**R Code for Examples in the book**



***“Statistics: The Art and Science of Learning from Data”***

**by Agresti, Franklin and Klingenberg, 5th edition**

**Chapter 3**

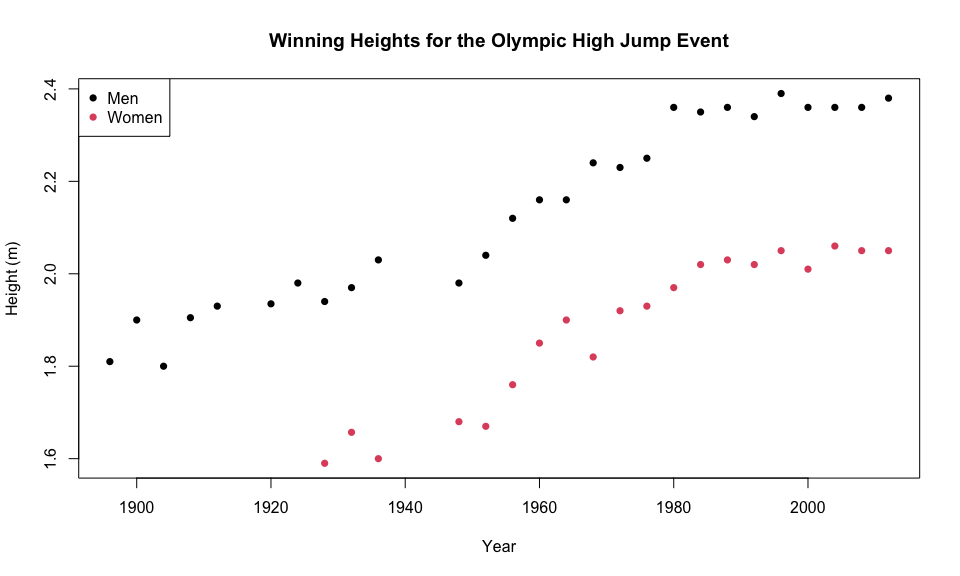
**Example 12: Olympic High Jumps – Exploring Multivariate Relationships**

## 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

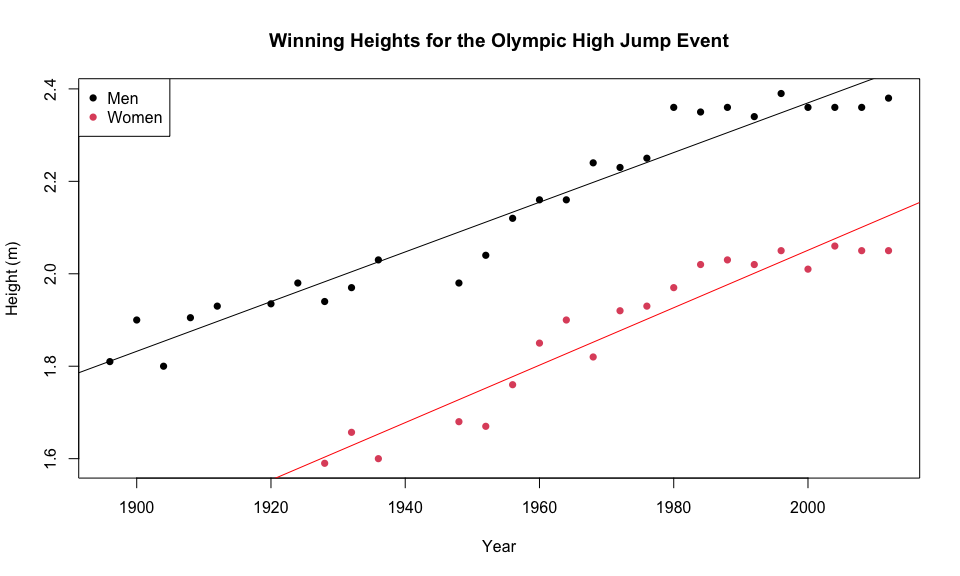
men\_observations <- subset(heights, Gender == 'Men')  
women\_observations <- subset(heights, Gender == 'Women')

## 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))

