Chapter 2:

Collection of data
Presentation of statistical data

Prepared by, ***Jannatul Ferdous,*** *Assistant Professor, Dept. of CSE, Metropolitan University, Sylhet*

Collection of data

Data are characteristics or information, usually numerical, that are collected through observation. In a more technical sense, data is a set of values
of *qualitative* or *quantitative* variables about one or more persons or objects, while a **datum** (singular of data) is a single value of a single variable. Data is measured, collected and reported, and analyzed, whereupon it can be visualized using graphs, images or other analysis tools.

Types of data based on source

Questionnaire ……???

Questionnaire **is** a technique for gathering statistical information about the attributes, attitudes, or actions of a population by a structured set of questions.

Sample Questionnaire

Auckland Transport
An Auckland Council Organisation

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

The Development of a Multi-Faceted Evaluation Framework of Shared Spaces

On-Street Perception Survey

Please circle the option best describes your opinion towards the following five statements.

		Strongly Disagree	Disagree	Tend to Disagree	Tend to Agree	Agree	Strongly Agree
1)	"I like spending time in this street"	-3	-2	-1	1	2	3
2)	"I can freely move around on this street"	-3	-2	-1	1	2	3
3)	"Driver behaviour is appropriate in this street"	-3	-2	-1	1	2	3
4)	"This street complements the economic activity"	-3	-2	-1	1	2	3
5)	"I feel safe and secure in this street"	-3	-2	-1	1	2	3

MAIN SURVEY QUESTIONS OF PERFORMANCE MEASURES

Please further provide us with your impression about this street space and your background information

6) Of the five statements above, which are the most important (.....) and least important (.....) ?

7) Of the five statements above, which do you feel could be most improved, and why?

.....

8) What, in particular, do you like most about this street space?

.....

ADDITIONAL SURVEY QUESTIONS

9) How often do you visit this street?

☐ First visit / Very Infrequently ☐ Around once a month
☐ Around once a week ☐ Multiple times a week

10) Why have you visited this street today? (e.g. passing through, shopping or eating)

.....

11) Age: ☐ Under 20 ☐ 20 – 34 ☐ 35 – 49 ☐ 50 – 65 ☐ Over 65

12) Gender: ☐ Male ☐ Female

13) Ethnic group: ☐ NZ European ☐ Maori / Pacific Islands ☐ Asian
☐ Other .....

DEMOGRAPHIC QUESTIONS

Presentation of statistical data

Data are usually collected in a raw format and thus the inherent information is difficult to understand. Therefore, raw data need to be summarized, processed, and analyzed. However, no matter how well manipulated, the information derived from the raw data should be presented in an effective format, otherwise, it would be a great loss for both authors and readers.

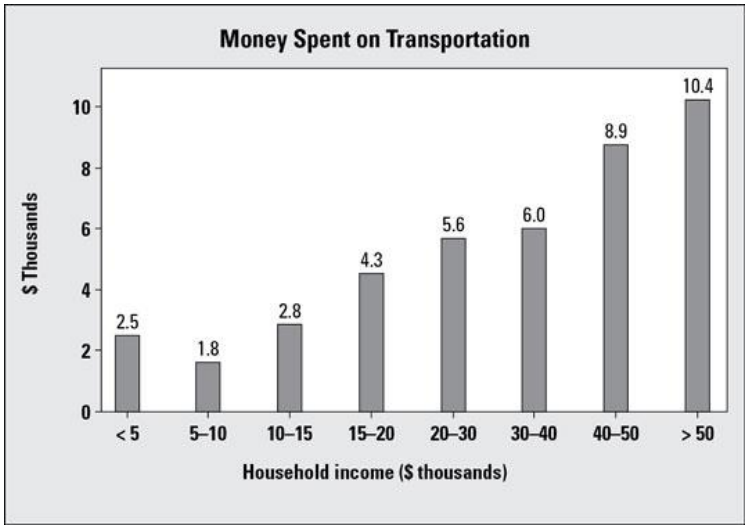
Data presentation

- **Textual:** Text is the principal method for explaining findings, outlining trends, and providing contextual information.
- **Tabular:** A table is best suited for representing individual information and represents both quantitative and qualitative information.
- **Graphical:** A graph is a very effective visual tool as it displays data at a glance, facilitates comparison, and can reveal trends and relationships within the data such as changes over time, frequency distribution, and correlation or relative share of a whole.

Textual Method	Tabular Method	Graphical Method
• Rearrangement from lowest to highest • Stem-and-leaf plot	• Frequency distribution table (FDT) • Relative FDT • Cumulative FDT • Contingency Table	• Bar Chart • Histogram • Frequency Polygon • Pie Chart • Less/greater than Ogive

Interpretation of Bar chart

The following bar graph shows how much money people in the United States spend on transportation to get back and forth to work:

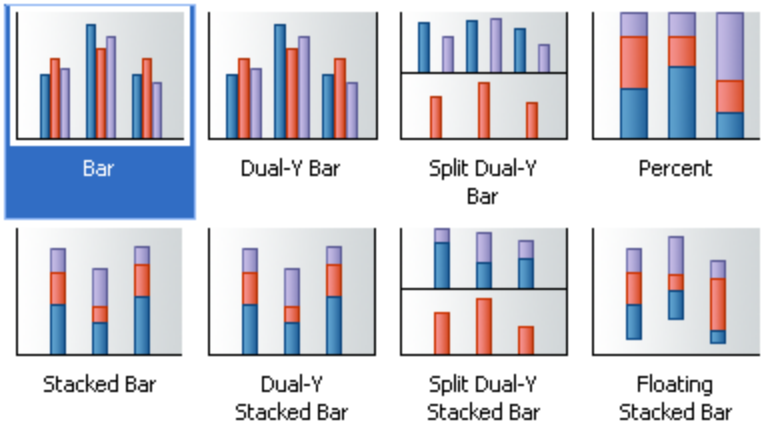


This particular bar graph shows how much money is spent on transportation for people in different household-income groups. It appears that as household income increases, the total expenditures on transportation also increase. This makes sense, because the more money people have, the more they have available to spend.

This $2,500 represents 50% of the annual income of those who make $5,000 per year; the percentage of the total income is even higher for those who make less than $5,000 per year. The households earning $30,000–$40,000 per year pay $6,000 per year on transportation, which is between 15% and 20% of their household income. So, although the people making more money spend more dollars on transportation, they don't spend more as a percentage of their total income. Depending on how you look at expenditures, the bar graph can tell two somewhat different stories. Another point to check out is the groupings on the graph.

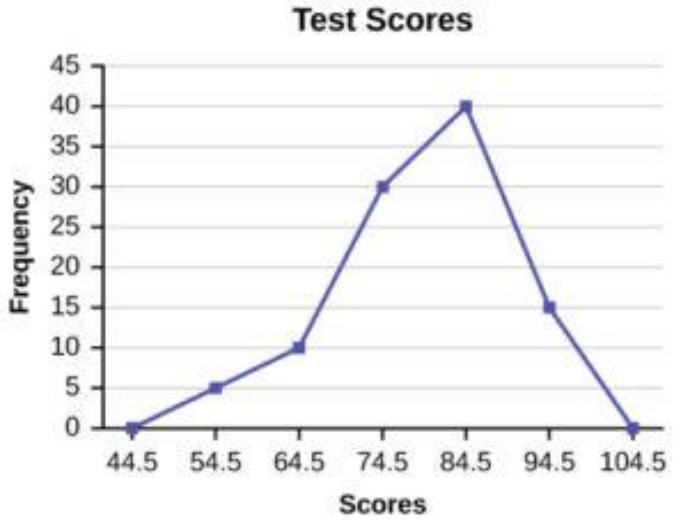
The categories for household income as shown aren't equivalent. For example, each of the first four bars represents household incomes in intervals of $5,000, but the next three groups increase by $10,000 each, and the last group contains every household making more than $50,000 per year. Bar graphs using different-sized intervals to represent numerical values (such as the above image) make true comparisons between groups more difficult. (However, the government probably has its reasons for reporting the numbers this way; for example, this may be the way income is broken down for tax-related purposes.)

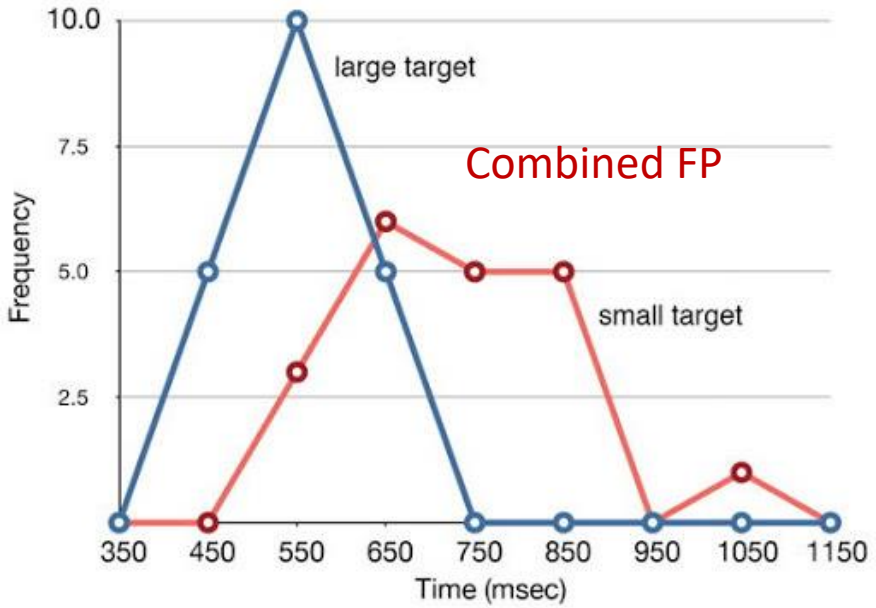
Different types of Bar chart



Interpretation of Frequency polygon

The frequencies corresponding to the class marks are plotted against each class mark as shown in the following graph. This makes sense as the frequency for class marks 44.5 and 104.5 are zero and touching the x-axis. These plot points are used only to give a closed shape to the polygon. Frequency polygons provide us with an understanding of the shape of the data and its trends.





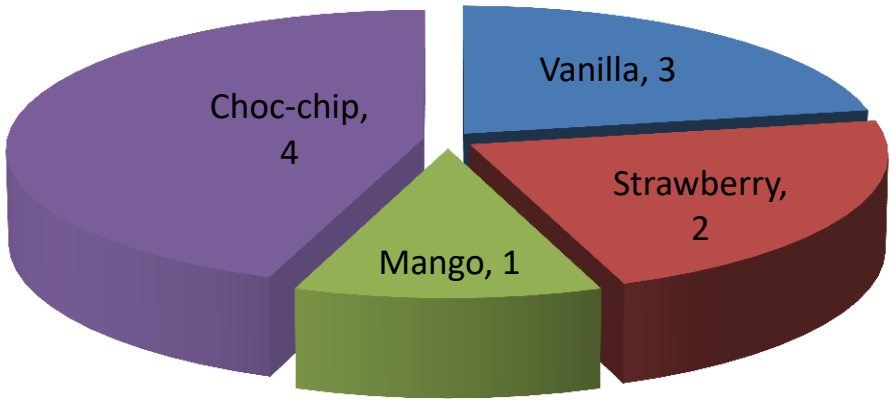
The frequency histogram has the similarity to a column graph without the presence of spaces between columns. The frequency polygon happens to be a special line graph whose use takes place in statistics. One can draw these graphs either separately or combined. The major difference between a **frequency polygon** and **frequency curve** is that the drawing of a frequency polygon by joining points by a straight line while the drawing of a frequency curve takes place by a smooth hand.

Interpretation of Pie Chart

A pie chart shows data as parts of a whole. The circle represents the total amount while the segments are the parts. When we compare the parts to the whole, we're looking at proportion. This is often written as a fraction.

This pie chart shows the favourite ice cream flavours of 10 people.

Favourite ice cream flavours of 10 people



The table below summarises the information displayed on this graph.

Category	Amount	Fraction
Vanilla	3	$\frac{3}{10}$
Strawberry	2	$\frac{2}{10}$
Mango	1	$\frac{1}{10}$
Choc-chip	4	$\frac{4}{10}$
Total	10	$\frac{10}{10}$

Usually it is used to present qualitative variables. Combined Pie Chart is not possible.

Interpretation of Ogive or Cumulative Frequency Polygon

The Ogive is defined as the frequency distribution graph of a series. The Ogive is a graph of a cumulative distribution, which explains data values on the horizontal plane axis and either the cumulative relative frequencies, the cumulative frequencies or cumulative percent frequencies on the vertical axis. Cumulative frequency is defined as the sum of all the previous frequencies up to the current point. To find the popularity of the given data or the likelihood of the data that fall within the certain frequency range, Ogive curve helps in finding those details accurately. Create the Ogive by plotting the point corresponding to the cumulative frequency of each class interval. Most of the Statisticians use Ogive curve, to illustrate the data in the pictorial representation. It helps in estimating the number of observations which are less than or equal to the particular value.

Ogive Graph

The graphs of the frequency distribution are frequency graphs that are used to exhibit the characteristics of discrete and continuous data. Such figures are more appealing to the eye than the tabulated data. It helps us to facilitate the comparative study of two or more frequency distributions. We can relate the shape and pattern of the two frequency distributions. The two methods of Ogives are

- **Less than Ogive:** The frequencies of all preceding classes are added to the frequency of a class.
- **Greater than or more than Ogive:** The frequencies of the succeeding classes are added to the frequency of a class

Marks	Frequency	More than Cumulative Frequency
More than 1	3	60
More than 11	8	57
More than 21	12	49
More than 31	14	37
More than 41	10	23
More than 51	6	13
More than 61	5	7
More than 71	2	2

Marks	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
Frequency	3	8	12	14	10	6	5	2