Psychology (PSY) 8960
<u>Data Science in Psychology</u>

Instructor: Richard N. Landers, Ph.D. (rlanders@umn.edu)
Location: Elliott Hall N483
Office Hours: By appointment

Catalog 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 basic statistics. Instead, we will learn how to conduct 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. With the exception of the first week and weeks with only one meeting time, weeks are generally structured as follows:

On Thursdays:
1) Instructions for the next week's assignment will be made available on Canvas. This will consist of some combination of:
   a. interactive online learning content on the website datacamp.com, which is due *according to Datacamp's timekeeping* the following Tuesday by 12PM
   b. outside readings, which are due the following Tuesday before class
   c. a project to be completed incorporating what you learn on datacamp, in class, and in readings, which is due the following Thursday by 12PM
2) In class, we will "debrief" the previous week's assigned project. This means that I will complete the project live while you watch and (hopefully) ask for clarification questions. Later in the semester, I will also use this time to demonstrate alternative approaches to the assignment and sometimes provide supplemental skill training.

On Tuesdays, I will provide instruction on any additional skills being targeted that week and will answer questions related to the project due that week. Sometimes new content will be introduced on Tuesdays that you need for the project, so you should not expect to complete the project before Tuesday's class.

On weeks where we meet one day but not the other, watch Canvas for specific instructions. We will also discuss these exceptions in class.

Evaluation Scale

Percent	Letter	Percent	Letter
93.4% or higher	A	73.4% - 76.6%	C
90.0% - 93.3%	A-	70.0% - 73.3%	C-
86.7% - 89.9%	B+	66.7% - 69.9%	D+
83.4% - 86.6%	B	65.0% - 66.6%	D
80.0% - 83.3%	B-	65% or lower	F
76.7% - 79.9%	C+		

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 to 12 hours may be a more 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 Projects

Each Thursday, the previous week's project, 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. 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 Canvas.

Comprehensive Project

On Thursday of Week 15, a comprehensive project will be assigned integrating many of the skills learned over the previous weeks. This will be due on the final exam day, May 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 analyses, which assesses your ability to generalize lessons learned from course content.

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-
66.7% - 69.9%	D+
65.0% - 66.6%	D
64.9% or lower	F

Course Policies

The following policies were written to ensure fair, reasonable, and equitable treatment across students. These policies are inflexible and non-negotiable and should be referenced when relevant.

1. *Attendance:* You are expected to attend all classes; however, attendance is not explicitly graded. However, discussion and activity grades do require you be present in class on those days.   If you need to miss something for a non-emergency event, you must tell Richard **at least 24 hours in advance**, or you will be unable to make up the grade. Students will not be penalized for absence during the semester due to unavoidable or legitimate circumstances. Such circumstances include verified illness, participation in intercollegiate athletic events, subpoenas, jury duty, military service, bereavement, and religious observances. Such circumstances do not include voting in local, state, or national elections. For complete information, please see the official university policy. In these cases, make-up assignments similar in content but with different scope and method will be used.
2. *Policies and penalties for late work:* There are no excuses for late work without extreme extenuating circumstances. You must complete these assignments on time for any credit unless special accommodations are made with Richard **at least one week in advance**.
3. *Disputing a grade:* Although I strive to be as fair and accurate as possible, you may not feel your grade represents your work. If so, please schedule a meeting with me as soon as possible to discuss your grade.
4. *Incompletes:* Incompletes are permitted at Richard's discretion when, due to extraordinary circumstances (as determined by Richard), a student who has successfully completed a substantial portion of the course's work with a passing grade was prevented from completing the work of the course on time. In general, this requires at least 75% of the course has elapsed; earlier in the semester, it is typically recommended that you instead withdraw from the course. You may consult One Stop for more information regarding incompletes.
5. *Academic integrity:* You are expected to do your own academic work and cite sources as necessary. Failing to do so is scholastic dishonesty. Scholastic dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to

falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis (see the [Student Conduct Code](#)). If it is determined that a student has cheated, the student may be given an "F" or an "N" for the course and may face additional sanctions from the University. For additional information, please see the [official university policy](#). The Office for Community Standards has also compiled a useful list of [Frequently Asked Questions](#) pertaining to scholastic dishonesty. If you have additional questions, please email or speak with Richard directly.

6. *Sexual harassment:* The University prohibits sexual misconduct, and encourages anyone experiencing sexual misconduct to access resources for personal support and reporting. If you want to speak confidentially with someone about an experience of sexual misconduct, please [contact your campus resources](#) including the Aurora Center, Boynton Mental Health or Student Counseling Services. If you want to report sexual misconduct, or have questions about the University's policies and procedures related to sexual misconduct, please contact your campus Title IX office or relevant policy contacts. Instructors are required to share information they learn about possible sexual misconduct with the campus Title IX office that addresses these concerns. This allows a Title IX staff member to reach out to those who have experienced sexual misconduct to provide information about personal support resources and options for investigation. You may talk to instructors about concerns related to sexual misconduct, and they will provide support and keep the information you share private to the extent possible given their University role. See [university policy regarding sexual harassment](#) for more information.

7. *Equity, diversity, equal opportunity and affirmative action:* The University provides equal access to and opportunity in its programs and facilities, without regard to race, color, creed, religion, national origin, gender, age, marital status, disability, public assistance status, veteran status, sexual orientation, gender identity, or gender expression. For more information, please consult the [Board of Regents Policy](#).

8. *Disability accommodations.* The University views disability as an important aspect of diversity, and is committed to providing equitable access to learning opportunities for all students. The Disability Resource Center (DRC) is the campus office that collaborates with students who have disabilities to provide and/or arrange reasonable accommodations. If you have, or think you have, a disability in any area such as, mental health, attention, learning, chronic health, sensory, or physical, please contact the DRC office on your campus (612-626-1333) to arrange a confidential discussion regarding equitable access and reasonable accommodations. Students with short-term disabilities, such as a broken arm, can often work with instructors to minimize classroom barriers. In situations where additional assistance is needed, students should contact the DRC as noted above. If you are registered with the DRC and have a disability accommodation letter dated for this semester or this year, please contact your instructor early in the semester to review how the accommodations will be applied in the course. If you are registered with the DRC and have questions or concerns about your accommodations please contact your (access consultant/disability specialist).

9. *Mental health.* As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance and may reduce your ability to participate in daily activities. University services are available to assist you. You can learn more about the broad range of confidential mental health services available on campus via the [Student Mental Health Website](#).

10. *Academic freedom.* Academic freedom is a cornerstone of the University. Within the scope and content of the course as defined by the instructor, it includes the freedom to discuss relevant matters in the classroom and conduct relevant research. Along with this freedom comes responsibility. Students are encouraged to develop the capacity for critical judgment and to engage in a sustained and independent search for truth. Students are free to take reasoned exception to the views offered in any course of study and to reserve judgment about matters of opinion, but they are responsible for learning the content of any course of study for which they are enrolled. When conducting research, pertinent institutional approvals must be obtained and the research must be consistent with University policies. Reports of concerns about academic freedom are taken seriously, and there are individuals and offices available for help. Contact in increasing order of escalation: Richard, your adviser, the [Chair of Psychology](), the [College of Liberal Arts Associate Dean for Social Sciences](), or the [Vice Provost for Faculty and Academic Affairs]().

Course Schedule – Spring 2020

Module	Week	Day	Topic(s)
Fundamental R Programming	1	Jan 21	Introduction to Data Science
		Jan 23	Setting up R, R Studio, Git, and Github
	2	Jan 28	Data Types and Variable Manipulation
		Jan 30	
	3	Feb 4	Conditionals, Loops, and Apply
		Feb 6	
	4	Feb 11	Data Import, Formatting, and *tidyverse*
		Feb 13	
	5	Feb 18	Manipulating Data and Deeper *tidyverse*
		Feb 20	
	6	Feb 25	Manipulating Strings with *tidyverse*
		Feb 27	
Data Science	7	Mar 3	Data Visualization with *ggplot2*
		Mar 5	
	8	Mar 10	**SPRING BREAK (NO CLASS)**
		Mar 12	
	9	Mar 17	R Markdown, Notebooks, and Web Apps
		Mar 19	
	10	Mar 24	APIs and Web Scraping
		Mar 26	
	11	Mar 31	Machine Learning I
		Apr 2	**NO CLASS**
	12	Apr 7	Machine Learning II
		Apr 9	
	13	Apr 14	Shell, Parallelization, and Supercomputing
		Apr 16	
	14	Apr 21	Text Mining and NLP
		Apr 23	**NO CLASS**
	15	Apr 28	SQL Databases
		Apr 30	

Final Project Due **Friday, May 8**