

# **SYLLABUS: GRADTDA 5402 DATA ANALYTICS FOUNDATIONS 2 SPRING 2022**

## **Course overview**

### **Instructor**

Instructor: Dr. Laura Kubatko

Email address: [kubatko.2@osu.edu](mailto:kubatko.2@osu.edu)

Office hours: Virtual Hours via Carmen Zoom: Mondays 4-5pm

Office Location: 219 Cockins Hall

### **Grader or Teaching Assistant**

TBA

### **Course description**

This is the second course in a two-semester sequence comprised of two, 3 credit-hour courses focusing on R programming and data analysis using R. The sequence is intended to be taken in parallel with the Big Data Computing Foundations sequence focusing on fundamental CS methods for data science. The focus of this second course is on more advanced R programming, distribution theory via simulation methods, statistical modeling (A/B testing, ANOVA, multiple linear regression, logistic regression, multivariate analysis), and communication of results.

Building on the computational and interpretative skills developed in Data Analytics Foundations I, the students will explore formal, theoretically motivated frameworks for data analysis. The initial portion of the course will stress the use of simulation-based experimentation to teach the basic elements of distribution theory and develop a practical understanding of the sampling variability of estimators. The remainder of the course will cover specific modeling strategies for different types of data and will illustrate the data analysis strategies with examples drawn from real world applications.

### **Course learning outcomes**

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

1. Understand and quantify the sampling variability of data and parameter estimators under a variety of data generating mechanisms.
2. Translate real world problems and scientific questions of interest into well-

formulated inferential models for the available data.

3. Select the variables that are most useful to address the inferential task at hand.
4. Use the inferential models developed in parts 2) and 3) to understand the problems of interest, provide answers to domain-specific questions and forecast the values for future occurrences of the data.
5. Assess and quantify the uncertainty in the conclusions drawn from inferential tasks.

## Course materials

### Recommended text:

Andy Field, Jeremy Miles, and Zoe Field (2012) *Discovering Statistics Using R*, SAGE Publishing (e-version available here:

<https://us.sagepub.com/en-us/nam/discovering-statistics-using-r/book236067> )

### Supplemental text:

Murray Aitkin, Brian Francis, John Hinde, and Ross Darnell (2014), *Statistical Modelling in R* (1st Edition), Oxford Statistical Science Series;

(<https://www.amazon.com/Statistical-Modelling-Oxford-Science/dp/0199219133>)

### Required Software:

- R ([www.r-project.org](http://www.r-project.org))
- RStudio ([www.rstudio.com](http://www.rstudio.com))

## 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](mailto: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

- CarmenZoom
- Collaborating in CarmenWiki

- Recording a slide presentation with audio narration
- Recording, editing, and uploading video

### 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>
  - Hands-on tutorials are available in the Swirl system, which you can learn about at <http://swirlstats.com/>. In particular, “R Programming: The basics of programming in R” is an appropriate first tutorial for students who have never used R.
- 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.**
- [Microsoft Office 365 ProPlus](#) All Ohio State students are now eligible for free Microsoft Office 365 ProPlus through Microsoft’s Student Advantage program. Each student can install Office on five PCs or Macs, five tablets (Windows, iPad® and Android™) and five phones.
  - Students are able to access Word, Excel, PowerPoint, Outlook and other programs, depending on platform. Users will also receive 1 TB of OneDrive for Business storage.
  - Office 365 is installed within your BuckeyeMail account. Full instructions for downloading and installation can be found <https://ocio.osu.edu/kb04733>.

## Course delivery

Each week several lecture videos, totaling approximately 2 hours of lecture, will be posted on the course website. 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 the class website. You will be given ample time to complete the assignments.

The instructor may elect to deliver live CarmenZoom lectures in place of videos for certain topics. In this case, the lecture will be recorded and posted on the class website soon after.

On each Wednesday of the semester, we will have one hour of live discussion of the course material presented in the week's lectures. Attendance is encouraged, but is not required, and the discussion will be recorded and posted for later viewing. All of the material covered in these discussion sessions, as well as that presented in the lecture, will be assessed via exams, projects, and homework.

The instructor will hold weekly office hours via CarmenZoom. The dates and times are listed above and are posted on the Carmen website. The instructor will also initiate and manage active discussion boards, also via Carmen.

The majority of the course can be completed asynchronously, meaning that you will be able to study materials and work on assessments according to your own schedule. However, students are strongly encouraged to take advantages of opportunities for synchronous course activities, such as attending the live weekly discussion sessions and making regular visits to office hours. Appointments with the instructor can also be easily arranged – please email the instructor to set up an individual meeting.

## Grading and faculty response

### Grades

Assignment or category	Percentage
Homework	30%
Projects	40%
Exams	30%
Total	100

### Assignment information

Homework will be assigned weekly. Each assignment will carry equal weight in the overall grade. The lowest homework score will be dropped. Late homework will be penalized 5% per day, and will not be accepted if it is more than 2 days late, except in exceptional circumstances requiring documentation.

Two projects will be assigned during the semester. The projects will involve analysis and presentation of a complex data set using the tools discussed in the course. The first project will be due approximately mid-way through the course, and the second project will be due during the final week of the course.

Two exams will be given (see the schedule below).

## Late assignments

Late homework will be penalized 5% per day, and will not be accepted if it is more than 2 days late, except in exceptional circumstances requiring documentation. However, the lowest homework score will be dropped from the grade calculation.

## 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

## 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 large weekly assignments, you can generally expect feedback within **7 days**.

### E-mail

I will reply to e-mails within **24 hours on school days**.

### Discussion board

I will check and reply to messages in the discussion boards every **24 hours on school days**.

# Attendance, participation, and discussions

## 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*.
- **Office hours and live sessions: OPTIONAL OR FLEXIBLE**  
All live, scheduled events for the course, including my office hours, are optional. For live presentations, I will provide a recording that you can watch later. If you would like 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.

## Discussion and communication guidelines

The following are my expectations for how we should communicate as a class. Above all, please remember to be respectful and thoughtful.

- **Writing style:** While there is no need to participate in class discussions as if you were writing a research paper, you should remember to write using good grammar, spelling, and punctuation. Informality (including an occasional emoticon) is fine for non-academic topics.
- **Tone and civility:** We will maintain a supportive learning community where everyone feels safe and where people can disagree amicably. Remember that sarcasm doesn't always come across online.
- **Citing your sources:** When we have academic discussions, please cite your sources to back up what you say. (For the textbook or other course materials, list at least the title and page numbers. For online sources, include a link.)
- **Backing up your work:** Consider composing your academic posts in a word processor, where you can save your work, and then copying into the Carmen discussion.

## 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>). They include the following:

- “A daily health check to report body temperature and health status will be required for all faculty, staff and students each day they intend to be on Ohio State’s campuses in the autumn.”
- Face masks must be worn in indoor settings, including classrooms.
- Members of the campus community will be required to sign a pledge “to affirm their understanding of what is needed to help fight the spread of the virus and their intention to do their part.”
- “Accountability measures will be in place for those who refuse to abide by required health and safety guidelines.”

## Potential disruptions to instruction

If you encounter any health-related or other issues affecting your ability to participate fully in this course, please contact your instructor as soon as possible so that appropriate accommodations can be made. Please note that documentation may be required in order to promote a fair environment for all students.

## 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

- **Exams:** You must complete all exams yourself, without any external help or communication.
- **Written assignments:** Your written assignments, including discussion posts, should be your own original work. In formal assignments, you should use a professional style (i.e., MLA, APA) to cite the ideas and words of your research sources. You are encouraged to

ask a trusted person to proofread your assignments before you turn them in--but no one else should revise or rewrite your work.

- **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 library research look more successful than it was.
- **Collaboration and informal peer-review:** The course includes many opportunities for formal collaboration with your classmates. While study groups and peer-review of major written projects is encouraged, remember that comparing answers on an exam is not permitted. If you're unsure about a particular situation, please feel free just to ask ahead of time.
- **Group projects:** This course includes group projects, which can be stressful for students when it comes to dividing work, taking credit, and receiving grades and feedback. I have attempted to make the guidelines for group work as clear as possible for each activity, but please let me know if you have any questions.

## 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 (Recommended)

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](mailto: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](mailto: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](http://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](http://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)

Week	Dates	Topics, Readings, Assignments, Deadlines
1	1/10-1/14	Random variables, statistical distributions
2	1/17-1/21	Estimators and sampling distributions
3	1/24-1/28	Simulation-based evaluation of statistical estimators
4	1/31-2/4	Hypothesis testing overview; one- and two-sample tests
5	2/7-2/11	A/B testing; principles of experimental design
6	2/14-2/18	Review of linear and multiple regression
7	2/21-2/25	Logistic regression
8	2/28-3/4	One-way ANOVA
9	3/7-3/11	One-way ANOVA; multiple comparisons
10	3/21-3/25	Two-way ANOVA
11	3/28-4/1	Generalized linear models
12	4/4-4/8	Categorical data 1: test of independence
13	4/11-4/15	Categorical data analysis 2: log linear models
14	4/18-4/22	Project presentations
15	4/25-4/29	Wrap-up (no lectures)

**Exam 1: Week of February 14**

**Final exam: Week of April 25**