Data represntations

Data divided into two types

- *Categorical data*
- *Numerical data*

Categorical data Represntations:

statement:

In a class 30 *members Girls are there and* 20 *Boys are there*

How can we represent the above statement in table format

Tabular representation

Gender	Count
Boys	20
Girls	30

Gender ==== > *class*

class means category

Count ===== > *class frequency*

Frequency table

Class	Class frequency
Boys	20
Girls	30

G	E	Grad
B	M	
G	B	
B	B	
B	M	
G	t2	
G	t2	
G		

1 ... √20

C	cf
B	10
G	10

C	f
M	5
B	10
t2	5

Frequency table has two columns

- *class : categorical clumn*
- *Class frequncy: Numerical column*

<mark>Graphical representation:</mark>
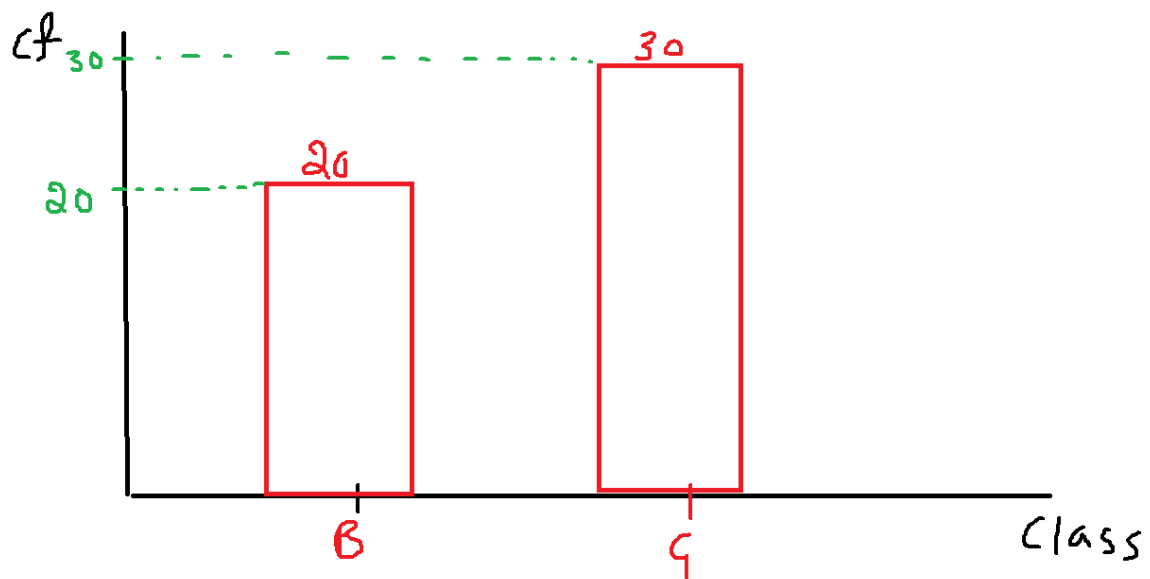
In order to plot Frequency table we will use bar chart

Bar chart or Bar plot

Bar chart will take $x-axis$: class (categorical column)

$y-axis$: class frequency (numerical column)

Note:

If we want plot a bar chart we required one categorical column

we required one numerical column

The chart above shows:
- cf axis with values 20 and 30
- Bar labeled B with value 20
- Bar labeled G with value 30
- x-axis labeled "Class"

:

Relative frequency ther percentage of the class members

In the above example

Boys are 20 out of total 50

Girls are 30 out of total 50

Class	Class frequency	Relative Frequency
Boys	20	0.4 = 40%
Girls	30	0.6 = 60%
	50	100%

==Graphical Represenations==

Pie chart

25 0 25
5
12 5
12 5 5 0 25

6 0 4
40 B
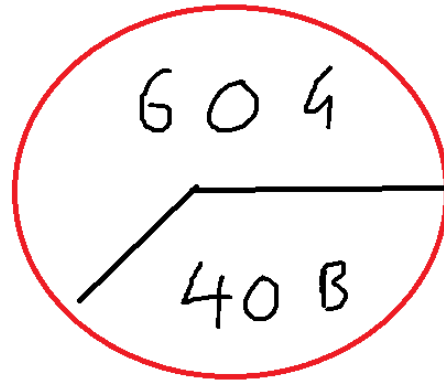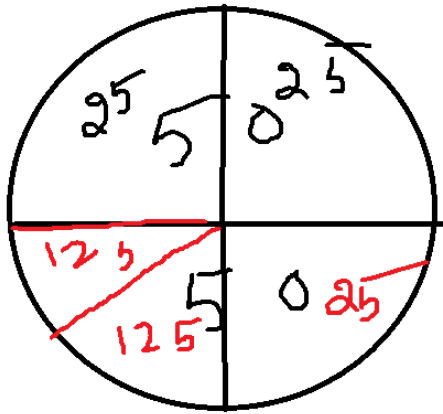
All together:

Tabular representation	Graphical represntations
Frequency table	Bar chart $(x-axis: cat, y-axis: numerical)$
Relative frequency	Pie chart

Numerical column representation:

In a class 10 *members are there* (*Boys and Girls*)

raw data : *marks*

$5, 8, 12, 15, 20$

6,9,11,4,17

we can make raw data into some groups

we can get a count of each group

Frequency distribution table

- *It has two columns*
- *One column : class interval*
- *Another column : class interval frequncy*

5, 8, 12, 15, 20,6,9,11,4,17

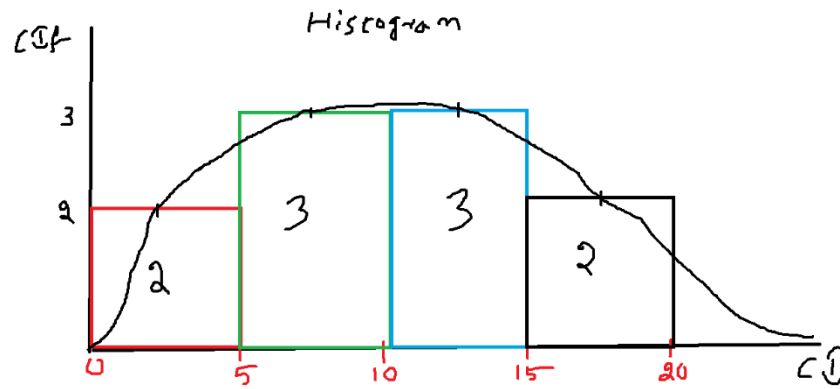| class interval	class interval frequncy
$0 - 5$	2
$5 - 10$	3
$10 - 15$	3
$15 - 20$	2
	$total = 10$

If you want to add 5 in first interval , then dont add in second interval

Histogram

- *It is a graphical representation of Numerical data*
- $X - axis$: *Class interval*
- $Y - axis$: *Interval frequency*
- *Numerical vs Numerical*
- *Class intervals we called bins*

-

  whenever you see any distribution plot
- *Immediatly you need to recall Histogram*
- *Whenever you are thinking about Histofgram*
  - *Immediatly you need to recall CI vs CIF (Frequency distribution table)*
- *Whenever you are thinking about Freqiency distribution table*
  - *Immediatly you need to recall Raw data*

Distribution == Histogram == Freq dist table == CI vs CIF == raw data
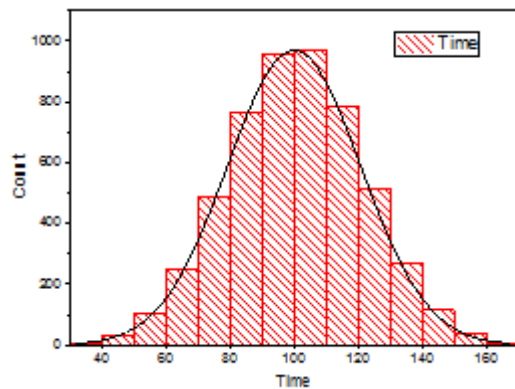
How to choose the Intervals

what happens If I choose less intervals:

- *The information will miss*
- *we can not descriminate the exact catogiry*

What happens If I choose More intervals:

- *un wanted information occures*

TABLE 2–4 Prices of Vehicles Sold Last Month at Whitner Autoplex

						Lowest
$23,197	$23,372	$20,454	$23,591	$26,651	$27,453	$17,266
18,021	28,683	30,872	19,587	23,169	35,851	19,251
20,047	24,285	24,324	24,609	28,670	15,546	15,935
19,873	25,251	25,277	28,034	24,533	27,443	19,889
20,004	17,357	20,155	19,688	23,657	26,613	20,895
20,203	23,765	25,783	26,661	32,277	20,642	21,981
24,052	25,799	15,794	18,263	35,925	17,399	17,968
20,356	21,442	21,722	19,331	22,817	19,766	20,633
20,962	22,845	26,285	27,896	29,076	32,492	18,890
21,740	22,374	24,571	25,449	28,337	20,642	23,613
24,220	30,655	22,442	17,891	20,818	26,237	20,445
21,556	21,639	24,296				

Highest

how to create class intervals and class width

From Raw data === > Frequency distribution table

Class intervals and Class frequency

It is very important how to choose Class interval and width of the interval

$$raw\ data\ observations\ \ n = 80$$

k	2^k
0	$2^0 = 1$
1	$2^1 = 2$
2	$2^2 = 4$
3	$2^3 = 8$
4	$2^4 = 16$
5	$2^5 = 32$
6	$2^6 = 64$
7	$2^7 = 128$

$$In\ \ which\ values\ 80\ occures$$

$$between\ k = 6\ \ and\ \ k = 7$$

$$we\ can\ not\ take\ k = 6\ because\ 64\ only\ , It\ is\ not\ equal\ to\ 80$$

$$so\ will\ consider\ k = 7$$

$$\text{The number of intervals are } k = 7$$

$Interval\ width$:

$$w = \frac{Highest\ value - Lowest\ value}{no\ of\ intervals}$$

$$w = \frac{36000 - 15000}{7} = \frac{21000}{7} = 3000$$

Class interval	Interval frequency
$15000 - 18000$	
$18 - 21$	
$21 - 24$	
$24 - 27$	
$27 - 30$	
$30 - 33$	
$33 - 36$	

Selling Prices ($ thousands)	Frequency
15 up to 18	8
18 up to 21	23
21 up to 24	17
24 up to 27	18
27 up to 30	8
30 up to 33	4
33 up to 36	2
Total	80

$$23 + 17 + 18 = 58 = 58 * \frac{100}{80} = 72\%$$

72% of cars price range between $18k$ to $27k$