Assignment 4

MSCI 718

March 19, 2019

Due: Friday, April 5, 2019 at 5:00pm

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.***

Assignment 4: Mini-project (20pts)

Choose a data set in any domain. This could be from your research, from a domain you're passionate amount, an open science dataset, or a sample data set from Data Science websites like Kaggle (https://www.kaggle.com/). Conduct an analysis using R/Rmd and the process described in the course. This analysis *must* involve different statistical techniques than assignment 3: either the techniques taught later in the textbook/course, or, an entirely different technique. You may want to employ multiple tests or answer multiple questions. Ask the instructional staff if you're unsure about your mini-project idea or want support in picking a topic.

You must complete the **online quiz by March 29** declaring your project idea and group members. Otherwise, the assignment will be docked marks.

The assignment will be marked based on how well it demonstrates an effective, complete, logical, and clearly-explained analysis. Use this analysis to demonstrate how well you can employ statistical methods for data analytics.