9.A – Analyzing Data

- **Measures of Central Tendency:**
  - Mean: average value
  - Medium: middle value
  - Mode: most frequency occurring value(s). Bimodal means data has two modes.
- **Range:** the difference between the greatest and least values given a set of data.
- **Quartiles:** the four parts that make up the range
- **Interquartile range:** the difference between the first and third quartiles.
- **Outlier**: a value that is substantially different from the rest of the data that can be misleading and affect the measure of central tendency.
- **Percentile**: a number from 0 to 100 that you can associate with a value x from a data set. It show the percent of the data that are less than or equal to x. If x is at the 63rd percentile, then 63% of the data are less than or equal to x.

Finding Measures of Central Tendency

Career **The frequency table shows the number of job offers received by each student within two months of graduating with a mathematics degree from a small college. What are the mean, median, and mode for the job offers per student?**

Job Offers	0	1	2	3	4
Students	2	2	4	5	2

Mean: $\overline{x} = \frac{2(0) + 2(1) + 4(2) + 5(3) + 2(4)}{15}$

$= \frac{33}{15} = 2.2$

The symbol $\overline{x}$, read "*x bar*," represents the mean.

The mean is 2.2.

Median: 0, 0, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4
The median is 2.

List each value the number of times it occurs. Arrange them in order. Find the middle value.

Mode: Five students received 3 job offers each.
The mode is 3.

The mode is the number of job offers received by most students.

Finding Percentiles

Testing **Here is an ordered list of midterm test scores for a Spanish class. What value is at the 65th percentile?**

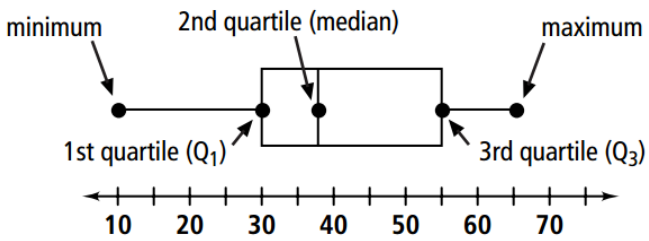
41	54	61	65	67	73	74
77	77	77	79	80	82	88
89	93	97	98	98	100	

Of the 20 values, 65% fall at or below the value at the 65th percentile.

$20 \cdot 65\% = 20 \cdot 0.65 = 13$

13 values fall at or below 82, the value at the 65th percentile.

Box and Whisker Plot



Calculating Outliers: 1.5 X IQR Rule

- Suspected **low outlier**: any value $< Q_1 - 1.5 \times IQR$
- Suspected **high outlier**: any value $> Q_3 + 1.5 \times IQR$

Comparing Data Sets

Temperature: The tables shows average monthly water temperatures for four locations on the Gulf of Mexico. How can you compare the 12 water temperatures from St. Petersburg with the 12 water temperatures from Key West?

Gulf of Mexico Eastern Coast Water Temperatures (°F)

Location	J	F	M	A	M	J	J	A	S	O	N	D
St. Petersburg, Florida	62	64	68	74	80	84	86	86	84	78	70	64
Key West, Florida	69	70	75	78	82	85	87	87	86	82	76	72
Dauphin Island, Alabama	51	53	60	70	75	82	84	84	80	72	62	56
Grand Isle, Louisiana	61	61	64	70	77	83	85	85	83	77	70	65

Source: National Oceanographic Data Center

St. Petersburg:

$$\bar{x} = \frac{62 + 64 + 64 + 68 + 70 + 74 + 78 + 80 + 84 + 84 + 86 + 86}{12}$$

$$= \frac{900}{12} = 75 \text{ (mean water temperature)}$$

Modes: 64, 84, and 86

Min.: 62; Max.: 86; Range: $86 - 62 = 24$

Median (Q_2) = 76

62 64 64 68 70 74 78 80 84 84 86 86

Median of lower part (Q_1) = 66

Median of upper part (Q_3) = 84

Interquartile range:
$Q_3 - Q_1 = 84 - 66 = 18$

Key West:

$$\bar{x} = \frac{69 + 70 + 72 + 75 + 76 + 78 + 82 + 82 + 85 + 86 + 87 + 87}{12}$$

$$= \frac{949}{12} \approx 79.1 \text{ (mean water temperature)}$$

Modes: 82 and 87

Min.: 69; Max.: 87; Range: $87 - 69 = 18$

Median (Q_2) = 80

69 70 72 75 76 78 82 82 85 86 87 87

Median of lower part (Q_1) = 73.5

Median of upper part (Q_3) = 85.5

Interquartile range:
$Q_3 - Q_1 = 85.5 - 73.5 = 12$

The range and the interquartile range show the temperatures varying less at Key West than at St. Petersburg. Also, the temperatures at Key West are generally higher.

Practice

1) Find the mean, median, and mode of each set of values

a. Time spent on Internet per day (in minutes): ~~75~~, ~~68~~, ~~43~~, 120, ~~65~~, 180, ~~95~~, 225, 140.

43 65 68 75 95 120 140 180 225

No mode

↑ median

$\frac{43+65+68+75+95+120+140+180+225}{9} \approx 112.\overline{3}$

↗ mean

b.

Age (years)	13	14	15	16	17	18	19
Frequency	7	12	18	9	5	4	2

$\frac{13(7)+14(12)+15(18)+16(9)+17(5)+18(4)+19(2)}{57} \approx 15.23 = \text{mean}$

mode = 15

median = 15 → middle # if all 57 teenagers are listed out least to greatest.

2) The table shows the average monthly temperatures of two cities. How can you compare the temperatures?

	J	F	M	A	M	J	J	A	S	O	N	D
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
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
mean = 67.99
median = 68.4
min = 52.4
max = 81.6
range = 29.2

Texas
mean = 68.58
median = 70.5
min = 48.8
max = 84.8
range = 36

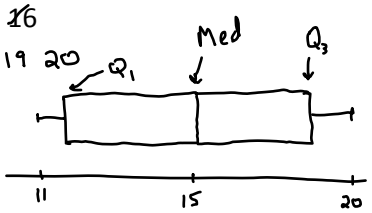
The measures of mean and median are similar, but Texas has a noticeably larger range of temperatures compared to Florida.

3) Make a box and whisker plot for each set of values:

a. 12, 11, 15, 12, 19, 20, 19, 14, 18, 15, 16

11 12 12 14 15 15 16 18 19 19 20

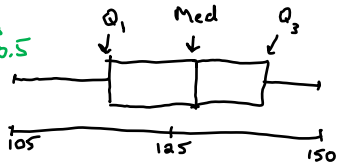
Q_1 = 12 (median Q_2 = 15, Q_3 = 19)



b. 120, 145, 133, 105, 117, 150, 130, 136, 128

105 117 120 128 130 133 136 145 150

Q_1 = 118.5, Med Q_2 = 130, Q_3 = 140.5



4) Find the outlier(s), if any, of the following data sets:

a. 50, 60, 73, 77, 80, 81, 82, 83, 84, 84, 84, 85, 88, 95, 100

$Q_1 = 77$, Q_2, $Q_3 = 85$, $IQR = 8$

Outlier $< Q_1 - 1.5(IQR)$
$< 77 - 1.5(8)$
Outlier < 65

Outlier $> Q_3 + 1.5(IQR)$
$> 85 + 1.5(8)$
Outlier > 97

Outliers: 50, 60, 100

b. 10.2, 14.1, 14.4, 14.4, 14.4, 14.5, 14.5, 14.6, 14.7, 14.7, 14.7, 14.9, 15.1, 15.9, 16.4

$Q_1 = 14.4$, Q_2, $Q_3 = 14.9$, $IQR = 0.5$

Outlier $< Q_1 - 1.5(IQR)$
$< 14.9 - 1.5(0.5)$
Outlier < 13.65

Outlier $> Q_3 + 1.5(IQR)$
$> 14.9 + 1.5(0.5)$
Outlier > 15.65

Outliers: 10.2, 15.9, 16.4

5) Find the values at the 30th and 90 percentile for each data set.

a. 6283, 5700, 6381, 6274, 5700, 5896, 5972, 6075, 5993, 5581

5581, 5700, 5700, 5896, 5972, 5993, 6075, 6274, 6283, 6381

10 terms (.30) = 3rd term
5700 is at the 30th percentile

10 terms (.90) = 9th term
6283 is at the 90th percentile

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

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

20 terms (.30) = 6th term
5 is at the 30th percentile

20 terms (.90) = 18th term
17 is at the 90th percentile