**R4: The Virtuous Cycle of a Data Ecosystem**

**Quiz Type:** Graded Quiz

**Shuffle Answers:** Yes

**Time Limit:** No Time Limit

**Multiple Attempts:** Yes

**Score to Keep:** Latest

**Attempts:** 3

**View Responses:** Always

**Show Correct Answers:**

From {1 MINUTE AFTER DUE} at 12:02am to {ONE WEEK AFTER} at 12:01am

**One Question at a Time:** No

**Require Respondus LockDown Browser:** No

**Required to View Quiz Results:** No

**Webcam Required:** No

Complete the following reading(s) and do your best to think critically while doing so:

*The Virtuous Cycle of a Data Ecosystem*, by yours truly. (Direct PDF download here.)

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005037

https://github.com/NeuralDataScience/Readings/blob/main/Voytek-PLOS\_Computational\_Biology2016.pdf

After you've completed the reading, complete the reading quiz. These quizzes will help hold you accountable for reading the assigned papers and articles. Then, the discussions in section are meant to really help to ensure you're able to discuss these papers and their nuances with classmates, professors, and any data scientists you work with in your future careers.

**Question 1 (1 pt)**

In the context of this article, what is the Data Ecosystem?

***Correct***

An environment wherein may different datasets are openly shared for reanalysis and remixing.

***Incorrects***

A journal wherein researchers can publish their datasets.

A software platform for integrating heterogeneous datasets.

An organic community of data scientists that interact with one another openly.

**Question 2 (1 pt)**

What is the primary argument here, in favor of data sharing among scientists?

***Correct***

Data can be remixed and reanalyzed in new ways to further scientific discovery.

***Incorrects***

It improves replication and reproducibility.

It is ethical to share data funded by the public.

Storage and hosting costs are rapidly dropping.

**Question 3 (1 pt) (matching)**

Match the concepts below:

***Pairs***

Reanalyzing old data using new methods:

If data are locked away they can only analyzed once, by the original research group, limiting their future potential.

Text mining for scientific discovery:

Large amounts of unstructured freeform text can be leveraged for scientific research.

Data remixing and combination:

Combining heterogeneous datasets—text, images, video, spatial, etc.—can facilitate novel scientific discoveries.

Semi-automated, or algorithmic, hypothesis generation:

Aligning multiple datasets allows for finding gaps in information, which can be leveraged to generate entirely new research paths.

***Foils***

Integrating many heterogeneous datasets minimizes bias.

Peer-reviewed summaries of research results are statistically combined to find evidence of *p*-hacking.

**Question 4 (1 pt) (multiple answers)**

How does the author argue that sharing data can potentially reduce bias in the scientific literature? (Select all that apply.)

***Corrects***

Shared data can be analyzed to identify sampling biases.

Sharing data can reduce cultural bias in interpreting the results.

***Incorrects***

Sharing data provides estimates of statistical power for future studies.

Shared open data tend to be cleaner, and more carefully curated, than closed data.

**Question 5 (1 pt) (essay question)**

*“By basing the decision of whether to share data or not, solely on whether replication and reproducibility is worth the cost of curation and storage, we are limiting the opportunities for future scientists to make novel use of our data in ways that we could never predict. By sharing the raw data, we can create a virtuous cycle that allows researchers to remix and reanalyze data in new and interesting ways. It is our duty to preserve our data so that future generations will not be hindered by our prejudiced interpretations and analytical limitations.”*

Given what you've learned so far in this class, what is your take on this perspective? (This will be graded very permissively: write something honest and thoughtful, and you'll get the 1 pt. Be honest; don't just answer what you *think* I want to hear, or try to just accept my perspective. I truly want to hear what you think!)

Your Answer: