

**B – Graphs and Statistics, Lesson 3, Frequency Histograms, Box Plots and Dot Plots (r. 2018)**

**GRAPHS AND STATISTICS**

**Frequency Histograms, Box Plots and Dot Plots**

<p><b>Common Core Standard</b></p> <p><b>S-ID.A.1</b> Represent data with plots on the real number line (dot plots, histograms, and box plots).</p>	<p><b>Next Generation Standard</b></p> <p><b>AI-S.ID.1</b> Represent data with plots on the real number line (dot plots, histograms, and box plots).</p>
---	--

**Overview of Lesson**

<p><b>Teacher Centered Introduction</b></p> <p>Overview of Lesson</p> <ul style="list-style-type: none"> <li>- activate students’ prior knowledge</li> <li>- vocabulary</li> <li>- learning objective(s)</li> <li>- big ideas: direct instruction</li> <li>- modeling</li> </ul>	<p><b>Student Centered Activities</b></p> <p>guided practice ←Teacher: anticipates, monitors, selects, sequences, and connects student work</p> <ul style="list-style-type: none"> <li>- developing essential skills</li> <li>- Regents exam questions</li> <li>- formative assessment assignment (exit slip, explain the math, or journal entry)</li> </ul>
--	--

**LEARNING OBJECTIVES**

Students will be able to:

- 1) Construct and label dot plots, histograms, and box plots above a number line to represent *univariate* data sets.

**VOCABULARY**

univariate  
bivariate  
dot plot

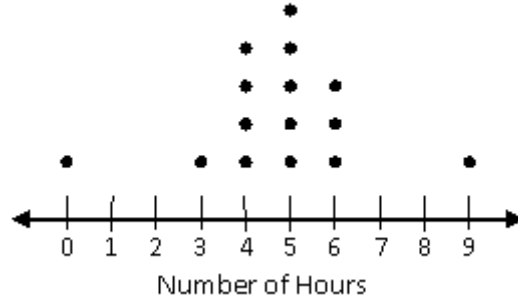
histogram  
box plot  
quartile

**BIG IDEAS**

**Dot Plots**

A dot plot consists of data points plotted on a simple scale. Dot plots are used for continuous, quantitative, *univariate* data. Data points may be labelled if there are few of them. The horizontal axis is a number line that displays the data in *equal intervals*. The frequency of each bar is shown by the number of dots on the vertical axis. Example: This dot plot shows how many hours students exercise each week. Fifteen students were asked how many hours they exercise in one week.

## Hours Exercising per Week

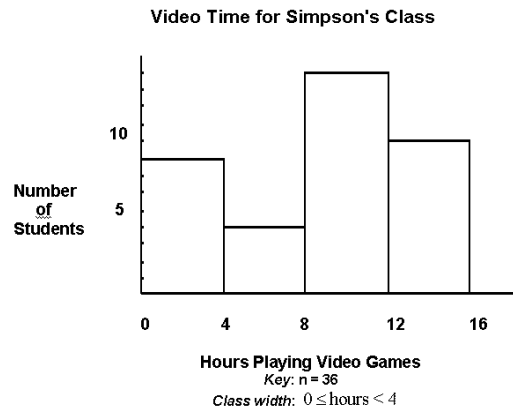


To create a dot plot, draw a number line, then draw one dot above the number line to represent each value in the data set.

### Histograms

A **histogram** is a frequency distribution for continuous, quantitative, univariate data. The horizontal axis is a number line that displays the data in equal intervals. The frequency of each bar is shown on the vertical axis.

**Example:** This histogram shows the number of students in Simpson's class that are in each interval. The students were asked how many hours they spent playing video games in one week.



To create a histogram, first complete a frequency table to show the number of values in intervals of *equal* size. Then draw a number line with equal intervals. Then, plot the frequency for each interval on the vertical axis.

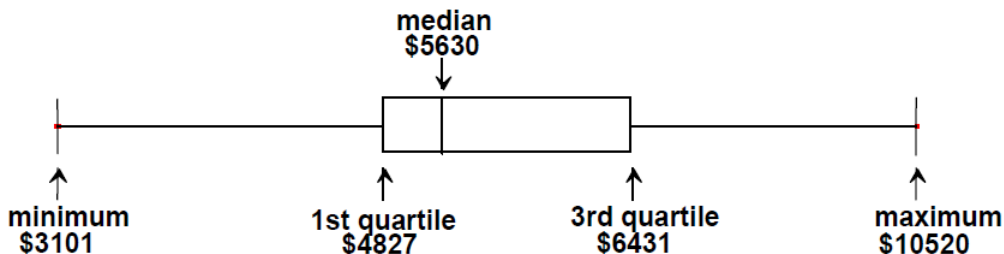
Interval	Tally	Frequency
0-4	III	8
4.01-8		4
8.01-12	III	13
12.01-16	I	11

### Box Plots

A **box plot**, also known as a **box and whiskers chart**, is a visual display of a set of data showing the five number summary: minimum, first quartile, median, third quartile, and maximum. A **box plot** shows the range of scores within *each quarter* of the data. It is useful for examining the variation in a set of data and comparing the variation of more than one set of data.

**Example:**

### Annual food expenditures per household in the U.S. in 2005



To create a box plot, use one-variable stats in a graphing calculator and plot the minimum, Q1, Q2, Q3, and maximum values on a *number line*. Draw boxes around the middle two quartiles. Connect the boxes to the minimum and maximum using lines.

### DEVELOPING ESSENTIAL SKILLS

1. Create a dot plot to represent the following information.

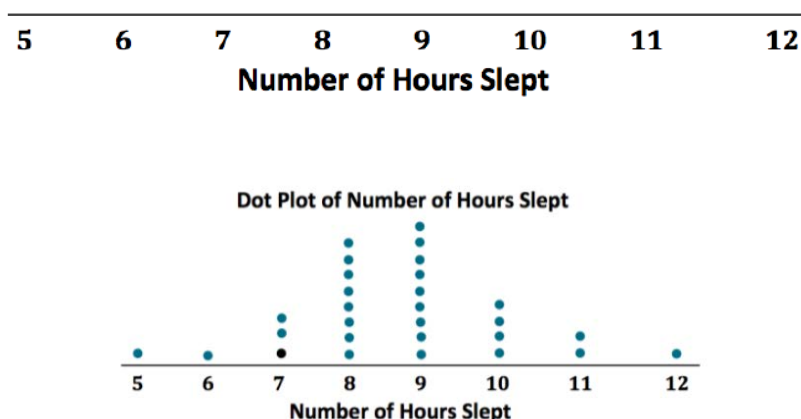
Robert, a sixth grader at Roosevelt Middle School, usually goes to bed around 10:00 p.m. and gets up around 6:00 a.m. to get ready for school. That means he gets about **8** hours of sleep on a school night. He decided to investigate the statistical question: How many hours per night do sixth graders usually sleep when they have school the next day?

Robert took a survey of **29** sixth graders and collected the following data to answer the question.

**7 8 5 9 9 9 7 7 10 10 11 9 8 8 8 12 6 11 10 8 8 9 9 9 8 10 9 9 8**

Robert decided to make a dot plot of the data to help him answer his statistical question. Robert first drew a number line and labeled it from **5** to **12** to match the lowest and highest number of hours slept. Robert's datum is not included.

#### Dot Plot of Number of Hours Slept



SOURCE: Engage New York

2. Create a histogram to represent the following data table.

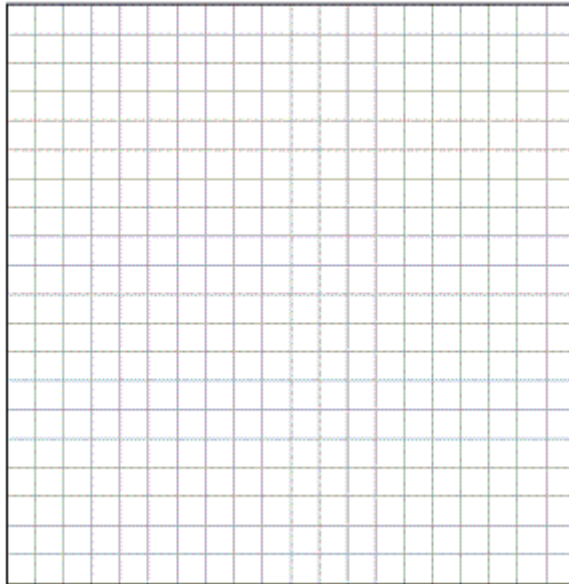
The Fahrenheit temperature readings on 30 April mornings in Stormville, New York, are shown below.

41°, 58°, 61°, 54°, 49°, 46°, 52°, 58°, 67°, 43°,  
 47°, 60°, 52°, 58°, 48°, 44°, 59°, 66°, 62°, 55°,  
 44°, 49°, 62°, 61°, 59°, 54°, 57°, 58°, 63°, 60°

Using the data, complete the frequency table below.

Interval	Tally	Frequency
40–44		
45–49		
50–54		
55–59		
60–64		
65–69		

On the grid below, construct and label a frequency histogram based on the table.

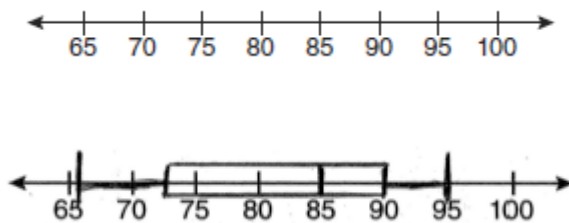


3. Create a box plot to represent the following data.

The test scores from Mrs. Gray's math class are shown below.

72, 73, 66, 71, 82, 85, 95, 85, 86, 89, 91, 92

Construct a box-and-whisker plot to display these data.



**REGENTS EXAM QUESTIONS**

## S.ID.A.1: Frequency Histograms, Box Plots and Dot Plots

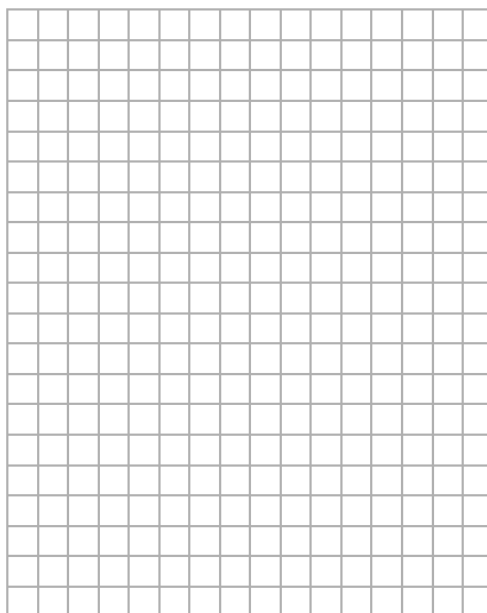
18) The heights, in feet, of former New York Knicks basketball players are listed below.

6.4 6.9 6.3 6.2 6.3 6.0 6.1 6.3 6.8 6.2  
6.5 7.1 6.4 6.3 6.5 6.5 6.4 7.0 6.4 6.3  
6.2 6.3 7.0 6.4 6.5 6.5 6.5 6.0 6.2

Using the heights given, complete the frequency table below.

Interval	Frequency
6.0-6.1	
6.2-6.3	
6.4-6.5	
6.6-6.7	
6.8-6.9	
7.0-7.1	

Based on the frequency table created, draw and label a frequency histogram on the grid below.



Determine and state which interval contains the upper quartile. Justify your response.

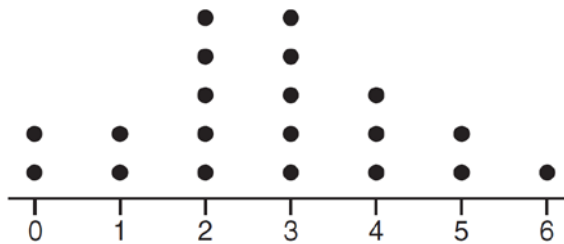
19) Robin collected data on the number of hours she watched television on Sunday through Thursday nights for a period of 3 weeks. The data are shown in the table below.

	Sun	Mon	Tues	Wed	Thurs
Week 1	4	3	3.5	2	2
Week 2	4.5	5	2.5	3	1.5
Week 3	4	3	1	1.5	2.5

Using an appropriate scale on the number line below, construct a box plot for the 15 values.

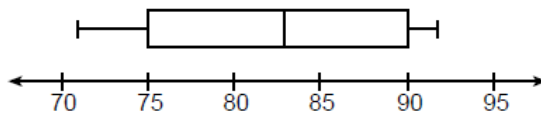


- 20) Which statistic can not be determined from a box plot representing the scores on a math test in Mrs. DeRidder's algebra class?
- 1) the lowest score
  - 2) the median score
  - 3) the highest score
  - 4) the score that occurs most frequently
- 21) The dot plot shown below represents the number of pets owned by students in a class.



Which statement about the data is *not* true?

- 1) The median is 3.
  - 2) The interquartile range is 2.
  - 3) The mean is 3.
  - 4) The data contain no outliers.
- 22) The box plot below summarizes the data for the average monthly high temperatures in degrees Fahrenheit for Orlando, Florida.



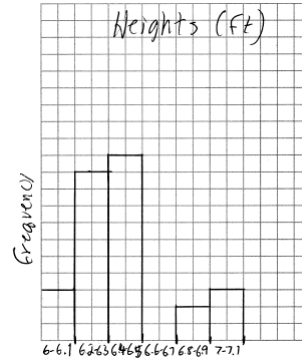
The third quartile is

- 1) 92
- 2) 90
- 3) 83
- 4) 71

### SOLUTIONS

18) ANS:

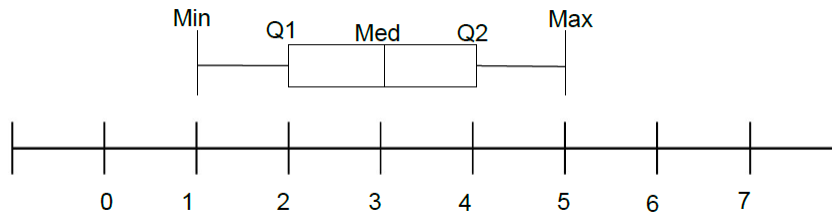
Interval	Frequency
6.0 - 6.1	3
6.2 - 6.3	10
6.4 - 6.5	11
6.6 - 6.7	0
6.8 - 6.9	2
7.0 - 7.1	3



Each quartile contains  $\frac{1}{4}$  or 25% of the data values. There are a total of 29 data values in the data set, so each quartile will contain  $\frac{29}{4} = 7.25$  values. The upper quartile will begin 7.25 values from the maximum, which places the upper quartile in the 6.4-6.5 interval.

PTS: 4                    NAT: S.ID.A.1                    TOP: Frequency Histograms  
 KEY: frequency histograms

19) ANS:



Strategy #1: Input all the numbers from the table in a TI 83+ graphing calculator, then calculate 1 variable stats, then use the calculator output to construct the box and whiskers plot.

Strategy #2 Follow these step-by-step procedures for creating a box and whiskers plot.

STEP 1. Organize the data set in ascending order, as follows. Be sure to include all the data.:  
 1, 1.5, 1.5, 2, 2, 2.5, 2.5, 3, 3, 3, 3.5, 4, 4, 4.5, 5

STEP 2. Plot a scale on the number line. In this case, the scale is 0 to five in equal intervals of .5 units.

STEP 3. Plot the minimum and maximum values: minimum = 1 and maximum = 5.

STEP 4. Identify the median. In this problem, there are fifteen numbers and the median is the middle number, which is 3. There are seven numbers to the left of 3 and seven numbers to the right of 3.

STEP 5. Plot and label the median = 3 (also known as Q2 or the second quartile).

STEP 6. Identify Q1, which is the *median of the bottom half* of the organized data set. The bottom half of the data includes all numbers below the median, which in this problem, includes the following numbers  
 1, 1.5, 1.5, 2, 2, 2.5, 2.5

The middle number in an organized list of seven numbers is the fourth number, which in this case is a 2.

STEP 7. Plot and label  $Q1 = 2$ .

STEP 8. Identify  $Q3$ , which is the *median of the top half* of the organized data set. The top half of the data includes all numbers above the median, which in this problem, includes the following numbers

3, 3, 3.5, 4, 4, 4.5, 5

Again, the middle number in an organized list of seven numbers is the fourth number, which in this case is a 4.

STEP 9. Plot and label  $Q3 = 4$ .

STEP 10. Finish the box plot by drawing boxes between the plotted points for  $Q1$ ,  $Q2$ , and  $Q3$ .

PTS: 2                    NAT: S.ID.A.1            TOP: Box Plots

20) ANS: 4

A box plot is also known as a box and whiskers chart and shows the following five statistics:

1. The minimum score.
2.  $Q1$ , which is the top of the first quartile.
3.  $Q2$ , which is also the median score and the top of the second quartile.
4.  $Q3$ , which is the top of the third quartile.
5. The maximum score.

The interquartile range can be determined by subtracting  $Q1$  from  $Q2$ .

PTS: 2                    NAT: S.ID.A.1

21) ANS: 3

Step 1. Understand that the problem is asking you to apply different statistical measures to the data in the dot plot and find the one answer choice that is not true.

Step 2. Strategy: Evaluate each answer choice and eliminate wrong answers.

Step 3. Execution of Strategy

a) To evaluate this answer choice, the median (middle) of the ordered data elements must be identified. There are 20 dots, so the middle is somewhere between the 10th and 11th dots. Counting 10 dots from either end, the median will occur in the 3 column. The median is 3, so answer a) must be eliminated.

b) To evaluate this answer, the interquartile range must be calculated. The interquartile range is defined as the distance between the first and third quartiles in an ordered distribution. The dot plot has 20 dots. Since each quartile contains 25% of the dots, each quartile will contain 25% of 20 dots, which equals 5 dots.

$Q1$  ends after five dots, so  $Q1=2$ .

$Q2$  ends after 10 dots, so  $Q2=10$ .

$Q3$  ends after 15 dots, so  $Q3=4$ .

The interquartile range is computed as  $Q3-Q2$ . In this dot plot, the interquartile range is 2, so answer b) is true and must be eliminated.

c) The mean for this data plot can be calculated as follows:

$$\bar{X} = \frac{0+0+1+1+2+2+2+2+2+2+3+3+3+3+3+4+4+4+5+5+6}{20}$$

$$\bar{X} = \frac{55}{20}$$

$$\bar{X} = 2.75$$

Answer c) is not true, because the mean of this data set is 2.75. Therefore, answer choice c) is the correct answer.

d) The data has no outliers. This is true by inspection. All the data is close together and there are no large gaps between the data. Hence, there are no outliers and choice d) must be eliminated.



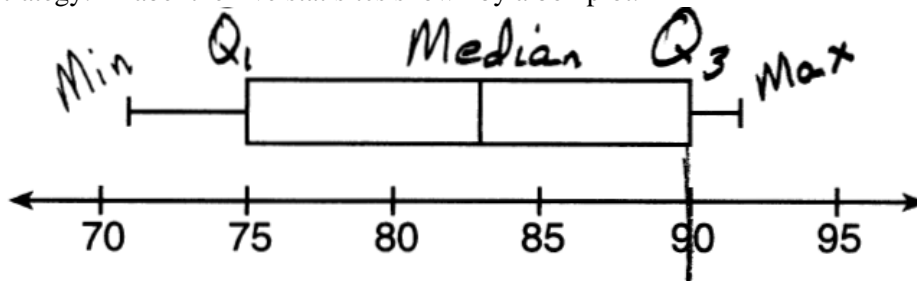
Step 4. Does it make sense? Yes. Three answer choices have been shown to be true and one answer choice has been shown to be false. The statement that is not true is choice c).

median = 3, IQR = 4 - 2 = 2,  $\bar{x} = 2.75$ . An outlier is outside the interval  $[Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$ .  
 $[2 - 1.5(2), 4 + 1.5(2)]$   
 $[-1, 7]$

PTS: 2 NAT: S.ID.A.1 TOP: Dot Plots

22) ANS: 2

Strategy: Label the five statistics shown by a box plot.



PTS: 2 NAT: S.ID.A.1 TOP: Box Plots KEY: interpret

23) ANS: 2

The number of pages a paper will have does not depend on how fast the student types.

PTS: 2 NAT: S.ID.C.9 TOP: Analysis of Data

24) ANS: 2

Strategy: Eliminate wrong answers.

Observe: Both variables (number of pages and amount of ink) increase together, so the correlation is positive.

Eliminate answer choices with negative correlation.

Reason: Printing causes ink to be used, so the relationship is causal. Eliminate answer choices with non-causal.

- a) positive correlation, but ~~not~~ causal
- b) positive correlation, and causal
- e) ~~negative~~ correlation, but not causal
- d) ~~negative~~ correlation, and causal

PTS: 2 NAT: S.ID.C.9 TOP: Analysis of Data