

eSoc 214: Introduction to Data Science
Tuesdays and Thursdays, 9:30am – 10:45am

Professor Yotam Shmargad

Email: yotam@email.arizona.edu

Office Location: SIRLS Building, Room 2A

Office Hours: By appointment (email me)

Course Homepage: D2L

Course Description:

As data continue to penetrate everything we do in contemporary work across many professions, employers are seeking data savvy people to extract meaning and patterns from data. This course provides an introduction to the various skills and considerations required for data management and analysis in business, education, and science. Particular attention will be given to learning how to use the free and open-source computing environment R, as well as its data visualization package ggplot2.

Upon completion of this course, students should be able to:

1. Know about big data
2. Think critically about how data are stored, used, managed, analyzed, and presented
3. Engage in computational thinking and modeling
4. Experiment with skills required for large-scale data analysis

A Few Words on Technology:

1. You will have access to and will be required to retrieve all course materials from the course page in D2L. Please prepare now for this experience by familiarizing yourself with D2L, the web-based courseware supporting this course. Training for D2L can be found online at: <http://help.d2l.arizona.edu/students>.
2. We will be using R to analyze and visualize data in this course. If you do not have a personal computer, you will need to use one of the various computers on campus or a friend's computer in order to access the software.

Let me know ASAP if you are unable to get access to a personal computer

3. Finally, always have a back-up plan. If your 'default' or most preferred computing location fails, be prepared to find a computer to use when you need a backup machine. It would be to your benefit to familiarize yourself with services offered at the university libraries and the ILC on campus.

Readings:

There is only one book that is required for *purchase* in this class: "Doing Data Science: Straight Talk from the Frontline", by Cathy O'Neil and Rachel Schutt. The authors are both top-notch industry data scientists, and the book will serve as a good reference for some of the more technical material that I will cover in the course. We will also be reading from the book, "An Introduction to Data Science", by Jeffrey Stanton, which is *free* and provides a fantastic introduction to R. You can find it here: https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf. Aside from these books, there will be other required readings. I will post these readings for you on D2L. Some come from academic journals, and others are news articles that appear in many of the newspapers you read in print and online. It is *crucial* that you read all assigned readings to do well in this class. Anyone who has not done the reading will simply not be able to participate. Needless to say, this will not do good things for your grade. Come to class well-prepared and you will be well-rewarded.

Complete List of Assignments with Grade Breakdown and Due Dates:

- Participation (10%)
- R Exercises (5% each, 25% total):
 - Exercise 1: Due September 15 at 9:00am (D2L Dropbox)
 - Exercise 2: Due September 22 at 9:00am (D2L Dropbox)
 - Exercise 3: Due October 6 at 9:00am (D2L Dropbox)
 - Exercise 4: Due October 27 at 9:00am (D2L Dropbox)
 - Exercise 5: Due November 3 at 9:00am (D2L Dropbox)
- Group Case 1 (15%): Due October 13 at 9:00am (D2L Dropbox)
- Group Case 2 (15%): Due November 10 at 9:00am (D2L Dropbox)
- Final Presentation (10%): Due December 1 at 9:00am (D2L Dropbox)
- Final Paper (25%): Due December 8 at 9:00am (D2L Dropbox)

Grade Distribution:

90-100% = A “exemplary, far beyond reqs/expectations”

80-89% = B “exceeds requirements/expectations”

70-79% = C “meets requirements/expectations”

60-69% = D “falls short of requirements/expectations”

< 60% = E “repeat of course needed”

All written work will be evaluated for format, organization, style, grammar, and punctuation as well as content and argument. Written work turned in for this course is expected to be formatted in accordance with the American Psychological Association. The instructions for course assignments will be posted on D2L well before the due dates.

Attendance, Due Dates, and Missing Work:

1. The UA’s policy concerning Class Attendance and Administrative Drops is available at:
<http://catalog.arizona.edu/2015-16/policies/classatten.htm>
2. Attendance is “voluntary,” but there is no way you will pass the class without showing up for face-to-face meetings. If you have a lot going on in your personal or professional life that will conflict with your attendance, this may not be a good time to enroll in this course! Attendance is critical.
3. All holidays or special events observed by organized religions will be honored for those students who show affiliation with that particular religion.
4. Absences pre-approved by the UA Dean of Students (or Dean designee) will be honored.
5. Arriving late and leaving early is extremely disruptive to others in the class. Please avoid this kind of disruption.
6. Missed class assignments or exams cannot be made up without a well-documented, verifiable, excuse (for example, a physician’s medical excuse). The validity of such excuses will be assessed by the professor during a face-to-face meeting. Indeed, *due dates are firm*, and late work will be accepted only with a verifiable and valid excuse.

Course Conduct and Campus Policies (be familiar with all campus policies):

1. Food and technologies are issues in classrooms. Cellular telephones are distracting, so please put them away. Laptops can be utilized, but only for note-taking purposes. Please follow classroom rules regarding food and beverages in the classroom.
2. This is a safe environment for sharing and generating unique ideas. Please try to be “open” to diverse perspectives and learn from others who may pose views that differ from your own.
3. Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: <http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity>.

4. It is the University's goal that learning experiences be as accessible as possible. If you anticipate or experience physical or academic barriers based on disability or pregnancy, please let me know immediately so that we can discuss options. You are also welcome to contact Disability Resources (520-621-3268) to establish reasonable accommodations. For additional information on Disability Resources and reasonable accommodations, please visit <http://drc.arizona.edu/>.
5. The Arizona Board of Regents' Student Code of Conduct, ABOR Policy 5-308, prohibits threats of physical harm to any member of the University community, including to one's self. See: <http://policy.arizona.edu/threatening-behavior-students>.
6. All student records will be managed and held confidentially. <http://www.registrar.arizona.edu/ferpa/default.htm>
7. Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.
8. UA Non-discrimination and Anti-harassment policy: <http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy>.
9. Confidentiality of Student Records: <http://www.registrar.arizona.edu/ferpa/default.htm>.
10. Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.

Requirements for the Course:

To succeed in this course, 2-3 hours of study time per hour of formal class time (or per unit) are required. This means that in addition to our three hours of formal class meeting time, 6-9 hours a week of study time are needed in order to meet course expectations. These hours should be spent on reading texts, writing papers, researching for new information, or thinking about course content. College-level reading and writing abilities are assumed.

Honors Students' Requirements:

Students wishing to contract this course for Honors Credit should email me to set up an appointment to discuss the terms of the contract and to sign the Honors Course Contract Request Form. The form is available at <http://www.honors.arizona.edu/documents/students/ContractRequestForm.pdf>. Students earning credit with the University of Arizona Honors College will be held to the following enhancements:

1. Honors students will be required to replicate and extend the data analyses in a respectable academic publication, and to critically evaluate the decisions made by the authors of the paper. Specifically, students will need to present their replication, discuss various issues that the authors had to consider, then conclude by addressing a particular limitation and extending the analyses in a way that addresses the limitation. Students' positions will need to be supported with relevant research. *This completed paper will be roughly 8-10 pages.*
2. Honors students will also be expected to informally 'journal' about the course each week. Each week, that is, students will be required to write a five-sentence paragraph reflecting on some issue or moment that has arisen in our readings or discussions (e.g., the problem with particular terms or some philosophical or practical dilemma). Ultimately, if offering a paragraph each week, honors students will have written roughly *15 reflective paragraphs for the semester.*

Section 1. Fundamentals

In this first section, we will take a close look at some of the societal implication of big data and think critically about how data can be used to produce insight. We will get our hand dirty with R, and will learn about some of the fundamentals of data science: functions, packages, and exploratory data analysis.

1. Tuesday, August 25: Welcome and Course Overview

Assignment for Next Class:

- Read: Lazer, D. 2015. "The Rise of the Social Algorithm." *Science*.
- Read: Sandvig, C. 2015. "What Facebook's 'It's not our Fault' Study Really Means." *Wired*.

2. Thursday, August 27: Data Science and Society

Assignment for Next Class:

- Read: boyd, d. and K. Crawford. 2012. "Critical Questions for Big Data." *Information, Communication, and Society*.
- Read: Tufekci, Z. 2014. "Engineering the Public: Big Data, Surveillance, and Computational Politics." *First Monday*.

3. Tuesday, September 1: Ethics of Big Data

4. Thursday, September 3: Technology of Big Data

Assignments for Next Class:

- Read: Chapter 1 of "Doing Data Science."
- Read: Chapters 1-2 of "Introduction to Data Science."

5. Tuesday, September 8: The Roles of a Data Scientist

6. Thursday, September 10: R Workshop 1

Assignments for Next Class:

- Read: Chapter 3-5 of "Introduction to Data Science."
- Complete: R Exercise 1

7. Tuesday, September 15: Functions and Packages

Due Today: R Exercise 1 (D2L Dropbox by 9:00am)

8. Thursday, September 17: R Workshop 2

Assignment for Next Class:

- Read: Chapter 6-9 of "Introduction to Data Science."
- Complete: R Exercise 2

9. Tuesday, September 22: Exploratory Data Analysis

Due Today: R Exercise 2 (D2L Dropbox by 9:00am)

10. Thursday, September 24: R Workshop 3

Assignment for Next Class:

- Read: Chapter 2 of "Doing Data Science." (excluding the RealDirect case study)
- Complete: R Exercise 3

11. Tuesday and Thursday, September 29, October 1: **No Class**

Section 2. Special Topics

In this second section, we will encounter a variety of special topics in data science: data visualization, APIs, text analysis, and algorithms. I will simulate real world data science environments by having you work on case studies in small groups.

12. Tuesday, October 6: Data Visualization

Due Today: R Exercise 3 (D2L Dropbox by 9:00am)

13. Thursday, October 8: Group Case 1
Assignment for Next Class:
 - Complete: Group Case 1
14. Tuesday, October 13: Case Presentations
Due Today: Group Case 1 (D2L Dropbox by 9:00am)
15. Thursday, October 15: APIs
16. Tuesday, October 20: **No Class**
17. Thursday, October 22: R Workshop 4
Assignment for Next Class:
 - Read: Chapter 10-11 of "Introduction to Data Science."
 - Complete: R Exercise 4
18. Tuesday, October 27: Text Analysis
Due Today: R Exercise 4 (D2L Dropbox by 9:00am)
19. Thursday, October 29: R Workshop 5
Assignment for Next Class
 - Read: Chapter 12-13 of "Introduction to Data Science."
 - Complete: R Exercise 5
20. Tuesday, November 3: Algorithms
Due Today: R Exercise 5 (D2L Dropbox by 9:00am)
21. Thursday, November 5: Group Case 2
Assignment for Next Class
 - Read: Chapter 3 of "Doing Data Science."
 - Complete: Group Case 2
22. Tuesday, November 10: Case Presentations
Due Today: Group Case 2 (D2L Dropbox by 9:00am)

Section 3. Final Projects

In this final section, you will have a chance to be creative and to show your data prowess by conducting your own novel data science investigation. You will collect your own data, use the tools you acquired in the course to analyze these data, and produce a write-up and presentation to outline your methods and findings.

23. Thursday, November 12: Final Workshop 1
24. Tuesday, November 17: Final Workshop 2
25. Thursday, November 19: Final Workshop 3
26. Tuesday and Thursday, November 24, 26: **No Class (Happy Thanksgiving!)**
27. Tuesday, December 1: Final Presentations 1
Due Today: Final Presentation (D2L Dropbox by 9:00am)
28. Thursday, December 3: Final Presentations 2
29. Tuesday, December 8: Course Overview
Due Today: Final Paper (D2L Dropbox by 9:00am)