

## Answer Key: Data Mastery: Conquer the Information Age in 11th Grade

Analyze complex visual patterns and identify algorithmic bias to make smarter decisions in policy, tech, and marketing careers.

**1. A public health researcher tracks a disease outbreak using 'proxy data' such as an increase in searches for 'fever' rather than clinical reports. What is the primary risk of using this data source?**

**Answer:** B) Correlation does not imply causation; searches may spike due to media coverage.

In 11th-grade data literacy, analyzing the 'why' behind data is key. Search spikes can be caused by curiosity or news cycles (confounding variables) rather than actual illness.

**2. When building a machine learning model, using a 'representative sample' means that the demographics of the training data should match the demographics of the population it affects.**

**Answer:** A) True

If a sample is not representative, it creates algorithmic bias, leading to skewed results that can unfairly impact specific groups in the real world.

**3. In the context of data ethics and privacy, the process of removing personally identifiable information (PII) from a dataset is known as \_\_\_\_\_.**

**Answer:** B) Anonymization

Anonymization is a critical step in data management to ensure that individuals' identities are protected while the data remains useful for research.

**4. You are examining a 'Heat Map' used by a city council to show crime rates. If the map uses very large geographic blocks, it might mask high-crime 'hotspots' or exaggerate safety in others. This is an example of:**

**Answer:** A) The Modifiable Areal Unit Problem (MAUP)

Higher-level data literacy involves understanding how the scale of aggregation (how data is grouped) can drastically change the visual interpretation of the same facts.

**5. Data 'veracity' refers specifically to the speed at which new data is generated and transmitted across a network.**

Name: \_\_\_\_\_

Date: \_\_\_\_\_

**Answer:** B) False

False. 'Velocity' refers to speed; 'Veracity' refers to the uncertainty, quality, and trustworthiness of the data.

**6. A specialized software tool used to store, manipulate, and analyze geographic or spatial data is called a \_\_\_\_\_.**

**Answer:** B) GIS (Geographic Information System)

GIS is a fundamental tool in data literacy for career paths like urban planning, environmental science, and logistics.

**7. Which of the following describes 'Survivor Bias' in data interpretation?**

**Answer:** B) Focusing only on the people or things that made it past a selection process.

Survivor bias occurs when you study successful outcomes while ignoring the 'silent' data of failures, leading to overly optimistic conclusions.

**8. Open Data initiatives are projects where governments and organizations make their datasets freely available for anyone to use and republish.**

**Answer:** A) True

Open Data promotes transparency and innovation by allowing citizens to find and use data from public institutions.

**9. The 'unit of analysis' in a study refers to the \_\_\_\_\_ unit that is being investigated (e.g., an individual, a city, or a country).**

**Answer:** C) Smallest

Defining the unit of analysis is a critical first step in data research to ensure the conclusions match the level of the data collected.

**10. When evaluating a data visualization, which feature is most likely used to 'lie' or mislead the viewer without changing the raw data values?**

**Answer:** B) Truncating the y-axis (not starting it at zero).

Truncating an axis can exaggerate small differences, making them look much more significant than they actually are.