

Chapter 2

Methods of Data Collection and Presentation

2.1 Method of data collection

There are two types of data: Primary and Secondary

- **Primary data:** the investigator himself collects the data.

Primary data collection methods: includes observation, personal interview, self administered questionnaire, mailed questionnaire etc.

- **Secondary data:** is not investigated by the investigator himself, but he obtains from someone else records.

Secondary data collection methods: obtained from published or unpublished documents: reports, journals, magazines, articles e t c.

2.2 Method of data organization and presentation

- The collected data will be too large, thus it should be organized.
- **Data presentation is used to:**
 - to display the points of similarity and dissimilarity
 - to condensation and suppression of irrelevant detail
 - to enable one to form a mental picture of objects of perception
 - to prepare the ground for comparison and inference
- **Classification of data:** used to classify data according to some characteristics

2.2.1 Presentation of data classified in categories.

- **Tabular Presentation**
- **Diagrammatic and Graphic presentation**

Definitions:

- **Raw data:** recorded information in its original collected form, whether it is counts or measurements, is referred to as raw data.
- **Frequency:** is the number of values in a specific class of the distribution.
- **Frequency distribution:** is the organization of raw data in table form using classes and frequencies

There are three basic Types of frequency distributions

- Categorical frequency distribution
- Ungrouped frequency distribution
- Grouped frequency distribution
-

Categorical frequency distributions:

Example: The following data are on the political party affiliations of sample of 40 students. D, R, and O stand for Democratic, Republican and Other, respectively.

D D D D O R O R O R O R O
D D R D D D R R O R D R R
O R R R R R O O R R D R D D

D=13, R=18 O=9

Construct ungrouped frequency distribution.

Solution

Step 1: Make a table

Step 2: Tally the data

Step 3: Count the tally and place the result

Step 4: Find the percentages of values in each class

by using;

$$\% = \frac{f}{n} * 100$$

Where f= frequency of the class, n=total number of value.

Class (1)	Tally (2)	Frequency (3)	Percent (4)
D	///// ///// ///	13	32.5%
R	///// ///// ///// ///	18	45%
O	//// ////	9	22.5

Ungrouped frequency distributions

Constructing ungrouped frequency distribution:

- First find the smallest and largest raw score in the collected data.
- Arrange the data in order of magnitude and count the frequency.
- To facilitate counting one may include a column of tallies.

Example: The following data represent the mark of 20 students.

80	76	90	85	80
70	60	62	70	85
65	60	63	74	75
76	70	70	80	85

Solution:

Step 1: Find the range, $\text{Range} = \text{Max} - \text{Min} = 90 - 60 = 30$.

Step 2: Make a table as shown

Step 3: Tally the data.

Step 4: Compute the frequency.

Mark	Tally	Frequenc y
60	//	2
62	/	1
63	/	1
65	/	1
70	////	4
74	/	1
75	//	2
76	/	1
80	///	3
85	///	3
90	/	1

Grouped frequency Distribution:

When the range of the data is large, the data must be grouped in to classes that are more than one unit in width.

Definitions:

Grouped Frequency Distribution: a frequency distribution when several numbers are grouped in one class.

Class limits: Separates one class in a grouped frequency distribution from another. The limits could actually appear in the data and have gaps between the upper limits of one class and lower limit of the next.

Units of measurement (U): the distance between two possible consecutive measures. It is usually taken as 1, 0.1, 0.01, 0.001, -----.

- **Class boundaries:** The boundaries have one more decimal places than the row data and therefore do not appear in the data. There is no gap between the upper boundary of one class and lower boundary of the next class. The lower class boundary is found by subtracting $U/2$ from the corresponding lower class limit and the upper class boundary is found by adding $U/2$ to the corresponding upper class limit.
- **Class mark (Mid points):** it is the average of the lower and upper class limits.
- **Cumulative frequency:** is the number of observations less than/more than or equal to a specific value.
- **Relative frequency (rf):** it is the frequency divided by the total frequency.

Example: Construct a frequency distribution for the following data.

11	29	6	33	14	31	22	27	19	20
18	17	22	38	23	21	26	34	39	27

■ Solutions:

Step 1: Find the highest and the lowest value $H=39$, $L=6$

Step 2: Find the range; $R=H-L=39-6=33$

Step 3: Select the number of classes desired using Sturges formula;

$$k = 1 + 3.32 \log n = 1 + 3.32 \log(20) = 5.32 = 6$$

Step 4: Find the class width; $w=R/k=33/6=5.5=6$ (rounding up)

- Step 5: Select the starting point, let it be the minimum observation.

6, 12, 18, 24, 30, 36 are the lower class limits.

Step 6: Find the upper class limit; e.g. the first upper class= $12 - U = 12 - 1 = 11$

11, 17, 23, 29, 35, 41 are the upper class limits.

Step 7: Find the class boundaries;

E.g. for class 1 Lower class boundary= $6 - U/2 = 5.5$

Upper class boundary
 $= 11 + U/2 = 11.5$

Step 8: tally the data.

Step 9: Write the numeric values for the tallies in the frequency column.

Step 10: Find cumulative frequency.

Step 11: Find relative frequency or/and relative cumulative frequency.

Class limit	Class boundar y	Class Mark	Freq .	Cf (less than type)	Cf (more than type)	rf.	rcf (less than type)
6 - 11	5.5 - 11.5	8.5	2	2	20	0.1 0	0.10
12 - 17	11.5 - 17.5	14.5	2	4	18	0.1 0	0.20
18 - 23	17.5 - 23.5	20.5	7	11	16	0.3 5	0.55
24 - 29	23.5 - 29.5	26.5	4	15	9	0.2 0	0.75
30 - 35	29.5 - 35.5	32.5	3	18	5	0.1 5	0.90
36 - 41	35.5 - 41.5	38.5	2	20	2	0.1 0	1.00

2.2.3 Diagrammatic and graphic presentation of data

Graphs for quantitative data

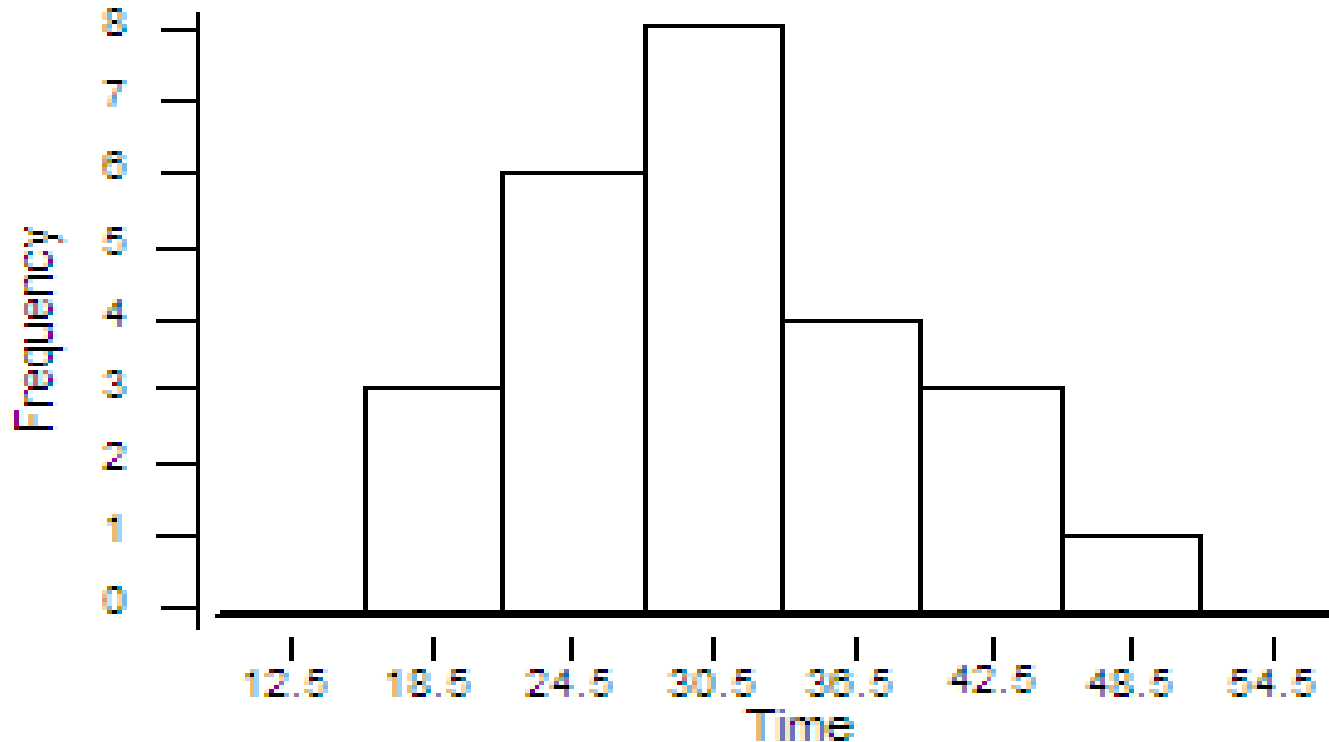
1. Histogram: its bases are marked by class boundaries and its heights are proportional to the frequencies.

Example: Construct a histogram for the frequency distribution of the time spent by the automobile workers.

The frequency distribution is:

Time (class boundaries)	Class mark	Number of workers
15.5-21.5	18.5	3
21.5-27.5	24.5	6
27.5-33.5	30.5	8
33.5-39.5	36.5	4
39.5-45.5	42.5	3
45.5-51.5	48.5	1

Figure 1: The time in minutes spent by automobile workers to travel from home to WO



2. Frequency polygon: class marks against class frequencies and joining them by a set of line segments

- Add two classes with zero frequencies at the two ends of the frequency distribution

Example: Construct a frequency polygon for the frequency distribution of the time spent by the automobile workers.

Figure 2: The time in minutes spent by automobile workers to travel from home to



3. Ogive (Cumulative frequency graph):

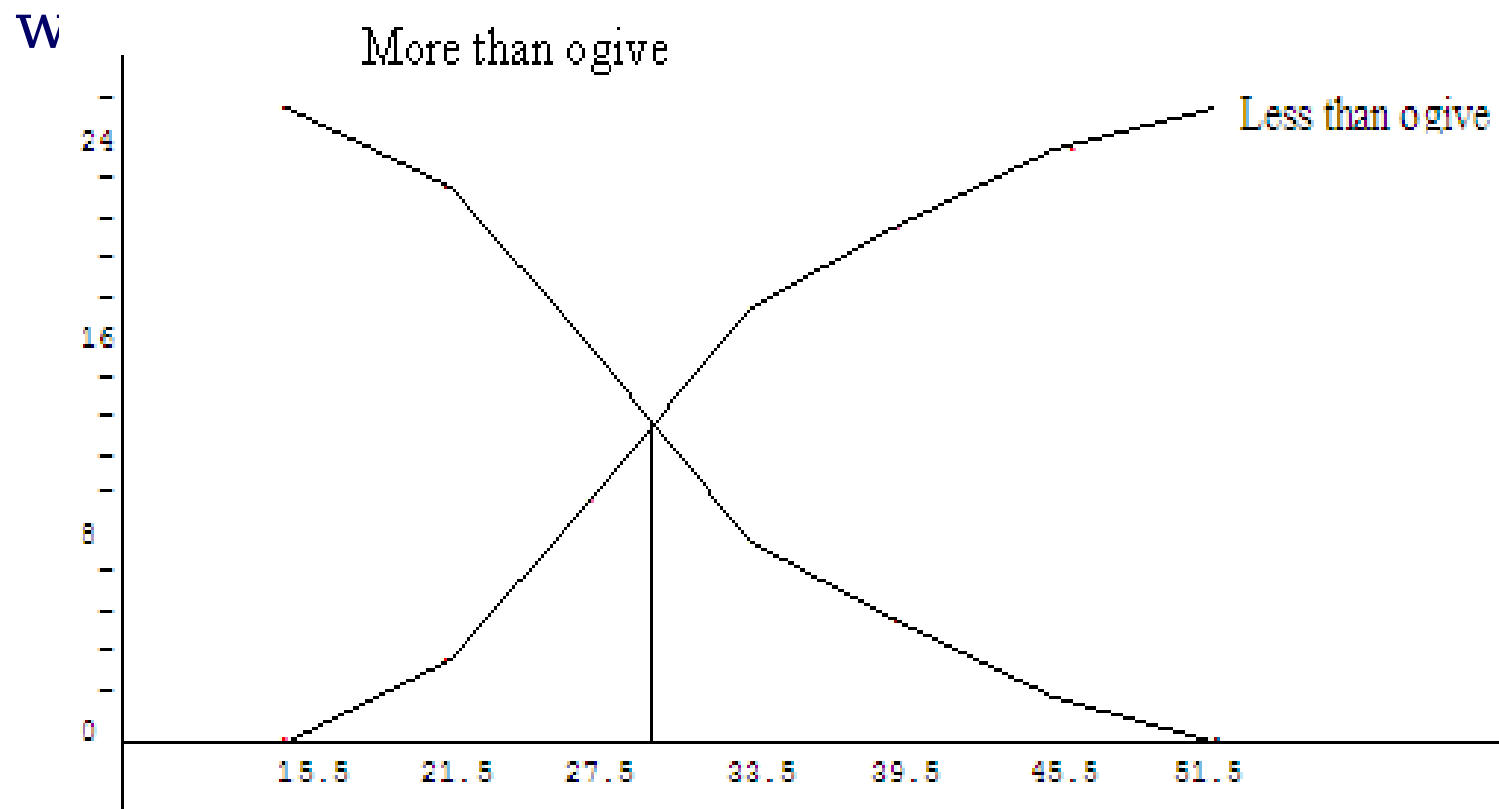
plotting the cumulative frequencies of a distribution against the boundaries

Example: Construct an ogive for the time spent by the automobile workers.

The frequency distribution is:

Class boundaries	LCF	MCF
15.5	0	25
21.5	3	22
27.5	9	16
33.5	17	8
39.5	21	4
45.5	24	1
51.5	25	0

Figure 3: The time in minutes spent by automobile workers to travel from home to



Graphs for qualitative data

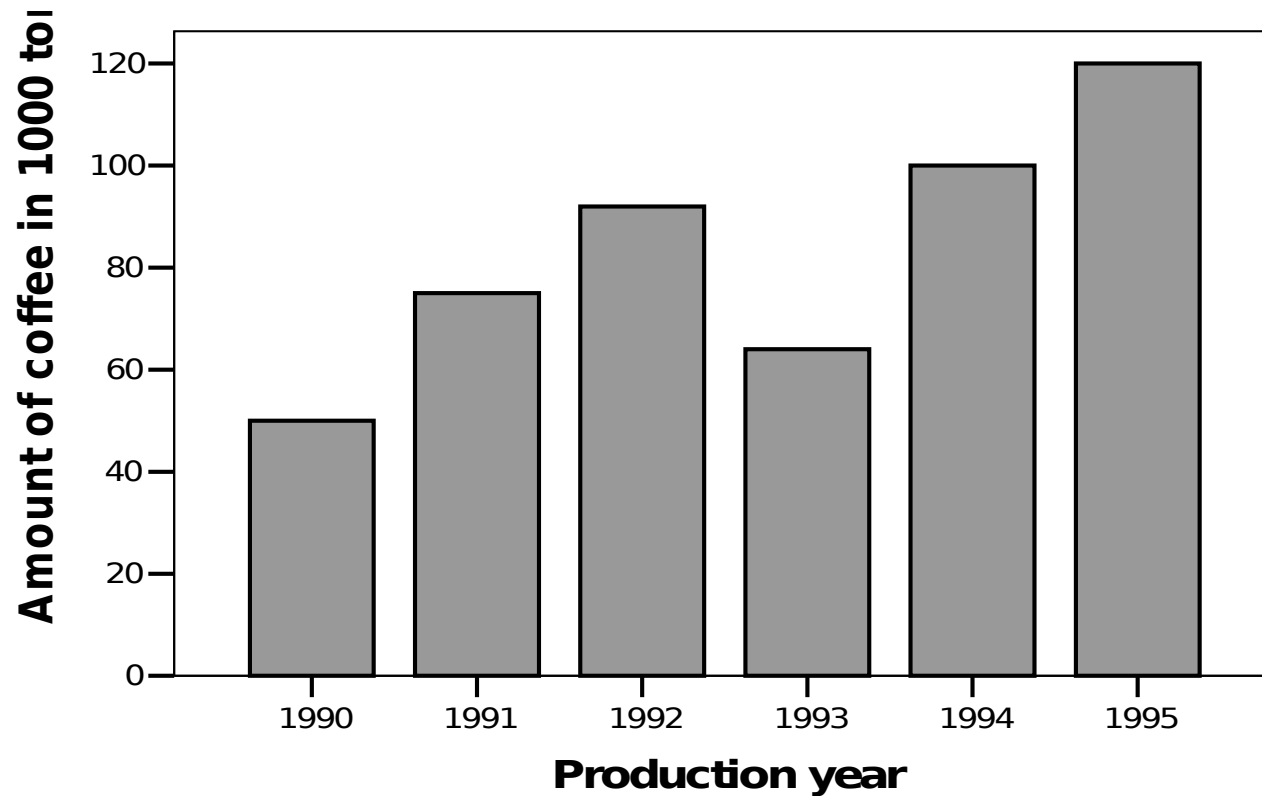
1. Bar charts: data are represented by a series of vertical or horizontal bars

i. Simple bar chart:

Example: Draw a bar chart for the following coffee production data from 1990 to 1995.

Year	1990	1991	1992	1993	1994	1995
Amount (in 1000 tones)	50	75	92	64	100	120

Figure : Production of coffee from 1990 to 1995

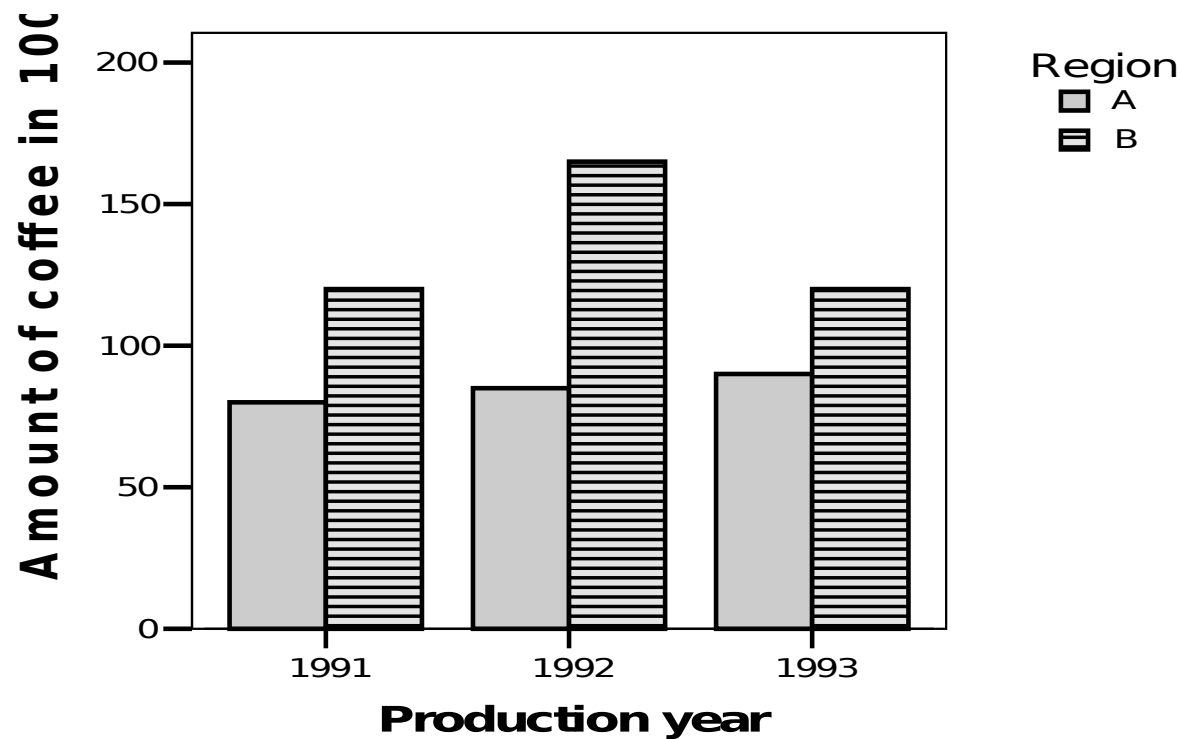


ii. Multiple bar charts:

Example : Draw a multiple bar chart for the data on production of coffee (in 1000 tons) from 1991 to 1993 by region.

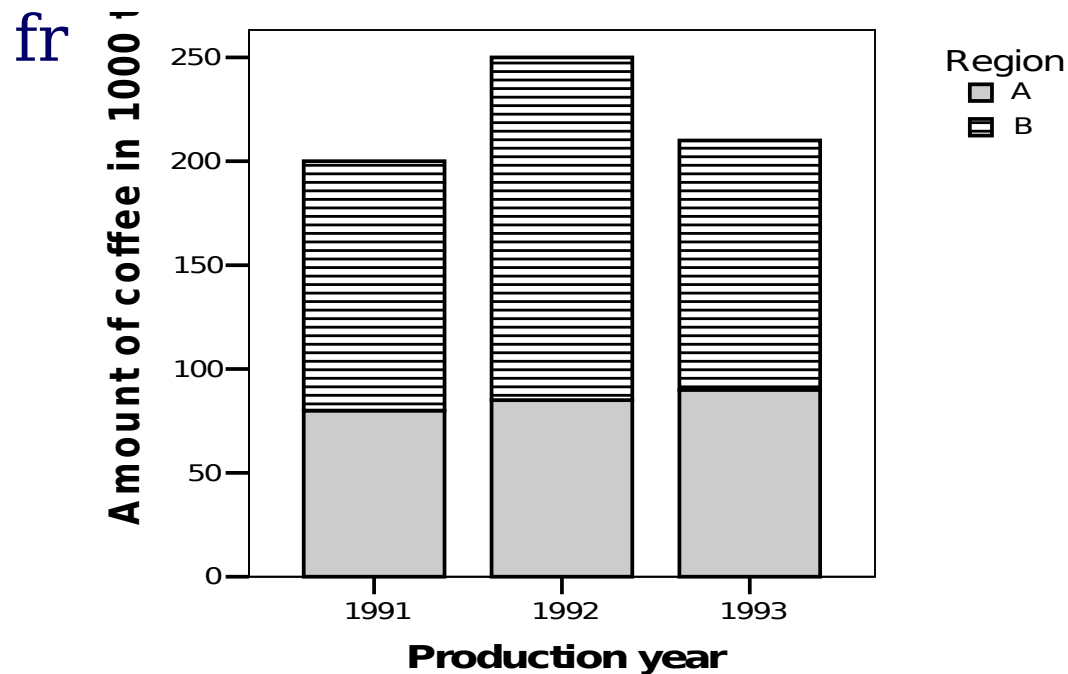
Region	Production year		
	1991	1992	1993
Region A	80	85	90
Region B	120	165	120
Total	200	250	210

Figure : Production of coffee from 1991 to 1993 in two regions.



iii. Component bar charts:

Example : Draw a component bar chart for the data on production of coffee (in 1000 tons)



2. Pie charts: a circle divided in to sectors to present the portion of the data set

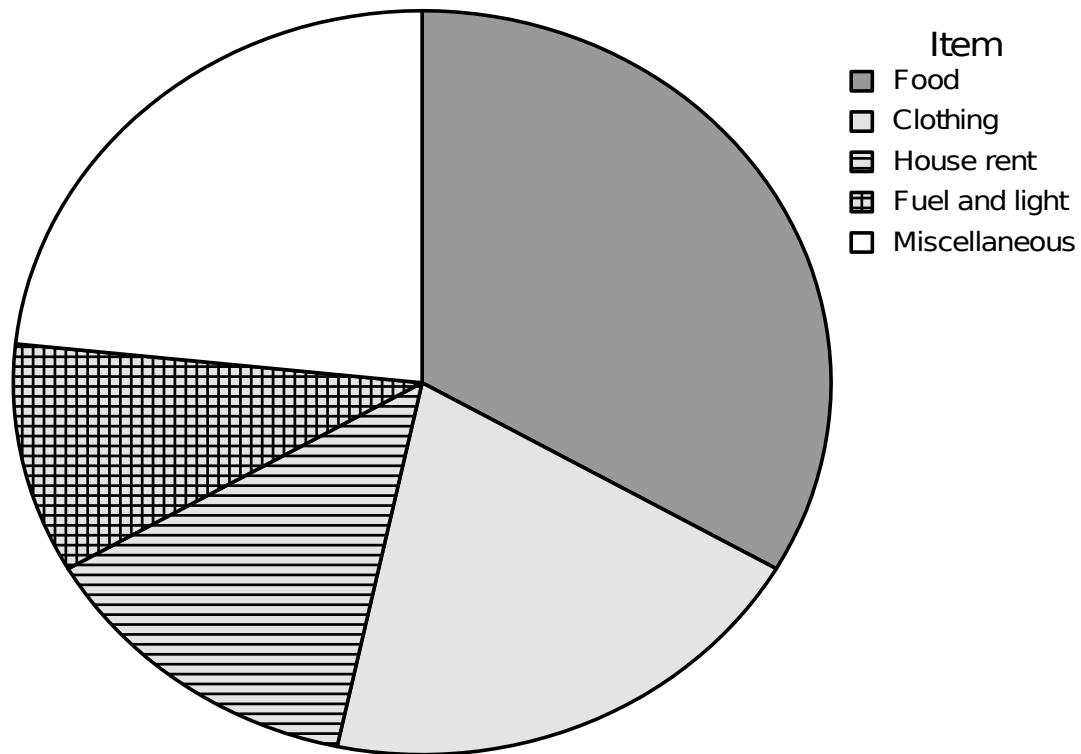
Steps to construct pie charts

- Calculate percentage frequency of each component
$$(f_i/n)*100$$
- Calculate the degree measures of each sector
$$(f_i/n)*360$$
- Draw the circle using protractor and compass

Example: Draw a pie-chart to represent the following data on a certain family expenditure.

Item	Food	Clothing	House rent	Fuel & light	Miscellaneous	Total
Expenditure	50	30	20	15	35	150
% frequency	33.33	20	13.33	10	23.33	100
Angles	120°	72°	48°	36°	84°	360°

Figure: Pie-chart of the data on family expenditure.



3. Pictogram: is a device used to represent data by means of pictures or small symbols.

Example 2.23: The following table shows the orange production in a plantation from production year 1990-1993. Represent the data by a pictogram.

Production	1990	1991	1992	1993
-------------------	-------------	-------------	-------------	-------------

year				
Amount (in kg)	3000	3850	3500	5000

Figure: Pictogram of the data on Orange productions from 1990 to 1993.

