University of Kentucky
School of Information Science (SIS)

LIS/ICT 661: Introduction to Data Science
Fall 2020 (August 17 to November 24)

Instructor   **Office Hours**
Youngseek Kim   *Virtual office hours are available via Zoom*
Email: youngseek.kim@uky.edu
Response Time: Within 24 hours during weekdays (expect a delay during weekends and holidays)

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 S. Saltz & Jeffrey M. Stanton (2018). An Introduction to Data Science (First Edition). Sage Publications. ISBN-13: 978-1506377537; ISBN-10: 150637753X

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 (PPT slides), *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 <u>Monday as the first day of class each week</u>. Also, readings should be completed by at least Wednesday in order to participate in online discussions. I will put up course notes (PPT slides) no later than midnight on Sunday of each week. 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 Learning 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 Information Technology Services (ITS) at (859) 218-HELP (4357) or email 218help@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. <u>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.</u>

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 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
	Assignment #1: Developing Research Questions	5.0	35.0
	Assignment #2: Descriptive Statistics Exercise	10.0	
	Assignment #3: Inferential Statistics Exercise	10.0	
	Assignment #4: Text Analysis Exercise	10.0	
Project	Final Project: Data Analysis Report	30.0	30.0
Participation	Topical Presentation	5.0	35.0
	Participation & Discussions (2 points per each week)	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 a few 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 0.7 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. You can earn up to 2.0 points (not 2.1 points) for each week's discussion (and/or exercises). This means that you need to post at least "three" quality postings in order to get a full credit (i.e., 2.0 points) for each week's discussion.

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

Very often in the assignments and project, you need to offer opinions. Simply stating your opinion does not constitute a complete response. You must support any opinion with arguments and evidence. For example, the question "compare and contrast different organization structures" might be interpreted by a student as follows (note the associated grades A, B, C, or E):

- A: Describe commonly used organization structures, noting where they are similar and different. Identify the important characteristics of an organization structure and systematically evaluate each of the listed organization structures on these dimensions. Suggest situations in which each might be most or least appropriate.
- B: Describe commonly used organization structures, noting where they are similar and different.
- C: Describe commonly used organization structures, one at a time.
- E: Write down anything you can think of about organization structure in no particular order. Avoid giving conclusions, but if you do, be sure they are not supported by anything you have written.

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 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.

Technology Information and Requirements
Technology Requirements
Minimum technical requirements for UK courses and suggested hardware, software, and internet connections are available at ITS Student Hardware & Software Guidelines.

Technical Support
For account help, contact UK's Information Technology Customer Services online, by email, or by phone at 859-218-HELP (4357).

Resources
Distance Learning Library Services
Carla Cantagallo, Distance Learning Librarian, 859-218-1240

Midterm Grades for Undergraduate Students (Senate Rules 6.1.3.A)
Mid-term grades will be posted in myUK by the deadline established in the Academic Calendar.

Excused Absences (Senate Rules 5.2.4.2)

Senate Rules 5.2.4.2 defines the following as acceptable reasons for excused absences: (a) significant illness, (b) death of a family member, (c) trips for members of student organizations sponsored by an educational unit, trips for University classes, and trips for participation in intercollegiate athletic events, (d) major religious holidays, (e) interviews for graduate/professional school or full-time employment post-graduation, and (f) other circumstances found to fit "reasonable cause for nonattendance" by the instructor of record. Students should notify the professor of absences prior to class when possible.

If a course syllabus requires specific interactions (e.g., with the instructor or other students), in situations where a student's total EXCUSED absences exceed 1/5 (or 20%) of the required interactions for the course, the student shall have the right to request and receive a "W," or the Instructor of Record may award an "I" for the course if the student declines a "W." (Senate Rules 5.2.4.2.1)

(If an attendance/interaction policy is not stated in the course syllabus or the policy does not include a penalty to the student, the Instructor cannot penalize the student for any unexcused absences.)

Verification of Absences (Senate Rules 5.2.4.2.A, B, C, and E)
Students may be asked to verify their absences in order for them to be considered excused. *Senate Rule 5.2.4.2* states that faculty have the right to request appropriate verification when students claim an excused absence due to: significant illness; death in the household, trips for classes, trips sponsored by an educational unit and trips for participation related to intercollegiate athletic events; and interviews for full-time job opportunities after graduation and interviews for graduate and professional school. (Appropriate notification of absences due to University-related trips is required prior to the absence when feasible and in no case more than one week after the absence.)

Academic Policies in relation to COVID-19
Fall Academic Calendar and Reading Days
Several modifications to the academic calendar have been made in response to the COVID-19 situation. The current calendar is available here.
The calendar features a Reading Day. The current Dead Week restrictions on certain instructional activities would continue to apply to Reading Days. An additional restriction would apply to Reading Days, namely no required class meetings or, more generally, no "required interactions." Reading Days are not academic holidays.

Religious Observances (Senate Rules 5.2.4.2.D)
Students anticipating an absence for a major religious holiday are responsible for notifying the instructor in writing of anticipated absences due to their observance of such holidays. Please check the course syllabus for the notification requirement. If no requirement is specified, two weeks prior to the absence is reasonable and should not be given any later. Information regarding major religious holidays may be obtained through the Ombud's website or calling 859-257-3737.

Make-Up Work (Senate Rule 5.2.4.2)
Students missing any graded work due to an excused absence are responsible: for informing the Instructor of Record about their excused absence within one week following the period of the excused absence (except where prior notification is required); and for making up the missed work. The instructor must give the student an opportunity to make up the work and/or the exams missed due to the excused absence, and shall do so, if feasible, during the semester in which the absence occurred. The instructor shall provide the student with an opportunity to make up the graded work and may not simply calculate the student's grade on the basis of the other course requirements, unless the student agrees in writing.

Accommodations Due to Disability
If you have a documented disability that requires academic accommodations, please see me as soon as

possible during scheduled office hours. In order to receive accommodations in this course, you must provide me with a Letter of Accommodation from the Disability Resource Center (DRC). The DRC coordinates campus disability services available to students with disabilities. Visit the DRC website, email the DRC, contact them by phone at (859) 257-2754, or visit their office on the corner of Rose Street and Huguelet Drive in the Multidisciplinary Science Building, Suite 407.

Non-Discrimination Statement and Title IX Information

UK is committed to providing a safe learning, living, and working environment for all members of the University community. The University maintains a comprehensive program which protects all members from discrimination, harassment, and sexual misconduct. For complete information about UK's prohibition on discrimination and harassment on aspects such as race, color, ethnic origin, national origin, creed, religion, political belief, sex, and sexual orientation, please see the electronic version of UK's *Administrative Regulation 6:1* ("Policy on Discrimination and Harassment") . In accordance with Title IX of the Education Amendments of 1972, the University prohibits discrimination and harassment on the basis of sex in academics, employment, and all of its programs and activities. Sexual misconduct is a form of sexual harassment in which one act is severe enough to create a hostile environment based on sex and is prohibited between members of the University community and shall not be tolerated. For more details, please see the electronic version of *Administrative Regulations 6:2* ("Policy and Procedures for Addressing and Resolving Allegations of Sexual Assault, Stalking, Dating Violence, Domestic Violence, and Sexual Exploitation"). Complaints regarding violations of University policies on discrimination, harassment, and sexual misconduct are handled by the Office of Institutional Equity and Equal Opportunity (IEEO), which is located in 13 Main Building and can be reached by phone at (859) 257-8927. You can also visit the IEEO's website.

Faculty members are obligated to forward any report made by a student related to IEEO matters to the Office of Institutional Equity and Equal Opportunity. Students can *confidentially* report alleged incidences through the Violence Intervention and Prevention Center, Counseling Center, or University Health Services.

Academic Integrity– Prohibition on Plagiarism (Senate Rules 6.3.1)

Per University policy, students shall not plagiarize, cheat, or falsify or misuse academic records. Students are expected to adhere to University policy on cheating and plagiarism in all courses. The minimum penalty for a first offense is a zero on the assignment on which the offense occurred. If the offense is considered severe or the student has other academic offenses on their record, more serious penalties, up to suspension from the University may be imposed.

Plagiarism and cheating are serious breaches of academic conduct. Each student is advised to become familiar with the various forms of academic dishonesty as explained in the Code of Student Rights and Responsibilities. Complete information can be found on the Academic Ombud page. A plea of ignorance is not acceptable as a defense against the charge of academic dishonesty. It is important that you review this information as all ideas borrowed from others need to be properly credited.

Senate Rule 6.3.1 (see current *Senate Rules*) states that all academic work, written or otherwise, submitted by students to their instructors or other academic supervisors, is expected to be the result of their own thought, research, or self-expression. In cases where students feel unsure about a question of plagiarism involving their work, they are obliged to consult their instructors on the matter before submission.

When students submit work purporting to be their own, but which in any way borrows ideas, organization, wording, or content from another source without appropriate acknowledgment of the fact, the students are guilty of plagiarism.

Plagiarism includes reproducing someone else's work (including, but not limited to a published article, a

book, a website, computer code, or a paper from a friend) without clear attribution. Plagiarism also includes the practice of employing or allowing another person to alter or revise the work, which a student submits as his/her own, whoever that other person may be. Students may discuss assignments among themselves or with an instructor or tutor, but when the actual work is done, it must be done by the student, and the student alone.

When a student's assignment involves research in outside sources or information, the student must carefully acknowledge exactly what, where and how he/she has employed them. If the words of someone else are used, the student must put quotation marks around the passage in question and add an appropriate indication of its origin. Making simple changes while leaving the organization, content, and phraseology intact is plagiaristic. However, nothing in these Rules shall apply to those ideas, which are so generally and freely circulated as to be a part of the public domain.

Please note: Any assignment you turn in may be submitted to an electronic database to check for plagiarism.

Academic Integrity – Prohibition on Cheating (Senate Rules 6.3.2)

Cheating is defined by its general usage. It includes, but is not limited to, the wrongfully giving, taking, or presenting any information or material by a student with the intent of aiding himself/herself or another on any academic work which is considered in any way in the determination of the final grade. The fact that a student could not have benefited from an action is not by itself proof that the action does not constitute cheating. Any question of definition shall be referred to the University Appeals Board.

Academic Integrity – Prohibition on Falsification/Misuse of Academic Records (SR 6.3.3)

Maintaining the integrity, accuracy, and appropriate privacy of student academic records is an essential administrative function of the University and a basic protection of all students. Accordingly, the actual or attempted falsification, theft, misrepresentation or other alteration or misuse of any official academic record of the University, specifically including knowingly having unauthorized access to such records or the unauthorized disclosure of information contained in such records, is a serious academic offense. As used in this context, "academic record" includes all paper and electronic versions of the partial or complete permanent academic record, all official and unofficial academic transcripts, application documents and admission credentials, and all academic record transaction documents. The minimum sanction for falsification, including the omission of information, or attempted falsification or other misuse of academic records as described in this section is suspension for one semester.

Youngseek Kim

TENTATIVE CLASS SCHEDULE (AS OF 8/16/2020)

Week	Date	Lecture	Reading (Chapters)	Due Date
1	8/17	Course Introduction		
2	8/24	Basics of Data Science	1 & 2	
3	8/31	Data Modeling	3, 4, & 5	Assignment #1: Developing Research Questions (9/6)
4	9/7	Data Collection and Measurement	6, 7, & 8	
5	9/14	Descriptive Statistics	9	
6	9/21	Probability and Statistics	10	Assignment #2: Descriptive Statistics Exercise (9/27)
7	9/28	Statistical Inference	11	
8	10/5	Linear Regression	16	
9	10/12	Data Visualization	12 & 13	Assignment #3: Inferential Statistics Exercise (10/18)
10	10/19	Text Mining	14 & 15	
11	10/26	Cluster Analysis	17	
12	11/2	Data Mining	18	Assignment #4: Text Analysis Exercise (11/8)
13	11/9	Data Curation	19	
14	11/16	Privacy, Ethics, and Transparency	20	
15	11/23-24	Course Summary and Reflection		Final Project (11/29)

*The due dates for the Topical Presentation varies by individual students.
**All the submission time is 11:59 p.m.