# **Exploring Big Data**

### ENGAGE: Get the Picture?

### Data...

• Which phone data plan do you prefer—1 gigabyte or 8 gigabytes?

### Data in our Everyday World

Image with description	What is it? How do you know?







# How do you know when you have enough data?"











### Data...

- 1. What are data?
- 2. Why is the amount of data important?
- 3. Think of one or more everyday circumstances in which increasing amounts of data help you understand something better. Describe these circumstances.

# How are the pixels in each photo data?

### EXPLORE: The Big Deal about Big Data

### **Using Analogies**

- What is an analogy?
- How can analogies help us understand something new?

any of a class of solids (as germanium) that have an ability to conduct electricity between that of a conductor and that of an insulator

dissolved oxygen

temperature

Caption:

### Write a Caption

- Below your scatter plot, write a caption that includes
  - What I See: things you notice about the data on the scatter plot
  - What it Means: what you think the pattern (or not) in the data means.

### **Using Analogies**

Analogy Map		
Photo	O <sub>2</sub> vs temp scatter plot	How they are alike

### Summarize our Learning

- How much data did they need to see a pattern or trend?
- Is there a certain point where more data don't really add to recognition?
- Scientific inquiry has a lot to do with thinking and learning. How do analogies help you conduct scientific inquiry?
- So what is "Big Data" anyway? How is it different than regular data sets?

EXPLAIN: Oyster Populations: A Case Study for Big Data

### Data in our world

Can you think of an example of two things that might be related to one another in your life?

Example: How many text messages you send and your battery life on your smartphone.

## The uses of Big Data

• Let's generate a list of ways in which Big Data might be helpful for scientists.

• How might scientists collect this kind of data?

### **Types of Scientific Questions**

- What is the relationship between variables?
- For a given dependent variable, is there a difference between two groups?

### Trends in Oyster Populations in Chesapeake Bay



Graph: US Fish and Wildlife Service; Data source: EPA Chesapeake Bay Program

### Practice

- Look at Visualizing Trends and Differences
  - What do you think? Is there a trend or difference?
  - How do you you know?

### Chesapeake Bay Water Quality

### Part 1

 For each graph, (1) write on and next to the graph things you notice about the graph (What I see), then (2) below each graph describe what you think it means (What it means).

#### Graph 1: Salinity versus Conductivity

Conductivity measures how easily electricity flows through water. Typical values for conductivity are rain water: 2 to 100  $\mu$ S/cm, ground water: 50 to 50,000  $\mu$ S/cm, ocean: 50,000  $\mu$ S/cm, drainage from landfill: 10,000  $\mu$ S/cm.



#### Graph 2. Dissolved Oxygen and Conductivity



#### Graph 3. pH versus Conductivity



New York Lake Erie Ссорн Bath Binghamtor ★ Chesapeake Bay Scrantog Ohio Watershed Pennsylvania State College New Jersey Harrisburg Potomac River Maryland Watershed Baltmore Petersburg West Virginia Washington DC Delaware × Salisbury Virginia Richmond Atlantic Ocean Angie 1997 0 25 50 100 150 Kilometers

Geography of large watersheds draining into Chesapeake Bay

Interstate Commission on the Potomac River Basin

Salinity in River Outlets



Population Density of the Chesapeake Bay Watershed



### Chesapeake Bay Water Quality

### Part 2

For each histogram, (1) write on and next to the histogram things you notice (What I see), and then (2) below each histogram describe what you think it means (What it means).



#### 1. Conductivity Levels in the Lower Chesapeake

#### 2. Conductivity in the Potomac Watershed





#### 3. Conductivity in the Upper Chesapeake Watershed



#### 4. Conductivity in the Susquehanna Watershed

## **TASK: Developing an Argument**

In teams...

- Develop explanations related to the decline of the oyster populations. You may argue for either A: Urbanization has caused oyster decline, or B: Urbanization has NOT caused oyster decline.
- Use the data sources and maps to fully develop their explanations of what they think is happening.
- In particular, pay close attention to the relationships between salinity and urban areas within the Bay.
- All team members must understand all aspects of the argument.
- Teams will each design and construct a visual aid. This visual aid will be used during the argument.
- Visual aid must be a map or graph with a summary or caption. The summary can be in either paragraph or bullet point format.
- Teams may search for appropriate "extra" information to bolster their argument.

### Whole Class Debrief

As we look at the evidence for each argument, two team members will do the following:

- Team Spokesperson: Present the main points of the OPPOSITE argument.
- Team Recorder: Make a master list of bullets for EACH argument.

### Summarize our Learning

- In general, what is the relationship between the number of data points and the ease of recognizing a trend or difference?
- In general, what is the relationship between how "spread out" data are and the ease of recognizing a trend or difference?
- Why would you choose one visualization (graph) over another?
- Why are these graphs helpful for people working with big data?
- How did using data and graphs help you to understand the oyster population decline in the Chesapeake Bay?

ELABORATE: Using Big Data to Answer Scientific Questions

## Forming a Scientific Question

- Let's think
  - What kinds of scientific questions could we ask that would help us understand trends between two variables?
  - What kind of scientific questions could we ask that would help us understand differences in populations?

### Now it's your turn

- With a partner, think about a scientific question you want to investigate in FieldScope.
  - Write your question down.
  - Decide which data (variables) you need to answer your questions.
  - Decide how you might visualize your data.
- Expectations:
  - A questions, with a graph, clearly labeled axis, and a caption below explaining observations in the data and what you think it means regarding your question.

### **Team Sharing**

- Presenters: Share your question, your graph, and your interpretation with another team.
- Audience: Listen carefully and provide feedback and questions to the presenters.
- Be ready to share the OTHER team's presentation!

### Whole Class

- Do all trend questions result in an obvious trend? Does an answer of "no" mean you have made a mistake?
- Do all difference questions result in an obvious difference? Does an answer of "no" mean you have made a mistake?
- What do you write for a claim when there is no obvious trend or difference? What does the evidence for each of these situations look like?

### Summarize our Learning

- How did using crowdsourced data in FieldScope help you answer your question (or not)?
- Could you have answered it in a different way using other sources of evidence?