STATISTICAL METHODS

Unit I

Introduction: Application of statistics in Business & Management : Basic concept of statistical studies: Population, Variable, Parameter, Sample ; Classification of Data; Diagrammatic & Graphical Presentation of Data : Bar Diagram, Histogram, Pie-Diagram Frequency Polygons and Ogives.

Unit II

Summary Statistics: Measures of Central Tendency: Arithmetic Mean, Weighted Mean, Median and Mode. Index Numbers: Concept & Applications

Unit III

Measures of Dispersion: Range, Average Deviation, Standard Deviation, Variance and Coefficient of Variation

Unit IV

Forecasting Techniques: Simple Correlation & Regression Analysis, Time Series Analysis-Trend Analysis, Cyclical Analysis, Seasonal Analysis, Irregular Variation.

Unit V

Probability: Introduction of Probability Theories, Concepts, Addition & Multiplication Theorems, Probability Distribution: Binomial, Poisson, Normal & Exponential

STATISTICS

- The word '**Statistics'** may have been derived from the French word- '**Statistique'** or the German word- '**Statistik'** or the Italian word- '**Statista'** or the Latin word- '**Status'.**

- In common language, Statistics is regarded as sets of numbers and grouping of numerical figures in tabular form.

- Statistics is the science of acquiring, classifying, organizing, analyzing, interpreting and presenting the numerical data, so as to make inferences about the population from the samples drawn. Statistics translates the complex facts & figures into simple & understandable form and boils up the complexities of details in a very short span of time.

- For Example, numerical data relating to population, production, price-level, national income, crimes, literacy, unemployment etc.

DEFINITION OF STATISTICS

- According to Croxton and Cowden "It is the science of collection, presentation, analysis and interpretation of numerical data from logical analysis"

Statistics in Business and Management:

- **ACCOUNTING** : Statistical sampling techniques are used during the conduction of audit for clients. It also helps in detecting the trend and make a projection for next year.

- **FINANCE & INVESTMENT**: Statistical information can be used to study the trend in securities and that can be used to provide investment recommendations. Statistical methods help in selecting securities which are safe and have the best prospects of yielding a good income.

- **MARKETING**: Statistical Analysis is frequently used in for making a decision in the field of marketing. It is first step to find out what can be sold and to whom. Then using statistical methods a suitable strategy is formulated. A statistical analysis of data on production purchasing power, manpower, habits of competitors, habits of consumer, transportation cost can be done before entering a new market.

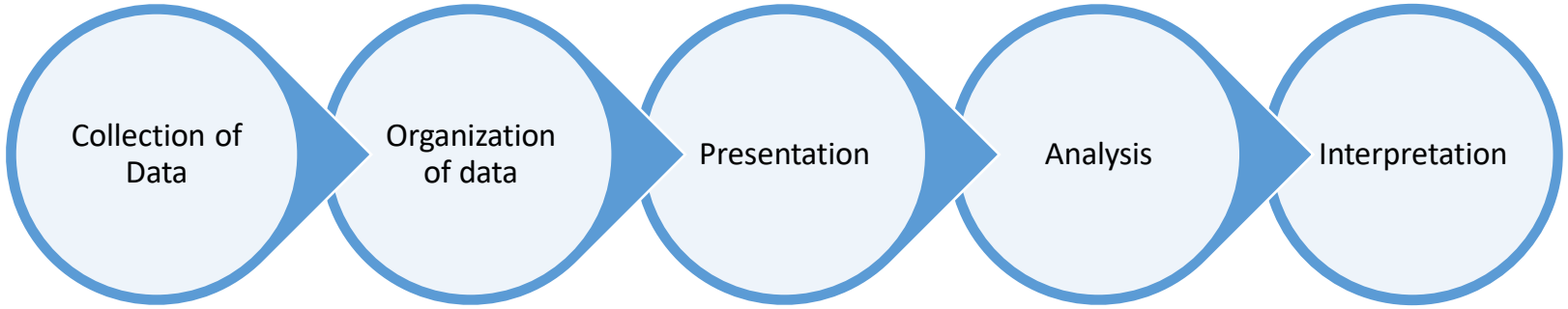
- **PRODUCTION**: Statistical methods are used in quality control during the production process. It is also used to control and manage the flow of production. Statistical methods are used in the scheduling of men and machines.

- **BANKING**: Statistical data gathering and analysis of the information help banks in their own business and also gives idea of the general economic situation of every segment of business in which they may have interest. Using this analysis they can formulate their lending policies.

- **PURCHASE**: Purchase department can fix their schedule of purchasing orders depending upon the trends in consumption of raw materials and inputs. Thus they decide what to buy? When to buy? And how much to buy?

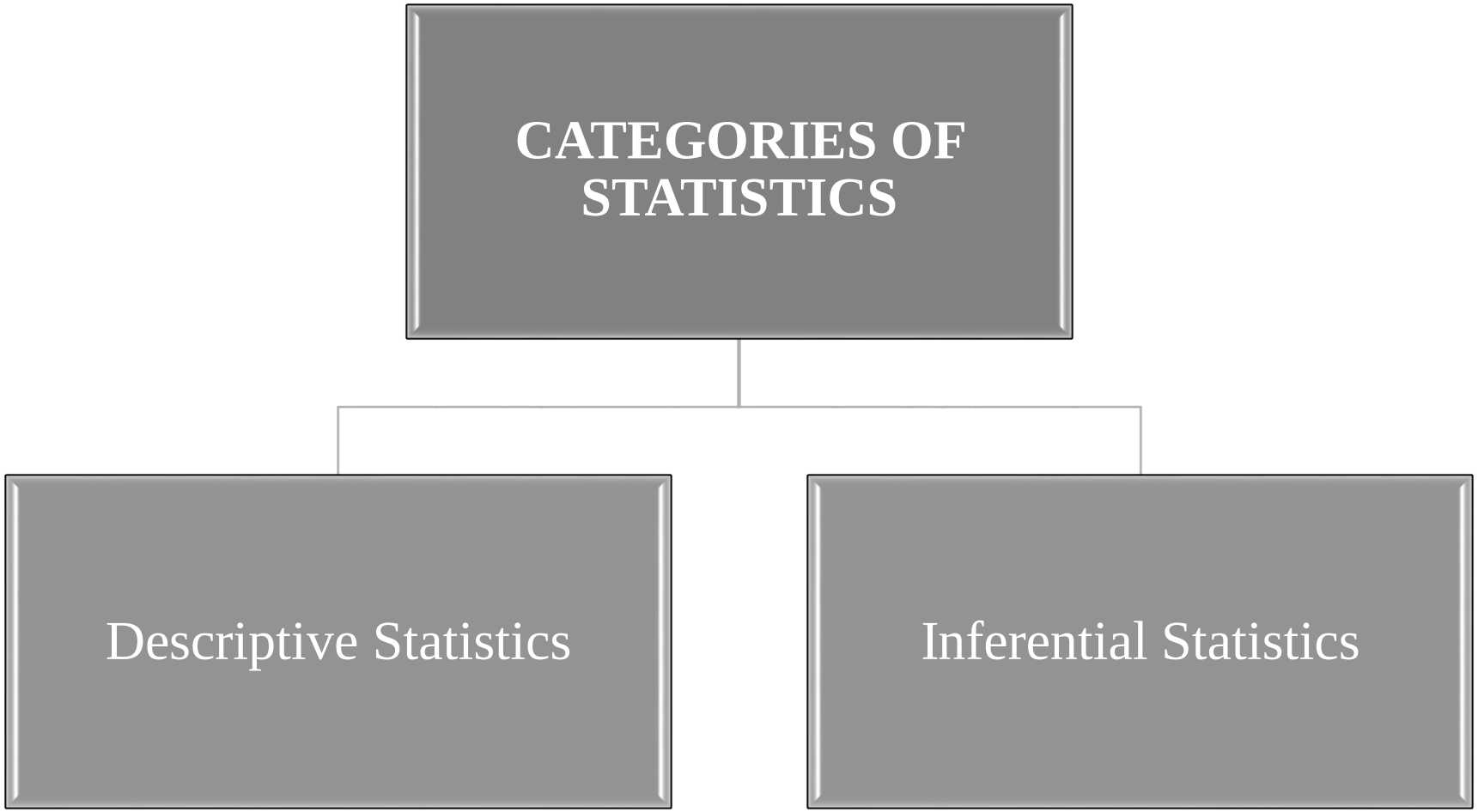
- **ECONOMICS**: Statistical techniques and analysis are used for forecasting the future of the economy. Time series like moving averages, indicators like inflation index are statistical methods. We can consider statistics as the backbone of economics.

STAGES OF STATISTICS



Functions of Statistics

1. Definitiveness
2. Comparison
3. Prediction
4. Systematic Collection
5. Facts numerically expressed


CATEGORIES OF STATISTICS
Descriptive Statistics
Inferential Statistics

DESCRIPTIVE STATISTICS

- It is the term given to analysis of data that helps to describe, show & summarize data in a meaningful way.
- It is that part of statistics which describes the characteristics of a particular dataset under study, with the help of brief summary about the sample.
- The methods used in statistics are preparing tables, drawing graphs, measuring central tendency, and the variation of the data from the central value.

Characteristics of Descriptive Statistics

1. Descriptive Statistics does not allow us to make conclusions beyond the data we have analyzed.
2. They are simply a way to describe our data.
3. It enables us to present the data in a more meaningful way which allows simpler interpretation of the data.
4. The data is represented by tables, graphs and charts.
5. This provides a quick method to make comparison between different data sets and to spot the smallest and largest values and trends or changes over a period of time.

LIMITATIONS OF DESCRIPTIVE STATISTICS

1. They only allow us to make summations about the people or objects that we have actually measured.
2. We cannot use the data we have collected to generalize to other people or objects. Thus no general inference can be obtained.

INFERENTIAL STATISTICS

- Inferential Statistics are techniques and methods that allow us to use samples to make generalizations about the population from which the samples were drawn.
- The data obtained in descriptive statistics are analyzed and a valid inference is made out of it for effective decision making for managers and professionals. In this type of deductions and conclusions are made regarding the population under study by collecting a sample from the population.

Characteristics of Inferential Statistics

1. They are produced through complex mathematical calculations that allow scientists to infer trends about a larger population based on a study of a sample taken from it.
2. Inferential statistics start with a sample and then generalizes to a sample and then generalizes to a population.

LIMITATIONS OF INFERENTIAL STATISTICS:

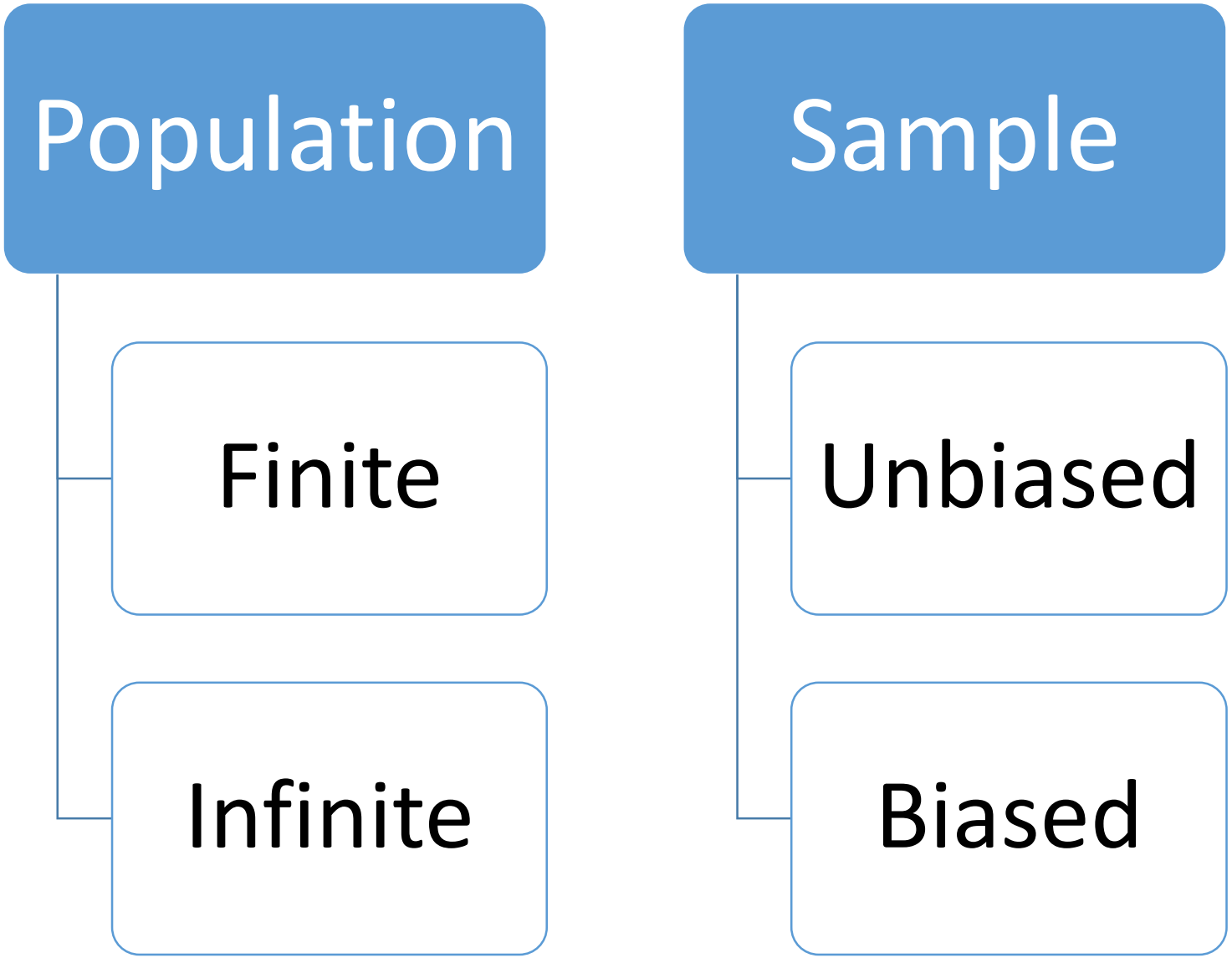
1. It depends on the data provided (sample selected) and therefore, we cannot ever be completely sure that the values/statistics calculate are correct. There will always be a degree of uncertainty in this study.
2. The inferential tests require the users to make educated guesses (based on theory) to run the inferential tests. Again there will be some uncertainty in this process.

DESCRIPTIVE V/S INFERENTIAL

BASIS	DESCRIPTIVE	INFERENTIAL
Meaning	It is a branch of statistics that deals with describing the population under study, to summarize sample.	Inferential statistics is a branch of statistics that aims at making deductions about the population, with the help of sample survey and observation.
Includes	Collection, organization, classification, summarization, analysis, interpretation and presentation of data.	Drawing conclusions, making comparisons, performing estimation, determining cause and effect relationship, hypothesis testing and making predictions.
Objective	To describe the characteristics of the sample	To draw conclusions of the data

BASIS	DESCRIPTIVE	INFERENTIAL
Usage	It describes the basic feature of the situation.	It explains the chance of occurrence of an event or activity
Tools used	Measures of central tendency, Measures of Dispersion	Hypothesis Test and Analysis of variance
Used when	Data set is small	Data set is large

BASIS	POPULATION	SAMPLE
Meaning	Population refers to the collection of all elements possessing common characteristics, that comprises universe.	Sample means a sub-group of the members of population chosen for participation in the study.
Includes	Each and every unit of the group.	Only a handful of units of population.
Characteristic	Parameter	Statistic
Data collection	Complete enumeration or census	Sample survey or sampling
Focus on	Identifying the characteristics	Making inferences about population.


Population
Finite
Infinite
Sample
Unbiased
Biased

BASIS	PARAMETER	STATISTIC
Meaning	• Parameter refers to a measure which describes population. • A fixed characteristic of population based on all the elements of the population is termed as the parameter.	• Statistic is a measure which describes a fraction of population. • A statistic is defined as a numerical value, which is obtained from a sample of data
Numerical Value	Fixed	Variable
Statistical Notation	μ = Population Mean σ = Population Standard Deviation P = Population Proportion X = Data Elements N = Size of Population ρ = Correlation coefficient	$\bar{x}$ = Sample Mean s = Sample Standard Deviation $\hat{p}$ = Sample Proportion x = Data Elements n = Size of sample r = Correlation coefficient

1) A researcher wants to know the average weight of females aged 22 years or older in India. The researcher obtains the average weight of 54 kg, from a random sample of 40 females.

Solution: In the given situation, the statistics are the average weight of 54 kg, calculated from a simple random sample of 40 females, in India while the parameter is the mean weight of all females aged 22 years or older.

2) A researcher wants to estimate the average amount of water consumed by male teenagers in a day. From a simple random sample of 55 male teens the researcher obtains an average of 1.5 liters of water.

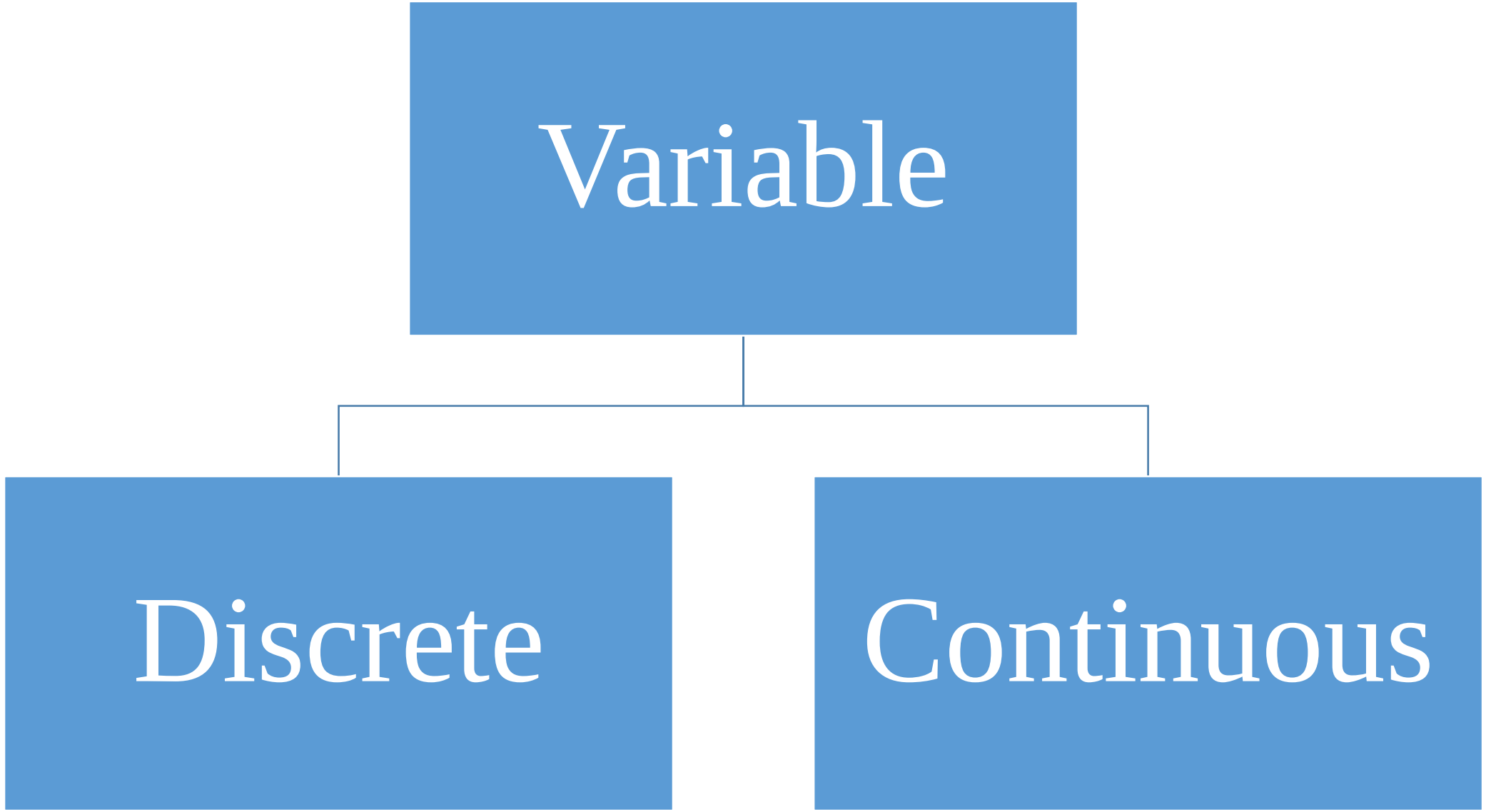
Solution: In this question, the parameter is the average amount of water consumed by all male teenagers, in a day whereas the statistic is the average 1.5 liters of water consumed in a day by male teens, obtained from a simple random sample of 55 male teens.

VARIABLE

- The term variable is derived from the word ‘vary’ which means to differ or change. Hence, variable means the characteristic which varies or differs or changes from person to person, time to time, place to place etc.

- A variable refers to quantity or attribute whose value varies from one investigation to another.

- For example: Age, height, weight, expenditure, hair colour etc.


Variable
Discrete
Continuous

DISCRETE VARIABLE

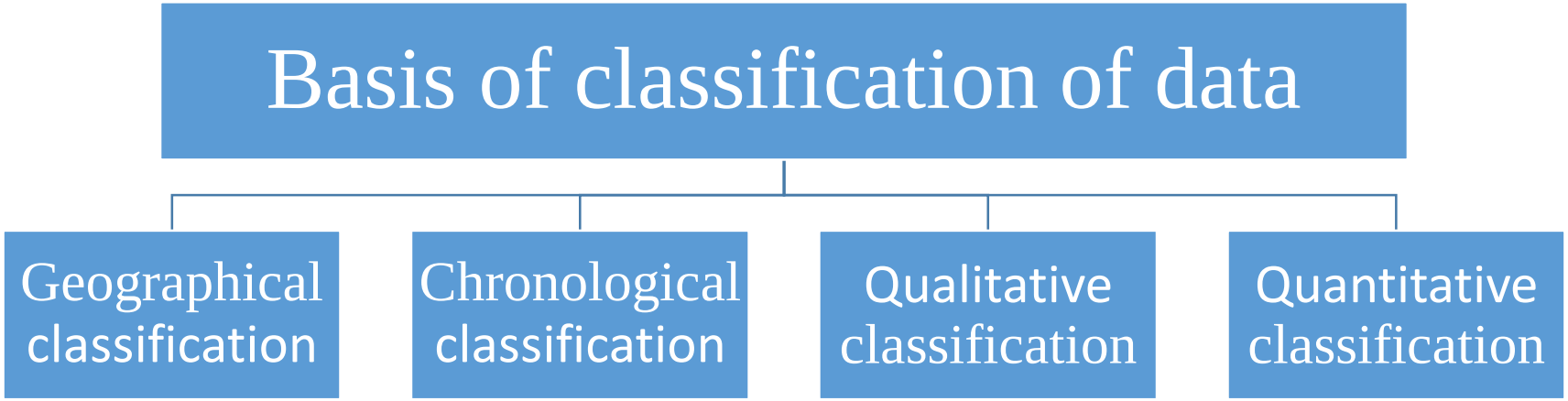
- Variables which are capable of taking an only exact value and not any fractional value are termed as discrete variables.
- For example: Number of workers in a factory, number of students in a class, number of children in a family etc.

CONTINUOUS VARIABLE

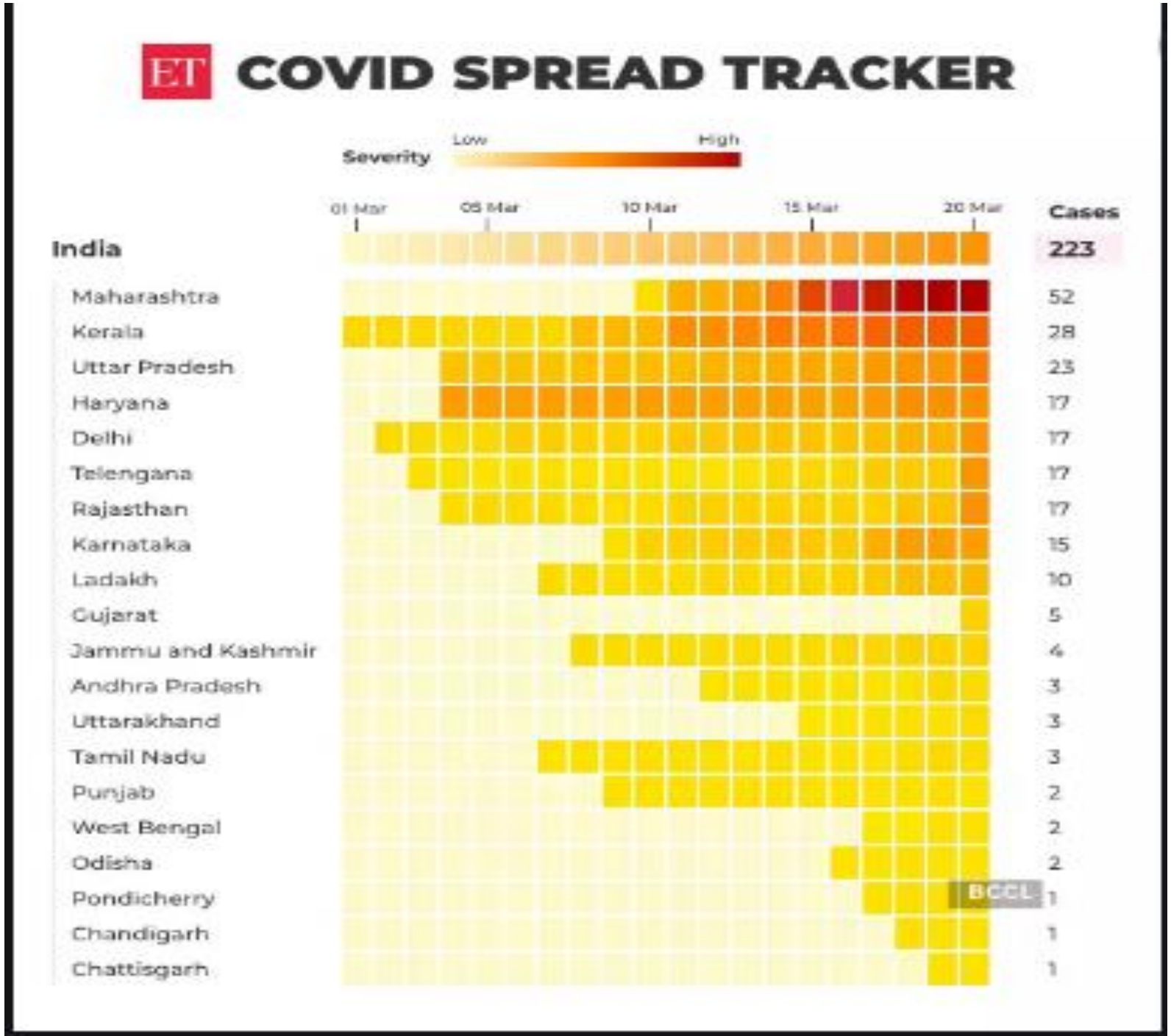
- Those variables which can take all the possible values in a given specified range are termed as continuous variable.
- For example: Temperature, Height, Weight, Marks etc.

CLASSIFICATION OF DATA

- It is the process of arranging data into homogeneous (similar) groups according to their common characteristics.
- Raw data cannot be easily understood and it is not fit for further analysis and interpretation. Classification of data helps users in comparison and analysis.
- For example: the population of a city can be grouped according to gender, age, marital status etc.


Basis of classification of data
Geographical classification
Chronological classification
Qualitative classification
Quantitative classification

- **Geographical Classification:** When data are classified with reference to geographical locations such as countries, states, cities, districts etc.


ET COVID SPREAD TRACKER
Severity
Low
High
01 Mar
05 Mar
10 Mar
15 Mar
20 Mar
Cases
India
223
Maharashtra
52
Kerala
28
Uttar Pradesh
23
Haryana
17
Delhi
17
Telengana
17
Rajasthan
17
Karnataka
15
Ladakh
10
Gujarat
5
Jammu and Kashmir
4
Andhra Pradesh
3
Uttarakhand
3
Tamil Nadu
3
Punjab
2
West Bengal
2
Odisha
2
Pondicherry
1
Chandigarh
1
Chattisgarh
1
BCCL

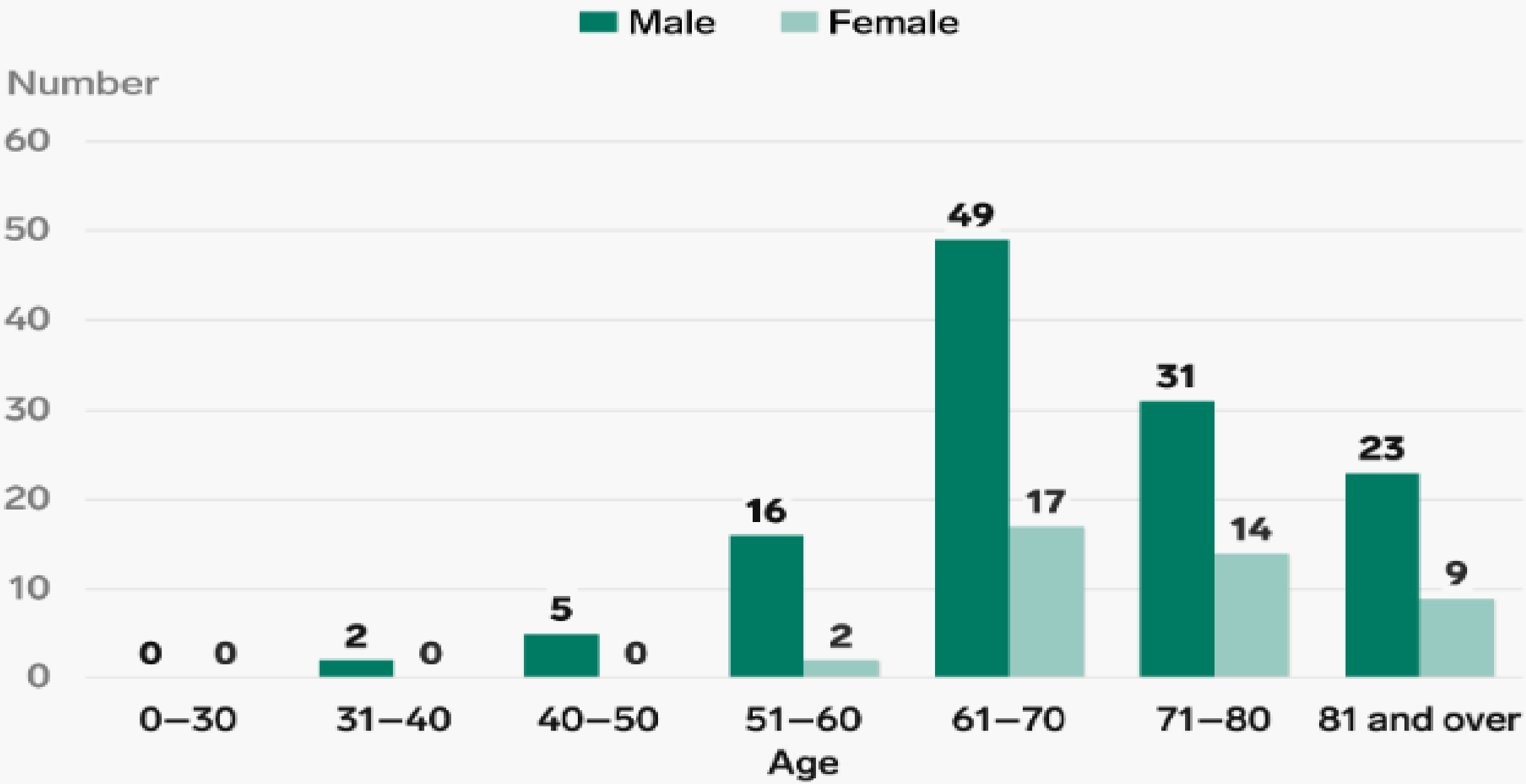
- **Chronological classification:** When data are classified either in ascending or in descending order with reference to time such as years, months, weeks etc.

Day-by-Day number of new cases

- **Qualitative classification:** When data are classified on the basis of some attributes or qualities such as honesty, beauty, intelligence, literacy, marital status etc.

- **Quantitative classification:** when data are classified on the basis of some measurable characteristics such as height, weight, age, income, marks etc.

COVID-19 deaths by age and sex in Wuhan

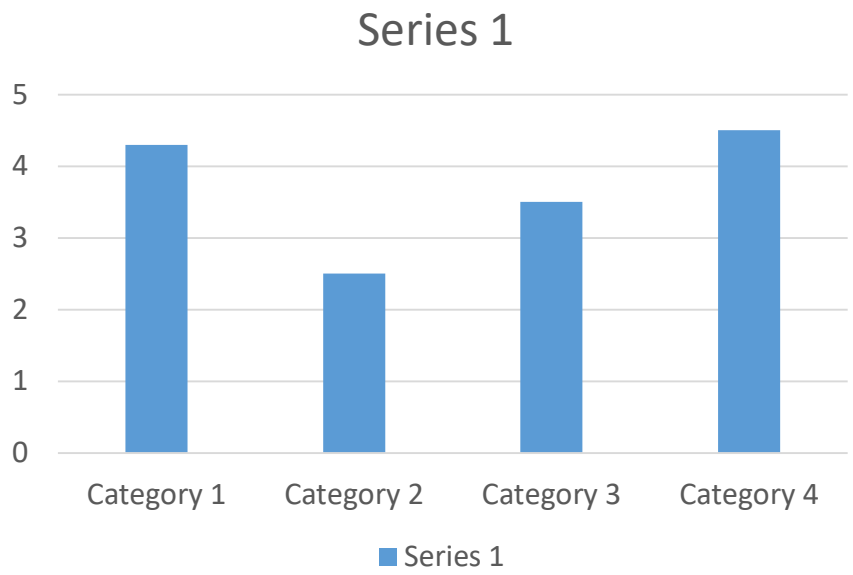


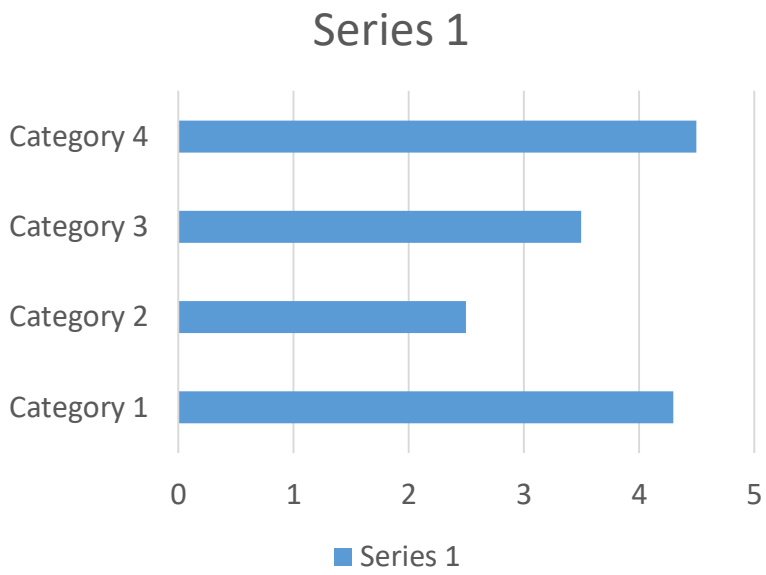
DIAGRAMMATIC PRESENTATION

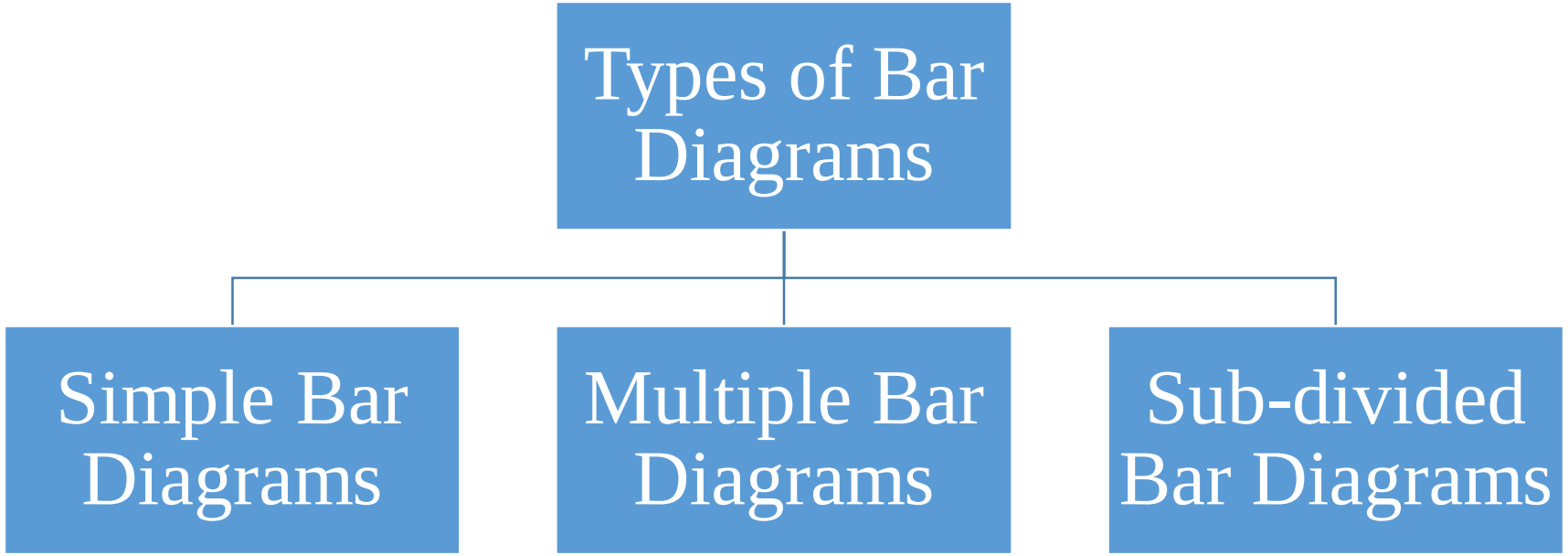
- Diagrammatic presentation is a technique of presenting numeric data.
- It translates pretty effectively highly complex ideas included in numbers into more concrete and quickly understandable form.
- It is the most attractive and appealing way to represent statistical data.

BAR DIAGRAM

- Length of various bars is proportionate to the magnitude of the data.
- As the width of the bars is not at all significant, width is uniform for all bars.
- The gap between various bars is identical.
- Bars can be prepared either vertically or horizontally.
- Both positive and negative values can be presented through bars.



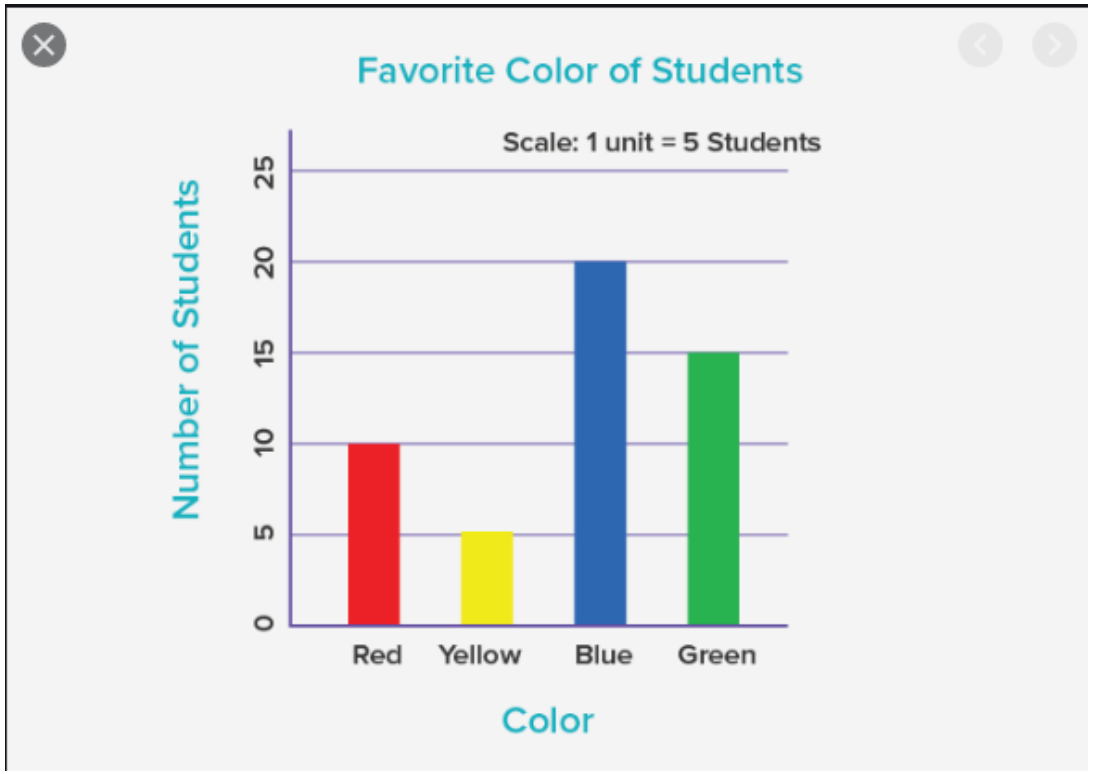



Types of Bar Diagrams
Simple Bar Diagrams
Multiple Bar Diagrams
Sub-divided Bar Diagrams

SIMPLE BAR DIAGRAMS:

- In Simple Bar Diagram, bar is constructed to represent one value of a given variable.
- The length of various bars is in the ratio of the magnitude of the given data.
- It is generally presented when the data indicates different values of a variable over a time period or when the data represents different situations.

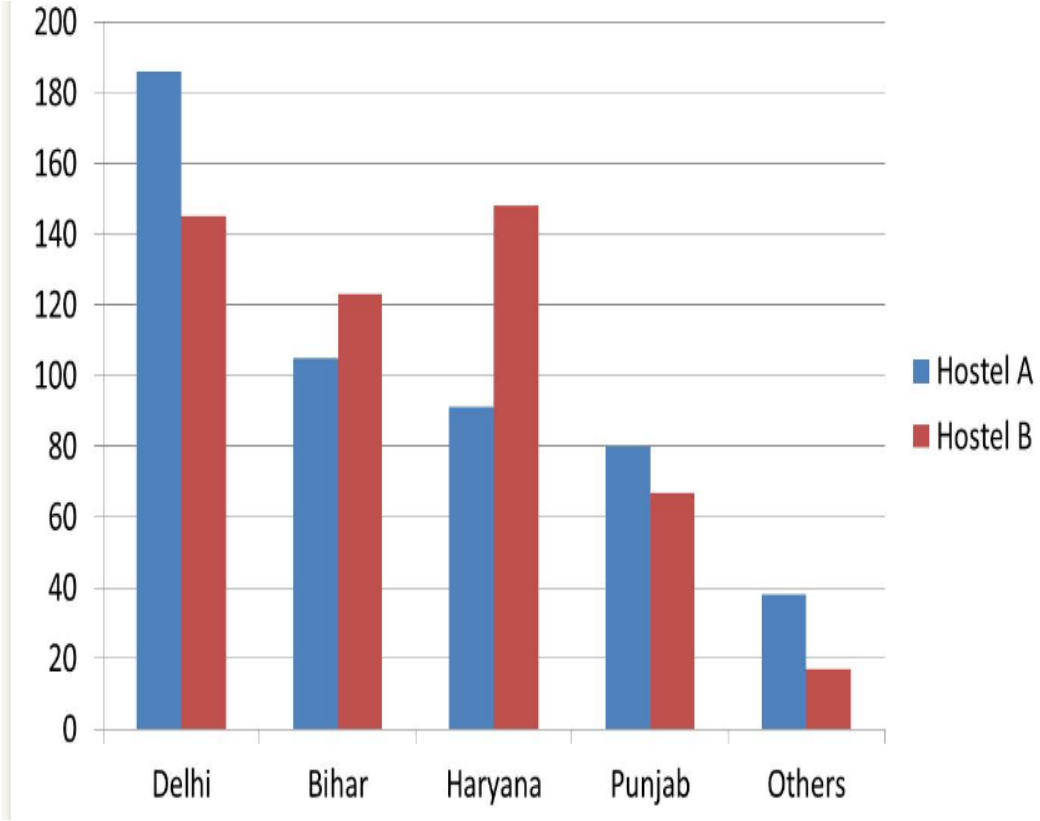
Favorite Color	Number of students
Red	10
Yellow	5
Blue	20
Green	15



MULTIPLE BAR DIAGRAM:

- In Multiple Bar Diagram, two or more bars adjoining each other are constructed to represent the values of different variables or the values of various components of the same variables.
- This diagram facilitates comparison of the values of different variables in a set and comparison of the values of the same variable over a period of time or different situations.

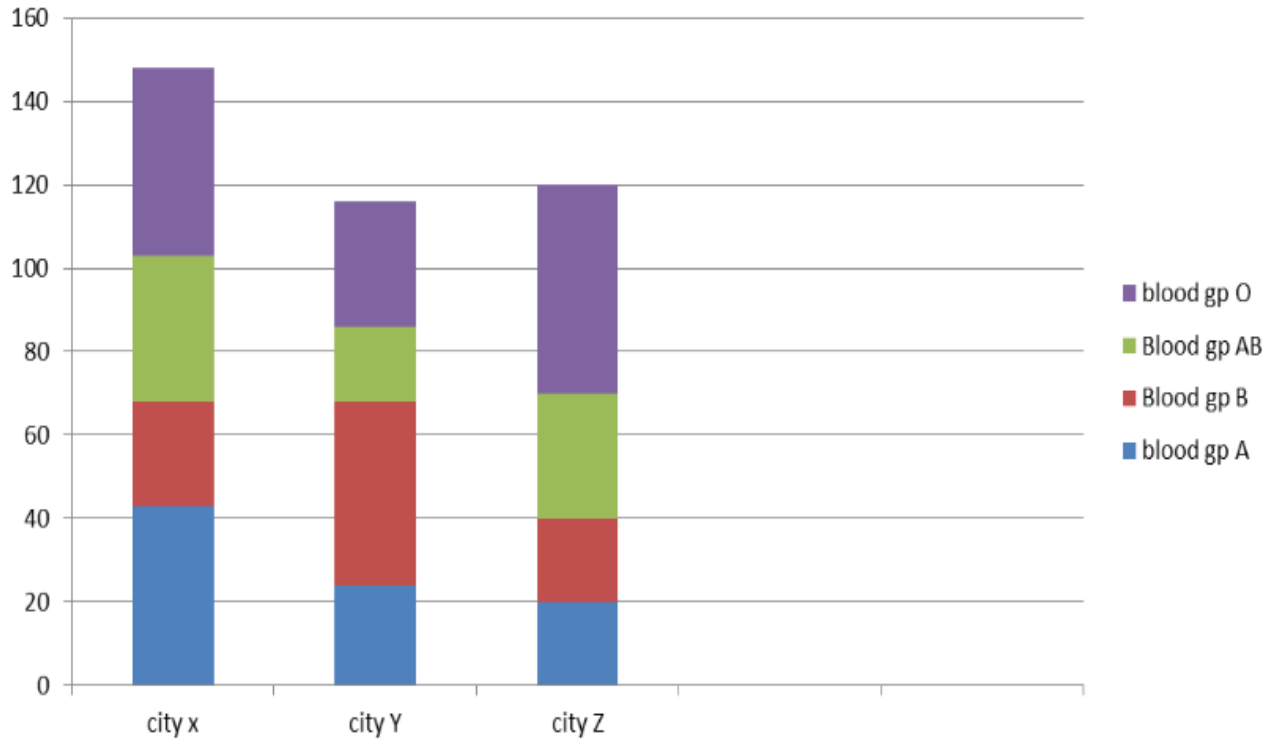
	Hostel A	Hostel B
Delhi	185	145
Bihar	105	122
Haryana	90	150
Punjab	80	65
Others	39	18



SUB-DIVIDED BAR DIAGRAM

- In Sub-divided Bar Diagram, one bar is constructed for total value of the variable and this bar is sub-divided in proportion to the values of the various components of the variable.
- The values of the different components are cumulated for constructing this bar diagram and the bar is to be sub-divided at these cumulated points.

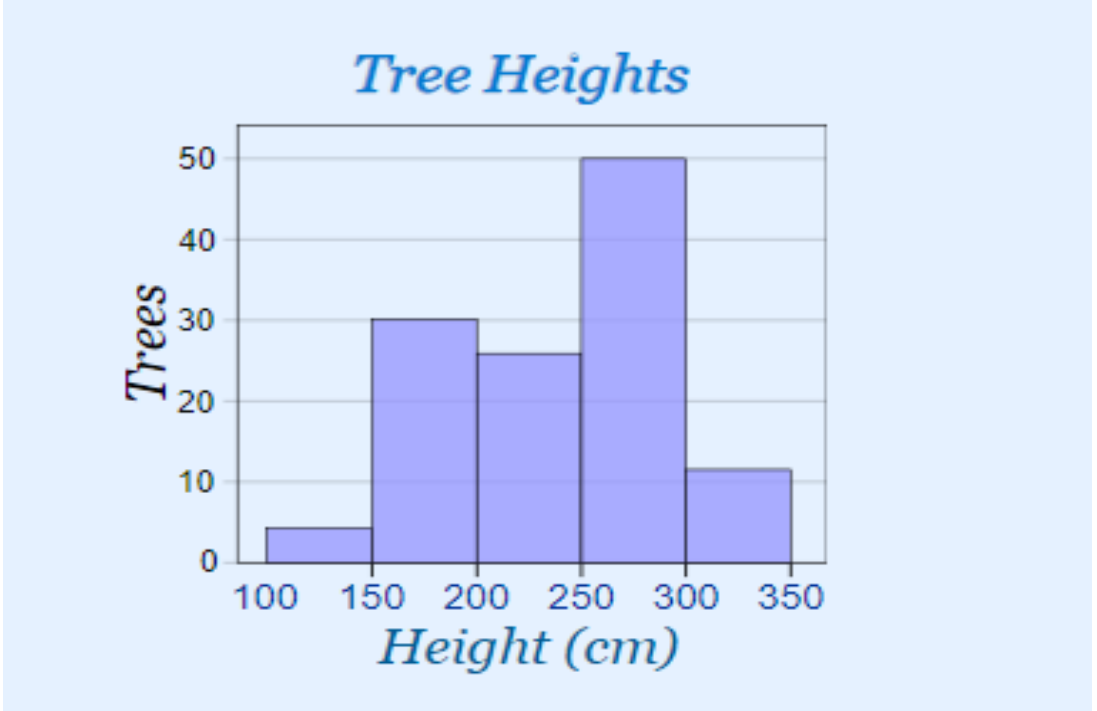
	City X	City Y	City Z
A	41	21	20
B	21	41	20
AB	41	21	30
O	41	35	50
Total	**144**	**118**	**120**



HISTOGRAM

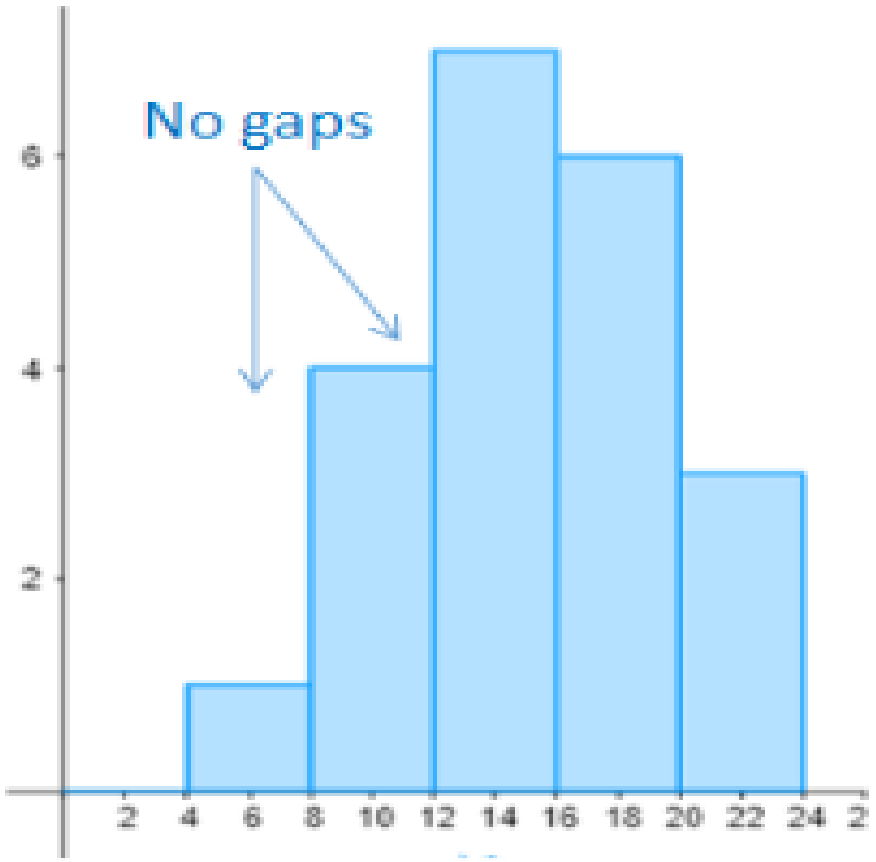
A histogram is the graphical representation of a frequency distribution in the form of rectangles with class intervals as bases and the corresponding frequencies as heights, there being no gap between any successive rectangles.

Height of trees	Number of trees
100-150	4
150-200	30
200-250	28
250-300	50
300-350	11

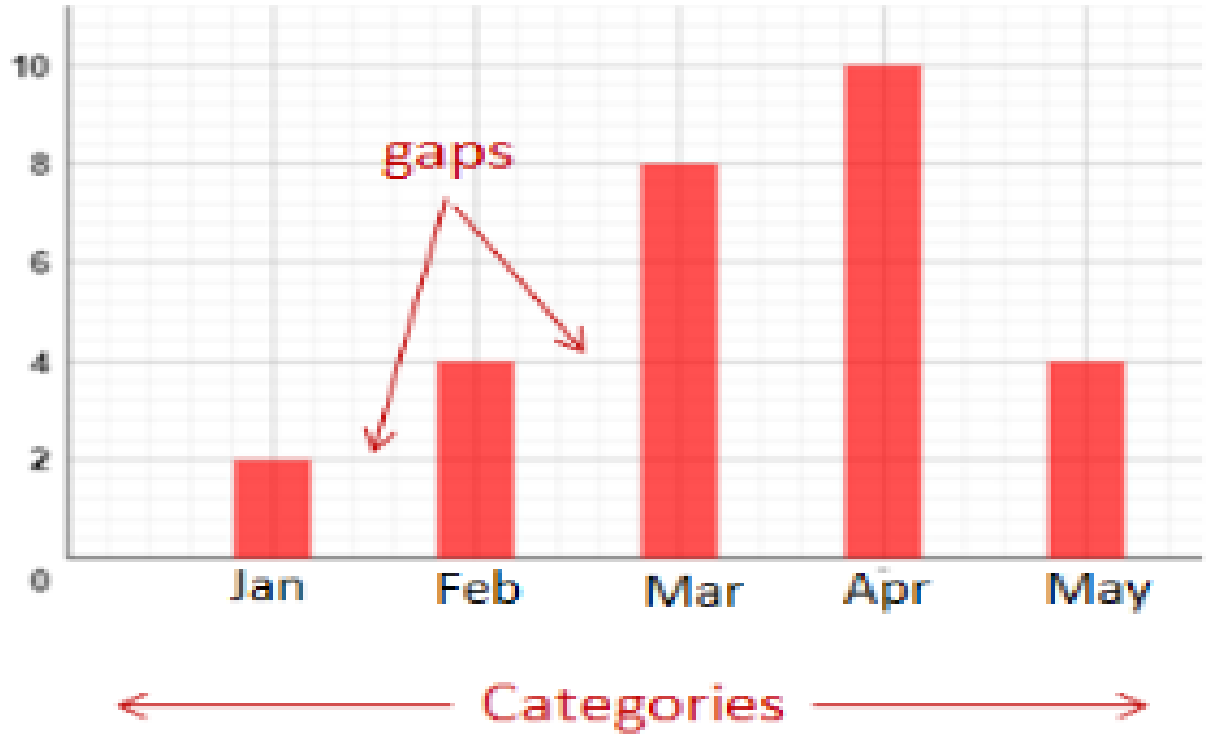


BASIS	HISTOGRAM	BAR GRAPH
Indicates	Distribution of continuous variables	Comparison of discrete variables
Spaces	Bars touch each other, hence there are no spaces between bars	Bars do not touch each other, hence there are spaces between bars.
Elements	Elements are grouped together, so that they are considered as ranges.	Elements are taken as individual entities.
Width of bars	Need not to be same	Same

Histogram vs. Bar Chart



Histogram



Bar Chart

PIE DIAGRAM

- It is also known as circular or angular diagram.
- It is a circle divided into various segments showing the percent values of a series. Each segment denotes a proportionate of the whole.
- Pie diagrams are constructed on percentage basis.
- The total value of the pie is always 100%.
- The pie chart formula is given as

(Given data /total value of data)*360°

A teacher surveys her class on the basis of their favourite sports:

Football	10
Hockey	5
Cricket	5
Basketball	10
Badminton	10

Step 1: Add all the values in the table to get the total.
i.e. Total students are 40 in this case.

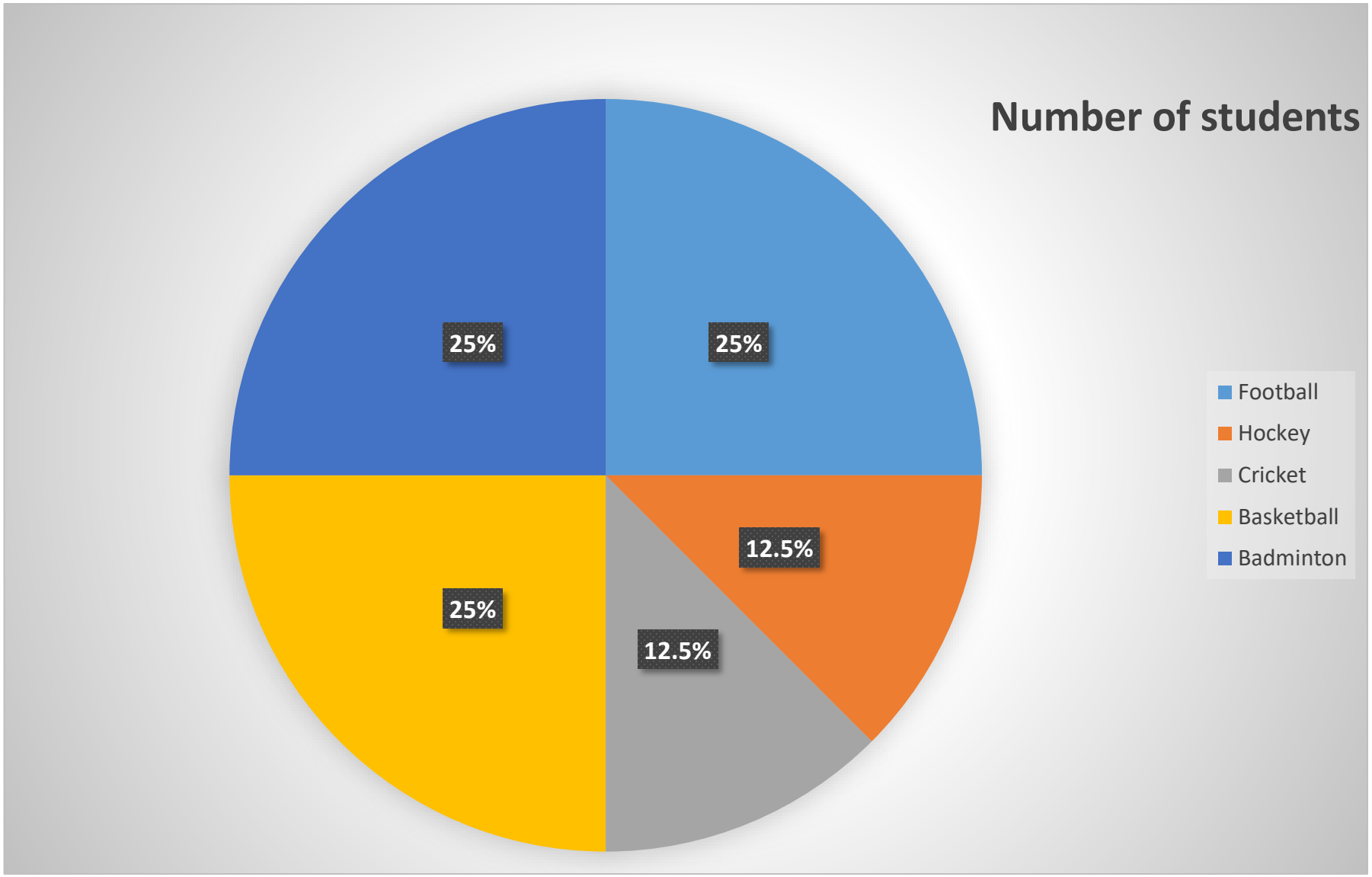
Step 2: Divide each value by the total and multiply by 100 to get a percent.

Favourite Sport	Number of students	Percent
Football	10	25% (10/40*100)
Hockey	5	12.5% (5/40*100)
Cricket	5	12.5% (5/40*100)
Basketball	10	25% (10/40*100)
Badminton	10	25% (10/40*100)
TOTAL	**40**	

Step 3: Apply the formula of pie chart i.e. (Given data /total value of data)*360°

Favourite Sport	Number of students	Percent	Pie Sector
Football	10	25% (10/40*100)	90° (10/40*360°)
Hockey	5	12.5% (5/40*100)	45° (5/40*360°)
Cricket	5	12.5% (5/40*100)	45° (5/40*360°)
Basketball	10	25% (10/40*100)	90° (10/40*360°)
Badminton	10	25% (10/40*100)	90° (10/40*360°)
TOTAL	**40**		

Step 4: Draw a circle and use the protractor to measure the degree of each sector.

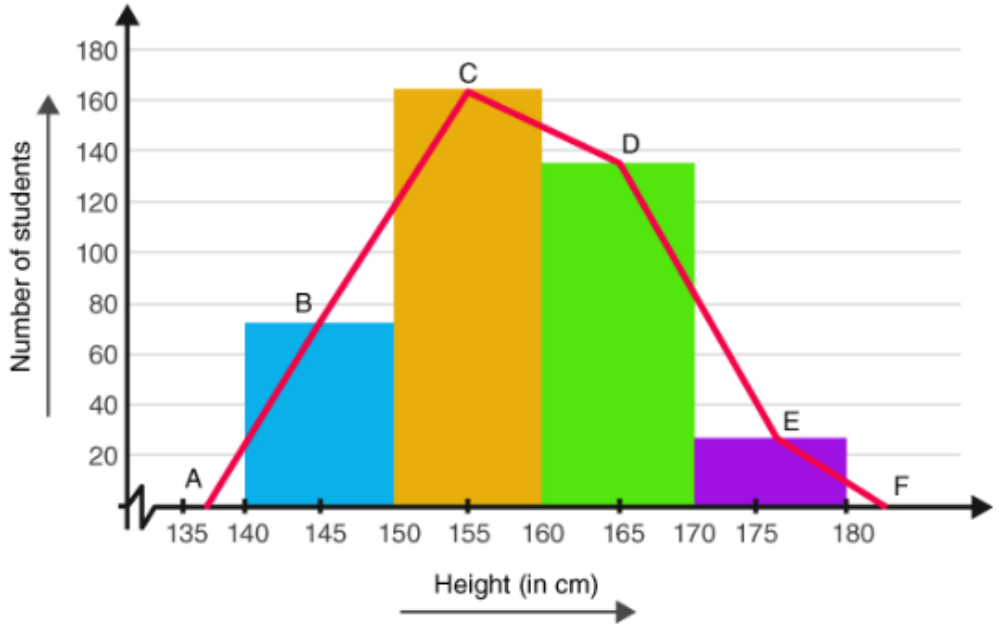

Number of students
25%
25%
12.5%
25%
12.5%
Football
Hockey
Cricket
Basketball
Badminton

FREQUENCY POLYGON:

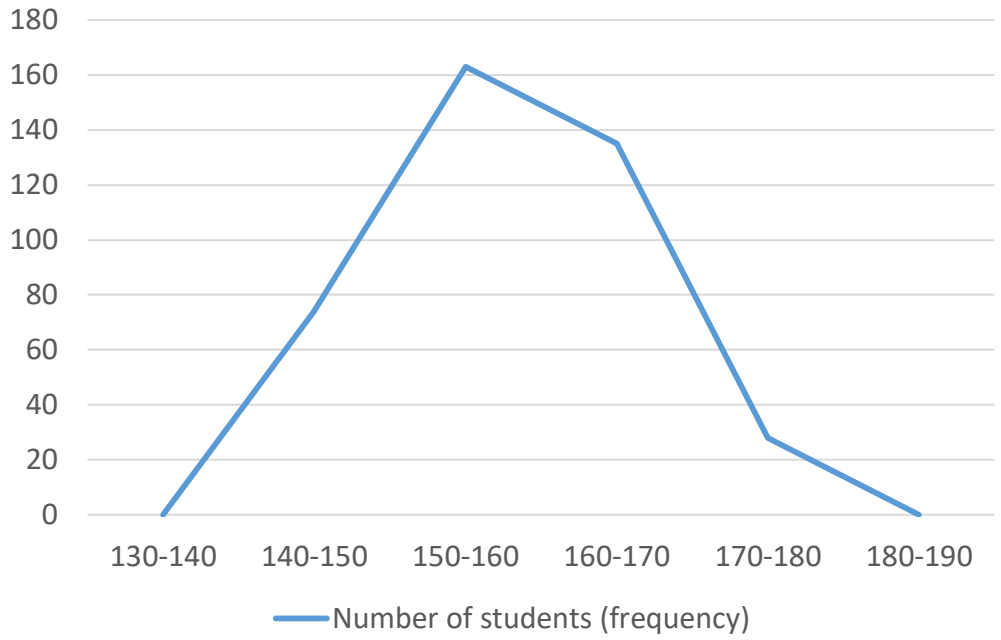
- A frequency polygon is a graph constructed by using lines to join the midpoints of each interval.
- The heights of the points represent the frequencies.
- A frequency polygon can be created from the histogram or by calculating the midpoints of the class intervals from the frequency distribution table.
- The midpoint of a class interval is calculated by adding the upper and lower boundary values of the class interval and dividing the sum by 2.

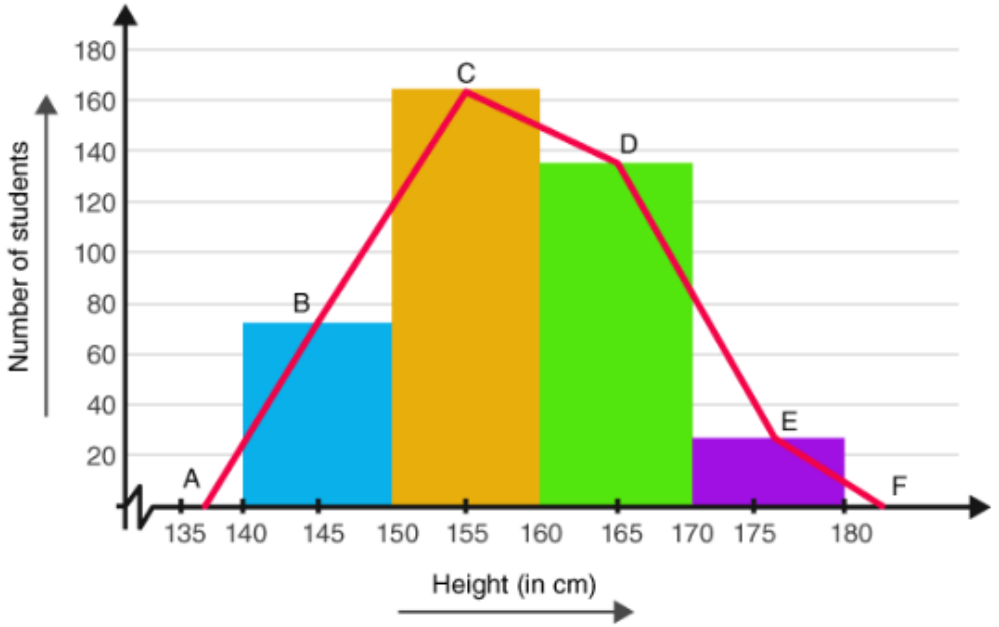
- **EXAMPLE: In a batch of 400 students, the height of students is given in the following table. Represent it through a frequency polygon.**

Height (in cms)	Number of students (frequency)
140-150	74
150-160	163
160-170	135
170-180	28
Total	400



Height (in cms)	Number of students (frequency)
130-140	0
140-150	74
150-160	163
160-170	135
170-180	28
180-190	0
TOTAL	400





OGIVES:

- Ogives are graphs that are used to estimate how many numbers lie below or above a particular variable or value in data.
- To construct an Ogive, firstly, the cumulative frequency of the variables is calculated using a frequency table. It is done by adding the frequencies of all the previous variables in the given dataset.
- The result or the last number in the cumulative frequency table is always equal to the total frequencies of the variables.
- The two methods of Ogives are:
    - Less than ogive
    - Greater or more than ogive

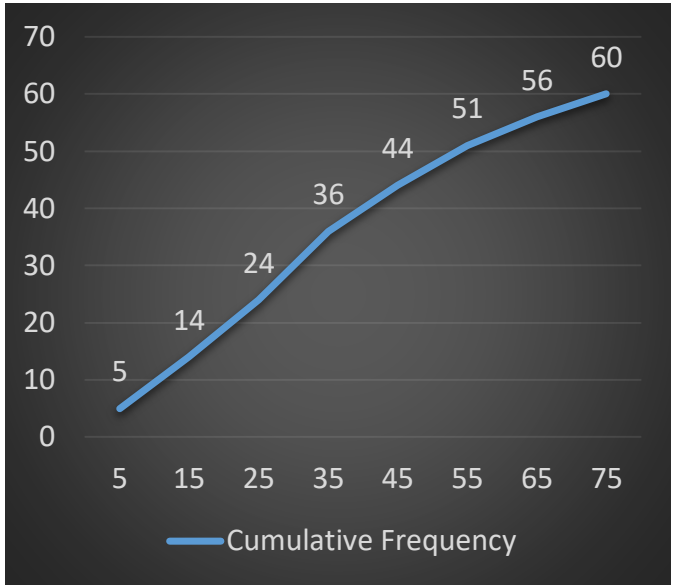
- **LESS THAN OGIVES:**
- The frequencies of all preceding classes are added to the frequency of a class. This series is called the less than cumulative series.
- It is constructed by adding the first class frequency to the second class frequency and so on.
- The downward cumulation results in the less than cumulative series.

- **HOW TO DRAW LESS THAN OGIVE CURVE:**
- Draw and mark the horizontal and vertical axes.
- take the cumulative frequencies along the y-axis and the upper class limits on the x-axis.
- Against each upper-class limit, plot the cumulative frequencies.
- Connect the points with a continuous curve.

EXAMPLE: Draw a ‘less than ogive' for the following data:

Class	Frequency
0-10	5
10-20	9
20-30	10
30-40	12
40-50	8
50-60	7
60-70	5
70-80	4

Class	MID POINTS	Frequency	CUMULATIVE FREQUENCY
0-10	5	5	5
10-20	15	9	14
20-30	25	10	24
30-40	35	12	36
40-50	45	8	44
50-60	55	7	51
60-70	65	5	56
70-80	75	4	60
		(TOTAL) 60	



- **MORE THAN OGIVE:**
- The frequencies of the succeeding classes are added to the frequency of a class. This series is called the more than or greater than cumulative series.
- It is constructed by subtracting the first class, second class frequency from the total, third class frequency from that and so on.
- The upward cumulation result is greater than or more than the cumulative series.

- **HOW TO DRAW MORE THAN OGIVE:**
- Draw and mark the horizontal and vertical axes.
- Take the cumulative frequencies along the y-axis and the lower class limits on the x-axis.
- Against each lower class limit, plot the cumulative frequencies.
- Connect the points with a continuous curve.

Class	MID POINTS	Frequency	CUMULATIVE FREQUENCY
0-10	5	5	60
10-20	15	9	55
20-30	25	10	46
30-40	35	12	36
40-50	45	8	24
50-60	55	7	16
60-70	65	5	9
70-80	75	4	4
		(TOTAL) 60