Investigation 3.2
Who Has the Longest First Name?

Overview

This investigation is based on one found in the Appendix for Level A in *Guidelines for Assessment and Instruction in Statistics Education (GAISE): A Pre-K–12 Curriculum Framework*. During the first week of school, a third-grade teacher is trying to help her students learn one another's names by playing various games. During one of the games, a student named MacKenzie noticed that she and her classmate Zacharius each have nine letters in their names. MacKenzie conjectured that their names were longer than everyone else's names, which gave the teacher an opportunity to introduce a statistics lesson.

In this investigation, students analyze the length (number of letters) of their first names. The data will be organized and displayed in **dotplots** to develop the **median** as a measure of center and the **range** as a measure of variability of first name lengths.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level A activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Collect data and organize the results in a dotplot

- Find measures of center (median and mode) for the data

- Consider what measures of center are appropriate for categorical versus quantitative data (addressed in the extensions)

- Find a measure of spread (range) for the data

Common Core State Standards
for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

Common Core State Standards
Grade Level Content

6.SP.1 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.2 Understand that a set of data collected to answer a statistical question has a distribution that can be described by its center, spread, and overall shape.

6.SP.3 Recognize that a measure of center for a numerical data set summarizes all its values with a single number, while a measure of variation describes how its values vary with a single number.

6.SP.4 Display numerical data in plots on a number line, including dotplots, histograms, and box plots.

NCTM Principles and Standards
for School Mathematics

Data Analysis and Probability

Grades 3-5 All students should design investigations to address a question and consider how data collection methods affect the nature of the data set; represent data using tables and graphs such as line plots, bar graphs, and line graphs; use measure of center, focusing on the median, and understand what each does and does not indicate about the data set.

Materials

- Sticky notes

- Masking tape

Estimated Time

Two days

Instructional Plan

✏️ Formulate a Statistical Question

1. Have your students discuss what they would like to know about their first names. Make a list of their responses. The following are some possible responses:

    Do any of us have the same name?

    What is our most common first name?

    Who has the longest name?

    Who has the shortest name?

    What is the most common first letter?

    What is the most common length for our first names?

2. This investigation focuses on the statistical question, "How do the lengths of first names vary in our class?"

✏️ Collect Appropriate Data

1. Hand out a sticky note to each student. Have your students write their first name on a sticky note, as well as the number of letters in their first name. For example:

2. Have your students place their sticky notes on the board in no particular order. They are placed randomly so that the class will be able to observe

how a dotplot can be used to organize the notes. Figure 3.2.1 is an example of the data for one class.

Figure 3.2.1 Example of sticky notes with names and lengths

✏ **Analyze the Data**

Have your students discuss how they may be able to organize all these sticky notes. They may suggest grouping the notes by the number of letters, as shown in Figure 3.2.2.

Figure 3.2.2 Example of sticky notes organized by number of letters

Reorganize the sticky notes into a preliminary dotplot. Note that, as much as possible, there should be no gaps or overlaps between each note in a column. See Figure 3.2.3.
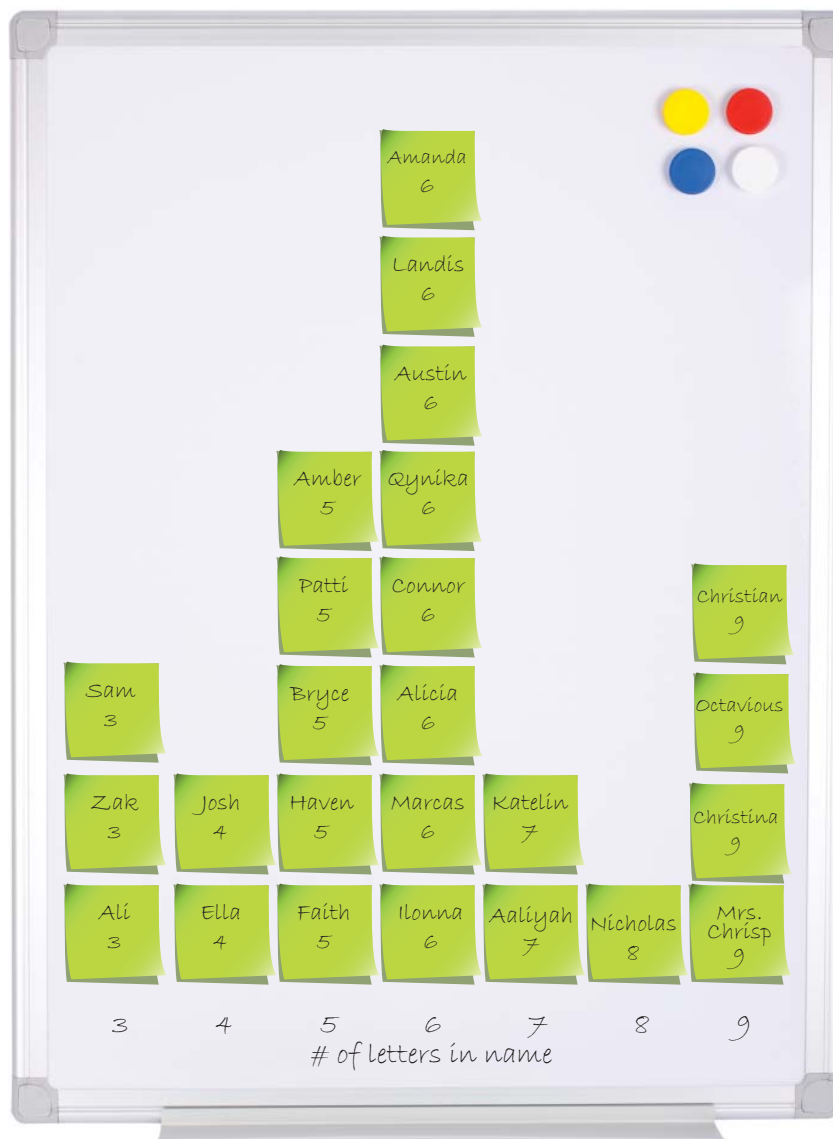

Figure 3.2.3 Dotplot of length of first names

Ask the following questions:

    a.   What is the shortest name? How many letters does it have?

    b.   What is the longest name? How many letters does it have?

c. Is it possible for someone to have a name shorter than our shortest name?

d. Is it possible for someone to have a name longer than our longest name?

e. What is the range of the data? The range of the data is the largest number minus the smallest number. (9 – 3 = 6 for this example)

f. What is the most common number of letters in the first names for our class? This value is called the mode. (6 for this example)

To introduce the concept of the median, students need to be lined up side by side according to the number of letters in their name. Note that it might be helpful to have your students form a human dotplot first. That will get all the students together who have the same number of letters.

a. Put a piece of masking tape on the floor labeled with numbers from the smallest number of letters to the largest number of letters in their names.
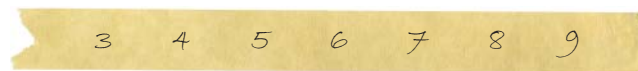

Figure 3.2.4 Labeled masking tape

b. Have your students line up corresponding to the length of their name (e.g., Josh and Ella would both line up above the 4) in either order. Once they have created a human dotplot, have them form a single-file line side-by-side, keeping them in order based on the length of their name.

c. Ask your students who they think is exactly in the middle of this display.

d. Have one student each at either end of the line sit down until one student is left standing in the middle. If there is an even number of students in the class, you may want to include yourself in the display to ensure students will be finding the median of an odd number of data points. For example, have the first student above 3 (Zak) and the last student above 9 (Octavious) sit down at the same time. Have the second student above 3 (Sam) and the second-to-last student above 9 (Christian) sit down at the same time. Continue this process until you are left with one person standing. The number of letters in that person's name is the median of the data set. Note that if the number of students is even, there will be two students left standing. The median is the value half-way between their values.

e. Ask your students how they could use the dotplot on the board with the sticky notes to model what they did in the human dotplot in

order to find the median length of their names. Have your students make suggestions. They should realize that they could either remove sticky notes one at a time from both ends until one was left or put an X through the notes until they get to the middle one, in the case of an odd number of sticky notes.

f. A new student arrives after the above has been completed. Her name is Seraphinia. How would adding her to the data set affect the center measures of mode and median and the spread measure of range? Note that Seraphinia has 10 letters, so the mode remains at 6. Also note that, with Seraphinia, there is now an even number of names. An even number of data points creates two "middles." The median is taken to be the value half-way between the two middles. In this example, both middle values are 6, so the median remains at 6. The range is now $10 – 3 = 7$, an increase of one letter.

✏ Interpret the Results in the Context of the Original Question

1. Have your students recall the original question: How do the lengths of first names vary in our class? Ask them to write a report that answers the question, along with providing a justification of it using their analysis.

2. Ask your students the following questions:

    a. Based on our data, what is the most common or typical length of our first names?

    b. Do you think this would be the same in the classroom across the hall? Why or why not?

    c. Do you think this would be the same in a middle- or high-school classroom? Why or why not?

    d. Do you think this would be the same in Mexico? In China?

Example of 'Interpret the Results' 💿

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

On the first day of school, our teacher had us play games to learn each other's names. After that, she was showing us some statistics by having us study the lengths of our first names. The question was, "How do the lengths of first names vary in our class?"

We used sticky notes with our names and the number of letters in our names written on them. After putting all of the sticky notes on the board all messed up, we organized them into a dotplot with the number of letters on the horizontal axis. But, we didn't do the graph the right way the first time, because we didn't keep the columns nice and straight and in line with the other columns. We had to remember to keep the rows in line, also. When we corrected that, we saw that there were more of us whose first names had six letters than any other number. It was the highest in the dotplot. That's called the mode number of letters.

We also calculated the middle number of letters by lining up from fewest number of letters to most and then having low and high sit down until we got to one person left. That number is called the median. It's the middle number of letters, 6 (Alicia), with 12 of us below Alicia and 12 of us above Alicia.

The day after we did that analysis, we got a new student in class, Seraphinia. When we added her, she had the longest name. The range of letters was 9 − 3 = 6 before Seraphinia, but 10 − 3 = 7 letters with her. The mode stayed at 6 because it was still the highest. To find the median, we sat down like before, but now there were two middles, Alicia and Connor. They both have 6 letters in their names, so the median is still 6.

We want to continue doing this study by looking at names from different countries to see if their number of letters differs from ours. We think that maybe Chinese names are shorter.
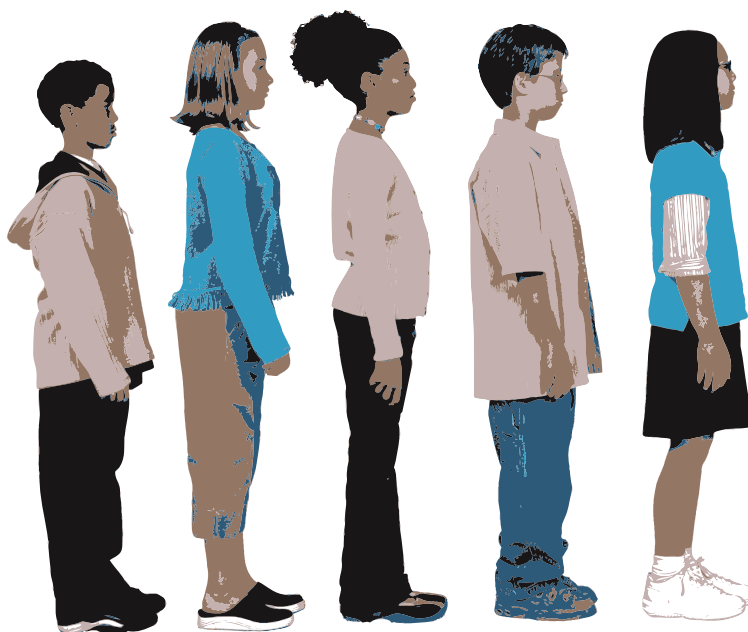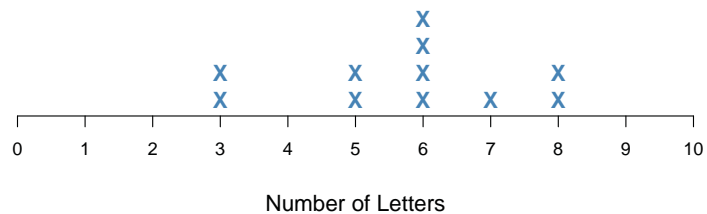
Assessment with Answers 💿

With the help of his family and friends, Jose collected data regarding the lengths of first names of his family and friends. Table 3.2.1 shows the data Jose collected.

Table 3.2.1 Length of First Name

	Family and Friends First Names	Number of Letters in First Name
1	Hector	6
2	Amada	5
3	Che	3
4	Ricardo	6
5	Camila	6
6	Roberto	7
7	Carlos	6
8	Raymundo	8
9	Gabriela	8
10	Diego	5
11	Tia	3

1. Make a dotplot of the length of the first names of Jose's family and friends.

Length of First Names

```
                            X
                            X
                  X     X   X         X
                  X     X   X   X     X
    |---|---|---|---|---|---|---|---|---|---|---|
    0   1   2   3   4   5   6   7   8   9   10
```

Number of Letters

2. Find the value of each of the following:

Maximum value: 8

Minimum value: 3

Mode: 6

Median: 6

Range: 5

3. Write a summary of what you observed about the length of the first names of Jose's family and friends. Your summary should include reference to the dotplot and the measures of center and spread that you found. The dotplot shows that all of the name lengths are from 3–8 letters long. The most common length was 6 letters (four people had this length) and the median length was also 6 letters, meaning there were five names above the median (6, 6, 7, 8, 8) and five below (3, 3, 5, 5, 6).

Extensions

1. Have your students do an analysis of the length of their last names.

2. Your students may have had the experience of having to write their full name (first, last, and space between) on a form that has a set of boxes. Have your students combine the length of their first and last names, including the space between their names, and investigate how many boxes the form should have so most of the students in class can fill in their entire name.

   Note: Discuss with your students what they feel "most of the students" means in this context. Some students will think just over half, while others will want to say all except for one or two students. Some may even interpret it as "most often" and want to use the mode. Allow students to answer the question based on their definition of "most."

3. Students may be tempted to find the median of a categorical set of data. The following exercise will demonstrate why the median is a measure of center for numerical data only.

   a. Consider the statistical question, "What is your favorite type of pet?" Put a piece of masking tape on the floor labeled as follows:

   Cat   Dog   Bird   Fish   Other

   Figure 3.2.5 Favorite pet

   Have your students line up according to their favorite type of pet, thus forming a human bar graph. They have to choose only one pet category.

b. Ask your students which type of pet is the most popular? Who thinks their favorite type of pet is in the middle?

    Note: The first question is asking for the mode, while the second question is designed to begin a discussion about why categorical data do not have a median. The following parts, c through e, should help with the discussion.

c. Recalling the process of creating a human dotplot to determine the median length of first names, have your students at either end sit down until one is left standing in the middle (two if the number of students is even). Ask your students what the median is. Students might respond with the name of the student or the type of pet.

d. Change the order in which the pets are listed on the floor and have your students make a new bar graph.

Figure 3.2.6 Favorite pet

e. Have your students find the middle again by sitting down starting at each end. Ask what the median or middle is. It is likely that this answer will be different from the first bar graph. Discuss with your students that changing the order of the categories has changed the median response, but the data are exactly the same in both cases. Have them do the process a third time. Yet another median appears for the same data. Discuss with your students that a median cannot have different values for the same data set. So, finding a median for categorical data does not make sense. To find a median, data must be able to be ordered from smallest to largest. That can be done for numerical data, but it cannot be done for categorical data.

4. Ask the class to work together to determine if the following data sets can have a median: number of pets, eye color, number of siblings, ways to get to school. Then, have your students work in pairs to create an example of a data set that could have a median and one that could not have a median. They should write a sentence or two with their examples, explaining why their data set can or cannot have a median.

References

Bereska, C., L. C. Bolster, C. A. Bolster, and R. Scheaffer. 1998. *Exploring statistics in the elementary grades: Book one, grades k–6*. White Plains, NY: Dale Seymour.

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics. www.corestandards.org.