

LIS 490: INTRODUCTION TO DATA SCIENCE

FALL 2016

Professor

Victoria Stodden

Office: 314 GSLIS Building at 501 E. Daniel St.

Email: vcs@stodden.net

Office Hours: Tuesdays 4 - 5pm

TAs

Kenney Guo

Office: Shared TA office Room 204 (or Computer Lab Room 12) in the GSLIS Building

Email: jguo24@illinois.edu

Office Hours: Mondays and Wednesdays 9:30am - 10:30am

Xiaomian Wu

Office: Shared TA office Room 204 (or Computer Lab Room 12) in the GSLIS Building

Email: xwu64@illinois.edu

Office Hours: Mondays and Wednesdays 4pm - 5pm

Class meets Tuesdays, 9 – 11:50am, in Room 126 of the GSLIS Building

Course Description: This course is intended to introduce students to modern programs and technologies that are useful for organizing, manipulating, analyzing, and visualizing data. We start with an overview of the R language, which will become the foundation for your work in this class. Then we'll move on to other useful tools, including working with regular expressions, basic UNIX tools, XML, and SQL. We'll also introduce supervised and unsupervised statistical learning techniques made possible by recent advances in computing power. This course is very computer-oriented, so it's very important to take the time outside of class to learn by doing – to explore the software we'll be covering in class, and try out new skills on real datasets in the homework assignments. Note: the course has a fast pace and moves quickly week to week. You will be expected to stay with the material throughout the semester.

SYLLABUS (subject to revision)

The goal of the course is to familiarize students with computing tools and methods to conduct data science in an applied setting.

The course will include computer exercises using *R* and other relevant computing languages. No experience or comfort with programming or computing is assumed nor is it necessary. Instruction will be given from basic steps.

Week 1: (Aug 23) What is data science? / Getting started with R (Readings and Lecture Notes)

- Week 2: (Aug 30) Data Types / Vectors / Data Manipulation (Lecture Notes)
- Week 3: (Sept 6) Graphics and Data Visualization (Lecture Notes)
- Week 4: (Sept 13) Lists and Apply Functions (Lecture Notes)
- Week 5: (Sept 20) Supervised learning 1: Linear Methods for Regression (ESL 3.1 and 3.2)
- Week 6: (Sept 27) Writing Functions in R / Life Cycle of Data (Lecture Notes, TDO chapter 5)
- Week 7: (Oct 4) Simulation / Bootstrap (Lecture Notes)
- Week 8: **(Oct 11) In-class Midterm (online students write in class too)**
- Week 9: (Oct 18) Regular Expressions / Unix Shell Scripting (Lecture Notes)
- Week 10: (Oct 25) Web Scraping / Unsupervised Learning: Association Rules, Clustering, / Organizing Information (Lecture Notes, ESP 14.1 - 14.3, TDO chapter 8)
- Week 11: (Nov 1) Representation of Information / Text Data Classification (Lecture Notes, TDO chapter 7)
- Week 12: (Nov 8) Introduction to SQL (Lecture Notes)
- Week 13: (Nov 15) Supervised Learning 2: Classification, Decision Trees (TDO chapter 8)
- Week 14: (Nov 22) No class - Thanksgiving
- Week 15: (Nov 29) Revisiting the Life Cycle of Data / Review (Lecture Notes)
- Week 16: **(Dec 6) Final Exam in class (online students write in class)**

Note: You must attend the final. No travel exceptions will be given.

Course Resources

There is no textbook for the class. Notes and reference materials will be made available throughout the semester.

We will use R as the primary computational environment. Some downloadable R manuals:

[Using R for Data Analysis and Graphics - Introduction, Examples and Commentary](#) by John Maindonald

[R for Beginners](#) by Emmanuel Paradis
[An Introduction to R](#)
[R Data Import/Export](#)
[R Language Definition](#)

We will use parts of these books, downloadable and/or available in the library:

[Elements of Statistical Learning](#), 2nd Edition, by Hastie, Tibshirani, Friedman (ESL)
[The Discipline of Organizing: Informatics Edition](#), 4th Edition, by Robert J. Glushko (TDO)

Course Requirements

- Weekly homeworks, 40% (your lowest homework score will be dropped)
- In-class Midterm 30%
- Final 30%
- Active class participation may bump you to the next letter grade if you're close (ie. B+ to A-)

Both the midterm and final are closed notes.

Academic Integrity

Please review and reflect on the academic integrity policy of the University of Illinois, http://admin.illinois.edu/policy/code/article1_part4_1-401.html to which we subscribe. By turning in materials for review, you certify that all work presented is your own and has been done by you independently.

If, in the course of your work, you use the words or ideas of another writer, proper acknowledgement must be given. Not to do so is to commit plagiarism, a form of academic dishonesty. If you are not absolutely clear on what constitutes plagiarism and how to cite sources appropriately, now is the time to learn. Please ask me!

Please be aware that the consequences for plagiarism or other forms of academic dishonesty will be severe. Students who violate university standards of academic integrity are subject to disciplinary action, including a reduced grade, failure in the course, and suspension or dismissal from the University.

Statement of Inclusion

<http://www.inclusiveillinois.illinois.edu/chancellordivstmtswf.html#ValueSt>

As the state's premier public university, the University of Illinois at Urbana-Champaign's core mission is to serve the interests of the diverse people of the state of Illinois and beyond. The institution thus values inclusion and a pluralistic learning and research environment, one which we respect the varied perspectives and lived experiences of a diverse community and global workforce. We support diversity of world views, histories, and cultural knowledge across a range of social groups including race, ethnicity, gender identity, sexual orientation, abilities, economic class, religion, and their intersections.

Disability Statement

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak St., Champaign, call 333-4603 (V/TTY), or e-mail a message to disability@uiuc.edu.