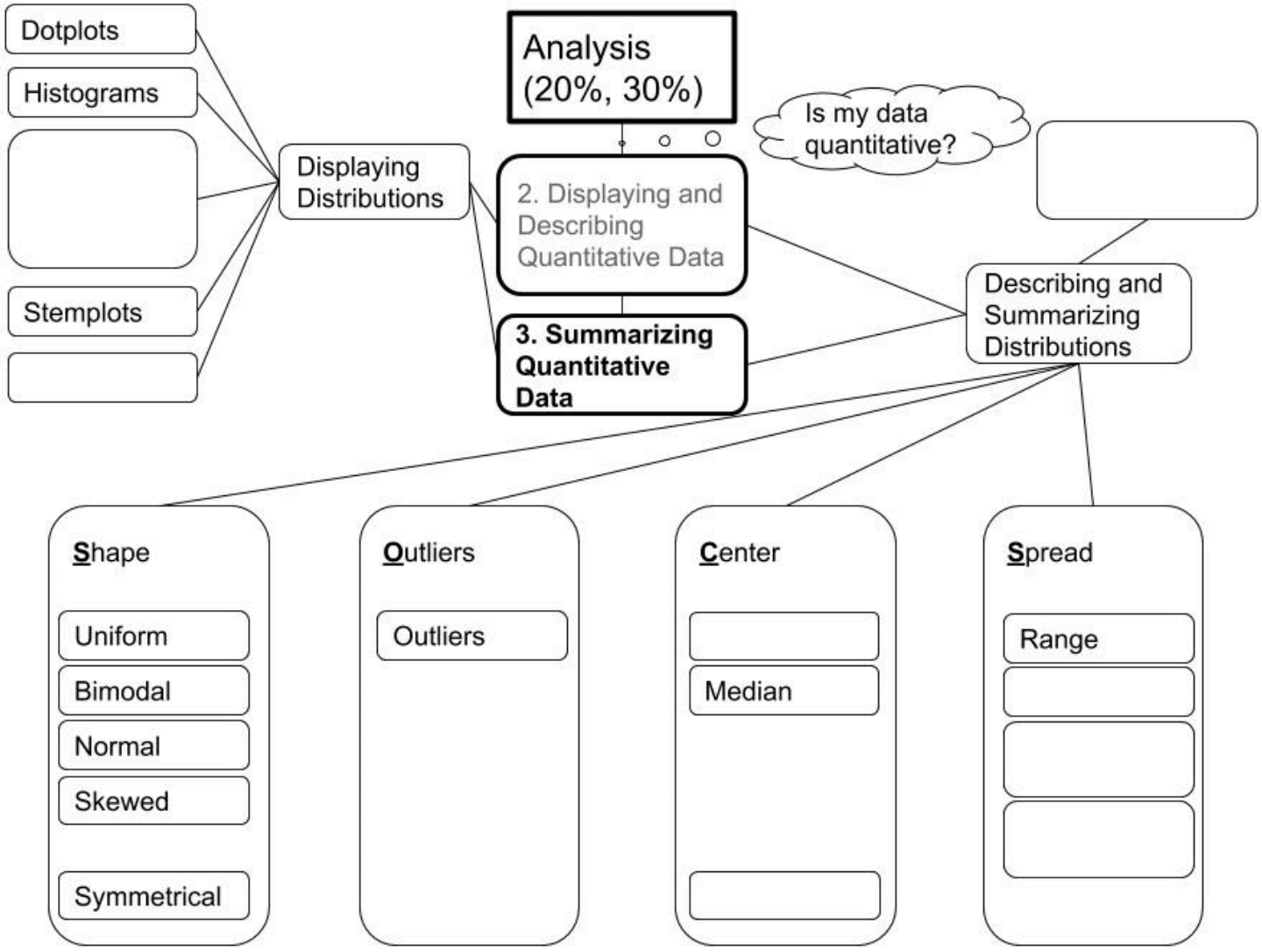
Unit 03: Summarizing Quantitative Data

Unit Objectives

- Gain additional parameters for describing center and spread of distributions
- Draw box plots
- Identify outliers



Unit 03 Lesson 01: Calculate Means

What is a second way of describing the center of a distribution? [Source: Khan Academy]

- Calculate means of data sets
- Calculate a missing value given the mean

1.

The following table shows the distance from Tessa's house to her favorite locations.

Location	School	Library	Park	Movie theater
Distance from Tessa's house (in kilometers)	3	5	4	4

Find the mean distance.

2.

Anduin casts a variety of healing spells. The number of health points that each of his healing spells restores is listed below.

$$43, 57, 30, 18, 26, 12$$

Find the mean number of health points.

3.

The following table shows the number of innings pitched by each of the Greenbury Goblins' starting pitchers during the Rockbottom Tournament.

Pitcher	Calvin	Thom	Shawn	Kris	Brantley
Number of innings pitched	11	12	7	3	?

If the mean of the data set is 8 innings, find the number of innings Brantley pitched.

~~~U03L01 Homework~~~

1. Summarizing quantitative data: Calculating the mean
2. Summarizing quantitative data: Missing value given the mean

~~~U03L01 Classwork~~~

1.

The following data points represent the volume of gas in each race car driver's tank (in liters).

$$4, 5, 13, ?$$

If the mean of the data set is 7 L, find the missing volume of gas.

2.

Farmer Holstein weighed each of his cows. He found that Jo weighs 711 kilograms, Marty weighs 533 kilograms, and Jennifer weighs 664 kilograms.

Find the mean weight.

3.

Guilherme observed the number of minutes his dormmates spent on social media sites while they were at the library. He reported his data in the following table.

Dormmate	Number of minutes on social media
Vera	13
Timothy	0
Manueala	14
Khaled	36
Yash	18
Fernanda	12

Find the mean number of minutes on social media.
~~~

4. **AP**: A dog breeder has computed the mean weight for both male and female dogs that are either Bulldogs or Retrievers within his kennel. The results are displayed in the table below in pounds. When males and females are grouped together, could Bulldogs have a higher mean weight than Retrievers?

	Bulldogs	Retrievers
Male	25.3	28.9
Female	31.3	33.6
Overall	?	?

A. No. The bulldogs mean weight for males and females must be 28.3 pounds, while the Retrievers mean weight must be 31.25 pounds.
B. No. There is not enough information to determine the mean weight for each breed, but it must be higher for retrievers than for bulldogs.
C. Yes. It could happen. Whether it does happen depends on the number of male and female dogs in each breed.
D. Yes. It could happen. Whether it does happen depends on the variability of the weights within each of the four dog groups.
E. Yes. It could happen. Whether it does happen depends on the shapes of the distributions of the weights within each of the four dog groups.

Unit 03 Lesson 02: Calculate Medians and Means from Data Displays

What if instead of being given a list of data, I'm given a data display?

- Calculate medians and means using charts

1.

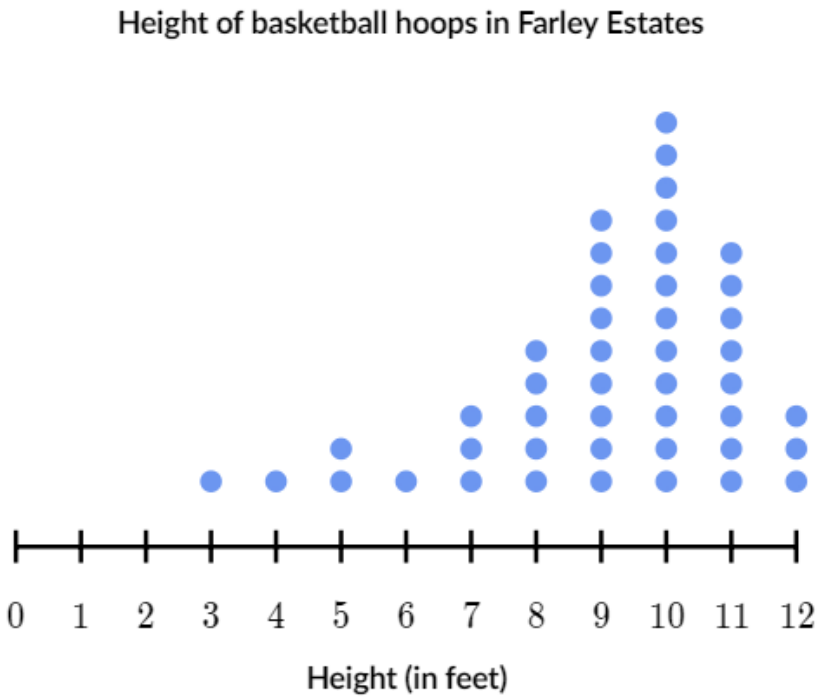
Many people require more than one attempt to pass the **Unit 1** test for a **certain teacher's class**
The dot plot gives the number of times 21 people took the **Unit 1** test in order to earn their **desired grade.**

a. Describe the shape of the distribution.

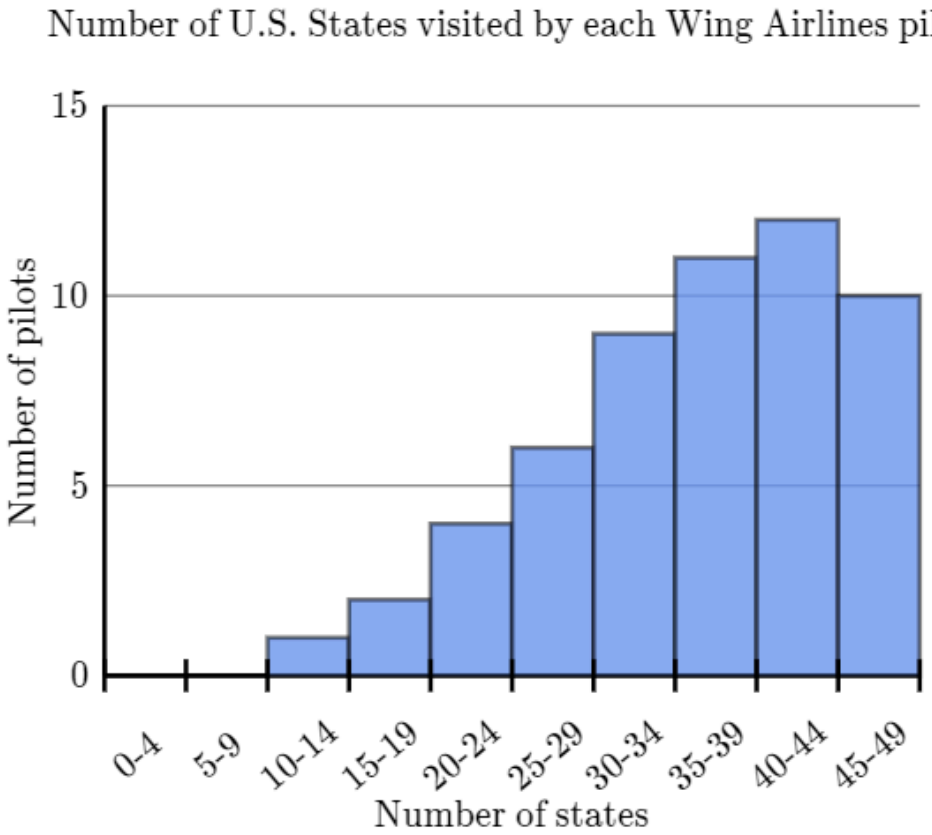
b. Calculate the mean and median of the distribution.

c. OOOPS. The 6 was actually supposed to be a 50. Calculate the new mean and median.

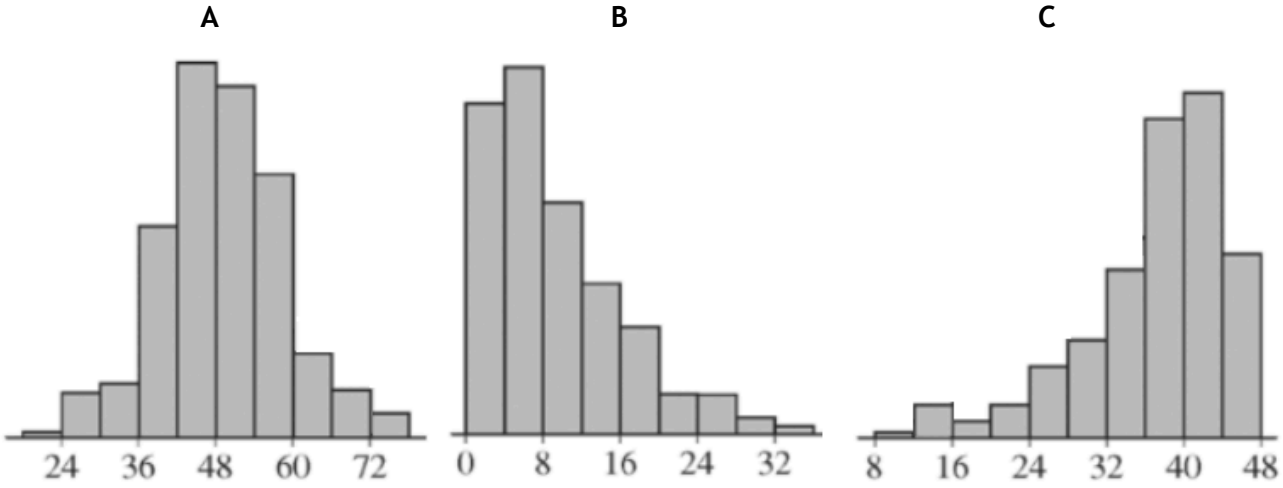
~~~U03L02 Homework~~~ 1. Summarizing quantitative data: Calculating mean and median from data displays

~~~U03L02 Classwork~~~

1.



a. Describe the shape of the distribution.

b. Calculate the mean and median of the distribution.

c. Oddly enough, someone bought a new hoop that is -30 feet tall. Calculate the new mean and median.

d. Using your observations from the notes and classwork, fill in the blanks.

Positively skewed distributions	Mean is always _____ than the median. Altering a data point usually affects the mean _____ than the median.
Negatively skewed distributions	Mean is always _____ than the median. Altering a data point usually affects the mean _____ than the median.
~~~

2. For the distribution below, which is greater, the median or the mean?



3. **AP**: For the 3 histograms above, which of the following correctly orders the histograms from the one with the smallest proportion of data above its mean to the one with the largest proportion of data above its mean.



a. A, B, C
b. A, C, B
c. B, A, C
d. B, C, A
e. C, A, B
f. C, B, A

Unit 03 Lesson 03: Calculate Variances and Standard Deviations

What is a more meaningful way to describe spread in distributions? [Source: Khan Academy]

- Calculate variances
- Calculate standard deviations

1.

A veterinarian weighed a sample of 6 puppies. Here are each of their weights (in kilograms):

$1, 2, 7, 7, 10, 15$

a. What kind of distribution is this?

b. Calculate the variance of the distribution.

c. Calculate the standard deviation of the distribution

2. Last year, Mr. Rose had 8 students in his AP Calculus class. Here are their scores for the AP test.
   2 1 2 2 2 2 3 1

   a. What kind of distribution is this?

   b. Calculate the variance of the distribution.

   c. Calculate the standard deviation of the distribution.

3.

Each dot plot below represents a different set of data.

A **B** **C**

Which of the following correctly orders the histograms from the one with the smallest standard deviation to the one with the largest standard deviation?

   a. A, B, C
   b. A, C, B
   c. B, A, C
   d. B, C, A
   e. C, A, B
   f. C, B, A

~~~U03L03 Homework~~~

1. Summarizing quantitative data: Sample standard deviation
2. Summarizing quantitative data: Visually assessing standard deviation

~~~U03L03 Classwork~~~

1.

Jin recorded the ages (in years) for a sample of 4 zebras from a wildlife preserve. Here are the ages:

$2, 9, 14, 15$

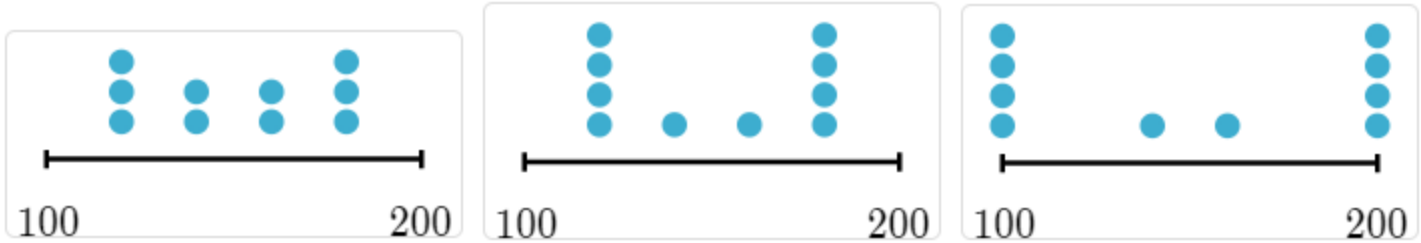
a. What kind of distribution is this?

b. Calculate the variance of the distribution.

c. Calculate the standard deviation of the distribution

2.

Each dot plot below represents a different set of data.

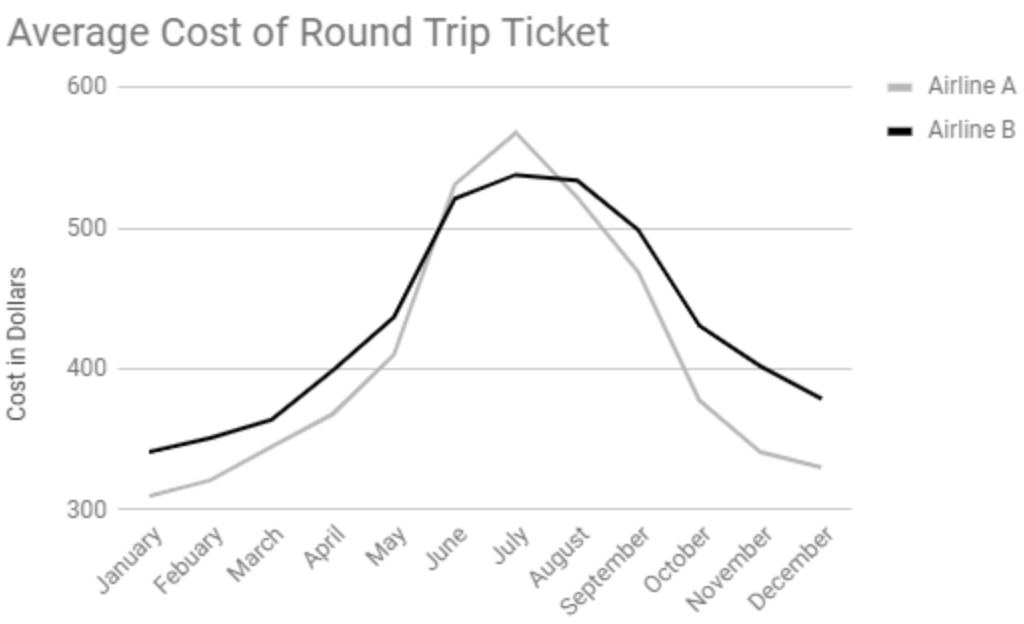


A **B** **C**

Which of the following correctly orders the histograms from the one with the smallest standard deviation to the one with the largest standard deviation

a. A, B, C
b. A, C, B
c. B, A, C
d. B, C, A
e. C, A, B
f. C, B, A

3. **AP**



A. Based on the graph, compare the two airlines with respect to the monthly average prices over the year. Address both similarities and differences in the patterns.
Similarities:

Differences:

B. For which of the two airlines is the standard deviation of the 12 monthly average prices greater? Justify your answer without performing any calculations.

4. **AP**: A meteorologist determines the mean and standard deviation of the number of rainy days in Chowchilla for each of the last ten years. Which of the following is the best description of the standard deviation?
    A. Approximately the median distance between the number of rainy days in the individual years and the mean number of rainy days of the last ten years.
    B. The distance between the greatest number of rainy days in the individual years and the mean number of rainy days in the last ten years.
    C. The number of days separating the fewest rainy days and the most rainy days when considering each year in the last ten years.
    D. The number of days separating the fewest rainy days and the most rainy days when considering the middle 50 percent of the distribution.
    E. Approximately the mean distance between the number of rainy days in the individual years and the mean number of rainy days of the last ten years.

Unit 03 Lesson 04: Effects of Shifting, Adding, or Removing a Data Point

How are the key aspects of a distribution influenced by a single data point?

- Compare the reactions of different parameters to the altering of data [Source: Khan Academy]

1.

Chucky grabbed 11 items in the grocery store that each had a different price and had a mean cost of about $\$4.44$. On his way to the register, he gave in to an impulse to add a 12^{th} item: an entire wheel of cheese that cost $\$39.99$.

Item	Cost
elephant garlic	$\$1.29$
Italian parsley	$\$1.92$
Kosher salt	$\$3.19$
bulgur wheat	$\$3.79$
matzoh meal	$\$3.99$
Anjou pears	$\$4.79$
tahini	$\$5.19$
mozzarella pearls	$\$5.29$
clam stock	$\$5.49$
rose water	$\$6.75$
beef tongue	$\$7.19$
gorgonzola cheese wheel	$\$39.99$

How will adding the wheel of cheese affect the mean and median?

Choose 1 answer:

(A) Both the mean and median will increase, but the median will increase by more than the mean.

(B) Both the mean and median will increase, but the mean will increase by more than the median.

(C) The mean will increase, and the median will decrease.

(D) The median will increase, and the mean will decrease.

2.

A small accounting firm has 4 accountants who each earn a different salary between $\$50{,}000$ and $\$60{,}000$, and a 5^{th} accountant who works part-time for tax season and earns $\$10{,}000$.

The firm decides to get rid of the part-time accountant and keep the other 4 salaries the same.

How will getting rid of the part-time accountant affect the mean and median?

Choose 1 answer:

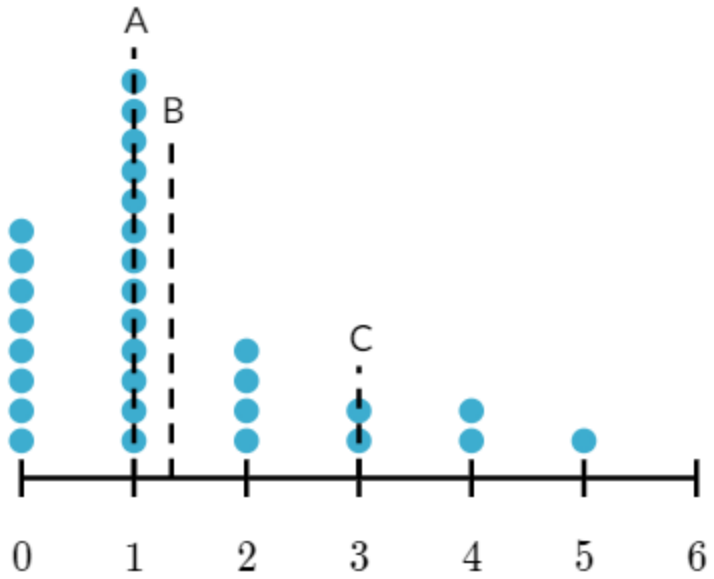
(A) Both the mean and median will increase, but the median will increase by more than the mean.

(B) Both the mean and median will increase, but the mean will increase by more than the median.

(C) The mean will increase, and the median will decrease.

(D) The median will increase, and the mean will decrease.

3. Use the dotplot below to estimate the following parameters.

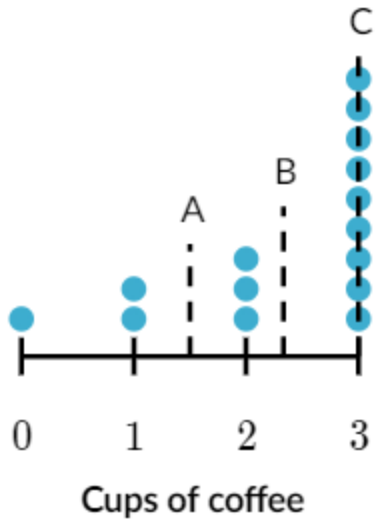


a. Which line represents the value of the mean?

b. Which line represents the value of the median?

~~~U03L04 Homework~~~

1. Summarizing quantitative data: Effects of shifting, adding, & removing a data point
2. Summarizing quantitative data: Estimating mean and median in data displays

~~~U03L04 Classwork~~~

1. Use the dotplot below to estimate the following parameters.



   a. Which line represents the value of the mean?

   b. Which line represents the value of the median?

2.

A group of 4 friends likes to bowl together, and each friend keeps track of his all-time highest score in a single game. Their high scores are all between 180 and 220, except for Adam, whose high score is 250.

Adam then bowls a great game and has a new high score of 290.

How will increasing Adam's high score affect the mean and median?

Choose 1 answer:

(A) Both the mean and median will increase.

(B) The median will increase, and the mean will stay the same.

(C) The mean will increase, and the median will stay the same.

(D) The mean will increase, and the median will decrease.
~~~

3. A student wanted to investigate the miles per gallon of several popular cars. The following data collected by the student show the amounts of miles per gallon for 10 popular cars.

   54 31 26 32 44 29 47 33 24 22

   a. Construct an appropriate graphical display of the amounts of miles per gallon found in the 10 cars.

   b. Use the graph in part A to write a few sentences describing the distribution of miles per gallon ratings for the 10 cars.

   c. An experimental car that has not entered production yet claims to get over 250 miles per gallon. If this value was added to the data set of ten numbers above, how would the mean and median of the data set above compare with the mean and median of the new data set with the 11th numbers? Explain how this comparison could be made without performing any computations.

4. An online journalist used the median instead of the mean when it reported the incomes of the average worker in Silicon Valley. A graphical display of incomes from people who worked in Silicon Valley indicated that the data were strongly skewed to the right. Which of the following explains why, in this situation, the median is a more accurate indicator of the income of the average worker than the mean is?
   A. The mean is affected by the skewness, whereas the median is not.
   B. The median is always the preferred statistic.
   C. The mean will be less than the median when the data are strongly skewed to the right.
   D. The mean should be used only when the data are strongly skewed to the left.
   E. The median is equal to to one-half the sum of the maximum and minimum incomes in Silicon Valley.

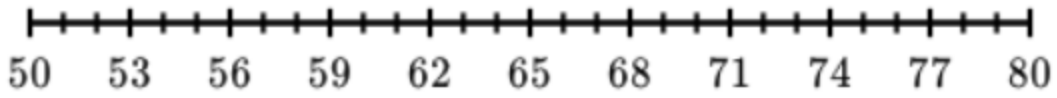
Unit 03 Lesson 05: Creating Box Plots

Is there a competitor to histograms?

- Draw box plots from given data [Source: Khan Academy]

1. The following data represent the test scores, in percent, of a sample of AP Chemistry students. Create a box plot to summarize the data.

66 73 75 76 54 67 69 70 60 65 72



2. **AP**: Which of the following statements would be true if a probability distribution is symmetric?
   a. The distribution is normal.
   b. The distribution is uniform.
   c. The distribution is bimodal.
   d. The mean of the distribution is equal to the median of the distribution.
   e. The interquartile range of the distribution is equal to the standard deviation of the distribution.

3. Create a box and whisker plot for the following data

29 35 35 30 34 29 25 28 37 38

4. Data were collected on the number of miles that individual drivers in a specific city drove for the month of March. The quartiles for these data are given below.

Q1	Q2	Q3
6,389 miles	7,454 miles	8,274 miles

Declare the following statements as true or false, and justify with a logically conclusive statement.

a. At least half of the drivers drove less than or equal to 7,454 miles and at least half drove greater than or equal to 7,454 miles.

b. Seventy-five percent of the drivers drove between 6,389 miles and 8,274 miles.

c. Twenty-five percent of the driver drove less than or equal to 8,274 miles and the remaining 75 percent drover greater than or equal to 8,274 miles.

d. The mean number of miles driven is 7,454 miles.

e. A majority of drivers drove 7,454 miles.

~~~U03L05 Homework~~~ 1. Summarizing quantitative data: Interquartile range (IQR)
2. Summarizing quantitative data: Creating box plots

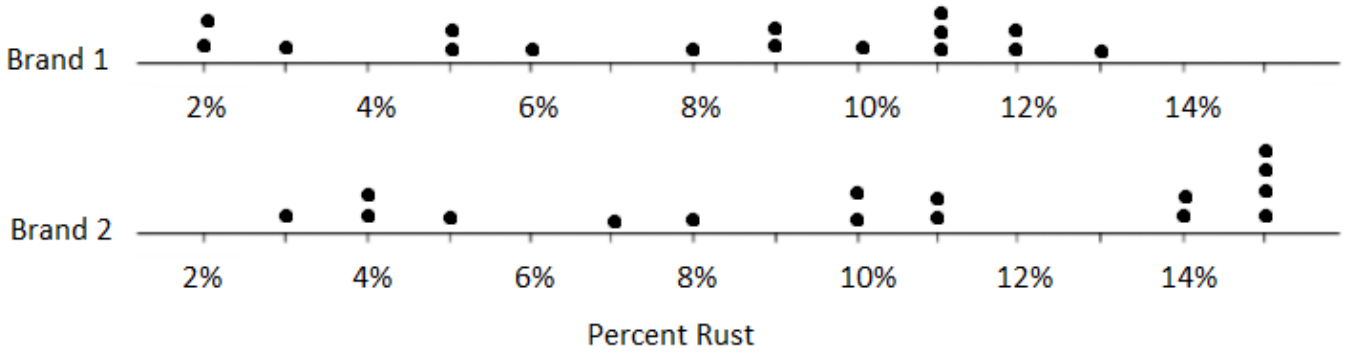
~~~U03L05 Classwork~~~

1. Mr. Rose recorded the number of leashless dogs he saw while walking his dog for the past 9 days. Mr. Rose passed no judgement on their owners, even though it's really not that hard to just buy a leash so your dog isn't running around marauding innocent passersby who leash their dogs responsibly. Display this distribution using a box and whisker plot.

   12, 15, 20, 22, 24, 15, 17, 19, 19

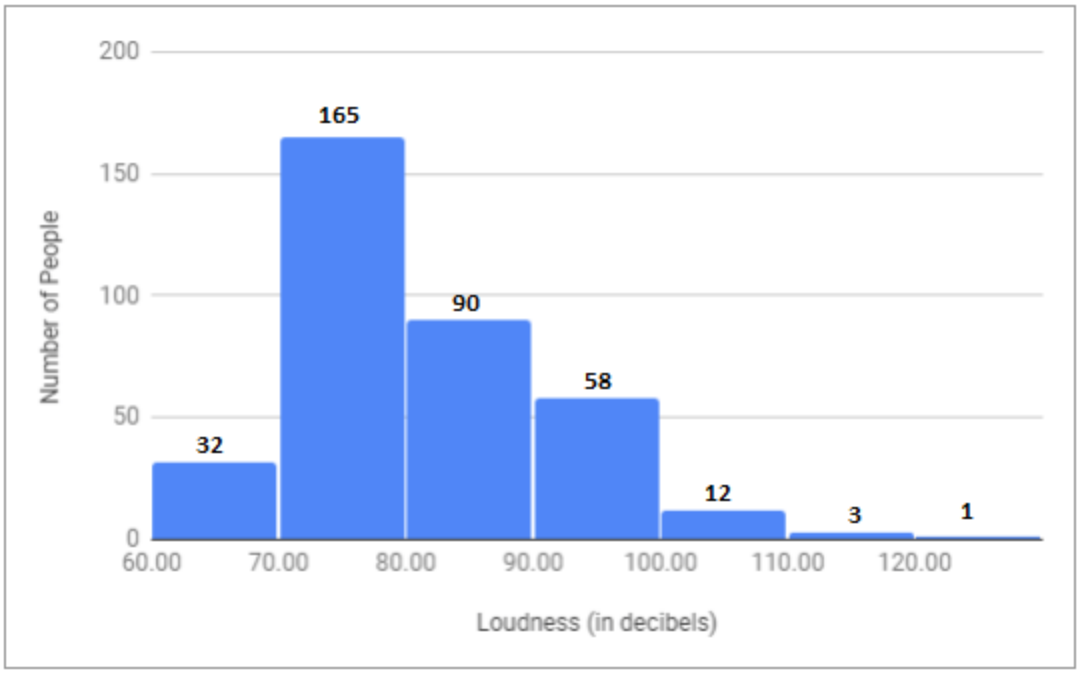
2. **AP**: What is the best synonym for 50th percentile?
   a. The mean
   b. The median
   c. The mode
   d. The interquartile range
   e. The standard deviation

3. **AP**: A random sample of cars from 2 different brands were inspected for frame rust after 10 years of operation. There were 16 cars sampled for each brand.



   Based on the dotplots, which of the following is greater for the percent rust of brand 1 than for the percent rust of brand 2?
   a. The first quartile
   b. The median
   c. The third quartile
   d. The range
   e. The maximum minus the third quartile.

4. **AP**: The histogram below shows the preferred volume of music for a sample of college students.



Which of the following must be true?

a. The minimum decibels is 60.
b. The maximum decibels is 120.
c. The median decibel level is not greater than 80.
d. The mean decibel level is between 70 and 80.
e. The upper quartile is greater than 100 decibels.

5. **AP**: A random sample of readers of a doomsday prepper website were asked how many days worth of food they had stored up for the apocalypse. Summary statistics are given below.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Days	120	115.7	48	78.9	289.2	31.6

Variable	Minimum	Maximum	Q1	Q3
Days	0.0	2900	29	93

Which of the following statements is true?

a. Seventy-five percent of the readers in the sample have more than 93 days worth of food stored.
b. Fifty percent of the readers in the sample have between 0 and 115.7 days worth of food stored.
c. The distribution of the number of days worth of food stored appears to be approximately symmetric.
d. The interquartile range of the number of days worth of food stored is 64.
e. The distribution of the number of days worth of food stored contains outliers on both the low side and the high side.

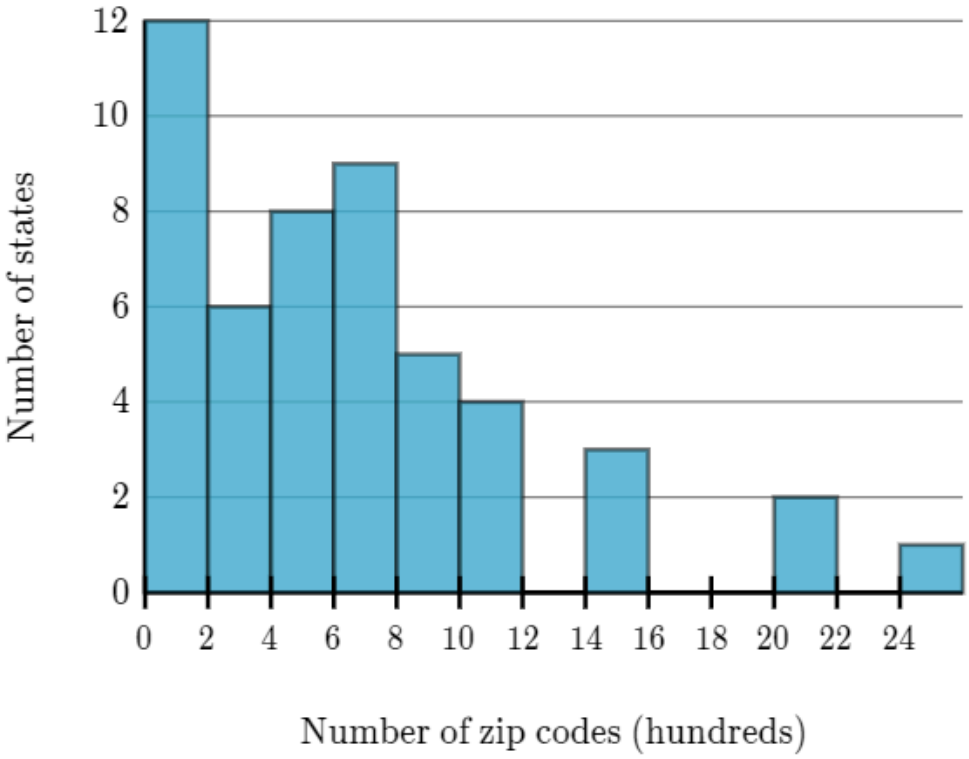
Unit 03 Lesson 06: Identify Outliers

What do we do with data points that don't seem to fit in with the others?

- Mathematically calculate if a data point would be considered an outlier [Source: Khan Academy]
- Draw box plots from given data, including outliers

1. According to the 1.5IQR rule, how many points are outliers?

Zip codes are numerical designations of locations for the purpose of mail delivery. The following histogram summarizes the number, in hundreds, of zip codes assigned to each state in the United States.



Here is the five-number summary for these data:

Five-number summary

min	Q_1	median	Q_3	max
1	275	533	933	2596

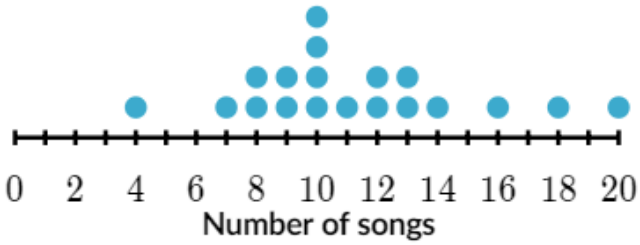
Flake Weight (g)	Frequency	Relative Frequency	Cumul Freq	Cumul Rel Freq
0.1	0	0	0	0.0000
0.2	0	0	0	0.0000
0.3	0	0	0	0.0000
0.4	28	0.0918	28	0.0918
0.5	27	0.0885	55	0.1803
0.6	28	0.0918	83	0.2721
0.7	30	0.0983	113	0.3704
0.8	31	0.1016	144	0.4721
0.9	30	0.0983	174	0.5704
1.0	31	0.1016	205	0.6721
1.1	30	0.0983	235	0.7704
1.2	34	0.1114	269	0.8819
1.3	20	0.0655	289	0.9475
1.4	5	0.0163	294	0.9639
1.5	2	0.0065	296	0.9704
1.6	0	0.0	296	0.9704
1.7	0	0.0	296	0.9704
1.8	1	0.0032	297	0.9737
1.9	0	0.0	297	0.9737
2.0	1	0.0032	298	0.9770
2.1	1	0.0032	299	0.9803
2.2	1	0.0032	300	0.9836
2.3	1	0.0032	301	0.9868
2.4	1	0.0032	302	0.9901
2.5	1	0.0032	303	0.9934
2.6	2	0.0064	305	1.0000

2. **AP**: An outlier may be defined as a data point that is more than 1.5 times the interquartile range below the lower quartile or is more than 1.5 times the interquartile range above the upper quartile. According to this definition, what is the diameter, in inches, of the smallest flake that is an outlier?
   A) 2.0 B) 1.9 C) 1.8 D) 0.4 E) 2.1

~~~U03L06 Homework~~~ 1. Summarizing quantitative data: Identifying outliers

1. According to the 1.5IQR rule, which points are outliers?

The following dotplot shows the number of songs on each album in Sal's collection. Each dot represents a different album.



Here is the five-number summary for these data:

Five-number summary

min	Q_1	median	Q_3	max
4	9	10	13	20

2. **AP**: The number of plays per month for a popular YouTube video was recorded for each of the last 24 months. Summary measures are shown below.

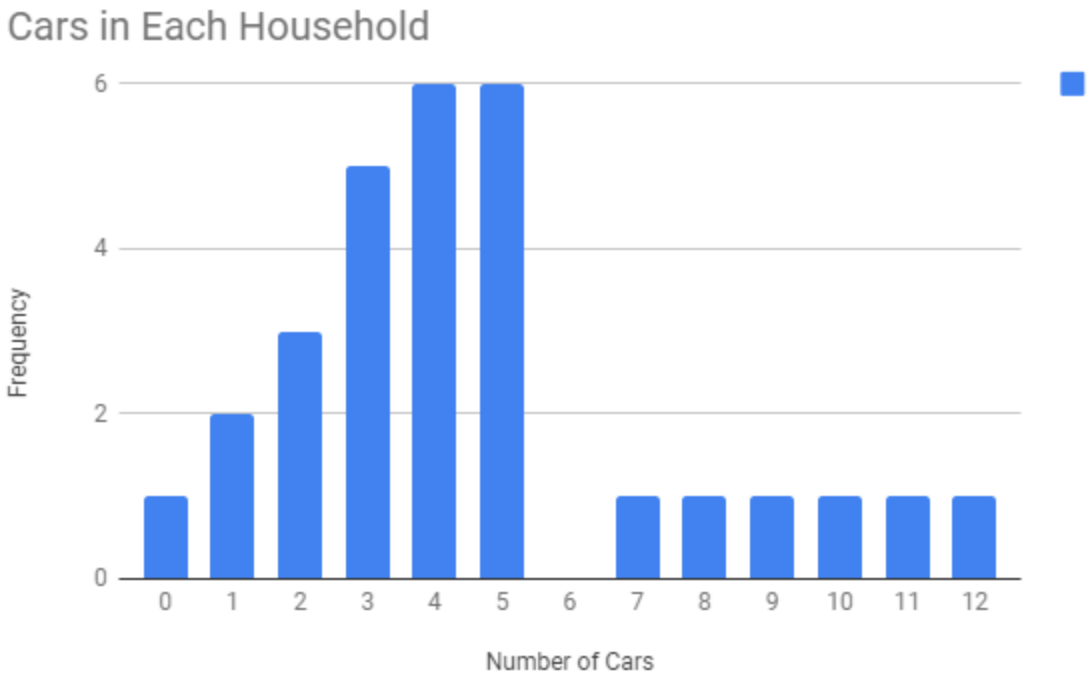
Min = 1,004,048
Lower quartile = 2,189,045
Median = 3,583,939
Max = 9,143,270
Upper quartile = 4,389,304
n = 24

Which of the following statements is true?

a. The smallest number of plays is 1,004,048 and is an outlier. No other months in the data set could be outliers.
b. The largest number of plays is 9,143,270 and it is an outlier. No other months in the data set could be outliers.
c. Both 1,004,048 and 9,143,270 are outliers. It is possible there could be other outliers.
d. 1,004,048 is an outlier and it is possible that there are other outliers at the low end of the data set. There are no outliers at the high end of the data set.
e. 9,143,270 is an outlier and it is possible that there are other outliers at the high end of the data set. There are no outliers at the low end of the data set.

3. **AP**: Ahmad surveyed a sample of 29 people in Hollywood, asking how many cars they had.

Mean	Standard Deviation	Q1	Q2	Q3
2.231	2.088	3	4	5



An outlier may be defined as a data point that is more than 1.5 times the interquartile range below the lower quartile or is more than 1.5 times the interquartile range above the upper quartile. According to this definition, which of the following is true?

a. The median is less than the mean, and the distribution has 1 outlier.
b. The median is less than the mean, and the distribution has 2 outliers.
c. The median is greater than the mean, and the distribution has 3 outliers.
d. The median is greater than the mean, and the distribution has 4 outliers.
e. The median is greater than the mean, and the distribution has 5 outliers.

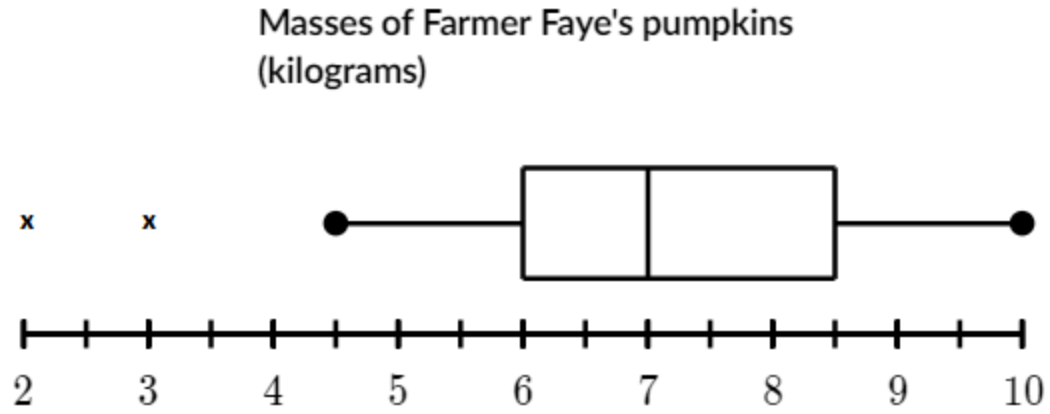
4. **AP**: A chemist recorded 20 measurements of the voltage of an experimental battery. One of the measurements was an outlier when compared with the other 19 measurements. Which of the following must be true about the 19 measurements, excluding the outlier, when compared with the 20 measurements? (Note: An outlier may be defined as a data point that is more than 1.5 times the interquartile range below the lower quartile or is more than 1.5 times the interquartile range above the upper quartile.)
   a. The median of the 19 measurements is less than the median of the 20 measurements.
   b. The median of the 19 measurements is greater than the median of the 20 measurements.
   c. The maximum of the 19 measurements is less than the maximum of the 20 measurements.
   d. The maximum of the 19 measurements is greater than the maximum of the 20 measurements.
   e. The standard deviation of the 19 measurements is less than the standard deviation of the 20 measurements.

Unit 03 Lesson 07: Using Box Plots

For what questions are box plots a useful aid?

- Use box plots to answer a variety of questions about data [Source: Khan Academy]

1. Determine the following parameters of the distribution of Faye's pumpkins' masses. If it cannot be determined using the box and whisker plot below, write "cannot be determined".



S

a. Shape

O

b. Outliers

C

c. Mean

d. Median

e. Size of Harambe's favorite pumpkin

S

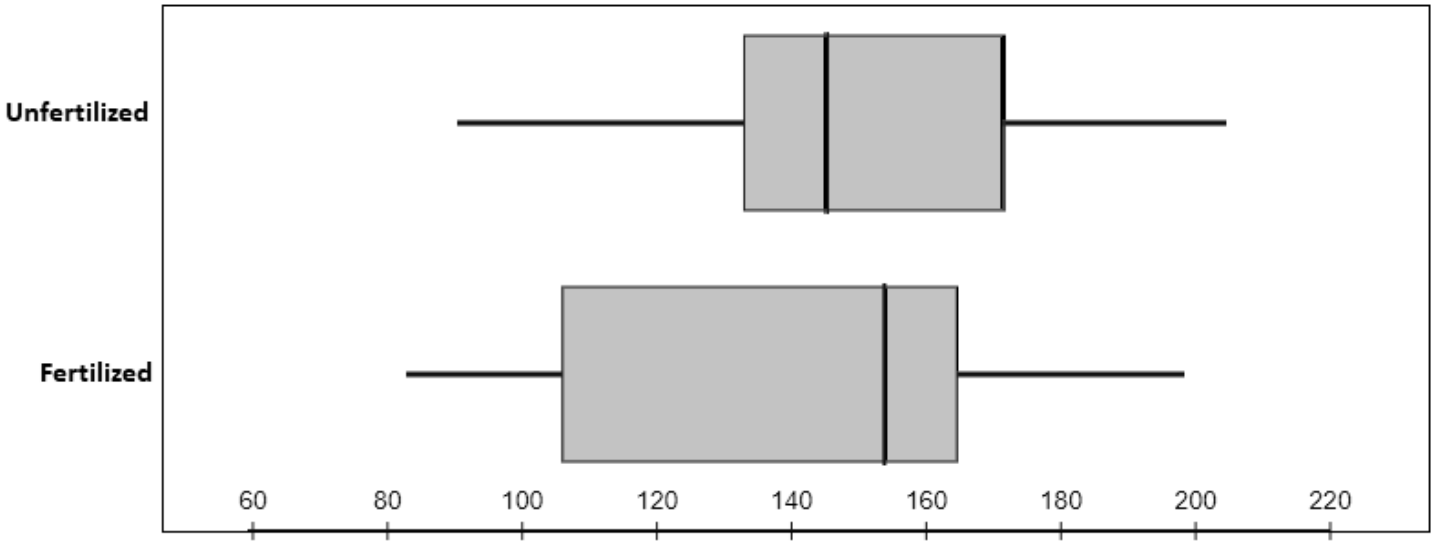
f. Range

g. Interquartile Range

h. Variance

i. Standard Deviation

2. **AP**



The figure above summarizes the heights, in centimeters, of approximately 200 tomatoes plants 3 months after they were planted in a field. Approximately half of the plants were allowed to be fertilized, and the remaining plants were never fertilized. Which of the following statements about the medians and the interquartile ranges (IQRs) of the heights of the two groups of plants 3 months after being planted is true?

a. The medians and IQRs are the same for the fertilized and unfertilized plants.
b. The median for the unfertilized plants is greater than the median for the fertilized plants, and the IQR is also greater for the unfertilized plants.
c. The median for the unfertilized plants is the same as the median for the fertilized plants, and the IQR is greater for the unfertilized plants.
d. The median for the unfertilized plants is less than the median for the fertilized plants, and the IQR is greater for the unfertilized trees.
e. The median for the unfertilized plants is less than the median for the fertilized plant, and the IQR is less for the unfertilized trees.

3.

The five-number summary for the number of accounts managed by each sales manager at Force Inc. is shown in the following table.

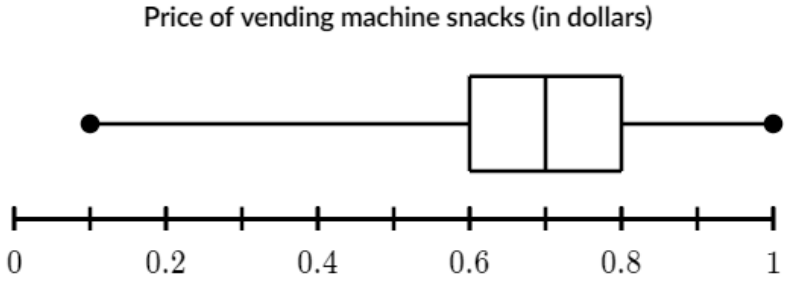
Min	Q_1	Median	Q_3	Max
35	45	50	65	85

The five-number summary suggests that about 50% of sales managers at Force Inc. manage fewer than what number of accounts?

~~~U03L07 Homework~~~

1. Summarizing quantitative data: Reading box plots
2. Summarizing quantitative data: Interpreting quartiles

1. Identify, with as much specificity as possible, the following values. Some may not be possible at all.



   a. Outliers

   b. Range

   c. Variance

   d. Mean

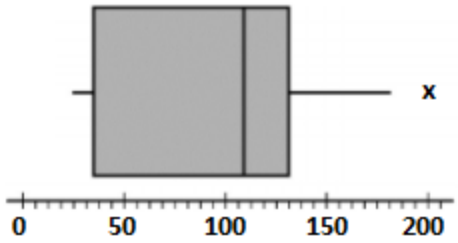
   e. Interquartile Range

   f. Median

   g. Shape

   h. Standard Deviation

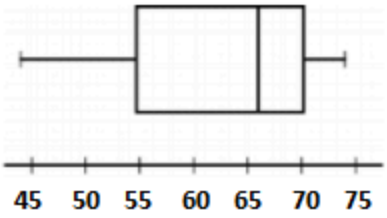
2. **AP**: Itzel surveyed a large group of seniors asking how many Snapchat messages they sent throughout the previous week. She created a boxplot summarizing her data below.



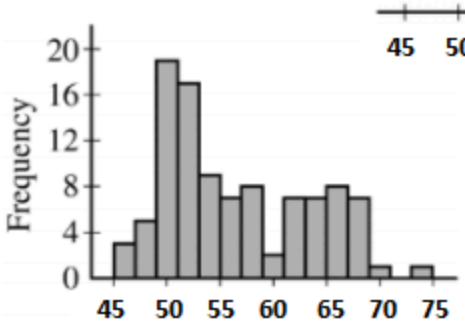
   Based on her boxplot, which of the following statements is the most reasonable conclusion?

   a. There are more students with data values between the first quartile and the median than between the median and the third quartile.
   b. There are less students with data values between the first quartile and the median than between the median and the third quartile.
   c. The data are more spread out between the third quartile and the median than between the median and the first quartile.
   d. There are approximately the same number of students with data values between the third quartile and the minimum as there are students with data values between the first quartile and the maximum.
   e. There are fewer students with data values above the median than there are students with data values below the median

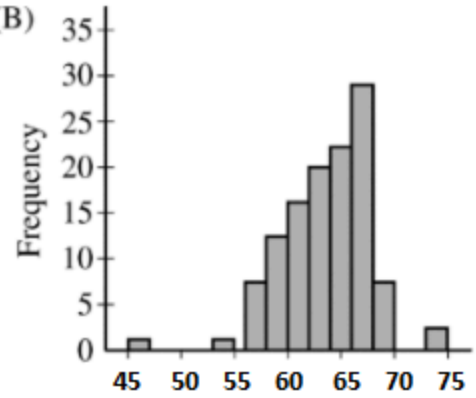
3. **AP**: Which histogram could have been produced from the same dataset as the the box and whisker plot shown? Show something logical to justify your choice.



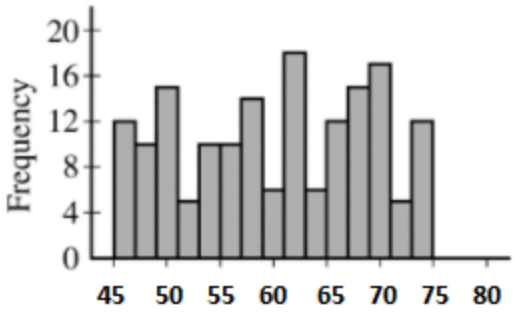
(A)



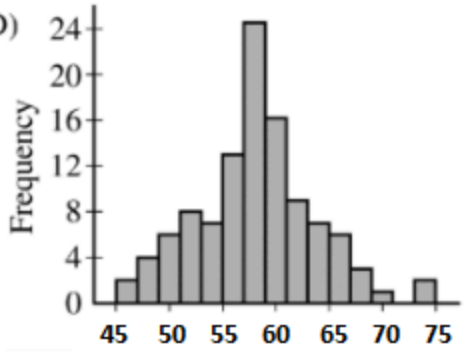
(B)



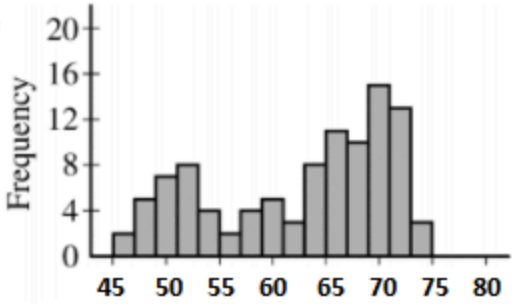
(C)



(D)

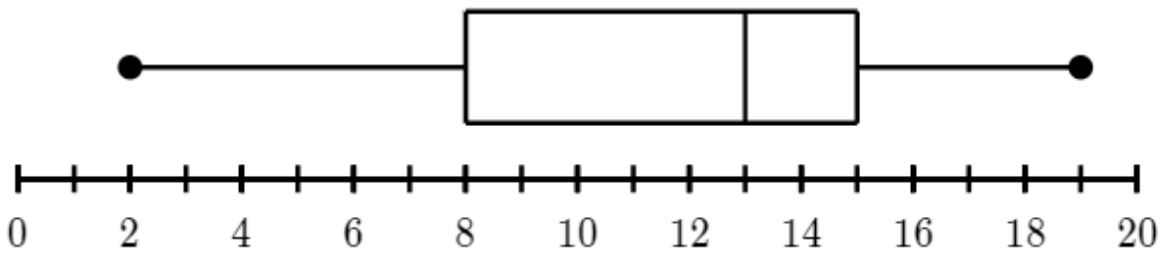


(E)



4.

Number of surfers at each beach in Miami



The box plot suggests that 75% of beaches in Miami have more than what number of surfers?

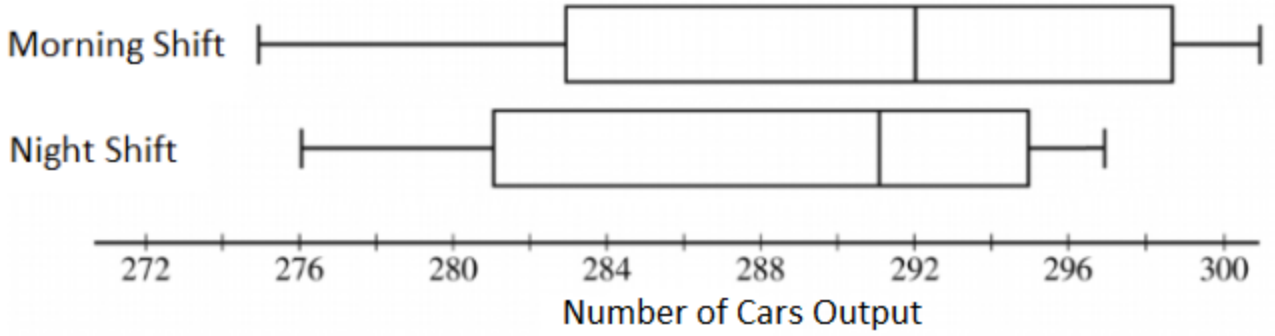
Unit 03 Lesson 08: Review for Test on Summarizing Quantitative Data

Can I handle any situation that involves means, medians, standard deviations, or box plots?

- Ensure you've mastered the concepts and skills of summarizing quantitative data
- Ensure you've retained mastery of preview units

Unit 03

1. The manager of a car factory keeps track of the total production output during both the morning shift and the night shift every day for a month. The boxplots below summarize the total number of cars manufactured by each shift of workers.



Based on the boxplots, which of the following statements must be true?

a. More cars were produced during the morning shift than the night shift.
b. For both shifts, at least half the days made less than 292 cars.
c. The number of days with an output of less than 294 cars for the night shift was greater than the number of days with an output of less than 294 cars for the morning shift.
d. The range of outputs for the morning shift is less than the range of outputs for the night shift.
e. The output of the day with the least output during the morning shift was greater than the output of the day with the least output during the night shift.

Unit 01

2. Adri wanted to research the connection between musical instruments played and GPA. He surveyed a simple random sample of 146 students at his high school of 3,000 students.

Musical Instruments Played	GPA ≤ 3.0	3.0 < GPA ≤ 3.5	3.5 < GPA ≤ 4.0	4.0 < GPA
Zero	50	34	21	1
One	5	3	12	10
Two	2	1	1	4
Three or more	0	1	0	1

a. What proportion of students play one instrument and have a GPA above 4.0?

b. What proportion of students play one instrument or have a GPA above 4.0?

c. What proportion of students who play one instrument have a GPA above 4.0?

d. What proportion of students have a GPA above 4.0?

Unit 02

3. Contestants on The Price is Right each try to guess the price of a new microwave oven. The contestant who guesses closest to the actual price without going over wins a prize. The guess of each contestant is shown in the back to back stemplot below.

Women		Men
9 7 5 5 3 2 1	2	2
7 5 3 2 1 1	3	
3 2 1 1	4	1 6
7 2	5	1 4 5
7	6	1 5 8 9
	7	1 4 4 8 9
2	8	1 5 5 5 9
	9	0 2 4 5 6 7 8 9

Compare the distributions of men and women (!!Note: I've been teaching this wrong. You can't just describe the two distributions separately. You must have a sentence comparing shape to shape, another sentence comparing center to center, etc. Moreover, you must use a comparative adjective, such as "higher", "greater", or "lower" when comparing center and and spread.)

Unit 03

Be able to

- Calculate/Estimate means and medians, either given a list of numbers or a graphical display
- Calculate variances and standard deviations
- Compare standard deviations of distributions given graphical displays
- Describe the effects of adding/removing/altering a data point, especially an extreme value
- Draw and/or analyze box plots
- Classify outliers using the "1.5*IQR rule"