Data — Progress Check

Name: _____ Period: _____ Date: _____

Choose the best answer for each question. One question asks you to select *two* answers.

1. Which best describes *information* in the context of data?

(A) The raw, unprocessed values collected by a sensor
(B) The collection of facts and patterns extracted from data
(C) The hardware used to store a data set
(D) Data about data, such as a file's creation date

2. A data set shows that towns with more firefighters also have more fires. What can you correctly conclude?

(A) There is a correlation, but it does not prove one thing causes the other
(B) The data must be wrong
(C) Firefighters cause fires
(D) Fewer firefighters would mean fewer fires

3. *Metadata* are best described as:

(A) A visualization of a data set
(B) The most important data in a data set
(C) Data about data, such as a photo's size or creation date
(D) Data that has been cleaned

4. Users typed a state name into an open field, producing `NY`, `New York`, and `ny`. Making these uniform without changing their meaning is called:

(A) Encrypting the data
(B) Sampling
(C) Cleaning the data
(D) Compressing the data

5. A team's results are biased. Which statement is correct?

(A) Bias only occurs in small data sets
(B) Bias cannot exist in digital data
(C) Collecting more data will always remove the bias
(D) Bias is often caused by the type or source of the data and is not removed just by collecting more

 CC BY-NC 4.0

6. Which is a reason to use a *visualization* (chart, diagram) when analyzing data?

(A) It permanently deletes the original data
(B) It communicates patterns and insight more clearly than raw numbers
(C) It guarantees the data has no bias
(D) It encrypts the data for security

7. *Crowdsourcing* is best defined as:

(A) Compressing large files before sending them
(B) Storing data across many servers for redundancy
(C) Cleaning a data set automatically
(D) Obtaining input or information from a large number of people via the Internet

8. How can bias enter a computing innovation?

(A) It cannot — computers are always neutral
(B) Only when the data set is too small
(C) Only through hardware failures
(D) Through biases written into its algorithms or present in the data it uses

9. Which *two* are challenges of processing real-world data sets?

(Select two answers.)

(A) Very large data sets may require parallel systems to process
(B) Metadata makes data impossible to organize
(C) Data always fits easily on a single computer
(D) Data may be incomplete, invalid, or not uniform

10. *Citizen science* refers to:

(A) Research conducted in whole or part by distributed individuals who contribute data, many of whom are not scientists
(B) The metadata attached to a scientific data set
(C) Science performed only by professional scientists in a lab
(D) A method of encrypting scientific data

CC BY-NC 4.0