STAT312 R Lab 2

Max Tjen

7/8/21

Problem 1:

The energy consumption for 90 gas-heated homes during a winter heating season is given in the file "Energy_Usage.csv" (see Data module on canvas). The variable reported is BTU/number of heating degree day.

a)

Use R codes to find the five point summary, mean and standard devaition of the data. (4 points)

```
usage = read.csv('Energy_Usage.csv')
summary(usage)
```

```
##        BTU
##  Min.   : 2.970
##  1st Qu.: 7.947
##  Median : 9.835
##  Mean   :10.038
##  3rd Qu.:12.045
##  Max.   :18.260
```

```
sd(usage$BTU)
```

```
## [1] 2.86799
```

Ans:
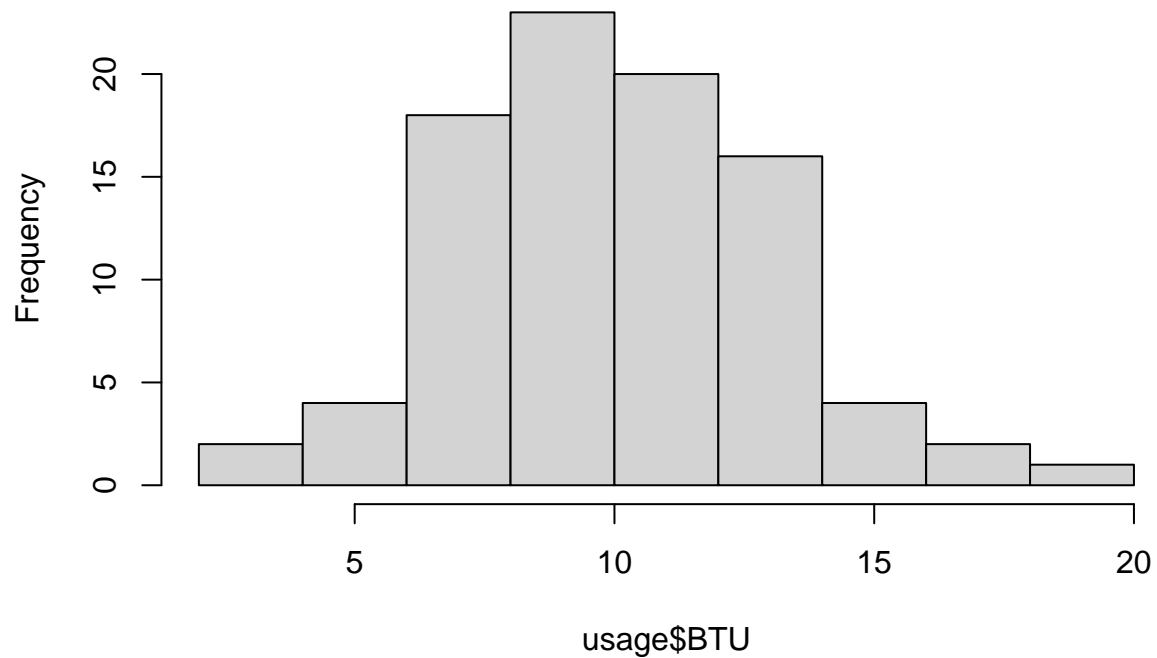5 point summary: min = 2.970, Q1 = 7.947, Median/Q2 = 9.835, Q3 = 12.045, max = 18.260
mean: 10.038
standard deviation: 2.868

b)

Construct a histogram and a stem-and-leaf diagram of energy usage. Use appropriate scale for the stem-and-leaf diagram. Then use SOCS method to describe the data. (6 points)
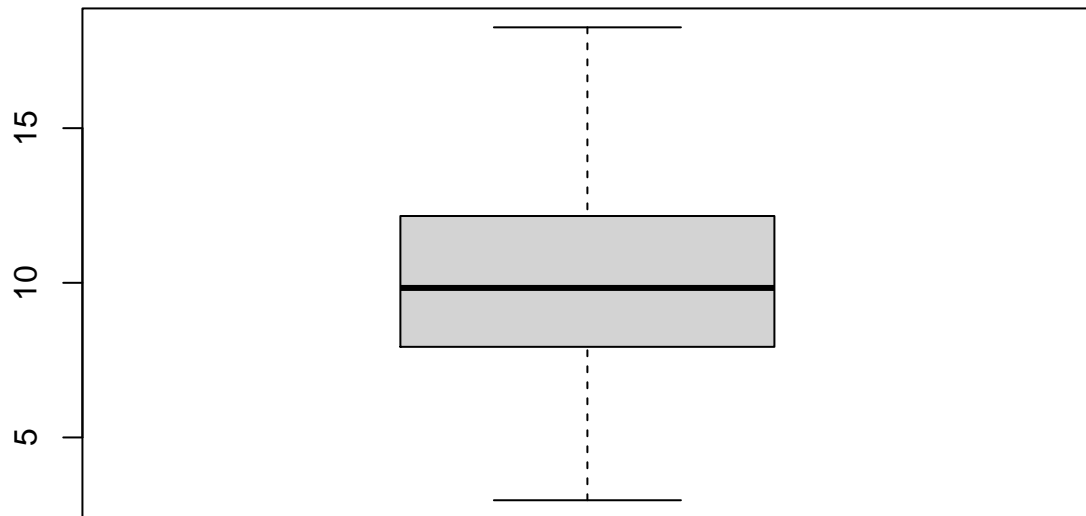
```
hist(usage$BTU)
```

Histogram of usage$BTU


```r
stem(usage$BTU, scale = 2)
```

```
##
##   The decimal point is at the |
##
##    2 |
##    3 | 0
##    4 | 0
##    5 | 269
##    6 | 04678899
##    7 | 2223667799
##    8 | 03345566778
##    9 | 134456688888
##   10 | 002333444556
##   11 | 011234677
##   12 | 223367799
##   13 | 144456
##   14 | 024
##   15 | 12
##   16 | 19
##   17 |
##   18 | 3
```

```r
boxplot(usage)
```

```r
range(usage)
```

```
## [1]  2.97 18.26
```

```r
range(usage)[2] - range(usage)[1]
```

```
## [1] 15.29
```

Ans:
The energy usage data is right skewed as while the histogram looks relatively symmetrical, the mean is greater than the median. There are not any outliers in the data and the center values of it consist of a mean = 10.038 and a median of 9.835. For the spread of the data, the range is 15.29, the standard deviation is 2.868, and the quartile values are Q1 = 7.947, Q2 = 9.835, Q3 = 12.045