

Provide a summary of the different descriptive statistics you looked at and WHY. The main variables I am looking at as possible predictors of Olympic performance are sex, age, height, and weight. I first looked at the distributions of the age, height, and weight -both across all athletes and across medalists only - using the describe() function to get mean, median, Q1 and Q3, as well as histograms to visualize the distribution. Then, in order to get a sense of how these variables differ between men and women and between different events, I created a new table in Pandas. The table groups entries in the original data by sex and sport, averaging age, height, weight, and BMI for all relevant athletes as well as all relevant medalists, and also includes columns to display the differences between these averages for all athletes vs. medalists.

Note: Due to Athletics including track and field events (which are very different), I distinguished between them in my new table.

Submit 2-3 key points you may have discovered about the data, e.g. new relationships?

Aha's! Did you come up with additional ideas for other things to review? 1. Across all athletes, the distributions for age, height, weight, and BMI are close to normal (although the age, weight, and BMI distributions are slightly right-skewed). Age is centered at 25.1 years, height at 175.4 cm, weight at 70.7 kg, and BMI at 22.8.

2. Across medalists only, these distributions are centered slightly higher: age at 25.4 years, height at 177.6 cm, weight at 73.8 kg (this is the biggest difference from that of all athletes), and BMI at 23.16.

3. For both men and women, athletes in these sports tend to have especially extreme values for age, height, and weight:

Low age: Gymnastics, Swimming, Figure Skating, Diving, Short Track Speed Skating, Alpine Skiing

High age: Shooting, Curling, Equestrianism

Low height: Gymnastics, Diving, Weightlifting

High height: Volleyball, Water Polo, Basketball, Rowing

Low weight: Gymnastics, Figure Skating, Diving

High weight: Volleyball, Water Polo, Basketball, Rowing, Bobsleigh

4. In most sports, women are about 1 year younger, 11.5 cm shorter, and 15.7 kg lighter than men. The most extreme differences are in these sports:

Gymnastics: women 5.1 years younger

Shooting: women 3.6 years younger

Sailing: women 3.3 years younger

Figure Skating: women 2.6 years younger

Cycling: men 2.2 years younger

5. The biggest differences between average age/height/weight between all athletes and medalists only are in these sports:

Medalists younger than average: M/F Table Tennis, M/F Shooting, M/F Archery, M Rugby Sevens

Older than average: F Curling, M/F Figure Skating, M Equestrianism, M/F Cross-Country Skiing, M Alpine Skiing, F Speed Skating

Shorter: M/F Gymnastics, M/F Diving, M Short Track Speed Skating, M Table Tennis, M Badminton

Taller: Many sports, but the most extreme examples are F Luge, F Modern Pentathlon, M Taekwondo, F Rhythmic Gymnastics, M/F Swimming

Lighter: M Ski Jumping, M Table Tennis, M Short Track Speed Skating, M/F Diving, M/F Gymnastics, M Badminton

Heavier: M/F Athletics (Field), M/F Cycling, M Basketball **Did you prove or disprove any of your initial hypotheses? If so, which one and what do you plan to do next?**

Hypothesis 1 (The best-performing athletes will be those in their mid-to-late 20's) seems at least partially correct, as the distribution shows medalists to be in this range but also in the early-to-mid 20's.

Hypothesis 2 (BMI of the best athletes will generally be about 20-23) also seems partially correct, as there are also quite a few medalists with BMIs above 23.

Hypothesis 4 (Height/weight/BMI will be lower in women) is confirmed by the above data.

Hypothesis 5 (medalists will have low height/weight in speed/agility sports, and high height/weight in strength sports) is interesting. Sports like Gymnastics, Diving, and Athletics (Field) fit my expectations. However, others, such as Weightlifting (a strength sport for which height is low) as well as Volleyball, Water Polo, and Basketball (team sports for which (expectedly) height AND (unexpectedly) weight are high) surprised me.

I still have yet to more deeply analyze Hypothesis 3 (correlation between height and weight).

What additional questions are you seeking to answer? Going forward, I want to:

1. Analyze correlation between height and weight (Hypothesis 3)
2. Think about why the trends I noticed make sense in given sports (for example, why do medalists' age/height/weight differ so much from other athletes in certain sports)?
3. Possibly categorize similar types of sports (team sports, races, etc.) and analyze trends within these categories
4. Possibly analyze trends over time