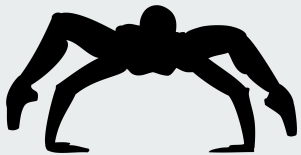
120 Years of Olympic History

Bayesian Statistics W19 Final Group Project

Jessica George, Shawna Tuli, Thanasi Bakis



Background

Historical dataset on Summer and Winter Olympic Games (Athens 1896 to Rio 2016)

How does a US Olympian's age, sex, height, & weight affect their chances of winning a medal?

Dataset

1051

120 years of Olympic history: athletes and results

basic bio data on athletes and medal results from Athens 1896 to Rio 2016

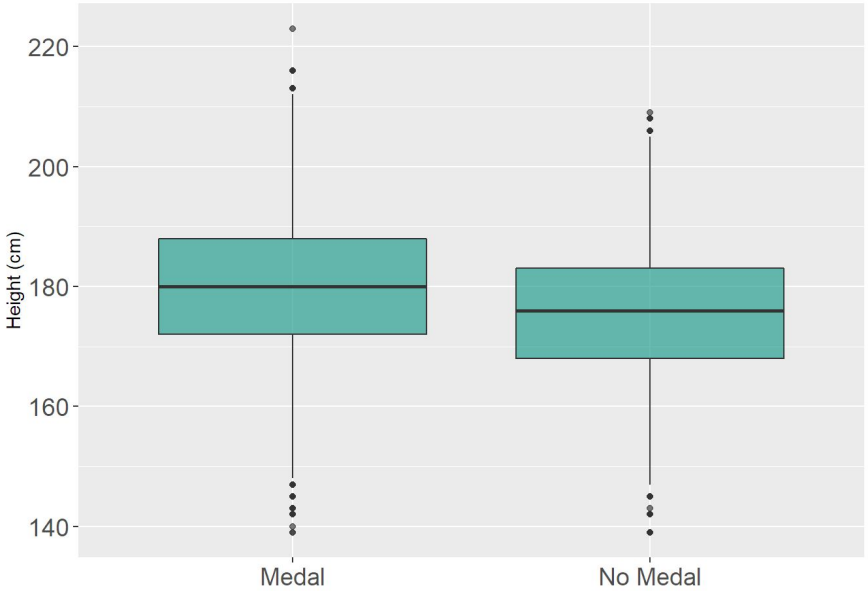
rgriffin • updated 2 years ago (Version 2)

Data | Tasks (1) | Kernels (154) | Discussion (6) | Activity | Metadata | Download (40 MB) | New Notebook

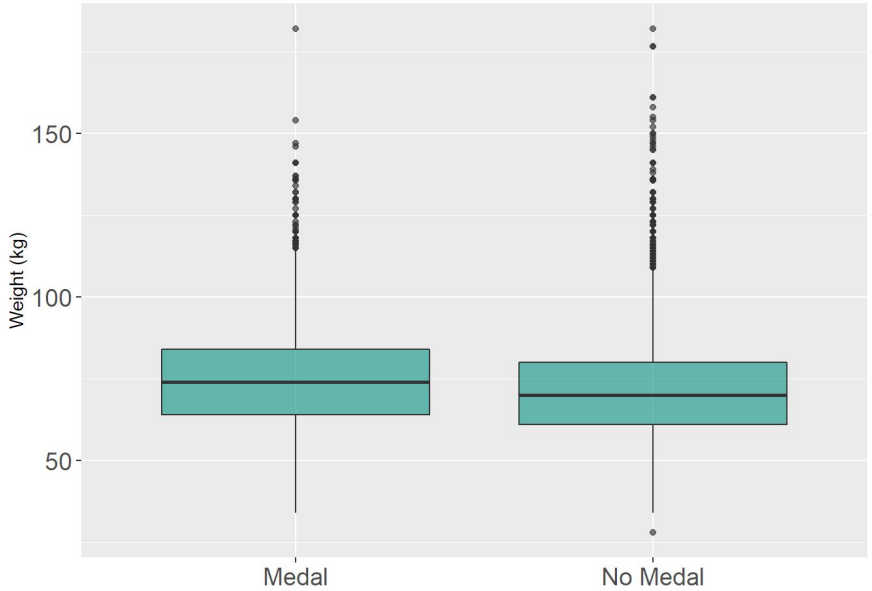
ID	Name	**Age *(yr)***
Sex *(M/F)*	**Height *(cm)***	**Weight *(kg)***
Team	NOC *(Country Code)*	Games *(Summer/Winter)*
Year	Season	City
Sport	Event	**Medal**

Exploratory Data Analysis

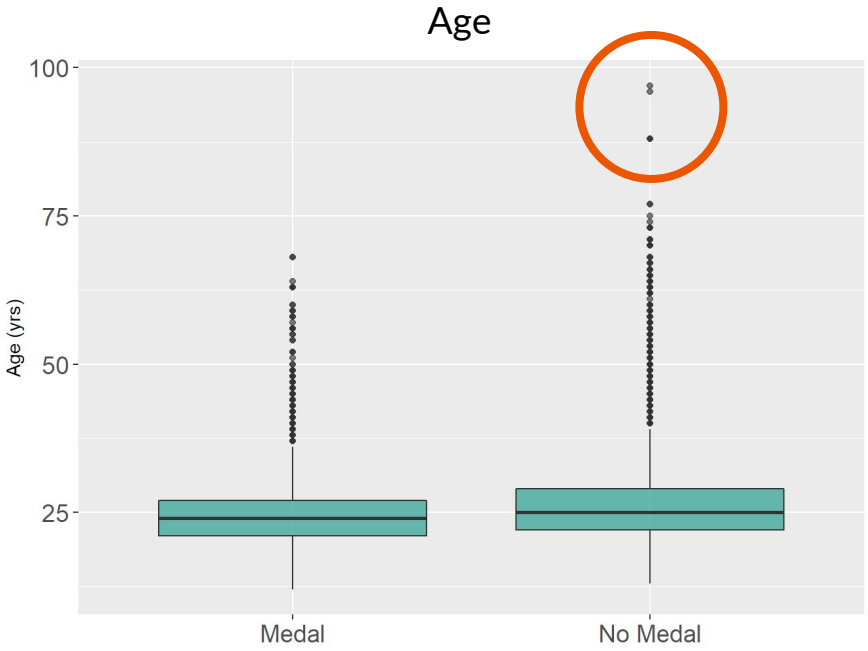
Height

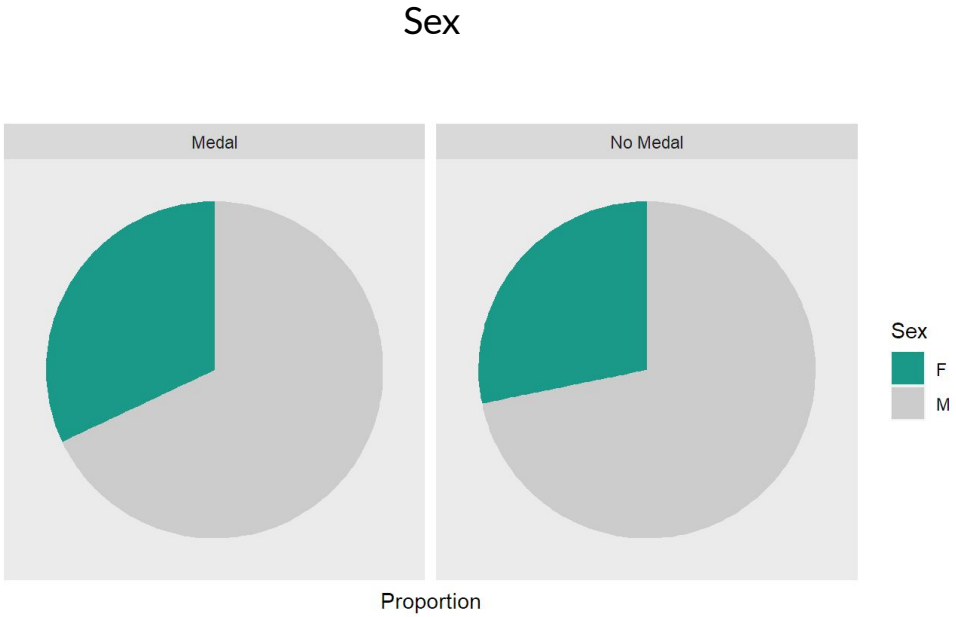


Weight



Exploratory Data Analysis





Wait... A 90-Year-Old Olympian?

```
USA %>%
    filter(Age > 90) %>%
    select(Name, Age, Year, Sport)
```

```
## # A tibble: 2 x 4
##   Name                    Age  Year Sport
##   <chr>                 <dbl> <dbl> <chr>
## 1 Winslow Homer            96  1932 Art Competitions
## 2 John Quincy Adams Ward   97  1928 Art Competitions
```

Fitting the Logistic Regression Model

$$\log(\frac{Y}{1-Y}) = \beta_0 + \beta_1 * Age + \beta_2 * I(Sex = M) + \beta_3 * Height + \beta_4 * Weight$$

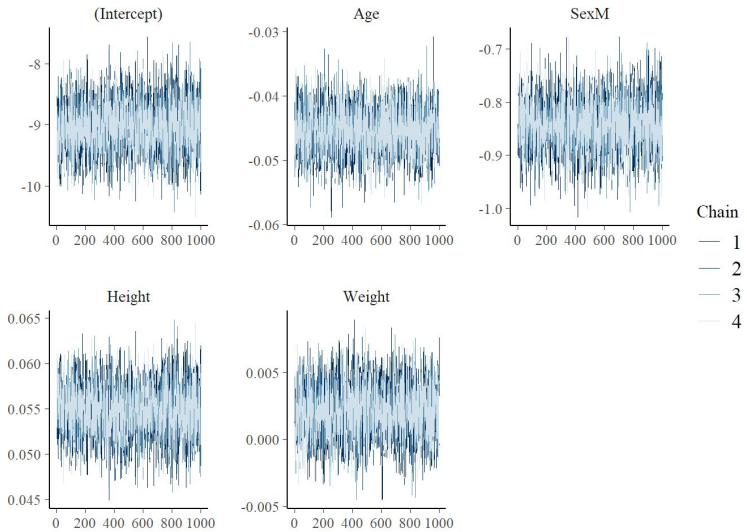
```r
normalize ← function(vec)
    (vec - mean(vec, na.rm = T)) / sd(vec, na.rm = T)

USA_normalized ← USA %>%
    mutate(Age    = normalize(Age),
           Height = normalize(Height),
           Weight = normalize(Weight))
```

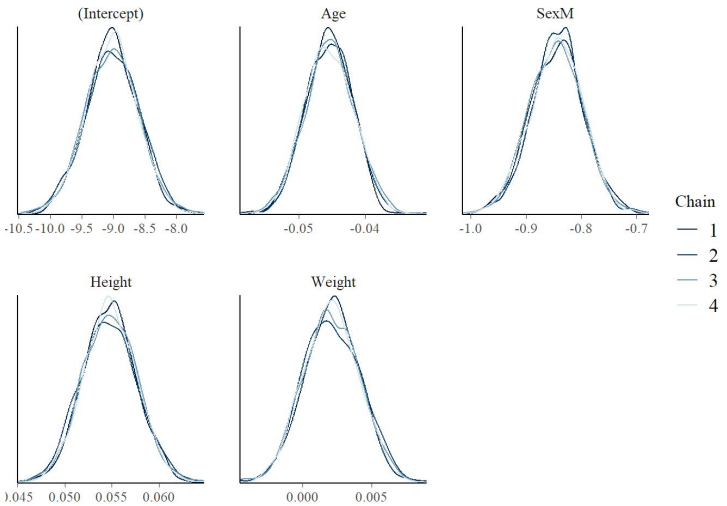
```r
model ← stan_glm(
    Won_Medal ~ Age + Sex + Height + Weight,
    data = USA,
    family = binomial(link = "logit")
)

model_normalized ← stan_glm(
    Won_Medal ~ Age + Sex + Height + Weight,
    data = USA_normalized,
    family = binomial(link = "logit")
)
```

```r
coef(model)
```

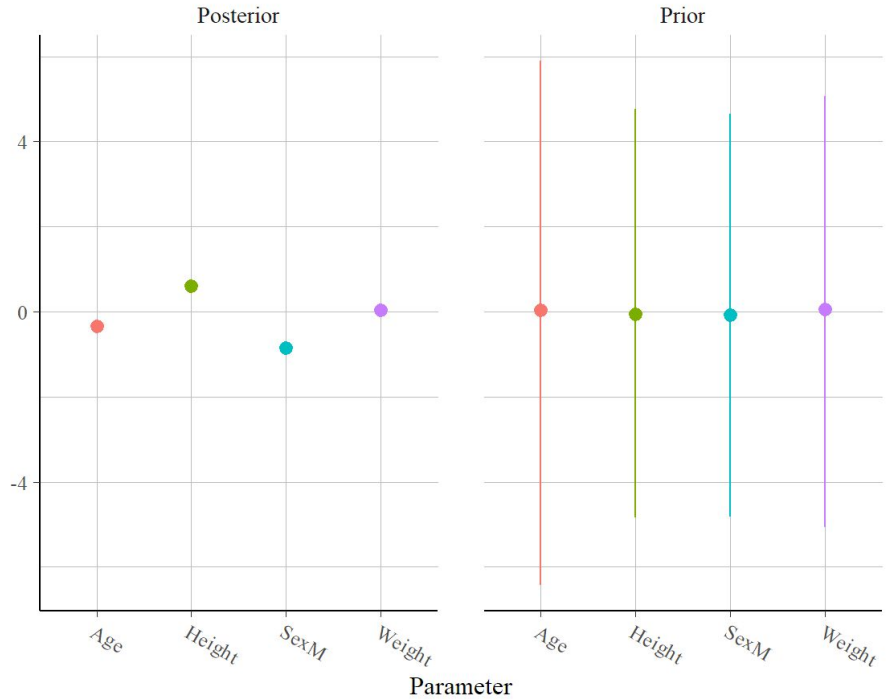
```
##  (Intercept)          Age         SexM       Height       Weight
## -9.024852149 -0.045428742 -0.845416285  0.054764970  0.002027598
```

Parameter Approximations





Parameter Distribution Intervals



Could We Win a 🏅 Too?

```
pred ← function(Age, Sex, Height, Weight)
{
    y_dist ← posterior_predict(model, newdata = data.frame(Age, Sex, Height, Weight))

    mean(y_dist) ⩾ 0.5
}
```

Thank You!