**OPTION A: RESEARCH PROPOSAL**

**Instructions**

Submit a compiled PDF report to Gradescope. Excluding references, the report should be at most 1200 words. Include all figures in the main text.

**Description**

In this option, you will prepare a short, accessible article describing a recent research paper on the statistical analysis of multi-omics data. The article should target a general audience (e.g., an intellectually curious college senior) and be similar in spirit to the reports that appear in venues like [eLife](https://elifesciences.org/articles/insight), [Nature](https://www.nature.com/research-analysis) , or [Quanta](https://www.quantamagazine.org/) magazines.

Structurally, your article should include a main text and an explanatory illustration. In terms of content, the main text should provide some background about the research community surrounding the selected paper and the specific problem that the paper addressed. You should also include caveats to prevent the misinterpretations a general audience might fall into. Consider closing with some directions researchers are now exploring, given the paper's result. The explanatory illustration should describe the main ideas of the paper using simplified figures and annotation. For this assignment, a hand-drawn figure is sufficient (in the real world, you could send your drawing to a professional illustrator).

I encourage you to be creative with this assignment. For example, many articles in eLife or Quanta have interesting opening "hooks" (specific images or descriptions of people that catch our attention) or valuable metaphors. You could even make up quotes from researchers, to make it seem more like a genuine article.

This is a playful exercise, but I have more serious reasons for assigning it. First, researchers need to be able to explain their work to audiences that are not immediately familiar with their topic — this is important in grant applications and seminar presentations, for example. Second, the long-term vitality of a field depends on how well it can spark the interest of beginners. Finally, to understand the essence of a research problem, it is helpful to step back in the way this exercise encourages.

**Rubric**

Depth [10 points]: Fully developed discussion of the relevant research community and results.

Organization [10 points]: Includes a logical opening, development, and conclusion.

Accessibility [5 points]: Discusses the topic in a way that is accessible to a general audience.

Illustration [5 points]: Provides an appealing and self-explanatory illustration of the research.

**OPTION B: RESEARCH PROPOSAL**

**Instructions**

Submit a compiled PDF report to Gradescope. Excluding references, the report should be at most 1200 words. Clearly distinguish the different prompts (e.g., “Motivation [1]:” for the first group in the motivation section).

**Description**

The most interesting statistical research projects often go through a long reflection period before a single line of code (or math) is written. This exercise walks you through some of the types of planning that can be helpful when structuring a new research project. If you are currently working on a research project, you can use this exercise to analyze the motivation and approach of your project. If this is your first time working in research, you can pick an area of interest and use this exercise to explore it systematically.

Your submission will be a write-up with the following structure. First, prepare a single-sentence description of the problem your research project should address. Then, prepare short responses to the prompts below. Note that we do not require you to propose a specific solution strategy in any question. Instead, the emphasis is on clearly articulating the problem and describing the existing literature from which you could draw potential solutions.

*Motivation*

1. Why does your proposed multi-omics data analysis problem matter to the community? What are the most effective existing approaches? Consider briefly describing existing papers and R packages that produce artifacts similar to your intended output.
2. Imagine a simple data analysis highlighting a critical limitation of existing approaches. For example, this can arise in a failure to respect inferential guarantees, poor prediction performance, or inadequate goodness-of-fit.

*Approach – Conceptual*

1. Once you have a well-defined problem, you should be able to draw from your statistics vocabulary to identify potential directions for a solution. What are the most important statistical/computational concepts that you can draw from in your problem-solving approach? Can you give a common-sense explanation for why these ideas would lead to an improvement over current methods? Describe 1 - 2 key references.
2. Imagine preparing a 30-minute live-coding demo for computationally savvy graduate students in the project's application domain. What dataset would you use in your example? What code snippets would you share? What would be the key takeaways?

*Approach – Implementation*

1. What evaluation techniques would you use to argue that any new approach is an improvement over the state-of-the-art? These evaluation criteria can be either quantitative (e.g., error rates, computation time) or qualitative (e.g., interpretations facilitated by your method).
2. What is one simulation or data analysis setting that you could use to implement the evaluation method described above? What are the proposed setup's potential weaknesses, and how could you mitigate them? If you choose a simulation, what would make it a convincing proxy for reality? If you select a data analysis, what would demonstrate your method's effectiveness?

If any of these prompts don't fit your current project, you can send me an email, and we can brainstorm substitutes.

**Rubric**

Motivation [10 points]: Fully developed discussion of the research context.

Specificity [10 points]: Gives examples and in-depth discussion for the proposed approach.

References [10 points]: Demonstrates familiarity with relevant research literature.