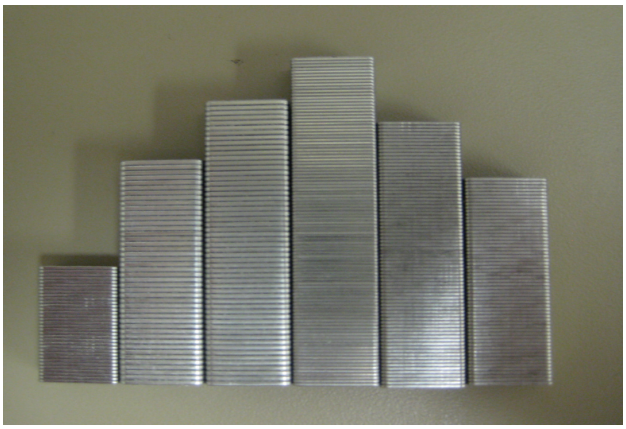
9.2 Data Distribution

A Practice Understanding Task



A lot of information can be obtained from looking at data plots and their distributions. It is important when describing data that we use context to communicate the **shape, center,** and **spread.**

Shape and spread:

- **Modes:** uniform (evenly spread- no obvious mode), unimodal (one main peak), bimodal (two main peaks), or multimodal (multiple locations where the data is relatively higher than others).
- **Skewed distribution:** when most data is to one side leaving the other with a 'tail'. Data is skewed to side of tail. (if tail is on left side of data, then it is skewed left).
- **Normal distribution and standard deviation:** curve is unimodal and symmetric. Data that has a normal distribution can also describe the data by how far it is from the mean using standard deviation.
- **Outliers:** values that stand away from the body of the distribution. For a box-and-whisker outliers determined if they are more than 1.5 times the interquartile range (length of box) beyond quartiles 1 and 3. Also considered an outliner if data is more than two standard deviations from the center of a normal distribution.
- **Variability:** values that are close together have low variability; values that are spread apart have high variability.

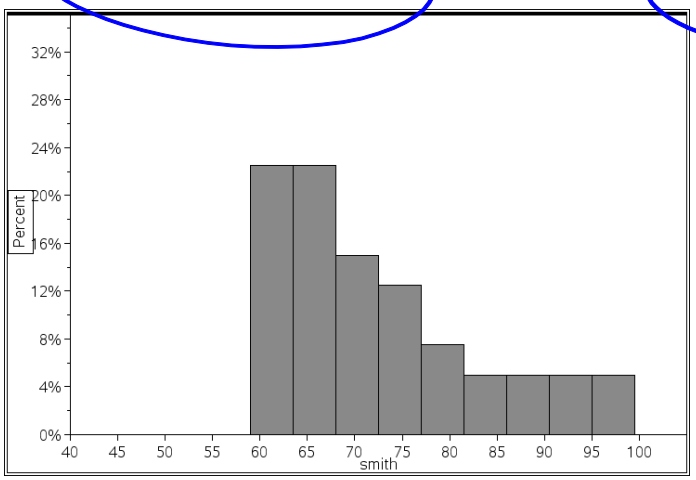
Center:

- Analyze the data and see if one value can be used to describe the data set. Normal distributions make this easy. If not a normal distribution, determine if there is a 'center' value that best describes the data. Bimodal or multimodal data may not have a center that would provide useful data.

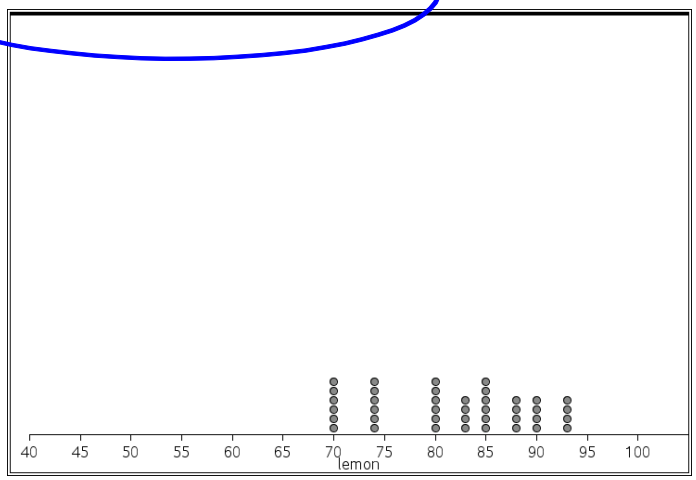
There are representations of test scores from six different classes found below, for each:

1. Describe the data distribution.
2. Compare data distributions between Anderson and Williams.
3. Compare data distributions between Williams and Lemon.
4. Compare data distributions between Croft and Hurlea.
5. Compare data distributions between Jones, Spencer, and Anderson.
6. Compare data distributions between Spencer and the other histograms.
7. Which distributions are most similar? Different? Explain your answer.

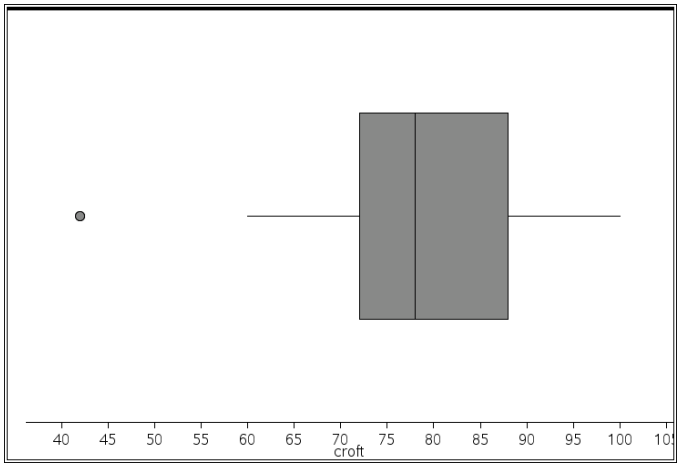
Data set I: Williams's class



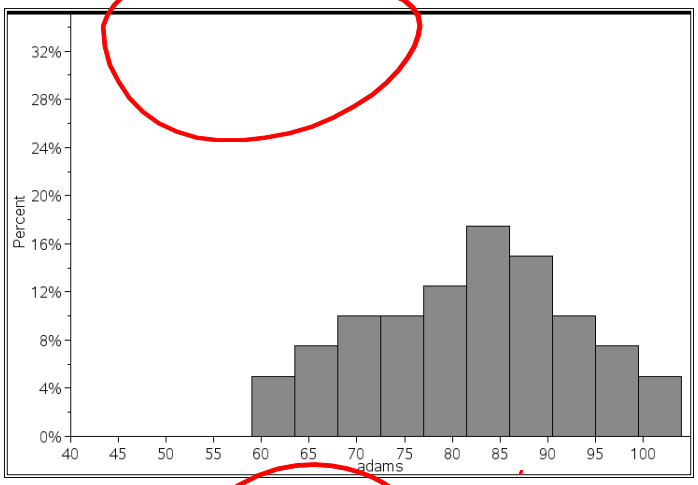
Data set II: Lemon's class



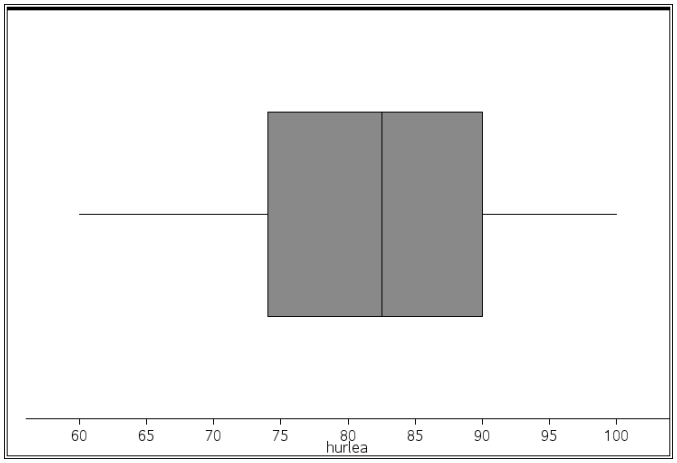
Data set III: Croft's Class



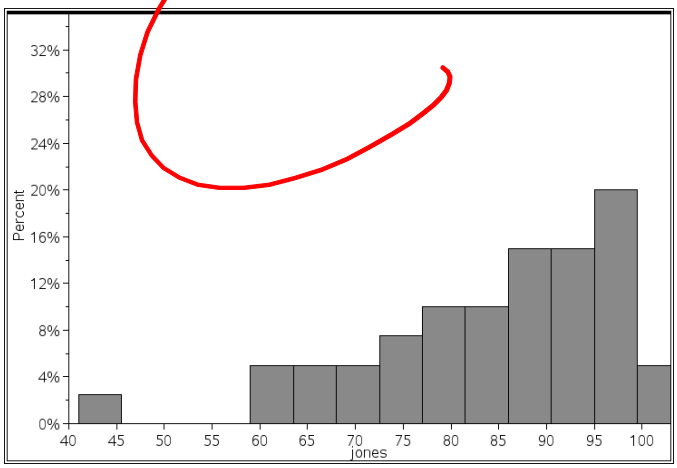
Data set IV: Anderson's Class



Data set V: Hurlea's class

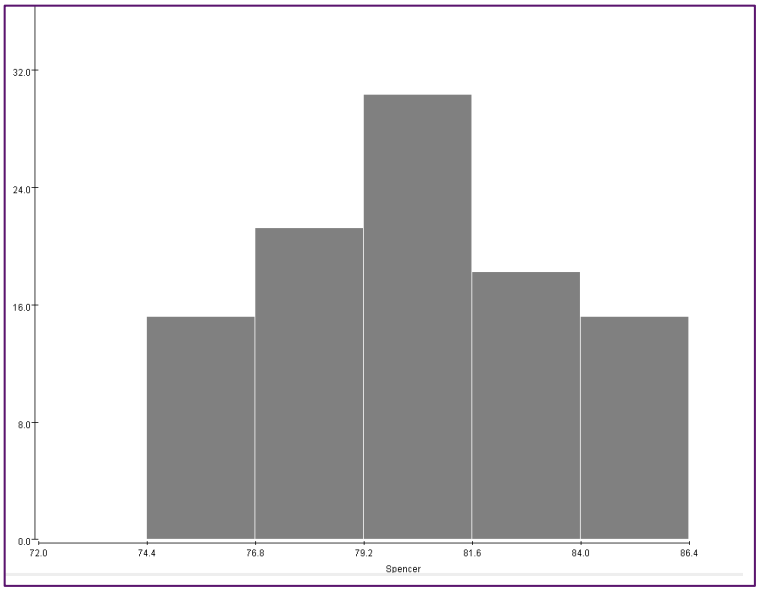


Data set VI: Jones' class

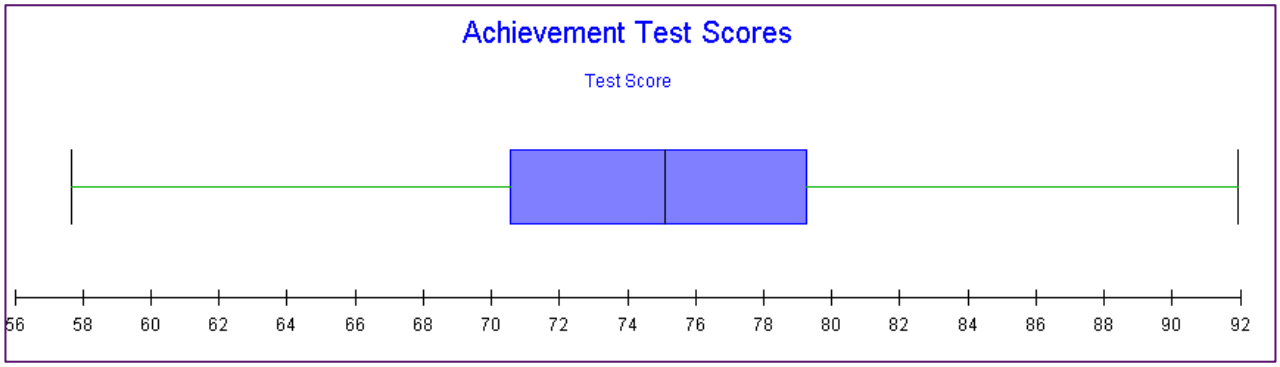


mathematics vision project

Data set VII: Spencer's class



Data set VIII: Overall Achievement Test Scores



9.2 Data Distribution – Teacher Notes

A Practice Understanding Task

Purpose: Students are already familiar with dot plots, box plots, and histograms. This task has them describe data distributions and compare shape, center, and spread of two or more sets of data.

Core Standards Focus:

> **S.ID.2** Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
>
> **S.ID.3** Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Related Standards: S.ID.1

Standards for Mathematical Practice:

> **SMP 3 – Construct viable arguments and critique the reasoning of others**
> This task provides opportunities for students to construct meaning of various sets of data and to compare/contrast data sets. Students share how data is similar and different. At this point, students should be comfortable with the academic vocabulary to describe and compare data sets. Throughout the task, students refine their thinking by listening and critiquing the reasoning of others, but they are also building fluency with comparing sets of univariate data and knowing what information is highlighted and what information is hidden.
>
> **SMP 4 – Model with mathematics**
> The task asks students to interpret test scores by identifying important quantities in each data representation, then to compare data sets. Students will routinely interpret results in terms of the context and also use their knowledge of what each representation highlights to make decisions.
>
> **SMP 8 – Look for and express regularity in repeated reasoning**
> Students interpret data with the given representation. By this time, students recognize that a box plot tells us information about spread and median values, but can be deceiving when it comes to whether or not data is skewed or normal. Students also recognize that histograms tell us about shape and center, but does not tell us details about specific data values the way a dot plot does. Students use this knowledge to make sense of univariate data,

The Teaching Cycle:

Essential Question for Students: *How do I compare two or more data sets based on the statistics appropriate to the shape of the data distribution?*

Note: It would be good to have the data you want to compare in a format that is large and visible for the whole group discussion. For example, you could copy the two data sets you wish to compare and place them next to each other in a format that can be projected so that when students are sharing during whole group, the visual representation is available for everyone to see.

Note: Students have been asked to identify and interpret univariate data using dot plots, histograms, and box plots since sixth grade. In this course, students are asked to **compare** data sets using their knowledge of shape, center, and spread and have become more comfortable with these attributes. Outliers, skewed data, and normal distribution may be new this year as well.

Launch (Whole Class):
Have students read the vocabulary to describe data distributions and ask them to underline information that is new to them. Have them work individually for a while on question 1 that has them describe each data set before having them work together with a partner or small group to answer the remaining questions (where they compare data sets).

Explore (Small Group):
Give students time to answer the questions comparing data sets. Listen for students to use vocabulary in describing a given data set, and to compare shape, center, and spread of two or more data sets. Listen for students to compare data sets, not just list attributes of each. Press students to make comparisons showing they understand when to use data to describe and compare shape, center, and spread between data sets. Examples include noticing outliers, variability and spread between data (notice that Hurlea and Spencer have a scale that is different than the others), and other trends. Again, make sure students do not just list characteristics of each distribution and think they are 'comparing'.

Discuss (Whole Class):
Begin the whole group discussion by selecting problems from questions that compare data sets. Based on small group conversations, choose which comparisons to share out in whole group. The focus of the whole group discussion is to do the following:

- Show student understanding of using statistics appropriate to the shape of the data distribution to compare center and spread.
- Show student understanding of what information is provided when given a histogram, box plot, dot plot.

Exit ticket for students: *Which distributions are most similar? Different? Explain your answer.*

Instructional Supports:
Compare/Contrast: This task provides opportunities for student to recognize the statistics appropriate to the shape of the data distribution to compare center and spread of two or more data sets. Students have visual/graphical representations of test scores from various classes where they can compare spread, range, center (including the relationship between mean and median in both data sets), and variability.

Visual: Have students answer the first question by describing each data set using the visual model given. For each class of data, students can describe the attributes that are evident from the model provided.

Instructional Adaptations:

Intervention Activity: For students who need additional support, have them start by answering questions two through five by comparing like representations (questions that compare two histograms or two box plots). Be explicit about how the two data sets can be compared based on the attributes of each representation (for example, spread: interquartile range for box plots, distribution for histograms). Next have students compare spread and variability between data sets presented in different ways. Finally, have students compare data sets and what, if any, information can be derived about the center.

Challenge Activity: Write a paragraph explaining to a friend how to compare data using shape, center, and spread.

Aligned Ready, Set, Go: Modeling Data 9.2

9.2 Data Distribution – Answer Key

A Practice Understanding Task

1. Answers to include the following:
   Data set I: Williams
   Shape- unimodal, skewed right
   Center- mean is closer to left of data while median would be to the right of the mean.
   Spread- data has a spread of about 60 – 100, meaning the range is approximately 40 points.
   Data set II: Lemon
   Shape- uniform (or multimodal, if argued).
   Center- mean is 81.75, median is 83
   Spread- data has a spread of about 70-93, meaning the range is approximately 23 points.
   Data set III: Croft
   Center- The median appears to be 78. Due to data not being symmetric, the mean is not the same as the median.
   Spread- data has a spread of about 40-100, meaning the range is approximately 60 points. While 50% of the data lies between 70 and 90 (interquartile range of 20), the remaining data has a greater span than most other class data.
   Data set IV: Anderson
   Shape- Unimodal, somewhat bell shaped (close to a normal distribution but not quite symmetric)
   Center- mean is close to 80, with the median falling to the right of the mean
   Spread- data has a spread of about 58 - 104, meaning the range is approximately 46 points (the histogram does not show endpoints like a box plot or dot plot does)
   Data set V: Hurlea
   Center- median appears to be 83 with an interquartile range of 16 points
   Spread- data has a spread of 60- 100, meaning the range is 40 points. Interquartile range is 16 points (spread from 74 to 90)
   Data set VI: Jones
   Shape- unimodal, skewed left
   Center- mean is toward left of data while median would be to the right of the mean.
   Spread- data has an overall range of approximately 60, however most data seems to fall between 78 and 95 (the histogram does not show endpoints like a box plot or dot plot does)

Data set VII: Spencer
Shape- unimodal, normal distribution (mostly symmetric)
Center- mean and median are close to each other (approximately 80.4) due to symmetric nature of graph. Mean may be slightly left.
Spread- data has a spread of about 74-86, meaning the range is approximately 12 points (the histogram does not show endpoints like a box plot or dot plot does). This class has a smaller range than any other class meaning the variability of this class is small compared to others.

2. **Compare Anderson and Williams:** Both classes have about the same spread and variability, however, Williams had more students score toward the bottom end of the data (in the 60's) and Anderson was more bell curved (more students mostly scoring between 80 and 90, with some students scores below and above this amount). Anderson's class would have a higher mean and median score.

3. **Compare Williams and Lemon:** Williams's class has a greater spread and more variability than Lemon's class, yet also has more students scoring low. Williams's class has a range of 40 points, yet more than 50% of the students scored below 75 compared to Lemons range of 23 points but more equally dispersed (33% of students scoring below 75%, 33% of students scoring above 86%, and the remaining 33% of students scoring somewhere in between).

4. **Compare Croft and Hurlea:** While both classes have students who scored high, the range of Croft's data is 60, compared to Hurlea's range of 40 points. If you did not count the one student who scored a 40 in Croft's class, the two classes would have a very similar range. Croft's median of 78 is less than Hurlea's median of 82. The variability and spread among the middle 50% of students in both courses is very similar (from about 73 – 90), although the variability looks greater in Hurlea's class due to the smaller viewing window in Hurlea's class.

5. **Compare Jones, Spencer, and Anderson:** Anderson and Spencer have a more normal, unimodal distribution while Jones has a unimodal, skewed left distribution. Spencer's data has less variability than the others, with a range of 12 points. Anderson and Jones have similar range values if you do not include the scores over the interval of 41-46. Overall, while Spencer's class did not have anyone who scored below 75% (Anderson's low score is between 58 and 63 while Jones had a score in the 40's), Spencer also did not have as many higher scores (his highest score is below an 87%). Jones had more students than Spencer score higher, with over 50% of his students scoring above 87%. Anderson high scores were between the other two classes, with over 40% of students scoring above an 85%.

6. **Compare Spencer and the overall achievement data:** Spencer has less spread, but both sets of data seem to have the same mean and median. This is evident in Spencer's data, whose histogram of test scores resembles a normal curve with a small amount of variability. The mean and median being the same is also evident in the box plot with the data being symmetric. If the achievement test scores have a normal distribution, the standard deviation would be larger than in Spencer's class due to greater variability in scores.

7. Similar distributions are those with similar characteristics of shape, center, and/or spread. Answers may vary for this question, so have students argue for who they believe to have the most similar or different sets of data. Be sure they use data as evidence.

   Examples of those most similar: Anderson and Spencer both with somewhat normal distributions are similar, but not as similar as Anderson and Hurlea, whose median values are both close to 82. Anderson's histogram would look very similar to Hurlea's box plot.

   Examples of those most different: Williams and Jones have similar spreads, however, the majority of scores in Jones class is high while the majority of William's are low. Lemon's class, by contrast of these two, is more evenly distributed. Spencer's class is also different from the others in that there is little variability.

For Use Only in WCPSS

READY, SET, GO! **Name** **Period** **Date**

READY

Topic: Drawing conclusions from data.

In problems 1 – 4 you are to select the best answer based on the given data. Below your chosen answer is a confidence scale. Circle the statement that best describes your confidence in the correctness of the answer you chose. *The goal is to gain awareness of how it seems easier to draw conclusions in some cases than in others.*

1. Data: 1, 2, 4, 8, 16, 32, The next number in the list will be: **c**

a. larger than 32 b. positive c. exactly 64 d. less than 32

I am certain I am correct. I am a little unsure. I had no idea so I guessed.

What about the data made you feel the way you did about the answer you marked?

Answers may vary. Possible answer: Each number in the sequence follows the pattern of multiplying by 2.

Answers may vary.

2. Data: 47, -13, -8, 9, -23, 14, The next number in the list will be: **c**

a. positive b. negative c. less than 100 d. less than -100

I am certain I am correct. I am a little unsure. I had no idea so I guessed.

What about the data made you feel the way you did about the answer you marked?

Answers may vary. Possible answer: Since all the rest of the numbers are less than 100, the next number should probably be less than 100.

Answers may vary.

3. Data: -10, ¾, 38, -10, ½, -81, -10, ¼, 93, -10, The next number in the list will be: **c or d**

a. more than 93 b. negative c. a fraction d. a whole number

I am certain I am correct. I am a little unsure. I had no idea so I guessed.

Answers may vary. Possible answer: The number -10 is always followed by a fraction so the next one could be a fraction, but it could also be 0 since the fractions keep decreasing by ¼ . The last fraction is 1/4; decreased by ¼ is 0.

Answers may vary.

4. Data: 50, -43, 36, -29, 22, -15 The next number in the list will be: **b**

a. odd b. less than 9 c. two-digits d. greater than -15

I am certain I am correct. I am a little unsure. I had no idea so I guessed.

What about the data made you feel the way you did about the answer you marked?

Answers may vary. Possible answer: The next number should be 8. The odd terms are subtracted by 14 and the even terms are added by 14. 22 (the 5th term) subtract 14 is 8, the 7th term.

SET

Topic: Creating histograms.

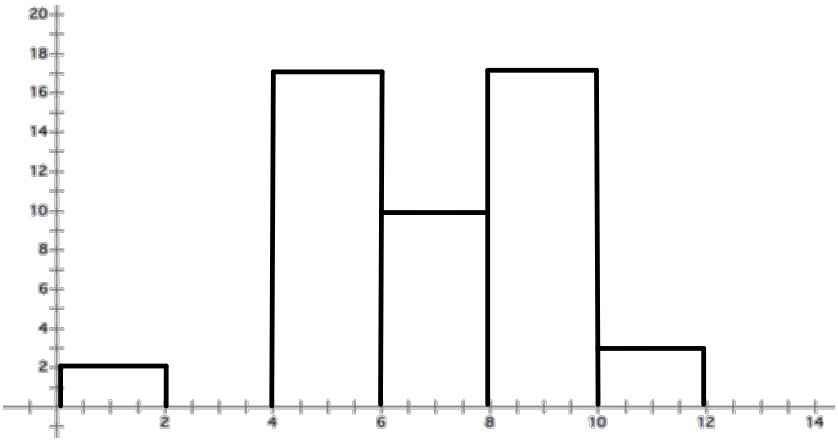
Mr. Austin gave a ten-point quiz to his 9th grade math classes. A total of 50 students took the quiz. Mr. Austin scored the quizzes and listed the scores alphabetically as follows.

1st Period Math	2nd Period Math	3rd Period Math
6, 4, 5, 7, 5, 9, 5, 4, 6, 6, 8, 5, 7, 5, 8, 1, 8, 7, 10, 9	4, 5, 8, 6, 8, 9, 5, 8, 5, 1, 5, 5, 7, 5, 7	9, 8, 10, 5, 9, 7, 8, 9, 8, 5, 8, 10, 8, 8, 5

5. Use ALL of the quiz data to make a frequency table with intervals. Use an interval of 2.

Score	Frequency
0 - 1	**2**
2 – 3	**0**
4 – 5	**17**
6 – 7	**10**
8 – 9	**17**
10-11	**3**

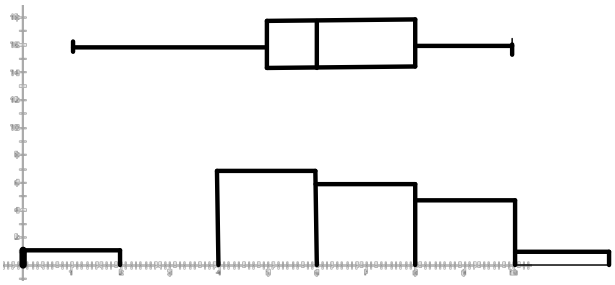
6. Use your frequency table to make a histogram for the data

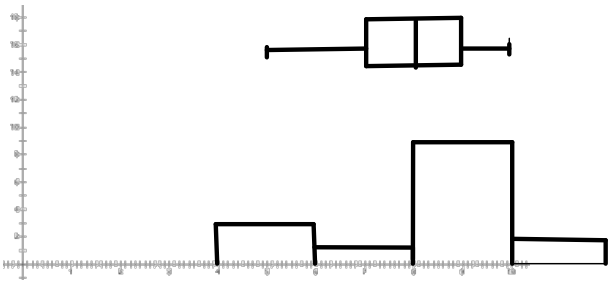


7. Describe the data distribution of the histogram you created. Include words such as: *mode, skewed, outlier, normal, symmetric, center,* and *spread,* if they apply. (Hint: Don't forget standard deviation.)

Answers will vary. Possible answers: Bimodal, symmetric, possible outliers at 1. Range in histogram appears to be 11 or 12 but is actually 9. Mean and median are not very useful since the data is bimodal.

8. Create a graph of your choice (histogram, boxplot, dotplot) for 1st and 3rd period.





Answer: Examples of boxplot and histograms.

9. Which class performed better? Justify your answer by comparing the shape, center, and spread of the two classes. (Hint: Don't forget standard deviation.)

Answers will vary. Possible answers: The 3rd period class performed better. 75% scored above a 7 and their median was 8 while 1st period had a 6 median. 3rd period has a smaller standard deviation since its range is smaller.

GO

Topic: Figuring percentages

10. What percent of 97 is 11?

Answer: 11.3%

11. What percent of 88 is 132?

Answer: 150%

12. What percent of 84 is 9?

Answer: 10.7%

13. What percent of 88.6 is 70?

Answer: 79%

14. What is 270% of 60?

Answer: 162

15. What is 84% of 25?

Answer: 21