



Big Data: A Different Kind of Science

Core Concepts:

- “Big data” is a term for extremely large and complex data sets that may be stored and analyzed computationally to reveal patterns, trends, and associations.
- Big data research studies may be less expensive and less time consuming than traditional science research.
- Big data research studies may be easier to do for long-term research or for complex diseases affected by many interacting factors.
- Big data research involves computer science, mathematics, science, and medicine to analyze big data sets to find better ways to prevent, diagnose, treat, and cure human diseases.
- Mining big data involves analyzing big data to look for patterns or correlations between complex interacting factors.
- To answer medical questions, doctors and scientists are increasingly analyzing big data that already exists in the electronic medical records of millions of patients.
- The science of big data research is evolving to link electronic medical data bases to other sources of data such as DNA analysis, social media, online purchases, and wearable devices.

Class Time Required:

1 forty-minute class period

Teacher Provides:

For each student

- One copy of student handout entitled **Big Data: A Different Kind of Science**
- One copy of **Big Data Stories**. Laminate or put in a sheet protector so that this can be used for multiple classes.

Suggested Class Procedure:

1. Students may work individually or in pairs for this activity.
2. Give each student a copy of the **Big Data: A Different Kind of Science** student handout and a copy of the **Big Data Stories**.
3. Tell students that they should read the questions on the **Big Data: A Different Kind of Science** sheet so that they know what information they should be looking for in stories.
4. Students read the **Big Data Stories** and use that information to answer the questions on the **Big Data: A Different Kind of Science** handout.
5. Discuss the answers to the questions with the class.

Teacher Resources:

- **The Big Data Conundrum: How to Define It?**
<https://www.technologyreview.com/s/519851/the-big-data-conundrum-how-to-define-it/>
- **The Role of Big Data In Medicine**
<http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-role-of-big-data-in-medicine>
- **Big Data Coming In Faster Than Biomedical Researchers Can Process It**
<http://www.npr.org/sections/health-shots/2016/11/28/503035862/big-data-coming-in-faster-than-biomedical-researchers-can-process-it>
- **The Power of Big Data Must Be Harnessed for Medical Progress**
<http://www.nature.com/news/the-power-of-big-data-must-be-harnessed-for-medical-progress-1.21026>
- **Big Data Sets Available for Free**
<http://www.datasciencecentral.com/profiles/blogs/big-data-sets-available-for-free>
- **Millions of Tweets Are a Goldmine for Data Mining** (*NOTE - This article has links to other interesting articles*)
<http://www.rochester.edu/newscenter/twitter-provides-a-rich-vein-of-data-218722/>
- **How Target Knew a High School Girl Was Pregnant Before Her Parents Did**
<http://techland.time.com/2012/02/17/how-target-knew-a-high-school-girl-was-pregnant-before-her-parents/>

Big Data Stories

1. Noelle's Big Data Story

Noelle is a big data analyst. Big data is a term for collecting and analyzing data sets that are so large or complex that traditional data processing applications are inadequate. Big data research uses extremely large and complex data sets that can be collected, stored, and analyzed to reveal patterns, trends, and associations.

Noelle used her background in computer science, statistics, and mathematics to do big data analysis for businesses. First, she worked for Amazon where she used big data to target advertising to people who were most likely to purchase specific products. Then, she worked for a large grocery store chain to identify variables that affect worker efficiency and store operating costs.

Now, Noelle is applying for a job with a large healthcare company that wants to identify cost effective ways to improve the quality of medical care for people who live in rural communities. If she gets the job, she will be “mining” big data to discover patterns and trends in variables that affect healthcare costs and quality.

2. Jason's Big Data Story

Jason has Parkinson's disease—a nervous system disease that causes increasingly severe problems with involuntary shaking, difficulty walking, depression, and dementia. He has given permission for big data researchers to access his electronic insurance records and electronic medical records, including his brain imaging tests and his blood tests.

Jason answered an extensive survey about his childhood and he donated DNA samples for genetic analysis. He is participating in a clinical trial that is studying the safety and effectiveness of a new Parkinson's disease drug. Jason wears a smart watch that measures his movements and sends information to the researchers via an app on his cell phone.

Parkinson's disease researchers are “mining” big data to search for patterns or trends that make sense of the data sets collected from Jason and thousands of other Parkinson's disease patients.

3. Frank's Big Data Story

Frank is a medical research scientist. Last year, Frank completed an expensive 5-year traditional scientific study to investigate whether taking fish oil supplements reduced people's risk for heart attacks. He compared an experimental group of people who took fish oil supplements with a control group of people who did not take fish oil supplements. He found that the heart attack rate in both groups was the same.

Heart disease is a complex disease that may be associated with many different interacting factors. Big data science is more likely to uncover the answers for complex or uncommon medical problems because it explores data involving multiple variables collected from many people and multiple sources.

Frank knows that big data research studies are less expensive and less time-consuming than traditional science research because scientists are increasingly analyzing big data that already exists in the electronic medical records of millions of patients. These electronic medical records can be linked to data from other sources such as from social media, online purchases, wearable devices, and other electronic sources.

4. Geri's Big Data Story

Geri is a 19 year-old college student. Several people in Geri's family have been diagnosed with breast cancer. She is considering participating in big data research that will explore ways to prevent, diagnose, and treat breast cancer. If she participates in this research study, she will need to grant access to all of her medical records.

Geri will be paid to wear a smart watch that measures her sleep patterns and activity levels. The wrist band on this watch will be made of a special polymer that samples the chemicals in the environment. Geri will also answer questions on a monthly survey and allow researchers to use the GPS on her phone to track where she travels. She has heard that big data analysts can access her social media, her online purchases, and her use of store credit cards.

Geri's mother is concerned that big data research looks at medical records and other information that is personal and protected by privacy laws. She is concerned that the data collected may be invasion of Geri's privacy and could be misused.

Big Data: A Different Kind of Science

The four **Big Data Stories** provide examples of how “big data” can be used to improve medicine or healthcare. Read the questions below and then use the information in the **Big Data Stories** answer the questions.

1. What is meant by the term “big data?”

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate.

2. Big data research combines knowledge from many different fields. List three different subjects that would be important in collecting and analyzing big data to improve the prevention, diagnosis, and treatment of diseases.

Computer science, statistics, mathematics, biology, and medicine

3. People who work in the big data field collect and organize different types of data from a variety of sources, such as doctor’s office records and social media. Use the information in the stories to make a list of five sources for data that could be used in a big data set.

Hospital data bases, hospital records, insurance records, medical tests, surveys, DNA or genetic analysis, clinical trial data, GPS, smart watches, health Apps social media, online purchase records, store reward cards, income records.

4. People who work in the field of medical informatics “mine” big data. What does “mining” big data mean?

Data mining refers to the activity of searching for patterns or trends that make sense of the data sets collected from thousands of people.

5. Describe two ways that big data research is different from traditional large research studies, such as controlled experiments.

Traditional large research studies are expensive and time consuming. Big data research uses existing data involving multiple variables collected from many people and multiple sources. Mining big data is more likely to discover new connections (trends and patterns) between many different variables.

6. Explain why some people are concerned that big data research violates laws that ensure an individual’s right to privacy.

Big data research looks at medical records and other information that is personal and protected by privacy laws.