my\_solutions

Hunter GIles

3/6/2022

# Assignment 1: Getting Started

### Question 1

3 + 4

## [1] 7

(-3 + 5 + 7 + 8)/4

## [1] 4.25

-3 + 5 + 7 + 8/4

## [1] 11

We can see that R has built in basic calculator functions. Also, R supports ‘order of operation’ syntax.

### Question 2

# assigning vector to variable1  
variable1 <- c(-3, 5, 7, 8)  
  
# running basic methods on vector for mean and standard deviation  
mean(variable1)

## [1] 4.25

sd(variable1)

## [1] 4.99166

Here we can see a numeric vector being assigned to variable1. This vector is saved in the environment for future use. We can use the built in functions *mean()* and *sd()*.

### Question 3

library(readr)  
library(mosaic)

Installing the necessary dependencies for the remainder of the assignment.

### Question 4

# assigning readmill data to hunterData  
hunterData <- read\_csv("Data/treadmill.csv")  
  
# viewing the first and and last 6 observations  
head(hunterData)  
tail(hunterData)

Above I ran *read\_csv()* to download a data-frame. Then I used *head()* and *tail()* to examine a piece of my data-frame.

### Question 5

# Basic mean and standard deviation of the RunTime column  
mean(hunterData$RunTime)

## [1] 10.58613

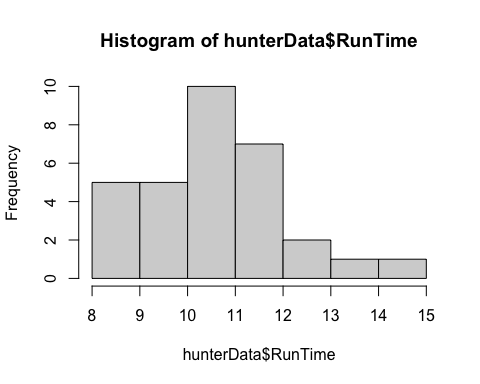
sd(hunterData$RunTime)

## [1] 1.387414

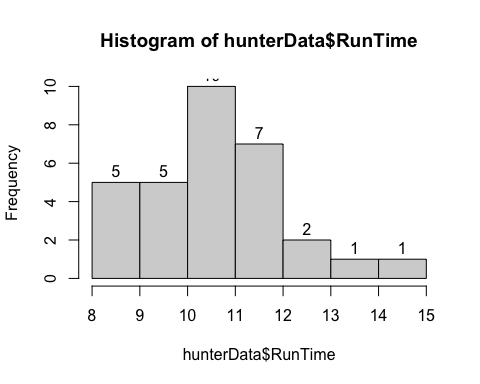
The chunk above displays the mean and standard deviation of the *RunTime* column. The $ is a characteristic of a tibble. From this we can receive a specific column (vector) of a tibble by calling dataframe$column.

### Question 6

# min, Q1, median, Q3, max, mean, std, n, missing  
favstats(hunterData$RunTime)  
  
# histogram  
hist(hunterData$RunTime)



hist(hunterData$RunTime, labels=TRUE)



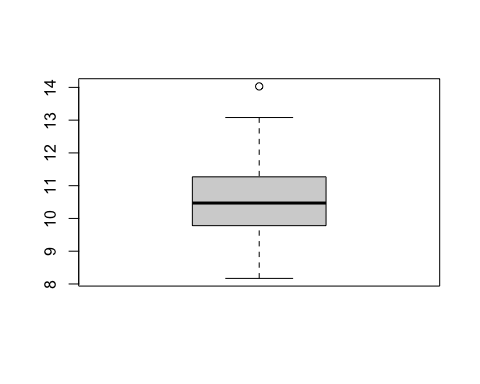
Above returns 1 table and 2 histograms. This table displays descriptive statistics of RunTime. The histograms display the distribution of Runtime. The last graph includes the label argument.

### Question 7

# difference between Q1 and Q3 for RunTime  
IQR <- 11.37 - 9.78  
  
# the upper limit for outliers  
11.27 + 1.5\*IQR

## [1] 13.655

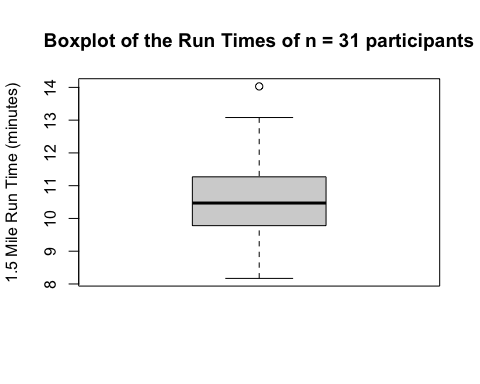
# boxplot  
boxplot(hunterData$RunTime)



Boxplots are also useful to display data distribution. Instead of frequency, boxplots show quartile rage. Boxplots treat outliers differently. If an observation is 1.5 IQRs above Q3, then it is an upper limit outlier.

### Question 8

boxplot(hunterData$RunTime, ylab = "1.5 Mile Run Time (minutes)",  
 main = "Boxplot of the Run Times of n = 31 participants")



A picture is only worth a thousand words, if well labeled.