For our first major class assignment, you will undertake a project tailored to you - working on/improving at least two statistical or communication skills identified through reflection on your portfolio of your statistical learning/progress so far (at least one of the skills must be statistical). The idea is that everyone in class has different strengths/weaknesses and different things they want to work on. This project is your chance to do that (and you can continue that in our other projects as well). By having you pick the goals and create the assignment, you take ownership of your own learning.

1 Project Process

Here is the outline of the process for this assignment (detailed timeline below):

1. Create your portfolio. (Should already be complete.)
2. Complete the portfolio reflection. Look over your responses. (Handout posted.) You will need to upload this completed handout to Gradescope.
3. Identify 2-4 skills/methods/topics you want to work on for your tailored project.
4. Brainstorm ways you could work on those skills in a three week project. The reflection will help with this. A data set that you can use to demonstrate a variety of things is provided below in case you don't want to find your own data.
5. Draft a proposal/plan for your project, and bring your draft to the discussion (next).
6. Have a discussion with me to share what skills you want to work on and how you plan to work on them. The intent is for me to help you revise/adjust your plan so that it will align with what you want to work on and fit within the time frame. This is a required oral discussion. Sign-ups are available online after class on Sept. 5th. If no posted time works for you due to a conflicting commitment, be in touch with me ASAP. Once someone signs up for a slot, it is taken, but if someone cancels their slot, it should re-open. Bear with me, trying this within Google calendar for the first time.
7. Submit your tailored project proposal/plan after the discussion to create a record and give yourself a road map for the project. I will be in touch if the proposal/plan needs adjusted. The proposal/plan will also help with assessment, as you will need to state what you'd like factored into the project assessment (besides just completion).
8. Complete the project over the following three weeks.
9. Submit the project.

Examples of skills and potential projects are below. These lack context (i.e. yours should be more in-depth), but I wanted to give you an idea of how to connect skills with how you might work on them.

2 Identifying Skills

In the portfolio reflections, you'll see questions about what skills you are practicing. Those lists of skills are NOT exhaustive, and many are broad/vague. It's entirely possible that you conceive of a project that includes a skill not listed there but which is completely fine for the project. One reason for the discussion with me is to help identify these so you are comfortable having them in your plan even if they weren't explicitly listed in the reflections.

3 Potential Data Set

An easily available data set that may be of use to you is the College data set from the ISLR text. Accessing this data set is as simple as (installing and) loading the ISLR library in R and then typing data(College). It has 777 observations of 18 variables. There are few categorical variables, but you could create them from existing numeric variables if needed for your particular project. Its help file in R will have more information.

Feel free to use any data set (with appropriate citation) of interest to you. This one is provided for a quick reference and because a number of potential analyses are possible with it. However, it may not call to you, so you can of course find your own data set!

There is also not a requirement to have a data set for this project. You could do a simulation study, where you are generating the underlying data through a model, or potentially a theoretical project. On the other hand, you could make the project entirely about data, such as generating a Shiny app with beautifully commented code, if that was what you wanted to work on.

4 Examples

4.1 Example 1

Suppose during your portfolio reflection, you felt that you wanted more practice with creating appropriate graphical displays (or maybe making a super awesome visual for something complicated), data wrangling, and communicating results of an exploratory data analysis (EDA).

You'd need to identify a data set that would challenge your wrangling skills (not the provided one above), lay out what type of visuals you'd like to work on, and prepare a write-up of your wrangling and EDA that includes your visuals. You could add a focus on good programming practices if interested as well.

For this, you might consider that you want assessment based on presentation of the wrangling (including good programming practices), explanation of the findings in the EDA, choice and presentation of the graphs, and clarity of the writeup in relation to the graphs. You'd submit a report (pdf), but the .Qmd would also be in your repo.

Bear in mind that project needs to be at an appropriate level for our class, so doing the above but only making a single histogram or just submitting the visuals (no write-up in

addition) is not appropriate.

4.2 Example 2

Suppose during your portfolio reflection, you saw the words "randomization test" and realized you were unsure on this topic (what is it, why would I do one, don't remember this, etc.). This could be paired with regression and hypothesis testing as skills to work on (there are lots of options here).

For this one, use of the College data set might work, unless you wanted to find your own data set.

The project could be an investigation of randomization tests (so a write-up about what they are and when/how to use them) with a demonstration on the College data set in a regression setting doing hypothesis testing. You could compare traditional inference results to those from the randomization procedure, and present all that in a write-up. You'd likely have a report (pdf) and .Qmd like the example above for deliverables.

For assessment, you could imagine it here is based on accuracy and clear presentation of the information about randomization tests, clarity of presentation in the application, and quality of the discussion comparing the two approaches in the write-up.

4.3 Example 3

Maybe looking over your portfolio, you are happy with your data analytic skills, but decide you might want to revisit some STAT 370 topics, say, maximum likelihood estimation (MLE) as an example. Even if that's just one skill item, you can build other skills into the project. (And I'm happy to help you brainstorm about this.)

You could envision a project like the one where we tackled a TAS article. Find an article on MLEs in a journal of your choice (which connects to the skill of reading about methods in literature), prepare a summary of it, with added exposition on what MLEs are (communication of theoretical results) and maybe formal solutions to one or two MLE problems (which I could help you come up with). Here, you might just have the report/write-up as a deliverable.

For assessment, you could imagine quality and accuracy of presentation on MLEs, clarity of the article summary, and presentation and correctness of the example problems as key items.

4.4 Example 4

Suppose looking over your portfolio, you realize that you want to revisit and majorly improve your STAT 230 course project. You could likely identify the relevant skills (depending on what methods it uses), and potentially tie in a communication skill (which would add parts below).

The project would be to redo the analysis, submit an updated project (perhaps with a particular emphasis on some aspect of communication), and a commentary on what was done

differently and why. That latter part is important because it shows where you improved. Here, the deliverable would likely be a pdf of the report and a .Qmd.

You could imagine assessment on the quality of the update, presentation of the updated project, and clarity/correctness of the commentary on the differences.

5 The Proposal/Plan

You'll need to prepare and submit a Tailored Project Proposal/Plan which is approved by me for the project. A template will be provided for you to fill in. The major items you'll need to include are:

1. What skills/methods/topics you want to work on for your tailored project (at least 2, probably more like 3-4).
2. What you envision doing for the project to work on those skills.
3. What your project deliverables are with details. (Is it just a report? What's in the report? Are you submitting multiple files, such as wrangling code? I'm imagining most of you will end up with at least one file to submit (probably pdf of something), but there could be more!)
4. How do you want the project to be assessed? Note that "based on completion" is not an option. I want you to think about how you'd like your work on these specific skills to be assessed. And am happy to brainstorm ideas here with you (even after the plan is submitted, as needed).
5. What is your rough road map/ timeline to complete the project? This is your chance to lay out a plan for yourself to tackle the project.
6. Any questions for me.

As you brainstorm what your proposal is for the project, I recommend reviewing the Examples above.

6 Project Timeline

1. Project Introduced on Friday, Sept. 5th in class
2. Sign Up for a Discussion Appt with Prof. Wagaman after class on Friday, Sept. 5th (15 minutes, in provided slots)
3. Engage in the Portfolio Reflection between class on the 5th and your discussion with Prof. Wagaman (will submit to Gradescope)

4. Review the project proposal/plan document. Bring a draft of your plan to your discussion.
5. Have your Discussion with Prof. Wagaman (most slots on Thursday, Sept. 11th)
6. Submit your Tailored Project Proposal/Plan by midnight Friday, Sept. 12th to Gradescope
7. Work on your tailored project over the next three weeks
8. The Tailored Project is due by midnight, Friday Oct. 3rd. At least one (pdf) file should be submitted to Gradescope, with any remaining deliverables and all your work in your course repositories.
9. Assuming it fits in the schedule, you'll do a short reflection on the project, most likely to see if you accomplished what you wanted to or what else remains that you might incorporate in our other projects.

7 Additional Information

Some other items are important to note for this project (and our future projects).

7.1 Assessment and Organization

First, to complete assessment of this project, I will examine both your submission in Gradescope and pull your repository to examine your work there. This means your repository needs to contain your work on the project, as well as all your deliverables, in an organized fashion. Only one deliverable (a pdf of something) can be uploaded to Gradescope. I should be able to find all items associated with the tailored project easily in the repo. This is easy if you just have a folder for your tailored project work and keep it organized.

Because I'm going to be looking at the repo, it's important that the files have names that make sense. Providing a readme file in the tailored project folder describing what each file is would be even better. For example, if the file "Partofwork" is actually the main submission, that filename doesn't indicate that. You want names like "FinalReport" or "DataWrangling", not "Tuesdaywork". If you have multiple deliverables, I should be able to figure out which repo files belong to (or are) each one.

Additionally, everyone will have the following assessment components in addition to whatever you list on your proposal/plan document:

1. Signed up for Discussion Appt
2. Submitted a completed Portfolio Reflection to Gradescope

3. Came to Discussion Appt Prepared and Engaged in Discussion (Brought a draft plan, able to engage in discussion, able to answer questions about your portfolio reflection - See Oral discussion prep sheet for more information)

4. Submitted Tailored Project Plan on time and completed in full, and

5. Submitted Tailored Project on time.

6. Engaged in Responsible / Reasonable commit behavior (see next below).

7.2 Commit Behavior in Github

The commits in your repository history should show your work, continually, over this project over the three week period (and technically, your work on the draft plan in the week prior). No, this is not something you have to work on every day, but the idea is your repo shouldn't just suddenly show all the files appear in final form in a single (or few) commit(s). Your commits should show your progress over time. If you are used to committing at just the end of a project, it's time to change your habits. That's not how commits are supposed to work (and is dangerous - you could lose work!).

For example, to create this assignment, I have a commit where I copied over a different project assignment (from a past project) to create the file draft, then a commit where I cut a bunch out, not relevant to this project, a third commit where I got it a bit more organized, and then a fourth commit related to putting in the timeline and this section. (And it will have more by the time it's finished, just giving you an idea here - on 8th commit or so now, and may be ready to post it for you all finally).

Your first commit related to the project will likely be pulling the assignment and tailored project plan template file. A second commit might be your draft work on the plan. A third commit might come after your meeting with me, resulting in a pdf you submit as the plan. A fourth commit might be pulling together files you need (from an old project, for example) or adding a timeline for yourself in your repo to help keep you on track. A fifth commit might be starting to work on some piece, with only a tiny bit of work, etc. Be sure you commit and push at the end of a working session!

I'm not going to mandate a specific number of commits, but I do want to be sure we are all practicing responsible/reasonable habits with commits. You should commit often. You can have several commits in a working session before a push. Commits should show the history of your work on the project. All (or many) files appearing at once (or in just a few commits) suggests issues with the intended process and/or submission itself.

If I have concerns re: this process (commits) or in the assessment process (on your submission itself), I reserve the right to have additional oral discussion(s) with you about any or all aspects of the project.

7.3 Making Changes to the Proposal/Plan

You'll submit your proposal/plan and I'll give you feedback so that (hopefully quickly) we arrive at an agreed upon project.

Suppose while you are working on the project that you decide it makes sense to adjust/alter/change something about it. What do you do?

Well, if you are adding something, that is, going beyond something you listed in the accepted proposal/plan, go for it. For example, suppose your plan said you would include two examples of a randomization procedure, but you'd like to include a third example. Go for it. You don't need to run that by me, though you can if you like.

If you are changing/subtracting something, you must talk to me to get it approved (which may mean adjusting your proposal in some way), or it may be construed as not satisfying the project proposal/plan. For example, suppose you said you would include 2 example problems on MLEs, but you've done one but worked really hard on the exposition and think it's amazing. Changing to just one might be reasonable here if the exposition really shines (but I may tell you no or ask you to include assessment of the exposition in an updated proposal/plan). For another example, if you said you wanted to practice logistic regression but are trying to revert back to multiple linear regression, I'm likely to tell you no, without more information. Why? MLR is already practiced in lots of classes, so you'd need to be more specific about what about it you were trying to practice, etc.

Why this structure? Everyone will be working on a different tailored project and I'll be doing my best to keep them at comparable levels (and at an appropriate level for our course). You can add on and go beyond and will still keep the project at a comparable "base" level and appropriate for class no problem. However, some changes or subtractions might jeopardize that, which is why I need to approve them.

Finally, in terms of extensions, since that would be a change to the proposal/plan, if there is any extension on the project, it would be given to the entire class and would be the one additional weekend plus Monday to complete the project. So, a potential due date of Oct. 6th. However, we have both the midterm (and a homework due) and the introduction of our second project that week, so I feel its in the best interests of the class to use the October 3rd deadline for smoother transitions and to give you time to study. I will evaluate this closer to time and with your input, but expect everyone to set their plan for the deadline of the 3rd.

7.4 Generative AI Guidance

There is no requirement to engage with generative AI for this or any course activity.

If you choose to engage with generative AI for any part of this project, remember to follow the class framework. Generative AI may be used only as a thought partner, and only in places that don't jeopardize your learning/our course learning outcomes.

You should immediately export, save, and commit any chats with generative AI about the project as you have them to keep a complete record to submit.

Adding the information to the integrity page as you go is also strongly recommended, rather than trying to do it at the end based on your exports. You could keep a running pdf

addendum of all the chats for easy reference too. Programs that merge pdfs (like Adobe Acrobat) can be useful for that.

Examples of what generative AI might be useful for in the context of the project and that are in line with the class framework are:

1. Asking AI for help coming up with a timeline that works for your proposal.
2. Asking AI for help with a code error, making sure you can understand how it fixed it and could re-explain it later.
3. Asking AI to help you outline a deliverable you describe in your plan.
4. Asking AI to tighten up writing in a particular paragraph or suggest improvements in a section for clarity. (Remember to review it to be sure the AI didn't add anything it shouldn't have!)
5. Asking AI to help you format a deliverable a particular way. (You really wanted a folded handout for some reason.)

Note how these are uses that you have available human support for, without concerns for academic integrity. Me, your peers, and the SLC (Strategic Learning Center) are resources for learning how to make timelines. Google (or any other search engine), me, and your peers are good resources for code issues or formatting challenges. Outlining and writing skills have support in the Writing Center, or you could ask a peer to review a submission.

Contrast that with the uses below, where academic integrity concerns arise. These show issues where the AI use is jeopardizing class learning outcomes / your learning experience. Obviously, the uses below (and others like them) are not allowed.

Examples of how generative AI might be used in the context of the project but that are not in line with the class framework (and therefore, not allowed):

1. Asking AI to fill in the plan template for you (especially if not based on your actual thinking about what skills you want to work on).
2. Asking AI to create a deliverable (partial or entirely) by creating much of the content in it that you describe in your plan. (You should be creating the content.)
3. Asking AI to do the data wrangling for you for the project.

Neither of these lists are exhaustive. Please refer back to the course syllabus or talk to me if you want to engage with Generative AI tools and have questions about appropriate use.