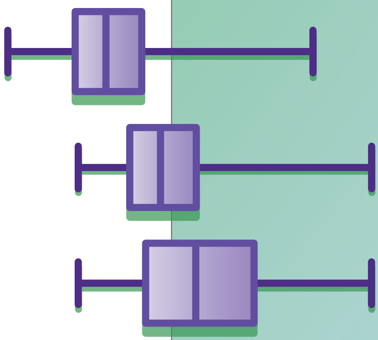
Unit 8

Probability and Sampling

Have you ever wondered what the chances of something happening are? Is the game you are playing a fair one? Do middle schoolers have more homework per night than high schoolers? These are the types of questions we can answer using statistics and probability, which help us make better predictions and decisions.



Essential Questions

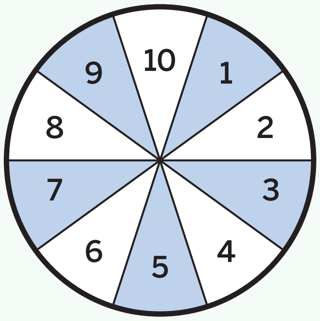
- How can we determine how likely an event is to happen?
- How can we simulate events in the world to make predictions?
- When is a sample representative of a population? How might this affect our analysis?

Summary | Lesson 1

How likely is an **event** to occur? One way to learn more is to look at the outcomes of **experiments**. An **outcome** is any of the possible results that can happen when you perform an experiment at **random**.

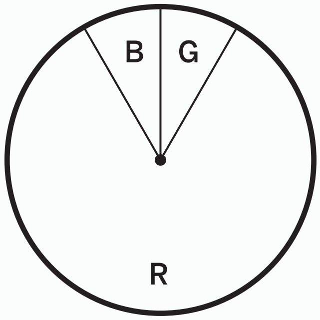
For example, when you spin a spinner with 10 equal sections labeled 1–10, one possible outcome is that the spinner will land on the number 5. You can describe the likelihood of events using these phrases: impossible, unlikely, equally likely as not, likely, or certain.

Event	How likely is it?
Lands on 5	Unlikely
Lands on an even number	Equally likely as not
Lands on a number that is *not* 5	Likely
Lands on a whole number	Certain
Lands on 12	Impossible



Try This

Write the letter that matches the likelihood of getting each letter on one spin.



a. Impossible ________ R

b. Unlikely ________ B

c. Equally likely as not ________ Y

d. Likely ________ R, B, or G

e. Certain ________ G

Summary | Lesson 2

The **probability** of an event is a number that represents how likely the event is to occur. One way to calculate the probability is to look at all of the possible outcomes for an experiment, which is known as the **sample space**.

When all of the outcomes are equally likely, the probability of an event is a ratio.

$$\frac{\textit{number of favorable outcomes}}{\textit{total possible number of outcomes}}$$

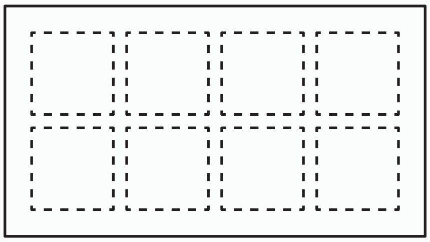
Probabilities are numbers between 0 and 1 written as fractions, decimals, or *percentages*. A probability of 1 means the event will always happen. A probability of 0 means the event will never happen.

Here are some examples of events and their probabilities.

Example	Probability
Picking a green marble out of a bag that contains only red and yellow marbles	0
Rolling a 1 on a number cube	$\frac{1}{6}$ (or equivalent)
Flipping a coin and it landing heads up	50% (or equivalent)
Picking a yellow marble from a bag of 10 marbles, where 8 of the marbles are yellow	0.8 (or equivalent)
Picking a green marble in a bag that only contains green marbles	1

Try This

A mystery box contains a secret set of letter tiles.

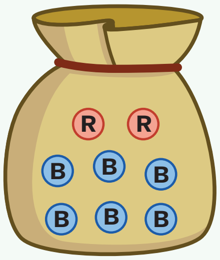


Label the secret letter tiles using these three probability clues:

- The probability of picking a P is $\frac{1}{4}$.
- The probability of picking a B is 0.
- The probability of picking a G is equal to the probability of picking an R.

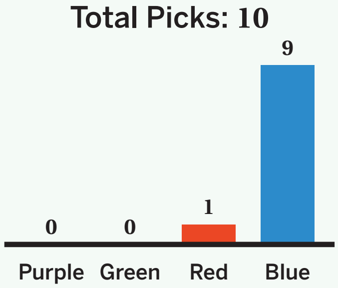
Summary | Lesson 3

In situations where you don't know the sample space, you can use data from experiments and proportional reasoning to predict what the sample space is. Repeating an experiment more times can help make your prediction more accurate.



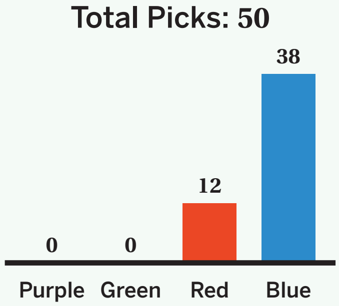
For example, here are the results from picking a marble out of a mystery bag 10 times. The bag has 8 marbles in it.

Only 1 out of the 10 marbles was red, so the *constant of proportionality* is 0.1. Multiplying 0.1 times the number of marbles in the bag (8), may lead someone to predict that there is only $0.1 \cdot 8 = 0.8$ or 1 red marble in the bag.



After 50 picks, the constant of proportionality (0.24) times the number of marbles in the bag is $0.24 \cdot 8 = 1.92$, which is close to 2.

This is a more accurate prediction of the number of red marbles in the bag.



Try This

There are 6 blocks in a bag. The table shows results from 100 picks.

Block Color	Number of Picks
Purple	32
Red	68

Based on these results, how many of the 6 blocks are likely to be red?

Explain your thinking.

Summary | Lesson 4

Graphing the outcomes of a repeated experiment can help you make sense of how likely an event is to occur. The fraction of experiments that result in a certain outcome is called the **relative frequency** of that outcome.

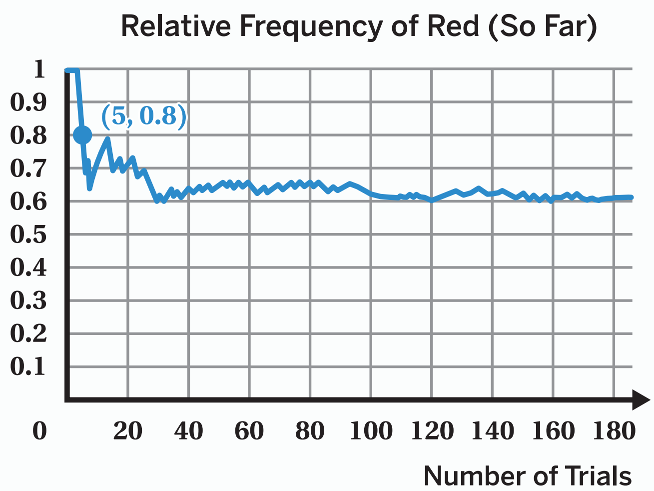
When an experiment is only repeated a few times, the results may surprise you because they may be far from the probability you expected. As the number of experiments increases, the relative frequency of each outcome gets close to its probability.

For example, the probability of a flipped coin landing heads up is $\frac{1}{2}$ or 0.5. This means that if the coin is flipped many times, we expect it to land heads up about half of the time.

If the coin landed tails up three times in a row, this does not mean that the next flip is more likely to be heads up. The chances of landing heads up are the same on each flip regardless of the previous outcomes.

Try This

This graph shows the results of 200 spins of a color spinner.



a There is a point on the graph at $(5, 0.8)$. What does this point represent?

b Based on these results, what is the probability that the spinner lands on red?

Explain your thinking.

Summary | Lesson 5

Repeated experiments can help you decide if an object is fair.

If an experiment is repeated only a few times, the results may not be what you expect, even if the object is fair. The more times you repeat the experiment (i.e., hundreds or thousands of times), the closer the relative frequency should get to the probability. This allows you to make a better decision about whether the object is fair.

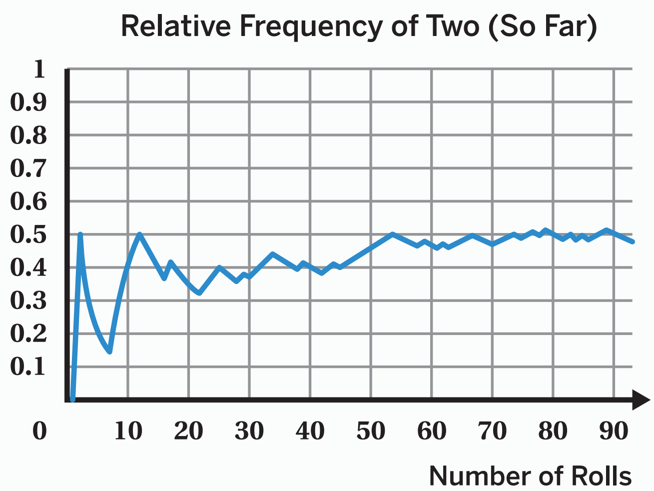
For example, here is a fair coin. The probability of this coin landing heads up is $\frac{1}{2}$.

If you flip the coin only 3 times, it may land heads up all 3 times. You may think the coin is unfair, but continuing to repeat the experiment can change your perspective.

If you flipped the coin 1,000 times, it would land heads up about half of the time because the sample space of this event is "heads" and "tails."

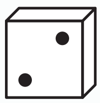
Try This

This graph shows the results of 100 rolls of a number cube.



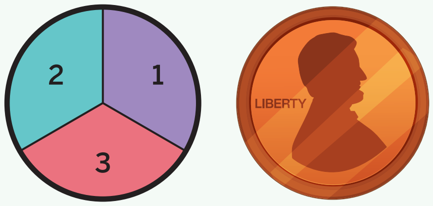
a Based on these results, what is the probability of rolling a 2 with this number cube?

b Describe or draw what the rest of this number cube could look like.



Summary | Lesson 6

There are several different ways to make sense of **compound events**, or events that involve multiple steps.



Here is one example: Let's spin a spinner and flip a fair coin. There are 6 outcomes in the *sample space* of this multistep event, which you can see in a list, a table, and a **tree diagram**.

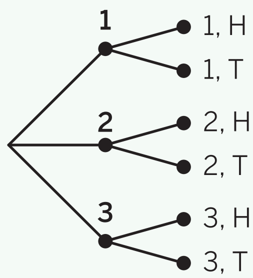
List

1, Heads　　1, Tails
2, Heads　　2, Tails
3, Heads　　3, Tails

Table

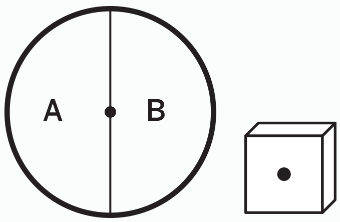
	H	T
1	1, H	1, T
2	2, H	2, T
3	3, H	3, T

Tree Diagram



Try This

Juana is playing her favorite board game, Dice & Doors, which uses a spinner and a standard number cube.



a Make a tree, table, or list to represent the sample space.

b What is the probability of Juana getting an A and an even number?

Summary | Lesson 7

Simulations are experiments that are used to estimate the probability of a real-world event. They are especially useful for estimating the probabilities of compound events, such as determining the probability that it will rain at least once over a three day period. In order to design a good simulation, first determine the probability of the individual events occuring.

For example, you could use a coin, number cube, or spinner to simulate a 50% probability of rain.

Flipping a Coin	Rolling a Number Cube	Using a Spinner
Landing heads up $\left(\frac{1}{2} = 50\%\right)$	Rolling an even number $\left(\frac{3}{6} = 50\%\right)$	Spinning a raindrop $\left(\frac{5}{10} = 50\%\right)$

To simulate the probability of rain over three days where each day has a 50% chance of rain, you can use three coins, number cubes, or spinners and repeat the experiment many times.

Try This

Brianna loves monkeys, seals, and elephants, so she designed a simulation with spinners to help her estimate the probability of these 3 animals being awake when she visits the zoo.

This table shows the results of 300 experiments with the spinners.

Estimate the probability that at least 2 of Brianna's favorite animals will be awake when she visits the zoo.

Experiments With . . .	Count	Percentage (%)
No animals awake	12	4
1 animal awake	171	57
2 animals awake	105	35
3 animals awake	12	4

Summary | Lesson 8

You can use simulations to estimate the probability of an event. Simulations are especially useful for estimating the probability of compound events. The more simulations you perform, the closer the probability in the relative frequency of each outcome should be to the probability.

Many professionals such as scientists, computer programmers, financial analysts, and sports analysts create simulations to model the outcomes of complicated real-world events. Using computer software, they are able to perform thousands of simulations to answer questions about everyday situations. Similarly, you can create a simulation using items such as spinners or bags to determine the probability of a real-world question you may want to answer.

Try This

Mayra is a basketball player who makes about 50% of her free throws.

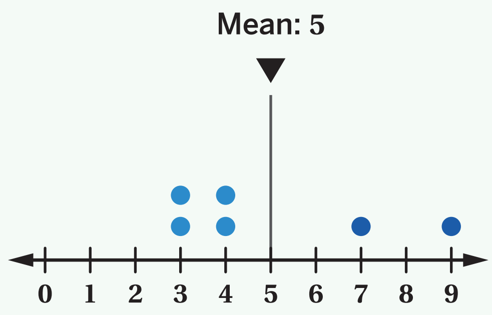
Design a simulation that you could use to answer the question: *What is the probability that Mayra will make exactly 3 out of 5 free throws in her next basketball game?*

Describe your simulation precisely enough that someone else could perform it.

Summary | Lesson 9

When comparing two sets of data, you can compare their centers, shapes, and spreads. Data sets can have the same center but very different shapes and spreads.

The *mean* is a measure of center. One way to measure the spread of data is to use the distances between each value in the data set and the mean. The average of those distances is called the *mean absolute deviation (MAD)*. In this example, the MAD is 2 units.



Data Point	3	3	4	4	7	9
Distance from Mean	2	2	1	1	2	4

$$\frac{2+2+1+1+2+4}{6} = 2$$

MAD: 2

Try This

Prisha eats cheese and crackers as a snack every day.

Here is some data Prisha collected about the number of crackers in her snack bags.

Number of crackers per bag:
$[15, 19, 27, 23, 21]$

a Calculate the mean.

b Calculate the MAD.

Summary | Lesson 10

To answer a question about a population, it is sometimes not realistic to collect data from the entire **population**. Instead, you can collect data from a **sample** of the population.

- A *population* is a set of people or things that we want to study.
- A *sample* is part of a population.

The sample you choose should be large enough to be able to draw conclusions about the population.

Here are some examples of populations and samples.

Population	Sample
All of the people who watch basketball	The people at a basketball game
All 7th grade students in your school	The 7th graders in your school who are in a band
All oranges grown in the U.S.	The oranges in your local grocery store

Try This

Alisha wants to survey the other 7th graders at her school to learn how many minutes they spend on their cell phones each day.

a What is the population for Alisha's question?

b What is a sample Alisha could use to help answer this question?

Summary | Lesson 11

Samples are useful when a population is too large to survey or measure. Depending on the strategy you use to sample, your sample might or might not be **representative** of the population. Some samples are not good representations of the population.

A representative sample has a distribution that closely resembles the distribution of the population. Representative samples are useful for making predictions about the whole population.

For example, if you were curious about all middle school students' favorite sport to play, the population would be all middle schoolers. A representative sample of this population might be randomly selecting 5 students from each class or 15 students from each grade to ask. A sample that is *not* representative of this population would be asking students in the tennis club because their responses might lead someone to believe that tennis is the favorite sport among all middle schoolers.

Try This

Match each headline with the sampling method that most likely produced it.

Headline	Sampling Method
a. One Quarter of Working Americans Spend Time Working From Home!	________ Ask all 100 employees at one technology company.
b. Most Americans Spend Time Working From Home!	________ Ask all the employees at 100 random grocery stores.
c. Almost Nobody Works From Home!	________ Call random phone numbers until you ask 100 people.

Summary | Lesson 12

How can we use data from a sample to make claims about a population? One way is to use *proportional* reasoning.

Let's say someone is wondering how many students at their school might vote for a candidate for student council. It would be challenging and time consuming to ask all 500 students at the school, so they collect a sample of 25 students. It is important to gather the sample in a way that makes sure the sample is likely to be representative, like asking one student from each homeroom or asking 25 students at random.

If 10 out of the 25 students in the sample said they would vote for this candidate, there are several strategies for making a prediction about the population.

Strategy A

10 out of 25 is equal to $\frac{10}{25}$ or 40% of the sample.

40% of the population (500 students) would be $0.4 \cdot 500 = 200$ students.

Strategy B

The population is $\frac{500}{25} = 20$ times as large as the sample, so multiply the number of votes by 20 to determine the number of students in the population who would vote for them.

Votes	Total Students
10	25
200	500

Try This

Cameron bought a bag of White Flower Seed Mix and is curious about how many flowers of each type there are. They planted 25 seeds, and these were the results.

a Complete the table with the percentage of each flower type.

Flower Type	Count	Percentage (%)
Daisy	14	
White Zinnia	5	
Aster	6	
Total	25	

b Estimate how many of the 600 seeds in the bag will be asters.

Show your calculations.

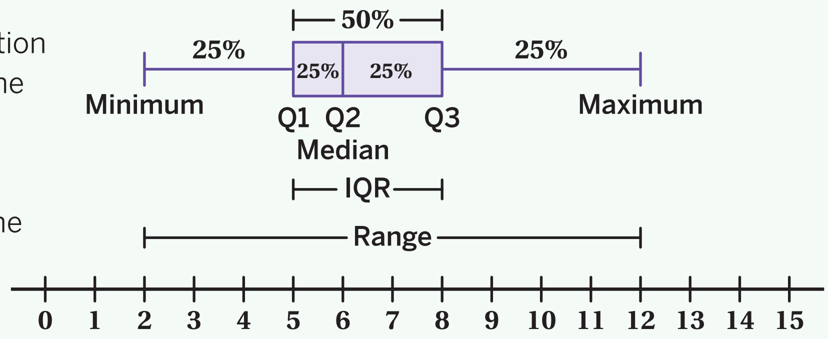
c What could you do to be more confident in your estimate?

Summary | Lesson 13

Box plots are one way of displaying data that can be helpful for comparing samples or populations.

They show the data split up into quartiles, with each section representing about 25% of the data in the data set.

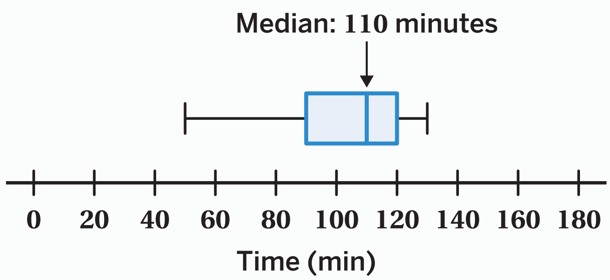
The measure of center on a box plot is the *median* and the measure of spread is the *interquartile range (IQR)*, or the distance between the middle 50% of the data.



One way to make sense of data is to collect and compare multiple samples from the same population. In general, when samples have similar medians or interquartile ranges, predictions about the population are more likely to be accurate. If the medians for two samples are very different, you may be less confident that your predictions about the population are accurate.

Try This

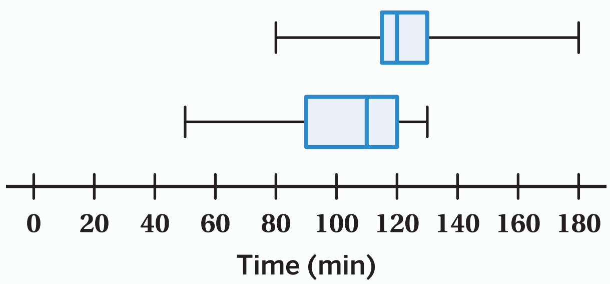
Sai asked 10 random people at their school how many minutes each person exercised every week. Then Sai created a box plot of the answers.



a What is the interquartile range (IQR)?

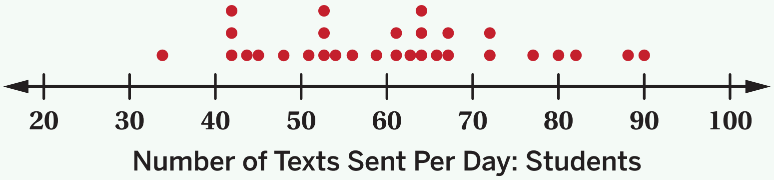
b Sai asked another 10 people. Here are both sets of results.

Estimate the population median.

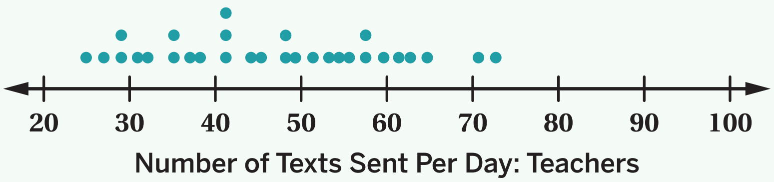


In general, it is easier to compare two individuals or objects than it is to compare two populations. For example, you can answer the question, "Which 7th grader is taller?" by measuring the heights of two 7th grade students and comparing them directly. However, to answer the question, "Do middle school students send more texts per day than their teachers?" you need to collect samples and analyze the measures of center and variability.

Mean: 60.5 **MAD: 11.4**



Mean: 46.3 **MAD: 11**



Here are the results of a random sample of 30 students and 30 teachers surveyed about the number of texts they send per day. To decide if the data sets are very different from each other, we can calculate the difference in their means and compare it to the larger MAD. The difference is $60.5 - 46.3 = 14.2$, which is about 1.25 times the MAD of 11.4. When the difference is more than 1 times the larger MAD, the data sets are very different. This suggests that students do send more texts than teachers.

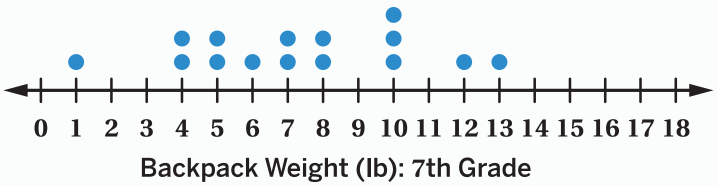
Try This

Nathan compared the weight of backpacks from students in 7th grade and 8th grade. After collecting data from a random sample of 15 students in each grade, he claimed that 7th graders' backpacks are much heavier.

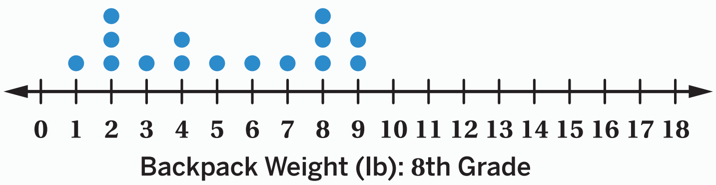
Do you agree with Nathan's claim?

Use the data to support your thinking.

Mean: 7.3 **MAD: 2.8**



Mean: 5.2 **MAD: 2.8**



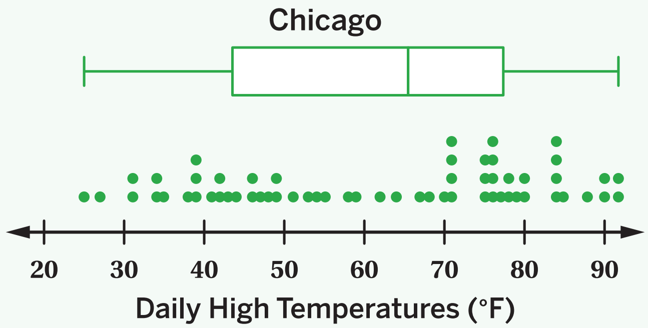
Summary | Lesson 15

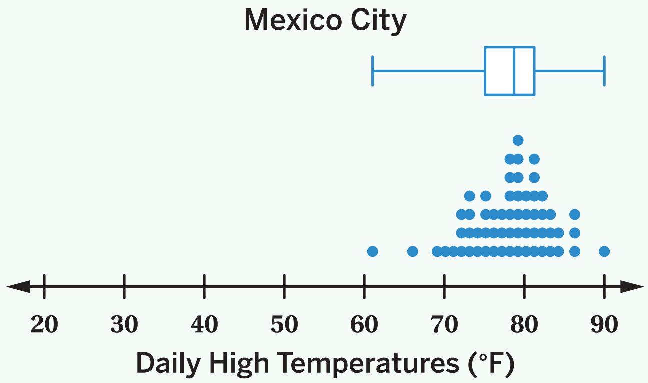
Statistics and sampling, along with visual representations of data, can help you to make sense of real-world topics. The more we understand the world around us, the more we can take action to improve it.

Let's say you want to compare the weather of Chicago and Mexico City. Here is a random sample of the daily high temperature for 60 days in 2023.

The box plots and dot plots let us see patterns in the data. We can also study measures of center and spread.

The mean and median temperatures are higher in Mexico City than in Chicago, so we might say it is generally hotter there. The IQR and MAD add more detail to our comparison. Chicago has greater measures of spread, so it has more variability in its temperature over the year.





	Mean	Median	IQR	MAD
Chicago	61.5	65.5	34	17.4
Mexico City	77.9	78.5	6	3.9

Try This

Here is a sample of asthma rates in adults from Emmanuel's city.

a One of the asthma rates is 105. What does this number mean?

b Why might someone be interested in this data?

c Based on this data, what is the average asthma rate in adults from Emmanuel's city?

Number of Adults Who Have Asthma (per 1,000)
112
105
93
129
127
125
93

Try This | Answer Key

Lesson 1

a. Impossible ___d___ R

b. Unlikely ___b___ B

c. Equally likely as not ___a___ Y

d. Likely ___e___ R, B, or G

e. Certain ___b___ G

Lesson 2

Responses vary.

P	P	G	G
R	R	R	G

Lesson 3

4 blocks. *Explanations vary.* $\frac{68}{100}$ is 68% and 68% of 6 blocks is about 4 blocks.

Lesson 4

a After 5 spins, the fraction of red spins was 0.8 or $\frac{4}{5}$. In the first 5 spins, 4 were red and 1 was not.

b 0.6 or 60%. *Explanations vary*. As you spin more times, the fraction of spins gets closer to the true probability. On the graph, the fraction of spins is getting closer to 0.6 (or a 60% chance of landing on red).

Lesson 5

a $\frac{1}{2}$ (or equivalent)

b *Responses vary*. Half of the sides are 2s. The other half of the sides are not 2s.

Lesson 6

a *Responses vary.*

	1	2	3	4	5	6
A	A, 1	A, 2	A, 3	A, 4	A, 5	A, 6
B	B, 1	B, 2	B, 3	B, 4	B, 5	B, 6

b $\frac{3}{12}$ (or equivalent)

Lesson 7

39%

Caregiver Note: Here is a strategy to estimate the probability. Add the number of experiments that resulted in 2 or 3 animals being awake. Then divide the sum by the total number of experiments.

$105 + 12 = 117$

$\frac{117}{300} = 0.39 = 39\%$

Lesson 8

Responses vary. Create 5 identical spinners to simulate 5 free throws in a game. Give each spinner two sections: one that says "Make" and one that says "Miss." Spin each spinner and write down how many free throws Mayra makes in the simulation. Repeat this many times and keep track of the percent of the experiments where "Make" comes up 3 out of 5 times.

Lesson 9

a 21 crackers

Caregiver Note: Here is a strategy to find the mean. Divide the sum of the values by the number of values. $\frac{15+19+27+23+21}{5} = 21$

b 3.2 crackers

Caregiver Note: Here is a strategy to find the MAD. First, find the distances between each value in the data set and the mean. Then, calculate the mean of those distances.
$\frac{6+2+6+2+0}{5} = 3.2$

Lesson 10

a **The population is all the 7th graders at Alisha's school.**

b ***Responses vary.* A sample could be all the 7th graders in Alisha's math class.**

Lesson 11

	Headline		Sampling Method
a.	One Quarter of Working Americans Spend Time Working From Home!	b	Ask all 100 employees at one technology company.
b.	Most Americans Spend Time Working From Home!	c	Ask all the employees at 100 random grocery stores.
c.	Almost Nobody Works From Home!	a	Call random phone numbers until you ask 100 people.

Lesson 12

a

Flower Type	Count	Percentage (%)
Daisy	14	**56**
White Zinnia	5	**20**
Aster	6	**24**
Total	25	**100**

b **144**

Caregiver Note: Here is one strategy. First, find the percentage of the sample that is asters by dividing the count by the sample size. Then find that percent of the population.

$\frac{6}{25} = 0.24 = 24\%$

$0.24 \cdot 600 = 144$

c ***Responses vary.***

- **You could plant more samples of 25 seeds and see if the results are similar between samples.**
- **You could plant a larger sample, like 100 seeds, because a larger sample might be more accurate.**

Lesson 13

a 30 minutes. Responses between 25 and 35 are considered correct.

b *Responses vary*. The population median is about 115 minutes because the median of both samples is close to 2 hours.

Lesson 14

Responses vary.

- Nathan's claim is incorrect because the difference between the means is $7.3 - 5.2 = 2.1$, which is less than the MAD.
- Nathan's claim is correct because the mean of the 7th graders' data is 2.1 pounds heavier than the mean of the 8th graders' data.

Lesson 15

a This number means that 105 out of every $1{,}000$ adults in one community in Emmanuel's city have asthma.

b *Responses vary.*

- They might want to know how much asthma medication is needed in the city.
- They could be curious about how this affects them and their children.

c 112 adults per $1{,}000$