Syllabus - Read on GitHub

Course Title

Developing Data Products Course

Instructor(s)

The primary instructor of this class is Brian Caffo

Brian is a professor at Johns Hopkins Biostatistics and co-directs the SMART working group

This class is co-taught by Roger Peng and Jeff Leek.

Course Description

A data product is the production output from a statistical analysis. Data products automate complex analysis tasks or use technology to expand the utility of a data informed model, algorithm or inference. This course covers the basics of creating data products using Shiny, R packages, and interactive graphics. The course will focus on the fundamentals of creating a data product that can be used to tell a story about data to a mass audience.

In this class students will learn a variety of core tools for creating data products in R and R Studio in specific. Students will be evaluated via quizzes and a culminating project.

Course Content

The lectures will be taught over four weeks with the third week dedicated to creating R packages.

The weeks are organized as follows

1. Shiny, rCharts, manipulate, googleVis

2. Presenting data analysis, slidify, R Studio presenter.

3. Students creating and deploying their projects

4. Creating R packages, classes and methods, yhat.

Github repository

The most up to date information on the course lecture notes will always be in the Github repository

The data science specialization is here

https://github.com/DataScienceSpecialization/Developing_Data_Products

Please issue pull requests so that we may improve the materials.

YouTube

If you'd prefer to watch the videos on YouTube, most of them can be found here:

https://www.youtube.com/playlist?list=PLpl-gQkQivXhr9PyOWSA3aOHf4ZNTrs90

Book: Developing Data Products in R

This book introduces the topic of Developing Data Products in R. A data product is the ideal output of a Data Science experiment. This book is based on the Coursera Class "Developing Data Products" as part of the Data Science Specialization. Particular emphasis is paid to developing Shiny apps and interactive graphics.

The book is available here: https://leanpub.com/ddp

It's variable pricing, including free! It also includes some content (like leaflet) that was not covered in the class and omits some other. It's a little rough, but as I work on it you'll get all of the updates.

Weekly quizzes

- There are three weekly quizzes.
- 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

Course Project

The Course Project is an opportunity to demonstrate the skills you have learned during the course. It is graded through peer assessment. You must earn a grade of at least 80% to pass the course project.

Grading policy

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

Your final grade will be calculated as follows:

- Quiz 1 = 20%
- Quiz 2 = 20%
- Quiz 3 = 20%
- Course project = 40%

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.