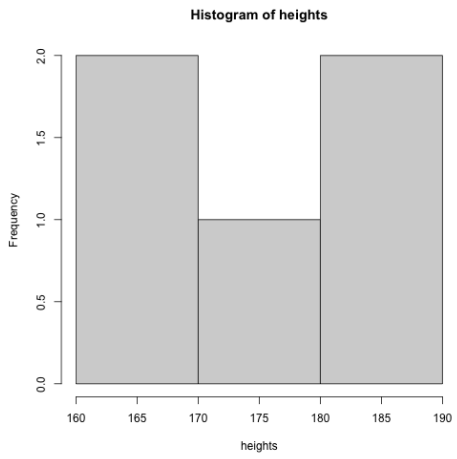
# Chapter 1

Example 1: Let's manually enter the heights data set (183, 170, 160, 175, 187) from chapter 1 and the following R commands will do the calculations we completed by hand.
Note: Lines of code that begin with # are comments and are not executed by the computer. Lines of code that begin with # denote the output.

```
# manually enter your data by using the "c" command
heights = c(183,170,160,175,187)

# calculate sample statistics on data set
mean(heights)

## [1] 175

median(heights)

## [1] 175

var(heights)

## [1] 114.5

sd(heights)

## [1] 10.70047
```

We can also create a histogram in R. Note that we can specify the breaks for the bins. In this case, the breaks are defined at 160, 170, 180, 190.

```
hist(heights, breaks = c(160, 170, 180, 190))
```



Example 2: Let's create a randomized dataset of eye colors and analyze the data.

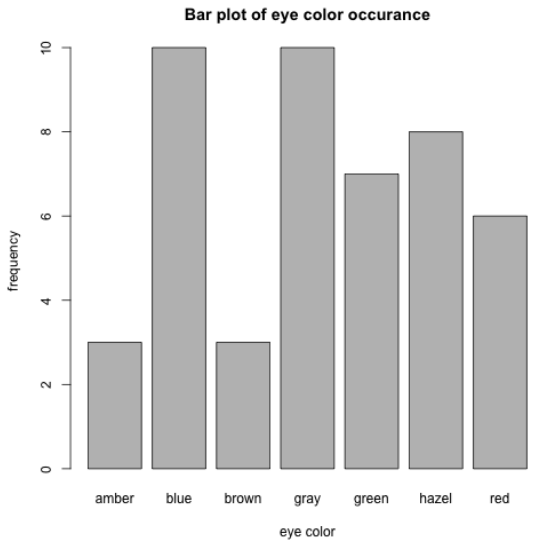
Note: When we generate random data in this class, we will use the **set.seed** function, so that we all generate the same random data. If we don't use the **set.seed** command before generating random data, we will all get different random data.

Here, we generate 47 random eye colors and take a look at the first few observations.

```
set.seed(2020)
eye.colors <- sample(c("amber", "blue", "brown", "gray", "green", "hazel",
    "red"), 47, replace = T)

# the 'head' function will give the first few data values or the
# 'header' of the dataset
head(eye.colors)

## [1] "gray"  "gray"  "red"   "hazel" "amber" "amber"
```

We can create a frequency table of these values and create a barplot.

```
table(eye.colors)

## eye.colors
## amber  blue brown  gray green hazel   red
##     3    10     3    10     7     8     6

barplot(table(eye.colors), xlab = "eye color",
        ylab = "frequency", main = "Bar plot of eye color occurance")
```



We can also calculate proportion and percentage tables.

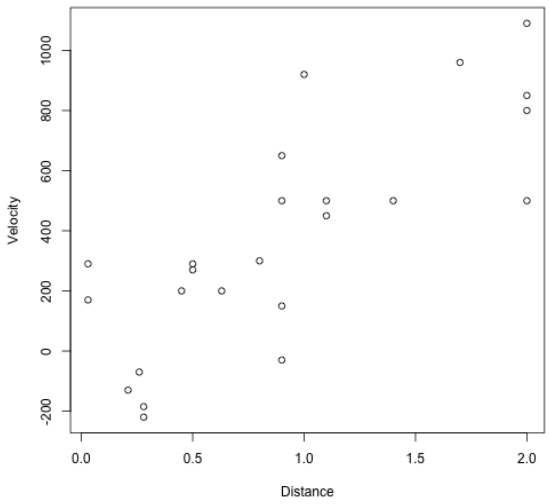
```
table(eye.colors) # shows how many of each value exist

## eye.colors
## amber  blue brown  gray green hazel   red
##     3    10     3    10     7     8     6

length(eye.colors) # tells how many data points there are

## [1] 47

table(eye.colors)/length(eye.colors) # proportion

## eye.colors
##      amber       blue      brown       gray      green      hazel
## 0.06382979 0.21276596 0.06382979 0.21276596 0.14893617 0.17021277
##        red
## 0.12765957

table(eye.colors)/length(eye.colors)*100 # percent

## eye.colors
##     amber      blue     brown      gray     green     hazel       red
##  6.382979 21.276596  6.382979 21.276596 14.893617 17.021277 12.765957
```

Example 3: We can easily load datasets into R Studio. Let's load **hubble.csv** which are data on distances and velocities of 24 galaxies containing Cepheid stars, from the Hubble space telescope key project to measure the Hubble constant.

First, click **File** in the menu bar, then **Import Dataset**, then **From Text (base)**, and finally find and select **hubble.csv**.

Let's plot distance versus velocity. Notice that we can identify a possible relationship between the two variables in the plot.

```
plot(Velocity~Distance,data=hubble)
```




If you want to examine one of the variables defined under the **hubble** dataset, you can use the **$** to refer to the variable of interest. For example, we can calculate the mean of Distance and variance of Velocity as follows.

```
mean(hubble$Distance)

## [1] 0.91125

var(hubble$Velocity)

## [1] 137830
```