HW1_Report

Student ID: 107023058

Pre-process data

```
raw_data = read.table("D:/辰浩/NTHU/課程講義/計算統計於商業分析之應用/HW1/customers.txt",header = T,sep="")
# read data from the given file
ages <- raw_data$age # convert raw_data into list
```

Question 1

- Description: 5th element in the original list
- Solution:

```
ages[5] # the 5th age
```

```
## [1] 45
```

Question 2

- Description: fifth lowest age
- solution:

```
sorted_ages <- sort(ages) # sorting the original data
sorted_ages[5] # the 5th age in the sorted_ages is the answer
```

```
## [1] 19
```

Question 3

- Description: extract five lowest age
- Solution:

```
extract_list <- c(sorted_ages[1:5]) # extract the five lowest age into a list
extract_list
```

```
## [1] 18 19 19 19 19
```

Question 4

- Description: five highest age by sorting in decreasing
- Solution:

```
sorted_ages_d <- sort(ages,decreasing = TRUE) # sorting the original data in increasing
sorted_ages_d[1:5] # 5 highest ages
```

```
## [1] 85 83 82 82 81
```

Question 5

- Description: find average of age
- Solution:

```
average_age <- mean(ages) # find average by the mean() function
average_age
```

```
## [1] 46.80702
```

Question 6

- Description: find the standard deviation of age
- Solution:

```
sd_data <- sd(ages) # find the standard deviation by the sd() function
sd_data
```

```
## [1] 16.3698
```

Question 7

- Description: find the difference between each age and mean
- Solution:

```
age_diff <- ages - mean(ages) # subtracting the mean to each element in
 the data set
age_diff[1:75]
```

```
##  [1]   2.1929825  22.1929825  -5.8070175  26.1929825  -1.8070175  24.
1929825
##  [7]   3.1929825  -3.8070175  23.1929825 -14.8070175   0.1929825  30.
1929825
## [13]  17.1929825   3.1929825   3.1929825  -1.8070175   2.1929825   0.
1929825
## [19]  15.1929825   3.1929825   0.1929825  25.1929825   0.1929825  16.
1929825
## [25] -25.8070175   2.1929825   3.1929825   1.1929825 -11.8070175  30.
1929825
## [31]   1.1929825   1.1929825   3.1929825   0.1929825 -17.8070175  -4.
8070175
## [37]  -4.8070175  38.1929825  -1.8070175   2.1929825  -1.8070175  -3.
8070175
## [43]   2.1929825  21.1929825  -4.8070175   1.1929825  25.1929825  32.
1929825
## [49]   1.1929825   3.1929825   0.1929825  -1.8070175 -16.8070175  29.
1929825
## [55] -15.8070175   2.1929825  27.1929825  25.1929825   1.1929825   2.
1929825
## [61]  26.1929825   3.1929825   0.1929825   0.1929825  36.1929825  25.
1929825
## [67]  28.1929825   3.1929825   3.1929825   2.1929825   1.1929825  -1.
8070175
## [73]   2.1929825   2.1929825   2.1929825
```

Question 8

- Description: find average of the result of question 8
- Solution:

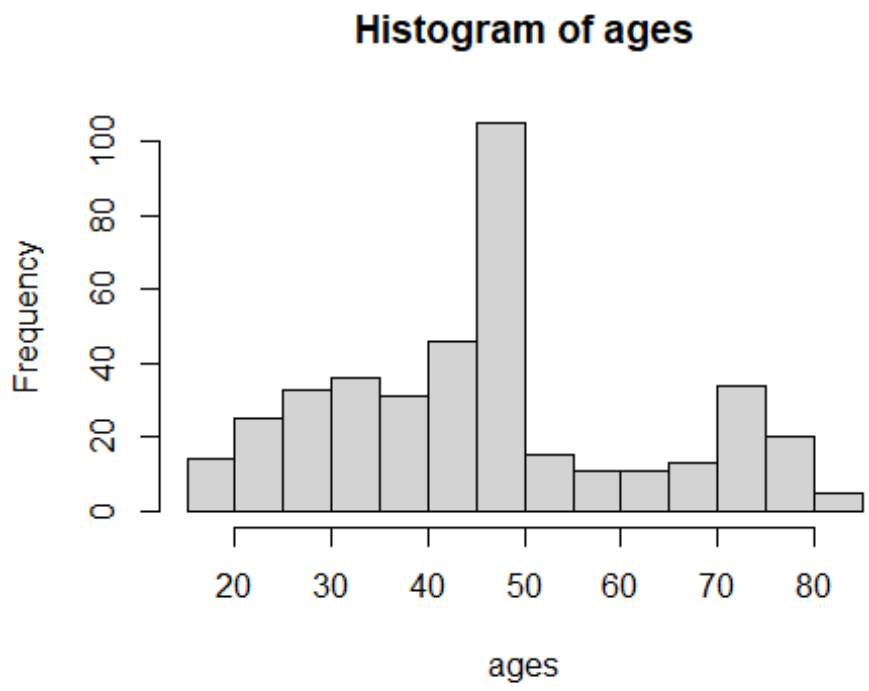
```
average_diff <- mean(age_diff) # find the average by the mean() functio
n
average_diff

## [1] -1.623275e-15
```

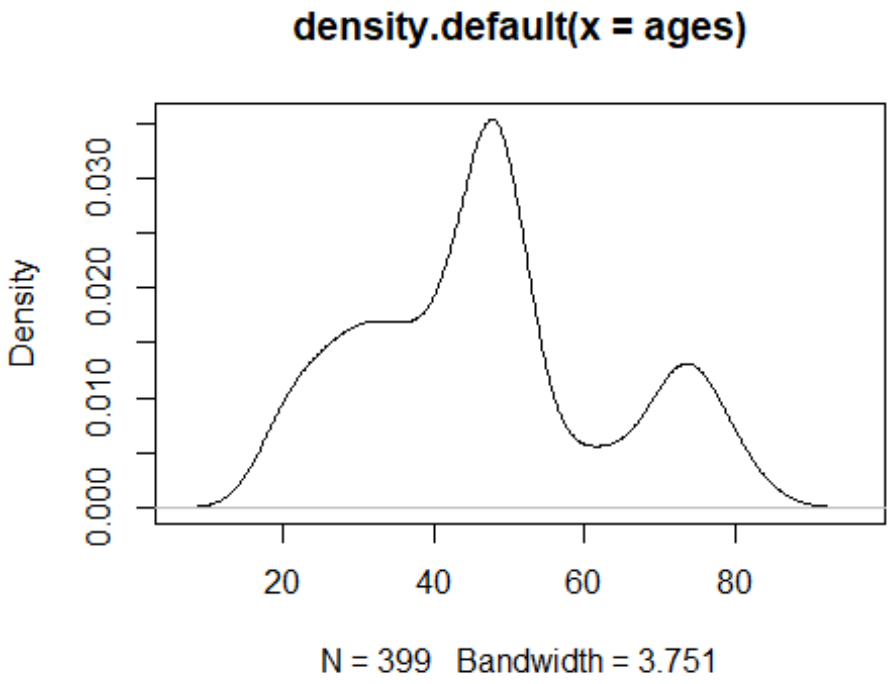
Question 9

- Description: visualize the raw data with three graphs
- Solution:

```
# a. histogram
hist(ages)
```




```
# b. density plot
plot(density(ages))
```




```
# c. boxplot + stripchart
boxplot(ages)
stripchart(ages,add = TRUE)
```