EDS 222 Final Project Guidelines
Professor: Tamma Carleton
Fall 2022

Goal of the final project:
The goal of this project is to explore an environmental data science question for which you do not know the answer. You will construct a research question, collect relevant data, and design a statistical analysis to answer your question. Your analysis must apply at least some of the statistical concepts you have learned in this course, including concepts covered in the second half of the class. You will summarize your findings and communicate clearly how they have or have not helped you answer your question.

Your results do not need to be conclusive! Most data science questions cannot be fully answered with one small analysis. If you carefully conduct your analysis, yet cannot conclude anything concrete, *that is perfectly fine*. There are many reasons you might find yourself in this situation, including, among many others: you have null results (i.e., you fail to reject the null hypothesis); your results are uncertain (e.g., one test suggests one answer while another test suggests another answer); there are obvious issues with the analysis due to limited data availability; there are obvious issues with the analysis because of violations of OLS assumptions, etc. Regardless of the final results, make sure to carefully describe your data and analysis, and to clearly articulate the limitations of your findings.

The submission has three parts:

1. **Brief proposal [Due date: 11/10, 9am]**

   Please write a short paragraph (4-5 sentences) describing the project you propose to pursue. Motivate the question, describe possible data sources, and suggest possible analyses you may conduct to answer your question. Please **email Sandy your proposal** at sandysum@ucsb.edu by 9am on November 10th.

2. **Technical blog post [Due date: 12/9, 5pm]**

   This blog post is a write up summarizing in text and with figures and/or tables your question, the data you have collected, your analysis plan, and your results. Your target audience should be other quantitative scientists and practitioners familiar with the basics of statistics and data science, but not necessarily experts in environmental science or the details of the methods studied in this course.

   Please **email me your blog post** in PDF format *and* as a live link by 5pm on December 9.

   Some guidelines for the blog post:
   - 3-5 pages in length, including figures, tables, captions, and references list
   - 1-3 tables or figures, each carefully labeled and captioned so that they are easily interpretable
   - Include scientific references when applicable

- Include links to the underlying data you use. If your data cannot be shared publicly, note this in a short "data availability" statement at the end of your post.
- [Recommended, but optional] If you can, include a link to a repository with your replication code. I will ***not*** evaluate your code as part of your grade, but this is good practice for reproducibility and transparency and will make your blog post more exciting to the outside world.

3. **3 minute in-class presentation [Presentations: 12/6, 8-11am]**

This is a short presentation for the class in which you motivate your question, describe your data and analysis plan, and summarize your results. It should be a fun way to practice sharing data science with an audience through public speaking. *You will be cut off after 3 minutes.* Don't worry, we will have breaks – no one can pay attention for three hours straight, no matter how exciting the data science is!

General guidelines:
- *Motivate your question.* Why is this important? Is there existing evidence on this question? If so, why is it inconclusive? If not, why not?
- *Describe your data.* Where did you access it? What are its spatial and temporal features? What are its limitations? What do you know about the sampling strategy and what biases that may introduce? If helpful, you can use a histogram, scatterplot, or summary statistics table to describe your data.
- *Clearly describe your analysis plan.* What is your analysis plan? Why did you choose this analysis, given your data and question? What are the limitations?
- *Summarize your results visually and in words.* Show us your results in figure(s) and/or table(s) that are carefully labeled and captioned. Describe in the text (and orally when presenting) what you found, and how these results either do or do not help you answer your question.
- *What might you do next?* One short analysis cannot fully answer an interesting scientific question. If you had time to collect more data or conduct more analysis, what would help you answer this question better?