Course Final Project

This project serves as an opportunity to apply the techniques presented in the course videos through an end-to-end, completely reproducible research project. This assignment will contain multiple parts. You will need to write a report that resembles a typical article in a biomedical journal and includes an abstract, introduction, methods, results and discussion sections, as well as any appropriate tables/charts/graphs and a list of references (see sections 1-6 below). You will need to provide all code with useful comments written in a literate programming style (section 7). The ‘report’ sections and code should be interspersed throughout a document like RMarkdown or Jupyter notebook. Finally, you will need to upload all documents relating to the project (section 7) to the class repository. Keep in mind best reproducibility practices throughout the project, and see the sections below for more details and expectations.

1. Abstract (typically 250 words or less)
   1. Motivation: Why do we care about the problem and the results? What is the knowledge gap your work fills?
   2. Problem statement: What problem are you trying to solve? What is the scope of your work (a generalized approach, or for a specific situation)?
   3. Approach: How did you go about solving or making progress on the problem? Did you use simulation, analytic models, prototype construction, or analysis of field data for an actual product?
   4. Results: What did you find? What did you answer?
   5. Conclusions: What are the implications of your results?
2. Introduction (1-2 paragraphs)
   1. Project topic background and importance (more detailed than the abstract)
   2. Short summary of previous work (literature review)
   3. The knowledge gap your work will be filling (more detailed than the abstract)
   4. Be sure to cite previous work where appropriate
3. Methods (2-4 paragraphs with code interspersed)
   1. Data description
      1. The source of the data
      2. The date (or time period) of collection
      3. Method of data collection
      4. Sample size
      5. Outcome of interest (include type, i.e. binary, continuous, categorical, time to event, etc.)
      6. List of predictors (covariates)
   2. Brief data cleaning (munging) summary
      1. Were there missing values? How did you handle them?
   3. Statistical methods used for analysis
4. Results
   1. Include appropriate tables, charts and figures
   2. Report important outcome values or measurements (p-values, odds ratios, etc.)
   3. Include all code used to generate the visuals and results
5. Discussion (1-2 paragraphs)
   1. Discuss your findings (Did you answer your scientific question/hypothesis?)
   2. Include any conclusions you can make from your results and how they advance work in your field including any implications your findings will have in practice
   3. Briefly describe possible future directions this work could take
6. References
   1. Include a list of references and work cited in the paper
7. Project Code
   1. Document and comment all code files
      1. Include a README file that details the content of each file and how to run the analysis
   2. Provide clean data set(s) and/or synthetic data sets for the user with variable labels and descriptions
   3. Upload code and all project documents to a repository