PSYCHOLOGY 896 - DATA SCIENCE

Fall 2017, 3 credits | Lab: TR 9:30AM – 10:45AM, BAL 1013B

Instructor: Richard N. Landers, Ph.D. (rnlanders@odu.edu)
Office: Mills Godwin Building 346E
Off. Hours: By Appointment (e-mail me!)

Course Description

"Data science" is a catch-all term used to describe the practice of working with and analyzing messy data sources to draw meaningful conclusions using techniques developed by computer scientists and computational statisticians. The purpose of this class is to give students who are training as quantitative social scientists a broad introduction to this skillset via the statistical programming language, R. In contrast to a data science course you might take in a statistics or computer science department, this class assumes you already have a working knowledge of univariate statistics (e.g., ANOVA, correlation, regression), as I will not be teaching you statistics. Instead, we will learn how to conduct many statistical analyses in R that you have already done in SPSS, Excel, or another such program. Additionally, we will build on this foundation to explore new skillsets uncommon in the social sciences, such as natural language processing, automated data curation, and machine learning.

Course Objective

By the end of this course, you should be able to write R code at an intermediate level. This means you will be able to import data into R using a variety of approaches, analyze it in a variety of ways, and create statistical output, visualizations or interactive outputs from those analyses. You will also learn "best practices" in terms of code organization, version control, and professionalism. We will tackle these goals with a flipped classroom approach to language learning: daily practice at home with reinforcement and exploration of key concepts in the classroom.

Required Materials

1. You must bring a laptop to class each day with R and R Studio installed. This laptop must be capable of accessing the Internet during class over a wireless connection, and you must be able to install software on it. All class demonstrations will be made using a Microsoft Windows system. It is recommended you also bring a Windows system, but this is not required, so long as you can install required software and translate Windows native approaches into your own operating system (e.g., folder organization and paths). Software-supported systems are Windows Vista/7/8/10, OS X 10.6+, Fedora 19+/RedHat 7+, Debian 8+/Ubuntu 12.04+, and openSUS 13.1+. However, I will be less able to help you for certain problems if you are not using Windows.
2. You must have reliable, convenient Internet access outside of class to complete course assignments and access course materials. Most content in this course is delivered via real-time interactive webpages that requires a live internet connection for the duration of the instruction.

Overall Course Expectations

This course meets twice per week. Aside from the first week, most weeks will be structured as follows. On Thursdays, you will receive an assignment to complete a set of designated instructional content on the website datacamp.com as well as a project to be completed using what you learn there to solve problems related to psychology. Completion of assigned instructional content is required by 6AM the

following Tuesday. In class on Tuesdays, I will provide instruction on any additional skills being targeted that week and will answer questions related to the project, which is due at 6AM on Thursdays. Weekly projects will integrate skills learned on DataCamp and also some new skills introduced during class on Tuesdays. On Thursdays, we will debrief the projects and look at an example "correct program" and walk through it as a class. Additionally, preparatory information regarding the course to be assigned that day may be provided.

Assessment and Evaluation

DataCamp Lessons

Each Thursday, lessons will be assigned from datacamp.com. Some of these lessons will be complete DataCamp courses, and others will be specific modules of courses. This will all be managed by assigning content to your (free) DataCamp account. The amount of content assigned will vary between one and two courses of content. DataCamp considers a course to be roughly 4 hours of lessons, which includes practice time. Realistically, the time you need will depend upon how intuitive you find R to be. For students already familiar with other programming languages and those with previous R experience, "8 hours" of courses is realistically closer to 2 hours; for complete novices that also find the material difficult, 8 hours is a realistic estimate. It is strongly recommended that you stretch out DataCamp lessons across the assignment period, for example, allocating 1 hour each day between Thursday and Tuesday. You will gain the most by treating this like a foreign language immersion course by using R every day, including for your own research.

Semester Grade

	% of Final Grade
DataCamp Lessons	20%
Weekly Assignments	30%
Comprehensive Project	30%
Final Exam	20%
Total	100%

Weekly Assignments

Each Thursday, the previous week's assignment, which is designed to test how well you can apply that week's DataCamp lessons in combination with any material introduced Tuesday, will be due by 6AM. In general, these assignments are designed to take less than 4 hours if you are on top of the DataCamp lessons, and the files you create should be submitted via Blackboard.

Evaluation Scale

Percent	**Letter**
93.4% or higher	A
90.0% - 93.3%	A-
86.7% - 89.9%	B+
83.4% - 86.6%	B
80.0% - 83.3%	B-
76.7% - 79.9%	C+
73.4% - 76.6%	C
70.0% - 73.3%	C-
69.9% or lower	F

Comprehensive Project

On Thursday of Week 14, a comprehensive project will be assigned integrating many of the skills learned over the previous 14 weeks. This will be due the last day of classes this semester, December 8, at 11:59PM. This project will be designed to mimic a research project, involving obtaining and importing data, processing it in a variety of requested ways, and creating several of visualizations and output tables of the type that might be included in an APA-style paper. It will also include learning a new R package and using it correctly in your dataset, which assesses your ability to generalize lessons learned from course content.

Final Exam

The final exam for this course is a 3-hour live coding session. This exam will test your ability to do basic coding tasks in a reasonable amount of time (of much less complexity than the kind you'll demonstrate in

the comprehensive project). You will be required to bring a laptop to the final exam and submit your final exam on Blackboard by the end of the exam period.

Course Policies

The following policies were written to ensure fair, reasonable, and equitable guidelines in the classroom. Richard will apply these policies to all students equally. **These policies are inflexible and non-negotiable and should be referenced should special circumstances arise.**

1. *Attendance:* You are expected to attend all classes. Because this course utilizes a flipped classroom model, if you miss a class, you will miss the opportunity to determine whether your understanding of DataCamp material assigned the previous Thursday or the project assigned the previous Tuesday is accurate. I will not meet with you individually if you have missed classes without valid justification.
2. *Policies and penalties for late work:* If you anticipate something that will restrict the times in which you will be available to complete assignments (e.g., religious observance, National Guard service, subpoenas, jury duty, or a confirmed medical illness), **you must let Richard know as soon as you know that you will be unable to complete the assignment**. Accommodations will be made then. Accommodations will not be made after an assignment due date has passed. Assignments will be accepted late with a 20% penalty per day late as of the due time (6AM) up to a 50% penalty, which will be in place up to 1 week late (for example, if you complete an assigned DataCamp course at 10AM the day it was due, you would receive a 20% penalty. Assignments have a 50% late penalty as of three days after the due date (at midnight) but can be turned in up to a week late at that penalty rate. After a week late, assignments cannot be turned in. Additionally, the comprehensive project can only be turned in until the final exam period.
3. *Disputing a grade:* Most assignments have specific, objectively defined goals that must be achieved that are directly tied to course grades. However, if you feel you have been graded unfairly, you are encouraged to submit your rationale to Richard for reconsideration.
4. *Incompletes:* Incompletes will only be granted in the case of medical or personal emergencies. Per ODU policy, incompletes will only be granted if 80% of the total course time has elapsed and **if you are passing the course at the time of the request** for an incomplete. Contact Richard as soon as you suspect you might need to take an incomplete in the course. See http://www.odu.edu/academics/academic-records/grades/incompletes-withdraws-zgrades for more detail.
5. *Academic integrity:* **Neither plagiarism nor cheating of any kind will be tolerated in this course.** You are responsible for doing all written work independently, without unauthorized collaboration, inadequate citation, or plagiarizing. Plagiarizing includes duplicating substantial portions of code from the Internet, duplicating another student's code, or collaborating with another student. Any incidents of plagiarism or cheating will be dealt with severely, potentially including automatic failure of this course and a permanent notation on your transcript of "Academic Dishonesty." In graduate courses, dismissal from the university is seriously considered even for first offenses. See http://www.odu.edu/about/monarchcitizenship/student-conduct for more details. Code may only be copied wholesale from DataCamp or assigned readings. As a general rule, it is safest if you never *look* at another student's code, although you can discuss assignments more generally. Because code is a type of language, it is as easy to automatically detect plagiarism as it is in English.
6. *Sexual harassment:* Sexual harassment is defined as ANY unwelcome sexual advance, request, or verbal, written, or physical conduct of a sexual nature. University policy prohibits sexual

harassment in any form. Concerns or complaints about sexual harassment should be directed to the Office of Institutional Equity and Diversity: see http://www.odu.edu/equity for more detail.

7. *Disability:* Students are encouraged to self-disclose disabilities that have been verified by the Office of Educational Accessibility by providing Accommodation Letters to their instructors early in the semester in order to start receiving accommodations. Accommodations will not be made until the Accommodation Letters are provided to instructors each semester.

Schedule

Module	Week	Day	Topic(s)
Fundamental R Programming	1	Aug 29	Introduction to the Course
		Aug 31	Setting up R and R Studio
	2	Sep 5	Data Types and Variable Manipulation
		Sep 7	
	3	Sep 12	Conditionals, Loops, and Apply
		Sep 14	
	4	Sep 19	Data Import and Formatting
		Sep 21	
	5	Sep 26	Data Manipulation
		Sep 28	
	6	Oct 3	String Manipulation
		Oct 5	
	7	Oct 10-12	**FALL BREAK (NO CLASS)**
Psychological Statistics	8	Oct 17	Data Visualization
		Oct 19	
	9	Oct 24	ANOVA and RM-ANOVA
		Oct 26	
	10	Oct 31	Correlation and Regression
		Nov 2	
Data Science	11	Nov 7	R Markdown
		Nov 9	
	12	Nov 14	APIs and Web-based Data
		Nov 16	
	13	Nov 21	Machine Learning
		Nov 23	**THANKSGIVING (NO CLASS)**
	14	Nov 28	Text Mining and NLP
		Nov 30	
Wrap-up	15	Dec 5	Project Work Week
		Dec 7	
	END	Dec 8 11:59 PM	**COMPREHENSIVE PROJECT**
		Dec 12 8:30AM	**FINAL EXAM**