Homework One

Rachel Meredith

1/26/2022

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

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

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(datasets)
```

R Markdown

Q14

a.) My Data set

```
data_set <- read.csv("C:/Users/rache/Documents/STAT 3010/Hw1_Q14_data.csv", header = T)
```

Stem Plot

```
stem(data_set$shower_flow_rate)
```

```
##
##   The decimal point is at the |
##
##    2 | 23
##    3 | 2344567789
##    4 | 01356889
##    5 | 00001114455666789
##    6 | 000012222334445666778999 9
##    7 | 00012233455555668
##    8 | 02233448
##    9 | 012233335666788
##   10 | 2344455688
##   11 | 2335999
##   12 | 37
##   13 | 8
##   14 | 36
##   15 | 0035
##   16 |
##   17 |
##   18 | 9
```

capture.output(stem(data_set$score)) file <- "C:/Users/rache/Documents/STAT 3010/Ex.1.2_Q14_data.txt"

b.) Typical flow rate is the flow rate that appears the most. So we would take one of the values from 6. ex = 6.7

c.) Highly concentrated, only one outlier.

d.) distribution is left skewed since there are more values below the typical flow rate.

e.) Outlier would be 18.9.

22.)

a.) sample size = 90 + 190 + 180 + 160 + 120 + 80 + 60 + 40 + 30 + 20 = 970

runners = 10

p = (# of runners) / (sample size)

p = 10 / 970 = 0.01

24.)

My data set

```
setwd("C:/Users/rache/Documents/STAT 3010")
data_set2 <- read.csv("Hw1_Q24_data.csv")

head(data_set2)
```

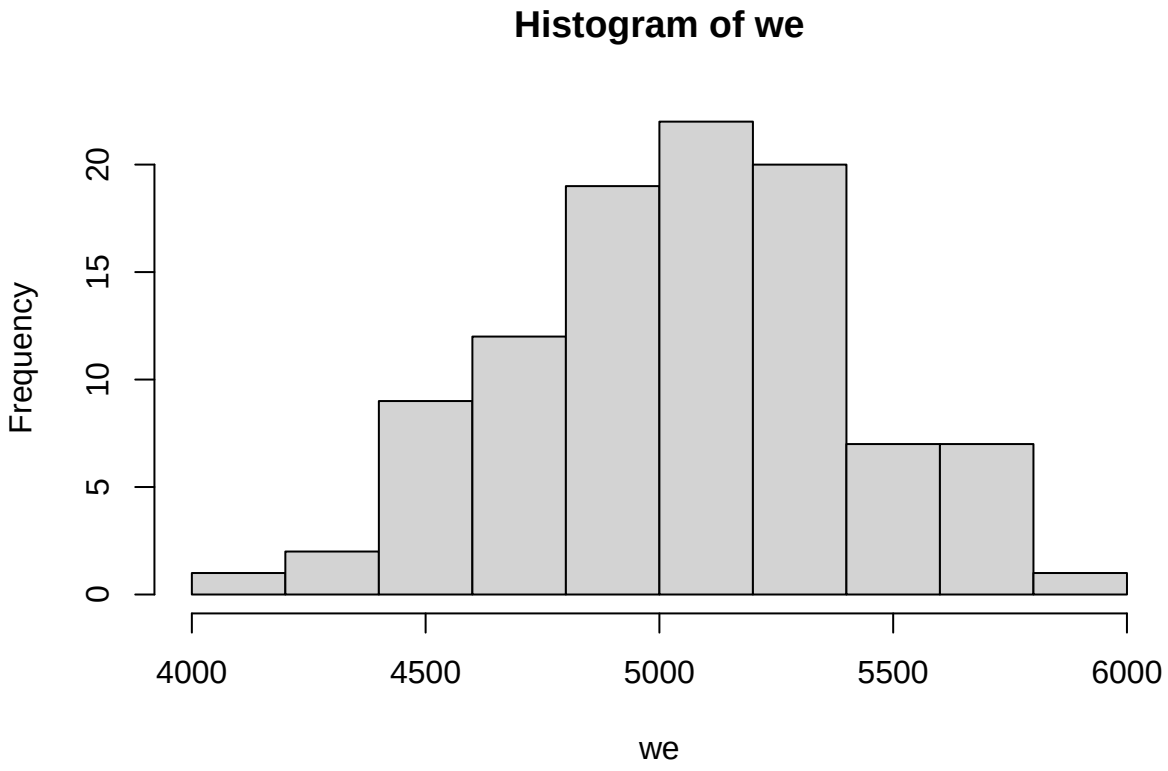
```
##   shear_strength
## 1           5434
## 2           5112
```

```
## 3           4820
## 4           5378
## 5           5027
## 6           4848
```

```
we = data_set2[,1]
```

My Histogram

```
hist(we)
```



34.)

```
setwd("C:/Users/rache/Documents/STAT 3010")
data_set3 <- read.csv("Hw1_Q34_data.csv")
```

a.) Sample mean homes = (6 + 5 + 11 + 33 + 4 + 5 + 80 + 18 + 35 + 17 + 23)/11= 21.55EU/mg

b.) farm homes = (2+15+12+8+8+7+6+19+3+9.8+22+9.6+2+2+0.5)/15 = 8.39

c.) ascending order 4,5,5,6,11,17,18,23,33,35,80 n = 11 ((11+1)/2)th position = median is 17 EU/mg for urban homes.

ascending order 0.5,2,2,2,3,6,7,8,8,9.6,12,15,19,22 n=15 ((15+1)/2)th position = median is 8 EU/mg for farm houses.

c.) trimmed mean = (6+5+11+33+5+18+35+17+23)/9= 17.00 for urban houses

trimmed mean = (2+15+12+8+8+7+6+19+3+9.8+9.6+2+2)/13 = 7.95 for farm houses.

(1 * 100)/11= 9.09% <-urban home trimming percentage (1 * 100)/15 = 6.67% <- farmhouse trimming percentage

35.)

```
data_set4 <- read.csv("C:/Users/rache/Documents/STAT 3010/Hw1_Q35_data.csv.xls")

# a.)

x <- c(data_set4)
result.mean <- mean(data_set4,trim = 0)
```

```
## Warning in mean.default(data_set4, trim = 0): argument is not numeric or
## logical: returning NA
```

```
print(result.mean)
```

```
## [1] NA
```

53

a.) n is odd so median is 2.74

Upper fourth meadian is 3.88

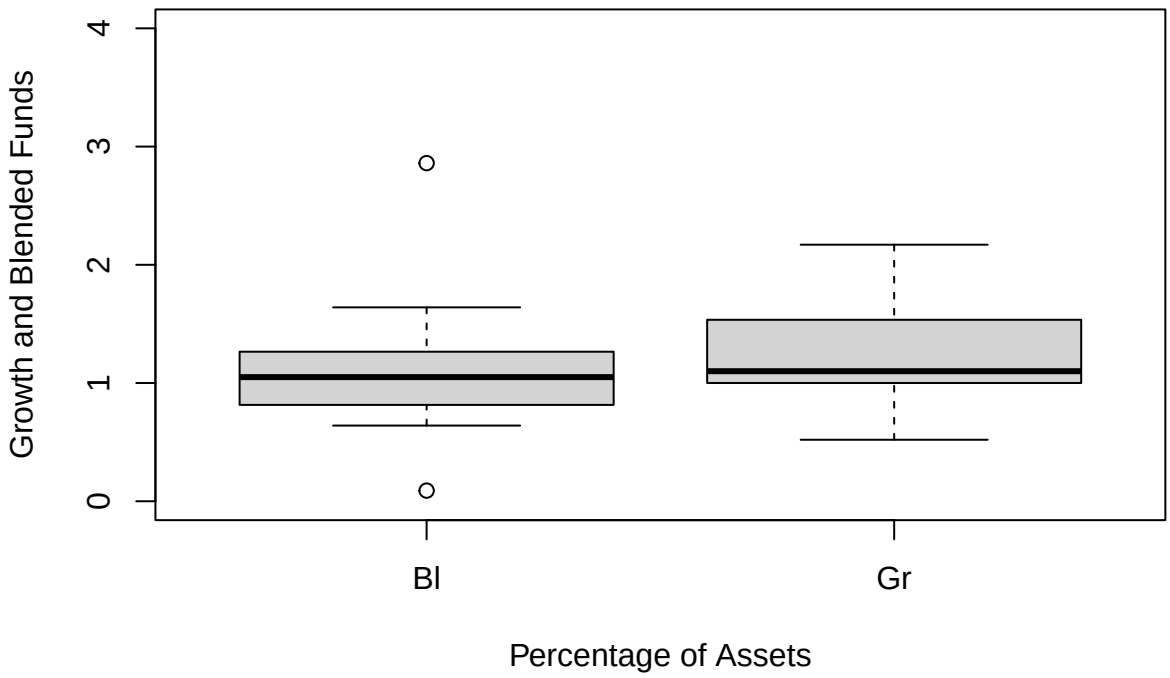
3.88 - 2.74 = 1.14

b.) my data set

```
data_set53 <- read.csv("C:/Users/rache/Documents/STAT 3010/Hw1_Q53_data.csv")
```

boxplot

```
boxplot(data_set53, xlab= "Percentage of Assets", ylab = "Growth and Blended Funds", ylim= c(0,4))
```

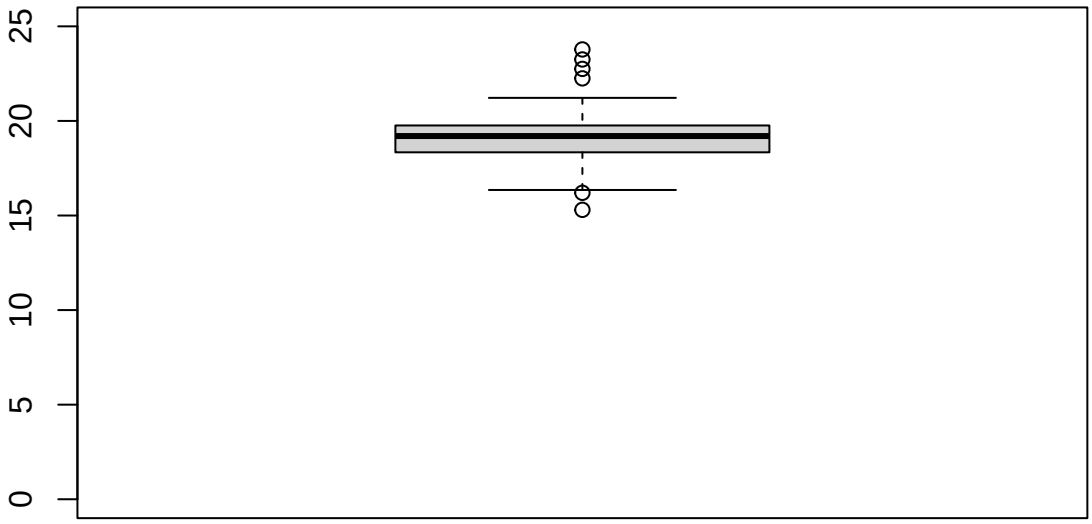


#56 data set:

```
data_set56 <- read.csv("C:/Users/rache/Documents/STAT 3010/Hw1_Q56_data.csv")
```

boxplot

```
boxplot(data_set56, xlab= "", ylab = "", ylim= c(0,25))
```



58

a.)

Machine 1 has a small data variation but machine 2 has a high data variation. Machine 1 also has an outlier but machine 2 does not.