

Histogram

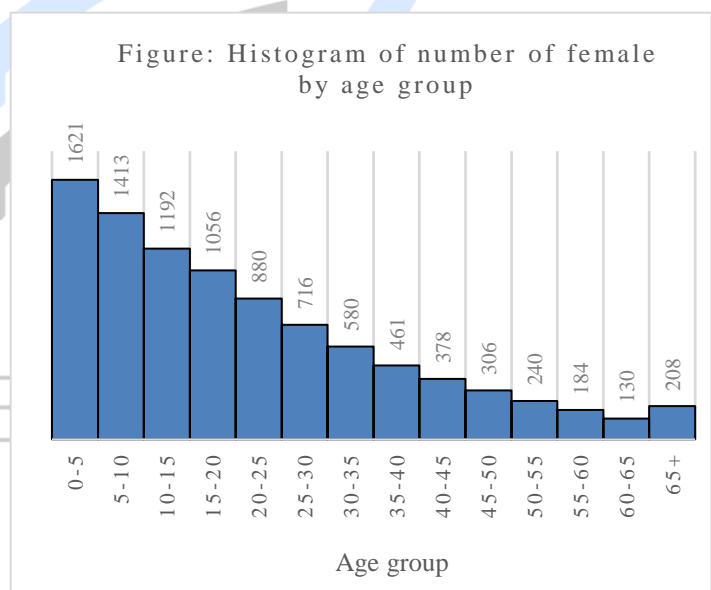
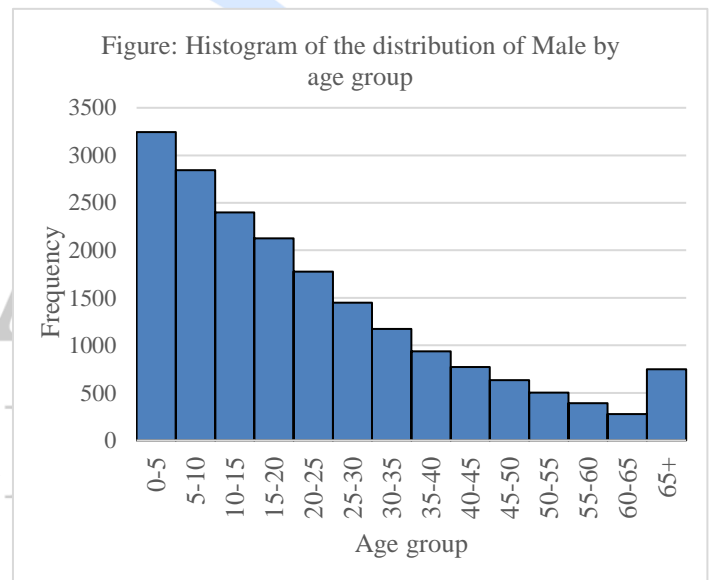
A **histogram** is a bar chart for **grouped numerical** data in which the frequencies or percentages of each group of numerical data are presented as individual vertical bars. In a histogram there are no gaps between adjacent bars as there is in a bar chart of categorical data.

Usually in histogram

- The variable of interest is displayed or plotted along the horizontal (X) axis.
- Frequency or the percentage of the values per class is displayed or plotted along the vertical (Y) axis.

Example:

Table : frequency distribution of male and female by age group		
Age group	Male	Female
0-5	3243	1621
5-10	2842	1413
10-15	2398	1192
15-20	2125	1056
20-25	1776	880
25-30	1450	716
30-35	1173	580
35-40	936	461
40-45	773	378
45-50	633	306
50-55	503	240
55-60	391	184
60-65	278	130
65+	749	208



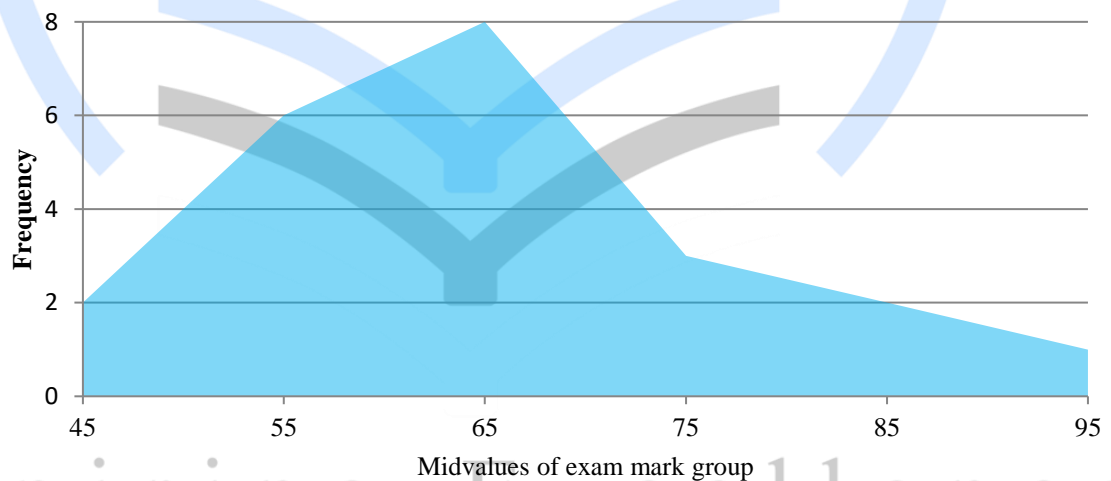
**The Polygon-
The Frequency Polygon**

In constructing frequency polygon, the mid values of the class intervals of the frequency distribution are placed on the horizontal (X) axis and the corresponding frequencies are represented on the vertical (Y) axis. The co-ordinates points thus obtained joined by straight line. The left most point is to be joined with the mid value of the immediate previous interval and the right most co- ordinate point is to be joined with the mid value of the immediate next interval. Thus, we obtain a polygon known as frequency polygon.

Table 1.4: Frequency distribution of students by age group

Marks	Mid value	Frequency
40-50	45	2
50-60	55	6
60-70	65	8
70-80	75	3
80-90	85	2
90-100	95	1

Figure : frequency distribution of marks obtained by students



Inspiring Excellence

The Percentage Polygon

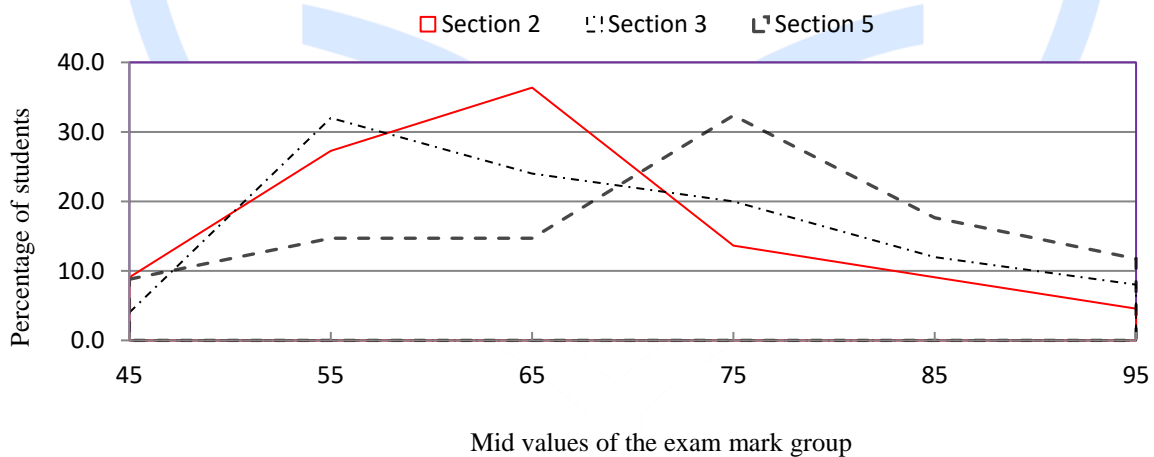
Constructing multiple histograms on the same graph to compare two or more data sets often gets confusing. Super imposing the vertical bars of one histogram on another histogram makes interpretation difficult. When there are two or more groups, one should use a percentage polygon.

A percentage polygon is formed by having the midpoint of each class represent the data in that class and then connecting the sequence of midpoint at their respective class percentages. The following table and figure illustrates the construction of the percentage polygon.

Table: Frequency distribution of Marks obtained by students taught by "X"

Mark Group	Mid value	Frequency of students			Percentage of Students		
		Section 2	Section 3	Section 5	Section 2	Section 3	Section 5
40-50	45	2	1	3	9.1	4.0	8.8
50-60	55	6	8	5	27.3	32.0	14.7
60-70	65	8	6	5	36.4	24.0	14.7
70-80	75	3	5	11	13.6	20.0	32.4
80-90	85	2	3	6	9.1	12.0	17.6
90-100	95	1	2	4	4.5	8.0	11.8
Total		22	25	34	100	100	100

Figure: Comparison of percentage distribution of grades obtained by students of taught by "X"

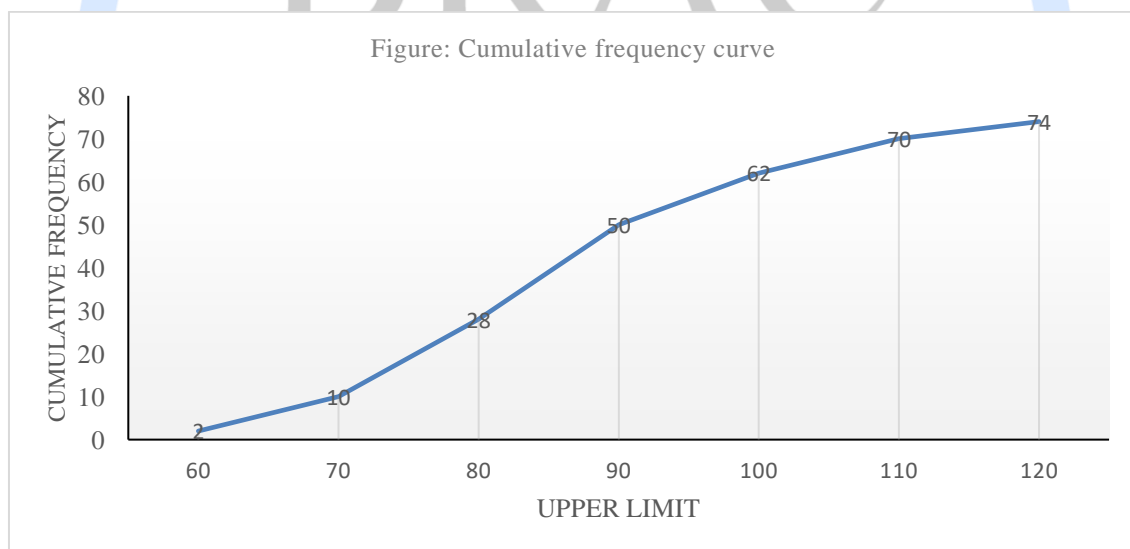


Inspiring Excellence

Ogive or Cumulative frequency curve:

Ogive or a cumulative frequency graphs is drawn by plotting the cumulative frequency of a class against the upper limit of the corresponding class. In these graphs the exact limits of the class intervals are plotted along the X-axis and the cumulative frequencies are plotted along the Y-axis. Coordinate points thus obtained is then joined with a line drawing smoothly. This will result in curve called ogive or cumulative frequency curve. Such curve is usually used to determine the values of certain quantities such as median, quartile, percentile etc.

Class interval of the amount of sell	upper limit	frequency	Cumulative frequency
50 - 60	60	2	2
60 - 70	70	8	10
70 - 80	80	18	28
80 - 90	90	22	50
90 - 100	100	12	62
100 - 110	110	8	70
110 - 120	120	4	74



Inspiring Excellence

The Stem and Leaf Display

To construct a Stem and Leaf plot each numerical value is divided into two parts. The leading digit(s) becomes the stem and the trailing digit the leaf. The stems are located along the vertical axis and the leaf values are staked against each other along the horizontal axis

Stem and leaf plot are a graphical technique of representing quantitative data that can be used to examine the shape of a frequency distribution, the range of the values and point of concentration of the values. This is, in essence a display technique taken from the area of statistics called exploratory data analysis (EDA).

Tukey (1977) first proposed the technique. It allows us to use the information contained in a frequency distribution to show

- The range of score
- Concentration of scores
- The shape of the distribution
- Presence of any specific values or scores not represented in the entire data set
- Whether there are any stray or extreme values in the distribution.

Example:

1. The following data represented the marks obtained by 20 students in a statistics test.

84 17 78 45 47 53 76 54 75 22
66 65 55 54 51 33 39 19 54 72

Use the stem leaf plot to display the data.

The stem leaf plot for the given data

Stem	Leaf
1	7,9
2	2
3	3,9
4	5,7
5	3,4,5,4,1,4
6	6,5
7	8,6,5,2
8	4

After arranging the stem leaf plot we get for the given data

Stem	Leaf
1	7,9
2	2
3	3,9
4	5,7
5	1,3,4,4,4,5
6	5,6
7	2,5,6,8
8	4

2. Form an ordered array, given the following data from a sample of n=8 midterm exam scores in math:

63 99 68 72 79 83 71 62

3. Form an stem and leaf display, given the following data from a sample of n=7 midterm exam scores in physics:

70 44 79 88 83 73 84

Inspiring Excellence

Which method to use tabulation or graphical representation

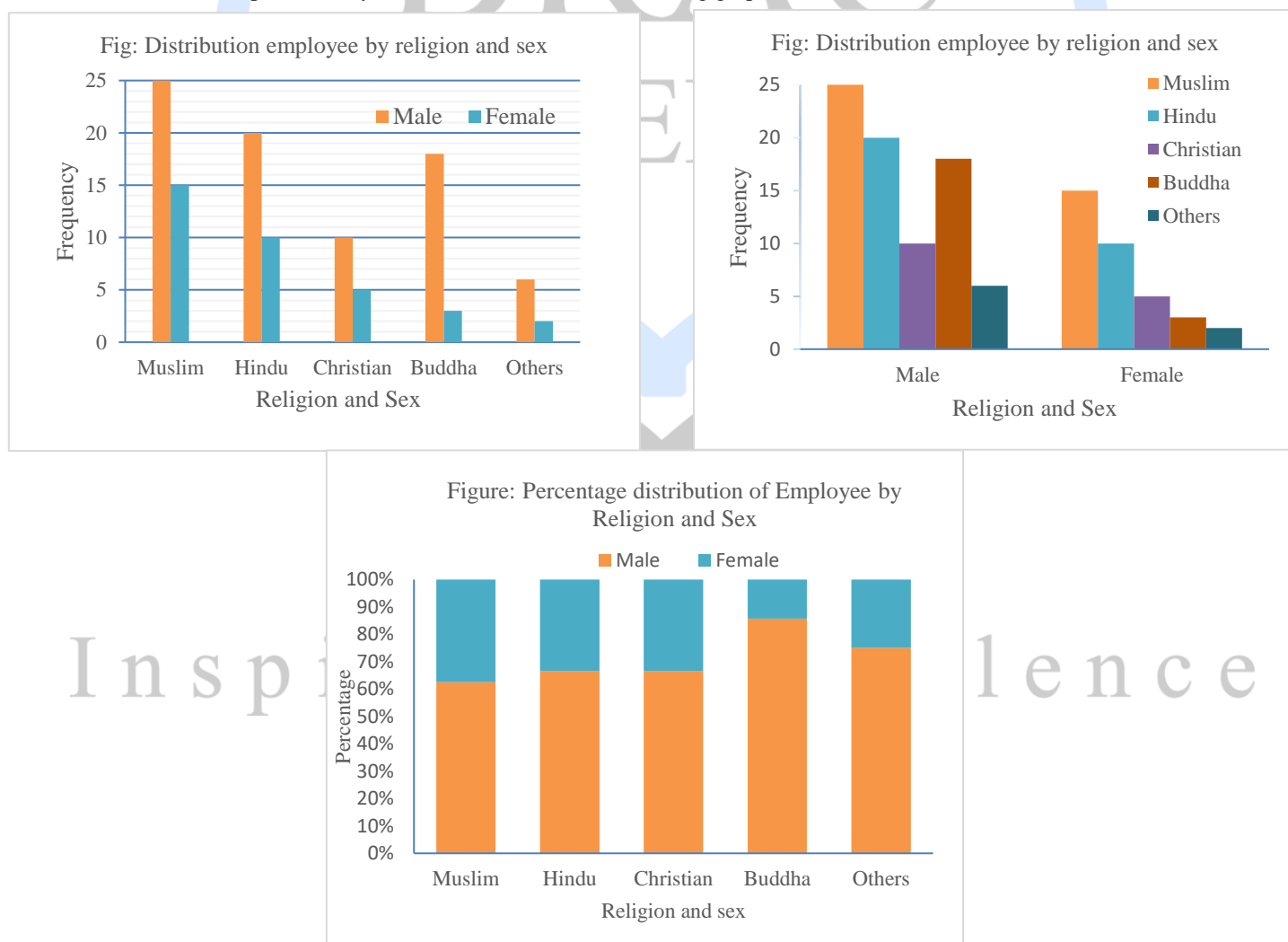
Both the tabulation and the graphical representation are means of data condensation. Researcher can use any of the method to present the data before the audience. A single data can be presented either by tabular form or by graphical form. Also, there can be different options to represent the data in either ways. It is up to the research, who needs to decide on which method to use. With this regards the researcher will decide based on the fact which method is best suited to express the objective of the researcher.

For example –

The following contingency table presents the summarization of two categorical variables.

Religion	Sex		Total
	Male	Female	
Muslim	25	15	40
Hindu	20	10	30
Christian	10	5	15
Buddha	18	3	21
Others	6	2	8
Total	79	35	114

This data can be represented by either or more the of the following graphs -



But depending on the objective, it is up to the researcher which method s/he is going to use.

Test yourself
Assignment 004

The following table represents the information of 40 individuals collected in a socio-economic survey. Using the information given in table 1 answer question A - D

Table 1: Summary information of 40 individuals

Sl. #	Sex	Religion	Previous month's Income	Division	Marital Status
1	M	Islam	1500	Dhaka	Married
2	F	Hindu	3100	Rajshahi	Married
3	M	Buddha	4400	Sylhet	Married
4	M	Christian	5600	Khulna	Unmarried
5	F	Hindu	3858	Dhaka	Divorced
6	M	Islam	9250	Rajshahi	Married
7	M	Islam	7475	Chittagong	Married
8	M	Hindu	7900	Khulna	Unmarried
9	F	Buddha	6600	Rangpur	Divorced
10	F	Islam	7300	Dhaka	Unmarried
11	M	Islam	6100	Barishal	Married
12	M	Buddha	6400	Rajshahi	Married
13	M	Christian	6900	Sylhet	Married
14	F	Islam	9980	Khulna	Unmarried
15	M	Islam	8050	Dhaka	Divorced
16	M	Christian	4500	Rajshahi	Married
17	M	Islam	4950	Chittagong	Married
18	M	Hindu	3865	Dhaka	Unmarried
19	F	Hindu	4800	Rajshahi	Divorced
20	M	Buddha	6200	Sylhet	Unmarried
21	F	Islam	5100	Barishal	Married
22	M	Islam	8789	Rajshahi	Married
23	M	Christian	5556	Sylhet	Married
24	F	Islam	8855	Khulna	Unmarried
25	M	Buddha	7155	Dhaka	Divorced
26	M	Islam	1800	Rajshahi	Married
27	F	Islam	6100	Chittagong	Married
28	M	Christian	4859	Khulna	Married
29	M	Islam	5550	Rangpur	Married
30	F	Christian	4980	Dhaka	Unmarried
31	M	Hindu	6100	Barishal	Divorced
32	F	Islam	6480	Rajshahi	Married
33	M	Christian	6999	Sylhet	Married
34	M	Islam	1200	Khulna	Unmarried
35	F	Christian	8050	Dhaka	Divorced
36	F	Hindu	6500	Rajshahi	Unmarried
37	M	Christian	7050	Chittagong	Married
38	F	Islam	6780	Khulna	Married
39	M	Hindu	4790	Rangpur	Married
40	M	Buddha	6480	Barishal	Married

Question A:

- How many variables are listed in table I?
- Classify the variables according to their types (Qualitative / Quantitative).

Question B:

Construct a frequency distribution table to represent the summary information of the variable "Division" and determine proportion of respondent from Dhaka.

Question C:

Complete the following table # 3 and answer (a) & (b)

Table 3: Frequency distribution of sex by Religion

Sex	Religion				Total
	Islam	Hindu	Christian	Buddha	
Male					
Female					
Total					

- What is the modal response for the variable “Sex”? (*Hint: Most frequent value of sex category*)
- What proportion of respondents are “Buddha”?
- What proportion of males are Christian?
- What proportion Christians are male?
- Construct a side by side bar chart to represent the information given in the table 3.

Question D:

Complete the following table # 4 and answer a), b) & c)

Table 4: Frequency distribution of previous month's income

Income Group	Tally	Frequency	Relative frequency	Cumulative relative frequency
Below – 3000				
3000 – 5000				
5000 – 7000				
7000 – 9000				
9000 +				

- What proportion (Percentage) of people had previous month's income between 3000 - 7000
- What proportion (Percentage) of people had previous month's income less than 7000
- Construct Histogram to display the data represented in table 4.

Inspiring Excellence

For further details of graph study

1. https://www.sheffield.ac.uk/polopoly_fs/1.96442!/file/graphical-presentation-06-07.pdf
2. <https://www.yourarticlelibrary.com/education/statistics/graphic-representation-of-data-meaning-principles-and-methods/64884>



For any queries related to this presentation please contact

IFTEKHAR Mohammad Shafiqul Kalam

Assistant Professor

Department of Mathematics and Natural Sciences

Email: imskalam@bracu.ac.bd

Inspiring Excellence