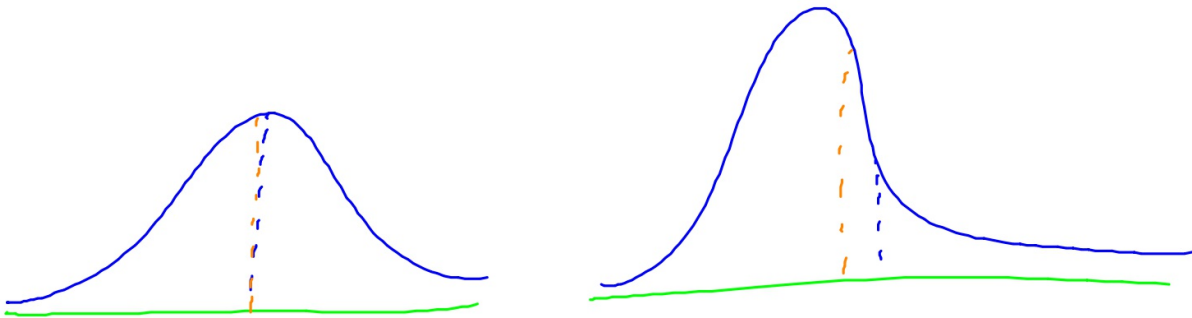


Warm Up



In a graph, when are the mean and median close to each other? When are they not?

1) when the graph is symmetric
2) when the graph is skewed



Objective: SWBAT display data in a boxplot

Agenda:

- Warm Up
- HW Huddle (Ch. 2.2)
- Aside: mean vs. median
- Notes
- Practice
- Reflection

Aside: Mean vs. Median

One advantage of the mean is that if you have multiple samples, the means can combine but the medians can't.

	n	Mean	Median
Sample 1	21	42.6	45.0
Sample 2	26	49.2	48.5

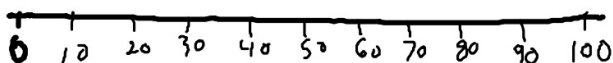
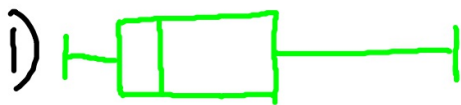
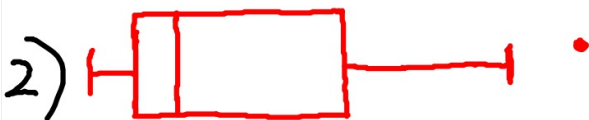
for 47 measurements, $\bar{x} = \frac{42.6(21) + 49.2(26)}{47} \approx 46.3$

Notes: Boxplots

A boxplot is the graphical form of the 5-number summary.

(min, Q₁, median, Q₃, max)

- the box is made from Q₁, the median, and Q₃
- the whiskers extend up to (IQR*1.5) units from the outer quartiles
- outliers (past the whiskers) are marked as dots, stars, etc.
- boxplots *hide information about shape!*



Notes: Standard Deviation

The *sample standard deviation*, s_x , measures the average distance of a data point from the sample mean. It is similar to the population standard deviation, σ . The square of s_x is called the *variance*.

$$s_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$$
$$s_x = \sqrt{\frac{1}{(n-1)} \sum (x_i - \bar{x})^2}$$

the standard deviation is used with the mean, and is *not* resistant to outliers.

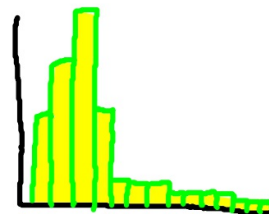
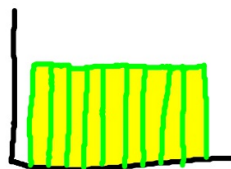
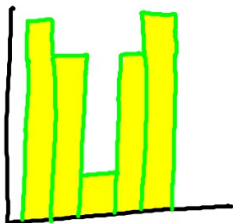
What does it mean?

Large s_x : data is spread out from the mean

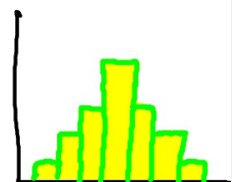
Small s_x : data is close to the mean

rearrange by std. dev.

largest



smallest



Practice

1. Provide a 5-number summary of the following data:

338 318 353 313 318 326 307 317 311 311

307 311 311 313 317 | 318 318 326 338 353

2. Find the IQR and any outliers.

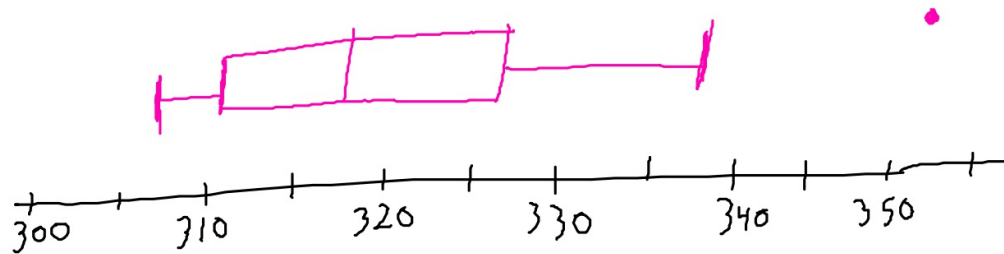
$$IQR = 15$$

$$\text{outliers: } 311 - 22.5 = 288.5$$

$$326 + 22.5 = 348.5$$

3. Draw a boxplot.

min: 307
Q1: 311
med: 317.5
Q3: 326
max: 353



Practice

on our Google class

Reflection

What is an advantage of using the median and IQR instead of the mean & standard deviation? Give a specific example for context.

HW: pg. 71: 90-94