STAT511 HW1

Amy Fox

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.0       v purrr   0.3.1
## v tibble  2.0.1       v dplyr   0.8.0.1
## v tidyr   0.8.3       v stringr 1.4.0
## v readr   1.3.1       v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Question 1: Describe the data in Problem 3.3

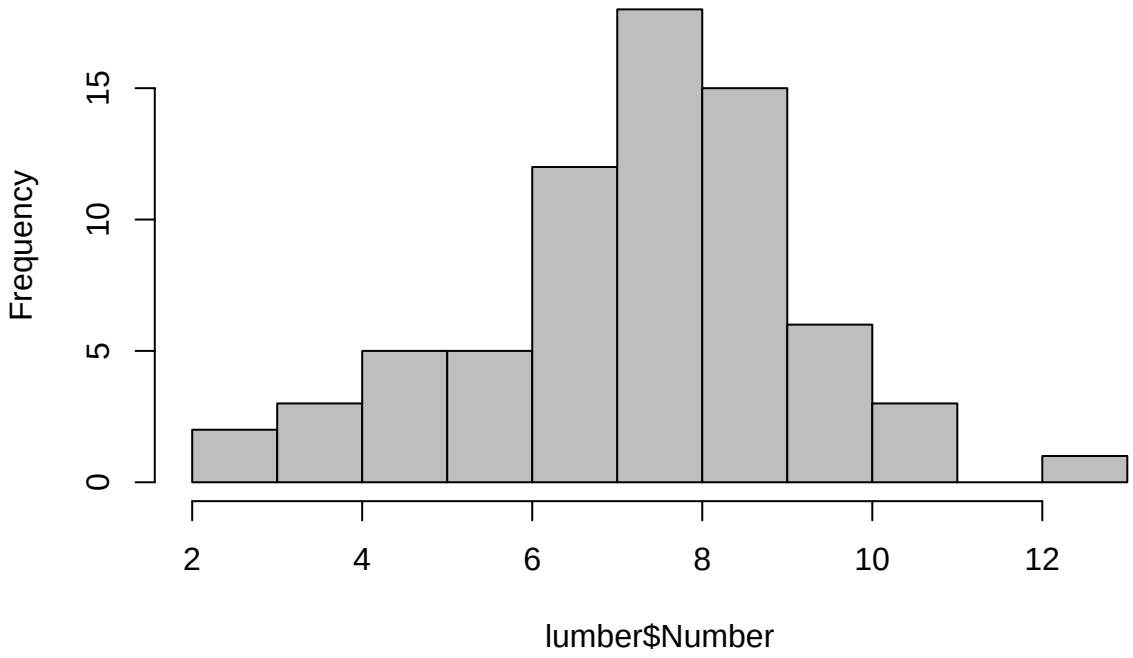
Code

```
# read in the data
lumber <- read.csv("../Data/OTT_Final/ASCII-comma/CH03/ex3-30.txt", quote = "'")

# look at the structure of the data
str(lumber)
```

```
## 'data.frame':    70 obs. of  1 variable:
##  $ Number: int  7 8 6 4 9 11 9 9 9 10 ...
```

```
# print the histogram
hist(lumber$Number, col = "grey", main = "Question 1A: Lumber Histogram")
```




```
# find the average of the lumber data
mean(lumber$Number)
```

```
## [1] 7.728571
```

```
# find the median of the lumber data
median(lumber$Number)
```

```
## [1] 8
```

Answer

A) The histogram is shown above.
B) The mean for the lumber data is 7.7285714. The median for the lumber data is 8.
C) Based on the histogram, it appears as though the data is normally distributed because the data appears as though there is centered around a number (7 in this plot). Also, the rest of the data does not seem to be left- or right- skewed.

Question 2: Describe the data in Problem 3.7

Code

```
# read in the survival data
survival <- read.csv("../Data/OTT_Final/ASCII-comma/CH03/ex3-7.txt", quote = "'")

# look at the structure of the data
str(survival)
```

```
## 'data.frame':    28 obs. of  2 variables:
##  $ StandardTherapy: int  4 14 29 6 15 2 6 13 24 16 ...
##  $ NewTherapy     : int  5 17 27 9 20 15 14 18 29 19 ...
```

```r
# print the median and standard deviation of the Standard therapy
mean(survival$StandardTherapy)
```

```
## [1] 15.67857
```

```r
sd(survival$StandardTherapy)
```

```
## [1] 9.630405
```

```r
# print the median and standard deviation of the New therapy
mean(survival$NewTherapy)
```

```
## [1] 20.71429
```

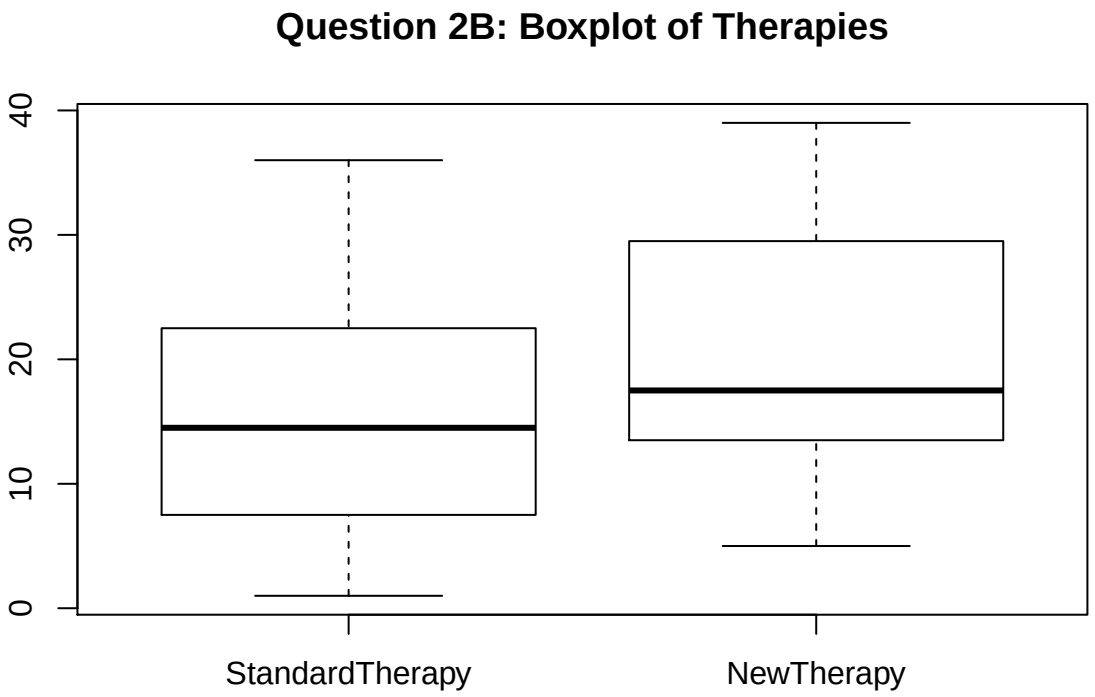
```r
sd(survival$NewTherapy)
```

```
## [1] 9.808753
```

```r
# tidy data for median and mean of both of the therapies
survival %>%
  gather(key = Treatment, value = data) %>%
  group_by(Treatment) %>%
  summarise(mean = mean(data),
            sd = sd(data))
```

```
## # A tibble: 2 x 3
##   Treatment        mean    sd
##   <chr>           <dbl> <dbl>
## 1 NewTherapy       20.7  9.81
## 2 StandardTherapy  15.7  9.63
```

```r
# print the boxplot
boxplot(survival, main= "Question 2B: Boxplot of Therapies")
```



Answer

A) The Standard Therapy mean is 15.6785714 and standard deviation is 9.6304054. The New Therapy mean is 20.7142857 and standard deviation is 9.8087532.

B) The boxplot is shown above.