LIS662-401/ICT662-401 (Cross-listed)
Data Analysis and Visualization
Fall 2018

Instructor:	Dr. Soohyung Joo
Office Address:	#337 Little Library Bldg
Email:	soohyung.joo@uky.edu
Office Phone:	859-257-5942
Office hours:	Tuesday and Thursday, 9:00 AM to 11:00 AM
Preferred method of contact:	Email/ Canvas.
	I usually respond to questions within 24 hours. (expect a delay during weekends and holidays).
Class meetings:	Thursday 6pm – 9pm, Rm 311 Little Library Bldg.

Course Description

This course examines three major categories of topics in relation to data analysis and visualization. First, this course will cover the basic ways that data can be obtained from various sources, such as raw text files, web APIs, and data repositories. It will also cover the techniques of data cleaning and how to organize data for analysis. Second, the course will cover the essential techniques for analyzing quantitative data. It will teach prediction and clustering methods that are useful to solve various real data analysis tasks. In addition, students will learn major theories and recent methods in text analysis. Third, this course teaches how to create visualizations that effectively communicate the meanings behind data and information. The course will cover key practical skills in information visualization, such as plotting, mapping, and network visualization. This course will not be mathematically intensive. Instead, the course will use existing computational tools and programming libraries to solve various problems. You will use the R language and environment intensively for data analysis and visualization.

Prerequisites

None

Student Learning Outcomes

After completing this course, the student will be able to:
1. Explain basic concepts and major methods in data analytics.
2. Obtain different types of data from various sources, including data repositories, social media and web and publicly open data (e.g., governmental websites).
3. Think critically about data and identify appropriate methods to solve given problems.
4. Apply appropriate data analysis procedures and visualization techniques to draw conclusions from such analyses.
5. Utilize graphical and numerical summaries to effectively represent analysis results.
6. Identify appropriate visualization methods and apply them to different types of data.
7. Implement a variety of R packages to analyze and visualize data.
8. Interpret and author advanced level R programming.

Required Materials
No textbook.
Extensive lecture notes will be provided every week. All course reading materials including lecture notes will be uploaded to Canvas.
A student will need to install the R software toolkit in his/her own computer (https://www.r-project.org/). The instructor will help students install the R software in the first week class.

Class Information
This is a face-to-face course. You are required to attend scheduled classroom sessions. You will need access to an appropriate laptop computer with a Wi-Fi Internet connection. (Please contact the instructor as soon as possible, if you cannot bring your own laptop. The instructor can help you check out a laptop computer from the School of Information Science.)

Course Activities and Assignments
Your final grade is determined by your performance on the following items: (1) six assignments (60%), (2) a final project – presentation and report (30%), and (3) class participation (10%).

Summary Description of Course Assignments
- Assignment #1 (10%) – Collecting data from Web APIs
- Assignment #2 (10%) – Text processing, term frequency, and word clouds
- Assignment #3 (10%) – Topic modeling
- Assignment #4 (10%) – Clustering analysis
- Assignment #5 (10%) – Network analysis and visualization
- Assignment #6 (10%) – Nonlinear analysis and visualization
- Final Project (30%) – Presentation (5%) and Report (25%).
- Class participation (10%)

Course Grading
Grades are based on the quality of the submitted work, not upon how well others performed. The following are grade expectations and divisions. Grading scale for this course:

    90-100% = A
    80 – 89.9% = B
    70 – 79.9% = C
    Below 70%= E

Tentative Course Schedule (Subject to Change)

Week	Date	Topics
1	8/23	Course overview, Introduction to R programing

2	8/30	Getting and cleaning data (Data format in R, XML)
3	9/6	Getting and cleaning data (Web APIs, JSON)
4	9/13	Text processing in R Visualization for text mining
5	9/20	LDA topic model and sentiment analysis
6	9/27	Clustering analysis and visualization
7	10/4	Classification – Naïve Bayes, Support vector machine
8	10/11	Classification – Logistics regression, Neural networks
9	10/18	Plotting linear and non-linear data
10	10/25	Network analysis and visualization (1)
11	11/1	Network analysis and visualization (2)
12	11/8	Geospatial data visualization
13	11/15	Case studies in data science Introduction to Culturomics
14	11/22	Thanksgiving
15	11/29	Final project presentation and feedback
16	12/6	Final project presentation and feedback

Submission of Assignments
All assignments should be submitted online via Canvas. An overdue assignment will get a penalty of 10% of total points for each day late. No assignment and project will be accepted after seven days.

Attendance Policy
This is an online course, which requires asynchronous class discussion via Canvas to facilitate a sense of community. Canvas will also be used for making course announcements, distributing reading materials, submitting assignments and posting grades. Please visit https://www.uky.edu/canvas (Canvas) for course homepage.

Course related communications will occur mainly in Canvas (online discussions, questions and answers, etc.). In addition, all announcements will be posted in Canvas. Students thus are required to check the course shell in 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. Please feel free to email me via soohyung.joo@uky.edu if you have any questions or concerns. In ordinary circumstances, it

is expected that the instructor will respond within 24 hours during weekdays. Please expect a delay during weekends and holidays.

Excused Absences
Students need to notify the professor of absences prior to class when possible. *Senate Rules 5.2.4.2* defines the following as acceptable reasons for excused absences: (a) serious illness, (b) illness or death of family member, (c) University-related trips, (d) major religious holidays, and (e) other circumstances found to fit "reasonable cause for nonattendance" by the professor.

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 no later than the last day in the semester to add a class. Two weeks prior to the absence is reasonable, but should not be given any later. Information regarding major religious holidays may be obtained through the Ombud (859-257-3737, http://www.uky.edu/Ombud/ForStudents_ExcusedAbsences.php.

Per *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 professor must give the student an opportunity to make up the work and/or the exams missed due to an excused absence, and shall do so, if feasible, during the semester in which the absence occurred.

Verification of Absences
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 because of illness, or death in the family. 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 Integrity
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 at the following website: http://www.uky.edu/Ombud. 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 Rules 6.3.1 (see http://www.uky.edu/Faculty/Senate/ for the current set of *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.

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. It is located on the corner of Rose Street and Huguelet Drive in the Multidisciplinary Science Building, Suite 407. You can reach them via phone at (859) 257-2754 and via email at drc@uky.edu. Their web address is http://www.uky.edu/StudentAffairs/DisabilityResourceCenter/.

Policies concerning academic integrity, excused absences and academic accommodations due to disability are available online at:
https://ci.uky.edu/sis/sites/default/files/policies.pdf

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: http://download.uky.edu/.

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

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

For more resources about online classes and student resources, visit http://www.uky.edu/ukonline/

The School of Information Science has a page with a comprehensive list of technology resources here: http://ci.uky.edu/sis/students/techtips

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/veteransfor more available resources.