



R Code for Examples in the book  
*"Statistics: The Art and Science of Learning from Data"*  
 by Agresti, Franklin and Klingenberg, 5<sup>th</sup> edition

## Chapter 3

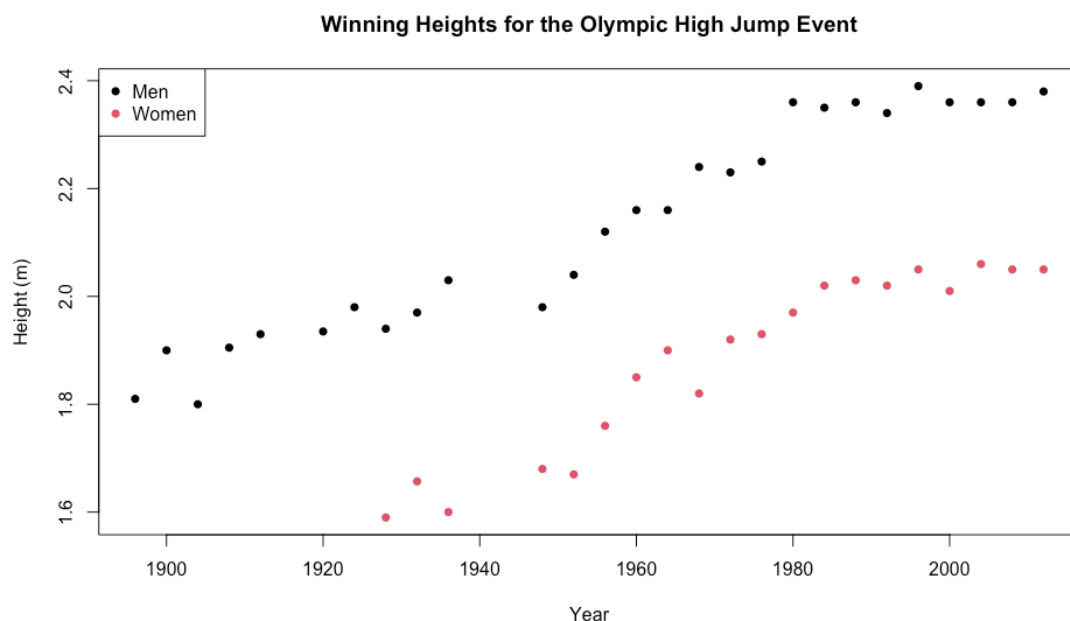
### Example 12: Olympic High Jumps – Exploring Multivariate Relationships

#### Reading in the data

```
heights <-  
read.csv(file='https://raw.githubusercontent.com/artofstat/data/master/Chapter3/high_jump.csv')  
attach(heights) # so we can refer to variable names
```

#### Basic scatterplot

```
plot(x = Year, y = Winning.Height..m., pch = 16,  
     col = factor(Gender),  
     main = 'Winning Heights for the Olympic High Jump Event',  
     xlab = 'Year', ylab = 'Height (m)')  
  
legend("topleft",  
      legend = levels(factor(Gender)),  
      pch = 16,  
      col = factor(levels(factor(Gender))))
```



## Separating observations for men and women

```
men_observations <- subset(heights, Gender == 'Men')
women_observations <- subset(heights, Gender == 'Women')
```

## Fitting in regression model for observations for men and women

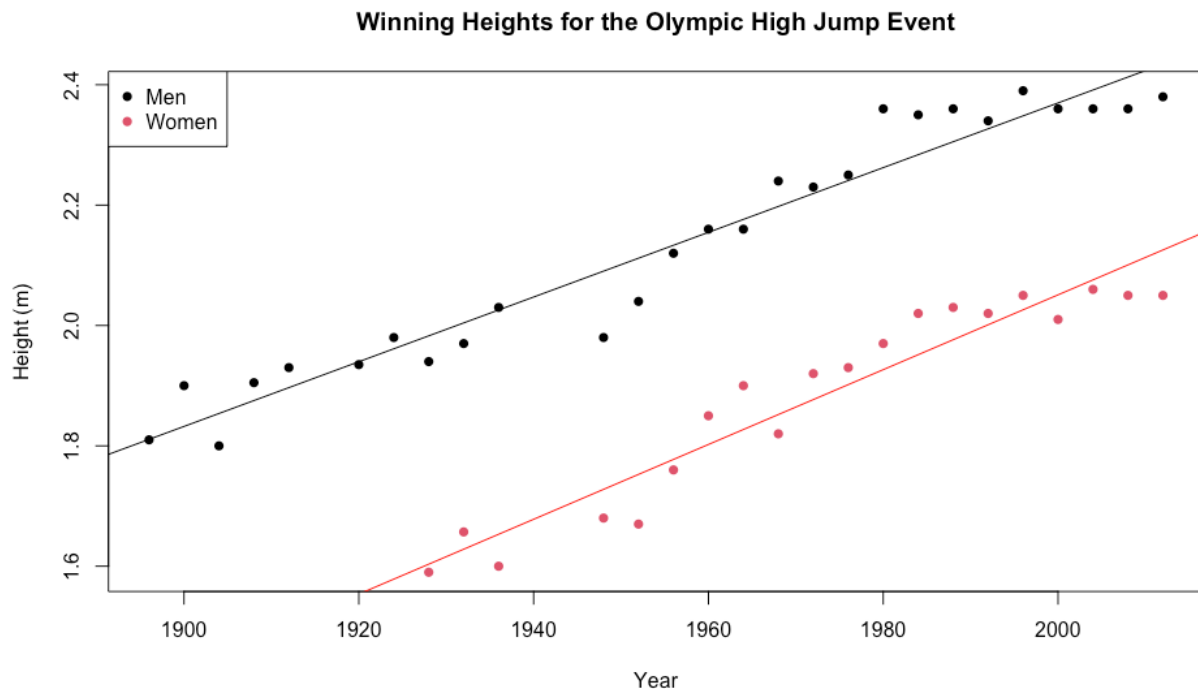
```
lm_men <- lm(Winning.Height..m. ~ Year, data = men_observations)
lm_women <- lm(Winning.Height..m. ~ Year, data = women_observations)
```

## Adding the regression equations to the plot

```
plot(x = Year, y = Winning.Height..m., pch = 16,
     col = factor(Gender),
     main = 'Winning Heights for the Olympic High Jump Event',
     xlab = 'Year', ylab = 'Height (m)')
```

```
legend("topleft",
      legend = levels(factor(Gender)),
      pch = 16,
      col = factor(levels(factor(Gender))))
```

```
abline(lm_men, col = 'black')
abline(lm_women, col = 'red')
```



## Scatterplot using ggplot2

```
library(ggplot2)
ggplot(heights,
       aes(x = Year, y = Winning.Height..m.)) +
  geom_point(aes(shape = Gender, color = Gender)) +
  geom_smooth(method=lm, se=FALSE, fullrange= TRUE,
             aes(color=Gender)) +
  labs(title = 'Winning Heights for the Olympic High Jump Event',
       x = 'Year', y = 'Height (m)') +
  theme_bw() +
  scale_x_continuous(limits = c(1890,2030), breaks = seq(1900,2020,20))
```

