SC212: The Project

Overview: In this project, you will select or collect a data set of interest to you, analyze the data, summarize your preliminary results in a report, and present your results in poster format.

Groups: comprised of 2 or 3 people.

Grading:

	Percent
Research Proposal	15
Finalized Dataset	5
Preliminary Analysis & Summary	40
Poster Presentation	40

The purpose of this assignment is to provide you with experience in:

- formulating a specific research question
- finding and evaluating peer-reviewed journal articles relevant to your research question
- writing an introduction to a scientific research paper that uses peer-reviewed journal articles to motivate and provide background for the proposed research question
- determining the type of data that are required to answer a specific research question
- reporting the findings of a statistical analysis
- presenting your findings to a large audience

Stage I: Research Proposal

The Research Proposal will be a formal submission no longer than 2 double-spaced typed pages. Be sure to use proper grammar and complete sentences throughout. The following elements should be included in your proposal.

To complete your Study Proposal you should:

- Identify your research question(s). Your questions should have clear explanatory and response variables (e.g. Is variable X associated with variable Y, etc.)
- The motivation for your research question(s)
    - You must explain why the questions you are investigating are relevant and/or important. Use your references to inform this portion of your proposal.
- Read and reference at least two articles in the refereed literature that are relevant to your question of interest
    - You should use these references in the introduction of your proposal. Briefly mention what the authors did and cite their relevant findings. Pay close attention to the figures, tables, and methods sections of the papers

you select as they can give you an idea of what I'll be expecting from you later on. Try to avoid articles that are too technical as well as those that appear in the popular press (e.g. The New York Times, etc.). Articles that appear in the popular press do not count as refereed references, although they may help motivate ideas for your project.
    - YOU WILL LOSE POINTS IF YOU FAIL TO INCLUDE VALID REFERENCES
    - Make SURE you report the full citation – not just a website (for journal articles).
- Specify the dataset that you are planning to use (it's ok if this changes later on)
    - The dataset can come from research you have conducted, friends or professors who have collected scientific data, reputable internet sites, etc. Some useful websites for obtaining data will be posted on the course moodle site. If you're having problems finding data, come and see me.
    - Describe how you plan to use the variables in your data set to address your research questions. What relationships should you see to confirm your hypotheses? How will you measure potentially abstract concepts?
    - Make sure you have enough data that you can analyze. You should have at least 40 records/individuals/rows of data. If you don't, you may need to find another source.

Important: If you are conducting your own survey, you must conform to the regulations set out for studies involving human subjects by Colby's Institutional Review Board.

Stage II: Gather and Finalize Your Dataset

After you've turned in your study proposal you will have a few weeks to gather and finalize your data

- Clean/recode your dataset
    - You should examine your data and identify any possible outliers/errors.
    - You may need to create new variables based on existing variables in your dataset.
    - You may decide to reduce the overall number of observations or variables in your dataset – however you should have AT LEAST 40 observations and, ideally, many more.

- Select at least five (but try to keep it under ten) specific variables from the data that would address your question of interest. Minimum requirements:
    - Include at least one **response** variable that is **continuous**
    - Include at least one **response** variable that is **categorical** (if you only have a continuous response variable, you may create a categorical response by recoding your continuous variable)

  - Include at least one **explanatory** variable that is **categorical** and has at least 3 categories (you may recode a continuous variable, e.g. age, so that it is categorical if you wish)
  - Include at least one **explanatory** variable that is **continuous**
  - Include at least three explanatory variables
  - Complete a "variable description chart" that contains the following information: data source, primary research question, variable names, variable roles (response, explanatory, confounder/lurking), variable type (continuous, categorical, etc.), how the variable is coded in your dataset (e.g. "female" =1, "male"=2) and unit of measurement (if applicable). **The template you should use for your chart is posted on our Moodle page.**
  - Your dataset MUST meet all of the above specifications prior to the due date for your group to receive full credit. DO NOT wait until the last minute to find a dataset. Come see me if you have problems/questions.

- **I strongly recommend that you meet** with me to discuss your project and data. The meeting should occur AFTER you've gathered and entered your data – I'll want to see it and most likely will have some suggestions for you about how you should clean/modify it.

- Once you've finalized your dataset, save it as a csv (comma separated values) dataset. Email me your **csv dataset AND your variable description sheet**.

Stage III: Preliminary Analysis and Summary

You should begin your data analysis as soon as you have a finalized dataset. Your summary should be **NO LONGER than 5 double-spaces pages**, and at a minimum, include the following items:

- Appropriate numeric and graphical summaries (e.g. histograms, tables, descriptive statistics, etc.) for the most *important/relevant* variables. At a minimum you should include summaries of your primary response and primary explanatory variables. Summarize the data using methods that are appropriate to the data type (e.g. continuous, categorical, etc.).
- Appropriate graphs and/or tables that ***explore the relationships* between pairs** of variables of interest (e.g. scatterplots, side-by-side boxplots, contingency tables etc.) Be sure to include informative titles on all of your plots/figures and appropriately labeled axis. **At a minimum, you must examine the relationships between three pairs of variables** – one categorical and one quantitative, two categorical, and two quantitative. Your explorations should uncover and display trends which shed light on your research questions and potentially suggest further questions.
- Results (and interpretations) from **at least two different statistical tests**

- **In the body of your paper you SHOULD NOT simply cut and paste R output; present results in well labeled and formatted tables/figures.** You should provide raw R-Studio output from your tests in your appendix – see the next bullet point).
- **An appendix** that contains details from two statistical tests (e.g. null and alternative hypotheses, assumption checking, diagnostic plots, etc.). Provide enough detail so that I can tell what you did and why you did it!
- **A bibliography** that lists any references/resources that you used.

Write up your report using clear, concise sentences. **Your paper should have a logical order**: start with a description of your research question and motivation, then describe your dataset, briefly describe your methods, include the results from your data analysis, your interpretations, and your preliminary conclusions. Also be sure to preview what your next steps will be – e.g. linear regression, anova. When you are describing your results, you might choose to focus on your statistical tests, but do not neglect to describe the simple features of your primary variables (e.g. distributions of primary explanatory and response variables). Don't forget your appendix and your bibliography. **See the posted example on our moodle page.**

Stage IV: The Poster

We will present posters virtually over Zoom during your final exam period, which will be like poster sessions at academic conferences (except more fun). It's a chance to mingle with your classmates and learn about their research findings (there will be evaluation forms to fill out regarding other posters). Your poster should convey your research question and results to readers in a clear, well-organized, concise, and attractive manner.

At a minimum, you should report and correctly interpret the most relevant results from your preliminary analysis AND report findings from an **AVONA OR regression analysis**. Be sure to **include your final model (if you choose regression)** on your poster and **results of any significance tests (for ANOVA or regression)**. During the semester, I'll bring in example posters to give you an idea of what you should prepare.