Syllabus and Course Policies

University of Wisconsin-Madison

STAT 240 LEC 001: Introduction to Data Modeling I

Description: Introduces students to reproducible data management, modeling, and analysis through a practical, hands-on case studies approach. Topics include the use of an integrated statistical computing environment, data wrangling, the R programming language, data graphics and visualization, random variables and concepts of probability, data modeling, and report generation using R Markdown with applications to a wide variety of data to address open-ended questions.

Designations

Credits: 4

Natural Science — N

Quantitative Reasoning B

Elementary Level (note: this course does not satisfy L&S mathematical sciences requirements)

Honors - students have the option to take the course for honors (See **this site**   **(https://stat.wisc.edu/honors/)** for more information about Honors in the Statistics Major)

Instructional Mode

Lecture: in-person

Discussion: in-person

Requisites

Satisfied Quantitative Reasoning (QR) A requirement

Course Meeting Times and Locations

Lecture: Tuesdays/Thursdays 9:30 - 10:45 AM in 331 Service Memorial Institute

Discussion 311: Wednesdays 3:30 - 4:20 PM in 133 Service Memorial Institute (Pritchard)

Discussion 312: Wednesdays 4:35 - 5:25 PM in 133 Service Memorial Institute (Pritchard)

Discussion 313: Thursdays 7:45 - 8:35 AM in 133 Service Memorial Institute (Pritchard)

Discussion 314: Tuesdays 4:35 - 5:25 PM in 1339 Sterling Hall (Huang)

Discussion 315: Wednesdays 4:35 - 5:25 PM in 206 Van Hise Hall (Huang)

Discussion 316: Thursdays 7:45 - 8:35 AM in 387 Van Hise Hall (Huang)

Discussion 317: Wednesdays 7:45 - 8:35 AM in 133 Service Memorial Institute (Gruenloh)

Discussion 318: Wednesdays 8:50 - 9:40 AM in 133 Service Memorial Institute (Gruenloh)

Discussion 319: Wednesdays 9:55 - 10:45 AM in 133 Service Memorial Institute (Gruenloh)

Instructor: Professor Jessi Kehe

Web: https://stat.wisc.edu/staff/cisewski-kehe-jessi/

Email: jjkehe@wisc.edu

TA: Tim Gruenloh (Sections 317-319)

Email: gruenloh@wisc.edu

TA: Shane Huang (Sections 314-316)

Email: shuang457@wisc.edu

TA: Nathaniel Pritchard (Sections 311-313)

Email: npritchard@wisc.edu

Course Websites:

Canvas: **https://canvas.wisc.edu/courses/293926 (https://canvas.wisc.edu/courses/293926)**

Piazza: **piazza.com/wisc/spring2022/sp22_stat240/home**   **(http://piazza.com/wisc/spring2022/sp22_stat240/home)** (or connect directly through the Piazza sidebar on our Canvas site)

This term we will be using Piazza for class discussion and questions. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and the instructor. Rather than emailing questions to the teaching staff, please post your questions on Piazza. If you have not used it before, please checkout our course-specific Piazza page and try it out!

Statistics Learning Center: The Statistics Learning Center   (https://stat.wisc.edu/learning-center/) is available to help you with this course. It is a free, drop-in tutoring service to help support your learning of statistics. The Statistics Learning Center will offer in-person and online tutoring at the follow times and locations:

- In-Person – 133 Service Memorial Institute (SMI)
  - Mondays 3:30pm-5:30pm
  - Tuesdays 3:00pm-7:00pm
  - Thursdays 11:00am-3:00pm
  - Fridays 12:00pm-2:00pm
- Online – Canvas (self-enroll at https://canvas.wisc.edu/enroll/3X6LJX (https://canvas.wisc.edu/enroll/3X6LJX))
  - Wednesdays 4:00pm-7:00pm
  - Thursdays 3:00pm-7:00pm
  - Fridays 11:00-2:00pm

The Statistics Learning Center will begin on January 31st and will end on May 6th. It will be closed the week of Spring Break (March 12th – March 20th). *Note that not all of the Learning Center tutors have experience with all the material covered in this course; be prepared to work with the tutors to locate course materials relevant for addressing your question.*

UW-Madison Credit Hour Policy: One hour (i.e. 50 minutes) of classroom or direct faculty/instructor instruction and a minimum of two hours of out of class student work each week over approximately 15 weeks, or an equivalent amount of engagement over a different number of weeks.

Note: This class meets for a total of three class periods, two 75-minute lectures and one 50-minute discussion period, each week over the semester and carries the expectation that students will work on course learning activities (reading, problem sets, projects, studying, quizzes, exams, and so one) for about 2 hours out of the classroom for every 50 minutes of classroom or direct faculty/instructor instruction. The syllabus includes additional information about meeting times and expectations for student work.

Learning Outcomes: Statistics 240 aims to provide a thorough introduction to the concepts and methods of statistical data science and data modeling. It is the first course in a two-course sequence. Students can expect to spend time each week viewing the posted lecture content, reading from the textbook or other assigned readings, working through examples, doing the homework exercises, and generally becoming skilled in statistical computing and thinking statistically.

Students who complete this course successfully will learn to:

- **wrangle data**: transform data, possibly from multiple sources, into a form convenient for analysis;
- **explore data**: visualize and summarize data, generate questions/hypotheses, and address them;
- **program**: write R code using the RStudio integrated statistical computing environment to carry out data wrangling, graphical data exploration, and analysis that is reproducible;
- **model data**: provide low-dimensional summaries of data that capture signal and quantify the noise; assess the adequacy of the model; understand random variables and probability concepts associated with the models;
- **interpret data**: explain what can be inferred from the data analysis and make predictions;
- **communicate**: use R Markdown to integrate prose, visualizations, code, interpretation, and results;
- **collaborate**: work with other students to solve data challenges.

Furthermore, students will learn about:

- **statistical inference**: the construction and interpretation of confidence intervals and the calculation and interpretation of p-values for hypothesis tests for a number of settings including one- and two-sample proportions and means
- **simple linear regression**: the construction and interpretation of regression models for two quantitative variables

Course Materials
Textbook:

  R for Data Science (https://r4ds.had.co.nz)
  by Wickham and Grolemund (freely available **online** (https://r4ds.had.co.nz))

  Statistics 240 Course Notes and Case Studies (https://bookdown.org/bret_larget/stat-240-case-studies/)
  by Bret Larget (freely available **online** (https://bookdown.org/bret_larget/stat-240-case-studies/))

Software:

  R: **https://cran.r-project.org/** (https://cran.r-project.org/)
  R Studio: **https://www.rstudio.com/** (https://www.rstudio.com/)

Grading: The final course grade will be determined by a score made up from the *weighted* sources below.

Component	Percent of Overall Grade
Short Online Assessments (SOAs) (lowest two scores dropped)	10%
Discussion group assignments (lowest score dropped)	10%

Component	Percent of Overall Grade
Individual weekly assignments (lowest score dropped)	20%
Final Project (18% for team project, 2% for peer review)	20%
Midterm Examination	20%
Final Examination	20%

Letter grades will be assigned using the following grading scale; if adjustments are made to this scale, it will only be in a direction that improves your grade.

A (93 - 100%), AB (87 - 93%), B (83 - 87%), BC (77 - 83%), C (70 - 77%), D (60 - 70%), and F (Below 60%)

Short Online Assessments (SOAs): Each week will include reading assignments and most will include an online quiz (SOA) which will examine your understanding of the assigned reading. Reading assignments will come from the online textbook and the course notes, and will be administered through Canvas. Lectures will assume that you have read and comprehended assigned reading. Anticipate spending approximately 30 - 60 minutes outside of class for each reading assignment and quiz.

Discussion Sessions: Weekly discussion sessions will include a discussion assignment. Discussion assignments are short group assignments meant to be completed during the 50-minute discussion period. Groups will be assigned early in the semester (prior to the third week of classes), and new groups will be assigned after the midterm exam. Only one group member needs to turn in each group assignment, but all group members are expected to contribute to doing the work. Discussion assignments are intended to practice a single concept. Each discussion assignment will involve editing an R Markdown file to answer several questions, knitting the document to HTML, and uploading the knitted document to the Canvas web page.

Discussion assignments turned in within 24 hours after the due date may receive up to half credit. Discussion assignments turned in more than 24 hours late will not be graded and will receive no credit.
Your lowest single discussion assignment score is dropped.

Each student is expected to attend their discussion section meeting every week and meaningfully participate with their assigned group. Students who are not present to contribute to the discussion assignment completion will not receive credit for the assignment. Students will not be allowed to complete a discussion session assignment individually.

Individual Assignments: In addition to group assignments in discussion sessions, there will be individual assignments, typically due on the Friday after they are assigned. These are assignments are longer than the group assignments, and each individual assignment may take 4–6 hours to complete.

Each homework assignment should be completed by creating an R Markdown document that combines text with chunks of code to produce graphics, carry out data analyses, and interpret results. Homework assignments should be well organized with all problems clearly marked and in order. Each problem solution should include a brief description of the problem (that may be paraphrased from the actual problem) as well as your work. A typical weekly assignment includes a series of short problems aimed at demonstrating conceptual understanding and mastery of basic skills from the week plus a longer more open-ended data analysis project directed toward one or more questions of interest.

Late homework assignments turned in within 24 hours of the due date will receive half credit; assignments turned in over 24 hours late will receive no credit and will not receive feedback. Any exceptions must be approved by prior arrangement with Professor Kehe, whose judgment on the acceptability of reasons for exceptions is final.

Completed homework assignments will be uploaded on Canvas and the submission typically will include (i) the completed R Markdown document and (ii) the knitted HTML document.

Each homework assignment will carry the same weight on your overall grade (regardless of the number of points each assignment is worth). For well-organized and neat assignments, each graded question on an assignment will be worth 5 points using the rubric below. Questions on homework assignments may be graded for accuracy, for completeness, or not graded. When completing an assignment, you should assume each question will graded for accuracy.

Points per problem	Characteristics
3	The solution is complete and correct.
2	A serious attempt was made to solve the problem, but the solution is not correct.

Points per problem	Characteristics
1	An attempt was made to solve some part of the problem, but the solution is incomplete or contains major errors.
0	The problem was not attempted or there was no evidence that the student put forth serious effort on the problem.

Exams: There will be an midterm exam and a cumulative final exam.

The midterm exam is an in-class exam during a lecture period and may also include a take-home portion to be completed for a limited amount of time during a 24-hour period. You must work independently. Evidence of collaboration on an exam may lead to receiving zero credit and will be reported to campus authorities for potential further sanctions.

At Professor Kehe's sole discretion, alternative examination times may be permitted for students who give her ample prior notice of an acceptable reason.

The date of the midterm exam is Tuesday, March 1.
The date of the final exam is Thursday, May 12, 2022 from 7:25 PM - 9:25 PM

Final group project: Each student will complete a project with an assigned group of students. The project will include the acquisition of a data from a novel source, an interesting question to address, the creation of an R Markdown document that contains a reproducible process of data manipulation, graphical exploration, modeling, analysis, and interpretation, using the appropriate methods from the course. The project will culminate in a report displayed as an HTML document.

Honors: Students may opt to take Statistics 240 for honors. Students who do so will be placed into groups to do a separate project using methods from the course on data the group selects to make a video presentation to share with the instructors and other honors groups. The honors project is separate and distinct from the final project (which will involve a different group of students). The honors project is graded on a satisfactory/unsatisfactory basis. Professor Kehe will form groups for honors projects after the second week of class. There will be a number of opportunities for feedback on the project during the semester. The honors project will be due on Friday, April 22 at 11:59pm.

Quarantine or Isolation Due to COVID-19
Please see **https://covidresponse.wisc.edu/**   **(https://covidresponse.wisc.edu/)** for information about COVID-19 and the university's response.
Students should continually monitor themselves for COVID-19 symptoms and get tested for the virus if they have symptoms or have been in close contact with someone with COVID-19. Student should reach out to instructors as soon as possible if they become ill or need to isolate or quarantine, in order to make alternate plans for how to proceed with the course. Students are strongly encouraged to communicate with their instructor concerning their illness and the anticipated extent of their absence from the course (either in-person or remote). The instructor will work with the student to provide alternative ways to complete the course work.

Copyright Statement – Note Taking and Recording Lectures
This syllabus and all lectures and materials are copyrighted 2021 J. J. C. Kehe. As a student in this course, you may take notes for solely your personal use. Students are prohibited from providing or selling notes to anyone else or being paid for taking notes by any person or commercial firm without my express written permission. Unless you are considered by the university to be a qualified student with a disability requiring accommodation, you are not authorized to record my lectures without my permission. [Regent Policy Document 4-1]. Unauthorized use of my copyrighted materials constitutes copyright infringement and may be addressed under the university's policies, UWS Chapters 14 and 17, governing student academic and non-academic misconduct.

Rules, Rights, and Responsibilities
See details **https://guide.wisc.edu/undergraduate/#rulesrightsandresponsibilitiestext**
(https://guide.wisc.edu/undergraduate/#rulesrightsandresponsibilitiestext) .

Academic Calendar and Religious Observances
See details **https://secfac.wisc.edu/academic-calendar/#religious-observances**   **(https://secfac.wisc.edu/academic-calendar/#religious-observances)** .

Academic Integrity and Data Ethics

By virtue of enrollment, each student agrees to uphold the high academic standards of the University of Wisconsin-Madison; academic misconduct is behavior that negatively impacts the integrity of the institution. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these previously listed acts are examples of misconduct which may result in disciplinary action. Examples of disciplinary action include, but is not limited to, failure on the assignment/course, written reprimand, disciplinary probation, suspension, or expulsion.

The members of the faculty of the Department of Statistics at UW-Madison uphold the highest ethical standards of teaching, data, and research. They expect their students to uphold the same standards of ethical conduct. Standards of ethical conduct in data analysis and data privacy are detailed on the ASA website (**https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx**
(https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx)), and include:

· Use methodology and data that are relevant and appropriate; without favoritism or prejudice; and in a manner intended to produce valid, interpretable, and reproducible results.

· Be candid about any known or suspected limitations, defects, or biases in the data that may affect the integrity or reliability of the analysis. Obviously, never modify or falsify data.

· Protect the privacy and confidentiality of research subjects and data concerning them, whether obtained from the subjects directly, other persons, or existing records.

By registering for this course, you are implicitly agreeing to conduct yourself with the utmost integrity throughout the semester.

Accommodations for Students with Disabilities
McBurney Disability Resource Center **syllabus statement** **(https://mcburney.wisc.edu/instructor/)** :

> The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA.

Diversity and Inclusion
Institutional **statement on diversity** **(https://diversity.wisc.edu/)** :

> Diversity is a source of strength, creativity, and innovation for UW–Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals.
>
> The University of Wisconsin–Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background — people who as students, faculty, and staff serve Wisconsin and the world.

Netiquette

- Any comment or answers must be on topic, concise, polite, and respectful of others.
- Assume the best intentions of others in the class and be forgiving when you think that the tone of someone's post is offensive. It is easy to misread the tone of someone's written communication. If in doubt, ask an open, honest question about what the person meant so that you can clarify before making assumptions that damage your perception of your colleague.
- Students must not post answers to homework or exam problems. Questions should be sent using Piazza. Students should not send questions to the TA or instructor via email, unless for personal or sensitive issues. Students should not expect an immediate answer to a question posted outside normal business hours, including late at night before an assignment due date.

Complaints
If you have a complaint about a TA or course instructor, you should feel free to discuss the matter directly with the TA or instructor. If the complaint is about the TA and you do not feel comfortable discussing it with him or her, you should discuss it with the course instructor. Complaints about mistakes in grading should be resolved with the instructor in the great majority of cases. If the complaint is about the instructor (other than ordinary grading questions) and you do not feel comfortable discussing it with him or her, contact the Director for Undergraduate Studies, Professor Cecile Ane, cecile.ane@wisc.edu.

If your complaint concerns sexual harassment, please see campus resources listed at **https://compliance.wisc.edu/titleix/resources/ (https://compliance.wisc.edu/titleix/resources/)** . In particular, there are a number of options to speak to someone confidentially.

If you have concerns about climate or bias in this class, or if you wish to report an incident of bias or hate that has occurred in class, you may contact the Chair of the Statistics Department Climate & Diversity Committee, Professor Karl Rohe (karl.rohe@wisc.edu). You may also use the University's bias incident reporting system, which you can reach at **https://doso.students.wisc.edu/services/bias-reporting-process/ (https://doso.students.wisc.edu/services/bias-reporting-process/)** .