Topic: Data Presentation

- By the end of this session, the students should be able to
- Define the data
- Know the different types of data
- Know the different ways to present the data scientifically and systematically

Definition of Data

Any observation collected in respect of any characteristic or event is called data.

Information

- Raw data carry/convey little meaning, when it is considered alone.
- The data is minimized, processed/analyzed and then presented systematically. So that it is converted into **Information**.
- It is important to note that data, that is not converted into information is of little value for evaluation and planning and can not be used by those who are involved in decision making.

Types of Data

- To give a holistic picture of classification data can divided into two types

- **Quantitative data** (numerical)
- **Qualitative data** (descriptive, categorical/frequency count)

Quantitative data

The data , that can be expressed in numbers/ figures is called quantitative data. In this the exact measures are possible.

It has two types

(a) Discrete:

Discrete variables can take only certain values and none in between e. g number of patients in a hospital census may be 178 or 179, but it cannot be in between these two, similarly the number of syringes used in a clinic in one day or number of children in a family.

It is expressed in whole number.

(b) Continuous:

Continuous variables may take any value (typically between certain limits). For example age (25.5 years), weight (70.5 kg), height (1.5 meter) , hemoglobin (12.5 gm), blood pressure (135/95). It can be expressed in decimals.

Qualitative Data

- Also called descriptive/ categorical data/ frequency count.
- When the data are arranged in categories on the basis of their quality and there is gap between two values, it is called qualitative data, e.g name, religion, marital status, socioeconomic status, awareness.
- Qualitative data cannot be expressed in numerical forms.

Types of qualitative data

Nominal data:

Nominal scale data are divided into categories, that are only distinguished by their name and labels and cannot be classified one above another e.g race, name , sex, name of country, name of crops, type of blood. In this type of data there is no implication of order or ratio.

Nominal data that falls into two groups are called dichotomous data e.g male/ female, black/white, rural/ urban.

Ordinal data:

When the categorical data can be placed in meaningful order on the basis of their quality, it is known as ordinal data. In this the exact difference between the two groups cannot be estimated e.g pain categorized as mild, moderate and severe. Similarly scoring of students categorized as A (70% and above), B (60-69 %), C (50-59 %). In this the exact difference between the students placed in grade A and B cannot be estimated.

Interval:

Interval scale data are like ordinal data in that they can be placed in a meaningful order.

- The categories are arranged in equally spaced units and there is no absolute zero point e.g temperature where 0 ˚ C does not mean no temperature but is equal to 32 ˚ F or 273 K (Kelvin scale).
- In addition they have meaningful intervals between items, which are usually measured quantities. For example on the Celsius scale the difference between 100 ˚ C and 90 ˚ C is the same as the difference between 50 ˚ C and 40 ˚ C.
- However because interval scales do not have an absolute zero, ratio of scores are not meaningful e.g 100 ˚ C is not twice as hot as 50 ˚ C, because 0 ˚ C does not indicate a complete absence of heat.

-

Ratio

- A ratio scale has the same properties as an interval scale; however, because it has an absolute zero, meaningful ratios do exist.
- Most biomedical variables form a ratio scale e.g weight in grams or pounds, time in seconds or days, blood pressure in millimeters of mercury, and pulse rate in beats per minute are all ratio scale data.
- The only ratio scale of temperature is the Kelvin scale, in which zero degree indicates an absolute absence of heat, just as a zero pulse rate indicates an absolute lack of heartbeat. Therefore it is correct to say that a pulse rate of 120 beats/min is twice as fast as pulse rate of 60 beats / min, or that 300 K is twice as hot as 150 K.

Data Presentation

- Principals of data presentation

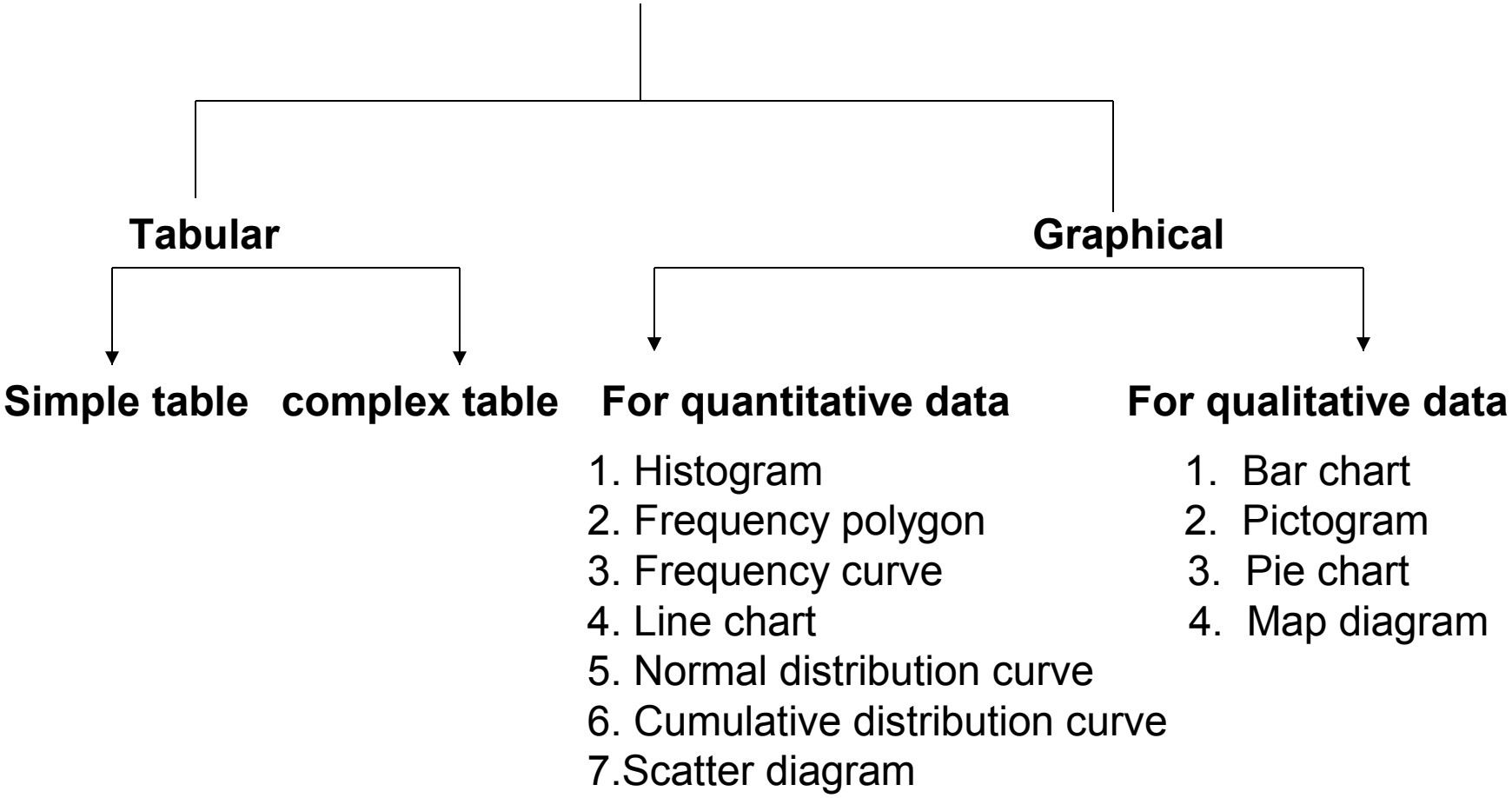
  (a) To arrange the data in such a way that it should create interest in the reader's mind at the first sight.

  (b) To present the information in a compact and concise form without losing important details.

  (c) To present the data in a simple form so as to draw the conclusion directly by viewing at the data.

  (d) To present it in such away that it can help in further statistical analysis.

Presentation of data



Tabulation

Tables are the devices, that are used to present the data in a simple form. It is probably the first step before the data is used for analysis or interpretation.

General principals of designing tables

a) The tables should be numbered e.g table 1, table 2 etc.
b) A title must be given to each table, which should be brief and self explanatory.
c) The headings of columns or rows should be clear and concise.
d) The data must be presented according to size or importance chronologically, alphabetically, or geographically.
e) If percentages or averages are to be compared, they should be placed as close as possible.
f) No table should be too large
g) Most of the people find a vertical arrangement better than a horizontal one because, it is easier to scan the data from top to bottom than from left to right
h) Foot notes may be given, where necessary, providing explanatory notes or additional information.

Types of tables

1) Simple tables :Measurements of single set are presented
2) Complex tables :Measurements of multiple sets are presented

Simple Table

When characteristics with values are presented in the form of table, it is known as simple table e.g

Table 4.4

Infant mortality rate of selected countries in 2004

Name of country	Infant mortality rate
Pakistan	90
Bangladesh	60
Sri Lanka	26
India	60

Frequency distribution table

- In the frequency distribution table, the data is first split up into convenient groups (class interval) and the number of items (frequency) which occur in each group is shown in adjacent columns.
- Hence it is a table showing the frequency with which the values are distributed in different groups or classes with some defined characteristics.

Rules for construction of frequency table

1) The class interval should not be too large or too small
2) The number of classes to be formed more than 8 and less than 15
3) The class interval should be equal and uniform through out the classification.
4) After construction of table, proper and clear heading should be given to it
5) The base or source of data should be mentioned with the pattern of analysis in footnote at the end of table

Frequency distribution table

Table 3

Age distribution of polio patients

Age	Number of patients
O-4	35
5-9	18
10-14	11
15-19	8
20-24	6

Table1.2
Grouped, relative, and cumulative frequency distributions if serum cholesterol levels in 200 men

Interval	Frequency f	relative f	cumulative f
251-260	5	2.5	100.0
241-250	13	6.5	97.5
231-240	19	9.5	91.0
221-230	18	9.0	81.5
211-220	38	19.0	72.5
201-210	72	36.0	53.5
191-200	14	7.0	17.5
181-190	12	6.0	10.5
171-180	5	2.5	4.5
161-170	4	2.0	2.0

Charts and Diagrams

Charts and diagrams are useful methods of presenting simple data.

- They have powerful impact on imagination of people.
- Gives information at a glance.
- Diagrams are better retained in memory than statistical table.
- However graphs cannot be substituted for statistical table, because the graphs cannot have mathematical treatment where as tables can be treated mathematically.
- Whenever graphs are compared , the difference in the scale should be noted.
- It should be remembered that a lot of details and accuracy of original data is lost in charts and diagrams, and if we want the real study, we have to go back to the original data.

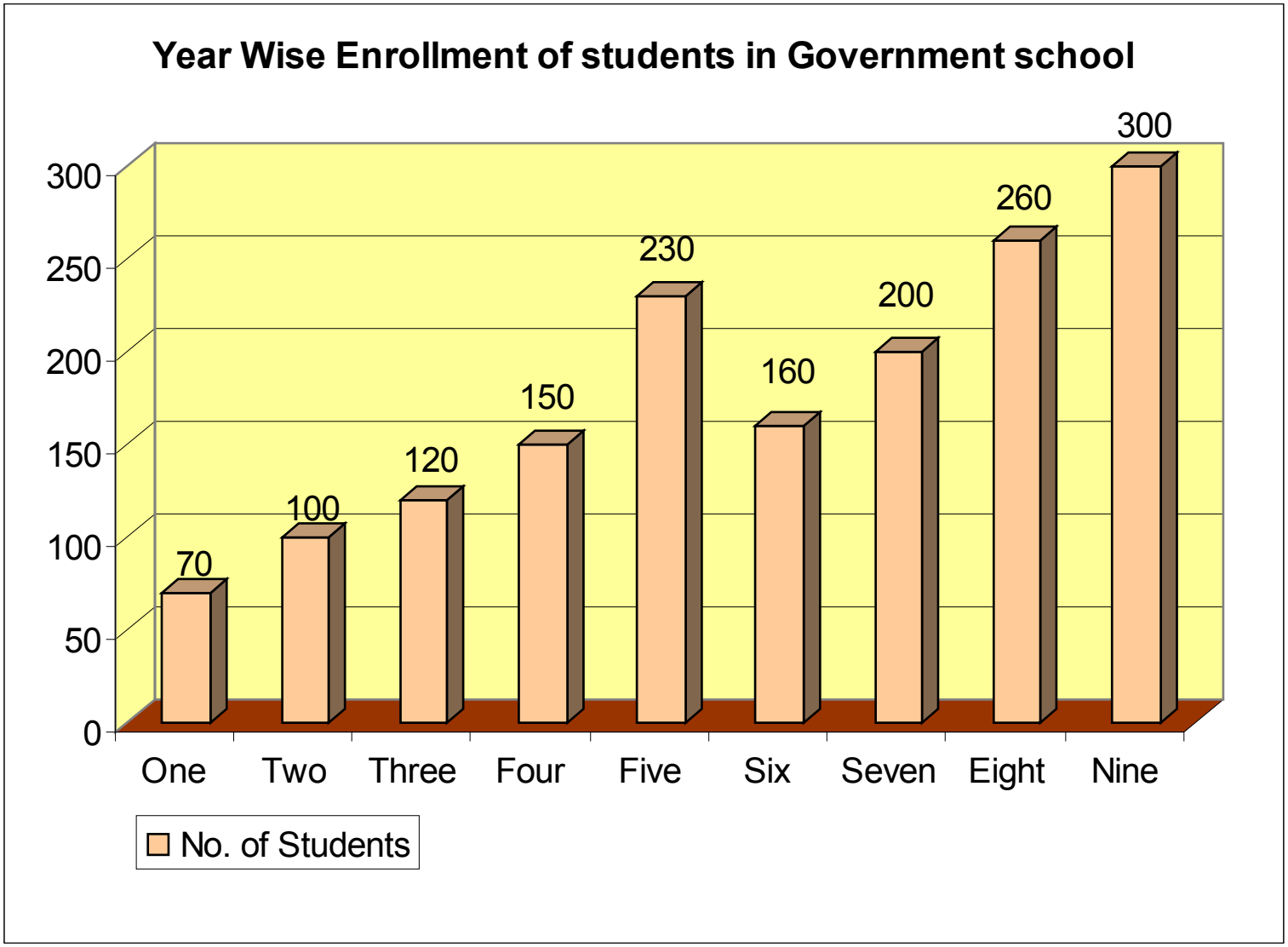
Common diagrams

- Pie chart
- Simple bar diagram
- Multiple bar diagram
- Component bar diagram or subdivided bar diagram
- Histogram
- Frequency polygon
- Frequency curve
- O give curve
- Scatter diagram
- Line diagram
- Pictogram
- Statistical maps

Bar charts

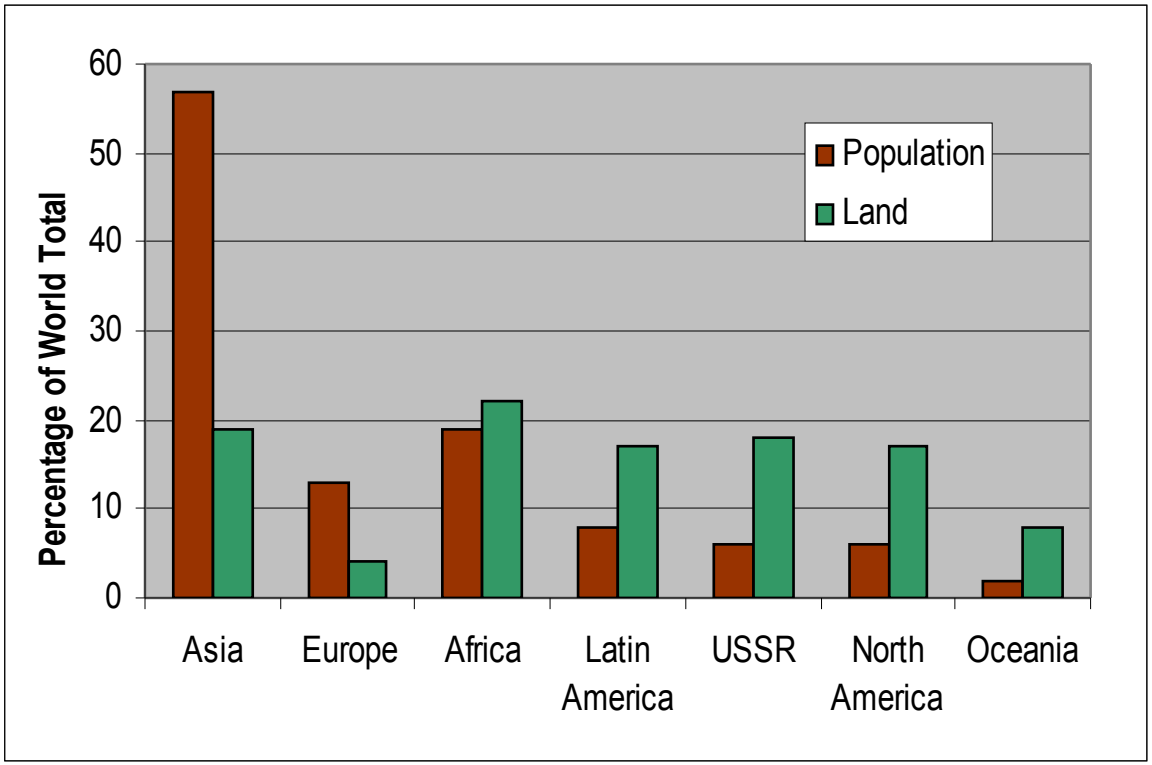
- The data presented is categorical
- Data is presented in the form of rectangular bar of equal breadth.
- Each bar represent one variant /attribute.
- Suitable scale should be indicated and scale starts from zero.
- The width of the bar and the gaps between the bars should be equal throughout.
- The length of the bar is proportional to the magnitude/ frequency of the variable.
- The bars may be vertical or horizontal.

Bar charts


Year Wise Enrollment of students in Government school
300
250
200
150
100
50
0
70
100
120
150
230
160
200
260
300
One
Two
Three
Four
Five
Six
Seven
Eight
Nine
No. of Students

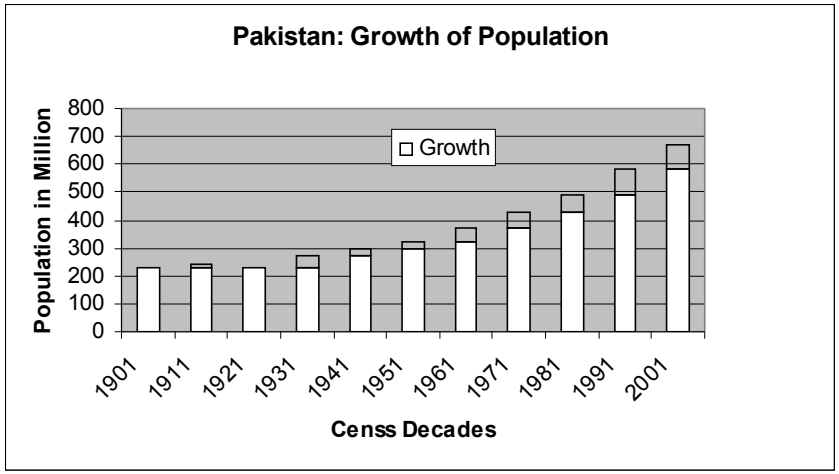
Multiple Bar Charts

- Also called compound bar charts
- More then one sub-attribute of variable can be expressed

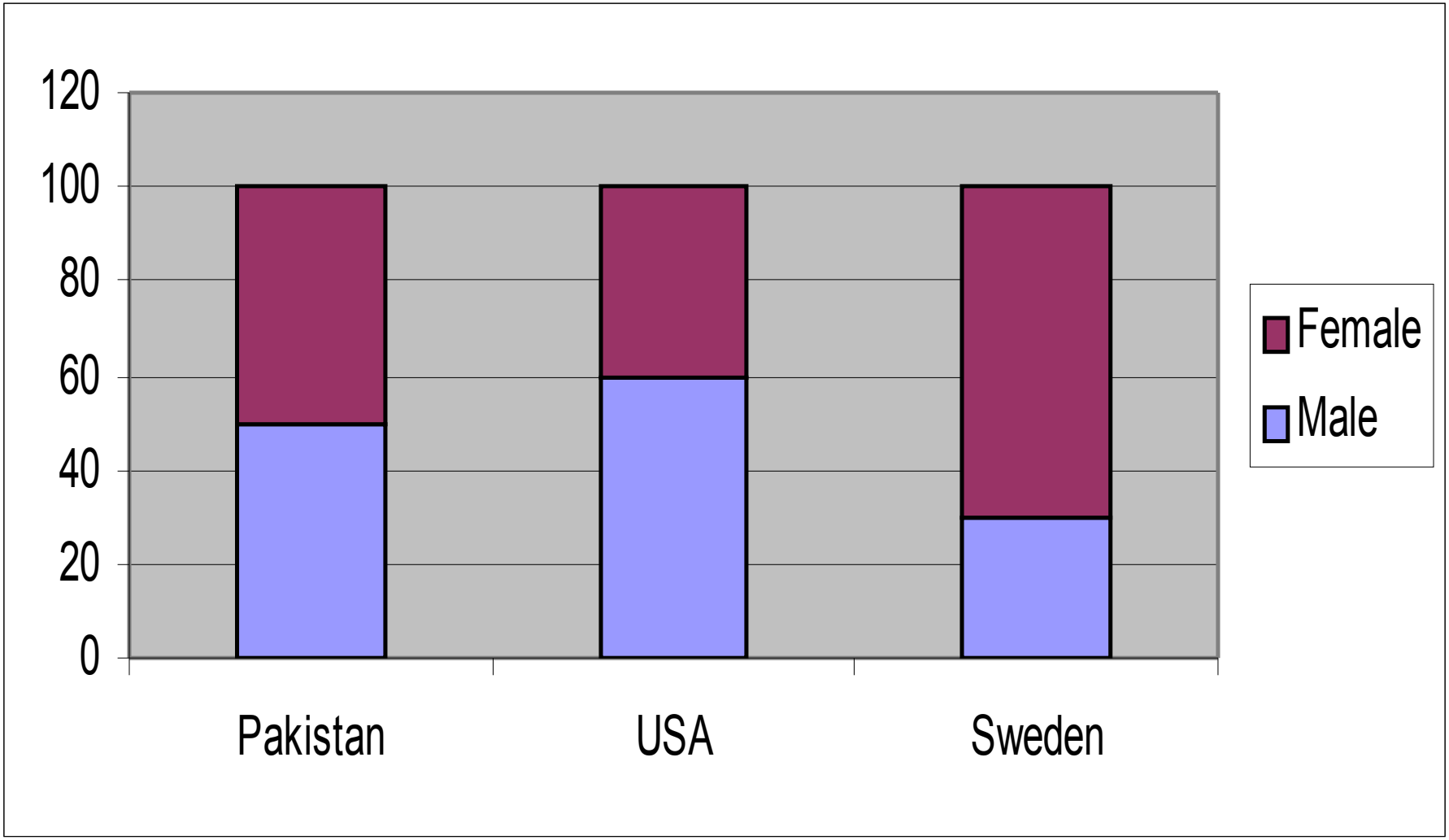


Component bar charts

- When there are many categories on X-axis (more than 5) and they have further subcategories, then to accommodate the categories, the bars may be divided into parts, each part representing a certain item and proportional to the magnitude of that particular item

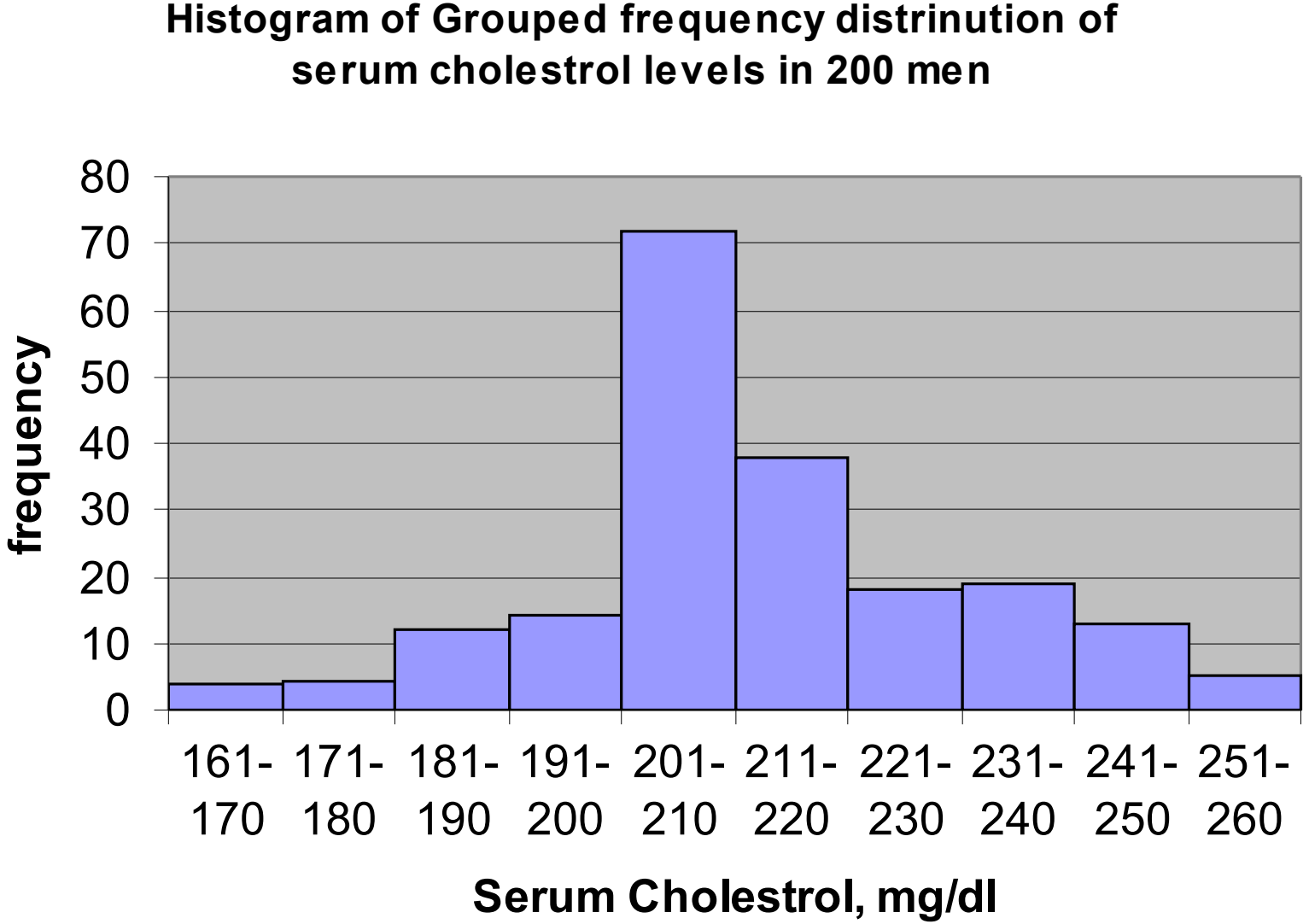


Component Bar Chart



Histogram

- Used for Quantitative, Continuous, Variables.
- It is used to present variables which have no gaps e.g age, weight, height, blood pressure, blood sugar etc.
- It consist of a series of blocks. The class intervals are given along horizontal axis and the frequency along the vertical axis.

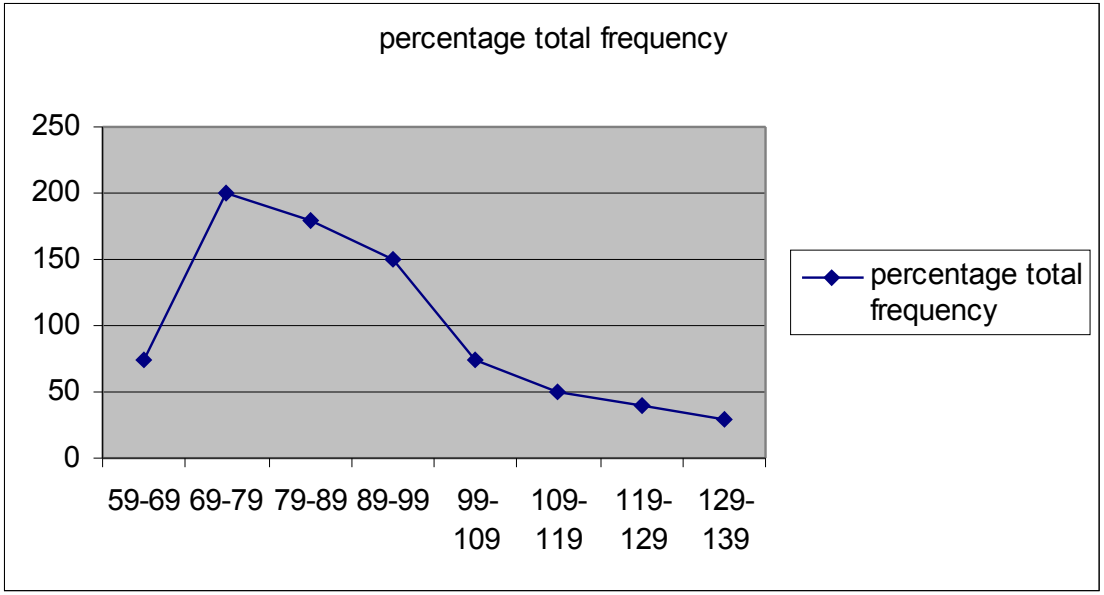

Histogram of Grouped frequency distrinution of serum cholestrol levels in 200 men
frequency
80
70
60
50
40
30
20
10
0
161-170
171-180
181-190
191-200
201-210
211-220
221-230
231-240
241-250
251-260
Serum Cholestrol, mg/dl

Frequency polygon

Frequency polygon is an area diagram of frequency distribution over a histogram.

It is a linear representation of a frequency table and histogram, obtained by joining the mid points of the hitogram blocks.

Frequency is plotted at the central point of a group



Normal frequency distribution curve

- Frequency polygons may take many different shapes, but many naturally occurring phenomena are approximately distributed according to the symmetrical, bell-shaped normal or Gaussian distribution.
- In normal distribution curve, the three measures of central tendency are identical. approximately 68% of the distributions falls within +_

  1 standard deviation of the mean .

  approximately 95% of the distributions falls within +_

  2 standard deviation of the mean

  approximately 99.7% of the distributions falls within +_

  3 standard deviation of the mean

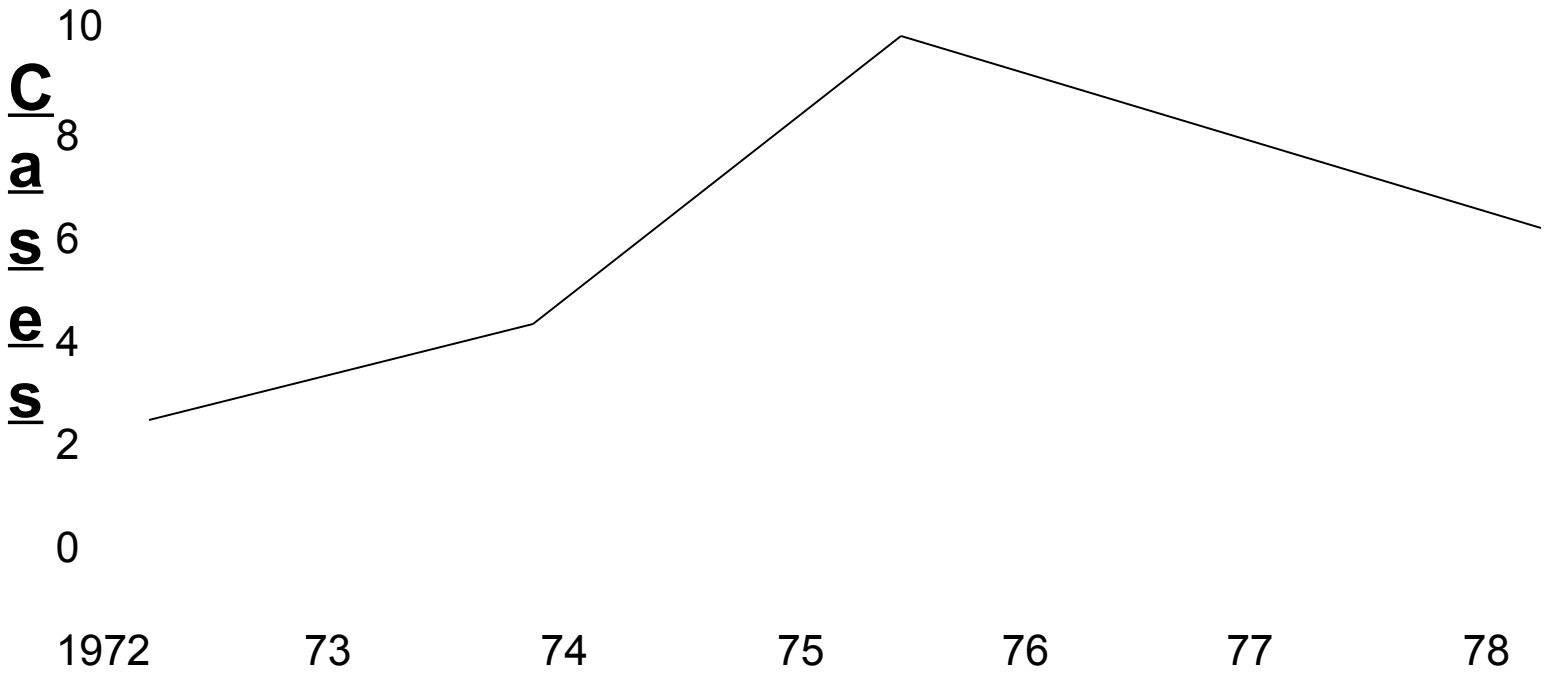
men and women (each gender forms its own distribution around a different midpoint).

Asymmetrical distribution are called skewed distributions. The three measures of central tendency differ. Mode is highest point on curve, the mean is pulled up or down by the influence of a relatively small number of very high or very low scores and the median lies between the two.

- n Positively (or right) skewed distributions and negatively (or left) skewed distributions cabe identified by the location of the tail of the curve.
- Positively skewed distributions have a relatively large number of low scores and a small number of very high scores.
- Negatively skewed distributions have relatively large number of high scores and a small number of low scores.
- Bimodal distributions are sometimes a combination of two underlying normal distributions, such as the heights of a large number of men and women (each gender forms its own distribution around a different midpoint).

Line diagram

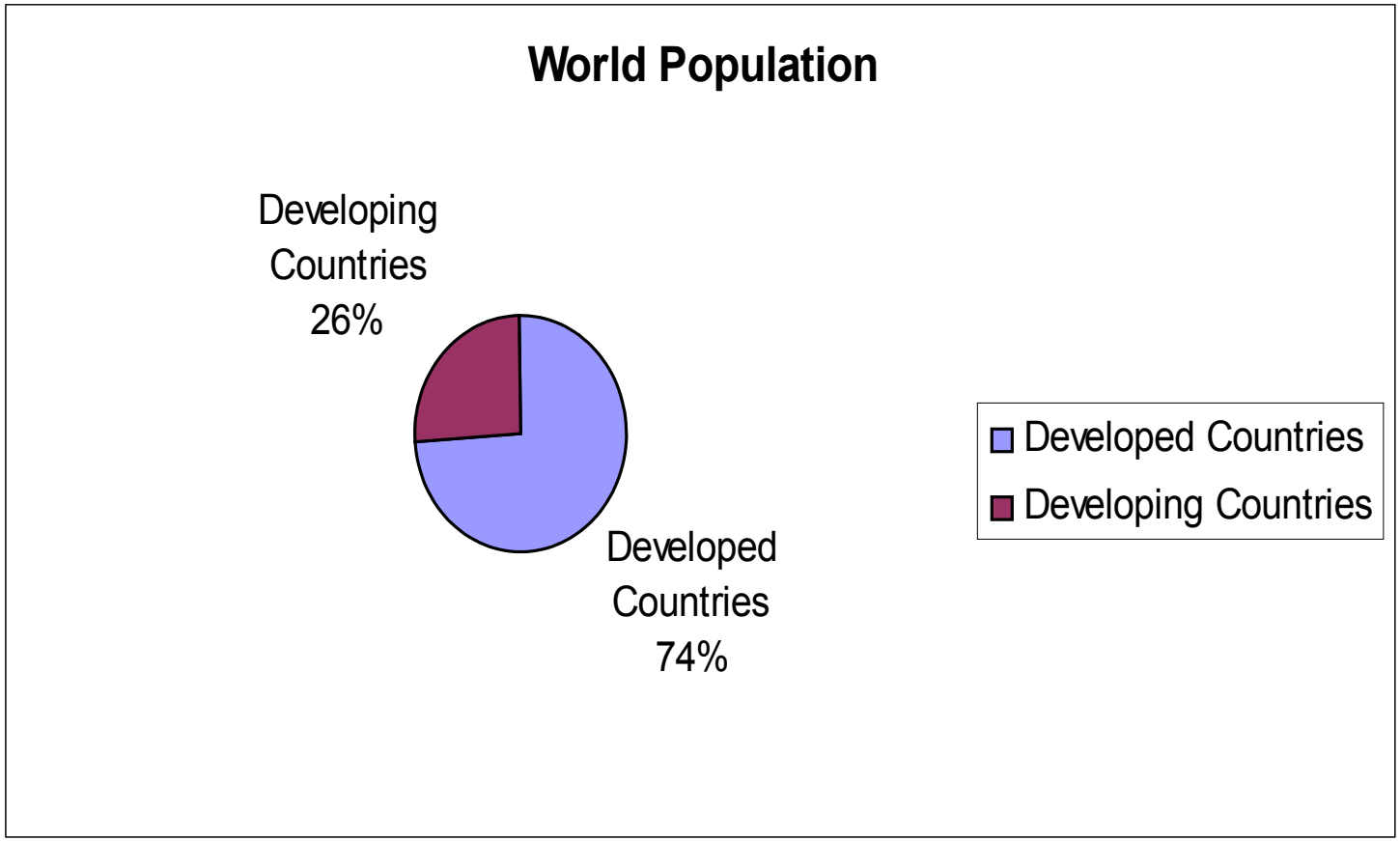
- Line diagrams are used to show the trend of events with the passage of time.
- Line diagram showing the malaria cases reported throughout the word excluding African region during 1972-78



Pie charts

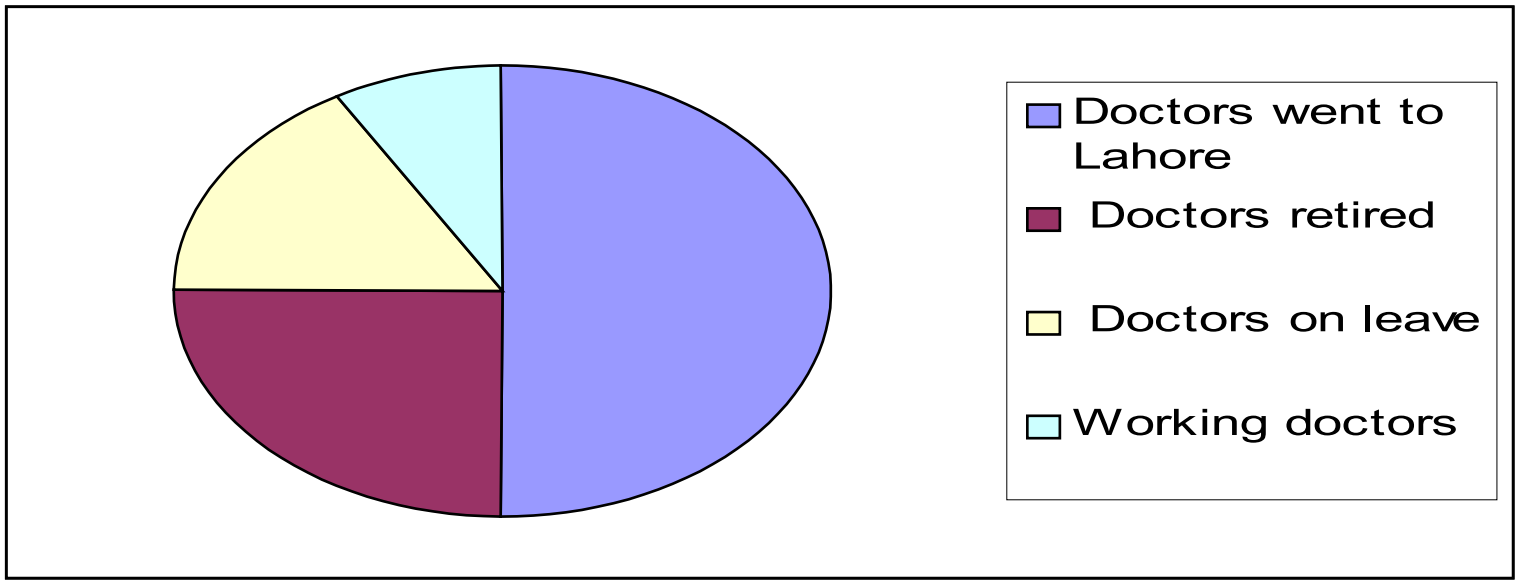
- Most common way of presenting data
- The value of each category is divided by the total values and then multiplied by 360 and then each category is allocated the respective angle to present the proportion it has.
- It is often necessary to indicate percentages in the segment as it may not be sometimes very easy virtually, to compare the areas of segments.

Pie Charts


World Population
Developing Countries 26%
Developed Countries 74%
Developed Countries
Developing Countries

Pie Chart

- **Question-:** In a DHQ Hospital 120 Doctors are working.60 doctors went to Lahore to attend a workshop.20 doctors went on long leave.30 doctors were retired.
- Show this data by Pie chart.



Pictogram

- Popular method of presenting data to those who cannot understand orthodox charts.
- Small pictures or symbols are used to present the data,e.g a picture of a doctor to represent the population physician.
- Fraction of the picture can be used to represent numbers smaller than the value of whole symbol

Statistical maps

- When statistical data refers to geographic or administrative areas, it is presented either as statistical map or dot map.
- The shaded maps are used to present data of varying size. The areas are shaded with different colour or different intensities of the same colour, which is indicated in the key.

Scatter diagram

- Scatter diagrams show the relationship between the two variables e.g a positive correlation/ association between the intake of fat and sugar in the average diets of 41 countries.
- If the dots cluster round a straight line, it shows evidence of a relationship of a linear nature.
- If there is no such cluster, it is probable that there is no relationship between the variables.