Optimizing event order of swim meets

Spencer Anderson

December 15, 2023

1. Abstract

Competitive swimming is a highly developed and highly standardized sport spanning every habitable continent and virtually every country. With multiple distances and strokes to optimize, swimmers often gravitate towards a small selection of events that they prefer. For example, Great Britain swimmer Adam Peaty has only competed in the 50 and 100 Breaststroke on the world stage for the past decade and USA's Katie Ledecky has dominated the distance events at the Olympics since she was 15. This paper aims to both research what groupings are most common as well as optimize the order of events to allow the most number of swimmers the most amount of recovery between events.

2. Introduction

An important consideration for swimmers when signing up for a meet is to spread out their events so that the performance of later events is not hindered by earlier ones. At championship meets events are typically split between multiple days allowing the athletes even more time to rest between swims. The focus of this paper will cover the typical layout of high school swim meets in the United States: 8 individual events, 3 relays, and diving, all within a single evening (between 1 and 3 hours). By providing a starting point for optimization, further research can be done to hopefully model more complex meets such as NCAA Championships, World Championships, and even the Olympics. From a broad overview, the issue lies in swimmers competing in back-to-back events. Thus, the objective function would try to maximize the amount of time between events for the most number of swimmers. Although rudimentary, an exhaustive search can be performed for the 12 high school events in a reasonable amount of time ($12! \approx 479,000,000$). A separate but related result of being able to numerically rank specific orders is being able to compare the best result returned against the order that is currently being used.

3. Methods

3.1. **Background.** As this model is only being applied to high-school meets, it is necessary to specify which events are in question. Due to the organization of swimming governing bodies in the US, USA Swimming (in addition to World Aquatics) only specifies officially recognized events but the National Federation of State High School Associations (NFHS) specifies the order for high school swimming [1]. A table of the events is below (Table 1).

Order	Event
1	200 Yard Medley Relay
2	200 Yard Freestyle
3	200 Yard Individual Medley
4	50 Yard Freestyle
5	Diving (1-meter, 6 dives)
6	100 Yard Butterfly
7	100 Yard Freestyle
8	500 Yard Freestyle
9	200 Yard Freestyle Relay
10	100 Yard Backstroke
11	100 Yard Breaststroke
12	400 Yard Freestyle Relay

Table 1. Order of events specified by the NFHS.

To keep further discussion of these events more succinct, the widely used abbreviations of Free, Back, Breast, Fly, IM, FR, and MR will be used.

One of the best ways to collect the necessary data is to record the entries for a championship-level meet. The New York State Public High School Athletic Association (NYSPHSAA) Federation Championships were chosen for this reason. More specifically, the NYSPHSAA 2022-2023 Boys [2] and 2022-2023 Girls [3] meet entries were recorded. As this is a relatively small sample to collect, the model should be re-run with data from the bigger swim states such as California, Texas, and Florida. However, the New York State meet has above-average qualifying standards meaning the diversity of participating swimmers is enough to highlight the sought-after event groupings.

Although these championships hold diving concurrently with the swimming events, the objective function is calculated with diving as one of these events. This allows the calculated event order to be used for the widest use case possible. Most championship meets often include a 15 to 20-minute break at the time diving would be to allow swimmers the rest they would typically get in a local or dual meet. It is important to point out a minor drawback from pulling data from a championship meet; there is no overlap between swimmers and divers. This only comes into effect when a diver needs to swim in a race or vice versa, something that happens very rarely.

3.2. **Mathematical Model.** The foundation of the model is to calculate how much rest each swimmer in the sample would get for a given permutation of events, and take the sum to find the value of the objective function for that permutation.

The model uses three inputs, E, A, P, defined below.

- $E := \{e_1, e_2, ..., e_{12}\}$
  The set of events to permute where e_k is an event listed in Table 1.

- $P := \{p_1, p_2, ..., p_{12!}\}$
  The set of permutations of E, where each p_k is an ordered 12-tuple of events.

- $A := \{a_1, a_2, ..., a_{711}\}$
  The set of athletes collected, where each a_k is the set of events each athlete had swum, $1 \leq |a_k| \leq 4$.

For a given "rest" function R and $p \in P$, the objective function T would be the sum of R for each athlete:

$$T(p) = \sum_{i=1}^{|A|} R(a_i, p).$$

As this is a maximization problem, the final result desired is

$$P_{max} = \max\{T(p), \forall p \in P\}.$$

The above equations summarize the general model but the choice of the rest function determines the properties of the resulting P_{max}.

3.3. **Rest functions.** The first rest function is a naive one, as it simply calculates the sum of the number of events between the events that an athlete is competing in.

For example, if the permutation was the usual order (see Table 1) and the given athlete's events were 100 Breast, 200 MR, 200 IM, 400 FR the following distances would be as follows

- 1 between the 200 MR and 200 IM
- 7 between the 200 IM and 100 Breast
- 0 between the 100 Breast and 400 FR

for a total of 8 events of rest.

The major drawback to the rest function described above is that it assumes that each event takes the same length. This is almost always not true, especially when comparing events like the 50 Free (about 30 seconds) and the 500 Free (about 5 minutes). However, the majority of events all have a duration of a couple of minutes and with several heats of each event, the relative time difference ends up being smaller than one might think.

A more comprehensive function would take into account the number of heats of each event and their respective average duration. One might even go as far as to calculate the actual heat that each swimmer is in, allowing the total rest to be calculated to within a few minutes. While not much more computationally expensive than the former method, the number of iterations that this calculation would need to perform would most likely add to the total run time of a program, not to mention values for each event would need to be collected.

3.4. **Solution Process.** As mentioned above a naive way to run the model on the collected data would be to loop through each of the 479,001,600 permutations, summing the result of the rest function for each of the 711 athletes every time.

While still terribly inefficient, there are two main ways to speed this up. Because of the assumption that the intersection of swimmers and divers is empty (and the data collected supports this), if diving is ever the first or last event in a permutation, it can be thrown out. Since diving can only ever add to the result of the rest function, if it shows up as the first or last it will always be adding 0 rest to the result.

Another major time-save allowed by the model is the symmetry of permutations. Simply put, the rest function for a given athlete and permutation returns the same result as when the event order has been reversed. This, combined with the previously mentioned time-save, shrinks the input space down to $5 \cdot 11! = 199,584,000$ permutations to try.

Using a computer with multiple cores further speeds up the process as each set of 11! events can be computed in parallel. Running the model using the naive rest function on a 12-core HP Z800 workstation took just under 12 hours.

4. Results

From just the data collected, the question about the most popular event groupings can be answered (see Figure 1 and Figure 2).

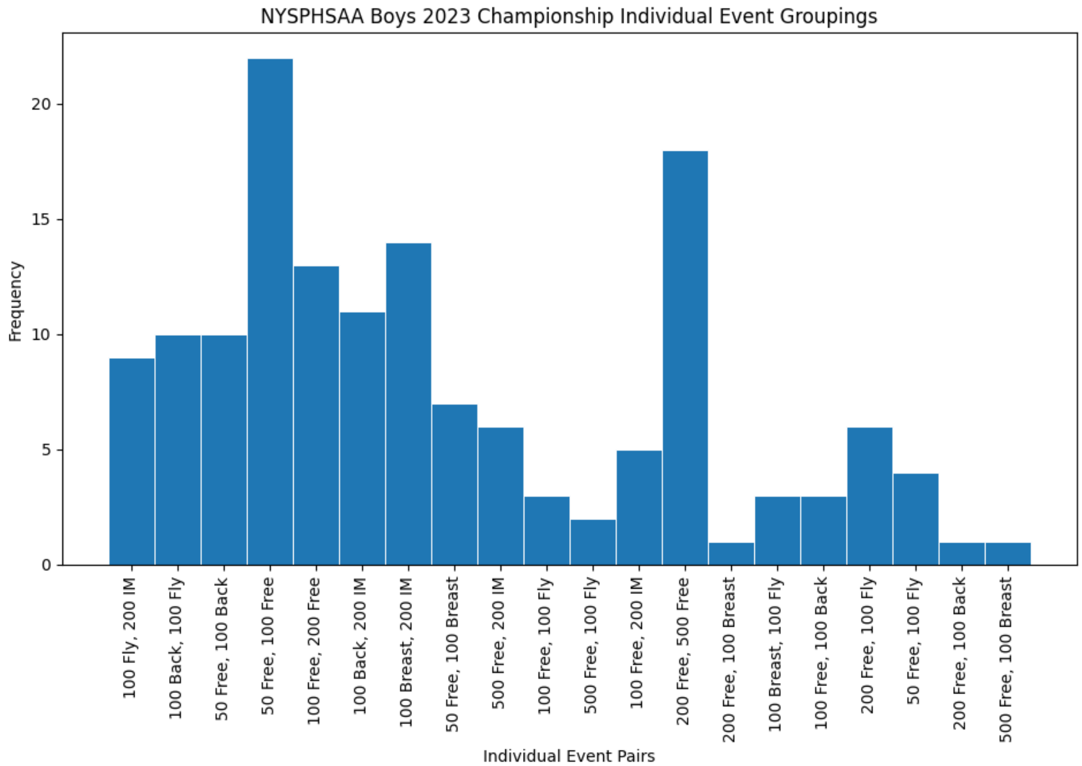

FIGURE 1. Of the swimmers at the boys meet who qualified in two individual events, this figure shows the popularity of each pair. The top five are as follows: 50 Free/100 Free (22), 200 Free/500 Free (18), 100 Breast/200 IM (14), 100 Free/200 Free (13), 100 Back/200 IM (11).

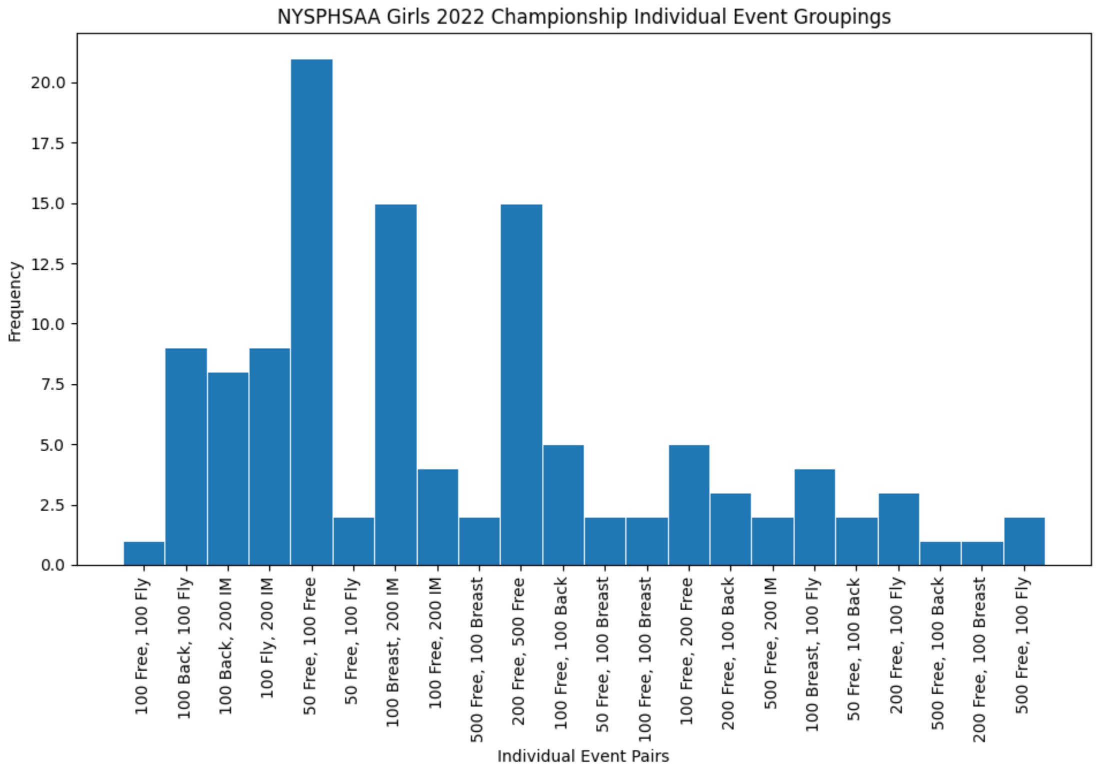

FIGURE 2. Of the swimmers at the girls meet who qualified in two individual events, this figure shows the popularity of each pair. The top five are as follows: 50 Free/100 Free (21), 200 Free/500 Free (15), 100 Breast/200 IM (15), 100 Back/100 Fly (9), 100 Fly/200 IM (9).

Although expected, it is interesting to see that certain pairs are indeed more popular, with the sprint events being first for both the girls and boys meets. This could potentially indicate that these events have a relatively easier qualification time, or New York State is a more sprint focused state or even just that high school swimmers tend to focus on shorter events. Regardless, it comes at no surprise that when combining the data, the same top three pairs emerge (see Figure 3).

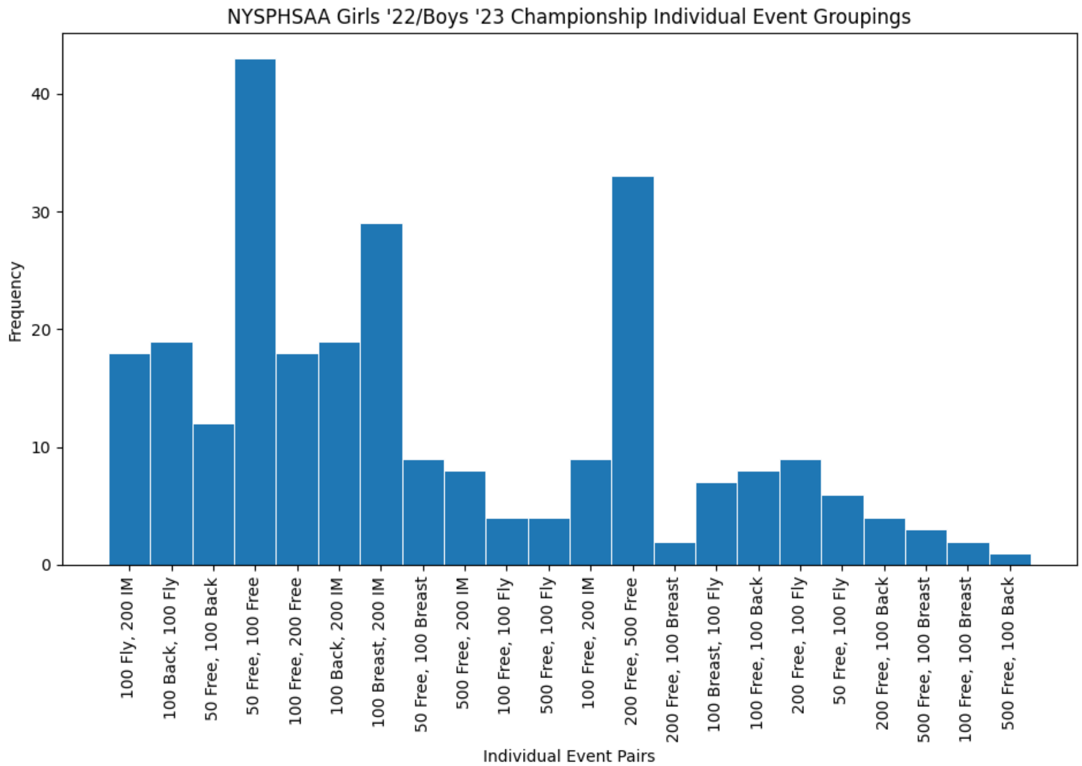

FIGURE 3. Plot generated from combining the totals presented in Figures 1 and 2.

The permutation that maximized the objective function with the naive rest function is provided below (see Table 2). The order of events used presently is listed again to right as a reference. The solution was originally found in the reversed order due to the order in which permutations were generated but is displayed in this order to more easily compare to the typical order. To give context to the listed scores, approximate 50% of scores are less than 1800 and 90% are less than 2200 [1].

Furthermore, it is noteworthy how the calculated solution keeps clusters of events from the typical order together. More specifically, the model result reverses the the 50 FR, 200 IM, 200 Free from their normal order, keeps Diving and the 100 Fly the same, swaps the 500 Free and 100 Free, and reverses 100 Breast, 100 Back, 200 FR from the typical order. Other than keeping clusters

[1] Taken from an IID sample of $n = 100000$ that returned the following 10% wide quantiles: 884, 1449, 1566, 1654, 1732, 1807, 1883, 1966, 2067, 2204, 2797.

Event	Model Result	Typical Order
1	200 MR	200 MR
2	400 FR	200 Free
3	50 Free	200 IM
4	200 IM	50 Free
5	200 Free	Diving
6	Diving	100 Fly
7	100 Fly	100 Free
8	500 Free	500 Free
9	100 Free	200 FR
10	100 Breast	100 Back
11	100 Back	100 Breast
12	200 FR	400 FR
score	2857	2574

Table 2. Model result compared to order specified by the NFHS.

of events together, it is important to highlight two other interesting facts.

(I) Putting the 200 MR and 400 FR back to back is bizarre from a traditional meet organization standpoint. Typically relays are spread out by at least a couple events as they serve as a way to break up the meet and make it more exciting for spectators, something the model has no sense of.

(II) The other characteristic that stands out is the relative positioning of each event cluster compared to the typical order. Despite the events within clusters being reversed, keeping the clusters in the same position was objectively best.

Even though the typical order is technically sub-optimal from the perspective of the objective function, it still shares an uncanny number of similarities to the model-calculated optimal solution.

5. Discussion

A few weaknesses of the model stand out immediately. Due to the exhaustive search, the solution process is not scalable to multi-day meets with many more events. Single day-college dual meets however might still be able to be optimized in this fashion.

A similar study done by Kevin Hallman, a collegiate swimmer [4] also used an exhaustive search but instead tried to minimize the number of events athletes would have to do on the same day (for the NCAA Championships), reducing the input space considerably.

If this model or similar was to be used for larger events, the way data is collected should be treated with more care. Often high-school age group meet standards are set to keep the total number of swimmers at the meet constant. However, sampling from a certain college or even college conference might skew the results as certain schools can be known for a particular demographic of swimmers.

Swimming is such a data rich sport that no detail goes unnoticed in the pursuit of performance. Even the International Olympic Committee is searching for better event orders; a new layout was announced for the upcoming 2024 Paris Games [5].

6. Conclusion

Just by collecting the data necessary, this project was able to answer which event groupings were most popular. This was to be expected as specialization happens in any field but it is always good to back up experience with data. Performing a similar study for collegiate and professional meets would hopefully further solidify the concepts here as well as bring to light the popular groupings out of a much larger selection of events.

Even though a relatively simple model was used to evaluate this problem, a reasonable solution that supports the use of the current event order was found. Despite a "better" order being found, it is not supported with enough evidence to make a case for it to be used as the national standard.

7. Code

All code and collected data is publicly available on GitHub here. Python 3.11 along with the Pandas data analysis library was used to clean the data as well as run the model. Figures were created with Matplotlib.

References

[1] NFHS. Swimming and Diving Distance Conversion Chart. https://www.nfhs.org/sports-resource-content/swimming-and-diving-distance-conversion-chart/. Accessed: 2023-12-09.

[2] Swimcloud. NYSPHSAA Boys Federation Championship, Mar 3–4, 2023. https://www.swimcloud.com/results/237390/. Accessed: 2023-12-09.

[3] Swimcloud. NYSPHSAA Girls Federation Championship, Nov 18–19, 2022. https://www.swimcloud.com/results/237195/. Accessed: 2023-12-09.

[4] SwimSwam, Kevin Hallman. Building a better meet schedule, May 8th, 2020. Accessed: 2023-12-09.

[5] Paris 2024. Olympic Competition Schedule by session v4.1. https://medias.paris2024.org/uploads/2023/04/Competition-schedule-by-sessionevents-V4.1.pdf. Accessed: 2023-12-09.