ACTIVITY 2

Signal and Noise

DATA ANALYSIS

ACTIVITY 2

Signal and Noise

ACTIVITY SUMMARY

Graphs from real-world studies provide students with the opportunity to evaluate the accuracy of a claim in an article. Students look for the signal in the noise as they investigate indoor vs. outdoor air quality during a wildfire.

ACTIVITY TYPE
DATA ANALYSIS

NUMBER OF
40–50 MINUTE
CLASS PERIODS
1-2

KEY CONCEPTS & PROCESS SKILLS

1 When there is scientific uncertainty in data, probabilistic reasoning is a method for determining the likelihood of different outcomes on which to base a decision.

2 Probabilistic reasoning can be used to identify meaningful patterns in data (*signal*) about a phenomenon being investigated. Variations in the data (*noise*) can increase scientific uncertainty by distorting or hiding the signal.

NEXT GENERATION SCIENCE STANDARDS (NGSS) CONNECTION:
Analyze data using tools, technologies, and/or models (e.g., computational, mathematical) in order to make valid and reliable scientific claims or determine an optimal design solution. (*Science and Engineering Practice: Analyzing and Interpreting Data*)

CONCEPTUAL TOOLS

VOCABULARY DEVELOPMENT

noise
information that hides, distracts from, or falsely resembles the meaningful information that is being investigated

signal
meaningful information about the phenomenon that is being investigated

TEACHER BACKGROUND INFORMATION

Signal and Noise in Science

In science, the term *signal* refers to the meaningful information or pattern you are trying to detect; the term *noise* represents the random, unwanted fluctuations or variations that obscure the signal, making it difficult to accurately interpret the data. Essentially, the signal is the important information you want to measure, and the noise is everything else that interferes with that measurement. A signal can be a specific trend, correlation, or change in a variable within the data. For example, consider an experiment to measure the effect of a new drug on blood pressure. The signal would be a change in blood pressure specifically due to the drug. The noise would be variations in blood pressure caused by factors such as individual differences, time of day, or stress levels. Noise can be caused by factors such as instrument error, environmental variations, or individual differences in subjects.

Indoor Air Quality

Indoor air quality refers to the quality of the air in a home, school, office, or other building environment. On average, people in the United States spend approximately 90 percent of their time indoors. The concentrations of some pollutants are often 2–5 times higher indoors than typical outdoor concentrations. Indoor concentrations of some pollutants have increased in recent decades due in part to energy-efficient building construction that lacks sufficient mechanical ventilation to ensure adequate air exchange and the increased use of synthetic materials in building materials, furnishings, personal-care products, pesticides, and household cleaners.

Most pollutants affecting indoor air quality come from sources inside buildings, although some originate outdoors and can enter buildings through open doors, open windows, ventilation systems, and cracks in structures. Indoor sources of air pollution include combustion sources such as tobacco, wood and coal from heating, cooking appliances, and fireplaces; cleaning supplies, paints, insecticides, and other commonly used products; degrading building materials or new building materials (e.g., chemical off-gassing from pressed wood products); radon; mold; and pet dander.

MATERIALS & ADVANCE PREPARATION

FOR EACH GROUP OF FOUR

- COLORED PENCILS (3 different colors)

FOR EACH STUDENT

- STUDENT SHEET 2.1 "Outdoor vs. Indoor Air Quality Measurements"
- STUDENT SHEET 1.3 "Unit Concepts and Skills" (OPTIONAL)
- 2 STUDENT SHEETS 2.2 "Frayer Model" (OPTIONAL)

TEACHING NOTES

Suggestions for **discussion questions** are highlighted in gold.

Strategies for the **equitable inclusion of diverse students** are highlighted in mint.

GETTING STARTED (15-30 MIN)

1 **Introduce the concepts of signal and noise with a tapping game.**

- Introduce the concepts of signal and noise by explaining that scientists try to separate the important information they are looking for—the signal—from other factors that might interfere with it. These interfering factors are called noise. You may want to have students record these concepts on optional Student Sheet 1.3, "Unit Concepts and Skills."

- As a class, you will model how to distinguish a signal from the noise. In this model, noise does not refer to the term *noise* in its more everyday use when it usually refers to unpleasant sounds. Noise is any information that hides, distracts from, or falsely resembles the meaningful information that is being investigated.

  1) Ask one student, the Tapper, to hold their hand under their desk so no one else can see it except the person next to them, known as the Observer. Instruct the Tapper to tap the underside of the desk hard enough for all students to hear (using an object or their knuckles). Instruct all the other students to raise their hands every time they hear the Tapper tap. Each time anyone raises their hand, the Observer tells everyone whether the Tapper really did tap or not. Do this a few times. If students are quiet and the Tapper taps hard enough, it should be fairly easy for students to hear the tap only when the Tapper taps.

  2) Instruct all students to record three column headings in their notebooks: "Heard Tap Correctly," "Misheard Tap," and "Missed Tap." Explain that as they continue to listen, they will keep track of their guesses by making a tally mark under each column according to the following instructions:

     - Heard Tap Correctly: Student raised their hand, and the tap is confirmed by the Observer.
     - Misheard Tap: Student raised their hand, but the Observer says there was no tap.
     - Missed Tap: Student did not raise their hand, but the Observer says there was a tap.

  Conduct a few more rounds of the tapping game while students track their guesses.

3) Next, have students draw a horizontal line under their last entry in their notebooks that goes across the columns. Then, instruct half the students to tap lightly on the top of their desks while still raising their other hands when they hear the Tapper tap. Conduct a few more rounds with students tracking their guesses. This should make hearing the original Tapper difficult, with more Misheard Taps and more Missed Taps.

4) Have students start a third section in their notebooks to record a new set of guesses. Now, instruct all students to join in and tap lightly on the top of their desks. Conduct a few more rounds with students tracking their guesses. This should make hearing the original Tapper even more difficult, with more Misheard Taps and more Missed Taps.

5) Finally, have students start a fourth section in their notebooks to record their guesses and continue the game with everyone tapping as loudly as they can. This should make accurately hearing the original Tapper practically impossible.

- Ask students to explain what they just did in terms of signal and noise. Ask, Did the signal change or the noise? The signal did not change, but the noise increased when others were tapping, which made it harder to distinguish the signal of the original Tapper. Ask, How did this affect the number of errors you made in accurately identifying the original Tapper? The increase in noise increased the number of errors. More noise led to more errors.

- Further develop the concept of signal and noise by using the example of listening to one student say something when everyone else is talking. In this case, the signal (student talking) can be difficult to distinguish from the noise (everyone else talking). Point out that so far, the examples used to differentiate these concepts have used sound. Refer to the Introduction in the Student Book to explain how identifying a signal among noise can also be done visually, such as in a search-and-find book. In the activity, students will differentiate between signal and noise in graphs.

2 Use the literacy strategy of a Frayer Model to support students' understanding of the concepts of signal and noise.

- To support the development of new vocabulary and concepts during the activity, consider using a Frayer Model with students as shown on optional Student Sheet 2.2, "Frayer Model." For the concepts of signal and noise, the model can be introduced in the beginning of the activity, filled out as the activity unfolds, and reviewed at the end of the activity. For more information about the Frayer Model, see Appendix 1: Literacy Strategies. A sample student response of the Frayer Model for the terms *signal* and *noise* is shown at the end of this activity.

- If you have begun a word wall, support students, particularly emerging multilingual learners, in sensemaking and language acquisition by adding the terms *signal* and *noise*.

3 Review how to interpret graphs (optional).

- This activity focuses on data analysis and requires the ability to interpret a graph. Depending on your student population, you may want to review how to interpret graphs. Support students, particularly emerging multilingual learners, in sensemaking and language acquisition by reviewing relevant language associated with a graph, such as x-axis, y-axis, title, key, and line graph.

PROCEDURE SUPPORT (45 MIN)

4 Present the article found in Procedure Step 1.

- As in Activity 1, students are presented with an article based on real-world events. The article presented in Step 1 can be shared with the class in multiple ways. Read the article aloud to the class or have individual students read it aloud while others follow along with the text (either as a whole class or in small groups).
- Reading the article aloud can better support comprehension for many students, including neurodiverse students and emerging multilingual learners who often have more highly developed listening and oral skills than reading comprehension skills. Alternatively, students can read the article independently.

5 Students examine and interpret two graphs.

- Each graph on Student Sheet 2.1, "Outdoor vs. Indoor Air Quality Measurements," provides data about air quality from air quality sensors located on the same wall of a home during a wildfire. The top graph provides data about outdoor air quality, and the bottom graph provides data about indoor air quality. If needed, support your students in interpreting the graphs.
- In Procedure Step 2, distribute 1 copy of Student Sheet 2.1 to each student and 1 set of 3 different-colored pencils to each group. A sample student response is shown at the end of this activity.
- You may wish to model how to complete Procedure Step 3a on Student Sheet 2.1.
- In Procedure Step 3b, students' hypotheses may vary.

Sample Student Response, Procedure Step 3b

- Some students may observe a correlation between wildfire smoke and increased indoor PM2.5 levels, as supported by data on September 13 night, September 15 night, September 17 day, and September 18 night where there are spikes in indoor PM2.5 levels that correspond to spikes in outdoor PM2.5 levels.
- Other students may conclude that there is not a strong correlation between wildfire smoke and indoor increased PM2.5 levels, as supported by data on September 15 day, September 18 morning, and September 19 evening where there are spikes in indoor PM2.5 levels but not in outdoor PM2.5 levels.

- In Procedure Step 5, students use Student Sheet 2.1 to look for and mark correlations in the data as well as mark when additional events occurred.
- In Procedure Step 6, the signal is the data about indoor PM2.5 levels during the wildfire that is not affected by noise. Any sources of variation are noise because they make it harder to determine the relationship between outdoor PM2.5 levels due to wildfire smoke and indoor PM2.5 levels. Students should observe the following:

  Sample Student Response, Procedure Step 6

  - Each graph shows a correlation between a spike in indoor PM2.5 levels and the activity being described (cooking, smoking, house fan).
  - Both cooking and smoking are noise in the graph for indoor air quality when looking for the effect of wildfire smoke.
  - The house fan being on near the indoor sensor made it difficult to gather data about the signal.

- In Procedure Step 7, students identify the signal and the noise in determining whether outdoor air quality is affecting indoor air quality during a wildfire by labeling each vertical box on Student Sheet 2.1 as either signal or noise.
- Students should use the additional information to adjust their hypotheses as needed. Remind students that, in science, the goal is not to be right or wrong, but to have an explanation that is supported by evidence. As more evidence is gathered, explanations may need to be revised. This is part of the process of probabilistic reasoning.

6 **In Procedure Step 9, groups brainstorm ways to improve indoor air quality.**

- Possible student responses include: avoid burning indoors, including fireplaces, wood stoves, gas appliances, and candles; avoid the use of airborne chemicals such as air fresheners, sprays, and cleaning products; vacuum rugs and carpets regularly; routinely wipe down household surfaces; and use an air purifier. If the outdoor air quality is good, opening windows and doors can also increase ventilation and improve indoor air quality.
- Regularly changing furnace filters and reducing humidity levels have also been shown to improve indoor air quality.

SYNTHESIS OF IDEAS (10 MIN)

7 Highlight that identifying the signal vs. the noise depends on the information that you are interested in.

- Revisit student responses to Procedure Step 7, in which they identified the signal and the noise on Student Sheet 2.1. Identifying the signal vs. the noise depends on the information that you are interested in. Since the focus was on indoor vs. outdoor air quality during a wildfire, data that showed a correlation between those two variables was the signal, while data that interfered with determining any potential relationship (such as smoking or cooking) was noise. Ask, When looking at the graphs, what would have been the signal if you were looking at the relationship between indoor air quality and cooking? In this case, data that showed a possible relationship between changes in indoor air quality during cooking would be the signal. Ask, When looking at the graphs, what would have been the noise if you were looking at the relationship between indoor air quality and cooking? In this case, any effect of wildfire smoke on indoor air quality would be noise because it would interfere with the signal (cooking).
- Use Connections to Everyday Life item 3 to highlight the idea that identifying the signal vs. the noise depends on the information that you are interested in.

8 Revisit the article presented in Step 1.

- Have students share their growing understanding of the relationship between air quality and health by asking, Do you agree or disagree with the title of the article: "Is Cooking Worse than Wildfires for Indoor Air Quality?" Point out that using the graphs to answer this question requires identifying cooking as the signal and the other possible effects on indoor air quality as noise. Many students may agree with the title because they saw spikes in the indoor PM2.5 levels on the graph for indoor air quality on Student Sheet 2.1, which matched the days when large meals were cooked indoors. On those days, the outdoor PM2.5 levels were not as high as the indoor levels.
- If you have not done so, have groups share their ideas for improving indoor air quality with the class.

9 Explain that distinguishing the signal from noise is an essential part of data analysis.

- Ask, How do the concepts of signal and noise affect data analysis? Noise makes it harder to interpret the data (the signal). Noise can be a source of error in interpreting data because it can be difficult to tell it apart from the signal.
- You may wish to use optional Student Sheet 1.3, "Unit Concepts and Skills," to help students organize their learning about the concepts of signal and noise.

SAMPLE STUDENT RESPONSES

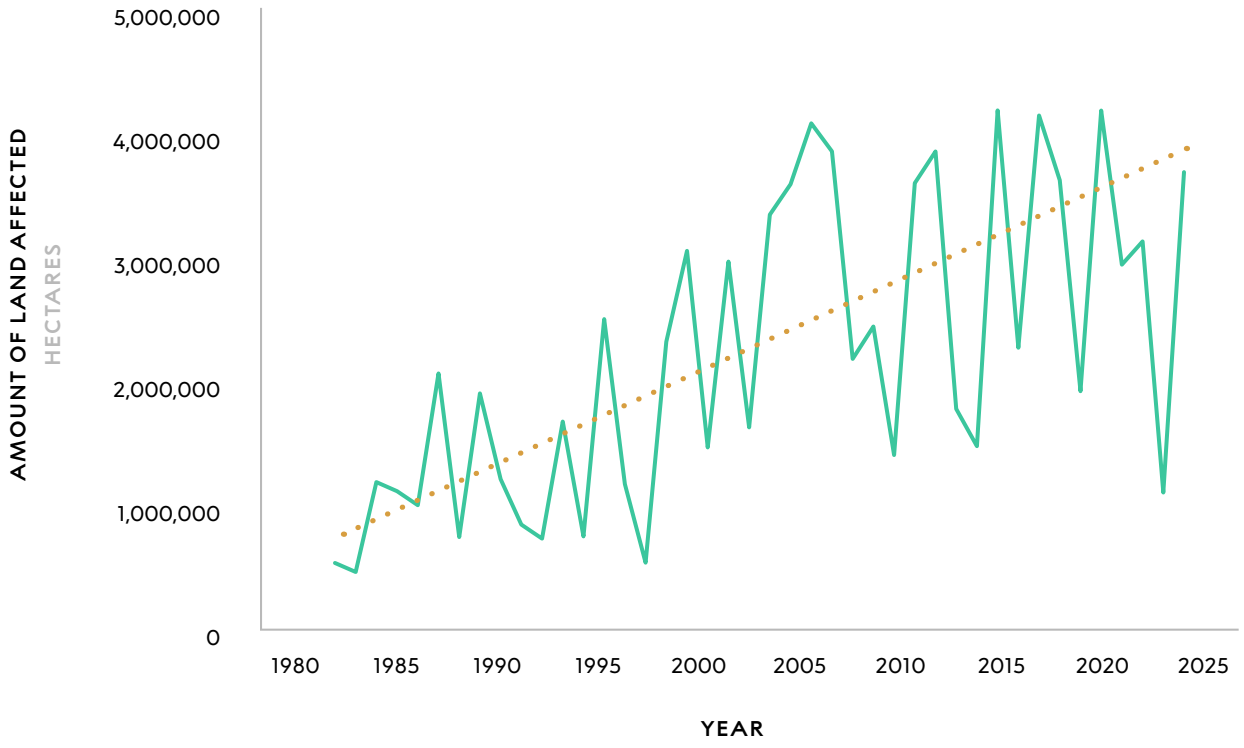
BUILD UNDERSTANDING

① **Would you prefer to stay indoors or be outdoors on a day when there are extreme wildfires in your area and your family is cooking a large meal? Explain your reasoning and support your answer with data from this activity.**

I would prefer to remain indoors because the graphs show that there were many days with spikes in outdoor PM2.5 levels that exceed indoor levels, such as September 13th night and September 17th early morning. I would have more control over indoor air quality. I could reduce activities that increased PM2.5 and use an indoor air purifier.

② **Figure 2.1 is a graph that shows the land area affected by wildfires in the United States since 1983. A hectare is a metric unit for measuring land area, equal to 2.47 acres or 10,000 square meters.**

FIGURE 2.1
Land Area Affected by Wildfires in the United States, 1983–2024



a What is the signal?

The trend line provides the signal to differentiate it from individual events (noise).

b Describe what the noise looks like in the graphed data.

The noise is the lines (the number of hectares burned) going up and down a lot each year.

c Based on the signal, what can you conclude about land area in the United States burned by wildfires over time?

The land area being burned each year is increasing. (The land area affected by wildfires in the United States has increased from around 800,000 hectares—2 million acres—in the early 1980s to around 3 million hectares—8 million acres—by 2020.)

d How might this pattern be similar or different from the area in which you live? ?

I think this pattern is similar to where I live because there seems to be a lot more fires reported in the news now than before.

CONNECTIONS TO EVERYDAY LIFE

③ You are watching a movie with friends. Another friend is joining you later.

a Your dog is barking, making it difficult for you to hear what's happening in the movie. What is the signal and what is the noise?

The sound from the movie is the signal, and the dog's barking is noise.

b You move your dog to another room. He later barks to let you know that someone is at the door. Is your dog's bark a signal or a noise?

The dog's bark is now a signal, telling you it is likely that your friend has arrived.

c Explain how identifying the signal vs. the noise depends on the information you are interested in.

The dog's barking was both a signal and noise, depending on what information you were interested in. While watching the movie, the signal was the sound from the movie. While waiting for a friend, the signal was the arrival of the friend, as indicated by the dog barking.

4. **You may have heard warnings about contaminated food, such as to avoid eating lettuce that has been recalled. In many cases, scientists gather data from individuals who fall sick to determine the common cause of their illness. The U.S. Centers for Disease Control and Prevention (CDC) uses this information and has estimated the annual number of foodborne illnesses in the United States to be 47.8 million cases. In the case of detecting foodborne illness:**

   a **What is the signal?**

   People falling sick from the food-borne illness that is being investigated

   b **What is a possible source of noise in the data? (What other explanation[s] could there be for the data?)**

   People may be sick from other causes.

   c **For its estimate, the CDC accounted for scientific uncertainty and identified that the range of actual cases of foodborne illness in the United States per year may be as low as 28.7 million and as high as 71.1 million cases. Would you expect that reducing noise in this data would increase or decrease the estimated range of the data? Explain your reasoning.**

   Reducing noise would likely decrease the range of data. This would be because other factors that could be mistaken for signal would be reduced. For example, fewer people would report being sick from other causes.

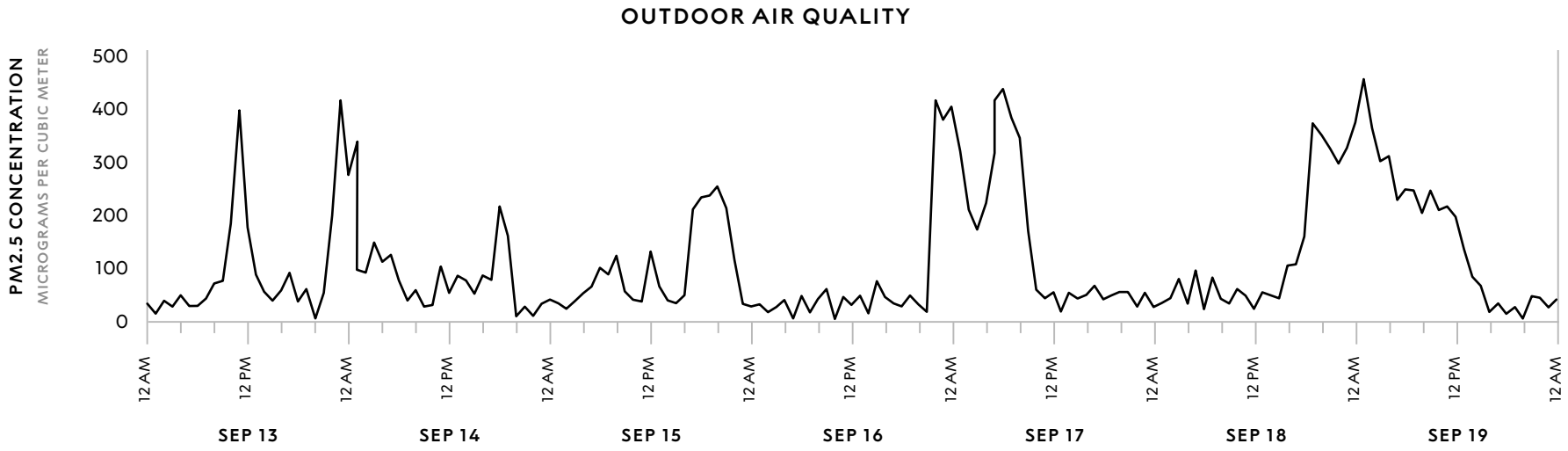
REFERENCES

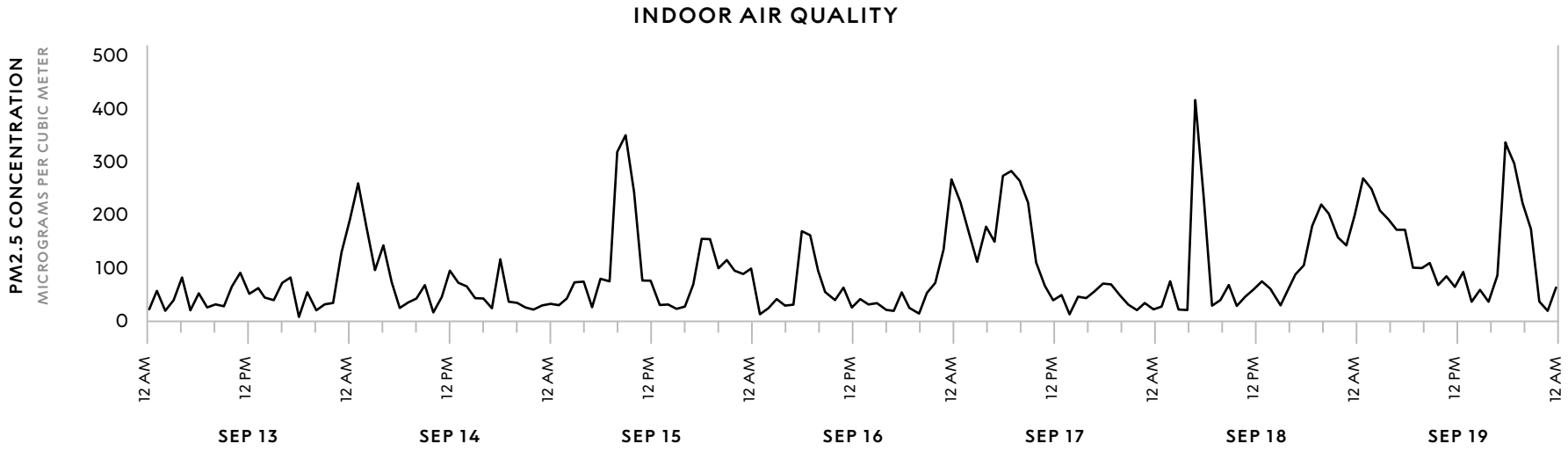
Centers for Disease Control and Prevention. (November 5, 2018.) *Estimates of foodborne illness in the United States: Findings*. https://www.cdc.gov/food-safety/php/data-research/foodborne-illness-burden/

Jones, Ben. (May 13, 2024). *Wildfires in the US: Alarming statistics and trends*. https://www.dryad.net/post/wildfires-usa-statistics

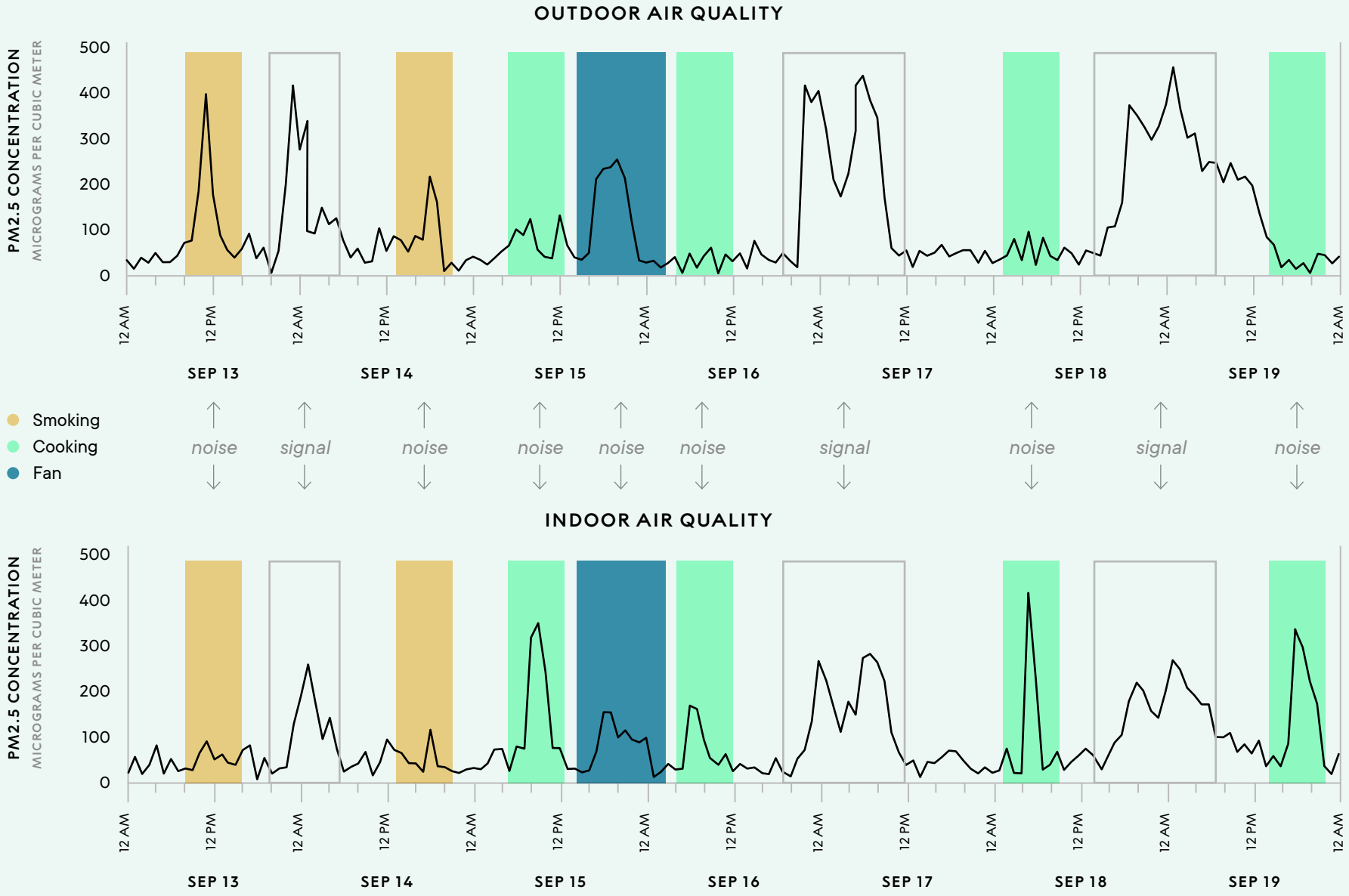
Liang, Y., Sengupta, D., Campmier, M. J., Lunderberg, D. M., Apte, J. S., Goldstein, A. H. (2021). Wildfire smoke impacts on indoor air quality assessed using crowdsourced data in California. *Proceedings of the National Academy of Science of the United States of America, 118*(36) e2106478118. https://doi.org/10.1073/pnas.2106478118

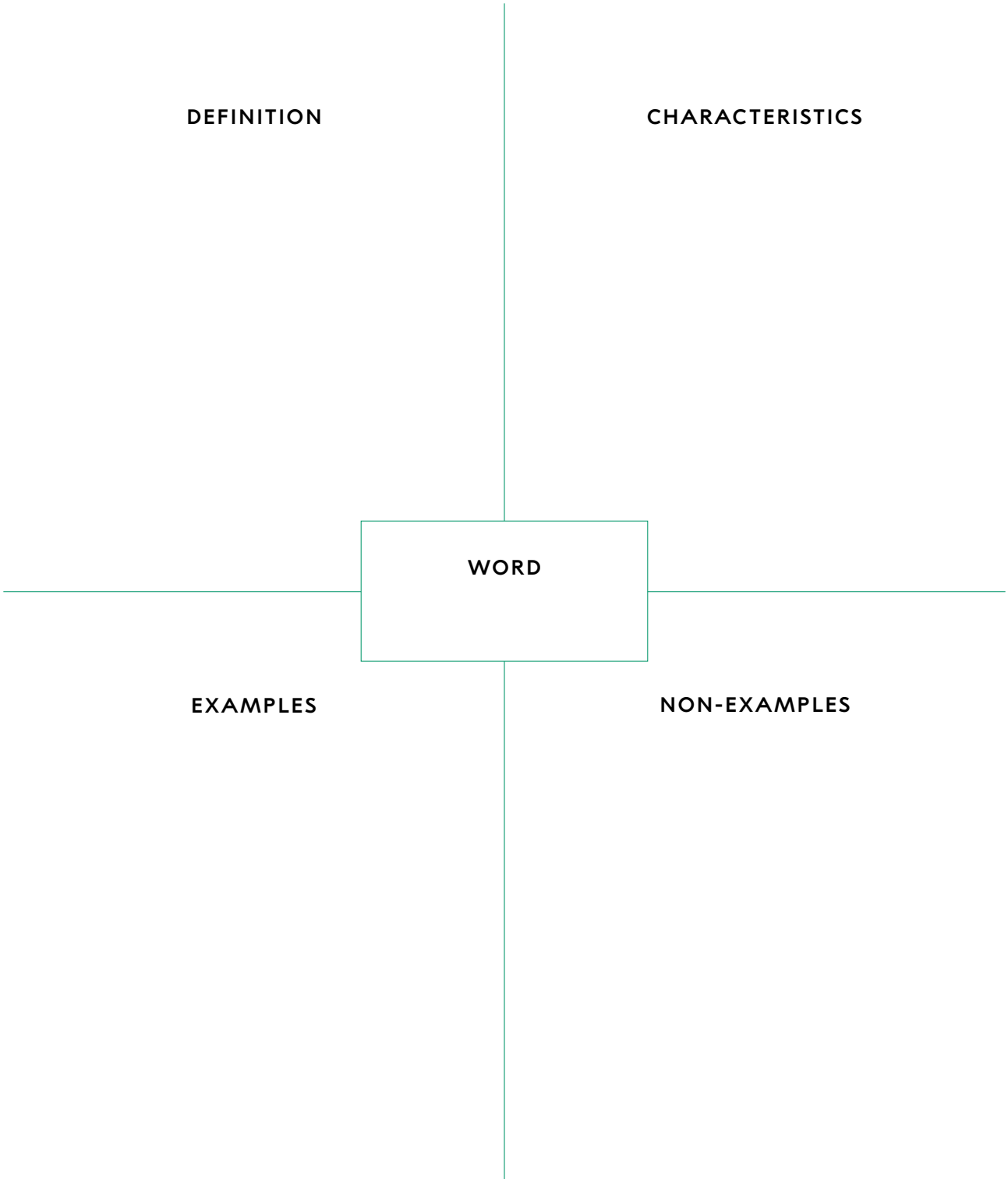
National Interagency Fire Center. (2024). *Wildfires and acres: Total wildland fires and acres (1983–2024)*. https://www.nifc.gov/fire-information/statistics/wildfires


OUTDOOR AIR QUALITY
PM2.5 CONCENTRATION
MICROGRAMS PER CUBIC METER
500
400
300
200
100
0
12 AM
12 PM
SEP 13
SEP 14
SEP 15
SEP 16
SEP 17
SEP 18
SEP 19


INDOOR AIR QUALITY
PM2.5 CONCENTRATION
MICROGRAMS PER CUBIC METER
500
400
300
200
100
0
12 AM
12 PM
SEP 13
SEP 14
SEP 15
SEP 16
SEP 17
SEP 18
SEP 19

SAMPLE STUDENT RESPONSE




DEFINITION
CHARACTERISTICS
WORD
EXAMPLES
NON-EXAMPLES

DEFINITION

meaningful information about the phenomenon that is being investigated

CHARACTERISTICS

- *what the signal looks like depends on the question I'm trying to answer*
- *could be relevant peaks, increase/ decrease in data, gaps in data*
- *signal depends on what I'm studying and how I'm trying to measure it*

WORD

signal

EXAMPLES

- *looking at PM measurements to think about how traffic patterns affect air pollution*

Everyday example
When I do an internet search, relevant and credible sites about what I'm looking for is a signal.

NON-EXAMPLES

- *looking at the data about precipitation when I'm trying to think about how traffic patterns affect air pollution*
- *false patterns*
- *random data, mistaken data points, or irrelevant data*

Everyday example
Ads, sponsored websites, and clickbait articles are not signals.

DEFINITION

information that hides, distracts from, or falsely resembles the meaningful information that is being investigated

CHARACTERISTICS

- *caused by sources of uncertainty such as scientific error, irrelevant data, mistakes in data collection, malfunctioning equipment*
- *makes it hard to see signal*
- *what counts as noise depends on what type of signal you are looking for*

WORD

noise

EXAMPLES

- *data that fluctuates a lot*
- *possible outliers on a graph*
- *a graph trying to show too many types of data at once*
- *false patterns*
- *random data, mistaken data points, or irrelevant data*

Everyday example
When I do an internet search, ads, sponsored websites, and clickbait articles are noise.

NONEXAMPLES

- *best-fit line on a graph*
- *an average of the data*

Everyday example
When I do an internet search, relevant and credible sites about what I'm looking for is not noise.