

## 12.2 Analyzing Data

### Measures of Central Tendency

\* any of these can be called an average

	Definition	Best used when...
Mean $\bar{x}$ (sample mean)	Average, add together all data points & divide by # of data points	
Median	The middle data point (ordered from least to greatest)	... there is an outlier
Mode	Most frequent data point	... most of the data is the mode

1) Find the mean, median, and mode for each data set. Then circle which measure of center most accurately represents the data.

a. Carl tracked how many text messages he sent per day for a week.

21, 25, 17, 168, 21, 50, 43  
17, 21, 21, 25, 43, 50, 168

Mean: 49.3

Median: 25

Mode: 21

Calculator

Stat 1: Edit, enter data

Stat → Calc, 1-Var Stat

Enter

b. A sample of 10 employee salaries was taken from a company.

\$45,000 \$50,000 \$75,000 \$50,000 \$50,000 \$50,000 \$50,000 \$50,000 \$112,000 \$50,000 \$52,000  
\$45k \$50k \$75k \$50k \$50k \$50k \$50k \$50k \$52k \$75k \$112k

Mean: \$58,400

Median: \$50,000

Mode: \$50,000

**Outlier:** A point that lies outside most of the data

2) Suppose the values 56, 65, 73, 59, 98, 65, and 59 are the data for the situations below. Would you discard the outlier? Explain.

a. Water temperature of a lake at seven locations

Discard - some freak spot, doesn't reflect lake as whole

b. The number of customers in a restaurant each night in a week

Keep - 98 could be weekend

Identify the outlier of each set of values.

a. 3.4, 4.5, 2.3, 5.9, 9.8, 3.3, 2.1, 3.0, 2.9

b. 17, 21, 19, 10, 15, 19, 14, 0, 11, 16

### Box and Whisker Plots

Box and whisker plots are graphical representations of a data set using the 5-number summary. They give us a more visual representation of the median, the range, and the interquartile range. They can also show what happens to a data set with the inclusion of outliers and the removal of outliers.

5-number summary: Min  $Q_1$  Med  $Q_3$  Max

$Q_1$ : median of lower half of data

$Q_3$ : median of upper half of data

Interquartile range:  $Q_3 - Q_1$

Range: Max - Min

When comparing data, you can talk about: Means, range, median, modes, IQR

3) Make a box and whisker plot to represent the data.

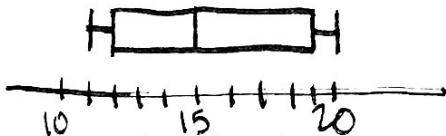
a. 12 11 15 12 19 20 19 14 18 15 16 (x-window)

b. 120 145 133 105 117 150 130 136 128

Calculator: **Stat** 1: Edit, enter data

**2nd** **Y=** **Enter** **ON**

Choose box plot, Adjust window



4) Make two box and whisker plots to represent the average monthly temperature in two cities. Then use the box and whisker plots to compare the data.

L1	Jacksonville, Florida	52.4	55.2	61.1	67.0	73.4	79.1	81.6	81.2	78.1	69.8	61.9	55.1
L2	Austin, Texas	48.8	52.8	61.5	69.9	75.6	81.3	84.5	84.8	80.2	71.1	60.9	51.6

Florida  $\bar{x}$ :  $68.0^\circ$

Median:  $68.4$

(Used calculator to compare stat plots)

Texas  $\bar{x}$ :  $68.6^\circ$

Median:  $70.5$

Texas has higher mean & median temperature, but has a larger range of temperatures than Florida.

A percentile is a number from 0 to 100 that shows the percentage of the data that is less than or equal to a value in the data. For example, if  $x$  is at the 63<sup>rd</sup> percentile, then 63% of the data are less than or equal to  $x$ .

Here is an ordered list of midterm test scores for a Spanish class. Circle the value of the 65<sup>th</sup> percentile.

41 54 61 65 67 73 74 77 77 77 79 80 (82) 88 89 93 97 98 98 100

Total: 20  
Each value is 5% of data ( $\frac{100}{20}$ )

Find the values at the 30<sup>th</sup> and 90<sup>th</sup> percentiles for each data set.

a) 6283 5700 6381 6274 5700 5896 5972 6075 5993 5581

Total: 10, each is 10%

30<sup>th</sup>: 5700 90<sup>th</sup>: 6283

b) 7 12 3 14 17 20 5 3 17 4 13 2 15 9 15 18 16 9 1 6

Sort Data  
**Stat** 1: Edit, enter data

**Stat** 2: Sort A

**2nd** **|** **Enter**