Data description + the problem

I will be working primarily with the dataset found on The Fencing Database (https://www.fencingdatabase.com/). This primary source of data here are clips of fencing phrases (i.e., from when the referee commands “fence” to when a touch is scored), the names of both fencers, which weapon is being used, and the event that the clip is sourced from. From these, community efforts are made to answer several questions, including:

- Where was the hit fencer standing?
- Where was the hit scored?
- Who initiated the action?

All three modern Olympic fencing weapons are featured in the database. While there are many differences between each, all three weapons have the same data detailed above, meaning any analytical pipelines can easily be reused (although weapons should always be kept separate because of their fundamental differences). The primary questions this project seeks to address are:

- Are fencers who initiated the action more likely to score? What does this say about aggression within each weapon?
- Where within the fencing area do most actions score? What about second most? Can a probability map be created showing likelihood of success?
- Does where the hit was scored correlate to where the hit fencer was standing?

By answering the primary questions, this project can provide a data-driven view on fencing that is sorely lacking. As an added benefit, quantitative comparisons can be made between the weapons where ordinarily only qualitative comparisons exist. This project is relevant because many countries are beginning to use analytical and data science approaches as a way to better develop their athletes and prepare them for fencing internationally. Currently, Team USA is already using data science to evaluate the efficacy of certain tactical decisions (although they do not address the questions asked here). While still early, Olympic team decisions often have downstream ramifications with NCAA teams and local clubs using the same or similar approaches.

One drawback of this database is that it requires an API key. While the database is regularly maintained (the last update was made March 14, 2024), the developer may be hesitant to allow access to their data for analysis purposes or may simply be unresponsive. If this happens, a backup database in the form of Quarte Riposte (https://actions.quarte-riposte.com/browse?u=actions) already exists and can be scraped.

Stakeholders

- Kyle Sozanski (myself): as a coach, the results of this project are going to inform my coaching practices and strategies
- Buckeye Fencing Club: since this is the club I coach at, the results of this project will have ramifications for how fencers develop
- The fencing community across all levels

Key Performance Indicators

- Inputs: Data quality and uniquenessGathering dataset, cleaning to remove missing values, separating each weapon
- Process: External team gathered, Github created, pipeline to address each question created and tested
- Output: Figures representing each question and interpretations of figures, dissemination to the community (e.g., via social media)