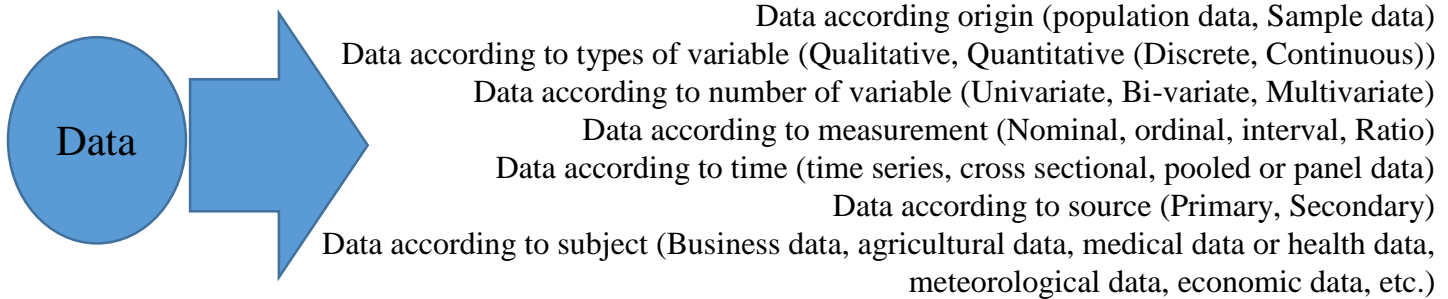


**In the previous lecture we have discussed:**

Data  
 Raw data  
 Different component of data condensation  
 Classification, tabulation and graphical representation  
 Different components of table  
 Class limit, class interval, Relative frequency, cumulative relative frequency etc.



**Contingency table** (also known as a cross tabulation or crosstab): In statistics, a *contingency table* (also known as a cross tabulation or crosstab) is a type of *table* in a matrix format that displays the (multivariate) frequency distribution of the variables. They are heavily used in survey research, business intelligence, engineering and scientific research.

A contingency table presents the results of two or more categorical variables. The joint responses are classified so that the categories of one variable are located in the rows and the categories of the other variable are located in the columns.

The values located at the intersections of the rows and columns are called **cells**. Depending on the type of contingency table constructed, the cells for each row-column combination contain the frequency, the percentage of the overall total, the percentage of the row total, or the percentage of the column total.

Table 1: Frequency distribution of students by religion and sex

Religion	Sex		Total
	Male	Female	
Muslim	25	20	45
Hindu	12	12	24
Christian	8	6	14
Buddha	5	3	8
Others	2	2	4
Total	52	43	95

**Test yourself**  
**W2D1\_ Assignment 002**

**Problem:**

A sample of 500 shoppers was selected in a large metropolitan area to determine various information concerning consumer behavior. Among the questions asked was “do you enjoy shopping for clothing?” the results are summarized in the following cross classified table:

Table 2: Frequency distribution of preference of shopping for  
clothing of the consumer

Enjoy shopping for clothing	Sex		Total
	Male	Female	
Yes	136	224	360
No	104	36	140
Total	240	260	500

- a. Construct contingency tables based on total percentages, row percentages and column percentages.
- b. Determine the proportion
  - i. Male among total respondents
  - ii. Female among total respondents
  - iii. Proportion of people who enjoys shopping
  - iv. Proportion of people who does not enjoys shopping
- c. Determine the proportion (or percentage) of
  - i. Male among those who enjoy shopping.
  - ii. People who enjoys shopping among male

**Test yourself:**  
**W2D1\_ Assignment 002**

**Solution:**

The table was given as -

Table 2: Frequency distribution of preference of shopping for clothing of the consumer

Enjoy shopping for clothing	Sex		Total
	Male	Female	
Yes	136	224	360
No	104	36	140
Total	240	260	500

Workout:

a.

Table 2: Frequency distribution of preference of shopping for clothing of the consumer			
Enjoy shopping	Sex		Total
	Male	Female	
Yes	<b>0.377 or 37.7%</b>	<b>0.622 or 62.2%</b>	<b>100%*</b>
	<b>0.566 or 56.6%</b>	<b>0.861 or 86.1%</b>	
No	<b>0.7428 or 74.28%</b>	<b>0.257 or 25.7%</b>	<b>100%*</b>
	<b>0.433 or 43.3%</b>	<b>0.138 or 13.8%</b>	
Total	<b>100%*</b>	<b>100%*</b>	

\*: ignoring rounding error

b. The proportion of

i. Male among total respondents	$\frac{(136 + 104)}{360} = \frac{240}{500} = 0.48$
ii. Female among total respondents	<i>Do your self</i>
iii. People who enjoys shopping	<i>Do your self</i>
iv. People who does not enjoys shopping	$\frac{(\text{---} + \text{---})}{500} = \frac{140}{500} = \text{---}$

c. Determine the proportion (or percentage) of

i. Male, among those who enjoy shopping.	$\frac{136}{360} = 0.37$
ii. People who enjoys shopping among male	$\frac{136}{240} = \text{---}$

## Graphical Representation of Data

### General Rules for Graphical Representation of Data

There are certain rules to effectively present the data and information in the graphical representation. They are:

- Appropriate title: *The title of the graph should be self-explanatory*. So that audience do not have any confusion in understanding the content of the graph.
- Axis should have appropriate label.
- Categories represented by different axis should be clearly mentioned.
- Measurements units (if any) must be mentioned clearly.
- Use appropriate scaling is a vital issue.
- Legends used in the graph should be duly labeled.
- Data Sources should be included where ever needed.
- Keep it Simple: Construct a graph in an easy way that everyone can understand.

**Different types of charts and graphs:** Data charts are available in a wide variety of maps, diagrams, and graphs. Selection of the most appropriate chart depends on a variety of different factors, such as –

- i Objective the researcher
- ii The nature of the data,
- iii The purpose of the chart, and
- iv Types of variable etc.

- For example, some of the most frequently graphs are used for

Graphical Representation	Situation of uses
Bar Chart Pie-diagram Pareto diagram Pictogram	Usually used for qualitative or categorical or Discrete variable
Stem and Leaf Plot	For frequency distribution of Quantitative data
Box and whisker plot, Normal Q-Q plot	For checking the normality of frequency distribution
Histogram, Frequency polygon, Ogive curve, Dot plot, Frequency curve	For checking nature of continuous data
Lorenz Curve, range or variation curve	For measures of dispersion
Scatter diagram, Venn diagram, Correlogram	For measuring correlation in bivariate data
Line graph, Z-curve, Band graph	For time series data.

**Bar Chart:** In a bar chart, a bar shows each category. The length of the bar represents the amount, frequency or percentage of values falling into a category. **Types of Bar Diagrams.** The following are the various types of bar diagrams in common use:

- (a) Simple bar diagram.
- (b) Sub-divided or component bar diagram.
- (c) Percentage bar diagram.
- (d) Multiple bar diagram.
- (e) Deviation or Bilateral bar diagram.

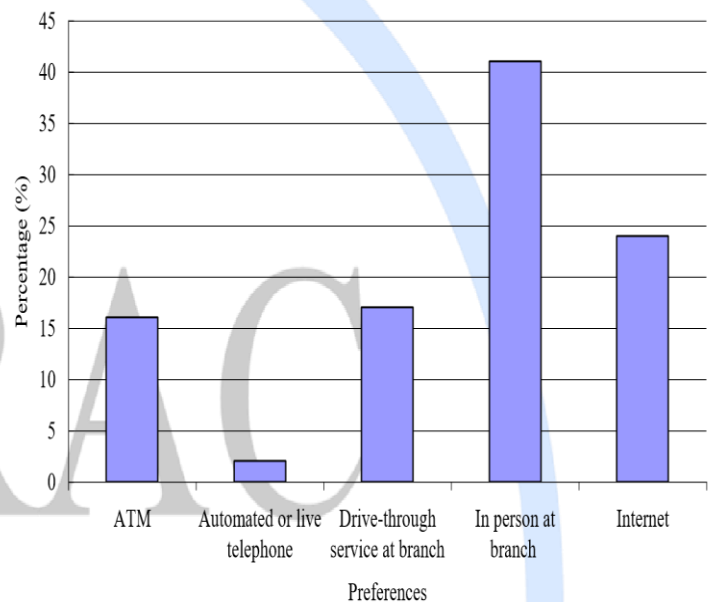
Example 1: Figure 1 displays the bar chart for the people's preference to do their banking as depicted in table 1. Bar chart allows researchers to compare the percentages in different categories. In figure 1:

respondents are most likely to bank in person at a branch and on the internet, followed by drive through service at a branch and ATM. Very few respondents mentioned automated or live telephone.

Table 1: Table of percentage distribution of banking preference of the customer of BANK XYZ

Banking Preference	Frequency (%)
ATM	56(16%)
Automated or live telephone	7 (7%)
Drive-through service at branch	63 (18%)
In person at branch	140 (40%)
Internet	84 (24%)
Total	350 (100%)

Figure 1: Bar chart for Banking Preference



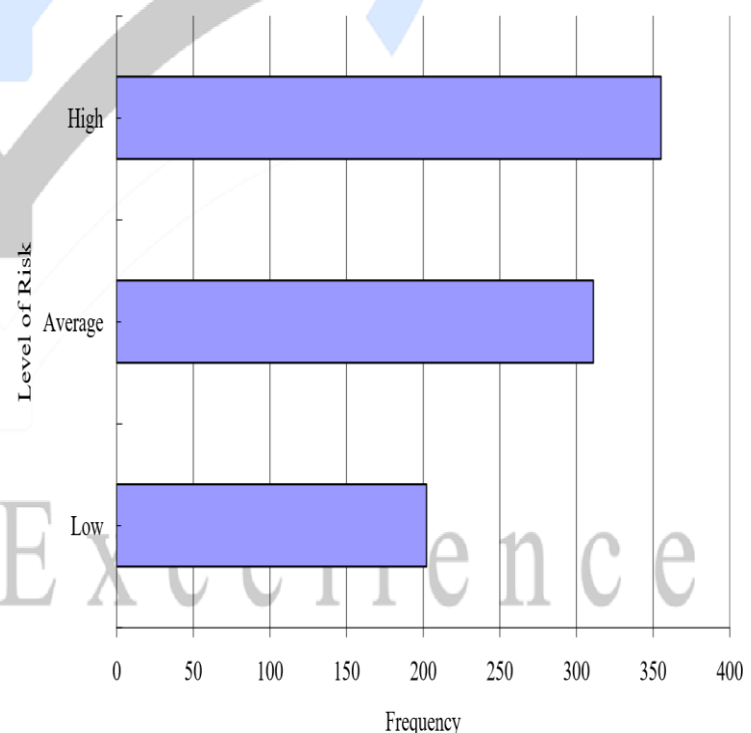
### Example 2:

A sample of 868 mutual funds has been selected and questions were asked to assess and categories the risk associated with the customer's investments in mutual funds. Of the 868 mutual funds 202 funds are classified as the low risk funds, 311 funds are classified as average-risk fund and the rest of 355 funds are categorized as high- risk. Hence the *summary table*<sup>1</sup> of levels of risk of mutual funds is given below.

Table 2: Frequency and Percentage Summary table of Risk Level for 868 Mutual Funds

Fund Risk Level	Number of funds	Percentage of funds (%)
Low	202	23.37
Average	311	35.83
High	355	40.89
Total	864	100.00

Figure 2: Bar Chart for Level of Risk

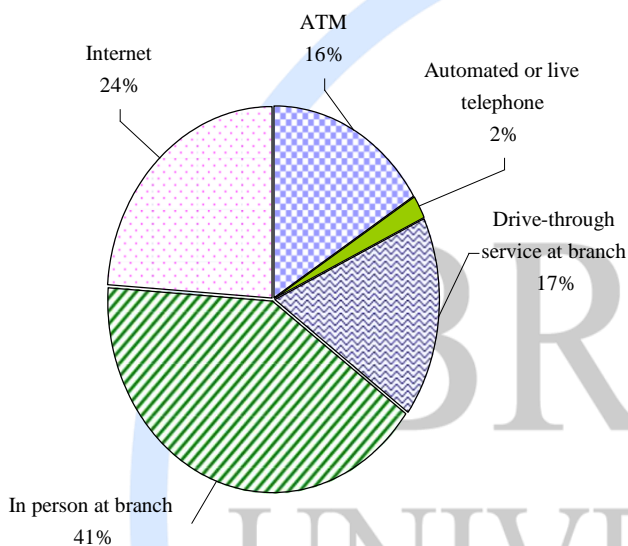


- A **summary table** indicates the frequency, amount or percentage of items in a set of categories so that one can see differences between the categories. A **summary table** lists the categories in one column and the frequency/ percentage / amount in a different column.

The Pie Chart: The Pie Chart is a Circle broken up into slices that represent categories. The size of each slice of the pie varies according to the percentage in each category.

In table 1 of this lecture 16% of the respondents stated that they prefer to bank using ATM. Thus in constructing the pie chart, the 360 degrees that makes up a circle is multiplied by 0.16, resulting in a slice of the pie that takes up 57.6 degrees of the 360 degrees of the circle. In this figure, bank in person at the branch takes 41% of the pie and automated or live telephone takes only 2%.

**Figure 3: Pie Chart for Banking**



In case of pie diagram

Angle of the slice of pie for a particular Category  $\propto$  Frequency (or Percentage) of that particular item.

If the frequency / value / percentage of any component is  $f$  from the whole  $N$  then the angle

of pie for that particular component is

$$\theta^\circ = \frac{f}{N} * 360^\circ$$

**Which chart should one use – a bar chart or a pie chart?**

The selection of a particular chart often depends on the intention of the researcher. If a comparison of categories is most important, one should use a bar chart. If observing the portion of the whole that is in a particular category is most important, one should use a pie chart.

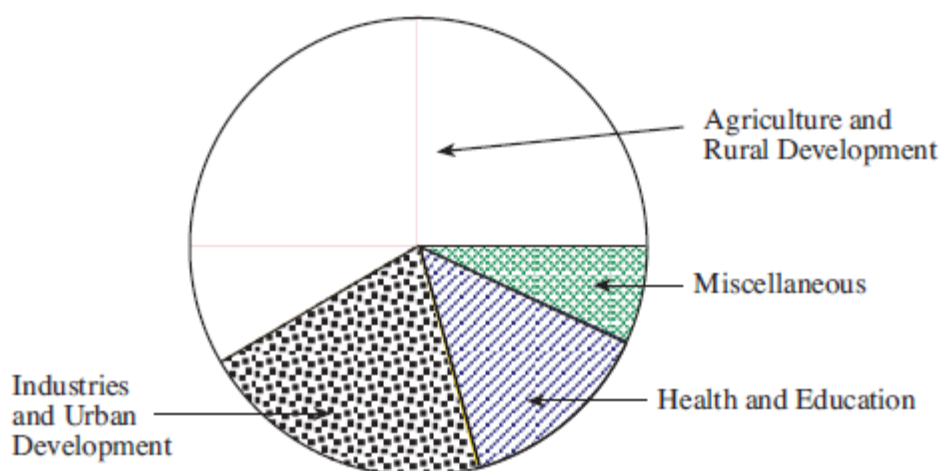
**Draw a pie diagram to represent the following data of proposed expenditure by a Government for the year 2017-18.**

Items	Agriculture & Rural Development	Industries & Urban Development	Health & Education	Miscellaneous
Proposed Expenditure (in million Rs.)	4,200	1,500	1,000	500

**Solution.**

**CALCULATIONS FOR PIE CHART**

Items	Proposed expenditure (in million Rs.)	Angle at the centre
(1)	(2)	(3) = $\frac{(2)}{7200} \times 360^\circ$
Agriculture and Rural Development	4,200	$\frac{42}{72} \times 360^\circ = 210^\circ$
Industries and Urban Development	1,500	$\frac{15}{72} \times 360^\circ = 75^\circ$
Health and Education	1,000	$\frac{10}{72} \times 360^\circ = 50^\circ$
Miscellaneous	500	$\frac{5}{72} \times 360^\circ = 25^\circ$
Total	7,200	360°



**Fig: PIE DIAGRAM REPRESENTING PROPOSED EXPENDITURE BY GOVERNMENT ON DIFFERENT ITEMS FOR 2017-18**

**Histogram:** A histogram can be defined as a set of rectangles with bases along with the intervals between class boundaries. Each rectangle bar depicts some sort of data and all the rectangles are adjacent. The heights of rectangles are proportional to corresponding frequencies of similar as well as for different classes.

A **histogram** is the graphical representation of data where data is grouped into continuous number ranges and each range corresponds to a vertical bar.

- The horizontal axis displays the number range.
- The vertical axis (frequency) represents the amount of data that is present in each range.

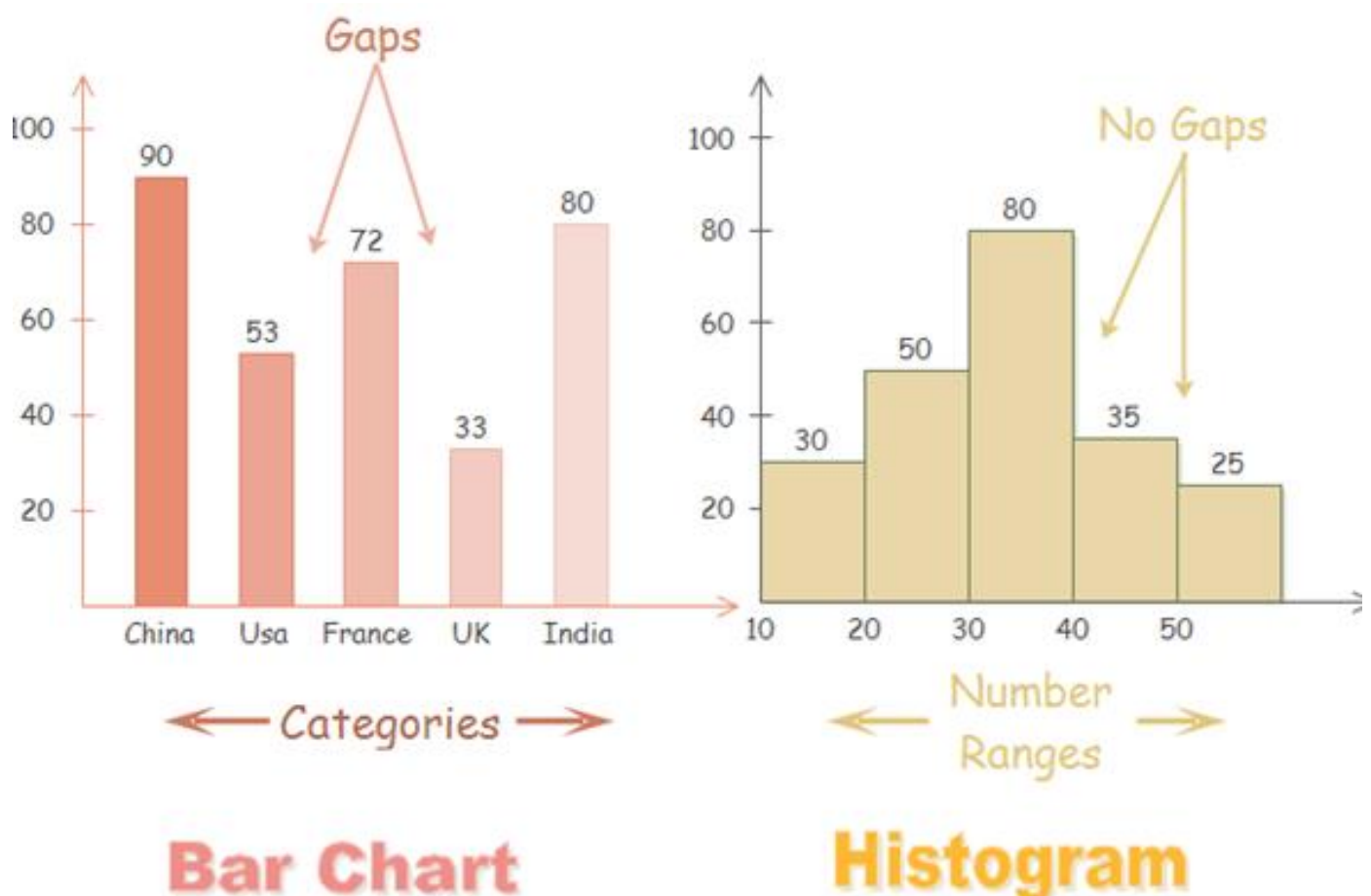
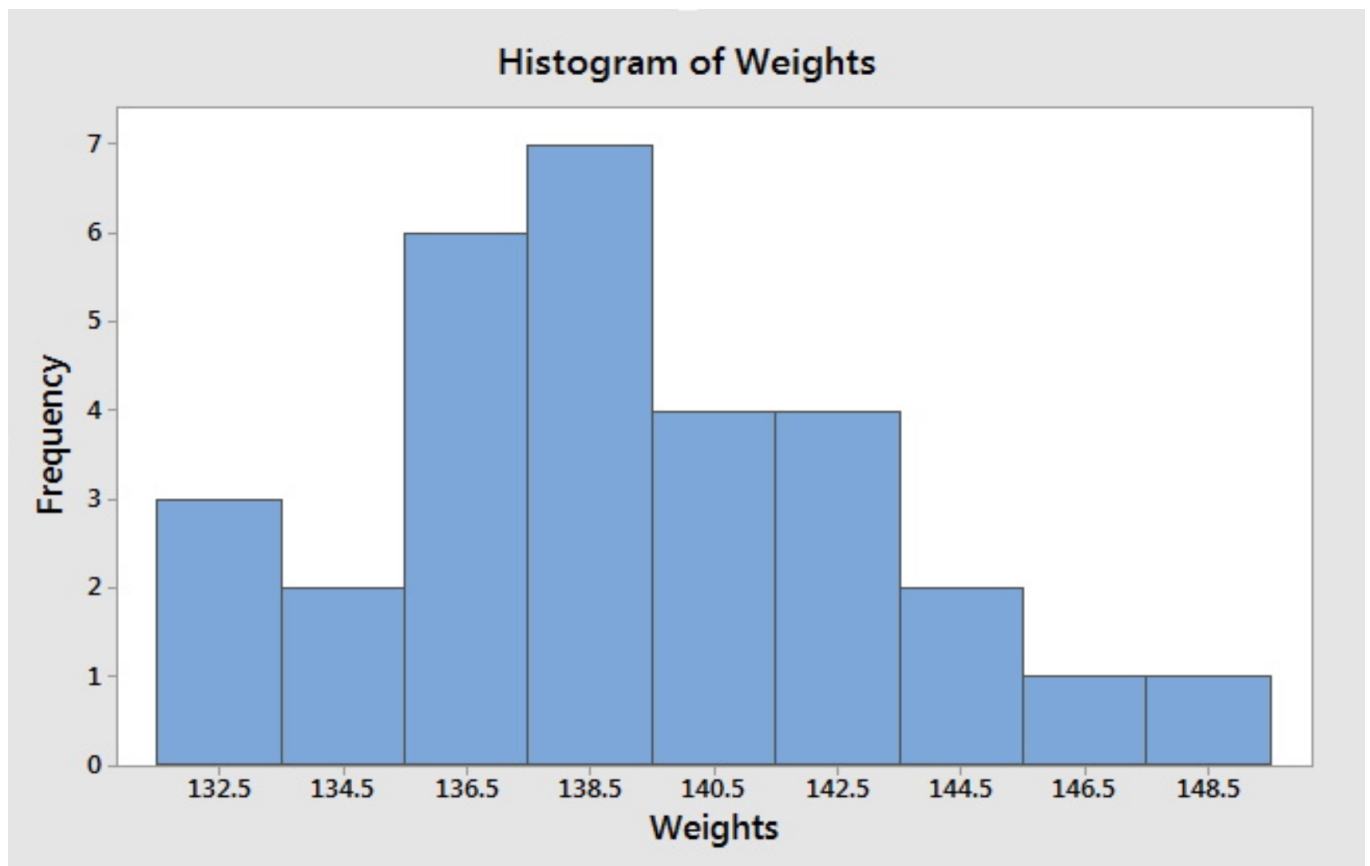
The number ranges depend upon the data that is being used.

**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:





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

Figure: Histogram of the distribution of Male by age group

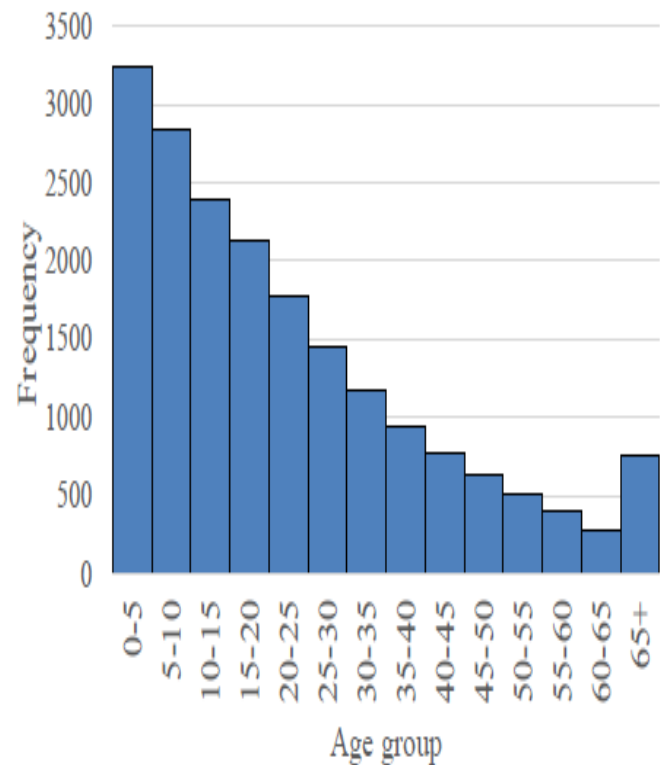
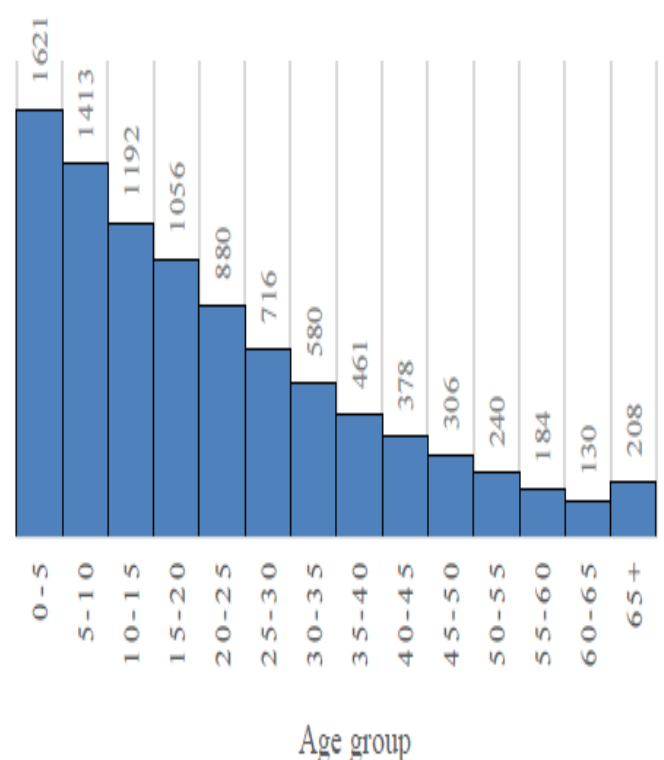


Figure: Histogram of number of female by age group



1. Complete the following table.

Grades on Statistics examination	Frequency	Relative Frequency	Percentage
A: 90 – 100		0.08	
B: 80 – 89	36		
C: 65 – 79	90		
D: 50 – 64	30		
F: Below 50	28		
Total	200	1.00	

2. Following are the different brands of laptops sold form a computer shop.

Y	X	X	Z	X	Y	Y	Y	X	X
Z	X	Y	Y	X	Z	Y	Y	Y	X

- Construct a frequency distribution table for the information given above.
  - Compute the relative frequency for each of the brands and comment on the market share.
  - Display the results, part a, in a frequency bar graph.
  - Display the results, part b, in a pie chart.
3. Assume telecommunication companies in Bangladesh spent about BDT 300 million in advertising. The spending is as follows:

Media	Amount (\$ millions)	Percentage (%)
Radio	20	6.67
Internet	30	10.00
Cinema	5	1.67
Direct mail	15	5.00
Magazines	35	11.67
Newspapers	65	21.67
Outdoor	45	15.00
TV	35	11.67
Other	50	16.67
	300	100

- Construct a bar chart and a pie chart.
  - Which graphical method do you think is best to portray these data?
4. The following data set represents the scores on intelligence quotient (IQ) examinations of 40 sixth-grade students at a particular school:

114	122	103	118	99	105	134	125	117	106
109	104	111	127	133	111	117	103	120	98
100	130	141	119	128	106	109	115	113	121
100	130	125	117	119	113	104	108	110	102

- Organize the data in classes such as 90 – 100, 100 – 110 and so on.
- Present the data set in a frequency histogram.

5. The international Rhino Federation estimates that there are 25280 rhinoceroses living in the wild in Africa and Asia. A breakdown of the number of rhinos of each species is reported in the accompanying table.

Rhino Species	Population Estimate
White rhino	18000
Black rhino	4240
Greater One-horned rhino	2800
Sumatran Rhino	200
Javan Rhino	40
Total	25280

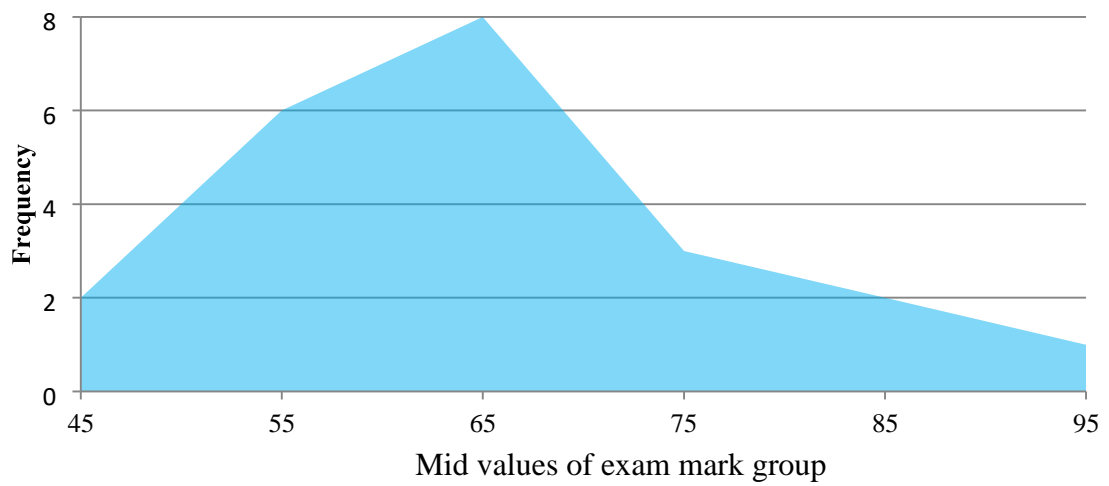
- Construct a relative frequency table for the data.
- Display the frequencies in a bar graph.
- Display the frequencies in a pie chart.
- What proportion of the 25280 rhinos are White rhinos? Black?

- **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

Table: frequency distribution of marks obtained by students



### The Percentage Polygon

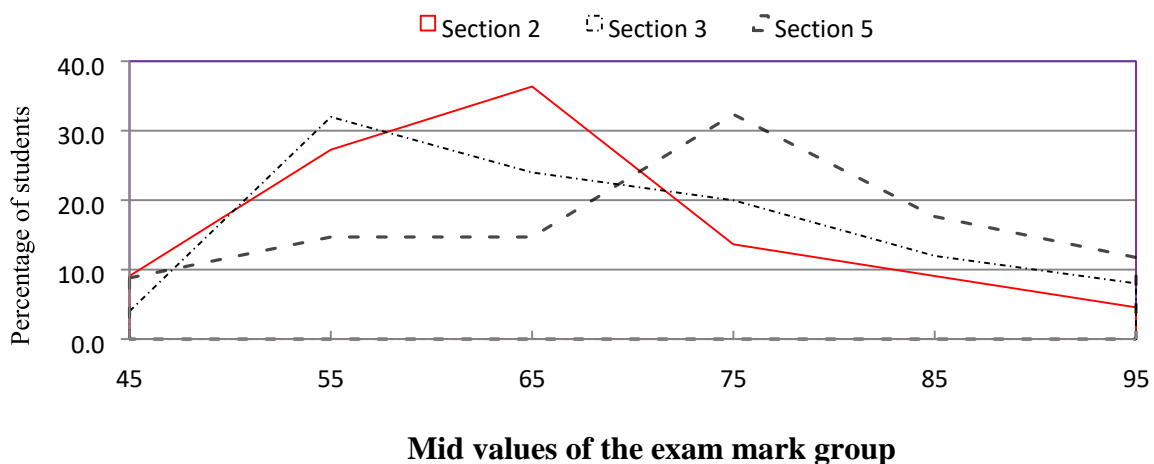
**Constructing multiple histograms on the same graph to compare two or more data sets often gets confusing.** Super imposing the vertical bars of one histogram on another histogram makes interpretation difficult. When there are two or more groups, one should use a percentage polygon.

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

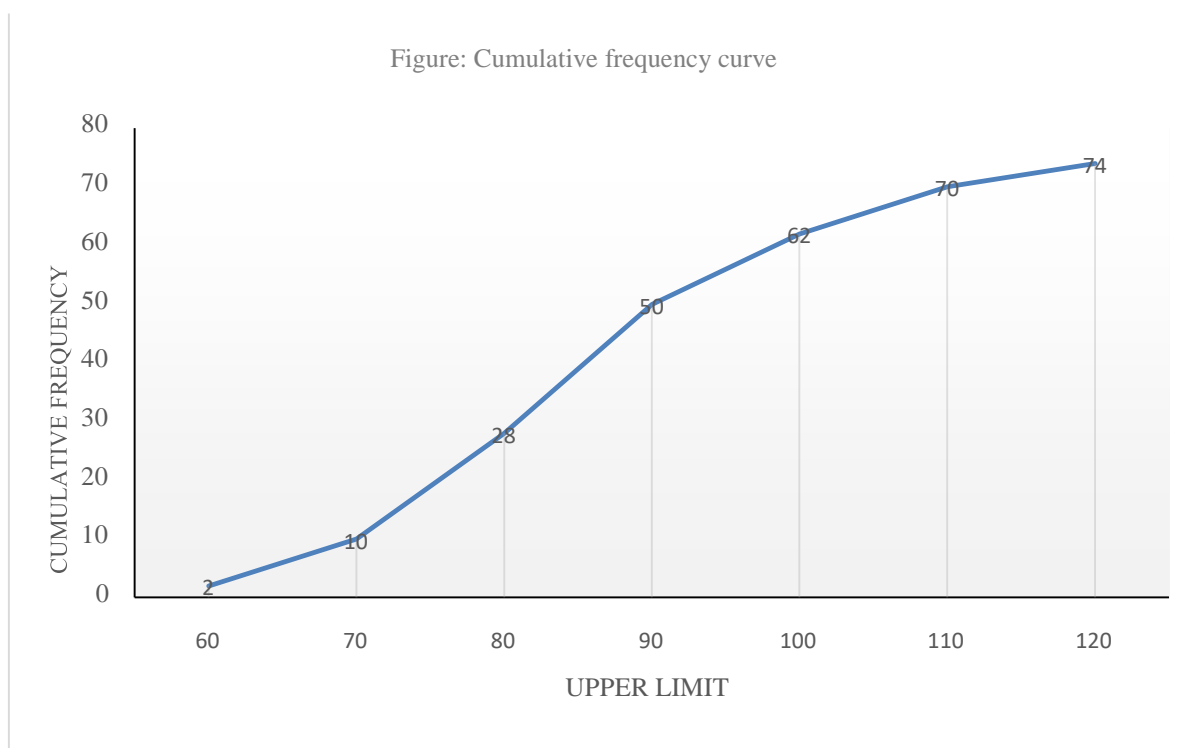
Figure: Comparison of percentage distribution of grades obtained by students of taught by "X"



### 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



### 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

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	3,4,5,4,1,4
6	6,5
7	8,6,5,2
8	4

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

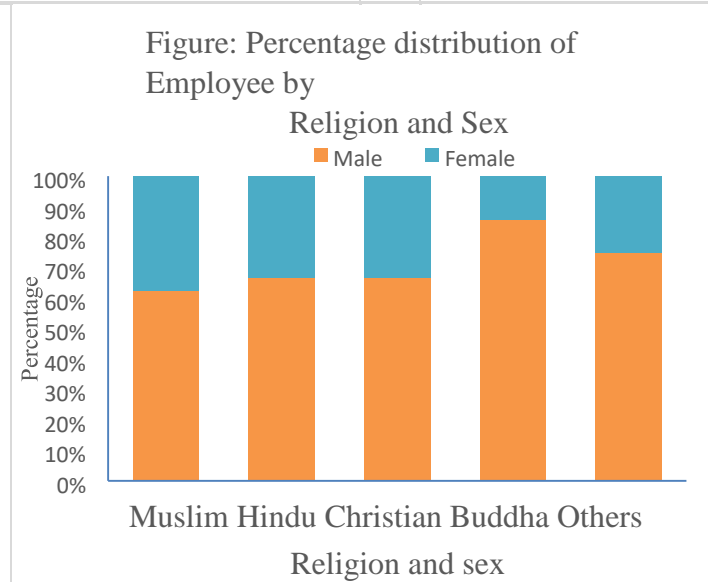
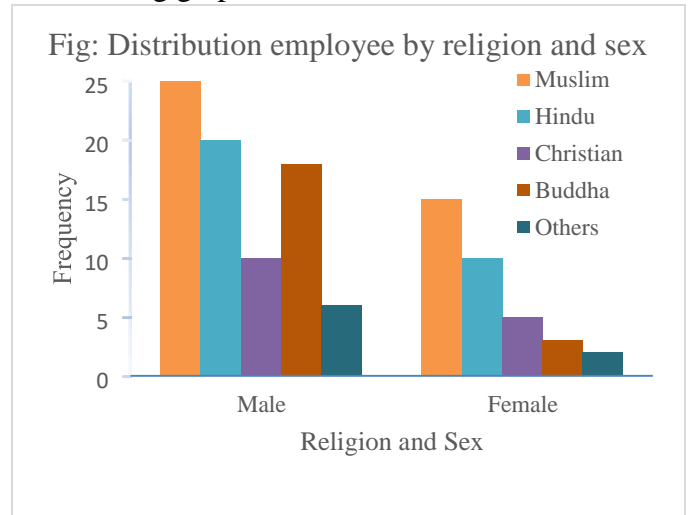
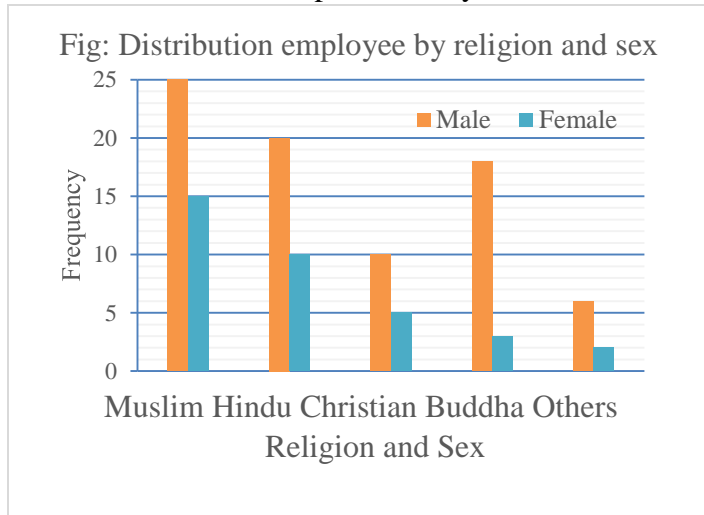
### Which method to use tabulation or graphical representation

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.

For further details of graph study

- [https://www.sheffield.ac.uk/polopoly\\_fs/1.96442!/file/graphical-presentation-06-07.pdf](https://www.sheffield.ac.uk/polopoly_fs/1.96442!/file/graphical-presentation-06-07.pdf)
- <https://www.yourarticlelibrary.com/education/statistics/graphic-representation-of-data-meaning-principles-andmethods/64884>