

Case Study 2: Olympic Team Selection

2023-11-01

Introduction

As the Paris 2024 Olympic Games draw closer, the United States gymnastics program is getting ready to continue their dominance. The US women's artistic gymnastics team is one of the most decorated nations in the sport. The program has won three team gold medals at the 1996, 2012, and 2016 Olympics seven team gold medals at the World Championships. The program also trained the most decorated female gymnast - Simone Biles - who holds 37 world and Olympic medals. Beyond just their success in the gym, the program has created an international hype around them with their Olympic gold medal teams receiving the nicknames of the Magnificent Seven, the Fierce Five, and the Final Five. At their last Olympic showing, they attracted an average of 17.4 million total viewers for their appearance in the finals. While receiving less of a media frenzy than the women's team, the US men's artistic gymnastics team are still strong competitors. Although they received no medals in Tokyo, the program has achieved 6 medals in their history and look to make a strong appearance in Paris.

In order to make a strong showing from both teams and continue the legacy of the women's program, choosing the best possible team is essential. USA Gymnastics has nearly 200,000 members from their beginning training to their elite levels, but few can be taken to the international stage to represent their country. Both teams will only be allowed to bring five members to the Paris 2024 Olympics. The women's team will compete on four apparatuses: vault, uneven bars, balance beam, and floor exercise. The men's team will compete on six apparatuses: floor exercise, pommel horse, still rings, vault, parallel bars, and high bar. Both teams will also compete in the team and individual all-around, allowing for 6 opportunities to medal for the women and 8 for the men. Out of these 5 member teams, competition will follow a 5-4-3 format (4 members out of the 5 compete on each apparatus, and the top 3 scores count). In the team all-around this changes to a 5-3-3 format, so every score counts for the 3 members chosen to compete. For the individual competitions, only 2 gymnasts max per country can qualify to move on to the finals.

With all of these considerations in terms of format and competition stages, selecting a team is very challenging. Selection takes into consideration not only scores from previous competitions, but also protecting the ability to defend a medal or title. It is more than just looking at who has the highest score, since a team has to be selected that can compete across all of these events and do their best at each apparatus. This selection is normally done by USA Gymnastics (for the women it was done by the Karolyis before the recent scandal), but we will create a more analytical way to select the team. Our analysis is centered around the goal of selecting 5-member men and women's gymnastics teams that will maximize medal counts at the Paris 2024 Olympic Games. This analysis is inspired by the UCSAS 2024 Data Challenge run by the University of Connecticut.

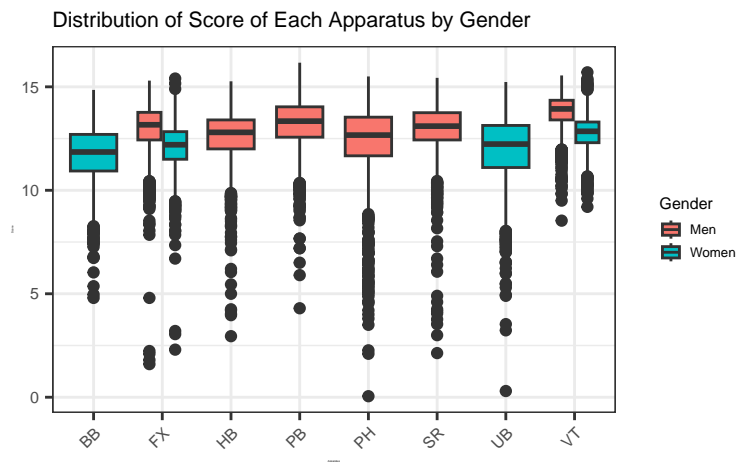
To achieve this goal, we are assessing data from past performances at recent competitions. Each of these competitions have scores for a gymnast's performance at a competition on one apparatus. These scores are a combination of an execution score (E Score) and a difficulty score (D Score). There also is a possible penalty that can be applied to a gymnasts score.

While previous literature on the topic of statistical analysis of Olympic team selection does not exist, there have been other attempts to apply statistical methods to the issue of team selection within the English Premier League. A 2013 paper by Jarvandi et. al used semi-Markov modeling with the issue of player selection in club soccer. Their analysis worked to predict the outcome of transfers within the League, to determine how compatible a player was with the rest of the team. This paper worked to predict the goal differential of this new addition to the team, much like we aim to predict the total score achieved by a gymnast. A much more recent paper by Nouraie et. al tackled player selection and position assignment using a deep neural

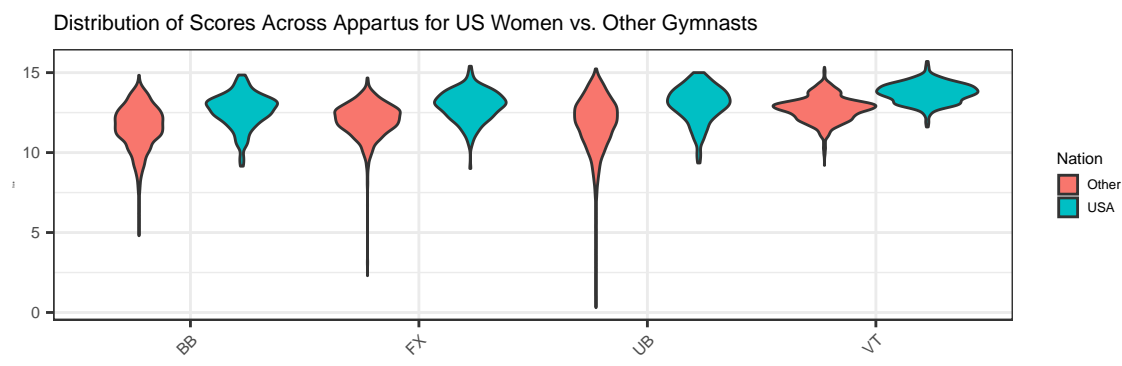
network approach. This study involved much more involved data than we have access to (such as GPS data to analyze player position on the field), but nonetheless motivates the importance of this work - across all sports that require assembling a team. Applying such a research task to gymnastics is a new area of study, but is important to be able to optimize the performance of the United States' 2024 gymnasts.

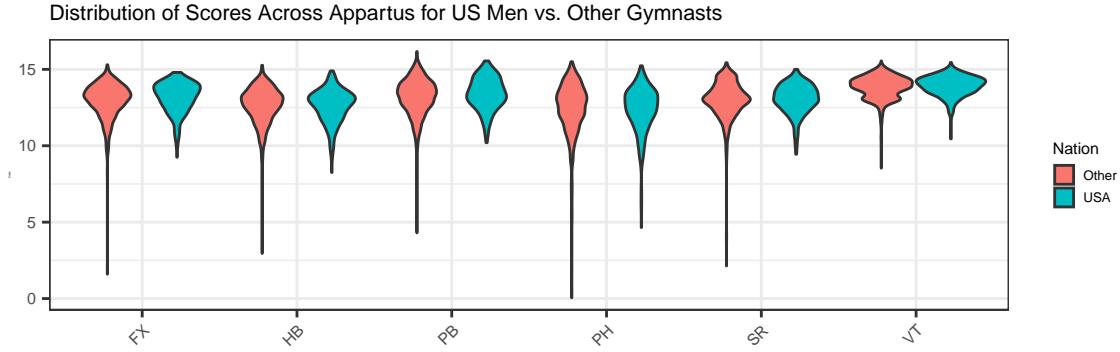
Exploratory Data Analysis

We started by looking at the distribution of total scores (the combination of both the E and D scores) for all gymnasts of both genders across all apparatuses. For most of the apparatuses, the distributions had the largest number of outliers towards the lower end of the distribution and a left skew. This means that there were many gymnasts who had unusually low scores compared to the rest of the score distribution. Scores were more clustered towards the upper end of the distribution, with the average scores for all of the apparatuses being around 13 points. For both genders, the vault and floor events had the smallest range of scores.



We also assessed the distribution of scores for US men and women gymnasts against the distributions for competitors for other countries. For the women gymnasts, across all apparatuses the average score was higher for US women. This makes sense given their impressive history in the sport. The US men's distributions were on par with the distributions for gymnasts from other countries. The averages for each apparatus were mostly the same.





Methodology

1. Determining Constraint Level

In order to narrow down the field from the 163 men and 59 women American gymnasts included in our dataset, we aimed to set a constraint level that a gymnast must meet to be considered. To do this, we first filtered out any gymnast that hadn't competed in 2023. This was due to both the need for relevant data to assess what state the gymnast will be in during the 2024 competitions, but also to weed out anyone that has been injured so couldn't recently compete or who has retired after the 2022 competitions. Taking out gymnasts who hadn't competed in the past year reduced the size of our selection pool to 142 men and 47 women.

We also filtered out gymnasts that had too low of an average score to warrant further consideration and even further reduce our selection pool. To do this, we calculated the average score across all competitions and apparatuses for each US gymnast (both male and female). We then found the first quartile value for each gender's average score. This means the average score that 25% of the average scores for male and female gymnasts respectively fell below this value. The 25th percentile score for US women gymnasts was 12.15156 and 12.15000 for men. After filtering out any competitors who had average scores below that, we were left with 106 men and 35 women.

2. Heuristically Analyzing Top Performers

Next, we aimed to heuristically narrow down the teams even more by considering how the process works in real life. The top two all-around finishers are automatically named to the Olympic team, which inspired our current analysis to find two individuals that should be part of a strong lineup for both the mens and womens team. We did this by taking the previous data for each performer and creating a weight metric and adjusted score that created an "adjusted" composite score to help us determine a narrower set of individuals that should, in expectation, be on the final 5.

First, we designed a formula that would weight certain historic results more/less than others based on it: 1) being in a final round, and 2) being in a more recent performance. Final round performances contribute 1.5x the weight and we defined a linear function where outcomes carry between an additional 1 to 2.5x weighing factor depending on the month and year of the tournament (January of 2022 was set as 1x and December of 2023 was set as 2.5x).

Second, we proposed a new composite score to create a new individual scoring system based on past performances that heavily increase the positive importance of top 5 ranks and also the importance (negatively) of ranking outside the top 10. We believe that consistently scoring within the top 10 is crucial for ensuring Olympic placements as the competition is much higher and even scoring within the 10-20 range will be insufficient for improving Team USA's score. For our adjusted score, gold, silver, and bronze medals contribute 10x, 7.5x, and 5x the weight of a rank outside the top 10, while placement within the 6th-10th range contributes a 2.5 weight.

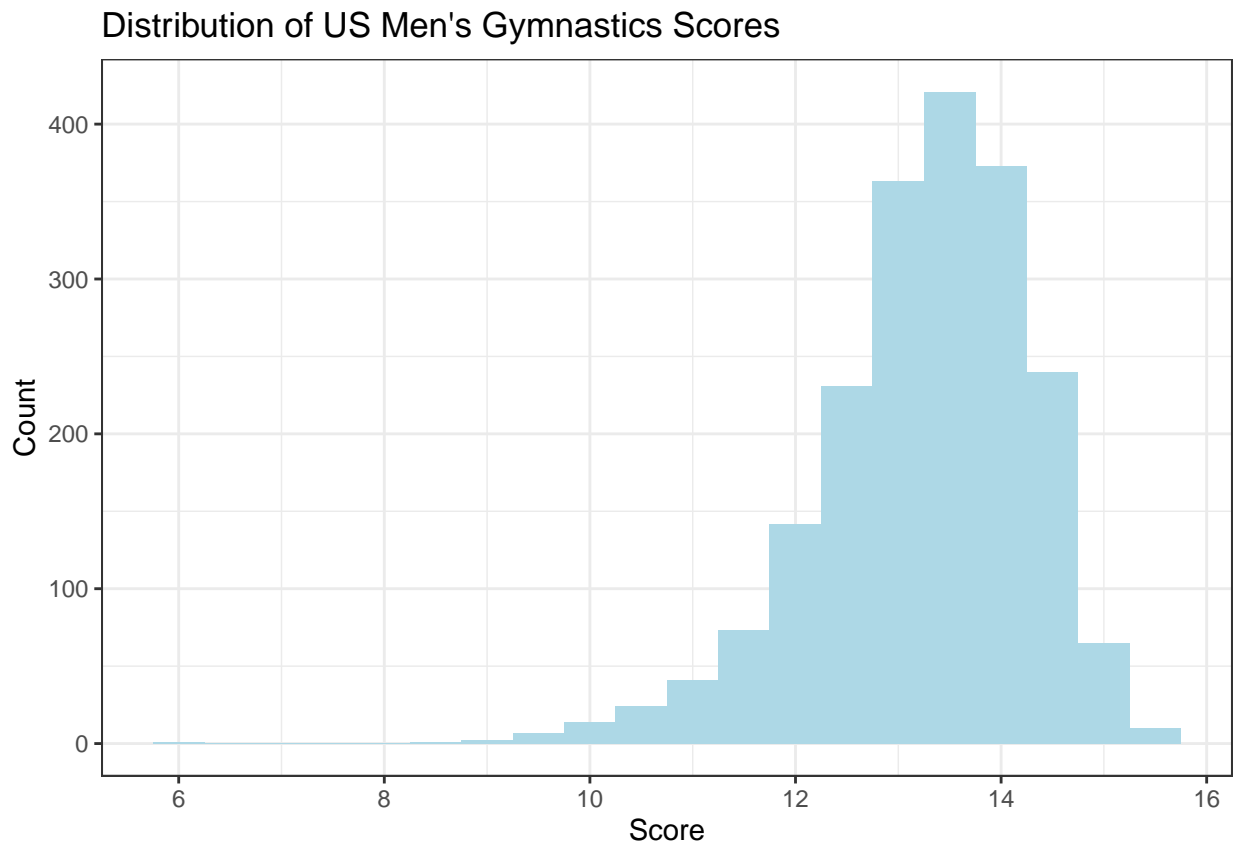
With these adjusted metrics, we produce an adjusted score for each event & apparatus that a gymnast performed in, and consider three different methods of analyzing these metrics:

1. Total sum of adjusted scores: this metric will provide us with a comprehensive way to determine a ranking of the top all-around performers, based on the sum of their weighted adjusted scores in each apparatus they perform in, that we would like to be staples on any team.
2. Best individual apparatus score: this metric will provide a ranking of the highest weighted score a competitor has in any individual apparatus.
3. Best two-apparatus score: this metric will provide a ranking of the average of the two highest scores a competitor has in their two top apparatus.

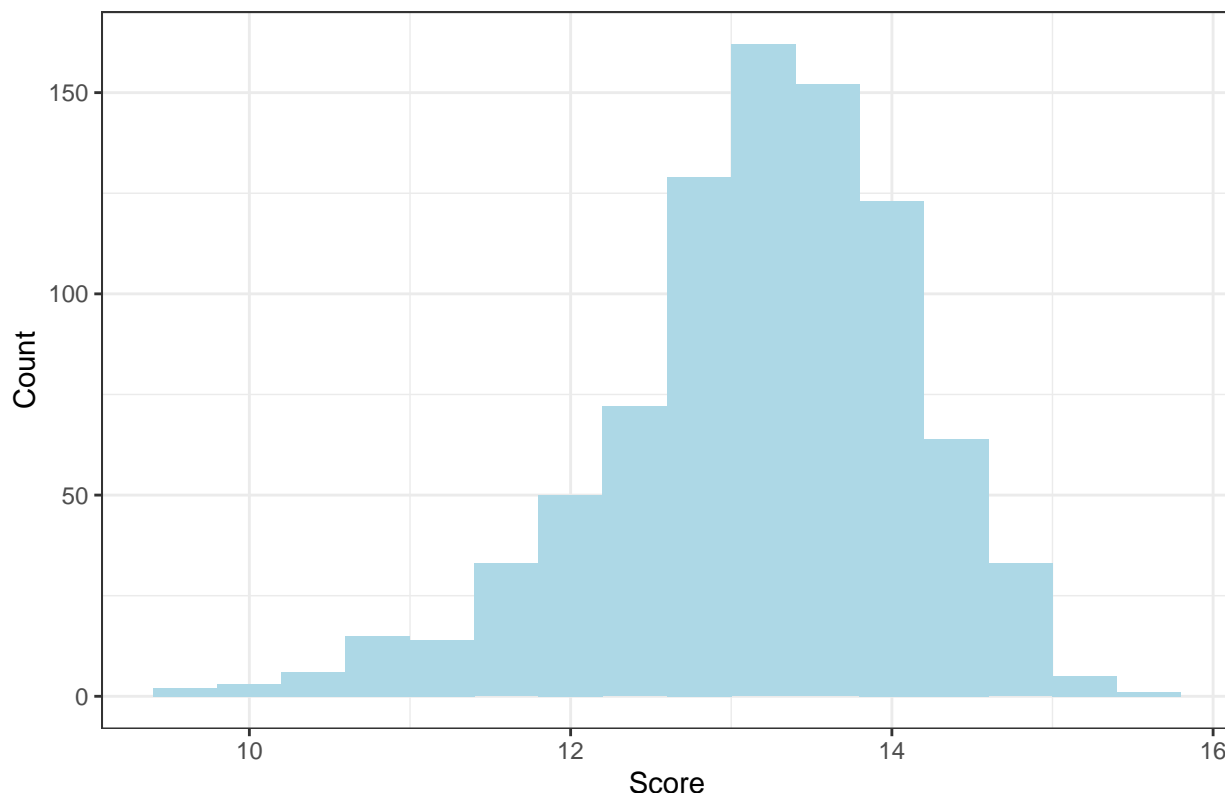
We aimed to use the 2nd and 3rd metrics to determine individuals that would be very likely to excel in a few apparatuses, alongside the two best all-around performers.

3. Simulating Composite Scores

The simulation of composite team scores our methodology can be broken down into two parts. To prepare the data for our simulation of composite scores, we must fit a distribution to describe the qualifying scores of each gymnast. Because each gymnast in the dataset has varying amounts of qualifying data, we believe these individual distributions will be more descriptive than a summary statistic, such as just a mean score. We anticipate that an understanding of the standard deviation of scores will help optimize the accuracy of our simulation. To assess which distribution class would best fit our score data, we first displayed the histograms for both men's and women's qualifier scores.



Distribution of US Women's Gymnastics Scores



Despite the fact that our data is bounded (scores are between 0-16), the concentration of scores in our histogram revealed symmetry around approximately the 13 score range. We therefore elected to fit the scores of each gymnast to a normal distribution, as opposed to a beta distribution or a truncated normal.

To fit these distributions, we will start by iterating through the rows of our US Mens and USA Womens data separately. For each name in the data, we will take the corresponding subset of the data containing all qualifying scores. Using these scores, we will fit a distribution to the associated gymnast, and store the mean and standard deviation, as well as the name of the gymnast into a data frame.

To eliminate redundant information in our final table, we will first use regular expressions to extract the relevant name information (i.e. Simone-BILES instead of Simone-BILES [1]), and then use the distinct function so that each gymnast has just one row in our data frame. We will repeat this process so that we will have two separate data frames containing the names, mean scores, and standard deviation of scores for both men and women USA gymnasts. We can now use these data frames to run our composite score simulation.

The second step in our methodology is running our simulation to generate composite team scores for randomly sampled groups of six USA gymnasts. Using the output of the second step in our methodology, we can specifically sample teams from a smaller subset of 10 male gymnasts and 10 female gymnasts that were selected to either be the best all-around gymnasts, or gymnasts that were highly proficient in a few events. We start our simulation by randomly sampling five different gymnast names, means, and standard deviations without replacement (men and women separately). For each gymnast sampled, we will generate one score based on the gymnasts' mean and standard deviation of qualifying scores using the `rnorm()` function. We will then add this value to our variable for the composite score. In cases where our standard deviation was NA, we assigned a generic deviation of zero. When our standard deviation value was less than $1e-10$, we simply assigned the mean score for that gymnast as our simulated score, as these scores often resulted in run-time errors. We will run this simulation process 10,000 times for the men's data set, as well as 5,000 for the women's data set (as it is much smaller), storing the composition of gymnasts and the associated composite score. Using this data frame, we can observe the combinations in descending order to reveal which

gymnasts contributed to the highest composite score.

Table 1: Top 10 Male Groups of 5

Gymnasts	Composite_Score
Taylor-BURKHART, Asher-HONG, Khoi-YOUNG, Brody-MALONE, Stephen-NEDOROSCIK	76.92
Stephen-NEDOROSCIK, Khoi-YOUNG, Joshua-KARNES, Curran-PHILLIPS, Frederick-RICHARD	76.49
Frederick-RICHARD, Brody-MALONE, Donnell-WHITTENBURG, Stephen-NEDOROSCIK, Khoi-YOUNG	76.22
Paul-JUDA, Asher-HONG, Brody-MALONE, Frederick-RICHARD, Stephen-NEDOROSCIK	75.97
Stephen-NEDOROSCIK, Asher-HONG, Paul-JUDA, Brody-MALONE, Khoi-YOUNG	75.78
Curran-PHILLIPS, Taylor-BURKHART, Paul-JUDA, Brody-MALONE, Khoi-YOUNG	75.50
Stephen-NEDOROSCIK, Curran-PHILLIPS, Brody-MALONE, Paul-JUDA, Khoi-YOUNG	75.45
Frederick-RICHARD, Curran-PHILLIPS, Asher-HONG, Brody-MALONE, Stephen-NEDOROSCIK	75.24
Brody-MALONE, Stephen-NEDOROSCIK, Asher-HONG, Donnell-WHITTENBURG, Joshua-KARNES	75.20
Joshua-KARNES, Stephen-NEDOROSCIK, Paul-JUDA, Curran-PHILLIPS, Brody-MALONE	75.17

Table 2: Top 10 Female Groups of 5

Gymnasts	Composite_Score
Zoe-MILLER, Jade-CAREY, Simone-BILES, Shilese-JONES, Jordan-CHILES	74.24
Shilese-JONES, Ashlee-SULLIVAN, Jordan-CHILES, Kaliya-LINCOLN, Zoe-MILLER	74.13
Jade-CAREY, Skye-BLAKELY, Kayla-DICELLO, Zoe-MILLER, Shilese-JONES	74.06
Skye-BLAKELY, Jade-CAREY, Shilese-JONES, Simone-BILES, Jordan-CHILES	73.98
Shilese-JONES, Kayla-DICELLO, Kaliya-LINCOLN, Simone-BILES, Jade-CAREY	73.68
Zoe-MILLER, Jade-CAREY, Joscelyn-ROBERSON, Jordan-CHILES, Simone-BILES	73.67
Joscelyn-ROBERSON, Jordan-CHILES, Jade-CAREY, Simone-BILES, Kaliya-LINCOLN	73.60
Ashlee-SULLIVAN, Jade-CAREY, Kayla-DICELLO, Simone-BILES, Joscelyn-ROBERSON	73.59
Kayla-DICELLO, Jordan-CHILES, Shilese-JONES, Jade-CAREY, Kaliya-LINCOLN	73.39
Simone-BILES, Jordan-CHILES, Jade-CAREY, Kaliya-LINCOLN, Zoe-MILLER	73.37

4. Simulating Competitions in the Olympic with All Other Countries

After narrowing down the set of potential 5-member teams for men and women, we ran simulations to see how each candidate US team would perform if placed in the pool with all other teams from the world. To do this, we need to create participating teams for other countries. Here we assumed all countries had a team consisting of 5 members for men and women respectively. If the number of distinct gymnasts for a gender from a country is less than 5, we selected all the gymnasts from that country in the data set to be in the Olympic team. The men/women teams for other countries were randomly selected, so the teams may not be their “best” teams. We used random selections here because 1) it is very unlikely that all countries use identical method as the US to select the “best” team to participate in the Olympic and 2) our aim in this part is to compare the performance of narrowed-down US teams, so randomization in teams from other countries serves as a control.

We then placed each potential US team and all teams from other countries into simulated competitions. The simulated score of a gymnast in an apparatus was generated based on the mean and standard deviations of their historical scores using the `rnorm()` function. With each individual score, we then used the same rules as the Olympic gymnastic competitions to get the final rankings of US team in each of the 8 events for men and 6 events for women: 5-4-3 format for the qualifying round, 5-3-3 format for the team all-round, maximum of two gymnasts per country for the individual all-round, and maximum of 2 gymnasts per country for each apparatus final, except that we integrated the 5-4-3 and 5-3-3 format together, removing the need of taking the precedence of the qualifying rounds over finals into consideration.

For each simulated competition result, we used a weighted metrics to calculate a “total medal score” for the US team - a score of 3 for gold, 2 for silver, and 1 for bronze. We then computed the mean of this total medal score from all iterations of simulation for each potential US team and compared their scores.

Results

Subsetting

After our initial filtering removed participants that were well outside of a top 5 range, we used our heuristics described in the methodology to further subset a top 10 sample of mens participants and womens participants using a new weighted composite score. We used 3 different metrics to understand this data: the sum of composite scores across each apparatus, the maximum individual apparatus score, and the average top 2 apparatus score for each individual.

Starting with men’s results, we were able to pinpoint Brody Malone as being the clear top overall performer in the men’s division, without an obvious number 2 all-around performer (and opted to use our simulations to fill in the rest of the team).

Below, is the overall composite score top 10.

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `across(where(is.numeric), round, 2)`.
```

Caused by warning:

```
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
##   # Previously
##   across(a:b, mean, na.rm = TRUE)
##
##   # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))
```

Table 3: Overall Top 10

Name	Score
Brody MALONE	249.80
Donnell WHITTENBURG	195.52
Frederick Nathaniel RICHARD	190.83
Asher HONG	189.27
Paul JUDA	178.88
Taylor BURKHART	174.99
Joshua Andrew KARNES	171.18
Khoi YOUNG	169.03
Colt WALKER	165.77
Cameron BOCK	158.15

Followed by the individual apparatus max score top 10.

Table 4: Single Apparatus Top 10

Name	Score
Brody MALONE	84.76
Curran PHILLIPS	77.29
Donnell WHITTENBURG	67.79
Stephen NEDOROSCIK	60.96
Frederick Nathaniel RICHARD	49.17
Ian SKIRKEY	48.39
Khoi YOUNG	47.75
Taylor BURKHART	47.43
Paul JUDA	47.26
Colt WALKER	47.09

And finally the best two apparatus max score top 10.

Table 5: Two Apparatus Avg Top 10

Name	Score
Brody MALONE	70.18
Curran PHILLIPS	58.98
Donnell WHITTENBURG	52.35
Khoi YOUNG	46.47
Taylor BURKHART	45.37
Paul JUDA	44.40
Frederick Nathaniel RICHARD	44.31
Asher HONG	43.30
Riley LOOS	39.99
Joshua Andrew KARNES	38.70

From these results, we are able to get insight into how each athlete would fit on a team. Brody Malone and Donnell Whittenberg seem to consistently perform well across multiple events and score very highly, which means that we could expect at least one of them to anchor a team as an all-around performer. Certain athletes like Curran Phillips, Khoi Young, and Taylor Purkhart excel at specific apparatuses, which may make them valuable to earn a medal in a single event.

Next, moving onto the women’s results. We were able to pinpoint Simone Biles as the clear top performer as an all-around athlete, but following was Jordan Chiles with a high composite score that also secures a spot as an all-around athlete.

Simulating Composite Scores

Combining our analysis in parts two and three, we were able to calculate the combination of mens and womens USA teams that elicited the highest composite scores in our simulation. After setting a random seed and running a simulation on the men’s data 10,000 times, our results yielded that the highest composite scoring men’s team includes Taylor Burkhart, Asher Hong, Khoi Young, Brody Malone, and Stephen Nedoroscik. This team combination generated a composite score of approximately 76.925. In other words, the average score for each member of this team is 15.385. This mean score is much higher than the individual means of the gymnasts. It is possible that the number of samples in our simulation slightly favors gymnasts with higher variances, distorting top scores to be slightly higher than expected. This holds true for the next closest team (as well as in our women’s data set), consisting of Stephen Nedoroscik, Khoi Young, Joshua Karnes, Curran Phillips, and Frederick Richard, with a score of 76.486 (A score of 15.297 per gymnast on average).

Our USA women’s output from our simulation yielded that the highest composite scoring team has a roster of Zoe Miller, Jade Carey, Simone Biles, Shilese Jones, and Jordan Chiles. The composite score for this roster was 74.240. Under this total score assumption, each gymnast would score a 14.848 on average. The second closest team features three of the same gymnasts in its overall lineup of Shilese Jones, Ashlee Sullivan, Jordan Chiles, Kaliya Lincoln, and Zoe Miller, with a total score of approximately 74.128 (about 14.826 on average). The overlap between gymnasts in top-composite-scoring teams for both men and women is a source of confidence in our model.

While average scores are ultimately high, and our number of simulations may occasionally favor gymnasts with high variances, our outputs give us a reasonable estimate of a starting combination for both the men’s and women’s 2024 Olympics lineup.

Simulating Medals

Our preliminary simulation for medaling provides us a way to simulate how our team would perform against other countries. Using the methodology proposed to create an international team to test the US team against, we are able to see how our 2 lineups would score against other teams.

Using a scoring system of 3 points for Gold, 2 points for Silver, and 1 point for Bronze, our men’s team most frequently earned 9 points, with little variance (a minimum of 6 points and a maximum of 13 points). Our women’s team performed stronger and averaged closer to 14 points in the simulation with higher variance (from 11 points to 22 points).

Discussion

In our analysis we aimed to address the goal of determining both a mens and womens artistic gymnastics team for the United States of America to send to the Paris 2024 Olympics. We wanted this team to be the optimal selection, maximizing the medal count at the games. We were able to achieve this through our four-step methodology. First, we set a constraint criteria that limited our selection pool. We then heuristically analyzed our top performs mirroring the selection process used for past games. After that, we simulated composite scores for our US Olympics teams and finally compared those against simulated scores for other nations’ Olympic teams.

There were a few limitations with the data. One was that the dataset itself did not include information on defending champions - something that is common for Olympic selection committees to weigh highly in their decision making. Although we were able to take this into account by making sure to include defending champions such as Simone Biles (and the men’s team had no defending champions to address), it would be an interesting addition to our work to have data on this. Another limitation is that this data just included scores in competition. There was no additional information about the athletes’ performances, such as if they are currently healthy or injured or even if they plan to or want to compete in the Paris 2024 Olympics. While we mitigated this by selecting only gymnasts with recent performances, this additional information would have improved our analysis.

We were able to address our ultimate goal by forming two teams that stand a strong chance in Paris. The men’s team that we selected consisted of Taylor Burkhart, Asher Hong, Khoi Young, Brody Malone, and Stephen Nedoroscik. Our results showed that this team had an average score of 15.385 per member and a composite score of 76.925. These scores were much higher than average, and after simulating scores for other nations show that this team has the strongest chance of bringing home the most medals.

Our womens’ team selection further addressed our ultimate goal of Paris 2024 dominance. Our team of Zoe Miller, Jade Carey, Simone Biles, Shilese Jones, and Jordan Chiles features two defending medalists. Their simulated average score of 14.848 per member gives them a competitive advantage which is supported by their composite score of 74.240.

Our analysis has produced teams with the highest simulated average and composite scores. Compared against simulated scores for teams from other nations, our selected US teams stand to perform better than scores from those teams. Our selection process has created the strongest chances for Team USA in the 2024 Olympics,

and provides an exciting new way that statistical analysis can be applied to this previously subjective process of team selection.

Sources

Artistic Gymnastics. Paris 2024. (2023, June 1). <https://www.paris2024.org/en/sport/artistic-gymnastics/>

Jarvandi, Ali, Sarkani, Shahram and Mazzuchi, Thomas. “Modeling team compatibility factors using a semi-Markov decision process: a data-driven approach to player selection in soccer” *Journal of Quantitative Analysis in Sports*, vol. 9, no. 4, 2013, pp. 347-366. <https://doi.org/10.1515/jqas-2012-0054>

Nouraie, M., Eslahchi, C. & Baca, A. Intelligent team formation and player selection: a data-driven approach for football coaches. *Appl Intell* (2023). <https://doi.org/10.1007/s10489-023-05150-x>

Wikimedia Foundation. (2023a, October 8). United States men’s National Artistic Gymnastics Team. Wikipedia. https://en.wikipedia.org/wiki/United_States_men%27s_national_artistic_gymnastics_team

Wikimedia Foundation. (2023, November 13). United States women’s National Artistic Gymnastics Team. Wikipedia. https://en.wikipedia.org/wiki/United_States_women%27s_national_artistic_gymnastics_team