Syllabus

Course Title

Data Scientist's Toolbox

Course Instructor(s)

Jeff Leek

Course Description

In this course you will get an introduction to the main tools and ideas in the data scientists toolbox. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like version control, markdown, git, Github, R, and Rstudio.

Course Content

- Track motivation
- Getting help
- Introduction to basic tools
    - R
    - Rstudio
    - Git
    - Github
- Types of data questions
- Steps in a data analysis
- Putting the science in data science

Lecture Materials

Lecture videos will be released at the beginning of the course and will be available thereafter. You are welcome to view them at your convenience. Accompanying each video lecture will be a PDF copy of the slides and a link to an HTML5 version of the slides.

Weekly quizzes

There are three weekly quizzes. You may begin submitting them as soon as the course opens. Quiz 1 is due at the end of the first week, Quiz 2 is due at the end of the second week, and Quiz 3 is due at the end of the third week. See the Quiz pages for exact due dates. The hard deadline is 5 days after the quiz due date.

Quiz Scoring

You may attempt each quiz up to 3 times. The score from your most successful attempt will count toward your grade.

Hard deadlines and soft deadlines

The reported due date is the soft deadline for each quiz. The hard deadline is five days after the soft deadline. Each day late will incur a 10% penalty, but if you use a late day, the penalty will not be applied to that day.

Late Days for Quizzes

You are permitted 5 total late days for quizzes in the course. If you use a late day, your quiz grade will not be affected.

The Course Project

The Course Project is available from the first day that the class is open. It is due BEFORE 11:30 PM UTC on the Sunday that ends the third week of the class. Peer evaluation of Course Project submissions opens immediately after the assignment is due and runs during the fourth week. You are required to review at least four of your classmates' projects. To access more detailed Course

Project instructions and the submission interface, click the Course Project link in the left navigation bar.

For this course, the project can be evaluated with a series of yes/no answers to determine whether people completed the required installations.

Screenshots and privacy

Part of the course project includes submitting a screenshot to demonstrate you have installed the relevant software. Be sure not to take a screenshot with other applications open that may reveal personal information or anything else you don't want others to see.

Grading policy

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

You must receive a final grade of 70% or better to pass the course and achieve the certificate. You must receive a final grade of 90% to pass the course with distinction.

Typos

- We are prone to a typo or two - please report them and we will try to update the notes accordingly.
- In some cases, the videos may still contain typos that have been fixed in the lecture notes. The lecture notes represent the most up-to-date version of the course material.

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.

Peer Assessment of Course Projects

For many of the Course Projects, peer assessment will be necessary to evaluate the completion of the assignments. We have created and tested rubrics for each assignment. They are not perfect and will not be perfectly applied. However, we believe that the Course Project feedback from peer assessment adds value above simple multiple choice assessments.

- We have tried to make the criteria as objective as possible, do your best to apply them to the best of your abilities.
- If you have questions or suggestions about the rubrics, please report them in the forum, "Rubric Issues".
- If you disagree with the Course Project scores you received through peer assessment, you may report those issues in the "Grading Issues" forum. Please note that it will be impossible for us to revise peer-grades, but we will attempt to use reports to improve future versions of the rubric.

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.

Technical Information

Regardless of your platform (Windows or Mac) you will need a high-speed Internet connection in order to watch the videos on the Coursera web site. It is possible to download the video files and watch them on your computer rather than stream them from Coursera and this may be preferable for some of you.

Here is some platform-specific information:

Windows

The Coursera web site seems to work best with either the Chrome or the Firefox web browsers. In particular, you may run into trouble if you use Internet Explorer. The Chrome and Firefox browsers can be downloaded from: *Chrome: http://www.google.com/chrome* Firefox: http://www.mozilla.org

Mac

The Coursera site appears to work well with Safari, Chrome, or Firefox, so any of these browsers should be fine.
