STAT 479: Sports Analytics
Fall 2025
University of Wisconsin–Madison

Description: Illustrates the use of statistical modeling and data science techniques to derive actionable insights from sports data. Emphasizes not only technical calculation of advanced metrics but also on written and oral communication to other data scientists and to non-technical audience. Topics may include: deriving team rankings from paired competitions; measuring an individual player's contribution to their team's overall success; assessing player performance and team strategy in terms of expected outcomes; forecasting the impact of new rule changes using simulation; and creating new metrics using high-resolution player tracking data.

Learning outcomes: By the end of the course, you will

- Implement appropriate statistical methods to assess player and team performance
- Work with play-by-play and high-resolution tracking data
- Provide constructive and actionable feedback on your peers' analytical reports
- Build a personal portfolio of sports data analysis projects
- Effectively communicate analytical processes, results, and conclusions to diverse audiences through written reports and oral presentation

Location & schedule: Tuesdays & Thursdays 11:00am - 12:15pm in Morgridge Hall 1524.

Mode of instruction: In-person lectures.

Instructor: Sameer Deshpande. Office Hours: Monday 11:00am - 12:00pm (Morgridge Hall 5586); Wednesday 3:00pm - 4:00pm (Morgridge Hall 5586); Friday 3:00pm - 4:30pm (Morgridge Hall 5618).

TA: Zhexuan Liu. Office Hours: Tuesdays and Thursdays 9:15am - 10:45am (Morgridge Hall 2515).

Requisites: STAT 333 or 340. Details on assumed background are available below.

Course website: https://skdeshpande91.githug.io/stat479_fall2025. We will also use Canvas and Piazza (please see below for links).

Textbook: There is no required textbook. Please see below for a list of helpful references.

Grading: Grades will be based on performance on three group projects and class participation.

Diversity & inclusion: My goal is that students of all backgrounds, perspectives, and experiences be well-served by this course. I will do my best to create an environment (both in and out of the classroom) that welcomes everyone. I expect that we will engage with each other in a way that respects and celebrates diversity in all forms including, but not limited to, age, culture, disability, gender, language, national origin, and sexual orientation. Your suggestions towards these goals are strongly encouraged and appreciated.

Additional course information: The remainder of this document contains much more detailed information about the course structure, assignments, and policies and procedures.

Requisites: The formal requisites are STAT 333 or 340. The course will make extensive use of R. Accordingly, I will assume fluency with basic R functionality (e.g., assignment, writing and executing scripts, saving data objects, setting environments, installing and loading packages); data manipulation with **dplyr** and other **tidyverse** packages; and visualization using either base R graphics or **ggplot2**. I will additionally assume some familiarity with fitting statistical models in R and interpreting their output (e.g., using `lm` and `glm`); fundamental probability and statistical concepts (e.g., random variables, expectations, variance, p-values); and some basic programming concepts (e.g., writing `for` loops and conditional statements).

The course does not pre-suppose deep knowledge about specific sports. It does, however, assume a willingness to learn about sports that you may not necessarily enjoy watching or playing.

Tentative Schedule:

Week	Tuesday	Thursday
9/1–9/5	(no meeting)	Overview
9/8 – 9/12	Expected goals (XG)	Building our own XG model
9/15 – 9/19	Adjusted Plus/Minus	Regularized Adjusted Plus/Minus
9/22 – 9/26	Expected Runs	Wins Above Replacement (WAR) I
9/29 – 10/3	WAR II	Multilevel modeling
10/6 – 10/10*	NFL WAR	Pitch Framing
10/13 – 10/17	Rankings I	Rankings II
10/20 – 10/24	Markov Chains I	Markov Chains II
10/27 – 10/31	Plackett-Luce Models	Simulating Effects of New Rules
11/3 – 11/7*	Intro. to tracking data	Intro. to tracking data
11/10 – 11/14	Space ownership I	Space ownership II
11/17 – 11/21	Forecasting trajectories I	Forecasting trajectories II
11/24 – 11/28	TBD	Thanksgiving (no meeting)
12/1 – 12/5*	TBD	TBD
12/8 – 12/12	Project Presentations	(no meeting)

Group project reports and presentation recordings are due on the starred ($\star$) dates: October 10, November 7, and December 5.

Assignments & grading: Your final grade will be based on your performance on three group projects and your overall participation.

Group Projects. There will be three group projects. Each project consists of two parts, a written report and a group presentation. The project due dates are October 10, November 7, and December 5.

The written report consists of a non-technical executive summary and a technical report. The executive summary, which should not exceed 500 words, should describe the overall goals, analytic approach, and main conclusions in non-technical language. The executive summary should be free from jargon, code listings, figures, tables, and charts. It should be written to be read and understood by a front office executive, coach, player, or fan with little data science experience. The rest of written report should

- Clearly state the problem being studied and provide sufficient background details and to motivate why the problem is important and interesting.

- Describe the data and major steps of the analysis

- Presents the main results within the context of the relevant sport(s) and supports the results with figures, tables, charts, and other statistical software output as appropriate.

- Discusses the limitations of the analysis and outlines concrete steps for further development.

The technical section of the report should contain enough detail and code that another data scientist could replicate your analysis verify its soundness. Code listings and output (e.g., figures, tables, charts, and numerical summaries) should be tightly integrated with the written exposition. ***The use of Quarto or RMarkdown is highly recommended for preparing the written report.***

Each team will also record an 8–10 minute presentation (e.g., using Zoom that provides an overview of their analysis. Each presentation should include the following elements

- Background (2–4 slides): clearly motivate and state the main problem being studied. Explain why it is interesting and important. Present just enough background to motivate the problem, while taking care not to overwhelm the audience with extraneous details. If appropriate, comment on the limitations of existing solutions to the problem or closely-related problems

- Analysis overview (2–4 slides): present only the main steps of your analysis. Be sure to explain why each step was necessary and how these steps contribute to the overall solution. Focus more on the high-level ideas and motivation for each step rather than the specific implementation or software syntax

- Main results (2–3 slides): distill your results into a few key points. Use figures, tables, charts, and other statistical software output to support your findings.

- Conclusion (1 slide): briefly summarize your analysis and findings and outline between 1 and 3 specific directions for future development, improvement or refinement.

Project groups can earn up to 100 points on their written report and up to 100 points on their presentation.

Project Peer Reviews. After every group project, every student is expected to peer-review the written report and presentations of 3 other teams. The primary purpose of this peer review is to *practice* providing constructive feedback on both technical and non-technical writing to your fellow data scientists. Peer reviews will be completed on Canvas and a structured evaluation rubric will be provided for each assignment. The peer review process will be single-blinded so teams will not know the identities of their reviewers. Students can earn a total of 20 points for each peer review by (i) completing the provided rubric and (ii) leaving a constructive comment that identities strengths and areas for potential improvement. Peer reviews are due one week after the project assignment due dates; that is, they are due on October 17, November 14, and December 12. A penalty will be assessed for submitting late reviews. In total, students can earn up to 180 points for completing peer reviews.

Group Participation: After each project submission, each student will fill out a team accountability survey. Students will earn up to 40 points per project submission: 20 for completing the survey and 20 based on the feedback from their teammates and their own self-assessment. In total, students can earn up to 120 points based on their level of participation in their project groups.

Participation: An additional 100 points will be awarded based on participation during class, office hours, and online discussions on Piazza.

Final grades. Your final grade will be based on how many of the 1000 total points you earn during the semester. Over 925 points is at least an A, over 875 points is at least an AB, over 800 points is at least a B, over 700 is at least a BC, over 600 points is at least a C, and over 500 point is at least a D. Final grade boundaries will be announced no later than December 5, 2025.

Course website. All lecture notes and code will be posted at

https://skdeshpande91.github.io/stat479_fall2025.

We will use Canvas for submitting assignments and completing peer reviews:

https://canvas.wisc.edu/courses/476839 .

We will also use Piazza for on-line discussions:

https://piazza.com/wisc/fall2025/fa25stat479001/info

Required textbook, software, and other course materials: There is no assigned textbook.

The course will make extensive use of R. The following are excellent references and I highly encourage you to read and consult them as needed:

- *R for Data Science* by Wickham, Çentikaya-Rundel, & Grolemud (link)

- *RMarkdown: The Definitive Guide* by Xie, Allaire, & Grolemud (link)

- *An Introduction to Statistical Learning* by James, Witten, Hastie, and Tibshirani (link)

- *Data Science: A First Introduction* by Timbers, Campbell, and Lee (link).

- *Beyond Linear Modeling: Applied Generalized Linear Models and Multilevel Models in R* by Roback and Leger (link)

You may also find the following websites, blogs, and podcasts useful as sources of inspiration as you develop analyses of your own.

- The online textbook *Analyzing Baseball Data with R* (link) and the associated blog, which contains lots of helpful resources for analyzing tabular box score data to high-resolution ball tracking data and everything in between

- Ron Yurko's Substack "Statistical Thinking in Sports Analytics" (link)

- The *Journal of Quantitative Analysis of Sports* (link): the premiere venue for academic sports analytics articles. You should have access to all articles through the UW Libraries. Each issue also features a publicly available "Editor's Choice" article.

- The Open Source Sports podcast (link): Each episode of this podcast focuses on a single academic research paper featuring authors as guests, with discussions about the statistical methodology, relevance and future directions of the research

- Hockey Graphs (link): a blog that's developed really innovative public-facing hockey analytics

- The Wharton Moneyball Post Game Podcast (link): a podcast version of the popular Sirius XM radio show hosted by several Wharton professors and featuring interviews with sport analytics thought leaders

Academic misconduct. I take academic integrity extremely seriously. While I encourage you to collaborate with your classmates on the problem sets, you are expected to write up your own solutions. You **must** acknowledge **any and all** sources that you consulted, whether it be a reference book, online resource, or another person (in the class or otherwise). You **may not** post any of the course material to a website like StackOverflow or Chegg to solicit answers from people outside the class. You **may not** post any course material to ChatGPT or a similar generative AI service. Doing so will result in disciplinary action, as detailed in UWS 14. Further, you are not permitted to post or share any course material (e.g., lecture notes, code, project reports, presentation recordings) without my written permission.

Policy on the use of generative artificial intelligence (AI) models. While the Statistics Department recognizes the potential benefits of AI, its use in academic work can be problematic. In this course, three rules regarding the use of ChatGPT and other generative AI models will be enforced:

1. Passing off AI-generated responses as original student work constitutes plagiarism and is strictly prohibited. Any students found to be engaging in this practice will be cited for academic misconduct.

2. Unless explicitly authorized by the instructor to do so, any use of AI-generated responses as sources of information, even with documentation and attribution, is prohibited.

3. You may not, under any circumstances, upload any course material to a generative AI model or agent. This includes lecture slides and notes and reports or videos uploaded by other student groups.

How credit hours are met: This class meets for two 75-minute class periods each week over the semester and carries the expectation that students will work on course learning activities (reading, writing, problem sets, studying, etc.) for about three hours out of the classroom for every class period.

Regular and substantive student-instructor interaction. Substantive interaction occurs via two channels: (i) regular class meetings in which students are encouraged to engage in discussion and (ii) weekly office hours.

Ethics. The members of the faculty of the Department of Statistics at UW–Madison uphold the highest ethical standards of teaching, data, and research. We expect our students to uphold the same standards of ethical conduct. The American Statistical Association's standards for ethical conduct in data analysis and data privacy are available at this website 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 the 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.

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 (UW-855) require the university to provide reasonable accommodations to students with disabilities to access and participate in its academic programs and educational services. Faculty and students share responsibility in the accommodation process. Students are expected to inform faculty of their need for instructional accommodations during the beginning of the semester, or as soon as possible after being approved for accommodations. Faculty will work either directly with the student or in coordination with the McBurney Center to provide reasonable instructional and course-related accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA. (See: McBurney Disability Resource Center).

Diversity & inclusion. 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.

Privacy of student records & the use of audio recorded lectures. Lecture materials and recordings for this course are protected intellectual property at UW–Madison. Students in courses may use the materials and recordings for their personal use related to participation in class. Students may also take notes solely for their personal use. If a lecture is not already recorded, students are not authorized to record lectures without permission unless they are considered by the university to be a qualified student with a disability who has an approved accommodation that includes recording (Regent Policy Document 4-1). Students may not copy or have lecture materials and recordings outside of class, including posting on internet sites or selling to commercial entities, with the exception of sharing copies of personal notes as a notetaker through the McBurney Disability Resource Center. Students are otherwise prohibited from providing or selling their personal notes to anyone else or being paid for taking notes by any person or commercial firm without the instructor's express written permission. Unauthorized use of these copyrighted lecture materials and recordings constitutes copyright infringement and may be addressed under the university's policies, UWS Chapters 14 and 17, governing student academic and non-academic misconduct.

Religious observances. Students are responsible for notifying the instructor within the first two weeks of classes about any need for flexibility due to religious observances.

Further information and policies. Please visit this link for additional information about student privacy, course evaluations, and student rights and responsibilities.