

CMPSC 300 Bioinformatics Fall 2019

Project Assignment



Figure 1: Bioinformatics contains many different types of research.

GitHub Starter Link for Groups

STOP! STOP!

Not everyone will be clicking this link at this time!

Only the team leader will be clicking the link to create the repository!!

<https://classroom.github.com/g/7h96fjD7>

Creating your repository

We will use a group assignment functionality of GitHub Classroom for this assignment. For group assignments **only one person will be creating the team while the other team members will join that team**. Please form a team of **no more than three people** and select one person to create the repository.

The selected person of the team should go into the link of the assignment sheet. Copy this link and paste it into your web browser. Now, you should accept the assignment and create a new team with a unique and descriptive team name (under the field marked, “Create a new team.”)

Now the other members of the team can click on the assignment link and select their team from the list under “Join an Existing Team.” When other team members join their group in GitHub Classroom, a team is created in our GitHub organization. Every team member will be able to push and pull to their teams repository.

Please work in groups and leave no-one out of the loop: Working in groups is one of the best ways to add focus to your project and to develop ideas (by brainstorming.) This focus can only happen when you include everyone in your group. Please do not leave any group member behind in your thinking and planning of your project.

To push your changes, you can use the following commands to add a single file, you must be in the directory where the file is located (or add the path to the file in the command):

HANDED OUT: 4th NOVEMBER 2019

- `git commit <nameOfFile> -m 'Your notes about commit here'`
- `git push`

Alternatively, you can use the following commands to add multiple files from your repository:

- `git add -A`
- `git commit -m 'Your notes about commit here'`
- `git push`

Reading Assignment

To help you work, please read the covered chapters from the course book, consult the week's slides and your class notes. You can also find useful information in the Python community by performing online research for programming help. Please take some time to gain experience with using Markdown to complete your writing assessments. **Your written deliverables (are to be written in Markdown.**

See *Mastering Markdown* <https://guides.github.com/features/mastering-markdown/> for more details about Markdown. Another good reference may be found at: <https://markdown-it.github.io/>.

Be sure to read the `README.md` file in the GitHub Classroom repository for instructions on how to complete your first assignment.

Summary

Throughout the semester, you have studied many of the concepts listed in Figure 1, and have also used various computational tools such as alignment and BLAST to explore a number of biological phenomena. 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. Please be sure to include all names of your group members in each deliverable that you submit. Remember, groups are up to three (3) members.

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 3 or 4 one-week lab assignments), but not too extensive so as not to be completed 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: Tuesday, 12th November 2019 by 2:30pm**; Develop an idea for your project including preliminary research on the importance of the problem and the availability of the data that your group will require to complete the work. For your data, please be sure to use publicly available data. Discuss your project idea with your group (and your instructor as necessary) to build on your original ideas and to brainstorm. Your proposal should include at least three references from academic peer-reviewed articles to provide foundation and motivate the importance of the problem. *Please do not include unsubstantiated resources to your list of citations such as websites, blogs, YouTube videos, popular news articles and similar.* 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. For these items, consider adding separate references to academic articles which have already incorporated them into projects.
2. **Progress report** (3-4 pages) **Deadline: Tuesday, 19th November, 2019 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. **Class presentation** (beginning on) **Monday, 9th December, 2019, during the lab session**; 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 is to contribute to the presentation talk. At the end of the presentation give a demonstration of your project.

Note: since your report is due *after* your talk, your talk does not need to conclude the project, rather it should provide a very informed series of events to describe your work.

4. **Final report** (Six or more pages of text, more pages will be necessary with included graphics) **Deadline: Friday, December 13th, 2019 by 9:00am**; Incorporate any feedback from the progress report and the presentation session. Your final report should be clear, concise and, most importantly, well written, this includes no typos or grammatical errors. Your report should be written in a professional manner and should include the following main sections.

- 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. Please provide code and algorithms as necessary.
- Analysis of your results. Make graphs, tables, snapshots of output, or anything else that can help the instructor understand your results and your interpretation.
- Conclusion. Give a short overview of your project and its results. Describe the potential impacts of having this knowledge, discuss the biggest challenges and the biggest rewards of your work.

5. **Ethical Component** (1 page, to be added to its own section of your final report document): You are to find two extra peer-reviewed research articles which have a similar theme to your own work. Discuss /introduce the methods, and then respond to the following questions of ethical responsibility concerning the main technologies and /or methodologies discussed in the two articles.

- Who created the technologies or methods which runs along a similar theme to your own project?
- Who are the users of the technologies or methods?
- Who is not supposed to use the technologies or methods?
- How could the the technologies or methods can cause harm?
- What solutions do you think should be in place to avoid or fix the harm caused by the technologies or methods?

Grading rubric

15 points: **Proposal**

20 points: **Progress report**

25 points: **Presentation and demonstration**

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

20 points: **Ethical component**

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

Submission Guidelines

For each deliverable, you are to submit a Markdown file of your report (or presentation slides). For your final report you are to submit any necessary and supplementary material. This includes programs, data sets, a *README.txt* file documenting what everything is (i.e., a justification of the existence of the files that you have left for the instructor in your repository). Finally, for your code, you will need to write up documentation to instruct how the code is to be used and what its expected inputs and outputs should be. Please write up these details so that they will be useful to the instructor.

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!