3.4 Exploratory Data Analysis

stem-and-leaf display is one of the exploratory data analysis.
Now we consider another exploratory data analysis (five-number summary)

Five number Summary:

1. Smallest value
2. First quartile (Q_1)
3. Median (Q_2) "second quartile"
4. Third quartile (Q_3)
5. largest value.

Example: (Q_{36} page 106) Consider a sample with data values
27, 25, 20, 15, 30, 34, 28, 25
Provide the five number summary.

first we order data ascending:
15, 20, 25, 25, 27, 28, 30, 34

1. Smallest value 15
5. largest value 34 "must be in the end"

2. Q1 "25th percentile" : $i = \left(\frac{25}{100}\right) 8 = 2$ ⇒ we take the averge of the 2nd and 3rd position

$$Q_1 = \frac{20+25}{2} = 22.5$$

3. Q_2 "50th percentile" : $i = \left(\frac{50}{100}\right) \times 8 = 4$ ⇒ we take the averge of the 4th and 5th position

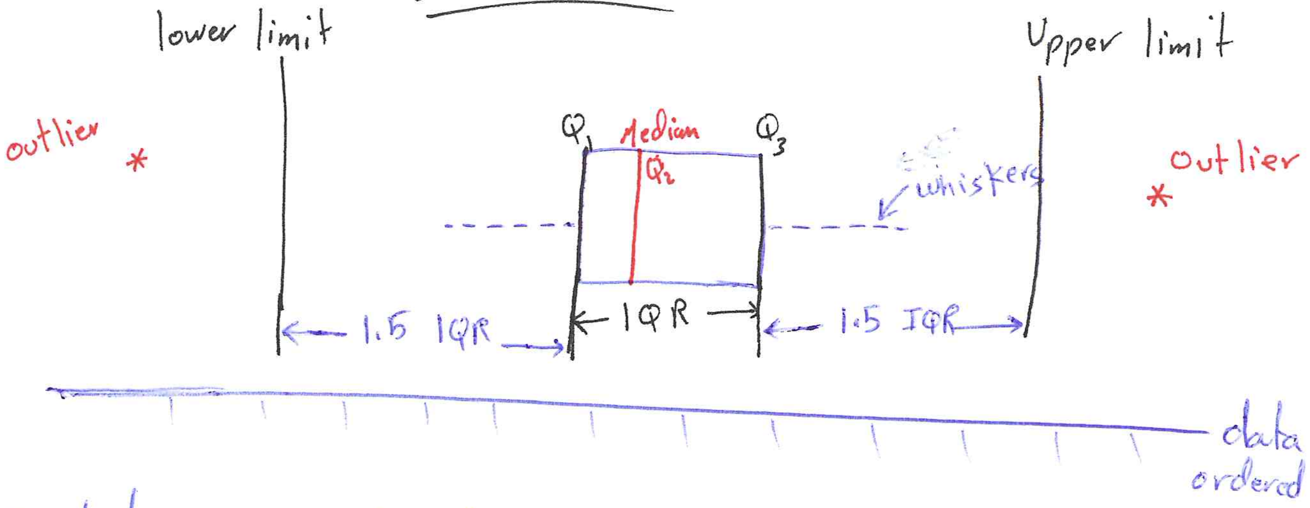
$$Q_2 = \frac{25+27}{2} = 26$$

4. Q_3 "75th percentile" : $i = \left(\frac{75}{100}\right) \times 8 = 6$ ⇒ we take the average of the 6th and 7th position

$$Q_3 = \frac{28+30}{2} = 29$$

The five numbers summary are 15, 22.5, 26, 29, 34

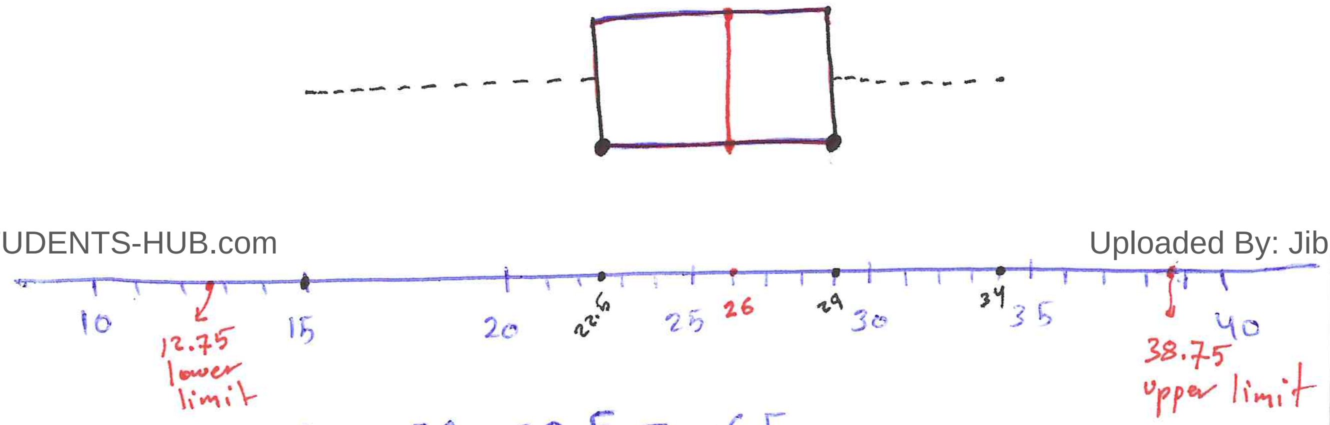
Box Plot



- Box plot is a graphical summary of data that is based on the five number summary.
- Box plot can be used to identify outliers.
- whiskers (dashed lines) are drawn from the ends of the box to the smallest and largest values inside the limits.

Example: (Q37 page 106) show the box plot for the data in Q 36.

Five numbers summary are $\underset{\text{smallest value}}{15}, \overset{Q_1}{22.5}, \underset{\text{Median}}{\overset{Q_2}{26}}, \overset{Q_3}{29}, \underset{\text{largest value}}{34}$



$$IQR = Q_3 - Q_1 = 29 - 22.5 = 6.5$$

To find limits $= 1.5(IQR) = 1.5(6.5) = 9.75$

Upper limit $= Q_3 + 9.75 = 38.75$

lower limit $= Q_1 - 9.75 = 12.75$

> we don't have outliers