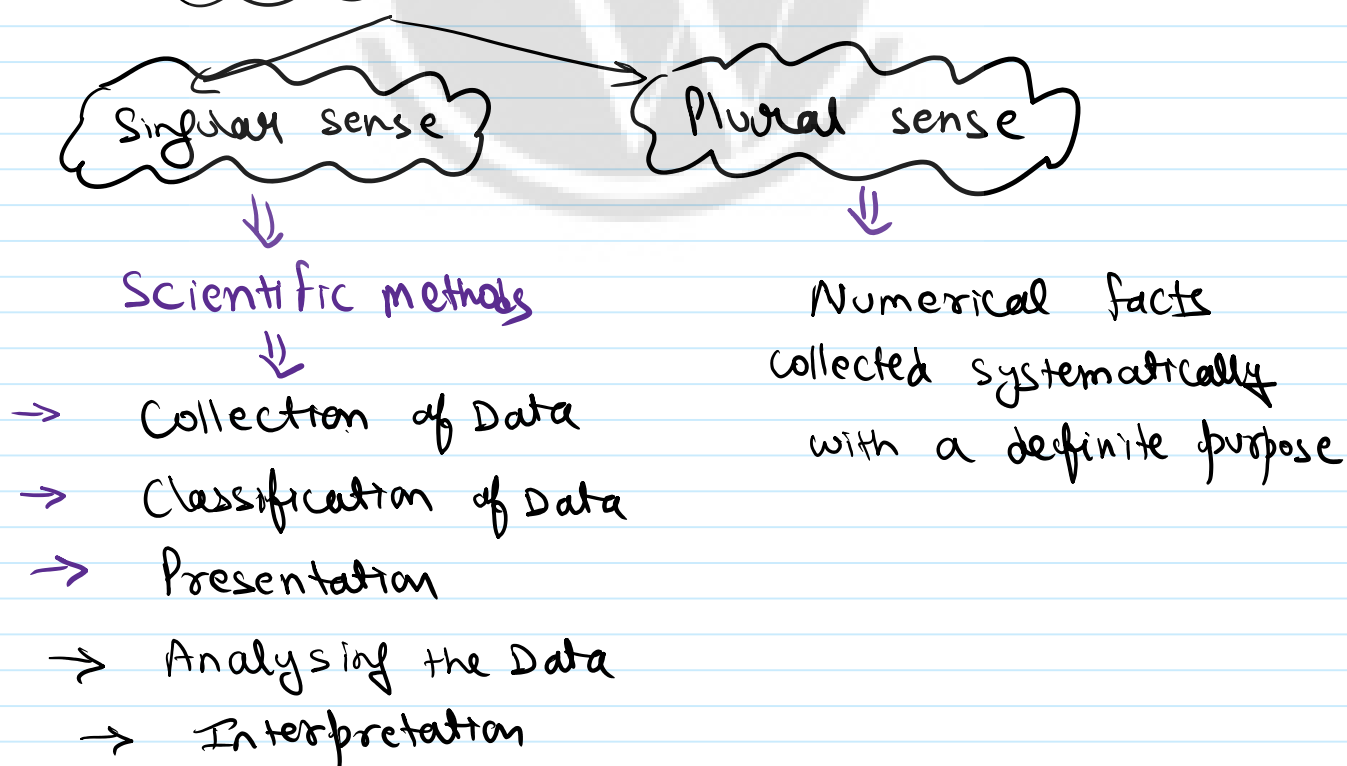


Statistical Description of Data

Statistics is originated from

Latin word	-	"Status"
Italian word	-	"Statista"
German word	-	"Statistik"
French word	-	"Statistique"

Definition of statistics



Application of statistics

Applications of Statistics

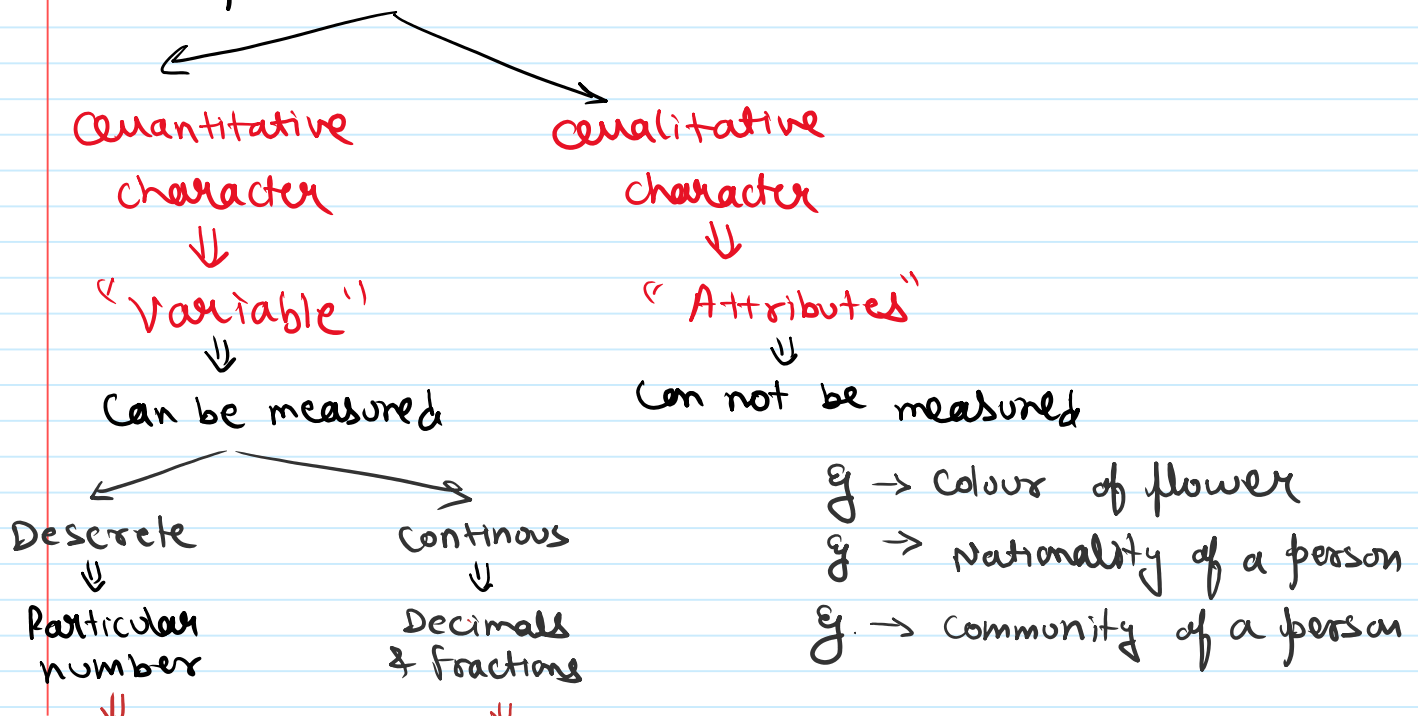
- Economics
- Business management
- Commerce & Industry

Limitation of Stats

- Deals in aggregates only
- Does not deal in qualitative data
- Not always true
- Expert consultation required
- Stats can be misused

Data

Information about some particular character



Particular number



g No. of Accidents

g Goals in a match

g Salary

Decimals & Fractions



g weight

g Height

g Age

g Temperature.

g → community of a person

Collection of Data

If Data is Primary

" This is collected for first time "

↓
Methods

#1 Interview method

Direct (Personal interview)

↓
Natural calamity

g Cyclone

g Earthquake

Indirect

↓
Train accident

Telephonic (fastest)

#1] International sources

g WHO, IMF
World Bank

#2] Govt sources

g surveys from different ministries

#3] Private sources

#4] Unpublished sources

#2] Questionnaire method

mailed
questionnaire



- wide Area
- maximum non responses

Enumeration
method



- for uneducated informants

#3] Observation Method

used for research
to gather data about
people, objects, events
& behaviours etc.

- Accurate data
- Time consuming

{Scrutiny of Data}



Verification

Accuracy
of data

consistency
of data

{Classification of Data}

1) Chronological or Temporal - on the Basis of time

g

Year	2010	2011	2012
Sales	100	120	150

g

Year	2010	2011	2012
Sales	100	120	150

2> **Geographical or Spatial** — on the basis of Location (Place)

g

States	UP	Punjab	Bihar	Goa
Rice Production	100	80	120	60

3> **Qualitative (or ordinal)**

g

Caste	Rajput	Brahman	Siga	Sunni
Population	1000	1200	600	400

4> **Quantitative or Cardinal**

g

Marks	1	2	3	4	5
No. of Students	6	8	4	10	20

Presentation of Data

Textual

↓
Paragraph

↓
worst method

Tabular Presentation

↓
Table
ii)

Diagrammatic Presentation

↓
Line Diagram

↓
Bar Diagram

↓
Pie Chart

↓
worst method

Table
↓
→ Best Method
→ most accurate

Line Diagram Bar Diagram Pie chart
↓
most attractive
(Can notice hidden trend in data)

Textual method

In 2009, out of a total of five thousand workers of Roy Enamel Factory, four thousand and two hundred were members of a Trade Union. The number of female workers was twenty per cent of the total workers out of which thirty per cent were members of the Trade Union. In 2010, the number of workers belonging to the trade union was increased by twenty per cent as compared to 2009 of which four thousand and two hundred were male. The number of workers not belonging to trade union was nine hundred and fifty of which four hundred and fifty were females.

Table

- 1> Table Number
- 2> Title (which explains content of tables)
- 3> Headnote (given in brackets information about the units)
- 4> Caption - Describes column & subcolumns
- 5> Stubs - Row Head
- 6> Body of Table - Numeric information
- 7> Source of Data
- 8> Footnote - specific feature which is not explained

Box Head
↓
upper part
of Table
includes column
& subcolumn number

Table No 25

number of Teachers

Table No 25

number of Teachers

Subjects	Class 11 th		Class 12 th		Total
	male	Female	male	Female	
Maths	3	1	2	2	8
A/c	2	2	3	2	9
Eco.	1	1	2	1	5
English	1	2	2	4	9
Total	7	6	9	9	31

Source :

Footnote : new joiners are not recorded here.

Diagrammatic method

⇒ 1) Line Diagram → can be used for time series

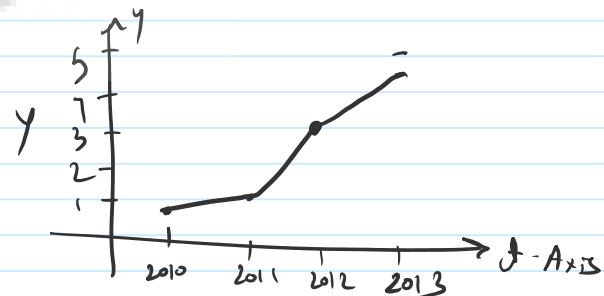
↓

d-y Plane

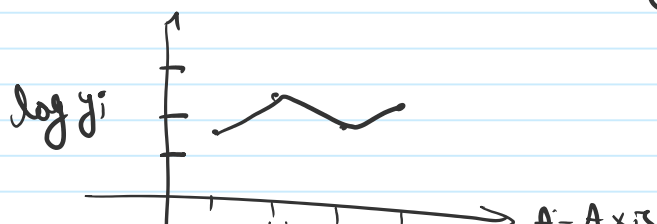
" There is no x-Axis,
there is t-Axis
& y-Axis "

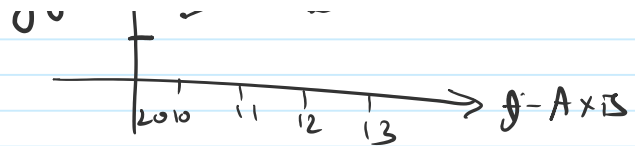
↓

one dimensional



Ratio chart : when time series have fluctuations we use ratio chart (use $\log y_i$)

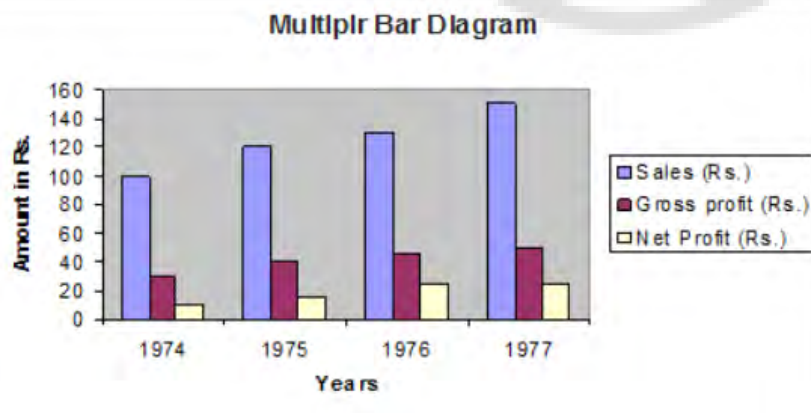
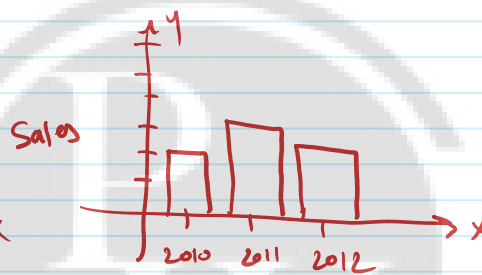
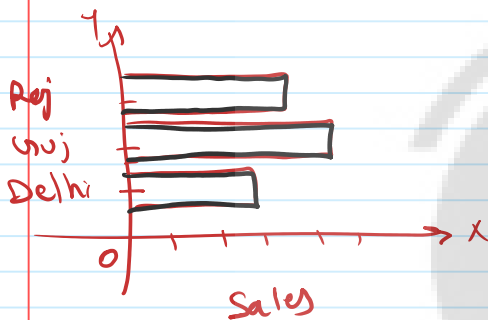


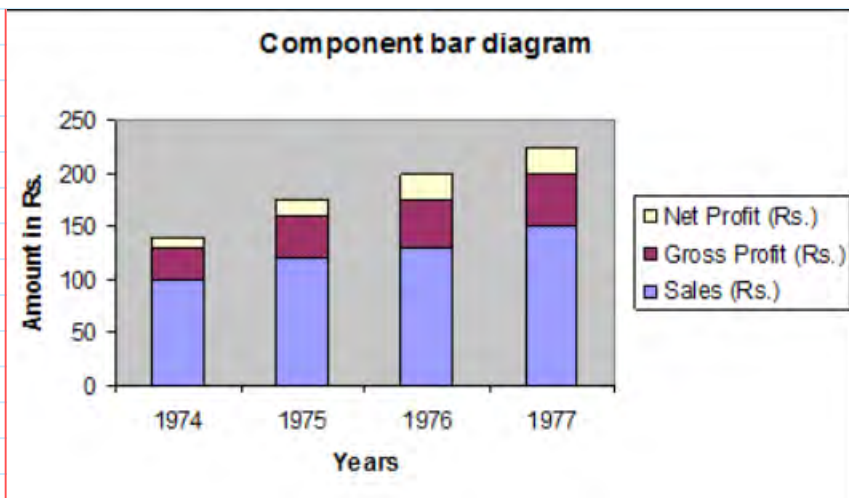


⇒ 2] **Bar Diagram** → Rectangle with Equal width
→ one Dimensional

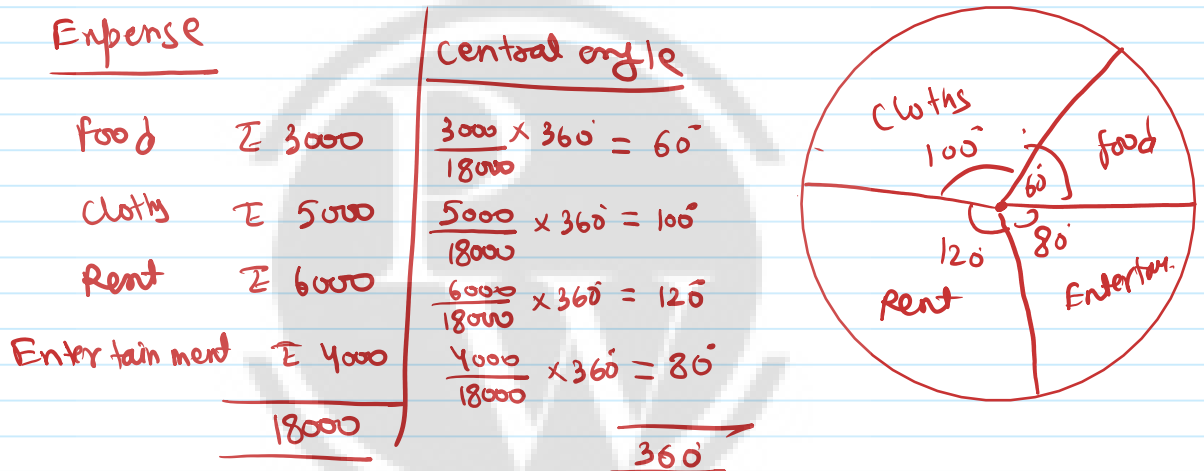
Horizontal
↓
Qualitative
&
vary over space

vertical
↓
Quantitative
&
vary over time





⇒ 3] Pie chart



#

Frequency Distribution



Tabular presentation
of statistical data

No. of Accidents in Last 15 Days

1, 2, 1, 1, 1, 2, 3, 3, 2, 1, 1, 1, 2, 3, 2

No. of Accidents	Tally Bar	Frequency
1		7

No. of accidents	Carry over	Frequency
1	III I	7
2	III	5
3	III	3
		$N=15$

Type of series

Individual

g marks: 1, 2, 3, 1,
4, 3, 1, 5, 2, 1, 3
1, 1, 2, 1, 5, 4, 3, 2

Discrete

g marks	fi
1	7
2	4
3	4
4	2
5	2
$N=19$	

Continuous series

Inclusive

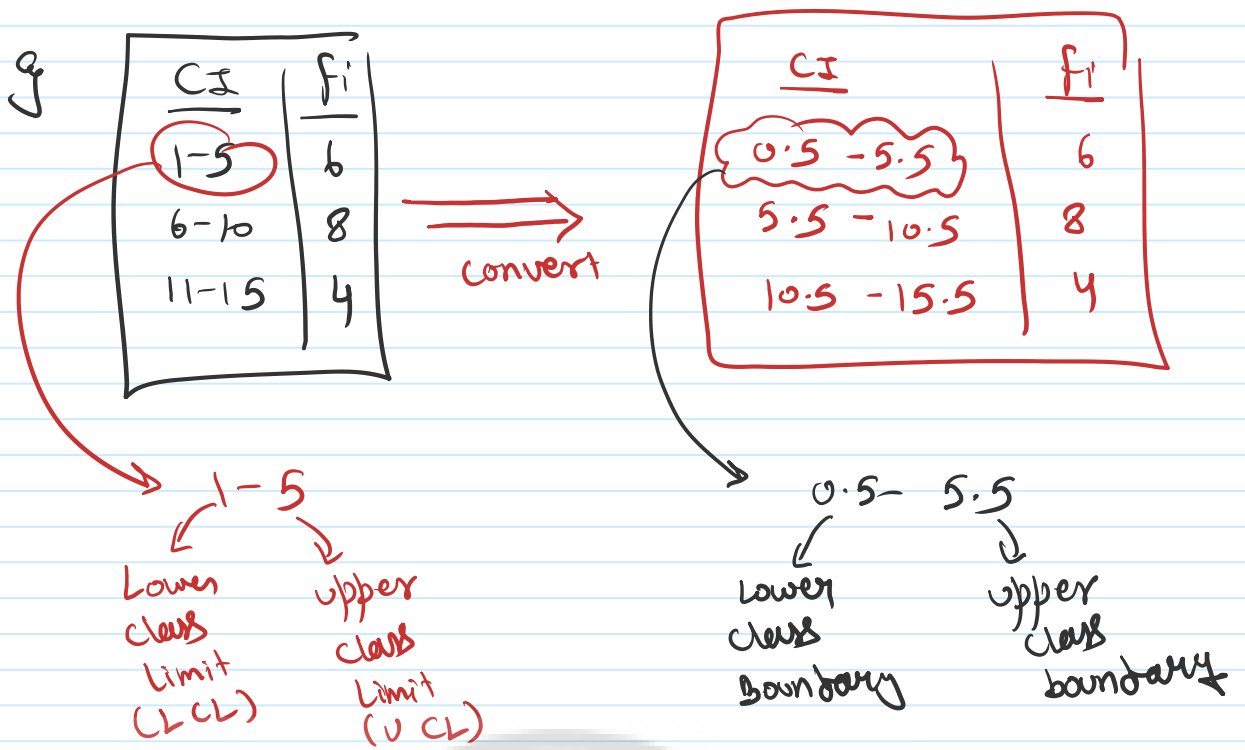
CI	fi
1-5	2
6-10	3
11-15	6
16-20	4

Exclusive

CI	fi
0-5	6
5-10	4
10-15	2
15-20	8
20	

Open ended

CI	fi
less than 10	2
10-12	3
12-14	6
14-16	3
more than 16	2
16	



Mid point of LCL & UCL is known as class mark

Frequency density of a class interval = $\frac{\text{frequency of class interval}}{\text{class length}}$

Relative frequency = $\frac{\text{class frequency}}{\text{Total frequency}}$

CI	fi	freq. Density	Relative frequency	% Frequency
0-5	2	$\frac{2}{5} = 0.4$	$\frac{2}{20} = 0.1$	10%
5-10	12	$\frac{12}{5} = 2.4$	$\frac{12}{20} = 0.6$	60%
10-15	6	$\frac{6}{5} = 1.2$	$\frac{6}{20} = 0.3$	30%
	20		1	100

##

Graphical representation of a frequency curve

Histogram

↓
2-D

Frequency Polygon

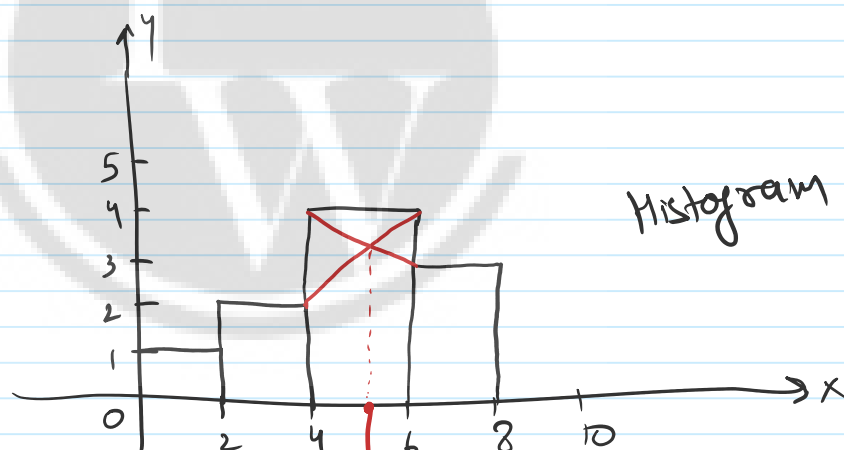
↓
1-D

Ogive

↓
1-D

Histogram

CI	f_i
0-2	1
2-4	2
4-6	4
6-8	3

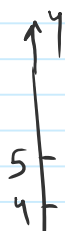


5.2 $\Rightarrow$ mode

we can find mode graphically using Histogram

Frequency Polygon

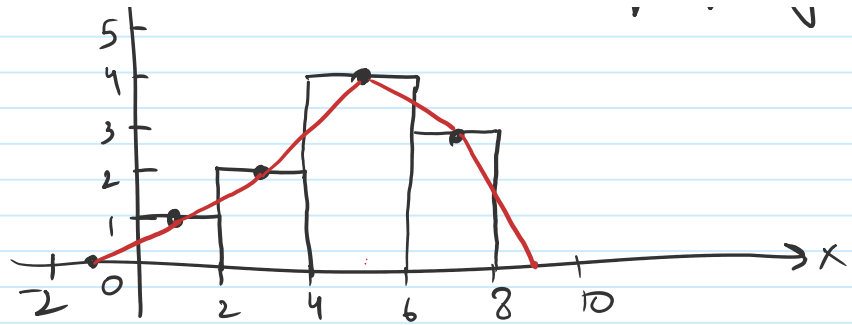
CI	f_i
0-2	1
2-4	2



Join mid point of each Histogram.
& we obtain frequency polygon



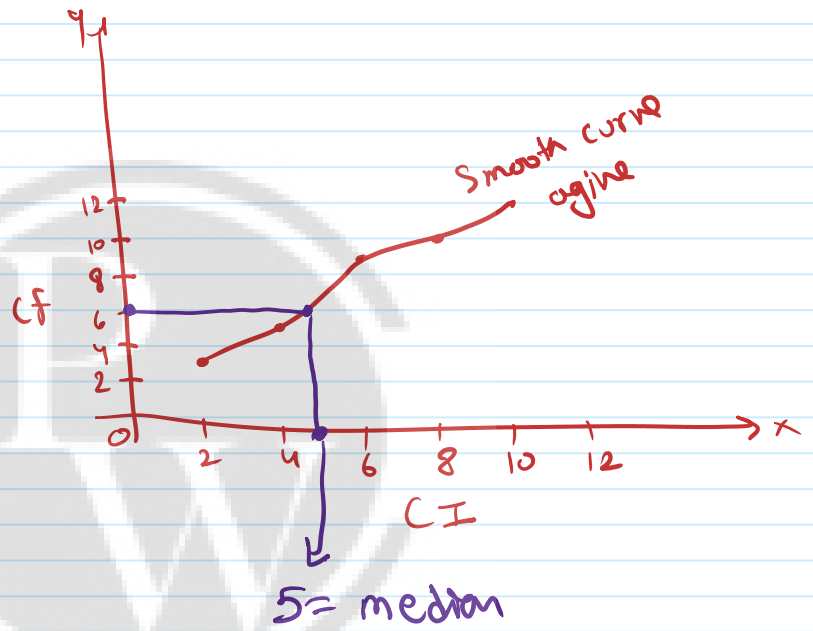
0-2	1
2-4	2
4-6	4
6-8	3



ogive

CI	f_i	cf
0-2	3	3
2-4	2	5
4-6	4	9
6-8	1	10
8-10	2	12
	12	

cumulative frequency



$$N=12$$

$$\frac{N}{2}=6$$

5 = median

"we can find median graphically using ogive"

Frequency curve

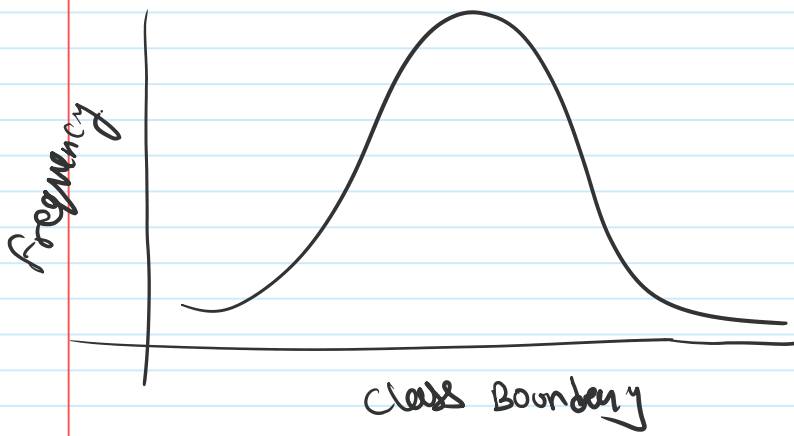


Smooth curve obtained by joining midpoint of upper side of rectangles in histogram

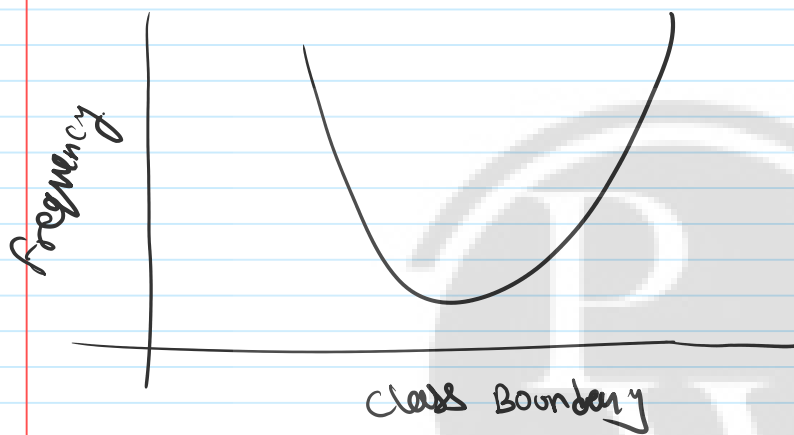


Area under frequency curve is one unit

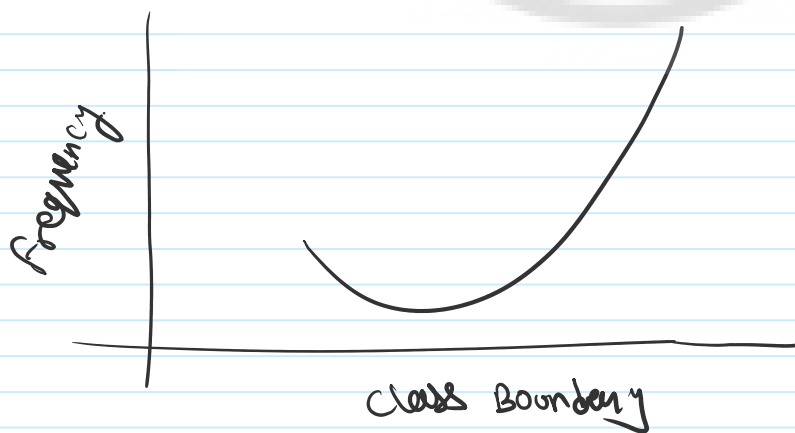
Area under frequency curve is one unit



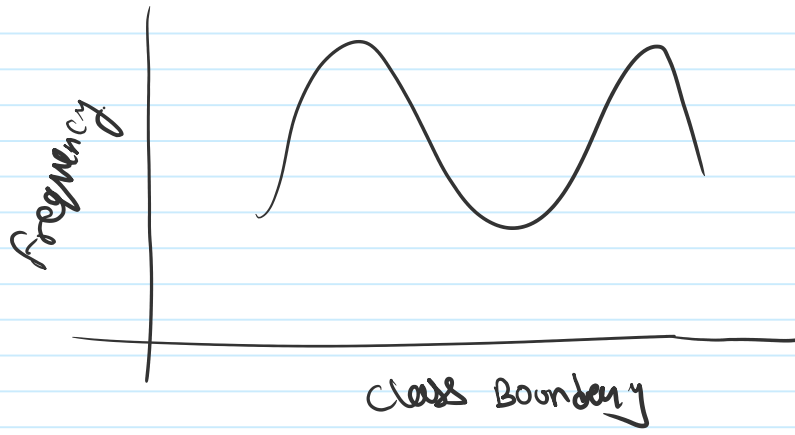
Bell Shape
(weight, profit)



U-Shape
(passengers in a metro)



J-Shape



mixed curve

