Definition 1. For whole numbers P (where $1 \leq P \leq 99$), the P^{th} percentile of a distribution is a value such that P% of the data fall at or below it and (100-P) % of the data fall at or above it.

Example 1. Alex took a general aptitude test and scored in the 91th percentile for aptitude in accounting.

1. What percentage of the scores were at or below her score? **91%**
2. What percentage were above? **9%**

Definition 2 (Interquartile Range (IQR))**.**

$Q1$ = ~~lower quartile~~
= 25th percentile
= median of the lower half of the data

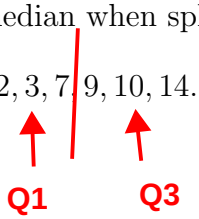
$Q3$ = ~~upper quartile~~
= 75th percentile
= median of the upper half of the data

IQR = how much territory the middle half of the data covers
= Q3 - Q1

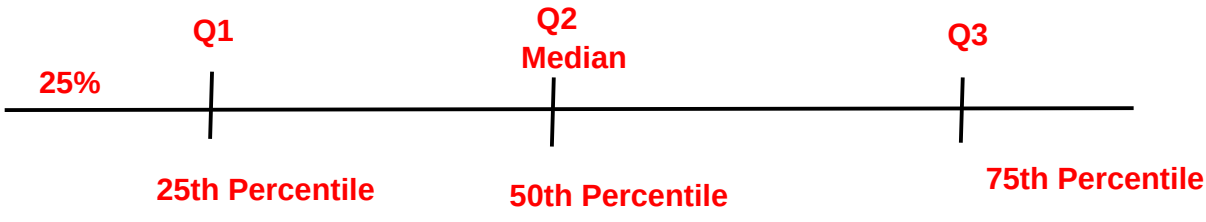
Note that when n is odd, you can omit the median when splitting the data.

Example 2. Suppose we have the datapoints $2, 3, 7, 9, 10, 14$. Find the IQR.

Q1 = 3
Q3 = 10



IQR = Q3 - Q1 = 10 - 3 = 7



Q 1. If we have to report a value for the spread, which value is best and why?

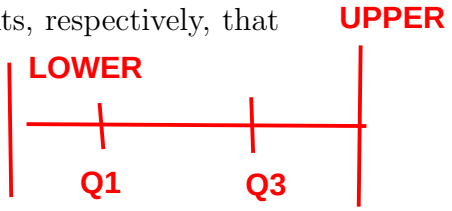
1. If SYMMETRIC with NO OUTLIERS → USE VARIANCE OR STANDARD DEVIATION
2. If SKEWED or WITH OUTLIERS → USE IQR

Definition 3 (Five Number Summary)**.** A summary of values consisting of the minimum, $Q1$, the median, $Q3$, and the maximum values for a dataset.

1 2 3 4 5

Exam Question

Definition 4 (Boxplot)**.** Uses information about the median and IQR to help determine outliers. The middle portion is in the shape of a box, and the lines drawn indicate Q1, the median, and Q3. The whiskers go out to the smallest and largest points, respectively, that are within the lower and upper fences.



$$\text{Lower Fence} = Q1 - 1.5(IQR)$$
$$\text{Upper Fence} = Q3 + 1.5(IQR)$$

Points that fall outside the fences are **outliers**, and are indicated by circles.

- The **upper whisker**: Is the largest point in the dataset inside the upper fence.
- The **lower whisker**: Is the smallest point in the dataset inside the lower fence.

Example 3. Suppose we have the datapoints $2, 3, 7, 9, 10, 14$. Draw a boxplot for this dataset.

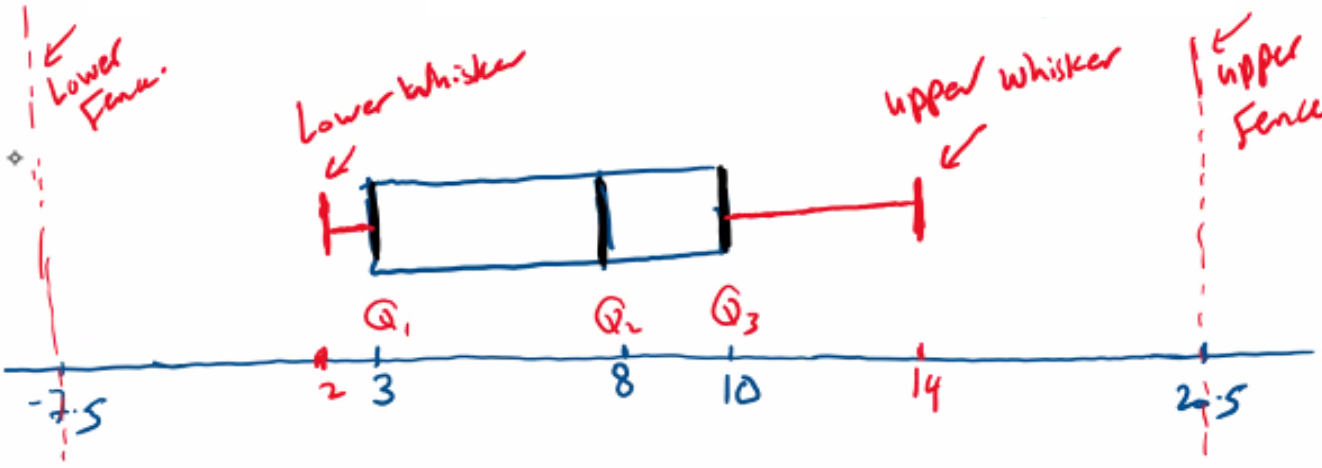
Q1 = 3, Q3 = 10, Q2 = median = $\frac{7+9}{2} = 8$

IQR = Q3 - Q1 = 10 - 3 = 7

Lower Fence = Q1 - 1.5(IQR) = 3 - 1.5(7) = -7.5

Upper Fence = Q3 + 1.5 * IQR = 3 + 1.5(7) = 20.5

If Point is above UPPER FENCE or below LOWER FENCE that point would be an OUTLIER



Example 4 (Car Batteries). The following data specifies the "life" of 40 similar car batteries recorded to the nearest tenth of a year. The batteries are guaranteed to last 3 years. (Data from pg 21 of our textbook.)

1.6	2.6	3.1	3.2	3.4	3.7	3.9	4.3
1.9	2.9	3.1	3.3	3.4	3.7	3.9	4.4
2.2	3.0	3.1	3.3	3.5	3.7	4.1	4.5
2.5	3.0	3.2	3.3	3.5	3.8	4.1	4.7
2.6	3.1	3.2	3.4	3.6	3.8	4.2	4.7

a) Find the following values:

- Sample Mean **(add all numbers and divide by 40)**

$$\text{mean} = \bar{x} = \frac{\Sigma x}{n} = \frac{1.6 + 1.9 + 2.2 + \cdots + 4.7}{40} = \boxed{3.4125}$$

- Sample Median

$$\frac{1}{2}(3.4 + 3.4) = 3.4$$

- Mode **(The number that repeats itself the most)**

  = 3.1

- Range **MAX - MIN = 4.7 - 1.6 = 3.1**

- Q1 **= 3.1**

- Q3

$$\frac{3.8 + 3.9}{2} = 3.85$$

- IQR **Q3 - Q1 = 3.85 - 3.1 = 0.75**

b) Write down the 5-number summary for this dataset.

min	Q1	Q2	Q3	MAX
1.6	3.1	3.4	3.85	4.7

c) Draw a boxplot for this dataset.

Lower Fence = Q1 - 1.5(IQR) = 3.1 - 1.5(0.75) = 1.975

Upper Fence = Q3 + 1.5(IQR) = 3.85 + 1.5(0.75) = 4.975

d) Is there any outlier(s)? Explain

Yes, 1.6 and 1.9 are outliers because they are below the lower fence

Example 5 (YOU TRY). We are given data on the pulse rate/minute for women:

56	66	72	78	83
61	67	73	79	84
62	68	74	81	84
63	69	76	81	88
64	71	77	82	106

13 25

Center
Skewed → median and IQR
Symmetric → mean and standard deviation.

a) Create a Histogram for the data with bins of size 5 starting at 55.

b) What things should we use to summarize this data? (Circle)

Mean (Median) Standard Deviation (IQR)

c) Find the 5-number summary of this dataset. This includes the minimum, Q1, the median, Q3, the maximum.

$Q_2 = 74$

$Q_1 = 67$, $Q_3 = 81$

d) Find the Mean.

e) Find the Range.

f) Find the IQR.

$IQR = Q_3 - Q_1 = 81 - 67 = 14$

g) Find the lower fence for a boxplot.

h) Find the upper fence for a boxplot.

Lower Fence $= Q_1 - 1.5 * IQR$

$= 67 - 1.5(14) = \boxed{46}$

Upper Fence $= Q_3 + 1.5\,IQR = 81 + 1.5(14) = 102$

i) Draw the boxplot for this dataset.

Upper whisker: 88 , lower whisker $= \boxed{56}$