

Case Study 2: Paris 2024 Team USA Olympics Gymnastics Team

Aimi Wen, Katie Nash, Nidhi Dhupati, Wendy Shi

Introduction

Background Information

The Olympic Games are an international multi-sport event held every four years, and are largely considered to be the most prestigious sporting competition in the world. These Games are therefore a popular spectator event across the world, with an average viewership of more than 15 million people in past years [1]. The 2024 Olympic Games will be held in Paris, France, and will include both a Men's and Women's Artistic Gymnastics event. Winning a medal in an Olympic gymnastics event is an honor for both the individual athlete and for the country that athlete represents; the medal brings a sense of national pride to citizens across the country and promotes the global recognition of the winning country. The USA is known for its men and women gymnastics teams, which have consistently earned medals in past Olympic games. Because of the high honor and importance of upholding the legacy of the United States in these competitions, it is important to carefully select the men and women that will represent our country as Team USA. The importance of and interest surrounding the Olympics provides a rich foundation for analytics and statistical interpretations of the athletes in the competitions, as well as predictions about who will represent Team USA in the many Olympic sports. There have already been analyses, published in major news outlets, about who Team USA will consist of for non-gymnastics sports in 2024 [2]. The objective of this project is to extend this analysis to the Artistic Gymnastics events, and to select the 5 athletes for each of the men's and women's events that will optimize performance for Team USA.

The format of the Olympic Artistic Gymnastics event includes a qualifying round and final round. Scores from qualifying rounds do not impact medal determinations, but rather decide which athletes advance to the finals. The final round, which is what determines medal placements, consists of team all-around, individual all-around, and individual apparatus events. Men's opportunities for medals are: 1) team all-around, 2) individual all-around, 3) floor exercise, 4) pommel horse, 5) still rings, 6) vault, 7) parallel bars, and 8) high bar. Women's opportunities for medals are: 1) team all-around, 2) individual all-around, 3) vault, 4) uneven bars, 5) balance beam, and 6) floor exercise. The athletes representing Team USA will compete in a pool of 96 men and 96 women, which will include teams of five from other countries as well as the option for individual entries for countries without full team qualifications.

Our hypothesis is:

- Who are the 5 men and 5 women who can make Team USA achieve the highest level of success at the Paris 2024 Artistic Gymnastics Event?

Data Description

Gymnastics competitions from 2017-2023 were included in the dataset. These competitions were split into two separate datasets by the seasons leading up to the 2021 Tokyo Games and the seasons leading up to the 2024 Paris Games (the 2022 and 2023 seasons). Competitions were included from around the globe, including world competitions, non-world competitions that included the United States as a competitor, and non-world competitions that did not include the United States as a competitor. The competition data included scoring information, athlete names, athlete demographic information, and athlete scores. Each entry in the dataset

is one event from one athlete, which includes the country affiliation of the athlete, the competition and competition location, the difficulty, execution, penalty, and overall scores, the apparatus that the athlete competed on, and their ranking. Additionally, the countries who qualified as a team, which are the countries that will be allowed to compete in the Team All-Around event for the men's and women's Olympic events, were provided. The dataset we used for analysis was last synced with the central repository on November 2nd.

EDA - separated by type of visualization, not gender

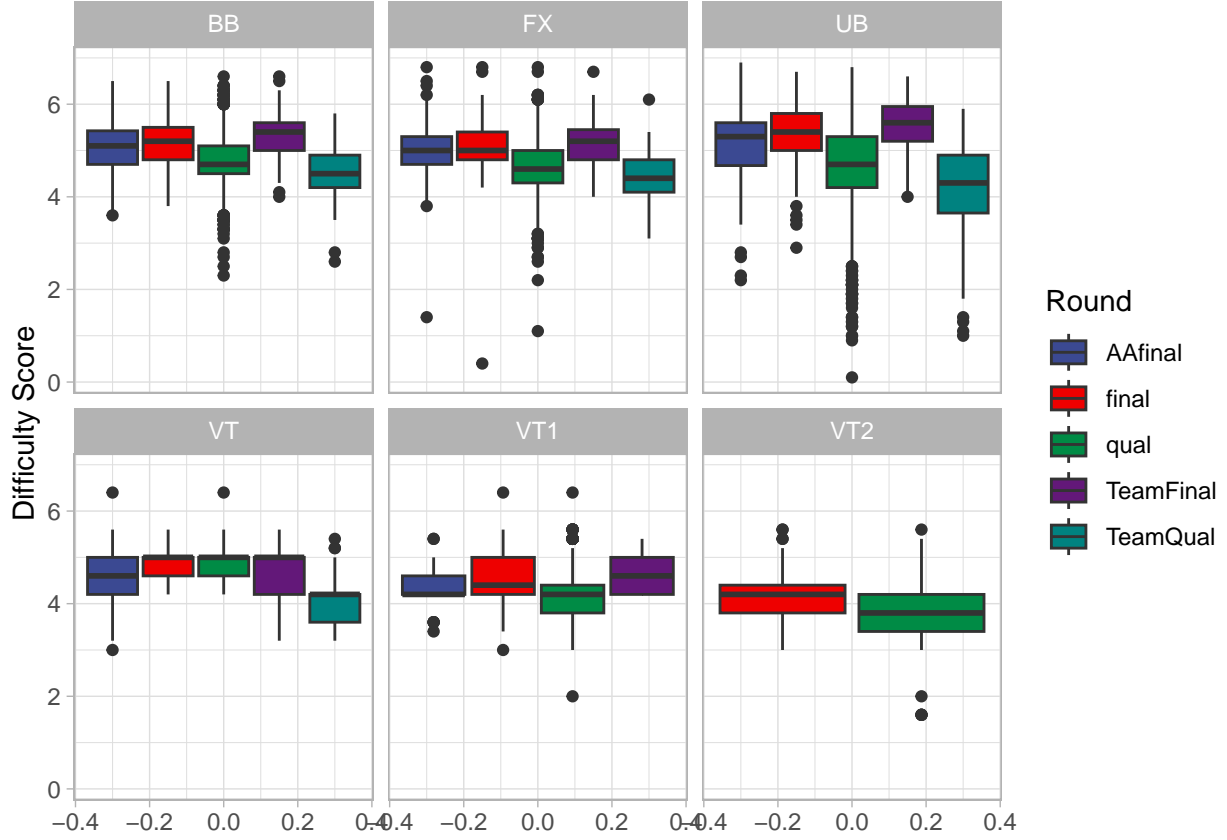


Figure 1: Distribution of the Difficulty Score by Round and Apparatus for Women

showing that most penalties are near 0: For both men's and women's competitions, the 95th percentile for the penalties is less than 0.5 (it is 0.5 for men and 0.4 for women). This means that 95% of penalties for both genders are not more than 0.5. The visualizations show that this does not change drastically based on the scope of the competition. Additionally, since the 51st percentile of penalties for both genders is 0.1, this means that the majority of entries had a penalty score of 0.1 or less.

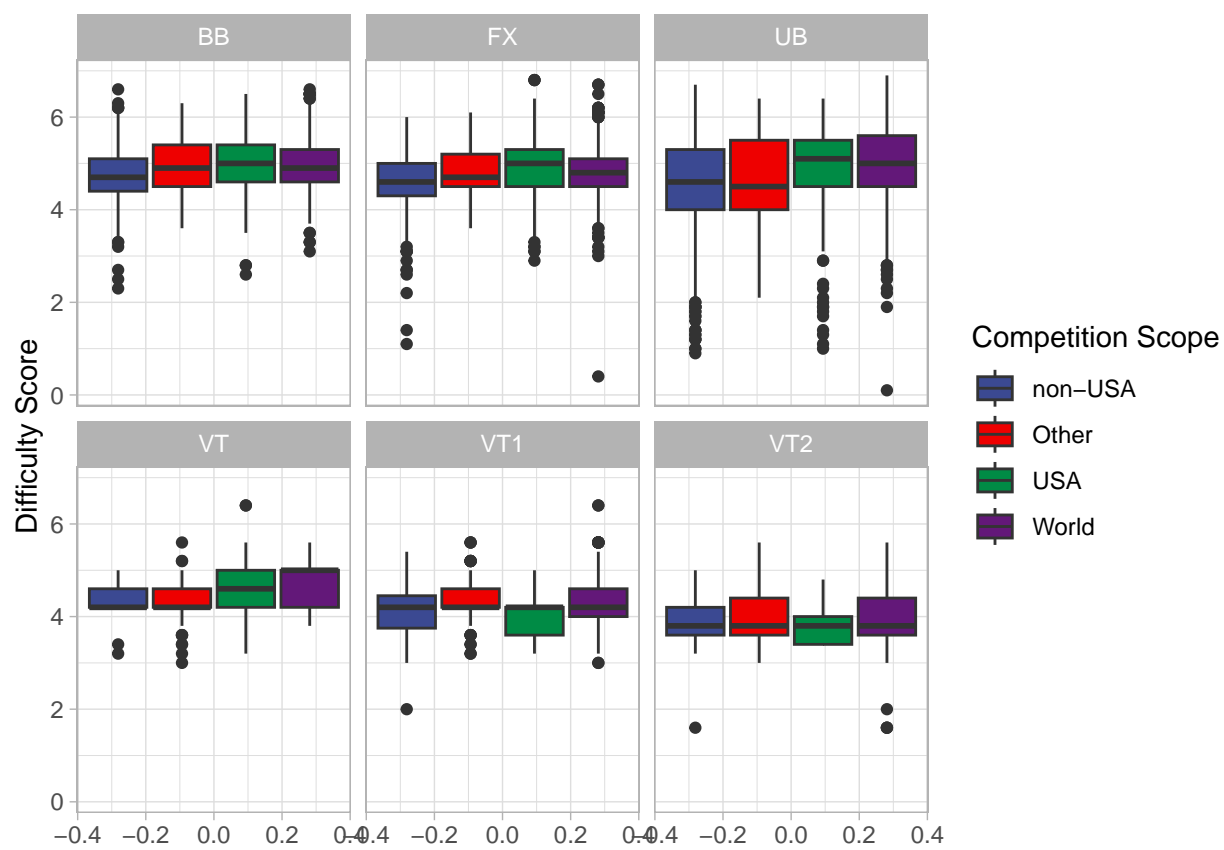


Figure 2: Distribution of Difficulty scores by the Scope of Competition and Apparatus for Women

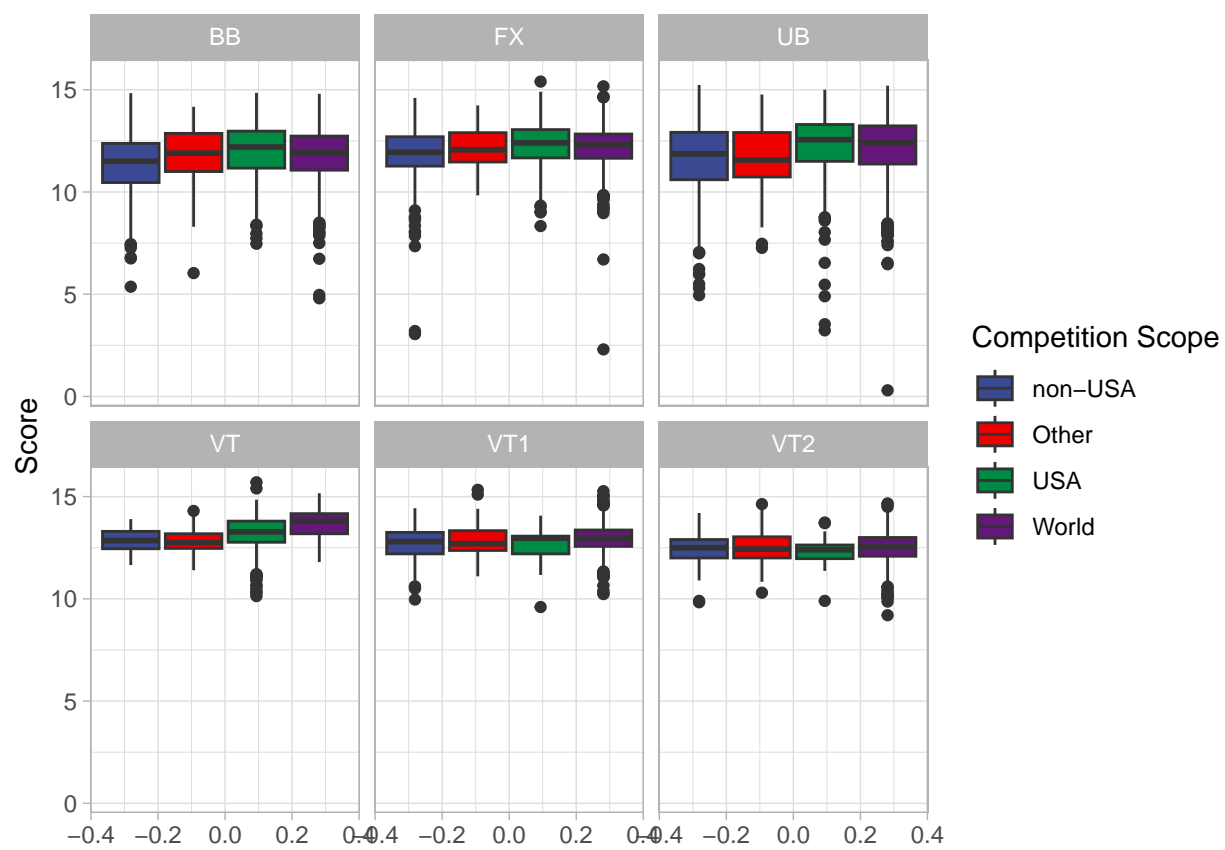
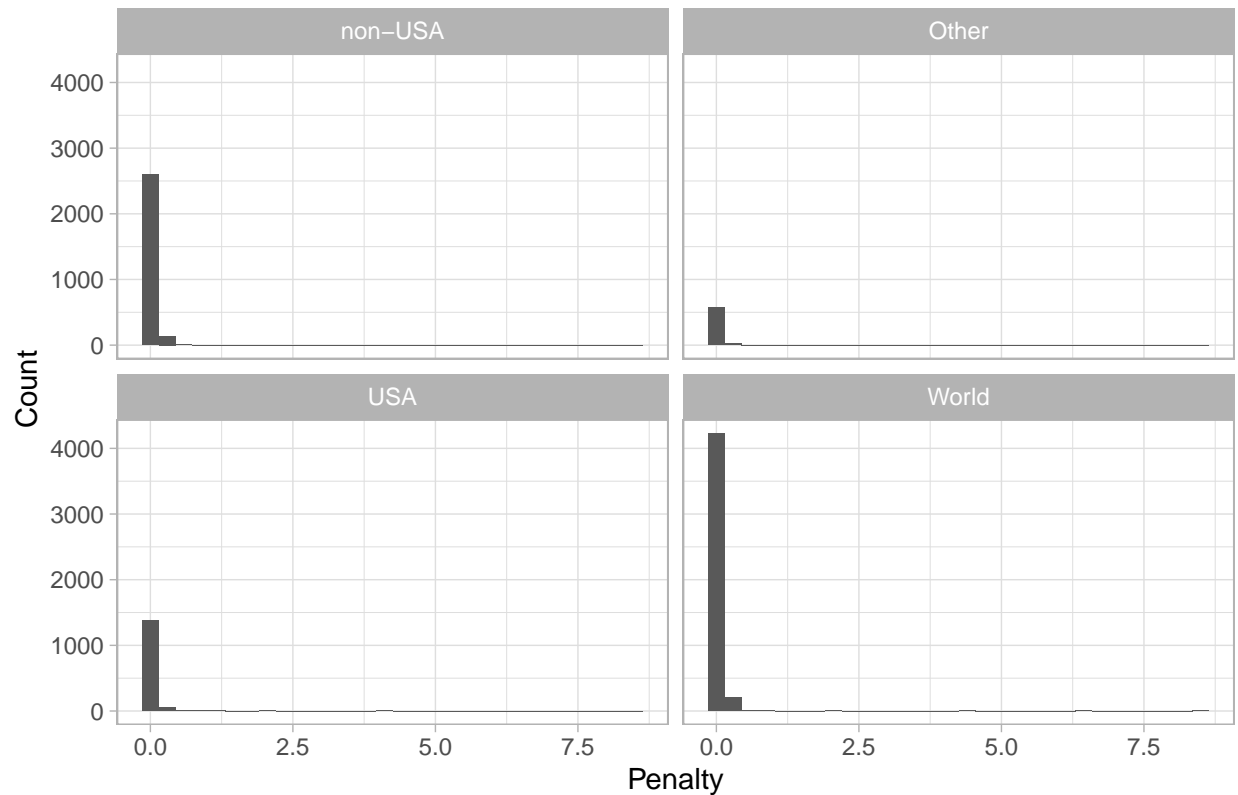


Figure 3: Distribution of Scores by Scope of Competition and Apparatus for Women

Penalty Distribution for Women Athletes



Why we're using mean for bootstrapping: The distribution of event scores for women and men are both approximately normally distributed, with the mean and median scores less than 0.5 apart within each gender. The mean of women's scores is 12.207, and the median of women's scores is 12.367. The mean of men's scores is 13.048, and the median of men's scores is 13.233.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Whether scores vary by type of round in competition: Doesn't seem to vary much; make the assumption that qualification scores are representative of an athlete's scores in the medal rounds

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

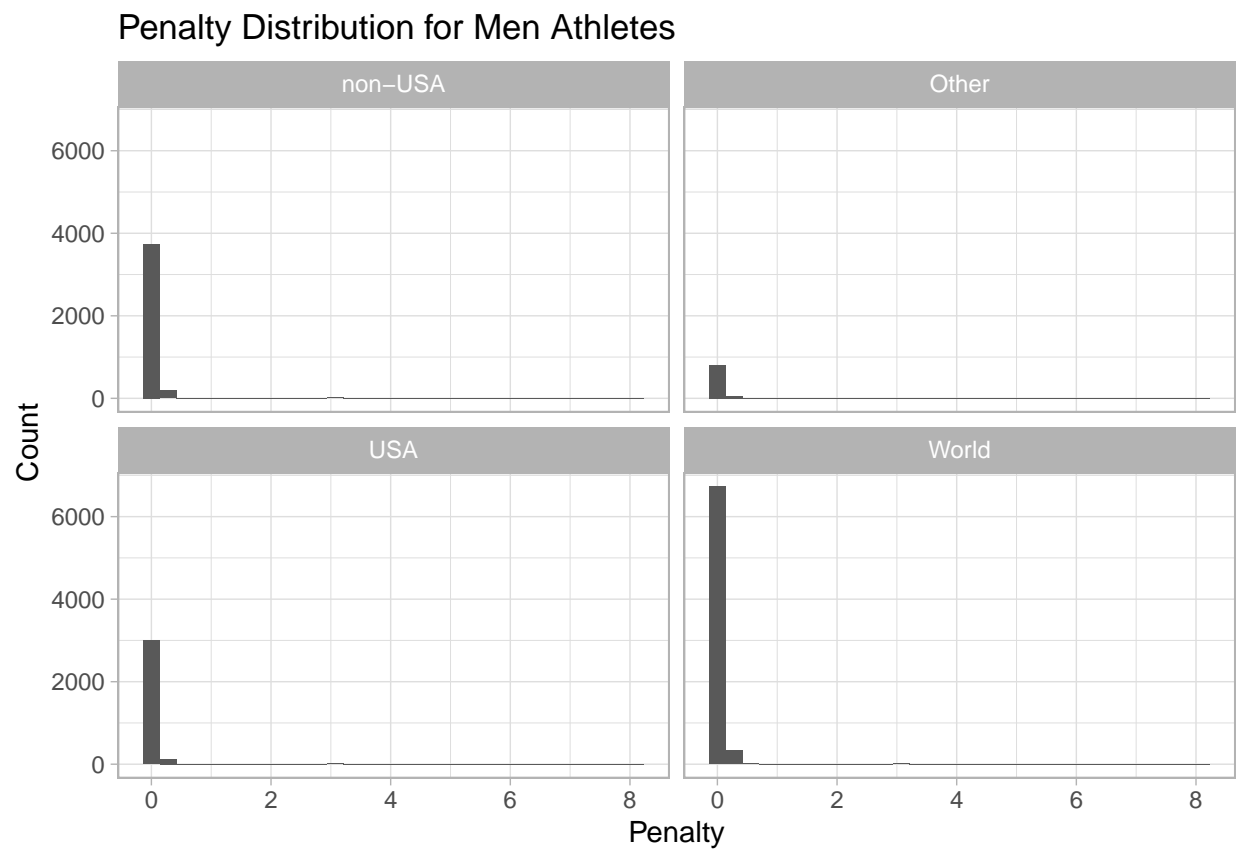


Figure 4: Distribution of Penalties by competition scope for Men

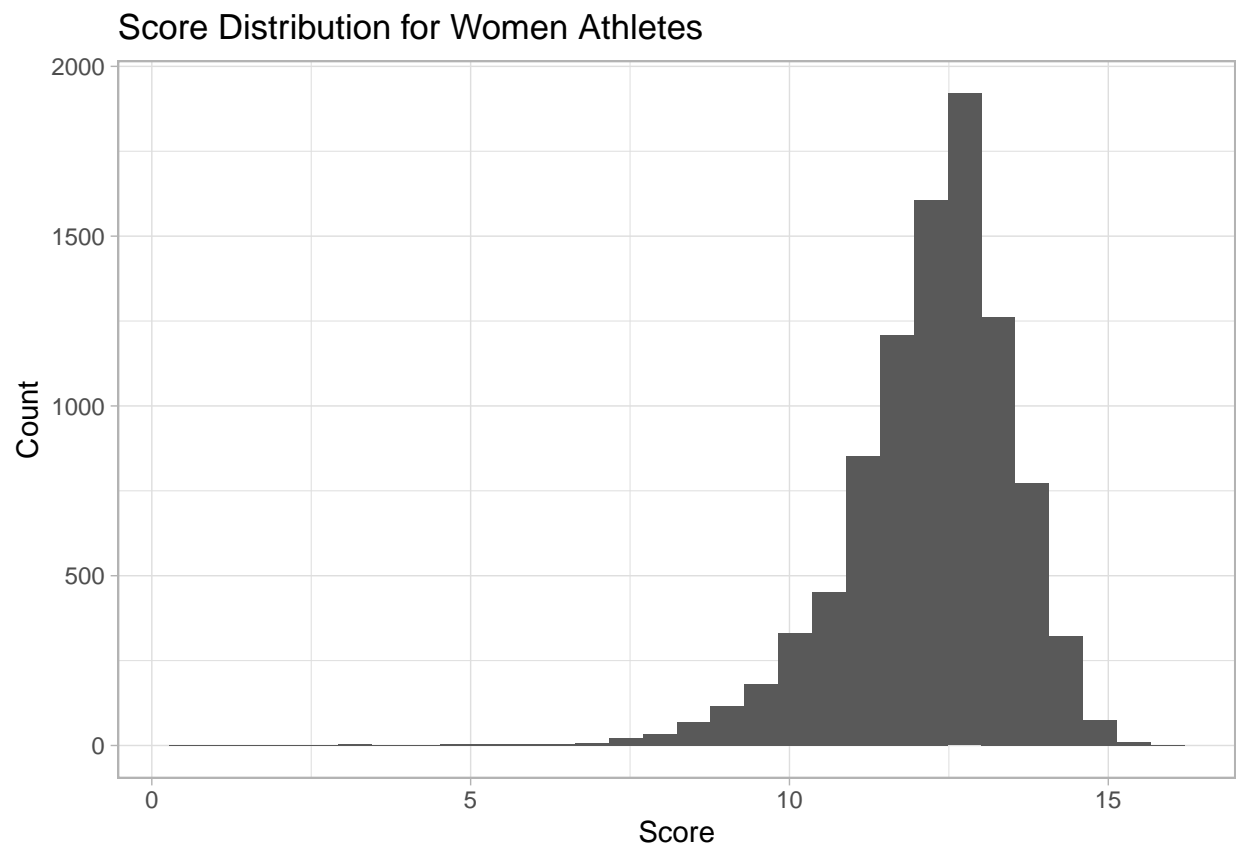


Figure 5: Score Distribution for Women Athletes

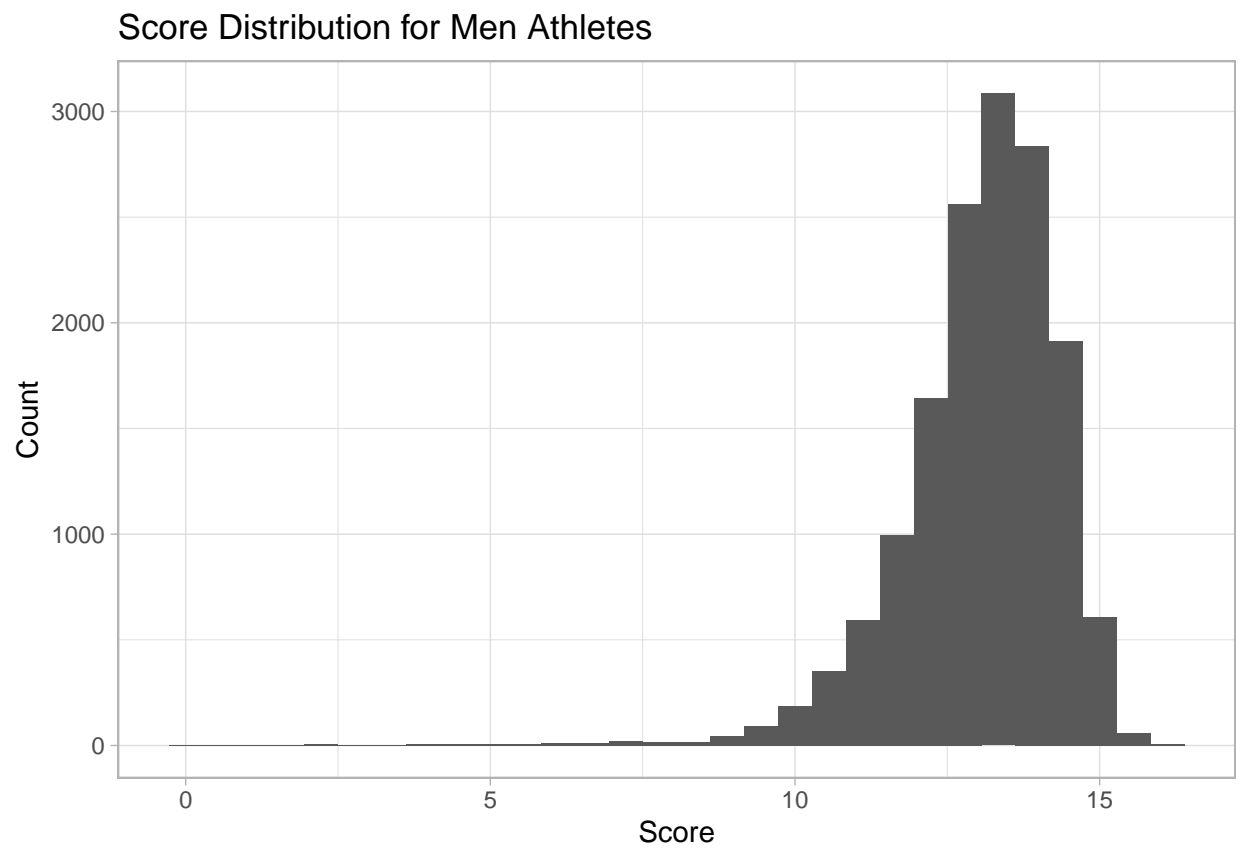
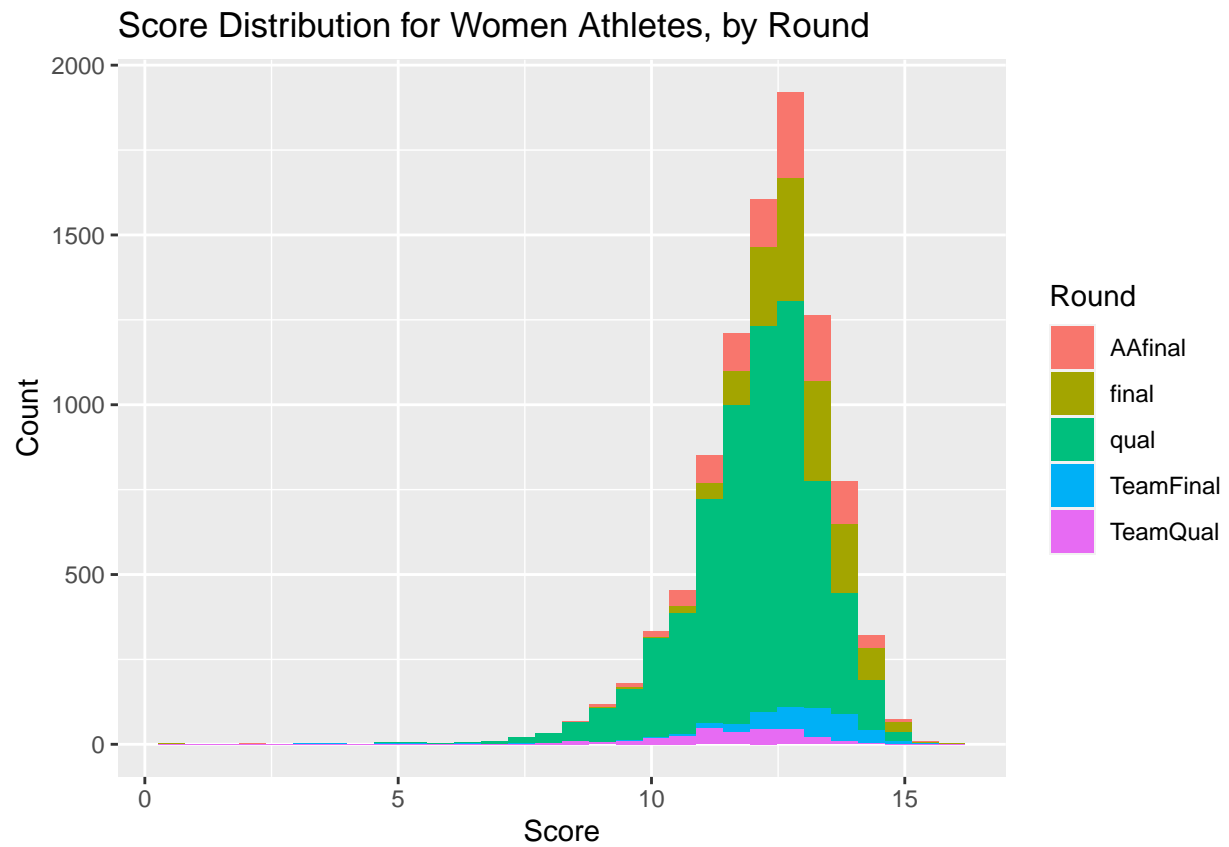
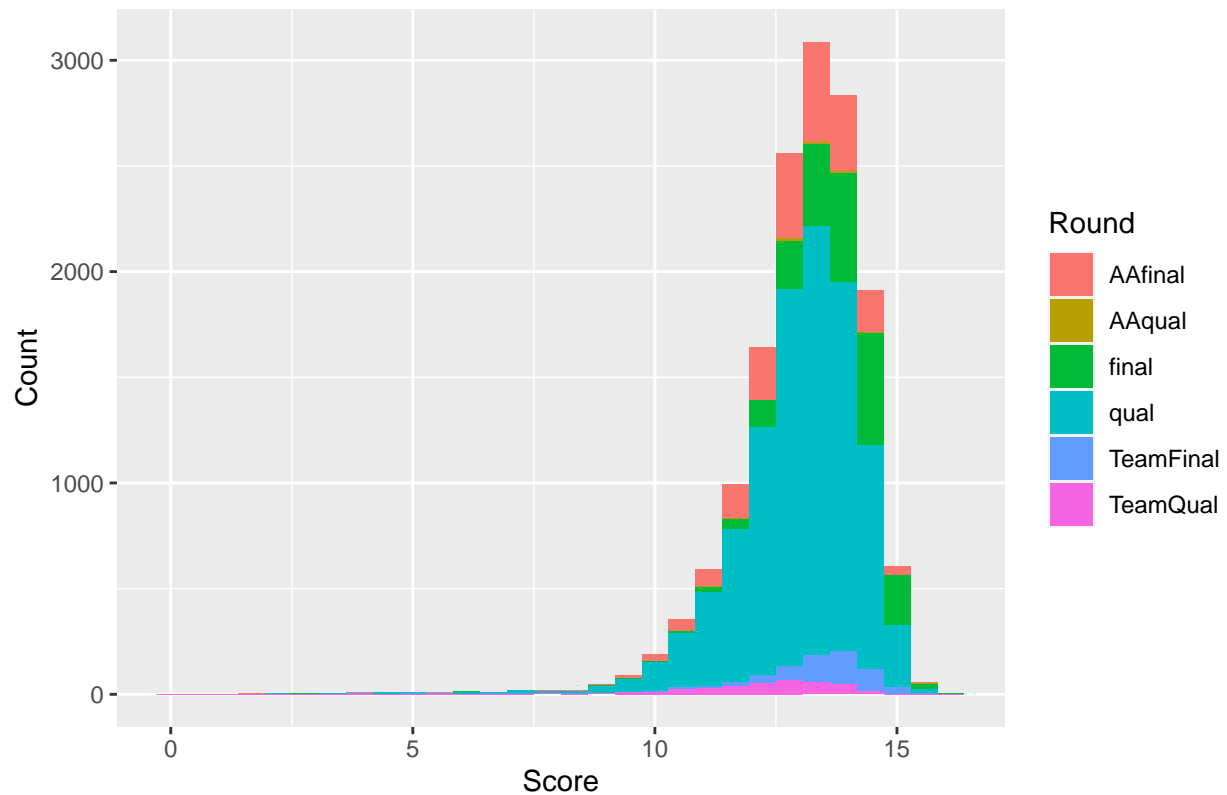


Figure 6: Score Distribution for Male Athletes



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Score Distribution for Men Athletes, by Round



```
## [1] 12.21275
```

```
## [1] 12.4
```

```
## [1] 13.04569
```

```
## [1] 13.233
```

Execution scores have a higher range than difficulty scores, for both men and women athletes (colored by round but doesn't seem to be much difference based on round):

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

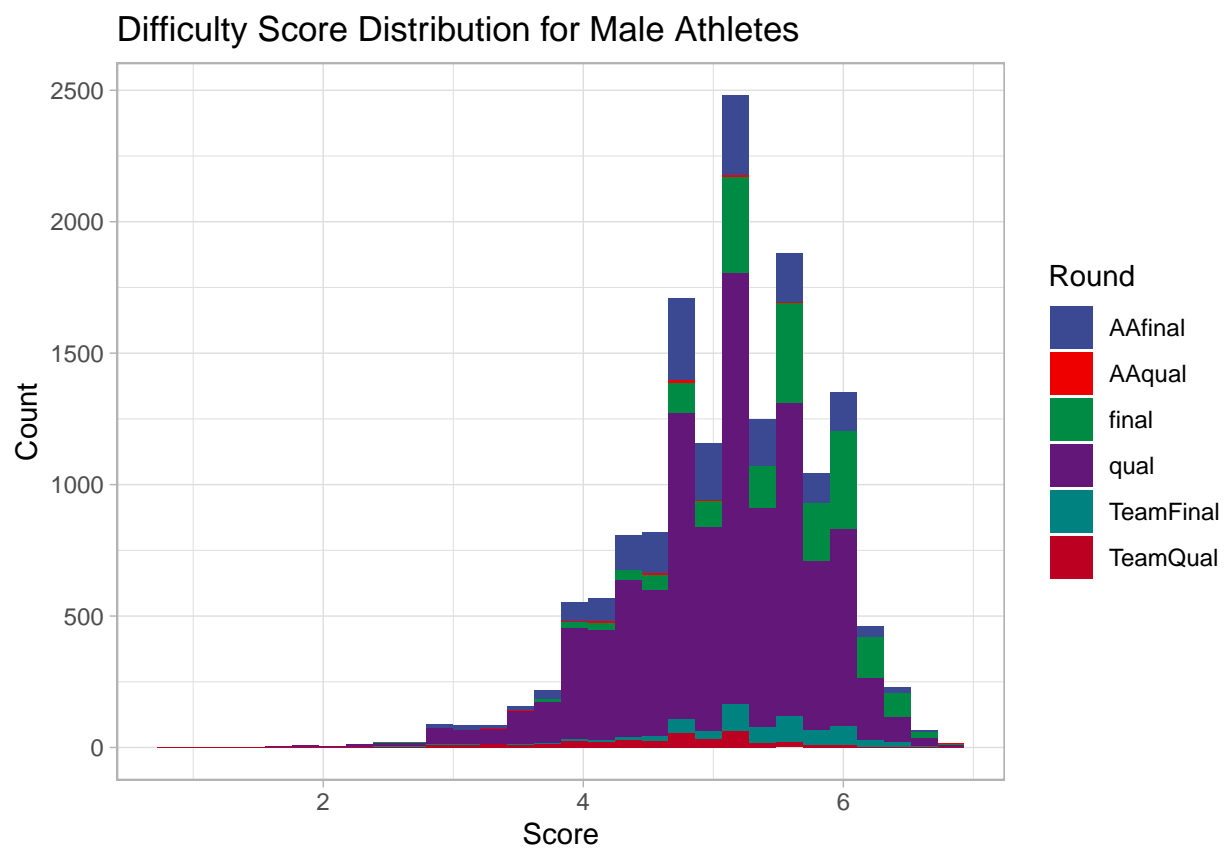


Figure 7: Difficulty Score Distribution for Male Athletes

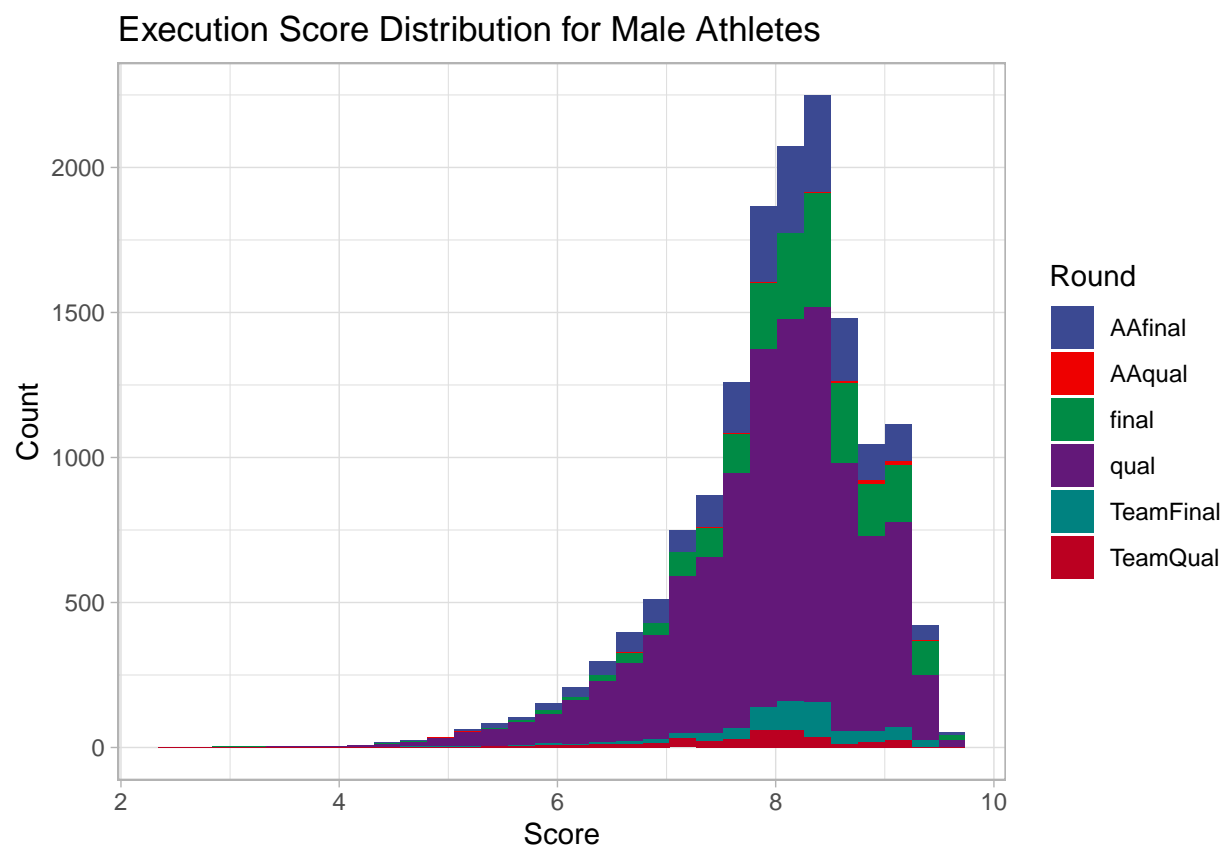
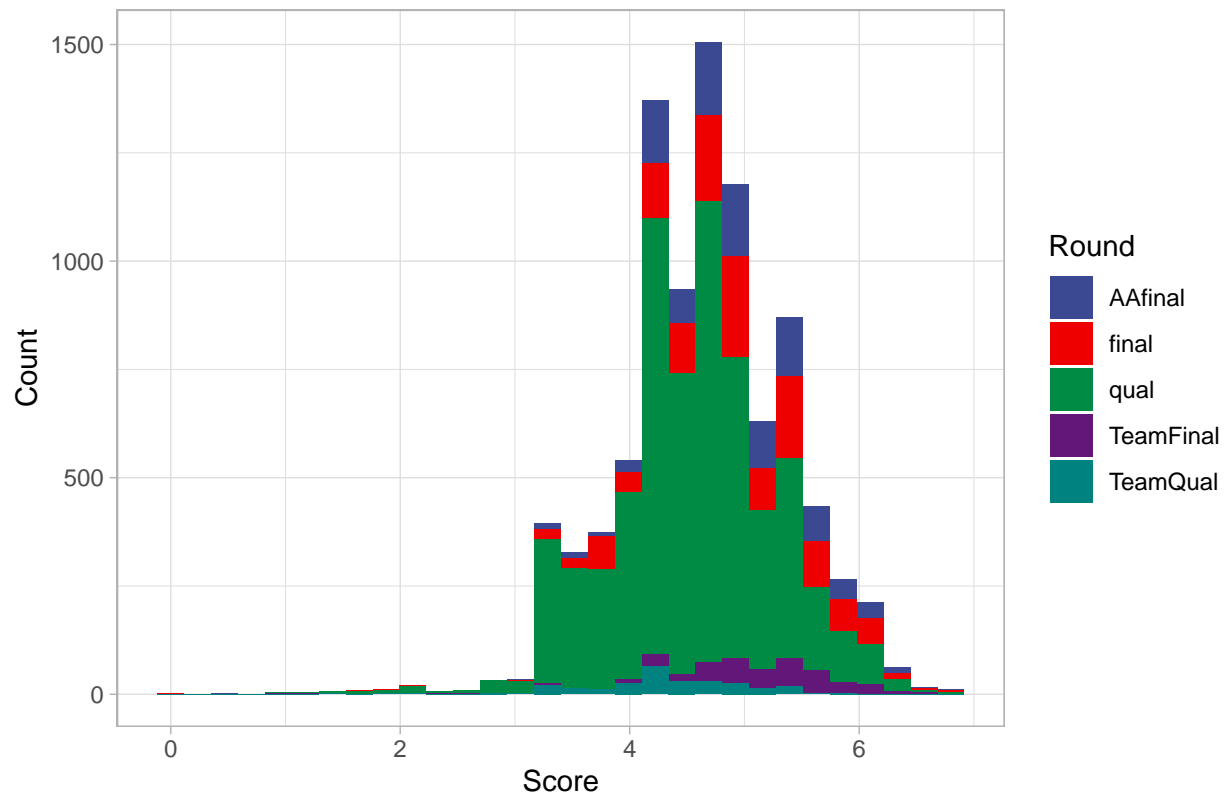


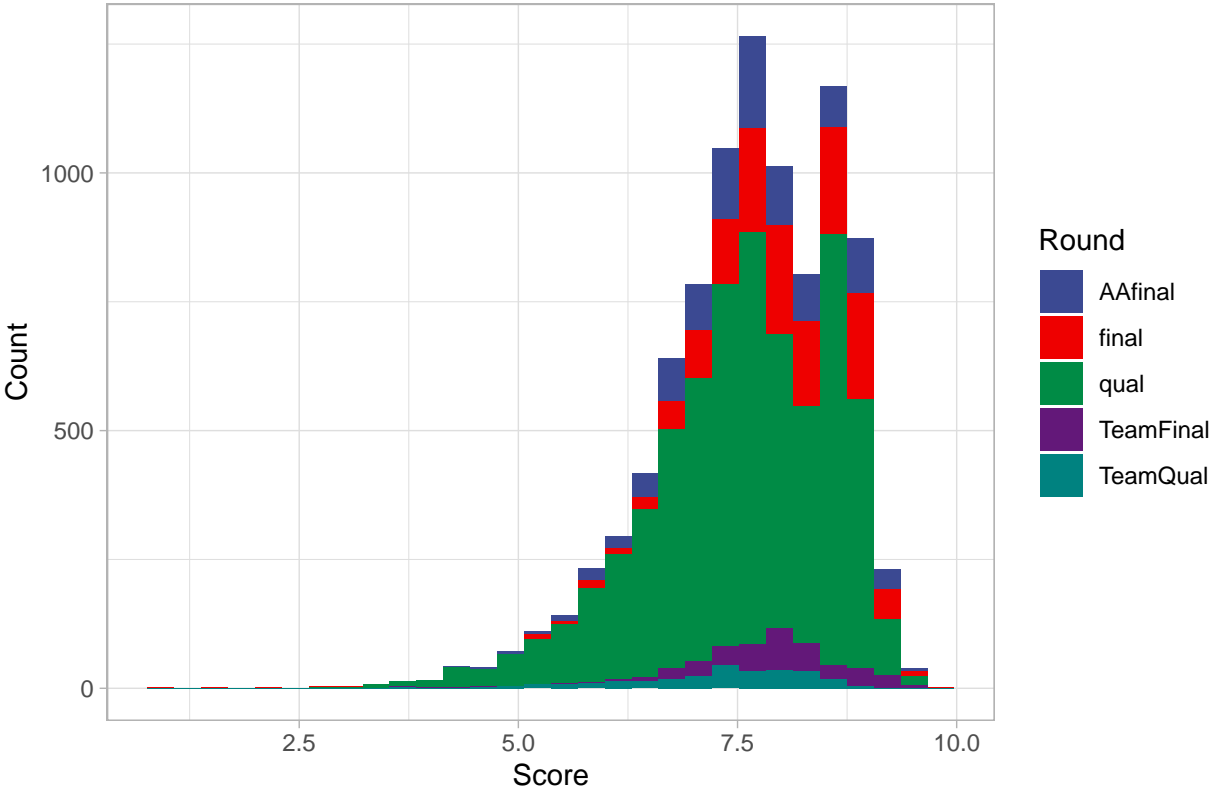
Figure 8: Execution Score Distribution for Male Athletes

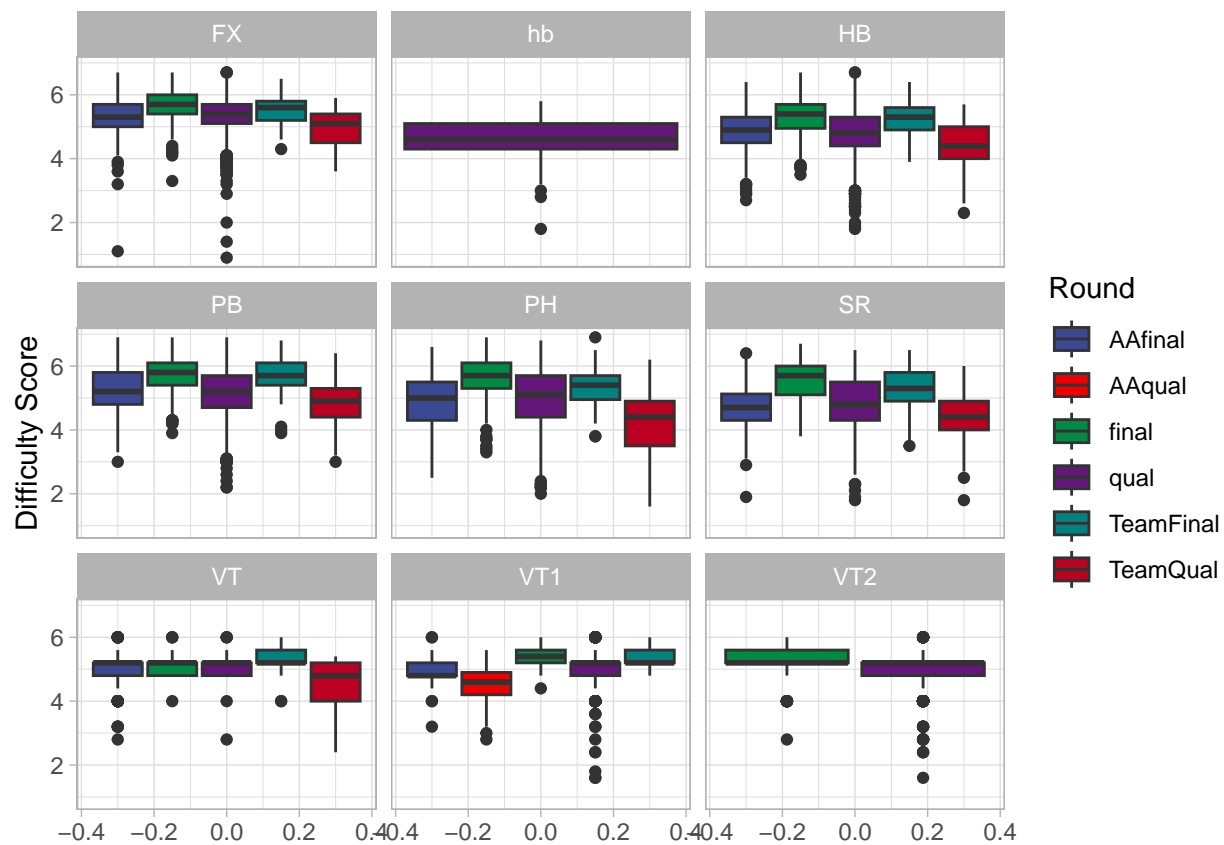
Difficulty Score Distribution for Women Athletes

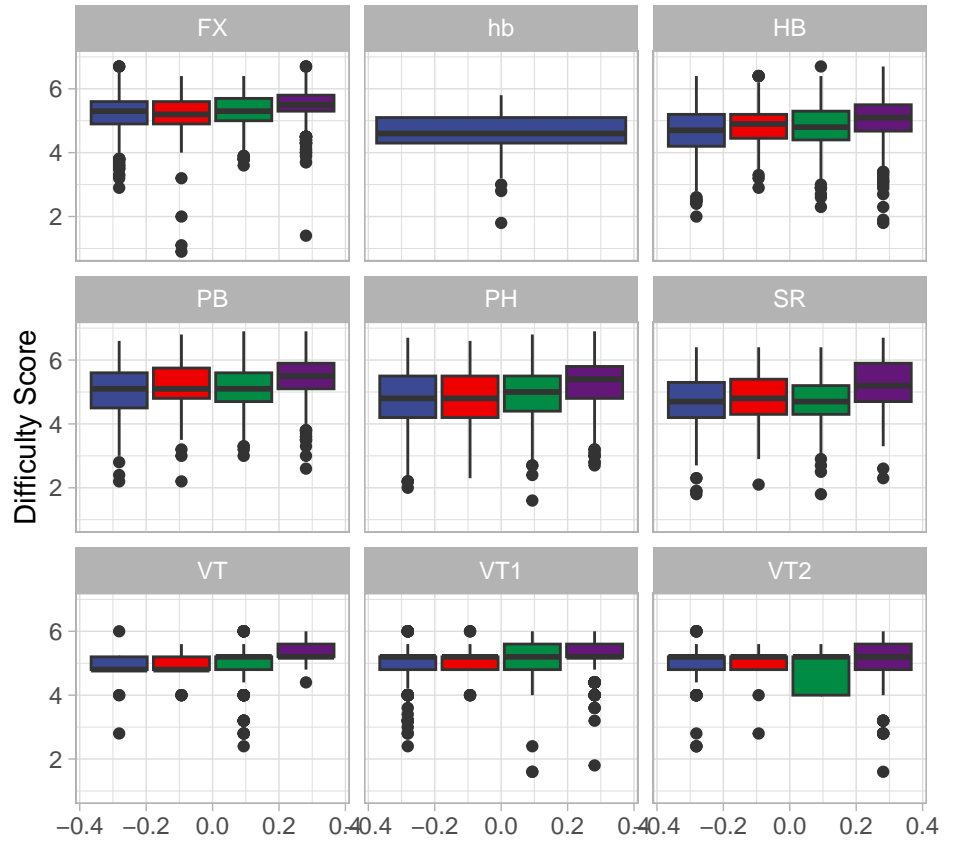


```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Execution Score Distribution for Women Athletes







Waiting for competition scope for men

Citations

[1] <https://www.cnbc.com/2021/08/04/simone-biles-return-helped-olympics-viewership-average-16point8-million.html> [2] <https://www.cbssports.com/nba/news/2024-paris-olympics-projecting-team-usas-12-man-roster-with-joel-embiid-joining-lebron-james-stephen-curry/>

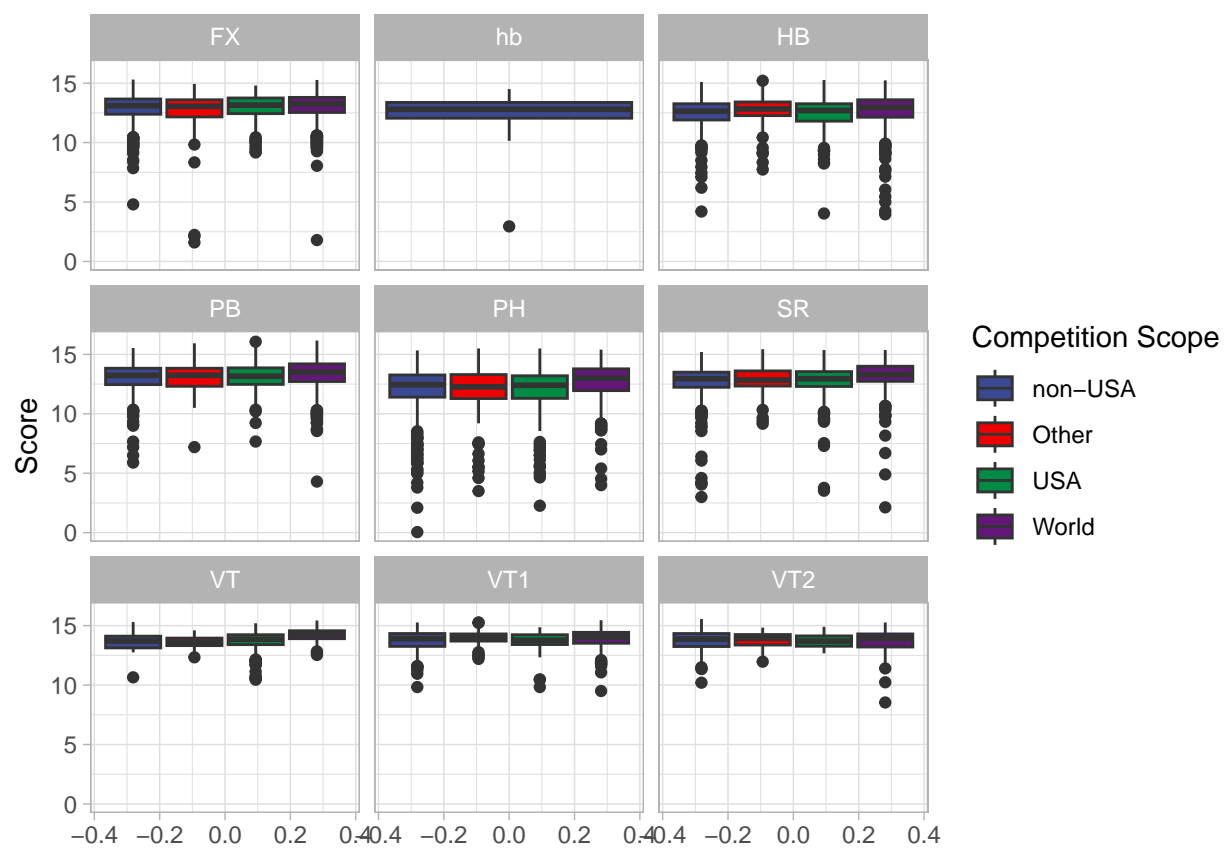


Figure 9: Distribution of Scores by Scope of Competition and Apparatus for Men