Syllabus
STAT 340: Data Modeling II
Fall 2021, 3 Credits

Description
Students will learn how to explore and analyze data using R, as well as how to present their findings and analyses clearly. Topics include basic probability models; the central limit theorem; Monte Carlo simulation; one- and two-sample hypothesis testing; Bayesian inference; linear and logistic regression; the boostrap; random forests and cross-validation. Students will learn how to present their findings in a clear and reproducible manner in a project setting by applying their skills to analyze a data set of their choice.

Prerequisites
STAT240: Introduction to Data Modeling I and one or more of MATH 217, 221, or 275. Specifically, students should have a broad familiarity with the R programming language and should be comfortable with basic concepts from calculus.

Instructors
Keith Levin (`kdlevin@wisc.edu`); office hours: Wednesdays 1:00pm to 3:00pm in MSC 6170
Bi Cheng Wu (`bwu62@wisc.edu`); office hours: Mondays 1:10pm to 2:10pm, Wednesdays 3:25pm-4:25 in MSC 1235B

Teaching Assistants
Joseph Salzer (`jsalzer@wisc.edu`); office hours: Mondays and Wednesdays 11:00am-12:30pm location TBA. Nathan Aviles (`naviles@wisc.edu`); office hours: Tuesday, Wednesday and Thursday 8:45am-9:45am location TBA Bowen Zhang (`bzhang327@wisc.edu`); office hours: Tuesday 1:00pm-2:00pm and 3:30pm-4:30pm location TBA

Meetings
Lecture: Monday, Wednesday, Friday 9:55AM-10:45AM in Chamberlin 2103
Discussion sections: Discussion sections are optional, but are highly recommended, especially during the last third of the course when you will be working on your projects. Refer to your schedule for time and location.

Textbook, Readings & Online Resources
There is no physical textbook required for this course. We will make frequent reference to *An Introduction to Statistical Learning with Applications in R* (ISLR) by James, Witten, Hastie and Tibshirani. The textbook is available online at `https://www.statlearning.com/` and in print from Springer. Other required readings will be made available as we cover relevant material, and supplemental readings will be suggested for those who are interested in learning more.

All class resources will be made available on the course web page, `https://kdlevin-uwstat.github.io/STAT340-Fall2021/` and on the course Canvas page. Please contact the instructors if any resources are missing from either of these websites. The instructors will make an effort to post lecture notes and demo code a few days ahead of lecture.

Course Topics

- **Random variables and models**. Basic probability models, conditional probability, Monte Carlo simulation, the strong law of large numbers, central limit theorem.
- **Estimation**. Confidence intervals, point estimates, the bootstrap.
- **Testing**. One- and two-sample hypothesis testing, test statistics, permutation tests.
- **Prediction**. Linear and logistic regression, random forests, cross-validation.
- **Exploratory Data Analysis**. Clustering, unsupervised learning, visualization.

Learning Outcomes

By the end of this course, you will be able to:

- Use the R programming language, especially tools from the tidyverse, to gather, clean and analyze data.
- Understand and apply basic concepts in probability; combine basic probability models to build more complicated ones; and critique models and their assumptions.
- Formulate statistical hypotheses for different kinds of research questions and test those hypotheses using both classical and Monte Carlo methods.
- Understand and apply principles of statistical estimation and prediction, including fitting models and assessing model quality.
- Perform basic exploratory data analysis and present findings visually using `ggplot2`.
- Apply statistical tools to answer research questions using real-world data and present these findings clearly in both spoken and written form to non-experts.

Grading, Exams, Homeworks & Late Days

Grades will be based on cumulative performance on a set of approximately five homeworks, two exams and a final project. Homeworks will review material from lecture and will be primarily programming-based. Exams will include both programming and short answer questions designed to assess how well students can explain and apply the concepts and methods discussed in lecture. The project will involve students working in groups to collect, analyze and explore a data set of their choice. Specifically, these will count toward your final grade as follows:

Homeworks:	30%
Exam 1:	20%
Exam 2:	20%
Project:	30%

Note that the exact number of homework assignments is subject to change depending on factors such as lecture cancellations and the speed with which we cover material. The instructors reserve the right to curve scores in the event of skewed class performance. Students may contest their grade on an assignment up to two (2) weeks from the day that an assignment's grades are released, after which grades may not be changed. Homework due dates are strict, and you may turn in work late only with the use of "late days", of which you have five (5) to use over the course of the semester. For each late day you spend, you may extend the deadline of a homework by up to 24 hours. You may spend multiple late days per homework. Once you have turned in your homework you may not spend more late days to turn in your homework again after the deadline (you may, of course, turn in multiple versions of your homework assignment through Canvas prior to the deadline). Late days will be deducted automatically, and there is no need to notify the instructors that you wish to spend a late day. Homeworks turned in late with no remaining late days to spend will receive a zero. The purpose of this late day policy is to give you a way to deal with unexpected circumstances (e.g., illness, family emergencies, job interviews) without having to come to the instructors. Of course, if dire circumstances arise (e.g., long-term illness that causes you to miss multiple weeks of lecture), please speak with the instructors as promptly as possible.

Key Dates First lecture: Wednesday, September 8, 2021

Exam 1: Monday, October 18, 2021

Exam 2: Friday, November 19, 2021

Last lecture: Wednesday, December 15, 2021

Project presentations: Saturday, December 18, 2:45PM-4:45PM (the scheduled final)

Ethics and class policies

Academic misconduct includes such actions as copying code from the web or from your fellow students, providing code to your fellow students, looking up solutions online, turning in assignments from other classes or previous iterations of this course, and hiring others to complete your work for you. You are welcome to discuss homeworks with your classmates, but the work that you turn in must be yours and yours alone, and you must disclose in your homework the names of those with whom you collaborated.

From the Office of Student Conduct and Community Standards:

> [A]cademic 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.

See `https://conduct.students.wisc.edu/academic-misconduct/` for more information.

Violations of these or other university ethical standards surrounding academic honesty will be met with serious consequences and disciplinary action. At a minimum, cheating on an assignment will result in a 0 for that assignment and the incident will be reported to the appropriate office. At the instructors' discretion, depending on the circumstances, an additional full letter grade may be deducted from the student's final grade in the course.

COVID-19 Preparation and Policies The university is, as of the time of this writing, requiring that masks be worn indoors at all times. We expect that students attending lecture will abide by this policy. The COVID-19 pandemic is, of course, still in progress and is inherently unpredictable and the university is updating its policies accordingly. Please be sure to follow all university guidelines surrounding masks and vaccines. Should the university decide to return to remote instruction, we are prepared to switch to online instruction.

Accommodations for Students with Disabilities

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 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 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.