Lesson 18: Using data to solve problems

Goals

- Recognise that different graphical displays offer different insights into a distribution. Choose an appropriate graphical display to represent a data set, and justify the choice (orally and in writing).
- Recognise that different measures of centre and variability offer different insights into a data set. Choose an appropriate measure of centre and variability to describe a data set, and justify the choice (orally and in writing).

Learning Targets

- I can decide whether mean and range or median and IQR would be more appropriate for describing the centre and spread of a data set.
- I can draw an appropriate graphical representation for a set of data.
- I can explain what the mean and range or the median and IQR tell us in the context of a situation and use them to answer questions.

Lesson Narrative

This lesson is a good opportunity for students to use the information they have learned in the unit and apply it to different situations, but may be shortened to fit time constraints.

In this lesson, students compare the centre and spread of different distributions. They determine what these different measures (mean and range or median and IQR) represent in context. They select an appropriate representation for the distribution based on the structure of the data, an appropriate set of measures of centre and spread, and interpret their meaning in the context.

For students who are curious why we are asking them to calculate measures of centre and variation by hand when computers would be more efficient and accurate, tell them that understanding the meaning of the values and knowing what questions to ask are skills computers have not yet mastered. By practising with calculations on small data sets, students are becoming familiar with these measures as well as questioning skills so they can correctly interpret results from computers in the future. If students do not raise the question themselves, this point may be left until the topics are revisited in later grades.

Addressing

- Understand that a set of data collected to answer a statistical question has a distribution which can be described by its centre, spread, and overall shape.
- Summarise and describe distributions.
- Giving quantitative measures of centre (median and/or mean) and variability (interquartile range and/or range), as well as describing any overall pattern and any

striking deviations from the overall pattern with reference to the context in which the data were gathered.

- Relating the choice of measures of centre and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

- Group Presentations
- Collect and Display
- Discussion Supports
- Notice and Wonder
- Think Pair Share

Required Materials

Straightedges
A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Tools for creating a visual display
Any way for students to create work that can be easily displayed to the class. Examples: chart paper and markers, whiteboard space and markers, shared online drawing tool, access to a document camera.

Required Preparation

Provide access to straightedges for students to use when drawing box plots. Preview the background information about the yellow perch fish for the main activity. Prepare tools for creating a visual display, one set for every 3–4 students.

Student Learning Goals

Let's compare data sets using visual displays.

18.1 Wild Bears

Warm Up: 5 minutes

This warm-up allows students to review two important ideas of this unit: interpreting data in a box plot and writing statistical questions based on a data set. Students write statistical questions based on given box plots, then swap questions to answer questions written by another student.

Instructional Routines

- Think Pair Share

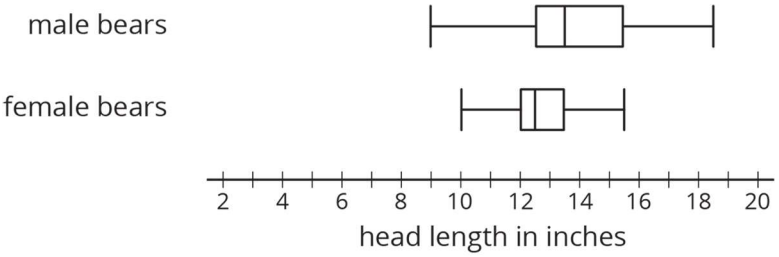
Launch

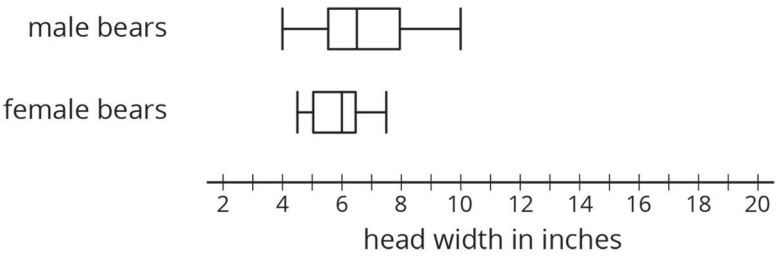
Arrange students in groups of 2. Tell students that, for the first question, one partner should write two questions about the head lengths and the other partner should write two questions about the head widths. For the second question, they should exchange and review each other's questions. If their partner's question does not seem to be a statistical question, suggest a revision so that it becomes a statistical question, and then answer the question. Remind students to consider units of measurement.

Give students 2 minutes of quiet work time for the first question and 2 minutes for collaboration afterwards.

Student Task Statement

In one study on wild bears, researchers measured the head lengths and head widths, in inches, of 143 wild bears. The ages of the bears ranged from newborns (0 years) to 15 years. The box plots summarise the data from the study.





1. Write four statistical questions that could be answered using the box plots: two questions about the head length and two questions about the head width.
2. Swap questions with your partner.
    a. Decide if each question is a statistical question.
    b. Use the box plots to answer each question.

Student Response

1. Answers vary. Sample statistical questions:
    - What is a typical head length, in inches, for male bears in the data set?
    - Do female bears generally have longer heads than male bears?

- Which data set shows more variability in head widths: male bears or female bears?
- What is the widest head width, in inches, for male bears?
- How do the ranges of head widths compare? Which group—male bears or female bears—has a larger range, and by how much?

2. Answers vary. Sample responses to the questions above:
   - A typical head length for male bears is about 13.5 inches.
   - No, female bears generally have shorter heads than male bears.
   - The data for male bears show more variability in head width.
   - The widest head width for male bears is 10 inches.
   - Male bears have a larger range of head widths; it is nearly twice as large as female bears' range of head widths. The range for male bears is about 6 inches, and the range for female bears is a little over 3 inches.

Activity Synthesis

Ask several students to share their questions about the head width and head length. Record and display their responses for all to see. After each student shares, ask the class if they agree or disagree it is a statistical question. If they agree, ask how they would find the answer, or for the answer itself. If they disagree, ask how they could rewrite the question so it is a statistical question.

18.2 Maths Homework (Part 1)

Optional: 15 minutes (there is a digital version of this activity)

In this activity, students compare and contrast different measures of centre and variability for data sets that have gaps and are not symmetrical. They interpret mean, range, median, and IQR in the context of a situation. Unlike many of the data sets students have seen so far, this one shows values that could roughly divide into three parts: the days when there is little or no homework, the days when there is a moderate number of homework problems, and the days when the assignment is relatively large. Because of this distribution, finding a typical number of homework problems (or whether it would be helpful to identify a typical number) is not obvious, prompting students to interpret measures of centre and spread more carefully.

As students work and discuss, identify at least one student or group that decides that the mean and range are appropriate measures of centre and spread and can explain their reasoning, and another that decides to go with the median and IQR and could support their choice. Invite them to share during whole-class discussion.

Instructional Routines

- Notice and Wonder

Launch

Keep students in groups of 2. Give students a moment of quiet time to look at the data on homework problems and identify at least one thing they notice and one thing they wonder. Give them another brief moment to share their observation and question with their partner. Then, ask a few students to share their responses with the class.

Students are likely to notice that the data values are quite different, that there are some days with no homework and others with quite a few problems, that there is not an obvious cluster, or that the number of problems could be roughly grouped into three kinds (a little, moderate, and a lot). They are likely to wonder why the numbers are so spread out and varied.

Briefly discuss the following questions to encourage students to think about the data contextually:

- "Why might the homework assignment data show this distribution? What are some possible explanations?" (When only one problem was assigned, the problem might be particularly challenging or might require considerable work or collaboration. Another possibility: there might be an upcoming exam, so the homework load was reduced. When many problems were assigned, the problems might be quick exercises with short answers, or the assignment might be review materials for an entire chapter.)
- "How might we describe 'a typical number of homework problems' in this case?"
- "Which do you predict would be higher: the mean or the median number of problems? Why?"

Next, give students 8–10 minutes to complete the task, either independently or collaboratively. Ask students to think quietly about the last question before discussing their response with their partner.

If students are using the digital activities, they will need to enter the data points in the column "A" for the applet to "list", "sort", etc. The applet allows for students to populate their own mean, Q1 and Q3 values.

Student Task Statement

Over a two-week period, Mai recorded the number of maths homework problems she had each school day.

2 15 20 0 5 25 1 0 10 12

1. Calculate the following. Show your reasoning.
    a. The mean number of maths homework problems
    b. The range
2. Interpret the mean and range. What do they tell you about the number of homework problems Mai had over these two weeks?
3. Find or calculate the following values and show your reasoning.
    a. The median, quartiles, maximum, and minimum of Mai's data
    b. The interquartile range (IQR)
4. Which pair of measures of centre and variability—mean and range, or median and IQR—do you think summarises the distribution of Mai's maths homework assignments better? Explain your reasoning.

Student Response

1.
    a. The mean is 9 homework problems per day. $\frac{2+15+20+0+5+25+1+0+10+12}{10} = \frac{90}{10} = 9$.
    b. The range is 25 homework problems per day. $25 - 0 = 25$.
2. The mean tells us that a typical number of homework problems given in a day is 9. Since the range is 25, which is much larger than the mean, the data is widely spread out. There is a lot of variation in the number of homework problems given each day.
    a. The data listed in order is: 0, 0, 1, 2, 5, 10, 12, 15, 20, 25. This means that Q1 is 1, Q2, or the median, is 7.5 (the average of the 5th and 6th data points), and Q3 is 15. The minimum is 0, and the maximum is 25.
    b. The IQR is 14, which is the difference between Q3 and Q1.
3. Answers vary. Sample response: Both the range and IQR show the large variability in the data, and the mean and median are pretty close in value. I think the median and IQR better summarise the centre and spread of the data. There are more days where not much homework was given, and the median is lower than the mean.

Activity Synthesis

Briefly discuss students' interpretations of the measures they just calculated:

- "What do the mean of 9 and range of 25 tell us? How can we interpret them in this context?"
- "What do the median of 7.5 and IQR of 14 tell us?"

Then, select two or more previously identified students to share their responses about which measures of centre and spread are appropriate for summarising the data set. After each person shares, briefly poll the class to see if others reasoned about the measures the same way. Sum up by asking:

- "Now that you have two pairs of measures of centre and spread, how would you respond if someone asked you, 'What is a typical number of homework problems for Mai's class?' Is the question easier to answer now?"

Students should walk away with increased awareness that, in some cases, measures of centre and spread do not always paint a full picture of what the actual data set entails, and that the measures should be interpreted with care.

18.3 Maths Homework (Part 2)

Optional: 15 minutes (there is a digital version of this activity)

In the previous activity, students considered appropriate measures of centre and spread for describing distributions. Here, they show the same data set using three different kinds of graphical representations—a dot plot, a box plot, and histograms using different bin sizes—and decide which are more useful or more appropriate for communicating the distribution.

As students work and discuss, identify those who draw clear graphical displays, those who noticed that the different displays offer different insights about the data distribution, and those who advocate for using different representations to display Jada's data. Ask them to share with the class later.

Instructional Routines

- Collect and Display

Launch

Arrange students in groups of 3–4. Provide access to straightedges.

Explain to students that they will now represent Jada's homework data graphically and think about which representation(s) might appropriately communicate the distribution of her data. Give students 4–5 quiet minutes to draw a dot plot and a box plot (the first two questions), and then another 4–5 minutes to collaborate on drawing histograms with different bin sizes. Ask each student in a group to be in charge of one histogram with a particular bin size. After all representations are drawn, students should analyse them and discuss the last question in their group.

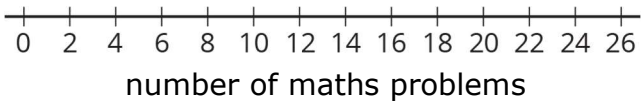
Classes using the digital version have an applet to create the statistical graphs. Data must be entered as a list, in curved brackets, separated by commas. Choices for histogram settings appear when that graph is selected.

Student Task Statement

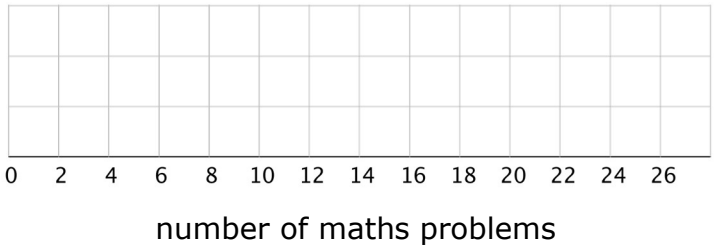
Jada wanted to know whether a dot plot, a histogram, or a box plot would best summarise the centre, variability, and other aspects of her homework data.

2 15 20 0 5 25 1 0 10 12

1. Use the axis to make a dot plot to represent the data. Mark the position of the mean, which you calculated earlier, on the dot plot using a triangle (Δ). From the triangle, draw a horizontal line segment to the left and right sides to represent the range.

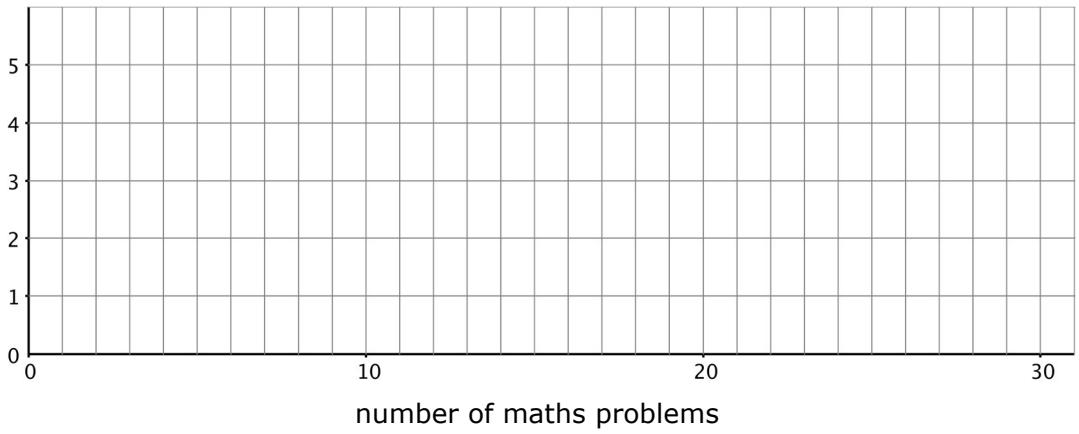


2. Draw a box plot that represents Jada's homework data.

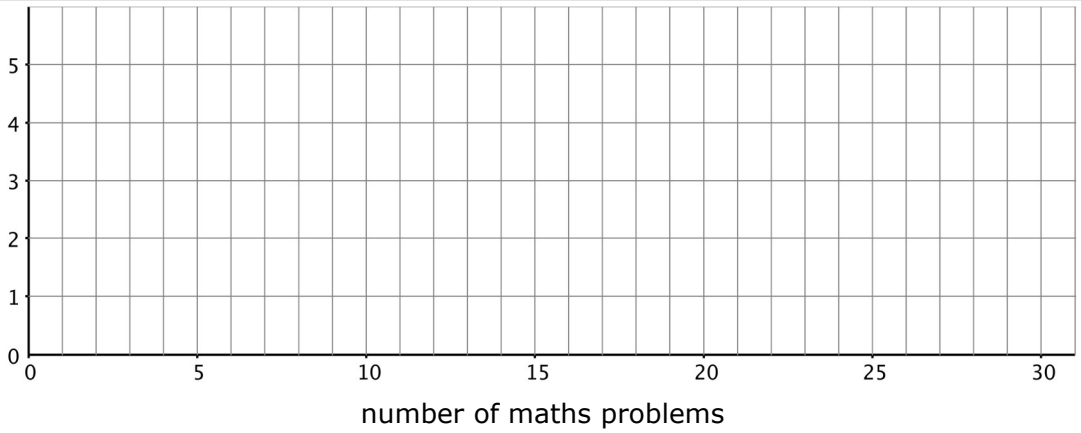


3. Work with your group to draw three histograms to represent Jada's homework data. The width of the bars in each histogram should represent a different number of homework problems, which are specified as follows.

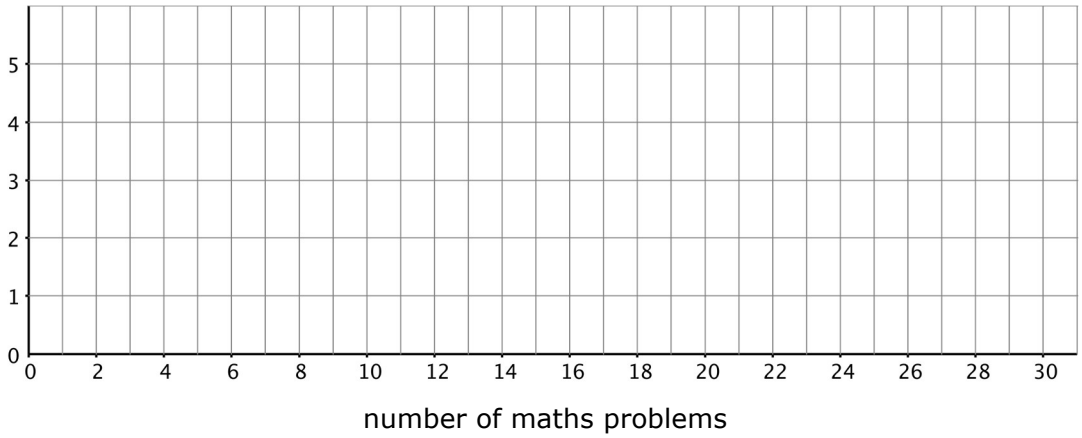
   a. The width of one bar represents 10 problems.



   b. The width of one bar represents 5 problems.



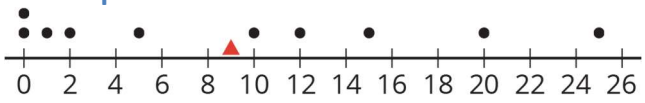
c. The width of one bar represents 2 problems.



4. Which of the five representations should Jada use to summarise her data? Should she use a dot plot, box plot, or one of the histograms? Explain your reasoning.

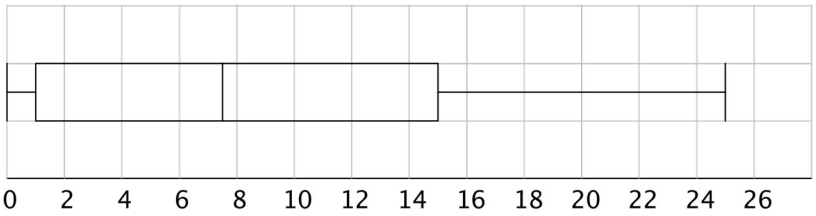
Student Response

1.

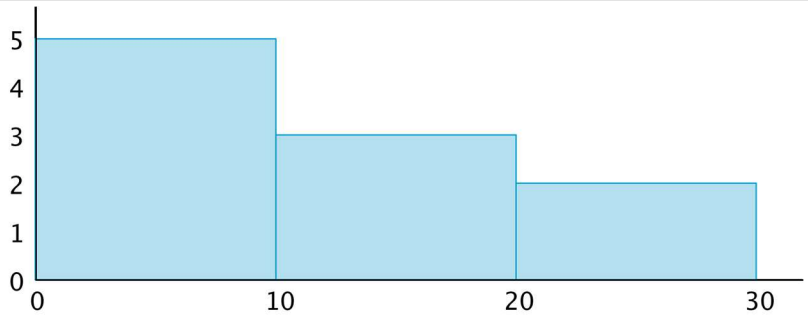


number of math problems

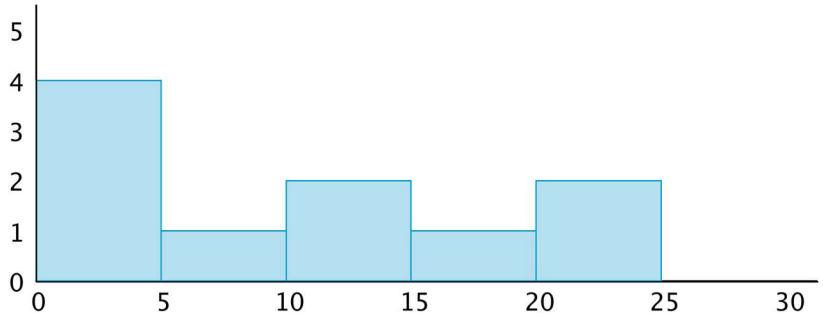
2.

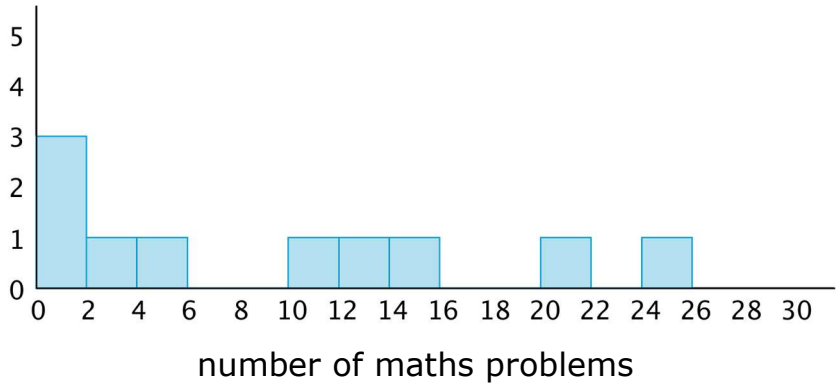


number of maths problems



3.





4. Answers vary. Sample responses:

- Jada should use the dot plot. Because it shows all the data points, we could use it to find the mean and range, or the median and IQR, to summarise the data.
- Jada should use the box plot. It shows that for about three-quarters of the days Jada's class had 15 or fewer problems, so we know that days with more than 15 problems are *not* typical.
- Jada should use the histogram with a bin size of 10 or 5. All the histograms more or less communicate the distribution, but the last one (where one bar represents 2 problems) is not much different from the dot plot. The first histogram (where one bar represents 10 problems) is helpful because tells us

that on most days Jada's class had fewer than 20 problems. The second histogram (where one bar represents 5 problems) gives us more detail but is at the same time harder to summarise or describe. We can see that one half of the days the class had fewer than 10 problems, and the other half they had 10 or more problems. (A sample counter-argument: With only 10 values in the data set, the individual values should be used to gain accuracy rather than combining them into groups and losing that information.)

Activity Synthesis

Invite previously identified students to share their dot plot, box plot, and histograms. Display their drawings for all to see. Then, select several students or groups to share their response to the last question (which representation should Jada choose?) and their explanation. If not already mentioned by students, discuss the different insights that each display offers, or different challenges it poses. (Some possible observations are listed under Student Response section.) For instance, consider asking the following questions about each data display:

- "What information can we get from this display?"
- "Does it give us a meaningful snapshot of the distribution?"
- "What characteristics of a different data set would make this representation more useful?"

Help students see that, in this case, none of the representations here are ill-suited to represent the data set, but a couple of them (e.g. the box plot, or the first histogram with a bin size of 10) allow us to describe the distribution the data set more easily because of how they summarise the data values in some ways.

Representation: Internalise Comprehension. Use colour coding and annotations to highlight differences between how dot plots, box plots, and histograms represent the data.
Supports accessibility for: Visual-spatial processing Representing, Listening: Collect and Display. As students discuss which representation Jada should choose, collect students' responses in a graphic organiser, such as a Venn diagram, and display for all to see. Throughout the remainder of the lesson, continue to update collected student language and remind students to borrow language from the display as needed. Chart language related to dot plot, box plot, and histogram representations. This will help students to use mathematical language during paired and group discussions.
Design Principle(s): Support sense-making; Maximise meta-awareness

18.4 Will the Yellow Perch Survive?

Optional: 30 minutes

In this culminating activity, students use what they have learned in the unit to answer statistical questions about a species of fish in the Great Lakes region. They use a histogram to represent the given data distribution, decide on appropriate measures of

centre and variability, and use their analyses to draw conclusions about a certain fish population.

Instructional Routines

- Group Presentations
- Discussion Supports

Launch

Tell students that they will now look at an example in which data analysis could be used to help conservation efforts. Provide students with the following background information.

The yellow perch is a freshwater fish that is a popular food for people in the Great Lakes region (Minnesota, Wisconsin, Michigan, Illinois, Indiana, Ohio, Pennsylvania, and New York). In past research, samples of yellow perch taken from the Great Lakes seemed to be mostly male and mostly old. People worried that yellow perch might not survive and efforts were made to limit commercial and individual fishing in order to try to increase in the number of younger fish. An important part of these efforts is to periodically check the typical age of the fish in the Great Lakes.

The Wisconsin Department of Natural Resources and the Great Lakes Water Institute collected data from samples of yellow perch in Lake Michigan. Students at Rufus King High School in Milwaukee, Wisconsin participated in the research. They evaluated the data and presented their findings in a student-conducted press conference. Explain to students that, in this task, they will investigate some of the same questions that these students addressed in their research.

Arrange students in groups of 3–4. Provide access to straightedges. Give students 7–8 minutes of quiet work time for the first three questions, and then 10–12 minutes to discuss their responses, complete the remainder of the task, and prepare a brief presentation on their response to the last set of questions.

Give each group access to tools for creating a visual display. Ask them to support their conclusions with specific pieces of evidence, such as their histogram, their analysis of the distribution, measures of centre and spread, etc.

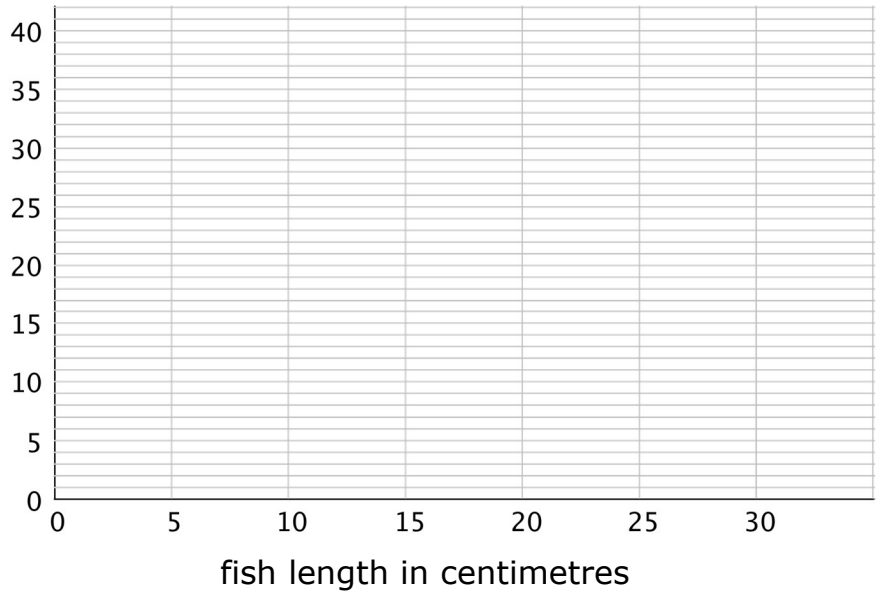
Student Task Statement

Scientists studying the yellow perch, a species of fish, believe that the length of a fish is related to its age. This means that the longer the fish, the older it is. Adult yellow perch vary in size, but they are usually between 10 and 25 centimetres.

Scientists at the Great Lakes Water Institute caught, measured, and released yellow perch at several locations in Lake Michigan. The following summary is based on a sample of yellow perch from one of these locations.

length of fish in centimetres	number of fish
0 to less than 5	5
5 to less than 10	7
10 to less than 15	14
15 to less than 20	20
20 to less than 25	24
25 to less than 30	30

1. Use the data to make a histogram that shows the lengths of the captured yellow perch. Each bar should contain the lengths shown in each row in the table.



2. How many fish were measured? How do you know?

3. Use the histogram to answer the following questions.

   a. How would you describe the shape of the distribution?

   b. Estimate the median length for this sample. Describe how you made this estimate.

   c. Predict whether the mean length of this sample is greater than, less than, or nearly equal to the median length for this sample of fish? Explain your prediction.

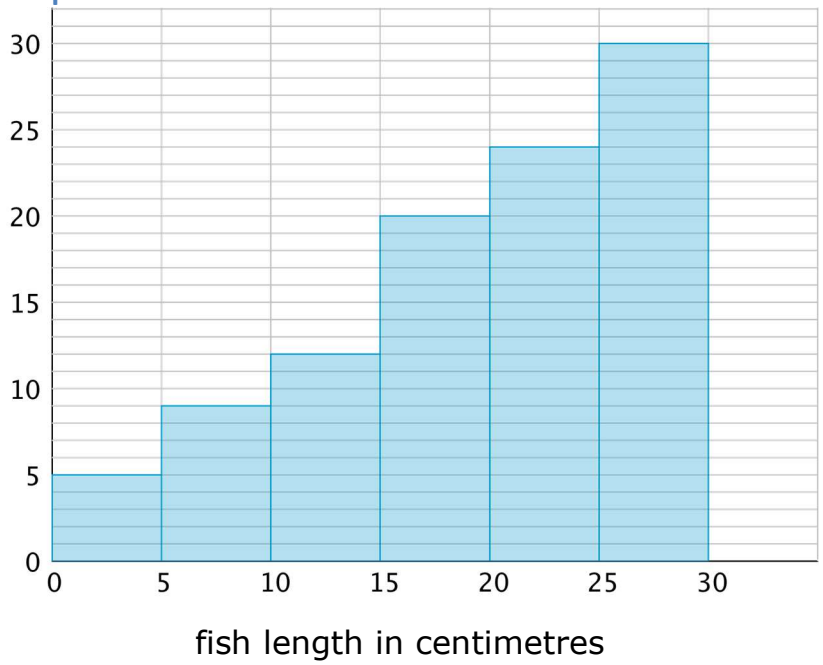
   d. Would you use the mean or the median to describe a typical length of the fish being studied? Explain your reasoning.

4. Based on your work so far:

a. Would you describe a typical age for the yellow perch in this sample as: "young," "adult," or "old"? Explain your reasoning.

b. Some researchers are concerned about the survival of the yellow perch. Do you think the lengths (or the ages) of the fish in this sample are something to worry about? Explain your reasoning.

Student Response



1.

2. 100 fish were measured. The numbers of fish in all length groups add up to 100.

3. Answers vary. Sample responses:

   a. The data is not symmetrical and has a peak in the range 25–30 cm.

   b. I estimate the median to be 22–23 cm. I look in the table for where the 50th and 51st value would be and see that it is in the "20 to less than 25 cm" group.

   c. I predict the mean to be less than the median. Because the data has a peak on the right, the average should be closer to the right as well, but the values that are to the left of the median (or less than the median) would pull the average down.

   d. I would use the median, because it would better describe where the centre of the data is (the data is not symmetric).

   a. I would describe a typical fish as "old" because its length is on the higher end of the range of adult sizes.

b. The lengths seem to be something to worry about. If the yellow perch in the Great Lakes tend to be old, and there are not many young fish around, the species might not survive.

Activity Synthesis

To allow all groups a chance to present, consider putting 2–3 groups together and asking them to present their work to each other. Groups that are not the first to present should focus on sharing new insights that have not been mentioned by the preceding groups. Invite students who are not presenting to attend carefully to the reasoning of the presenting group and to ask clarifying questions.

If time permits, highlight some conclusions that students drew about whether the fish in the sample were young, adult age, or old, and whether researchers should be worried.

Tell students that several years after the students at Rufus King High School participated in the research, newer samples of yellow perch showed more favourable length-age distributions: more of the fish were smaller or younger.

Speaking: Discussion Supports. During the group presentations, provide students with sentence frames such as: "Based on the histogram, I think _____ because . . ." or "I think _____ is/is not something to worry about because . . .". This will help students articulate their ideas, and use mathematical language such as mean, median, range, and interquartile range as supporting evidence during their presentation.
Design Principle(s): Support sense-making; Maximise meta-awareness

Lesson Synthesis

In this lesson we practise finding measures of centre and variability (mean, range, median, and IQR) and making sense of them in the context of the given situation. We notice that they give us different insights into the distribution of a data set.

- "What do the mean and range tell us?" (The mean tells us the fair share or balance point of the distribution and the range tells us the full extent of the data.)
- "How do we interpret this statement: 'Noah's mean number of homework problems per day is 10 and the range is 20.'?" (If we were to distribute Noah's assignments so that the number of problems he has each day is the same, he would have 10 per day. The range of 20 tells us that there is some variability in the number of problems assigned, so not all days have exactly 10 problems assigned. Noah has at least 20 problems on one of the days.)
- "What do the median and IQR tell us?" (The median tells us the value for which half the data set is equal to or greater and half the data set is equal to or less and the IQR tells us the range for the middle half of the data set.)
- "How do we interpret this statement: 'Lin's median number of homework problems per day is 10 and the IQR is 6.'?" (One half of Lin's assignments involve 10 or fewer

problems, and the other half involve 10 or more problems. The IQR tells us that half of Lin's assignments are between 7 and 13 problems.)

We also looked at different ways to graphically represent a numerical distribution.

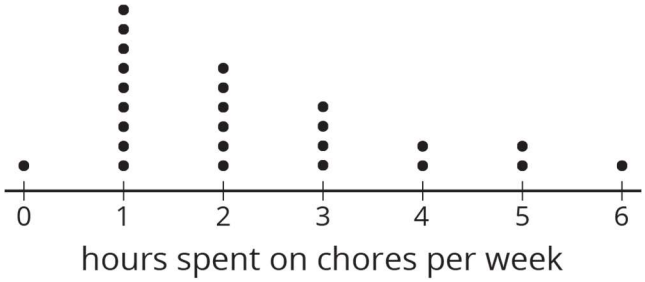
- "What are the ways we can represent a data set?" (Dot plot, histogram, box plot.)
- "Which representations are helpful for summarising a distribution?" (It varies depending on the distribution we're studying and what information we want to know.)

18.5 Time Spent on Chores

Cool Down: 5 minutes

Student Task Statement

Lin surveyed her classmates on the number of hours they spend doing chores each week. She represented her data with a dot plot and a histogram.



1. Lin thinks that she could find the median, the minimum, and the maximum of the data set using both the dot plot and the histogram. Do you agree? Explain your reasoning.
2. Should Lin use the mean and range or the median and IQR to summarise her data? Explain your reasoning.

Student Response

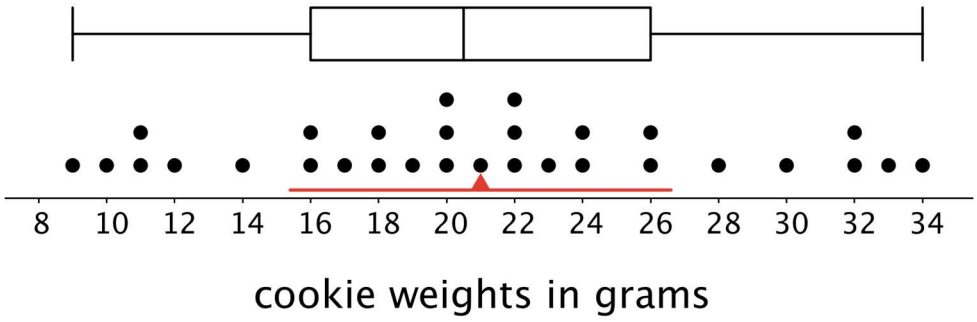
Answers vary. Samples responses.

1. Disagree. The dot plot makes it possible to find the median, the minimum, and the maximum fairly easily since it shows each data value individually. The histogram makes it possible to estimate these values, but it is impossible to tell the exact values because the data points are grouped together.

2. Lin should use the median and IQR as the data is not approximately symmetrical and has values far from the centre. There are a few larger values are not similar to most of the other values.

Student Lesson Summary

The dot plot shows the distribution of 30 cookie weights in grams.

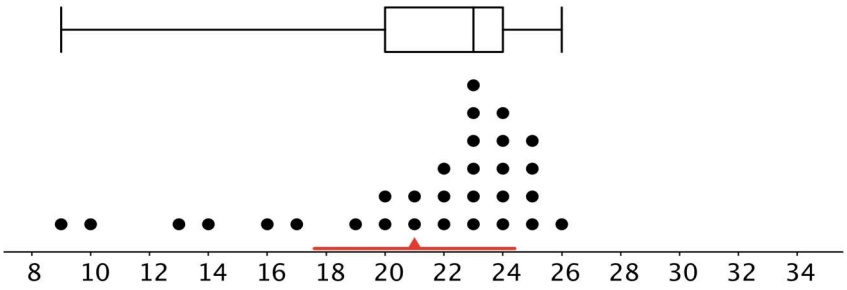


The mean cookie weight, marked by the triangle, is 21 grams. This tells us that if the weights of all of the cookies were redistributed so they all had the same weight, each cookie would weigh 21 grams. The range is 25 grams, which suggests that a cookie roughly weighs between 9 grams and 33 grams.

The box plot for the same data set is shown above the dot plot. The median shows that half of the weights are greater than or equal to 20.5 grams, and half are less than or equal to 20.5 grams. The box shows that the IQR is 10 and that the middle half of the cookies weigh between 16 and 26 grams.

In this case, the median weight is very close to the mean weight, and the IQR is about half the range.

Now let's look at another example of 30 different cookies.



Here the mean is 21 grams, and the range is 17 grams. The median cookie weight is 23 grams, and the box plot shows that the middle half of the data are between 20 and 24 grams. These two pairs of measures paint very different pictures of the variability of the cookie weights.

The median (23 grams) is closer to the middle of the big cluster of values. If we were to ignore the smaller cookies, the median and IQR would give a more accurate picture of how much a cookie typically weighs.

When a distribution is not symmetrical, the median and IQR are often better measures of centre and spread than the mean and range. However the decision on which pair of measures to use depends on what we want to know about the group we are investigating.