Quantitative Analysis (QA)

What Is Quantitative Analysis (QA)?

Quantitative analysis is often associated with numerical analysis where data is collected, classified, and then computed for certain findings using a set of statistical methods. Data is chosen randomly in large samples and then analyzed.

Quantitative analysis (QA) is the process data scientists and other professionals use to collect, evaluate and analyze data to understand patterns in the past, present and future.
Its tools can be applied to financial situations, such as investment opportunities, chemistry to turn hard data into numerical values.
QA results are logical, statistical and unbiased.

Understanding Quantitative Analysis (QA)

Quantitative analysis provides analysts with tools to examine and analyze past, current, and anticipated future events. Any subject involving numbers can be quantified; thus, QA is used in many fields including analytical chemistry, financial analysis, social science, and organized sports. In the financial world, analysts who rely strictly on QA are frequently referred to as "quants" or "quant jockeys.“

What is quantitative analysis used for?
Generally used for three purpose.

⇒ To measure differences between groups.
⇒ To assess relationships between variable
⇒ To test hypotheses in a scientifically rigorous.

Understanding Quantitative Analysis (QA)

A quantitative analyst's main task is to present a given hypothetical situation in terms of numerical values. Quantitative analysis helps in evaluating performance, assessing financial instruments, and making predictions.

It encompasses three main techniques of measuring data: regression analysis, linear programming, and data mining.

regression analysis:

It involves using statistical equation to predict or estimate the impact of one variable on another.

linear programming:

is quantitative method that determine how to achieve such an optimal solution.

data mining:

techniques are used to evaluate very large sets of data to find patterns or corelation concealed within them.

Quantitative analysis focuses on what. What happened?

Like: How many people bought this product?

What percentage of people considered this brand?

Quantitative analysis and research methods often include:

- **Closed-ended questionnaires and surveys**
- **Large-scale data sets**
- **Analytics gathered by machines**
- **Random sampling**
- **Structured data**
- **Tracking software such as CRMs, marketing automation, advertising**

Quantitative analysis tends to look very broadly at many things to understand the what.

Qualitative Analysis

What is a Qualitative analysis?

Qualitative analysis is concerned with the analysis of data that cannot be quantified.

This type of data is about the understanding and insights into the properties and attributes of objects (participants).

A qualitative analysis is an assessment of a certain event or a situation by the means of inquiry or observation.

Qualitative analysis focuses on why did that happen.

Like: Why do people behave in certain ways?

Why do they make decisions?

Qualitative analysis and research methods often include:

- **Focus group discussion**
- **Open-ended questionnaires and surveys**
- **Unstructured interviews**
- **Unstructured observations (like reading social media posts)**
- **Case studies**

Qualitative analysis tends to look very deeply at a few things to understand the why.

Key differences between qualitative and quantitative analysis

- **Definition of qualitative and quantitative analysis**
- **Data collection for qualitative and quantitative analysis**
- **Research methodology involved in qualitative and quantitative analysis**
- **Research findings**

QUALITATIVE ANALYSIS
VERSUS
QUANTITATIVE ANALYSIS

Qualitative analysis	Quantitative analysis
It is a subjective analysis that is more concerned with non-statistical data that cannot be computed	It is an objective analysis that quantifies data
Typical data include color, gender, nationality, religion and many more	Typical data include measurable quantities such as length, size, weight, mass and many more
The analysis is used to understand why a certain phenomenon occurs	The analysis is concerned with how many or how much a certain phenomenon occurs
Sample is small and is non-representative of the entire population	The sample is large and can be generalized to cover the entire population
Interprets and understands social interactions	Test hypotheses and give future predictions
Research methodology is exploratory	Research methodology is often conclusive

Qualitative Data (Categorical)	Quantitative Data (Numerical)
Gender	Age
Religion	Height
Qualification	Income
Native Language	University Size
Method of Treatment	Percent of Lecture Attended
Type of Teaching Approach	Group Size
Problem Solving Strategy Used	Number of Errors

Introduction to

Statistics

Introduction

- **Statistics means numerical description.**

- **The purpose of statistics is to manipulate, summarize and investigate data so that useful decision-making information results.**

ORIGIN OF STATISTICS

- **The word *'Statistics'* is derived from Latin word *'Status'*, German word *'Statistik'*, Italian word *'Statista'*, French word *'Statisttique'* each of which means political state.**

- ***In ancient periods, the beginning of statistics was made to meet the administrative needs of the state.***

- ***In modern times, statistics is not related to the administration of the alone, but it has close relation with almost all those activities of our lives which can be expressed in quantitative terms.***

Why Statistics?

- **It is estimating the present , predicting the future.**
- **Study methods that can be used to solve problems ,build knowledge.**
- **Statistics make data into information**
- **Statistics is very important in every aspects of society (Govt., People or Business)**
- **how it effect product ,process and system.**

Meaning of Statistics

The term Statistics conveys different meanings in singular and plural sense.

In plural: statistics means numerical set of data.

Definitions:

“ Statistics are numerical statements of facts in any department of enquiry placed in relation to each other ”

- Bowley

“ By Statistics we mean quantitative data affected to a market extent by multiplicity of causes ”

- Yule and Kendall

This definition includes all aspects of Statistics.

This definition has the following features.

- **Aggregates of facts:** ***Single or isolated figure cannot constitute statistics because such figures are unrelated and incomparable.***
- **Affected by multiplicity of causes:** ***The facts and figures should be affected by a set of causes.***
- **Numerically expressed:** ***information or facts which are reducible to quantitative form***
- **Enumerated or estimated according to reasonable standards of accuracy:** ***data will be exact and accurate or data may not be accurate.***
- **Collected in a systematic manner:** ***well plan of data collection should be prepared***
- **Collected for a pre-determined purpose:** ***purpose of the enquiry should be stated***
- **Placed in relation to each other:** ***The numerical data collected constitutes statistics if they are comparable***

By: Prof. Horace Secrist

In singular: statistics means the science of COPAI of the data.

Definitions:

" Statistics may be defined as the collection, Organization, presentation, analysis and interpretation of the numerical data."

-Croxton and Cowdon

" Statistics is the science which deals with the collection, classification and tabulation of numerical facts as a basis for the explanation, description and comparison of phenomena."

- Lovitt

Features of Statistics as Science

1. **Collection of data:** ***It is decided as to how, when and where and what kind of data are to be collected.***
2. **Organization of data:** ***organize the collected data.***
3. **Presentation of data:** ***To make the data intelligible, brief and attractive .***
4. **Analysis of data:** ***Various statistical tools like average, dispersion, correlation, test of significance etc.***
5. **Interpretation of data:** ***To interpret means to draw the valid conclusion from the data which has been analysed***

Functions of Statistics

1. **Statistics simplifies complexity:** ***Statistic consists of aggregate of numerical facts.***
2. **Statistics presents fact in a definite form:** ***The conclusion stated numerically is definite and hence more convincing than the conclusions stated qualitatively.***
3. **Statistics facilities comparison:** ***The science of statistics does not mean only counting but also comparison.***
4. **To help in formulation of policies:** ***in different fields mainly in economics, business etc.***
5. **Statistics helps in forecasting:** ***Statistical methods provide helpful means in forecasting the future by studying and analysing the tendencies based on passed records.***
6. **Statistics helps in formulating and testing hypothesis:** ***in formulating and testing the hypothesis for the development of new theories.***

Importance and *Uses* of Statistics

The importance of statistics in the following major areas:

1. **Statistics in Business and Industry:** ***decisions regarding business.***
2. **Statistics in the Field of Science and Research:** ***Statistics has great significance in the field of physical and natural sciences.***
3. **Statistics in the Field of Banking:** ***banking industry.***
4. **Statistics in State:** ***welfare of its people.***
5. **Statistics in planning:** ***faster rate of growth through the best use of a nation's resources.***
6. **Statistics in Economics:** ***to measure aggregates such as investment, saving, consumption, expenditure and changes in the value of money.***

Limitations of Statistics

Statistics is considered to be a science as well as an art, which is used as an instrument of research in almost every sphere of our activities.

Some of the limitations of statistics are as follows:

1. Statistics Suits to the Study of Quantitative Data Only

2. Statistical Results are not Exact: ***The task of statistical analysis is performed under certain conditions***

3. Statistics Deals with Aggregates Only: ***Statistics does not recognise individual items.***

4. Statistics is Useful for Experts Only

Classification, Tabulation and Graphical Representation

- **The collected data in any statistical investigation are usually voluminous, crude in form and are known as raw data.**
- **As such they cannot be easily understood by person and are not directly fit for further analysis and interpretation.**
- **After having collected and edited the data, the next important step is to organize it in a systematic manner.**
- **By organization of data we mean the classification and presentation of data in such a way that the data becomes easy and convenient to use and handle.**

Definitions

“ The placement of data in different groups, formed based on some characteristics or criteria, is called *classification of data*.”

“ Classification is a process of arranging data in groups or classes according to their resemblances and affinities and gives expressions to the unity of attributes that may subset among a diversity of individuals.” - R.L Conner

“ Classification is the process of arranging data into sequences and groups according to their common characteristics, or separating them into different but related parts.” -- Secrist

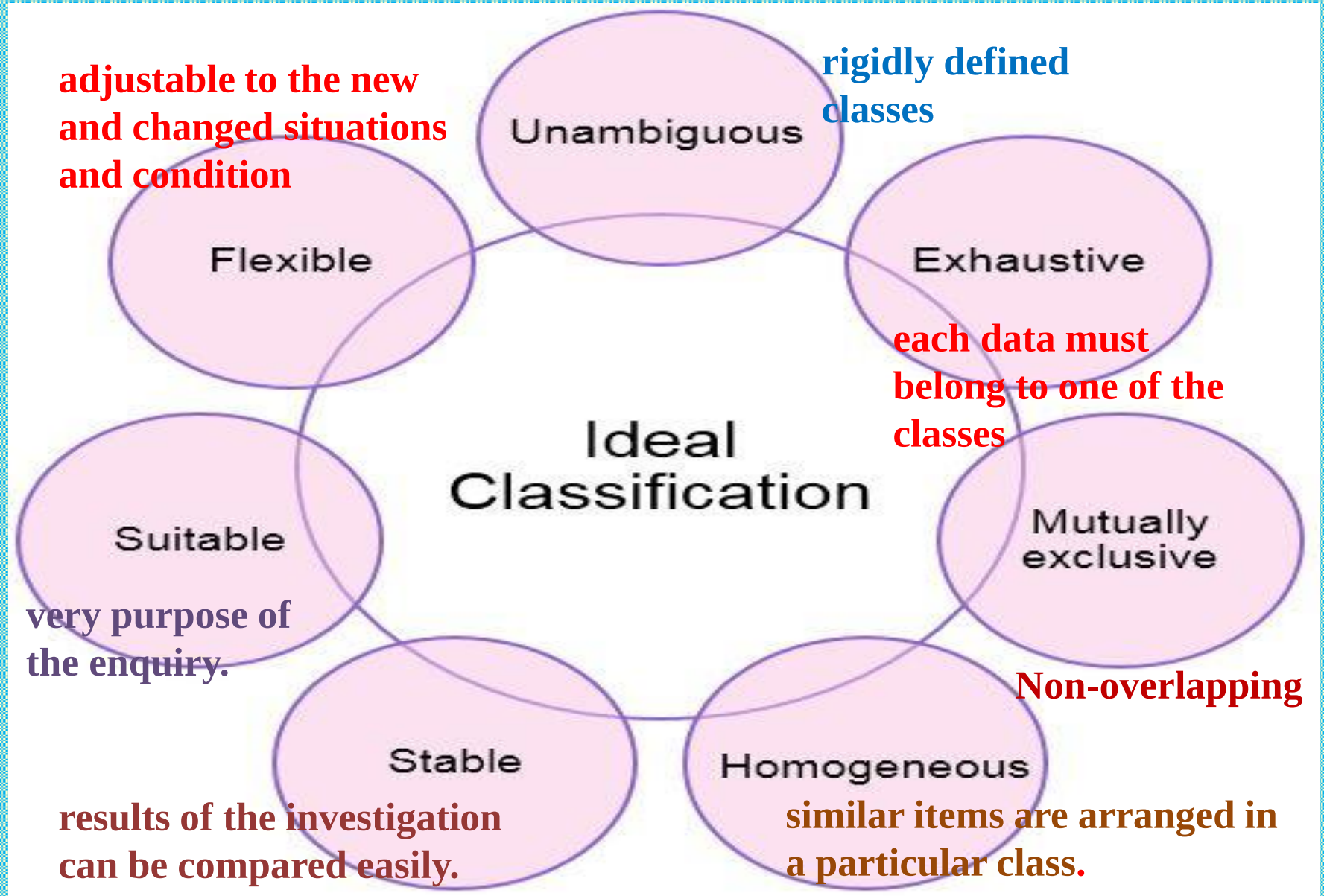
Ex: the people may be divided into different age groups like

< 10, 11 - 20, 21 - 30, 31 - 40 etc.

Functions of Classification

- ***It condenses the data:*** **Classification presents the huge unwieldy raw data in a condensed form.**
- ***It facilitates comparisons:*** **Classification enables us to make meaningful comparisons depending on the basis or criterion of classification.**
- ***It helps to study the relationships:*** **The classification of the given data *w.r.t.* two or more criteria.**
- ***It facilitates the statistical treatment of the data:*** **The arrangement of the voluminous heterogeneous data into relatively homogeneous groups.**

Rules of Classification



Important types of Classification

Geographical (i.e. based on area or region wise)
e.g.: village, city etc.

Chronological (based on time of occurrence, starting from the earliest period to the latest period/ Temporal)
e.g.: national income, annual output of rice, monthly expenditure of a household etc.

Qualitative (based on character / attributes)
e.g.: gender, qualification, color etc.

Quantitative (on the basis of magnitude i.e. numerical)
e.g.: height, weight, age etc.

Geographical Classification

In geographical classification, the classification is based on the geographical regions.

***Ex*: Sales of the company (In Million Rupees) (region – wise)**

Region	Sales
North	285
South	300
East	185
west	235

Chronological Classification

If the statistical data are classified according to the time of its occurrence, the type of classification is called chronological classification.

Annual sales reported by a departmental store

Month	Sales (Rs. in lakh)
January	22
Feb	26
Mar	32
Apr	25
May	27
Jun	29
Jul	30
Aug	30

Qualitative Classification

In qualitative classifications, the data are classified according to the presence or absence of attributes in given units. Thus, the classification is based on some quality characteristics / attributes.

e.g.: ***Gender, Literacy, Education, Class grade etc.***

Further, it may be classified as

a) Simple classification (Dichotomy or Two-fold Classification)

b) Manifold classification

***Simple classification (*Dichotomy or Two-fold Classification):**

based on presence or absence of an attribute, the data are classified into two classes- one possessing that attribute, and the other not possessing that attribute, it is called two-fold or dichotomous classification.

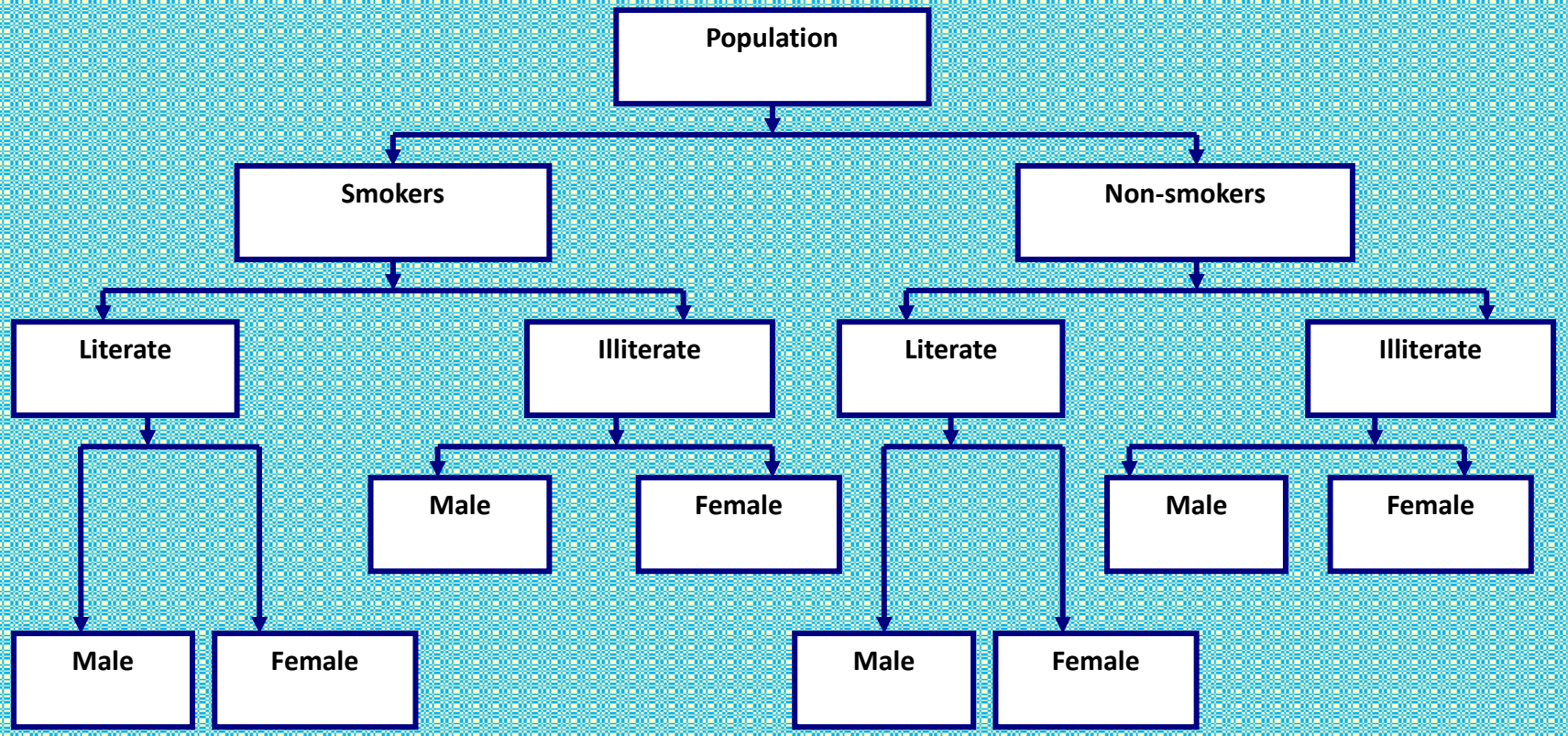
OR **If the classification is done into only two classes, then classification is known as simple classification.**

e.g.: **a) Population into Male / Female**

b) Population into Educated / Uneducated

Manifold classification:

In this classification, the classification is based on more than one attribute at a time.



Quantitative Classification

In Quantitative classification, the classification is based quantitative measurements of some characteristics,

Ex: age, marks, income, production, sales etc.

The quantitative phenomenon under study is known as variable and hence this classification is also called as classification by variable.

e.g.: For a 50 marks test, Marks obtained by students as classified as follows

Marks	No. of students
0 – 10	5
11 – 20	7
21 – 30	10
31 – 40	25
41 – 50	3
Total Students = 50	

Variable Classification: **The quantitative phenomena under study is known as variable and classification on this basis is known as variable classification.**

(a) Continuous variable (b) Discrete Variable

Continuous Variable: Those variables which can take all the possible values in a give specified range are termed as continuous variable.

Classification based on the basis of Continuous values	
Income (Rs.)	No. of Employees
1000-1500	15
1501-2000	33
2001-2500	22
2501-3000	18
3001-3500	12
Total	100

***Discrete Variable*: Those variables which cannot take all the possible values within a given specified range are termed as discrete variables.**

Example: the marks in a test (out of 100) of a group of students is a discrete variable since in this case marks can take only integral values from 0 to 100. It can not take all the values between 0 to 100.

Classification based on Discrete Values	
Height (cms.)	No. of Students
154	8
155	10
156	6
157	2
158	12
159	12
Total	50

Meaning and Definition of Tabulation

- **Tabulation may be defined as systematic arrangement of data is column and rows.**
- **It is designed to simplify presentation of data for the purpose of analysis and statistical inferences.**
- **It is the logical listing of related quantitative data in vertical columns and horizontal rows with explanatory and qualifying words, phrases and statements along with titles, heading and notes.**
- **It attempts to furnish the maximum information in the minimum possible space, without sacrificing the quality and usefulness of the data.**
- **It is an intermediate process between the collection of the data and the statistical analysis.**

Major Objectives of Tabulation

- To simplify the complex data
- To facilitate comparison
- To economies the space
- To draw valid inference / conclusions
- To help for further analysis

“ Table involves the orderly and systematic presentation of numerical data in a form designed to elucidate the problem under consideration”

Prof. L.R.Connor

“Table in its broadest sense is an orderly arrangement of data in columns and rows.”

Prof. M.M. Blaire

Parts of a Table

1. Table Number:-

A table should be numbered for easy reference and identification.

This number, if possible, should be written in the center at the top of the table.

Sometimes it is also written just before the title of the table.

2.Title of the Table:-

A table should have a suitable title which is placed centrally on the top of a table just below the table number or just after table number in the same line in bold letters.

The title describes the contents of the table briefly.

As far as possible, the title should be complete and unambiguous as regards the subject matter of the data.

It should be clear, properly worded and self-explanatory.

3. Captions or Column Headings:-

A word or phrase which explains the contents of a column of a table is called the caption.

A caption should be placed at the middle of the column.

Under a caption, there may be subheads.

When the items in the different columns are expressed in different units of measurement, the corresponding unit should be mentioned with the captions.

4.Stubs or Row Designations:-

Stubs stands for brief and self-explanatory headings of horizontal rows.

Normally, a relatively more important classification is given in rows.

Also, a variable with many classes is usually represented in rows.

5. Head Notes :–

It is a statement given below the title which clarifies the contents of the table.

It provides an explanation concerning the entire table or main parts of it.

6. Body:-

The body of the table is the most important part of the table.

The data is arranged in this part according to the description given by the captions and stubs.

7. Footnotes:-

Footnotes are given at the foot of the table for explanation of any fact or information included in the table which needs some explanation. Thus, they are meant for explaining or providing further details about the data, that have not been part of the table.

The data is arranged in this part according to the description given by the captions and stubs.

8. Source Note:-

It appears below the footnotes if they are used.

The source notes are indicated, in case secondary data are used.

These state the sources form which the data has been obtained.

Or The source note should give the name of the author, table number, volume number, page number and the place of the publisher of the source from which data have been collected.

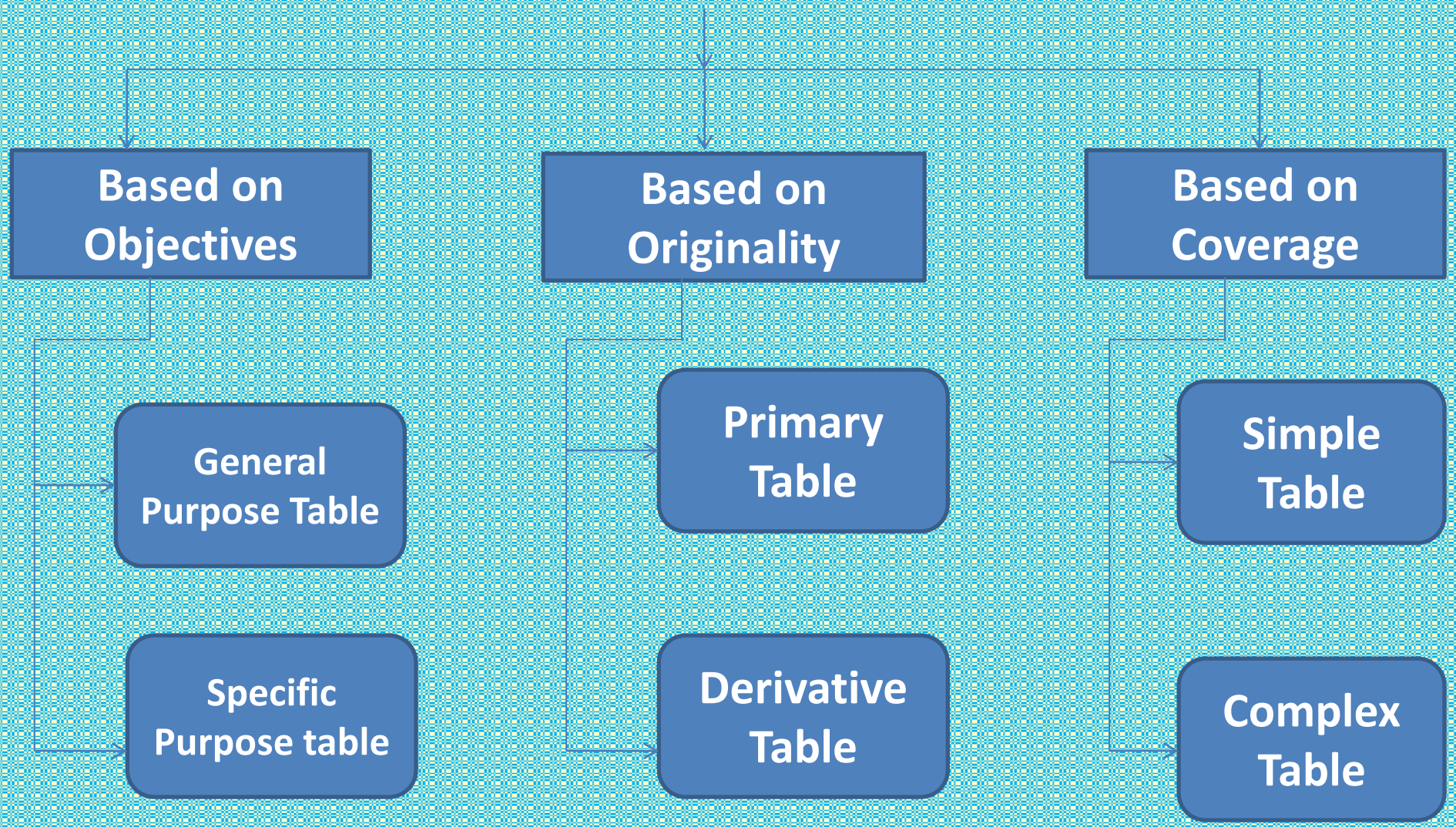
Table 5. YOUTH ACTIVITIES
Pune youth, Jan 2023 and Mumbai youth, 2020

	Pune youth, Jan 2023	Mumbai youth, 2020
Read Book and News Papers almost daily	75%	40%
Active on social media	70%	90%
Watch TV almost daily	50%	35%
Get together with friends almost weekly	60%	80%
Exercise almost daily	75%	70%

A Study of Lifestyle and values of youth,2020 n=2000.

Source: India Today, Monitoring of future.

Kinds of Tables



Based on Objective

General Purpose Tables: They are reference tables which are also called "information tables". They are designed for use by research workers, statisticians and government agencies. They serve as a repository of information and are arranged for ease reference. They tell facts in general with no particular purpose.

Specific purpose tables: They are summary tables which are also called "interpretative tables". They are analytical in nature and prepared with an idea of making comparative studies.

Based on Originality

Primary Tables: The primary tables are those tables in which facts are expressed in the original form. They contain actual and absolute figures. They are also called "original tables".

Derived or Derivative Tables: Derived table is one which contains figures and results derived from the original or primary data.

It expresses the information in terms of ratios, percentages, aggregates, averages and other statistical measures.

Based on Coverage

Simple Tables:

The simple tables are those tables which furnish information about only a single characteristic of the data.

They are popularly known as "one-way tables".

The following is an example of simple table:

Table Showing the Number of Employees in Dena Bank according to Age Group	
Age in Years	No. of employees
Below 30	17
30-40	20
40-50	10
50 and above	3
Total	50

Based on Coverage

Complex Tables: If the data are grouped based on two or more characteristics or criterion simultaneously, then we get a complex table. The complex table may be of the following types

(a) Two-way Tables

(b) Three-way Tables

(c) Manifold Tables

(a) Two-way Tables:- They are called "doubled tables" which furnish information about two characteristics of the data. The two attributes of the data are interrelated for e.g., the number of students—male and female, male– married and unmarried, and female– married and unmarried.

Table Showing the Number of Employees in Dena Bank according to Age, Gender

Gender / Age in Years	No. of employees		Total
	Male	Female	
Below 30	10	7	17
30-40	11	9	20
40-50	3	7	10
50 and above	1	2	3
Total	25	25	50

(b) Three-way Tables:- The three-way tables are those in which the data are classified with reference to three characteristics.

They furnish information regarding three interrelated attributes of the same data.

e.g., the number of students male and female, married and unmarried, and graduates and undergraduates.

The following is an example of 'Three Way table':

Table Showing the Number of Employees in Dena Bank according to Age and Gender and Martial Status

Gender / Martial Status	No. of employees						Total
	Male			Female			
Age Group	Married	Unmarried	Total	Married	Unmarried	Total	
Below 30	3	7	10	3	4	7	17
30-40	3	8	11	7	2	9	20
40-50	2	1	3	7	--	7	10
50 and above	1	--	1	2	--	2	3
Total	9	16	25	19	6	25	50

(c) Manifold Tables:- These are high order tables having more than three characteristics of the same data which are interrelated. Manifold or higher order tables are commonly used in presenting population census data.

Two-way Table	Three-way Table	Manifold Table
Data is classified on the basis of two inter-related characteristics of a specific phenomenon.	Classification is done according to three characteristics of the data that are inter-related.	More than three characteristics of data of a particular phenomenon is required in this data classification.
It is also known as Double Table.	This table is also called as Treble Table.	This is also known as Higher Order Table.
Groups of data cannot be present.	Groups of data can be present.	This is considered as the most complex form of a table.

STEM and LEAF Display of Data

Suppose we have a set of *n* observations *x1, x2, x3,..., xn* •

A simple and easy way to represent these data is the stem and leaf display.

***e.g.*: we have eight two digits values as 43, 48, 52, 66, 53, 47, 68, 57. The set of observations can be presented in the form of stem and leaf display as given below by taking the digits in the tenth place as steam and m the unit place as leaf.**

Stem	Leaf
4	3,8,7
5	2,3,7
6	6,8

***Example*. The data below give the monthly output of coal in India after independence from July 1948 to June 1951.**

Monthly Output (Million Tones)

22.3, 22.4, 25.4, 24.5, 26.7, 23. 4, 23.6, 25.7

Stem	Leaf
22	.3, .4
23	.6
24	.5
25	.4, .7
26	.7

Differences between Classification and Tabulation

BASIS FOR COMPARISON	CLASSIFICATION	TABULATION
Meaning	Classification is the process of grouping data into different categories, on the basis of nature, behavior, or common characteristics.	Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.
Order	After data collection	After classification
Arrangement	Attributes and variables	Columns and rows
Purpose	To analyse data	To present data
Splitting (data into two)	Categories and sub-categories	Headings and sub-headings

Ex: Tabulate the given information:

(i) In 2018 , out of a total of 1,500 applicants in a college, 900 were from Commerce background. The number of girls was 600 , out of which 230 were from Science stream.

(i) In 2019 , the total number of applicants was 2,500 of which 1,800 were boys. The number of students from Science stream was 1,600 of which 450 were girls.

Year: 2018
Applicants: 1500
Commerce : 900
Girls: 600
Boys: 1500-600=900

Girls:
Science: 230
Commerce:600-230=370

Boys:
600-230=370
900-370= 530

Year: 2019
Applicants: 2500
Boys: 1800
Science: 1600
Girls: 450

Girls:
Science: 450
Commerce:900-650=250

Boys:
1600-450=1150
1800-1150= 650

Table 1. ADMISSION STREAM WISE

Total number of applicants in a college (according to gender and stream and year of admission)

STREAM	2018			2019			TOTAL		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
Science	370	230	600	1150	450	1600	1520	680	2200
Commerce	530	370	900	650	250	900	1180	620	1800
Total	900	600	1500	1800	700	2500	2700	1300	4000

Ex:

Present the following information in a suitable tabular form, suppling the figures.

In 1995, out of total 2000 workers in a factory, 1550 were members at a trade union. The number of women workers employed was 250, out of which 200 did not belong to any trade union.

In 2000, the number of union workers was 1725 of which 1600 were men. The number of non-union was 380 among which 155 were women.

Table 1. TRADE UNION

Comparative study of the membership of trade union in a factory in 1995 and 2000.

Year →	1995			2000		
Tread Union ↓	Male	Female	Total	Male	Female	Total
Members	1500	50	1550	1600	125	1725
Non-members	250	200	450	225	155	380
Total	1750	250	2000	1825	280	2105

RULES FOR DRAWING GRAPHS AND DIAGRAMS:

- First choose the form of diagrams /graphs which is capable of representing the given set of data.
- Title- gives information of diagrams or graphs contain.
- Scale – selection of scale should be neither too small or too large. The scale should also specify the size of unit and what it represents. (eg: No. of persons in thousands).
- Neatness
- Attractive – different types of lines or shades, colours etc can be used to make the pictures more attractive.
- Originality – helps the observer to see the details with accuracy
- Simplicity –good diagram depends upon ease with which the observer can interpret it.
- Economy – cost and labour should be exercised drawing a diagram.

Diagrammatic and Graphic Representation

Presentation of Data

A visual form for presentation of statistical data.

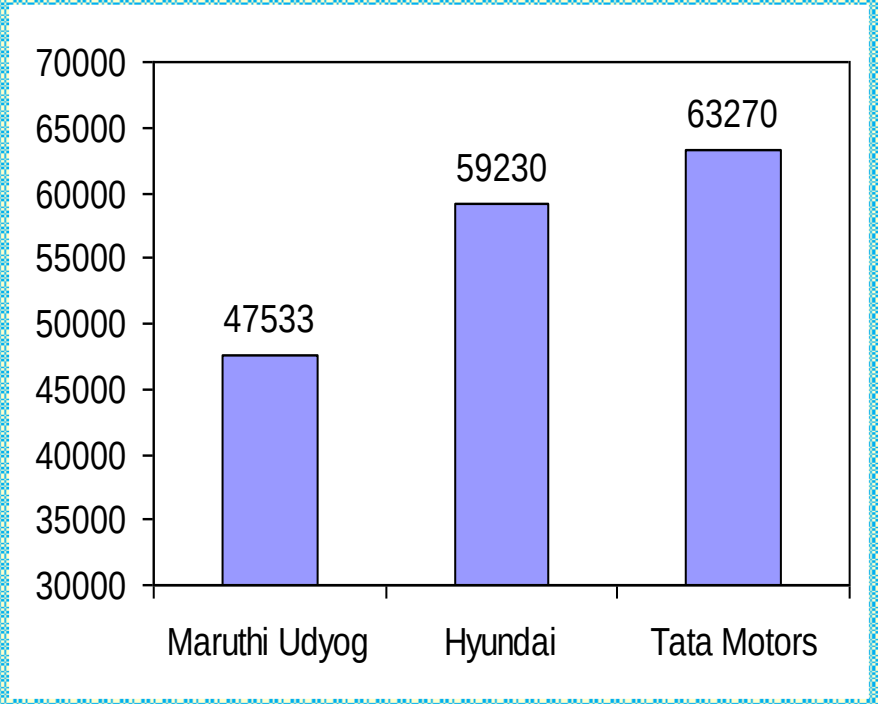
Diagrams	Graphs
Bar Diagram ✓ Simple Bar Diagram ✓ Multiple Bar Diagram ✓ Component Diagram ✓ Percentage Bar Diagram ✓ Deviation Bar Diagram Pie Diagram	✓ Histogram ✓ Frequency ✓ Polygon ✓ Line graph

Simple Bars Diagram

The annual expresses of maintaining the car of various types are given below. Draw the vertical bar diagram. The annual expenses of maintaining includes
(fuel + maintenance + repair + assistance + insurance).

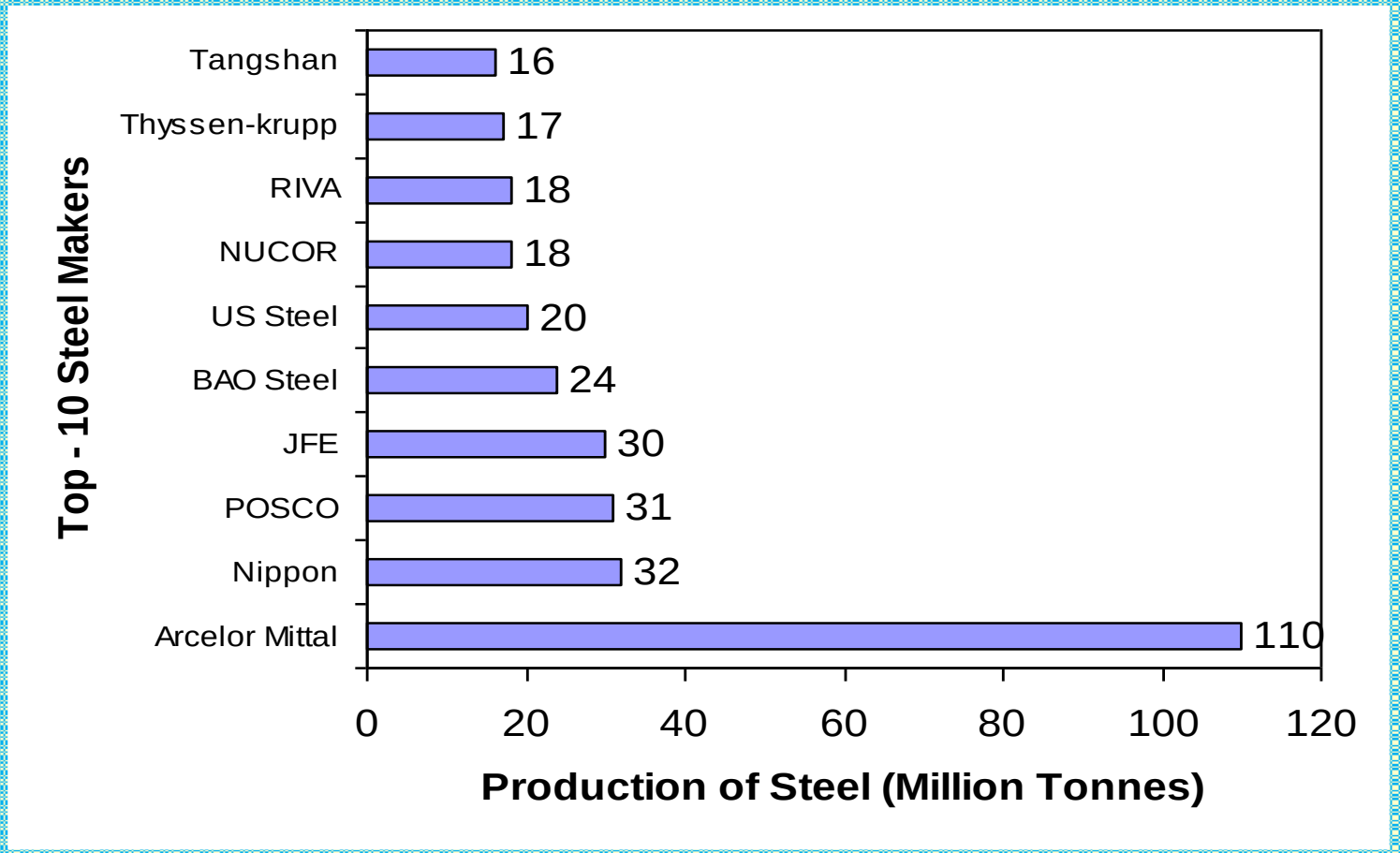
Type of the car	Expense in Rs. / Year
Maruthi Udyog	47533
Hyundai	59230
Tata Motors	63270

Source: TCS Study



Horizontal bar diagram

World biggest top 10 steel makers are data are given below.
Draw horizontal bar diagram.



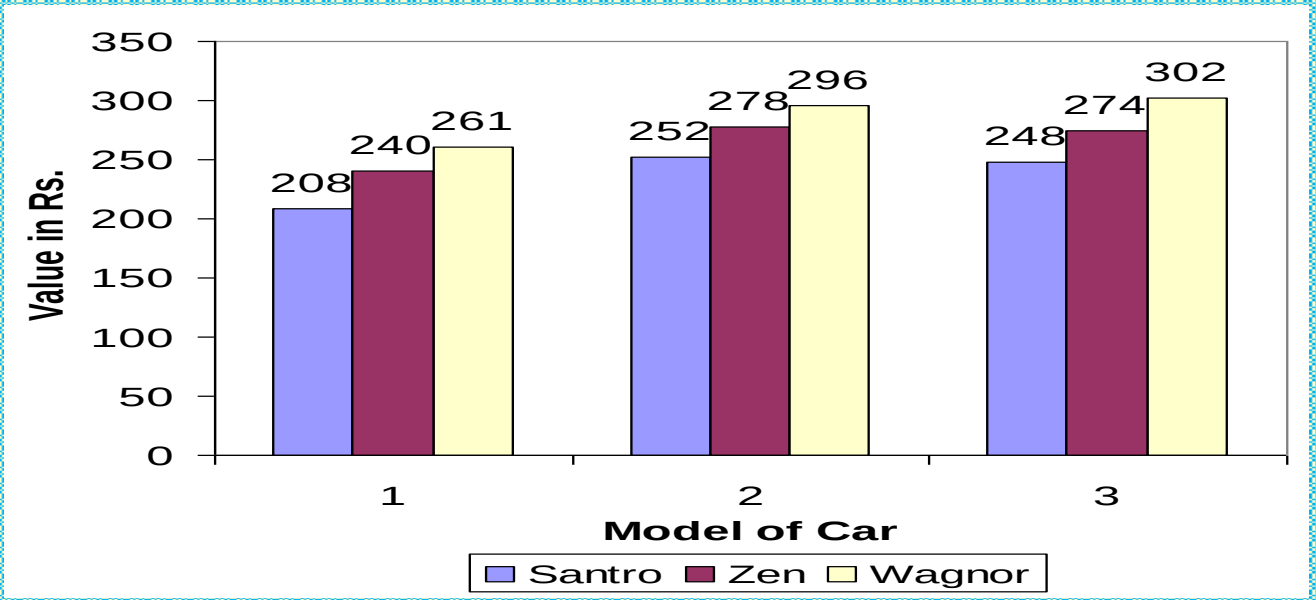
Source: **India Today**

Compound bar diagram (Multiple bar diagram)

Ex: Draw the bar diagram for the following data.

Resale value of the cars (Rs.) are as follows.

Year (Model)	Santro	Zen	Wagonr
2003	208	252	248
2004	240	278	274
2005	261	296	302

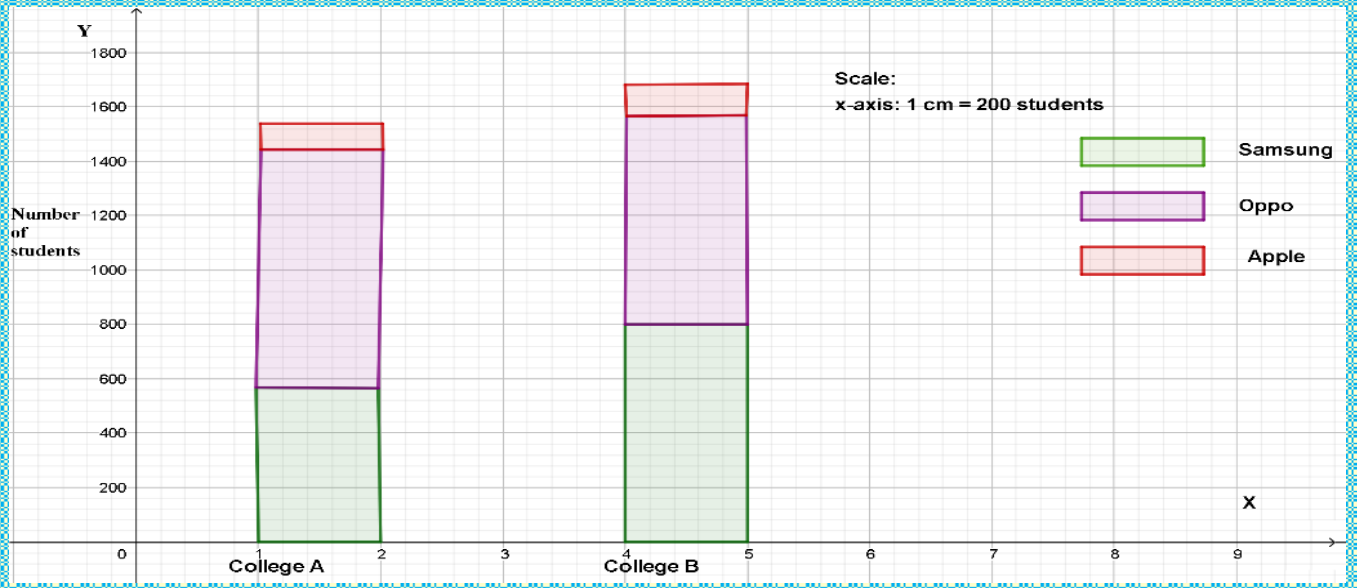


***Source*: True value used car purchase data**

Subdivided or Component Bar Diagram

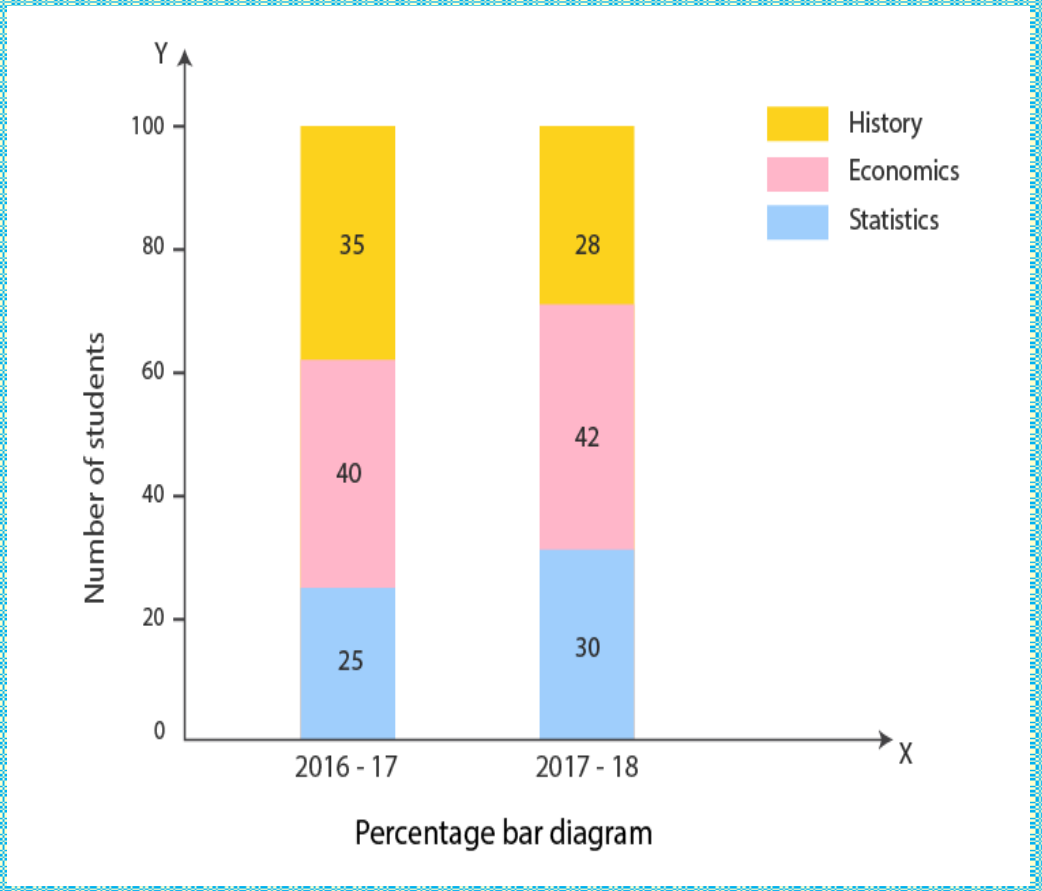
A sub-divided bar diagram is used to present data having 2 or more components. Sub-divided or component bar chart is used to represent data in which the total magnitude is divided into different or components.

Products	Number of Students	
	College A	College B
Samsung	590	800
Oppo	880	750
Apple	100	150



PERCENTAGE BAR DIAGRAM

Subject	2016–17		2017–18	
	Number of students (%)	Cumulative percentage	Number of students (%)	Cumulative percentage
Statistics	25	25	30	30
Economics	40	60	42	72
History	35	100	28	100
Total	**100**		**100**	



DEVIATION BARS

Deviation bars are specially useful for graphic presentation of net quantities *viz*., surplus or deficit, *e.g*., net profit or loss, net of imports and exports which have both positive and negative values. The positive deviations (*e.g*., profits, surplus) are presented by bars above the base line while negative deviations (loss, deficit) are represented by bars below the base line.

Pie Diagram

Steps in the construction of pie diagram

Step 1. Find the value of each category or component or group as percentage of total of all categories or components or groups.

Step 2. Calculate degree of the angle formed by each category or component or group by the formula given below.

Degree for a particular category/component/group

= Value of the group / Total of all groups * 360^0

Step 3. Take a circle of a suitable size and draw radius.

Step 4. Now draw angles calculated in step 2 with the help of a protractor.

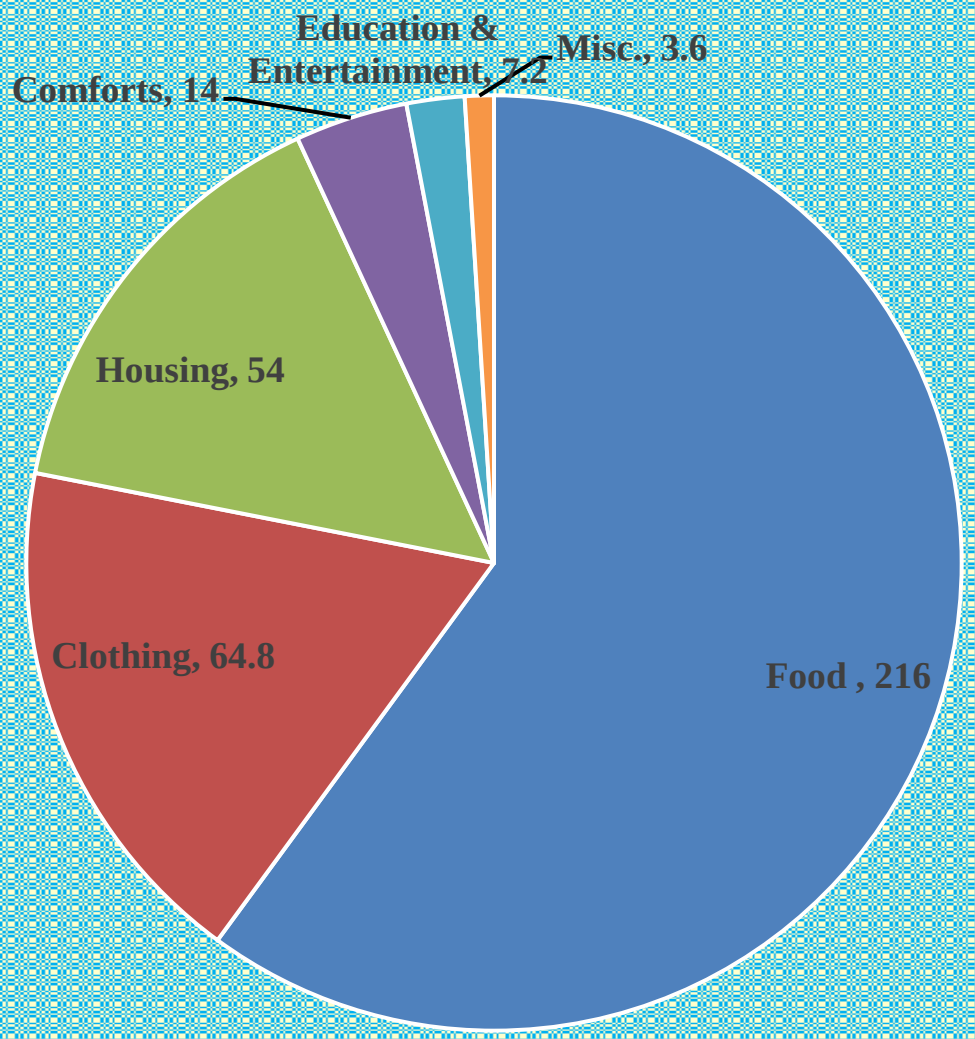
Step 5. Shade or color different segments suitably or make the distinctions between different categories or components or groups.

Step 6. For each category or component or group put the percentage in the pie diagram

Example: From the monthly budget of an industrial worker of Mumbai Industrial Area, it was found that the family spent . 360 on food, 108 on clothing, 90 on housing, 24 on comforts, 12 on education and entertainment and 6 on miscellaneous items. Construct a pie-diagram

Table 7. MONTHLY FAMILY BUDGET
Monthly Family Budget of an Industrial Worker of Mumbai

Items	Expenditure	Percentage	Degree = Value of items / Total value *360
Food	360	360 / 600 * 100 = 60	$60/100*360^{o}= 216.0^{0}$
Clothing	108	108 / 600 * 100 = 18	$18/100*360^{o}= 64.8^{0}$
Housing	90	90 / 600 * 100 =15	$15/100*360^{o}= 54.0^{0}$
Comforts	24	24 / 600 * 100 = 4	$4/100*360^{o}= 14.4^{0}$
Education & Entertainment	12	12 / 600 * 100 = 2	$2/100*360^{o}= 7.2^{0}$
Misc.	6	6 / 600 * 100 = 1	$1/100*360^{o}= 3.6^{0}$
Total	600	100	360.0^{0}


Education & Entertainment, 7.2
Misc., 3.6
Comforts, 14
Housing, 54
Clothing, 64.8
Food , 216

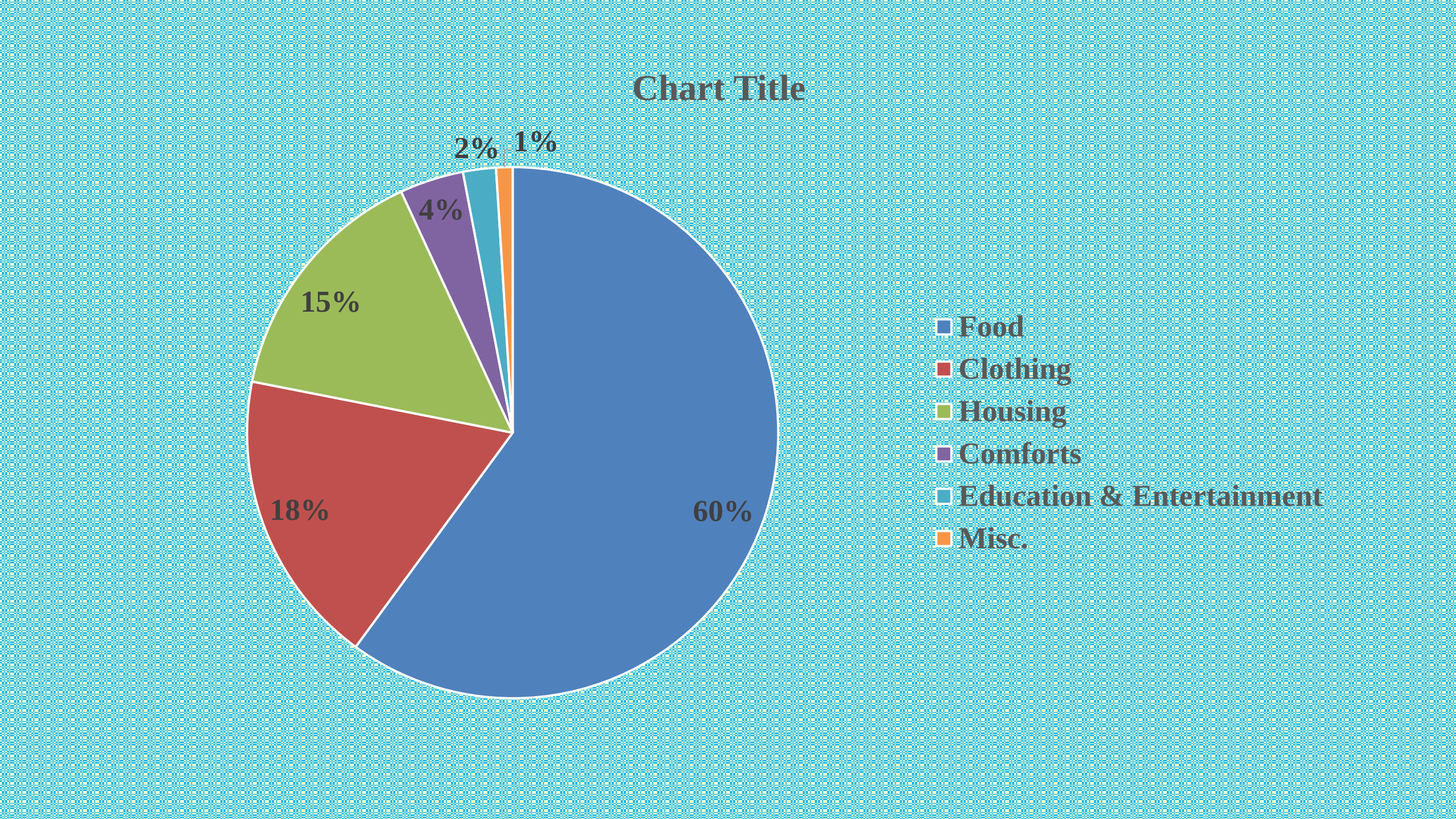


Chart Title
2%
1%
4%
15%
18%
60%
Food
Clothing
Housing
Comforts
Education & Entertainment
Misc.

Histogram:

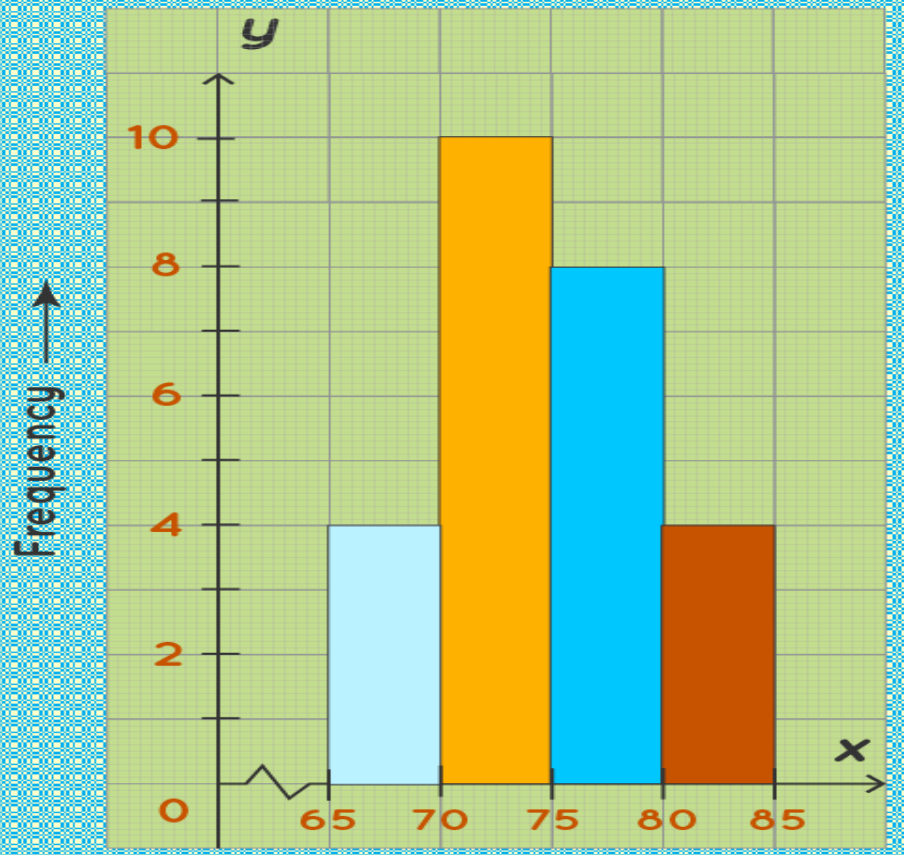
- Represented by a set of rectangular bars
- Variables (Class) is taken along the X-axis & frequency along the Y-axis.
- With the class intervals as base, rectangles with height proportional to class frequency are drawn.
- The set of rectangular bars so obtained gives histogram.

Note :

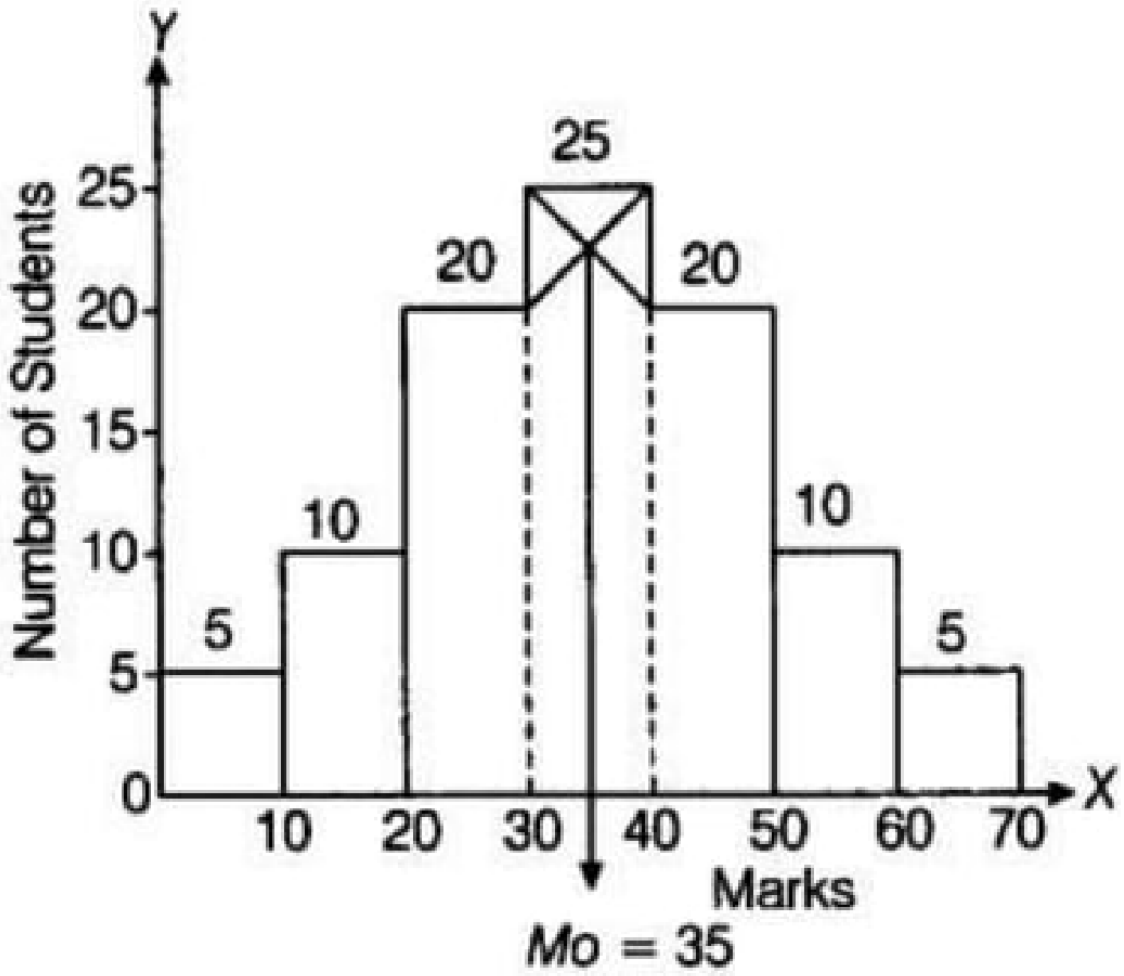
- The total area of the rectangles in a histogram represent total frequency.
- If the frequency distribution has inclusive class intervals, they should be converted into exclusive type

- Mode of the distribution can be obtained from the histogram (from the highest rectangular bar)

Weights (in lbs)	Frequency (Number of students)
65 - 70	4
70 - 75	10
75 - 80	8
80 - 85	4

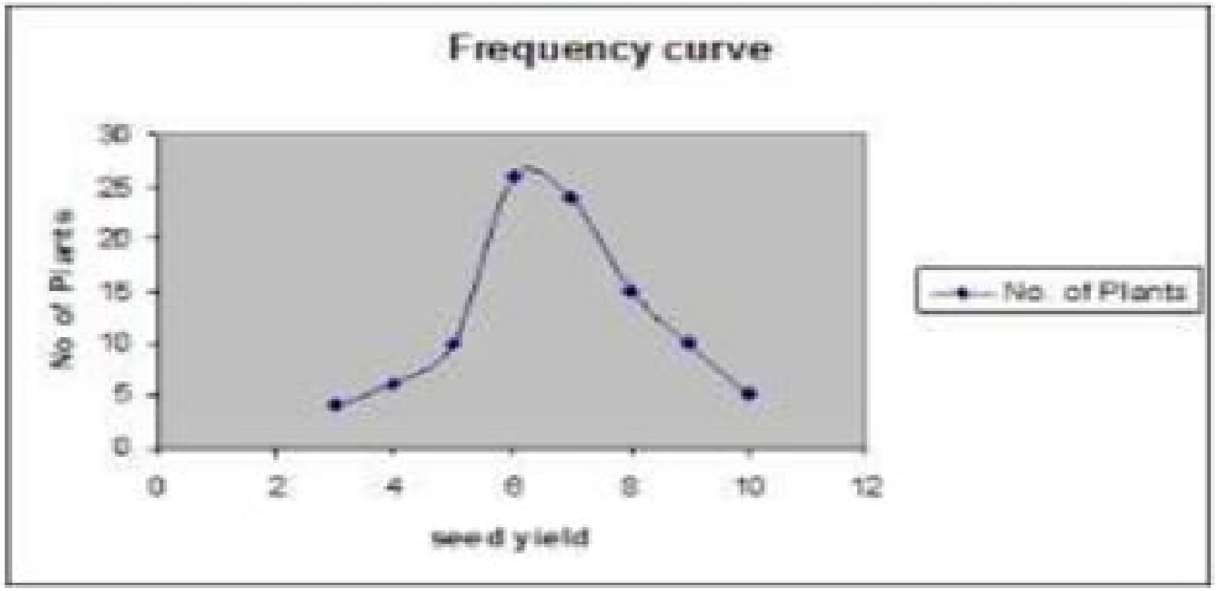


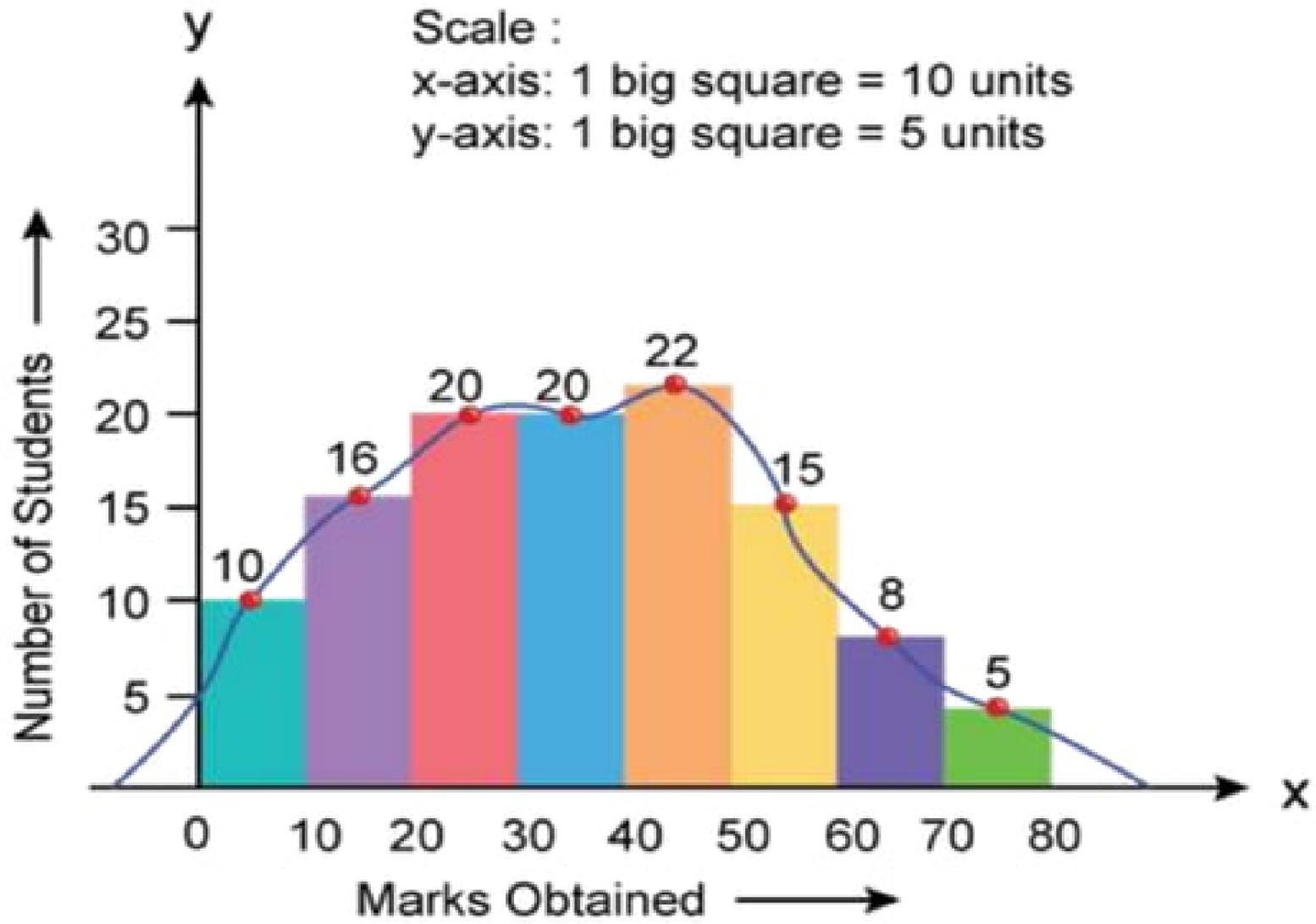
Mode of the distribution can be obtained from the histogram



FREQUENCY CURVE:

- Variables is taken along the X-axis and frequencies along Y-axis. .
- Frequencies are plotted against the class mid-values and then, these points are joined by a smooth curve.
- The curve so obtained is the frequency curve.
- Total area under the frequency curve represents total frequency.

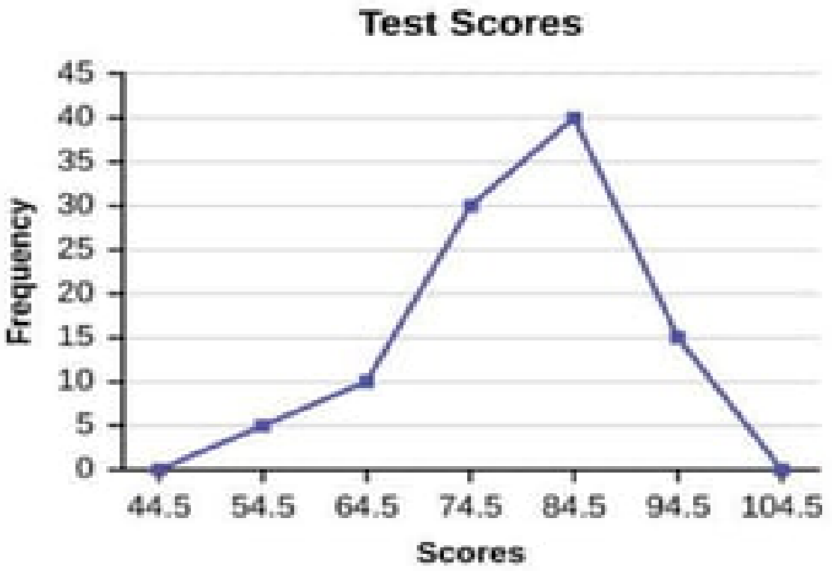



Scale :
x-axis: 1 big square = 10 units
y-axis: 1 big square = 5 units
y
x
Number of Students
Marks Obtained
30
25
20
15
10
5
0
10
20
30
40
50
60
70
80
10
16
20
20
22
15
8
5

Frequency Polygon:

- Variables is taken along the X-axis and frequencies along the Y-axis.
- Class frequencies are plotted against the class mid-values and then,

  these points are joined by Straight line.
- The figure so obtained is the frequency polygon.
- Total area under the frequency curve represents total frequency.

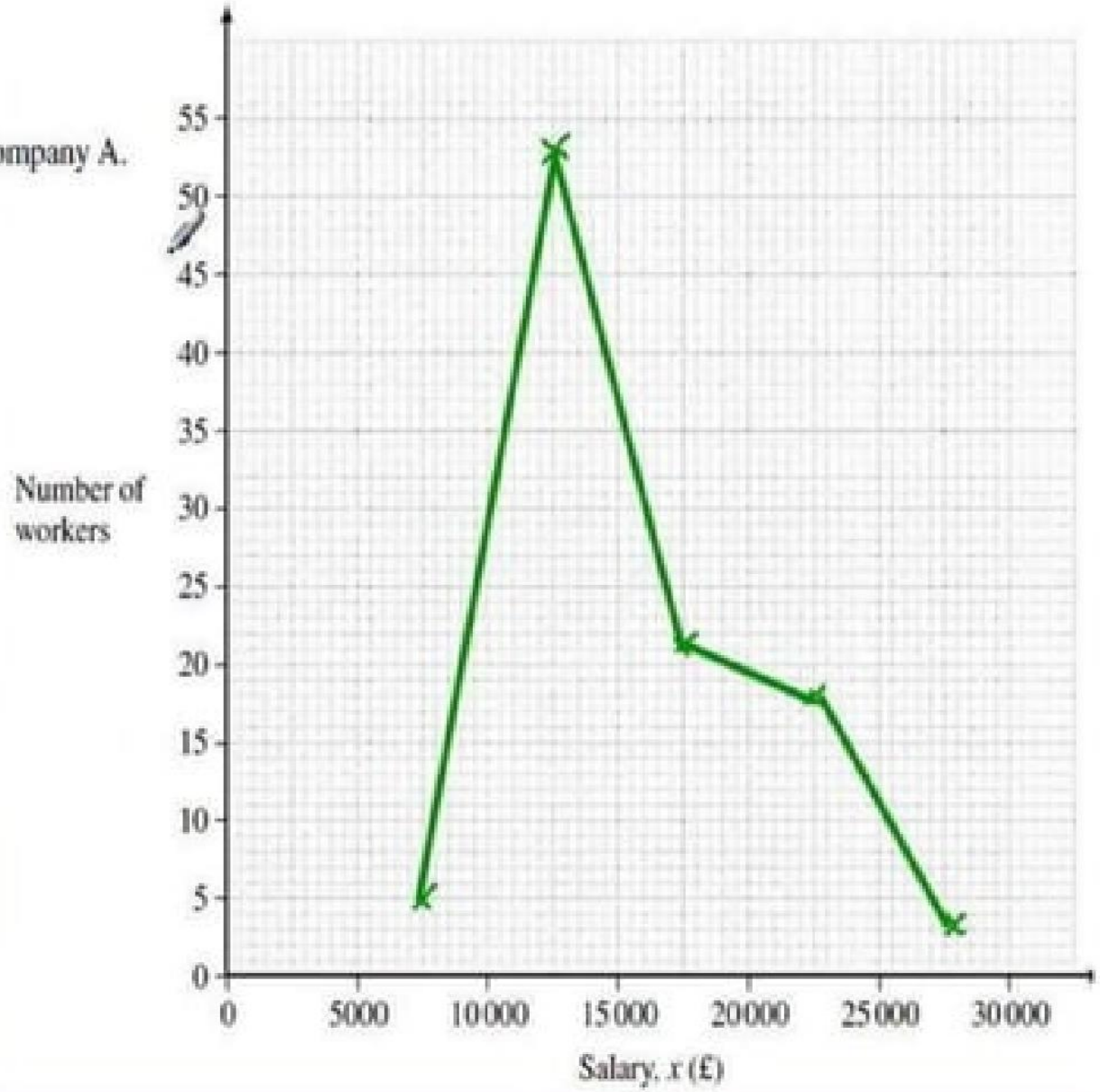


Frequency Polygon

The table shows the annual salaries of 100 office workers in Company A.

Salary, x (£)	Number of workers
$5000 < x \leqslant 10000$	5
$10000 < x \leqslant 15000$	53
$15000 < x \leqslant 20000$	21
$20000 < x \leqslant 25000$	18
$25000 < x \leqslant 30000$	3

Draw a frequency polygon to represent this data.



Line Graph: (Time series graph)

· Line graphs are used to display the comparison between two variables which are plotted on the X-axis and Y-axis.

• The X-axis represents measures of time, while the Y-axis represents percentage or measures of quantity.

•They organize and present data in a clear manner and show relationships between the data.

•Line graphs displays a change in direction

•It shows trend of an event occurring over a period of time to know whether it is increased or decreased. Eg: IMR, Cancer deaths etc

Some important types of Diagrams

Line diagram

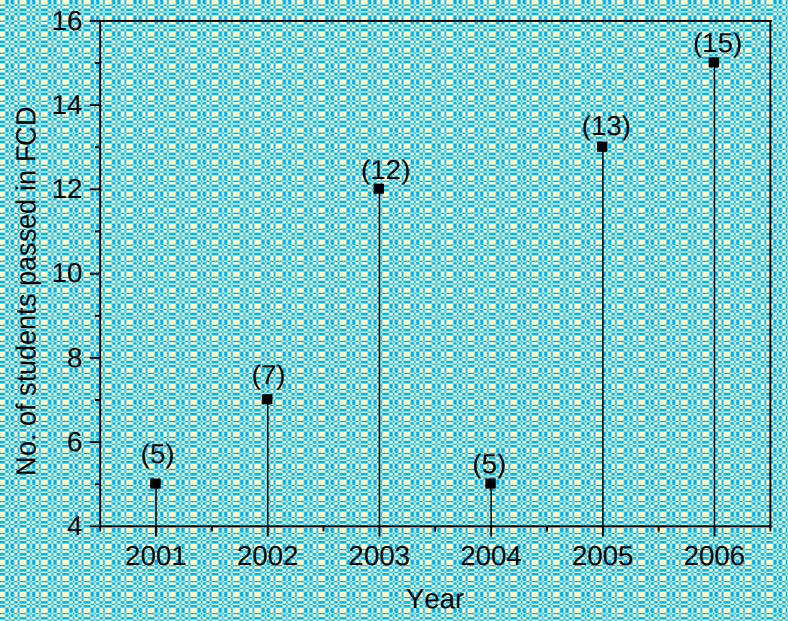
This is simplest type of one-dimensional diagram.
Based on size of the figures, heights of the bar / lines are drawn.
The distance between bars are kept uniform.
The limitation of this diagram are it is not attractive cannot provide more than one information.

***Ex*: Draw the line diagram for the following data**

Year: 2001 2002 2003 2004 2005 2006

No. of students passed in first class with distinction

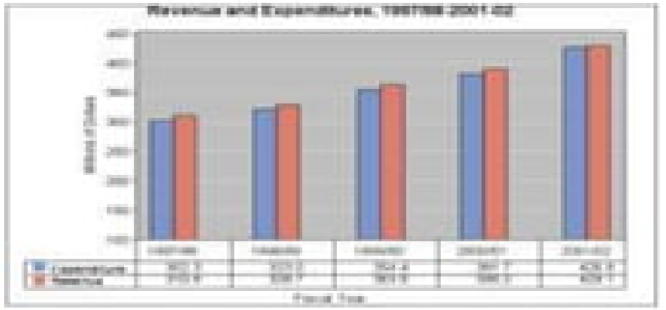
5 7 1 2 5 1 3 1 5



Difference between Graphs and Diagrams:

- To construct a graph, graph paper is generally used whereas a diagram is constructed on a plain paper.
- A graph represents mathematical relationship between two variables where as a diagram does not.
- Graphs are more appropriate than diagrams to represent frequency distributions and time series. Diagrams are not at all used for representing frequency distributions.
- Diagrams are more attractive to the eyes and as such are better suited for publicity and propaganda.
- Diagrams do not add anything to the meaning of the data and hence they are not helpful in analysis of data.
- Graphs are very much used by the statisticians and the research workers in their analysis.

Limitations:



- ✓They are visual aids. They cannot be considered as alternatives for numerical data.
- ✓Though theories and results could be easily visualized by diagrams and graphs, mathematical rigour cannot be brought in
- ✓Diagrams and graphs are not accurate as tabular data. Only tabular data can be used for further analysis.
- ✓By diagrammatical and graphical misrepresentation observers can be misled easily. It is possible to create wrong impressions using diagrams and graphs.

1. Collection of Data
2. Tabulation of Data
3. Analysis of Data
4. Interpretation of Data

Collection of data

TYPES OF DATA

There are two categories of data namely

(i) primary data and
(ii) secondary data.

PROCESSING OF DATA

Before tabulation of primary data, it should be scrutinized for

(i) Completeness
(ii) Consistency
(iii) Accuracy
(iv) Editing.

ACCURACY OF MEASUREMENT

ROUNDING OF FIGURES

ABSOLUTE AND RELATIVE ERROR

An absolute error (A.E.) is the absolute difference between the actual value (X) and its estimated value (x).

$$A.E. = |X-x| \qquad ...(1.1)$$

The relative error (R.E.) is the ratio of the absolute error to the actual value. i.e.,

$$R.E. = |X-x|/X \qquad ...(1.2)$$

Very often the relative error is given in per cent, i.e.,

$$R.E=|X-x|/X * 100 \quad (1.2.1)$$

The smaller the relative error, the better is the result.

METHODS OF ENQUIRY

Then keeping in view the purpose and importance of the investigation and types of respondents, the statistician has to choose one of the three methods of enquiry given below:

1. Personal enquiry method.

2.Correspondence, i.e., mailed questionnaire method.

3. Direct observational method.

TABULATION OF DATA

A table in general consists of the following parts.

Title
Footnote
Source Note

ANALYSIS OF DATA

INTERPRETATION OF DATA

Main Rules of Classification

- ***Unambiguity or Exactness:*** **rigidly defined Classes so that they should no lead to any ambiguity.**
- ***Exhaustive and Mutually Exclusive:*** **each data must belong to one of the classes and non overlapping.**
- ***Stability:*** **the same pattern of classification must be adopted throughout analysis.**
- ***Flexibility:*** **so that it can be adjustable to the new and changed situations and condition**
- ***Suitable for the purpose:***

PRESENTATION OF DATA

Before giving the description of diagrams, it is essential to familiarize oneself with certain terms.

Variable:

Commonly a factor or character which can take different values is called a variable.

A variable is any characteristics, number, or quantity that can be measured or counted. A variable may also be called a data item.

It is called a variable because the value may vary between data units in a population and may change in value over time.

e.g.: height, length, .weight, age, income and expenditure etc.

e.g.: 'income' is a variable that can vary between data units in a population (i.e. the people or businesses being studied may not have the same incomes) and can also vary over time for each data unit (i.e. income can go up or down).

Random Variable

Is a variable whose value is unknown or a function that assigns values to each of an experiment's outcomes.

A random variable can be either discrete (having specific values) or continuous (any value in a continuous range).

Continuous Random Variable

Observations can take any value between a certain set of real numbers. The value given to an observation for a continuous variable can include values as small as the instrument of measurement allows.

e.g. : height, time, age, and temperature.

Discrete Random Variable

Observations can take a value based on a count from a set of distinct whole values. A discrete variable cannot take the value of a fraction between one value and the next closest value.

e.g.: number of registered cars, number of business locations, and number of children in a family, all of which measured as whole units (i.e. 1, 2, 3 cars).

Frequency

Number of times a variate value is repeated is called frequency of the variate value.

Ex: 1. suppose there are 7 girl students who have secured 54 marks, 7 is the frequency of 54 marks.

2: If there are 12 people with monthly income of Rs. 50000 -70000, 12 is the frequency of the income group 50000-70000.

Frequency Distribution

Frequency distribution is a table used to organize the data. The left column (called classes or groups) includes numerical intervals on a variable under study. The right column contains the list of frequencies, or number of occurrences of each class/group. Intervals are normally of equal size covering the sample observations range. It is simply a table in which the gathered data are grouped into classes and the number of occurrences, which fall in each class, is recorded.

A frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups with their corresponding frequencies.

- Croxton and Cowden

The organisation of the data pertaining to a quantitative phenomenon involves the following four stages :

(i) The set or series of individual observations - unorganised (raw) or organised (arrayed) data.

(ii) Discrete or ungrouped frequency distribution.

(iii) Grouped frequency distribution.

(iv) Continuous frequency distribution.

Frequency Array

If the data is arranged in ascending / descending in the order of their magnitude, which gives better presentation then, it is called arraying of data.

If the individual items or values of a variable are given along with their corresponding frequencies, it is called a frequency array.

Ex: The wages per month and the number of persons in a small-scale industry are presented below:

wages per month (Rs)	350	490	600	780	800	1000
number of persons	4	5	7	8	4	-2

Such a presentation of data is called frequency array.

Discrete (ungrouped) Frequency Distribution:

If the data series are presented in such away that indicating its exact measurement of units, then it is called as discrete frequency distribution. Discrete variable is one where the variants differ from each other by definite amounts.

e.g.: Assume that a survey has been made to know number of post-graduates in 10 families at random; the resulted raw data could be as follows.

0, 1, 3, 1, 0, 2, 2, 2, 2, 4

This data can be classified into an ungrouped frequency distribution.

Ex:

In a survey, the age of 29 women at the time of their marriage by eight district administrative was reported as given below:

The data can be presented in the form of frequency distribution with the help of tally mark:

24, 25, 26, 24, 23, 27, 25, 22, 24, 23, 27, 28, 22, 23, 27, 26, 24, 23, 22, 22, 23, 23 , 23, 23, 23, 22, 25, 26, 27

Age (i)	Tally Marks (ii)	No. of women (iii)
22	𝍸	5
23	𝍸 ||||	9
24	||||	4
25	||	2
26	|||	3
27	||||	4
28	|	1

The number of post-graduates becomes variable (x) for which we can list the frequency of occurrence (f) in a tabular from as follows;

Number of postgraduates (x)	Frequency (f)
0	0
1	2
2	4
3	1
4	1

The above example shows a discrete frequency distribution, where the variable has discrete numerical values.

Continuous frequency distribution (grouped frequency distribution) :

Continuous data series is one where the measurements are only approximations and are expressed in class intervals within certain limits. In continuous frequency distribution the class interval theoretically continuous from the starting of the frequency distribution till the end without break. According to Boddington „the variable which can take very intermediate value between the smallest and largest value in the distribution is a continuous frequency distribution.

Ex: Marks obtained by 20 students in students" exam for 50 marks are as given below – convert the data into continuous frequency distribution form.

18	23	28	29	44	28	48	33	32	43
24	29	32	39	49	42	27	33	28	29

By grouping the marks into class interval of 10 following frequency distribution tables can be formed.

Marks No. of students	Marks No. of students
0-5	0
5-10	0
10-15	0
15-20	1
20-25	2
25-30	7
30-35	4
35-40	1
40-45	3
45-50	2

***Ex*: The birth weights (kilogram) of 30 children were recorded as follows:**

2.0, 3.7, 2.3, 3.0, 3.1, 2.7, 2.3, 3.5, 3.1, 3.7, 4.0, 2.3, 3.5, 4.2, 3.7, 3.2, 2.7, 2.5, 2.7, 3.8, 3.1, 3.0, 2.6, 2.8, 3.5, 4.1, 3.9, 2.8, 2.2, 4.2

Frequency distribution can be formed in the manner described so far, using various class intervals.

The width of the classes and the number of classes will be found out by Sturge's formula .

The range of data is 2.0 to 4.2 i.e., L = 4.2, S = 2.0

The class interval:

$$i = \frac{4.2 - 2.0}{(1 + 3.322 \log_{10} 30)}$$

$$\frac{2.2}{(1 + 3.322 * 1.4771)} = 2.2 / 5.91 = 0.37 = 0.4$$

and $k = 5.91 = 6.$

Hence, six classes with a width of 0.4 kg are to be taken in the frequency distribution.

Notes:

The lower limit of a class is included in that class.

It is not necessary to choose the smallest value as the lower limit of the lowest class or the largest value as upper limit of the highest class.

One may choose the classes as 1.0 -1.4, 1.4 -1.8 and so on.

Classes (Weight in kg)	Tally Marks	No. of children (ferquency)
2.0-2.4	𝍸	5
2.4-2.8	𝍸	5
2.8-3.2	𝍸 ||||	9
3.2-3.6	||||	4
3.6-4.0	||||	4
4.0-4.4	|||	3

Cumulative Frequency Distribution

(abbreviated as **c.f**)

It is the number of observations less than (more than) or equal to a specified value.

Less Than Cumulative Frequency:

Less than cumulative frequency for any value of the variable (or class) is obtained on adding successively the frequencies of all the previous values (or classes), including the frequency of variable (class) against which the totals are written, provided the values (classes) are arranged in ascending order of magnitude.

Example:

the total number of students with marks less than, say, 40 is 5 + 10 = 15; 'less than 50' is the sum of all the previous frequencies upto and including the class 45—50 i.e., 5 + 10 + 15 + 30 = 60 and so on.

‘LESS THAN’ CUMULATIVE FREQUENCY DISTRIBUTION OF MARKS OF 70 STUDENTS

Marks	Frequency (f) No. of students	‘Less than’ c.f.
30-35	5	5
35-40	10	5+10=15
40-45	15	15+15=30
45-50	30	30+30=60
50-55	5	60+5=65
55-60	5	65+5=70

LESS THAN c. f. DISTRIBUTION

Marks	Frequency (f)
Less than 30	0
Less than 35	5
Less than 40	15
Less than 45	30
Less than 50	60
Less than 55	65
Less than 60	70

More Than Cumulative Frequency:

The ‘more than cumulative frequency’ is obtained similarly by finding the cumulative totals of frequencies starting from the highest value of the variable (class) to the lowest value (class). Thus in the above illustration the number of students with marks ‘more than 50’ is 5 + 5 = 10, and ‘more than 40’ is 15 + 30 + 5 + 5 = 55 and so on.

‘MORE THAN’ CUMULATIVE FREQUENCY DISTRIBUTION OF MARKS OF 70 STUDENTS

Marks	Frequency (f) No. of students	‘more than’ c.f.
30-35	5	65+5=70
35-40	10	60+5=65
40-45	15	30+30=60
45-50	30	15+15=30
50-55	5	5+10=15
55-60	5	5

MORE THAN c. f. DISTRIBUTION

Marks	Frequency (f)
More than 30	70
More than 35	65
More than 40	55
More than 45	30
More than 50	15
More than 55	5
More than 60	0