Chapter 1

2024-08-19

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

This is for the review from chapter 1

library(readr)

data

treadmill <- read\_csv('http://www.math.montana.edu/courses/s217/documents/treadmill.csv')

## Rows: 31 Columns: 8  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (8): Subject, TreadMillOx, TreadMillMaxPulse, RunTime, RunPulse, RestPul...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

treadmill

## # A tibble: 31 × 8  
## Subject TreadMillOx TreadMillMaxPulse RunTime RunPulse RestPulse BodyWeight  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 60.0 186 8.63 170 48 81.9  
## 2 2 59.6 172 8.17 166 40 68.2  
## 3 3 54.6 155 8.92 146 48 70.9  
## 4 4 54.3 168 8.65 156 45 85.8  
## 5 5 51.8 170 10.3 166 50 83.1  
## 6 6 50.6 155 9.93 148 49 59.1  
## 7 7 50.5 168 10.1 168 45 73.0  
## 8 8 50.4 168 10.1 168 67 73.4  
## 9 9 49.9 180 9.22 178 55 89.0  
## 10 10 49.2 185 8.95 180 44 81.4  
## # ℹ 21 more rows  
## # ℹ 1 more variable: Age <dbl>

downloading his package

library(remotes)  
remotes::install\_github('greenwood-stat/catstats2')

## Using GitHub PAT from the git credential store.

## Skipping install of 'catstats2' from a github remote, the SHA1 (02be2a5c) has not changed since last install.  
## Use `force = TRUE` to force installation

library(catstats2)  
data(treadmill)  
  
library(mosaic)

## Registered S3 method overwritten by 'mosaic':  
## method from   
## fortify.SpatialPolygonsDataFrame ggplot2

##   
## The 'mosaic' package masks several functions from core packages in order to add   
## additional features. The original behavior of these functions should not be affected by this.

##   
## Attaching package: 'mosaic'

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following object is masked from 'package:Matrix':  
##   
## mean

## The following object is masked from 'package:ggplot2':  
##   
## stat

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,  
## quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

explore with this package ####

data <- treadmill

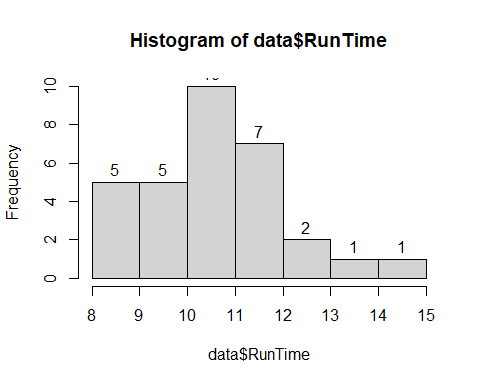
summary stats

favstats(data$RunTime)

## min Q1 median Q3 max mean sd n missing  
## 8.17 9.78 10.47 11.27 14.03 10.58613 1.387414 31 0

visuals

hist(data$RunTime)  
hist(data$RunTime, labels = T)



outlier? the standard for base R boxplot is that an outlier is 1.5 times greater than the IQR (inter-quartile range) IQR = Q3 - Q1

IQR <- 11.27 - 9.78  
IQR

## [1] 1.49

11.27 + 1.5\*IQR

## [1] 13.505

it appears that the max of 14.03 is an outlier

boxplot(data$RunTime,  
 ylab = '1.5 Mile Run Time (minutes)',  
 main = 'Boxplot of run times of n=31')

