Syllabus - Honors Introduction to Data Science (IDS 3932) 3 credits

Instructor: Kevin Lanning and Terje Hoim
Term: Spring 2019
KL Office: WB 213; Office hours TBA
TH Office: HC 150, Office hours TBA
Class Meeting Days/Time: MW 930-1150
Class Location: AD 102
Phone: 561-799-8652
Email: lanning@fau.edu, thoim@fau.edu
instructor website: wise.fau.edu/~lanning
course website http://bit.ly/IntroDataSci2019

course description and objectives

IDS 3932 is an introductory seminar on data science. Hochster (in Hicks & Irizarry, 2017) describes two broad types of data scientists: Type A (Analysis) data scientists, whose skills are like those of an applied statistician, and Type B (Building) data scientists, whose skills lie in problem solving or coding, using the skills of the computer scientist. Our course is closer to a Type A than a Type B treatment, one which is closer to Statistics than to Computer Science, but it is also essentially concerned with content in the concentrations in the arts, humanities, and natural and social sciences. It is thus best understood as a third (Type C) approach, one which has as its objectives progress not just in the understanding of statistics and computing, but also in skills such as collaboration and communication, in exposure to the methods and tools of reproducible science, and in fostering a heightened sensitivity to the ethical challenges of the digital age.

The course will be taught using the statistical and graphical language R. In addition to R, we'll use a range of other tools, including the Slack platform for communication and collaboration, markdown editors such as Typora for writing, and spreadsheets such as Excel or Google Sheets. As noted on the syllabus, these tools will be used in service of a hierarchy of goals, ranging from literacy through proficiency and fluency to leadership.

note of honors distinction

The course receives honors credit by virtue of its small class size, by virtue of a dialectic approach in the classroom structure, and by the fact that students receive extensive exposure to supplementary materials and primary sources. This course differs substantially from a non-Honors course in that (a) the expectations for participation in class discussions will be greater than in a typical undergraduate course with a larger number of students, (b) class projects will be undertaken in heterogeneous groups in which students will be teaching and learning from their peers as well as the instructor, (c) assignments and expectations will be, to some extent, tailored to the backgrounds and interests of the individual student (d) the data sets we will collaboratively examine will be chosen to foster disciplinary breadth.

course prerequisites/co-requisites

STA 2023 (or equivalent) or permission of instructor is a prerequisite.

required texts and materials

Wickham, H. & Grolemund, G. (2016) *R for data science*. Sebastopol, CA: O'Reilly or online at http://r4ds.had.co.nz. (Referred to in syllabus as R4DS).

Lanning, K. (in prep). *Data Science for the Liberal Arts.* https://kevinlanning.github.io/DataSciLibArts. (Referred to in schedule as *dsla*).

A laptop computer running either Windows or Mac OS. (If you don't have access to a laptop for everyday use, please see the instructor as soon as possible).

In addition, there are a number of sources which we are likely to access at least occasionally (see references, below). **All of these materials are presently available online, as well as in a Dropbox directory which will be made available to you.**

requirements / course evaluation and grading

Grades will be based on a 100point scale, with points earned by participation, homework and quizzes, a term project, and a final exam.

Participation (25 points). Attendance is a necessary but not sufficient part of class participation. Your participation grade will be based also on the extent to which you contribute to our class by asking constructive questions and helping us solve the numerous challenges which we collectively will face.

Homework/quizzes (35 points). These are linked: most homework projects will be submitted as in-class quizzes on the assigned date. The first example of this will be in the second class.

A term project (25 points). Learning is social. The term project will be a collaborative, data-based project, which you will undertake with two to four of your peers and which you will submit as a fully-contained R markdown document, that is, as a reproducible document which includes your argument, commented code, and the results of statistically appropriate analyses. These results will typically include tables and/or figures (data visualizations). The project will be empirical, typically from data that I provide you with or we find together. The datasets that we will be working with will be small enough to analyze on your laptops in R.

In order for us to assess your individual contributions and to minimize social loafing, I ask that all meetings and communications among group members be undertaken on the Slack platform, and that, in addition to the paper, all group members sign a 1-page cover page describing the primary contribution and percent effort of each person. We'll work together on creating groups that will, hopefully, maximize synergies among you, that is, how much you learn from each other and the quality of the final project. Groups and paper topics will be developed in class. You'll present your project in class in the final week of the term.

A final exam (15 points). This will include questions about the other group projects as well as what you have learned in class and from the readings.

Above and beyond (10 points). You'll have the opportunity to earn an additional 10 points by solving one or more **data challenges** that we will develop as the class goes forward.

<table>
<tr><th colspan="3">course grading scale</th></tr>
<tr><td>A 93 or more</td><td>C 73 to 76</td><td rowspan="6">note that in borderline cases, students may receive the higher of two grades if there is evidence of sustained effort and/or improvement over the course of the term</td></tr>
<tr><td>A- 90 to 92</td><td>C- 70 to 72</td></tr>
<tr><td>B+ 87 to 89</td><td>D+ 67 to 69</td></tr>
<tr><td>B 83 to 86</td><td>D 63 to 66</td></tr>
<tr><td>B- 80 to 82</td><td>D- 60 to 62</td></tr>
<tr><td>C+ 77 to 79</td><td>F 59 or less</td></tr>
</table>

classroom etiquette policy

In order to enhance and maintain a productive atmosphere for education, personal communication devices, such as cell phones, are to be disabled in class sessions. Laptops and tablets should be used only for appropriate purposes.

attendance policy

As noted above, attendance is expected and will contribute to the portion of grades assessed as "class participation."

policy on accommodations

In compliance with the Americans with Disabilities Act (ADA), students who require reasonable accommodations to properly execute coursework must register with Student Accessibility Services (SAS) -- in Boca Raton, SU 131(561-297-3880); in Davie, LA 131 (954-236-1222); in Jupiter and all Northern Campuses, SR 111F (561-799-8585) – and follow all SAS procedures.

policy on late work, makeup tests, and incompletes

late work will not be accepted. Make-up examinations will be provided in the case of medical excuse or dire need. Incomplete grades will not be granted unless judged absolutely necessary.

code of academic integrity policy statement

Students at Florida Atlantic University are expected to maintain the highest ethical standards. Academic dishonesty is considered a serious breach of these ethical standards, because it interferes with the university mission to provide a high quality education in which no student enjoys an unfair advantage over any other. Academic dishonesty is also destructive of the university community, which is grounded in a system of mutual trust and places high value on personal integrity and individual responsibility. Harsh penalties are associated with academic dishonesty. For more information, see University Regulation 4.001.

policy recognizing that we are human (counseling and psychological services)

Life as a university student can be challenging physically, mentally and emotionally. Students who find stress negatively affecting their ability to achieve academic or personal goals may wish to consider utilizing FAU's Counseling and Psychological Services (CAPS) Center. CAPS provides FAU students a range of services – individual counseling, support meetings, and psychiatric services, to name a few – offered to help improve and maintain emotional well-being. For more information, go to http://www.fau.edu/counseling/

topics / schedule

A screenshot of the current schedule is given below, which you should disregard. (Instead, see **http://bit.ly/DataSciSched2019** for the latest updates). **With the exception of scheduled exam and paper due dates, all content is subject to change.**

Please do the readings prior to class. This is particularly important when we have visitors, either in person or via Skype who will be discussing their work with us.

date	section	module(s) in *Data Science for the*	other readings and links	other notes
1/7	Intro	0 - preface and 1 - data science for the		
1/9	Intro	2 - getting started	dsla 1 and 2	simple quiz and pretest
1/14	Intro	3 - an introduction to R; begin 4 - draw	r4ds 1	another possible quiz
1/16	Literacy	4 - draw the rest of the owl (continued)		" etc.
1/23	Literacy	5 - principles of data visualization		
1/28	Literacy	6 - visualization in R with ggplot	r4ds 2 and 3	
1/30	Literacy	6 - ggplot continued		
2/4	Literacy	7 - Statistics: Probability and inference		
2/6	Literacy	8 - the most dangerous equation	Wainer (2009)	
2/11	Literacy	9 - Reproducibility and Open Science	https://www.nature.com/news/	
2/13	Proficiency	10 - Literate programming with R markdown	Four chapters in r4ds (several are short): 6, 8, 27 and 3	assignment0214
2/18	Proficiency	11 - Tricks from the tidyverse	https://github.com/hadley/data	assignment0219
2/20	Proficiency	12 - Messy data: Cleaning and curation	R4DS 7; 9 (an introduction), 10.1 and 10.2 (introducing tibbles), and 11.1 - 11.2 (reading	
2/25	Proficiency	13 - Selecting data in R with dplyr	R4DS 5.1 - 5.4. (Be prepared to answer qs from exercises)	Data challenge I: Most dangerous?
2/27	Proficiency	14 - Manipulating data in R with dplyr	https://github.com/fivethirtyeight/data	Begin reviewing datasets on 538.com or elsewhere for your group project
3/11	Proficiency	15 - Working with relational data		Groups formed for projects
3/13	Proficiency	16 - Working with strings, factors, dates,	r4ds chapters 19-21	
3/18	Fluency	17 - working with lists		see March19.rmd on slack
3/20	Fluency	18 - Writing loops and functions		see March21.rmd on slack
3/25	Fluency	19 - Multiple regression		
3/27	Fluency	20 - From regression to classification	r4ds chapters 22-23	
4/1	Fluency	21 - Classification and machine learning		see April 2.rmd on slack
4/3	Fluency	22 - Modeling: A computational		review, annotate, and
4/8	Fluency	23 - Making interactive visualizations	https://shiny.rstudio.com/tutori	
4/10	Leadership	24 - Humility and teamwork, 25 - Ethics		
4/15	Leadership	presentations		papers due :)
4/17	Leadership	presentations		
4/22	Leadership	presentations		
5/2	final exam			
add schedule - MPFI, PB Post, Magic Leap, WHC alums, Pat McDonald (New College), Brian MacDonald (Panthers)				