STEM 680: Foundations of Educational Data Science I

Fall 2024

Key Information

Meeting Time and Place

Mondays, 12:00 - 2:30 pm (online and synchronous; see schedule below for specific class dates)

Credit Hours

3

Faculty Contact Information

Dr. Hanall Sung, hanallsung@utk.edu

Course Links

Zoom: https://tennessee.zoom.us/my/hanall (no password!)

Canvas: https://utk.instructure.com/courses/213028

Slack: eddatascienceatutk.slack.com

Course Description

Intended to support graduate-level students to be able to apply data science methods to topics of teaching, learning, and educational systems. Introduces students to the data science software and programming language R. Course activities focusing on preparing, using, and visualizing complex data sources for analysis using the tidyverse suite of R packages. Data ethics are foregrounded. Includes an introduction to text analysis/Natural Language Processing. No prerequisites or programming experience is required.

A key element of this class is that students will have the opportunity to bring their own data from their research projects for use in this class. In this way, they will have immediate application for the concepts learned in the course. If no data is immediately available from the student's research, students can use one of hundreds of freely available datasets to complete coursework or students can use datasets provided to them.

In all, this course will provide scaffolding to help students become proficient in a few sophisticated data science techniques, and it will give students sufficient foundational knowledge to pick up new data science skills on their own after the course is through. This course will serve as a foundation for later data science in education, including the second foundations class, data visualization, and machine learning and the capstone course.

Learning Objectives

The objectives for the proposed course are for students to be able to:

- Install, set up, and use R and RStudio
- Use reproducible workflows (so that analyses can easily be modified and then carried out again by the analyst or others) with R Markdown

- Develop foundational skills - focused around the tidyverse R packages - to prepare and explore data sources for analysis

- Understand how issues of equity, privacy, and ethics are central to data science in education

- Develop a personal learning and development plan related to data science in education

- Begin a portfolio of work from this class that you can add to later

- Pursue an independent project to work toward a relevant professional goal

Format and Learning Environment

This class will be taught in a fully-online format. We will use Zoom for synchronous (or at-the-same-time) class. We will also use a number of tools for asynchronous communication, including a) Slack, and b) features of the Canvas course learning management system.

Communication and Late policy

You will generally receive a response to messages within 24 hours during the work week (Monday - Friday). We ask for you to please try to respond within 24 hours during the work week, too. You can contact me via email (above) or Slack (preferred).

Don't hesitate to ask questions! Learning to do data science is challenging for everyone, and reaching out for support and assistance is imperative.

Late assignments will deduct 5% of your final grade for each day they are received after the due date, with a maximum penalty of 50%. If an assignment is submitted 10 days after its due date (resulting in a 50% penalty), the late penalty will not be increased.

Required Equipment

You will need a computer (Mac, Windows, or Linux are fine!) on which you can install applications, but you do not need a computer with any particular specifications (speed, storage, etc.) beyond what you use for other courses: whatever you have will work for this course.

Grading Scale and Course Grading Scheme

Grading Scale

LETTER GRADE	PERCENTAGE
A	93.01-100
A-	90.01-93
B+	87.01-90
B	83.01-87
B-	80.01-83
C+	77.01-80
C	73.01-77
C-	70.01-73
D	60.01-70
F	60 and below

Course Grading Scheme

ASSIGNMENT	PERCENT OF GRADE	POINTS
Readings	12.5%	10 / week for 15 weeks = 150 points
Weekly Assignments	32.5%	30 / week for 13 weeks = 390 points
Professional Development Plan	5%	60
Mini Project	12.5%	150
Data Ethics Statement	12.5%	150
Final Project	25%	300
Total:	100%	1,200

Learning Activities

- **Participation: weekly classes.** Each class will have a consistent structure.

  - *Complete reading, discussion, and any assignment(s)* before class

  - *Answer the eliciting question* in groups at the beginning of each class and discuss as a class (30 minutes)

  - Listen, answer questions, and code-along with the *introducing new ideas* portion of class (30 minutes)

  - *Code-along* using built-in data to get a feel for the code you will be using (30 minutes)

  - Time to *start on the assignment for the next week with a peer/peers* (30 minutes)

  - Ask any questions or have independent work time (30 minutes)

- **In-class programming.** At the core of this class is programming in R. We will develop R programming skills for data wrangling, exploration, and visualization together by doing various in-class programming activities. We will complete these activities as a whole class, in small groups, or in pairs. Semi-structured activities will give us a chance to discuss, better understand, and practice our programming skills.

- **Weekly assignments**: Weekly tasks that involve combining reading about relevant theory and prior research, working through fundamentals in a guided practice model. These will be submitted as R documents, HTML files, or images.

- **Mini project:** This independent project will involve the application of theory and programming to create various visualizations from an already-existing data set. Your work will be shared with other students and the instructor to provide you with opportunities to provide and receive constructive critique (and to revise your work, as is the case with all visualizations!).

- **Data ethics statement:** You will explore visualizations created by others for #tidytuesday and apply newly learned skills together with the theory and programming learned in class to a provided data set to create various visualizations from a provided data set. Your work will be shared with other students and the instructor to provide you with opportunities to provide and receive constructive critique and revise your work.

- **Professional development plan:** Develop a plan for your continued professional data science learning.

- **Final project:** You will complete a final project that involves developing visualizations for your own data or a data set of your choice. The goal of this project is to create a publication-ready visualization that demonstrates what you have learned throughout the course.

Class and University Policies

Generative Artificial Intelligence

Generative artificial intelligence (GAI) tools can be incredibly helpful when programming in R and other languages, offering quick solutions and code suggestions. However, in the early stages of learning to program, it is essential to understand the underlying principles and logic of the language. This foundational knowledge empowers you to critically evaluate and audit the code generated by GAI, as these tools can sometimes produce errors or inefficient code. By mastering the basics yourself, you can use GAI tools effectively while ensuring the accuracy and quality of your work. Therefore, I do not prohibit their use, but ask that you include a note with any submissions in which you have used a GAI tool, explaining how you utilized it. This will provide context to me and help me learn how these tools can be most useful.

Academic Integrity

An essential feature of the University of Tennessee, Knoxville is a commitment to maintaining an atmosphere of intellectual integrity and academic honesty. As a student of the university, I pledge that I will neither knowingly give nor receive any inappropriate assistance in academic work, thus affirming my own personal commitment to honor and integrity.

Title IV Policy

University of Tennessee faculty are committed to supporting our students and upholding gender equity laws as outlined by Title IX.

Please be aware that if you choose to confide in a faculty member regarding an issue of sexual harassment (including sexual assault, dating violence, domestic violence, and stalking), sexual exploitation, and retaliation (prohibited conduct) we are obligated to inform the University's Office of Title IX. They can assist you in connecting with all possible resources both on- and off-campus.

If you would like to speak with someone confidentially, the Student Counseling Center (865-974-2196) and the Student Health Center (865-974-3135) are both confidential resources.

For additional resources and information, visit titleix.utk.edu.

University Civility Statement

Civility is genuine respect and regard for others: politeness, consideration, tact, good manners, graciousness, cordiality, affability, amiability and courteousness. Civility enhances academic freedom and integrity, and is a prerequisite to the free exchange of ideas and knowledge in the learning community. Our community consists of students, faculty, staff, alumni, and campus visitors. Community members affect each other's well-being and have a shared interest in creating and sustaining an environment where all community members and their points of view are valued and respected. Affirming the value of each member of the university community, the campus asks that all its members adhere to the principles of civility and community adopted by the campus: http://civility.utk.edu/.

Disability Services

Any student who feels s/he may need an accommodation based on the impact of a disability should contact Student Disability Services in Dunford Hall, at 865-974-6087, or by video relay at, 865-622-6566, to coordinate reasonable academic accommodations.

Your Role in Improving Teaching and Learning Through Course Assessment

It is our collective responsibility to improve the state of teaching and learning. During the semester, you may be requested to assess aspects of this course either during class or at the completion of the class. You are encouraged to respond to these various forms of assessment as a means of continuing to improve the quality of the UT learning experience.

Basic Needs

Any student who faces challenges securing their food or housing and believes they may affect their performance in the course is urged to contact the Dean of Students (974-HELP or viahttps://dos.utk.edu/) for support. Furthermore, please contact the instructor if you are comfortable doing so.

Weekly Schedule and Readings

Please find the weekly schedules and readings here: https://docs.google.com/document/d/19V-QpNu5_Qa rRMWMaWPSlp7wT9jryQrWm___mFqzO-rg/edit?usp=sharing

F24-Founds-EDS Weekly Schedule and Readings

Semester Week	Date	Focal Question and Topic	Substantive Reading(s) to be Complete *Before* Class	Tutorial Reading(s) or Tutorial Activities Due *Before* Class	Assignments Due *By the Next* Class
1	August 19 (Pre-Work)	What's R? Getting Setup	Estrellado, R. A., Freer, E. A., Mostipak, J., Rosenberg, J. M., & Velásquez, I. C. (2020). Data science in education using R. Routledge. Chapter 3, What does data science in education look like? McFarland, D. A., Khanna, S., Domingue, B. W., & Pardos, Z. A. (2021). Education data science: Past, present, future. AERA Open, 7, 233 2858 4211052055.	Primer 1: The Basics Primer 2: Work with Data	Weekly assignment Weekly reading and discussion
2	August 26	What's this course about? Introducing the course and ourselves	Salganik, M. J. (2019). Bit by bit: Social research in the digital age. Princeton University Press. Chapters 2 (Observing Behavior) and 3 (Asking Questions).	Install RStudio R for Data Science, Chapter 8: Cleaning names with Janitor	Weekly assignment Weekly reading and discussion

3	September 2 (Labor day: **Asynchronous**)	Extending what we know	[McCandless, D. (2010). The beauty of data visualization.](#) [Schwabish, J., Feng, A. Do No Harm Guide: Applying Equity Awareness in Data Visualization (summary only)](#) [Spreadsheet error led to Edinburgh hospital opening delay](#)	[R for Data Science, Chapter 8](#): [Cleaning names with Janitor](#)	Weekly assignment Weekly reading and discussion
4	September 9	What's all the fuss about data science? Doing things we could not easily do without *graphics*	Lang, C., Wise, A. F., Merceron, A., Gasevic, D., & Siemens, G. (2022). What is learning analytics? The handbook of learning analytics, 8-18.	[R for Data Science, Chapter 2 ggplot2 tutorial](#) [Data Visualization introduction from Modern Dive](#)	Weekly assignment Weekly reading and discussion
5	September 16	What's all the fuss about data science? Doing things we could not easily do without *coding*	Worsley, M., Martinez-Maldonado, R., & D'Angelo, C. (2021). A New Era in Multimodal Learning Analytics: Twelve Core Commitments to Ground and Grow MMLA. Journal of Learning	[R 4 Data Science, Chapter 20, Joining tables](#) [Joining tables tutorial](#)	Weekly assignment Weekly reading and discussion Weekly reading and

			Analytics, 8(3), 10-27.		discussion
6	September 23	How do I access and describe data? Using an R package and an API	Rutherford et al. (2022) Leveraging mathematics software data to understand student learning and motivation during the COVID-19 pandemic, *Journal of Research on Technology in Education, 54*:sup1, S94-S131, DOI: 10.1080/15391523.2021.1920520 Rosenberg et al. (2022) Posts about students on Facebook: A data ethics perspective	[education data package](#) [R for Data Science, Chapter 3](#)	Weekly assignment Weekly reading and discussion
7	September 30	Creating a good college ranking system	How U.S. News Calculated the 2024 Best Colleges RankingsLinks to an external site. A different kind of college rankingLinks to an external site. With a New Formula, U.S. News Rankings Boost Some State	[Using School-Level Aggregate Data to Illuminate Educational Inequities](#)	Weekly assignment Weekly reading and discussion

			Universities		
8	October 7 (Fall Break: **No Class**)	N/A	N/A	N/A	Professional Development Plan
9	October 14	How do I group and summarize data?	Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., ... & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. Review of Research in Education, 44(1), 130-160. Hakimi, L., Eynon, R., & Murphy, V. A. (2021). The ethics of using digital trace data in education: A thematic review of the research landscape. Review of Educational Research, 91(5), 671-717.	[Summarizing Data by Groups](#)	Weekly assignment Weekly reading and discussion
10	October 21	Do the pivot	Kimmons, R., & Veletsianos, G. (2018). Public internet data mining methods in instructional design, educational technology,	[Data tidying](#)	Weekly assignment Weekly reading and

			and online learning research. TechTrends, 62(5), 492-500.		discussion
11	October 28	Mapping, part 1	Mann, B., & Saultz, A. (2019). The role of place, geography, and geographic information systems in educational research. Aera Open, 5(3), 2332858419869340.	5 Census geographic data and applications in R (Read 5 & 5.1)	Mini Project Weekly assignment Weekly reading and discussion
12	November 4	Mapping, part 2	Mann, B., & Saultz, A. (2019). The role of place, geography, and geographic information systems in educational research. Aera Open, 5(3), 2332858419869340.	5.2 Plotting geographic data (Read 5.2)	Weekly assignment Weekly reading and discussion
13	November 11	Introduction to Text Analysis	Lucy L., Demszky D., Bromley P., Jurafsky D. (2020). Content analysis of textbooks via natural language processing: Findings on gender, race, and ethnicity in Texas US history textbooks. AERA Open, 6(3). <<https://doi>.org /10.1177/23	Introduction to text analysis with the tidytext package	Weekly assignment Weekly reading and discussion

			32858 420940312>		
14	November 18	Topic modeling	Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2021). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. Sociological Methods & Research, 50(1), 202-237.	Topic modeling	Weekly assignment Weekly reading and discussion
15	November 25	Functions and loops	TBD	25 Functions	Data Ethics Statement Weekly assignment Weekly reading and discussion
16	December 2	Final projects	N/A	N/A	Final Project