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.

3. It is likely that the **Variables in a Big Data Set** does <u>not</u> 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.

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

_____

_____

_____

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

_____

_____

_____