Course Title: Exploratory Data Analysis

Course Instructor: Roger D. Peng

Course Description:

This course covers the essential exploratory techniques for summarizing data. These techniques are typically applied before formal modeling commences and can help inform the development of more complex statistical models. Exploratory techniques are also important for eliminating or sharpening potential hypotheses about the world that can be addressed by the data. We cover in detail the plotting systems in R as well as some of the basic principles of constructing data graphics. We also cover some of the common multivariate statistical techniques used to visualize high-dimensional data.

Course Content:

- Making exploratory graphs
- Principles of analytic graphics
- Plotting systems and graphics devices in R
- The base, lattice, and ggplot2 plotting systems in R
- Clustering methods
- Dimension reduction techniques

Course Textbook

The book *Exploratory Data Analysis with R* covers the lecture material in this course.

Weekly quizzes

- There are two quizzes (Weeks 1 and 2).
- You must earn a grade of at least 80% to pass a quiz.
- You may attempt each quiz up to 3 times in 8 hours.
- The score from your most successful attempt will count toward your final grade.

The Course Projects

The two course projects will be assessed via peer assessment. In these projects, you will be asked to construct or reproduce certain plots, the purpose of which is to make you familiar with various plotting options. You will be evaluated on the plot that you produce and the code that you write to construct the plot. Course projects evaluated via peer assessment will make use of your GitHub account.

For each Course Project, you are required to evaluate and grade at least four of your classmates' projects.

Grading policy

You must score at least 80% on all assignments (Quizzes & Projects) to pass the course.

Your final grade will be calculated as follows:

- Quiz 1 = 20%
- Quiz 2 = 20%
- Course Project 1 = 25%
- Course Project 2 = 35%

Differences of opinion

Keep in mind that currently data analysis is as much art as it is science - so we may have a difference of opinion - and that is ok! Please refrain from angry, sarcastic, or abusive comments on the message boards. Our goal is to create a supportive community that helps the learning of all students, from the most advanced to those who are just seeing this material for the first time.

Plagiarism

Johns Hopkins University defines plagiarism as "...taking for one's own use the words, ideas, concepts or data of another without proper attribution. Plagiarism includes both direct use or paraphrasing of the words, thoughts, or concepts of another without proper attribution." We take plagiarism very seriously, as does Johns Hopkins University.

We recognize that many students may not have a clear understanding of what plagiarism is or why it is wrong. Please see the JHU referencing guide for more information on plagiarism.

It is critically important that you give people/sources credit when you use their words or ideas. If you do not give proper credit -- particularly when quoting directly from a source -- you violate the trust of your fellow students.

The Coursera Honor code includes an explicit statement about plagiarism:

I will register for only one account. My answers to homework, quizzes and exams will be my own work (except for assignments that explicitly permit collaboration). I will not make solutions to homework, quizzes or exams available to anyone else. This includes both solutions written by me, as well as any official solutions provided by the course staff. I will not engage in any other activities that will dishonestly improve my results or dishonestly improve/hurt the results of others.

Reporting plagiarism on course projects

One of the criteria in the project rubric focuses on plagiarism. Keep in mind that some components of the projects will be very similar across terms and so answers that appear similar may be honest coincidences. However, we would appreciate if you do a basic check for obvious plagiarism and report it during your peer assessment phase.

It is currently very difficult to prove or disprove a charge of plagiarism in the MOOC peer assessment setting. We are not in a position to evaluate whether or not a submission actually constitutes plagiarism, and we will not be able to entertain appeals or to alter any grades that have been assigned through the peer evaluation system.

But if you take the time to report suspected plagiarism, this will help us to understand the extent of the problem and work with Coursera to address critical issues with the current system.