BIOS 7400: Data Management and Computing
Final Project Guidelines

This document provides specifics about the final project for BIOS 7400. The final report should follow the formatting guidelines listed in section D.

A. Dataset (10 points)

Obtain one or more open datasets from public sources (e.g. NIH, CDC, etc.), use the SAS skills you learned from the course to manage the datasets the way you would like to analyze (section B and C), and then write a report on your findings (section F). Introduce the dataset in this section, assuming that the reader is not familiar with the data.

B. Data Management (50 points)

1. Read in the datasets in their original format (e.g. .xls, .txt, .dat, .asc, etc. or .sas7bdat if you are lucky) using SAS. Assign variable names if needed. Attach labels to make the variables easy to be interpreted.
2. Check the initial dataset attributes, use SAS to change or create variables if needed.
3. Apply SAS skills learned from the course to manipulate the datasets (e.g. combining data sets, array, macro, conditional execution, iterative operation, etc.). Output a ready-to-use dataset and save it as a permanent data file in SAS format.
4. For SAS executions made to change the dataset in B1-B3, write down your justifications.
5. Ensure your data management process is appropriate and meaningful.

C. Analysis (15 points)

1. Analyze the manipulated dataset from section B, produce a series of analytical results (e.g. tables and graphs showing descriptive statistics, frequency tabulates, model fitting results from SAS procs, etc.).
2. All tables and graphs should be clearly labeled. Variables should be displayed with meaningful labels. In general, assume that the reader is not familiar with the data.
3. Write a section which summarizes your findings.
4. Ensure your analysis is appropriate and meaningful.

D. Code Quality (20 points)

Ensure your SAS code is efficient and easily and readily maintainable (not just by its author!)

F. Formatting (5 points)

1. Font size 11-12 (tables and figures excluded), single space.
2. Tables and graphs should be inserted in the right place in the report, and in a form as you would see in a journal.
3. Tables and graphs should be numbered (i.e. Table 1, Figure 3) so that you can refer to them in the text.
4. The final report should be **no more than 6 pages** (excluding references). No final report with more than 6 pages will be accepted.
5. The report, datasets, SAS code, log, and output should be submitted online. Name the files the same way as for the homework assignments.