The OLDEST PERSON You've Known

A Super Bowl commercial became the impetus for engaging students in a meaningful data collection project.

Sarah B. Bush, Karen S. Karp, Judy Albanese, and Fred Dillon

What better way to interest students in mathematics than using a Super Bowl® commercial? A Prudential® insurance commercial aired during the Super Bowl in 2013 was the impetus for our lesson (see it here on YouTube™: http://www.youtube.com/watch?v=IsNiKGMSHUQ). In the commercial, four hundred people were polled on "How Old Is the Oldest Person You've Known?" and each was given a sticker to place on a larger-than-life dot plot marking the age of the oldest person they knew. We re-created this activity for sixth-grade and seventh-grade students to engage them in collecting meaningful data, creating organized data displays, and analyzing and interpreting the data to draw generalizations. We also posed a series of questions about measures of central tendency, the shape of data, the interquartile range, absolute mean deviation, representativeness, predictions, and more. Our activity was tested in the sixth-grade and seventh-grade classes taught by one of the authors.

LESSON DESCRIPTION

Overview

Our activity incorporates a model adapted from the GAISE report (Franklin et al. 2005), which recommends that students engage in statistical problem-solving tasks in which they—

FUNSTOCK/THINKSTOCK

1. formulate questions;
2. collect data;
3. analyze data; and
4. interpret results (p. 11).

We emphasize the latter three because the Prudential commercial provided the initial question, which launched the activity: How old is the oldest person you've known?

Collecting the Data

We wanted students to work with one data set consisting of subsets of data, so that we could ask students questions that required them to make comparative inferences among three sample groups. Therefore, students needed to collect three pieces of data. First, they needed the age and name of the oldest person they knew. When students asked their parent or guardian and then their grandparent or other older relative the same question, three data points resulted. Our only stipulation was that the "oldest person you've known" must be someone who was still living and who was known personally. Students or family members were not allowed to choose celebrities or historical figures. Students used a data collection sheet (see **fig. 1**) and had five days to submit their three data points. Then each student made a prediction of the mean for the oldest person from each group (see **fig. 2**).

We anticipated that the student data set would have a spread and measure of center that was smaller than the other two sets of data (parent and grandparent). We also predicted that the grandparent data would have a small spread and a much higher measure of center than either of the other data sets.

Creating a Dot Plot

After students finalized predictions about the average ages, they created a dot plot similar to the one that appeared on the television commercial. Students wrote the ages from their own, their parents', and their grandparents' oldest person choices

Fig. 1 Students kept track of information in a data collection sheet.

How Old Is the Oldest Person You've Known? Data Collection Sheet

	Age	Name	How Do You Know This Person?
Student			
Parent/guardian			
Someone else two generations older (grandparent, neighbor, etc.)			

Fig. 2 Students predicted the average of three data sets.

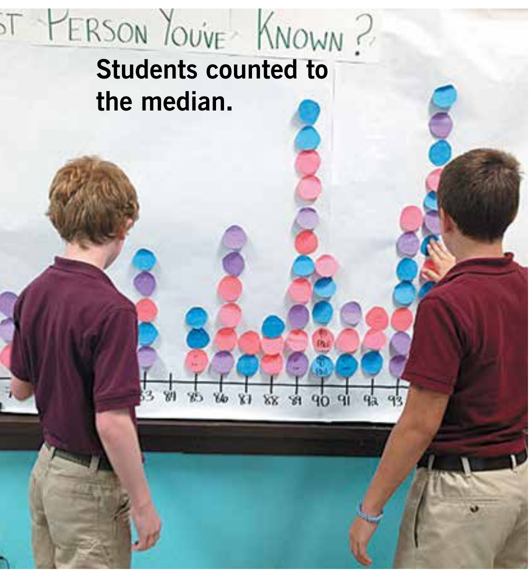
How old do you think the *average* age will be for the three groups? Write your predictions here:

Student average: ______________

Parent/guardian average: __________

Grandparent/other older relative average: __________

on purple, pink, and blue sticky notes, respectively, and placed them on the wall-size graph. Because the graph was so large, it was prepared beforehand with ages 54–106 years along the x-axis and with frequencies on the y-axis. We felt comfortable preparing the graph ahead of time because students had already shown proficiency in creating graphs with appropriately labeled axes, scales, and titles. We carefully spaced the intervals far enough apart so that sticky notes would not overlap.



Students counted to the median.

SARAH B. BUSH

Analyzing and Interpreting the Data

The next day, students watched the commercial. We explained that the commercial was the catalyst for this activity and although the subject matter involved saving for retirement, our focus was on using the dot plot with their data to analyze and interpret the results of their data. We hoped that students would generate links to real-world situations that were similar to the retirement scenario.

Students were eager to examine their data and discover if their predictions (see **fig. 2**) were accurate. They used an activity sheet (see **fig. 3**) containing questions that centered on grade-level statistics concepts and worked in mixed-ability groups of three or four.

Question 1, complete the table, asked students to calculate measures of central tendency for each of the three groups and for the combined data set. Some groups were assigned the mean and range, and other groups, the median. Students walked up to the dot plot to observe the data and calculate these measures. We noticed that some students were initially unsure about counting the same age more than once when there was more than one data point for the same age. Student groups then shared their descriptive statistics with the class. After every student had completed the table, we were ready for data analysis.

Using responses from question 1 and their prior knowledge, students worked on questions 2–10. Students also strategically selected tools and resources to use, such as calculators, the dot plot, the vocabulary word wall, and the teachers.

Immediately, students noticed that the range of ages for the parent group was narrower than for the other two groups (as prompted in question 2). One student wrote, "The parents know more people that are in the same age group while everyone else is spread apart, that says they [students and grandparents] know a wider variety of peoples' ages." Another student wrote, "The ranges are differing from grandparents at 40, parents at 21, and students at 43. The parents have more clustered data, where the students are spread out." Some students used terms such as "dense" and "scattered" to describe the data. Hearing these student answers gave us an opportunity to review mathematical vocabulary of deviation including "spread" and "range." We required that students use precise terminology in their explanations with their groups. It was surprising that the range was so large for the grandparent group. This situation provoked a rich discussion, particularly about one value being lower than the other blue

sticky notes. We discussed the concept of an outlier and how it affects the mean and the range.

In examining responses for question 3, students determined that our dot plot looked like a combination of "skewed left" or "mound shaped." We heard a variety of thoughtful conversations in groups, such as this one:

Student 1: I think it is skewed left because there are more [points] to the right.
Student 2: I think mound shaped because the end [student points to the data on the right of the graph] is making it a mound.
Student 3: I think it is skewed right or mound.
Student 1: Why do you think it is skewed right?
Teacher: Another word we use when describing mound-shaped graphs is *symmetry*. Does that change your opinion about the shape of this graph? Why?

To answer this question, the students discussed how the language of "skewed left" and "skewed right" could be confusing because it is the opposite of the direction that one may expect. Although student groups were not satisfied describing the data using just one shape, they realized that this graph did not meet the requirement of line symmetry for a mound shape. Therefore, they decided that the best descriptor was "skewed left." These conversations engaged students in one of the Common Core's Standards for Mathematical Practice (SMP) as they explained their thinking and justified their reasoning to others (SMP 3).

For question 4, student groups used the dot plot to help them find the interquartile range for the combined data set. They first counted in from the ends to find the median of all data points. Next, they identified the upper and lower quartiles by finding the median of the lower half of data (from the least value to the median) and the median of the upper half of data (from the median to the greatest value) using this counting method. Students explained that the difference of the first and third quartile represented the interquartile range.

The responses for questions 7 and 8 showed not only that students understood the algorithm for calculating the mean absolute deviation (MAD) but also that they understood the concept of the MAD.

Question 7 asked students to find the MAD (average distance of the data points from the mean) for each group. Previously, students had discussed the definition of mean absolute deviation and had practiced finding the MAD

Fig. 3 Students recorded their data analysis on an activity sheet.

How Old Is the Oldest Person You've Known?

1. Complete the table.

Sample	Mean	Median	Range
Students			
Parents/guardian			
Grandparents/other older relative			
Combined			

2. Are there differences in the ranges of the three groups? If so, describe the differences.
3. Examine the dot plot for the entire data set. Describe the shape of the data using the words *uniform, mound-shaped, skewed left,* or *skewed right.*
4. What is the interquartile range for the entire data set?
5. Are there any values for the entire data set that can be considered outliers? If so, what are they? How did you determine whether they were outliers or not?
6. Which measure of central tendency describes the data most accurately? Why?
7. Find the mean absolute deviation for each group. What information do you learn by comparing the three statistics?
8. Do you think the results of your data collection are representative of students at other schools? Why or why not? Would it be representative of students in other countries? Why or why not?
9. What overlap do you notice on the dot plot among the three colors (groups)? What does this overlap represent?
10. What was the difference between your predicted average (before plotting the data) and the actual average for each group? Are you surprised? Why or why not?

Fig. 4 These thoughtful student responses were in answer to question 10, "What was the difference between your predicted average (before plotting the data) and the actual average for each group? Are you surprised? Why or why not?"

I thought that the parents would know
people who were older than what the
parents actually knew. I thought the
average for the parents would be around
ages 95, but the mean for the parents
was 89. My predictions for the students
It didn't surprise me. → was pretty similar but it was only
3 numbers too high compared to the mean.

(a)

The average I predicted for students was very
close. My prediction for parents was over, and I
was surprised, because I would think parents
would know older people than students and
have a higher average. My prediction for
grandparents was lower, because the oldest
person my grandma knew was younger
than the person I knew.

(b)

for small data sets. Students were initially challenged by this larger data set (31 data points for each group; 93 combined). Students knew they must start by recalling the mean for that data set (referring back to question 1).

Next, students decided to work in groups of three at the dot plot, with one student calling out a data point, another student quickly computing the difference between each individual data point and mean, and a third student recording that difference to keep a running log. Students knew to take the absolute value of the differences (or subtract the smaller number from the larger number) because the distance could not be negative. After students recorded all the differences between the data points and the mean, they went back to their desks and found the mean of those differences, thus calculating the MAD for their data sat. Another option to confirm this value was to ask students to make a list of data-point values below and above the mean, then sum each set to see if they were equal. Because the means above and below summed to zero, we used the absolute value. When asked what the MAD told them about their data, one student said, "It's the average distance each one is away from the mean." Students realized that MAD represented the average number of years that the reported age for each data point was from the mean. After deep pondering, another student explained that he just realized the MAD was probably lowest for the parent group because its narrower range might be a predictor of how large or how small the MAD would be. Students were not only calculating the MAD but also made conjectures about what could cause the MAD to be a smaller or a larger number, such as occurred in its connection to the range of the data set or how an outlier can affect the MAD.

We wrote question 8 to build cultural connections and encourage students to think about the generalizability of their data. Students considered many factors when thinking about their responses and asked questions about the ages of students at other schools, whether they were students in our city or elsewhere, and whether they should think about other countries. We received the following responses to question 8:

> For other schools yes because if they do the same grade they are most likely the same age as us. For other countries no because their life expectancy might not be as long.

> In other schools kids would be our age so yes. But in other countries no because the death rate could be different.

We felt that this question required students to consider the problem's context, thus aligning well with SMP 4: Model with mathematics. As students considered the context and the idea of asking the same question of students in other countries, they explained that life expectancies might be different internationally. Additionally, students discussed the idea that a sample at an individual school was relatively small and not random and that this could cause the means to vary.

The final question asked students to think about their original predictions of the mean for student, parent, and grandparent groups and compare that prediction with the actual results. Students provided thoughtful responses, such as those in **figure 4**. Most

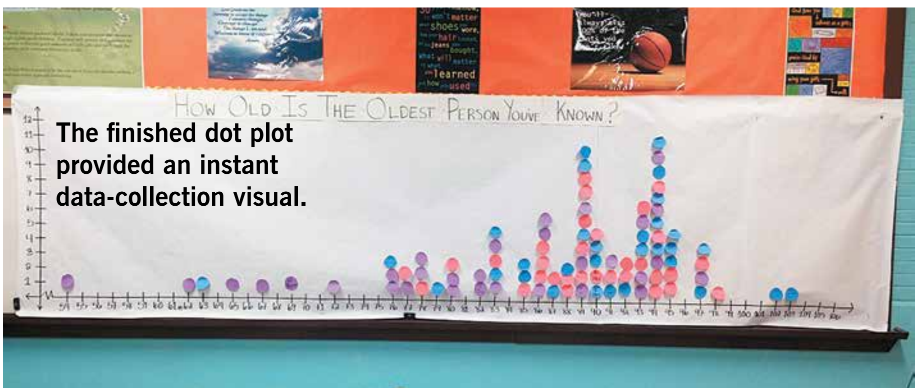
Fig. 5 Students held a rope to show the box plots for student (**a**), parent (**b**), and grandparent (**c**) data.

(a)

(b)

(c)

SARAH B. BUSH



The finished dot plot provided an instant data-collection visual.

SARAH B. BUSH

students were close on at least one of their predictions. Some students predicted the average for the grandparent group would be much higher than it actually was, because they assumed that their grandparents were likely to know the oldest people. This was in line with our original thoughts, so we were surprised about the results for the three groups, as well.

AN EXTENSION: EXPLORING BOX PLOTS

The final extension task asked students to create three box plots representing the student, parent, and grandparent data. Instead of making the plots with paper and pencil or technology, students stood in for the data (see **fig. 5**).

We taped a number line to the floor that used a limited domain from 50 to 105 years to align with our data, using 5-year intervals as the scale, and asked students to look at their original data collection sheets (see **fig. 1**) to recall their data points. First, students lined up silently on the number line according to the data point of the oldest person they knew (student data set). Students used gestures to communicate with one another as they identified others' numbers and figured out their position on the number line. Students who had the same data point were confused about where to stand. They acknowledged that for their specific data point to be accurately represented on the number line, students with the same values should line up behind, rather than beside, one another. We have also seen this activity completed in which students stood beside one another, rather than behind, but we respected the students' reasoning.

We asked students how they could find the median, and they said they needed to count toward the middle. We gave small flags to the two individuals who represented the smallest and greatest data points, and students called "pass" as they systematically moved

Fig. 6 Question 3 of the box plot activity sheet was the impetus for students' responses in (a) and (b).

Box Plot Questions

1. Examine the length of the whiskers of the three data sets. What does each of their lengths tell you about the data?

2. Which box is the shortest? What does that tell you about the data points inside the box?

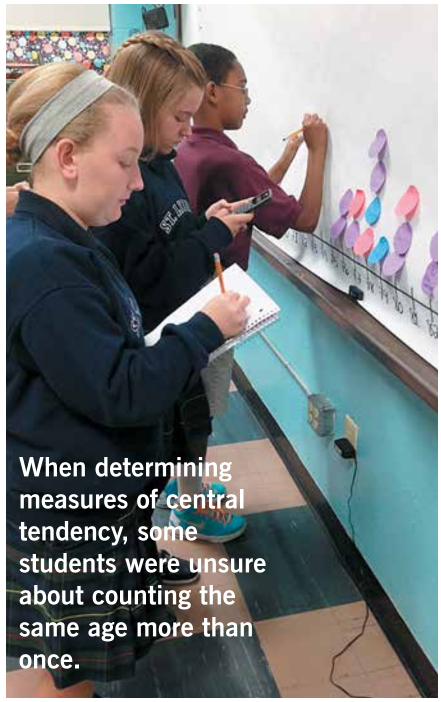
3. What is the most surprising thing you notice when you compare the three box plots? What does that tell you about the data?

we are suprised that the parents were more clustered than the grandparents. I think this is suprising because since they are old, they would have people or friends that are older than them just like we have friends that are older than us. maybe, most of their friends have passed away.

(a)

What suprises me is that the grandparents box is even more spreadout than the parents because it seems like the grand-parents would know people at the same age group and people who are older than who the parents and students know.

(b)



When determining measures of central tendency, some students were unsure about counting the same age more than once.

SARAH B. BUSH

them toward the middle until the flags met at the median. When there were two middle numbers at the completion of the passing, we asked students what to do. After discussion, they suggested averaging those two numbers.

Next, we asked students to follow a similar flag-passing process to determine the upper quartile and lower quartile. This time, the flags moved from the middle back to the ends, with students on each end passing additional flags inward toward the middle. The two quartiles were determined when each set of flags met. Students noticed that the numbers from the lower quartile to the upper quartile were the same as the "box" part of a box plot. Students were then given a rope. The students at the 1st and 3rd quartile changes (and everyone in between) were asked to lift the rope. The remaining students held the rope at their waist, thus representing the "whiskers" of the plot. The biggest challenge that students had with making the box was when there were two middle numbers at the upper or lower quartile, causing both students to want to raise the rope. However, the actual upper and lower quartile fell between those two students, which confused some students. After the human graph was complete for the student data, we took a photograph (see **figs. 5a–c**). We repeated this process for the parent and grandparent data.

We displayed all three photographs on the overhead projector. Students were given an activity sheet, which asked them to make comparative inferences about the three different groups, using the three box plots displayed (see **fig. 6**).

For the first question, most students agreed that the data from the student and grandparent box plots were more spread out, with some students attributing this to their larger ranges (from their previous work). Other students mentioned that outliers caused the length of the whiskers. On question 2, students explained that the parent data made the shortest box. Student responses were similar to this student comment, "The parents' box is the smallest, which tells us that the range is small, the data is clustered, and that the ages of the people that the parents

know is similar." When asked what surprised them, students had a variety of interesting responses (see **fig. 6**). Most students originally thought that the grand-parent box plot would have been more clustered (and include older ages) because they should have friends their age and older. However, some students noted that perhaps some of their grandparents' friends may have passed away, thus meaning that they might have known fewer older people. Students also thought that the parent data set would have been more spread out and found it strange that it was so tightly clustered. However, they were unable to infer possible causes.

STUDENTS: WE "WERE THE DATA"

This activity provided an engaging and relevant context that allowed students to use their knowledge of statistics to collect, display, analyze, and interpret data in meaningful ways. Students who took part in this activity were excited to "be the data," survey two different generations about the oldest person they knew, and compare the three sets of data by answering a variety of questions.

This lesson provided a relevant avenue to involve sixth-grade and seventh-grade students in multiple CCSSM content standards. What was most important was this activity's student-centered focus because students were doing the work of thinking, modeling, justifying, and reasoning and were authentically engaged in several mathematical practices. Specifically, students modeled with mathematics as they worked within this real-life context and were personally connected to the data and represented the data in multiple ways.

We placed a strong focus on asking students to explain their reasoning; to justify their answers, considering the data and the context of the problem; and to discuss their thoughts in their groups, thus helping them construct arguments and critique the reasoning of others. Finally, we continually reinforced the idea of being precise in all of their spoken, written, and calculated work, which emphasized the importance of giving attention to precision in all mathematical endeavors. Students found ways to connect to their families and the real world as they gathered information for an engaging learning experience.

CCSSM Practices in Action

SMP 3: Construct viable arguments and critique the reasoning of others.
SMP 4: Model with mathematics.
SMP 6: Attend to precision.

Statistics and Probability, both sixth-grade and seventh-grade domains: 6.SP.1, 6.SP.2, 6.SP.3, 6.SP.4, and 6.SP.5; 7.SP.3 and 7.SP.4

REFERENCES

Common Core State Standards Initiative (CCSSI). 2010. Common Core State Standards for Mathematics. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. http://www.corestandards.org/wp-content/uploads/Math_Standards.pdf

Franklin, Christine, Gary Kader, Denise Mewborn, Jerry Moreno, Roxy Peck, Mike Perry, and Richard Scheaffer. 2005. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. Alexandria, VA: American Statistical Association.

YouTube. 2013. Prudential—Stickers. http://www.youtube.com/watch?v=IsNiKGMSHUQ

Any thoughts on this article? Send an email to ***mtms@nctm.org.***—*Ed*.




Sarah B. Bush, sbush@bellarmine.edu, is an assistant professor of mathematics education at Bellarmine University in Louisville, Kentucky. She is a former middle-grades mathematics teacher who is interested in interdisciplinary and relevant and engaging math activities. **Karen S. Karp**, karen@louisville.edu, is a professor of mathematics education at the University of Louisville in Kentucky. She is a former member of the NCTM Board of Directors and a former president of the Association of Mathematics Teacher Educators (AMTE). She continues to work in classrooms to support teachers of students with disabilities in their mathematics instruction. **Judy Albanese**, jalbanese@stleonard louisville.org, is a middle-grades mathematics teacher at St. Leonard School in Louisville, Kentucky. She seeks to develop her students' conceptual understanding of mathematics by implementing instruction and activities that are engaging and relevant to her students. **Fred Dillon**, dillon314@sbcglobal.net, is a mathematics teacher from Strongsville, Ohio, and a former member of the NCTM Board of Directors and *MTMS* Editorial Panel. He is interested in helping teachers use engaging tasks and student involvement in their classrooms.