STAT 206
Final Project Guideline

Due Dates

- **Project Proposal**: Due on November 27 (Mon.), 11:59PM
- **Project Report**: Due on December 11 (Mon.), 11:59PM

Submission Guideline

- **Proposal**: Submit a file (.pdf or .docx) following the name format "STA206-XX-YY-ZZ-ProjectProposal", where XX, YY, ZZ stands for group member names (FirstName LastName) using the "file upload link".
- **Project**: Submit a file (.pdf or .docx) following the name format "STA206-XX-YY-ZZ-Project", where XX, YY, ZZ stands for group member names (FirstName LastName) using the "file upload link".
- **File Upload Link**: `https://ucdavis.app.box.com/f/cdcee3c5707f46e88f5541c22c57a438`
- **Each group should only make one submission**. *By submission of the project files, each member of the group acknowledges that they have worked together on the project and they have agreed to submit it in its submitted form.*

About the Final Project

Your task is to conduct a start-to-finish statistical analysis of a real data set aided by **regression techniques** learned in this class (or beyond). This includes formulating questions, building models, performing analysis and diagnostics, interpreting results and wrapping everything up in a paper format report.

Project Proposal

Project proposal will not be graded. It will be used to make sure you have a project of appropriate scope. The proposal should include the following information.

1. **Group Membership:**

   You can form a group with no more than 3 members. You can also do an individual project, then the group will only have one member. **In the proposal, you need to list all group member(s) and their email address(es).**

   You are strongly encouraged to form a group with fellow classmate(s). The project will be graded solely based on the report.

2. **Data Set to Analyze:**

   Each group need to choose a data set to analyze. You may choose a data set on canvas under "Files/Project", or you may use your own data (e.g., from your lab or from online). In the latter case, make sure that the data from yourself is at least as complex (e.g., in terms of number of variables and number of cases) as the data sets on canvas. If you are not sure, you should come to my office hours to discuss the data set with me.

   In your proposal, you need to indicate which data you want to analyze. If you choose to use a data set on canvas, you only need to indicate the data set name. If you will use your own data set, then you need to provide a short description of your data, including:

   (i) Where do you obtain the data? What the data is about?

   (ii) The number of variables in the data set.

   (iii) Brief descriptions of each variable in the data set (e.g., What are they measuring? Are they categorical or quantitative variables?).

   (iv) The number of observations (cases) in your data set.

3. **Questions of Interest:**

   In the proposal, you should list several questions that you want to answer by analyzing this data set.

4. **Plan for Data Analysis:**

   In the proposal, you should provide a brief plan for data analysis. For example: How would you explore the data (e.g., graphs, summaries statistics) ? Which models are you considering? What are potential pitfalls and how would you deal with them? **Note that, you may change the actual analysis strategies in your final report.**

Project Report

Page Limits:

- The main text (not including appendices and references) of the report should be 4 to 6 pages.
- The report should be prepared using 12pt font, single spacing on US letter paper.
- The written part of the report should be prepared by a commonly used word processor such as *Latex* or *MS Word*.

Project Report Format:

The project report should follow a standard publication format and should include the following parts:

1. **Title:** The title of the project and the names of all group members and their email addresses.

2. **Abstract :** A brief summary of the project including data set, questions of interest and findings.

3. **Introduction:** A description of (i) the data and relevant background; (ii) questions of interest; and (iii) the motivation of the project, i.e., why other people should be interested in your project.

4. **Methods and Results:** A detailed description of your analysis, supported by numerical results, figures and tables. In the main text, you should only include the most important figures and tables. Other (supporting) figures/tables, and R codes and outputs should be included in appendices and referred to in the main text as needed. **You should appropriately number equations, figures and tables, etc. for easy reference.**

   For example: Describe in details how you reached your final model and then write down the final model and the assumptions, parameters estimates and relevant statistics, diagnostics, etc. You should include an exploratory data analysis step to explore the data before conducting more sophisticated analysis. For each step of your analysis, you should provide a motivation (Why do you do it? What goals you intend to achieve by this analysis?), a justification (Is the analysis reasonable for the data and the goals?), and a description of results (What does the analysis tell you?).

5. **Conclusions and Discussion:** A brief summary of main findings and discussion.

   For example: What do your results mean for the questions you were trying to answer? What is your conclusion? How general are your results, to what extent do they apply or not apply? What are the limitations of your analysis? What you might have done to address these limitations? And any other comments and discussions that you deem important.

6. **Appendices:** The following should be included in the appendices (do not include these in the main text):

   - Supporting figures and tables.
   - `R` codes and outputs. **You should organize the codes and outputs in a legible and easy to follow format (consider using R Markdown). You can include a limited amount of raw output from R, but a flood of unprocessed output from R may result in penalties.**

7. **References:** papers and other sources that being used in your report.

Please carefully proofread your report and correct typos, grammar errors, etc., before submission.