

---

# Project Proposal Guidelines

## 1 Logistics

The project is intended for teams of 2-3. However, email me if you'd like an exception, which should be possible. All submissions will be on Gradescope. There should be only one submission per team: use the group submit option to tag your teammates.

## 2 Important dates

1. **February 8:** deadline for submitting project proposal
2. **February 22:** deadline for submitting progress report
3. **March 14:** Poster presentations during lecture in week 10
4. **March 18:** deadline for submitting final write-up (including access to code and data [as long as data is not private])

## 3 Project Proposal

The proposal should be a **one-page** (not more!) summary of your plan for the project. It should include the **names of all team members** and answer

- What is your research question, and why is it important?
- What data will you use, and how will you procure them?
- What methods will you use to answer your research question?

Please feel free to post on Ed (project tab) with any ideas for projects if you are looking for partners. Please feel free to come to any OH to brainstorm the projects/get help.

## 4 Project ideas

The project is meant to allow you the freedom for you to tackle a question in sports analytics that you care about. Treat the project more like an opportunity to explore sports data, educate yourself on a particular topic, and pursue directions that interest you.

It is good to present “negative results” as well. “Negative results” includes things like: there is no meaningful differences between these methods, teams, or players; or this approach is not effective for a reason that you explore and specify.

All of last year's projects are up on the final project blog here: <https://stats100blog.wordpress.com/about/>. These give a good sense of the scale and types of projects students have done in the past.

The following list of topics is by no means exhaustive or restrictive, but merely hopes to get your creative juices flowing:

- **Paper replication and extension.** Look at an edition of [The Journal of Quantitative Analysis in Sports \(JQAS\)](#) from the last two years. JQAS should be available through the Stanford library. Read and understand the paper, and replicate at least one of their numerical analyses (ideally using their data, or a similar dataset). Extend their paper in some way, such as by carrying out related analysis; extending their methods to a novel dataset or sport; or critiquing their methods or conclusions and proposing a better alternative.
- **Consultant for a player or team.** Imagine you are a analytics consultant for your favorite player or team (including possibly a Stanford team!). What recommendations would you give them? Which players should a team trade or hire? What strategy or schedule should a player pursue? Back your recommendations with ample data and convincing analysis, and quantify and convey the value of your advice.
- **Applying quantitative analysis to “folk wisdom.”** Is there such a thing as “the hot hand”? Do “trap games” really cause upsets? Is LeBron James the greatest NBA player of all time? Does your coach have beliefs that are just plain wrong? Consider an article of faith in one of your favorite sports, and subject it to rigorous quantitative analysis. It can be good either to corroborate or refute the “folk wisdom”—as long as you have solid data to back it up.
- **Following up on lectures.** Was there a lecture or guest lecture that you really enjoyed, piqued your curiosity, or inspired you? Or did you hear something in lecture that struck you as odd, counterintuitive, or just plain wrong? Dive into your curiosity or dig into your objections.
- **Applications to Other Sports** Similarly, are there methods from lecture that you thought might be adaptable to another sport, especially sports with less existing data analysis? In many cases you could do the first analysis of the kind for a given sport!
- **Unique Data.** Do you have access to a unique or novel dataset? Perhaps it is a new data modality that has come out within the last two years on which there is little published work. Or perhaps it is a dataset that you have unique access to you because of your background or interests. See what new questions your dataset opens up!
- **Fantasy Champ.** Eager to win your fantasy league? Concoct a method to ensure you’ll come out on top next season, and back it up with data and analysis so that we believe it’ll work.
- **Sports Betting.** Propose a method to make money based on sports betting. Ground it in data and analysis. Give convincing evidence that it works (just don’t use real money!).
- **Rankings.** Survey ranking algorithms used in a sport of interest to you and in related sports. Compare their benefits and their deficiencies. Is any ranking clearly better than the others? Can you propose a better ranking system? It may be interesting to consider both the evaluative and predictive functions of rankings, as well as their social implications (especially if you can find convincing data about the latter).
- **Theoretical Extensions.** Perhaps you are very interested in a specific statistical technique covered in the course, like empirical Bayes or regularization. Research into the theoretical extensions of these approaches. Such a project has scope to be much more mathematical or theoretical; just be sure to have some small data analysis tying it back to sports.

- Any question about sports that you care about and can be tackled with data science.

## 5 Possible data sources

- General sports statistics from many different sports can be found on [Sports Reference](#)
- [NCAA Statistics](#) from various different college sports
- ESPN has a wealth of sports data, such as can be scraped from this [API](#)
- You can also get data from the MLB such as with this [API](#)
- And of course you can get data from the NFL such as with this [API](#) (you get the point...)
- The [retrosheets](#) package in R provides play-by-play baseball data, which can be useful for analyzing specific players and decisions
- Stacast data from [Baseball savant](#)
- Hustle! In the past a student got data by twitter DMing the American Ultimate Disc League. Be bold and reach out; ask and you shall receive!
- Many other datasets and R packages/APIs on the pinned Ed post