

Vaulting into Victory: Optimizing for US Gymnastics Medal Count at the Paris 2024 Olympics

Sean Li, Christopher Tsai, Benjamin Thorpe, Jerry Xin

2023-11-21

Reponse to Review

Looking at the feedback provided in Gradescope and talking to the instructor and TAs, we have responded to feedback in our new report. Our changes are summarized below.

Because our introduction gained full points, we decided not to change anything within the introduction. We decided that it gave a great intro to USA and Olympic Gymnastics, the research objective of the case study we wanted to solve, our dataset, as well as exploratory data analysis. We added a more creative title to our case study.

Looking at the methodology section, we were able to perform a more “qualitative” analysis of choosing a team in our final report. In our initial report, we mentioned that our results are aligned with the “eye-test” in our conclusion but provided no formal methodology for this “eye-test”. This qualitative analysis strives to provide more structure to this. We mention how we did our qualitative analysis and add to our findings in the discussion section.

We can also justify simulating the mean score for each gymnast in response to your comment if that is the correct statistic to use. We think that the mean is still reasonable to do here for 2 reasons. One, the distribution of scores for each gymnast is still roughly normal, which means that the mean is a good summary statistic, and because it makes sense with our methodology. As a reminder, we take the top 24 candidates for each event/apparatus, simulate their mean scores, and then rank them. Next, we compare the mean of our selected candidate for the specific apparatus to find a predicted medal count. Since we use the mean to find the predicted medal cutoffs, it only makes sense to use the mean score for our selected candidate (for the specific apparatus). We simulate the mean by drawing this score repeatedly from a T-distribution, which also will account for individual variability around the mean, and potentially can obtain scores that are not within our dataset. We considered the median, but it will exclude outlier performances, which we actually want. Outliers in any sport are indicative of a real performance outcome that is possible. We also considered using something such as Q3, but there is no sufficient evidence to assume that athletes will perform at their own 75th percentile at the Olympics. Furthermore, it eases the consistency during comparison since we are simulating predicted medal cutoffs and counts with the means, it only makes sense to use the mean to compare against these medal cutoffs. Lastly, our method is robust to outliers because takes the top 24 athletes for each event apparatus, and draws samples repeatedly from this elite distribution.

Looking at the results section, the only comment was to add table headers, and we successfully added table headers for the tables given in our appendix. We also double checked each graph for formatting.

Lastly, because we got full points for the conclusion and discussion section, we thought that our alternative approaches, limitations, and discussion of final results was adequate and not to change anything significantly, except adding in a qualitative analysis of the team to better supplement an “eye-test” with background research. We also read over our paper for grammatical errors.

Introduction

Background

The Olympics brings about a sense of national pride seldom felt by athletes and viewers alike. The sheer amount of viewers adds to the pressure and importance of an event like this for national athletic committees. In every Olympic games since 2008, over 3 billion unique viewers tuned in worldwide to view the games. To put that in perspective, the Super Bowl, the most watched individual sporting event in America, garners around 115 million viewers. Americans, in particular, are fascinated by the sport of artistic gymnastics with the highest percentage of those who watch the Olympics claiming to be interested in the event (Tracy, 2021). It follows that success in this sport has long been a source of international recognition for the United States. The breathtaking performances, grace, and unparalleled athleticism displayed by American gymnasts have not only brought numerous Olympic medals home, but also served as an inspiration for generations of aspiring athletes. Simone Biles, Gabby Douglas, and Nastia Liukin are all household names due to their success in Olympic gymnastics. As we approach the next Olympic Games in Paris 2024, it is extremely vital that team USA finds success and brings home as many medals as possible in artistic gymnastics. In the era of data-driven decision-making, the role of predictive analytics in the world of sports has gained unprecedented significance and the need for innovation and precision in athlete selection has never been more crucial. Our goal is to use data analytics to predict the best five-member lineups for both the USA Men's and Women's artistic gymnastics teams, optimizing for total medal count.

First, understanding Olympic gymnastics and its scoring system will be vital to understanding the decision making process later in this paper. Women's artistic gymnastics and Men's artistic gymnastics vary slightly, with women competing in four apparatuses and men competing in six. Each country, if they have qualified for the team all-around competition (which the US has), can send five athletes per gender; otherwise a country can send a maximum of three individuals. The event begins with a qualification round that is also the qualifying round for the team all-around, individual all-around, and individual apparatus competitions. A team will send four of their five athletes to compete for each of the apparatuses. For the team score, only the top three of the four scores will count. If an individual competes on every apparatus in the qualifying round they are then eligible for the individual all-around final, regardless of if they are on a team. However, only the top 24 athletes will be in the individual all-around final. If an individual is in the top eight in qualifying for a given apparatus they also qualify for that individual apparatus final. For all individual finals, however, only a maximum of two athletes per country can qualify. All things considered, this gives six medal opportunities on the women's side (team all-around, individual all-around, each of the four apparatuses) and eight on the men's side (team all-around, individual all-around, each of six apparatuses).

The gymnast selection process has, historically, relied heavily on the expertise and intuition of coaches and selection committees and thus research into data-driven decision making in gymnastics is sparse. While there has been plenty of prior research that has explored the use of data analytics in sports, they often fall short in the context of artistic gymnastics. Many of these studies are limited in scope and do not consider the nuanced aspects of gymnastics performance and individual variance from round to round. Also, we plan to address the fact that some apparatuses (like pommel horse) are more difficult to judge by taking these higher variances in scores that earn medals into account when optimizing our team (Guston, 2023). We will explore how to mathematically account for this in our analysis.

Research Objective

This study aims to overcome the limitations listed above by analyzing athlete-specific data and incorporating domain expertise from the gymnastics community. It will offer a comprehensive and customized approach that addresses the unique competition format presented by artistic gymnastics and provides the USA Olympics Committee with a robust tool to make statistically informed decisions. Our goal is to recommend a team of five women and five men for the Olympic Games in Paris 2024, with an expected individual medal count for each team. We did not include the team medals in our predicted medal counts, but because the team competition scores are calculated through individual performances, optimizing for these should at minimum get us close to the best squad for the team event. In the future, we want to eventually predict for the total medal count that includes the expected team competition performance from the five gymnasts chosen.

Dataset Overview

The data is taken from major domestic and international competitions worldwide. We were provided two datasets, one from the competitions leading up to the Tokyo Games in 2020 (2017 to 2021), and one from the competitions leading up to the Paris Games in 2024 (2022 to 2023). We only used the 2022 to 2023 data, as the competition format changed between 2021 and 2022.

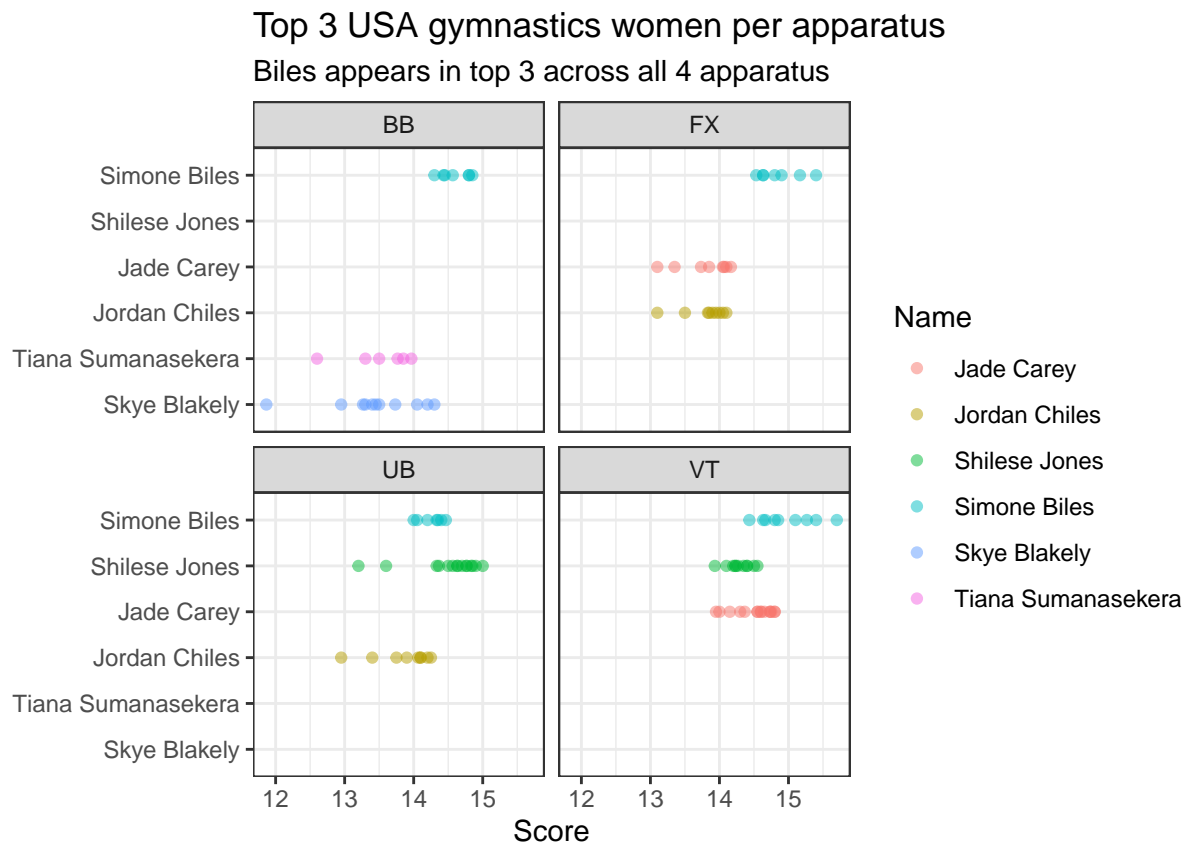
The data consists of competition results from many gymnasts from all different countries. Our categorical variables are FirstName and LastName, for the first and last names of the competitors, gender, country (of the competitor), date (date range of the competition), competition, round (either qualifier or final), location, and apparatus. Our numerical variables include rank (place in competition), D Score (Difficulty Score), E Score (Execution Score), Penalty, and Score. It is important to note that Score is simply the combination of D Score + E Score - Penalty.

The global data had 23,891 observations, and the dataset, when filtered for the USA only, had 3,362 observations. There were 12 features given in the dataset.

Exploratory Data Analysis: Score distributions & top USA candidates

Our primary takeaways from our EDA (appendix A) were that the distribution of scores for each apparatus for men and women in the USA, as well as worldwide, was roughly normal, with somewhat of a left skew and moderately fat tails. First, this implies that there are some under performers we should simply filter out for certain events. We do this by sampling from the top 24 athletes to create our predicted medal threshold. Next, this implies that for each athlete, the distribution of their performance can be approximated by a t-distribution. This in turn inspired us to simulate drawing from a t-distribution, and then comparing these draws with our predicted medal threshold. This threshold was derived from also sampling from a t-distribution in order to simulate each athlete's chance of medaling. This will account for the individual variance of an athlete on a specific apparatus.

If our goal is to put forth the best lineup to maximize medal count, we need to look into who the top US candidates are in each apparatus and all-around. Here, we graph dot plots of the top three US gymnasts in each event across all the events, using the mean score from their 2022-2023 data.



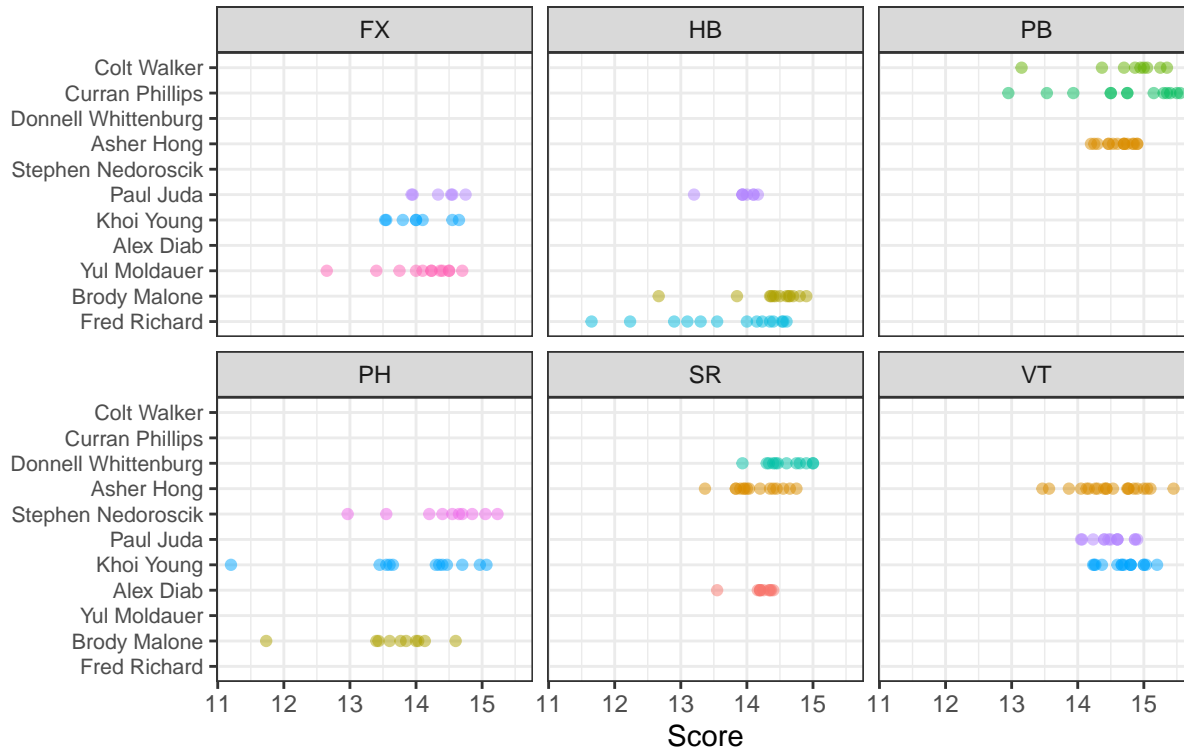
To clarify why some athletes appear but don't have score dots in an apparatus, that athlete was in the top 3 for another apparatus but not this current one. We keep them in just to show the "depth" of US gymnastics. A small

number of total athletes that appear in the top 3 across all apparatus signals that most elite athletes are great at more than 1 event.

Taking a glance at the top US women, we can see that Simone Biles by far and away is the best American gymnast. That is quite unsurprising, given that she has the most gymnastics medals in the world and has consistently performed at an elite level in all events since her return to the sport. We can almost lock her in to make the Paris 2024 roster. Shilese Jones is the only gymnast have a higher mean score in an apparatus than Simone, specifically in the uneven bars (UB). She is someone to keep an eye out for, especially considering her recent all-around Bronze medal in the World Championships in October 2023. There are only six women to place in the top three across four apparatuses, a telling sign that the US women's team selections for Paris 2024 might be top heavy.

Top 3 USA gymnastics men per apparatus

Juda, Young, Hong appear in top 3 in half of the apparatus'



The US mens data is not quite as clear cut. There are a total of 11 American gymnasts who place in the top three of at least one of the six events. Asher Hong, Khoi Young, and Yul Moldauer place in the top three for three out of the events. Additionally, Brody Malone places in the top five for four out of six events, and Fred Richard most recently won bronze in the 2023 World Championships in the all round. Those are athletes we might expect to be placed into our USA mens team.

Our EDA is important as it allows us to sanity check our decision making, and will provide a sturdy foundation to any USA team lineups.

Methodology

Breakdown of All Round + Specialists

Treating this as an optimization problem where we are maximizing for total medal count, we have developed nuanced heuristics based off analyzing past Olympic gymnastic event results to pick our top five-person lineups. Historically, we've seen that the USA, along with most other participating countries, send a combination of all-round gymnasts and event specialists in a five person lineup. Specifically, the USA Olympic Team Selection Procedure outlines that the top two overall all-round gymnasts will automatically qualify for the squad and then the committee will select the other three to complement them. As discussed previously, a team can qualify a maximum of two individual gymnasts for each apparatus final as well as the all-around final. By this logic, it makes the most sense to send two all-around participants, while then having three specialists that would give that country the best shot at medaling

in each individual apparatus. We wanted to mimic this approach, leading us to implement a two-tiered approach in our team selection process. First, we choose the top two overall all-round gymnasts and then we select three more gymnasts to round out the team in an attempt to maximize total medal count.

Selection Method

In selecting the two all-round gymnasts, we will choose the two candidates that have the simulated highest total score average across all the events they have participated in leading up to the 2024 Olympics. Specifically, we weighted each score by artificially determining its importance by how recent it was and the “stage” of the competition in which it took place. We averaged the weighted scores across all the events a US gymnast has participated in. We then sum the average weighted scores in each apparatus to find a composite all-around score.

Clutch Factor and Recency Considerations

We chose to put weights on competition date and competition stage, specifically, because we believe a heavier weight should be placed on performances that take place under pressure in the final round and are more recent. This weighting is our attempt to measure the “clutch” factor and peak/prime performance of an individual athlete. The way our weights will work is we will duplicate rows of individual scores in the dataset given the assigned weight of that score. Competition date will be split into three categories by how recent the competition was: very recent (< 6 months), somewhat recent (6 - 18 months), not recent (18+ months). A multiplier of 1 will be given to not recent competitions, a multiplier of 2 to somewhat recent competitions and a multiplier of 3 for very recent competitions. As for the competition stage, there are two stages of competition: qualifiers and finals. Qualifying round scores will receive a multiplier of 1 and final round scores will receive a multiplier of 2. For example, a row containing a score from a not recent competition and qualifying round will remain as one singular row. However, a row containing a qualifying round score, but from very recent, will be multiplied three times to get three rows containing that score. If that score was instead from the finals, it would be multiplied $2 \times 3 = 6$ times.

Taking the average of an individual athlete’s scores after this duplication process will give us a mean expected score on each apparatus for that athlete. Still, we are aware that each gymnast will have variance in their scores for each apparatus and we wanted to take that into account. For each USA gymnast, we will also develop an individual total score distribution from their weighted scores to account for the variance for each individual on each apparatus. A density plot of all the scores after the duplication process will give us a distribution centered around the mean score for an athlete as well as a variance.

Obtaining Medaling Probabilities

Because the main goal of analysis is to determine the probability that an individual athlete medals, we also had to determine a way to get medal probabilities. We will accomplish this goal by determining medal thresholds, or thresholds necessary to achieve a certain medal. This will first require us to get a distribution of scores on individual apparatuses. We do this by creating a T-distribution from the top 24 individuals’ weighted mean score. Of the 24 people, however, no more than 4 competitors from the same country will be included in this distribution as a maximum of 4 individuals from each country can compete in qualifying for individual apparatuses. From the EDA, we saw that the distribution of scores was roughly normal, with somewhat of a tail towards the left side of the distribution, and somewhat fat tails. This inspired us to take the top 24 people (the lower scoring competitors would be dropped), and from there create a T-distribution to sample from in our simulation, since we have normality and somewhat fat tails.

In order to find an individual’s probability of earning a medal we will use a two step simulation and sampling process. We will simulate the following process 1,000 times. We first will sample from the distribution of the top 24 individuals 24 times. The top three scores of the 24 will represent the medal cutoffs with the 3rd highest sample score corresponding to a Bronze medal, the 2nd highest sample score corresponding to a Silver medal cutoff, and the highest sample score corresponding to a Gold medal cutoff. Because we are only worried about the placing of gymnasts from the USA, we will sample once from each of the top 8 USA gymnasts in each individual apparatus given their unique distributions. If that gymnast’s sampled score surpasses one of the cutoffs they will have ‘earned’ that medal for the given simulation. For example, if when sampling from the 24 gymnasts distribution we get medal cutoffs of 14.5, 14.8, and 14.9 and gymnast A gets a sampled score of 14.6 they will “earn” a Bronze medal. A sampled score of 14.81 would “earn” a Silver medal and a sampled score of 14.91 would “earn” a Gold medal. The only caveat is if multiple American gymnasts fall within the threshold for a certain medal. In that case the gymnasts

would be ordered and the top score earns that medal and the next highest score would earn the next highest medal if it is available. We again simulate this two-tiered sampling approach 1,000 times to get each individual gymnast’s probability of getting a specific medal in a specific apparatus. We will use this approach for the individual all-around as well, and the two men and two women with the highest medal probabilities will be automatically selected for our team. The assumptions for our simulation require an adequate amount of repeating sampling. 1000 is surely enough for each athlete, who on average has 10-20 observations. We also assume that past performance is somewhat indicative of future performance.

Optimization through a qualitative lens

From there we will use optimization to determine the 5-person lineup that maximizes expected medals earned including the team all-around medal. In determining the team all-around medal we will use a system identical to the way the Olympics will work in 2024. We will use many permutations of 3 specialists along with our two guaranteed all-around gymnasts and sample from every gymnast’s individual distributions to get all-around scores. For each apparatus, the top three US gymnasts on the teams distributions will be sampled from (because only the top 3 gymnasts participate in the team all-around finals). The scores will be aggregated to get a total team all-around score. The same method of medal cutoffs will be used in the team all-around medal cutoffs, but using empirical team all-around score data to create a distribution to be sampled from. The 5-person team that maximizes total medals earned will be the team we recommend the USA send to Paris for the 2024 Olympics.

We also would like to supplement our quantitative results with a few lineups that we picked up through intensive background research on recent news, gymnastic community forums, and common expectations. Our selections were made through meticulous video study and reading of the latest competition results and quality news articles. These shall be detailed further in the discussion.

Results

Our Results are compiled in the table below, with predicted individual medals. We used predicted total individual apparatus medals to select our teams, and categorized each competitor as either All Around or Specialist:

Table 1: Proposed Paris 2024 USA Gymnastics Team, with Expected Medal Counts (EMC)

Category	USA_Women	EMC_Women	USA_Men	EMC_Men
All Around	Simone Biles	3.409	Paul Juda	0.663
All Around	Shilese Jones	0.466	Fred Richard	0.491
Specialist	Kaliya Lincoln	0.439	Donnell Whittenburg	0.264
Specialist	Jade Carey	0.190	Khoi Young	0.564
Specialist	Zoe Miller	0.061	Curran Phillips	0.310
Total Exp. Medals		4.565		2.402

To maximize for total medal count at the Paris 2024 Olympics, the USA Gymnastics team should select Simone Biles, Skye Blakely, Shilese Jones, Jade Carey, and Zoe Miller for the women’s team, and Fred Richard, Paul Juda, Donnell Whittenburg, Khoi Young, and Curran Phillips for the men’s team. Our total predicted medal counts are 4.565 individual medals for women, and 2.402 for the men’s team.

Biles and Jones are the clear all-round selection, expecting to contribute 3.409 and 0.466 medals in total respectively, and having 99.9% and 23.9% to medal in the individual all-around finals. Onto the three specialists, Carey is an outstanding performer in Vault and Floor Exercise, having 19.0% and 19.9% to medal in those apparatus. Lincoln has 43.9% chance medaling in Floor Exercise. Miller is our last selection, and she brings value in the Uneven Bars, having a 6.1% percent of medaling there.

Juda and Richard were the all-round selections on the men’s side. They have the highest chances of medaling in our individual all-round, with 5.1% and 4.9% respectively. The first specialist we choose was Khoi Young for his ability in vault and pommel horse, having 44.9% and 11.5% of medaling respectively. Then we choose veteran Whittenburg for his consistency on the still rings (SR), having the highest percentage chance to help the US medal in that apparatus with 26.4%. Lastly, we went with Curran Phillips for parallel bars as he had 26.3% to medal there.

Information on the percentages chances of medaling for each athlete can be in found in Appendix B.

Discussion

Goals, Eye-test, Outside Research

Our goal was to recommend a team of 5 women and 5 men for the Olympic Games in Paris 2024, with an expected medal count for each team. We accomplished our goal of selecting athletes by first selecting 2 all-around athletes, and then finding 3 specialists to complement them, while adjusting for ‘clutch’ factor and recency by duplicating observations based on final vs qualifying rounds, and recency of competition. We then account for individual variance by creating a T-distribution for each competitor for each apparatus, and sampling from that distribution 1000 times. We also then find the predicted medal thresholds using repeated sampling from a T-distribution, taking the top 24 individual scorers, and compare our sampled individual results with the predicted thresholds to find predicted individual medal counts.

The ‘eye’ test confirms some of the choices made by our model. The choices of Simone Biles, Skye Blakely, Shilese Jones align with the choices of the US team in the most recent World Championships where the US women won a gold medal. Both Jones and Biles also medaled in the individual all-around event justifying them being chosen as the two guaranteed all-around competitors. Furthermore, an independent level 10 gymnast committed to a D1 university chose a team identical to that of the US women’s world championship squad leading us to believe the eye test gives ground for our choices. On the men’s side the model’s choices do align with the US men’s team that won bronze in the most recent World Championship. Paul Juda, Fred Richard, and Khoi Young are both on our suggested Olympic team and the 2023 World Championship team. Hong and Richard were the selected all-around competitors also justifying our model’s choices.

Still, there is important information that the USA Olympic committee as well as some of the general public have that our model does not. For example, Konnor McClain may look like a strong choice, but she left Elite (which is a program designed to be a pathway to the USA national team) in 2023 in favor of college. Although she plans to continue training for the Olympics, her non-Elite status likely keeps her from being in the USA olympic committee’s player pool. A similar situation exists for Sunisa Lee, the 2020 Olympic all-around champion. Although her weighted scores are quite high and the model would select her as one of the all-around locks, she has not competed Elite since her enrollment in college in 2022 and a recent injury in which she gained over 50 pounds likely keeps her from making the USA national team. Injury concerns will also keep Brody Malone off the team on the men’s side. Although he eyes a spring 2024 comeback from a knee injury suffered in the spring of 2023, he will likely not be ready or in form for the Olympics.

A Qualitative Analysis of Team USA

A qualitative analysis of Team USA’s women’s gymnastics team has determined that Team USA will most likely pick from a pool of 7 women to choose their team. Only 7 women placed in the top three at any event in the most recent USA national gymnastics championships. These 7 women are Simone Biles, Shilese Jones, Leanne Wong, Skye Blakely, Kaliya Lincoln, Joscelyn Roberson, and Jade Carey. The two all-around staples are likely to be Simone Biles and Shilese Jones as both placed first and second in the USA championships as well as were the top 2 USA finishers in the World Championships. Leanne Wong has a case for being an all-around competitor, but in both championships she finished behind Biles and Jones.

As for the specialists there is support for Leanne Wong being a third all-around athlete as that was the way the USA utilized her in the World Championships. However, she did not qualify for any individual final so her utilization as such an athlete in the Olympics may not be the most optimal strategy. If we choose three specialists, Skye Blakely is a lock to be a specialist. Not only did she place second in the vault, uneven bars, and balance beam, but she is far and away the USA’s third best uneven bars athlete along with Biles and Jones. Beyond these three athletes, the biggest hole lies in the Vault and Floor Exercise events. Of the three athletes remaining only Kaliya Lincoln fills the gap in the floor exercise as she not only silver medaled at the national championships but also won gold at the Pan-American games. Kayla DiCello is another option at Vault for the USA. While not as strong as Lincoln, if the USA is looking for another all-around gymnast for added depth, Dicello provides just that, not to mention her second place finish in floor at the Pan-American games. Still, Kaliya Lincoln’s floor abilities would make her a good fourth addition. That leaves Joscelyn Roberson and Jade Carey who are the two best options for the Vault. Given her recent form, Joscelyn Roberson has a slight edge over Jade Carey. Roberson won gold at the National

Championships and had a good chance to medal at the World Championships before tweaking her left leg before the finals (which also prevented her from competing in the team all-around final).

All in all, a qualitative analysis would suggest a five-woman team of Simone Biles, Shilese Jones, Skye Blakely, Kaliya Lincoln, and Joscelyn Roberson. Alternative lineups would feature either Kayla Dicello over Kaliya Lincoln or Jade Carey over Joscelyn Roberson.

A qualitative analysis for the Men's Olympic team is much less clear. Frederick Richard would be a lock for the team. His bronze medal at the World Championships in the individual all-around while also topping the USA in all-around score in qualifying cements his spot as our top all-around athlete. The second spot would go to either Asher Hong or Khoi Young. Because Asher Hong won the National Championship and was chosen above Young in the World Championships the edge goes to him.

Hong and Richard are both skillful at 4 of the 6 events, but neither covers the base of horizontal bar, so selecting an athlete for that event is important. The best option for this event is Paul Juda who made it to the horizontal bar finals round at the World Championships. Juda has also shown high levels of success in the floor exercise (won the National Championship) and the vault (made the World Championship final) making him an ideal third choice for the Olympic team. Khoi Young is the next best option to make the team. Not only did he place second in the individual all-around at the National Championships exemplifying his strength at all events, but he also got a silver medal in both the vault and pommel horse at the World Championships. Given the results of the National Championships two men are in consideration for the final spot: Donnell Whittenberg and Yul Moldauer. Both placed first in an event at the National Championships and third in another event. The slight edge, however, goes to Whittenberg who placed first in the rings, the only event in which the USA did not send a single athlete to the finals of the World championships, at the National Championships. While Moldauer has Olympic experience, Whittenberg's expertise in the rings gives him a slight edge.

That gives a final five-man team of Frederick Richard, Asher Hong, Paul Juda, Khoi Young, and Donnell Whittenberg.

Limitations

There were also several limitations in our statistical methods. Firstly, we arbitrarily assigned weights to each of the decided 'cutoffs' it is more than possible that our weightings could result in the selection of a different team than if no weights were used. However, the weights used were for good reason given how recent performance affects future performance and our desire to pick more 'clutch' individuals. Another potential limitation in our method is when determining projected scores for athletes that have not competed in some time. Take Sunisa Lee as an example. Due to competing in college and injuries, Lee has very few scores since her Olympic performance in 2021. Because those performances account for so many of her competitions, her lower recent scores are severely overshadowed by her high, older scores. Even with weights her older scores account for a much greater proportion of her data. Thus in our methodology her projected score is unreasonably high for what recent scores suggest. This could remain true for many other athletes where their projected score does not resemble what their recent form might be. The final issue with our methodology and sampling in general is because there is technically no maximum score a gymnast could receive due to more challenging tricks being completed in every successive competition the sample could possibly get too high scores as there is no maximum cutoff in the distributions.

Some alternatives for methodology could include using a Kernel Density Estimation method to estimate the distribution, and then sampling, or obtaining all possible combinations of 5 athletes and then running an olympics simulation. We could also introduce more factors to weighting such as the age of the competitor, skill level of the competition, etc. For future work, and potentially in the final draft, we could also include a sensitivity analysis on our weights, since they were selected somewhat arbitrarily. We plan in the future to find the predicted total medal counts, which include individual all around, individual specialist, and team medals. We did not include the team medals in our predicted medal counts, but also because team medals are calculated through individual performances, optimizing for the best individual performances would necessarily optimize for the best team medal percentage.

The future of gymnastics: gymnast player grades, flexible optimization criterion, and rising stars

As the Olympics has become synonymous with national pride, success in Olympic gymnastics is vital to maintaining a sense of American pride. Given the high expectations laid out for the USA due to past success in the Olympics

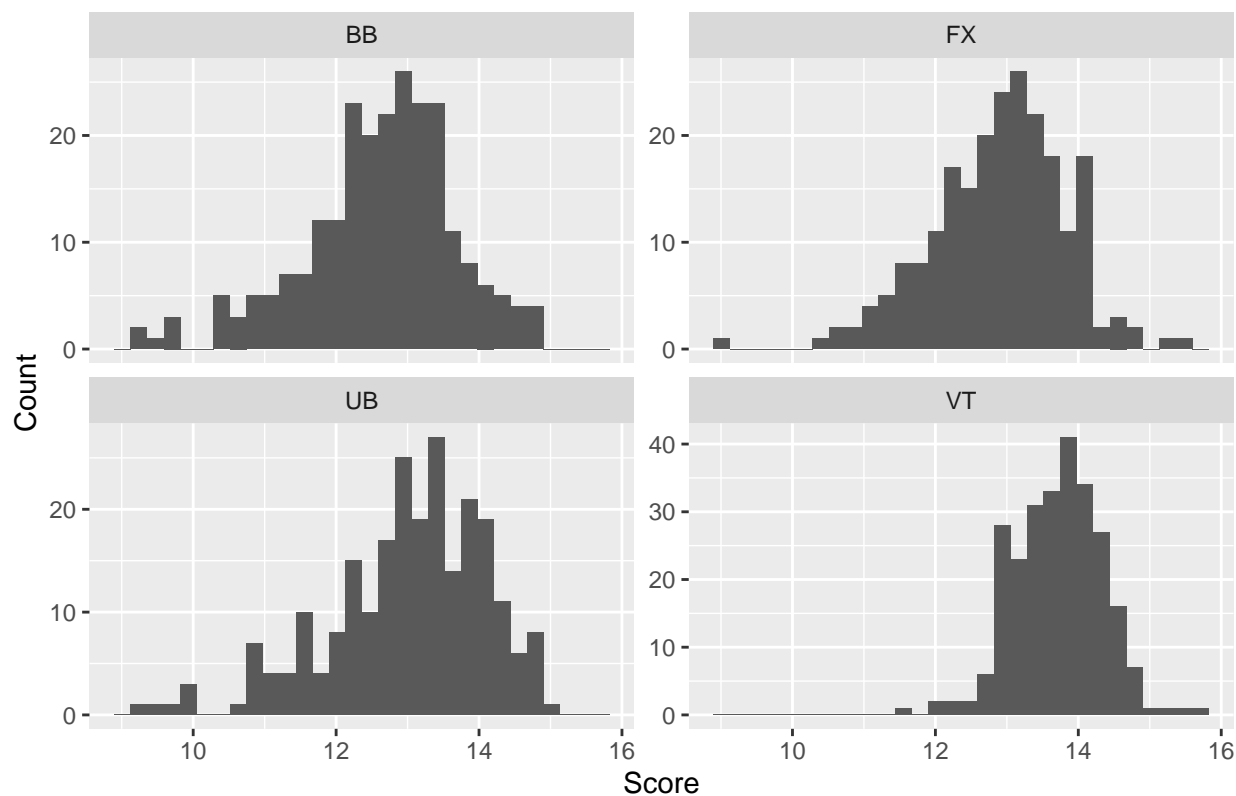
and as other countries have become more competitive on the gymnastics scene, it has become evermore important to select the team that gives the USA the best chance at being successful in Paris 2024.

The statistical analysis done in this paper gives one-of-many statistical ways to determine that “best” team to optimize for total medal count. Our nuanced two-tier approach provides the USA Olympic committee with tangible percentages and reasonable estimates for a candidate’s estimated medal contribution, and a percentage chance to medal in a specific apparatus. Using our methodology, others might be able to further optimize for things such as team medal or overall total scores given the chance to run more simulations or adding additional data. Another use case would be replicating this for many countries, and compiling an elite database with “player grades” of sorts attached to Olympic performance. As more advanced data is measured and the intersection between sports and data analytics continue to grow, our methods and data can create a competitive advantage over the rest of the field.

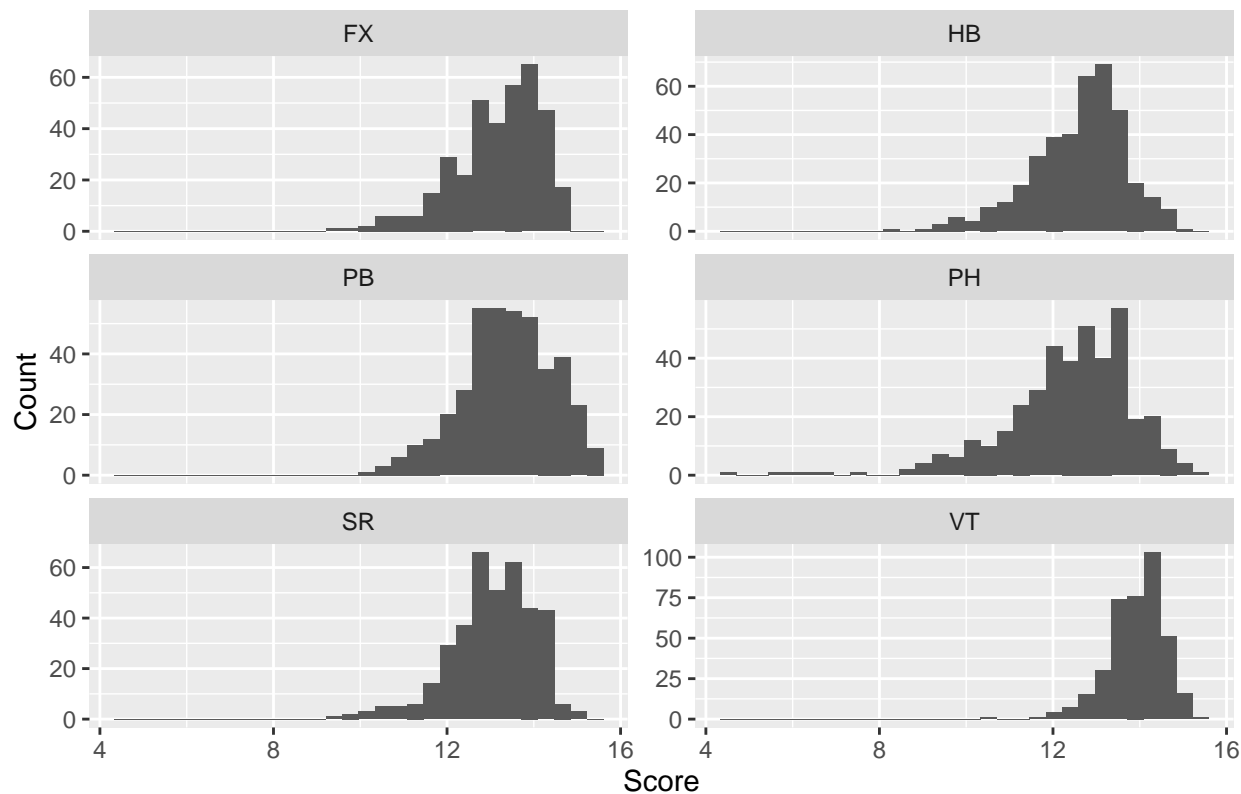
The potential for future research is endless. Our next step would be looking into optimizing lineups based on different criteria. That is, if a person wants to put a higher weighting on having the most consistent gymnasts or on the gymnast with highest potential than the model would be able to put out a lineup based on that criteria. Building out something more interactive for a final product is in the works. We will also look into more data related to the Olympics and gymnastics, and maybe interview a couple of professional gymnasts or their coaches to get further insight into Olympics level gymnastics. Finally, in the future we hope to build off similar tools to the ones created during this data analysis to predict or scout rising stars competing on the USA Elite pathway to keep an eye on for future Olympic cycles.

Appendix A

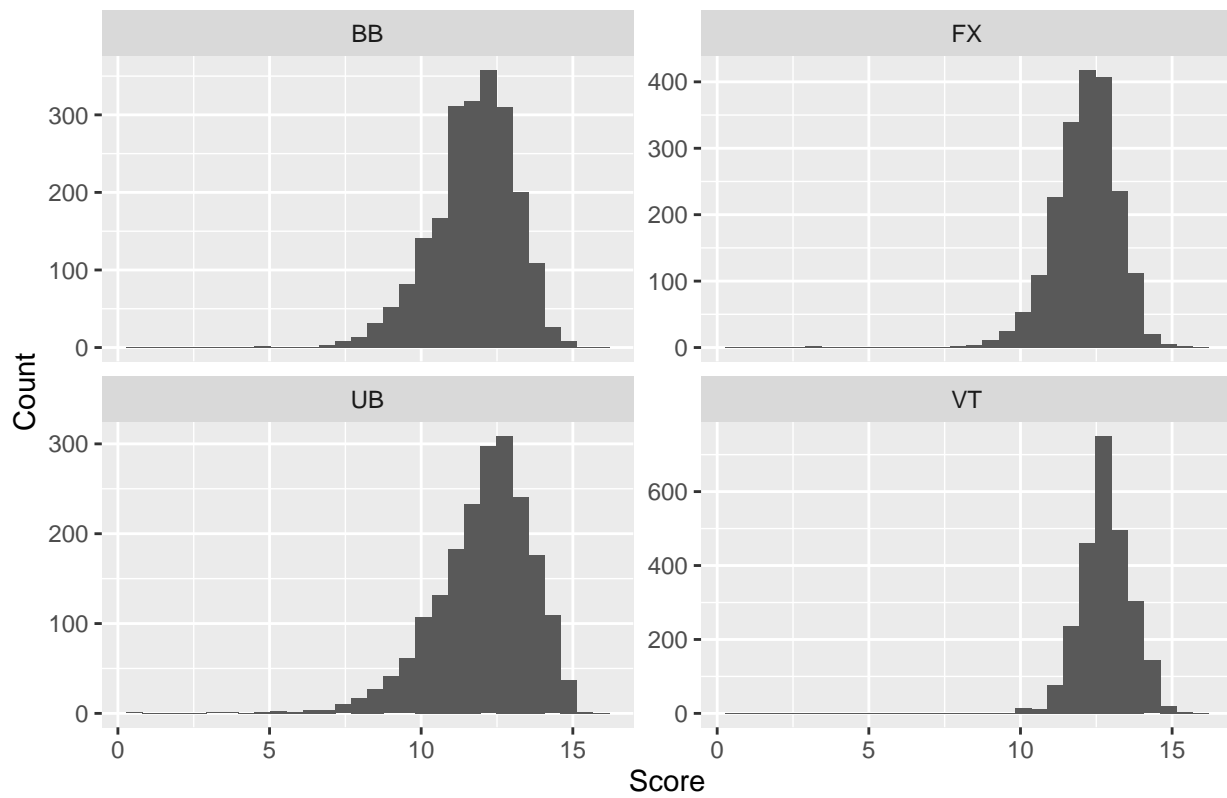
Distribution of Scores for USA Women for Different Apparatuses



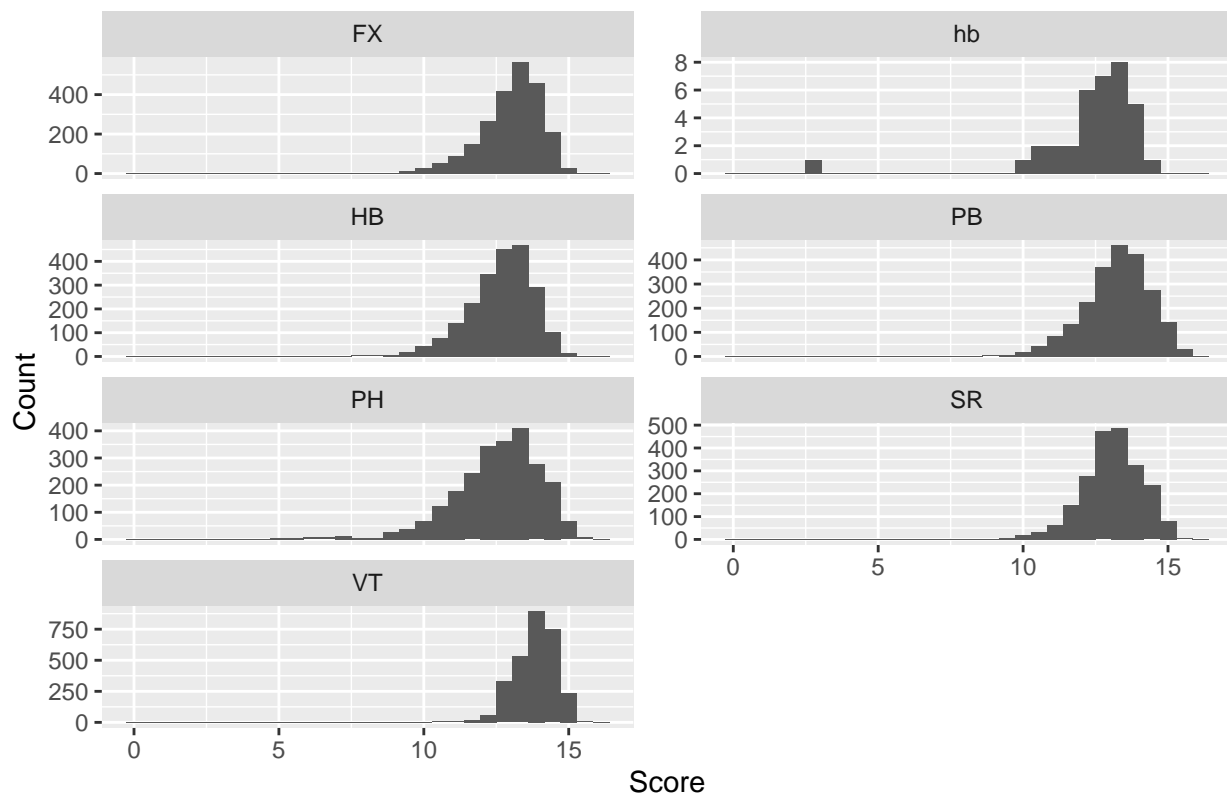
Distribution of Scores for USA Men for Different Apparatuses



Distribution of Scores for All Women for Different Apparatuses



Distribution of Scores for All Men for Different Apparatuses



Appendix B

Name	Total_AA	Total_BB	Total_FX	Total_VT	Total_UB
Simone Biles	0.999	0.715	0.993	0.702	0.000
Shilese Jones	0.239	0.035	0.182	0.005	0.192
Konnor McClain	0.154	0.485	0.082	0.004	0.000
Skye Blakely	0.028	0.043	0.000	0.000	0.003
Jordan Chiles	0.019	0.000	0.203	0.004	0.000
Leanne Wong	0.010	0.000	0.028	0.005	0.000
Jade Carey	0.008	0.000	0.199	0.190	0.000
Kayla Dicello	0.002	0.001	0.028	0.001	0.000
Addison Fatta	0.000	0.000	0.000	0.000	0.000
Alicia Zhou	0.000	0.000	0.000	0.000	0.000
Amelia Disidore	0.000	0.000	0.000	0.000	0.000
Annalisa Milton	0.000	0.000	0.000	0.000	0.000
Ashlee Sullivan	0.000	0.001	0.000	0.000	0.000
Brooke Pierson	0.000	0.000	0.000	0.000	0.000
Charlotte Booth	0.000	0.000	0.000	0.000	0.000
Chloe Cho	0.000	0.000	0.000	0.000	0.000
Ciena Alipio	0.000	0.003	0.000	0.000	0.000
Dulcy Caylor	0.000	0.000	0.000	0.000	0.000
Elle Mueller	0.000	0.000	0.000	0.000	0.000
Eveylynn Lowe	0.000	0.000	0.000	0.000	0.000
Joscelyn Roberson	0.000	0.016	0.107	0.029	0.000
Kaliya Lincoln	0.000	0.000	0.439	0.000	0.000
Katelyn Jong	0.000	0.000	0.000	0.000	0.000
Katelyn Rosen	0.000	0.000	0.000	0.000	0.000
Kelise Woolford	0.000	0.000	0.000	0.000	0.000
Lauren Little	0.000	0.000	0.000	0.000	0.000
Levi Jung-Ruivivar	0.000	0.000	0.000	0.000	0.000
Lexi Zeiss	0.000	0.000	0.000	0.000	0.000
Madray Johnson	0.000	0.000	0.000	0.000	0.000
Malea Milton	0.000	0.000	0.000	0.000	0.000
Marissa Neal	0.000	0.000	0.000	0.000	0.000
Michelle Pineda	0.000	0.000	0.000	0.000	0.000
Myli Lew	0.000	0.000	0.000	0.000	0.000
Nola Matthews	0.000	0.000	0.000	0.000	0.000
Norah Christian	0.000	0.000	0.000	0.000	0.000
Sunisa Lee	0.000	0.208	0.000	0.000	0.000
Tiana Sumanasekera	0.000	0.012	0.000	0.000	0.000
Zoe Miller	0.000	0.000	0.000	0.000	0.061

Name	Total_AA	Total_SR	Total_FX	Total_PB	Total_PH	Total_VT	Total_HB
Brody Malone	0.272	0.000	0.168	0.042	0.010	0.045	0.435
Frederick Richard	0.051	0.000	0.235	0.025	0.000	0.049	0.180
Paul Juda	0.049	0.000	0.393	0.000	0.000	0.221	0.044
Khoi Young	0.030	0.000	0.229	0.000	0.115	0.449	0.000
Yul Moldauer	0.028	0.000	0.211	0.163	0.000	0.075	0.000
Asher Hong	0.024	0.034	0.123	0.026	0.000	0.296	0.000
Colt Walker	0.014	0.000	0.038	0.208	0.000	0.090	0.000
Shane Wiskus	0.009	0.000	0.150	0.068	0.000	0.015	0.015
Donnell Whittenburg	0.004	0.264	0.186	0.000	0.000	0.194	0.000
Cameron Bock	0.002	0.000	0.000	0.044	0.000	0.000	0.000
Fred Richard	0.002	0.000	0.073	0.000	0.000	0.000	0.123
Alex Karadzhov	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Alex Tapanes	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Anthony Koppie	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Asher Cohen	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Blake Sun	0.000	0.000	0.000	0.149	0.000	0.000	0.000
Brandon Briones	0.000	0.000	0.000	0.000	0.000	0.044	0.000
Brandon Nguyen	0.000	0.000	0.000	0.000	0.000	0.011	0.000
Caden Clinton	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Caleb Melton	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Chase Davenport-Mills	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cole Partridge	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Colin Flores	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Crew Bold	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Curran Phillips	0.000	0.000	0.000	0.263	0.000	0.204	0.047
Dallas Hale	0.000	0.000	0.015	0.008	0.000	0.354	0.000
Daniel Simmons	0.000	0.000	0.000	0.000	0.000	0.002	0.000
Drake Andrews	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Evan Hymanson	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Evan Manivong	0.000	0.000	0.000	0.000	0.000	0.021	0.000
Fuzzy Benas	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Garrett Brauntun	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Garrett Schooley	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ian Gunther	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ian Lasic-Ellis	0.000	0.000	0.026	0.000	0.000	0.000	0.000
Ian Sandoval	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Isaiah Drake	0.000	0.000	0.000	0.000	0.000	0.002	0.000
Izaiha Mlay	0.000	0.000	0.000	0.000	0.000	0.000	0.000
J.R. Chou	0.000	0.000	0.000	0.116	0.000	0.000	0.000
Jack Freeman	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Jackson Harrison	0.000	0.000	0.050	0.000	0.000	0.000	0.000
Jeremy Bischoff	0.000	0.000	0.046	0.000	0.000	0.000	0.000
Joshua Karnes	0.000	0.000	0.033	0.000	0.000	0.032	0.000
Kameron Nelson	0.000	0.000	0.001	0.000	0.000	0.037	0.000
Kazuki Hayashi	0.000	0.000	0.000	0.000	0.000	0.043	0.000
Landen Blixt	0.000	0.000	0.000	0.000	0.000	0.059	0.000
Landon Simpson	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Matt Cormier	0.000	0.000	0.086	0.000	0.000	0.000	0.000
Maxim Bereznev	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Michael Artlip	0.000	0.000	0.000	0.000	0.000	0.001	0.000
Michael Jaroh	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mike Fletcher	0.000	0.000	0.000	0.000	0.000	0.011	0.000
Noah Newfeld	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Oliver Zavel	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Riley Loos	0.000	0.000	0.023	0.000	0.000	0.018	0.000
Rithik Puri	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ryan Swatscheno	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Name	Total_AA	Total_SR	Total_FX	Total_PB	Total_PH	Total_VT	Total_HB
Samuel Phillips	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Taylor Burkhart	0.000	0.000	0.077	0.000	0.000	0.139	0.000
Taylor Christopulos	0.000	0.000	0.000	0.000	0.000	0.012	0.000
Toby Liang	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Vahe Petrosyan	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Will Fleck	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Works Cited

Brody Malone eyes spring 2024 return to gymnastics from leg surgeries. 17 May 2023. NBC Sports. <https://www.nbcsports.com/olympics/news/brody-malone-gymnastics-injury-comeback>. Accessed 20 Nov. 2023.

Duffy, P. Konnor McClain enrolling at LSU this fall with hopes of balancing NCAA and elite - Gymnastics Now. *Gymnastics-Now.com*. 13 Jul. 2023. <https://gymnastics-now.com/konnor-mcclain-enrolling-at-lsu-this-fall-with-hopes-of-balancing-ncaa-and-elite/>. Accessed 20 Nov. 2023.

Gunston, Jo. "Judging the Judges – How Statistical Analysis Evaluates Fairness And ..." *Olympics.Com*, 19 Oct. 2023, olympics.com/en/news/how-statistical-analysis-evaluates-fairness-accuracy-gymnastics. Accessed 06 Nov. 2023.

Most popular sports in the Summer Olympics US 2021. (n.d.). Statista. <https://www.statista.com/statistics/1245746/summer-olympics-most-followed-sports-us/>. Accessed 20 Nov. 2023.

Olympic Summer Games: global broadcast audience. (n.d.). Statista. <https://www.statista.com/statistics/280502/total-number-of-tv-viewers-of-olympic-summer-games-worldwide/#:~:text=Between%202012%20and%202020%2C%20the,3.2%>. Accessed 20 Nov. 2023.

Olympic champ Sunisa Lee gained 45 pounds due to kidney issue. "It was so scary." 17 Nov. 2023. *USA TODAY*. <https://www.usatoday.com/story/sports/olympics/2023/11/17/sunisa-lee-olympic-champion-kidney-health-paris/71616483007/>. Accessed 20 Nov. 2023.

Tracy, Jeff. "Poll: Gymnastics Is Americans' Most Highly Anticipated Olympic Sport." *Axios, Morning Consult*, 19 July 2021, www.axios.com/2021/07/19/olympics-favorite-sport-gymnastics. Accessed 04 Dec. 2023.