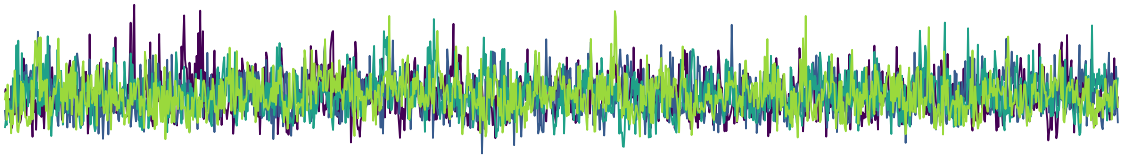
EEEB GR5005
| Graduate Level Research Project

Columbia University, Fall 2021



Research Project Overview

For the graduate level version of this course (EEEB GR5005), all students will prepare a short research paper. The emphasis of this research project is on the data you choose to analyze and your analytic methods. Basically, this project is a scientific paper largely reduced to its analytic core.

I envision the final research project format as follows:

- A short introduction of the relevant background information for the study. This should be roughly a 1-2 paragraph section (not an extensive literature review) ending with your specific project objectives or hypotheses.
- Detailed sections on analytic methods and results. These should be modeled after the methods and results sections of a scientific paper. As such, you'll likely want to include some tables or figures. Similarly, describe your own work with the clarity and rigor you would expect to find in the scientific literature.
- A short discussion on your data analysis (1-2 paragraphs). Your analyses should make use of concepts and techniques learned in this class, i.e. follow Bayesian methodology. If you think that one or more aspects of your analyses require methods not discussed in this class, please discuss this with me as soon as you realize. For some projects, the move beyond Bayesian methods may be justified based on the lack of comparable Bayesian models or the fact that Bayesian implementations are well beyond the scope

of this class. However, complex non-Bayesian alternatives may have a steep learning curve as well and are not easily implemented or understood. Thus, when in doubt, keep it simple and attempt to stay within the realms of what we have covered in this class.

There is no strict length minimum or maximum for the project paper, but most students, depending on the scope of their research question(s), should be able to address all of the requirements in about 5-10 pages of written text (double-spaced), not including figures.

I expect that most students will choose a research project topic that is related to their thesis, and I hope that the work you do for this project can directly benefit your thesis efforts. However, a connection to a larger research project is not a strict requirement. If you are not preparing a thesis, you can choose any topic of interest. If you have some question independent of a thesis that you would like to explore, you are also welcome to choose that. If you are in the beginning phases of your thesis research and have ideas but no data of your own, feel free to use data from your advisor, collaborators, and/or publicly available datasets. Running statistical analyses on data that are similar to what you might collect can be a good "test run" of the approaches you will employ when you do have your own data. It is even possible to simulate data for the purposes of this term project if you're so inclined.

I will also review the R code you produce as part of this project. It's increasingly expected that published scientific papers are accompanied by relevant statistical code, so the general aim here is to encourage you to carefully prepare your code as you would any other form of scientific communication. The purpose of sharing code is to enable reproducibility. As such, your code should be well-organized and well-commented. I strongly recommend keeping all your code in one self-contained mark-down document that starts with the loading of all required libraries and all the data you will need, similar to what you are used to for your course assignments. You will know if it is fully self-contained when you try to knit it to a PDF.

All research project submissions should be uploaded to CourseWorks by 11:59 pm on the dates indicated below.

Part 1: Prospectus (5% of Overall Project Grade)

As a first step in the research project process, each of you will submit a short prospectus outlining your project (1-2 paragraphs total). First, describe the main question(s) you want to answer and the kinds of analyses you intend to use. **Note that you are not limited to using statistical techniques that will be taught in this course.** Part of being an independent scientist is being able to identify analytical methods that are most appropriate for your work and learning to implement those approaches. However, I also understand that at the time the project prospectus is due, we will still be working through relatively basic material as a class. Thus, if you know what you want to accomplish in a broad sense but don't know the best statistical methods for the job, indicate that in your prospectus, and I will provide some guidance. Second, the prospectus should also include a brief description

of the data you plan to use. Will you use your own data? Data your advisor has? Data that is publicly available? How much data do you plan to have? If it isn't all collated in spreadsheets, how will you prepare the data for analysis in R? If the data exist but you don't have it yourself, what steps do you need to take to acquire the data? What are the primary variables of interest for your question(s)?

The overall goal of the project prospectus is simply to ensure that you have a plan for your research project early on in the semester. It will also allow me to best assist you towards successful completion of the project.

- **Refer to Canvas, syllabus and/or email announcements for due date!**

Part 2: Rough Draft (10% of Overall Project Grade)

In order to help everyone make good progress towards their project goals, I will also review one rough draft of each student's research project. The research project rough draft should include the text of your report (at whatever state you have it in), a copy of the R code you have been developing, and your data. If your dataset(s) are very large, talk to me about how to get them to me.

In addition, **please include a short paragraph (2-5 sentences) at the end of your project rough draft with your own evaluation of how your project does or does not meet the standards outlined in the rubric below.** This paragraph does not need to appear in your project final draft. It is intended to help you think at this intermediate project stage about which parts of the project you already have a good handle on, and, by contrast, those elements that may need more work. I want to make sure you have a good understanding of the project expectations *before* you turn in the final draft.

- **Refer to Canvas, syllabus and/or email announcements for due date!** But you're welcome to submit your draft earlier than the official deadline, if you'd like, and I can try to give you more rapid feedback.

Part 3: Final Draft (85% of Overall Project Grade)

- **Refer to Canvas, syllabus and/or email announcements for due date!**

Research Project Final Draft Grading Rubric

1) Background and Objectives/Hypotheses (10 points)

- provides a brief (1-2 paragraph) summary of the context for the analyses using relevant literature
- clearly states the objectives/hypotheses for the analyses

2) Methods (25 points)

- describes data collection and processing procedures
- describes the statistical techniques and R packages used for the analyses
- methods are described at a level of detail such that a scientifically-savvy reader could replicate the analyses

3) Results (25 points)

- accurately and comprehensively describes the results of all analyses outlined in the methods section
- where relevant, results not only describe point estimates for statistical measures of interest but also communicate uncertainty

4) Data Presentation, Figures, Tables (10 points)

- where relevant, data that are described in text are also presented visually, whether by figures or tables
- figures and/or tables are captioned, appropriately labeled (i.e., column headings, axis labels, etc.), and feature readable font sizes

5) Discussion (10 points)

- provides a brief (1-2 paragraph) summary and interpretation of the statistical results that speak to the motivating study objectives/hypotheses

6) Literature Cited (5 points)

- all primary literature and software referenced in the main text appear in the bibliography using a consistent citation style (use the *Ecology* journal style as described here if you don't have another preferred style)

7) Overall Writing Style (5 points)

- writing is free from grammatical errors
- writing demonstrates clear thinking with smooth transitions between ideas

8) Supplemental Code (10 points)

- code is styled to be neat and readable
- code is commented to outline at least the major steps in the analysis
- code is complete such that it enables reproducibility