

University of Kentucky
School of Library & Information Science (SLIS)

LIS 690: Introduction to Data Science

Spring 2016 (January 13 to April 29)

Instructor

Youngseek Kim

Assistant Professor

Office: 331 Little Library Building

Phone: (859) 218 – 2295

Email: youngseek.kim@uky.edu and Canvas Messages (Preferred)

Response Time: Within 24 hours during weekdays (expect a delay during weekends and holidays)

Office Hours

Wednesdays: 9:00 a.m. – 1:00 p.m.

and by appointments

Virtual office hours are available during the regular office hours via Adobe Connect

Course Information

This is an online course. Please visit <https://uk.instructure.com> (Canvas) for course homepage.

COURSE DESCRIPTION

This course will provide a foundation in the area of data science based on data curation and statistical analysis. The primary goal of this course is for students to learn data analysis concepts and techniques that facilitate making decisions from a rich data set. Students will investigate data concepts, metadata creation and interpretation, general linear method, cluster analysis, and basics of information visualization. At the beginning, this course will introduce fundamentals about data and data standards and methods for organizing, curating, and preserving data for reuse. Then, we will focus on the inferential statistics: drawing conclusions and making decisions from data. This course will help students understand how to use data analysis tools, and especially, provide an opportunity to utilize an open source data analysis tool, *R*, for data manipulation, analysis, and visualization. Finally, in this course we will discuss diverse issues around data including technologies, behaviors, organizations, policies, and society.

COURSE OBJECTIVES

Upon successful completion of this course, students should have developed some or all of the following areas of skills and knowledge:

- An understanding of how the nature of the data collection, the data itself, and the analysis processes relate to the kinds of inferences that can be drawn
- Understand the limitations of data sets based on their contents and provenance
- Knowledge of data organization, management, preservation, and reuse
- Knowledge of what statistical analysis techniques to choose, given particular demands of inference and available data
- Knowledge of general linear models and cluster analysis methods for statistical analysis
- Skills and knowledge in preparing data for analysis, including cleaning data, manipulating data, and dealing with missing data
- Skills in actually analyzing data using open source data analysis tools
- Skills in scripting for data manipulation, analysis, and visualization using *R*, *R-Studio*, and a variety of add on packages.

COURSE MATERIALS

Required Textbook:

Jeffrey M. Stanton (2013). Introduction to Data Science. Available for free in the iTunes bookstore or as a PDF download at <http://surface.syr.edu/istpub/165/>.

Additional readings:

For some weeks, there will be additional readings to the textbook, and they will be made available on the Canvas course site.

COURSE CONDUCT

The course includes *online* lectures, *online* discussions and exercises, case studies, assignments, and project. As this is an online course with no set meeting times, the question arises when are assignments due each week. For the purposes of this class we will treat Monday as the first day of class each week. Also, readings should be completed by at least Wednesday in order to participate in online discussions. Regarding the weekly discussion, you need to answer some discussion questions and discuss them (see more information below). The discussion board is asynchronous, meaning that you can join in the discussion whenever it is convenient for you to do so.

Canvas as the Main Platform of the Course

This class uses Canvas as a required and main part of the course. Teaching materials (syllabus, course notes, discussions, assignments, resources, etc.) will be made available in the Canvas. All assignments should also be submitted to the Canvas. Students can check grading status and progress in the Canvas. Please visit the Canvas Information Pages at <http://www.uky.edu/canvas/> to learn about the Canvas. For technical support, call the UKIT Service Desk at (859) 218-HELP (4357) or email helpdesk@uky.edu.

Communications:

All course related communications (online discussions, queries on assignments, etc.) will occur within the Canvas. Please post your questions on the Canvas discussion board because other students may have the same questions and receive the benefits from answers. Important announcements will be made inside the Canvas. Students thus are required to check Canvas on a regular basis. Failure to receive such announcements cannot be used as an excuse for not being informed.

I welcome emails sent to my UKY.EDU email account and/or Canvas Messages account. I prefer to use the Canvas Messages in order to keep all course related emails in one place to facilitate communication; however, please feel free to send me any email message to my UKY.EDU email account. Please do not expect an immediate response on your email message. However, in ordinary circumstances, it is expected that the instructor will respond within 24 hours during weekdays. Please expect a delay during weekends and holidays.

METHODS OF ASSESSMENT

Your final grade is determined by your performance on the items in the table below. Assignments and individual project will be assigned as the course continues. Lastly, we will have a weekly online discussion.

Each assignment will be posted at least two weeks before its due date. Please submit your assignments and project through Canvas assignment drop-box. The table below shows the list of assignments, project, and participation and the points for each component.

Type	Topic	Points	Sub Total
Assignments	Assignment #1: Two Topical Presentations (5.0 points × 2)	10.0	45.0
	Assignment #2: Locating Large Datasets	5.0	
	Assignment #3: Developing Research Questions	5.0	
	Assignment #4: Data Science for Library	15.0	
	Assignment #5: Conducting Statistical Analysis with R	10.0	
Project	Final Project: Data Analysis Report	25.0	25.0
Participation	Online Discussions (2 points per each week)	30.0	30.0
		Total: 100.0	

Online Discussions:

Students are expected to participate in the weekly discussion (or exercise) via Canvas Discussions. I will pose several discussion questions (and/or one or two exercises), so you can have the discussions based on the questions (and/or exercises). In addition, there will be a course content discussion thread, where you can talk about each week's course material.

The discussion will start on Monday morning, and it will end on the following Sunday night. I will leave the discussion forum open later, but I am going to evaluate your postings during the week ONLY. You can earn up to 2.0 points for each week's discussion (and/or exercises). You can earn up to 0.5 point for any valuable posting regarding the discussion questions which are posed for each week, your own question regarding each week's course content, answers for the questions posed by other students. This means that you need to post at least 4 posting a week in order to get 2.0 points for that week.

A valuable posting or quality posting includes a substantive and thoughtful contribution to each week's discussion topics, during that week. No credit will be given for posts that occur after the week. A quality posting is both substantive (in most instances this means at least 100 words) and thoughtful ("I agree with the author" only is not a credit-worthy response). Also, please write each discussion posting concisely (100 to 150 words – 1 or 2 paragraphs). I encourage you to complete your discussion posts and other work in Word and then paste it to Canvas. If you compose online and there is a technology-related failure, you will likely lose your work.

Grade Expectations:

Grades are based on the quality of the submitted work, not upon how well others performed. The following are grade expectations and divisions.

Grade	Score (Percentage)	Expectation
A	90% - 100%	Exceptional Achievement
B	80% - 89%	High Achievement
C	70% - 79%	Average Achievement
E	0% - 69%	Failing

Late assignment policy:

Some of the assignments will be discussed in following week's online discussions after the assignments are due. Most assignments will help build a base for future assignments and the project. Thus all assignments should be turned in on time as specified. An overdue assignment will get a penalty of 20% of total points for each day late. No assignment and project will be accepted after five days except 'excused absences' (see more information in Ethics & Policies below).

RE-GRADING REQUESTS

The grade for each assignment is recorded in the Canvas before the assignment is returned to the student. It is a student's responsibility to ask questions or request re-grading of an assignment within five business days from the time the assignment is returned. No re-grading requests will be accepted after the five business day period.

ETHICS & POLICIES

Excused Absences and Verification: Please refer to Student Rights and Responsibilities, Part II, Section 5.2.4.2 (<http://www.uky.edu/StudentAffairs/Code/part2.html>) for UK's policy on excused absences. You can request verification for excused absences.

Excused absences include (as defined at the web site above):

- Significant illness of student or serious illness of household member or immediate family
- Death of a household member or immediate family
- Trips for members of student organizations, class excursions or participation in intercollegiate athletic events
- Major religious holidays
- Any other circumstance that the instructor finds reasonable cause for nonattendance

Academic Accommodations: If you have a documented disability that requires academic accommodations, please see me as soon as possible. In order to receive accommodations in this course, you must provide me with a Letter of Accommodation from the Disability Resource Center (Room 2, Alumni Gym, 257-2754, jkarnes@uky.edu) for coordination of campus disability services available to students with disabilities. We can then collaborate on the best solution.

Academic Integrity, Cheating and Plagiarism: You are expected to submit your own original work for all assignments in this course. See the home page for the Office of Academic Ombud Services (<http://www.uky.edu/Ombud>) for a definition of plagiarism, how to avoid plagiarism and UK's new academic offense policy. Please refer to Student Rights and Responsibilities, Part II, Section 6.3 (<http://www.uky.edu/StudentAffairs/Code/part2.html>) for UK's policy on academic integrity.

Classroom Behavior, Decorum and Civility: Please be respectful to others in the class and engage in civil discourse when we discuss topics that have a diversity of perspectives. Please help me maintain the most courteous environment by using a little peer pressure if necessary.

TECHNOLOGY INFORMATION & RESOURCES

Distance Learning Students are expected to have a minimum level of technological acumen and the availability of technological resources. Students must have regular access a computer with a reliable Internet connection and audio capabilities. Internet Explorer 7 (IE) or Firefox 2.x are the recommended browsers for those using a Windows-based PC. Those using Firefox 3.x may encounter problems with assignment uploads. Those using an Apple computer with MAC OS X (10.5.x) may use Firefox 3.x or Safari 3.x.

Please be certain that your computer and/or browser allow you to view Adobe Reader documents (.pdf). Microsoft Office and other software products are free for students: <https://iweb.uky.edu/MSDownload/>.

As your instructor, I am your first go-to person for technology problems. If you need more immediate

assistance, please contact UKIT.

Contact information for Distance Learning programs (<http://www.uky.edu/DistanceLearning>) and Information Technology Customer Service Center (<http://www.uky.edu/UKIT/Help/>; 859-218-HELP).

Information Technology Customer Service Center (UKIT)

<http://www.uky.edu/UKIT/>; 859-218-4357

Library Services

Distance Learning Services

<http://www.uky.edu/Libraries/DLLS>

- Carla Cantagallo, DL Librarian
- Local phone number: (859) 257-0500, ext. 2171; long-distance phone #: (800) 828-0439 (option #6)
- Email: dllservice@email.uky.edu
- DL Interlibrary Loan Service: http://www.uky.edu/Libraries/libpage.php?lweb_id=253&llib_id=16

GENERAL COURSE POLICIES

Policies concerning academic integrity, excused absences and academic accommodations due to disability are available online at:

<http://ci.uky.edu/sis/sites/default/files/policies.pdf>

Military Members and Veterans

We recognize the complexities of being a member of the military community and also a student. If you are a member of the military or a military veteran or dependent, please inform your instructor if you are in need of special accommodations. Drill schedules, calls to active duty, mandatory training exercises, complications with GI Bill disbursement, and other unforeseen military and veteran related developments can complicate your academic life. If you are aware of a complication, we will work with you and put you in contact with university staff members who are trained to assist you. Please contact the Coordinator of the University of Kentucky Veterans Resource Center at (859) 257-1148 for additional assistance. Visit <http://www.uky.edu/veterans> for more available resources.

TENTATIVE CLASS SCHEDULE (AS OF 1/1/2016)

Week	Date	Topic	Reading	Assignments & Project
1	1/13	Course Introduction		
2	1/19	Definition of Data Science (1/18: Martin Luther King Birthday)	pp. 3-7	
3	1/25	Essential Concepts of Data	Ch. 1	
4	2/1	Data Problems and Solutions	Ch. 2	Assignment #2: Locating Large Datasets (2/7)
5	2/8	R Installation and Basics	Ch. 3	
6	2/15	Data Modeling and Relationship	Ch. 4	
7	2/22	Data Structure and Variables	Ch. 5	Assignment #3: Developing Research Questions (2/28)
8	2/29	Descriptive Statistics	Ch. 6	
9	3/7	Sampling Distribution	Ch. 7	
10	3/14	Spring Break – Enjoy!		Assignment #4: Data Science for Library (3/20)
11	3/21	Big Data and Statistics (t-test)	Ch. 8	
12	3/28	Linear Regression	Ch. 16	
13	4/4	Cluster Analysis	Ch. 17-18	Assignment #5: Conducting Statistical Analysis with R (4/10)
14	4/11	Analysis of Text (Unstructured Data)	Ch. 10-13	
15	4/18	Relevant Issues in Data Science	Ch. 14-15	
16	4/25	Course Summary & Reflections		Final Project (5/1)

* The due dates for the Assignment #1: Topical Presentation varies by individual students.

** All the submission time is 11:59 p.m.