S.ID.A.1 Dot Plots, Histograms, and Box Plots

GRAPHS AND STATISTICS
S.ID.A.l: Dot Plots, Histograms, and Box Plots

A. Summarize, represent, and interpret data on a single count or measurement variable
1. Represent data with plots on the real number line (dot plots, histograms, and box plots).
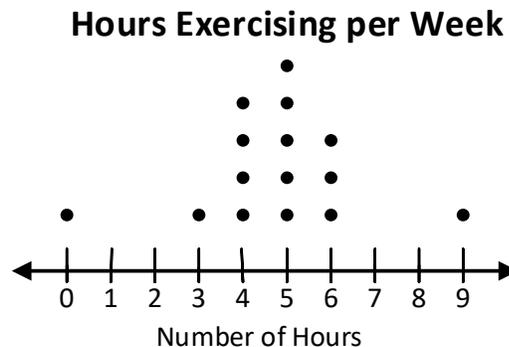
Overview of Lesson
- activate prior knowledge and review learning objectives (see above)
- explain vocabulary and/or big ideas associated with the lesson
- connect assessment practices with curriculum
- model an assessment problem and solution strategy
- facilitate guided discussion of student activity
- facilitate guided practice of student activity
          Selected problem set(s)
- facilitate a summary and share out of student work
          Homework – Write the Math Assignment

univariate A set of data involving one variable.

multivariate A set of data involving more than one variable.

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.
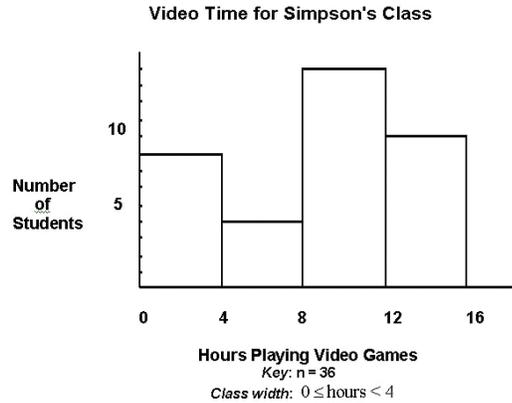
Example: This **dot plot** shows how many hours students excercise each week. Fifteen students were asked how many hours they exercise in one week.

Hours Exercising per Week

```
                      •
                 •    •
                 •    •    •
                 •    •    •
        •        •    •    •              •
   <----+----+----+----+----+----+----+----+----+---->
        0    1    2    3    4    5    6    7    8    9
                      Number of Hours
```

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.

Video Time for Simpson's Class

Number
of
Students

10

5

0   4   8   12   16

Hours Playing Video Games
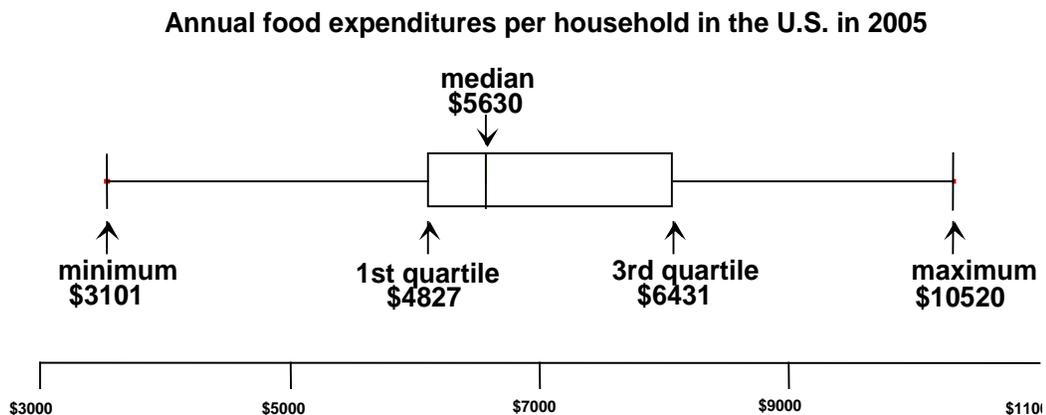*Key*: n = 36
Class width: $0 \leq hours < 4$

A **box plot**, also known as a **box and whiskers chart**, is a visual display of a set of data above a number line showing the five number summary:

1. Minimum – this is the smallest number in the data set.
2. First Quartile (Q1) – this is the midpoint of the *lower* half of the data set.
3. Median (Q2) – this is the midpoint of the data set.
4. Third Quartile (Q3) – this is the midpoint of the *upper* half of the data set.
5. Maximum – this is the largest number in the data set.

A **box plot** shows the range of scores within each quarter of the data. It is useful for examining the variation within a set of data and comparing the variation of more than one set of data.

The **interquartile range** is the distance on the number line between Q1 and Q2.
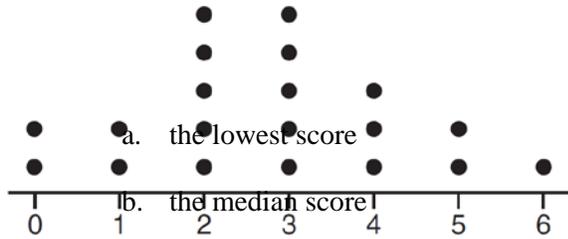
Example:

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

median
$5630

minimum
$3101

1st quartile
$4827

3rd quartile
$6431

maximum
$10520

$3000 $5000 $7000 $9000 $1100

REGENTS PROBLEMS TYPICAL OF THIS STANDARD

1. The dot plot shown below represents the number of pets owned by students in a class.

Lesson Plan

a. the lowest score   c. the highest score

b. the median score   d. the score that occurs most frequently

Which statement about the data is *not* true?
a. The median is 3.   c. The mean is 3.
b. The interquartile range is 2.   d. The data contain no outliers.

2.   Which statistic can not be determined from a box plot representing the scores on a math test in Mrs. DeRidder's algebra class?

3.   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.

S.ID.A.1 Dot Plots, Histograms, and Box Plots
Answer Section

1. ANS: C
    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.
    $$Q1 \text{ ends after five dots, so } Q1=2.$$
    $$Q2 \text{ ends after 10 dots, so } Q2=10.$$
    $$Q3 \text{ ends after 15 dots, so } Q3=4.$$
    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:

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

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

    $$\overline{X} = 2.75$$

    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).

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

    $[-1, 7]$

    PTS: 2   REF: 061620ai   NAT: S.ID.A.1   TOP: Dot Plots
2. ANS: D
    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.
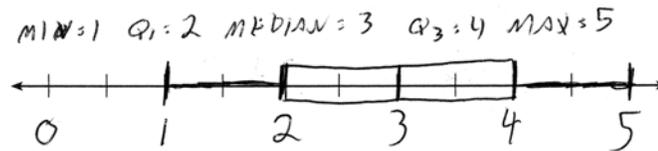   The interquartile range can be determined by subtracting Q1 from Q2.

    PTS: 2   REF: 081603ai   NAT: S.ID.A.1

Lesson Plan

3.   ANS:

MIN = 1   Q₁ = 2   MEDIAN = 3   Q₃ = 4   MAX = 5

$MIN = 1 \quad Q_1 = 2 \quad MEDIAN = 3 \quad Q_3 = 4 \quad MAX = 5$

0   1   2   3   4   5

Strategy: 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 = 2.

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   REF: 061432ai   NAT: S.ID.A.1   TOP: Box Plots

Homework - Write the Math Assignment

START   Write your name, date, topic of lesson, and class on your paper.
       NAME: Mohammed Chen
       DATE: December 18, 2015
       LESSON: Missing Number in the Average
       CLASS: Z

PART 1a.   Copy **the problem** from the lesson and underline/highlight key words.
PART 1b.   State your understanding of **what the problem is asking**.
PART 1c.   **Answer** the problem.
PART 1d.   Explanation of **strategy** with all work shown.

PART 2a.   Create **a new problem** that addresses the same math idea.
PART 2b.   State your understanding of **what the new problem is asking**.
PART 2c.   **Answer** the new problem.
PART 2d.   Explanation of **strategy** used in solving the new problem with all work shown.

Clearly label each of the eight parts.

Grading Rubric
Each homework writing assignment is graded using a four point rubric, as follows:

Part 1. The Original Problem	**Up to 2** points will be awarded for: a) correctly restating the original problem; b) explicitly stating what the original problem is asking; c) answering the original problem correctly; and d) explaining the math.
Part 2. My New Problem	**Up to 2** points will be awarded for: a) creating a new problem similar to the original problem; b) explicitly stating what the new problem is asking; c) answering the new problem correctly; and d) explaining the math.

This assignment/activity is designed to incorporate elements of [Polya's four step universal algorithm](#) for problem solving with the idea that writing is thinking. Polya's four steps for solving any problem are:
1. Read and understand the problem.
2. Develop a strategy for solving the problem.
3. Execute the strategy.
4. Check the answer for reasonableness.

Lesson Plan

EXEMPLAR OF A WRITING THE MATH ASSIGNMENT

Part 1a. The Problem
TOP Electronics is a small business with five employees. The mean (average) weekly salary for the five employees is $360. If the weekly salaries of four of the employees are $340, $340, $345, and $425, what is the salary of the fifth employee?

Part 1b. What is the problem asking?
Find the salary of the fifth employee.

Part 1c. Answer
The salary of the fifth employee is $350 per week.

Part 1d. Explanation of Strategy
The arithmetic mean or average can be represented algebraically as:

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n}$$

I put information from the problem into the formula. The problem says there are 5 employees, so $n = 5$. The problem also gives the mean (average) salary and the salaries of 4 of the employees. These numbers can be substituted into the formula as follows:

$$360 = \frac{340 + 340 + 345 + 425 + x_5}{5}$$

$$1800 = 340 + 340 + 345 + 425 + x_5$$

$$1800 = 1450 + x_5$$

$$1800 - 1450 = x_5$$

$$350 = x_5$$

Check: $360 = \dfrac{340 + 340 + 345 + 425 + 350}{5} = \dfrac{1800}{5} = 360$

Part 2a. A New Problem
Joseph took five math exams this grading period and his average score on all of the exams is 88. He remembers that he received test scores of 78, 87, 94, and 96 on four of the examinations, but he has lost one examination and cannot remember what he scored on it. What was Joseph's score on the missing exam?

Part 2b What is the new problem asking?
Find Joseph's score on the missing exam.

Part 2c. Answer to New Problem
Joseph received a score of 85 on the missing examination.

Part 2d. Explanation of Strategy
I substitute information from the problem into the formula for the arithmetic mean, as follows:

$$88 = \frac{78 + 87 + 94 + 96 + x_5}{5}$$

$$440 = 355 + x_5$$

$$85 = x_5$$

$$88 = \frac{78 + 87 + 94 + 96 + 85}{5} = \frac{440}{5} = 88$$

The answer makes sense.