Prof. Brennan Klein
b.klein@northeastern.edu
Network Science Institute
Northeastern University

CNET 5442
Sports Analytics through Data and Networks

Spring 2026; 4 credits
Mon/Wed: 2:50 – 4:30pm
Jan. 7 – Apr. 24, 2026
Richards Hall #140

Summary

Sports provide a powerful lens for learning quantitative thinking. Every game generates a stream of measurable events—goals, passes, shots, substitutions—that can be modeled as time series, networks, probability distributions, or decision processes. In this course, sports analytics is more than an industry tool: it is a *Trojan horse* for teaching statistics, modeling, scientific computing, and complex systems. Students will work hands-on with real datasets and Python notebooks to practice regression, hypothesis testing, Bayesian inference, classification, and causal inference. These methods expand into the tools of complex systems, from passing networks and player embeddings to tournament forecasting and the diffusion of tactics. The semester culminates in a group project where students apply course methods to a dataset of their choice (preferably, a novel dataset collected for this course). This course demonstrates how analytics reshapes modern sports and provides the opportunity to learn a transferable set of skills for analyzing complex systems in any domain.

Course Learning Outcomes

By the end of this course, students should be able to:

1. Apply core statistical and data science techniques—including regression, hypothesis testing, Bayesian inference, classification, and causal inference—to real sports datasets.
2. Represent sports and complex systems data in multiple forms (e.g., tables, networks, sequences, tensors) and evaluate the assumptions and limitations of each representation.
3. Construct and analyze network models of play, including passing networks, ranking systems, and player embeddings, using appropriate computational tools.
4. Develop and implement forecasting and simulation models to predict matches, tournaments, and tactical trends, and assess their accuracy.
5. Collect, clean, and process raw sports data; generate reproducible analyses in Python; and communicate findings through effective visualizations, written reports, and oral presentations.
6. Critically evaluate how analytics influences sports and broader complex systems, and formulate evidence-based arguments that test intuition against data.

Coursework, Class Structure, Grading

This course meets twice weekly and combines short lectures with interactive coding and discussion. Lectures will introduce the core ideas, followed by live demonstrations and guided exercises in Python. Students should bring laptops to every class, as many sessions will include in-class coding, data exploration, or short activities. We will occasionally host guest speakers from academia and industry to provide perspective on applied sports analytics. A short mid-class break will be built in to allow for questions and informal interaction. Class time will be complemented by structured weekly assignments, which focus on applying methods to real datasets, and a semester-long group project that allows students to design and carry out their own analysis. The project emphasizes the full research cycle: posing a question, collecting data, building models, evaluating results, and communicating findings clearly.

Students can expect timely communication about assignments via Canvas, feedback on major assessments within two weeks, and responses to emails within 48 hours on weekdays. In turn, I expect students to come prepared to participate in coding and discussion, to collaborate respectfully, and to communicate promptly if challenges arise that may affect their coursework.

Grading will be based on the following:

- **Attendance and Participation: 10%**
  Active participation in discussions, coding labs, and peer feedback sessions.
- **Weekly Assignments: 45%**
  Six coding and analysis assignments that build/evaluate technical skills.
- **Midterm Project Proposal and Presentation: 10%**
  Groups present their research plan and preliminary analysis to receive structured feedback.
- **Final Project Report and Presentation: 35%**
  A group project involving data collection/analysis, culminating in a report and presentation.

Final Project Details

The semester culminates in a group project that applies course methods to a real sports or complex systems dataset. Projects may involve analyzing an existing dataset, scraping or collecting new data, or extending techniques introduced in class. Projects will be evaluated on originality, rigor, use of course methods, clarity of communication, and reproducibility of analysis. Deliverables include:

- **Proposal & Intermediate Presentation (Week 7/8):** A short written description of your question, dataset, and plan and a 5-7 minute presentation of your project design for feedback.
- **Final Report and Presentation (due Finals Week):** A 8-12 page write-up that presents your methods, results, and interpretation, and a group presentation delivered during the final week.

Course Materials

There is no single textbook that covers the scope of this course. Instead, students will work with a combination of open-source texts, research articles, and software tools. All required readings will be made available through the course website.

Resources

- *Python and Data Science:*
  - VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, Inc. `https://jakevdp.github.io/PythonDataScienceHandbook/`
  - Severance, C. (2016). Python for Everybody: Exploring Data using Python 3. Charles Severance. `https://do1.dr-chuck.com/pythonlearn/EN_us/pythonlearn.pdf`
  - Downey, A. (2012). Think Python: How to Think Like a Computer Scientist. `https://www.greenteapress.com/thinkpython/thinkpython.pdf`
- *Sports Analytics and Statistics*
  - Albert, J., Bennett, J., & Cochran, J.J. (Eds.). (2005). Anthology of Statistics in Sports. *Society for Industrial and Applied Mathematics*. `https://epubs.siam.org/doi/abs/10.1137/1.9780898718386.ch37`
  - Lebed, F. (2017). Complex Sport Analytics. Routledge. Available online at: `https://doi.org/10.4324/9781315692920`

  - Miller, T.W. (2015). Sports Analytics and Data Science: Winning the Game with Methods and Models. FT Press. https://github.com/mtpa/sads
  - Beggs, C. (2024). Soccer Analytics: An Introduction Using R. Chapman and Hall/CRC. https://doi.org/10.1201/9781003328568

- *Advanced Topics & Methods*
  - Barabási, A.L. & Pósfai, M. (2016). *Network Science*. Cambridge University Press. https://networksciencebook.com/
  - Menczer, F., Fortunato, S., & Davis, C. A. (2020). *A First Course in Network Science*. Cambridge University Press. https://doi.org/10.1017/9781108653947
  - Thurner, S., Hanel, R., & Klimek, P. (2018). Introduction to the Theory of Complex Systems. Oxford University Press. https://academic.oup.com/book/25504
  - Klein, B., Smith, A., Chinazzi, M., Zhang, Q., et al. (2025) Network Science Data & Models Python Textbook — https://network-science-data-and-models.github.io/phys7332_fa25/README.html

- Research papers and case studies in sports analytics (provided as PDFs).

Software and Data

- Python (e.g. `numpy`, `pandas`, `matplotlib`, `networkx`, `statsmodels`, `scikit-learn`, `statsbombpy`, among others) and Jupyter notebooks, distributed through the course GitHub.
- Open datasets, including StatsBomb open event data, NBA play-by-play data, NFL tracking data, ATP Tennis datasets, and more.
- Students are encouraged to collect or scrape additional data for their final projects.

Instructor

Brennan Klein is core faculty at the Network Science Institute and Assistant Teaching Professor in the Department of Physics. He is the program director of the MS in Complex Network Analysis at Northeastern University. Prof. Klein is also the director of the Complexity & Society Lab, which is focused on two broad research areas: 1) Information, emergence, and inference in complex systems: developing tools and theory for characterizing dynamics, structure, and scale in networks, and 2) Public health and public safety: drawing on complex systems science to document—and fight against—emergent or systemic disparities in society, especially as they relate to public health and public safety. As of 2025, he is also the director of NetSI Sport, an interdisciplinary research group focusing on complex systems-inspired approaches to sports analytics. In 2023, Prof. Klein was awarded the René Thom Young Researcher Award, given to a researcher to recognize substantial early career contributions and leadership in research in Complex Systems-related fields. Prof. Klein is the Data for Justice Fellow at the Institute on Policing, Incarceration & Public Safety at Harvard University's Hutchins Center for African & African American Research. He received a PhD in Network Science in 2020 from Northeastern University and earned his BA in Cognitive Science & Psychology from Swarthmore College in 2014. Website: `brennanklein.com`.

Office Hours

There will be weekly office hours to support coding assignments, data collection, and group projects. Students are encouraged to attend office hours for technical support, debugging help, and early feedback on project ideas. Mondays, 4:45-5:45pm, Network Science Institute (177 Huntington Ave, 10th floor) or Zoom link posted on Canvas.

Accessibility and Accommodations

Northeastern is committed to providing equal educational opportunities for all students. Students who require accommodations for a documented disability should contact the Disability Resource Center as early as possible to ensure that appropriate arrangements can be made. Once you have documentation, please share your accommodation letter with me so we can discuss how best to support your learning.

Late Work Policy

Assignments are due on the dates listed in the schedule. Each student has a 48-hour grace period across the semester that can be applied to any assignment without penalty. After this, late work will be marked down 10% per day, up to three days. Extensions for serious circumstances will be considered.

Academic Integrity

All students are expected to uphold Northeastern University's Academic Integrity Policy, which prohibits cheating, plagiarism, fabrication, unauthorized collaboration, and other forms of academic dishonesty. You are responsible for ensuring that your work reflects your own effort and analysis, even when you consult outside resources such as peers, published materials, or AI tools. Proper citation is required whenever you use code, data, text, or ideas that are not your own. Questions about what counts as appropriate collaboration or citation should be raised with me directly. Suspected violations will be referred to the Office of Student Conduct and Conflict Resolution. More information can be found here: `https://osccr.sites.northeastern.edu/academic-integrity-policy/`.

All student records and coursework in this class are handled in compliance with the Family Educational Rights and Privacy Act. Please use your Northeastern email account for all course communications.

Policy on Artificial Intelligence and Large Language Models

This course recognizes the potential of artificial intelligence (AI) tools—such as ChatGPT, Copilot, Claude, and other text or code generators—to support learning, creativity, and efficiency. You are encouraged to use AI when it adds value to your learning process, provided that its use is transparent, relevant, and critically evaluated. AI can help brainstorm ideas, debug code, generate visualizations, or give writing feedback, but it is not a substitute for your own analysis or reasoning.

Guidelines for Use

- AI use will vary depending on the assignment. Labels will be provided to indicate whether AI use is prohibited, permitted, encouraged, or required, depending on the learning objectives.
- For assignments where AI use is allowed: cite the tool, include information about the prompt or queries you used, and briefly explain how it contributed to your work. This is not meant to police your prompts, but rather to crowdsource and share effective strategies for navigating the tool.
- You remain responsible for the accuracy, originality, and integrity of all submitted work. AI tools are known to make errors, invent references, or introduce bias. Verification is your responsibility.

Learning Orientation

Think of AI as a *ladder, not a crutch*. Its purpose is to extend your abilities, not to replace the productive struggle of problem-solving. Over-reliance on AI will limit your growth, while thoughtful use can accelerate your improvement on a range of quantitative and qualitative skills. Throughout the semester, we will highlight best practices for integrating AI into analysis, coding, and communication in ways that strengthen—not weaken—your understanding.

Title IX and Mandatory Reporting

At any point of the semester, if an issue arises that affects your ability to fully participate and learn in this class, please talk with me as you feel comfortable. Please note that I am a university-mandated reporter, meaning that if I am made aware of an incident related to sexual assault, sexual harassment, gender-based harassment, dating or domestic violence, sexual exploitation, or stalking, I am required to report it to the Northeastern University's Office of University Equity and Compliance/Title IX Coordinator.

Many campus resources exist to support you, including the following:

- **University Health and Counseling Services*:** Provides confidential medical and mental health services.
- **Find@Northeastern*:** Offers 24/7 mental health support to full-time students and connection to unlimited free counseling sessions. Call 877-233-9477 (US), 855-229-8797 (Canada), or +1-781-457-7777 (International).
- **We Care:** Creates a support network for interpersonal problems, academic concerns, or personal tragedies.
- **Disability Access Services:** Ensures that students with disabilities have equal access to the academic experience at Northeastern and advocates for students with learning differences.
- **OPEN (Alcohol and Other Drug and Sexual Violence Services)*:** Provides confidential, non-judgmental check-ins for students.
- **Center for Spirituality, Dialogue, and Service*:** Attends to spiritual needs and offers confidential support and guidance.
- **Office for University Equity and Compliance:** Investigates reports of discrimination, harassment, and retaliation at Northeastern.
- **Wellness Days:** As part of a pilot program, Boston-based undergraduate students can take up to two wellness days each semester. `https://wellnessdays.studentlife.northeastern.edu/overview/`

*Indicates confidential resource

Schedule

Schedule and topics may be adjusted with reasonable notice.

Week 1: Introduction

Monday, January 5, 2026 – No class

Wednesday, January 7, 2026 – Class 1: Introduction – Sports as Complex Systems

- Sports as complex systems; why sports are data-rich laboratories.

Week 2: Data Types Across Sports

Monday, January 12, 2026 – Class 2: Data Types Across Sports

- Core data modalities across sports: event data, tracking data, outcomes, contextual covariates.

Wednesday, January 14, 2026 – Class 3: Tournament Structures

- Leagues, knockout tournaments, groups, Swiss systems, and how structure shapes inference.

Week 3: Ranking and Competition Structures

Monday, January 19, 2026 – No class, Martin Luther King Jr. Day

Wednesday, January 21, 2026 – Class 4: Distributions, Odds, & Surprises

- Heavy tails, streaks, upsets; calibration and surprises across sports.

Week 4: Distributions and Regression

Monday, January 26, 2026 – Class 5: Regression Pt. 1 – Moneyball Replication

- Replicating a Moneyball-style analysis; model specification and evaluation.

Wednesday, January 28, 2026 – Class 6: Regression Pt. 2 – Expectation & Measures of Likelihood

- Expected value, likelihood, loss functions, and interpreting probabilistic predictions.

Week 5: Regression Continued

Monday, February 2, 2026 – Class 7: Regression Pt. 3 – Survival Analysis & Logistic Regression

- Survival/time-to-event modeling and logistic regression for binary outcomes in sports.

Wednesday, February 4, 2026 – Class 8: Regression Pt. 4 – Bayesian Statistics & the Hot Hand

- Bayesian framing of uncertainty; priors/posteriors; hot-hand style questions.

Week 6: Classification and Embedding

Monday, February 9, 2026 – Class 9: Classification & Clustering

- Supervised vs. unsupervised learning; clustering players/teams; evaluation and pitfalls.

Wednesday, February 11, 2026 – Class 10: Multidimensional Data & Embedding

- Dimensionality reduction and embeddings for high-dimensional sports features.

Week 7: Invited Speaker

Monday, February 16, 2026 – No class, Presidents' Day

Wednesday, February 18, 2026 – Class 11: Invited Speaker (TBD)

- Guest talk and discussion (details to be announced).

Week 8: Causality

Monday, February 23, 2026 – Class 12: Introduction to Causality Pt. 1

- Causal questions in sports data; experimental vs. observational designs

Wednesday, February 25, 2026 – Class 13: Introduction to Causality Pt. 1 & Intermediate Presentations

- Applications of causal inference techniques.

Week 9: Spring Break

Monday, March 2, 2026 – No class, Spring Break

Wednesday, March 4, 2026 – No class, Spring Break

Week 10: Machine Learning

Monday, March 9, 2026 – Class 14: Machine Learning Pt. 1 – Introduction

- Problem setup, features/labels, training/validation/testing, and common baselines.

Wednesday, March 11, 2026 – Class 15: Machine Learning Pt. 2 – March Madness

- Bracket prediction as a modeling case study; tournament prediction and evaluation.

Week 11: Spatiotemporal Data and Networks

Monday, March 16, 2026 – Class 16: Spatiotemporal Data Analysis: Hockey

- Working with space and time: tracking-derived features, rates, and movement patterns.

Wednesday, March 18, 2026 – Class 17: Introduction to Network Science Through Sports

- Nodes/edges in sports; projections; weighted/temporal networks; basic measures.

Week 12: Networks in Soccer I

Monday, March 23, 2026 – Class 18: Networks in Soccer – Passing Network Analysis

- Match-level passing networks; centrality, cohesion, and interpretation.

Wednesday, March 25, 2026 – Class 19: Networks in Soccer – Pitch Passing & Spatial Networks

- Spatially grounded passing networks; zones, geometry, and positional structure.

Week 13: Networks in Soccer II

Monday, March 30, 2026 – Class 20: Networks in Soccer – Sequences of Events Pt. 1

- Event sequences as networks; representations and basic modeling ideas.

Wednesday, April 1, 2026 – Class 21: Networks in Soccer – Sequences of Events Pt. 2

- Higher-order structure, motifs, and sequence-based prediction tasks.

Week 14: Roles and Transfer Networks

Monday, April 6, 2026 – Class 22: Networks in Soccer – Roles and Motifs

- Role discovery, motifs, and mesoscale structure in match networks.

Wednesday, April 8, 2026 – Class 23: Transfer, Trade, and Scouting Networks

- Player movement and scouting as networks; markets, intermediaries, and pathways.

Week 15: Information and Wrap-Up

Monday, April 13, 2026 – Class 24: Information Theory or Ranking with Networks?

- Topic TBD (information-theoretic views of sport, or network-based ranking methods).

Wednesday, April 15, 2026 – Class 25: Invited Speaker (TBD)

- Guest talk and discussion (details to be announced).

Week 16: Final Presentations

Monday, April 20, 2026 – No class, Patriot's Day

Wednesday, April 22, 2026 – Class 26: Final Presentations

- Final project presentations.
