95778 R for Data Science Syllabus Spring 2021

Instructor

John K. Ostlund

jostlund@andrew.cmu.edu (or jkostlund@gmail.com)

Remote "Office" Hours: 9:00 am to 9:00 pm, via email, TeamViewer, Google Hangouts ...

I will be available to you as much as I can, 7 days per week, unless I'm traveling.

Physical Office Hours: None during this course

PLEASE NOTE: We will have no physical lectures during this course, due to Covid-19 concerns. Instead, I will be posting video lectures online on a weekly basis. We will have quizzes during weeks 3, 5, and 7, but these are through Canvas and you will not need to be in the classroom for these: you can do them from home or from wherever else is convenient for you. We will have group based Final Projects, but with coordination among each group's members I hope you will be able to accomplish these remotely, rather than face-to-face with each other.

Teaching Assistants

To Be Announced

- The primary job of the TAs will be to help me with grading homework assignments and quizzes
- They may also hold office hours, to be determined

Course Information*	Course Title: **95778 R for Data Science, A4 Mini 4** Instructor: John K. Ostlund (jostlund@andrew.cmu.edu or jkostlund@gmail.com)
Prerequisites (if applicable)	90711, 94842, or 95796 (with R experience/training) You must have prior courses/experience in statistics, and prior courses/experience in R, the statistical programming language/system. If you have limited training in statistics or limited experience with R, you will not be successful in this course.

Description*	This course introduces students to advanced features in R to prepare them for a career in Data Science. Students will be exposed to the Tidyverse framework, the advanced method of manipulating data within R as well as the data science lifecycle that is encompassed within RStudio. The course will cover data wrangling, advanced data visualization (including d3), and the modeling paradigm of Machine Learning. These are the skills that allow data science to happen, and students will get the practices by doing each of these things with R. Students will also be exposed to daily routines of data science and will be given problems from industry to solve. Further, the course will expose students to data science application lifecycle using Git for version control and GitLab for communication and collaboration of data science projects.
Course Materials	**Required textbook (available online):** **R for Data Science** (http://r4ds.had.co.nz/) by Garrett Grolemund and Hadley Wickham Other books to which we are likely to refer (all available online): **R Packages** (http://r-pkgs.had.co.nz/) by Hadley Wickham **GGPlot2** (http://ggplot2.org/book/) by Hadley Wickham **Advanced R** (http://adv-r.had.co.nz/) by Hadley Wickham **Elements of Statistical Learning** (https://web.stanford.edu/~hastie/ElemStatLearn/) by Hastie, Tibshirani and Friedman Other resources and references: Tidyverse (https://www.tidyverse.org/) RStudio Cheatsheets (https://www.rstudio.com/resources/cheatsheets/) Stackoverflow (https://stackoverflow.com/) GitHub (https://github.com/) UCLA Institute for Digital Research and Education (https://stats.idre.ucla.edu/)
Evaluation Method	The final grade will be out of 100%. The grading breakdown is listed below: • Weekly Homework 20%

	• Bi-Weekly Quizzes 30% • Final Project 50% I assign letter grades to each homework, quiz, and Group Project deliverable, then compute a course grade from these letter grades. On homework assignments, 90 or better out of 100 points is an A. Quizzes and the Final Project deliverables are graded according to the Grading Scale below. Letter grades have these Heinz standard QPA points: **A**+ 4.33 **A** 4.00 **A-** 3.67 **B**+ 3.33 **B** 3.00 **B-** 2.67 **C**+ 2.33 **C** 2.00 **C-** 1.67 **D**+ 1.33 **D** 1.00 **D-** 0.67 Suppose your Homework average grades were A (which is more or less expected), that your three Quiz grades were A, B+, and B-, and that your Final Project grade was A-. Then your course grade QPA would be: 4.00 * 0.20 + [(4.00 + 3.33 + 2.67) / 3] * 0.30 + 3.67 * 0.50 == 3.635 I round rather than truncating, so this becomes a course letter grade of A-. In order to get an A+ course letter grade, your course grade QPA must be > 4.0 (strictly greater than 4.0).
Learning/Course Objectives*	By the end of the class, students learn to: - Use RStudio (http://www.rstudio.com) projects and advanced features. - Use Git for version control. - Use R to perform modeling. - Use RStudio Shiny to present your work. - Import and export data from various sources. - Perform data wrangling. - Produce visualizations with ggplot2 and other libraries. - Perform machine learning models in R.

	- Use R Notebook to perform data science analysis. - Create simple Shiny dashboards.
Heinz standard Grading Scale*	A+ 97.00 – 100.00% A 93.00 – 96.99% A- 90.00 – 92.99% B+ 87.00 – 89.99% B 83.00 – 86.99% B- 80.00 – 82.99% C+ 77.00 – 79.99% C 73.00 – 76.99% C- 70.00 – 72.99% * No A+ grades will be given for homework: A is the maximum for these (and more or less expected). If a quiz turns out to be more difficult than expected, I may apply a gentler grading scheme than this Heinz standard.
Course/Topical Outline:*	· Week 1 – Data Science Workflow · Week 2 – Coding in R; Communication with Graphs · Week 3 – About the Tidyverse · Week 4 – Loading Data; Vectorization · Week 5 – Model Workflow · Week 6 – Data Science Notebook · Week 7 – R Shiny · Week 8 – Finals Week: Final Project due * Subject to change
Schedule*	Homework is due prior to each lecture with no exceptions granted. Online quizzes will be given during weeks 3, 5 and 7.

	· Week 1 - lecture · Week 2 - homework 1 due/lecture · Week 3 - homework 2 due/quiz 1/lecture · Week 4 - homework 3 due/lecture · Week 5 - homework 4 due/quiz 2/lecture · Week 6 - homework 5 due/lecture · Week 7 - homework 6 due/quiz 3/Group Project presentations * Subject to change.
Course Policies & Expectations	**Assignment Submission:** Everything must be submitted in Canvas by the due date/time. If you experience upload problems with Canvas, email me your work for grading IMMEDIATELY, AND PRIOR TO, THE DUE DATE/TIME, along with a screenshot of the upload error. When emailing me your work, I also need you to email technical information to validate the issue (type out what the error message is that you are receiving, computer information, network information, file information, date/time of attempted upload, and screenshot of error) prior to the due date/time via email to me or you will receive a 0% on the corresponding assignment. I need the error information so I can validate your excuse with Canvas administration – it must be validated by error logging. If you contact me about Canvas submission issues after the due/date time, I cannot help you. **Late Policy:** Unless otherwise stated, no assignments will be accepted late. On the rare occasion that an assignment is announced that it can be submitted late, the assignment will be accepted with a penalty of 10% of the total worth of the assignment per day late, up to and including the late deadline announced. Do *not* ask me to make special exceptions for you and you alone – that is NOT fair to the rest of the class. NO assignments may ever be delivered by email. Please do not ask to have a Canvas assignment re-opened online for late submission. Budget for upload time to Canvas. All assignments are due by the start time of the class which it is due (unless otherwise noted).

Students with Disabilities:

Our community values diversity and seeks to promote meaningful access to educational opportunities for all students. CMU and your instructors are committed to your success and to supporting Section 504 of the Rehabilitation Act of 1973 as amended and the Americans with Disabilities Act (1990). This means that in general no individual who is otherwise qualified shall be excluded from participation in, be denied benefits of, or be subjected to discrimination under any program or activity, solely by reason of having a disability.

If you believe that you need accommodations for a disability, please contact us ASAP, and we will work together to ensure that you have the correct access to resources on campus to assist you through your coursework and time at CMU.

Academic Integrity:

Carnegie Mellon University sets high standards for academic integrity. Those standards are supported and enforced by students, including those who serve as academic integrity hearing panel members and hearing officers. The presumptive sanction for a first offense is course failure, accompanied by the transcript notation "Violation of the Academic Integrity Policy." The standard sanction for a first offense by graduate students may be suspension or expulsion. Please see http://www.cmu.edu/academic-integrity/ for any questions.

Cell Phones, Smartphones and other handheld wireless devices:

Other than during class breaks, please silence ring tones and refrain from engaging in calls, messaging or other use during class time. All devices must not be visible during quizzes.

Policy Regarding Students Using English as a Foreign Language:

Assignments in this course are graded with reference to evidence of the acquisition of concepts, presentation format, and accuracy of information.

Having done business in countries that use languages other than English, we understand that the use of an unfamiliar language can result in unusual word choices or grammatical errors that are not critical to the overall understanding of the information. Therefore, we will take into account your need to function in a language that may be unfamiliar to you. We will provide feedback as appropriate if we feel that language or grammar you have used in assignments would be best if it were configured in a different way.

Use of Canvas System for this course:

The Heinz School uses Carnegie Mellon University's Canvas system to facilitate distance learning as well as to enhance main campus courses. In this course, we will use the Canvas system generally to post lecture notes and related documents and to receive assignments electronically from students.

Take care of yourself:

Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

- CaPS: 412-268-2922

	• Re:solve Crisis Network: 888-796-8226
	• If the situation is life threatening, call the police:
	  ◦ On campus: CMU Police: 412-268-2323
	  ◦ Off campus: 911
	
	If you have questions about this or your coursework, please let me know.