MATH 141: Introduction to Probability and Statistics, Fall 2021
Project Guidelines

Project Timeline		
Phase	**Description**	**Due**
1	Group Formation	9/10
2	Data selection, study design, data exploration, and framing research questions	10/1
3	Hypothesis formulation, more data exploration, and testing out methods as appropriate	11/15
4	Testing out the methods, methods fine-tuning, and writing the interpretation, discussion, conclusions of the results in the context of the data. Putting the entire report together as one scientific narrative.	12/8

Purpose: Writing a scientific report/paper is essential for any researcher. Your project is a way for you to get exposure to the rigors of writing a paper that involves data exploration and statistical analysis.

Datasets: Each group must choose a dataset or multiple related datasets from the list (see Data Sets list) for their project. If you need/want to have additional related data sets that are not listed but can help with your project, please let the instructor know for a discussion.

Submission Requirement: The most important aspect of your project is that you must write a reproducible report. That means that you must include code into your report - similar to your lab assignments. This can be done easily using RMarkdown. Your submission should be a pdf file with the names of your group members.

<u>Project Phases</u>

Phase 1 - Group Formation: Collaboration is an important part of any scientific pursuit. Each group should consist of 4-5 people. For people who want to form their own groups, please send a list of names for your group by the due date. The rest will be randomly assigned. Final group formations will be announced.

Phase 2 - Data selection, study design, data exploration, and framing research questions: This part of your project involves exploring your chosen dataset. Your dataset must be rich enough with information that you can ask questions and perform statistical analysis. A rich data set involves many rows and columns with a mixture of numerical and categorical variables. Your first project report document must include these three parts.

1. Dataset: Put together your first data frame. What is it? Where does it come from? What is the description of each variable? What are the types of variables? What kind of sampling strategy (or data collection) is used?
2. Exploration: Exploring the dataset involves creating plots such as scatterplots, barplots, box plots, histograms, etc. Once you see what you are interested in, pick plots that are relevant to your research questions.
3. Research questions: Formulate your research question and include relevant background research.

Phase 3 - Hypothesis formulation, more data exploration, and testing out methods as appropriate: This phase of your project involves more data exploration, testing out methods that we discussed in class, and formulating your hypothesis, which includes statements and mathematical symbols. Also, make sure you make action items that address the comments provided to you in your Phase 1 submission.

1. More Exploration: In order to create a hypothesis, you need to explore more and consider different ways to look at your data. What other variables can you include in your analysis?
2. Hypothesis Formulation: Based on your research questions, formulate a hypothesis - or hypotheses. The hypothesis statements must be full sentences with its corresponding mathematical notations.
3. Applying Statistical Methods: Depending on the hypothesis, what statistical method can you apply in order to answer your research question? Are you using proportions, means, difference in proportions, difference in means, etc? What are your point estimates? What are your initial results?

Phase 4 - Testing out the methods, methods fine-tuning, and writing the interpretation, discussion, conclusions of the results in the context of the data. Putting the entire report together as one scientific narrative. The last phase of your project involved carrying out the method you described from phase 3 and addressing the feedback comments.

1. Performing the statistical analysis: Perform the appropriate statistical method to your analysis. Interpret the results and discuss your conclusions.
2. Put your report together as one scientific narrative: Your final report should contain these three parts below. Note that you already wrote most of these from Phases 2-3. You just need to organize them in a more scientific narrative fashion.
    a. Introduction: This section should include your background research, research questions, hypothesis statements, and data description.
    b. Data Exploration: This section should include your exploration of your data that leads to your question and hypotheses. Include details on how you wrangle your data or if made additional computations.
    c. Methods: This section should include the processes you took on performing your statistical method.
    d. Discussions and Conclusions: This section should include your interpretation and analysis of the results and conclusions.