Lesson 17: Using box plots

Goals

- Compare and contrast (orally and in writing) box plots that represent different data sets, including ones with the same median but very different IQRs and vice versa.
- Determine what information is needed to solve problems about comparing box plots. Ask questions to elicit that information.
- Interpret a box plot to answer (orally) statistical questions about a data set.

Learning Targets

- I can use a box plot to answer questions about a data set.
- I can use medians and IQRs to compare groups.

Lesson Narrative

In the previous lesson, students analysed a dot plot and a box plot in order to study the distribution of a data set. They saw that, while the box plot summarises the distribution of the data and highlights some key measures, it was not possible to know all the data values of the distribution from the dot plot alone. In this lesson, students use box plots to make sense of the data in context, compare distributions, and answer statistical questions about them.

Students compare box plots for distributions that have the same median but different IQRs, as well as box plots with the same IQRs but different medians. They recognise and articulate that the centres are the same but the spreads are different in the first case, and the centres are different but the spreads are the same in the second case. They use this understanding to compare typical members of different groups in terms of the context of the problem.

Addressing

- Recognise a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students' ages.
- Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
- Summarise numerical data sets in relation to their context, such as by:

Instructional Routines

- Anticipate, Monitor, Select, Sequence, Connect
- Information Gap Cards

- Compare and Connect
- Think Pair Share

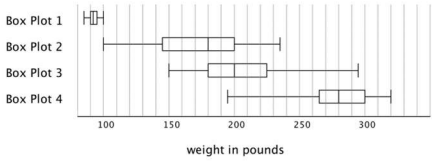
Required Materials

Pre-printed slips, cut from copies of the blackline master

Info Gap: Sea Turtles
Problem Card 1

The box plots represent data on the weights of several species of sea turtles that nest at the Outer Banks of North Carolina. Two of the species are the olive ridley and the hawksbill.

Which two box plots represent the data for the weights of olive ridley and hawksbill sea turtles?



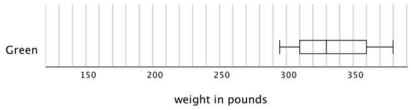
Info Gap: Sea Turtles
Data Card 1

- The median for the hawksbill data is about twice the median for the olive ridley data.
- The range for the hawksbill data is 135 pounds.
- The IQR for the olive ridley data is 5 pounds and the range is 15 pounds.
- The heaviest weight in the olive ridley data is 100 pounds.

Info Gap: Sea Turtles
Problem Card 2

Loggerhead and green sea turtles are two species that also nest at the Outer Banks. Here is the box plot for a set of data on weights of green sea turtles.

a. Which of the two species has a heavier typical weight? Explain how you know.
b. Which species shows more variation in weight? Explain how you know.



Info Gap: Sea Turtles
Data Card 2

- The IQR for the loggerhead data is 45 pounds.
- The lightest weight in the loggerhead data is 150 pounds.
- The median for the loggerhead data is 130 pounds less than the median for the green turtle data.
- The heaviest weight in the loggerhead data is 295 pounds.

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.

Required Preparation

Print and cut up slips from the Sea Turtles Info Gap blackline master. Prepare 1 set for every 2 students. Provide access to straightedges for drawing box plots. Consider creating a few paper planes of different sizes or styles to fly for the Paper Planes activity.

Student Learning Goals

Let's use box plots to make comparisons.

17.1 Hours of Slumber

Warm Up: 5 minutes

This warm-up allows students to practice creating a box plot from a five-number summary and think about the types of questions that can be answered using the box plot. To develop questions based on the box plot prompts students to put the numbers of the five-number summary into context.

As students work, identify a student who has clearly and correctly drawn the box plot to share during the whole-class discussion.

For the second question, some students may write decontextualised questions that are simply about parts of the box plot (e.g., "What is the IQR?" or "What is the range?"). Others might write contextualised questions that the box plot could help to answer (e.g., "What is the least amount of sleep in this data set?" or "What is the median number of hours of sleep for this group?"). Identify a few students from each group so that they can share later.

Instructional Routines

- Anticipate, Monitor, Select, Sequence, Connect
- Think Pair Share

Launch

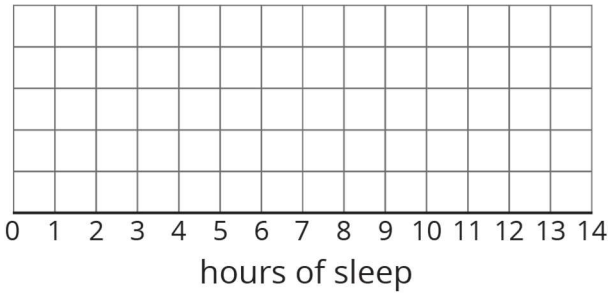
Arrange students in groups of 2. Give students 2 minutes of quiet work time, followed by a whole-class discussion.

Student Task Statement

Ten year 7 students were asked how much sleep, in hours, they usually get on a school night. Here is the five-number summary of their responses.

- Minimum: 5 hours
- Lower quartile: 7 hours
- Median: 7.5 hours
- Upper quartile: 8 hours
- Maximum: 9 hours

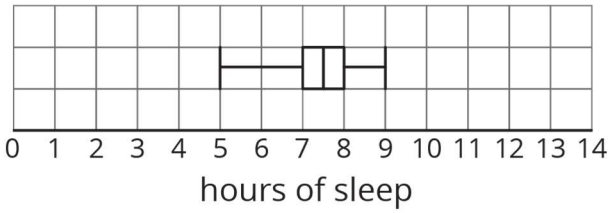
1. On the grid, draw a box plot for this five-number summary.



2. What questions could be answered by looking at this box plot?

Student Response

1.



2. Answers vary. Possible responses:
   - What is the least amount of sleep a student usually gets on a school night?
   - How many students usually get at least 7 hours of sleep on a school night?

Activity Synthesis

Select the previously identified student with a correct box plot to display it for all to see. If that is not possible, ask them to share how they drew the box plot and record and display the drawing based on their directions.

Select other previously identified students to share questions that could be answered by looking at the box plot first those that can be answered without the context followed by questions that rely on the context. Record and display these questions for all to see. After each question, ask the rest of the class if they agree or disagree that the answer can be found using the box plot. If time permits, ask students for the answer to each shared question.

Point out questions that are contextualised versus those that are not. Explain that a box plot can help to make sense of a data set in context and answer questions about a group or a characteristic of a group in which we are interested. The different measures that we learned to identify or calculate help to make sense of data distribution in context.

17.2 Info Gap: Sea Turtles

15 minutes

In this info gap activity, students practice analysing box plots and thinking carefully about what questions they could help to answer. They connect features of box plots to information about IQR, range, median, and minimum and maximum values of several data sets on sea turtles. Along the way, they see the usefulness of box plots in comparing distributions and characteristics of different populations.

The info gap structure requires students to make sense of problems by determining what information is necessary, and then to ask for information they need to solve it. This may take several rounds of discussion if their first requests do not yield the information they need. It also allows them to refine the language they use and ask increasingly precise questions until they get the information they need.

Instructional Routines

- Information Gap Cards

Launch

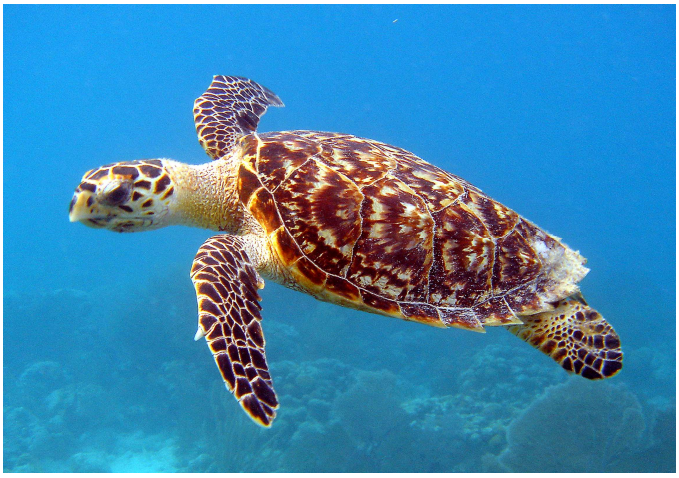
Tell students that they will now practise using box plots to answer questions about populations of sea turtles. Provide students with the following background information: Sea turtles are air-breathing amphibians that spend most of their time floating in seaweed beds. Females come on shore to lay their eggs. One nesting place for several species of sea turtles is the Outer Banks of North Carolina. All sea turtles are considered to be endangered and only about 1 in 1 000 newly hatched sea turtles survives, which is why their nesting areas are protected.

Arrange students in groups of 2. Give each group a pair of cut-up cards from the blackline master—a Problem Card for one partner and a Data Card for the other. Remind students of the protocol for asking and answering questions, as shown on the task statement. Give students 10 minutes to complete the activity. If any groups finish early, offer a second pair of cards and ask the partners to switch roles.

Engagement: Develop Effort and Persistence. Display or provide students with a physical copy of the written directions. Check for understanding by inviting students to rephrase directions in their own words. Keep the display of directions visible throughout the activity.
Supports accessibility for: Memory; Organisation Conversing: This activity uses *Information Gap* to give students a purpose for discussing information necessary to connect features of a box plot to information about the five-number summary. Display questions or question starters for students who need a starting point such as: "Can you tell me . . . (specific piece of information)", and "Why do you need to know . . . (that piece of information)?"
Design Principle(s): Cultivate Conversation

Student Task Statement

Your teacher will give you either a Problem Card or a Data Card about sea turtles that nest on the Outer Banks of North Carolina. Do not show or read your card to your partner.



If your teacher gives you the *problem card*:

1. Silently read your card and think about what information you need to be able to answer the question.
2. Ask your partner for the specific information that you need.
3. Explain how you are using the information to solve the problem.

   Continue to ask questions until you have enough information to solve the problem.
4. Share the *problem card* and solve the problem independently.
5. Read the *data card* and discuss your reasoning.

If your teacher gives you the *data card*:

1. Silently read your card.
2. Ask your partner *"What specific information do you need?"* and wait for them to *ask* for information.

   If your partner asks for information that is not on the card, do not do the calculations for them. Tell them you don't have that information.
3. Before sharing the information, ask "*Why do you need that information?*" Listen to your partner's reasoning and ask clarifying questions.
4. Read the *problem card* and solve the problem independently.
5. Share the *data card* and discuss your reasoning.

Pause here so your teacher can review your work. Ask your teacher for a new set of cards and repeat the activity, swapping roles with your partner.

Student Response

Problem Card 1: Box plot 1 represents the data for the olive ridley sea turtles and box plot 2 represents the data for the hawksbill.

Problem Card 2:

1. The green sea turtles have a heavier typical weight because their median weight is higher.
2. The loggerhead sea turtles show greater variability. Their IQRs are about the same (50 pounds for the green sea turtles and 45 for the loggerhead), but the range of weights for the loggerhead sea turtles is much larger (145 pounds, compared to 85 pounds for the green sea turtles).

Activity Synthesis

Invite students to share their experiences with the Info Gap activity. Consider discussing some of the following questions.

- For students who had a Problem Card:
  - "How did you decide what information to ask for? How did the information on your card help?"
  - "How easy or difficult was it to explain why you needed the information you were asking for?"
  - "Give an example of a question that you asked, the clue you received, and how you made use of it."
  - "How many questions did it take for you to be able to solve the problem? What were those questions?"
  - "Was anyone able to solve the problem with a different set of questions?"
- For students who had a Data Card:
  - "When you asked your partner why they needed a specific piece of information, what kind of explanations did you consider acceptable?"
  - "Were you able to tell from their questions what statistical question they were trying to answer? If so, how? If not, why might that be?"

17.3 Paper Planes

15 minutes

In this lesson, students continue to create box plots from data sets. They compare and interpret box plots for distributions with the same median but very different IQRs, and use the plots to answer questions.

As students work, make sure that they correctly identify the five-number summary of each data set. If students have trouble making comparisons, prompt them to study the medians, IQRs, and ranges the data sets. Then, notice how they compare the box plots and whether they interpret the different measures in the context of the given situation. If they make comparisons only in abstract terms (e.g., "The median for both data sets are the same"), push them to specify what the comparisons mean in this situation (e.g., "What does the equal median tell us in this context?"). Identify students who made sense of these numbers in terms of typical distances and consistency of the flights of each person's plane. Ask them to share later.

Instructional Routines

- Compare and Connect

Launch

Tell students that they will analyse data sets about flight distances of paper airplanes. To familiarise students with the context of this activity, consider preparing a few different styles or sizes of paper airplanes. Before students begin working, fly each paper plane a couple of times and ask students to observe their flight distances.

Arrange students in groups of 3–4. Provide access to straightedges. Give groups 8–10 minutes to complete the activity. Ask each group member to find the five-number summary and draw the box plot for *one* student (Andre, Lin, or Noah) and then share their summaries and drawings. Ask them to pause and have their summaries and drawings reviewed before answering the last two questions. Consider posting somewhere in the classroom the five-number summaries and the box plots so that students can check their answers. Ask students to be prepared to explain how Andre, Lin, and Noah's flight distances are alike or different.

Engagement: Develop Effort and Persistence. Display or provide students with a physical copy of the written directions. Check for understanding by inviting students to rephrase directions in their own words. Keep the display of directions visible throughout the activity.
Supports accessibility for: Memory; Organisation

Student Task Statement

Andre, Lin, and Noah each designed and built a paper airplane. They launched each plane several times and recorded the distance of each flight in yards.

Andre

25	26	27	27	27	28	28	28	29	30	30

Lin

20	20	21	24	26	28	28	29	29	30	32

Noah

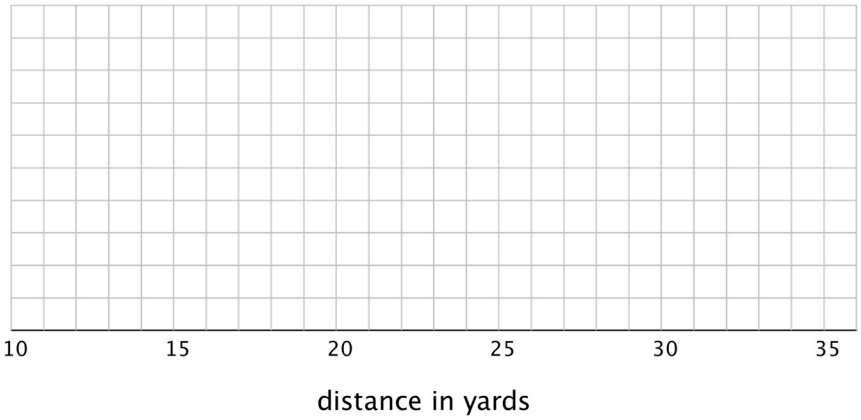
13	14	15	18	19	20	21	23	23	24	25

Work with your group to summarise the data sets with numbers and box plots.

1. Write the five-number summary for the data for each airplane. Then, calculate the interquartile range for each data set.

min	Q1	median	Q3	max	IQR

2. Draw three box plots, one for each paper airplane. Label the box plots clearly.



3. How are the results for Andre and Lin's planes the same? How are they different?

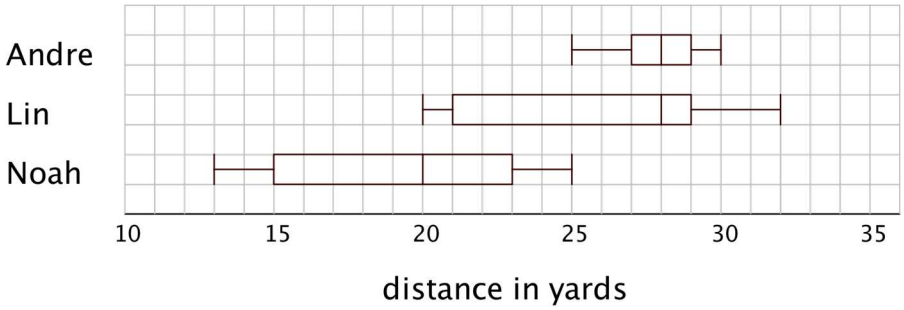
4. How are the results for Lin and Noah's planes the same? How are they different?

Student Response

1.

min	Q1	Q2 (median)	Q3	max	IQR
25	27	28	29	30	2
20	21	28	29	32	8
13	15	20	23	25	8

2.



3. Andre and Lin's results were similar in that they had the same median flight length of 28 yards. Their outcomes were different because Andre's plane travelled more

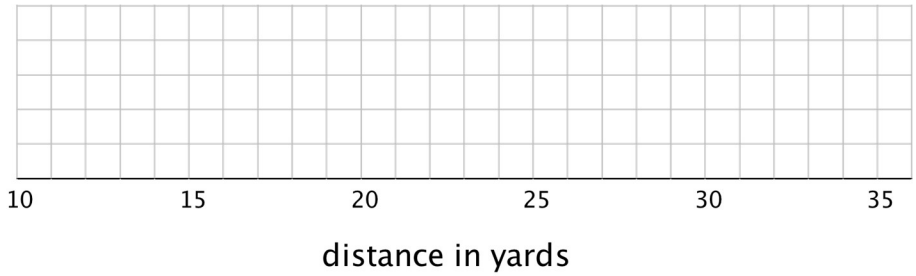
consistent distances than Lin's plane, which travelled more variable distances, as shown by her plane's larger IQR.

4. Lin and Noah's planes had the same IQR of 8 yards, meaning that their planes had a similar amount of variability in flight distance. Noah's median was much lower than Lin's, however, so generally speaking, his planes fly a much shorter distance.

Are You Ready for More?

Priya joined in the paper-plane experiments. She launched her plane eleven times and recorded the lengths of each flight. She found that her maximum and minimum were equal to Lin's. Her IQR was equal to Andre's.

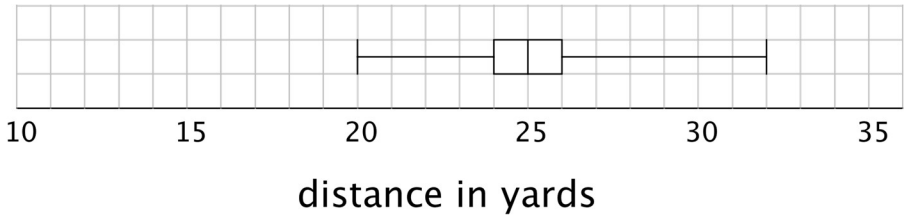
Draw a box plot that could represent Priya's data.



With the information given, can you estimate the median for Priya's data? Explain your reasoning.

Student Response

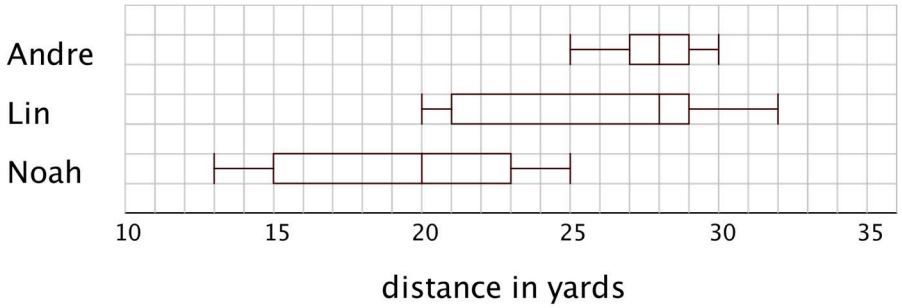
Box plots vary. Sample box plot:



Answers vary. Sample response: No, I cannot estimate the median. I know the IQR, but I don't have enough information to tell where the first, second, or third quartiles are. The box could be anywhere between 20 and 32, and the median could be anywhere in that box.

Activity Synthesis

Focus the whole-class discussion on students' analyses and interpretations of the box plots. Display the box plots for all to see.



Select a few students or groups to share their responses comparing Andre's and Lin's data. Be sure to discuss what it means when two data sets have the same median but different IQRs, as in Andre's and Lin's cases. If no students connect these values to the centre and spread and data, ask them to do so.

- "What can you say about the centre of Andre's data and that of Lin's data?" (They have the same centre of 28 yards.)
- "What does the same centre tell us in this context?" (The same number could be used to describe a typical flight distance for both Andre's and Lin's flight distances.)
- "What can you say about the spread of Andre's data and that of Lin's data?" (Andre's data are far more concentrated than Lin's, which are quite spread out. The range of her data is 12 yards, which tells us there is much more variability in her flight distances. For Andre the range is only 5 yards, less than half of Lin's, and his IQR is a quarter of Lin's. Overall, there is much less variability in his data.)
- "Which of the two planes—Andre's or Lin's—flies a more consistent distance? How do you know?"(Andre's, because the spread of his data is much smaller.)

Then, select a few other students or groups to compare Lin's and Noah's data. Be sure to discuss what it means when two data sets have the same spread (IQR) but different medians.

- "What do the two very different centres tell us in this context?" (Generally speaking, a typical flight distance for Lin's plane is quite different than that for Noah's.)
- "What can you say about the spreads of Lin's and Noah's data?" (Both sets have the same range and the same IQR, though the values of the quartiles are different for the two sets of data.)
- "What does the same range tell us in this case?" (The difference between the shortest flight distance and the longest one is the same for both data sets.)
- "What does the same IQR tell us in this case?" (The middle half of the two sets of data cover the same distance.)
- "Whose plane—Lin's or Noah's—flies a more consistent distance?" (Their planes fly with very similar consistency, or inconsistency. The identical IQR and range tell us that their data have very similar variability.)

Representing, Listening: Compare and Connect. Use this routine to support students as they interpret the box plot representations of the data. Ask students, "What is the same and what is different among the box plots?" Invite students to use the language of median, IQR, distribution, centre, spread as they discuss comparisons. Ask students to make connections in the box plot diagrams to the values in the 5-number summary table. For example, "Andre's IQR is 2 in the table which is the length of the rectangle in Andre's box plot." These exchanges strengthen students' mathematical language related to representations of data sets used and help them connect what the box plots tell us about the centre and spread of data sets.
Design Principle(s): Optimise output (for comparison); Maximise meta-awareness

Lesson Synthesis

In this lesson, we see that box plots can tell us stories about the centre and spread of data sets.

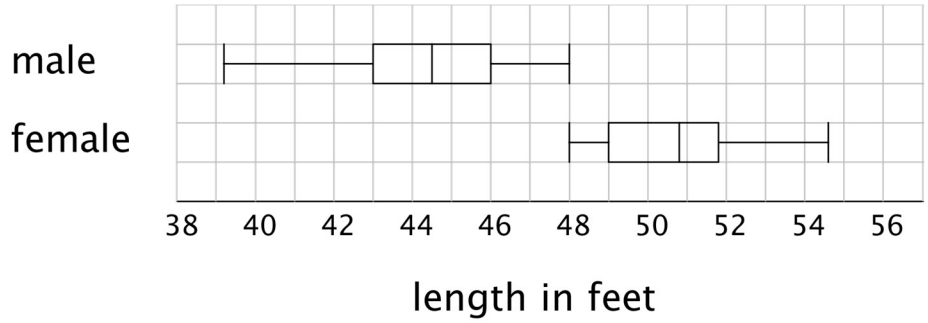
- "What are some questions you can ask to match box plots to data?" (Questions about the 3 quartiles, maximum, or minimum will help distinguish the different box plots.)
- "What does it mean when two box plots show the same median but different IQRs?" (The data they summarise have the same centre but different spreads.)
  - "How can we see this in a box plot?" (The lines inside the box will be at the same place, but the widths of the boxes will be different.)
  - "What does the same median, different IQRs' mean in context?" (It means that we can use the same number to describe what is typical for each group, but the variability in the data is different.)
- "What does it mean when two box plots show the same IQR but different medians?" (The data they summarise have different centres but the same spread.)
  - "How can we see this in a box plot?" (The Q2 segments in the boxes are located in different positions along the number line, but the boxes have the same width.)
  - "What does 'the same IQR, different medians' tell us in context?" (It means that what is typical for each group is different, but the variability is similar for the two groups.)

17.4 Humpback Whales

Cool Down: 5 minutes

Student Task Statement

Researchers measured the lengths, in feet, of 20 male humpback whales and 20 female humpback whales. Here are two box plots that summarise their data.



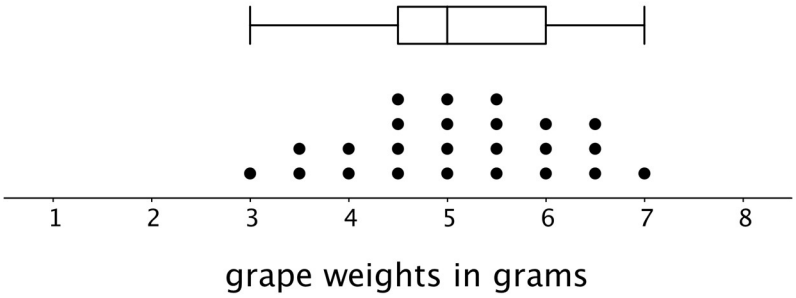
1. How long was the longest whale measured? Was this whale male or female?
2. What was a typical length for the male humpback whales that were measured?
3. Do you agree with each of these statements about the whales that were measured? Explain your reasoning.
    a. More than half of male humpback whales measured were longer than 46 feet.
    b. The male humpback whales tended to be longer than female humpback whales.
    c. The lengths of the male humpback whales tended to vary more than the lengths of the female humpback whales.

Student Response

1. The longest whale was about 55 feet long and was a female.
2. A typical male humpback whale was about 44.5 feet long.
    a. Disagree. Sample explanation: The upper quartile of the data for the male humpbacks is 46 feet, which means a quarter of the whales are longer than 46 feet.
    b. Disagree. Sample explanation: The entire distribution for the lengths of female humpbacks is greater than that for male humpbacks, so female humpbacks tend to be longer than their male counterparts.
    c. Agree. Sample explanation: The IQR of the data for male humpbacks is slightly greater than that for female humpbacks, and the range of the data for the males is larger than that for females, so the lengths of male humpbacks tend to vary more.

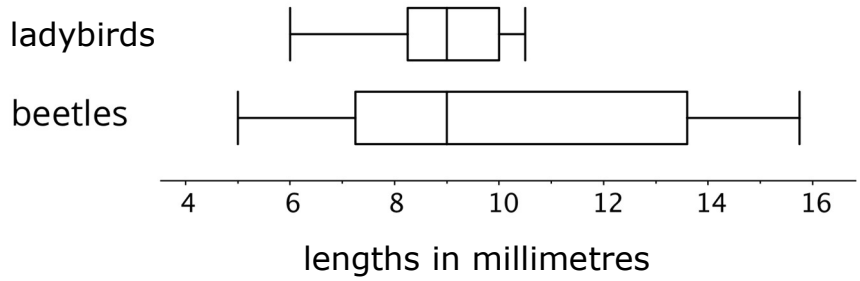
Student Lesson Summary

Box plots are useful for comparing different groups. Here are two sets of plots that show the weights of some berries and some grapes.



Notice that the median berry weight is 3.5 grams and the median grape weight is 5 grams. In both cases, the IQR is 1.5 grams. Because the grapes in this group have a higher median weight than the berries, we can say a grape in the group is typically heavier than a berry. Because both groups have the same IQR, we can say that they have a similar variability in their weights.

These box plots represent the length data for a collection ladybirds and a collection of beetles.

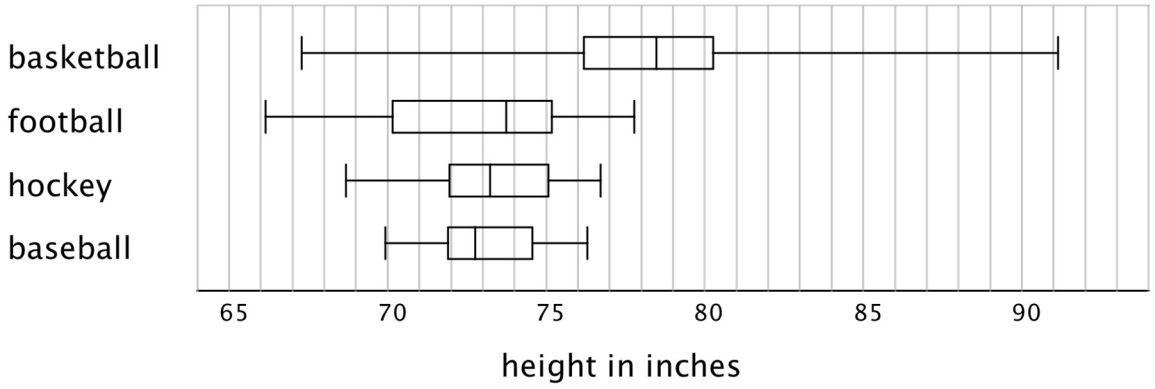


The medians of the two collections are the same, but the IQR of the ladybirds is much smaller. This tells us that a typical ladybird length is similar to a typical beetle length, but the ladybirds are more alike in their length than the beetles are in their length.

Lesson 17 Practice Problems

Problem 1 Statement

Here are box plots that summarise the heights of 20 professional male athletes in basketball, football, hockey, and baseball.



a. In which two sports are the players' height distributions most alike? Explain your reasoning.

b. Which sport shows the greatest variability in players' heights? Which sport shows the least variability?

Solution

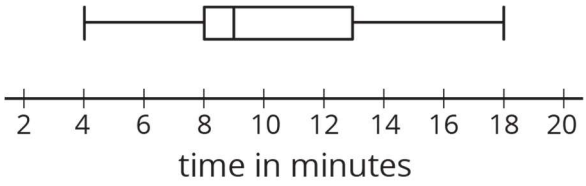
a. Hockey and baseball players are most alike. Sample explanation: The two medians are very close (around 73 inches each), their IQRs differ by only about $\frac{1}{2}$ inch.

b. Overall, basketball players show the greatest variability in height (indicated by the largest range). Variability for the middle half of data is the greatest for football players (shown by the largest IQR). Baseball players show the least variability in height (shown by the smallest range and IQR).

Problem 2 Statement

Here is a box plot that summarises data for the time, in minutes, that a fire department took to respond to 100 emergency calls.

Select **all** the statements that are true, according to the dot plot.



a. Most of the response times were under 13 minutes.

b. Fewer than 30 of the response times were over 13 minutes.

c. More than half of the response times were 11 minutes or greater.

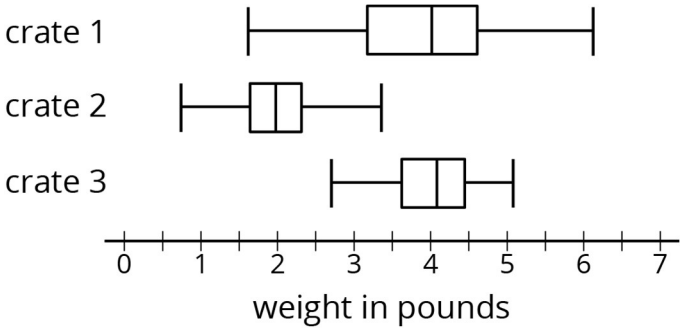
d. There were more response times that were greater than 13 minutes than those that were less than 9 minutes.

e. About 75% of the response times were 13 minutes or less.

Solution ["A", "B", "E"]

Problem 3 Statement

Pineapples were packed in three large crates. For each crate, the weight of every pineapple in the crate was recorded. Here are three box plots that summarise the weights in each crate.



Select **all** of the statements that are true, according to the box plots.

a. The weights of the pineapples in crate 1 were the most variable.

b. The heaviest pineapple was in crate 1.

c. The lightest pineapple was in crate 1.

d. Crate 3 had the greatest median weight and the greatest IQR.

e. More than half the pineapples in crate 1 and crate 3 were heavier than the heaviest pineapple in crate 2.

Solution ["A", "B", "E"]

Problem 4 Statement

Two TV shows each asked 100 viewers for their ages. For one show, the mean age of the viewers was 35 years and the range was 60 years. For the other show, the mean age of the viewers was 30 years and the range was 15 years.

A year 7 student says he watches one of the shows. Which show do you think he watches? Explain your reasoning.

Solution

The first show. Explanations vary. Sample explanation: Even though the second show has a lower mean, the much higher range of the first show means its viewers have a wider age range.