GRADTDA 5401: DATA ANALYTICS FOUNDATIONS 1

Instructor Contact Information

Name: **Thomas Metzger**
Office Location: **Cockins 327**
Email: **metzger.181@osu.edu**
Office Hours: **TBD via Zoom**

Course Meeting Information

Method:
1) Required weekly asynchronous video lectures will be posted approximately a week in advance, totaling about one to two hours of content in total each week
2) Optional but recommended synchronous sessions will be held Tuesdays from 6-7 pm for supplementary Q&A, discussion, and questions; these sessions will be recorded and posted within 24 hours

Course Prerequisites

Good standing in the MTDA program. Course enrollment is restricted to graduate students enrolled in the program.

Description of Course:

The successful working professional engaged in modern, real world, data science must be able to extract useful information from data and use that information to address work-related challenges. Central to this endeavor is the focus on data-driven decision making under uncertainty. Data come from multiple sources and in all types of formats. Proficiency in identifying, sourcing, manipulating, and interpreting data is paramount.

This is the first course in a two-semester sequence comprised of two, 3 credit-hour courses focusing on R programing and data analysis using R. The sequence is intendent to be taken in parallel with the Big Data Computing Foundations sequence focusing on fundamental CS methods for data science.

Course Learning Goals:

Upon successful completion of the course, students will be able to:

1. Use the statistical programing language and software environment R and the companion integrated development environment RStudio to import, manipulate and visualize data
2. Clean the data and transform them into

formats amenable to statistical analysis and visualization
3. Derive and interpret numerical and visual summaries of the data
4. Conduct application-driven, exploratory analyses that point in the direction of meaningful, application-specific structure in the data, and facilitate the separation of the information "signal" in the data from the "noise" reflecting pure randomness
5. Use R programing skills to organize the data analysis steps into a streamlined, efficient workflow.
6. Perform basic programing tasks in R
7. Author effective summary reports of the performed analysis using R Markdown

Course Materials and Texts:

Open Source and Free Online Textbooks:

Garrett Grolemund and Hadley Wickham (2017), *R for Data Science* (First Edition), O'Reilly; (https://r4ds.had.co.nz/) "Wickham Text"

Barbara Illowsky and Susan Dean (2013), *Introductory Statistics*, openstax; (https://openstax.org/details/books/introductory-statistics) "Illowsky Text"

Required Software:

- R (www.r-project.org)
- RStudio (www.rstudio.com)

Grading Information:

Grade Breakdown:

Homeworks: 50%
Project 1: 15%
Project 2: 15%
Readings, Discussions, and Reflections: 20%

Grading Scale:

>90%	A/A-
80-90%	B-/B/B+
70-80%	C-/C/C+
60-70%	D/D+
<60%	E

Course Delivery:

Asynchronous (pre-recorded and viewable at students' leisure): each week several lecture videos, totaling approximately 1 to 2 hours of content, will be posted to Carmen. You are responsible for watching the videos and studying the material that is assigned each week. In addition to the lecture videos, weekly assignments will be posted on Carmen.

Synchronous (live): there will also be weekly group sessions over Zoom, Tuesdays from 6-7 pm. The purpose of these is for me to provide feedback on things I have observed in submissions, allow questions and answers that students have, and foster group discussions about the content. I have found that students often lead this time according to their curiosity, needs, and suggestions. These sessions are recommended but optional, and will be recorded and posted to Carmen for students to view at their convenience. Office hours and meetings or discussions between individual students and myself can be scheduled in advance and are optional as needed.

Attendance and Participation:

Because this is a distance-education course, your attendance is based on your online activity and participation. Please check your email and log into Carmen at least once per weekday.

Discussion and Communication:

When posting content visible to myself and/or to other students, please be respectful, thoughtful, and professional.

- Tone and civility: all discussions must contribute to a sense of safety and civility to ensure all students and myself feel comfortable and welcome as part of the course.
- Writing style: good grammar, spelling, and punctuation are important. They reflect on your professionalism and the care you have put into your submissions.
- Cite your sources: in academic discussions, please cite your sources to back up what you say. This may be as simple as links to resources you've found online, or references to page numbers in the textbook.

Assignments and Grades:

Homework will account for 60% of your course grade. They will be assigned and submitted via Carmen. Homeworks will usually comprise reflections, data analysis, coding, and reports summarizing findings.

Projects will account for 40% of your course grade. Projects will be completed individually or in small groups and will entail data analysis, coding, and presentation components.

Health and Safety:

The Ohio State University Wexner Medical Center's Coronavirus Outbreak site (https://wexnermedical.osu.edu/features/coronavirus) includes the latest information about COVID-19 as well as guidance for students, faculty and staff. Guidelines and requirements for campus safety from the University's COVID-19 Transition Task Force were published on July 1, 2020 on the Safe and Healthy website (https://safeandhealthy.osu.edu). They include the following:

Potential Disruptions to Education:

If you are unable to complete content or complete assignments because of positive diagnosis or symptoms, please reach out to me as soon as possible to let me know so we can develop a contingency plan. Typically this will include postponing due dates or developing alternative assignments, and will be done on a case by case basis.

If I am unable to create and post new content due to positive diagnosis or symptoms, I will alert the class as quickly as possible to develop a contingency plan.

Lecture Breakdown:

Week	Statistical Topic	R and Data Science Topic	Textbook Readings
1	**Sampling and Data Populations, parameters, samples, and statistics**	**Introduction to R Booleans, variables, and vectors**	**Illowsky Chapter 1 Wickham Chapters 1, 2, 20**
2	**Descriptive Statistics**	**Loading and exploring data sets Dataframes Simple base R visualizations R Markdown reports**	**Illowsky Chapter 2 Wickham Chapters 3, 26, 27**
3	**Discrete random variables Common discrete probability distributions**	**Factors in R Data tables and matrices Simple ggplot visualizations**	**Illowsky Chapter 4 Wickham Chapters 4, 5, 6, 7**
4	**Continuous random variables Common continuous probability distributions Working with strings in R Writing R functions Introduction to hypothesis testing on means**	**R projects Reproducibility for() loops and apply()**	**Illowsky Chapter 5 Wickham Chapters 8, 9, 10, 11, 12, 21**
5	**The normal distribution**	**Relational data Merges and joins**	**Illowsky Chapter 6 Wickham Chapter 13**
6	**The central limit theorem Sampling distributions**	**Strings and factors in R Introduction to t-tests in R**	**Illowsky Chapter 7 Wickham Chapters 14, 15**

7	**The t distribution Confidence intervals**	**Time and dates in R Confidence intervals in R**	**Illowsky Chapter 8 Wickham Chapter 16**
8	**One sample hypothesis tests Project 1**	**Functions Hypothesis tests in R**	**Illowsky Chapter 9 Wickham Chapters 17, 18, 19**
9	**Two sample hypothesis tests Chi-square distribution and categorical data**	**Data ethics**	**Illowsky Chapters 10, 11**
10	**Type-I Error Type-II Error and power**	**Simulation studies in R**	**Wickham Chapters 22, 23, 24, 25**
11	**Linear regression and correlation**	**Hypothesis testing philosophy and criticism**	**Illowsky Chapter 12**
12	**F distribution and one-way ANOVA**		**Illowsky Chapter 13**
13	**Multiple comparisons Multiple regression and collinearity**		
14	**Variable selection**		
15	**Bootstrapping and cross-validation Project 2**		

Collaboration and Academic Misconduct:

Students are encouraged to **collaborate** remotely on homework assignments, projects, and discussions, but ultimately the work you submit must be your own. Students should work individually on exams. For all assignments, course and online resources may be used but do not solicit specific help from others, such as by posting specific problems online.

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct http://studentlife.osu.edu/csc/.

Disability Services:

The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know

immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

Title IX and Sexual Misconduct:

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at http://titleix.osu.edu or by contacting the Ohio State Title IX Coordinator, Kellie Brennan, at titleix@osu.edu

Diversity and Inclusion:

The Ohio State University affirms the importance and value of diversity in the student body. Our programs and curricula reflect our multicultural society and global economy and seek to provide opportunities for students to learn more about persons who are different from them. We are committed to maintaining a community that recognizes and values the inherent worth and dignity of every person; fosters sensitivity, understanding, and mutual respect among each member of our community; and encourages each individual to strive to reach his or her own potential. Discrimination against any individual based upon protected status, which is defined as age, color, disability, gender identity or expression, national origin, race, religion, sex, sexual orientation, or veteran status, is prohibited.

Mental Health:

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via the Office of Student Life's Counseling and Consultation Service (CCS) by

visiting ccs.osu.edu or calling 614-292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614-292-5766 and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1-800-273-TALK or at suicidepreventionlifeline.org.