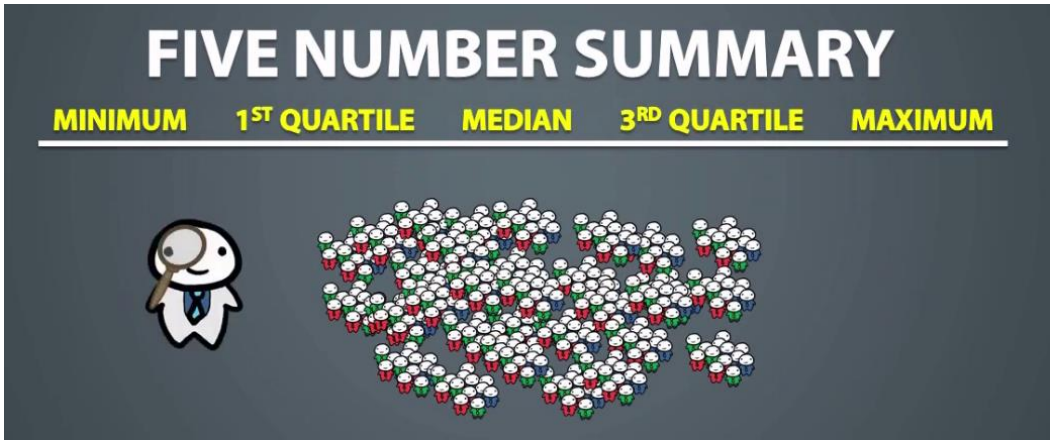
3. Maths - Statistics – PART – 4

Contents

4. Maths - Statistics – PART – 4

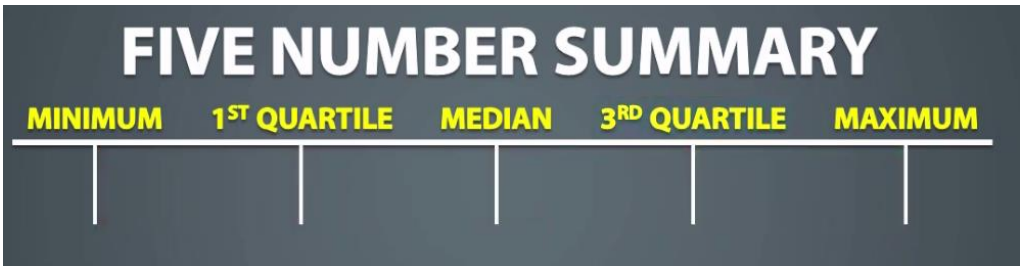
1. What is five number summary?

- ✓ The five number summary gives a way to describe the distribution

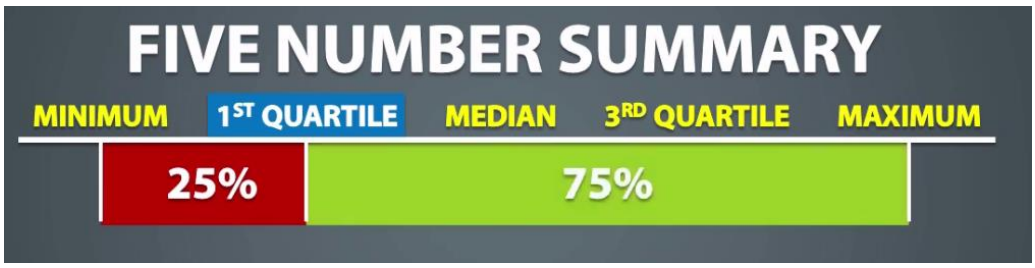


2. Explanation

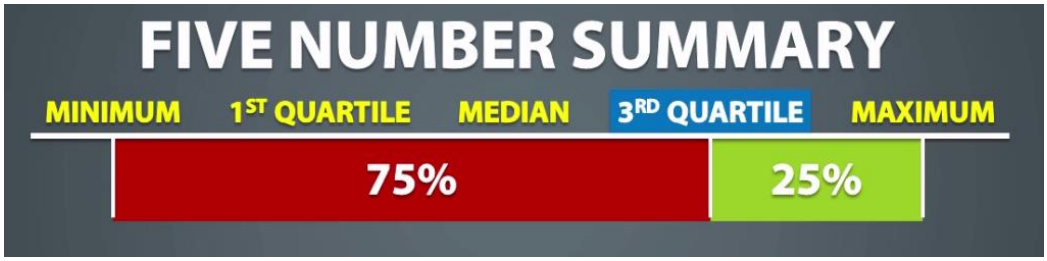
- ✓ So, if we took the data then we can find these five numbers in that data
- ✓ Minimum is smallest value in a dataset
- ✓ Maximum is largest value in a dataset
- ✓ Median is middle data value
    - It is the point 50% data values is below the median and 50% data values is above the median



- ✓ The median of bottom half is called as 1st Quartile
    - It is the position 25% of the data values are below this and 75% values are above the point
    - The 1st Quartile is essentially as median of the median

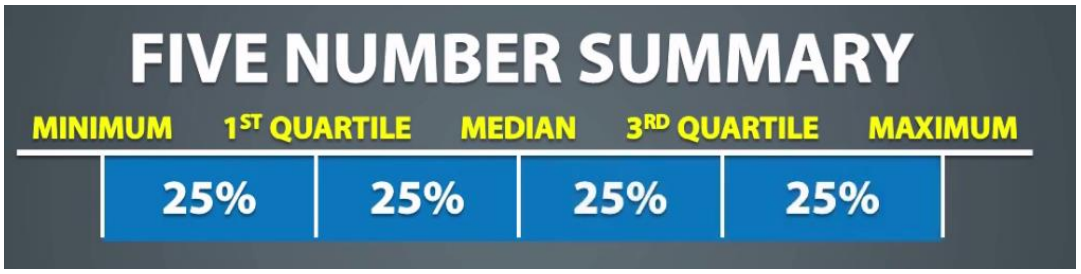


- ✓ The median of top half is called as 3rd Quartile
  - It is the position 75% of the data values are below this and 25% values are above the point

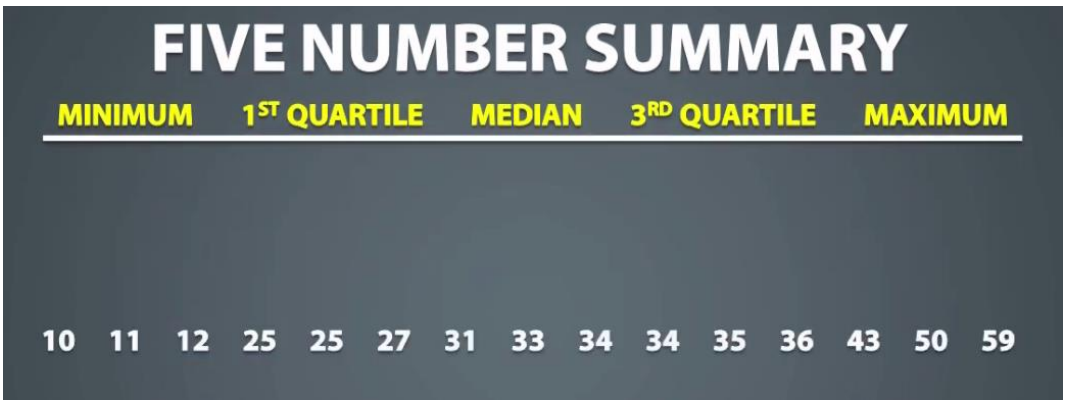


3. 4 equal quarters

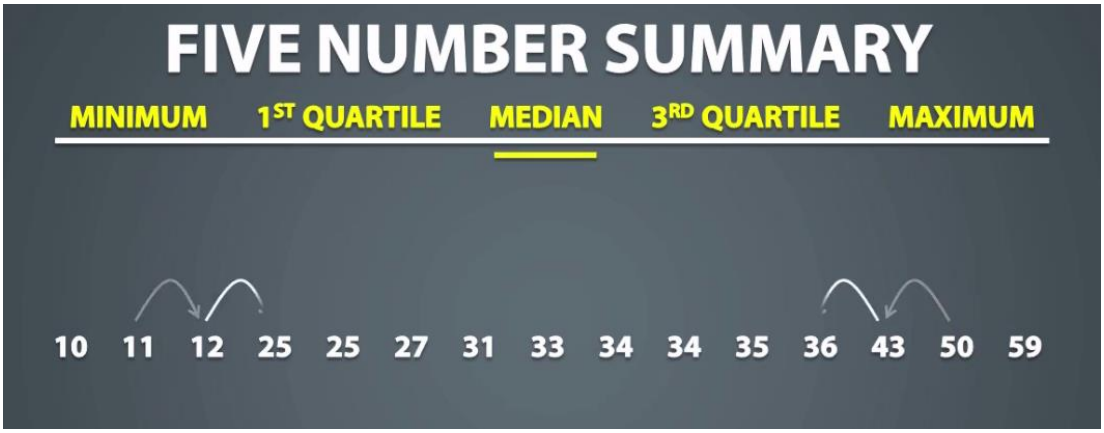
✓ The five number summary is divides the data into equal quarters

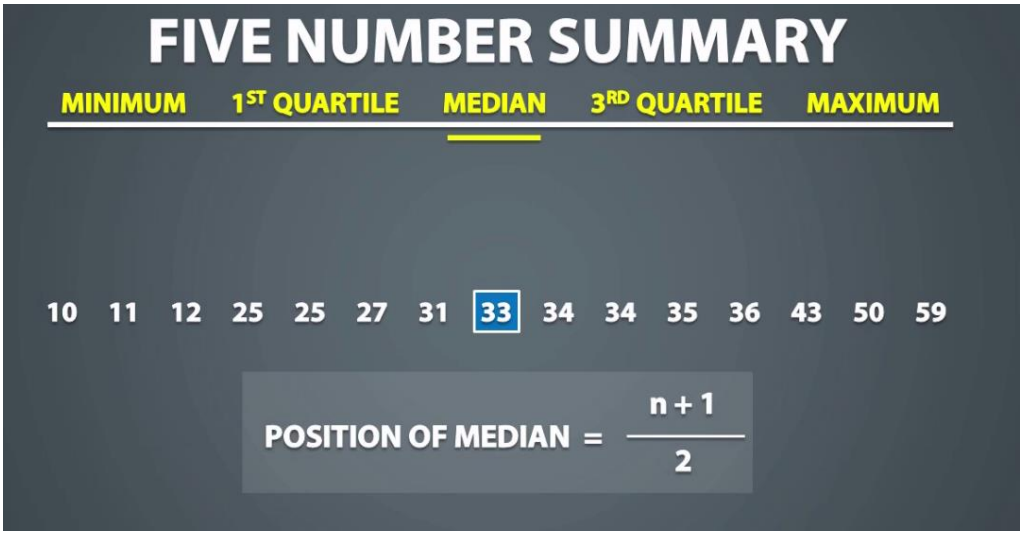


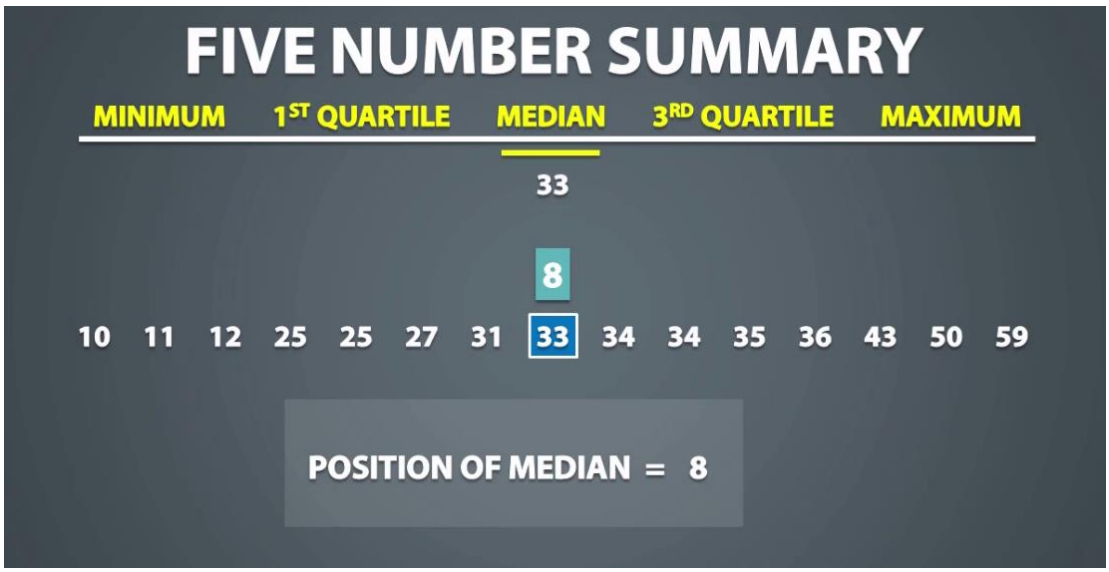
✓ Let's take one example to determine the five number summary



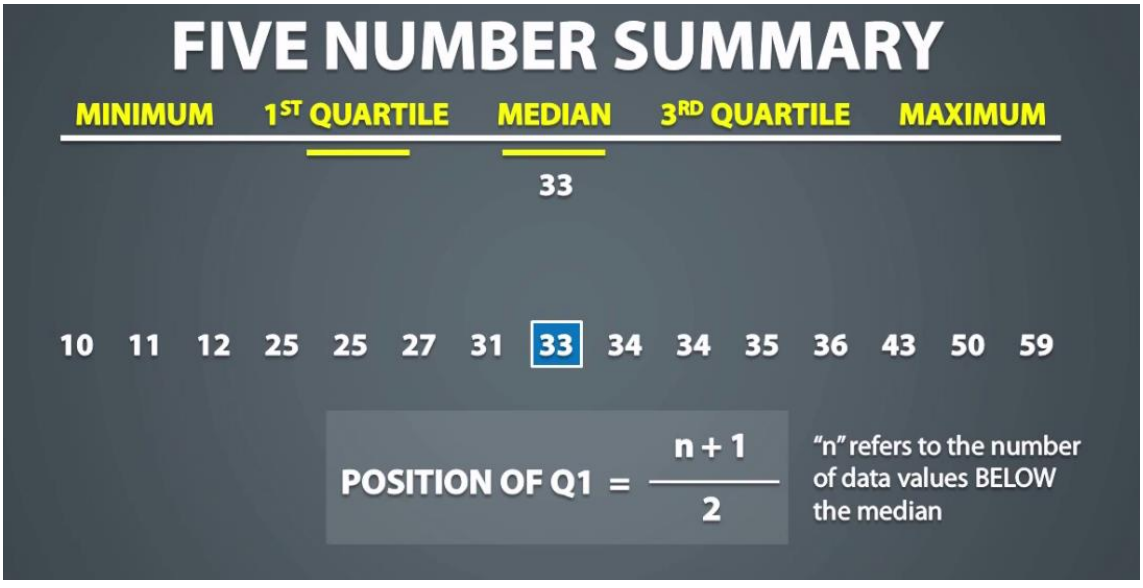
Median: Middle of the value

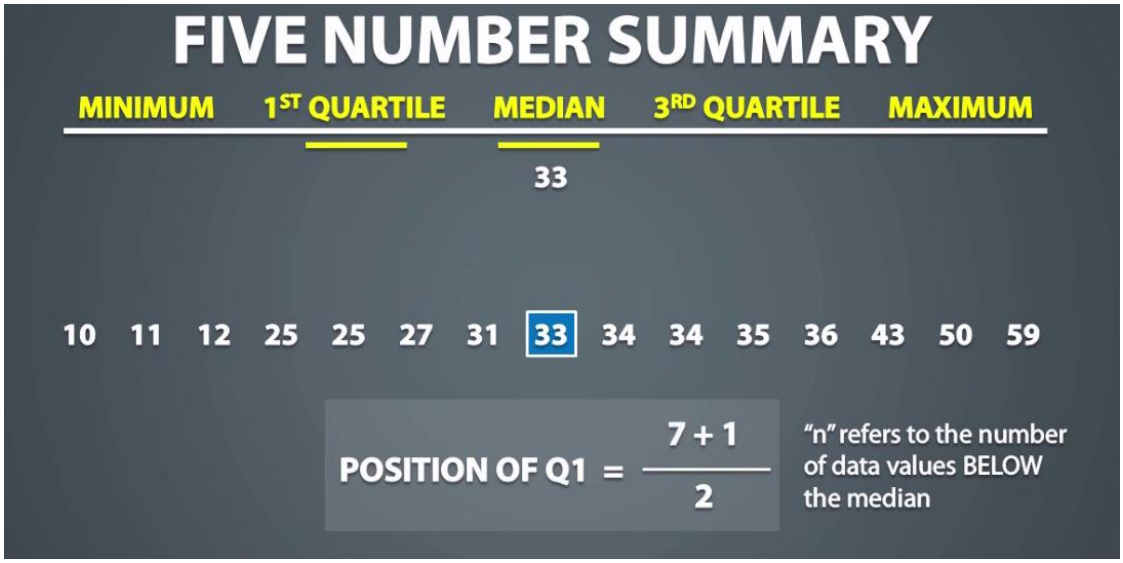


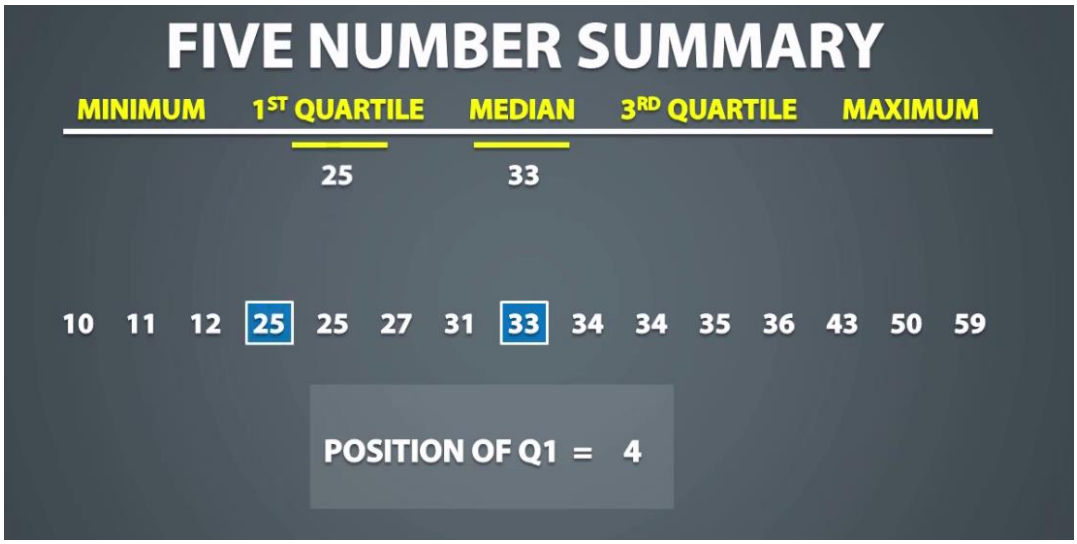




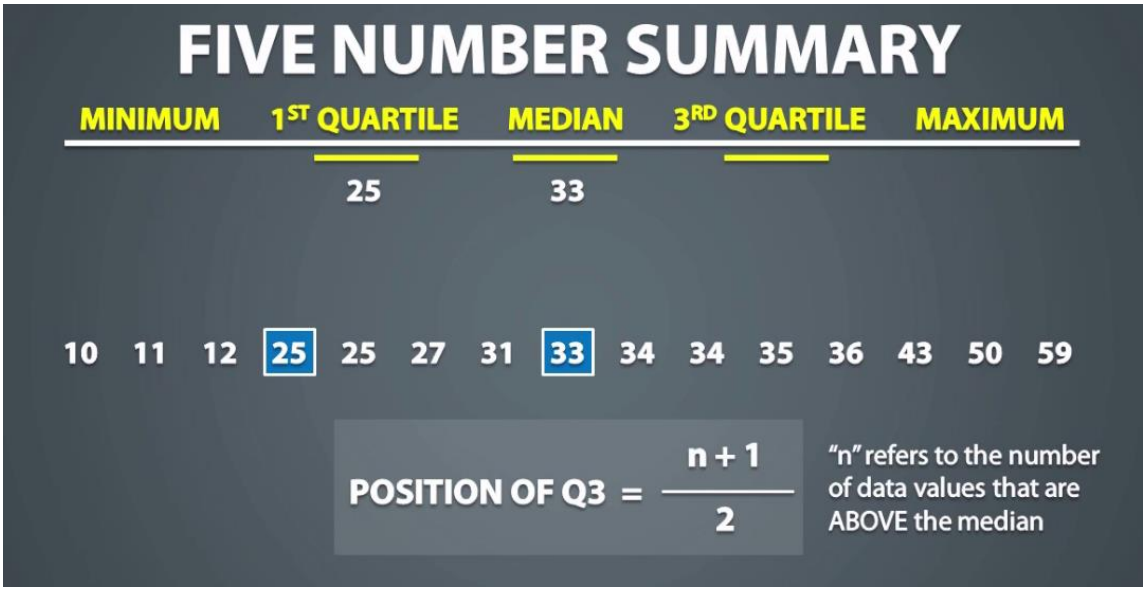
4. 1st Quartile

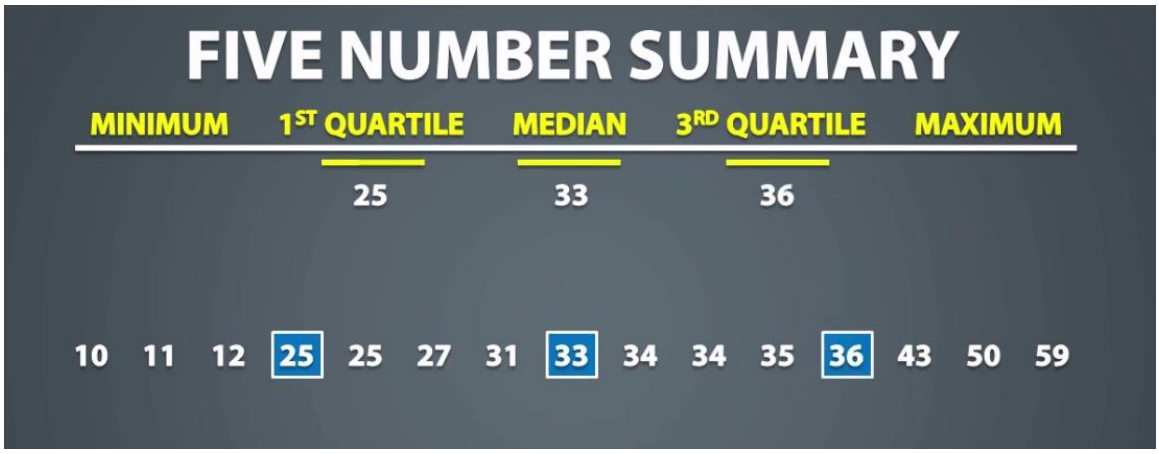



FIVE NUMBER SUMMARY
MINIMUM
1ST QUARTILE
MEDIAN
3RD QUARTILE
MAXIMUM
33
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
POSITION OF Q1 = 7 + 1 / 2
"n" refers to the number of data values BELOW the median

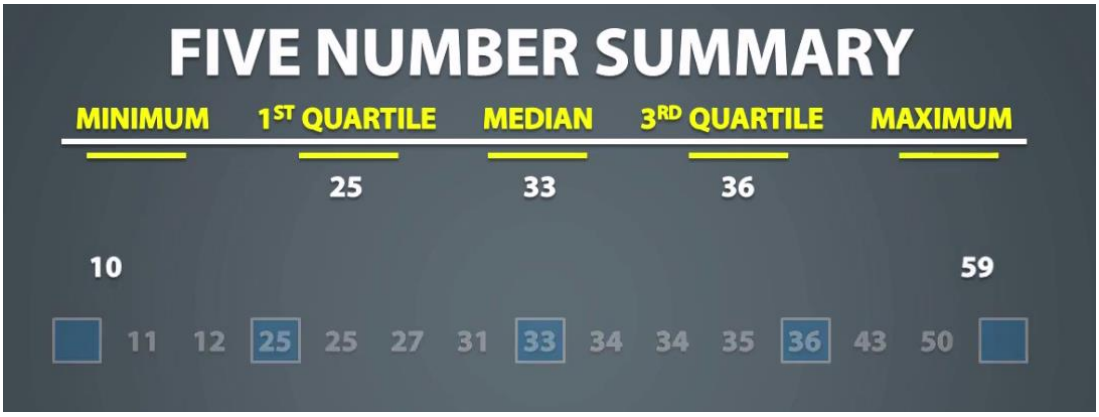

FIVE NUMBER SUMMARY
MINIMUM
1ST QUARTILE
MEDIAN
3RD QUARTILE
MAXIMUM
25
33
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
POSITION OF Q1 = 4

5. 3rd Quartile



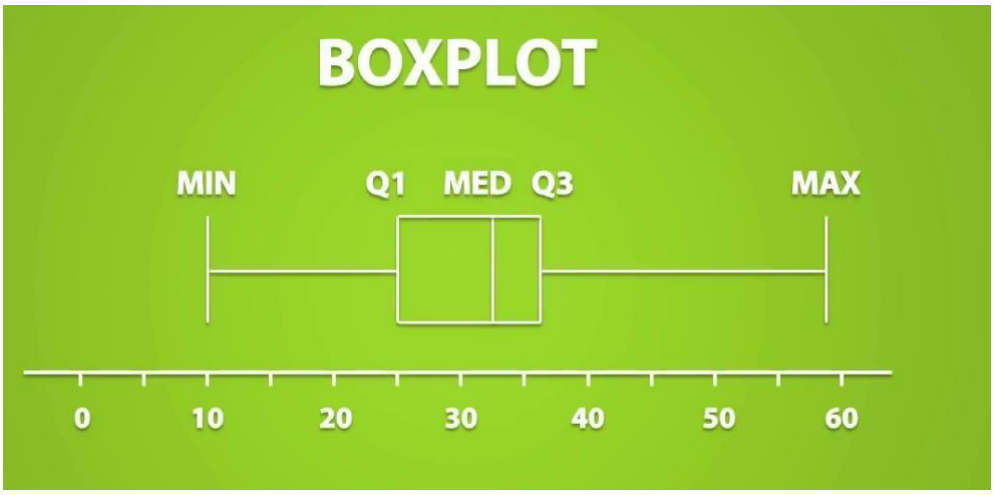


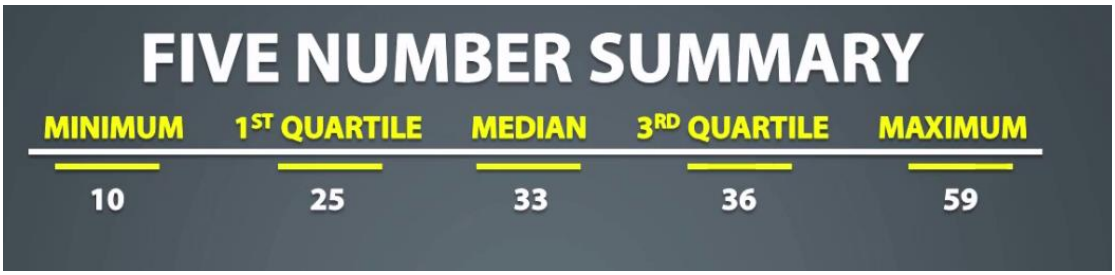
6. Smallest and largest

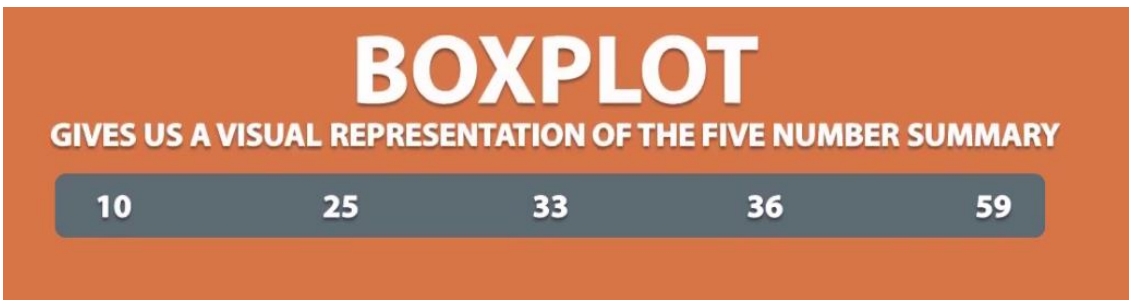


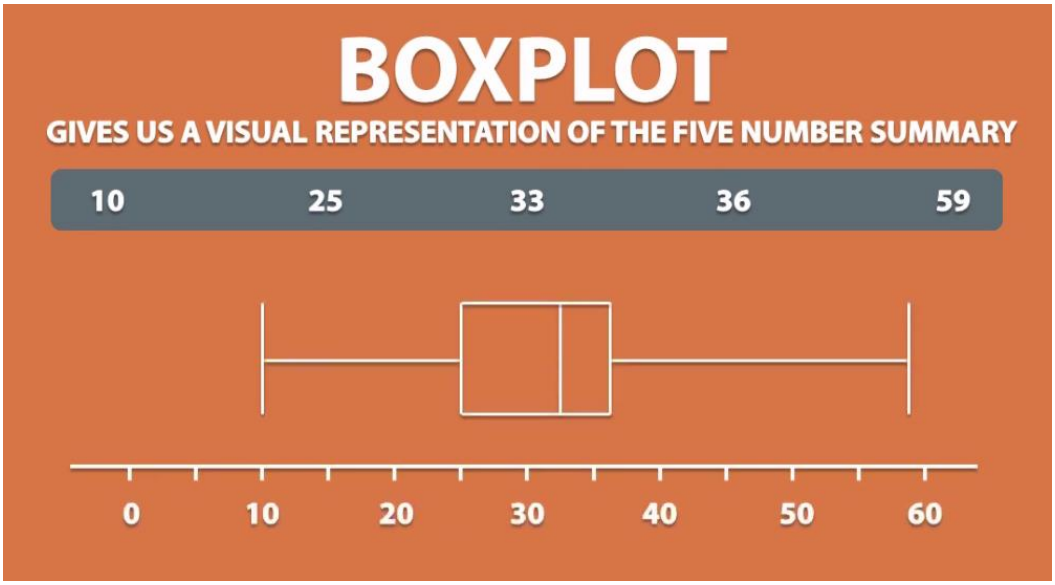
7. Box plot

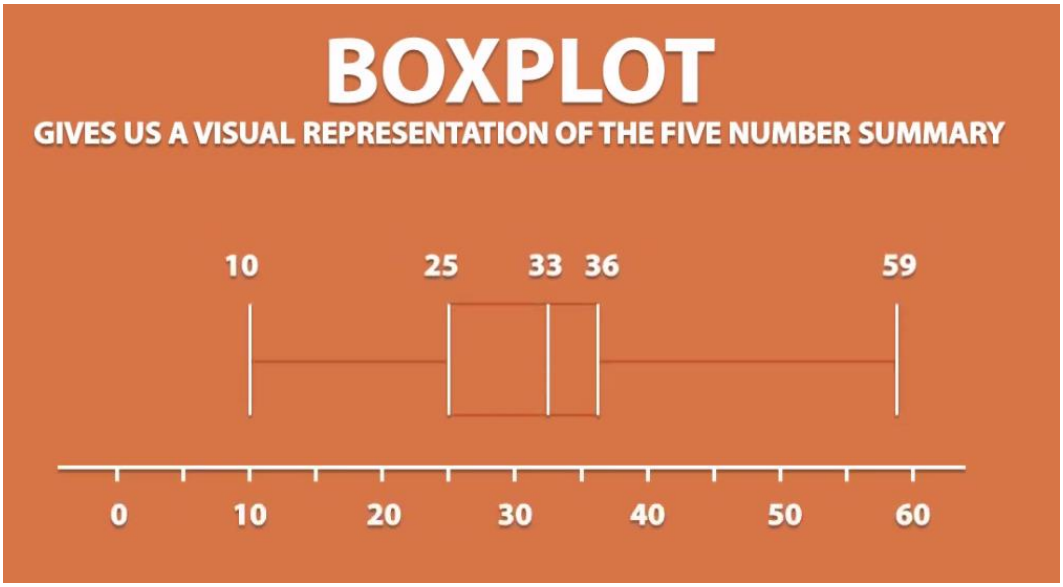
- ✓ We can take above five values summary and create a box plot
- ✓ A box plot gives us visual representation of the five numbers summary





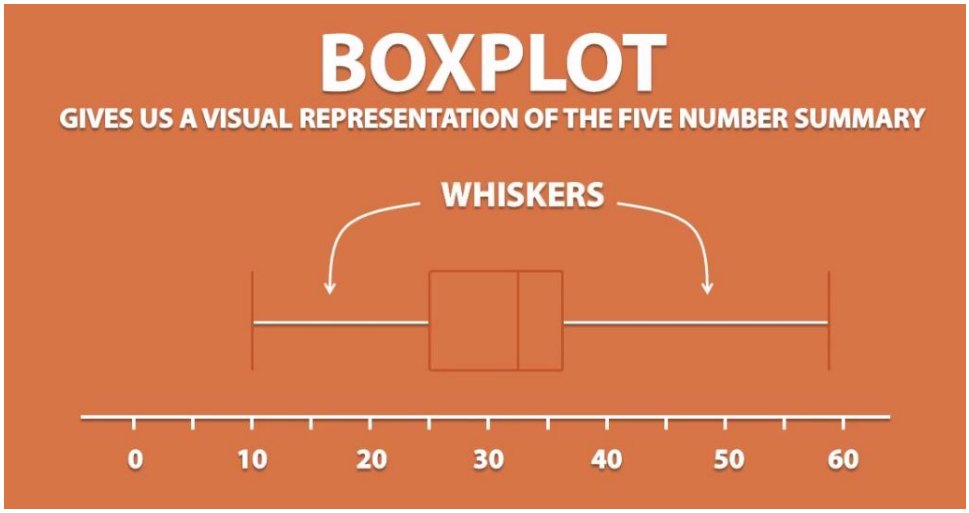



BOXPLOT
GIVES US A VISUAL REPRESENTATION OF THE FIVE NUMBER SUMMARY
10
25
33
36
59
0
10
20
30
40
50
60


BOXPLOT
GIVES US A VISUAL REPRESENTATION OF THE FIVE NUMBER SUMMARY
10
25
33
36
59
0
10
20
30
40
50
60

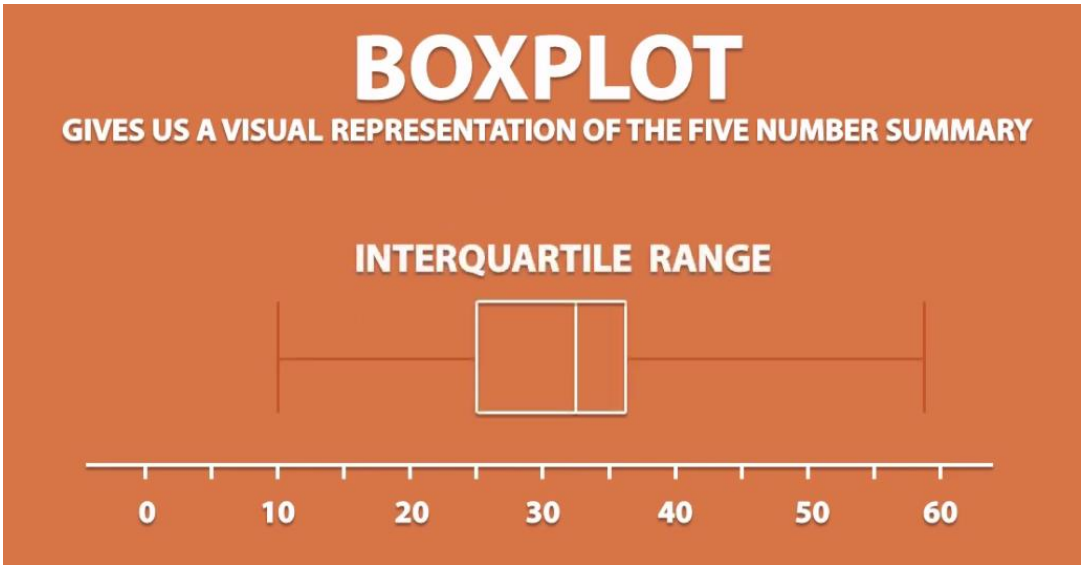
8. Whiskers

- ✓ The horizontal line that extends out from the box is called as whisker

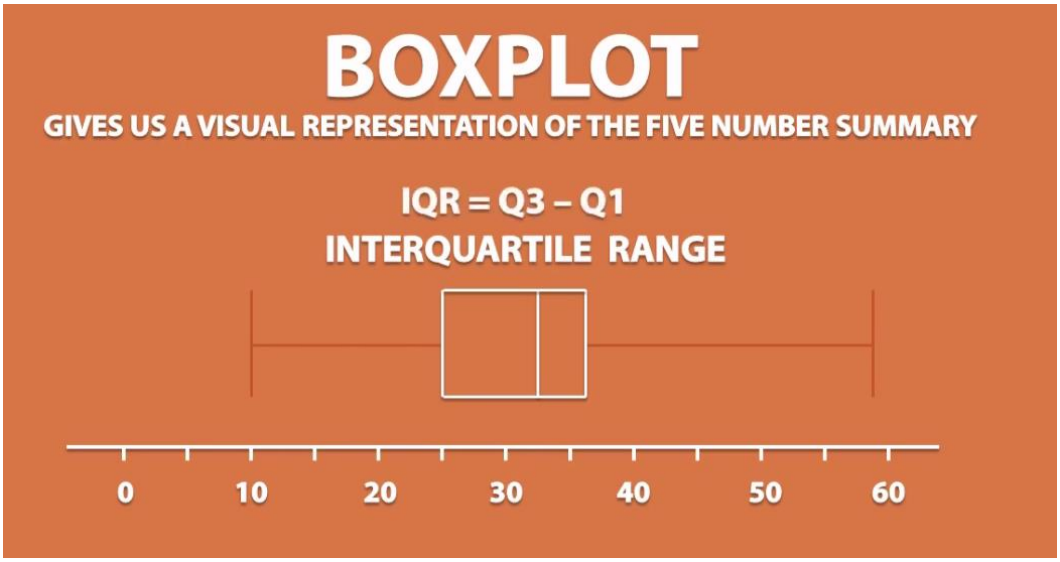


9. Interquartile range

- ✓ The actual box is called as Interquartile range

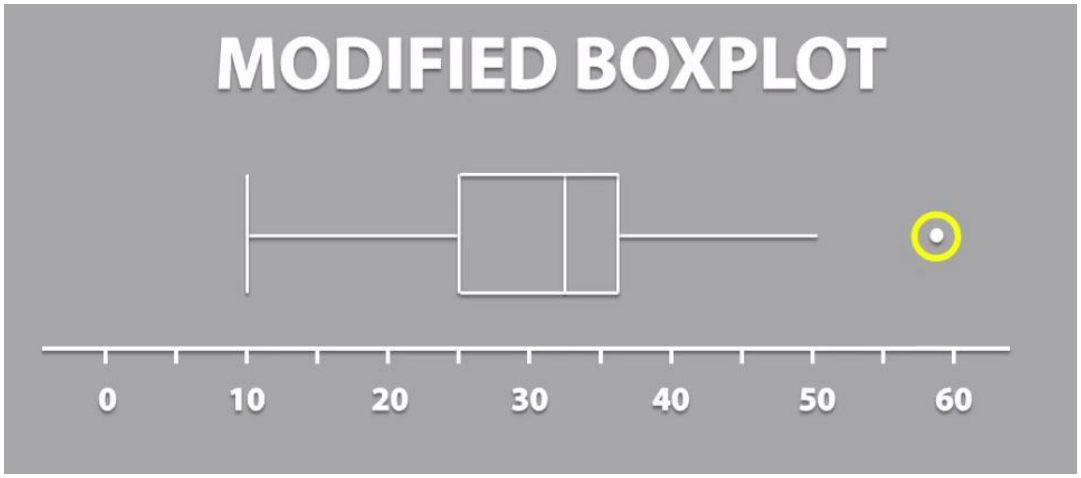


10. IQR = Q3 – Q1

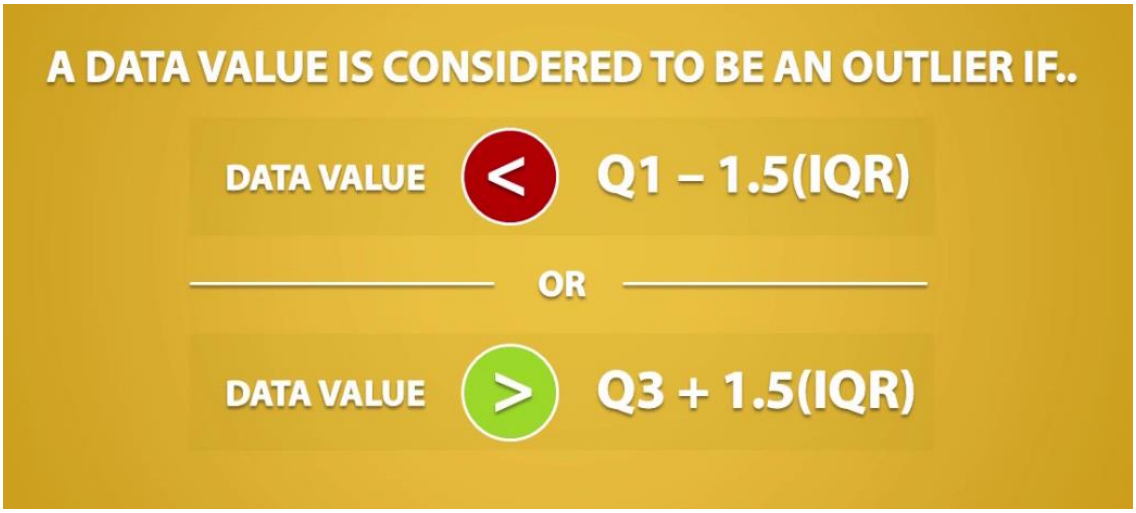

BOXPLOT
GIVES US A VISUAL REPRESENTATION OF THE FIVE NUMBER SUMMARY
IQR = Q3 – Q1
INTERQUARTILE RANGE
0
10
20
30
40
50
60

11. Outliers (modified box plot)

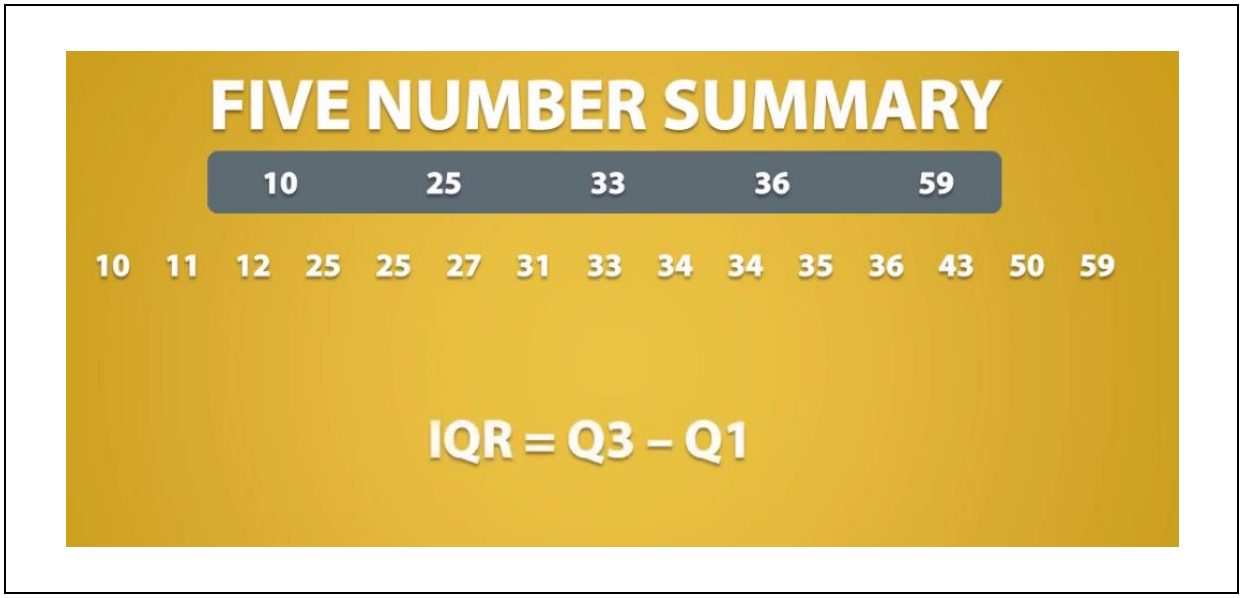
- ✓ A box plot with outlier is called as modified box plot

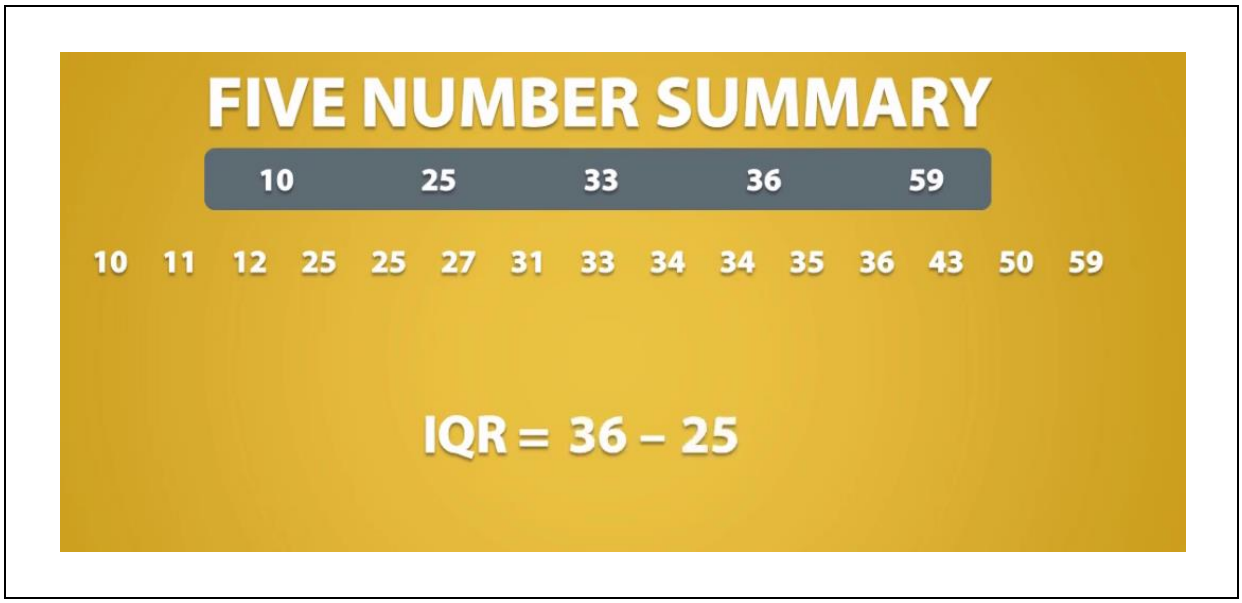


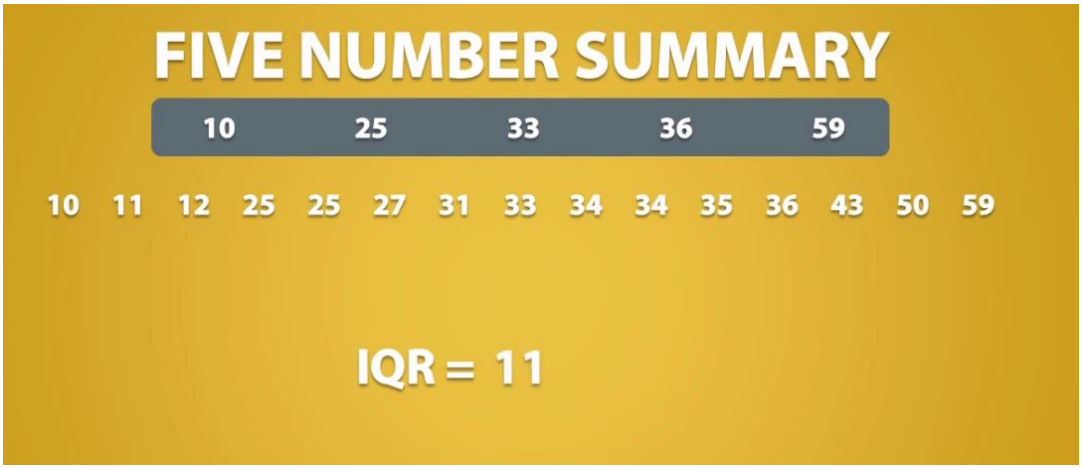
12. Outlier recognizing

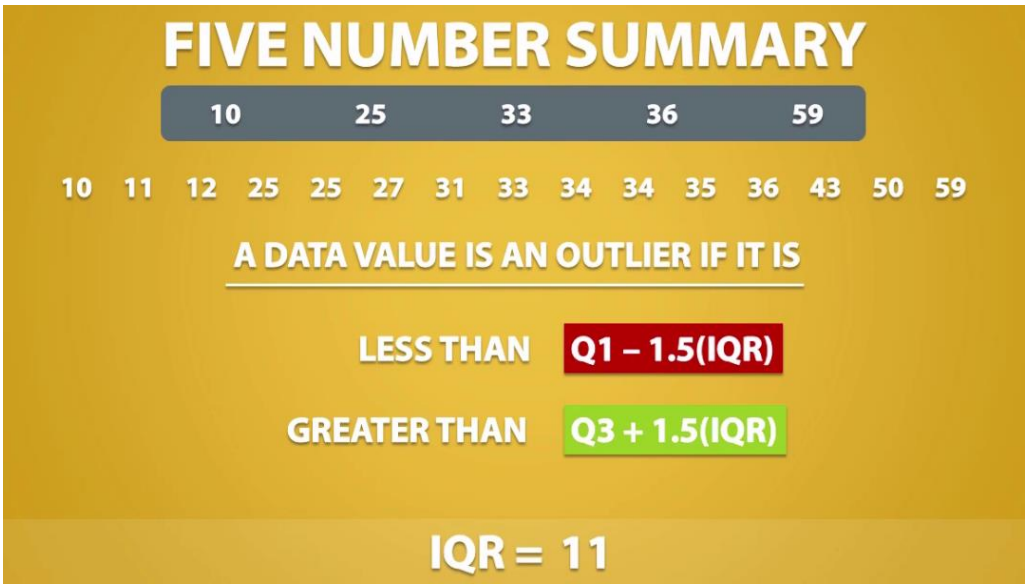


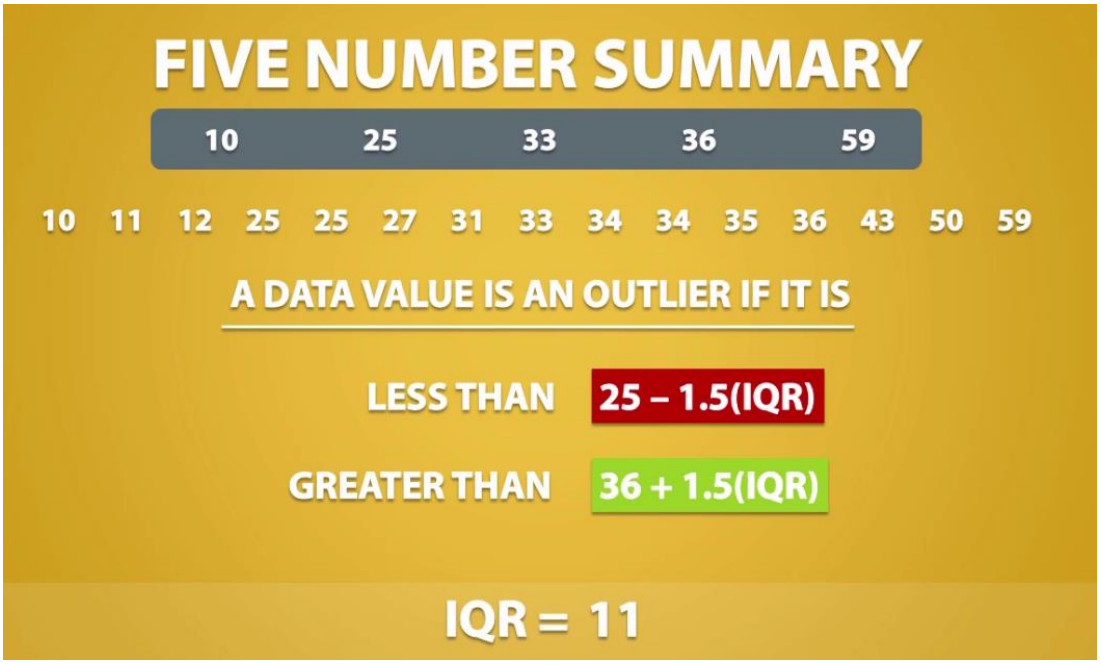
13. Checking outliers

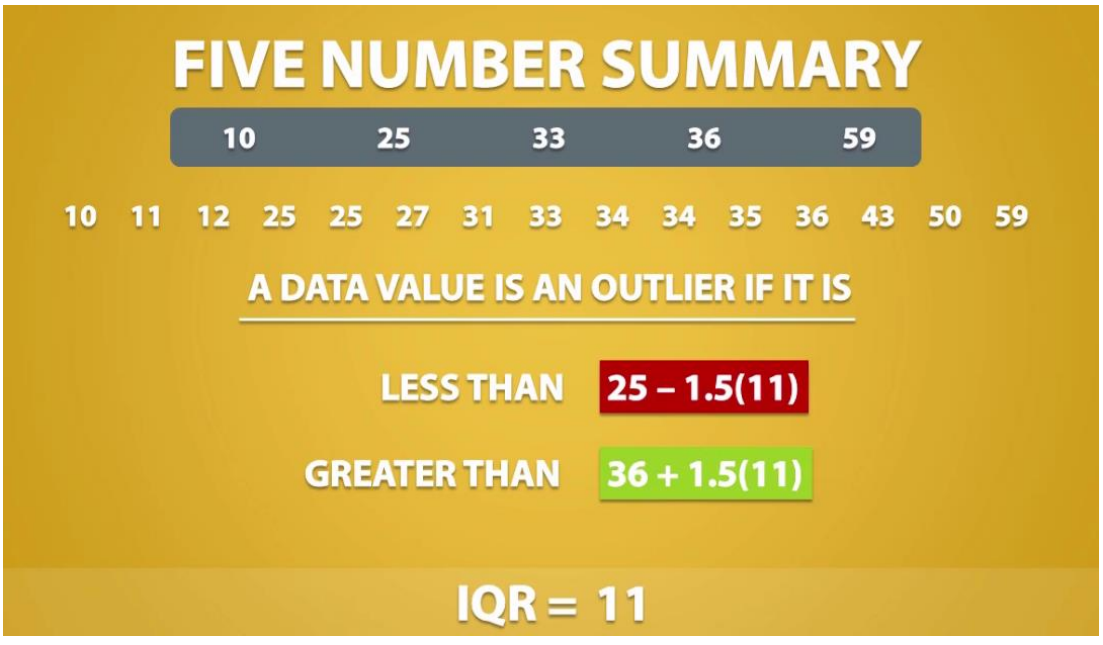


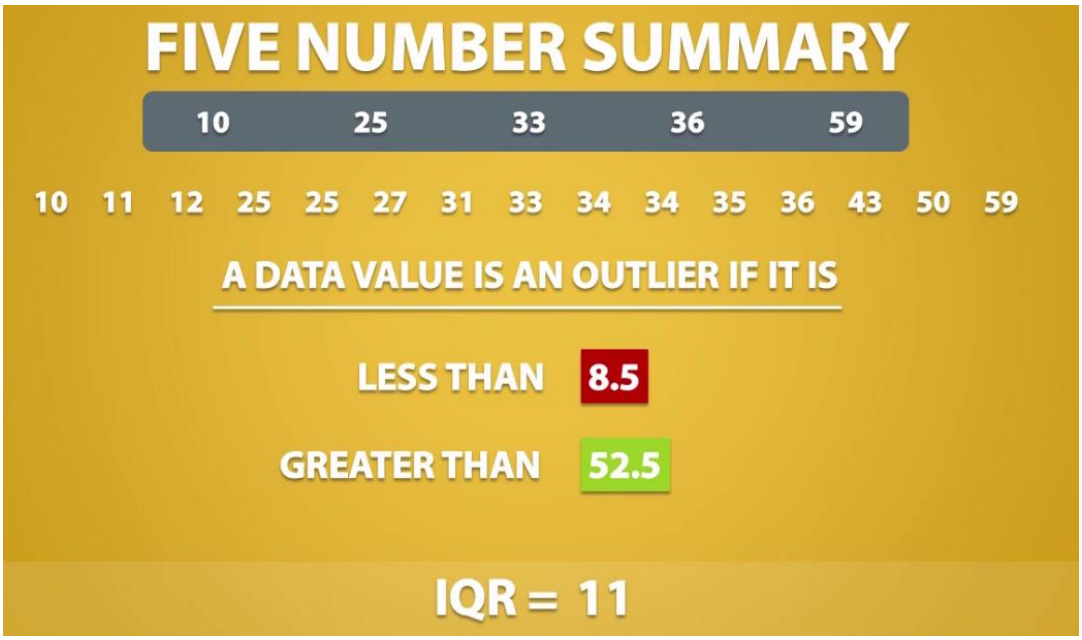


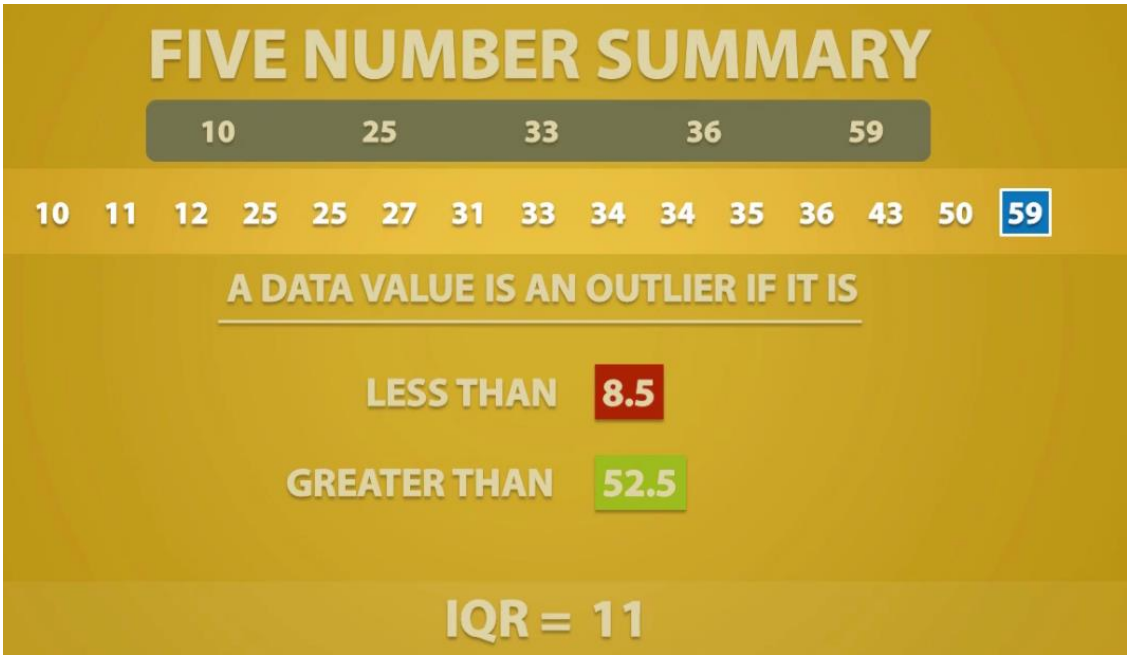

FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
IQR = 11

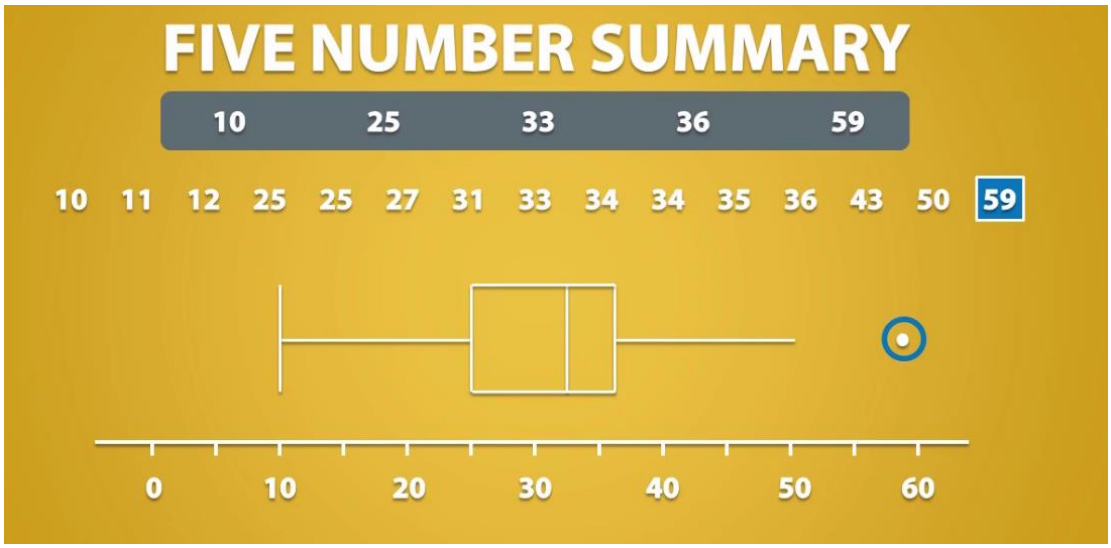

FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
A DATA VALUE IS AN OUTLIER IF IT IS
LESS THAN Q1 – 1.5(IQR)
GREATER THAN Q3 + 1.5(IQR)
IQR = 11


FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
A DATA VALUE IS AN OUTLIER IF IT IS
LESS THAN 25 – 1.5(IQR)
GREATER THAN 36 + 1.5(IQR)
IQR = 11


FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
A DATA VALUE IS AN OUTLIER IF IT IS
LESS THAN 25 – 1.5(11)
GREATER THAN 36 + 1.5(11)
IQR = 11


FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
A DATA VALUE IS AN OUTLIER IF IT IS
LESS THAN 8.5
GREATER THAN 52.5
IQR = 11


FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
A DATA VALUE IS AN OUTLIER IF IT IS
LESS THAN 8.5
GREATER THAN 52.5
IQR = 11


FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
0 10 20 30 40 50 60