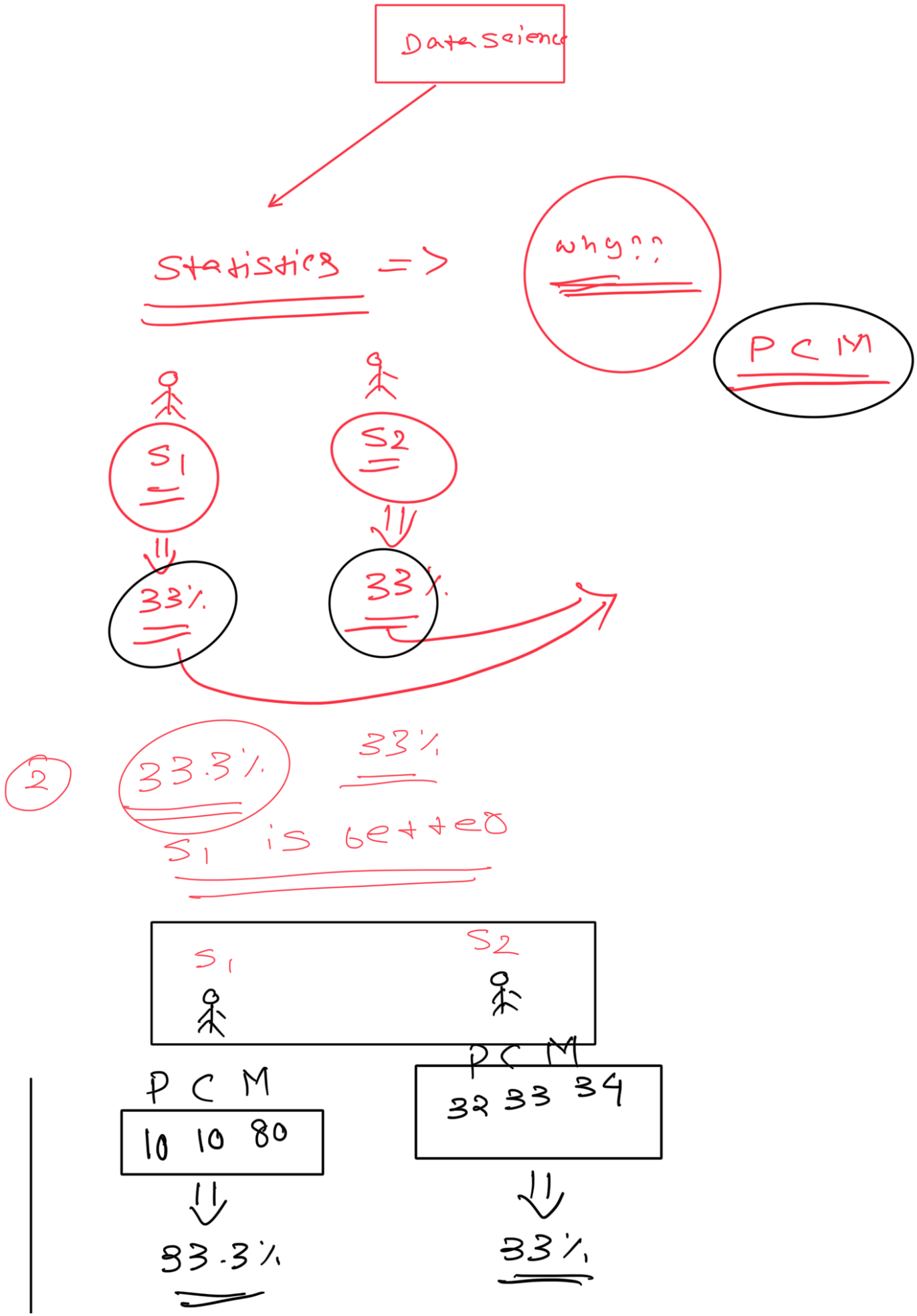
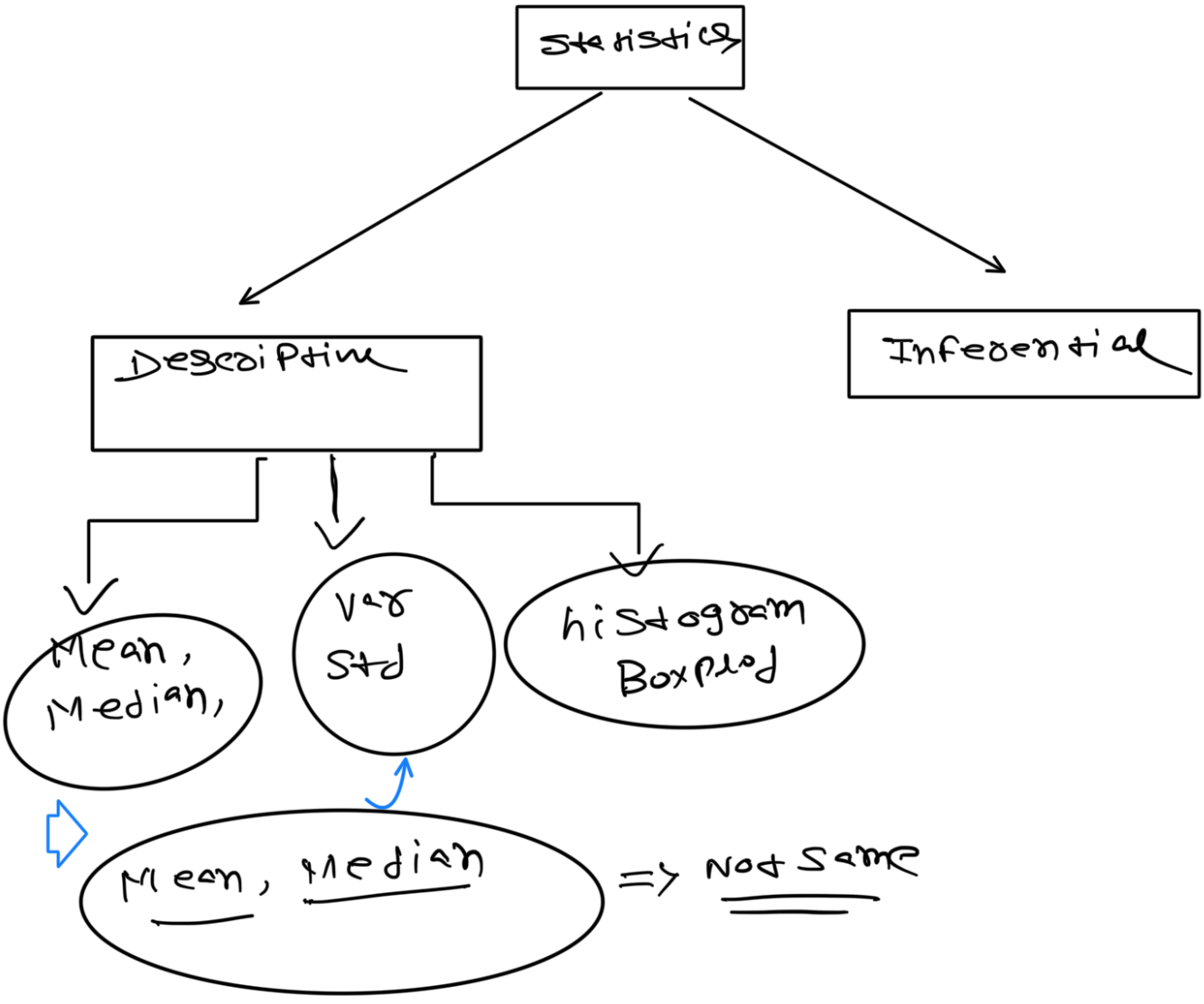

Data science
Statistics =>
why??
PCM
S1
S2
33%
33%
2
33.3%
33%
S1 is getted
S1
S2
PCM
PCM
10 10 80
32 33 34
33.3%
33%

Statistics => why ??

S_1 =

S_2 =



5, 7, 3, 9

Data =>

Mean = $\frac{\text{sum}}{\#\text{count}} = \frac{25}{5} = 5$

Median => 1, 3, 5, 7, 9

Median is 5

1, 3, 7, 5

1, 3, 5, 7

$\frac{3+5}{2} = 4$

If Data => mean => Median

Next ?? Different

=> 1, 5, 8, 12

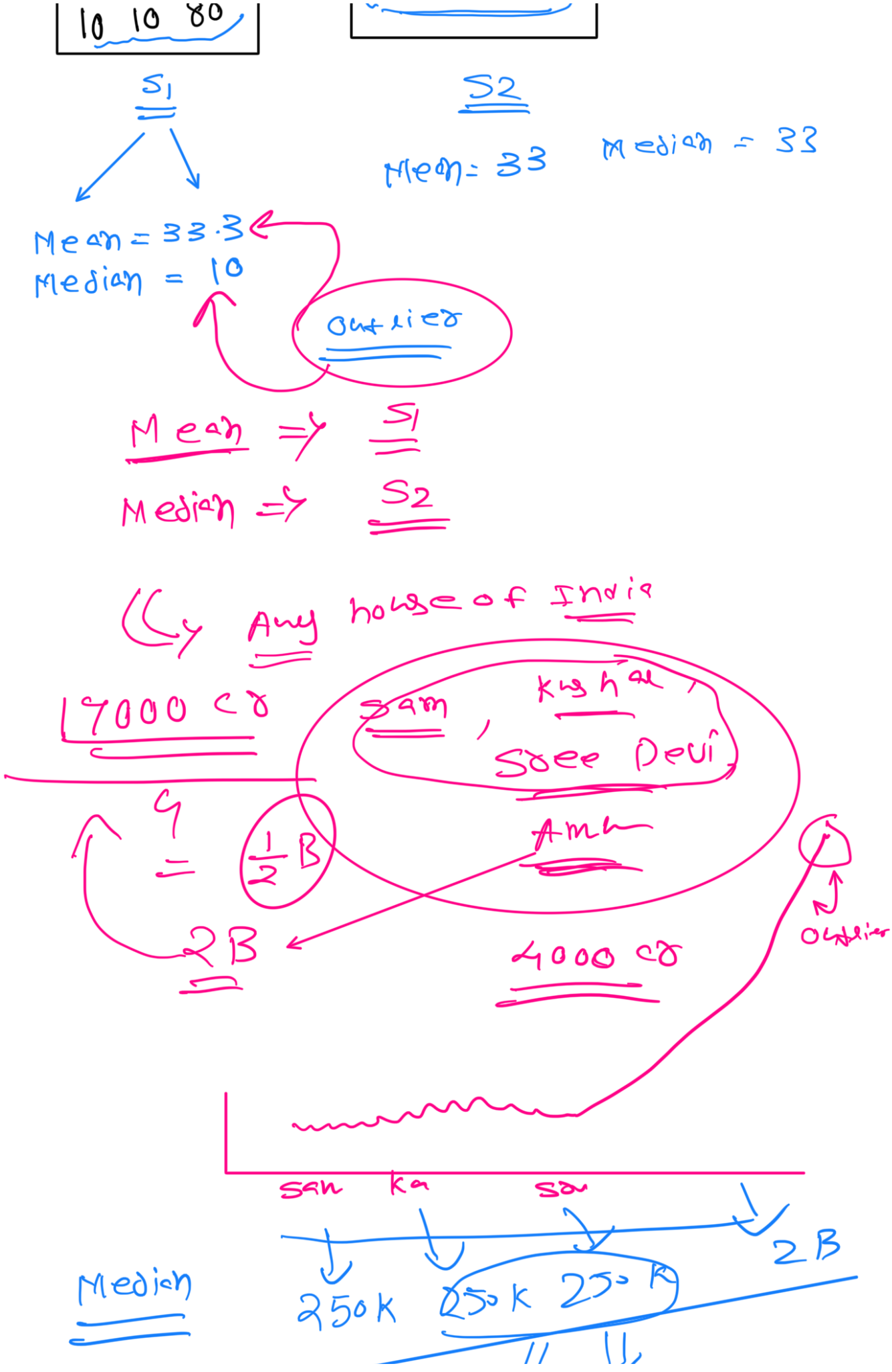
Mean & Median

6.5

PCM

PCM

32 33 34


10 10 80
S1
S2
Mean = 33
Median = 33
Mean = 33.3
Median = 10
Outlier
Mean => S1
Median => S2
Any house of India
17000 cr
Ram, Kushal, Sita Devi
Amt
1/2 B
2B
4000 cr
Outlier
Ram Ka Sar
2B
Median
250K 250K 250K

250K

S1

10%

S2

12%

=> Median

10, 11, 12

10, 40, 100

Variance =>

Data => 1, 3, 5, 7, 9

5

$(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2$

5 - 1

count

count - 1

-4 -2

$$= \frac{16+4+0+4+16}{4} = \frac{40}{4} = 10$$

10 13, 15, 17, 20

15 →

$$\frac{(10-15)^2 + (13-15)^2 + (15-15)^2 + (17-15)^2 + (20-15)^2}{5-1}$$

$$= \frac{25+4+0+4+25}{4} = \frac{58}{4}$$

$$= 14.5$$

STD : $\sqrt{\text{variance}}$

9

10, 10, 79

33

9

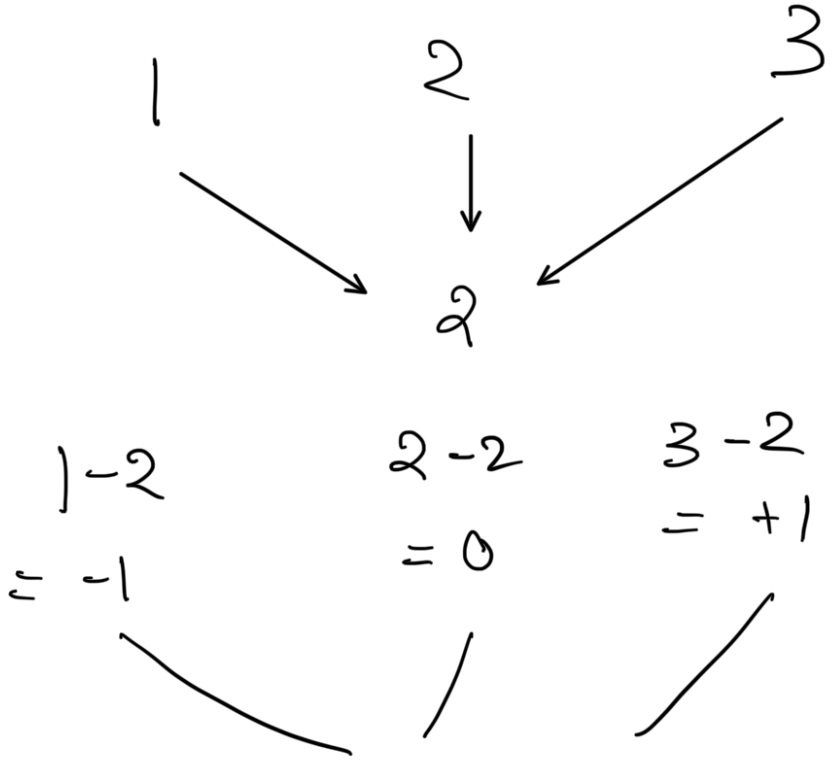
32, 33, 34 → 4

33

$$\frac{(10-33)^2 + (10-33)^2 + (79-33)^2}{2 \Leftarrow (\text{count} - 1)} \Rightarrow$$

$$\frac{(32-33)^2 + (33-33)^2 + (34-33)^2}{2} = \frac{1+0+1}{2}$$

$$= 1$$

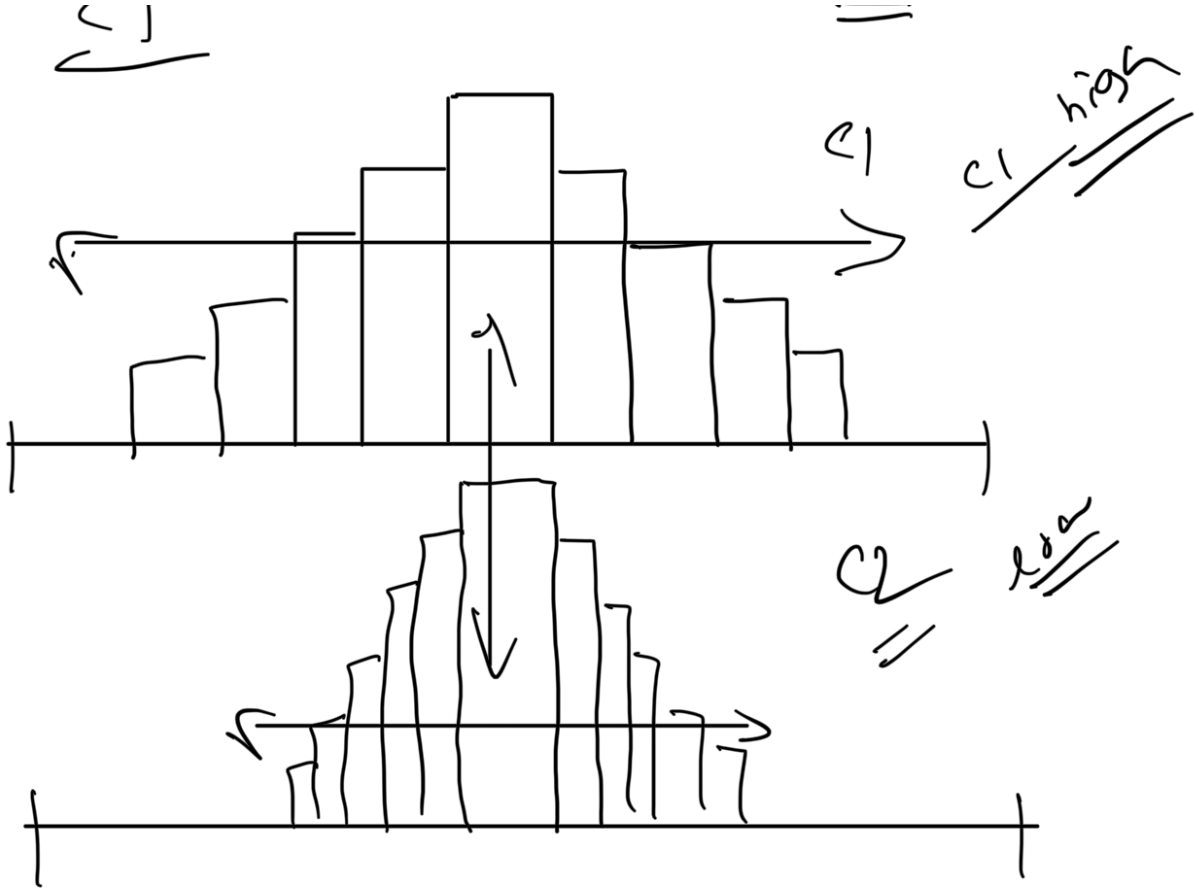


histogram

Boxplot

Plot

C2

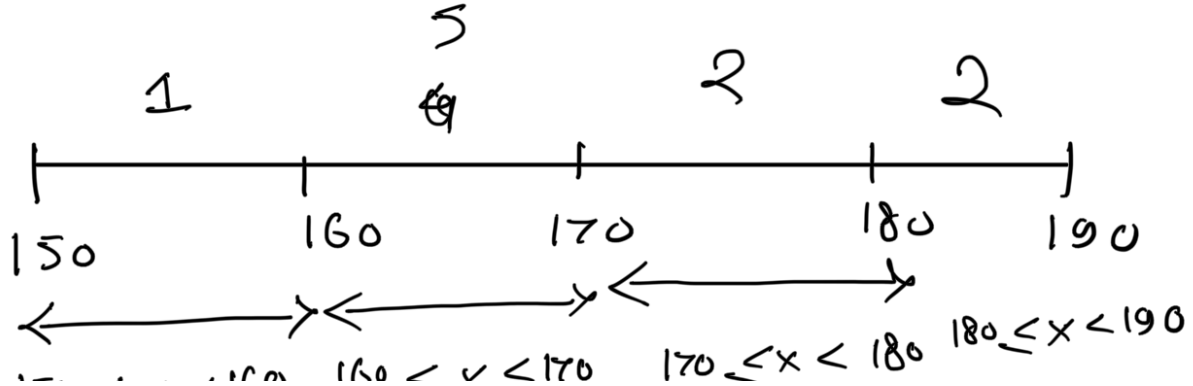


10 Marc wt

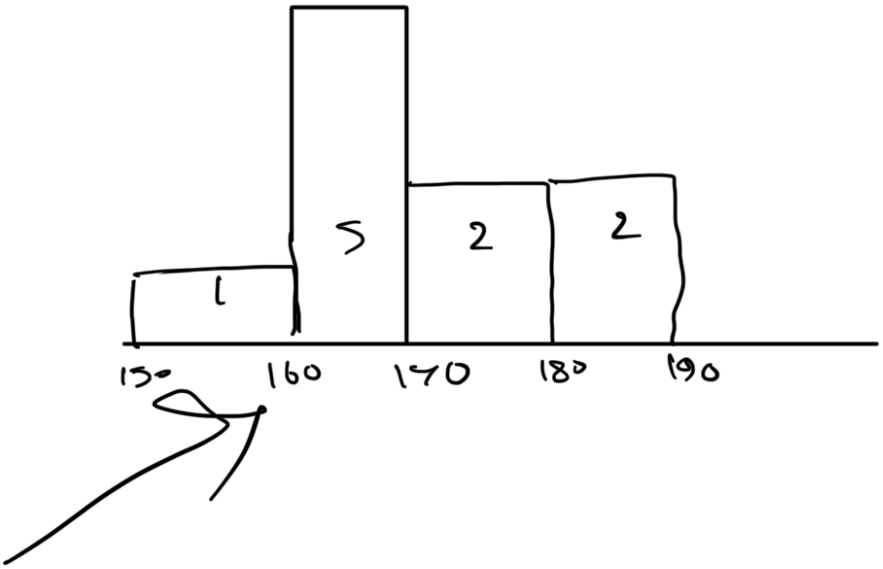
150, 165, 162, 170, 164, 171, 180, 185

161, 169

10 150 185 190



$150 \leq x < 190$



Boxplot

Outliers

S_1 → 10, 10, 80

OBJ

(i) => outliers ( )

=> Quantile => percentile

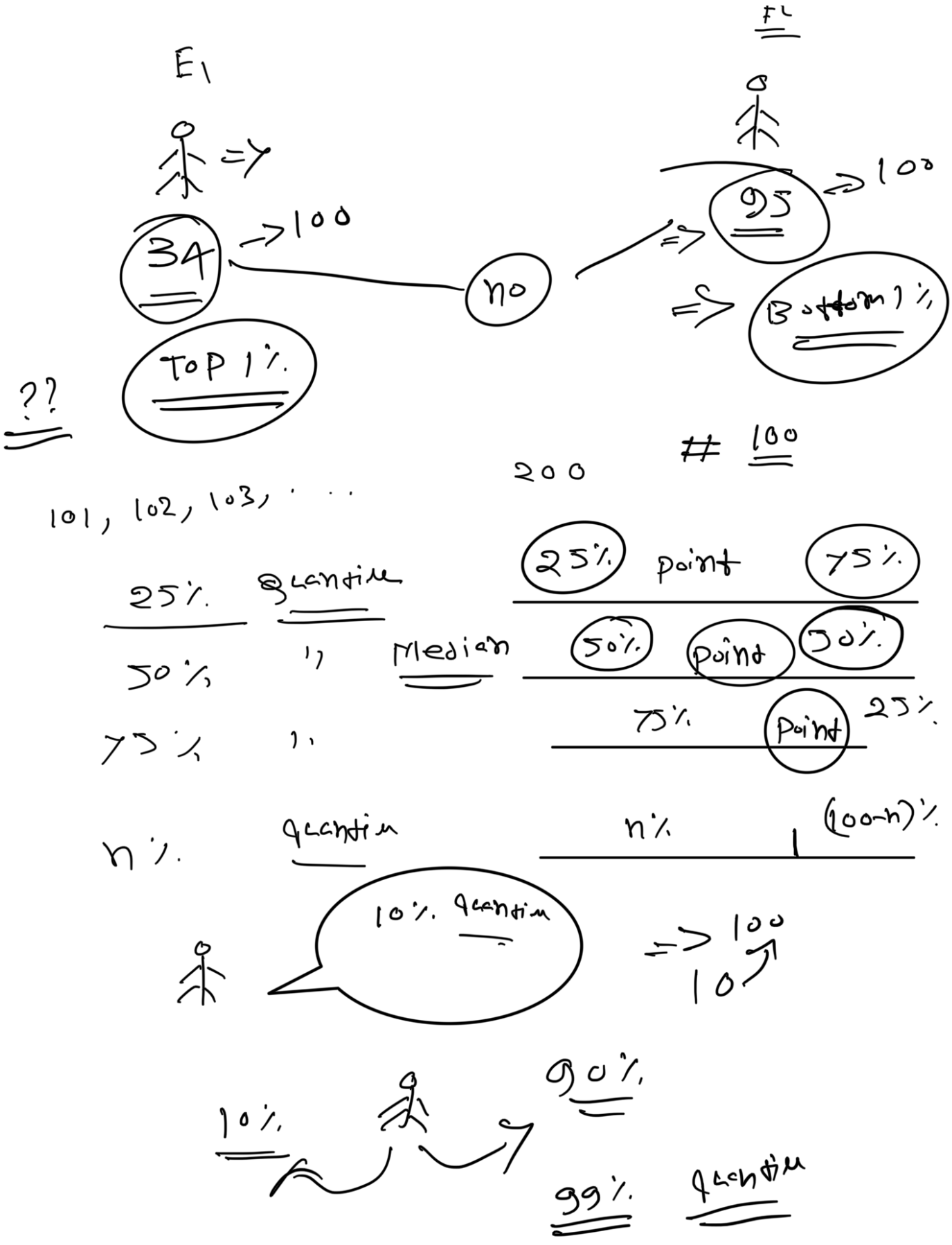
MBA

=> 98%

99%

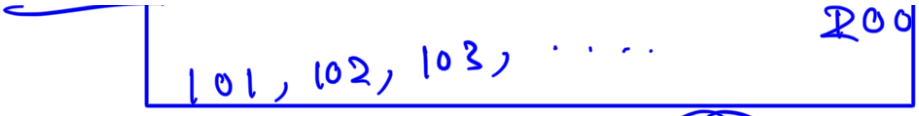
35 → 100

higher


E1
E2
34
100
95
100
no
TOP 1%
Bottom 1%
??
100
200
101, 102, 103, ...
25% quantile
50% "
75% "
n%
quantile
Median
25% point
50% point
75% point
n%
(100-n)%
10% quantile
=> 100
10
10%
90%
99%
quantile

Date

100



(i) 25%

↓

1st

50%

2nd

75%

3rd

count of the data

$\frac{125+126}{2}$

$= 125.5$

25% => $100 \times 25\% = 25$ → 25th 26th

125 126

50% => $100 \times 50\% = 50$ → 50th 51th

150.5 150 151

75% => $100 \times 75\% = 75$ → 75th 76th

175 176

175.5

=

101, 102, 103, ... 200, 201

n = 101

25% → 101 × 25% → 25.25 → 26th 126

50% → 101 × 50% → 50.5 → 51th 151

75% → 101 × 75% → 75.75 → 76th 176

10, 12, 11, 25, 17, 15, 18, 20, 10, 11

25% 50% 75%

Data => 10, 11, (11), 12, 15, 17, 18, (19), 20, 25

count = 10

25% → 10 × 25% => (2.5) → 3 → (11)

50% → 10 × 50% => (5) → 5 & 6 → 15+17/2 = 16

75% → 10 × 75% => (7.5) → 8 → 19

25% => 11

50% => 16

75% => 19

Box Plot

S1

Data: 101, 102, ... 199, $\frac{250, 300}{2}$

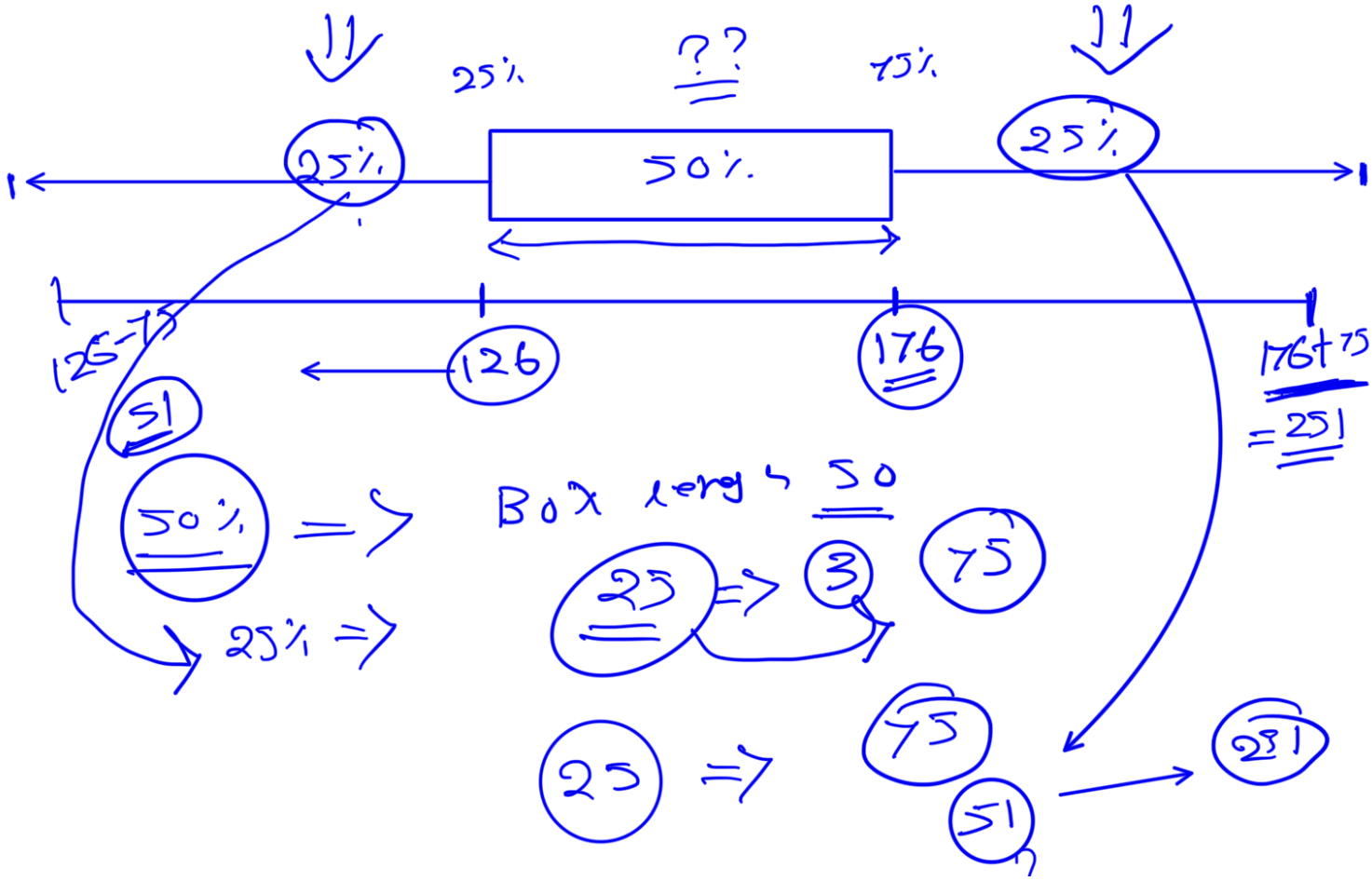
count => 25% 75%

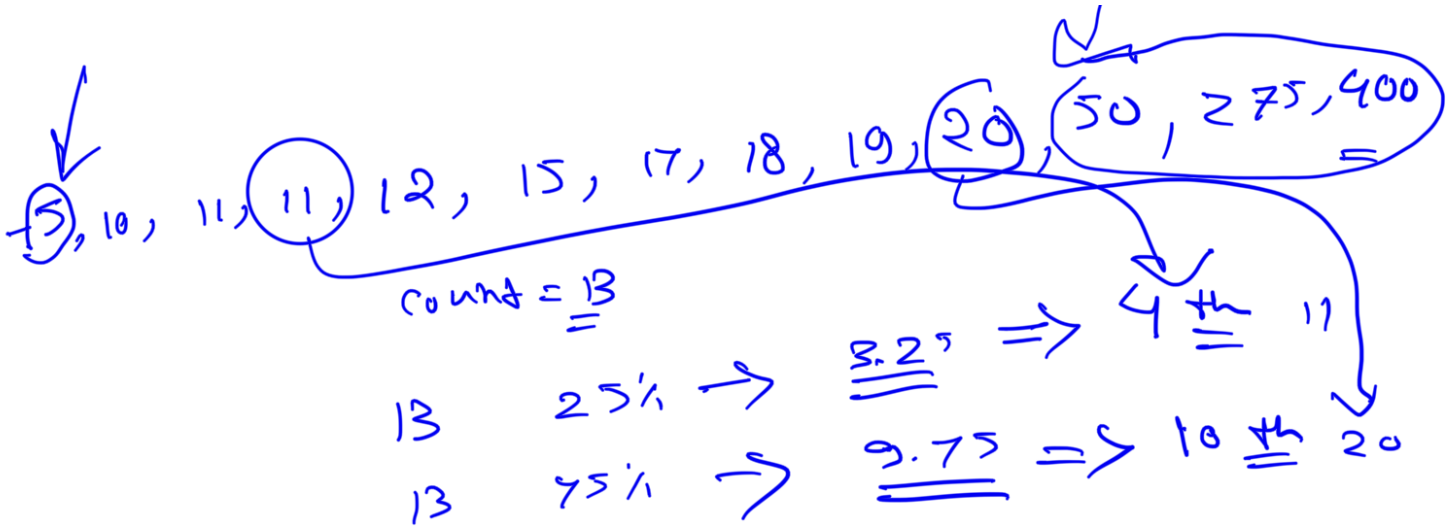
101

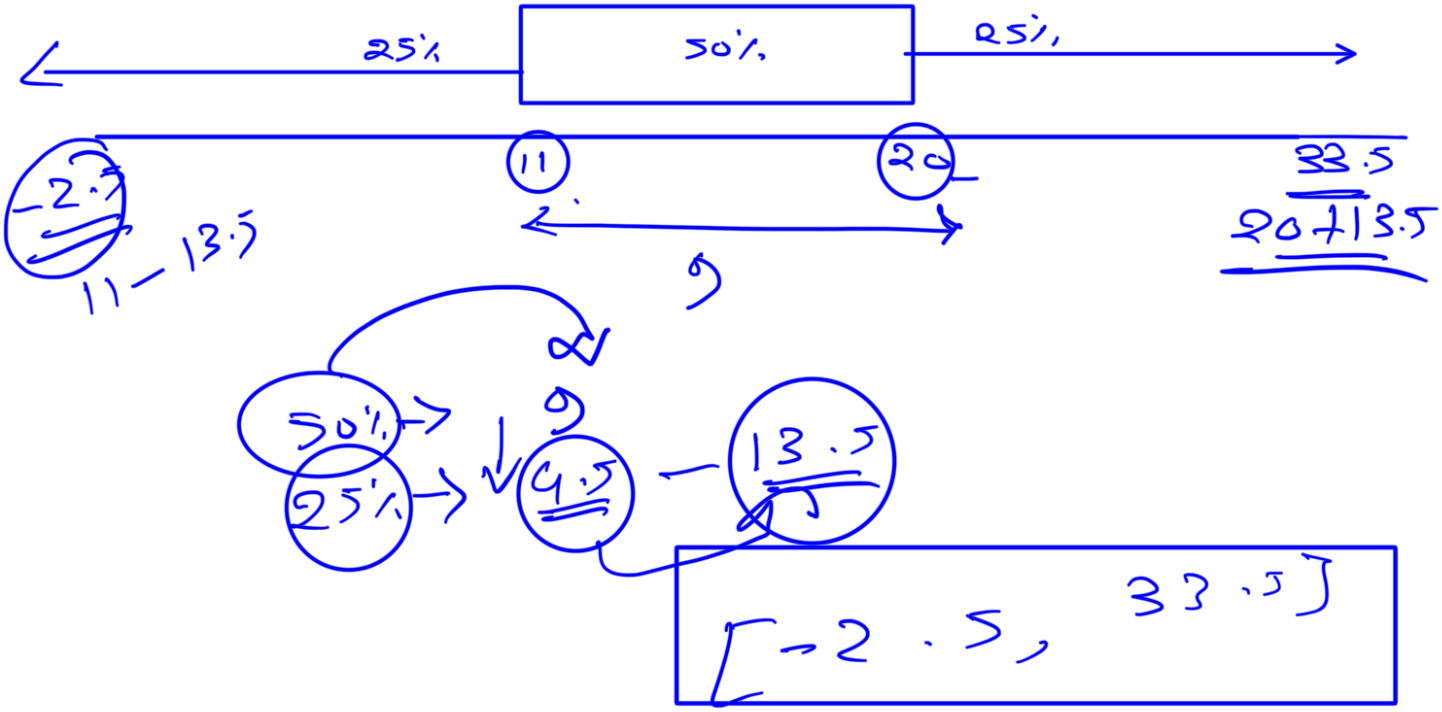
$101 \times 25\% \Rightarrow 25.25 \sim$ 26th 126

$101 \times 75\% \Rightarrow 75.75 \sim$ 76th 176

Boxplot




5, 10, 11, 11, 12, 15, 17, 18, 19, 20, 50, 275, 400
count = 13
13 25% → 3.25 ⇒ 4th 11
13 75% → 9.75 ⇒ 10th 20


25%
50%
25%
11
20
-2.5
11 − 13.5
33.5
20 + 13.5
50%
25%
9.5
13.5
[-2.5, 33.5]