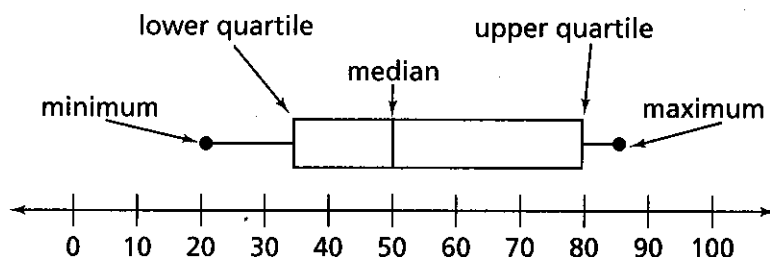


CC Investigation 5: Histograms and Box Plots

© DOMAIN: Statistics and Probability

A box plot is constructed from the **five-number summary**: the minimum value, lower quartile, median, upper quartile, and maximum value.



Problem 5.1

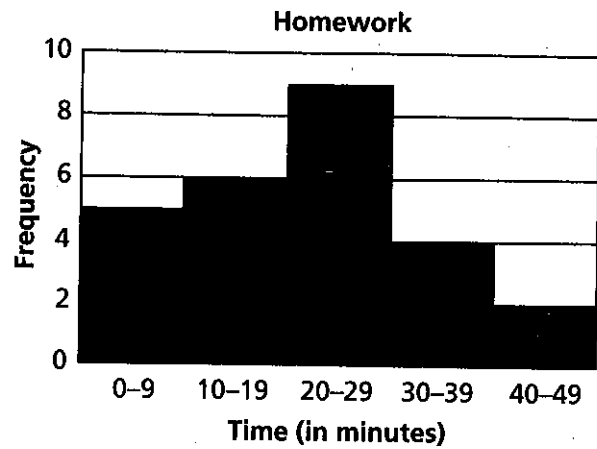
The Panthers scored the following numbers of points during games this season:

68	91	86	89	88	82	95	85	80	78	82
68	86	96	73	68	91	80	90	86	72	87

- Order the set of data from the least to the greatest. What are the minimum and maximum values?
- Find the median. Explain how you found this value.
- The **lower quartile** is the median of the lower half of the scores. What is the lower quartile of the data?
- The **upper quartile** is the median of the upper half of the scores. What is the upper quartile of the data?
- Find the difference between the upper quartile and the lower quartile. This difference is called the **interquartile range**.
 - What does the interquartile range represent?
- Draw a number line from 60 to 100.
 - Above your number line, draw vertical line segments at the values you found for the median, the lower quartile, and the upper quartile.
 - Connect the vertical lines to form a rectangle.
 - Locate the value you identified as the minimum and draw a line to the left from the rectangle to meet that point.
 - Locate the value you identified as the maximum and draw a line to the right from the rectangle to meet that point.

- G.**
1. Where is a score of 75 found on your box plot?
 2. What does the location of 75 tell you about the performance of the team this season?
- H.** The Panthers scored 96 points in the first game of the playoffs.
1. How do you think this change will affect the mean, median, and mode of the numbers of points scored?
 2. Find the mean, median, and mode for the 23 games played this season.
 3. How did the mean, median, and mode change?
 4. Make a new box plot to include the data for all 23 games.
 5. Can you see the changes to the mean, median, or mode between the first and second box plots? Explain why or why not.
 6. Which of the five-number-summary values changed the most? Explain your answer.

A **histogram** is a type of bar graph in which the bars represent numerical intervals. Each interval must be the same size, and there can be no gaps between them. In this histogram, there are 5 equal intervals of 10 minutes each.



Problem 5.2

A. The table shows the winning scores in the first round of the basketball tournament.

90	69	70	89	62	97	64	68
79	67	77	66	65	89	99	82
100	81	78	53	80	62	86	77
73	77	69	72	51	68	80	73

1. What are the greatest and least winning scores?
 2. Divide the range of the data into equal intervals that will be represented by bars on the histogram. Give the range for each interval.
 3. Explain why you chose that number of intervals.
 4. Make a table to show the frequency of scores in each interval.
 5. Make a histogram of the data. Draw a bar for each interval to represent the frequency.
 6. Summarize what the histogram shows about the data.
- B. The table shows the scores of all of the games in the football playoffs.

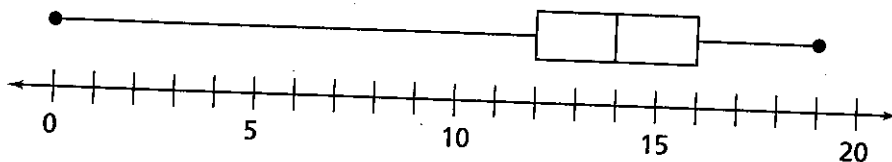
33	24	51	34	34	45	17	20
30	31	31	17	17	28	3	14
14	3	14	45	14	14		

1. Make a histogram of the data.
2. Summarize what the histogram shows about the data.
3. Compare the histogram to the one you made in Part A. Explain the differences the graphs illustrate about the data.

You can summarize data sets using measures of variability. **Variability** is the degree to which data are spread out around a center value.

Problem 5.3

The box plot shows the number of points Grant scored in each game.



- What are the median, lower quartile, and upper quartile of the data?
- What is the interquartile range? What does that number tell you about the how consistent Grant's scoring was this season?
- Describe the overall pattern of the data.
- Do there appear to be any scores that do not follow the pattern of the rest of the data? Explain what those values represents and what makes them unusual.

A data set's **mean absolute deviation** is the average distance of all data values from the mean of the set. First, find the mean of the data set. Then find the distance of each value in the set from that mean and find the average of those distances.

Problem 5.4

Paige's lacrosse team scored the following numbers of goals in the first six games of the season: 7, 6, 17, 8, 7, 9.

- What is the mean number of goals scored? Show your work.
- What is the distance of each data value from the mean?
- What is the total distance of all of the data points from the mean?
 - The mean absolute deviation is the average of these distances. What is the mean absolute deviation of the data?
- What does the mean absolute deviation tell you about the numbers of goals scored?
- Do you notice any value that does not follow the pattern of the rest of the data? Explain what makes that value unusual.

Exercises

For Exercises 1–3, use a five-number summary to draw a box plot for each set of data.

1. 12 7 3 11 13 18 8 4 3 10

2. 26 16 25 30 29 21 18 32 25 15 20

3. 4.2 3.8 6.2 7.8 8.3 2.9 6.8 9.3 4.3

4. A farmer starts 9 tomato plants in a greenhouse several weeks before spring. The seedlings look a little small this year so the farmer decides to compare this year's growth with last year's growth.

This year's growth is measured in inches as:

12 8.4 10 9.8 14 7.9 11 12.7 13.7

- a. Use a five-number summary to draw a box plot for this set of data. Mark your number line from 0 to 20.
- b. Last year, the five-number summary for the tomato plants was 9, 11, 13.4, 16, 17. Draw a box plot for this set of data. Mark your number line from 0 to 20.
- c. Write this year's summary above last year's summary. Is the farmer's concern justified? Why or why not?
5. a. Explain why you would use a box plot when you have similar data to compare.
- b. Explain why you would not use a box plot if you needed to show the mean of the data.
6. CJ scored 85, 88, 94, 90, and 64 on math tests so far this grading period. His teacher allows students to retake the test with the lowest score and substitute the new test score. CJ scores a 98 on the retest. How does substituting the new test score affect the mean, median, and mode?
7. During a dance competition, Laura's dance team received scores of 9, 9, 8, 9, 10, 8, 3, and 8 from the judges. For each team, the highest and lowest scores are removed. The remaining scores are then averaged to find the team's final score. How was the team's final score affected when the highest and lowest scores were removed?
8. a. During the first 8 games of the basketball season, Rita made the following number of free throws: 0, 3, 5, 5, 4, 8, 5, and 6. During the next 7 games she made 8, 8, 8, 7, 9, 8, and 3 free throws. Make a box plot showing the data for the first 8 games and then the data for all of the games.
- b. How did the mean, median, and mode of the free-throw data change from the first 8 games compared to all of the games?

For Exercises 9–11, use the information below.

Teams of two competed in the egg-toss distance competition. If the egg breaks, the distance is 0. The results for Dave and Paul's team are shown below for the first round.

Dave and Paul's Egg Toss Results								
Toss	1	2	3	4	5	6	7	8
Distance (in meters)	3.5	0	6	6	0	5	8.5	11

In a bonus round, each team can replace 1 toss from the first round. Dave and Paul make a toss of 12 meters.

9. How did the mean, median, and mode change after the toss from the bonus round?
10. Which measure—mean, median, or mode—changed the most?
11. Make a box plot using the data after the first round and then using the data after the bonus round.
12. Faye is writing an article in the school newspaper about the school's paper airplane flying competition. She records Wheeler's first flights in the table below.

Wheeler's Paper Airplane Competition Results								
Flight	1	2	3	4	5	6	7	8
Distance (in feet)	20	18	20	22	21	19	21	20

After the first 8 flights, Wheeler adds a paper clip to the nose of his airplane. Faye records the results of his next 8 flights in the table below.

Wheeler's Paper Airplane Competition Results								
Flight	9	10	11	12	13	14	15	16
Distance (in feet)	35	40	44	41	48	47	47	46

- a. Make a box plot showing the data for the first 8 flights and then the data for the second 8 flights.
- b. How did the mean, median, and mode of the flight distances change from the first 8 flights to the second 8 flights?

For Exercises 13–15, make a histogram to display the set of data.

13. ages of mall shoppers:

23	33	21	18	17	45	40	23
12	31	27	27	29	24	14	40
19	18	25	17	36	40	38	20

14. class grades on a history exam:

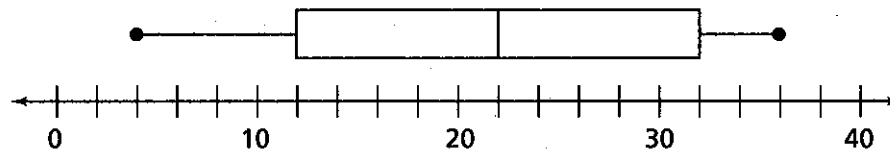
97	84	93	76	87	100	92	90
70	85	83	99	90	89	84	91
100	96	76	74	73	87	80	93

15. weights of pumpkins (in lb):

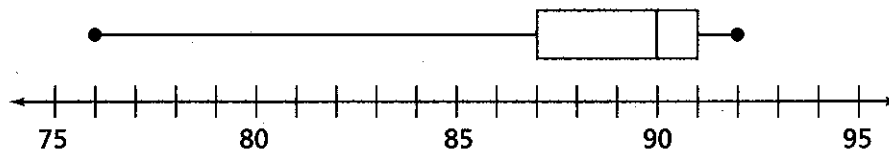
5	16	23	8	7	9	12	15
20	15	7	18	6	6	21	16
8	11	12	16	10	20	23	9
14	24	17	7	6	18	9	10

16. Multiple Choice What is the interquartile range of the data?

- A. 8
- B. 20
- C. 10
- D. 24



17. a. Describe the overall pattern of the data shown in the box plot.



b. Identify any data value that is far outside the pattern and explain why it is outside the pattern.

For Exercises 18–20, find the mean absolute deviation for the set of data.

18. 21 23 18 27 30 24

19. 88 89 86 89 90 82

20. 2.4 2.8 2.1 2.7 13.0 2.5

21. Describe the overall pattern of data in the following set. Identify any data value that is far outside the pattern and explain why it is outside the pattern.

11 13 9 12 14 12 10 12
7 9 13 11 12 10 45 13

22. After the eighth game and for the rest of the season, Brandon spent an hour after each practice working on 3-point baskets. He also practiced for another half-hour when he got home. His 3-point basket data for the entire season are shown below.

Game	1	2	3	4	5	6	7	8	9	10
Points	0	3	0	0	6	3	3	0	6	6
Game	11	12	13	14	15	16	17	18	19	20
Points	12	12	12	15	6	12	15	9	12	12

- How did the data change after the first 8 games, if at all?
- Why do you think the data did, or did not, change? Explain.
- When data in a data set change, changes in which measures (mean, median, mode) will be shown in box plots? Explain your thinking.
- Display the data from Brandon's first 8 games in a box plot. Display the data from all 20 games in another box plot. Explain how differences between the two groups of data are shown in the plots.