



THE OHIO STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

SYLLABUS: GRADTDA 5401 DATA ANALYTICS FOUNDATIONS 1 AUTUMN 2023

Course overview

Instructor

Instructor: Subhadeep Paul

Email address: paul.963@osu.edu

Lectures: Asynchronous video lectures shared via Carmen (1 hour 50 minutes per week) and Synchronously on Zoom **Tuesdays 6:00–6:55 PM EST**

Office hours: Virtual Hours via Zoom (Wednesdays 7:00 – 8:00 PM EST)

Course Website

Carmen (<https://carmen.osu.edu/>)

All course related materials including lecture notes, videos, assignments and announcements will be posted here)

Teaching Assistant

Yumo Peng (peng.704@osu.edu)

TA office hours: TBA

Course description

This course is the first of a sequence of two courses designed to teach you statistical analysis and data science using R. The first part of the course will focus more on visualizing data, performing exploratory statistical analysis and programming with R, while the second part will introduce core concepts of statistics. The course will take a hands-on approach where you will learn by performing data analysis yourself!

Tentative list of topics we will cover:

1. Introduction to R and data science
 - a. Basic R programming, Visualization in base R and ggplot
 - b. Data transformations and manipulations
 - c. Apply, for loops, iterations
 - d. Tibbles, data frames, importing data, merging multiple data tables
 - e. Advanced R programming – custom functions
2. Basic statistical inference
 - a. Graphical and numerical summaries of data
 - b. Introduction to probability, conditional probability and Bayes theorem
 - c. Discrete and continuous random variables, probability distributions
 - d. Simulation using R
 - e. Population, samples, sampling distributions, confidence intervals
3. Regression Analysis
 - a. Simple linear regression
 - b. Multiple linear regression
 - c. Logistic regression

Prerequisite or corequisite:

Must be a graduate student enrolled in the MTDA program.

Course learning outcomes

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

1. Use statistical programming language R and Rstudio to import, clean, and visualize data
2. Derive and interpret graphical and numerical summaries from data
3. Perform exploratory as well as in-depth statistical analysis on data to gain insights about the problem or business process
4. Perform programming tasks using R
5. Estimate population quantities using a sample
6. Quantify uncertainty in estimating parameters and test statistical hypotheses
7. Use regression analysis to understand relationships between two or more variables.

Course materials

Required

We have two required books for the course. Both the books are available to students freely.

- *The R for data science book:* <https://r4ds.had.co.nz/index.html>
- *Introductory Statistics book:* <https://openstax.org/details/books/introductory-statistics>

Additional references (not a required text). These two books are also freely available from their websites:

- *Hands on Programming with R* by Garrett Golemund <https://rstudio-education.github.io/hopr/index.html>
- *R programming for Data Science* by Roger D Peng <https://bookdown.org/rdpeng/rprogdatascience/>.

I will also post lecture notes in Carmen.

Course technology

For help with your password, university e-mail, Carmen, or any other technology issues, questions, or requests, contact the OSU IT Service Desk. Standard support hours are available at <https://ocio.osu.edu/help/hours>, and support for urgent issues is available 24x7.

- **Self-Service and Chat support:** <http://ocio.osu.edu/selfservice>
- **Phone:** 614-688-HELP (4357)
- **Email:** 8help@osu.edu
- **TDD:** 614-688-8743

Baseline technical skills necessary for online courses

- Basic computer and web-browsing skills
- Navigating Carmen

Technology skills necessary for this specific course

- Zoom
- Recording a slide presentation with audio narration
- Typing a solution and submitting a PDF file through Carmen.

Necessary equipment

- Computer: current Mac (OS X) or PC (Windows 10+) with high-speed internet connection
- Webcam: built-in or external webcam, fully installed
- Microphone: built-in laptop or tablet mic or external microphone

Necessary software

- This class requires you to use the statistical software package called R (The R Project for Statistical Computing; <http://www.r-project.org/>). This software package is available as Free Software.
 - You can download R for Windows, Mac, and Linux, from the CRAN archive at <https://cran.r-project.org>.
 - An in-depth introduction to R is available at <http://cran.r-project.org/doc/manuals/R-intro.pdf>
 - This course will teach everything you need to know about R – no background in R is assumed.
- An easier to use interface to R is available in the software package RStudio. This package is available for Windows, Mac, and Linux and can be downloaded for free from <http://rstudio.org>. **Note that RStudio requires R to be installed.**

Course delivery

The course will be **delivered online in two modes**. Every week I will record and upload video lectures of approximately 1 hour 50 minutes in duration. We will also have a 55 minutes synchronous (live) lecture over zoom during the lecture times listed above. The asynchronous content will be posted as videos on the class website in carmen.

You will be responsible for attending live classes and the recorded videos as well as studying the material that is assigned. For the live lectures, attendance is encouraged, but is not required. The lecture will be recorded and posted in carmen for viewing later.

In addition to the lecture videos, assignments will be posted on the class website. You will be given ample time to complete the assignments. The instructor will hold weekly office hours via Zoom. The times are given above.

Grades

Assignment or category	Percentage
Homeworks	40
Projects	45
Take home Exam	15
Total	100

Grading scale

93–100: A
 90–92.9: A-
 87–89.9: B+
 83–86.9: B
 80–82.9: B-
 77–79.9: C+
 73–76.9: C
 70–72.9: C-
 67–69.9: D+
 60–66.9: D
 Below 60: E

Assignment information

Homework: Homework will be assigned regularly (about 5-6 assignments over the semester). It will consist of data analysis and computer programming/simulation problems. You may discuss the problems with your fellow classmates, tutors/TA, and me, but each student must hand in their own work, written in their own words. Do not copy any part of another student's homework including computer output. Use of homework solutions distributed in previous offerings of the course or available on the web constitutes academic misconduct and will be handled according to university rules. **All homework must be typed and submitted online as a PDF file through the class website (carmen).** Please be sure that the questions are clearly labeled, all supporting work (including computer code) can be easily identified, and that all figures/tables are referenced and interpreted in the text.

Projects: In addition to homework assignments, there will be **two (2) data analysis projects** assigned over the semester. The assignment will consist of an oral presentation of results obtained from analyzing a dataset and will require use of the R software. The students are expected to work in groups of 3-4 and each group can select a dataset of their choice to analyze (preferably from the work/education domain of the students to bring more diversity of topics). There will be a fun opportunity to collaborate with the students from Statistical Consulting Course (Stat 6750) on the projects. More details on this will be given later in the class.

Please note assignments submitted **late** will be penalized at 5% per day it is late. Assignments **will not be accepted beyond 3 days of their due date unless prior exception has been sought.** if you are unable to complete an assignment on time, please get in touch with me as soon as possible so we can discuss your situation.

Exams: There will be **one take home exam.** Coverage includes lecture material, assigned reading, and homework. Further details will be given in advance of the take home exam. The format of the exam is similar to other homework assignments and consists of data analysis problems. The rules regarding academic integrity will also be the same as for other homework assignments: you are allowed to discuss the problems with your fellow classmates, tutors/TA, and me. However, you must write your own solution in your own words.

Tentative date of take home exam: Nov 27, 2023- Dec 06, 2023.

Tentative date of Project presentations.

Project 1: Tuesday Oct 10: 6:00pm-6:55 pm.

Project 2: Tuesday Dec 05: 6:00 pm – 6:55 pm.

Faculty feedback and response time

I am providing the following list to give you an idea of my intended availability throughout the course. (Remember that you can call **614-688-HELP** at any time if you have a technical problem.)

Grading and feedback

For weekly assignments, you can generally expect graded homeworks back within **14 days**.

E-mail

I will reply to e-mails related to course logistics within **72 hours on school days**. However, if you have questions on understanding the course materials or regarding the homework assignments please visit my Zoom office hours and the TA's office hours.

Attendance, participation, and discussions

Students may miss class, for a variety of reasons related to COVID-19. As much as possible, please stay in contact with the instructor so that we can discuss accommodations should they be needed.

Student participation requirements

Because this is a distance-education course, your attendance is based on your online activity and participation. The following is a summary of everyone's expected participation:

- **Logging in: AT LEAST ONCE PER WEEK**
Be sure you are logging in to the course in Carmen each week, including weeks with holidays or weeks with minimal online course activity. (During most weeks you will probably log in many times.) If you have a situation that might cause you to miss an entire week of class, discuss it with me *as soon as possible*.
- **In live lectures: REGULAR ATTENDANCE ENCOURAGED BUT NOT REQUIRED.** Students are encouraged to participate, discuss, and answer questions in online live lectures.
- **Recorded video lectures: REQUIRED.** Students are expected to watch the video lectures I will upload regularly at their own time.
- **Office hours: OPTIONAL OR FLEXIBLE**
All office hours, are optional. If you are required to discuss an assignment with me, please contact me at the beginning of the week if you need a time outside my scheduled office hours.

Other course policies

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 on the Safe and Healthy website (<https://safeandhealthy.osu.edu>).

Potential disruptions to instruction

- As much as is possible, students will have access to material online if they are unable to attend class because of positive diagnosis, symptoms, or quarantine required following contact tracing.
- If the instructor is unable to be present in person because of positive diagnosis, symptoms, or quarantine following contact tracing a new instructor will be assigned to the course. Details will be given on the course website

Student academic services

Student academic services offered on the OSU main campus
<http://advising.osu.edu/welcome.shtml>.

Student support services

Student support services offered on the OSU main campus <http://ssc.osu.edu>.

Academic integrity policy

Policies for this online course

- **Exam:** You must complete the exam yourself, without any external help or communication.

- **Written assignments:** Your written assignments should be your own original work. You are of course permitted to work together in assignments as long as you write your solutions yourself.
- **Reusing past work:** In general, you are prohibited in university courses from turning in work from a past class to your current class, even if you modify it. If you want to build on past research or revisit a topic you've explored in previous courses, please discuss the situation with me.
- **Falsifying research or results:** All research you will conduct in this course is intended to be a learning experience; you should never feel tempted to make your results or your research look more successful than it was.
- **Collaboration and informal peer-review:** The course includes many opportunities for formal collaboration with your classmates. The group projects give you opportunity to work together in a group on a project. You are also encouraged to form study groups and peer-review for the projects. While group study for assignments is permitted, remember that your answers on homework assignments should not be an identical copy of another students' answers. If you're unsure about a particular situation, please feel free just to ask ahead of time.

Ohio State's academic integrity policy

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/>.

Copyright disclaimer

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Statement on title IX

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 at titleix@osu.edu

Accessibility accommodations for students with disabilities

The university strives to make all learning experiences as accessible as possible. In light of the current pandemic, students seeking to request COVID-related accommodations may do so through the university's request process, managed by Student Life Disability Services. 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; <http://slds.osu.edu>; 098 Baker Hall, 113 W. 12th Avenue.

Accessibility of course technology

This online course requires use of Carmen (Ohio State's learning management system) and other online communication and multimedia tools. If you need additional services to use these technologies, please request accommodations with your instructor.

- [Carmen \(Canvas\) accessibility](#)
- Streaming audio and video
- Synchronous course tools

Your 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

Disclaimer

This syllabus should be taken as a fairly reliable guide for the course content. However, you cannot claim any rights from it and in particular we reserve the right to change due dates or the methods of grading and/or assessment if necessary. Any changes will be communicated to you through official course announcements.

Course schedule (tentative)

The dates indicate the week over which the material will be delivered while the last date is the live lecture date.

Week	Dates	Topics, Readings, Assignments, Deadlines
1	Aug 23-Aug 29	Introduction to R and base R programming
2	Aug 30-Sep 5	Types of data, Graphical and numerical summaries of data
3	Sep 6- Sep 12	Data transformation using R tidyverse/dplyr
4	Sep 13 - Sep19	Visualization using R ggplot
5	Sep 20- Sep 26	Exploratory data analysis in R and data import
6	Sep 27 – Oct 3	R topics: import and tidy messy data, strings
7	Oct 4- Oct 10	R topics: merging multiple data tables, Project 1 presentation
8	Oct 11- Oct 17	Sample correlation, Autumn break
9	Oct 18- Oct 24	Introduction to probability, methods for counting
10	Oct 25- Oct 31	Conditional probability, Bayes theorem
11	Nov 1- Nov 7	Discrete random variables, expectations, simulation
12	Nov 8- Nov 14	Advanced programming in R – loops, apply, custom functions
13	Nov 15- Nov 21	Simple Linear regression – model, estimation, inference on parameters
14	Nov 22 – Nov 18	Simple Linear regression – diagnostics and remedies, inference on model
15	Nov 29- Dec 5	Multiple linear regression, Take home exam, Project 2 presentation