



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 2

### Example 14: Female Student Heights – Empirical Rule

#### Reading in values from file:

```
student_heights <-  
read.csv(file='http://www.artofstats.com/data/chapter2/heights.csv')  
attach(student_heights) # so we can refer to variable names
```

The original dataset contains height measurements for men and women. You can use the `subset()` function to filter out height measurements for men and omit the measurement of 92 inches.

```
heights_women <- subset(HEIGHT, GENDER == 'Female' & HEIGHT != 92)
```

#### Sample Size

```
length(heights_women)
```

```
## [1] 261
```

#### Mean

```
mean(heights_women)
```

```
## [1] 65.28352
```

#### Standard Deviation

```
sd(heights_women)
```

```
## [1] 2.952847
```

#### 5 Number Summary

```
summary(heights_women)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##   56.00   64.00   65.00   65.28   67.00   77.00
```

## Creating Histogram using ggplot2

```
library(ggplot2)
ggplot(data.frame(heights_women),
       aes(x = heights_women,
           y = 100 * (..count.. / sum(..count..)))) +
  geom_histogram(center = 0, binwidth = 1, color = 'black', fill = 'tan') +
  labs(x = 'Height (in)', y = 'Percent (%)',
       title = 'Histogram of Female Student Heights') +
  theme_bw() +
  scale_y_continuous(limits = c(0,20), breaks = seq(0,16,4), expand = c(0,0))
+
  scale_x_continuous(breaks = seq(56,78,2)) +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank())
)
```

