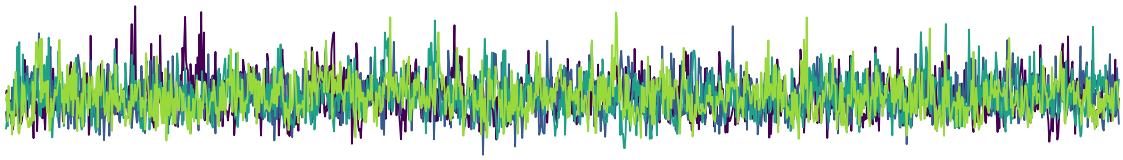
EEEB GR5005
Graduate Level Research Project

Columbia University - Spring 2019



Research Project Outline

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 analytical methods. Basically, this project is a scientific paper largely reduced to its analytical core.

Generally, I envision the project format as follows:

- A short explanation of the background behind the study. This should be roughly a 1-2 paragraph introduction (not an extensive literature review) ending with hypotheses.
- Detailed sections on analytical 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 patterns in the data (1-2 paragraphs). If your analysis uses statistical techniques not taught in class, I am interested in having some informal discussion about those here as well. Briefly, how did you come to the conclusion this was the best way to analyze your data? How did you implement the techniques in R? If you used functions from an R package, how did you decide on which package(s) to use? Conceptually, do the statistical methods you chose have some relation to linear modeling?

There is no strict length minimum or maximum for this project, but most students, pending 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 "dry 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.

As the first step in the 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.

Research Project Submission Details and Dates

- **The research project prospectus will be due on Thursday, March 07.**
- If you so choose, **I will review one rough draft of each student's research project.** In order to give me time to review drafts and provide feedback before final drafts are due, **any rough draft is due by Thursday, April 11 at the latest.** You're welcome to submit a draft earlier, if you'd like. A research project draft should include the text of your report (at whatever state you have it in), a copy of the R code

you have been using/developing, and your data. If your dataset(s) are very large, talk to me about how to get them to me.

- For all students, **research project final drafts will be due on Thursday, May 02**, with an identical format to the project rough drafts outlined above.
- All project submissions should be uploaded to CourseWorks by 11:59 pm on the date indicated.

Research Project 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/or software 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