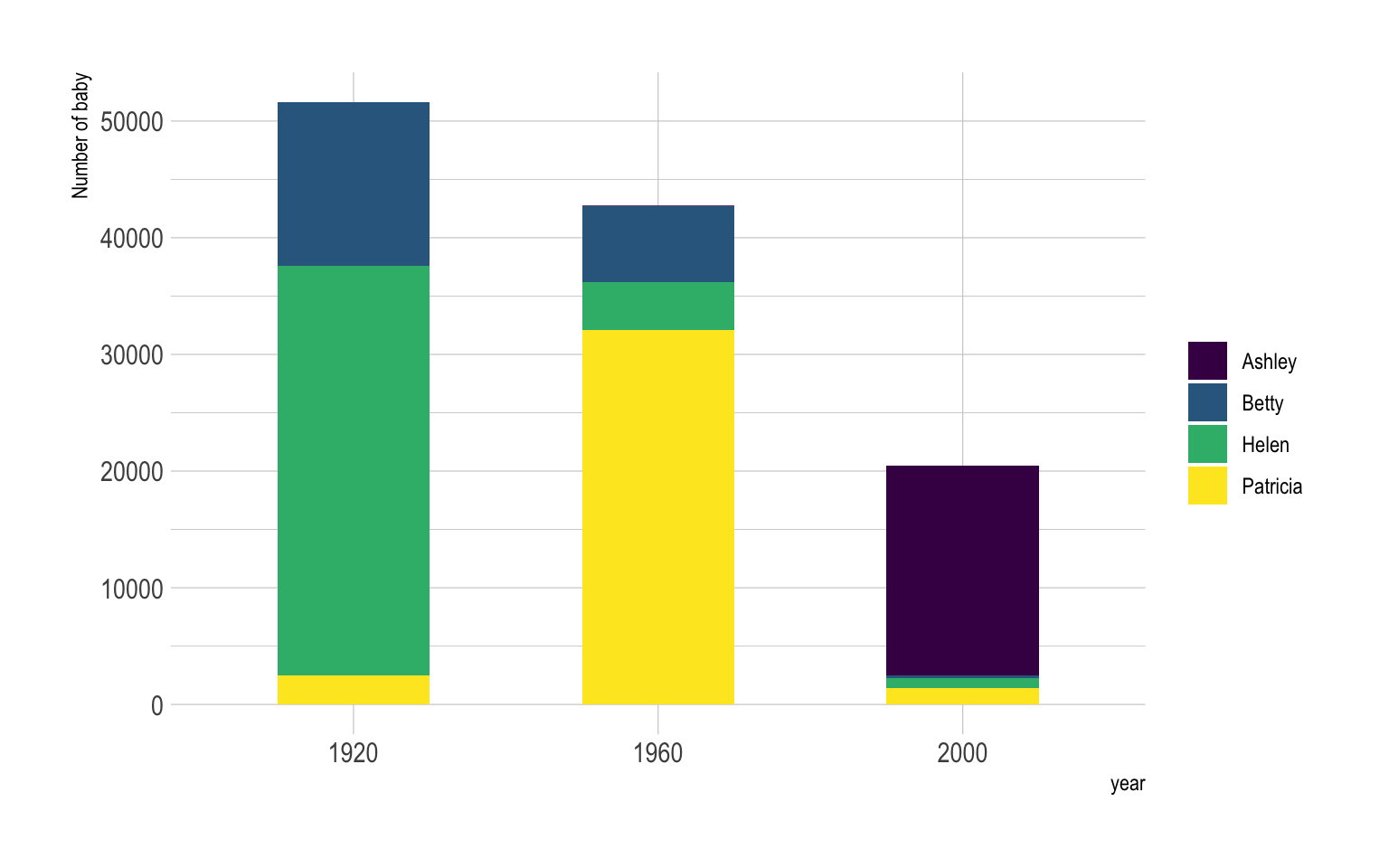
# What for

A barplot shows the relationship between a numeric and a categoric variable. In the previous graphic, each country is a level of the categoric variable, and the quantity of weapon sold is the numeric variable. An ordered barplot is a very good choice here since it displays both the ranking of countries and their specific value.

A barplot can also display values for several levels of grouping. In the following graphic, the number of given baby name is provided by name (level1) and per year (level2). With this kind of information it is possible to build a grouped barplot



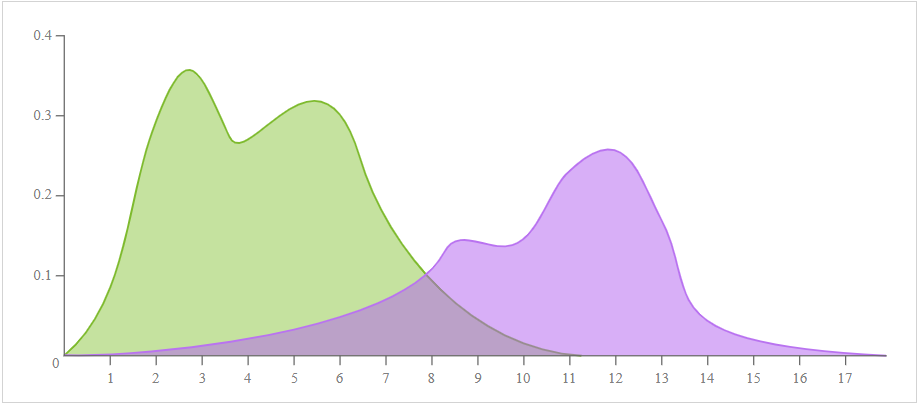
Instead of putting the bars one beside each other it is possible to stack them, resulting in a stacked bar plot



# Common mistakes

* Do not confound barchart with [histogram](https://www.data-to-viz.com/graph/histogram.html). A histogram has only a numeric variable as input and shows its distribution.
* [Order your bars](http://www.data-to-viz.com/caveat/order_data.html). If the levels of your categoric variable have no obvious order, order the bars following their values.
* Several values per group? [Don’t use a barplot](http://www.data-to-viz.com/caveat/error_bar.html). Even with error bars, it hides information and other type of graphic like [boxplot](https://www.data-to-viz.com/caveat/boxplot.html) or [violin](https://www.data-to-viz.com/graph/violin.html) are much more appropriate.

**Density Plot**

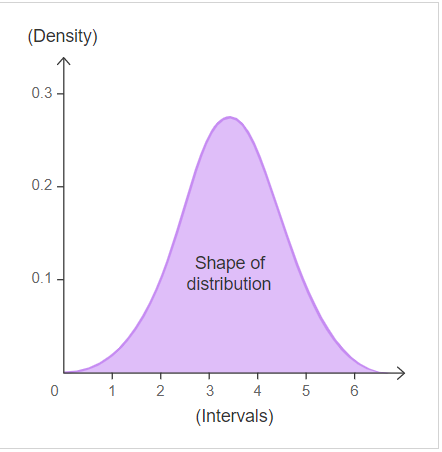


**Description**

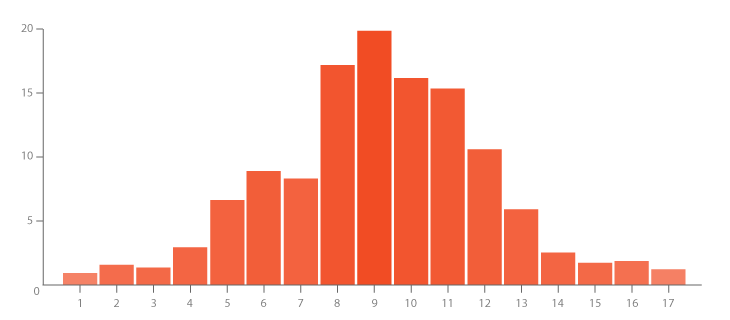
As known as *Kernel Density Plots, Density Trace Graph*.

A Density Plot visualises the distribution of data over a continuous interval or time period. This chart is a variation of a [Histogram](http://datavizcatalogue.com/methods/histogram.html) that uses [kernel smoothing](https://en.wikipedia.org/wiki/Kernel_smoother) to plot values, allowing for smoother distributions by smoothing out the noise. The peaks of a Density Plot help display where values are concentrated over the interval.

An advantage Density Plots have over Histograms is that they're better at determining the [distribution shape](https://en.wikipedia.org/wiki/Shape_of_the_distribution)because they're not affected by the number of bins used (each bar used in a typical histogram). A Histogram comprising of only 4 bins wouldn't produce a distinguishable enough shape of distribution as a 20-bin Histogram would. However, with Density Plots, this isn't an issue.



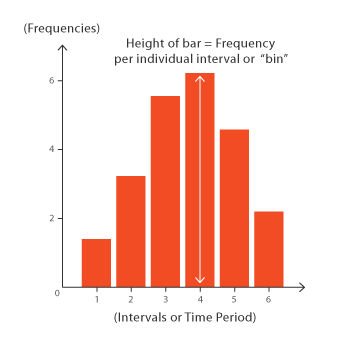
**Histogram**



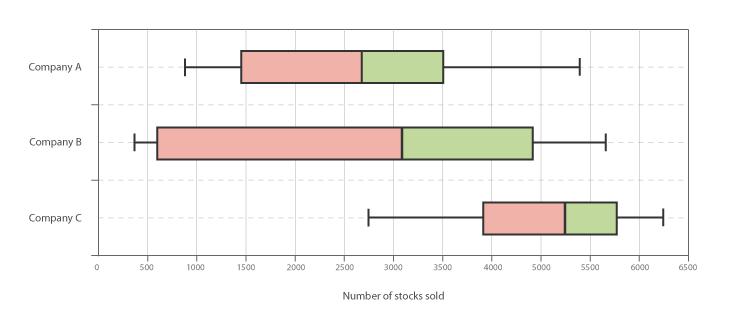
**Description**

A Histogram visualises the distribution of data over a continuous interval or certain time period. Each bar in a histogram represents the tabulated frequency at each interval/bin.

Histograms help give an estimate as to where values are concentrated, what the extremes are and whether there are any gaps or unusual values. They are also useful for giving a rough view of the probability distribution.



**Box and Whisker Plot**



**Description**

A Box and Whisker Plot (or Box Plot) is a convenient way of visually displaying the data distribution through their quartiles.

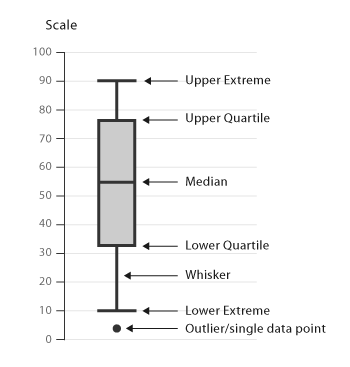
The lines extending parallel from the boxes are known as the “whiskers”, which are used to indicate variability outside the upper and lower quartiles. Outliers are sometimes plotted as individual dots that are in-line with whiskers. Box Plots can be drawn either vertically or horizontally.

Although Box Plots may seem primitive in comparison to a [Histogram](https://datavizcatalogue.com/methods/histogram.html) or [Density Plot](https://datavizcatalogue.com/methods/density_plot.html), they have the advantage of taking up less space, which is useful when comparing distributions between many groups or datasets.

Here are the types of observations one can make from viewing a Box Plot:

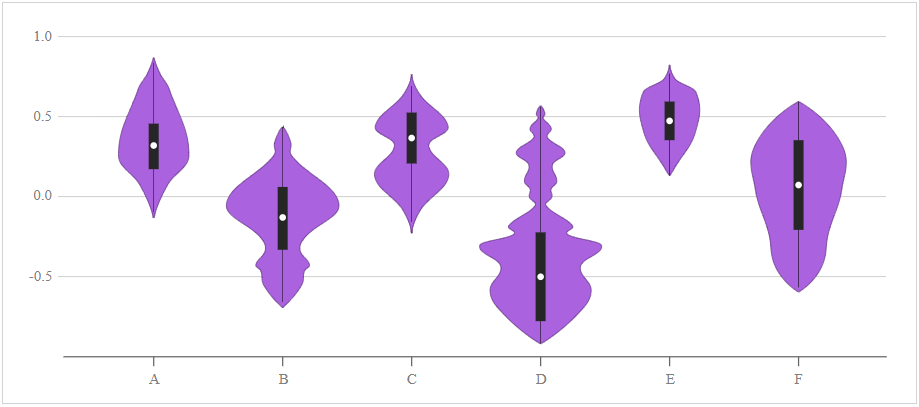
* What the key values are, such as: the average, median 25th percentile etc.
* If there are any outliers and what their values are.
* Is the data symmetrical.
* How tightly is the data grouped.
* If the data is skewed and if so, in what direction.

Two of the most commonly used variation of Box Plot are: variable-width Box Plots and notched Box Plots.



Reference : <https://datavizcatalogue.com/methods/box_plot.html>

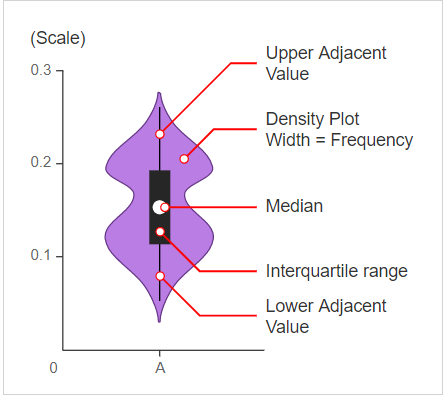
**Violin Plot**



**Description**

A Violin Plot is used to visualise the distribution of the data and its [probability density](https://en.wikipedia.org/wiki/Probability_density_function).

This chart is a combination of a [Box Plot](https://datavizcatalogue.com/methods/box_plot.html) and a [Density Plot](https://datavizcatalogue.com/methods/density_plot.html)that is rotated and placed on each side, to show the [distribution shape](https://en.wikipedia.org/wiki/Shape_of_the_distribution) of the data. The white dot in the middle is the median value and the thick black bar in the centre represents the interquartile range. The thin black line extended from it represents the upper (max) and lower (min) adjacent values in the data. Sometimes the graph marker is clipped from the end of this line.



Box Plots are limited in their display of the data, as their visual simplicity tends to hide significant details about how values in the data are distributed. For example, with Box Plots, you can't see if the distribution is [bimodal or multimodal](https://en.wikipedia.org/wiki/Multimodal_distribution). While Violin Plots display more information, they can be noisier than a Box Plot.

**Reference :** <https://datavizcatalogue.com/methods/violin_plot.html>

**Violin Plot VS Box Plot**

A violin plot is a method of plotting numeric data. It is similar to a [box plot](https://en.wikipedia.org/wiki/Box_plot), with the addition of a rotated [kernel density plot](https://en.wikipedia.org/wiki/Kernel_density_estimation) on each side

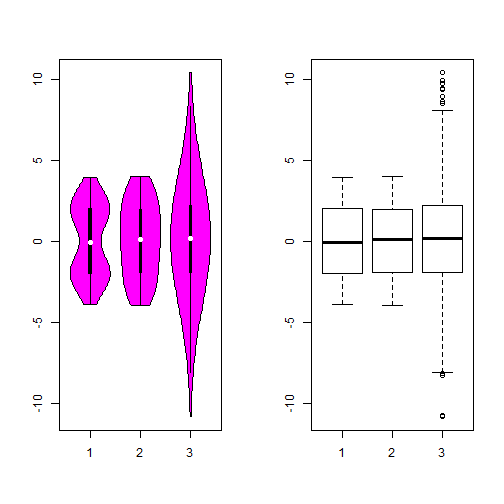
So a violin plot is more informative than a plain box plot. While a box plot only shows summary statistics such as mean/median and interquartile ranges, the violin plot shows the full distribution of the data. The difference is particularly useful when the data distribution is multimodal (more than one peak). In this case a violin plot shows the presence of different peaks, their position and relative amplitude.

Like box plots, violin plots are used to represent comparison of a variable distribution (or sample distribution) across different "categories" (for example, temperature distribution compared between day and night, or distribution of car prices compared across different car makers).

Bar plot would be better to compare grouped discrete categorical data over their size/frequency of occurrence not the frequency of distribution.

A violin plot can have multiple layers. For instance, the outer shape represents all possible results. The next layer inside might represent the values that occur 95% of the time. The next layer (if it exists) inside might represents the values that occur 50% of the time.

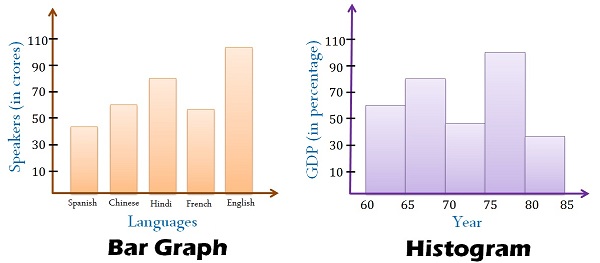
Although more informative than box plots, they are less popular. Because of their unpopularity, their meaning can be harder to grasp for many readers not familiar with the violin plot representation. In this case, a more accessible alternative can be plotting a series of stacked histograms or [kernel density distributions](https://en.wikipedia.org/wiki/Kernel_density_estimation).



So obviously, the violin plot can show more information than box plot. When we perform an exploratory analysis, nothing about the samples could be known. So the distribution of the samples can not be assumed to a normal distribution and usually when you get a big data, the normal distribution will show some out liars in box plot. The violin plot combines the box plot and the density trace, so it seems that the box plot may give the place to the violin plot

* the violin plot can’t show a better curve with small samples. In Hintze’s paper, he thought a smooth curve with at least 30 observations. But the box plot may stand for a smaller observations. Also the bandwidth need to be choosen carefully.
* the modification box plot could show the number of observations in the groups using the var width while the violin plot couldn’t. When we make some comparison between different groups, the violin plot will hide this information.
* Another problem is the notch in the box plot to compare the median. In the violin plot, we get a better understanding of distribution of violin plot but less with comparisone with ‘strong evidence

**Difference Between Histogram and Bar Graph**

[](https://keydifferences.com/wp-content/uploads/2016/04/bar-graph-vs-histogram.jpg)The fundamental difference between histogram and bar graph will help you to identify the two easily is that there are gaps between bars in a bar graph but in the histogram, the bars are adjacent to each other.

After the collection and verification of data, it needs to be compiled and displayed in such a way that it highlights the essential features clearly to the users. The statistical analysis can only be performed if it is properly presented. There are three modes of presentation of data i.e. textual presentation, tabular presentation, and diagrammatic presentation. The diagrammatic representation of data is one of the best and attractive way of presenting data as it caters both educated and uneducated section of the society.

**Bar Graph** and**Histogram** are the two ways to display data in the form of a diagram. As they both use bars to display data, people find it difficult to differentiate the two.

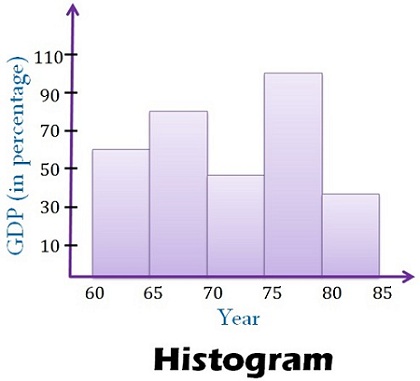
### **Comparison Chart**

| BASIS FOR COMPARISON | HISTOGRAM | BAR GRAPH |
| --- | --- | --- |
| Meaning | Histogram refers to a graphical representation, that displays data by way of bars to show the frequency of numerical data. | Bar graph is a pictorial representation of data that uses bars to compare different categories of data. |
| Indicates | Distribution of non-discrete variables | Comparison of discrete variables |
| Presents | Quantitative data | Categorical data |
| Spaces | Bars touch each other, hence there are no spaces between bars | Bars do not touch each other, hence there are spaces between bars. |
| Elements | Elements are grouped together, so that they are considered as ranges. | Elements are taken as individual entities. |
| Can bars be reordered? | No | Yes |
| Width of bars | Need not to be same | Same |

### **Definition of Histogram**

In statistics, Histogram is defined as a type of bar chart that is used to represent statistical information by way of bars to show the frequency distribution of continuous data. It indicates the number of observations which lie in-between the range of values, known as class or bin.

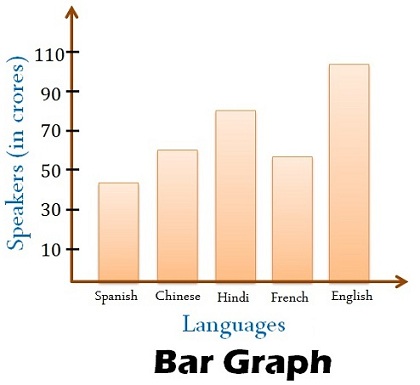
The first step, in the construction of histogram, is to take the observations and split them into logical series of intervals called bins. X-axis indicates, independent variables i.e. classes while the y-axis represents dependent variables i.e. occurrences. Rectangle blocks i.e. bars are depicted on the x-axis, whose area depends on the classes. See figure given below:

[](https://keydifferences.com/wp-content/uploads/2016/04/histogram.jpg)

### **Definition of Bar graph**

A bar graph is a chart that graphically represents the comparison between categories of data. It displays grouped data by way of parallel rectangular bars of equal width but varying the length. Each rectangular block indicates specific category and the length of the bars depends on the values they hold. The bars in a bar graph are presented in such a way that they do not touch each other, to indicate elements as separate entities.

Bar diagram can be horizontal or vertical, where a horizontal bar graph is used to display data varying over space whereas the vertical bar graph represents time series data. It contains two axis, where one axis represents the categories and the other axis shows the discrete values of the data. See figure given below:

**[](https://keydifferences.com/wp-content/uploads/2016/04/bargraph.jpg)**

**Key Differences Between Histogram and Bar graph**

The differences between histogram and bar graph can be drawn clearly on the following grounds:

1. Histogram refers to a graphical representation; that displays data by way of bars to show the frequency of numerical data. A bar graph is a pictorial representation of data that uses bars to compare different categories of data.
2. A histogram represents the frequency distribution of continuous variables. Conversely, a bar graph is a diagrammatic comparison of discrete variables.
3. Histogram presents numerical data whereas bar graph shows categorical data.
4. The histogram is drawn in such a way that there is no gap between the bars. On the other hand, there is proper spacing between bars in a bar graph that indicates discontinuity.
5. Items of the histogram are numbers, which are categorised together, to represent ranges of data. As opposed to the bar graph, items are considered as individual entities.
6. In the case of a bar graph, it is quite common to rearrange the blocks, from highest to lowest. But with histogram, this cannot be done, as they are shown in the sequences of classes.
7. The width of rectangular blocks in a histogram may or may not be same while the width of the bars in a  bar graph is always same.

### **Conclusion**

Prima facie both the two graphs seem alike, as both bar graph, and histogram has an x-axis and y-axis and uses vertical bars to display data. The height of the bars is decided by its relative frequency of the amount of data in the element. Moreover, skewness does matter in histogram but not in the case of a bar graph.

### **Histograms VS. Bar Charts**

### **Definition of Histograms and Bar Charts**

Bar charts and histograms can both be used to compare the sizes of different groups. A Bar chart is made up of bars plotted on a graph.

Histogram is a chart representing a frequency distribution; heights of the bars represent observed frequencies. In other words a histogram is a graphical display of data using bars of different heights. Usually, there is no space between adjacent Bars.

From the definition we can see their common point: The height of the column indicates the size of the group defined by the column label.

### **Difference between Histograms and Bar Charts**

**Bar Chart**

* The columns are positioned over a label that represents a categorical variable.
* The height of the column indicates the size of the group defined by the categories.

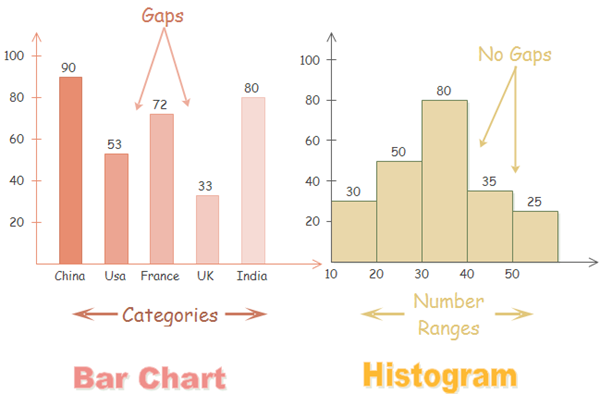
**Histogram**

* The columns are positioned over a label that represents a quantitative variable.
* The column label can be a single value or a range of values.

Here is the main difference between them. With bar charts, each column represents a group defined by a categorical variable; and with histograms, each column represents a group defined by a quantitative variable.

One indication of this distinction: it is always appropriate to talk about the skewness of a histogram; that is, the tendency of the observations to fall more on the low end or the high end of the X axis.

Differently, bar charts' X axis does not have a low end or a high end; because the labels on the X axis are categorical - not quantitative. Therefore, it is less appropriate to comment on the skewness of a bar chart. Refer to the following diagram to see a visual comparison between them.



### **Why Use Histograms and Bar Charts**

Like many other visuals, histograms and bar charts are gaining increasing popularity for the following benefits.

* Ease data analysis.
* Simplify complicated statistics and ideas.
* Add interest and fun to presentation (reports, essays and so on).
* Make presentation more compelling and comprehensible.
* Enhance effective information communication.