SNS COLLEGE OF TECHNOLOGY

An Autonomous Institution
Coimbatore-35

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

23ADT202 – FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

II YEAR IV SEM

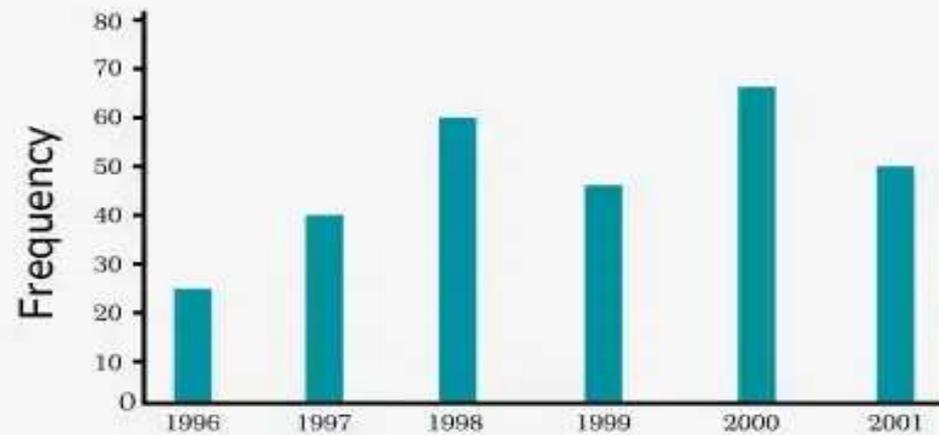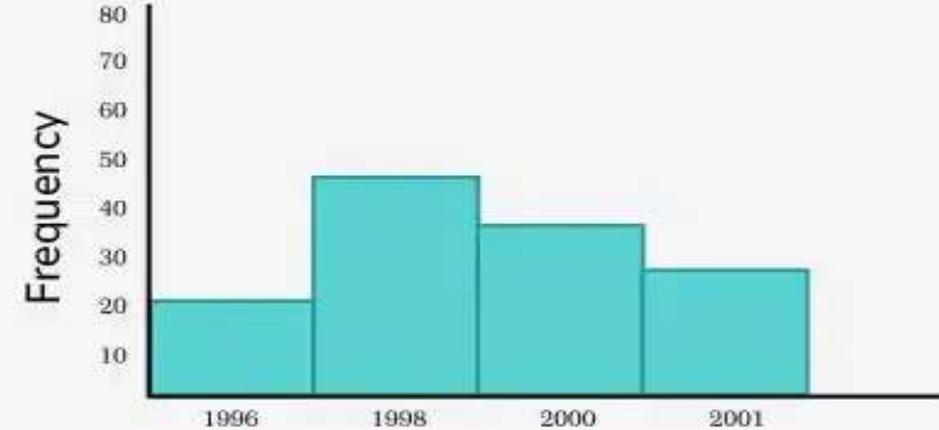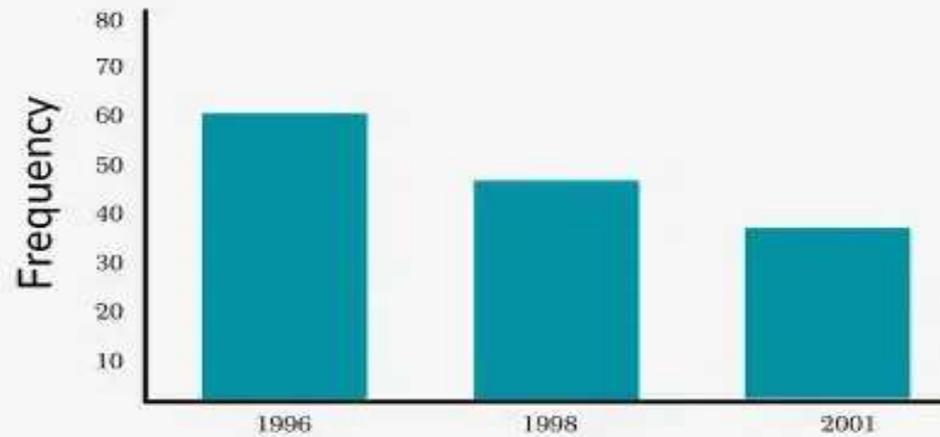## UNIT II – FREQUENCY DISTRIBUTIONS

FREQUENCY DISTRIBUTIONS

EMPATHY:

- A **frequency distribution** is a way to organize data and see how often each value appears.

- It shows how many times each value or range of values occurs in a dataset.

- This helps us understand patterns, like which values are common and which are rare.

- Frequency distributions are often shown in tables or graphs, and make it easier to analyze and conclude data.

FREQUENCY DISTRIBUTION

Frequency Distribution is a tool in statistics that helps us organize the data and also helps us reach meaningful conclusions.
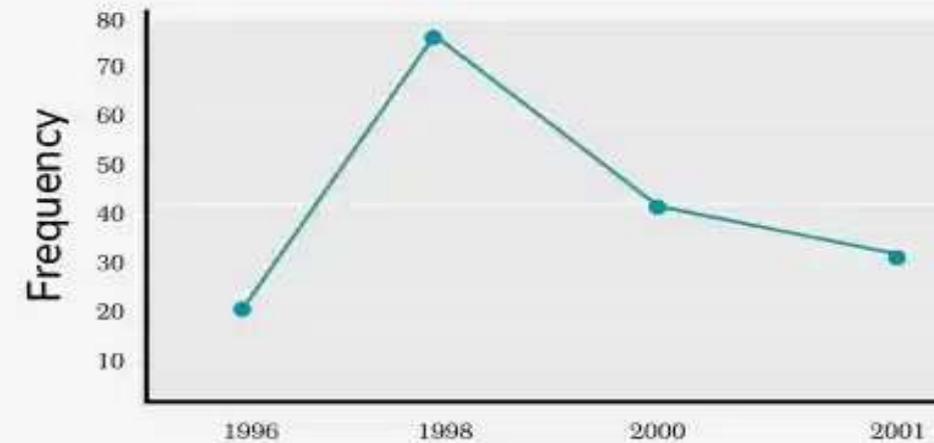
Frequency Distribution Graphs

Bar Graph

Histogram

Pie Chart
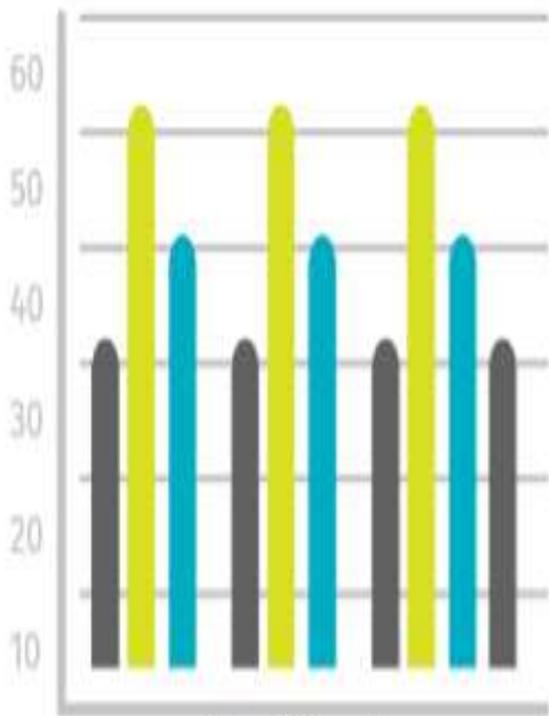

Frequency Polygon

DEFINE:
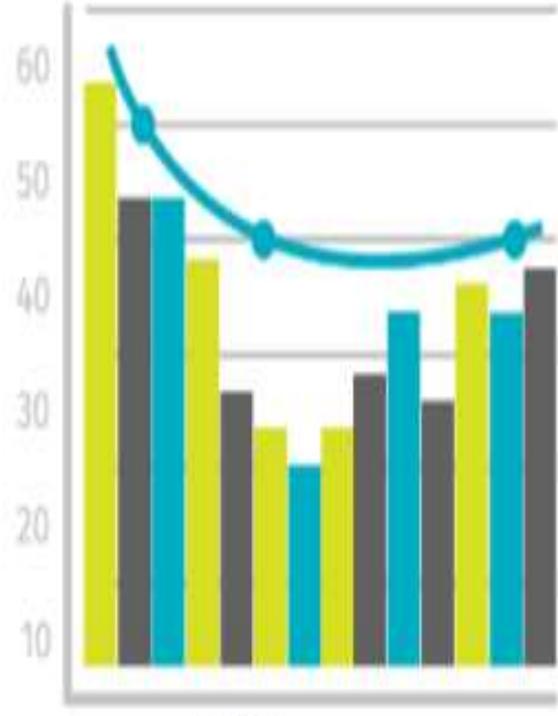
To represent the Frequency Distribution, there are various methods such as a

❖ Histogram
❖ Bar Graph
❖ Frequency Polygon and
❖ Pie Chart.

A brief description of all these graphs is as follows:
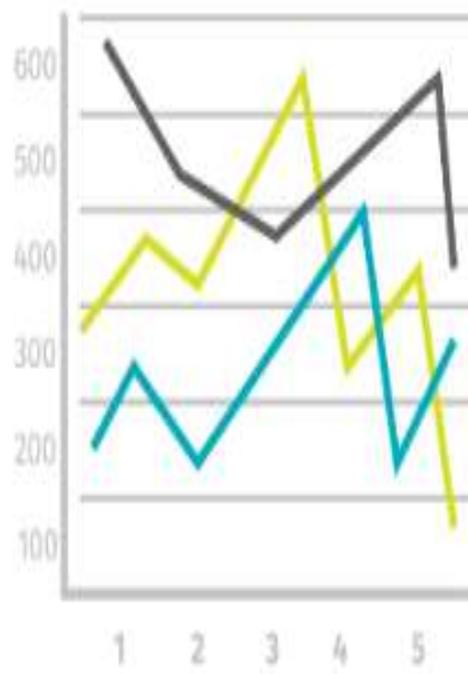
Graph Type	Description	Use Cases
Histogram	Represents the frequency of each interval of continuous data using bars of equal width.	Continuous data distribution analysis.
Bar Graph	Represents the frequency of each interval using bars of equal width; it can also represent discrete data.	Comparing discrete data categories.

Frequency Polygon	Connects midpoints of class frequencies using lines, similar to a histogram but without bars.	Comparing various datasets.
Pie Chart	Circular graph showing data as slices of a circle, indicating the proportional size of each slice relative to the whole dataset.	Showing relative sizes of data portions.

Made with GAMMA

Frequency Distribution Table

• A frequency distribution table is a way to organize and present data in a tabular form, showing which helps us summarize a large dataset into a concise table.

• In the frequency distribution table, there are two columns: one representing the data either in the form of a range or an individual data set and the other column showing the frequency of each interval or individual.

For example, let's say we have a dataset of students' test scores in a class.

Test Score	Frequency
0-20	6
20-40	12
40-60	22
60-80	15
80-100	5

Types of Frequency Distribution
There are four types of frequency distribution:

Grouped Frequency Distribution

Ungrouped Frequency Distribution

Relative Frequency Distribution

Cumulative Frequency Distribution

Grouped Frequency Distribution

- In Grouped Frequency Distribution observations are divided between different intervals known as class intervals and then their frequencies are counted for each class interval.

- This Frequency Distribution is used mostly when the data set is very large.

Example: Make the Frequency Distribution Table for the ungrouped data given as follows:

23, 27, 21, 14, 43, 37, 38, 41, 55, 11, 35, 15, 21, 24, 57, 35, 29, 10, 39, 42, 27, 17, 45, 52, 31, 36, 39, 38, 43, 46, 32, 37, 25

Grouped Frequency Distribution Table

Class Interval	Frequency (f)
10 – 19	5
20 – 29	8
30 – 39	11
40 – 49	6
50 – 59	3
Total	**33**

Histogram for Grouped Data

UnGrouped Frequency Distribution

In Ungrouped Frequency Distribution, all distinct observations are mentioned and counted individually. This Frequency Distribution is often used when the given dataset is small.

Example: Make the Frequency Distribution Table for the ungrouped data given as follows:

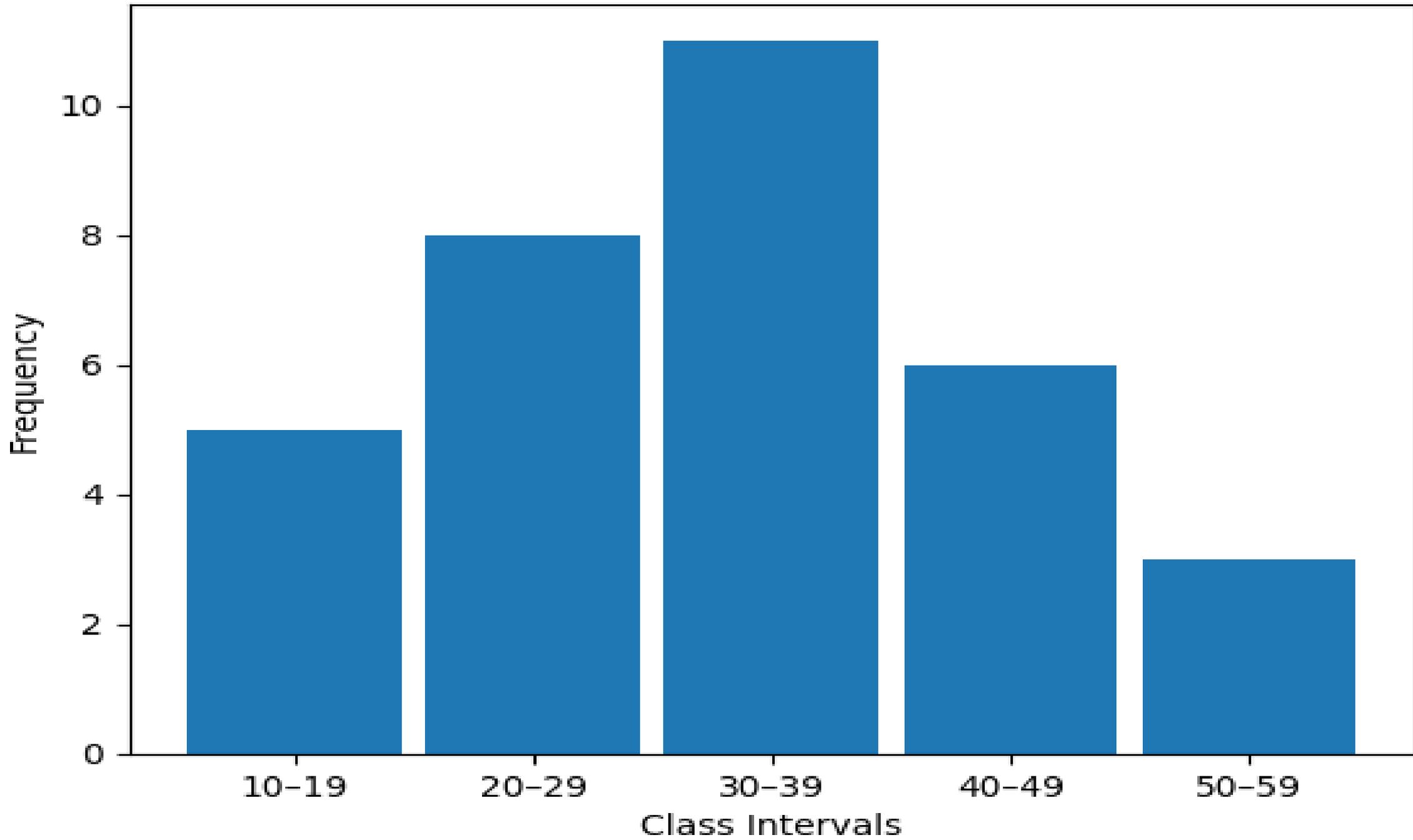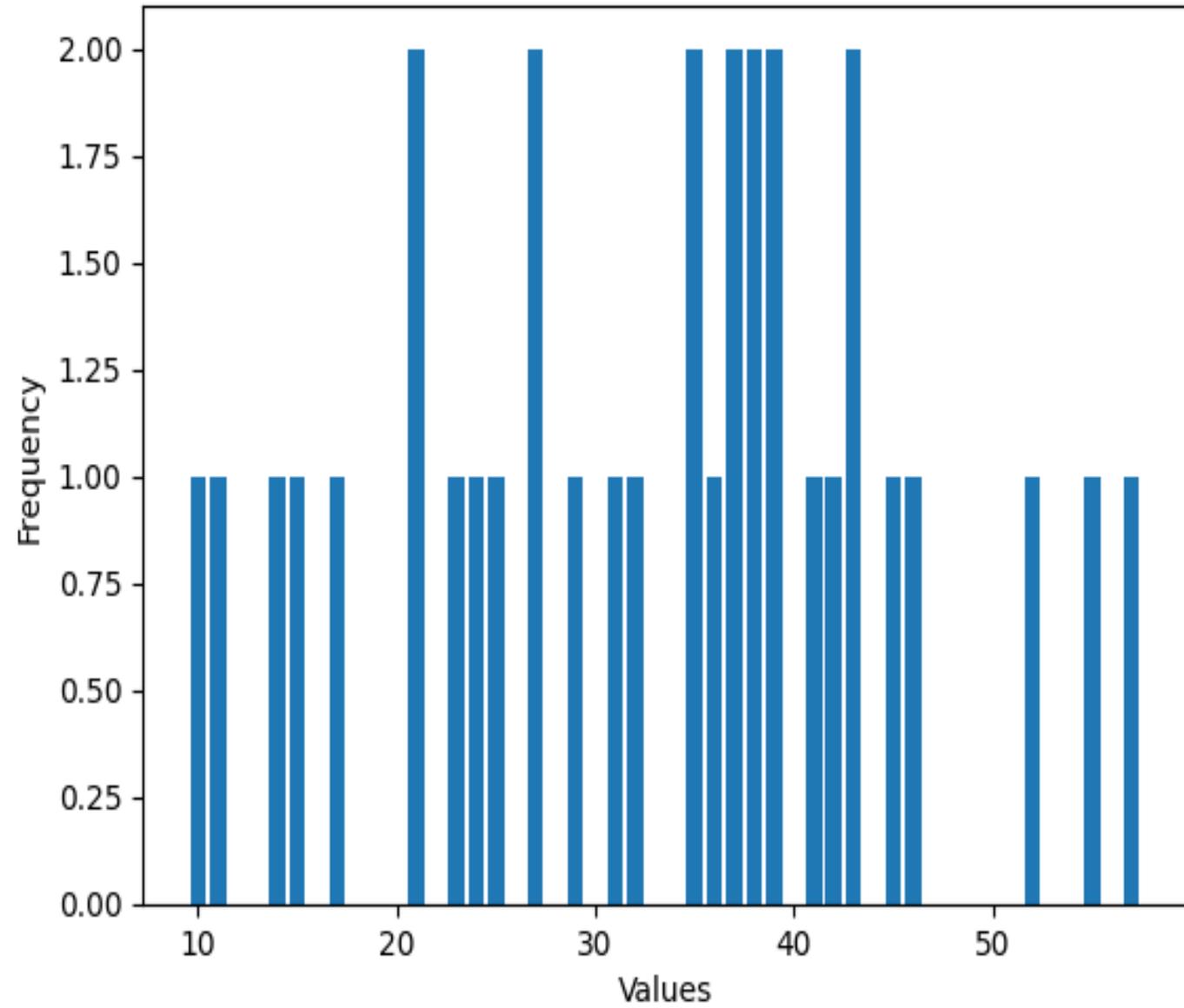10, 20, 15, 25, 30, 10, 15, 10, 25, 20, 15, 10, 30, 25
23, 27, 21, 14, 43, 37, 38, 41, 55, 11, 35, 15, 21, 24, 57,
35, 29, 10, 39, 42, 27, 17, 45, 52, 31, 36, 39, 38, 43, 46,
32, 37, 25

Ungrouped Frequency Distribution Table

Value (x)	Frequency (f)
10	1
11	1
14	1
15	1
17	1
21	2
23	1
24	1
25	1
27	2
29	1
31	1
32	1
35	2
36	1
37	2
38	2
39	2
41	1
42	1
43	2
45	1
46	1
52	1
55	1
57	1
Total	**33**

Bar Chart of Ungrouped Data

Histogram of Ungrouped Data
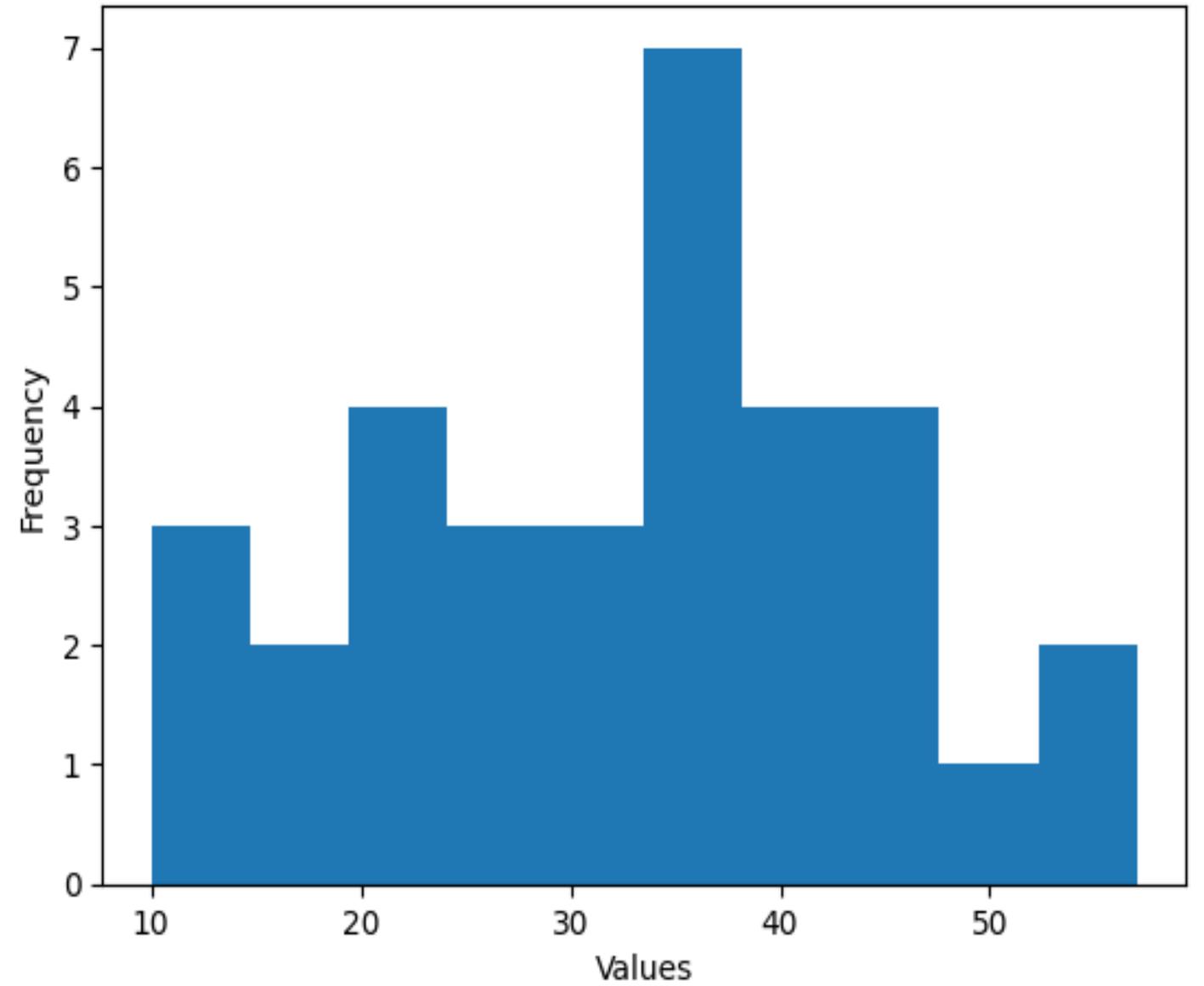
Relative Frequency Distribution

- This distribution displays the proportion or percentage of observations in each interval or class.

- It is useful for comparing different data sets or for analyzing the distribution of data within a set.

- Relative Frequency is given by:

- Relative Frequency = (Frequency of Event)/(Total Number of Events)

Relative Frequency Distribution Table

Class Interval	Frequency (f)	Relative Frequency (f/33)
10 – 19	5	0.15
20 – 29	8	0.24
30 – 39	11	0.33
40 – 49	6	0.18
50 – 59	3	0.09
Total	**33**	**1.00**

Formula

$$\text{Relative } Frequency = \frac{Class\ Frequency}{Total\ Frequency}$$

Total frequency = **33**

Relative Frequency Distribution (Grouped Data)
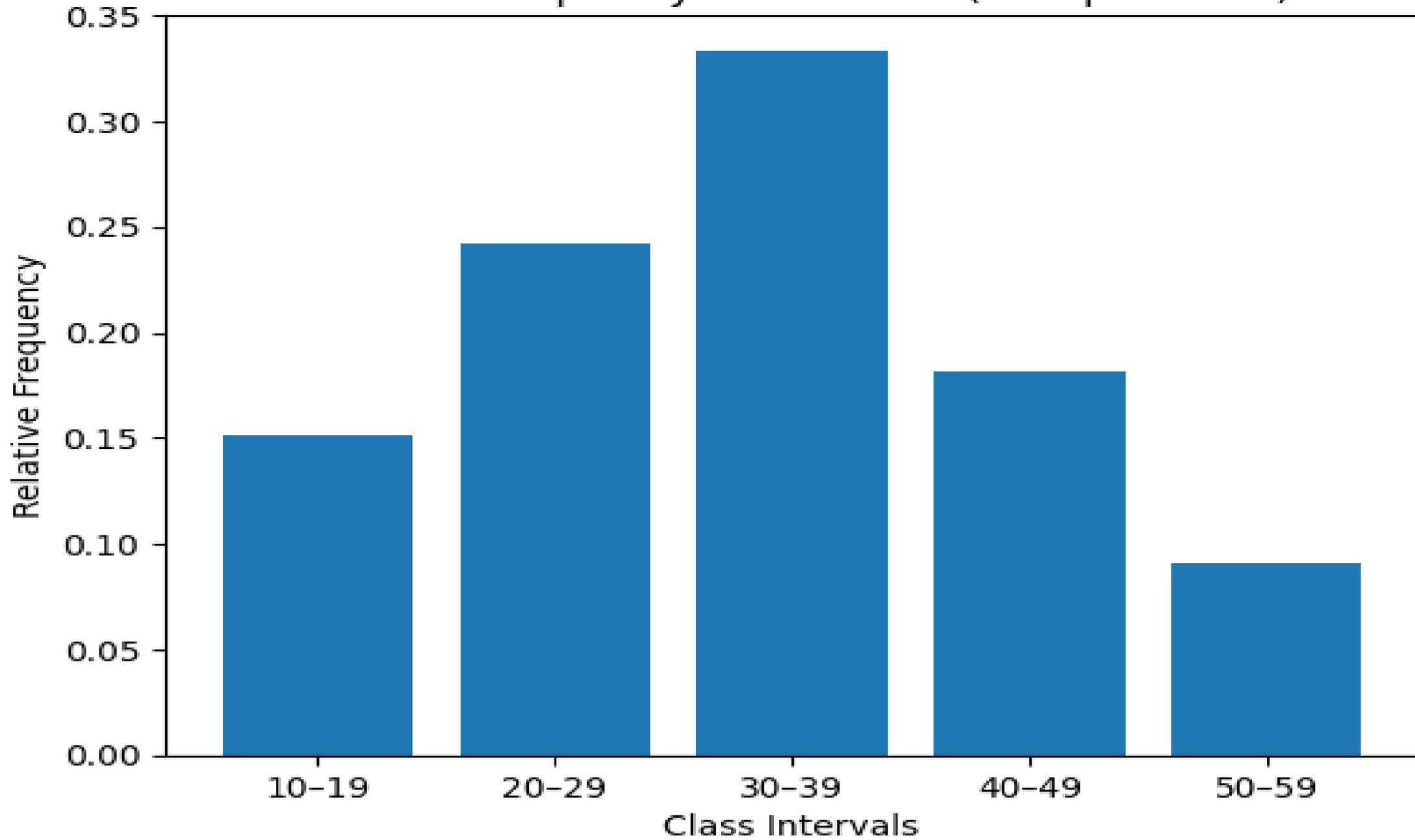
Cumulative Frequency Distribution

- A **cumulative frequency distribution** shows the total number of observations **up to and including** a certain value or class.

- It helps in understanding how values accumulate across a dataset:

- **Less than Type:** We sum all the frequencies before the current interval.

- **More than Type:** We sum all the frequencies after the current interval.

Cumulative Frequency Table (Less-than)

Class Interval	Frequency (f)	Cumulative Frequency (cf)
10 – 19	5	5
20 – 29	8	13
30 – 39	11	24
40 – 49	6	30
50 – 59	3	33

Normal Distribution Curve

Skewed Distribution Curve

Bimodal Distribution Curve

Multimodal Distribution Curve

Uniform Distribution Curve

Exponential Distribution Curve

Log-Normal Distribution Curve

Power Law Distribution Curve

Frequency Distribution Curve

PROTOTYPING:

❖ A frequency distribution curve, also known as a frequency curve, is a graphical representation of a data set's frequency distribution.

❖ It is used to visualize the distribution and frequency of values or observations within a dataset.

❖ Let's understand its different types based on the shape of it, as follows:

10

Made with GAMMA
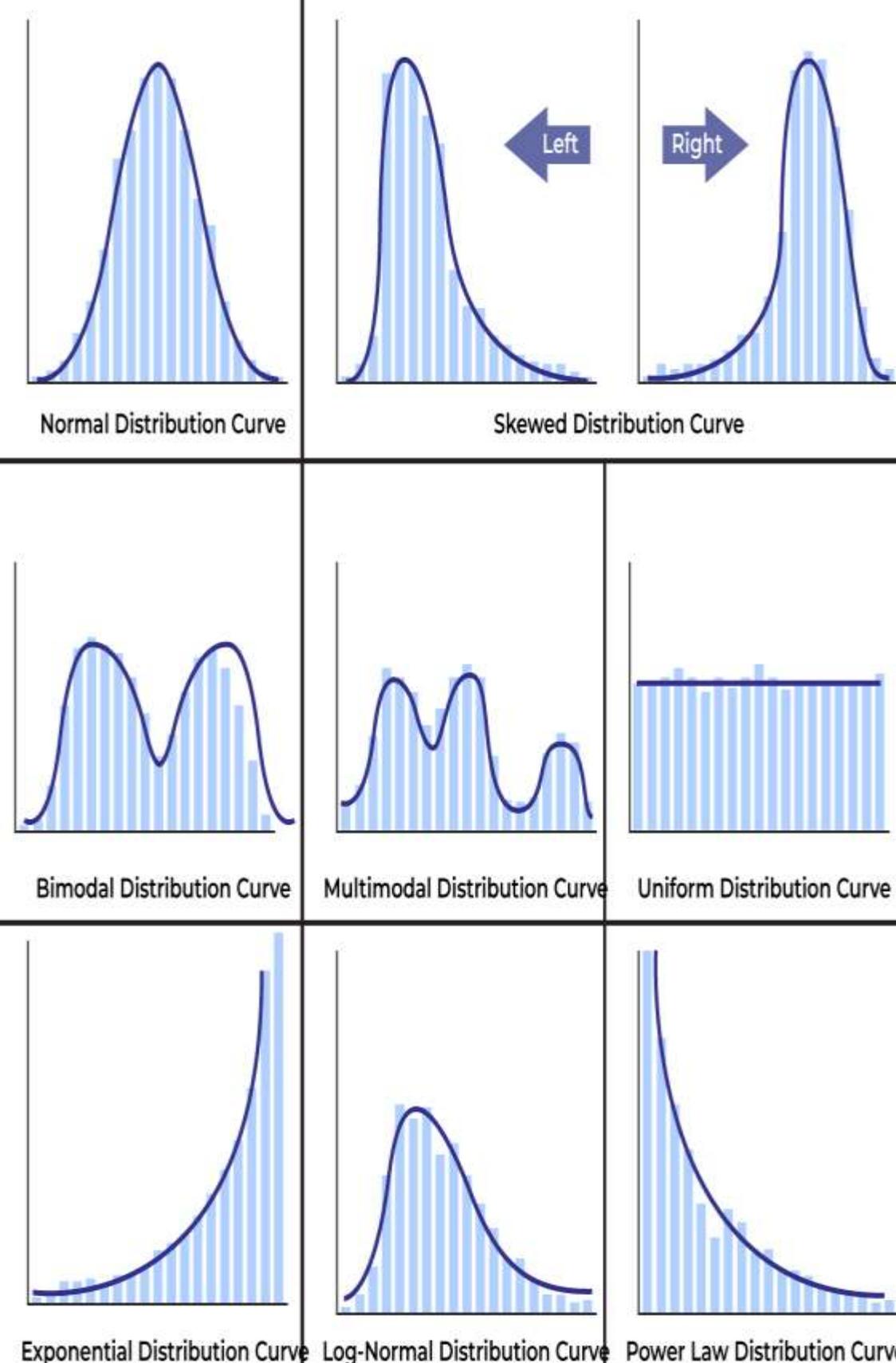
Frequency Distribution Formula

- Various formulas can be learned in the context of Frequency Distribution, one such formula is the coefficient of variation.

- This formula for Frequency Distribution is discussed below in detail.

- Frequency (f)= Number of times a value occurs

Coefficient of Variation

- We can use mean and standard deviation to describe the dispersion in the values.

- But sometimes while comparing the two series or frequency distributions becomes a little hard as sometimes both have different units.

The coefficient of Variation is defined as,

$$\frac{\sigma}{\bar{x}} \times 100$$

Where,

- σ represents the standard deviation
- $\bar{x}$ represents the mean of the observations

Note: Data with greater C.V. is said to be more variable than the other.

The series having lesser C.V. is said to be more consistent than the other.