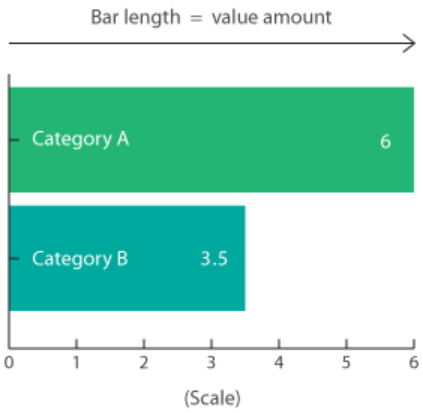
Bar Chart

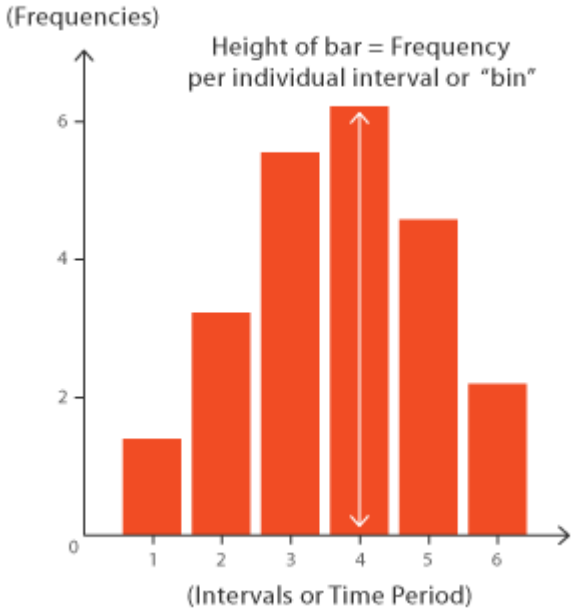
- The Bar Chart uses either horizontal or vertical bars (column chart) to show discrete, numerical comparisons across categories. One axis of the chart shows the specific categories being compared and the other axis represents a discrete value scale.



- One major flaw with Bar Charts is that labelling becomes problematic when there are a large number of bars.

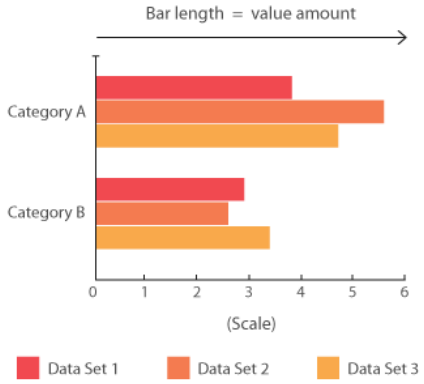
Histogram

- A Histogram shows the distribution of data over a continuous interval or certain time period.
- Intervals are plotted on x-axis and frequencies on y-axis
- 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
- They are also useful for giving a rough view of the probability distribution.



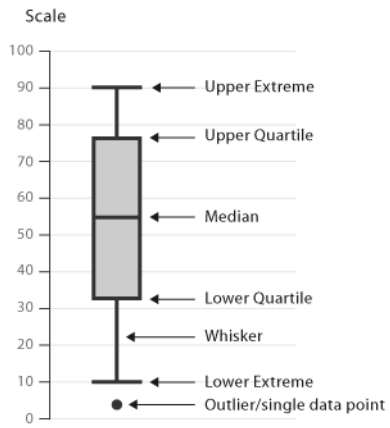
Multi-set Bar Chart

- This variation of a Bar Chart is used when two or more data series are plotted side-by-side and grouped together under categories, all on the same axis.
- Like a Bar Chart, the length of each bar is used to show discrete, numerical comparisons amongst categories. Each data series is assigned an individual colour or a varying shade of the same colour, in order to distinguish them
- The use of Multi-set Bar Charts is usually to compare grouped variables or categories to other groups with those same variables or category types
- The downside of Multi-set Bar Charts is that they become harder to read the more bars you have in one group



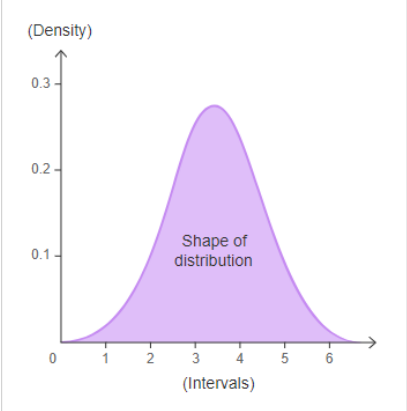
Box and Whisker Plot

- A box plot is mainly used to show the distributions of the data (1st quartile, median , 3rd quartile)
- Box plot consume less space when compared to Histogram and Density plot
- 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.
- 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.



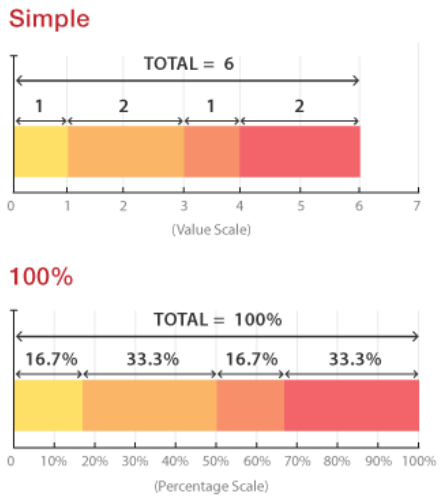
Density Plot

- A Density Plot visualises the distribution of data over a continuous interval or time period.
- This chart is a variation of a Histogram that uses kernel smoothing 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 because they're not affected by the number of bins used



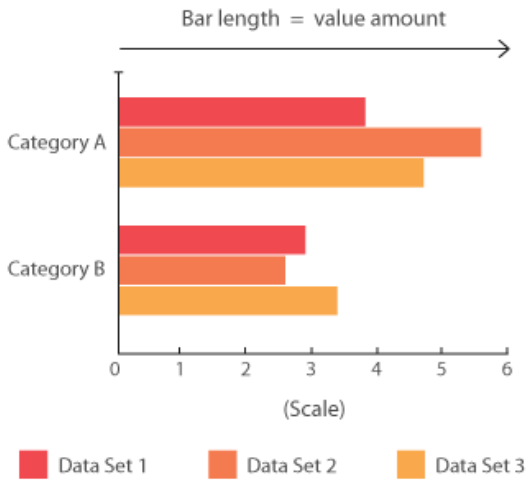
Stacked Bar Graph

- They are used to show how a larger category is divided into smaller categories and what the relationship of each part has on the total amount.
- **There are two types of Stacked Bar Graphs:**
- Simple Stacked Bar Graphs **place each value for the segment after the previous one. The total value of the bar is all the segment values added together.**
- 100% Stack Bar Graphs **show the percentage-of-the-whole of each group and are plotted by the percentage of each value to the total amount in each group.**



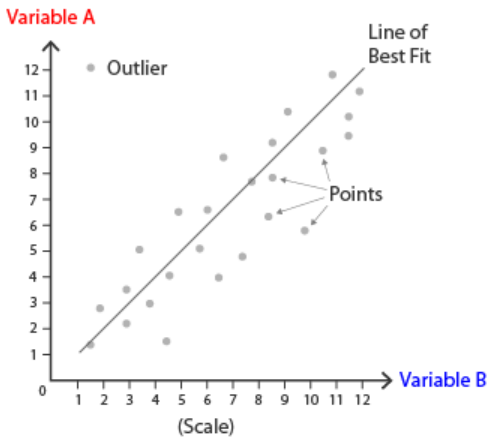
Multi-set Bar Chart

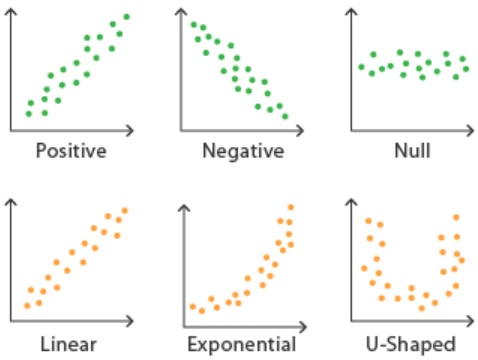
- This variation of a Bar Chart is used when two or more data series are plotted side-by-side and grouped together under categories, all on the same axis.
- Like a Bar Chart, the length of each bar is used to show discrete, numerical comparisons amongst categories.
- The use of Multi-set Bar Charts is usually to compare grouped variables or categories to other groups with those same variables or category types.



Scatterplot

- By displaying a variable in each axis, you can detect if a relationship or correlation between the two variables exists.
- Various types of correlation can be interpreted through the patterns displayed on Scatterplots. These are: positive(values increase together), negative (one value decreases as the other increases), null (no correlation), linear, exponential and U-shaped.
- The strength of the correlation can be determined by how closely packed the points are to each other on the graph. Points that end up far outside the general cluster of points are known as outliers.
- Scatterplots are ideal when you have paired numerical data and you want to see if one variable impacts the other.

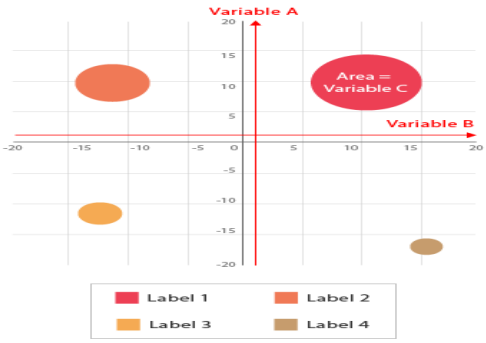




Correlation Strength:

Bubble Chart

- Unlike a Scatterplot, each point is assigned a label or category
- Each plotted point then represents a third variable by the area of its circle. Colours can also be used to distinguish between categories or used to represent an additional data variable.
- Bubble Charts are typically used to compare and show the relationships between categorised circles, by the use of positioning and proportions.
- The overall picture of Bubble Charts can be use to analyse for patterns/correlations.
- Too many bubbles can make the chart hard to read, so Bubble Charts have a limited data size capacity.

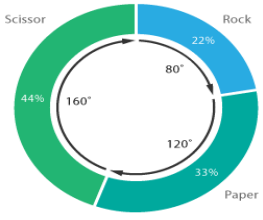


Circle Area = $\pi \times \text{Radius}^2$

Circle Diameter = (SQRT(Area / π)) x 2

Donut Chart

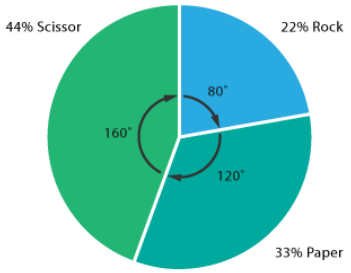
- A donut chart is essentially a Pie Chart with an area of the centre cut out.
- Pie Charts are sometimes criticised for focusing readers on the proportional areas of the slices to one another and to the chart as a whole. This makes it tricky to see the differences between slices, especially when you try to compare multiple Pie Charts together.
- A Donut Chart somewhat remedies this problem by de-emphasizing the use of the area. Instead, readers focus more on reading the length of the arcs, rather than comparing the proportions between slices.
- Also, Donut Charts are more space-efficient than Pie Charts because the blank space inside a Donut Chart can be used to display information inside it.



Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
2/9 = **22%**	3/9 = **33%**	4/9 = **44%**	100%
Degrees for each "donut slice"			
(2/9) x 360 = **80°**	(3/9) x 360 = **120°**	(4/9) x 360 = **160°**	360°

Pie Charts

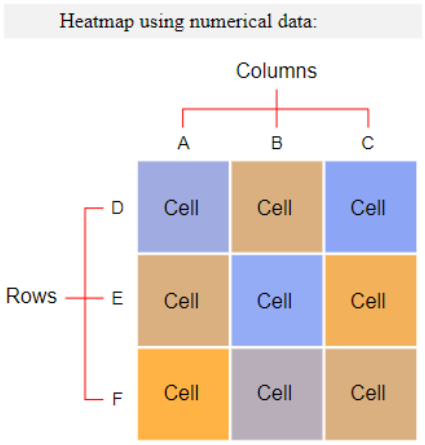
- Pie Charts help show proportions and percentages between categories, by dividing a circle into proportional segments.
- Each arc length represents a proportion of each category, while the full circle represents the total sum of all the data, equal to 100%.
- Pie Charts are ideal for giving the reader a quick idea of the proportional distribution of the data.
- However the major downsides to pie charts are:
- Maximum number of levels that can be represented is 8-10 levels. More levels makes the graph clumsy
- They are not great for making accurate comparisons between groups of Pie Charts. This being that it is harder to distinguish the size of items via area when it is for length.



Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
2/9=22%	3/9=33%	4/9=44%	100%
Degrees for each "pie slice"			
(2/9) x 360 = 80°	(3/9) x 360 = 120°	(4/9) x 360 = 160°	360°

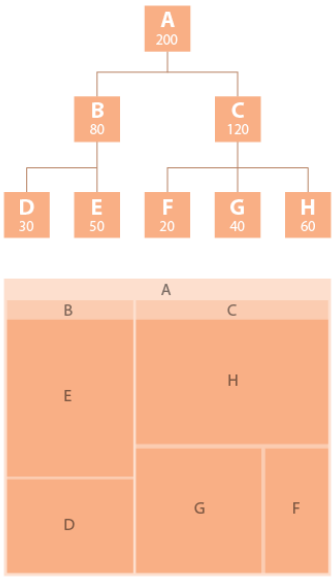
Heatmap (Matrix)

- Heatmaps are good for showing variance across multiple variables, revealing any patterns, displaying whether any variables are similar to each other, and for detecting if any correlations exist in-between them
- The cells contained within the table either contain colour-coded categorical data or numerical data, that is based on a colour scale.
- The data contained within a cell is based on the relationship between the two variables in the connecting row and column.
- An example of this would be to use a Heatmap to compare the temperature changes across the year in multiple cities, to see where's the hottest or coldest places. So the rows could list the cities to compare, the columns contain each month and the cells would contain the temperature values.



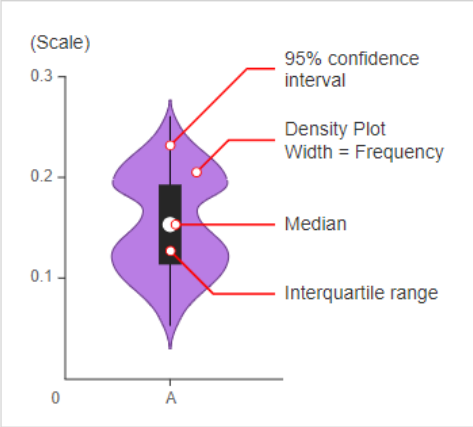
Treemap

- Way of visualising the hierarchical structure of a Tree Diagram while also displaying quantities for each category via area size.
- Each category is assigned a rectangle area with their subcategory rectangles nested inside of it.
- When a quantity is assigned to a category, its area size is displayed in proportion to that quantity and to the other quantities within the same parent category in a part-to-whole relationship.
- The way rectangles are divided and ordered into sub-rectangles is dependent on the tiling algorithm used.



Violin Plot

- A Violin Plot is used to visualise the distribution of the data and its probability density.
- This chart is a combination of a Box Plot and a Density Plotthat is rotated and placed on each side, to show the distribution shape of the data.
- The thick black bar in the centre represents the interquartile range, the thin black line extended from it represents the 95% confidence intervals, and the white dot is the median.
- For example, with Box Plots, you can't see if the distribution is bimodal or multimodal. While Violin Plots display more information, they can be noisier than a Box Plot



Word Cloud

- A visualisation method that displays how frequently words appear in a given body of text, by making the size of each word proportional to its frequency. All the words are then arranged in a cluster or cloud of words.
- Although being simple and easy to understand, Word Clouds have some major flaws:
- Long words are emphasised over short words.
- Words whose letters contain many ascenders and descenders may receive more attention.
- They're not great for analytical accuracy, so used more for aesthetic reasons instead.

Sunburst Diagram

- This type of visualisation shows hierarchy through a series of rings, that are sliced for each category node.
- Each ring corresponds to a level in the hierarchy, with the central circle representing the root node and the hierarchy moving outwards from it.