Unit Title: Statistics & Data Analysis

Previous Knowledge: This unit builds off students' knowledge of representing and interpreting data first introduced in the fourth- and fifth-grade standards in the Measurement & Data domain of the Texas Essential Knowledge and Skills for Mathematics (TEKS) and Common Core State Standards (CCSS), as well as the Data Analysis & Probability domain of National Council of Teachers of Mathematics (NCTM) standards for grades 3-5.

Unit Overview: In sixth grade, students will develop a deeper understanding of data representations and variability. While some students may already be familiar with some measures of central tendency, this will be the first time that mean, median, and mode are officially introduced as a means of summarizing the center of a data set.

Guiding Questions:
- What is a statistical question, and what does it mean that a question can be answered by numerical data?
- How can you represent numerical data in the most useful way?
- How can you use measures of central tendency and variability to interpret data?

Unit Project Idea: Students will formulate and investigate their own statistical question in groups of 2-3. After gathering data on this question, they will represent their findings graphically and interpret the results using center and spread. These results will be presented as a news story where students act as journalists reporting about their question. This could be video recorded. Their analysis should address the following ideas:
- Measures of central tendency (mean, median, and mode) and spread (mean absolute deviation, range, and interquartile range)
- The five-number summary and its relationship to box plots
- Advantages and disadvantages of displaying data as dot plots, box plots, and histograms

Help students formulate and select appropriate questions for their project by using the resources provided by Open Education Resources Commons. A student-facing introduction handout and rubric is provided.

Objectives:
- ● Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers (6.SP.A.1).

Key Ideas:
- ● Numerical data are data values that are measured using digits, while categorical data are values that can be sorted into groups or categories (usually, but not always, non-numeric).
- ● A statistical question can be answered by collecting data where there will be variability. The phrasing of the question makes a difference in the possible answers being variable or deterministic. For example, "How many text messages did students in this class send yesterday?" is a statistical question because there is anticipated variability in the number of text messages sent by each student. "How many text messages did I send yesterday?" is a non-statistical question because there is only possible answer.

Vocabulary: statistical question, categorical data, numerical data

Lesson: Describing Shape, Center, and Spread

Audience: Sixth grade

Standards:

CCSS
- ● 6.SP.A.2 - Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- ● 6.SP.B.4 - Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

NCTM
- ● Select, create, and use appropriate graphical representations of data, including histograms, box plots, and scatterplots.

TEKS
- ● 6.12.B - Use the graphical representation of numeric data to describe the center, spread, and the shape of the data distribution.

Teacher Background Information: Data can be organized in a variety of ways, including (but not limited to) box plots, dot plots, histograms, frequency charts/tables, and scatterplots. Students will first focus on analyzing data that is represented as a dot plot. Sixth-grade students should already be familiar with making line plots to represent whole and fractional data (covered in CCSS.4.MD.B.4 and CCSS.5.MD.B.2). The terms "dot plot" and "line plot" are often used interchangeably to mean a graph that shows the frequency of data on a number line. Each data value is represented by a dot. The number of dots over a particular value tells you how many data points share that value. Therefore, this representation is most helpful for determining the frequency of data at a particular value. Note that this representation is probably not the best when representing large data sets because the process would be time-consuming and have a high potential for error. Although students will not be independently creating a dot plot in this particular lesson, they will be expected to do so in future lessons.

Lesson Objectives: In this lesson, students will learn that numerical questions that generate variability can be further analyzed to determine shape (symmetric, left skew, right skew), center (where most of the data is clustered), and spread (how the data varies and if there are outliers). Representing data on a dot plot makes it easier to see how data is spread out or clustered. Calculating mean and median is new to sixth-grade standards as it had previously been included in many fifth-grade standards prior to CCSS. Because learners are not expected to be familiar with these terms yet, they will focus on describing the center of a data set by observing where most of the data is clustered within the dot plot.

Prerequisite Skills:
- Number lines
- Statistical questions

Materials:
- Projector
- Computer with Internet access
- Pom-pom balls or another small soft item that students can easily throw
- Empty cans/cups/plastic containers
- Yard sticks

Introduction (10-12 minutes)

Dot Plots Warm-Up
a. Display or handout page 4 of the [EngageNY lesson on displaying data distributions,](#) showing exercises 11-14.
b. Ask students to match each numerical question with its corresponding dot plot. Stress that they should be ready to share *why* they matched the dot plots and questions as they did. You may also ask students to explain why these are considered numerical statistical questions to review the definitions of "statistical question" and "numerical data."
c. Go over the answers as a class, selecting students to explain their choices and reasoning. Allow students to question one another rather than jumping to address incorrect/insufficient responses.

Concept Overview
a. Explain that numerical data can be described using **shape**, **center**, and **spread**. Graphical representations, like dot plots, make it easier to see the organization of the data.
b. Introduce these descriptors using the video from "Skill Instruction" on card 2 of the "Describe the center, spread, and shape of data displays" [LearnZillion lesson plan](#).
c. Explain when the graph displays a "cluster" of data (larger frequency of data bunched together) on the right-hand side of the graph and the "tail" (smaller frequency of data) is to the left, the skew is to the left. When the graph displays a "cluster" of data on the left-hand side of the graph and the "tail" is to the right, the skew is to the right.

d. Introduce the term **gap** to indicate a place on the graph where no data values are present. Explain that identifying gaps help determine outliers.

Exploration (15-20 minutes)

Pom-Pom Toss Activity: This activity is adapted from the [Data Analysis Using Mean, Median, Mode, and Range lesson from PBS LearningMedia](#).

a. Pre-arrange students in groups of 2-3.
b. Model what you'd like the students to do by placing a can on the ground and situating yourself 4 feet away from the can (using the yardstick to measure). Try to toss the pom-pom into the can, asking students to keep track of the number of attempts.
    i. Remind students that they should only use underhand tosses when throwing the pom-pom.
    ii. You may wish to change how far the thrower is from the can, depending on the size of your classroom.
c. Ask students to brainstorm a statistical question that is related to the Pom-Pom Toss Activity. Allow groups to share their ideas in case there are differing statistical questions. The goal, however, is to arrive at and agree upon a statistical question similar to: "How many attempts does it take students in this class to get the pom-pom into the can?"
d. Tell students they will be working in their groups to collect data to answer this question. Each student will try to toss the pom-pom into the can, while the partners count their attempts. Stress that all students should stand the agreed-upon distance away from the can so that the data they collect is valid.
e. As a class, create a dot plot that models the data from the Pom-Pom Activity using a Dot Plot Tool, like this one from [Geogebra](#). Each student can enter his or her own data value into the spreadsheet on the applet as they finish the activity, or you may decide to ask students to verbally share their results and enter all data values yourself.
    i. The tool displays mean, median, range, and mean absolute deviation on the dot plot. You may choose to address these terms now or explain that you will learn about calculating these measures later in the unit.
f. Using the class-created dot plot, ask students to work with their group members to describe the shape, center, and spread of the data set. These will vary depending on the data set.

Scaffold Opportunity - Suggest possible sentence frames to assist students.
    i. Shape - "Because most of the data is on the ___ side of the graph, the

tail is on the ___ side, so the data is skewed ____."
   ii.   Center - "The data clusters around ____, so this is the center."
   iii.   Spread - "Most of the data clusters between ___ and ___." "There is a gap in the data from ___ to ___, which means __ and ___ are outliers."

g. Facilitate a class discussion, allowing groups to share their descriptions of shape, center, and spread. You may wish to have students come up to the board to use the projected dot plot when explaining how they determined these descriptors. Several ideas for discussion and extension questions are:
   i.   Why is a dot plot useful in determining shape, center, and spread of a data set?
   ii.   If we added additional values of _____, how would this affect the shape of the data? (Provides the opportunity to discuss all three shape options)
   iii.   Are you familiar with any other terms that represent the center of a data set?
   iv.   How large must a gap be to decide if a data value is an outlier?
   v.   How does the data help you answer: "How many attempts do you think it would take most people, on average, to get the pom-pom into the can?"

Conclusion (5 minutes)

Exit Ticket - Use questions #1-4 of [Illustrative Mathematics' Describing Distributions task](#) to assess students' understanding of shape, center, and spread of data represented in a dot plot.

Possible Enrichment Activity - Assign #5-8 of [Illustrative Mathematics' Describing Distributions task](#), asking students to apply their knowledge of shape, center, and spread to data displayed in a histogram instead.

Material to Follow Lesson
Objectives:
   ● Summarize numerical data sets using a five-number summary to describe center and variability (6.SP.B.5.C)
   ● Display numerical data in a box plot on a number line (6.SP.B.4)

Key Ideas:
   ● The median is a measure of center of a data set. If there are an odd number of data values, the median is the middle number of a data set organized in numerical order. If there is an even number of data values, the median is the average of the two data values in the middle.
   ● A box plot displays a data set using the five-number summary, which includes the minimum, first quartile, median, third quartile, and maximum values. The first quartile

is the middle number between the minimum and the median (sometimes referred to as the median as the lower half of the data set). The third quartile is the middle number between the median and the maximum (sometimes referred to as the median of the upper half of the data set.

- Unlike dot plots, box plots are a good representation for a large set of data. It is not as easy, however, to determine the shape or skew of the data set. Also, the number of data values cannot be determined from a box plot alone.

Vocabulary: box plot, five-number summary, median, minimum, maximum, first quartile, third quartile, range, interquartile range