8.4 Measures of dispersion

Measures of central tendency (mean, median, and mode) explore the middle of a data set.

When you describe a data set you should give at least of measure of central tendency and one of dispersion.

Measures of dispersion describe the spread of the data around a central value.

The range is the simplest measure of dispersion to calculate.

What is the range? highest value – lowest value

- It can be effected by extreme values.
- It doesn't' tell you how the remaining data is distributed.

Quartiles:

- The median of a set of data separates the data into two halves – half less than the median, half greater.
- **Quartiles** separate the original set of data into four equal parts.
  - Each contains one-quarter (25%) of the data

First quartile	The **first quartile** is the value one-quarter of the way into the data. One quarter of the data lies below the first quartile and three-fourths lies above. It is also called the 25th percentile and often has the symbol Q_1.
Second quartile	The **second quartile** is another name for the median of the entire set of data and is also called the 50th percentile.
Third quartile	The **third quartile** is three-quarters of the way in. Three-fourths of the data lies below the third quartile and one-fourth lies above. It is also called the 75th percentile and has the symbol Q_3.

← it's the median of the bottom half

← it's the median of the top half

$Q_1 = \frac{1}{4}(n+1)$th value and $Q_3 = \frac{3}{4}(n+1)$th value where n is the number of data values in the data set.

You can get a sense of a data set's distribution by examining a **five statistical summary**:

1. minimum
2. 1st Quartile
3. median
4. 3rd Quartile
5. maximum

Show in graphing calc

Here is a list of combined NFL scores for two weeks of the season.

32	56	79	59	23	35	37	24	48	49	37	45	47	50	16	22
35	46	37	44	59	38	39	25	44	44	39	65	41	72	33	40

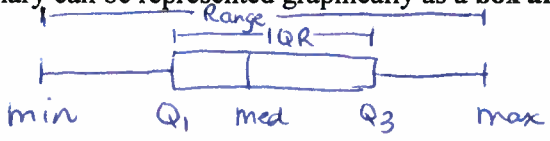
Find the five statistical summary for the data.

min → 16
Q_1 → 35
med → 40.5
Q_3 → 48.5
max → 79

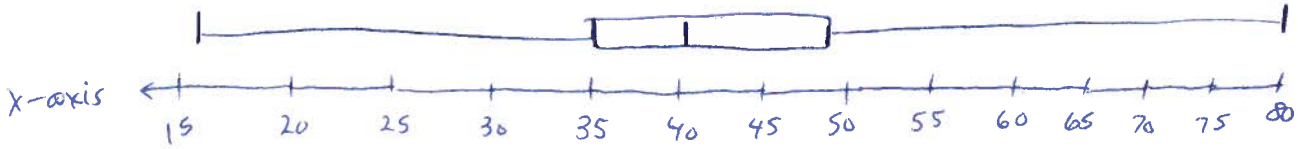
Then difference between the third and first quartile is called the…

interquartile range (IQR) $= Q_3 - Q_1$

A five statistical summary can be represented graphically as a **box and whisker** plot.



Draw a box and whisker plot for the NFL data above.



*Extreme or distant data values are called **outliers**.

An outlier is any value at least 1.5 IQR above Q_3 or below Q_1.

Are there any outliers for the NFL data? $Q_3 - Q_1 = 48.5 - 35 = 13.5$

$Q_1 - 1.5(IQR) = 35 - 1.5(13.5) = 14.75$

$Q_3 + 1.5(IQR) = 48.5 + 1.5(13.5) = 68.75$

2 outliers 72, 79

Exercise 8F