Problem Statement

Sailing Rankings Brainstorm, June 9, 2019

General

Our goal in this brainstorming session is to come up with several possible rankings approaches that use the quality of regatta participants present to determine the quality of the regatta. We will work through the three below cases: a regional system with existing rankings, a regional system without existing rankings, and a national system without existing rankings. For each case, we'll follow the below process:

1. Pair off for 15 minutes to come up with a rough approach
2. Discuss approaches as a group for 30 minutes
3. Spend 30 minutes formalizing your solution with your partner

If you are unfamiliar with college racing, there is a quick start summary at the end of this document that has sufficient information to approach the problem and would be helpful to read prior to the rest of the document.

Cases

Existing rankings

To start, we consider world with an existing ranking system which only considers NEISA teams at NEISA regattas. Our question: given the existing rankings, how do we update the values after each regatta?

Start condition

Once we have an understanding of how to update the rankings given existing rankings, the question remains of how to assign initial rankings. Develop a start condition for your ranking system. What happens when all of the boats at a regatta don't have scores? What happens if only one or two don't have scores?

National problem

Within a relatively compact system like NEISA, it is easier for the rankings to stay in sync, as schools compete frequently and amongst themselves, but as the system expands we risk

rankings in one region not being comparable to rankings in another region. In your ranking system, is this a justified concern? If so, how can we prevent the situation?

Data

Race Results Data: General

The data provided is deidentified fleet racing data from Techscore. Only data pertaining to double handed dinghies (FJs, 420s, Fireflies, and Larks) are included. Names of schools, regattas, years, and specific dates are masked to avoid risks of introducing bias into the model. We are looking at the total score for each school and make no note of individual division scores. If a school fielded multiple teams at the regatta, only the school name is noted; this means that at a given regatta in our dataset a school could be listed more than one.

Variables:

- **regatta_id**, integer, uniquely identifies regattas within each dataset (is not consistent between the two)
- **day**, integer, number of days since September 1
- **school_coded**, string, uniquely identifies school, consistent between datasets
- **place**, integer, the place awarded to the school at the regatta
- **score**, integer, the number of points earned by the school at the regatta

Race Results Data: NEISA

Only contains NEISA schools, schools from other regions are excluded; however, the place number noted is not updated. (i.e. If College of Charleston got first at a NEISA regatta and Yale got second, there would be no record for first place and Yale's place would be recorded as second.) NEISA schools are coded as cereals.

Race Results Data: National

All fall regattas meeting the criteria above for a particular year on Techscore. Non-NEISA schools are coded as either a fruit, vegetable or candy.

Rankings Data Set

These rankings are from a previous method of ranking regattas and are used here to help us work through early questions in the problem. Rankings exist for all NEISA schools and have the following variables:

- **school_coded**, string, uniquely identifies school, consistent between datasets
- **score**, double, the numeric school given from the old ranking system. The higher a school's score, the higher their ranking. The score is normalized between 0 and 1, but you are welcome to scale it any way that simplifies your logic.

- **ranking**, int, the ranking of the school from 1 to n schools

College Racing Basics

- Two people sail together in a small boat
  - The skipper steers and controls the main sail
  - The crew focuses on keeping the boat flat (the force of the wind in the sails tips the boat) and the jib (small sail at the front) full
- There are two main types of college racing: fleet and team.
- We are focusing on fleet racing where individual boats try to sail around a course a fast as possible. The fastest boat wins.
- In a college regatta, several individual races are run often in multiple divisions (usually called A and B).
- A boat's place in each race is summed for the teams final score; the lowest score wins.
- A certain regattas, a school may send separate teams which are scored separately.
- In a given season, there are harder regattas and easier regattas.
  - Generally, the schools present determine the difficulty of the regatta
  - Better sailing schools make for harder racing/regattas
- National college sailing is divided into geographic regions
  - We are looking a NEISA, the New England division. NEISA is broadly considered one of the more competitive regions in college racing as many of the schools have been competing since the 1930s.
- Techscore is a website used to track college racing scores.