Journey to Paris 2024: A Bayesian Approach to Finding the Best Men's and Women's U.S. Gymnastics Teams

Chris Liang, Enzo Moraes Mescall, Mitchelle Mojekwu, Zoe Svec

Introduction

The Olympic Games are a highly anticipated world-renowned multi-sporting event that takes place every four years. Particularly the Summer Olympic Games tend to have a wider variety of 32 sports and more viewers than that of the Winter Olympics (Olympics, 2021). Athletes from all over the world can participate granted they meet the criteria established by their nation's Olympic committees and the international sports federations. With female qualifying gymnasts from the United States placing with medals in the team all-around, individual all-around, and each individual apparatus in the 2020 Tokyo Olympics game, there has been a surge in media attention on the United States gymnastics teams (Olympics, 2020).

As the Paris 2024 Summer Olympic Games is approaching, the United States Olympic Men's and Women's Artistic Gymnastics aims to put together a team of 5 each that best represents the country on the world's sporting stage by optimizing medal success amongst the team all-around, individual all-around, and individual apparatus events. At the Paris Olympics, there are specific rules about the number of athletes and countries allowed to compete in the events, the low number of athletes that qualify for the finals suggests there must be thoughtful crafting of the team of 5 (UCSAS, 2023). This study aims to use the most recent Olympic Games and other world competitions' qualifying and final round results data to best assemble a team that is likely to produce optimal success in terms of medals within the Olympic qualifiers and final criteria (UCSAS, 2023).

The UConn Sports Analytics Symposium provisioned two clean data sets of the accumulation of results of teams worldwide that participated in the major domestic and international gymnastic qualifying and final competition events leading up to the 2024 Summer Olympic Games. The first data set includes includes the results of the 2020 (taking place in 2021) Tokyo Summer Olympics qualifying and final rounds, and the second data set includes competitions in the 2022 and 2023 seasons. Observations for both data sets are at the athlete- and apparatus-level score for an event in a round at a gymnastics competition–for example, Simone Biles's final uneven bars score at the 2023 US Gymnastics Championships. It is worth noting, however, that the data from the Tokyo Olympics only include results for women's gymnastics, while the data from 2022-2023 include results for both men's and women's gymnastics. The data are collected from the results on each corresponding competition's official website, which are results from the officially judged scores of each competition. Variables in the data sets include first and last names of each athlete, gender, country, date of competition, name of competition, the round of the competition (e.g. qualifier or final of an individual apparatus, individual all-around, or team event), the location of the competition, apparatus (women compete in "BB": balance beam, "FX": floor exercise, "UB": uneven bars, and "VT": vault; men compete in "FX": floor exercise, HB": high bar, "PB": parallel bars, "PH": pommel horse, "SR": still rings, "VT": vault; beyond the floor and vault overlap, both men and women may compete in vaults "VT1" and "VT2", which are 2 different vaults required in individual apparatus qualifications and finals), the execution score, difficulty score, penalty, and final score for that athlete on that apparatus, and the rank of that athlete in that apparatus and round.

We decided to not proceed in using the data set of results from the Tokyo Summer Olympics since the data consisted only of female athletes and one competition (the Olympic Games). Additionally, in the context of Olympic gymnastics, athletes of age 16 and older are eligible to compete but gymnastics is a sport in which most athletes retire in their early to mid-twenties. Specifically in the summer 2020 Tokyo Olympics only three female athletes aged 27 or older qualified to compete (Camenker, 2021). Furthermore, the average age for female gymnasts in the 2020 Olympics was approximately 22 years of age, meaning we assume that many of the competitors in the older data set will not be competing in the 2024 Paris Summer Olympics (Meyers, 2021).

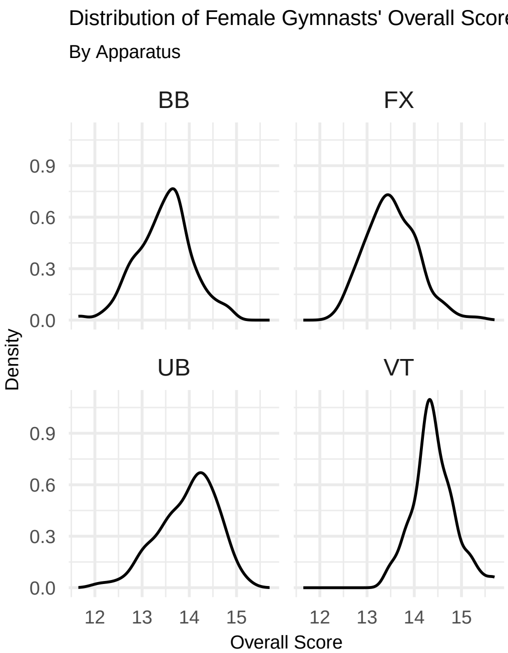
We have the following objectives for this study: (UCSAS, 2023)

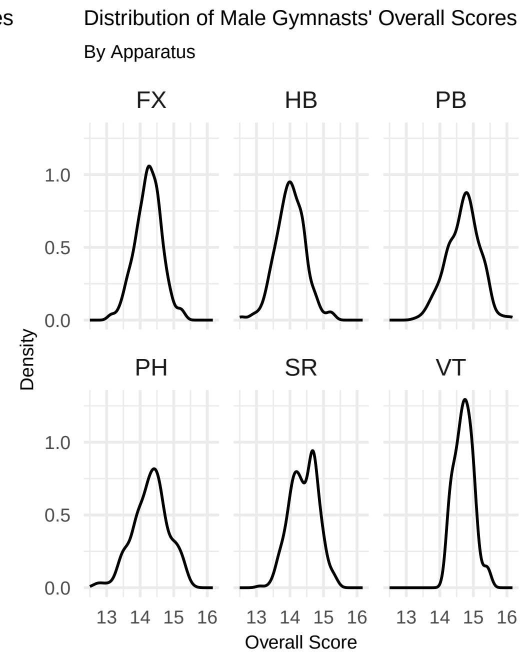
1) Decide on whether to maximize total medal count, gold medal count, or a weighted medal count
2) Decide on whether to value the medals of an event (team, individual all-around, individual apparatus) over others.
3) Decide on whether Team USA should maximize its total medal count by selecting a team of five gymnasts who are all-around gymnasts, event specialists (gymnasts who focus on 1 or more apparatus but not all apparatus), or a combination of those.
4) Identify the group of five athletes who will most likely enable the Team USA Olympic Men's and Women's Artistic Gymnastics team to maximize medals won in the Paris 2024 Summer Olympics.

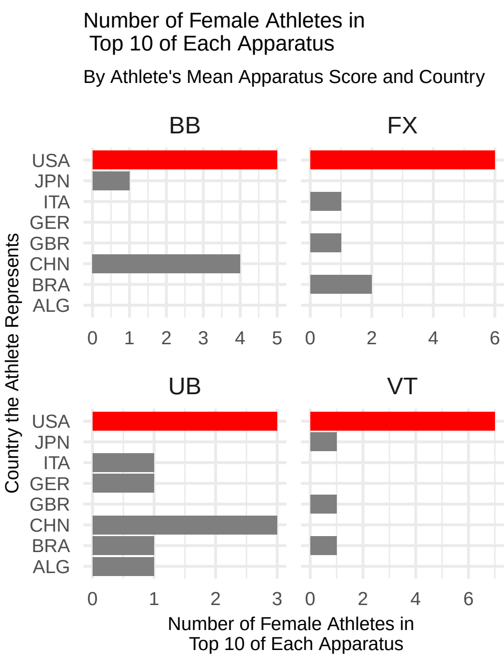
Addressing these objectives will assist the national Olympic Artistic Gymnastics teams in best approaching the Olympic gymnastics events in totality by offering recommended strategies to best approach team selection. In our analysis of the best fit US male and female gymnastics teams for the Paris Olympics, we will undertake a Bayesian approach to simulate outcomes of individual athletes' scores in an apparatus. Bayesian frameworks in sports analytics to simulate athlete's results are well-documented and have seen a rise in popularity in the past decade (Santos-Fernandez, et. al., 2019). For instance, Yang and Swartz use Monte Carlo Markov Chains to simulate the outcomes of baseball games (Yang et. al., 2022). We will build upon these analyses and choose the appropriate Bayesian method to simulate outcomes of gymnast results in each apparatus, after which we will analyze the top performers in each apparatus, assign medals, and find the best combination of athletes.

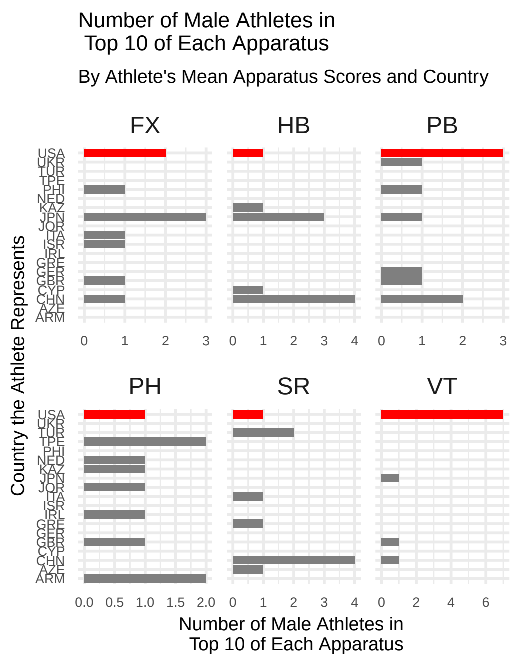
Exploratory Data Visualizations

We created some exploratory data visualizations to examine the distributions of athlete scores and examine the performance of US athletes relative to other countries. We find that from the density plots of male and female athletes' overall scores (difficulty score + execution score - penalty = overall score) per apparatus that the scores are approximately normally distributed for the apparatuses for both genders. There are some slight deviations from normality, namely women's vault scores and uneven bars scores are slightly skewed left and men's still rings scores have a slight dip in the peak. Nonetheless, the approximate normality of the distribution of athlete's scores by apparatus informs our bayesian approach, as we may use normal priors for our data. Furthermore, we plotted the number of athletes per country in the top 10 of each apparatus internationally, top 10 meaning the athletes with top 10 highest mean scores for each apparatus. For example, we see that for women's vault, 6 of the top 10 athletes in the world are US gymnasts. These plots help inform us of if we should be thinking about specialists or generalists in the US team combinations. We see that the top 10 for each apparatus have a high concentration of US female gymnasts, so we may want specialists in our team makeup, whereas that case does not transfer to the US male gymnasts, as there are few US male gymnasts in the top 10 for the floor exercise, high bar, pommel horse, and still rings apparatuses. In the men's case, we may not want to send specialists to take up a spot on team of five, and we shall explore this conjecture in our simulations.









Methodology

Prior to conducting simulations, we cleaned our data set on the 2022-2023 gymnastics competition results. There were several cases of missing or inconsistent athlete first and last names, so we created unique athlete IDs using string methods by using the first three letters of an athlete's first name, the first three letters of an athlete's last name, and the country code. We made sure to add in missing names and account for names with less than three characters. Furthermore, the apparatus code for high bar was inconsistent across the Commonwealth Games and all other competitions, so we made sure to consolidate high bar into one apparatus code. Because individual apparatus qualifying vaults needed athletes to compete in two different vaults (VT1, VT2) as opposed to one vault in the finals or team or all-around events, we decided to take the higher of the two vault scores for an athlete for a competition, if there were two vaults completed, and consolidated that score as one vault apparatus code. We decided to keep the higher score given the vaults were different, and athletes likely compete with the vault that gives them the higher score during event finals.

We then proceeded to cut down on the number of observations in our data set for two purposes: 1) so that our simulations later could run more quickly, and 2) we felt that it was unnecessary to simulate scores for athletes that had little to no chance of medaling given their previous records, given the scarcity of athletes that will attend the Olympic finals. To filter the observations, we first removed any individual athletes entirely who had never made the finals in any event in any competition in the data set. Afterward, we created quantiles of 20% increments and 10% increments for each round in a competition for each apparatus, separated by gender because men and women compete separately. We checked the number of unique athletes that competed at each competition in each round (see Appendix), and found that for all rounds in competitions other than the Oceania Championships, at least 36 unique athletes participated. For some rounds, hundreds of athletes participated–so we decided to filter for: if more than 100 athletes competed in a round in an apparatus, then we filtered for athletes' scores in the top 10%; if less than 100 athletes competed in a round in an apparatus, then we filtered for athletes' scores in the top 20%; for the Oceania Championships, we filtered for the top 40% (four athletes). The reason we adopted a quantile-based filtering approach is because of the variation in number of athletes who competed at different competitions, so simply taking the top 20 athletes, for example, of each competition may not account for that variation. Our last method of filtering was to remove observations of athletes' scores for apparatuses if an athlete had not competed more than twice in that apparatus in the entire 2022-2023 data set. Our rationale was that there were 37 distinct competitions in the data set, so if an athlete has not competed more than twice in the past two years in an apparatus, they are likely not that active in that apparatus. Additionally, filtering for athletes' scores when an athleted has competed in that apparatus more than twice allows us to find a variance and mean for that athlete's performance on an apparatus. We are left with 2210 observations in our filtered data set to conduct simulations, with 157 unique male athletes and 88 unique female athletes.

Bayesian Monte Carlo Approach

In this study, we present a Bayesian statistical Monte Carlo approach to select the top male and female American gymnast candidates for participation in the 2024 Olympics. Our method involves the creation of prior distributions based on historical performance data, conditioning these distributions on individual competition results, and simulating medal outcomes by predicting scores for each gymnast in each apparatus event. The approach offers a robust framework for incorporating both prior beliefs and observed data to make informed predictions about athletes' performances in simulated events (Hoff, 2009). Utilizing Bayes' law for probability density functions, where x is a vector of all the data from an apparatus and gender combination, x_i represents the vector of observed data for athlete i, and θ represents the parameters of the distribution we will be using to model the competition. We are under the assumption that all gymnastic scores are independent and identically distributed for every athlete and that every athlete's scores come from the same distribution type. Furthermore, for modeling purposes, we assume a common prior $p(x|\theta)$ for all athletes such that $p(x_i|\theta) = p(x|\theta)$ for all i. This allows us to write the posterior distribution as:

$$p(\theta|x_i) \propto p(x|\theta)p(\theta)$$

To simulate a score for an athlete we sample $\theta^{(s)} \sim p(\theta|x_i)$ from the posterior distribution and then a new value, and then sample $\tilde{x_i} \sim p(x|\theta^{(s)})$. This represents a new predicted data point for athlete i using the posterior distribution, a common practice for estimating values from a posterior predictive distribution in Bayesian Monte Carlo Simulations (Hoff, 2009). Thus, given we can simulate an athlete's scores, we can then simulate a competition between all candidate athletes and allocate gold, silver, and bronze medals to the top three athletes.

Prior Distribution Creation and Conditioning on Individual Results

We began by creating prior distributions for each apparatus' total score. Splitting the data set by apparatus and gender resulted in multiple separate smaller data sets, see Exploratory Data Visualizations for the apparatus-gender level distribution. Observing the combined score distributions, they appear to be unimodal and slightly right-skewed. Various distributions were considered, like the beta distribution which is conveniently upper and lower-bounded, but the normal distribution was chosen for its simplicity, ease of use, and effective fit.

For ease of use, we depended on conjugacy to derive the parameters of the normal distribution. Since both the mean and variance of the normal distribution are unknown, we used normal-inverse gamma priors for both parameters. At this point, the issue arises that a plurality of athletes has only ever done a single competition in any given apparatus, resulting in a large number of athletes with 0 variances and skewing the distribution. Due to the unlikely nature of athletes who do not compete regularly participating in the Olympics, we have decided to truncate the apparatus level data set to exclude athletes with less than three competition appearances for that given apparatus.

To estimate prior parameters, we fit a normal distribution to the distribution of athletes' means from the data using the maximum likelihood method and the `fitdist()` function (Muller, 2023). The maximum likelihood estimates for the mean and variance were then used as the prior parameters for the normal part of the normal-inverse gamma distribution. Similarly, for the inverse-gamma parameters, we fit an inverse-gamma distribution to the distribution of athlete's variances from the data using a similar method and used these parameters as the prior parameters for the inverse-gamma part of the normal-inverse gamma distribution. This process was done independently for all apparatus-gender combinations to produce a different set of prior parameters for each independently.

Following the establishment of the prior distributions, we updated these distributions based on individual competition results. We rely on existing literature for the formulas for the posterior parameters (Hoff, 2009). We then employ a Monte Carlo method to sample new data, simulating the posterior parameters for mean and variance and then using these to simulate individual gymnastic scores.

Simulation of Gymnastics Events

To simulate gymnastics events, we performed 500 iterations for each apparatus event. For every iteration, we simulated a score for each athlete in the event by sampling $\tilde{x_i}$ from the posterior distribution. Furthermore, since the normal distribution is unbounded, we truncated the distribution at 0 and 20 to reflect the scoring system. We then ranked the athletes by their simulated scores and awarded gold, silver, and bronze medals to the top three athletes. Notably, we chose not to go with a qualification structure and had a simple one-shot round for victory. This decision was made due to computational constraints and introduced more variance in the medal distribution, which we take into account when identifying which athletes to pick for Team USA. We repeated the simulation process for each apparatus event, resulting in about 11 million simulated scores across all competitions.

Assumptions

Inherent in our methodology are several assumptions. Firstly, we assume that gymnastic scores are normally distributed, justifying the use of the normal distribution for both prior and posterior distributions. Additionally, we assume independence between events, allowing us to treat each apparatus event as a separate and identically distributed random variable. We also assumed that athletes prioritize all stages of every event identically. Furthermore, we assume that historical performance data adequately represents the gymnasts' true abilities and that changing age is not a factor in gymnastic ability. While this assumption simplifies the modeling process, it may not fully capture the complexities of individual development and improvements over time.

Results

We ran simulations for each apparatus for each gender (women's 4 apparatuses and men's 6 apparatuses) 500 times, then outputted the top athletes by medal count per apparatus.

Female Athletes' Results

For women's apparatuses, we outputted the tables of simulation outcomes for floor exercise, balance beam, and vault because of the high presence of medals for US gymnasts. The table of simulation outcomes for other apparatuses are in the appendix.

Table 1: Women's Floor Exercise Simulation Results

Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Simone Biles: USA	186	85	57	328
Rebeca Andrade: BRA	38	45	38	121
Kaliya Lincoln: USA	30	37	29	96
Jessica Gadirova: GBR	14	27	43	84
Flavia Saraiva: BRA	27	25	18	70
Jade Carey: USA	22	19	23	64
Martina Maggio: ITA	11	18	29	58
Jordan Chiles: USA	12	20	19	51
Joscelyn Roberson: USA	10	17	20	47
Sabrina Maneca Voinea: ROU	11	15	14	40

For the women's gymnastics team, we select Simone Biles, Zoe Miller, Shilese Jones, Konnor McClain, and Jade Carey to represent the US at the Paris Olympics. Our rationale for this combination and our framework for optimizing gold medal count, total medal count, and apparatus vs. individual all-around vs. team event wins are explored further in the discussion.

Based on our 500 simulations, Simone Biles has the highest count of total medals in balance beam (170 total medals), floor exercise (328 total medals), and vault (328 total medals), as well as highest gold medal count in floor exercise and vault (186 and 151 gold medals expected of 500 simulations, respectively). Therefore, Biles is expected to be a strong US contender in the individual apparatus events, individual all-around events, as well as a the team all-around since only three members compete in each apparatus for that event. Simone Biles places ninth in simulation outcomes for uneven bars, and we note that US athletes Zoe Miller and Shilese Jones come in third (35 gold medals) and fifth place (32 gold medals), respectively, for number of gold medals won in the uneven bars simulation outcomes (see appendix.) Additionally, Konnor McClain comes second highest for gold medal count for balance beam (65 gold medals) in simulation outcomes, after Yaqin Zhou of China, and Jade Carey comes in third for highest count of total medals (199) and gold medals (61 medals) for vault. Among the simulation results, we generally see the US women gymnasts place very highly in individual apparatuses by both total medal count and gold medal count.

Male Athletes' Results

For men's apparatuses, we outputted the tables of simulation outcomes for vault and parallel bars because of the high presence of medals for US gymnasts relative to the other apparatuses. The tables of simulation outcomes for floor exercise, still rings, pommel horse, and high bar are in the appendix.

Table 2: Men's Parallel Bars Simulation Results

Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Jingyuan Zou: CHN	133	75	49	257
Lukas Dauser: GER	43	40	36	119
Boheng Zhang: CHN	25	26	27	78
Curran Phillips: USA	24	20	24	68
Colt Walker: USA	20	28	19	67
Kaito Sugimoto: JPN	23	25	17	65
Joe Fraser: GBR	24	17	19	60
Carlos Yulo: PHI	16	20	22	58
Cong Shi: CHN	15	17	12	44
Blake Sun: USA	15	10	19	44

For the men's gymnastics team, we select Asher Hong, Curran Phillips, Donnell Whittenburg, Colt Walker, and Brody Malone to represent the US at the Paris Olympics. Our rationale for choosing this combination and optimizing for all-around gymnasts as opposed to event specialists are explored further in the discussion.

We see from our simulation outcomes tables that the US men gymnasts do not come in the top 3 simulated rankings by total medal count for any apparatuses other than Asher Hong in vault, who comes in a simulated second place with 183 total vault medals and 67 total vault gold medals. As a result, it may not be fruitful to pick individual US male athletes by their apparatus-specific performance given athletes from other countries may place ahead of them on individual apparatuses–instead, it's worth noting the overlap of US men gymnasts who place in the top 10 of simulated total medals across the apparatuses. Curran Phillips places in sixth in simulated total medals for vault (112 medals) and fourth in simulated total medals for parallel bars (68 medals). Donnell Whittenburg places seventh in simulated total medals for still rings (56 medals), eighth for simulated total medals in floor exercise (43 medals), and seventh for simulated total medals for vault (111 medals). Colt Walker places ninth in simulated vault total medals (85 medals), fifth in parallel bars simulated total medals (67 medals), ninth in floor exercise simulated total medals (40 medals). Lastly, Brody Malone places fourth in simulated high bar total medals (72 medals) and tenth in simulated floor exercise total medals (39 medals). Because we see that these athletes are still able to place well in multiple events, all the apparatuses except for pommel horse are covered with at least one athlete who performs well in it, but who can also place well in other apparatuses.

Discussion

Objective 1: Choice of Medal Success Metric (Total Number of Gold Medals)

From the dot plot visualizations of the women's simulation of the three considered success metrics (gold medal count, total medal count, and weighted medal count) for each apparatus by USA and non-USA teams, there looks to be at least one USA athlete that places higher than of all non-USA athletes in each medal metric for each apparatus except uneven bars (Appendix: Image 5). The women's USA team makes up 51% of the total women's gold medals in the simulation which is a higher proportion than the 47% of the total medal count and 48% of the weighted medals (Appendix: Image 7). From the dot plot visualizations of the men's simulation of the three considered success metrics, for each apparatus by USA and non-USA teams, there are non-USA athletes for each apparatus that exceed the USA in each medal success metric (Appendix: Image 6). The men's USA team makes up 24% of the total medal count in the simulation which is a higher proportion than the 21% of the total gold medal count and 23% of the weighted medals. (Appendix: Image 8) When viewing the top 5 most successful female athletes (top 5 most decorated by that medal metric) in each apparatus for each medal success metric, the USA makes a good portion of these athletes. There tend to be 2-4 USA athletes in the top 5 depending on the success metric and apparatus (Appendix: Image 7). When viewing the top 5 most successful male athletes in each apparatus for each medal success metric, there tend to be 0-3 (mostly 0) US male athletes present (Appendix: Image 8).

Considering that female USA medalists tend to represent a much higher proportion of medal successes (no matter the success metric) than male USA athletes, it is best to prioritize the success metric that the female team performs the best in. Also viewing the male top 5 most decorated athlete by each metric for each apparatus, the men's USA team has a higher proportion of athletes in the top 5 when using the total number of gold medals as a success metric (Appendix: Image 8). Therefore, the success metric that we aim to maximize to best ensure the USA team's success is the total number of gold medals.

Objective 2: Value of Medals for Each Event Type (Team AA > Individual AA > Individual Apparatus)

From the table of the top 10 most decorated gold medal female athletes by apparatus from the simulated data, the USA, China, Brazil, and Great Britain make multiple appearances. The USA has athletes in the top 10 most decorated gold medalists for each apparatus as well as the top 5, but other countries do not (Appendix: Image 9). This allows us to assume that the USA has great potential in winning the team all-around since it is the only country with many of the most successful athletes in each apparatus in terms of the number of gold medals. In this case, valuing the team's all-around medal more than the individual all-around and individual apparatus will hopefully increase medal success in terms of gold medal count. Also when viewing the top 10 most decorated gold medal female athletes by apparatus, the USA's Simone Biles, appears in the balance beam as first, in floor exercise as first, in uneven bars as ninth, and in vault as first. Valuing the individual all-around events higher also may help team USA increase in our metric of success. Furthermore, since these events are harder to achieve than individual apparatuses because of the multiple sections within the event that need to also meet a standard, it will be harder for other countries to also benefit from this increased value.

From the table of the top 10 most decorated gold medal male athletes by apparatus, the USA, Japan, and China make multiple appearances. The only country that has an athlete in each apparatus for the top 10, is the USA (Appendix: Image 11). It could be slightly beneficial to the men's team to value the team's all-around success more than the other events. The US men's team also does not have a well-rounded athlete that places in the top 10 most decorated gold male athletes for each apparatus so we can assume valuing individual all-around successes over the other events would not help the US men's team but it also would not hurt it since other countries also do not have a highly decorated well-rounded competitor.

In the dot plots of the top 5 decorated gold medal female athletes' countries by number of gold medals for each apparatus, US athletes make multiple appearances (Appendix: Image 10). In the dot plots of the top 5 decorated gold medal male athlete's countries by number of gold medals for each apparatus, US athletes are present in multiple apparatuses but not many athletes are well decorated within each apparatus. But in vault there are two US athletes in the top 5 (Appendix: Image 12). Valuing individual apparatus events as regular events of weight 1 would best suit both the male and female teams' success against their competitors. Weighing the team all-around as 3 points is viable because not only do both the men's and women's USA have the potential to win based on this simulation, but there is less reliance and pressure on one singular person. Weighing the individual all-around as 2 will hopefully benefit the women's team with Simone Biles as the potential representative for this event. These weights will in hope best accommodate the male and female athletes and give them the best chance at success against other countries in terms of the total number of gold medals.

Objective 3: All-Around vs Event Specialist vs Mixture (Women: Even Specialist, Men: Mixture)

In our metric of success, we chose the total count of gold medals and we decided to weigh team all-around events as greater than individual all-around events and individual all-around great than the individual apparatuses. For the women's team, we believe it is best to select a team of five female athletes who are event-specialist gymnasts. The US women's team has a strong shot at winning the individual all-around and many individual apparatus events with multiple-apparatus specialist and highly decorated gold medalist athlete Simone Biles as well as win team all-around with other highly decorated gold medalists who specialize in their apparatus. So focusing on athletes that specialize in apparatuses would be the best strategy (Appendix: Image 10). For the men's team, we believe it is best to select a team of five male athletes who are all-around gymnasts. The top male competitors for each apparatus from Team USA are almost always severely overshadowed by top male competitors from other countries by number of gold medal count. Team USA does not have comparable male competitors who specialize. But in vault the US makes up 7 of the top 10 most decorated gold medalists in the apparatus, so including as vault specialist could help the US men's gymnastics team increase chances of success. (Appendix: Image 12)

Objective 4: Identifying 5 Athletes

Considering the conclusions of the previous objectives, we predict that the following athletes would best optimize gold medal count success for both the male and female US gymnastics team in the 2024 Paris Summer Olympics:

Women's USA Gymnastics Team: Event Specialists Preference

- Simone Biles
- Shilese Jones
- Zoe Miller
- Konnor Mcclain
- Jade Carey

In our results, we saw that Simone Biles was a strong individual apparatus (particularly in floor exercise, vault, and balance beam), individual all-around, and team all-around top-finisher contender–she is both a specialist and a generalist. To supplement Biles's performance and add strength to the US uneven bars performance, we choose Shilese Jones and Zoe Miller for their strong uneven bars rankings as specialists in the event. Furthermore, we choose Konnor McClain as a specialist in balance beam, as she has more simulated gold medals than does Simone Biles, but she may also compete on vault given fifth overall simulated ranking in vault total medals and fourth overall simulated ranking in vault gold medals. Lastly, Jade Carey we choose as another vault specialist given her third overall simulated ranking in vault, but she may also compete in floor exercise given her sixth overall simulated medal ranking in floor exercise.This selection provides two of the strongest USA specialists by medal count for almost each apparatus besides floor exercise since Simone received a whopping count of 192 gold medals in comparison the second place Brazilian contender with 58 gold medals in the simulation (Appendix: Image 9).

Men's USA Gymnastics Team: Mixture of Specialist and All-Around, All-Around Preference

- Asher Hong
- Curran Phillips
- Donnell Whittenburg
- Colt Walker
- Brody Malone

In order to find the expected best all-around male athletes from the simulation, we gave rankings of total count of gold medals for each apparatus, took the average of those rankings for each athlete and picked the top five athletes with the highest average ranking (Appendix: Image 13). This method also matches our simulation results, which shows these five athletes (except for Asher Hong as a vault specialist, yet Asher can also rank high in all-around male athletes) as all-around gymnasts who can place in the top 10 of simulated total medals in at least two apparatuses, and cover all the apparatuses except for pommel horse. We see that other countries have event specialists that may do better than one specific US athlete, but the ability of each selected athlete to cover multiple apparatuses well bodes well for individual all-around and team all-around event medals.

Methodology Evalution

The baseline choices in the methodology: taking a Bayesian approach and relying on the Monte Carlo method, provide a robust foundation and allow for generally informed predictions. However, there are several areas where the implementation of this methodology could be improved. Firstly, the normal distribution is a simple and effective choice for the posterior distributions, but it is not a perfect fit for the data. The normal distribution is unbounded, while gymnastics scores are bounded between 0 and 20. This discrepancy is addressed by truncating the distribution at 0 and 20, but this is not a perfect solution. Furthermore, when determining the prior parameters we use the mean and standard deviation of athlete results in a given apparatus, but this does not account for the fact that athletes are not equally sampled and some have far more data points than others.

We use a single distribution to model final scores, when in reality we should be sampling from a distribution of difficulty scores, then conditioning on difficulty to sample from a distribution of execution scores and penalty scores, and then summing these values to get a final score. This would allow us to more accurately model the data and incorporate the fact that difficulty and execution scores are not independent. Additionally, the simulated competitions include every gymnast in the dataset, so the results are not representative of the actual competition. In reality we should sample the best athletes from every nation and then simulate the multiple individual and team rounds in a gymnastics competition. The choice to avoid this modelling process, alongside the choice to only run 500 competition simulations, were done due to computational constraints. Finally, we assume that the athletes' abilities are constant over time, which is not necessarily true. This assumption is addressed by using the most recent data, but it is still a simplification of the true process.

Data Limitations

We run our Monte Carlo experiments on a subsetted group of observations from 2022-2023 gymnastics competition data, and it is possible that this subset of data are not representative of the athletes' overall performances. Data prior to 2022 was not used because of the short athletic life-cycle of gymnasts and data only on women, so unfortunately we may have discarded pertinent data to some athletes' performances. Additionally, we had to remove data for athletes' scores when the athlete did not compete in more than 2 competitions for an apparatus, which may have excluded some potentially strong athletes if they had not competed in the 2022-2023 seasons due to injury, medical, or other reasons. Finally, given there are a limited number of international competitions per season, most athletes in the data set were competing in less than ten competitions–some only competing in 1 or 2 competitions across 2022-2023. The limited number of score results per athlete may lead to hard to generalize results.

Implications and Conclusions

An iteration of this study would be stronger if it addressed the methodological limitations mentioned above. Additionally, including more historical data and athletes' age in the model would create a much more robust model. The current methodology also lacks a selection mechanism, relying on us to manually select the top five athletes. A more robust model would include a selection mechanism that would automatically select the top five athletes based on the results of the Monte Carlo simulations. Another change that would have advanced the specificity in the vault apparatus scores within our simulations would have been handling vault 1 and vault 2 as separate entities. Within individual apparatus events, two different vaults are required in the qualification and final event but in our simulation we used the mean of both vault events instead. Lastly, another change in our methodology would be to run 1000 competition simulations to produce more generalizable resulting scores for each athlete. Instead 500 competition simulations were used to be computationally conservative considering there were a multitude of athlete- and apparatus-level score observations. Overall, this study provides valuable insights on an optimal team selection strategy that would best aid the United States artistic gymnastics male and female team achieve success in the 2024 Paris Summer Olympics Games.

Appendix

Works Cited

- Camenker, Jacob. "How Old Is Simone Biles? Why Elite Olympic Gymnasts Typically Retire at a Young Age." Sporting News, 18 Sept. 2021, www.sportingnews.com/us/athletics/news/simone-biles-retire-age-olympics/1laom4i4u4wh1thcgta4nun2x. Accessed 20 Nov. 2023.
- Hoff, Peter D. A First Course in Bayesian Statistical Methods. Springer, 2009.
- Meyers, Dvora. "Time for the End of the Teen Gymnast." FiveThirtyEight, 27 July 2021, fivethirtyeight.com/features age-olympics/. Accessed 20 Nov. 2023.
- Muller, Marie Laure Delignette, and Christophe Dutang. "Overview of the Fitdistrplus Package." R-Project, CRAN, 25 Apr. 2023, cran.r-project.org/web/packages/fitdistrplus/vignettes/fitdistrplus_vignette.html.
- "Paris 2024 Olympic Games: How Do Athletes Qualify?" Olympics, 8 Aug. 2021, olympics.com/en/news/paris-2024-olympic-games-how-do-athletes-qualify. Accessed 20 Nov. 2023.
- Santos-Fernandez, Edgar, et al. "Bayesian Statistics Meets Sports: A Comprehensive Review." Journal of Quantitative Analysis in Sports, vol. 15, no. 4, Dec. 2019, pp. 289–312. www.degruyter.com, https://doi.org/10.1515/jqas-2018-0106.
- "Tokyo 2020 Artistic Gymnastics - Olympic Results by Discipline." Olympics, olympics.com/en/olympic-games/tokyo-2020/results/artistic-gymnastics. Accessed 20 Nov. 2023.
- "UCSAS 2024 USOPC DATA CHALLENGE." UConn Sports Analytics Symposium (UCSAS), statds.org/events/ucsa Accessed 20 Nov. 2023.
- Yang, Tae Young, and Tim Swartz. "A Two-Stage Bayesian Model for Predicting Winners in Major League Baseball." Journal of Data Science, vol. 2, no. 1, Aug. 2022, pp. 61–73. jds-online.org, https://doi.org/10.6339/JDS.2004.02(1).142.

Additional Simulation Results

Given women compete on 4 apparatuses and men compete on 4 apparatuses, we have tables of the simulation outcomes for all 10 apparatuses, and the 5 simluation outcome tables where US gymnasts do not perform as well relative to athletes from other countries are outputted below.

Table 3: Women's Balance Beam Simulation Results

Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Simone Biles: USA	49	71	50	170
Yaqin Zhou: CHN	68	49	42	159
Konnor McClain: USA	65	47	29	141
Qingying Zhang: CHN	54	32	41	127
Sunisa Lee: USA	31	35	24	90
Huan Luo: CHN	24	15	17	56
Yushan Ou: CHN	12	21	17	50
Urara Ashikawa: JPN	15	15	11	41
Rebeca Andrade: BRA	11	14	13	38
Skye Blakely: USA	10	12	16	38

Table 4: Women's Vault Simulation Results

Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Simone Biles: USA	151	94	83	328
Rebeca Andrade: BRA	103	79	63	245
Jade Carey: USA	61	78	60	199
Jordan Chiles: USA	28	34	43	105
Konnor McClain: USA	40	32	31	103
Shilese Jones: USA	22	38	39	99
Ondine Achampong: GBR	14	31	46	91
Joscelyn Roberson: USA	15	29	34	78
Shokyo Miyata: JPN	26	21	30	77
Tiana Sumanasekera: USA	18	31	26	75

Table 5: Women's Uneven Bars Simulation Results

Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Kayla Neymour: ALG	82	55	42	179
Qiyan Qiu: CHN	51	43	51	145
Shilese Jones: USA	32	40	36	108
Alice D'Amato: ITA	28	37	36	101
Xijing Tang: CHN	33	29	30	92
Xiaoyuan Wei: CHN	30	28	31	89
Zoe Miller: USA	35	29	24	88
Rebeca Andrade: BRA	22	19	21	62
Elisabeth Seitz: GER	17	17	23	57
Simone Biles: USA	15	16	21	52

Table 6: Men's Vault Simulation Results

Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Jake Jarman: GBR	91	70	48	209
Asher Hong: USA	67	58	58	183
Daiki Hashimoto: JPN	55	55	46	156
Boheng Zhang: CHN	54	44	37	135
Khoi Young: USA	29	43	47	119
Curran Phillips: USA	33	42	37	112
Donnell Whittenburg: USA	35	38	38	111
Dallas Hale: USA	34	29	47	110
Colt Walker: USA	21	28	36	85
Taylor Burkhart: USA	25	27	27	79

Table 7: Men's Floor Exercise Simulation Results

Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Carlos Yulo: PHI	48	33	46	127
Ryosuke Doi: JPN	26	26	31	83
Artem Dolgopyat: ISR	21	22	26	69
Paul Juda: USA	28	21	13	62
Daiki Hashimoto: JPN	16	25	16	57
Boheng Zhang: CHN	19	18	12	49
Nicola Bartolini: ITA	13	18	12	43
Donnell Whittenburg: USA	16	18	9	43
Colt Walker: USA	16	13	12	41
Brody Malone: USA	11	14	15	40

Table 8: Men's High Bar Simulation Results

Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Daiki Hashimoto: JPN	44	45	35	124
Boheng Zhang: CHN	50	26	33	109
Cong Shi: CHN	35	38	27	100
Brody Malone: USA	18	29	25	72
Weide Su: CHN	23	20	16	59
Wei Sun: CHN	26	20	13	59
Shohei Kawakami: JPN	16	23	19	58
Ilias Georgiou: CYP	18	15	20	53
Milad Karimi: KAZ	25	11	15	51
Arthur Mariano: BRA	11	19	16	46

Table 9: Men's Pommel Horse Simulation Results

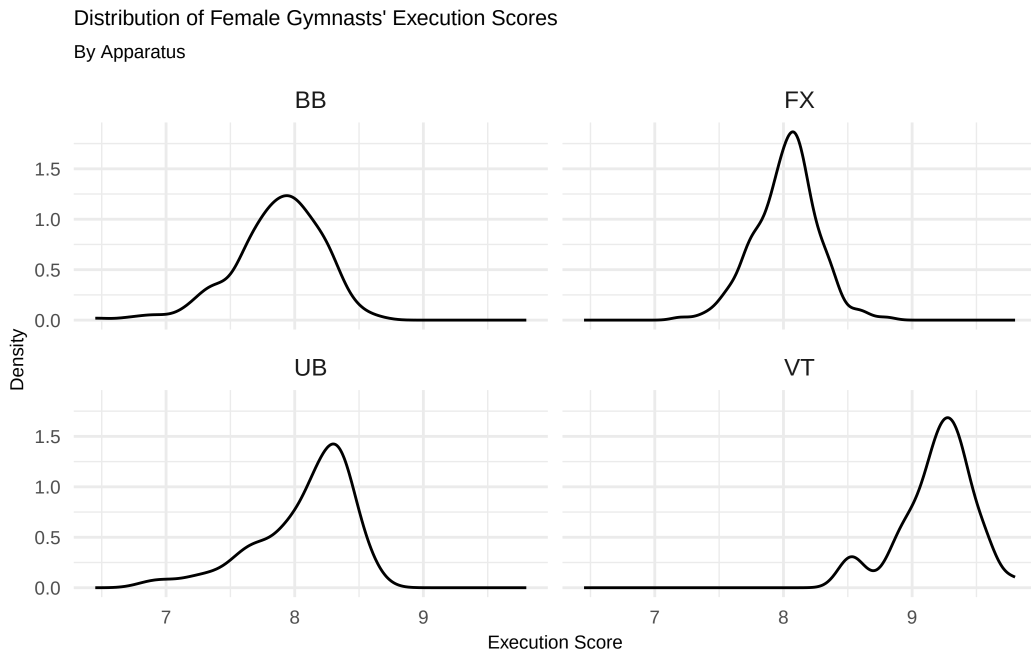
Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Max Whitlock: GBR	84	40	29	153
Chih Lee: TPE	57	45	27	129
Nariman Kurbanov: KAZ	31	52	35	118
Ahmad Abu Al Soud: JOR	15	26	25	66
Rhys McClenaghan: IRL	22	24	19	65
Stephen Nedoroscik: USA	20	15	18	53
Loran De Munck: NED	21	16	13	50
Gagik Khachikyan: ARM	7	18	22	47
Yu-Jan Shiao: TPE	11	16	18	45
Kakeru Tanigawa: JPN	11	12	21	44

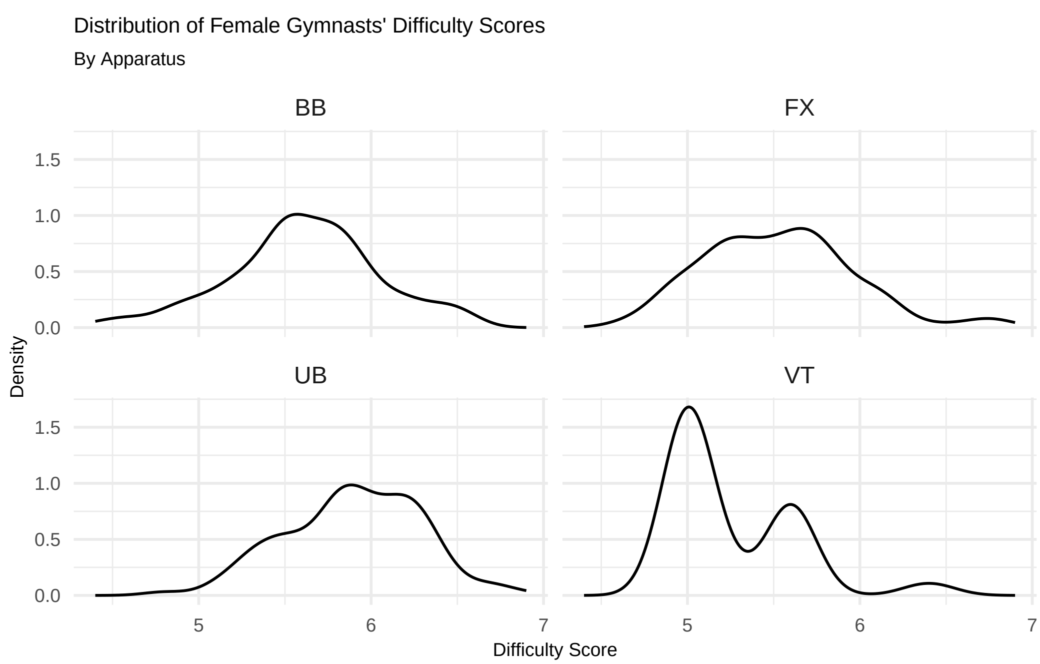
Table 10: Men's Still Rings Simulation Results

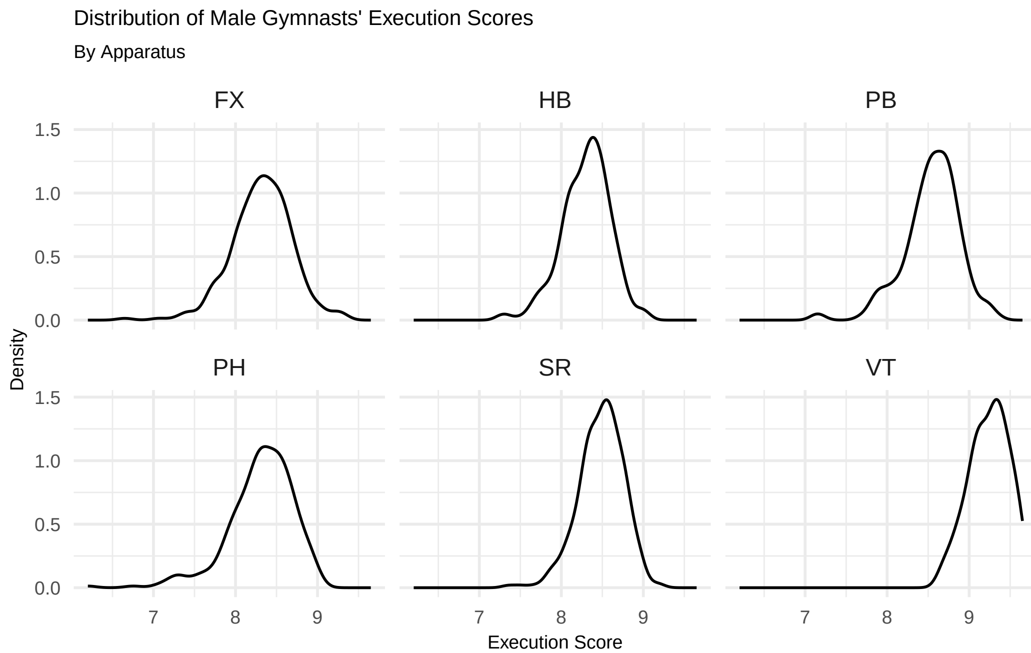
Athlete & Country	Golds	Silvers	Bronzes	Total Medals
Yang Liu: CHN	69	57	40	166
Xingyu Lan: CHN	41	47	29	117
Eleftherious Petrounias: GRE	38	34	36	108
Jingyuan Zou: CHN	44	25	23	92
Hao You: CHN	18	19	24	61
Ibrahim Colak: TUR	18	22	18	58
Donnell Whittenburg: USA	16	12	28	56
Adem Asil: TUR	10	17	22	49
Salvatore Maresca: ITA	12	24	13	49
Boheng Zhang: CHN	17	19	9	45

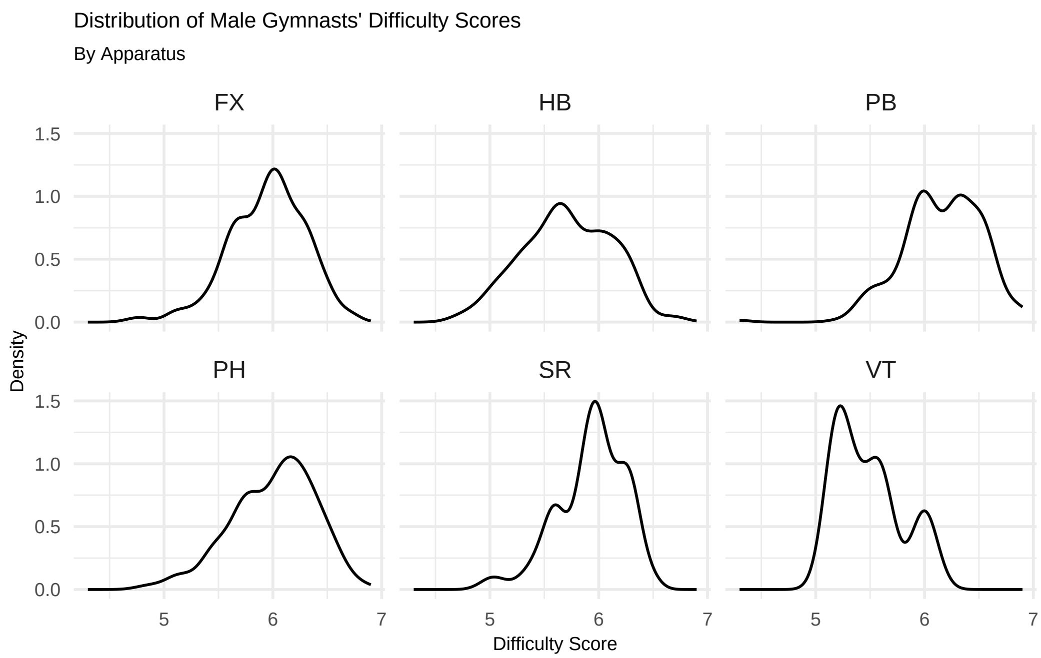
Extra Visualizations

The following visualizations show the distribution of difficulty and execution scores by apparatus for male and female gymnasts, which are still approximately normal but do show more drastic deviations from normality than do the overall scores for each gymnast at an apparatus in a competition round. So, we thought it would be more fitting to fit normal-inverse gamma priors on the means and variances of the overall scores.

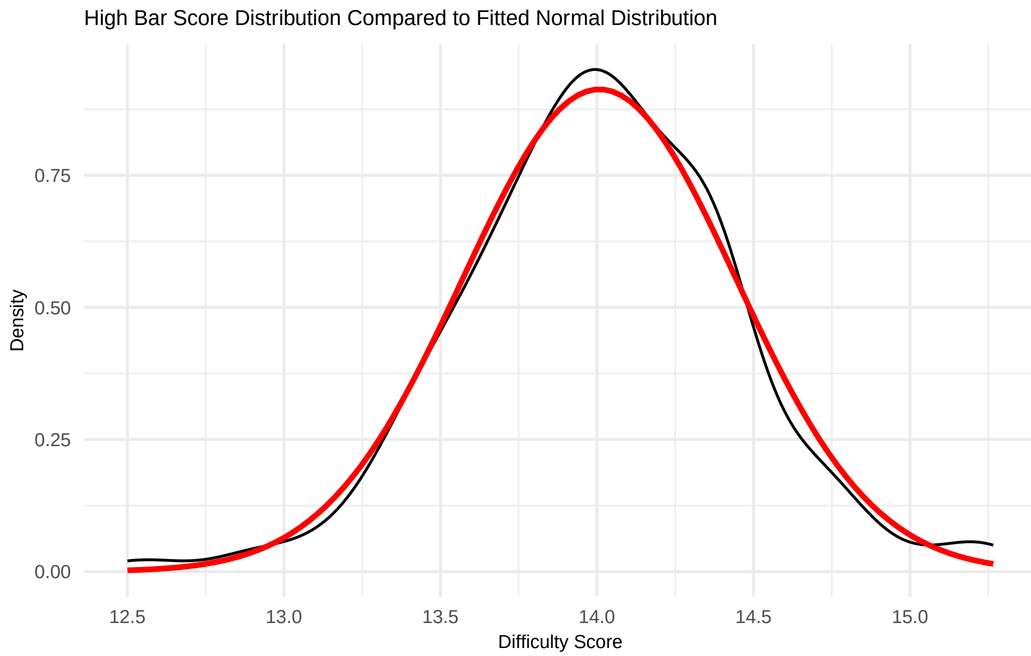

Distribution of Female Gymnasts' Execution Scores
By Apparatus
BB
FX
UB
VT
Density
1.5
1.0
0.5
0.0
7
8
9
Execution Score

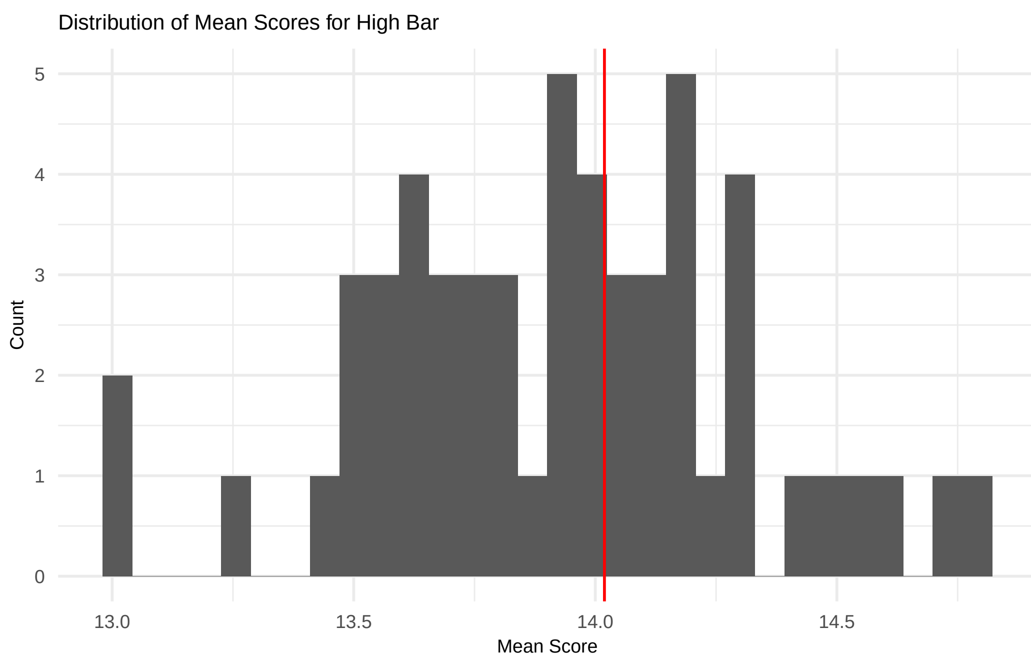

Distribution of Female Gymnasts' Difficulty Scores
By Apparatus
BB
FX
UB
VT
Density
1.5
1.0
0.5
0.0
5
6
7
Difficulty Score

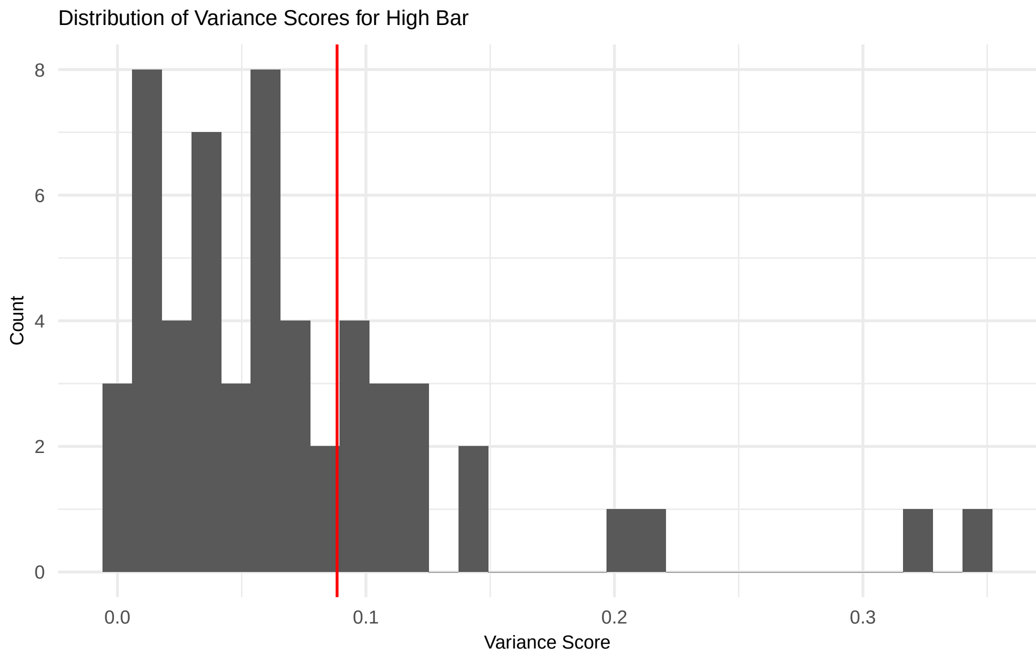




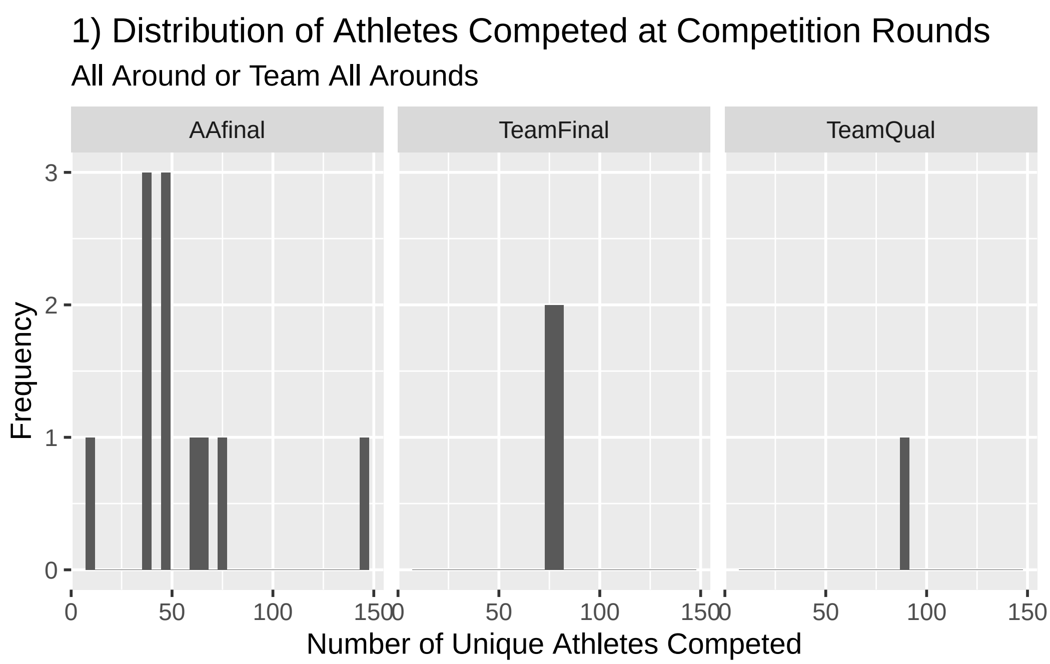
Additionally, we look at the distribution of mean scores and the distribution of variance scores for men's high bar apparatus as an example of how we will fit normal-inverse gamma priors for the means and variances. We can see that the distribution of mean scores follows an approximately normal distribution and the distribution of variance scores follow an approximately inverse-gamma distribution.


High Bar Score Distribution Compared to Fitted Normal Distribution
0.75
0.50
0.25
0.00
Density
12.5
13.0
13.5
14.0
14.5
15.0
Difficulty Score

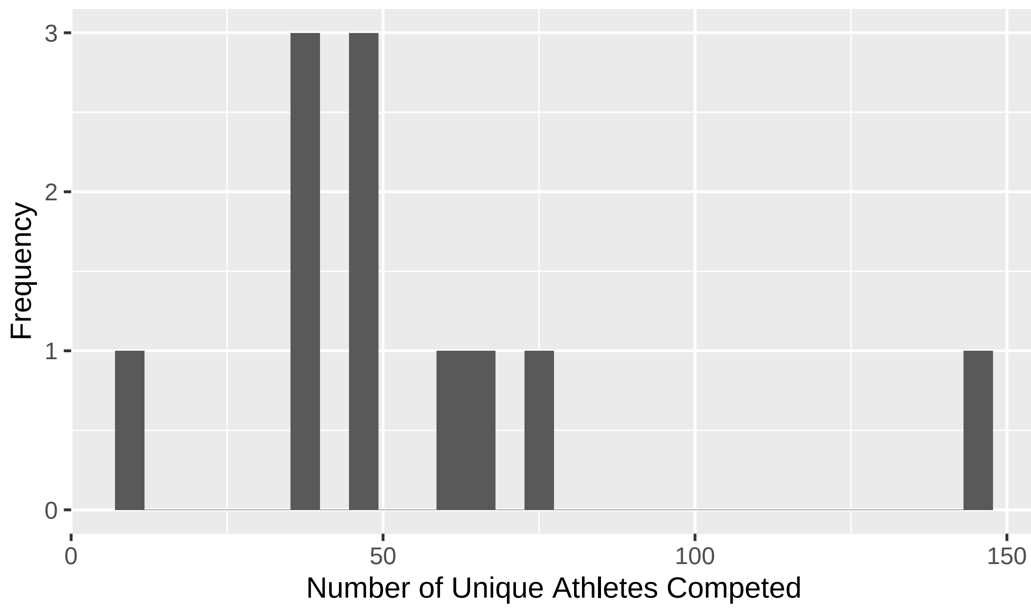

Distribution of Mean Scores for High Bar
5
4
3
2
1
0
Count
13.0
13.5
14.0
14.5
Mean Score



The below plots visualize how many unique athletes are competing at each round in a competition, spearated by gender, so that we can understand sample size for when we filter out data.

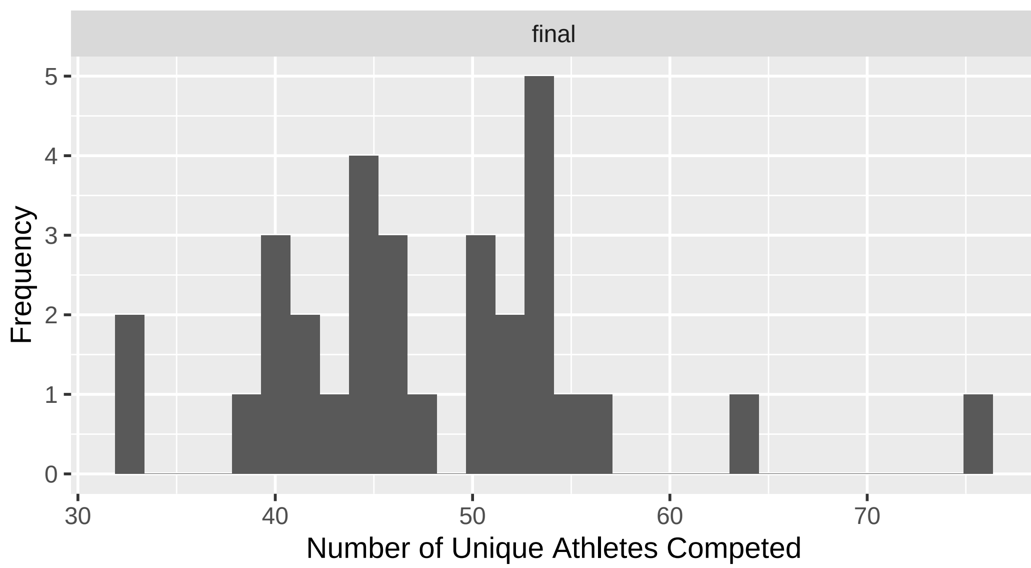


2) Distribution of Athletes Competed at AA Finals



3) Distribution of Athletes Competed at Final Rounds

Individual Apparatuses



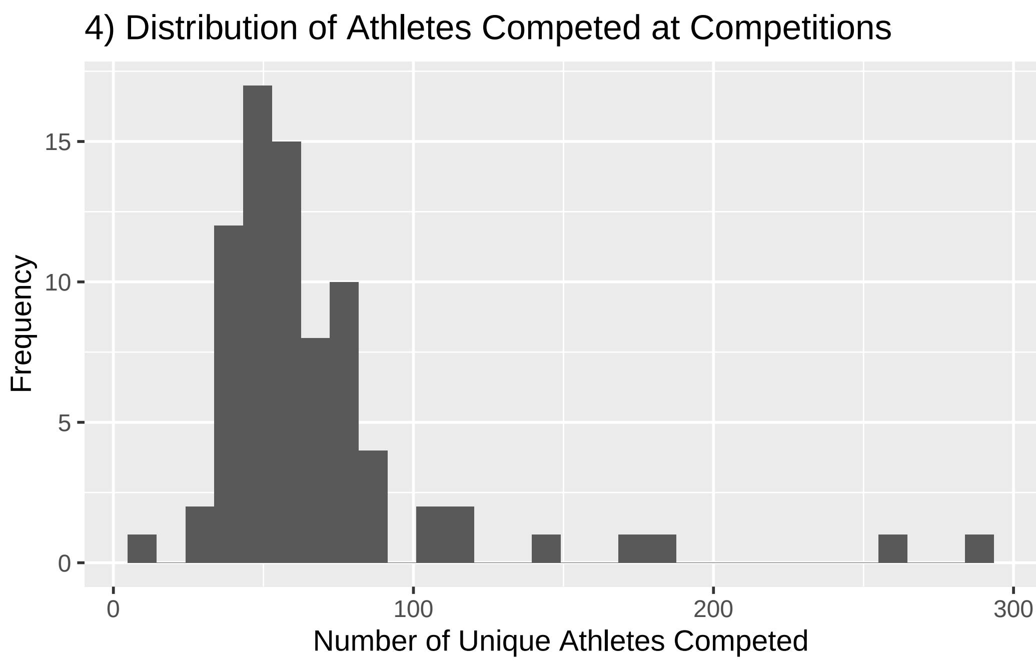
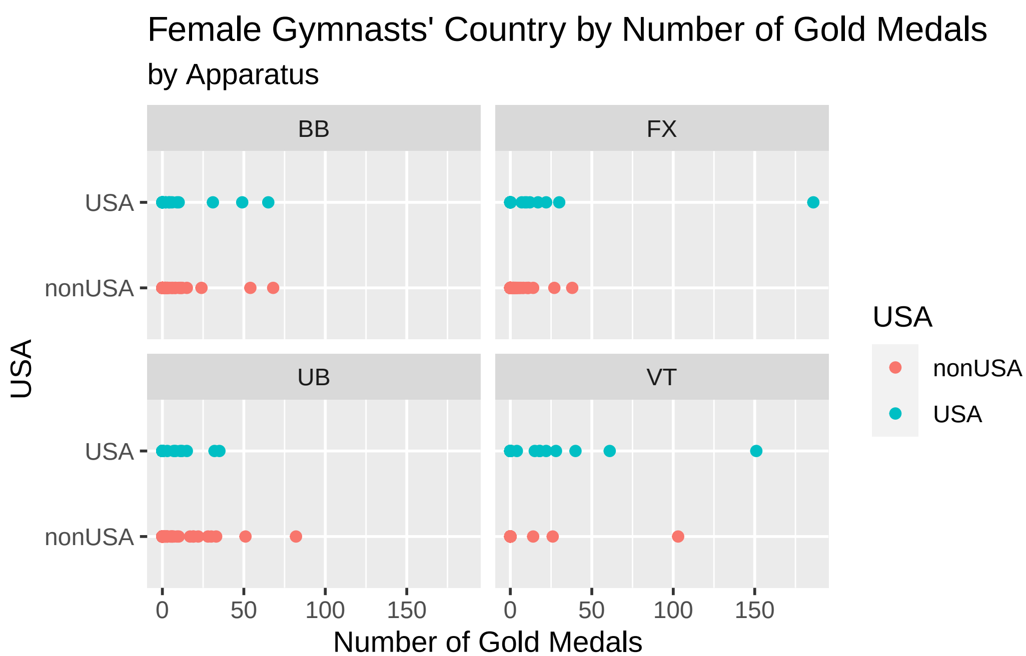
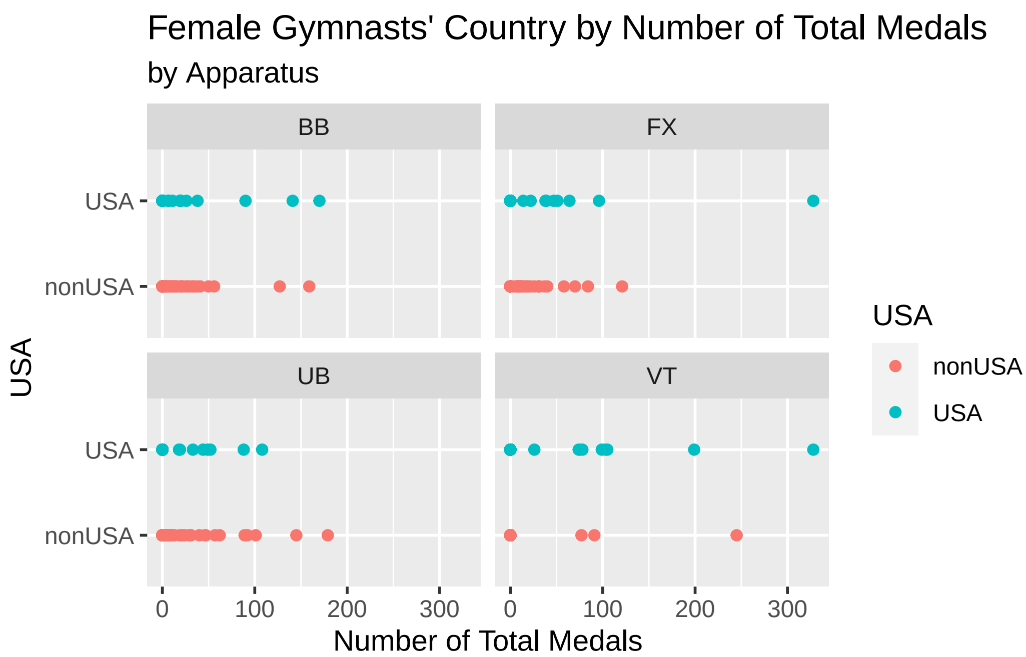

4) Distribution of Athletes Competed at Competitions
Frequency
15
10
5
0
0
100
200
300
Number of Unique Athletes Competed

Image 5)


Female Gymnasts' Country by Number of Gold Medals
by Apparatus
BB
FX
UB
VT
USA
nonUSA
USA
0
50
100
150
Number of Gold Medals
USA
nonUSA
USA


Female Gymnasts' Country by Number of Total Medals
by Apparatus
BB
FX
UB
VT
USA
nonUSA
0
100
200
300
Number of Total Medals
USA
nonUSA
USA

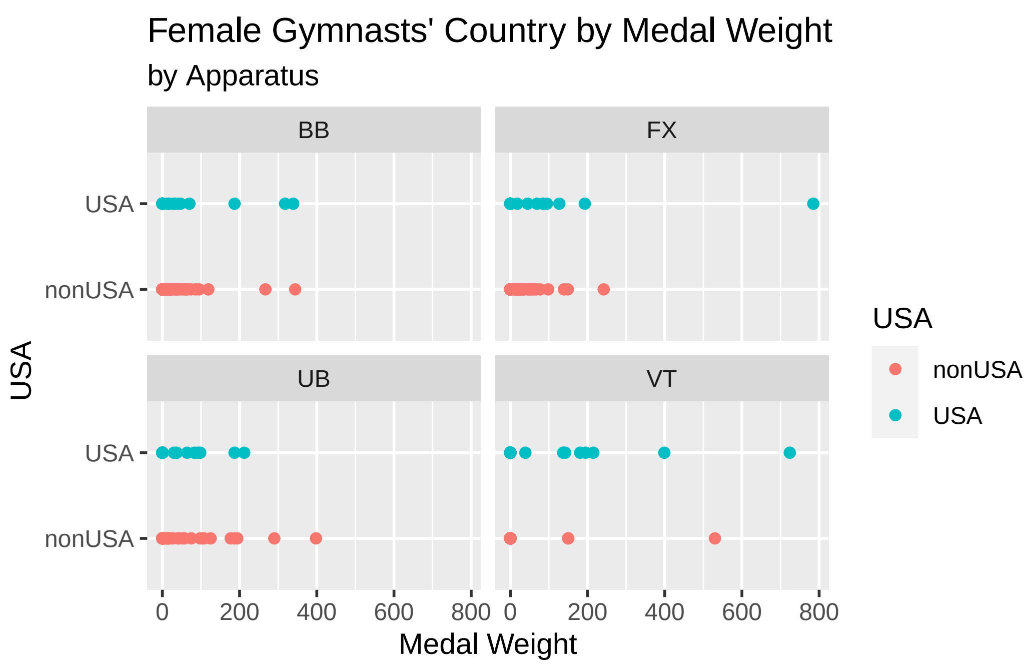
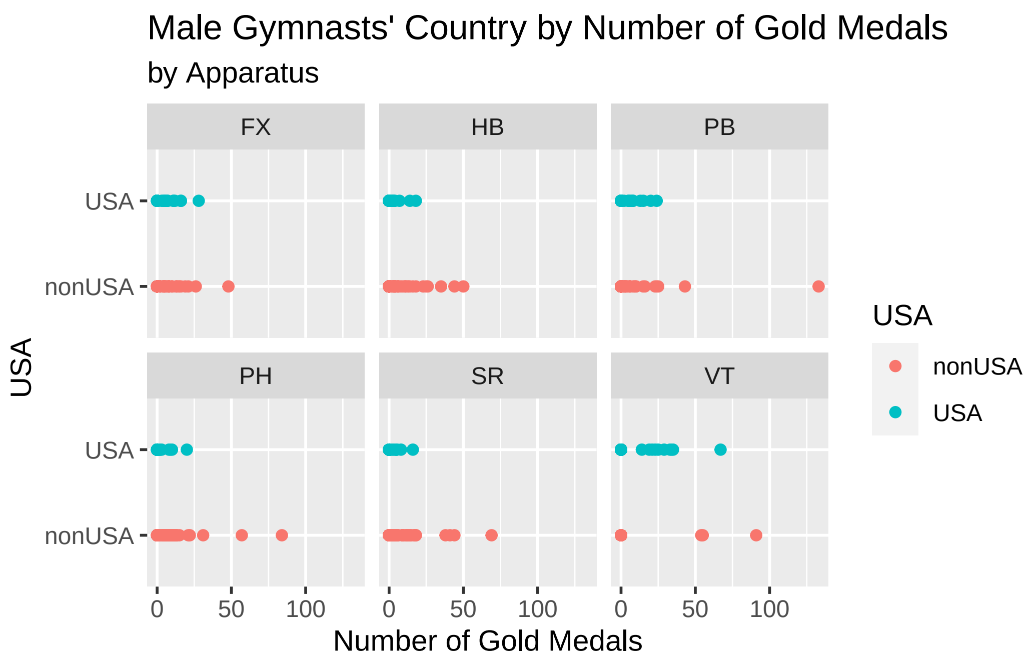
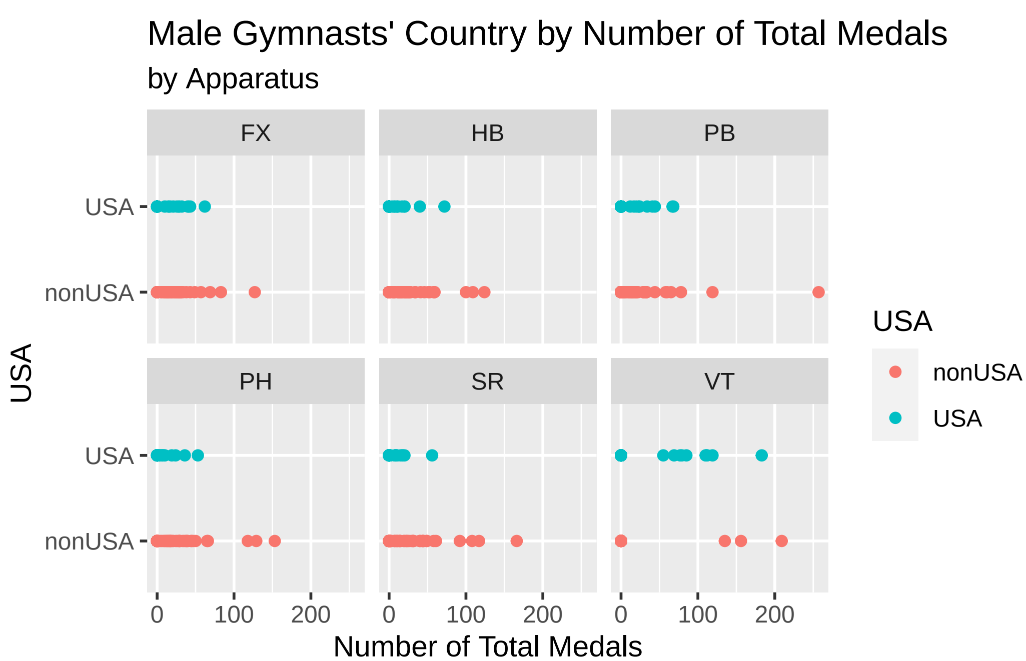

Female Gymnasts' Country by Medal Weight
by Apparatus
BB
FX
UB
VT
USA
nonUSA
0
200
400
600
800
Medal Weight
USA
nonUSA
USA

Image 6)


Male Gymnasts' Country by Number of Gold Medals
by Apparatus
FX
HB
PB
PH
SR
VT
USA
nonUSA
0
50
100
Number of Gold Medals
USA
nonUSA
USA


Male Gymnasts' Country by Number of Total Medals
by Apparatus
FX
HB
PB
PH
SR
VT
USA
nonUSA
0
100
200
Number of Total Medals
USA
nonUSA
USA

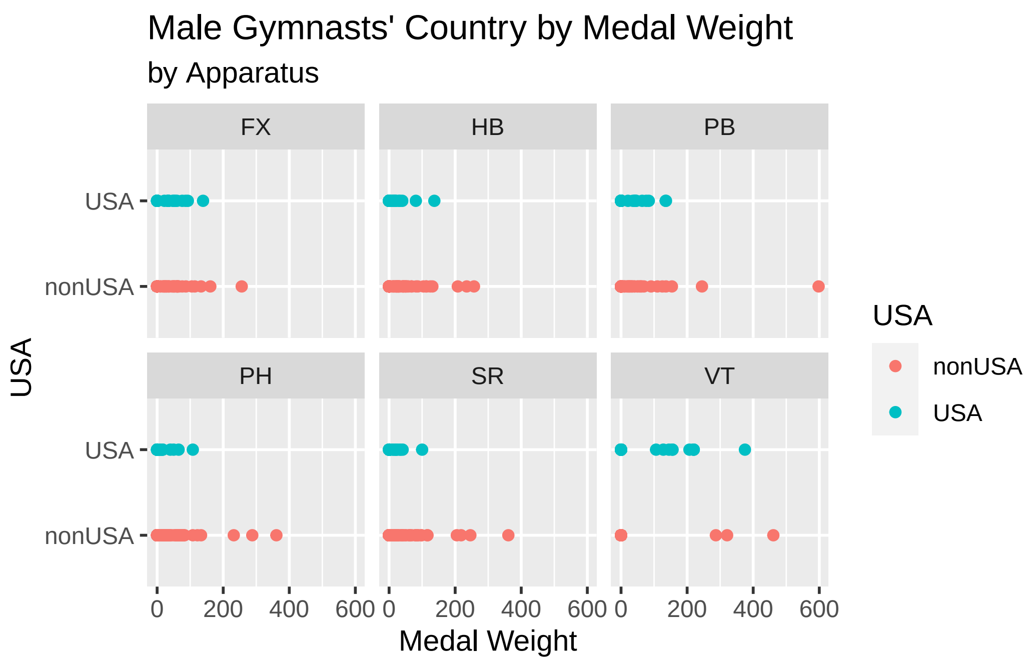


Image 7)

Women:

- Top 5 athletes by apparatus for each of the 3 success metrics
- Sum of each of the 3 metrics made by athletes from the US and non-US countries

unique_id	Golds	Silvers	Bronzes	Total_Medals	Country	Medal_Weight	Apparatus	USA
YaqZHO_CHN	68	49	42	159	CHN	344	BB	nonUSA
KonMCC_USA	65	47	29	141	USA	318	BB	USA
QinZHA_CHN	54	32	41	127	CHN	267	BB	nonUSA
SimBIL_USA	49	71	50	170	USA	339	BB	USA
SunLEE_USA	31	35	24	90	USA	187	BB	USA
SimBIL_USA	186	85	57	328	USA	785	FX	USA
RebAND_BRA	38	45	38	121	BRA	242	FX	nonUSA
KalLIN_USA	30	37	29	96	USA	193	FX	USA
FlaSAR_BRA	27	25	18	70	BRA	149	FX	nonUSA
JadCAR_USA	22	19	23	64	USA	127	FX	USA
KayNEM_ALG	82	55	42	179	ALG	398	UB	nonUSA
QiyQIU_CHN	51	43	51	145	CHN	290	UB	nonUSA
ZoeMIL_USA	35	29	24	88	USA	187	UB	USA
XijTAN_CHN	33	29	30	92	CHN	187	UB	nonUSA
ShiJON_USA	32	40	36	108	USA	212	UB	USA
SimBIL_USA	151	94	83	328	USA	724	VT	USA
RebAND_BRA	103	79	63	245	BRA	530	VT	nonUSA
JadCAR_USA	61	78	60	199	USA	399	VT	USA
KonMCC_USA	40	32	31	103	USA	215	VT	USA
JorCHI_USA	28	34	43	105	USA	195	VT	USA

unique_id	Golds	Silvers	Bronzes	Total_Medals	Country	Medal_Weight	Apparatus	USA
SimBIL_USA	49	71	50	170	USA	339	BB	USA
YaqZHO_CHN	68	49	42	159	CHN	344	BB	nonUSA
KonMCC_USA	65	47	29	141	USA	318	BB	USA
QinZHA_CHN	54	32	41	127	CHN	267	BB	nonUSA
SunLEE_USA	31	35	24	90	USA	187	BB	USA
SimBIL_USA	186	85	57	328	USA	785	FX	USA
RebAND_BRA	38	45	38	121	BRA	242	FX	nonUSA
KalLIN_USA	30	37	29	96	USA	193	FX	USA
JesGAD_GBR	14	27	43	84	GBR	139	FX	nonUSA
FlaSAR_BRA	27	25	18	70	BRA	149	FX	nonUSA
KayNEM_ALG	82	55	42	179	ALG	398	UB	nonUSA
QiyQIU_CHN	51	43	51	145	CHN	290	UB	nonUSA
ShiJON_USA	32	40	36	108	USA	212	UB	USA
AliD A_ITA	28	37	36	101	ITA	194	UB	nonUSA
XijTAN_CHN	33	29	30	92	CHN	187	UB	nonUSA
SimBIL_USA	151	94	83	328	USA	724	VT	USA
RebAND_BRA	103	79	63	245	BRA	530	VT	nonUSA
JadCAR_USA	61	78	60	199	USA	399	VT	USA
JorCHI_USA	28	34	43	105	USA	195	VT	USA
KonMCC_USA	40	32	31	103	USA	215	VT	USA

unique_id	Golds	Silvers	Bronzes	Total_Medals	Country	Medal_Weight	Apparatus	USA
YaqZHO_CHN	68	49	42	159	CHN	344	BB	nonUSA
SimBIL_USA	49	71	50	170	USA	339	BB	USA
KonMCC_USA	65	47	29	141	USA	318	BB	USA
QinZHA_CHN	54	32	41	127	CHN	267	BB	nonUSA
SunLEE_USA	31	35	24	90	USA	187	BB	USA
SimBIL_USA	186	85	57	328	USA	785	FX	USA
RebAND_BRA	38	45	38	121	BRA	242	FX	nonUSA
KalLIN_USA	30	37	29	96	USA	193	FX	USA
FlaSAR_BRA	27	25	18	70	BRA	149	FX	nonUSA
JesGAD_GBR	14	27	43	84	GBR	139	FX	nonUSA
KayNEM_ALG	82	55	42	179	ALG	398	UB	nonUSA
QiyQIU_CHN	51	43	51	145	CHN	290	UB	nonUSA
ShiJON_USA	32	40	36	108	USA	212	UB	USA
AliD A_ITA	28	37	36	101	ITA	194	UB	nonUSA
XijTAN_CHN	33	29	30	92	CHN	187	UB	nonUSA
ZoeMIL_USA	35	29	24	88	USA	187	UB	USA
SimBIL_USA	151	94	83	328	USA	724	VT	USA
RebAND_BRA	103	79	63	245	BRA	530	VT	nonUSA
JadCAR_USA	61	78	60	199	USA	399	VT	USA
KonMCC_USA	40	32	31	103	USA	215	VT	USA
JorCHI_USA	28	34	43	105	USA	195	VT	USA

USA	sumGolds	sumTotal	sumWeighted
nonUSA	1040	3236	6380
USA	960	2764	5620

For the women's simulation when looking at the top 5 athletes by:

- *Gold Medal Count* for each apparatus there are 10 out of 20 from the US: balance beam (BB): 3, floor exercise (FX): 3, uneven bars (UB): 2, and vault (VT): 2

- USA makes up 51% of the total women's gold medals in the simulation.

- *Total Medal Count* for each apparatus there are 12 out of 20 from the US: balance beam (BB): 3, floor exercise (FX): 4, uneven bars (UB): 1, vault (VT): 4
  - USA makes up 47% of the total women's medals in the simulation.
- *Weighted Medal Count* for each apparatus there are 10 out of 20 from the US: balance beam (BB): 3, floor exercise (FX): 2, uneven bars (UB): 1, vault (VT): 4
  - USA makes up 48% of the weight of women's medals in the simulation.

Image 8)

Men:

- Top 5 athletes by apparatus for each of the 3 success metrics
- Sum of each of the 3 metrics made by athletes from the US and non-US countries

unique_id	Golds	Silvers	Bronzes	Total_Medals	Country	Medal_Weight	Apparatus	USA
CarYUL_PHI	48	33	46	127	PHI	256	FX	nonUSA
PauJUD_USA	28	21	13	62	USA	139	FX	USA
RyoDOI_JPN	26	26	31	83	JPN	161	FX	nonUSA
ArtDOL_ISR	21	22	26	69	ISR	133	FX	nonUSA
BohZHA_CHN	19	18	12	49	CHN	105	FX	nonUSA
BohZHA_CHN	50	26	33	109	CHN	235	HB	nonUSA
DaiHAS_JPN	44	45	35	124	JPN	257	HB	nonUSA
ConSHI_CHN	35	38	27	100	CHN	208	HB	nonUSA
WeiSUN_CHN	26	20	13	59	CHN	131	HB	nonUSA
MilKAR_KAZ	25	11	15	51	KAZ	112	HB	nonUSA
JinZOU_CHN	133	75	49	257	CHN	598	PB	nonUSA
LukDAU_GER	43	40	36	119	GER	245	PB	nonUSA
BohZHA_CHN	25	26	27	78	CHN	154	PB	nonUSA
CurPHI_USA	24	20	24	68	USA	136	PB	USA
JoeFRA_GBR	24	17	19	60	GBR	125	PB	nonUSA
MaxWHI_GBR	84	40	29	153	GBR	361	PH	nonUSA
ChiLEE_TPE	57	45	27	129	TPE	288	PH	nonUSA
NarKUR_KAZ	31	52	35	118	KAZ	232	PH	nonUSA
RhyMCC_IRL	22	24	19	65	IRL	133	PH	nonUSA
LorDE _NED	21	16	13	50	NED	108	PH	nonUSA
YanLIU_CHN	69	57	40	166	CHN	361	SR	nonUSA
JinZOU_CHN	44	25	23	92	CHN	205	SR	nonUSA
XinLAN_CHN	41	47	29	117	CHN	246	SR	nonUSA
ElePET_GRE	38	34	36	108	GRE	218	SR	nonUSA
HaoYOU_CHN	18	19	24	61	CHN	116	SR	nonUSA
IbrCOL_TUR	18	22	18	58	TUR	116	SR	nonUSA
JakJAR_GBR	91	70	48	209	GBR	461	VT	nonUSA
AshHON_USA	67	58	58	183	USA	375	VT	USA
DaiHAS_JPN	55	55	46	156	JPN	321	VT	nonUSA
BohZHA_CHN	54	44	37	135	CHN	287	VT	nonUSA
DonWHI_USA	35	38	38	111	USA	219	VT	USA

unique_id	Golds	Silvers	Bronzes	Total_Medals	Country	Medal_Weight	Apparatus	USA
CarYUL_PHI	48	33	46	127	PHI	256	FX	nonUSA
RyoDOI_JPN	26	26	31	83	JPN	161	FX	nonUSA
ArtDOL_ISR	21	22	26	69	ISR	133	FX	nonUSA
PauJUD_USA	28	21	13	62	USA	139	FX	USA
DaiHAS_JPN	16	25	16	57	JPN	114	FX	nonUSA
DaiHAS_JPN	44	45	35	124	JPN	257	HB	nonUSA
BohZHA_CHN	50	26	33	109	CHN	235	HB	nonUSA
ConSHI_CHN	35	38	27	100	CHN	208	HB	nonUSA
BroMAL_USA	18	29	25	72	USA	137	HB	USA
WeiSU___CHN	23	20	16	59	CHN	125	HB	nonUSA
WeiSUN_CHN	26	20	13	59	CHN	131	HB	nonUSA
JinZOU_CHN	133	75	49	257	CHN	598	PB	nonUSA
LukDAU_GER	43	40	36	119	GER	245	PB	nonUSA
BohZHA_CHN	25	26	27	78	CHN	154	PB	nonUSA
CurPHI_USA	24	20	24	68	USA	136	PB	USA
ColWAL_USA	20	28	19	67	USA	135	PB	USA
MaxWHI_GBR	84	40	29	153	GBR	361	PH	nonUSA
ChiLEE_TPE	57	45	27	129	TPE	288	PH	nonUSA
NarKUR_KAZ	31	52	35	118	KAZ	232	PH	nonUSA
AhmABU_JOR	15	26	25	66	JOR	122	PH	nonUSA
RhyMCC_IRL	22	24	19	65	IRL	133	PH	nonUSA
YanLIU_CHN	69	57	40	166	CHN	361	SR	nonUSA
XinLAN_CHN	41	47	29	117	CHN	246	SR	nonUSA
ElePET_GRE	38	34	36	108	GRE	218	SR	nonUSA
JinZOU_CHN	44	25	23	92	CHN	205	SR	nonUSA
HaoYOU_CHN	18	19	24	61	CHN	116	SR	nonUSA
JakJAR_GBR	91	70	48	209	GBR	461	VT	nonUSA
AshHON_USA	67	58	58	183	USA	375	VT	USA
DaiHAS_JPN	55	55	46	156	JPN	321	VT	nonUSA
BohZHA_CHN	54	44	37	135	CHN	287	VT	nonUSA
KhoYOU_USA	29	43	47	119	USA	220	VT	USA

unique_id	Golds	Silvers	Bronzes	Total_Medals	Country	Medal_Weight	Apparatus	USA
CarYUL_PHI	48	33	46	127	PHI	256	FX	nonUSA
RyoDOI_JPN	26	26	31	83	JPN	161	FX	nonUSA
PauJUD_USA	28	21	13	62	USA	139	FX	USA
ArtDOL_ISR	21	22	26	69	ISR	133	FX	nonUSA
DaiHAS_JPN	16	25	16	57	JPN	114	FX	nonUSA
DaiHAS_JPN	44	45	35	124	JPN	257	HB	nonUSA
BohZHA_CHN	50	26	33	109	CHN	235	HB	nonUSA
ConSHI_CHN	35	38	27	100	CHN	208	HB	nonUSA
BroMAL_USA	18	29	25	72	USA	137	HB	USA
WeiSUN_CHN	26	20	13	59	CHN	131	HB	nonUSA
JinZOU_CHN	133	75	49	257	CHN	598	PB	nonUSA
LukDAU_GER	43	40	36	119	GER	245	PB	nonUSA
BohZHA_CHN	25	26	27	78	CHN	154	PB	nonUSA
CurPHI_USA	24	20	24	68	USA	136	PB	USA
KaiSUG_JPN	23	25	17	65	JPN	136	PB	nonUSA
MaxWHI_GBR	84	40	29	153	GBR	361	PH	nonUSA
ChiLEE_TPE	57	45	27	129	TPE	288	PH	nonUSA
NarKUR_KAZ	31	52	35	118	KAZ	232	PH	nonUSA
RhyMCC_IRL	22	24	19	65	IRL	133	PH	nonUSA
AhmABU_JOR	15	26	25	66	JOR	122	PH	nonUSA
YanLIU_CHN	69	57	40	166	CHN	361	SR	nonUSA
XinLAN_CHN	41	47	29	117	CHN	246	SR	nonUSA
ElePET_GRE	38	34	36	108	GRE	218	SR	nonUSA
JinZOU_CHN	44	25	23	92	CHN	205	SR	nonUSA
HaoYOU_CHN	18	19	24	61	CHN	116	SR	nonUSA
IbrCOL_TUR	18	22	18	58	TUR	116	SR	nonUSA
JakJAR_GBR	91	70	48	209	GBR	461	VT	nonUSA
AshHON_USA	67	58	58	183	USA	375	VT	USA
DaiHAS_JPN	55	55	46	156	JPN	321	VT	nonUSA
BohZHA_CHN	54	44	37	135	CHN	287	VT	nonUSA
KhoYOU_USA	29	43	47	119	USA	220	VT	USA
CurPHI_USA	33	42	37	112	USA	220	VT	USA

USA	sumGolds	sumTot	sumWeighted
nonUSA	2293	6667	13490
USA	707	2333	4510

For the men's simulation when looking at the top 5 athletes by:

- *Gold Medal Count* for each apparatus there are 5 out of 30 from the US: floor exercise (FX): 1, high bar (HB): 1, parallel bars (PB): 1 pommel horse (PH): 0, still rings (SR): 0, vault (VT): 2
  - USA makes up 21% of the total men's gold medals in the simulation.
- *Total Medal Count* for each apparatus there are 4 out of 30 from the US: floor exercise (FX): 1, high bar (HB): 1, parallel bars (PB): 0, pommel horse (PH): 0, still rings (SR): 0, vault (VT): 2
  - USA makes up 24% of the total men's medals in the simulation.
- *Weighted Medal Count* for each apparatus there are 4 out of 30 from the US: floor exercise (FX): 1, high bar (HB): 1, parallel bars (PB): 0, pommel horse (PH): 0, still rings (SR): 0, vault (VT): 2
  - USA makes up 23% of the weight of men's medals in the simulation.

Image 9) Top ten most successful female gymnast using total gold medal count by apparatus

unique_id	Golds	Country	Apparatus	USA
YaqZHO_CHN	68	CHN	BB	nonUSA
KonMCC_USA	65	USA	BB	USA
QinZHA_CHN	54	CHN	BB	nonUSA
SimBIL_USA	49	USA	BB	USA
SunLEE_USA	31	USA	BB	USA
HuaLUO_CHN	24	CHN	BB	nonUSA
UraASH_JPN	15	JPN	BB	nonUSA
YusOU___CHN	12	CHN	BB	nonUSA
ShoMIY_JPN	12	JPN	BB	nonUSA
RebAND_BRA	11	BRA	BB	nonUSA
EmmMAL_GER	11	GER	BB	nonUSA
SimBIL_USA	186	USA	FX	USA
RebAND_BRA	38	BRA	FX	nonUSA
KalLIN_USA	30	USA	FX	USA
FlaSAR_BRA	27	BRA	FX	nonUSA
JadCAR_USA	22	USA	FX	USA
KonMCC_USA	17	USA	FX	USA
JesGAD_GBR	14	GBR	FX	nonUSA
JorCHI_USA	12	USA	FX	USA
MarMAG_ITA	11	ITA	FX	nonUSA
SabMAN_ROU	11	ROU	FX	nonUSA
YusOU___CHN	11	CHN	FX	nonUSA
KayNEM_ALG	82	ALG	UB	nonUSA
QiyQIU_CHN	51	CHN	UB	nonUSA
ZoeMIL_USA	35	USA	UB	USA
XijTAN_CHN	33	CHN	UB	nonUSA
ShiJON_USA	32	USA	UB	USA
XiaWEI_CHN	30	CHN	UB	nonUSA
AliD A_ITA	28	ITA	UB	nonUSA
RebAND_BRA	22	BRA	UB	nonUSA
YunLEE_KOR	19	KOR	UB	nonUSA
RebDOW_GBR	19	GBR	UB	nonUSA
SimBIL_USA	151	USA	VT	USA
RebAND_BRA	103	BRA	VT	nonUSA
JadCAR_USA	61	USA	VT	USA
KonMCC_USA	40	USA	VT	USA
JorCHI_USA	28	USA	VT	USA
ShoMIY_JPN	26	JPN	VT	nonUSA
ShiJON_USA	22	USA	VT	USA
TiaSUM_USA	18	USA	VT	USA
SkyBLA_USA	18	USA	VT	USA
JosROB_USA	15	USA	VT	USA

Image 10)

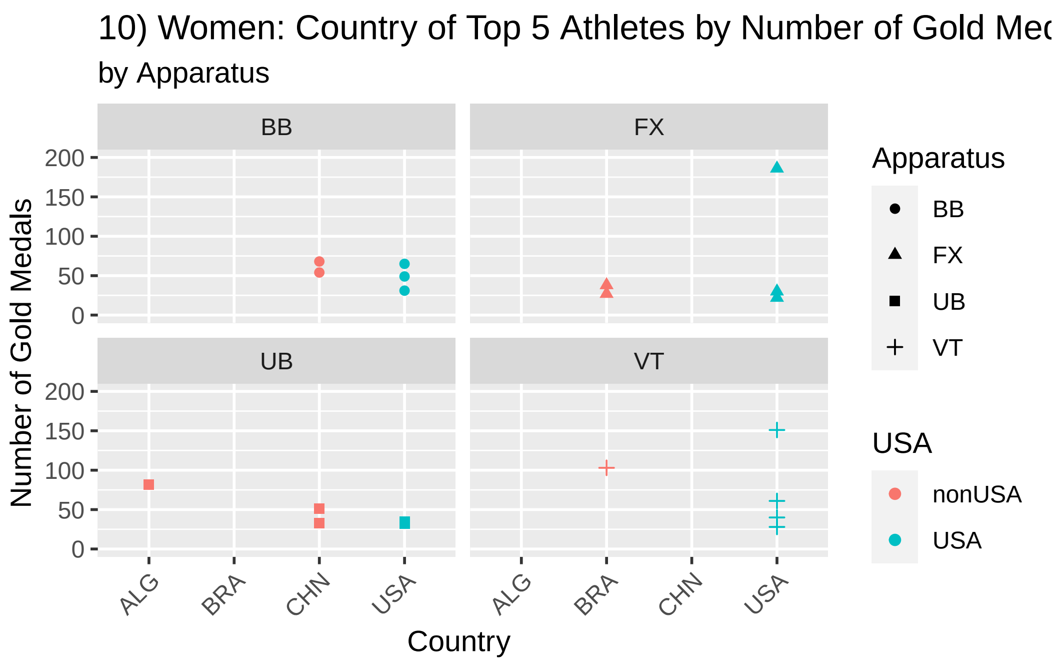
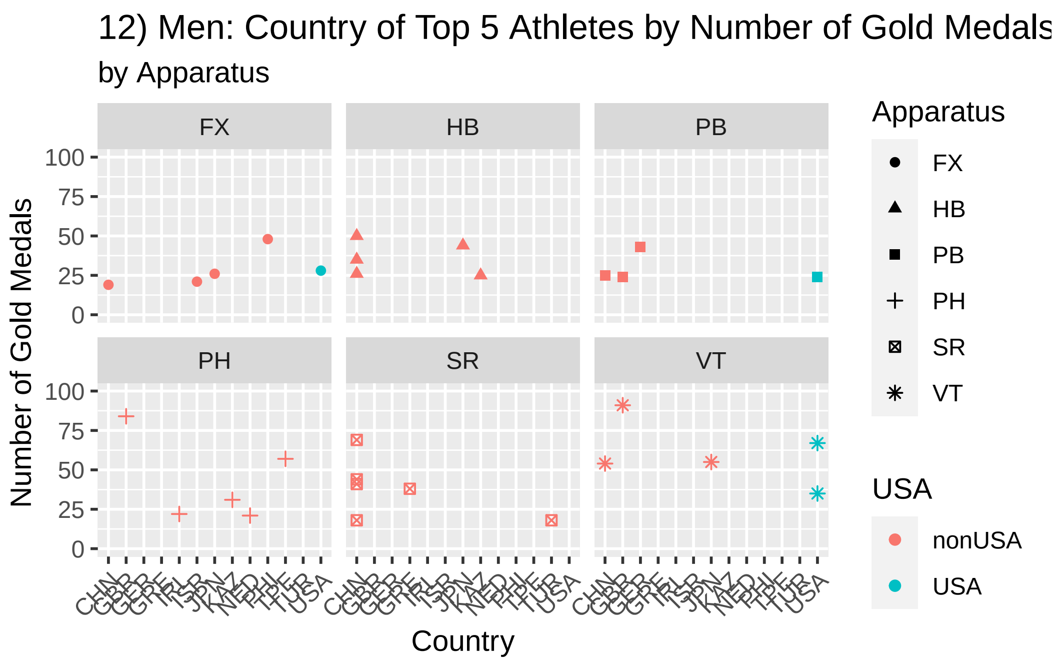


Image 11) Top ten most successful male gymnast using total gold medal count by apparatus

unique_id	Golds	Country	Apparatus	USA
CarYUL_PHI	48	PHI	FX	nonUSA
PauJUD_USA	28	USA	FX	USA
RyoDOI_JPN	26	JPN	FX	nonUSA
ArtDOL_ISR	21	ISR	FX	nonUSA
BohZHA_CHN	19	CHN	FX	nonUSA
DaiHAS_JPN	16	JPN	FX	nonUSA
DonWHI_USA	16	USA	FX	USA
ColWAL_USA	16	USA	FX	USA
HanKIM_KOR	15	KOR	FX	nonUSA
WilEMA_CAN	15	CAN	FX	nonUSA
BohZHA_CHN	50	CHN	HB	nonUSA
DaiHAS_JPN	44	JPN	HB	nonUSA
ConSHI_CHN	35	CHN	HB	nonUSA
WeiSUN_CHN	26	CHN	HB	nonUSA
MilKAR_KAZ	25	KAZ	HB	nonUSA
WeiSU__CHN	23	CHN	HB	nonUSA
BroMAL_USA	18	USA	HB	USA
IliGEO_CYP	18	CYP	HB	nonUSA
ShoKAW_JPN	16	JPN	HB	nonUSA
FreRIC_USA	14	USA	HB	USA
ChaLIN_CHN	14	CHN	HB	nonUSA
JinZOU_CHN	133	CHN	PB	nonUSA
LukDAU_GER	43	GER	PB	nonUSA
BohZHA_CHN	25	CHN	PB	nonUSA
CurPHI_USA	24	USA	PB	USA
JoeFRA_GBR	24	GBR	PB	nonUSA
KaiSUG_JPN	23	JPN	PB	nonUSA
ColWAL_USA	20	USA	PB	USA
CarYUL_PHI	16	PHI	PB	nonUSA
ConSHI_CHN	15	CHN	PB	nonUSA
BlaSUN_USA	15	USA	PB	USA
MaxWHI_GBR	84	GBR	PH	nonUSA
ChiLEE_TPE	57	TPE	PH	nonUSA
NarKUR_KAZ	31	KAZ	PH	nonUSA
RhyMCC_IRL	22	IRL	PH	nonUSA
LorDE _NED	21	NED	PH	nonUSA
SteNED_USA	20	USA	PH	USA
AhmABU_JOR	15	JOR	PH	nonUSA
JamLEW_GBR	13	GBR	PH	nonUSA
DaiHAS_JPN	13	JPN	PH	nonUSA
DehYIN_CHN	13	CHN	PH	nonUSA
YanLIU_CHN	69	CHN	SR	nonUSA
JinZOU_CHN	44	CHN	SR	nonUSA
XinLAN_CHN	41	CHN	SR	nonUSA
ElePET_GRE	38	GRE	SR	nonUSA
HaoYOU_CHN	18	CHN	SR	nonUSA
IbrCOL_TUR	18	TUR	SR	nonUSA
BohZHA_CHN	17	CHN	SR	nonUSA
NikSIM_AZE	17	AZE	SR	nonUSA
DonWHI_USA	16	USA	SR	USA
MahAHM_IRI	15	IRI	SR	nonUSA
JakJAR_GBR	91	GBR	VT	nonUSA
AshHON_USA	67	USA	VT	USA
DaiHAS_JPN	55	JPN	VT	nonUSA

Image 12)



Note: The excessive number of countries display that there is not much overlap in the top 5 most gold medal decorated athletes on the men's team and therefore the lack of well-rounded gymnasts.

Image 13)

Average of total gold medal count ranking by apparatus of all US male gymnasts in ascending order (to find best all around gymnast for men's team)

unique_id	avgRank
DonWHI_USA	13.66667
AshHON_USA	14.00000
BroMAL_USA	14.00000
ColWAL_USA	14.33333
PauJUD_USA	15.50000
CurPHI_USA	15.66667
FreRIC_USA	16.00000
KhoYOU_USA	16.16667
ShaWIS_USA	16.83333
YulMOL_USA	17.00000
SteNED_USA	17.33333
BlaSUN_USA	17.50000
IanLAS_USA	17.50000
DalHAL_USA	17.66667
CamBOC_USA	18.00000
ConMCC_USA	18.00000
IanSKI_USA	18.16667
PatHOO_USA	18.50000
TayBUR_USA	18.50000
BraBRI_USA	18.66667
JosKAR_USA	18.66667
RilLOO_USA	18.66667
JavALF_USA	19.00000
EvaHYM_USA	19.16667
KamNEL_USA	19.16667
MicJAR_USA	19.33333
ZacNUN_USA	19.33333
AleDIA_USA	19.50000
MatCOR_USA	19.50000
TayCHR_USA	19.66667
AddFAT_USA	19.83333
AshSUL_USA	19.83333
JadCAR_USA	19.83333
JorCHI_USA	19.83333
JosROB_USA	19.83333
KalLIN_USA	19.83333
KatJON_USA	19.83333
KonMCC_USA	19.83333
LeaWON_USA	19.83333
MadJOH_USA	19.83333
NolMAT_USA	19.83333
ShiJON_USA	19.83333
SimBIL_USA	19.83333
SkyBLA_USA	19.83333
SunLEE_USA	19.83333
TiaSUM_USA	19.83333
ZoeMIL_USA	19.83333