AMOD-5240H Project EDA

Yameen

Summary Statistics

Table 1: Summary Statistics for Age, Height, and Weight

Variable	Mean	Median	Min	Max	SD	IQR
Age	25.56	24	10	97	6.39	7
Height	175.34	175	127	226	10.52	15
Weight	70.70	70	25	214	14.35	19

Table 2: Proportion of Athletes by Sex

Sex	Count	Proportion (%)
F	74522	27.49
M	196594	72.51

Table 3: Proportion of Medal Winners

Medal Type	Count	Proportion (%)
Bronze	13295	4.90
Gold	13372	4.93
Silver	13116	4.84
NA	231333	85.33

Table 4: Medal Distribution by Medal Type & Sex

Medal Type	Count (Female)	Count (Male)	Proportion (Female %)	Proportion (Male %)
Bronze	3771	9524	1.39	3.51
Gold	3747	9625	1.38	3.55
Silver	3735	9381	1.38	3.46
NA	63269	168064	23.34	61.99

Visualization

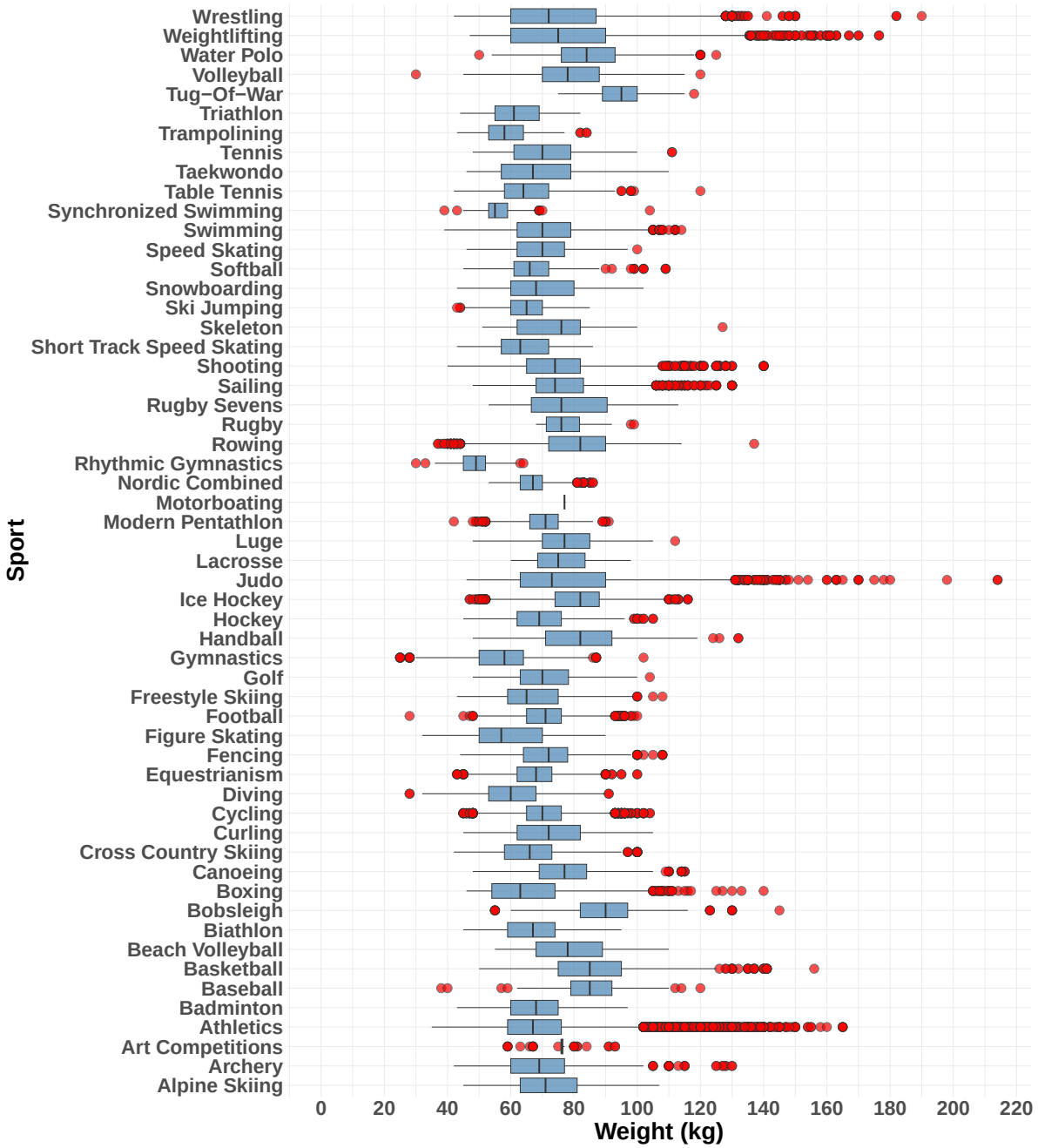

Figure 1: Box Plot of Weight Distribution by Sport

Discussion

Summary Statistics (Table 1)

Table 1 presents descriptive statistics for three continuous variables: Age, Height, and Weight of athletes.

Key Insights:

Age:

1. The mean age is 25.56 years, with a median of 24 years, indicating a slightly right-skewed distribution.
2. The minimum age is 10, suggesting participation of very young athletes.
3. The maximum age is 97, indicating that some sports allow much older competitors.
4. The IQR is 7, showing that most athletes fall within a 7-year age range around the median.

Height:

1. The mean height is 175.34 cm, with a median of 175 cm, suggesting a fairly symmetric distribution.
2. Heights range from 127 cm to 226 cm, showing significant variation, likely due to different sports requiring different physical attributes.
3. The IQR is 15 cm, indicating moderate dispersion around the median.

Weight:

1. The mean weight is 70.70 kg, with a median of 70 kg, showing a balanced distribution.
2. The minimum weight is 25 kg, while the maximum is 214 kg, showing extreme variation due to different body requirements across sports.
3. The IQR is 19 kg, indicating a relatively wider spread of weight compared to height.

Overall, these statistics show substantial variation in athlete characteristics, reflecting the diversity of physical demands across different sports.

Proportion of Athletes by Sex (Table 2)

Table 2 provides an overview of the distribution of athletes based on sex in the dataset.

Key Insights:

- Male athletes (M) make up 72.51% of the dataset, while female athletes (F) account for only 27.49%.

- This indicates a historical gender disparity in Olympic participation, with men being overrepresented.
- This imbalance could be due to historical factors where fewer women competed in early Olympic events.
- Further analysis could examine trends over different decades to see if this gender gap has narrowed over time.

Proportion of Medal Winners (Table 3)

Table 3 breaks down the proportion of medals won, including athletes who did not receive a medal (NA category).

Key Insights:

- Only a small proportion of participants win medals:
  - Gold: 4.93%
  - Silver: 4.84%
  - Bronze: 4.90%
  - No Medal (NA): 85.33%
- The vast majority (85.33%) of athletes do not win any medals, highlighting the competitive nature of the Olympics.
- The distribution among the three medal types is fairly even, with no significant bias toward one type.
- Further exploration could examine whether certain sports have a higher likelihood of medal wins.

Medal Distribution by Medal Type and Sex (Table 4)

Table 4 provides a gender-based breakdown of medal distributions across different medal types.

Key Insights:

- Male athletes have won significantly more medals than female athletes, reflecting historical participation disparities.
- The proportions for each medal type between male and female athletes are similar, meaning that once competing, both sexes have similar success rates.
- NA category (athletes without medals) is notably high for both males and females, aligning with the previous table showing only ~15% of participants win medals.
- Future analysis could investigate if participation in certain sports influences medal-winning rates for each gender.

Visualization with Box Plot of Weight Distribution by Sport (Figure 1)

The box plot in Figure 1 provides an overview of the weight distribution of Olympic athletes across different sports. It allows us to analyze the spread, central tendency, and potential outliers for each sport in terms of weight.

Key Observations:

1. Sports with Higher Median Weight

- Weightlifting, Wrestling, and Rugby have the highest median weights, with most athletes in these categories weighing significantly more than athletes in other sports.
- Water Polo, Ice Hockey, and Rowing also show relatively higher weight distributions, likely due to the physical strength required in these sports

2. Sports with Lower Median Weight

- Gymnastics, Diving, and Figure Skating show some of the lowest weight distributions among all sports.
- Rhythmic Gymnastics and Synchronized Swimming exhibit very compact distributions with lower variability, indicating that most athletes in these sports fall within a narrow weight range.

3. Outliers and Extreme Values

- Several sports, such as Wrestling, Weightlifting, and Rugby, contain upper outliers, indicating that some athletes weigh significantly more than their peers in the same sport.
- Archery, Equestrianism, and Art Competitions show less variation, meaning the athlete weights in these sports are relatively uniform.

4. Variability in Weight Across Sports

- Sports like Athletics and Swimming have a wide spread, indicating a large range of athlete weights, which is expected due to the variety of events within these sports (e.g., sprinters vs. long-distance runners, or sprinters vs. swimmers).
- Basketball, Volleyball, and Rowing also show a large IQR, reflecting diversity in body composition within the sport.