Representing Data Using Boxplots

Strand: Probability and Statistics

Topic: Collect and analyze data using a boxplot.

Primary 2023 SOL: **8.PS.2 The student will apply the data cycle (formulate questions; collect or acquire data; organize and represent data; and analyze data and communicate results) with a focus on boxplots.**

a) Formulate questions that require the collection or acquisition of data with a focus on boxplots.
b) Determine the data needed to answer a formulated question and collect the data (or acquire existing data) using various methods (e.g., observations, measurement, surveys, experiments).
c) Determine how statistical bias might affect whether the data collected from the sample is representative of the larger population.
d) Organize and represent a numeric data set of no more than 20 items, using boxplots, with and without the use of technology.
e) Identify and describe the lower extreme (minimum), upper extreme (maximum), median, upper quartile, lower quartile, range, and interquartile range given a data set, represented by a boxplot.
f) Describe how the presence of an extreme data point (outlier) affects the shape and spread of the data distribution of a boxplot.
g) Analyze data represented in a boxplot by making observations and drawing conclusions.
h) Compare and analyze two data sets represented in boxplots.
i) Given a contextual situation, justify which graphical representation (e.g., pictographs, bar graphs, line graphs, line plots/dot plots, stem-and-leaf plots, circle graphs, histograms, and boxplots) best represents the data.
j) Identify components of graphical displays that can be misleading.

Materials

- Different Data Representations activity sheet (attached)
- Problem 1: Creating a Boxplot (Box-and-Whisker Plot) activity sheet (attached)
- Problem 2: Boxplots activity sheet (attached)
- Problem 3: Boxplots activity sheet (attached)
- Rulers
- Calculators

Vocabulary

- From earlier grades*: extreme data values, mean, median, range*
- *interquartile range, lower quartile, maximum, minimum, upper quartile*

Student/Teacher Actions: What should students be doing? What should teachers be doing?

1. Distribute the Data Representations activity sheet.
   a. Part 1: Have students complete the analysis of the line plots independently or with a partner. Students will have prior experience with line plots. As a class, discuss the answers to questions 2 and 3. In particular, note that the RANGE of the data is the same, but the distribution or spread of the data is different. Discuss how a line plot helps you see this. Discuss the mean found for each data set. Discuss how this measure of center is helpful or not in describing this set of data.
   b. Part 2: Students may or may not have seen a boxplot before – that is okay for now. As a class, discuss and answer the questions related to the boxplots. Have students share questions (or conclusions that you form into questions) related to the boxplots. Highlight what types of questions you would NOT be able to answer with this representation as well. *You want students to recognize that a boxplot does not show individual data points like a line plot did, but you can see the range of the data. You want them to make some hypothesis about what the different "lines" on the boxplot represent in relation to the data.*

2. Distribute the Boxplot activity sheets (1, 2, and 3). Have students follow along as you complete the first problem in creating a boxplot (box-and-whisker plot) with the whole class. Students will not need familiarity with boxplots in order to complete the first page. As a class, discuss the answers and reasoning behind the answers to the first problem. *Note: This problem has an odd number of data entries. Make sure to point that out to the students.*

3. Have students complete the second problem in pairs or small groups. As a class, discuss the answers and reasoning behind the answers to the second problem. As students are discussing how they compared the two boxplots, listen for and encourage talk about percentages (e.g., 75 percent of the students scored higher than 77 percent), the spread of the data (e.g., interquartile range, range), Class 2 not having a right whisker, and the boxplot not showing the mean. *Note: This problem has odd-numbered and even-numbered sets of data. Make sure you go over the differences between the two.*

4. *Have students complete the third problem in groups of four. To reduce the time spent on constructing the boxplots, direct two students in each group to create boxplots of the men's data and check accuracy, while the other two create boxplots of the women's data. In addition, consider using this data set as an opportunity to use technology to create the boxplots for this data set. Specifically, the teacher can model how to do this with one set of data and have student groups use technology such as Desmos to create the box plots for the other two sets of data. Make sure all students are using the same number line in order to make it possible to compare boxplots later. As groups make conclusions about the experiment, encourage them to use of the boxplots to justify their conclusions. Have each group write their conclusions on chart paper for whole-*

class discussion. During class discussion of the third problem, allow each group to present and justify a conclusion they made. Encourage discourse by asking students to restate their conclusions and the reasoning behind them and asking other students whether they agree or disagree with the conclusions and/or reasoning. *Note: This problem has four data sets, all with even numbers of data.*

Assessment

- **Questions**
    - What information can you obtain from a boxplot?
    - What is visible in a boxplot versus what is not visible?
    - For information that is not visible, can you determine it from the boxplot?
    - What information can you *not* obtain from a boxplot?
    - In what situation might you want to use a boxplot?
    - What conclusions can be drawn about the distribution of data in a boxplot and what conclusions cannot be drawn?

- **Journal/Writing Prompts**
    - Explain the advantages and disadvantages of using boxplots.
    - Describe how you might compare two boxplots.

- **Other**
    - Have students gather data related to two groups and present the data in boxplots. Then, have them analyze and compare the plots, noting their observations. Finally, have them exchange plots, analyze each other's plots, and compare observations. Data collection could involve students gathering information from their classmates and then splitting the data into two groups based on gender. Other possibilities include gathering data about two sports teams, salary data of men and women, life expectancies in two different areas of the world, etc.

Extensions and Connections (for all students)

- After gathering and displaying data in boxplots, have students eliminate an outlier to see what effect that has on the plot.
- Have students explore what happens to the mean when an outlier is removed.
- Have students represent the same data with a different graphical representation. Have them explore the difference between the representations and list any advantages or disadvantages of one over the other.
- Have students create boxplots of their quiz grades. Then, have them compare these plots to plots of their grades in other subjects or to plots of their grades during a previous marking period.

Strategies for Differentiation

- Have students create a graphic organizer illustrating the process of constructing a boxplot.
- Allow students to use graph paper when constructing boxplots.

- Provide number lines (*x*-axis) and straight edges to those students who need them.
- Clarify and simplify the description of data and directions for problem 3.

Note: The following pages are intended for classroom use for students as a visual aid to learning.

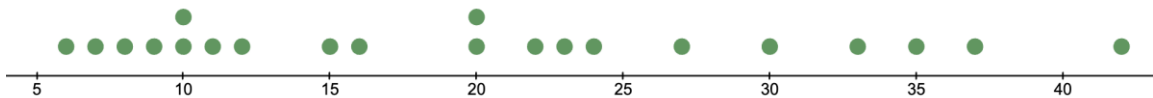Virginia Department of Education © 2018 (Revised December 2023)

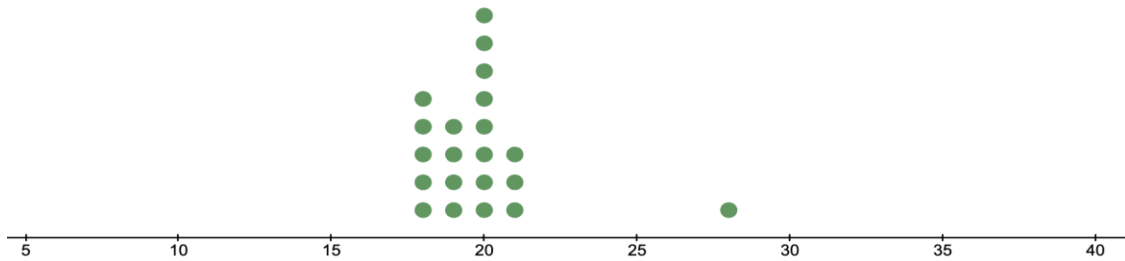# Different Data Representations

Name _____ **Date**_____

Part One:

The data below shows the number of points scored in each game across a season for Player A and Player B.
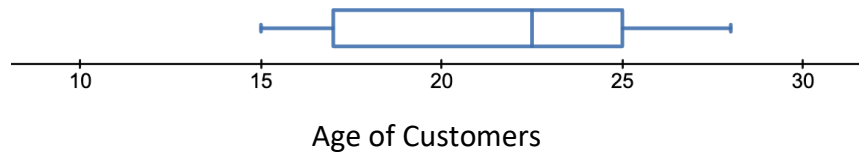
Player A

Player B

1) Find the mean of each data set. Circle it on the line plot.

2) Look at the data represented by the two line plots above. What do you notice about the distribution of the data in the two plots?
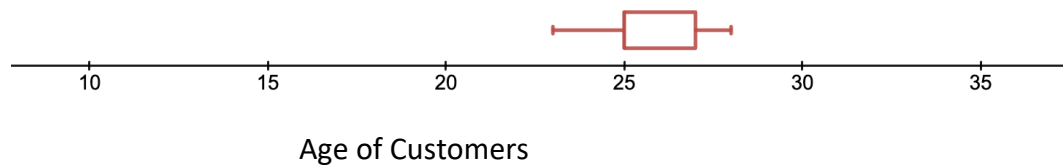
Part Two:

The boxplots below show the ages of customers at two stores, Store A and Store B.

Store A:

Age of Customers

Store B:

Age of Customers

1) What questions could you answer by looking at the boxplots?

2) What do you think the different lines tell you about the data that was used to create these boxplots?

3) How is the representation of the data in the boxplot and the questions you can answer similar to and different from the representation of the data in the line plots in Problem 1?

Problem 1: Creating a Boxplot (Box-and-Whisker Plot)

Name _____ **Date**_____

A restaurant owner wants to know how far his customers are traveling to eat at his restaurant. He recorded their miles traveled for a day.

19, 2, 3, 7, 3, 1, 23, 5, 10, 15, 11, 18, 3, 2, 5, 12, 14

Use the steps below to make a boxplot for this data.

1. Start by drawing an **evenly scaled** number line horizontally along the bottom of a lined piece of paper. This should represent your data appropriately. It should start just below the minimum piece of data and end just after the maximum piece of data. It should be evenly scaled, by an amount that is most helpful.

2. Label or title this horizontal axis.

3. Sort your data from the smallest to the largest value. Then find these five pieces of data: **minimum (min), lower (first) quartile (LQ), median, upper (third) quartile (UQ), and maximum (max)**. *See note below to help with calculations.

4. Two lines **above** the number line, place points at the min and max.
5. From one to three lines **above** the number line, draw a vertical line at the each of the following: lower quartile (LQ), median, and upper quartile (UQ).

6. To form the "**box,**" draw horizontal lines connecting the top of the LQ to the top of the UQ and the bottom of the LQ to the bottom of the UQ. Note: these lines will also touch the top and bottom of the median vertical line.

7. The "**whiskers**" are drawn from the center of the sides of the box to the points above the min and max.

8. Title the graph.

Note: To find the minimum (min), lower (or first) quartile (LQ), median, upper (or third) quartile (UQ), and the maximum (max):
*   To find the **median**, line up the data from least to greatest. The median is the middle piece of data. If there are an even number of data values, take a mean of the middle two values.
*   The median divides the data into two halves. To divide the data into quarters (finding the upper and lower quartiles), you must find the medians of these two halves. Note: If you have an even number of values (meaning the first median was the average of the two middle values), then you include the middle values in your submedian computations. If you have an odd number of values (meaning the first median was an actual data point), then you do not include that value in your submedian computations.
*   The **minimum**, or min, is the smallest value.
*   The **maximum**, or max, is the largest value.

Problem 2: Boxplots

Name _____ Date_____

Scores on a mathematics test for two classes are as follows:

Class 1

Student	A	B	C	D	E	F	G	H	I	J
Score	55	64	83	92	100	77	86	95	80	98

Class 2

Student	A	B	C	D	E	F	G	H	I
Score	52	79	71	100	100	76	100	78	76

1) Make a boxplot of each set of data below.

2) Identify the min, lower quartile (Q1), median, upper quartile (Q3), and max of each.

	Class 1	**Class 2**
min		
Q1		
median		
Q3		
max		

3) Which class did better? Justify your answer.

4) What is the average (mean) score for each class?

5) Does this change your opinion about which class did better? Why, or why not?

6) Can the wording affect how one perceives the overall picture of "which class did better?" If so, in what way?

Problem 3: Boxplots

Name _____ Date_____

An experiment found a significant difference between men and women regarding their abilities to identify unseen objects held in their left hands. The left hand is controlled by the right side of the brain, while the right hand is controlled by the left side of the brain. The test involved 20 small objects that participants were not allowed to see. First, they held 10 of the objects, one by one, in their left hands and guessed what they were holding. Then, they held the other 10 objects, one by one, in their right hands and guessed what they were holding.

Correct Guesses

Women, Left	Women, Right	Men, Left	Men, Right
8	4	7	10
9	1	8	6
10	8	7	10
6	9	5	10
10	6	7	7
8	10	8	9
9	4	10	10
7	9	4	8
9	8	10	10
10	9	8	9

1) Make boxplots that will allow you to compare the data.

2) After analyzing the boxplots and comparing them to one another, draw conclusions about the experiment.

3) Explain why a boxplot is or is not an appropriate representation to show this data set. Justify your reasoning.