

## STAT 240, Introduction to Data Modeling I

- 4 credits
- Canvas: <https://canvas.wisc.edu/courses/450178> (<https://canvas.wisc.edu/courses/450178>)
- Designations
  - Natural Science — N
  - Quantitative Reasoning B (QR-B)
  - Intermediate Level
  - Honors - students have the option to take the course for honors
- Instructional Mode:
  - lecture: in-person
  - discussion: in-person

## Course Description

Introduces reproducible data management, modeling, analysis, and statistical inference 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 including the binomial and normal distributions, data modeling, statistical inference in one- and two- sample settings for proportions and means, simple linear regression, and report generation using R Markdown with applications to a wide variety of data to address open-ended questions.

## Requisites

- Satisfied Quantitative Reasoning (QR) A requirement

## Course Meeting Times and Locations

- **Lecture:** MWF 3:30 PM - 4:20 PM, 150 minutes per week
  - **Sahifa Siddiqua** ([ssiddiqua@wisc.edu](mailto:ssiddiqua@wisc.edu) (<mailto:ssiddiqua@wisc.edu>))
  - 125 Agricultural Hall
- **Discussion:** 50 minutes per week
  - **Xiaobin Zhou** (Discussion Sections 341, 342, 343) ([xiaobin.zhou@wisc.edu](mailto:xiaobin.zhou@wisc.edu) (<mailto:xiaobin.zhou@wisc.edu>))
  - **Jun Hyung Chang** (Discussion Sections 344, 345, 346) ([jchang243@wisc.edu](mailto:jchang243@wisc.edu) (<mailto:jchang243@wisc.edu>))

## UW-Madison Credit Hour Policy

**45 Hours Per Credit** – This class meets for a total of four class periods, three 50-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 class period. The syllabus includes additional information about meeting times and expectations for student work.

## Learning Outcomes

The primary objective of Statistics 240 is to provide students with a thorough introduction to the concepts and methods of statistical data science and data modeling. This course is the first course in a two-course sequence. Successful students can expect to spend many hours outside of class each week reading and working examples from the textbook, working on exercises, becoming skilled in statistical computing, and mastering the ability to think 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 R Studio integrated statistical computing environment to carry out reproducible data analysis
  - **model data:** use probability and random variables in statistical computing environment to carry out reproducible data analysis
  - **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

- **make statistical inferences:** employ confidence intervals and hypothesis tests using both computational and standard methods such as binomial and t-tests

## Grading

Your absolute score in the class will be determined by the following weighted categories.

Source	Percentages
Discussion Assignments (lowest 2 dropped)	10%
Individual Homework Assignments (lowest 1 dropped)	20%
Final Project	10%
Midterm 1	20%
Midterm 2	20%
Final Exam	20%

Your letter grade in the class will be assessed based on your absolute score in the class.

### Absolute Score Scale

Grade	Range
A	92-100%
AB	88-92%
B	80-88%
BC	76-80%
C	64-76%
D	60-64%
F	<60%

## Discussion Assignments

Discussion assignments are short group assignments meant to be completed during the 50-minute discussion period. Groups will be assigned early prior to the third week of classes. Only one group member needs to turn in each group assignment, but all ought 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 should be uploaded at the end of the discussion period.** Full credit is given to all students who are present and actively participate in the group assignment, even if the assignment is incomplete.

Your lowest two discussion assignment scores are dropped. This may not show up in Canvas but will be accounted for in our calculations of your final grade.

**Students who are not present to contribute to the discussion assignment completion will not receive credit for the assignment.**

## Individual Assignments

There will be individual assignments which are longer than the group assignments. 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.

Late homework assignments do not receive full credit. Assignments received up to one day late will earn **70%** of the credit. Assignments received up to one day late will earn **50%** of the credit. Assignments later than that will receive no credit and will not receive feedback. Any exceptions must be approved by prior arrangement with the instructor, whose judgment on the acceptability of reasons for exceptions is final.

Your lowest homework assignment score is dropped. This may not show up in Canvas but will be accounted for in our calculations of your final grade.

## Exams

There will be two exams during the semester and one final exam.

The first exam will be at the end of Week 6 (**Friday, February 28**) and will be an **in-person evening exam** of *90 minutes*. The room and timing of the exam will be announced closer to the exam date. You **must show up in-person to take the exam**. Similarly to homeworks, you will download a .Rmd file containing questions, which you will edit with your answers, then finally knit and submit both the .Rmd and .html file.

The second exam will be at the end of Week 11 (**Friday, April 4**) and will be an **in-person exam** of *50 minutes* during normally scheduled lecture time in the normally scheduled lecture room. You **must show up in-person to take the exam**. This test will be a paper and pen test, more details will be announced later.

The final exam will be **in-person** on **Friday, May 9th**, from **5:05PM - 7:05PM**. The room will be announced closer to the exam date. You **must show up in-person to take the exam**.

**If alternative testing dates are needed due to a conflict please let the instructor know as soon as possible.** You must work independently on all exams. Evidence of collaboration or use of external resources (like ChatGPT, google and so on) other than the ones allowed on an exam may lead to academic misconduct sanctions from campus authorities.

## Final Group Project

Each student will complete a project with an assigned group of students. The project must 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.

## Communication

### Lecture:

Participation in regularly scheduled lectures each week will include the opportunity for direct interaction between students and the instructor, and regularly scheduled discussion sections offer the opportunity for direct interaction with the TA.

Office hours are another avenue for direct communication with both instructor and TAs. Please see the Canvas page (linked at the top of this syllabus) for more information on office hours.

### Canvas:

Canvas announcements (which should automatically send an email to your university supplied address) will be used to share all critical information. It is imperative that your @wisc.edu email is active and working, and that you check it regularly or have it forwarded to the account that you use most regularly. In addition, the Canvas course page contains the course schedule and links to all assignments, required reading, and lecture notes.

## Textbook and Software

### Textbooks:

These textbooks are optional, free, supplementary references you may consult if you wish, but are not required.

- *R for Data Science*, by Wickham and Grolemund.
  - Free online at <https://r4ds.had.co.nz/> (<https://r4ds.had.co.nz/>)
- *Statistics 240 Course Notes*
  - Free online at <https://bwu62.github.io/stat240-revamp/index.html> (<https://bwu62.github.io/stat240-revamp/index.html>)

### Software:

- R: <https://cran.r-project.org/> (<https://cran.r-project.org/>)
- R Studio: <https://posit.co/> (<https://posit.co/>)
  - The company formerly known as RStudio has been renamed Posit. We will continue to use the RStudio product for the course.

### Hardware

All students need a computer with the course software for almost all assignments. There are many lectures where you will want to execute code and do examples as the lecture proceeds, so having a laptop computer is helpful, if not essential. Discussion sessions assume that you have a laptop computer.

## Rules, Rights, and Responsibilities

See <https://guide.wisc.edu/undergraduate/#rulesrightsandresponsibilitiestext> (<https://guide.wisc.edu/undergraduate/#rulesrightsandresponsibilitiestext>)

## Course Evaluations

Students will be provided with an opportunity to evaluate this course and your learning experience.

You will receive an official email two weeks prior to the end of the semester when your course evaluation is available. You will receive a link to log into the course evaluation with your NetID where you can complete the evaluation and submit it, anonymously. Your participation is an integral component of this course, and your feedback is important to me. I strongly encourage you to participate in the course evaluation.

## Academic Calendar and Religious Observances

See: <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. For detailed information, please see <https://conduct.students.wisc.edu/academic-misconduct/> (<https://conduct.students.wisc.edu/academic-misconduct/>).

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.

## Netiquette on Piazza and online communication

See <https://kb.wisc.edu/50548> (<https://kb.wisc.edu/50548>) for a general netiquette. Specifically:

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 problems. Almost all questions should be sent via Piazza. For personal or sensitive issues, use private messages to instructor by email. Students should not expect an immediate answer to a question posted 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.

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. 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/>). 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, and include:

# Accommodations for Students with Disabilities

McBurney Disability Resource Center syllabus statement:

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 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. I will work either directly with the student 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:

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.

<https://diversity.wisc.edu/> (<https://diversity.wisc.edu/>)

## Overlapping course time statement

The Department of Statistics strongly discourages students from enrolling in any courses whose regular class meeting dates & times overlap with each other. This policy is in alignment with the College of Letters and Sciences Course Attendance Policy. It is also consistent with the Class Attendance Policy for Students at UW-Madison (<https://kb.wisc.edu/lis/24628> (<https://kb.wisc.edu/lis/24628>)), whose first sentence reads, "It is expected that every student will be present at all classes." Statistics instructors may opt not to make any alternative arrangements in the event any

conflict arises due to a student taking a course with class meetings that overlap with a Statistics course, including a conflict between two Statistics courses. Note that final exams occasionally are scheduled simultaneously for courses which meet at different times; in this situation, please contact your instructor well before the exam date about potential accommodations.