

# Scientific Programming & Workflow Management: Final Assessment

**Instructions:** Please select the single best answer for each of the following questions. You have 60 minutes to complete the quiz.

---

## Section 1: Foundations of Scientific Programming & R

**1. The "Problem Sorter" task distinguished between scientific and general-purpose programming. Which statement best summarizes the core mission of a scientific programmer?**

- a) To build complex, scalable software applications for millions of users.
- \*b) To use a computer as a lab instrument to conduct transparent, reproducible, and verifiable experiments on data.
- c) To create visually appealing user interfaces and websites for presenting data.
- d) To write code that runs faster than any other programming language.

**2. In the debugging module, the "red error text" was described as your friend. What is the most effective professional mindset a researcher should adopt when their code fails?**

- a) Assume the software is broken and try restarting the program.
- b) View the error as a personal failure and avoid asking for help.
- \*c) Carefully read the error message as the first and most important clue to diagnosing the problem.
- d) Immediately delete the code and try to rewrite it from memory.

**3. What is the crucial difference between the `install.packages("ggplot2")` and `library(ggplot2)` commands in R?**

- \*a) `install.packages()` downloads the package from the internet to your computer, while `library()` loads the package's functions into your active R session.
- b) `library()` downloads the package from the internet, while `install.packages()` loads it into your session.
- c) Both commands do the same thing, but `library()` is the more modern syntax.
- d) You only need to use `install.packages()`; the library is loaded automatically.

**4. A vector in R has a strict rule that was demonstrated in the "Organizing Your Samples" activity. What happens if you try to create a vector like `new_vector <- c(5, "hello", TRUE)`?**

- a) R will produce an error because the data types are mixed.
- \*b) R will convert all elements to the most flexible type (in this case, character), resulting in `c("5", "hello", "TRUE")`.
- c) R will store each element with its original data type inside the vector.
- d) R will drop the non-numeric elements, resulting in a vector containing only the number 5.

**5. What is the primary advantage of writing your own function in R, as demonstrated in the "Building Your Own Instruments" module?**

- a) It makes your code run significantly faster than using a `for` loop.
- \*b) It allows you to create a reusable block of code, which makes your analysis more modular, readable, and easier to update.
- c) It is the only way to perform mathematical calculations in R.
- d) It automatically creates a plot of your results.

**6. According to the "Grammar of Graphics," `ggplot2` builds plots in layers. Which of the following contains the three essential components for any basic plot?**

- a) A title, a legend, and a color scheme.
- \*b) A dataset (`data`), aesthetic mappings (`aes`), and at least one geometric layer (`geom_`).
- c) A `for` loop, a function, and a data frame.
- d) `install.packages()`, `library()`, and the `plot()` function.

## **Section 2: Data & Workflow Management**

**7. Best practices in data management require separating raw and processed data. What is the most important reason for keeping your original raw data in a "read-only" state?**

- a) To save hard drive space, because raw data files are always smaller.
- \*b) To ensure you can always return to the original, untouched data, guaranteeing the integrity and reproducibility of your entire workflow.
- c) To make the data easier to share with collaborators.
- d) To prevent R from loading it, as R can only read from a "processed" folder.

**8. Your project folder contains a file named `data-final-revised-v2.csv`. Which data management principle does this file name violate?**

- \*a) It is not descriptive and does not clearly state its contents or origin, making it hard to understand later.
- b) It correctly uses the CSV format, which is a good practice.
- c) It is too long and will cause errors when loading the file in R.
- d) It should be stored in a `scripts/` folder, not a `data/` folder.

**9. In the "Unannounced Audit" module, your group critiqued a set of confusing instructions. This was an analogy for a poorly documented analysis. What is the definition of a scientific workflow?**

- a) The specific R script that produces the final plot for a paper.
- \*b) The entire, end-to-end process from raw data to final conclusion, documented so clearly that anyone can follow it.
- c) A list of all the R packages needed for an analysis.
- d) The process of submitting a paper to a journal for peer review.

**10. What is the primary function of a version control system like Git?**

- a) It automatically debugs your code and fixes any errors.
- \*b) It acts as a "logbook," tracking every change made to your project files, noting who made the change, when, and why.
- c) It is a tool for creating high-quality data visualizations.
- d) It encrypts your project files to prevent others from seeing them.

**11. What is the key advantage of using a reproducible reporting tool like R Markdown?**

- a) It automatically writes the narrative text for your report based on your code.
- b) It prevents you from making any errors in your R code.
- \*c) It combines your code, its results (plots, tables), and your narrative into a single document that can be automatically updated if the data or code changes.
- d) It allows you to use Microsoft Word to write your R code.

**Section 3: Principles and Protocols**

**12. The FAIR principles (Findable, Accessible, Interoperable, Reusable) are a cornerstone of modern data management. What does "Interoperable" mean in practice?**

- a) The data is stored in a secure, password-protected location.
- \*b) The data uses standard formats and vocabularies, allowing it to be easily combined with other datasets and read by different software.
- c) The data is published in a high-impact journal.
- d) The data is easy to find through a Google search.

**13. The ODD protocol is used for documenting a specific type of scientific work. Which of the following would be most appropriately described using ODD?**

- a) The statistical analysis of a clinical trial.
- b) The results of a chemistry experiment.
- \*c) An agent-based model simulating forest ecology.
- d) The assembly of a new genome.

**14. What is the central idea behind the TRACE protocol?**

- \*a) To document model credibility through a transparent and comprehensive evaluation of its purpose, assumptions, and validation.
- b) To ensure a project uses at least four different programming languages.
- c) To provide a standardized way to request funding for a research project.
- d) To track the physical location of lab samples using GPS.

**Section 4: Ethics in Science & Open Science**

**15. In the "Sky Shield" geoengineering scenario, a private foundation from a wealthy nation considers an experiment that could negatively affect weather patterns in poorer nations. This highlights the core concern of which ethical viewpoint?**

- a) The Scientific Freedom Advocate, focused on the need for unrestricted innovation.

- b) The Universal Ethics View, focused on applying the Belmont Report principles.
- \*c) The Justice-Oriented View, focused on how research can perpetuate structural inequalities and power imbalances.
- d) The Contextual Ethics Perspective, focused on how consent norms differ between cultures.

**16. Imagine a scenario where researchers adapt their informed consent process to align with a community's collective decision-making, rather than requiring individual written consent from every participant. This approach is most representative of which ethical view?**

- \*a) The Contextual Ethics Perspective.
- b) The Universal Ethics View.
- c) The Justice-Oriented View.
- d) The Scientific Freedom Advocate.

**17. The "Nightingale Pre-Print" case study, where an early-career researcher's work was "scooped" by a better-funded lab after she shared it openly, is a perfect illustration of the concerns raised by which perspective on Open Science?**

- a) The Innovation Catalyst View.
- b) The Democratization Perspective.
- c) The Quality and Transparency Advocate.
- \*d) The Researcher Protection Critique.

**18. A proponent of the "Democratization Perspective" of Open Science would be most passionate about which of the following goals?**

- a) Ensuring that all published research is 100% reproducible.
- b) Protecting the intellectual property of researchers to encourage commercialization.
- \*c) Making scientific knowledge freely accessible to eliminate barriers for researchers in the Global South and at underfunded institutions.
- d) Accelerating the pace of discovery by allowing large, well-funded labs to rapidly build on new ideas.

## **Section 5: Collaboration and Professional Skills**

**19. In the "Sandwich Review" exercise, the "Bad Review" was unhelpful. What is a core principle of giving good, constructive feedback?**

- a) Focus your critique on the personality of the person who did the work.
- b) Be as vague as possible to avoid hurting the person's feelings.
- \*c) Start with positive aspects, be specific in your critique, and focus on the work itself, not the person.
- d) Only provide feedback if you can prove the work is objectively wrong.

**20. During the "Project Clinic" (Peer Consultation) exercise, what was the specific rule for the "Open Brainstorm" phase?**

- a) The case giver leads the discussion, correcting any wrong ideas.

- b) Only the moderator is allowed to speak.
- \*c) The case giver must remain silent, allowing the consultants to brainstorm freely without influence or interruption.
- d) Everyone must write their ideas on sticky notes instead of speaking.

**21. When engaging in a peer review of a colleague's code, you see a variable named  $x$ .**

**According to the principles of clean code, what is the best feedback you can give?**

- a) "You should not use the variable  $x$ ."
- b) "This code is confusing."
- \*c) "I don't understand what the variable  $x$  represents. Could we give it a more descriptive name so the code is clearer?"
- d) "My code is much better than this."

**22. In the Project Management module, the concept of "hats" was introduced. If a team member is focused on ensuring the project is on schedule and meeting its overall goals, which hat are they wearing?**

- a) The Data Engineer.
- b) The Analyst.
- c) The Storyteller.
- \*d) The Coordinator.

**23. Why is it important to work backwards from a final deadline and set intermediate milestones for a research project?**

- \*a) It makes a large, overwhelming project feel more manageable and helps ensure steady progress.
- b) It is a requirement for using the Git version control system.
- c) It guarantees that you will not encounter any unexpected bugs or problems.
- d) It allows you to skip the data cleaning and documentation steps.

**24. The 4Rs of reproducibility are Repeatability, Reproducibility, Replicability, and Reusability. Which 'R' describes a different team using the *same* algorithm and data to get the same results in a *different* environment?**

- a) Repeatability.
- \*b) Reproducibility.
- c) Replicability.
- d) Reusability.

**25. A key message of the entire course is that modern science is collaborative. What is the main purpose of adopting standardized tools and practices like Git, R Markdown, and clear data management?**

- \*a) To make your work understandable, transparent, and trustworthy for your most important collaborator: your future self, and for others.
- b) To ensure that only scientists with advanced programming skills can participate in research.

- c) To create a single, universal programming language that all scientists must use.
- d) To automate every aspect of the research process, removing the need for human scientists.