Lesson 7: Using histograms to answer statistical questions

Goals

- Compare and contrast (in writing) histograms that represent two different data sets measuring the same quantity.
- Critique (orally) a description of a distribution, recognising that there are multiple valid ways to describe its centre and spread.
- Describe (orally and in writing) the distribution shown on a histogram, including making claims about the centre and spread.

Learning Targets

- I can draw a histogram from a table of data.
- I can use a histogram to describe the distribution of data and determine a typical value for the data.

Lesson Narrative

In this lesson, students create, read, and interpret histograms. They characterise the distribution displayed in a histogram in terms of its shape and spread, and identify a measurement that is typical for the data set by looking for the centre in a histogram. Students also use histograms to make comparisons and to better understand what different spreads and values of the centre mean in a given context.

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.
- 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.
- Display quantitative data in plots on a number line, including dot plots, histograms, and box plots.
- Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

Instructional Routines

- Stronger and Clearer Each Time
- Compare and Connect

- Think Pair Share
- Which One Doesn't Belong?

Required Materials

Rulers marked with centimetres

Student Learning Goals

Let's draw histograms and use them to answer questions.

7.1 Which One Doesn't Belong: Questions

Warm Up: 5 minutes

The purpose of this warm-up is to encourage students to connect the ideas they learned earlier about statistical questions and types of data (qualitative and quantitative) to the work on describing distributions (centre and spread).

There are many ways to interpret the questions and identify how each one is unique. For example, they could say that Question A is about retirement, Question B about jobs, etc. If students begin to depart from thinking in statistical terms, remind them to think about how we might go about answering the questions and what the answers might involve.

Instructional Routines

- Which One Doesn't Belong?

Launch

Arrange students in groups of 2–4. Display the questions for all to see. Give students 1 minute of quiet think time and ask students to indicate when they have noticed one question that does not belong and can explain why. When the minute is up, give students 2 minutes to share with their group their reasoning on why a question doesn't belong, and then, together, find at least one reason each question doesn't belong.

Student Task Statement

Here are four questions about the population of Alaska. Which question does not belong? Be prepared to explain your reasoning.

1. In general, at what age do Alaska residents retire?
2. At what age can Alaskans vote?
3. What is the age difference between the youngest and oldest Alaska residents with a full-time job?
4. Which age group is the largest part of the population: 18 years or younger, 19–24 years, 25–34 years, 35–44 years, 45–54 years, 55–64 years, or 65 years or older?

Student Response

Answers vary. Sample responses:

1. It is the only question that asks for the centre of data (containing the ages at which Alaska residents retire).
2. It is the only non-statistical question.
3. It is the only question that asks for the spread of data.
4. It is the only one that requires qualitative data (or requires rearranging quantitative data into categories). All the other questions require quantitative data.

Activity Synthesis

Ask each group to share one reason why a particular question does not belong. Record and display the responses for all to see. After each response, ask the class if they agree or disagree. Since there is no single correct answer for why a question does not belong, attend to students' explanations and ensure the reasons given are correct.

If students use terms that are essential in this unit (such as centre, spread, statistical, non-statistical, quantitative data, qualitative data, etc.), ask them to explain their meanings in their own words; these are opportunities to reinforce their understanding of the terms and to note any misconceptions. If students give unsubstantiated claims, ask them to substantiate them.

7.2 Measuring Earthworms

20 minutes

In a previous lesson, students had a lesson on histograms—how they are drawn, how they differ from dot plots, and what information they can tell us. In this activity, students practise drawing a histogram for a given data set and using it to answer statistical questions. To help students understand the lengths involved in the data set, students are asked to draw various lengths used to group the worms in the first histogram.

As students organise the data set and draw their histogram, notice any challenges or questions students come across. If a question is raised by multiple groups, consider discussing it with the whole class. Also pay attention to how students use the histogram to identify a "typical" length—some might describe it in terms of the size of the bins (e.g., "a typical length is between 20 and 40 mm"); others might choose a value within a bin or a boundary between bins. Invite them to share their reasonings later.

Instructional Routines

- Compare and Connect
- Think Pair Share

Launch

Arrange students in groups of 2. Provide access to centimetre rulers.

Consider giving students a brief overview of the context for the problems in the activity. Tell students that there are nearly 6 000 species of earthworms in the world. Some earthworms help the environment, while others (generally not native to the region in which they are found) may harm the environment. Earthworms that are native to a particular region of the world are often raised, by farmers, in terrariums (a container or bin similar to an aquarium but it contains soil and leaves). The terrarium-raised earthworms provide bait for people who fish, provide food for various wildlife, and decompose food waste into soil. Food waste and water are added to the terrariums as food for raising and growing worms. Soil produced by the worms as they eat the food waste is often used as fertiliser.

Explain that the lengths of the worms in the bins provide information about the ages of the worms, which can be useful for the farmer. In this activity, students will organise the lengths of the earthworms in several terrariums or bins.

Give students 8–10 minutes of quiet work time, and then 3–4 minutes to discuss their work and complete the activity with a partner.

Action and Expression: Internalise Executive Functions. Activate or supply background knowledge by demonstrating how to use a ruler to draw a line segment for a given length. *Supports accessibility for: Memory; Conceptual processing*

Anticipated Misconceptions

When determining frequencies of data values, students might lose track of their counting. Suggest that they use tally marks to keep track of the number of occurrences for each bin.

When drawing the histogram, students might mistakenly use bar graphs as a reference and leave spaces between the bars. Ask them to look at the bars in other histograms they have seen so far and to think about what the gaps might mean considering that the bars are built on a number line.

Student Task Statement

An earthworm farmer set up several containers of a certain species of earthworms so that he could learn about their lengths. The lengths of the earthworms provide information about their ages. The farmer measured the lengths of 25 earthworms in one of the containers. Each length was measured in millimetres.

1. Using a ruler, draw a line segment for each length:
   - 20 millimetres
   - 40 millimetres
   - 60 millimetres
   - 80 millimetres
   - 100 millimetres

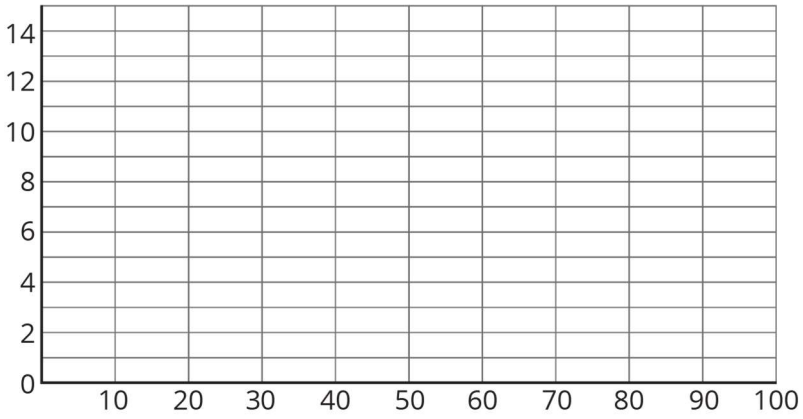
2. Here are the lengths, in millimetres, of the 25 earthworms.

6	11	18	19	20	23	23	25
25	26	27	27	28	29	32	33
41	42	48	52	54	59	60	77
93							

   Complete the table for the lengths of the 25 earthworms.

length	frequency
0 millimetres to less than 20 millimetres	
20 millimetres to less than 40 millimetres	
40 millimetres to less than 60 millimetres	
60 millimetres to less than 80 millimetres	
80 millimetres to less than 100 millimetres	

3. Use the grid and the information in the table to draw a histogram for the worm length data. Be sure to label the axes of your histogram.



4. Based on the histogram, what is a typical length for these 25 earthworms? Explain how you know.

5. Write 1–2 sentences to describe the spread of the data. Do most of the worms have a length that is close to your estimate of a typical length, or are they very different in length?

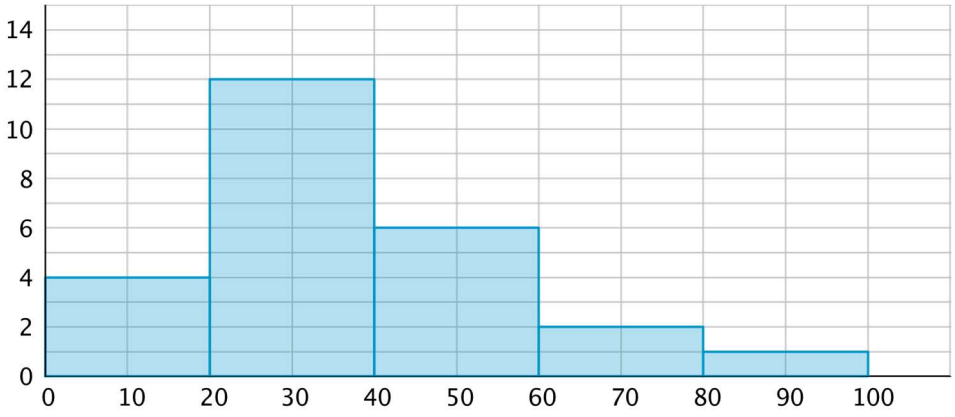
Student Response

1. Drawings should show segments of 20 mm, 40 mm, 60 mm, 80 mm, and 100 mm.

2.

length	frequency
0 millimetres to less than 20 millimetres	4
20 millimetres to less than 40 millimetres	12
40 millimetres to less than 60 millimetres	6
60 millimetres to less than 80 millimetres	2
80 millimetres to less than 100 millimetres	1

3.

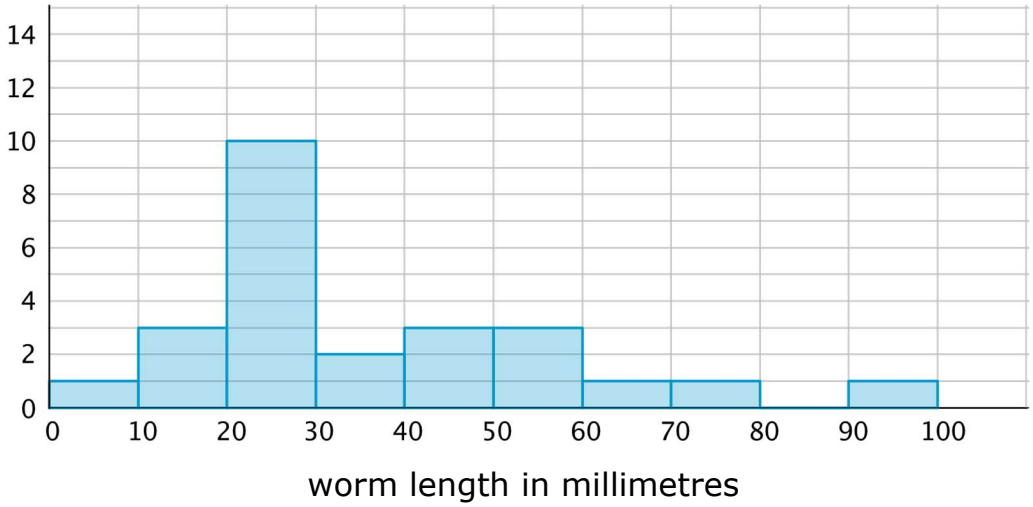


4. Answers vary. Sample responses: Between 20 and 40 mm; about 30 mm; about 35 mm.

5. Answers vary. Sample response: Most of the worms appear to be shorter than my estimate of the typical value, but they are all pretty close to it. There are worms that are longer than my estimate, but not as many.

Are You Ready for More?

Here is another histogram for the earthworm measurement data. In this histogram, the measurements are in different groupings.



1. Based on this histogram, what is your estimate of a typical length for the 25 earthworms?

2. Compare this histogram with the one you drew. How are the distributions of data summarised in the two histograms the same? How are they different?

3. Compare your estimates of a typical earthworm length for the two histograms. Did you reach different conclusions about a typical earthworm length from the two histograms?

Student Response

Answers vary. Sample responses:

1. About 30mm.

2. Both histograms have clusters of data around 20 and 30mm. The one I drew has more data values between 30 and 60mm, and this histogram has data that is more spread out.

3. The typical earthworm lengths are both around 30mm, but the typical length from this histogram is a little smaller.

Activity Synthesis

Ask one or two students to display their completed histograms for all to see and briefly describe the overall distribution. Then, select a few other previously identified students to share their responses and explanations for the last two questions.

Focus the discussion on how identifying centre and spread using a histogram is different than doing so using a dot plot. Discuss:

- "In a histogram, are we able to see clusters of values in the distribution?"
- "Can we see the largest and smallest values? Can we tell the overall spread?"
- "How do we identify the centre of a distribution?"

From the various estimates that students give for a typical earthworm length (and from earlier exercises), students should begin to see that identifying a typical value of a distribution is not a straightforward or precise process so far. Explain that in upcoming lessons they will explore how to describe a typical value and characterise a distribution more systematically.

Representing, Conversing: Compare and Connect. As the selected students show and describe their histograms, invite pairs to discuss: "What is the same and what is different?" about their own histograms and reasoning. This will help students better understand there are multiple valid ways for describing centre and spread.
Design Principle(s): Maximise meta-awareness; Support sense-making

7.3 Tall and Taller Players

10 minutes

Now that students have some experience drawing and interpreting histograms, they use histograms to compare distributions of two populations. In a previous activity, students compared the two dot plots of students in a keyboarding class—one for the typing speeds at the beginning of the course and the other showing the speeds at the end of the course. In this activity, they recognise that we can compare distributions displayed in histograms in a similar way—by studying shapes, centres, and spreads.

Instructional Routines

- Stronger and Clearer Each Time
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 4–5 minutes of quiet work time and 1–2 minutes to share their responses with a partner.

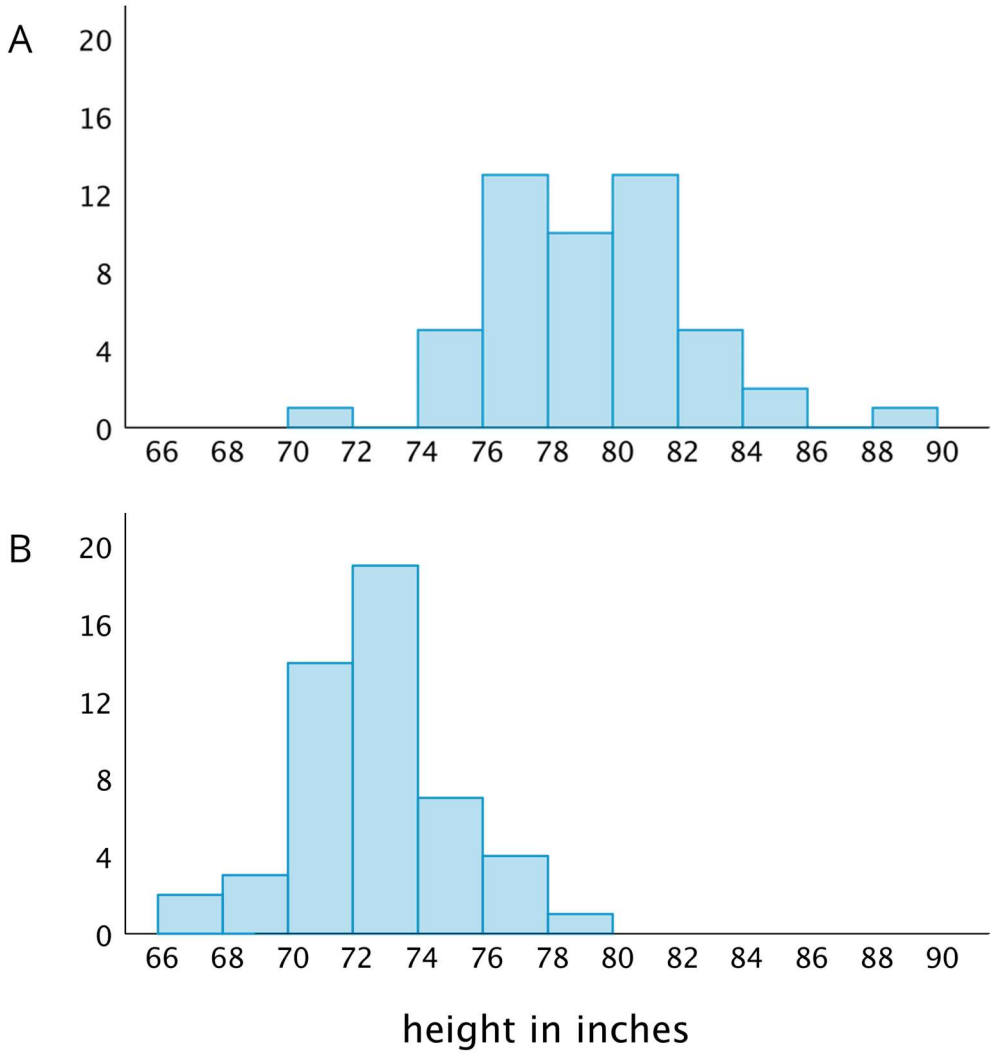
Engagement: Develop Effort and Persistence. Encourage and support opportunities for peer interactions. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their reasoning. For example, "Histogram ___ represents the heights of _____ players because . . .", and "I agree/disagree because . . ."
Supports accessibility for: Language; Social-emotional skills

Student Task Statement

Professional basketball players tend to be taller than professional baseball players.

Here are two histograms that show height distributions of 50 male professional baseball players and 50 male professional basketball players.

1. Decide which histogram shows the heights of baseball players and which shows the heights of basketball players. Be prepared to explain your reasoning.



2. Write 2–3 sentences that describe the distribution of the heights of the basketball players. Comment on the centre and spread of the data.

3. Write 2–3 sentences that describe the distribution of the heights of the baseball players. Comment on the centre and spread of the data.

Student Response

1. Histogram A shows the heights of professional basketball players. Sample explanation: The cluster of bars are located higher on the number line compared to those in histogram B, which means they represent taller players.

2. Answers vary. Sample description: The basketball players' data distribution is centred at around 80–81 inches. Except for two players at the shortest and tallest ends, most players are within 6–7 inches of the centre.

3. Answers vary. Sample description: The baseball players' data distribution is centred at around 72 inches. The spread is also similar to the basketball players' data. All players are within 6 inches of the centre of the data.

Activity Synthesis

Select a few students to share their descriptions about basketball players and baseball players. After each student shares, ask others if they agree with the descriptions and, if not, how they might revise or elaborate on them. In general, students should recognise that the distributions of the two groups of athletes are quite different and be able to describe how they are different.

Highlight the fact that students are using approximations of centre and different adjectives to characterise a distribution or a typical height, and that, as a result, there are variations in our descriptions. In some situations, these variations might make it challenging to compare groups more precisely. We will study specific ways to measure centre and spread in upcoming lessons.

Representing, Writing: Stronger and Clearer Each Time. Use this routine to give students a structured opportunity to revise and refine their written responses to the questions about describing the distribution of heights of basketball and baseball players. Ask students to meet with 2–3 partners, to share and get feedback on their responses. Provide listeners with prompts for feedback that will clarify the descriptions about the distributions shown in the histograms (e.g., "What details about the histograms are important are important to describe?"). Finally, give students time to revise their initial written responses to reflect input they received. This will help students write descriptions of the centre and spread of distributions displayed in histograms using precise language.
Design Principle(s): Optimise output (for explanation); Maximise meta-awareness

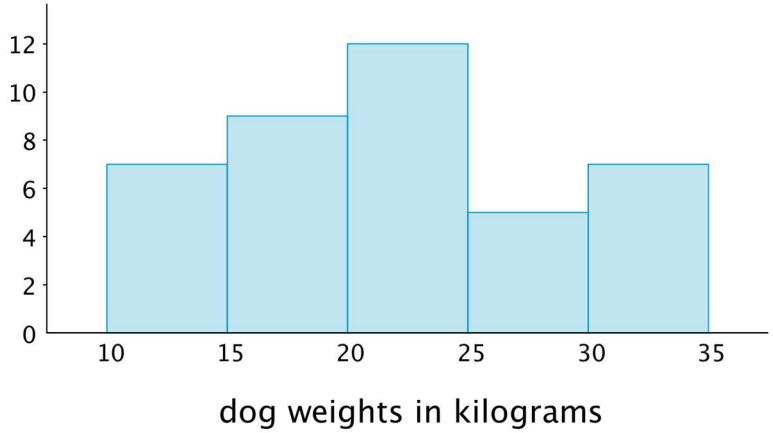
Lesson Synthesis

In this lesson, we learn how to draw a histogram and how to use it to describe characteristics of a data set.

- "What are some decisions we should think about and make before drawing a histogram?"
- "Does the width of each bar have to represent a distance of 5 units, or can it represent other number of units?"
- "What does the horizontal axis of a histogram tell us? What about the vertical axis?"
- "How do we know how tall to make each bar?"

Once we have a histogram drawn, we can use it to answer some questions about a data set.

- "How would you describe a typical weight for this group of dogs?"
- "What can we say about the spread of the dog weights based on this histogram?"

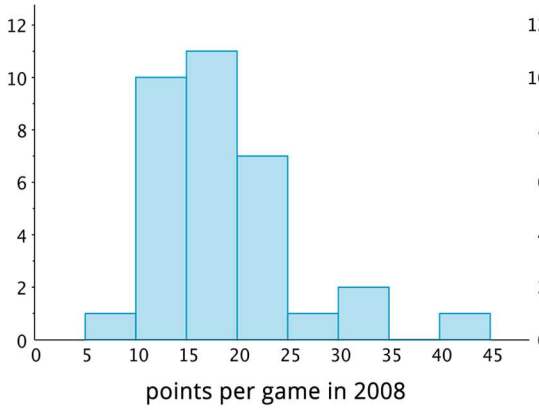


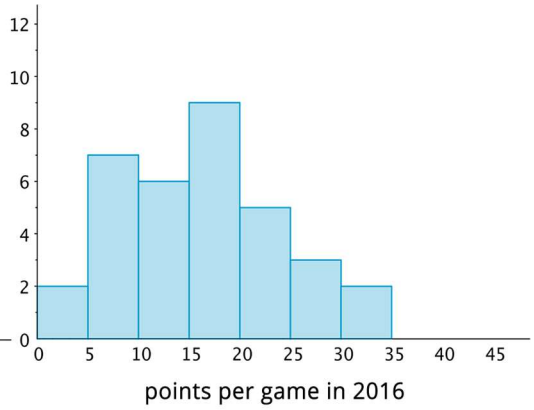
7.4 A Tale of Two Seasons

Cool Down: 5 minutes

Student Task Statement

The two histograms show the points scored per game by a basketball player in 2008 and 2016.





1. What is a typical number of points per game scored by this player in 2008? What about in 2016? Explain your reasoning.

2. Write 2–3 sentences that describe the spreads of the two distributions, including what spreads might tell us in this context.

Student Response

1. Answers vary. Sample response: In both seasons, the player typically scored around 15 to 20 points in a game. In each histogram the typical score should be around the centre of the spread of data.

2. Answers vary. Sample response: The spread of data was more narrow in 2016 than in 2008. Although the player was more consistent in 2016, she had a really great game in 2008 that changed the spread of the scores.

Student Lesson Summary

Here are the weights, in kilograms, of 30 dogs.

10	11	12	12	13	15	16	16
17	18	18	19	20	20	20	21
22	22	22	23	24	24	26	26
28	30	32	32	34	34		

Before we draw a histogram, let's consider a couple of questions.

- What are the smallest and largest values in our data set? This gives us an idea of the distance on the number line that our histogram will cover. In this case, the minimum is 10 and the maximum is 34, so our number line needs to extend from 10 to 35 at the very least.

  (Remember the convention we use to mark off the number line for a histogram: we include the left boundary of a bar but exclude the right boundary. If 34 is the right boundary of the last bar, it won't be included in that bar, so the number line needs to go a little greater than the maximum value.)

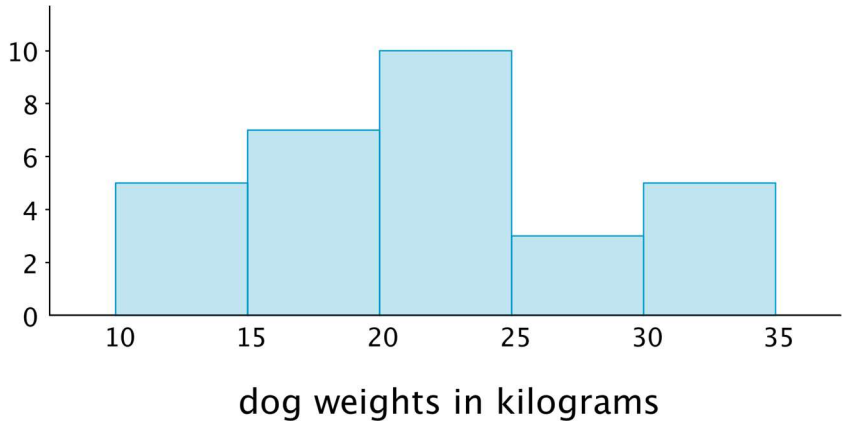
- What group size or bin size seems reasonable here? We could organise the weights into bins of 2 kilograms (10, 12, 14, . . .), 5 kilograms, (10, 15, 20, 25, . . .), 10 kilograms (10, 20, 30, . . .), or any other size. The smaller the bins, the more bars we will have, and vice versa.

Let's use bins of 5 kilograms for the dog weights. The boundaries of our bins will be: 10, 15, 20, 25, 30, 35. We stop at 35 because it is greater than the maximum.

Next, we find the frequency for the values in each group. It is helpful to organise the values in a table.

weights in kilograms	frequency
10 to less than 15	5
15 to less than 20	7
20 to less than 25	10
25 to less than 30	3
30 to less than 35	5

Now we can draw the histogram.

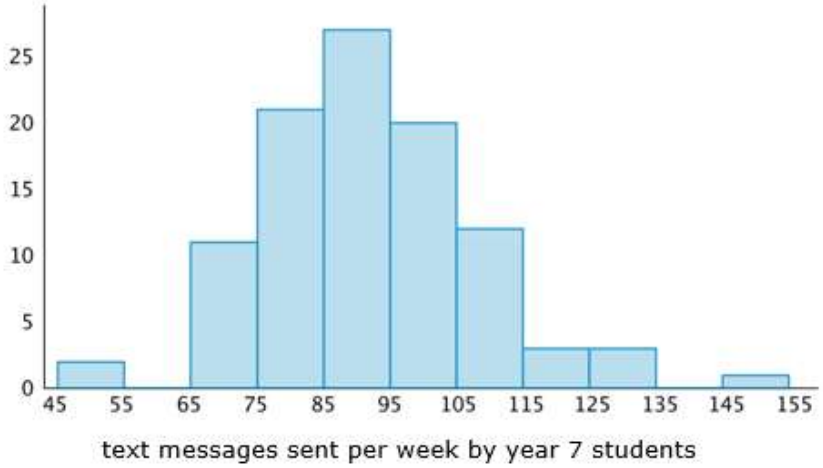


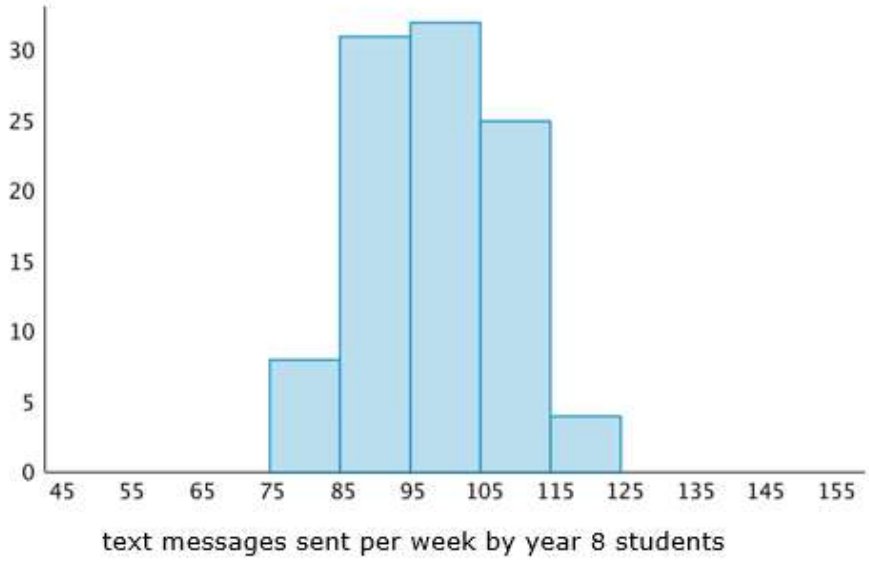
The histogram allows us to learn more about the dog weight distribution and describe its centre and spread.

Lesson 7 Practice Problems

Problem 1 Statement

These two histograms show the number of text messages sent in one week by two groups of 100 students. The first histogram summarises data from year 7 students. The second histogram summarises data from year 8 students.





a. Do the two data sets have approximately the same centre? If so, explain where the centre is located. If not, which one has the greater centre?

b. Which data set has greater spread? Explain your reasoning.

c. Overall, which group of students — year 7 or year 8 — sent more text messages?

Solution

a. Yes, both are centred around 100 text messages.

b. The year 7 students have a wider spread, much of their data is outside the 75–125 range, while no year 8 students are outside this range.

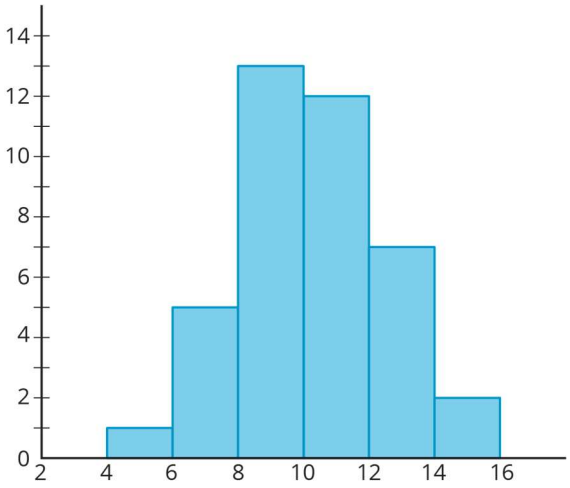
c. Neither. Both send about the same number of text messages, because the centre of the two data sets is close.

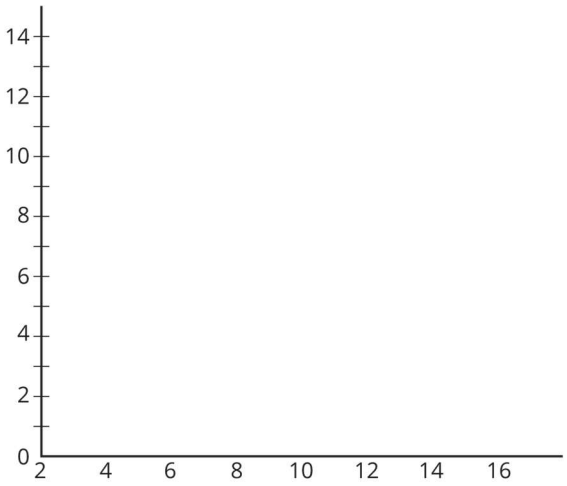
Problem 2 Statement

Forty year 7 students ran 1 mile. Here is a histogram that summarises their times, in minutes. The centre of the distribution is approximately 10 minutes.

On the blank axes, draw a second histogram that has:

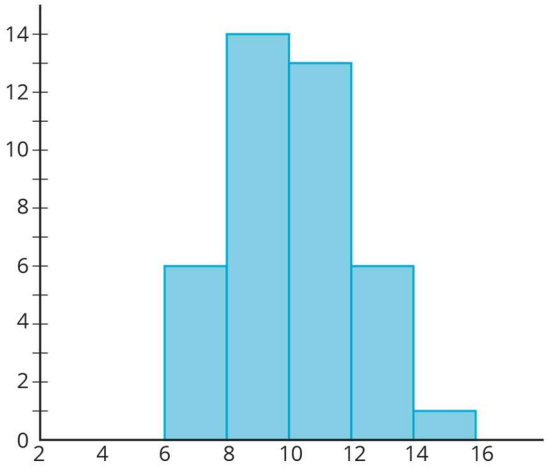
- a distribution of times for a different group of 40 year 7 students.
- a centre at 10 minutes.
- less variability than the distribution shown in the first histogram.





Solution

Responses vary. Sample response:



Problem 3 Statement

Jada has d 10p coins. She has more than 30 pence but less than a pound.

a. Write two inequalities that represent how many 10p coins Jada has.

b. Can d be 10?

c. How many possible solutions make both inequalities true? If possible, describe or list the solutions.

Solution

a. $d > 3$ and $d < 10$ (or $d \geq 4$ and $d \leq 9$)

b. No, this does not make $d < 10$ true. 10 10p coins is a pound, which is too much.

c. 6 possible solutions: 4, 5, 6, 7, 8, and 9

Problem 4 Statement

Order these numbers from greatest to least: -4, $\frac{1}{4}$, 0, 4, $-3\frac{1}{2}$, $\frac{7}{4}$, $-\frac{5}{4}$

Solution

4, $\frac{7}{4}$, $\frac{1}{4}$, 0, $-\frac{5}{4}$, $-3\frac{1}{2}$, -4