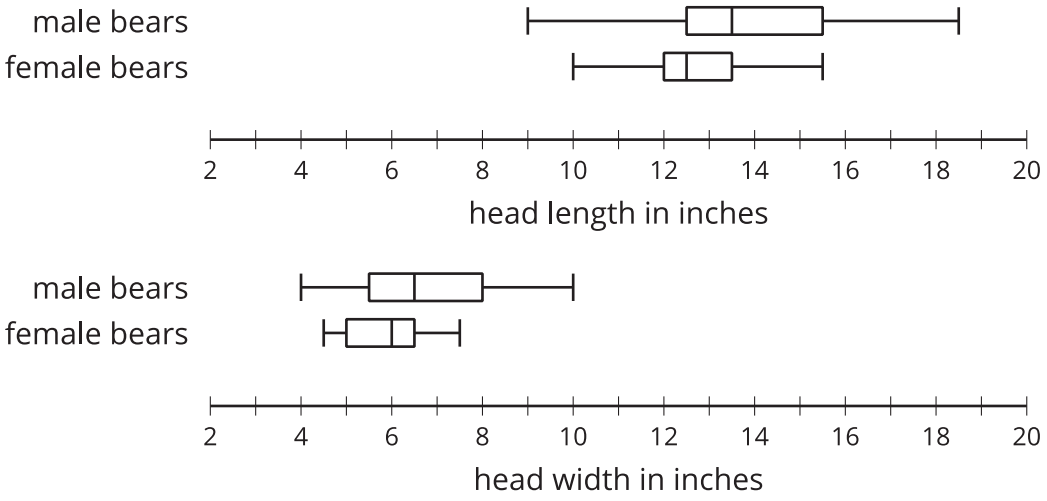
Using Data to Solve Problems

Let's compare data sets using visual displays.

18.1 Wild Bears

In one study on wild bears, researchers measured the head lengths and head widths, in inches, of 143 wild bears. The box plots summarize 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. Trade questions with your partner.

   a. Decide if each question is a statistical question.

   b. Use the box plots to answer each question.

18.2 Homework Analysis (Part 1)

Over a two-week period, Mai records the number of math homework problems she has each school day.

2 15 20 0 5 25 1 0 10 12

1. Calculate these values. Show your reasoning.
   a. The mean number of math homework problems

   b. The mean absolute deviation (MAD)

2. Interpret the mean and MAD. What do they tell you about the number of math 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 center and variability—mean and MAD, or median and IQR—do you think summarizes the distribution of Mai's math homework assignments better? Explain your reasoning.

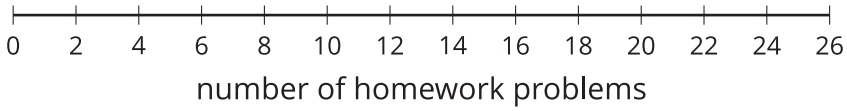
CC BY NC Illustrative Mathematics, based on IM 6–8 Math, CC BY Open Up Resources.

18.3 Homework Analysis (Part 2)

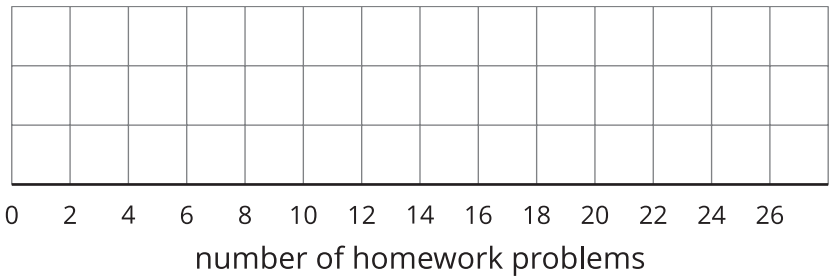
Jada wants to know whether a dot plot, a histogram, or a box plot would best show the distribution of her homework data.

0 0 4 4 5 5 7 8 10 23

1. Use the axis to make a dot plot to represent the data, and indicate the mean of 6.6 with a triangle. The MAD is 4.32.

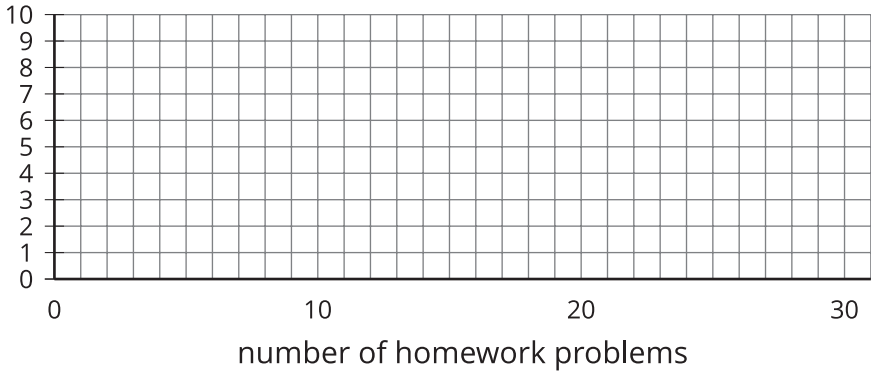


2. Draw a box plot that represents Jada's homework data.

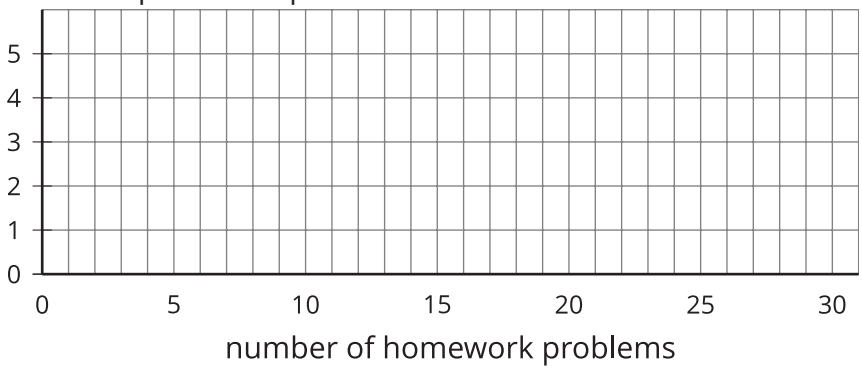


3. Work with your group to draw three histograms to represent Jada's homework data. The width of the bars in each histogram should represent a different number of homework problems.

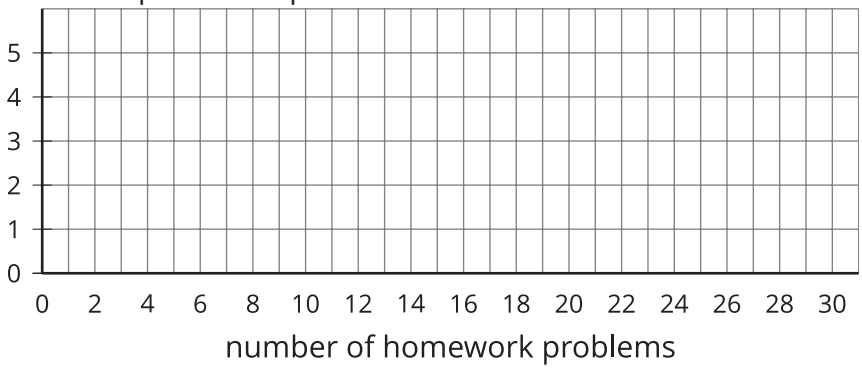
   a. The width of one bar represents 10 problems.



   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 summarize her data? Explain your reasoning.

CC BY NC Illustrative Mathematics, based on IM 6–8 Math, CC BY Open Up Resources.

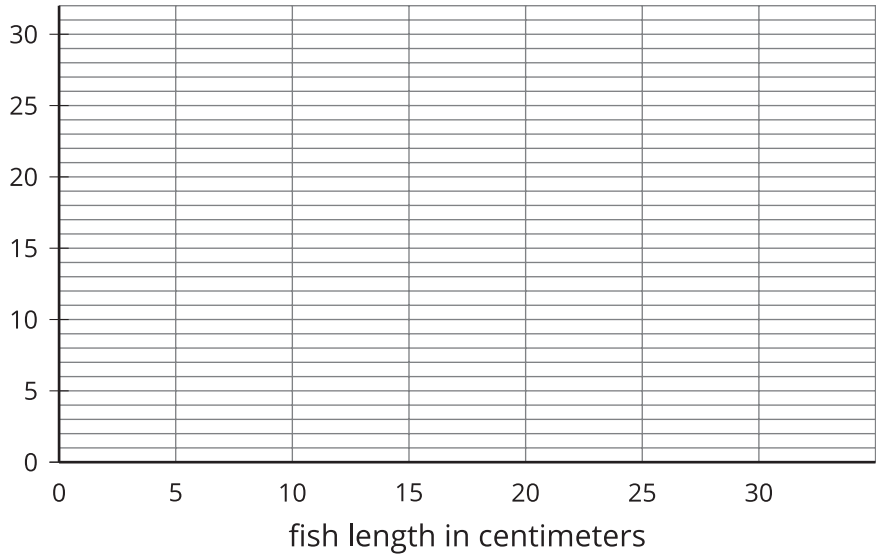
18.4 Will the Yellow Perch Survive?

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 centimeters.

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

length of fish in centimeters	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 these 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.

CC BY NC Illustrative Mathematics, based on IM 6–8 Math, CC BY Open Up Resources.

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.

18.5 Butterfly Measurements

Navigate to this activity in the digital version of the materials.

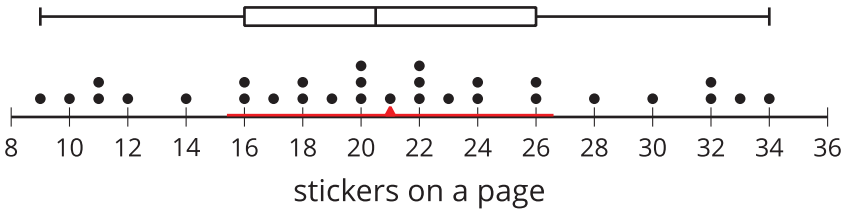
The applet contains two data sets, each containing 127 measurements of wingspan or body length of butterflies captured in an area.

Select one data set to analyze. Use the applet to help create a display that summarizes the information, including any values calculated. Make sure that your display contains:

- At least one display of the distribution discussed in this unit, as well as a description of why you chose that method to display the distribution.
- A value for a measure of center, along with an explanation of why you chose that measure.
- A value for a measure of variability, along with an explanation of why you chose that measure.
- A few sentences describing the distribution, including any additional features you notice.

Lesson 18 Summary

The data displays show the distribution of stickers on 30 pages.

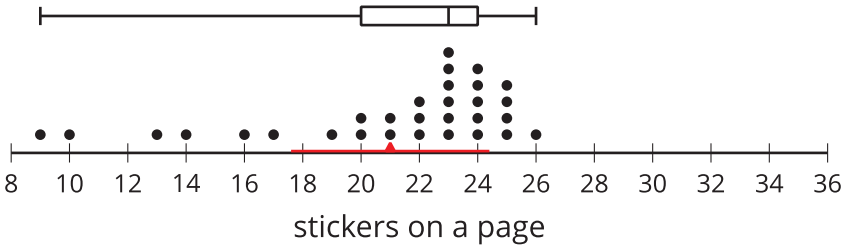


The mean number of stickers, marked by the triangle, is 21. This tells us that if the total number of stickers were redistributed so that each page had the same number of stickers, then each page would have 21. The MAD is 5.6 stickers, which suggests that a page typically has between 15.4 stickers and 26.6 stickers.

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

In this case, the median number of stickers is very close to the mean number of stickers, and the IQR is about twice the MAD. This tells us that the two pairs of measures of center and spread are very similar.

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



Here the mean is 21 stickers, and the MAD is 3.4 stickers. This suggests that a page typically has between 17.6 and 24.4 stickers. The median number of stickers is 23, and the box plot shows that the middle half of the data are between 20 and 24 stickers. These two pairs of measures paint very different pictures of the variability of the number of stickers on a page.

From the dot plot, we can see that the median (23 stickers) is closer to the middle of the big cluster of values. If we were to ignore the pages with very few stickers, the median and IQR would give a more accurate picture of how many stickers are typically on a page.

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

CC BY NC Illustrative Mathematics, based on IM 6–8 Math, CC BY Open Up Resources.