University of Kentucky
School of Information Science (SIS)

[ICT/LIS 661-201]: Introduction to Data Science

Instructor: Spencer Greenhalgh, PhD

Contact Information: *Office*—341 Lucille Little Library Bldg.
Phone—859.218.2294
Email—spencer.greenhalgh@uky.edu
Zoom—https://uky.zoom.us/my/greenhalgh

Response Time: During the work week (but not the weekend!), I commit to respond to all emails within 24 hours. I expect you to regularly check Canvas and email for messages from me and to respond quickly.

Office Hours: I will hold office hours in person or on Zoom (see above):
Tuesday, 1pm to 3pm,
Wednesday, 12pm to 2pm, or
by appointment

Meeting Schedule: This course takes place asynchronously on Canvas

Required Materials: *An Introduction to Data Science*
Jeffrey S. Saltz & Jeffrey M. Stanton
(ISBN: 978-1506377537)

Any other course materials will be available on Canvas.

COVID-19 (and More) Statement [inspired by Dr. Andrew Heiss]

Grad school can be stressful enough during normal times, but it's likely worse during a global pandemic. You may know (or be) someone who has lost their job, tested positive for COVID-19, been hospitalized, or taken on new family responsibilities. Not only this, but there have been several events over the past 24 months that are reminders of inequalities and injustices people here in Kentucky and around the world are facing. This is a difficult time to be enrolled in classes.

Despite these difficulties, I am fully committed to making sure that you learn everything you were hoping to learn from this class! My late policy and willingness to make accommodations are generous even during normal times, and if this pandemic (or anything else) is turning your life upside down, I'm willing to be as flexible as you need me to be.

If you feel like you're behind, not understanding everything, or just plain stressed, **do not suffer in silence!** I'm usually quick to respond to email and more than happy to meet with you.

Basic Needs Statement [inspired by Dr. Sara Goldrick-Rab]

Any student who has difficulty affording or accessing food to eat every day or who lacks a safe and stable place to live and believes this may affect their performance in the course is urged to contact the Dean of Students and to explore the resources listed at the bottom of this page. Furthermore, please notify me if you are comfortable in doing so.

Course Information

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, the general linear model, 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:
- understanding of how the nature of the data collection, the data itself, and analysis processes relate to the kinds of inferences that can be drawn
- understanding of 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, RStudio, and a variety of add on packages.

Course Assessment
Your grade for this course will be based on 100 points:

　　　　90 points – 100.0 points = A
　　　　80 points –　89.9 points = B
　　　　70 points –　79.9 points = C
　　　　　0 points –　69.9 points = E

These 100 points come from the following assessment activities, which should all be completed honestly and individually on Canvas:

Projects
Throughout the semester, you will complete five projects worth a total of 65 points:
- Project #1: Developing Research Questions (5 points)
- Project #2: Descriptive Statistics (10 points)
- Project #3: Inferential Statistics (10 points)
- Project #4: Text Analysis (10 points)
- Final Project: Data Analysis Report (30 points)

Detailed instructions for these projects can be found on Canvas.

Participation
Throughout the semester, you will earn 35 points from a series of participation activities:
- once during the semester, you will submit a Topical Presentation (a 3-5 paragraph overview of a topic related to data science or data curation) for me and your classmates to read and discuss (5 points)
- during each of the fifteen modules of the semester, you will complete two one-point participation activities that will help you extend your understanding of or practice applying course content (30 points total)

Late Work Policy
Officially, each assignment is due at 11:59pm on the Sunday night indicated in Canvas. Practically speaking, however, I will grade without penalty (for graded assessments) and provide feedback on (for all assessments) any assessment that is turned in by the time I begin looking over that assessment. However, I will not grade or provide feedback on any work that is completed after this time unless you have made other arrangements with me. Naturally, because my schedule varies from week to week and because I try to provide feedback as quickly as possible, your best bet is to turn in your work by the official deadline or—if life has thrown you a curveball—to get in touch with me ahead of time to make other arrangements.

Course Policies

All of the policies listed on this page are in effect for this course. Also in effect are these rules related to academic offenses; plagiarism is a serious concern, and I will not tolerate it. It is your responsibility to ensure that you understand what plagiarism is and how to avoid it.

Diversity, Equity and Inclusion

The University of Kentucky is committed to our core values of diversity and inclusion, mutual respect and human dignity, and a sense of community (Governing Regulations XIV). We acknowledge and respect the seen and unseen diverse identities and experiences of all members of the university community (https://www.uky.edu/regs/gr14). These identities include but are not limited to those based on race, ethnicity, gender identity and expressions, ideas and perspectives, religious and cultural beliefs, sexual orientation, national origin, age, ability, and socioeconomic status. We are committed to equity and justice and providing a learning and engaging community in which every member is engaged, heard, and valued.

We strive to rectify and change behavior that is inconsistent with our principles and commitment to creating a safe, equitable, and anti-racist environment. If students encounter such behavior in a course, they are encouraged to speak with the instructor of record or the college's diversity officer, who is charged with addressing concerns about diversity, equity, and inclusiveness (uky.edu/inclusiveexcellence/college-diversity-inclusion-officers). Students may also contact a faculty member within the department, program director, the director of undergraduate or graduate studies, the department chair, or the dean. To submit an official report of bias, hatred, racism, or identity-based violence, visit the Bias Incident Support Services website (https://www.uky.edu/biss/report-bias-incident).

Please also consider the following resources related to diversity, equity, and inclusion:

Bias Incident Support Services
Bias Incident Support Services (BISS) provides confidential support and advocacy for any student, staff, or faculty member impacted by bias, hatred, and/or an act of identity-based violence. BISS staff aid impacted parties in accessing campus and community resources, including the Bias Incident Response Team, the University's official reporting system for acts that negatively impact a sense of belonging. Campus and community consultation and educational opportunities centered on inclusion, diversity, equity and belonging is a resource also provided by BISS. For more detailed information please visit the BISS website (uky.edu/biss) or contact them via email (biss@uky.edu).

Counseling Center
The UK Counseling Center (UKCC) provides a range of confidential psychological services to students enrolled in 6 credit hours or more, psychoeducational outreach programming (including QPR suicide prevention), and consultation to members of the UK community (students, faculty, staff, administrators, parents, concerned others). Please visit the UKCC's website (uky.edu/counselingcenter) for more detailed information or call (859) 257-8701.

Disability Resource Center

If you have a documented disability that requires academic accommodations, please inform your instructor as soon as possible during scheduled office hours. In order to receive accommodations in this course, you must provide your instructor 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, via email (drc@uky.edu) or visit their website (uky.edu/DisabilityResourceCenter).

Martin Luther King Center

The Martin Luther King Center (MLKC) supports an inclusive learning environment where diversity and individual differences are understood, respected, and appreciated as a source of strength. The MLKC's year-round programs and activities that focus on the importance of cultural awareness and cross-cultural understanding support its three primary goals: 1) sponsoring cultural and educational programming; 2) offering opportunities for student support and development; and 3) through programmatic linkages with a wide variety of civic and community agencies, promoting community outreach, engagement, and collaboration. Students can reach the MLKC via phone at (859) 257-4130, by visiting them in Gatton Student Center Suite A230, via email (mlkc@uky.edu), and by visiting the MLKC website (uky.edu/mlkc).

If there are aspects within your experience here at UK that result in barriers to your inclusion or accurate assessment of achievement, please notify the instructor as soon as possible and/or email the Office for Institutional Diversity via email (vpid@uky.edu).

Non-Discrimination / Title IX

In accordance with federal law, 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") (https://www.uky.edu/regs/ar6-1). 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") (https://www.uky.edu/regs/ar6-2). Complaints regarding violations of University policies on discrimination, harassment, and sexual misconduct are handled by the Office of Institutional Equity and Equal Opportunity (Institutional Equity), which is located in 13 Main Building and can be reached by phone at (859) 257-8927. You can also visit Institutional Equity's website (https://www.uky.edu/eeo).

Faculty members are obligated to forward any report made by a student related to discrimination, harassment, and sexual misconduct to the Office of Institutional Equity. Students can confidentially report alleged incidences through the [Violence Intervention and Prevention Center](https://www.uky.edu/vipcenter) (https://www.uky.edu/vipcenter), [Counseling Center](https://www.uky.edu/counselingcenter) (https://www.uky.edu/counselingcenter), or [University Health Service](https://ukhealthcare.uky.edu/university-health-service/student-health) (https://ukhealthcare.uky.edu/university-health-service/student-health).

Reports of discrimination, harassment, or sexual misconduct may be made to Institutional Equity [here](#).

Office of LGBTQ* Resources

UK is committed to supporting students and upholding the University's efforts to promote inclusion among our community. UK faculty and staff employees support inclusion and diversity throughout the University, including the ways in which faculty structure classroom conversations and manage those dynamics. To assist in these efforts, students are welcome to provide the names and pronouns they prefer. One easy way to do this is by using the pronoun feature of UK's Name Change Form. (More information about the form can be found on the [Office of LGBTQ*'s website](#) (uky.edu/lgbtq/forms-and-resources).) Otherwise, students can provide this information to instructors directly.

Discrimination based on sexual orientation, gender expression, and gender identity is prohibited at UK. If you have questions about support, advocacy, and community-building services related to sexual orientation, gender expression, or gender identity, students are encouraged to visit the [website of the Office of LGBTQ* Resources](#) (uky.edu/lgbtq/forms-and-resources).

Veterans Resource Center (VRC)

Being both a member of the military community and a student can bring some complexities. If you are a member of the military or a military veteran or dependent, please let instructors know when these challenges arise. Drill schedules, calls to active duty, mandatory training exercises, issues with GI Bill disbursement, etc. can complicate your academic life. Let your instructor know if you experience complications.

The VRC is a great resource for members of our military family. If you have questions regarding your VA benefits or other related issues, the VRC has a full complement of staff to assist you. The VRC also provides study and lounge space, as well as free printing. Please visit the [VRC website](#) (uky.edu/veterans), [email the VRC](mailto:vetcenter@uky.edu) (vetcenter@uky.edu), visit them in the basement of Erikson Hall, or call the director, Colonel Tony Dotson, at (859) 257-1148.

If you are a military student serving in the National Guard or Reserve, it is in your best interest to let all of your instructors know that immediately. You might also consider sharing a copy of your training schedule.

If you are a military student who is a member of the National Guard or Military Reserve and are called to duty for one-fifth or less of this semester, provide a copy of your military orders to the Director of the Veterans Resource Center (contact information above) once you become aware of the call to duty. (Please also provide the Director with a list of all your current courses and

instructors.) The Director will verify the orders with the appropriate military authority and will, on the military student's behalf, notify their instructors as to the known extent of the absence.

Your absences will not be penalized and instructors will work with military students to create reasonable accommodations for making up missed assignments, quizzes, and tests.

Violence Intervention and Prevention (VIP) Center

If you experience an incident of sex- or gender-based discrimination or interpersonal violence, we encourage you to report it. While you may talk to a faculty member or TA/RA/GA, understand that as a "Responsible Employee" of the University these individuals MUST report any acts of violence (including verbal bullying and sexual harassment) to the University's Title IX Coordinator in the Institutional Equity Office. If you would like to speak with someone who may be able to afford you confidentiality, you can visit the [Violence Intervention and Prevention (VIP) Center's website](uky.edu/vipcenter/content/faq) (uky.edu/vipcenter/content/faq) (offices located in Bosworth Hall, 1st Floor; (859) 257-3574), the [Counseling Center's (CC) website](uky.edu/counselingcenter/student-resources) (uky.edu/counselingcenter/student-resources), and the [University Health Services (UHS) website](uky.edu/university-health-service/student-health/our-student-services) (uky.edu/university-health-service/student-health/our-student-services). The VIP Center, CC, and UHS are confidential resources on campus. **The VIP Center accepts Zoom, phone, and walk-in appointments.**

Course Schedule

Module (Dates)	Topic	Reading Activities	Other Activities
Module 1 (Aug 23 - Aug 29)	Course Introduction	- read and discuss the syllabus	- introduce yourself to the class! - sign up for a topical presentation group
Module 2 (Aug 30 - Sep 5)	Basics of Data Science	- read and discuss "An Introduction to Data Science," chapters 1-2 - read about and discuss the basics of data science	- install R and RStudio - find a dataset
Module 3 (Sep 6 - Sep 12)	Privacy, Ethics, and Transparency	- read about and discuss privacy, ethics, and transparency	- set up GitHub - discuss ethics and privacy in data science
Module 4 (Sep 13 - Sep 19)	Data and Models	- read and discuss "An Introduction to Data Science," chapters 3-5 - read about and discuss data and models - post and discuss topical presentations	- create a vector - create a dataframe - submit Project #1: Developing Research Questions
Module 5 (Sep 20 - Sep 26)	Data Collection and Management	- read and discuss "An Introduction to Data Science," chapters 6-8 - read about and discuss data collection and management - post and discuss topical presentations	- practice manipulating data in R - practice data munging
Module 6 (Sep 27 - Oct 3)	Descriptive Statistics	- read and discuss "An Introduction to Data Science," chapter 9 - read about and discuss descriptive statistics - post and discuss topical presentations	- write a function in R to calculate descriptive statistics - create a histogram
Module 7 (Oct 4 - Oct 10)	Probability and Statistics	- read and discuss "An Introduction to Data Science," chapter 10 - read about and discuss probability and statistics - post and discuss topical presentations	- calculate and plot a sample mean - review chapter 10 procedures

Module 8 (Oct 11 - Oct 17)	Statistical Inference	- read and discuss "An Introduction to Data Science," chapter 11 - read about and discuss statistical inference - post and discuss topical presentations	- perform a t-test with your own data - review chapter 11 procedures
Module 9 (Oct 18 - Oct 24)	Linear Regression	- read and discuss "An Introduction to Data Science," chapter 16 - read about and discuss linear regression - post and discuss topical presentations	- perform a linear regression - review chapter 16 procedures
Module 10 (Oct 25 - Oct 31)	Data Visualization	- read and discuss "An Introduction to Data Science," chapters 12-13 - read about and discuss data visualization - post and discuss topical presentations	- practice using ggplot2 - practice using ggmap
Module 11 (Nov 1 - Nov 7)	Text Mining	- read and discuss "An Introduction to Data Science," chapters 14-15 - read about and discuss text mining - post and discuss topical presentations	- create a word cloud in R - perform a sentiment analysis in R
Module 12 (Nov 8 - Nov 14)	Cluster Analysis	- read and discuss "An Introduction to Data Science," chapters 17 - read about and discuss cluster analysis - post and discuss topical presentations	- perform cluster analysis in R - perform association rules mining in R
Module 13 (Nov 15 - Nov 21)	Data Mining	- read and discuss "An Introduction to Data Science," chapter 18 - read about and discuss data mining - post and discuss topical presentations	- develop a Support Vector Machine (SVM) - research data mining applications
Module 14 (Nov 22 - Nov 28)	Data Curation	- read and discuss "An Introduction to Data Science," chapter 19 - read about and discuss data curation - post and discuss topical presentations	- develop (hypothetical) metadata - reflect on the data lifecycle
Module 15 (Nov 29 - Dec 8)	Course Summary and Reflection	- read and discuss "An Introduction to Data Science," chapter 20 - read and discuss a course summary - post and discuss topical presentations	- reflect on what you've learned this semester