STEM 680: Foundations of Educational Data Science I

Fall 2023

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. Isabella Velasquez

Dr. Maryrose Weatherton

Course Links

GitHub: https://github.com/utkeds/f23-founds-eds-1

Canvas: https://utk.instructure.com/courses/184613

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 pre-requisites 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, b) GitHub, and c) features of the Canvas course learning management system.

Communication and Late Submission 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 us 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.

Our late assignment policy is that as long as you submit the assignment before we grade it, you will receive full credit. However, we may grade assignments very soon after they are due. For assignments received after the due date, 5% from the grade you otherwise would earn will be subtracted from your final grade for each day late.

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

LETTER GRADE	PERCENTAGE
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 *no reading in Week 16
Weekly Assignments	32.5%	30 / week for 13 weeks = 390 points *no assignment Weeks 1, 9, and 16
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.

Weekly Schedule and Readings

Please find the weekly schedules and readings here: https://docs.google.com/document/d/1cUrX_bpHFNdqvSZU_cw0mwEWuz_3m2Sn7TjIVILAv_E/edit?usp=sharing

Class and University Policies

Generative Artificial Intelligence

Open AI's GPT-4 and other generative artificial intelligence tools like it (e.g., Google's Bard) can be immensely helpful when it comes to programming in R and other languages. We encourage their use, but ask that you add a note to anything you submit in which you have used a generative artificial intelligence tool about how you have used it to provide context to us and to help us to learn how these tools are 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.

F23-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 21 (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 28	What's this course	Salganik, M. J. (2019). Bit by bit: Social research in	Install RStudio	Weekly assignment

		about? Introducing the course and ourselves	the digital age. Princeton University Press. Chapters 2 (Observing Behavior) and 3 (Asking Questions).	R for Data Science, Chapter 8: Cleaning names with Janitor	Weekly reading and discussion
3	September 4	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 2 ggplot2 tutorial Data Visualization introduction from Modern Dive	Weekly assignment Weekly reading and discussion
4	September 11	What's all the fuss about data science? Doing things we could not	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

		easily do without advanced *graphics*			
5	September 18	What's all the fuss about data science? Doing things we could not easily do without *coding*	Worsley, M., Mart inez -Maldonado, R., & D'Angelo, C. (2021). A New Era in Multimodal Learning Analytics: Twelve Core Commitments to Ground and Grow MMLA. Journal of Learning Analytics, 8(3), 10-27.	R 4 Data Science, Chapter 20, Joining tables Joining tables tutorial	Weekly assignment Weekly reading and discussion Weekly reading and discussion
6	September 25	How do I access and describe data? Using an R package and an API	Reich, J. (2022). Learning analytics and learning at scale. Lang, C., Siemens, G. Wise, AF, Gašević, D., Merceron, A.,(eds.) Handbook of Learning Analytics, 2nd ed. Vancouver, Canada: SOLAR. White, R. S. (2023).	educationdata package Optional: Analyzing NCES Data Using EdSurvey: A User's Guide - Chapter 4.11	Weekly assignment Weekly reading and discussion

			Ceilings Made of Glass and Leaving En Masse? Examining Superintendent Gender Gaps and Turnover Over Time Across the United States. Educational Researcher, 0013189X231163139.		
7	October 2	How do I preparing data for analysis F oundational skills, Part 1	Shelton, C. C., Koehler, M. J., Greenhalgh, S. P., & Carpenter, J. P. (2022). Lifting the veil on Teach ersP ayTeachers. com: An in vestigation of educational marketplace offerings and downloads. Learning, Media and Technology, 47(2), 268-287. Rosenberg, J. M., Borchers, C., Burchfield, M. A., Anderson, D., Stegenga, S. M., & Fischer, C. (2022). Posts about students on Facebook: a data ethics p	R for Data Science, Chapter 4	Weekly assignment Weekly reading and discussion

			erspective. Educational researcher, 51(8), 547-550.		
8	October 9 (Fall Break)	How do I group and summarize data? F oundational skills, Part 2	Park, J. J., Kim, B. H., Wong, N., Zheng, J., Breen, S., Lo, P., ... & Poon, O. (2023). Inequality Beyond S tandardized Tests: Trends in Extr acurricular Activity Reporting in College A pplications Across Race and Class. Technical Report. Annenberg Institute at Brown University. <https ://edworkin g papers.>co m/ai23-749.	Data Science in Education, Chapter 9	Weekly assignment Weekly reading and discussion
9	October 16	N/A	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	Animation Plot with gganimate gganimate package	Professional Development Plan

			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.		
10	October 23	How do I access public Internet data? Web scraping	Kimmons, R., & Veletsianos, G. (2018). Public internet data mining methods in in structional design, educational technology, and online learning research. TechTrends, 62(5), 492-500. Staudt Willet, K. B., & Carpenter, J. P. (2021). A tale of two subreddits: Change and continuity in t each ing-related online spaces. British Journal of	Towards Data Science, Web Scraping rvest introduction	Weekly assignment Weekly reading and discussion

			Educational Technology, 52(2), 714-733.		
11	October 30	What have I done?!? Sharing work (Quarto, Rmd and its friends) and Presenting Mini Project	D'ignazio, C., & Klein, L. F. (2020). Data feminism. MIT press. Chaper 1 (In troduction: Why Data Science Needs Feminism)	Quarto with R Creating table with kable and kableExtra	Mini Project Weekly assignment Weekly reading and discussion
12	November 6	What are the more complex skills I need to work with data? F oundational skills, Part 3	Dowell, N. M., Nixon, T. M., & Graesser, A. C. (2019). Group co mmunication analysis: A co mputational linguistics approach for detecting soc iocognitive roles in multiparty in teractions. Behavior Research Methods, 51, 1007-1041.	Iteration using the purrr package, R for Data Science, Chapter 27	Weekly assignment Weekly reading and discussion
13	November	What are the core	Nguyen, H., Lopez, J., Homer, B., Ali, A., & Ahn,	Pivoting Tables, R for Data Science, Chapters 6.3-6.4	Weekly assignment

	13	skills I need to work with data? F oundational skills, Part 4	J. (2023). Reminders, r eflections, and rel ationships: insights from the design of a chatbot for college advising. Information and Learning Sciences, 124(3/4), 128-146.	Using flexdashboard in R	Weekly reading and discussion
14	November 20	How can I prepare for the final project? Learning more and creating data products	D'Mello, S. (2017). Emotional learning analytics. Handbook of learning analytics, 115.	Creating a map with ggplot2	Weekly assignment Weekly reading and discussion
15	November 27	How can I use text as data? Tokens and frequencies	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	Introduction to text analysis with the tidytext package	Data Ethics Statement Weekly assignment Weekly reading and discussion

			Open, 6(3). <<https://doi> .org /10.1177/23 32858 420940312>		
16	December 4	How can I use text as data? Sentiment and term c ooccurences	Rosenberg, J. M., Borchers, C., Dyer, E. B., Anderson, D., & Fischer, C. (2021). Un derstanding public sentiment about educational reforms: The next generation science standards on Twitter. AERA open, 7, 233 2858 4211024261.	Sentiment analysis with tidytext	Weekly assignment Weekly reading and discussion
		Final Projects	N/A		Final Project