Statistics And Probability For Data Scientist And Data Analyst

Chapter 2: - Graphical Representation: -

Graphical Representation **is a way of analysing numerical data. It exhibits the relation between data, ideas, information and concepts in a diagram. It is easy to understand and it is one of the most important learning strategies. It always depends on the type of information in a particular domain. There are different types of graphical representation.**

❖ General Rules For Graphical Representation: -

1. The title of the graph should be appropriate that indicate the main subject of the presentation.
2. The unit of measurement in the graph should be mentioned properly.
3. Represent the data in an accurate manner with the help of proper scaling.
4. To understand the graphical representation, index the appropriate colours, shades, lines and designs in graphs.
5. Data sources should be included wherever it is necessary at the bottom of the graph.
6. The construction of graphs should be in a simple way that everyone can understand.
7. Always choose the correct size, fonts, colours, etc so that they should be a visual aid for presenting the information.

❖ Principle Of Graphical Representation: -

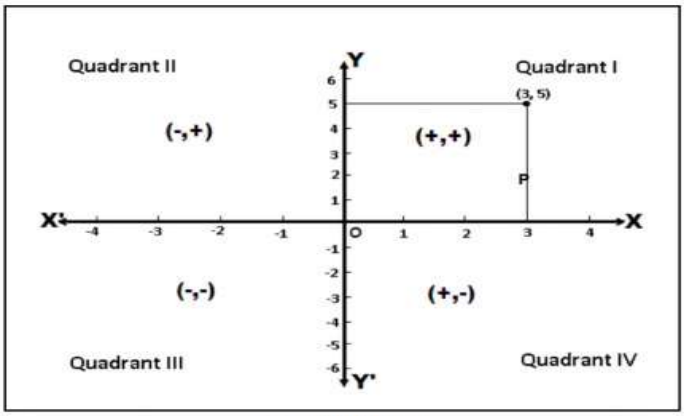
The principles of graphical representation of data are algebraic and are applicable to all types of it. In every graph, a data point is displayed using variables representing coordinate axes of the coordinate plane.

Coordinates are of two types, horizontal axes, and vertical axes, and they are placed perpendicular to each other with both the coordinates intersecting at the origin. The horizontal axis is known as X-axis and the vertical axis is known as Y-axis.

When the X-axis and Y-axis are intersected with each other at the origin it divides the coordinate plane into four parts which are called Quadrant I, Quadrant II, Quadrant III and Quadrant IV.

Following are some principles regarding the graphical representation of data given below:

- In the coordinate system, any point above the horizontal axis is taken as a positive value while any point below the horizontal axis is taken as a negative value.
- In the coordinate system, any point to the right of the vertical axis is taken as a positive value and any point to the left of the vertical axis is taken as a negative value.



Graphical Plane

Statistics And Probability For Data Scientist And Data Analyst

❖ **Frequency Distribution: -**

A **frequency distribution** describes the number of observations for each possible value of a variable.

Example: - In the 2022 Winter Olympics, Team USA won 25 medals. This frequency table gives the medals' values (gold, silver, and bronze) and frequencies:

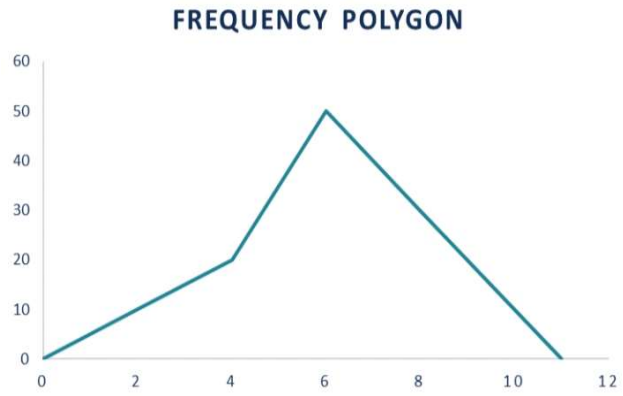
Frequency table of the 25 medals Team USA won during the 2022 Winter Olympics

Medal	Frequency
Gold	8
Silver	10
Bronze	7

➢ **Types Of Frequency Distribution: -** There are four types of frequency distributions.

1. **Ungrouped frequency distribution: -** The number of observations of each **value** of a variable. You can use this type of frequency distribution for categorical variables.
2. **Grouped frequency distribution: -** The number of observations of each **class interval** of a variable. Class intervals are ordered groupings of a variable's values. You can use this type of frequency distribution for quantitative variables.
3. **Relative frequency distribution: -** The proportion of observations of each value or class interval of a variable. You can use this type of frequency distribution for **any type of variable** when you're more interested in **comparing frequencies** than the actual number of observations.
4. **Cumulative frequency distribution: -** The sum of the frequencies less than or equal to each value or class interval of a variable. You can use this type of frequency distribution for **ordinal or quantitative variables** when you want to understand **how often observations fall below certain values**.

❖ **Frequency Polygon: -** A frequency polygon is a type of line graph where a line segment curves to join the midpoints of all the class intervals.



Statistics And Probability For Data Scientist And Data Analyst

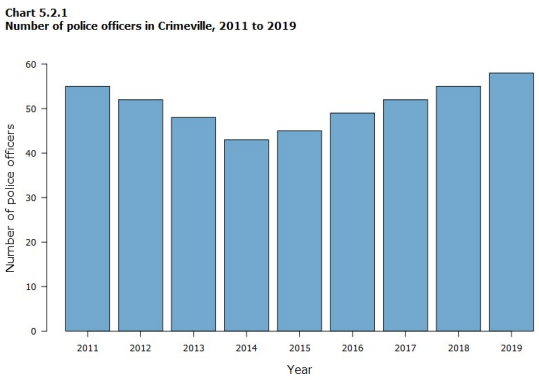
❖ Distinguish Between Frequency Polygon And Histogram: -

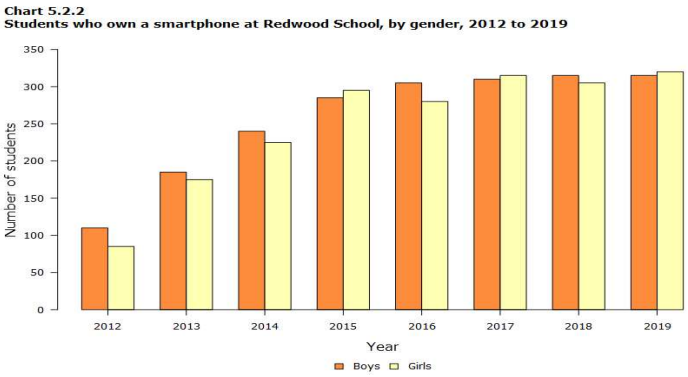
Frequency Polygons	Histograms
A frequency polygon graph is a curve that is depicted by a line segment.	A histogram is a graph that depicts data through rectangular-shaped bars with no spaces between them.
In a frequency polygon graph, the midpoint of the frequencies is used.	In a histogram, the frequencies are evenly spread over the class intervals.
The accurate points in a frequency polygon graph represent the data of the particular class interval.	The height of the bars in a histogram only depicts the quantity of the data.
Comparison of data is visually more accurate in a frequency polygon graph.	Comparison of data is not visually appealing in a histogram graph.

❖ Statistical Graph: -

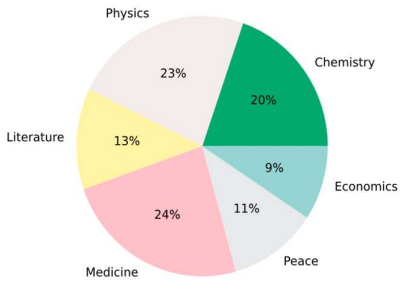
Bar Graph: - A **bar chart** or **bar graph** is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent.

Bar charts usually present categorical variables, discrete variables or continuous variables grouped in class intervals.





Pie Chart: - A pie chart, sometimes called a circle chart, Pie graphs are used to show the distribution of qualitative (categorical) data. It shows the **frequency** or **relative frequency** of values in the data.



Statistics And Probability For Data Scientist And Data Analyst

Scatter Plot: - **Scatter plots** are the graphs that present the relationship between two variables in a data-set.

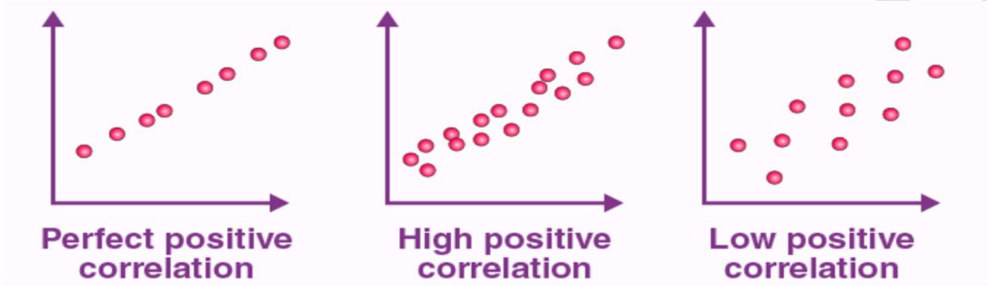
- **Types of Scatter plot Correlation: -**

1]Positive Correlation: - When the points in the graph are rising, moving from left to right, then the scatter plot shows a positive correlation.

Positive correlation: - Which represents a perfectly straight line

High Positive: - All points are nearby

Low Positive: - When all the points are scattered

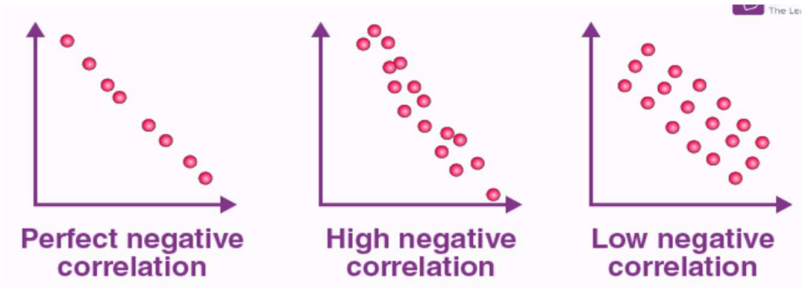


2]Negative Correlation: - When the points in the scatter graph fall while moving left to right, then it is called a negative correlation.

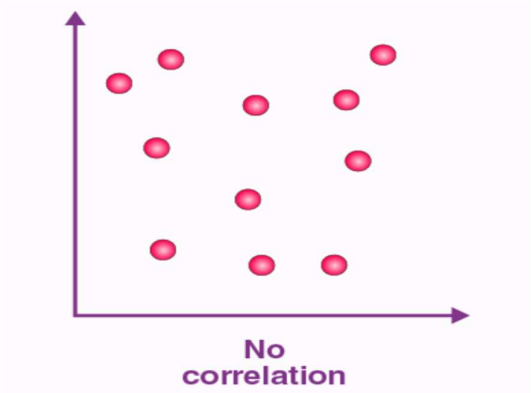
Perfect Negative: -Which form almost a straight line.

High Negative: - When points are near to one another.

Low Negative: - When points are in scattered form.



3]No Correlation: -When the points are scattered all over the graph and it is difficult to conclude whether the values are increasing or decreasing, then there is no correlation between the variables.



Histogram: - The histogram provides a visual representation of the distribution of the data, showing the number of observations that fall within each bin. A histogram is a widely used in quantitative (numerical) data.

Chart 5.7.1
Distribution of salaries of the employees of ABC Corporation