Assignment 3

MSCI 718

March 12, 2019

Due: Thursday, March 28, 2019 at 11:30pm

Instructions: Please complete this assignment and submit online to the LEARN Dropbox as a PDF; the PDF will be considered complete for marking. Also upload any source files, for example, any .R or .Rmd files that you used.

This assignment may be completed ***individually or as a group of size 2, 3, or 4****. If you work in a group, please email the TAs as soon as you can with your names, student numbers, WatIAm IDs, and group name so we can setup your dropbox. In the report, please clearly state the group name, and each group member's name, student number, and contribution. If you work individually, please state that the work submitted is your own, and report the nature of any discussion with other students (e.g., on Piazza, in a study group). Remember, you can discuss approaches to problems with other groups, but* ***the work you submit must be your own or that of your group.***

The Assignment (10 pts per analysis)

Download the red and white wine data sets from `https://archive.ics.uci.edu/ml/datasets/Wine+Quality`, and merge them together into one while encoding the color of wine (i.e. red or white).

Analyze the data by investigating it, forming one or more questions or hypotheses, then testing those hypotheses. You must run **N+1 analyses, where N is the number of people in your group** (i.e., an individual assignment must involve two analyses, a group of 3 must run four analyses). Run at least one multiple linear regression, one simple or multiple logistic regression, and one comparison of two means between different categories or variables (if you do an individual assignment, you need to only do two of these).

Some ideas: find out...

a) whether red or white wine hold higher quality,
b) what factors impact wine quality,
c) what factors predict the color of the wine (i.e. red or white), or
d) additional questions between the variables

Make sure to clearly state your hypothesis for each analysis and why you chose it, check assumptions, interpret the results, and provide clear reasons for each step that you follow. Full marks for a logical, interesting, well-presented analysis.