



Mining Big Data: Osteoporosis

Core Concepts:

- Big data research typically involves “mining” data bases (sets) to look for trends, patterns, associations, or correlations among variables that have not been previously discovered.
- The overall goal of the data mining process is to extract meaningful information from a data set.

Class Time Required:

1 forty-minute class period

Teacher Provides:

For each student

- One copy of student handout entitled **Mining Big Data: Osteoporosis**

For each pair of students

- Computer with internet access
- Posters that students made in the **Collecting Big Data** lesson and/or **Variables in a Big Data Set** sheet

Teacher Resources:

- **Thinking Twice About Calcium Supplements**
<https://well.blogs.nytimes.com/2013/04/08/thinking-twice-about-calcium-supplements-2/>
- **What is Data Mining in Healthcare?**
<https://www.healthcatalyst.com/wp-content/uploads/2014/06/What-is-data-mining-in-healthcare.pdf>
- **10 Great Healthcare Data Sets**
<http://www.datasciencecentral.com/profiles/blogs/10-great-healthcare-data-sets>
- **Big Data: 33 Brilliant and Free Data Sources for 2016**
<http://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-data-sources-for-2016/#545ac0567961>

Copyright © 2018 by University of Rochester. All rights reserved. May be copied for classroom use.

This lesson was developed with support from the National Institutes of Health under Award Number R25OD010494-03S1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Suggested Class Procedure:

1. Prior knowledge of osteoporosis is important for this activity. This can be obtained by completing the **A Case of Osteoporosis** activity or allowing time for students to do Internet research on osteoporosis.
2. This activity is more engaging for students if they have completed the **Collecting Big Data** activity and can use the posters created by their class.
3. Read the “Teens and Osteoporosis” information at the top of page 1 aloud to the class.
4. Assign students to work with a partner. Allow the partners to select the factor that they would like to study.
5. Distribute the **Variables in a Big Data Set** sheet and/or hang the posters that students made during the **Collecting Big Data** activity.

*Note: The student instructions are written as if they should use the **Variables in a Big Data Set**. If you have posters from the **Collecting Big Data** lesson, you will need to instruct students to use the posters instead of (or in addition to) the **Variables in a Big Data Set**.*

6. Read question 2 and model how you could answer question 2 for the factor “Calcium intake.”
For example, some people claim that low calcium intake causes osteoporosis. You might include these variables: amount of milk and cheese, lactose intolerant, ethnicity, supplements purchased, grocery store purchases, vegan/vegetarian. Be sure to also explain how each variable might be connected to the factor of “Calcium intake”.
7. Students work with their partners and then with other pairs who are working on the same factor to complete the chart in question 2.
8. Read question 3 and model how you would answer question 3 for the factor “Calcium intake”.
For example, you might include these variables: digestive system disorders, calcium tablets, calcium rich foods purchased, antacids use, multivitamin and mineral tablets, etc. Be sure to explain how each variable might be connected to the factor of “Calcium intake”.
9. Students work with their partners to complete the chart in question 3.
10. Allow time for students to answer and discuss questions 4 and 5.
11. Students meet with other pairs who are working on the same factor. They share and discuss the charts that they made for questions 2 and 3.

Variables in a Big Data Set

Health insurance	Favorite candy	Antacids purchased
Soft drink purchase	Lactose intolerant	Supplements purchased
Bone fractures	Gluten intolerant	Hours of weight-bearing exercise
Time spent outdoors	Vaccinations	Bone density T-Score
Use of sunscreen	Allergy test results	Shoe size
Skin color	X-ray results	Blood test results
Amount of milk and cheese	Chocolate consumption	Hours of sleep per night
GPS tracking for locations	MRI results	Hours of computer use per day
Vacation location	Favorite color	Time spent on social media
Amount of savings	Eye color	Sports participation
Health related web searches	Ethnicity	Twitter mention of bone density
Hiking	Pets in home	Cycling
Fitness tracker data	Hours of employment	Coffee shop reward card use
Age reached puberty	Type of employment	Grocery store purchases
Age reached menopause	Bedtime	Vegan or vegetarian
Prescription medicines	Waking time	Height
Over-the-counter medicines	Sleep problems	Weight
Health of your children	Depression	BMI (body mass index)
Health of your parents	Fat content of diet	Alcohol consumption
Health of your siblings	Salt intake	Smoking
Birth date	Number of children	Current address
Blood type	Health problems	Brain scans
Hours of sun exposure	Vitamins purchased	Tanning parlor use
Chores/jobs	Time spent sitting	Yearly family income

Mining Big Data: Osteoporosis

Teens and Osteoporosis

Experts are concerned that the lifestyle choices that today’s teens are making may result in an “osteoporosis epidemic” when they reach old age. The greatest amount of bone building occurs before age 18. Therefore, the best time to reduce the risks of osteoporosis and low bone density comes during the teenage years.

The things that teens do during their bone building years may affect their risk for osteoporosis when they get older. Unfortunately, very little research has been done to provide evidence that specific teen lifestyle choices increase or decrease the risks for developing osteoporosis later in life.

Analyzing the factors that increase a teen’s risk for developing osteoporosis is complicated because:

- Researchers need to collect data for individuals over many years (from teens to old age).
- Osteoporosis is likely due to complex interactions between many different variables.

Data mining may be used to overcome these challenges. **Data mining** is a process used to turn the information in big data sets into useful information. Data mining involves using computer software to look for associations, patterns, and trends in large data sets.

1. Select one of the four **factors** below that you would like to study to determine whether it affects teens’ risks for developing osteoporosis. Circle the factor that you selected.

Vitamin D / UV light	Exercise	Caffeine	Sleep
----------------------	----------	----------	-------

2. The **Variables in a Big Data Set** sheet shows some variables (types of data) in a big data set. You have been asked to “mine” the data set to select variables that will help you determine if the factor you circled affects a teen’s risk for developing osteoporosis later in life/
 - Select 5 variables (types of data) that might be related to the factor you circled.
 - Give a brief explanation for why each variable might be connected to the factor you chose.

Variables In the Big Data Set	Explanation for how the variable might be connected to the factor that you chose.
	<i>Student answers will vary but most should show a logical relationship to the factor being investigated</i>

3. It is likely that the **Variables in a Big Data Set** does not contain all of the possible data related to the factor that you chose to study. There may be additional variables (types of data) that you would like to add to the data set. *Hint: These may be completely new types of data or more specifics related to your answers in question 2.*
- List 5 new variables (types of data) that might be related to the factor you circled.
 - Give a brief explanation for why each variable might be connected to the factor you chose.

New Variables	Explanation for how the variable might be connected to the factor that you chose.
	<i>Student answers will vary but most should show a logical relationship to the factor being investigated</i>

4. Explain why a background in computer science would be very important for scientists who do big data research on human health issues.

The amount of data available for processing is so gigantic that computer hardware and software is essential for collecting and analyzing big data.

5. Explain why a background in mathematics and statistics would be very important for scientists who do big data research on human health issues.

Mathematics and statistics are essential for understanding associations, trends, and patterns.