Q2 Difference b/w histogram & barplot:

Histogram visualize quantitative data or numerical data.

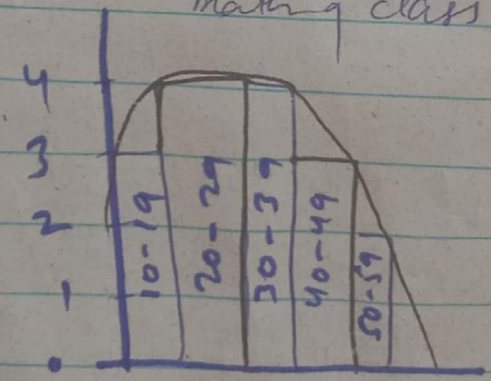
whereas, bar displays categorical value.

Ages: {10, 12, 12, 14, 23, 20, 32, 19, 21, 29, 36, 56, 45, 52, 41, 40, 31, 39}

10, 12, 12, 14, 19, 20, 21, 23, 29, 31, 32, 36, 39, 40, 41, 45, 52, 56

10 - 19	3
20 - 29	4
30 - 39	4
40 - 49	3
50 - 59	2
	18

∴ Barchart deals every value as an individual histogram deals it by making class.



Histogram.

Discrete:
1.4, 3.1, 6.4

Continuous:
1, 2, 3, 4

Measure of Central Tendency:
Mean, Median, Mode

Measure of disperson:

Standard Deviation
Variance

ages = [13, 14, 15, 19, 20, 45]

Mean:

μ = population mean. $\mu = \Sigma x / N$
$\bar{x}$ = Sample mean $\bar{x} = \Sigma x / n$

Σx = sum of all data values, N = number of data items in population, n = number of data items in sample.

Mode: [13, 14, 15, 19, 20, 45]

∴ Most repeated value, All are unique, so every value up here, is mode.

Median: the number in the middle.

For odd =

$$\left(\frac{n+1}{2}\right)^{Th}$$

For Even =

$$\frac{\left(\frac{n}{2}\right)^{Th} + \left(\frac{n}{2}+1\right)^{Th}}{2}$$

Even = 2, 4, 6, 8, 10
Odd = 1, 3, 5, 7, 9

2/3

"The more away from mean, The data is more unvaluable"

"The more close the mean, The data is more valuable"

Measure of dispersion:

- Standard Deviation
- Variance

Mean: [21, 21, 21, 21, 21, 21, 21] = 21

Mode: 21

Median: 21

- Variance: MOD that takes into account the spread of all data points in a data set.

Population:

$$\sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2 / N$$

σ^2 = population variance

x_i = value of i^{th} value/element

μ = population mean

N = population size.

Sample:

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)$$

s^2 = sample variance

x_i = value of i^{th} element

$\bar{x}$ = sample mean

n = sample size

:= Degree of Freedom!.

-1:

"Sample must be 1 less than population"

Minusing 1 in the end, to eliminate the factor of dependency.

[13, 14, 19, 20, 45] → Sample.

Standard deviation:

$s = 12.083046$ unit of dispersion

Variance::

$s^2 = 146$ variance.

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / n - 1$$

count = $n = 6$

Mean = $\bar{x} = 21$

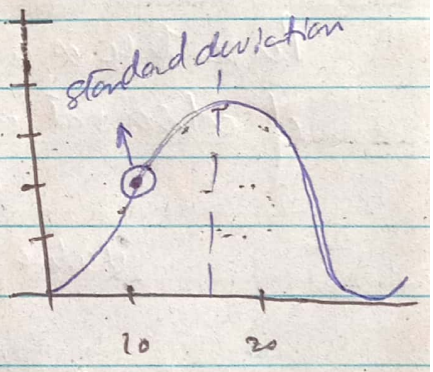
Sum of Squares: $SS = 730$

$s^2 = SS / n - 1$

$s^2 = 730 / 6 - 1$

$s^2 = 730 / 5$

$\boxed{s^2 = 146}$



How do we normalize a data:

$$\frac{x - \bar{x}_i}{sd} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Subtracting it with mean & dividing by it by its standard deviation

- 5 number theory and boxplot:
  (Box & Whisker Plot)

Percentage:
a rate, number or amount in each hundred.

Percentile:
In Terms of performance, of a student get 70 percentile, thats mean you are better than 70 and worst than 30.

or

Percentile & Quantile used to find outliers.

or

Percentile is a value below which a cartain percentage of observation lies.

or

value below which a percentage of data fall

Data = [2,2,3,4,5,5,5,6,7,8,8,8,8,8,9,9,10,11,11,12]

Percentile ranking of 10: $\frac{\text{Num of value below}}{n} * 100$

16/20 * 100 = 80.1.

$\frac{n}{20} * 100 = 70\%$.

$n = 0.7 \times 20$

$n = 14^{th}$ value = 18

Five Number Summary:

Minimum, First Quartile, Median, Third Quartile, Maximum.

$x = [1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 7, 7, 8, 8, 8, 7]$

Lower fence $\neq$ Min $= Q1 - 1.5\,IQR$

Upper fence $\neq$ Max $= Q3 + 1.5\,IQR$

$IQR = Q3 - Q1$

$Q1 = (20 * 25)/100 = 5^{th}$ Value $= 3$

$Q3 = (20 * 75)/100 = 15^{th}$ Value $= 7$

$IQR = Q3 - Q1$

$= 7 - 3$

$= 4$

Lower fence: $Q1 - 1.5(4) = 3 - (1.5)4 = -3^{th}$ Value

Upper $= Q3 - 1.5(4) = 7 + 1.5(4) = 13^{th}$ val