

# ***STATISTICS AND ANALYTICS***

## ***Unit-2, Session-1***

# ***SUMMARIZATION OF DATA***



**Presented By:**

**Umesha D K**

Lecturer, Dept of CSE

Shivasharane Akkamahadevi Govt Women's  
Polytechnic, Shiralakoppa

# CONTENTS

- Basic Concepts
- Data Tabulation
- Chats

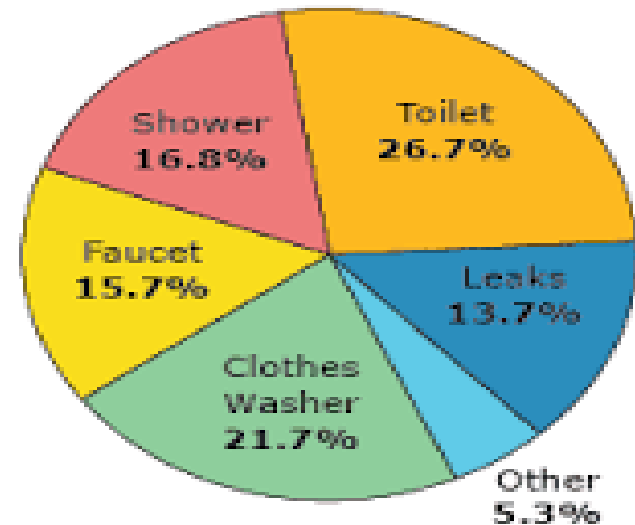


# BASIC CONCEPTS

- **Descriptive statistics** : Descriptive statistics is a way to organise, represent and describe a collection of data using tables, graphs, and summary measures.

Marks	Number of Students		Total
	Males	Females	
30 – 40	8	6	14
40 – 50	16	10	26
50 – 60	14	16	30
60 – 70	12	8	20
70 – 80	6	4	10
Total	56	44	100

How Much Water Do We Use?



Source: American Water Works Association Research Foundation, "Residential End Uses of Water." 1999

# DATA TABULATION

- ❖ Tabulation is a process of systematic arrangement of the classified data in rows and columns, in the form of table.

## Example 1:

Number of oranges in the box	5	6	7	8	9	10	Total
Number of boxes	5	8	10	6	3	13	45

## Example 2:

Height (cm)	140-150	150-160	160-170	170-180	Total
No. of students	6	24	18	2	50

The above two types example are **Frequency Distribution or Frequency table.**

# DATA TABULATION

- **Frequency:** The frequency of a particular data value is the number of times the data value occurs.
- **Frequency Distribution :** A frequency distribution table is a chart that summarizes values and their frequency.
- **Class Frequency:** The number of observations corresponding to a particular class is known as class frequency.
- **Relative Frequency :** Relative frequency is the ratio of frequency of the value of the variable to the total frequency.

# DATA TABULATION

**Ex:**

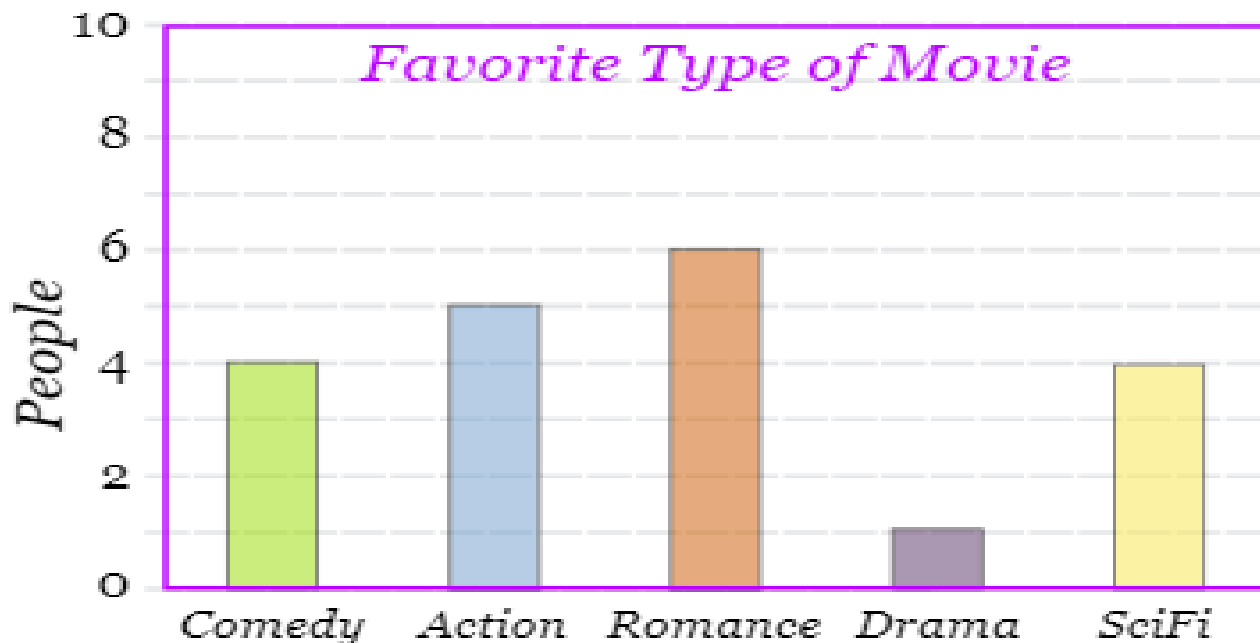
No of apples Per box	No of boxes	Relative Frequency $R.f=f/N$
5	5	$5/45 = 0.111$
6	8	$8/45 = 0.178$
7	13	$13/45 = 0.289$
8	10	$10/45 = 0.222$
9	6	$6/45 = 0.133$
10	3	$3/45 = 0.067$
Total (N)	45	1

# CHATS

- **Ungrouped data** is the data you first gather from an experiment or study. The data is raw — that is, it's not sorted into categories, classified.
- **Grouped data** is data that has been bundled together in categories.
- **Charts** convey information about our data faster than tables.
- Ex: BAR GRAPH , PIE CHART , LINE GRAPH, etc...

# BAR GRAPH

- A bar graph (also known as a bar chart or bar diagram) is a visual tool that uses bars to compare data among categories.
- A bar graph may run horizontally or vertically.

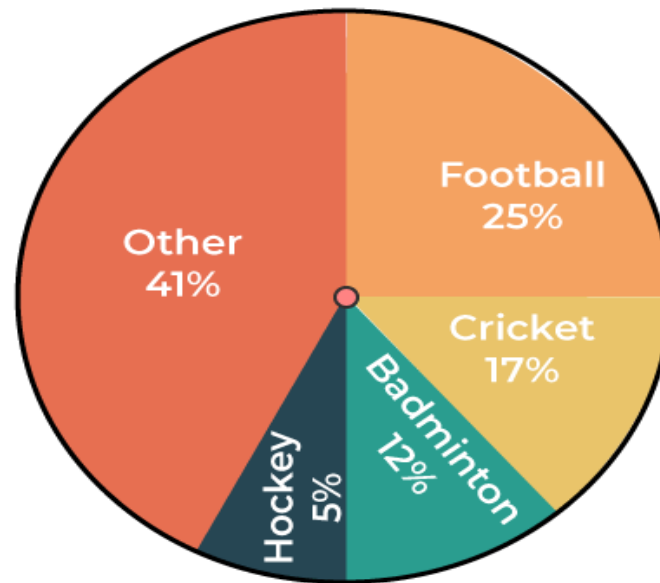




# PIE CHART

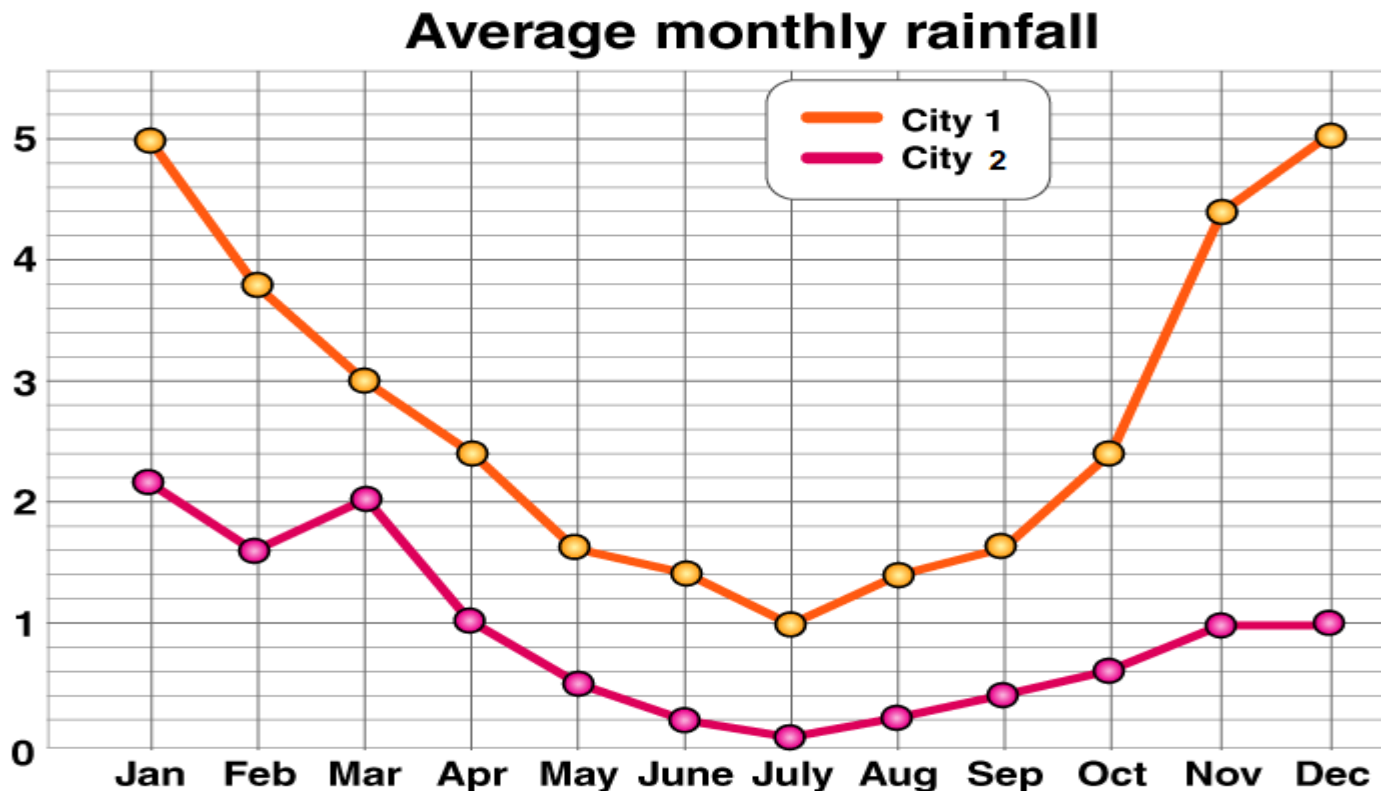
- ❖ The “pie chart” also is known as “circle chart”, that divides the circular statistical graphic into sectors or slices in order to illustrate the numerical problems.

**Number of Students**



# LINE GRAPH:

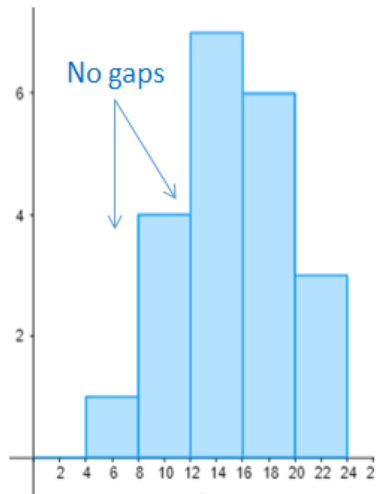
- ❖ Line Graph: A line graph is a type of chart used to show information that changes over time. We plot line graphs using several points connected by straight lines.



# HISTOGRAM?

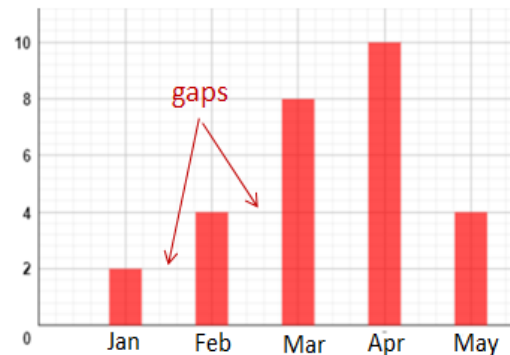
It can be defined as a set of rectangles with bases along with the intervals between class boundaries and with areas proportional to frequencies in the corresponding classes.

## Histogram vs. Bar Chart



← Number Ranges →

**Histogram**

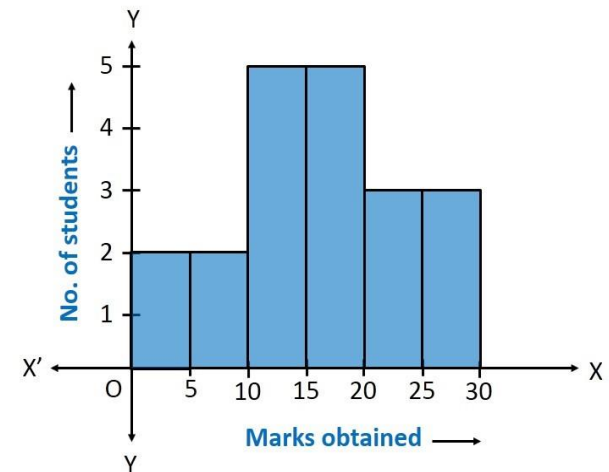


← Categories →

**Bar Chart**

Intervals	Frequency
0 - 5	2
5 - 10	2
10 - 15	5
15 - 20	5
20 - 25	3
25 - 30	3

teachoo.com

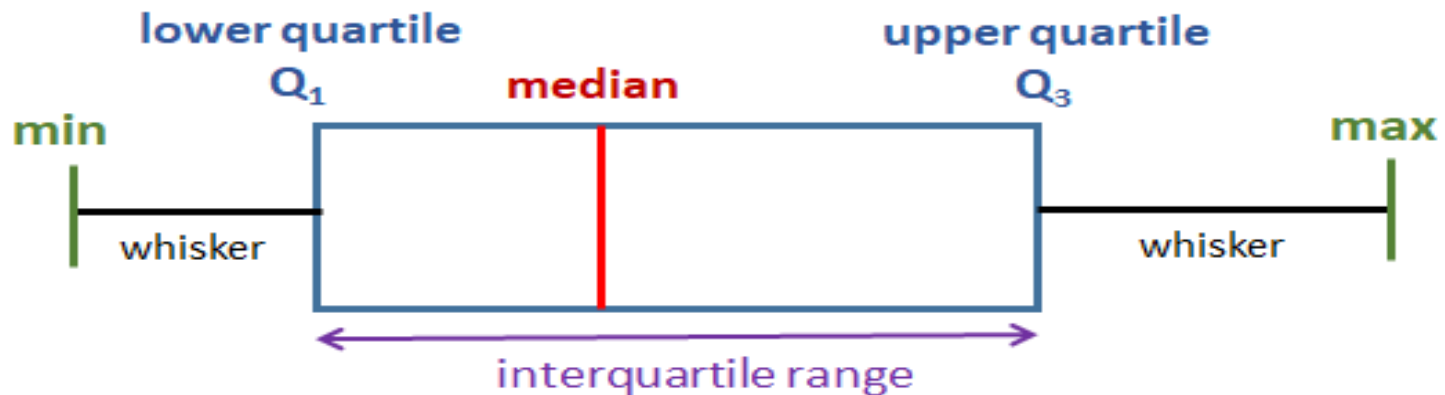


# BOXPLOT:

- ❖ In descriptive statistics, a boxplot is a method for graphically depicting groups of numerical data through their quartiles.

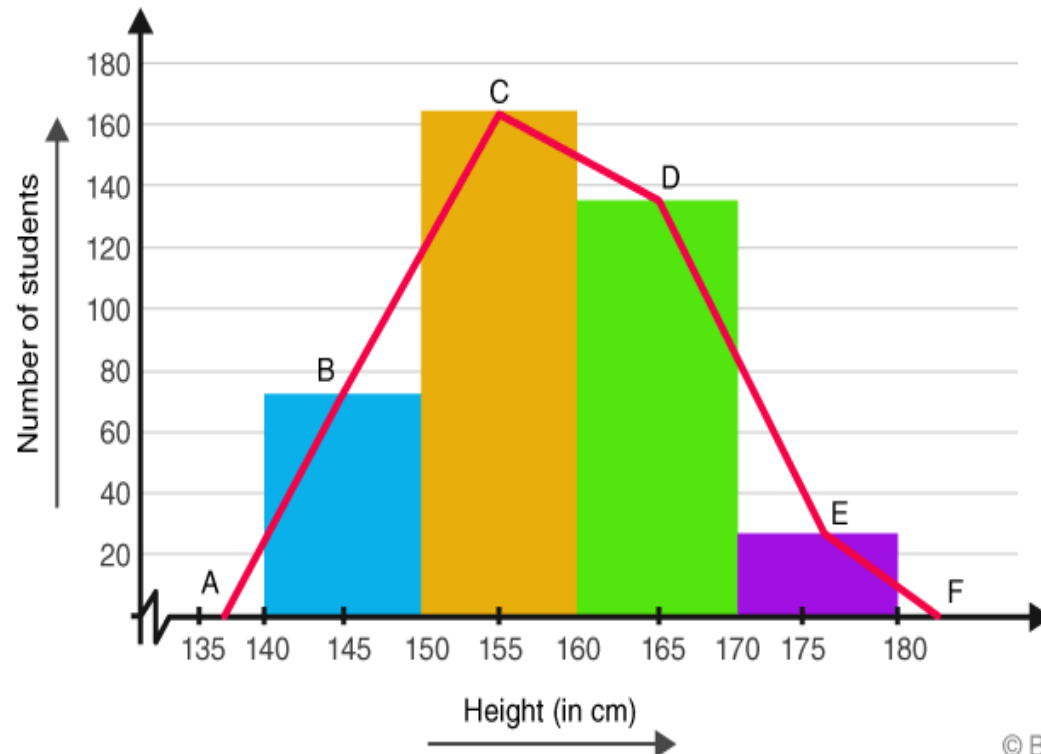
## Box and Whisker Plot

A box and whisker plot (also called a box plot) shows the five-number summary of a set of data: **minimum**, **lower quartile**, **median**, **upper quartile**, and **maximum**.



# FREQUENCY POLYGON

A frequency polygon is almost identical to a histogram, which is used to compare sets of data or to display a cumulative frequency distribution. It uses a line graph to represent quantitative data.



***THANK YOU***

*Unit-2, Session-1*  
*Assessment-MCQ*

# Assessment-MCQ

## 1) What is data tabulation?

- a) **Organizing data into rows and columns**
- b) Summarizing data using charts and graphs
- c) Calculating the average of a dataset
- d) Identifying the outliers in a dataset

## 2) What does frequency refer to in statistics?

- a) **The number of times an event occurs in a dataset**
- b) The range of values in a dataset
- c) The average of a dataset
- d) The spread of data in a dataset



# Assessment-MCQ

## 3) What is a frequency class?

- a) **A group of values in a dataset with similar characteristics**
- b) The highest occurring value in a dataset
- c) The midpoint of a frequency distribution
- d) The total number of values in a dataset

## 4) What does the mean represent in statistics?

- a) The most frequently occurring value in a dataset
- b) The middle value in a dataset
- c) **The average value of a dataset**
- d) The spread of data in a dataset

# Assessment-MCQ

5) Which of the following summarizes categorical data by displaying the number of occurrences for each category?

- a) Histogram    b) Scatter plot    **c) Frequency table**    d) Box plot

6 ) In a frequency table, the total of all frequencies must be equal to

- a) The number of categories    b) The range of the data  
c) The mean of the data    **d) The total number of observations**

7) The relative frequency for a category in a relative frequency table is calculated by dividing the category frequency by:

- a) The total number of categories    b) The maximum frequency  
**c) The total number of observations**    d) The mode of the data

# Assessment-MCQ

8) Which type of graph is best suited to display the distribution of a single continuous variable?

- a) Bar graph      b) Pie chart      c) Line graph      **d) Box plot**

9) Which type of graph is commonly used to represent categorical data?

- a) Bar graph**      b) Pie chart      c) Line graph      d) Box plot

10) Which type of graph is commonly used to represent trends over time?

- A      a) Bar graph      b) Pie chart      **c) Line graph**      d) Box plot

# *Assessment-MCQ*

**11. Which type of graph is most suitable for displaying the distribution of a single categorical variable?**

- a) **Bar graph**      b) Pie chart      c) Line graph      d) Box plot

**12. Which type of graph is used to show the median, quartiles, and outliers of a dataset?**

- a) Bar graph      b) Pie chart      c) Line graph      d) **Box plot**

**13. Which type of graph is used to display the composition of a whole?**

- a) Bar graph      b) **Pie chart**      c) Line graph      d) Box plot

***THANK YOU***