**R Code for Examples in the book**



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

**by Agresti, Franklin and Klingenberg, 5th 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()  
 )

