

**CMPSC 300 and BIO 300
Introduction to Bioinformatics
Spring 2016**

Janyl Jumadinova and Kristen Webb

<http://cs.allegheny.edu/sites/jjumadinova/300>

**Course Project
Given: February 26, 2016**

Summary

Throughout the semester, you have used various computational tools to explore a number of biological questions. The final project invites you to explore, in greater detail, a real-world application of bioinformatics. You will research in more depth a real-world bioinformatics project that is interesting to you and carry out a computational investigation through the use, implementation, testing, and evaluation of different types of software.

For your final project, you can work individually or in groups of two or three. If you decide to work in a group, each member of the group will be evaluated separately based on his or her contributions to the project. This evaluation will be determined largely from the feedback of the group members. If you work in a group, you have to create a new project repository, named `300s2016-groupName`, that can be shared between your team members and your instructors.

Assignment Specifications

For the project assignment you may choose a biological research question that is of the most interest to you and can be investigated through computational technique(s). Pick something realistic and useful, for example, you can choose specific question(s) related to the topics we discussed during the first week of the course. You may use anything and everything we have learned (or will learn) in class and also you should research additional resources beyond of what we discussed in class. You may also extend any of the programs or concepts we have developed in the labs or in class. However, the problem that you choose should not just be a copy of one of the lab assignments, or the class exercises, or the programs in the book with slight modifications. Your project must be extensive enough to qualify as a project (think of work for at least 5 one-week lab assignments), but not too extensive so that you can not finish it by the due date. Remember, you must adhere to the Honor code! Please be original!

Requirements

1. Select a real biological question(s) to investigate.
2. Research relevant background.
3. Identify and utilize computational techniques for answering your question(s) (you may use web tools and/or develop programs).

Timeline: Deliverables

1. **Proposal** (at least one page) **Deadline: Friday, March 11, 2016 by 2:30pm:**
Develop an idea for your project including preliminary research on the importance of the problem and data availability (e.g., database of SNP data from GWAS). Discuss your project idea with your instructors during the lab session on March 4. After receiving a verbal approval, write a one page description of what you propose to do for your project and share it with your instructors through your Bitbucket repository. Your proposal should include at least three references that motivate the importance of the problem. You do not need to include any specifications on how exactly you will solve your proposed problem at this point, however you should discuss potential tools or algorithms you maybe able to utilize for your project.
2. **Progress report** (3-4 pages) **Deadline: April 8, 2016 by 2:30pm:**
Describe everything you have done so far in your progress report. By this point, you should have conducted necessary research on the background of the problem, decided on the approach you will use to solve it, and made a significant progress towards implementing the solution to your proposed project. Describe anything new that you have learned so far and any unexpected challenges that you have encountered.
3. **Presentation** **April 29, 2016, during the lab session (another date maybe added depending on the number of projects):** In the presentation, you should describe the motivation, problem definition, challenges, approaches, and results and analysis. Use diagrams and a few bullet points rather than long sentences and equations. The goal of the presentation is to convey the important high-level ideas and give intuition rather than be a formal specification of everything you did. Prepare for ~ 10 minute presentation. Design at least five slides, including a slide with the title of your project and group members' names. Every member of the group needs to contribute to the presentation talk. At the end of the presentation give a demonstration of your project.
4. **Final report** (6 or more pages) **Deadline: May 5, 2016 by 7pm:** Incorporate any feedback from the progress report and the presentation session. Your final report should be clearly and well written, this includes no typos or grammatical errors. Your report should be written in a professional manner. Your report should include:
 - The motivation for your project. Why is the question you decided to address important and useful?
 - Background for the proposed biological investigation. What have others done for it already? Include references.
 - Detailed description of the work you completed for this project. Provide algorithms if necessary.
 - Analysis of your results. Make graphs, tables, snapshot of output, or anything else that can help us understand your results.
 - Conclusion. Give a short overview of your project and its results. Describe what you learned, what were the biggest challenges and the biggest rewards.

Grading rubric

15 points: **Proposal**

20 points: **Progress report**

25 points: **Presentation and demonstration:**

40 points: **Final report and project implementation:**

10 points: **Extra credit:** does the project present interesting and novel ideas (i.e., would this or its extension be publishable at a conference)?

For each deliverable, you need to submit a PDF with your report (or presentation slides), for your final report you need to submit any supplementary material (programs, data, a README file documenting what everything is, and how to run your program) via your repository.

In adherence to the Honor Code, students should complete this assignment while exclusively collaborating with the other member of their team. While it is appropriate for students in this class who are not in the same team to have high-level conversations about the assignment, it is necessary to distinguish carefully between the team that discusses the principles underlying a problem with another team and the team that produces an assignment that is identical to, or merely a variation on, the work of another team. Deliverables from one team that are nearly identical to the work of another team will be taken as evidence of violating Allegheny College's Honor Code. Do not be tempted to look online for possible problems and solutions, that institutes a violation to the Honor code! Please be original!