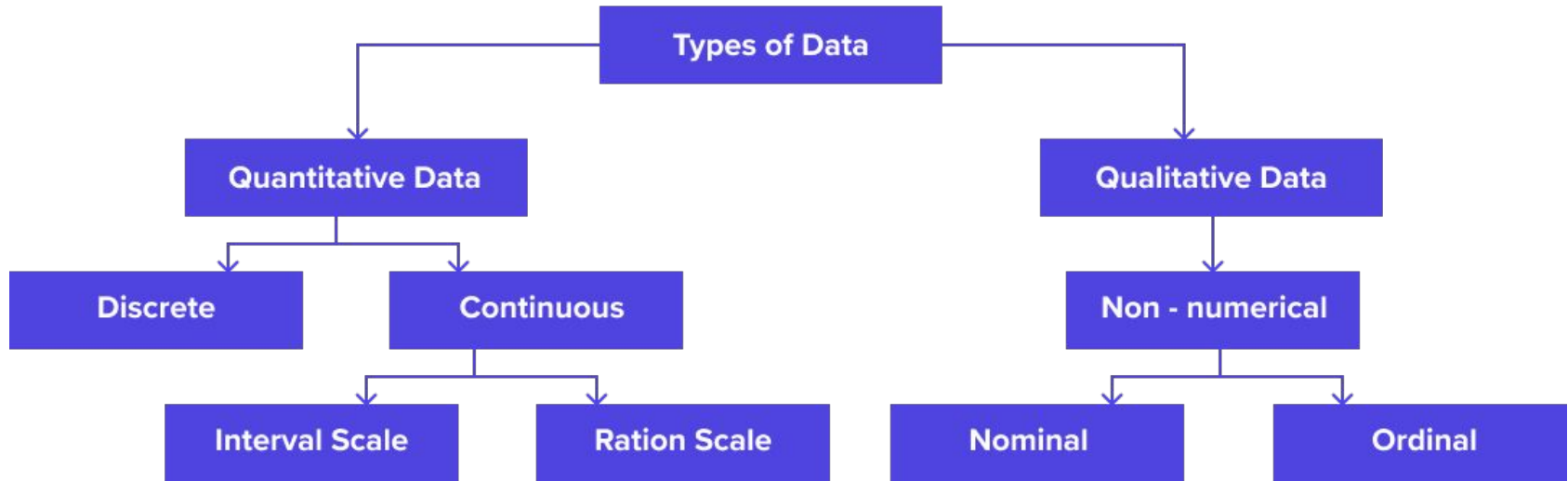# What is statistics?

Statistics is a branch of mathematics involving data collection, analysis, interpretation, and presentation. It takes data and turns it into insights we can use to make decisions.

Real World Data → Statistics → Decision

Relevel
by Unacademy
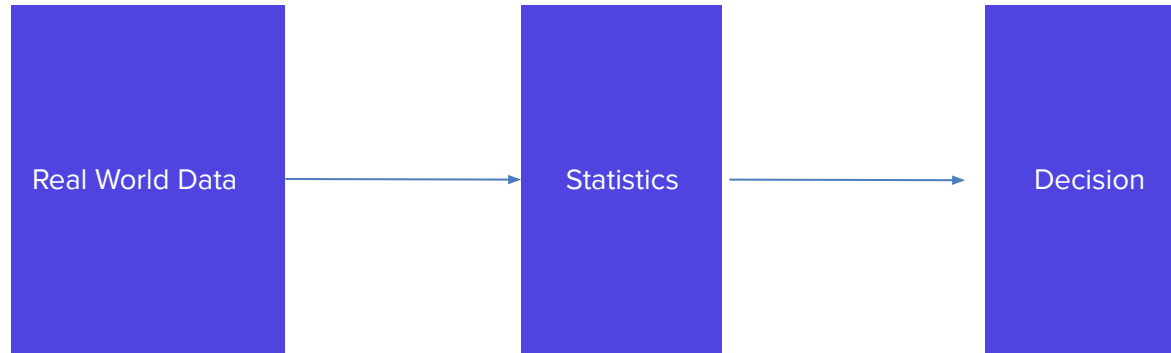
Various Types of Data

Before deep-diving into understanding data, it is important to understand various kinds of data types.

Data Types and Scales of Measurement

```
                          Types of Data
                 ┌─────────────┴─────────────┐
          Quantitative Data            Qualitative Data
          ┌──────┴──────┐                    │
      Discrete      Continuous          Non - numerical
                    ┌────┴────┐          ┌────┴────┐
              Interval Scale  Ration Scale  Nominal  Ordinal
```

Categorical Data

Characteristics are represented through categorical data. As a result, it can indicate things like a person's gender or language. Numeral values can also be assigned to categorical data (Example: 1 for female and 0 for male, 1 for True and 0 for False). Categorical data falls under non-numerical data (qualitative) from the previous slide.

There are primarily two kinds of categorical data:

1 . **Nominal Data:** Nominal values represent discrete units and label variables that do not have quantitative value. Just think of them as "labels ". Nominal data has no order. Therefore if you change the order of its values, the meaning would not change.

2. **Ordinal Data:** Ordinal values represent discretely ordered units. It is almost the same as nominal data, except that here ordering matters.

Examples of Nominal Data

Are you married?

► Yes

► No

What languages do you speak?

► English

► French

► German

► Spanish

The first example has two categories. The other example has four categories.

Examples of Ordinal Data

What is your Educational Background?

► Elementary

► High School

► Undergraduate

► Graduate

Note that the difference between Elementary and High School is different than the difference between High School and College.

This is the main limitation of ordinal data; the differences between the values is not really known.

Numerical Data (Quantitative Data)

Information that may be measured is known as numerical data. Although other sorts of data can also exist in numerical form, it is always collected in that format. The number of persons that visited the cinema over a month is an example of numerical data.

There are two types of numerical data:
1. Discrete
2. Continuous

Numerical Data – Discrete data

Discrete Data: Data is said to be discrete if its values are distinct and separate. In other words: Data is said to be discrete data if the data can only take on certain values. This data type **cannot be measured, but it can be counted.**

It represents information that can be categorized into a classification. An example is the number of tails in 10 coin flips.

You can check by asking the following two questions whether you are dealing with discrete data or not:

- Can you count it and,

- Can it be divided up into smaller parts?

Numerical Data – Continuous Data

Since continuous data represent measurements, its values can be measured but not counted.

A person's height is something you could represent on a real number line by utilising intervals.

Why are Data types important?

Because some data types cannot be employed with statistical methods, datatypes are a crucial notion. Continuous data must be analysed differently from categorical data, or the result could be incorrect. Knowing the different sorts of data you are working with, will help you select the best analytical strategy.
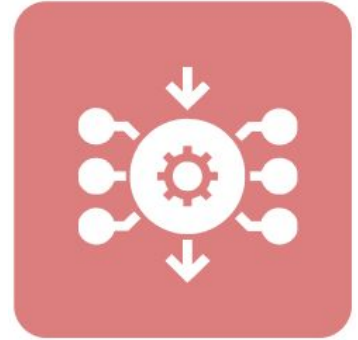
Practice Problem

Classify the data into nominal, ordinal, discrete, and continuous data:

1. Name of students in a call

2. Size of shirts in a shop (S, M, L , XL, XXL)

3. Outcome of a roll of dice

4. Height of students

Practice Problem - Solution

1. Name of students in a call **- Nominal**

2. Size of shirts in a shop (S, M, L , XL, XXL) **- Ordinal**

3. Outcome of a roll of dice **- Discrete**

4. Height of students **- Continuous**

Introduction to Graphs

In this section, we will cover :

1. Intro to graphs

2. Various types of graphs

3. Importance of charts

4. Plotting on google sheets

5. Plotting a scatter plot

6. Correlation

Introduction to Graphs

A graphical representation in statistics is the visual presentation of data as a diagram or graph. The data is graphically represented in a chart by symbols like bars in a bar chart, lines in a line chart, or slices in a pie chart. It accurately depicts the collection of data. When data is provided, decisions can be made considerably more effectively. The following are a few examples of the numerous graphical representations:

- Line Graphs

- Bar Graphs

- Histograms

- Pie charts

- Scatterplot

Relevel
by Unacademy

Importance of graphs

A graph's function is to see a set of data quickly. Understanding and spotting patterns and trends in the ever-increasing flow of data requires graphic visual representations of information. Drawing line segments or charting the points in the graphs comprises a wide range of ways to clarify, understand, and evaluate the data.

Relevel
by Unacademy
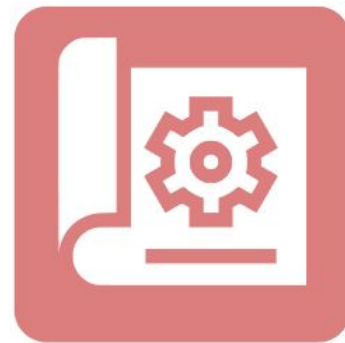
Scatterplots

The graphs that show the association between two variables in a data set are called **scatter plots**. It displays data points either on a **Cartesian system** or a two-dimensional plane. The X-axis is used to represent the independent variable or attribute, while the Y-axis is used to plot the dependent variable. These diagrams or graphs are frequently used to describe these plots.

This chart type is generally used for numerical data – preferably discrete.
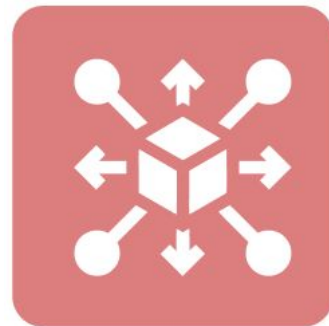
Developing a Scatterplot in Google Spreadsheet

Refer to the dataset in the tab "demo":

Download the file from the link below

https://docs.google.com/spreadsheets/d/1p6Pp5n4Um6bSOurGmP7AUGOs9aly0gzZFQyqkllfGf8/edit?usp=sharing

Relevel
by Unacademy

Developing a Scatterplot in Google Spreadsheet

1. Go to Insert and select chart

2. Select the dataset

Developing a Scatterplot in Google Spreadsheet
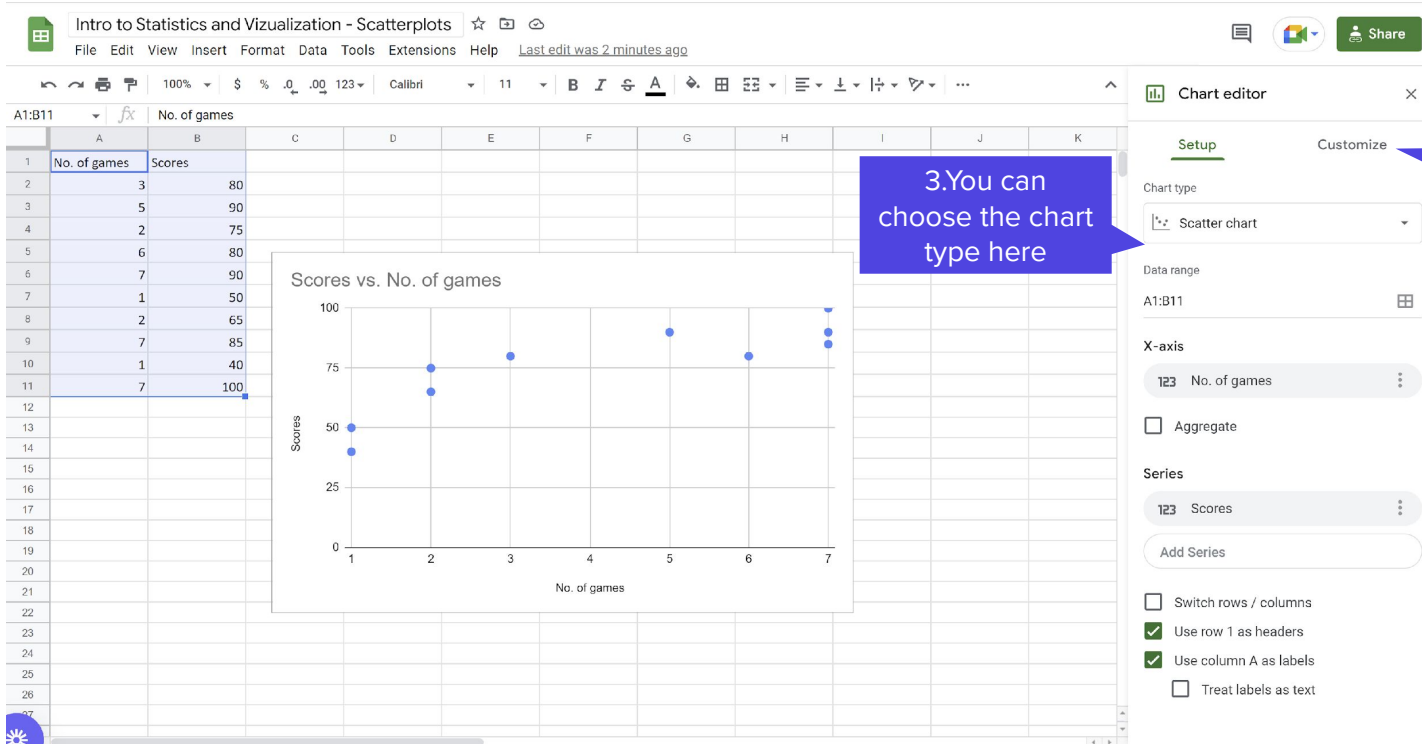
3.You can choose the chart type here

4.We can customize the chart here

Developing a Scatterplot in Google Spreadsheet

#150DaysofPurpose Relevel by Unacademy
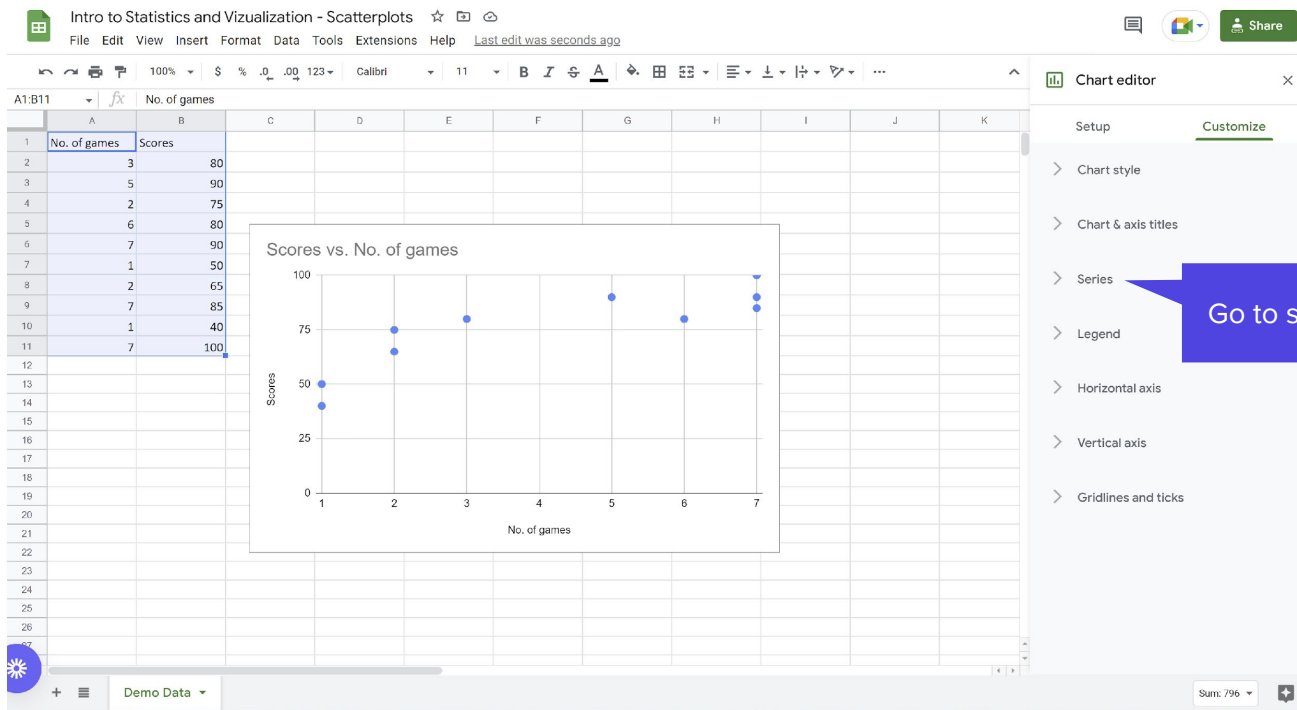
Developing a Scatterplot in Google Spreadsheet
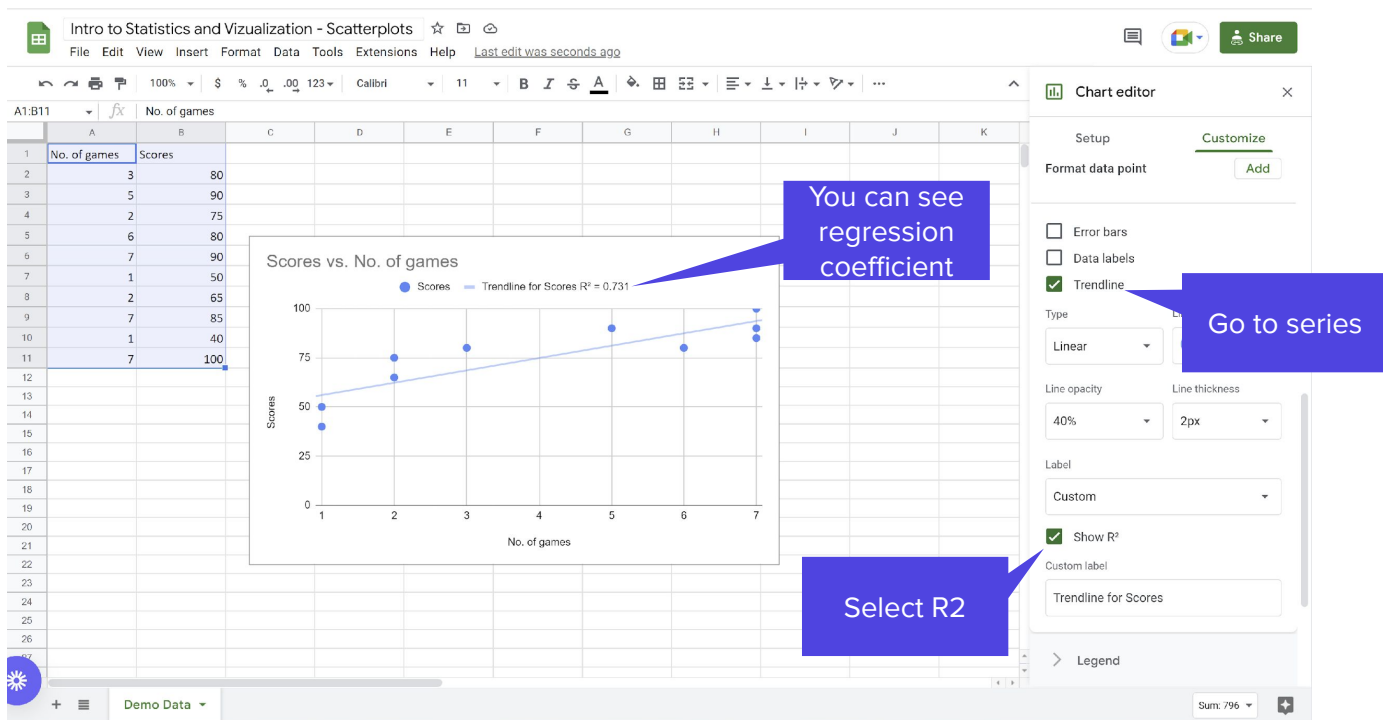
Practice Problem in Scatter Plot

Refer to the dataset in the tab "Practice Problem":

Download the dataset from the link below

https://docs.google.com/spreadsheets/d/1p6Pp5n4Um6bSOurGmP7AUGOs9aly0gzZFQyqkllfGf8/edit?usp=sharing

Practice Problem in Scatter Plot - Solution

Refer to the dataset in the tab "Practice Problem - solution":

Download the dataset fron the link below.

https://docs.google.com/spreadsheets/d/1p6Pp5n4Um6bSOurGmP7AUGOs9aly0gzZFQyqkllfG

f8/edit?usp=sharing

Scatter plot Correlation

Correlation is a statistical measure of the relationship between the two variables' relative movements. The points will fall along a line or curve if the variables are correlated. The better the correlation, the closer the points will touch the line. This cause examination tool is considered one of the seven essential quality tools.

Types of correlation

The scatter plot explains the correlation between two attributes or variables. It represents how closely the two variables are connected. There can be three such situations to see the relation between the two variables –

1. Positive Correlation   Direct proportionality
2. Negative Correlation   Inverse Proportionality
3. No Correlation

Positive Correlation

- The scatter plot displays a positive association when the graph's points rise and travel from left to right. It indicates that one variable's values are rising about another. The three categories of positive correlation are as follows:

- **The Perfect Positive represents a perfectly straight line.**

- **High Positive – Nearby for all points**

- **When all the points are dispersed, the positive is low.**

Perfect positive correlation

High positive correlation

Low positive correlation
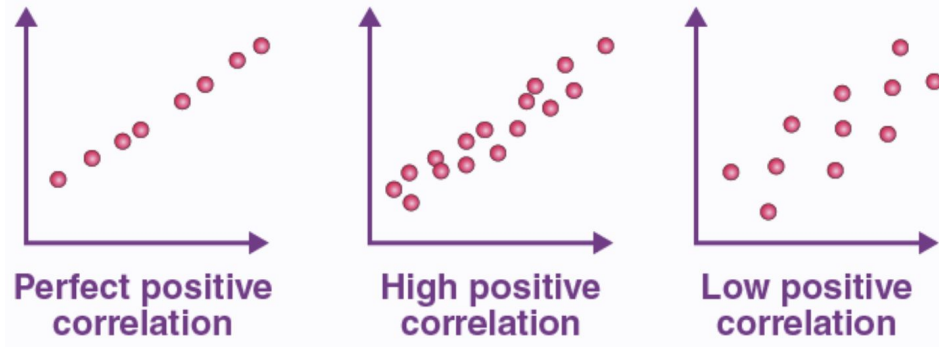
Negative Correlation

- A negative correlation occurs when the points in the scatter graph go from left to right. It indicates that one variable's values are falling about another. These too come in three varieties:

- **They almost create a straight line in the perfect negative.**

- **High Negative: When two points are nearby.**

- **When points are dispersed, the value is low negative.**

Perfect negative correlation   High negative correlation   Low negative correlation
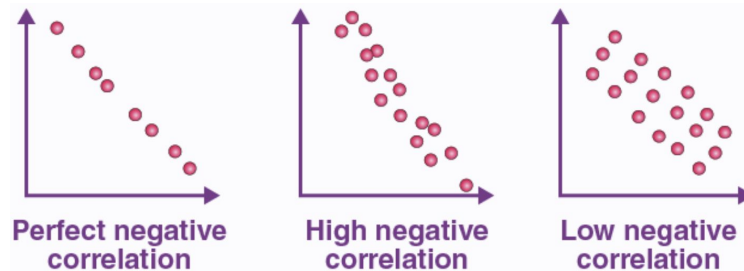
No Correlation

There is no association between the variables if the points are dispersed around the graph, and it is impossible to determine whether the values are rising or falling.

No correlation
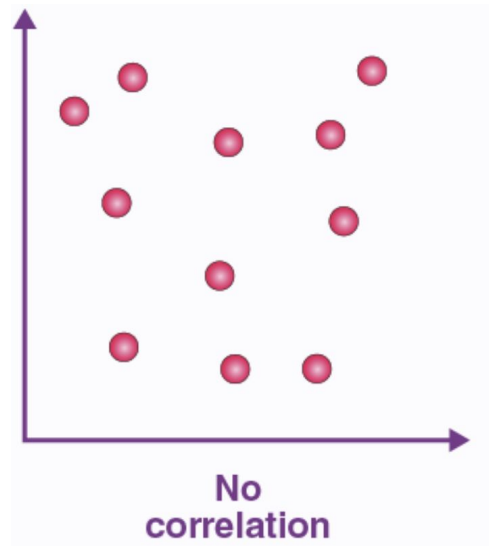
Frequency Distribution and Histogram

In this section, we will cover :

1. Intro to Frequency distribution

2. Building Frequency distribution in google sheets

3. Intro to histogram

4. Plotting histogram in google sheets

Frequency Distribution

A frequency distribution table is a thorough representation of the arrangement of the raw data for a quantitative variable in statistics. This table displays the frequency distributions of different values of a variable. We can make two frequency distribution tables:

(i) Discrete frequency distribution

(ii) Continuous frequency distribution (Grouped frequency distribution)

Frequency Distribution – Discrete (Example)

Discrete Frequency Calculation is similar to using the COUNT function in SQL. This helps calculate the

number of times an instance repeats during the period of measurement.

In a quiz, the marks obtained by 20 students out of 30 are given as:

12,15,15,29,30,21,30,30,15,17,19,15,20,20,16,21,23,24,23,21

Frequency Distribution – Discrete (Example)

Marks obtained in quiz	Number of students (Frequency)
12	1
15	4
16	1
17	1
19	1
20	2
21	3
23	2
24	1
29	1
30	3
Total	20

Refer to the 'Discrete' tab:

https://docs.google.com/spreadsheets/d/1u-Gx-4lgrs5os7eVb61UQY69y_V_Ef7TP2Q8W-astEo/edit?usp=sharing

Relevel
by Unacademy

Frequency Distribution – Continuous (Example)

The heights of 50 students, measured to the nearest centimetres, have been found to be as follows:

161, 150, 154, 165, 168, 161, 154, 162, 150, 151, 162, 164, 171, 165, 158, 154, 156, 172, 160, 170, 153, 159, 161,

170, 162, 165, 166, 168, 165, 164, 154, 152, 153, 156, 158, 162, 160, 161, 173, 166, 161, 159, 162, 167, 168,

159, 158, 153, 154, 159

Frequency Distribution – Continuous (Example)

Frequency bin	Frequency
150-154	12
155-159	9
160-164	14
165-169	10
170-174	5

Refer to the 'Continuous' tab:

https://docs.google.com/spreadsheets/d/1u-Gx-4lgrs5os7eVb61UQY69y_V_Ef7TP2Q8W-astEo/edit?usp=sharing

Practice Problem : Frequency Distribution – Discrete

In a quiz, the marks obtained by 33 students out of 100 are given as:

69, 48, 84, 58, 48, 73, 83, 48, 66, 58, 84,66, 64, 71, 64, 66, 69, 66, 83, 66, 69, 71, 81, 71, 73, 69, 66, 66,

64, 58, 64, 69, 69

Refer to the 'Practice - Discrete' tab:

https://docs.google.com/spreadsheets/d/1u-Gx-4lgrs5os7eVb61UQY69y_V_Ef7TP2Q8W-astEo/edit?usp=sharing

#150DaysofPurpose

Relevel
by Unacademy

Practice Problem : Frequency Distribution – Discrete

Marks obtained in quiz	Number of students (Frequency)
12	1
15	4
16	1
17	1
19	1
20	2
21	3
23	2
24	1
29	1
30	3
Total	20

Refer to the 'Practice – Discrete Solution' tab:

https://docs.google.com/spreadsheets/d/1u-Gx-4lgrs5os7eVb61UQY69y_V_Ef7TP2Q8W-astEo/edit?usp=sharing

#150DaysofPurpose

Relevel
by Unacademy

Practice Problem : Frequency Distribution – Continuous

The following are the marks (out of 100) of 60 students in mathematics.

16, 13, 5, 80, 86, 7, 51, 48, 24, 56, 70, 19, 61, 17, 16, 36, 34, 42, 34, 35, 72, 55, 75, 31, 52, 28,72,

97, 74, 45, 62, 68, 86, 35, 85, 36, 81, 75, 55, 26, 95, 31, 7, 78, 92, 62, 52, 56, 15, 63,25, 36, 54,

44, 47, 27, 72, 17, 4, 30

Refer to the 'Practice - Continuous' tab:

https://docs.google.com/spreadsheets/d/1u-Gx-4lgrs5os7eVb61UQY69y_V_Ef7TP2Q8W-astEo/edit?usp=sharing

Practice Problem : Frequency Distribution – Continuous

Frequency bin	Frequency
0-9	4
10-19	7
20-29	5
30-39	10
40-49	5
50-59	8
60-69	5
70-79	8
80-89	5
90-99	3

Refer to the 'Practice – Discrete Solution' tab:

https://docs.google.com/spreadsheets/d/1iuDkGgsLISO5tfYkoXELhzWYhTdQwgxMIywcDJQ0uRU/edit?usp=sharing

Practice Problem : Frequency Distribution – Cumulative

The following are the marks (out of 100) of 60 students in mathematics.

16, 13, 5, 80, 86, 7, 51, 48, 24, 56, 70, 19, 61, 17, 16, 36, 34, 42, 34, 35, 72, 55, 75, 31, 52, 28,72, 97, 74,

45, 62, 68, 86, 35, 85, 36, 81, 75, 55, 26, 95, 31, 7, 78, 92, 62, 52, 56, 15, 63,25, 36, 54, 44, 47, 27, 72,

17, 4, 30.

After plotting frequency distribution, calculate the cumulative sum.

Refer to the 'Practice - Continuous' tab:

https://docs.google.com/spreadsheets/d/1u-Gx-4lgrs5os7eVb61UQY69y_V_Ef7TP2Q8W-astEo/edit?usp=sharing

Practice Problem : Frequency Distribution – Continuous

Frequency bin	Frequency	Cum Frequency
0-9	4	4
10-19	7	11
20-29	5	16
30-39	10	26
40-49	5	31
50-59	8	39
60-69	5	44
70-79	8	52
80-89	5	57
90-99	3	60

Refer to the 'Practice Cumulative – Continuous Solution' tab:

https://docs.google.com/spreadsheets/d/1u-Gx-4lgrs5os7eVb61UQY69y_V_Ef7TP2Q8W-astEo/edit?usp=sharing

Histogram

A histogram is a pictorial distribution of frequency distribution. It represents a range of outcomes in column formation along the x-axis. In the same histogram, the number count or multiple occurrences in the data for each column is represented by the y-axis.

Histograms help visualize how readings are distributed and understand readings that occur more frequently than others.

- The horizontal axis displays the number range.
- The vertical axis (frequency) represents the amount of data present in each range.

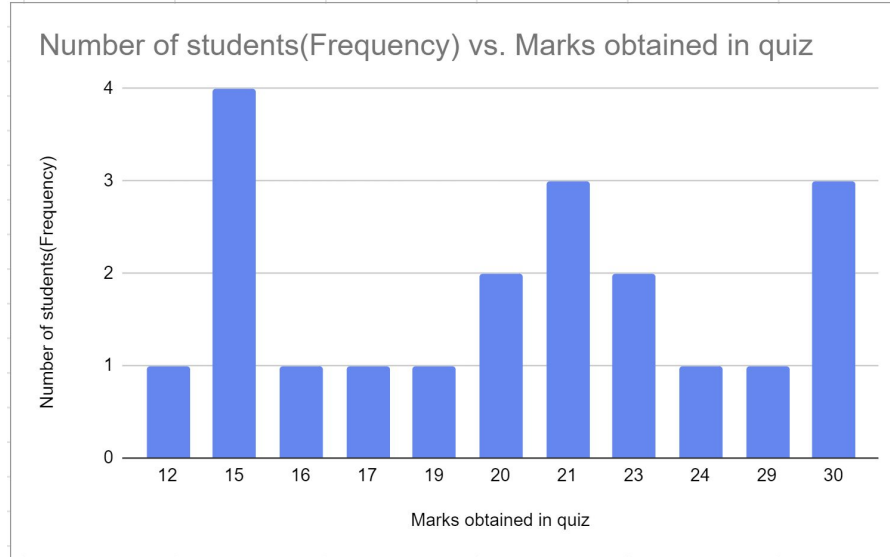The best way to plot is: to go to Google Sheet -> insert -> charts->histogram.

Histogram – Discrete (Example)

In a quiz, the marks obtained by 20 students out of 30 are given as:

12,15,15,29,30,21,30,30,15,17,19,15,20,20,16,21,23,24,23,21

Histogram – Discrete (Example)

Number of students(Frequency) vs. Marks obtained in quiz

Refer to the 'Histogram-Discrete' tab:

https://docs.google.com/spreadsheets/d/18TeuIVZ_-ClCQ4rnjl-v2xGqvsrXJ0uB5OF1dx25O_M/edit?usp=sharing
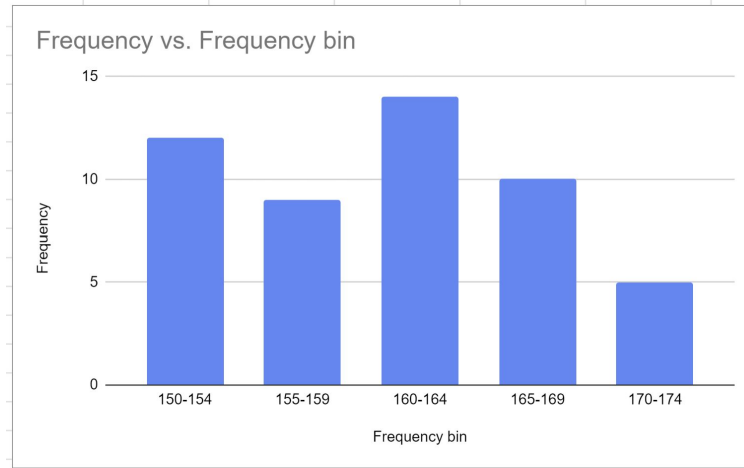
Histogram – Continuous (Example)

The heights of 50 students, measured to the nearest centimetres, have been found to be as follows:

161, 150, 154, 165, 168, 161, 154, 162, 150, 151, 162, 164, 171, 165, 158, 154, 156, 172, 160, 170, 153, 159, 161, 170, 162, 165, 166, 168, 165, 164, 154, 152, 153, 156, 158, 162, 160, 161, 173, 166, 161, 159, 162, 167, 168, 159, 158, 153, 154, 159

Histogram – Continuous (Example)

Refer to the 'Histogram - Continuous' tab: Refer to the 'Histogram-Discrete' tab:

https://docs.google.com/spreadsheets/d/18TeuIVZ_-ClCQ4rnjl-v2xGqvsrXJ0uB5OF1dx25O_M/edit?usp=sharing
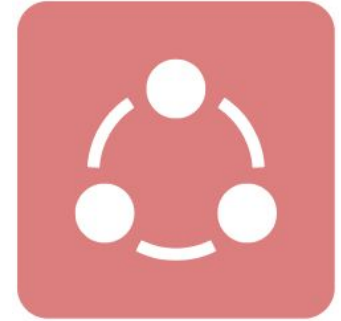
Practice Problem : Histogram – Discrete

In a quiz, the marks obtained by 33 students out of 100 are given as:

69, 48, 84, 58, 48, 73, 83, 48, 66, 58, 84,66, 64, 71, 64, 66, 69, 66, 83, 66, 69, 71, 81, 71, 73, 69, 66, 66, 64, 58, 64, 69, 69

Refer to the 'Histogram Practice - Discrete' tab:

https://docs.google.com/spreadsheets/d/18TeuIVZ_-ClCQ4rnjl-v2xGqvsrXJ0uB5OF1dx25O_M/edit?usp=sharing

#150DaysofPurpose

Relevel
by Unacademy

Practice Problem : Histogram – Discrete

Refer to the 'Histogram Practice – Discrete Solution' tab:

https://docs.google.com/spreadsheets/d/18TeuIVZ_-ClCQ4rnjl-v2xGqvsrXJ0uB5OF1dx25O_M/edit?usp=sharing

Number of students(Frequency) vs. Marks obtained in quiz

Relevel
by Unacademy
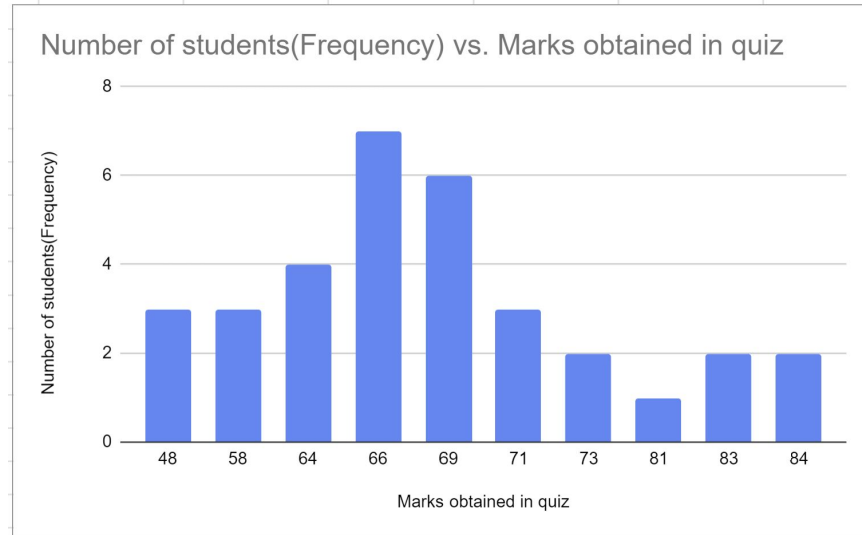
Practice Problem : Histogram – Continuous

The following are the marks (out of 100) of 60 students in mathematics.

16, 13, 5, 80, 86, 7, 51, 48, 24, 56, 70, 19, 61, 17, 16, 36, 34, 42, 34, 35, 72, 55, 75, 31, 52, 28,72, 97, 74, 45, 62, 68, 86, 35, 85, 36, 81, 75, 55, 26, 95, 31, 7, 78, 92, 62, 52, 56, 15, 63,25, 36, 54, 44, 47, 27, 72, 17, 4, 30

Refer to the 'Histogram Practice - Continuous' tab:

https://docs.google.com/spreadsheets/d/18TeuIVZ_-ClCQ4rnjl-v2xGqvsrXJ0uB5OF1dx25O_M/edit?usp=sharing

Relevel
by Unacademy

Practice Problem : Histogram – Continuous

Refer to the 'Histogram Practice – Continuous Solution' tab:

https://docs.google.com/spreadsheets/d/18TeuIVZ_-ClCQ4rnjl-v2xGqvsrXJ0uB5OF1dx25O_M/edit?usp=sharing
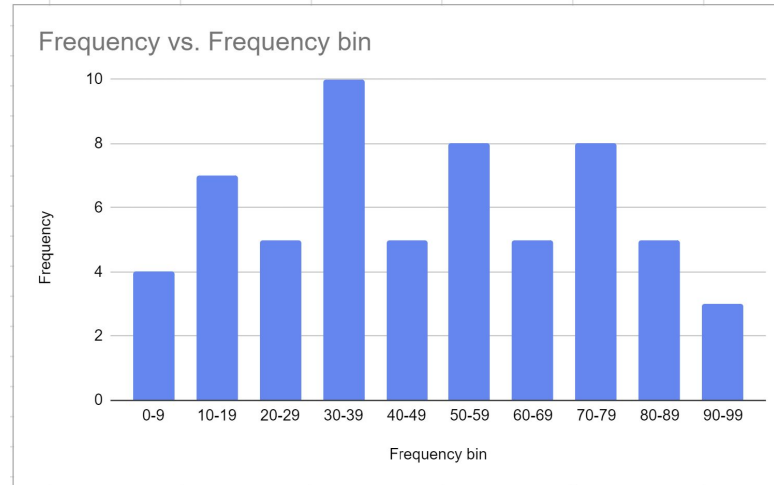
Introduction to Pie-Chart

In this section, we will cover :

1. Intro to pie-charts
2. SOP to pie-charts
3. Plotting pie-charts on google sheets

Relevel
by Unacademy

Pie Chart

A form of chart called a pie chart uses a circular graph to display data visually. It is one of the most popular graphs used to depict data because it uses the characteristics of spheres, circles, and angular data to represent information from the actual world. A pie chart is **circular**, with the pie representing the total amount of data and the slices representing its component elements, which are separately recorded.

An essential sort of data representation is the pie chart. It has several segments and sectors, where each pie chart section and sector represents a particular percentage of the total (percentage). 360° is equal to the total of all the data.

The pie's overall value is always 100%.

Pie Charts are useful in cases where we need to compare parts of a whole - showing percentages of types of customers, percentage of revenue from different products, and profits from different countries. We identify the total value and plot the contribution of each segment to that whole.

Relevel
by Unacademy

SOP for Pie Chart

To work out with the percentage for a pie chart, follow the steps given below:

- Categorize the data

- Calculate the total

- Divide the categories

- Convert into percentages

- Finally, calculate the degrees

Therefore, the pie chart formula is given as:

(Given Data/Total value of Data) × 360°

Alternatively, we can just use a google sheet. Go to insert, select chart, and use pie-chart

Pie Chart - Example

Imagine that a teacher surveys her class on the basis of the favourite sports of her students:

Foodball	Hockey	Cricket	Basketball	Badminton
10	5	5	10	10

Step 1: First, Enter the data into the table.

Foodball	Hockey	Cricket	Basketball	Badminton
10	5	5	10	10

Step 2: Add all the values in the table to get the total.

I.e. Total students are 40 in this case.

Pie Chart - Example

Step 3: Next, divide each value by the total and multiply by 100 to get a percent:

Foodball	Hockey	Cricket	Basketball	Badminton
(10/40) x 100 = 25%	(5/40) x 100 = 12.5%	(5/40) x 100 = 12.5%	(10/40) x 100 = 25%	(10/40) x 100 = 25%

Step 4: Next, we will take a whole circle of 360° and perform the calculations below to determine how many degrees are required for each "pie sector":

Each component's central angle equals (each component's value/the sum of all the component's values)x360°.

Foodball	Hockey	Cricket	Basketball	Badminton
(10/40) x 360° = 90°	(5/40) x 360° = 45°	(5/40) x 360° = 45°	(10/40) x 360° = 90°	(10/40) x 360° = 90°

Pie Chart - Example

Step 5: Draw a circle and use the protractor to measure the degree of each sector.

Favourite Sports Percentage

- Football
- Basketball
- Badminton
- Hockey
- Cricket
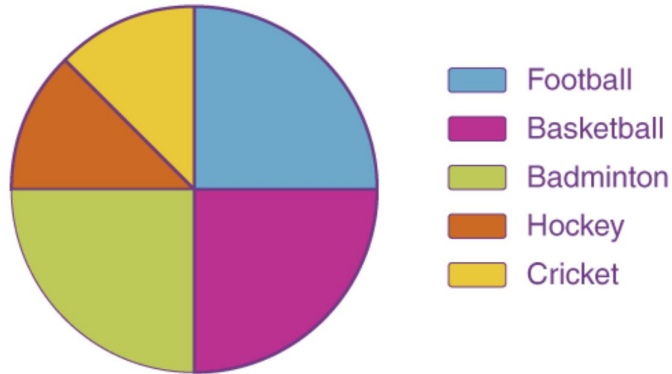
Practice Problem : Frequency Distribution – Continuous

Question-1: The percentages of various crops cultivated in a village of a particular district are given in the following table.

Items	Wheat	Pulses	Jowar	Groundnuts	Vegetables	Total
Percentage of cops	125/3	126/6	25/2	50/3	25/3	100

Refer to the 'Pie-Chart Problem-1' tab:

https://docs.google.com/spreadsheets/d/1Kuz56oAkOfwpOCTJ8sTgi8sstx_qR98heARihhK_f0M/edit?usp=sharing

Refer to the 'Pie-Chart Problem-1: Solution' tab

https://docs.google.com/spreadsheets/d/1Kuz56oAkOfwpOCTJ8sTgi8sstx_qR98heARihhK_f0M/edit?usp=sharing

Relevel
by Unacademy

Practice Problem : Frequency Distribution – Continuous

Question-1: The pie-chart shows the marks obtained by a student in an examination. If the student secures 440 marks in all, calculate the marks in each of the given subjects.

Subject	Mathematics	Science	English	Social Science	Total
Marks	132	99	88	55	440

Refer to the 'Pie-Chart Problem-2' tab:

https://docs.google.com/spreadsheets/d/1Kuz56oAkOfwpOCTJ8sTgi8sstx_qR98heARihhK_f0M/edit?usp=sharing

Refer to the 'Pie-Chart Problem-2: Solution' tab:

https://docs.google.com/spreadsheets/d/1Kuz56oAkOfwpOCTJ8sTgi8sstx_qR98heARihhK_f0M/edit?usp=sharing

THANK YOU