Predicting Olympic Swimming Medals

—

Team_199: Luke Francese, Amalia Desir, Ella Davis, and Ruben Mesalles

Introduction to the Topic

Examining what makes a successful Olympic swimmer

- Research Question:
  - Which traits can best predict Olympic swimmers' performance (between nationality, height, weight, age, and BMI)?
- Initial Hypothesis:
  - Height and nationality will impact Olympic swimmers the most

Introduction to the Data

Overview- olympics <- read.csv("data/olympics.csv")

- The dataset contains information on athletes, events, and countries participating in the Olympic Games.

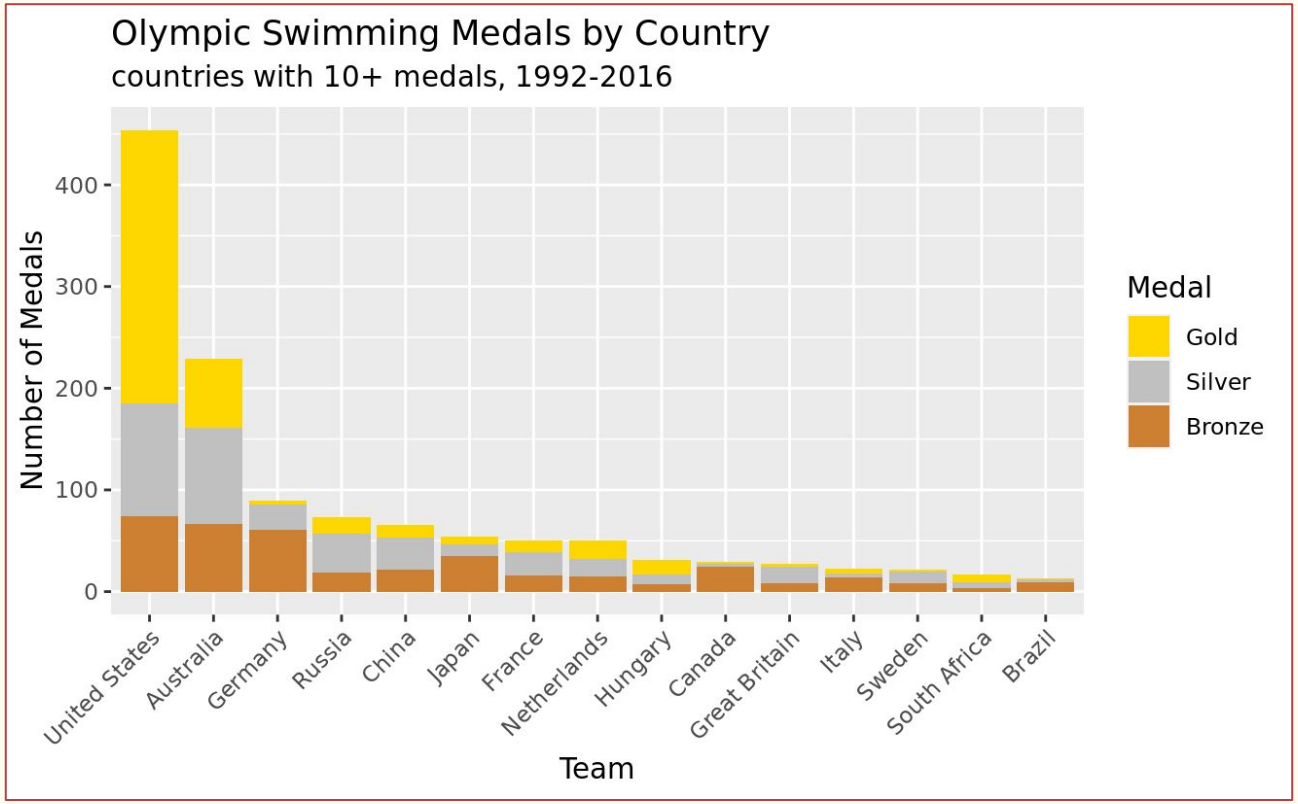
Common Variables-

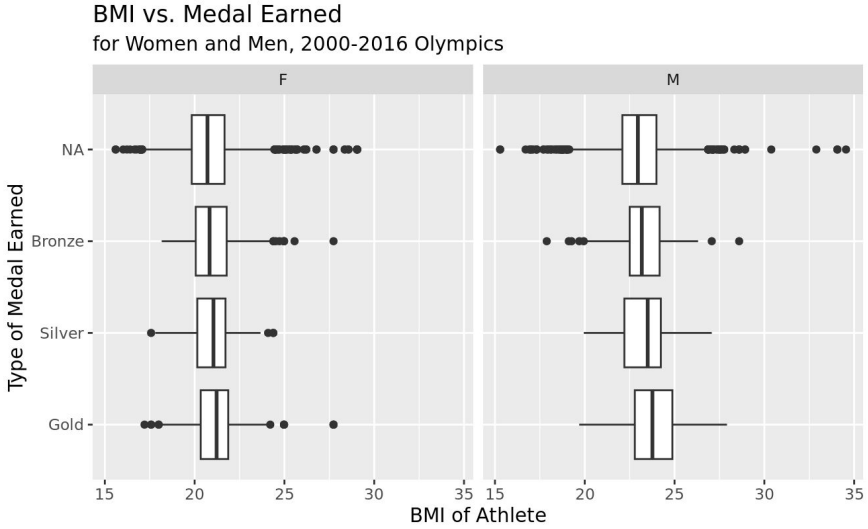
- Athlete,Sex, Age, Height, Weight, Team, NOC, Games, Year, Season, City, Sport, Event, and Medal.

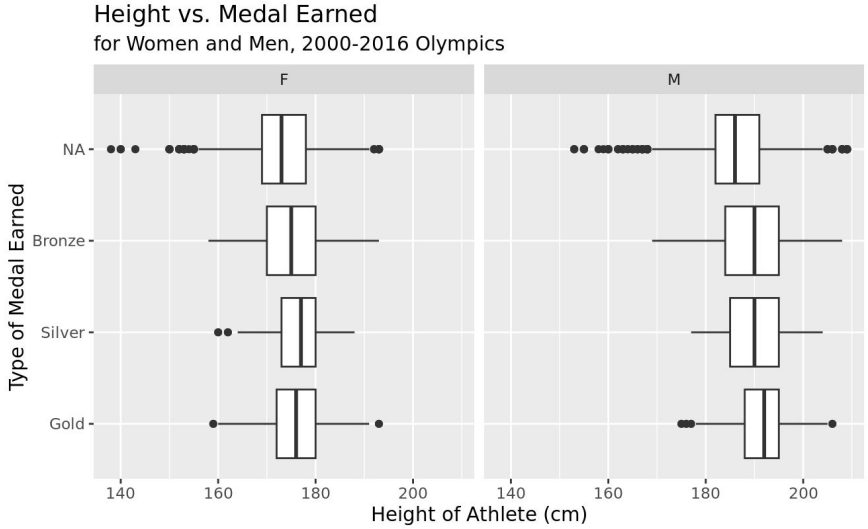
22 15 8	11 67 1	Suzanne Brigit "Sue" Bird	F	35	175	66	United States	USA	2016 Summer	2016	Summer	Rio de Janeiro	Basketball	Basketball Women's Basketball	Gold

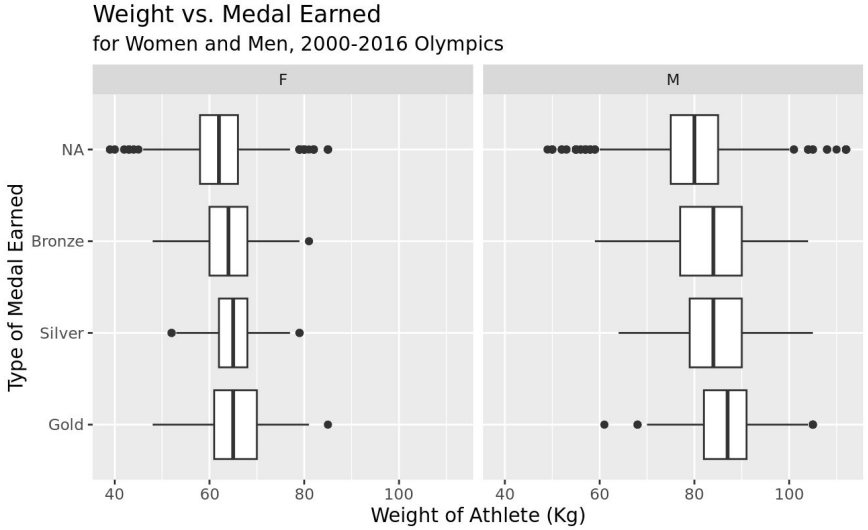
To learn more about the structure, summary statistics, and first few rows of the dataset, use R functions like str(), glimpse(), and head().

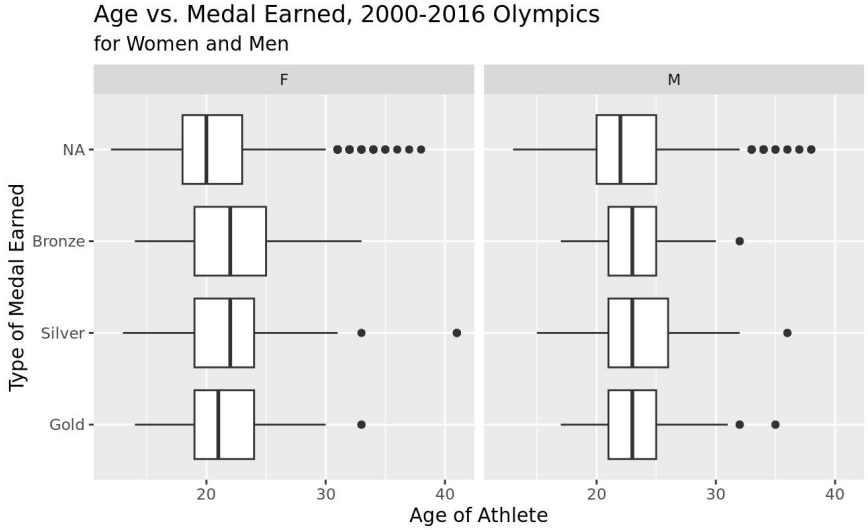
Exploratory Data Analysis




BMI vs. Medal Earned
for Women and Men, 2000-2016 Olympics
F
M
NA
Bronze
Silver
Gold
Type of Medal Earned
15
20
25
30
35
BMI of Athlete


Height vs. Medal Earned
for Women and Men, 2000-2016 Olympics
F
M
NA
Bronze
Silver
Gold
Type of Medal Earned
140
160
180
200
Height of Athlete (cm)


Weight vs. Medal Earned
for Women and Men, 2000-2016 Olympics
F
M
NA
Bronze
Silver
Gold
Type of Medal Earned
40
60
80
100
Weight of Athlete (Kg)


Age vs. Medal Earned, 2000-2016 Olympics
for Women and Men
F
M
NA
Bronze
Silver
Gold
Type of Medal Earned
20
30
40
Age of Athlete

Logistic Regression: Predicting Medals

```{r}

logistic_swim7 <- logistic_reg() |>
    set_engine("glm") |>
    fit(medalyesno ~ Height + Team,
        data = olympics3, family = "binomial")

glance(logistic_swim7)
tidy(logistic_swim7) |>
  knitr::kable()
```

$$\log\left(\frac{p}{1-p}\right) = -7.5364541 + 0.0401009 \times Height + [TeamCoef.]$$

$$p = \frac{\exp(-7.5364541 + 0.0401009 \times Height + [TeamCoef.])}{1 + \exp(-7.5364541 + 0.0401009 \times Height + [TeamCoef.])}$$

term	estimate
(Intercept)	-7.5364541
Height	0.0401009
TeamCanada	-2.4016185
TeamChina	-1.8478142
TeamFrance	-1.7962859
TeamGermany	-1.4709515
TeamGreat Britain	-3.1181233
TeamHungary	-2.0786095
TeamItaly	-1.6192694
TeamJapan	-2.5335523
TeamNetherlands	-1.5083302
TeamRussia	-1.9544265
TeamSouth Africa	-1.8619358
TeamSweden	-1.1646447
TeamUnited States	1.6256412

$p < 0.05$ except for Team Sweden

Conclusion

- Initial Hypothesis: The traits of nationality and height will impact Olympic swimmers' performance the most.
- Result: Logistic Model Height/Nationality produced the lowest AIC compared to other traits.
- Significance
- Future Research