Main

Wendy

2023-10-30

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
##
## Attaching package: 'vroom'
##
##
## The following objects are masked from 'package:readr':
##
##     as.col_spec, col_character, col_date, col_datetime, col_double,
##     col_factor, col_guess, col_integer, col_logical, col_number,
##     col_skip, col_time, cols, cols_condense, cols_only, date_names,
##     date_names_lang, date_names_langs, default_locale, fwf_cols,
##     fwf_empty, fwf_positions, fwf_widths, locale, output_column,
##     problems, spec
##
##
## Rows: 24387 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (9): LastName, FirstName, Gender, Country, Date, Competition, Round, Loc...
## dbl (5): Rank, D_Score, E_Score, Penalty, Score
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
##  [1] "2022 51st FIG Artistic Gymnastics World Championships"
##  [2] "2022 Cairo World Cup"
##  [3] "2022 Paris World Challenge Cup"
##  [4] "2023 52nd FIG Artistic Gymnastics World Championships"
##  [5] "2023 Baku World Cup"
##  [6] "2023 Cairo World Cup"
##  [7] "2023 Doha World Cup"
##  [8] "2022 British Gymnastics Championships"
##  [9] "BIRMINGHAM 2022 Commonwealth Games"
## [10] "2023 Artistic Gymnastics Senior Pan American Championships"
```

```
## [11] "2022 Senior European Championships"
## [12] "2023 British Gymnastics Championships"
## [13] "2023 Senior European Championships"
## [14] "2022 Osijek World Challenge Cup"
## [15] "SANTIAGO 2023 XIX Pan American Games"
## [16] "2022 U.S. Championships"
## [17] "2022 U.S. Classic"
## [18] "2022 Winter Cup"
## [19] "2022 9th Senior Artistic Gymnastics Asian Championships"
## [20] "2023 10th Senior Artistic Gymnastics Asian Championships"
## [21] "2023 Varna World Challenge Cup"
## [22] "HANGZHOU 2022 19th Asian Games"
## [23] "2022 Mersin World Challenge Cup"
## [24] "2023 Tel Aviv World Challenge Cup"
## [25] "2023 Core Hydration Classic"
## [26] "2023 FISU World University Games"
## [27] "2023 Cottbus World Cup"
## [28] "2022 Baku World Cup"
## [29] "2022 Varna World Challenge Cup"
## [30] "2023 Central American and Caribbean Games"
## [31] "2022 Doha World Cup"
## [32] "2023 Osijek World Challenge Cup"
## [33] "EnBW DTB Pokal Team Challenge 2023"
## [34] "2022 Cottbus World Cup"
## [35] "2022 Koper World Challenge Cup"
## [36] "2022 Szombathely World Challenge Cup"
## [37] "2023 U.S. Championships"
## [38] "2023 Winter Cup"
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `competition_number = as.numeric(competition_number)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `competition_number = as.numeric(competition_number)`.
## Caused by warning:
## ! NAs introduced by coercion
```

Introduction (WORK IN PROGRESS - Nidhi)

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 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.

The objective of this project is to select the 5 athletes for each of men's and women's events that will optimize performance for Team USA across the 8 medal events for men and the 6 medal events for women.

Data Description

EDA - separated by type of visualization, not gender

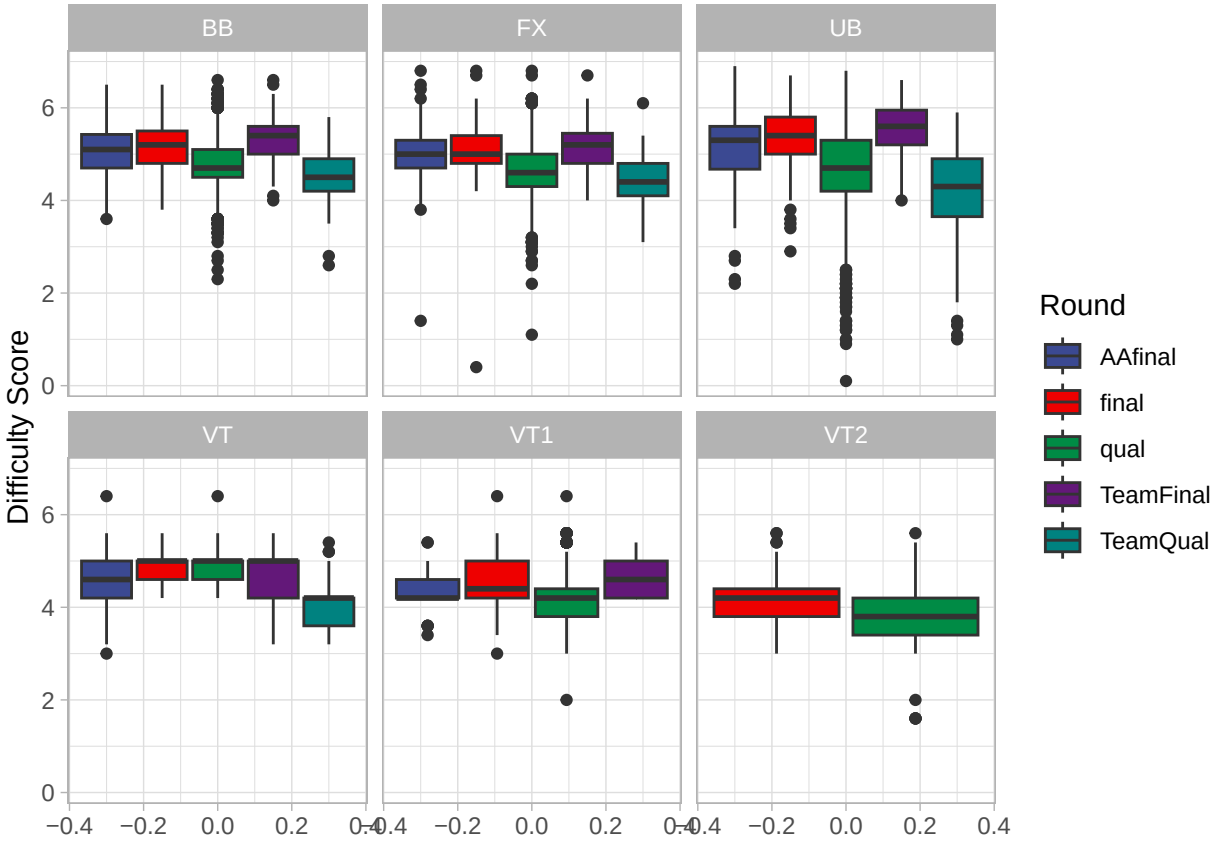


Figure 1: Distribution of the Difficulty Score by Round and Apparatus for Women

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.

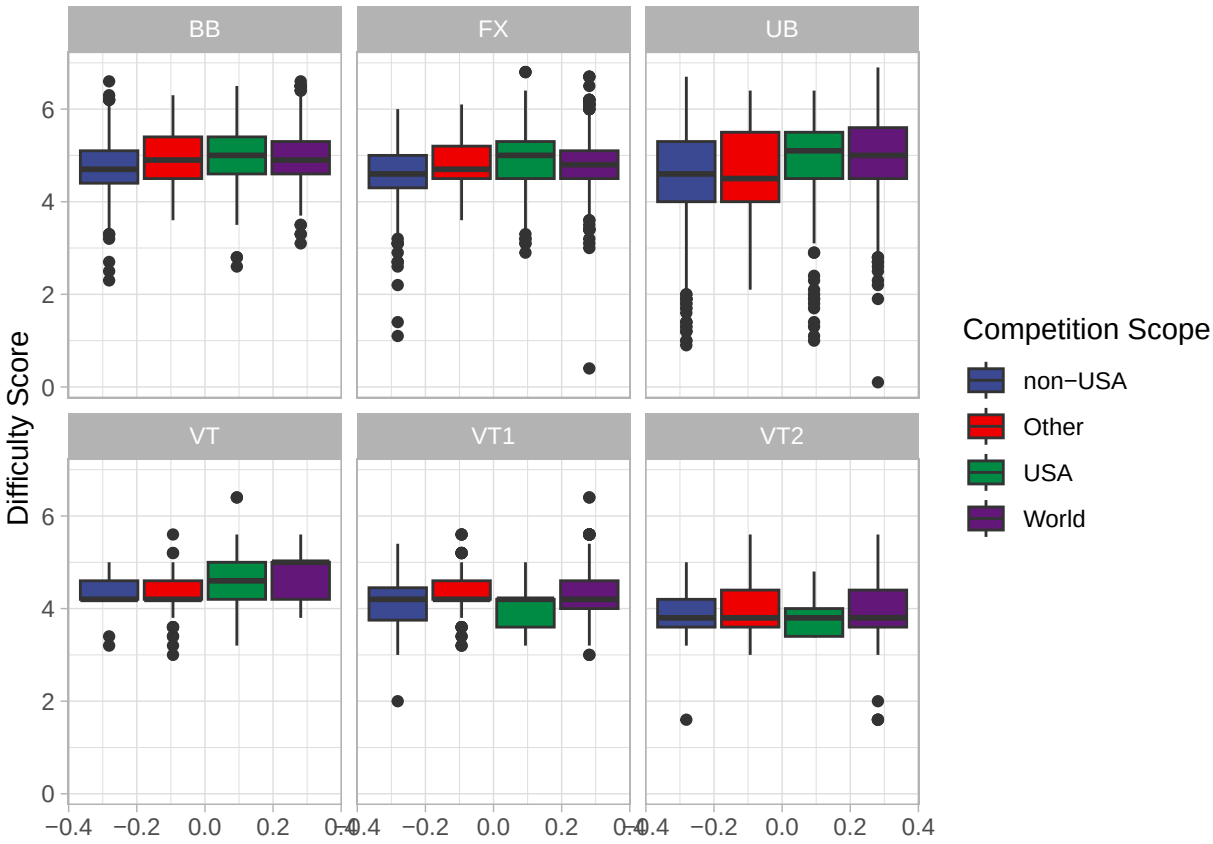


Figure 2: Distribution of Difficulty scores by the Scope of Competition and Apparatus for Women

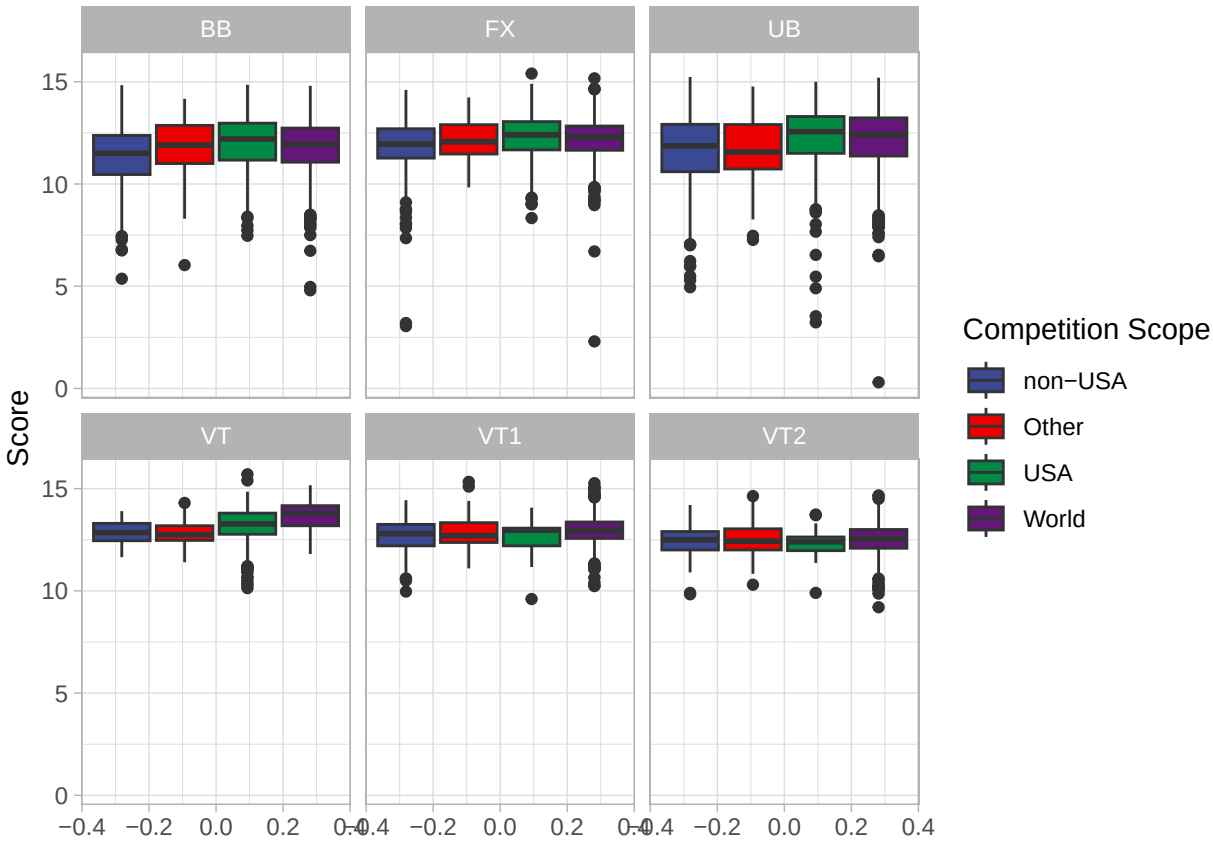
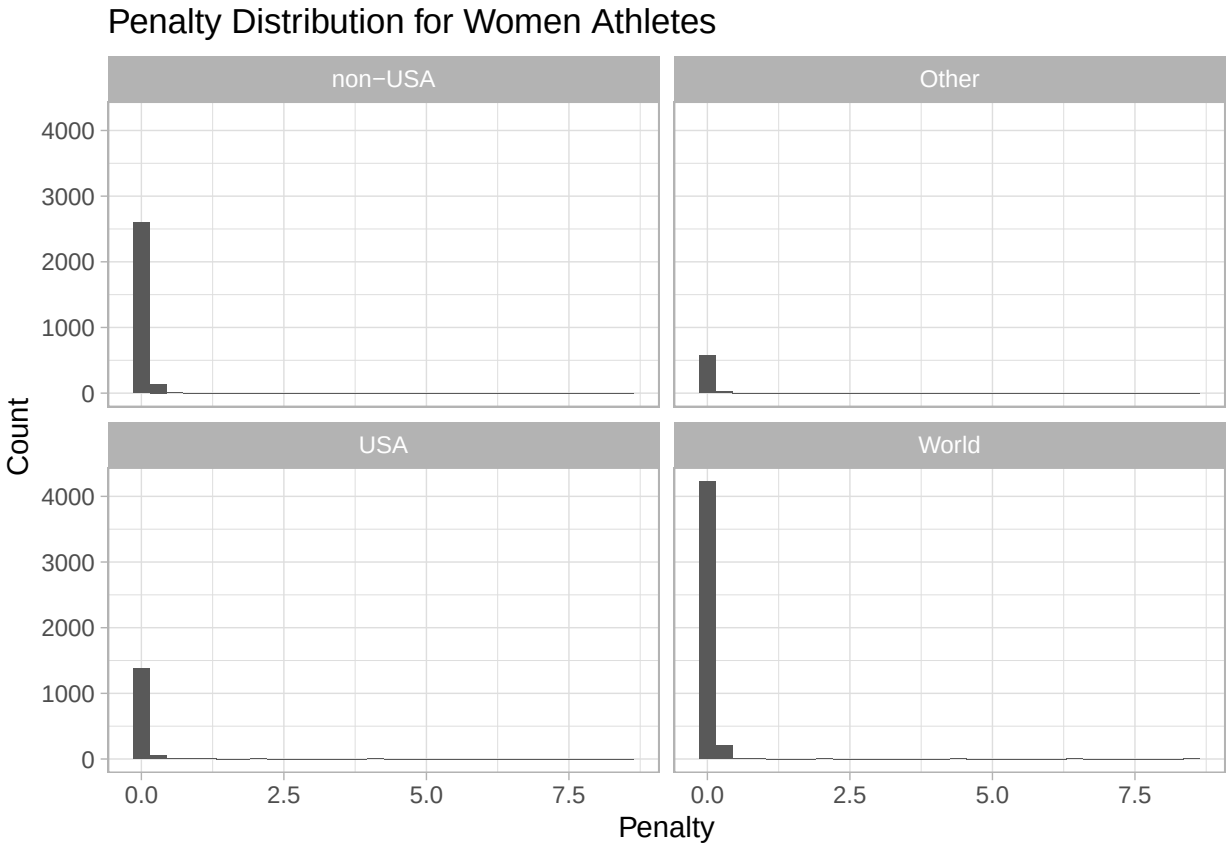


Figure 3: Distribution of Scores by Scope of Competition and Apparatus for Women



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`.
```

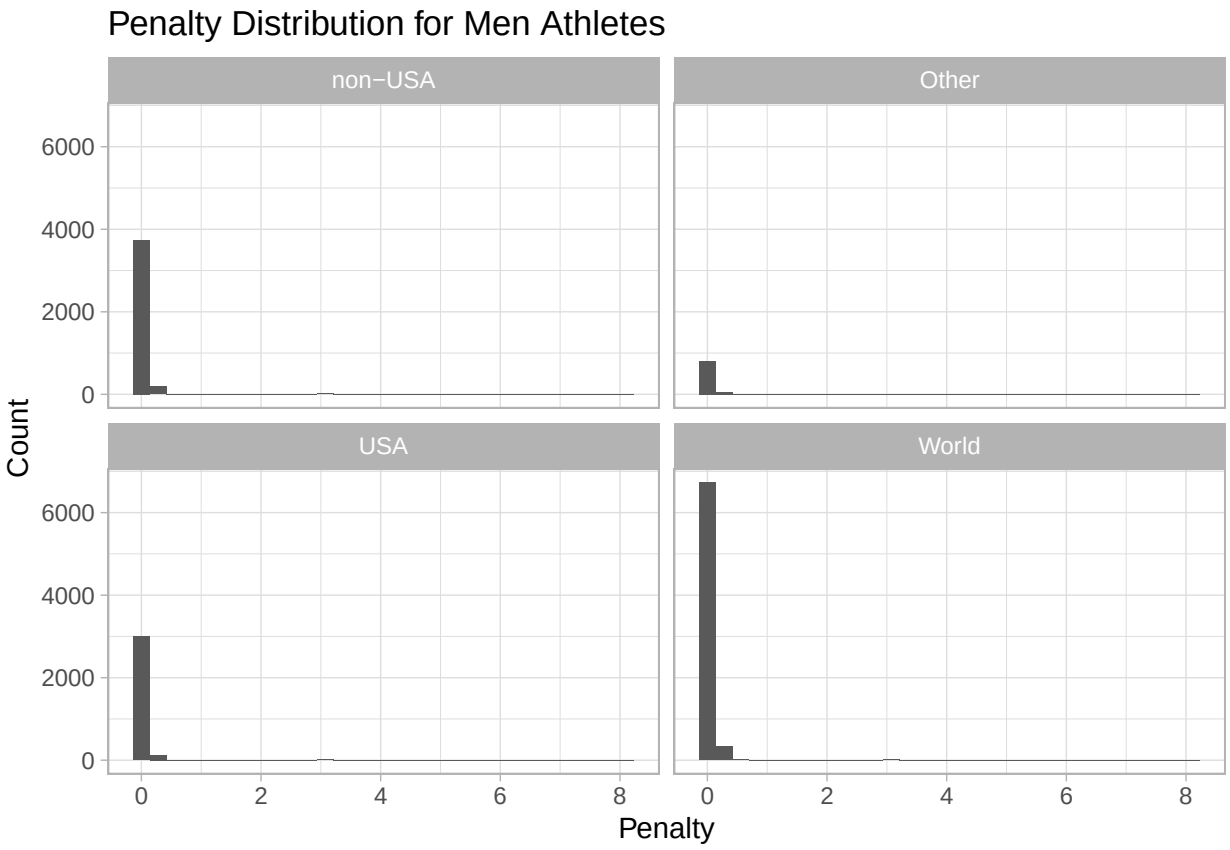


Figure 4: Distribution of Penalties by competition scope for Men

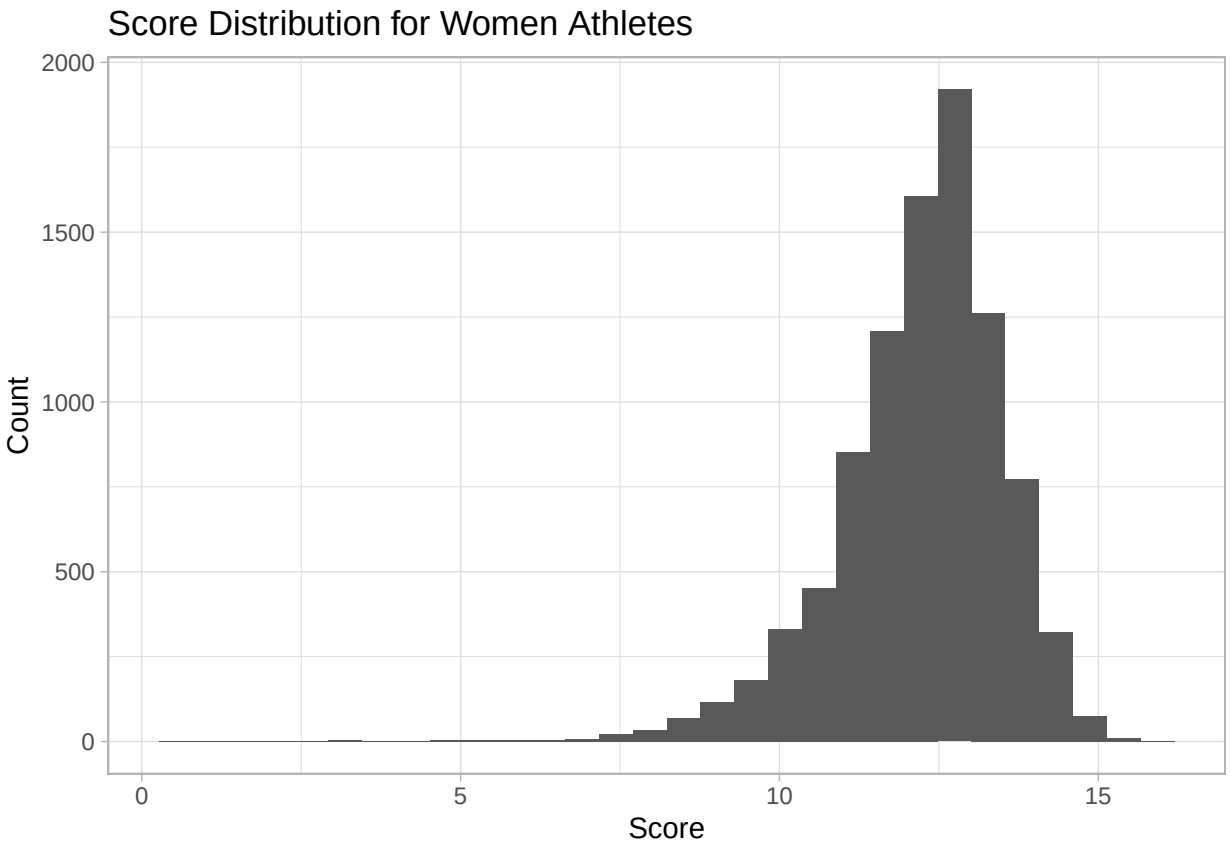


Figure 5: Score Distribution for Women Athletes

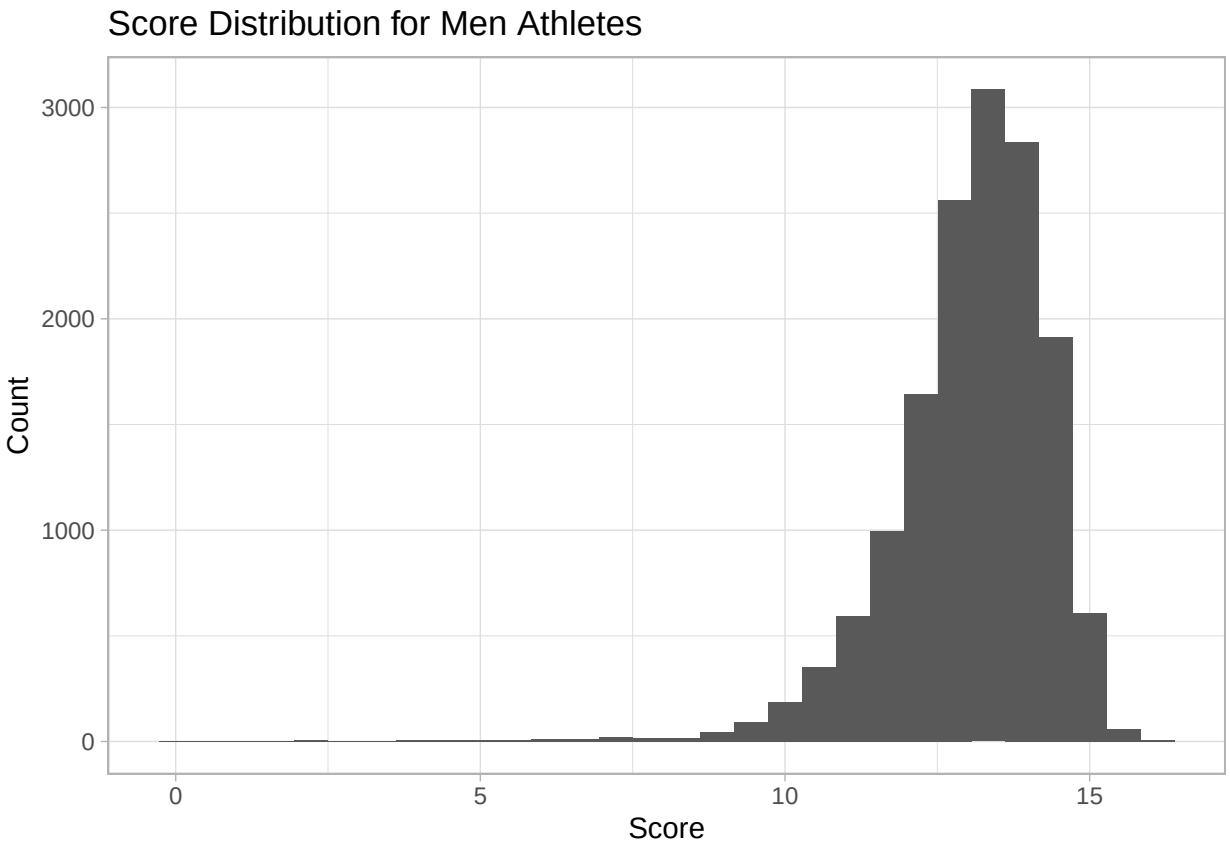
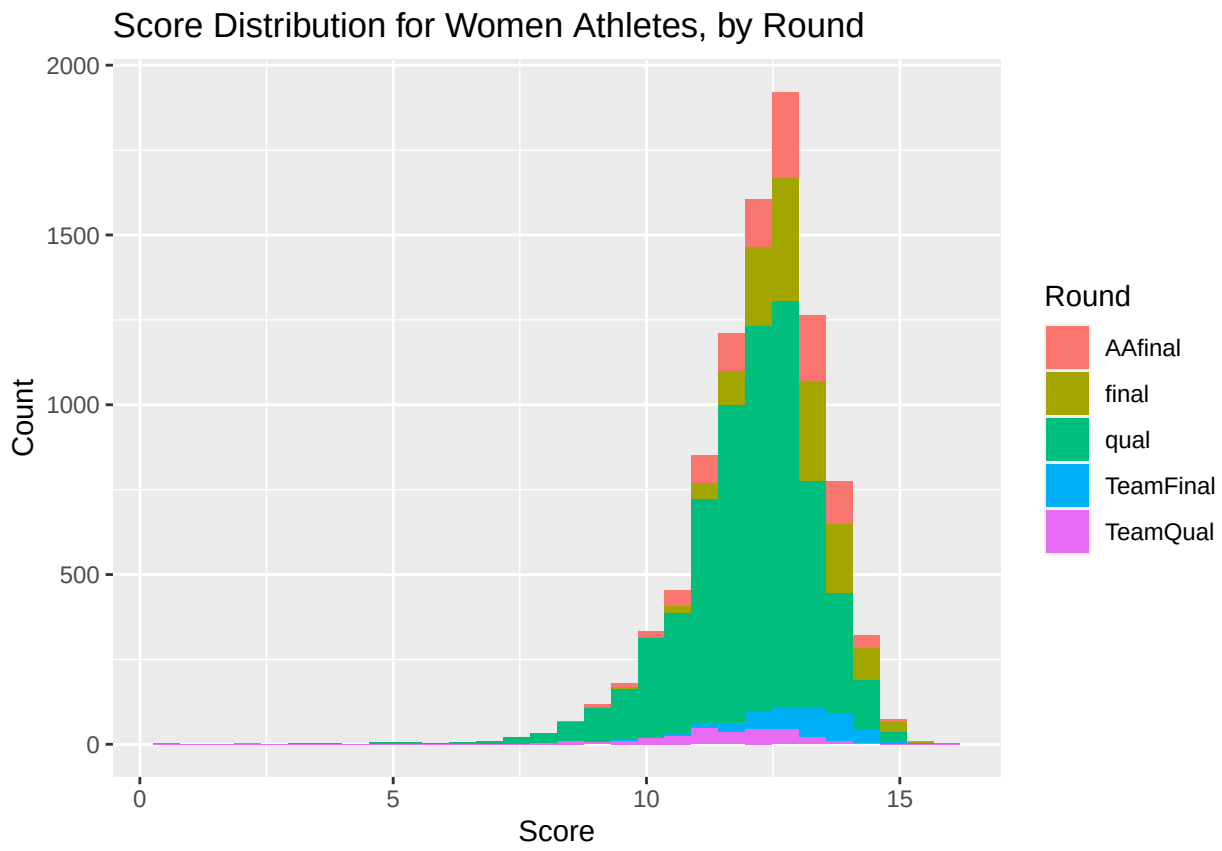
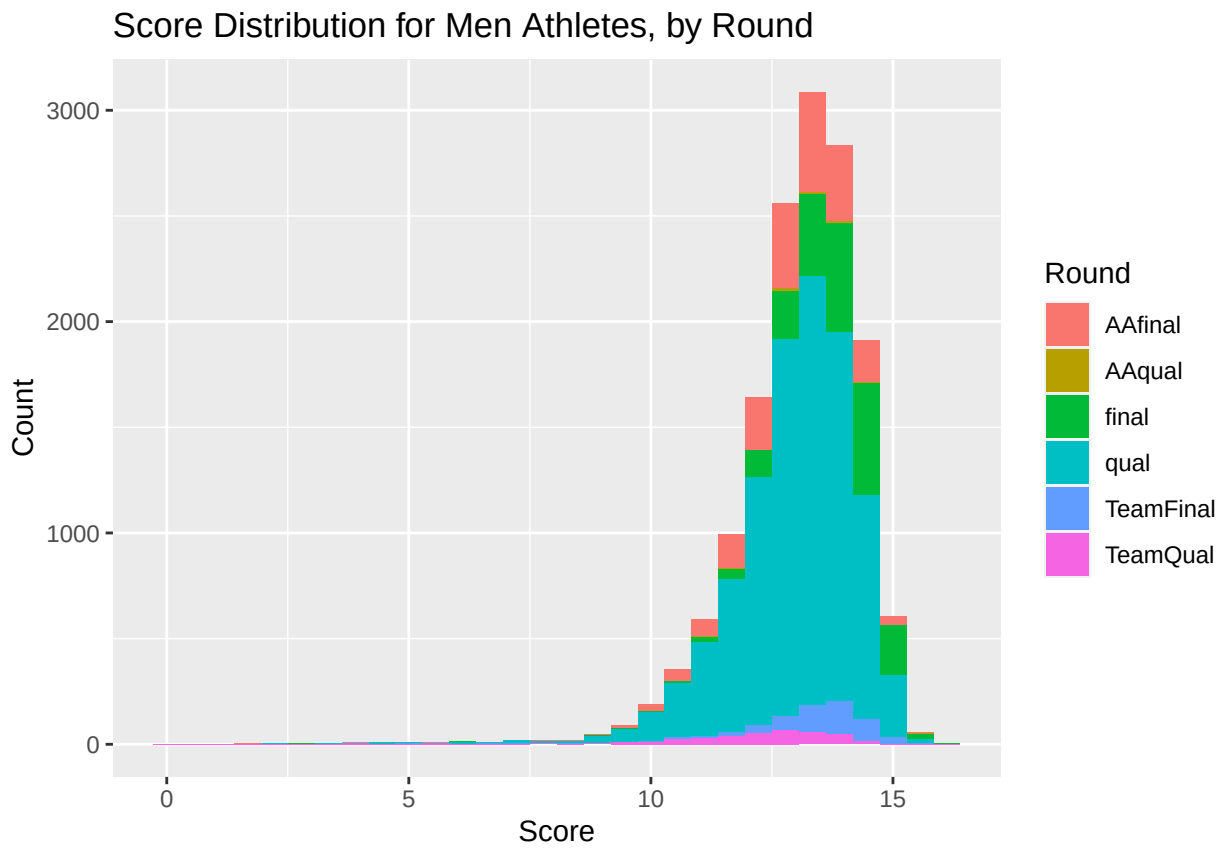


Figure 6: Score Distribution for Male Athletes




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




```
## [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`.
```

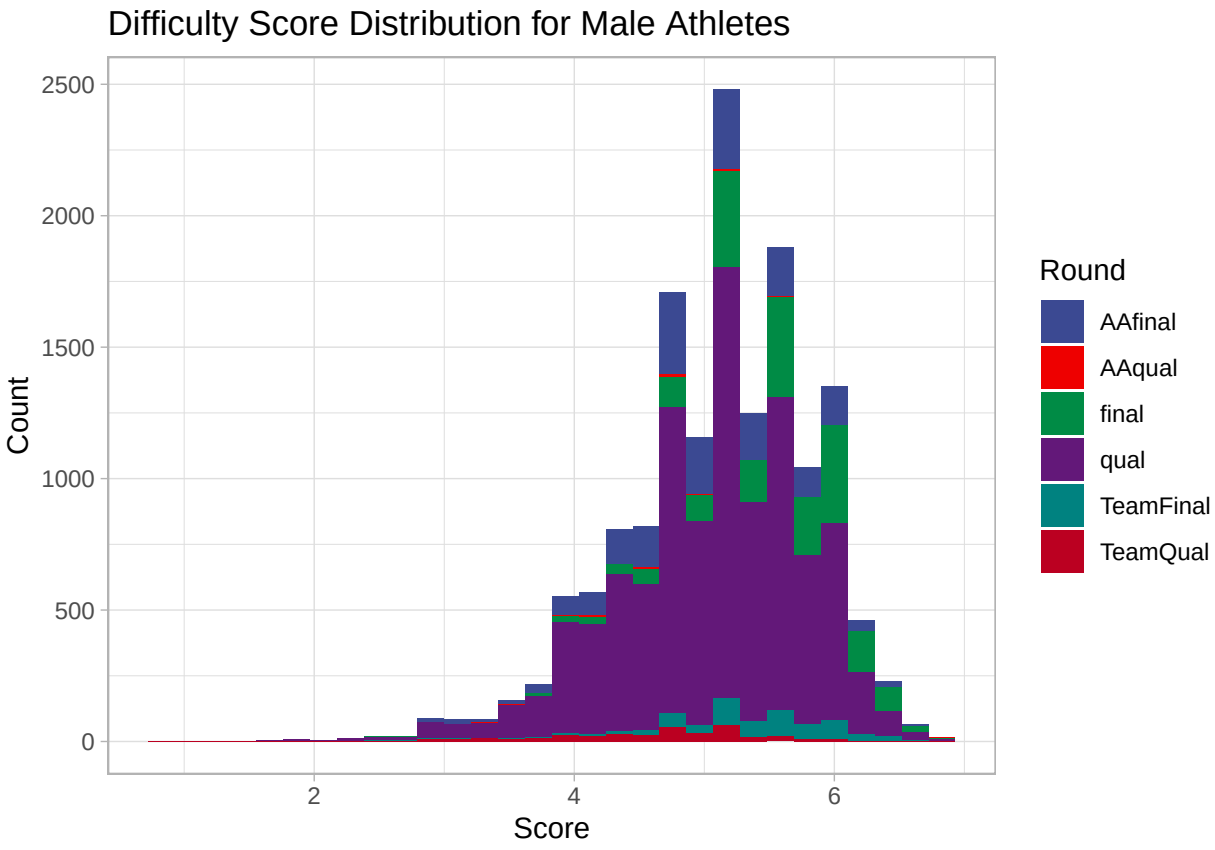


Figure 7: Difficulty Score Distribution for Male Athletes

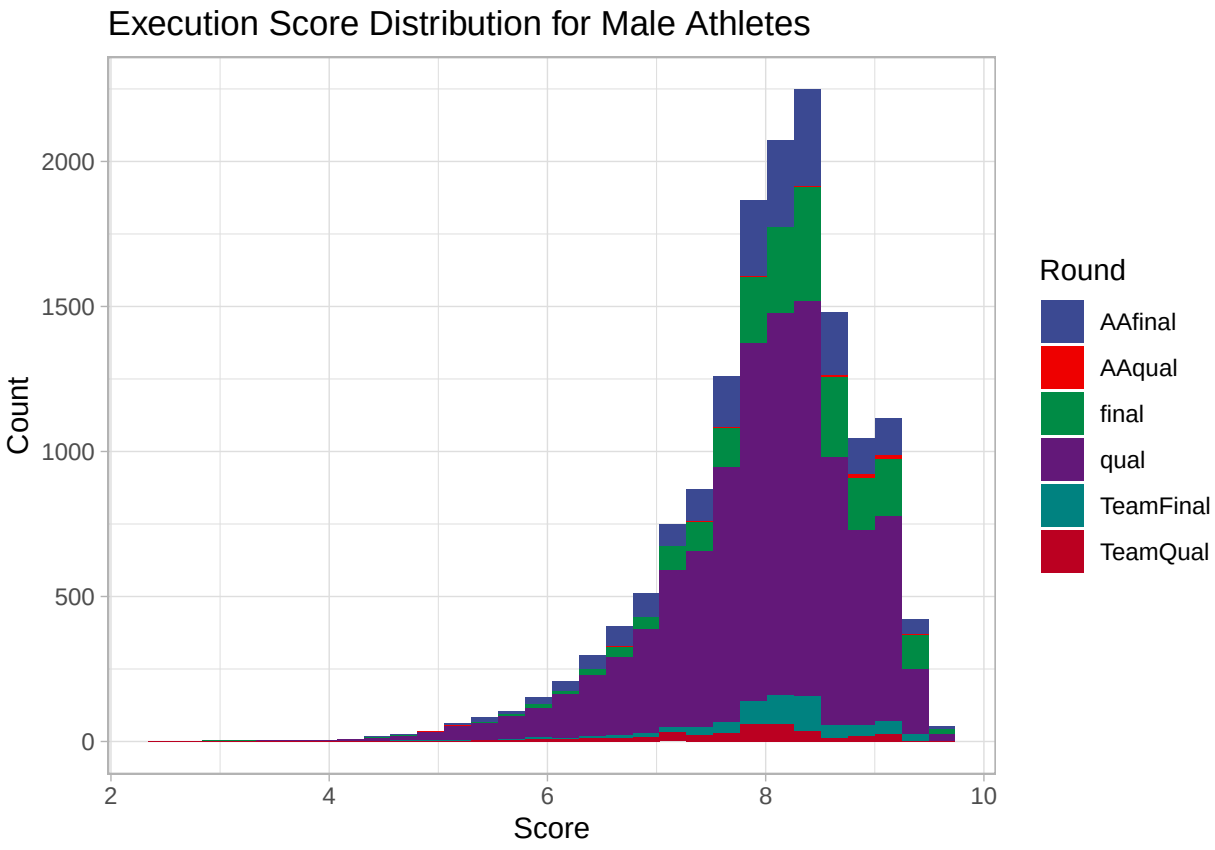
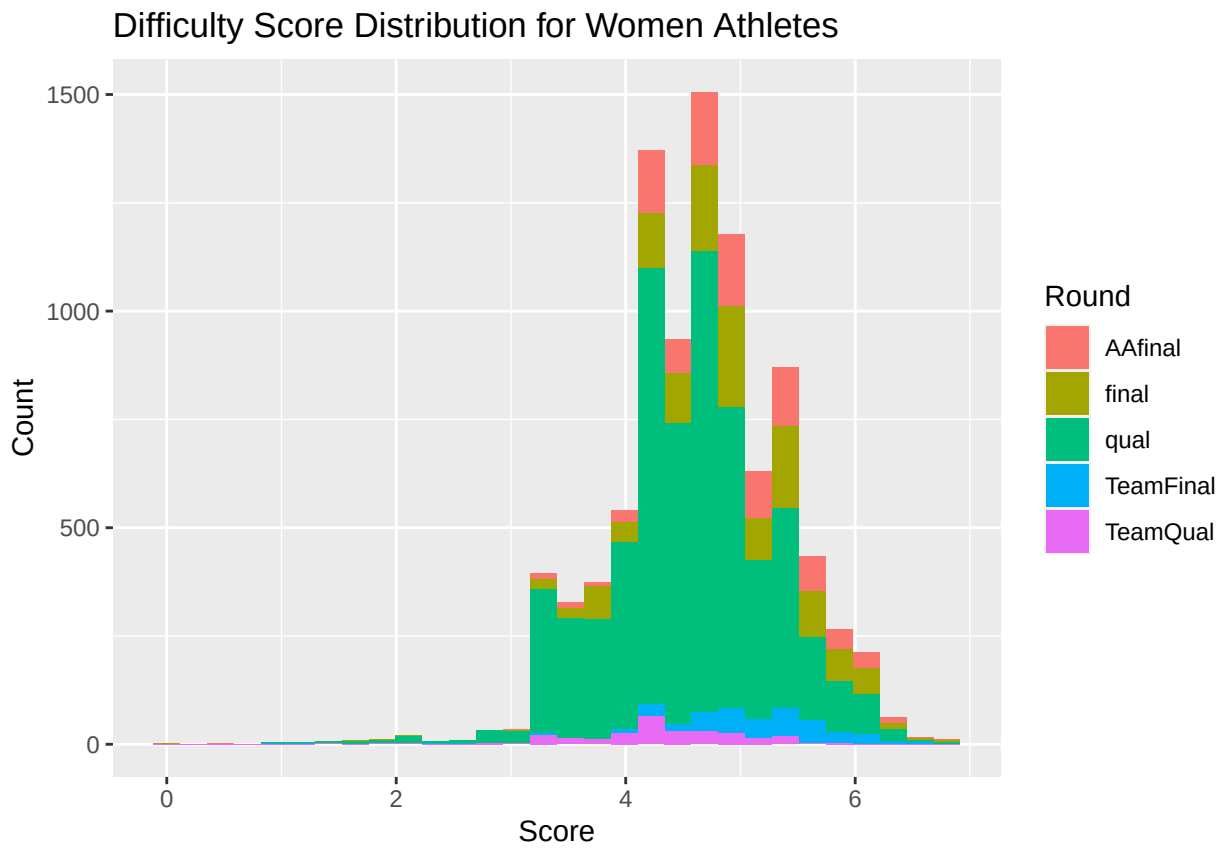
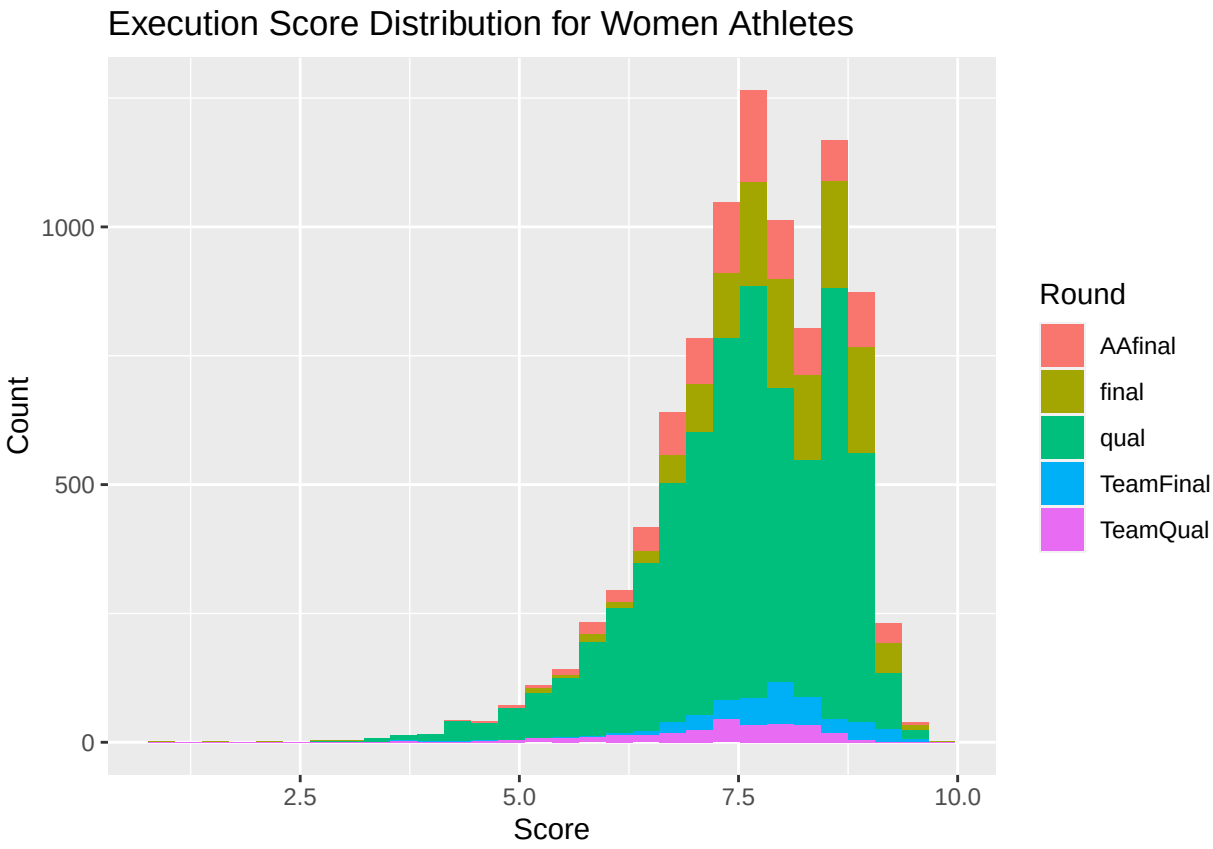
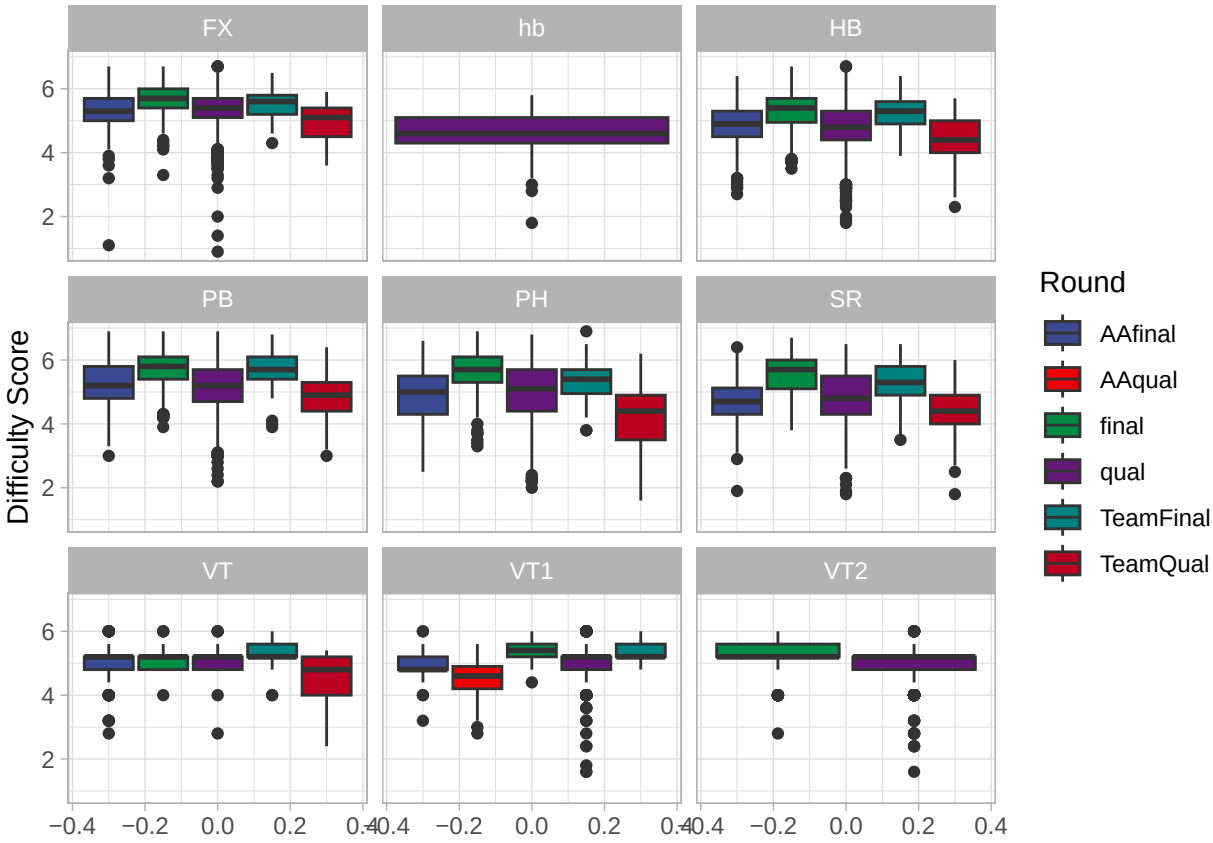


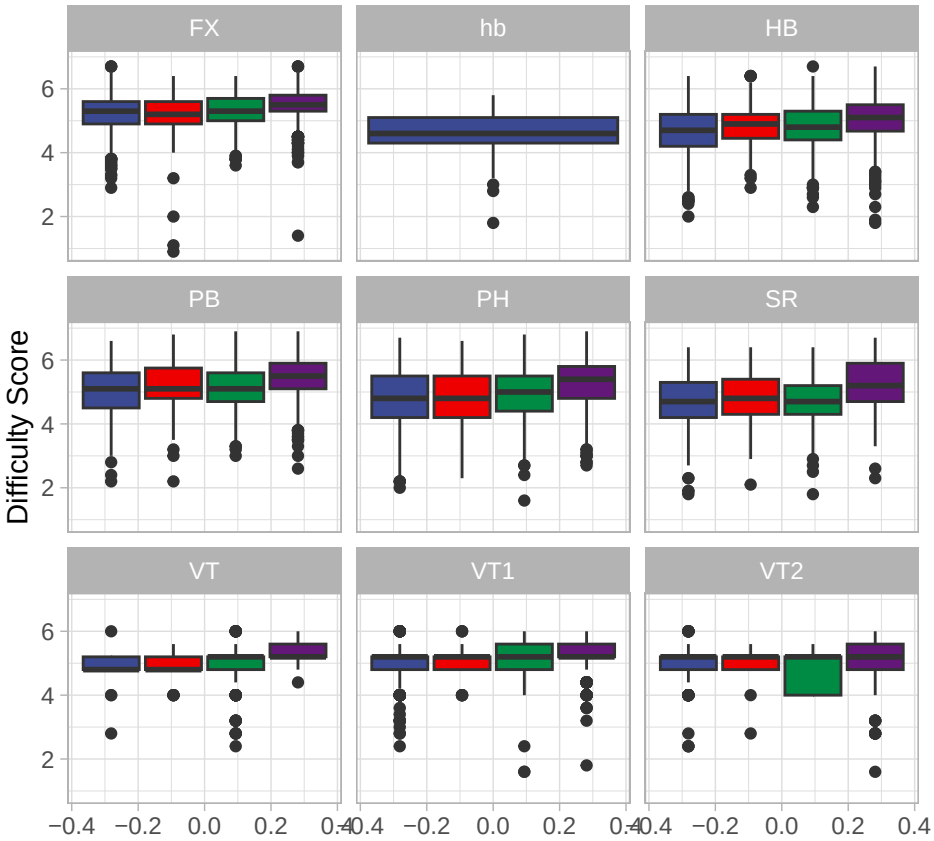
Figure 8: Execution Score Distribution for Male Athletes




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


Execution Score Distribution for Women Athletes
Count
1000
500
0
2.5
5.0
7.5
10.0
Score
Round
AAfinal
final
qual
TeamFinal
TeamQual


FX
hb
HB
PB
PH
SR
VT
VT1
VT2
Difficulty Score
2
4
6
−0.4
−0.2
0.0
0.2
0.4
Round
AAfinal
AAqual
final
qual
TeamFinal
TeamQual



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

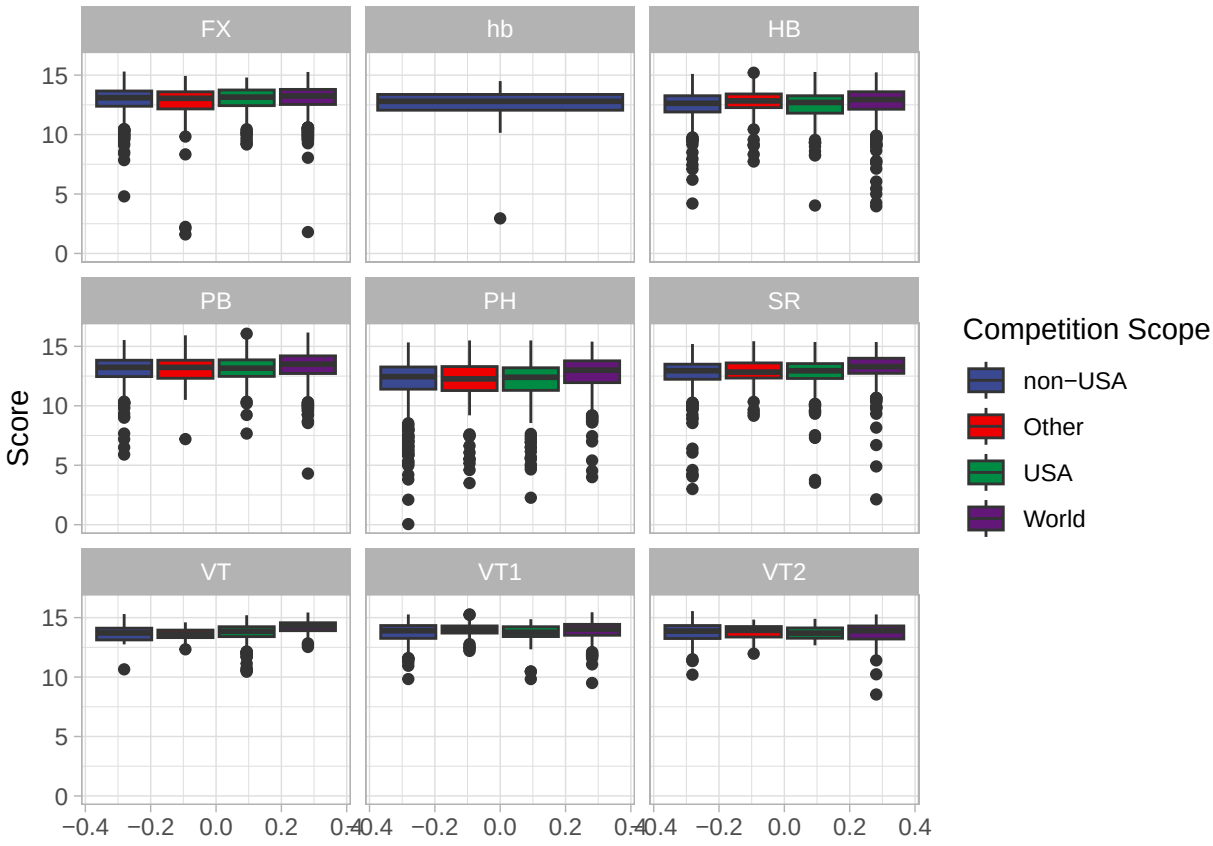


Figure 9: Distribution of Scores by Scope of Competition and Apparatus for Men