Chapter 16: Histogram of Grouped Data

OBJECTIVES

At the end of the chapter, students should be able to:

1. State the need for grouping data.
2. Form a grouped frequency table.
3. Calculate class boundaries, class intervals and class marks (mid-marks).

I. Need for Grouping Data

It is necessary to group data when such data contain a large number of values and the range of entries is very high.

The variables need to be grouped into classes in order to have a grouped frequency distribution.

In grouping values, it is very important to let each class contain an equal spread of values, i.e. the class interval.

To make available detailed information about data the classes should not be too many or too few. However, a suitable number of classes should be a minimum of 6 and a maximum of 12.

II. Forming a Grouped
Frequency Distribution

Worked Example 1

Construct a frequency distribution for the table of marks scored by 50 students in an Economics examination.

31	34	4	48	30	39	17	29	5	12
39	45	46	4	41	43	40	34	7	43
18	10	17	34	11	39	29	21	39	35
24	12	30	19	29	44	17	20	46	18
51	17	36	21	16	11	29	7	22	31

SOLUTION

(a) When the number of classes is 8

Range = 51 âˆ’ 4 = 47

$$\text{Class width} = \frac{\text{Range}}{\text{Number of classes}}$$

$$= \frac{47}{8} = 5.875 = 6$$

(nearest whole number)

Hence, the 8 class intervals are 4–9, 10–15, 16–21, 22–27, 28–33, 34–39, 40–45, 46–51.

So, the grouped frequency distribution is as follows:

Class interval	Tally	Frequency
4–9	////	5
10–15	////	5
16–21	//// //// /	11
22–27	//	2
28–33	//// ///	8
34–39	//// ////	9
40–45	//// /	6
46–51	////	4
		50

(b) When the number of classes is 6
Range = 51 âˆ’ 4 = 47

$$\text{Class width} = \frac{47}{6} = 7.83 = 8$$

(nearest whole number)
Hence, the 6 class intervals are 4–11, 12–19, 20–27, 28–35, 36–43, 44–51.
So, the grouped frequency distribution is presented as follows:

Class interval	Tally	Frequency
4–11	//// ///	8
12–19	//// ////	10
20–27	////	5
28–35	//// //// //	12
36–43	//// ////	9
44–51	//// /	6
		50

Need for a class boundary
Finding the class boundary is necessary to take care of the approximated values that may arise when calculating the class intervals. For instance, a value such as 35.4 could be rounded down to 35 while 35.5 could be rounded up to 36.

III. Calculation of Class Boundaries, Class Intervals and Class Marks

(I) Calculating the class boundaries

If 4–9 is the first class interval of a frequency distribution, then 4 is the lower limit and 9 is the upper limit.
To get the lower class boundary of the class interval 4–9, imagine that there is a class of equal width, that is lower than class 4–9, whose upper limit is 1 less than 4 (lower limit of class 4–9). Such upper class limit will be 3. Divide the sum of 3 and 4 (upper limit of class lower than class 4–9) by 2 to get:

$$\frac{3+4}{2} = \frac{7}{2} = 3.5$$

So, 3.5 is the lower class boundary of class 4–9. To determine the upper class boundary of class 4–9, add the upper limit 9 to the lower limit of the next higher class
10–15, that is 10, and divide the result by 2to get

$$\frac{9+10}{2} = \frac{19}{2} = 9.5.$$

Meanwhile, the class boundary of class 4–9 is 3.5–9.5.

Apply the same method to find the class boundaries of the rest of the class.
Inclusion of class boundaries in the two frequency tables will now be as follows:

1.

Class interval	Frequency	Class boundaries
4–9	5	3.5–9.5
10–15	5	9.5–15.5
16–21	11	15.5–21.5
22–27	2	21.5–27.5
28–33	8	27.5–33.5
34–39	9	33.5–39.5
40–45	6	39.5–45.5
46–51	4	45.5–51.5

2.

Class interval	Frequency	Class boundaries
4–11	8	3.5–11.5
12–19	10	11.5–19.5
20–27	5	19.5–27.5
28–35	12	27.5–35.5
36–43	9	35.5–43.5
44–51	6	43.5–51.5
	50	

`

(II) Class width

Class width is calculated as the positive difference between the upper class boundary and the lower class boundary.
For instance, the width for class 4–9 is
9.5–3.5 = 6.0. This method is applicable to other classes.

(III) Class limit

In class interval 4–9, 4 is regarded as the lower class limit while 9 is the upper class limit. This method is applicable to other class intervals.

(IV) Class mark

The calculation of the class mark of a class interval is the division of the sum of the lower class limit and the upper class limit by 2. From the class interval 4–9, the class mark is calculated as

$$\frac{4+9}{2}=\frac{13}{2}=6.5.$$

The same process applies to other class intervals.

IV. Histogram

(i) Meaning

A histogram is a graph containing bars which are positioned side by side. It is a set of rectangular bars constructed to display the proportion of items it represents.
Histogram is the graphical or pictorial representation of the frequency distribution.
In constructing a histogram, the frequencies are located along the vertical axis (ordinate) or *y*-axis while the score intervals are located along the horizontal axis (abscissa) or *x*-axis. The width of each bar represents the size of the class interval. In plotting a histogram, it is advisable to use a graph sheet.

There are two ways in which the bases of the rectangles of a histogram are marked:
(a) By marking the class boundaries of the lines which make up the rectangle.
(b) By marking the mid-points of the bases of the rectangles and then fixing the class marks.
WORKED Example 2
The table below shows the mass in kg of 50 boys in a particular village.

Mass (kg)	41	42	43	44	45	46
Frequency	6	10	5	8	13	8

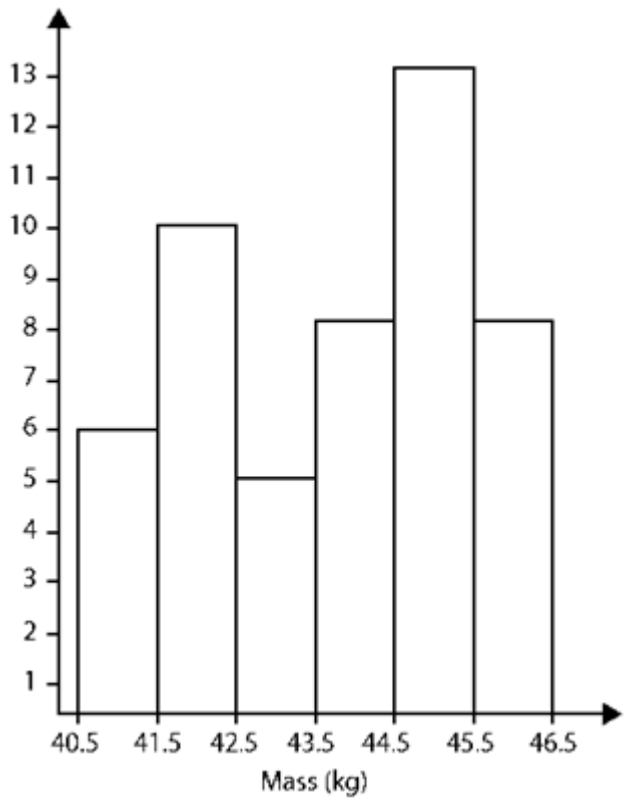
Draw a histogram to represent the information.
SOLUTION
(a) Method 1: Marking the class boundaries
The class boundaries of an ungrouped data can be found as follows:

1. Subtract the first mass 41 kg from the second mass 42 kg to get 42 kg–41 kg = 1 kg.
2. Divide the result of step 1 by 2 to get $\frac{1}{2}$ kg = 0.5 kg.
3. Take away 0.5 kg from each mass to obtain the lower class boundary and add 0.5 kg to each mass to obtain the upper class boundary.

Hence, the table will be written as

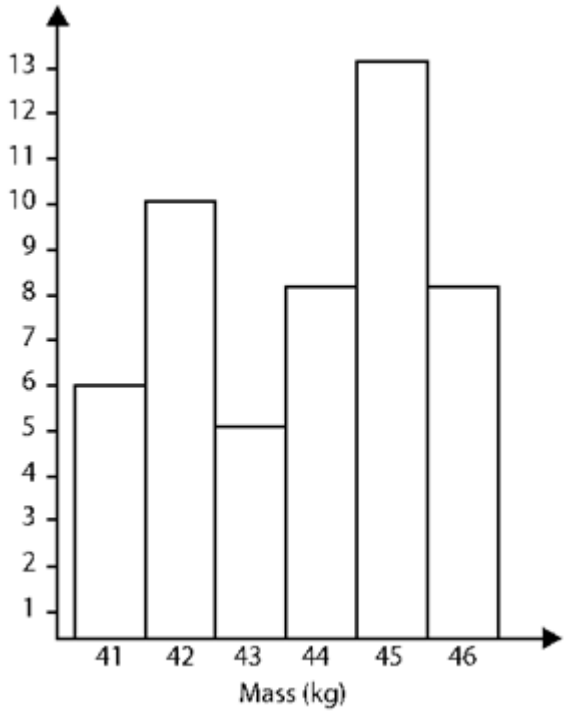
Mass (kg)	Frequency	Class boundaries
41	6	40.5–41.5
42	10	41.5–42.5
43	5	42.5–43.5
44	8	43.5–44.5
45	13	44.5–45.5
46	8	45.5–46.5

However, represent the frequency distribution on the histogram. Thus, the following is a histogram showing the mass, in kg, of 50 boys in a village.



(c) Method 2: Taking the class mark
Taking the class mark instead of the class boundaries is another important method used in representing information on the histogram. It is often used for discrete data. The mark is located on the variable axis at the middle of the rectangular bar. The information in the worked example can be represented on the histogram. Thus, the following is a histogram showing the mass, in kg, of 50

boys in a village.



V. Histogram of a Grouped Frequency Distribution

A grouped frequency distribution can be constructed in two ways similar to that of the ungrouped data.

Worked Example 3

The following data shows the marks of 40 students in a history examination:

41	52	37	56	63	48	65	46
54	32	51	66	74	23	35	61
58	44	49	53	45	57	56	38
59	28	50	49	67	56	36	45
70	68	43	56	26	47	55	71

(a) Form a grouped frequency table with the class interval 20–29, 30–39, 40–49, etc.

(b) Present the information on a histogram.

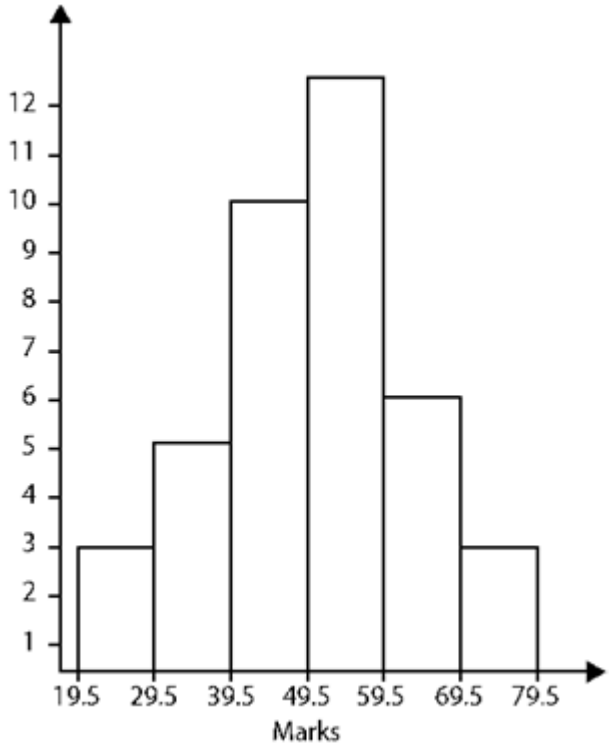
SOLUTION

(I) The following is a frequency table with columns, class interval, tally, frequency, class mark and class boundary.

Class interval	Tally	Frequency	Class mark	Class boundary
20–29	///	3	24.5	19.5–29.5
30–39	~~////~~	5	34.5	29.5–39.5
40–49	~~//// ////~~	10	44.5	39.5–49.5
50–59	~~//// //// ///~~	13	54.5	49.5–59.5
60–69	~~//// /~~	6	64.5	59.5–69.5
70–79	///	3	74.5	69.5–79.5

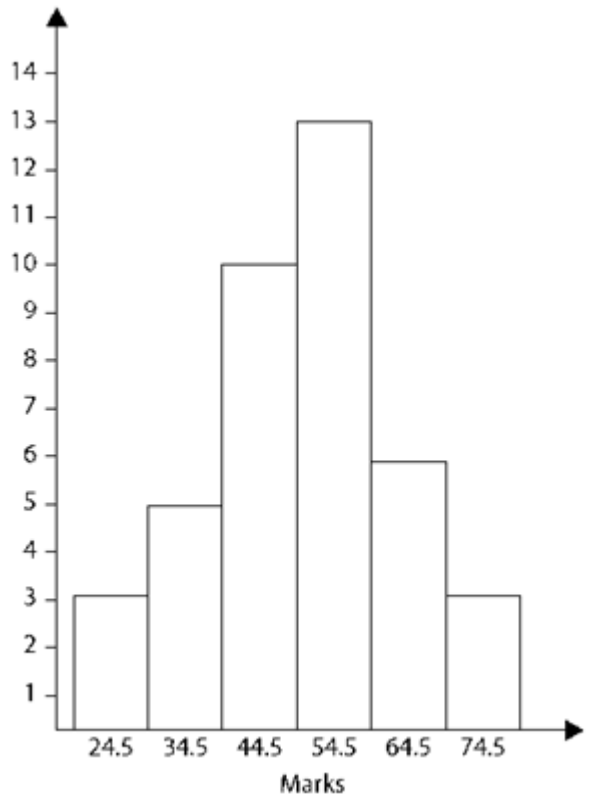
(II) (i) Method 1: Marking the class boundaries

Histogram showing the marks of 40 students in a history examination



(ii) Method 2: Marking the class marks or mid-values

Histogram showing the marks of 40 students in a history examination



Worked Example 4

The marks obtained by 40 students in an examination are as follows:

85	77	87	74	77	78	79	89	85	90
78	73	86	83	91	74	84	81	83	75
77	70	81	69	75	63	76	87	61	78
69	96	65	80	84	80	77	74	88	72

(a) Copy and complete the following table for the distribution using the above data:

Class boundaries	Tally	Frequency
59.5–64.5		
64.5–69.5		
69.5–74.5		
74.5–79.5		
79.5–84.5		
84.5–89.5		
89.5–94.5		
94.5–99.5		

(b) Draw a histogram to represent the distribution.

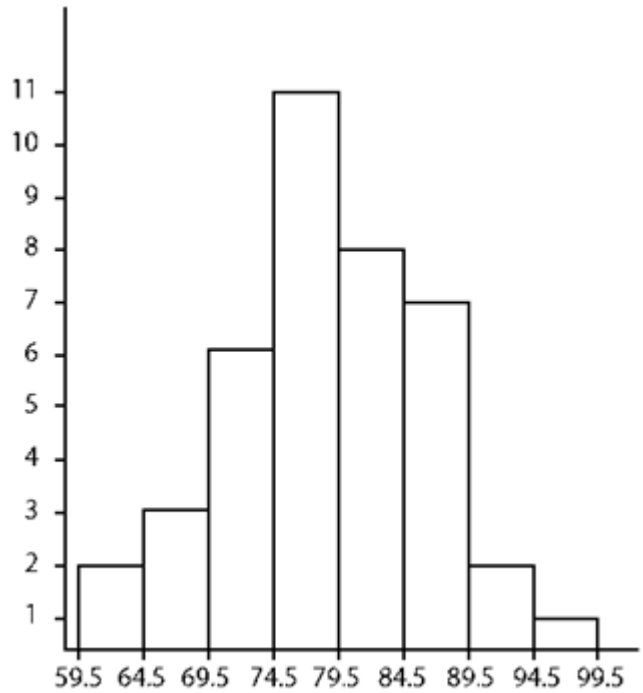
(c) Using your histogram, estimate the modal mark. (WAEC Jun 2003:9)

SOLUTION

(a)

Class boundaries	Tally	Frequency
59.5–64.5	//	2
64.5–69.5	///	3
69.5–74.5	~~###~~ /	6
74.5–79.5	~~###~~ ~~###~~ /	11
79.5–84.5	~~###~~ ///	8
84.5–89.5	~~###~~ //	7
89.5–94.5	//	2
94.5–99.5	/	1

(c) Histogram representing the marks obtained by 40 students in an examination



Exercise 1

1. In an examination, the percentage marks scored by 50 students were as follows:

77	62	77	94	92	66	53	70	80	34
56	52	67	51	43	89	85	81	63	85
39	60	62	72	62	34	74	51	71	43
57	47	51	70	40	35	42	66	27	24
97	42	53	55	72	71	59	51	87	45

(a) Prepare a frequency distribution table, taking class intervals 21–30, 31–40, etc.

(b) Represent the above data on a histogram and use it to estimate the mode.

(c) Calculate the mean mark.

(WAEC Jun 81:5)

2. The following are the marks scored in a class of 30 boys in an examination: 45, 19, 62, 29, 42, 27, 40, 62, 56, 51, 11, 70, 68, 60, 81, 15, 60, 30, 41, 37, 9, 54, 35, 37, 45, 49, 85, 46, 76 and 45.

(a) Form the frequency table.

(b) Identify the modal class.

(c) Calculate the mean mark.

(d) Draw a histogram to represent the distribution.

(WAEC JUN 82:17)

3. The marks scored by 50 students in an examination are as follows:

30	45	48	55	39	25	31	12	18	21
54	59	51	33	43	44	10	38	19	26
41	35	37	41	46	33	51	37	58	48
17	19	23	26	29	38	57	36	35	44
43	27	31	43	22	31	47	34	18	15

Prepare a frequency table with the class intervals 10–19, 20–29, 30–39, etc. and draw a histogram to illustrate the information.

(WAEC JUN 86:12)

4. The marks scored by 30 students in a particular subject are as follows:

39	31	50	18	51	63
10	34	42	89	73	11
33	31	41	25	76	13
26	23	29	30	51	91
37	64	19	86	9	20

(a) Prepare a frequency table, using class interval 1–20, 21–40, etc.

(b) Use the table to draw a histogram.

(c) Estimate the mode from the histogram. (WAEC NOV 89:9)

5. The table below shows the marks obtained by 40 pupils in a Mathematics test.

Marks	0–9	10–19	20–29	30–39	40–49	50–59
No. of pupils	4	5	6	12	8	5

(a) Draw a histogram for the mark distribution.

(b) Use your histogram to estimate the mode. (WAEC JUN 96:3)

6. In an examination, the marks scored by 50 candidates were as follows:

23	47	32	41	29	61	35	45	30	38
37	48	49	47	52	55	62	54	59	58
57	58	56	53	71	59	60	53	95	82
79	83	78	63	75	72	73	65	77	77
92	91	84	64	70	69	67	76	84	83

(a) Prepare a frequency distribution table, using class intervals 21–30, 31–40, etc.

(b) Represent the data on a histogram.

7. The following are the lengths in cm of 50 planks cut by a machine in a saw mill.

33	49	60	58	59	71	42	88	68	91
54	32	81	59	41	55	38	56	86	62
50	69	50	84	77	33	71	42	69	93
61	51	23	76	63	96	26	70	66	80
44	52	46	33	68	39	61	71	48	66

(a) Using class intervals of 21–30, 31–40, etc., construct the frequency table.

(b) Draw a histogram for the distribution.
(c) Identify the modal class.
(d) Use the histogram to estimate the mode of the distribution.
(WAEC NOV. 1996:11)

8. The ages, in years, of 50 teachers in a school are as follows:

21	37	49	27	49	42	26	33	46	40
50	29	23	24	29	31	36	22	27	38
30	26	42	39	34	23	21	32	41	46
46	31	33	29	28	43	47	40	34	44
26	38	34	49	45	27	25	33	39	40

(a) Form a frequency distribution for the data using the intervals 21–25, 26–30, 31–35, etc.
(b) Draw the histogram of the distribution.
(c) Use the histogram to estimate the mode.

9. The table below shows the frequency distribution of marks obtained by 30 students in a test.

Mark	1–5	6–10	11–15	16–20	21–25
Frequency	4	6	11	8	1

(a) Draw a histogram for the distribution.
(b) Use the histogram to estimate the mode.
(WAEC NOV. 06:3)

10. The table below shows the age distribution of members of a club.

Age (years)	10–14	15–19	20–24	25–29	30–34	35–39
Frequency	7	18	25	17	9	4

(a) Draw a histogram to illustrate the information.
(b) Use the histogram to estimate the modal age.
(WAEC JUN 2005:11b)

VI. Estimation of Mode from a Histogram

The mode of a grouped data can be obtained in two ways:

1. By estimation from the histogram.
2. By calculation using formula.

Estimation of mode from a histogram shall be discussed in this chapter while calculation of the mode of a grouped data using formula shall be discussed in Chapter 18 of this book.
Recall that the mode of an ungrouped data is the number that occurs most in a set of given numbers.
Estimation of mode from a histogram is as follows:

1. Construct a histogram from a given frequency table.
2. Identify the highest rectangle from the histogram.
3. Write the class interval that contains the highest rectangle as the modal class.
4. Join the right corner of the bar of the highest rectangle to the right corner of the bar of the left adjacent rectangle with a straight line.
5. Also, join the left corner of the bar of the highest rectangle to the left corner of the bar of the right adjacent rectangle with a straight line.
6. Mark the meeting point of the two straight lines.
7. Draw a perpendicular line from the meeting point of the two straight lines to the base of the rectangle.
8. The point at which the perpendicular line meets the bar of the rectangle that contains the modal class is called the mode.

Worked Example 5
The following table shows the age distribution of members of a club:

Table 16.1

Age (years)	10–14	15–19	20–24	25–29	30–34	35–39
Frequency	7	10	12	2	9	4

1. Draw a histogram to illustrate the information.
2. Use the histogram to estimate the modal age.

SOLUTION

Table 16.2

Class interval	Frequency (*f*)	Class boundaries
10–14	7	9.5–14.5
15–19	10	14.5–19.5
20–24	12	19.5–24.5
25–29	2	24.5–29.5
30–34	9	29.5–34.5
30–39	4	34.5–39.5

Histogram showing the age distribution of members of a club

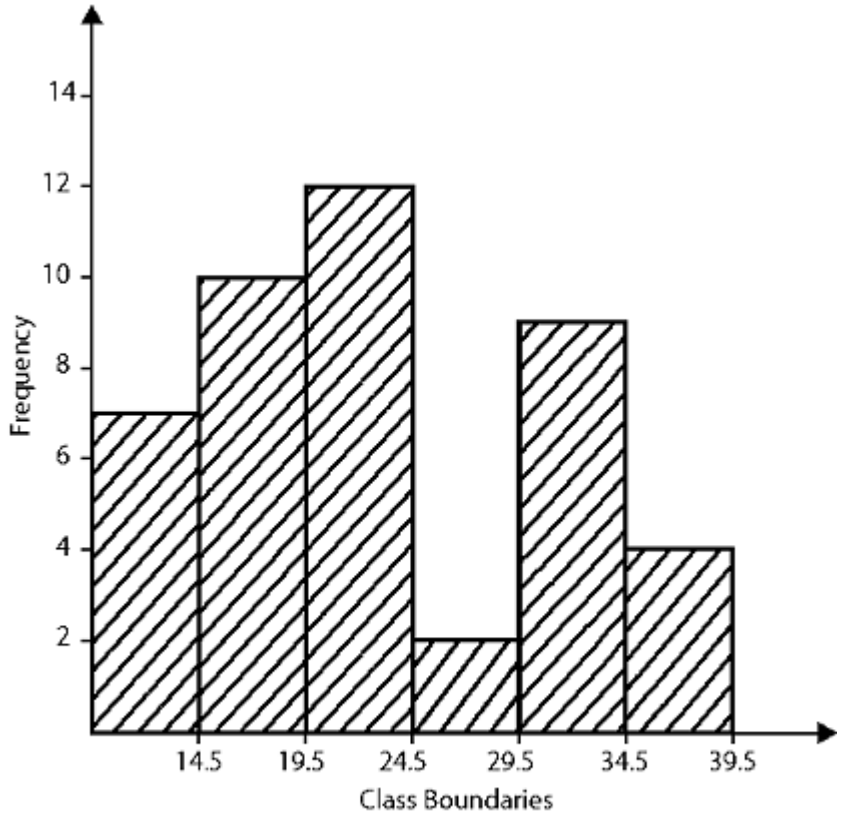


Figure 16.1

SUMMARY

In this chapter, we have learnt the following:

â– Grouping data is necessary where there are a large number of values and the range is very high.

â– In forming a grouped data frequency distribution, suitable number of classes should be a minimum of 6 and a maximum of 12.

â– Finding the class boundary is important in that it takes care of the approximated values that may arise in the process of calculating the class intervals.

â– Class width is the difference between the upper class boundary and the lower class boundary.

â– Class mark is the division of the sum of the lower class limit and upper class limit by 2.

â– Histogram is a graph containing bars which are positioned side by side. It is a set of rectangular bars constructed to display the proportion of items it represents.

â– Marking the bases of a histogram is done by marking the class boundary of lines constituting the rectangle or the mid-point of the base of the rectangles where the class mark is fixed.

â– Mode can be estimated from the histogram.