Olympic Medals: Matter of Nerves

Dr Preet Deep Singh

07/08/2021

Abstract

Using a dataset of over 120 years of Olympic history, Ifind that after 1950 the mean age of medal winners is higher than that of non-medal winners. This relationship is more pronounced in the case of Gold medalists. I hypothesize that Age is a proxy for experience of Olympic Games and show that the probability of getting a medal in the second appearance is higher than that in the first one. This could be due to experience, better handling of pressure and better preparation.

Visualising Age and Success

I look at the age of all athletes that have competed at Olympics in the past 120 years. It has a mean of 25.5568984. The mean age of all athletes that won a medal over the same time period is 25.9251748. The mean age of all athletes who did not win a medal during the same period is 25.4922885. As evident, there is very little difference in the mean age of atheletes who won and who didnt win. One reason for the same could be the quality of the competition where the difference is overcome by skill, training and practice.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   10.00   21.00   24.00   25.56   28.00   97.00    9474
```

Low Age

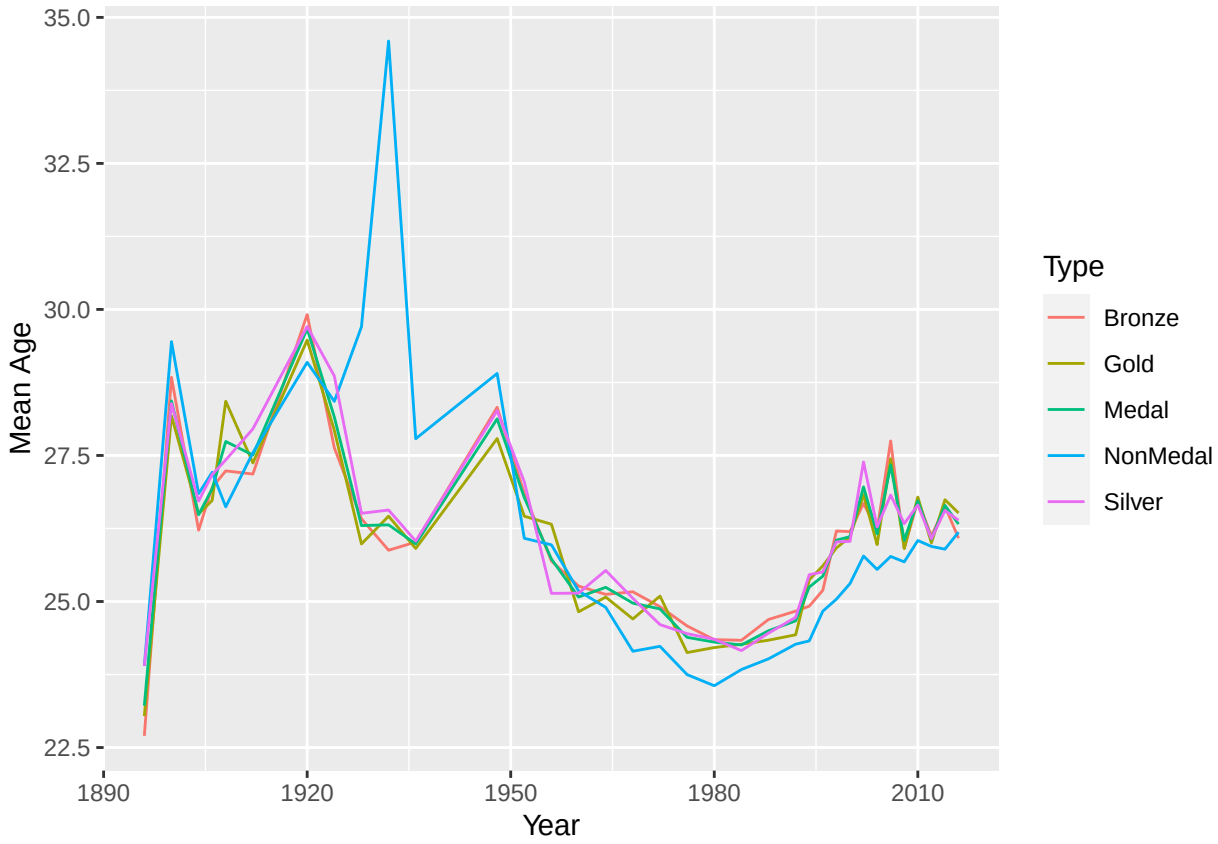
A total of 12508 participations have been seen at Olympics by athletes who were less than 18 years of age. Out of these 6712 have been after 1980. This number includes multiple participation by the same athlete as well. 5959 unique athletes have participated till date and those who competed after 1980 are 3136.

High Age

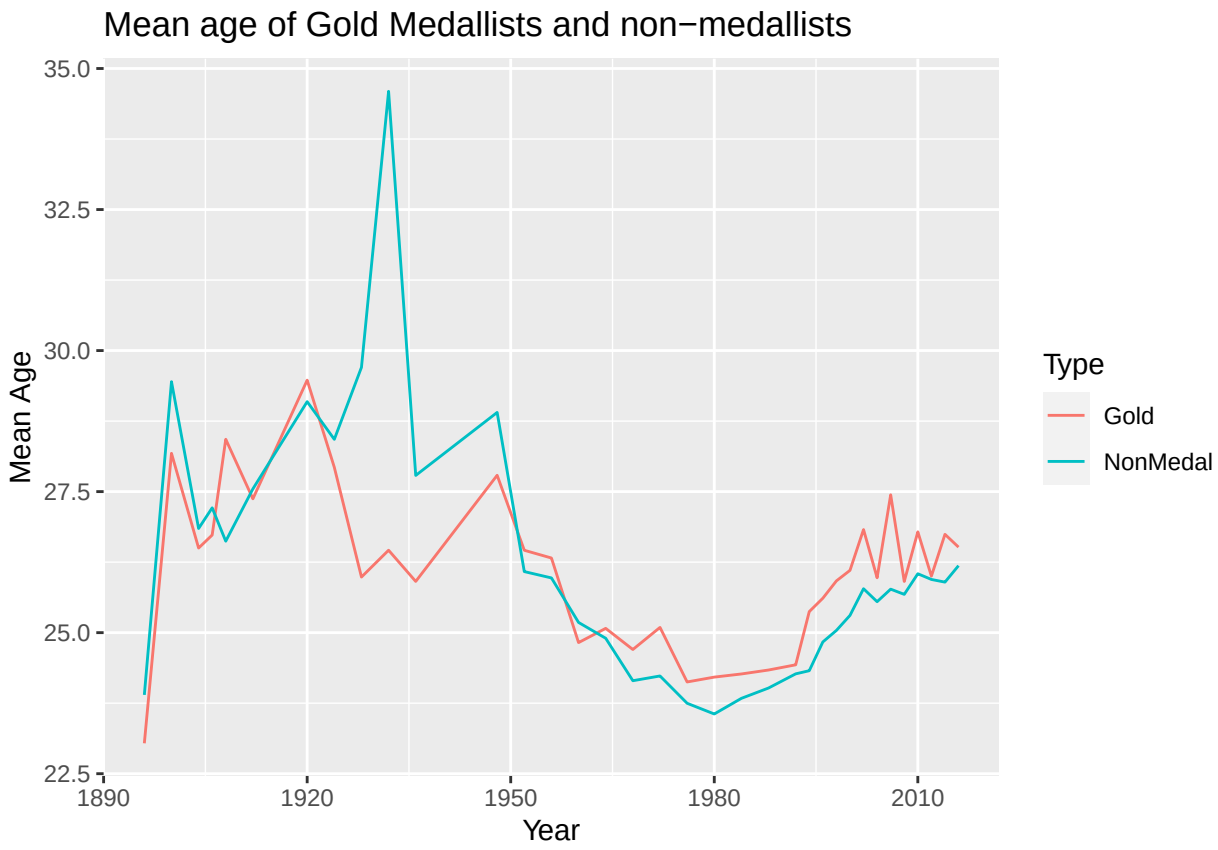
A total of 570 participations have been seen at Olympics by athletes who were more than 60 years of age. Out of these 26 have been after 1980. This number includes multiple participation by the same athlete as well. 235 unique athletes have participated till date and those who competed after 1980 are 16.

Pooled Plot

A pooled plot of the age of athletes over the years shows much more variation before 1950. The mean age of 1928 Olympics is abnormally high. While the age has not been constant in any two years, the mean age has remained between 25 and 28 in this century.


35.0
32.5
30.0
27.5
25.0
22.5
Mean Age
1890
1920
1950
1980
2010
Year
Type
Bronze
Gold
Medal
NonMedal
Silver

Gold Vs Non Medal



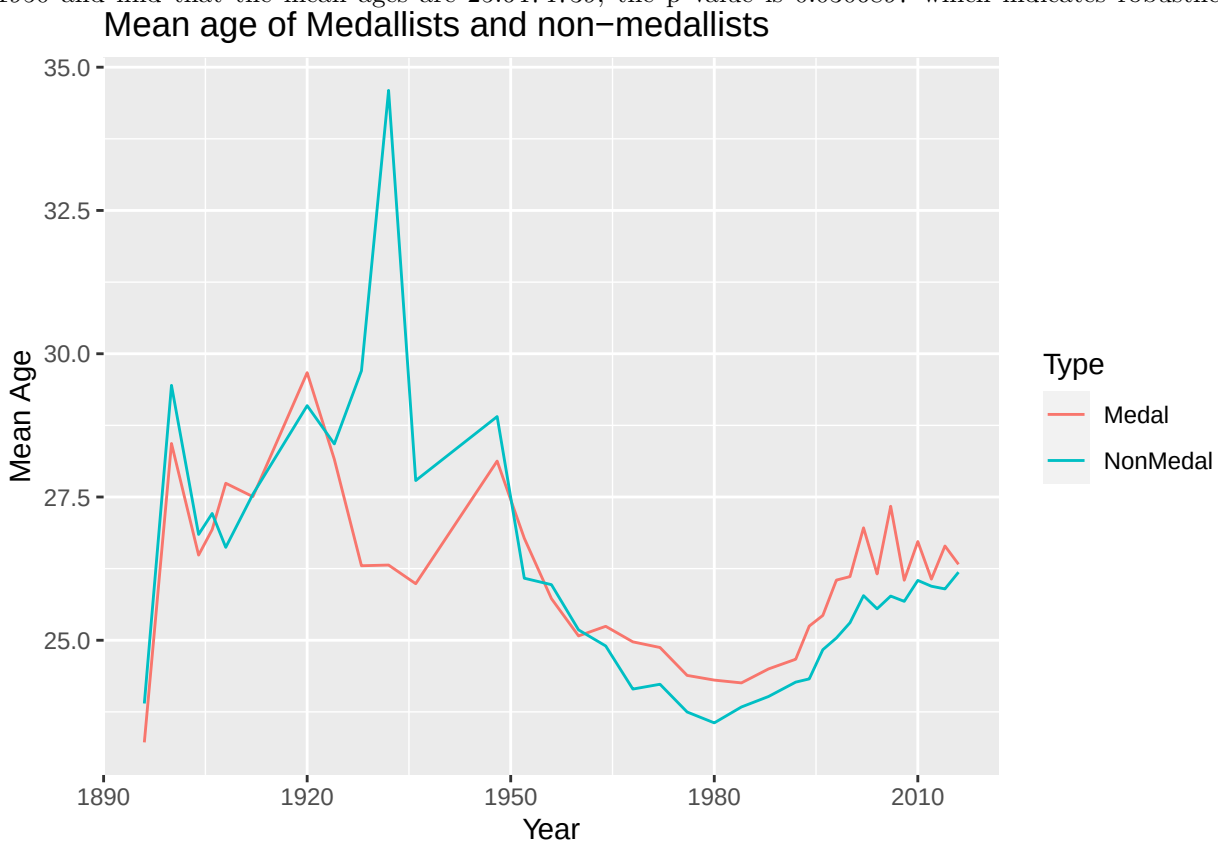

```
##
##  Welch Two Sample t-test
##
## data:  agelater$meanage[agelater$Type == "Gold"] and agelater$meanage[agelater$Type == "NonMedal"]
## t = 2.0255, df = 43.343, p-value = 0.049
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.002545472 1.106511716
## sample estimates:
## mean of x mean of y
##  25.61096  25.05643
```

In order to look at the difference between the Gold Medal winners and the non-medallists, we plot their mean age over the years. Starting 1950, the mean age of Gold Medallists has been higher than that of non-medallists other than in one instance. This could be because age represents experience. Olympics is an important event and there is a fair amount of pressure on the participants. Age or experience might assist in dealing with that pressure. We conduct a Welch Two Sample t-test, the p-value of the difference in mean of age of gold medallists and non-medallists is 0.8222277, thereby indicating that the difference in age is not statistically significant. This could be because of the anomally in 1928. We limit our sample to post 1950. We find that the p value is not 0.0489982 which is fairly robust.

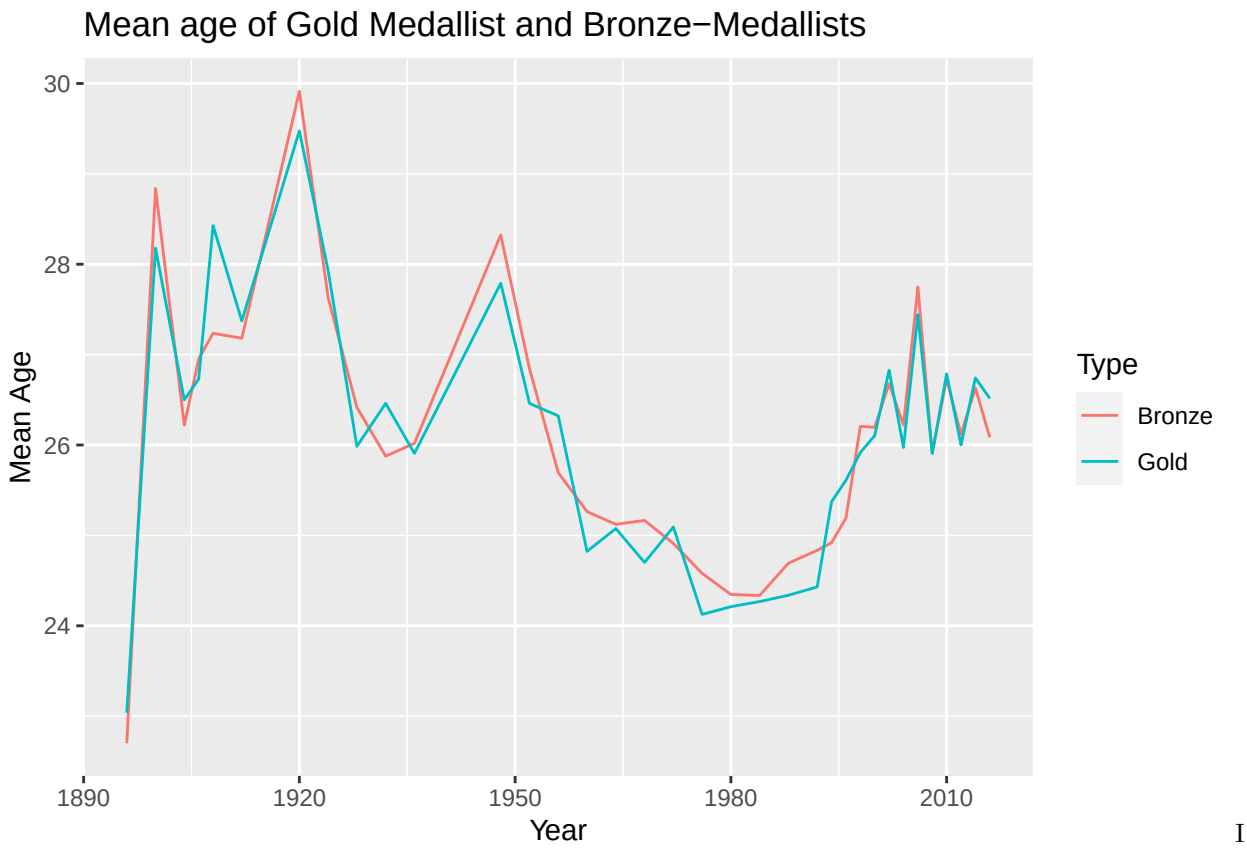
Medal Vs Nonmedal

I also look at the difference in mean age between medallists and non-medallists. The mean ages are 26.1360254, the p value is 0.9175656 which indicates randomness in the results. Similar to above, I look at values after

1950 and find that the mean ages are 25.6474739, the p value is 0.0306897 which indicates robustness.



Gold Vs Bronze



I look at values after 1950 for the mean age of Gold medal winners and Bronze medal winners. and find that the mean ages are 25.6109562 and 25.6709731 , the p value is 0.8315852 which implies that the observed difference is random. This points to the tough competition at the Olympics.

Correlation

The correlation between Whether or not someone won a medal and their age is

```
##         cor
## 0.02412649
```

```
## [1] 5.33573e-35
```

Regression

We use Logistic regression to model the dependent variable of Success defined as getting a medal.

We use a pooled regression to model success as defined by winning a medal, depedent on age, while controlling for height, weight, and gender. We see that the p value for Age is extremely significant at $2.2834577 \times 10^{-14}$, the estimate of 0.0011046 is not materially significant. Change of an year in age would lead to 0.001% increase in the probability of a medal.

Controlling for Sport

In order to better understand the role age plays, we run regressions on subgroups of sports. We present a table of estimates of Age where the p value is less than 0.05.

```
##                    Sport          Age
## 22 Synchronized Swimming  0.028225430
## 19        Figure Skating  0.017698655
## 1               Football  0.012985071
## 21            Volleyball  0.008602487
## 3   Cross Country Skiing  0.008078978
## 18     Modern Pentathlon  0.007682225
## 24       Nordic Combined  0.007364293
## 2          Speed Skating  0.006880085
## 9                Rowing  0.006592649
## 8                  Luge  0.005885674
## 7          Alpine Skiing  0.004301481
## 15               Diving  0.004222864
## 16             Canoeing  0.003996503
## 26           Ski Jumping  0.003963489
## 17               Tennis  0.003735010
## 10             Bobsleigh  0.003688796
## 6              Gymnastics  0.002898858
## 11         Equestrianism  0.002512693
## 14               Cycling  0.001596447
## 5                Sailing  0.001516626
## 4               Athletics -0.001516380
## 12              Shooting -0.002323429
## 20               Archery -0.003920322
## 23          Table Tennis -0.004818834
## 13             Taekwondo -0.010226588
## 25           Rugby Sevens -0.013282973
```

We see that Age has a different role to play across sports. While Synchronized Swimming sees the maximum positive impact of age, Rugby Sevenssees the maximum negative impact of age. All sports that have a negative sign, have a negative impact of increasing age, on the likelihood of a medal.

Controlling for Sport and Year

```
##                    Sport          Age
## 19        Figure Skating  0.020969979
## 1               Football  0.013678369
## 3   Cross Country Skiing  0.008957542
## 11                Rowing  0.008131567
## 18     Modern Pentathlon  0.007966956
## 10                  Luge  0.006420489
## 2          Speed Skating  0.006227560
## 8          Alpine Skiing  0.005060836
## 9               Handball  0.004994668
## 17                Tennis  0.004951198
## 16              Canoeing  0.004122026
## 12             Bobsleigh  0.003746188
## 13         Equestrianism  0.003252834
## 7             Gymnastics  0.003143856
```

```
## 4              Swimming  0.002764415
## 6              Biathlon  0.002239077
## 5               Sailing  0.001447830
## 14             Shooting -0.002204774
## 15            Taekwondo -0.010230013
```

I also control for Years as factors and repeat the analysis. While Figure Skating sees the maximum positive impact of age, Taekwondosees the least impact of age. While controlling for years, we see that there is no sport that has a negative impact on likelihood of medal due to age.

Controlling for Event and Year

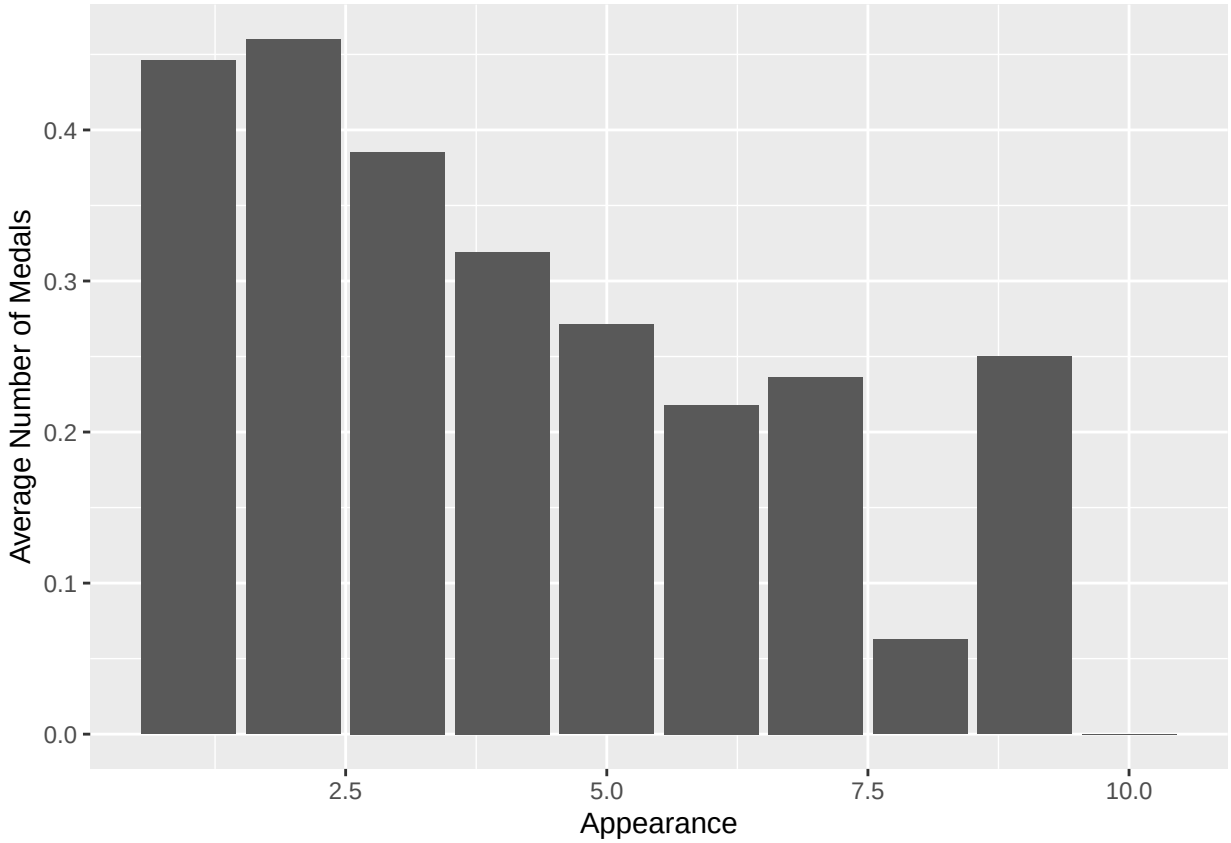
```
##                                                Event         Age
## 5 Cross Country Skiing Men's 4 x 10 kilometres Relay 0.019524927
## 6             Athletics Women's 4 x 100 metres Relay 0.007907996
## 2                   Speed Skating Women's 500 metres 0.007900925
## 1                             Football Men's Football 0.006739393
## 3           Cross Country Skiing Men's 10 kilometres 0.006147448
## 4           Cross Country Skiing Men's 50 kilometres 0.005955757
## 7                         Ice Hockey Men's Ice Hockey 0.004066803
```

I go one level deeper and look at the same analysis at event level. An event is one rendition of a sport. For example Swimming is a sport but 100m, 200m and 100m backstroke are different events. While sees the maximum positive impact of age, sees the least impact of age. As my data has become granular, there are fewer observations per regression set and the impact of Age has also becomes significant for fewer events.

Medals Across Outings

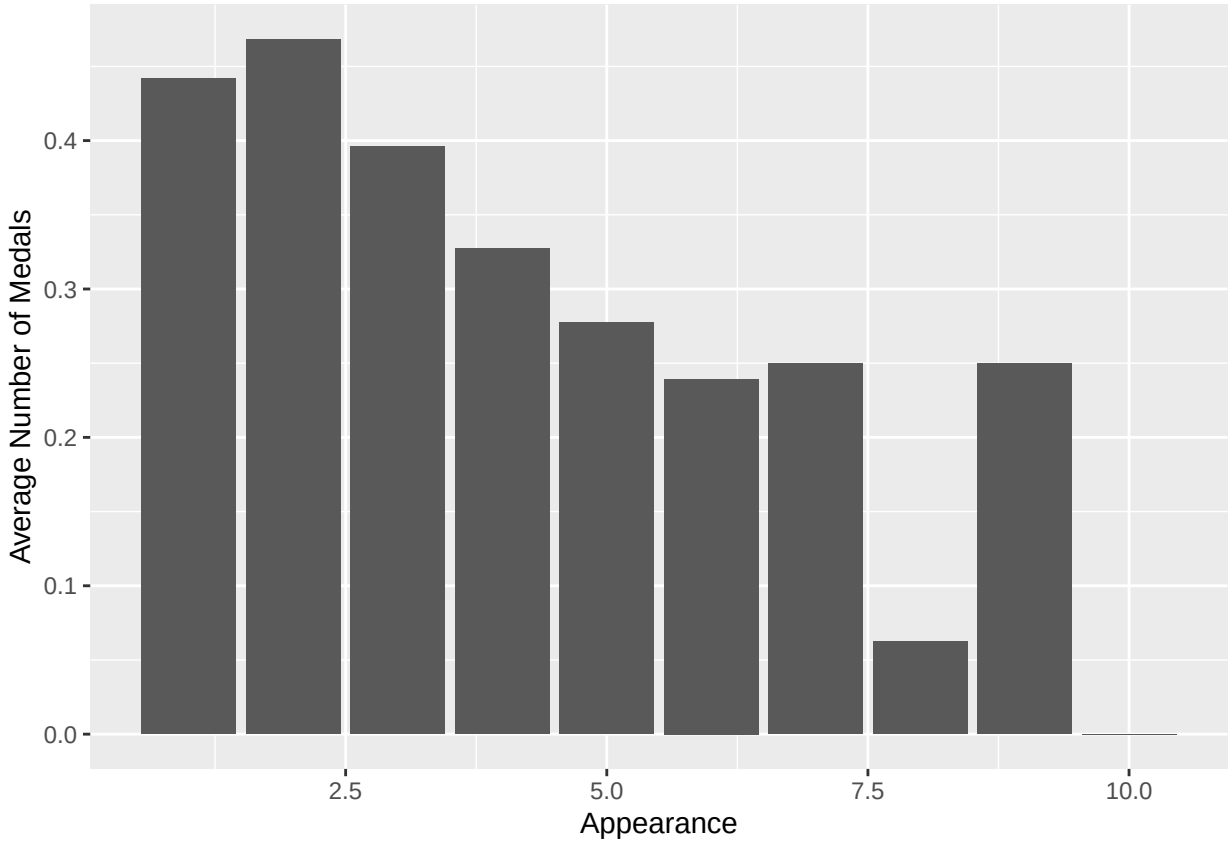
I look at Olympians who have participated in multiple versions of Olympics across years and have bagged at least one medal. I filter this to see whether the medal tally in subsequent appearances. I hypothesize that it is hard to maintain supremacy in such a competitive tournament and athletes would not be able to win in subsequent versions. We see a list of 12994 athletes that have competed in multiple Olympics with the highest being Ian Millar, Ian Millar Olympics. We do not consider breaks. For example if an athlete competed in two Olympics across 8 years, we still consider that as two participations irrespective of whether it was consecutive or not.

Pooled



If we look at the average number of medals (any color) for each iteration of appearance at the Olympics we find that the second outing is more likely to result in a medal than the first. I run a Welch Two Sample t-test on the medal victories in the first appearance versus a medal in the second appearance. The mean values are 0.4459932 and 0.4599895 with a p value of 0.0042754 which indicates that this relationship is statistically highly significant.

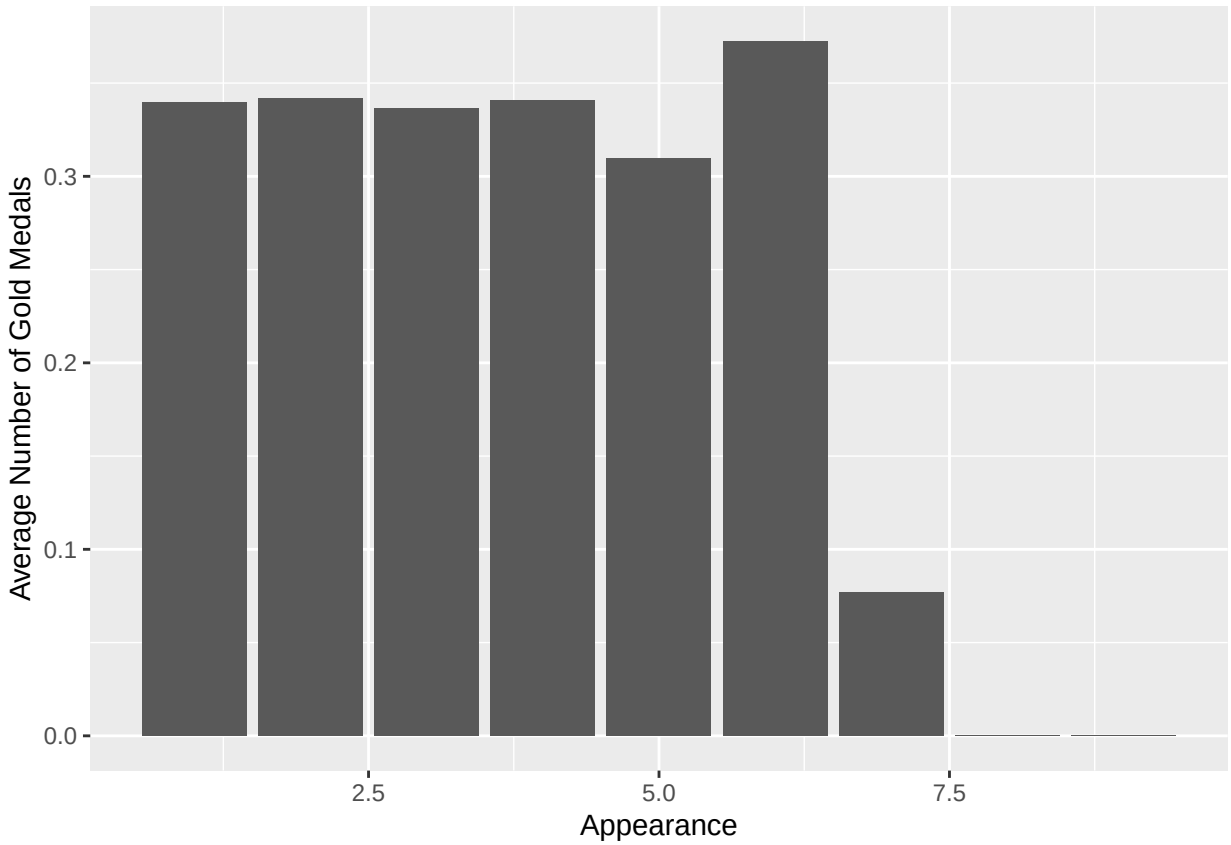
Post 1950



Similarly when I look at results from after 1950, I see that the man number of medals in the first appearance are 0.4419787 as compared to a mean value of 0.4682504 in the second outing. The p-value for this difference is highly significant at 8.932506×10^{-7}.

Gold Medals

```
## Warning: Removed 1 rows containing missing values (position_stack).
```



While the maximum appearances of an athlete to have won a gold medal are 7, I am interested in looking at how gold medals stack across appearances. The graph shows that the average number of Gold Medals are higher in subsequent appearances, though by a very thin margin. I conduct a Welch Two Sample t-test and see that the p value of 0.7828514 is not significant, thereby signalling that the difference in success based on the iteration of appearance is not there. We look at the same analysis for a more recent period, that is after 1950 and find a similar p value.

We conclude that the difference in success defined as winning a gold medal is not influenced by whether it is the first of the second appearance but when defined as winning a medal, the sequence of the appearance matters.

It coincides with our reasoning on why higher age seems to be correlated with medals even when controlling for height, weight, and gender.

Gold Medals, post 1950

ToDo

The source of this data is https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results/metadata.

- Limitations
- Titles for all graphs