UNIT-2 SUMMARIZATION of DATA

① Descriptive statistics :-

Datatabulation (frequency table)

- Descriptive statistics :- It refers to a set of methods used to summarise and describe the main features of data set, such as its central tendency, variability and distribution.

- frequency table :- A frequency table is a way to present data. The data are counted and ordered to summarise bigger sets of data.
with a frequency table we can analyze the way in which the data is distributed across different values.

frequency :- means the no. of times a value appear in data. A table can quickly show us how many times each value appear.

examples :-

ungrouped

Marks obtained	frequency
25	6
26	9
27	10
28	2
29	3
30	4

* used for small data set

Grouped.

class-interval	frequency
0-5	6
5-10	12
10-15	02
15-20	01
20-25	11
25-30	06

* used for large data set.

- Relative frequency table :-

→ The word "relative" is used to indicate that an event is being considered in relation or in proportion to something else.

→ Relative frequency is a way to measure how often a particular event occurs against total occurences

formula :-

$$\text{Relative frequency} = \frac{f}{N}$$

where 'f' is the frequency of a specific group and 'N' is the total frequency.

Example :- This is a frequency table to see how many students have got marks betⁿ given interval in maths.

Groped data

Marks	frequency	Relative frequency
45-50	3	3/40 = 0.075
50-55	1	1/40 = 0.025
55-60	1	1/40 = 0.025
60-65	6	6/40 = 0.15
65-70	8	8/40 = 0.2
70-75	3	3/40 = 0.075
75-80	11	11/40 = 0.275
80-85	7	7/40 = 0.175
	Total = 40	

practical No. - 2

prepare a Google form for a specified problem statement to collect the data set. (for example questionnaires to conduct online quiz)

To set up a new form or quiz.

Step 1 :- Go to forms.google.com

Step 2 :- Click on 'Blank +'

Step 3 :- Name your ~~untitled~~ ~~united~~ form

Step 4 :- Add questions and suitable answer types.

various answer types include

* Short answer (≡)
* Paragraph (≡)
* Multiple-choice (◉)
* Check boxes (☑)
* Dropdown (⌄)

UNTITLED FORM

B I U (-) T̸

FORM DESCRIPTION

B I U (-) 1. 2. 3. T̸

UNTITLED QUESTIONS

B I U (-) T̸

☐ ◉ Multiple choice ▼

- option 1
- Add option or add "other"

Required

for example :- To conduct an online quiz

step 1 :- Go to forms.google.com

step 2 :- click on 'Blank +'

step 3 :- Name the untitled form as "online quiz".

step 4 :- Add questions No.1 on the untitled question box

ex:- i) who is the PM of India?

step 5 :- Select appropriate answer type according to the question.

for ex:- Multiple choice.

step 6 :- Add options from option tag

- option 1
  ex:- Amit Shah
- option 2
  ex:- Rahul Gandhi
- option 3
  ex:- Narendra Modi

we can add untitled options "other" answer types

step 7:- Ensure the correct an option as answer.

click on the ☑ option and then select the right option

you can again marks for each correct answer by assigning points.

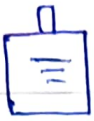
As you click ☑. the dialoge box appears:-

☑ choose the correct answer.

who is the PM of India?

..2.. points ⟵ This option 3 helps to provide marks for each ooo correct answer.

- Amit Shah
- Rahul Gandhi
- Narendra Modi ⟶ select the right option



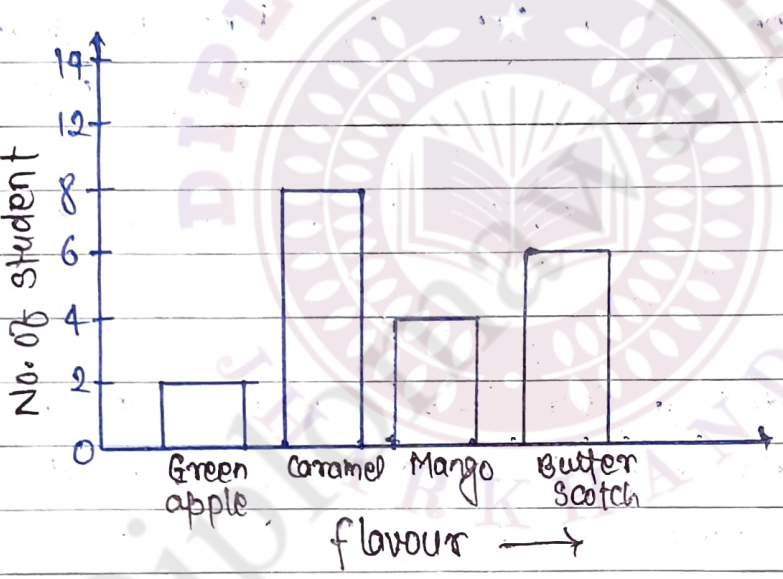
Add answer feedback ← this helps to show feedback for correct or incorrect options.

* we can add such more questions and answer (options) by clicking on ⊕ options.

(b) Grouped Data :-

* Bar graph :-

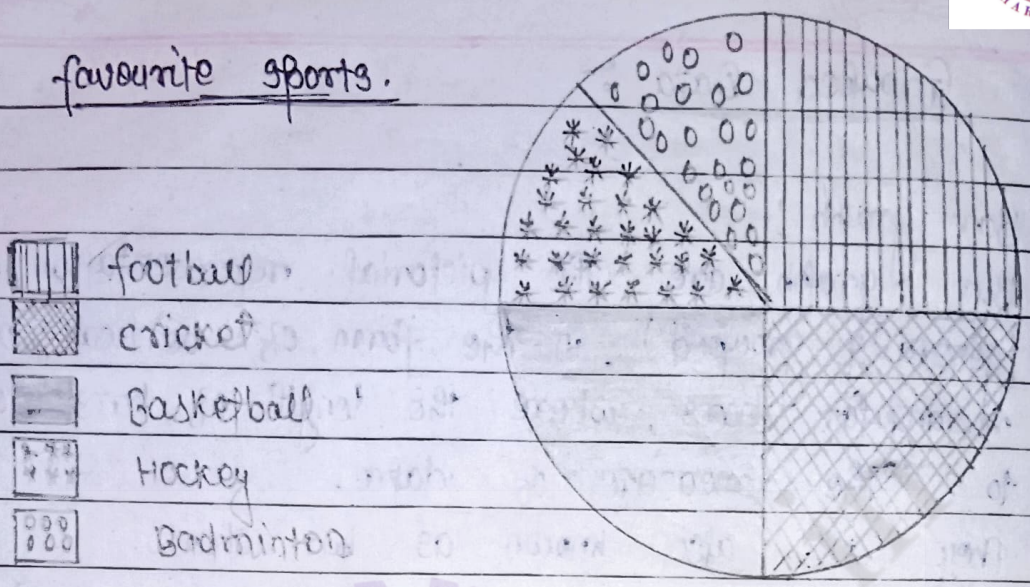
- Bar graph are the pictorial representation of data (generally grouped), in the form of vertical or horizontal rectangular bars, where the length of bars are proportional to the measure of data.
- They are also known as bar charts.
- Bar graphs are one of the means of data handling in statistics.



* Pie chart :-

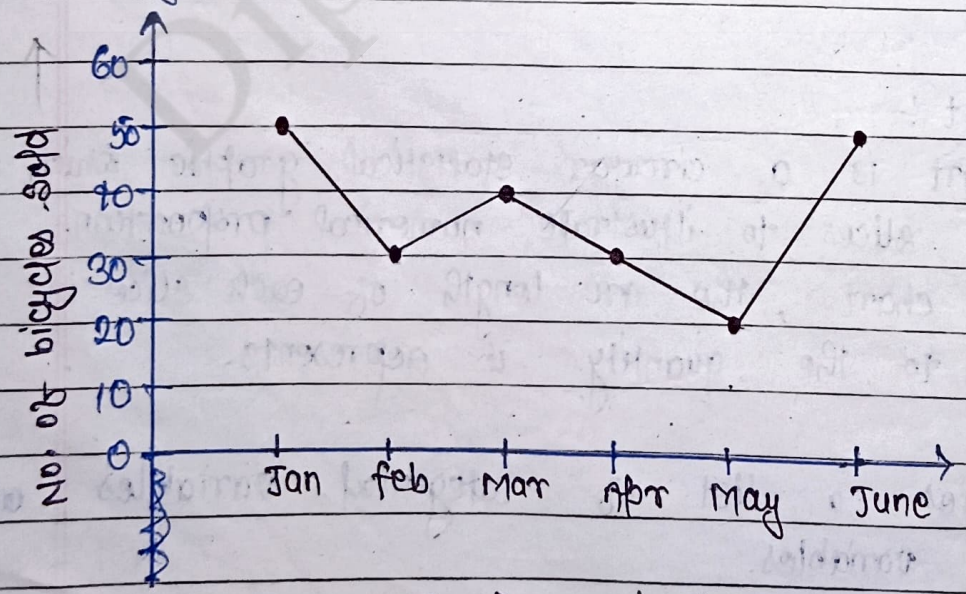
- A pie chart is a circular statistical graphic which is divided into slices to illustrate numerical proportion.
- In a pie chart, the arc length of each slice is proportional to the quantity it represents.
- It requires a list of categorial variables and numerical variables.

e.g:- favourite sports.



* Line graph :-

- A line graph is a type of chart or graph that is used to show information that changes over time.
- A line graph can be plotted using several points connected by straight lines.
- It is used to display change over time as a series of data points connected by straight line segments on two axes.

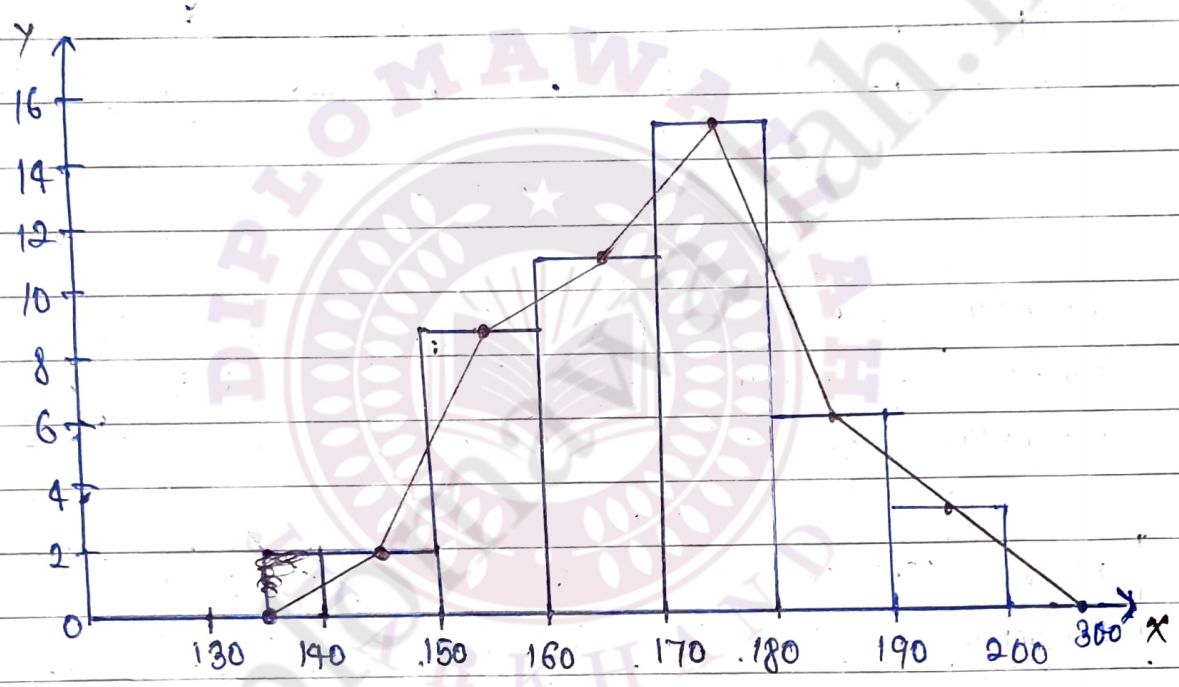


* frequency polygon :-

- A frequency polygon is a line graph of class frequency plotted against class mid-point.
- The accurate points in a frequency polygon graph represent the data of the particular class interval.

formula :-

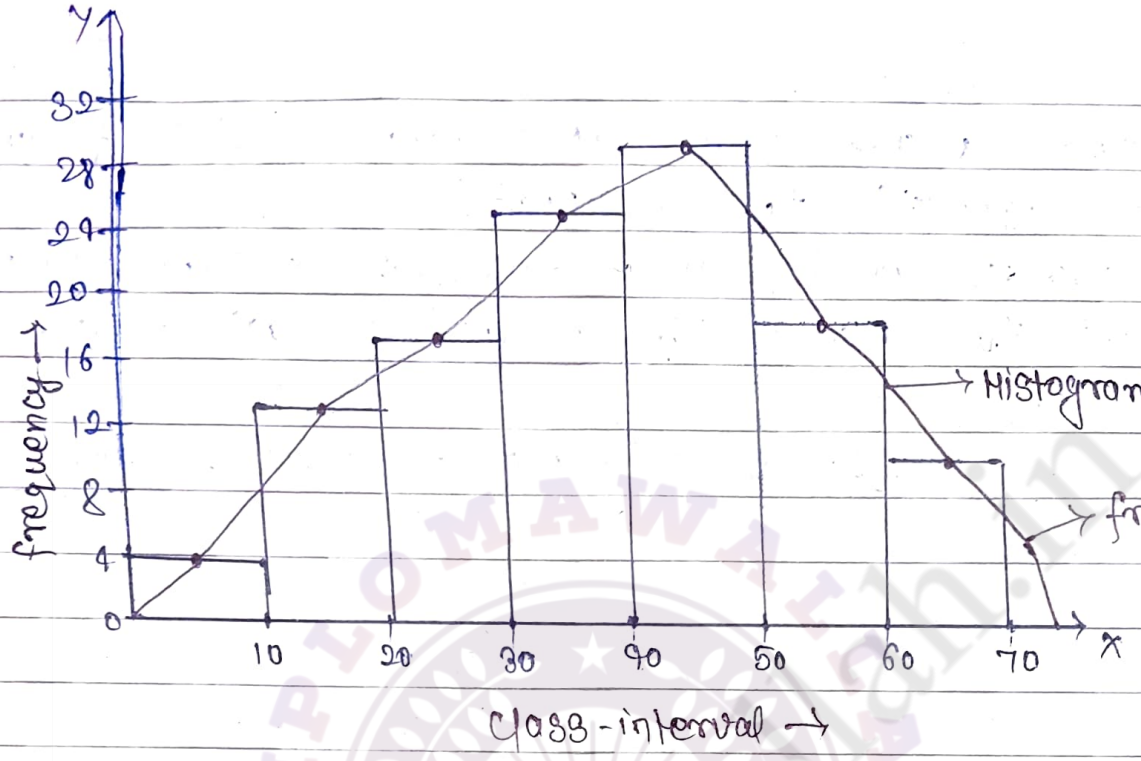
classmark (mid-point) = $\frac{\text{upper limit} + \text{lower limit}}{2}$



* frequency curve :-

- A frequency curve is a limiting form of a histogram or a frequency polygon.
- A frequency curve for a given distribution can be obtained by drawing a smooth, free hand curve through the mid-points of the upper sides of the rectangles forming the histogram.

Recieved

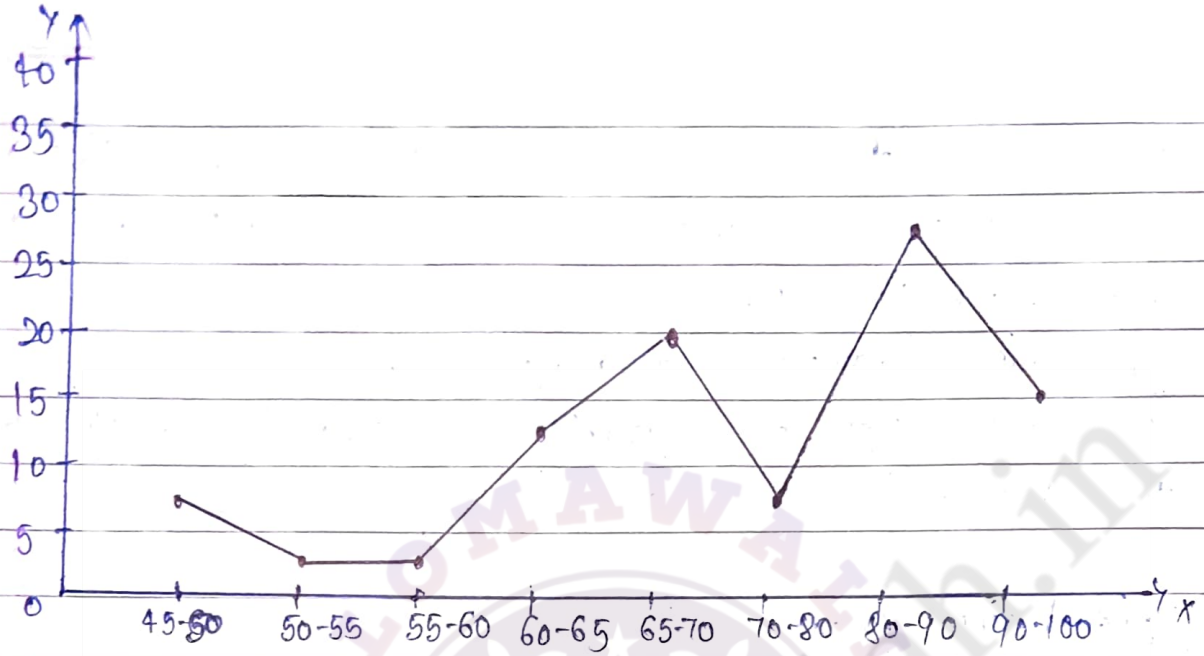


* Relative frequency polygon :-
- Relative frequency polygon has peaks that represent the percentage of total data point falling within the interval.

eg:-

Marks	45-50	50-55	55-60	60-65	65-70	70-80	80-90	90-100
frequency	3	1	1	6	8	3	11	7

Marks	frequency	Relative frequency
45-50	3	$3/40 \times 100 = 7.5$
50-55	1	$1/40 \times 100 = 2.5$
55-60	1	$1/40 \times 100 = 2.5$
60-65	6	$6/40 \times 100 = 15$
65-70	8	$8/40 \times 100 = 20$
70-80	3	$3/40 \times 100 = 7.5$
80-90	11	$11/40 \times 100 = 27.5$
90-100	7	$7/40 \times 100 = 17.5$

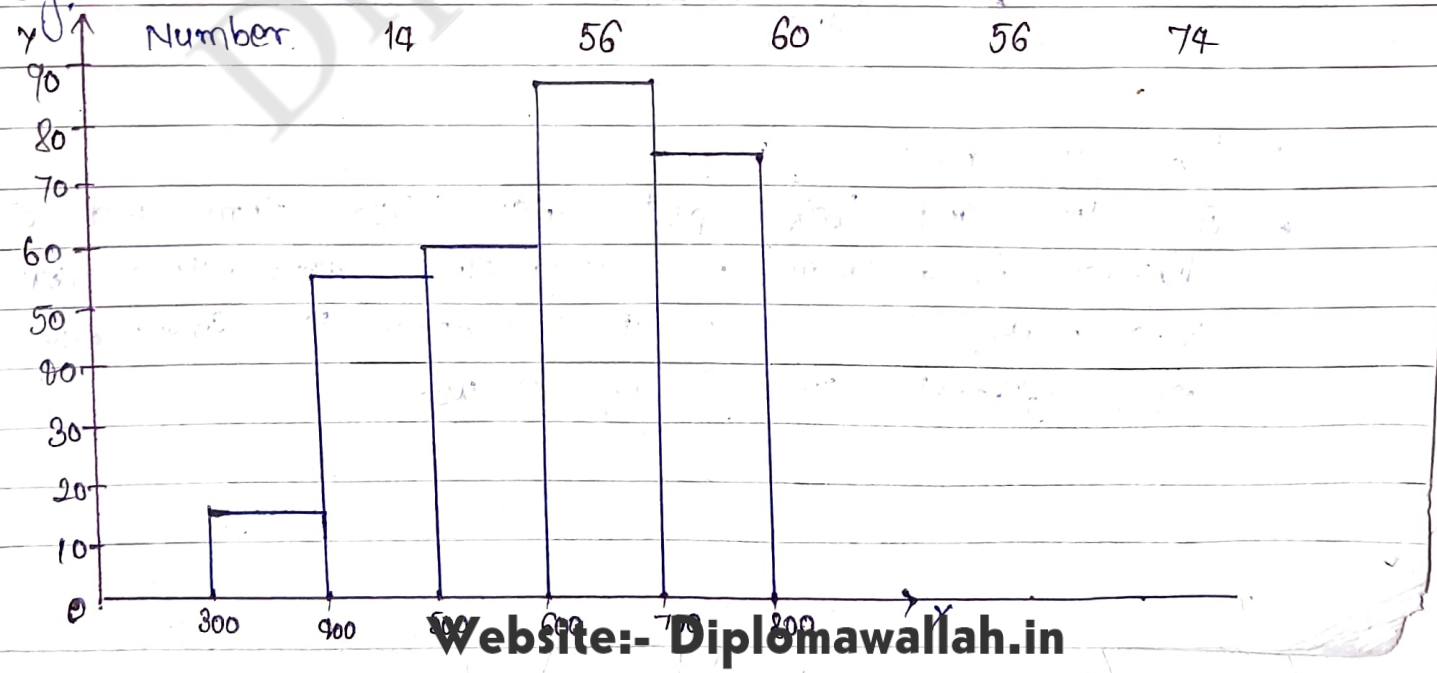


* **Histogram** :-

- A histogram is a graphical representation of the distribution of data. The histogram is represented by a set of rectangles, adjacent to each other. It is an area diagram and can be defined as a set of rectangle with bases along with the interval - betⁿ class boundaries and with area proportional to frequencies in the corresponding classes.

e.g:-

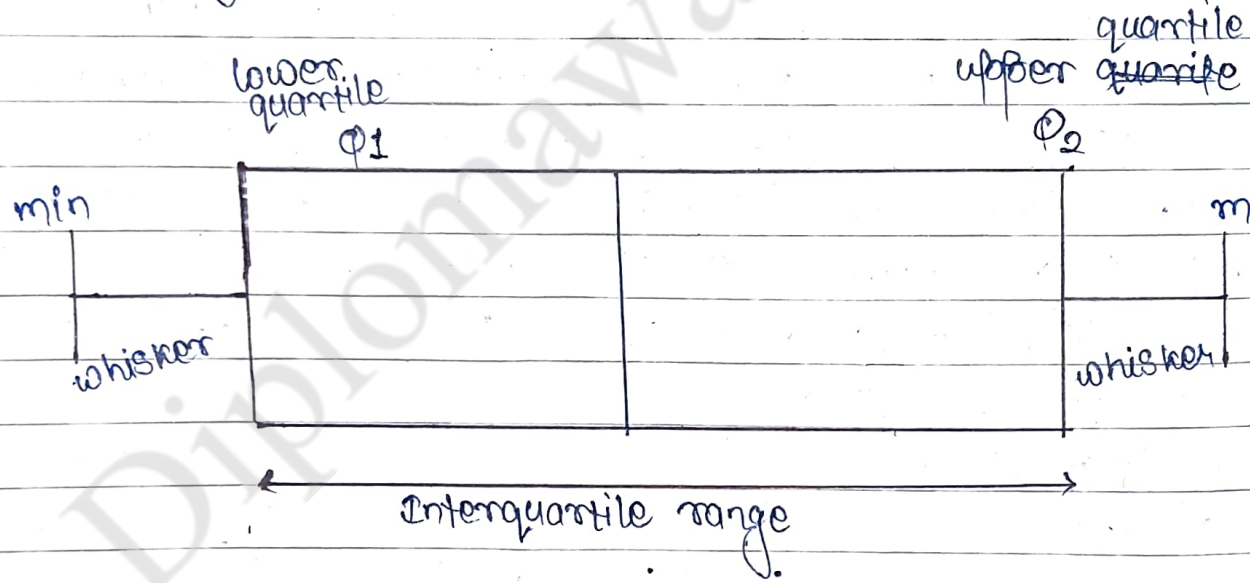
Life-time	300-400	400-500	500-600	600-700	700-800
Number	14	56	60	56	74



* Box - plot :-

- The method to summarize a set of data that is measured using an interval state is called a box and whisker plot.
- These are maximum used for data analysis, we use these type of graphs or graphical representation to know.

1) Distribution shape
2) central value of it
3) variability of it.



* Leaf - stem plot :-

- A stem and leaf plot also called stem and leaf diagram is a way of organising data into a form that make it easy to observe the frequency of different types of value.

A stem and leaf plot is represented in form of a special table where each first digit or digit of data value is split into a stem and the last digit of data in a leaf.

Stem	Leaf
2	0012 57
3	1 4 8
4	5 2
5	1

key : 2|0 = 20