Review: OSEMN (pronounced “Awesome”)

1. What does the acronym OSEMN stand for in the data science pipeline?

2. Which stage of the OSEMN framework typically takes up 80% of the work in data science?

3. What is one reason the OSEMN data science pipeline was described as a "pipe-mess"?

4. What are some common tasks in the “Scrub” step of the data science pipeline?

5. What are some common activities in the “Explore” stage of the OSEMN pipeline?

6. Give an example of a modeling technique from the OSEMN pipeline.

7. Why is the command-line considered “agile” for data science?

8. How does the command-line support scalability in data science workflows?

9. What makes the command-line “extensible”?

10. List three reasons why the command-line is considered ubiquitous in data science.

Bonus question: which fun Unix programs were introduced? Name at least one!