v 1.0

ACTIVITY 3

Scientific Uncertainty in Data

COMPUTER INVESTIGATION

Crowdsourced air sensors can easily be placed around homes and neighborhoods.

3 : SCIENTIFIC UNCERTAINTY IN DATA

GUIDING QUESTION

What are some sources of scientific uncertainty in data?

INTRODUCTION

In 2015, Utah resident Adrian Dybwad noticed wind blowing dust from a nearby gravel pit into the air. Curious about the effect this had on air quality, he bought air quality sensors. However, the sensors gave very different readings, leaving him unsure about the air quality. In science, the term *error* doesn't mean a mistake, as it does in everyday language. Scientific error is the difference between a measured or observed value and the true value of a quantity. The true value is the actual number that would be found if the measurement could be made without error.

Dybwad, an engineer, started experimenting with air quality sensors to reduce scientific error. He created more reliable sensors than those typically available for purchase and began sharing them with others. Dybwad went on to found PurpleAir, a website that tracks particulate matter data from sensors that people can buy and place in their communities. This type of data collection (by many people in a community) is called *crowdsourcing*. The data collected can help people make informed decisions about their exposure to particulate matter in the air.

In this activity, you will examine air quality data to better understand how close you can get to the true value and to look for sources of uncertainty in the measurements.

CONCEPTUAL TOOLS

MATERIALS LIST

FOR EACH PAIR OF STUDENTS

- COMPUTER WITH INTERNET ACCESS

FOR EACH STUDENT

- STUDENT SHEET 3.1 “Analyzing Crowdsourced Air Quality Data”

PROCEDURE

PART A: CROWDSOURCED SENSOR DATA

1 Review the following Air Quality Index, which is a general guide to air quality based on several air pollutants, including particulate matter, ozone, nitrogen dioxide, sulfur dioxide, and carbon monoxide.

FIGURE 3.1
Air Quality Index (AQI)

AQI CATEGORY (COLOR)	INDEX VALUE	DESCRIPTION OF AIR QUALITY
Good (green)	0–50	Air quality is satisfactory, and air pollution poses little or no risk.
Moderate (yellow)	51–100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive* to air pollution.
Unhealthy for Sensitive* Groups (orange)	101–150	Members of sensitive* groups may experience health effects. The general public is less likely to be affected.
Unhealthy (red)	151–200	Some members of the general public may experience health effects. Members of sensitive* groups may experience more serious health effects.
Very Unhealthy (purple)	201–300	Health alert: The risk of health effects is increased for everyone.
Hazardous (maroon)	301 AND HIGHER	Health warning of emergency conditions: Everyone is more likely to be affected.

* According to the American Lung Association, sensitive groups include children under 18, adults over 65, people with chronic heart or lung disease, people who are pregnant, and people with diabetes. Adults who are active outdoors, including outdoor workers and frequent outdoor exercisers, can be considered sensitive because of prolonged exposure to outside air.

2 With your partner, explore the air quality of your state today by using more widely available crowdsourced air quality data, such as from the PurpleAir website.

3 On Student Sheet 3.1, "Analyzing Crowdsourced Air Quality Data," record today's date and the name of your state in the appropriate places in Table 1. Then work with your partner to complete the first row of Table 1.

a Record the range of PM2.5 measurements in your state today.

b Use this air quality data and the Air Quality Index to determine the general air quality category in your state today and then record today's air quality and how you made your determination.

c Work with your partner to describe possible reason(s) for scientific uncertainty in the data and then record your ideas.

HINT: Think about reasons that the air quality data might not represent the true value and/or possible sources of scientific error in the data.

4 Work with your partner to identify an area of your state where there are *few* sensors and then complete the second row of Table 1.

a Record the name of the area and the available data.

b Discuss what you can conclude about the air quality in that area, using the Air Quality Index, and record today's air quality and how you made your determination.

c Work with your partner to describe possible reason(s) for scientific uncertainty in the data and record your ideas.

5 Work with your partner to identify an area of your state that has *many* sensors and then complete the third row of Table 1.

a Record the name of the area and the range of data available in this area.

b Discuss what you can conclude about the air quality in that area, using the Air Quality Index, and record today's air quality and explain how you made your determination.

c Work with your partner to describe possible reason(s) for scientific uncertainty in the data and record your ideas.

d Discuss whether you are more sure of your determination of air quality in the area with fewer sensors or more sensors. Explain your reasoning.

6 Work with your partner to gather more data from an area with many sensors and then complete the "Crowdsourced Data" column of Table 2 on your student sheet.

a Select and record five measurements.

b Identify and record any number(s) that differs significantly among the five measurements. With your partner, discuss possible explanations why this data point(s) is significantly different from the others.

c Calculate and record the average of the five measurements.

d Use the Air Quality Index (from Procedure Step 1) to determine the air quality for this area and record it.

e Compare your air quality findings based on the average of five data points with your determination from many sensors (Procedure Step 5b). Describe to your partner how your results are similar or different, which air quality determination is more likely to be closer to the true value, and why.

PART B: HIGHER-QUALITY SENSOR DATA

7 a Go to a site that provides data from higher-quality sensors, such as the Fire and Smoke Map at *AirNow*. *AirNow* is a partnership of federal, state, local, and tribal air quality agencies in the United States. It shows data from permanent and temporary air monitoring stations (which have higher-quality sensors) and crowdsourced sensors such as PurpleAir.

b If you are using *AirNow*, remove the data from sensors such as PurpleAir by clicking on the MAP SETTINGS button on the top right of the screen and then turning off the AIR SENSORS toggle (seen under the LAYERS bar).

Environmental Protection Agency (EPA) scientist Andrea Clemens (standing) examines an air quality sensor at an EPA research site in North Carolina.

8 Find the area on the map with many sensors that you investigated in Procedure Steps 5 and 6. Work with your partner to gather more data to complete the "Higher-Quality Data" column of Table 2 on your student sheet.

 a Select up to five measurements from the available higher-quality sensors and record them. (NOTE: There may not be five high-quality sensors in the area you selected. If not, use the data available in the area.)

 b Identify and record any number(s) that differs significantly among the five measurements. If there are no differing data points, record "none."

 c Calculate and record the average of these measurements.

 d Use this air quality data and the Air Quality Index (from Procedure Step 1) to determine the air quality category for this area and record it.

9 Compare your air quality findings from high-quality sensors such as *AirNow* with sensors such as PurpleAir. Describe to your partner:

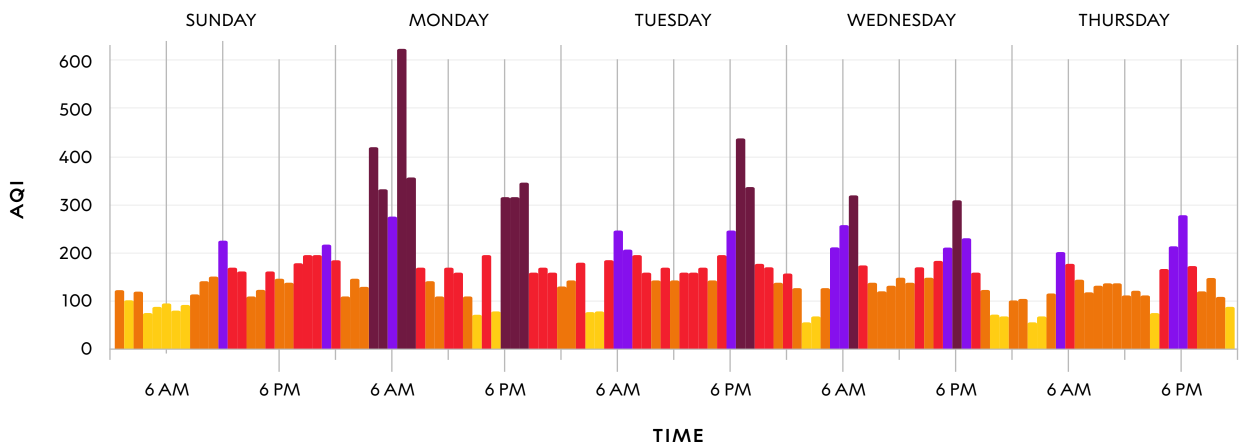
- how the data are similar or different.
- which data set is likely to be closer to the true value (have less scientific error) and why.
- how probabilistic reasoning was involved in making your determination of air quality.

BUILD UNDERSTANDING

1. Review your analysis of local air quality data from both sites.
   - **a** List two possible sources of scientific uncertainty in air quality data.
   - **b** Brainstorm how each source of scientific uncertainty could be reduced.

2. The graph in Figure 3.2 shows several days of hourly AQI measurements for the capital city of N'Djamena in the country of Chad. Explain what conclusions you can make based on the data in the graph. Refer to Figure 3.1, "Air Quality Index (AQI)" in Procedure Step 1 as needed. In your explanation, be sure to include the following:
   - **a** Describe what patterns you observe in the air quality over time.
   - **b** Explain what conclusions you can make about local air quality.
   - **c** Explain at least two possible sources of scientific uncertainty, including possible scientific errors, that may have affected the data.

FIGURE 3.2
Air Quality Index (AQI) Data for Five Days in N'Djamena, Chad



3. Based on the data in Figure 3.2, which day would be better to be outdoors: Monday or Tuesday? Support your answer with evidence.

CONNECTIONS TO EVERYDAY LIFE

4. The local Air Quality Index (AQI) on the day of your team's soccer semifinals is reported as 135. A sensor near your home shows an AQI of 100. The best player on your soccer team has asthma. Would you recommend that she play in the semifinal game? Support your answer with evidence and identify the trade-offs of your decision.

5. People rely on data such as air quality from scientific tools and technology to make decisions. Yet data from such technology can sometimes be inaccurate. In the case of an air quality sensor, is it worse to have a false positive or a false negative? Explain your reasoning.

EXTENSION

Use an online air quality map (such as one from PurpleAir) to look at data in other parts of the world over a period of several days. Look for any patterns in the data and develop a tentative explanation for what you observe. Brainstorm how you could further investigate your ideas while reducing scientific uncertainty.

KEY SCIENTIFIC TERMS

scientific error
true value