LESSON

How Do You Investigate and Represent Data?

Understanding what data to collect and how to represent it to best answer a research question.

What We Are Hoping For: Learning Goals

- Nature of Science
  - A-E
- Data Representation
  - A, B, C

Learning Goals:

- Investigate
- Represent data
- Create an appropriate scale for the dataset
- Graphing

CHECKLIST

How Do You Investigate and Represent Data?

In this lesson, students will:

- Discuss quote: "Water is the reflection of all the bad water quality things we do on land." What does that mean? **(3 min)**
- Relate quote to why Dr. Kaushal tested water in a forested area. **(3 min)**
- Use distributed datasets to predict salt levels for urban, suburban, and forested steams (**4 min)**
- Watch prepared slideshow. **(20 min)**
    - Discuss how population density impacts salt levels.
    - Discuss how scientists choose how to represent their data.
    - Discuss the importance of a proper scale on a graph.
- Complete *How do you investigate and represent data?* section of the Investigation Booklet. (**5 min**)
- Set up graphs including: **(10 min – continue in next lesson)**
- Select a set of benchmarks for their graphs.
    - Indicate population densities of each area on their graphs.
    - Choose appropriate checkboxes in title
    - Begin to graph data (if time allows)

(Times indicated are approximate.)

1. Reintroduce Dr. Sujay Kaushal's story with the following quote from the video: "Water is the reflection of all the bad water quality things we do on land." Relate the quote to why Dr. Kaushal tested streams in forested areas. (6 minutes)

Land Affects the Water.

What happens on land affects the water.

Discussion

Key Idea: What happens on land affects the water.

Question: What does this statement mean?
Answer: This quote means that whatever happens on land ends up affecting the water, because everything eventually ends up running off into the water.

Question: Looking at your local surroundings, what do you think the water quality is like in your neighborhood?
Answer: Probably not very high. New York City is very densely populated with lots of asphalt, trucks, and cars, all of which affect nearby water quality.

Discussion

Key Idea: Dr. Sujay Kaushal chose to compare urban, suburban, and forested streams to gauge human impact on streams.

Question: So if Dr. Kaushal wanted to determine what we were doing on land by testing the water, why does he bother to test forested areas where people do not even live?
Answer: Dr. Kaushal tested both areas because he was interested in comparing water in areas with high levels of human disturbance to areas with low human impact.

Question: Would his research design have been strong if he had only tested the urban stream water? Why or why not?
Answer: If Dr. Kaushal had only studied the urban stream, he would be unable to gauge human impact because he would not know the salt level normally found in Baltimore streams that are undisturbed by people. By studying urban and forested areas, he was able to compare and give meaning to his data.

2. Distribute the complete datasets to the class and lead a discussion about the difference in the three areas, and why it is important for the class/scientists to look at data from forested, suburban, and urban water areas. (20 minutes)

Where to Collect Data

Finding differences that make a difference.

Discussion
Before the slideshow when students are just holding the data:

"It is now your turn to work with the data. You will graphically represent and analyze the data that Dr. Kashual collected by studying forested, urban, and suburban areas."

Question: What type of data are we looking at? How are the datasets different?
Answers: 1) Seasonal data
2) Forested/Suburban/Urban – different density of roads

3. Present slide show, which begins with a discussion of how population density impacts salt levels. (20 minutes)

How to Represent Data

What is the best way to represent the data in order to compare different data sets?

Begin slideshow. Slides 1-5 discuss how population density impacts salt levels.

Discussion

Key Idea: Salt will be most abundant in areas that have many roads and high traffic levels due to differences in population densities.

Question: Where do you expect to find the saltiest water - urban, suburban or forested locations?
Answer: In an urban location because there are more roads

Question: Where do you expect to find the least salty water - urban, suburban or forested locations?
Answer: In a forested areas because there aren't any roads

Question: How would you test these predictions? Look back at your data. What type of data are we using?
Answer: Seasonal and location data, therefore, we would need to analyze data from forested, suburban, urban locations to see whether water in urban locations is saltier and whether salt levels increase in the winter.

Slide 6 summarizes the data and research the students are analyzing. It says:

- Dr. Kaushal analyzed data from streams near Baltimore considered to be urban, suburban, and forested.
- He analyzed samples during each season to determine how salt levels changed.
- He was lucky. The dataset was enormous, going back over 30 years. You are only looking at a subset of the data he and his colleagues analyzed.

Key Idea: The annual salt levels might differ in forested, suburban, and urban streams (because of roads)

Each area has a different population density that directly reflects how many roads are near the study streams - the greater the density of people, the greater the number of roads.

- Forest: *Although people walk and drive to the forest, no people live in it.*
- Suburban stream: *population density* ~3000 people/mile^2
- Urban stream *population density ~8000 people/mile^2*

***Note for perspective**: The population density of New York City is 26,403 people/square mile, more than any other city in the U.S. The next highest population density is San Francisco with 17,323 people/square mile.*

Give students time to examine the datasets to determine how to best represent the data. This examination will lead to a discussion of how best to represent the data for analysis. The slideshow also contains slides of the datasets, so the class can more easily discuss them as a group.

Discussion

Key Idea: Scientists carefully choose ways to represent data based on needs for analysis.

Question: Data tables are a form of raw data, which must be analyzed by scientists to draw conclusions. What is the best way to represent the data so that the different datasets can be easily compared? *In Slideshow*
Answer: A graph, because graphs make it easy to look at patterns in data.

Question: When creating a graph, what data goes on the x-axis? What data goes on the y-axis? Look at the columns in your data tables to see what type of data you have.
Answer: x-axis: Time (years, seasons) (independent variable)
y-axis: Salt levels (dependent variable)

Note: This would be a good time to introduce the concept of dependent / independent variables, but it is not necessary for the completion of the unit.

the slideshow.)

Discussion

Key Idea: The importance of the proper scale on a graph.

Question: How do you create an appropriate scale for the dataset?
Answer: (Use the sample graphs of growth of money over time—Maya and John's investment income—in the slideshow.)

Scaling Misconception Alert

The data ranges of the forested (high of 3 mg/L) and urban (high of almost 1800 mg/L) datasets are very different. It therefore will be tempting for students to use different scales for them. Yet, in order to compare the datasets, students must use of a single scale. It will not be possible to compare data, if they use different scales for the various datasets.

Discussion
Key Idea: Scientists must choose the right scale for a graph so the data can be correctly analyzed.

Slide 11: The two graphs show Maya and John's investment income from 2009. The graphs show the same growth in income, but are different scales, which make it look like Maya made more money than John. (*Teachers Note: They both earned the same income)*

Question: Who made more money, Maya or John?
Answer: Most students will answer Maya because of the differences in scale, however they both made the same amount.

Slide 12: The next slide shows both investments on the same scale.

Question: Who really made more money?
Answer: Both Maya and John made the same amount of money.

Question: Explain how it is possible for both of them to be making the same income with this graph, but for it to appear as if Maya is making more money on the previous graph if both graphs are of the SAME data.
Answer: The scale was different on both graphs, which made it look like Maya had a bigger increase, even though her total investment income was the same as John's.

Question: Why is it important to choose an appropriate scale for your data?
Answer: If the scales are different, you cannot compare the data.

Question: Examine your datasets from forested to suburban, and decide what scale you would like to use for your graphs. What should be the biggest number on the Y axis? Remember you need to take into account the forested, suburban, and URBAN data.
Answer: Come to a conclusion as a class, each student **MUST** have the same scale, so that the graphs can be compared.

Note: Numbers vary from as low as 2 mg/L of salt for forested areas and as high as 1800 mg/L for urban areas. However, **ALL** graphs, even the forested area, must use the same scale that reaches 1800 mg/L. A labeled graphing template is included in the materials for this lesson.

4. Complete the How do you investigate and represent data? section of the Investigation Booklet. (5 minutes)

5. After completing the Investigation Booklet, students can begin setting up graphs to represent their data for the rest of the period. (10 minutes)

Setting Up Graphs

Getting a start on graphing the salt level data.

If they don't already have the datasets, distribute the complete datasets to the class, from the "How to Represent Data" activity.

Graphing Directions
A guide to group work on the graphs

1) Hand out the graphing template to each group of 4, which is prepared with a preset scale and a benchmark from the salt and ecosystems case study.
2) Assign each student one of the datasets: 1-forested, 2-suburban, 3-urban, and 4-annual data.
3) Before graphing ask students to use the salt and ecosystem case studies (see the "Salt and Ecosystems" activity) to add 2-3 more benchmarks onto their graphs. The students can use the sample benchmark as a guide for how to put the benchmarks onto the graphs.
4) Groups can choose a bar or line graph, but each person in the group must do the same type of graph for ease of comparison.
5) Ask students to fill in different population density levels for each area, which can be found on their datasets on the top of the graph next to the key.
6) Ask students to choose the appropriate check boxes in the title
    - Pick one checkbox for annual or seasonal
    - Pick all checkboxes that apply for forested, suburban, and urban streams
7) Remind students to individually label their lines or bars, if they are plotting data from more than one stream on their graphs (e.g. the annual dataset includes data from forested, suburban, and urban streams).
8) Check student work.
9) Ask students to begin graphing. If a student finishes early, they can begin to graph other datasets onto their graph. *Note: Make sure students delineate between the different datasets by using different colors or style of lines.*

Scaffold — Rounding Numbers

Some students might need help determining how to round off numbers. For more instruction on rounding, try these websites:

Math is Fun, Rounding Numbers
http://www.mathsisfun.com/rounding-numbers.html

My Schoolhouse — Rounding Numbers
http://www.myschoolhouse.com/Teachers/Resources/rounding%20numbers.htm

Tips for Graphing

It is important for the students to complete the graphs correctly because their analysis hinges on the correct plotting of their data.

- When students plot the salt and freshwater benchmarks, they should use a straightedge to draw a straight line at the correct level of salt.
- When student groups choose to do a line or bar graph, make sure that everyone makes the same type of graph for ease of comparison.
- Ask students to complete the graph in pencil first and check their points before having them connect their lines or create their bars.