Due: 2022/02/20

1. The code of this Homework:

```r
# Read the txt file
customer <- scan("customers.txt", skip=1)

# Problem 1-7
cat("1. The 5th element is ", customer[5], '\n')
customer_sorted <- sort(customer, decreasing = FALSE) # sort increasing
cat("2. The fifth lowest age is ", customer_sorted[5], '\n')
cat("3. Extracting the five lowest ages together: ", customer_sorted[1:5], '\n')
customer_sorted <- rev(customer_sorted) # decreasingly
cat("4. The five highest ages: ", customer_sorted[1:5], '\n')
cat("5. The average (mean) age: ", mean(customer), '\n')
cat("6. The standard deviation of ages: ", sd(customer), '\n')
age_diff <- customer - mean(customer) # problem 7
cat("8. The average of age_diff is ", mean(age_diff))

# Problem 9a
hist(customer, xlab = "age", breaks = 5)

# Problem 9b
den <- density(customer)
plot(den, frame = FALSE, col = "blue", main = "Density of Customer")

# Problem 9c
boxplot(customer, horizontal=TRUE)
stripchart(customer, method="stack", add=TRUE)
```

Use the command `source("HW1.R")` in console of RStudio to run the code.

2. After running the R script, the result shows:

```
> setwd("HW1")
> source("HW1.R")
Read 399 items
1. The 5th element is  45
2. The fifth lowest age is  19
3. Extracting the five lowest ages together:  18 19 19 19 19
4. The five highest ages:  85 83 82 82 81
5. The average (mean) age:  46.80702
6. The standard deviation of ages:  16.3698
8. The average of age_diff is  -1.623275e-15
>
```
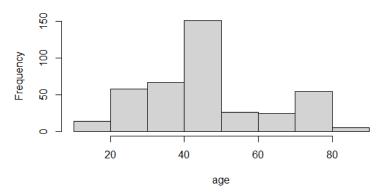
The answer of problem 1 8 is shown. Particularly, we should note the the answer of problem 8 is, theoretically, 0. That's because

$$\sum_{i=1}^{n} \frac{1}{n}(X_i - \bar{X}) = \frac{1}{n}(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X}) = \frac{1}{n}(n\bar{X} - n\bar{X}) = 0$$

However, constraint to the precision of floating-point (when computing mean), the answer of 8 may be a very small positive number.

For problem 9, the diagrams are shown below:

Histogram of customer

Density of Customer

N = 399   Bandwidth = 3.751